Linear Algebra: An Interactive Introduction

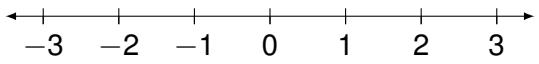
June 5, 2026

Contents

1 Preliminaries

2 A Brief Introduction to $\mathbb{R}^n$

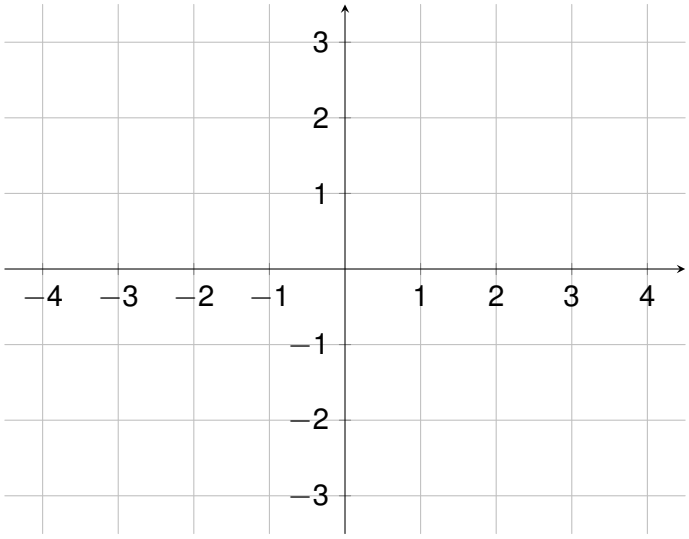
The set of all real numbers is denoted by $\mathbb{R}$. It is convenient to associate real numbers with points on a line, called the *real number line*.



The set of all ordered pairs (x, y), where x and y are real numbers, is called $\mathbb{R}^2$. Using set notation we write:

$$\mathbb{R}^2 = \{(x, y) : x, y \in \mathbb{R}\}$$

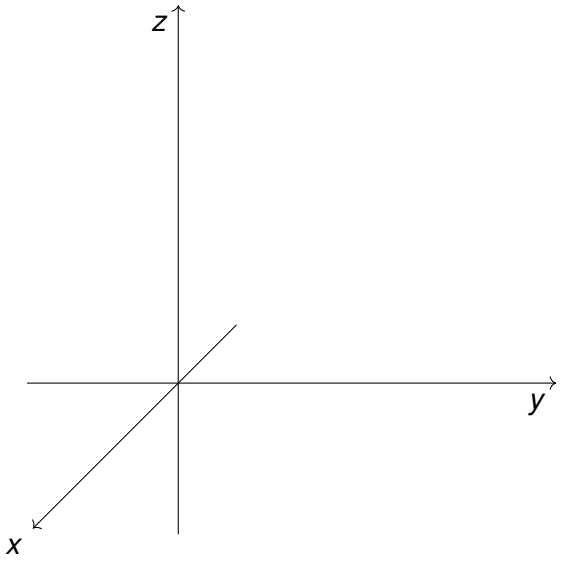
Geometrically speaking, $\mathbb{R}^2$ can be associated with a coordinate plane in which we refer to each point by its x and y coordinates.



The set of all ordered triples (x, y, z), where x, y and z are real numbers, is called $\mathbb{R}^3$.

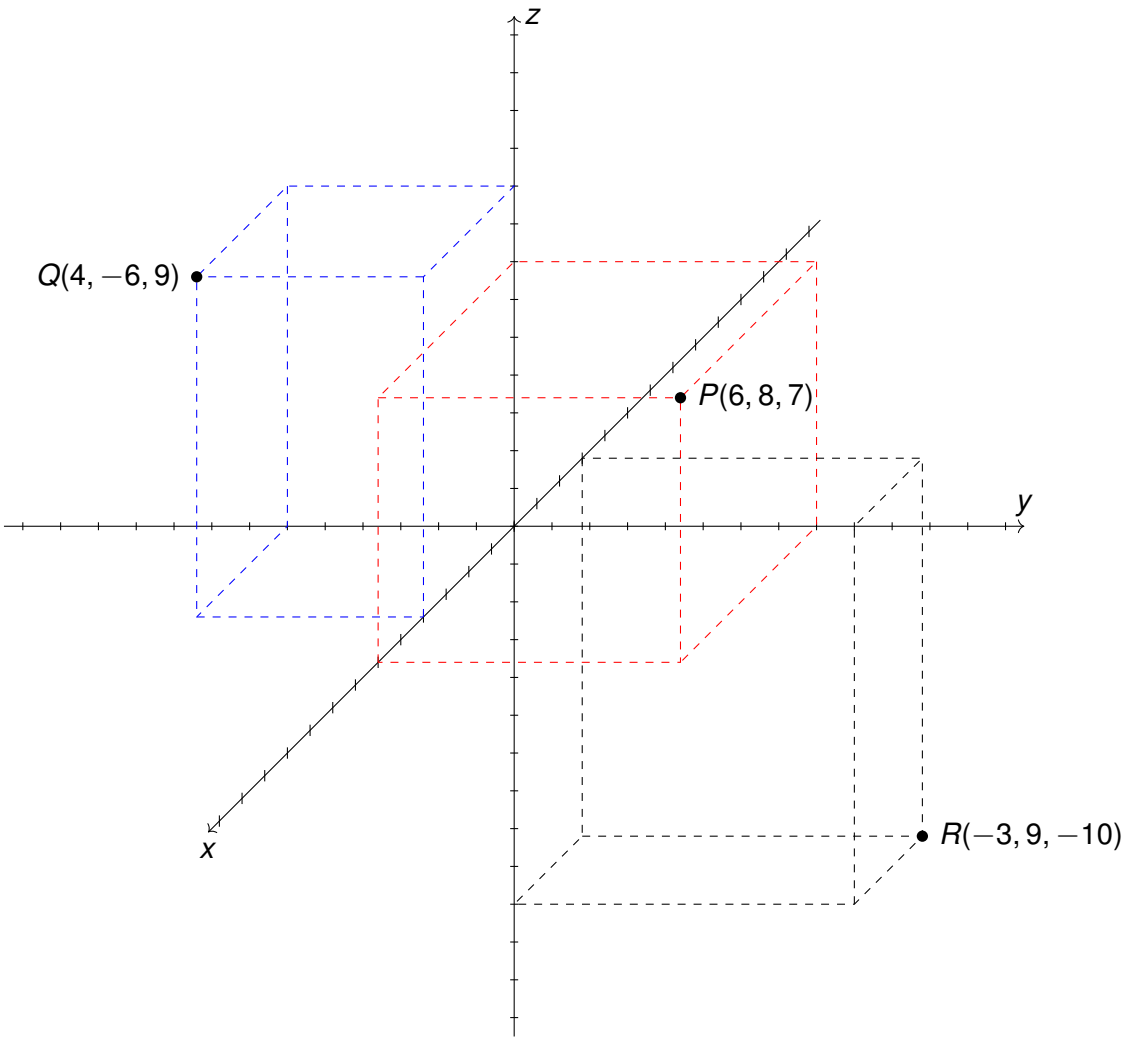
$$\mathbb{R}^3 = \{(x, y, z) : x, y, z \in \mathbb{R}\}$$

Geometrically, an ordered triple of $\mathbb{R}^3$ is associated with a point of a three-dimensional space whose position is given by x, y and z coordinates.

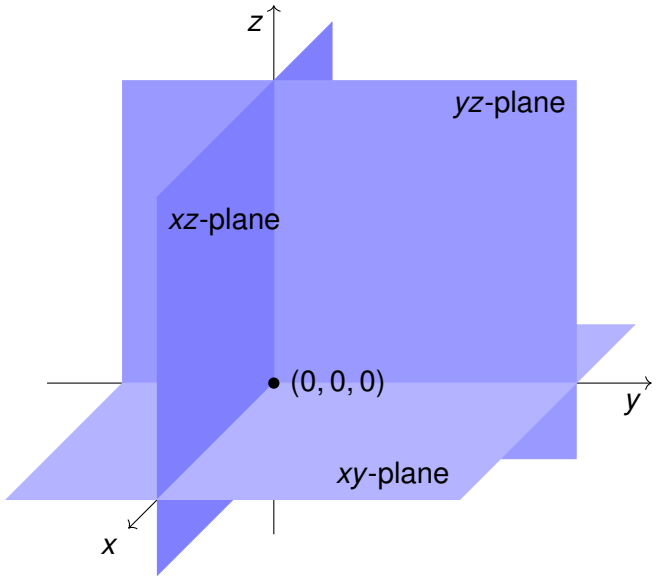


Example 1. The following points are shown plotted in $\mathbb{R}^3$.

1. $P(6, 8, 7)$
2. $Q(4, -6, 9)$
3. $R(-3, 9, -10)$



Each pair of axes in $\mathbb{R}^3$ determines a plane. The resulting three planes are called *coordinate planes*. Each coordinate plane is named after the axes that determine it. Thus, we have the *xy*-plane, *xz*-plane, and *yz*-plane. Coordinate planes intersect at the point $(0, 0, 0)$, called the *origin*, and subdivide $\mathbb{R}^3$ into eight regions, called *octants*.



The set of all ordered n-tuples $(x_1, x_2, \ldots, x_n)$, where x_i is a real number for $1 \leq i \leq n$, is called $\mathbb{R}^n$.

$$\mathbb{R}^n = \{(x_1, x_2, \ldots, x_n) | x_i \in \mathbb{R} \text{ for } 1 \leq i \leq n\}$$

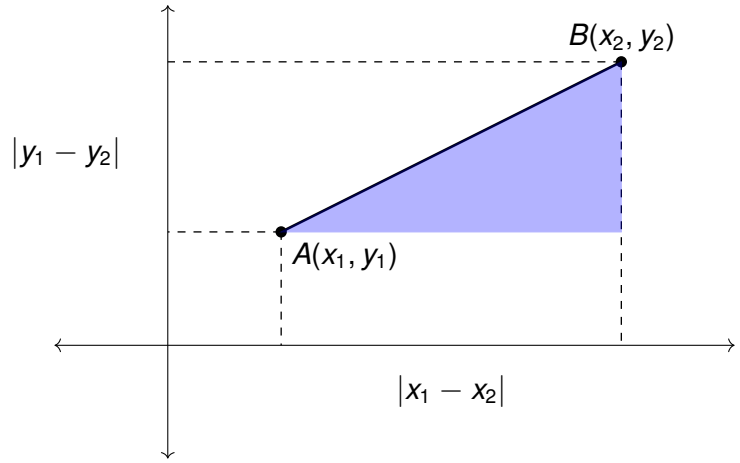
The point $(0, 0, \ldots, 0)$ in $\mathbb{R}^n$ is called the *origin*.

$\mathbb{R}^n$ cannot be visualized for $n > 3$, but many familiar ideas, such as the distance formula, can be generalized to $\mathbb{R}^n$.

Distance in $\mathbb{R}^n$

In this section we will establish a formula for the distance between two points in $\mathbb{R}^n$. We begin by observing that the distance between two numbers (points) x_1 and x_2 on the number line is given by $|x_1 - x_2|$. (Why do we use the absolute value brackets?).

We can use the Pythagorean Theorem to establish the distance formula for points of $\mathbb{R}^2$.



Let $A(x_1, y_1)$ and $B(x_2, y_2)$ be points in $\mathbb{R}^2$. By the Pythagorean Theorem we have

$$AB^2 = (x_1 - x_2)^2 + (y_1 - y_2)^2$$

$$AB = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$

Why were we able to drop the absolute value brackets?

The distance formula for points in $\mathbb{R}^3$ can also be derived using the Pythagorean Theorem. Let $A(x_1, y_1, z_1)$ and $B(x_2, y_2, z_2)$ be points of $\mathbb{R}^3$. Use the following GeoGebra interactive to walk through the steps of the derivation of the distance formula. RIGHT-CLICK and DRAG to rotate the image.

Access GeoGebra interactives through the online version of this text at

https://ximera.osu.edu/linearalgebradzv3/LinearAlgebraInteractiveIntro.

Let points $A'(x_1, y_1, 0)$ and $B'(x_2, y_2, 0)$ be projections of A and B onto the xy-plane. By the distance formula in $\mathbb{R}^2$, the distance between A' and B' is

$$A'B' = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$

Let $C = (x_2, y_2, z_1)$. Observe that $\triangle ABC$ is a right triangle with $AC = A'B'$, and $BC = |z_1 - z_2|$. By the Pythagorean theorem we have

$$AB^2 = AC^2 + BC^2 = \left(\sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}\right)^2 + (z_1 - z_2)^2 = (x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2$$

This gives us the following formula.

Formula 1. Let $A(x_1, y_1, z_1)$ and $B(x_2, y_2, z_2)$ be points in $\mathbb{R}^3$. The distance between A and B is given by

$$AB = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2}$$

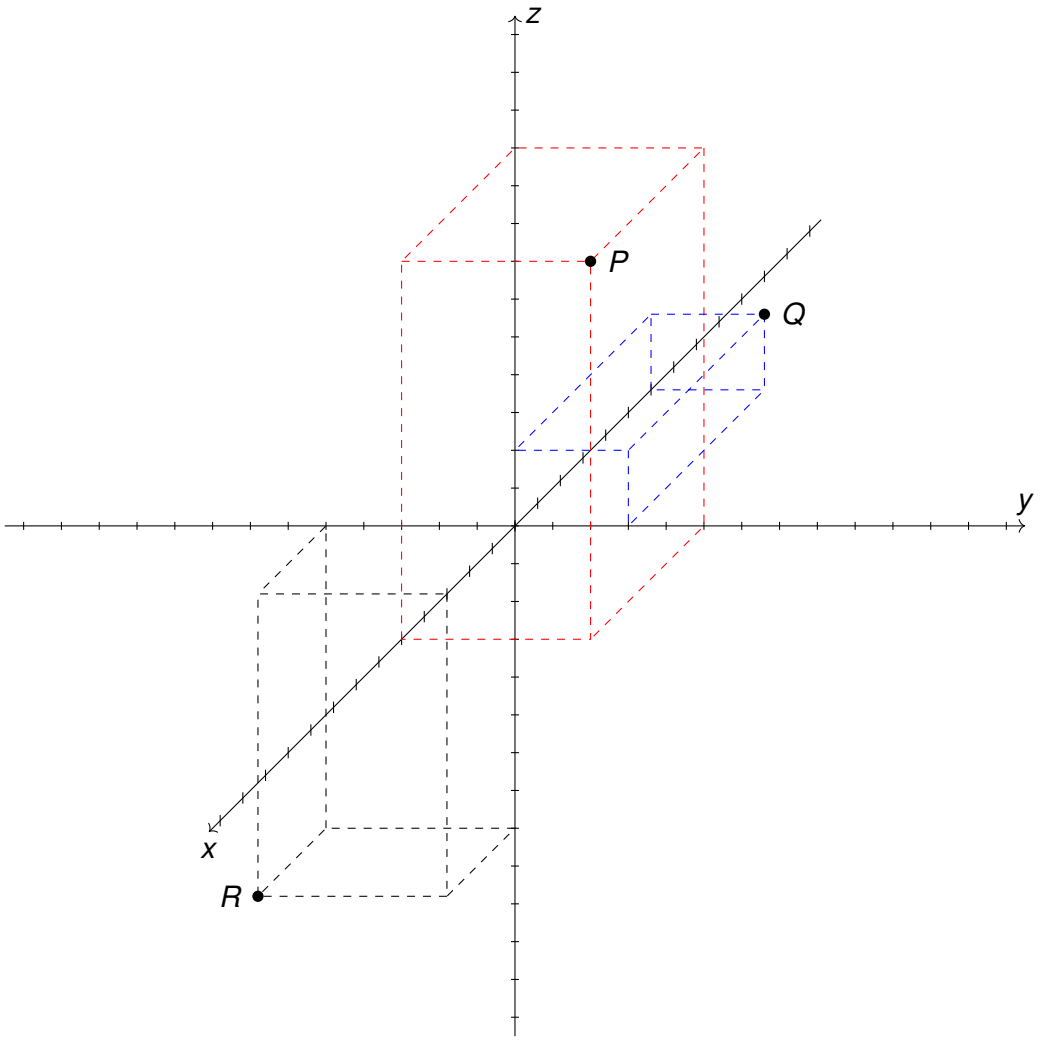
Observe the similarity of pattern in the distance formulas for $\mathbb{R}^1$, $\mathbb{R}^2$ and $\mathbb{R}^3$. We will take advantage of this pattern to define the distance between two points of $\mathbb{R}^n$.

Formula 2. Let $A(a_1, a_2, \ldots, a_n)$ and $B(b_1, b_2, \ldots, b_n)$ be points in $\mathbb{R}^n$. The distance between A and B is given by

$$AB = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2 + \ldots + (a_n - b_n)^2}$$

Practice Problems

Problem 1 *Find the coordinates of each point.*



Answer:

$P(\ldots\ldots, \ldots\ldots, \ldots\ldots)$

$Q(\ldots\ldots, \ldots\ldots, \ldots\ldots)$

$R(\ldots\ldots, \ldots\ldots, \ldots\ldots)$

Problem 2 *Find the coordinates of each point. RIGHT-CLICK and DRAG to rotate the image. Access GeoGebra interactives through the online version of this text at https://ximera.osu.edu/linearalgebradzv3/LinearAlgebraInteractiveIntro.*

Answer:

$A(\ldots\ldots, \ldots\ldots, \ldots\ldots)$

$B(\ldots\ldots, \ldots\ldots, \ldots\ldots)$

$C(\ldots\ldots, \ldots\ldots, \ldots\ldots)$

$D(\ldots\ldots, \ldots\ldots, \ldots\ldots)$

Problem 3 *Find the distance between* $A(-2, -1, 4)$ *and* $B(1, -5, -8)$*.*

$AB = \ldots\ldots$

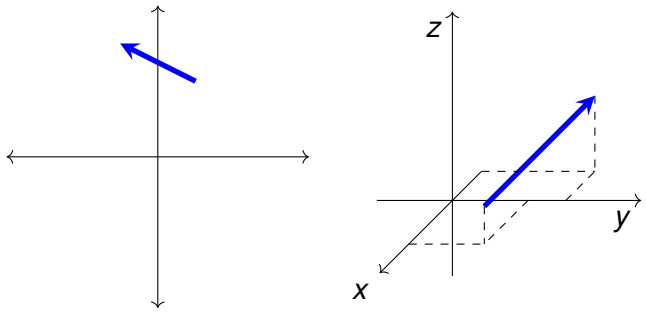
3 Introduction to Vectors

What is a Vector?

A *scalar* is a quantity that has size, often called magnitude, but no direction. For example, temperature, mass and speed are scalars. In this course, scalars will typically be real numbers, but we will also see complex numbers on a few occasions.

A *vector* has magnitude and direction. For example, velocity is a vector because it tells us how fast the object is traveling and also the direction of travel.

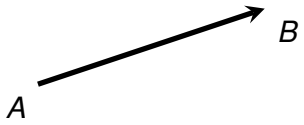
If an object is traveling along a number line, the direction of travel is given by the sign of its velocity (positive or negative), while the speed is given by the absolute value of the velocity. If the object is traveling in a plane or in space, direction of travel can be described by an arrow, while the speed can be represented by the length of the arrow. Graphically speaking, vectors in $\mathbb{R}^2$ and $\mathbb{R}^3$ look like this:



A vector can be denoted by a lower-case letter with an arrow over the top (like this: $\overrightarrow{u}$), or a bold lower-case letter (like this: **u**).

The magnitude, or length, of a vector is denoted by double absolute value brackets. For example, the magnitude of **u**, is denoted by $\|\mathbf{u}\|$. A vector of zero length and no direction is called the *zero vector.* We denote the zero vector by $\overrightarrow{0}$ or **0**. Going forward, we will use the terms magnitude of a vector and length of a vector interchangeably.

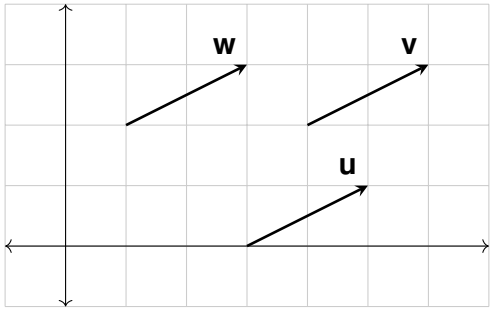
Sometimes it is convenient to refer to a vector by naming the endpoints of the arrow. In the figure below, point *A* is the *tail*, and point *B* is the *head* of the vector.



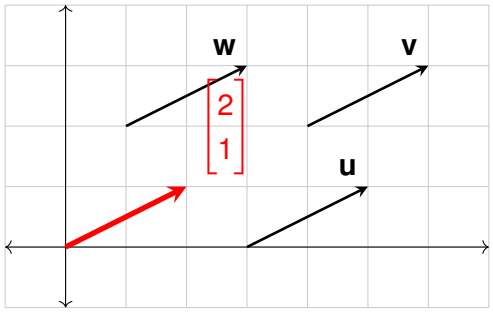
We refer to this vector as $\overrightarrow{AB}$.

Vectors in Standard Position

Vectors that point in the same direction and have the same length are said to be *equivalent*. For example, vectors **u**, **v** and **w** in the figure below are equivalent. We write **u** = **v** = **w**.



For the purpose of developing standard, convenient notation, we observe that every vector is equivalent to some vector whose tail is at the origin. Vectors with tails at the origin are said to be in *standard position*. We will refer to each vector in standard position by the coordinates of its head. For example, a vector in standard position whose head is located at the point $(2, 1)$ will be referred to as $\begin{bmatrix} 2 \\ 1 \end{bmatrix}$.



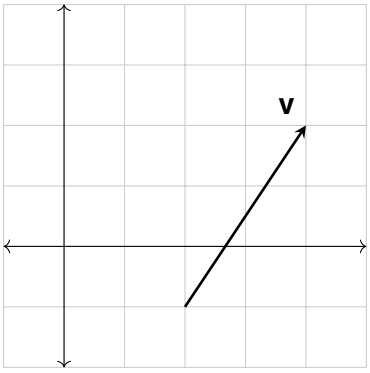
Vectors **u**, **v** and **w** in the figure are equivalent to vector $\begin{bmatrix} 2 \\ 1 \end{bmatrix}$. We write **u** = **v** = **w** = $\begin{bmatrix} 2 \\ 1 \end{bmatrix}$. Number 2 is called the *first component* of the vector (or the x-component) while number 1 is the *second component* (or the y-component). The form $\begin{bmatrix} 2 \\ 1 \end{bmatrix}$ is called the *component form*.

Vector $\begin{bmatrix} 2 \\ 1 \end{bmatrix}$ is an example of a *column vector*. Occasionally, we will find that representing this vector as a *row vector* [2, 1] is more convenient.

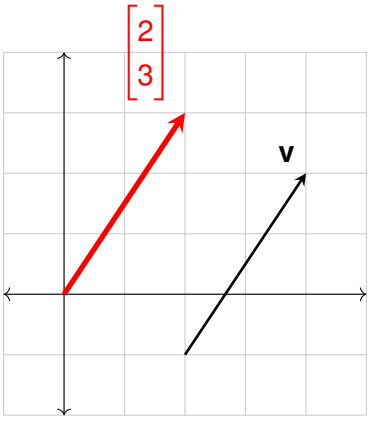
Column (or row) representation of vectors in component form allows us to go beyond the physical and geometric definition, and think of vectors more abstractly as arrays of numbers.

Our next goal is to find a process for writing any vector in the coordinate plane in component form.

Example 2. Consider vector **v** shown below. We will express **v** in component form.



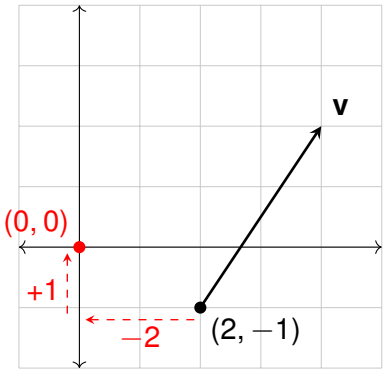
Explanation. Note that the vector has a "run" of 2 and a "rise" of 3. If we construct a vector with tail at the origin, a "run" of 2 and a "rise" of 3, we will have a vector in standard position equivalent to vector **v**.



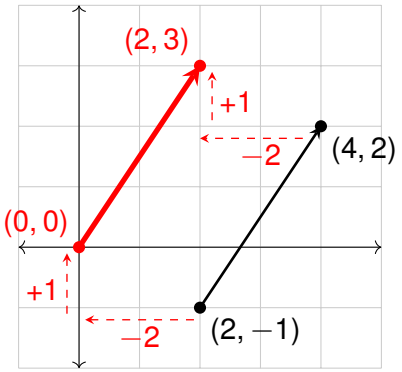
The component form for the vector we constructed is $\begin{bmatrix} 2 \\ 3 \end{bmatrix}$. This gives us $\mathbf{v} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$.

The approach we used in Example **??** is applicable to specific vectors that can easily be visualized. What we need now is an algebraic approach that can be generalized to higher dimensions and more abstract situations.

Let's return to vector **v** of Example **??**. Suppose we were to slide vector **v** into standard position. Consider what would happen to the tail of **v** as we do so.



What happens to the tail of the vector has to happen to the head

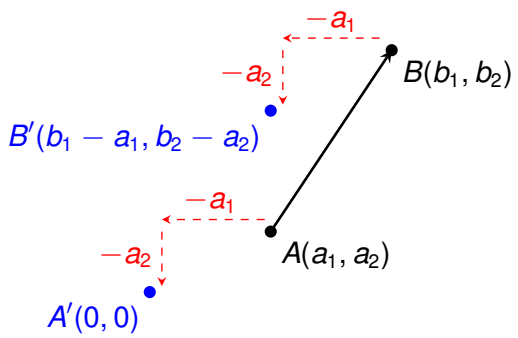


We subtracted 2 from the x-coordinate and added 1 to the y-coordinate of the tail. To find the new location of the head we subtract 2 from the x-coordinate of the head, and add 1 to the y-coordinate of the head. This gives us $(4-2, 2+1)$. So, the new location of the head is $(2, 3)$, and $\mathbf{v} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$.

If you look back at what we did you will find that the components of **v** were computed by subtracting the coordinates of the tail from the coordinates of the head

$$\mathbf{v} = \begin{bmatrix} 4-2 \\ 2-(-1) \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$$

The following diagram summarizes and generalizes our findings.



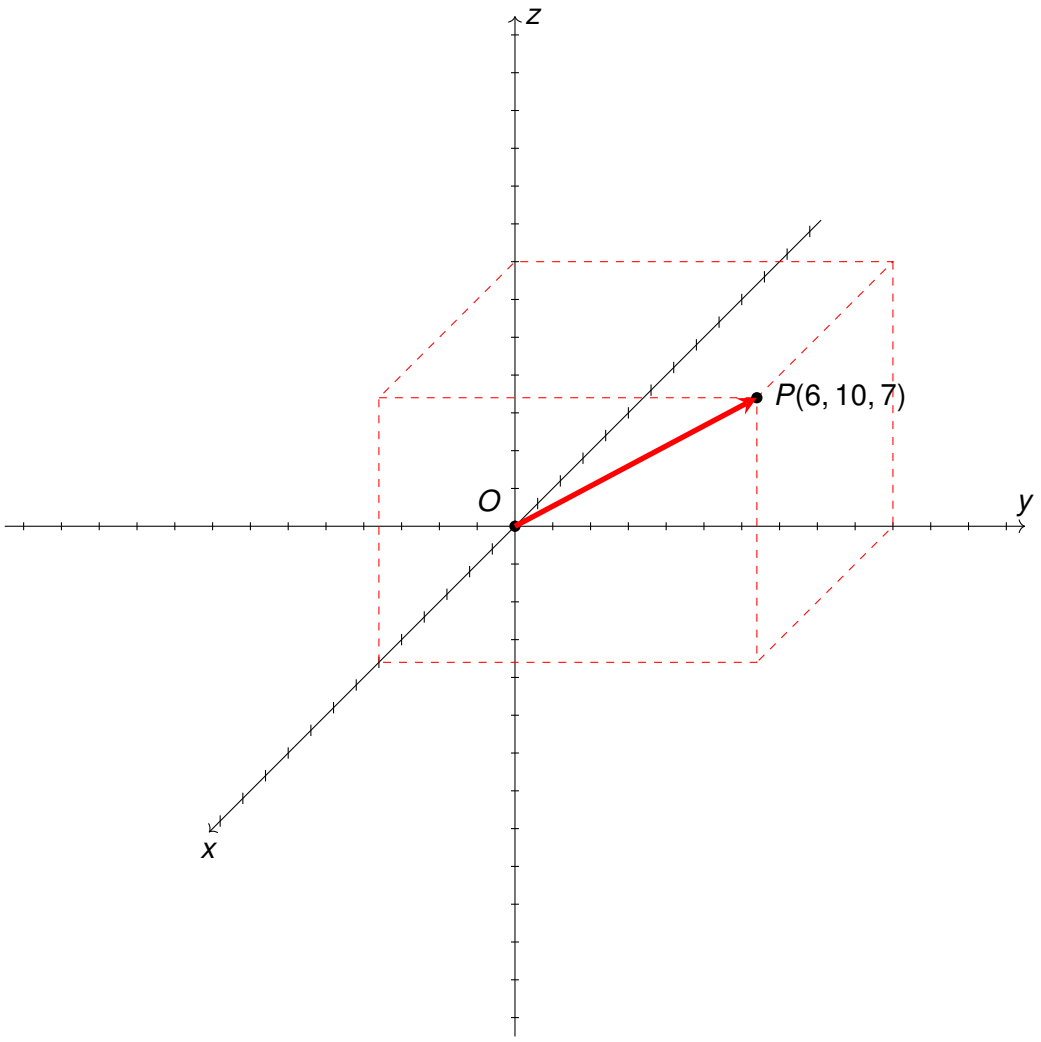
Let $\overrightarrow{AB}$ be a vector in $\mathbb{R}^2$, with tail at point $A(a_1, a_2)$ and head at point $B(b_1, b_2)$. As we slide $\overrightarrow{AB}$ into standard position by moving point A to the origin, point B travels along with point A by undergoing the same horizontal and vertical shifts. We now have an equivalent vector $\overrightarrow{A'B'}$ in standard position. The diagram suggests the following formula.

Formula 3 ("Head - Tail" Formula in $\mathbb{R}^2$)**.** Suppose a vector's tail is at point $A(a_1, a_2)$ and the vector's head is at $B(b_1, b_2)$, then

$$\overrightarrow{AB} = \begin{bmatrix} b_1 - a_1 \\ b_2 - a_2 \end{bmatrix} \tag{1}$$

Vectors in $\mathbb{R}^3$

Definitions of standard position and component form for vectors in $\mathbb{R}^3$ are analogous to their counterparts for vectors in $\mathbb{R}^2$. For example, vector $\overrightarrow{OP}$ in the figure below, is in standard position and can be written in component form as $\overrightarrow{OP} = \begin{bmatrix} 6 \\ 10 \\ 7 \end{bmatrix}$.



If a vector is not in standard position but the location of its head and tail are known, a three-dimensional version of the "Head - Tail" formula can be used to express the vector in component form.

Formula 4 ("Head - Tail" Formula in $\mathbb{R}^3$)**.** Suppose a vector's tail is at point $A(a_1, a_2, a_3)$ and the vector's head is at $B(b_1, b_2, b_3)$, then

$$\overrightarrow{AB} = \begin{bmatrix} b_1 - a_1 \\ b_2 - a_2 \\ b_3 - a_3 \end{bmatrix} \tag{2}$$

Vectors in $\mathbb{R}^n$

We cannot see $\mathbb{R}^n$ for $n > 3$, but we can conceptualize it by generalizing what we know about $\mathbb{R}^2$ and $\mathbb{R}^3$. A vector $\mathbf{v}$ in standard position whose head is located at $(v_1, v_2, \ldots, v_n)$ can be written in component form as $\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$.

Recall that we defined the *zero vector* as a vector that has length 0 and no direction. In component form, the zero vector is a vector all of whose components are 0.

We conclude this section by stating the generalized "Head - Tail" formula.

Formula 5 ("Head - Tail" Formula in $\mathbb{R}^n$)**.** Suppose a vector's tail is at point $A(a_1, a_2, \ldots, a_n)$ and the vector's head is at $B(b_1, b_2, \ldots, b_n)$, then

$$\overrightarrow{AB} = \begin{bmatrix} b_1 - a_1 \\ b_2 - a_2 \\ \vdots \\ b_n - a_n \end{bmatrix} \tag{3}$$

Practice Problems

Problem 4 *Express each vector in component form.*

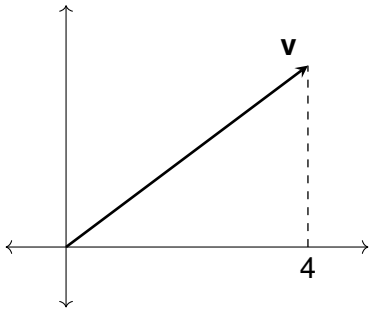
1. *Vector* $\mathbf{v}$ *has a tail at* $(-2, 4)$ *and a head at* $(4, 5)$*.*

$$\mathbf{v} = \begin{bmatrix} \ldots \ldots \\ \ldots \ldots \end{bmatrix}$$

2. *Vector* $\mathbf{p}$ *has a tail at* $(2, 4, 5)$ *and a head at* $(-2, 4, 8)$*.*

$$\mathbf{p} = \begin{bmatrix} \ldots \ldots \\ \ldots \ldots \\ \ldots \ldots \end{bmatrix}$$

Problem 5 *Express vector* $\mathbf{v}$ *shown in the figure in component form if the magnitude (length) of* $\mathbf{v}$ *is* 5*.*



Answer:

$$\mathbf{v} = \begin{bmatrix} \ldots\ \ldots \\ \ldots\ \ldots \end{bmatrix}$$

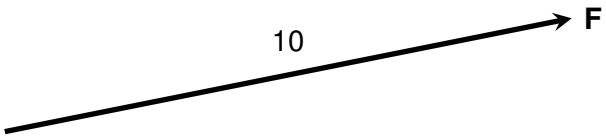
Problem 6 *Let* $A(-2,3)$*,* $B(3,1)$ *and* $C(-1,-1)$*. Find point D such that* $\overrightarrow{AB} = \overrightarrow{CD}$*.*

Answer:

$$D(\ldots\ \ldots\ ,\ \ldots\ \ldots)$$

4 Length of a Vector

Vector quantities, such as velocity and force, have magnitude and direction. The magnitude of a vector quantity is the length of the vector. For example, if a force of 10 Newtons is applied to an object, we would represent the force by a 10-unit-long vector.



The magnitude of a vector is denoted by double absolute value brackets. In the case of force **F**, we write

$$\|\mathbf{F}\| = 10$$

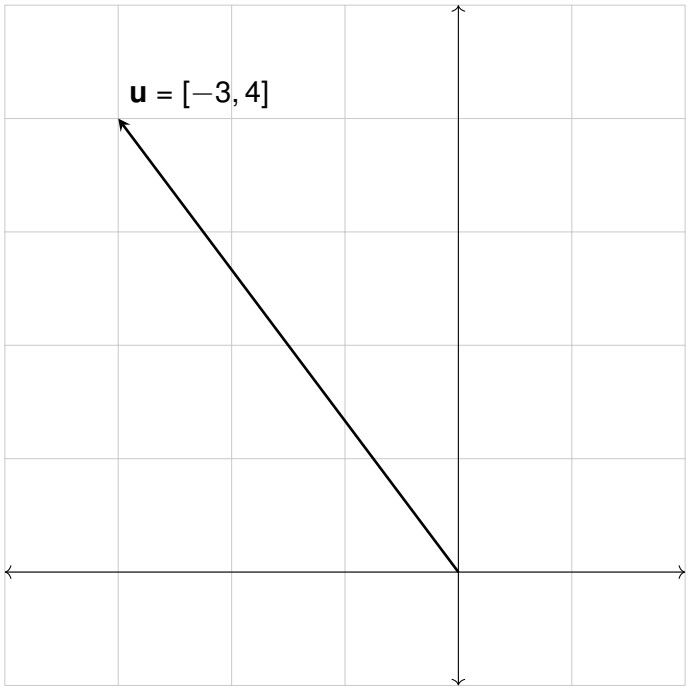
To find the length of a vector, we need to find the distance between the tail of the vector and its head. Recall that in $\mathbb{R}^2$, the distance between $A(a_1, a_2)$ and $B(b_1, b_2)$ is given by

$$d_{AB} = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2}$$

A vector $\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$ has the length of the vector in standard position with its head at (v_1, v_2) and tail at $(0, 0)$. We find the length of **v** using the distance formula

$$\|\mathbf{v}\| = \sqrt{(v_1 - 0)^2 + (v_2 - 0)^2} = \sqrt{v_1^2 + v_2^2} \tag{4}$$

Example 3. Find the magnitude of $\mathbf{u} = \begin{bmatrix} -3 \\ 4 \end{bmatrix}$.



Explanation.

$$\|\mathbf{u}\| = \sqrt{(-3)^2 + (4)^2} = 5$$

The distance formula for points in $\mathbb{R}^3$ is analogous to the distance formula in $\mathbb{R}^2$. Given two points $A(a_1, a_2, a_3)$ and $B(b_1, b_2, b_3)$, the distance between them is given by

$$d_{AB} = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2 + (a_3 - b_3)^2}$$

To find the length of vector $\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}$, we find the distance between (v_1, v_2, v_3) and $(0, 0, 0)$.

$$\|\mathbf{v}\| = \sqrt{(v_1 - 0)^2 + (v_2 - 0)^2 + (v_3 - 0)^2} = \sqrt{v_1^2 + v_2^2 + v_3^2} \tag{5}$$

Example 4. The magnitude of $\mathbf{v} = \begin{bmatrix} 6 \\ -2 \\ 3 \end{bmatrix}$ is given by

$$\|\mathbf{v}\| = \sqrt{(6)^2 + (-2)^2 + (3)^2} = 7$$

Distance formulas for $\mathbb{R}^2$ and $\mathbb{R}^3$ motivate the following definition of distance between two points in $\mathbb{R}^n$.

Definition 1. Let $A(a_1, a_2, \ldots, a_n)$ and $B(b_1, b_2, \ldots, b_n)$ be points in $\mathbb{R}^n$. The *distance* between A and B is defined to be

$$d_{AB} = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2 + \ldots + (a_n - b_n)^2}$$

The following definition follows directly from the distance formula for $\mathbb{R}^n$ in the same way that expressions (**??**) and (**??**) followed from distance formulas in $\mathbb{R}^2$ and $\mathbb{R}^3$.

Definition 2. Let $\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$ be a vector in $\mathbb{R}^n$, then the *length*, or the *magnitude*, of $\mathbf{v}$ is given by

$$\|\mathbf{v}\| = \sqrt{v_1^2 + v_2^2 + \ldots + v_n^2} \tag{6}$$

Example 5. The magnitude of $\mathbf{w} = \begin{bmatrix} 5 \\ -1 \\ -2 \\ 4 \end{bmatrix}$ is given by

$$\|\mathbf{w}\| = \sqrt{5^2 + (-1)^2 + (-2)^2 + 4^2} = \sqrt{46}$$

Practice Problems

Problems **??**-**??**

Find the length of the following vectors. Enter your answers in exact, simplified form. (e.g. $5\sqrt{3}$)

Problem 7

$$\mathbf{u} = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$$

Answer:

$$\|\mathbf{u}\| = \sqrt{\ldots\ldots}$$

Problem 8

$$\mathbf{v} = \begin{bmatrix} -4 \\ -6 \end{bmatrix}$$

Answer: Enter your answer in the most simplified exact form. (e.g. $a\sqrt{b}$)

$$\|\mathbf{v}\| = \ldots\ldots\sqrt{\ldots\ldots}$$

Problem 9

$$\mathbf{w} = \begin{bmatrix} 2 \\ -4 \\ 6 \end{bmatrix}$$

Answer: Enter your answer in the most simplified exact form. (e.g. $a\sqrt{b}$)

$$\|\mathbf{w}\| = \ldots\ldots\sqrt{\ldots\ldots}$$

Problem 10

$$\mathbf{q} = \begin{bmatrix} -10 \\ 4 \\ -7 \\ 5 \\ 3 \end{bmatrix}$$

Answer:

$$\|\mathbf{q}\| = \sqrt{\ldots\ldots}$$

Problem 11 *Find the component form of vector* $\mathbf{v}$ *in* $\mathbb{R}^2$ *if we know that* $\|\mathbf{v}\| = 15$*, the x component of* $\mathbf{v}$ *is* -9 *and the vector is located in the third quadrant.*

Answer:

$$\mathbf{v} = \begin{bmatrix} -9 \\ \ldots\ldots \end{bmatrix}$$

Problem 12 *For a vector in* $\mathbb{R}^2$ *with a length of 26 and the y component of 10, what are the possibilities for the x component? List the possibilities in an increasing order.*

Answer:

$\ldots\ldots$, *and* $\ldots\ldots$

Problem 13 *Let* $\mathbf{v} = \begin{bmatrix} 1 \\ y \\ 4 \end{bmatrix}$. *Find all possible values of* y *if* $\|\mathbf{v}\| = 9$. *List the possibilities in an increasing order.*

Answer:

..... , *and* .....

Problem 14 *For a vector in* $\mathbb{R}^4$*, what are the possibilities for the fourth component if the length of the vector is 14, and the* x*,* y *and* z *components are 1, 5 and 13, respectively? List the possibilities in an increasing order.*

Answer:

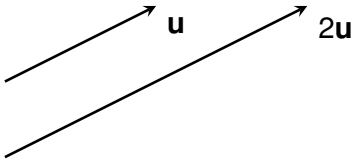
..... , *and* .....

Problem 15 *The points* $A(2, 1, 4)$, $B(4, 2, -1)$*, and* $C(6, 8, 7)$ *form a triangle in* $\mathbb{R}^3$*. Is it a right triangle?*

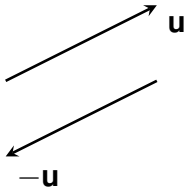
5 Vector Arithmetic

Geometry of Scalar Multiplication

The product of vector **u** with a positive scalar k, is a vector $k\mathbf{u}$ that points in the same direction as **u**, and whose length is equal to the length of **u** multiplied by k. For example, the figure below shows vectors **u** and $2\mathbf{u}$. The vectors point in the same direction but the magnitude of $2\mathbf{u}$ is twice the magnitude of **u**.



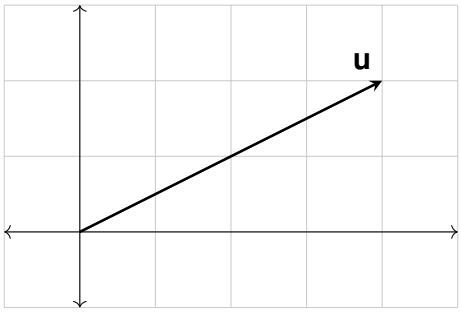
If a vector **u** is multiplied by -1, the resulting vector is denoted by $-\mathbf{u}$. It has the same length as vector **u**, but points in the opposite direction.



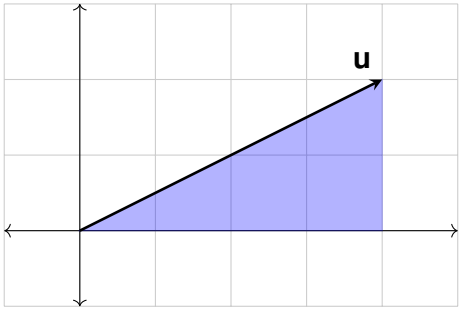
Algebra of Scalar Multiplication

We know what scalar multiplication accomplishes geometrically. Our goal now is to translate this idea to an algebraic operation.

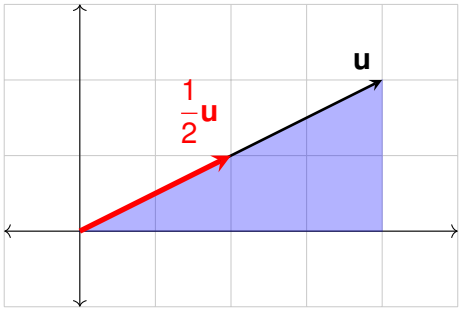
Exploration 16 *Consider vector* $\mathbf{u} = \begin{bmatrix} 4 \\ 2 \end{bmatrix}$. *We will find an algebraic approach for multiplying* **u** *by* $\frac{1}{2}$.



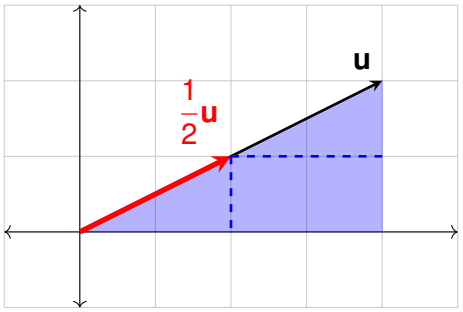
Consider **u** *to be the hypotenuse of a right triangle.*



The head of $\frac{1}{2}\mathbf{u}$ *should be the midpoint of the hypotenuse.*



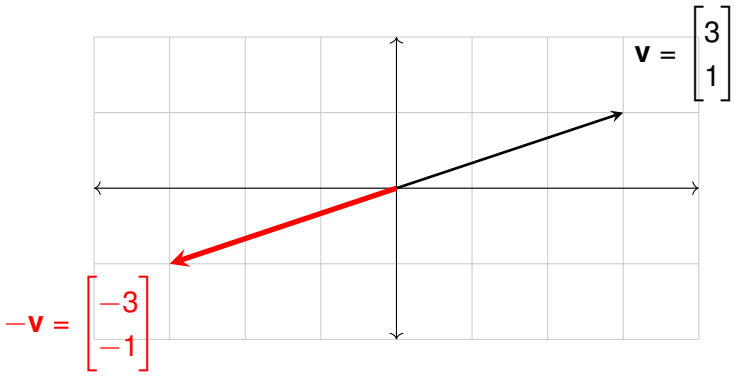
From our study of similar triangles in geometry, we know that if we drop perpendiculars from the midpoint of the hypotenuse to the two legs of the triangle, the perpendiculars will bisect the legs.



This tells us that to find x and y components of $\frac{1}{2}\mathbf{u}$ we must multiply each component of $\mathbf{u}$ *by* $\frac{1}{2}$.

$$\frac{1}{2}\mathbf{u} = (1/2)\begin{bmatrix}4\\2\end{bmatrix} = \begin{bmatrix}(1/2)(4)\\(1/2)(2)\end{bmatrix} = \begin{bmatrix}2\\1\end{bmatrix}$$

Exploration 17 *Consider vector* $\mathbf{v} = \begin{bmatrix}3\\1\end{bmatrix}$ *It is clear that multiplying the components of* $\mathbf{v}$ *by* -1 *reverses the direction of* $\mathbf{v}$ *while preserving its magnitude.*



Explorations **??** and **??** give rise to the following definition of scalar multiplication.

Definition 3. Let $\mathbf{u} = \begin{bmatrix}u_1\\u_2\\\vdots\\u_n\end{bmatrix}$ be a vector in $\mathbb{R}^n$, and let k be a scalar, then

$$k\mathbf{u} = k\begin{bmatrix}u_1\\u_2\\\vdots\\u_n\end{bmatrix} = \begin{bmatrix}ku_1\\ku_2\\\vdots\\ku_n\end{bmatrix}$$

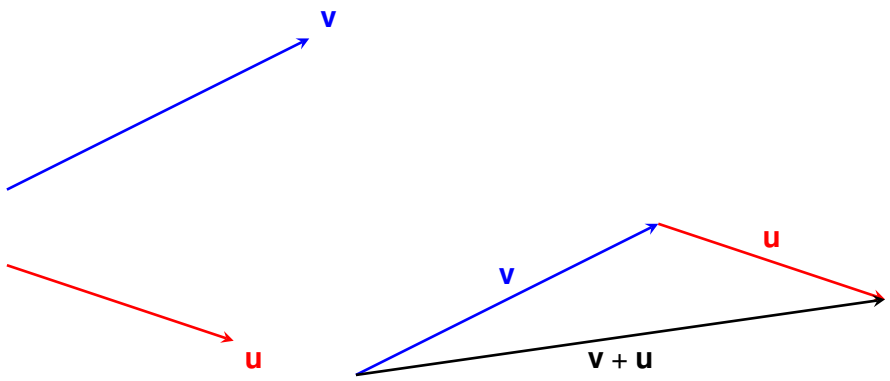
If $\mathbf{v} = k\mathbf{u}$ ($k \neq 0$), then $\mathbf{u} = \frac{1}{k}\mathbf{v}$, and we say that $\mathbf{v}$ and $\mathbf{u}$ are *scalar multiples of each other*.

Geometry of Vector Addition

There are two ways to add vectors geometrically.

"Head-to-Tail" Addition Method

Given vectors $\mathbf{v}$ and $\mathbf{u}$, we can find the sum $\mathbf{v} + \mathbf{u}$ by sliding $\mathbf{u}$ so as to place its tail at the head of vector $\mathbf{v}$. The vector connecting the tail of $\mathbf{v}$ with the head of $\mathbf{u}$ is the sum $\mathbf{v} + \mathbf{u}$, as shown in the figure below.



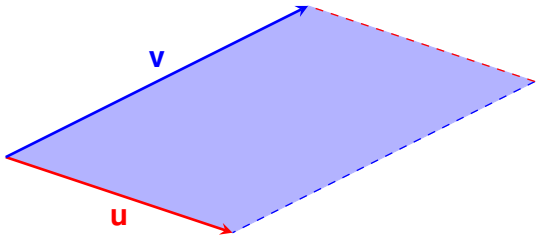
This sum can be interpreted as the total displacement that occurs when traveling along the two vectors starting at the tail of $\mathbf{v}$ and finishing at the head of $\mathbf{u}$.

Note that if we place the tail of $\mathbf{v}$ at the head of $\mathbf{u}$ instead, the sum vector $\mathbf{u} + \mathbf{v}$ will be the same as $\mathbf{v} + \mathbf{u}$. Thus, addition of vectors is commutative.

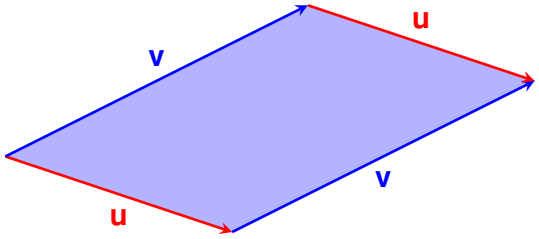
Parallelogram Addition Method

Most of the time we deal with vectors in standard position. So all vector tails are located at the origin. This motivates the *parallelogram method* for adding vectors.

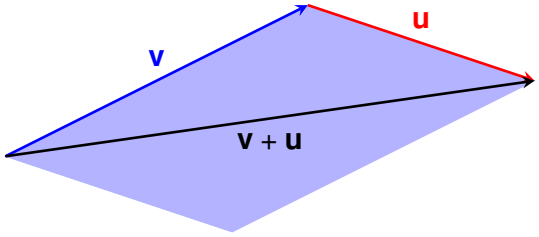
Observe that if we slide vectors $\mathbf{u}$ and $\mathbf{v}$ so that their tails are together, the two vectors determine a parallelogram.



Opposite sides of a parallelogram are congruent and parallel.



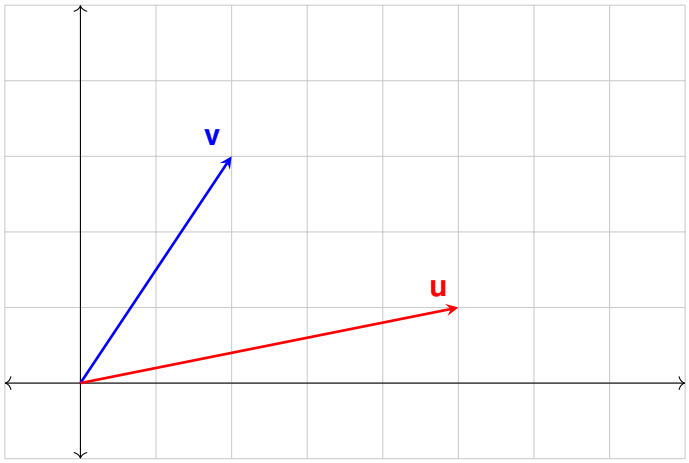
Applying the "head-to-tail" addition method shows that the sum **v** + **u** is the diagonal of the parallelogram determined by **v** and **u**.



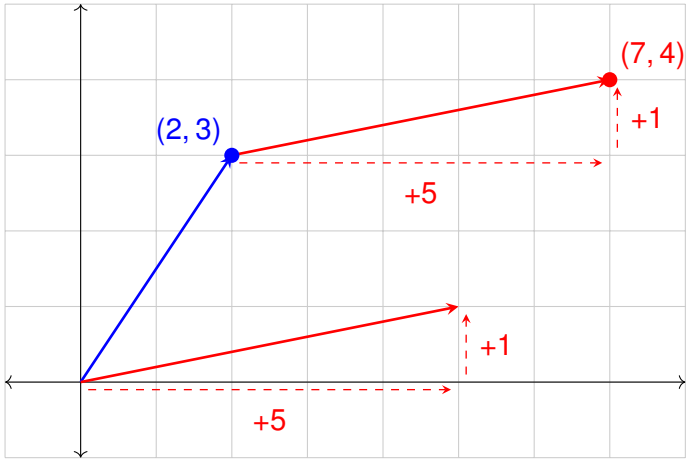
Algebra of Vector Addition

We now know how to add vectors geometrically. Our next goal is to translate this idea to an algebraic operation.

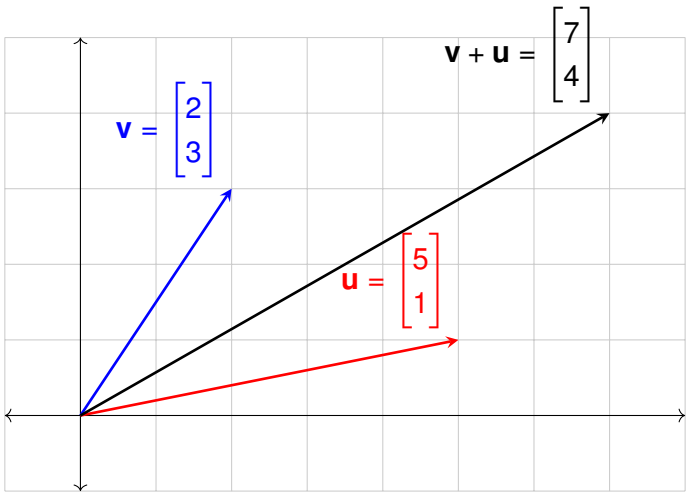
Exploration 18 *In this problem we will find the sum of* $\mathbf{u} = \begin{bmatrix} 5 \\ 1 \end{bmatrix}$ *and* $\mathbf{v} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$.



To use "head-to-tail" addition method, or to construct the side of a parallelogram opposite of $\mathbf{u}$*, we want to slide* $\mathbf{u}$ *so that its tail is at the point* $(2, 3)$*. Observe that* $\mathbf{u}$ *has a "run" of* 5 *and a "rise" of* 1*. If we start at* $(2, 3)$*, go over* 5 *then up* 1*, we will land on* $(7, 4)$*.*



The sum $\mathbf{u} + \mathbf{v}$ *is shown below.*



We see that the components of $\mathbf{v} + \mathbf{u}$ *can be found by adding the components of* $\mathbf{v}$ *and* $\mathbf{u}$*.*

$$\mathbf{v} + \mathbf{u} = \begin{bmatrix} 2 \\ 3 \end{bmatrix} + \begin{bmatrix} 5 \\ 1 \end{bmatrix} = \begin{bmatrix} 7 \\ 4 \end{bmatrix}$$

Exploration **??** motivates the following definition.

Definition 4. Let $\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$ be vectors in $\mathbb{R}^n$. We define $\mathbf{u} + \mathbf{v}$ by

$$\mathbf{u} + \mathbf{v} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} + \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = \begin{bmatrix} u_1 + v_1 \\ u_2 + v_2 \\ \vdots \\ u_n + v_n \end{bmatrix}$$

Geometry of Vector Addition in $\mathbb{R}^3$

Vectors in $\mathbb{R}^1$, $\mathbb{R}^2$, and $\mathbb{R}^3$ have the advantage in that we can gain insight into their behavior through visualization. Vectors in $\mathbb{R}^1$ and $\mathbb{R}^2$ are the easiest to visualize. Vectors in $\mathbb{R}^3$ are a little trickier. The following exploration will help you visualize addition of vectors in $\mathbb{R}^3$.

Exploration 19 *Adding two vectors amounts to finding the diagonal of a parallelogram determined by placing the two vectors tail to tail. This process is not limited to vectors of* $\mathbb{R}^2$*. Use the following GeoGebra interactive to add multiple vectors in* $\mathbb{R}^3$*, two vectors at a time, by constructing diagonals of parallelograms. To use the interactive*

- *Define vectors* **u**, **v** *and* **w**.
- *Use check-boxes at the bottom of the right panel to display the parallelograms.*
- *RIGHT-CLICK and DRAG the left panel to rotate the graph.*

Access GeoGebra interactives through the online version of this text at

https://ximera.osu.edu/linearalgebradzv3/LinearAlgebraInteractiveIntro.

The sum of two vectors can be visualized as the diagonal of a parallelogram. The sum of three (non-co-planar) vectors is the diagonal of a three-dimensional counterpart of a parallelogram, called a parallelepiped*. Each face of the parallelepiped is a parallelogram determined by two out of the three given vectors. The following GeoGebra exercise will help you visualize the sum of three vectors as the diagonal of a parallelepiped.*

Define vectors **u**, **v** *and* **w**. *The sum is the diagonal of the parallelepiped. RIGHT-CLICK and DRAG the left panel to rotate the graph.*

Access GeoGebra interactives through the online version of this text at

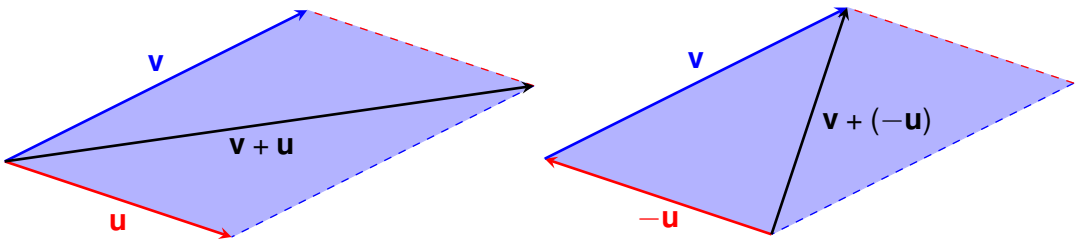
https://ximera.osu.edu/linearalgebradzv3/LinearAlgebraInteractiveIntro.

Vector Subtraction

We can find the difference of two vectors by interpreting subtraction as "addition of the opposite". Thus,

$$\mathbf{v} - \mathbf{u} = \mathbf{v} + (-\mathbf{u})$$

Vector subtraction has an interesting geometric interpretation. As shown in the figure below, if $\mathbf{v} + \mathbf{u}$ is a diagonal of the parallelogram determined by $\mathbf{v}$ and $\mathbf{u}$, the difference $\mathbf{v} - \mathbf{u}$ is the other diagonal of the same parallelogram.



Properties of Vector Addition and Scalar Multiplication

Theorem 1. The following properties hold for vectors $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{w}$ in $\mathbb{R}^n$ and scalars k and p in $\mathbb{R}$.

1. Commutative Property of Addition
$$\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$$
2. Associative Property of Addition
$$(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$$
3. Existence of Additive Identity: There exists a vector $\mathbf{0}$ such that
$$\mathbf{u} + \mathbf{0} = \mathbf{u}$$
4. Existence of Additive Inverse: For every vector $\mathbf{u}$, there exists a vector $-\mathbf{u}$ such that
$$\mathbf{u} + (-\mathbf{u}) = \mathbf{0}$$
5. Distributive Property over Vector Addition
$$k(\mathbf{u} + \mathbf{v}) = k\mathbf{u} + k\mathbf{v}$$
6. Distributive Property over Scalar Addition
$$(k + p)\mathbf{u} = k\mathbf{u} + p\mathbf{u}$$
7. Associative Property for Scalar Multiplication
$$k(p\mathbf{u}) = (kp)\mathbf{u}$$
8. Multiplication by 1
$$1\mathbf{u} = \mathbf{u}$$

We will prove Properties **??** and **??**. Proofs of the remaining properties are left to the reader.

Proof of Property ?? For any vector $\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}$ in $\mathbb{R}^n$, let

$$-\mathbf{u} = (-1)\mathbf{u} = \begin{bmatrix} -u_1 \\ -u_2 \\ \vdots \\ -u_n \end{bmatrix}$$

Then $\mathbf{u} + (-\mathbf{u}) = \mathbf{0}$. ■

Proof of Property ??

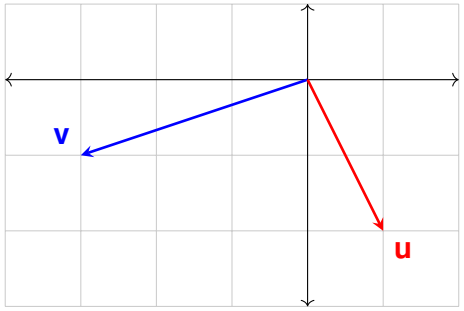
$$k(\mathbf{u}+\mathbf{v}) = k\left(\begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} + \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}\right) = k\begin{bmatrix} u_1+v_1 \\ u_2+v_2 \\ \vdots \\ u_n+v_n \end{bmatrix} = \begin{bmatrix} k(u_1+v_1) \\ k(u_2+v_2) \\ \vdots \\ k(u_n+v_n) \end{bmatrix} =$$

$$= \begin{bmatrix} ku_1+kv_1 \\ ku_2+kv_2 \\ \vdots \\ ku_n+kv_n \end{bmatrix} = \begin{bmatrix} ku_1 \\ ku_2 \\ \vdots \\ ku_n \end{bmatrix} + \begin{bmatrix} kv_1 \\ kv_2 \\ \vdots \\ kv_n \end{bmatrix} = k\begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} + k\begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = k\mathbf{u}+k\mathbf{v}$$

■

Practice Problems

Problem 20 *The figure below shows vectors* $\mathbf{u}$ *and* $\mathbf{v}$*. Sketch each of the following in the same coordinate plane.*



1. $\mathbf{u}+\mathbf{v}$
2. $2\mathbf{u}-\mathbf{v}$
3. $3\mathbf{v}$
4. $-2\mathbf{u}$

Problems **??-??**

Let

$$\mathbf{u} = \begin{bmatrix} -2 \\ 1 \\ 4 \end{bmatrix}, \quad \mathbf{v} \begin{bmatrix} 3 \\ -1 \\ 0 \end{bmatrix}$$

Find each of the following

Problem 21

$$2\mathbf{u} - \mathbf{v} = \begin{bmatrix} \ldots \ldots \\ \ldots \ldots \\ \ldots \ldots \end{bmatrix}$$

Problem 22 *The additive inverse of* $\mathbf{v}$ *is*

$$\begin{bmatrix} \ldots \ldots \\ \ldots \ldots \\ \ldots \ldots \end{bmatrix}$$

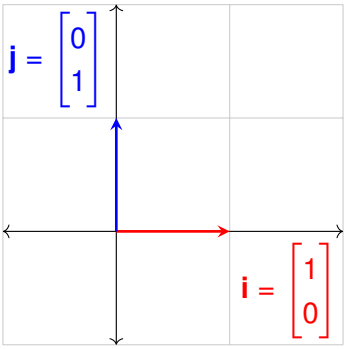
Problem 23 *Prove Properties* **??**-**??** *of Theorem* **??**.

Problem 24 *Prove Properties* **??**-**??** *of Theorem* **??**.

6 Standard Unit Vectors in $\mathbb{R}^n$

Standard Unit Vectors in $\mathbb{R}^2$ and $\mathbb{R}^3$

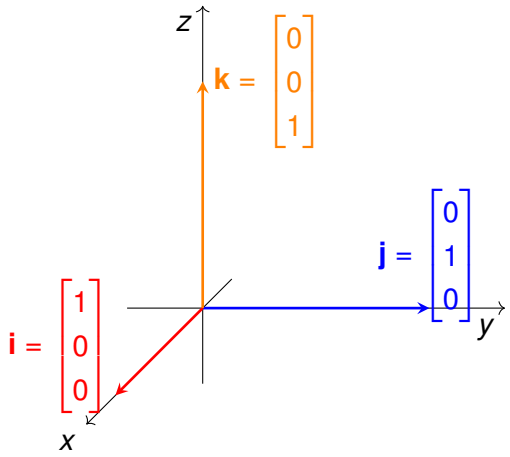
A *unit vector* is a vector of length 1. A unit vector in the positive direction of a coordinate axis is called a *standard unit vector*. There are two standard unit vectors in $\mathbb{R}^2$. The vector $\mathbf{i} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ is parallel the x-axis, and the vector $\mathbf{j} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ is parallel the y-axis.



Vector names **i** and **j** are reserved for standard unit vectors in the direction of x and y axes, respectively. We chose to express **i** and **j** as column vectors, instead of row vectors, because the context in which we will encounter them in the future will require them to be column vectors. You may see them presented as row vectors in a different course.

There are three standard unit vectors in $\mathbb{R}^3$:

$$\mathbf{i} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{j} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad \mathbf{k} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$



A Vector as a Linear Combination of Standard Unit Vectors

Every vector in $\mathbb{R}^2$ and $\mathbb{R}^3$ can be written as a sum of scalar multiples of **i**, **j** and **k**. For example, if $\mathbf{v} = \begin{bmatrix} 3 \\ -2 \\ 7 \end{bmatrix}$, then

$$\mathbf{v} = \begin{bmatrix} 3 \\ -2 \\ 7 \end{bmatrix} = \begin{bmatrix} 3 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ -2 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 7 \end{bmatrix} = 3 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + (-2) \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + 7 \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = 3\mathbf{i} - 2\mathbf{j} + 7\mathbf{k}$$

The expression $3\mathbf{i} - 2\mathbf{j} + 7\mathbf{k}$ is called a *linear combination* of **i**, **j** and **k**.

Standard Unit Vectors in $\mathbb{R}^n$

When working with vectors in $\mathbb{R}^n$, we often use a different notation to denote the standard unit vectors.

Definition 5. Let $\mathbf{e}_i$ denote a vector that has 1 as the i^{th} component and zeros elsewhere. In

other words,

$$\mathbf{e}_i = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix}$$

where 1 is in the i^{th} position. We say that $\mathbf{e}_i$ is a *standard unit vector of* $\mathbb{R}^n$.

Practice Problems

Problem 25 *Express the given vector as a linear combination of appropriate standard unit vectors.*

$$\mathbf{u} = \begin{bmatrix} 0 \\ 4 \\ -3 \end{bmatrix}$$

Answer:

$$\mathbf{u} = \ldots\ldots\mathbf{i} + \ldots\ldots\mathbf{j} + \ldots\ldots\mathbf{k}$$

Problem 26 *Express the given vector as a linear combination of appropriate standard unit vectors.*

$$\mathbf{w} = \begin{bmatrix} 5 \\ -3 \\ 1 \\ 7 \end{bmatrix}$$

Answer:

$$\mathbf{w} = \ldots\ldots\mathbf{e}_1 + \ldots\ldots\mathbf{e}_2 + \ldots\ldots\mathbf{e}_3 + \ldots\ldots\mathbf{e}_4$$

Problem 27 *Express the given vector in component form.*

$$\mathbf{u} = \mathbf{i} + 3\mathbf{j}$$

is a vector in $\mathbb{R}^2$.

Answer:

$$\mathbf{u} = \begin{bmatrix} \ldots\ldots \\ \ldots\ldots \end{bmatrix}$$

Problem 28 *Express the given vector in component form.*

$$\mathbf{v} = -\mathbf{j} + 5\mathbf{k}$$

is a vector in $\mathbb{R}^3$.

Answer:

$$\mathbf{v} = \begin{bmatrix} \dots \\ \dots \\ \dots \end{bmatrix}$$

Problem 29 *Express the given vector in component form.*

$$\mathbf{w} = \mathbf{e}_1 - 2\mathbf{e}_3 + 4\mathbf{e}_4$$

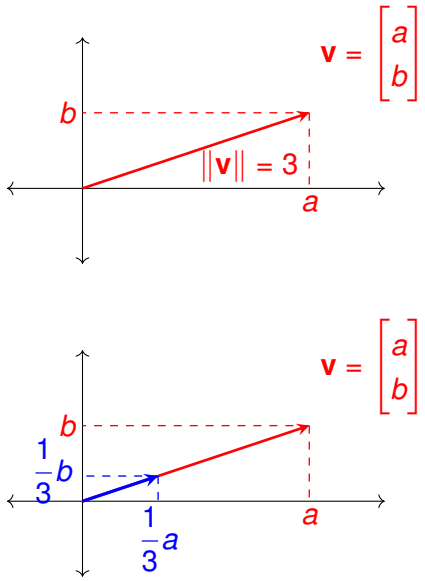
is a vector in $\mathbb{R}^4$.

Answer:

$$\mathbf{w} = \begin{bmatrix} \dots \\ \dots \\ \dots \\ \dots \end{bmatrix}$$

7 Unit Vector in the Direction of a Given Vector

Recall that a *unit* vector is a vector of length 1. Given a non-zero vector $\mathbf{v}$, we can find a unit vector in the same direction by multiplying $\mathbf{v}$ by an appropriate scalar. For example, if $\mathbf{v} = \begin{bmatrix} a \\ b \end{bmatrix}$ and $\|\mathbf{v}\| = 3$, then a unit vector $\mathbf{u}$ in the same direction is given by $\mathbf{u} = \begin{bmatrix} a/3 \\ b/3 \end{bmatrix} = \begin{bmatrix} a/\|\mathbf{v}\| \\ b/\|\mathbf{v}\| \end{bmatrix}$.



In general, dividing a non-zero vector by its own magnitude produces a unit vector in the same direction. We summarize this observation in a theorem.

Theorem 2. Let $\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$ be a non-zero vector in $\mathbb{R}^n$. Vector $\mathbf{u}$ given by

$$\mathbf{u} = \frac{1}{\|\mathbf{v}\|}\mathbf{v} = \begin{bmatrix} v_1/\|\mathbf{v}\| \\ v_2/\|\mathbf{v}\| \\ \vdots \\ v_n/\|\mathbf{v}\| \end{bmatrix}$$

is a unit vector in the direction of $\mathbf{v}$.

Proof Because $\mathbf{u}$ is a positive scalar multiple of $\mathbf{v}$, $\mathbf{u}$ points in the direction of $\mathbf{v}$. We now show that $\|\mathbf{u}\| = 1$.

$$\begin{aligned}\|\mathbf{u}\| &= \sqrt{\left(\frac{v_1}{\|\mathbf{v}\|}\right)^2 + \left(\frac{v_2}{\|\mathbf{v}\|}\right)^2 + \ldots + \left(\frac{v_n}{\|\mathbf{v}\|}\right)^2} \\ &= \frac{1}{\|\mathbf{v}\|}\sqrt{v_1^2 + v_2^2 + \ldots + v_n^2} \\ &= \frac{\|\mathbf{v}\|}{\|\mathbf{v}\|} = 1\end{aligned}$$

■

Example 6. Find a unit vector in the direction of $\mathbf{v} = \begin{bmatrix} 2 \\ -3 \\ 1 \\ 0 \\ 1 \end{bmatrix}$.

Explanation. We first compute $\|\mathbf{v}\|$.

$$\|\mathbf{v}\| = \sqrt{4 + 9 + 1 + 1} = \sqrt{15}$$

By Theorem **??**,

$$\mathbf{u} = \begin{bmatrix} 2/\sqrt{15} \\ -3/\sqrt{15} \\ 1/\sqrt{15} \\ 0 \\ 1/\sqrt{15} \end{bmatrix} = \frac{1}{\sqrt{15}} \begin{bmatrix} 2 \\ -3 \\ 1 \\ 0 \\ 1 \end{bmatrix}$$

Practice Problems

Problem 30 *Find a unit vector in the direction of the given vector* $\mathbf{v}$.

$$\mathbf{v} = \begin{bmatrix} -3 \\ 4 \end{bmatrix}$$

Answer: $\mathbf{u} = \begin{bmatrix} \ldots \ldots \\ \ldots \ldots \end{bmatrix}$

Problem 31 *Find a unit vector in the direction of the given vector* $\mathbf{v}$.

$$\mathbf{v} = \begin{bmatrix} 2 \\ -3 \\ 6 \end{bmatrix}$$

Enter exact answers. No decimal approximations.

Answer: $\mathbf{u} = \begin{bmatrix} \ldots \ldots \\ \ldots \ldots \\ \ldots \ldots \end{bmatrix}$

Problem 32 *Let* $\mathbf{v} = \begin{bmatrix} -1 \\ 1 \\ \sqrt{7} \end{bmatrix}$. *Apply the concepts from this section to find a vector* $\mathbf{w}$ *that points in the same direction as* $\mathbf{v}$ *and whose length is 5.*

Enter exact answers. No decimal approximations.

Answer: $\mathbf{w} = \begin{bmatrix} \ldots \ldots \\ \ldots \ldots \\ \ldots \ldots \end{bmatrix}$

8 Dot Product and its Properties

Definition 6. Let $\mathbf{u}$ and $\mathbf{v}$ be vectors in $\mathbb{R}^n$. The *dot product* of $\mathbf{u}$ and $\mathbf{v}$, denoted by $\mathbf{u} \cdot \mathbf{v}$, is given by

$$\mathbf{u} \cdot \mathbf{v} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} \cdot \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = u_1 v_1 + u_2 v_2 + \ldots + u_n v_n.$$

Example 7. Find $\mathbf{u} \cdot \mathbf{v}$ if $\mathbf{u} = \begin{bmatrix} -2 \\ 0 \\ 1 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} 3 \\ 2 \\ -4 \end{bmatrix}$.

Explanation.

$$\mathbf{u} \cdot \mathbf{v} = \begin{bmatrix} -2 \\ 0 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} 3 \\ 2 \\ -4 \end{bmatrix} = (-2)(3) + (0)(2) + (1)(-4) = -6 - 4 = -10$$

Note that the dot product of two vectors is a scalar. For this reason, the dot product is sometimes called a *scalar product*.

Properties of the Dot Product

A quick examination of Example **??** will convince you that the dot product is *commutative*. In other words, $\mathbf{u} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{u}$. This and other properties of the dot product are stated below.

Theorem 3. The following properties hold for vectors $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{w}$ in $\mathbb{R}^n$ and scalar k in $\mathbb{R}$.

1. $\mathbf{u} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{u}$
2. $(\mathbf{u} + \mathbf{v}) \cdot \mathbf{w} = \mathbf{u} \cdot \mathbf{w} + \mathbf{v} \cdot \mathbf{w}$
3. $\mathbf{u} \cdot (\mathbf{v} + \mathbf{w}) = \mathbf{u} \cdot \mathbf{v} + \mathbf{u} \cdot \mathbf{w}$
4. $(k\mathbf{u}) \cdot \mathbf{v} = k(\mathbf{u} \cdot \mathbf{v}) = \mathbf{u} \cdot (k\mathbf{v})$
5. $\mathbf{u} \cdot \mathbf{u} \geq 0$, and $\mathbf{u} \cdot \mathbf{u} = 0$ if and only if $\mathbf{u} = \mathbf{0}$.

6. $\|\mathbf{u}\|^2 = \mathbf{u} \cdot \mathbf{u}$

We will prove Property **??**. The remaining properties are left as exercises.

Proof of Property ??:

$$
\begin{aligned}
(\mathbf{u}+\mathbf{v}) \cdot \mathbf{w} &= \left(\begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} + \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} \right) \cdot \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix} = \begin{bmatrix} u_1 + v_1 \\ u_2 + v_2 \\ \vdots \\ u_n + v_n \end{bmatrix} \cdot \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix} \\
&= (u_1 + v_1)w_1 + (u_2 + v_2)w_2 + \ldots + (u_n + v_n)w_n \\
&= u_1w_1 + v_1w_1 + u_2w_2 + v_2w_2 + \ldots + u_nw_n + v_nw_n \\
&= (u_1w_1 + u_2w_2 \ldots + u_nw_n) + (v_1w_1 + v_2w_2 + \ldots + v_nw_n) \\
&= \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} \cdot \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix} + \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} \cdot \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix} = \mathbf{u} \cdot \mathbf{w} + \mathbf{v} \cdot \mathbf{w}
\end{aligned}
$$

■

We will illustrate Property **??** with an example.

Example 8. Let $\mathbf{u} = \begin{bmatrix} -2 \\ 3 \end{bmatrix}$. Use $\mathbf{u}$ to illustrate Property **??** of Theorem **??**.

Explanation.

$$\|\mathbf{u}\|^2 = (-2)^2 + 3^2 = (-2)(-2) + (3)(3) = \begin{bmatrix} -2 \\ 3 \end{bmatrix} \cdot \begin{bmatrix} -2 \\ 3 \end{bmatrix} = \mathbf{u} \cdot \mathbf{u}$$

If we take the square root of both sides of the equation in Property **??**, we get an alternative way to think of the length of a vector.

Corollary 1 (Length of a Vector)**.** Let $\mathbf{v}$ be a vector in $\mathbb{R}^n$, then the *length*, or the *magnitude*, of $\mathbf{v}$ is given by

$$\|\mathbf{v}\| = \sqrt{\mathbf{v} \cdot \mathbf{v}}$$

Practice Problems

Problem 33 *Find the dot product of* $\mathbf{u}$ *and* $\mathbf{v}$ *if*

$$\mathbf{u} = \begin{bmatrix} -1 \\ -2 \\ 5 \\ 4 \end{bmatrix}, \quad \mathbf{v} = \begin{bmatrix} 2 \\ -2 \\ -3 \\ 1 \end{bmatrix}$$

Answer:

$$\mathbf{u} \cdot \mathbf{v} = \ldots \ldots$$

Problem 34 *Find the dot product of* $\mathbf{u}$ *and* $\mathbf{v}$ *if*

$$\mathbf{u} = \begin{bmatrix} 1 \\ 1/2 \end{bmatrix}, \quad \mathbf{v} = \begin{bmatrix} -2 \\ 4 \end{bmatrix}$$

Answer:

$$\mathbf{u} \cdot \mathbf{v} = \ldots \ldots$$

Problem 35 *Use vector* $\mathbf{u} = \begin{bmatrix} 2 \\ 5 \\ -7 \end{bmatrix}$ *to illustrate Property* **??** *of Theorem* **??**.

Problem 36 *Prove Properties* **??**, **??**, **??**, **??** *and* **??** *of Theorem* **??**.

Problem 37 *From the given list of vector pairs, identify ALL pairs of vectors that lie on perpendicular lines.*

Select All Correct Answers:

1. $\mathbf{u} = \begin{bmatrix} 1 \\ 1 \\ \overline{2} \end{bmatrix}$, $\mathbf{v} = \begin{bmatrix} -2 \\ 4 \end{bmatrix}$
2. $\mathbf{u} = \begin{bmatrix} -1 \\ 1 \\ \overline{2} \end{bmatrix}$, $\mathbf{v} = \begin{bmatrix} -2 \\ 4 \end{bmatrix}$

3. $\mathbf{u} = \begin{bmatrix} 1 \\ \frac{1}{2} \end{bmatrix}$, $\mathbf{v} = \begin{bmatrix} 1 \\ -2 \end{bmatrix}$

4. $\mathbf{u} = \begin{bmatrix} -1 \\ -\frac{1}{2} \end{bmatrix}$, $\mathbf{v} = \begin{bmatrix} -2 \\ 4 \end{bmatrix}$

Compute $\mathbf{u} \cdot \mathbf{v}$ *for each pair. What do you observe?*

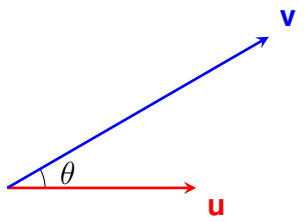
Problem 38 1. *Vector* $\mathbf{u}$ *parallel to the line* $y = mx$ *has the form* $\mathbf{u} = k \begin{bmatrix} 1 \\ m \end{bmatrix}$. *Assuming that* $m \neq 0$, *find the general form for a vector* $\mathbf{v}$ *perpendicular to* $y = mx$.

2. *Find* $\mathbf{u} \cdot \mathbf{v}$.

3. *What is true about the dot product of perpendicular vectors?*

9 Dot Product and Angle

Given two vectors **u** and **v**, let θ be the angle between them such that $0 \leq \theta \leq \pi$. We will refer to θ as the *included angle*.

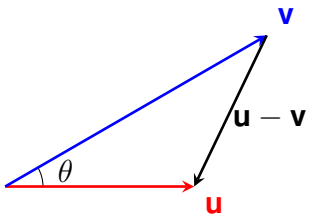


The following theorem establishes a relationship between the dot product and the included angle.

Theorem 4. Let **u** and **v** be vectors in $\mathbb{R}^n$, and let θ be the included angle. Then

$$\mathbf{u} \cdot \mathbf{v} = \|\mathbf{u}\| \, \|\mathbf{v}\| \cos\theta$$

Proof Consider the triangle formed by **u**, **v** and $\mathbf{u} - \mathbf{v}$.



By the Law of Cosines we have:

$$\|\mathbf{u} - \mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 - 2\|\mathbf{u}\| \, \|\mathbf{v}\| \cos\theta$$

By Theorem **????** of Dot Product and its Properties

$$(\mathbf{u} - \mathbf{v}) \cdot (\mathbf{u} - \mathbf{v}) = \mathbf{u} \cdot \mathbf{u} + \mathbf{v} \cdot \mathbf{v} - 2\|\mathbf{u}\| \, \|\mathbf{v}\| \cos\theta$$

By Theorem **????** of Dot Product and its Properties

$$(\mathbf{u} - \mathbf{v}) \cdot \mathbf{u} - (\mathbf{u} - \mathbf{v}) \cdot \mathbf{v} = \mathbf{u} \cdot \mathbf{u} + \mathbf{v} \cdot \mathbf{v} - 2\|\mathbf{u}\| \, \|\mathbf{v}\| \cos\theta$$

By Theorem **????** of Dot Product and its Properties

$$\mathbf{u}\cdot\mathbf{u}-\mathbf{v}\cdot\mathbf{u}-\mathbf{u}\cdot\mathbf{v}+\mathbf{v}\cdot\mathbf{v}=\mathbf{u}\cdot\mathbf{u}+\mathbf{v}\cdot\mathbf{v}-2\,\|\mathbf{u}\|\,\|\mathbf{v}\|\cos\theta$$

By Theorem **????** of Dot Product and its Properties

$$-2(\mathbf{u}\cdot\mathbf{v})=-2\,\|\mathbf{u}\|\,\|\mathbf{v}\|\cos\theta$$
$$\mathbf{u}\cdot\mathbf{v}=\|\mathbf{u}\|\,\|\mathbf{v}\|\cos\theta$$

■

Example 9. Find the included angle between vectors $\mathbf{u}=\begin{bmatrix}2\\-1\\4\end{bmatrix}$ and $\mathbf{v}=\begin{bmatrix}1\\2\\-3\end{bmatrix}$.

Explanation. By Theorem **??**, $\cos\theta=\dfrac{\mathbf{u}\cdot\mathbf{v}}{\|\mathbf{u}\|\,\|\mathbf{v}\|}$.

$$\mathbf{u}\cdot\mathbf{v}=2-2-12=-12$$
$$\|\mathbf{u}\|=\sqrt{4+1+16}=\sqrt{21}$$
$$\|\mathbf{v}\|=\sqrt{1+4+9}=\sqrt{14}$$
$$\cos\theta=\frac{-12}{\sqrt{21}\sqrt{14}}$$
$$\theta\approx 134.4^\circ$$

Orthogonal Vectors

Definition 7. Let $\mathbf{u}$ and $\mathbf{v}$ be vectors in $\mathbb{R}^n$. We say $\mathbf{u}$ and $\mathbf{v}$ are *orthogonal* if $\mathbf{u}\cdot\mathbf{v}=0$.

We can use Theorem **??** to show that two non-zero *orthogonal* vectors of $\mathbb{R}^n$ are simply *perpendicular* vectors (the included angle is 90°). To see this, suppose that $\mathbf{u}\cdot\mathbf{v}=0$ for nonzero vectors $\mathbf{u},\mathbf{v}$. Then from Theorem **??** we have

$$0=\|\mathbf{u}\|\,\|\mathbf{v}\|\cos\theta.$$

Since $\mathbf{u},\mathbf{v}$ are nonzero vectors, we have $0=\cos\theta$, which implies $\theta=90^\circ$. The converse also holds. If $\theta=90^\circ$, then the dot product is clearly 0.

The reason we prefer the term “orthogonal” to “perpendicular” in this course is because $\mathbb{R}^n$ is only one example of a *vector space*, and the dot product is only one example of a more general

product, called an *inner product*. For vectors in $\mathbb{R}^n$ a zero dot product happens to coincide with the geometric idea of perpendicularity, but there are many vector spaces that do not possess the visual geometry of $\mathbb{R}^n$. (Later in the text, you will encounter vector spaces whose vectors are polynomial functions!) In these more abstract settings, a zero inner product still signals a special relationship between vectors. The term *orthogonal* captures this relationship.

Remark 1. By our definition, the zero vector is orthogonal to any vector. However, we will not use the word *perpendicular* when the zero vector is involved, as it is not possible to talk about an "included angle".

Practice Problems

Problem 39 *Find the degree measure of the included angle, θ for each pair of vectors. Round your answers to the nearest tenth.*

1. $\begin{bmatrix}1\\2\end{bmatrix}$ *and* $\begin{bmatrix}-3\\-1\end{bmatrix}$.
   Answer: $\theta = \ldots\ldots^\circ$
2. $\begin{bmatrix}-1\\2\\4\end{bmatrix}$ *and* $\begin{bmatrix}-2\\1\\-1\end{bmatrix}$
   Answer: $\theta = \ldots\ldots^\circ$
3. $\begin{bmatrix}0\\-3\\1\end{bmatrix}$ *and* $\begin{bmatrix}-5\\-2\\4\end{bmatrix}$
   Answer: $\theta = \ldots\ldots^\circ$

Problem 40 *What does the sign of the dot product tell us about the included angle?*

Problem 41 *Find all values of a so that* $\begin{bmatrix}a^2\\2a\\1\end{bmatrix}$ *is orthogonal to* $\begin{bmatrix}1\\2\\3\end{bmatrix}$. *List your answers in increasing order.*

Answer: $\ldots\ldots, \ldots\ldots$

Problem 42 *Find the value of x for which the vector $\begin{bmatrix} x \\ -4 \end{bmatrix}$ is parallel to the vector $\begin{bmatrix} 3 \\ 2 \end{bmatrix}$. What is the measure of the included angle, θ? Find the measure of the included angle using Theorem* **??**. *Do the two results agree?*

Answer:

$$x = \ldots\ldots$$

$$\theta = \ldots\ldots^\circ$$

Problem 43 *Prove that if* $\mathbf{u}$ *is a unit vector, then* $\mathbf{u} \cdot \mathbf{u} = 1$.

Problem 44 *Prove that if* $\mathbf{u}_1$ *and* $\mathbf{u}_2$ *are unit vectors, then* $-1 \leq \mathbf{u}_1 \cdot \mathbf{u}_2 \leq 1$. *In what cases are the extreme values of 1 and* -1 *attained?*

Problem 45 *Imagine a clock with hands represented by vectors* $\mathbf{m}$ *and* $\mathbf{h}$*, as shown below. At what whole hour will* $\mathbf{m} \cdot \mathbf{h}$ *attain its maximum value? At what whole hour will* $\mathbf{m} \cdot \mathbf{h}$ *be as small as possible?*



Answer:

$\mathbf{m} \cdot \mathbf{h}$ *is greatest at* ..... : 00 *o'clock*

$\mathbf{m} \cdot \mathbf{h}$ *is smallest at* ..... : 00 *o'clock*

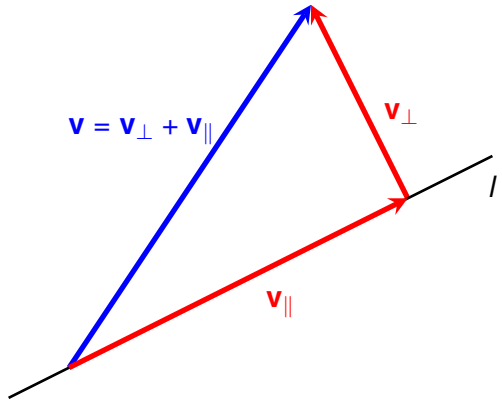
Photo Credits

The following images are courtesy of Wikimedia Commons

Hannes Grobe, *Wall clock manufactured by Telefonbau & Normalzeit.* CC-BY 3.0

10 Orthogonal Projections

Given a line l and a vector $\mathbf{v}$ emanating from a point on l, it is sometimes convenient to express $\mathbf{v}$ as the sum of a vector $\mathbf{v}_{\parallel}$, parallel to l, and a vector $\mathbf{v}_{\perp}$, perpendicular to l. If you have taken a physics course, you may have seen a force vector decomposed into the sum of two components: one parallel and one perpendicular to the direction of motion.



Suppose $\mathbf{d}$ is a direction vector for l. Then $\mathbf{v}_{\parallel} = k\mathbf{d}$ for some scalar k. Our goal is to find k.

$$\begin{aligned}\mathbf{v}\cdot\mathbf{d} &= (\mathbf{v}_{\parallel}+\mathbf{v}_{\perp})\cdot\mathbf{d}\\ &= (k\mathbf{d}+\mathbf{v}_{\perp})\cdot\mathbf{d}\\ &= k\mathbf{d}\cdot\mathbf{d}+\mathbf{v}_{\perp}\cdot\mathbf{d}\\ &= k\left\|\mathbf{d}\right\|^2+0\\ &= k\left\|\mathbf{d}\right\|^2\end{aligned}$$

We conclude that

$$k = \frac{\mathbf{v}\cdot\mathbf{d}}{\left\|\mathbf{d}\right\|^2}$$

and

$$\mathbf{v}_{\parallel} = k\mathbf{d} = \left(\frac{\mathbf{v}\cdot\mathbf{d}}{\left\|\mathbf{d}\right\|^2}\right)\mathbf{d}$$

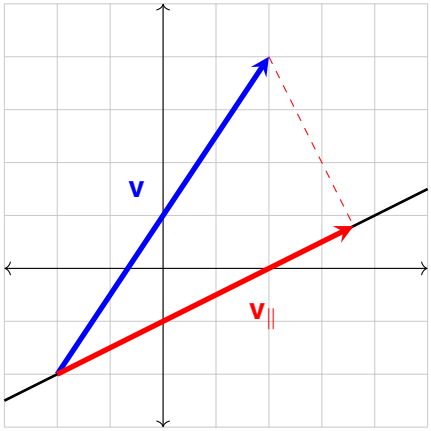
The vector $\mathbf{v}_{\parallel} = \left(\dfrac{\mathbf{v}\cdot\mathbf{d}}{\left\|\mathbf{d}\right\|^2}\right)\mathbf{d}$ is called the *projection of* $\mathbf{v}$ *onto* $\mathbf{d}$. In our discussion, $\mathbf{d}$ is a direction vector for line l. So, we can also say that $\mathbf{v}_{\parallel}$ is the *projection of* $\mathbf{v}$ *onto* l.

To find $\mathbf{v}_{\perp}$, observe that $\mathbf{v}_{\perp} = \mathbf{v} - \mathbf{v}_{\parallel}$.

Definition 8. Let $\mathbf{v}$ be a vector, and let $\mathbf{d}$ be a non-zero vector. The *projection of* $\mathbf{v}$ *onto* $\mathbf{d}$ is given by

$$\text{proj}_{\mathbf{d}}\mathbf{v} = \left(\frac{\mathbf{v}\cdot\mathbf{d}}{\|\mathbf{d}\|^2}\right)\mathbf{d}$$

Example 10. Find the projection of $\mathbf{v}$, shown below, onto the line given by $y = \frac{1}{2}x - 1$.



Explanation. We begin by finding vectors $\mathbf{v}$ and $\mathbf{d}$. The tail of $\mathbf{v}$ is located at $(-2,-2)$, and the head of $\mathbf{v}$ is at $(2,4)$. Using the "head-tail" formula we get

$$\mathbf{v} = \begin{bmatrix} 2-(-2) \\ 4-(-2) \end{bmatrix} = \begin{bmatrix} 4 \\ 6 \end{bmatrix}$$

The direction vector for the line $y = \frac{1}{2}x - 1$ is

$$\mathbf{d} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

We find that $\mathbf{v}\cdot\mathbf{d} = 14$ and $\|\mathbf{d}\|^2 = 5$. Thus

$$\text{proj}_{\mathbf{d}}\mathbf{v} = \left(\frac{\mathbf{v}\cdot\mathbf{d}}{\|\mathbf{d}\|^2}\right)\mathbf{d} = \frac{14}{5}\begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 28/5 \\ 14/5 \end{bmatrix}$$

Distance from a Point to a Line

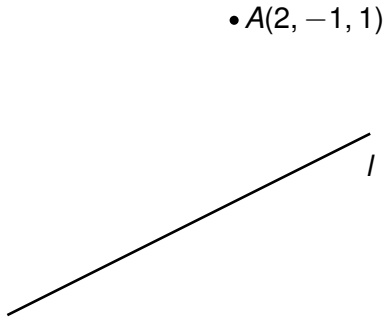
The shortest distance from a point to a line is the length of the perpendicular line segment dropped from the point to the line. Vector projection formula will help us find the length of such a perpendicular.

Example 11. Let $A(2, -1, 1)$ be a point in $\mathbb{R}^3$. Suppose line l is given by parametric equations

$$x = t + 3$$

$$y = -t + 1$$

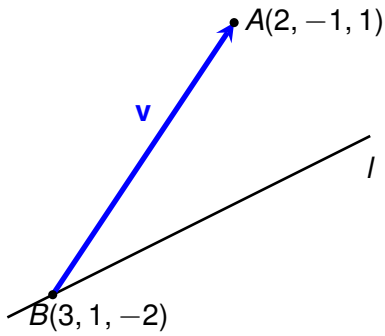
$$z = t - 2$$



Find the distance from A to l.

Explanation. We will first construct a vector $\mathbf{v}$ by picking an arbitrary point B on l to be the tail of $\mathbf{v}$ and using point A as the head of $\mathbf{v}$. An easy point to choose on line l is the point $(3, 1, -2)$ that corresponds to $t = 0$. Now we have

$$\mathbf{v} = \overrightarrow{BA} = \begin{bmatrix} 2-3 \\ -1-1 \\ 1-(-2) \end{bmatrix} = \begin{bmatrix} -1 \\ -2 \\ 3 \end{bmatrix}$$

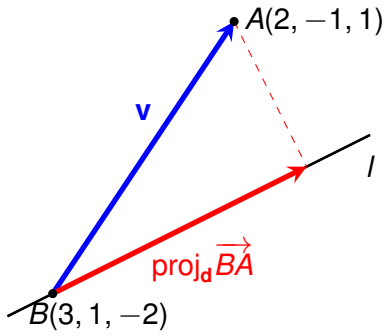


The line has a direction vector

$$\mathbf{d} = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$$

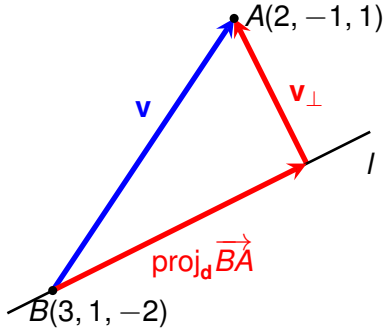
We will now find the projection of $\overrightarrow{BA}$ onto l

$$\text{proj}_{\mathbf{d}}\overrightarrow{BA} = \left(\frac{\mathbf{v}\cdot\mathbf{d}}{\|\mathbf{d}\|^2}\right)\mathbf{d} = \frac{4}{3}\begin{bmatrix}1\\-1\\1\end{bmatrix} = \begin{bmatrix}4/3\\-4/3\\4/3\end{bmatrix}$$



Next, we find $\mathbf{v}_\perp$.

$$\mathbf{v}_\perp = \mathbf{v} - \mathbf{v}_\parallel = \begin{bmatrix}-1\\-2\\3\end{bmatrix} - \begin{bmatrix}4/3\\-4/3\\4/3\end{bmatrix} = \begin{bmatrix}-7/3\\-2/3\\5/3\end{bmatrix}$$



Finally, to find the distance between point A and line l, we find the magnitude of $\mathbf{v}_\perp$.

$$\|\mathbf{v}_\perp\| = \frac{1}{3}\sqrt{49+4+25} = \frac{\sqrt{78}}{3}$$

Practice Problems

Problems **??-??**

Find $\text{proj}_{\mathbf{d}}\mathbf{v}$.

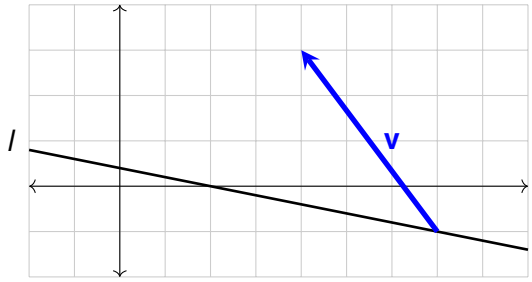
Problem 46 *If* $\mathbf{d} = \begin{bmatrix} -1 \\ 3 \end{bmatrix}$ *and* $\mathbf{v} = \begin{bmatrix} 1 \\ 4 \end{bmatrix}$ *then*

$$proj_{\mathbf{d}}\mathbf{v} = \begin{bmatrix} \ldots \ldots \\ \ldots \ldots \end{bmatrix}$$

Problem 47 *If* $\mathbf{d} = \begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix}$ *and* $\mathbf{v} = \begin{bmatrix} -1 \\ -4 \\ 2 \end{bmatrix}$ *then*

$$proj_{\mathbf{d}}\mathbf{v} = \begin{bmatrix} \ldots \ldots \\ \ldots \ldots \\ \ldots \ldots \end{bmatrix}$$

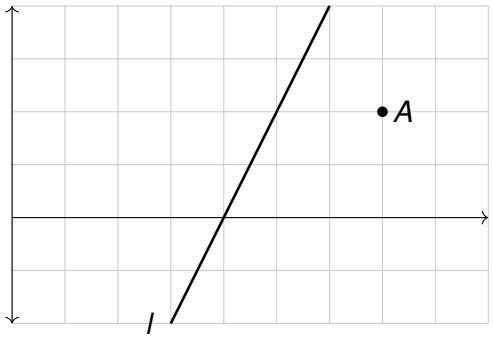
Problem 48 *Find the projection of vector* $\mathbf{v}$ *onto line l. (If entering answers in decimal form, round to the nearest one hundredth.)*



Answer:

$$\begin{bmatrix} \ldots \ldots \\ \ldots \ldots \end{bmatrix}$$

Problem 49 *Find the distance between point A and line l.*



Answer: $\sqrt{\ldots\ldots}$.

Problem 50 *Show that* $\text{proj}_{\mathbf{d}}\mathbf{v}$ *does not depend on the length of* $\mathbf{d}$ *by proving that* $\text{proj}_{\mathbf{d}}\mathbf{v} = \text{proj}_{k\mathbf{d}}\mathbf{v}$ *for* $k \neq 0$. *What does this result mean geometrically? Illustrate your response with a diagram.*

Problem 51 *Find the radius of a circle centered at* $(4, 2)$ *if the line* $y = \frac{3}{2}x + 3$ *is tangent to the circle. Enter your response as a fraction.*

Answer:

$$r = \sqrt{\ldots\ldots}$$

The graph below shows the line $y = \frac{3}{2}x + 3$ *together with a circle of radius* 1. *Change the value of* r *to the radius you have found to visualize the correct answer.*

Access GeoGebra interactives through the online version of this text at

https://ximera.osu.edu/linearalgebradzv3/LinearAlgebraInteractiveIntro.

11 Cross Product and its Properties

In Dot Product and its Properties we introduced the *dot product*, one of two important products for vectors. We will now introduce the second type of product, called the *cross product*. There are several important distinctions to keep in mind. First, the dot product is defined for two vectors of $\mathbb{R}^n$, for any natural number n; the cross-product will only be defined for vectors of $\mathbb{R}^3$. Second, the dot product is a scalar; the cross product of two vectors will be a vector. Finally, we will find that unlike the dot product, the cross product is not commutative.

The cross product has many applications in physics and engineering. It also has important geometric properties which will be addressed in this section and in Determinants as Areas and Volumes.

Preliminaries

In order to define the cross product in a convenient way we need to define 2×2 and 3×3 determinants. If you know how to find such determinants you may skip this section and proceed directly to the definition.

Definition 9 (2×2 Determinant)**.** A 2×2 *determinant* is a number associated with a 2×2 matrix

$$\det \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$$

Example 12.

$$\begin{vmatrix} 2 & 4 \\ -5 & 3 \end{vmatrix} = (2)(3) - (4)(-5) = 6 + 20 = 26$$

Definition 10 (3×3 Determinant)**.** A 3×3 *determinant* is a number associated with a 3×3 matrix

$$\det \begin{bmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{bmatrix} = \begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix} = a_1 \begin{vmatrix} b_2 & b_3 \\ c_2 & c_3 \end{vmatrix} - a_2 \begin{vmatrix} b_1 & b_3 \\ c_1 & c_3 \end{vmatrix} + a_3 \begin{vmatrix} b_1 & b_2 \\ c_1 & c_2 \end{vmatrix}$$

Example 13.

$$\begin{vmatrix} 4 & 2 & 1 \\ 5 & -3 & 0 \\ -2 & 3 & 2 \end{vmatrix} = 4\begin{vmatrix} -3 & 0 \\ 3 & 2 \end{vmatrix} - 2\begin{vmatrix} 5 & 0 \\ -2 & 2 \end{vmatrix} + 1\begin{vmatrix} 5 & -3 \\ -2 & 3 \end{vmatrix}$$

$$= 4\Big((-3)(2) - (0)(3)\Big) - 2\Big((5)(2) - (0)(-2)\Big) + \Big((5)(3) - (-3)(-2)\Big)$$

$$= (4)(-6) - (2)(10) + (15 - 6)$$

$$= -35$$

For more help on 3×3 determinants, see Example **??** and the video located there.

Definition of the Cross Product

Definition 11. Let $\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}$ be vectors in $\mathbb{R}^3$. The *cross product* of $\mathbf{u}$ and $\mathbf{v}$, denoted by $\mathbf{u} \times \mathbf{v}$, is given by

$$\mathbf{u} \times \mathbf{v} = (u_2v_3 - u_3v_2)\mathbf{i} - (u_1v_3 - u_3v_1)\mathbf{j} + (u_1v_2 - u_2v_1)\mathbf{k}$$

$$= (u_2v_3 - u_3v_2)\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} - (u_1v_3 - u_3v_1)\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + (u_1v_2 - u_2v_1)\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} u_2v_3 - u_3v_2 \\ -u_1v_3 + u_3v_1 \\ u_1v_2 - u_2v_1 \end{bmatrix}$$

This formula is much easier to remember when stated symbolically in terms of determinants.

$$\mathbf{u} \times \mathbf{v} = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} \times \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{vmatrix} = \mathbf{i}\begin{vmatrix} u_2 & u_3 \\ v_2 & v_3 \end{vmatrix} - \mathbf{j}\begin{vmatrix} u_1 & u_3 \\ v_1 & v_3 \end{vmatrix} + \mathbf{k}\begin{vmatrix} u_1 & u_2 \\ v_1 & v_2 \end{vmatrix}$$

Example 14. Find the cross product of $\mathbf{u} = \begin{bmatrix} 3 \\ -10 \\ 2 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} -2 \\ 4 \\ 7 \end{bmatrix}$.

Explanation.

$$\mathbf{u}\times\mathbf{v}=\begin{vmatrix}\mathbf{i}&\mathbf{j}&\mathbf{k}\\3&-10&2\\-2&4&7\end{vmatrix}=\mathbf{i}\begin{vmatrix}-10&2\\4&7\end{vmatrix}-\mathbf{j}\begin{vmatrix}3&2\\-2&7\end{vmatrix}+\mathbf{k}\begin{vmatrix}3&-10\\-2&4\end{vmatrix}$$
$$=\mathbf{i}\Big((-10)(7)-(2)(4)\Big)-\mathbf{j}\Big((3)(7)-(2)(-2)\Big)+\mathbf{k}\Big((3)(4)-(-10)(-2)\Big)$$
$$=-78\mathbf{i}-25\mathbf{j}-8\mathbf{k}=\begin{bmatrix}-78\\-25\\-8\end{bmatrix}$$

Properties of the Cross Product

Exploration 52 *What would happen if we took the cross product of the vectors in Example* **??** *but reversed the order?*

Let $\mathbf{v}=\begin{bmatrix}-2\\4\\7\end{bmatrix}$ *and* $\mathbf{u}=\begin{bmatrix}3\\-10\\2\end{bmatrix}$. *Recall that* $\mathbf{u}\times\mathbf{v}=\begin{bmatrix}-78\\-25\\-8\end{bmatrix}$. *We need to compute* $\mathbf{v}\times\mathbf{u}$. *If you have already studied the effect that switching two rows of a matrix has on its determinant, you should be able to guess the outcome of the upcoming computation.*

$$\mathbf{v}\times\mathbf{u}=\begin{vmatrix}\mathbf{i}&\mathbf{j}&\mathbf{k}\\-2&4&7\\3&-10&2\end{vmatrix}=\mathbf{i}\begin{vmatrix}4&7\\-10&2\end{vmatrix}-\mathbf{j}\begin{vmatrix}-2&7\\3&2\end{vmatrix}+\mathbf{k}\begin{vmatrix}-2&4\\3&-10\end{vmatrix}$$
$$=\mathbf{i}\Big((4)(2)-(7)(-10)\Big)-\mathbf{j}\Big((-2)(2)-(7)(3)\Big)+\mathbf{k}\Big((-2)(-10)-(4)(3)\Big)$$
$$=78\mathbf{i}+25\mathbf{j}+8\mathbf{k}=\begin{bmatrix}78\\25\\8\end{bmatrix}=-(\mathbf{u}\times\mathbf{v})$$

This computation shows that the cross product is an operation that is not commutative. It also suggests that switching the order of the vectors changes the sign of the result.

Warning 1. The Cross Product is not a commutative operation.

Theorem 5. Let $\mathbf{u}$ and $\mathbf{v}$ be vectors in $\mathbb{R}^3$, then

$$\mathbf{u}\times\mathbf{v}=-(\mathbf{v}\times\mathbf{u})$$

Proof The proof is left to the reader. (See Practice Problem **??**) ■

The next theorem lists two additional properties of the cross product. Proofs of these properties are routine and are left to the reader.

Theorem 6. Let **u**, **v** and **w** be vectors in $\mathbb{R}^3$, and k be a scalar, then

1. Scalar Associativity
$$(k\mathbf{u}) \times \mathbf{v} = \mathbf{u} \times (k\mathbf{v}) = k(\mathbf{u} \times \mathbf{v})$$
2. Distributivity
$$(\mathbf{u} + \mathbf{v}) \times \mathbf{w} = \mathbf{u} \times \mathbf{w} + \mathbf{v} \times \mathbf{w}$$
$$\mathbf{u} \times (\mathbf{v} + \mathbf{w}) = \mathbf{u} \times \mathbf{v} + \mathbf{u} \times \mathbf{w}$$

The cross product has several important geometric properties. The following problems give us a glimpse of these properties.

Exploration 53 *Compute the following products:*

$$\mathbf{i} \times \mathbf{j} = \begin{bmatrix} \ldots \\ \ldots \\ \ldots \end{bmatrix}, \quad \mathbf{j} \times \mathbf{k} = \begin{bmatrix} \ldots \\ \ldots \\ \ldots \end{bmatrix}$$

For the two vectors in each product, sketch the vectors together with the product vector. What do you observe about the relationship between the cross product and the plane determined by the two vectors in the product?

Orthogonality Property

Exploration 54 *In this problem we will return to vectors of* $\mathbf{v} = \begin{bmatrix} -2 \\ 4 \\ 7 \end{bmatrix}$ *and* $\mathbf{u} = \begin{bmatrix} 3 \\ -10 \\ 2 \end{bmatrix}$ *of Example* **??** *and Exploration Problem* **??**. *We know that*

$$\mathbf{u} \times \mathbf{v} = \begin{bmatrix} -78 \\ -25 \\ -8 \end{bmatrix} \quad \textit{and} \quad \mathbf{v} \times \mathbf{u} = \begin{bmatrix} 78 \\ 25 \\ 8 \end{bmatrix}$$

We will now compute the dot product of $\mathbf{v} \times \mathbf{u}$ *with each of the original vectors* $\mathbf{u}$ *and* $\mathbf{v}$*.*

$$\mathbf{u} \cdot (\mathbf{v} \times \mathbf{u}) = \begin{bmatrix} 3 \\ -10 \\ 2 \end{bmatrix} \cdot \begin{bmatrix} 78 \\ 25 \\ 8 \end{bmatrix} = 0$$

$$\mathbf{v} \cdot (\mathbf{v} \times \mathbf{u}) = \begin{bmatrix} -2 \\ 4 \\ 7 \end{bmatrix} \cdot \begin{bmatrix} 78 \\ 25 \\ 8 \end{bmatrix} = 0$$

It is also easy to verify that $\mathbf{u} \cdot (\mathbf{u} \times \mathbf{v}) = 0$ *and* $\mathbf{v} \cdot (\mathbf{u} \times \mathbf{v}) = 0$*. Recall that the dot product of two vectors is* 0 *if and only if the two vectors are orthogonal. (Definition* **??***) We conclude that, at least in this case, the cross product of two vectors is orthogonal to each of the vectors.*

It turns out that the orthogonality property illustrated by Exploration Problems **??** and **??** holds in general. We state it as a theorem.

Theorem 7. Let $\mathbf{u}$ and $\mathbf{v}$ be vectors of $\mathbb{R}^3$, then $\mathbf{u} \times \mathbf{v}$ is orthogonal to both $\mathbf{u}$ and $\mathbf{v}$.

Proof This proof can be done by direct computation and is left to the reader. (See Practice Problem **??**) ■

Cross Product and the Angle between Vectors

Recall that the dot product of $\mathbf{u}$ and $\mathbf{v}$ is related to the angle θ between $\mathbf{u}$ and $\mathbf{v}$ by the following formula

$$\mathbf{u} \cdot \mathbf{v} = \|\mathbf{u}\| \, \|\mathbf{v}\| \cos\theta \tag{7}$$

We will derive an analogous result for the cross product. To do so, we will need the following Lemma.

Lemma 1. Let $\mathbf{u}$ and $\mathbf{v}$ be vectors in $\mathbb{R}^3$. Then

$$\|\mathbf{u} \times \mathbf{v}\|^2 = \|\mathbf{u}\|^2 \|\mathbf{v}\|^2 - (\mathbf{u} \cdot \mathbf{v})^2$$

Proof The proof is left to the reader. ■

The following theorem establishes a relationship between the magnitude of the cross product, the magnitudes of the two vectors involved in the cross product and the angle between the two vectors. It is important to note that the identity in this theorem involves the magnitude of the cross product, not the cross product itself.

Theorem 8. Let $\mathbf{u}$ and $\mathbf{v}$ be vectors in $\mathbb{R}^3$. Let θ be the angle between $\mathbf{u}$ and $\mathbf{v}$ such that $0 \leq \theta \leq \pi$. Then

$$\|\mathbf{u} \times \mathbf{v}\| = \|\mathbf{u}\| \, \|\mathbf{v}\| \sin\theta$$

Proof By Lemma **??** and (**??**) we have

$$\begin{aligned} \|\mathbf{u} \times \mathbf{v}\|^2 &= \|\mathbf{u}\|^2 \|\mathbf{v}\|^2 - (\mathbf{u} \cdot \mathbf{v})^2 \\ &= \|\mathbf{u}\|^2 \|\mathbf{v}\|^2 - (\|\mathbf{u}\| \, \|\mathbf{v}\| \cos\theta)^2 \\ &= \|\mathbf{u}\|^2 \|\mathbf{v}\|^2 - \|\mathbf{u}\|^2 \|\mathbf{v}\|^2 \cos^2\theta \\ &= \|\mathbf{u}\|^2 \|\mathbf{v}\|^2 (1 - \cos^2\theta) \\ &= \|\mathbf{u}\|^2 \|\mathbf{v}\|^2 \sin^2\theta \end{aligned}$$

Observe that all magnitudes are non-negative. Also, $\sin\theta \geq 0$ because $0 \leq \theta \leq \pi$. Taking the square root of both sides give us

$$\|\mathbf{u} \times \mathbf{v}\| = \|\mathbf{u}\| \, \|\mathbf{v}\| \sin\theta$$

■

Practice Problems

Problem 55 *Find the cross product* $\mathbf{u} \times \mathbf{v}$*, and verify that* $\mathbf{u} \times \mathbf{v}$ *is orthogonal to both* $\mathbf{u}$ *and* $\mathbf{v}$*.*

$\mathbf{u} = \begin{bmatrix} 2 \\ -1 \\ 4 \end{bmatrix}$, $\mathbf{v} = \begin{bmatrix} 0 \\ 3 \\ 1 \end{bmatrix}$.

Answer:

$$\mathbf{u} \times \mathbf{v} = \begin{bmatrix} \ldots \ldots \\ \ldots \ldots \\ \ldots \ldots \end{bmatrix}$$

Problem 56 *Let* $\mathbf{u}$ *and* $\mathbf{v}$ *be vectors in* $\mathbb{R}^3$*, prove that*

$$\mathbf{u} \times \mathbf{v} = -(\mathbf{v} \times \mathbf{u})$$

Problem 57 *Let* $\mathbf{u}$ *and* $\mathbf{v}$ *be vectors of* $\mathbb{R}^3$*. Prove that* $\mathbf{u} \times \mathbf{v}$ *is orthogonal to both* $\mathbf{u}$ *and* $\mathbf{v}$*. (Theorem* **??***)*

Problem 58 *Prove that the cross product of any vector with itself is the zero vector.*

12 Parametric Equations of Lines

At this point you should be very familiar with graphing linear equations of the form $y = mx + b$, where m is the slope of the line and b is the y-intercept. Unfortunately, there is no equivalent way of representing lines in $\mathbb{R}^3$, or better yet, $\mathbb{R}^n$. In this section we will develop an alternative way of representing lines.

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

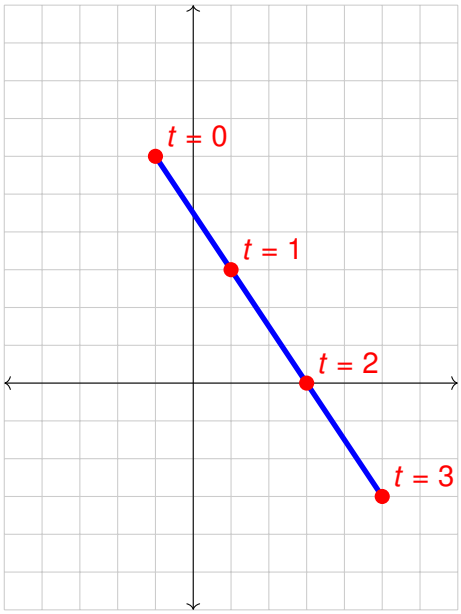
Exploration 59 *Imagine a ladybug crawling around a plane. At every instant in time, the ladybug's position in the plane can be described by an ordered pair* (x, y)*. Coordinates* x *and* y *are functions of time. Suppose the position of the ladybug at time* t *is given by:*

$$x = 2t - 1$$
$$y = -3t + 6$$

Our goal is to sketch the path of the ladybug during the time interval $0 \leq t \leq 3$*. To do this we will make a table of values, just like you did when you first started sketching graphs of functions.*

t	x	y
0	-1	6
1	1	3
2	3	0
3	5	-3

The point (x, y)*, corresponding to each value of* t *is the location of the ladybug in the coordinate plane at time* t*.*



It appears that the points corresponding to $t = 0, 1, 2, 3$ *lie on a line with slope* $-\frac{3}{2}$*. In the next problem we will find the equation of the line that contains the path of the ladybug.*

Exploration 60 *In Exploration* **??** *we considered equations*

$$x = 2t - 1$$
$$y = -3t + 6$$

that described the path of a ladybug in the plane. We will now consider these equations in a broader context as simply describing a curve in the plane. We conjectured that the curve described by these equations is a line with slope $-\frac{3}{2}$*. We will now find the equation of this line.*

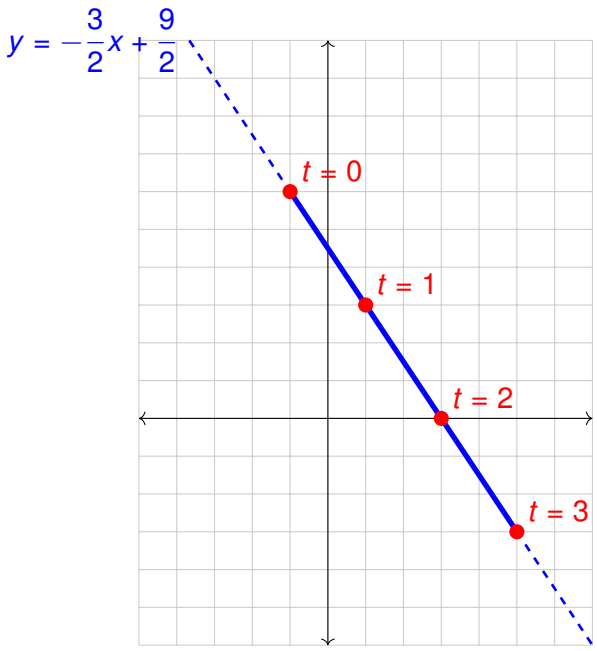
One approach to finding the equation is to solve one of the given equations for t*, then substitute into the other equation. Solving* $x = 2t - 1$ *for* t *gives us*

$$t = \frac{1}{2}x + \frac{1}{2}$$

Substituting this expression into $y = -3t + 6$ *we get*

$$y = -3\left(\frac{1}{2}x + \frac{1}{2}\right) + 6 = -\frac{3}{2}x + \frac{9}{2}$$

Plotting the line $y = -\frac{3}{2}x + \frac{9}{2}$ *together with the path we plotted earlier, we see that our original path lies on the line.*



If we do not restrict t to values between 0 and 3, we can get every point on the line $y = -\frac{3}{2}x + \frac{9}{2}$.

Equations such as

$$x = 2t - 1$$
$$y = -3t + 6$$

are called *parametric equations*, and t is called a *parameter*.

When given an equation of the form $y = mx + b$, we recognize it as an equation whose graph is a line and we don't need to make a table of values to sketch the graph of the equation. We should be able to do the same for parametric equations of lines. In the next Exploration we will examine our equations carefully to see if we can discern any patterns that would help us plot the line without making a table of values.

Exploration 61 *Let's return to parametric equations from the previous problem.*

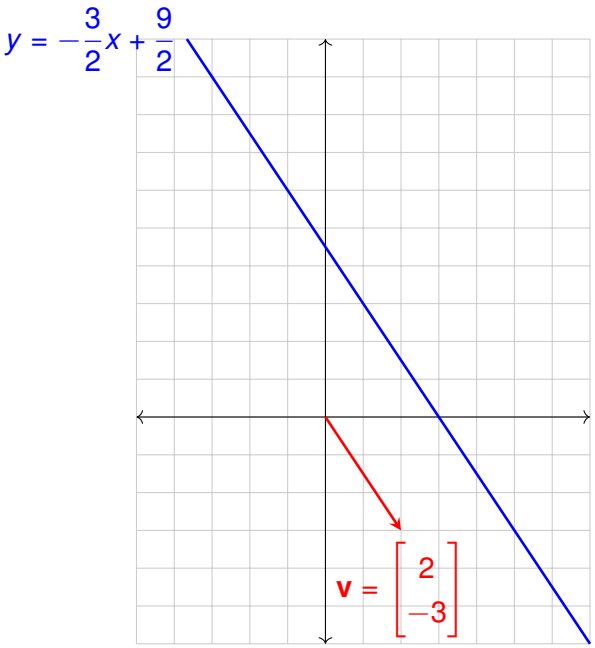
$$x = 2t - 1$$
$$y = -3t + 6$$

Consider the table of values we constructed in Exploration **??**.

t	x	y
0	-1	6
1	1	3
2	3	0
3	5	-3

Observe that every time t increases by 1, x increases by 2 and y decreases by 3. An increase by 2 in the x-coordinate combined with a decrease of 3 in the y-coordinate corresponds to the coefficients 2 and -3 of t in the parametric equations. Our findings are in agreement with the fact that the slope of the line is $-\frac{2}{3}$.

We will capture the "rise" and "run" aspect of the line by using vector $\mathbf{v} = \begin{bmatrix} 2 \\ -3 \end{bmatrix}$. *Sketching* $\mathbf{v}$ *together with the line, we find that the vector is parallel to the line.*



Vector $\mathbf{v}$ *is called a direction vector. Observe that the components of* $\mathbf{v}$ *are the coefficients of t in the parametric equations.*

$$x = 2t - 1$$
$$y = -3t + 6$$

Next, let's turn our attention to constants -1 *and* 6*. They are the values of* x *and* y *when* $t = 0$*, and correspond to the point* $(-1, 6)$*.*

Recall that in Exploration **??** *we said that the given parametric equations describe the position of a ladybug crawling in the coordinate plane. If we are concerned with the position of the ladybug at time* t*, the point* $(-1, 6)$ *is very important - this is where the bug is located when* $t = 0$*. But, if we are using parametric equations simply to describe the line* $y = -\frac{3}{2}x + \frac{9}{2}$*, without regard for when the bug is located at each point, then the point* $(-1, 6)$ *is not any more special than any other point on the line. In fact, we can use any other point on this line to find another set of parametric equations that describe the same line. In Practice Problem* **??** *you will be asked to show that equations*

$$x = 2t + 1$$
$$y = -3t + 3$$

describe the same line. Thus, parametric representations are not unique.

We can generalize our observations in Exploration **??** as follows.

Formula 6. Let $\mathbf{v} = \begin{bmatrix} a \\ b \end{bmatrix}$ be a direction vector for line l, and let (x_0, y_0) be an arbitrary point on l. Then the following parametric equations describe l:

$$x = at + x_0$$
$$y = bt + y_0$$

Example 15. Find parametric equations for line l if l contains the point $(2, -5)$ and is parallel to line k with parametric equations

$$x = 3t + 4$$
$$y = -8t + 1$$

Explanation. To find a set of parametric equations for l, we need a point on l and a direction vector. Point $(2, -5)$ is on l. Line l is parallel to k, so we can use $\begin{bmatrix} 3 \\ -8 \end{bmatrix}$, a direction vector for k, as a direction vector for l. This gives us the following parametric equations for l:

$$x = 3t + 2$$
$$y = -8t - 5$$

Parametric Lines in $\mathbb{R}^3$

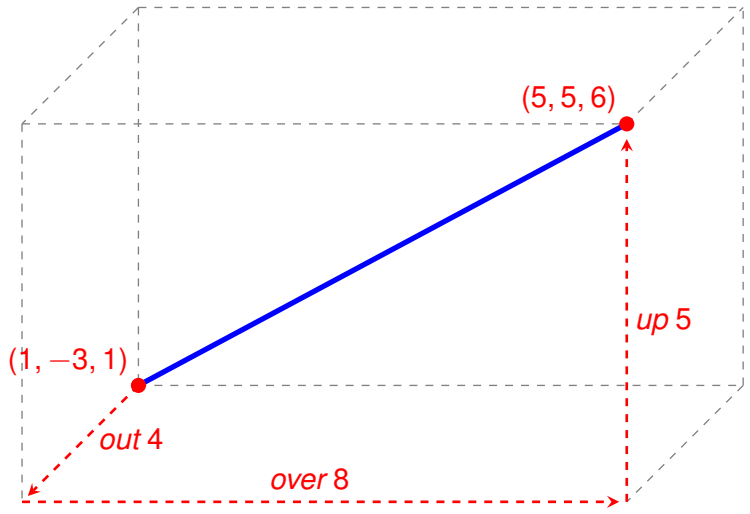
Exploration 62 *Consider the following set of parametric equations*

$$\begin{aligned} x &= 4t + 1 \\ y &= 8t - 3 \\ z &= 5t + 1 \end{aligned}$$

We will begin by making a table of values.

t	x	y	z
-1	-3	-11	-4
0	1	-3	1
1	5	5	6
2	9	13	11

Observe that every time t increases by 1, x increases by 4, y increases by 8, and z increases by 5. The pattern "out 4, over 8, up 5" is illustrated in the diagram.



We will use a direction vector $\mathbf{v} = \begin{bmatrix} 4 \\ 8 \\ 5 \end{bmatrix}$ *to represent the direction of the line. Note that the components of* $\mathbf{v}$ *are the same as the coefficients of t in the parametric equations.*

$$\begin{aligned} x &= 4t + 1 \\ y &= 8t - 3 \\ z &= 5t + 1 \end{aligned}$$

Also, observe that the constants $1, -3, 1$ *correspond to the coordinates of the point on the line for which* $t = 0$*.*

$$\begin{aligned} x &= 4t + 1 \\ y &= 8t + (-3) \\ z &= 5t + 1 \end{aligned}$$

As before, the point $(1, -3, 1)$ *may be replaced with any other point on the line to produce a different set of parametric equations describing the same line.*

We can generalize our observations in Exploration **??** as follows.

Formula 7. Let $\mathbf{v} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$ be a direction vector for line l, and let (x_0, y_0, z_0) be an arbitrary point on l. Then the following parametric equations describe l:

$$\begin{aligned} x &= at + x_0 \\ y &= bt + y_0 \\ z &= ct + z_0 \end{aligned}$$

Example 16. Find a set of parametric equations for a line in $\mathbb{R}^3$ that passes through $(1, -2, 4)$ and $(-3, 1, 1)$.

Explanation. We need a point and a direction vector. To find the direction vector we will use the "head-tail" formula (Formula **??**) to find the vector whose tail is at $(1, -2, 4)$ and whose head is at $(-3, 1, 1)$. The direction vector is

$$\begin{bmatrix} -3 - 1 \\ 1 - (-2) \\ 1 - 4 \end{bmatrix} = \begin{bmatrix} -4 \\ 3 \\ -3 \end{bmatrix}$$

For our point, we can pick either of the two given points. The equations will differ depending on the point we pick, but they will describe the same line. Remember, parametric representations are not unique! If we choose $(1, -2, 4)$ for our point we get the following set of parametric equations:

$$\begin{aligned} x &= -4t + 1 \\ y &= 3t - 2 \\ z &= -3t + 4 \end{aligned}$$

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

After working with parametric equations of lines in $\mathbb{R}^2$ and $\mathbb{R}^3$, it should be easy to surmise what parametric equations of lines look like for lines in $\mathbb{R}^n$.

Formula 8. Let $\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$ be a direction vector for line l in $\mathbb{R}^n$, and let $(a_1, a_2, \ldots, a_n)$ be an arbitrary point on l. Then the following parametric equations describe l:

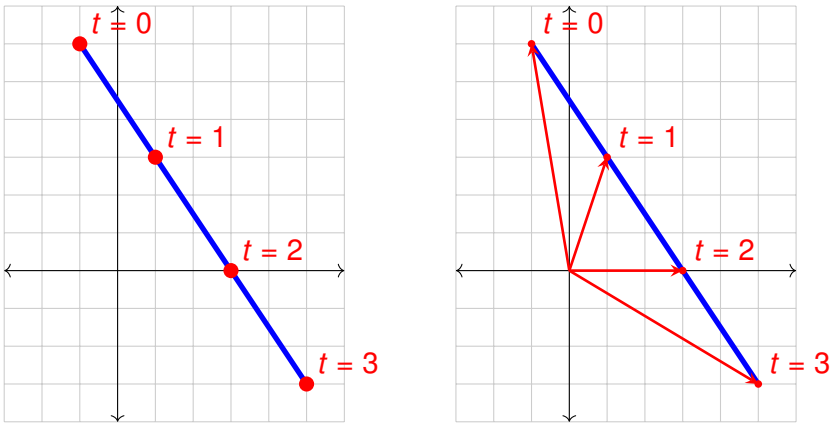
$$\begin{aligned} x_1 &= v_1 t + a_1 \\ x_2 &= v_2 t + a_2 \\ &\vdots \\ x_n &= v_n t + a_n \end{aligned}$$

From Parametric Equations to Vector Equations

Consider again the parametric equations from Exploration **??**.

$$\begin{aligned} x &= 2t - 1 \\ y &= -3t + 6 \end{aligned}$$

These equations describe the line by providing coordinates of individual points on the line, as shown in the left figure below.



If we consider the points on the line to be heads of vectors, then we can think of the line as being "traced" by a vector as t changes (See above, right). We can capture this algebraically by using a *vector equation* for the line:

$$\mathbf{x}(t) = \begin{bmatrix} 2t-1 \\ -3t+6 \end{bmatrix} = \begin{bmatrix} 2 \\ -3 \end{bmatrix} t + \begin{bmatrix} -1 \\ 6 \end{bmatrix}$$

Note that the direction vector $\begin{bmatrix} 2 \\ -3 \end{bmatrix}$ and the point $(-1, 6)$, now vector $\begin{bmatrix} -1 \\ 6 \end{bmatrix}$, are easy to identify in the equation.

We can use vector equations to describe lines in $\mathbb{R}^n$ using the following formula.

Formula 9. Let $\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$ be a direction vector for line l in $\mathbb{R}^n$, and let $(a_1, a_2, \ldots, a_n)$ be an arbitrary point on l. Then the following vector equation describes l:

$$\mathbf{x}(t) = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} t + \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}$$

Practice Problems

Problem 63 *In Explorations* **??**, **??** *and* **??** *you encountered a set of parametric equations*

$$x = 2t - 1$$
$$y = -3t + 6$$

Show that the following parametric equations describe the same line.

$$x = 2t + 1$$
$$y = -3t + 3$$

Problem 64 *Find two different parametric representations for a line with slope* 3 *and y-intercept* $(0, -4)$.

Problem 65 *Suppose line l has a parametric representation:*

$$x_1 = t + 3$$
$$x_2 = t - 4$$
$$x_3 = 3t$$
$$x_4 = 7t + 1$$

Find the missing components for the two vectors below if we know that both vectors can be considered direction vectors for line l.

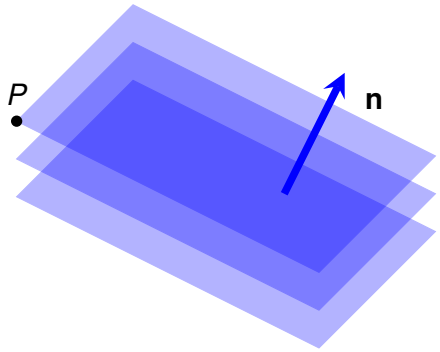
$$\begin{bmatrix} \dots \dots \\ \dots \dots \\ \dots \dots \\ 7 \end{bmatrix}, \begin{bmatrix} -2 \\ \dots \dots \\ \dots \dots \\ \dots \dots \end{bmatrix}$$

13 Planes in $\mathbb{R}^3$

You are probably familiar with the expression "two points determine a line." This means that given two distinct points, there is exactly one line that passes through both of them. There are other ways to describe a line. For example, as we saw in Parametric Equations of Lines, a single point together with a direction vector also determine a line.

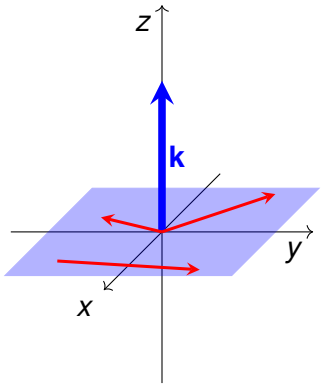
You are probably also familiar with the fact that three non-collinear points determine a plane. (This is why photographers use tripods for stability, while four-legged chairs often wobble!) Is there another way to determine a plane?

The diagram below shows several planes perpendicular to vector **n**. In fact, there are infinitely many such planes. Vector **n** does not determine a plane, but if we know of a point P contained in the plane, then together, P and **n** describe a unique plane.

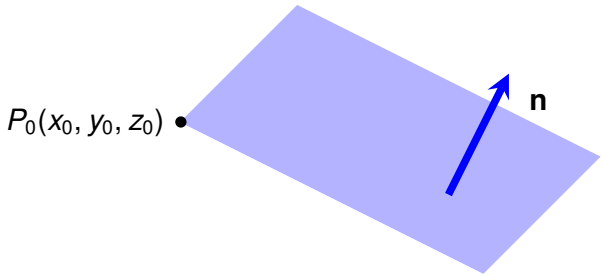


Definition 12 (Normal Vector)**.** A nonzero vector **n** is called a *normal* for a plane if it is orthogonal to every vector in the plane.

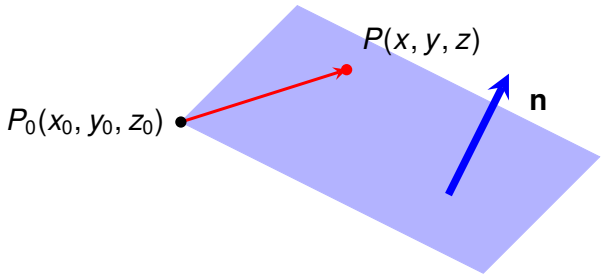
Example 17. The standard unit vector **k** is orthogonal to every vector in the xy-plane, therefore it is a normal for the xy-plane.



Given a point $P_0 = P_0(x_0, y_0, z_0)$ and a nonzero vector $\mathbf{n}$, there is a unique plane through P_0 with normal $\mathbf{n}$.



This fact can be used to give a very simple description of a plane. Observe that a point $P = P(x, y, z)$ lies on this plane if and only if the vector $\overrightarrow{P_0P}$ is orthogonal to $\mathbf{n}$ (i.e. P lies in the plane if and only if $\mathbf{n} \cdot \overrightarrow{P_0P} = 0$).



Let $\mathbf{n} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$. By "head-tail" formula (Formula **??**), we have: $\overrightarrow{P_0P} = \begin{bmatrix} x - x_0 \\ y - y_0 \\ z - z_0 \end{bmatrix}$. So, $P(x, y, z)$ lies in the plane if and only if

$$\begin{bmatrix} a \\ b \\ c \end{bmatrix} \cdot \begin{bmatrix} x - x_0 \\ y - y_0 \\ z - z_0 \end{bmatrix} = a(x - x_0) + b(y - y_0) + c(z - z_0) = 0$$

We summarize this result as a theorem.

Theorem 9. The plane through $P_0(x_0, y_0, z_0)$ with a normal vector

$$\mathbf{n} = \begin{bmatrix} a \\ b \\ c \end{bmatrix} \neq \mathbf{0}$$

is given by

$$a(x - x_0) + b(y - y_0) + c(z - z_0) = 0 \tag{8}$$

Example 18. Find an equation of the plane through $P_0(1,-1,3)$ with $\mathbf{n} = \begin{bmatrix} 3 \\ -1 \\ 2 \end{bmatrix}$ as a normal vector.

Explanation. Here the equation becomes

$$3(x-1)-(y+1)+2(z-3)=0$$

This simplifies to $3x-y+2z=10$.

GeoGebra interactive below shows the plane, together with point P_0 and the normal vector. RIGHT-CLICK and DRAG to rotate the image for a better view.

Access GeoGebra interactives through the online version of this text at

https://ximera.osu.edu/linearalgebradzv3/LinearAlgebraInteractiveIntro.

As demonstrated in Example **??**, we can distribute coefficients a, b and c of equation (**??**) as follows:

$$ax-ax_0+by-by_0+cz-cz_0=0$$

Setting $d = ax_0+by_0+cz_0$, shows that every plane with a normal vector $\mathbf{n} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$ has a linear equation of the form

$$ax+by+cz=d \tag{9}$$

for some constant d. Conversely, the graph of this equation is a plane with $\mathbf{n} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$ as a normal vector (assuming that a, b, and c are not all zero).

Example 19. Find an equation of the plane through $P_0(3,-1,2)$ that is parallel to the plane with equation $2x-3y=6$.

Explanation. The plane with equation $2x-3y=6$ has a normal vector $\mathbf{n} = \begin{bmatrix} 2 \\ -3 \\ 0 \end{bmatrix}$. Because the two planes are parallel, $\mathbf{n}$ serves as a normal for the plane we seek, so the equation is $2x-3y=d$ for some d by (**??**). Insisting that $P_0(3,-1,2)$ lies on the plane determines d; that is, $d = 2\cdot 3-3(-1)=9$. Hence, the equation is $2x-3y=9$.

GeoGebra interactive below shows the two planes, together with point P_0 and the normal vector. RIGHT-CLICK and DRAG to rotate the image for a better view.

Access GeoGebra interactives through the online version of this text at
https://ximera.osu.edu/linearalgebradzv3/LinearAlgebraInteractiveIntro.

Linear Equations and their Graphs: From Lines to Hyperplanes

An equation of the form $ax + by = d$ is a linear equation whose graph is a line in $\mathbb{R}^2$. If we solve for y in terms of x, we obtain a more familiar form of this equation, $y = -\frac{a}{b}x + \frac{d}{b}$. The slope of the corresponding line is $m = -\frac{a}{b}$. Observe that a line perpendicular to this line has the slope $\frac{b}{a}$. If we interpret a as a horizontal "run" and b as a vertical "rise", we see that the vector $\begin{bmatrix} a \\ b \end{bmatrix}$ is perpendicular to the line $ax + by = d$. You can use the following GeoGebra interactive to solidify your understanding of this.

Access GeoGebra interactives through the online version of this text at
https://ximera.osu.edu/linearalgebradzv3/LinearAlgebraInteractiveIntro.

The idea of coefficients in front of variables forming components of a normal vector should be very familiar to you. Recall that the graph of $ax + by + cz = d$ is a plane in $\mathbb{R}^3$ with a normal vector $\mathbf{n} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$.

Lines and planes may seem very different, but they are all graphs of linear equations, just in different dimensions. In general, an equation of the form

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = d$$

is called a *linear equation* in n variables. The graph of such an equation, for $n > 3$, is called a *hyperplane*. The vector in $\mathbb{R}^n$ whose components are the coefficients $a_1, a_2, \ldots, a_n$ is orthogonal to every vector in the hyperplane. Unfortunately, hyperplanes are impossible to see, but we can often use insights we gain from working with lines and planes and generalize them to the invisible world of higher dimensions.

Practice Problems

Problem 66 *Find an equation of the form $ax + by + cz = d$ for the plane described below.*

The plane has a normal vector $\mathbf{n} = \begin{bmatrix} 2 \\ -3 \\ 1 \end{bmatrix}$ *and the plane passes through the point* $(4, 4, -3)$.

Answer:

$$\ldots\ldots x + \ldots\ldots y + \ldots\ldots z = \ldots\ldots$$

Problem 67 *Find an equation of the form* $ax + by + cz = d$ *for the plane described below.*

The plane contains the point $(-1, 3, 0)$ *and is parallel to the plane described by* $2x - 5y + 4z = 7$.

Answer:

$$\ldots\ldots x + \ldots\ldots y + \ldots\ldots z = \ldots\ldots$$

Problem 68 *Find an equation of the form* $ax + by + cz = d$ *for the plane described below.*

The plane contains the point $(2, 0, 5)$ *and is parallel to the* xz*-plane.*

Answer:

$$\ldots\ldots x + \ldots\ldots y + \ldots\ldots z = \ldots\ldots$$

Problem 69 *How many planes satisfying the given conditions are there?*

Containing points $(1, 2, 3)$ *and* $(4, 5, 6)$*, and having a normal vector* $\mathbf{n} = \begin{bmatrix} -9 \\ -9 \\ -9 \end{bmatrix}$.

Multiple Choice:

1. *0*
2. *1*
3. *Infinitely Many*

Problem 70 *How many planes satisfying the given conditions are there?*

Containing $(0, 2, 4)$*,* $(-1, 5, 3)$*,* $(4, 9, 1)$.

Multiple Choice:

1. *0*
2. *1*

3. *Infinitely Many*

Problem 71 *How many planes satisfying the given conditions are there?*

Planes containing $(-1, 3, 5)$*,* $(1, 3, 6)$*,* $(3, 3, 7)$*.*

Multiple Choice:

1. *0*
2. *1*
3. *Infinitely Many*

Text Source

The text in this section is an adaptation of Section 4.2 of Keith Nicholson's *Linear Algebra with Applications*. (CC-BY-NC-SA)

W. Keith Nicholson, *Linear Algebra with Applications*, Lyryx 2018, Open Edition, p 233-234.

14 Systems of Linear Equations

15 Introduction to Systems of Linear Equations

You are probably familiar with the concept of a *system of linear equations* and with some methods for solving such systems. In this section, we will look at the algebra and geometry of finding and interpreting solutions of systems of linear equations. We will start with two-variable and three-variable systems, then move on to systems involving more variables.

Algebra of Linear Systems

When you were first introduced to systems of equations, you learned to solve for one variable in terms of the other(s), then substitute. Here, we will introduce another method. This alternative method involves adding multiples of one equation to another equation in order to eliminate one of the variables. This method will form the foundation for an algorithm we will develop for solving linear systems and performing other computations related to systems. Exploration **??** illustrates how the second method works.

Exploration 72 *The purpose of this problem is to formalize what you may already know (perhaps under a different name) about elementary row operations as means of solving systems of linear equations. Consider the system*

$$\begin{aligned} 2x &- y &&= -4 \\ 3x &+ 2y &&= 1 \end{aligned}$$

We will begin by adding twice the first row to the second row, and replacing the second row with the sum.

$$R_2 + 2R_1 \to R_2$$

$$\begin{aligned} 2x &- y &&= -4 \\ 7x &+ 0y &&= -7 \end{aligned}$$

Note that this step eliminates y from the second equation.

Next we divide both sides of the second equation by 7.

$$\frac{1}{7}R_2 \to R_2$$

$$\begin{aligned} 2x &- y = -4 \\ x & \quad\ = -1 \end{aligned}$$

We now know what x is. Our next goal is to eliminate x from the first equation. To this end, we subtract twice the second row from the first row and replace the first row with the difference.

$$R_1 - 2R_2 \to R_1$$

$$\begin{aligned} 0x &- y = -2 \\ x & \quad\ = -1 \end{aligned}$$

Next we multiply both sides of the first equation by -1.

$$-R_1 \to R_1$$

$$\begin{aligned} & y = 2 \\ x & \quad\ = -1 \end{aligned}$$

Finally, we can switch the order of equations in order to display x in the top row.

$$R_1 \leftrightarrow R_2$$

$$\begin{aligned} x & \quad\ = -1 \\ & y = 2 \end{aligned}$$

This solution can be written as an ordered pair $(-1, 2)$.

In Exploration **??** we introduced elementary row operations and the notation associated with them. We now make these definitions formal.

Definition 13 (Elementary Row Operations)**.** The following three operations performed on a linear system are called *elementary row operations*.

1. Switching the order of equations (rows) i and j:

$$R_i \leftrightarrow R_j$$

2. Multiplying both sides of equation (row) i by the same non-zero constant, k, and replacing equation i with the result:

$$kR_i \to R_i$$

3. Adding k times equation (row) i to equation (row) j, and replacing equation j with the result:

$$R_j + kR_i \to R_j$$

As we applied elementary row operations to the system in Exploration **??**, the system changed, but a quick check will convince you that all six systems have the same solution: $(-1, 2)$. The six systems are said to be *equivalent*.

It turns out that if a system of equations is transformed into another system through a sequence of elementary row operations, the new system will be equivalent to the original system, in other words, both systems will have the same solution(s). We will formalize this statement as Theorem **??** at the end of this section.

Example 20. Solve the system of equations using elementary row operations.

$$\begin{array}{rcrcrcr} 3x & - & y & + & z & = & 0 \\ 2x & + & y & + & 2z & = & 2 \\ x & + & 4y & - & 2z & = & 11 \end{array}$$

Explanation. It may be daunting to think about how to begin. But keep in mind the desired end-result. What we want is to use elementary row operations to transform the given system into something like this

$$\begin{array}{rrrcr} x & & & = & a \\ & y & & = & b \\ & & z & = & c \end{array}$$

We will accomplish this by using a convenient variable in one row to "wipe out" this variable from the other two rows. For example, we can use x in the third equation to wipe out $3x$ in the first equation and $2x$ in the second equation. To do this, multiply the third row by -3 and add it to the top row, then multiply the third row by -2 and add it to the second row. We now have:

$$\begin{array}{rcrcrcr} \xrightarrow{R_1-3R_3} & 0x & - & 13y & + & 7z & = & -33 \\ \xrightarrow{R_2-2R_3} & 0x & - & 7y & + & 6z & = & -20 \\ & x & + & 4y & - & 2z & = & 11 \end{array}$$

In the previous step x was a convenient variable to use because the coefficient in front of x was 1. We no longer have a variable with coefficient 1. We could create a coefficient of 1 using division, but that would lead to fractions, making computations cumbersome. Instead, we will subtract twice the second row from the first row. This gives us:

$$\begin{array}{rcrcrcr} \xrightarrow{R_1-2R_2} & 0x & + & y & - & 5z & = & 7 \\ & 0x & - & 7y & + & 6z & = & -20 \\ & x & + & 4y & - & 2z & = & 11 \end{array}$$

Next we add seven times the first row to the second row, and subtract four times the first row from the third row.

$$\begin{array}{rcl} & \begin{array}{rcrcrcr} 0x & + & y & - & 5z & = & 7 \end{array} \\ \xrightarrow{R_2+7R_1} & \begin{array}{rcrcrcr} 0x & + & 0y & - & 29z & = & 29 \end{array} \\ \xrightarrow{R_3-4R_1} & \begin{array}{rcrcrcr} x & + & 0y & + & 18z & = & -17 \end{array} \end{array}$$

Now we divide both sides of the second row by -29.

$$\begin{array}{rcl} & \begin{array}{rcrcrcr} 0x & + & y & - & 5z & = & 7 \end{array} \\ \xrightarrow{-\frac{1}{29}R_2} & \begin{array}{rcrcrcr} 0x & + & 0y & + & z & = & -1 \end{array} \\ & \begin{array}{rcrcrcr} x & + & 0y & + & 18z & = & -17 \end{array} \end{array}$$

Adding 5 times the second row to the first row and subtracting 18 times the second row from the third row gives us

$$\begin{array}{rcl} \xrightarrow{R_1+5R_2} & \begin{array}{rcrcrcr} 0x & + & y & + & 0z & = & 2 \end{array} \\ & \begin{array}{rcrcrcr} 0x & + & 0y & + & z & = & -1 \end{array} \\ \xrightarrow{R_3-18R_2} & \begin{array}{rcrcrcr} x & + & 0y & + & 0z & = & 1 \end{array} \end{array}$$

Finally, rearranging the rows gives us

$$\begin{array}{rcrcrcr} x & + & 0y & + & 0z & = & 1 \\ 0x & + & y & + & 0z & = & 2 \\ 0x & + & 0y & + & z & = & -1 \end{array}$$

$$\begin{array}{rcrcr} x & & & = & 1 \\ & y & & = & 2 \\ & & z & = & -1 \end{array}$$

Thus the system has a unique solution $(1, 2, -1)$.

At this point you may be wondering whether it will always be possible to take a system of three equations and three unknowns and use elementary row operations to transform it to a system of the form

$$\begin{array}{rcrcr} x & & & = & a \\ & y & & = & b \\ & & z & = & c \end{array}$$

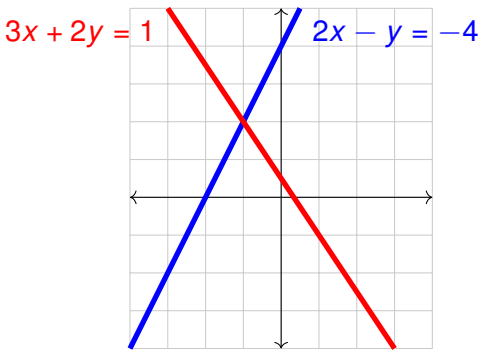
The short answer to this question is NO. The existence of an equivalent system of this form implies that the original system has a unique solution (a, b, c). However, it is possible for a system to have no solutions or to have infinitely many solutions. We will study these different possibilities from an algebraic perspective in subsequent sections. For now, we will attempt to gain insight into existence and uniqueness of solutions through geometry.

Geometry of Linear Systems in Two Variables

Exploration Problem **??** offers an example of a linear system of two equations and two unknowns (variables) with a unique solution.

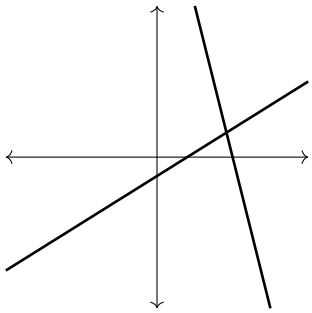
$$\begin{aligned} 2x &- y &&= -4 \\ 3x &+ 2y &&= 1 \end{aligned}$$

Geometrically, the graph of each equation is a line in $\mathbb{R}^2$. The point $(-1, 2)$ is a solution to both equations, so it must lie on both lines. The graph below shows the two lines intersecting at $(-1, 2)$.

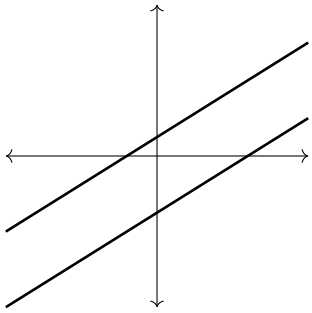


Given a system of two equations with two unknowns, there are three possible geometric outcomes.

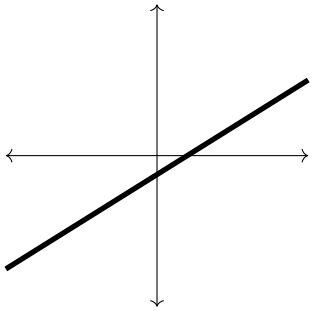
- First, the graphs of the two equations intersect at a point. If this is the case, the system has exactly one solution. We say that the system is *consistent* and has a *unique solution*.



- Second, the two lines may have no points in common. If this is the case, the system has no solutions. We say that the system is *inconsistent*.



- Finally, the two lines may coincide. In this case, there are infinitely many points that satisfy both equations simultaneously. We say that the system is consistent and has infinitely many solutions.



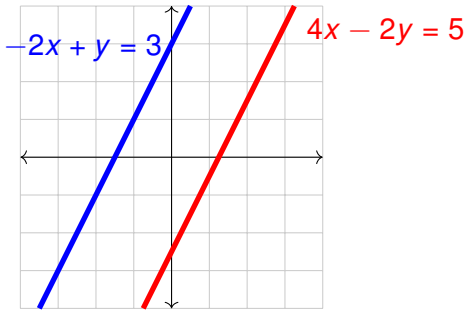
Example 21. Solve the system of equations and interpret your results geometrically.

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

Explanation. We will use elementary row operations. Adding twice the first equation to the second equation gives us

$$\begin{array}{rcrcll} -2x & + & y & = & 3 & \\ 0x & - & 0y & = & 11 & \leftarrow \text{Never true} \end{array}$$

This is where we run into a problem: there are no values of x and y that satisfy the second equation. We conclude that the system is inconsistent. Plotting the two lines in the same coordinate plane shows that the two lines are parallel.



Example 22. Solve the system of equations and interpret your results geometrically.

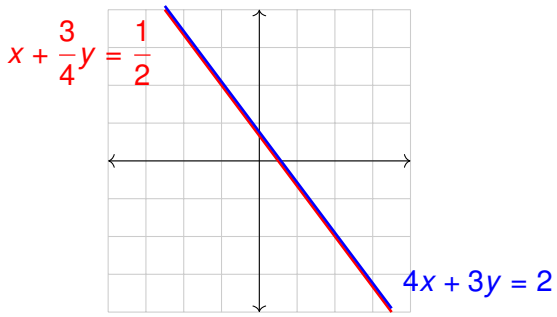
$$\begin{array}{rcrcl} 4x & + & 3y & = & 2 \\ x & + & \frac{3}{4}y & = & \frac{1}{2} \end{array}$$

Explanation. To eliminate x from the second equation, we subtract one quarter of the first equation from the second. This gives us

$$\begin{array}{rcrcll} 4x & + & 3y & = & 2 & \\ 0x & + & 0y & = & 0 & \leftarrow \text{Always true} \end{array}$$

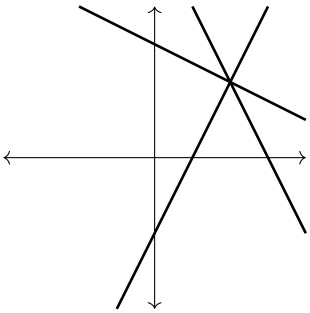
Unlike the situation in Example **??**, any combination of x and y satisfies the second equation. So, any ordered pair (x, y) that satisfies the first equation will satisfy the second equation. Thus, the solution set for this system is the same as the set of all solutions of $4x + 3y = 2$.

When we plot the two equations of the original system, we find that the two lines coincide.

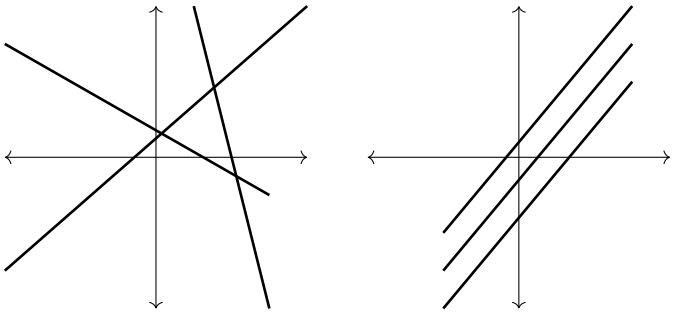


Given a linear system in two variables and more than two equations, we have a variety of geometric possibilities. In terms of the number of solutions, there are three possibilities.

- First, it is possible for the graphs of all equations in the system to intersect at a single point, giving us a unique solution.



- Second, it is possible for the graphs to have no points common to all of them. If this is the case, the system is inconsistent.



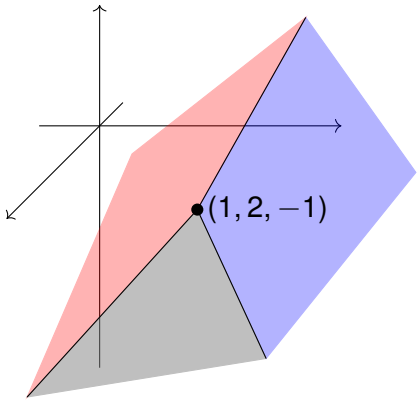
- Finally, it is possible for all of the lines to coincide, giving us infinitely many solutions.

Geometry of Linear Systems in Three Variables

In Example **??** we solved the following linear system of three equations and three unknowns

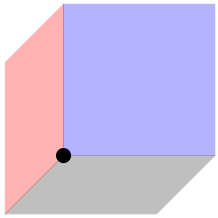
$$\begin{array}{rcrcrcr} 3x & - & y & + & z & = & 0 \\ 2x & + & y & + & 2z & = & 2 \\ x & + & 4y & - & 2z & = & 11 \end{array}$$

We found that the system has a unique solution $(1, 2, -1)$. The graph of each equation is a plane. The three planes intersect at a single point, as shown in the figure.

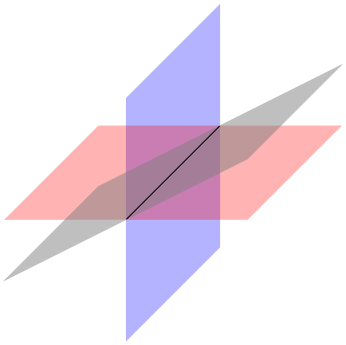


Given a linear system of three equations in three variables, there are three ways in which the system can be consistent.

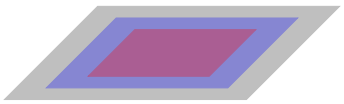
- First, the three planes could intersect at a single point, giving us a unique solution.



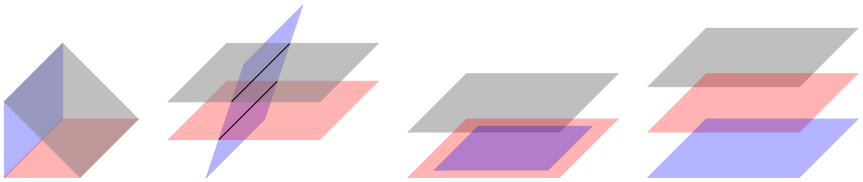
- Second, the three planes can intersect in a line, forming a paddle-wheel shape. In this case, every point along the line of intersection is a solution to the system, giving us infinitely many solutions.



- Finally, the three planes can coincide. If this is the case, there are infinitely many solutions.



There are four ways for a system to be inconsistent. They are depicted below.



General Systems of Linear Equations

Definition 14. A *linear equation* in variables $x_1, \ldots, x_n$ is an equation that can be written in the form

$$a_1x_1 + a_2x_2 + \ldots + a_nx_n = b$$

where $a_1, \ldots, a_n$ and b are constants.

An n-tuple $(x_1, x_2, \ldots, x_n)$ is a *solution* to the equation $a_1x_1 + a_2x_2 + \ldots + a_nx_n = b$ provided that it turns the equation into a true statement. The set of all n-tuples that are solutions to a given equation is called the *graph* of the equation. The graph of a linear equation in two variables is a line in $\mathbb{R}^2$. The graph of a linear equation in three variables is a plane in $\mathbb{R}^3$. In $\mathbb{R}^n$, for $n > 3$, we say that the graph of a linear equation is a *hyperplane*. A hyperplane cannot be visualized, but we can still talk about intersections of hyperplanes and their other attributes in algebraic terms.

A linear system of m equations and n unknowns is typically written as follows

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

A *solution to a system* of linear equations in n variables is an n-tuple that satisfies every equation in the system. All solutions to a system of equations, taken together, form a *solution set*.

Definition 15. Two systems of linear equations are said to be equivalent if they have the same solution set.

Recall that to solve systems of equations in this section, we utilized three *elementary row operations*. These operations are:

1. Switching the order of two equations
2. Multiplying both sides of an equation by the same non-zero constant
3. Adding a multiple of one equation to another

Theorem 10. Given a system of linear equations, any of the three elementary row operations performed on the system produces an equivalent system.

Proof Clearly, the order of equations does not affect the solution set, so **??** produces an equivalent system. Next, you learned years ago that multiplying both sides of an equation by a non-zero constant does not change its solution set, which establishes that **??** produces an equivalent system. To see that **??** produces an equivalent system, note that if we add a multiple of an equation to another equation in the system, we are adding the same thing to both sides, which does not change the solution set of that equation, nor of the system. ■

Practice Problems

Problem 73 *Give a graphical illustration (draw a picture) for each of the following scenarios for a system of THREE equations and TWO unknowns:*

1. *The system of three equations is inconsistent, but a combination of any two of the three equations forms a consistent system.*
2. *The system is consistent and has a unique solution.*

3. *The system is consistent and has infinitely many solutions.*

4. *The system is inconsistent and no two equations form a consistent system.*

Problem 74 *Solve the given system of linear equations algebraically or algebraically demonstrate that a solution does not exist. Interpret your results geometrically.*

$$\begin{aligned} x &+ 3y = 4 \\ x &- 2y = -6 \end{aligned}$$

Solution:

(... ... , ...)

Problem 75 *Solve the given system of linear equations algebraically or algebraically demonstrate that a solution does not exist. Interpret your results geometrically.*

$$\begin{aligned} -3x &+ 2y = 7 \\ 6x &- 4y = 5 \end{aligned}$$

Problem 76 *Solve the given system of linear equations algebraically or algebraically demonstrate that a solution does not exist. Interpret your results geometrically.*

$$\begin{aligned} x &- 2y + z = 0 \\ 3x &- 2y + 4z = 2 \\ 2x &- y + 2z = 3 \end{aligned}$$

Solution:

(... ... , ... , ...)

Problem 77 *Consider the following system of equations.*

$$\begin{aligned} kx &+ 8y = 4 \\ 2x &+ ky = -2 \end{aligned}$$

1. *Find all possible values of k such that this system has no solutions.*

   Answer:

   $k =$ ... ...

2. *Find all possible values of k such that this system has infinitely many solutions.*

   Answer:

$$k = \ldots\ldots$$

Problem 78 *Why is there a non-zero provision in Part* **??** *of Definition* **??***? Why is there not a non-zero provision in Part* **??***?*

Problem 79 *Suppose the following system was obtained from system* (A) *by adding twice the second row of* (A) *to the first row.*

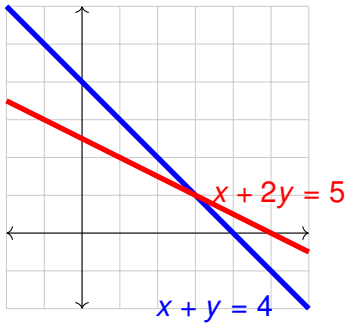
$$\begin{aligned} 8x &+ 3y = 11 \\ 3x &+ 2y = 5 \end{aligned}$$

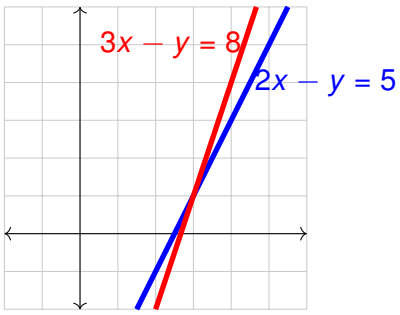
Find system (A)*.*

Answer:

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

Problem 80 *The following figures show a geometric depiction of two equivalent systems. (The systems are equivalent because they have the same solution set.) Can the first system be transformed into the second system by elementary row operations? If so, how?*





Problem 81 *Consider the system of equations*

$$\begin{aligned} ax &+ by = e \\ cx &+ dy = f \end{aligned}$$

Show that if (x_0, y_0) *is a solution to this system, and if we apply elementary row operation* **??** *to the system, then* (x_0, y_0) *will be a solution to the new system of equations.*

Problem 82 *Demonstrate that elementary row operations are reversible by answering the following questions. Be specific about the elementary row operation that you would use.*

1. *Suppose we obtained system (B) from system (A) by swapping two equations. How would we obtain system (A) from system (B)?*
2. *Suppose we obtained system (B) from system (A) by multiplying one of the equations of (A) by a non-zero constant k. How would we obtain system (A) from system (B)?*
3. *Suppose we obtained system (B) from system (A) by adding a multiple of one of the equations of (A) to another. How would we obtain system (A) from system (B)?*

16 Augmented Matrix Notation and Elementary Row Operations

Augmented Matrix Notation

Recall that the following three operations performed on a linear system are called *elementary row operations*

- Switching the order of two equations
- Multiplying both sides of an equation by the same non-zero constant
- Adding a multiple of one equation to another

In Introduction to Systems of Linear Equations we discussed why applying elementary row operations to a linear system results in an equivalent system - a linear system with the same solution set.

In this section we seek an efficient method for recording our computations as we perform elementary row operations.

Exploration 83 *Consider the linear system*

$$\begin{array}{rcrcrcrcl} x & - & y & & & & & = & 0 \\ 2x & - & 2y & + & z & + & 2w & = & 4 \\ & & y & & & + & w & = & 0 \\ & & & & 2z & + & w & = & 5 \end{array} \tag{10}$$

Our goal is to use elementary row operations to transform this system into an equivalent system of the form

$$\begin{array}{rcrcrcl} x & & & & & = & a \\ & & y & & & = & b \\ & & & z & & = & c \\ & & & & w & = & d \end{array} \tag{11}$$

We have to keep in mind that given an arbitrary system, an equivalent system of this form may not exist (we will talk a lot more about this later). However, it does exist in this case, and we would like to find a more efficient way of getting to it than having to write and rewrite our equations at each step.

*In this problem, we prompt you to perform elementary row operations on (**??**) and ask you to fill in the coefficients in the resulting equations. This is a multi-step process. Steps will unfold automatically as you enter correct answers.*

Problem 83.1 *We start by subtracting twice row 1 from row 2. ($R_2 - 2R_1 \to R_2$)*

$$\begin{array}{rcrcrcrcl} x & - & y & + & 0z & + & 0w & = & 0 \\ 0x & + & 0y & + & 1z & + & 2w & = & 4 \\ 0x & + & y & + & 0z & + & w & = & 0 \\ 0x & + & 0y & + & 2z & + & w & = & 5 \end{array}$$

Next, we add row 3 to row 1. ($R_1 + R_3 \to R_1$)

Problem 83.1.1 Problem 83.1.1.1 Problem 83.1.1.1.1

We now exchange rows 2 and 3. ($R_2 \leftrightarrow R_3$)

$$\begin{array}{rcrcrcrcl} x & + & 0y & + & 0z & + & 0w & = & -1 \\ 0x & + & y & + & 0z & + & 0w & = & -1 \\ 0x & + & 0y & + & z & + & 0w & = & 2 \\ 0x & + & 0y & + & 0z & + & w & = & 1 \end{array} \tag{12}$$

If we drop all of the zero terms, we have:

$$\begin{array}{rrrrrcl} x & & & & & = & -1 \\ & y & & & & = & -1 \\ & & z & & & = & 2 \\ & & & & w & = & 1 \end{array} \tag{13}$$

Now we see that $(-1, -1, 2, 1)$ *is the solution.*

*Observe that throughout the entire process, variables x, y, z and w remained in place; only the coefficients in front of the variables and the entries on the right changed. Let's try to recreate this process without writing down the variables. We can capture the original system in (**??**) as follows:*

$$\left[\begin{array}{cccc|c} 1 & -1 & 0 & 0 & 0 \\ 2 & -2 & 1 & 2 & 4 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 2 & 1 & 5 \end{array}\right]$$

The side to the left of the vertical bar is called the coefficient matrix, while the side to the right of the bar is a vector that consists of constants on the right side of the system. The coefficient matrix, together with the vector, is called an augmented matrix.

We can capture all of the elementary row operations we performed earlier as follows:

$$\left[\begin{array}{cccc|c} 1 & -1 & 0 & 0 & 0 \\ 2 & -2 & 1 & 2 & 4 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 2 & 1 & 5 \end{array}\right] \xrightarrow{R_2-2R_1} \left[\begin{array}{cccc|c} 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 2 & 4 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 2 & 1 & 5 \end{array}\right] \xrightarrow{R_1+R_3}$$

$$\left[\begin{array}{cccc|c} 1 & 0 & 0 & 1 & 0 \\ 2 & -2 & 1 & 2 & 4 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 2 & 1 & 5 \end{array}\right] \xrightarrow{R_4-2R_2} \left[\begin{array}{cccc|c} 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 2 & 4 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & -3 & -3 \end{array}\right] \xrightarrow{(-1/3)R_4}$$

$$\left[\begin{array}{cccc|c} 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 2 & 4 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{array}\right] \begin{array}{l} \xrightarrow{R_1-R_4} \\ \xrightarrow{R_2-2R_4} \\ \xrightarrow{R_3-R_4} \end{array} \left[\begin{array}{cccc|c} 1 & 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 & 2 \\ 0 & 1 & 0 & 0 & -1 \\ 0 & 0 & 0 & 1 & 1 \end{array}\right] \xrightarrow{R_2\leftrightarrow R_3}$$

$$\left[\begin{array}{cccc|c} 1 & 0 & 0 & 0 & -1 \\ 0 & 1 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 & 2 \\ 0 & 0 & 0 & 1 & 1 \end{array}\right] \tag{14}$$

*The last augmented matrix corresponds to systems in (**??**) and (**??**), and we can easily see the solution.*

Exploration **??** introduced us to some vocabulary terms. Let's formalize our definitions. Every linear system

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

can be written in the *augmented matrix form* as follows:

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

The array to the left of the vertical bar is called the *coefficient matrix* of the linear system and is often given a capital letter name, like A. The vertical array to the right of the bar is called a *constant vector*.

$$A = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{bmatrix} \quad \text{and} \quad \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$$

We will sometimes use the following notation to represent an augmented matrix.

$$\left[\begin{array}{c|c} A & \mathbf{b} \end{array}\right]$$

The same elementary row operations that we perform on a system of equations can be performed on the corresponding augmented matrix, or any matrix for that matter. If a matrix can be obtained from another matrix by means of elementary row operations, we say that the two matrices are *row-equivalent*.

Exploration 84 *Consider the system*

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

Recall that in Exploration **??** *we converted the given system to an augmented matrix form, then performed elementary row operations until we arrived at a "convenient" form. We then converted the "convenient" augmented matrix back to a system of equations and identified the solution. The term "convenient" is open to interpretation. In this problem we will explore two "convenient" forms. Each one will lead to a definition.*

$$
\begin{array}{r}
\\
\xrightarrow{R_2-4R_1} \\
\xrightarrow{R_3+R_1} \\
\\
\xrightarrow{(1/2)R_2} \\
\\
\xrightarrow{R_1+R_2} \\
\\
\xrightarrow{R_1+(1/2)R_3} \\
\xrightarrow{R_2-(5/2)R_3} \\
\\
\end{array}
\left[\begin{array}{ccc|c}
1 & -1 & -3 & 1 \\
4 & -2 & -7 & 1 \\
-1 & 1 & 4 & 1
\end{array}\right]
\left[\begin{array}{ccc|c}
1 & -1 & -3 & 1 \\
0 & 2 & 5 & -3 \\
0 & 0 & 1 & 2
\end{array}\right] \quad (15)
$$

$$
\left[\begin{array}{ccc|c}
1 & -1 & -3 & 1 \\
0 & 1 & 5/2 & -3/2 \\
0 & 0 & 1 & 2
\end{array}\right]
\left[\begin{array}{ccc|c}
1 & 0 & -1/2 & -1/2 \\
0 & 1 & 5/2 & -3/2 \\
0 & 0 & 1 & 2
\end{array}\right]
$$

$$
\left[\begin{array}{ccc|c}
1 & 0 & 0 & 1/2 \\
0 & 1 & 0 & -13/2 \\
0 & 0 & 1 & 2
\end{array}\right] \quad (16)
$$

*The augmented matrix in (**??**) has the same convenient form as the one in (**??**). This augmented matrix corresponds to the system*

$$
\begin{aligned}
x_1 &= 1/2 \\
x_2 &= -13/2 \\
x_3 &= 2
\end{aligned}
$$

This gives us the solution $(\frac{1}{2}, -\frac{13}{2}, 2)$.

*While the augmented matrix in (**??**) was certainly "convenient", we could have converted back to the equation format a little earlier. Let's take a look at the augmented matrix in (**??**). Converting (**??**) to a system of equations gives us*

$$
\begin{aligned}
x_1 - x_2 - 3x_3 &= 1 \\
2x_2 + 5x_3 &= -3 \\
x_3 &= 2
\end{aligned}
$$

Substituting $x_3 = 2$ *into the second equation and solving for* x_2 *gives us*

$$
x_2 = \frac{-3 - 5(2)}{2} = -\frac{13}{2}
$$

Now substituting $x_3 = 2$ *and* $x_2 = -\dfrac{13}{2}$ *into the first equation results in*

$$x_1 = 1 + \left(-\frac{13}{2}\right) + 3(2) = \frac{1}{2}$$

This process is called back substitution and it produces the same solution as we obtained earlier.

Observe that the coefficient matrices in (**??**) and (**??**) have the same format: 1's along the diagonal, zeros above and below the 1's. The other "convenient" format, exhibited by the coefficient matrix in (**??**), also has zeros below the diagonal, but not all of the diagonal entries are 1's and some of the entries above the diagonal are not zero. Each of these formats gives rise to a definition. These definitions are the topic of the next section.

Row-Echelon and Reduced Row-Echelon Forms

Definition 16. The first non-zero entry in a row of a matrix (when read from left to right) is called the *leading entry*. When the leading entry is 1, we refer to it as a *leading 1*.

Definition 17 (Row-Echelon Form)**.** A matrix is said to be in *row-echelon form* if:

1. All entries below each leading entry are 0.
2. Each leading entry is in a column to the right of the leading entries in the rows above it.
3. All rows of zeros, if there are any, are located below non-zero rows.

The term row-echelon form can be applied to matrices whether or not they are augmented matrices (matrices with the vertical bar). For example, both the coefficient matrix and the augmented matrix in (**??**) are in row-echelon form. Note that the leading entries form a staircase pattern. All entries below the leading entries are zero, but the entries above the leading entries are not all zero.

Below are two more examples of matrices in row-echelon form. The leading entries of each matrix are boxed.

$$\begin{bmatrix} \boxed{1} & 2 & -1 & -5 & 0 & 2 \\ 0 & 0 & \boxed{3} & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 & \boxed{1} & 0 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} \boxed{4} & 2 \\ 0 & \boxed{1} \\ 0 & 0 \end{bmatrix}$$

The difference between the coefficient matrix in (**??**) and the coefficient matrix in (**??**) is that the leading entries of the matrix in (**??**) are all 1's, and the matrix has zeros above each leading 1. This motivates our next definition.

Definition 18 (Reduced Row-Echelon Form)**.** A matrix that is already in *row-echelon* form is said to be in *reduced row-echelon form* if:

1. Each leading entry is 1
2. All entries *above* and below each leading 1 are 0

The following two matrices are in *reduced row-echelon form*. Note that there are 0' s below and above each leading 1.

$$\begin{bmatrix} \boxed{1} & 0 & -1 & 0 & 1 \\ 0 & \boxed{1} & 0 & 0 & 2 \\ 0 & 0 & 0 & \boxed{1} & 2 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} \boxed{1} & 0 \\ 0 & \boxed{1} \\ 0 & 0 \end{bmatrix}$$

When solving linear systems using the augmented matrix notation, our goal will be to transform the augmented matrix $M = [A|\mathbf{b}]$ into a row-echelon or reduced row-echelon form. The reduced row-echelon form of M is denoted by rref(M). As we transform the augmented matrix M to its reduced row-echelon form, the coefficient matrix A (the matrix to the left of the bar) also gets transformed to its reduced row-echelon form, rref(A).

Example 23. Solve the system of equations or determine that the system is inconsistent.

$$\begin{array}{rcrcrcrcl} x & - & 2y & + & z & & & = & 2 \\ -2x & + & y & - & z & + & w & = & 0 \\ x & + & 4y & & & + & w & = & 1 \end{array}$$

Explanation. We begin by rewriting the system in the augmented matrix form.

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

Our goal is to convert this matrix to its reduced row-echelon form by means of elementary row operations. To do this, we will proceed from left to right and use leading entries to wipe out all entries above and below them.

$$\left[\begin{array}{cccc|c} 1 & -2 & 1 & 0 & 2 \\ -2 & 1 & -1 & 1 & 0 \\ 1 & 4 & 0 & 1 & 1 \end{array}\right] \begin{array}{l} \xrightarrow{R_2+2R_1} \\ \xrightarrow{R_3-R_1} \end{array} \left[\begin{array}{cccc|c} 1 & -2 & 1 & 0 & 2 \\ 0 & -3 & 1 & 1 & 4 \\ 0 & 6 & -1 & 1 & -1 \end{array}\right] \xrightarrow{R_3+2R_2}$$

$$\left[\begin{array}{cccc|c} 1 & -2 & 1 & 0 & 2 \\ 0 & -3 & 1 & 1 & 4 \\ 0 & 0 & 1 & 3 & 7 \end{array}\right] \xrightarrow{(-1/3)R_2} \left[\begin{array}{cccc|c} 1 & -2 & 1 & 0 & 2 \\ 0 & 1 & -1/3 & -1/3 & -4/3 \\ 0 & 0 & 1 & 3 & 7 \end{array}\right] \xrightarrow{R_1+2R_2}$$

$$\left[\begin{array}{cccc|c} 1 & 0 & 1/3 & -2/3 & -2/3 \\ 0 & 1 & -1/3 & -1/3 & -4/3 \\ 0 & 0 & 1 & 3 & 7 \end{array}\right] \begin{array}{c} \xrightarrow{R_1-(1/3)R_3} \\ \xrightarrow{R_2+(1/3)R_3} \end{array} \left[\begin{array}{cccc|c} 1 & 0 & 0 & -5/3 & -3 \\ 0 & 1 & 0 & 2/3 & 1 \\ 0 & 0 & 1 & 3 & 7 \end{array}\right]$$

Our final matrix may not be quite as nice as the one in (**??**), but it is in reduced row-echelon form. Our next step is to convert our augmented matrix back to a system of equations. We have:

$$\begin{array}{rcrcrcl} x & & & - & (5/3)w & = & -3 \\ & y & & + & (2/3)w & = & 1 \\ & & z & + & 3w & = & 7 \end{array}$$

We will rewrite the system as follows:

$$\begin{array}{rcrcl} x & & & = & -3 + (5/3)w \\ & y & & = & 1 - (2/3)w \\ & & z & = & 7 - 3w \end{array}$$

Now we see that we can assign any value to w, then compute x, y and z to obtain a solution to the system. For example, let $w = 0$, then $x = -3$, $y = 1$ and $z = 7$, so $(-3, 1, 7, 0)$ is a solution. If we let $w = 3$, then $x = 2$, $y = -1$ and $z = -2$, so $(2, -1, -2, 3)$ is also a solution. To capture all possibilities, we will let $w = t$, where t is an arbitrary *parameter*.

We can think of the solution set in two different ways. First, the solution set is the set of all points of the form

$$\left(-3 + \frac{5}{3}t, 1 - \frac{2}{3}t, 7 - 3t, t\right)$$

We can also think of the solution set in geometric terms by observing that

$$\begin{aligned} x &= -3 + \frac{5}{3}t \\ y &= 1 - \frac{2}{3}t \\ z &= 7 - 3t \\ w &= t \end{aligned}$$

is a set of parametric equations that describes a line in $\mathbb{R}^4$. (See Formula **??**) This means that the three hyperplanes given by the equations in the system intersect in a line, producing infinitely many solutions to the system.

Observe that in Example **??**, variables x, y and z correspond to the leading $1's$ in the reduced row-echelon. We say that x, y and z are the *leading variables*. Variable w is not a leading variable; we refer to it as a *free variable* and assign a *parameter* t to it.

Example 24. Solve the system of equations or determine that the system is inconsistent.

$$\begin{array}{rcrcrcr} 2x & - & y & & & = & 2 \\ x & + & y & - & 2z & = & 0 \\ -3x & & & + & 2z & = & -1 \end{array}$$

Explanation. We rewrite the system in augmented matrix form and transform it to reduced row-echelon form. We leave the details of the elementary row operations to the reader and state the final result.

$$\left[\begin{array}{ccc|c} 2 & -1 & 0 & 2 \\ 1 & 1 & -2 & 0 \\ -3 & 0 & 2 & -1 \end{array}\right] \rightsquigarrow \left[\begin{array}{ccc|c} 1 & 0 & -2/3 & 0 \\ 0 & 1 & -4/3 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right] \tag{17}$$

Converting back to a system of linear equations, we get

$$\begin{array}{rcrcrcr} 1x & + & 0y & - & (2/3)z & = & 0 \\ 0x & + & 1y & - & (4/3)z & = & 0 \\ 0x & + & 0y & + & 0z & = & 1 \end{array} \tag{18}$$

The last equation in this system clearly has no solutions. We conclude that this system (and the original system) is inconsistent.

Observation 1. Note that the last row of the reduced row-echelon form in (**??**) looks like this

$$\left[\begin{array}{ccc|c} 0 & 0 & 0 & 1 \end{array}\right]$$

This row corresponds to the equation

$$0x + 0y + 0z = 1$$

which clearly has no solutions.

In general, if the reduced row-echelon form of the augmented matrix contains a row

$$\left[\begin{array}{ccc|c} 0 & \ldots & 0 & 1 \end{array}\right]$$

we can conclude that the system is inconsistent.

Example 25. Solve the system of equations or determine that the system is inconsistent.

$$\begin{array}{rcrcrcr} x & - & 3y & - & z & = & 3 \\ -3x & + & 5y & & & = & -2 \\ x & + & y & + & 2z & = & -4 \end{array}$$

Explanation. We rewrite the system in the augmented matrix form and transform it to reduced row-echelon form. We leave the details of the elementary row operations to the reader and state the final result.

$$\left[\begin{array}{ccc|c} 1 & -3 & -1 & 3 \\ -3 & 5 & 0 & -2 \\ 1 & 1 & 2 & -4 \end{array}\right] \rightsquigarrow \left[\begin{array}{ccc|c} 1 & 0 & 5/4 & -9/4 \\ 0 & 1 & 3/4 & -7/4 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

Converting the augmented matrix back to a system of equations we get

$$\begin{array}{rcrcrcr} x & + & 0y & + & (5/4)z & = & -9/4 \\ 0x & + & y & + & (3/4)z & = & -7/4 \\ 0x & + & 0y & + & 0z & = & 0 \end{array}$$

Unlike the last equation in (**??**), the last equation in this system has infinitely many solutions because all values of x, y and z satisfy it. Since the last equation contributes nothing, we will remove it and rewrite the system as

$$\begin{array}{rcl} x & = & -9/4 - (5/4)z \\ y & = & -7/4 - (3/4)z \end{array}$$

Variables x and y correspond to leading $1's$ in the reduced row-echelon form. So, x and y are the leading variables. Variable z is a free variable. We let $z = t$. Solutions to this system are points of the form

$$\left(-\frac{9}{4} - \frac{5}{4}t, -\frac{7}{4} - \frac{3}{4}t, t\right)$$

We can also interpret the solutions as lying on a line in $\mathbb{R}^3$ given by parametric equations

$$\begin{aligned} x &= -\frac{9}{4} - \frac{5}{4}t \\ y &= -\frac{7}{4} - \frac{3}{4}t \\ z &= t \end{aligned}$$

This line is the line of intersection of the three planes.

Practice Problems

Problem 85 *Is the augmented matrix shown below in reduced row-echelon form? Explain.*

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

Multiple Choice:

1. *Yes*
2. *No*

Problem 86 *Is the augmented matrix shown below in reduced row-echelon form? Explain.*

$$\left[\begin{array}{cc|c} 1 & 1 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{array}\right]$$

Multiple Choice:

1. *Yes*
2. *No*

Problem 87 *Is the augmented matrix shown below in reduced row-echelon form? Explain.*

$$\left[\begin{array}{cc|c} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{array}\right]$$

Multiple Choice:

1. *Yes*
2. *No*

Problem 88 *Is the augmented matrix shown below in reduced row-echelon form? Explain.*

$$\left[\begin{array}{ccc|c} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right]$$

Multiple Choice:

1. *Yes*
2. *No*

Problem 89 *Is the augmented matrix shown below in reduced row-echelon form? Explain.*

$$\left[\begin{array}{ccc|c} 1 & 1 & 0 & 2 \\ 0 & 1 & 0 & 9 \\ 0 & 0 & 1 & -1 \end{array}\right]$$

Multiple Choice:

1. *Yes*
2. *No*

Problem 90 *Fill in the steps that lead to the reduced row-echelon form in Example* **??**.

Problem 91 *Fill in the steps that lead to the reduced row-echelon form in Example* **??**.

Problem 92 *Suppose a system of equations has the following reduced row-echelon form*

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

What can you say about the system?

Multiple Choice:

1. *The system is inconsistent*
2. *The system has infinitely many solutions*
3. *The system has a unique solution*
4. *We would have to examine the original system to make the final determination*

Problem 93 *Solve the system using augmented matrix notation.*

$$\begin{array}{rcrcrcr} x & + & 3y & - & 2z & = & -11 \\ 2x & + & y & + & 4z & = & 12 \\ x & - & y & - & z & = & 0 \end{array}$$

Solution:

$$(\ldots\ldots, \ldots\ldots, \ldots\ldots)$$

Problem 94 *Solve the system using augmented matrix notation.*

$$\begin{array}{rcrcrcr} 3x & - & y & + & z & = & -5 \\ x & + & 2y & - & z & = & -3 \\ x & - & 5y & + & 3z & = & 1 \end{array}$$

Solution: (Enter your answers in fraction form)

$$(\ldots\ldots - \ldots\ldots t, \ldots\ldots + \ldots\ldots t, t)$$

Example Source

Linear system in Exploration **??** comes from Jim Hefferon's *Linear Algebra*. (CC-BY-NC-SA)

Jim Hefferon, *Linear Algebra*, 3rd edition, page 6.

17 Gaussian Elimination and Rank

Row Echelon and Reduced Row Echelon Forms

In *Augmented Matrix Notation and Elementary Row Operations*, we learned to write linear systems in *augmented matrix* form and use elementary row operations to transform an augmented matrix to *row-echelon form* and the *reduced row-echelon form* in order to solve linear systems.

Recall that a matrix (or augmented matrix) is in *row-echelon form* if:

- All entries *below* each leading entry are 0.
- Each leading entry is in a column to the right of the leading entries in the rows above it.
- All rows of zeros, if there are any, are located below non-zero rows.

A matrix in row-echelon form is said to be in *reduced row-echelon form* if it has the following additional properties

- Each leading entry is a 1
- All entries *above* each leading 1 are 0

Matrices in row-echelon form and reduced row-echelon form are "convenient" because their corresponding systems of equations are easy to solve.

In this module we turn to the question of existence and uniqueness of the two echelon forms. Can every matrix (augmented matrix) be carried to a row-echelon form? If so, is the row-echelon form unique? Can every matrix be reduced to a reduced row-echelon form? If so, is the reduced row-echelon form unique? The answers we find will have long-lasting implications.

Exploration 95 *Solve the following system of equations.*

$$\begin{aligned} x &+ 2y - 3z = 1 \\ -5x &+ 2y - 3z = 1 \\ x &- 2y + z = 1 \end{aligned}$$

We create an augmented matrix corresponding to the system and apply row operations until the matrix is in row-echelon form.

$$\left[\begin{array}{ccc|c} 1 & 2 & -3 & 1 \\ -5 & 2 & -3 & 1 \\ 1 & -2 & 1 & 1 \end{array}\right] \begin{array}{l} \xrightarrow{R_2+5R_1} \\ \xrightarrow{R_3-R_1} \end{array} \left[\begin{array}{ccc|c} 1 & 2 & -3 & 1 \\ 0 & 12 & -18 & 6 \\ 0 & -4 & 4 & 0 \end{array}\right] \xrightarrow{R_3+\frac{1}{3}R_2}$$

$$\left[\begin{array}{ccc|c} 1 & 2 & -3 & 1 \\ 0 & 12 & -18 & 6 \\ 0 & 0 & -2 & 2 \end{array}\right] \tag{19}$$

*Note that the elementary row operations that lead to (**??**) were not prescribed. We may employ row-operations in a different manner and obtain a different matrix in row-echelon form. For example, suppose for some reason we had begun by switching the first and third rows.*

$$\left[\begin{array}{ccc|c} 1 & 2 & -3 & 1 \\ -5 & 2 & -3 & 1 \\ 1 & -2 & 1 & 1 \end{array}\right] \xrightarrow{R_1 \leftrightarrow R_3} \left[\begin{array}{ccc|c} 1 & -2 & 1 & 1 \\ -5 & 2 & -3 & 1 \\ 1 & 2 & -3 & 1 \end{array}\right]$$

Next we would reduce this matrix to row-echelon form, perhaps in this way:

$$\left[\begin{array}{ccc|c} 1 & -2 & 1 & 1 \\ -5 & 2 & -3 & 1 \\ 1 & 2 & -3 & 1 \end{array}\right] \begin{array}{l} \xrightarrow{R_2+5R_1} \\ \xrightarrow{R_3-R_1} \end{array} \left[\begin{array}{ccc|c} 1 & -2 & 1 & 1 \\ 0 & -8 & 2 & 6 \\ 0 & 4 & -4 & 0 \end{array}\right] \xrightarrow{R_3+\frac{1}{2}R_2}$$

$$\left[\begin{array}{ccc|c} 1 & -2 & 1 & 1 \\ 0 & -8 & 2 & 6 \\ 0 & 0 & -3 & 3 \end{array}\right] \tag{20}$$

*The augmented matrices in (**??**) and (**??**) are clearly not the same, but both are in row-echelon form.*

*If we write the systems of equations corresponding to (**??**) and (**??**), we can employ back substitution to solve them. The matrix in (**??**) corresponds to*

$$\begin{aligned} x + 2y - 3z &= 1 \\ 12y - 18z &= 6 \\ -2z &= 2 \end{aligned}$$

*The matrix in (**??**) corresponds to*

$$\begin{aligned} x - 2y + z &= 1 \\ -8y + 2z &= 6 \\ -3z &= 3 \end{aligned}$$

Because both systems are equivalent to the original system, it is not surprising that back substitution yields the same solution for both systems.

$$x = 0, \quad y = -1, \quad z = -1$$

It is clear from Exploration **??** that a row-echelon form corresponding to a matrix is not unique. But what about the reduced row-echelon form?

Exploration 96 *In this problem we revisit the system*

$$\begin{array}{rcrcrcl} x & + & 2y & - & 3z & = & 1 \\ -5x & + & 2y & - & 3z & = & 1 \\ x & - & 2y & + & z & = & 1 \end{array}$$

*Following the steps we took to get (**??**), but taking the process a little further, we get the reduced row-echelon form.*

$$\left[\begin{array}{ccc|c} 1 & 2 & -3 & 1 \\ -5 & 2 & -3 & 1 \\ 1 & -2 & 1 & 1 \end{array}\right] \begin{array}{l} \xrightarrow{R_2+5R_1} \\ \xrightarrow{R_3-R_1} \end{array} \left[\begin{array}{ccc|c} 1 & 2 & -3 & 1 \\ 0 & 12 & -18 & 6 \\ 0 & -4 & 4 & 0 \end{array}\right] \xrightarrow{R_3+\frac{1}{3}R_2}$$

$$\left[\begin{array}{ccc|c} 1 & 2 & -3 & 1 \\ 0 & 12 & -18 & 6 \\ 0 & 0 & -2 & 2 \end{array}\right] \begin{array}{l} \xrightarrow{\frac{1}{12}R_2} \\ \xrightarrow{-\frac{1}{2}R_3} \end{array} \left[\begin{array}{ccc|c} 1 & 2 & -3 & 1 \\ 0 & 1 & -\frac{3}{2} & \frac{1}{2} \\ 0 & 0 & 1 & -1 \end{array}\right] \begin{array}{l} \xrightarrow{R_1+3R_3} \\ \xrightarrow{R_2+\frac{3}{2}R_3} \end{array}$$

$$\left[\begin{array}{ccc|c} 1 & 2 & 0 & -2 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \end{array}\right] \xrightarrow{R_1-2R_2} \left[\begin{array}{ccc|c} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \end{array}\right] \tag{21}$$

*Do you think it is possible to start with (**??**) and obtain the same reduced row-echelon form? Try to justify your response. If possible, find the elementary row operations that take (**??**) to the reduced row-echelon form in (**??**).*

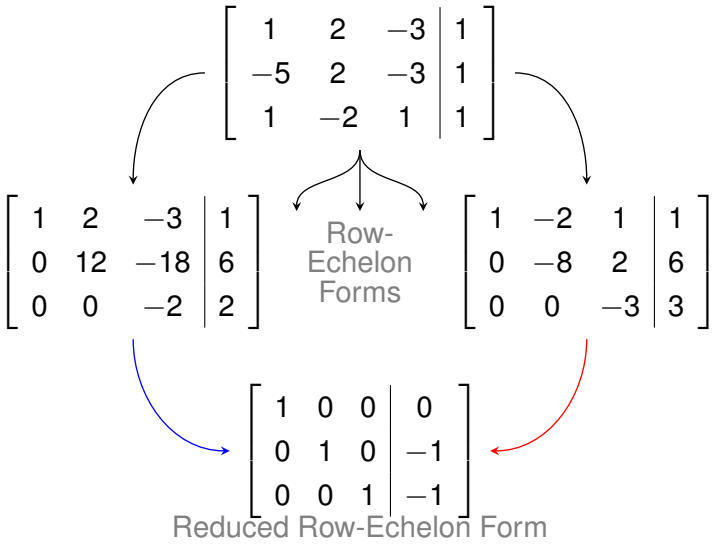
$$\left[\begin{array}{ccc|c} 1 & 2 & -3 & 1 \\ -5 & 2 & -3 & 1 \\ 1 & -2 & 1 & 1 \end{array}\right] \xrightarrow{R_1 \leftrightarrow R_3} \left[\begin{array}{ccc|c} 1 & -2 & 1 & 1 \\ -5 & 2 & -3 & 1 \\ 1 & 2 & -3 & 1 \end{array}\right]$$

$$\begin{array}{l} \xrightarrow{R_2+5R_1} \\ \xrightarrow{R_3-R_1} \end{array} \left[\begin{array}{ccc|c} 1 & -2 & 1 & 1 \\ 0 & -8 & 2 & 6 \\ 0 & 4 & -4 & 0 \end{array}\right] \xrightarrow{R_3+\frac{1}{2}R_2}$$

$$\left[\begin{array}{ccc|c} 1 & -2 & 1 & 1 \\ 0 & -8 & 2 & 6 \\ 0 & 0 & -3 & 3 \end{array}\right] \overset{\text{Elementary Row Ops.}}{\rightsquigarrow} \left[\begin{array}{ccc|c} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \end{array}\right]$$

You will be asked to fill in the elementary row operations in Practice Problem **??**.

Our observations in Exploration **??** are summarized in the following diagram.



We observed that a row-echelon form associated with a matrix is not unique. In contrast, we also saw how different sequences of elementary row operations lead to the same solution set and the same reduced row-echelon form. It turns out that the reduced row-echelon form of a matrix is unique.

Theorem 11. The reduced row-echelon form of a matrix is unique.

A proof of this result can be found in [Yuster].

The reduced row-echelon form of a matrix is an instance of a row-echelon form of the matrix. While a given matrix may have multiple row-echelon forms, all row-echelon forms will share one characteristic: the number of nonzero rows in a row-echelon form of the given matrix will be the same. We will prove this result in Theorem **??**.

Gaussian and Gauss-Jordan Elimination

Definition 19 (Gaussian Elimination)**.** The process of using the elementary row operations on a matrix to transform it into row-echelon form is called *Gaussian Elimination*.

As we saw in the previous section, it is possible to follow different sequences of row operations to arrive at various row-echelon forms. However, it was not clear whether it is *always possible* to find a row-echelon form. The following algorithm takes any matrix (or augmented matrix) and transforms it into row-echelon form:

Algorithm 1 (Gaussian Algorithm)**.** Let A be an $m \times n$ matrix.

Set $i = 1$ initially.

Step 1. If A consists entirely of zeros, stop; A is already in row-echelon form.

Step 2. Otherwise, find the first column from the left containing a nonzero entry in row i or below row i. This column will be called a *pivot column*. Go down the pivot column, beginning with row i. Pick the topmost nonzero entry and call it a. If a is not in row i, switch rows so that a moves to row i. Now a is the *leading entry* in its row. We will also refer to a as a *pivot*.

Step 3. By subtracting multiples of the row containing a from rows below it, make each entry below a zero.

Step 4. Set $i = i + 1$. If $i > m$ then stop; A is in row-echelon form.

Repeat steps 1–4 on the matrix consisting of the remaining rows. When the process stops, A will be in row echelon form.

Gaussian Algorithm guarantees that every matrix will have a row-echelon form.

Example 26. Use the Gaussian Algorithm to find a row-echelon form of A if

$$A = \begin{bmatrix} 2 & 4 \\ 1 & 2 \\ -1 & 1 \\ 3 & 5 \end{bmatrix}$$

Explanation. Following Step 2, we choose the first entry, 2, as our pivot. We then perform step 3, using the top row to get zeros in all entries below the 2.

$$\begin{bmatrix} \boxed{2} & 4 \\ 1 & 2 \\ -1 & 1 \\ 3 & 5 \end{bmatrix} \begin{array}{l} \xrightarrow{R_2-(1/2)R_1} \\ \xrightarrow{R_3+(1/2)R_1} \\ \xrightarrow{R_4-(3/2)R_1} \end{array} \begin{bmatrix} 2 & 4 \\ 0 & 0 \\ 0 & 3 \\ 0 & -1 \end{bmatrix}$$

The first row is now complete, and we repeat the process on the rows below it. We identify 3 as a pivot entry in the second column and move the row containing 3 to be directly below the first completed row. We then use the 3 to make each entry below the 3 a zero.

$$\begin{bmatrix} 2 & 4 \\ 0 & 0 \\ 0 & \boxed{3} \\ 0 & -1 \end{bmatrix} \xrightarrow{R_2 \leftrightarrow R_3} \begin{bmatrix} 2 & 4 \\ 0 & \boxed{3} \\ 0 & 0 \\ 0 & -1 \end{bmatrix} \xrightarrow{R_4 + \frac{1}{3}R_2} \begin{bmatrix} 2 & 4 \\ 0 & 3 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

This time the algorithm terminates since row 3 and row 4 are zero rows.

Definition 20 (Gauss-Jordan Elimination)**.** The process of using the elementary row operations on a matrix to transform it into reduced row-echelon form is called *Gauss-Jordan elimination*.

Given a matrix in row-echelon form, it is easy to bring it the reduced row-echelon form. For example, continuing with Example **??**, we can start where we left off and compute rref(A). From our earlier computations we have:

$$\begin{bmatrix} 2 & 4 \\ -1 & 1 \\ 3 & 5 \\ 1 & 2 \end{bmatrix} \rightsquigarrow \begin{bmatrix} 2 & 4 \\ 0 & 3 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

Now we create leading $1's$ and use them to to wipe out all non-zero entries above them.

$$\begin{bmatrix} 2 & 4 \\ 0 & 3 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{matrix} \xrightarrow{(1/2)R_1} \\ \xrightarrow{(1/3)R_2} \end{matrix} \begin{bmatrix} 1 & 2 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \xrightarrow{R_1 - 2R_2} \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} = \text{rref}(A)$$

The following modification to the Gaussian Algorithm produces the reduced row-echelon form of a matrix. This algorithm guarantees the existence of the reduced row-echelon form.

Algorithm 2 (Gauss-Jordan Algorithm)**.** Let A be an $m \times n$ matrix. Follow the steps of the Gaussian Algorithm but modify Step 2 to create leading $1's$ by multiplying the row containing a by $\frac{1}{a}$.

When the Gaussian Algorithm terminates, subtract multiples of the rows containing leading $1's$ from the rows above to make all entries above the pivots zero.

Example 27. Use the Gauss-Jordan Algorithm to solve the system

$$\begin{aligned} 3x &+ y &+ 7z &= 7 \\ 5x &+ 3y &+ 9z &= 13 \\ 2x &+ y &+ 4z &= 5 \end{aligned}$$

Explanation.

$$\left[\begin{array}{ccc|c} \boxed{3} & 1 & 7 & 7 \\ 5 & 3 & 9 & 13 \\ 2 & 1 & 4 & 5 \end{array}\right]$$

$$\xrightarrow{\frac{1}{3}R_1} \left[\begin{array}{ccc|c} \boxed{1} & 1/3 & 7/3 & 7/3 \\ 5 & 3 & 9 & 13 \\ 2 & 1 & 4 & 5 \end{array}\right]$$

$$\xrightarrow{R_2-5R_1} \left[\begin{array}{ccc|c} \boxed{1} & 1/3 & 7/3 & 7/3 \\ 0 & 4/3 & -8/3 & 4/3 \\ 2 & 1 & 4 & 5 \end{array}\right]$$

$$\xrightarrow{R_3-2R_1} \left[\begin{array}{ccc|c} 1 & 1/3 & 7/3 & 7/3 \\ 0 & \boxed{4/3} & -8/3 & 4/3 \\ 0 & 1/3 & -2/3 & 1/3 \end{array}\right]$$

$$\xrightarrow{\frac{3}{4}R_2} \left[\begin{array}{ccc|c} 1 & 1/3 & 7/3 & 7/3 \\ 0 & \boxed{1} & -2 & 1 \\ 0 & 1/3 & -2/3 & 1/3 \end{array}\right]$$

$$\xrightarrow{R_3-\frac{1}{3}R_2} \left[\begin{array}{ccc|c} 1 & 1/3 & 7/3 & 7/3 \\ 0 & \boxed{1} & -2 & 1 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

$$\xrightarrow{R_1-\frac{1}{3}R_2} \left[\begin{array}{ccc|c} 1 & 0 & 3 & 2 \\ 0 & 1 & -2 & 1 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

We convert the reduced row-echelon form to a system of equations and find the solution. The last equation contributes nothing to the system so we omit writing it down.

$$\begin{aligned} x &+ 3z &= 2 \\ y &- 2z &= 1 \end{aligned}$$

The solution is

$$x = 2 - 3t, \quad y = 1 + 2t, \quad z = t$$

The Gauss-Jordan Algorithm guarantees the existence of the reduced row-echelon form for all matrices. When doing computations by hand, however, the algorithm may not always be the optimal method of finding a row-echelon form or the reduced row-echelon form because the procedure often leads to fractions early in the process. The following video shows how to arrive at the same reduced row-echelon form for the matrix in Example **??** without doing any fraction arithmetic. You will see that we still employ row operations, but in a different order.

YouTube link: https://www.youtube.com/watch?v=76Y41ncuLeQ

The video highlights the fact that regardless of what sequence of elementary row operations we take to arrive at the reduced row-echelon form, the end result is the same.

Rank

As stated in Theorem **??**, the reduced row-echelon form of a matrix A is uniquely determined by A. That is, no matter which series of row operations is used to transform A to its reduced row-echelon matrix, the result will always be the same matrix. In contrast, this is not true for row-echelon matrices: different sequences of row operations can transform the same matrix A to different row-echelon matrices. However, it is true that the number r of nonzero rows must be the same in each of these row-echelon matrices, as we will see in Theorem **??**. Hence, the number r depends only on A and not on the way in which A is carried to row-echelon form.

Example 28. Matrices (**??**) and (**??**) of Exploration **??** are both row-echelon forms of A. Both matrices have three nonzero rows. The same is true for rref(A).

Definition 21. The *rank* of matrix A, denoted by rank(A), is the number of nonzero rows that remain after we reduce A to row-echelon form by elementary row operations.

Example 29. Compute the rank of

$$A = \begin{bmatrix} 1 & 1 & -1 & 4 \\ 2 & 1 & 3 & 0 \\ 0 & 1 & -5 & 8 \end{bmatrix}$$

Explanation. A reduction of A to row-echelon form is

$$A = \begin{bmatrix} 1 & 1 & -1 & 4 \\ 2 & 1 & 3 & 0 \\ 0 & 1 & -5 & 8 \end{bmatrix} \rightsquigarrow \begin{bmatrix} 1 & 1 & -1 & 4 \\ 0 & -1 & 5 & -8 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Because the row-echelon form has two nonzero rows, rank(A) = 2.

Theorem 12. Suppose a system of m equations in n variables is consistent, and that the rank of the *coefficient* matrix is r.

1. The set of solutions involves exactly $n-r$ parameters, corresponding to $n-r$ free variables.
2. If $r < n$, the system has infinitely many solutions.
3. If $r = n$, the system has a unique solution.

Proof The fact that the rank of the coefficient matrix is r means that there are exactly r leading variables in the coefficient matrix, and hence exactly $n - r$ nonleading variables. The nonleading variables are called free variables. All free variables are assigned parameters, so the set of solutions involves exactly $n - r$ parameters. Hence if $r < n$, there is at least one parameter, and so infinitely many solutions. If $r = n$, there are no parameters and the resulting solution is unique. ■

Theorem **??** shows that, for any system of linear equations, exactly three possibilities exist:

1. Unique solution. This occurs when every variable is a leading variable.
2. Infinitely many solutions. This occurs when the system is consistent and there is at least one nonleading variable, so at least one parameter is involved.
3. No solution. This occurs when a row $\left[\begin{array}{cccc|c} 0 & 0 & \ldots & 0 & 1 \end{array}\right]$ appears in the row-echelon form. Such a row corresponds to an equation with no solutions. (See Example **??**.)

Practice Problems

Problem 97 *Follow the indicated steps of the Gauss-Jordan algorithm to transform the matrix to its reduced row-echelon form. Steps will unfold automatically as you enter correct answers.*

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

Problem 97.1 Problem 97.1.1 Problem 97.1.1.1 Problem 97.1.1.1.1 Problem 97.1.1.1.1.1 Probl

Problem 98 *The following reduction steps do not follow the algorithm but are easier for a human to carry out because they use fewer fractions. Follow the indicated steps to transform the matrix to its reduced row-echelon form. Steps will unfold automatically as you enter correct answers.*

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

Problem 98.1 Problem 98.1.1 Problem 98.1.1.1 Problem 98.1.1.1.1 Problem 98.1.1.1.1.1 Proble

Problem 99 *Find the rank A.*

$$A = \begin{bmatrix} 4 & 3 & -1 \\ -8 & -6 & 2 \end{bmatrix}$$

Answer:

$$rank(A) = \ldots\ldots$$

Problem 100 *Find the rank B.*

$$B = \begin{bmatrix} 1 & 1 \\ 2 & -2 \\ 3 & -1 \end{bmatrix}$$

Answer:

$$rank(B) = \ldots\ldots$$

Problem 101 *Find the rank C.*

$$C = \begin{bmatrix} 1 & 0 & 1 \\ 2 & 1 & 3 \\ 0 & 1 & -2 \end{bmatrix}$$

Answer:

$$rank(C) = \ldots\ldots$$

Problem 102 *Find the rank D.*

$$D = \begin{bmatrix} 1 & 1 & 2 \\ -1 & -2 & 1 \\ 1 & 0 & 5 \end{bmatrix}$$

Answer:

$$rank(D) = \ldots\ldots$$

Problem 103 *Suppose A is a matrix with 5 rows and 7 columns. Which of the following is/are possible? Select ALL that apply.*

Select All Correct Answers:

1. $rank(A) = 7$
2. $rank(A) = 6$
3. $rank(A) = 5$
4. $rank(A) = 4$

Problem 104 *Suppose a linear system has* 4 *equations and* 5 *unknowns. Which of the following is NOT a possibility?*

Multiple Choice:

1. *The system has a unique solution*
2. *The system has no solutions*
3. *The system has infinitely many solutions*

Problem 105 *Suppose A is a matrix such that rref*(A) *has* 5 *leading* $1's$*. What do we know to be true about A? Select ALL that apply.*

Select All Correct Answers:

1. $rank(A) = 5$
2. *A has at least* 25 *entries*

3. *Any row-echelon form of A will have exactly* 5 *nonzero rows*
4. *Some row-echelon forms of A may have more than* 5 *nonzero rows*
5. *Some row-echelon forms of A may have fewer than* 5 *nonzero rows*

Problem 106 *In this problem we will consider how the rank of the coefficient matrix associated with a linear system compares to the rank of the augmented matrix associated with the system.*

1. *Explain why the rank of the augmented matrix has to be greater than or equal to the rank of the coefficient matrix.*
2. *Prove that for a consistent system the rank of the coefficient matrix will be the same as the rank of the augmented matrix.*
3. *Give an example of an inconsistent system for which the rank of the augmented matrix is greater than the rank of the coefficient matrix.*
4. *Can the rank of an augmented matrix be greater than the number of variables?*
5. *Is the following statement true?*

   "If the rank of the augmented matrix associated with a linear system is greater than the rank of the coefficient matrix, then the system is inconsistent."

Text Source

The section on Rank is an adaptation of Section 1.2 of Keith Nicholson's *Linear Algebra with Applications*.

W. Keith Nicholson, *Linear Algebra with Applications*, Lyryx 2018, Open Edition, p 15-17.

Bibliography

[Yuster] Thomas Yuster, The Reduced Row Echelon Form of a Matrix is Unique: a Simple Proof, Mathematics Magazine, vol. 57, no. 2 (Mar. 1984), pp. 93-94.

18 Iterative Methods for Solving Linear Systems

The approaches to solving linear systems we have studied up to this point are often called *direct methods*, to set them apart from *iterative methods*, which we study in this section. When we solve linear systems using *iterative methods*, we repeat the same procedure (called an *iteration*) many times (usually using a computer), and we look for the outputs to "converge to" the actual solution. Let's begin with an exploration to illustrate this idea.

Exploration 107 *In the following GeoGebra interactive, the red line is the graph of $ax + by = r$, and the blue line is the graph of $cx + dy = s$. Point A, referred to as our "initial guess," may be dragged anywhere on the plane before starting the iterative procedure. Use the navigation bar on the bottom to proceed through several iterations. One iteration will consist of sliding horizontally from the current point until we hit the red line, followed by sliding vertically until we hit the blue line. These two steps are then repeated from the new point on the blue line.*

Access GeoGebra interactives through the online version of this text at

https://ximera.osu.edu/linearalgebradzv3/LinearAlgebraInteractiveIntro.

1. *Try solving the system of equations*

$$\begin{aligned} 5x &- 2y = 10 \\ x &+ 3y = 36 \end{aligned} \tag{22}$$

   from various starting points A. Does the iterative method "converge" to the actual solution of this system?

2. *Now try solving the same system of equations, but changing which equation is the red line and which equation is the blue line. In other words, try*

$$\begin{aligned} x &+ 3y = 36 \\ 5x &- 2y = 10 \end{aligned} \tag{23}$$

   from various starting points A. Now what happens when we try the iterative method?

3. *Try other 2×2 systems of equations and see how the iterative method works.*

Iterative methods have become more important as computers have become more ubiquitous. When we execute an iterative method, we are implementing the same sequence of steps repeatedly, and this is something computers do very well.

In this section we will explore two different iterative methods for solving a system of linear equations. Exploration **??** was a geometric representation of the *Gauss-Seidel* method for a system of two equations with two unknowns. We will show how this may be implemented numerically to approximate the solution to a system of equations. However, we first explore a related method known as *Jacobi's method*.

Jacobi's method

As you saw in the Exploration above, iterative methods do not give an exact answer, but an approximate answer. More iterations give a more accurate answer.

To use *Jacobi's method* for solving a linear system of n equations in n variables, we isolate the first variable in the first equation, we isolate the second variable in the second equation, and in general, we isolate the ith variable in the ith equation.

Let's return to the system of equations from Exploration **??**. To use Jacobi's method to approximate the solution to the system, we solve for a different variable in each of the two equations. This gives us two formulas, one for each unknown. We begin with our initial guess $\mathbf{v}_0$, and use these formulas to compute the next iteration $\mathbf{v}_1$. Taking the vector $\mathbf{v}_1$, and we again apply these formulas to compute the next iteration to arrive at $\mathbf{v}_2$.

$$\begin{array}{rcrcl} 5x & - & 2y & = & 10 \\ x & + & 3y & = & 36 \end{array} \rightsquigarrow \text{isolating one variable in each equation} \rightsquigarrow \begin{array}{rcl} x & = & \dfrac{10+2y}{5} \\ y & = & \dfrac{36-x}{3} \end{array}$$

Let's begin with the initial guess $\mathbf{v}_0 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$. Applying the pair of formulas gives

$$\mathbf{v}_1 = \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \dfrac{10+2(0)}{5} \\ \dfrac{36-0}{3} \end{bmatrix} = \begin{bmatrix} 2 \\ 12 \end{bmatrix}.$$

For the next iteration, we apply these same formulas to the x and y values of $\mathbf{v}_1$. We get

$$\mathbf{v}_2 = \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \dfrac{10+2(12)}{5} \\ \dfrac{36-2}{3} \end{bmatrix} = \begin{bmatrix} 34/5 \\ 34/3 \end{bmatrix} \approx \begin{bmatrix} 6.80 \\ 11.33 \end{bmatrix}.$$

We use this to perform another iteration, giving us

$$\mathbf{v}_3 = \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \dfrac{10 + 2(11.33)}{5} \\ \dfrac{36 - 6.80}{3} \end{bmatrix} = \begin{bmatrix} 34/5 \\ 38/3 \end{bmatrix} \approx \begin{bmatrix} 6.53 \\ 9.93 \end{bmatrix}.$$

We observe that at each iteration our approximate solution gets closer to the actual solution to the system, $x = 6$, $y = 10$.

The Gauss-Seidel method

To implement the *Gauss-Seidel method* for solving a linear system of n equations in n variables, we use the same set of n formulas as we use in Jacobi's method. The difference is that we implement the formulas *sequentially* rather than simultaneously.

To demonstrate, we return to the same system of equations examined above.

$$\begin{array}{rcrcl} 5x & - & 2y & = & 10 \\ x & + & 3y & = & 36 \end{array} \rightsquigarrow \text{isolating variables as before} \rightsquigarrow x = \frac{10 + 2y}{5}, y = \frac{36 - x}{3}$$

Once again we start with the initial guess $\mathbf{v}_0 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$. We apply the first formula and get

$$x = \frac{10 + 2(0)}{5} = 2.$$

However, we will compute y using this new value $x = 2$ rather than using $x = 0$. We get

$$y = \frac{36 - 2}{3} = 34/3 \approx 11.33.$$

So after our first iteration we have $\mathbf{v}_1 = \begin{bmatrix} 2 \\ 11.33 \end{bmatrix}$. We repeat this procedure to perform another iteration. We get

$$x = \frac{10 + 2(11.33)}{5} \approx 6.53$$

followed by

$$y = \frac{36 - 6.53}{3} \approx 9.82,$$

so we have $\mathbf{v}_2 \approx \begin{bmatrix} 6.53 \\ 9.82 \end{bmatrix}$.

We observe that in this example the second iteration of the Gauss-Seidel method is closer to the actual solution of $x = 6, y = 10$ than the third iteration of Jacobi's method. The Gauss-Seidel method is more efficient than Jacobi's method because it makes use of a better approximation of one coordinate to find another.

Exploration 108 *The Jacobi and Gauss-Seidel methods are easily implemented on a computer. Below is a link to a spreadsheet which can be used to solve a linear system of 2 equations in 2 unknowns using each of these methods.*

To use the spreadsheet, SAVE A COPY of the sheet. Then you will be able to change the yellow boxes to input a system of equations as well as the initial values. The video below the link to the spreadsheet gives additional information.

LINK TO SPREADSHEET

YouTube link: `https://www.youtube.com/watch?v=rb45dYb0fp0`

Both techniques described above can be used to solve linear systems with more variables.

Example 30. Use direct methods to solve the linear system. Then try using Jacobi's Method and the Gauss-Seidel method.

$$\begin{aligned} 5x &+ y &+ 2z &= 1 \\ 2x &+ 5y &+ z &= 1 \\ 2x &+ y &+ 5z &= 1 \end{aligned}$$

Explanation. We perform Gauss-Jordan elimination by applying elementary row operations as described in Gaussian Elimination and Rank.

$$\left[\begin{array}{ccc|c} 5 & 1 & 2 & 1 \\ 2 & 5 & 1 & 1 \\ 2 & 1 & 5 & 1 \end{array}\right] \rightsquigarrow \text{ becomes (after a number of steps) } \rightsquigarrow \left[\begin{array}{ccc|c} 1 & 0 & 0 & 1/8 \\ 0 & 1 & 0 & 1/8 \\ 0 & 0 & 1 & 1/8 \end{array}\right]$$

This shows that the three planes in $\mathbb{R}^3$ share a common intersection point at $\left(\frac{1}{8}, \frac{1}{8}, \frac{1}{8}\right)$, which is easily verified by looking at the original equations, and can also be seen in the GeoGebra graph below (RIGHT-CLICK AND DRAG TO ROTATE).

Access GeoGebra interactives through the online version of this text at

https://ximera.osu.edu/linearalgebradzv3/LinearAlgebraInteractiveIntro.

Now we solve the same linear system using our iterative methods. In order to use Jacobi's method or the Gauss-Seidel method, we solve for x in the first equation, we solve for y in the

second equation, and we solve for z in the third equation. We get three formulas.

$$\begin{array}{rcrcrcl} 5x & + & y & + & 2z & = & 1 \\ 2x & + & 5y & + & z & = & 1 \\ 2x & + & y & + & 5z & = & 1 \end{array} \rightsquigarrow \text{isolating these variables} \rightsquigarrow \begin{array}{l} x = \dfrac{1-y-2z}{5}, \\ y = \dfrac{1-2x-z}{5}, \\ z = \dfrac{1-2x-y}{5} \end{array}$$

Below is a link to a spreadsheet which can be used to solve a linear system of 3 equations in 3 unknowns using the Jacobi method and the Gauss-Seidel method.

To use the spreadsheet, SAVE A COPY of the sheet. Then you will be able to change the yellow boxes to input a system of equations as well as the initial values.

Access spreadsheet link through the online version of this text at

https://ximera.osu.edu/linearalgebradzv3/LinearAlgebraInteractiveIntro.

Notice how much more efficient the Gauss-Seidel method is than Jacobi's method. Both methods converge toward the exact solution we computed earlier using Gauss-Jordan elimination ((0.125, 0.125.0.125) in decimal form). Starting with an initial guess of $(1, 2, 3)$, when we use the Gauss-Seidel method, each coordinate gets to within 0.001 of the corresponding value of the exact solution. Using Jacobi's method, 15 iterations are necessary to achieve the same level of accuracy with the same initial guess. This is not surprising, since the Gauss-Seidel Method uses new values as we find them, rather than waiting until the subsequent iteration, as is done with the Jacobi Method.

Convergence of Iterative Methods

We observed in Exploration **??** that the Gauss-Seidel method converged for the linear system **??**. However, when we switched the order of the equations, writing the same linear system using **??**, the method failed to converge.

If you experiment with the spreadsheets and the GeoGebra interactive, you can find other systems of equations for which these iterative methods fail to converge to the solution. At this point you may be wondering why we are studying iterative methods when we cannot be certain that they will converge to the solution!

There are many situations where iterative methods are, in fact, the best methods we have for performing computations. We will see a nice example of this later when we study The Power Method and the Dominant Eigenvalue.

Also, it turns out that sometimes we can be certain that an iterative method will converge to the solution. To do so, we examine the *coefficient matrix* A of a linear system $A\mathbf{x} = \mathbf{b}$. To be clear, when we write a linear system as a single augmented matrix, the coefficient matrix A is the part of an augmented matrix to the left of the vertical line.

We need the following definition, which will also be useful when we study Gershgorin's Theorem.

Definition 22. Let $A = [a_{ij}]$ be the $n \times n$ matrix which is the coefficient matrix of the linear system $A\mathbf{x} = \mathbf{b}$. Let

$$r_i(A) := \sum_{j \neq i} |a_{ij}|$$

denote the sum of the absolute values of the non-diagonal entries in row i. We say that A is *strictly diagonally dominant* if

$$|a_{ii}| > r_i(A)$$

for all values of i from $i = 1$ to $i = n$.

Example 31. For each of the linear systems examined above, determine whether or not the coefficient matrix is strictly diagonally dominant.

Explanation. 1. In Example **??** we considered the linear system

$$\begin{array}{rcrcrcl} 5x & + & y & + & 2z & = & 1 \\ 2x & + & 5y & + & z & = & 1 \\ 2x & + & y & + & 5z & = & 1 \end{array}.$$

The associated coefficient matrix is

$$A = \begin{bmatrix} 5 & 1 & 2 \\ 2 & 5 & 1 \\ 2 & 1 & 5 \end{bmatrix}.$$

We see that A is strictly diagonally dominant, for proceeding row by row, we note that

$$\begin{aligned} 5 &> 1 + 2 \\ 5 &> 2 + 1 \\ 5 &> 2 + 1. \end{aligned}$$

2. In Exploration **??** we considered the linear system

$$\begin{array}{rcrcl} 5x & - & 2y & = & 10 \\ x & + & 3y & = & 36 \end{array}.$$

The associated coefficient matrix is

$$A = \begin{bmatrix} 5 & -2 \\ 1 & 3 \end{bmatrix}.$$

We see that A is strictly diagonally dominant, for proceeding row by row, we note that

$$\begin{aligned} 5 &> |-2| \\ 3 &> 1 \end{aligned}$$
.

3. Later in the same Exploration **??**, we switched the order of the equations, giving us the linear system

$$\begin{array}{rcrcl} x & + & 3y & = & 36 \\ 5x & - & 2y & = & 10 \end{array}.$$

The associated coefficient matrix is

$$A = \begin{bmatrix} 1 & 3 \\ 5 & -2 \end{bmatrix}.$$

We see that A is *NOT* strictly diagonally dominant, for proceeding row by row, we note that

$$\begin{aligned} 1 &< 3 \\ 5 &> |-2|. \end{aligned}$$

.

Now that we have explained the definition of *strict diagonal dominance*, we state (but do not prove) the following convergence theorem.

Theorem 13. Let $A = [a_{ij}]$ be the $n \times n$ matrix which is the coefficient matrix of the linear system $A\mathbf{x} = \mathbf{b}$, and suppose A is strictly diagonally dominant. Then both the Jacobi method and the Gauss-Seidel method will converge to the solution to the linear system.

The proof of this theorem at this point in the course would take us too far astray. Instead, we refer the interested reader to this 1995 paper by Roberto Bagnara.

Note that the converse of this theorem is false. Sometimes these iterative methods converge even though the coefficient matrices are not strictly diagonally dominant. Practice Problem **??** gives two examples where both methods converge to the solution even though the coefficient matrices are not strictly diagonally dominant. We say that strict diagonal dominance is *sufficient* to guarantee convergence, but it is not a *necessary* condition for convergence.

Practice Problems

Problems **??-??**

For each of the following, use Jacobi's method and the Gauss-Seidel method with an initial guess of $x = 0, y = 0$ to find an approximate solution where each coordinate is within 0.1 of the exact solution. You can make a copy of the spreadsheets above and modify them to help. Note that the order of the equations may affect convergence.

Problem 109

$$\begin{aligned} -5x + y &= 17 \\ x + y &= 5 \end{aligned}$$

Problem 110

$$\begin{aligned} -3x + 4y &= 5 \\ -x + 2y &= 1 \end{aligned}$$

Problems **??-??**

For each of the following, use Jacobi's method and the Gauss-Seidel method with an initial guess of $x = 0, y = 0, z = 0$ to find an approximate solution where each coordinate is within 0.1 of the exact solution. You can make a copy of the spreadsheets above and modify them to help. Note that the order of the equations may affect convergence.

Problem 111

$$\begin{aligned} 3x - 2y + z &= -5 \\ x + 3y - z &= 12 \\ x + y + 2z &= 0 \end{aligned}$$

Problem 112

$$\begin{aligned} 8x + 3y + 3z &= 7 \\ 4x + 9y + 3z &= 7 \\ 2x + 6y + 9z &= 9 \end{aligned}$$

Problem 113 *The lines, $x + 3y = 1$ and $4x - y = 3$. intersect at $x = \frac{10}{13}, y = \frac{1}{13}$. Use Jacobi's method and the Gauss-Seidel method with an initial guess of $x = 0, y = 0$ to find an approximate solution where each coordinate is within 0.1 of the exact solution. Write the coefficient matrix so it is strictly diagonally dominant in order to ensure convergence.*

Problem 114 *What happens if the Gauss-Seidel method is to find the solution to a system of 2 equations in 2 unknowns when the slopes of the two lines are m and $-m$?*

Practice Problem Source

Problem **??** comes from the end of Chapter 1 of Ken Kuttler's *A First Course in Linear Algebra*. (CC-BY)

Ken Kuttler, *A First Course in Linear Algebra*, Lyryx 2017, Open Edition, p. 42.

19 Big Ideas about Vectors

20 Linear Combinations of Vectors

When studying vectors, the two main operations we have learned about are vector addition and scalar multiplication. Both are involved in the important concept of a *linear combination* of vectors.

Definition 23. A vector $\mathbf{v}$ is said to be a *linear combination* of vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ if

$$\mathbf{v} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \ldots + a_n\mathbf{v}_n$$

for some scalars $a_1, a_2, \ldots, a_n$.

For example, $\begin{bmatrix} -4 \\ 9 \\ -10 \\ -1 \end{bmatrix}$ is a linear combination of $\begin{bmatrix} -1 \\ 3 \\ -3 \\ 0 \end{bmatrix}$, $\begin{bmatrix} 2 \\ 0 \\ 1 \\ 4 \end{bmatrix}$ and $\begin{bmatrix} 0 \\ 1 \\ -1 \\ 1 \end{bmatrix}$ because

$$\begin{bmatrix} -4 \\ 9 \\ -10 \\ -1 \end{bmatrix} = 2\begin{bmatrix} -1 \\ 3 \\ -3 \\ 0 \end{bmatrix} + (-1)\begin{bmatrix} 2 \\ 0 \\ 1 \\ 4 \end{bmatrix} + 3\begin{bmatrix} 0 \\ 1 \\ -1 \\ 1 \end{bmatrix}$$

In this section we will focus on vectors in $\mathbb{R}^2$ and $\mathbb{R}^3$.

Visualizing Linear Combinations in $\mathbb{R}^2$ and $\mathbb{R}^3$

Let's start by visualizing linear combinations of two vectors in $\mathbb{R}^2$.

Exploration 115 *Answer the questions below using the GeoGebra interactive. To use the interactive, you can*

1. *Change vectors* $\mathbf{v}$ *and* $\mathbf{w}$ *by dragging the tips of these vectors.*
2. *Change the coefficients* k_1 *and* k_2 *of the linear combination by using sliders.*

Access GeoGebra interactives through the online version of this text at

https://ximera.osu.edu/linearalgebradzv3/LinearAlgebraInteractiveIntro.

1. *Let* $\mathbf{w} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ *and* $\mathbf{v} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$. *Find* k_1 *and* k_2 *such that* $k_1\mathbf{w} + k_2\mathbf{v} = \begin{bmatrix} 4 \\ -1 \end{bmatrix}$

   $$k_1 = \ldots\ldots, \quad k_2 = \ldots\ldots$$

2. *Let* $\mathbf{w} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ *and* $\mathbf{v} = \begin{bmatrix} -2 \\ 0 \end{bmatrix}$. *Find* k_1 *and* k_2 *such that* $k_1\mathbf{w} + k_2\mathbf{v} = \begin{bmatrix} 3 \\ -2 \end{bmatrix}$

   $$k_1 = \ldots\ldots, \quad k_2 = \ldots\ldots$$

3. *Use the same vectors* $\mathbf{w}$ *and* $\mathbf{v}$ *as in the previous part. Do you think it is possible to express any vector in* $\mathbb{R}^2$ *as a linear combination of* $\mathbf{w}$ *and* $\mathbf{v}$*? (Yes, / No)*

4. *Let* $\mathbf{w} = \begin{bmatrix} 4 \\ 2 \end{bmatrix}$ *and* $\mathbf{v} = \begin{bmatrix} -2 \\ -1 \end{bmatrix}$. *Do you think it is possible to express any vector in* $\mathbb{R}^2$ *as a linear combination of* $\mathbf{w}$ *and* $\mathbf{v}$*? (Yes, / No)*

Visualizing linear combinations of vectors in $\mathbb{R}^3$ is more difficult than doing so in $\mathbb{R}^2$. The following Exploration will help you do this.

Exploration 116 *We will start by visualizing linear combinations of two vectors,* $\mathbf{u}$ *and* $\mathbf{v}$*, in* $\mathbb{R}^3$*. To use the interactive below, define vectors* $\mathbf{u}$ *and* $\mathbf{v}$*. Use sliders to change the coefficients* k_1 *and* k_2 *of the linear combination. You will see the linear combination* $k_1\mathbf{u} + k_2\mathbf{v}$ *as the pink vector along the diagonal of the parallelogram determined by* $\mathbf{u}$ *and* $\mathbf{v}$*.*

RIGHT-CLICK and DRAG the left panel to rotate the image.

Access GeoGebra interactives through the online version of this text at

https://ximera.osu.edu/linearalgebradzv3/LinearAlgebraInteractiveIntro.

We will now consider three vectors. Define vectors $\mathbf{u}$*,* $\mathbf{v}$ *and* $\mathbf{w}$*. Use sliders to change the coefficients* k_1, k_2 *and* k_3 *of the linear combination. The linear combination* $k_1\mathbf{u} + k_2\mathbf{v} + k_3\mathbf{w}$ *is shown as the pink vector along the diagonal of the parallelepiped.*

RIGHT-CLICK and DRAG the left panel to rotate the image.

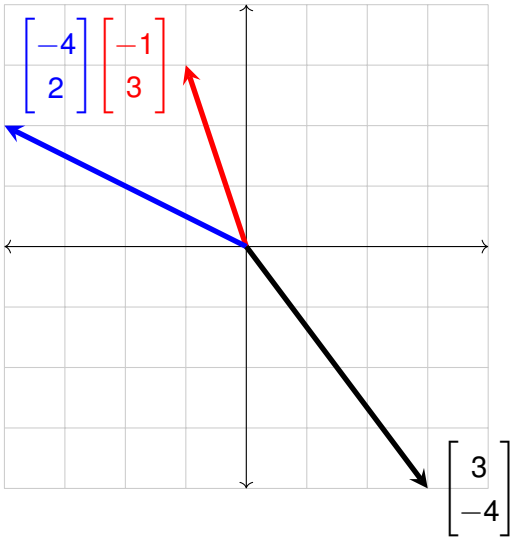
Access GeoGebra interactives through the online version of this text at

https://ximera.osu.edu/linearalgebradzv3/LinearAlgebraInteractiveIntro.

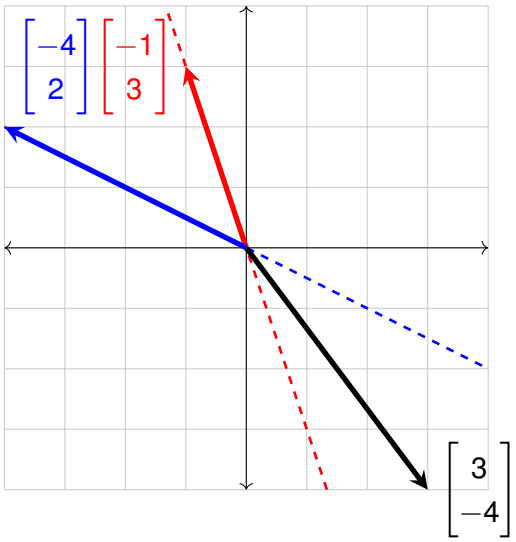
Geometry of Linear Combinations

Example 32. Use geometry to express $\begin{bmatrix} 3 \\ -4 \end{bmatrix}$ as a linear combination of $\begin{bmatrix} -1 \\ 3 \end{bmatrix}$ and $\begin{bmatrix} -4 \\ 2 \end{bmatrix}$.

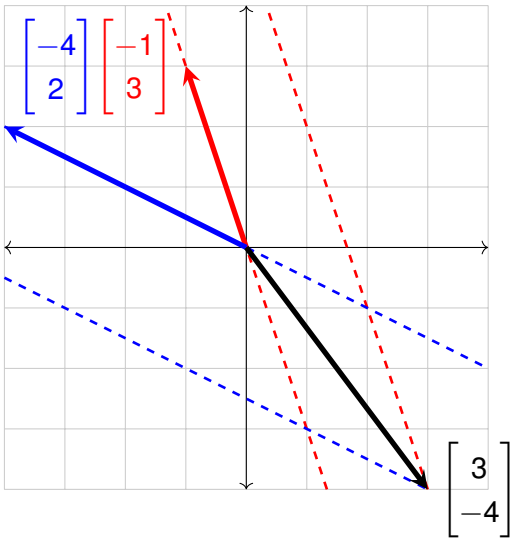
Explanation. We are looking for $\begin{bmatrix} 3 \\ -4 \end{bmatrix}$ to be the diagonal of a parallelogram determined by scalar multiples of $\begin{bmatrix} -1 \\ 3 \end{bmatrix}$ and $\begin{bmatrix} -4 \\ 2 \end{bmatrix}$.



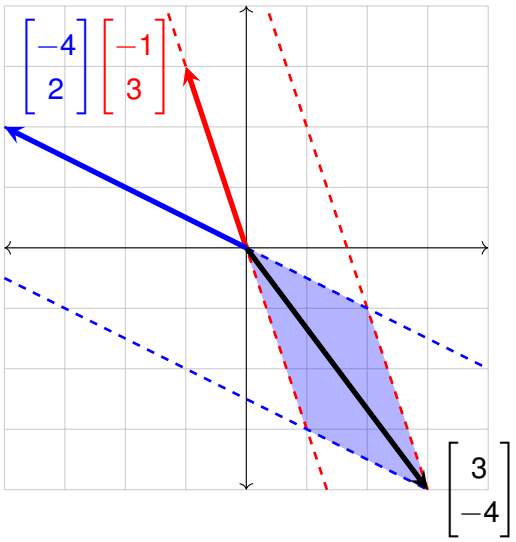
Because a scalar multiple of a vector can point in the same direction as the vector or in the opposite direction, we will start by drawing straight lines determined by the two vectors.



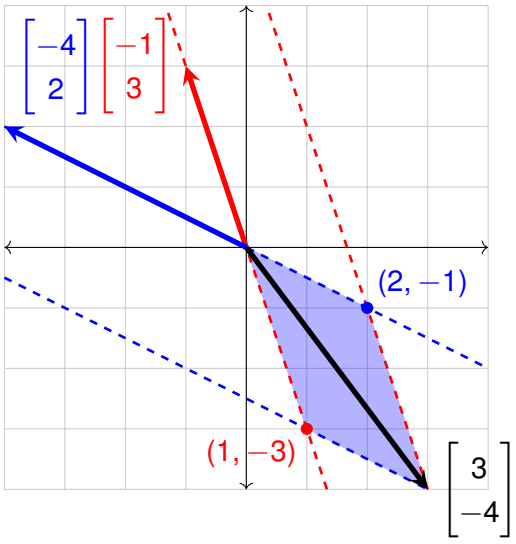
The two lines that we drew will contain the sides of the parallelogram we are looking for. To find the other two sides we will draw lines parallel to $\begin{bmatrix}-1\\3\end{bmatrix}$ and $\begin{bmatrix}-4\\2\end{bmatrix}$ through the head of vector $\begin{bmatrix}3\\-4\end{bmatrix}$.



Now the parallelogram is clearly visible.



The last remaining task is to identify the sides of the parallelogram as scalar multiples of $\begin{bmatrix}-1\\3\end{bmatrix}$ and $\begin{bmatrix}-4\\2\end{bmatrix}$. Observe that vectors $\begin{bmatrix}1\\-3\end{bmatrix}$ and $\begin{bmatrix}2\\-1\end{bmatrix}$ determine the parallelogram.



Vector $\begin{bmatrix}2\\-1\end{bmatrix}$ is half the length of $\begin{bmatrix}-4\\2\end{bmatrix}$ and points in the opposite direction, while the vector

$\begin{bmatrix} 1 \\ -3 \end{bmatrix}$ is the same length as $\begin{bmatrix} -1 \\ 3 \end{bmatrix}$ and also points in the opposite direction.

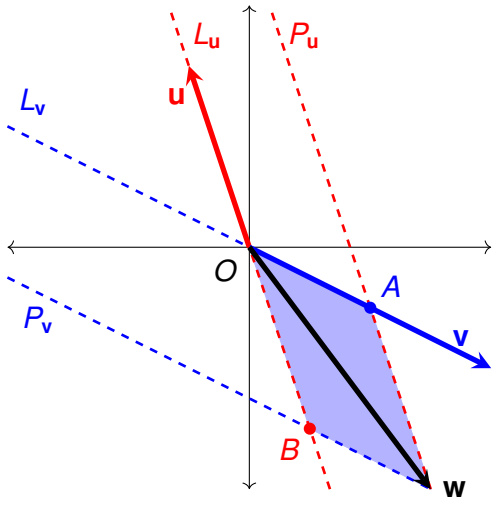
We write $\begin{bmatrix} 3 \\ -4 \end{bmatrix}$ as a linear combination of $\begin{bmatrix} -1 \\ 3 \end{bmatrix}$ and $\begin{bmatrix} -4 \\ 2 \end{bmatrix}$ as follows

$$\begin{bmatrix} 3 \\ -4 \end{bmatrix} = \begin{bmatrix} 1 \\ -3 \end{bmatrix} + \begin{bmatrix} 2 \\ -1 \end{bmatrix} = (-1)\begin{bmatrix} -1 \\ 3 \end{bmatrix} + \left(-\frac{1}{2}\right)\begin{bmatrix} -4 \\ 2 \end{bmatrix}$$

The method we used in Example **??** to express the given vector as a linear combination of two other vectors is sufficiently useful that we summarize the steps.

Procedure 1. Given two non-collinear vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^2$, and a vector $\mathbf{w}$, we can express $\mathbf{w}$ as a linear combination of $\mathbf{u}$ and $\mathbf{v}$ as follows:

1. Draw lines $L_{\mathbf{u}}$ and $L_{\mathbf{v}}$ determined by $\mathbf{u}$ and $\mathbf{v}$, respectively.
2. Through the head of vector $\mathbf{w}$, draw lines $P_{\mathbf{u}}$ and $P_{\mathbf{v}}$, parallel to $L_{\mathbf{u}}$ and $L_{\mathbf{v}}$, respectively.
3. Let A be the point of intersection of $P_{\mathbf{u}}$ and $L_{\mathbf{v}}$.
4. Let B be the point of intersection of $P_{\mathbf{v}}$ and $L_{\mathbf{u}}$.
5. Let O denote the origin. Then $\overrightarrow{OA} = k_1\mathbf{v}$ and $\overrightarrow{OB} = k_2\mathbf{u}$ for some scalars k_1 and k_2.
6. We have $\mathbf{w} = k_2\mathbf{u} + k_1\mathbf{v}$.



Exploration 117 *This GeoGebra interactive will allow you to go through the steps given in Procedure* **??** *for a combination of vectors of your choice. To use the interactive*

1. *Enter components of vectors* **u** *and* **v**.
2. *Enter components of vector* **w** *that you want to express as a linear combination of* **u** *and* **v**.
3. *Use the navigation bar to go through the steps of Procedure* **??**

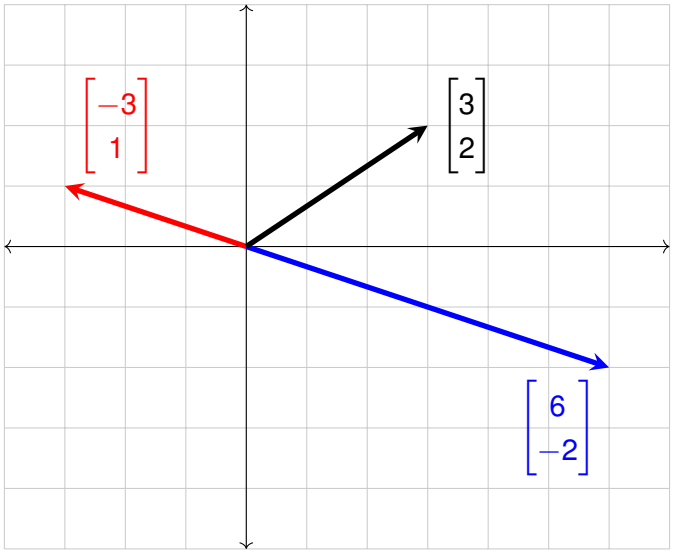
Access GeoGebra interactives through the online version of this text at
https://ximera.osu.edu/linearalgebradzv3/LinearAlgebraInteractiveIntro.

From Geometry to Algebra of Linear Combinations

One of the stipulations in Procedure **??** is that vectors **u** and **v** should be non-collinear. You can use the interactive in Exploration **??** to investigate what happens when **u** and **v** are collinear. The following example examines what happens from a geometric as well as an algebraic standpoint.

Example 33. Can the vector $\begin{bmatrix} 3 \\ 2 \end{bmatrix}$ be written as a linear combination of vectors $\begin{bmatrix} -3 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} 6 \\ -2 \end{bmatrix}$?

Explanation. We will start with a geometric approach.



Observe that $\begin{bmatrix} -3 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} 6 \\ -2 \end{bmatrix}$ are scalar multiples of each other and lie on the same line.

A linear combination of $\begin{bmatrix} -3 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} 6 \\ -2 \end{bmatrix}$ has the form:

$$a\begin{bmatrix} 6 \\ -2 \end{bmatrix} + b\begin{bmatrix} -3 \\ 1 \end{bmatrix} = a(-2)\begin{bmatrix} -3 \\ 1 \end{bmatrix} + b\begin{bmatrix} -3 \\ 1 \end{bmatrix} = (-2a + b)\begin{bmatrix} -3 \\ 1 \end{bmatrix}$$

This shows that all linear combinations of $\begin{bmatrix} -3 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} 6 \\ -2 \end{bmatrix}$ will be scalar multiples of $\begin{bmatrix} -3 \\ 1 \end{bmatrix}$, and therefore lie on the same line as $\begin{bmatrix} -3 \\ 1 \end{bmatrix}$. Since $\begin{bmatrix} 3 \\ 2 \end{bmatrix}$ does not lie on the line determined by $\begin{bmatrix} -3 \\ 1 \end{bmatrix}$ it cannot be expressed as a linear combination of $\begin{bmatrix} -3 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} 6 \\ -2 \end{bmatrix}$.

We can also address this question algebraically. To express $\begin{bmatrix} 3 \\ 2 \end{bmatrix}$ as a linear combination of $\begin{bmatrix} -3 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} 6 \\ -2 \end{bmatrix}$, we need to solve the equation.

$$a\begin{bmatrix} -3 \\ 1 \end{bmatrix} + b\begin{bmatrix} 6 \\ -2 \end{bmatrix} = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$$

This gives us a system of equations

$$\begin{aligned} -3a + 6b &= 3 \\ a - 2b &= 2 \end{aligned}$$

When you try to solve this system, you will find that the system is inconsistent. Thus, $\begin{bmatrix} 3 \\ 2 \end{bmatrix}$ cannot be written as a linear combination of $\begin{bmatrix} -3 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} 6 \\ -2 \end{bmatrix}$.

We know that there is no way to express $\begin{bmatrix} 3 \\ 2 \end{bmatrix}$ as a linear combination of vectors $\begin{bmatrix} -3 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} 6 \\ -2 \end{bmatrix}$. What would happen if we tried to apply Procedure **??** to these vectors? You can use the GeoGebra interactive in Exploration **??** to find out.

Example 34. Express $\begin{bmatrix} 2 \\ 4 \end{bmatrix}$ as a linear combination of $\begin{bmatrix} 2 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} 2 \\ -2 \end{bmatrix}$. Interpret your results geometrically.

Explanation. We need to find scalars a and b such that

$$\begin{bmatrix} 2 \\ 4 \end{bmatrix} = a \begin{bmatrix} 2 \\ 1 \end{bmatrix} + b \begin{bmatrix} 2 \\ -2 \end{bmatrix}$$

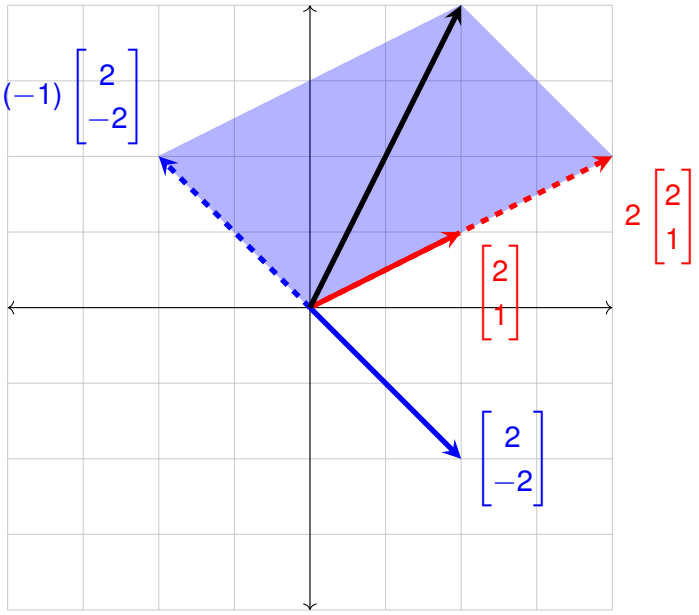
This amounts to solving a system of linear equations

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

Use your favorite method to solve this system. (Hint: adding the second equation to the first will work well for this system.) You will find that $a = 2$ and $b = -1$. Now we can write $\begin{bmatrix} 2 \\ 4 \end{bmatrix}$ as a linear combination of $\begin{bmatrix} 2 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} 2 \\ -2 \end{bmatrix}$ as follows:

$$\begin{bmatrix} 2 \\ 4 \end{bmatrix} = 2 \begin{bmatrix} 2 \\ 1 \end{bmatrix} + (-1) \begin{bmatrix} 2 \\ -2 \end{bmatrix}$$

Geometrically speaking, this means that the vector $\begin{bmatrix} 2 \\ 4 \end{bmatrix}$ is the diagonal of the parallelogram determined by $2 \begin{bmatrix} 2 \\ 1 \end{bmatrix}$ and $(-1) \begin{bmatrix} 2 \\ -2 \end{bmatrix}$. The original vectors $\begin{bmatrix} 2 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} 2 \\ -2 \end{bmatrix}$ are shown below together with the parallelogram and its diagonal.



Example 35. If possible, express $\begin{bmatrix}7\\4\\-5\end{bmatrix}$ as a linear combination of $\begin{bmatrix}1\\-2\\1\end{bmatrix}$ and $\begin{bmatrix}3\\0\\-1\end{bmatrix}$.

Explanation. We are looking for coefficients a and b such that

$$a\begin{bmatrix}1\\-2\\1\end{bmatrix} + b\begin{bmatrix}3\\0\\-1\end{bmatrix} = \begin{bmatrix}7\\4\\-5\end{bmatrix}$$

This translates into a system of equations

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

Solve this system for a and b, and enter your answers below:

$$a = \ldots\ldots \quad \text{and} \quad b = \ldots\ldots$$

We conclude that $\begin{bmatrix}7\\4\\-5\end{bmatrix}$ is a linear combination of $\begin{bmatrix}1\\-2\\1\end{bmatrix}$ and $\begin{bmatrix}3\\0\\-1\end{bmatrix}$, and write:

$$\begin{bmatrix}7\\4\\-5\end{bmatrix} = \ldots\ldots\begin{bmatrix}1\\-2\\1\end{bmatrix} + \ldots\ldots\begin{bmatrix}3\\0\\-1\end{bmatrix}$$

Example 36. Set up a system of equations that can be used to express $\begin{bmatrix} 2 \\ -1 \\ 3 \\ 0 \end{bmatrix}$ as a linear combination of $\begin{bmatrix} 1 \\ 0 \\ 4 \\ -2 \end{bmatrix}$, $\begin{bmatrix} -2 \\ -1 \\ 1 \\ -1 \end{bmatrix}$, $\begin{bmatrix} 0 \\ 4 \\ -3 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} 1 \\ 1 \\ -1 \\ 4 \end{bmatrix}$, or to determine that such a combination does not exist. Do not solve the system.

Explanation. We are looking for x_1, x_2, x_3 and x_4 such that

$$x_1 \begin{bmatrix} 1 \\ 0 \\ 4 \\ -2 \end{bmatrix} + x_2 \begin{bmatrix} -2 \\ -1 \\ 1 \\ -1 \end{bmatrix} + x_3 \begin{bmatrix} 0 \\ 4 \\ -3 \\ 1 \end{bmatrix} + x_4 \begin{bmatrix} 1 \\ 1 \\ -1 \\ 4 \end{bmatrix} = \begin{bmatrix} 2 \\ -1 \\ 3 \\ 0 \end{bmatrix}$$

This translates into the following system of equations:

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

Practice Problems

Problem 118 *Solve a system of linear equations to express* $\begin{bmatrix} -1 \\ 7 \end{bmatrix}$ *as a linear combination of* $\begin{bmatrix} 1 \\ 2 \end{bmatrix}$ *and* $\begin{bmatrix} -1 \\ 1 \end{bmatrix}$.

System of linear equations:

$$\begin{array}{rcrcr} \ldots\ldots a & + & \ldots\ldots b & = & \ldots\ldots \\ \ldots\ldots a & + & \ldots\ldots b & = & \ldots\ldots \end{array}$$

Values of a and b:

$$a = \ldots\ldots \quad \textit{and} \quad b = \ldots\ldots$$

Linear Combination:

$$\begin{bmatrix} -1 \\ 7 \end{bmatrix} = \ldots \ldots \begin{bmatrix} 1 \\ 2 \end{bmatrix} + \ldots \ldots \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

Problem 119 *Use Procedure* **??** *to express* $\begin{bmatrix} -3 \\ 0 \end{bmatrix}$ *as a linear combination of* $\begin{bmatrix} 2 \\ 4 \end{bmatrix}$ *and* $\begin{bmatrix} -1 \\ 1 \end{bmatrix}$.

Linear Combination:

$$\begin{bmatrix} -3 \\ 0 \end{bmatrix} = \ldots \ldots \begin{bmatrix} 2 \\ 4 \end{bmatrix} + \ldots \ldots \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

Problem 120 *Use two different approaches (algebraic and geometric) to explain why the vector* $\begin{bmatrix} 5 \\ 1 \end{bmatrix}$ *cannot be expressed as a linear combination of vectors* $\begin{bmatrix} 2 \\ -1 \end{bmatrix}$ *and* $\begin{bmatrix} -4 \\ 2 \end{bmatrix}$.

Problem 121 *We have seen Procedure* **??** *applied to vectors in* $\mathbb{R}^2$*. The same process can, in certain cases be applied to vectors in* $\mathbb{R}^3$*. Follow the steps in Procedure* **??** *to express one vector as a linear combination of two given vectors.*

The following GeoGebra interactive shows vectors $\mathbf{u} = \begin{bmatrix} 3 \\ 0 \\ -1 \end{bmatrix}$, $\mathbf{v} = \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}$ *and* $\mathbf{w} = \begin{bmatrix} 7 \\ 4 \\ -5 \end{bmatrix}$*. Use the navigation bar to click through the construction steps. RIGHT-CLICK and DRAG to rotate the image.*

Access GeoGebra interactives through the online version of this text at

https://ximera.osu.edu/linearalgebradzv3/LinearAlgebraInteractiveIntro.

1. *Can* **w** *be expressed as a linear combination of* **u** *and* **v***?*

   Multiple Choice:

   (i) *No, because* **w** *is not between* **u** *and* **v***.*

   (ii) *Yes, because all three vectors are in the same plane.*

   (iii) *Yes, because all three vectors are in the same plane, AND* **u** *and* **v** *are not collinear.*

2. *Use the navigation bar at the bottom of the interactive window to view construction steps of Procedure* **??** *applied to vectors* **u**, **v** *and* **w***. (Right-click and drag to rotate the image.) Use the final image to express* **w** *as a linear combination of* **v** *(blue) and* **u** *(red).*

   $$\mathbf{w} = \ldots \ldots \mathbf{v} + \ldots \ldots \mathbf{u}$$

Problem 122 *The following GeoGebra interactive shows vectors* $\mathbf{v} = \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}$, $\mathbf{u} = \begin{bmatrix} 3 \\ 0 \\ -1 \end{bmatrix}$, *and* $\mathbf{w} = \begin{bmatrix} 7 \\ 4 \\ 0 \end{bmatrix}$. *RIGHT-CLICK and DRAG to rotate the image. Use geometry to explain why* $\mathbf{w}$ *cannot be expressed as a linear combination of* $\mathbf{v}$ *and* $\mathbf{u}$.

Access GeoGebra interactives through the online version of this text at

https://ximera.osu.edu/linearalgebradzv3/LinearAlgebraInteractiveIntro.

We can also show that $\mathbf{w}$ *is not a linear combination of* $\mathbf{v}$ *and* $\mathbf{u}$ *algebraically by attempting to solve a system of equations corresponding to*

$$x_1\mathbf{v} + x_2\mathbf{u} = \mathbf{w}$$

Set up the system of equations

$$\begin{aligned} \ldots\ldots x_1 &+ \ldots\ldots x_2 = \ldots\ldots \\ \ldots\ldots x_1 &+ \ldots\ldots x_2 = \ldots\ldots \\ \ldots\ldots x_1 &+ \ldots\ldots x_2 = \ldots\ldots \end{aligned}$$

Find the reduced row echelon form.

$$\left[\begin{array}{cc|c} \ldots\ldots & \ldots\ldots & \ldots\ldots \\ \ldots\ldots & \ldots\ldots & \ldots\ldots \\ \ldots\ldots & \ldots\ldots & \ldots\ldots \end{array}\right]$$

21 Span

Linear Combinations Revisited

Recall that a vector $\mathbf{v}$ is said to be a linear combination of vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ if

$$\mathbf{v} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \ldots + a_n\mathbf{v}_n$$

for some scalars $a_1, a_2, \ldots, a_n$.

Example 37. If possible, express the given vector as a linear combination of

$$\mathbf{v}_1 = \begin{bmatrix} -2 \\ -3 \\ 4 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 2 \\ 3 \\ 2 \end{bmatrix}$$

Interpret your results geometrically.

1. $$\mathbf{u} = \begin{bmatrix} 2 \\ 3 \\ 5 \end{bmatrix}$$

2. $$\mathbf{w} = \begin{bmatrix} 5 \\ 5 \\ 1 \end{bmatrix}$$

Explanation. ?? We need to find coefficients a_1 and a_2 such that $\mathbf{u} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2$. To do this we need to solve the vector equation:

$$a_1 \begin{bmatrix} -2 \\ -3 \\ 4 \end{bmatrix} + a_2 \begin{bmatrix} 2 \\ 3 \\ 2 \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \\ 5 \end{bmatrix}$$

This equation translates into the following system:

$$\begin{aligned} -2a_1 &+ 2a_2 = 2 \\ -3a_1 &+ 3a_2 = 3 \\ 4a_1 &+ 2a_2 = 5 \end{aligned}$$

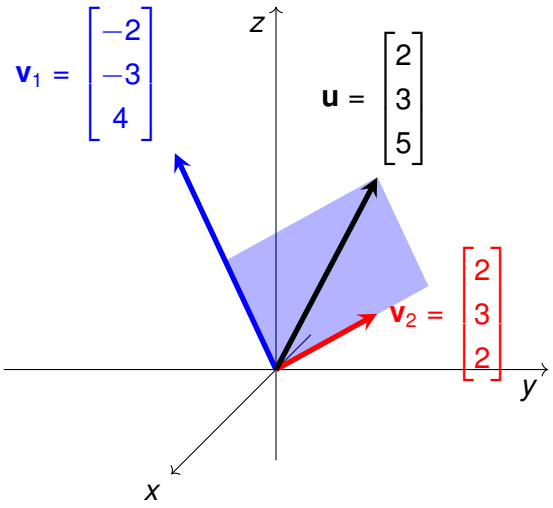
We write the system in augmented matrix form and apply elementary row operations to bring it to reduced row-echelon form.

$$\left[\begin{array}{cc|c} -2 & 2 & 2 \\ -3 & 3 & 3 \\ 4 & 2 & 5 \end{array}\right] \rightsquigarrow \left[\begin{array}{cc|c} 1 & 0 & 1/2 \\ 0 & 1 & 3/2 \\ 0 & 0 & 0 \end{array}\right]$$

This shows that $a_1 = \frac{1}{2}$ and $a_2 = \frac{3}{2}$, and we can express $\mathbf{u}$ as a linear combination of $\mathbf{v}_1$ and $\mathbf{v}_2$ as follows:

$$\frac{1}{2}\begin{bmatrix} -2 \\ -3 \\ 4 \end{bmatrix} + \frac{3}{2}\begin{bmatrix} 2 \\ 3 \\ 2 \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \\ 5 \end{bmatrix}$$

Observe that because vector $\mathbf{u}$ is a linear combination of $\mathbf{v}_1$ and $\mathbf{v}_2$, $\mathbf{u}$ is the diagonal of a parallelogram whose sides are scalar multiples of $\mathbf{v}_1$ and $\mathbf{v}_2$. As such, $\mathbf{u}$ lies in the same plane as $\mathbf{v}_1$ and $\mathbf{v}_2$, as illustrated below.



?? We need to solve the following vector equation:

$$a_1\begin{bmatrix} -2 \\ -3 \\ 4 \end{bmatrix} + a_2\begin{bmatrix} 2 \\ 3 \\ 2 \end{bmatrix} = \begin{bmatrix} 5 \\ 5 \\ 1 \end{bmatrix}$$

This equation corresponds to the system:

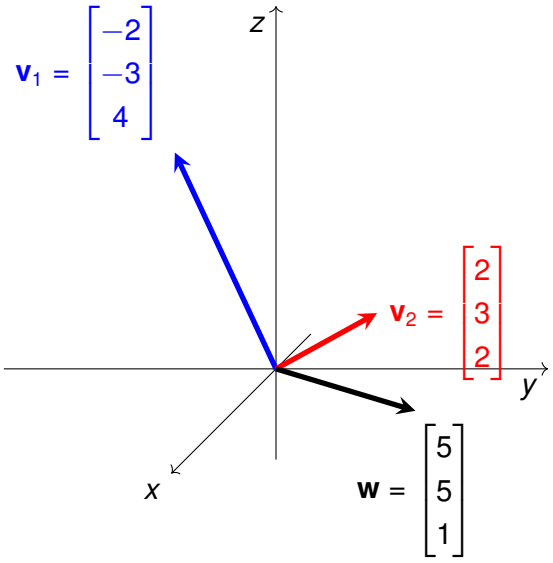
$$\begin{aligned} -2a_1 &+ 2a_2 &= 5 \\ -3a_1 &+ 3a_2 &= 5 \\ 4a_1 &+ 2a_2 &= 1 \end{aligned}$$

Writing the system in augmented matrix form and applying elementary row operations gives us the following reduced row-echelon form:

$$\left[\begin{array}{cc|c} -2 & 2 & 5 \\ -3 & 3 & 5 \\ 4 & 2 & 1 \end{array}\right] \rightsquigarrow \left[\begin{array}{cc|c} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{array}\right]$$

We conclude that there are no solutions, and $\mathbf{w}$ is not a linear combination of $\mathbf{v}_1$ and $\mathbf{v}_2$.

Geometrically, this means that $\mathbf{w}$ is not the diagonal of any parallelogram whose sides are scalar multiples of $\mathbf{v}_1$ and $\mathbf{v}_2$. Thus, $\mathbf{w}$ does not lie in the plane determined by $\mathbf{v}_1$ and $\mathbf{v}_2$.



In part **??** of Example **??** we expressed $\mathbf{u}$ as a linear combination of $\mathbf{v}_1$ and $\mathbf{v}_2$, and concluded that $\mathbf{u}$ lies in the plane determined by $\mathbf{v}_1$ and $\mathbf{v}_2$. We say that $\mathbf{u}$ is *in the span* of $\mathbf{v}_1$ and $\mathbf{v}_2$. In fact, every vector in the plane determined by $\mathbf{v}_1$ and $\mathbf{v}_2$ is in the span of $\mathbf{v}_1$ and $\mathbf{v}_2$. We say that $\mathbf{v}_1$ and $\mathbf{v}_2$ *span the plane*.

In contrast, vector $\mathbf{w}$ of part **??** of Example **??** is not a linear combination of $\mathbf{v}_1$ and $\mathbf{v}_2$. We say that $\mathbf{w}$ is not in the span of $\mathbf{v}_1$ and $\mathbf{v}_2$.

The following video takes another look at Example **??** using our new vocabulary.

YouTube link: `https://www.youtube.com/watch?v=Sum3qoOEXhY`

Definition of Span

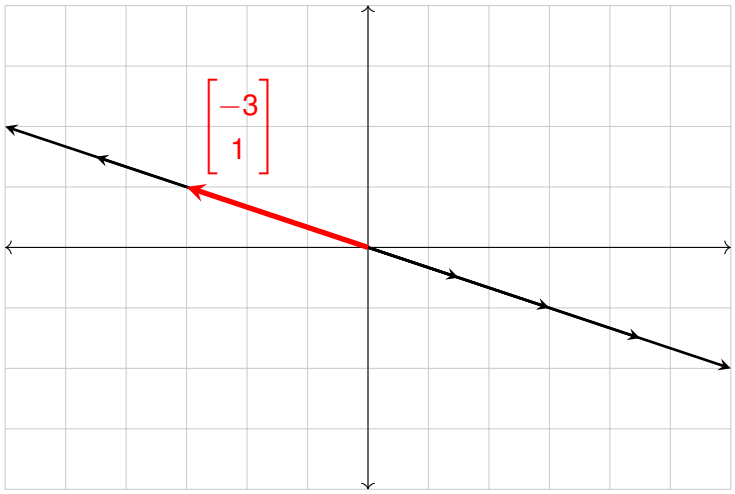
Definition 24. Let $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p$ be vectors in $\mathbb{R}^n$. The set S of all linear combinations of $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p$ is called the *span* of $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p$. We write

$$S = \text{span}(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p)$$

and we say that vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p$ *span* S. Any vector in S is said to be *in the span* of $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p$. The set $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p\}$ is called a *spanning set* for S.

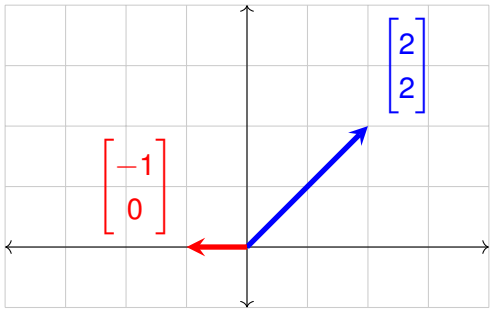
Example 38. Describe span $\left(\begin{bmatrix} -3 \\ 1 \end{bmatrix} \right)$.

Explanation. The span of $\begin{bmatrix} -3 \\ 1 \end{bmatrix}$ is the set of all linear combinations of $\begin{bmatrix} -3 \\ 1 \end{bmatrix}$. Since we are looking for linear combinations of only one vector, we are really looking for all of its scalar multiples. So, the span will be the set of all vectors of the form $\mathbf{v} = a \begin{bmatrix} -3 \\ 1 \end{bmatrix}$. All such vectors lie on the line determined by $\begin{bmatrix} -3 \\ 1 \end{bmatrix}$.



Example 39. Describe span $\left(\begin{bmatrix} 2 \\ 2 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \end{bmatrix} \right)$.

Explanation. First, observe that $\begin{bmatrix} 2 \\ 2 \end{bmatrix}$ and $\begin{bmatrix} -1 \\ 0 \end{bmatrix}$ are not scalar multiples of each other.



Geometrically, we can use Procedure **??** to express any vector of $\mathbb{R}^2$ as a linear combination of $\begin{bmatrix}2\\2\end{bmatrix}$ and $\begin{bmatrix}-1\\0\end{bmatrix}$, indicating that the two vectors span all of $\mathbb{R}^2$.

To verify this claim algebraically we will show that an arbitrary vector $\begin{bmatrix}s\\t\end{bmatrix}$ of $\mathbb{R}^2$ can be written as a linear combination of $\begin{bmatrix}2\\2\end{bmatrix}$ and $\begin{bmatrix}-1\\0\end{bmatrix}$.

Consider the vector equation:

$$a_1\begin{bmatrix}2\\2\end{bmatrix} + a_2\begin{bmatrix}-1\\0\end{bmatrix} = \begin{bmatrix}s\\t\end{bmatrix}$$

This corresponds to the system:

$$\begin{aligned} 2a_1 - a_2 &= s \\ 2a_1 &= t \end{aligned}$$

Writing the system in augmented matrix form and applying elementary row operations gives us the following reduced row-echelon form:

$$\left[\begin{array}{cc|c} 2 & -1 & s \\ 2 & 0 & t \end{array}\right] \rightsquigarrow \left[\begin{array}{cc|c} 1 & 0 & t/2 \\ 0 & 1 & t-s \end{array}\right]$$

This shows that every vector of $\mathbb{R}^2$ can be written as a linear combination of $\begin{bmatrix}2\\2\end{bmatrix}$ and $\begin{bmatrix}-1\\0\end{bmatrix}$:

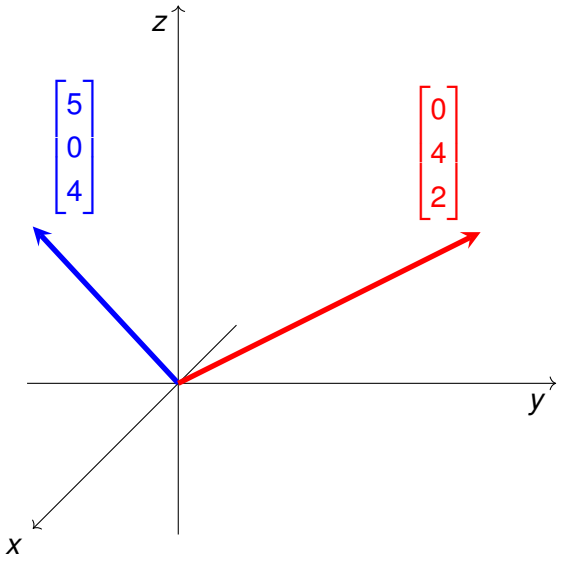
$$(t/2)\begin{bmatrix}2\\2\end{bmatrix} + (t-s)\begin{bmatrix}-1\\0\end{bmatrix} = \begin{bmatrix}s\\t\end{bmatrix}$$

We conclude that

$$\text{span}\left(\begin{bmatrix}2\\2\end{bmatrix},\begin{bmatrix}-1\\0\end{bmatrix}\right)=\mathbb{R}^2$$

Example 40. Describe span $\left(\begin{bmatrix}5\\0\\4\end{bmatrix},\begin{bmatrix}0\\4\\2\end{bmatrix}\right)$.

Explanation. First, observe that $\begin{bmatrix}5\\0\\4\end{bmatrix},\begin{bmatrix}0\\4\\2\end{bmatrix}$ are not scalar multiples of each other.

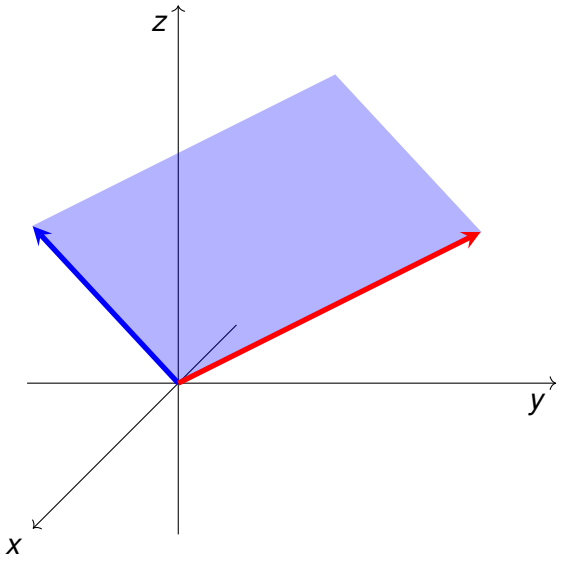


The span of $\begin{bmatrix}5\\0\\4\end{bmatrix}$ and $\begin{bmatrix}0\\4\\2\end{bmatrix}$ consists of elements of the form

$$a_1\begin{bmatrix}5\\0\\4\end{bmatrix}+a_2\begin{bmatrix}0\\4\\2\end{bmatrix}$$

Geometrically, we can interpret all such linear combinations as diagonals of parallelograms determined by scalar multiples of $\begin{bmatrix}5\\0\\4\end{bmatrix}$ and $\begin{bmatrix}0\\4\\2\end{bmatrix}$. All such diagonals will lie in the plane determined

by $\begin{bmatrix} 5 \\ 0 \\ 4 \end{bmatrix}$ and $\begin{bmatrix} 0 \\ 4 \\ 2 \end{bmatrix}$. Let this plane be called p. A portion of p is shown below.



Because Procedure **??** can be applied to vectors that lie in p just as easily as it can be applied to vectors of $\mathbb{R}^2$, we conclude that every vector in p can be expressed as a linear combination of $\begin{bmatrix} 5 \\ 0 \\ 4 \end{bmatrix}$ and $\begin{bmatrix} 0 \\ 4 \\ 2 \end{bmatrix}$. Thus,

$$\text{span}\left(\begin{bmatrix} 5 \\ 0 \\ 4 \end{bmatrix}, \begin{bmatrix} 0 \\ 4 \\ 2 \end{bmatrix}\right) = p$$

Redundant Vectors

If a friend told you that they have a line spanned by $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} 2 \\ 2 \end{bmatrix}$ and $\begin{bmatrix} 3 \\ 3 \end{bmatrix}$, you would probably think that your friend's description is a little excessive. Isn't one of the above vectors sufficient to describe the line? A line can be described as a span of one vector, but it can also be described as a span of two or more vectors. There are many advantages, however, to using the most efficient description possible. In this section we will begin to explore what makes a description "more efficient."

Exploration 123 *Consider the following collection of vectors:*

$$\left\{\begin{bmatrix}2\\-1\end{bmatrix}, \begin{bmatrix}-4\\2\end{bmatrix}, \begin{bmatrix}2\\1\end{bmatrix}\right\}$$

What is the span of these vectors? (A line, / $\mathbb{R}^2$, / A parallelogram, / A parallelepiped)

In this Exploration we will examine what can happen to the span of a collection of vectors when a vector is removed from the collection.

First, let's remove $\begin{bmatrix}2\\1\end{bmatrix}$ *from* $\left\{\begin{bmatrix}2\\-1\end{bmatrix}, \begin{bmatrix}-4\\2\end{bmatrix}, \begin{bmatrix}2\\1\end{bmatrix}\right\}$.

Which of the following is true?

Multiple Choice:

1. $\text{span}\left(\begin{bmatrix}2\\-1\end{bmatrix}, \begin{bmatrix}-4\\2\end{bmatrix}\right) = \text{span}\left(\begin{bmatrix}2\\-1\end{bmatrix}, \begin{bmatrix}-4\\2\end{bmatrix}, \begin{bmatrix}2\\1\end{bmatrix}\right)$
2. $\text{span}\left(\begin{bmatrix}2\\-1\end{bmatrix}, \begin{bmatrix}-4\\2\end{bmatrix}\right)$ *is a line*
3. $\text{span}\left(\begin{bmatrix}2\\-1\end{bmatrix}, \begin{bmatrix}-4\\2\end{bmatrix}\right) = \mathbb{R}^2$
4. $\text{span}\left(\begin{bmatrix}2\\-1\end{bmatrix}, \begin{bmatrix}-4\\2\end{bmatrix}\right)$ *is a parallelogram.*

Removing $\begin{bmatrix}2\\1\end{bmatrix}$ *from* $\left\{\begin{bmatrix}2\\-1\end{bmatrix}, \begin{bmatrix}-4\\2\end{bmatrix}, \begin{bmatrix}2\\1\end{bmatrix}\right\}$ *(changed, / did not change) the span.*

Now let's remove $\begin{bmatrix}-4\\2\end{bmatrix}$ *from the* original *collection of vectors.*

Which of the following is true?

Multiple Choice:

1. $\text{span}\left(\begin{bmatrix}2\\-1\end{bmatrix}, \begin{bmatrix}2\\1\end{bmatrix}\right) = \text{span}\left(\begin{bmatrix}2\\-1\end{bmatrix}, \begin{bmatrix}-4\\2\end{bmatrix}, \begin{bmatrix}2\\1\end{bmatrix}\right)$

2. *span* $\left(\begin{bmatrix} 2 \\ -1 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \end{bmatrix}\right)$ *is a line*

3. *span* $\left(\begin{bmatrix} 2 \\ -1 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \end{bmatrix}\right)$ *is the right side of the coordinate plane.*

4. *span* $\left(\begin{bmatrix} 2 \\ -1 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \end{bmatrix}\right)$ *is a parallelogram.*

Removing $\begin{bmatrix} -4 \\ 2 \end{bmatrix}$ *from* $\left\{\begin{bmatrix} 2 \\ -1 \end{bmatrix}, \begin{bmatrix} -4 \\ 2 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \end{bmatrix}\right\}$ *(changed, / did not change) the span.*

As you just discovered, removing a vector from a collection of vectors may or may not affect the span of the collection. We will refer to vectors that can be removed from a collection without changing the span as *redundant*. In Exploration **??**, $\begin{bmatrix} -4 \\ 2 \end{bmatrix}$ is *redundant*, while $\begin{bmatrix} 2 \\ 1 \end{bmatrix}$ is not.

Definition 25. Let $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ be a set of vectors in $\mathbb{R}^n$. If we can remove one vector without changing the span of this set, then that vector is *redundant*. In other words, if

$$\text{span}\,(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k) = \text{span}\left(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{j-1}, \mathbf{v}_{j+1}, \ldots, \mathbf{v}_k\right)$$

we say that $\mathbf{v}_j$ is a redundant element of $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$, or simply redundant.

Our next goal is to see what causes $\begin{bmatrix} -4 \\ 2 \end{bmatrix}$ of Exploration **??** to be redundant. The answer lies not in the vector itself, but in its relationship to the other vectors in the collection. Observe that $\begin{bmatrix} -4 \\ 2 \end{bmatrix} = -2\begin{bmatrix} 2 \\ -1 \end{bmatrix}$. In other words, $\begin{bmatrix} -4 \\ 2 \end{bmatrix}$ is a scalar multiple of another vector in the set. To see why this matters, let's pick an arbitrary vector $\mathbf{w} = \begin{bmatrix} 0 \\ 2 \end{bmatrix}$ in span $\left(\begin{bmatrix} 2 \\ -1 \end{bmatrix}, \begin{bmatrix} -4 \\ 2 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \end{bmatrix}\right)$. Vector $\mathbf{w}$ is in the span because it can be written as a linear combination of the three vectors as follows

$$\mathbf{w} = \begin{bmatrix} 0 \\ 2 \end{bmatrix} = \begin{bmatrix} 2 \\ -1 \end{bmatrix} + \begin{bmatrix} -4 \\ 2 \end{bmatrix} + \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

But $\begin{bmatrix} -4 \\ 2 \end{bmatrix}$ is not essential to this linear combination because it can be replaced with $-2\begin{bmatrix} 2 \\ -1 \end{bmatrix}$, as shown below.

$$\begin{bmatrix}0\\2\end{bmatrix} = \begin{bmatrix}2\\-1\end{bmatrix} + \begin{bmatrix}-4\\2\end{bmatrix} + \begin{bmatrix}2\\1\end{bmatrix} = \begin{bmatrix}2\\-1\end{bmatrix} + (-2)\begin{bmatrix}2\\-1\end{bmatrix} + \begin{bmatrix}2\\1\end{bmatrix} = -\begin{bmatrix}2\\-1\end{bmatrix} + \begin{bmatrix}2\\1\end{bmatrix}$$

Regardless of what vector **w** we write as a linear combination of $\begin{bmatrix}2\\-1\end{bmatrix}, \begin{bmatrix}-4\\2\end{bmatrix}$ and $\begin{bmatrix}2\\1\end{bmatrix}$, we will always be able to replace $\begin{bmatrix}-4\\2\end{bmatrix}$ with $-2\begin{bmatrix}2\\-1\end{bmatrix}$, placing **w** into the span of $\begin{bmatrix}2\\-1\end{bmatrix}$ and $\begin{bmatrix}2\\1\end{bmatrix}$, and making $\begin{bmatrix}-4\\2\end{bmatrix}$ redundant. (Note that we can just as easily write $\begin{bmatrix}2\\-1\end{bmatrix} = -\frac{1}{2}\begin{bmatrix}-4\\2\end{bmatrix}$, and argue that $\begin{bmatrix}2\\-1\end{bmatrix}$ is redundant.) We conclude that only one of $\begin{bmatrix}-4\\2\end{bmatrix}$ and $\begin{bmatrix}2\\-1\end{bmatrix}$ is needed to maintain the span of the original three vectors. We have

$$\text{span}\left(\begin{bmatrix}2\\-1\end{bmatrix}, \begin{bmatrix}-4\\2\end{bmatrix}, \begin{bmatrix}2\\1\end{bmatrix}\right) = \text{span}\left(\begin{bmatrix}2\\-1\end{bmatrix}, \begin{bmatrix}2\\1\end{bmatrix}\right) = \text{span}\left(\begin{bmatrix}-4\\2\end{bmatrix}, \begin{bmatrix}2\\1\end{bmatrix}\right)$$

The left-most collection in this expression contains redundant vectors; the other two collections do not.

In Exploration **??** we found one vector to be redundant because we could replace it with a scalar multiple of another vector in the set. The following Exploration delves into what happens when a vector in a given set is a linear combination of the other vectors.

Exploration 124 *Consider the set of vectors*

$$\left\{\begin{bmatrix}1\\2\\-1\end{bmatrix}, \begin{bmatrix}2\\0\\1\end{bmatrix}, \begin{bmatrix}4\\4\\-1\end{bmatrix}\right\}$$

The three vectors are shown below. RIGHT-CLICK and DRAG to rotate the interactive graph.

Access GeoGebra interactives through the online version of this text at

https://ximera.osu.edu/linearalgebradzv3/LinearAlgebraInteractiveIntro.

span $\left(\begin{bmatrix}1\\2\\-1\end{bmatrix}, \begin{bmatrix}2\\0\\1\end{bmatrix}, \begin{bmatrix}4\\4\\-1\end{bmatrix}\right)$ *is (A line, / A plane, /* $\mathbb{R}^3$*, / A parallelepiped)*

Can we remove one of the vectors from the set without changing the span? Observe that we can write $\begin{bmatrix} 4 \\ 4 \\ -1 \end{bmatrix}$ *as a linear combination of the other two vectors*

$$\begin{bmatrix} 4 \\ 4 \\ -1 \end{bmatrix} = 2 \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix} + 1 \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix} \tag{24}$$

This means that we can write any vector in span $\left(\begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 4 \\ 4 \\ -1 \end{bmatrix} \right)$ *as a linear combination of only* $\begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}$ *and* $\begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix}$ *by replacing* $\begin{bmatrix} 4 \\ 4 \\ -1 \end{bmatrix}$ *with the expression in (**??**). For example,*

$$\begin{bmatrix} 7 \\ 6 \\ -1 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix} + \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix} + \begin{bmatrix} 4 \\ 4 \\ -1 \end{bmatrix} = \ldots \ldots \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix} + \ldots \ldots \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix}$$

We have

$$span \left(\begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 4 \\ 4 \\ -1 \end{bmatrix} \right) = span \left(\begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix} \right)$$

We conclude that vector $\begin{bmatrix} 4 \\ 4 \\ -1 \end{bmatrix}$ *is redundant. Can each of the other two vectors in the set* $\left\{ \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 4 \\ 4 \\ -1 \end{bmatrix} \right\}$ *be considered redundant? You will address this question in the problem set.*

Collections of vectors that do not contain redundant vectors are very important in linear algebra. In subsequent sections, we will formally introduce such collections as *linearly independent*. Collections of vectors that contain redundant vectors will be called *linearly dependent*.

Practice Problems

Problem 125 *Choose the best description for the set below.*

$$\textit{span}\left(\begin{bmatrix}1\\1\\-2\end{bmatrix},\begin{bmatrix}2\\2\\-4\end{bmatrix}\right)$$

Multiple Choice:

1. *Plane in* $\mathbb{R}^3$
2. *Line in* $\mathbb{R}^2$
3. *Line in* $\mathbb{R}^3$
4. $\mathbb{R}^3$
5. $\mathbb{R}^2$

Problem 126 *Choose the best description for the set below.*

$$\textit{span}\left(\begin{bmatrix}1\\-2\end{bmatrix},\begin{bmatrix}1\\0\end{bmatrix}\right)$$

Multiple Choice:

1. *Plane in* $\mathbb{R}^3$
2. *Line in* $\mathbb{R}^2$
3. *Line in* $\mathbb{R}^3$
4. $\mathbb{R}^3$
5. $\mathbb{R}^2$

Problem 127 *Choose the best description for the set below.*

$$\textit{span}\left(\begin{bmatrix}-3\\1\end{bmatrix},\begin{bmatrix}6\\-2\end{bmatrix},\begin{bmatrix}3\\-1\end{bmatrix}\right)$$

Multiple Choice:

1. *Plane in* $\mathbb{R}^3$
2. *Line in* $\mathbb{R}^2$
3. *Line in* $\mathbb{R}^3$
4. $\mathbb{R}^3$
5. $\mathbb{R}^2$

Problem 128 *Which of the following pairs of sets are equal? Select all that apply.*

Select All Correct Answers:

1. $V = span\left(\begin{bmatrix}5\\0\\0\end{bmatrix}, \begin{bmatrix}10\\0\\0\end{bmatrix}, \begin{bmatrix}0\\0\\-4\end{bmatrix}\right)$ *and* $W = span\left(\begin{bmatrix}-2\\0\\0\end{bmatrix}, \begin{bmatrix}0\\0\\1\end{bmatrix}\right)$
2. $V = span\left(\begin{bmatrix}5\\3\end{bmatrix}, \begin{bmatrix}10\\-1\end{bmatrix}, \begin{bmatrix}0\\2\end{bmatrix}\right)$ *and* $W = span\left(\begin{bmatrix}-2\\0\end{bmatrix}, \begin{bmatrix}0\\1\end{bmatrix}\right)$
3. $V = span\left(\begin{bmatrix}1\\-2\\4\end{bmatrix}\right)$ *and* $W = span\left(\begin{bmatrix}-1\\2\\-4\end{bmatrix}, \begin{bmatrix}0\\0\\1\end{bmatrix}\right)$
4. $V = span\left(\begin{bmatrix}5\\0\end{bmatrix}\right)$ *and* $W = span\left(\begin{bmatrix}-2\\0\end{bmatrix}, \begin{bmatrix}1\\0\end{bmatrix}, \begin{bmatrix}-4\\0\end{bmatrix}, \begin{bmatrix}0\\0\end{bmatrix}\right)$

Problem 129 *Let* $\mathbf{v} = \begin{bmatrix}3\\4\\5\end{bmatrix}$*. Give an example of at least one vector* $\mathbf{w}$ *such that* $\mathbf{v}$*,* $\mathbf{w}$ *do NOT span a plane in* $\mathbb{R}^3$*. Describe span*$(\mathbf{v}, \mathbf{w})$*.*

Problem 130 *Prove or disprove. The zero vector of* $\mathbb{R}^n$ *is contained in the span of any collection of vectors of* $\mathbb{R}^n$*.*

Problem 131 *In Exploration* **??** *we considered the following set of vectors*

$$\left\{ \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 4 \\ 4 \\ -1 \end{bmatrix} \right\}$$

and demonstrated that $\begin{bmatrix} 4 \\ 4 \\ -1 \end{bmatrix}$ *is redundant by using the fact that it is a linear combination of the other two vectors.*

1. *Express each of* $\begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}$ *and* $\begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix}$ *as a linear combination of the remaining vectors.*

$$\begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix} = \ldots \ldots \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix} + \ldots \ldots \begin{bmatrix} 4 \\ 4 \\ -1 \end{bmatrix}$$

$$\begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix} = \ldots \ldots \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix} + \ldots \ldots \begin{bmatrix} 4 \\ 4 \\ -1 \end{bmatrix}$$

2. *Which of the following is NOT true?*

Multiple Choice:

(i) *If* $\mathbf{w}$ *is in span* $\left(\begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 4 \\ 4 \\ -1 \end{bmatrix} \right)$*, then* $\mathbf{w}$ *is in span* $\left(\begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 4 \\ 4 \\ -1 \end{bmatrix} \right)$.

(ii) *Both* $\begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}$ *and* $\begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix}$ *are redundant in* $\left\{ \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 4 \\ 4 \\ -1 \end{bmatrix} \right\}$.

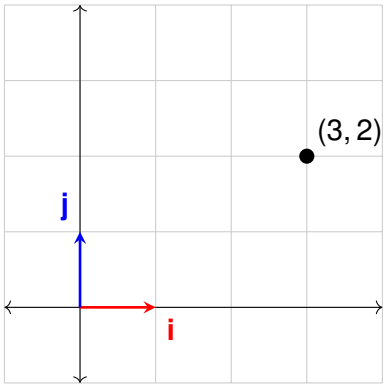
(iii) *We can remove* $\begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}$ *and* $\begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix}$ *from* $\left\{ \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 4 \\ 4 \\ -1 \end{bmatrix} \right\}$ *at the same time without affecting the span.*

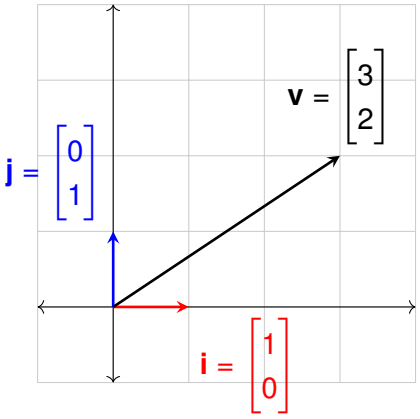
Problem 132 *Show that if the zero vector is part of a collection of two or more vectors, the zero vector is redundant.*

22 What Makes a Good Coordinate System?

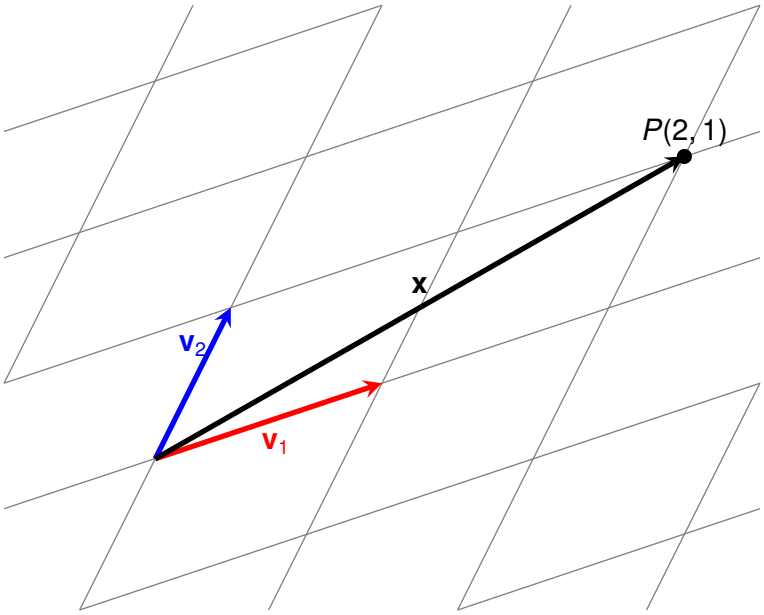
How to Read a Coordinate System

Let's take a fresh look at the familiar rectangular coordinate system. We can impose such a coordinate system on a plane by picking the origin, and two orthogonal unit vectors **i** and **j** to determine the axes. In the diagram on the left, the point $P = (3, 2)$ can be located by moving 3 units from the origin in the direction of **i**, then moving 2 units in the direction of **j**. In this coordinate system, vector **v** drawn from the origin to point P, is represented by $\begin{bmatrix} 3 \\ 2 \end{bmatrix}$. We can express **v** as a linear combination of **i** and **j** as $\mathbf{v} = 3\mathbf{i} + 2\mathbf{j}$. Note that the coefficients of the linear combination are the first and second components of $\begin{bmatrix} 3 \\ 2 \end{bmatrix}$.





Orthogonal unit vectors are often a convenient choice for establishing a coordinate system, but occasionally, using other vectors is preferable. An example of an alternative coordinate grid determined by vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ is shown below.



Fortunately, finding points and constructing vectors in such coordinate systems relies on the same principles used in the familiar rectangular coordinate system. To reach point P, we start at the origin, then travel 2 copies of $\mathbf{v}_1$ in the direction of $\mathbf{v}_1$, followed by one copy of $\mathbf{v}_2$ in the direction of $\mathbf{v}_2$. We can say that in this coordinate system, P has coordinates $(2, 1)$. Also, observe that vector $\mathbf{x}$ can be written as a linear combination of $\mathbf{v}_1$ and $\mathbf{v}_2$ as follows

$$\mathbf{x} = 2\mathbf{v}_1 + \mathbf{v}_2$$

Idea 1. We take it for granted that the x-coordinate (the coefficient in front of $\mathbf{i}$) is the first component in the ordered pair of coordinates, and the y-coordinate (the coefficient in front of $\mathbf{j}$) is the second component.

In the above setup, we chose to list the coefficient in front of $\mathbf{v}_1$ as the first coordinate of P and the coefficient in front of $\mathbf{v}_2$ as the second coordinate.

How would the coordinates of P be impacted if we had chosen a different order? Think about why establishing (or knowing) the order in which $\mathbf{v}_1$ and $\mathbf{v}_2$ are used is important.

Exploration 133 *The following interactive demonstrates how to find coordinates for a point using a coordinate grid defined using vectors $\overrightarrow{OA}$ and $\overrightarrow{OB}$ (in that order).*

To use the interactive

- *Move the sliders to change the coordinates of P within this coordinate system;*
- *Move the tips of vectors $\overrightarrow{OA}$ and $\overrightarrow{OB}$ to change the coordinate system.*

Question 133.1 *Move the tips of vectors $\overrightarrow{OA}$ and $\overrightarrow{OB}$ to change the coordinate system. Explain why the coordinates of P do not change while P moves around the plane.*

GeoGebra link: https://www.geogebra.org/m/qxjtdgkc

Existence and Uniqueness of Representation

We now understand that we can take two vectors and use them to impose a coordinate system on a plane. In this section we will explore the properties of vectors that determine a "good" coordinate system. The following questions will be central to our discussion:

- Can ANY two vectors be used to form a coordinate system in a plane?
- Can we use more than two vectors to determine a coordinate system in a plane?
- What constitutes a "good" coordinate system?

Exploration 134 *Use the interactive below to answer the questions.*

Tip: To better navigate between the interactive and the questions, we suggest that you use a second device to manipulate the interactive. Refresh your browser to reset the interactive to default settings.

GeoGebra link: https://www.geogebra.org/m/qw5dpmqq

Question 134.1 *Find the coordinates for each P_i with respect to the given coordinate system. Your coordinates should be of the form (A-coordinate, B-coordinate).*

$P_1 = (\ldots\ldots, \ldots\ldots)$

$P_2 = (\ldots\ldots, \ldots\ldots)$

$P_3 = (\ldots\ldots, \ldots\ldots)$

$P_4 = (\ldots\ldots, \ldots\ldots)$

$P_5 = (\ldots\ldots, \ldots\ldots)$

Question 134.2 *Move point B to coincide with P_3. List the coordinates for each P_i with respect to the new coordinate system. Your coordinates should be of the form (A-coordinate, B-coordinate).*

$P_1 = (\ldots\ldots, \ldots\ldots)$

$$P_2 = (\ldots\ldots, \ldots\ldots)$$
$$P_3 = (\ldots\ldots, \ldots\ldots)$$
$$P_4 = (\ldots\ldots, \ldots\ldots)$$
$$P_5 = (\ldots\ldots, \ldots\ldots)$$

How do these coordinates compare to the coordinates in the previous question? Explain why this is happening. (Hint: you can refresh your browser to return to the original coordinate system for comparison.)

Question 134.3 *Refresh your browser to return to the original coordinate system. Move point A to coincide with P_1. List the coordinates for each P_i with respect to the new coordinate system. Your coordinates should be of the form (A-coordinate, B-coordinate).*

$$P_1 = (\ldots\ldots, \ldots\ldots)$$
$$P_2 = (\ldots\ldots, \ldots\ldots)$$
$$P_3 = (\ldots\ldots, \ldots\ldots)$$
$$P_4 = (\ldots\ldots, \ldots\ldots)$$
$$P_5 = (\ldots\ldots, \ldots\ldots)$$

How do these coordinates compare to the coordinates in the original coordinate system? Explain why this is happening. (Hint: you can refresh your browser to return to the original coordinate system for comparison.)

Question 134.4 *Refresh your browser to return to the original coordinate system. Express each $\overrightarrow{OP}_i$ as a linear combination of $\overrightarrow{OA}$ and $\overrightarrow{OB}$.*

$$\overrightarrow{OP}_1 = \ldots\ldots \overrightarrow{OA} + \ldots\ldots \overrightarrow{OB}$$
$$\overrightarrow{OP}_2 = \ldots\ldots \overrightarrow{OA} + \ldots\ldots \overrightarrow{OB}$$
$$\overrightarrow{OP}_3 = \ldots\ldots \overrightarrow{OA} + \ldots\ldots \overrightarrow{OB}$$
$$\overrightarrow{OP}_4 = \ldots\ldots \overrightarrow{OA} + \ldots\ldots \overrightarrow{OB}$$
$$\overrightarrow{OP}_5 = \ldots\ldots \overrightarrow{OA} + \ldots\ldots \overrightarrow{OB}$$

Discuss the relationship between your answers to the first question and your answers here. Observe that each vector $\overrightarrow{OP}_i$ is in span $\left(\overrightarrow{OA}, \overrightarrow{OB}\right)$.

Question 134.5 *Move point B to coincide with P_2. What do you observe? Can we express every point in the plane as a linear combination of $\overrightarrow{OA}$ and $\overrightarrow{OB}$? Do vectors $\overrightarrow{OA}$ and $\overrightarrow{OB}$ determine a good coordinate system for the plane? Can we express some points in the plane as linear combinations of $\overrightarrow{OA}$ and $\overrightarrow{OB}$?*

Write a short paragraph to recast your answers above in terms of span. Can we express every point in the plane as a linear combination of $\overrightarrow{OA}$ and $\overrightarrow{OB}$? Do vectors $\overrightarrow{OA}$ and $\overrightarrow{OB}$ span the plane? Can we express some points in the plane as linear combinations of $\overrightarrow{OA}$ and $\overrightarrow{OB}$? What is the span of $\overrightarrow{OA}$ and $\overrightarrow{OB}$?

Exploration 135 *We will use the same set-up as in the previous exploration but introduce an additional vector $\overrightarrow{OC}$. Tips: To better navigate between the interactive and the questions, we suggest that you use a second device to manipulate the interactive. Refresh your browser to reset the interactive to default settings.*

GeoGebra link: https://www.geogebra.org/m/b3k96x2w

Question 135.1 *Let's attempt to define a coordinate system using three vectors: $\overrightarrow{OA}$, $\overrightarrow{OB}$, and $\overrightarrow{OC}$. If the coordinates are to be of the form (A-coordinate, B-coordinate, C-coordinate), how many ways do you think there would be to express P_1? Fill in the missing coordinates for P_1 below to get a sense of the variety of possibilities.*

$$P_1 = (\ldots\ldots, \ldots\ldots, 0)$$

$$P_1 = (0, \ldots\ldots, \ldots\ldots)$$

$$P_1 = (\ldots\ldots, 0, \ldots\ldots)$$

$$P_1 = (\ldots\ldots, -1, -1)$$

Question 135.2 *Based on your work above, express $\overrightarrow{OP_1}$ as a linear combination of $\overrightarrow{OA}$, $\overrightarrow{OB}$, and $\overrightarrow{OC}$. How many ways do you think there are to do this?*

Question 135.3 *Compare and contrast the coordinate systems in this exploration and Exploration* ***??****.*

We now return to the question of what makes a good coordinate system. In this section we restricted our explorations to the plane but the same principles apply in other settings. (What does a coordinate grid look like in $\mathbb{R}^3$?)

Idea 2. What makes a good coordinate system?

- Every point has coordinates associated with it;
- Each point has exactly one coordinate representation.

We can recast the above idea in terms of vectors. First, the vectors that determine the axes must span the plane. This would guarantee that every vector in the plane can be written as a linear combination of the axes vectors. Second, we want that linear combination to be unique. Exploration **??** demonstrates how picking two vectors that do not span the plane results in a bad coordinate system. Exploration **??** shows that even when the axes vectors span the plane, using too many vectors as axes vectors results in loss of uniqueness of representation.

In the previous section, you learned how to address questions related to span. This will help you tackle existence of representation. In the next section, we will address the question of uniqueness.

23 Linear Independence

In the last section we considered properties a set of vectors should possess in order to be used to define a "good" coordinate system for a plane. Let S be a set of vectors in the plane. We have already established that in order to use vectors in S to define the axes of a "good" coordinate system, vectors in S must span the plane. This would guarantee that every vector in the plane can be written as a linear combination of elements of S.

In addition, we observed that we want each vector in the plane to be represented as a *unique* linear combination of elements of S. This gives us motivation to consider the question of uniqueness. We will now step away from coordinate systems in the plane and turn to this question in a more general setting.

Let $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ be vectors of $\mathbb{R}^n$. Let $V = \text{span}(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k)$. Then every vector in V can be written as a linear combination of $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ in at least one way. Our interest here is in spanning sets where each vector in V has exactly one representation as a linear combination of $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$.

Suppose that two linear combinations are equal

$$a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_k\mathbf{v}_k = b_1\mathbf{v}_1 + b_2\mathbf{v}_2 + \cdots + b_k\mathbf{v}_k$$

We are looking for a condition on the set $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ that guarantees that this representation is unique. This amounts to showing that $a_i = b_i$ for each i. Taking all terms to the left side gives us

$$(a_1 - b_1)\mathbf{v}_1 + (a_2 - b_2)\mathbf{v}_2 + \cdots + (a_k - b_k)\mathbf{v}_k = \mathbf{0}$$

So the condition that guarantees uniqueness is that all coefficients $(a_i - b_i)$ must be zero. This motivates the following definition.

Definition 26 (Linear Independence)**.** Let $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ be vectors of $\mathbb{R}^n$. We say that the set $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ is *linearly independent* if the only solution to

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \ldots + c_p\mathbf{v}_k = \mathbf{0} \tag{25}$$

is the *trivial solution* $c_1 = c_2 = \ldots = c_k = 0$.

If, in addition to the trivial solution, any *non-trivial solutions* (not all $c_1, c_2, \ldots, c_k$ are zero) exist, then we say that the set $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ is *linearly dependent*.

Observation 2. The trivial solution $c_1 = c_2 = \ldots = c_k = 0$ is always a solution to $c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \ldots + c_p\mathbf{v}_k = \mathbf{0}$ because $0\mathbf{v}_1 + 0\mathbf{v}_2 + \ldots + 0\mathbf{v}_k = \mathbf{0}$. So, when checking for linear independence, the question is not whether the trivial solution is a solution, but whether it is the ONLY solution.

Remark 2. Given a set of vectors $X = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ we can now ask the following questions:

1. Are the vectors in X linearly dependent according to Definition **??**?
2. Can we write one element of X as a linear combination of the others?
3. Does X contain redundant vectors?

It turns out that these questions are equivalent. In other words, if the answer to one of them is "YES", the answer to the other two is also "YES". Conversely, if the answer to one of them is "NO", then the answer to the other two is also "NO". We will start by illustrating this idea with an example, then conclude this section by formally proving the equivalency in Theorem **??**.

Example 41. What can we say about the following sets of vectors in light of Remark **??**?

1. $$\left\{ \begin{bmatrix} 2 \\ -3 \end{bmatrix}, \begin{bmatrix} 0 \\ 3 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ -2 \end{bmatrix} \right\}$$
2. $$\left\{ \begin{bmatrix} 2 \\ 1 \\ 4 \end{bmatrix}, \begin{bmatrix} -3 \\ 1 \\ 1 \end{bmatrix} \right\}$$

Explanation. ?? We will start by addressing linear independence. To do so, we will solve the vector equation

$$c_1 \begin{bmatrix} 2 \\ -3 \end{bmatrix} + c_2 \begin{bmatrix} 0 \\ 3 \end{bmatrix} + c_3 \begin{bmatrix} 1 \\ -1 \end{bmatrix} + c_4 \begin{bmatrix} 1 \\ -2 \end{bmatrix} = \mathbf{0} \tag{26}$$

Clearly $c_1 = c_2 = c_3 = c_4 = 0$ is a solution to the equation. The question is whether another solution exists.

The vector equation translates into the following system:

$$\begin{array}{rcrcrcrcl} 2c_1 & & & + & c_3 & + & c_4 & = & 0 \\ -3c_1 & + & 3c_2 & - & c_3 & - & 2c_4 & = & 0 \end{array}$$

Writing the system in augmented matrix form and applying elementary row operations gives us the following reduced row-echelon form:

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

This shows that (**??**) has infinitely many solutions:

$$c_1 = -\frac{1}{2}s - \frac{1}{2}t, \quad c_2 = -\frac{1}{6}s + \frac{1}{6}t, \quad c_3 = s, \quad c_4 = t$$

Letting $t = s = 6$, we obtain the following:

$$-6\begin{bmatrix}2\\-3\end{bmatrix} + 0\begin{bmatrix}0\\3\end{bmatrix} + 6\begin{bmatrix}1\\-1\end{bmatrix} + 6\begin{bmatrix}1\\-2\end{bmatrix} = \mathbf{0} \tag{27}$$

We conclude that the vectors are linearly dependent.

Observe that (**??**) allows us to solve for one of the vectors and express it as a linear combination of the others. For example,

$$\begin{bmatrix}2\\-3\end{bmatrix} = 0\begin{bmatrix}0\\3\end{bmatrix} + \begin{bmatrix}1\\-1\end{bmatrix} + \begin{bmatrix}1\\-2\end{bmatrix} \tag{28}$$

This would not be possible if a nontrivial solution to the equation

$$c_1\begin{bmatrix}2\\-3\end{bmatrix} + c_2\begin{bmatrix}0\\3\end{bmatrix} + c_3\begin{bmatrix}1\\-1\end{bmatrix} + c_4\begin{bmatrix}1\\-2\end{bmatrix} = \mathbf{0}$$

did not exist.

Using the linear combination in (**??**) and the argument of Exploration **??**, we conclude that $\begin{bmatrix}2\\-3\end{bmatrix}$ is redundant in

$$\left\{\begin{bmatrix}2\\-3\end{bmatrix}, \begin{bmatrix}0\\3\end{bmatrix}, \begin{bmatrix}1\\-1\end{bmatrix}, \begin{bmatrix}1\\-2\end{bmatrix}\right\}$$

We find that the answer to all questions in Remark **??** is “YES”.

?? To address linear independence, we need to solve the equation

$$c_1\begin{bmatrix}2\\1\\4\end{bmatrix} + c_2\begin{bmatrix}-3\\1\\1\end{bmatrix} = \mathbf{0}$$

Converting the equation to augmented matrix form and performing row reduction gives us

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

This shows that $c_1 = c_2 = 0$ is the only solution. Therefore the two vectors are linearly independent.

Furthermore, we cannot write one of the vectors as a linear combination of the other. (Do you see that the only way this would be possible with a set of two vectors is if they were scalar multiples of each other?)

Finally, we observe that removing either vector would change the span from a plane in $\mathbb{R}^3$ to a line in $\mathbb{R}^3$, so the answer to all three questions in Remark **??** is "NO".

We now formalize Remark **??** as a theorem.

Theorem 14. Let $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ be a set of vectors in $\mathbb{R}^n$ containing two or more vectors. The following conditions are equivalent.

1. $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ are linearly dependent.
2. One of $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ can be expressed as a linear combination of the others.
3. The set $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ contains redundant vectors.

Proof **??** $\implies$ **??** If $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ are linearly dependent, then

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \ldots + c_j\mathbf{v}_j + \ldots + c_k\mathbf{v}_k = \mathbf{0}$$

has a non-trivial solution. In other words at least one of the constants, say c_j, does not equal zero. This allows us to solve for $\mathbf{v}_j$:

$$-c_j\mathbf{v}_j = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \ldots + c_k\mathbf{v}_k$$

$$\mathbf{v}_j = -\frac{c_1}{c_j}\mathbf{v}_1 - \frac{c_2}{c_j}\mathbf{v}_2 - \ldots - \frac{c_k}{c_j}\mathbf{v}_k$$

(Do you see why it was important to have one of the constants nonzero?) This shows that $\mathbf{v}_j$ may be expressed as a linear combination of the other vectors.

?? $\implies$ **??** First, suppose $\mathbf{v}_j$ is a linear combination of $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$. We will show that $\mathbf{v}_j$ is redundant by showing that

$$\text{span}\,(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k) = \text{span}\left(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{j-1}, \mathbf{v}_{j+1}, \ldots, \mathbf{v}_k\right).$$

To show equality of the two spans we will pick a vector in the left span and show that it is also an element of the span on the right. Then, we will pick a vector in the right span and show that it is also an element of the span on the left, and we will conclude that the sets are equal.

Observe that if $\mathbf{w}$ is in span $\left(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{j-1}, \mathbf{v}_{j+1}, \ldots, \mathbf{v}_k\right)$, then it has to be in span $(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k)$. (Why?)

Now suppose $\mathbf{w}$ is in span $(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k)$. We need to show that $\mathbf{w}$ is also in span $\left(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{j-1}, \mathbf{v}_{j+1}, \ldots, \mathbf{v}_k\right)$.

By assumption, we can write $\mathbf{v}_j$ as

$$\mathbf{v}_j = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_{j-1}\mathbf{v}_{j-1} + a_{j+1}\mathbf{v}_{j+1} + \cdots + a_k\mathbf{v}_k. \tag{29}$$

Since $\mathbf{w}$ is in span $(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k)$, we have

$$\mathbf{w} = b_1\mathbf{v}_1 + b_2\mathbf{v}_2 + \cdots + b_j\mathbf{v}_j + \cdots + b_k\mathbf{v}_k.$$

Substituting the expression in (**??**) for $\mathbf{v}_j$ and simplifying, we obtain the following

$$\begin{aligned}\mathbf{w} = (b_1 + b_j a_1)\mathbf{v}_1 + (b_2 + b_j a_2)\mathbf{v}_2 \quad &+ \quad \ldots \\ &+ \quad (b_{j-1} + b_j a_{j-1})\mathbf{v}_{j-1} \\ &+ \quad (b_{j+1} + b_j a_{j+1})\mathbf{v}_{j+1} \\ &+ \quad \ldots \\ &+ \quad (b_k + b_j a_k)\mathbf{v}_k.\end{aligned}$$

This shows that $\mathbf{w}$ is in span $\left(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{j-1}, \mathbf{v}_{j+1}, \ldots, \mathbf{v}_k\right)$. We now have

$$\text{span}\,(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k) = \text{span}\left(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{j-1}, \mathbf{v}_{j+1}, \ldots, \mathbf{v}_k\right),$$

which shows that $\mathbf{v}_j$ is redundant.

?? $\Longrightarrow$ **??** Suppose that $\mathbf{v}_j$ is redundant, so that

$$\text{span}\,(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k) = \text{span}\left(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{j-1}, \mathbf{v}_{j+1}, \ldots, \mathbf{v}_k\right).$$

Consider a vector $\mathbf{w}$ in span $(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k)$

$$\mathbf{w} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_j\mathbf{v}_j + \cdots + a_k\mathbf{v}_k \tag{30}$$

Since the span contains ALL possible linear combinations of $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$, we may choose $\mathbf{w}$ such that $a_j \neq 0$.

By assumption, $\mathbf{w}$ is also in span $\left(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{j-1}, \mathbf{v}_{j+1}, \ldots, \mathbf{v}_k\right)$. Therefore, we can express $\mathbf{w}$ as a linear combination

$$\mathbf{w} = b_1\mathbf{v}_1 + b_2\mathbf{v}_2 + \cdots + b_{j-1}\mathbf{v}_{j-1} + b_{j+1}\mathbf{v}_{j+1} + \cdots + b_k\mathbf{v}_k. \tag{31}$$

We complete the proof by showing there exists a non-trivial solution to

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \ldots + c_j\mathbf{v}_j + \ldots + c_k\mathbf{v}_k = \mathbf{0}. \tag{32}$$

Subtracting expression (**??**) from (**??**) we obtain

$$\begin{aligned}\mathbf{0} = \mathbf{w} - \mathbf{w} = (a_1 - b_1)\mathbf{v}_1 &+ \ldots \\ &+ (a_{j-1} - b_{j-1})\mathbf{v}_{j-1} + a_j\mathbf{v}_j + (a_{j+1} - b_{j+1})\mathbf{v}_{j+1} \\ &+ \ldots \\ &+ (a_k - b_k)\mathbf{v}_k\end{aligned}$$

Recall that we ensured that $a_j \neq 0$. This implies that we have a non-trivial solution to Equation **??**.

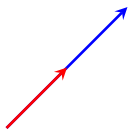
These three parts of the proof show that if one of the conditions is true, all three must be true. It is a logical consequence that if one of the three conditions is false, all three must be false. ■

Geometry of Linearly Dependent and Linearly Independent Vectors

Theorem **??** gives us a convenient ways of looking at linear dependence/independence geometrically. When looking at two or more vectors, we ask, "can one of the vectors be written as a linear combination of the others?" We can also ask, "is one of the vectors redundant?" If the answer to either of these questions is "YES", then the vectors are linearly dependent.

A Set of Two Vectors

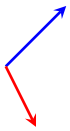
Two vectors are linearly dependent if and only if one is a scalar multiple of the other. Two nonzero linearly dependent vectors may look like this:



or like this:



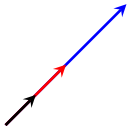
Two linearly independent vectors will look like this:



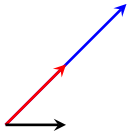
A Set of Three Vectors

Given a set of three nonzero vectors, we have the following possibilities:

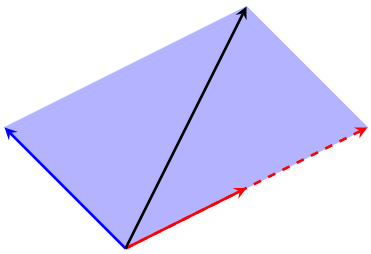
- (Linearly Dependent Vectors) The three vectors are scalar multiples of each other.



- (Linearly Dependent Vectors) Two of the vectors are scalar multiples of each other.



- (Linearly Dependent Vectors) One vector can be viewed as the diagonal of a parallelogram determined by scalar multiples of the other two vectors. All three vectors lie in the same plane.



- (Linearly Independent Vectors) A set of three vectors is linearly independent if the vectors do not lie in the same plane. For example, vectors **i**, **j** and **k** are linearly independent.

Practice Problems

Problem 136 *Are the vectors below linearly independent?*

$$\begin{bmatrix} -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 4 \\ -1 \end{bmatrix}$$

Multiple Choice:

1. *Yes*
2. *No*

Problem 137 *Are the vectors below linearly independent?*

$$\begin{bmatrix} 1 \\ 0 \\ 5 \end{bmatrix}, \begin{bmatrix} 2 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$$

Multiple Choice:

1. *Yes*
2. *No*

Problem 138 *Are the vectors below linearly independent?*

$$\begin{bmatrix} 3 \\ 0 \\ 5 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \\ 2 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ -5 \end{bmatrix}$$

Multiple Choice:

1. *Yes*
2. *No*

Problem 139 *Are the vectors below linearly independent?*

$$\begin{bmatrix} 3 \\ 1 \\ 4 \\ 1 \end{bmatrix}, \begin{bmatrix} -2 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

Multiple Choice:

1. *Yes*
2. *No*

Problem 140 *(True or False?) Any set containing the zero vector is linearly dependent.*

Multiple Choice:

1. *TRUE*
2. *FALSE*

Problem 141 *(True or False?) A set containing five vectors in $\mathbb{R}^2$ is linearly dependent.*

Multiple Choice:

1. *TRUE*
2. *FALSE*

Problem 142 *Given that*

$$0\mathbf{v}_1 + 0\mathbf{v}_2 + 0\mathbf{v}_3 = \mathbf{0}$$

what (if anything) can we conclude about linear independence of vectors $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$?

Multiple Choice:

1. *The vectors are linearly independent*
2. *The vectors are linearly dependent*
3. *There is not enough information given to make a determination*

Problem 143 *Given that*

$$3\mathbf{v}_1 + 4\mathbf{v}_2 - \mathbf{v}_3 = \mathbf{0}$$

what (if anything) can we conclude about linear independence of vectors $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$?

Multiple Choice:

1. *The vectors are linearly independent*
2. *The vectors are linearly dependent*
3. *There is not enough information given to make a determination*

Problem 144 *Given that*

$$2\mathbf{v}_1 + 0\mathbf{v}_2 + 0\mathbf{v}_3 = \mathbf{0}$$

what (if anything) can we conclude about linear independence of vectors $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$*?*

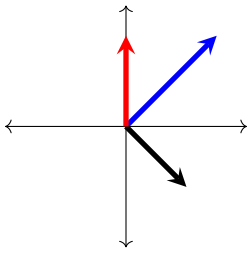
Multiple Choice:

1. *The vectors are linearly independent*
2. *The vectors are linearly dependent*
3. *There is not enough information given to make a determination*

Problem 145 *Are the vectors in the diagram linearly dependent or independent?*

Multiple Choice:

1. *The vectors are linearly independent*
2. *The vectors are linearly dependent*
3. *There is not enough information given to make a determination*

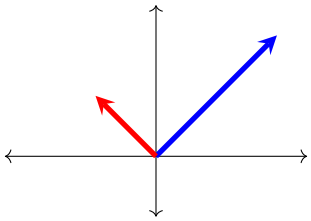


Problem 146 *Are the vectors in the diagram linearly dependent or independent?*

Multiple Choice:

1. *The vectors are linearly independent*

2. *The vectors are linearly dependent*
3. *There is not enough information given to make a determination*



Problem 147 *Suppose* $\{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$ *is a linearly independent set in* $\mathbb{R}^n$*, and that* $\mathbf{v}_{m+1}$ *is not in span* $(\mathbf{v}_1, \ldots, \mathbf{v}_m)$*. Prove that* $\{\mathbf{v}_1, \ldots, \mathbf{v}_m, \mathbf{v}_{m+1}\}$ *is also linearly independent.*

Problem 148 *Suppose* $\{\mathbf{u}, \mathbf{v}\}$ *is a linearly independent set of vectors. Prove that the set* $\{\mathbf{u} - \mathbf{v}, \mathbf{u} + 2\mathbf{v}\}$ *is also linearly independent.*

Problem 149 *Suppose* $\{\mathbf{u}, \mathbf{v}\}$ *is a linearly independent set of vectors in* $\mathbb{R}^3$*. Is the following set dependent or independent* $\{\mathbf{u} - \mathbf{v}, \mathbf{u} + 2\mathbf{v}, \mathbf{u} + \mathbf{v}\}$*? Prove your claim.*

Text Source

A portion of this section was adapted from Section 5.2 of Keith Nicholson's *Linear Algebra with Applications*. (CC-BY-NC-SA)

W. Keith Nicholson, *Linear Algebra with Applications*, Lyryx 2018, Open Edition, p. 271.

24 Matrices

25 Addition and Scalar Multiplication of Matrices

Introduction to Matrices

A *matrix* is a rectangular array of numbers. The plural form of *matrix* is *matrices*. You have encountered matrices before in the context of augmented matrices and coefficient matrices associate with linear systems.

Consider the matrix

$$M = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 2 & 8 & 7 \\ 6 & -9 & 1 & 2 \end{bmatrix}$$

The *dimension of a matrix* is defined as $m \times n$ where m is the number of rows and n is the number of columns. The above matrix is a 3×4 matrix because there are three rows and four columns.

A *column vector* in $\mathbb{R}^n$ is an $n \times 1$ matrix. A *row vector* in $\mathbb{R}^n$ is a $1 \times n$ matrix.

The individual *entries* in the matrix are identified according to their position. The (i, j)-entry of a matrix is the entry in the i^{th} row and j^{th} column. For example, in matrix M above, 8 is called the $(2, 3)$-entry because it is in the second row and the third column.

We denote the entry in the i^{th} row and the j^{th} column of matrix A by a_{ij}, and write A in terms of its entries as

$$A = \begin{bmatrix} a_{ij} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1j} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2j} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots & & \vdots \\ a_{i1} & a_{i2} & \dots & a_{ij} & \dots & a_{in} \\ \vdots & \vdots & & \vdots & & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mj} & \dots & a_{mn} \end{bmatrix}$$

.

Occasionally it will be convenient to talk about columns and rows of a matrix A as vectors. We

will use the following notation:

$$A = \begin{bmatrix} | & | & & | \\ \mathbf{c}_1 & \mathbf{c}_2 & \dots & \mathbf{c}_n \\ | & | & & | \end{bmatrix} \quad \text{or} \quad A = \begin{bmatrix} \mathbf{c}_1 & \mathbf{c}_2 & \dots & \mathbf{c}_n \end{bmatrix}$$

$$A = \begin{bmatrix} - & \mathbf{r}_1 & - \\ - & \mathbf{r}_2 & - \\ & \vdots & \\ - & \mathbf{r}_m & - \end{bmatrix} \quad \text{or} \quad A = \begin{bmatrix} \mathbf{r}_1 \\ \mathbf{r}_2 \\ \vdots \\ \mathbf{r}_m \end{bmatrix}$$

A matrix is called a *square matrix* if it has the same number of rows and columns. If $B = \begin{bmatrix} b_{ij} \end{bmatrix}$ is an $n \times n$ square matrix, the entries of the form b_{ii} are said to lie on the *main diagonal*. For example, if

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

then the main diagonal consists of entries $b_{11} = 1$, $b_{22} = 5$ and $b_{33} = 9$.

There are various operations which are done on matrices of appropriate sizes. Matrices can be added to and subtracted from other matrices, multiplied by a scalar, and multiplied by other matrices. We will never divide a matrix by another matrix, but we will see later how multiplication by a matrix inverse (if an inverse exists) plays a similar role to division.

In doing arithmetic with matrices, we often define the action by what happens in terms of the entries (or components) of the matrices. Before looking at these operations in depth, consider a few general definitions.

Definition 27 (The Zero Matrix)**.** The $m \times n$ *zero matrix* is the $m \times n$ matrix having every entry equal to zero. The zero matrix is denoted by O.

Definition 28 (Equality of Matrices)**.** Let $A = \begin{bmatrix} a_{ij} \end{bmatrix}$ and $B = \begin{bmatrix} b_{ij} \end{bmatrix}$ be two $m \times n$ matrices. Then $A = B$ means that $a_{ij} = b_{ij}$ for all $1 \leq i \leq m$ and $1 \leq j \leq n$.

Addition of Matrices

Given two matrices of the same dimensions, we can add them together by adding their corresponding entries.

Definition 29 (Addition of Matrices)**.** Let $A = \left[a_{ij}\right]$ and $B = \left[b_{ij}\right]$ be two $m \times n$ matrices. Then the *sum of matrices* A and B, denoted by $A + B$, is an $m \times n$ matrix given by

$$A + B = \left[a_{ij} + b_{ij}\right]$$

Example 42. Find the sum of A and B, if possible.

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 0 & 4 \end{bmatrix}, B = \begin{bmatrix} 5 & 2 & 3 \\ -6 & 2 & 1 \end{bmatrix}$$

Explanation. Notice that both A and B are of size 2×3. Since A and B are of the same size, addition is possible.

$$\begin{aligned} A + B &= \begin{bmatrix} 1 & 2 & 3 \\ 1 & 0 & 4 \end{bmatrix} + \begin{bmatrix} 5 & 2 & 3 \\ -6 & 2 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 1+5 & 2+2 & 3+3 \\ 1+-6 & 0+2 & 4+1 \end{bmatrix} \\ &= \begin{bmatrix} 6 & 4 & 6 \\ -5 & 2 & 5 \end{bmatrix} \end{aligned}$$

Going forward, whenever we write $A + B$ it will be assumed that the two matrices are of equal size and addition is possible.

Theorem 15 (Properties of Matrix Addition)**.** Let A, B and C be matrices. Then, the following properties hold.

1. Commutative Law of Addition

$$A + B = B + A$$

2. Associative Law of Addition

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

3. Additive Identity

There exists a zero matrix such that

$$A + O = A$$

4. Additive Inverse

There exists a matrix, $-A$, such that

$$A + (-A) = O$$

We will prove Properties **??** and **??**. The remaining properties are left as exercises.

Proof of Property ??: The (i, j)-entry of $A + B$ is given by

$$a_{ij} + b_{ij}$$

The (i, j)-entry of $B + A$ is given by

$$b_{ij} + a_{ij}$$

Since $a_{ij} + b_{ij} = b_{ij} + a_{ij}$, for all i, j, we conclude that $A + B = B + A$. ■

Proof of Property ??: Let $-A$ be defined by

$$-A = \left[-a_{ij}\right]$$

Then $A + (-A) = O$. ■

You will recognize the zero matrix of Theorem **????** as the zero matrix of Definition **??**.

Scalar Multiplication of Matrices

When a matrix is multiplied by a scalar, the new matrix is obtained by multiplying every entry of the original matrix by the given scalar.

Definition 30 (Scalar Multiplication of Matrices)**.** If $A = \left[a_{ij}\right]$ and k is a scalar, then $kA = \left[ka_{ij}\right]$.

Example 43. Find $7A$ if

$$A = \begin{bmatrix} 2 & 0 \\ 1 & -4 \end{bmatrix}$$

Explanation. By Definition **??**, we multiply each entry of A by 7. Therefore,

$$7A = 7\begin{bmatrix} 2 & 0 \\ 1 & -4 \end{bmatrix} = \begin{bmatrix} 7(2) & 7(0) \\ 7(1) & 7(-4) \end{bmatrix} = \begin{bmatrix} 14 & 0 \\ 7 & -28 \end{bmatrix}$$

Theorem 16 (Properties of Scalar Multiplication)**.** Let A, B be matrices, and k, p be scalars. Then, the following properties properties of scalar multiplication hold.

1. Distributive Law over Matrix Addition

$$k(A + B) = kA + kB$$

2. Distributive Law over Scalar Addition

$$(k + p)A = kA + pA$$

3. Associative Law for Scalar Multiplication

$$k(pA) = (kp)A$$

4. Multiplication by 1

$$1A = A$$

The proof of this theorem is similar to the proof of Theorem **??** and is left as an exercise.

Practice Problems

Problem 150 *If*

$$A = \begin{bmatrix} 2 & -1 \\ 3 & 4 \end{bmatrix} \quad \textit{and} \quad B = \begin{bmatrix} 3 & 0 \\ -2 & 1 \end{bmatrix}$$

then

$$2(A + B) = \begin{bmatrix} \ldots & \ldots \\ \ldots & \ldots \end{bmatrix}$$

$$3A - 2B = \begin{bmatrix} \ldots & \ldots \\ \ldots & \ldots \end{bmatrix}$$

Text Source

The text in this section is an adaptation of Section 2.1 of Ken Kuttler's *A First Course in Linear Algebra*. (CC-BY)

Ken Kuttler, *A First Course in Linear Algebra*, Lyryx 2017, Open Edition, p. 53-58.

26 Matrix Multiplication

We will introduce matrix multiplication by first considering the special case of a matrix-vector product. In other words, the first matrix is $m \times n$ and the second matrix is $n \times 1$ for some positive integers m, n.

Matrix-Vector Multiplication

Example 44. Let

$$A = \begin{bmatrix} 2 & -1 & 3 & 2 \\ 0 & 3 & -2 & 1 \\ -2 & 4 & 1 & 0 \end{bmatrix} \quad \text{and} \quad \mathbf{x} = \begin{bmatrix} 3 \\ -1 \\ 4 \\ 1 \end{bmatrix}$$

One way to understand the matrix-vector product $A\mathbf{x}$ is by thinking of it as a linear combination of the columns of A, using the entries in $\mathbf{x}$ as our coefficients.

$$A\mathbf{x} = \begin{bmatrix} 2 & -1 & 3 & 2 \\ 0 & 3 & -2 & 1 \\ -2 & 4 & 1 & 0 \end{bmatrix} \begin{bmatrix} 3 \\ -1 \\ 4 \\ 1 \end{bmatrix} = 3 \begin{bmatrix} 2 \\ 0 \\ -2 \end{bmatrix} + (-1) \begin{bmatrix} -1 \\ 3 \\ 4 \end{bmatrix} + 4 \begin{bmatrix} 3 \\ -2 \\ 1 \end{bmatrix} + \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 21 \\ -10 \\ -6 \end{bmatrix}$$

We can also compute the product one entry at a time. First, let's focus on the first row of A.

$$\begin{bmatrix} 2 & -1 & 3 & 2 \\ 0 & 3 & -2 & 1 \\ -2 & 4 & 1 & 0 \end{bmatrix} \begin{bmatrix} 3 \\ -1 \\ 4 \\ 1 \end{bmatrix} = \begin{bmatrix} (2)(3) + (-1)(-1) + (3)(4) + (2)(1) \\ \\ \end{bmatrix} = \begin{bmatrix} 21 \\ \\ \end{bmatrix}$$

Next, let's look a the second row of A.

$$\begin{bmatrix} 2 & -1 & 3 & 2 \\ 0 & 3 & -2 & 1 \\ -2 & 4 & 1 & 0 \end{bmatrix} \begin{bmatrix} 3 \\ -1 \\ 4 \\ 1 \end{bmatrix} = \begin{bmatrix} 21 \\ (0)(3) + (3)(-1) + (-2)(4) + (1)(1) \\ \end{bmatrix} = \begin{bmatrix} 21 \\ -10 \\ \end{bmatrix}$$

Finally, let's do the third row of A.

$$\begin{bmatrix} 2 & -1 & 3 & 2 \\ 0 & 3 & -2 & 1 \\ -2 & 4 & 1 & 0 \end{bmatrix} \begin{bmatrix} 3 \\ -1 \\ 4 \\ 1 \end{bmatrix} = \begin{bmatrix} 21 \\ -10 \\ (-2)(3) + (4)(-1) + (1)(4) + (0)(1) \end{bmatrix} = \begin{bmatrix} 21 \\ -10 \\ -6 \end{bmatrix}$$

Definition 31. Let A be an $m \times n$ matrix, and let $\mathbf{x}$ be an $n \times 1$ vector. The product $A\mathbf{x}$ is the $m \times 1$ vector given by:

$$A\mathbf{x} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & \dots & \dots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = x_1 \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{bmatrix} + x_2 \begin{bmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{bmatrix} + \cdots + x_n \begin{bmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{bmatrix}$$

or, equivalently,

$$A\mathbf{x} = \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n \end{bmatrix}$$

We can now make a couple of observations about the matrix-vector product. The first observation is part of the definition, but it is still worth pointing out.

Observation 3. In order for the product $A\mathbf{x}$ to exist, A and $\mathbf{x}$ must have compatible dimensions. In particular, vector $\mathbf{x}$ must have as many components as the number of columns of A. (Otherwise, we would not be have a well-defined linear combination of the columns.) So, if A is an $m \times n$ matrix, $\mathbf{x}$ must be an $n \times 1$ vector. If we write these dimensions next to each other, we will notice that the inner dimensions (n) must match, while the outer dimensions, m and 1, give us the dimensions of the product.

$$\begin{array}{ccccc} A & & \mathbf{x} & = & A\mathbf{x} \\ (m \times n) & & (n \times 1) & & (m \times 1) \end{array}$$

Observation 4. If you are familiar with the *dot product* (see Dot Product and its Properties), you may have noticed that each individual entry in the product matrix $A\mathbf{x}$ is the dot product of a row of A with $\mathbf{x}$. Thus, if the rows of A are vectors $\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_n$ we can restate Definition **??** as follows:

$$A\mathbf{x} = \begin{bmatrix} - & \mathbf{r}_1 & - \\ - & \mathbf{r}_2 & - \\ & \vdots & \\ - & \mathbf{r}_n & - \end{bmatrix} \mathbf{x} = \begin{bmatrix} \mathbf{r}_1 \cdot \mathbf{x} \\ \mathbf{r}_2 \cdot \mathbf{x} \\ \vdots \\ \mathbf{r}_n \cdot \mathbf{x} \end{bmatrix}$$

Let's find another matrix-vector product.

Example 45. Let

$$A = \begin{bmatrix} 1 & -1 \\ 2 & 3 \\ -2 & 1 \\ 4 & 0 \end{bmatrix} \quad \text{and} \quad \mathbf{x} = \begin{bmatrix} -3 \\ 5 \end{bmatrix}$$

Find $A\mathbf{x}$.

Explanation. Fill in the blanks below.

$$A\mathbf{x} = \begin{bmatrix} 1 & -1 \\ 2 & 3 \\ -2 & 1 \\ 4 & 0 \end{bmatrix} \begin{bmatrix} -3 \\ 5 \end{bmatrix} = \begin{bmatrix} \dots \dots \\ \dots \dots \\ \dots \dots \\ \dots \dots \end{bmatrix}$$

Matrix-Matrix Multiplication

Matrix-matrix multiplication is simply an extension of the idea of matrix-vector multiplication. In order for the product definition to work, matrix dimensions must be compatible. Let A be an $m \times n$ matrix, and let B be an $n \times p$ matrix, then the product AB will be an $m \times p$ matrix.

$$\begin{array}{ccccc} A & & B & = & AB \\ (m \times n) & & (n \times p) & & (m \times p) \end{array}$$

Just like with vector products, the inner dimensions must be the same, while the outer dimensions, m and p, give us the dimensions of the product.

Definition 32. Let A be an $m \times n$ matrix whose rows are vectors $\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_n$. Let B be an $n \times p$ matrix with columns $\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_p$. Then the entries of the matrix product AB are given by the dot

products

$$AB = \begin{bmatrix} - & \mathbf{r}_1 & - \\ - & \mathbf{r}_2 & - \\ & \vdots & \\ - & \mathbf{r}_i & - \\ & \vdots & \\ - & \mathbf{r}_m & - \end{bmatrix} \begin{bmatrix} | & | & & | & & | \\ \mathbf{b}_1 & \mathbf{b}_2 & \cdots & \mathbf{b}_j & \cdots & \mathbf{b}_p \\ | & | & & | & & | \end{bmatrix}$$

$$= \begin{bmatrix} \mathbf{r}_1 \cdot \mathbf{b}_1 & \mathbf{r}_1 \cdot \mathbf{b}_2 & \cdots & \mathbf{r}_1 \cdot \mathbf{b}_j & \cdots & \mathbf{r}_1 \cdot \mathbf{b}_p \\ \mathbf{r}_2 \cdot \mathbf{b}_1 & \mathbf{r}_2 \cdot \mathbf{b}_2 & \cdots & \mathbf{r}_2 \cdot \mathbf{b}_j & \cdots & \mathbf{r}_2 \cdot \mathbf{b}_p \\ \vdots & \vdots & & \vdots & & \vdots \\ \mathbf{r}_i \cdot \mathbf{b}_1 & \mathbf{r}_i \cdot \mathbf{b}_2 & \cdots & \mathbf{r}_i \cdot \mathbf{b}_j & \cdots & \mathbf{r}_i \cdot \mathbf{b}_p \\ \vdots & \vdots & & \vdots & & \vdots \\ \mathbf{r}_m \cdot \mathbf{b}_1 & \mathbf{r}_m \cdot \mathbf{b}_2 & \cdots & \mathbf{r}_m \cdot \mathbf{b}_j & \cdots & \mathbf{r}_m \cdot \mathbf{b}_p \end{bmatrix}$$

So the (i, j)-entry of AB is the dot product of the i^{th} row of A and the j^{th} column of B.
In terms of components, if the i^{th} row of A is

$$\begin{bmatrix} a_{i1} & a_{i2} & \cdots & a_{in} \end{bmatrix}$$

and the j^{th} column of B is

$$\begin{bmatrix} b_{1j} \\ b_{2j} \\ \vdots \\ b_{nj} \end{bmatrix}$$

then the (i, j)-entry of AB is given by

$$a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{in}b_{nj} = \sum_{k=1}^{n} a_{ik}b_{kj} \tag{33}$$

Example 46. Let

$$A = \begin{bmatrix} 3 & 2 & -1 & 1 \\ 0 & 3 & 1 & 1 \\ 1 & 4 & -1 & 0 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 1 & -2 \\ -1 & 4 \\ 2 & 3 \\ 2 & 0 \end{bmatrix}$$

Find AB.

Explanation. Observe that A is a 3×4 matrix and B is a 4×2 matrix. So, we expect the product AB to be a 3×2 matrix.

$$\begin{aligned} AB &= \begin{bmatrix} 3 & 2 & -1 & 1 \\ 0 & 3 & 1 & 1 \\ 1 & 4 & -1 & 0 \end{bmatrix} \begin{bmatrix} 1 & -2 \\ -1 & 4 \\ 2 & 3 \\ 2 & 0 \end{bmatrix} \\ &= \begin{bmatrix} (3)(1)+(2)(-1)+(-1)(2)+(1)(2) & (3)(-2)+(2)(4)+(-1)(3)+(1)(0) \\ (0)(1)+(3)(-1)+(1)(2)+(1)(2) & (0)(-2)+(3)(4)+(1)(3)+(1)(0) \\ (1)(1)+(4)(-1)+(-1)(2)+(0)(2) & (1)(-2)+(4)(4)+(-1)(3)+(0)(0) \end{bmatrix} \\ &= \begin{bmatrix} 1 & -1 \\ 1 & 15 \\ -5 & 11 \end{bmatrix} \end{aligned}$$

Remark 3. It is possible to use linear combinations rather than dot products to compute a matrix-matrix product. If the columns of matrix B are given by $\mathbf{b}_1, \mathbf{b}_2, ..., \mathbf{b}_p$, then the matrix product consists of p columns, each of which is a matrix-vector product:

$$AB = \begin{bmatrix} | & | & ... & | \\ A\mathbf{b}_1 & A\mathbf{b}_2 & ... & A\mathbf{b}_p \\ | & | & ... & | \end{bmatrix}.$$

If we return to Example **??** above, we may use this approach to compute

$$A\mathbf{b}_1 = \begin{bmatrix} 3 \\ 0 \\ 1 \end{bmatrix} - \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix} + 2 \begin{bmatrix} -1 \\ 1 \\ -1 \end{bmatrix} + 2 \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ -5 \end{bmatrix}$$

and

$$A\mathbf{b}_2 = -2 \begin{bmatrix} 3 \\ 0 \\ 1 \end{bmatrix} + 4 \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix} + 3 \begin{bmatrix} -1 \\ 1 \\ -1 \end{bmatrix} + 0 = \begin{bmatrix} -1 \\ 15 \\ 11 \end{bmatrix}.$$

Example 47. Let

$$A = \begin{bmatrix} -2 & 1 \\ 3 & 0 \\ 1 & -3 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 4 & -2 & 3 & -2 & 2 \\ 1 & -3 & 1 & 1 & 0 \end{bmatrix}$$

Find AB.

Explanation. Fill in the blanks below.

$$AB = \begin{bmatrix} -2 & 1 \\ 3 & 0 \\ 1 & -3 \end{bmatrix} \begin{bmatrix} 4 & -2 & 3 & -2 & 2 \\ 1 & -3 & 1 & 1 & 0 \end{bmatrix} = \begin{bmatrix} \ldots & \ldots & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots \end{bmatrix}$$

Example 48. Suppose that we know that $\mathbf{x} = \begin{bmatrix} -2 \\ 5 \end{bmatrix}$ satisfies

$$\begin{bmatrix} -2 & 3 \\ 4 & -1 \end{bmatrix} \mathbf{x} = \begin{bmatrix} 19 \\ -13 \end{bmatrix}$$

Use this information to express $\begin{bmatrix} 19 \\ -13 \end{bmatrix}$ as a linear combination of the columns of $\begin{bmatrix} -2 & 3 \\ 4 & -1 \end{bmatrix}$.

Explanation.

$$-2 \begin{bmatrix} -2 \\ 4 \end{bmatrix} + 5 \begin{bmatrix} 3 \\ -1 \end{bmatrix} = \begin{bmatrix} 19 \\ -13 \end{bmatrix}$$

Properties of Matrix Multiplication

Exploration 151 *Let*

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \quad \textit{and} \quad B = \begin{bmatrix} 5 & 6 \\ 7 & 8 \end{bmatrix}$$

Observe that both AB and BA are defined, and both products are 2×2 *matrices. Let's compute the two products*

$$AB = \begin{bmatrix} 19 & 22 \\ 43 & 50 \end{bmatrix} \quad \textit{and} \quad BA = \begin{bmatrix} 23 & 34 \\ 31 & 46 \end{bmatrix}$$

Clearly $AB \neq BA$. *We say that A and B do not commute.*

While it is possible to find specific matrices that commute, matrix multiplication is not commutative in general.

Warning 2. Matrix multiplication is not commutative.

One example of an $n \times n$ square matrix that commutes with all $n \times n$ matrices is the matrix I_n defined by

$$I_n = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix}$$

I_n has 1's along the main diagonal and 0's everywhere else. It is often useful to think of I_n as a matrix whose j^{th} column (and j^{th} row) is $\mathbf{e}_j$, the j^{th} standard unit vector of $\mathbb{R}^n$. When the dimensions of I_n are clear from the context, or irrelevant, we will omit the subscript n and simply refer to this matrix as I.

You can easily convince yourself that I commutes with all square matrices of appropriate dimensions. Let

$$A = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} \quad \text{and} \quad I = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Verify that $AI = A$ and $IA = A$.

Because I acts like the multiplicative identity 1 in regular multiplication, I (or I_n) is called the *identity matrix*.

Next we list several important properties of matrix multiplication. These properties hold only when matrix sizes are such that the products are defined.

Theorem 17 (Properties of Matrix Multiplication)**.** The following hold for matrices A, B, and C and for scalar c,

1. Left Distributive Property
$$A(B + C) = AB + AC$$
2. Right Distributive Property
$$(B + C)A = BA + CA$$
3. Associativity
$$A(BC) = (AB)C$$
4.
$$c(AB) = (cA)B = A(cB)$$
5. Multiplicative Identity
$$AI = IA = A$$

Proof We prove **??** using the expression in (**??**) for the (i, j)-entry of a matrix product. (The proof of **??** is similar.) The (i, j)-entry of $A(B + C)$ is given by

$$\sum_k a_{ik}(b_{kj} + c_{kj}) = \sum_k a_{ik}b_{kj} + \sum_k a_{ik}c_{kj}$$

We recognize the right hand side as the (i, j)-entry of $AB + AC$. Thus $A(B + C) = AB + AC$.

For **??**,

$$c\left(\sum_k a_{ik}b_{kj}\right) = \sum_k (ca_{ik})\, b_{kj} = \sum_k a_{ik}\left(b_{kj}c\right).$$

For **??**, the (i, j)-entry of the product AI is given by the dot product of the i^{th} row of A with the standard unit vector $\mathbf{e}_j$. Clearly, this dot product is a_{ij}. Because the (i, j)-entry of the product AI is equal to the (i, j)-entry A, we conclude that $AI = A$. The proof that $IA = A$ is similar.

Note that we skipped the proof of **??**, which is quite cumbersome using sigma notation. We will easily tackle this proof later in the course when we cover Composition of Linear Transformations. ■

Practice Problems

Problem 152 *Explain why the following product is not defined.*

$$\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}$$

Problem 153 *Express the given product as a linear combination of the columns of the matrix.*

$$\begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \end{bmatrix} \begin{bmatrix} -2 \\ 3 \\ -5 \\ 7 \end{bmatrix} = \ldots \ldots \begin{bmatrix} 1 \\ 5 \end{bmatrix} + \ldots \ldots \begin{bmatrix} 2 \\ 6 \end{bmatrix} + \ldots \ldots \begin{bmatrix} 3 \\ 7 \end{bmatrix} + \ldots \ldots \begin{bmatrix} 4 \\ 8 \end{bmatrix}$$

Problem 154 *Find the dimensions of the product.*

$$\begin{bmatrix} 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix}$$

Dimensions of product: ... ... $\times$ ... ...

Problem 155 *Find the dimensions of the product.*

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \end{bmatrix}$$

Dimensions of product: ... ... $\times$ ... ...

Problem 156 *Find the product.*

$$\begin{bmatrix} 1 & 3 & -2 & 1 \\ -2 & 1 & 0 & 4 \end{bmatrix} \begin{bmatrix} 4 \\ -2 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} \ldots \ldots \\ \ldots \ldots \end{bmatrix}$$

Problem 157 *Find the product.*

$$\begin{bmatrix} 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} -1 & 2 & -3 & 1 \\ 1 & 1 & -1 & -2 \\ 0 & 1 & 2 & 1 \end{bmatrix} = \begin{bmatrix} \ldots \ldots & \ldots \ldots & \ldots \ldots & \ldots \ldots \end{bmatrix}$$

Problem 158 *Find the product.*

$$\begin{bmatrix} 1 & 2 \\ -1 & 0 \\ 2 & 3 \\ -4 & -1 \end{bmatrix} \begin{bmatrix} -1 & 1 \\ 2 & -3 \end{bmatrix} = \begin{bmatrix} \ldots \ldots & \ldots \ldots \\ \ldots \ldots & \ldots \ldots \\ \ldots \ldots & \ldots \ldots \\ \ldots \ldots & \ldots \ldots \end{bmatrix}$$

Text Source

The section on Properties of Matrix Multiplication is an adaptation of Section 2.1 of Ken Kuttler's *A First Course in Linear Algebra*. (CC-BY)

Ken Kuttler, *A First Course in Linear Algebra*, Lyryx 2017, Open Edition, p. 67-68.

27 Block Matrix Multiplication

It is often useful to partition a matrix into smaller matrices, called *blocks*. A matrix viewed in this way is said to be *partitioned into blocks*. For example, each column of a matrix can be considered to be a block. Writing a matrix B in the form

$$B = \begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \dots & \mathbf{b}_k \end{bmatrix} \text{ where the } \mathbf{b}_j \text{ are the columns of } B$$

is a block partition of B.

Here is another example. Consider matrix A.

$$A = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 2 & -1 & 4 & 2 & 1 \\ 3 & 1 & -1 & 7 & 5 \end{bmatrix}$$

There is a natural way to partition A. Observe the 2×2 identity matrix, I_2, in the upper left corner. There is also a 2×3 zero matrix, O_{23}, in the upper right corner. We will take advantage of these features to partition A as follows.

$$A = \left[\begin{array}{cc|ccc} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ \hline 2 & -1 & 4 & 2 & 1 \\ 3 & 1 & -1 & 7 & 5 \end{array}\right] = \begin{bmatrix} I_2 & O_{23} \\ P & Q \end{bmatrix}$$

This notation is particularly useful when we are multiplying two matrices because the product can be computed in block form.

Consider matrix B.

$$B = \begin{bmatrix} 4 & -2 \\ 5 & 6 \\ 7 & 3 \\ -1 & 0 \\ 1 & 6 \end{bmatrix}$$

We will compute the product AB by ordinary matrix multiplication, using blocks as entries. The only requirement is that the blocks be compatible. That is, the sizes of the blocks must be such that all matrix products of blocks that occur make sense. This means that the number of columns in each block of A must equal the number of rows in the corresponding block of B.

To find the product AB, we need to partition B so that block I_2 corresponds to a 2×2 block of B, and block O_{23} of A corresponds to a 3×2 block of B. We partition B as follows.

$$B = \left[\begin{array}{cc} 4 & -2 \\ 5 & 6 \\ \hline 7 & 3 \\ -1 & 0 \\ 1 & 6 \end{array}\right] = \begin{bmatrix} X \\ Y \end{bmatrix}$$

Then,

$$AB = \begin{bmatrix} I & O \\ P & Q \end{bmatrix} \begin{bmatrix} X \\ Y \end{bmatrix} = \begin{bmatrix} IX + OY \\ PX + QY \end{bmatrix} = \begin{bmatrix} X \\ PX + QY \end{bmatrix} = \left[\begin{array}{cc} 4 & -2 \\ 5 & 6 \\ \hline 30 & 8 \\ 8 & 27 \end{array}\right]$$

This is easily checked to be the product AB, computed in the conventional manner.

Theorem 18. If matrices A and B are partitioned compatibly into blocks, the product AB can be computed by matrix multiplication using blocks as entries.

We omit the proof.

Block multiplication has theoretical uses, as we shall see later. It is also useful in computing products of matrices when using a computer with limited memory capacity. The matrices are partitioned into blocks in such a way that each product of blocks can be handled. Then the blocks are stored in auxiliary memory and their products are computed one by one.

Practice Problems

Problem 159 *Compute AB, using the indicated block partitioning.*

$$A = \left[\begin{array}{cc|cc} 2 & -1 & 3 & 1 \\ 1 & 0 & 1 & 2 \\ \hline 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right] \quad B = \left[\begin{array}{cc|c} 1 & 2 & 0 \\ -1 & 0 & 0 \\ \hline 0 & 5 & 1 \\ 1 & -1 & 0 \end{array}\right]$$

Answer:

$$AB = \left[\begin{array}{cc|c} \ldots\ldots & \ldots\ldots & \ldots\ldots \\ \ldots\ldots & \ldots\ldots & \ldots\ldots \\ \hline \ldots\ldots & \ldots\ldots & \ldots\ldots \\ \ldots\ldots & \ldots\ldots & \ldots\ldots \end{array}\right]$$

Problem 160 *In each case give formulas for all powers $A, A^2, A^3, \ldots$ of A using the block decomposition indicated.*

1. $A = \left[\begin{array}{c|cc} 1 & 0 & 0 \\ \hline 1 & 1 & -1 \\ 1 & -1 & 1 \end{array}\right]$

2. $A = \left[\begin{array}{cc|cc} 1 & -1 & 2 & -1 \\ 0 & 1 & 0 & 0 \\ \hline 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & 1 \end{array}\right]$

   Click the arrow to see the answer. $A^{2k} = \left[\begin{array}{cc|cc} 1 & -2k & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \hline 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right]$ *for* $k = 0, 1, 2, \ldots$

   $A^{2k+1} = A^{2k}A = \left[\begin{array}{cc|cc} 1 & -(2k+1) & 2 & -1 \\ 0 & 1 & 0 & 0 \\ \hline 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & 1 \end{array}\right]$ *for* $k = 0, 1, 2, \ldots$

Problem 161 *Compute the following using block multiplication (all blocks are $k \times k$).*

1. $\left[\begin{array}{cc} I & X \\ -Y & I \end{array}\right]\left[\begin{array}{cc} I & 0 \\ Y & I \end{array}\right]$

2. $\left[\begin{array}{cc} I & X \\ 0 & -I \end{array}\right]^n$ *any* $n \geq 1$

3. $\left[\begin{array}{cc} 0 & X \\ I & 0 \end{array}\right]^n$ *any* $n \geq 1$

Click the arrow to see the answer. $\begin{bmatrix} X^m & 0 \\ 0 & X^m \end{bmatrix}$ *if* $n = 2m$; $\begin{bmatrix} 0 & X^{m+1} \\ X^m & 0 \end{bmatrix}$ *if* $n = 2m + 1$

Text Source

The text in this section is an adaptation of Section 2.3 of Keith Nicholson's *Linear Algebra with Applications*. (CC-BY-NC-SA)

W. Keith Nicholson, *Linear Algebra with Applications*, Lyryx 2018, Open Edition, p 73-74.

28 Transpose of a Matrix

Another important operation on matrices is that of taking the *transpose*. For a matrix A, we denote the *transpose of* A by A^T. Before formally defining the transpose, we explore this operation on the following matrix.

$$\begin{bmatrix} 1 & 4 \\ 3 & 1 \\ 2 & 6 \end{bmatrix}^T = \begin{bmatrix} 1 & 3 & 2 \\ 4 & 1 & 6 \end{bmatrix}$$

What happened? The first column became the first row and the second column became the second row. Thus the 3×2 matrix became a 2×3 matrix. The number 4 was in the first row and the second column and it ended up in the second row and first column.

The definition of the transpose is as follows.

Definition 33 (The Transpose of a Matrix)**.** Let $A = \begin{bmatrix} a_{ij} \end{bmatrix}$ be an $m \times n$ matrix. Then the *transpose of* A, denoted by A^T, is the $n \times m$ matrix given by

$$A^T = \begin{bmatrix} a_{ij} \end{bmatrix}^T = \begin{bmatrix} a_{ji} \end{bmatrix}$$

The (i, j)-entry of A becomes the (j, i)-entry of A^T.

Example 49. Calculate A^T for the following matrix

$$A = \begin{bmatrix} 1 & 2 & -6 \\ 3 & 5 & 4 \end{bmatrix}$$

Explanation.

$$A^T = \begin{bmatrix} 1 & 3 \\ 2 & 5 \\ -6 & 4 \end{bmatrix}$$

Note that A is a 2×3 matrix, while A^T is a 3×2 matrix. The columns of A are the rows of A^T, and the rows of A are the columns of A^T.

Theorem 19 (Properties of the Transpose of a Matrix)**.** Let A be an $m \times n$ matrix, B an $n \times p$ matrix, and k a scalar. Then

1. $\left(A^T\right)^T = A$

2. $(AB)^T = B^T A^T$ (Shoes and Socks rule)
3. $(A + B)^T = A^T + B^T$
4. $(kA)^T = kA^T$

We will prove Property **??**. The remaining properties are left as exercises.

Proof of Property ??: Note that A and B have compatible dimensions, so that AB is defined and has dimensions $m \times p$. Thus, $(AB)^T$ has dimensions $p \times m$. On the right side of the equality, A^T has dimensions $n \times m$, and B^T has dimensions $p \times n$. Therefore $B^T A^T$ is defined and has dimensions $p \times m$. Now we know that $(AB)^T$ and $B^T A^T$ have the same dimensions.

To show that $(AB)^T = B^T A^T$ we need to show that their corresponding entries are equal. Recall that the (i, j)-entry of AB is given by the dot product of the i^{th} row of A and the j^{th} column of B. The same dot product is also the (j, i)-entry of $(AB)^T$.

The (j, i)-entry of $B^T A^T$ is given by the dot product of the j^{th} row of B^T and the i^{th} column of A^T. But the j^{th} row of B^T is has the same entries as the j^{th} column of B, and the i^{th} column of A^T has the same entries as the i^{th} row of A. Therefore the (j, i)-entry of $B^T A^T$ is also equal to the (i, j)-entry of AB.

Thus, the corresponding components of $(AB)^T$ are equal and we conclude that $(AB)^T = B^T A^T$. ■

Symmetric and Skew Symmetric Matrices

Definition 34 (Symmetric and Skew Symmetric Matrices)**.** An $n \times n$ matrix A is said to be *symmetric* if $A = A^T$. It is said to be *skew symmetric* if $A = -A^T$.

We will explore these definitions in the following examples.

Example 50. Let

$$A = \begin{bmatrix} 2 & 1 & 3 \\ 1 & 5 & -3 \\ 3 & -3 & 7 \end{bmatrix}$$

Show that A is symmetric.

Explanation.

$$A^T = \begin{bmatrix} 2 & 1 & 3 \\ 1 & 5 & -3 \\ 3 & -3 & 7 \end{bmatrix}$$

Hence, $A = A^T$, so A is symmetric.

Example 51. Let

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

Show that A is skew symmetric.

Explanation.

$$A^T = \begin{bmatrix} 0 & -1 & -3 \\ 1 & 0 & -2 \\ 3 & 2 & 0 \end{bmatrix}$$

Each entry of A^T is equal to -1 times the same entry of A. Hence, $A^T = -A$ and so by Definition **??**, A is skew symmetric.

A special case of a symmetric matrix is a *diagonal matrix*. A diagonal matrix is a square matrix whose entries outside of the main diagonal are all zero. The identity matrix I is a diagonal matrix. Here is another example.

$$\begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & -3 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 4 \end{bmatrix}$$

Practice Problems

Problem 162 *Prove Properties* **??**, **??** *and* **??** *of Theorem* **??**.

Problem 163 *Let A be an arbitrary matrix. What can you say about the dimensions of the product* $A^T A$*?*

Problems **??**-**??**

Classify each matrix as symmetric, skew symmetric, or neither.

Problem 164

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

Multiple Choice:

1. *Symmetric*
2. *Skew symmetric*
3. *Neither*

Problem 165

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

Multiple Choice:

1. *Symmetric*
2. *Skew symmetric*
3. *Neither*

Problem 166

$$\begin{bmatrix} 1 & 1 & 3 \\ 1 & -1 & 2 \\ 3 & 2 & 4 \end{bmatrix}$$

Multiple Choice:

1. *Symmetric*
2. *Skew symmetric*
3. *Neither*

Problem 167 *Give your own example of a* 4×4 *skew symmetric matrix.*

Problem 168 *Make a conjecture about the main diagonal entries of a skew symmetric matrix. Prove your conjecture.*

Text Source

The text in this section is an adaptation of Section 2.1 of Ken Kuttler's *A First Course in Linear Algebra*. (CC-BY)

Ken Kuttler, *A First Course in Linear Algebra*, Lyryx 2017, Open Edition, p. 68-70.

29 Linear Systems as Matrix and Linear Combination Equations

A Linear System as a Matrix Equation

Exploration 169 *Consider the linear system*

$$\begin{array}{rcrcrcrcr} 3x_1 & - & 2x_2 & + & 4x_3 & + & x_4 & = & 5 \\ -x_1 & & & + & 5x_3 & - & 2x_4 & = & 1 \\ 2x_1 & + & x_2 & - & x_3 & + & 3x_4 & = & -4 \end{array}$$

Let's construct the coefficient matrix A and multiply it by $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}$ *on the right.*

$$A\mathbf{x} = \begin{bmatrix} 3 & -2 & 4 & 1 \\ -1 & 0 & 5 & -2 \\ 2 & 1 & -1 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 3x_1 - 2x_2 + 4x_3 + x_4 \\ -x_1 + 5x_3 - 2x_4 \\ 2x_1 + x_2 - x_3 + 3x_4 \end{bmatrix}$$

Observe that each component of the product vector corresponds to one of the equations in the system. Let $\mathbf{b} = \begin{bmatrix} 5 \\ 1 \\ -4 \end{bmatrix}$*. Then*

$$A\mathbf{x} = \mathbf{b}$$

$$\begin{bmatrix} 3 & -2 & 4 & 1 \\ -1 & 0 & 5 & -2 \\ 2 & 1 & -1 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 5 \\ 1 \\ -4 \end{bmatrix}$$

is a matrix equation that corresponds to our system of equations.

In general, a system of linear equations

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

can be written as a matrix equation as follows:

$$\begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & \ldots & \ldots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$$

Solving this matrix equation (or showing that a solution does not exist) amounts to finding the reduced row-echelon form of the augmented matrix

$$\left[\begin{array}{cccc|c} a_{11} & a_{12} & \ldots & a_{1n} & b_1 \\ a_{21} & a_{22} & \ldots & a_{2n} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{m1} & \ldots & \ldots & a_{mn} & b_m \end{array}\right]$$

Example 52. Given a linear system

$$\begin{array}{rcrcl} x & + & 2y & = & 0 \\ -x & + & y & = & -3 \\ & & y & = & -1 \\ x & & & = & 2 \end{array}$$

1. Write the system as a matrix equation
2. Solve the system and the matrix equation

Explanation. ?? The matrix equation that corresponds to the system is

$$\begin{bmatrix} \ldots & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \ldots & \ldots \\ \ldots & \ldots \\ \ldots & \ldots \\ \ldots & \ldots \end{bmatrix}$$

?? The augmented matrix that corresponds to the original system and its reduced row-echelon form are

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

This shows that the ordered pair $(2, -1)$ is a solution to the system. We conclude that $\mathbf{x} = \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 2 \\ -1 \end{bmatrix}$ is a solution to the matrix equation in **??**. A quick verification confirms this

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

Example 53. Let

$$A = \begin{bmatrix} 2 & 1 & -1 & 2 \\ 1 & 1 & 0 & 3 \end{bmatrix} \quad \text{and} \quad \mathbf{b} = \begin{bmatrix} 0 \\ -2 \end{bmatrix}$$

Solve $A\mathbf{x} = \mathbf{b}$.

Explanation. We write the equation $A\mathbf{x} = \mathbf{b}$ in augmented matrix form and apply elementary row operations to find its reduced row-echelon form.

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

One way to obtain a solution is to convert this to a system of equations. It is not necessary to write the system down, but it helps to think about it as you write out your solution vector.

$$\begin{aligned} x_1 \quad &- \quad x_3 \quad - \quad x_4 \quad = \quad 2 \\ x_2 \quad &+ \quad x_3 \quad + \quad 4x_4 \quad = \quad -4 \end{aligned}$$

We see that x_1 and x_2 are leading variables because they correspond to leading 1s in the reduced row-echelon form , while x_3 and x_4 are free variables. We start by assigning parameters s and t

to x_3 and x_4, respectively, then solve for x_1 and x_2.

$$\begin{aligned} x_1 &= 2 + s + t \\ x_2 &= -4 - s - 4t \\ x_3 &= s \\ x_4 &= t \end{aligned}$$

We can now write the solution vector as follows

$$\mathbf{x} = \begin{bmatrix} 2+s+t \\ -4-s-4t \\ s \\ t \end{bmatrix} = \begin{bmatrix} 2 \\ -4 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 1 \\ -1 \\ 1 \\ 0 \end{bmatrix} s + \begin{bmatrix} 1 \\ -4 \\ 0 \\ 1 \end{bmatrix} t \tag{34}$$

The solution given in (**??**) is an example of a *general solution* because it accounts for all of the solutions to the system. Letting s and t take on specific values produces *particular solutions*. For example, $\begin{bmatrix} 2 \\ -1 \\ 1 \\ -1 \end{bmatrix}$ is a particular solution that corresponds to $s = 1$, $t = -1$.

Singular and Nonsingular Matrices

Our examples so far involved non-square matrices. Square matrices, however, play a very important role in linear algebra. This section will focus on square matrices.

Example 54. Let

$$A = \begin{bmatrix} 3 & -1 & 1 \\ 0 & 1 & 2 \\ 1 & 2 & 2 \end{bmatrix} \quad \text{and} \quad \mathbf{b} = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}$$

Solve $A\mathbf{x} = \mathbf{b}$.

Explanation. We apply elementary row operations to bring the augmented matrix to its reduced row-echelon form.

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

We can immediately see that the solution vector is

$$\mathbf{x} = \begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix}$$

Observe that the left-hand side of the augmented matrix in Example **??** is the identity matrix I. This means that $\text{rref}(A) = I$.

The elementary row operations that carried A to I were not dependent on the vector $\mathbf{b}$. In fact, the same row reduction process can be applied to the matrix equation $A\mathbf{x} = \mathbf{b}$ for *any* vector $\mathbf{b}$ to obtain a unique solution.

$$\left[\begin{array}{ccc|c} 3 & -1 & 1 & a \\ 0 & 1 & 2 & b \\ 1 & 2 & 2 & c \end{array}\right] \rightsquigarrow \left[\begin{array}{ccc|c} 1 & 0 & 0 & a^* \\ 0 & 1 & 0 & b^* \\ 0 & 0 & 1 & c^* \end{array}\right]$$

$$\mathbf{x} = \begin{bmatrix} a^* \\ b^* \\ c^* \end{bmatrix}$$

Given a matrix A such that $\text{rref}(A) = I$, the system $A\mathbf{x} = \mathbf{b}$ will never be inconsistent because we will never have a row like this: $\left[\begin{array}{cccc|c} 0 & 0 & \ldots & 0 & 1 \end{array}\right]$. Neither will we have infinitely many solutions because there will never be free variables. Matrices such as A deserve special attention.

Definition 35. A square matrix A is said to be *nonsingular* provided that $\text{rref}(A) = I$. Otherwise we say that A is *singular*.

Non-singular matrices have many useful properties.

Theorem 20. The following statements are equivalent for an $n \times n$ matrix A.

1. A is nonsingular
2. $A\mathbf{x} = \mathbf{b}$ has a unique solution for any $\mathbf{b}$ in $\mathbb{R}^n$
3. $A\mathbf{x} = \mathbf{0}$ has only the trivial solution $\mathbf{x} = \mathbf{0}$

We will prove equivalence of the three statements by showing that

??⇒??⇒??⇒??

Proof of ??⇒?? Suppose rref(A) = I. Given any vector $\mathbf{b}$ in $\mathbb{R}^n$, the augmented matrix $[A|\mathbf{b}]$ can be carried to its reduced row-echelon form $[I|\mathbf{b}^*]$. Uniqueness of the reduced row-echelon form guarantees that $\mathbf{b}^*$ is the unique solution of $A\mathbf{x} = \mathbf{b}$. ■

Proof of ??⇒?? Suppose $A\mathbf{x} = \mathbf{b}$ has a unique solution for all vectors $\mathbf{b}$. Then $A\mathbf{x} = \mathbf{0}$ has a unique solution. But $\mathbf{x} = \mathbf{0}$ is always a solution to $A\mathbf{x} = \mathbf{0}$. Therefore $\mathbf{x} = \mathbf{0}$ is the only solution. ■

Proof of ??⇒?? Suppose $A\mathbf{x} = \mathbf{0}$ has only the trivial solution. This means that $x_1 = 0, x_2 = 0, \ldots, x_n = 0$ is the only solution of $A\mathbf{x} = \mathbf{0}$. But then, we know that the augmented matrix $[A|\mathbf{0}]$ can be reduced to $[I|\mathbf{0}]$. The same row operations will carry A to I. ■

Not all square matrices are nonsingular. For example,

$$\text{rref}\left(\begin{bmatrix} 2 & -1 & 1 \\ 1 & 1 & 1 \\ 3 & 0 & 2 \end{bmatrix}\right) = \begin{bmatrix} 1 & 0 & 2/3 \\ 0 & 1 & 1/3 \\ 0 & 0 & 0 \end{bmatrix} \neq I$$

By Theorem **??**, a matrix equation $A\mathbf{x} = \mathbf{b}$ involving a singular matrix A cannot have a unique solution. The following example illustrates the two scenarios that arise when solving equations that involve singular matrices.

Example 55. Let

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

Solve the equation $A\mathbf{x} = \mathbf{b}$ or show that the equation is inconsistent.

1. $$\mathbf{b} = \begin{bmatrix} 4 \\ -1 \\ 3 \end{bmatrix}$$

2. $$\mathbf{b} = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$$

Explanation. ?? Row reduction gives us

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

There are infinitely many solutions.

$$\mathbf{x} = \begin{bmatrix} 1-(2/3)t \\ -2-(1/3)t \\ t \end{bmatrix} = \begin{bmatrix} 1 \\ -2 \\ 0 \end{bmatrix} + \begin{bmatrix} -2/3 \\ -1/3 \\ 1 \end{bmatrix} t$$

?? When the vector **b** is changed, the row operations that take A to its reduced row-echelon form produce a 1 in the last row of the vector on the right, which shows that the system is inconsistent.

$$\left[\begin{array}{ccc|c} 2 & -1 & 1 & 1 \\ 1 & 1 & 1 & -1 \\ 3 & 0 & 2 & 1 \end{array}\right] \rightsquigarrow \left[\begin{array}{ccc|c} 1 & 0 & 2/3 & 0 \\ 0 & 1 & 1/3 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right]$$

A Linear System as a Linear Combination Equation

Recall that the product of a matrix and a vector can be interpreted as a linear combination of the columns of the matrix. For example,

$$\begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = x_1 \begin{bmatrix} 1 \\ 5 \end{bmatrix} + x_2 \begin{bmatrix} 2 \\ 6 \end{bmatrix} + x_3 \begin{bmatrix} 3 \\ 7 \end{bmatrix} + x_4 \begin{bmatrix} 4 \\ 8 \end{bmatrix}$$

Example 56. For each given matrix A and vector **b**, determine whether **b** is a linear combination of the columns of A. If possible, express **b** as a linear combination of the columns of A.

1. $$A = \begin{bmatrix} 3 & 1 & -2 \\ 1 & 0 & 3 \\ -2 & 1 & 1 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} -2 \\ 7 \\ -1 \end{bmatrix}$$

2. $$A = \begin{bmatrix} 1 & -1 & 3 \\ 2 & -1 & 1 \\ 0 & 1 & -5 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 4 \\ -1 \\ 2 \end{bmatrix}$$

Explanation. ?? We are looking for x_1, x_2, x_3 such that

$$x_1 \begin{bmatrix} 3 \\ 1 \\ -2 \end{bmatrix} + x_2 \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} + x_3 \begin{bmatrix} -2 \\ 3 \\ 1 \end{bmatrix} = \begin{bmatrix} -2 \\ 7 \\ -1 \end{bmatrix}$$

Solving this equation amounts to finding $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$ such that $A\mathbf{x} = \mathbf{b}$. The augmented matrix corresponding to this equation, together with its reduced row-echelon form are

$$\left[\begin{array}{ccc|c} 3 & 1 & -2 & -2 \\ 1 & 0 & 3 & 7 \\ -2 & 1 & 1 & -1 \end{array}\right] \rightsquigarrow \left[\begin{array}{ccc|c} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 2 \end{array}\right]$$

So, $\mathbf{x} = \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}$ is a solution to the matrix equation. We conclude that $\mathbf{b}$ is a linear combination of the columns of A, and write

$$\mathbf{b} = \begin{bmatrix} -2 \\ 7 \\ -1 \end{bmatrix} = \begin{bmatrix} 3 \\ 1 \\ -2 \end{bmatrix} - \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} + 2 \begin{bmatrix} -2 \\ 3 \\ 1 \end{bmatrix}$$

?? We begin by attempting to solve the matrix equation $A\mathbf{x} = \mathbf{b}$. The augmented matrix corresponding to this equation, together with its reduced row-echelon form are

$$\left[\begin{array}{ccc|c} 1 & -1 & 3 & 4 \\ 2 & -1 & 1 & -1 \\ 0 & 1 & -5 & 2 \end{array}\right] \rightsquigarrow \left[\begin{array}{ccc|c} 1 & 0 & -2 & 0 \\ 0 & 1 & -5 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right]$$

This shows that this matrix equation has no solutions. We conclude that $\mathbf{b}$ is not a linear combination of the columns of A.

Practice Problems

Problem 170 *Given a system of linear equations, write (a) the corresponding matrix equation, and (b) the corresponding linear combination equation. DO NOT SOLVE.*

$$\begin{array}{rcrcrcr} 3x & - & y & - & 2z & = & 4 \\ -x & & & + & z & = & -1 \\ & & -y & + & 5z & = & 0 \end{array}$$

Problem 171 *Use an augmented matrix and elementary row operations to find coefficients x_1 and x_2 that make the expression true, or demonstrate that such coefficients do not exist.*

$$x_1 \begin{bmatrix} 1 \\ -2 \end{bmatrix} + x_2 \begin{bmatrix} 1 \\ 3 \end{bmatrix} = \begin{bmatrix} 1 \\ 8 \end{bmatrix}$$

If coefficients x_1 and x_2 do not exist, enter NA in each answer box.

$x_1 =$ ... ... , $x_2 =$ ... ...

Problem 172 *Use an augmented matrix and elementary row operations to find coefficients x_1 and x_2 that make the expression true, or demonstrate that such coefficients do not exist.*

$$x_1 \begin{bmatrix} 4 \\ -1 \end{bmatrix} + x_2 \begin{bmatrix} -8 \\ 2 \end{bmatrix} = \begin{bmatrix} 0 \\ 3 \end{bmatrix}$$

If coefficients x_1 and x_2 do not exist, enter NA in each answer box.

$x_1 =$ ... ... , $x_2 =$ ... ...

Problem 173 *Determine whether vector* **b** *is in the span of the given set of vectors.*

$$\mathbf{b} = \begin{bmatrix} 2 \\ 14 \\ 7 \end{bmatrix}$$

$$\left\{ \begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix}, \begin{bmatrix} -3 \\ 4 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ -3 \\ -2 \end{bmatrix} \right\}$$

Multiple Choice:

1. **b** *is in the span of the vectors.*
2. **b** *is NOT in the span of the vectors.*

Problem 174 *Determine whether vector* $\mathbf{b}$ *is in the span of the given set of vectors.*

$$\mathbf{b} = \begin{bmatrix} 5 \\ 2 \\ 4 \end{bmatrix}$$

$$\left\{ \begin{bmatrix} 4 \\ 2 \\ 4 \end{bmatrix}, \begin{bmatrix} 4 \\ 5 \\ 1 \end{bmatrix}, \begin{bmatrix} 3 \\ 2 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} \right\}$$

Multiple Choice:

1. $\mathbf{b}$ *is in the span of the vectors.*
2. $\mathbf{b}$ *is NOT in the span of the vectors.*

Problem 175 *Determine whether vector* $\mathbf{b}$ *is in the span of the given set of vectors.*

$$\mathbf{b} = \begin{bmatrix} 2 \\ 4 \\ -7 \\ -5 \end{bmatrix}$$

$$\left\{ \begin{bmatrix} 1 \\ -1 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 4 \\ 1 \\ -3 \\ 1 \end{bmatrix} \right\}$$

Multiple Choice:

1. $\mathbf{b}$ *is in the span of the vectors.*
2. $\mathbf{b}$ *is NOT in the span of the vectors.*

30 Homogeneous Linear Systems

Definition 36. A system of linear equations is called *homogeneous* if the system can be written in the form

$$\begin{array}{ccccccccc} a_{11}x_1 & + & a_{12}x_2 & + & \ldots & + & a_{1n}x_n & = & 0 \\ a_{21}x_1 & + & a_{22}x_2 & + & \ldots & + & a_{2n}x_n & = & 0 \\ & & & & \vdots & & & & \\ a_{m1}x_1 & + & a_{m2}x_2 & + & \ldots & + & a_{mn}x_n & = & 0 \end{array}$$

A homogeneous linear system is always consistent because $x_1 = 0, x_2 = 0, \ldots, x_n = 0$ is a solution. This solution is called the *trivial solution*. Geometrically, a homogeneous system can be interpreted as a collection of lines or planes (or hyperplanes) passing through the origin. Thus, they will always have the origin in common, but may have other points in common as well.

If A is the coefficient matrix for a homogeneous system, then the system can be written as a matrix equation $A\mathbf{x} = \mathbf{0}$. The augmented matrix that represents the system looks like this

$$\left[\begin{array}{c|c} A & 0 \end{array}\right]$$

As we perform elementary row operations, the entries to the right of the vertical bar remain 0. It is customary to omit writing them down and apply elementary row operations to the coefficient matrix only.

Example 57. Solve the given homogeneous system and interpret your solution geometrically.

$$\begin{array}{ccccccc} 4x & + & 5y & - & z & = & 0 \\ x & - & 4y & - & 2z & = & 0 \\ 3x & - & 6y & - & 4z & = & 0 \end{array}$$

Explanation. We start by rewriting this system as a matrix equation

$$\begin{bmatrix} 4 & 5 & -1 \\ 1 & -4 & -2 \\ 3 & -6 & -4 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \mathbf{0}$$

We will proceed to find the reduced row-echelon form of the matrix as usual, but will omit writing the zeros to the right of the vertical bar.

$$\begin{bmatrix} 4 & 5 & -1 \\ 1 & -4 & -2 \\ 3 & -6 & -4 \end{bmatrix} \rightsquigarrow \begin{bmatrix} 1 & 0 & -2/3 \\ 0 & 1 & 1/3 \\ 0 & 0 & 0 \end{bmatrix}$$

x and y are leading variables, and z is a free variable. We let $z = t$ and solve for x and y.

$$\begin{aligned} x &= \frac{2}{3}t \\ y &= -\frac{1}{3}t \\ z &= t \end{aligned}$$

Each of the equations in the original system represents a plane through the origin in $\mathbb{R}^3$. The system has infinitely many solutions. Geometrically, we can interpret these solutions as points lying on the line shared by the three planes. The above solution is a parametric representation of this line. Use the GeoGebra demo below to take a better look at the system. (RIGHT-CLICK and DRAG to rotate the image.)

Access GeoGebra interactives through the online version of this text at

https://ximera.osu.edu/linearalgebradzv3/LinearAlgebraInteractiveIntro.

General and Particular Solutions

Definition 37. Given any linear system $A\mathbf{x} = \mathbf{b}$, the system $A\mathbf{x} = \mathbf{0}$ is called the *associated homogeneous system*.

It turns out that there is a relationship between solutions of $A\mathbf{x} = \mathbf{b}$ and solutions of the associated homogeneous system.

Exploration 176 *Let*

$$A = \begin{bmatrix} 1 & 2 & 4 \\ 3 & -7 & -1 \\ -1 & 4 & 2 \end{bmatrix} \quad \textit{and} \quad \mathbf{b} = \begin{bmatrix} -2 \\ 7 \\ -4 \end{bmatrix}$$

Consider the matrix equation $A\mathbf{x} = \mathbf{b}$. Row reduction produces the following.

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

We conclude that $\mathbf{x} = \begin{bmatrix} -2t \\ -1 - t \\ t \end{bmatrix}$.

Let's take a more careful look at $\mathbf{x}$*.*

$$\mathbf{x} = \begin{bmatrix} -2t \\ -1-t \\ t \end{bmatrix} = \begin{bmatrix} 0 \\ -1 \\ \dots \dots \end{bmatrix} + \begin{bmatrix} \dots \dots \\ \dots \dots \\ \dots \dots \end{bmatrix} t$$

We now see that the solution vector $\mathbf{x}$ *is made up of two distinct parts:*

- *one specific vector* $\begin{bmatrix} 0 \\ -1 \\ 0 \end{bmatrix}$
- *infinitely many scalar multiples of* $\begin{bmatrix} -2 \\ -1 \\ 1 \end{bmatrix}$*.*

The vector $\begin{bmatrix} 0 \\ -1 \\ 0 \end{bmatrix}$ *is an example of a particular solution. This particular "particular solution" corresponds to* $t = 0$*. We can find any number of particular solutions by letting* t *assume different values. For example, the particular solution that corresponds to* $t = 1$ *is* $\begin{bmatrix} -2 \\ -2 \\ 1 \end{bmatrix}$*. Let* $\mathbf{x}_p$ *be any particular solution of* $A\mathbf{x} = \mathbf{b}$*. It turns out that all vectors of the form*

$$\mathbf{x} = \mathbf{x}_p + \begin{bmatrix} -2 \\ -1 \\ 1 \end{bmatrix} t$$

are solutions of $A\mathbf{x} = \mathbf{b}$*. We can verify this as follows*

$$A\mathbf{x} = A\left(\mathbf{x}_p + \begin{bmatrix} -2 \\ -1 \\ 1 \end{bmatrix} t\right) = A\mathbf{x}_p + \begin{bmatrix} 1 & 2 & 4 \\ 3 & -7 & -1 \\ -1 & 4 & 2 \end{bmatrix} \begin{bmatrix} -2 \\ -1 \\ 1 \end{bmatrix} t = A\mathbf{x}_p + \mathbf{0} = \mathbf{b}$$

This shows that the specific vector $\begin{bmatrix} 0 \\ -1 \\ 0 \end{bmatrix}$ *is not very special, as any solution of* $A\mathbf{x} = \mathbf{b}$ *can be used in its place.*

The vector $\begin{bmatrix} -2 \\ -1 \\ 1 \end{bmatrix}$*, however, is special. Note that*

$$A\begin{bmatrix} -2 \\ -1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 4 \\ 3 & -7 & -1 \\ -1 & 4 & 2 \end{bmatrix}\begin{bmatrix} -2 \\ -1 \\ 1 \end{bmatrix} = \mathbf{0}$$

So $\begin{bmatrix} -2 \\ -1 \\ 1 \end{bmatrix}$ *and all of its scalar multiples are solutions to the associated homogeneous system.*

Observation 5. In Exploration **??** we found that the general solution of the equation $A\mathbf{x} = \mathbf{b}$ has the form:

$$\mathbf{x} = (\text{Any Particular Solution of } A\mathbf{x} = \mathbf{b}) + (\text{General Solution of } A\mathbf{x} = \mathbf{0})$$

It turns out that the general solution of any linear system can be written in this format. Theorem **??** formalizes this result.

Theorem 21. Suppose $\mathbf{x}_p$ is a particular solution of $A\mathbf{x} = \mathbf{b}$.

1. If $\mathbf{x}_h$ is a solution of the associated homogeneous system, then $\mathbf{x}_p + \mathbf{x}_h$ is a solution of $A\mathbf{x} = \mathbf{b}$.
2. If $\mathbf{x}_1$ is a solution of $A\mathbf{x} = \mathbf{b}$, then there exists a solution of the associated homogeneous system, $\mathbf{x}_h$, such that $\mathbf{x}_1 = \mathbf{x}_p + \mathbf{x}_h$.

We will prove part **??**. The proof of part **??** is left to the reader.

Proof of ?? Let $\mathbf{x}_h = \mathbf{x}_1 - \mathbf{x}_p$, then

$$A\mathbf{x}_h = A(\mathbf{x}_1 - \mathbf{x}_p) = A\mathbf{x}_1 - A\mathbf{x}_p = \mathbf{b} - \mathbf{b} = \mathbf{0}$$

and

$$\mathbf{x}_1 = \mathbf{x}_p + \mathbf{x}_h$$

■

Example 58. Let

$$A = \begin{bmatrix} 2 & -4 & -2 \\ 1 & -2 & -1 \end{bmatrix} \quad \text{and} \quad \mathbf{b} = \begin{bmatrix} 8 \\ 4 \end{bmatrix}$$

If possible, find a solution of $A\mathbf{x} = \mathbf{b}$ and express it as a sum of a particular solution and the general solution of the associated homogeneous system. ($\mathbf{x} = \mathbf{x}_p + \mathbf{x}_h$)

Explanation. First, we obtain the reduced row-echelon form

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

So

$$\mathbf{x} = \begin{bmatrix} 4 + 2s + t \\ s \\ t \end{bmatrix} = \begin{bmatrix} 4 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} s + \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} t$$

In this case

$$\mathbf{x}_p = \begin{bmatrix} 4 \\ 0 \\ 0 \end{bmatrix}$$

$$\mathbf{x}_h = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} s + \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} t$$

Practice Problems

Problems **??**-**??**

For each matrix A and vector $\mathbf{b}$ below, find a solution to $A\mathbf{x} = \mathbf{b}$ and express your solution as a sum of a particular solution and a general solution to the associated homogeneous system.

Problem 177

$$A = \begin{bmatrix} 1 & 1 & 3 & 1 \\ 3 & 4 & 2 & 1 \end{bmatrix} \quad \textit{and} \quad \mathbf{b} = \begin{bmatrix} 6 \\ 1 \end{bmatrix}$$

Problem 178

$$A = \begin{bmatrix} 3 & 2 & 1 \\ 1 & -1 & 1 \\ 4 & 1 & 1 \end{bmatrix} \quad \textit{and} \quad \mathbf{b} = \begin{bmatrix} 10 \\ 2 \\ 12 \end{bmatrix}$$

Problem 179 *Prove that a consistent system has infinitely many solutions if and only if the associated homogeneous system has infinitely many solutions.*

Problems **??**-**??**

If possible, construct an example of each of the following. If not possible, explain why.

Problem 180 *An inconsistent system with an associated homogeneous system that has infinitely many solutions.*

Problem 181 *An inconsistent system with an associated homogeneous system that has a unique (trivial) solution.*

Problem 182 *Prove that a linear combination of any number of solutions of a homogeneous equation is a solution of the same equation.*

Problem 183 *Prove Part* **??** *of Theorem* **??***.*

31 The Inverse of a Matrix

Definition and Properties of Matrix Inverses

Consider the equation $2x = 6$. It takes little time to recognize that the solution to this equation is $x = 3$. In fact, the solution is so obvious that we do not think about the algebraic steps necessary to find it. Let's take a look at these steps in detail.

$$2x = 6$$

$$\frac{1}{2} \times (2x) = \frac{1}{2} \times 6$$

$$(\frac{1}{2} \times 2)x = 3$$

$$1x = 3$$

$$x = 3$$

This process utilizes many properties of real-number multiplication. In particular, we make use of existence of *multiplicative inverses*. Every non-zero real number a has a multiplicative inverse $a^{-1} = \frac{1}{a}$ with the property that $\frac{1}{a} \times a = a \times \frac{1}{a} = 1$. We say that 1 is the *multiplicative identity* because $a \times 1 = 1 \times a = a$.

Given a matrix equation $A\mathbf{x} = \mathbf{b}$, we would like to follow a process similar to the one above to solve this matrix equation for $\mathbf{x}$.

Observe that the role of the multiplicative identity for $n \times n$ square matrices is filled by I_n because $AI_n = I_nA = A$. Given an $n \times n$ matrix A, a multiplicative inverse of A would have to be some $n \times n$ matrix B such that

$$BA = AB = I_n$$

Assuming that such an inverse B exists, this is what the process of solving the equation $A\mathbf{x} = \mathbf{b}$ would look like:

$$A\mathbf{x} = \mathbf{b}$$

$$B(A\mathbf{x}) = B\mathbf{b}$$

$$(BA)\mathbf{x} = B\mathbf{b}$$

$$I\mathbf{x} = B\mathbf{b}$$

$$\mathbf{x} = B\mathbf{b}$$

Definition 38. Let A be an $n \times n$ matrix. An $n \times n$ matrix B is called an *inverse* of A if

$$AB = BA = I$$

where I is an $n \times n$ identity matrix. If such an inverse matrix exists, we say that A is *invertible*. If an inverse does not exist, we say that A is not invertible.

It follows directly from the way the definition is stated that if B is an inverse of A, then A is an inverse of B. We say that A and B are inverses of each other.

The following theorem shows that matrix inverses are unique.

Theorem 22. Suppose A is an invertible matrix, and B is an inverse of A. Then B is unique.

Proof Because B is an inverse of A, we have:

$$AB = BA = I$$

Suppose there exists another $n \times n$ matrix C such that

$$AC = CA = I$$

Then

$$C(AB) = CI$$
$$(CA)B = C$$
$$IB = C$$
$$B = C$$

■

Now that we know that a matrix A cannot have more than one inverse, we can safely refer to the inverse of A as A^{-1}.

Example 59. Let

$$A = \begin{bmatrix} 1 & -1 \\ -5 & 2 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} -2/3 & -1/3 \\ -5/3 & -1/3 \end{bmatrix}$$

Verify that A and B are inverses of each other.

Explanation. We will show that $AB = BA = I$.

$$AB = \begin{bmatrix} 1 & -1 \\ -5 & 2 \end{bmatrix} \begin{bmatrix} -2/3 & -1/3 \\ -5/3 & -1/3 \end{bmatrix} = \begin{bmatrix} -2/3+5/3 & -1/3+1/3 \\ 10/3-10/3 & 5/3-2/3 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$BA = \begin{bmatrix} -2/3 & -1/3 \\ -5/3 & -1/3 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ -5 & 2 \end{bmatrix} = \begin{bmatrix} -2/3+5/3 & 2/3-2/3 \\ -5/3+5/3 & 5/3-2/3 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Example 60. Use what we found in Example **??** to solve the matrix equation:

$$\begin{bmatrix} 1 & -1 \\ -5 & 2 \end{bmatrix} \mathbf{x} = \begin{bmatrix} -3 \\ 1 \end{bmatrix}$$

Explanation. We multiply both sides of the equation by the inverse of $\begin{bmatrix} 1 & -1 \\ -5 & 2 \end{bmatrix}$.

$$\begin{bmatrix} -2/3 & -1/3 \\ -5/3 & -1/3 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ -5 & 2 \end{bmatrix} \mathbf{x} = \begin{bmatrix} -2/3 & -1/3 \\ -5/3 & -1/3 \end{bmatrix} \begin{bmatrix} -3 \\ 1 \end{bmatrix}$$

$$\mathbf{x} = \begin{bmatrix} 5/3 \\ 14/3 \end{bmatrix}$$

We now prove several useful properties of matrix inverses.

Theorem 23. The following properties are stated for square matrices of appropriate sizes.

1. $I^{-1} = I$
2. For an invertible matrix A, $(A^{-1})^{-1} = A$.
3. If A and B are invertible matrices, then (AB) is invertible and $(AB)^{-1} = B^{-1}A^{-1}$ (Shoes and socks rule).
4. If A is an invertible matrix and k is a non-zero number, then $(kA)^{-1} = \frac{1}{k}A^{-1}$.
5. If A is an invertible matrix, then A^T is also invertible, and $(A^T)^{-1} = (A^{-1})^T$.

We will prove **??**. The remaining properties are left as exercises.

Proof of Property ??: We will check to see if $B^{-1}A^{-1}$ is the inverse of AB.

$$(B^{-1}A^{-1})(AB) = B^{-1}(A^{-1}A)B = B^{-1}IB = B^{-1}B = I$$

$$(AB)(B^{-1}A^{-1}) = A(BB^{-1})A^{-1} = AIA^{-1} = AA^{-1} = I$$

Thus (AB) is invertible and $(AB)^{-1} = B^{-1}A^{-1}$. ■

Computing the Inverse

We now turn to the question of how to find the inverse of a square matrix, or determine that the inverse does not exist.

Given a square matrix A, we are looking for a square matrix B such that

$$AB = I \quad \text{and} \quad BA = I$$

We will start by attempting to satisfy $AB = I$. Let $\mathbf{v}_1, \ldots, \mathbf{v}_n$ be the columns of B, then

$$A \begin{bmatrix} | & | & & | \\ \mathbf{v}_1 & \mathbf{v}_2 & \ldots & \mathbf{v}_n \\ | & | & & | \end{bmatrix} = \begin{bmatrix} | & | & & | \\ \mathbf{e}_1 & \mathbf{e}_2 & \ldots & \mathbf{e}_n \\ | & | & & | \end{bmatrix}$$

where each $\mathbf{e}_i$ is a standard unit vector of $\mathbb{R}^n$. This gives us a system of equations $A\mathbf{v}_i = \mathbf{e}_i$ for each $1 \leq i \leq n$. If each $A\mathbf{v}_i = \mathbf{e}_i$ has a unique solution, then finding these solutions will give us the columns of the desired matrix B.

First, suppose that $\text{rref}(A) = I$, then we can use elementary row operations to carry each $[A|\mathbf{e}_i]$ to its reduced row-echelon form.

$$[A|\mathbf{e}_i] \rightsquigarrow [I|\mathbf{v}_i]$$

Observe that the row operations that carry A to I will be the same for each $A\mathbf{v}_i = \mathbf{e}_i$. We can, therefore, combine the process of solving n systems of equations into a single process

$$[A|I] \rightsquigarrow [I|B]$$

Each $\mathbf{v}_i$ is a unique solution of $A\mathbf{v}_i = \mathbf{e}_i$, and we conclude that

$$B = \begin{bmatrix} | & | & & | \\ \mathbf{v}_1 & \mathbf{v}_2 & \ldots & \mathbf{v}_n \\ | & | & & | \end{bmatrix}$$

is a solution to $AB = I$.

By Problem **??**, we can reverse the elementary row operations to obtain

$$[I|B] \rightsquigarrow [A|I]$$

But the same row operations would also give us

$$[B|I] \rightsquigarrow [I|A]$$

We conclude that $BA = I$, and $B = A^{-1}$.

Next, suppose that rref(A) $\neq I$. Then rref(A) must contain a row of zeros. Because one of the rows of A was completely wiped out by elementary row operations, one of the rows of A must be a linear combination of the other rows. Suppose row p is a linear combination of the other rows. Then row p can be carried to a row of zeros. But then the system $A\mathbf{v}_p = \mathbf{e}_p$ is inconsistent. This is because $\mathbf{e}_p$ has a 1 as the p^{th} entry and zeros everywhere else. The 1 in the p^{th} spot will not be affected by elementary row operations, and the p^{th} row will eventually look like this

$$[0 \ldots 0|1]$$

This shows that a matrix B such that $AB = I$ does not exist, and A does not have an inverse.

We have just proved the following theorem.

Theorem 24 (Row-reduction Method for Computing the Inverse of a Matrix)**.** Let A be a square matrix. If it is possible to use elementary row operations to carry the augmented matrix $[A|I]$ to $[I|B]$, then $B = A^{-1}$. If such a reduction is not possible, then A does not have an inverse.

Corollary 2. A square matrix A has an inverse if and only if rref(A) = I.

Example 61. Find A^{-1} or demonstrate that A^{-1} does not exist.

$$A = \begin{bmatrix} 1 & -1 & 2 \\ 1 & 1 & 1 \\ 1 & 3 & -1 \end{bmatrix}$$

Explanation. We start with the augmented matrix

$$\left[\begin{array}{ccc|ccc} 1 & -1 & 2 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 3 & -1 & 0 & 0 & 1 \end{array}\right] \begin{array}{l} \xrightarrow{R_2 - R_1} \\ \xrightarrow{R_3 - R_1} \end{array}$$

$$\left[\begin{array}{ccc|ccc} 1 & -1 & 2 & 1 & 0 & 0 \\ 0 & 2 & -1 & -1 & 1 & 0 \\ 0 & 4 & -3 & -1 & 0 & 1 \end{array}\right] \xrightarrow{R_3 - 2R_2}$$

$$\left[\begin{array}{ccc|ccc} 1 & -1 & 2 & 1 & 0 & 0 \\ 0 & 2 & -1 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 & -2 & 1 \end{array}\right] \xrightarrow{(-1)R_3}$$

$$\left[\begin{array}{ccc|ccc} 1 & -1 & 2 & 1 & 0 & 0 \\ 0 & 2 & -1 & -1 & 1 & 0 \\ 0 & 0 & 1 & -1 & 2 & -1 \end{array}\right] \begin{array}{l} \xrightarrow{R_1 - 2R_3} \\ \xrightarrow{R_2 + R_3} \end{array}$$

$$\left[\begin{array}{ccc|ccc} 1 & -1 & 0 & 3 & -4 & 2 \\ 0 & 2 & 0 & -2 & 3 & -1 \\ 0 & 0 & 1 & -1 & 2 & -1 \end{array}\right] \xrightarrow{\frac{1}{2}R_2}$$

$$\left[\begin{array}{ccc|ccc} 1 & -1 & 0 & 3 & -4 & 2 \\ 0 & 1 & 0 & -1 & 3/2 & -1/2 \\ 0 & 0 & 1 & -1 & 2 & -1 \end{array}\right] \xrightarrow{R_1+R_2}$$

$$\left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 2 & -5/2 & 3/2 \\ 0 & 1 & 0 & -1 & 3/2 & -1/2 \\ 0 & 0 & 1 & -1 & 2 & -1 \end{array}\right]$$

We conclude that

$$A^{-1} = \begin{bmatrix} 2 & -5/2 & 3/2 \\ -1 & 3/2 & -1/2 \\ -1 & 2 & -1 \end{bmatrix}$$

Example 62. Find A^{-1} or demonstrate that A is not invertible.

$$A = \begin{bmatrix} 2 & 3 & -1 \\ 0 & 2 & 1 \\ 4 & 4 & -3 \end{bmatrix}$$

Explanation. We start with the augmented matrix

$$\left[\begin{array}{ccc|ccc} 2 & 3 & -1 & 1 & 0 & 0 \\ 0 & 2 & 1 & 0 & 1 & 0 \\ 4 & 4 & -3 & 0 & 0 & 1 \end{array}\right] \xrightarrow{\frac{1}{2}R_1} \xrightarrow{\frac{1}{2}R_2}$$

$$\left[\begin{array}{ccc|ccc} 1 & 3/2 & -1/2 & 1/2 & 0 & 0 \\ 0 & 1 & 1/2 & 0 & 1/2 & 0 \\ 4 & 4 & -3 & 0 & 0 & 1 \end{array}\right] \xrightarrow{R_3-4R_1}$$

$$\left[\begin{array}{ccc|ccc} 1 & 3/2 & -1/2 & 1/2 & 0 & 0 \\ 0 & 1 & 1/2 & 0 & 1/2 & 0 \\ 0 & -2 & -1 & -2 & 0 & 1 \end{array}\right] \xrightarrow{R_3+2R_2}$$

$$\left[\begin{array}{ccc|ccc} 1 & 3/2 & -1/2 & 1/2 & 0 & 0 \\ 0 & 1 & 1/2 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & -2 & 1 & 1 \end{array}\right]$$

At this point we see that the left-hand side cannot be turned into I through elementary row operations. We conclude that A^{-1} does not exist.

Remark 4. Recall that a square matrix whose reduced row-echelon form is the identity matrix is called *nonsingular*. (Definition **??**) According to Corollary **??**, a matrix is invertible if and only if it is nonsingular. For this reason many linear algebra texts use the terms *invertible* and *nonsingular* as synonyms.

Inverse of a 2×2 Matrix

We will conclude this section by discussing the inverse of a nonsingular 2×2 matrix. Let $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ be a nonsingular matrix. We can find A^{-1} by using the row reduction method described above, that is, by computing the reduced row-echelon form of $[A|I]$. Row reduction yields the following:

$$\left[\begin{array}{cc|cc} a & b & 1 & 0 \\ c & d & 0 & 1 \end{array}\right] \rightsquigarrow \left[\begin{array}{cc|cc} 1 & 0 & d/(ad-bc) & -b/(ad-bc) \\ 0 & 1 & -c/(ad-bc) & a/(ad-bc) \end{array}\right]$$

Note that the denominator of each term in the inverse matrix is the same. Factoring it out, gives us the following formula for A^{-1}.

Formula 10.

$$A^{-1} = \frac{1}{ad-bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

Clearly, the expression for A^{-1} is defined, if and only if $ad - bc \neq 0$. So, what happens when $ad - bc = 0$? In Practice Problem **??** you will be asked to fill in the steps of the row reduction procedure that produces this formula, and show that if $ad - bc = 0$ then A does not have an inverse.

Practice Problems

Problem 184 *Verify that the matrix* $\begin{bmatrix} 2 & -1 \\ 3 & -2 \end{bmatrix}$ *is its own inverse.*

Problem 185 *Use the row-reduction method for computing matrix inverses to explain why the given matrix does not have an inverse.*

$$\begin{bmatrix} 1 & 1 & 1 \\ 2 & 2 & 2 \\ 3 & 3 & 3 \end{bmatrix}$$

Problem 186 *Let*

$$A = \begin{bmatrix} 1 & -1 & 2 \\ 2 & 1 & 1 \\ -3 & 1 & 0 \end{bmatrix} \quad \textit{and} \quad B = \begin{bmatrix} -1/12 & 1/6 & -1/4 \\ -1/4 & 1/2 & 1/4 \\ 5/12 & 1/6 & 1/4 \end{bmatrix}$$

Are A and B inverses of each other?

Multiple Choice:

1. *Yes, A and B are inverses.*
2. *No, A and B are not inverses.*

Problem 187 *Use the following invertible matrices to illustrate Properties* **??**, **??**, **??** *and* **??** *of Theorem* **??**.

$$A = \begin{bmatrix} 2 & 0 \\ 4 & -3 \end{bmatrix} \quad \textit{and} \quad B = \begin{bmatrix} 2 & -1 \\ 1 & 5 \end{bmatrix}$$

Problem 188 *In Example* **??** *we found the inverse of*

$$A = \begin{bmatrix} 1 & -1 & 2 \\ 1 & 1 & 1 \\ 1 & 3 & -1 \end{bmatrix}$$

to be

$$A^{-1} = \begin{bmatrix} 2 & -5/2 & 3/2 \\ -1 & 3/2 & -1/2 \\ -1 & 2 & -1 \end{bmatrix}$$

Use A^{-1} to solve the equation

$$A\mathbf{x} = \begin{bmatrix} 3 \\ -2 \\ 4 \end{bmatrix}$$

Answer:

$$\mathbf{x} = \begin{bmatrix} \ldots\ldots \\ \ldots\ldots \\ \ldots\ldots \end{bmatrix}$$

Problem 189 *Find the inverse of each matrix by using the row-reduction procedure.*

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

$$A^{-1} = \begin{bmatrix} \ldots\ldots & \ldots\ldots & \ldots\ldots \\ \ldots\ldots & \ldots\ldots & \ldots\ldots \\ \ldots\ldots & \ldots\ldots & \ldots\ldots \end{bmatrix}$$

Problem 190 *Use Formula* **??** *to find the inverse of*

$$A = \begin{bmatrix} 2 & -3 \\ 1 & 5 \end{bmatrix}$$

Use A^{-1} to solve the equation

$$A\mathbf{x} = \begin{bmatrix} -3 \\ 2 \end{bmatrix}$$

Answer:

$$\mathbf{x} = \begin{bmatrix} \ldots\ldots \\ \ldots\ldots \end{bmatrix}$$

Problem 191 *Find the value of x for which the matrix is not invertible.*

$$\begin{bmatrix} 1 & 2 \\ 4 & x \end{bmatrix}$$

Answer: $x =$ ... ...

Problem 192 *Suppose $AB = AC$ and A is an invertible $n \times n$ matrix. Does it follow that $B = C$? Explain why or why not.*

Problem 193 *Suppose $AB = AC$ and A is a non-invertible $n \times n$ matrix. Does it follow that $B = C$? Explain why or why not.*

Problem 194 *Suppose A and B are invertible $n \times n$ matrices. Does it follow that $(A + B)$ is invertible?*

32 Elementary Matrices

Definition and Examples

Exploration 195 *Consider the matrices*

$$B = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad C = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{bmatrix}$$

The two matrices have something in common. Can you figure out what it is? (The answer will be given later in the problem.)

$$\text{Let} \quad A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$$

Let's compute BA and CA.

$$BA = \begin{bmatrix} \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots \end{bmatrix} \quad \text{and} \quad CA = \begin{bmatrix} \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots \end{bmatrix}$$

Observe that multiplying A by B on the left results in multiplying the second row of A by 5*, while multiplying A by C on the left results in multiplying the third row of A by* -2*.*

Now we need to return to the question of what B and C have in common. Both matrices were obtained from the identity matrix by multiplying one row of the identity by a non-zero constant. Matrices B and C were obtained from I by multiplying one row of I by 5 *and* -2 *respectively. Multiplying A by B (or C) on the left affects A in the same way.*

Matrix A does not have to be a square matrix. Try finding BA' *and* CA' *for*

$$A' = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix}$$

$$BA' = \begin{bmatrix} \ldots & \ldots \\ \ldots & \ldots \\ \ldots & \ldots \end{bmatrix} \quad \text{and} \quad CA' = \begin{bmatrix} \ldots & \ldots \\ \ldots & \ldots \\ \ldots & \ldots \end{bmatrix}$$

Observe that B and C have the same effect on A' as they did on A.

In general, if a square matrix E is obtained from the identity matrix I by multiplying row j of I by a non-zero constant k, then multiplying an appropriately sized matrix A on the left by E results in row j of A being multiplied by k.

Recall that multiplication of a row of a matrix by a non-zero constant is one of three *elementary row operations*. Applying such an elementary row operation to I in order to produce E, results in applying the same elementary row operation to A when A is multiplied by E on the left.

Exploration 196 *Consider the matrices*

$$D = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \textit{and} \quad F = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

As in the previous Exploration, the two matrices have something in common. Both D and F were obtained from the identity matrix by adding a multiple of one row to another row. Can you guess what will happen if we multiply a matrix A by D and F on the left?

$$\textit{Let} \quad A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$$

Let's compute DA and FA.

$$DA = \begin{bmatrix} \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots \end{bmatrix} \quad \textit{and} \quad FA = \begin{bmatrix} \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots \end{bmatrix}$$

As you had probably guessed, multiplication by D resulted in the third row of A being added to the first, and multiplication by F produced a matrix by adding -2 times the first row to the second row of A. The elementary row operations performed on A mimic the elementary row operations performed on I in order to obtain D and F.

In general, if a square matrix E is obtained from the identity matrix I by adding k times row j of I to row i, then multiplying an appropriately sized matrix A on the left by E results in k times row j of A being added to row i of A.

Recall that adding a scalar multiple of one row to another row of a matrix is one of three *elementary row operations*. Applying such an elementary row operation to I in order to produce E, results in applying the same elementary row operation to A when A is multiplied by E on the left.

Exploration 197 *The matrices B, C, D, F above are special because when we multiply them by any appropriately sized matrix A, we are performing row operations on A. Can you construct a matrix G such that GA is the same as A except that its first and third rows are switched?*

$$G = \begin{bmatrix} \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots \end{bmatrix}$$

The matrices B, C, D, F, G of Explorations **??**, **??** and **??** are known as *elementary matrices* because they perform elementary row operations on appropriately sized matrices.

Definition 39. An *elementary matrix* is a square matrix formed by applying a single elementary row operation to the identity matrix.

Suppose A is an $m \times n$ matrix. If E is an $m \times m$ elementary matrix formed by performing a certain row operation on the $m \times m$ identity matrix, then multiplying any matrix A on the left by E is equivalent to performing that same row operation on A. As there are three types of elementary row operations, there are three types of elementary matrices.

Elementary matrices give us a new way of looking at Gauss-Jordan elimination. Suppose it takes j elementary row operations to transform A into R, its reduced row-echelon form. Then we can represent this reduced row-echelon form as

$$R = E_j \cdots E_2 E_1 A$$

where each E_i is the elementary matrix corresponding to the ith row operation performed on A.

Inverses of Elementary Matrices

It is easy to see that any elementary matrix E is invertible, because if E is formed by applying a certain row operation to the identity matrix I, then there is a single row operation that may be applied to E to get I back. For example, in Exploration **??**, F is formed by adding -2 times the first row of the identity to the second row of the identity. It follows that F^{-1} should be the matrix formed by adding 2 times the first row of the identity to the second row of the identity, i.e.

$$F^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

And indeed we can check

$$FF^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

and also $F^{-1}F = I$.

As part of the Practice Problem set you are asked to find the inverse of each of the other elementary matrices in Explorations **??**, **??** and **??**. Once we have accounted for each of the three types of elementary matrices, we will have proven the following theorem.

Theorem 25. Elementary matrices are invertible, and the inverse of an elementary matrix is an elementary matrix.

Proof Suppose E is obtained from I by switching rows i and j. To find the inverse of E, we need to find a matrix F such that $FE = I$. To get from E back to I, rows i and j of E must be switched. This can be accomplished by multiplying E by itself on the left. So, E is its own inverse.

We can use the same line of reasoning to show that the other two types of elementary matrices are also invertible, and their inverses are also elementary matrices. The details are left to the reader. ■

Elementary Matrices and Nonsingular Matrices

Recall that a square matrix A is called *nonsingular* provided that rref(A) = I.

Theorem 26. The following statements are equivalent for an $n \times n$ matrix A.

1. A is nonsingular
2. A is a product of elementary matrices
3. A is invertible

We will prove equivalence of the three statements by showing that

??$\Rightarrow$**??**$\Rightarrow$**??**$\Rightarrow$**??**

Proof of **??**$\Rightarrow$**??** Suppose rref(A) = I. Then A can be carried to the identity by elementary row operations. So, there exist elementary matrices $E_1, E_2, \ldots, E_k$ such that

$$E_k \ldots E_2 E_1 A = I$$

By Theorem **??**, elementary matrices are invertible and their inverses are also elementary matrices. Thus, we can write A as a product of elementary matrices as follows:

$$A = E_1^{-1} E_2^{-1} \ldots E_k^{-1}$$

■

***Proof of* ??⇒??** Suppose $A = E_1 E_2 \ldots E_k$, where $E_1, E_2, \ldots, E_k$ are elementary matrices. In Theorem **??** we proved that $(BC)^{-1} = C^{-1} B^{-1}$. By repeated applications of this theorem we have

$$(E_1 E_2 \ldots E_k)^{-1} = E_k^{-1} \ldots E_2^{-1} E_1^{-1} = A^{-1}$$

We conclude that A is invertible. ■

***Proof of* ??⇒??** See the corollary to Theorem **??**. ■

Practice Problems

Problem 198 *Determine the elementary row operation that results from multiplying some* $3 \times n$ *matrix A by E on the left. Write down* E^{-1} *without going through the row-reduction procedure. (Hint: Think of an elementary row operation that would undo the row operation caused by E.)*

$$E = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Answer:

$$E^{-1} = \begin{bmatrix} \ldots\ldots & \ldots\ldots & \ldots\ldots \\ \ldots\ldots & \ldots\ldots & \ldots\ldots \\ \ldots\ldots & \ldots\ldots & \ldots\ldots \end{bmatrix}$$

Problem 199 *Determine the elementary row operation that results from multiplying some* $3 \times n$ *matrix A by E on the left. Write down* E^{-1} *without going through the row-reduction procedure. (Hint: Think of an elementary row operation that would undo the row operation caused by E.)*

$$E = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 5 \end{bmatrix}$$

Answer:

$$E^{-1} = \begin{bmatrix} \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots \end{bmatrix}$$

Problem 200 *Determine the elementary row operation that results from multiplying some* $3 \times n$ *matrix A by E on the left. Write down* E^{-1} *without going through the row-reduction procedure. (Hint: Think of an elementary row operation that would undo the row operation caused by E.)*

$$E = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 4 & 1 \end{bmatrix}$$

Answer:

$$E^{-1} = \begin{bmatrix} \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots \end{bmatrix}$$

Problem 201 *Express A as a product of elementary matrices.*

$$A = \begin{bmatrix} 1 & 2 & 0 \\ -1 & 0 & 1 \\ 0 & 4 & 0 \end{bmatrix}$$

Is this representation unique? Prove your claim.

Problem 202 *If possible, express*

$$A = \begin{bmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{bmatrix}$$

as a product of elementary matrices.

33 LU Factorization

Exploration 203 *Consider the equation:*

$$\begin{bmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ -2 & 2 & 1 \end{bmatrix} \begin{bmatrix} 2 & -1 & 2 \\ 0 & 4 & -1 \\ 0 & 0 & 3 \end{bmatrix} \mathbf{x} = \begin{bmatrix} 1 \\ 8 \\ 5 \end{bmatrix}$$

This equation is unusual in that it involves two matrices on the left-hand side. If we multiply the matrices together, we get

$$\begin{bmatrix} 2 & -1 & 2 \\ 6 & 1 & 5 \\ -4 & 10 & -3 \end{bmatrix} \mathbf{x} = \begin{bmatrix} 1 \\ 8 \\ 5 \end{bmatrix}$$

Gaussian elimination yields:

$$\left[\begin{array}{ccc|c} 2 & -1 & 2 & 1 \\ 6 & 1 & 5 & 8 \\ -4 & 10 & -3 & 5 \end{array}\right] \rightsquigarrow \left[\begin{array}{ccc|c} 1 & 0 & 0 & 2 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & -1 \end{array}\right]$$

We conclude that

$$\mathbf{x} = \begin{bmatrix} 2 \\ 1 \\ -1 \end{bmatrix}.$$

Now that we know the answer, we will turn to the original question and consider the advantages of the format of the original problem. Observe that the two matrices have a distinct pattern. The matrix on the left has zeros above the main diagonal, while the matrix on the right has zeros below the main diagonal. Matrices that follow such a pattern are called lower triangular and upper triangular matrices, respectively.

Let $\mathbf{y} = \begin{bmatrix} 2 & -1 & 2 \\ 0 & 4 & -1 \\ 0 & 0 & 3 \end{bmatrix} \mathbf{x}$*. Then we can write the original equation as*

$$\begin{bmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ -2 & 2 & 1 \end{bmatrix} \left(\begin{bmatrix} 2 & -1 & 2 \\ 0 & 4 & -1 \\ 0 & 0 & 3 \end{bmatrix} \mathbf{x} \right) = \begin{bmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ -2 & 2 & 1 \end{bmatrix} \mathbf{y} = \begin{bmatrix} 1 \\ 8 \\ 5 \end{bmatrix}$$

The equation

$$\begin{bmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ -2 & 2 & 1 \end{bmatrix} \mathbf{y} = \begin{bmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ -2 & 2 & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 8 \\ 5 \end{bmatrix}$$

corresponds to the system

$$\begin{array}{rrrcl} y_1 & & & = & 1 \\ 3y_1 & +y_2 & & = & 8 \\ -2y_1 & +2y_2 & +y_3 & = & 5 \end{array}$$

This system can be easily solved by substitution, giving us

$$y_1 = 1$$

$$3(1) + y_2 = 8 \rightsquigarrow y_2 = 5$$

$$-2(1) + 2(5) + y_3 = 5 \rightsquigarrow y_3 = -3$$

So,

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 5 \\ -3 \end{bmatrix}$$

Recall that we let $\mathbf{y} = \begin{bmatrix} 2 & -1 & 2 \\ 0 & 4 & -1 \\ 0 & 0 & 3 \end{bmatrix} \mathbf{x}$*. The equation*

$$\begin{bmatrix} 2 & -1 & 2 \\ 0 & 4 & -1 \\ 0 & 0 & 3 \end{bmatrix} \mathbf{x} = \begin{bmatrix} 1 \\ 5 \\ -3 \end{bmatrix}$$

corresponds to the system

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

which can be easily solved by back-substitution:

$$x_3 = \ldots \ldots$$

$$x_2 = \ldots \ldots$$

$$x_1 = \ldots \ldots$$

Thus, $\mathbf{x} = \begin{bmatrix} 2 \\ 1 \\ -1 \end{bmatrix}$*. Observe that this is the same answer that we obtained in the beginning of the problem using Gaussian elimination.*

Given a matrix equation such as

$$\begin{bmatrix} 2 & -1 & 2 \\ 6 & 1 & 5 \\ -4 & 10 & -3 \end{bmatrix} \mathbf{x} = \begin{bmatrix} 1 \\ 8 \\ 5 \end{bmatrix}$$

of Exploration **??**, often in practice we express the matrix on the left as a product of an upper triangular matrix U and a lower triangular matrix L and use substitution to solve the equation instead of using Gaussian elimination to find the solution. In Exploration **??**,

$$\begin{bmatrix} 2 & -1 & 2 \\ 6 & 1 & 5 \\ -4 & 10 & -3 \end{bmatrix} = LU = \begin{bmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ -2 & 2 & 1 \end{bmatrix} \begin{bmatrix} 2 & -1 & 2 \\ 0 & 4 & -1 \\ 0 & 0 & 3 \end{bmatrix}$$

The process of taking a matrix A and expressing it as a product $A = LU$ of a lower triangular matrix L and an upper triangular matrix U is called *LU factorization*. We also adopt the (common) convention that L is a matrix with a diagonal consisting entirely of 1's, which is often called a *lower unit triangular* matrix.

Not every matrix has an LU factorization, but we will see that we can correct for that.

Example 63. Consider the system

$$\begin{array}{rcrcrcl} & & 12x_2 & - & 18x_3 & = & 6 \\ x_1 & + & 2x_2 & - & 3x_3 & = & 1 \\ x_1 & - & 2x_2 & + & x_3 & = & 1 \end{array}.$$

We can express this system as the matrix equation $A\mathbf{x} = \mathbf{b}$, and we get

$$\begin{bmatrix} 0 & 12 & -18 \\ 1 & 2 & -3 \\ 1 & -2 & 1 \end{bmatrix} \mathbf{x} = \begin{bmatrix} 6 \\ 1 \\ 1 \end{bmatrix}.$$

Unfortunately, for reasons we will explain below, it is not possible to find an *LU* factorization of this coefficient matrix. However, if we simply swap the first two equations:

$$\begin{array}{rcrcrcl} x_1 & + & 2x_2 & - & 3x_3 & = & 1 \\ & & 12x_2 & - & 18x_3 & = & 6 \\ x_1 & - & 2x_2 & + & x_3 & = & 1 \end{array},$$

we are able to find an *LU* factorization of the coefficient matrix.

$$\begin{bmatrix} 1 & 2 & -3 \\ 0 & 12 & -18 \\ 1 & -2 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & -1/3 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & -3 \\ 0 & 12 & -18 \\ 0 & 0 & -2 \end{bmatrix}$$

(How we actually come up with an *LU* factorization of *A* will be discussed after this example.)

To understand how an *LU* factorization can be used, it is helpful to think of this system of equations as a matrix equation.

$$A\mathbf{x} = \mathbf{b}$$
$$(LU)\mathbf{x} = \mathbf{b}$$
$$L(U\mathbf{x}) = \mathbf{b}$$

So if we let $\mathbf{y} = U\mathbf{x}$ and also write $\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}$, then we are able to solve for $y_1, y_2,$ and y_3 using *forward-substitution*.

$$L\mathbf{y} = \mathbf{b}$$
$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & -1/3 & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 6 \\ 1 \end{bmatrix}$$

See if you can do it!

$$y_1 = \ldots\ldots, \quad y_2 = \ldots\ldots, \quad y_3 = \ldots\ldots$$

Once we have the values of $\mathbf{y}$, since $\mathbf{y} = U\mathbf{x}$, all that remains is to solve the following matrix equation using back-substitution.

$$U\mathbf{x} = \mathbf{y}$$

$$\begin{bmatrix} 1 & 2 & -3 \\ 0 & 12 & -18 \\ 0 & 0 & -2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 6 \\ 2 \end{bmatrix}$$

By now you are quite used to solving this kind of problem...

$$x_1 = \ldots\ldots, \quad x_2 = \ldots\ldots, \quad x_3 = \ldots\ldots$$

Finding an *LU* factorization by Inspection

Example 64. Find an *LU* factorization of $A = \begin{bmatrix} 1 & 2 & 0 & 2 \\ 1 & 3 & 2 & 1 \\ 2 & 3 & 4 & 0 \end{bmatrix}$.

One way to find the *LU* factorizationis to simply look for it directly. We need

$$\begin{bmatrix} 1 & 2 & 0 & 2 \\ 1 & 3 & 2 & 1 \\ 2 & 3 & 4 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ x & 1 & 0 \\ y & z & 1 \end{bmatrix} \begin{bmatrix} a & d & h & j \\ 0 & b & e & i \\ 0 & 0 & c & f \end{bmatrix}$$

By multiplying the latter matrices we get

$$\begin{bmatrix} a & d & h & j \\ xa & xd+b & xh+e & xj+i \\ ya & yd+zb & yh+ze+c & yj+iz+f \end{bmatrix}$$

and from this, it is a simple exercise to determine each of the unknowns. For example, from the first column, we get $a = 1$ and then $x = 1, y = 2$.

See if you can continue to determine the entire *LU* factorization of *A*.

$$\begin{bmatrix} 1 & 2 & 0 & 2 \\ 1 & 3 & 2 & 1 \\ 2 & 3 & 4 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 2 & \ldots\ldots & 1 \end{bmatrix} \begin{bmatrix} 1 & \ldots\ldots & \ldots\ldots & \ldots\ldots \\ 0 & \ldots\ldots & \ldots\ldots & \ldots\ldots \\ 0 & 0 & \ldots\ldots & \ldots\ldots \end{bmatrix}$$

Finding an *LU* factorization by the Multiplier Method

In the following example we demonstrate a method for computing an *LU* factorization known as the multiplier method.

Example 65. Find an *LU* factorization for

$$\begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \\ -2 & 3 & -2 \end{bmatrix}$$

Explanation. Write the matrix as the following product.

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

In the matrix on the right, we begin with the left row and zero out the entries below the top using the row operation which involves adding a multiple of a row to another row. You do this and also update the matrix on the left so that the product will be unchanged.

Here is the first step. We take -2 times the top row and add to the second. Then take 2 times the top row and add to the second in the matrix on the left. This gives us

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

The next step is to take 2 times the top row and add to the bottom in the matrix on the right. To ensure that the product is unchanged, we place a -2 in the bottom left corner in the matrix on the left. Thus the second step yields

$$\begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ -2 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 0 & -1 & -5 \\ 0 & 7 & 4 \end{bmatrix}$$

For our final step, we take 7 times the middle row on right and add to bottom row. Updating the matrix on the left in the manner we did earlier, we get

$$\begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ -2 & -7 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 0 & -1 & -5 \\ 0 & 0 & -31 \end{bmatrix}$$

At this point, we can stop. We have an LU factorization of the original matrix. We can always multiply the matrices to check, if we wish.

Notice that when we perform the multiplier method, we are making repeated use of a certain type of elementary row operation, namely, we are adding a scalar multiple of one row to a row below it. The reason this helps to create an LU factorization depends upon the fact that the elementary matrices corresponding to such operations are lower triangular. To understand how this works, we begin with a lemma.

Lemma 2. Let L be a lower (upper) triangular matrix $m \times m$ which has ones down the main diagonal. Then L^{-1} also is a lower (upper) triangular matrix which has ones down the main diagonal. In the case that L is of the form

$$L = \begin{bmatrix} 1 & & & \\ a_1 & 1 & & \\ \vdots & & \ddots & \\ a_n & & & 1 \end{bmatrix}, \tag{35}$$

where all entries are zero except for the left column and main diagonal, it is also the case that L^{-1} is obtained from L by simply multiplying each entry below the main diagonal in L by -1. The same is true if the single nonzero column is in another position.

Proof Consider the usual setup for finding the inverse $\left[\begin{array}{c|c} L & I \end{array}\right]$. Each row operation used on L to transform this matrix to reduced row echelon form changes only the entries in I below the main diagonal. Whether we have the special case of L given in **??** where the nonzero nondiagonal entries are in the left column, or if the single nonzero column is in another position, it is clear that multiplication by -1 as described in the lemma gives us L^{-1}. ■

For a simple illustration of the lemma, observe:

$$\left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & a & 1 & 0 & 0 & 1 \end{array}\right] \rightsquigarrow \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & -a & 1 \end{array}\right]$$

Now let A be an $m \times n$ matrix, say

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

and assume A can be row reduced to an upper triangular form using only row operation 3. Thus, in particular, $a_{11} \neq 0$. Multiply on the left by

$$E_1 = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ -\dfrac{a_{21}}{a_{11}} & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ -\dfrac{a_{m1}}{a_{11}} & 0 & \cdots & 1 \end{bmatrix}$$

This is the product of elementary matrices which make modifications in the first column only. It is equivalent to taking $-a_{21}/a_{11}$ times the first row and adding to the second. Then taking $-a_{31}/a_{11}$ times the first row and adding to the third and so forth. The quotients in the first column of the above matrix are the multipliers. Thus the result is of the form

$$E_1 A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a'_{1n} \\ 0 & a'_{22} & \cdots & a'_{2n} \\ \vdots & \vdots & & \vdots \\ 0 & a'_{m2} & \cdots & a'_{mn} \end{bmatrix}$$

By assumption, $a'_{22} \neq 0$ and so it is possible to use this entry to zero out all the entries below it in the matrix on the right by multiplication by a matrix of the form

$$E_2 = \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & E \end{bmatrix}$$

where E is an $(m-1) \times (m-1)$ matrix of the form

$$E = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ -\dfrac{a'_{32}}{a'_{22}} & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ -\dfrac{a'_{m2}}{a'_{22}} & 0 & \cdots & 1 \end{bmatrix}$$

Again, the entries in the first column below the 1 are the multipliers. Continuing this way, zeroing out the entries below the diagonal entries, finally leads us to

$$E_{m-1}E_{m-2}\cdots E_1 A = U$$

where U is upper triangular. Each E_j has all ones down the main diagonal and is lower triangular. Now we can multiply both sides by the inverses of the E_j in the reverse order. This yields

$$A = E_1^{-1}E_2^{-1}\cdots E_{m-1}^{-1}U$$

By Lemma **??**, this implies that the product of those E_j^{-1} is a lower triangular matrix having all ones down the main diagonal.

The above discussion and lemma explain how the multiplier method works. The expressions

$$-\frac{a_{21}}{a_{11}}, -\frac{a_{31}}{a_{11}}, \cdots, -\frac{a_{m1}}{a_{11}},$$

which we obtained in building E_1, and which we denote respectively by $M_{21}, \cdots, M_{m1}$ to save notation, are the multipliers. Therefore, according to the lemma, to find E_1^{-1} we simply write

$$\begin{bmatrix} 1 & 0 & \cdots & 0 \\ -M_{21} & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ -M_{m1} & 0 & \cdots & 1 \end{bmatrix}$$

Similar considerations apply to the other E_j^{-1}. Thus L is a product of the form

$$\begin{bmatrix} 1 & 0 & \cdots & 0 \\ -M_{21} & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ -M_{m1} & 0 & \cdots & 1 \end{bmatrix} \cdots \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & 0 & \ddots & \vdots \\ 0 & \cdots & -M_{m(m-1)} & 1 \end{bmatrix},$$

where each factor has at most one nonzero column, the position of which moves from left to right as we scan the above product of matrices from left to right. It follows from what we know about the effect of multiplying on the left by an elementary matrix that the above product is of the form

$$\begin{bmatrix} 1 & 0 & \cdots & 0 & 0 \\ -M_{21} & 1 & \cdots & 0 & 0 \\ \vdots & -M_{32} & \ddots & \vdots & \vdots \\ -M_{(M-1)1} & \vdots & \cdots & 1 & 0 \\ -M_{M1} & -M_{M2} & \cdots & -M_{MM-1} & 1 \end{bmatrix}$$

To sum up the procedure, beginning at the left column and moving toward the right, you simply insert, into the corresponding position in the identity matrix, -1 times the multiplier which was used to zero out an entry in that position below the main diagonal in A, while retaining the main diagonal which consists entirely of ones. This will give us L as we create U using row operation 3.

We now return to Example **??**, to understand why we could not use the multiplier method to find an LU factorization of the coefficient matrix. Suppose we had written

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 12 & -18 \\ 1 & 2 & -3 \\ 1 & -2 & 1 \end{bmatrix}.$$

Our first step of Gaussian elimination would require a row swap to get a nonzero entry into the top left corner (we swapped rows 1 and 2 in Example **??**). Unfortunately, the elementary matrix that accomplishes this, $P = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ is NOT lower triangular. So we cannot apply Lemma **??** to generate a lower triangular L.

If the elementary matrices used to reduce our matrix to row-echelon form are all lower triangular, then we can find an LU factorization. But what about the general case?

Theorem 27. Suppose an $m \times n$ matrix A is transformed to a row-echelon matrix U using Gaussian elimination. Let $P_1, P_2, \ldots, P_s$ be the elementary matrices corresponding (in order) to the row interchanges used, and write $P = P_s \cdots P_2 P_1$. (If no interchanges are used take $P = I_M$.) Then:

1. PA is the matrix obtained from A by doing these interchanges (in order) to A.
2. PA has an LU factorization.

For a proof of the above theorem, the reader is referred to [Nicholson].

The matrix P in the above theorem is called a *permutation matrix*. These matrices have other important applications, as we will see later.

Practice Problems

Problem 204 *Use the given LU factorization of the coefficient matrix to solve the system of equations.*

$$\begin{array}{rcrcrcl} x & + & 2y & + & 3z & = & 5 \\ 2x & + & 3y & + & z & = & 6 \\ 3x & + & 5y & + & 4z & = & 11 \end{array}$$

Observe that an LU factorization of the coefficient matrix is

$$\begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \\ 3 & 5 & 4 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 3 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 0 & -1 & -5 \\ 0 & 0 & 0 \end{bmatrix}$$

Problems **??-??**

Find the LU factorization of the coefficient matrix and use it to solve the system of equations.

Problem 205

$$\begin{array}{rcrcl} x & + & 2y & = & 5 \\ 2x & + & 3y & = & 6 \end{array}$$

$$L = \begin{bmatrix} \ldots\ldots & 0 \\ \ldots\ldots & \ldots\ldots \end{bmatrix}, \quad U = \begin{bmatrix} \ldots\ldots & \ldots\ldots \\ 0 & \ldots\ldots \end{bmatrix}$$

$$x = \ldots\ldots, y = \ldots\ldots$$

Problem 206

$$\begin{array}{rcrcrcl} x & + & 2y & + & z & = & 1 \\ & & y & + & 3z & = & 2 \\ 2x & + & 3y & & & = & 6 \end{array}$$

$$L = \begin{bmatrix} \ldots\ldots & 0 & 0 \\ \ldots\ldots & \ldots\ldots & 0 \\ \ldots\ldots & \ldots\ldots & \ldots\ldots \end{bmatrix}, \quad U = \begin{bmatrix} \ldots\ldots & \ldots\ldots & \ldots\ldots \\ 0 & \ldots\ldots & \ldots\ldots \\ 0 & 0 & \ldots\ldots \end{bmatrix}$$

$$x = \ldots\ldots, y = \ldots\ldots, z = \ldots\ldots$$

Problem 207 *Is there only one LU factorization for a given matrix?*

Problem 208 *Can you show that every permutation matrix is invertible? If so, What does the inverse of a permutation matrix look like? (Recall that a permutation matrix is matrix P of Theorem* **??***.)*

Text Source

The text in this section is an adaptation of Section 2.2 of Ken Kuttler's *A First Course in Linear Algebra*. (CC-BY)

Ken Kuttler, *A First Course in Linear Algebra*, Lyryx 2017, Open Edition, p. 99-106.

Bibliography

[Nicholson] W. Keith Nicholson, *Linear Algebra with Applications*, Lyryx 2018, Open Edition, pp. 123–127.

34 Subspaces of Rn

35 $\mathbb{R}^n$ and Subspaces of $\mathbb{R}^n$

We are familiar with two operations that can be applied to vectors in $\mathbb{R}^n$, namely, addition and scalar multiplication. We learned that addition and scalar multiplication satisfy many nice properties (see Theorem **??**). These properties give $\mathbb{R}^n$ an algebraic structure. We begin this section by introducing another property, called *closure*. Adding closure to the properties we studied earlier allows us to show that $\mathbb{R}^n$ satisfies all of the properties of a *vector space*.

Closure

Definition 40. A set V is said to be *closed under addition* if for each element $\mathbf{u} \in V$ and $\mathbf{v} \in V$ the sum $\mathbf{u} + \mathbf{v}$ is also in V.

Definition 41. A set V is said to be *closed under scalar multiplication* if for each element $\mathbf{v} \in V$ and for each scalar $k \in \mathbb{R}$ the product $k\mathbf{v}$ is also in V.

Example 66. Let E be the set of positive even integers. Then E is closed under addition, because the sum of two even integers is again an even integer.

Example 67. Let D be the set of positive odd integers. Then D is not closed under addition, for the sum of two odd integers need not be an odd integer (in fact, it will always be even).

Example 68. Let X_1 be defined as follows.

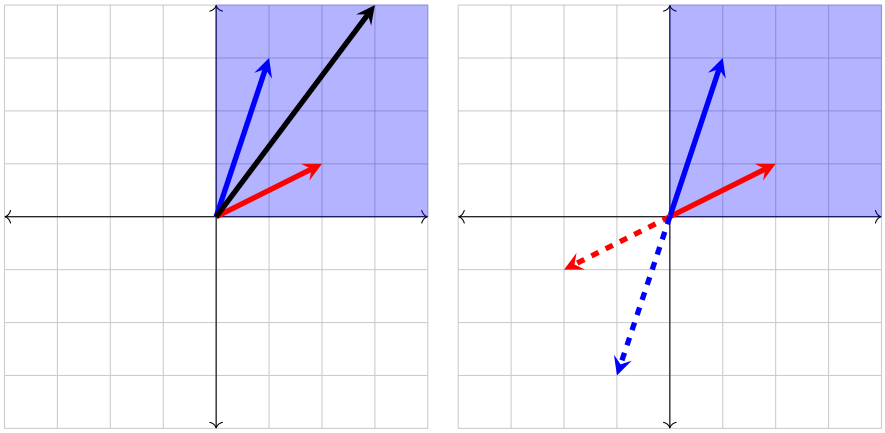
$$X_1 = \left\{ \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \in \mathbb{R}^2 : x_1 \geq 0, x_2 \geq 0 \right\}$$

In other words, X_1 is the set of vectors in $\mathbb{R}^2$ in the first quadrant (or on an axis on the boundary of the first quadrant). Show that X_1 is closed under addition, but X_1 is not closed under scalar multiplication.

Explanation. Suppose $\mathbf{v}_1 = \begin{bmatrix} a \\ b \end{bmatrix}$ and $\mathbf{v}_2 = \begin{bmatrix} c \\ d \end{bmatrix}$ are in X_1. This means that $a, b, c, d \geq 0$. But then we have $a + c, b + d \geq 0$. Therefore $\mathbf{v}_1 + \mathbf{v}_2 = \begin{bmatrix} a + c \\ b + d \end{bmatrix}$ is also in X_1. We conclude that X_1 is closed under addition.

X_1 is not closed under scalar multiplication because $(-1)\mathbf{v}_1 = \begin{bmatrix} -a \\ -b \end{bmatrix}$ is not in X_1.

The figure below helps us see that the sum of any two vectors in X_1 also lies in X_1, but any negative scalar multiple of a vector in X_1 does not lie in X_1.



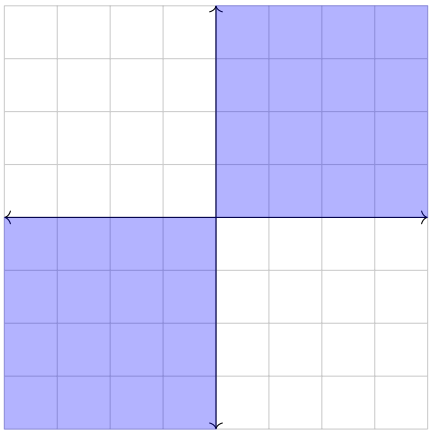
Example 69. Let X_1 be as in Example **??**. Let

$$X_3 = \left\{ \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \in \mathbb{R}^2 : x_1 \leq 0, x_2 \leq 0 \right\}$$

In other words, X_3 is the set of vectors in $\mathbb{R}^2$ in the third quadrant (or on an axis on the boundary of the third quadrant). Now let

$$Y = X_1 \bigcup X_3$$

In other words, Y is the set of vectors in $\mathbb{R}^2$ that are either in the first quadrant, the third quadrant, or lie along one of the axes, as shown below.



Then Y is closed under scalar multiplication, but Y is not closed under addition.

$\mathbb{R}^n$ as a Vector Space

In Theorem **??** we learned that vector addition and scalar multiplication in $\mathbb{R}^n$ satisfy the following eight properties:

For all vectors $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w} \in \mathbb{R}^n$, and scalars $k, p \in \mathbb{R}$,

1. Commutative Property of Addition: $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$
2. Associative Property of Addition: $(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$
3. Existence of Additive Identity: $\mathbf{u} + \mathbf{0} = \mathbf{u}$
4. Existence of Additive Inverse: $\mathbf{u} + (-\mathbf{u}) = \mathbf{0}$
5. Distributive Property over Vector Addition: $k(\mathbf{u} + \mathbf{v}) = k\mathbf{u} + k\mathbf{v}$
6. Distributive Property over Scalar Addition: $(k + p)\mathbf{u} = k\mathbf{u} + p\mathbf{u}$
7. Associative Property for Scalar Multiplication: $k(p\mathbf{u}) = (kp)\mathbf{u}$
8. Multiplication by 1: $1\mathbf{u} = \mathbf{u}$

In addition, observe that

$\mathbb{R}^n$ is closed under addition (Why?)

$\mathbb{R}^n$ is closed under scalar multiplication (Why?)

The eight properties of vector operations, together with closure, constitute the criteria for a set with two operations to be considered a *vector space*. So, $\mathbb{R}^n$ is a vector space.

We will encounter other vector spaces later. Any vector space must be closed under both of its operations, and must satisfy the other eight properties in the list above. We will see (in Abstract Vector Spaces, for instance) that a wide variety of sets, with a wide variety of operations, are vector spaces (one important reason to study linear algebra). As we shall see, these sets and their operations may look very different, but the behavior of the elements under the two operations makes them vector spaces. For now, we simply focus on $\mathbb{R}^n$.

Subspaces of $\mathbb{R}^n$

Now that we understand what it means for a set to be closed under addition and scalar multiplication, we are ready for the main definition.

Definition 42. Suppose that V is a nonempty subset of $\mathbb{R}^n$ that is closed under addition and closed under scalar multiplication. Then V is a *subspace* of $\mathbb{R}^n$.

We use the term *subspace* because it turns out that any subset of $\mathbb{R}^n$ closed under both addition and scalar multiplication is also a vector space. In other words, by inheriting vector addition and scalar multiplication from $\mathbb{R}^n$, and satisfying the properties of closure, a subset of $\mathbb{R}^n$ will automatically satisfy all vector space properties. We will prove this in Theorem **??** of Abstract Vector Spaces.

Example 70. Let V be the set of vectors in $\mathbb{R}^3$ on the $y-$axis. Then V is a subspace of $\mathbb{R}^3$.

Explanation. To verify this, note that any vector in V is of the form $\begin{bmatrix} 0 \\ y \\ 0 \end{bmatrix}$. If we multiply such a vector by a scalar c, we get a vector of the form $\begin{bmatrix} 0 \\ cy \\ 0 \end{bmatrix}$, which is clearly still in V. This proves V is closed under scalar multiplication. Next, note that $\begin{bmatrix} 0 \\ y_1 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ y_2 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ y_1 + y_2 \\ 0 \end{bmatrix}$, so V is closed under addition.

Recall that the *span* of a set of vectors is the set of all linear combinations of those vectors (see Definition **??**). It is easy to see from this definition, that the span of any set of vectors in $\mathbb{R}^n$ must be closed under both addition and scalar multiplication, and therefore the span of those vectors is a subspace of $\mathbb{R}^n$. This argument proves the following result, giving us an abundance of examples of subspaces:

Theorem 28. Let S be any set of vectors in $\mathbb{R}^n$. Then span(S) is a subspace of $\mathbb{R}^n$.

In particular, if we take S to be the single vector $\mathbf{v}$, we have that span($\mathbf{v}$) is a subspace of $\mathbb{R}^n$. Geometrically, this subspace is a line with a direction vector $\mathbf{v}$. Similarly, the span of two vectors is a subspace of $\mathbb{R}^n$. If the two vectors are linearly independent, then the subspace is a plane in $\mathbb{R}^n$.

Not every line or plane in $\mathbb{R}^n$ is a subspace, however. The following important result provides us with a quick way to determine that some subsets are not subspaces.

Theorem 29. If V is a subspace of $\mathbb{R}^n$, then the zero vector $\mathbf{0}$ is in V.

Proof Take any vector $\mathbf{v}$ in V, and note that $0\mathbf{v} = \mathbf{0}$ is in V because V is closed under scalar multiplication. ■

Theorem **??** shows that the only lines in $\mathbb{R}^n$ that are subspaces are those that pass through the origin. The same holds true for planes and hyperplanes. For example, the plane $z = 3$ in $\mathbb{R}^3$ is not a subspace of $\mathbb{R}^3$, while any plane containing the origin is a subspace.

Theorem 30. If V is a subspace of $\mathbb{R}^n$, then for any vector $\mathbf{v} \in V$, the opposite vector, $-\mathbf{v}$, is also in V.

The proof is similar to what was done for the previous theorem and is left as an exercise.

Practice Problems

Problem 209 *Let Y^+ be the set of all vectors in $\mathbb{R}^2$ whose y components are non-negative. Is Y^+ closed under vector addition?*

Multiple Choice:

1. *Yes*
2. *No*

Problem 210 *Let Y^+ be the set of all vectors in $\mathbb{R}^2$ whose y components are non-negative. Is Y^+ closed under scalar multiplication?*

Multiple Choice:

1. *Yes*
2. *No*

Problem 211 *Let X be the set of all vectors in $\mathbb{R}^3$ that lie on either the x-axis, the y-axis, or the z-axis. Is X closed under vector addition?*

Multiple Choice:

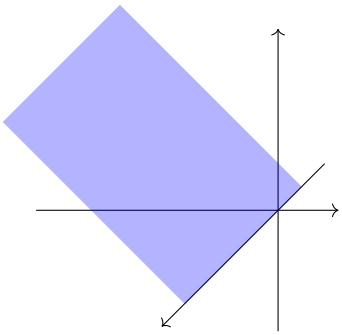
1. *Yes*
2. *No*

Problem 212 *Let X be the set of all vectors in $\mathbb{R}^3$ that lie on either the x-axis, the y-axis, or the z-axis. Is X closed under scalar multiplication?*

Multiple Choice:

1. *Yes*
2. *No*

Problem 213 *Determine whether the set V of vectors shown in the figure is closed under vector addition and scalar multiplication. Justify your responses. V consists of all vectors in $\mathbb{R}^3$ in a slanted half-plane which has the x-axis as a boundary.*



Is V closed under scalar multiplication?

Multiple Choice:

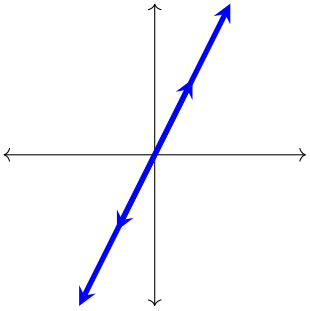
1. *Yes*
2. *No*

Is V closed under addition?

Multiple Choice:

1. *Yes*
2. *No*

Problem 214 *Determine whether the set V of vectors shown in the figure is closed under vector addition and scalar multiplication. Justify your responses. V consists of all vectors along the line, as shown.*



Is V closed under scalar multiplication?

Multiple Choice:

1. *Yes*
2. *No*

Is V closed under addition?

Multiple Choice:

1. *Yes*
2. *No*

Problem 215 *Prove that if V is a subspace of $\mathbb{R}^n$, then for any vector $\mathbf{v} \in V$, the opposite vector, $-\mathbf{v}$, is also in V. (Theorem **??**)*

Problem 216 *Let A be an $m \times n$ matrix. Let V be the subset of $\mathbb{R}^n$ consisting of all vectors $\mathbf{x}$ such that $A\mathbf{x} = \mathbf{0}$. Prove that V is a subspace of $\mathbb{R}^n$. (This subspace is called the null space of the matrix A. We will denote it null(A).)*

36 Introduction to Bases

Coordinate Vectors

When we first introduced vectors we learned to represent them using component notation. If $\mathbf{u} = \begin{bmatrix} 2 \\ 5 \end{bmatrix}$ we know that the head of $\mathbf{u}$ is located at the point $(2, 5)$. But there is another way to look at the component form. Observe that $\mathbf{u}$ can be expressed as a linear combination of the standard unit vectors $\mathbf{i}$ and $\mathbf{j}$:

$$\mathbf{u} = 2\begin{bmatrix} 1 \\ 0 \end{bmatrix} + 5\begin{bmatrix} 0 \\ 1 \end{bmatrix} = 2\mathbf{i} + 5\mathbf{j}$$

In fact, any vector $\mathbf{v} = \begin{bmatrix} a \\ b \end{bmatrix}$ of $\mathbb{R}^2$ can be written as a linear combination of $\mathbf{i}$ and $\mathbf{j}$:

$$\mathbf{v} = \begin{bmatrix} a \\ b \end{bmatrix} = a\mathbf{i} + b\mathbf{j}$$

This gives us an alternative way of interpreting the component notation:

$$\begin{bmatrix} a \\ b \end{bmatrix} \begin{array}{ll} \longleftarrow & \text{coefficient in front of } \mathbf{i} \\ \longleftarrow & \text{coefficient in front of } \mathbf{j} \end{array}$$

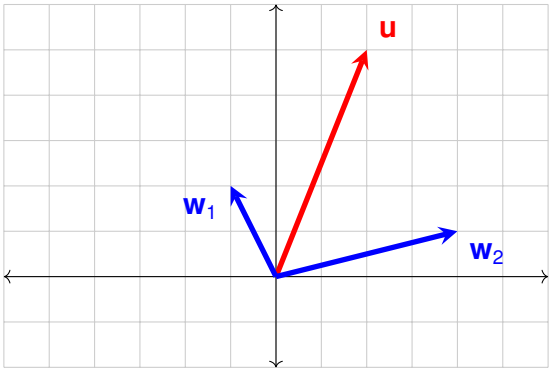
We say that a and b are *coordinates* of $\mathbf{v}$ with respect to $\{\mathbf{i}, \mathbf{j}\}$, and $\begin{bmatrix} a \\ b \end{bmatrix}$ is said to be the *coordinate vector* for $\mathbf{v}$ with respect to $\{\mathbf{i}, \mathbf{j}\}$. Every vector $\mathbf{v}$ of $\mathbb{R}^2$ can be thus represented using $\mathbf{i}$ and $\mathbf{j}$. Moreover, such representation in terms of $\mathbf{i}$ and $\mathbf{j}$ is unique for each vector, meaning that we will never have two different coordinate vectors representing the same vector. We will refer to $\{\mathbf{i}, \mathbf{j}\}$ as a *basis* of $\mathbb{R}^2$.

The order in which the basis elements are written matters. For example, $\mathbf{u}$ is represented by the coordinate vector $\begin{bmatrix} 2 \\ 5 \end{bmatrix}$ with respect to $\{\mathbf{i}, \mathbf{j}\}$, but changing the basis to $\{\mathbf{j}, \mathbf{i}\}$ would change the coordinate vector to $\begin{bmatrix} 5 \\ 2 \end{bmatrix}$.

$$\begin{bmatrix} 5 \\ 2 \end{bmatrix} \begin{array}{ll} \longleftarrow & \text{coefficient in front of the first basis element} \\ \longleftarrow & \text{coefficient in front of the second basis element} \end{array}$$

Clearly, standard unit vectors **i** and **j** are very convenient, but other vectors can also be used in place of **i** and **j** to represent **u**.

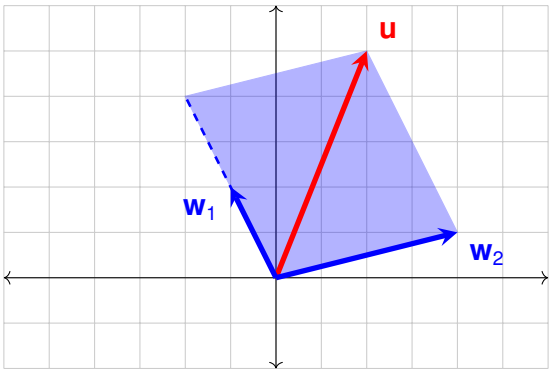
Exploration 217 *The diagram below shows* **u** *together with vectors* $\mathbf{w}_1$ *and* $\mathbf{w}_2$*.*



It is easy to see that

$$\mathbf{u} = 2\mathbf{w}_1 + \mathbf{w}_2$$

as shown below.



If we declare $\{\mathbf{w}_1, \mathbf{w}_2\}$ *to be a basis of* $\mathbb{R}^2$*, then we can say that the coordinate vector for* **u** *with respect to* $\{\mathbf{w}_1, \mathbf{w}_2\}$ *is* $\begin{bmatrix} 2 \\ 1 \end{bmatrix}$ *.*

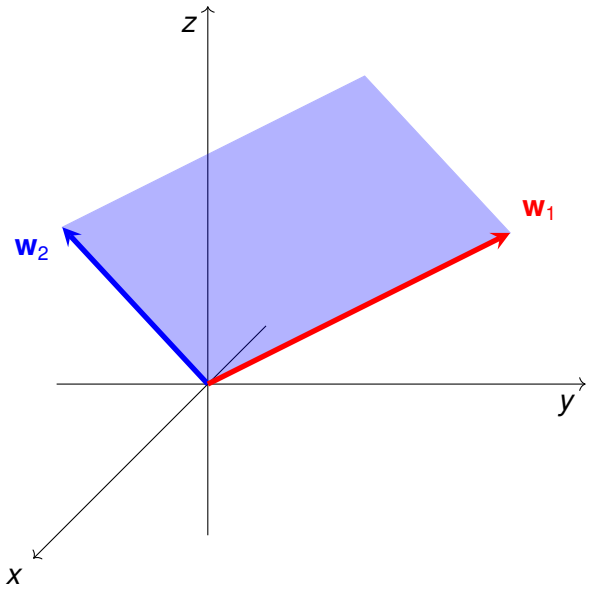
$$\begin{bmatrix} 2 \\ 1 \end{bmatrix} \begin{array}{ll} \longleftarrow & \textit{coefficient in front of the first basis element} \\ \longleftarrow & \textit{coefficient in front of the second basis element} \end{array}$$

What Constitutes a Basis?

In the previous section we had used the term *basis* without defining it. Now is the time to pause and think about what we *want* a basis to do. Let's focus on $\mathbb{R}^n$ and subspaces of $\mathbb{R}^n$. What we establish here will easily generalize to other vector spaces.

Based on our previous discussion, given any vector $\mathbf{v}$ of $\mathbb{R}^n$ (or a subspace V of $\mathbb{R}^n$), we want to be able to write a coordinate vector for $\mathbf{v}$ with respect to the given basis of $\mathbb{R}^n$ (or V). Based on this condition, we will require that basis vectors span $\mathbb{R}^n$ (or V).

For example, consider $\mathbf{w}_1$ and $\mathbf{w}_2$ shown below.



The set $\{\mathbf{w}_1, \mathbf{w}_2\}$ cannot be a basis for $\mathbb{R}^3$ because $\mathbf{w}_1$ and $\mathbf{w}_2$ span a plane in $\mathbb{R}^3$, and any vector not in that plane cannot be written as a linear combination of $\mathbf{w}_1$ and $\mathbf{w}_2$.

On the other hand, the plane spanned by $\mathbf{w}_1$ and $\mathbf{w}_2$ is a subspace of $\mathbb{R}^3$. Because every vector in that plane can be written as a linear combination of $\mathbf{w}_1$ and $\mathbf{w}_2$, the set $\{\mathbf{w}_1, \mathbf{w}_2\}$ can potentially be a basis for the plane, provided that the set satisfies our second requirement.

Our second requirement is that for a fixed basis of $\mathbb{R}^n$ (or V), the coordinate vector for each $\mathbf{v}$ in $\mathbb{R}^n$ (or V) should be unique. Uniqueness of representation in terms of the basis elements will play an important role in our future study of functions that map vector spaces to vector spaces.

The following theorem shows that the uniqueness requirement is equivalent to the requirement that the basis vectors be linearly independent.

Theorem 31. Suppose $\{\mathbf{w}_1, \mathbf{w}_2, ..., \mathbf{w}_p\}$ is a spanning set for a subspace V of $\mathbb{R}^n$. Then every element $\mathbf{v}$ of V has a unique representation as a linear combination of $\mathbf{w}_1, \mathbf{w}_2, ..., \mathbf{w}_p$ if and only if the vectors $\mathbf{w}_1, \mathbf{w}_2, ..., \mathbf{w}_p$ are linearly independent.

Proof Suppose that every $\mathbf{v}$ in V can be expressed as a unique linear combination of $\mathbf{w}_1, \mathbf{w}_2, ..., \mathbf{w}_p$. This means that $\mathbf{0}$ has a unique representation as a linear combination of $\mathbf{w}_1, \mathbf{w}_2, ..., \mathbf{w}_p$. But

$$\mathbf{0} = 0\mathbf{w}_1 + 0\mathbf{w}_2 + ... + 0\mathbf{w}_p$$

is a representation of $\mathbf{0}$ in terms of $\mathbf{w}_1, \mathbf{w}_2, ..., \mathbf{w}_p$. Since we are assuming that such a representation is unique, we conclude that there is no other. This means that the vectors $\mathbf{w}_1, \mathbf{w}_2, ..., \mathbf{w}_p$ are linearly independent.

Conversely, suppose that vectors $\mathbf{w}_1, \mathbf{w}_2, ..., \mathbf{w}_p$ are linearly independent. An arbitrary element $\mathbf{v}$ of V can be expressed as a linear combination of $\mathbf{w}_1, \mathbf{w}_2, ..., \mathbf{w}_p$:

$$\mathbf{v} = a_1\mathbf{w}_1 + a_2\mathbf{w}_2 + ... + a_p\mathbf{w}_p$$

Suppose this representation is not unique. Then there may be another linear combination that is also equal to $\mathbf{v}$:

$$\mathbf{v} = b_1\mathbf{w}_1 + b_2\mathbf{w}_2 + ... + b_p\mathbf{w}_p$$

But then

$$a_1\mathbf{w}_1 + a_2\mathbf{w}_2 + ... + a_p\mathbf{w}_p = b_1\mathbf{w}_1 + b_2\mathbf{w}_2 + ... + b_p\mathbf{w}_p$$

This gives us

$$(a_1 - b_1)\mathbf{w}_1 + (a_2 - b_2)\mathbf{w}_2 + ... + (a_p - b_p)\mathbf{w}_p = \mathbf{0}$$

Because we assumed that $\mathbf{w}_1, \mathbf{w}_2, ..., \mathbf{w}_p$ are linearly independent, we must have

$$a_1 - b_1 = 0,\ a_2 - b_2 = 0,\ ...,\ a_p - b_p = 0$$

so that

$$a_1 = b_1,\ a_2 = b_2,\ ...,\ a_p = b_p$$

This proves the representation of $\mathbf{v}$ in terms of $\mathbf{w}_1, \mathbf{w}_2, ..., \mathbf{w}_p$ is unique.

■

Example 71. Use $V = \text{span}(\mathcal{S})$, where $\mathcal{S} = \left\{ \begin{bmatrix} 5 \\ 2 \\ 4 \end{bmatrix}, \begin{bmatrix} 4 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} -3 \\ 0 \\ 2 \end{bmatrix} \right\}$ to illustrate why a set of linearly dependent vectors cannot be used as a basis for a subspace by showing that linearly dependent vectors fail to ensure uniqueness of coordinate vectors for vectors in V.

Explanation. We will first show that the elements of $\mathcal{S}$ are linearly dependent. Let A be a matrix whose columns are the vectors in $\mathcal{S}$.

$$A = \begin{bmatrix} 5 & 4 & -3 \\ 2 & 1 & 0 \\ 4 & 1 & 2 \end{bmatrix}$$

We find that

$$\text{rref}(A) = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & -2 \\ 0 & 0 & 0 \end{bmatrix}$$

Therefore the matrix equation $A\mathbf{x} = \mathbf{0}$ has infinitely many solutions:

$$\mathbf{x} = \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix} t$$

This tells us that there are infinitely many nontrivial linear relations among the elements of $\mathcal{S}$. Letting $t = 1$ gives us one such nontrivial relation.

$$-\begin{bmatrix} 5 \\ 2 \\ 4 \end{bmatrix} + 2\begin{bmatrix} 4 \\ 1 \\ 1 \end{bmatrix} + \begin{bmatrix} -3 \\ 0 \\ 2 \end{bmatrix} = \mathbf{0}$$

Now let's pick an arbitrary vector $\mathbf{v}$ in V. Any vector will do, so let

$$\mathbf{v} = \begin{bmatrix} 5 \\ 2 \\ 4 \end{bmatrix} + (-1)\begin{bmatrix} 4 \\ 1 \\ 1 \end{bmatrix} + 0\begin{bmatrix} -3 \\ 0 \\ 2 \end{bmatrix}$$

Based on this representation of $\mathbf{v}$, the coordinate vector for $\mathbf{v}$ with respect to $\mathcal{S}$ is

$$\begin{bmatrix} \ldots \\ \ldots \\ \ldots \end{bmatrix}$$

But

$$\begin{bmatrix} 5 \\ 2 \\ 4 \end{bmatrix} = 2\begin{bmatrix} 4 \\ 1 \\ 1 \end{bmatrix} + \begin{bmatrix} -3 \\ 0 \\ 2 \end{bmatrix}$$

So, by substitution, we have:

$$\mathbf{v} = \left(2 \begin{bmatrix} 4 \\ 1 \\ 1 \end{bmatrix} + \begin{bmatrix} -3 \\ 0 \\ 2 \end{bmatrix} \right) + (-1) \begin{bmatrix} 4 \\ 1 \\ 1 \end{bmatrix} + 0 \begin{bmatrix} -3 \\ 0 \\ 2 \end{bmatrix} = 0 \begin{bmatrix} 5 \\ 2 \\ 4 \end{bmatrix} + 1 \begin{bmatrix} 4 \\ 1 \\ 1 \end{bmatrix} + 1 \begin{bmatrix} -3 \\ 0 \\ 2 \end{bmatrix}$$

Based on this representation, the coordinate vector for $\mathbf{v}$ with respect to $\mathcal{S}$ is

$$\begin{bmatrix} \dots \dots \\ \dots \dots \\ \dots \dots \end{bmatrix}$$

The set $\mathcal{S}$ is linearly dependent. As a result, coordinate vectors for elements of V are not unique and we do not want to use $\mathcal{S}$ as a basis for V.

Definition of a Basis

Definition 43. A set $\mathcal{S}$ of vectors is called a *basis* of $\mathbb{R}^n$ (or a basis of a subspace V of $\mathbb{R}^n$) provided that

1. $\text{span}(\mathcal{S}) = \mathbb{R}^n$ (or V)
2. $\mathcal{S}$ is linearly independent.

Example 72. The standard unit vectors $\mathbf{e}_1, \dots, \mathbf{e}_n$ are linearly independent and span $\mathbb{R}^n$. Thus $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ is a basis of $\mathbb{R}^n$.

Definition 44. The set $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ is called the *standard basis* of $\mathbb{R}^n$.

Bases are not unique. For example, we know that vectors **i** and **j** form the standard basis of $\mathbb{R}^2$. But, as we discussed in Example **??**, vectors

$$\begin{bmatrix} 2 \\ 2 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \end{bmatrix}$$

are linearly independent vectors that span $\mathbb{R}^2$. Therefore

$$\left\{ \begin{bmatrix} 2 \\ 2 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \end{bmatrix} \right\}$$

is also a basis for $\mathbb{R}^2$.

Any linearly independent spanning set in $\mathbb{R}^n$ (or a subspace of $\mathbb{R}^n$) is a basis of $\mathbb{R}^n$ (or the subspace). The plural form of the word *basis* is *bases*. It is easy to see that $\mathbb{R}^n$ and its subspaces each has infinitely many bases.

Example 73. Let $V = \text{span}\left(\begin{bmatrix}-2\\1\\3\end{bmatrix}, \begin{bmatrix}2\\-4\\1\end{bmatrix}\right)$. The set

$$\mathcal{B} = \left\{\begin{bmatrix}-2\\1\\3\end{bmatrix}, \begin{bmatrix}2\\-4\\1\end{bmatrix}\right\}$$

is a basis for V because the two vectors in $\mathcal{B}$ are linearly independent and span V. Find the coordinate vector for $\mathbf{v} = \begin{bmatrix}2\\-10\\9\end{bmatrix}$ with respect to $\mathcal{B}$.

Explanation. We need to express $\begin{bmatrix}2\\-10\\9\end{bmatrix}$ as a linear combination of the elements of $\mathcal{B}$. To this end, we need to solve the vector equation:

$$a_1\begin{bmatrix}-2\\1\\3\end{bmatrix} + a_2\begin{bmatrix}2\\-4\\1\end{bmatrix} = \begin{bmatrix}2\\-10\\9\end{bmatrix}$$

The augmented matrix and the reduced row-echelon form are:

$$\left[\begin{array}{cc|c}-2&2&2\\-1&-4&-10\\3&1&9\end{array}\right] \rightsquigarrow \left[\begin{array}{cc|c}1&0&2\\0&1&3\\0&0&0\end{array}\right]$$

We conclude that $a_1 = 2$, $a_2 = 3$. This gives us

$$2\begin{bmatrix}-2\\1\\3\end{bmatrix} + 3\begin{bmatrix}2\\-4\\1\end{bmatrix} = \begin{bmatrix}2\\-10\\9\end{bmatrix}$$

The coefficient in front of the first basis vector is 2, the coefficient in front of the second basis vector is 3. This means that the coordinate vector for $\begin{bmatrix}2\\-10\\9\end{bmatrix}$ with respect to $\mathcal{B}$ is $\begin{bmatrix}2\\3\end{bmatrix}$.

Remark 5. It may seem strange to you that the coordinate vector for a vector in $\mathbb{R}^3$ only has two components. But remember that subspace V is a plane. When viewed as a vector in the plane, it makes sense that the coordinate vector for $\begin{bmatrix} 2 \\ -10 \\ 9 \end{bmatrix}$ only requires two components. This issue is related to the question of dimension, which will be addressed in Bases and Dimension.

Remark 6. To construct the coordinate vector for $\begin{bmatrix} 2 \\ -10 \\ 9 \end{bmatrix}$ with respect to $\mathcal{B}$, we had to be mindful of the order of the elements in $\mathcal{B}$. Ordinarily, the order of elements in a set is irrelevant, and the basis

$$\left\{ \begin{bmatrix} -2 \\ 1 \\ 3 \end{bmatrix}, \begin{bmatrix} 2 \\ -4 \\ 1 \end{bmatrix} \right\}$$

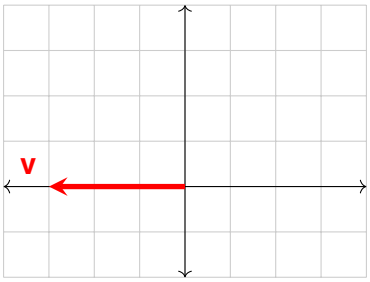
is considered to be the same as

$$\left\{ \begin{bmatrix} 2 \\ -4 \\ 1 \end{bmatrix}, \begin{bmatrix} -2 \\ 1 \\ 3 \end{bmatrix} \right\}$$

When dealing with coordinate vectors, however, the order of the elements dictates the order of the components of the coordinate vector coefficients. If we switch the order of the elements in $\mathcal{B}$, the coordinate vector becomes $\begin{bmatrix} 3 \\ 2 \end{bmatrix}$. For this reason, when we come back to studying coordinate vectors in more detail, we will use the term *ordered basis* to avoid confusion.

Practice Problems

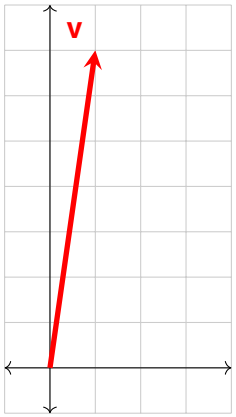
Problem 218 *Let* $\mathcal{B} = \left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 2 \end{bmatrix} \right\}$ *be a basis for* $\mathbb{R}^2$*. Find the coordinate vector for* $\mathbf{v}$ *with respect to* $\mathcal{B}$*.*



Answer:

$$\begin{bmatrix} \dots & \dots \\ \dots & \dots \end{bmatrix}$$

Problem 219 *Let* $\mathcal{B} = \left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 2 \end{bmatrix} \right\}$ *be a basis for* $\mathbb{R}^2$*. Find the coordinate vector for* $\mathbf{v}$ *with respect to* $\mathcal{B}$*.*



Answer:

$$\begin{bmatrix} \dots & \dots \\ \dots & \dots \end{bmatrix}$$

Problem 220 *Let* $\mathcal{B} = \left\{ \begin{bmatrix} 1 \\ -1 \\ 3 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ -1 \end{bmatrix} \right\}$ *be a basis for span* $\left(\begin{bmatrix} 1 \\ -1 \\ 3 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ -1 \end{bmatrix} \right)$*. Find the coordinate vector for* $\begin{bmatrix} -4 \\ -2 \\ 2 \end{bmatrix}$ *with respect to* $\mathcal{B}$*.*

Answer:

$$\begin{bmatrix} \dots & \dots \\ \dots & \dots \end{bmatrix}$$

Problem 221 *Suppose* $\mathcal{B} = \left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \mathbf{w} \right\}$ *is a basis for* $\mathbb{R}^3$*. Find* $\mathbf{w}$ *if the coordinate vector for* $\begin{bmatrix} -2 \\ -7 \\ 4 \end{bmatrix}$ *is* $\begin{bmatrix} -1 \\ 2 \\ -3 \end{bmatrix}$.

Answer:

$$\begin{bmatrix} \dots & \dots \\ \dots & \dots \\ \dots & \dots \end{bmatrix}$$

Problem 222 *Which of the following is a basis for* $\mathbb{R}^2$*?*

Select All Correct Answers:

1. $\left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \end{bmatrix} \right\}$
2. $\left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \end{bmatrix} \right\}$
3. $\left\{ \begin{bmatrix} 3 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} -4 \\ 3 \end{bmatrix} \right\}$
4. $\left\{ \begin{bmatrix} 1 \\ -3 \end{bmatrix}, \begin{bmatrix} -2 \\ 6 \end{bmatrix} \right\}$

Problem 223 *Which of the following is a basis for* $V = \text{span}\left(\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix} \right)$*?*

Select All Correct Answers:

1. $\left\{ \begin{bmatrix} 2 \\ -1 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix} \right\}$
2. $\left\{ \begin{bmatrix} 0 \\ 3 \\ 0 \end{bmatrix}, \begin{bmatrix} 3 \\ -3 \\ 3 \end{bmatrix} \right\}$
3. $\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}$
4. $\left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ -1 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix} \right\}$

37 Bases and Dimension

Recall that a *basis* of a subspace V of $\mathbb{R}^n$ is a subset of V that is linearly independent and spans V. A basis allows us to uniquely express every element of V as a linear combination of the elements of the basis. Several questions may come to mind at this time. Does every subspace of $\mathbb{R}^n$ have a basis? We know that bases are not unique. If there is more than one basis, what, if anything, do they have in common?

Exploring Dimension

Exploration 224 *How would you describe*

$$V = span\left(\begin{bmatrix}1\\-2\\3\end{bmatrix}, \begin{bmatrix}-2\\4\\-6\end{bmatrix}\right)?$$

If you answered that V is a line in $\mathbb{R}^3$, you are correct. While the two vectors span the line, it is not necessary to have both of them in the spanning set to describe the line.

What is the minimum number of vectors needed to span a line?

Answer: ... .

Observe also that the vectors in the given spanning set are not linearly independent, so they do not form a basis for V. How many vectors would a basis for V have?

Answer: ... .

Now consider another subspace of $\mathbb{R}^3$:

$$W = span\left(\begin{bmatrix}1\\0\\2\end{bmatrix}, \begin{bmatrix}0\\-3\\0\end{bmatrix}\right)$$

Geometrically, W is a plane in $\mathbb{R}^3$. Note that the vectors in the spanning set are linearly independent. Can we remove one of the vectors and have the remaining vector span the plane?

What is the minimum number of vectors needed to span a plane?

Answer: ... .

How many vectors would a basis for a plane have?

Answer: ... .

Our observations in Exploration **??** hint at the idea of dimension. We know that a line is a one-dimensional object, a plane is a two-dimensional object, and the space we reside in is three-dimensional.

Based on our observations in Exploration **??**, it makes sense for us to define dimension of a vector space (or a subspace) as the minimum number of vectors required to span the space (subspace). We can accomplish this by defining dimension as the number of elements in a basis.

We have to proceed carefully because we don't want the dimension to depend on our choice of a basis. So, before we state our definition, we need to make sure that every basis for a given vector space (or subspace) has the same number of elements.

Theorem 32. Suppose $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_t\}$ and $\mathcal{C} = \{\mathbf{w}_1, \mathbf{w}_2, ..., \mathbf{w}_s\}$ be two bases of $\mathbb{R}^n$ (or a subspace V of $\mathbb{R}^n$). Then $s = t$.

Proof Suppose $s \neq t$. Without loss of generality, assume that $s > t$. Because $\mathcal{B}$ spans V, every $\mathbf{w}_i$ of $\mathcal{C}$ can be written as a linear combination of elements of $\mathcal{B}$:

$$\mathbf{w}_i = a_{1i}\mathbf{v}_1 + a_{2i}\mathbf{v}_2 + ... + a_{ti}\mathbf{v}_t$$

Consider the vector equation

$$b_1\mathbf{w}_1 + b_2\mathbf{w}_2 + ... + b_s\mathbf{w}_s = \mathbf{0} \tag{36}$$

By substitution, we have:

$$\begin{aligned}
& b_1\mathbf{w}_1 + b_2\mathbf{w}_2 + ... + b_s\mathbf{w}_s = \\
&= b_1(a_{11}\mathbf{v}_1 + a_{21}\mathbf{v}_2 + ... + a_{t1}\mathbf{v}_t) + b_2(a_{12}\mathbf{v}_1 + a_{22}\mathbf{v}_2 + ... + a_{t2}\mathbf{v}_t) + ... \\
&\quad + b_s(a_{1s}\mathbf{v}_1 + a_{2s}\mathbf{v}_2 + ... + a_{ts}\mathbf{v}_t) \\
&= (b_1a_{11} + b_2a_{12} + ... + b_sa_{1s})\mathbf{v}_1 + (b_1a_{21} + b_2a_{22} + ... + b_sa_{2s})\mathbf{v}_2 + ... \\
&\quad + (b_1a_{t1} + b_2a_{t2} + ... + b_sa_{ts})\mathbf{v}_t \\
&= \mathbf{0}
\end{aligned}$$

Because $\mathbf{v}_j$'s are linearly independent, we must have

$$b_1a_{j1} + b_2a_{j2} + ... + b_sa_{js} = 0$$

For all $1 \leq j \leq t$. This gives us a system of t equations and s unknowns. We can write the system as a matrix equation.

$$\begin{bmatrix} a_{11} & a_{12} & ... & a_{1s} \\ a_{21} & a_{22} & ... & a_{2s} \\ \vdots & \vdots & \ddots & \vdots \\ a_{t1} & a_{t2} & ... & a_{ts} \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_s \end{bmatrix} = \mathbf{0}$$

Recall our assumption that $s > t$. By Theorem **??**, we know that the system has infinitely many solutions. This shows that equation (**??**) has a nontrivial solution (in fact, infinitely many of them). But this shows that $\{\mathbf{w}_1, \mathbf{w}_2, ..., \mathbf{w}_s\}$ is linearly dependent and contradicts our assumption that $\mathcal{C}$ is a basis of V. We conclude that $s = t$. ■

Definition 45. Let V be a subspace of $\mathbb{R}^n$. The *dimension* of V is the number, m, of elements in any basis of V. We write

$$\dim(V) = m$$

Example 74. We know that vectors $\mathbf{e}_1, ..., \mathbf{e}_n$ form a basis of $\mathbb{R}^n$. Therefore $\dim(\mathbb{R}^n) = n$.

Remark 7. It is easy to verify that $\{\mathbf{0}\}$ is a subspace of $\mathbb{R}^n$. Because $\{\mathbf{0}\}$ is linearly dependent, it does not have a basis. We define $\dim\{\mathbf{0}\} = 0$.

The following section will guarantee that dimension is defined for every subspace of $\mathbb{R}^n$.

Every Subspace of $\mathbb{R}^n$ has a Basis

Lemma 3. If a linearly independent subset of $\mathbb{R}^n$ contains m vectors, then $m \leq n$.

Proof See Practice Problem **??**. ■

Lemma 4. Let $\{\mathbf{v}_1, ..., \mathbf{v}_k\}$ be a linearly independent subset of $\mathbb{R}^n$. If $\mathbf{u}$ is not in span$(\mathbf{v}_1, ..., \mathbf{v}_k)$, then $\{\mathbf{u}, \mathbf{v}_1, ..., \mathbf{v}_k\}$ is linearly independent.

Proof Consider the equation

$$a\mathbf{u} + a_1\mathbf{v}_1 + ... + a_k\mathbf{v}_k = \mathbf{0} \tag{37}$$

We need to show that $a = a_1 = ... = a_k = 0$. Suppose $a \neq 0$, then $\mathbf{u} = \frac{-a_1}{a}\mathbf{v}_1 + ... + \frac{-a_k}{a}\mathbf{v}_k$. But this contradicts the assumption that $\mathbf{u}$ is not in the span of $\mathbf{v}_1, ..., \mathbf{v}_k$. So, $a = 0$. But $a_1 = ... = a_k = 0$ because $\mathbf{v}_1, ..., \mathbf{v}_k$ are linearly independent.

This means that (**??**) has only the trivial solution, and $\{\mathbf{u}, \mathbf{v}_1, ..., \mathbf{v}_k\}$ is linearly independent. ■

Theorem 33. Let V be a subspace of $\mathbb{R}^n$. Any linearly independent subset of V can be expanded to a basis of V.

Proof Suppose that $X = \{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ is a linearly independent subset of V. If $\text{span}(X) = V$ then X is already a basis of V. If $\text{span}(X) \neq V$, choose $\mathbf{u}_1$ in V such that $\mathbf{u}_1$ is not in $\text{span}(X)$. The set $\{\mathbf{u}_1, \mathbf{v}_1, \ldots, \mathbf{v}_k\}$ is linearly independent by Lemma **??**.

If $\text{span}(\mathbf{u}_1, \mathbf{v}_1, \ldots, \mathbf{v}_k) = V$ we are done; otherwise choose $\mathbf{u}_2 \in V$ such that $\mathbf{u}_2$ is not in $\text{span}(\mathbf{u}_1, \mathbf{v}_1, \ldots, \mathbf{v}_k)$. Then $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{v}_1, \ldots, \mathbf{v}_k\}$ is linearly independent, and the process continues. We claim that a basis of V will be reached eventually. If no basis of V is ever reached, the process creates arbitrarily large independent sets in $\mathbb{R}^n$. But this is impossible by Lemma **??**. ■

Practice Problems

Problem 225 *Find dim(span(S)).*

$$S = \left\{ \begin{bmatrix} 1 \\ 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \\ 1 \end{bmatrix} \right\}$$

Answer: dim(span(S)) = ...

Problem 226 *Find dim(span(S)).*

$$S = \left\{ \begin{bmatrix} 3 \\ -2 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 3 \\ 3 \\ -2 \end{bmatrix}, \begin{bmatrix} 1 \\ -5 \\ -2 \\ 3 \end{bmatrix} \right\}$$

Answer: dim(span(S)) = ...

Problem 227 *Find dim(span(S)).*

$$S = \left\{ \begin{bmatrix} 1 \\ 1 \\ -3 \end{bmatrix}, \begin{bmatrix} -3 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 5 \\ -2 \\ 4 \end{bmatrix} \right\}$$

Answer: dim(span(S)) = ...

Problem 228 *Prove that if a linearly independent subset of $\mathbb{R}^n$ contains m vectors, then $m \leq n$. (Lemma **??**)*

Exercise Source

Practice Problem **??** is adopted from Keith Nicholson's Example 5.2.12 *Linear Algebra with Applications*. (CC-BY-NC-SA)

W. Keith Nicholson, *Linear Algebra with Applications*, Lyryx 2018, Open Edition, p 277.

38 Subspaces of $\mathbb{R}^n$ Associated with Matrices

Row Space of a Matrix

Recall that in Gaussian Elimination and Rank, we claimed that every row-echelon form of a given matrix has the same number of nonzero rows. This result suggests that there are certain characteristics associated with the rows of a matrix that are not affected by elementary row operations. We are now in the position to examine this question and to supply the proof we omitted earlier.

Definition 46. Let A be an $m \times n$ matrix. The *row space* of A, denoted by row(A), is the subspace of $\mathbb{R}^n$ spanned by the rows of A.

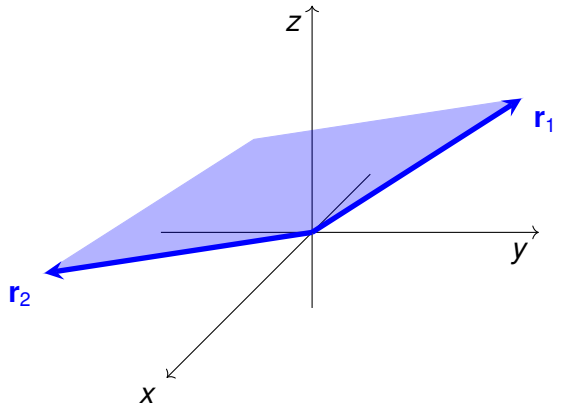
Exploration 229 *Consider the matrix*

$$A = \begin{bmatrix} -2 & 2 & 1 \\ 4 & -2 & 1 \end{bmatrix}$$

Let $\mathbf{r}_1$ *and* $\mathbf{r}_2$ *be the rows of A:*

$$\mathbf{r}_1 = \begin{bmatrix} -2 & 2 & 1 \end{bmatrix}, \quad \mathbf{r}_2 = \begin{bmatrix} 4 & -2 & 1 \end{bmatrix}$$

Then row(A) = *span*$(\mathbf{r}_1, \mathbf{r}_2)$ *is a plane through the origin containing* $\mathbf{r}_1$ *and* $\mathbf{r}_2$.



We will use elementary row operations to reduce A to rref(A).

$$\begin{bmatrix} -2 & 2 & 1 \\ 4 & -2 & 1 \end{bmatrix} \rightsquigarrow \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 3/2 \end{bmatrix}$$

Let ρ_1 and ρ_2 be the rows of rref(A):

$$\rho_1 = \begin{bmatrix} 1 & 0 & 1 \end{bmatrix}, \quad \rho_2 = \begin{bmatrix} 0 & 1 & 3/2 \end{bmatrix}$$

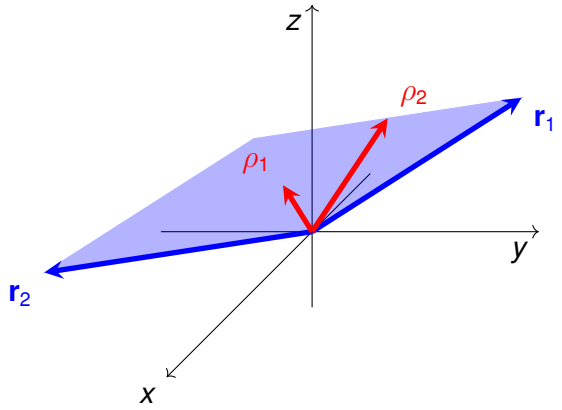
What do you think span(ρ_1, ρ_2) looks like?

The following video will help us visualize span(ρ_1, ρ_2) and compare it to span($\mathbf{r}_1, \mathbf{r}_2$).

YouTube link: `https://www.youtube.com/watch?v=6wz-5G14jo8`

Based on what we observed in the video, we may conjecture that

$$span(\rho_1, \rho_2) = span(\mathbf{r}_1, \mathbf{r}_2)$$



But why does this make sense? Vectors ρ_1 and ρ_2 were obtained from $\mathbf{r}_1$ and $\mathbf{r}_2$ by repeated applications of elementary row operations. At every stage of the row reduction process, the rows of the matrix are linear combinations of $\mathbf{r}_1$ and $\mathbf{r}_2$. Thus, at every stage of the row reduction process, the rows of the matrix lie in the span of $\mathbf{r}_1$ and $\mathbf{r}_2$. Our next video shows a step-by-step row reduction process accompanied by sketches of vectors.

YouTube link: `https://www.youtube.com/watch?v=KpoeWUQ3wkY`

Exploration **??** makes a convincing case for the following theorem.

Theorem 34. If matrix B was obtained from matrix A by applying an elementary row operation to A then

$$\text{row}(B) = \text{row}(A)$$

Proof Let $\mathbf{r}_1, \ldots, \mathbf{r}_m$ be the rows of A.

There are three elementary row operations. Clearly, switching the order of vectors in span($\mathbf{r}_1, \ldots, \mathbf{r}_m$) will not affect the span.

Suppose that B was obtained from A by multiplying the i^{th} row of A by a non-zero constant k. We need to show that

$$\text{span}(\mathbf{r}_1, \ldots, k\mathbf{r}_i, \ldots, \mathbf{r}_m) = \text{span}(\mathbf{r}_1, \ldots, \mathbf{r}_i, \ldots, \mathbf{r}_m)$$

To do this we will assume that some vector $\mathbf{v}$ is in $\text{span}(\mathbf{r}_1, \ldots, k\mathbf{r}_i, \ldots, \mathbf{r}_m)$, and show that $\mathbf{v}$ is in $\text{span}(\mathbf{r}_1, \ldots, \mathbf{r}_i, \ldots, \mathbf{r}_m)$. We will then assume that some vector $\mathbf{w}$ is in $\text{span}(\mathbf{r}_1, \ldots, \mathbf{r}_i, \ldots, \mathbf{r}_m)$ and show that $\mathbf{w}$ must be in $\text{span}(\mathbf{r}_1, \ldots, k\mathbf{r}_i, \ldots, \mathbf{r}_m)$.

Suppose that $\mathbf{v}$ is in $\text{span}(\mathbf{r}_1, \ldots, k\mathbf{r}_i, \ldots, \mathbf{r}_m)$. Then

$$\mathbf{v} = a_1\mathbf{r}_1 + \ldots + a_i(k\mathbf{r}_i) + \ldots + a_m\mathbf{r}_m$$

But then

$$\mathbf{v} = a_1\mathbf{r}_1 + \ldots + (a_ik)\mathbf{r}_i + \ldots + a_m\mathbf{r}_m$$

So $\mathbf{v}$ is in $\text{span}(\mathbf{r}_1, \ldots, \mathbf{r}_i, \ldots, \mathbf{r}_m)$.

Now suppose $\mathbf{w}$ is in $\text{span}(\mathbf{r}_1, \ldots, \mathbf{r}_i, \ldots, \mathbf{r}_m)$, then

$$\mathbf{w} = b_1\mathbf{r}_1 + \ldots + b_i\mathbf{r}_i + \ldots + b_m\mathbf{r}_m$$

But because $k \neq 0$, we can do the following:

$$\mathbf{w} = b_1\mathbf{r}_1 + \ldots + \frac{b_i}{k}(k\mathbf{r}_i) + \ldots + b_m\mathbf{r}_m$$

So $\mathbf{w}$ is in $\text{span}(\mathbf{r}_1, \ldots, k\mathbf{r}_i, \ldots, \mathbf{r}_m)$.

We leave it to the reader to verify that adding a multiple of one row of A to another does not change the row space.

■

Corollary 3. If matrix B was obtained from matrix A by applying a sequence of elementary row operations to A then

$$\text{row}(B) = \text{row}(A)$$

Proof This follows from repeated applications of Theorem **??**. ■

Corollary 4.

$$\text{row}(A) = \text{row}(\text{rref}(A))$$

Example 75. Let

$$A = \begin{bmatrix} 2 & -1 & 1 & -4 & 1 \\ 1 & 0 & 3 & 3 & 0 \\ -2 & 1 & -1 & 5 & 2 \\ 4 & -1 & 7 & 2 & 1 \end{bmatrix}$$

Find two distinct bases for row(A).

Explanation. By Corollary **??** a basis for row(rref(A)) will also be a basis for row(A). Row reduction gives us:

$$\begin{bmatrix} 2 & -1 & 1 & -4 & 1 \\ 1 & 0 & 3 & 3 & 0 \\ -2 & 1 & -1 & 5 & 2 \\ 4 & -1 & 7 & 2 & 1 \end{bmatrix} \rightsquigarrow \begin{bmatrix} 1 & 0 & 3 & 0 & -9 \\ 0 & 1 & 5 & 0 & -31 \\ 0 & 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} = \text{rref}(A)$$

Since the zero row contributes nothing to the span, we conclude that the nonzero rows of rref(A) span row(rref(A)). Therefore

$$\text{row}(A) = \text{span}\left(\begin{bmatrix} 1 & 0 & 3 & 0 & -9 \end{bmatrix}, \begin{bmatrix} 0 & 1 & 5 & 0 & -31 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 & 1 & 3 \end{bmatrix} \right)$$

By Problem **??**, the nonzero rows of rref(A) are linearly independent. It follows that the nonzero rows of rref(A) form a basis for row(A).

To find a second basis for row(A), observe that by Corollary **??** the row space of *any* row-echelon form of A will be equal to row(A). Matrix A has many row-echelon forms. Here is one of them:

$$B = \begin{bmatrix} 1 & 0 & 3 & 3 & 0 \\ 0 & -1 & -5 & -10 & 1 \\ 0 & 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

The nonzero rows of B span row(A). By Theorem **??**, the nonzero rows of B are linearly independent. Thus the nonzero rows of B form a basis for row(A).

Our observations in Example **??** can be generalized to all matrices. Given any matrix A,

1. The nonzero rows of rref(A) are linearly independent (Why?) and span row(A) (Corollary **??**).
2. The nonzero rows of any row-echelon form of A are linearly independent (Why?) and span row(A) (Corollary **??**).

Therefore nonzero rows of rref(A) or the nonzero rows of any row-echelon form of A constitute a basis of row(A). Since all bases for row(A) must have the same number of elements (Theorem **??**), we have just proved the following theorem.

Theorem 35. All row-echelon forms of a given matrix have the same number of nonzero rows.

This result was first introduced without proof in Gaussian Elimination and Rank where we used it to define the *rank* of a matrix as the number of nonzero rows in its row-echelon forms. We can now update the definition of rank as follows.

Definition 47. For any matrix A,

$$\text{rank}(A) = \dim\Big(\text{row}(A)\Big)$$

.

Column Space of a Matrix

Definition 48. Let A be an $m \times n$ matrix. The *column space* of A, denoted by col(A), is the subspace of $\mathbb{R}^m$ spanned by the columns of A.

Exploration 230 *Let*

$$B = \begin{bmatrix} 2 & -1 & 3 & 1 \\ 1 & -1 & 2 & 2 \\ 1 & 3 & -2 & -3 \end{bmatrix}$$

Our goal is to find a basis for col(B). To do this we need to find a linearly independent subset of the columns of B that spans col(B).

Consider the linear relation:

$$a_1 \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix} + a_2 \begin{bmatrix} -1 \\ -1 \\ 3 \end{bmatrix} + a_3 \begin{bmatrix} 3 \\ 2 \\ -2 \end{bmatrix} + a_4 \begin{bmatrix} 1 \\ 2 \\ -3 \end{bmatrix} = \mathbf{0} \tag{38}$$

Solving this equation amounts to finding rref(B).

$$\begin{bmatrix} 2 & -1 & 3 & 1 \\ 1 & -1 & 2 & 2 \\ 1 & 3 & -2 & -3 \end{bmatrix} \rightsquigarrow \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \textit{rref}(B)$$

*We see that (**??**) has infinitely many solutions. This means that the columns of B are linearly dependent, and contain a redundant vector. Which vector can we remove from the set without changing the span?*

Choosing a non-zero value for the free variable a_3, we can express the third column as a linear combination of the other columns. We conclude that

$$\textit{col}(B) = \textit{span}\left(\begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ -1 \\ 3 \end{bmatrix}, \begin{bmatrix} 3 \\ 2 \\ -2 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ -3 \end{bmatrix}\right) = \textit{span}\left(\begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ -1 \\ 3 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ -3 \end{bmatrix}\right)$$

Removing the third column also gives us

$$\begin{bmatrix} 2 & -1 & 3 & 1 \\ 1 & -1 & 2 & 2 \\ 1 & 3 & -2 & -3 \end{bmatrix} \rightsquigarrow \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

This implies that the remaining columns are linearly independent, and the set

$$\left\{ \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ -1 \\ 3 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ -3 \end{bmatrix} \right\}$$

is a basis for col(B).

The approach we took to find a basis for col(B) in Exploration **??** uses the reduced row-echelon form of B. It is true, however, that *any* row-echelon form of B could have been used in place of rref(B). (Why?). We generalize the steps as follows:

Procedure 2. Given a matrix B, a basis for col(B) can be found as follows:

1. Find rref(B) (or any row-echelon form B' of B.)
2. Identify the pivot columns of rref(B) (or B').
3. The columns of B corresponding to the pivot columns of rref(B) (or B') form a basis for col(B).

Proof Let $\mathbf{b}_1, \ldots, \mathbf{b}_n$ be the columns of B, and let $\mathbf{b}'_1, \ldots, \mathbf{b}'_n$ be the columns of rref(B) (or B'). Observe that the equations

$$a_1\mathbf{b}_1 + \ldots + a_n\mathbf{b}_n = \mathbf{0} \tag{39}$$

$$a_1\mathbf{b}'_1 + \ldots + a_n\mathbf{b}'_n = \mathbf{0} \tag{40}$$

have the same solution set. This means that any non-trivial relation among the columns of rref(B) (or B') translates into a non-trivial relation among the columns of B. Likewise, any collection of linearly independent columns of rref(B) (or B') corresponds to linearly independent columns of B.

By Problem **??**, the pivot columns of rref(B) (or B') are linearly independent. Therefore the corresponding columns of B are linearly independent. Non-pivot columns can be expressed as linear combinations of the pivot columns, therefore they contribute nothing to the span and can be removed from the spanning set. (See Problem **??**) ■

The proof of Procedure **??** shows that the number of basis elements for the column space of a matrix is equal to the number of pivot columns. But the number of pivot columns is the same as the number of pivots in a row-echelon form, which is equal to the number of nonzero rows and the rank of the matrix. This gives us the following important result.

Theorem 36. Let A be a matrix.

$$\dim\Big(\text{row}(A)\Big) = \text{rank}(A) = \dim\Big(\text{col}(A)\Big)$$

Example 76. We will return to matrix A of Example **??** and find a basis for col(A).

Explanation. We begin by finding rref(A).

$$\begin{bmatrix} 2 & -1 & 1 & -4 & 1 \\ 1 & 0 & 3 & 3 & 0 \\ -2 & 1 & -1 & 5 & 2 \\ 4 & -1 & 7 & 2 & 1 \end{bmatrix} \rightsquigarrow \begin{bmatrix} 1 & 0 & 3 & 0 & -9 \\ 0 & 1 & 5 & 0 & -31 \\ 0 & 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} = \text{rref}(A)$$

Columns 1, 2 and 4 of rref(A) contain leading $1's$. Therefore columns 1, 2 and 4 of A form a basis for col(A).

The Null Space

Definition 49. Let A be an $m \times n$ matrix. The *null space* of A, denoted by null(A), is the set of all vectors $\mathbf{x}$ in $\mathbb{R}^n$ such that $A\mathbf{x} = \mathbf{0}$.

Example 77. Find null(A) if

$$A = \begin{bmatrix} 3 & -1 \\ -6 & 2 \end{bmatrix}$$

Explanation. We need to solve the equation $A\mathbf{x} = \mathbf{0}$. Row reduction gives us

$$\begin{bmatrix} 3 & -1 \\ -6 & 2 \end{bmatrix} \rightsquigarrow \begin{bmatrix} 1 & -1/3 \\ 0 & 0 \end{bmatrix} = \text{rref}(A)$$

We conclude that $\mathbf{x} = \begin{bmatrix} 1/3 \\ 1 \end{bmatrix} t$. Thus null($A$) consists of all vectors of the form $\begin{bmatrix} 1/3 \\ 1 \end{bmatrix} t$. We might write

$$\text{null}(A) = \left\{ \begin{bmatrix} 1/3 \\ 1 \end{bmatrix} t \right\}$$

or

$$\text{null}(A) = \text{span}\left(\begin{bmatrix} 1/3 \\ 1 \end{bmatrix}\right)$$

Example **??** allows us to make an important observation. Note that every scalar multiple of $\begin{bmatrix} 1/3 \\ 1 \end{bmatrix}$ is contained in null(A). This means that null(A) is closed under vector addition and scalar multiplication. Recall that this property makes null(A) a *subspace* of $\mathbb{R}^n$. This result was first presented as Practice Problem **??**. We now formalize it as a theorem.

Theorem 37. Let A be an $m \times n$ matrix. Then null(A) is a subspace of $\mathbb{R}^n$.

Proof To show that null(A) is closed under vector addition and scalar multiplication we will show that a linear combination of any two elements of null(A) is contained in null(A).

Suppose $\mathbf{x}_1$ and $\mathbf{x}_2$ are in null(A). Then $A\mathbf{x}_1 = \mathbf{0}$ and $A\mathbf{x}_2 = \mathbf{0}$. But then

$$A(a_1\mathbf{x}_1 + a_2\mathbf{x}_2) = a_1A\mathbf{x}_1 + a_2A\mathbf{x}_2 = \mathbf{0}$$

We conclude that $a_1\mathbf{x}_1 + a_2\mathbf{x}_2$ is also in null(A). ■

Example 78. Find a basis for null(A), where A is the matrix in Example **??**.

Explanation. Elements in the null space of A are solutions to the equation

$$\begin{bmatrix} 2 & -1 & 1 & -4 & 1 \\ 1 & 0 & 3 & 3 & 0 \\ -2 & 1 & -1 & 5 & 2 \\ 4 & -1 & 7 & 2 & 1 \end{bmatrix} \mathbf{x} = \mathbf{0}$$

Row reduction yields rref(A)

$$\begin{bmatrix} 2 & -1 & 1 & -4 & 1 \\ 1 & 0 & 3 & 3 & 0 \\ -2 & 1 & -1 & 5 & 2 \\ 4 & -1 & 7 & 2 & 1 \end{bmatrix} \rightsquigarrow \begin{bmatrix} 1 & 0 & 3 & 0 & -9 \\ 0 & 1 & 5 & 0 & -31 \\ 0 & 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Therefore, elements of null(A) are of the form

$$\mathbf{x} = \begin{bmatrix} 9t - 3s \\ 31t - 5s \\ s \\ -3t \\ t \end{bmatrix} = \begin{bmatrix} -3 \\ -5 \\ 1 \\ 0 \\ 0 \end{bmatrix} s + \begin{bmatrix} 9 \\ 31 \\ 0 \\ -3 \\ 1 \end{bmatrix} t$$

Thus

$$\text{null}(A) = \text{span}\left(\begin{bmatrix} -3 \\ -5 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 9 \\ 31 \\ 0 \\ -3 \\ 1 \end{bmatrix}\right)$$

To find a basis for null(A) we need to find linearly independent vectors that span null(A). Take a closer look at the vectors

$$\begin{bmatrix} -3 \\ -5 \\ \boxed{1} \\ 0 \\ \boxed{0} \end{bmatrix}, \begin{bmatrix} 9 \\ 31 \\ \boxed{0} \\ -3 \\ \boxed{1} \end{bmatrix}$$

Because of the locations of $1's$ and $0's$, it is clear that one vector is not a scalar multiple of the other. Therefore the two vectors are linearly independent. We conclude that

$$\left\{\begin{bmatrix} -3 \\ -5 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 9 \\ 31 \\ 0 \\ -3 \\ 1 \end{bmatrix}\right\}$$

is a basis of null(A), and $\dim\Big(\text{null}(A)\Big) = 2$.

It is not a coincidence that the steps we used in Example **??** produced linearly independent vectors, and it is worth while to try to understand why this procedure will always produce linearly independent vectors.

Take a closer look at the elements of the null space:

$$\mathbf{x} = \begin{bmatrix} 9t - 3s \\ 31t - 5s \\ \boxed{s} \\ -3t \\ \boxed{t} \end{bmatrix} = \begin{bmatrix} -3 \\ -5 \\ \boxed{1} \\ 0 \\ \boxed{0} \end{bmatrix} s + \begin{bmatrix} 9 \\ 31 \\ \boxed{0} \\ -3 \\ \boxed{1} \end{bmatrix} t$$

The parameter s in the third component of $\mathbf{x}$ produces a 1 in the third component of the first vector and a 0 in the third component of the second vector, while parameter t in the fifth component of

x produces a 1 in the fifth component of the second vector and a 0 in the fifth component of the first vector. This makes it clear that the two vectors are linearly independent.

This pattern will hold for any number of parameters, each parameter producing a 1 in exactly one vector and $0's$ in the corresponding components of the other vectors.

$$\begin{bmatrix} \vdots \\ t_1 \\ \vdots \\ t_2 \\ \vdots \\ t_3 \\ \vdots \\ t_n \\ \vdots \end{bmatrix} = \begin{bmatrix} \vdots \\ 1 \\ \vdots \\ 0 \\ \vdots \\ 0 \\ \vdots \\ 0 \\ \vdots \end{bmatrix} t_1 + \ldots + \begin{bmatrix} \vdots \\ 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \\ \vdots \\ 0 \\ \vdots \end{bmatrix} t_2 + \ldots + \begin{bmatrix} \vdots \\ 0 \\ \vdots \\ 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \\ \vdots \end{bmatrix} t_3 + \ldots + \begin{bmatrix} \vdots \\ 0 \\ \vdots \\ 0 \\ \vdots \\ 0 \\ \vdots \\ 1 \\ \vdots \end{bmatrix} t_n$$

Therefore, vectors obtained in this way will always be linearly independent.

Rank and Nullity Theorem

Definition 50. Let A be a matrix. The dimension of the null space of A is called the *nullity* of A.

$$\dim\Big(\text{null}(A)\Big) = \text{nullity}(A)$$

We know that the dimension of the row space and the dimension of the column space of a matrix are the same and are equal to the rank of the matrix (or the number of nonzero rows in any row-echelon form of the matrix).

As we observed in Example **??**, the dimension of the null space of a matrix is equal to the number of free variables in the solution vector of the homogeneous system associated with the matrix. Since the number of pivots and the number of free variables add up to the number of columns in a matrix (Theorem **??**) we have the following significant result.

Theorem 38. Let A be an $m \times n$ matrix. Then

$$\text{rank}(A) + \text{nullity}(A) = n$$

We will see the geometric implications of this theorem when we study linear transformations.

Practice Problems

Problem 231 *Let A be a matrix. Prove that non-zero rows of rref(A) are linearly independent.*

Problem 232 *Let A be a matrix. Prove that the pivot columns of rref(A) are linearly independent.*

Problem 233 *Let A be a matrix. Show that the non-pivot columns of rref(A) can be expressed as linear combinations of the pivot columns.*

Problem 234 *Let*

$$A = \begin{bmatrix} 2 & 0 & 2 & 4 \\ 1 & 3 & -2 & -1 \\ -1 & -2 & 1 & 0 \end{bmatrix}$$

Question 234.1 *Find rref(A). (You may use technology.)*

$$rref(A) = \begin{bmatrix} \ldots\ldots & \ldots\ldots & \ldots\ldots & \ldots\ldots \\ \ldots\ldots & \ldots\ldots & \ldots\ldots & \ldots\ldots \\ \ldots\ldots & \ldots\ldots & \ldots\ldots & \ldots\ldots \end{bmatrix}$$

Question 234.2 *Complete the statement.*

$$rank(A) = dim(row(A)) = dim(col(A)) = \ldots\ldots$$

Question 234.3 *Use rref(A) and the procedure outlined in Example* **??** *to find a basis for row(A).*

Basis for row(A) : $\left\{ \begin{bmatrix} \ldots\ldots & \ldots\ldots & \ldots\ldots & \ldots\ldots \end{bmatrix}, \begin{bmatrix} \ldots\ldots & \ldots\ldots & \ldots\ldots & \ldots\ldots \end{bmatrix} \right\}$

Question 234.4 *Use Procedure* **??** *to find a basis for col(A).*

Basis for col(A) : $\left\{ \begin{bmatrix} \ldots\ldots \\ \ldots\ldots \\ \ldots\ldots \end{bmatrix}, \begin{bmatrix} \ldots\ldots \\ \ldots\ldots \\ \ldots\ldots \end{bmatrix} \right\}$

Question 234.5 *Find a basis for null(A).*

Basis for null(A) : $\left\{ \begin{bmatrix} \dots \dots \\ \dots \dots \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} \dots \dots \\ \dots \dots \\ 0 \\ 1 \end{bmatrix} \right\}$

Demonstrate that the Rank-Nullity Theorem (Theorem **??***) holds for A.*

Explain how you can quickly tell that the two vectors you selected for your basis are linearly independent.

Problem 235 *Let*

$$B = \begin{bmatrix} 1 & 2 & 3 \\ -1 & 1 & 3 \\ 2 & 0 & -2 \\ 1 & -2 & -5 \\ 0 & 1 & 2 \end{bmatrix}$$

Question 235.1 *Find rref(B). (You may use technology.)*

$$rref(B) = \begin{bmatrix} \dots \dots & \dots \dots & \dots \dots \\ \dots \dots & \dots \dots & \dots \dots \\ \dots \dots & \dots \dots & \dots \dots \\ \dots \dots & \dots \dots & \dots \dots \\ \dots \dots & \dots \dots & \dots \dots \end{bmatrix}$$

Question 235.2 *Complete the statement.*

$$rank(B) = dim(row(B)) = dim(col(B)) = \dots \dots$$

Question 235.3 *Use rref(B) and the procedure outlined in Example* **??** *to find a basis for row(B).*

Basis for row(B) : $\left\{ \begin{bmatrix} \dots \dots & \dots \dots & \dots \dots \end{bmatrix}, \begin{bmatrix} \dots \dots & \dots \dots & \dots \dots \end{bmatrix} \right\}$

Question 235.4 *Use Procedure* **??** *to find a basis for col(B).*

Basis for col(B) : $\left\{ \begin{bmatrix} \dots \dots \\ \dots \dots \\ \dots \dots \\ \dots \dots \\ \dots \dots \end{bmatrix}, \begin{bmatrix} \dots \dots \\ \dots \dots \\ \dots \dots \\ \dots \dots \\ \dots \dots \end{bmatrix} \right\}$

Question 235.5 *Find a basis for null(B).*

Basis for null(B) : $\left\{ \begin{bmatrix} \dots \dots \\ \dots \dots \\ 1 \end{bmatrix} \right\}$

Demonstrate that the Rank-Nullity Theorem (Theorem **??***) holds for B.*

Problem 236 *Suppose matrix M is such that*

$$rref(M) = \begin{bmatrix} 1 & 0 & 2 & 0 & 3 & 1 \\ 0 & 1 & -1 & 0 & 1 & -2 \\ 0 & 0 & 0 & 1 & -2 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Question 236.1 *Follow the process used in Example* **??** *to find a basis for null(M). Explain why the basis elements obtained in this way are linearly independent.*

Basis of null(M) : $\left\{ \begin{bmatrix} \dots \dots \\ \dots \dots \\ 1 \\ \dots \dots \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \dots \dots \\ \dots \dots \\ 0 \\ \dots \dots \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} \dots \dots \\ \dots \dots \\ 0 \\ \dots \dots \\ 0 \\ 1 \end{bmatrix} \right\}$

Question 236.2 *Let $\mathbf{v}_1, \dots, \mathbf{v}_6$ denote the columns of M. Express $\mathbf{v}_3$ as a linear combination of $\mathbf{v}_1$ and $\mathbf{v}_2$.*

Answer:

$$\mathbf{v}_3 = \dots \dots \mathbf{v}_1 + \dots \dots \mathbf{v}_2$$

Problem 237 *Suppose A is a 3×5 matrix. Which of the following statements could be true?*

Select All Correct Answers:

1. *dim(col(A)) = 5*
2. *dim(row(A)) = 3*
3. *dim(null(A)) = 1*
4. *dim(null(A)) = 2*

5. $dim(null(A)) = 3$

Problem 238 *Suppose A is a 7×3 matrix. Which of the following statements could be true?*

Select All Correct Answers:

1. $dim(col(A)) = 3$
2. $dim(row(A)) = 3$
3. $dim(row(A)) = 7$
4. $dim(null(A)) = 0$
5. $dim(null(A)) = 4$

Problem 239 *Complete the proof of Theorem* **??** *by showing that adding a scalar multiple of one row of a matrix to another row does not change the row space.*

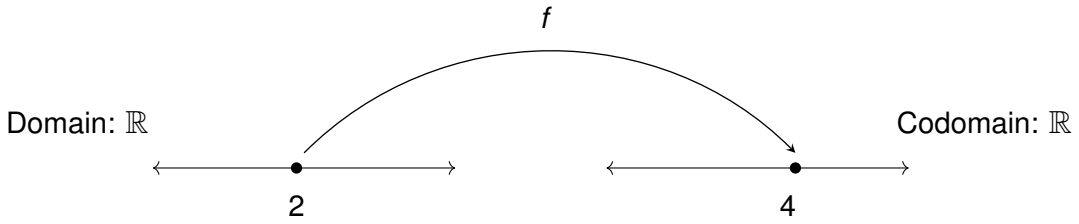
Problem 240 *Prove that $rank(A) = rank(A^T)$*

39 Linear Transformations

40 Matrix Transformations

Functions from $\mathbb{R}^n$ into $\mathbb{R}^m$

In the past you have worked with functions $f : \mathbb{R} \longrightarrow \mathbb{R}$. Most of the time such functions were defined algebraically. For example, we can define f by $f(x) = x^2$. This function takes a number in the *domain* ($\mathbb{R}$) and *maps* it to the square of the number in the *codomain* (also $\mathbb{R}$). Previously, you might have visualized function f by looking at its graph, the set of all points of the form $(x, f(x))$ in $\mathbb{R}^2$. In this course, we will find it more useful to look at functions diagrammatically. For instance, the diagram below shows that f maps 2 to 4. We say that 4 is the *image of* 2 *under* f.

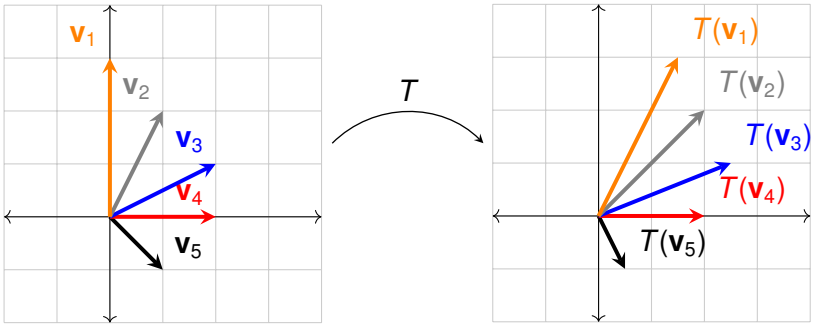


We will now consider functions that map $\mathbb{R}^n$ into $\mathbb{R}^m$. We will refer to such functions as *transformations*. There are two ways of thinking of transformations. A transformation $T : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ can take a vector in $\mathbb{R}^n$ and map it to a vector in $\mathbb{R}^m$, or it can map a point in $\mathbb{R}^n$ to a point in $\mathbb{R}^m$. We will think of transformations as acting on vectors or points interchangeably because every point $(x_1, x_2, \ldots, x_n)$ of $\mathbb{R}^n$ can be interpreted as the tip of a vector $\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$ in $\mathbb{R}^n$. Matrix multiplication will provide us with initial tools for defining some transformations.

Examples of Matrix Transformations

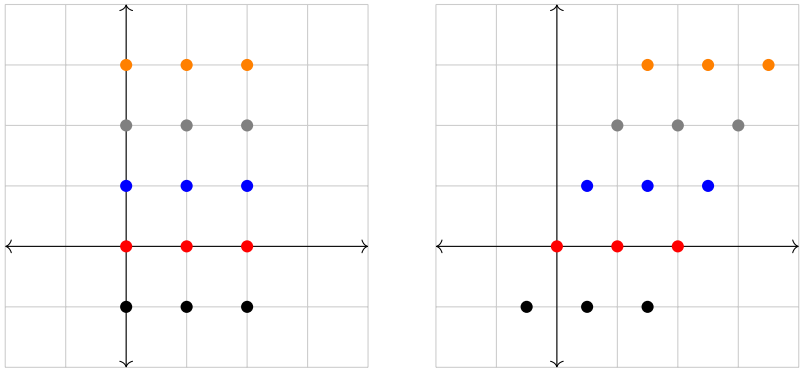
Consider the matrix $A = \begin{bmatrix} 1 & 0.5 \\ 0 & 1 \end{bmatrix}$. The product of A with a 2×1 vector is a 2×1 vector. We can define a transformation $T : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$ by $T(\mathbf{x}) = A\mathbf{x}$. This transformation can be applied to every

vector of $\mathbb{R}^2$. We will look at what it does to five vectors.



Vectors $\mathbf{v}_1, \ldots, \mathbf{v}_5$. Images of $\mathbf{v}_1, \ldots, \mathbf{v}_5$.

Even after looking at a handful of vectors it is often difficult to tell what the transformation actually accomplishes. This is why sometimes looking at points instead of vectors can be beneficial. If we consider every point in the left grid below as a tip of a vector, we can apply the transformation to each point to obtain the grid on the right.



Applying T to a grid of points helps us see that the entire plane was *sheared* by the transformation.

We can also analyze the action of T algebraically. Start by finding the image of a generic vector $\begin{bmatrix} x \\ y \end{bmatrix}$.

$$T\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} 1 & 0.5 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x + 0.5y \\ y \end{bmatrix}$$

We immediately see that the y component of the vector remains unchanged. We also see that the x component increases (or decreases) by an increment that depends on y. When considering T as a transformation acting on points, we see that points located 1 unit above the x-axis, get shifted to the right by 0.5. Points located 2 units above, get shifted to the right by 1. The higher the point, the greater the shift. Points with negative y-coordinates get shifted to the left. In this fashion T shears the entire plane.

Now that we have seen the effect of functions defined via matrix multiplication, we can better appreciate the term *transformation*, as such functions distort the domain and the shapes located in it. The following Exploration will help you visualize this.

Exploration 241 *Make your own shape by moving points A, B, C, D, E, F, G in the left pane. (You can also move the entire figure by clicking and dragging the whole polygon.) The images of the points and the polygon under the transformation induced by M are shown on the right.*

Access GeoGebra interactives through the online version of this text at

https://ximera.osu.edu/linearalgebradzv3.

Try each of the following matrices to determine what each transformation accomplishes. (Type pi into GeoGebra to get π.)

$$M_1 = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}, \quad M_2 = \begin{bmatrix} 1/2 & 0 \\ 0 & 1 \end{bmatrix}, \quad M_3 = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}$$

$$M_4 = \begin{bmatrix} \cos(\pi) & -\sin(\pi) \\ \sin(\pi) & \cos(\pi) \end{bmatrix}, \quad M_5 = \begin{bmatrix} \cos\left(\pi/4\right) & -\sin\left(\pi/4\right) \\ \sin\left(\pi/4\right) & \cos\left(\pi/4\right) \end{bmatrix}$$

$$M_6 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \quad M_7 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad M_8 = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

Match the description of each transformation with the matrix that induces it.

Horizontal Shear: (M_1/ M_2/ M_3/ M_4/ M_5/ M_6/ M_7/ M_8)

Rotation by 45° counterclockwise: (M_1/ M_2/ M_3/ M_4/ M_5/ M_6/ M_7/ M_8)

Reflection about the x-axis: (M_1/ M_2/ M_3/ M_4/ M_5/ M_6/ M_7/ M_8)

Vertical Stretch: (M_1/ M_2/ M_3/ M_4/ M_5/ M_6/ M_7/ M_8)

Maps everything to a straight line: (M_1/ M_2/ M_3/ M_4/ M_5/ M_6/ M_7/ M_8)

Rotation through a 180° angle: (M_1/ M_2/ M_3/ M_4/ M_5/ M_6/ M_7/ M_8)

Horizontal compression: (M_1/ M_2/ M_3/ M_4/ M_5/ M_6/ M_7/ M_8)

Reflection about the line $y = x$: (M_1/ M_2/ M_3/ M_4/ M_5/ M_6/ M_7/ M_8)

A 2×2 matrix induces a transformation from $\mathbb{R}^2$ into $\mathbb{R}^2$. An $m \times n$ matrix can be multiplied by an $n \times 1$ vector on the right, with the resulting product being an $m \times 1$ vector. Therefore we can use an $m \times n$ matrix A to define a transformation $T : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ by $T(\mathbf{x}) = A\mathbf{x}$.

Example 79. Let $A = \begin{bmatrix} 1 & 2 & -1 \\ 3 & 2 & -2 \end{bmatrix}$. Define a transformation $T : \mathbb{R}^3 \longrightarrow \mathbb{R}^2$ by $T(\mathbf{x}) = A\mathbf{x}$. Find all vectors in the domain that map to $\mathbf{0}$.

Explanation. We need to solve the system $A\mathbf{x} = \mathbf{0}$. We begin by forming an augmented matrix and finding its reduced row echelon form

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

There are infinitely many solutions

$$\mathbf{x} = \begin{bmatrix} 1/2 \\ 1/4 \\ 1 \end{bmatrix} t$$

This means that as T transforms $\mathbb{R}^3$ into $\mathbb{R}^2$, all points along the line $x = \frac{1}{2}t; y = \frac{1}{4}t; z = t$ map to the origin.

Linearity of Matrix Transformations

Restating properties **??** and **??** of Section Matrix Multiplication in terms of matrix-vector multiplication gives us

$$k(A\mathbf{v}) = A(k\mathbf{v}) \tag{41}$$

$$A(\mathbf{v} + \mathbf{w}) = A\mathbf{v} + A\mathbf{w} \tag{42}$$

These two properties of matrix multiplications translate into analogous properties of matrix transformations. Suppose $T : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ is a matrix transformation, then for all vectors $\mathbf{v}$, $\mathbf{w}$ in $\mathbb{R}^n$ and all constants k in $\mathbb{R}$,

$$T(k\mathbf{v}) = kT(\mathbf{v}) \tag{43}$$

$$T(\mathbf{v} + \mathbf{w}) = T(\mathbf{v}) + T(\mathbf{w}) \tag{44}$$

In general, any transformation that satisfies (**??**) and (**??**) is called a *linear transformation*. As we have just seen, all matrix transformations are linear. We will study linear transformations in depth later in this chapter.

Where did i Go?

In this section we will look at the images of standard unit vectors under a matrix transformation, and discuss why this information is helpful.

Exploration 242 *Let* $A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$*. Find the following products:*

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

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

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

Let $T : \mathbb{R}^3 \longrightarrow \mathbb{R}^3$ *be a matrix transformation induced by A, then we can say that T maps* **i**, **j** *and* **k** *to the first, second and third columns of A, respectively.*

This nice property is not limited to transformations induced by square matrices. Let $T : \mathbb{R}^3 \to \mathbb{R}^2$ *be a linear transformation induced by*

$$A = \begin{bmatrix} 1 & -2 & 4 \\ 0 & 3 & 5 \end{bmatrix}$$

We will examine the effect of T on the standard unit vectors **i**, **j** *and* **k**.

$$T(\mathbf{i}) = A\mathbf{i} = \begin{bmatrix} \dots \dots \\ \dots \dots \end{bmatrix}, \quad T(\mathbf{j}) = A\mathbf{j} = \begin{bmatrix} \dots \dots \\ \dots \dots \end{bmatrix}, \quad T(\mathbf{k}) = A\mathbf{k} = \begin{bmatrix} \dots \dots \\ \dots \dots \end{bmatrix}$$

Observe that the image of **i** *is the first column of A, the image of* **j** *is the second column of A, and the image of* **k** *is the third column.*

We formalize our findings in Exploration **??** as follows.

Observation 6. In general, the linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$, induced by an $m \times n$ matrix A maps the standard unit vectors $\mathbf{e}_1 \ldots \mathbf{e}_n$ to the columns of A. We summarize this observation by expressing columns of A as images of vectors $\mathbf{e}_1 \ldots \mathbf{e}_n$ under T.

$$A = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & \ldots & \ldots & a_{mn} \end{bmatrix} = \begin{bmatrix} | & | & & | \\ T(\mathbf{e}_1) & T(\mathbf{e}_2) & \ldots & T(\mathbf{e}_n) \\ | & | & & | \end{bmatrix}$$

Why is it that knowing the images of standard unit vectors under a matrix transformation is helpful? Consider the following example.

Example 80. Let $T : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$ be a matrix transformation such that $T(\mathbf{i}) = \begin{bmatrix} -2 \\ 3 \end{bmatrix}$ and $T(\mathbf{j}) = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$. Find $T\left(\begin{bmatrix} -4 \\ 10 \end{bmatrix}\right)$.

Explanation. We will make use of linearity of matrix transformations.

$$T\left(\begin{bmatrix} -4 \\ 10 \end{bmatrix}\right) = T(-4\mathbf{i} + 10\mathbf{j}) = -4T(\mathbf{i}) + 10T(\mathbf{j}) = -4\begin{bmatrix} -2 \\ 3 \end{bmatrix} + 10\begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 18 \\ -12 \end{bmatrix}$$

Example **??** illustrates that a matrix transformation $T : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ is *completely determined* by where it maps the standard unit vectors. This is true because we can express every vector $\mathbf{v}$ in $\mathbb{R}^n$ as a linear combination of the standard unit vectors, then use (**??**) and (**??**) to find the image of $\mathbf{v}$.

Practice Problems

Problem 243 *Let $T : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$ be a matrix transformation induced by the matrix A. The GeoGebra window on the left shows the domain of T, with standard unit vectors* **i** *and* **j***, and a vector* **x***. The window on the right shows the codomain of T, with the images of* **i***,* **j** *and* **x** *plotted. To use this interactive, you can*

- *Change the entries of matrix A;*
- *Change vector* **x** *by dragging its tip.*

Access GeoGebra interactives through the online version of this text at
https://ximera.osu.edu/linearalgebradzv3.
Choose your matrix A. Visually verify the following claims:

- *The image of* **i** *is the first column of matrix A.*
- *The image of* **j** *is the second column of matrix A.*

Let $\mathbf{x} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$*. Complete the following statement by filling the blanks.*

$$T(\mathbf{x}) = T(\ldots\ldots\mathbf{i} + \ldots\ldots\mathbf{j}) = \ldots\ldots T(\mathbf{i}) + \ldots\ldots T(\mathbf{j})$$

Choose a different matrix A, but keep vector **x** *the same. Does the above relationship still hold? Change vector* **x** *by dragging its tip. Observe the image of* **x** *and its relationship to the images of* **i** *and* **j***. Complete the following statement for all vectors* $\begin{bmatrix} a \\ b \end{bmatrix}$ *in* $\mathbb{R}^2$*.*

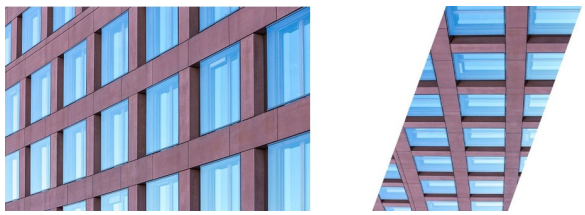
$$T\left(\begin{bmatrix} a \\ b \end{bmatrix}\right) = T(\ldots\ldots\mathbf{i} + \ldots\ldots\mathbf{j}) = \ldots\ldots T(\mathbf{i}) + \ldots\ldots T(\mathbf{j})$$

Problem 244 *Show that a matrix transformation* $T : \mathbb{R}^n \to \mathbb{R}^m$ *maps* $\mathbf{0}$ *to* $\mathbf{0}$*. In other words,* $T(\mathbf{0}) = \mathbf{0}$*.*

Problem 245 *Show that a matrix transformation* $T : \mathbb{R}^n \to \mathbb{R}^m$ *maps a line in* $\mathbb{R}^n$ *to a line (or the origin) in* $\mathbb{R}^m$*.*

41 Geometric Transformations of the Plane

In Exploration **??** we saw how geometric shapes in the plane can be manipulated with matrix transformations. In this section we will study matrix transformations of the plane in more detail by applying matrix transformations to pixels in photos. We will treat every pixel of a picture as a point or the head of a vector in $\mathbb{R}^2$. When a transformation is applied to each pixel, the output pixel is colored the same color as the input pixel. The figure below shows the result of one such transformation.

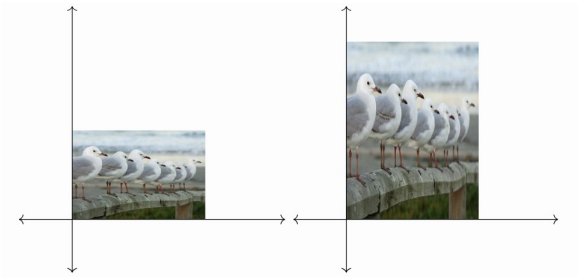


Not all transformations are matrix transformations. Practice Problem **??** offers a cautionary tale of what happens when we try to find a matrix for a transformation when such a matrix does not exist. Luckily, many familiar transformations, including rotations, scalings, reflections, and shears, are matrix transformations. We will focus our attention on those.

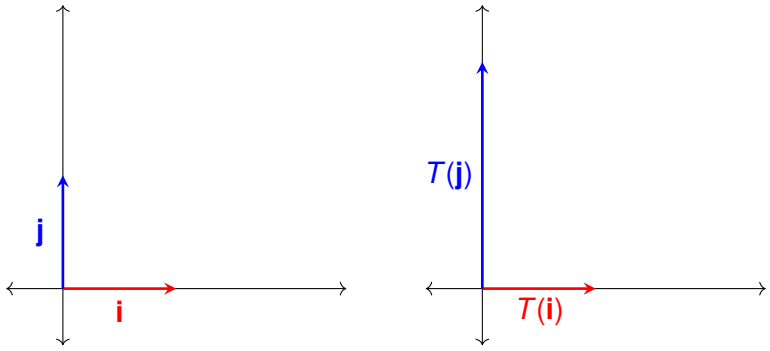
Recall that by Observation **??** of Matrix Transformations, we can find a matrix of a matrix transformation by examining what the transformation does to the standard basis vectors **i** and **j**.

Horizontal and Vertical Scaling

Exploration 246 *Let us attempt to find a matrix M for the transformation T that stretches an image vertically by a factor of 2, as shown in the figure below.*



Consider what this transformation does to the standard unit vectors. We observe that $T(\mathbf{i}) = \mathbf{i}$ *and* $T(\mathbf{j}) = 2\mathbf{j}$.



This allows us to construct a candidate for the transformation matrix M, by making the images of **i** *and* **j** *the columns of M. Thus,*

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

We can now check to see what this matrix does to an arbitrary point (a, b)*. Treating this point as a vector* $\begin{bmatrix} a \\ b \end{bmatrix}$*, we compute*

$$M = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} a \\ 2b \end{bmatrix}$$

Thus, this transformation takes point (a, b) *to point* $(a, 2b)$*. So, the proposed transformation doubles all y-coordinates resulting in a vertical stretch by a factor of 2.*

In general, a vertical stretch (or compression) leaves **i** unchanged, and scales the vector **j** while preserving its vertical direction. Thus, a vertical stretch (or compression) maps **i** to **i**, and maps **j** to k**j** for some positive number k. Similarly, a horizontal stretch (or compression) maps **i** to k**i**, and maps **j** to **j**.

Formula 11 (Horizontal and Vertical Scaling)**.** A matrix transformation induced by

$$M_v = \begin{bmatrix} 1 & 0 \\ 0 & k \end{bmatrix}, \tag{45}$$

where $(k > 0)$, scales objects in the plane vertically by a factor of k.

A matrix transformation induced by

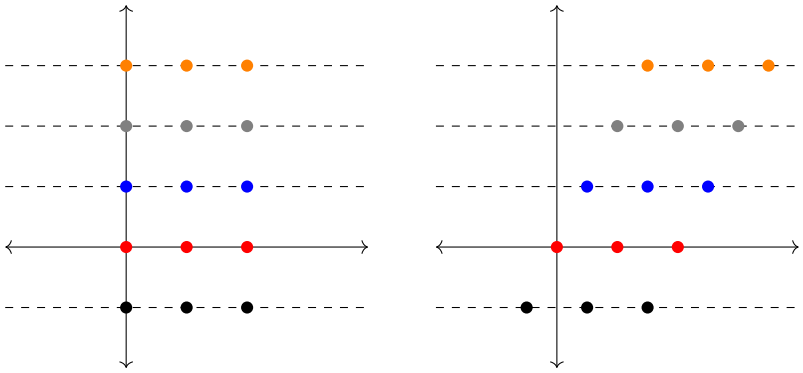
$$M_h = \begin{bmatrix} k & 0 \\ 0 & 1 \end{bmatrix}, \tag{46}$$

where $(k > 0)$, scales objects in the plane horizontally by a factor of k.

In stating the above formula we stipulated that $k > 0$. If we were to allow k to be zero, what would the resulting transformations accomplish? In what way would the resulting matrices be fundamentally different from matrices M_v and M_h? What would happen if k were allowed to be negative? (See Practice Problem **??**)

Horizontal and Vertical Shears

A *horizontal shear* is a transformation that takes an arbitrary point (a, b) and maps it to the point $(a + kb, b)$. The effect of this transformation is that all points along a fixed horizontal line slide to the left or to the right by a fixed amount. Note that the higher the point (a, b) is above the x-axis, the greater is the magnitude of kb, resulting in a greater amount of horizontal slide.



Adding a scalar multiple of the y component to the x component can be accomplished by matrix multiplication. Observe that

$$\begin{bmatrix} 1 & k \\ 0 & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} a + kb \\ b \end{bmatrix}$$

A *vertical shear* is a transformation that takes an arbitrary point (a, b) and maps it to the point $(a, b + ka)$. This too, can be accomplished by matrix multiplication.

$$\begin{bmatrix} 1 & 0 \\ k & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} a \\ b + ka \end{bmatrix}$$

Formula 12 (Horizontal and Vertical Shears)**.** A matrix transformation induced by

$$M_{hs} = \begin{bmatrix} 1 & k \\ 0 & 1 \end{bmatrix} \tag{47}$$

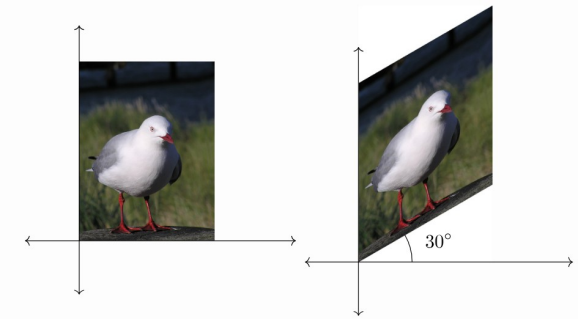
shears the plane horizontally.

A matrix transformation induced by

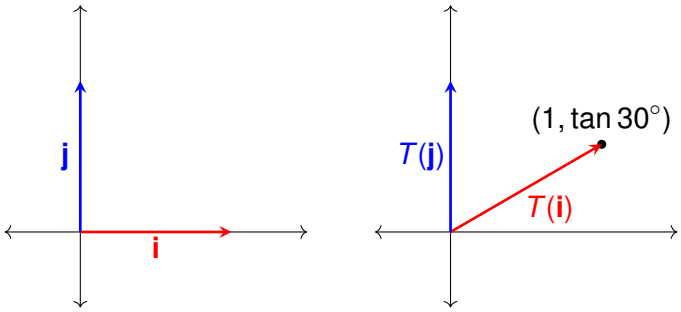
$$M_{vs} = \begin{bmatrix} 1 & 0 \\ k & 1 \end{bmatrix} \tag{48}$$

shears the plane vertically.

Example 81. Find the matrix M_{vs} of a matrix transformation T that *shears* the image of a seagull, as shown in the figure below.



Explanation. Consider what this transformation does to the standard unit vectors.



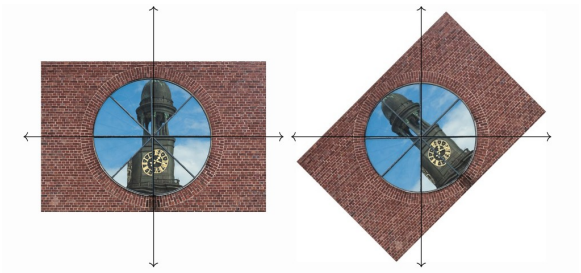
The tip of the vector **i** slides up a vertical line and its x-component remains the same. Vector **j** stays fixed. We observe that $T(\mathbf{i}) = \begin{bmatrix} 1 \\ \tan 30^\circ \end{bmatrix} = \begin{bmatrix} 1 \\ \sqrt{3}/3 \end{bmatrix}$ and $T(\mathbf{j}) = \mathbf{j}$. This allows us to construct M_{vs}, by making the images of **i** and **j** the columns of M_{vs}. Thus,

$$M_{vs} = \begin{bmatrix} 1 & 0 \\ \sqrt{3}/3 & 1 \end{bmatrix}$$

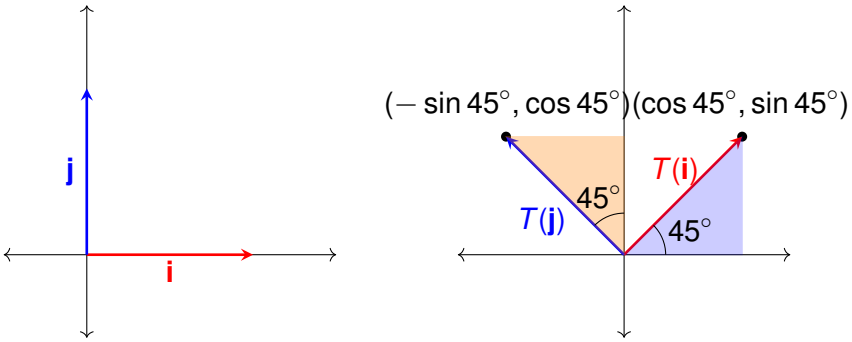
Rotations about the Origin

It turns out that rotations about the origin are also matrix transformations. You will have an opportunity to revisit and prove this claim in Challenge Problems for Chapter 6.

Example 82. Let $R_{45^\circ} : \mathbb{R}^2 \to \mathbb{R}^2$ be a transformation that rotates $\mathbb{R}^2$ countercolockwise about the origin by 45°. Find a matrix that induces R_{45°.



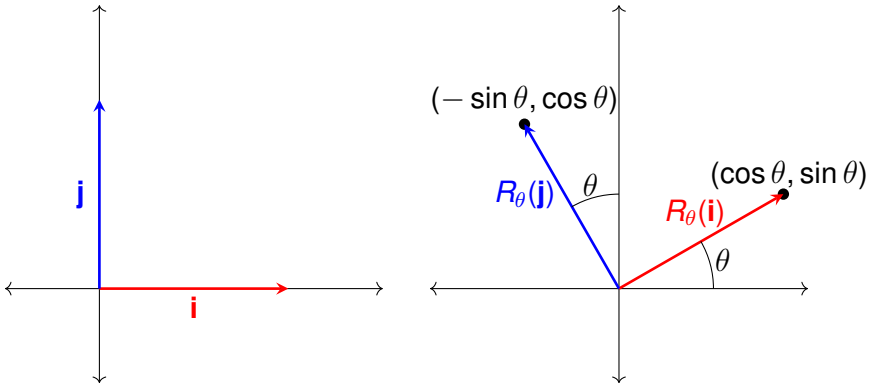
Explanation. Consider the action of R_θ on the standard unit vectors.



We observe that $R_\theta(\mathbf{i}) = \begin{bmatrix} \cos 45^\circ \\ \sin 45^\circ \end{bmatrix} = \begin{bmatrix} \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{bmatrix}$ and $R_\theta(\mathbf{j}) = \begin{bmatrix} -\sin 45^\circ \\ \cos 45^\circ \end{bmatrix} = \begin{bmatrix} -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{bmatrix}$. This allows us to construct the matrix M_θ, by making the images of $\mathbf{i}$ and $\mathbf{j}$ the columns of M_θ. Thus,

$$M_\theta = \begin{bmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{bmatrix}$$

In general, we find the rotation matrix by determining the images of vectors $\mathbf{i}$ and $\mathbf{j}$.



Formula 13 (Counterclockwise Rotation)**.** A matrix transformation induced by

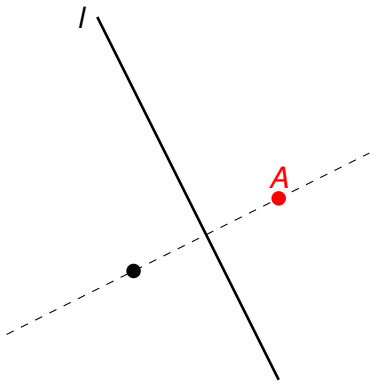
$$M_\theta = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \tag{49}$$

rotates the plane counterclockwise through angle θ about the origin.

Reflections about Lines of the Form $y = mx$

When a point is reflected about a line, its image is located on the opposite side of the line and the same distance away from the line as the original point. This is another example of a matrix transformation. (We will prove that this is a matrix transformation in Challenge Problems for Chapter 6.)

For example, the figure below shows the reflection of point *A* about line *l*. Note that the reflection lies on a line through *A* perpendicular to *l*.



Our task is to find the matrix of a reflection of the plane about an arbitrary line through the origin.

Exploration 247 *In this problem we will find the matrix for the reflection about the x and y axes. You can easily do this on your own by finding the images of vectors* **i** *and* **j**.

We will start with the reflection about the x-axis.

$$\mathbf{i} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad \textit{maps to} \quad \begin{bmatrix} \ldots & \ldots \\ \ldots & \ldots \end{bmatrix}$$

$$\mathbf{j} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad \textit{maps to} \quad \begin{bmatrix} \ldots & \ldots \\ \ldots & \ldots \end{bmatrix}$$

So, the matrix that induces the reflection about the x-axis is

$$\begin{bmatrix} \ldots & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots \end{bmatrix}$$

Next, we will consider the reflection about the y-axis.

$$\mathbf{i} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad \textit{maps to} \quad \begin{bmatrix} \ldots & \ldots \\ \ldots & \ldots \end{bmatrix}$$

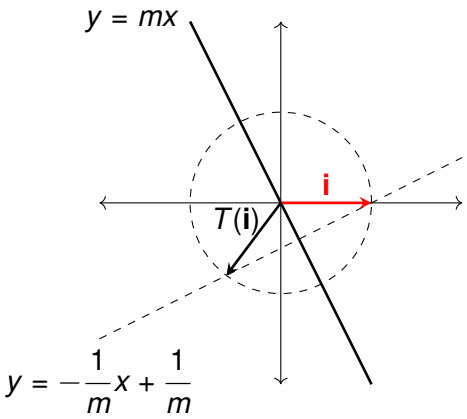
$$\mathbf{j} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad \textit{maps to} \quad \begin{bmatrix} \ldots & \ldots \\ \ldots & \ldots \end{bmatrix}$$

Thus, the matrix that induces the reflection about the y-axis is

$$\begin{bmatrix} \ldots & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots \end{bmatrix}$$

Now we will turn our attention to transformations that reflect the plane about the line $y = mx$. We will assume that $m \neq 0$.

Consider the vector **i** and its reflection.



Observe that the head of the image vector, $T(\mathbf{i})$, will lie on the line that passes through $(1, 0)$ and is perpendicular to the line $y = mx$. The equation of this line is given by

$$y = -\frac{1}{m}x + \frac{1}{m} \tag{50}$$

The head of $T(\mathbf{i})$ will also lie on the circle with equation

$$x^2 + y^2 = 1$$

To find the image of $\mathbf{i}$ we need to determine where the line $y = -\frac{1}{m}x + \frac{1}{m}$ intersects the circle. Substitution gives us

$$x^2 + \left(-\frac{1}{m}x + \frac{1}{m}\right)^2 = 1$$

After a little algebra we get

$$\left(1 + \frac{1}{m^2}\right)x^2 - \left(\frac{2}{m^2}\right)x + \left(\frac{1}{m^2} - 1\right) = 0$$

The quadratic formula yields

$$x = 1 \quad \text{and} \quad x = \frac{1 - m^2}{m^2 + 1}$$

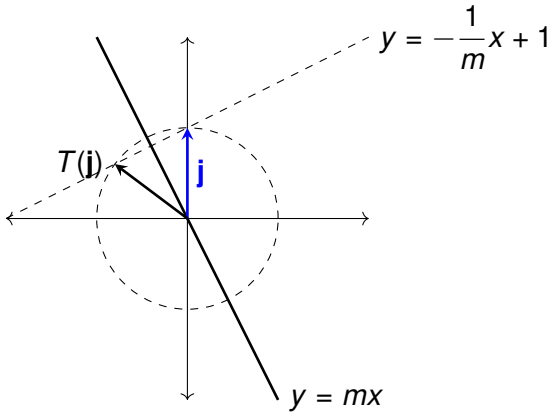
The solution $x = 1$ corresponds to the head of the vector $\mathbf{i}$. So, the x-component of $T(\mathbf{i})$ is $x = \frac{1 - m^2}{m^2 + 1}$. We find the y-component of $T(\mathbf{i})$ by substituting $x = \frac{1 - m^2}{m^2 + 1}$ into Equation **??**.

$$y = -\frac{1}{m}\left(\frac{1 - m^2}{m^2 + 1}\right) + \frac{1}{m} = \frac{2m}{m^2 + 1}$$

Thus, the image of **i** under this reflection is given by

$$T(\mathbf{i}) = \begin{bmatrix} \dfrac{1-m^2}{m^2+1} \\ \dfrac{2m}{m^2+1} \end{bmatrix}$$

Next we need to find the image of **j**. The head of $T(\mathbf{j})$ is located at one of the intersections of line $y = -\dfrac{1}{m}x + 1$ and the circle $x^2 + y^2 = 1$.



We leave it to the reader to verify that

$$T(\mathbf{j}) = \begin{bmatrix} \dfrac{2m}{m^2+1} \\ \dfrac{m^2-1}{m^2+1} \end{bmatrix} \tag{51}$$

This reflection is induced by the matrix

$$M_{y=mx} = \begin{bmatrix} \dfrac{1-m^2}{m^2+1} & \dfrac{2m}{m^2+1} \\ \dfrac{2m}{m^2+1} & \dfrac{m^2-1}{m^2+1} \end{bmatrix} = \frac{1}{1+m^2}\begin{bmatrix} 1-m^2 & 2m \\ 2m & m^2-1 \end{bmatrix}$$

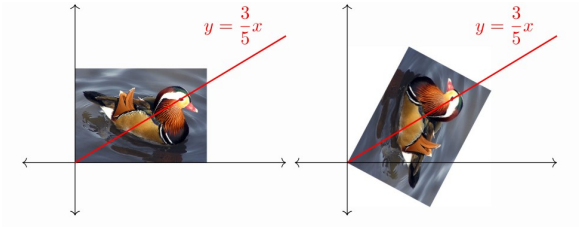
Formula 14 (Reflection about the line $y = mx$)**.** A matrix transformation induced by

$$M_{y=mx} = \frac{1}{1+m^2}\begin{bmatrix} 1-m^2 & 2m \\ 2m & m^2-1 \end{bmatrix} \tag{52}$$

reflects the plane about the line $y = mx$.

Note that when $m = 0$, expression (**??**) is consistent with the reflection matrix you found in Exploration **??**.

Example 83. Find matrix M that reflects the image of the duck about the line $y = \frac{3}{5}x$.



Explanation.

$$M = \frac{1}{1+9/25}\begin{bmatrix} 1-9/25 & 6/5 \\ 6/5 & 9/25-1 \end{bmatrix} = \frac{25}{34}\begin{bmatrix} 16/25 & 6/5 \\ 6/5 & -16/25 \end{bmatrix} = \begin{bmatrix} 8/17 & 15/17 \\ 15/17 & -8/17 \end{bmatrix}$$

Note that the eye of the duck in Example **??** is located on the line $y = \frac{3}{5}x$. The reflection leaves the eye fixed in place. The eye is an example of a *fixed point*. In Practice Problem **??** you will be asked to show that every point along the line $y = \frac{3}{5}x$ is a fixed point.

Composition of Linear Transformations

If a matrix transformation is followed by another matrix transformation, the resulting transformation can be represented as a product of the two matrices that induce the individual transformations. Thus, if T_1 is induced by M_1 and T_2 is induced by M_2, then $T = T_2 \circ T_1$ is induced by $M = M_2M_1$.

$$\mathbb{R}^2 \xrightarrow{T_1} \mathbb{R}^2 \xrightarrow{T_2} \mathbb{R}^2 \qquad (T)$$

Remember that matrix multiplication is not commutative, so the order in which the matrices are multiplied is of utmost importance.

Consider the following example which incorporates a reflection as well as a rotation of vectors.

Example 84. 1. Find the matrix of the linear transformation which is obtained by first rotating all vectors counterclockwise by an angle of $\pi/6$ and then reflecting across the x-axis.

2. Multiply the vector $\mathbf{v} = \begin{bmatrix} 1/2 \\ \sqrt{3}/2 \end{bmatrix}$ by the matrix you found in part (a) to illustrate how the transformation works.

Explanation. 1. Using Formula **??**, the matrix of the transformation which involves rotating through an angle of $\pi/6$ is

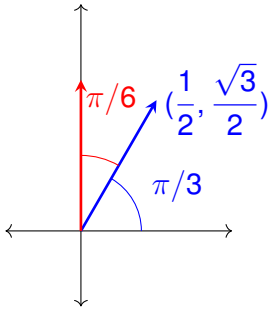
$$M_1 = \begin{bmatrix} \cos(\pi/6) & -\sin(\pi/6) \\ \sin(\pi/6) & \cos(\pi/6) \end{bmatrix} = \begin{bmatrix} \frac{1}{2}\sqrt{3} & -\frac{1}{2} \\ \frac{1}{2} & \frac{1}{2}\sqrt{3} \end{bmatrix}$$

As we learned in Exploration **??**, the matrix for the transformation which reflects all vectors about the x-axis is $M_2 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$.

Therefore, the matrix of the linear transformation which first rotates through $\pi/6$ and then reflects through the x-axis is given by

$$M = M_2 M_1 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} \frac{1}{2}\sqrt{3} & -\frac{1}{2} \\ \frac{1}{2} & \frac{1}{2}\sqrt{3} \end{bmatrix} = \begin{bmatrix} \frac{1}{2}\sqrt{3} & -\frac{1}{2} \\ -\frac{1}{2} & -\frac{1}{2}\sqrt{3} \end{bmatrix}$$

2. From unit circle trigonometry, we know that $\mathbf{v}$ makes a $\pi/3$ angle with the x-axis. If we rotate $\mathbf{v}$ counterclockwise by $\pi/6$, it will point straight up.



Following the rotation with reflection across the x-axis will make it point straight down. Sure enough,

$$M\mathbf{v} = \begin{bmatrix} \frac{1}{2}\sqrt{3} & -\frac{1}{2} \\ -\frac{1}{2} & -\frac{1}{2}\sqrt{3} \end{bmatrix} \begin{bmatrix} 1/2 \\ \sqrt{3}/2 \end{bmatrix} = \begin{bmatrix} 0 \\ -1 \end{bmatrix}.$$

Exploration 248 *Use the GeoGebra interactive below to decompose a matrix into a product of two matrices corresponding to the basic transformations we discussed above: scalings, rotations, shears and reflections.*

Consider the matrix $M = \begin{bmatrix} 1 & 0 \\ -1 & -1 \end{bmatrix}$. *This matrix induces a transformation that can be broken into two parts: (1) a reflection followed by (2) a shear. Find matrices* M_1 *and* M_2 *that induce the reflection and the shear respectively. Verify that the product of the two matrices is equal to M (be careful about the order of the product!).*

$$M = M_2 M_1 = \begin{bmatrix} \ldots & \ldots \\ \ldots & \ldots \end{bmatrix} \begin{bmatrix} \ldots & \ldots \\ \ldots & \ldots \end{bmatrix}$$

Access GeoGebra interactives through the online version of this text at

https://ximera.osu.edu/linearalgebradzv3/LinearAlgebraInteractiveIntro.

Let $M = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}$. *Use the GeoGebra interactive above to visually examine the transformation induced by M. What sequence of transformations is equivalent to the transformation induced by M?*

Multiple Choice:

1. *Rotation by 180 degrees.*
2. *Reflection about the x-axis, followed by a reflection about the y-axis.*
3. *Reflection about the y-axis, followed by a reflection about the x-axis.*
4. *All of the above.*

Let $M = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$. *Use the GeoGebra interactive above to visually examine the transformation induced by M. What sequence of transformations is equivalent to the transformation induced by M?*

Multiple Choice:

1. *Reflection about the x-axis, followed by a reflection about the line* $y = -x$.
2. *Reflection about the y-axis, followed by a reflection about the x-axis.*

3. *All of the above.*

4. *None of the above.*

Example 85. In this problem we will consider compositions of two reflections and use geometry to illustrate non-commutativity of matrix multiplication. Let

$$T_{y=-2x} : \mathbb{R}^2 \to \mathbb{R}^2$$

be a reflection about the line $y = -2x$. Let

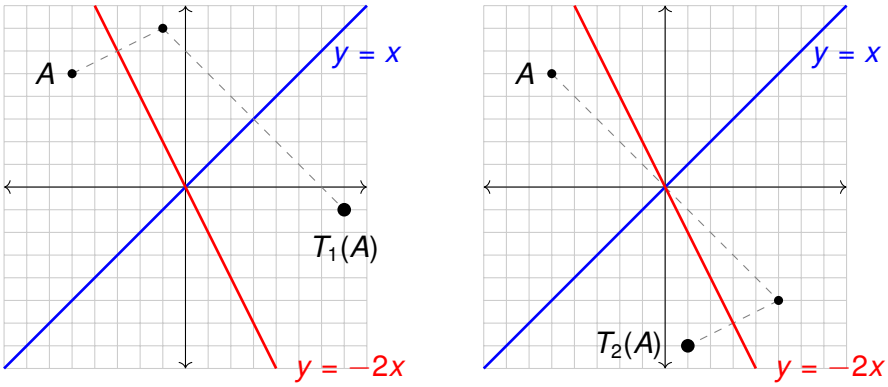
$$T_{y=x} : \mathbb{R}^2 \to \mathbb{R}^2$$

be a reflection about the line $y = x$. We will denote the standard matrices for these transformations by $M_{y=-2x}$ and $M_{y=x}$, and use geometry to demonstrate that $M_{y=x}M_{y=-2x} \neq M_{y=-2x}M_{y=x}$.

To do this, consider transformations $T_1 = T_{y=x} \circ T_{y=-2x}$ and $T_2 = T_{y=-2x} \circ T_{y=x}$. Transformation T_1 is induced by $M_{y=x}M_{y=-2x}$, and T_2 is induced by $M_{y=-2x}M_{y=x}$.

The figure on the left illustrates the action of T_1 on a single point A. First, A is reflected about the line $y = -2x$, then A is reflected about the line $y = x$.

The figure on the right shows the action of T_2 on the same point A. The point is first reflected about the line $y = x$, followed by a reflection about the line $y = -2x$. The final images of point A under T_1 and T_2 are clearly different.



Since $T_1 \neq T_2$, we conclude that $M_{y=x}M_{y=-2x} \neq M_{y=-2x}M_{y=x}$.

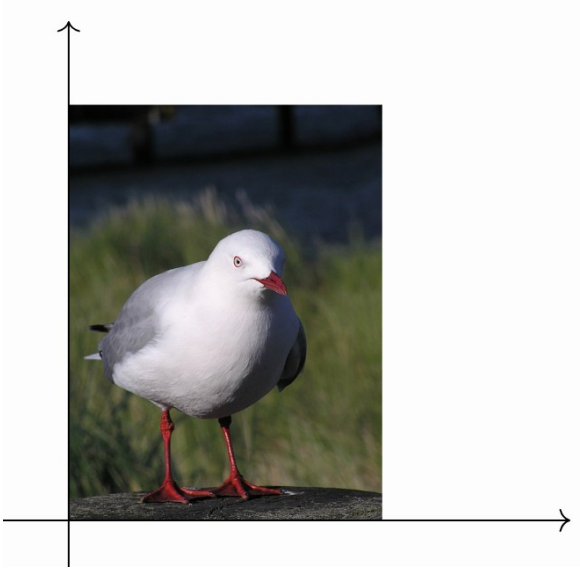
Example 86. Some pairs of matrices do commute. For example, geometry makes it is easy to see that two rotation matrices commute.

Practice Problems

Problem 249 *Consider matrices M_v and M_h in (**??**) and (**??**).*

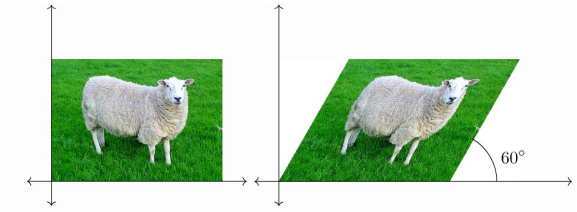
1. *If we were to allow k to be zero, what would the resulting transformations accomplish?*
2. *If $k = 0$, in what way would the resulting matrices be fundamentally different from matrices M_v and M_h ($k \neq 0$)?*
3. *Do M_v and M_h ($k \neq 0$) have inverses? What about M_v and M_h ($k = 0$)?*
4. *What would happen if we allowed k to be negative?*

Problem 250 *Find a matrix M that would double the length of a photo horizontally, and triple the height of the photo.*



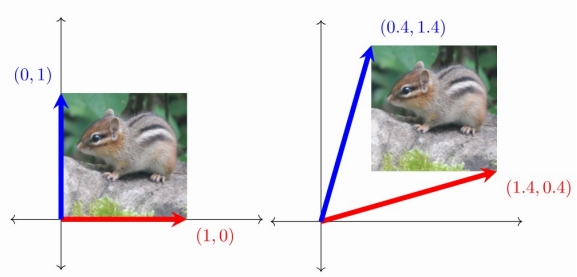
$$M = \begin{bmatrix} \ldots & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots \end{bmatrix}$$

Problem 251 *(Sheared Sheep) Find a matrix that induces the transformation shown in the figure.*



$$M = \begin{bmatrix} \cdots & \cdots \\ \cdots & \cdots \end{bmatrix}$$

Problem 252 *Suppose a 1 by 1 photo of a chipmunk was shifted as shown in the figure.*



Suppose we tried to construct a standard matrix M for this transformation by making the images of **i** *and* **j** *the columns of M. We would obtain*

$$M = \begin{bmatrix} 1.4 & 0.4 \\ 0.4 & 1.4 \end{bmatrix}$$

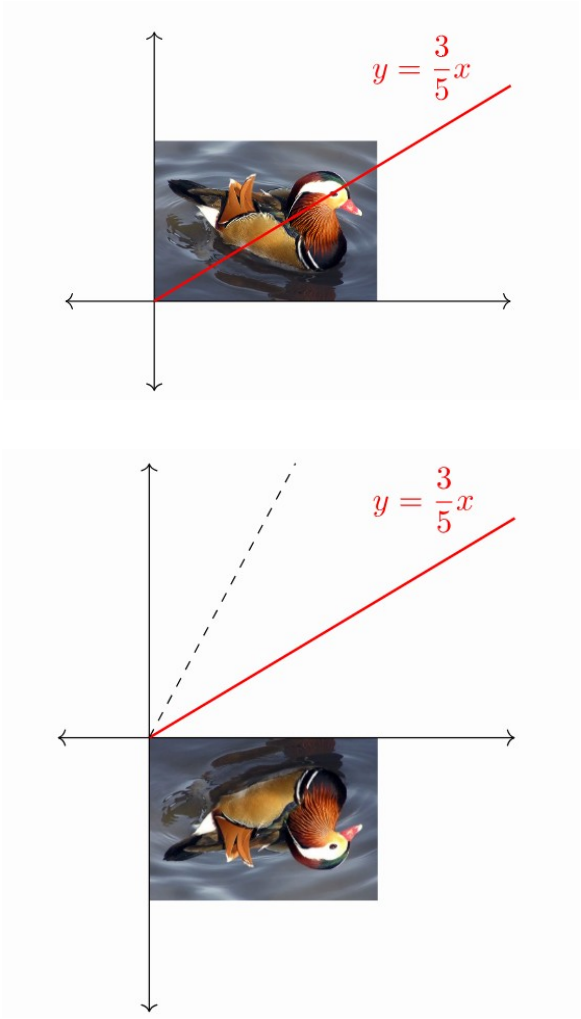
Does this matrix describe the transformation? If so, prove it. If not, explain why not.

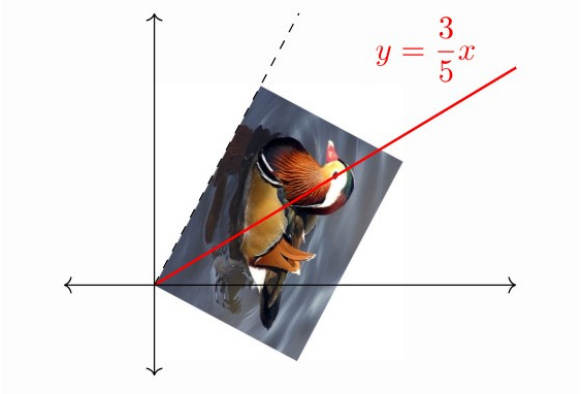
Problem 253 *A transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$ that shifts all points in the plane horizontally or vertically by a fixed amount is called a translation. Is T a matrix transformation? Prove your claim.*

Problem 254 *A reflection about the line $y = mx$ followed by another reflection about the same line, returns all points to their original position. Prove this using matrix multiplication.*

Problem 255 *Prove that every point along the line $y = \frac{3}{5}x$ in Example* **??** *is a fixed point.*

Problem 256 *The figure below shows a sequence of two matrix transformations that accomplishes a reflection about the line* $y = \frac{3}{5}x$. *The first transformation is a reflection of the plane about the* x*-axis. The second transformation is a rotation of the plane about the origin. Find matrices that induce the two transformations and verify that their product (in the correct order) is the reflection matrix of Example* **??**.





Example Source

Example **??** was adapted from Example 5.27 of Ken Kuttler's *A First Course in Linear Algebra*. (CC-BY)

Ken Kuttler, *A First Course in Linear Algebra*, Lyryx 2017, Open Edition, p. 290.

Photo Credits

The following images are courtesy of Wikimedia Commons

Adrian Pingstone, *A male Mandarin Duck at Slimbridge Wildfowl and Wetlands Centre, Gloucestershire, England.* Public Domain

Ansgar Koreng, *Facade of ARD-Hauptstadtstudio in Berlin-Mitte.* CC-BY 4.0

Christoph Braun, *Reflection of St. Michaelis Church in a window of St. Ansgar in Hamburg, Germany.* Public Domain.

Daniel Gammert, *Red-billed Gulls Chroicocephalus novaehollandiae scopulinus. Brighton Beach, New Zealand.* Public Domain

I, Tony Wills, *Red billed gull in Wellington Harbour, Wellington, New Zealand.* CC-BY

Jackhynes, *Lleyn sheep taken with a Sony Digital Camera at 3.2 megapixels in Devon, UK.* Public Domain

Erik Davis, *Chipmunk.* CC-BY

42 Introduction to Linear Transformations

We start by reviewing the definition of a function.

Definition 51. Let V and W be sets. A *function* f from V into W, denoted by

$$f : V \to W$$

assigns to each element x of V, a single element $y = f(x)$ of W.

The set V is called the *domain* of f, and the set W is called the *codomain*.

If $y = f(x)$, we say that x *maps to* y, and y is the *image* of x.

The collection of images of all points of V is called the *image* of V under f, or the *image* of f. (It is also known as the *range* of f.)

In algebra and calculus you worked with functions $f : \mathbb{R} \to \mathbb{R}$ whose domain and codomain were each the set of all real numbers. In linear algebra, we call our functions *transformations*. The domain and codomain of a transformation are *vector spaces*.

Exploration 257 *In this exercise we will introduce a very special type of transformation by contrasting the effects of two transformations on vectors of* $\mathbb{R}^2$*. We will see that some transformations have "nice" properties, while others do not. Define* T_1 *and* T_2 *as follows:*

$$T_1 : \mathbb{R}^2 \to \mathbb{R}^2$$

$$T_1\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} x - y \\ x \end{bmatrix}$$

$$T_2 : \mathbb{R}^2 \to \mathbb{R}^2$$

$$T_2\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} -x + y + 1 \\ y - 2 \end{bmatrix}$$

Each of these transformations takes a vector in $\mathbb{R}^2$*, and maps it to another vector in* $\mathbb{R}^2$*. To see if you understand how these transformations are defined, see if you can determine what these transformations do to the vector* $\begin{bmatrix} 4 \\ 3 \end{bmatrix}$.

$$T_1\left(\begin{bmatrix}4\\3\end{bmatrix}\right) = \begin{bmatrix}\cdots\ \cdots\\ \cdots\ \cdots\end{bmatrix} \quad \text{and} \quad T_2\left(\begin{bmatrix}4\\3\end{bmatrix}\right) = \begin{bmatrix}\cdots\ \cdots\\ \cdots\ \cdots\end{bmatrix}.$$

Now, let's take the vector $\begin{bmatrix}4\\3\end{bmatrix}$ *and multiply it by a scalar, say* 7*.*

$$7\begin{bmatrix}4\\3\end{bmatrix} = \begin{bmatrix}28\\21\end{bmatrix}$$

.

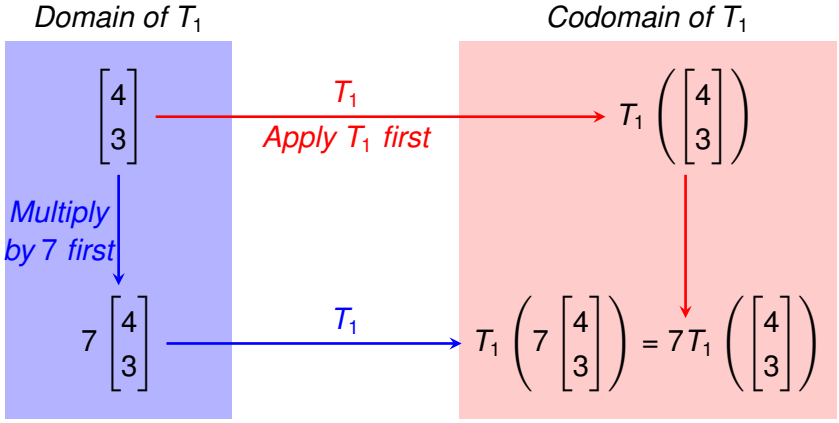
Now let's compare how T_1 *and* T_2 *"handle" this product. Starting with* T_1*, we compute:*

$$T_1\left(7\begin{bmatrix}4\\3\end{bmatrix}\right) = T_1\left(\begin{bmatrix}28\\21\end{bmatrix}\right) = \begin{bmatrix}7\\28\end{bmatrix}$$

Observe that multiplying the original vector by 7*, then applying* T_1*, has the same effect as applying* T_1 *to the original vector, then multiplying the image by* 7*. In other words,*

$$T_1\left(7\begin{bmatrix}4\\3\end{bmatrix}\right) = \begin{bmatrix}7\\28\end{bmatrix} = 7\begin{bmatrix}1\\4\end{bmatrix} = 7T_1\left(\begin{bmatrix}4\\3\end{bmatrix}\right)$$

Diagrammatically, this can be represented as follows.



You should try to verify that this property does not hold for transformation T_2*. In other words,*

$$T_2\left(7\begin{bmatrix}4\\3\end{bmatrix}\right) \neq 7T_2\left(\begin{bmatrix}4\\3\end{bmatrix}\right)$$

There is nothing special about the number 7*, and it is not hard to prove that for any scalar* k *and vector* $\mathbf{u}$ *of* $\mathbb{R}^2$*,* T_1 *satisfies*

$$kT_1(\mathbf{u}) = T_1(k\mathbf{u}). \tag{53}$$

It turns out that T_1 *satisfies another important property. For all vectors* $\mathbf{u}$ *and* $\mathbf{v}$ *of* $\mathbb{R}^2$ *we have:*

$$T_1(\mathbf{u}+\mathbf{v}) = T_1(\mathbf{u}) + T_1(\mathbf{v}) \tag{54}$$

We leave it to the reader to illustrate this property with a specific example (see Practice Problem **??***). We will show that* T_1 *satisfies (***??***) in general.*

Let $\mathbf{u} = \begin{bmatrix}u_1\\u_2\end{bmatrix}$ *and* $\mathbf{v} = \begin{bmatrix}v_1\\v_2\end{bmatrix}$*, then*

$$\begin{aligned} T_1(\mathbf{u}+\mathbf{v}) &= T_1\left(\begin{bmatrix}u_1\\u_2\end{bmatrix} + \begin{bmatrix}v_1\\v_2\end{bmatrix}\right) = T_1\left(\begin{bmatrix}u_1+v_1\\u_2+v_2\end{bmatrix}\right) = \begin{bmatrix}u_1+v_1-u_2-v_2\\u_1+v_1\end{bmatrix} \\ &= \begin{bmatrix}u_1-u_2\\u_1\end{bmatrix} + \begin{bmatrix}v_1-v_2\\v_1\end{bmatrix} = T_1\left(\begin{bmatrix}u_1\\u_2\end{bmatrix}\right) + T_1\left(\begin{bmatrix}v_1\\v_2\end{bmatrix}\right) \\ &= T_1(\mathbf{u}) + T_1(\mathbf{v}) \end{aligned}$$

It turns out that T_2 *fails to satisfy this property. Can you prove that this is the case? Remember that to prove that a property DOES NOT hold, it suffices to find a counter-example. See if you can find vectors* $\mathbf{u}$ *and* $\mathbf{v}$ *such that*

$$T_2(\mathbf{u}+\mathbf{v}) \neq T_2(\mathbf{u}) + T_2(\mathbf{v}). \tag{55}$$

(See Practice Problem **??***.)*

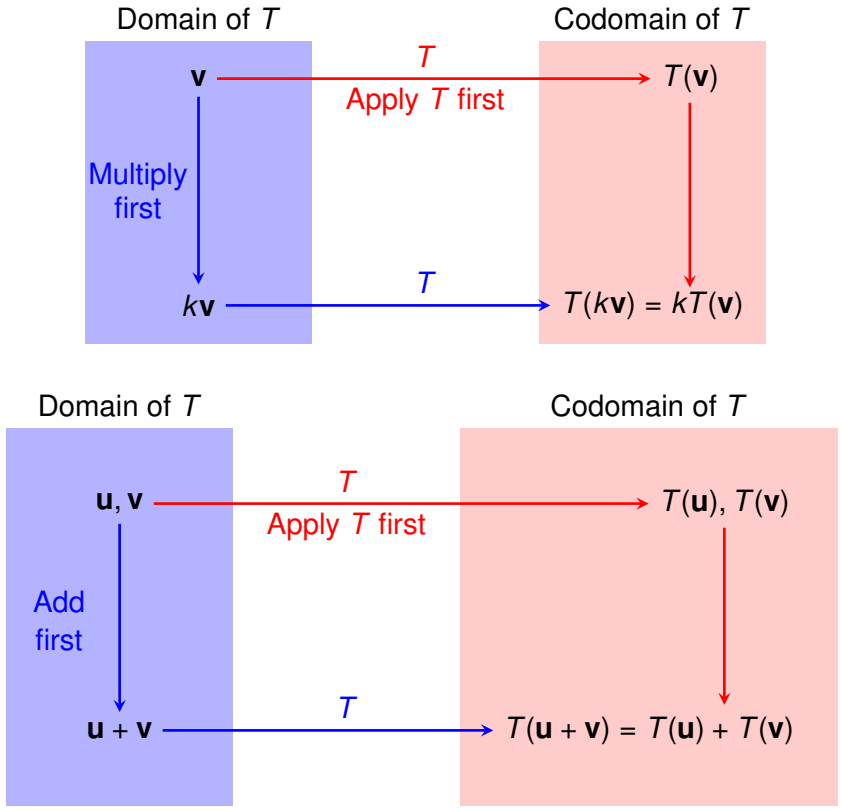
Transformations satisfying (**??**) and (**??**), like T_1, are called *linear transformations*. Transformations like T_2 are not linear. You have already encountered several linear transformations in the form of matrix transformations in sections Matrix Transformations and Geometric Transformations of the Plane.

Definition 52. A transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ is called a *linear transformation* if the following are true for all vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$, and scalars k.

$$T(k\mathbf{u}) = kT(\mathbf{u}) \tag{56}$$

$$T(\mathbf{u}+\mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v}) \tag{57}$$

Equations (**??**) and (**??**) of the above definition can be illustrated diagrammatically as follows.



Remark 8. Properties (**??**) and (**??**) are often combined into a single property.

$$T(k_1\mathbf{u} + k_2\mathbf{v}) = k_1 T(\mathbf{u}) + k_2 T(\mathbf{v}) \tag{58}$$

Example 87. Suppose $T : \mathbb{R}^2 \to \mathbb{R}^3$ is a linear transformation such that

$$T\left(\begin{bmatrix}1\\2\end{bmatrix}\right) = \begin{bmatrix}-1\\0\\3\end{bmatrix} \quad \text{and} \quad T\left(\begin{bmatrix}0\\-1\end{bmatrix}\right) = \begin{bmatrix}2\\-1\\0\end{bmatrix}$$

Find each of the following:

1. $T\left(\begin{bmatrix}2\\5\end{bmatrix}\right) = T\left(2\begin{bmatrix}1\\2\end{bmatrix} - \begin{bmatrix}0\\-1\end{bmatrix}\right)$
2. $T\left(\begin{bmatrix}1\\1\end{bmatrix}\right)$

Explanation. ?? Because T is a linear transformation, it satisfies (**??**). We compute:

$$\begin{aligned} T\left(\begin{bmatrix}2\\5\end{bmatrix}\right) &= T\left(2\begin{bmatrix}1\\2\end{bmatrix} - \begin{bmatrix}0\\-1\end{bmatrix}\right) \\ &= 2T\left(\begin{bmatrix}1\\2\end{bmatrix}\right) - T\left(\begin{bmatrix}0\\-1\end{bmatrix}\right) \\ &= 2\begin{bmatrix}-1\\0\\3\end{bmatrix} - \begin{bmatrix}2\\-1\\0\end{bmatrix} \\ &= \begin{bmatrix}-2\\0\\6\end{bmatrix} - \begin{bmatrix}2\\-1\\0\end{bmatrix} = \begin{bmatrix}-4\\1\\6\end{bmatrix} \end{aligned}$$

?? Observe that $\begin{bmatrix}1\\1\end{bmatrix} = \begin{bmatrix}1\\2\end{bmatrix} + \begin{bmatrix}0\\-1\end{bmatrix}$. By (**??**) we have:

$$\begin{aligned} T\left(\begin{bmatrix}1\\1\end{bmatrix}\right) &= T\left(\begin{bmatrix}1\\2\end{bmatrix} + \begin{bmatrix}0\\-1\end{bmatrix}\right) \\ &= T\left(\begin{bmatrix}1\\2\end{bmatrix}\right) + T\left(\begin{bmatrix}0\\-1\end{bmatrix}\right) \\ &= \begin{bmatrix}-1\\0\\3\end{bmatrix} + \begin{bmatrix}2\\-1\\0\end{bmatrix} = \begin{bmatrix}1\\-1\\3\end{bmatrix} \end{aligned}$$

Observation 7. In Example **??** we were given the images of two vectors, $\begin{bmatrix}1\\2\end{bmatrix}$ and $\begin{bmatrix}0\\-1\end{bmatrix}$, under a linear transformation T. Based on this information, we were able to determine the images of two additional vectors: $\begin{bmatrix}2\\5\end{bmatrix}$ and $\begin{bmatrix}1\\1\end{bmatrix}$. The reason we were able to determine $T\left(\begin{bmatrix}2\\5\end{bmatrix}\right)$ and $T\left(\begin{bmatrix}1\\1\end{bmatrix}\right)$ is because $\begin{bmatrix}2\\5\end{bmatrix}$ and $\begin{bmatrix}1\\1\end{bmatrix}$ can be written as unique linear combinations of $\begin{bmatrix}1\\2\end{bmatrix}$ and $\begin{bmatrix}0\\-1\end{bmatrix}$.

Can every vector of $\mathbb{R}^2$ be written as a linear combination of $\begin{bmatrix}1\\2\end{bmatrix}$ and $\begin{bmatrix}0\\-1\end{bmatrix}$?

Multiple Choice:

1. Yes
2. No

Is the information provided in Example **??** sufficient to determine the image of every vector in $\mathbb{R}^2$ under T?

Multiple Choice:

1. Yes
2. No

Example 88. Suppose $T : \mathbb{R}^2 \to \mathbb{R}^2$ is a transformation such that

$$T\left(\begin{bmatrix}2\\1\end{bmatrix}\right) = \begin{bmatrix}3\\2\end{bmatrix}, \quad T\left(\begin{bmatrix}-1\\0\end{bmatrix}\right) = \begin{bmatrix}1\\1\end{bmatrix}, \quad T\left(\begin{bmatrix}-1\\1\end{bmatrix}\right) = \begin{bmatrix}2\\-4\end{bmatrix}$$

Determine whether T is a linear transformation.

Explanation. Observe that

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

If T were a linear transformation, then we would have:

$$T\left(\begin{bmatrix}-1\\1\end{bmatrix}\right) = T\left(\begin{bmatrix}2\\1\end{bmatrix} + 3\begin{bmatrix}-1\\0\end{bmatrix}\right) = \begin{bmatrix}3\\2\end{bmatrix} + 3\begin{bmatrix}1\\1\end{bmatrix} = \begin{bmatrix}6\\5\end{bmatrix}$$

But according to the given,

$$T\left(\begin{bmatrix}-1\\1\end{bmatrix}\right) = \begin{bmatrix}2\\-4\end{bmatrix}$$

Since $\begin{bmatrix}6\\5\end{bmatrix} \neq \begin{bmatrix}2\\-4\end{bmatrix}$ we conclude that transformation T is not linear.

In Exploration **??** we introduced a transformation T_2 which turned out to be non-linear. It took some work to show that T_2 is not linear. The following theorem would have made our work easier.

Theorem 39. Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Then

1. $T(\mathbf{0}) = \mathbf{0}$. In other words, linear transformations map the zero vector to the zero vector.
2. T maps any line in $\mathbb{R}^n$ to a line (or the zero vector) in $\mathbb{R}^m$.

Proof To prove part **??**, let $\mathbf{v}$ be any vector in $\mathbb{R}^n$. By linearity of T, we have:

$$T(\mathbf{0}) = T(\mathbf{v} - \mathbf{v}) = T(\mathbf{v}) - T(\mathbf{v}) = \mathbf{0}$$

Part **??** will become evident after we prove Theorem **??** in Standard Matrix of a Linear Transformation and combine it with Practice Problem **??**. ■

Example 89. Use Theorem **??** to show that transformation T_2 of Exploration **??** is not linear.

Explanation. Recall that $T_2 : \mathbb{R}^2 \to \mathbb{R}^2$ was defined by

$$T_2\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} -x + y + 1 \\ y - 2 \end{bmatrix}$$

We evaluate T_2 at $\mathbf{0}$:

$$T_2(\mathbf{0}) = \begin{bmatrix} -0 + 0 + 1 \\ 0 - 2 \end{bmatrix} = \begin{bmatrix} 1 \\ -2 \end{bmatrix} \neq \mathbf{0}$$

Since $T_2(\mathbf{0}) \neq \mathbf{0}$, T_2 is not linear.

Linear Transformations Induced by Matrices

Recall that a transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ defined by $T(\mathbf{v}) = A\mathbf{v}$, where A is some $m \times n$ matrix, is called a matrix transformation (or transformation induced by A). As we had discovered in Matrix Transformations, all matrix transformations are linear. We now formalize this result as a theorem.

Theorem 40. Let A be an $m \times n$ matrix. Define $T : \mathbb{R}^n \to \mathbb{R}^m$ by $T(\mathbf{v}) = A\mathbf{v}$. Then T is a linear transformation.

Proof Let $\mathbf{u}$ and $\mathbf{v}$ be vectors in $\mathbb{R}^n$, and let k be a scalar. By properties of matrix multiplication we have:

$$T(\mathbf{u} + \mathbf{v}) = A(\mathbf{u} + \mathbf{v}) = A\mathbf{u} + A\mathbf{v} = T(\mathbf{u}) + T(\mathbf{v})$$

$$T(k\mathbf{u}) = A(k\mathbf{u}) = kA\mathbf{u} = kT(\mathbf{u})$$

Therefore T is a linear transformation. ■

Example 90. Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation induced by

$$A = \begin{bmatrix} 2 & 0 \\ 1 & 4 \\ 0 & 1 \end{bmatrix}$$

1. Find n and m.
2. Find the image of T.

Explanation. ?? A is a 3×2 matrix, so for the expression $T(\mathbf{x}) = A\mathbf{x}$ to make sense, $\mathbf{x}$ has to be a 2×1 vector. Thus, the domain of T is $\mathbb{R}^2$ ($n = 2$). The product $A\mathbf{x}$ is a 3×1 vector. The codomain of T is $\mathbb{R}^3$ ($m = 3$).

?? By Definition **??**, the image of T consists of images of all individual vectors in $\mathbb{R}^2$ under T. Every vector $\mathbf{v}$ in $\mathbb{R}^2$ can be written as $\mathbf{v} = a\mathbf{i} + b\mathbf{j}$ for some real numbers a and b. Consider the image of $\mathbf{v}$

$$T(\mathbf{v}) = T(a\mathbf{i} + b\mathbf{j}) = aT(\mathbf{i}) + bT(\mathbf{j}) = a \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} + b \begin{bmatrix} 0 \\ 4 \\ 1 \end{bmatrix}$$

This shows that the range, or the image, of T consists of all linear combinations of the columns of A. In other words, the image of T is the span of vectors $\begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 0 \\ 4 \\ 1 \end{bmatrix}$. The two vectors are not scalar multiples of each other, therefore they span a plane in $\mathbb{R}^3$.

Example 91. Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation induced by

$$A = \begin{bmatrix} -2 & 1 & 3 \\ 4 & -2 & -6 \end{bmatrix}$$

1. Find n and m.
2. Find and draw the image of T.

Explanation. ??

$$n = \ldots\ldots, \quad m = \ldots\ldots$$

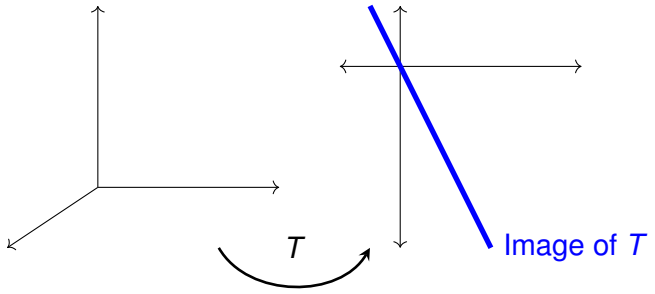
?? To find the image of T, we will take a slightly different approach from what we did in Example **????**.

Let $\mathbf{v} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$ be an arbitrary vector of $\mathbb{R}^3$. The image of $\mathbf{v}$ is given by

$$T(\mathbf{v}) = \begin{bmatrix} -2 & 1 & 3 \\ 4 & -2 & -6 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = a \begin{bmatrix} -2 \\ 4 \end{bmatrix} + b \begin{bmatrix} 1 \\ -2 \end{bmatrix} + c \begin{bmatrix} 3 \\ -6 \end{bmatrix}$$

$$= (a(-2) + b + c(3)) \begin{bmatrix} 1 \\ -2 \end{bmatrix}$$

This shows that the image of every vector in $\mathbb{R}^3$ is a scalar multiple of $\begin{bmatrix} 1 \\ -2 \end{bmatrix}$. This means that the image of T is a line in $\mathbb{R}^2$.



Linear Transformations of Subspaces of $\mathbb{R}^n$

Definition **??** defines a linear transformation as a map from $\mathbb{R}^n$ into $\mathbb{R}^m$. We will now make this definition more general by allowing the domain and the codomain of the transformation to be subspaces of $\mathbb{R}^n$ and $\mathbb{R}^m$. Eventually, a linear transformation will be defined as a mapping between vector spaces.

Definition 53. Let V and W be subspaces of $\mathbb{R}^n$ and $\mathbb{R}^m$. A transformation $T : V \to W$ is called a *linear transformation* if the following are true for all vectors $\mathbf{u}$ and $\mathbf{v}$ in V, and scalars k.

$$T(k\mathbf{u}) = kT(\mathbf{u})$$

$$T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$$

Example 92. Let V be a subspace of $\mathbb{R}^3$ consisting of all vectors in the xy-plane. Let W be a subspace of $\mathbb{R}^3$ consisting of all vectors along the z-axis. (Do a quick verification that V and W are subspaces of $\mathbb{R}^3$.) Define a transformation $T : V \to W$ by

$$T\left(\begin{bmatrix} a \\ b \\ 0 \end{bmatrix}\right) = \begin{bmatrix} 0 \\ 0 \\ a+b \end{bmatrix}$$

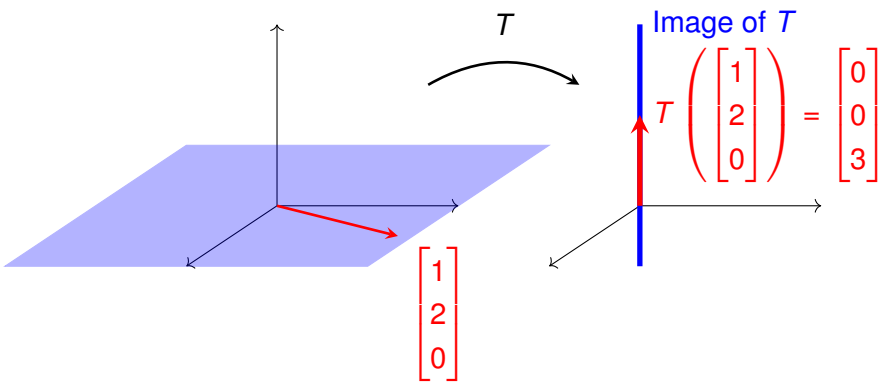
Show that T is a linear transformation, and describe its action geometrically.

Explanation. Consider two arbitrary elements $\begin{bmatrix} a_1 \\ b_1 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} a_2 \\ b_2 \\ 0 \end{bmatrix}$ of V.

$$\begin{aligned} T\left(\begin{bmatrix} a_1 \\ b_1 \\ 0 \end{bmatrix} + \begin{bmatrix} a_2 \\ b_2 \\ 0 \end{bmatrix}\right) &= T\left(\begin{bmatrix} a_1 + a_2 \\ b_1 + b_2 \\ 0 \end{bmatrix}\right) \\ &= \begin{bmatrix} 0 \\ 0 \\ (a_1 + a_2) + (b_1 + b_2) \end{bmatrix} \\ &= \begin{bmatrix} 0 \\ 0 \\ (a_1 + b_1) + (a_2 + b_2) \end{bmatrix} \\ &= \begin{bmatrix} 0 \\ 0 \\ a_1 + b_1 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ a_2 + b_2 \end{bmatrix} \\ &= T\left(\begin{bmatrix} a_1 \\ b_1 \\ 0 \end{bmatrix}\right) + T\left(\begin{bmatrix} a_2 \\ b_2 \\ 0 \end{bmatrix}\right) \end{aligned}$$

Verification of the fact that $T\left(k\begin{bmatrix} a_1 \\ b_1 \\ 0 \end{bmatrix}\right) = kT\left(\begin{bmatrix} a_1 \\ b_1 \\ 0 \end{bmatrix}\right)$ is similar, and we omit the details.

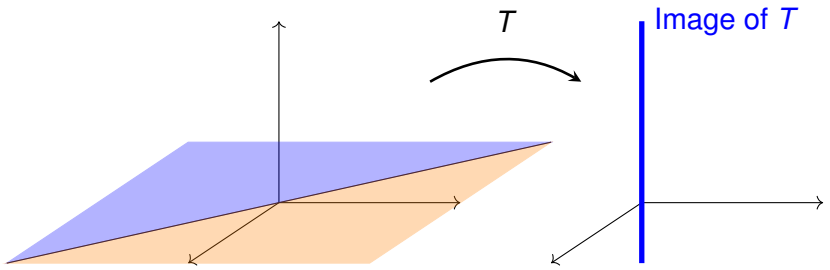
We have shown that T is a linear transformation. T maps all vectors in the xy-plane to the z-axis. The following diagram helps us visualize the action of T on a specific vector.



We can investigate further. Recall that T is defined by

$$T\left(\begin{bmatrix} a \\ b \\ 0 \end{bmatrix}\right) = \begin{bmatrix} 0 \\ 0 \\ a+b \end{bmatrix}$$

Complete each of the following statements referring to the diagram below.



The image of the line $y = -x$ under T is (a line/ a plane/ the zero vector/ all of $\mathbb{R}^3$)

The image of the orange part of the domain (the front triangle) is (the positive z-axis/ the negative z-axis/ the zero vector/ the entire z-axis)

The image of the purple part of the domain (the back triangle) is (the positive z-axis/ the negative z-axis/ the zero vector/ the entire z-axis)

We conclude this section by introducing two simple but important transformations.

Definition 54. The *identity transformation* on V, denoted by id_V, is a transformation that maps each element of V to itself.

In other words,

$$\mathrm{id}_V : V \to V$$

is a transformation such that

$$\mathrm{id}_V(\mathbf{v}) = \mathbf{v} \quad \text{for all} \quad \mathbf{v} \in V$$

Definition 55. The *zero transformation*, Z, maps every element of the domain to the zero vector. In other words,

$$Z : V \to W$$

is a transformation such that

$$Z(\mathbf{v}) = \mathbf{0} \quad \text{for all} \quad \mathbf{v} \in V$$

Theorem 41. The identity transformation is linear.

Proof Left to the reader. (See Practice Problem **??**) ■

Theorem 42. The zero transformation is linear.

Proof Left to the reader. (See Practice Problem **??**) ■

Practice Problems

Problem 258 *Define T_1 as follows*

$$T_1 : \mathbb{R}^2 \to \mathbb{R}^2$$

$$T_1\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} x - y \\ x \end{bmatrix}$$

*(See Exploration **??**) Use vectors* $\begin{bmatrix} 3 \\ 4 \end{bmatrix}$ *and* $\begin{bmatrix} -2 \\ 1 \end{bmatrix}$ *to illustrate that* $T_1(\mathbf{u} + \mathbf{v}) = T_1(\mathbf{u}) + T_1(\mathbf{v})$.

Problem 259 *Define T_2 as follows*

$$T_2 : \mathbb{R}^2 \to \mathbb{R}^2$$

$$T_2\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} -x + y + 1 \\ y - 2 \end{bmatrix}$$

*(See Exploration **??**)*

Use a counter-example to prove that $T_2(\mathbf{u} + \mathbf{v}) \neq T_2(\mathbf{u}) + T_2(\mathbf{v})$

Problem 260 *Suppose* $T : \mathbb{R}^{10} \to \mathbb{R}^2$ *is a linear transformation such that* $T(\mathbf{u}) = \begin{bmatrix} 2 \\ -1 \end{bmatrix}$ *and* $T(\mathbf{v}) = \begin{bmatrix} -5 \\ 3 \end{bmatrix}$. *Find the image of* $3\mathbf{u} - \mathbf{v}$.

$$T(3\mathbf{u} - \mathbf{v}) = \begin{bmatrix} \dots \dots \\ \dots \dots \end{bmatrix}$$

Problem 261 *Let* $\mathbf{u}$ *be a fixed non-zero vector. Define* $T_{\mathbf{u}} : \mathbb{R}^2 \to \mathbb{R}^2$, *by* $T_{\mathbf{u}}(\mathbf{x}) = \mathbf{u} - \mathbf{x}$.

1. *Describe the effect of this transformation by sketching* $\mathbf{x}$ *and* $T_{\mathbf{u}}(\mathbf{x})$ *for at least four vectors* $\mathbf{x}$ *and a fixed vector* $\mathbf{u}$ *of your choice.*
2. *Is* $T_{\mathbf{u}}$ *a linear transformation?*

Problem 262 *Define* $P_{xy} : \mathbb{R}^3 \to \mathbb{R}^2$, *by* $P_{xy}\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) = \begin{bmatrix} x \\ y \\ 0 \end{bmatrix}$. *This transformation is called an orthogonal projection onto the xy-plane. Show that* P_{xy} *is a linear transformation.*

Problem 263 *Suppose a linear transformation* $T : \mathbb{R}^3 \to \mathbb{R}^3$ *maps* $\mathbf{i}$ *to* $\begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix}$, $\mathbf{j}$ *to* $\begin{bmatrix} -2 \\ 4 \\ 1 \end{bmatrix}$, *and* $\mathbf{k}$ *to* $\begin{bmatrix} 3 \\ 0 \\ -5 \end{bmatrix}$. *Find the image of* $\begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix}$ *under* T.

$$T\left(\begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix}\right) = \begin{bmatrix} \dots \dots \\ \dots \dots \\ \dots \dots \end{bmatrix}$$

Problem 264 *Prove that the identity transformation is linear. (Theorem* **??***)*

Problem 265 *Prove that the zero transformation is linear. (Theorem* **??***)*

Problem 266 *Find the domain and the codomain of the linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ induced by A; find and draw the image of T.*

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

Domain: $\mathbb{R}^n$, where $n =$ ... .

Codomain: $\mathbb{R}^m$, where $m =$ ... .

Problem 267 *Find the domain and the codomain of the linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ induced by A; find and draw the image of T.*

$$A = \begin{bmatrix} 3 & -1 \\ -3 & 1 \end{bmatrix}$$

Exercise Source

Practice Problem **??** is adopted from Problem 5.1.3 of Ken Kuttler's *A First Course in Linear Algebra*. (CC-BY)

Ken Kuttler, *A First Course in Linear Algebra*, Lyryx 2017, Open Edition, p. 272.

43 Standard Matrix of a Linear Transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$

In Matrix Transformations and Introduction to Linear Transformations we learned several important properties of matrix transformations of $\mathbb{R}^n$ and subspaces of $\mathbb{R}^n$. Let's summarize the main points.

Summary 1. For a matrix transformation $T : \mathbb{R}^n \to \mathbb{R}^m$, induced by an $m \times n$ matrix A we have the following results:

- T is linear. (Theorem **??**) This means that

$$T(k_1\mathbf{u} + k_2\mathbf{v}) = k_1 T(\mathbf{u}) + k_2 T(\mathbf{v})$$

  for vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$ and scalars k_1 and k_2 in $\mathbb{R}$.

- Columns of A are the images of the standard unit vectors of $\mathbb{R}^n$ under T. (Observation **??**)

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & \dots & \dots & a_{mn} \end{bmatrix} = \begin{bmatrix} | & | & & | \\ T(\mathbf{e}_1) & T(\mathbf{e}_2) & \dots & T(\mathbf{e}_n) \\ | & | & & | \end{bmatrix}$$

- The action of T on all of the elements of $\mathbb{R}^n$ is completely determined by where T maps the standard unit vectors. (See Examples **??** and **??**)

The last point in the summary is so important that it is worth illustrating again.

Example 93. Let $T : \mathbb{R}^3 \to \mathbb{R}^2$ be a linear transformation. Suppose that the only information we have about this transformation is that $T(\mathbf{i}) = \begin{bmatrix} 3 \\ -1 \end{bmatrix}$, $T(\mathbf{j}) = \begin{bmatrix} 0 \\ 4 \end{bmatrix}$ and $T(\mathbf{k}) = \begin{bmatrix} -2 \\ 1 \end{bmatrix}$. Is this information sufficient to determine the image of $\mathbf{w} = \begin{bmatrix} 1 \\ -3 \\ 6 \end{bmatrix}$?

Explanation. Observe that

$$\mathbf{w} = \mathbf{i} - 3\mathbf{j} + 6\mathbf{k}$$

We find $T(\mathbf{w})$ by using the fact that T is linear.

$$T(\mathbf{w}) = T(\mathbf{i} - 3\mathbf{j} + 6\mathbf{k}) = T(\mathbf{i}) - 3T(\mathbf{j}) + 6T(\mathbf{k}) = \begin{bmatrix} 3 \\ -1 \end{bmatrix} - 3\begin{bmatrix} 0 \\ 4 \end{bmatrix} + 6\begin{bmatrix} -2 \\ 1 \end{bmatrix} = \begin{bmatrix} -9 \\ -7 \end{bmatrix}$$

Because of properties of linear transformations, the information about the images of the standard unit vectors proved to be sufficient for us to determine the image of $\mathbf{w}$.

In Example **??**, there was nothing special about the vector $\mathbf{w}$. Any vector $\mathbf{x}$ of $\mathbb{R}^n$ can be written as a unique linear combination of the standard unit vectors $\mathbf{e}_1, \ldots, \mathbf{e}_n$. Therefore, the image of any vector $\mathbf{x}$ under a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ is uniquely determined by the images of $\mathbf{e}_1, \ldots, \mathbf{e}_n$. Knowing $T(\mathbf{e}_1), \ldots, T(\mathbf{e}_n)$ allows us to construct a matrix A, with $T(\mathbf{e}_1), \ldots, T(\mathbf{e}_n)$ as columns, that induces transformation T. We formalize this idea in a theorem.

Theorem 43. Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Then T is a matrix transformation with

$$A = \begin{bmatrix} | & | & & | \\ T(\mathbf{e}_1) & T(\mathbf{e}_2) & \ldots & T(\mathbf{e}_n) \\ | & | & & | \end{bmatrix}$$

as a matrix that induces T.

Proof Observe that

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = x_1 \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix} + \cdots + x_n \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix} = x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + \cdots + x_n\mathbf{e}_n$$

Because T is linear, we have

$$T(\mathbf{x}) = T(x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + \cdots + x_n\mathbf{e}_n) = x_1T(\mathbf{e}_1) + x_2T(\mathbf{e}_2) + \cdots + x_nT(\mathbf{e}_n)$$

$$= \begin{bmatrix} | & | & & | \\ T(\mathbf{e}_1) & T(\mathbf{e}_2) & \ldots & T(\mathbf{e}_n) \\ | & | & & | \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = A\mathbf{x}$$

Thus, for every $\mathbf{x}$ in $\mathbb{R}^n$, we have $T(\mathbf{x}) = A\mathbf{x}$. ■

Theorem **??** shows that every matrix transformation is linear. Theorem **??** states that every linear transformation from $\mathbb{R}^n$ into $\mathbb{R}^m$ is a matrix transformation. We combine these results in a corollary.

Corollary 5. A transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation if and only if it is a matrix transformation.

The results of this section rely on the fact that every vector of $\mathbb{R}^n$ can be written as a unique linear combination of the standard unit vectors $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n$. These vectors form the *standard basis* for $\mathbb{R}^n$. We will see in Matrix of a Linear Transformation with Respect to Arbitrary Bases that the matrix used to represent a linear transformation depends on a choice of basis. Because we are using the standard basis, it is natural to name the matrix in Theorem **??** accordingly.

Definition 56. The matrix in Theorem **??** is known as the *standard matrix of the linear transformation T*.

Example 94. The standard matrix of a linear transformation $T : \mathbb{R}^3 \to \mathbb{R}^2$ such that $T(\mathbf{i}) = \begin{bmatrix} 2 \\ -1 \end{bmatrix}$, $T(\mathbf{j}) = \begin{bmatrix} -1 \\ 3 \end{bmatrix}$ and $T(\mathbf{k}) = \begin{bmatrix} 0 \\ 4 \end{bmatrix}$ is

$$A = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 3 & 4 \end{bmatrix}$$

Example 95. Find the standard matrix of a linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$ such that $T(\mathbf{i}) = 2\mathbf{i}$ and $T(\mathbf{j}) = 2\mathbf{j}$.

Explanation. We use the images of $\mathbf{i}$ and $\mathbf{j}$ as columns of the matrix. The standard matrix of T is

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

Example 96. Find the standard matrix of a linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^4$ if $T\left(\begin{bmatrix} 3 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} 6 \\ 1 \\ 13 \\ -1 \end{bmatrix}$ and $T\left(\begin{bmatrix} -2 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} -2 \\ 0 \\ -8 \\ 2 \end{bmatrix}$.

Explanation. In this example we are not given the images of the standard basis vectors $\mathbf{i}$ and $\mathbf{j}$. However, we can find the images of $\mathbf{i}$ and $\mathbf{j}$ by expressing $\mathbf{i}$ and $\mathbf{j}$ as linear combinations of $\begin{bmatrix} 3 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} -2 \\ 0 \end{bmatrix}$, then apply the fact that T is linear.

Let's start with the easy one.

$$\mathbf{i} = -\frac{1}{2}\begin{bmatrix} -2 \\ 0 \end{bmatrix}$$

Therefore, by linearity of T, we have:

$$T(\mathbf{i}) = T\left(-\frac{1}{2}\begin{bmatrix} -2 \\ 0 \end{bmatrix}\right) = -\frac{1}{2}T\left(\begin{bmatrix} -2 \\ 0 \end{bmatrix}\right) = -\frac{1}{2}\begin{bmatrix} -2 \\ 0 \\ -8 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 4 \\ -1 \end{bmatrix}$$

This gives us the first column of the standard matrix for T.

You can solve the vector equation

$$a\begin{bmatrix} 3 \\ 1 \end{bmatrix} + b\begin{bmatrix} -2 \\ 0 \end{bmatrix} = \mathbf{j}$$

to express $\mathbf{j}$ as a linear combination of $\begin{bmatrix} 3 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} -2 \\ 0 \end{bmatrix}$ as follows:

$$\mathbf{j} = \begin{bmatrix} 3 \\ 1 \end{bmatrix} + \frac{3}{2}\begin{bmatrix} -2 \\ 0 \end{bmatrix}$$

By linearity of T,

$$T(\mathbf{j}) = T\left(\begin{bmatrix} 3 \\ 1 \end{bmatrix} + \frac{3}{2}\begin{bmatrix} -2 \\ 0 \end{bmatrix}\right) = T\left(\begin{bmatrix} 3 \\ 1 \end{bmatrix}\right) + \frac{3}{2}T\left(\begin{bmatrix} -2 \\ 0 \end{bmatrix}\right)$$
$$= \begin{bmatrix} 6 \\ 1 \\ 13 \\ -1 \end{bmatrix} + \frac{3}{2}\begin{bmatrix} -2 \\ 0 \\ -8 \\ 2 \end{bmatrix} = \begin{bmatrix} 3 \\ 1 \\ 1 \\ 2 \end{bmatrix}$$

This gives us the second column of the standard matrix. Putting all of the information together, we get the following standard matrix for T:

$$A = \begin{bmatrix} 1 & 3 \\ 0 & 1 \\ 4 & 1 \\ -1 & 2 \end{bmatrix}$$

Practice Problems

Problem 268 *Suppose that a linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^3$ is such that $T(\mathbf{i}) = \begin{bmatrix} -4 \\ 2 \\ 1 \end{bmatrix}$ and $T(\mathbf{j}) = \begin{bmatrix} 0 \\ -1 \\ 5 \end{bmatrix}$. Find $T\left(\begin{bmatrix} 4 \\ -1 \end{bmatrix} \right)$.*

$$T\left(\begin{bmatrix} 4 \\ -1 \end{bmatrix} \right) = \begin{bmatrix} \ldots \ldots \\ \ldots \ldots \\ \ldots \ldots \end{bmatrix}$$

Problem 269 *Suppose that a linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^3$ is such that $T\left(\begin{bmatrix} 1 \\ -1 \end{bmatrix} \right) = \begin{bmatrix} 1 \\ 4 \\ -1 \end{bmatrix}$ and $T\left(\begin{bmatrix} 2 \\ 0 \end{bmatrix} \right) = \begin{bmatrix} 0 \\ 6 \\ 4 \end{bmatrix}$. Find the standard matrix A of T.*

$$A = \begin{bmatrix} \ldots \ldots & \ldots \ldots \\ \ldots \ldots & \ldots \ldots \\ \ldots \ldots & \ldots \ldots \end{bmatrix}$$

Problem 270 *Find the standard matrix A of the linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$ if T doubles the x component of every vector and triples the y component.*

$$A = \begin{bmatrix} \ldots \ldots & \ldots \ldots \\ \ldots \ldots & \ldots \ldots \end{bmatrix}$$

Problem 271 *Find the standard matrix A of the linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$ if T reverses the direction of each vector.*

$$A = \begin{bmatrix} \ldots \ldots & \ldots \ldots \\ \ldots \ldots & \ldots \ldots \end{bmatrix}$$

Problem 272 *Find the standard matrix A of the linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$ if T doubles the length of each vector.*

$$A = \begin{bmatrix} \ldots \ldots & \ldots \ldots \\ \ldots \ldots & \ldots \ldots \end{bmatrix}$$

Problem 273 *Find the standard matrix A of the linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$ if T projects each vector onto the x-axis. (e.g. $T\left(\begin{bmatrix}4\\5\end{bmatrix}\right) = \begin{bmatrix}4\\0\end{bmatrix}$)*

$$A = \begin{bmatrix} \ldots\ldots & \ldots\ldots \\ \ldots\ldots & \ldots\ldots \end{bmatrix}$$

Problem 274 *Find the standard matrix A of the linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$ if T projects each vector onto the y-axis. (e.g. $T\left(\begin{bmatrix}4\\5\end{bmatrix}\right) = \begin{bmatrix}0\\5\end{bmatrix}$)*

$$A = \begin{bmatrix} \ldots\ldots & \ldots\ldots \\ \ldots\ldots & \ldots\ldots \end{bmatrix}$$

44 Composition and Inverses of Linear Transformations

Note to Student: In this section we will often use U, V and W to denote subspaces of $\mathbb{R}^n$, or any other finite-dimensional vector space, such as those we study in Vector Spaces.

Composition of Linear Transformations

Definition 57. Let U, V and W be vector spaces, and let $T : U \to V$ and $S : V \to W$ be linear transformations. The *composition* of S and T is the transformation $S \circ T : U \to W$ given by

$$(S \circ T)(\mathbf{u}) = S(T(\mathbf{u}))$$

Example 97. Define

$$T : \mathbb{R}^2 \to \mathbb{R}^2 \quad \text{by} \quad T\left(\begin{bmatrix} u_1 \\ u_2 \end{bmatrix}\right) = \begin{bmatrix} u_1 + u_2 \\ 3u_1 + 3u_2 \end{bmatrix}$$

$$S : \mathbb{R}^2 \to \mathbb{R}^2 \quad \text{by} \quad S\left(\begin{bmatrix} v_1 \\ v_2 \end{bmatrix}\right) = \begin{bmatrix} 3v_1 - v_2 \\ -3v_1 + v_2 \end{bmatrix}$$

(You should be able to verify that both transformations are linear.) Examine the effect of $S \circ T$ on vectors of $\mathbb{R}^2$.

Explanation. From the computational standpoint, the situation is simple.

$$\begin{aligned}(S \circ T)\left(\begin{bmatrix} u_1 \\ u_2 \end{bmatrix}\right) &= S\left(T\left(\begin{bmatrix} u_1 \\ u_2 \end{bmatrix}\right)\right) = S\left(\begin{bmatrix} u_1 + u_2 \\ 3u_1 + 3u_2 \end{bmatrix}\right) \\ &= \begin{bmatrix} 3(u_1 + u_2) - (3u_1 + 3u_2) \\ -3(u_1 + u_2) + (3u_1 + 3u_2) \end{bmatrix} \\ &= \mathbf{0}\end{aligned}$$

This means that $S \circ T$ maps all vectors of $\mathbb{R}^2$ to $\mathbf{0}$.

In addition to the computational approach, it is also useful to visualize what happens geometrically.

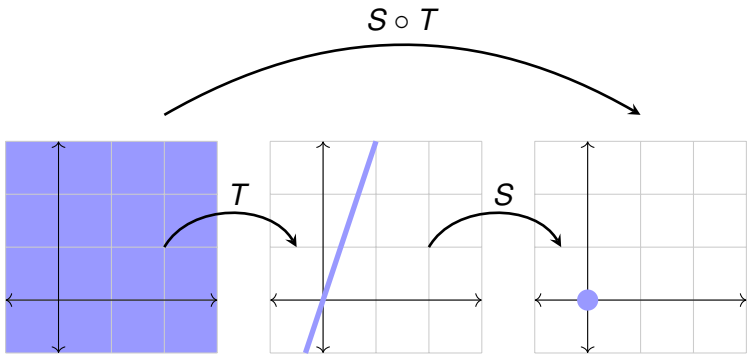
First, observe that

$$T\left(\begin{bmatrix} u_1 \\ u_2 \end{bmatrix}\right) = \begin{bmatrix} u_1 + u_2 \\ 3u_1 + 3u_2 \end{bmatrix} = (u_1 + u_2)\begin{bmatrix} 1 \\ 3 \end{bmatrix}$$

Therefore the image of any vector of $\mathbb{R}^2$ under T lies on the line determined by the vector $\begin{bmatrix}1\\3\end{bmatrix}$.

Even though S is defined on all of $\mathbb{R}^2$, we are only interested in the action of S on vectors along the line determined by $\begin{bmatrix}1\\3\end{bmatrix}$. Our computations showed that all such vectors map to $\mathbf{0}$.

The actions of individual transformations, as well as the composite transformation are shown below.



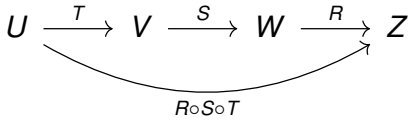
Theorem 44. The composition of two linear transformations is linear.

Proof Let $T : U \to V$ and $S : V \to W$ be linear transformations. We will show that $S \circ T$ is linear. For all vectors $\mathbf{u}_1$ and $\mathbf{u}_2$ of U and scalars a and b we have:

$$\begin{aligned}(S \circ T)(a\mathbf{u}_1 + b\mathbf{u}_2) &= S(T(a\mathbf{u}_1 + b\mathbf{u}_2))\\ &= S(aT(\mathbf{u}_1) + bT(\mathbf{u}_2))\\ &= aS(T(\mathbf{u}_1)) + bS(T(\mathbf{u}_2))\\ &= a(S \circ T)(\mathbf{u}_1) + b(S \circ T)(\mathbf{u}_2)\end{aligned}$$

■

Theorem 45. Composition of linear transformations is associative. In other words, for linear transformations T, S and R



We have $(R \circ S) \circ T = R \circ (S \circ T)$.

Proof For all $\mathbf{u}$ in U we have:

$$\begin{aligned}((R \circ S) \circ T)(\mathbf{u}) &= (R \circ S)(T(\mathbf{u})) = R(S(T(\mathbf{u}))) \\ &= R((S \circ T)(\mathbf{u})) = (R \circ (S \circ T))(\mathbf{u})\end{aligned}$$

■

Composition and Matrix Multiplication

In this section we will consider linear transformations of $\mathbb{R}^n$ and their standard matrices.

Theorem 46. Let $T : \mathbb{R}^n \to \mathbb{R}^m$ and $S : \mathbb{R}^m \to \mathbb{R}^p$ be linear transformations with standard matrices M_T and M_S, respectively. Then the composite transformation $S \circ T : \mathbb{R}^n \to \mathbb{R}^p$ has a standard matrix given by the product $M_S M_T$.

Proof For all $\mathbf{v}$ in $\mathbb{R}^n$ we have:

$$(S \circ T)(\mathbf{v}) = S(T(\mathbf{v})) = S(M_T\mathbf{v}) = M_S(M_T\mathbf{v}) = (M_S M_T)\mathbf{v}$$

■

Example 98. In Example **??**, we discussed a composite transformation $S \circ T : \mathbb{R}^2 \to \mathbb{R}^2$ given by:

$$T\left(\begin{bmatrix}u_1\\u_2\end{bmatrix}\right) = \begin{bmatrix}u_1 + u_2\\3u_1 + 3u_2\end{bmatrix} \quad \text{and} \quad S\left(\begin{bmatrix}v_1\\v_2\end{bmatrix}\right) = \begin{bmatrix}3v_1 - v_2\\-3v_1 + v_2\end{bmatrix}$$

Express $S \circ T$ as a matrix transformation.

Explanation. The standard matrix for $T : \mathbb{R}^2 \to \mathbb{R}^2$ is

$$\begin{bmatrix}1 & 1\\3 & 3\end{bmatrix}$$

and the standard matrix for $S : \mathbb{R}^2 \to \mathbb{R}^2$ is

$$\begin{bmatrix}3 & -1\\-3 & 1\end{bmatrix}$$

The standard matrix for $S \circ T$ is the product

$$\begin{bmatrix}3 & -1\\-3 & 1\end{bmatrix}\begin{bmatrix}1 & 1\\3 & 3\end{bmatrix} = \begin{bmatrix}0 & 0\\0 & 0\end{bmatrix}$$

We conclude this section by revisiting the associative property of matrix multiplication. At the time matrix multiplication was introduced, we skipped the cumbersome proof that for appropriately sized matrices A, B and C, we have $(AB)C = A(BC)$. (See Theorem **??**.) We are now in a position to prove this result with ease.

Every matrix induces a linear transformation. The product of two matrices can be interpreted as a composition of transformations. Since function composition is associative, so is matrix multiplication. We formalize this observation as a theorem.

Theorem 47 (Associativity of Matrix Multiplication)**.** Let A, B and C be matrices of appropriate dimensions so that the product $(AB)C$ is defined. Then

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

Inverses of Linear Transformations

Exploration 275 *Define a linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$ by $T(\mathbf{v}) = 2\mathbf{v}$. In other words, T doubles every vector in $\mathbb{R}^2$. Now define $S : \mathbb{R}^2 \to \mathbb{R}^2$ by $S(\mathbf{v}) = \frac{1}{2}\mathbf{v}$. What happens when we compose these two transformations?*

$$(S \circ T)(\mathbf{v}) = S(T(\mathbf{v})) = S(2\mathbf{v}) = \left(\frac{1}{2}\right)(2)\mathbf{v} = \mathbf{v}$$

$$(T \circ S)(\mathbf{v}) = T(S(\mathbf{v})) = T(\frac{1}{2}\mathbf{v}) = (2)\left(\frac{1}{2}\right)\mathbf{v} = \mathbf{v}$$

Both composite transformations return the original vector $\mathbf{v}$. In other words, $S \circ T = id_{\mathbb{R}^2}$ and $T \circ S = id_{\mathbb{R}^2}$. We say that S is an inverse of T, and T is an inverse of S.

Definition 58. Let V and W be vector spaces, and let $T : V \to W$ be a linear transformation. A transformation $S : W \to V$ that satisfies $S \circ T = \mathrm{id}_V$ and $T \circ S = \mathrm{id}_W$ is called an *inverse* of T. If T has an inverse, T is called *invertible*.

Example 99. Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be a transformation defined by $T\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} x+y \\ x-y \end{bmatrix}$. (How would you verify that T is linear?) Show that $S : \mathbb{R}^2 \to \mathbb{R}^2$ given by $S\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} 0.5x+0.5y \\ 0.5x-0.5y \end{bmatrix}$ is an inverse of T.

Explanation. We will show that $S \circ T = \mathrm{id}_{\mathbb{R}^2}$.

$$(S \circ T)\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = S\left(T\left(\begin{bmatrix} x \\ y \end{bmatrix}\right)\right) = S\left(\begin{bmatrix} x+y \\ x-y \end{bmatrix}\right)$$
$$= \begin{bmatrix} 0.5(x+y)+0.5(x-y) \\ 0.5(x+y)-0.5(x-y) \end{bmatrix} = \begin{bmatrix} x \\ y \end{bmatrix}$$

We leave it to the reader to verify that $T \circ S = \mathrm{id}_{\mathbb{R}^2}$.

Linearity of Inverses of Linear Transformations

Definition **??** does not specifically require an inverse S of a linear transformation T to be linear, but it turns out that the requirement that $S \circ T = \mathrm{id}_V$ and $T \circ S = \mathrm{id}_W$ is sufficient to guarantee that S is linear.

Theorem 48. Suppose $T : V \to W$ is an invertible linear transformation. Let S be an inverse of T. Then S is linear.

Proof The proof of this result is left to the reader. (See Practice Problem **??**) ■

Linear Transformations of $\mathbb{R}^n$ and the Standard Matrix of the Inverse Transformation

Every linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ is a matrix transformation. (See Theorem **??**.) If T has an inverse S, then by Theorem **??**, S is also a matrix transformation. Let M_T and M_S denote the standard matrices of T and S, respectively. We see that $S \circ T = \mathrm{id}_{\mathbb{R}^n}$ and $T \circ S = \mathrm{id}_{\mathbb{R}^m}$ if and only if $M_S M_T = I_n$ and $M_T M_S = I_m$. In other words, T and S are inverse transformations if and only if M_T and M_S are matrix inverses.

Note that if S is an inverse of T, then M_T and M_S are square matrices, and $n = m$.

Theorem 49. Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be a linear transformation, and let M be the standard matrix of T.

1. (Existence of Inverses.) T is invertible if and only if M is invertible. If T is invertible, then the inverse is induced by M^{-1}.
2. (Uniqueness of Inverses.) If S is an inverse of T, then S is unique.

Proof Part **??** follows directly from the preceding discussion. Part **??** follows from uniqueness of matrix inverses. (Theorem **??**.) ■

Please note that Theorem **??** is only applicable in the context of linear transformations of $\mathbb{R}^n$ and their standard matrices. The following example provides us with motivation to investigate inverses further, which we will do in Existence of the Inverse of a Linear Transformation.

Exploration 276 *Let*

$$V = span\left(\begin{bmatrix}1\\0\\0\end{bmatrix}, \begin{bmatrix}1\\1\\1\end{bmatrix}\right)$$

Define a linear transformation

$$T : V \to \mathbb{R}^2$$

by

$$T\left(\begin{bmatrix}1\\0\\0\end{bmatrix}\right) = \begin{bmatrix}1\\1\end{bmatrix} \quad \text{and} \quad T\left(\begin{bmatrix}1\\1\\1\end{bmatrix}\right) = \begin{bmatrix}0\\1\end{bmatrix}$$

Observe that $\left\{\begin{bmatrix}1\\0\\0\end{bmatrix}, \begin{bmatrix}1\\1\\1\end{bmatrix}\right\}$ *is a basis of V (why?). The information about the images of the basis vectors is sufficient to define a linear transformation. This is because every vector* $\mathbf{v}$ *in V can be expressed as a linear combination of the basis elements in a unique way. The image,* $T(\mathbf{v})$*, can be found by applying the linearity properties. At this point we know what transformation T does, but it is still unclear what the matrix of this linear transformation is.*

Geometrically speaking, the domain of T is a plane in $\mathbb{R}^3$ *and its codomain is* $\mathbb{R}^2$*.*

Does T have an inverse? We are not in a position to answer this question right now because Theorem **??** *does not apply to this situation.*

Exploration **??** highlights the necessity of further discussion of inverses and of matrices associated with linear transformations. In Example **??**, we will prove that T has an inverse, and in Example **??**, we will address the issue of matrices associated with T and its inverse.

Practice Problems

Problem 277 *Let* $T : \mathbb{R}^2 \to \mathbb{R}^2$ *and* $S : \mathbb{R}^2 \to \mathbb{R}^2$ *be linear transformations with standard matrices*

$$M_T = \begin{bmatrix}2 & -4\\1 & 2\end{bmatrix} \quad \text{and} \quad M_S = \begin{bmatrix}1 & -1\\2 & 1\end{bmatrix}$$

respectively. Describe the actions of T, S, and $S \circ T$ geometrically, as in Example **??**.

Problem 278 *Let $T : \mathbb{R}^3 \to \mathbb{R}^2$ and $S : \mathbb{R}^2 \to \mathbb{R}^2$ be linear transformations with standard matrices*

$$M_T = \begin{bmatrix} 1 & 0 & -1 \\ 2 & 1 & 0 \end{bmatrix} \quad \text{and} \quad M_S = \begin{bmatrix} -1 & 2 \\ 1 & -2 \end{bmatrix}$$

respectively. Describe the actions of T, S, and $S \circ T$ geometrically, as in Example **??**.

Problem 279 *Complete the Explanation of Example* **??** *by verifying that $T \circ S = id_{\mathbb{R}^2}$.*

Problem 280 *Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be a linear transformation given by*

$$T\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} 2x - 5y \\ -x + 3y \end{bmatrix}$$

Propose a candidate for the inverse of T and verify your choice using Definition **??**.

$$T^{-1}\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} \ldots\ldots x + \ldots\ldots y \\ x + \ldots\ldots y \end{bmatrix}$$

Problem 281 *Explain why linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$ given by*

$$T\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} 2x + 2y \\ -3x - 3y \end{bmatrix}$$

does not have an inverse.

Problem 282 *Suppose $T : U \to V$ and $S : V \to W$ are linear transformations with inverses T' and S' respectively. Prove that $T' \circ S'$ is the inverse of $S \circ T$.*

45 Image and Kernel of a Linear Transformation

Note to Student: In this section we will often use U, V and W to denote subspaces of $\mathbb{R}^n$, or any other finite-dimensional vector space, such as those we study in Vector Spaces.

The Image of a Linear Transformation

Definition 59. Let V and W be vector spaces, and let $T : V \to W$ be a linear transformation. The *image* of T, denoted by $\text{im}(T)$, is the set

$$\text{im}(T) = \{ T(\mathbf{v}) : \mathbf{v} \in V \}$$

In other words, the image of T consists of individual images of all vectors of V.

Example 100. Consider the linear transformation $T : \mathbb{R}^3 \to \mathbb{R}^2$ with standard matrix

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \end{bmatrix}$$

1. Find $\text{im}(T)$.
2. Illustrate the action of T with a sketch.

Explanation. ?? Let $\mathbf{v} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$ then

$$T(\mathbf{v}) = A\mathbf{v} = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = a \begin{bmatrix} 1 \\ 2 \end{bmatrix} + b \begin{bmatrix} 2 \\ 4 \end{bmatrix} + c \begin{bmatrix} 3 \\ 6 \end{bmatrix}$$

Thus, every element of the image can be written as a linear combination of the columns of A. We conclude that

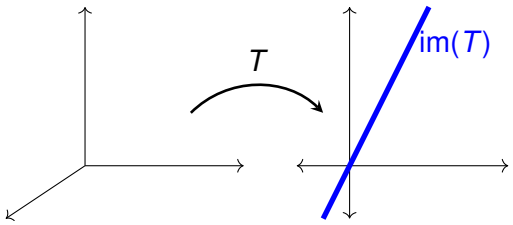
$$\text{im}(T) = \text{span}\left(\begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 2 \\ 4 \end{bmatrix}, \begin{bmatrix} 3 \\ 6 \end{bmatrix} \right) = \text{col}(A)$$

Every column of A is a scalar multiple of $\begin{bmatrix} 1 \\ 2 \end{bmatrix}$. Thus,

$$\text{im}(T) = \text{span}\left(\begin{bmatrix}1\\2\end{bmatrix}, \begin{bmatrix}2\\4\end{bmatrix}, \begin{bmatrix}3\\6\end{bmatrix}\right) = \text{span}\left(\begin{bmatrix}1\\2\end{bmatrix}\right)$$

The image of T is a line in $\mathbb{R}^2$ determined by the vector $\begin{bmatrix}1\\2\end{bmatrix}$.

?? The action of T can be illustrated with a sketch.



In Example **??** we observed that the image of the linear transformation was equal to the column space of its standard matrix. In general, it is easy to see that if $T : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation with standard matrix A then the following relationship holds:

$$\text{im}(T) = \text{col}(A)$$

In addition, by Theorem **??**, we know that

$$\dim(\text{im}(T)) = \dim(\text{col}(A)) = \text{rank}(A)$$

Example 101. Let $T : \mathbb{R}^5 \to \mathbb{R}^4$ be a linear transformation with standard matrix

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

Find $\text{im}(T)$ and $\dim(\text{im}(T))$.

Explanation. As in Example **??**, the image of T is given by

$$\text{im}(T) = \text{span}\left(\begin{bmatrix}1\\-1\\3\\1\end{bmatrix}, \begin{bmatrix}2\\3\\0\\-1\end{bmatrix}, \begin{bmatrix}2\\1\\0\\1\end{bmatrix}, \begin{bmatrix}-1\\0\\3\\-2\end{bmatrix}, \begin{bmatrix}0\\-1\\6\\-1\end{bmatrix}\right) = \text{col}(A)$$

This time it is harder to detect the vectors that can be eliminated from the spanning set without affecting the span. We have to rely on the reduced row-echelon form of A.

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

We can see that $\text{rank}(A) = 3$, so $\dim(\text{im}(T)) = 3$.

To identify vectors that span $\text{im}(T)$, we turn to Procedure **??**. We identify the first three columns as pivot columns. These columns are linearly independent and span $\text{col}(A)$. Therefore,

$$\text{im}(T) = \text{col}(A) = \text{span}\left(\begin{bmatrix} 1 \\ -1 \\ 3 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 3 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ 0 \\ 1 \end{bmatrix}\right)$$

By Theorem **??** and Definition **??**, we know that for an $m \times n$ matrix A, $\text{col}(A)$ is a subspace of $\mathbb{R}^m$. However, when vector spaces other than $\mathbb{R}^m$ are involved, it is not yet clear that $\text{im}(T)$ is a subspace of the codomain. The following theorem resolves this issue.

Theorem 50. Let $T : V \to W$ be a linear transformation. Then $\text{im}(T)$ is a subspace of W.

Proof To show that $\text{im}(T)$ is a subspace, we need to show that $\text{im}(T)$ is closed under addition and scalar multiplication.

Suppose $\mathbf{w}_1$ and $\mathbf{w}_2$ are in $\text{im}(T)$. Then there are vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ in V such that $T(\mathbf{v}_1) = \mathbf{w}_1$ and $T(\mathbf{v}_2) = \mathbf{w}_2$. Then

$$\mathbf{w}_1 + \mathbf{w}_2 = T(\mathbf{v}_1) + T(\mathbf{v}_2) = T(\mathbf{v}_1 + \mathbf{v}_2)$$

This shows that $\mathbf{w}_1 + \mathbf{w}_2$ is in $\text{im}(T)$.

For any scalar a, we have:

$$a\mathbf{w}_1 = aT(\mathbf{v}_1) = T(a\mathbf{v}_1)$$

This shows that $a\mathbf{w}_1$ is in $\text{im}(T)$.

■

We can now define the rank of a linear transformation.

Definition 60. The *rank* of a linear transformation $T : V \to W$, is the dimension of the image of T.

$$\text{rank}(T) = \dim(\text{im}(T))$$

This definition gives us the following relationship between the rank of a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ and the rank of the standard matrix A associated with it.

Formula 15.

$$\text{rank}(A) = \dim(\text{col}(A)) = \dim(\text{im}(T)) = \text{rank}(T)$$

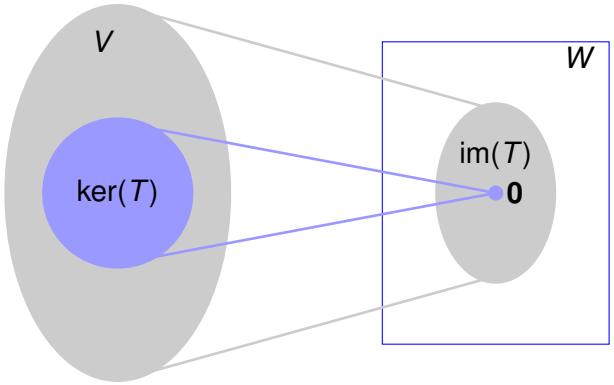
The Kernel of a Linear Transformation

Definition 61. Let V and W be vector spaces, and let $T : V \to W$ be a linear transformation. The *kernel* of T, denoted by $\ker(T)$, is the set

$$\ker(T) = \{\mathbf{v} : T(\mathbf{v}) = \mathbf{0}\}$$

In other words, the kernel of T consists of all vectors of V that map to $\mathbf{0}$ in W.

It is important to pay attention to the locations of the kernel and the image. We already proved that $\text{im}(T)$ is a subspace of the codomain. In contrast, $\ker(T)$ is located in the domain. (We will prove shortly that it is a subspace of the domain.)



Example 102. Let $T : \mathbb{R}^5 \to \mathbb{R}^4$ be a linear transformation with standard matrix

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

1. Find $\ker(T)$
2. Is $\ker(T)$ a subspace of $\mathbb{R}^5$? If so, find $\dim(\ker(T))$.

Explanation. ?? To find the kernel of T, we need to find all vectors of $\mathbb{R}^5$ that map to $\mathbf{0}$ in $\mathbb{R}^4$. This amounts to solving the equation $A\mathbf{x} = \mathbf{0}$.

Gauss-Jordan elimination yields:

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

Thus, the kernel of T consists of all elements of the form:

$$\begin{bmatrix} -1 \\ -1 \\ 2 \\ 1 \\ 0 \end{bmatrix} s + \begin{bmatrix} -2 \\ -1 \\ 2 \\ 0 \\ 1 \end{bmatrix} t$$

We conclude that

$$\ker(T) = \text{span}\left(\begin{bmatrix} -1 \\ -1 \\ 2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -2 \\ -1 \\ 2 \\ 0 \\ 1 \end{bmatrix} \right)$$

?? Since $\ker(T)$ is the span of two vectors of $\mathbb{R}^5$, we know that $\ker(T)$ is a subspace of $\mathbb{R}^5$. (See Theorem **??**.) Observe that the two vectors in the spanning set are linearly independent. (How can we see this without performing computations?) Therefore $\dim(\ker(T)) = 2$.

Recall that the *null space* of a matrix A is defined to be set of all solutions to the homogeneous equation $A\mathbf{x} = \mathbf{0}$. This means that if $T : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation with standard matrix A then

$$\ker(T) = \text{null}(A)$$

We know that $\text{null}(A)$ of an $m \times n$ matrix is a subspace of $\mathbb{R}^n$. (See Theorem **??**.) We conclude this section by showing that even when vector spaces other than $\mathbb{R}^n$ are involved, the kernel of a linear transformation is a subspace of the domain of the transformation.

Theorem 51. Let $T : V \to W$ be a linear transformation, then $\ker(T)$ is a subspace of V.

Proof To show that $\ker(T)$ is a subspace, we need to show that $\ker(T)$ is closed under addition and scalar multiplication.

Suppose that $\mathbf{v}_1$ and $\mathbf{v}_2$ are in $\ker(T)$. Then,

$$T(\mathbf{v}_1 + \mathbf{v}_2) = T(\mathbf{v}_1) + T(\mathbf{v}_2) = \mathbf{0} + \mathbf{0} = \mathbf{0}$$

This shows that $\mathbf{v}_1 + \mathbf{v}_2$ is in $\ker(T)$.

For any scalar a we have:

$$T(a\mathbf{v}_1) = aT(\mathbf{v}_1) = a\mathbf{0} = \mathbf{0}$$

This shows that $a\mathbf{v}_1$ is in $\ker(T)$.

■

Definition 62. The *nullity* of a linear transformation $T : V \to W$, is the dimension of the kernel of T.

$$\text{nullity}(T) = \dim(\ker(T))$$

This definition gives us the following relationship between nullity of a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ and the nullity of the standard matrix A associated with it.

Formula 16.

$$\text{nullity}(A) = \dim(\text{null}(A)) = \dim(\ker(T)) = \text{nullity}(T)$$

Rank-Nullity Theorem for Linear Transformations

In Examples **??** and **??**, we found the image and the kernel of the linear transformation $T : \mathbb{R}^5 \to \mathbb{R}^4$ with standard matrix

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

We also found that

$$\text{rank}(T) = \dim(\text{im}(T)) = \dim(\text{col}(A)) = \text{rank}(A) = 3$$

and

$$\text{nullity}(T) = \dim(\ker(T)) = \dim(\text{null}(A)) = \text{nullity}(A) = 2$$

Because of Rank-Nullity Theorem for matrices (Theorem **??**), it is not surprising that

$$\text{rank}(T) + \text{nullity}(T) = 3 + 2 = 5 = \dim(\mathbb{R}^5)$$

The following theorem is a generalization of this result.

Theorem 52. Let $T : V \to W$ be a linear transformation. Suppose $\dim(V) = n$, then

$$\text{rank}(T) + \text{nullity}(T) = n$$

Proof By Theorem **??**, $\text{im}(T)$ is a subspace of W. There exists a basis for $\text{im}(T)$ of the form $\{T(\mathbf{v}_1), ..., T(\mathbf{v}_r)\}$. By Theorem **??**, $\ker(T)$ is a subspace of V. Let $\{\mathbf{u}_1, ..., \mathbf{u}_s\}$ be a basis for $\ker(T)$.

We will show that $\{\mathbf{u}_1, ..., \mathbf{u}_s, \mathbf{v}_1, ..., \mathbf{v}_r\}$ is a basis for V.

For any vector $\mathbf{v}$ in V, we have:

$$T(\mathbf{v}) = c_1 T(\mathbf{v}_1) + ... + c_r T(\mathbf{v}_r)$$

for some scalars c_i $(1 \leq i \leq r)$. Thus,

$$T(\mathbf{v}) - \left(c_1 T(\mathbf{v}_1) + ... + c_r T(\mathbf{v}_r)\right) = \mathbf{0}$$

By linearity,

$$T((\mathbf{v} - (c_1\mathbf{v}_1 + ... + c_r\mathbf{v}_r)) = \mathbf{0}$$

Therefore $\mathbf{v} - (c_1\mathbf{v}_1 + ... + c_r\mathbf{v}_r)$ is in $\ker(T)$.

Hence there are scalars a_i $(1 \leq i \leq s)$ such that

$$\mathbf{v} - (c_1\mathbf{v}_1 + ... + c_r\mathbf{v}_r) = a_1\mathbf{u}_1 + ... + a_s\mathbf{u}_s$$

Thus,

$$\mathbf{v} = (c_1\mathbf{v}_1 + ... + c_r\mathbf{v}_r) + (a_1\mathbf{u}_1 + ... + a_s\mathbf{u}_s)$$

We conclude that

$$V = \text{span}(\mathbf{u}_1, ..., \mathbf{u}_s, \mathbf{v}_1, ..., \mathbf{v}_r)$$

Now we need to show that $\{\mathbf{u}_1, ..., \mathbf{u}_s, \mathbf{v}_1, ..., \mathbf{v}_r\}$ is linearly independent.

Suppose

$$c_1\mathbf{v}_1 + ... + c_r\mathbf{v}_r + a_1\mathbf{u}_1 + ... + a_s\mathbf{u}_s = \mathbf{0} \tag{59}$$

Applying T to both sides, we get

$$T(c_1\mathbf{v}_1 + ... + c_r\mathbf{v}_r + a_1\mathbf{u}_1 + ... + a_s\mathbf{u}_s) = T(\mathbf{0})$$

$$c_1T(\mathbf{v}_1) + ... + c_rT(\mathbf{v}_r) + a_1T(\mathbf{u}_1) + ... + a_sT(\mathbf{u}_s) = \mathbf{0}$$

But $T(\mathbf{u}_i) = \mathbf{0}$ for $1 \leq i \leq s$, thus

$$c_1T(\mathbf{v}_1) + ... + c_rT(\mathbf{v}_r) = \mathbf{0}$$

Since $\{T(\mathbf{v}_1), ..., T(\mathbf{v}_r)\}$ is linearly independent, it follows that each $c_i = 0$.

But then Equation (**??**) implies that $a_1\mathbf{u}_1 + ... + a_s\mathbf{u}_s = \mathbf{0}$. Because $\{\mathbf{u}_1, ..., \mathbf{u}_s\}$ is linearly independent, it follows that each $a_i = 0$.

We conclude that $\{\mathbf{u}_1, ..., \mathbf{u}_s, \mathbf{v}_1, ..., \mathbf{v}_r\}$ is a basis for V. Thus,

$$\dim(\ker(T)) + \dim(\text{im}(T)) = s + r = n$$

■

Practice Problems

Problem 283 *Describe the image and find the rank of the linear transformation* $T : \mathbb{R}^5 \to \mathbb{R}^2$ *induced by* $A = \begin{bmatrix} 3 & 2 & 4 & 7 & 1 \\ -1 & -9 & 7 & 6 & 8 \end{bmatrix}$.

Multiple Choice:

1. $im(T) = \mathbb{R}^2$
2. $im(T)$ *is a line in* $\mathbb{R}^2$
3. $im(T) = \{\mathbf{0}\}$
4. $im(T) = \mathbb{R}^5$
5. $im(T)$ *is a plane in* $\mathbb{R}^5$

$rank(T) =$ ... ...

Problem 284 *Describe the image and find the rank of the linear transformation* $T : \mathbb{R}^2 \to \mathbb{R}^3$ *induced by* $A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{bmatrix}$

Multiple Choice:

1. $im(T) = \mathbb{R}^3$
2. $im(T)$ *is a line in* $\mathbb{R}^2$
3. $im(T)$ *is a line in* $\mathbb{R}^3$
4. $im(T) = \{\mathbf{0}\}$
5. $im(T)$ *is a plane in* $\mathbb{R}^3$

$rank(T) =$ ... ...

Problem 285 *Suppose linear transformations* $T : \mathbb{R}^2 \to \mathbb{R}^2$ *and* $S : \mathbb{R}^2 \to \mathbb{R}^2$ *are such that* $im(T) = im(S) = span\left(\begin{bmatrix} 1 \\ -3 \end{bmatrix}\right)$. *Does this mean that* T *and* S *are the same transformation? Justify your claim.*

Problem 286 *Describe the kernel and find the nullity of the linear transformation* $T : \mathbb{R}^3 \to \mathbb{R}^2$ *induced by* $A = \begin{bmatrix} 2 & 1 & 0 \\ -1 & 1 & -3 \end{bmatrix}$.

Multiple Choice:

1. $ker(T) = \mathbb{R}^3$
2. $ker(T) = \{\mathbf{0}\}$
3. $ker(T) = \mathbb{R}^2$
4. $ker(T)$ *is a plane in* $\mathbb{R}^3$
5. $ker(T)$ *is a line in* $\mathbb{R}^3$

$nullity(T) =$ ... ...

Problem 287 *Describe the kernel and find the nullity of the linear transformation* $T : \mathbb{R}^2 \to \mathbb{R}^2$ *induced by* $A = \begin{bmatrix} 2 & -1 \\ 3 & 0 \end{bmatrix}$.

Multiple Choice:

1. $ker(T) = \mathbb{R}^2$
2. $ker(T) = \{\mathbf{0}\}$
3. $ker(T)$ *is a line in* $\mathbb{R}^2$

$nullity(T) =$ ... ...

Problem 288 *Describe the kernel and find the nullity of the linear transformation* $T : \mathbb{R}^3 \to \mathbb{R}^5$ *induced by* $A = \begin{bmatrix} 1 & 2 & -1 \\ 1 & 2 & -1 \\ 1 & 2 & -1 \\ 1 & 2 & -1 \\ 1 & 2 & -1 \end{bmatrix}$

Multiple Choice:

1. $ker(T)$ *is a plane in* $\mathbb{R}^3$
2. $ker(T)$ *is a line in* $\mathbb{R}^3$
3. $ker(T)$ *is a line in* $\mathbb{R}^5$
4. $ker(T) = \mathbb{R}^3$
5. $ker(T) = \{\mathbf{0}\}$

$nullity(T) =$ ... ...

Problem 289 *Suppose a linear transformation* $T : \mathbb{R}^3 \to \mathbb{R}^3$ *is such that* $im(T)$ *is a plane in* $\mathbb{R}^3$. *Then*

$$rank(T) = \text{... ...}$$

$$nullity(T) = \text{... ...}$$

Problem 290 *Suppose a linear transformation $T : \mathbb{R}^5 \to \mathbb{R}^5$ is such that $T(\mathbf{v}) = \mathbf{0}$ for all $\mathbf{v}$ in $\mathbb{R}^5$. Then*

$$rank(T) = \ldots\ldots$$

$$nullity(T) = \ldots\ldots$$

Problem 291 *Let $T : \mathbb{R}^6 \to \mathbb{R}^4$ be a linear transformation with standard matrix*

$$A = \begin{bmatrix} 2 & -1 & 1 & -2 & 1 & 1 \\ 1 & 2 & 3 & 6 & -4 & 1 \\ 0 & 2 & 2 & 4 & -2 & -1 \\ 1 & 3 & 2 & 6 & -3 & 2 \end{bmatrix}$$

Find $im(T)$ and $ker(T)$ if the reduced row-echelon form of A is

$$rref(A) = \begin{bmatrix} 1 & 0 & 0 & 1 & -1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Problem 292 *Let $V = span\left(\begin{bmatrix} 1 \\ 1 \end{bmatrix}\right)$, and let $T : V \to \mathbb{R}^2$ be a linear transformation defined by $T(\mathbf{v}) = 2\mathbf{v}$. Find $im(T)$ and $ker(T)$.*

Problem 293 *Suppose a linear transformation T is induced by a 4×6 matrix A. Let S be a linear transformation induced by A^T. Find $nullity(S)$, if $nullity(T) = 3$. Prove your claim.*

46 The Determinant

47 Finding the Determinant

In this section we will define a function that assigns to each square matrix A a scalar output called the *determinant of A*. We will denote the determinant of A by $\det A$. For a matrix with real number entries, the output of the determinant function will always be a real number.

One important property of the determinant is its connection to matrix inverses. We will find that a matrix A is singular if and only if $\det A = 0$.

Geometrically speaking, the determinant of a matrix of a linear transformation is the factor by which the area (or volume or hypervolume) is scaled by the transformation.

Cofactor Expansion Along the Top Row

To start from the beginning, let us define the determinant of a 1×1 matrix.

Definition 63. Let $A = \begin{bmatrix} a \end{bmatrix}$. Define the *determinant* of A by $\det A = a$.

It is important to note that this definition is consistent with our goal of making a connection between determinants and invertibility. Observe that $A^{-1} = \begin{bmatrix} a^{-1} \end{bmatrix}$ exists if and only if $a \neq 0$.

Now we proceed to 2×2 matrices. According to Formula **??**, the inverse of a nonsingular matrix $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is given by

$$A^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

Observe that A^{-1} exists if and only if $ad - bc \neq 0$. We will call the number $ad - bc$ the determinant of A.

Definition 64. Let $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$. The determinant of A is defined by

$$\det A = \det \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc \tag{60}$$

Note the distinction between the square bracket notation associated with the matrix $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ and the vertical bar notation $\begin{vmatrix} a & b \\ c & d \end{vmatrix}$ used to denote the determinant in expression (**??**).

Example 103.

$$\det \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} = \begin{vmatrix} 1 & 2 \\ 3 & 4 \end{vmatrix} = (1)(4) - (2)(3) = -2$$

The easiest way to understand the definition of the determinant for a 3×3 matrix is to start with an example.

Example 104. Find $\det A$ if

$$A = \begin{bmatrix} 3 & -2 & 1 \\ 5 & -1 & 2 \\ 1 & 4 & 1 \end{bmatrix}$$

Explanation. We will use three color-coded copies of the matrix A

$$A = \begin{bmatrix} 3 & -2 & 1 \\ 5 & -1 & 2 \\ 1 & 4 & 1 \end{bmatrix}, \quad A = \begin{bmatrix} 3 & -2 & 1 \\ 5 & -1 & 2 \\ 1 & 4 & 1 \end{bmatrix}, \quad A = \begin{bmatrix} 3 & -2 & 1 \\ 5 & -1 & 2 \\ 1 & 4 & 1 \end{bmatrix}$$

$$\begin{aligned} \det A &= (3) \begin{vmatrix} -1 & 2 \\ 4 & 1 \end{vmatrix} - (-2) \begin{vmatrix} 5 & 2 \\ 1 & 1 \end{vmatrix} + (1) \begin{vmatrix} 5 & -1 \\ 1 & 4 \end{vmatrix} \\ &= (3)(-1 - 8) - (-2)(5 - 2) + (1)(20 + 1) \\ &= -27 + 6 + 21 \\ &= 0 \end{aligned}$$

We now formalize what we learned in Example **??**.

Definition 65. Let

$$A = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}$$

The *determinant* of A is given by

$$\det A = |A| = a \begin{vmatrix} e & f \\ h & i \end{vmatrix} - b \begin{vmatrix} d & f \\ g & i \end{vmatrix} + c \begin{vmatrix} d & e \\ g & h \end{vmatrix} \tag{61}$$

We will now reiterate several important features of this definition and introduce some vocabulary:

- The coefficients a, b and c are the entries of the first row of matrix A. Coefficients in the formula follow an alternating sign pattern: $+a$, $-b$, $+c$. This pattern will persist in the determinant formulas for determinants of larger matrices.
- When using equation (**??**), we compute determinants of three matrices:

$$\begin{bmatrix} e & f \\ h & i \end{bmatrix}, \quad \begin{bmatrix} d & f \\ g & i \end{bmatrix}, \quad \begin{bmatrix} d & e \\ g & h \end{bmatrix}$$

  These matrices are called *minor* matrices. To form each minor matrix, cross out the row and column that the corresponding coefficient is in. For example, the minor matrix corresponding to coefficient b is found by crossing out the row and column that b is in.

$$\begin{bmatrix} \cancel{a} & b & \cancel{c} \\ d & \cancel{e} & f \\ g & \cancel{h} & i \end{bmatrix}$$

- The process for finding the determinant described in Definition **??** is referred to as a *cofactor expansion along the top row*.

Example 105. Find $\det A$ if

$$A = \begin{bmatrix} 4 & 3 & -2 \\ 1 & -5 & 3 \\ -4 & 1 & 1 \end{bmatrix}$$

Explanation.

$$\det A = (4)\begin{vmatrix} \dots & \dots \\ \dots & \dots \end{vmatrix} - (3)\begin{vmatrix} \dots & \dots \\ \dots & \dots \end{vmatrix} + (-2)\begin{vmatrix} \dots & \dots \\ \dots & \dots \end{vmatrix} = \dots \dots$$

We are starting to observe a certain pattern in the process of computing the determinant. This pattern will persist for larger matrices. Let's take a look at a 4×4 matrix.

Example 106. Find $\det A$ if

$$A = \begin{bmatrix} 2 & 3 & -2 & -5 \\ 0 & 1 & -2 & 0 \\ 1 & 3 & 0 & -1 \\ 2 & 0 & 1 & 1 \end{bmatrix}$$

Explanation. As you watch the video below, pay particular attention to the same patterns as you saw in the case of 3×3 matrices: the alternating sign pattern and the process of forming minor matrices.

$$\begin{aligned}\det A &= 2\begin{vmatrix}1 & -2 & 0\\ 3 & 0 & -1\\ 0 & 1 & 1\end{vmatrix} - 3\begin{vmatrix}0 & -2 & 0\\ 1 & 0 & -1\\ 2 & 1 & 1\end{vmatrix} - 2\begin{vmatrix}0 & 1 & 0\\ 1 & 3 & -1\\ 2 & 0 & 1\end{vmatrix} - (-5)\begin{vmatrix}0 & 1 & -2\\ 1 & 3 & 0\\ 2 & 0 & 1\end{vmatrix}\\ &= 2(7) - 3(6) - 2(-3) + 5(11)\\ &= 57\end{aligned}$$

Example 107. Find $\det A$ if

$$A = \begin{bmatrix}4 & -1 & 2 & 1\\ 3 & 0 & 1 & -2\\ 2 & 1 & 5 & 1\\ -2 & 1 & 3 & -1\end{bmatrix}$$

Explanation. We will use the entries in the top row as coefficients in front of 3×3 determinants. As before, we will use the alternating sign pattern for the coefficients:

$$+(4), -(-1), +(2), -(1)$$

Just like in the case of a 3×3 matrix in Example **??**, each of the smaller determinants is obtained by crossing out the row and the column where the coefficient is located.

$$\begin{aligned}\det A &= 4\begin{vmatrix}0 & 1 & -2\\ 1 & 5 & 1\\ 1 & 3 & -1\end{vmatrix} + \begin{vmatrix}3 & 1 & -2\\ 2 & 5 & 1\\ -2 & 3 & -1\end{vmatrix} + 2\begin{vmatrix}3 & 0 & -2\\ 2 & 1 & 1\\ -2 & 1 & -1\end{vmatrix} - \begin{vmatrix}3 & 0 & 1\\ 2 & 1 & 5\\ -2 & 1 & 3\end{vmatrix}\\ &= 4(\ldots\ \ldots) + (\ldots\ \ldots) + 2(\ldots\ \ldots) - (\ldots\ \ldots)\\ &= \ldots\ \ldots\end{aligned}$$

Cofactor Expansion Along the First Column

We defined the determinant of a matrix in terms of cofactor expansion along the top row. We will now see what happens when we expand along the first *column* instead. We will refer to this process as *cofactor expansion along the first column*. Surprisingly, both expansions yield the same result. To illustrate this, let's revisit Examples **??** and **??**.

Example 108. Let

$$A = \begin{bmatrix} 3 & -2 & 1 \\ 5 & -1 & 2 \\ 1 & 4 & 1 \end{bmatrix}$$

In Example **??** we found that $\det A = 0$. Let's try to mimic what we did earlier, but instead of expanding along the first row, we will expand along the fist column.

$$\begin{aligned} &(3)\begin{vmatrix} -1 & 2 \\ 4 & 1 \end{vmatrix} - (5)\begin{vmatrix} -2 & 1 \\ 4 & 1 \end{vmatrix} + (1)\begin{vmatrix} -2 & 1 \\ -1 & 2 \end{vmatrix} \\ &= (3)(-1-8) - (5)(-2-4) + (1)(-4+1) \\ &= -27 + 30 - 3 \\ &= 0 \\ &= \det A \end{aligned}$$

Let's go through this process again for a larger matrix.

Exploration 294 *Let*

$$A = \begin{bmatrix} 4 & -1 & 2 & 1 \\ 3 & 0 & 1 & -2 \\ 2 & 1 & 5 & 1 \\ -2 & 1 & 3 & -1 \end{bmatrix}$$

In Example **??** *we found that* $\det A = -58$*. We will now try to expand along the fist column.*

When computing determinants of the four 3×3 *matrices below, try different approaches. You might want to expand along the first row for some of them, and along the first column for others. Looking for where zeros are located will help you decide what to try.*

$$\begin{aligned} &4\begin{vmatrix} 0 & 1 & -2 \\ 1 & 5 & 1 \\ 1 & 3 & -1 \end{vmatrix} - 3\begin{vmatrix} -1 & 2 & 1 \\ 1 & 5 & 1 \\ 1 & 3 & -1 \end{vmatrix} + 2\begin{vmatrix} -1 & 2 & 1 \\ 0 & 1 & -2 \\ 1 & 3 & -1 \end{vmatrix} - (-2)\begin{vmatrix} -1 & 2 & 1 \\ 0 & 1 & -2 \\ 1 & 5 & 1 \end{vmatrix} \\ &= 4(\ldots\ldots) - 3(\ldots\ldots) + 2(\ldots\ldots) + 2(\ldots\ldots) \\ &= \ldots\ldots \\ &= \det A \end{aligned}$$

In Example **??** and Exploration **??** we were careful not to claim at the outset that we were finding the determinant of the matrix by cofactor expansion along the first column; we merely observed that the resulting value was equal to the determinant. It is possible to prove that both expansions produce the same result (Theorem **??**). Therefore the determinant of a matrix can be defined in terms of cofactor expansion along the first row or column.

Cofactor Expansion Along Any Row or Column

We originally defined the determinant of a matrix via expansion along the top row of the matrix. We later observed that expansion along the first column produces the same result. It turns out that the value of the determinant can be computed by expanding along any row or column. This result is known as the *Laplace Expansion Theorem* (**??**).

When expanding along an arbitrary row or column, we will continue to follow the two patterns we observed earlier.

- The alternating sign pattern for coefficients will follow the checkerboard pattern below.

$$\begin{bmatrix} + & - & + & - & + & \dots \\ - & + & - & + & - & \dots \\ + & - & + & - & + & \dots \\ - & + & - & + & - & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

- Minor matrices will be formed by eliminating the row and column that the corresponding coefficient is in.

To illustrate this, let's take another look at matrix A from Example **??**.

Example 109. Let

$$A = \begin{bmatrix} 4 & -1 & 2 & 1 \\ 3 & 0 & 1 & -2 \\ 2 & 1 & 5 & 1 \\ -2 & 1 & 3 & -1 \end{bmatrix}$$

Follow the rules described above to expand along the second row. Compare your result with the determinant you found in Example **??**.

Explanation. The second row has the advantage over other rows in that it contains a zero. This will simplify our calculations. Following the checkerboard sign pattern along the second row we get

$$\begin{aligned} \det A &= -3\begin{vmatrix} -1 & 2 & 1 \\ 1 & 5 & 1 \\ 1 & 3 & -1 \end{vmatrix} - \begin{vmatrix} 4 & -1 & 1 \\ 2 & 1 & 1 \\ -2 & 1 & -1 \end{vmatrix} + (-2)\begin{vmatrix} 4 & -1 & 2 \\ 2 & 1 & 5 \\ -2 & 1 & 3 \end{vmatrix} \\ &= -3(\dots\ \dots) - (\dots\ \dots) - 2(\dots\ \dots) \\ &= \dots\ \dots \end{aligned}$$

This answer is the same as the answer we got using expansion along the first row in Example **??**.

It is clear that having zeros as entries in the matrix reduces the number of computations necessary to find the determinant. The following example demonstrates how to use zeros to our advantage.

Example 110. Find $\det A$ if

$$A = \begin{bmatrix} 4 & 0 & 0 & 0 & 2 \\ 0 & -1 & 1 & 0 & 0 \\ 2 & 0 & 0 & -5 & 3 \\ 0 & 1 & 4 & 0 & -1 \\ 1 & 1 & 5 & 0 & 0 \end{bmatrix}$$

Explanation. The fourth column contains the most zeros, so we will expand along that column. The $(3,4)$-entry is the only non-zero entry in the fourth column. Following the checkerboard pattern, we see that the sign in front of -5 is a minus.

$$\det A = -(-5)\begin{vmatrix} 4 & 0 & 0 & 2 \\ 0 & -1 & 1 & 0 \\ 0 & 1 & 4 & -1 \\ 1 & 1 & 5 & 0 \end{vmatrix}$$

Next we will expand the minor matrix along the top row.

$$\det A = 5\left(4\begin{vmatrix} -1 & 1 & 0 \\ 1 & 4 & -1 \\ 1 & 5 & 0 \end{vmatrix} - 2\begin{vmatrix} 0 & -1 & 1 \\ 0 & 1 & 4 \\ 1 & 1 & 5 \end{vmatrix}\right)$$

Try the next step on your own. We suggest that you expand the first matrix along the last column and expand the second matrix along the first column.

$$\det(A) = 5\big(4(\ldots\ldots) - 2(\ldots\ldots)\big) = \ldots\ldots$$

A Note on Equivalency

We initially introduced the determinant of a matrix via cofactor expansion along the top row. We later observed that cofactor expansion along any row or column produces the same result. We have to be careful, however, not to use a few examples as "proof" that all cofactor expansions are equivalent. Such claims need to be carefully supported with general proofs. Unfortunately, in this case, the proofs are tedious and conceptually unenlightening. An interested reader can find them in Tedious Proofs Concerning Determinants.

Determinants of Some Special Matrices

We know that we can find the determinant of a matrix by cofactor expansion along the top row or the first column. (See Theorem **??** of Tedious Proofs Concerning Determinants for proof.) This property gives rise to a useful result.

Theorem 53. Let A be a square matrix, then

$$\det A^T = \det A$$

Proof See Practice Problem **??**. ■

As we observed earlier, having zeros in a matrix makes it easier for us to compute its determinant. Recall that that a square matrix is *upper-triangular* if all of the entries below the main diagonal are zero. Similarly, a square matrix is called *lower-triangular* if all of the entries above the main diagonal are zero. Together, upper and lower triangular matrices are categorized as *triangular* matrices.

Theorem 54. If A is a triangular matrix, then $\det A$ is equal to the product of its diagonal entries.

Proof We proceed by induction on n, where A is an $n \times n$ matrix. It is easy to see that this result holds for $n = 1, 2$. Suppose that the result holds for $(n-1) \times (n-1)$ triangular matrices. We need to show that it holds for $n \times n$ triangular matrices.

Suppose $A = [a_{ij}]$ is a triangular matrix. Then, with the exception of a_{11}, the entries in the first row (or column) of A are guaranteed to be zeros. We will take advantage of these zeros and expand along the first row (or column) of A. As we do so, we obtain a single product of a_{11} and the determinant of a minor matrix obtained by crossing out the first row and column of A. But this minor $(n-1) \times (n-1)$ matrix is also a triangular matrix with diagonal etries $a_{22}, a_{33}, \ldots, a_{nn}$. By induction hypothesis, its determinant is equal to the product of its diagonal entries, $a_{22} \cdot a_{33} \cdot \ldots \cdot a_{nn}$. Therefore

$$\det A = a_{11}(a_{22} \cdot a_{33} \cdot \ldots \cdot a_{nn})$$

This completes the proof. ■

As an immediate consequence of this theorem, we have the following result.

Corollary 6. Let I be the identity matrix, then

$$\det I = 1$$

We first introduced block matrices in Block Matrix Multiplication. Matrices of the form $\begin{bmatrix} A & C \\ O & B \end{bmatrix}$ and $\begin{bmatrix} A & O \\ D & B \end{bmatrix}$, where A, B are square matrices and O is the zero matrix, are said to be *block triangular*. The following theorem makes it easy to compute determinants of such matrices.

Theorem 55. Consider block triangular matrices $\begin{bmatrix} A & C \\ O & B \end{bmatrix}$ and $\begin{bmatrix} A & O \\ D & B \end{bmatrix}$, where A and B are square matrices. Then

$$\det \begin{bmatrix} A & C \\ O & B \end{bmatrix} = \det A \det B \quad \text{and} \quad \det \begin{bmatrix} A & O \\ D & B \end{bmatrix} = \det A \det B$$

Proof Write $T = \begin{bmatrix} A & C \\ O & B \end{bmatrix}$ and proceed by induction on k, where A is $k \times k$. If $k = 1$, then the result follows from cofactor expansion along the first column. In general, let $S_i(T)$ denote the matrix obtained from T by deleting row i and column 1. Then the cofactor expansion along the first column is

$$\det T = a_{11} \det S_1(T) - a_{21} \det S_2(T) + \cdots + (-1)^{k+1} \det S_k(T)$$

where $a_{11}, a_{21}, \ldots, a_{k1}$ are the entries in the first column of A. Observe that

$$S_i(T) = \begin{bmatrix} S_i(A) & C_i \\ O & B \end{bmatrix}$$

where $i = 1, 2, \ldots, k$, $S_i(A)$ denotes matrix A with column 1 and row i deleted, and C_i denotes matrix C with with row i deleted.

Since $S_i(A)$ is a $(k-1) \times (k-1)$ matrix, by the induction hypothesis,

$$\det S_i(T) = \det S_i(A) \cdot \det B$$

This gives us

$$\begin{aligned} \det T =& a_{11} \det S_1(T) - a_{21} \det S_2(T) + \cdots + (-1)^{k+1} a_{k1} \det S_k(T) \\ =& a_{11} \det S_1(A) \cdot \det B - a_{21} \det S_2(A) \cdot \det B + \ldots \\ & \cdots + (-1)^{k+1} a_{k1} \det S_k(A) \cdot \det B \\ =& \Big(a_{11} \det S_1(A) - a_{21} \det S_2(A) + \cdots + (-1)^{k+1} a_{k1} \det S_k(A) \Big) \det B \\ =& \det A \det B \end{aligned}$$

The lower triangular case is similar. ■

Example 111. Find det A if

$$A = \begin{bmatrix} 2 & 1 & 3 & 3 \\ 1 & -1 & -2 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 4 & 1 \end{bmatrix}$$

Explanation.

$$\begin{vmatrix} 2 & 1 & 3 & 3 \\ 1 & -1 & -2 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 4 & 1 \end{vmatrix} = \begin{vmatrix} 2 & 1 \\ 1 & -1 \end{vmatrix} \begin{vmatrix} 1 & 1 \\ 4 & 1 \end{vmatrix} = \ldots \ldots$$

Practice Problems

Problem 295 *Find the determinant.*

$$A = \begin{bmatrix} 4 & -2 \\ 3 & 7 \end{bmatrix}$$

Answer:

$$\det A = \ldots \ldots$$

Problem 296 *Find the determinant.*

$$B = \begin{bmatrix} 5 & -1 & 0 \\ 0 & 3 & -2 \\ 1 & -1 & 2 \end{bmatrix}$$

Answer:

$$det(B) = \ldots \ldots$$

Problem 297 *Find the determinant.*

$$C = \begin{bmatrix} 1 & -2 & 0 & 0 & 0 \\ 0 & -4 & 1 & 1 & 0 \\ 3 & 0 & -1 & 0 & 1 \\ 0 & 0 & 4 & 1 & 0 \\ -1 & -2 & 0 & 0 & 0 \end{bmatrix}$$

Answer:

$$\det(C) = \ldots \ldots$$

Problem 298 *Prove that for a square matrix A,* $\det A^T = \det A$. *(Theorem* **??***)*

Problem 299 *Let*

$$A = \begin{bmatrix} -2 & 1 & 3 \\ 0 & 4 & 0 \\ -1 & 1 & 2 \end{bmatrix}$$

1. $$detA = \ldots \ldots$$

2. *Let A_1 be a matrix obtained from A by switching the first and the second row of A.*

   $$detA_1 = \ldots \ldots$$

   What do you observe?

3. *Let A_2 be a matrix obtained from A by multiplying the bottom row by -3.*

   $$detA_2 = \ldots \ldots$$

   What do you observe?

4. *Let A_3 be a matrix obtained from A by multiplying A by 2.*

   $$detA_3 = \ldots \ldots$$

   What do you observe? Relate your observations to what you found in the previous part.

5. *Let A_4 be a matrix obtained from A by adding twice the third row to the first.*

   $$detA_4 = \ldots \ldots$$

   What do you observe?

Problem 300 *Is it true that* $\det(A + B) = \det A + \det B$*?*

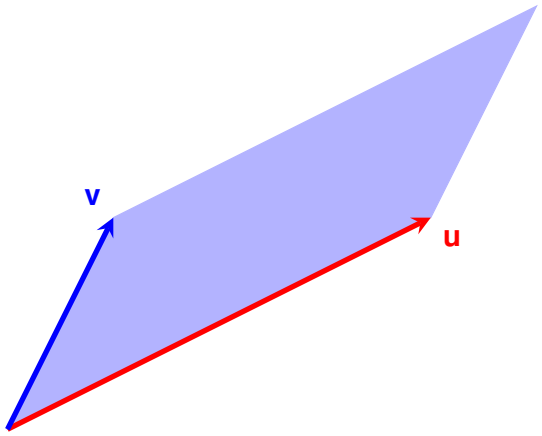
Text Source

The text in this section partially comes from Section 3.1 of Keith Nicholson's *Linear Algebra with Applications* (CC-BY-NC-SA).

48 Determinants, Areas, and Volumes

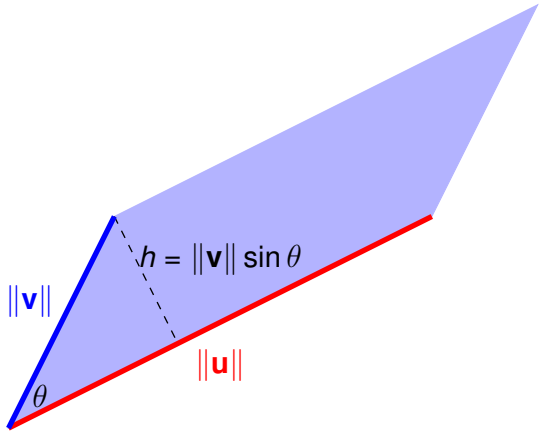
2×2 Determinant and the Area of a Parallelogram

Consider the parallelogram determined by vectors **u** and **v** in $\mathbb{R}^3$.



Recall that the area of a parallelogram is given by the product of the length of the base and the height. As shown in the diagram below, the length of the base is the magnitude of **u**. The height, h, can be found using trigonometry

$$h = \|\mathbf{v}\| \sin\theta$$



Using the area of a parallelogram formula together with Theorem **??** we get

$$\text{Area} = (\text{base})(\text{height}) = \|\mathbf{u}\|\, h = \|\mathbf{u}\|\, \|\mathbf{v}\| \sin\theta = \|\mathbf{u} \times \mathbf{v}\|$$

We have established the following formula.

Formula 17. The area of a parallelogram determined by vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^3$ is given by

$$\text{Area} = \|\mathbf{u} \times \mathbf{v}\|$$

Example 112. Use Formula **??** to find the area of a parallelogram determined by vectors

$$\mathbf{u} = \begin{bmatrix} -2 \\ 1 \\ 3 \end{bmatrix}, \quad \mathbf{v} = \begin{bmatrix} 1 \\ 6 \\ 2 \end{bmatrix}$$

Explanation. We can start by visualizing the parallelogram in GeoGebra. RIGHT-CLICK and DRAG to rotate the image below. The area of the parallelogram, rounded to two decimal places, is displayed inside the parallelogram.

Access GeoGebra interactives through the online version of this text at

https://ximera.osu.edu/linearalgebradzv3.

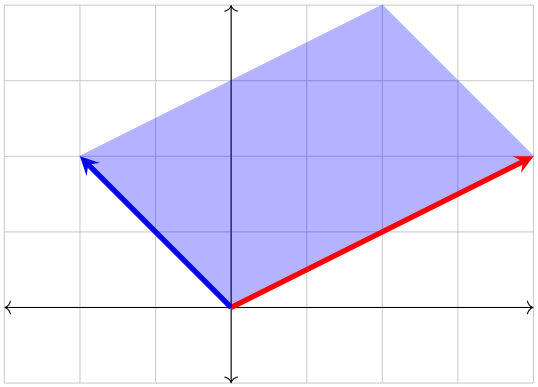
To find the exact area we compute

$$\mathbf{u} \times \mathbf{v} = \begin{bmatrix} -2 \\ 1 \\ 3 \end{bmatrix} \times \begin{bmatrix} 1 \\ 6 \\ 2 \end{bmatrix} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ -2 & 1 & 3 \\ 1 & 6 & 2 \end{vmatrix} = \begin{bmatrix} \dots \dots \\ \dots \dots \\ \dots \dots \end{bmatrix}$$

$$\text{Area} = \|\mathbf{u} \times \mathbf{v}\| = \sqrt{\dots \dots}$$

Formula **??** can be easily adapted to parallelograms determined by vectors in $\mathbb{R}^2$, as illustrated by the following example.

Example 113. Find the area of the parallelogram in the diagram.



Explanation. The vectors that determine the parallelogram are

$$\begin{bmatrix} 4 \\ 2 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} -2 \\ 2 \end{bmatrix}$$

The problem we run into is that these vectors are in $\mathbb{R}^2$, whereas the cross product is defined only for vectors in $\mathbb{R}^3$. We will get around this difficulty by "padding" our vectors with zeros on the bottom. In other words, we will consider them as vectors sitting in the xy-coordinate plane in $\mathbb{R}^3$. This allows us to compute the cross product

$$\begin{bmatrix} 4 \\ 2 \\ 0 \end{bmatrix} \times \begin{bmatrix} -2 \\ 2 \\ 0 \end{bmatrix} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ 4 & 2 & 0 \\ -2 & 2 & 0 \end{vmatrix} = \mathbf{k}\Big((4)(2) - (2)(-2)\Big) = 12\mathbf{k} = \begin{bmatrix} 0 \\ 0 \\ 12 \end{bmatrix}$$

The area of the parallelogram is then given by

$$\text{Area} = \left\| \begin{bmatrix} 0 \\ 0 \\ 12 \end{bmatrix} \right\| = 12$$

Example **??** illustrates an important phenomenon. Observe that the zeros in the last column of the determinant ensure that the **i** and **j** components of the cross product are zero, while the last component is the determinant of the 2×2 matrix whose rows (or columns) are the two vectors that determine the parallelogram in $\mathbb{R}^2$. In general, if the parallelogram is determined by vectors

$$\begin{bmatrix} a \\ b \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} c \\ d \end{bmatrix}$$

then the area of the parallelogram can be computed as follows:

$$\begin{bmatrix} a \\ b \\ 0 \end{bmatrix} \times \begin{bmatrix} c \\ d \\ 0 \end{bmatrix} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ a & b & 0 \\ c & d & 0 \end{vmatrix} = \mathbf{k} \begin{vmatrix} a & b \\ c & d \end{vmatrix} = \mathbf{k}\Big((a)(d) - (b)(c)\Big) = \begin{bmatrix} 0 \\ 0 \\ ad - bc \end{bmatrix}$$

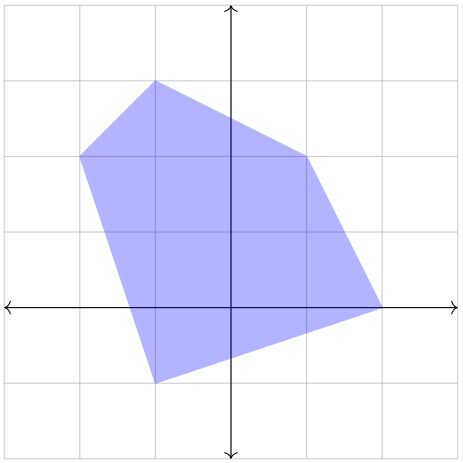
$$\text{Area} = \left\| \begin{bmatrix} 0 \\ 0 \\ ad - bc \end{bmatrix} \right\| = |ad - bc| = \left| \det \begin{bmatrix} a & b \\ c & d \end{bmatrix} \right| = \left| \det \begin{bmatrix} a & c \\ b & d \end{bmatrix} \right|$$

So the area of the parallelogram turns out to be the absolute value of the determinant of the matrix whose rows (or columns) are the two vectors that determine the parallelogram. The following formula summarizes our discussion.

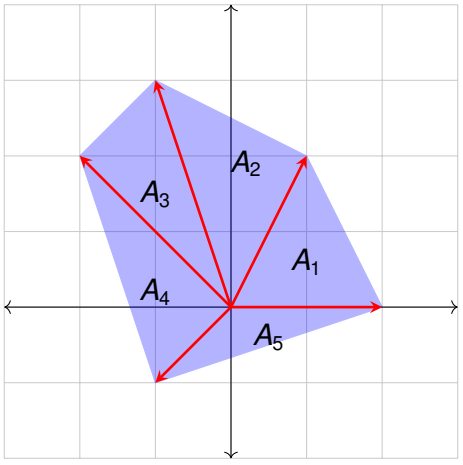
Formula 18. Let $\mathbf{u} = \begin{bmatrix} a \\ b \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} c \\ d \end{bmatrix}$ be vectors of $\mathbb{R}^2$. The area of the parallelogram determined by $\mathbf{u}$ and $\mathbf{v}$ is given by

$$\text{Area} = \left|\det \begin{bmatrix} a & b \\ c & d \end{bmatrix}\right| = \left|\det \begin{bmatrix} a & c \\ b & d \end{bmatrix}\right|$$

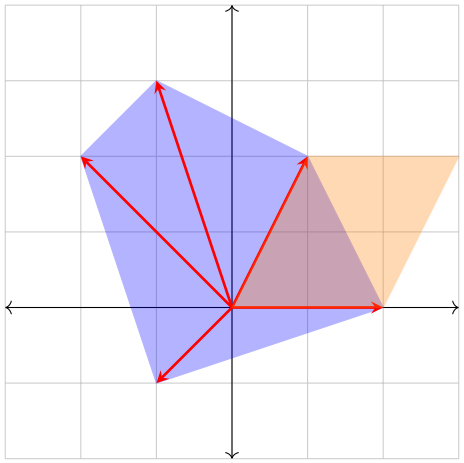
Example 114. Use Formula **??** to find the area of the polygon shown below.



Explanation. We will start by splitting this region into triangles.



We can find the total area of the polygon by finding the area of each triangle. The area of each triangle is half of the area of the corresponding parallelogram. For instance, A_1 is half of the area of the parallelogram depicted below.



We compute

$$A_1 = \frac{1}{2}\left|\det\begin{bmatrix}2 & 1\\ 0 & 2\end{bmatrix}\right| = \ldots\ldots$$

$$A_2 = \frac{1}{2}\left|\det\begin{bmatrix}1 & -1\\ 2 & 3\end{bmatrix}\right| = \ldots\ldots$$

$$A_3 = \frac{1}{2}\left|\det\begin{bmatrix}-1 & -2\\ 3 & 2\end{bmatrix}\right| = \ldots\ldots$$

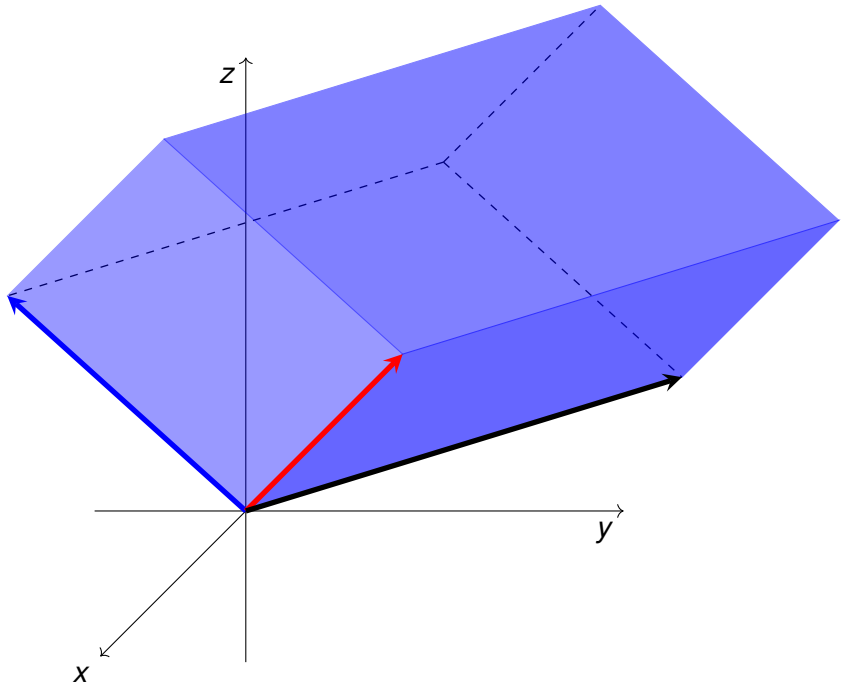
$$A_4 = \frac{1}{2}\left|\det\begin{bmatrix}-2 & -1\\ 2 & -1\end{bmatrix}\right| = \ldots\ldots$$

$$A_5 = \frac{1}{2}\left|\det\begin{bmatrix}-1 & 2\\ -1 & 0\end{bmatrix}\right| = \ldots\ldots$$

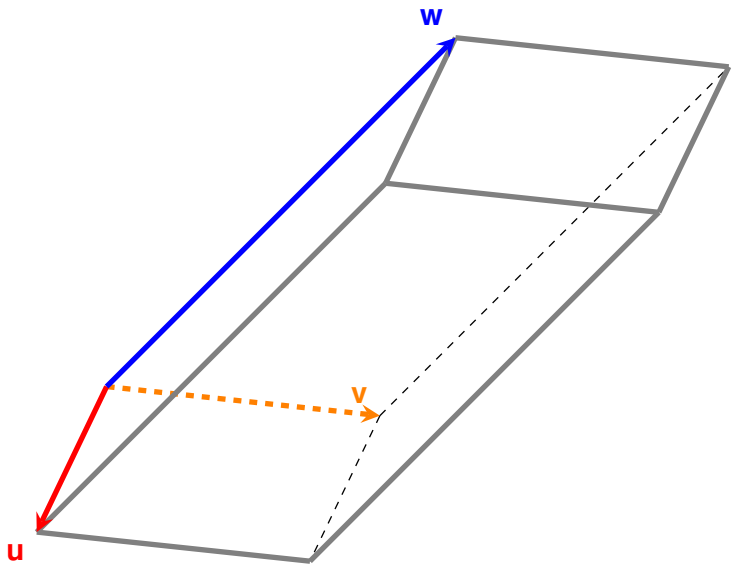
The total area of the polygon is ..... .

3×3 Determinant and the Volume of a Parallelepiped

Our next goal is to find the volume of a three-dimensional figure called a *parallelepiped*. A parallelepiped is a six-faced figure whose opposite faces are congruent parallelograms located in parallel planes. A parallelepiped is a three-dimensional counterpart of a parallelogram, and is determined by three non-coplanar vectors in $\mathbb{R}^3$. The figure below shows a parallelepiped determined by three vectors.



Consider a parallelepiped determined by vectors **u**, **v** and **w**, as shown below.

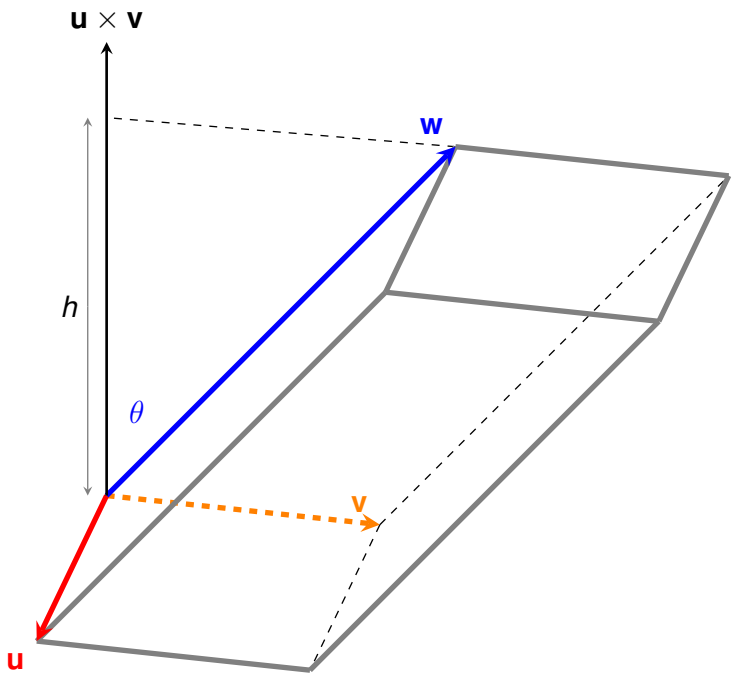


The volume of a parallelepiped is given by

$$\text{Volume} = (\text{area of base})(\text{height})$$

We will consider the parallelogram determined by **u** and **v** to be the base of the parallelepiped. Thus, the area of the base is given by

$$\text{Area of Base} = \|\mathbf{u} \times \mathbf{v}\|$$



The height of the parallelepiped is measured along a line perpendicular to the base. By Theorem **??**, $\mathbf{u} \times \mathbf{v}$ lies on such a line. Let θ be the angle between **w** and $\mathbf{u} \times \mathbf{v}$, $0 \leq \theta \leq \pi$. Then the height, h, of the parallelepiped is given by

$$h = \|\mathbf{w}\| \, |\cos\theta|$$

It may be difficult to visualize this in two dimensions. Below is a replica of of the above diagram in GeoGebra. RIGHT-CLICK and DRAG to rotate the image.

Access GeoGebra interactives through the online version of this text at

https://ximera.osu.edu/linearalgebradzv3.

This gives us the following formula for the volume of the parallelepiped

$$\text{Volume} = \|\mathbf{u} \times \mathbf{v}\| \, \|\mathbf{w}\| \, |\cos\theta| = |(\mathbf{u} \times \mathbf{v}) \cdot \mathbf{w}|$$

We have established the following formula.

Formula 19. The volume of a parallelepiped determined by vectors **u**, **v** and **w** in $\mathbb{R}^3$ is given by

$$\text{Volume} = |(\mathbf{u} \times \mathbf{v}) \cdot \mathbf{w}|$$

Our next goal is to show that this expression for the volume is equal to the determinant of a 3×3 matrix whose rows are the vectors that determine the parallelepiped.

Let

$$\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix}, \quad \mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}, \quad \mathbf{w} = \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix}$$

then

$$(\mathbf{u} \times \mathbf{v}) \cdot \mathbf{w} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{vmatrix} \cdot \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} = \begin{vmatrix} w_1 & w_2 & w_3 \\ u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{vmatrix} \tag{62}$$

The expression in (**??**) is sometimes referred to as the *box product* or the *scalar triple product*.

Recall that $\det(A) = \det(A^T)$ (Theorem **??**). Therefore, the three vectors that determine the parallelogram can be used to form rows or columns of the determinant on the right side of (**??**). This gives us the following formula.

Formula 20. Let $\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix}, \quad \mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}, \quad \mathbf{w} = \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix}$ be vectors in $\mathbb{R}^3$. Then the volume of the parallelepiped determined by **u**, **v** and **w** is given by

$$\text{Volume} = \left| \det \begin{bmatrix} w_1 & w_2 & w_3 \\ u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{bmatrix} \right| = \left| \det \begin{bmatrix} w_1 & u_1 & v_1 \\ w_2 & u_2 & v_2 \\ w_3 & u_3 & v_3 \end{bmatrix} \right|$$

Determinants and Linear Transformations

We will now turn our attention to the determinant of a matrix of a linear transformation.

Exploration 301 *The following GeoGebra interactive shows a polygon P located in the domain of a linear transformation T induced by the matrix M. The right-hand side shows the image of P under T. The number inside each polygon indicates its area.*

Access GeoGebra interactives through the online version of this text at

https://ximera.osu.edu/linearalgebradzv3.

Question 301.1 *Let* $M = \begin{bmatrix} 1 & 1 \\ -1 & 2 \end{bmatrix}$*. Find the determinant of M.*

$$\det M = \ldots \ldots$$

Drag the vertices of P to change the polygon. Make a note of how the area of P and the area of the image change. How are the areas related to each other?

$$Area(T(P)) = \ldots \ldots Area(P)$$

Question 301.2 *Change the matrix M to a matrix whose determinant is 1. Compare the areas of P and T(P). Try matrices whose determinant is 0 or negative. What do you observe about the areas?*

Formulate a conjecture about the relationship between the area of the polygon and the area of its image under a linear transformation.

We will not prove your conjecture in Exploration **??** for arbitrary figures as it is beyond the scope of this text. However, we can tackle the problem of how linear transformations affect areas of parallelograms. This is the topic of our next example.

Example 115. Let $T : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$ be a linear transformation induced by matrix M. Suppose $\mathbf{u}$ and $\mathbf{v}$ are vectors in $\mathbb{R}^2$. Let P be a parallelogram determined by $\mathbf{u}$ and $\mathbf{v}$. Show that

$$\text{Area}(T(P)) = |\det M| \, \text{Area}(P)$$

Explanation. Let $\mathbf{u} = \begin{bmatrix} a \\ b \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} c \\ d \end{bmatrix}$, and let $M = \begin{bmatrix} m & n \\ p & q \end{bmatrix}$. By Formula **??**, $\text{Area}(P) = ad - bc$. Applying M to $\mathbf{u}$ and $\mathbf{v}$ we get

$$T(\mathbf{u}) = \begin{bmatrix} am + bn \\ ap + bq \end{bmatrix} \quad \text{and} \quad T(\mathbf{v}) = \begin{bmatrix} cm + dn \\ cp + dq \end{bmatrix}$$

Using Formula **??**, we compute

$$\begin{aligned} \text{Area}(T(P)) &= |(am + bn)(cp + dq) - (ap + bq)(cm + dn)| \\ &= |(mq - np)(ad - bc)| \\ &= |\det M| \text{Area}(P) \end{aligned}$$

Practice Problems

Problem 302 *Let S be a square determined by $\begin{bmatrix}2\\0\end{bmatrix}$ and $\begin{bmatrix}0\\2\end{bmatrix}$. Let P be a parallelogram determined by vectors $\begin{bmatrix}2\\5\end{bmatrix}$ and $\begin{bmatrix}0\\2\end{bmatrix}$.*

1. *Sketch both figures in the same coordinate plane, and use geometry to explain why S and P have the same area. Compute the area of P using Formula* **??**.

   Area of P = ...

2. *Suppose M is the standard matrix of a linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$ such that $T(S) = P$. Find* det M.

   det M = ...

Problem 303 *Find the volume of a parallelepiped determined by*

$$\begin{bmatrix}0\\5\\4\end{bmatrix} \quad \begin{bmatrix}3\\1\\2\end{bmatrix} \quad \begin{bmatrix}1\\1\\6\end{bmatrix}$$

Answer: Volume = ...

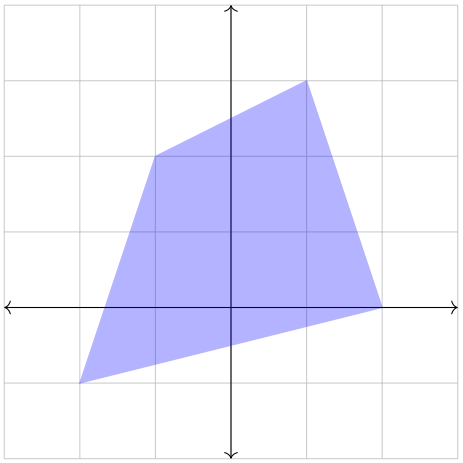
Problem 304 *Find the volume of a parallelepiped determined by*

$$\begin{bmatrix}1\\4\\-1\end{bmatrix} \quad \begin{bmatrix}3\\-2\\4\end{bmatrix} \quad \begin{bmatrix}5\\6\\2\end{bmatrix}$$

Explain your result geometrically.

Answer: Volume = ...

Problem 305 *Find the area of the figure using determinants.*



Answer: Area = ...

49 Elementary Row Operations and the Determinant

When we first introduced the determinant we motivated its definition for a 2×2 matrix by the fact that the value of the determinant is zero if and only if the matrix is singular. We will soon be able to generalize this result to larger matrices, and will eventually establish a formula for the inverse of a nonsingular matrix in terms of determinants.

Recall that we can find the inverse of a matrix or establish that the inverse does not exist by using elementary row operations to carry the given matrix to its reduced row-echelon form. In order to start relating determinants to inverses we need to find out what elementary row operations do to the determinant of a matrix.

The Effects of Elementary Row Operations on the Determinant

Recall that there are three elementary row operations:

1. Switching the order of two rows
2. Multiplying a row by a non-zero constant
3. Adding a multiple of one row to another

Elementary row operations are used to carry a matrix to its reduced row-echelon form. In Practice Problem **??** we established that elementary row operations are reversible. In other words, if we know what elementary row operations carried A to rref(A), we can undo each operation with another elementary row operation to carry rref(A) back to A. This will prove useful for computing the determinant. Computing the determinant of rref(A) is easy. (Why?) If we know what elementary row operations carry rref(A) back to A, and what effect each of these operations has on the determinant of rref(A), we could find the determinant of A.

Exploration 306 *Let*

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

Find det A.

$$\det A = \ldots \ldots$$

Construct matrix B by switching the first and the third rows of A.

$$B = \begin{bmatrix} -1 & 4 & 2 \\ 3 & 1 & -2 \\ 3 & -1 & -1 \end{bmatrix}$$

Find det *B.*

$$\det B = \ldots\ldots$$

Next, try switching consecutive rows. Construct matrix B' by interchanging the second and third rows of A.

$$B' = \begin{bmatrix} 3 & -1 & -1 \\ -1 & 4 & 2 \\ 3 & 1 & -2 \end{bmatrix}$$

Find det B'.

$$\det B' = \ldots\ldots$$

It appears that switching any two rows of a matrix produces a determinant that is negative of the determinant of the original matrix.

Next, construct matrix C by multiplying the last row of A by k.

$$C = \begin{bmatrix} 3 & -1 & -1 \\ 3 & 1 & -2 \\ -k & 4k & 2k \end{bmatrix}$$

Find det *C.*

$$\det C = \ldots\ldots k$$

It turns out that multiplying the first or the second row of A by k yields exactly the same result as this.

Finally, construct matrix D by adding twice row 3 to row 1.

$$D = \begin{bmatrix} 1 & 7 & 3 \\ 3 & 1 & -2 \\ -1 & 4 & 2 \end{bmatrix}$$

Find det *D.*

$$\det D = \ldots\ldots$$

This result is particularly surprising. Try a few more variations of this example to convince yourself that adding a multiple of one row to another row does not appear to affect the determinant.

The following theorem generalizes our observations.

Theorem 56. Let $A = \left[a_{ij}\right]$ be an $n \times n$ matrix.

1. If B is obtained from A by interchanging two different rows, then
$$\det B = -\det A$$
2. If B is obtained from A by multiplying one of the rows of A by a non-zero constant k. Then
$$\det B = k \det A$$
3. If B is obtained from A by adding a multiple of one row of A to another row, then
$$\det B = \det A$$

The proof of this theorem is relegated to Tedious Proofs Concerning Determinants. For a sketch of the proof, you can watch this video: YouTube link: `https://www.youtube.com/watch?v=y0XVK-ey30c`

The following lemma is a useful consequence of parts **??** and **??** of Theorem **??**.

Lemma 5. Let A be an $n \times n$ matrix.

1. If A has a row of zeros, then $\det A = 0$.
2. If two rows of A are the same, then $\det A = 0$.
3. If one row of A is a scalar multiple of another row, then $\det A = 0$.

We will prove Part **??**. Parts **??** and **??** are left as exercises.

Proof of Part ?? Suppose rows p and q of A are the same. Let B be a matrix obtained from A by switching p and q. By Theorem **????** we know that $\det B = -\det A$. But p and q are the same, so $A = B$. But then $\det A = -\det A$. We conclude that $\det A = 0$. ■

Because $\det A = \det A^T$, we have the following counterpart of Theorem **??** for columns.

Theorem 57 (Elementary Column Operations and the Determinant)**.** Let A be an $n \times n$ matrix.

1. If B is obtained from A by interchanging two different columns, then
$$\det B = -\det A$$

2. If B is obtained from A by multiplying one of the columns of A by a non-zero constant k. Then
$$\det B = k \det A$$

3. If B is obtained from A by adding a multiple of one column of A to another column, then
$$\det B = \det A$$

Computing the Determinant Using Elementary Row Operations

What we discovered about the effects of elementary row operations on the determinant will allow us to compute determinants without using the cumbersome process of cofactor expansion.

Example 116. Suppose that a 6×6 matrix A is carried to the identity matrix by a sequence of elementary row operations listed below. Find $\det A$.

$$A \xrightarrow{R_2 - 2R_4} A_1 \xrightarrow{R_1 \leftrightarrow R_3} A_2 \xrightarrow{\frac{1}{2}R_6} A_3 \xrightarrow{R_5 + R_6} I$$

Explanation. Let's take a look at what happens to the determinant of A one step at a time.

$$\det A_1 = \det A \qquad \text{by Theorem } \textbf{????}$$
$$\det A_2 = -\det A_1 = -\det A \qquad \text{by Theorem } \textbf{????}$$
$$\det A_3 = \frac{1}{2}\det A_2 = -\frac{1}{2}\det A \qquad \text{by Theorem } \textbf{????}$$
$$\det I = \det A_3 = -\frac{1}{2}\det A \qquad \text{by Theorem } \textbf{????}$$

Recall that $\det I = 1$ (Lemma **??**). This gives us

$$-\frac{1}{2}\det A = 1$$

Therefore $\det A = -2$.

Example 117. Let

$$A = \begin{bmatrix} 3 & 0 & -9 \\ 10 & 5 & 2 \\ 8 & 4 & 2 \end{bmatrix}$$

Find $\det A$ by applying elementary row operations to reduce A to its row-echelon form.

Explanation.

$$A = \begin{bmatrix} 3 & 0 & -9 \\ 10 & 5 & 2 \\ 8 & 4 & 2 \end{bmatrix}$$

$$\xrightarrow{(1/3)R_1} \begin{bmatrix} 1 & 0 & -3 \\ 10 & 5 & 2 \\ 8 & 4 & 2 \end{bmatrix} \tag{63}$$

$$\xrightarrow{(1/2)R_3} \begin{bmatrix} 1 & 0 & -3 \\ 10 & 5 & 2 \\ 4 & 2 & 1 \end{bmatrix} \tag{64}$$

$$\begin{matrix} \xrightarrow{R_2 - 10R_1} \\ \xrightarrow{R_3 - 4R_1} \end{matrix} \begin{bmatrix} 1 & 0 & -3 \\ 0 & 5 & 32 \\ 0 & 2 & 13 \end{bmatrix} \tag{65}$$

$$\xrightarrow{(1/5)R_2} \begin{bmatrix} 1 & 0 & -3 \\ 0 & 1 & 32/5 \\ 0 & 2 & 13 \end{bmatrix} \tag{66}$$

$$\xrightarrow{R_3 - 2R_2} \begin{bmatrix} 1 & 0 & -3 \\ 0 & 1 & 32/5 \\ 0 & 0 & 1/5 \end{bmatrix} \tag{67}$$

We stop when we get to a row-echelon form of A because we can see that its determinant is $\frac{1}{5}$ (Theorem **??**).

The following table summarizes the effect of each elementary row operation on the determinant.

Matrix	Determinant
A	$\det A$
(1)	$\frac{1}{3}\det A$
(2)	$\frac{1}{2}\cdot\frac{1}{3}\det A$
(3)	$\frac{1}{2}\cdot\frac{1}{3}\det A$
(4)	$\frac{1}{5}\cdot\frac{1}{2}\cdot\frac{1}{3}\det A$
(5)	$\frac{1}{5}\cdot\frac{1}{2}\cdot\frac{1}{3}\det A$

Since the determinant of the row-echelon form of A in (**??**) is $\frac{1}{5}$, we have

$$\frac{1}{5} = \frac{1}{5}\cdot\frac{1}{2}\cdot\frac{1}{3}\det A$$

Therefore

$$\det A = 6$$

You should verify this result by direct computation using cofactors.

Practice Problems

Problem 307 *Let A be an $n \times n$ matrix. Show that*

$$\det kA = k^n \det A$$

Problem 308 *Prove that if one row of A is a scalar multiple of another row, then* $\det A = 0$. *(Lemma* **????***)*

Problem 309 *Prove that if one row of a matrix is a linear combination of two other rows of the matrix, then the determinant of the matrix is 0.*

Problem 310 *Find* $\det A$ *using elementary row operations.*

$$A = \begin{bmatrix} 3 & 1 & -3 \\ 12 & 5 & -5 \\ 4 & 2 & 1 \end{bmatrix}$$

Answer:

$$\det A = \ldots\ldots$$

Problem 311 *Find* det *A using elementary row operations.*

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

Answer:

$$\det A = \ldots \ldots$$

Problem 312 *The matrix below is an elementary matrix.*

1. *What elementary row operation does this matrix perform?*
2. *Compute the determinant of the matrix in two different ways:*
   (i) *By cofactor expansion.*
   (ii) *By thinking about how the given matrix was obtained from the identity matrix.*

$$E_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Answer:

$$\det E_1 = \ldots \ldots$$

Problem 313 *The matrix below is an elementary matrix.*

1. *What elementary row operation does this matrix perform?*
2. *Compute the determinant of the matrix in two different ways:*
   (i) *By cofactor expansion.*
   (ii) *By thinking about how the given matrix was obtained from the identity matrix.*

$$E_2 = \begin{bmatrix} 1 & 0 & 4 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Answer:

$$\det E_2 = \ldots \ldots$$

Problem 314 *The matrix below is an elementary matrix.*

1. *What elementary row operation does this matrix perform?*
2. *Compute the determinant of the matrix in two different ways:*
   (i) *By cofactor expansion.*
   (ii) *By thinking about how the given matrix was obtained from the identity matrix.*

$$E_3 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

Answer:

$$\det E_3 = \ldots \ldots$$

50 Properties of the Determinant

We begin by summarizing the properties of determinants we introduced in previous sections.

Summary 2.

- The determinant of the identity matrix is equal to 1. (Lemma **??**)
- The determinant of a triangular matrix is equal to the product of the main diagonal entries. (Lemma **??**)
- The determinant of the transpose is equal to the determinant of the matrix. (Theorem **??**)
- If a matrix contains a row of zeros, then its determinant is equal to 0. (Lemma **??**)
- If two rows of a matrix are the same, then the determinant of the matrix is equal to 0. (Lemma **??**)
- If one row of a matrix is a scalar multiple of another row, then the determinant of the matrix is equal to 0. (Lemma **??**)
- If B is obtained from A by interchanging two different rows, then

$$\det B = -\det A$$

- If B is obtained from A by multiplying one of the rows of A by a non-zero constant k. Then

$$\det B = k \det A$$

- If B is obtained from A by adding a multiple of one row of A to another row, then

$$\det B = \det A$$

(The last three properties comprise Theorem **??**)

In this section we will prove the following important results:

1. A square matrix is singular if and only if its determinant is equal to 0.
2. The determinant of a product is the product of the determinants.

To get us started, we need the following lemma.

Lemma 6. Let A be a square matrix, and let E be an elementary matrix, then

$$\det EA = \det E \det A$$

Proof Recall that if E is obtained from I using an elementary row operation, then the same elementary row operation carries A to EA. There are three types of elementary row operations and three types of elementary matrices, so we will have to consider three cases.

Case 1. Suppose E is obtained from I by interchanging two rows, then

$$\det E = -1 \quad \text{and} \quad \det EA = -\det A$$

so

$$\det EA = \det E \det A$$

Case 2. Suppose E is obtained from I by multiplying one of the rows of I by a non-zero constant k, then

$$\det E = k \quad \text{and} \quad \det EA = k \det A$$

so

$$\det EA = \det E \det A$$

Case 3. Suppose E is obtained from I by adding a scalar multiple of one row to another row, then

$$\det E = 1 \quad \text{and} \quad \det EA = \det A$$

so

$$\det EA = \det E \det A$$

■

Invertibility and the Determinant

Recall that we first introduced determinants in the context of invertibility of 2×2 matrices. Specifically, we found that $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is invertible if and only if $\det A \neq 0$. (A logically equivalent statement is: A is singular if and only if $\det A = 0$.) We are now in the position to prove this result for all square matrices.

Theorem 58. A square matrix A is singular if and only if $\det A = 0$.

Proof Let A be a square matrix. To determine whether A is singular we need to find rref(A). In Elementary Matrices we found that there exist elementary matrices $E_1, \ldots, E_k$ such that

$$E_k \ldots E_2 E_1 A = \text{rref}(A)$$

so

$$\det(E_k \ldots E_2 E_1 A) = \det\big(\text{rref}(A)\big)$$

By repeated application of Lemma **??**, we find that

$$\det E_k \ldots \det E_2 \det E_1 \det A = \det\big(\text{rref}(A)\big)$$

Suppose that A is singular, then rref(A) $\neq I$. But then rref(A) contains a row of zeros, and $\det\big(\text{rref}(A)\big) = 0$. (Lemma **??**) Since determinants of elementary matrices are non-zero, we conclude that $\det A = 0$.

Conversely, suppose $\det A = 0$, then

$$\det\big(\text{rref}(A)\big) = \det E_k \ldots \det E_2 \det E_1 \det A = 0$$

But then rref(A) $\neq I$, so A is singular. ■

Example 118. Determine whether A is an invertible matrix without using elementary row operations.

$$A = \begin{bmatrix} 3 & 4 & -8 \\ 1 & 8 & -10 \\ 1 & -2 & 1 \end{bmatrix}$$

Explanation. Compute the determinant of A. You will find that $\det A = 0$. By Theorem **??** we conclude that A is not invertible.

Determinant of a Product

Theorem 59. Let A and B be square matrices, then

$$\det AB = \det A \det B$$

Proof Suppose A is invertible, then A can be written as a product of elementary matrices. (Theorem **??**)

$$A = E_1 E_2 \ldots E_k$$

Then, by repeated application of Lemma **??**, we get

$$\begin{aligned}\det AB &= \det(E_1 E_2 \ldots E_k B)\\ &= \det E_1 \det E_2 \ldots \det E_k \det B\\ &= \det(E_1 E_2 \ldots E_k)\det B\\ &= \det A \det B\end{aligned}$$

Now suppose that A is not invertible. Then AB is also not invertible. So, $\det A = 0$ and $\det AB = 0$. Thus $\det AB = 0 = \det A \det B$. ■

The following theorem is a nice consequence of Theorem **??**. We leave the proof to the reader. (Practice Problem **??**)

Theorem 60. Let A be a nonsingular matrix, then

$$\det A^{-1} = \frac{1}{\det A}$$

Practice Problems

Problem 315 *Without doing written computations, determine whether matrix A is singular.*

$$A = \begin{bmatrix} 0 & 0 & -3 \\ 2 & 3 & -1 \\ 0 & 0 & 10 \end{bmatrix}$$

Multiple Choice:

1. *A is singular*
2. *A is nonsingular*

Problem 316 *Without doing written computations, determine whether matrix A is singular.*

$$A = \begin{bmatrix} 2 & -3 & -1 \\ 0 & 4 & 5 \\ 0 & 0 & -3 \end{bmatrix}$$

Multiple Choice:

1. *A is singular*
2. *A is nonsingular*

Problem 317 *Show that all matrices of the form*

$$\begin{bmatrix} x & x+1 & x+2 \\ x+3 & x+4 & x+5 \\ x+6 & x+7 & x+8 \end{bmatrix}$$

are singular.

Problem 318 *Find values of x for which the given matrix is singular.*

$$\begin{bmatrix} 2-x & 1 \\ 5 & 6-x \end{bmatrix}$$

List values of x in an increasing order.

Answer:

$x =$ ... ... $\quad x =$ ... ...

Problem 319 *Suppose A and B are 2×2 matrices such that* $\det A = 2$ *and* $\det B = \frac{1}{3}$. *Then,* $\det AB^{-1} =$ ... ...

Problem 320 *Suppose A and B are 2×2 matrices such that* $\det A = 2$ *and* $\det B = \frac{1}{3}$. *Then,* $\det (AB)^{-1} =$ ... ...

Problem 321 *Suppose A and B are 2×2 matrices such that* $\det A = 2$ *and* $\det B = \frac{1}{3}$. *Then,* $\det 2AB =$ ... ...

Problem 322 *Prove or give a counterexample.*

$$\det (A + B) = \det (A^T + B^T)$$

Problem 323 *Prove that for a nonsingular matrix A*

$$\det A^{-1} = \frac{1}{\det A}$$

(Theorem **??***)*

Problem 324 *Suppose A is an invertible matrix such that*

$$A^{-1} = A^T$$

What are the possible values for $\det A$*? List your answers in increasing order.*

Answer: $\det A =$ ... ... , ...

51 Tedious Proofs Concerning Determinants

In Finding the Determinant we described the determinant as a function that assigns a scalar to every square matrix. The value of the function in the original definition was given by cofactor expansion along the first row of the matrix. We also observed, through examples, that cofactor expansion along any row or column produces the same value. Examples, however, do not constitute a sufficient proof of equivalency of different cofactor expansions. In this section we will prove that cofactor expansions along any row or column produce the same outcome. This result is known as the Laplace Expansion Theorem. We will also prove several results concerning elementary row operations.

Cofactor Expansion Along the Top Row

Let

$$A = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}$$

Definition **??** of cofactor expansion along the top row for a 3×3 matrix requires three *minor* matrices associated with A.

- A_{11} is obtained from A by deleting the first row and the first column of A.

$$\begin{bmatrix} \cancel{a} & \cancel{b} & \cancel{c} \\ \cancel{d} & e & f \\ \cancel{g} & h & i \end{bmatrix}$$

$$A_{11} = \begin{bmatrix} e & f \\ h & i \end{bmatrix}$$

- A_{12} is obtained from A by deleting the first row and the second column of A.

$$\begin{bmatrix} \cancel{a} & \cancel{b} & \cancel{c} \\ d & \cancel{e} & f \\ g & \cancel{h} & i \end{bmatrix}$$

$$A_{12} = \begin{bmatrix} d & f \\ g & i \end{bmatrix}$$

- A_{13} is obtained from A by deleting the first row and the third column of A.

$$\begin{bmatrix} a & b & c \\ \boxed{\begin{matrix} d & e \\ g & h \end{matrix}} & & f \\ & & i \end{bmatrix}$$

$$A_{13} = \begin{bmatrix} d & e \\ g & h \end{bmatrix}$$

The determinant of A is given by

$$\begin{aligned} \det A = |A| &= a \begin{vmatrix} e & f \\ h & i \end{vmatrix} - b \begin{vmatrix} d & f \\ g & i \end{vmatrix} + c \begin{vmatrix} d & e \\ g & h \end{vmatrix} \\ &= a\big(\det A_{11}\big) - b\big(\det A_{12}\big) + c\big(\det A_{13}\big) \end{aligned}$$

Now we are ready for an $n \times n$ matrix. Let

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1j-1} & a_{1j} & a_{1j+1} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2j-1} & a_{2j} & a_{2j+1} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nj-1} & a_{nj} & a_{nj+1} & \dots & a_{nn} \end{bmatrix}$$

Define A_{1j} to be an $(n-1) \times (n-1)$ matrix obtained from A by deleting the first row and the j^{th} column of A. We say that A_{1j} is the $(1, j)$-*minor* of A.

$$\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1j-1} & a_{1j} & a_{1j+1} & \dots & a_{1n} \\ \boxed{\begin{matrix} a_{21} & a_{22} & \dots & a_{2j-1} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nj-1} \end{matrix}} & & & & a_{2j} & \boxed{\begin{matrix} a_{2j+1} & \dots & a_{2n} \\ \vdots & & \vdots \\ a_{nj+1} & \dots & a_{nn} \end{matrix}} \\ & & & & a_{nj} & & & \end{bmatrix}$$

For a 3×3 matrix A we have

$$\det A = a_{11}\big(\det A_{11}\big) - a_{12}\big(\det A_{12}\big) + a_{13}\big(\det A_{13}\big)$$

We want to follow the same pattern to define the determinant of a larger matrix. A distinct feature of this expression is the alternating sign pattern. We want to preserve this feature as we increase matrix size.

Definition 66. Let $A = \left[a_{ij}\right]$ be an $n \times n$ matrix. Define the *determinant* of A by

$$\begin{aligned}\det A = (-1)^{1+1} a_{11} \det A_{11} + (-1)^{1+2} a_{12} \det A_{12} + ... \\ ... + (-1)^{1+j} a_{1j} \det A_{1j} + ... \\ ... + (-1)^{1+n} a_{1n} \det A_{1n} \\ = \sum_{j=1}^{n} (-1)^{1+j} a_{1j} \det A_{1j}\end{aligned}$$

Naturally, we would like to condense this formula. To accomplish this, let

$$C_{1j} = (-1)^{1+j} \det A_{1j}$$

We will refer to C_{1j} as the $(1, j)$-*cofactor of A*. When we use the cofactor notation, the expression in Definition **??** turns into the following:

$$\det A = a_{11} C_{11} + a_{12} C_{12} + ... + a_{1n} C_{1n} = \sum_{j=1}^{n} a_{1j} C_{1j}$$

This process of computing the determinant is called the *cofactor expansion along the first row*.

Cofactor Expansion Along the First Column

As we have observed in several examples in Finding the Determinant, cofactor expansion along the first column produces the same result as cofactor expansion along the top row. We will now formalize the process of cofactor expansion along the first column for an $n \times n$ matrix and prove that this process produces the same result as our original definition of the determinant. Let A be an $n \times n$ matrix.

$$A = \begin{bmatrix} a_{11} & a_{12} & ... & a_{1n} \\ a_{21} & a_{22} & ... & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{i1} & a_{i2} & ... & a_{in} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & ... & a_{nn} \end{bmatrix}$$

Define A_{i1} to be an $(n-1) \times (n-1)$ matrix obtained from A by deleting the first column and the i^{th} row of A. We say that A_{i1} is the $(i, 1)$-*minor* of A.

$$\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{i1} & a_{i2} & \dots & a_{in} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix}$$

Define $C_{i1} = (-1)^{i+1} \det A_{i1}$ to be the $(i,1)$-*cofactor of A*.

Definition 67. Let $A = \left[a_{ij}\right]$ be an $n \times n$ matrix. Define the *determinant* of A by

$$\begin{aligned} \det A &= (-1)^{1+1} a_{11} \det A_{11} + (-1)^{2+1} a_{21} \det A_{21} + \dots \\ &\qquad \dots + (-1)^{i+1} a_{i1} \det A_{i1} + \dots \\ &\qquad \dots + (-1)^{n+1} a_{n1} \det A_{n1} \\ &= \sum_{i=1}^{n} (-1)^{i+1} a_{i1} \det A_{i1} \end{aligned}$$

or

$$\det A = a_{11} C_{11} + a_{21} C_{21} + \dots + a_{n1} C_{n1} = \sum_{i=1}^{n} a_{i1} C_{i1}$$

Proof of Definition Equivalence

We will now show that cofactor expansion along the first row produces the same result as cofactor expansion along the first column.

Theorem 61. Let $A = \left[a_{ij}\right]$ be an $n \times n$ matrix. Then

$$\sum_{j=1}^{n} (-1)^{1+j} a_{1j} \det A_{1j} = \sum_{i=1}^{n} (-1)^{i+1} a_{i1} \det A_{i1}$$

Proof We will proceed by induction on n. Clearly, the result holds for $n = 1$. Just for practice you should also verify the equality for $n = 2, 3$. (See Practice Problem **??**.) We will assume that the result holds for $(n-1) \times (n-1)$ matrices and show that it must hold for $n \times n$ matrices.

You will find the following matrix a useful reference as we proceed.

$$A = \begin{bmatrix} a_{11} & a_{12} & ... & a_{1j} & ... & a_{1n} \\ a_{21} & a_{22} & ... & a_{2j} & ... & a_{2n} \\ \vdots & \vdots & & \vdots & & \vdots \\ a_{i1} & a_{i2} & ... & a_{ij} & ... & a_{in} \\ \vdots & \vdots & & \vdots & & \vdots \\ a_{n1} & a_{n2} & ... & a_{nj} & ... & a_{nn} \end{bmatrix}$$

For convenience, we will refer to the Right-Hand Side (RHS) and the Left-Hand Side (LHS) of the equality we are trying to prove.

$$\text{LHS} = \sum_{j=1}^{n} (-1)^{1+j} a_{1j} \det A_{1j} \stackrel{?}{=} \sum_{i=1}^{n} (-1)^{i+1} a_{i1} \det A_{i1} = \text{RHS}$$

Note that the first term $a_{11}(-1)^{1+1} \det A_{11}$ is the same for LHS and RHS, so we will only need to consider $i, j \geq 2$.

We will start by analyzing RHS. Consider an arbitrary entry a_{i1} of the fist column. This entry will only appear in the term $a_{i1}(-1)^{i+1} \det A_{i1}$. We will find $\det A_{i1}$ by cofactor expansion along the first row. As we proceed, we have to pay special attention to the subscripts. Because the first column of A was removed, the j^{th} column of A contains the $(j-1)$ column of A_{i1}.

$$\begin{aligned} & a_{i1}(-1)^{i+1} \det A_{i1} = \\ =& a_{i1}(-1)^{i+1} \left(a_{12}(-1)^{1+1} \det (A_{i1})_{11} + ... + a_{1j}(-1)^{1+(j-1)} \det (A_{i1})_{1(j-1)} + ... \right) \end{aligned}$$

Note that the entry a_{1j} will only appear in the term

$$a_{1j}(-1)^{1+(j-1)} \det (A_{i1})_{1(j-1)}$$

So, after we distribute $a_{i1}(-1)^{i+1}$, RHS will contain only one term of the form

$$a_{i1} a_{1j} (-1)^{p+q} \det (A_{st})_{pq}$$

We will perform a similar analysis on LHS. Consider an arbitrary entry a_{1j} of the fist row. This entry will only appear in the term $a_{1j}(-1)^{1+j} \left(\det A_{1j} \right)$. Invoking the induction hypothesis, we will find $\det A_{1j}$ by cofactor expansion along the first column.

$$\begin{aligned} & a_{1j}(-1)^{1+j} \det A_{1j} = \\ =& a_{1j}(-1)^{1+j} \left(a_{21}(-1)^{1+1} \det (A_{1j})_{11} + ... + a_{i1}(-1)^{(i-1)+1} \det (A_{1j})_{(i-1)1} + ... \right) \end{aligned}$$

The entry a_{i1} will only appear in the term

$$a_{i1}(-1)^{(i-1)+1}\det(A_{1j})_{(i-1)1}$$

So, after we distribute $a_{1j}(-1)^{1+j}$, LHS will contain only one term of the form

$$a_{i1}a_{1j}(-1)^{p+q}\det(A_{st})_{pq}$$

But RHS also has only one term of this form. We now need to show that these two terms are equal. The two terms are

$$a_{i1}(-1)^{i+1}a_{1j}(-1)^{1+(j-1)}\det(A_{i1})_{1(j-1)} = a_{i1}a_{1j}(-1)^{i+j+1}\det(A_{i1})_{1(j-1)}$$

and

$$a_{1j}(-1)^{1+j}a_{i1}(-1)^{(i-1)+1}\det(A_{1j})_{(i-1)1} = a_{i1}a_{1j}(-1)^{i+j+1}\det(A_{1j})_{(i-1)1}$$

Observe that $(A_{i1})_{1(j-1)}$ and $(A_{1j})_{(i-1)1}$ are the same matrix because both were obtained from matrix A by deleting the first and the i^{th} rows of A, and the first and the j^{th} columns of A. Therefore $\det(A_{i1})_{1(j-1)} = \det(A_{1j})_{(i-1)1}$.

We conclude that the terms of LHS and RHS match. This establishes the desired equality. ■

Now we know that cofactor expansion along the first row and cofactor expansion along the first column produce the same result, so either expansion can be used to find the determinant.

Proof of Results Concerning Elementary Row Operations

In Elementary Row Operations we observed, without proof, the following properties of the determinant. (See Theorem **??**.)

Summary 3. Let $A = \left[a_{ij}\right]$ be an $n \times n$ matrix.

1. If B is obtained from A by interchanging two different rows, then

$$\det B = -\det A$$

2. If B is obtained from A by multiplying one of the rows of A by a non-zero constant k. Then

$$\det B = k\det A$$

3. If B is obtained from A by adding a multiple of one row of A to another row, then

$$\det B = \det A$$

We now prove these properties.

Proof of Theorem ???? We will start by showing that the result holds if two consecutive rows are interchanged. Suppose B is obtained from A by swapping rows p and $p+1$ of A.

We proceed by induction on n. The result is not applicable for $n = 1$. In Practice Problem **??**, you will be asked to verify that the result holds for 2×2 matrices. Suppose that the result holds for $(n-1) \times (n-1)$ matrices. We need to show that it holds for $n \times n$ matrices. You may find the following diagram useful throughout the proof.

$$\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ \vdots & \vdots & & \vdots \\ a_{p1} & a_{p2} & \dots & a_{pn} \\ a_{(p+1)1} & a_{(p+1)2} & \dots & a_{(p+1)n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix} \qquad \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ \vdots & \vdots & & \vdots \\ a_{(p+1)1} & a_{(p+1)2} & \dots & a_{(p+1)n} \\ a_{p1} & a_{p2} & \dots & a_{pn} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix}$$

Observe that for $i \neq p, p+1$ we have:

$$b_{ij} = a_{ij}$$

Because B_{ij} is obtained from A_{ij} by switching two rows of A_{ij}, our induction hypothesis give us:

$$\det B_{ij} = -\det A_{ij}$$

For $i = p$ and $i = p+1$ we have:

$$b_{p1} = a_{(p+1)1} \quad \text{and} \quad b_{(p+1)1} = a_{p1}$$

$$B_{p1} = A_{(p+1)1} \quad \text{and} \quad B_{(p+1)1} = A_{p1}$$

We compute the determinant of B by cofactor expansion along the first column.

$$\begin{aligned}
\det B &= b_{11}(-1)^{1+1}\det B_{11} + ... + b_{i1}(-1)^{i+1}\det B_{i1} + ... \\
&\quad + b_{p1}(-1)^{p+1}\det B_{p1} + b_{(p+1)1}(-1)^{(p+1)+1}\det B_{(p+1)1} + ... \\
&\quad + b_{n1}(-1)^{n+1}\det B_{n1} \\
&= a_{11}(-1)^{1+1}(-1)\det A_{11} + ... + a_{i1}(-1)^{i+1}(-1)\det A_{i1} + ... \\
&\quad + a_{(p+1)1}(-1)^{p+1}\det A_{(p+1)1} + a_{p1}(-1)^{(p+1)+1}\det A_{p1} + ... \\
&\quad + a_{n1}(-1)^{n+1}(-1)\det A_{n1} \\
&= a_{11}(-1)^{1+1}(-1)\det A_{11} + ... + a_{i1}(-1)^{i+1}(-1)\det A_{i1} + ... \\
&\quad + a_{(p+1)1}(-1)^{(p+1)+1}(-1)\det A_{(p+1)1} + a_{p1}(-1)^{p+1}(-1)\det A_{p1} + ... \\
&\quad + a_{n1}(-1)^{n+1}(-1)\det A_{n1} \\
&= (-1)\Big(a_{11}(-1)^{1+1}\det A_{11} + ... + a_{i1}(-1)^{i+1}\det A_{i1} + ... \\
&\quad + a_{(p+1)1}(-1)^{(p+1)+1}\det A_{(p+1)1} + a_{p1}(-1)^{p+1}\det A_{p1} + ... \\
&\quad + a_{n1}(-1)^{n+1}\det A_{n1}\Big) \\
&= -\det A
\end{aligned}$$

If two non-adjacent rows are switched, then the switch can be carried out by performing an odd number of adjacent row interchanges (See Practice Problem **??**), so the result still holds. ■

Proof of Theorem ???? We proceed by induction on n. Clearly the statement is true for $n = 1$. Just for fun, you might want to verify directly that it holds for 2×2 matrices. Now suppose the statement is true for all $(n-1) \times (n-1)$ matrices. We will show that it holds for $n \times n$ matrices.

Suppose B is obtained from A by multiplying the p's row of A by k.

$$A = \begin{bmatrix} a_{11} & a_{12} & ... & a_{1n} \\ \vdots & \vdots & & \vdots \\ a_{p1} & a_{p2} & ... & a_{pn} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & ... & a_{nn} \end{bmatrix} \quad B = \begin{bmatrix} a_{11} & a_{12} & ... & a_{1n} \\ \vdots & \vdots & & \vdots \\ ka_{p1} & ka_{p2} & ... & ka_{pn} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & ... & a_{nn} \end{bmatrix}$$

We compute the determinant of B by cofactor expansion along the first column.

$$\begin{aligned}
\det B &= a_{11}(-1)^{1+1}\det B_{11} + ... + ka_{p1}(-1)^{p+1}\det B_{p1} + ... \\
&\quad + a_{n1}(-1)^{n+1}\det B_{n1} \\
&= ka_{11}(-1)^{1+1}\det A_{11} + ... + ka_{p1}(-1)^{p+1}\det A_{p1} + ... \\
&\quad + ka_{n1}(-1)^{n+1}\det A_{n1} \\
&= k\Big(a_{11}(-1)^{1+1}\det A_{11} + ... + a_{p1}(-1)^{p+1}\det A_{p1} + ... \\
&\quad + a_{n1}(-1)^{n+1}\det A_{n1}\Big) \\
&= k\det A
\end{aligned}$$

■

Before we tackle the proof of Part **??** of Theorem **??** we will need to prove the following lemma.

Lemma 7. Let A, B and C be $n \times n$ matrices. Suppose A, B and C are identical, except for the p^{th} row. If the p^{th} row of C is the sum of the p^{th} rows of A and B, then

$$\det C = \det A + \det B$$

Proof We will proceed by induction on n. We leave it to the reader to verify cases $n = 1, 2$. We will assume that the statement holds for all $(n-1) \times (n-1)$ matrices and show that it holds for $n \times n$ matrices.

You may find the following representations of A, B and C helpful. Identical entries in A, B and C are labeled d_{ij}.

$$A = \begin{bmatrix} d_{11} & d_{12} & ... & d_{1n} \\ \vdots & \vdots & & \vdots \\ a_{p1} & a_{p2} & ... & a_{pn} \\ \vdots & \vdots & & \vdots \\ d_{n1} & d_{n2} & ... & d_{nn} \end{bmatrix} \quad B = \begin{bmatrix} d_{11} & d_{12} & ... & d_{1n} \\ \vdots & \vdots & & \vdots \\ b_{p1} & b_{p2} & ... & b_{pn} \\ \vdots & \vdots & & \vdots \\ d_{n1} & d_{n2} & ... & d_{nn} \end{bmatrix}$$

$$C = \begin{bmatrix} d_{11} & d_{12} & \dots & d_{1n} \\ \vdots & \vdots & & \vdots \\ a_{p1} + b_{p1} & a_{p2} + b_{p2} & \dots & a_{pn} + b_{pn} \\ \vdots & \vdots & & \vdots \\ d_{n1} & d_{n2} & \dots & d_{nn} \end{bmatrix}$$

Observe that

$$A_{p1} = B_{p1} = C_{p1}$$

For $i \neq p$, the induction hypothesis gives us

$$\det C_{i1} = \det A_{i1} + \det B_{i1}$$

We now compute the determinant of C by cofactor expansion along the first column.

$$\begin{aligned} \det C &= d_{11}(-1)^{1+1} \det C_{11} + \dots + (a_{p1} + b_{p1})(-1)^{p+1} \det C_{p1} + \dots \\ &\quad + d_{n1}(-1)^{n+1} \det C_{n1} \\ \\ &= d_{11}(-1)^{1+1} \Big(\det A_{11} + \det B_{11} \Big) + \dots \\ &\quad + a_{p1}(-1)^{p+1} \det C_{p1} + b_{p1}(-1)^{p+1} \det C_{p1} + \dots \\ &\quad + d_{n1}(-1)^{n+1} \Big(\det A_{n1} + \det B_{n1} \Big) \\ \\ &= d_{11}(-1)^{1+1} \det A_{11} + d_{11}(-1)^{1+1} \det B_{11} + \dots \\ &\quad + a_{p1}(-1)^{p+1} \det A_{p1} + b_{p1}(-1)^{p+1} \det B_{p1} + \dots \\ &\quad + d_{n1}(-1)^{n+1} \det A_{n1} + d_{n1}(-1)^{n+1} \det B_{n1} \\ \\ &= \det A + \det B \end{aligned}$$

■

We are now ready to finish the proof of Theorem **??**

Proof of Theorem ???? Suppose B is obtained from A by adding k times row p to row q.

($p \neq q$ and $k \neq 0$) You may find the following representations of A and B useful.

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ \vdots & \vdots & & \vdots \\ a_{p1} & a_{p2} & \dots & a_{pn} \\ \vdots & \vdots & & \vdots \\ a_{q1} & a_{q2} & \dots & a_{qn} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix} \qquad B = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ \vdots & \vdots & & \vdots \\ a_{p1} & a_{p2} & \dots & a_{pn} \\ \vdots & \vdots & & \vdots \\ a_{q1} + ka_{p1} & a_{q2} + ka_{p2} & \dots & a_{qn} + ka_{pn} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix}$$

We will form another matrix A' by replacing the q^{th} row of A with k times the p^{th} row.

$$A' = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ \vdots & \vdots & & \vdots \\ a_{p1} & a_{p2} & \dots & a_{pn} \\ \vdots & \vdots & & \vdots \\ ka_{p1} & ka_{p2} & \dots & ka_{pn} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix}$$

Observe that matrices A, A' and B are identical except for the q^{th} row, and the q^{th} row of matrix B is the sum of the q^{th} rows of A and A'. Thus, by Lemma **??**, we have:

$$\det B = \det A + \det A'$$

Since one row of A' is a scalar multiple of another row, we know $\det A' = 0$ (see Practice Problem **??**). Therefore $\det B = \det A$. ■

The Laplace Expansion Theorem

As we have seen in examples, the value of the determinant can be computed by expanding along any row or column. This result is known as the Laplace Expansion Theorem. We begin by generalizing some earlier definitions.

Given an $n \times n$ matrix

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1(j-1)} & a_{1j} & a_{1(j+1)} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2(j-1)} & a_{2j} & a_{2(j+1)} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ a_{(i-1)1} & a_{(i-1)2} & \dots & a_{(i-1)(j-1)} & a_{(i-1)j} & a_{(i-1)(j+1)} & \dots & a_{(i-1)n} \\ a_{i1} & a_{i2} & \dots & a_{i(j-1)} & a_{ij} & a_{i(j+1)} & \dots & a_{in} \\ a_{(i+1)1} & a_{(i+1)2} & \dots & a_{(i+1)(j-1)} & a_{(i+1)j} & a_{(i+1)(j+1)} & \dots & a_{(i+1)n} \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \dots & a_{n(j-1)} & a_{nj} & a_{n(j+1)} & \dots & a_{nn} \end{bmatrix}$$

define the *minor* A_{ij} to be an $(n-1) \times (n-1)$ matrix obtained from A by deleting the i^{th} row and the j^{th} column of A.

$$\begin{bmatrix} \boxed{\begin{matrix} a_{11} & a_{12} & \dots & a_{1(j-1)} \\ a_{21} & a_{22} & \dots & a_{2(j-1)} \\ \vdots & \vdots & & \vdots \\ a_{(i-1)1} & a_{(i-1)2} & \dots & a_{(i-1)(j-1)} \end{matrix}} & \begin{matrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{(i-1)j} \end{matrix} & \boxed{\begin{matrix} a_{1(j+1)} & \dots & a_{1n} \\ a_{2(j+1)} & \dots & a_{2n} \\ \vdots & & \vdots \\ a_{(i-1)(j+1)} & \dots & a_{(i-1)n} \end{matrix}} \\ \begin{matrix} a_{i1} & a_{i2} & \dots & a_{i(j-1)} \end{matrix} & a_{ij} & \begin{matrix} a_{i(j+1)} & \dots & a_{in} \end{matrix} \\ \boxed{\begin{matrix} a_{(i+1)1} & a_{(i+1)2} & \dots & a_{(i+1)(j-1)} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \dots & a_{n(j-1)} \end{matrix}} & \begin{matrix} a_{(i+1)j} \\ \vdots \\ a_{nj} \end{matrix} & \boxed{\begin{matrix} a_{(i+1)(j+1)} & \dots & a_{(i+1)n} \\ \vdots & & \vdots \\ a_{n(j+1)} & \dots & a_{nn} \end{matrix}} \end{bmatrix}$$

Define the (i,j)-*cofactor* of A by

$$C_{ij} = (-1)^{i+j} \det(A_{ij})$$

Note that the sign of $(-1)^{i+j}$ follows a checkerboard pattern.

$$\begin{bmatrix} + & - & + & - & + & \dots \\ - & + & - & + & - & \dots \\ + & - & + & - & + & \dots \\ - & + & - & + & - & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

Theorem 62 (Laplace Expansion Theorem)**.** Let $A = \left[a_{ij}\right]$ be an $n \times n$ matrix. Then each of the following computations produces $\det A$.

1. *Cofactor Expansion along the i^{th} row*

$$\begin{aligned}\det A &= a_{i1}C_{i1} + a_{i2}C_{i2} + ... + a_{in}C_{in}\\ &= a_{i1}(-1)^{i+1}\det A_{i1} + a_{i2}(-1)^{i+2}\det A_{i2} + ... + a_{in}(-1)^{i+n}\det A_{in}\\ &= \sum_{j=1}^{n} a_{ij}(-1)^{i+j}\det A_{ij}\end{aligned}$$

2. *Cofactor Expansion along the j^{th} column*

$$\begin{aligned}\det A &= a_{1j}C_{1j} + a_{2j}C_{2j} + ... + a_{nj}C_{nj}\\ &= a_{1j}(-1)^{1+j}\det A_{1j} + a_{2j}(-1)^{2+j}\det A_{2j} + ... + a_{nj}(-1)^{n+j}\det A_{nj}\\ &= \sum_{i=1}^{n} a_{ij}(-1)^{i+j}\det A_{ij}\end{aligned}$$

Proof We will start by showing that cofactor expansion along column j produces the same result as cofactor expansion along the first column. Observe that column j can be shifted into the first column position by $j-1$ consecutive row switches. Let $A' = \left[a'_{ij}\right]$ be the matrix obtained from A by performing the necessary column switches. Then

$$\begin{aligned}\det A &= (-1)^{j-1}\det A'\\ &= (-1)^{j-1}\sum_{i=1}^{n} a'_{i1}(-1)^{i+1}\det A'_{i1}\\ &= (-1)^{j-1}\sum_{i=1}^{n} a_{ij}(-1)^{i+1}\det A_{ij}\\ &= \sum_{i=1}^{n} a_{ij}(-1)^{j-1}(-1)^{i+1}\det A_{ij}\\ &= \sum_{i=1}^{n} a_{ij}(-1)^{i+j}\det A_{ij}\end{aligned}$$

To show that the determinant of A can also be computed by cofactor expansion along any row follows from the fact that $\det A = \det A^T$. (Theorem **??**) ■

Practice Problems

Problem 325 *Complete the proof of Theorem* **????** *by showing that the result holds for a* 2×2 *matrix.*

Problem 326 *Let p and q be two rows of a matrix, with $p < q$. Show that the switch of p and q requires $2(q-p)-1$ adjacent row interchanges.*

52 Determinants and Inverses of Nonsingular Matrices

Combining results of Theorem **??** and Theorem **??** shows that the following statements about matrix A are equivalent:

- A^{-1} exists
- Any equation $A\mathbf{x} = \mathbf{b}$ has a unique solution
- $\det A \neq 0$

In this section we will take a closer look at the relationship between the determinant of a nonsingular matrix A, solution to the system $A\mathbf{x} = \mathbf{b}$, and the inverse of A.

Cramer's Rule

We begin by establishing a formula that allows us to express the unique solution to the system $A\mathbf{x} = \mathbf{b}$ in terms of the determinant of A, for a nonsingular matrix A. This formula is called *Cramer's rule*.

Consider the system

$$\begin{aligned} ax &+ by = e \\ cx &+ dy = f \end{aligned}$$

The system can be written as a matrix equation

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} e \\ f \end{bmatrix}$$

Using one of our standard methods for solving systems we find that

$$x = \frac{ed - bf}{ad - bc} \quad \text{and} \quad y = \frac{af - ec}{ad - bc}$$

Observe that the denominators in the expressions for x and y are the same and equal to $\det \begin{bmatrix} a & b \\ c & d \end{bmatrix}$.

A close examination shows that the numerators of expressions for x and y can also be interpreted as determinants of matrices. The numerator of the expression for x is the determinant of the

matrix that is formed by replacing the first column of $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ with $\begin{bmatrix} e \\ f \end{bmatrix}$. The numerator of the expression for y is the determinant of the matrix that is formed by replacing the second column of $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ with $\begin{bmatrix} e \\ f \end{bmatrix}$. Thus, x and y can be written as

$$x = \frac{ed - bf}{ad - bc} = \frac{\begin{vmatrix} e & b \\ f & d \end{vmatrix}}{\begin{vmatrix} a & b \\ c & d \end{vmatrix}} \quad \text{and} \quad y = \frac{af - ec}{ad - bc} = \frac{\begin{vmatrix} a & e \\ c & f \end{vmatrix}}{\begin{vmatrix} a & b \\ c & d \end{vmatrix}}$$

Note that a unique solution to the system exists if and only if the determinant of the coefficient matrix is not zero.

It turns out that a solution to any square system $A\mathbf{x} = \mathbf{b}$ can be expressed using ratios of determinants, provided that A is nonsingular. The general formula for the i^{th} component of the solution vector is

$$x_i = \frac{\det(\text{matrix } A \text{ with column } i \text{ replaced by } \mathbf{b})}{\det A}$$

To formalize this expression, we need to introduce some notation. Given a matrix

$$A = \begin{bmatrix} | & | & & | \\ \mathbf{a}_1 & \mathbf{a}_2 & \dots & \mathbf{a}_n \\ | & | & & | \end{bmatrix}$$

and a vector $\mathbf{b}$ we use

$$A_i(\mathbf{b})$$

to denote the matrix obtained from A by replacing the i^{th} column of A with $\mathbf{b}$. In other words,

$$A_i(\mathbf{b}) = \begin{bmatrix} | & | & & | & | & | & & | \\ \mathbf{a}_1 & \mathbf{a}_2 & \dots & \mathbf{a}_{i-1} & \mathbf{b} & \mathbf{a}_{i+1} & \dots & \mathbf{a}_n \\ | & | & & | & | & | & & | \end{bmatrix} \tag{68}$$

Using our new notation, we can write the i^{th} component of the solution vector as

$$x_i = \frac{\det A_i(\mathbf{b})}{\det A}$$

We will work through a couple of examples before proving this result as a theorem.

Example 119. Solve $A\mathbf{x} = \mathbf{b}$ using Cramer's rule if

$$A = \begin{bmatrix} 3 & -1 \\ 2 & 5 \end{bmatrix} \quad \text{and} \quad \mathbf{b} = \begin{bmatrix} -2 \\ 4 \end{bmatrix}$$

Explanation. We start by computing the determinant of A.

$$\det A = \begin{vmatrix} 3 & -1 \\ 2 & 5 \end{vmatrix} = (3)(5) - (-1)(2) = 17$$

Next, we compute $\det A_1(\mathbf{b})$ and $\det A_2(\mathbf{b})$.

$$\det A_1(\mathbf{b}) = \begin{vmatrix} -2 & -1 \\ 4 & 5 \end{vmatrix} = (-2)(5) - (-1)(4) = -6$$

$$\det A_2(\mathbf{b}) = \begin{vmatrix} 3 & -2 \\ 2 & 4 \end{vmatrix} = (3)(4) - (-2)(2) = 16$$

We now compute the components of the solution vector.

$$x_1 = \frac{-6}{17} \quad \text{and} \quad x_2 = \frac{16}{17}$$

Finally, it is a good idea to verify that what we found is a solution to the system.

$$\begin{bmatrix} 3 & -1 \\ 2 & 5 \end{bmatrix} \begin{bmatrix} -6/17 \\ 16/17 \end{bmatrix} = \begin{bmatrix} -18/17 - 16/17 \\ -12/17 + 80/17 \end{bmatrix} = \begin{bmatrix} -34/17 \\ 68/17 \end{bmatrix} = \begin{bmatrix} -2 \\ 4 \end{bmatrix}$$

Example 120. Solve $A\mathbf{x} = \mathbf{b}$ using Cramer's rule if

$$A = \begin{bmatrix} 1 & 2 & -1 \\ -1 & 1 & 1 \\ 0 & 3 & 1 \end{bmatrix} \quad \text{and} \quad \mathbf{b} = \begin{bmatrix} -2 \\ 3 \\ 1 \end{bmatrix}$$

Explanation. Find the determinant of A.

$$\det A = \begin{vmatrix} 1 & 2 & -1 \\ -1 & 1 & 1 \\ 0 & 3 & 1 \end{vmatrix} = \ldots \ldots$$

Next, we compute $\det A_i(\mathbf{b})$ for $i = 1, 2, 3$.

$$\det A_1(\mathbf{b}) = \begin{vmatrix} -2 & 2 & -1 \\ 3 & 1 & 1 \\ 1 & 3 & 1 \end{vmatrix} = \ldots\ldots$$

$$\det A_2(\mathbf{b}) = \begin{vmatrix} 1 & -2 & -1 \\ -1 & 3 & 1 \\ 0 & 1 & 1 \end{vmatrix} = \ldots\ldots$$

$$\det A_3(\mathbf{b}) = \begin{vmatrix} 1 & 2 & -2 \\ -1 & 1 & 3 \\ 0 & 3 & 1 \end{vmatrix} = \ldots\ldots$$

This gives us the solution vector

$$\mathbf{x} = \begin{bmatrix} \ldots\ldots \\ \ldots\ldots \\ \ldots\ldots \end{bmatrix}$$

You should verify that what you found really is a solution.

We are now ready to state and prove Cramer's rule as a theorem.

Theorem 63. Let A be a nonsingular $n \times n$ matrix, and let $\mathbf{b}$ be an $n \times 1$ vector. Then the components of the solution vector $\mathbf{x}$ of $A\mathbf{x} = \mathbf{b}$ are given by

$$x_i = \frac{\det A_i(\mathbf{b})}{\det A}$$

where $A_i(\mathbf{b})$ is defined as in (**??**).

Proof For this proof we will need to think of matrices in terms of their columns. Thus,

$$A = \begin{bmatrix} | & | & & | \\ \mathbf{a}_1 & \mathbf{a}_2 & \ldots & \mathbf{a}_n \\ | & | & & | \end{bmatrix}$$

We will also need the identity matrix I. The columns of I are standard unit vectors.

$$I = \begin{bmatrix} | & | & & | \\ \mathbf{e}_1 & \mathbf{e}_2 & \ldots & \mathbf{e}_n \\ | & | & & | \end{bmatrix}$$

Recall that

$$A_i(\mathbf{b}) = \begin{bmatrix} | & | & & | & | & | & & | \\ \mathbf{a}_1 & \mathbf{a}_2 & \dots & \mathbf{a}_{i-1} & \mathbf{b} & \mathbf{a}_{i+1} & \dots & \mathbf{a}_n \\ | & | & & | & | & | & & | \end{bmatrix}$$

Similarly,

$$I_i(\mathbf{x}) = \begin{bmatrix} | & | & & | & | & | & & | \\ \mathbf{e}_1 & \mathbf{e}_2 & \dots & \mathbf{e}_{i-1} & \mathbf{x} & \mathbf{e}_{i+1} & \dots & \mathbf{e}_n \\ | & | & & | & | & | & & | \end{bmatrix}$$

Observe that x_i is the only non-zero entry in the i^{th} row of $I_i(\mathbf{x})$. Cofactor expansion along the i^{th} row gives us

$$\det I_i(\mathbf{x}) = x_i \tag{69}$$

Now, consider the product $A\Big(I_i(\mathbf{x})\Big)$

$$\begin{aligned} A\Big(I_i(\mathbf{x})\Big) &= \begin{bmatrix} | & | & & | \\ \mathbf{a}_1 & \mathbf{a}_2 & \dots & \mathbf{a}_n \\ | & | & & | \end{bmatrix} \begin{bmatrix} | & | & & | & | & | & & | \\ \mathbf{e}_1 & \mathbf{e}_2 & \dots & \mathbf{e}_{i-1} & \mathbf{x} & \mathbf{e}_{i+1} & \dots & \mathbf{e}_n \\ | & | & & | & | & | & & | \end{bmatrix} \\ &= \begin{bmatrix} | & | & & | & | & | & & | \\ \mathbf{a}_1 & \mathbf{a}_2 & \dots & \mathbf{a}_{i-1} & A\mathbf{x} & \mathbf{a}_{i+1} & \dots & \mathbf{a}_n \\ | & | & & | & | & | & & | \end{bmatrix} \\ &= \begin{bmatrix} | & | & & | & | & | & & | \\ \mathbf{a}_1 & \mathbf{a}_2 & \dots & \mathbf{a}_{i-1} & \mathbf{b} & \mathbf{a}_{i+1} & \dots & \mathbf{a}_n \\ | & | & & | & | & | & & | \end{bmatrix} = A_i(\mathbf{b}) \end{aligned}$$

This gives us

$$AI_i(\mathbf{x}) = A_i(\mathbf{b})$$

$$\det AI_i(\mathbf{x}) = \det A_i(\mathbf{b})$$

$$\det A \det I_i(\mathbf{x}) = \det A_i(\mathbf{b})$$

By our earlier observation in (**??**) we have

$$\det Ax_i = \det A_i(\mathbf{b})$$

A is nonsingular, so $\det A \neq 0$. Thus

$$x_i = \frac{\det A_i(\mathbf{b})}{\det A}$$

■

Finding the determinant is computationally expensive. Because Cramer's rule requires finding many determinants, it is not a computationally efficient way of solving a system of equations. However, Cramer's rule is often used for small systems in applications that arise in economics, natural, and social sciences, particularly when solving for only a subset of the variables.

Adjugate Formula for the Inverse of a Matrix

In Practice Problem **??** we used the row reduction algorithm to show that if

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

is nonsingular then

$$A^{-1} = \frac{1}{\det A} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \tag{70}$$

This formula is a special case of a general formula for the inverse of a nonsingular square matrix. Just like the formula for a 2×2 matrix, the general formula includes the coefficient $\dfrac{1}{\det A}$ and a matrix related to the original matrix. We will now derive the general formula using Cramer's rule.

Let A be an $n \times n$ nonsingular matrix. When looking for the inverse of A, we look for a matrix X such that $AX = I$. We will think of matrices in terms of their columns

$$I = \begin{bmatrix} | & | & & | \\ \mathbf{e}_1 & \mathbf{e}_2 & \dots & \mathbf{e}_n \\ | & | & & | \end{bmatrix} \quad \text{and} \quad X = \begin{bmatrix} | & | & & | \\ \mathbf{x}_1 & \mathbf{x}_2 & \dots & \mathbf{x}_n \\ | & | & & | \end{bmatrix}$$

If $AX = I$ then we must have

$$\begin{aligned} A\mathbf{x}_1 &= \mathbf{e}_1 \\ A\mathbf{x}_2 &= \mathbf{e}_2 \\ &\vdots \\ A\mathbf{x}_n &= \mathbf{e}_n \end{aligned}$$

This gives us n systems of equations. Solution vectors to these systems are the columns of X. Thus, the j^{th} column of X is

$$\mathbf{x}_j = \begin{bmatrix} x_{1j} \\ x_{2j} \\ \vdots \\ x_{nj} \end{bmatrix} \quad \text{such that} \quad A\mathbf{x}_j = \mathbf{e}_j$$

By Cramer's rule

$$x_{ij} = \frac{\det A_i(\mathbf{e}_j)}{\det A}$$

But

$$A_i(\mathbf{e}) = \begin{bmatrix} | & | & & | & | & | & & | \\ \mathbf{a}_1 & \mathbf{a}_2 & \dots & \mathbf{a}_{i-1} & \mathbf{e}_j & \mathbf{a}_{i+1} & \dots & \mathbf{a}_n \\ | & | & & | & | & | & & | \end{bmatrix}$$

To find $\det A_i(\mathbf{e}_j)$, we can expand along the i^{th} column of $A_i(\mathbf{e}_j)$. But the i^{th} column of $A_i(\mathbf{e}_j)$ is the vector $\mathbf{e}_j$ which has 1 in the j^{th} spot and zeros everywhere else. Thus

$$\det A_i(\mathbf{e}_j) = (-1)^{i+j} \det A_{ji} = C_{ji}$$

We now have

$$\mathbf{x}_j = \begin{bmatrix} C_{j1}/\det A \\ C_{j2}/\det A \\ \vdots \\ C_{jn}/\det A \end{bmatrix} = \frac{1}{\det A} \begin{bmatrix} C_{j1} \\ C_{j2} \\ \vdots \\ C_{jn} \end{bmatrix}$$

Thus,

$$A^{-1} = X = \frac{1}{\det A} \begin{bmatrix} C_{11} & C_{21} & \dots & C_{n1} \\ C_{12} & C_{22} & \dots & C_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ C_{1n} & C_{2n} & \dots & C_{nn} \end{bmatrix}$$

The matrix of cofactors of A is called the *adjugate* of A. We write

$$\text{adj}(A) = \begin{bmatrix} C_{11} & C_{21} & \dots & C_{n1} \\ C_{12} & C_{22} & \dots & C_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ C_{1n} & C_{2n} & \dots & C_{nn} \end{bmatrix}$$

Warning 3. Note the order of subscripts of C in the adjugate matrix. The (i, j)-entry of the adjugate matrix is C_{ji}.

We summarize our result as a theorem

Theorem 64. Let A be a nonsingular square matrix, then

$$A^{-1} = \frac{1}{\det A}\text{adj}(A)$$

Example 121. Use Theorem **??** to find A^{-1} if

$$A = \begin{bmatrix} 1 & -1 & 2 \\ 1 & 1 & 1 \\ 1 & 3 & -1 \end{bmatrix}$$

Explanation. We begin by finding det A

$$\det A = \ldots\ldots$$

The first column of adj(A) has entries C_{11}, C_{12} and C_{13}.

$$C_{11} = (-1)^{1+1}\begin{vmatrix} 1 & 1 \\ 3 & -1 \end{vmatrix} = \ldots\ldots$$

$$C_{12} = (-1)^{1+2}\begin{vmatrix} 1 & 1 \\ 1 & -1 \end{vmatrix} = \ldots\ldots$$

$$C_{13} = (-1)^{1+3}\begin{vmatrix} 1 & 1 \\ 1 & 3 \end{vmatrix} = \ldots\ldots$$

The second column of adj(A) has entries C_{21}, C_{22} and C_{23}.

$$C_{21} = (-1)^{2+1}\begin{vmatrix} \ldots\ldots & \ldots\ldots \\ \ldots\ldots & \ldots\ldots \end{vmatrix} = \ldots\ldots$$

$$C_{22} = (-1)^{2+2}\begin{vmatrix} \ldots\ldots & \ldots\ldots \\ \ldots\ldots & \ldots\ldots \end{vmatrix} = \ldots\ldots$$

$$C_{23} = (-1)^{2+3}\begin{vmatrix} \ldots\ldots & \ldots\ldots \\ \ldots\ldots & \ldots\ldots \end{vmatrix} = \ldots\ldots$$

Now we compute the third column of adj(A).

$$C_{31} = \ldots\ldots$$

$$C_{32} = \ldots\ldots$$

$$C_{33} = \ldots\ldots$$

This gives us

$$A^{-1} = \begin{bmatrix} 2 & -5/2 & 3/2 \\ -1 & 3/2 & -1/2 \\ -1 & 2 & -1 \end{bmatrix}$$

Compare this result to the answer in Problem **??**.

Practice Problems

Problems **??**-**??**

Use Cramer's rule to solve each of the following systems.

Problem 327

$$\begin{aligned} -3x + 2y &= 1 \\ 2x - y &= 4 \end{aligned}$$

Answer:

$$x = \ldots\ldots \quad y = \ldots\ldots$$

Problem 328

$$\begin{bmatrix} 2 & -5 & 1 \\ 6 & 0 & -2 \\ -3 & 1 & 1 \end{bmatrix} \mathbf{x} = \begin{bmatrix} 1 \\ 4 \\ 3 \end{bmatrix}$$

Answer:

$$\mathbf{x} = \begin{bmatrix} \ldots\ldots \\ \ldots\ldots \\ \ldots\ldots \end{bmatrix}$$

Problem 329 *Consider the equation*

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 3 & 2 & -1 \\ 3 & -2 & 3 & -2 \\ 2 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 10 \\ 10 \\ 0 \\ 11 \end{bmatrix}$$

1. *Solve for x_2 using Cramer's Rule.*

   Answer:

   $$x_2 = \ldots\ldots$$

2. *If you had to solve for all four variables, which method would you use? Why?*

Problems **??**-**??**

Use Theorem **??** to find the inverse of each of the following matrices.

Problem 330

$$A = \begin{bmatrix} 2 & 7 \\ 1 & 3 \end{bmatrix}$$

Answer:

$$A^{-1} = \begin{bmatrix} \ldots\ldots & \ldots\ldots \\ \ldots\ldots & \ldots\ldots \end{bmatrix}$$

Problem 331

$$A = \begin{bmatrix} 2 & 1 & 4 \\ 4 & -2 & 1 \\ 0 & 3 & -1 \end{bmatrix}$$

Answer:

$$A^{-1} = \frac{1}{\ldots\ldots} \begin{bmatrix} \ldots\ldots & \ldots\ldots & \ldots\ldots \\ \ldots\ldots & \ldots\ldots & \ldots\ldots \\ \ldots\ldots & \ldots\ldots & \ldots\ldots \end{bmatrix}$$

Problem 332 *Show that the formula in (***??***) is a special case of the formula in Theorem* **??** *by showing that*

$$adj\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

53 Eigenvalues

54 Describing Eigenvalues and Eigenvectors

At several places in this course it has been valuable to restrict ourselves to square matrices, and we do so again when discussing eigenvalues and eigenvectors.

In Theorem **??**, we proved that any $n \times n$ matrix induces a linear transformation from $\mathbb{R}^n$ to itself. For our first few examples, let us consider the case $n = 2$.

Exploration 333 *Let* $A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$. *The following interactive helps us to visualize the matrix transformation associated with A. Given a vector* $\mathbf{x}$ *in* $\mathbb{R}^2$ *(left panel), its image,* $A\mathbf{x}$*, is also in* $\mathbb{R}^2$ *(right panel). Drag the head of the vector* $\mathbf{x}$ *to see its image.*

What do you observe about the relationship between vectors along the red line $y = -x$ *and their images?*

For many vectors $\mathbf{x}$ *of* $\mathbb{R}^2$*,* $A\mathbf{x}$ *does not point in the same direction as* $\mathbf{x}$ *but if we look at the vectors along the line* $y = -x$*, we notice that the transformation does not seem to affect their magnitude and direction. Such vectors are sometimes called fixed vectors of A. For example* $\begin{bmatrix} 1 \\ -1 \end{bmatrix}$ *is a fixed vector.*

What do you observe about the relationship between vectors along the blue line $y = x$ *and their images?*

We observe that the magnitudes of these vectors are changed by the transformation, but the direction in which the vectors point is unchanged.

In Exploration **??** we found that the images of vectors $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} -1 \\ 1 \end{bmatrix}$ stay parallel to the original vectors under the linear transformation induced by matrix A. Such vectors are examples of *eigenvectors* of A.

In general, any nonzero vector whose image under a matrix transformation is parallel to the original vector is called an *eigenvector* of the matrix that induced the transformation. The following definition captures this idea algebraically.

Definition 68. Let A be an $n \times n$ matrix. We say that a non-zero vector $\mathbf{x}$ is an *eigenvector* of A if

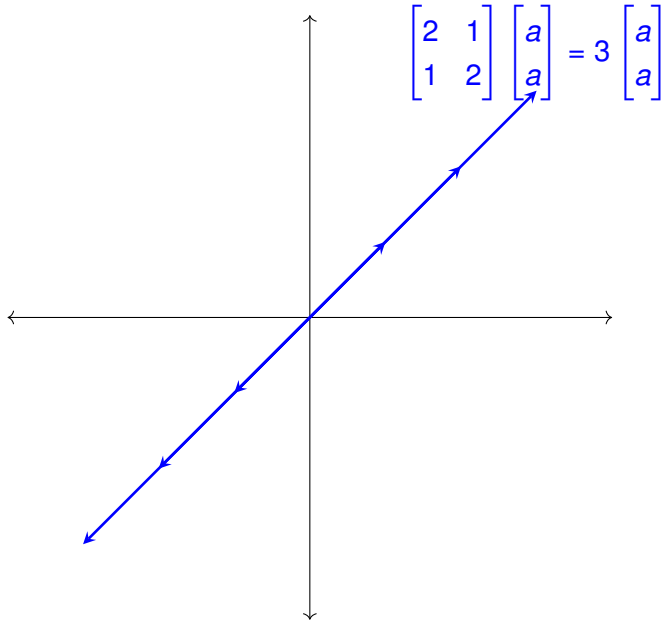
$$A\mathbf{x} = \lambda\mathbf{x}$$

for some scalar λ. We say that λ is an *eigenvalue* of A associated with the eigenvector $\mathbf{x}$.

Let's revisit Exploration **??** in light of Definition **??**. In the exploration, we observed visually that vectors parallel to $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ were eigenvectors associated with $A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$, as these vectors changed length but remained parallel to the original vector under the linear transformation induced by A. To verify this algebraically, observe that all vectors parallel to $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ can be written in the form $\begin{bmatrix} a \\ a \end{bmatrix}$, ($a \neq 0$). We compute

$$\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} a \\ a \end{bmatrix} = \begin{bmatrix} 3a \\ 3a \end{bmatrix} = 3 \begin{bmatrix} a \\ a \end{bmatrix}$$

This shows that any non-zero scalar multiple of $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ is an eigenvector of A which has a corresponding eigenvalue of 3.



Fixed vectors of Exploration **??** are also eigenvectors. For example,

$$\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \end{bmatrix} = 1 \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

This shows that $\begin{bmatrix} 1 \\ -1 \end{bmatrix}$ is a fixed vector and an eigenvector of A which has a corresponding eigenvalue of 1.

The above discussion leads us to the following result.

Theorem 65. If $\mathbf{x}$ is an eigenvector of matrix A and λ is the corresponding eigenvalue, then every non-zero scalar multiple of $\mathbf{x}$ is also an eigenvector of A and λ is the corresponding eigenvalue.

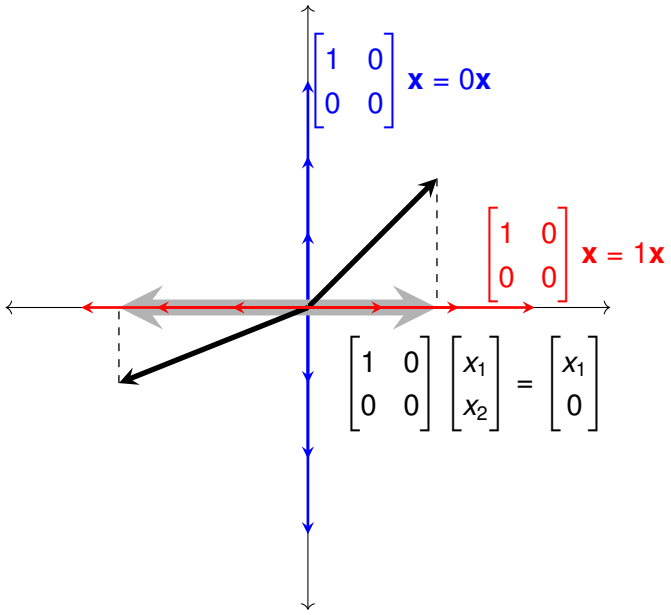
Proof See Practice Problem **??** ■

Observation 8. A couple of finer points of Definition **??** require clarification.

- The definition requires that eigenvectors be non-zero. Imagine what would happen if we allowed $\mathbf{x} = \mathbf{0}$ to be an eigenvector of A. Clearly $A\mathbf{0} = \lambda\mathbf{0}$ for all scalars λ. This means that every number would be an eigenvalue of every matrix. Because eigenvalues are supposed to capture certain information about the matrix, allowing every number to be an eigenvalue of every matrix would defeat the purpose.
- Until now, we had talked about eigenvectors as vectors whose images under a matrix transformation are parallel to the original vectors. But the algebraic definition allows non-zero vectors that map to zero to be considered eigenvectors. (What would an eigenvalue of such an eigenvector be?) The zero vector has no direction, so we cannot say that the image of such an eigenvector is parallel to the original vector. Example **??** will illustrate this point.

Example 122. Let $P = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$. Note that P takes a vector in $\mathbb{R}^2$ and projects it onto the x-axis, as we learned in Practice Problem **??**. Which vectors in $\mathbb{R}^2$ would be the eigenvectors, and what are the corresponding eigenvalues?

Explanation. Since $P\mathbf{x}$ is the projection of $\mathbf{x}$ onto the x-axis, in many cases $P\mathbf{x}$ and $\mathbf{x}$ are not parallel. Notice, however, that all of the red vectors located along the x-axis in the diagram are fixed by P. So, for any of the red vectors we have $P\mathbf{x} = \mathbf{x} = 1\mathbf{x}$, which means that each of the red vectors is an eigenvector of P with the corresponding eigenvalue of 1.



The blue vectors along the y-axis are also eigenvectors. To see this, note that each of the blue vectors is of the form $\mathbf{x} = \begin{bmatrix}0\\x_2\end{bmatrix}$. But then

$$P\mathbf{x} = \begin{bmatrix}1 & 0\\0 & 0\end{bmatrix}\begin{bmatrix}0\\x_2\end{bmatrix} = \begin{bmatrix}0\\0\end{bmatrix} = 0\begin{bmatrix}0\\x_2\end{bmatrix} = 0\mathbf{x}$$

So each of the blue vectors is an eigenvector of P with the corresponding eigenvalue of 0.

Exploration 334 *Let* $A = \begin{bmatrix}1 & 6\\1 & 0\end{bmatrix}$. *The GeoGebra interactive below shows vector* $\mathbf{x}$*, and its image,* $A\mathbf{x}$ *in the same coordinate plane. (Move the head of* $\mathbf{x}$ *to see the effect of* A *on different vectors.)*

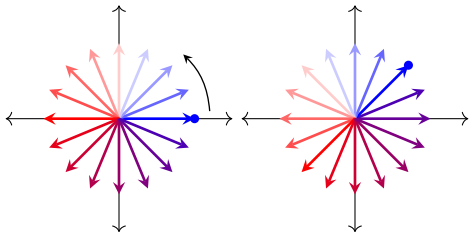
Question 334.1 *Note that vectors* $\mathbf{x}_1 = \begin{bmatrix}3\\1\end{bmatrix}$ *and* $\mathbf{x}_2 = \begin{bmatrix}-2\\1\end{bmatrix}$ *(and their scalar multiples) remain positioned along the same lines even as they change magnitude and direction. This indicates that* $\mathbf{x}_1$ *and* $\mathbf{x}_2$*, along with all of their scalar multiples, are eigenvectors of* A*. What are the eigenvalues associated with these eigenvectors?*

Eigenvalue associated with $\mathbf{x}_1$ *is* $\lambda_1 = \ldots\ldots$

Eigenvalue associated with $\mathbf{x}_2$ *is* $\lambda_2 = \ldots\ldots$

A natural question is this: does every square matrix have eigenvalues and eigenvectors? We will see later that the answer to this question is "yes", provided that we permit eigenvalues and entries of eigenvectors to be complex numbers. The next example illustrates this.

Example 123. Let $M = \begin{bmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{bmatrix}$. Note that M takes any vector $\mathbf{x}$ in $\mathbb{R}^2$ and rotates it 45°, as we saw in Example **??**.



Since M rotates every vector in $\mathbb{R}^2$, *every* nonzero vector changes direction, so there are no eigenvectors in the plane. It turns out that M does have eigenvectors and eigenvalues, but in order to find them we need to work with vectors whose entries are complex numbers. Since these vectors are not in $\mathbb{R}^2$, we cannot see them.

To follow the computation below, you need to recall that the imaginary unit i is defined by $i^2 = -1$.

Consider the vector $\begin{bmatrix} \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2}i \end{bmatrix}$. We compute:

$$\begin{bmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{bmatrix} \begin{bmatrix} \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2}i \end{bmatrix} = \begin{bmatrix} \frac{1}{2} - \frac{1}{2}i \\ \frac{1}{2} + \frac{1}{2}i \end{bmatrix} = (\frac{\sqrt{2}}{2} - \frac{\sqrt{2}}{2}i) \begin{bmatrix} \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2}i \end{bmatrix},$$

so $\begin{bmatrix} \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2}i \end{bmatrix}$ is an eigenvector of M. Its corresponding eigenvalue is $\frac{\sqrt{2}}{2} - \frac{\sqrt{2}}{2}i$.

We will continue to work with complex numbers as we study eigenvalues and eigenvectors.

Why All the Fuss About Eigenvalues and Eigenvectors?

Remark 9. The first in-depth study of eigenvalues can probably be attributed to Fourier as he studied partial differential equations early in the nineteenth century, and in particular when he studied what is known as the heat equation. [Trefethen and Embree] By the twentieth century mathematicians understood the connections between differential equations and eigenvalues. Systems of differential equations are often best represented by matrices, especially in the context of using computers to find numerical solutions. Most algorithms to solve these systems work by iterating some process, and eigenvalues along with their corresponding eigenvectors indicate what will happen to such a process after many repetitions.

The most famous modern example of a large-scale eigenvalue problem is the Google PageRank algorithm, which helped set Google apart from its competitors as a search engine. Some of the relevant mathematics can be learned by working through the paper, "The $25,000,000,000 Eigenvector, The Linear Algebra Behind Google", by Kurt Bryan and Tanya Leise.

Practice Problems

Problem 335 *Let* $C = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix}$.

Question 335.1 *Show that* $\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ *is an eigenvector of* C*. What is its corresponding eigenvalue?* ... ...

Question 335.2 *Show that* $\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}$ *is an eigenvector of* C*. What is its corresponding eigenvalue?* ... ...

Question 335.3 *Show that* $\begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix}$ *is an eigenvector of* C*. What is its corresponding eigenvalue?* ... ...

Problem 336 *Let* $Q = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$. *Note that Q takes any vector in* $\mathbb{R}^2$ *and projects it onto the y-axis. Which vectors in* $\mathbb{R}^2$ *would be eigenvectors, and what are the corresponding eigenvalues?*

Problem 337 *Returning to Example* **??**, *let* $M = \begin{bmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{bmatrix}$. *Show that* $\begin{bmatrix} \frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2}i \end{bmatrix}$ *is an eigenvector of M. What is its corresponding eigenvalue?*

Please enter an exact answer; no decimal approximations. To enter square roots, click inside the answer box, then use the "Math Editor" button that appears at the top of your XIMERA window.

..... + ... *i*

Problem 338 *Arguing geometrically, identify the linear transformation whose standard matrix has eigenvalues* $\lambda_1 = 1$ *and* $\lambda_2 = -1$.

Multiple Choice:

1. *Vertical Shear*
2. *Horizontal Shear*
3. *Counterclockwise Rotation through a* 90° *angle*
4. *Reflection About the line y = mx*
5. *Horizontal Stretch*
6. *Vertical Stretch*

Problem 339 *Let* $A = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 3 \end{bmatrix}$. *Can you find an eigenvector and its corresponding eigenvalue? Can you find another "eigenpair"? Can you find all of the eigenvectors of A?*

Problem 340 *The rotation matrix in Example* **??** *has complex eigenvectors and eigenvalues. Think geometrically to find an example of a (non-identity) rotation matrix with real eigenvectors and eigenvalues.*

Enter degree measure between 0 and 360.

Answer: Rotation through ..... degrees.

Problem 341 *Can an eigenvalue have multiple eigenvectors associated with it?*

Multiple Choice:

1. *Yes*
2. *No*

Can an eigenvector have multiple eigenvalues associated with it?

Multiple Choice:

1. *Yes*
2. *No*

Problem 342 *Prove that if* **x** *is an eigenvector of matrix A and* λ *is the corresponding eigenvalue, then every non-zero scalar multiple of* **x** *is also an eigenvector of A and* λ *is the corresponding eigenvalue. (Theorem* **??***)*

Bibliography

[Trefethen and Embree] Trefethen, Lloyd and Embree, Mark, *Spectra and Pseudospectra*, Princeton University Press, 2005, p. 5-6

55 The Characteristic Equation

Let A be an $n \times n$ matrix. In the previous section we learned that the eigenvectors and eigenvalues of A are vectors $\mathbf{x}$ and scalars λ that satisfy the equation

$$A\mathbf{x} = \lambda\mathbf{x} \tag{71}$$

We listed a few reasons why we are interested in finding eigenvalues and eigenvectors, but we did not give any process for finding them. In this section we will focus on a process which can be used for small matrices. For larger matrices, the best methods we have are iterative methods, and we will explore some of these in The Power Method and the Dominant Eigenvalue.

For an $n \times n$ matrix, we will see that the eigenvalues are the roots of a polynomial called the *characteristic polynomial*. So finding eigenvalues is equivalent to solving a polynomial equation of degree n. Finding the corresponding eigenvectors turns out to be a matter of computing the null space of a matrix, as the following exploration demonstrates.

Exploration 343 *If a vector* $\mathbf{x}$ *is an eigenvector satisfying Equation (**??**), then clearly it also satisfies* $A\mathbf{x} - \lambda\mathbf{x} = (0/\mathbf{0})$.

It seems natural at this point to try to factor. We would love to "factor out" $\mathbf{x}$*. Here is the procedure:*

$$A\mathbf{x} - \lambda\mathbf{x} = \mathbf{0}$$
$$A\mathbf{x} - \lambda I\mathbf{x} = \mathbf{0}$$
$$(A - \lambda I)\mathbf{x} = \mathbf{0}$$

This shows that any eigenvector $\mathbf{x}$ *of A is in the (row space/ column space/ null space) of the related matrix,* $A - \lambda I$.

Since eigenvectors are non-zero vectors, this means that A will have eigenvectors if and only if the null space of $A - \lambda I$ *is nontrivial. The only way that null*$(A - \lambda I)$ *can be nontrivial is if rank*$(A - \lambda I)$ $(=/</>)$ *n.*

If the rank of an $n \times n$ *matrix is less than n, then the matrix is singular. Since* $(A - \lambda I)$ *must be singular for any eigenvalue* λ*, we see that* λ *is an eigenvalue of A if and only if*

$$det(A - \lambda I) = \ldots\ldots \tag{72}$$

In theory, Exploration **??** offers us a way to find eigenvalues. To find the eigenvalues of A, one can solve Equation (**??**) for λ.

Eigenvalues

Definition 69. The equation

$$\det(A - \lambda I) = 0$$

is called the *characteristic equation* of A.

Example 124. Let $A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$. Compute the eigenvalues of this matrix using the characteristic equation.

Explanation.

$$\begin{aligned} \det(A - \lambda I) = \begin{vmatrix} 2-\lambda & 1 \\ 1 & 2-\lambda \end{vmatrix} &= (2-\lambda)^2 - 1 \\ &= \lambda^2 - 4\lambda + 3 \\ &= (\lambda - 1)(\lambda - 3) \end{aligned}$$

The characteristic equation $(\lambda - 1)(\lambda - 3) = 0$ has solutions $\lambda_1 = 1$ and $\lambda_2 = 3$. These are the eigenvalues of A.

Example 125. Let $B = \begin{bmatrix} 2 & 1 \\ 4 & 2 \end{bmatrix}$. Compute the eigenvalues of B using the characteristic equation. (List your answers in an increasing order.)

$\lambda_1 =$ ... ... and $\lambda_2 =$ ... ...

Example 126. Let $C = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix}$. Compute the eigenvalues of C using the characteristic equation.

Explanation.

$$\begin{aligned} \det(C - \lambda I) &= \begin{vmatrix} 2-\lambda & 1 & 1 \\ 1 & 2-\lambda & 1 \\ 1 & 1 & 2-\lambda \end{vmatrix} \\ &= (2-\lambda)\begin{vmatrix} 2-\lambda & 1 \\ 1 & 2-\lambda \end{vmatrix} - 1\begin{vmatrix} 1 & 1 \\ 1 & 2-\lambda \end{vmatrix} + 1\begin{vmatrix} 1 & 2-\lambda \\ 1 & 1 \end{vmatrix} \\ &= (2-\lambda)^3 - (2-\lambda) - ((2-\lambda) - 1) + 1 - (2-\lambda) \end{aligned}$$

$$
\begin{aligned}
&= 8 - 12\lambda + 6\lambda^2 - \lambda^3 - 2 + \lambda - 2 + \lambda + 1 + 1 - 2 + \lambda \\
&= 4 - 9\lambda + 6\lambda^2 - \lambda^3 \\
&= -(\lambda - 4)(\lambda - 1)^2
\end{aligned}
$$

Matrix C has eigenvalues $\lambda_1 = 1$ and $\lambda_2 = 4$.

Remark 10. The factor $(\lambda - 1)$ appears twice. This repeated factor gives rise to the eigenvalue $\lambda_1 = 1$. We say that the eigenvalue $\lambda_1 = 1$ has *algebraic multiplicity* 2.

Definition 70. *Algebraic multiplicity* of an eigenvalue is its multiplicity as a root of the characteristic equation.

It is not always possible to completely factor the characteristic polynomial using only real numbers. However, a fundamental fact from algebra is that every degree n polynomial has n roots (counting multiplicity) provided that we allow complex numbers. This is why sometimes eigenvalues and their corresponding eigenvectors involve complex numbers. The next example illustrates this point.

Example 127. Let $D = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & -1 & 1 \end{bmatrix}$. Compute the eigenvalues of this matrix.

Explanation.

$$
\begin{aligned}
\det(D - \lambda I) &= \begin{vmatrix} -\lambda & 0 & 0 \\ 0 & 1 - \lambda & 1 \\ 0 & -1 & 1 - \lambda \end{vmatrix} \\
&= -\lambda \begin{vmatrix} 1 - \lambda & 1 \\ -1 & 1 - \lambda \end{vmatrix} \\
&= -\lambda((1 - \lambda)^2 + 1) \\
&= -\lambda(\lambda^2 - 2\lambda + 2)
\end{aligned}
$$

So one of the eigenvalues of D is $\lambda = 0$. To get the other eigenvalues we must solve $\lambda^2 - 2\lambda + 2 = 0$. Using the quadratic formula, we compute that $\lambda_1 = 1 + i$ and $\lambda_2 = 1 - i$ are also eigenvalues of D.

Exploration 344 *Let* $T = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 5 & 6 \\ 0 & 0 & 9 \end{bmatrix}$. *Compute the eigenvalues of this matrix. (List your answers in an increasing order.)*

$$\lambda_1 = \ldots\ldots, \quad \lambda_2 = \ldots\ldots, \quad \textit{and} \quad \lambda_3 = \ldots\ldots$$

What do you observe about the eigenvalues?

What property of the matrix makes this "coincidence" possible?

The matrix in Exploration Problem **??** is a triangular matrix, and the property you observed holds in general.

Theorem 66. Let T be a triangular matrix. Then the eigenvalues of T are the entries on the main diagonal.

Proof See Practice Problem **??**. ■

Corollary 7. Let D be a diagonal matrix. Then the eigenvalues of D are the entries on the main diagonal.

One final note about eigenvalues. We began this section with the sentence, "In theory, then, to find the eigenvalues of A, one can solve Equation (**??**) for λ." In general, one does not attempt to compute eigenvalues by solving the characteristic equation of a matrix, as there is no simple way to solve this polynomial equation for $n > 4$. Instead, one can often approximate the eigenvalues using *iterative methods*. We will explore some of these techniques in The Power Method and the Dominant Eigenvalue.

Eigenvectors

Once we have computed an eigenvalue λ of an $n \times n$ matrix A, the next step is to compute the associated eigenvectors. In other words, we seek vectors $\mathbf{x}$ such that $A\mathbf{x} = \lambda\mathbf{x}$, or equivalently,

$$(A - \lambda I)\mathbf{x} = \mathbf{0} \tag{73}$$

As we have seen in the previous section, a (non-zero) scalar multiple of an eigenvector is an eigenvector. So for any given eigenvalue λ there are infinitely many eigenvectors associated with it. In fact, the eigenvectors associated with λ, together with the zero vector, form a subspace of $\mathbb{R}^n$.

Theorem 67. Let A be an $n \times n$ matrix and let λ be an eigenvalue of A. Then the set of all eigenvectors associated with λ, together with the zero vector, is a subspace of $\mathbb{R}^n$.

Proof See Practice Problems **??** and **??**. ■

This motivates the following definition.

Definition 71. The set of all eigenvectors associated with a given eigenvalue of a matrix, together with the zero vector, is known as the *eigenspace* associated with that eigenvalue.

Given an eigenvalue λ, there is an associated eigenspace $\mathcal{S}_\lambda$. Our goal is to find a basis of $\mathcal{S}_\lambda$. Then any eigenvector $\mathbf{x}$ associated with λ will be a linear combination of the vectors in that basis. In seeking a basis for $\mathcal{S}_\lambda$, we are trying to find a basis for the set of vectors that satisfy $(A - \lambda I)\mathbf{x} = \mathbf{0}$, which means we seek a basis for null$(A - \lambda I)$.

Let's return to the examples we did in the first part of this section.

Example 128. (Finding eigenvectors for Example **??**)

Recall that $A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$ has eigenvalues $\lambda_1 = 1$ and $\lambda_2 = 3$. Compute a basis for the eigenspace associated with each of these eigenvalues.

Explanation. Eigenvectors associated with the eigenvalue $\lambda_1 = 1$ are in the null space of $A - I$. So we seek a basis for null$(A - I)$. We compute:

$$\text{rref}(A - I) = \text{rref}\left(\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}\right) = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix},$$

From this we see that the eigenspace $\mathcal{S}_1$ associated with $\lambda_1 = 1$ consists of vectors of the form $\begin{bmatrix} -1 \\ 1 \end{bmatrix} t$. This means that $\left\{ \begin{bmatrix} -1 \\ 1 \end{bmatrix} \right\}$ is one possible basis for $\mathcal{S}_1$.

In a similar way, we compute a basis for $\mathcal{S}_3$, the subspace of all eigenvectors associated with the eigenvalue $\lambda_2 = 3$. Now we compute:

$$\text{rref}(A - 3I) = \text{rref}\left(\begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix}\right) = \begin{bmatrix} 1 & -1 \\ 0 & 0 \end{bmatrix},$$

Vectors in the null space have the form $\begin{bmatrix} 1 \\ 1 \end{bmatrix} t$ This means that $\left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\}$ is one possible basis for the eigenspace $\mathcal{S}_3$.

These computations reinforce the geometric insights into this problem that we gained in the previous section.

Example 129. (Finding eigenvectors for Example **??**) We know from Example **??** that $B = \begin{bmatrix} 2 & 1 \\ 4 & 2 \end{bmatrix}$ has eigenvalues $\lambda_1 = 0$ and $\lambda_2 = 4$. Compute a basis for the eigenspace associated with each of these eigenvalues.

Explanation. Let's begin by finding a basis for the eigenspace $\mathcal{S}_0$, which is the subspace of $\mathbb{R}^n$ consisting of eigenvectors corresponding to the eigenvalue $\lambda_1 = 0$. We need to compute a basis for null$(B - 0I)$ = null(B). We compute:

$$\text{rref}(B) = \text{rref}\left(\begin{bmatrix} 2 & 1 \\ 4 & 2 \end{bmatrix}\right) = \begin{bmatrix} 1 & \frac{1}{2} \\ 0 & 0 \end{bmatrix},$$

From this we see that an eigenvector in $\mathcal{S}_0$ has the form $\begin{bmatrix} -1/2 \\ 1 \end{bmatrix} t$. This means that $\left\{\begin{bmatrix} -1/2 \\ 1 \end{bmatrix}\right\}$ is one possible basis for the eigenspace $\mathcal{S}_0$. By letting $t = -2$, we obtain an arguably nicer-looking basis: $\left\{\begin{bmatrix} 1 \\ -2 \end{bmatrix}\right\}$.

To compute a basis for $\mathcal{S}_4$, the subspace of all eigenvectors associated to the eigenvalue $\lambda_2 = 4$, we compute:

$$\text{rref}(B - 4I) = \text{rref}\left(\begin{bmatrix} -2 & 1 \\ 4 & -2 \end{bmatrix}\right) = \begin{bmatrix} 1 & -\frac{1}{2} \\ 0 & 0 \end{bmatrix}$$

From this we find that $\left\{\begin{bmatrix} 1 \\ \dots \end{bmatrix}\right\}$ is one possible basis for the eigenspace $\mathcal{S}_4$.

Example 130. (Finding eigenvectors for Example **??**) We know from Example **??** that $C = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix}$ has eigenvalues $\lambda_1 = 1$ and $\lambda_2 = 4$. Compute a basis for the eigenspace associated to each of these eigenvalues.

Explanation. We first find a basis for the eigenspace $\mathcal{S}_1$. We need to compute a basis for null$(C - I)$. We compute:

$$\text{rref}(C - I) = \text{rref}\left(\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}\right) = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

Notice that there are two free variables. The eigenvectors in $\mathcal{S}_1$ have the form

$$\begin{bmatrix} -s - t \\ s \\ t \end{bmatrix} = s \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} + t \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$$

So one possible basis for the eigenspace $\mathcal{S}_1$ is given by $\left\{ \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \right\}$.

Next we find a basis for the eigenspace $\mathcal{S}_4$. We need to compute a basis for null$(C - 4I)$. We compute:

$$\text{rref}(C - 4I) = \text{rref}\left(\begin{bmatrix} -2 & 1 & 1 \\ 1 & -2 & 1 \\ 1 & 1 & -2 \end{bmatrix} \right) = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{bmatrix}$$

This time there is one free variable. The eigenvectors in $\mathcal{S}_4$ have the form $\begin{bmatrix} t \\ t \\ t \end{bmatrix}$, so a possible basis for the eigenspace $\mathcal{S}_4$ is given by $\left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right\}$.

Remark 11. Recall that $\lambda_1 = 1$ has algebraic multiplicity 2. Note that $\dim(\mathcal{S}_{\lambda_1}) = 2$. We say that λ_1 has geometric multiplicity 2. We have to be careful, and not draw hasty conclusions. Geometric and algebraic multiplicity are NOT always the same! We will revisit this issue later in the text.

Definition 72. The *geometric multiplicity* of an eigenvalue is the dimension of the eigenspace corresponding to it.

Example 131. (Finding eigenvectors for Example **??**) We know from Example **??** that $D = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & -1 & 1 \end{bmatrix}$ has eigenvalues $\lambda = 0$, $\lambda_1 = 1 + i$, and $\lambda_2 = 1 - i$. Compute a basis for the eigenspace associated with each eigenvalue.

Explanation. We first find a basis for the eigenspace $\mathcal{S}_0$. We need to compute a basis for null$(D - 0I)$ = null(D). We compute:

$$\text{rref}(D) = \text{rref}\left(\begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & -1 & 1 \end{bmatrix} \right) = \begin{bmatrix} 0 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix},$$

From this we see that for any eigenvector $\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$ in $\mathcal{S}_0$ we have $x_2 = 0$ and $x_3 = 0$, but x_1 is a free

variable. So one possible basis for the eigenspace $\mathcal{S}_0$ is given by

$$\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \right\}$$

Next we find a basis for the eigenspace $\mathcal{S}_{1+i}$. We need to compute a basis for null$(D - (1+i)I)$. We compute:

$$\begin{aligned} \text{rref}(D - (1+i)I) &= \text{rref}\left(\begin{bmatrix} -(1+i) & 0 & 0 \\ 0 & 1-(1+i) & 1 \\ 0 & -1 & 1-(1+i) \end{bmatrix} \right) \\ &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & i \\ 0 & 0 & 0 \end{bmatrix} \end{aligned}$$

There is one free variable. Setting $x_3 = t$, we get $x_1 = 0$ and $x_2 = ti$. From this we see that eigenvectors in $\mathcal{S}_{1+i}$ have the form $\begin{bmatrix} 0 \\ -i \\ 1 \end{bmatrix} t$, so a possible basis for the eigenspace $\mathcal{S}_{1+i}$ is given by $\left\{ \begin{bmatrix} 0 \\ -i \\ 1 \end{bmatrix} \right\}$. We ask you in Practice Problem **??** to show that $\left\{ \begin{bmatrix} 0 \\ i \\ 1 \end{bmatrix} \right\}$ is a basis for $\mathcal{S}_{1-i}$.

We conclude this section by establishing the significance of a matrix having an eigenvalue of zero.

Theorem 68. A square matrix has an eigenvalue of zero if and only if it is singular.

Proof A square matrix A is singular if and only if $\det A = 0$. But $\det A = 0$ if and only if $\det A - 0I = 0$, which is true if and only if zero is an eigenvalue of A. ■

Practice Problems

Problem 345 *Compute the eigenvalues of A, and find the corresponding eigenspaces.*

$$A = \begin{bmatrix} 4 & 1 \\ 8 & -3 \end{bmatrix}$$

Answer: (List the eigenvalues in an increasing order.)

$$\lambda_1 = \ldots\ldots, \quad \lambda_2 = \ldots\ldots$$

A basis for $\mathcal{S}_{\lambda_1}$ *is* $\left\{ \begin{bmatrix} \ldots\ldots \\ 1 \end{bmatrix} \right\}$. *A basis for* $\mathcal{S}_{\lambda_2}$ *is* $\left\{ \begin{bmatrix} \ldots\ldots \\ 1 \end{bmatrix} \right\}$.

Problem 346 *Compute the eigenvalues of A, and find the corresponding eigenspaces.*

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

Answer:

$$\lambda_1 = \ldots\ldots + \ldots\ldots i, \quad \lambda_2 = \ldots\ldots - \ldots\ldots i$$

A basis for $\mathcal{S}_{\lambda_1}$ *is* $\left\{ \begin{bmatrix} i \\ \ldots\ldots \end{bmatrix} \right\}$. *A basis for* $\mathcal{S}_{\lambda_2}$ *is* $\left\{ \begin{bmatrix} -i \\ \ldots\ldots \end{bmatrix} \right\}$.

Problem 347 *Let* $T = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 5 & 6 \\ 0 & 0 & 9 \end{bmatrix}$. *Compute a basis for each of the eigenspaces of this matrix,* $\mathcal{S}_1$, $\mathcal{S}_5$, *and* $\mathcal{S}_9$.

Problem 348 *Let*

$$A = \begin{bmatrix} 9 & 2 & 8 \\ 2 & -6 & -2 \\ -8 & 2 & -5 \end{bmatrix}$$

Question 348.1 *Compute the eigenvalues of A.*

Answer:

(List your answers in an increasing order.)

$$\lambda_1 = \ldots\ldots, \quad \lambda_2 = \ldots\ldots, \quad \lambda_3 = \ldots\ldots$$

Question 348.2 *Compute a basis for each of the eigenspaces of A,* $\mathcal{S}_{\lambda_1}$, $\mathcal{S}_{\lambda_2}$, *and* $\mathcal{S}_{\lambda_3}$.

Answer: A basis for $\mathcal{S}_{\lambda_1}$ *is* $\left\{\begin{bmatrix} 1 \\ \dots \\ \dots \end{bmatrix}\right\}$, *a basis for* $\mathcal{S}_{\lambda_2}$ *is* $\left\{\begin{bmatrix} -2 \\ \dots \\ \dots \end{bmatrix}\right\}$,

and a basis for $\mathcal{S}_{\lambda_3}$ *is* $\left\{\begin{bmatrix} 2 \\ \dots \\ \dots \end{bmatrix}\right\}$.

Problem 349 *Complete Example* **??** *by showing that a basis for* $\mathcal{S}_{1-i}$ *is given by* $\left\{\begin{bmatrix} 0 \\ i \\ 1 \end{bmatrix}\right\}$, *where* $\mathcal{S}_{1-i}$ *is the eigenspace associated with the eigenvalue* $\lambda = 1 - i$ *of the matrix* $D = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & -1 & 1 \end{bmatrix}$.

Problem 350 *Prove Theorem* **??**.

Problem 351 *Recall that a vertical stretch/compression of the plane is a linear transformation whose standard matrix is*

$$M_v = \begin{bmatrix} 1 & 0 \\ 0 & k \end{bmatrix}$$

Find the eigenvalues of M_v. *Find a basis for the eigenspace corresponding to each eigenvalue.*

Answer: A basis for $\mathcal{S}_1$ *is* $\left\{\begin{bmatrix} 1 \\ \dots \\ \dots \end{bmatrix}\right\}$ *and a basis for* $\mathcal{S}_k$ *is* $\left\{\begin{bmatrix} \dots \\ 1 \end{bmatrix}\right\}$

Sketch several vectors in each eigenspace and use geometry to explain why the eigenvectors you sketched make sense.

Problem 352 *Recall that a horizontal shear of the plane is a linear transformation whose standard matrix is*

$$M_{hs} = \begin{bmatrix} 1 & k \\ 0 & 1 \end{bmatrix}$$

Find the eigenvalue of M_{hs}.

Answer: $\lambda = \dots \dots$

Find a basis for the eigenspace corresponding to λ.

Answer: A basis for $\mathcal{S}$ is $\left\{ \begin{bmatrix} 1 \\ \dots \dots \end{bmatrix} \right\}$

What is the algebraic multiplicity of λ? ... ...

What is the geometric multiplicity of λ? ... ...

Problem 353 *Recall that a counterclockwise rotation of the plane through angle θ is a linear transformation whose standard matrix is*

$$M_\theta = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

Verify that the eigenvalues of M_θ are

$$\lambda = \cos\theta \pm \sqrt{\cos^2\theta - 1}$$

Use algebra, then use geometry to explain why λ is a real number if and only if θ is a multiple of π.

Suppose $\theta = \pi$. Then $\lambda =$ ... ... has algebraic multiplicity ... . Which of the following describes the corresponding eigenspace?

Multiple Choice:

1. *All vectors in $\mathbb{R}^2$.*
2. *All vectors along the x-axis.*
3. *All vectors along the y-axis.*
4. *All vectors along the line $y = x$.*

Problem 354 *Recall that a reflection of the plane about the line $y = mx$ is a linear transformation whose standard matrix is*

$$M_{y=mx} = \frac{1}{1+m^2} \begin{bmatrix} 1-m^2 & 2m \\ 2m & m^2-1 \end{bmatrix}$$

Verify that the eigenvalues of $M_{y=mx}$ are 1 and -1.

Find a basis for eigenspaces $\mathcal{S}_1$ and $\mathcal{S}_{-1}$. (For simplicity, assume that $m \neq 0$.)

Answer: A basis for $\mathcal{S}_1$ is $\left\{ \begin{bmatrix} 1 \\ \dots \dots \end{bmatrix} \right\}$ *and a basis for $\mathcal{S}_{-1}$ is* $\left\{ \begin{bmatrix} m \\ \dots \dots \end{bmatrix} \right\}$

Choose the best description of $\mathcal{S}_1$.

Multiple Choice:

1. *All vectors in $\mathbb{R}^2$.*
2. *All vectors with "slope" m.*
3. *All vectors with "slope" $1/m$.*
4. *All vectors with "slope" $-1/m$.*

Choose the best description of $\mathcal{S}_{-1}$.

Multiple Choice:

1. *All vectors along the line $y = mx$.*
2. *All vectors parallel to the x-axis.*
3. *All vectors parallel to the y-axis.*
4. *All vectors perpendicular to the line $y = mx$.*

Use geometry to explain why the eigenspaces you found make sense.

Problem 355 *Prove that the eigenvectors associated with an eigenvalue λ of an $n \times n$ matrix A, together with the zero vector, form a subspace of $\mathbb{R}^n$.*

Exercise Source

Practice Problem **??** is adopted from Problem 7.1.11 of Ken Kuttler's *A First Course in Linear Algebra*. (CC-BY)

Ken Kuttler, *A First Course in Linear Algebra*, Lyryx 2017, Open Edition, p. 361.

56 Similar Matrices and Their Properties

Let A and B be $n \times n$ matrices. Then the products AB and BA are both $n \times n$ matrices. In most cases the products AB and BA are not equal.

However, for some pairs of $n \times n$ matrices A and B, we are able to find an invertible matrix P such that $PB = AP$. This leads to the following definition.

Definition 73. If A and B are $n \times n$ matrices, we say that A and B are *similar*, if $B = P^{-1}AP$ for some invertible matrix P. In this case we write $A \sim B$.

The following theorem shows that *similarity* ($\sim$) satisfies *reflexive*, *symmetric*, and *transitive* properties.

Theorem 69. Similarity is an *equivalence relation*, i.e. for $n \times n$ matrices A, B, and C,

1. $A \sim A$ (reflexive)
2. If $A \sim B$, then $B \sim A$ (symmetric)
3. If $A \sim B$ and $B \sim C$, then $A \sim C$ (transitive)

Proof **??** It is clear that $A \sim A$ (let $P = I$).

?? If $A \sim B$, then for some invertible matrix P,

$$A = P^{-1}BP$$

and so

$$PAP^{-1} = B$$

But then

$$\left(P^{-1}\right)^{-1} AP^{-1} = B$$

which shows that $B \sim A$.

?? Now suppose $A \sim B$ and $B \sim C$. Then there exist invertible matrices P, Q such that

$$A = P^{-1}BP,\ B = Q^{-1}CQ$$

Then,

$$A = P^{-1}\left(Q^{-1}CQ\right)P = (QP)^{-1}C(QP)$$

showing that A is similar to C. ■

Any relation satisfying the reflexive, symmetric and transitive properties is called an *equivalence relation*. Theorem **??** proves that similarity between matrices is an *equivalence relation*. Practice Problem **??** gives a good example of a relation that is NOT an equivalence relation.

As we will see later, similar matrices share many properties. Before proceeding to explore these properties, we pause to introduce a simple matrix function that we will continue to use throughout the course.

Definition 74. The *trace* of an $n \times n$ matrix A, abbreviated tr$A$, is defined to be the sum of the main diagonal elements of A. In other words, if $A = [a_{ij}]$, then

$$\text{tr}(A) = a_{11} + a_{22} + \cdots + a_{nn}$$

We may also write $\text{tr}(A) = \sum_{i=1}^{n} a_{ii}$.

It is easy to see that $\text{tr}(A + B) = \text{tr}A + \text{tr}B$ and that $\text{tr}(cA) = c\,\text{tr}(A)$ holds for all $n \times n$ matrices A and B and all scalars c. The following fact is more surprising.

Theorem 70. Let A and B be $n \times n$ matrices. Then $\text{tr}(AB) = \text{tr}(BA)$.

Proof Write $A = [a_{ij}]$ and $B = [b_{ij}]$. For each i, the (i, i)-entry d_i of the matrix AB is given as follows: $d_i = a_{i1}b_{1i} + a_{i2}b_{2i} + \cdots + a_{in}b_{ni} = \sum_j a_{ij}b_{ji}$. Hence

$$\text{tr}(AB) = d_1 + d_2 + \cdots + d_n = \sum_i d_i = \sum_i \left(\sum_j a_{ij}b_{ji} \right)$$

Similarly we have $\text{tr}(BA) = \sum_i \left(\sum_j b_{ij}a_{ji} \right)$. Since these two double sums are the same, we have proven the theorem. ■

The following theorem lists a number of properties shared by similar matrices.

Theorem 71. If A and B are $n \times n$ matrices and $A \sim B$, then

1. $\det(A) = \det(B)$,
2. $\text{rank}(A) = \text{rank}(B)$,
3. $\text{tr}(A) = \text{tr}(B)$,

4. A and B have the same characteristic equations, and
5. A and B have the same eigenvalues.

Proof Let $B = P^{-1}AP$ for some invertible matrix P.

For **??**, $\det B = \det(P^{-1}) \det A \det P = \det A$ because $\det(P^{-1}) = 1/\det P$ (Theorem **??**).

Similarly, for **??** $\text{rank}B = \text{rank}(P^{-1}AP) = \text{rank}A$, because multiplication by an invertible matrix cannot change the rank. To see this, note that any invertible matrix is a product of elementary matrices. Multiplying by elementary matrices is equivalent to performing elementary row (column) operations on A, which does not change the row (column) space, nor the rank. It follows that similar matrices have the same rank.

For **??**, we make use of Theorem **??**:

$$\text{tr}(P^{-1}AP) = \text{tr}[P^{-1}(AP)] = \text{tr}[(AP)P^{-1}] = \text{tr}A.$$

As for **??**,

$$\begin{aligned}\det(B - \lambda I) &= \det\{P^{-1}AP - \lambda(P^{-1}P)\} \\ &= \det\{P^{-1}(A - \lambda I)P\} \\ &= \det(A - \lambda I),\end{aligned}$$

so A and B have the same characteristic equation. Finally, **??** implies **??** because the eigenvalues of a matrix are the roots of its characteristic polynomial. ■

Remark 12. Sharing the five properties in Theorem **??** does not guarantee that two matrices are similar. The matrices $A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ and $I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ have the same determinant, rank, trace, characteristic polynomial, and eigenvalues, but they are not similar because $P^{-1}IP = I$ for any invertible matrix P.

Even though the properties in Theorem **??** cannot be used to show two matrices are similar, these properties come in handy for showing that two matrices are NOT similar.

Example 132. Are the matrices $A = \begin{bmatrix} 2 & 1 \\ 1 & -1 \end{bmatrix}$ and $B = \begin{bmatrix} 3 & 0 \\ 1 & -1 \end{bmatrix}$ similar?

Explanation. A quick check shows us $\det A = \det B$, and both matrices are seen to be invertible, so they have the same rank. However, $\text{tr}A = 1$ and $\text{tr}B = 2$, so the matrices are not similar.

The next theorem shows that similarity is preserved under inverses, transposes, and powers:

Theorem 72. If A and B are $n \times n$ matrices and $A \sim B$, then

1. $A^{-1} \sim B^{-1}$,
2. $A^T \sim B^T$, and
3. $A^k \sim B^k$ for all integers $k \geq 1$.

Proof See Practice Problem **??**. ■

Practice Problems

Problem 356 *At the beginning of this section we mentioned that similarity of $n \times n$ matrices is an equivalence relation.*

An equivalence relation is a binary relation R on elements of a set S that has the following properties:

- *The reflexive property: xRx for every $x \in S$*
- *The symmetric property: If xRy, then yRx for every $x, y \in S$*
- *The transitive property: If xRy and yRz, then xRz for every $x, y, z \in S$*

Let S be the set of all positive integers. We can show that the relation "less than" (symbolized by $<$) is NOT an equivalence relation on this set. To see this, note that "less than" is not reflexive, because $s < s$ is not true for any positive integer s.

1. *Is the relation "less than" symmetric? (Yes/ No)*
2. *Is the relation "less than" transitive? (Yes/ No)*

Problem 357 *Another relation between matrices we have studied in this course is that two matrices can be "row equivalent". Is the relation "row equivalent"*

1. *reflexive? (Yes/ No)*
2. *symmetric? (Yes/ No)*
3. *transitive? (Yes/ No)*

Problem 358 *Use Theorem* **??** *to show that A and B are not similar.* $A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$, $B = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$

Problem 359 *Use Theorem* **??** *to show that A and B are not similar.* $A = \begin{bmatrix} 3 & 1 \\ -1 & 2 \end{bmatrix}$, $B = \begin{bmatrix} 2 & -1 \\ 3 & 2 \end{bmatrix}$

Problem 360 *Use Theorem* **??** *to show that A and B are not similar.* $A = \begin{bmatrix} 1 & 2 & -3 \\ 1 & -1 & 2 \\ 0 & 3 & -5 \end{bmatrix}$, $B = \begin{bmatrix} -2 & 1 & 3 \\ 6 & -3 & -9 \\ 0 & 0 & 0 \end{bmatrix}$

Problem 361 *Show that* $\begin{bmatrix} 1 & 2 & -1 & 0 \\ 2 & 0 & 1 & 1 \\ 1 & 1 & 0 & -1 \\ 4 & 3 & 0 & 0 \end{bmatrix}$ *and* $\begin{bmatrix} 1 & -1 & 3 & 0 \\ -1 & 0 & 1 & 1 \\ 0 & -1 & 4 & 1 \\ 5 & -1 & -1 & -4 \end{bmatrix}$ *are not similar.*

Problem 362 *Let*

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

Question 362.1 *The eigenvalues of A are known to be* $\lambda_1 = -3$, $\lambda_2 = 3$ *and* $\lambda_3 = 1$. *Find the corresponding eigenspaces.*

$$\mathcal{S}_{-3} = span\left(\begin{bmatrix} \ldots \ldots \\ \ldots \ldots \\ 1 \end{bmatrix} \right)$$

$$\mathcal{S}_{3} = span\left(\begin{bmatrix} \ldots \ldots \\ \ldots \ldots \\ 1 \end{bmatrix} \right)$$

$$\mathcal{S}_1 = span\left(\begin{bmatrix} \ldots \ldots \\ \ldots \ldots \\ 1 \end{bmatrix}\right)$$

We will refer to the eigenvectors you listed above as $\mathbf{x}_1$, $\mathbf{x}_2$ *and* $\mathbf{x}_3$.

Question 362.2 *Form matrix P whose columns are the eigenvectors* $\mathbf{x}_1$, $\mathbf{x}_2$, *and* $\mathbf{x}_3$ *you found in the previous part. Use technology to find* P^{-1}.

$$P^{-1} = \frac{1}{6}\begin{bmatrix} \ldots \ldots & \ldots \ldots & \ldots \ldots \\ \ldots \ldots & \ldots \ldots & \ldots \ldots \\ \ldots \ldots & \ldots \ldots & \ldots \ldots \end{bmatrix}$$

Question 362.3 *Find the product* $P^{-1}AP$. *What did you get? Which of the following is true?*

Multiple Choice:

1. $A \sim P$
2. $A \sim D$ *for ANY diagonal matrix* D
3. $A \sim D$ *for the diagonal matrix* D *whose diagonal entries are the eigenvalues of* A

Problem 363 *Prove that if* A *and* B *are* $n \times n$ *matrices and* $A \sim B$, *then*

1. $A^{-1} \sim B^{-1}$ *(provided that the inverses exist),*
2. $A^T \sim B^T$, *and*
3. $A^k \sim B^k$ *for all integers* $k \geq 1$.

(See Theorem **??***)*

Problem 364 *Let* λ *be an eigenvalue of* A *with corresponding eigenvector* $\mathbf{x}$. *If* $B = P^{-1}AP$ *is similar to* A, *show that* $P^{-1}\mathbf{x}$ *is an eigenvector of* B *corresponding to* λ.

Text Source

The text in this section is a compilation of material from Section 7.2.1 of Ken Kuttler's *A First Course in Linear Algebra* (CC-BY) and Section 5.5 of Keith Nicholson's *Linear Algebra with Applications* (CC-BY-NC-SA).

Ken Kuttler, *A First Course in Linear Algebra*, Lyryx 2017, Open Edition, p. 362-364.

Many of the Practice Problems are Exercises from W. Keith Nicholson, *Linear Algebra with Applications*, Lyryx 2018, Open Edition, pp. 298-310.

57 Diagonalizable Matrices and Multiplicity

Recall that a *diagonal matrix* D is a matrix containing a zero in every entry except those on the main diagonal. More precisely, if d_{ij} is the ij^{th} entry of a diagonal matrix D, then $d_{ij} = 0$ unless $i = j$. Such matrices look like the following.

$$D = \begin{bmatrix} * & & 0 \\ & \ddots & \\ 0 & & * \end{bmatrix}$$

where $*$ is a number which might not be zero.

Diagonal matrices have some nice properties, as we demonstrate below.

Exploration 365 *Let* $M = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$ *and let* $D = \begin{bmatrix} 2 & 0 & 0 \\ 0 & -5 & 0 \\ 0 & 0 & 10 \end{bmatrix}$. *Compute MD and DM using Octave.*

```
% Define M
M=[1 2 3; 4 5 6; 7 8 9]

% The function diag(vector) creates a square matrix with diagonal
%vector and the remaining entries zero.
D=diag([2 -5 10])

% The function diag() can also be used to return the diagonal of a matrix.
diag_M=diag(M)

% Understanding multiplication by a diagonal matrix
product_MD=M*D
product_DM=D*M

% Raising a diagonal matrix to a power
D5=D^5
```

Link to code.

Notice the patterns present in the product matrices.

- *Each row of DM is the same as its corresponding row of M multiplied by the scalar which is the corresponding diagonal element of D.*
- *In the product MD, it is the columns of M that have been multiplied by the diagonal elements.*
- *Raising D to the* 5^{th} *power amounts to raising each diagonal element to the* 5^{th} *power.*

These patterns hold in general for any diagonal matrix, and they are fundamental to understanding diagonalization *and its uses. We discuss diagonalization next.*

Definition 75. Let A be an $n \times n$ matrix. Then A is said to be *diagonalizable* if there exists an invertible matrix P such that

$$P^{-1}AP = D$$

where D is a diagonal matrix. In other words, a matrix A is diagonalizable if it is similar to a diagonal matrix, $A \sim D$.

If we are given a matrix A that is diagonalizable, then we can write $P^{-1}AP = D$ for some matrix P, or, equivalently,

$$AP = PD \tag{74}$$

If we pause to examine Equation (**??**), the work that we did in Exploration **??** can help us to understand how to find P that will diagonalize A. The product PD is formed by multiplying each column of P by a scalar which is the corresponding element on the diagonal of D. To restate this, if $\mathbf{x}_i$ is column i in our matrix P, then Equation (**??**) tells us that

$$A\mathbf{x}_i = \lambda_i \mathbf{x}_i, \tag{75}$$

where λ_i is the ith diagonal element of D.

Of course, Equation (**??**) is very familiar! We see that if we are able to diagonalize a matrix A, the columns of matrix P will be the eigenvectors of A, and the corresponding diagonal entries of D will be the corresponding eigenvalues of A. This is summed up in the following theorem.

Theorem 73. An $n \times n$ matrix A is diagonalizable if and only if there is an invertible matrix P given by

$$P = \begin{bmatrix} | & | & & | \\ \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_n \\ | & | & & | \end{bmatrix},$$

where the columns $\mathbf{x}_i$ are eigenvectors of A.

Moreover, if A is diagonalizable, the corresponding eigenvalues of A are the diagonal entries of the diagonal matrix D.

Proof Suppose P is given as above as an invertible matrix whose columns are eigenvectors of A. To show that A is diagonalizable, we will show

$$AP = PD,$$

which is equivalent to $P^{-1}AP = D$. We have

$$AP = \begin{bmatrix} | & | & & | \\ A\mathbf{x}_1 & A\mathbf{x}_2 & \cdots & A\mathbf{x}_n \\ | & | & & | \end{bmatrix},$$

while

$$PD = \begin{bmatrix} | & | & & | \\ \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_n \\ | & | & & | \end{bmatrix} \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix} = \begin{bmatrix} | & | & & | \\ \lambda_1\mathbf{x}_1 & \lambda_2\mathbf{x}_2 & \cdots & \lambda_n\mathbf{x}_n \\ | & | & & | \end{bmatrix}$$

We can complete this half of the proof by comparing columns, and noting that

$$A\mathbf{x}_i = \lambda_i\mathbf{x}_i \tag{76}$$

for $i = 1, \ldots, n$ since the $\mathbf{x}_i$ are eigenvectors of A and the λ_i are corresponding eigenvalues of A.

Conversely, suppose A is diagonalizable so that $P^{-1}AP = D$. Let

$$P = \begin{bmatrix} | & | & & | \\ \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_n \\ | & | & & | \end{bmatrix}$$

where the columns are the vectors $\mathbf{x}_i$ and

$$D = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix}$$

Then

$$AP = PD = \begin{bmatrix} | & | & & | \\ \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_n \\ | & | & & | \end{bmatrix} \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix}$$

and so

$$\begin{bmatrix} | & | & & | \\ A\mathbf{x}_1 & A\mathbf{x}_2 & \cdots & A\mathbf{x}_n \\ | & | & & | \end{bmatrix} = \begin{bmatrix} | & | & & | \\ \lambda_1\mathbf{x}_1 & \lambda_2\mathbf{x}_2 & \cdots & \lambda_n\mathbf{x}_n \\ | & | & & | \end{bmatrix}$$

showing the $\mathbf{x}_i$ are eigenvectors of A and the λ_i are eigenvalues. ■

Notice that because the matrix P defined above is invertible it follows that the set of eigenvectors of A, $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, , \mathbf{x}_n\}$, is a basis of $\mathbb{R}^n$.

We demonstrate the concept given in the above theorem in the next example. Note that not only are the columns of the matrix P formed by eigenvectors, but P must be invertible, and therefore must consist of a linearly independent set of eigenvectors.

Example 133. Let

$$A = \begin{bmatrix} 2 & 0 & 0 \\ 1 & 4 & -1 \\ -2 & -4 & 4 \end{bmatrix}$$

Find an invertible matrix P and a diagonal matrix D such that $P^{-1}AP = D$.

Explanation. We will use eigenvectors of A as the columns of P, and the corresponding eigenvalues of A as the diagonal entries of D. The eigenvalues of A are $\lambda_1 = 2, \lambda_2 = 2$, and $\lambda_3 = 6$. We leave these computations as exercises, as well as the computations to find a basis for each eigenspace.

One possible basis for $\mathcal{S}_2$, the eigenspace corresponding to 2, is $\left\{ \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \right\}$, while a basis for $\mathcal{S}_6$ is given by $\left\{ \begin{bmatrix} 0 \\ 1 \\ -2 \end{bmatrix} \right\}$.

We construct the matrix P by using these basis elements as columns.

$$P = \begin{bmatrix} -2 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & -2 \end{bmatrix}$$

You can verify (See Practice Problem) that

$$P^{-1} = \begin{bmatrix} -1/4 & 1/2 & 1/4 \\ 1/2 & 1 & 1/2 \\ 1/4 & 1/2 & -1/4 \end{bmatrix}$$

Thus,

$$\begin{aligned} P^{-1}AP &= \begin{bmatrix} -1/4 & 1/2 & 1/4 \\ 1/2 & 1 & 1/2 \\ 1/4 & 1/2 & -1/4 \end{bmatrix} \begin{bmatrix} 2 & 0 & 0 \\ 1 & 4 & -1 \\ -2 & -4 & 4 \end{bmatrix} \begin{bmatrix} -2 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & -2 \end{bmatrix} \\ &= \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 6 \end{bmatrix} \end{aligned}$$

You can see that the result here is a diagonal matrix where the entries on the main diagonal are the eigenvalues of A. Notice that eigenvalues on the main diagonal *must* be in the same order as the corresponding eigenvectors in P.

Exploration 366 *We will continue to use matrices from Example* **??** *to illustrate why diagonalizable matrices are easier to work with. Let* $A = \begin{bmatrix} 2 & 0 & 0 \\ 1 & 4 & -1 \\ -2 & -4 & 4 \end{bmatrix}$ *and let* $D = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 6 \end{bmatrix}$*. Would it be easier to compute* A^5 *or* D^5 *if you had to do so by hand, without a computer? Certainly* D^5 *is easier, due to the number of zero entries!*

Raising a diagonal matrix to a power amounts to simply raising each entry to that same power, whereas computing A^5 *requires many more calculations. However, we learned in Example* **??** *that A is similar to D. We will use this to make our computation easier as follows*

$$\begin{aligned} A^5 &= \left(PDP^{-1}\right)^5 \\ &= (PDP^{-1})(PDP^{-1})(PDP^{-1})(PDP^{-1})(PDP^{-1}) \\ &= PD(P^{-1}P)D(P^{-1}P)D(P^{-1}P)D(P^{-1}P)DP^{-1} \\ &= PD(I)D(I)D(I)D(I)DP^{-1} \\ &= PD^5P^{-1} \end{aligned}$$

With this in mind, it is not as daunting to calculate A^5 *by hand.*

$$A^5 = \begin{bmatrix} -2 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & -2 \end{bmatrix} \begin{bmatrix} 32 & 0 & 0 \\ 0 & 32 & 0 \\ 0 & 0 & 7776 \end{bmatrix} \begin{bmatrix} -1/4 & 1/2 & 1/4 \\ 1/2 & 1 & 1/2 \\ 1/4 & 1/2 & -1/4 \end{bmatrix} = \begin{bmatrix} 32 & 0 & 0 \\ 1936 & 3904 & -1936 \\ -3872 & -7744 & 3904 \end{bmatrix}$$

The savings in work would certainly be more pronounced for larger matrices or for powers larger that 5.

In Exploration **??**, because matrix A was diagonalizable, we were able to cut down on computations. When we chose to work with D and P instead of A we worked with the eigenvalues and eigenvectors of A. Each column of P is an eigenvector of A, and so we repeatedly made use of the following theorem (with $m = 5$).

Theorem 74. Let A be an $n \times n$ matrix and suppose $A\mathbf{x} = \lambda\mathbf{x}$. Then $A^m\mathbf{x} = \lambda^m\mathbf{x}$

Proof We prove this theorem by induction on m. Clearly $A^m\mathbf{x} = \lambda^m\mathbf{x}$ holds when $m = 1$, as that was given. For the inductive step, suppose that we know $A^{m-1}\mathbf{x} = \lambda^{m-1}\mathbf{x}$. Then

$$\begin{aligned} A^m\mathbf{x} &= (AA^{m-1})\mathbf{x} \\ &= A(A^{m-1}\mathbf{x}) \\ &= A(\lambda^{m-1}\mathbf{x}) \\ &= \lambda^{m-1}A\mathbf{x} \\ &= \lambda^{m-1}\lambda\mathbf{x} \\ &= \lambda^m\mathbf{x} \end{aligned}$$

as desired. ■

Matrix A from the Example **??** and Exploration **??** had a repeated eigenvalue of 2. The next theorem and corollary show that matrices which have *distinct* eigenvalues (where none are repeated) have desirable properties.

Theorem 75. Let A be an $n \times n$ matrix, and suppose that A has distinct eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_m$. For each i, let $\mathbf{x}_i$ be a λ_i-eigenvector of A. Then $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m\}$ is linearly independent.

Proof We prove this by induction on m, the number of vectors in the set. If $m = 1$, then $\{\mathbf{x}_1\}$ is a linearly independent set because $\mathbf{x}_1 \neq \mathbf{0}$. In general, suppose we have established that the theorem is true for some $m \geq 1$. Given eigenvectors $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{m+1}\}$, suppose

$$c_1\mathbf{x}_1 + c_2\mathbf{x}_2 + \cdots + c_{m+1}\mathbf{x}_{m+1} = \mathbf{0}. \tag{77}$$

We must show that each $c_i = 0$. Multiply both sides of (**??**) on the left by A and use the fact that $A\mathbf{x}_i = \lambda_i\mathbf{x}_i$ to get

$$c_1\lambda_1\mathbf{x}_1 + c_2\lambda_2\mathbf{x}_2 + \cdots + c_{m+1}\lambda_{m+1}\mathbf{x}_{m+1} = \mathbf{0}. \tag{78}$$

If we multiply (**??**) by λ_1 and subtract the result from (**??**), the first terms cancel and we obtain

$$c_2(\lambda_2 - \lambda_1)\mathbf{x}_2 + c_3(\lambda_3 - \lambda_1)\mathbf{x}_3 + \cdots + c_{m+1}(\lambda_{m+1} - \lambda_1)\mathbf{x}_{m+1} = \mathbf{0}.$$

Since $\mathbf{x}_2, \mathbf{x}_3, \ldots, \mathbf{x}_{m+1}$ correspond to distinct eigenvalues $\lambda_2, \lambda_3, \ldots, \lambda_{m+1}$, the set $\{\mathbf{x}_2, \mathbf{x}_3, \ldots, \mathbf{x}_{m+1}\}$ is linearly independent by the induction hypothesis. Hence,

$$c_2(\lambda_2 - \lambda_1) = 0, \quad c_3(\lambda_3 - \lambda_1) = 0, \quad \ldots, \quad c_{m+1}(\lambda_{m+1} - \lambda_1) = 0$$

and so $c_2 = c_3 = \cdots = c_{m+1} = 0$ because the λ_i are distinct. It follows that (**??**) becomes $c_1\mathbf{x}_1 = \mathbf{0}$, which implies that $c_1 = 0$ because $\mathbf{x}_1 \neq \mathbf{0}$, and the proof is complete. ■

The corollary that follows from this theorem gives a useful tool in determining if A is diagonalizable.

Corollary 8. Let A be an $n \times n$ matrix and suppose it has n distinct eigenvalues. Then it follows that A is diagonalizable.

Remark 13. Note that Corollary **??** is NOT an "if and only if statement". This means that if A has repeated eigenvalues it is still sometimes possible to diagonalize A, as seen in Example **??**.

Definition 76. If we are able to diagonalize A, say $A = PDP^{-1}$, we say that PDP^{-1} is an *eigenvalue decomposition* of A.

Not every matrix has an eigenvalue decomposition. Sometimes we cannot find an invertible matrix P such that $P^{-1}AP = D$. Consider the following example.

Example 134. Let

$$A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$$

If possible, find an invertible matrix P and a diagonal matrix D so that $P^{-1}AP = D$.

Explanation. We see immediately (how?) that the eigenvalues of A are $\lambda_1 = 1$ and $\lambda_2 = 1$. To find P, the next step would be to find a basis for the corresponding eigenspace $\mathcal{S}_1$. We solve the equation $(A - \lambda I)\,\mathbf{x} = \mathbf{0}$. Writing this equation as an augmented matrix, we already have a matrix in row echelon form:

$$\left[\begin{array}{cc|c} 0 & -1 & 0 \\ 0 & 0 & 0 \end{array}\right]$$

We see that the eigenvectors in $\mathcal{S}_1$ are of the form

$$\begin{bmatrix} t \\ 0 \end{bmatrix} = t \begin{bmatrix} 1 \\ 0 \end{bmatrix},$$

so a basis for the eigenspace $\mathcal{S}_1$ is given by $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$. It is easy to see that we cannot form an invertible matrix P, because any two eigenvectors will be of the form $\begin{bmatrix} t \\ 0 \end{bmatrix}$, and so the second row of P would have a row of zeros, and P could not be invertible. Hence A cannot be diagonalized.

We saw earlier in Corollary **??** that an $n \times n$ matrix with n distinct eigenvalues is diagonalizable. It turns out that there are other useful diagonalizability tests.

Recall that the *algebraic multiplicity* of an eigenvalue λ is the number of times that it occurs as a root of the characteristic polynomial.

Definition 77. The *geometric multiplicity* of an eigenvalue λ is the dimension of the corresponding eigenspace $\mathcal{S}_\lambda$.

Consider now the following lemma.

Lemma 8. Let A be an $n \times n$ matrix, and let $\mathcal{S}_{\lambda_1}$ be the eigenspace corresponding to the eigenvalue λ_1 which has algebraic multiplicity m. Then

$$\dim(\mathcal{S}_{\lambda_1}) \leq m$$

In other words, the geometric multiplicity of an eigenvalue is less than or equal to the algebraic multiplicity of that same eigenvalue.

Proof Let k be the geometric multiplicity of λ_1, i.e., $k = \dim(\mathcal{S}_{\lambda_1})$. Suppose $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k\}$ is a basis for the eigenspace $\mathcal{S}_{\lambda_1}$. Let P be any invertible matrix having $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k$ as its first k columns, say

$$P = \begin{bmatrix} | & | & & | & | & & | \\ \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_k & \mathbf{x}_{k+1} & \cdots & \mathbf{x}_n \\ | & | & & | & | & & | \end{bmatrix}.$$

In block form we may write

$$P = \begin{bmatrix} B & C \end{bmatrix} \quad \text{and} \quad P^{-1} = \begin{bmatrix} D \\ E \end{bmatrix}$$

where B is $n \times k$, C is $n \times (n-k)$, D is $k \times n$, and E is $(n-k) \times n$. We observe $I_n = P^{-1}P = \left[\begin{array}{c|c} DB & DC \\ \hline EB & EC \end{array}\right]$. This implies

$$DB = I_k, \quad DC = O_{k\ n-k}, \quad EB = O_{n-k\ k} \quad \text{and} \quad EC = I_{n-k}$$

Therefore,

$$P^{-1}AP = \begin{bmatrix} D \\ E \end{bmatrix} A \begin{bmatrix} B & C \end{bmatrix} = \left[\begin{array}{c|c} DAB & DAC \\ \hline EAB & EAC \end{array}\right]$$

$$= \left[\begin{array}{c|c} \lambda_1 DB & DAC \\ \hline \lambda_1 EB & EAC \end{array}\right] = \left[\begin{array}{c|c} \lambda_1 I_k & DAC \\ \hline O & EAC \end{array}\right]$$

We finish the proof by comparing the characteristic polynomials on both sides of this equation, and making use of the fact that similar matrices have the same characteristic polynomials.

$$\det(A - \lambda I) = \det(P^{-1}AP - \lambda I) = (\lambda_1 - \lambda)^k \det(EAC)$$

We see that the characteristic polynomial of A has $(\lambda_1 - \lambda)^k$ as a factor. This tells us that algebraic multiplicity of λ_1 is at least k, proving the desired inequality. ■

This result tells us that if λ is an eigenvalue of A, then the number of linearly independent λ-eigenvectors is never more than the multiplicity of λ. We now use this fact to provide a useful diagonalizability condition.

Theorem 76. Let A be an $n \times n$ matrix A. Then A is diagonalizable if and only if for each eigenvalue λ of A, the algebraic multiplicity of λ is equal to the geometric multiplicity of λ.

Proof Suppose A is diagonalizable and let $\lambda_1, \ldots, \lambda_t$ be the distinct eigenvalues of A, with algebraic multiplicities $m_1, \ldots, m_t$, respectively and geometric multiplicities $k_1, \ldots, k_t$, respectively. Since A is diagonalizable, Theorem **??** implies that $k_1 + \cdots + k_t = n$. By applying Lemma **??** t times, we have

$$n = k_1 + \cdots + k_t \leq m_1 + \cdots + m_t = n,$$

which is only possible if $k_i = m_i$ for $i = 1, \ldots, t$.

Conversely, if the geometric multiplicity equals the algebraic multiplicity of each eigenvalue, then obtaining a basis for each eigenspace yields n eigenvectors. Applying Theorem **??**, we know that these n eigenvectors are linearly independent, so Theorem **??** implies that A is diagonalizable. ■

Practice Problems

Problem 367 *Fill in the details of Example* **??***.*

Question 367.1 *Find the eigenvalues of matrix*

$$A = \begin{bmatrix} 2 & 0 & 0 \\ 1 & 4 & -1 \\ -2 & -4 & 4 \end{bmatrix}$$

Question 367.2 *Find a basis for each eigenspace of the matrix A.*

Question 367.3 *Compute the inverse of*

$$P = \begin{bmatrix} -2 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & -2 \end{bmatrix}$$

Question 367.4 *Compute* $D = P^{-1}AP$

Question 367.5 *Show that computing the inverse of P is not really necessary by comparing the products PD and AP.*

Problem 368 *Let* $A = \begin{bmatrix} 4 & -1 \\ -3 & 2 \end{bmatrix}$

Question 368.1 *Find the eigenvalues of A and list them in increasing order.*

$$\lambda_1 = \ldots\ldots, \quad \lambda_2 = \ldots\ldots$$

Question 368.2 *Find a diagonalization for A*

$$\begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} = P^{-1}AP = \frac{1}{4}\begin{bmatrix} 1 & \ldots\ldots \\ \ldots\ldots & 1 \end{bmatrix}\begin{bmatrix} 4 & -1 \\ -3 & 2 \end{bmatrix}\begin{bmatrix} 1 & \ldots\ldots \\ \ldots\ldots & 1 \end{bmatrix}$$

Question 368.3 *Find* A^6.

$$A^6 = \begin{bmatrix} \ldots\ldots & \ldots\ldots \\ \ldots\ldots & \ldots\ldots \end{bmatrix}$$

Problem 369 *Explain why the shearing matrix* $\begin{bmatrix} 1 & k \\ 0 & 1 \end{bmatrix}$*, where* $k \neq 0$ *is not diagonalizable.*

Problem 370 *Let*

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

Question 370.1 *Find eigenvalues of A and list them in increasing order.*

$$\lambda_1 = \ldots\ldots, \quad \lambda_2 = \ldots\ldots, \quad \lambda_3 = \ldots\ldots$$

Question 370.2 *Find a diagonalization for A.*

$$\begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix} = P^{-1}AP = \begin{bmatrix} \ldots & \ldots & \ldots \\ 0 & 0 & 1 \\ \ldots & \ldots & \ldots \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 1 & 2 & 1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \ldots & \ldots & \ldots \\ 1 & \ldots & 1 \\ \ldots & 1 & \ldots \end{bmatrix}$$

Problem 371 *Let*

$$A = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 2 & 2 \\ 2 & 3 & 1 \end{bmatrix}$$

$\lambda_1 = 4$ *is an eigenvalue of algebraic multiplicity ... .*

$$\mathcal{S}_{\lambda_1} = span\left(\begin{bmatrix} \ldots \\ \ldots \\ 1 \end{bmatrix} \right)$$

Geometric multiplicity of λ_1 *is ... .*

Explain why A is not diagonalizable.

Problem 372 *Let A denote an* $n \times n$ *upper triangular matrix. If all the main diagonal entries of A are distinct, show that A is diagonalizable.*

Problem 373 *Let A be a diagonalizable* $n \times n$ *matrix with eigenvalues* $\lambda_1, \lambda_2, \ldots, \lambda_n$ *(including multiplicities). Show that:*

1. $\det A = \lambda_1 \lambda_2 \cdots \lambda_n$
2. $trA = \lambda_1 + \lambda_2 + \cdots + \lambda_n$

Text Source

The text in this section is a compilation of material from Section 7.2.1 of Ken Kuttler's *A First Course in Linear Algebra* (CC-BY) and Section 5.5 of Keith Nicholson's *Linear Algebra with Applications* (CC-BY-NC-SA).

Ken Kuttler, *A First Course in Linear Algebra*, Lyryx 2017, Open Edition, p. 362-364.

Many of the Practice Problems are Exercises from W. Keith Nicholson, *Linear Algebra with Applications*, Lyryx 2018, Open Edition, pp. 298-310.

58 Gershgorin's Theorem

We have seen that eigenvalues are the roots of the characteristic polynomial, and therefore may be complex numbers, even when a matrix has entries that are real. This section is devoted to a remarkable theorem proved by S.A. Gershgorin in 1931. Gershgorin's theorem says that the n eigenvalues of an $n \times n$ matrix can be found in a region in the complex plane consisting of n disks. We explore the theorem for the case $n = 3$ before proving the result. If you have never plotted complex numbers before, you may wish to try Exploration **??** in Complex Numbers before proceeding.

Exploration 374 *In the GeoGebra exploration below, you can adjust any of the entries of the* 3×3 *matrix A. The three eigenvalues are plotted as dots in the complex plane. The Gershgorin region is the union of three closed disks whose boundaries are plotted in red. Notice that the eigenvalues are always in the Gershgorin region.*

As you experiment with the interactive, see if you can answer the following questions.

1. *Is it always true that each disk contains an eigenvalue? (Yes/ No)*
2. *What is the center of each disk?*
3. *What is the radius of each disk?*
4. *Are there some matrices for which one of the disks does not intersect the others? Where none of the disks intersect?*
5. *Try using a matrix with complex number entries. How is the visual pattern of eigenvalues for such a matrix different from a matrix whose entries are real numbers?*

Access GeoGebra interactives through the online version of this text at

https://ximera.osu.edu/linearalgebradzv3.

In the exploration above you were able to see the eigenvalues of a 3×3 matrix. The eigenvalues are contained in the Gershgorin region, which consists of three disks.

In general, for an $n \times n$ matrix $A = [a_{ij}]$, the Gershgorin region consists of n disks in the complex plane:

- The centers of the disks are given by the n diagonal entries of A.

- We get the radius of the disk with center a_{ii} by adding the absolute values of each of the other elements in row i.

We state this all formally in the following theorem.

Theorem 77 (Gershgorin)**.** Let $A = [a_{ij}]$ be an $n \times n$ complex matrix. Then for $1 < i < n$, we define the *absolute deleted row sum* to be

$$r_i(A) := \sum_{i \neq j} |a_{ij}|.$$

Now we can define the ith *Gershgorin disk* of A to be

$$\Gamma_i(A) := \{z \in \mathbb{C} : |z - a_{ii}| \leq r_i(A)\}. \tag{79}$$

The *Gershgorin region* of A is defined to be the union of the Gershgorin disks, i.e.,

$$\Gamma(A) := \bigcup_{1 \leq i \leq n} \Gamma_i(A).$$

Then the eigenvalues of A are contained in the Gershgorin region $\Gamma(A)$.

The fact that (**??**) represents a disk in the complex plane is a direct result of the complex distance formula (see Complex Numbers for more information). Before proving Gershgorin's theorem, we provide an example and a video to aid in understanding exactly what the theorem says.

YouTube link: `https://www.youtube.com/watch?v=D-87FPpLoXs`

Example 135. Plot the eigenvalues and the Gershgorin region of each of the following matrices.

1. $A = \begin{bmatrix} 1 & 1 \\ 2 & 0 \end{bmatrix}$
2. $B = \begin{bmatrix} 1+i & 1 & i & 0 \\ 0 & 4 & 1+i & 0 \\ 0 & 0 & 2 & -1 \\ 0 & 0 & 0 & -1 \end{bmatrix}$

Explanation. 1. $\Gamma(A)$ consists of two disks. The first is centered at 1 with a radius of 1. The second is centered at 0 with a radius of 2, so we see that the first disk is contained in the second. The eigenvalues, 2 and -1, are both contained in $\Gamma(A)$ (with 2 being on the boundary).

2. B is a matrix with complex entries, which makes for a more interesting Gershgorin region. See Complex Matrices for more on matrices like this one.

   Since B is upper triangular, its eigenvalues are its diagonal entries. These are the centers of the Gershgorin disks, which proves that the eigenvalues of B are contained in $\Gamma(B)$. $\Gamma(B)$ consists of four disks. The first is centered at $1 + i$ with a radius of 2, since the absolute value of i is 1. The second is centered at 4 with a radius of $\sqrt{2}$. The third disk is centered at 2 with a radius of 1. The fourth disk is centered at -1 with a radius of zero, so in fact, consists only of the eigenvalue -1.

Proof of Gershgorin's Theorem Let λ be an eigenvalue of A with corresponding eigenvector $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$ so that $A\mathbf{x} = \lambda\mathbf{x}$. Let x_k be an entry of $\mathbf{x}$ such that $|x_k|$ is greater than or equal to the absolute value of every other entry in $\mathbf{x}$. Then considering row k of the equation

$$A \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \lambda \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

gives us

$$\sum_{j=1}^{n} a_{kj}x_j = \lambda x_k.$$

Subtracting yields

$$\sum_{j \neq k} a_{kj}x_j = (\lambda - a_{kk})x_k.$$

Taking the absolute value of both sides, we have

$$|(\lambda - a_{kk})x_k| = |\lambda - a_{kk}||x_k| \tag{80}$$

$$= \left| \sum_{j \neq k} a_{kj}x_j \right| \tag{81}$$

$$\leq \sum_{j \neq k} |a_{kj}x_j| \tag{82}$$

$$= \sum_{j \neq k} |a_{kj}||x_j| \tag{83}$$

$$\leq \sum_{j \neq k} |a_{kj}||x_k| \tag{84}$$

Here equations **??** and **??** follow from property C2 of Complex Numbers, inequality **??** follows from the Triangle Inequality (C12 of Complex Numbers), and we arrive at inequality **??** by replacing each $\mathbf{x}_j$ in inequality **??** by $\mathbf{x}_k$.

These inequalties show that $|\lambda - a_{kk}| \leq \sum_{j \neq k} |a_{kj}|$, so $\lambda \in \Gamma_k(A)$, and hence $\lambda \in \Gamma(A)$, as desired.

■

Example 136. Sketch the Gershgorin region for each matrix and, if possible, use it as an aid for determining whether or not the matrix is singular.

1. $A = \begin{bmatrix} 4 & 1 & 2 \\ 2 & 5 & 1 \\ 2 & 1 & 6 \end{bmatrix}$
2. $B = \begin{bmatrix} 2 & -1 & 0 \\ 1 & 1 & -2 \\ -3 & 0 & 2 \end{bmatrix}$
3. $C = \begin{bmatrix} 3 & 4 & -1 \\ 0 & 1 & 2 \\ 0 & 0 & 6 \end{bmatrix}$

Explanation. We can check our sketches of the Gershgorin regions using the GeoGebra interactive in Exploration **??**.

1. If we sketch the Gershgorin region of matrix A, we see that the zero vector is not contained in it. It follows that zero cannot be an eigenvalue of matrix A, and we conclude that A is nonsingular.

2. The Gershgorin region of matrix B contains the zero vector, so there we cannot determine from the Gershgorin region whether or not B is singular. Instead, if we perform Gaussian elimination on matrix B, we will see it is not of full rank, and so B is singular.

3. The Gershgorin region of matrix C contains the zero vector, so again we cannot determine from the Gershgorin region whether or not C is singular. Observe that C is an upper triangular matrix with non-zero entries along the diagonal. Therefore $\det C \neq 0$. We conclude that C is non-singular.

Practice Problems

Problems **??-??**

Sketch the Gerhgorin region for each matrix. What, if anything, does the region tell us about the matrix being singular or nonsingular? If the test is inconclusive, test for singularity in some other way.

Problem 375 $A = \begin{bmatrix} 1 & i \\ -i & 1 \end{bmatrix}$

Problem 375.1 *Matrix A is (singular/ nonsingular).*

Problem 376 $B = \begin{bmatrix} 4 & 3-i \\ 3+i & 1 \end{bmatrix}$

Problem 376.1 *Matrix A is (singular/ nonsingular).*

Problem 377 $C = \begin{bmatrix} 2 & 1+i \\ 1-i & 3 \end{bmatrix}$

Problem 377.1 *Matrix A is (singular/ nonsingular).*

Problem 378 $D = \begin{bmatrix} 1 & 0 & 1+i \\ 0 & 2 & 0 \\ 1-i & 0 & 0 \end{bmatrix}$

Problem 378.1 *Matrix A is (singular/ nonsingular).*

Problem 379 $E = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1+i \\ 0 & 1-i & 2 \end{bmatrix}$

Problem 379.1 *Matrix A is (singular/ nonsingular).*

Problem 380 *Let $A = [a_{ij}]$ be an $n \times n$ matrix, and let*

$$r_i(A) := \sum_{j \neq i} |a_{ij}|$$

denote the sum of the absolute values of the non-diagonal entries in row i. We say that A is strictly diagonally dominant if

$$|a_{ii}| > r_i(A)$$

for $i = 1, \ldots, n$.

Prove that if a matrix is strictly diagonally dominant then it is nonsingular.

59 The Power Method and the Dominant Eigenvalue

We know eigenvalues as the roots of a characteristic polynomial of a matrix. In practice, the problem of finding eigenvalues is virtually never solved by finding roots of the characteristic polynomial, as this task is computationally prohibitive for large matrices. Iterative methods are used instead. In this section we describe the Power Method and some variations. In QR Factorization we will touch upon a better iterative method for approximating eigenvalues of a non-singular matrix called the QR-algorithm. The methods we present in this text are a mere introduction to the iterative methods used to compute eigenvalues and eigenvectors.

Finding the Dominant Eigenvalue using the Power Method

In Exploration **??** of Diagonalizable Matrices and Multiplicity, our initial rationale for diagonalizing matrices was to be able to compute the powers of a square matrix, and the eigenvalues were needed to do this. In this section, we are interested in efficiently computing eigenvalues, and it may come as no surprise that the first method we discuss uses the powers of a matrix.

Definition 78. An eigenvalue λ of an $n \times n$ matrix A is called a *dominant eigenvalue* if λ has multiplicity 1, and

$$|\lambda| > |\mu| \quad \text{for all eigenvalues } \mu \neq \lambda$$

Any corresponding eigenvector is called a *dominant eigenvector* of A.

When such an eigenvalue exists, one technique for finding it is as follows: Let $\mathbf{x}_0$ in $\mathbb{R}^n$ be a first approximation to a dominant eigenvector λ, and compute successive approximations $\mathbf{x}_1, \mathbf{x}_2, \ldots$ as follows:

$$\mathbf{x}_1 = A\mathbf{x}_0 \quad \mathbf{x}_2 = A\mathbf{x}_1 \quad \mathbf{x}_3 = A\mathbf{x}_2 \quad \cdots$$

In general, we define

$$\mathbf{x}_{k+1} = A\mathbf{x}_k \quad \text{for each } k \geq 0$$

We will see in this section that if the first estimate $\mathbf{x}_0$ is good enough, these vectors $\mathbf{x}_n$ will approximate the dominant eigenvector λ. This technique is called the *Power Method* (because $\mathbf{x}_k = A^k\mathbf{x}_0$ for each $k \geq 1$). Observe that if $\mathbf{z}$ is any eigenvector corresponding to λ, then

$$\frac{\mathbf{z} \cdot (A\mathbf{z})}{\|\mathbf{z}\|^2} = \frac{\mathbf{z} \cdot (\lambda\mathbf{z})}{\|\mathbf{z}\|^2} = \lambda$$

Because the vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n, \ldots$ approximate dominant eigenvectors, this suggests that we define the *Rayleigh quotients* as follows:

$$r_k = \frac{\mathbf{x}_k \cdot \mathbf{x}_{k+1}}{\|\mathbf{x}_k\|^2} \quad \text{for } k \geq 1$$

Then the numbers r_k approximate the dominant eigenvalue λ.

Exploration 381 *Use the power method to approximate a dominant eigenvector and eigenvalue of* $A = \begin{bmatrix} 1 & 1 \\ 2 & 0 \end{bmatrix}$.

The eigenvalues of A are 2 *and* -1*, with eigenvectors* $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ *and* $\begin{bmatrix} 1 \\ -2 \end{bmatrix}$*, as we can verify using Octave.*

```
A=[1 1; 2 0]
[EIGENVECTORS,EIGENVALUES] = eig(A)
```

Link to code.

For more information on using Octave to find eigenvalues, please refer to the Octave Tutorial section of this text.

If we take $\mathbf{x}_0 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ *as the first approximation and compute* $\mathbf{x}_1, \mathbf{x}_2, \ldots$ *successively, from* $\mathbf{x}_1 = A\mathbf{x}_0, \mathbf{x}_2 = A\mathbf{x}_1, \ldots$*, the result is*

$$\mathbf{x}_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \ \mathbf{x}_2 = \begin{bmatrix} 3 \\ 2 \end{bmatrix}, \ \mathbf{x}_3 = \begin{bmatrix} 5 \\ 6 \end{bmatrix}, \ \mathbf{x}_4 = \begin{bmatrix} 11 \\ 10 \end{bmatrix}, \ \mathbf{x}_3 = \begin{bmatrix} 21 \\ 22 \end{bmatrix}, \ \ldots$$

These vectors are approaching scalar multiples of the dominant eigenvector $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$*. Moreover, we will find that the Rayleigh quotients are*

$$r_1 = \frac{7}{5}, r_2 = \frac{27}{13}, r_3 = \frac{115}{61}, r_4 = \frac{451}{221}, \ldots$$

and these are approaching the dominant eigenvalue 2*.*

1. *Verify the above calculations using the following Octave code.*

2. *Try a different starting vector. Do we still get convergence?*
3. *What happens if the starting vector is changed to* $\mathbf{x}_0 = \begin{bmatrix} 1 \\ -2 \end{bmatrix}$*? Why doesn't this method work in this case?*
4. *Try using* $\mathbf{x}_0 = \begin{bmatrix} 1 \\ -2.01 \end{bmatrix}$*. You may need to increase the number of iterations to see convergence.*

```
% The Power Method
A=[1 1; 2 0]; %Input a square matrix
x_0=[1;0]; %Input an initial vector
maxit =5; %Choose number iterations
x = x_0;
count = 0;
r=0;
    while count < maxit
    y=x;
    x = A*y;
        if count > 0;
        r=(transpose(y)*x)/(transpose(y)*y);
        RQUOTIENTS(count)=r;
        end
    count = count + 1;
    VECTORS(:,count)=x;
end
VECTORS
RQUOTIENTS
```

Link to code.

To see why the power method works, let $\lambda_1, \lambda_2, \ldots, \lambda_m$ be eigenvalues of A with λ_1 dominant and let $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_m$ be corresponding eigenvectors. What is required is that the first approximation $\mathbf{x}_0$ be a linear combination of these eigenvectors:

$$\mathbf{x}_0 = a_1\mathbf{y}_1 + a_2\mathbf{y}_2 + \cdots + a_m\mathbf{y}_m \quad \text{with } a_1 \neq 0$$

If $k \geq 1$, the fact that $\mathbf{x}_k = A^k\mathbf{x}_0$ and $A^k\mathbf{y}_i = \lambda_i^k\mathbf{y}_i$ for each i gives

$$\mathbf{x}_k = a_1\lambda_1^k\mathbf{y}_1 + a_2\lambda_2^k\mathbf{y}_2 + \cdots + a_m\lambda_m^k\mathbf{y}_m \quad \text{for } k \geq 1$$

Hence

$$\frac{1}{\lambda_1^k}\mathbf{x}_k = a_1\mathbf{y}_1 + a_2\left(\frac{\lambda_2}{\lambda_1}\right)^k \mathbf{y}_2 + \cdots + a_m\left(\frac{\lambda_m}{\lambda_1}\right)^k \mathbf{y}_m$$

The right side approaches $a_1\mathbf{y}_1$ as k increases because λ_1 is dominant $\left(\left|\frac{\lambda_i}{\lambda_1}\right| < 1 \text{ for each } i > 1\right)$. Because $a_1 \neq 0$, this means that $\mathbf{x}_k$ approximates the dominant eigenvector $a_1\lambda_1^k\mathbf{y}_1$.

The power method requires that the first approximation $\mathbf{x}_0$ be a linear combination of eigenvectors. In Example **??**, the eigenvectors form a basis of $\mathbb{R}^2$. But even in this case, the method fails if $a_1 = 0$, where a_1 is the coefficient of the dominant eigenvector (as we saw in part **??**). In general, the rate of convergence is quite slow if any of the ratios $\left|\frac{\lambda_i}{\lambda_1}\right|$ is near 1 (as we saw in part **??**).

Exploration 382 *Use the Power Method to approximate a dominant eigenvector and eigenvalue of* $A = \begin{bmatrix} 9 & 2 & 8 \\ 2 & -6 & -2 \\ -8 & 2 & -5 \end{bmatrix}$

You can use Octave to employ the power method.

```
% The Power Method
A=[9 2 8; 2 -6 -2; -8 2 -5]; %Input a square matrix
xinit=[1;1;1];
%Input an initial vector
maxit =15;
%Choose number iterations
x = xinit;
count = 0;
r=0;
    while count < maxit
    y=x;
    x = A*y;
        if count > 0;
        r=(transpose(y)*x)/(transpose(y)*y);
        RQUOTIENTS(count)=r;
        end
    count = count + 1;
    VECTORS(:,count)=x;
end
```

```
VECTORS
RQUOTIENTS
[EIGENVECTORS, EIGENVALUES]=eig(A)
VECTORS(:,maxit)/norm(VECTORS(:,maxit))
```

Link to code.

We observe that the Rayleigh quotients approach -3, which is the dominant eigenvalue. We can confirm this using the command 'eig'. To see that our vectors are approaching the dominant eigenvector, in the last step we divide our final approximate vector by its norm to obtain a unit vector. Note that we get the same result as the 'eig' command. In practice, we achieve greater accuracy if we normalize the approximate eigenvector at each iteration rather than waiting until the end.

Finding Other Eigenvalues

If the Power Method converges to the dominant eigenvalue, how can we determine other eigenvalues of a matrix? This section is devoted to some variations of the Power Method that address this problem. Many of the ideas in this section can be employed in other settings.

The Inverse Power Method

The Inverse Power Method is based upon the following theorem.

Theorem 78. Let M be an invertible $n \times n$ matrix. Then the eigenvalues of M^{-1} are reciprocals of the eigenvalues of M.

Proof The proof is left as an exercise. See Practice Problem **??**. ■

We return to the matrix in Exploration **??**. Let $M = \begin{bmatrix} 9 & 2 & 8 \\ 2 & -6 & -2 \\ -8 & 2 & -5 \end{bmatrix}$ and let $A = M^{-1}$.

Now use Octave to verify that $A = \begin{bmatrix} 17/3 & 13/3 & 22/3 \\ 13/3 & 19/6 & 17/3 \\ -22/3 & -17/3 & -29/3 \end{bmatrix}$ and that the eigenvalues of A are indeed the reciprocals of the eigenvalues of M.

```
% Eigenvalues of the inverse are reciprocals of eigenvalues of the matrix
M=[9 2 8; 2 -6 -2; -8 2 -5];
eig(M)
A=inv(M)
eig(A)
```

Link to code.

If we now apply the Power Method to A, the Rayleigh quotients converge toward the dominant eigenvalue of A, which is -1. If we take the reciprocal of the dominant eigenvalue of A, we will have found the "least dominant" eigenvalue of M, i.e., the eigenvalue closest to zero. Sure enough, the reciprocal of -1 is -1, which is the eigenvalue of M closest to zero.

The Shifted Power Method

Again, we begin with a theorem.

Theorem 79. Let M be an invertible $n \times n$ matrix and let z be any number (real or complex). If λ is an eigenvalue of M, then $\lambda - z$ is an eigenvalue of $M - cI$.

Proof The proof is left as an exercise. See Practice Problem **??**. ■

Once again, let $M = \begin{bmatrix} 9 & 2 & 8 \\ 2 & -6 & -2 \\ -8 & 2 & -5 \end{bmatrix}$ and let $A = M - (-3)I$.

Using Octave, you can verify that $A = \begin{bmatrix} 12 & 2 & 8 \\ 2 & -3 & -2 \\ -8 & 2 & -2 \end{bmatrix}$ and that the eigenvalues of A are obtained by subtracting $c = -3$ from (i.e. by adding 3 to) each eigenvalue of M. The dominant eigenvalue of M gives us an eigenvalue of zero for A (you may find it gives a number very close to zero, due to rounding error). So if we apply the Power Method to A to find its dominant eigenvalue (which is 5), we can add -3 to get the "second largest" eigenvalue of M (i.e., the eigenvalue second-farthest from zero), which is 2.

To see this, we will use Octave.

```
% The shifted power method
M=[9 2 8;
2 -6 -2;
```

```
-8 2 -5];
eig(M)
c=-3;
A=M-c*eye(3)
eig(A)
xinit=[1;1;1]; %Input an initial vector
maxit =15; %Choose number iterations
x = xinit;
count = 0;
r=0;
    while count < maxit
    y=x;
    x = A*y;
        if count > 0;
        r=(transpose(y)*x)/(transpose(y)*y);
        RQUOTIENTS(count)=r;
        end
    count = count + 1;
    VECTORS(:,count)=x;
end
VECTORS
RQUOTIENTS
RQUOTIENTS(maxit-1)+c
```

Link to code.

The Shifted Inverse Power Method

The Shifted Inverse Power method combines the two ideas we just studied. Using the Inverse Power method, we know how to find the eigenvalue closest to zero. We also know that we can change which eigenvalue is the dominant eigenvalue using the Shifted Power method. With the Shifted Inverse Power method, we are able to find the eigenvalue closest to a target τ, where τ can be any number, real or complex. We begin with the following theorem.

Theorem 80. Let M be an $n \times n$ matrix and let λ be an eigenvalue of M. If τ is any number (real or complex) other than an eigenvalue of M, then $\dfrac{1}{\lambda - \tau}$ is an eigenvalue of $(M - \tau I)^{-1}$.

Proof Suppose $M\mathbf{x} = \lambda\mathbf{x}$. By Theorem **??** we have

$$(M - \tau I)\mathbf{x} = (\lambda - \tau)\mathbf{x}.$$

Since τ is not an eigenvalue, $\det(M - \tau I) \neq 0$, so $M - \tau I$ is invertible. Also, since τ is not an eigenvalue, $\lambda \neq \tau$. We multiply both sides on the left by $(M - \tau I)^{-1}$ and divide both sides by the nonzero scalar $\lambda - \tau$. We arrive at

$$\frac{1}{\lambda - \tau}\mathbf{x} = (M - \tau I)^{-1}\mathbf{x}$$

as desired. ■

Again we let $M = \begin{bmatrix} 9 & 2 & 8 \\ 2 & -6 & -2 \\ -8 & 2 & -5 \end{bmatrix}$. We will use the Shifted Inverse Power Method to compute the eigenvalue of M closest to the target $\tau = 1$.

Using Octave, you can verify that $A = \begin{bmatrix} 5.75 & 3.5 & 6.5 \\ 3.5 & 2 & 4 \\ -6.5 & -4 & -7.5 \end{bmatrix}$. When we apply the Power Method to A to find its dominant eigenvalue, our Rayleigh quotients approach 1. This means that the eigenvalue of M closest to our target $\tau = 1$ is $1 + \tau$, or 2.

We will use Octave.

```
% The shifted inverse power method
M=[9 2 8; 2 -6 -2; -8 2 -5];
eig(M)
tau=1;
A=inv(M-tau*eye(3)) %I love the fact that eye(n) is the command for the identity matrix
eig(A)
xinit=[1;1;1]; %Input an initial vector
maxit =15; %Choose number iterations
x = xinit;
count = 0;
r=0;
    while count < maxit
    y=x;
    x = A*y;
        if count > 0;
        r=(transpose(y)*x)/(transpose(y)*y);
        RQUOTIENTS(count)=r;
        end
    count = count + 1;
```

```
    VECTORS(:,count)=x;
end
VECTORS
RQUOTIENTS
RQUOTIENTS(maxit-1)+tau
```

Link to code.

Practice Problems

Problem 383 *1. Prove Theorem* **??**.

2. Prove Theorem **??**.

Problem 384 *In each case, find the exact eigenvalues and determine corresponding eigenvectors. Then start with* $\mathbf{x}_0 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ *and compute* $\mathbf{x}_4$ *and* r_3 *using the power method.*

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

2. $A = \begin{bmatrix} 5 & 2 \\ -3 & -2 \end{bmatrix}$

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

4. $A = \begin{bmatrix} 3 & 1 \\ 1 & 0 \end{bmatrix}$

Problem 385 *Apply the power method to* $A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$, *starting at* $\mathbf{x}_0 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$. *Does it converge? Explain.*

Text Source

This section was adapted from Section 8.5 of Keith Nicholson's *Linear Algebra with Applications*. (CC-BY-NC-SA)

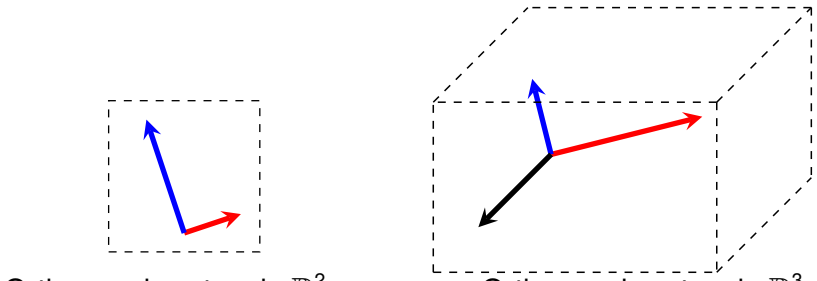
W. Keith Nicholson, *Linear Algebra with Applications*, Lyryx 2018, Open Edition, pp. 441–445.

60 Orthogonality

61 Orthogonality and Projections

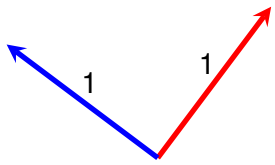
Orthogonal and Orthonormal Sets

In this section, we examine what it means for vectors (and sets of vectors) to be orthogonal and orthonormal. Recall that two non-zero vectors are orthogonal if their dot product is zero. A collection of non-zero vectors in $\mathbb{R}^n$ is called *orthogonal* if the vectors are pair-wise orthogonal. The diagram below shows two orthogonal vectors in $\mathbb{R}^2$ and three orthogonal vectors in $\mathbb{R}^3$.



Orthogonal vectors in $\mathbb{R}^2$ Orthogonal vectors in $\mathbb{R}^3$

If every vector in an orthogonal set of vectors is also a unit vector, then we say that the given set of vectors is *orthonormal*.



An orthonormal set of two vectors

Formally, we can define orthogonal and orthonormal vectors as follows.

Definition 79. Let $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_k\}$ be a set of nonzero vectors in $\mathbb{R}^n$. Then this set is called an *orthogonal set* if $\mathbf{v}_i \cdot \mathbf{v}_j = 0$ for all $i \neq j$. Moreover, if $\|\mathbf{v}_i\| = 1$ for $i = 1, \ldots, m$ (i.e. each vector in the set is a unit vector), we say the set of vectors is an *orthonormal set*.

An orthogonal set of vectors may not be orthonormal. To convert an orthogonal set to an orthonormal set, we need to divide each vector by its own length.

Definition 80. *Normalizing* an orthogonal set is the process of turning an orthogonal set into an orthonormal set. If $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ is an orthogonal subset of $\mathbb{R}^n$, then

$$\left\{ \frac{1}{\|\mathbf{v}_1\|}\mathbf{v}_1, \frac{1}{\|\mathbf{v}_2\|}\mathbf{v}_2, \ldots, \frac{1}{\|\mathbf{v}_k\|}\mathbf{v}_k \right\}$$

is an orthonormal set.

We illustrate this concept in the following example.

Example 137. Consider the vectors

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

Show that $\{\mathbf{v}_1, \mathbf{v}_2\}$ is an orthogonal set of vectors but not an orthonormal one. Find the corresponding orthonormal set.

Explanation. One easily verifies that $\mathbf{v}_1 \cdot \mathbf{v}_2 = 0$ and $\{\mathbf{v}_1, \mathbf{v}_2\}$ is an orthogonal set of vectors. On the other hand one can compute that $\|\mathbf{v}_1\| = \|\mathbf{v}_2\| = \sqrt{2} \neq 1$ and so the set is not orthonormal.

To find a corresponding orthonormal set, we need to normalize each vector.

$$\begin{aligned} \mathbf{q}_1 &= \frac{1}{\|\mathbf{v}_1\|}\mathbf{v}_1 \\ &= \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} \end{aligned}$$

Similarly,

$$\begin{aligned} \mathbf{q}_2 &= \frac{1}{\|\mathbf{v}_2\|}\mathbf{v}_2 \\ &= \frac{1}{\sqrt{2}} \begin{bmatrix} -1 \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} \end{aligned}$$

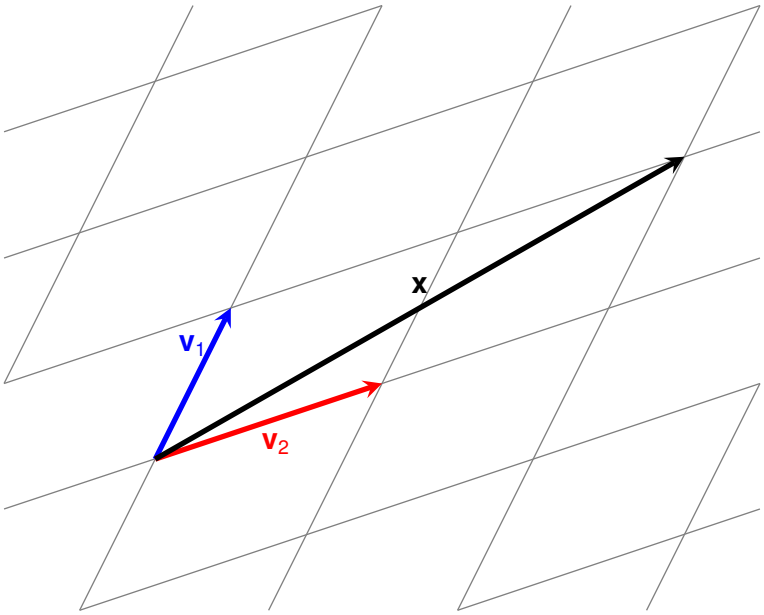
Therefore the corresponding orthonormal set is

$$\{\mathbf{q}_1, \mathbf{q}_2\} = \left\{ \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}, \begin{bmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} \right\}$$

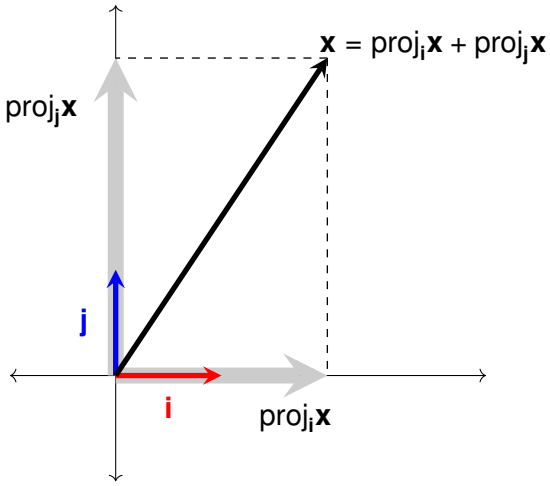
You can verify that this set is orthonormal.

Orthogonal and Orthonormal Bases

Recall that every basis of $\mathbb{R}^n$ (or a subspace W of $\mathbb{R}^n$) imposes a coordinate system on $\mathbb{R}^n$ (or W) that can be used to express any vector of $\mathbb{R}^n$ (or W) as a linear combination of the elements of the basis. For example, vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ impose a coordinate system onto the plane, as shown in the figure below. We readily see that $\mathbf{x}$, contained in the plane, can be written as $\mathbf{x} = \mathbf{v}_1 + 2\mathbf{v}_2$.



Vector $\mathbf{x}$ is visually easy to work with. In general, one way to express an arbitrary vector as a linear combination of the basis vectors is to solve a system of linear equations, which can be costly. One reason we like $\{\mathbf{i}, \mathbf{j}\}$ as a basis of $\mathbb{R}^2$ is because any vector $\mathbf{x}$ of $\mathbb{R}^2$ can be easily expressed as the sum of the orthogonal projections of $\mathbf{x}$ onto the basis vectors $\mathbf{i}$ and $\mathbf{j}$, as shown below.



We can see why an "upright" coordinate system with basis $\{\mathbf{i}, \mathbf{j}\}$ works well. What if we tilt this coordinate system while preserving the orthogonal relationship between the basis vectors? The following exploration allows you to investigate the consequences.

Exploration 386 *In the following GeoGebra interactive, vectors* $\mathbf{v}_1$ *and* $\mathbf{v}_2$ *are orthogonal (slopes of the lines containing them are negative reciprocals of each other). These vectors are clearly linearly independent and span* $\mathbb{R}^2$*. Therefore* $\{\mathbf{v}_1, \mathbf{v}_2\}$ *is a basis of* $\mathbb{R}^2$*.*

Let $\mathbf{x}$ *be an arbitrary vector. Orthogonal projections of* $\mathbf{x}$ *onto* $\mathbf{v}_1$ *and* $\mathbf{v}_2$ *are depicted in light grey.*

- *Use the tip of vector* $\mathbf{x}$ *to manipulate the vector and convince yourself that* $\mathbf{x}$ *is always the diagonal of the parallelogram (a rectangle!) determined by the projections.*
- *Use the tips of* $\mathbf{v}_1$ *and* $\mathbf{v}_2$ *to change the basis vectors. What happens when* $\mathbf{v}_1$ *and* $\mathbf{v}_2$ *are no longer orthogonal?*
- *Pick another pair of orthogonal vectors* $\mathbf{v}_1$ *and* $\mathbf{v}_2$*. Verify that* $\mathbf{x}$ *is the sum of its projections.*

Access GeoGebra interactives through the online version of this text at

https://ximera.osu.edu/linearalgebradzv3/LinearAlgebraInteractiveIntro.

As you have just discovered in Exploration **??**, we can express an arbitrary vector of $\mathbb{R}^2$ as the sum of its projections onto the basis vectors, provided that the basis is orthogonal. It turns out that this result holds for any subspace of $\mathbb{R}^n$, making a basis consisting of orthogonal vectors especially useful.

If an orthogonal set is a basis, we call it an *orthogonal basis*. Similarly, if an orthonormal set is a basis, we call it an *orthonormal basis*.

The following theorem generalizes our observation in Exploration **??**. As you read the statement of the theorem, it will be helpful to recall that the orthogonal projection of vector **x** onto a non-zero vector **d** is given by

$$\text{proj}_{\mathbf{d}}\mathbf{x} = \left(\frac{\mathbf{x}\cdot\mathbf{d}}{\|\mathbf{d}\|^2}\right)\mathbf{d} \tag{85}$$

(See Orthogonal Projections.)

Theorem 81. Let W be a subspace of $\mathbb{R}^n$ and suppose $\{\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_m\}$ is an orthogonal basis of W. Then for every **x** in W,

$$\mathbf{x} = \left(\frac{\mathbf{x}\cdot\mathbf{f}_1}{\|\mathbf{f}_1\|^2}\right)\mathbf{f}_1 + \left(\frac{\mathbf{x}\cdot\mathbf{f}_2}{\|\mathbf{f}_2\|^2}\right)\mathbf{f}_2 + \cdots + \left(\frac{\mathbf{x}\cdot\mathbf{f}_m}{\|\mathbf{f}_m\|^2}\right)\mathbf{f}_m. \tag{86}$$

Proof We may express **x** as a linear combination of the basis elements:

$$\mathbf{x} = c_1\mathbf{f}_1 + c_2\mathbf{f}_2 + \cdots + c_m\mathbf{f}_m.$$

We claim that $c_i = \dfrac{\mathbf{x}\cdot\mathbf{f}_i}{\|\mathbf{f}_i\|^2}$ for $i = 1, \ldots, m$. To see this, we take the dot product of each side with the vector $\mathbf{f}_i$ and obtain the following.

$$\mathbf{x}\cdot\mathbf{f}_i = (c_1\mathbf{f}_1 + c_2\mathbf{f}_2 + \cdots + c_m\mathbf{f}_m)\cdot\mathbf{f}_i$$

Our basis is orthogonal, so $\mathbf{f}_j\cdot\mathbf{f}_i = 0$ for all $j \neq i$, which means after we distribute the dot product, only one term will remain on the right-hand side. We have

$$\mathbf{x}\cdot\mathbf{f}_i = c_i\mathbf{f}_i\cdot\mathbf{f}_i$$

We now divide both sides by $\mathbf{f}_i\cdot\mathbf{f}_i = \|\mathbf{f}_i\|^2$, and since our claim holds for $i = 1, \ldots, m$, the proof is complete. ■

Theorem **??** shows one important benefit of a basis being orthogonal. With an orthogonal basis it is easy to represent any vector in terms of the basis vectors.

Example 138. Let $\mathbf{f}_1 = \begin{bmatrix}1\\-1\\2\end{bmatrix}$, $\mathbf{f}_2 = \begin{bmatrix}0\\2\\1\end{bmatrix}$, $\mathbf{f}_3 = \begin{bmatrix}5\\1\\-2\end{bmatrix}$, and let $\mathbf{x} = \begin{bmatrix}1\\1\\1\end{bmatrix}$.

Notice that $\mathcal{B} = \{\mathbf{f}_1, \mathbf{f}_2, \mathbf{f}_3\}$ is an orthogonal set of vectors, and $\mathcal{B}$ spans $\mathbb{R}^3$. Use this fact to write **x** as a linear combination of the vectors of $\mathcal{B}$.

Explanation. We first observe that $\mathcal{B}$ is a linearly independent set of vectors, and so $\mathcal{B}$ is a basis for $\mathbb{R}^3$. Next we apply Theorem **??** to express $\mathbf{x}$ as a linear combination of the vectors of $\mathcal{B}$. We wish to write:

$$\mathbf{x} = \left(\frac{\mathbf{x}\cdot\mathbf{f}_1}{\|\mathbf{f}_1\|^2}\right)\mathbf{f}_1 + \left(\frac{\mathbf{x}\cdot\mathbf{f}_2}{\|\mathbf{f}_2\|^2}\right)\mathbf{f}_2 + \left(\frac{\mathbf{x}\cdot\mathbf{f}_3}{\|\mathbf{f}_3\|^2}\right)\mathbf{f}_3.$$

We readily compute:

$$\frac{\mathbf{x}\cdot\mathbf{f}_1}{\|\mathbf{f}_1\|^2} = \frac{2}{6},\ \frac{\mathbf{x}\cdot\mathbf{f}_2}{\|\mathbf{f}_2\|^2} = \frac{3}{5},\ \text{and}\ \frac{\mathbf{x}\cdot\mathbf{f}_3}{\|\mathbf{f}_3\|^2} = \frac{4}{30}.$$

Therefore,

$$\begin{bmatrix}1\\1\\1\end{bmatrix} = \frac{1}{3}\begin{bmatrix}1\\-1\\2\end{bmatrix} + \frac{3}{5}\begin{bmatrix}0\\2\\1\end{bmatrix} + \frac{2}{15}\begin{bmatrix}5\\1\\-2\end{bmatrix}.$$

The formula from Theorem **??** is easy to use, and it becomes even easier when our basis is *orthonormal*.

Corollary 9. Let W be a subspace of $\mathbb{R}^n$ and suppose $\{\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_m\}$ is an orthonormal basis of W. Then for any $\mathbf{x}$ in W,

$$\mathbf{x} = (\mathbf{x}\cdot\mathbf{q}_1)\,\mathbf{q}_1 + (\mathbf{x}\cdot\mathbf{q}_2)\,\mathbf{q}_2 + \cdots + (\mathbf{x}\cdot\mathbf{q}_m)\,\mathbf{q}_m.$$

Proof This is a special case of Theorem **??**. Because $\|\mathbf{u_i}\| = 1$ for $i = 1, \ldots, m$, the terms are given by

$$\left(\frac{\mathbf{x}\cdot\mathbf{q}_i}{\|\mathbf{q}_i\|^2}\right)\mathbf{q}_i = (\mathbf{x}\cdot\mathbf{q}_i)\,\mathbf{q}_i.$$

■

Orthogonal Projection onto a Subspace

In the previous section we found that given a subspace W of $\mathbb{R}^n$ with an orthogonal basis $\mathcal{B}$, every vector $\mathbf{x}$ in W can be expressed as the sum of the orthogonal projections of $\mathbf{x}$ onto the elements of $\mathcal{B}$. Note that our premise was that $\mathbf{x}$ is in W. In this section, we look into the meaning of the sum of orthogonal projections of $\mathbf{x}$ onto the elements of an orthogonal basis of W for those vectors $\mathbf{x}$ of $\mathbb{R}^n$ that are *not* in W.

Exploration 387 *In the GeoGebra interactive below, W is a plane spanned by $\mathbf{v}_1$ and $\mathbf{v}_2$, in $\mathbb{R}^3$. W is subspace of $\mathbb{R}^3$. In the initial set up, $\mathbf{v}_1$ and $\mathbf{v}_2$ are orthogonal. Vector $\mathbf{x}$ is not in W.*

Use check-boxes to construct the sum of orthogonal projections of $\mathbf{x}$ onto $\mathbf{v}_1$ and $\mathbf{v}_2$. RIGHT-CLICK and DRAG to rotate the image.

Access GeoGebra interactives through the online version of this text at

https://ximera.osu.edu/linearalgebradzv3/LinearAlgebraInteractiveIntro.

Question 387.1 *If moved, return the basis vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ to their default position (set $s_1 = s_2 = 0$) to ensure that they are orthogonal.*

- *Rotate the image to convince yourself that the perpendiculars dropped from the tip of $\mathbf{x}$ to $\mathbf{v}_1$ and $\mathbf{v}_2$ are indeed perpendicular to $\mathbf{v}_1$ and $\mathbf{v}_2$ in the diagram. (You'll have to look at it just right to convince yourself of this.) Are both of these perpendiculars also necessarily perpendicular to the plane? (Yes, / No)*
- *Use sliders x_1, x_2 and x_3 to manipulate $\mathbf{x}$. Rotate the figure for a better view. What is true about about vector $\mathbf{p}$?*

  Multiple Choice:

  1. $\mathbf{p} = \mathbf{x} - (proj_{\mathbf{v}_1}\mathbf{x} + proj_{\mathbf{v}_2}\mathbf{x})$.
  2. *Vector $\mathbf{p}$ is orthogonal to W.*
  3. *All of the above.*

- *Rotate the figure so that you're looking directly down at the plane. If you're looking at it correctly, you will notice that (1) the parallelogram determined by the projections of $\mathbf{x}$ onto $\mathbf{v}_1$ and $\mathbf{v}_2$ is a rectangle; (2) the sum of projections, $proj_{\mathbf{v}_1}\mathbf{x} + proj_{\mathbf{v}_2}\mathbf{x}$, is located directly underneath $\mathbf{x}$, like a shadow at midday.*

Question 387.2 *Use sliders s_1 and s_2 to manipulate the basis vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ so that they are no longer orthogonal.*

- *Rotate the figure for a better view. Which of the following is true?*

  Multiple Choice:

  1. $\mathbf{p} = \mathbf{x} - (proj_{\mathbf{v}_1}\mathbf{x} + proj_{\mathbf{v}_2}\mathbf{x})$.
  2. *Vector $\mathbf{p}$ is orthogonal to W.*

3. *All of the above.*

- *Rotate your figure so that you're looking directly down at the plane. Which of the following is true?*

Multiple Choice:

1. *Parallelogram determined by* $\mathbf{v}_1$ *and* $\mathbf{v}_2$ *is a rectangle.*
2. $proj_{\mathbf{v}_1}\mathbf{x} + proj_{\mathbf{v}_2}\mathbf{x}$ *is located directly underneath* $\mathbf{x}$.
3. *None of the above.*

In Exploration **??**, you discovered that given a plane, spanned by orthogonal vectors $\mathbf{v}_1, \mathbf{v}_2$, in $\mathbb{R}^3$, and a vector $\mathbf{x}$, not in the plane, we can interpret the sum of orthogonal projections of $\mathbf{x}$ onto $\mathbf{v}_1$ and $\mathbf{v}_2$ as a "shadow" of $\mathbf{x}$ that lies in the plane directly underneath the vector $\mathbf{x}$. We say that this "shadow" is an *orthogonal projection of* $\mathbf{x}$ *onto* W. You have also found that if $\mathbf{v}_1, \mathbf{v}_2$ are not orthogonal, the parallelogram representing the sum of the orthogonal projections of $\mathbf{x}$ onto $\mathbf{v}_1$ and $\mathbf{v}_2$ will not be a rectangle. In this case, $\mathbf{x}$ minus this sum will NOT be orthogonal to the plane. It is essential that $\mathbf{v}_1, \mathbf{v}_2$ are orthogonal for $\text{proj}_{\mathbf{v}_1}\mathbf{x} + \text{proj}_{\mathbf{v}_2}\mathbf{x}$ to be considered an orthogonal projection.

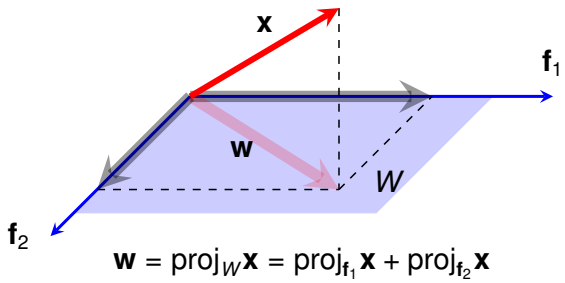
In general, we can define an orthogonal projection of $\mathbf{x}$ in $\mathbb{R}^n$ onto a subspace W of $\mathbb{R}^n$ as the sum of the orthogonal projections of $\mathbf{x}$ onto the elements of an orthogonal basis of W. Definition **??** and the subsequent diagram summarize this discussion.

Definition 81 (Projection onto a Subspace of $\mathbb{R}^n$)**.** Let W be a subspace of $\mathbb{R}^n$ with orthogonal basis $\{\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_m\}$. If $\mathbf{x}$ is in $\mathbb{R}^n$, the vector

$$\mathbf{w} = \text{proj}_W(\mathbf{x}) = \text{proj}_{\mathbf{f}_1}\mathbf{x} + \text{proj}_{\mathbf{f}_2}\mathbf{x} + \cdots + \text{proj}_{\mathbf{f}_m}\mathbf{x} \tag{87}$$

is called the *orthogonal projection* of $\mathbf{x}$ onto W.

An illustration of Definition **??** for a two-dimensional subspace W with orthogonal basis $\{\mathbf{f}_1, \mathbf{f}_2\}$ is shown below.



$\mathbf{w} = \text{proj}_W\mathbf{x} = \text{proj}_{\mathbf{f}_1}\mathbf{x} + \text{proj}_{\mathbf{f}_2}\mathbf{x}$

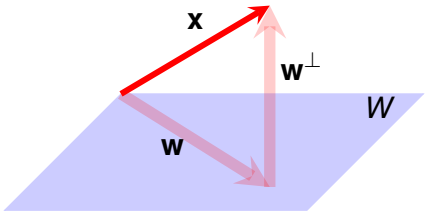
Using equation (**??**) multiple times, we can also express **w** in Definition **??** using the following formula.

Formula 21.

$$\mathbf{w} = \text{proj}_W(\mathbf{x}) = \frac{\mathbf{x}\cdot\mathbf{f}_1}{\|\mathbf{f}_1\|^2}\mathbf{f}_1 + \frac{\mathbf{x}\cdot\mathbf{f}_2}{\|\mathbf{f}_2\|^2}\mathbf{f}_2 + \cdots + \frac{\mathbf{x}\cdot\mathbf{f}_m}{\|\mathbf{f}_m\|^2}\mathbf{f}_m \tag{88}$$

Orthogonal Decomposition of x

Definition **??** allows us to express **x** as the sum of its orthogonal projection, $\mathbf{w} = \text{proj}_W\mathbf{x}$, located in W, and a vector we will call $\mathbf{w}^\perp$ (pronounced "W-perp"), given by $\mathbf{w}^\perp = \mathbf{x} - \mathbf{w}$. This decomposition of **x** is shown in the diagram below.



$$\mathbf{x} = \mathbf{w} + \mathbf{w}^\perp$$

You have already met $\mathbf{w}^\perp$, under the name of **p** in Exploration **??**, and observed that this vector is orthogonal to W. We will now prove that $\mathbf{w}^\perp$ is orthogonal to every vector in W. This will be accomplished in two steps. First, in Theorem **??** we will prove that $\mathbf{w}^\perp$ is orthogonal to all of the basis elements of W. Next, you will use this result to demonstrate that $\mathbf{w}^\perp$ is orthogonal to every vector in W.

Theorem 82. Let W be a subspace of $\mathbb{R}^n$ with orthogonal basis $\{\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_m\}$. Let **x** be in $\mathbb{R}^n$, and define $\mathbf{w}^\perp$ as

$$\mathbf{w}^\perp = \mathbf{x} - \text{proj}_W\mathbf{x} = \mathbf{x} - (\text{proj}_{\mathbf{f}_1}\mathbf{x} + \text{proj}_{\mathbf{f}_2}\mathbf{x} + \cdots + \text{proj}_{\mathbf{f}_m}\mathbf{x})$$

Then $\mathbf{w}^\perp$ is orthogonal to $\mathbf{f}_i$ for $1 \leq i \leq m$.

Proof We will use Formula **??** to show that $\mathbf{w}^\perp \cdot \mathbf{f}_i = 0$. Recall that $\{\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_m\}$ is an orthogonal basis. Therefore $\mathbf{f}_j \cdot \mathbf{f}_i = 0$ for $i \neq j$. This observation enables us to compute as follows.

$$\begin{aligned}
\mathbf{w}^\perp \cdot \mathbf{f}_i &= \left[\mathbf{x} - \left(\frac{\mathbf{x}\cdot\mathbf{f}_1}{\|\mathbf{f}_1\|^2}\mathbf{f}_1 + \cdots + \frac{\mathbf{x}\cdot\mathbf{f}_i}{\|\mathbf{f}_i\|^2}\mathbf{f}_i + \cdots + \frac{\mathbf{x}\cdot\mathbf{f}_m}{\|\mathbf{f}_m\|^2}\mathbf{f}_m\right)\right]\cdot\mathbf{f}_i \\
&= \mathbf{x}\cdot\mathbf{f}_i - \frac{\mathbf{x}\cdot\mathbf{f}_i}{\|\mathbf{f}_i\|^2}(\mathbf{f}_i\cdot\mathbf{f}_i) \\
&= \mathbf{x}\cdot\mathbf{f}_i - \frac{\mathbf{x}\cdot\mathbf{f}_i}{\|\mathbf{f}_i\|^2}\|\mathbf{f}_i\|^2 = \mathbf{x}\cdot\mathbf{f}_i - \mathbf{x}\cdot\mathbf{f}_i = 0
\end{aligned}$$

■

We leave the proof of the following Corollary as Practice Problem **??**

Corollary 10. Let W be a subspace of $\mathbb{R}^n$ with orthogonal basis $\{\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_m\}$. Let $\mathbf{x}$ be in $\mathbb{R}^n$, and define $\mathbf{w}^\perp$ as

$$\mathbf{w}^\perp = \mathbf{x} - \text{proj}_W \mathbf{x} = \mathbf{x} - (\text{proj}_{\mathbf{f}_1} \mathbf{x} + \text{proj}_{\mathbf{f}_2} \mathbf{x} + \cdots + \text{proj}_{\mathbf{f}_m} \mathbf{x})$$

Then $\mathbf{w}^\perp$ is orthogonal to every vector in W.

The fact that the decomposition of $\mathbf{x}$ into the sum of $\mathbf{w}$ and $\mathbf{w}^\perp$ is unique is the subject of the Orthogonal Decomposition Theorem which we will prove in Orthogonal Complements and Decompositions.

Throughout this section we have worked with orthogonal bases of subspaces. Does every subspace of $\mathbb{R}^n$ have an orthogonal basis? If so, how do we find one? These questions will be addressed in Gram-Schmidt Orthogonalization.

Practice Problems

Problem 388 *Retry Example* **??** *using Gaussian elimination.*

Problem 389 *Let* $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k \in \mathbb{R}^n$ *and suppose* $span\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k\} = \mathbb{R}^n$. *Furthermore, suppose that there exists a vector* $\mathbf{v} \in \mathbb{R}^n$ *for which* $\mathbf{v} \cdot \mathbf{x}_j = 0$ *for all* j, $1 \leq j \leq k$. *Show that* $\mathbf{v} = \mathbf{0}$.

Problem 390 *Let* $\mathbf{x} = \begin{bmatrix} 1 \\ -2 \\ 1 \\ 6 \end{bmatrix}$ *in* $\mathbb{R}^4$, *and let* $W = span\left(\begin{bmatrix} 2 \\ 1 \\ 3 \\ -4 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 0 \\ 1 \end{bmatrix} \right)$.

Question 390.1 *Compute* $proj_W(\mathbf{x})$.

Answer:

$$\frac{1}{10} \begin{bmatrix} \ldots \\ \ldots \\ \ldots \\ \ldots \end{bmatrix}$$

Question 390.2 *Show that* $\left\{ \begin{bmatrix} 1 \\ 0 \\ 2 \\ -3 \end{bmatrix}, \begin{bmatrix} 4 \\ 7 \\ 1 \\ 2 \end{bmatrix} \right\}$ *is another orthogonal basis of* W.

Question 390.3 *Use the basis you found in the previous part to compute* $proj_W(\mathbf{x})$.

Answer:

$$\frac{1}{70} \begin{bmatrix} \ldots \ldots \\ \ldots \ldots \\ \ldots \ldots \\ \ldots \ldots \end{bmatrix}$$

Problem 391 *Prove Corollary* **??**

Text Source

A portion of the text in this section is an adaptation of Section 4.11.1 of Ken Kuttler's *A First Course in Linear Algebra*. (CC-BY)

Ken Kuttler, *A First Course in Linear Algebra*, Lyryx 2017, Open Edition, p. 233-238.

62 Least-Squares Approximation

Often an exact solution to a problem in applied mathematics is difficult or impossible to obtain. However, it is usually just as useful to find an approximation to a solution. In particular, finding "linear approximations" is a powerful technique in applied mathematics. One basic case is the situation where a system of linear equations has no solution, and it is desirable to find a "best approximation" to a solution to the system.

We begin by defining the "best approximation" in a natural way, and showing that computing the best approximation reduces to solving a related system of linear equations called the *normal equations*. Next, we demonstrate a common application where a collection of data points is approximated by a line (or a curve). We conclude this section by showing that *QR*-factorization provides us with a more efficient way to solve the normal equations and compute the best approximation.

Best Approximate Solutions

Exploration 392 *Let*

$$A = \begin{bmatrix} 3 & 1 \\ 1 & 2 \\ 1 & 2 \end{bmatrix} \quad \text{and} \quad \mathbf{b} = \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix}$$

Consider the matrix equation $A\mathbf{x} = \mathbf{b}$*. A quick examination of the last two rows should convince you that this equation has no solutions. In other words,* $\mathbf{b}$ *is not in the span of the columns of A.*

If $\mathbf{z}$ *were an exact solution to* $A\mathbf{x} = \mathbf{b}$*, then* $\mathbf{b} - A\mathbf{z}$ *would be* $\mathbf{0}$*. Since the equation does not have a solution, we will attempt to find the next best thing to a solution by finding* $\mathbf{z}$ *such that* $\|\mathbf{b} - A\mathbf{z}\|$ *is as small as possible. The quantity* $\|\mathbf{b} - A\mathbf{z}\|$ *is called the* error*.*

The following GeoGebra interactive will help you understand the geometry behind finding $\mathbf{z}$*.*

RIGHT-CLICK and DRAG to rotate the image for a better view.

Access GeoGebra interactives through the online version of this text at

https://ximera.osu.edu/linearalgebradzv3/LinearAlgebraInteractiveIntro.

Record your best guess for $\mathbf{z}$ *– you will have a chance to check your answer in Example* **??***.*

What did you discover about the geometry of minimizing $\|\mathbf{b} - A\mathbf{z}\|$*? Select all that apply.*

Select All Correct Answers:

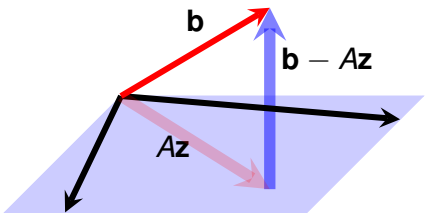
1. $\mathbf{z}$ *is orthogonal to the plane spanned by the columns of A.*

2. $\|\mathbf{b} - A\mathbf{z}\|$ *is orthogonal to col*(A).
3. $\mathbf{b} - A\mathbf{z}$ *is orthogonal to col*(A).
4. $A\mathbf{z}$ *is orthogonal to col*(A).
5. $A\mathbf{z}$ *is an orthogonal projection of* $\mathbf{b}$ *onto col*(A).

Our geometric observations will help us develop a method for finding $\mathbf{z}$.

Suppose A is an $m \times n$ matrix, and $\mathbf{b}$ is a column vector in $\mathbb{R}^m$. Consider the matrix equation $A\mathbf{x} = \mathbf{b}$. If this equation does not have a solution, we can attempt to find a best approximation by finding $\mathbf{z}$ which minimizes the *error*, $\|\mathbf{b} - A\mathbf{z}\|$. The expression $\|\mathbf{b} - A\mathbf{z}\|$ is also sometimes called the *residual*. The error (or the residual) is given in terms of a vector norm. Recall that our definition of the norm involves the sum of squares of the vector components. When we minimize the norm, we minimize the sum of squares. This is why the method we are describing is often referred to as *least squares*. We will explore this idea further later in this section.

In the case when col(A) is a subspace of $\mathbb{R}^3$, we can see geometrically that $\mathbf{z}$ is the best approximation if and only if $A\mathbf{z}$ is an orthogonal projection of $\mathbf{b}$ onto col(A), and the error is the magnitude of $\mathbf{b} - A\mathbf{z}$, as shown below.



What we observed above, holds in general. We will use this fact to find $\mathbf{z}$.

Every vector in col(A) can be written in the form $A\mathbf{x}$ for some $\mathbf{x}$ in $\mathbb{R}^m$. Our goal is to find $\mathbf{z}$ such that $A\mathbf{z}$ is the orthogonal projection of $\mathbf{b}$ onto col(A). By Corollary **??**, every vector $A\mathbf{x}$ in col(A) is orthogonal to $\mathbf{b} - A\mathbf{z}$.

Therefore, we have

$$\begin{aligned} A^T(\mathbf{b} - A\mathbf{z}) &= \mathbf{0} \\ A^T\mathbf{b} - A^TA\mathbf{z} &= \mathbf{0} \\ A^TA\mathbf{z} &= A^T\mathbf{b} \end{aligned} \tag{89}$$

Since $\mathbf{b} - A\mathbf{z}$ is normal to the subspace col(A), we call the system of linear equations in (**??**) the *normal equations for* $\mathbf{z}$. If A^TA is invertible, then we can write

$$\mathbf{z} = (A^TA)^{-1}A^T\mathbf{b} \tag{90}$$

We will return to the question of invertibility of A^TA in Theorem **??**. For now, let's revisit the problem posed in Exploration **??**.

Example 139. We now return to the matrix equation $A\mathbf{x} = \mathbf{b}$ of Exploration **??** to find $\mathbf{z}$ that best approximates a solution.

Explanation. Recall that

$$A = \begin{bmatrix} 3 & 1 \\ 1 & 2 \\ 1 & 2 \end{bmatrix} \quad \text{and} \quad \mathbf{b} = \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix}$$

In this case, $(A^TA)^{-1}$ exists. Applying equation (**??**), we compute

$$\begin{aligned} \mathbf{z} &= \left(\begin{bmatrix} 3 & 1 & 1 \\ 1 & 2 & 2 \end{bmatrix} \begin{bmatrix} 3 & 1 \\ 1 & 2 \\ 1 & 2 \end{bmatrix} \right)^{-1} \begin{bmatrix} 3 & 1 & 1 \\ 1 & 2 & 2 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix} \\ &= \begin{bmatrix} 11 & 7 \\ 7 & 9 \end{bmatrix}^{-1} \begin{bmatrix} 10 \\ 10 \end{bmatrix} = \begin{bmatrix} 0.18 & -0.14 \\ -0.14 & 0.22 \end{bmatrix} \begin{bmatrix} 10 \\ 10 \end{bmatrix} \\ &= \begin{bmatrix} 0.4 \\ 0.8 \end{bmatrix} \end{aligned}$$

Compare this answer to your guess in Exploration **??**. If your guess was correct, nice job! If your guess was different, try setting $\mathbf{z}$ to the correct answer and use the GeoGebra interactive in Exploration **??** to examine the geometry of the problem.

We now come back to the question of when A^TA is invertible.

Theorem 83. If columns of matrix A are linearly independent, then A^TA is invertible.

Proof Let A be a matrix with linearly independent columns. We will show that $\left(A^TA\right)\mathbf{x} = \mathbf{0}$ has only the trivial solution. For $\mathbf{x}$, a solution of $A^TA\mathbf{x} = \mathbf{0}$, we have

$$\begin{aligned} \|A\mathbf{x}\|^2 &= (A\mathbf{x}) \cdot (A\mathbf{x}) \\ &= (A\mathbf{x})^T A\mathbf{x} \\ &= \mathbf{x}^T A^T A\mathbf{x} \\ &= \mathbf{x}^T \cdot \mathbf{0} = 0 \end{aligned}$$

Therefore $A\mathbf{x} = \mathbf{0}$. By linear independence of the columns of A we conclude that $\mathbf{x} = \mathbf{0}$. ■

We summarize our findings in the following theorem.

Theorem 84. Let A be an $m \times n$ matrix, let $\mathbf{b}$ be a column vector in $\mathbb{R}^m$. Consider the matrix equation

$$A\mathbf{x} = \mathbf{b}$$

1. Any solution $\mathbf{z}$ to the normal equations

$$\left(A^T A\right)\mathbf{z} = A^T\mathbf{b}$$

   is a best approximation to a solution to $A\mathbf{x} = \mathbf{b}$ in the sense that $\|\mathbf{b} - A\mathbf{z}\|$ is minimized.

2. If the columns of A are linearly independent, then $A^T A$ is invertible and $\mathbf{z}$ is given uniquely by

$$\mathbf{z} = \left(A^T A\right)^{-1} A^T\mathbf{b}$$

Example 140. The sytem of linear equations

$$\begin{array}{rcrcl} 3x & - & y & = & 4 \\ x & + & 2y & = & 0 \\ 2x & + & y & = & 1 \end{array}$$

has no solution. Find the vector $\mathbf{z}$ that best approximates a solution.

Explanation. We have

$$A = \begin{bmatrix} 3 & -1 \\ 1 & 2 \\ 2 & 1 \end{bmatrix}, \quad \begin{bmatrix} 4 \\ 0 \\ 1 \end{bmatrix}.$$

The normal equations are $\left(A^T A\right)\mathbf{z} = A^T\mathbf{b}$. We compute

$$\begin{bmatrix} 3 & 1 & 2 \\ -1 & 2 & 1 \end{bmatrix} \begin{bmatrix} 3 & -1 \\ 1 & 2 \\ 2 & 1 \end{bmatrix} \mathbf{z} = \begin{bmatrix} 3 & 1 & 2 \\ -1 & 2 & 1 \end{bmatrix} \begin{bmatrix} 4 \\ 0 \\ 1 \end{bmatrix}.$$

$$\begin{bmatrix} 14 & 1 \\ 1 & 6 \end{bmatrix} \mathbf{z} = \begin{bmatrix} 14 \\ -3 \end{bmatrix}$$

We observe that is A^TA is invertible. Multiplying on the left by $(A^TA)^{-1}$ yields

$$\mathbf{z} = \frac{1}{83}\begin{bmatrix} 87 \\ -56 \end{bmatrix}$$

With this vector $\mathbf{z}$, the left sides of the equations become

$$\begin{array}{rcrcl} 3(87/83) & - & (-56/83) & \approx & 3.82 \\ (87/83) & + & 2(-56/83) & \approx & -0.30 \\ 2(87/83) & + & (-56/83) & \approx & 1.42 \end{array}$$

This is as close as possible to a solution.

Example 141. The average number g of goals per game scored by a hockey player seems to be related linearly to two factors: the number x_1 of years of experience and the number x_2 of goals in the preceding 10 games. The data on the following page were collected on four players. Find the linear function $g = a_0 + a_1x_1 + a_2x_2$ that best fits the data.

g	x_1	x_2
0.8	5	3
0.8	3	4
0.6	1	5
0.4	2	1

Explanation. If the relationship is given by $g = a_0 + a_1x_1 + a_2x_2$, then the data can be described as follows:

$$\begin{bmatrix} 1 & 5 & 3 \\ 1 & 3 & 4 \\ 1 & 1 & 5 \\ 1 & 2 & 1 \end{bmatrix}\begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} 0.8 \\ 0.8 \\ 0.6 \\ 0.4 \end{bmatrix}$$

Using Theorem **??**, we get

$$\mathbf{z} = \underbrace{\frac{1}{42}\begin{bmatrix} 119 & -17 & -19 \\ -17 & 5 & 1 \\ -19 & 1 & 5 \end{bmatrix}}_{(A^TA)^{-1}}\underbrace{\begin{bmatrix} 1 & 1 & 1 & 1 \\ 5 & 3 & 1 & 2 \\ 3 & 4 & 5 & 1 \end{bmatrix}}_{A^T}\begin{bmatrix} 0.8 \\ 0.8 \\ 0.6 \\ 0.4 \end{bmatrix} = \begin{bmatrix} 0.14 \\ 0.09 \\ 0.08 \end{bmatrix}$$

Hence the best-fitting function is $g = 0.14 + 0.09x_1 + 0.08x_2$.

Application of Least Squares to Curve Fitting

In Curve Fitting, we discussed how to fit a function to a set of data points, so that the graph of the function passes through each of the points. We also discussed why doing so is sometimes impossible (two points lie one above the other), and may not even be desirable (overfitting). In this section we will learn how to approximate a collection of data points with a line (or a curve) that fits the "trend" of the points. We will start with data that fit a linear pattern.

Exploration 393 *Consider the points* $(1,1)$, $(2,3)$ *and* $(4,4)$. *These points do not lie on a straight line, but they have a general upward linear trend. (Typically there would be many more points to consider, but we will limit our exploration to what we can do by hand.) Our goal is to find a line that fits these points as closely as possible.*

We are looking for a function f *of the form* $f(x) = ax + b$ *such that the following infeasible system is satisfied as closely as possible*

$$\begin{aligned} a(1) &+ b = 1 \\ a(2) &+ b = 3 \\ a(4) &+ b = 4 \end{aligned}$$

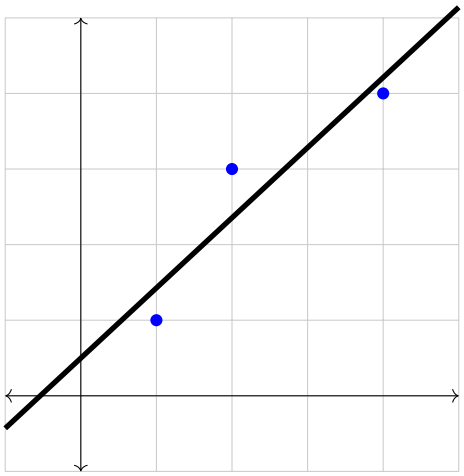
From the first part of this section we know how to find a best approximation. By Theorem **??**, *we have*

$$\begin{aligned} \mathbf{z} = \begin{bmatrix} a \\ b \end{bmatrix} &= \left(A^T A\right)^{-1} A^T \mathbf{b} \\ &= \left(\begin{bmatrix} 1 & 2 & 4 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 4 & 1 \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 & 2 & 4 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 3 \\ 4 \end{bmatrix} \\ &= \begin{bmatrix} 13/14 \\ 1/2 \end{bmatrix} \end{aligned}$$

According to our computations, the line that best fits the data is given by

$$f(x) = \frac{13}{14}x + \frac{1}{2}$$

Let's take a look.



We found this fit by minimizing $\|\mathbf{b} - A\mathbf{z}\|$*. We will now investigate the meaning of this expression in relation to the line and the data points.*

$$\mathbf{b} - A\mathbf{z} = \begin{bmatrix} 1 \\ 3 \\ 4 \end{bmatrix} - \begin{bmatrix} (13/14)(1) + 1/2 \\ (13/14)(2) + 1/2 \\ (13/14)(4) + 1/2 \end{bmatrix} \approx \begin{bmatrix} -0.43 \\ 0.64 \\ -0.21 \end{bmatrix} \tag{91}$$

Observe that each entry of $\mathbf{b} - A\mathbf{z}$ *is the signed vertical distance between a particular point and the line.*

Instead of computing the error, $\|\mathbf{b} - A\mathbf{z}\|$*, we will compute* $\|\mathbf{b} - A\mathbf{z}\|^2$ *to avoid the square root.*

$$\|\mathbf{b} - A\mathbf{z}\|^2 = (-0.43)^2 + 0.64^2 + (-0.21)^2 \approx 0.64 \tag{92}$$

Minimizing $\|\mathbf{b} - A\mathbf{z}\|$ *also minimizes* $\|\mathbf{b} - A\mathbf{z}\|^2$*. Therefore, what we have minimized is the sum of squares of the vertical distances between the data points and the line. The following GeoGebra interactive will help you explore this idea.*

Access GeoGebra interactives through the online version of this text at

https://ximera.osu.edu/linearalgebradzv3/LinearAlgebraInteractiveIntro.

In Exploration **??** we discovered that $\|\mathbf{b} - A\mathbf{z}\|^2$ is the sum of squares of vertical distances between the given data points and the proposed line. By minimizing $\|\mathbf{b} - A\mathbf{z}\|$, we minimize the sum of squares of vertical distances. This observation holds in general. Given a collection of points $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$, finding a linear function of the form $f(x) = ax + b$ that best fits

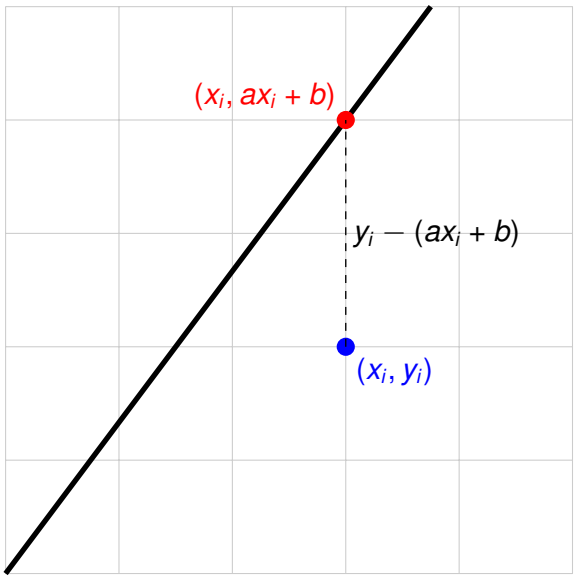
the points we would find a best solution to the system

$$\begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

by minimizing

$$\left\| \begin{bmatrix} y_1 - (ax_1 + b) \\ y_2 - (ax_2 + b) \\ \vdots \\ y_n - (ax_n + b) \end{bmatrix} \right\|^2 = (y_1 - (ax_1 + b))^2 + (y_2 - (ax_2 + b))^2 + \cdots + (y_n - (ax_n + b))^2$$

A geometric interpretation of $y_i - (ax_i + b)$ is shown below.



The line we obtain in this fashion is called a *line of best fit* or a *trendline*, and the method we used is referred to as the *method of least squares*.

We can apply the method of least squares to find best fitting non-linear functions.

Example 142. Find the least squares approximating quadratic polynomial of the form $f(x) = ax^2 + bx + c$ for the following points.

$$(-3, 3), (-1, 1), (0, 1), (1, 2), (3, 4)$$

Explanation. We are looking for an approximate solution to the system of equations

$$\begin{array}{rcrcrcl} a(-3)^2 & + & b(-3) & + & c & = & 3 \\ a(-1)^2 & + & b(-1) & + & c & = & 1 \\ a(0)^2 & + & b(0) & + & c & = & 1 \\ a(1)^2 & + & b(1) & + & c & = & 2 \\ a(3)^2 & + & b(3) & + & c & = & 4 \end{array}$$

This corresponds to the matrix equation

$$\begin{bmatrix} 9 & -3 & 1 \\ 1 & -1 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 1 \\ 9 & 3 & 1 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} 3 \\ 1 \\ 1 \\ 2 \\ 4 \end{bmatrix}$$

Multiplying on the left by A^T gives us the normal equations.

$$A^T A\mathbf{z} = A^T\mathbf{b}$$

$$\begin{bmatrix} 164 & 0 & 20 \\ 0 & 20 & 0 \\ 20 & 0 & 5 \end{bmatrix} \mathbf{z} = \begin{bmatrix} 66 \\ 11 \\ 4 \end{bmatrix}$$

It turns out that $A^T A$ is invertible, so it is easy to solve for $\mathbf{z}$. You can use technology to accomplish this.

We arrive at the solution

$$\mathbf{z} = \begin{bmatrix} 0.26 \\ 0.20 \\ 1.15 \end{bmatrix}$$

Therefore, the quadratic function of best fit is given by $f(x) = 0.26x^2 + 0.2x + 1.15$. You can see the graph and the points shown below.

For more information about doing least squares with Octave, please see Octave for Chapter 9.

Example 143. Given the data points $(-1, 0)$, $(0, 1)$, and $(1, 4)$, find the least squares approximating function of the form $f(x) = ax + b2^x$.

Explanation. We are looking for an approximate solution to the system of equations

$$\begin{array}{rcrcl} a(-1) & + & b(2^{-1}) & = & 0 \\ a(0) & + & b(2^{0}) & = & 1 \\ a(1) & + & b(2^{1}) & = & 4 \end{array}$$

This corresponds to the matrix equation

$$\begin{bmatrix} -1 & 1/2 \\ 0 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 4 \end{bmatrix}$$

Using the normal equations, we obtain

$$A^T A\mathbf{z} = A^T\mathbf{b}$$

$$\frac{1}{4}\begin{bmatrix} 8 & 6 \\ 6 & 21 \end{bmatrix}\mathbf{z} = \begin{bmatrix} 4 \\ 9 \end{bmatrix}$$

Solving for $\mathbf{z}$ yields

$$\mathbf{z} = \begin{bmatrix} 10/11 \\ 16/11 \end{bmatrix}$$

Therefore, the function of best fit (of the given form) is given by

$$f(x) = \frac{10}{11}x + \frac{16}{11}(2^x)$$

Practice Problems

Problem 394 *Find the best approximation to a solution to the system of equations.*

$$\begin{array}{rcrcrcl} 3x & + & y & + & z & = & 6 \\ 2x & + & 3y & - & z & = & 1 \\ 2x & - & y & + & z & = & 0 \\ 3x & - & 3y & + & 3z & = & 8 \end{array}$$

Enter answers in fraction form below.

$x =$ ..... , $y =$ ..... , $z =$ .....

Problem 395 *Find a linear function of best fit for the given set of data points. Examine how well your line fits the points by typing the equation of the line into the Desmos window.*

$$(2, 4), (4, 3), (7, 2), (8, 1)$$

Enter your answers in fraction form.

$$f(x) = \ldots \ldots x + \ldots \ldots$$

Problem 396 *Find a linear function of best fit for the given set of data points. Examine how well your line fits the points by typing the equation of the line into the Desmos window.*

$$(-2, 3), (-1, 1), (0, 0), (1, -2), (2, -4)$$

$$f(x) = \ldots \ldots x + \ldots \ldots$$

Problem 397 *Use Octave to find the least squares approximating quadratic function for the following data points.*

$$(-2, 1), (0, 0), (3, 2), (4, 3)$$

Round your answers to three decimal places.

$$f(x) = \ldots \ldots x^2 + \ldots \ldots x + \ldots \ldots$$

For more information about doing least squares with Octave, please see Octave for Chapter 9.

Text Source

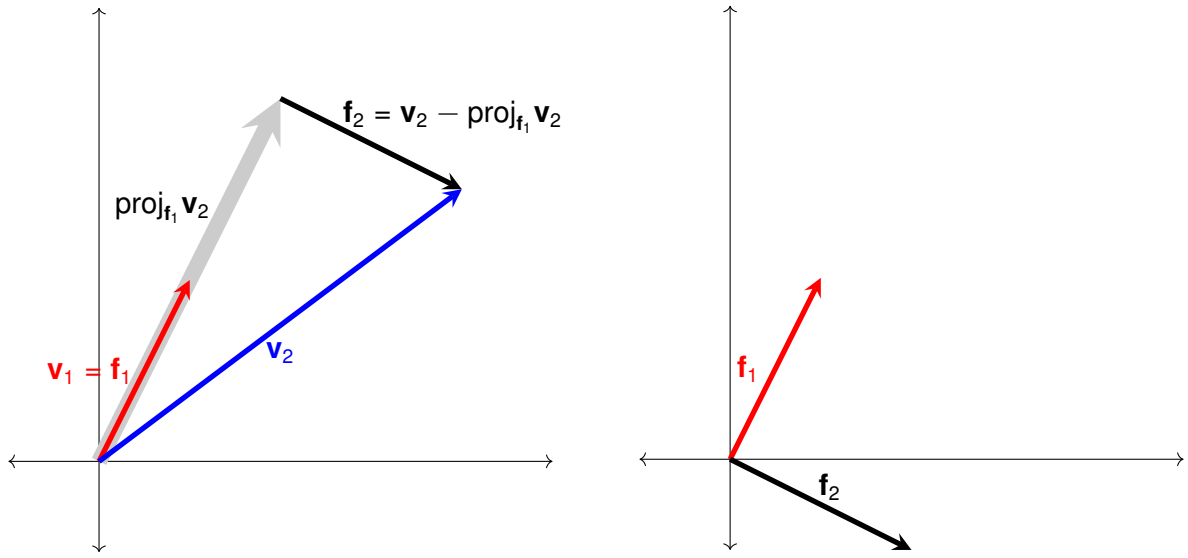
A portion of this section has been adapted from W. Keith Nicholson, *Linear Algebra with Applications*, Lyryx 2021-A, Open Edition, pp. 308-319.

63 Gram-Schmidt Orthogonalization

In Orthogonality and Projections we said that a set $\{\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_m\}$ of nonzero vectors in $\mathbb{R}^n$ is called an *orthogonal set* if $\mathbf{f}_i \cdot \mathbf{f}_j = 0$ for all $i \neq j$. In this section we will prove that every orthogonal set is linearly independent, and therefore it is a basis for its span. We have already seen that the expansion of a vector as a linear combination of orthogonal basis vectors is easy to obtain because formulas exist for the coefficients. Hence the orthogonal bases are the "nice" bases. Our next task is to show that every subspace of $\mathbb{R}^n$ *has* an orthogonal basis. We will start with intuitive explorations in lower dimensions, then proceed to formalize our results for subspaces of $\mathbb{R}^n$.

A Visual Guide to Creating an Orthogonal Basis

Given an arbitrary basis $\{\mathbf{v}_1, \mathbf{v}_2\}$ of $\mathbb{R}^2$, let's start building our orthogonal basis, $\{\mathbf{f}_1, \mathbf{f}_2\}$, by setting $\mathbf{f}_1 = \mathbf{v}_1$. To find the next element of our orthogonal basis, consider the orthogonal projection of $\mathbf{v}_2$ onto $\mathbf{f}_1$. (See the figure below.)



Next, let $\mathbf{f}_2 = \mathbf{v}_2 - \text{proj}_{\mathbf{f}_1}\mathbf{v}_2$. Observe that $\mathbf{f}_2$ is orthogonal to $\mathbf{f}_1$ (See Theorem **??** of Orthogonality and Projections). This gives us an orthogonal collection $\mathcal{B} = \{\mathbf{f}_1, \mathbf{f}_2\}$. It is intuitively clear that $\mathbf{f}_1$ and $\mathbf{f}_2$ are linearly independent. Therefore $\mathcal{B}$ is an orthogonal basis of $\mathbb{R}^2$.

The following exploration illustrates this process dynamically.

Exploration 398 *Choose an arbitrary basis $\{\mathbf{v}_1, \mathbf{v}_2\}$ of $\mathbb{R}^2$ by dragging the tips of vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ to desired positions. Use the navigation bar at the bottom of the interactive window to go through the steps of constructing an orthogonal basis of $\mathbb{R}^2$.*

Access GeoGebra interactives through the online version of this text at

https://ximera.osu.edu/linearalgebradzv3/LinearAlgebraInteractiveIntro.

We can apply this process to any two-dimensional subset of $\mathbb{R}^n$. The following exploration will guide you through the process of constructing an orthogonal basis for a plane spanned by two arbitrary vectors in $\mathbb{R}^3$.

Exploration 399 *Let $W = span(\mathbf{v}_1, \mathbf{v}_2)$. W is a plane through the origin in $\mathbb{R}^3$. Use the navigation bar at the bottom of the interactive window to go through the steps of constructing an orthogonal basis for W. RIGHT-CLICK and DRAG to rotate the image for a better view.*

Access GeoGebra interactives through the online version of this text at

https://ximera.osu.edu/linearalgebradzv3/LinearAlgebraInteractiveIntro.

In the next exploration, we take the process of constructing an orthogonal basis to the edge of the visual realm and construct an orthogonal basis for $\mathbb{R}^3$.

Exploration 400 *In the GeoGebra interactive below $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is a basis of $\mathbb{R}^3$. Use check boxes to go through the steps for constructing an orthogonal basis starting with the given basis. RIGHT-CLICK and DRAG to rotate the image for a better view.*

Access GeoGebra interactives through the online version of this text at

https://ximera.osu.edu/linearalgebradzv3/LinearAlgebraInteractiveIntro.

Gram-Schmidt Orthogonalization Algorithm

In Orthogonality and Projections we have repeatedly assumed that our subspaces of $\mathbb{R}^n$ have an orthogonal basis. We will now prove that this is indeed the case. Recall that to be a basis of a subspace, a set of vectors must be linearly independent and it must span the subspace. We will start by demonstrating that a set of orthogonal vectors must be linearly independent.

Theorem 85. Let $\{\mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_k\}$ be an orthogonal set of non-zero vectors in $\mathbb{R}^n$. Then this set is linearly independent.

Proof To show that this set is linearly independent, we need to demonstrate that the only solution to the following equation is the trivial solution.

$$a_1\mathbf{w}_1 + a_2\mathbf{w}_2 + \cdots + a_k\mathbf{w}_k = \mathbf{0}$$

To accomplish this, we need to show that all $a_i = 0$ for all $0 \leq i \leq k$. To do so we take the dot product of each side of the above equation with the vector $\mathbf{w}_i$ and obtain the following.

$$\begin{aligned} \mathbf{w}_i \cdot (a_1\mathbf{w}_1 + a_2\mathbf{w}_2 + \cdots + a_k\mathbf{w}_k) &= \mathbf{w}_i \cdot \mathbf{0} \\ a_1(\mathbf{w}_i \cdot \mathbf{w}_1) + a_2(\mathbf{w}_i \cdot \mathbf{w}_2) + \cdots + a_k(\mathbf{w}_i \cdot \mathbf{w}_k) &= 0 \end{aligned}$$

Now since the set is orthogonal, $\mathbf{w}_i \cdot \mathbf{w}_m = 0$ for all $m \neq i$, so we have:

$$a_1(0) + \cdots + a_i(\mathbf{w}_i \cdot \mathbf{w}_i) + \cdots + a_k(0) = 0$$

$$a_i \left\|\mathbf{w}_i\right\|^2 = 0$$

We know that $\left\|\mathbf{w}_i\right\|^2 \neq 0$, so it follows that $a_i = 0$. Since i was chosen arbitrarily, $a_i = 0$ for all i $(0 \leq i \leq k)$. This proves that $\{\mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_k\}$ is linearly independent.

■

The following theorem shows how to start with an arbitrary basis of a subspace W of $\mathbb{R}^n$ and find an orthogonal basis for W. To better understand the notation and the process presented in this theorem, you may want to match the steps of the theorem to the steps of Exploration **??**.

Theorem 86 (Gram-Schmidt Orthogonalization)**.** If $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m\}$ is any basis of a subspace W of $\mathbb{R}^n$, consider the following sequence of subspaces:

$$\begin{aligned} W_1 &= \text{span}\{\mathbf{v}_1\} \\ W_2 &= \text{span}\{\mathbf{v}_1, \mathbf{v}_2\} \\ W_3 &= \text{span}\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\} \\ &\vdots \\ W_m &= \text{span}\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \ldots, \mathbf{v}_m\} \end{aligned}$$

Then we can construct an orthogonal basis $\{\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_m\}$ for W_k for each $k = 2, 3, \ldots, m$ by adding one vector at a time successively as follows:

$$\begin{aligned} \mathbf{f}_1 &= \mathbf{v}_1 \\ \mathbf{f}_2 &= \mathbf{v}_2 - \text{proj}_{W_1}(\mathbf{v}_2) \\ \mathbf{f}_3 &= \mathbf{v}_3 - \text{proj}_{W_2}(\mathbf{v}_3) \\ &\vdots \\ \mathbf{f}_m &= \mathbf{v}_m - \text{proj}_{W_{m-1}}(\mathbf{v}_m) \end{aligned}$$

Then, $\{\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_m\}$ will be an orthogonal basis for W.

Proof Using the definition of projection onto a subspace, the iterative procedure above may be written:

$$\begin{aligned}
\mathbf{f}_1 &= \mathbf{v}_1 \\
\mathbf{f}_2 &= \mathbf{v}_2 - \frac{\mathbf{v}_2 \cdot \mathbf{f}_1}{\|\mathbf{f}_1\|^2}\mathbf{f}_1 \\
\mathbf{f}_3 &= \mathbf{v}_3 - \frac{\mathbf{v}_3 \cdot \mathbf{f}_1}{\|\mathbf{f}_1\|^2}\mathbf{f}_1 - \frac{\mathbf{v}_3 \cdot \mathbf{f}_2}{\|\mathbf{f}_2\|^2}\mathbf{f}_2 \\
&\vdots \\
\mathbf{f}_k &= \mathbf{v}_k - \frac{\mathbf{v}_k \cdot \mathbf{f}_1}{\|\mathbf{f}_1\|^2}\mathbf{f}_1 - \frac{\mathbf{v}_k \cdot \mathbf{f}_2}{\|\mathbf{f}_2\|^2}\mathbf{f}_2 - \cdots - \frac{\mathbf{v}_k \cdot \mathbf{f}_{k-1}}{\|\mathbf{f}_{k-1}\|^2}\mathbf{f}_{k-1} \\
&\vdots
\end{aligned} \tag{93}$$

We see immediately that $\text{span}\{\mathbf{f}_1\} = W_1$ and that $\text{span}\{\mathbf{f}_1, \mathbf{f}_2\} = W_2$ because $\mathbf{f}_2$ is a linear combination of $\mathbf{v}_1$ and $\mathbf{v}_2$. In fact, for any value of k, we see that $\text{span}\{\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_k\} = W_k$, because each $\mathbf{f}_k$ is a linear combination of the vectors $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{k-1}\}$.

Repeated application of Corollary **??** shows that the set $\{\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_m\}$ is orthogonal. Linear independence follows from orthogonality by Theorem **??**.

We conclude that $\{\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_m\}$ is a linearly independent orthogonal set that spans W. ■

Remark 14. Erhardt Schmidt (1876–1959) was a German mathematician who studied under the great David Hilbert. He first described the present algorithm in 1907. Jörgen Pederson Gram (1850–1916) was a Danish actuary.

Example 144. Find an orthogonal basis of the row space of $A = \begin{bmatrix} 1 & 1 & -1 & -1 \\ 3 & 2 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix}$.

Explanation. Let $\mathbf{v}_1$, $\mathbf{v}_2$, $\mathbf{v}_3$ denote the rows of A and observe that $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is linearly independent. Take $\mathbf{f}_1 = \mathbf{v}_1$. The algorithm gives

$$\begin{aligned}
\mathbf{f}_2 &= \mathbf{v}_2 - \frac{\mathbf{v}_2 \cdot \mathbf{f}_1}{\|\mathbf{f}_1\|^2}\mathbf{f}_1 = [3,2,0,1] - \frac{4}{4}[1,1,-1,-1] = [2,1,1,2] \\
\mathbf{f}_3 &= \mathbf{v}_3 - \frac{\mathbf{v}_3 \cdot \mathbf{f}_1}{\|\mathbf{f}_1\|^2}\mathbf{f}_1 - \frac{\mathbf{v}_3 \cdot \mathbf{f}_2}{\|\mathbf{f}_2\|^2}\mathbf{f}_2 = \mathbf{v}_3 - \frac{0}{4}\mathbf{f}_1 - \frac{3}{10}\mathbf{f}_2 = \frac{1}{10}[4,-3,7,-6]
\end{aligned}$$

Hence $\{[1,1,-1,-1], [2,1,1,2], \frac{1}{10}[4,-3,7,-6]\}$ is the orthogonal basis provided by the algorithm. In hand calculations it may be convenient to eliminate fractions (see the Remark below), so $\{[1,1,-1,-1], [2,1,1,2], [4,-3,7,-6]\}$ is also an orthogonal basis for row(A).

Remark 15. Observe that the vector $\dfrac{\mathbf{x}\cdot\mathbf{f}_i}{\|\mathbf{f}_i\|^2}\mathbf{f}_i$ is unchanged if a nonzero scalar multiple of $\mathbf{f}_i$ is used in place of $\mathbf{f}_i$. Hence, if a newly constructed $\mathbf{f}_i$ is multiplied by a nonzero scalar at some stage of the Gram-Schmidt algorithm, the subsequent **f**s will be unchanged. This is useful in actual calculations.

The Gram-Schmidt algorithm demonstrates in a constructive way that every subspace of $\mathbb{R}^n$ has an orthogonal basis. We formalize this in one final theorem.

Theorem 87. Let W be a subspace of $\mathbb{R}^n$. Then W has an orthogonal basis. In fact, every orthogonal subset $\{\mathbf{f}_1, \ldots, \mathbf{f}_m\}$ in W can be extended to an orthogonal basis for W.

Proof Suppose $\{\mathbf{f}_1, \ldots, \mathbf{f}_m\}$ is an orthogonal subset of W. If span $(\mathbf{f}_1, \ldots, \mathbf{f}_m) = W$, it is already a basis. Otherwise, there exists **x** in W outside span $(\mathbf{f}_1, \ldots, \mathbf{f}_m)$. Using the Gram-Schmidt procedure we define $\mathbf{f}_{m+1} = \mathbf{x} - \text{proj}_{W_m}(\mathbf{x})$, where $W_m = \text{span}\{\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_m\}$. If span $(\mathbf{f}_1, \ldots, \mathbf{f}_m, \mathbf{f}_{m+1}) = W$, we are done. Otherwise, the process continues to create larger and larger orthogonal subsets of W. They are all linearly independent by Theorem **??**, so we have a basis when we reach a subset containing dim(W) vectors. ■

The process described in the proof of this theorem is used in this final example.

Example 145. In Example **??**, given $A = \begin{bmatrix} 1 & 1 & -1 & -1 \\ 3 & 2 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix}$, we showed that an orthogonal basis for row(A) is given by

$$\{\mathbf{f}_1 = [1, 1, -1, -1], \mathbf{f}_2 = [2, 1, 1, 2], \mathbf{f}_3 = [4, -3, 7, -6]\}.$$

Choose any vector $\mathbf{v}_4 \in \mathbb{R}^4$ not in span$\{\mathbf{f}_1, \mathbf{f}_2, \mathbf{f}_3\}$, and apply the Gram-Schmidt algorithm to produce a vector $\mathbf{f}_4$ such that $\{\mathbf{f}_1, \mathbf{f}_2, \mathbf{f}_3, \mathbf{f}_4\}$ is an orthogonal basis for $\mathbb{R}^4$.

Explanation. Let $\mathbf{v}_4 = [1, 0, 0, 0]$. (How would you check that $\mathbf{v}_4$ is not in span$\{\mathbf{f}_1, \mathbf{f}_2, \mathbf{f}_3\}$?) To get a vector **f** orthogonal to the row space, we perform an iteration of Gram-Schmidt:

$$\begin{aligned} \mathbf{f} &= \mathbf{v}_4 - \frac{\mathbf{v}_4\cdot\mathbf{f}_1}{\|\mathbf{f}_1\|^2}\mathbf{f}_1 - \frac{\mathbf{v}_4\cdot\mathbf{f}_2}{\|\mathbf{f}_2\|^2}\mathbf{f}_2 - \frac{\mathbf{v}_4\cdot\mathbf{f}_3}{\|\mathbf{f}_3\|^2}\mathbf{f}_3 \\ &= [1, 0, 0, 0] - \frac{1}{4}[1, 1, -1, -1] - \frac{2}{10}[2, 1, 1, 2] - \frac{4}{110}[4, -3, 7, -6] \\ &= \frac{1}{44}[9, -15, -9, 3] \end{aligned}$$

Since any multiple of **f** will suffice, we are free to choose $\mathbf{f}_4 = 44\mathbf{f} = [9, -15, -9, 3]$ to get rid of the fraction. It is easy to check that $\{\mathbf{f}_1, \mathbf{f}_2, \mathbf{f}_3, \mathbf{f}_4\}$ is an orthogonal set, and it follows from Theorem **??** that this set is a basis for $\mathbb{R}^4$.

Remark 16. Suppose instead of $[1,0,0,0]$ we had started with $\mathbf{v}_4 = [7,-1,7,-5]$. This vector $\mathbf{v}_4$ is in $\text{span}\{\mathbf{f}_1,\mathbf{f}_2,\mathbf{f}_3\}$, as it is the sum of those three vectors. But if we were to try to proceed as above, we would get

$$\begin{aligned}\mathbf{f} &= \mathbf{v}_4 - \frac{\mathbf{v}_4\cdot\mathbf{f}_1}{\|\mathbf{f}_1\|^2}\mathbf{f}_1 - \frac{\mathbf{v}_4\cdot\mathbf{f}_2}{\|\mathbf{f}_2\|^2}\mathbf{f}_2 - \frac{\mathbf{v}_4\cdot\mathbf{f}_3}{\|\mathbf{f}_3\|^2}\mathbf{f}_3\\ &= [7,-1,7,-5] - \frac{4}{4}[1,1,-1,-1] - \frac{10}{10}[2,1,1,2] - \frac{110}{110}[4,-3,7,-6]\\ &= [0,0,0,0]\end{aligned}$$

We could not add a multiple of $\mathbf{f}$ to $\{\mathbf{f}_1,\mathbf{f}_2,\mathbf{f}_3\}$ to get an orthogonal basis for $\mathbb{R}^4$.

Practice Problems

Problem 401 *Try Example* **??** *again starting with some other vector* $\mathbf{v}_4 \in \mathbb{R}^4$.

Problems **??**-**??**

In each case, use the Gram-Schmidt algorithm to convert the given basis $\mathcal{B}$ of V to an orthogonal basis.

Problem 402 $V = \mathbb{R}^2$, $\mathcal{B} = \left\{\begin{bmatrix}1\\-1\end{bmatrix}, \begin{bmatrix}2\\1\end{bmatrix}\right\}$

Problem 403 $V = \mathbb{R}^2$, $\mathcal{B} = \left\{\begin{bmatrix}2\\1\end{bmatrix}, \begin{bmatrix}1\\2\end{bmatrix}\right\}$

Problem 404 $V = \mathbb{R}^3$, $\mathcal{B} = \left\{\begin{bmatrix}1\\-1\\1\end{bmatrix}, \begin{bmatrix}1\\0\\1\end{bmatrix}, \begin{bmatrix}1\\1\\2\end{bmatrix}\right\}$

Problem 405 $V = \mathbb{R}^3$, $\mathcal{B} = \left\{\begin{bmatrix}0\\1\\1\end{bmatrix}, \begin{bmatrix}1\\1\\1\end{bmatrix}, \begin{bmatrix}1\\-2\\2\end{bmatrix}\right\}$

Text Source

This section was adapted from the first part of Section 8.1 of Keith Nicholson's *Linear Algebra with Applications*. (CC-BY-NC-SA)

W. Keith Nicholson, *Linear Algebra with Applications*, Lyryx 2018, Open Edition, p. 415

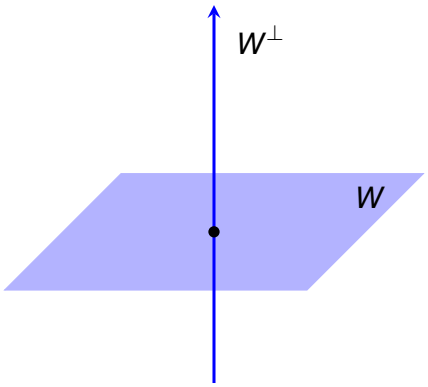
64 Orthogonal Complements and Decompositions

Orthogonal Complements

We will now consider the set of vectors that are orthogonal to every vector in a given subspace. As a quick example, consider the xy-plane in $\mathbb{R}^3$. Clearly, every scalar multiple of the standard unit vector $\mathbf{k}$ in $\mathbb{R}^3$ is orthogonal to every vector in the xy-plane. We say that the set $\{c\mathbf{k} \mid c \in \mathbb{R}\}$ is an *orthogonal complement* of $\{a\mathbf{i} + b\mathbf{j} \mid a, b \in \mathbb{R}\}$.

Definition 82 (Orthogonal Complement of a Subspace of $\mathbb{R}^n$)**.** If W is a subspace of $\mathbb{R}^n$, define the *orthogonal complement* $W^\perp$ of W (pronounced "W-perp") by

$$W^\perp = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{x} \cdot \mathbf{y} = 0 \text{ for all } \mathbf{y} \in W\}$$



The following theorem collects some useful properties of the orthogonal complement; the proof of **??** and **??** is left as Practice Problem **??**.

Theorem 88. Let W be a subspace of $\mathbb{R}^n$.

1. $W^\perp$ is a subspace of $\mathbb{R}^n$.
2. $\{\mathbf{0}\}^\perp = \mathbb{R}^n$ and $(\mathbb{R}^n)^\perp = \{\mathbf{0}\}$.
3. If $W = \text{span}\,(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k)$ then $W^\perp = \{\mathbf{x} \text{ in } \mathbb{R}^n \mid \mathbf{x} \cdot \mathbf{x}_i = 0 \text{ for } i = 1, 2, \ldots, k\}$.

Proof of Part ??: We must show that $W^\perp = \{\mathbf{x} \mid \mathbf{x} \cdot \mathbf{x}_i = 0 \text{ for each } i\}$. To show that two sets are equal, we must show that all elements of one set are included in the other set, and then we must show the reverse inclusion.

If $\mathbf{x}$ is in $W^\perp$ then $\mathbf{x} \cdot \mathbf{x}_i = 0$ for all i because each $\mathbf{x}_i$ is in W. This shows $W^\perp \subseteq \{\mathbf{x} \mid \mathbf{x} \cdot \mathbf{x}_i = 0 \text{ for each } i\}$. For the reverse inclusion, suppose that $\mathbf{x} \cdot \mathbf{x}_i = 0$ for all i; we need to show that $\mathbf{x}$ is in $W^\perp$. We need to show $\mathbf{x} \cdot \mathbf{y} = 0$ for each $\mathbf{y}$ in W. We can write $\mathbf{y} = c_1\mathbf{x}_1 + c_2\mathbf{x}_2 + \cdots + c_k\mathbf{x}_k$, where each c_i is in $\mathbb{R}$. Then

$$\mathbf{x} \cdot \mathbf{y} = c_1(\mathbf{x} \cdot \mathbf{x}_1) + c_2(\mathbf{x} \cdot \mathbf{x}_2) + \cdots + c_k(\mathbf{x} \cdot \mathbf{x}_k) = c_1 0 + c_2 0 + \cdots + c_k 0 = 0$$

as required, and the proof of equality is complete.

■

Example 146. Find a basis for $W^\perp$ if $W = \text{span}\left(\begin{bmatrix}1\\-1\\2\\0\end{bmatrix}, \begin{bmatrix}1\\0\\-2\\3\end{bmatrix}\right)$ in $\mathbb{R}^4$.

Explanation. By Theorem **??**, $\mathbf{x} = \begin{bmatrix}x\\y\\z\\w\end{bmatrix}$ is in $W^\perp$ if and only if $\mathbf{x}$ is orthogonal to both $\mathbf{v}_1 = \begin{bmatrix}1\\-1\\2\\0\end{bmatrix}$ and $\mathbf{v}_2 = \begin{bmatrix}1\\0\\-2\\3\end{bmatrix}$; that is, $\mathbf{x} \cdot \mathbf{v}_1 = 0$ and $\mathbf{x} \cdot \mathbf{v}_2 = 0$, or

$$\begin{aligned} x - y + 2z &= 0 \\ x - 2z + 3w &= 0 \end{aligned}$$

Using Gaussian elimination on this system gives $W^\perp = \text{span}\left(\begin{bmatrix}2\\4\\1\\0\end{bmatrix}, \begin{bmatrix}3\\3\\0\\-1\end{bmatrix}\right)$. You are asked to confirm this in Practice Problem **??** (which serves as a wonderful review of concepts we covered earlier in the course!).

Some of the important subspaces we studied earlier are orthogonal complements of each other. Recall the following definitions associated with an $m \times n$ matrix A.

1. The *null space* of A, $\text{null}(A) = \{\mathbf{x} \in \mathbb{R}^n \mid A\mathbf{x} = \mathbf{0}\}$, is a subspace of $\mathbb{R}^n$.

2. The *row space* of A, row(A) = span(the rows of A), is a subspace of $\mathbb{R}^n$.

3. The *column space* of A, col(A) = span(the columns of A), is a subspace of $\mathbb{R}^m$.

Exploration 406 *In the following GeoGebra interactive, you can change the coordinates of the vectors* $\mathbf{v}$ *and* $\mathbf{w}$ *using the sliders. (At this stage make sure that* $\mathbf{v}$ *and* $\mathbf{w}$ *are not collinear.) Let* $A = \begin{bmatrix} - & \mathbf{v} & - \\ - & \mathbf{w} & - \end{bmatrix}$. *Then row*($A$) = *span*($\mathbf{v}, \mathbf{w}$). *RIGHT-CLICK and DRAG to rotate the coordinate system for a better view.*

Access GeoGebra interactives through the online version of this text at

https://ximera.osu.edu/linearalgebradzv3/LinearAlgebraInteractiveIntro.

1. *Follow the prompts in the interactive to visualize row*(A) *and null*(A). *What relationships do you observe between row*(A) *and null*(A)*?*

2. *It is possible to "break" this interactive (for certain choices of* $\mathbf{v}$ *and* $\mathbf{w}$*). If* $\mathbf{v}$ *and* $\mathbf{w}$ *are scalar multiples of each other, then row*(A) *is a (Point, / Line,/ Plane), and the dimension of null*(A) *is (1, / 2,/ 3). The interactive does not accommodate this situation. To see what happens when* $\mathbf{v}$ *and* $\mathbf{w}$ *are scalar multiples of each other, see Practice Problem* **??**.

Theorem 89. Let A be an $m \times n$ matrix. Then we have:

1. null(A) = (row(A))$^{\perp}$;

2. null(A^T) = (col(A))$^{\perp}$.

Before proving this theorem, let's examine what it says about a couple of our examples. In Example **??**, we solved for the unknown vectors $\mathbf{x} = \begin{bmatrix} x \\ y \\ z \\ w \end{bmatrix}$. Notice that this is equivalent to creating a 2×4 matrix A whose rows are $\mathbf{v}_1$ and $\mathbf{v}_2$, and then finding the null space of that matrix A. You can check that a basis for null $\left(\begin{bmatrix} 1 & -1 & 2 & 0 \\ 1 & 0 & -2 & 3 \end{bmatrix} \right)$ is given by $\left\{ \begin{bmatrix} 2 \\ 4 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 3 \\ 3 \\ 0 \\ -1 \end{bmatrix} \right\}$.

Example 147. Let

$$A = \begin{bmatrix} 2 & -1 & 1 & -4 & 1 \\ 1 & 0 & 3 & 3 & 0 \\ -2 & 1 & -1 & 5 & 2 \\ 4 & -1 & 7 & 2 & 1 \end{bmatrix}$$

Verify each of the statements in Theorem **??**.

Explanation. We compute rref(A) to find a basis for null(A), row(A), and col(A). After some work we arrive at: $\text{null}(A) = \text{span}\left(\begin{bmatrix} -3 \\ -5 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 9 \\ 31 \\ 0 \\ -3 \\ 1 \end{bmatrix} \right)$ and

$$\text{row}(A) = \text{span}\left(\begin{bmatrix} 1 & 0 & 3 & 0 & -9 \end{bmatrix}, \begin{bmatrix} 0 & 1 & 5 & 0 & -31 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 & 1 & 3 \end{bmatrix} \right).$$

(See examples in Subspaces of $\mathbb{R}^n$ Associated with Matrices[1] for the details.) It is easy to check that each of the basis vectors of null(A) is orthogonal to each of the basis vectors of row(A), demonstrating the first part of Theorem **??**. You will be asked to demonstrate the second part of Theorem **??** for this example in Practice Problem **??**.

We now return to the proof of Theorem **??**.

Proof of Theorem ??: Let $\mathbf{x} \in \mathbb{R}^n$. $\mathbf{x} \in (\text{row}(A))^\perp$ if and only if x is orthogonal to every row of A. But this is true if and only if $A\mathbf{x} = \mathbf{0}$, which is equivalent to saying $\mathbf{x} \in \text{null}(A)$, which proves **??**. To prove **??**, we simply replace A with A^T, and we may apply **??** since $\text{col}(A) = \text{row}(A^T)$. ■

Orthogonal Decomposition Theorem

Now that we have defined the orthogonal complement of a subspace, we are ready to state the main theorem of this section. If you have studied physics or multi-variable calculus, you are familiar with the idea of expressing a vector in as the sum of its tangential and normal components. (If you haven't yet taken those courses, this section will help to prepare you for them!) The following theorem is a generalization of this idea.

[1] See Subspaces of $\mathbb{R}^n$ Associated with Matrices at https://ximera.osu.edu/linearalgebradzv3/LinearAlgebraInteractiveIntro/VSP-0040/main

Theorem 90 (Orthogonal Decomposition Theorem)**.** Let W be a subspace of $\mathbb{R}^n$ and let $\mathbf{x} \in \mathbb{R}^n$. Then there exist unique vectors $\mathbf{w} \in W$ and $\mathbf{w}^\perp \in W^\perp$ such that $\mathbf{x} = \mathbf{w} + \mathbf{w}^\perp$.

Proof This is an example of an "existence and uniqueness" theorem, so there are two things to prove. If we have an orthogonal basis $\{\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_m\}$ for W, then it is easy to show that our orthogonal decomposition exists for $\mathbf{x}$. We let $\mathbf{w} = \text{proj}_W(\mathbf{x})$, which is clearly in W, and we let $\mathbf{w}^\perp = \mathbf{x} - \mathbf{w}$, and we have $\mathbf{w} + \mathbf{w}^\perp = \mathbf{w} + (\mathbf{x} - \mathbf{w}) = \mathbf{x}$, so we need to see that $\mathbf{w}^\perp \in W^\perp$.

By Theorem **?? ??**, it suffices to show that $\mathbf{w}^\perp$ is orthogonal to each of the basis vectors $\mathbf{f}_i$, $i = 1, \ldots, m$. We compute for $i = 1, \ldots, m$

$$\begin{aligned}\mathbf{f}_i \cdot \mathbf{w}^\perp &= \mathbf{f}_i \cdot (\mathbf{x} - \mathbf{w}) \\ &= \mathbf{f}_i \cdot \mathbf{x} - \mathbf{f}_i \cdot \left(\frac{\mathbf{x} \cdot \mathbf{f}_1}{\|\mathbf{f}_1\|^2}\mathbf{f}_1 + \frac{\mathbf{x} \cdot \mathbf{f}_2}{\|\mathbf{f}_2\|^2}\mathbf{f}_2 + \cdots + \frac{\mathbf{x} \cdot \mathbf{f}_m}{\|\mathbf{f}_m\|^2}\mathbf{f}_m \right) \\ &= \mathbf{f}_i \cdot \mathbf{x} - \left(\frac{\mathbf{x} \cdot \mathbf{f}_1}{\|\mathbf{f}_1\|^2}\mathbf{f}_i \cdot \mathbf{f}_i \right) = \mathbf{f}_i \cdot \mathbf{x} - (\mathbf{x} \cdot \mathbf{f}_i) = 0.\end{aligned}$$

This proves that $\mathbf{w}^\perp \in W^\perp$.

The reason we need to prove this decomposition is unique is because we started with the orthogonal basis $\{\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_m\}$ for W, but what would happen if we chose a different orthogonal basis?

Suppose that $\{\mathbf{f}'_1, \mathbf{f}'_2, \ldots, \mathbf{f}'_m\}$ is another orthogonal basis of W, and let

$$\mathbf{w}' = \left(\frac{\mathbf{x} \cdot \mathbf{f}'_1}{\|\mathbf{f}'_1\|^2} \right) \mathbf{f}'_1 + \left(\frac{\mathbf{x} \cdot \mathbf{f}'_2}{\|\mathbf{f}'_2\|^2} \right) \mathbf{f}'_2 + \cdots + \left(\frac{\mathbf{x} \cdot \mathbf{f}'_m}{\|\mathbf{f}'_m\|^2} \right) \mathbf{f}'_m$$

As before, $\mathbf{w}' \in W$ and $\mathbf{x} - \mathbf{w}' \in W^\perp$, and we must show that $\mathbf{w}' = \mathbf{w}$. To see this, write the vector $\mathbf{w} - \mathbf{w}'$ as follows:

$$\mathbf{w} - \mathbf{w}' = (\mathbf{x} - \mathbf{w}') - (\mathbf{x} - \mathbf{w})$$

This vector is in W (because $\mathbf{w}$ and $\mathbf{w}'$ are in W) and it is in $W^\perp$ (because $\mathbf{x} - \mathbf{w}'$ and $\mathbf{x} - \mathbf{w}$ are in $W^\perp$), and so it must be the zero vector (it is orthogonal to itself!). This means $\mathbf{w}' = \mathbf{w}$ as desired. ■

Example 148. Let W be a subspace given by $W = \text{span}\left(\begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ 2 \end{bmatrix} \right)$, and let $\mathbf{x} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}$. Write $\mathbf{x}$ as the sum of a vector in W and a vector in $W^\perp$.

Explanation. Following the notation of Theorem **??**, we will write $\mathbf{x} = \mathbf{w}+\mathbf{w}^\perp$, where $\mathbf{w} = \text{proj}_W(\mathbf{x})$ and $\mathbf{w}^\perp = \mathbf{x} - \mathbf{w}$. Let $\mathbf{f}_1 = \begin{bmatrix}1\\0\\1\\0\end{bmatrix}$ and let $\mathbf{f}_2 = \begin{bmatrix}0\\1\\0\\2\end{bmatrix}$. We observe that we have the good fortune that $\mathbf{f}_1, \mathbf{f}_2$ is an orthogonal basis for W (otherwise, our first step would be to use the Gram-Schmidt procedure to create an orthogonal basis for W). We compute:

$$\mathbf{w} = \text{proj}_W(\mathbf{x}) = \text{proj}_{\mathbf{f}_1}(\mathbf{x}) + \text{proj}_{\mathbf{f}_2}(\mathbf{x}) = \frac{4}{2}\begin{bmatrix}1\\0\\1\\0\end{bmatrix} + \frac{10}{5}\begin{bmatrix}0\\1\\0\\2\end{bmatrix} = \begin{bmatrix}2\\2\\2\\4\end{bmatrix},$$

and then $\mathbf{w}^\perp = \begin{bmatrix}1\\2\\3\\4\end{bmatrix} - \begin{bmatrix}2\\2\\2\\4\end{bmatrix} = \begin{bmatrix}-1\\0\\1\\0\end{bmatrix}$.

This gives us

$$\mathbf{x} = \mathbf{w} + \mathbf{w}^\perp = \begin{bmatrix}2\\2\\2\\4\end{bmatrix} + \begin{bmatrix}-1\\0\\1\\0\end{bmatrix}$$

The final theorem of this section shows that projection onto a subspace of $\mathbb{R}^n$ is actually a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^n$.

Theorem 91. Let W be a fixed subspace of $\mathbb{R}^n$. If we define $T : \mathbb{R}^n \to \mathbb{R}^n$ by

$$T(\mathbf{x}) = \text{proj}_W(\mathbf{x}) \quad \text{for all } \mathbf{x} \text{ in } \mathbb{R}^n$$

1. T is a linear transformation. (See Introduction to Linear Transformations.)
2. $\text{im}(T)$ is W and $\ker(T)$ is $W^\perp$. (See Image and Kernel of a Linear Transformation.)
3. $\dim(W) + \dim(W^\perp) = n$.

Proof If $W = \{\mathbf{0}\}$, then $W^\perp = \mathbb{R}^n$, and so $T(\mathbf{x}) = \mathbf{0}$ for all $\mathbf{x}$. Thus $T = 0$ is the zero (linear) operator, so **??**, **??**, and **??** hold. Hence assume that $W \neq \{\mathbf{0}\}$.

1. If $\{\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_m\}$ is an orthonormal basis of W, then

$$T(\mathbf{x}) = (\mathbf{x} \cdot \mathbf{q}_1)\mathbf{q}_1 + (\mathbf{x} \cdot \mathbf{q}_2)\mathbf{q}_2 + \cdots + (\mathbf{x} \cdot \mathbf{q}_m)\mathbf{q}_m \quad \text{for all } \mathbf{x} \in \mathbb{R}^n \tag{94}$$

   by the definition of the projection. Thus T is a linear transformation because

$$(\mathbf{x} + \mathbf{y}) \cdot \mathbf{q}_i = \mathbf{x} \cdot \mathbf{q}_i + \mathbf{y} \cdot \mathbf{q}_i \quad \text{and} \quad (r\mathbf{x}) \cdot \mathbf{q}_i = r(\mathbf{x} \cdot \mathbf{q}_i) \quad \text{for each } i.$$

2. We have $\text{im}(T)$ is a subset of W by (**??**) because each $\mathbf{q}_i$ is in W. But if $\mathbf{x}$ is in W, then $\mathbf{x} = T(\mathbf{x})$ by (**??**) and Theorem **??** applied to the space W. This shows that W is a subset of $\text{im}(T)$, so $\text{im}(T)$ is W.

   Now suppose that $\mathbf{x}$ is in $W^\perp$. Then $\mathbf{x} \cdot \mathbf{q}_i = 0$ for each i (again because each $\mathbf{q}_i$ is in W) so $\mathbf{x}$ is in $\ker(T)$ by (**??**). Hence $W^\perp$ is in $\ker(T)$. On the other hand, Theorem **??** shows that $\mathbf{x} - T(\mathbf{x})$ is in $W^\perp$ for all $\mathbf{x}$ in $\mathbb{R}^n$, and it follows that $\ker(T)$ is in $W^\perp$. Hence $\ker(T)$ is $W^\perp$, proving **??**.

3. This follows from **??**, **??**, and the Rank-Nullity theorem (Theorem **??**).

■

Practice Problems

Problem 407 *Solve the linear system in Example* **??** *and use your result to find a basis for* $W^\perp$ *if* $W = \text{span}\left(\begin{bmatrix} 1 \\ -1 \\ 2 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ -2 \\ 3 \end{bmatrix}\right)$ *in* $\mathbb{R}^4$.

Problem 408 *In this problem we return to the GeoGebra interactive in Exploration* **??***, and we consider the case where the matrix A has rank 1 (which Exploration* **??** *could not handle). This time, the sliders define row 1 of matrix A, and row 2 will be 2 times row 1. Follow the prompts in the interactive to visualize row*(A) *and null*(A)*. What relationships do you observe between row*(A) *and null*(A)*?*

Access GeoGebra interactives through the online version of this text at

https://ximera.osu.edu/linearalgebradzv3/LinearAlgebraInteractiveIntro.

Problem 409 *In this problem you are asked to finish Example* **??**. *More specifically, for the matrix*

$$A = \begin{bmatrix} 2 & -1 & 1 & -4 & 1 \\ 1 & 0 & 3 & 3 & 0 \\ -2 & 1 & -1 & 5 & 2 \\ 4 & -1 & 7 & 2 & 1 \end{bmatrix},$$

show that $null(A^T) = (col(A))^{\perp}$. *It may be helpful to consult Subspaces of* $\mathbb{R}^n$ *Associated with Matrices, where we found a basis for the column space of this matrix.*

Problems **??**-**??**

In each case, write $\mathbf{x}$ as $\mathbf{x} = \mathbf{w} + \mathbf{w}^{\perp}$, where $\mathbf{w} = \text{proj}_W(\mathbf{x})$ and $\mathbf{w}^{\perp} = \mathbf{x} - \mathbf{w}$.

Problem 410 $\mathbf{x} = \begin{bmatrix} 2 \\ 1 \\ 6 \end{bmatrix}$, $W = span\left(\begin{bmatrix} 3 \\ -1 \\ 2 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \\ -3 \end{bmatrix}\right)$

Problem 411 $\mathbf{x} = \begin{bmatrix} 2 \\ 0 \\ 1 \\ 6 \end{bmatrix}$, $W = span\left(\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ -1 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ 1 \\ -1 \end{bmatrix}\right)$

Problem 412 $\mathbf{x} = \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix}$, $W = span\left(\begin{bmatrix} 1 \\ -1 \\ 2 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \\ 1 \\ 1 \end{bmatrix}\right)$

Problem 413 *Let* $W = span(\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_k)$, $\mathbf{w}_i \in \mathbb{R}^n$, *and let A be the* $k \times n$ *matrix with the* $\mathbf{w}_i$ *as rows.*

1. *Show that* $W^{\perp} = \{\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^n, A\mathbf{x}^T = \mathbf{0}\}$.
2. *Use part (a) to find* $W^{\perp}$ *if*

$$W = span\left(\begin{bmatrix} 1 \\ -1 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ -1 \\ 1 \end{bmatrix}\right).$$

Problem 414 *1. Prove part* **??** *of Theorem* **??**.

2. Prove part **??** *of Theorem* **??**.

Problem 415 *Let W be a subspace of $\mathbb{R}^n$. If $\mathbf{x}$ in $\mathbb{R}^n$ can be written in any way at all as $\mathbf{x} = \mathbf{p}+\mathbf{q}$ with $\mathbf{p}$ in W and $\mathbf{q}$ in $W^\perp$, show that necessarily $\mathbf{p} = proj_W(\mathbf{x})$.*

Problem 416 *Let W be a subspace of $\mathbb{R}^n$ and let $\mathbf{x}$ be a vector in $\mathbb{R}^n$. Using Practice Problem* **??***, or otherwise, show that $\mathbf{x}$ is in W if and only if $\mathbf{x} = proj_W(\mathbf{x})$.*

Problem 417 *If W is a subspace of $\mathbb{R}^n$, show that $proj_W(\mathbf{x}) = \mathbf{x}$ for all $\mathbf{x}$ in W.*

Problem 418 *If W is a subspace of $\mathbb{R}^n$, show that $\mathbf{x} = proj_W(\mathbf{x}) + proj_{W^\perp}(\mathbf{x})$ for all $\mathbf{x}$ in $\mathbb{R}^n$.*

Problem 419 *If $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is an orthogonal basis of $\mathbb{R}^n$ and $W = span(\mathbf{v}_1, \ldots, \mathbf{v}_m)$, $m < n$, show that $W^\perp = span(\mathbf{v}_{m+1}, \ldots, \mathbf{v}_n)$.*

Text Source

This section was adapted from the second part of Section 8.1 of Keith Nicholson's *Linear Algebra with Applications*. (CC-BY-NC-SA)

W. Keith Nicholson, *Linear Algebra with Applications*, Lyryx 2018, Open Edition, p. 415–423.

Example **??** was adapted from Example 4.148 of Ken Kuttler's *A First Course in Linear Algebra*. (CC-BY)

Ken Kuttler, *A First Course in Linear Algebra*, Lyryx 2017, Open Edition, p. 249.

65 Orthogonal Matrices and Symmetric Matrices

Recall that an $n \times n$ matrix A is diagonalizable if and only if it has n linearly independent eigenvectors. (see Diagonalizable Matrices and Multiplicity) Moreover, the matrix P with these eigenvectors as columns is a diagonalizing matrix for A, that is

$$P^{-1}AP \text{ is diagonal.}$$

As we have seen, the nice bases of $\mathbb{R}^n$ are the orthogonal ones, so a natural question is: which $n \times n$ matrices have n orthogonal eigenvectors, so that columns of P form an orthogonal basis for $\mathbb{R}^n$? These turn out to be precisely the *symmetric matrices* (matrices for which $A = A^T$), and this is the main result of this section.

Orthogonal Matrices

A collection of non-zero, pairwise orthogonal vectors in $\mathbb{R}^n$ is called an *orthogonal* set of vectors. An orthogonal set of vectors is called *orthonormal* if $\|\mathbf{q}\| = 1$ for each vector $\mathbf{q}$ in the set. A set of orthogonal vectors $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ can be "*normalized*", i.e. converted into an orthonormal set $\left\{ \frac{1}{\|\mathbf{v}_1\|}\mathbf{v}_1, \frac{1}{\|\mathbf{v}_2\|}\mathbf{v}_2, \ldots, \frac{1}{\|\mathbf{v}_k\|}\mathbf{v}_k \right\}$. In particular, if a matrix A has n orthogonal eigenvectors, they can (by normalizing) be taken to be orthonormal. The corresponding diagonalizing matrix (we will use Q instead of P) has orthonormal columns, and such matrices are very easy to invert.

Theorem 92. The following conditions are equivalent for an $n \times n$ matrix Q.

1. Q is invertible and $Q^{-1} = Q^T$.
2. The rows of Q are orthonormal.
3. The columns of Q are orthonormal.

Proof First note that condition **??** is equivalent to $Q^TQ = I$. Let $\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_n$ denote the columns of Q. Then $\mathbf{q}_i^T$ is the ith row of Q^T, so the (i, j)-entry of Q^TQ is $\mathbf{q}_i \cdot \mathbf{q}_j$. Thus $Q^TQ = I$ means that $\mathbf{q}_i \cdot \mathbf{q}_j = 0$ if $i \neq j$ and $\mathbf{q}_i \cdot \mathbf{q}_j = 1$ if $i = j$. Hence condition **??** is equivalent to **??**. The proof of the equivalence of **??** and **??** is similar. ■

Definition 83 (Orthogonal Matrices)**.** An $n \times n$ matrix Q is called an *orthogonal matrix* if it satisfies one (and hence all) of the conditions in Theorem **??**.

Example 149. The rotation matrix $\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$ is orthogonal for any angle θ.

Explanation. See Practice Problem **??**.

Remark 17. In view of **??** and **??** of Theorem **??**, *orthonormal matrix* might be a better name. But *orthogonal matrix* is standard.

It is not enough that the rows of a matrix *A* are merely orthogonal for *A* to be an orthogonal matrix. Here is an example.

Exploration 420 *Let* $A = \begin{bmatrix} 2 & 1 & 1 \\ -1 & 1 & 1 \\ 0 & -1 & 1 \end{bmatrix}$

1. *Check that matrix A has rows that are orthogonal.*
2. *Check that matrix A has columns that are NOT orthogonal.*
3. *Check that matrix A has rows that are NOT orthonormal.*
4. *Create a matrix Q by normalizing each of the rows of A.*
5. *Check that Q is an orthogonal matrix.*

Click the arrow to see the answer. You should get $Q = \begin{bmatrix} \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\ \frac{-1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ 0 & \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$*, and one can check that this is orthogonal in a number of ways.*

This exploration can certainly be done by hand (although it takes some time), but it also makes for a very nice Octave exercise.

```
%Exploration from Section 9.4 Orthogonal Matrices and Symmetric Matrices
A=[2 1 1; -1 1 1; 0 -1 1]
%Check that matrix A has rows that are orthogonal.
A(1,:)*transpose(A(2,:))
A(2,:)*transpose(A(3,:))
A(1,:)*transpose(A(3,:))
```

```
%Check that matrix A has columns that are NOT orthogonal.
transpose(A(:,1))*A(:,2)
%(This is 1 of 3 calculations to do.)
%Check that matrix A in the Octave window has rows that are NOT orthonormal.
%(See the results from the first question.)
%Create a matrix Q by normalizing each of the rows of A.
q1=A(1,:)/norm(A(1,:));
q2=A(2,:)/norm(A(2,:));
q3=A(3,:)/norm(A(3,:));
Q = [q1;q2;q3]
%Check that Q is an orthogonal matrix.
Q*transpose(Q)
%(You may get numbers close to zero in some places you expect
%to get zero due to rounding error)
```

Link to code.

For more information about using Octave, please see Octave for Chapter 9.

We studied the idea of closure when we studied Subspaces of $\mathbb{R}^n$. The next theorem tells us that orthogonal matrices are closed under matrix multiplication.

Theorem 93. 1. If P and Q are orthogonal matrices, then PQ is also orthogonal. (We say that the set of orthogonal matices is closed under matrix multiplication.)

2. If P is an orthogonal matrix, then so is $P^{-1} = P^T$.

Proof For **??**, P and Q are invertible, so PQ is also invertible and

$$(PQ)^{-1} = Q^{-1}P^{-1} = Q^TP^T = (PQ)^T$$

Hence PQ is orthogonal. For **??**,

$$(P^{-1})^{-1} = P = (P^T)^T = (P^{-1})^T$$

shows that P^{-1} is orthogonal. ■

Symmetric Matrices

We now shift our focus from orthogonal matrices to another important class of $n \times n$ matrices called *symmetric* matrices. A *symmetric matrix* is a matrix which is equal to its transpose. We saw a few examples of such matrices in Transpose of a Matrix.

When we began our study of eigenvalues and eigenvectors, we saw numerous examples of matrices with entries that were real numbers with eigenvalues that were complex numbers. It can be shown that symmetric matrices only have real eigenvalues. We also learned that some matrices are diagonalizable while other matrices are not. It turns out that every symmetric matrix is diagonalizable. In fact, we can say more, but first we need the following definition.

Definition 84. An $n \times n$ matrix A is said to be *orthogonally diagonalizable* if an orthogonal matrix Q can be found such that $Q^{-1}AQ = Q^TAQ$ is diagonal.

We have learned earlier that when we diagonalize a matrix A, we write $P^{-1}AP = D$ for some matrix P where D is diagonal, and the diagonal entries are the eigenvalues of A. We have also learned that the columns of the matrix P are the corresponding eigenvectors of A. So when a matrix is orthogonally diagonalizable, we are able to accomplish the diagonalization using a matrix Q consisting of n eigenvectors that form an orthonormal basis for $\mathbb{R}^n$. The following remarkable theorem shows that the matrices that have this property are precisely the symmetric matrices.

Theorem 94 (Real Spectral Theorem)**.** Let A be an $n \times n$ matrix. Then A is symmetric if and only if A is orthogonally diagonalizable.

Proof If A is orthogonally diagonalizable, then it is an easy exercise to prove that it is symmetric. You are asked to do this in Practice Problem **??**.

To prove the "only if" part of this theorem, we assume A is symmetric, and we need to show it is orthogonally diagonalizable. We proceed by induction on n, the size of the symmetric matrix. If $n = 1$, A is already diagonal. If $n > 1$, assume that we know the "only if" statement holds for $(n-1) \times (n-1)$ symmetric matrices. Let λ_1 be an eigenvalue of A, and let $A\mathbf{x}_1 = \lambda_1\mathbf{x}_1$, where $\|\mathbf{x}_1\| = 1$. Next, set $\mathbf{q}_1 = \mathbf{x}_1$, and use the Gram-Schmidt algorithm to find an orthonormal basis $\{\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_n\}$ for $\mathbb{R}^n$. Let $Q_1 = \begin{bmatrix} | & | & & | \\ \mathbf{q}_1 & \mathbf{q}_2 & \cdots & \mathbf{q}_n \\ | & | & & | \end{bmatrix}$, so that Q_1 is an orthogonal matrix. We have

$$Q_1^TAQ_1 = \begin{bmatrix} -- & \mathbf{q}_1^T & -- \\ -- & \mathbf{q}_2^T & -- \\ & \vdots & \\ -- & \mathbf{q}_n^T & -- \end{bmatrix} A \begin{bmatrix} | & | & & | \\ \mathbf{q}_1 & \mathbf{q}_2 & \cdots & \mathbf{q}_n \\ | & | & & | \end{bmatrix}$$

$$= \begin{bmatrix} -- & \mathbf{q}_1^T & -- \\ -- & \mathbf{q}_2^T & -- \\ & \vdots & \\ -- & \mathbf{q}_n^T & -- \end{bmatrix} \begin{bmatrix} | & | & & | \\ A\mathbf{q}_1 & A\mathbf{q}_2 & \cdots & A\mathbf{q}_n \\ | & | & & | \end{bmatrix}$$

$$= \begin{bmatrix} \lambda_1 & B \\ \mathbf{0} & A_1 \end{bmatrix},$$

where the block B has dimensions $1 \times (n-1)$, and the block under λ_1 is a $(n-1) \times 1$ zero matrix, because of the orthogonality of the basis vectors.

Next, using the fact that A is symmetric, we notice that

$$(Q_1^T A Q_1)^T = Q_1^T A^T (Q_1^T)^T = Q_1^T A Q_1,$$

so $Q_1^T A Q_1$ is symmetric. It follows that B is also a zero matrix and that A_1 is symmetric. Since A_1 is an $(n-1) \times (n-1)$ symmetric matrix, we may apply the inductive hypothesis, so there exists an $(n-1) \times (n-1)$ orthogonal matrix Q such that $Q^T A_1 Q = D_1$ is diagonal. We observe that $Q_2 = \begin{bmatrix} 1 & 0 \\ 0 & Q \end{bmatrix}$ is orthogonal, and we compute:

$$\begin{aligned} (Q_1 Q_2)^T A (Q_1 Q_2) &= Q_2^T (Q_1^T A Q_1) Q_2 \\ &= \begin{bmatrix} 1 & 0 \\ 0 & Q^T \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 \\ 0 & A_1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & Q \end{bmatrix} \\ &= \begin{bmatrix} \lambda_1 & 0 \\ 0 & D_1 \end{bmatrix} \end{aligned}$$

is diagonal. Because $Q_1 Q_2$ is orthogonal by Theorem **????**, this completes the proof. ■

Because the eigenvalues of a real symmetric matrix are real, Theorem **??** is called the *Real Spectral Theorem*, and the set of distinct eigenvalues is called the *spectrum* of the matrix. A similar result holds for matrices with complex entries (Theorem **??**).

Example 150. Find an orthogonal matrix Q such that $Q^{-1} A Q$ is diagonal, where $A = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 2 \\ -1 & 2 & 5 \end{bmatrix}$.

Explanation. The characteristic equation of A is:

$$\det \begin{bmatrix} 1-\lambda & 0 & -1 \\ 0 & 1-\lambda & 2 \\ -1 & 2 & 5-\lambda \end{bmatrix} = -\lambda(\lambda-1)(\lambda-6) = 0$$

Thus the eigenvalues are $\lambda = 0$, 1, and 6, and corresponding eigenvectors are

$$\mathbf{x}_1 = \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix} \quad \mathbf{x}_2 = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} \quad \mathbf{x}_3 = \begin{bmatrix} -1 \\ 2 \\ 5 \end{bmatrix}$$

respectively. Moreover, by what at first appears to be remarkably good luck, these eigenvectors are *orthogonal*. We have $\|\mathbf{x}_1\|^2 = 6$, $\|\mathbf{x}_2\|^2 = 5$, and $\|\mathbf{x}_3\|^2 = 30$, so

$$P = \begin{bmatrix} | & | & | \\ \frac{1}{\sqrt{6}}\mathbf{x}_1 & \frac{1}{\sqrt{5}}\mathbf{x}_2 & \frac{1}{\sqrt{30}}\mathbf{x}_3 \\ | & | & | \end{bmatrix} = \frac{1}{\sqrt{30}} \begin{bmatrix} \sqrt{5} & 2\sqrt{6} & -1 \\ -2\sqrt{5} & \sqrt{6} & 2 \\ \sqrt{5} & 0 & 5 \end{bmatrix}$$

is an orthogonal matrix. Thus $P^{-1} = P^T$ and

$$P^T A P = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 6 \end{bmatrix}.$$

Actually, the fact that the eigenvectors in Example **??** are orthogonal is no coincidence. These vectors certainly must be linearly independent (they correspond to distinct eigenvalues). We will see that the fact that the matrix is *symmetric* implies that the eigenvectors are orthogonal. To prove this we need the following useful fact about symmetric matrices.

Theorem 95. If A is an $n \times n$ symmetric matrix, then

$$(A\mathbf{x}) \cdot \mathbf{y} = \mathbf{x} \cdot (A\mathbf{y})$$

for all columns $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^n$.

Remark 18. The converse also holds (see Practice Problem **??**).

Proof Recall that $\mathbf{x} \cdot \mathbf{y} = \mathbf{x}^T\mathbf{y}$ for all columns $\mathbf{x}$ and $\mathbf{y}$. Because $A^T = A$, we get

$$(A\mathbf{x}) \cdot \mathbf{y} = (A\mathbf{x})^T\mathbf{y} = \mathbf{x}^T A^T\mathbf{y} = \mathbf{x}^T A\mathbf{y} = \mathbf{x} \cdot (A\mathbf{y})$$

■

Theorem 96. If A is a symmetric matrix, then eigenvectors of A corresponding to distinct eigenvalues are orthogonal.

Proof Let $A\mathbf{x} = \lambda\mathbf{x}$ and $A\mathbf{y} = \mu\mathbf{y}$, where $\lambda \neq \mu$. We compute

$$\lambda(\mathbf{x} \cdot \mathbf{y}) = (\lambda\mathbf{x}) \cdot \mathbf{y} = (A\mathbf{x}) \cdot \mathbf{y} = \mathbf{x} \cdot (A\mathbf{y}) = \mathbf{x} \cdot (\mu\mathbf{y}) = \mu(\mathbf{x} \cdot \mathbf{y})$$

Hence $(\lambda - \mu)(\mathbf{x} \cdot \mathbf{y}) = 0$, and so $\mathbf{x} \cdot \mathbf{y} = 0$ because $\lambda \neq \mu$. ■

Now the procedure for diagonalizing a symmetric $n \times n$ matrix is clear. Find the distinct eigenvalues and find orthonormal bases for each eigenspace (the Gram-Schmidt algorithm may be needed when there is a repeated eigenvalue). Then the set of all these basis vectors is orthonormal (by Theorem **??**) and contains n vectors. Here is an example.

Example 151. Orthogonally diagonalize the symmetric matrix $A = \begin{bmatrix} 8 & -2 & 2 \\ -2 & 5 & 4 \\ 2 & 4 & 5 \end{bmatrix}$.

Explanation. The characteristic equation is

$$\det \begin{bmatrix} 8-\lambda & -2 & 2 \\ -2 & 5-\lambda & 4 \\ 2 & 4 & 5-\lambda \end{bmatrix} = -\lambda(\lambda - 9)^2 = 0$$

Hence the distinct eigenvalues are 0 and 9 are of algebraic multiplicity 1 and 2, respectively. The geometric multiplicities must be the same, for A is diagonalizable, being symmetric. It follows that $\dim(\mathcal{S}_0) = 1$ and $\dim(\mathcal{S}_9) = 2$. Gaussian elimination gives

$$\mathcal{S}_0(A) = \text{span}\{\mathbf{x}_1\}, \quad \mathbf{x}_1 = \begin{bmatrix} 1 \\ 2 \\ -2 \end{bmatrix}, \quad \text{and} \quad \mathcal{S}_9(A) = \text{span}\left\{ \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix} \right\}$$

The eigenvectors in $\mathcal{S}_9$ are both orthogonal to $\mathbf{x}_1$ as Theorem **??** guarantees, but not to each other. However, an orthogonal basis can be found using projections. (See Gram-Schmidt Orthogonalization)

$$\{\mathbf{f}_2, \mathbf{f}_3\} \text{ of } \mathcal{S}_9(A) \quad \text{where} \quad \mathbf{f}_2 = \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix} \text{ and } \mathbf{f}_3 = \begin{bmatrix} 2 \\ 4 \\ 5 \end{bmatrix}$$

Normalizing gives orthonormal vectors $\{\frac{1}{3}\mathbf{x}_1, \frac{1}{\sqrt{5}}\mathbf{f}_2, \frac{1}{3\sqrt{5}}\mathbf{f}_3\}$, so

$$Q = \begin{bmatrix} | & | & | \\ \frac{1}{3}\mathbf{x}_1 & \frac{1}{\sqrt{5}}\mathbf{f}_2 & \frac{1}{3\sqrt{5}}\mathbf{f}_3 \\ | & | & | \end{bmatrix} = \frac{1}{3\sqrt{5}} \begin{bmatrix} \sqrt{5} & -6 & 2 \\ 2\sqrt{5} & 3 & 4 \\ -2\sqrt{5} & 0 & 5 \end{bmatrix}$$

is an orthogonal matrix such that $Q^{-1}AQ$ is diagonal.

It is worth noting that other, more convenient, diagonalizing matrices Q exist. For example, $\mathbf{y}_2 = \begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix}$ and $\mathbf{y}_3 = \begin{bmatrix} -2 \\ 2 \\ 1 \end{bmatrix}$ lie in $\mathcal{S}_9(A)$ and they are orthogonal. Moreover, they both have norm 3 (as does $\mathbf{x}_1$), so

$$\hat{Q} = \begin{bmatrix} | & | & | \\ \frac{1}{3}\mathbf{x}_1 & \frac{1}{3}\mathbf{y}_2 & \frac{1}{3}\mathbf{y}_3 \\ | & | & | \end{bmatrix} = \frac{1}{3}\begin{bmatrix} 1 & 2 & -2 \\ 2 & 1 & 2 \\ -2 & 2 & 1 \end{bmatrix}$$

is a nicer orthogonal matrix with the property that $\hat{Q}^{-1}A\hat{Q}$ is diagonal.

Theorem 97. Let A be an $n \times n$ matrix. A has an orthonormal set of n eigenvectors if and only if A is orthogonally diagonalizable.

Proof Let $\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_n$ be orthonormal eigenvectors of A with corresponding eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$. We must show A is orthogonally diagonalizable. Let $Q = \begin{bmatrix} | & | & & | \\ \mathbf{q}_1 & \mathbf{q}_2 & \cdots & \mathbf{q}_n \\ | & | & & | \end{bmatrix}$ so that Q is orthogonal. We have

$$AQ = \begin{bmatrix} | & | & & | \\ A\mathbf{q}_1 & A\mathbf{q}_2 & \cdots & A\mathbf{q}_n \\ | & | & & | \end{bmatrix} = \begin{bmatrix} | & | & & | \\ \lambda_1\mathbf{q}_1 & \lambda_2\mathbf{q}_2 & \ldots & \lambda_n\mathbf{q}_n \\ | & | & & | \end{bmatrix} = QD,$$

where D is the diagonal matrix with diagonal entries $\lambda_1, \lambda_2, \ldots, \lambda_n$. But then $Q^TAQ = D$, proving this half of the theorem.

For the converse, if A is orthogonally diagonalizable, then by Theorem **??** it is symmetric. But then Theorem **??** tells us that eigenvectors corresponding to distinct eigenvalues are orthogonal. Because A is (orthogonally) diagonalizable, we know the geometric multiplicity of each eigenvalue is equal to its algebraic multiplicity. This implies that we can use Gram-Schmidt on each eigenspace of dimension > 1 to get a full set of n orthogonal eigenvectors. ■

If we are willing to replace "diagonal" by "upper triangular" in the real spectral theorem, we can weaken the requirement that A is symmetric to insisting only that A has real eigenvalues.

Theorem 98 (Schur Triangularization Theorem)**.** If A is an $n \times n$ matrix with n real eigenvalues, an orthogonal matrix Q exists such that Q^TAQ is upper triangular.

Remark 19. There is also a lower triangular version of this theorem.

Proof See Practice Problem **??** ■

The eigenvalues of an upper triangular matrix are displayed along the main diagonal. Because A and Q^TAQ have the same determinant and trace whenever Q is orthogonal (for they are similar matrices), Theorem **??** gives:

Corollary 11. If A is an $n \times n$ matrix with real eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ (possibly not all distinct), then $\det A = \lambda_1\lambda_2 \ldots \lambda_n$ and $\text{tr}A = \lambda_1 + \lambda_2 + \cdots + \lambda_n$.

Remark 20. This corollary remains true even if the eigenvalues are not real.

Practice Problems

Problem 421 *Normalize the columns to make the following matrix orthogonal.* $A = \begin{bmatrix} 1 & -4 \\ 2 & 2 \end{bmatrix}$

$$\frac{1}{\sqrt{\ldots\ldots}} \begin{bmatrix} \ldots\ldots & \ldots\ldots \\ \ldots\ldots & \ldots\ldots \end{bmatrix}$$

Problem 422 *Normalize the columns to make the following matrix orthogonal.* $A = \begin{bmatrix} -1 & 2 & 2 \\ 2 & -1 & 2 \\ 2 & 2 & -1 \end{bmatrix}$

Problem 423 *For each matrix A, find an orthogonal matrix Q such that* $Q^{-1}AQ$ *is diagonal.*

Question 423.1

$$A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

$$Q = \frac{1}{\sqrt{\ldots\ldots}} \begin{bmatrix} 1 & -1 \\ \ldots\ldots & \ldots\ldots \end{bmatrix}$$

Question 423.2

$$A = \begin{bmatrix} 3 & 0 & 7 \\ 0 & 5 & 0 \\ 7 & 0 & 3 \end{bmatrix}$$

Problem 424 *Suppose A is orthogonally diagonalizable. Prove that A is symmetric. (This is the easy direction of the "if and only if" in Theorem* **??**.*)*

Problem 425 *Prove the converse of Theorem* **??**:

If $(A\mathbf{x}) \cdot \mathbf{y} = \mathbf{x} \cdot (A\mathbf{y})$ for all n-columns $\mathbf{x}$ and $\mathbf{y}$, then A is symmetric.

Problem 426 *Let Q be an orthogonal matrix.*

1. *Show that* $\det Q = 1$ *or* $\det Q = -1$.
2. *Give* 2×2 *examples of Q such that* $\det Q = 1$ *and* $\det Q = -1$.

Text Source

This section was adapted from Section 8.2 of Keith Nicholson's *Linear Algebra with Applications*. (CC-BY-NC-SA)

W. Keith Nicholson, *Linear Algebra with Applications*, Lyryx 2018, Open Edition, p. 424

66 Positive Definite Matrices

All the eigenvalues of any symmetric matrix are real (proved later in **??**); this section is about the case in which those eigenvalues are positive. These matrices, which arise whenever optimization (maximum and minimum) problems are encountered, have countless applications throughout science and engineering. They also arise in statistics (for example, in factor analysis used in the social sciences) and in geometry (quadratic forms, for instance). We will encounter them again in Inner Product Spaces when describing all inner products in $\mathbb{R}^n$.

Definition 85. A square matrix is called *positive definite* if it is symmetric and all its eigenvalues λ are positive. We write $\lambda > 0$ when eigenvalues are real and positive.

Because these matrices are symmetric, the Real Spectral Theorem (**??**) plays a central role in the theory.

Theorem 99. If A is positive definite, then it is invertible and $\det A > 0$.

Proof If A is $n \times n$ and the eigenvalues are $\lambda_1, \lambda_2, \ldots, \lambda_n$, then $\det A = \lambda_1\lambda_2 \cdots \lambda_n > 0$ by the Real Spectral Theorem (or, more specifically, **??**). ■

If $\mathbf{x}$ is a column in $\mathbb{R}^n$ and A is any real $n \times n$ matrix, we view the 1×1 matrix $\mathbf{x}^T A\mathbf{x}$ as a real number. With this convention, we have the following characterization of positive definite matrices.

Theorem 100. A symmetric matrix A is positive definite if and only if $\mathbf{x}^T A\mathbf{x} > 0$ for every column $\mathbf{x} \neq \mathbf{0}$ in $\mathbb{R}^n$.

Proof A is symmetric. Using the Real Spectral Theorem, let $Q^T AQ = D = \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$ where $Q^{-1} = Q^T$ and the λ_i are the eigenvalues of A. Given a column $\mathbf{x}$ in $\mathbb{R}^n$, write $\mathbf{y} = Q^T\mathbf{x} = \begin{bmatrix} y_1 & y_2 & \ldots & y_n \end{bmatrix}^T$. Then

$$\mathbf{x}^T A\mathbf{x} = \mathbf{x}^T(QDQ^T)\mathbf{x} = \mathbf{y}^T D\mathbf{y} = \lambda_1 y_1^2 + \lambda_2 y_2^2 + \cdots + \lambda_n y_n^2 \tag{95}$$

If A is positive definite and $\mathbf{x} \neq \mathbf{0}$, then $\mathbf{x}^T A\mathbf{x} > 0$ by (**??**) because some $y_j \neq 0$ and every $\lambda_i > 0$. Conversely, if $\mathbf{x}^T A\mathbf{x} > 0$ whenever $\mathbf{x} \neq \mathbf{0}$, let $\mathbf{x} = Q\mathbf{e}_j \neq \mathbf{0}$ where $\mathbf{e}_j$ is column j of I_n. Then $\mathbf{y} = \mathbf{e}_j$, so (**??**) reads $\lambda_j = \mathbf{x}^T A\mathbf{x} > 0$. ■

Note that Theorem **??** shows that the positive definite matrices are exactly the symmetric matrices A for which the quadratic form $q = \mathbf{x}^T A\mathbf{x}$ takes only positive values.

The next example of a positive definite matrix is one we will see again in SVD Decomposition

Example 152. If U is any invertible $n \times n$ matrix, show that $A = U^T U$ is positive definite.

Explanation. If $\mathbf{x}$ is in $\mathbb{R}^n$ and $\mathbf{x} \neq \mathbf{0}$, then

$$\mathbf{x}^T A\mathbf{x} = \mathbf{x}^T(U^T U)\mathbf{x} = (U\mathbf{x})^T(U\mathbf{x}) = \|U\mathbf{x}\|^2 > 0$$

because $U\mathbf{x} \neq \mathbf{0}$ (U is invertible). Hence Theorem **??** applies.

It is remarkable that the converse to Example **??** is also true. In fact every positive definite matrix A can be factored as $A = U^T U$ where U is an upper triangular matrix with positive elements on the main diagonal. However, before verifying this, we introduce another concept that is central to any discussion of positive definite matrices.

If A is any $n \times n$ matrix, let ${}^{(r)}A$ denote the $r \times r$ submatrix in the upper left corner of A; that is, ${}^{(r)}A$ is the matrix obtained from A by deleting the last $n - r$ rows and columns. The matrices ${}^{(1)}A, {}^{(2)}A, {}^{(3)}A, \ldots, {}^{(n)}A = A$ are called the *principal submatrices* of A.

Example 153. If $A = \begin{bmatrix} 10 & 5 & 2 \\ 5 & 3 & 2 \\ 2 & 2 & 3 \end{bmatrix}$ then ${}^{(1)}A = [10]$, ${}^{(2)}A = \begin{bmatrix} 10 & 5 \\ 5 & 3 \end{bmatrix}$ and ${}^{(3)}A = A$.

Lemma 9. If A is positive definite, so is each principal submatrix ${}^{(r)}A$ for $r = 1, 2, \ldots, n$.

Proof Write $A = \begin{bmatrix} {}^{(r)}A & P \\ Q & R \end{bmatrix}$ in block form. If $\mathbf{y} \neq \mathbf{0}$ in $\mathbb{R}^r$, write $\mathbf{x} = \begin{bmatrix} \mathbf{y} \\ \mathbf{0} \end{bmatrix}$ in $\mathbb{R}^n$.

Then $\mathbf{x} \neq \mathbf{0}$, so the fact that A is positive definite gives

$$0 < \mathbf{x}^T A\mathbf{x} = \begin{bmatrix} \mathbf{y}^T & \mathbf{0} \end{bmatrix} \begin{bmatrix} {}^{(r)}A & P \\ Q & R \end{bmatrix} \begin{bmatrix} \mathbf{y} \\ \mathbf{0} \end{bmatrix} = \mathbf{y}^T({}^{(r)}A)\mathbf{y}$$

This shows that ${}^{(r)}A$ is positive definite by Theorem **??**. (A similar argument shows that, if B is any matrix obtained from a positive definite matrix A by deleting certain rows and deleting the *same* columns, then B is also positive definite.) ■

If A is positive definite, Lemma **??** and Theorem **??** show that $\det({}^{(r)}A) > 0$ for every r. This proves part of the following theorem which contains the converse to Example **??**, and characterizes the positive definite matrices among the symmetric ones.

Theorem 101. The following conditions are equivalent for a symmetric $n \times n$ matrix A:

1. A is positive definite.

2. $\det({}^{(r)}A) > 0$ for each $r = 1, 2, \ldots, n$.

3. $A = U^T U$ where U is an upper triangular matrix with positive entries on the main diagonal.

Furthermore, the factorization in **??** is unique (called the *Cholesky factorization* of A).

Remark 21. Andre-Louis Cholesky (1875–1918), was a French mathematician who died in World War I. His factorization was published in 1924 by a fellow officer.

Proof First, **??** $\Rightarrow$ **??** by Example **??**, and **??** $\Rightarrow$ **??** by Lemma **??** and Theorem **??**.

?? $\Rightarrow$ **??**. Assume **??** and proceed by induction on n. If $n = 1$, then $A = [a]$ where $a > 0$ by **??**, so take $U = \left[\sqrt{a}\right]$. If $n > 1$, write $B = {}^{(n-1)}A$. Then B is symmetric and satisfies **??** so, by induction, we have $B = U^T U$ as in **??** where U is of size $(n-1) \times (n-1)$. Then, as A is symmetric, it has block form $A = \begin{bmatrix} B & \mathbf{p} \\ \mathbf{p}^T & b \end{bmatrix}$ where $\mathbf{p}$ is a column in $\mathbb{R}^{n-1}$ and b is in $\mathbb{R}$. If we write $\mathbf{x} = (U^T)^{-1}\mathbf{p}$ and $c = b - \mathbf{x}^T\mathbf{x}$, block multiplication gives

$$A = \begin{bmatrix} U^T U & \mathbf{p} \\ \mathbf{p}^T & b \end{bmatrix} = \begin{bmatrix} U^T & 0 \\ \mathbf{x}^T & 1 \end{bmatrix} \begin{bmatrix} U & \mathbf{x} \\ 0 & c \end{bmatrix}.$$

Taking determinants gives $\det A = \det(U^T)\det U \cdot c = c(\det U)^2$. Hence $c > 0$ because $\det A > 0$ by **??**, so the above factorization can be written

$$A = \begin{bmatrix} U^T & 0 \\ \mathbf{x}^T & \sqrt{c} \end{bmatrix} \begin{bmatrix} U & \mathbf{x} \\ 0 & \sqrt{c} \end{bmatrix}$$

Since U has positive diagonal entries, this proves **??**.

To prove uniqueness, suppose that $A = U^T U = U_1^T U_1$ are two Cholesky factorizations. Now write $D = UU_1^{-1} = (U^T)^{-1}U_1^T$. Then D is upper triangular, because $D = UU_1^{-1}$, and lower triangular, because $D = (U^T)^{-1}U_1^T$, and so it is a diagonal matrix. Thus $U = DU_1$ and $U_1 = DU$, so it suffices to show that $D = I$. But eliminating U_1 gives $U = D^2 U$, so $D^2 = I$ because U is invertible. Since the diagonal entries of D are positive (this is true of U and U_1), it follows that $D = I$. ■

The remarkable thing is that the matrix U in the Cholesky factorization is easy to obtain from A using row operations. The key is that Step 1 of the following algorithm is *possible* for any positive definite matrix A. A proof of the algorithm is given following Example **??**.

Theorem 102 (Algorithm for the Cholesky Factorization)**.** If A is a positive definite matrix, the Cholesky factorization $A = U^T U$ can be obtained as follows:

Step 1. Carry A to an upper triangular matrix U_1 with positive diagonal entries using row operations each of which adds a multiple of a row to a lower row.

Step 2. Obtain U from U_1 by dividing each row of U_1 by the square root of the diagonal entry in that row.

Example 154. Find the Cholesky factorization of $A = \begin{bmatrix} 10 & 5 & 2 \\ 5 & 3 & 2 \\ 2 & 2 & 3 \end{bmatrix}$.

Explanation. The matrix A is positive definite by Theorem **??** because $\det^{(1)}A = 10 > 0$, $\det^{(2)}A = 5 > 0$, and $\det^{(3)}A = \det A = 3 > 0$. Hence Step 1 of the algorithm is carried out as follows:

$$A = \begin{bmatrix} 10 & 5 & 2 \\ 5 & 3 & 2 \\ 2 & 2 & 3 \end{bmatrix} \to \begin{bmatrix} 10 & 5 & 2 \\ 0 & \frac{1}{2} & 1 \\ 0 & 1 & \frac{13}{5} \end{bmatrix} \to \begin{bmatrix} 10 & 5 & 2 \\ 0 & \frac{1}{2} & 1 \\ 0 & 0 & \frac{3}{5} \end{bmatrix} = U_1$$

Now carry out Step 2 on U_1 to obtain $U = \begin{bmatrix} \sqrt{10} & \frac{5}{\sqrt{10}} & \frac{2}{\sqrt{10}} \\ 0 & \frac{1}{\sqrt{2}} & \sqrt{2} \\ 0 & 0 & \frac{\sqrt{3}}{\sqrt{5}} \end{bmatrix}$.

The reader can verify that $U^T U = A$.

Proof of the Cholesky Algorithm If A is positive definite, let $A = U^T U$ be the Cholesky factorization, and let $D = \text{diag}(d_1, \ldots, d_n)$ be the common diagonal of U and U^T. Then $U^T D^{-1}$ is lower triangular with ones on the diagonal (call such matrices LT-1). Hence $L = (U^T D^{-1})^{-1}$ is also LT-1, and so $I_n \to L$ by a sequence of row operations each of which adds a multiple of a row to a lower row (modify columns right to left). But then $A \to LA$ by the same sequence of row operations. Since $LA = [D(U^T)^{-1}][U^T U] = DU$ is upper triangular with positive entries on the diagonal, this shows that Step 1 of the algorithm is possible.

Turning to Step 2, let $A \to U_1$ as in Step 1 so that $U_1 = L_1 A$ where L_1 is LT-1. Since A is symmetric, we get

$$L_1 U_1^T = L_1 (L_1 A)^T = L_1 A^T L_1^T = L_1 A L_1^T = U_1 L_1^T \tag{96}$$

Let $D_1 = \text{diag}(e_1, \ldots, e_n)$ denote the diagonal of U_1. Then (**??**) gives $L_1(U_1^T D_1^{-1}) = U_1 L_1^T D_1^{-1}$. This is both upper triangular (right side) and LT-1 (left side), and so must equal I_n. In particular, $U_1^T D_1^{-1} = L_1^{-1}$. Now let $D_2 = \text{diag}(\sqrt{e_1}, \ldots, \sqrt{e_n})$, so that $D_2^2 = D_1$. If we write $U = D_2^{-1} U_1$ we have

$$U^T U = (U_1^T D_2^{-1})(D_2^{-1} U_1) = U_1^T (D_2^2)^{-1} U_1 = (U_1^T D_1^{-1}) U_1 = (L_1^{-1}) U_1 = A$$

This proves Step 2 because $U = D_2^{-1}U_1$ is formed by dividing each row of U_1 by the square root of its diagonal entry (verify). ■

Practice Problems

Problem 427 *Find the Cholesky decomposition of each of the following matrices.*

1. $\begin{bmatrix} 4 & 3 \\ 3 & 5 \end{bmatrix}$

2. $\begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix}$

   Click the arrow to see the answer. $U = \dfrac{\sqrt{2}}{2}\begin{bmatrix} 2 & -1 \\ 0 & 1 \end{bmatrix}$

3. $\begin{bmatrix} 12 & 4 & 3 \\ 4 & 2 & -1 \\ 3 & -1 & 7 \end{bmatrix}$

4. $\begin{bmatrix} 20 & 4 & 5 \\ 4 & 2 & 3 \\ 5 & 3 & 5 \end{bmatrix}$

   Click the arrow to see the answer. $U = \dfrac{1}{30}\begin{bmatrix} 60\sqrt{5} & 12\sqrt{5} & 15\sqrt{5} \\ 0 & 6\sqrt{30} & 10\sqrt{30} \\ 0 & 0 & 5\sqrt{15} \end{bmatrix}$

Problem 428 *a. If A is positive definite, show that A^k is positive definite for all $k \geq 1$.*

b. Prove the converse to (a) when k is odd.

c. Find a symmetric matrix A such that A^2 is positive definite but A is not.

Problem 429 *Let* $A = \begin{bmatrix} 1 & a \\ a & b \end{bmatrix}$. *If $a^2 < b$, show that A is positive definite and find the Cholesky factorization.*

Problem 430 *If A and B are positive definite and $r > 0$, show that A+B and rA are both positive definite.*

Click the arrow to see the answer. If $\mathbf{x} \neq \mathbf{0}$, then $\mathbf{x}^T A\mathbf{x} > 0$ and $\mathbf{x}^T B\mathbf{x} > 0$. Hence $\mathbf{x}^T(A+B)\mathbf{x} = \mathbf{x}^T A\mathbf{x} + \mathbf{x}^T B\mathbf{x} > 0$ and $\mathbf{x}^T(rA)\mathbf{x} = r(\mathbf{x}^T A\mathbf{x}) > 0$, as $r > 0$.

Problem 431 *If A and B are positive definite, show that* $\begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix}$ *is positive definite.*

Problem 432 *If A is an $n \times n$ positive definite matrix and U is an $n \times m$ matrix of rank m, show that $U^T AU$ is positive definite.*

Problem 433 *If A is positive definite, show that each diagonal entry is positive.*

Problem 434 *Let A_0 be formed from A by deleting rows 2 and 4 and deleting columns 2 and 4. If A is positive definite, show that A_0 is positive definite.*

Problem 435 *If A is positive definite, show that $A = CC^T$ where C has orthogonal columns.*

Problem 436 *If A is positive definite, show that $A = C^2$ where C is positive definite.*

Click the arrow for answer. Let $P^T AP = D = diag(\lambda_1, \ldots, \lambda_n)$ where $P^T = P$. Since A is positive definite, each eigenvalue $\lambda_i > 0$. If $B = diag(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_n})$ then $B^2 = D$, so $A = PB^2P^T = (PBP^T)^2$. Take $C = PBP^T$. Since C has eigenvalues $\sqrt{\lambda_i} > 0$, it is positive definite.

Problem 437 *Let A be a positive definite matrix. If a is a real number, show that aA is positive definite if and only if $a > 0$.*

Problem 438

1. *Suppose an invertible matrix A can be factored in $\mathbb{M}_{nn}$ as $A = LDU$ where L is lower triangular with 1s on the diagonal, U is upper triangular with 1s on the diagonal, and D is diagonal with positive diagonal entries. Show that the factorization is unique: If $A = L_1D_1U_1$ is another such factorization, show that $L_1 = L$, $D_1 = D$, and $U_1 = U$.*
2. *Show that a matrix A is positive definite if and only if A is symmetric and admits a factorization $A = LDU$ as in (a).*

*Click the arrow for answer. If A is positive definite, use Theorem **??** to write $A = U^T U$ where U is upper triangular with positive diagonal D. Then $A = (D^{-1}U)^T D^2 (D^{-1}U)$ so $A = L_1D_1U_1$ is such a factorization if $U_1 = D^{-1}U$, $D_1 = D^2$, and $L_1 = U_1^T$. Conversely, let $A^T = A = LDU$ be such a factorization. Then $U^T D^T L^T = A^T = A = LDU$, so $L = U^T$ by **(a)**. Hence $A = LDL^T = V^T V$ where $V = LD_0$ and D_0 is diagonal with $D_0^2 = D$ (the matrix D_0 exists because D has positive diagonal entries). Hence A is symmetric, and it is positive definite by Example **??**.*

Problem 439 *Let A be positive definite and write $d_r = \det\left({}^{(r)}A\right)$ for each $r = 1, 2, \ldots, n$. If U is the upper triangular matrix obtained in step 1 of the algorithm, show that the diagonal elements $u_{11}, u_{22}, \ldots, u_{nn}$ of U are given by $u_{11} = d_1$, $u_{jj} = d_j / d_{j-1}$ if $j > 1$.*

Text Source

This section was adapted from Section 8.3 of Keith Nicholson's *Linear Algebra with Applications*. (CC-BY-NC-SA)

W. Keith Nicholson, *Linear Algebra with Applications*, Lyryx 2018, Open Edition, p. 433-437.

67 QR Factorization

One of the main virtues of orthogonal matrices is that they can be easily inverted—the transpose is the inverse. This fact, combined with the factorization theorem in this section, provides a useful way to simplify many matrix calculations (for example, in least squares approximation).

Definition 86. Let A be an $m \times n$ matrix with independent columns. A *QR-factorization* of A expresses it as $A = QR$ where Q is $m \times n$ with orthonormal columns and R is an invertible and upper triangular matrix with positive diagonal entries.

The importance of the factorization lies in the fact that there are computer algorithms that accomplish it with good control over round-off error, making it particularly useful in matrix calculations. The factorization is a matrix version of the Gram-Schmidt process.

Suppose $A = \begin{bmatrix} | & | & & | \\ \mathbf{c}_1 & \mathbf{c}_2 & \cdots & \mathbf{c}_n \\ | & | & & | \end{bmatrix}$ is an $m \times n$ matrix with linearly independent columns $\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_n$. The Gram-Schmidt algorithm can be applied to these columns to provide orthogonal columns $\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_n$ where $\mathbf{f}_1 = \mathbf{c}_1$ and

$$\mathbf{f}_k = \mathbf{c}_k - \frac{\mathbf{c}_k \cdot \mathbf{f}_1}{\|\mathbf{f}_1\|^2}\mathbf{f}_1 + \frac{\mathbf{c}_k \cdot \mathbf{f}_2}{\|\mathbf{f}_2\|^2}\mathbf{f}_2 - \cdots - \frac{\mathbf{c}_k \cdot \mathbf{f}_{k-1}}{\|\mathbf{f}_{k-1}\|^2}\mathbf{f}_{k-1}$$

for each $k = 2, 3, \ldots, n$. Now write $\mathbf{q}_k = \dfrac{1}{\|\mathbf{f}_k\|}\mathbf{f}_k$ for each k. Then $\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_n$ are orthonormal columns, and the above equation becomes

$$\|\mathbf{f}_k\|\,\mathbf{q}_k = \mathbf{c}_k - (\mathbf{c}_k \cdot \mathbf{q}_1)\mathbf{q}_1 - (\mathbf{c}_k \cdot \mathbf{q}_2)\mathbf{q}_2 - \cdots - (\mathbf{c}_k \cdot \mathbf{q}_{k-1})\mathbf{q}_{k-1}$$

Using these equations, express each $\mathbf{c}_k$ as a linear combination of the $\mathbf{q}_i$:

$$\begin{aligned}
\mathbf{c}_1 &= \|\mathbf{f}_1\|\,\mathbf{q}_1 \\
\mathbf{c}_2 &= (\mathbf{c}_2 \cdot \mathbf{q}_1)\mathbf{q}_1 + \|\mathbf{f}_2\|\,\mathbf{q}_2 \\
\mathbf{c}_3 &= (\mathbf{c}_3 \cdot \mathbf{q}_1)\mathbf{q}_1 + (\mathbf{c}_3 \cdot \mathbf{q}_2)\mathbf{q}_2 + \|\mathbf{f}_3\|\,\mathbf{q}_3 \\
&\vdots \qquad\qquad \vdots \\
\mathbf{c}_n &= (\mathbf{c}_n \cdot \mathbf{q}_1)\mathbf{q}_1 + (\mathbf{c}_n \cdot \mathbf{q}_2)\mathbf{q}_2 + (\mathbf{c}_n \cdot \mathbf{q}_3)\mathbf{q}_3 + \cdots + \|\mathbf{f}_n\|\,\mathbf{q}_n
\end{aligned}$$

These equations have a matrix form that gives the required factorization:

$$A = \begin{bmatrix} | & | & | & & | \\ \mathbf{c}_1 & \mathbf{c}_2 & \mathbf{c}_3 & \cdots & \mathbf{c}_n \\ | & | & | & & | \end{bmatrix}$$
$$= \begin{bmatrix} | & | & | & & | \\ \mathbf{q}_1 & \mathbf{q}_2 & \mathbf{q}_3 & \cdots & \mathbf{q}_n \\ | & | & | & & | \end{bmatrix} \begin{bmatrix} \|\mathbf{f}_1\| & \mathbf{c}_2 \cdot \mathbf{q}_1 & \mathbf{c}_3 \cdot \mathbf{q}_1 & \cdots & \mathbf{c}_n \cdot \mathbf{q}_1 \\ 0 & \|\mathbf{f}_2\| & \mathbf{c}_3 \cdot \mathbf{q}_2 & \cdots & \mathbf{c}_n \cdot \mathbf{q}_2 \\ 0 & 0 & \|\mathbf{f}_3\| & \cdots & \mathbf{c}_n \cdot \mathbf{q}_3 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \|\mathbf{f}_n\| \end{bmatrix} \tag{97}$$

Here the first factor $Q = \begin{bmatrix} | & | & | & & | \\ \mathbf{q}_1 & \mathbf{q}_2 & \mathbf{q}_3 & \cdots & \mathbf{q}_n \\ | & | & | & & | \end{bmatrix}$ has orthonormal columns, and the second factor is an $n \times n$ upper triangular matrix R with positive diagonal entries (and so is invertible). We record this in the following theorem.

Theorem 103 (QR-Factorization)**.** Every $m \times n$ matrix A with linearly independent columns has a QR-factorization $A = QR$ where Q has orthonormal columns and R is upper triangular with positive diagonal entries.

The matrices Q and R in Theorem **??** are uniquely determined by A; we return to this below.

Example 155. Find the QR-factorization of $A = \begin{bmatrix} 1 & 1 & 0 \\ -1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}$.

Explanation. Denote the columns of A as $\mathbf{c}_1$, $\mathbf{c}_2$, and $\mathbf{c}_3$, and observe that $\{\mathbf{c}_1, \mathbf{c}_2, \mathbf{c}_3\}$ is independent. If we apply the Gram-Schmidt algorithm to these columns, the result is:

$$\mathbf{f}_1 = \mathbf{c}_1 = \begin{bmatrix} 1 \\ -1 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{f}_2 = \mathbf{c}_2 - \frac{1}{2}\mathbf{f}_1 = \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \\ 1 \\ 0 \end{bmatrix}, \text{ and } \quad \mathbf{f}_3 = \mathbf{c}_3 + \frac{1}{2}\mathbf{f}_1 - \mathbf{f}_2 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}.$$

Write $\mathbf{q}_j = \dfrac{1}{\|\mathbf{f}_j\|^2}\mathbf{f}_j$ for each j, so $\{\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3\}$ is orthonormal. Then equation (**??**) preceding Theorem **??** gives $A = QR$ where

$$Q = \begin{bmatrix} | & | & | \\ \mathbf{q}_1 & \mathbf{q}_2 & \mathbf{q}_3 \\ | & | & | \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & 0 \\ \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & 0 \\ 0 & \frac{2}{\sqrt{6}} & 0 \\ 0 & 0 & 1 \end{bmatrix} = \frac{1}{\sqrt{6}} \begin{bmatrix} \sqrt{3} & 1 & 0 \\ -\sqrt{3} & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & \sqrt{6} \end{bmatrix}$$

$$R = \begin{bmatrix} \|\mathbf{f}_1\| & \mathbf{c}_2 \cdot \mathbf{q}_1 & \mathbf{c}_3 \cdot \mathbf{q}_1 \\ 0 & \|\mathbf{f}_2\| & \mathbf{c}_3 \cdot \mathbf{q}_2 \\ 0 & 0 & \|\mathbf{f}_3\| \end{bmatrix} = \begin{bmatrix} \sqrt{2} & \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \\ 0 & \frac{\sqrt{3}}{\sqrt{2}} & \frac{\sqrt{3}}{\sqrt{2}} \\ 0 & 0 & 1 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 2 & 1 & -1 \\ 0 & \sqrt{3} & \sqrt{3} \\ 0 & 0 & \sqrt{2} \end{bmatrix}$$

The reader can verify that indeed $A = QR$.

If a matrix A has independent rows and we apply QR-factorization to A^T, the result is:

Corollary 12. If A has independent rows, then A factors uniquely as $A = LP$ where P has orthonormal rows and L is an invertible lower triangular matrix with positive main diagonal entries.

Since a square matrix with orthonormal columns is orthogonal, we have

Theorem 104. Every square invertible matrix A has factorizations $A = QR$ and $A = LP$ where Q and P are orthogonal, R is upper triangular with positive diagonal entries, and L is lower triangular with positive diagonal entries.

We now take the time to prove the uniqueness of the QR-factorization.

Theorem 105. Let A be an $m \times n$ matrix with independent columns. If $A = QR$ and $A = Q_1R_1$ are QR-factorizations of A, then $Q_1 = Q$ and $R_1 = R$.

Proof Write $Q = \begin{bmatrix} | & | & & | \\ \mathbf{c}_1 & \mathbf{c}_2 & \cdots & \mathbf{c}_n \\ | & | & & | \end{bmatrix}$ and $Q_1 = \begin{bmatrix} | & | & & | \\ \mathbf{d}_1 & \mathbf{d}_2 & \cdots & \mathbf{d}_n \\ | & | & & | \end{bmatrix}$ in terms of their columns, and observe first that $Q^TQ = I_n = Q_1^TQ_1$ because Q and Q_1 have orthonormal columns. Hence

it suffices to show that $Q_1 = Q$ (then $R_1 = Q_1^T A = Q^T A = R$). Since $Q_1^T Q_1 = I_n$, the equation $QR = Q_1 R_1$ gives $Q_1^T Q = R_1 R^{-1}$; for convenience we write this matrix as

$$Q_1^T Q = R_1 R^{-1} = \left[\; t_{ij} \; \right]$$

This matrix is upper triangular with positive diagonal elements (since this is true for R and R_1), so $t_{ii} > 0$ for each i and $t_{ij} = 0$ if $i > j$. On the other hand, the (i, j)-entry of $Q_1^T Q$ is $\mathbf{d}_i^T \mathbf{c}_j = \mathbf{d}_i \cdot \mathbf{c}_j$, so we have $\mathbf{d}_i \cdot \mathbf{c}_j = t_{ij}$ for all i and j. But each $\mathbf{c}_j$ is in $\text{span}\{\mathbf{d}_1, \mathbf{d}_2, \ldots, \mathbf{d}_n\}$ because $Q = Q_1(R_1 R^{-1})$. We know how to write a vector as a linear combination of an orthonormal basis (using Corollary **??** from Orthogonality and Projections):

$$\mathbf{c}_j = (\mathbf{d}_1 \cdot \mathbf{c}_j)\mathbf{d}_1 + (\mathbf{d}_2 \cdot \mathbf{c}_j)\mathbf{d}_2 + \cdots + (\mathbf{d}_n \cdot \mathbf{c}_j)\mathbf{d}_n = t_{1j}\mathbf{d}_1 + t_{2j}\mathbf{d}_2 + \cdots + t_{jj}\mathbf{d}_i,$$

because $\mathbf{d}_i \cdot \mathbf{c}_j = t_{ij} = 0$ if $i > j$. The first few equations here are

$$\begin{aligned}
\mathbf{c}_1 &= t_{11}\mathbf{d}_1 \\
\mathbf{c}_2 &= t_{12}\mathbf{d}_1 + t_{22}\mathbf{d}_2 \\
\mathbf{c}_3 &= t_{13}\mathbf{d}_1 + t_{23}\mathbf{d}_2 + t_{33}\mathbf{d}_3 \\
\mathbf{c}_4 &= t_{14}\mathbf{d}_1 + t_{24}\mathbf{d}_2 + t_{34}\mathbf{d}_3 + t_{44}\mathbf{d}_4 \\
\vdots &\quad \vdots
\end{aligned}$$

The first of these equations gives $1 = \|\mathbf{c}_1\| = \|t_{11}\mathbf{d}_1\| = |t_{11}| \, \|\mathbf{d}_1\| = t_{11}$, whence $\mathbf{c}_1 = \mathbf{d}_1$. But then we have $t_{12} = \mathbf{d}_1 \cdot \mathbf{c}_2 = \mathbf{c}_1 \cdot \mathbf{c}_2 = 0$, so the second equation becomes $\mathbf{c}_2 = t_{22}\mathbf{d}_2$. Now a similar argument gives $\mathbf{c}_2 = \mathbf{d}_2$, and then $t_{13} = 0$ and $t_{23} = 0$ follows in the same way. Hence $\mathbf{c}_3 = t_{33}\mathbf{d}_3$ and $\mathbf{c}_3 = \mathbf{d}_3$. Continue in this way to get $\mathbf{c}_i = \mathbf{d}_i$ for all i. This proves that $Q_1 = Q$. ■

QR-Algorithm for approximating eigenvalues

In The Power Method and the Dominant Eigenvalue, we learned about an iterative method for computing eigenvalues. We also mentioned that a better method for approximating the eigenvalues of an invertible matrix A depends on the QR-factorization of A. While it is beyond the scope of this book to pursue a detailed discussion of this method, we give an example and conclude with some remarks on the QR-algorithm. The interested reader is referred to J. M. Wilkinson, *The Algebraic Eigenvalue Problem* (Oxford, England: Oxford University Press, 1965) or G. W. Stewart, *Introduction to Matrix Computations* (New York: Academic Press, 1973).

The *QR-algorithm* uses QR-factorization repeatedly to create a sequence of matrices $A_1 = A, A_2, A_3, \ldots$, as follows:

Algorithm 3. Let A be an invertible matrix.

Step 1: Define $A_1 = A$ and factor it as $A_1 = Q_1R_1$.

Step 2: Define $A_2 = R_1Q_1$ and factor it as $A_2 = Q_2R_2$.

Step 3: Define $A_3 = R_2Q_2$ and factor it as $A_3 = Q_3R_3$.

$\vdots$

Step $k+1$: Define $A_{k+1} = R_kQ_k$ (where $A_k = Q_kR_k$).

Note that A_{k+1} is similar to A_k (in fact, $A_{k+1} = R_kQ_k = (Q_k^{-1}A_k)Q_k$), and hence each A_k has the same eigenvalues as A. If the eigenvalues of A are real and have distinct absolute values, the sequence of matrices $A_1, A_2, A_3, \ldots$ converges to an upper triangular matrix with these eigenvalues on the main diagonal. We will discuss the complex case eigenvalue case below.

Example 156. If $A = \begin{bmatrix} 1 & 1 \\ 2 & 0 \end{bmatrix}$ as in Example **??** of The Power Method and the Dominant Eigenvalue, use the QR-algorithm to approximate the eigenvalues.

Explanation. The matrices A_1, A_2, and A_3 are as follows:

$$A_1 = \begin{bmatrix} 1 & 1 \\ 2 & 0 \end{bmatrix} = Q_1R_1 \quad \text{where } Q_1 = \frac{1}{\sqrt{5}}\begin{bmatrix} 1 & 2 \\ 2 & -1 \end{bmatrix} \text{ and } R_1 = \frac{1}{\sqrt{5}}\begin{bmatrix} 5 & 1 \\ 0 & 2 \end{bmatrix}$$

$$A_2 = \frac{1}{5}\begin{bmatrix} 7 & 9 \\ 4 & -2 \end{bmatrix} = \begin{bmatrix} 1.4 & -1.8 \\ -0.8 & -0.4 \end{bmatrix} = Q_2R_2$$

$$\text{where } Q_2 = \frac{1}{\sqrt{65}}\begin{bmatrix} 7 & 4 \\ 4 & -7 \end{bmatrix} \text{ and } R_2 = \frac{1}{\sqrt{65}}\begin{bmatrix} 13 & 11 \\ 0 & 10 \end{bmatrix}$$

$$A_3 = \frac{1}{13}\begin{bmatrix} 27 & -5 \\ 8 & -14 \end{bmatrix} = \begin{bmatrix} 2.08 & -0.38 \\ 0.62 & -1.08 \end{bmatrix}$$

This is converging to $\begin{bmatrix} 2 & * \\ 0 & -1 \end{bmatrix}$ and so is approximating the eigenvalues 2 and -1 on the main diagonal.

Shifting.

We first learned about the concept of *shifting* in The Power Method and the Dominant Eigenvalue. Convergence is accelerated if, at stage k of the algorithm, a number τ_k is chosen and $A_k - \tau_k I$ is

factored in the form $Q_k R_k$ rather than A_k itself. Then

$$Q_k^{-1} A_k Q_k = Q_k^{-1}(Q_k R_k + \tau_k I) Q_k = R_k Q_k + \tau_k I$$

so we take $A_{k+1} = R_k Q_k + \tau_k I$. If the shifts τ_k are carefully chosen, convergence can be greatly improved.

Preliminary Preparation.

A matrix such as

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

is said to be in *upper Hessenberg* form, and the QR-factorizations of such matrices are greatly simplified. Given an $n \times n$ matrix A, a series of orthogonal matrices $H_1, H_2, \ldots, H_m$ (called *Householder matrices* can be easily constructed such that

$$B = H_m^T \cdots H_1^T A H_1 \cdots H_m$$

is in upper Hessenberg form. Then the QR-algorithm can be efficiently applied to B and, because B is similar to A, it produces the eigenvalues of A.

Complex Eigenvalues.

If some of the eigenvalues of a real matrix A are not real, the QR-algorithm converges to a block upper triangular matrix where the diagonal blocks are either 1×1 (the real eigenvalues) or 2×2 (each providing a pair of conjugate complex eigenvalues of A).

Practice Problems

Problem 440 *In each case find the QR-factorization of A.*

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

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

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

4. $A = \begin{bmatrix} 1 & 1 & 0 \\ -1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & -1 & 0 \end{bmatrix}$

Click the arrow to see answer.

1. $Q = \frac{1}{\sqrt{5}} \begin{bmatrix} 2 & -1 \\ 1 & 2 \end{bmatrix}$, $R = \frac{1}{\sqrt{5}} \begin{bmatrix} 5 & 3 \\ 0 & 1 \end{bmatrix}$

2. $Q = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 & 1 & 0 \\ -1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & -1 & 1 \end{bmatrix}$,
$R = \frac{1}{\sqrt{3}} \begin{bmatrix} 3 & 0 & -1 \\ 0 & 3 & 1 \\ 0 & 0 & 2 \end{bmatrix}$

Problem 441 *If R is upper triangular and invertible, show that there exists a diagonal matrix D with diagonal entries ± 1 such that $R_1 = DR$ is invertible, upper triangular, and has positive diagonal entries.*

Problem 442 *If A has independent columns, let $A = QR$ where Q has orthonormal columns and R is invertible and upper triangular. (Some authors do not require a QR-factorization to have positive diagonal entries.) Show that there is a diagonal matrix D with diagonal entries ± 1 such that $A = (QD)(DR)$ is the QR-factorization of A.*

Problem 443 *In each case, find the exact eigenvalues and then approximate them using the QR-algorithm.*

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

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

Click the arrow to see answer.

1. *Eigenvalues* $\lambda_1 = \frac{1}{2}(3+\sqrt{13}) = 3.302776$, $\lambda_2 = \frac{1}{2}(3-\sqrt{13}) = -0.302776$

$A_1 = \begin{bmatrix} 3 & 1 \\ 1 & 0 \end{bmatrix}$, $Q_1 = \frac{1}{\sqrt{10}}\begin{bmatrix} 3 & -1 \\ 1 & 3 \end{bmatrix}$, $R_1 = \frac{1}{\sqrt{10}}\begin{bmatrix} 10 & 3 \\ 0 & -1 \end{bmatrix}$

$A_2 = \frac{1}{10}\begin{bmatrix} 33 & -1 \\ -1 & -3 \end{bmatrix}$,

$Q_2 = \frac{1}{\sqrt{1090}}\begin{bmatrix} 33 & 1 \\ -1 & 33 \end{bmatrix}$,

$R_2 = \frac{1}{\sqrt{1090}}\begin{bmatrix} 109 & -3 \\ 0 & -10 \end{bmatrix}$

$A_3 = \frac{1}{109}\begin{bmatrix} 360 & 1 \\ 1 & -33 \end{bmatrix}$

$= \begin{bmatrix} 3.302775 & 0.009174 \\ 0.009174 & -0.302775 \end{bmatrix}$

Problem 444 *If A is symmetric, show that each matrix A_k in the QR-algorithm is also symmetric. Deduce that they converge to a diagonal matrix.*

Click the arrow to see answer. Use induction on k. If $k = 1$, $A_1 = A$. In general $A_{k+1} = Q_k^{-1}A_kQ_k = Q_k^TA_kQ_k$, so the fact that $A_k^T = A_k$ implies $A_{k+1}^T = A_{k+1}$. The eigenvalues of A are all real (Theorem **??***), so the A_k converge to an upper triangular matrix T. But T must also be symmetric (it is the limit of symmetric matrices), so it is diagonal.*

Problem 445 *Apply the QR-algorithm to* $A = \begin{bmatrix} 2 & -3 \\ 1 & -2 \end{bmatrix}$. *Explain.*

Problem 446 *Given a matrix A, let A_k, Q_k, and R_k, $k \geq 1$, be the matrices constructed in the QR-algorithm. Show that $A_k = (Q_1Q_2\cdots Q_k)(R_k\cdots R_2R_1)$ for each $k \geq 1$ and hence that this is a QR-factorization of A_k.*

Text Source

This section was adapted from Sections 8.4 and 8.5 of Keith Nicholson's *Linear Algebra with Applications*. (CC-BY-NC-SA)

W. Keith Nicholson, *Linear Algebra with Applications*, Lyryx 2018, Open Edition, pp. 437–445.

68 SVD Decomposition

We begin this section with an important definition.

Definition 87. Let A be an $m \times n$ matrix. The *singular values* of A are the square roots of the positive eigenvalues of $A^T A$.

Singular Value Decomposition (SVD) can be thought of as a generalization of orthogonal diagonalization of a symmetric matrix to an arbitrary $m \times n$ matrix. This decomposition is the focus of this section.

The following is a useful result that will help when computing the SVD of matrices.

Lemma 10. Let A be an $m \times n$ matrix. Then $A^T A$ and AA^T have the same **nonzero eigenvalues**.

Proof Suppose A is an $m \times n$ matrix, and suppose that λ is a nonzero eigenvalue of $A^T A$. Then there exists a nonzero vector $\mathbf{x} \in \mathbb{R}^n$ such that

$$(A^T A)\mathbf{x} = \lambda \mathbf{x} \tag{98}$$

Multiplying both sides of this equation by A yields:

$$\begin{aligned} A(A^T A)\mathbf{x} &= A\lambda\mathbf{x} \\ (AA^T)(A\mathbf{x}) &= \lambda(A\mathbf{x}) \end{aligned}$$

Since $\lambda \neq 0$ and $\mathbf{x} \neq 0_n$, $\lambda\mathbf{x} \neq 0_n$, and thus by equation (**??**), $(A^T A)\mathbf{x} \neq 0_m$; thus $A^T(A\mathbf{x}) \neq 0_m$, implying that $A\mathbf{x} \neq 0_m$.

Therefore $A\mathbf{x}$ is an eigenvector of AA^T corresponding to eigenvalue λ. An analogous argument can be used to show that every nonzero eigenvalue of AA^T is an eigenvalue of $A^T A$, thus completing the proof. ■

Given an $m \times n$ matrix A, we will see how to express A as a product

$$A = U\Sigma V^T$$

where

- U is an $m \times m$ orthogonal matrix whose columns are eigenvectors of AA^T.
- V is an $n \times n$ orthogonal matrix whose columns are eigenvectors of $A^T A$.

- Σ is an $m \times n$ matrix whose only nonzero values lie on its main diagonal, and are the singular values of A.

How can we find such a decomposition? We are aiming to decompose A in the following form:

$$A = U \begin{bmatrix} \sigma & 0 \\ 0 & 0 \end{bmatrix} V^T$$

where σ is a block matrix of the form

$$\sigma = \begin{bmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_k \end{bmatrix}$$

Thus $A^T = V \begin{bmatrix} \sigma & 0 \\ 0 & 0 \end{bmatrix} U^T$ and it follows that

$$A^T A = V \begin{bmatrix} \sigma & 0 \\ 0 & 0 \end{bmatrix} U^T U \begin{bmatrix} \sigma & 0 \\ 0 & 0 \end{bmatrix} V^T = V \begin{bmatrix} \sigma^2 & 0 \\ 0 & 0 \end{bmatrix} V^T$$

and so $A^T AV = V \begin{bmatrix} \sigma^2 & 0 \\ 0 & 0 \end{bmatrix}$. Similarly, $AA^T U = U \begin{bmatrix} \sigma^2 & 0 \\ 0 & 0 \end{bmatrix}$. Therefore, you would find an orthonormal basis of eigenvectors for AA^T make them the columns of a matrix such that the corresponding eigenvalues are decreasing. This gives U. You could then do the same for $A^T A$ to get V.

We formalize this discussion in the following theorem.

Theorem 106 (Singular Value Decomposition)**.** Let A be an $m \times n$ matrix. Then there exist orthogonal matrices U and V of the appropriate size such that $A = U\Sigma V^T$ where Σ is of the form

$$\Sigma = \begin{bmatrix} \sigma & 0 \\ 0 & 0 \end{bmatrix}$$

and σ is a block matrix of the form

$$\sigma = \begin{bmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_k \end{bmatrix}$$

for the σ_i the singular values of A.

Proof There exists an orthonormal basis, $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ such that $A^TA\mathbf{v}_i = \sigma_i^2\mathbf{v}_i$ where $\sigma_i^2 > 0$ for $i = 1, \ldots, k$, $(\sigma_i > 0)$ and equals zero if $i > k$. Thus for $i > k$, $A\mathbf{v}_i = \mathbf{0}$ because

$$A\mathbf{v}_i \cdot A\mathbf{v}_i = (A\mathbf{v}_i)^T(A\mathbf{v}_i) = \mathbf{v}_i^T(A^TA\mathbf{v}_i) = \mathbf{0}$$

For $i = 1, \ldots, k$, define $\mathbf{u}_i \in \mathbb{R}^m$ by

$$\mathbf{u}_i = \sigma_i^{-1}A\mathbf{v}_i$$

Thus $A\mathbf{v}_i = \sigma_i\mathbf{u}_i$. Now

$$\begin{aligned}\mathbf{u}_i \cdot \mathbf{u}_j &= \sigma_i^{-1}A\mathbf{v}_i \cdot \sigma_j^{-1}A\mathbf{v}_j = \sigma_i^{-1}\mathbf{v}_i^T\sigma_j^{-1}A^TA\mathbf{v}_j \\ &= \sigma_i^{-1}\mathbf{v}_i \cdot \sigma_j^{-1}\sigma_j^2\mathbf{v}_j = \frac{\sigma_j}{\sigma_i}\left(\mathbf{v}_i \cdot \mathbf{v}_j\right)\end{aligned}$$

This means that $\mathbf{u}_i \cdot \mathbf{u}_j = 1$ when $i = j$ and $\mathbf{u}_i \cdot \mathbf{u}_j = 0$ when $i \neq j$. Thus $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ is an orthonormal set of vectors in $\mathbb{R}^m$. Also,

$$AA^T\mathbf{u}_i = AA^T\sigma_i^{-1}A\mathbf{v}_i = \sigma_i^{-1}AA^TA\mathbf{v}_i = \sigma_i^{-1}A\sigma_i^2\mathbf{v}_i = \sigma_i^2\mathbf{u}_i$$

Now, using Gram-Schmidt, extend $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ to an orthonormal basis for all of $\mathbb{R}^m$, $\{\mathbf{u}_1, \ldots, \mathbf{u}_m\}$ and let

$$U = \begin{bmatrix} \mathbf{u}_1 & \cdots & \mathbf{u}_m \end{bmatrix}$$

while $V = \begin{bmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_n \end{bmatrix}$. Thus U is the matrix which has the $\mathbf{u}_i$ as columns and V is defined as the matrix which has the $\mathbf{v}_i$ as columns. Then

$$U^TAV = \begin{bmatrix} \mathbf{u}_1^T \\ \vdots \\ \mathbf{u}_k^T \\ \vdots \\ \mathbf{u}_m^T \end{bmatrix} A \begin{bmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_n \end{bmatrix}$$

$$= \begin{bmatrix} \mathbf{u}_1^T \\ \vdots \\ \mathbf{u}_k^T \\ \vdots \\ \mathbf{u}_m^T \end{bmatrix} \begin{bmatrix} \sigma_1\mathbf{u}_1 & \cdots & \sigma_k\mathbf{u}_k & \mathbf{0} & \cdots & \mathbf{0} \end{bmatrix} = \begin{bmatrix} \sigma & 0 \\ 0 & 0 \end{bmatrix}$$

where σ is given in the statement of the theorem. ■

The SVD has as an immediate corollary which is given in the following interesting result.

Corollary 13. Let A be an $m \times n$ matrix. Then the rank of A (or of A^T) equals the number of singular values.

Let's compute the SVD of a simple matrix.

Example 157. Let $A = \begin{bmatrix} 1 & -1 & 3 \\ 3 & 1 & 1 \end{bmatrix}$. Find the SVD of A.

Explanation. To begin, we compute AA^T and A^TA.

$$AA^T = \begin{bmatrix} 1 & -1 & 3 \\ 3 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 3 \\ -1 & 1 \\ 3 & 1 \end{bmatrix} = \begin{bmatrix} 11 & 5 \\ 5 & 11 \end{bmatrix}$$

$$A^TA = \begin{bmatrix} 1 & 3 \\ -1 & 1 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} 1 & -1 & 3 \\ 3 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 10 & 2 & 6 \\ 2 & 2 & -2 \\ 6 & -2 & 10 \end{bmatrix}$$

Since AA^T is 2×2 while A^TA is 3×3, and AA^T and A^TA have the same *nonzero* eigenvalues (by Lemma **??**), we compute the characteristic polynomial $c_{AA^T}(x)$ (because it is easier to compute than $c_{A^TA}(x)$).

$$\begin{aligned} c_{AA^T}(z) &= \det(zI - AA^T) = \det \begin{bmatrix} z-11 & -5 \\ -5 & z-11 \end{bmatrix} \\ &= (z-11)^2 - 25 \\ &= z^2 - 22z + 121 - 25 \\ &= z^2 - 22z + 96 \\ &= (z-16)(z-6) \end{aligned}$$

Therefore, the eigenvalues of AA^T are $\lambda_1 = 16$ and $\lambda_2 = 6$.

The eigenvalues of A^TA are $\lambda_1 = 16$, $\lambda_2 = 6$, and $\lambda_3 = 0$, and the singular values of A are $\sigma_1 = \sqrt{16} = 4$ and $\sigma_2 = \sqrt{6}$. By convention, we list the eigenvalues (and corresponding singular values) in non increasing order (i.e., from largest to smallest).

To find the matrix V:

To construct the matrix V we need to find eigenvectors for A^TA. Since the eigenvalues of AA^T are distinct, the corresponding eigenvectors are orthogonal, and we need only normalize them.

$\lambda_1 = 16$: solve $(16I - A^TA)\mathbf{x}_1 = \mathbf{0}$.

$$\left[\begin{array}{ccc|c} 6 & -2 & -6 & 0 \\ -2 & 14 & 2 & 0 \\ -6 & 2 & 6 & 0 \end{array}\right] \rightarrow \left[\begin{array}{ccc|c} 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right], \text{ so } \mathbf{x}_1 = \begin{bmatrix} t \\ 0 \\ t \end{bmatrix} = t \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, t \in \mathbb{R}.$$

$\lambda_2 = 6$: solve $(6I - A^T A)\mathbf{x}_2 = \mathbf{0}$.

$$\left[\begin{array}{ccc|c} -4 & -2 & -6 & 0 \\ -2 & 4 & 2 & 0 \\ -6 & 2 & -4 & 0 \end{array}\right] \rightarrow \left[\begin{array}{ccc|c} 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right], \text{ so } \mathbf{x}_2 = \begin{bmatrix} -s \\ -s \\ s \end{bmatrix} = s \begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix}, s \in \mathbb{R}.$$

$\lambda_3 = 0$: solve $(-A^T A)\mathbf{x}_3 = \mathbf{0}$.

$$\left[\begin{array}{ccc|c} -10 & -2 & -6 & 0 \\ -2 & -2 & 2 & 0 \\ -6 & 2 & -10 & 0 \end{array}\right] \rightarrow \left[\begin{array}{ccc|c} 1 & 0 & 1 & 0 \\ 0 & 1 & -2 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right], \text{ so } \mathbf{x}_3 = \begin{bmatrix} -r \\ 2r \\ r \end{bmatrix} = r \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix}, r \in \mathbb{R}.$$

Let

$$\mathbf{v}_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \mathbf{v}_2 = \frac{1}{\sqrt{3}} \begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix}, \mathbf{v}_3 = \frac{1}{\sqrt{6}} \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix}$$

Then

$$V = \frac{1}{\sqrt{6}} \begin{bmatrix} \sqrt{3} & -\sqrt{2} & -1 \\ 0 & -\sqrt{2} & 2 \\ \sqrt{3} & \sqrt{2} & 1 \end{bmatrix}$$

Also,

$$\Sigma = \begin{bmatrix} 4 & 0 & 0 \\ 0 & \sqrt{6} & 0 \end{bmatrix}$$

and we use A, V^T, and Σ to find U. Since V is orthogonal and $A = U\Sigma V^T$, it follows that $AV = U\Sigma$. Let $V = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3 \end{bmatrix}$, and let $U = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 \end{bmatrix}$, where $\mathbf{u}_1$ and $\mathbf{u}_2$ are the two columns of U. Then we have

$$\begin{aligned} A \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3 \end{bmatrix} &= \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 \end{bmatrix} \Sigma \\ \begin{bmatrix} A\mathbf{v}_1 & A\mathbf{v}_2 & A\mathbf{v}_3 \end{bmatrix} &= \begin{bmatrix} \sigma_1\mathbf{u}_1 + 0\mathbf{u}_2 & 0\mathbf{u}_1 + \sigma_2\mathbf{u}_2 & 0\mathbf{u}_1 + 0\mathbf{u}_2 \end{bmatrix} \\ &= \begin{bmatrix} \sigma_1\mathbf{u}_1 & \sigma_2\mathbf{u}_2 & 0 \end{bmatrix} \end{aligned}$$

which implies that $A\mathbf{v}_1 = \sigma_1\mathbf{u}_1 = 4\mathbf{u}_1$ and $A\mathbf{v}_2 = \sigma_2\mathbf{u}_2 = \sqrt{6}\mathbf{u}_2$. Thus,

$$\mathbf{u}_1 = \frac{1}{4}A\mathbf{v}_1 = \frac{1}{4}\begin{bmatrix} 1 & -1 & 3 \\ 3 & 1 & 1 \end{bmatrix}\frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} = \frac{1}{4\sqrt{2}}\begin{bmatrix} 4 \\ 4 \end{bmatrix} = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

and

$$\mathbf{u}_2 = \frac{1}{\sqrt{6}}A\mathbf{v}_2 = \frac{1}{\sqrt{6}}\begin{bmatrix} 1 & -1 & 3 \\ 3 & 1 & 1 \end{bmatrix}\frac{1}{\sqrt{3}}\begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix} = \frac{1}{3\sqrt{2}}\begin{bmatrix} 3 \\ -3 \end{bmatrix} = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

Therefore,

$$U = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

and

$$\begin{aligned} A &= \begin{bmatrix} 1 & -1 & 3 \\ 3 & 1 & 1 \end{bmatrix} \\ &= \left(\frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}\right)\begin{bmatrix} 4 & 0 & 0 \\ 0 & \sqrt{6} & 0 \end{bmatrix}\left(\frac{1}{\sqrt{6}}\begin{bmatrix} \sqrt{3} & 0 & \sqrt{3} \\ -\sqrt{2} & -\sqrt{2} & \sqrt{2} \\ -1 & 2 & 1 \end{bmatrix}\right) \end{aligned}$$

Here is another example.

Example 158. Find an SVD for $A = \begin{bmatrix} -1 \\ 2 \\ 2 \end{bmatrix}$.

Explanation. Since A is 3×1, A^TA is a 1×1 matrix whose eigenvalues are easier to find than the eigenvalues of the 3×3 matrix AA^T.

$$A^TA = \begin{bmatrix} -1 & 2 & 2 \end{bmatrix}\begin{bmatrix} -1 \\ 2 \\ 2 \end{bmatrix} = \begin{bmatrix} 9 \end{bmatrix}$$

Thus A^TA has eigenvalue $\lambda_1 = 9$, and the eigenvalues of AA^T are $\lambda_1 = 9$, $\lambda_2 = 0$, and $\lambda_3 = 0$. Furthermore, A has only one singular value, $\sigma_1 = 3$.

To find the matrix V: To do so we find an eigenvector for A^TA and normalize it. In this case, finding a unit eigenvector is trivial: $\mathbf{v}_1 = \begin{bmatrix} 1 \end{bmatrix}$, and

$$V = \begin{bmatrix} 1 \end{bmatrix}$$

Also, $\Sigma = \begin{bmatrix} 3 \\ 0 \\ 0 \end{bmatrix}$, and we use A, V^T, and Σ to find U.

Now $AV = U\Sigma$, with $V = \begin{bmatrix} \mathbf{v}_1 \end{bmatrix}$, and $U = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \mathbf{u}_3 \end{bmatrix}$, where $\mathbf{u}_1$, $\mathbf{u}_2$, and $\mathbf{u}_3$ are the columns of U. Thus

$$\begin{aligned} A\begin{bmatrix} \mathbf{v}_1 \end{bmatrix} &= \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \mathbf{u}_3 \end{bmatrix}\Sigma \\ \begin{bmatrix} A\mathbf{v}_1 \end{bmatrix} &= \begin{bmatrix} \sigma_1\mathbf{u}_1 + 0\mathbf{u}_2 + 0\mathbf{u}_3 \end{bmatrix} \\ &= \begin{bmatrix} \sigma_1\mathbf{u}_1 \end{bmatrix} \end{aligned}$$

This gives us $A\mathbf{v}_1 = \sigma_1\mathbf{u}_1 = 3\mathbf{u}_1$, so

$$\mathbf{u}_1 = \frac{1}{3}A\mathbf{v}_1 = \frac{1}{3}\begin{bmatrix} -1 \\ 2 \\ 2 \end{bmatrix}\begin{bmatrix} 1 \end{bmatrix} = \frac{1}{3}\begin{bmatrix} -1 \\ 2 \\ 2 \end{bmatrix}$$

The vectors $\mathbf{u}_2$ and $\mathbf{u}_3$ are eigenvectors of AA^T corresponding to the eigenvalue $\lambda_2 = \lambda_3 = 0$. Instead of solving the system $(0I - AA^T)\mathbf{x} = 0$ and then using the Gram-Schmidt process on the resulting set of two basic eigenvectors, the following approach may be used.

Find vectors $\mathbf{u}_2$ and $\mathbf{u}_3$ by first extending $\{\mathbf{u}_1\}$ to a basis of $\mathbb{R}^3$, then using the Gram-Schmidt algorithm to orthogonalize the basis, and finally normalizing the vectors.

Starting with $\{3\mathbf{u}_1\}$ instead of $\{\mathbf{u}_1\}$ makes the arithmetic a bit easier. It is easy to verify that

$$\left\{\begin{bmatrix} -1 \\ 2 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}\right\}$$

is a basis of $\mathbb{R}^3$. Set

$$\mathbf{x}_1 = \begin{bmatrix} -1 \\ 2 \\ 2 \end{bmatrix}, \mathbf{x}_2 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \mathbf{x}_3 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

and apply the Gram-Schmidt algorithm to $\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\}$. This gives us

$$\mathbf{f}_1 = \begin{bmatrix} -1 \\ 2 \\ 2 \end{bmatrix}, \mathbf{f}_2 = \begin{bmatrix} 4 \\ 1 \\ 1 \end{bmatrix} \text{ and } \mathbf{f}_3 = \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}$$

Therefore,

$$\mathbf{u}_2 = \frac{1}{\sqrt{18}} \begin{bmatrix} 4 \\ 1 \\ 1 \end{bmatrix}, \mathbf{u}_3 = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}$$

and

$$U = \begin{bmatrix} -\frac{1}{3} & \frac{4}{\sqrt{18}} & 0 \\ \frac{2}{3} & \frac{1}{\sqrt{18}} & \frac{1}{\sqrt{2}} \\ \frac{2}{3} & \frac{1}{\sqrt{18}} & -\frac{1}{\sqrt{2}} \end{bmatrix}$$

Finally,

$$A = \begin{bmatrix} -1 \\ 2 \\ 2 \end{bmatrix} = \begin{bmatrix} -\frac{1}{3} & \frac{4}{\sqrt{18}} & 0 \\ \frac{2}{3} & \frac{1}{\sqrt{18}} & \frac{1}{\sqrt{2}} \\ \frac{2}{3} & \frac{1}{\sqrt{18}} & -\frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 3 \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} 1 \end{bmatrix}$$

Consider another example.

Example 159. Find an SVD for the matrix

$$A = \begin{bmatrix} \frac{2}{5}\sqrt{2}\sqrt{5} & \frac{4}{5}\sqrt{2}\sqrt{5} & 0 \\ \frac{2}{5}\sqrt{2}\sqrt{5} & \frac{4}{5}\sqrt{2}\sqrt{5} & 0 \end{bmatrix}$$

Explanation. First consider $A^T A$

$$\begin{bmatrix} \frac{16}{5} & \frac{32}{5} & 0 \\ \frac{32}{5} & \frac{64}{5} & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

What are some eigenvalues and eigenvectors? Some computing shows that the eigenvalues are 0 and 16. Furthermore, we can find a basis for each eigenspace.

$$\mathcal{S}_0 = \text{span}\left(\left[\begin{array}{c} 0 \\ 0 \\ 1 \end{array}\right], \left[\begin{array}{c} -\frac{2}{5}\sqrt{5} \\ \frac{1}{5}\sqrt{5} \\ 0 \end{array}\right]\right), \quad \mathcal{S}_{16} = \text{span}\left(\left[\begin{array}{c} \frac{1}{5}\sqrt{5} \\ \frac{2}{5}\sqrt{5} \\ 0 \end{array}\right]\right)$$

Thus the matrix V is given by

$$V = \left[\begin{array}{ccc} \frac{1}{5}\sqrt{5} & -\frac{2}{5}\sqrt{5} & 0 \\ \frac{2}{5}\sqrt{5} & \frac{1}{5}\sqrt{5} & 0 \\ 0 & 0 & 1 \end{array}\right]$$

Next consider AA^T

$$\left[\begin{array}{cc} 8 & 8 \\ 8 & 8 \end{array}\right]$$

Eigenvalues are 0 and 16, and eigenspaces are

$$\mathcal{S}_0 = \text{span}\left(\left[\begin{array}{c} -\frac{1}{2}\sqrt{2} \\ \frac{1}{2}\sqrt{2} \end{array}\right]\right), \quad \mathcal{S}_{16} = \text{span}\left(\left[\begin{array}{c} \frac{1}{2}\sqrt{2} \\ \frac{1}{2}\sqrt{2} \end{array}\right]\right)$$

Thus you can let U be given by

$$U = \left[\begin{array}{cc} \frac{1}{2}\sqrt{2} & -\frac{1}{2}\sqrt{2} \\ \frac{1}{2}\sqrt{2} & \frac{1}{2}\sqrt{2} \end{array}\right]$$

Let's check this. $U^T AV =$

$$\left[\begin{array}{cc} \frac{1}{2}\sqrt{2} & \frac{1}{2}\sqrt{2} \\ -\frac{1}{2}\sqrt{2} & \frac{1}{2}\sqrt{2} \end{array}\right] \left[\begin{array}{ccc} \frac{2}{5}\sqrt{2}\sqrt{5} & \frac{4}{5}\sqrt{2}\sqrt{5} & 0 \\ \frac{2}{5}\sqrt{2}\sqrt{5} & \frac{4}{5}\sqrt{2}\sqrt{5} & 0 \end{array}\right] \left[\begin{array}{ccc} \frac{1}{5}\sqrt{5} & -\frac{2}{5}\sqrt{5} & 0 \\ \frac{2}{5}\sqrt{5} & \frac{1}{5}\sqrt{5} & 0 \\ 0 & 0 & 1 \end{array}\right]$$

$$= \left[\begin{array}{ccc} 4 & 0 & 0 \\ 0 & 0 & 0 \end{array}\right]$$

This illustrates that if you have a good way to find the eigenvectors and eigenvalues for a Hermitian matrix which has nonnegative eigenvalues, then you also have a good way to find the SVD of an arbitrary matrix.

Practice Problems

Problem 447 *If A is a square matrix, show that $|det(A)|$ is the product of the singular values of A.*

Problem 448 *Find an SVD for each of the following matrices.*

$$A = \begin{bmatrix} 1 & -1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 1 & 1 \\ -1 & 0 & -2 \\ 1 & 2 & 0 \end{bmatrix}$$

Problem 449 *Find an SVD for $A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$.*

Text Source

This section was adapted from Section 8.11 of Keith Nicholson's *Linear Algebra with Applications*. (CC-BY-NC-SA)

W. Keith Nicholson, *Linear Algebra with Applications*, Lyryx 2018, Open Edition.

69 Abstract Vector Spaces

70 Abstract Vector Spaces

Earlier in the text we discussed $\mathbb{R}^n$ as a vector space and introduced the notion of a subspace of $\mathbb{R}^n$. In this section we will consider sets other than $\mathbb{R}^n$ that have two operations and satisfy the same properties as $\mathbb{R}^n$. Such sets, together with the operations of addition and scalar multiplication, will also be called vector spaces.

Properties of Vector Spaces

Recall that $\mathbb{R}^n$ is said to be a vector space because

$\mathbb{R}^n$ is closed under vector addition

$\mathbb{R}^n$ is closed under scalar multiplication

and satisfies the following properties:

1. Commutative Property of Addition: $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$
2. Associative Property of Addition: $(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$
3. Existence of Additive Identity: $\mathbf{u} + \mathbf{0} = \mathbf{u}$
4. Existence of Additive Inverse: $\mathbf{u} + (-\mathbf{u}) = \mathbf{0}$
5. Distributive Property over Vector Addition: $k(\mathbf{u} + \mathbf{v}) = k\mathbf{u} + k\mathbf{v}$
6. Distributive Property over Scalar Addition: $(k + p)\mathbf{u} = k\mathbf{u} + p\mathbf{u}$
7. Associative Property for Scalar Multiplication: $k(p\mathbf{u}) = (kp)\mathbf{u}$
8. Multiplication by 1: $1\mathbf{u} = \mathbf{u}$

In the next two examples we will explore two sets other than $\mathbb{R}^n$ endowed with addition and scalar multiplication and satisfying the same properties.

Example 160. Let $\mathbb{M}_{m,n}$ be the set of all $m \times n$ matrices. Matrix addition and scalar multiplication were defined in Addition and Scalar Multiplication of Matrices.

Observe that the sum of two $m \times n$ matrices is also an $m \times n$ matrix. Likewise, a scalar multiple of an $m \times n$ matrix is an $m \times n$ matrix. Thus

$\mathbb{M}_{m,n}$ is closed under matrix addition

$\mathbb{M}_{m,n}$ is closed under scalar multiplication

In addition, Theorems **??** and **??** give us the following properties of matrix addition and scalar multiplication. Note that these properties are analogous to the eight vector properties above.

1. Commutative Property of Addition: $A + B = B + A$
2. Associative Property of Addition: $(A + B) + C = A + (B + C)$
3. Existence of Additive Identity: $A + O = A$ where O is the $m \times n$ zero matrix
4. Existence of Additive Inverse: $A + (-A) = O$
5. Distributive Property over Matrix Addition: $k(A + B) = kA + kB$
6. Distributive Property over Scalar Addition: $(k + p)A = kA + pA$
7. Associative Property for Scalar Multiplication: $k(pA) = (kp)A$
8. Multiplication by 1: $1A = A$

Example 161. Consider the set $\mathbb{L}$ of all linear functions. This set includes all polynomials of degree 1 and degree 0. We will use addition and scalar multiplication of polynomials as the two operations, and show that $\mathbb{L}$ is closed under those operations and satisfies eight properties analogous to those of vectors of $\mathbb{R}^n$.

Explanation. Elements of $\mathbb{L}$ are functions f given by

$$f(x) = mx + b$$

(Note that m and b can be equal to zero.)

Given f_1 and f_2 in $\mathbb{L}$, it is easy to verify that $f_1 + f_2$ is also in $\mathbb{L}$. This gives us closure under function addition.

For any scalar k, we have

$$kf(x) = k(mx + b) = (km)x + (kb)$$

Therefore kf is in $\mathbb{L}$, and $\mathbb{L}$ is closed under scalar multiplication.

We now proceed to formulate eight properties analogous to those of vectors of $\mathbb{R}^n$.

Let f_1, f_2 and f_3 be elements of $\mathbb{L}$ given by $f_1(x) = m_1x + b_1$, $f_2(x) = m_2x + b_2$, and $f_3(x) = m_3x + b_3$. Let k and p be scalars.

1. Commutative Property of Addition: $f_1 + f_2 = f_2 + f_1$

   This property holds because

$$\begin{aligned} f_1(x) + f_2(x) &= (m_1x + b_1) + (m_2x + b_2) \\ &= (m_2x + b_2) + (m_1x + b_1) \\ &= f_2(x) + f_1(x) \end{aligned}$$

2. Associative Property of Addition: $(f_1 + f_2) + f_3 = f_1 + (f_2 + f_3)$

   This property is easy to verify.

3. Existence of Additive Identity: $f_1 + f_0 = f_1$

   The additive identity f_0 is given by $f_0(x) = 0$. Note that f_0 is in $\mathbb{L}$.

4. Existence of Additive Inverse: $f_1 + (-f_1) = f_0$

   The additive inverse of f_1 is a function $-f_1$ given by $-f_1(x) = -mx + (-b)$. Note that $-f_1$ is in $\mathbb{L}$.

5. Distributive Property over Vector Addition: $k(f_1 + f_2) = kf_1 + kf_2$

   This property holds because

$$\begin{aligned} k(f_1(x) + f_2(x)) &= k((m_1x + b_1) + (m_2x + b_2)) \\ &= k(m_1x + b_1) + k(m_2x + b_2) \\ &= kf_1(x) + kf_2(x) \end{aligned}$$

6. Distributive Property over Scalar Addition: $(k + p)f_1 = kf_1 + pf_1$

   This property holds because

$$\begin{aligned} (k + p)f_1(x) &= (k + p)(m_1x + b_1) \\ &= k(m_1x + b_1) + p(m_1x + b_1) \\ &= kf_1(x) + pf_1(x) \end{aligned}$$

7. Associative Property for Scalar Multiplication: $(k(pf_1)) = (kp)f_1$

   This property holds because

$$\begin{aligned} k(p(f_1(x))) &= k(p(m_1x + b_1)) \\ &= k(pm_1x + pb_1) \\ &= kpm_1x + kpb_1 \\ &= (kp)m_1x + (kp)b_1 \\ &= (kp)(m_1x + b_1) \\ &= (kp)f_1(x) \end{aligned}$$

8. Multiplication by 1

$$1f_1 = f_1$$

Definition of a Vector Space

Examples **??** and **??** show us that there are many times in mathematics when we encounter a set with two operations (that we call addition and scalar multiplication) such that the set is closed under the two operations, and satisfies the same eight properties as $\mathbb{R}^n$. We will refer to such sets as *vector spaces*.

Definition 88. Let V be a nonempty set. Suppose that elements of V can be added together and multiplied by scalars. The set V, together with operations of addition and scalar multiplication, is called a *vector space* provided that

V is closed under addition

V is closed under scalar multiplication

and the following properties hold for $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{w}$ in V and scalars k and p:

1. Commutative Property of Addition: $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$
2. Associative Property of Addition: $(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$
3. Existence of Additive Identity: $\mathbf{u} + \mathbf{0} = \mathbf{u}$
4. Existence of Additive Inverse: $\mathbf{u} + (-\mathbf{u}) = \mathbf{0}$
5. Distributive Property over Vector Addition: $k(\mathbf{u} + \mathbf{v}) = k\mathbf{u} + k\mathbf{v}$

6. Distributive Property over Scalar Addition: $(k + p)\mathbf{u} = k\mathbf{u} + p\mathbf{u}$
7. Associative Property for Scalar Multiplication: $k(p\mathbf{u}) = (kp)\mathbf{u}$
8. Multiplication by 1: $1\mathbf{u} = \mathbf{u}$

We will refer to elements of V as *vectors*.

When scalars k and p in the above definition are restricted to real numbers, as they are in this chapter, vector space V may be referred to as a *vector space over the real numbers*.

Example 162. $\mathbb{M}_{m,n}$ and $\mathbb{L}$ are vector spaces. (See Examples **??** and **??**)

Sets of polynomials provide an important source of examples, so we review some basic facts. A *polynomial with real coefficients* in x is an expression

$$p(x) = a_0 + a_1x + a_2x^2 + ... + a_nx^n$$

where $a_0, a_1, a_2, ..., a_n$ are real numbers called the *coefficients* of the polynomial. If all the coefficients are zero, the polynomial is called the *zero polynomial* and is denoted simply as 0. If $p(x) \neq 0$, the highest power of x with a nonzero coefficient is called the *degree* of $p(x)$ denoted as $\deg(p(x))$. The coefficient itself is called the *leading coefficient* of $p(x)$. Hence $\deg(3 + 5x) = 1$, $\deg(1 + x + x^2) = 2$, and $\deg(4) = 0$. (The degree of the zero polynomial is not defined.)

Let $\mathbb{P}$ denote the set of all polynomials and suppose that

$$p(x) = a_0 + a_1x + a_2x^2 + ...$$
$$q(x) = b_0 + b_1x + b_2x^2 + ...$$

are two polynomials in $\mathbb{P}$ (possibly of different degrees). Then $p(x)$ and $q(x)$ are called *equal* (written $p(x) = q(x)$) if and only if all the corresponding coefficients are equal—that is, $a_0 = b_0$, $a_1 = b_1$, $a_2 = b_2$, and so on. In particular, $a_0 + a_1x + a_2x^2 + ... = 0$ means that $a_0 = 0$, $a_1 = 0$, $a_2 = 0$, ...

The set $\mathbb{P}$ has an addition and scalar multiplication defined on it as follows: if $p(x)$ and $q(x)$ are as before and k is a real number,

$$p(x) + q(x) = (a_0 + b_0) + (a_1 + b_1)x + (a_2 + b_2)x^2 + ...$$
$$kp(x) = ka_0 + (ka_1)x + (ka_2)x^2 + ...$$

Example 163. $\mathbb{P}$ is a vector space.

Explanation. It is easy to see that the sum of two polynomials is again a polynomial, and that a scalar multiple of a polynomial is a polynomial. Thus, $\mathbb{P}$ is closed under addition and scalar multiplication. The other eight vector space properties are easily verified, and we conclude that $\mathbb{P}$ is a vector space.

Example 164. Let Y be the set of all degree two polynomials in x. In other words,

$$Y = \left\{ ax^2 + bx + c : a \neq 0 \right\}$$

We claim that Y is not a vector space.

Explanation. Observe that Y is not closed under addition. To see this, let $y_1 = 2x^2 + 3x + 4$ and let $y_2 = -2x^2$. Then y_1 and y_2 are both elements of Y. However, $y_1 + y_2 = 3x + 4$ is not an element of Y, as it is only a degree one polynomial. We require the coefficient a of x^2 to be nonzero for a polynomial to be in Y, and this is not the case for $y_1 + y_2$.

As an exercise, check the remaining vector space properties one-by-one to see which properties hold and which do not.

Set Y in Example **??** is not a vector space, but if we make a slight modification, we can make it into a vector space.

Example 165. Let $\mathbb{P}^2$ be the set of polynomials of degree two or less. In other words,

$$\mathbb{P}^2 = \left\{ ax^2 + bx + c : a, b, c \in \mathbb{R} \right\}$$

Note that $\mathbb{P}^2$ contains the zero polynomial (let $a = b = c = 0$).

Unlike set Y in Example **??**, $\mathbb{P}^2$ is closed under polynomial addition and scalar multiplication. It is easy to verify that all vector space properties hold, so $\mathbb{P}^2$ is a vector space.

Example 166. Let n be a natural number. Define $\mathbb{P}^n$ to be the set of polynomials of degree n or less than n, then by reasoning similar to Example **??**, $\mathbb{P}^n$ is a vector space.

Subspaces

Definition 89. A nonempty subset U of a vector space V is called a *subspace* of V, provided that U is itself a vector space when given the same addition and scalar multiplication as V.

Example 167. In Example **??** we demonstrated that $\mathbb{P}^2$ is a vector space. From Example **??** we know that $\mathbb{P}$ is a vector space. But $\mathbb{P}^2$ is a subset of $\mathbb{P}$, and uses the same operations of polynomial addition and scalar multiplication. Therefore $\mathbb{P}^2$ is a subspace of $\mathbb{P}$.

Checking all ten properties to verify that a subset of a vector space is a subspace can be cumbersome. Fortunately we have the following theorem.

Theorem 107 (Subspace Test)**.** Let U be a nonempty subset of a vector space V. If U is closed under the operations of addition and scalar multiplication of V, then U is a subspace of V.

Proof To prove that closure is a sufficient condition for U to be a subspace, we will need to show that closure under addition and scalar multiplication of V guarantees that the remaining eight properties are satisfied automatically.

Observe that Properties **??**, **??**, **??**, **??**, **??** and **??** hold for all elements of V. Thus, these properties will hold for all elements of U. We say that these properties are *inherited* from V.

To prove Property **??** we need to show that $\mathbf{0}$, which we know to be an element of V, is contained in U. Let $\mathbf{u}$ be an element of U (recall that U is nonempty). We will show that $0\mathbf{u} = \mathbf{0}$ in V. Then, by closure under scalar multiplication, we will be able to conclude that $0\mathbf{u} = \mathbf{0}$ must be in U.

$$0\mathbf{u} = (0 + 0)\mathbf{u} = 0\mathbf{u} + 0\mathbf{u}$$

Adding the additive inverse of $0\mathbf{u}$ to both sides gives us

$$0\mathbf{u} + (-0\mathbf{u}) = (0\mathbf{u} + 0\mathbf{u}) + (-0\mathbf{u})$$

By Properties **??** and **??**

$$\mathbf{0} = 0\mathbf{u} + (0\mathbf{u} + (-0\mathbf{u}))$$

By Properties **??** and **??**

$$\mathbf{0} = 0\mathbf{u} + \mathbf{0} = 0\mathbf{u}$$

Because U is closed under scalar multiplication $0\mathbf{u} = \mathbf{0}$ is in U.

We know that every element of U, being an element of V, has an additive inverse in V. We need to show that the additive inverse of every element of U is contained in U. Let $\mathbf{u}$ be any element of U. We will show that $(-1)\mathbf{u}$ is the additive inverse of $\mathbf{u}$. Then by closure, $(-1)\mathbf{u}$ will have to be contained in U. To show that $(-1)\mathbf{u}$ is the additive inverse of $\mathbf{u}$, we must show that $\mathbf{u} + (-1)\mathbf{u} = \mathbf{0}$. We compute:

$$\mathbf{u} + (-1)\mathbf{u} = 1\mathbf{u} + (-1)\mathbf{u} = (1 + (-1))\mathbf{u} = 0\mathbf{u} = \mathbf{0}$$

Thus $(-1)\mathbf{u}$ is the additive inverse of $\mathbf{u}$. By closure, $(-1)\mathbf{u}$ is in U. ■

Example 168. Let A be a fixed matrix in $\mathbb{M}_{n,n}$. Show that the set C_A of all $n \times n$ matrices that commute with A under matrix multiplication is a subspace of $\mathbb{M}_{n,n}$.

Explanation. The set C_A consists of all $n \times n$ matrices X such that $AX = XA$. First, observe that C_A is not empty because I_n is an element. Now we need to show that C_A is closed under matrix addition and scalar multiplication.

Suppose that X_1 and X_2 lie in C_A. Then $AX_1 = X_1A$ and $AX_2 = X_2A$. Then

$$A(X_1 + X_2) = AX_1 + AX_2 = X_1A + X_2A + (X_1 + X_2)A$$

Therefore $(X_1 + X_2)$ commutes with A. Thus $(X_1 + X_2)$ is in C_A. We conclude that C_A is closed under matrix addition.

Now suppose X is in C_A. Let k be a scalar, then

$$A(kX) = k(AX) = k(XA) = (kX)A$$

Therefore (kX) commutes with A. We conclude that (kX) is in C_A, and C_A is closed under scalar multiplication. Hence C_A is a subspace of $\mathbb{M}_{n,n}$.

Suppose $p(x)$ is a polynomial and a is a number. Then the number $p(a)$ obtained by replacing x by a in the expression for $p(x)$ is called the *evaluation* of $p(x)$ at a. For example, if $p(x) = 5-6x+2x^2$, then the evaluation of $p(x)$ at $a = 2$ is $p(2) = 5 - 12 + 8 = 1$. If $p(a) = 0$, the number a is called a *root* of $p(x)$.

Example 169. Consider the set U of all polynomials in $\mathbb{P}$ that have 3 as a root:

$$U = \{p(x) \in \mathbb{P} : p(3) = 0\}$$

Show that U is a subspace of $\mathbb{P}$.

Explanation. Observe that U is not empty because $r(x) = x - 3$ is an element of U. Suppose $p(x)$ and $q(x)$ lie in U. Then $p(3) = 0$ and $q(3) = 0$. We have $(p + q)(x) = p(x) + q(x)$ for all x, so $(p + q)(3) = p(3) + q(3) = 0 + 0 = 0$, and U is closed under addition. The verification that U is closed under scalar multiplication is similar.

Linear Combinations and Span

Definition 90. Let V be a vector space and let $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ be vectors in V. A vector $\mathbf{v}$ is said to be a *linear combination* of vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ if

$$\mathbf{v} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \ldots + a_n\mathbf{v}_n$$

for some scalars $a_1, a_2, \ldots, a_n$.

Definition 91. Let V be a vector space and let $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p$ be vectors in V. The set S of all linear combinations of $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p$ is called the *span* of $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p$. We write

$$S = \text{span}(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p)$$

and we say that vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p$ *span* S. Any vector in S is said to be *in the span* of $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p$. The set $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p\}$ is called a *spanning set* for S.

Example 170. Consider $p_1 = 1 + x + 4x^2$ and $p_2 = 1 + 5x + x^2$ in $\mathbb{P}^2$. Determine whether p_1 and p_2 lie in span$\{1 + 2x - x^2, 3 + 5x + 2x^2\}$.

Explanation. For p_1, we want to determine if a and b exist such that

$$p_1 = a(1 + 2x - x^2) + b(3 + 5x + 2x^2)$$

Expanding the right hand side gives us:

$$a + 2ax - ax^2 + 3b + 5bx + 2bx^2$$

Combining like terms, we get:

$$(a + 3b) + (2a + 5b)x + (-a + 2b)x^2$$

Setting this equal to $p_1 = 1 + x + 4x^2$ and equating coefficients of powers of x gives us a system of equations

$$1 = a + 3b, \quad 1 = 2a + 5b, \quad \text{and} \quad 4 = -a + 2b$$

This system has the solution $a = -2$ and $b = 1$, so p_1 is indeed in span$\{1 + 2x - x^2, 3 + 5x + 2x^2\}$.

Turning to $p_2 = 1 + 5x + x^2$, we are looking for a and b such that

$$p_2 = a(1 + 2x - x^2) + b(3 + 5x + 2x^2)$$

Again equating coefficients of powers of x gives equations $1 = a+3b$, $5 = 2a+5b$, and $1 = -a+2b$. But in this case there is no solution, so p_2 is not in span$\{1 + 2x - x^2, 3 + 5x + 2x^2\}$.

Theorem 108. Let V be a vector space. Let S be any subset of V. Then $U = \text{span}(S)$ is a subspace of V.

Proof See Practice Problem **??**. ■

Practice Problems

Problem 450 *Given the set of all points in $\mathbb{R}^2$, define the operations of addition and scalar multiplication as follows:*

Addition: $(a, b) + (c, d) = (a + d, b + c)$
Scalar Multiplication: $k(a, b) = (ka, kb)$

Is addition, as defined above, commutative?

Multiple Choice:

1. *Yes*
2. *No*

Problem 451 *Given the set of all points in $\mathbb{R}^2$, define the operations of addition and scalar multiplication as follows:*

Addition: $(a, b) + (c, d) = (0, b + d)$
Scalar Multiplication: $k(a, b) = (ka, kb)$

Is addition, as defined above, commutative?

Multiple Choice:

1. *Yes*
2. *No*

Is there an additive identity?

Multiple Choice:

1. *Yes*
2. *No*

Problem 452 *Given the set of all points in $\mathbb{R}^2$, define the operations of addition and scalar multiplication as follows:*

Addition: $(a, b) + (c, d) = (a + c, b + d)$
Scalar Multiplication: $k(a, b) = (a, kb)$

Is the Distributive Property over Vector Addition satisfied?

Multiple Choice:

1. *Yes*
2. *No*

Is the Distributive Property over Scalar Addition satisfied?

Multiple Choice:

1. *Yes*
2. *No*

Problem 453 *Let $\mathcal{F}$ be the set of all real-valued functions whose domain is all real numbers. Define addition and scalar multiplication as follows:*

$$(f+g)(x) = f(x) + g(x) \quad (cf)(x) = cf(x)$$

Verify that $\mathcal{F}$ is a vector space.

Problem 454 *Recall that for a fixed matrix A in $\mathbb{M}_{n,n}$, the set C_A of all $n \times n$ matrices that commute with A under matrix multiplication is a subspace of $\mathbb{M}_{n,n}$. (Example **??**)*

Describe all elements of C_I, where I is a 3×3 identity matrix.

Problem 455 *Is the subset of all invertible $n \times n$ matrices a subspace of $\mathbb{M}_{n,n}$?*

Multiple Choice:

1. *Yes*
2. *No*

Prove your claim.

Problem 456 *Is the subset of all symmetric $n \times n$ matrices a subspace of $\mathbb{M}_{n,n}$?*

Multiple Choice:

1. *Yes*
2. *No*

Prove your claim.

Problem 457 *Let Z be a subset of $\mathbb{M}_{n,n}$ that consists of $n \times n$ matrices that commute with every matrix in $\mathbb{M}_{n,n}$ under matrix multiplication. In other words,*

$$Z = \{B : BY = YB \text{ for all } Y \in \mathbb{M}_{n,n}\}$$

Is Z a subspace of $\mathbb{M}_{n,n}$?

Problem 458 *List several elements of span* $\left(\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \right)$. *Suggest a spanning set for* $\mathbb{M}_{2,2}$.

Problem 459 *Find span*$(1, x, x^2, x^3)$.

Problem 460 *Let V be a vector space. Let S be any subset of V. Prove that $U = span(S)$ is a subspace of V. (Theorem* **??***)*

Text Source

The discussion on polynomials was adapted from Section 6.1 of Keith Nicholson's *Linear Algebra with Applications*. (CC-BY-NC-SA)

W. Keith Nicholson, *Linear Algebra with Applications*, Lyryx 2018, Open Edition, p. 331-332

Example Source

Examples **??** and **??** were adapted from Examples 6.2.4 and 6.2.7 of Keith Nicholson's *Linear Algebra with Applications*. (CC-BY-NC-SA)

W. Keith Nicholson, *Linear Algebra with Applications*, Lyryx 2018, Open Edition, p. 340-341

Exercise Source

Practice Problems **??**-**??** is adopted from Problems 9.1.1-9.1.4 of Ken Kuttler's *A First Course in Linear Algebra*. (CC-BY)

Ken Kuttler, *A First Course in Linear Algebra*, Lyryx 2017, Open Edition, p. 469.

71 Bases and Dimension of Abstract Vector Spaces

When working with $\mathbb{R}^n$ and subspaces of $\mathbb{R}^n$ we developed several fundamental ideas including *span*, *linear independence*, *bases* and *dimension*. We will find that these concepts generalize easily to abstract vector spaces and that analogous results hold in these new settings.

Linear Independence

Definition 92 (Linear Independence)**.** Let V be a vector space. Let $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p$ be vectors of V. We say that the set $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p\}$ is *linearly independent* if the only solution to

$$a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \ldots + a_p\mathbf{v}_p = \mathbf{0}$$

is the *trivial solution* $a_1 = a_2 = \ldots = a_p = 0$.

If, in addition to the trivial solution, a *non-trivial solution* (not all $a_1, a_2, \ldots, a_p$ are zero) exists, then we say that the set $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p\}$ is *linearly dependent*.

Example 171. Show that $P = \{1 + x, 3x + x^2, 2 + x - x^2\}$ is linearly independent in $\mathbb{P}^2$.

Explanation. Consider the linear combination equation

$$\begin{aligned} a(1 + x) + b(3x + x^2) + c(2 + x - x^2) &= 0 \\ a + ax + 3bx + bx^2 + 2c + cx - cx^2 &= 0 \\ (a + 2c) + (a + 3b + c)x + (b - c)x^2 &= 0 \end{aligned}$$

The constant term, as well as the coefficients in front of x and x^2, must be equal to 0. This gives us the following system of equations.

$$\begin{array}{rcrcrcl} a & + & & + & 2c & = & 0 \\ a & + & 3b & + & c & = & 0 \\ & & b & - & c & = & 0 \end{array}$$

The only solution is $a = b = c = 0$. We conclude that P is linearly independent in $\mathbb{P}^2$.

Bases and Dimension

Recall that our motivation for defining a basis of a subspace of $\mathbb{R}^n$ was to have a collection of vectors such that every vector of the subspace can be expressed as a unique linear combination

of the vectors in that collection. Definition of a basis (**??**) generalizes to abstract vector spaces as follows.

Definition 93. Let V be a vector space. A set $\mathcal{B}$ of vectors of V is called a *basis* of V provided that

1. $\text{span}(\mathcal{B}) = V$
2. $\mathcal{B}$ is linearly independent.

Theorem 109. Let V be a vector space, and let $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n\}$ be a basis for V. Then every element $\mathbf{v}$ of V has a unique representation as linear combination of the elements of $\mathcal{B}$.

Proof By the definition of a basis, we know that $\mathbf{v}$ can be written as a linear combination of $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n$. Suppose there are two such representations. Then,

$$\mathbf{v} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + ... + a_n\mathbf{v}_n$$

$$\mathbf{v} = b_1\mathbf{v}_1 + b_2\mathbf{v}_2 + ... + b_n\mathbf{v}_n$$

But then we have:

$$a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + ... + a_n\mathbf{v}_n = b_1\mathbf{v}_1 + b_2\mathbf{v}_2 + ... + b_n\mathbf{v}_n$$
$$a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + ... + a_n\mathbf{v}_n - (b_1\mathbf{v}_1 + b_2\mathbf{v}_2 + ... + b_n\mathbf{v}_n) = \mathbf{0}$$
$$(a_1 - b_1)\mathbf{v}_1 + (a_2 - b_2)\mathbf{v}_2 + ... + (a_n - b_n)\mathbf{v}_n = \mathbf{0}$$

Because $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n$ are linearly independent, we have $a_i - b_i = 0$ for $1 \leq i \leq n$. Consequently $a_i = b_i$ for $1 \leq i \leq n$. ■

In Bases and Dimension we defined the dimension of a subspace of $\mathbb{R}^n$ to be the number of elements in a basis. (Definition **??**) We will adopt this definition for abstract vector spaces. As before, to ensure that *dimension* is well-defined we need to establish that this definition is independent of our choice of a basis. The proof of the following theorem is identical to the proof of its counterpart in $\mathbb{R}^n$. (Theorem **??**)

Theorem 110. Let V be a vector space. Suppose $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_t\}$ and $\mathcal{C} = \{\mathbf{w}_1, \mathbf{w}_2, ..., \mathbf{w}_s\}$ be two bases of V. Then $s = t$.

Now we can state the definition.

Definition 94. Let V be a subspace of a vector space W. The *dimension* of V is the number, m, of elements in any basis of V. We write

$$\dim(V) = m$$

In our discussions up to this point, we have always assumed that a basis is nonempty and hence that the dimension of the space is at least 1. However, the zero space $\{\mathbf{0}\}$ has *no* basis. To accommodate for this, we will say that the zero vector space $\{\mathbf{0}\}$ is defined to have dimension 0:

$$\dim\{\mathbf{0}\} = 0$$

Our insistence that $\dim\{\mathbf{0}\} = 0$ amounts to saying that the empty set of vectors is a basis of $\{\mathbf{0}\}$. Thus the statement that "the dimension of a vector space is the number of vectors in any basis" holds even for the zero space.

Example 172. Recall that the vector space $\mathbb{M}_{m,n}$ consists of all $m \times n$ matrices. (See Example **??** of Abstract Vector Spaces). Find a basis and the dimension of $\mathbb{M}_{m,n}$.

Explanation. Let $\mathcal{B}$ consist of $m \times n$ matrices with exactly one entry equal to 1 and all other entries equal to 0. It is clear that every $m \times n$ matrix can be written as a linear combination of elements of $\mathcal{B}$. It is also easy to see that the elements of $\mathcal{B}$ are linearly independent. Thus $\mathcal{B}$ is a basis for $\mathbb{M}_{m,n}$. The set $\mathcal{B}$ contains mn elements, so $\dim(\mathbb{M}_{m,n}) = mn$.

Example 173. Recall that $\mathbb{P}^n$ is the set of all polynomials of degree n or less. (See Example **??** of Abstract Vector Spaces) Show that $\dim(\mathbb{P}^n) = n+1$ and that $\{1, x, x^2, \ldots, x^n\}$ is a basis of $\mathbb{P}^n$.

Explanation. Each polynomial $p(x) = a_0 + a_1 x + \ldots + a_n x^n$ in $\mathbb{P}^n$ is clearly a linear combination of $1, x, \ldots, x^n$, so $\mathbb{P}^n = \text{span}\{1, x, \ldots, x^n\}$.

Suppose $a_0 1 + a_1 x + \cdots + a_n x^n = 0$, then $a_0 = a_1 = \ldots = a_n = 0$. So $\{1, x, \ldots, x^n\}$ is linearly independent and hence is a basis containing $n+1$ vectors. Thus, $\dim(\mathbb{P}^n) = n+1$.

Example 174. Consider the subset

$$C_A = \{X \in \mathbb{M}_{2,2} : AX = XA\}$$

of $\mathbb{M}_{2,2}$.

It was shown in Example **??** of Abstract Vector Spaces that C_A is a subspace for any choice of the matrix A.

Let $A = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$. Show that $\dim(C_A) = 2$ and find a basis of C_A.

Explanation. Suppose $X = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is in C_A. Then

$$\begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$$

$$\begin{bmatrix} a+c & b+d \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} a & a \\ c & c \end{bmatrix}$$

This gives us two relationships:

$$b+d=a \quad \text{and} \quad c=0$$

We can now express a generic element X of C_A as

$$X = \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} b+d & b \\ 0 & d \end{bmatrix} = \begin{bmatrix} b & b \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} d & 0 \\ 0 & d \end{bmatrix} = b \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} + d \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Let

$$\mathcal{B} = \left\{ \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right\}$$

The set $\mathcal{B}$ is linearly independent. (See Practice Problem **??**) Every element X of C_A can be written as a linear combination of elements of $\mathcal{B}$. Thus $C_A = \text{span}\mathcal{B}$. Therefore $\mathcal{B}$ is a basis of C_A, and $\dim(C_A) = 2$.

Example 175. In Practice Problem **??** of Abstract Vector Spaces you demonstrated that the set of all symmetric $n \times n$ matrices is a subspace of $\mathbb{M}_{n,n}$.

Let V be a subspace of $\mathbb{M}_{2,2}$ consisting of all 2×2 symmetric matrices. Find the dimension of V.

Explanation. A matrix A is symmetric if $A^T = A$. In other words, a matrix A is symmetric when entries directly across the main diagonal are equal, so each 2×2 symmetric matrix has the form

$$\begin{bmatrix} a & c \\ c & b \end{bmatrix} = a \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + b \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} + c \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

Hence the set $\mathcal{B} = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \right\}$ spans V. The reader can verify that $\mathcal{B}$ is linearly independent. Thus $\mathcal{B}$ is a basis of V, so $\dim(V) = 3$.

Finite-Dimensional Vector Spaces

Our definition of dimension of a vector space depends on the vector space having a basis. In this section we will establish that any vector space spanned by finitely many vectors has a basis.

Definition 95. A vector space is said to be *finite-dimensional* if it is spanned by finitely many vectors.

Given a finite-dimensional vector space V we will find a basis for V by starting with a linearly independent subset of V and expanding it to a basis. The following results are more general versions of Lemmas **??** and **??**, and Theorem **??** of Bases and Dimension. The proofs are identical and we will omit them.

Lemma 11. Let V be a vector space spanned by n vectors. If a linearly independent subset S of V contains m vectors, then $m \leq n$.

Lemma 12. Let V be a vector space. Let $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ be a linearly independent subset of V. If $\mathbf{u}$ is not in $\text{span}(\mathbf{v}_1, \ldots, \mathbf{v}_k)$, then $\{\mathbf{u}, \mathbf{v}_1, \ldots, \mathbf{v}_k\}$ is linearly independent.

Theorem 111. Let V be a finite-dimensional vector space. Any linearly independent subset of V can be expanded to a basis of V.

Coordinate Vectors

Recall that in the context of $\mathbb{R}^n$ (and subspaces of $\mathbb{R}^n$) the requirement that elements of a basis be linearly independent guarantees that every element of the vector space has a *unique* representation in terms of the elements of the basis. (See Theorem **??** of Introduction to Bases) We proved the same property for abstract vector spaces in Theorem **??**.

Uniqueness of representation in terms of the elements of a basis allows us to associate every element of a vector space V with a unique *coordinate vector* with respect to a given basis. Coordinate vectors were first introduced in Introduction to Bases. We now give a formal definition.

Definition 96. Let V be a vector space, and let $\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ be a basis for V. If $\mathbf{v} = a_1\mathbf{v}_1 + \ldots + a_n\mathbf{v}_n$, then the vector in $\mathbb{R}^n$ whose components are the coefficients $a_1, \ldots, a_n$ is said to be the *coordinate vector for* $\mathbf{v}$ *with respect to* $\mathcal{B}$. We denote the coordinate vector by $[\mathbf{v}]_\mathcal{B}$ and write:

$$[\mathbf{v}]_\mathcal{B} = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}$$

Remark 22. The order of in which vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ appear in $\mathcal{B}$ of Definition **??** is important. Switching the order of these vectors would switch the order of the coordinate vector components. For this reason, we will often use the term *ordered basis* to describe $\mathcal{B}$ in the context of coordinate vectors.

Example 176. The coordinate vector for $p(x) = 4 - 3x^2 + 5x^3$ in $\mathbb{P}^4$ with respect to the ordered

basis $\mathcal{B}_1 = \{1, x, x^2, x^3, x^4\}$ is

$$[p(x)]_{\mathcal{B}_1} = \begin{bmatrix} 4 \\ 0 \\ -3 \\ 5 \\ 0 \end{bmatrix}$$

Now let's change the order of the elements in $\mathcal{B}_1$. The coordinate vector for $p(x) = 4 - 3x^2 + 5x^3$ with respect to the ordered basis $\mathcal{B}_2 = \{x^4, x^3, x^2, x, 1\}$ is

$$[p(x)]_{\mathcal{B}_2} = \begin{bmatrix} 0 \\ 5 \\ -3 \\ 0 \\ 4 \end{bmatrix}$$

Example 177. Show that the set $\mathcal{B} = \{x, 1 + x, x + x^2\}$ is a basis for $\mathbb{P}^2$. Keep the order of elements in $\mathcal{B}$ and find the coordinate vector for $p(x) = 4 - x + 3x^2$ with respect to the ordered basis $\mathcal{B}$.

Explanation. We will begin by showing that the elements of $\mathcal{B}$ are linearly independent. Suppose

$$ax + b(1 + x) + c(x + x^2) = 0$$

then

$$b + (a + b + c)x + cx^2 = 0$$

This gives us the following system of equations:

$$\begin{array}{rcrcrcl} & & b & & & = & 0 \\ a & + & b & + & c & = & 0 \\ & & & & c & = & 0 \end{array}$$

The solution $a = b = c = 0$ is unique. We conclude that $\mathcal{B}$ is linearly independent.

Next, we need to show that $\mathcal{B}$ spans $\mathbb{P}^2$. To this end, we will consider a generic element $p(x) = \alpha + \beta x + \gamma x^2$ of $\mathbb{P}^2$ and attempt to express it as a linear combination of the elements of $\mathcal{B}$.

$$ax + b(1 + x) + c(x + x^2) = \alpha + \beta x + \gamma x^2$$

then

$$b + (a + b + c)x + cx^2 = \alpha + \beta x + \gamma x^2$$

Setting the coefficients of like terms equal to each other gives us

$$\begin{aligned} b &= \alpha \\ a + b + c &= \beta \\ c &= \gamma \end{aligned}$$

Solving this linear system of a, b and c gives us

$$a = \beta - \alpha - \gamma, \quad b = \alpha, \quad c = \gamma$$

(You should verify this.) This shows that every element of $\mathbb{P}^2$ can be written as a linear combination of elements of $\mathcal{B}$. Therefore $\mathcal{B}$ is a basis for $\mathbb{P}^2$.

To find the coordinate vector for $p(x) = 4 - x + 3x^2$ with respect to $\mathcal{B}$ we need to express $p(x)$ as a linear combination of the elements of $\mathcal{B}$. Fortunately, we have already done all the necessary work. For $p(x)$, $\alpha = 4$, $\beta = -1$ and $\gamma = 3$. This gives us the coefficients of the linear combination: $a = \beta - \alpha - \gamma = -8$, $b = \alpha = 4$, $c = \gamma = 3$. We now write $p(x)$ as a linear combination

$$p(x) = -8(x) + 4(1 + x) + 3(x + x^2)$$

The coordinate vector for $p(x)$ with respect to $\mathcal{B}$ is

$$[p(x)]_{\mathcal{B}} = \begin{bmatrix} -8 \\ 4 \\ 3 \end{bmatrix}$$

Example 178. Recall that the set V of all symmetric 2×2 matrices is a subspace of $\mathbb{M}_{2,2}$. In Example **??** we demonstrated that $\mathcal{B} = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \right\}$ is a basis for V. Let $A = \begin{bmatrix} 2 & -3 \\ -3 & 1 \end{bmatrix}$. Observe that A is an element of V.

1. Find the coordinate vector with respect to the ordered basis $\mathcal{B}$ for A.
2. Let $\mathcal{B}' = \left\{ \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \right\}$ be another ordered basis for V. Find the coordinate vector for A with respect to $\mathcal{B}'$.

Explanation. We write A as a linear combination of the elements of $\mathcal{B}$.

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

Thus, the coordinate vector with respect to $\mathcal{B}$ is

$$[A]_{\mathcal{B}} = \begin{bmatrix} 2 \\ 1 \\ -3 \end{bmatrix}$$

The coordinate vector with respect to $\mathcal{B}'$ is

$$[A]_{\mathcal{B}'} = \begin{bmatrix} \dots \dots \\ \dots \dots \\ \dots \dots \end{bmatrix}$$

Coordinate vectors will play a vital role in establishing one of the most fundamental results in linear algebra, that all n-dimensional vector spaces have the same structure as $\mathbb{R}^n$. In Example **??** of Isomorphic Vector Spaces, for instance, we will show that $\mathbb{P}^2$ is essentially the same as $\mathbb{R}^3$.

Practice Problems

Problem 461 *Consider the subset*

$$C_A = \{X \in \mathbb{M}_{2,2} : AX = XA\}$$

of $\mathbb{M}_{2,2}$. *(See Example* **??***)*

Prove that set $\mathcal{B} = \left\{ \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right\}$ *is a linearly independent subset of* C_A.

Problem 462 *Show that the following subset of* $\mathbb{P}^2$ *is linearly independent, and is a basis of* $\mathbb{P}^2$.

$$\{1 + x, 1 - x, x + x^2\}$$

Problem 463 *Show that the following subset of* $\mathbb{P}^2$ *is linearly independent, and is a basis of* $\mathbb{P}^2$.

$$\{x^2, x + 1, 1 - x - x^2\}$$

Problem 464 *Show that the following subset of* $\mathbb{M}_{2,2}$ *is linearly independent, and is a basis of* $\mathbb{M}_{2,2}$.

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

Problem 465 *Find the coordinate vector for* $p(x) = 6 - 2x + 4x^2$ *in* $\mathbb{P}^2$ *with respect to*

$$\mathcal{B} = \{1 + x, 1 - x, x + x^2\}$$

Answer:

$$[p(x)]_{\mathcal{B}} = \begin{bmatrix} \dots \\ \dots \\ \dots \end{bmatrix}$$

Problem 466 *Find the coordinate vector for* $p(x) = 6 - 2x + 4x^2$ *in* $\mathbb{P}^2$ *with respect to*

$$\mathcal{B} = \{x^2, x + 1, 1 - x - x^2\}$$

Answer:

$$[p(x)]_{\mathcal{B}} = \begin{bmatrix} \dots \\ \dots \\ \dots \end{bmatrix}$$

Problem 467 *Find the coordinate vector for* $A = \begin{bmatrix} 4 & -3 \\ 1 & 2 \end{bmatrix}$ *in* $\mathbb{M}_{2,2}$ *with respect to the ordered basis*

$$\mathcal{B} = \left\{ \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & -1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix} \right\}$$

Answer:

$$[A]_{\mathcal{B}} = \begin{bmatrix} \dots \\ \dots \\ \dots \\ \dots \end{bmatrix}$$

Text Source

The discussion of the zero space was adapted from Section 6.3 of Keith Nicholson's *Linear Algebra with Applications*. (CC-BY-NC-SA)

W. Keith Nicholson, *Linear Algebra with Applications*, Lyryx 2018, Open Edition, p. 349

Example Source

Examples **??** and **??** were adapted from Examples 6.3.1 and 6.3.10 of Keith Nicholson's *Linear Algebra with Applications*. (CC-BY-NC-SA)

W. Keith Nicholson, *Linear Algebra with Applications*, Lyryx 2018, Open Edition, p. 346, 350

Exercise Source

Practice Problems **??**, **??** and **??** are Exercises 6.3(a)(b)(c) from Keith Nicholson's *Linear Algebra with Applications*. (CC-BY-NC-SA)

W. Keith Nicholson, *Linear Algebra with Applications*, Lyryx 2018, Open Edition, p. 351

72 Linear Transformations of Abstract Vector Spaces

Recall that a transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ is called a *linear transformation* if the following are true for all vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$, and scalars k.

$$T(k\mathbf{u}) = kT(\mathbf{u})$$

$$T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$$

We generalize this definition as follows.

Definition 97. Let V and W be vector spaces. A transformation $T : V \to W$ is called a *linear transformation* if the following are true for all vectors $\mathbf{u}$ and $\mathbf{v}$ in V, and scalars k.

$$T(k\mathbf{u}) = kT(\mathbf{u})$$

$$T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$$

Example 179. Recall that $\mathbb{M}_{n,n}$ is the set of all $n \times n$ matrices. In Example **??** of Abstract Vector Spaces we demonstrated that $\mathbb{M}_{n,n}$ together with operations of matrix addition and scalar multiplication is a vector space.

Let $T_Q : \mathbb{M}_{n,n} \to \mathbb{M}_{n,n}$ be a transformation defined by $T_Q(A) = QA$, where Q is fixed $n \times n$ matrix. Show that T_Q is a linear transformation.

Explanation. We verify the linearity properties using properties of matrix-matrix and matrix-scalar multiplication. (See Theorem **??** of Matrix Multiplication.) For A and B in $\mathbb{M}_{n,n}$ and a scalar k we have:

$$T_Q(kA) = Q(kA) = k(QA) = kT_Q(A)$$

$$T_Q(A + B) = Q(A + B) = QA + QB = T_Q(A) + T_Q(B)$$

Example 180. Recall that $\mathbb{P}^3$ is the set of polynomials of degree 3 or less than 3. In Example **??** of Abstract Vector Spaces we showed that $\mathbb{P}^3$ together with operations of polynomial addition and scalar multiplication is a vector space.

Suppose $T : \mathbb{R}^3 \to \mathbb{P}^3$ is a linear transformation such that

$$T(\mathbf{i}) = 1 + x - 2x^2 + x^3$$

$$T(\mathbf{j}) = x + 2x^3$$

$$T(\mathbf{k}) = 3 + x^3$$

Find the image of $\begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}$ under T.

Explanation.

$$\begin{aligned} T\left(\begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}\right) &= T(\mathbf{i} - 2\mathbf{j} + \mathbf{k}) = T(\mathbf{i}) - 2T(\mathbf{j}) + T(\mathbf{k}) \\ &= (1 + x - 2x^2 + x^3) - 2(x + 2x^3) + (3 + x^3) \\ &= 4 - x - 2x^2 - 2x^3 \end{aligned}$$

Example 181. Let $T : \mathbb{M}_{3,3} \to \mathbb{R}$ be a transformation such that $T(A) = \text{rank}(A)$. Show that T is not linear.

Explanation. To show that T is not linear it suffices to find two matrices A and B such that $T(A + B) \neq T(A) + T(B)$. Observe that if we pick A and B so that each has rank 3 we would have $T(A)+T(B) = \text{rank}(A)+\text{rank}(B) = 6$ while $T(A+B) = \text{rank}(A+B) \leq 3$. Clearly $T(A+B) \neq T(A)+T(B)$. This argument is sufficient, but if we want to have a specific example, we can find one. Let

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix}$$

Then

$$T(A) = 3 \quad \text{and} \quad T(B) = 3$$

$$T(A + B) = T\left(\begin{bmatrix} 0 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 0 \end{bmatrix}\right) = 1$$

Thus, $1 = T(A + B) \neq T(A) + T(B) = 6$.

Linear Transformations and Bases

Exploration 468 *Suppose we want to define a linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$ by*

$$T(\mathbf{i}) = \begin{bmatrix} 3 \\ -2 \end{bmatrix} \quad \textit{and} \quad T(\mathbf{j}) = \begin{bmatrix} -1 \\ 2 \end{bmatrix}$$

Is this information sufficient to define T? To answer this question we will try to determine what T does to an arbitrary vector of $\mathbb{R}^2$.

If $\mathbf{v}$ is a vector in $\mathbb{R}^2$, then $\mathbf{v}$ can be uniquely expressed as a linear combination of $\mathbf{i}$ and $\mathbf{j}$

$$\mathbf{v} = a\mathbf{i} + b\mathbf{j}$$

By linearity of T we have

$$T(\mathbf{v}) = T(a\mathbf{i} + b\mathbf{j}) = aT(\mathbf{i}) + bT(\mathbf{j}) = a\begin{bmatrix} 3 \\ -2 \end{bmatrix} + b\begin{bmatrix} -1 \\ 2 \end{bmatrix}$$

This shows that the image of every vector of $\mathbb{R}^2$ under T is completely determined by the action of T on the standard unit vectors $\mathbf{i}$ and $\mathbf{j}$.

Vectors $\mathbf{i}$ and $\mathbf{j}$ form a standard basis of $\mathbb{R}^2$. What if we want to use a different basis?

Let $\mathcal{B} = \left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ -1 \end{bmatrix} \right\}$ be our basis of choice for $\mathbb{R}^2$. (How would you verify that $\mathcal{B}$ is a basis of $\mathbb{R}^2$?) And suppose we want to define a linear transformation $S : \mathbb{R}^2 \to \mathbb{R}^2$ by

$$S\left(\begin{bmatrix} 1 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} 0 \\ -1 \end{bmatrix} \quad \text{and} \quad S\left(\begin{bmatrix} 2 \\ -1 \end{bmatrix}\right) = \begin{bmatrix} 2 \\ 0 \end{bmatrix}$$

Is this enough information to define S?

Because $\begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ -1 \end{bmatrix}$ form a basis of $\mathbb{R}^2$, every element $\mathbf{v}$ of $\mathbb{R}^2$ can be written as a unique linear combination

$$\mathbf{v} = a\begin{bmatrix} 1 \\ 1 \end{bmatrix} + b\begin{bmatrix} 2 \\ -1 \end{bmatrix}$$

We can find $S(\mathbf{v})$ as follows:

$$S(\mathbf{v}) = S\left(a\begin{bmatrix} 1 \\ 1 \end{bmatrix} + b\begin{bmatrix} 2 \\ -1 \end{bmatrix}\right) = a\begin{bmatrix} 0 \\ -1 \end{bmatrix} + b\begin{bmatrix} 2 \\ 0 \end{bmatrix}$$

Again, we see how a linear transformation is completely determined by its action on a basis.

Theorem **??** *assures us that given a basis, every vector has a unique representation as a linear combination of the basis vectors. Imagine what would happen if this were not the case. In the first part of this exploration, for instance, we might have been able to represent $\mathbf{v}$ as $a\mathbf{i} + b\mathbf{j}$ and $c\mathbf{i} + d\mathbf{j}$ ($a \neq c$ or $b \neq d$). This would have resulted in $\mathbf{v}$ mapping to two different elements: $aT(\mathbf{i}) + bT(\mathbf{j})$ and $cT(\mathbf{i}) + dT(\mathbf{j})$, implying that T is not even a function.*

Let $\mathcal{B} = \{\mathbf{v}_1, ..., \mathbf{v}_p\}$ be a basis of a vector space V. To define a linear transformation $T : V \to W$, it is sufficient to state the image of each basis vector under T. Once the images of the basis vectors are established, we can determine the images of all vectors of V as follows:

Given any vector $\mathbf{v}$ of V, write $\mathbf{v}$ as a linear combination of the elements of $\mathcal{B}$

$$\mathbf{v} = a_1\mathbf{v}_1 + ... + a_p\mathbf{v}_p.$$

Then

$$T(\mathbf{v}) = T(a_1\mathbf{v}_1 + ... + a_p\mathbf{v}_p) = a_1T(\mathbf{v}_1) + ... + a_pT(\mathbf{v}_p).$$

Coordinate Vectors

Transformations that map vectors to their coordinate vectors will prove to be of great importance. In this section we will prove that such transformations are linear and give several examples.

If V is a vector space, and $\mathcal{B} = \{\mathbf{v}_1, ..., \mathbf{v}_n\}$ is an ordered basis for V then any vector $\mathbf{v}$ of V can be uniquely expressed as $\mathbf{v} = a_1\mathbf{v}_1 + ... + a_n\mathbf{v}_n$ for some scalars $a_1, ..., a_n$. Vector $[\mathbf{v}]_\mathcal{B}$ in $\mathbb{R}^n$ given by

$$[\mathbf{v}]_\mathcal{B} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}$$

is said to be the *coordinate vector for* $\mathbf{v}$ *with respect to the ordered basis* $\mathcal{B}$. (See Definition **??**.)

It turns out that the transformation $T : V \to \mathbb{R}^n$ defined by $T(\mathbf{v}) = [\mathbf{v}]_\mathcal{B}$ is linear. Before we prove linearity of T, consider the following examples.

Example 182. Consider $\mathbb{M}_{2,2}$. Let $\mathcal{B} = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\}$ be an ordered basis for $\mathbb{M}_{2,2}$. (You should do a quick mental check that $\mathcal{B}$ is a legitimate basis.) Define $T : \mathbb{M}_{2,2} \to \mathbb{R}^4$ by $T(A) = [A]_\mathcal{B}$. Find $T\left(\begin{bmatrix} -2 & 3 \\ 1 & -5 \end{bmatrix} \right)$.

Explanation. We need to find the coordinate vector for $\begin{bmatrix} -2 & 3 \\ 1 & -5 \end{bmatrix}$ with respect to $\mathcal{B}$.

$$\begin{bmatrix} -2 & 3 \\ 1 & -5 \end{bmatrix} = -2 \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + 3 \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} + (-5) \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

This gives us:

$$T\left(\begin{bmatrix} -2 & 3 \\ 1 & -5 \end{bmatrix}\right) = \left[\begin{bmatrix} -2 & 3 \\ 1 & -5 \end{bmatrix}\right]_{\mathcal{B}} = \begin{bmatrix} -2 \\ 3 \\ 1 \\ -5 \end{bmatrix}$$

Example 183. Recall that $\mathbb{P}^2$ is the set of polynomials of degree 2 or less than 2. In Example **??** of Abstract Vector Spaces we showed that $\mathbb{P}^2$ is a vector space.

1. Let $\mathcal{B}_1 = \{1, x, x^2\}$ be an ordered basis for $\mathbb{P}^2$. (It is easy to verify that $\mathcal{B}_1$ is a basis.) If $T : \mathbb{P}^2 \to \mathbb{R}^3$ is given by $T(p) = [p]_{\mathcal{B}_1}$, find $T(2x^2 - 3x)$.

2. Let $\mathcal{B}_2 = \{1 + x, 1 - x, x + x^2\}$ be an ordered basis for $\mathbb{P}^2$. (In Practice Problem **??**, you demonstrated that $\mathcal{B}_2$ is a basis.) If $T : \mathbb{P}^2 \to \mathbb{R}^3$ is given by $T(p) = [p]_{\mathcal{B}_2}$, find $T(2x^2 - 3x)$.

Explanation. ?? We express $2x^2 - 3x$ as a linear combination of elements of $\mathcal{B}_1$.

$$2x^2 - 3x = 0 \cdot 1 + (-3)x + 2x^2$$

Therefore

$$[2x^2 - 3x]_{\mathcal{B}_1} = \begin{bmatrix} 0 \\ -3 \\ 2 \end{bmatrix}$$

Note that it is important to keep the basis elements in the same order in which they are listed, as the order of components of the coordinate vector depends on the order of the basis elements. We conclude that

$$T(2x^2 - 3x) = \begin{bmatrix} 0 \\ -3 \\ 2 \end{bmatrix}$$

?? Our goal is to express $2x^2 - 3x$ as a linear combination of the elements of $\mathcal{B}_2$. Thus, we need to find coefficients a, b and c such that

$$2x^2 - 3x = a(1 + x) + b(1 - x) + c(x + x^2) = (a + b) + (a - b + c)x + cx^2$$

This gives us a system of linear equations:

$$\begin{array}{rcrcrcr} a & + & b & & & = & 0 \\ a & - & b & + & c & = & -3 \\ & & & & c & = & 2 \end{array}$$

Solving the system yields $a = -\frac{5}{2}$, $b = \frac{5}{2}$ and $c = 2$. Thus

$$T(2x^2 - 3x) = [2x^2 - 3x]_{\mathcal{B}_2} = \begin{bmatrix} -5/2 \\ 5/2 \\ 2 \end{bmatrix}$$

Theorem 112. Let V be an n-dimensional vector space, and let $\mathcal{B}$ be an ordered basis for V. Then $T : V \to \mathbb{R}^n$ given by $T(\mathbf{v}) = [\mathbf{v}]_{\mathcal{B}}$ is a linear transformation.

Proof First observe that Theorem **??** of Bases and Dimension of Abstract Vector Spaces guarantees that there is only one way to represent each element of V as a linear combination of elements of $\mathcal{B}$. Thus each element of V maps to exactly one element of $\mathbb{R}^n$, as long as the order in which elements of $\mathcal{B}$ appear is taken into account. This proves that T is a function, or a transformation. We will now prove that T is linear.

Let $\mathbf{v}$ be an element of V. We will first show that $T(k\mathbf{v}) = kT(\mathbf{v})$. Suppose $\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$, then $\mathbf{v}$ can be written as a unique linear combination:

$$\mathbf{v} = a_1\mathbf{v}_1 + \ldots + a_n\mathbf{v}_n$$

We have:

$$\begin{aligned} T(k\mathbf{v}) &= T(k(a_1\mathbf{v}_1 + \ldots + a_n\mathbf{v}_n)) \\ &= T((ka_1)\mathbf{v}_1 + \ldots + (ka_n)\mathbf{v}_n) \\ &= \begin{bmatrix} ka_1 \\ \vdots \\ ka_n \end{bmatrix} = k \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} = kT(\mathbf{v}) \end{aligned}$$

We leave it to the reader to verify that $T(\mathbf{v} + \mathbf{w}) = T(\mathbf{v}) + T(\mathbf{w})$. (See Practice Problem **??**.) ■

In our final example, we will consider T in the context of a basis of the codomain, as well as a basis of the domain. This will later help us tackle the question of the matrix of T associated with bases other than the standard basis of $\mathbb{R}^n$.

Example 184. Let

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix} \quad \text{and} \quad \mathbf{v}_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$$

$$\mathbf{w}_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \quad \text{and} \quad \mathbf{w}_2 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

Let

$$V = \text{span}(\mathbf{v}_1, \mathbf{v}_2) \quad \text{and} \quad W = \text{span}(\mathbf{w}_1, \mathbf{w}_2)$$

Because each of $\{\mathbf{v}_1, \mathbf{v}_2\}$ and $\{\mathbf{w}_1, \mathbf{w}_2\}$ is linearly independent, let

$$\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2\} \quad \text{and} \quad \mathcal{C} = \{\mathbf{w}_1, \mathbf{w}_2\}$$

be ordered bases of V and W, respectively.

Define a linear transformation $T : V \to W$ by

$$T(\mathbf{v}_1) = 2\mathbf{w}_1 - 3\mathbf{w}_2 \quad \text{and} \quad T(\mathbf{v}_2) = -\mathbf{w}_1 + 4\mathbf{w}_2$$

1. Verify that $\mathbf{v} = \begin{bmatrix} 2 \\ 5 \\ 1 \end{bmatrix}$ is in V and find the coordinate vector $[\mathbf{v}]_{\mathcal{B}}$.
2. Find $T(\mathbf{v})$ and the coordinate vector $[T(\mathbf{v})]_{\mathcal{C}}$.

Explanation. ?? We need to express $\mathbf{v}$ as a linear combination of $\mathbf{v}_1$ and $\mathbf{v}_2$. This can be done by observation or by solving the equation

$$\begin{bmatrix} 1 & 0 \\ 2 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 2 \\ 5 \\ 1 \end{bmatrix}$$

We find that $a = 2$ and $b = 1$, so $\mathbf{v} = 2\mathbf{v}_1 + \mathbf{v}_2$. Thus $\mathbf{v}$ is in V. The coordinate vector for $\mathbf{v}$ with respect to the ordered basis $\mathcal{B}$ is

$$[\mathbf{v}]_{\mathcal{B}} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

?? By linearity of T we have

$$\begin{aligned} T(\mathbf{v}) = T(2\mathbf{v}_1 + \mathbf{v}_2) &= 2T(\mathbf{v}_1) + T(\mathbf{v}_2) \\ &= 2(2\mathbf{w}_1 - 3\mathbf{w}_2) + (-\mathbf{w}_1 + 4\mathbf{w}_2) \\ &= 3\mathbf{w}_1 - 2\mathbf{w}_2 = \begin{bmatrix} 1 \\ 0 \\ 3 \end{bmatrix} \end{aligned}$$

The coordinate vector for $T(\mathbf{v})$ with respect to the ordered basis $\mathcal{C}$ is

$$[T(\mathbf{v})]_{\mathcal{C}} = \begin{bmatrix} 3 \\ -2 \end{bmatrix}.$$

Practice Problems

Problem 469 *Suppose $T : \mathbb{P}^2 \to \mathbb{M}_{2,2}$ is a linear transformation such that*

$$T(1) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad T(x) = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \quad T(x^2) = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

Find $T(4 - x + 3x^2)$.

Answer:

$$T(4 - x + 3x^2) = \begin{bmatrix} \ldots \ldots & \ldots \ldots \\ \ldots \ldots & \ldots \ldots \end{bmatrix}$$

Problems **??-??**

Define $T : \mathbb{M}_{3,3} \to \mathbb{R}$ by $T(A) = \text{tr}(A)$. (Recall that $\text{tr}(A)$ denotes the *trace* of A, which is the sum of the main diagonal entries of A.)

Problem 470 *Find* $T\left(\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}\right)$

Answer:

$$T\left(\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}\right) = \ldots \ldots$$

Problem 471 *Is T a linear transformation? If so, prove it. If not, give a counterexample.*

Problems **??-??**

Define $T : \mathbb{R}^2 \to \mathbb{M}_{2,2}$ by $T\left(\begin{bmatrix} a \\ b \end{bmatrix}\right) = \begin{bmatrix} a & 1 \\ 1 & b \end{bmatrix}$.

Problem 472 *Find* $T\left(\begin{bmatrix} 2 \\ -1 \end{bmatrix}\right)$.

Answer:

$$T\left(\begin{bmatrix} 2 \\ -1 \end{bmatrix}\right) = \begin{bmatrix} \ldots\ldots & \ldots\ldots \\ \ldots\ldots & \ldots\ldots \end{bmatrix}$$

Problem 473 *Is T a linear transformation? If so, prove it. If not, give a counterexample.*

Problems **??-??**

This problem requires the knowledge of how to compute a 3×3 determinant. (For a quick reminder, see Definition **??** of Cross Product and its Properties.) Define $T : \mathbb{M}_{3,3} \to \mathbb{R}$ by $T(A) = \det(A)$.

Problem 474 *Find* $T\left(\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}\right)$

Answer:

$$T\left(\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}\right) = \ldots\ldots$$

Problem 475 *Is T a linear transformation? If so, prove it. If not, give a counterexample.*

Problems **??-??**

Define $T : \mathbb{P}^3 \to \mathbb{P}^2$ by $T(p(x)) = p'(x)$. (In other words, T maps a polynomial to its derivative.)

Problem 476 *Find* $T(4x^3 - 2x^2 + x + 6)$.

Answer:

$$T(4x^3 - 2x^2 + x + 6) = \ldots\ldots x^2 - \ldots\ldots x + \ldots\ldots$$

Problem 477 *Is T a linear transformation? If so, prove it. If not, give a counterexample.*

Problem 478 *Recall that the set V of all symmetric 2×2 matrices is a subspace of $\mathbb{M}_{2,2}$. In Example* **??** *of Bases and Dimension of Abstract Vector Spaces we demonstrated that $\mathcal{B} =$*

$\left\{\begin{bmatrix}1&0\\0&0\end{bmatrix},\begin{bmatrix}0&0\\0&1\end{bmatrix},\begin{bmatrix}0&1\\1&0\end{bmatrix}\right\}$ *is a basis for* V*. Define* $T: V \to \mathbb{R}^3$ *by* $T(A) = [A]_{\mathcal{B}}$*. Find* $T(I_2)$ *and* $T\left(\begin{bmatrix}2&-3\\-3&1\end{bmatrix}\right)$.

Answer:

$$T(I_2) = \begin{bmatrix}\dots\dots\\\dots\dots\\\dots\dots\end{bmatrix}$$

$$T\left(\begin{bmatrix}2&-3\\-3&1\end{bmatrix}\right) = \begin{bmatrix}\dots\dots\\\dots\dots\\\dots\dots\end{bmatrix}$$

Problem 479 *Let* V *be a subspace of* $\mathbb{R}^3$ *with a basis* $\mathcal{B} = \left\{\begin{bmatrix}2\\1\\-1\end{bmatrix},\begin{bmatrix}0\\3\\2\end{bmatrix}\right\}$*. Find the coordinate vector,* $[\mathbf{v}]_{\mathcal{B}}$*, for* $\mathbf{v} = \begin{bmatrix}4\\-1\\-4\end{bmatrix}$.

$$[\mathbf{v}]_{\mathcal{B}} = \begin{bmatrix}\dots\dots\\\dots\dots\end{bmatrix}$$

Problem 480 *If the order of the basis elements in Problem* **??** *was switched to form a new basis*

$$\mathcal{B}' = \left\{\begin{bmatrix}0\\3\\2\end{bmatrix},\begin{bmatrix}2\\1\\-1\end{bmatrix}\right\}$$

How would this affect the coordinate vector?

$$[\mathbf{v}]_{\mathcal{B}'} = \begin{bmatrix}\dots\dots\\\dots\dots\end{bmatrix}$$

Problem 481 *In Practice Problem* **??** *of Bases and Dimension of Abstract Vector Spaces you demonstrated that* $\mathcal{B} = \{x^2, x+1, 1-x-x^2\}$ *is a basis for* $\mathbb{P}^2$*. Define* $T: \mathbb{P}^2 \to \mathbb{R}^3$ *by* $T(p(x)) = [p(x)]_{\mathcal{B}}$*. Find* $T(0)$*,* $T(x+1)$ *and* $T(x^2-3x+1)$.

Answer:

$$T(0) = \begin{bmatrix} \ldots & \ldots \\ \ldots & \ldots \\ \ldots & \ldots \end{bmatrix}$$

$$T(x+1) = \begin{bmatrix} \ldots & \ldots \\ \ldots & \ldots \\ \ldots & \ldots \end{bmatrix}$$

$$T(x^2 - 3x + 1) = \begin{bmatrix} \ldots & \ldots \\ \ldots & \ldots \\ \ldots & \ldots \end{bmatrix}$$

Problem 482 *Let V and W be vector spaces, and let $\mathcal{B}_V = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4\}$ and $\mathcal{B}_W = \{\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3\}$ be ordered bases of V and W, respectively. Suppose $T : V \to W$ is a linear transformation such that:*

$$T(\mathbf{v}_1) = \mathbf{w}_2$$

$$T(\mathbf{v}_2) = 2\mathbf{w}_1 - 3\mathbf{w}_2$$

$$T(\mathbf{v}_3) = \mathbf{w}_2 + \mathbf{w}_3$$

$$T(\mathbf{v}_4) = -\mathbf{w}_1$$

If $\mathbf{v} = -2\mathbf{v}_1 + 3\mathbf{v}_2 - \mathbf{v}_4$, express $T(\mathbf{v})$ as a linear combination of vectors of $\mathcal{B}_W$.

$$T(\mathbf{v}) = \ldots \ldots \mathbf{w}_1 - \ldots \ldots \mathbf{w}_2 + \ldots \ldots \mathbf{w}_3$$

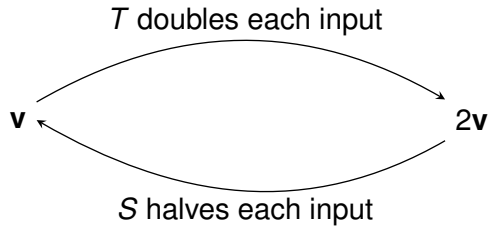
Find $[\mathbf{v}]_{\mathcal{B}_V}$ and $[T(\mathbf{v})]_{\mathcal{B}_W}$

$$[\mathbf{v}]_{\mathcal{B}_V} = \begin{bmatrix} \ldots & \ldots \\ \ldots & \ldots \\ \ldots & \ldots \\ \ldots & \ldots \end{bmatrix}, \quad [T(\mathbf{v})]_{\mathcal{B}_W} = \begin{bmatrix} \ldots & \ldots \\ \ldots & \ldots \\ \ldots & \ldots \end{bmatrix}$$

Problem 483 *Complete the proof of Theorem* **??**.

73 Existence of the Inverse of a Linear Transformation

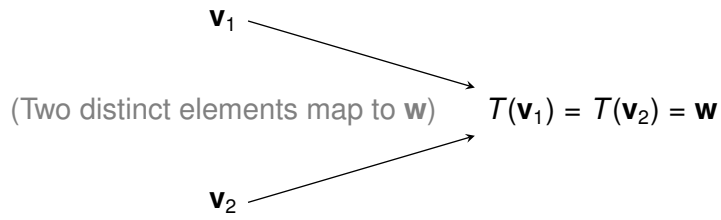
In Exploration **??** of Composition and Inverses of Linear Transformations we examined a linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$ that doubles all input vectors, and its inverse $S : \mathbb{R}^2 \to \mathbb{R}^2$, that halves all input vectors. We observed that the composite functions $S \circ T$ and $T \circ S$ are both identity transformations. Diagrammatically, we can represent T and S as follows:



This gives us a way of thinking about an inverse of T as a transformation that "undoes" the action of T by "reversing" the mapping arrows. We will now use these intuitive ideas to understand which linear transformations are invertible and which are not.

Given an arbitrary linear transformation $T : V \to W$, "reversing the arrows" may not always result in a transformation. Recall that transformations are functions. The figures below show two ways in which our attempt to "reverse" T may fail to produce a function.

First, if two distinct vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ map to the same vector $\mathbf{w}$ in W, then reversing the arrows gives us a mapping that is clearly not a function.



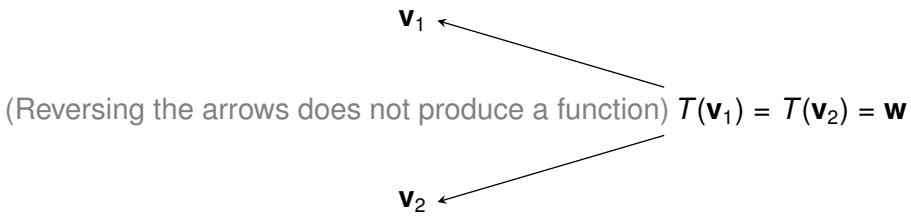


Figure 1. Two distinct elements map to **w**.

Second, observe that our definition of an inverse of $T : V \to W$ requires that the domain of the inverse transformation be W. (Definition **??**, Composition and Inverses of Linear Transformations) If there is a vector **b** in W that is not an image of any vector in V, then **b** cannot be in the domain of an inverse transformation.

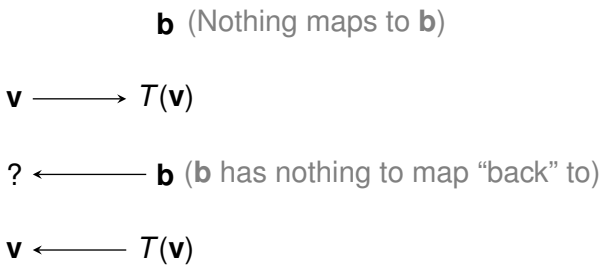


Figure 2. What happens when nothing maps to **b**.

We now illustrate these potential issues with specific examples.

Example 185. Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be a linear transformation whose standard matrix is

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

Does T have an inverse? Show that multiple vectors of the domain map to **0** in the codomain.

Explanation. The matrix $\begin{bmatrix} 1 & 1 \\ 2 & 2 \end{bmatrix}$ is not invertible, so T does not have an inverse.

We now dig a little deeper to get additional insights into why T does not have an inverse. Observe that all vectors of the form $\begin{bmatrix} k \\ -k \end{bmatrix}$ map to **0**. To verify this, use matrix multiplication:

$$\begin{bmatrix} 1 & 1 \\ 2 & 2 \end{bmatrix} \begin{bmatrix} k \\ -k \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

This shows that there are infinitely many vectors that map to **0**. So, "reversing the arrows" would not result in a function. (See Figure 1)

Example 186. Let $T : \mathbb{R}^2 \to \mathbb{R}^3$ be a linear transformation whose standard matrix is

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

Does T have an inverse? Show that there exists a vector $\mathbf{b}$ in $\mathbb{R}^3$ such that no vector of $\mathbb{R}^2$ maps to $\mathbf{b}$.

Explanation. The matrix $\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 2 & 0 \end{bmatrix}$ is not invertible (it's not even a square matrix!), so T does not have an inverse.

We now get another insight into why T is not invertible. To find a vector $\mathbf{b}$ such that no vector of $\mathbb{R}^2$ maps to $\mathbf{b}$, we need to find $\mathbf{b}$ for which the matrix equation

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 2 & 0 \end{bmatrix} \mathbf{x} = \mathbf{b} \tag{99}$$

has no solution.

Let $\mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$. Gauss-Jordan elimination yields:

$$\left[\begin{array}{cc|c} 1 & 0 & b_1 \\ 0 & 1 & b_2 \\ 2 & 0 & b_3 \end{array}\right] \rightsquigarrow \left[\begin{array}{cc|c} 1 & 0 & b_1 \\ 0 & 1 & b_2 \\ 0 & 0 & b_3 - 2b_1 \end{array}\right]$$

Equation (**??**) has a solution if and only if $b_3 - 2b_1 = 0$. Since we do not want (**??**) to have a solution, all we need to do is pick values b_1, b_2 and b_3 such that $b_3 - 2b_1 \neq 0$. Let $\mathbf{b} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$. Then no element of $\mathbb{R}^2$ maps to $\mathbf{b}$. This shows that we cannot "reverse the arrows" in an attempt to produce an inverse of T. (See Figure 2)

Our next goal is to develop vocabulary that would allow us to discuss issues illustrated in Figures 1 and 2.

One-to-one Linear Transformations

Figure 1 gave us a diagrammatic representation of a transformation that maps two distinct elements, $\mathbf{v}_1$ and $\mathbf{v}_2$ to the same element $\mathbf{w}$, making it impossible for us to "reverse the arrows" in an attempt to find the inverse transformation. Based on this example, it is reasonable to conjecture that for a transformation to be invertible, the transformation must be such that each output is the image of exactly one input. Such transformations are called *one-to-one*.

Definition 98 (One-to-One)**.** A linear transformation $T : V \to W$ is *one-to-one* if

$$T(\mathbf{v}_1) = T(\mathbf{v}_2) \quad \text{implies that} \quad \mathbf{v}_1 = \mathbf{v}_2$$

The transformation in Figure 1 is not one-to-one because $\mathbf{v}_1$ and $\mathbf{v}_2$ map to the same vector $\mathbf{w}$, (i.e. $T(\mathbf{v}_1) = T(\mathbf{v}_2)$), yet the diagram suggests that $\mathbf{v}_1 \neq \mathbf{v}_2$.

Example 187. Transformation T in Example **??** is not one-to-one.

Explanation. We can use any two vectors of the form $\begin{bmatrix} k \\ -k \end{bmatrix}$ to make our case.

$$T\left(\begin{bmatrix} 1 \\ -1 \end{bmatrix}\right) = \mathbf{0} = T\left(\begin{bmatrix} -2 \\ 2 \end{bmatrix}\right) \quad \text{but} \quad \begin{bmatrix} 1 \\ -1 \end{bmatrix} \neq \begin{bmatrix} -2 \\ 2 \end{bmatrix}$$

In other words, we have more than one vector that maps to the zero vector.

Example **??** provides us with an important insight. Recall that the collection of all vectors that map to zero under a linear transformation constitute the *kernel* of the linear transformation. (See Image and Kernel of a Linear Transformation) It is clear that when $\ker(T) \neq \{\mathbf{0}\}$ the transformation is not one-to-one. This implication goes the other way as well. In Problem **??** you will show that a linear transformation T is one-to-one if and only if $\ker(T) = \{\mathbf{0}\}$.

Example 188. Prove that the transformation in Example **??** is one-to-one.

Proof Suppose

$$T\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = T\left(\begin{bmatrix} y_1 \\ y_2 \end{bmatrix}\right)$$

Then

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$$

$$x_1 \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} + x_2 \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = y_1 \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} + y_2 \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

$$(x_1 - y_1) \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} + (x_2 - y_2) \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \mathbf{0}$$

It is clear that $\begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}$ and $\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$ are linearly independent. Therefore, we must have $x_1 - y_1 = 0$ and $x_2 - y_2 = 0$. But then $x_1 = y_1$ and $x_2 = y_2$, so

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$$

■

Since transformation in Example **??** is one-to-one but not invertible we can conjecture that being one-to-one is a necessary, but not a sufficient condition for a linear transformation to have an inverse. We will consider the other necessary condition next.

"Onto" Linear Transformations

Figure 2 makes a convincing case that for a transformation to be invertible every element of the codomain must have something mapping to it. Transformations such that every element of the codomain is an image of some element of the domain are called *onto*.

Definition 99 (Onto)**.** A linear transformation $T : V \to W$ is *onto* if for every element $\mathbf{w}$ of W, there exists an element $\mathbf{v}$ of V such that $T(\mathbf{v}) = \mathbf{w}$.

Example 189. The transformation in Example **??** is not onto.

Explanation. No element of $\mathbb{R}^2$ maps to $\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$.

Example 190. Prove that the linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$ whose standard matrix is

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

is onto.

Explanation. Let $\mathbf{b}$ be an element of the codomain ($\mathbb{R}^2$). We need to find $\mathbf{x}$ in the domain ($\mathbb{R}^2$) such that $T(\mathbf{x}) = \mathbf{b}$. Observe that A is invertible, and

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

Let $\mathbf{x} = \begin{bmatrix} 1 & 0 \\ -2 & 1 \end{bmatrix} \mathbf{b}$, then

$$T(\mathbf{x}) = \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix} \left(\begin{bmatrix} 1 & 0 \\ -2 & 1 \end{bmatrix} \mathbf{b} \right) = I\mathbf{b} = \mathbf{b}$$

Example 191. Prove that the linear transformation $T : \mathbb{R}^3 \to \mathbb{R}^2$ induced by

$$A = \begin{bmatrix} 1 & 1 & -1 \\ 2 & 3 & -1 \end{bmatrix}$$

is onto.

Explanation. Let $\mathbf{b}$ be an element of $\mathbb{R}^2$. We need to show that there exists $\mathbf{x}$ in $\mathbb{R}^3$ such that $T(\mathbf{x}) = A\mathbf{x} = \mathbf{b}$. Observe that

$$\text{rref}(A) = \begin{bmatrix} 1 & 0 & -2 \\ 0 & 1 & 1 \end{bmatrix}$$

This means that $A\mathbf{x} = \mathbf{b}$ has a solution (in fact, it has infinitely many solutions) for every $\mathbf{b}$ in $\mathbb{R}^2$. Therefore every $\mathbf{b}$ in $\mathbb{R}^2$ is an image of some $\mathbf{x}$ in $\mathbb{R}^3$. We conclude that T is onto.

Example 192. Let

$$V = \text{span}\left(\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right)$$

Define a linear transformation

$$T : V \to \mathbb{R}^2$$

by

$$T\left(\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \right) = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad \text{and} \quad T\left(\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right) = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

Show that T is one-to-one and onto.

Explanation. We will now show that T is one-to-one. Suppose

$$T(\mathbf{u}) = T(\mathbf{v})$$

for some $\mathbf{u}$ and $\mathbf{v}$ in V. Vectors $\mathbf{u}$ and $\mathbf{v}$ are in the span of $\begin{bmatrix}1\\0\\0\end{bmatrix}$ and $\begin{bmatrix}1\\1\\1\end{bmatrix}$, so

$$\mathbf{u} = a\begin{bmatrix}1\\0\\0\end{bmatrix} + b\begin{bmatrix}1\\1\\1\end{bmatrix} \quad \text{and} \quad \mathbf{v} = c\begin{bmatrix}1\\0\\0\end{bmatrix} + d\begin{bmatrix}1\\1\\1\end{bmatrix}$$

for some scalars a, b, c, d.

$$T(\mathbf{u}) = aT\left(\begin{bmatrix}1\\0\\0\end{bmatrix}\right) + bT\left(\begin{bmatrix}1\\1\\1\end{bmatrix}\right) = a\begin{bmatrix}1\\1\end{bmatrix} + b\begin{bmatrix}0\\1\end{bmatrix} = \begin{bmatrix}a\\a+b\end{bmatrix}$$

$$T(\mathbf{v}) = cT\left(\begin{bmatrix}1\\0\\0\end{bmatrix}\right) + dT\left(\begin{bmatrix}1\\1\\1\end{bmatrix}\right) = c\begin{bmatrix}1\\1\end{bmatrix} + d\begin{bmatrix}0\\1\end{bmatrix} = \begin{bmatrix}c\\c+d\end{bmatrix}$$

Thus,

$$\begin{bmatrix}a\\a+b\end{bmatrix} = \begin{bmatrix}c\\c+d\end{bmatrix}$$

This implies that $a = c$ which, in turn, implies $b = d$. This gives us $\mathbf{u} = \mathbf{v}$, and we conclude that T is one-to-one.

Next we will show that T is onto. The key observation is that vectors $\begin{bmatrix}1\\1\end{bmatrix}$ and $\begin{bmatrix}0\\1\end{bmatrix}$ span $\mathbb{R}^2$. This means that given a vector $\mathbf{v}$ in $\mathbb{R}^2$, we can write $\mathbf{v}$ as $\mathbf{v} = a\begin{bmatrix}1\\1\end{bmatrix} + b\begin{bmatrix}0\\1\end{bmatrix}$. But this means that $\mathbf{v} = T\left(a\begin{bmatrix}1\\0\\0\end{bmatrix} + b\begin{bmatrix}1\\1\\1\end{bmatrix}\right)$ We conclude that T is onto.

Existence of Inverses

Theorem 113. Let V and W be vector spaces, and let $T : V \to W$ be a linear transformation. Then T has an inverse if and only if T is one-to-one and onto.

Proof We will first assume that T is one-to-one and onto, and show that there exists a transformation $S : W \to V$ such that $S \circ T = \mathrm{id}_V$ and $T \circ S = \mathrm{id}_W$. Because T is onto, for every $\mathbf{w}$ in W, there exists $\mathbf{v}$ in V such that $T(\mathbf{v}) = \mathbf{w}$. Moreover, because T is one-to-one, vector $\mathbf{v}$ is the only vector that maps to $\mathbf{w}$. To stress this, we will say that for every $\mathbf{w}$, there exists $\mathbf{v_w}$ such that $T(\mathbf{v_w}) = \mathbf{w}$. (Since every $\mathbf{v}$ maps to exactly one $\mathbf{w}$, this notation makes sense for elements of V as well.) We can now define $S : W \to V$ by $S(\mathbf{w}) = \mathbf{v_w}$. Then

$$(S \circ T)(\mathbf{v_w}) = S(T(\mathbf{v_w})) = S(\mathbf{w}) = \mathbf{v_w}$$

$$(T \circ S)(\mathbf{w}) = T(S(\mathbf{w})) = T(\mathbf{v_w}) = \mathbf{w}$$

We conclude that $S \circ T = \mathrm{id}_V$ and $T \circ S = \mathrm{id}_W$. Therefore S is an inverse of T.

We will now assume that T has an inverse S and show that T must be one-to-one and onto. Suppose

$$T(\mathbf{v}_1) = T(\mathbf{v}_2)$$

then

$$S(T(\mathbf{v}_1)) = S(T(\mathbf{v}_2))$$

but then

$$\mathbf{v}_1 = \mathbf{v}_2$$

We conclude that T is one-to-one.

Now suppose that $\mathbf{w}$ is in W. We need to show that some element of V maps to $\mathbf{w}$. Let $\mathbf{v} = S(\mathbf{w})$. Then

$$T(\mathbf{v}) = T(S(\mathbf{w})) = (T \circ S)(\mathbf{w}) = \mathrm{id}_W(\mathbf{w}) = \mathbf{w}$$

We conclude that T is onto. ■

Example 193. Transformation T in Example **??** is invertible.

Explanation. We demonstrated that T is one-to-one and onto. By Theorem **??**, T has an inverse.

Recall that T was introduced in Exploration **??** of Composition and Inverses of Linear Transformations to demonstrate that Theorem **??** of Composition and Inverses of Linear Transformations is not always directly applicable. We now have additional tools. Theorem **??** assures us that T has an inverse, but does not help us find it. We will visit this problem again, in Matrices of Linear Transformations with Respect to Arbitrary Bases, and find an inverse of T.

Uniqueness of Inverses

Definition **??** of Composition and Inverses of Linear Transformations refers to S as *an* inverse of T, implying that there may be more than one such transformation S. We will now show that if such a transformation S exists, it is unique. This will allow us to refer to it as *the* inverse of T and to start using T^{-1} to denote the unique inverse of T.

Theorem 114. If T is a linear transformation, and S is an inverse of T. Then S is unique.

Proof Let $T : V \to W$ be a linear transformation. If S is an inverse of T, then S satisfies

$$S \circ T = \mathrm{id}_V \quad \text{and} \quad T \circ S = \mathrm{id}_W$$

Suppose there is another transformation, S', such that

$$S' \circ T = \mathrm{id}_V \quad \text{and} \quad T \circ S' = \mathrm{id}_W$$

We now show that $S = S'$.

$$S = S \circ \mathrm{id}_W = S \circ (T \circ S') = (S \circ T) \circ S' = \mathrm{id}_V \circ S' = S'$$

■

Practice Problems

Problem 484 *Show that a linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$ with standard matrix $A = \begin{bmatrix} 2 & -4 \\ -3 & 6 \end{bmatrix}$ is not one-to-one.*

Problem 485 *Show that a linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^3$ with standard matrix $A = \begin{bmatrix} 1 & 2 \\ -1 & 1 \\ 0 & 1 \end{bmatrix}$ is not onto.*

Problem 486 *Suppose that a linear transformation $T : \mathbb{R}^3 \to \mathbb{R}^3$ has a standard matrix A such that $\mathrm{rref}(A) = I$.*

Prove that T is one-to-one. Prove that T is onto.

Problem 487 *Define a transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$ by*

$$T\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} x \\ -2x + 4y \end{bmatrix}$$

Show that T is a linear transformation that has an inverse.

Problem 488 *Let* $V = \text{span}\left(\begin{bmatrix}1\\0\\1\end{bmatrix}, \begin{bmatrix}0\\1\\0\end{bmatrix}\right)$. *Define a linear transformation* $T : V \to \mathbb{R}^2$ *by*

$$T\left(\begin{bmatrix}1\\0\\1\end{bmatrix}\right) = \begin{bmatrix}0\\1\end{bmatrix} \quad \text{and} \quad T\left(\begin{bmatrix}0\\1\\0\end{bmatrix}\right) = \begin{bmatrix}-1\\1\end{bmatrix}$$

Prove that T *has an inverse.*

Problem 489 *Prove that a linear transformation* $T : V \to W$ *is one-to-one if and only if* $\ker(T) = \{\mathbf{0}\}$.

74 Isomorphic Vector Spaces

A vector space is defined as a collection of objects together with operations of addition and scalar multiplication that follow certain rules (Definition **??** of Abstract Vector Spaces). In our study of abstract vector spaces, we have encountered spaces that appeared very different from each other. Just how different are they? Does $\mathbb{L}$, a vector space whose elements have the form $mx+b$, have anything in common with $\mathbb{R}^2$? Is $\mathbb{P}^3$ fundamentally different from $\mathbb{M}_{2,2}$?

To answer these questions, we will have to look beyond the superficial appearance of the elements of a vector space and delve into its structure. The "structure" of a vector space is determined by how the elements of the vector space interact with each other through the operations of addition and scalar multiplication.

Let us return to the question of what $\mathbb{L}$ has in common with $\mathbb{R}^2$. Consider two typical elements of $\mathbb{L}$:

$$mx + b \quad \text{and} \quad qx + c \tag{100}$$

We can add these elements together

$$(mx + b) + (qx + c) = (m + q)x + (b + c) \tag{101}$$

or multiply each one by a scalar

$$k(mx + b) = kmx + kb \quad \text{and} \quad t(qx + c) = tqx + tc \tag{102}$$

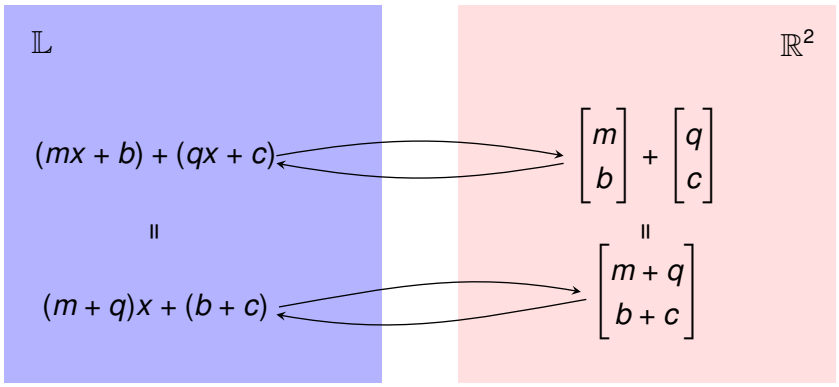
But suppose we get tired of having to write x down every time. Could we leave off the x and represent $mx + b$ by $\begin{bmatrix} m \\ b \end{bmatrix}$? If we do this, expressions (**??**), (**??**) and (**??**) would be mimicked by the following expressions involving vectors of $\mathbb{R}^2$:

$$\begin{bmatrix} m \\ b \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} q \\ c \end{bmatrix}$$

$$\begin{bmatrix} m \\ b \end{bmatrix} + \begin{bmatrix} q \\ c \end{bmatrix} = \begin{bmatrix} m+q \\ b+c \end{bmatrix}$$

$$k \begin{bmatrix} m \\ b \end{bmatrix} = \begin{bmatrix} km \\ kb \end{bmatrix} \quad \text{and} \quad t \begin{bmatrix} q \\ c \end{bmatrix} = \begin{bmatrix} tq \\ tc \end{bmatrix}$$

It appears that we should be able to switch back and forth between $\mathbb{L}$ and $\mathbb{R}^2$, translating questions and answers from one space to the other and back again.



We begin to suspect that $\mathbb{L}$ and $\mathbb{R}^2$ have the same "structure". Spaces such as $\mathbb{L}$ and $\mathbb{R}^2$ are said to be *isomorphic*. This term is derived from the Greek "iso," meaning "same," and "morphe," meaning "form." The term captures the idea that isomorphic vector spaces have the same structure. Before we present a precise definition of the term, we need to better understand what we mean by "switching back and forth" between spaces. The following Exploration will help us formulate this vague notion in terms of transformations.

Exploration 490 *Recall that the set of all polynomials of degree 2 or less, together with polynomial addition and scalar multiplication, is a vector space, denoted by* $\mathbb{P}^2$. *Let* $\mathcal{B} = \{1, x, x^2\}$. *You should do a quick mental check that* $\mathcal{B}$ *is a basis of* $\mathbb{P}^2$.

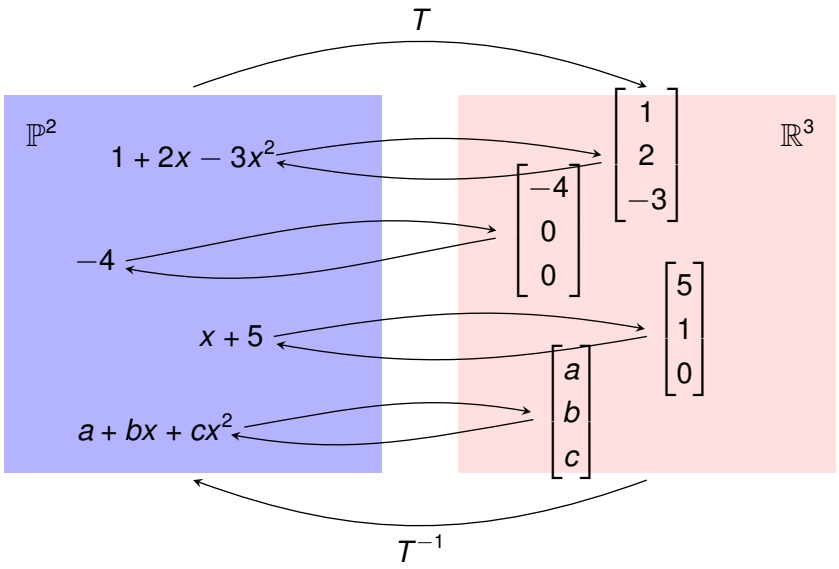
Define a transformation $T : \mathbb{P}^2 \to \mathbb{R}^3$ *by* $T(a + bx + cx^2) = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$. *You may have recognized* T *as the transformation that maps each element of* $\mathbb{P}^2$ *to its coordinate vector with respect to the ordered basis* $\mathcal{B}$.

Transformation T *has several nice properties:*

1. *By Theorem* **??** *of Linear Transformations of Abstract Vector Spaces,* T *is linear.*
2. *It is easy to verify that* T *is one-to-one and onto. (See Practice Problem* **??**.*)*
3. *By Theorem* **??** *of Existence of the Inverse of a Linear Transformation,* T *has an inverse.*

Our goal is to investigate and illustrate what these properties mean for transformation T, *and for the relationship between* $\mathbb{P}^2$ *and* $\mathbb{R}^3$.

First, observe that T *being one-to-one and onto establishes "pairings" between elements of* $\mathbb{P}^2$ *and* $\mathbb{R}^3$ *in such a way that every element of one vector space is uniquely matched with exactly one element of the other vector space, as shown in the diagram below.*



Second, the fact that T (and T^{-1}) are linear will allow us to translate questions related to linear combinations in one of the vector spaces to equivalent questions in the other vector space, then translate answers back to the original vector space. To make this statement concrete, consider the following problem:

Let

$$p_1(x) = 3 - x + 2x^2 \quad \text{and} \quad p_2(x) = -1 + 3x + x^2$$

find $p_1(x) + p_2(x)$.

The answer is, of course

$$p_1(x) + p_2(x) = 2 + 2x + 3x^2$$

Easy. But suppose for a moment that we did not know how to add polynomials, or that we found the process extremely difficult, or maybe instead of $\mathbb{P}^2$ we had another vector space that we did not want to deal with.

It turns out that we can use T and T^{-1} to answer the addition question. We will start by applying T to $p_1(x)$ and $p_2(x)$ separately:

$$T(p_1(x)) = \begin{bmatrix} 3 \\ -1 \\ 2 \end{bmatrix}, \quad T(p_2(x)) = \begin{bmatrix} -1 \\ 3 \\ 1 \end{bmatrix}$$

Next, we add the images of $p_1(x)$ and $p_2(x)$ in $\mathbb{R}^3$

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

This maneuver allows us to avoid the addition question in $\mathbb{P}^2$ and answer the question in $\mathbb{R}^3$ instead. We use T^{-1} to translate the answer back to $\mathbb{P}^2$:

$$T^{-1}\left(\begin{bmatrix} 2 \\ 2 \\ 3 \end{bmatrix}\right) = 2 + 2x + 3x^2$$

All of this relies on linearity. Here is a formal justification for the process. Try to spot where linearity is used.

$$p_1(x) + p_2(x) = (3 - x + 2x^2) + (-1 + 3x + x^2) \tag{103}$$

$$= T^{-1}\left(\begin{bmatrix} 3 \\ -1 \\ 2 \end{bmatrix}\right) + T^{-1}\left(\begin{bmatrix} -1 \\ 3 \\ 1 \end{bmatrix}\right) \tag{104}$$

$$= T^{-1}\left(\begin{bmatrix} 3 \\ -1 \\ 2 \end{bmatrix} + \begin{bmatrix} -1 \\ 3 \\ 1 \end{bmatrix}\right) \tag{105}$$

$$= T^{-1}\left(\begin{bmatrix} 2 \\ 2 \\ 3 \end{bmatrix}\right) \tag{106}$$

$$= 2 + 2x + 3x^2 \tag{107}$$

Which transition requires linearity?

Multiple Choice:

1. *From expresion (**??**) to expresion (**??**)*
2. *From expresion (**??**) to expresion (**??**)*
3. *From expresion (**??**) to expresion (**??**)*
4. *From expresion (**??**) to expresion (**??**)*

Invertible linear transformations, such as transformation T of Exploration **??**, are useful because they preserve the structure of interactions between elements as we move back and forth between two vector spaces, allowing us to answer questions about one vector space in a different vector space. In particular, any question related to linear combinations can be addressed in this fashion. This includes questions concerning linear independence, span, basis and dimension.

Definition 100. Let V and W be vector spaces. If there exists an invertible linear transformation $T : V \to W$ we say that V and W are *isomorphic* and write $V \cong W$. The invertible linear transformation T is called an *isomorphism*.

It is worth pointing out that if $T : V \to W$ is an isomorphism, then $T^{-1} : W \to V$, being linear and invertible, is also an isomorphism.

Example 194. Our earlier discussion suggests that $\mathbb{L} \cong \mathbb{R}^2$. We postpone the proof until Theorem **??**.

Example 195. Exploration **??** shows that $\mathbb{P}^2 \cong \mathbb{R}^3$.

Example 196. Show that $\mathbb{M}_{2,2}$ and $\mathbb{P}^3$ are isomorphic.

Explanation. We will start by finding a plausible candidate for an isomorphism. Define $T : \mathbb{M}_{2,2} \to \mathbb{P}^3$ by

$$T\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = a + bx + cx^2 + dx^3$$

We will first show that T is a linear transformation, then verify that T is invertible.

$$\begin{aligned} T\left(k\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) &= T\left(\begin{bmatrix} ka & kb \\ kc & kd \end{bmatrix}\right) \\ &= ka + kbx + kcx^2 + kdx^3 = k(a + bx + cx^2 + dx^3) \\ &= kT\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) \end{aligned}$$

$$\begin{aligned} T\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix} + \begin{bmatrix} m & n \\ p & q \end{bmatrix}\right) &= T\left(\begin{bmatrix} a+m & b+n \\ c+p & d+q \end{bmatrix}\right) \\ &= (a+m) + (b+n)x + (c+p)x^2 + (d+q)x^3 \\ &= (a + bx + cx^2 + dx^3) + (m + nx + px^2 + qx^3) \\ &= T\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) + T\left(\begin{bmatrix} m & n \\ p & q \end{bmatrix}\right) \end{aligned}$$

We can show that T is one-to-one and onto, and therefore has an inverse. We can also observe directly that T^{-1} is given by

$$T^{-1}(a + bx + cx^2 + dx^3) = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

We conclude that T is an isomorphism, and $\mathbb{M}_{2,2} \cong \mathbb{P}^3$.

Isomorphism T in Example **??** establishes the fact that $\mathbb{M}_{2,2} \cong \mathbb{P}^3$. However, there is nothing special about T, as there are many other isomorphisms from $\mathbb{M}_{2,2}$ to $\mathbb{P}^3$. Just for fun, try to verify that each of the following is an isomorphism.

$$\tau_1 : \mathbb{M}_{2,2} \to \mathbb{P}^3 \tag{108}$$
$$\tau_1\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = a - bx - cx^2 + dx^3$$

$$\tau_2 : \mathbb{M}_{2,2} \to \mathbb{P}^3 \tag{109}$$
$$\tau_2\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = a + (a + c)x + (b - c)x^2 + dx^3$$

The Coordinate Vector Isomorphism

In Exploration **??** we made good use of a transformation that maps every element of $\mathbb{P}^2$ to its coordinate vector in $\mathbb{R}^3$. We observed that this transformation is linear and invertible, therefore it is an isomorphism. The following example generalizes this result.

Theorem 115. Let V be an n-dimensional vector space, and let $\mathcal{B}$ be an ordered basis for V. Then $T : V \to \mathbb{R}^n$ given by $T(\mathbf{v}) = [\mathbf{v}]_{\mathcal{B}}$ is an isomorphism.

Proof We leave the proof of this result to the reader. (See Practice Problem **??**.) ■

Properties of Isomorphic Vector Spaces and Isomorphisms

In this section we will illustrate properties of isomorphisms with specific examples. Formal proofs of properties will be presented in the next section.

Exploration 491 *In Exploration* **??** *we defined a transformation* $T : \mathbb{P}^2 \to \mathbb{R}^3$ *by* $T(a + bx + cx^2) = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$. *We later observed that* T *is an isomorphism. We will now examine the effect of* T *on two different bases of* $\mathbb{P}^2$.

Let $\mathcal{B}_1 = \{1, x, x^2\}$ *and* $\mathcal{B}_2 = \{x, 1 + x, x + x^2\}$. *(Recall that* $\mathcal{B}_2$ *is a basis of* $\mathbb{P}^2$ *by Example* **??** *of Bases and Dimension of Abstract Vector Spaces.)*

First,

$$T(\mathcal{B}_1) = \{T(1), T(x), T(x^2)\} = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}$$

Clearly, the images of the elements of $\mathcal{B}_1$ *form a basis of* $\mathbb{R}^3$.

Now we consider $\mathcal{B}_2$.

$$T(\mathcal{B}_2) = \{T(x), T(1 + x), T(x + x^2)\} = \left\{ \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \right\}$$

It is easy to verify that $\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$ *are linearly independent and span* $\mathbb{R}^3$, *therefore the images of the elements of* $\mathcal{B}_2$ *from a basis of* $\mathbb{R}^3$.

We can try any number of bases of $\mathbb{P}^2$ and we will find that the image of each basis of $\mathbb{P}^2$ is a basis of $\mathbb{R}^3$. In general, we have the following result:

Fact 1. An isomorphism maps a basis of the domain to a basis of the codomain. (We will state this result more formally as Theorem **??** in the next section.)

Isomorphisms preserve bases, but more generally, they preserve linear independence.

Fact 2. If $T : V \to W$ is an isomorphism, then the subset $\{\mathbf{v_1}, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ of V is linearly independent if and only if $\{T(\mathbf{v_1}), T(\mathbf{v}_2), \ldots, T(\mathbf{v}_n)\}$ is linearly independent in W. (We will state and prove this result as Theorem **??**.)

Example 197. Let V be a vector space, and let $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4\}$ be an ordered basis of V. Let

$$\mathbf{w}_1 = 2\mathbf{v}_1 + 3\mathbf{v}_2 - 5\mathbf{v}_3 - 2\mathbf{v}_4$$

$$\mathbf{w}_2 = -\mathbf{v}_1 + 4\mathbf{v}_2 - 3\mathbf{v}_3 + \mathbf{v}_4$$
$$\mathbf{w}_3 = -3\mathbf{v}_1 - \mathbf{v}_2 + 4\mathbf{v}_3 + 3\mathbf{v}_4$$

Are $\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3$ linearly independent?

Explanation. We could approach this question head-on by considering the vector equation

$$a\mathbf{w}_1 + b\mathbf{w}_2 + c\mathbf{w}_3 = \mathbf{0}$$

to see if the only solution is the trivial one. (See Practice Problem **??**.)

Instead, we will use isomorphisms. Observe that we do not know anything about V aside from the fact that it has four basis vectors. Vectors $\mathbf{w}_1$, $\mathbf{w}_2$, $\mathbf{w}_3$ are given in terms of these basis vectors. This should give us an idea for constructing an isomorphism between V and $\mathbb{R}^4$. Consider $T : V \to \mathbb{R}^4$ such that $T(\mathbf{w}) = [\mathbf{w}]_{\mathcal{B}}$. Then

$$T(\mathbf{w}_1) = [\mathbf{w}_1]_{\mathcal{B}} = \begin{bmatrix} 2 \\ 3 \\ -5 \\ -2 \end{bmatrix}, \quad T(\mathbf{w}_2) = [\mathbf{w}_2]_{\mathcal{B}} = \begin{bmatrix} -1 \\ 4 \\ -3 \\ 1 \end{bmatrix}, \quad T(\mathbf{w}_3) = [\mathbf{w}_3]_{\mathcal{B}} = \begin{bmatrix} -3 \\ -1 \\ 4 \\ 3 \end{bmatrix}$$

By Theorem **??**, T is an isomorphism. This means that $\mathbf{w}_1$, $\mathbf{w}_2$, $\mathbf{w}_3$ are linearly independent if and only if their coordinate vectors are linearly independent. There are multiple ways of determining whether

$$\begin{bmatrix} 2 \\ 3 \\ -5 \\ -2 \end{bmatrix}, \begin{bmatrix} -1 \\ 4 \\ -3 \\ 1 \end{bmatrix}, \begin{bmatrix} -3 \\ -1 \\ 4 \\ 3 \end{bmatrix}$$

are linearly independent. One way is to find the reduced row echelon form of

$$\begin{bmatrix} 2 & -1 & -3 \\ 3 & 4 & -1 \\ -5 & -3 & 4 \\ -2 & 1 & 3 \end{bmatrix}$$

The matrix reduces as follows:

$$\begin{bmatrix} 2 & -1 & -3 \\ 3 & 4 & -1 \\ -5 & -3 & 4 \\ -2 & 1 & 3 \end{bmatrix} \rightsquigarrow \begin{bmatrix} 1 & 0 & -13/11 \\ 0 & 1 & 7/11 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

We see that the rank of the matrix is 2. We conclude that the column vectors are not linearly independent. Thus, the vectors $\mathbf{w}_1$, $\mathbf{w}_2$ and $\mathbf{w}_3$ are not linearly independent.

Proofs of Isomorphism Properties

Recall that a transformation T is *one-to-one* provided that

$$T(\mathbf{v}_1) = T(\mathbf{v}_2)$$

implies that

$$\mathbf{v}_1 = \mathbf{v}_2$$

We will show that images of linearly independent vectors under one-to-one linear transformations are linearly independent.

Theorem 116. Let $T : V \to W$ be a one-to-one linear transformation. Suppose $\{\mathbf{v}_1, ..., \mathbf{v}_n\}$ is linearly independent in V. Then $\{T(\mathbf{v}_1), ..., T(\mathbf{v}_n)\}$ is linearly independent in W.

Proof Suppose $a_1, ..., a_n$ satisfy

$$a_1 T(\mathbf{v}_1) + ... + a_n T(\mathbf{v}_n) = \mathbf{0} \tag{110}$$

We will show that for each i, we must have $a_i = 0$.

By linearity, we have:

$$\begin{aligned} \mathbf{0} &= a_1 T(\mathbf{v}_1) + ... + a_n T(\mathbf{v}_n) \\ &= T(a_1\mathbf{v}_1) + ... + T(a_n\mathbf{v}_n) \\ &= T(a_1\mathbf{v}_1 + ... + a_n\mathbf{v}_n) \end{aligned}$$

By Theorem **??**, $T(\mathbf{0}) = \mathbf{0}$. Therefore,

$$T(a_1\mathbf{v}_1 + ... + a_n\mathbf{v}_n) = T(\mathbf{0})$$

Because T is one-to-one, we conclude that

$$a_1\mathbf{v}_1 + ... + a_n\mathbf{v}_n = \mathbf{0} \tag{111}$$

By assumption, $\{\mathbf{v}_1, ..., \mathbf{v}_n\}$ is linearly independent. Therefore $a_i = 0$ for $1 \leq i \leq n$. ■

Recall that a transformation T is *onto* provided that every vector of the codomain of T is the image of some vector in the domain of T.

We will show that an onto linear transformation maps sets that span the domain to sets that span the codomain.

Theorem 117. Let $T : V \to W$ be an onto linear transformation. Suppose $V = \text{span}(\mathbf{v}_1, \ldots, \mathbf{v}_n)$. Then $W = \text{span}(T(\mathbf{v}_1), \ldots, T(\mathbf{v}_n))$.

Proof Suppose $\mathbf{w}$ is an element of W. To show that $\{T(\mathbf{v}_1), \ldots, T(\mathbf{v}_n)\}$ spans W, we will express $\mathbf{w}$ as a linear combination of $T(\mathbf{v}_1), \ldots, T(\mathbf{v}_n)$.

Because T is onto, $\mathbf{w} = T(\mathbf{v})$ for some $\mathbf{v}$ in V. But $V = \text{span}(\mathbf{v}_1, \ldots, \mathbf{v}_n)$. Therefore, $\mathbf{v} = a_1\mathbf{v}_1 + \ldots + a_n\mathbf{v}_n$ for some scalar coefficients $a_1, \ldots, a_n$. By linearity, we have:

$$\begin{aligned} \mathbf{w} = T(\mathbf{v}) &= T(a_1\mathbf{v}_1 + \ldots + a_n\mathbf{v}_n) \\ &= a_1T(\mathbf{v}_1) + \ldots + a_nT(\mathbf{v}_n) \end{aligned}$$

Thus, $\mathbf{w}$ is in the span of $T(\mathbf{v}_1), \ldots, T(\mathbf{v}_n)$. ■

We will now combine the results of Theorem **??** and Theorem **??** to obtain a result about the effect of isomorphisms on a basis.

Theorem 118. Let $T : V \to W$ be an isomorphism. Suppose $\mathcal{B}_V = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is a basis for V. Then $\{T(\mathbf{v}_1), \ldots, T(\mathbf{v}_n)\}$ is a basis for W.

Proof Left to the reader. (See Practice Problem **??**) ■

Theorem 119. Suppose $T : V \to W$ is an isomorphism, then the subset $\{\mathbf{v_1}, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ of V is linearly independent if and only if $\{T(\mathbf{v_1}), T(\mathbf{v}_2), \ldots, T(\mathbf{v}_n)\}$ is linearly independent in W.

Proof We have already proved one direction of this this "if and only if" statement as Theorem **??**. To prove the other direction, suppose that $T(\mathbf{v_1}), T(\mathbf{v}_2), \ldots, T(\mathbf{v}_n)$ are linearly independent vectors in W. We need to show that this implies that $\mathbf{v_1}, \mathbf{v}_2, \ldots, \mathbf{v}_n$ are linearly independent in V. Observe that if T is an isomorphism, then $T^{-1} : W \to V$ is also an isomorphism. Thus, by Theorem **??**, $T^{-1}(T(\mathbf{v_1})), T^{-1}(T(\mathbf{v}_2)), \ldots, T^{-1}(T(\mathbf{v}_n))$ are linearly independent. But this means that $\mathbf{v_1}, \mathbf{v}_2, \ldots, \mathbf{v}_n$ are linearly independent. ■

Theorem 120. Let U, V and W be vector spaces. Suppose that $T_1 : U \to V$ and $T_2 : V \to W$ are isomorphisms. Then $T_2 \circ T_1 : U \to W$ is an isomorphism.

Proof The proof is left to the reader. (See Practice Problem **??**.) ■

Finite-dimensional Vector Spaces

Theorem 121. Let V and W be finite-dimensional vector spaces. Then

$$V \cong W \quad \text{if and only if} \quad \dim(V) = \dim(W)$$

Proof First, assume that $V \cong W$. Then there exists an isomorphism $T : V \to W$. Suppose $\dim(V) = n$ and let $\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n\}$ be a basis for V. By Theorem **??** $\{T(\mathbf{v}_1), ..., T(\mathbf{v}_n)\}$ is a basis for W. Therefore $\dim(W) = n$.

Conversely, suppose $\dim(V) = \dim(W) = n$, and let $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n\}$, $\mathcal{C} = \{\mathbf{w}_1, \mathbf{w}_2, ..., \mathbf{w}_n\}$ be bases for V and W, respectively.

Define a linear transformation $T : V \to W$ by $T(\mathbf{v}_i) = \mathbf{w}_i$ for $1 \leq i \leq n$. To show that T is an isomorphism, we need to prove that T is one-to-one and onto.

Suppose $T(\mathbf{u}_1) = T(\mathbf{u}_2)$ for some vectors $\mathbf{u}_1$, $\mathbf{u}_2$ in V. We know that

$$\mathbf{u}_1 = a_1\mathbf{v}_1 + ... + a_n\mathbf{v}_n$$

$$\mathbf{u}_2 = b_1\mathbf{v}_1 + ... + b_n\mathbf{v}_n$$

for some scalars a_i's and b_i's. Thus,

$$T(a_1\mathbf{v}_1 + ... + a_n\mathbf{v}_n) = T(b_1\mathbf{v}_1 + ... + b_n\mathbf{v}_n)$$

By linearity of T,

$$a_1\mathbf{w}_1 + ... + a_n\mathbf{w}_n = b_1\mathbf{w}_1 + ... + b_n\mathbf{w}_n$$

$$(a_1 - b_1)\mathbf{w}_1 + ... + (a_n - b_n)\mathbf{w}_n = \mathbf{0}$$

But $\mathbf{w}_1, \mathbf{w}_2, ..., \mathbf{w}_n$ are linearly independent, so $a_i - b_i = 0$ for all $1 \leq i \leq n$. Therefore $a_i = b_i$ for all $1 \leq i \leq n$. We conclude that $\mathbf{u}_1 = \mathbf{u}_2$.

We now show that T is onto. Suppose that $\mathbf{w}$ is an element of W. Then $\mathbf{w} = c_1\mathbf{w}_1 + ... + c_n\mathbf{w}_n$ for some scalars c_i's. But then

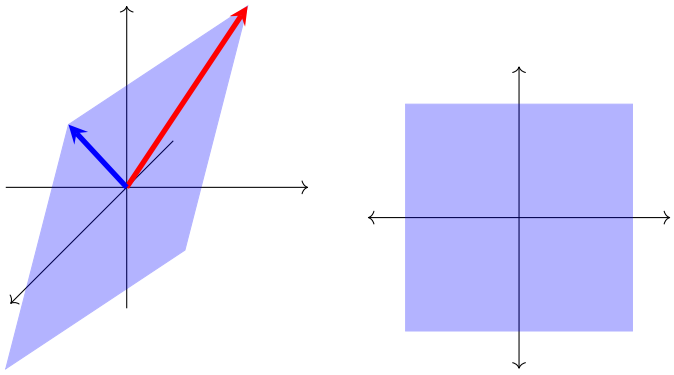
$$\mathbf{w} = c_1 T(\mathbf{v}_1) + ... + c_n T(\mathbf{v}_n) = T(c_1\mathbf{v}_1 + ... + c_n\mathbf{v}_n)$$

We conclude that $\mathbf{w}$ is an image of an element of V, so T is onto. ■

From this theorem follows an important corollary that shows why we spent so much time trying to understand $\mathbb{R}^n$ in this course.

Corollary 14. Every n-dimensional vector space is isomorphic to $\mathbb{R}^n$.

Example 198. The span of any two linearly independent vectors in $\mathbb{R}^3$ is isomorphic to $\mathbb{R}^2$.



Example 199. $\mathbb{P}^2 \ncong \mathbb{R}^2$

Explanation. Recall that $\dim(\mathbb{P}^2) = 3$. Since $\dim(\mathbb{P}^2) = 3 \neq 2 = \dim(\mathbb{R}^2)$, we conclude that $\mathbb{P}^2$ is not isomorphic to $\mathbb{R}^2$.

Practice Problems

Problem 492 *Prove that transformation T of Exploration* **??** *is one-to-one and onto.*

Problem 493 *Verify that*

$$\tau_1 : \mathbb{M}_{2,2} \to \mathbb{P}^3$$

given by

$$\tau_1\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = a - bx - cx^2 + dx^3$$

of Expression **??** *is an isomorphism.*

Problem 494 *Do Example* **??** *without using isomorphisms.*

Problem 495 *Let*

$$\mathcal{S} = \left\{ \begin{bmatrix} 1 & -3 \\ -2 & 2 \end{bmatrix}, \begin{bmatrix} 4 & -2 \\ 1 & 5 \end{bmatrix}, \begin{bmatrix} 5 & 5 \\ 8 & 4 \end{bmatrix} \right\}$$

Is $\mathcal{S}$ linearly independent in $\mathbb{M}_{2,2}$? (Yes/ No)

Problem 496 *Let*

$$\mathcal{S} = \left\{ \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}, \begin{bmatrix} 5 & 6 \\ 7 & 8 \end{bmatrix}, \begin{bmatrix} 9 & 10 \\ 11 & 12 \end{bmatrix}, \begin{bmatrix} 13 & 14 \\ 15 & 16 \end{bmatrix} \right\}$$

Is $\mathcal{S}$ a basis for $\mathbb{M}_{2,2}$? (Yes/ No)

Problem 497 *Prove Theorem* **??**.

Problem 498 *Let V be a vector space, and suppose $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4, \mathbf{v}_5\}$ is a basis for V. What can we conclude about V? Check ALL that apply.*

Select All Correct Answers:

1. *We cannot conclude anything about V because we don't know what V is.*
2. $V \cong \mathbb{R}^5$
3. $V \cong \mathbb{P}^5$
4. $V \cong \mathbb{R}^4$
5. $V \cong \mathbb{P}^4$
6. $V \cong \mathbb{M}_{5,5}$

Problem 499 *Which of the followng statements are true? Check ALL that apply.*

Select All Correct Answers:

1. $\mathbb{M}_{3,3} \cong \mathbb{R}^6$
2. $\mathbb{M}_{3,3} \cong \mathbb{R}^9$
3. $\mathbb{M}_{4,4} \cong \mathbb{P}^{16}$
4. $\mathbb{M}_{4,4} \cong \mathbb{P}^{15}$
5. $\mathbb{L} \cong \mathbb{R}^2$

Problem 500 *Verify that $T : V \to \mathbb{R}^n$ of Theorem* **??** *is an isomorphism.*

Problem 501 *Prove that the composition of two isomorphisms is an isomorphism. (Theorem* **??***.)*

75 Matrices of Linear Transformations with Respect to Arbitrary Bases

We know that every linear transformation from $\mathbb{R}^n$ into $\mathbb{R}^m$ is a matrix transformation (Theorem **??** of Standard Matrix of a Linear Transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$). What about linear transformations between vector spaces other than $\mathbb{R}^n$? In this section we will learn to represent linear transformations between arbitrary finite-dimensional vector spaces using matrices. To do so, we will use the fact that every n-dimensional vector space is isomorphic to $\mathbb{R}^n$ (Corollary **??** of Isomorphic Vector Spaces). What we do here will serve as yet another example of how isomorphisms can be used to translate problems in one vector space to another, more convenient, vector space.

Exploration 502 *In Isomorphic Vector Spaces we introduced the transformation*

$$\tau_2 : \mathbb{M}_{2,2} \to \mathbb{P}^3$$

defined by

$$\tau_2\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = a + (a+c)x + (b-c)x^2 + dx^3$$

*You should verify that τ_2 is linear. (See Practice Problem **??**.)*

We will examine τ_2 in an effort to find a way to represent it with a matrix. (In the process, we will also end up proving that τ_2 is an isomorphism, which is what you were challenged to do in Isomorphic Vector Spaces.)

We will start by selecting a basis for each of $\mathbb{M}_{2,2}$ and $\mathbb{P}^3$. We can choose any basis for either space, but we will choose bases that will make computations easier.

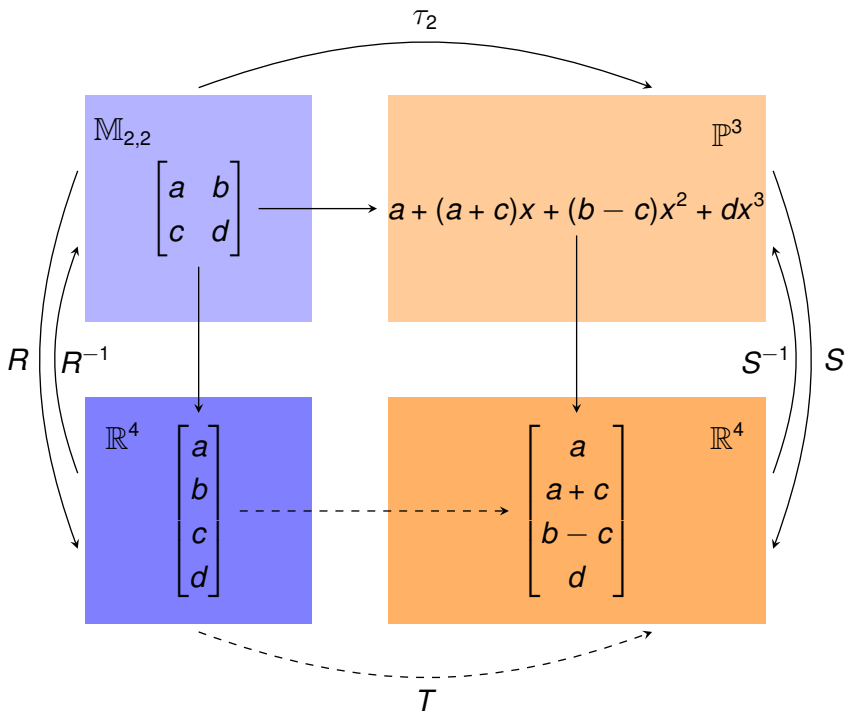
Let

$$\mathcal{B} = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\}$$

$$\mathcal{C} = \{1, x, x^2, x^3\}$$

be our ordered bases or choice for $\mathbb{M}_{2,2}$ and $\mathbb{P}^3$, respectively.

*Recall that a coordinate vector isomorphism maps a vector to its coordinate vector with respect to the given ordered basis (Theorem **??** of Isomorphic Vector Spaces). In the diagram below, let R and S be coordinate vector isomorphisms with respect to $\mathcal{B}$ and $\mathcal{C}$.*



Define $T : \mathbb{R}^4 \to \mathbb{R}^4$ *by*

$$T\left(\begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix}\right) = \begin{bmatrix} a \\ a + c \\ b - c \\ d \end{bmatrix}$$

Observe that $T = S \circ \tau_2 \circ R^{-1}$*. Because* T *is a composition of linear transformations,* T *itself is linear (Theorem* **??** *of Composition and Inverses of Linear Transformations). Thus, we should be able to find the standard matrix for* T*. To do this, find the images of the standard unit vectors and use them to create the standard matrix* A *for* T*.*

$$A = \begin{bmatrix} \ldots & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots \end{bmatrix}$$

We say that A *is the matrix of* τ_2 *with respect to ordered bases* $\mathcal{B}$ *and* $\mathcal{C}$*.*

As a side-note, observe that $\tau_2 = S^{-1} \circ T \circ R$*. Observe also that* T *is invertible because* A *is invertible. So,* T *is an isomorphism. As a composition of isomorphisms,* τ_2 *is an isomorphism*

*(Theorem **??** of Isomorphic Vector Spaces). While we could have proved this result directly, as you were challenged to do in Isomorphic Vector Spaces, this approach is much less tedious.*

Exploration 503 *Let*

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix} \quad \text{and} \quad \mathbf{v}_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$$

$$\mathbf{w}_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \quad \text{and} \quad \mathbf{w}_2 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

*In Example **??** we defined V and W as follows:*

$$V = \text{span}(\mathbf{v}_1, \mathbf{v}_2) \quad \text{and} \quad W = \text{span}(\mathbf{w}_1, \mathbf{w}_2)$$

Geometrically, V and W are planes in $\mathbb{R}^3$. We chose

$$\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2\} \quad \text{and} \quad \mathcal{C} = \{\mathbf{w}_1, \mathbf{w}_2\}$$

as ordered bases of V and W, respectively.

We also defined a linear transformation $T : V \to W$ by

$$T(\mathbf{v}_1) = 2\mathbf{w}_1 - 3\mathbf{w}_2 \quad \text{and} \quad T(\mathbf{v}_2) = -\mathbf{w}_1 + 4\mathbf{w}_2$$

Our goal now is to find a matrix for T with respect to $\mathcal{B}$ and $\mathcal{C}$.

*The information given in this problem is slightly different from the information in Exploration **??**. Instead of being given an expression for the image of a generic vector of V, we are only given the images of the two basis vectors of V. But this information is sufficient to determine the linear transformation.*

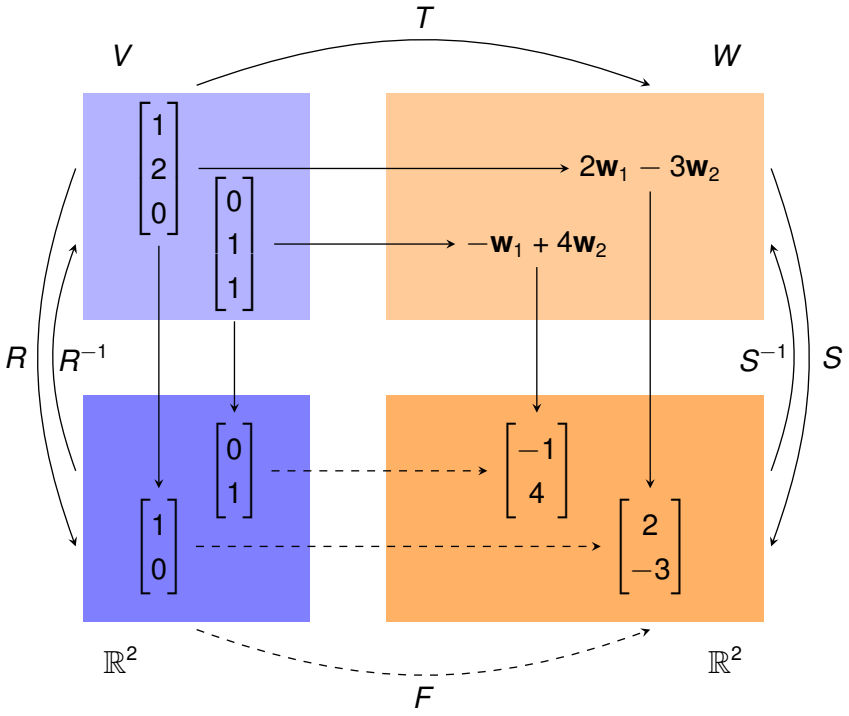
As before, we will map vectors of V and W to their coordinate vectors. Where are the coordinate vectors located?

1. *$[\mathbf{v}_1]_{\mathcal{B}}$ and $[\mathbf{v}_2]_{\mathcal{B}}$ are elements of ($\mathbb{R}$, / $\mathbb{R}^2$, / $\mathbb{R}^3$)*
2. *$[\mathbf{w}_1]_{\mathcal{C}}$ and $[\mathbf{w}_2]_{\mathcal{C}}$ are elements of ($\mathbb{R}$, / $\mathbb{R}^2$, / $\mathbb{R}^3$)*

Now we find the coordinate vectors.

1. $$[\mathbf{v}_1]_{\mathcal{B}} = \begin{bmatrix} \cdots \\ \cdots \end{bmatrix} \quad \text{and} \quad [\mathbf{v}_2]_{\mathcal{B}} = \begin{bmatrix} \cdots \\ \cdots \end{bmatrix}$$

2. $$[T(\mathbf{v}_1)]_{\mathcal{C}} = \begin{bmatrix} \cdots \\ \cdots \end{bmatrix} \quad \text{and} \quad [T(\mathbf{v}_2)]_{\mathcal{C}} = \begin{bmatrix} \cdots \\ \cdots \end{bmatrix}$$



Here is a diagram that summarizes this information. (Press the arrow on the right to expand.)

Define $F : \mathbb{R}^2 \to \mathbb{R}^2$ *by* $F = S \circ T \circ R^{-1}$. *F is a linear transformation that maps* $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ *and* $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$ *to* $\begin{bmatrix} 2 \\ -3 \end{bmatrix}$ *and* $\begin{bmatrix} -1 \\ 4 \end{bmatrix}$, *respectively. Thus, the standard matrix for F is:*

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

We say that A is a matrix for T with respect to $\mathcal{B}$ *and* $\mathcal{C}$.

Let's take a look at what this matrix can do for us. Recall that in Example **??** *we found that the image of* $\mathbf{v} = 2\mathbf{v}_1 + \mathbf{v}_2$ *is*

$$T(\mathbf{v}) = T(2\mathbf{v}_1 + \mathbf{v}_2) = 2T(\mathbf{v}_1) + T(\mathbf{v}_2) = 3\mathbf{w}_1 - 2\mathbf{w}_2$$

This result can be obtained by finding the product of A with the coordinate vector of $\mathbf{v}$*.*

$$\begin{bmatrix} 2 & -1 \\ -3 & 4 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ -2 \end{bmatrix}$$

Given any vector $\mathbf{u} = a\mathbf{v}_1 + b\mathbf{v}_2$ *of V, we can find* $T(\mathbf{u})$ *as follows:*

$$\begin{bmatrix} 2 & -1 \\ -3 & 4 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 2a - b \\ -3a + 4b \end{bmatrix}$$

This gives us

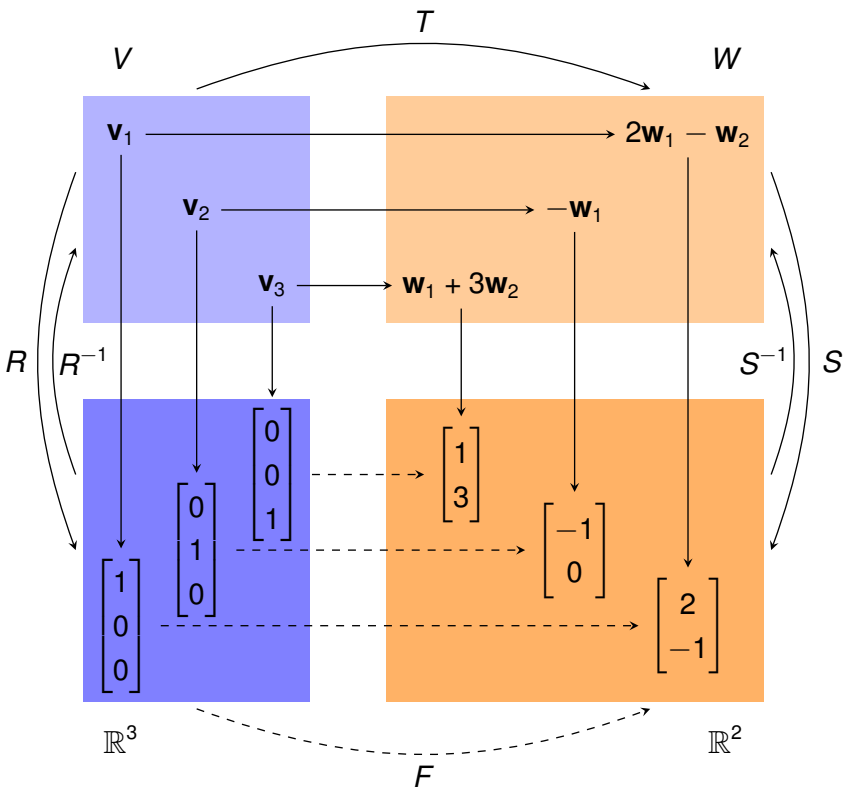
$$T(\mathbf{u}) = T(a\mathbf{v}_1 + b\mathbf{v}_2) = (2a - b)\mathbf{w}_1 + (-3a + 4b)\mathbf{w}_2$$

Example 200. Let V and W be vector spaces with ordered bases $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ and $\mathcal{C} = \{\mathbf{w}_1, \mathbf{w}_2\}$, respectively. Define a linear transformation $T : V \to W$ by

$$T(\mathbf{v}_1) = 2\mathbf{w}_1 - \mathbf{w}_2, \quad T(\mathbf{v}_2) = -\mathbf{w}_1, \quad T(\mathbf{v}_3) = \mathbf{w}_1 + 3\mathbf{w}_2$$

Find the matrix of T with respect to $\mathcal{B}$ and $\mathcal{C}$, and use it to find $T(2\mathbf{v}_1 - \mathbf{v}_2 + 3\mathbf{v}_3)$. Verify your answer by computing $T(2\mathbf{v}_1 - \mathbf{v}_2 + 3\mathbf{v}_3)$ directly.

Explanation. We start with a diagram:



Looking at the images of the standard unit vectors in $\mathbb{R}^3$, we can construct the standard matrix A of T with respect to $\mathcal{B}$ and $\mathcal{C}$.

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

Applying this matrix to the coordinate vector of $2\mathbf{v}_1 - \mathbf{v}_2 + 3\mathbf{v}_3$ we get

$$\begin{bmatrix} 2 & -1 & 1 \\ -1 & 0 & 3 \end{bmatrix} \begin{bmatrix} 2 \\ -1 \\ 3 \end{bmatrix} = \begin{bmatrix} 8 \\ 7 \end{bmatrix}$$

This means that $T(2\mathbf{v}_1 - \mathbf{v}_2 + 3\mathbf{v}_3) = 8\mathbf{w}_1 + 7\mathbf{w}_2$. We can verify this by direct computation as follows:

$$\begin{aligned} T(2\mathbf{v}_1 - \mathbf{v}_2 + 3\mathbf{v}_3) &= 2T(\mathbf{v}_1) - T(\mathbf{v}_2) + 3T(\mathbf{v}_3) \\ &= 2(2\mathbf{w}_1 - \mathbf{w}_2) - (-\mathbf{w}_1) + 3(\mathbf{w}_1 + 3\mathbf{w}_2) \\ &= 8\mathbf{w}_1 + 7\mathbf{w}_2 \end{aligned}$$

The Matrix of a Linear Transformation

In this section we will formalize the process for finding the matrix of a linear transformation with respect to arbitrary bases that we established through earlier examples.

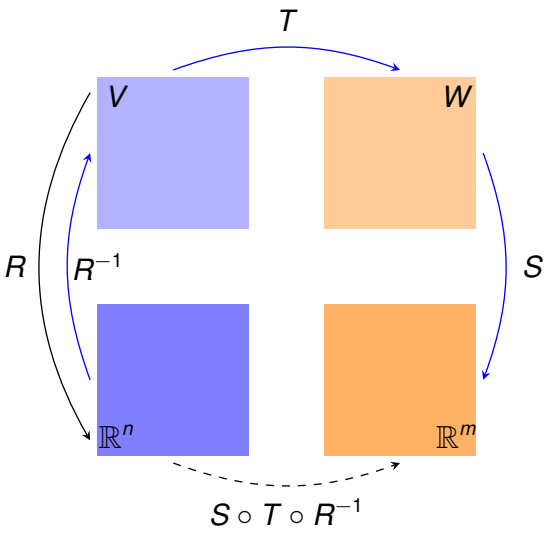
Let V and W be vector spaces with ordered bases $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ and $\mathcal{C} = \{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_m\}$, respectively. Suppose $T : V \to W$ is a linear transformation. Our goal is to find a matrix for T with respect to $\mathcal{B}$ and $\mathcal{C}$.

Observe that $\dim(V) = n = \dim(\mathbb{R}^n)$ and $\dim(W) = m = \dim(\mathbb{R}^m)$. Let $R : V \to \mathbb{R}^n$ and $S : W \to \mathbb{R}^m$ be coordinate vector isomorphisms defined by

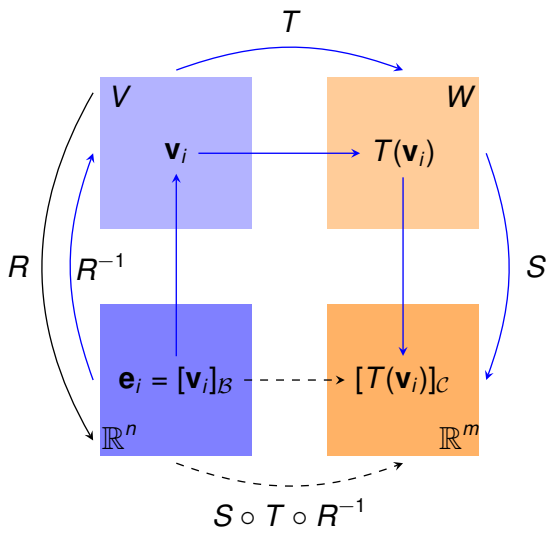
$$R(\mathbf{v}) = [\mathbf{v}]_{\mathcal{B}} \quad \text{and} \quad S(\mathbf{w}) = [\mathbf{w}]_{\mathcal{C}}$$

We know that R is an isomorphism and R^{-1} exists. Consider the transformation

$$S \circ T \circ R^{-1} : \mathbb{R}^n \to \mathbb{R}^m$$



As a composition of linear transformation, $S \circ T \circ R^{-1}$ is linear and thus has a standard matrix. To find it, we need to determine the images of standard unit vectors $\mathbf{e}_i$ under $S \circ T \circ R^{-1}$. We have the following:



Vectors $[T(\mathbf{v}_i)]_{\mathcal{C}}$ will become the columns of the standard matrix. We summarize this discussion as a theorem.

Theorem 122. Let V and W be finite-dimensional vector spaces with ordered bases $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ and $\mathcal{C}$, respectively. Suppose $T : V \to W$ is a linear transformation.

$$\text{Let} \quad A = \begin{bmatrix} | & | & & | \\ [T(\mathbf{v}_1)]_{\mathcal{C}} & [T(\mathbf{v}_2)]_{\mathcal{C}} & \cdots & [T(\mathbf{v}_n)]_{\mathcal{C}} \\ | & | & & | \end{bmatrix}$$

Then $A[\mathbf{v}]_{\mathcal{B}} = [T(\mathbf{v})]_{\mathcal{C}}$ for all vectors $\mathbf{v}$ in V.

Definition 101. Matrix A of Theorem **??** is called the matrix of T with respect to ordered bases $\mathcal{B}$ and $\mathcal{C}$.

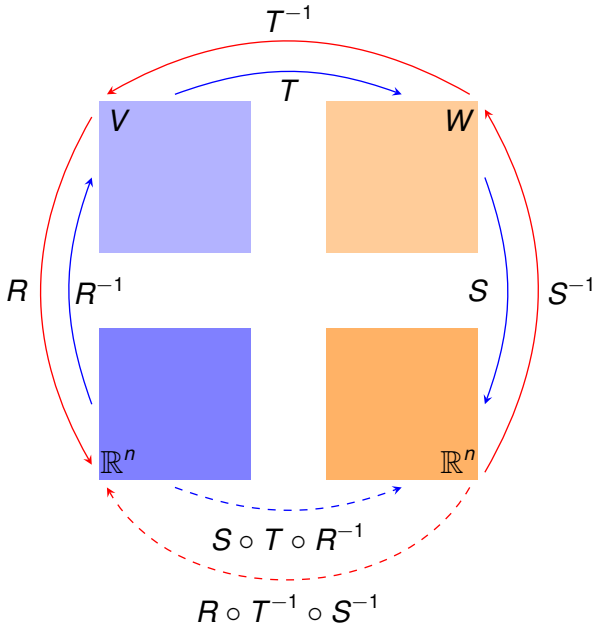
In conclusion, observe how isomorphisms helped us solve the matrix of a linear transformation problem. The coordinate mappings R and S are isomorphisms. This means that V and $\mathbb{R}^n$ are isomorphic and have the same structural properties. The same is true for W and $\mathbb{R}^m$. In this abstract discussion, we do not know anything about the elements of V and W, but isomorphisms allow us to take a problem that we do not know much about and transform it to a familiar problem involving familiar spaces.

The Inverse of a Linear Transformation and its Matrix

Let V and W be vector spaces. Suppose $T : V \to W$ is an invertible linear transformation. This, of course, means that T is an isomorphism, which means that

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

Let $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ and $\mathcal{C} = \{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_n\}$ be ordered bases of V and W, respectively. We can find the matrix of T^{-1} with respect to $\mathcal{C}$ and $\mathcal{B}$ by finding the standard matrix of the linear transformation $R \circ T^{-1} \circ S^{-1} : \mathbb{R}^n \to \mathbb{R}^n$.



Observe that $R \circ T^{-1} \circ S^{-1}$ is the inverse of $S \circ T \circ R^{-1}$. So, if A is the standard matrix of $S \circ T \circ R^{-1}$, then A^{-1} is the standard matrix of $R \circ T^{-1} \circ S^{-1}$. Thus, A^{-1} is the matrix of T^{-1} with respect to $\mathcal{C}$ and $\mathcal{B}$.

Example 201. In Exploration Problem **??** of Composition and Inverses of Linear Transformations, we introduced the following set up.

Let

$$V = \text{span}\left(\begin{bmatrix}1\\0\\0\end{bmatrix}, \begin{bmatrix}1\\1\\1\end{bmatrix}\right)$$

Define a linear transformation

$$T : V \to \mathbb{R}^2$$

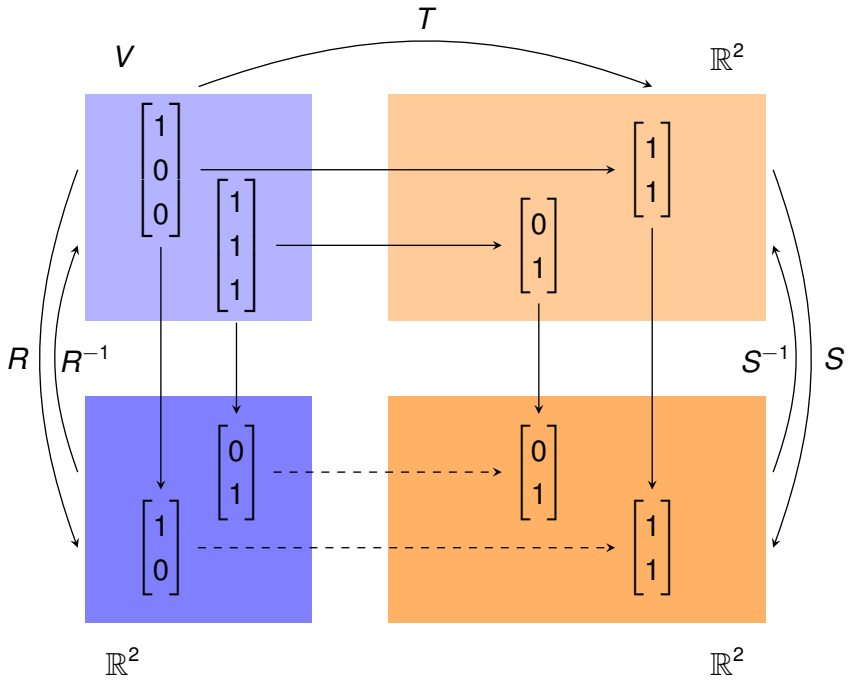
by

$$T\left(\begin{bmatrix}1\\0\\0\end{bmatrix}\right) = \begin{bmatrix}1\\1\end{bmatrix} \quad \text{and} \quad T\left(\begin{bmatrix}1\\1\\1\end{bmatrix}\right) = \begin{bmatrix}0\\1\end{bmatrix}$$

In Example **??** of Existence of the Inverse of a Linear Transformation, we proved that T is invertible.

Find the matrix of T^{-1} with respect to ordered bases $\mathcal{C} = \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}$ and $\mathcal{B} = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right\}$

Explanation. Consider the diagram:



This gives us the matrix of T with respect to $\mathcal{B}$ and $\mathcal{C}$:

$$A = \begin{bmatrix} \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \end{bmatrix}$$

We now find A^{-1}.

$$A^{-1} = \begin{bmatrix} \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \end{bmatrix}$$

A^{-1} is the matrix of T^{-1} with respect to $\mathcal{C}$ and $\mathcal{B}$.

Practice Problems

Problem 504 *Verify that τ_2 of Exploration* **??** *is linear.*

Problems **??**-**??**

Define $T : \mathbb{M}_{2,2} \to \mathbb{L}$ by

$$T\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = (a+b) + (c+d)x$$

Problem 505 *Show that T is linear.*

Problem 506 *Is T an isomorphism? (Yes, / No)*

Problem 507 *Let*

$$\mathcal{B} = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\}$$

$$\mathcal{C} = \{1, x\}$$

be ordered bases for $\mathbb{M}_{2,2}$ and $\mathbb{L}$, respectively.

Find the matrix for T with respect to $\mathcal{B}$ and $\mathcal{C}$.

Answer:

$$\begin{bmatrix} \ldots\ldots & \ldots\ldots & \ldots\ldots & \ldots\ldots \\ \ldots\ldots & \ldots\ldots & \ldots\ldots & \ldots\ldots \end{bmatrix}$$

Problem 508 *Let V and W be vector spaces with ordered bases $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2\}$ and $\mathcal{C} = \{\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3\}$, respectively. Define a linear transformation $T : V \to W$ by*

$$T(\mathbf{v}_1) = \mathbf{w}_1 + 2\mathbf{w}_2 - \mathbf{w}_3, \quad T(\mathbf{v}_2) = 3\mathbf{w}_2 + \mathbf{w}_3$$

Find the matrix A of T with respect to $\mathcal{B}$ and $\mathcal{C}$, and use it to find $T(-\mathbf{v}_1 - 3\mathbf{v}_2)$. Verify your answer by computing $T(-\mathbf{v}_1 - 3\mathbf{v}_2)$ directly.

$$A = \begin{bmatrix} \ldots\ldots & \ldots\ldots \\ \ldots\ldots & \ldots\ldots \\ \ldots\ldots & \ldots\ldots \end{bmatrix}$$

$$T(-\mathbf{v}_1 - 3\mathbf{v}_2) = \ldots\ldots\mathbf{w}_1 + \ldots\ldots\mathbf{w}_2 + \ldots\ldots\mathbf{w}_3$$

Problems **??-??**

Let

$$\mathbf{v}_1 = \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix} \quad \text{and} \quad \mathbf{v}_2 = \begin{bmatrix} 3 \\ 1 \\ 1 \end{bmatrix}$$

$$\mathbf{w}_1 = \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix} \quad \text{and} \quad \mathbf{w}_2 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$$

Let V and W be subspaces of $\mathbb{R}^3$ with ordered bases

$$\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2\} \quad \text{and} \quad \mathcal{C} = \{\mathbf{w}_1, \mathbf{w}_2\}$$

respectively.

Let $T : V \to W$ be a linear transformation such that

$$T(\mathbf{v}_1) = \begin{bmatrix} -1 \\ 3 \\ 2 \end{bmatrix} \quad \text{and} \quad T(\mathbf{v}_2) = \begin{bmatrix} 2 \\ 4 \\ 1 \end{bmatrix}$$

Problem 509 *Show that* $\begin{bmatrix} -1 \\ 3 \\ 2 \end{bmatrix}, \begin{bmatrix} 2 \\ 4 \\ 1 \end{bmatrix}$ *lie in W by expressing them as linear combinations of* $\mathbf{w}_1$ *and* $\mathbf{w}_2$.

$$\begin{bmatrix} -1 \\ 3 \\ 2 \end{bmatrix} = \ldots\ldots\mathbf{w}_1 + \ldots\ldots\mathbf{w}_2$$

$$\begin{bmatrix} 2 \\ 4 \\ 1 \end{bmatrix} = \ldots\ldots\mathbf{w}_1 + \ldots\ldots\mathbf{w}_2$$

Problem 510 *Find the matrix of T with respect to* $\mathcal{B}$ *and* $\mathcal{C}$.

$$\begin{bmatrix} \ldots\ldots & \ldots\ldots \\ \ldots\ldots & \ldots\ldots \end{bmatrix}$$

Problem 511 *Find the matrix of T^{-1} with respect to $\mathcal{C}$ and $\mathcal{B}$.*

$$\begin{bmatrix} \ldots\ldots & \ldots\ldots \\ \ldots\ldots & \ldots\ldots \end{bmatrix}$$

Problem 512 *Find* $T^{-1}\left(\begin{bmatrix} 10 \\ 10 \\ 0 \end{bmatrix}\right)$.

$$T^{-1}\left(\begin{bmatrix} 10 \\ 10 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} \ldots\ldots \\ \ldots\ldots \\ \ldots\ldots \end{bmatrix}$$

76 Applications

77 Applications

One of the beautiful aspects of linear algebra is its wide variety of applications, and it is our belief that a first course in linear algebra should include exposure to many of them. In this chapter we have six different applications, but there are so many others that may catch the interest of a particular student. Here we list a few resources where you can find other applications.

1. Feryal Alayont and Steven Schlicker, *Linear Algebra and Applications: An Inquiry-Based Approach*, Scholarworks at GVSU, 2019, Open Edition, available at https://scholarworks.gvsu.edu/books/2

   In this free book, each chapter opens with an application that is tackled later in the chapter. This makes for 37 different applications of elementary linear algebra!

2. Jim Hefferon, *Linear Algebra* , freely available at https://hefferon.net/linearalgebra/

   In this free book, each chapter contains 4-5 applications after covering the other content. This makes for 23 different applications of elementary linear algebra!

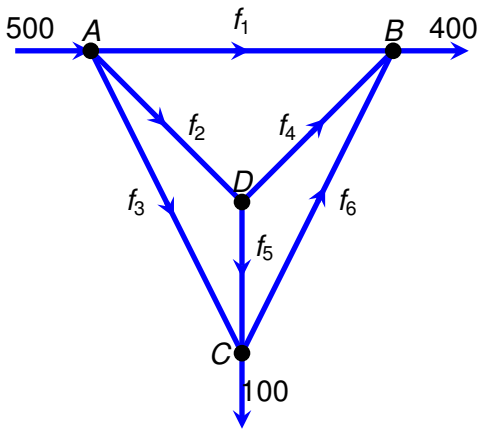
78 Application to Network Flow

There are many types of problems that concern a network of conductors along which some sort of flow is observed. Examples of these include an irrigation network and a network of streets or freeways. There are often points in the system at which a net flow either enters or leaves the system. The basic principle behind the analysis of such systems is that the total flow into the system must equal the total flow out. In fact, we apply this principle at every junction in the system.

Theorem 123 (Junction Rule)**.** At each of the junctions in the network, the total flow into that junction must equal the total flow out.

This requirement gives a linear equation relating the flows in conductors emanating from the junction.

Example 202. A network of one-way streets is shown in the accompanying diagram. The rate of flow of cars into intersection A is 500 cars per hour, and 400 and 100 cars per hour emerge from B and C, respectively. Find the possible flows along each street.



Explanation. Suppose the flows along the streets are f_1, f_2, f_3, f_4, f_5, and f_6 cars per hour in the directions shown.

Then, equating the flow in with the flow out at each intersection, we get

$$\begin{array}{lrl} \text{Intersection } A & 500 & = f_1 + f_2 + f_3 \\ \text{Intersection } B & f_1 + f_4 + f_6 & = 400 \\ \text{Intersection } C & f_3 + f_5 & = f_6 + 100 \\ \text{Intersection } D & f_2 & = f_4 + f_5 \end{array}$$

These give four equations in the six variables $f_1, f_2, \ldots, f_6$.

$$\begin{array}{rcrcrcrcrcr} f_1 & + & f_2 & + & f_3 & & & & & = & 500 \\ f_1 & & & & & + & f_4 & & f_6 & = & 400 \\ & & & & f_3 & & & + & f_5 - f_6 & = & 100 \\ & & f_2 & & & - & f_4 & - & f_5 & = & 0 \end{array}$$

The reduction of the augmented matrix is

$$\left[\begin{array}{cccccc|c} 1 & 1 & 1 & 0 & 0 & 0 & 500 \\ 1 & 0 & 0 & 1 & 0 & 1 & 400 \\ 0 & 0 & 1 & 0 & 1 & -1 & 100 \\ 0 & 1 & 0 & -1 & -1 & 0 & 0 \end{array}\right] \rightarrow \left[\begin{array}{cccccc|c} 1 & 0 & 0 & 1 & 0 & 1 & 400 \\ 0 & 1 & 0 & -1 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & -1 & 100 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

Hence, when we use f_4, f_5, and f_6 as parameters, the general solution is

$$f_1 = 400 - f_4 - f_6 \qquad f_2 = f_4 + f_5 \qquad f_3 = 100 - f_5 + f_6$$

This gives all solutions to the system of equations and hence all the possible flows.

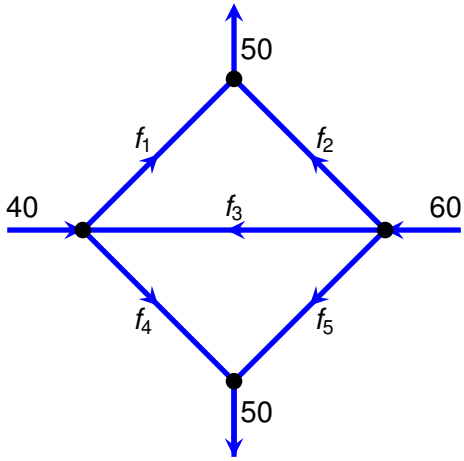
Of course, not all these solutions may be acceptable in the real situation. For example, the flows $f_1, f_2, \ldots, f_6$ are all *positive* in the present context (if one came out negative, it would mean traffic flowed in the opposite direction). This imposes constraints on the flows: $f_1 \geq 0$ and $f_3 \geq 0$ become

$$f_4 + f_6 \leq 400 \qquad f_5 - f_6 \leq 100$$

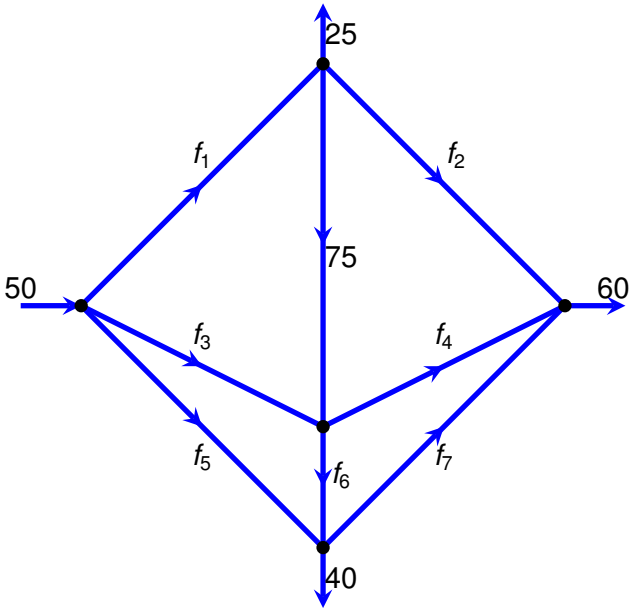
Further constraints might be imposed by insisting on maximum values on the flow in each street.

Practice Problems

Problem 513 *Find the possible flows in each of the following networks of pipes.*



1.



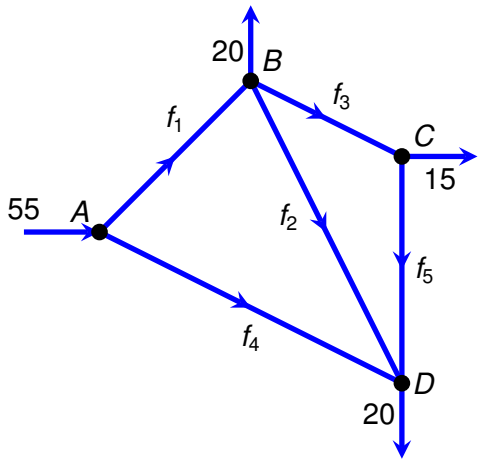
2.

Click the arrow to see the answer.

$$
\begin{aligned}
f_1 &= 85 - f_4 - f_7 \\
f_2 &= 60 - f_4 - f_7 \\
f_3 &= -75 + f_4 + f_6 \\
f_5 &= 40 - f_6 - f_7
\end{aligned}
$$

f_4, f_6, f_7 parameters

Problem 514 *A proposed network of irrigation canals is described in the accompanying diagram. At peak demand, the flows at interchanges A, B, C, and D are as shown.*



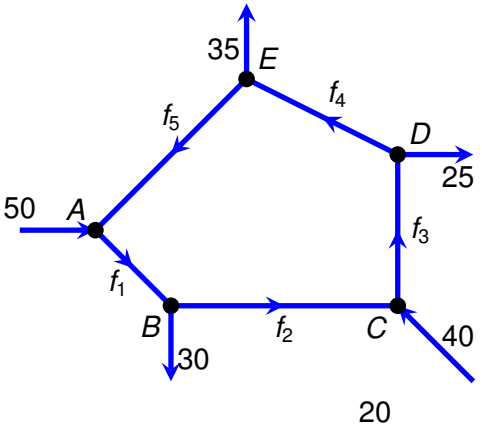
1. *Find the possible flows.*
2. *If canal BC is closed, what range of flow on AD must be maintained so that no canal carries a flow of more than 30?*

   Click the arrow to see the answer.

   $f_5 = 15$

   $25 \leq f_4 \leq 30$

Problem 515 *A traffic circle has five one-way streets, and vehicles enter and leave as shown in the accompanying diagram.*



1. *Compute the possible flows.*
2. *Which road has the heaviest flow?*

Multiple Choice:

1. *AB*
2. *BC*
3. *CD*
4. *DE*
5. *EA*

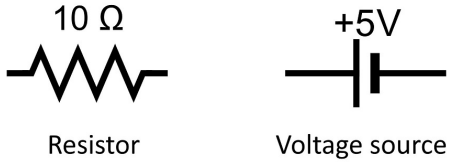
Text Source

This application was adapted from Section 1.4 of Keith Nicholson's *Linear Algebra with Applications*. (CC-BY-NC-SA)

W. Keith Nicholson, *Linear Algebra with Applications*, Lyryx 2018, Open Edition, p. 27

79 Application to Electrical Networks

In an electrical network it is often necessary to find the current in amperes (A) flowing in various parts of the network. These networks usually contain resistors that retard the current. Resistance is measured in ohms (Ω). The current is increased at various points by voltage sources (for example, a battery). The voltage of these sources is measured in volts (V). We use the following symbols to represent resistors and voltage sources:



We assume these voltage sources have no resistance. The flow of current is governed by the following principles.

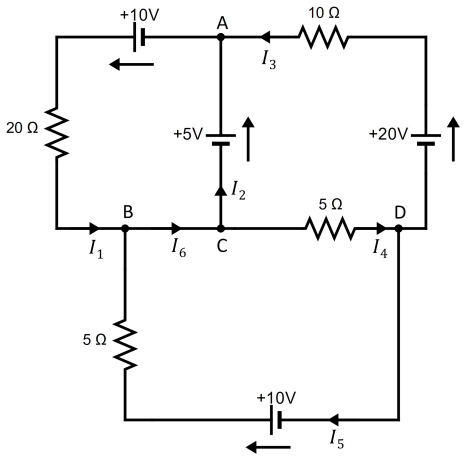
Theorem 124 (Ohm's Law)**.** The current I and the voltage drop V across a resistance R are related by the equation $V = RI$.

Theorem 125 (Kirchhoff's Laws)**.** Each of the following is true in an electric circuit.

1. (Junction Rule) The current flow into a junction equals the current flow out of that junction.
2. (Circuit Rule) The algebraic sum of the voltage drops (due to resistances) around any closed circuit of the network must equal the sum of the voltage increases around the circuit.

When applying the Circuit Rule, select a direction (clockwise or counterclockwise) around the closed circuit and then consider all voltages and currents positive when in this direction and negative when in the opposite direction. Here is an example.

Example 203. Find the various currents in the circuit shown.



Explanation. First apply the junction rule at junctions *A*, *B*, *C*, and *D* to obtain

$$\begin{aligned}
&\text{Junction } A: && I_1 &&= I_2 + I_3\\
&\text{Junction } B: && I_6 &&= I_1 + I_5\\
&\text{Junction } C: && I_2 + I_4 &&= I_6\\
&\text{Junction } D: && I_3 + I_5 &&= I_4
\end{aligned}$$

Note that these equations are not independent (in fact, the third is an easy consequence of the other three).

Next, the circuit rule insists that the sum of the voltage increases (due to the sources) around a closed circuit must equal the sum of the voltage drops (due to resistances). By Ohm's law, the voltage loss across a resistance *R* (in the direction of the current *I*) is *RI*. Going counterclockwise around three closed circuits yields

$$\begin{aligned}
&\text{Upper left:} && 10 + 5 &&= 20I_1\\
&\text{Upper right:} && -5 + 20 &&= 10I_3 + 5I_4\\
&\text{Lower:} && -10 &&= -5I_5 - 5I_4
\end{aligned}$$

Hence, disregarding the redundant equation obtained at junction *C*, we have six equations in the six unknowns $I_1, \ldots, I_6$. The solution is

$$I_1 = \frac{15}{20} \qquad I_2 = \frac{-1}{20}$$

$$I_3 = \frac{16}{20} \qquad I_4 = \frac{28}{20}$$

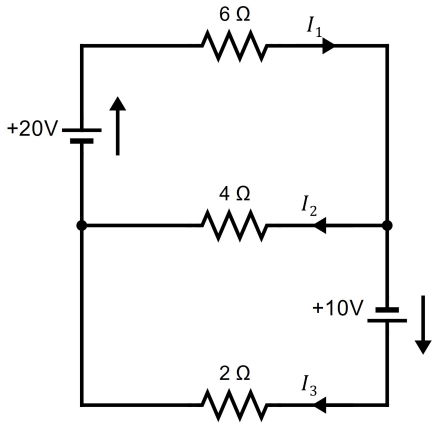
$$I_5 = \frac{12}{20} \qquad I_6 = \frac{27}{20}$$

The fact that I_2 is negative means, of course, that this current is in the opposite direction, with a magnitude of $\frac{1}{20}$ amperes.

Practice Problems

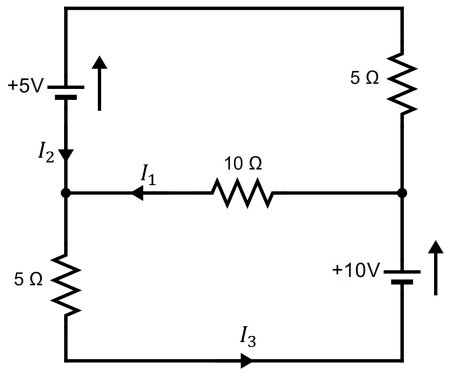
Problems **??-??**

Find the currents in the curcuits.



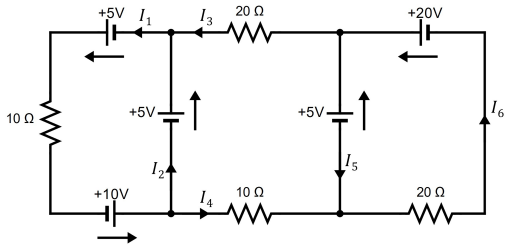
Problem 516

I_1 = ..., I_2 = ..., I_3 = ...



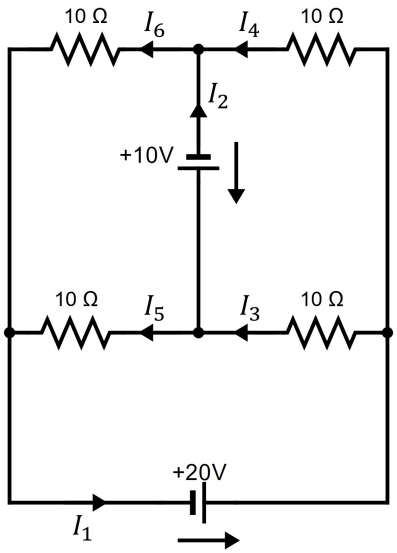
Problem 517

I_1 = ..., I_2 = ..., I_3 = ...



Problem 518

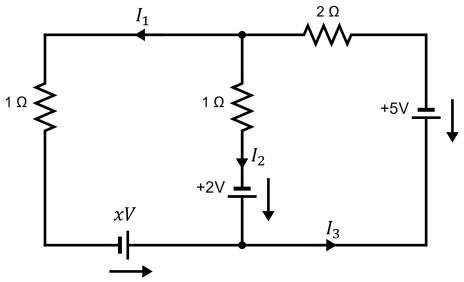
I_1 = ..., I_2 = ..., I_3 = ..., I_4 = ..., I_5 = ..., I_6 = ...



Problem 519

$I_1 = \ldots\ldots$, $I_2 = \ldots\ldots$, $I_3 = \ldots\ldots$, $I_4 = \ldots\ldots$, $I_5 = \ldots\ldots$, $I_6 = \ldots\ldots$

Problem 520 *Find the voltage x such that the current $I_1 = 0$.*



$x = \ldots\ldots$

Text Source

This application was adapted from Section 1.5 of Keith Nicholson's *Linear Algebra with Applications*. (CC-BY-NC-SA)

W. Keith Nicholson, *Linear Algebra with Applications*, Lyryx 2018, Open Edition, p. 29

Circuit Diagrams were made using *https://www.circuit-diagram.org/*

80 Application to Chemical Equations

When a chemical reaction takes place a number of molecules combine to produce new molecules. Hence, when hydrogen H_2 and oxygen O_2 molecules combine, the result is water H_2O. We express this as

$$H_2 + O_2 \to H_2O$$

Individual atoms are neither created nor destroyed, so the number of hydrogen and oxygen atoms going into the reaction must equal the number coming out (in the form of water). In this case the reaction is said to be *balanced*. Note that each hydrogen molecule H_2 consists of two atoms as does each oxygen molecule O_2, while a water molecule H_2O consists of two hydrogen atoms and one oxygen atom. In the above reaction, this requires that twice as many hydrogen molecules enter the reaction; we express this as follows:

$$2H_2 + O_2 \to 2H_2O$$

This is now balanced because there are 4 hydrogen atoms and 2 oxygen atoms on each side of the reaction.

Example 204. Balance the following reaction for burning octane C_8H_{18} in oxygen O_2:

$$C_8H_{18} + O_2 \to CO_2 + H_2O$$

where CO_2 represents carbon dioxide. We must find positive integers x, y, z, and w such that

$$xC_8H_{18} + yO_2 \to zCO_2 + wH_2O$$

Equating the number of carbon, hydrogen, and oxygen atoms on each side gives $8x = z$, $18x = 2w$ and $2y = 2z + w$, respectively. These can be written as a homogeneous linear system

$$\begin{array}{rcrcrcl} 8x & & - & z & & = & 0 \\ 18x & & & & - \ 2w & = & 0 \\ & 2y & - \ 2z & - & w & = & 0 \end{array}$$

which can be solved by gaussian elimination. In larger systems this is necessary but, in such a simple situation, it is easier to solve directly. Set $w = t$, so that $x = \frac{1}{9}t$, $z = \frac{8}{9}t$, $2y = \frac{16}{9}t + t = \frac{25}{9}t$. But x, y, z, and w must be positive integers, so the smallest value of t that eliminates fractions is 18. Hence, $x = 2$, $y = 25$, $z = 16$, and $w = 18$, and the balanced reaction is

$$2C_8H_{18} + 25O_2 \to 16CO_2 + 18H_2O$$

The reader can verify that this is indeed balanced.

It is worth noting that this problem introduces a new element into the theory of linear equations: the insistence that the solution must consist of positive integers.

Practice Problems

Problems **??**-**??**

Balance the chemical reaction.

Problem 521 $CH_4 + O_2 \rightarrow CO_2 + H_2O$. *This is the burning of methane* CH_4.

Problem 522 $\ldots\ldots NH_3 + \ldots\ldots CuO \rightarrow N_2 + \ldots\ldots Cu + \ldots\ldots H_2O$. *Here* NH_3 *is ammonia, CuO is copper oxide, Cu is copper, and* N_2 *is nitrogen.*

Problem 523 $CO_2 + H_2O \rightarrow C_6H_{12}O_6 + O_2$. *This is called the photosynthesis reaction—*$C_6H_{12}O_6$ *is glucose.*

Problem 524 $\ldots\ldots Pb(N_3)_2 + \ldots\ldots Cr(MnO_4)_2 \rightarrow \ldots\ldots Cr_2O_3 + \ldots\ldots MnO_2 + \ldots\ldots Pb_3O_4 + \ldots\ldots NO.$

Problem 525 $KNO_3 + H_2CO_3 \rightarrow K_2CO_3 + HNO_3$

Problem 526 $AgI + Na_2S \rightarrow Ag_2S + NaI$

Problem 527 $Ba_3N_2 + H_2O \rightarrow Ba(OH)_2 + NH_3$

Problem 528 $CaCl_2 + Na_3PO_4 \rightarrow Ca_3(PO_4)_2 + NaCl$

Text Source

This application was adapted from Section 1.4 of Keith Nicholson's *Linear Algebra with Applications*. (CC-BY-NC-SA)

W. Keith Nicholson, *Linear Algebra with Applications*, Lyryx 2018, Open Edition, p. 32 Practice Problems **??**, **??**, **??**, and **??** are Exercises 1.2.60 from Keith Nicholson's *Linear Algebra with Applications*. (CC-BY-NC-SA)

Ken Kuttler, *A First Course in Linear Algebra*, Lyryx 2017, Open Edition, p. 49.

81 Application to Input-Output Economic Models

In 1973 Wassily Leontief was awarded the Nobel prize in economics for his work on mathematical models. (See W. W. Leontief, "The world economy of the year 2000," *Scientific American,* Sept. 1980.) Roughly speaking, an economic system in this model consists of several industries, each of which produces a product and each of which uses some of the production of the other industries. The following example is typical.

Example 205. A primitive society has three basic needs: food, shelter, and clothing. There are thus three industries in the society—the farming, housing, and garment industries—that produce these commodities. Each of these industries consumes a certain proportion of the total output of each commodity according to the following table.

		OUTPUT		
		Farming	**Housing**	**Garment**
	Farming	0.4	0.2	0.3
INPUT	**Housing**	0.2	0.6	0.4
	Garment	0.4	0.2	0.3

Find the annual prices that each industry must charge for its income to equal its expenditures.

Explanation. Let p_1, p_2, and p_3 be the prices charged per year by the farming, housing, and garment industries, respectively, for their total output. To see how these prices are determined, consider the farming industry. It receives p_1 for its production in any year. But it *consumes* products from all these industries in the following amounts (from row 1 of the table): 40% of the food, 20% of the housing, and 30% of the clothing. Hence, the expenditures of the farming industry are $0.4p_1 + 0.2p_2 + 0.3p_3$, so

$$0.4p_1 + 0.2p_2 + 0.3p_3 = p_1$$

A similar analysis of the other two industries leads to the following system of equations.

$$\begin{aligned} 0.4p_1 + 0.2p_2 + 0.3p_3 &= p_1 \\ 0.2p_1 + 0.6p_2 + 0.4p_3 &= p_2 \\ 0.4p_1 + 0.2p_2 + 0.3p_3 &= p_3 \end{aligned}$$

This has the matrix form $E\mathbf{p} = \mathbf{p}$, where

$$E = \begin{bmatrix} 0.4 & 0.2 & 0.3 \\ 0.2 & 0.6 & 0.4 \\ 0.4 & 0.2 & 0.3 \end{bmatrix} \quad \text{and} \quad \mathbf{p} = \begin{bmatrix} p_1 \\ p_2 \\ p_3 \end{bmatrix}$$

The equations can be written as the homogeneous system

$$(I - E)\mathbf{p} = \mathbf{0}$$

where I is the 3×3 identity matrix, and the solutions are

$$\mathbf{p} = \begin{bmatrix} 2t \\ 3t \\ 2t \end{bmatrix}$$

where t is a parameter. Thus, the pricing must be such that the total output of the farming industry has the same value as the total output of the garment industry, whereas the total value of the housing industry must be $\frac{3}{2}$ as much.

In general, suppose an economy has n industries, each of which uses some (possibly none) of the production of every industry. We assume first that the economy is *closed* (that is, no product is exported or imported) and that all product is used. Given two industries i and j, let e_{ij} denote the proportion of the total annual output of industry j that is consumed by industry i. Then $E = \begin{bmatrix} e_{ij} \end{bmatrix}$ is called the *input-output* matrix. Clearly,

$$0 \leq e_{ij} \leq 1 \quad \text{for all } i \text{ and } j \tag{112}$$

Moreover, all the output from industry j is used by *some* industry (the model is closed), so

$$e_{1j} + e_{2j} + \cdots + e_{ij} = 1 \quad \text{for each } j \tag{113}$$

This condition asserts that each column of E sums to 1. Matrices satisfying conditions (**??**) and (**??**) are called *stochastic matrices*.

As in Example **??**, let p_i denote the price of the total annual production of industry i. Then p_i is the annual revenue of industry i. On the other hand, industry i spends $e_{i1}p_1 + e_{i2}p_2 + \cdots + e_{in}p_n$ annually for the product it uses ($e_{ij}p_j$ is the cost for product from industry j). The closed economic system is said to be in *equilibrium* if the annual expenditure equals the annual revenue for each industry—that is, if

$$e_{1j}p_1 + e_{2j}p_2 + \cdots + e_{ij}p_n = p_i \quad \text{for each } i = 1, 2, \ldots, n$$

If we write $\mathbf{p} = \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_n \end{bmatrix}$, these equations can be written as the matrix equation

$$E\mathbf{p} = \mathbf{p}$$

This is called the *equilibrium condition*, and the solutions **p** are called *equilibrium price structures*. The equilibrium condition can be written as

$$(I - E)\mathbf{p} = \mathbf{0}$$

which is a system of homogeneous equations for **p**. Moreover, there is always a nontrivial solution **p**. Indeed, the column sums of $I - E$ are all 0 (because E is stochastic), so the row-echelon form of $I - E$ has a row of zeros. In fact, more is true:

Theorem 126. Let E be any $n \times n$ stochastic matrix. Then there is a nonzero $n \times 1$ vector **p** with nonnegative entries such that $E\mathbf{p} = \mathbf{p}$. If all the entries of E are positive, the matrix **p** can be chosen with all entries positive.

Theorem **??** guarantees the existence of an equilibrium price structure for any closed input-output system of the type discussed here. The proof is beyond the scope of this book.[2]

Example 206. Find the equilibrium price structures for four industries if the input-output matrix is

$$E = \begin{bmatrix} 0.6 & 0.2 & 0.1 & 0.1 \\ 0.3 & 0.4 & 0.2 & 0 \\ 0.1 & 0.3 & 0.5 & 0.2 \\ 0 & 0.1 & 0.2 & 0.7 \end{bmatrix}$$

Find the prices if the total value of business is \$1000.

Explanation. If $\mathbf{p} = \begin{bmatrix} p_1 \\ p_2 \\ p_3 \\ p_4 \end{bmatrix}$ is the equilibrium price structure, then the equilibrium condition reads $E\mathbf{p} = \mathbf{p}$. When we write this as $(I - E)\mathbf{p} = \mathbf{0}$, we obtain the following family of solutions

$$\mathbf{p} = \begin{bmatrix} 44t \\ 39t \\ 51t \\ 47t \end{bmatrix}$$

where t is a parameter. If we insist that $p_1 + p_2 + p_3 + p_4 = 1000$, then $t = 5.525$. Hence

$$\mathbf{p} = \begin{bmatrix} 243.09 \\ 215.47 \\ 281.76 \\ 259.67 \end{bmatrix}$$

[2]The interested reader is referred to P. Lancaster's *Theory of Matrices* (New York: Academic Press, 1969) or to E. Seneta's *Non-negative Matrices* (New York: Wiley, 1973).

to five figures.

The Open Model

We now assume that there is a demand for products in the *open sector* of the economy, which is the part of the economy other than the producing industries (for example, consumers). Let d_i denote the total value of the demand for product i in the open sector. If p_i and e_{ij} are as before, the value of the annual demand for product i by the producing industries themselves is $e_{i1}p_1 + e_{i2}p_2 + \cdots + e_{in}p_n$, so the total annual revenue p_i of industry i breaks down as follows:

$$p_i = (e_{i1}p_1 + e_{i2}p_2 + \cdots + e_{in}p_n) + d_i \quad \text{for each } i = 1, 2, \ldots, n$$

The column $\mathbf{d} = \begin{bmatrix} d_1 \\ \vdots \\ d_n \end{bmatrix}$ is called the *demand matrix*, and this gives a matrix equation

$$\mathbf{p} = E\mathbf{p} + \mathbf{d}$$

or

$$(I - E)\mathbf{p} = \mathbf{d} \tag{114}$$

This is a system of linear equations for $\mathbf{p}$, and we ask for a solution $\mathbf{p}$ with every entry nonnegative. Note that every entry of E is between 0 and 1, but the column sums of E need not equal 1 as in the closed model.

Before proceeding, it is convenient to introduce a useful notation. If $A = \begin{bmatrix} a_{ij} \end{bmatrix}$ and $B = \begin{bmatrix} b_{ij} \end{bmatrix}$ are matrices of the same size, we write $A > B$ if $a_{ij} > b_{ij}$ for all i and j, and we write $A \geq B$ if $a_{ij} \geq b_{ij}$ for all i and j. Thus $P \geq 0$ means that every entry of P is nonnegative. Note that $A \geq 0$ and $B \geq 0$ implies that $AB \geq 0$.

Now, given a demand matrix $\mathbf{d} \geq \mathbf{0}$, we look for a production matrix $\mathbf{p} \geq \mathbf{0}$ satisfying equation (**??**). This certainly exists if $I - E$ is invertible and $(I - E)^{-1} \geq 0$. On the other hand, the fact that $\mathbf{d} \geq \mathbf{0}$ means any solution $\mathbf{p}$ to equation (**??**) satisfies $\mathbf{p} \geq E\mathbf{p}$. Hence, the following theorem is not too surprising.

Theorem 127. Let $E \geq 0$ be a square matrix. Then $I - E$ is invertible and $(I - E)^{-1} \geq 0$ if and only if there exists a column $\mathbf{p} > \mathbf{0}$ such that $\mathbf{p} > E\mathbf{p}$.

Heuristic Proof If $(I - E)^{-1} \geq 0$, the existence of $\mathbf{p} > \mathbf{0}$ with $\mathbf{p} > E\mathbf{p}$ is left as Practice Problem **??**. Conversely, suppose such a column $\mathbf{p}$ exists. Observe that

$$(I - E)(I + E + E^2 + \cdots + E^{k-1}) = I - E^k$$

holds for all $k \geq 2$. If we can show that every entry of E^k approaches 0 as k becomes large then, intuitively, the infinite matrix sum

$$U = I + E + E^2 + \cdots$$

exists and $(I - E)U = I$. Since $U \geq 0$, this does it. To show that E^k approaches 0, it suffices to show that $EP < \mu P$ for some number μ with $0 < \mu < 1$ (then $E^k P < \mu^k P$ for all $k \geq 1$ by induction). The existence of μ is left as Practice Problem **??**. ■

The condition $\mathbf{p} > E\mathbf{p}$ in Theorem **??** has a simple economic interpretation. If $\mathbf{p}$ is a production matrix, entry i of $E\mathbf{p}$ is the total value of all product used by industry i in a year. Hence, the condition $\mathbf{p} > E\mathbf{p}$ means that, for each i, the value of product produced by industry i exceeds the value of the product it uses. In other words, each industry runs at a profit.

Example 207. If $E = \begin{bmatrix} 0.6 & 0.2 & 0.3 \\ 0.1 & 0.4 & 0.2 \\ 0.2 & 0.5 & 0.1 \end{bmatrix}$, show that $I - E$ is invertible and $(I - E)^{-1} \geq 0$.

Explanation. Use $\mathbf{p} = (3, 2, 2)^T$ in Theorem **??**.

If $\mathbf{p}_0 = (1, 1, 1)^T$, the entries of $E\mathbf{p}_0$ are the row sums of E. Hence $\mathbf{p}_0 > E\mathbf{p}_0$ holds if the row sums of E are all less than 1. This proves the first of the following useful facts (the second is Practice Problem **??**).

Corollary 15. Let $E \geq 0$ be a square matrix. In each case, $I - E$ is invertible and $(I - E)^{-1} \geq 0$:

1. All row sums of E are less than 1.
2. All column sums of E are less than 1.

Practice Problems

Problems **??**-**??**

Find the possible equilibrium price structures for each given input-output matrix.

Problem 529 $\begin{bmatrix} 0.1 & 0.2 & 0.3 \\ 0.6 & 0.2 & 0.3 \\ 0.3 & 0.6 & 0.4 \end{bmatrix}$

Problem 530 $\begin{bmatrix} 0.5 & 0 & 0.5 \\ 0.1 & 0.9 & 0.2 \\ 0.4 & 0.1 & 0.3 \end{bmatrix}$

$\mathbf{p} = \begin{bmatrix} \dots \dots \\ \dots \dots \\ \dots \dots \end{bmatrix}$ *(Use t as the parameter.)*

Problem 531 $\begin{bmatrix} 0.3 & 0.1 & 0.1 & 0.2 \\ 0.2 & 0.3 & 0.1 & 0 \\ 0.3 & 0.3 & 0.2 & 0.3 \\ 0.2 & 0.3 & 0.6 & 0.5 \end{bmatrix}$

Problem 532 $\begin{bmatrix} 0.5 & 0 & 0.1 & 0.1 \\ 0.2 & 0.7 & 0 & 0.1 \\ 0.1 & 0.2 & 0.8 & 0.2 \\ 0.2 & 0.1 & 0.1 & 0.6 \end{bmatrix}$

$\mathbf{p} = \begin{bmatrix} \dots \dots \\ \dots \dots \\ \dots \dots \\ \dots \dots \end{bmatrix}$ *(Use t as the parameter.)*

Problem 533 *Three industries A, B, and C are such that all the output of A is used by B, all the output of B is used by C, and all the output of C is used by A. Find the possible equilibrium price structures.*

$\mathbf{p} = \begin{bmatrix} \dots \dots \\ \dots \dots \\ \dots \dots \end{bmatrix}$ *(Use t as the parameter.)*

Problem 534 *Find the possible equilibrium price structures for three industries where the input-output matrix is* $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$*. Discuss why there are two parameters here.*

Problem 535 *Prove Theorem* **??** *for a* 2×2 *stochastic matrix E by first writing it in the form* $E = \begin{bmatrix} a & b \\ 1-a & 1-b \end{bmatrix}$*, where* $0 \leq a \leq 1$ *and* $0 \leq b \leq 1$*.*

Problem 536 *If E is an $n \times n$ stochastic matrix and* $\mathbf{c}$ *is an $n \times 1$ matrix, show that the sum of the entries of* $\mathbf{c}$ *equals the sum of the entries of the $n \times 1$ matrix $E\mathbf{c}$.*

Problem 537 *Let $W = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \end{bmatrix}$. Let E and F denote $n \times n$ matrices with nonnegative entries.*

1. *Show that E is a stochastic matrix if and only if $WE = W$.*
2. *Use part (a.) to deduce that, if E and F are both stochastic matrices, then EF is also stochastic.*

Problem 538 *Find a 2×2 matrix E with entries between 0 and 1 such that:*

1. *$I - E$ has no inverse.*
2. *$I - E$ has an inverse but not all entries of $(I - E)^{-1}$ are nonnegative.*

Problem 539 *If E is a 2×2 matrix with entries between 0 and 1, show that $I - E$ is invertible and $(I - E)^{-1} \geq 0$ if and only if $tr(E) < 1 + \det E$. Here, if $E = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, then $tr(E) = a + d$ and $\det E = ad - bc$.*

Problems **??-??**

In each case show that $I - E$ is invertible and $(I - E)^{-1} \geq 0$.

Problem 540 $\begin{bmatrix} 0.6 & 0.5 & 0.1 \\ 0.1 & 0.3 & 0.3 \\ 0.2 & 0.1 & 0.4 \end{bmatrix}$

Problem 541 $\begin{bmatrix} 0.7 & 0.1 & 0.3 \\ 0.2 & 0.5 & 0.2 \\ 0.1 & 0.1 & 0.4 \end{bmatrix}$

Problem 542 $\begin{bmatrix} 0.6 & 0.2 & 0.1 \\ 0.3 & 0.4 & 0.2 \\ 0.2 & 0.5 & 0.1 \end{bmatrix}$

Problem 543 $\begin{bmatrix} 0.8 & 0.1 & 0.1 \\ 0.3 & 0.1 & 0.2 \\ 0.3 & 0.3 & 0.2 \end{bmatrix}$

Problem 544 *Prove that (1) implies (2) in the Corollary to Theorem* **??**.

Problem 545 *If* $(I - E)^{-1} \geq 0$*, find* $\mathbf{p} > 0$ *such that* $\mathbf{p} > E\mathbf{p}$.

Problem 546 *If* $E\mathbf{p} < \mathbf{p}$ *where* $E \geq 0$ *and* $\mathbf{p} > 0$*, find a number* μ *such that* $E\mathbf{p} < \mu\mathbf{p}$ *and* $0 < \mu < 1$.

Text Source

This application was adapted from Section 2.8 of Keith Nicholson's *Linear Algebra with Applications*. (CC-BY-NC-SA)

W. Keith Nicholson, *Linear Algebra with Applications*, Lyryx 2018, Open Edition, p. 128

82 Application to Markov Chains

Many natural phenomena progress through various stages and can be in a variety of states at each stage. For example, the weather in a given city progresses day by day and, on any given day, may be sunny or rainy. Here the states are "sun" and "rain," and the weather progresses from one state to another in daily stages. Another example might be a football team: The stages of its evolution are the games it plays, and the possible states are "win," "draw," and "loss."

The general setup is as follows: A "system" evolves through a series of "stages," and at any stage it can be in any one of a finite number of "states." At any given stage, the state to which it will go at the next stage depends on the past and present history of the system—that is, on the sequence of states it has occupied to date.

Definition 102. Markov Chain A *Markov chain* is such an evolving system wherein the state to which it will go next depends only on its present state and does not depend on the earlier history of the system.

(Note: The name honors Andrei Andreyevich Markov (1856–1922) who was a professor at the university in St. Petersburg, Russia.)

Even in the case of a Markov chain, the state the system will occupy at any stage is determined only in terms of probabilities. In other words, chance plays a role. For example, if a football team wins a particular game, we do not know whether it will win, draw, or lose the next game. On the other hand, we may know that the team tends to persist in winning streaks; for example, if it wins one game it may win the next game $\frac{1}{2}$ of the time, lose $\frac{4}{10}$ of the time, and draw $\frac{1}{10}$ of the time. These fractions are called the *probabilities* of these various possibilities. Similarly, if the team loses, it may lose the next game with probability $\frac{1}{2}$ (that is, half the time), win with probability $\frac{1}{4}$, and draw with probability $\frac{1}{4}$. The probabilities of the various outcomes after a drawn game will also be known.

We shall treat probabilities informally here: *The probability that a given event will occur is the long-run proportion of the time that the event does indeed occur*. Hence, all probabilities are numbers between 0 and 1. A probability of 0 means the event is impossible and never occurs; events with probability 1 are certain to occur.

If a Markov chain is in a particular state, the probabilities that it goes to the various states at the next stage of its evolution are called the *transition probabilities* for the chain, and they are assumed to be known quantities. To motivate the general conditions that follow, consider the following simple example. Here the system is a man, the stages are his successive lunches, and the states are the two restaurants he chooses.

Example 208. A man always eats lunch at one of two restaurants, *A* and *B*. He never eats at *A* twice in a row. However, if he eats at *B*, he is three times as likely to eat at *B* next time as at *A*. Initially, he is equally likely to eat at either restaurant.

1. What is the probability that he eats at *A* on the third day after the initial one?
2. What proportion of his lunches does he eat at *A*?

Explanation. The table of transition probabilities follows. The *A* column indicates that if he eats at *A* on one day, he never eats there again on the next day and so is certain to go to *B*.

		Present Lunch	
		A	*B*
Next Lunch	*A*	0	0.25
	B	1	0.75

The *B* column shows that, if he eats at *B* on one day, he will eat there on the next day $\frac{3}{4}$ of the time and switches to *A* only $\frac{1}{4}$ of the time.

The restaurant he visits on a given day is not determined. The most that we can expect is to know the probability that he will visit *A* or *B* on that day.

Let $\mathbf{s}_m = \begin{bmatrix} s_1^{(m)} \\ s_2^{(m)} \end{bmatrix}$ denote the *state vector* for day *m*. Here $s_1^{(m)}$ denotes the probability that he eats at *A* on day *m*, and $s_2^{(m)}$ is the probability that he eats at *B* on day *m*. It is convenient to let $\mathbf{s}_0$ correspond to the initial day. Because he is equally likely to eat at *A* or *B* on that initial day, $s_1^{(0)} = 0.5$ and $s_2^{(0)} = 0.5$, so $\mathbf{s}_0 = \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix}$. Now let

$$P = \begin{bmatrix} 0 & 0.25 \\ 1 & 0.75 \end{bmatrix}$$

denote the *transition matrix*. We claim that the relationship

$$\mathbf{s}_{m+1} = P\mathbf{s}_m$$

holds for all integers $m \geq 0$. This will be derived later; for now, we use it as follows to successively

compute $\mathbf{s}_1, \mathbf{s}_2, \mathbf{s}_3, \ldots$.

$$\mathbf{s}_1 = P\mathbf{s}_0 = \begin{bmatrix} 0 & 0.25 \\ 1 & 0.75 \end{bmatrix} \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix} = \begin{bmatrix} 0.125 \\ 0.875 \end{bmatrix}$$

$$\mathbf{s}_2 = P\mathbf{s}_1 = \begin{bmatrix} 0 & 0.25 \\ 1 & 0.75 \end{bmatrix} \begin{bmatrix} 0.125 \\ 0.875 \end{bmatrix} = \begin{bmatrix} 0.21875 \\ 0.78125 \end{bmatrix}$$

$$\mathbf{s}_3 = P\mathbf{s}_2 = \begin{bmatrix} 0 & 0.25 \\ 1 & 0.75 \end{bmatrix} \begin{bmatrix} 0.21875 \\ 0.78125 \end{bmatrix} = \begin{bmatrix} 0.1953125 \\ 0.8046875 \end{bmatrix}$$

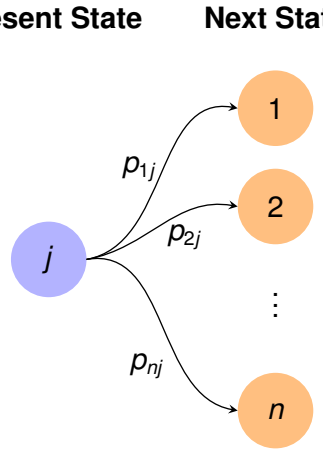
Hence, the probability that his third lunch (after the initial one) is at *A* is approximately 0.195, whereas the probability that it is at *B* is 0.805. If we carry these calculations on, the next state vectors are (to five figures):

$$\mathbf{s}_4 = \begin{bmatrix} 0.20117 \\ 0.79883 \end{bmatrix} \quad \mathbf{s}_5 = \begin{bmatrix} 0.19971 \\ 0.80029 \end{bmatrix}$$

$$\mathbf{s}_6 = \begin{bmatrix} 0.20007 \\ 0.79993 \end{bmatrix} \quad \mathbf{s}_7 = \begin{bmatrix} 0.19998 \\ 0.80002 \end{bmatrix}$$

Moreover, as m increases the entries of $\mathbf{s}_m$ get closer and closer to the corresponding entries of $\begin{bmatrix} 0.2 \\ 0.8 \end{bmatrix}$. Hence, in the long run, he eats 20% of his lunches at *A* and 80% at *B*.

Example **??** incorporates most of the essential features of all Markov chains. The general model is as follows: The system evolves through various stages and at each stage can be in exactly one of n distinct states. It progresses through a sequence of states as time goes on. If a Markov chain is in state j at a particular stage of its development, the probability p_{ij} that it goes to state i at the next stage is called the *transition probability*. The $n \times n$ matrix $P = \left[p_{ij}\right]$ is called the *transition matrix* for the Markov chain. The situation is depicted graphically in the diagram.



We make one important assumption about the transition matrix $P = \left[p_{ij}\right]$: It does *not* depend on which stage the process is in. This assumption means that the transition probabilities are *independent of time*—that is, they do not change as time goes on. It is this assumption that distinguishes Markov chains in the literature of this subject.

Example 209. Suppose the transition matrix of a three-state Markov chain is If, for example, the system is in state 2, then column 2 lists the probabilities of where it goes next. Thus, the probability is $p_{12} = 0.1$ that it goes from state 2 to state 1, and the probability is $p_{22} = 0.9$ that it goes from state 2 to state 2. The fact that $p_{32} = 0$ means that it is impossible for it to go from state 2 to state 3 at the next stage.

Consider the jth column of the transition matrix P.

$$\begin{bmatrix} p_{1j} \\ p_{2j} \\ \vdots \\ p_{nj} \end{bmatrix}$$

If the system is in state j at some stage of its evolution, the transition probabilities $p_{1j}, p_{2j}, \ldots, p_{nj}$ represent the fraction of the time that the system will move to state 1, state 2, ..., state n, respectively, at the next stage. We assume that it has to go to *some* state at each transition, so the sum of these probabilities is 1:

$$p_{1j} + p_{2j} + \cdots + p_{nj} = 1 \quad \text{for each } j$$

Thus, the columns of P all sum to 1 and the entries of P lie between 0 and 1. Hence P is called a *stochastic matrix*.

As in Example **??**, we introduce the following notation: Let $s_i^{(m)}$ denote the probability that the system is in state i after m transitions. The $n \times 1$ matrices

$$\mathbf{s}_m = \begin{bmatrix} s_1^{(m)} \\ s_2^{(m)} \\ \vdots \\ s_n^{(m)} \end{bmatrix} \qquad m = 0, 1, 2, ...$$

are called the *state vectors* for the Markov chain. Note that the sum of the entries of $\mathbf{s}_m$ must equal 1 because the system must be in *some* state after m transitions. The matrix $\mathbf{s}_0$ is called the *initial state vector* for the Markov chain and is given as part of the data of the particular chain. For example, if the chain has only two states, then an initial vector $\mathbf{s}_0 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ means that it started in state 1. If it started in state 2, the initial vector would be $\mathbf{s}_0 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$. If $\mathbf{s}_0 = \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix}$, it is equally likely that the system started in state 1 or in state 2.

Theorem 128. Let P be the transition matrix for an n-state Markov chain. If $\mathbf{s}_m$ is the state vector at stage m, then

$$\mathbf{s}_{m+1} = P\mathbf{s}_m$$

for each $m = 0, 1, 2, ...$.

Heuristic Proof Suppose that the Markov chain has been run N times, each time starting with the same initial state vector. Recall that p_{ij} is the proportion of the time the system goes from state j at some stage to state i at the next stage, whereas $s_i^{(m)}$ is the proportion of the time it is in state i at stage m. Hence

$$s_i^{m+1} N$$

is (approximately) the number of times the system is in state i at stage $m+1$. We are going to calculate this number another way. The system got to state i at stage $m+1$ through *some* other state (say state j) at stage m. The number of times it was *in* state j at that stage is (approximately) $s_j^{(m)} N$, so the number of times it got to state i via state j is $p_{ij}(s_j^{(m)} N)$. Summing over j gives the number of times the system is in state i (at stage $m+1$). This is the number we calculated before, so

$$s_i^{(m+1)} N = p_{i1} s_1^{(m)} N + p_{i2} s_2^{(m)} N + \cdots + p_{in} s_n^{(m)} N$$

Dividing by N gives $s_i^{(m+1)} = p_{i1} s_1^{(m)} + p_{i2} s_2^{(m)} + \cdots + p_{in} s_n^{(m)}$ for each i, and this can be expressed as the matrix equation $\mathbf{s}_{m+1} = P\mathbf{s}_m$. ■

If the initial probability vector $\mathbf{s}_0$ and the transition matrix P are given, Theorem **??** gives $\mathbf{s}_1, \mathbf{s}_2, \mathbf{s}_3, \ldots$, one after the other, as follows:

$$\begin{aligned} \mathbf{s}_1 &= P\mathbf{s}_0 \\ \mathbf{s}_2 &= P\mathbf{s}_1 \\ \mathbf{s}_3 &= P\mathbf{s}_2 \\ &\vdots \end{aligned}$$

Hence, the state vector $\mathbf{s}_m$ is completely determined for each $m = 0, 1, 2, \ldots$ by P and $\mathbf{s}_0$.

Example 210. A wolf pack always hunts in one of three regions R_1, R_2, and R_3. Its hunting habits are as follows:

1. If it hunts in some region one day, it is as likely as not to hunt there again the next day.
2. If it hunts in R_1, it never hunts in R_2 the next day.
3. If it hunts in R_2 or R_3, it is equally likely to hunt in each of the other regions the next day.

If the pack hunts in R_1 on Monday, find the probability that it hunts there on Thursday.

Explanation. The stages of this process are the successive days; the states are the three regions. The transition matrix P is determined as follows (see the table): The first habit asserts that $p_{11} = p_{22} = p_{33} = \frac{1}{2}$. Now column 1 displays what happens when the pack starts in R_1: It never goes to state 2, so $p_{21} = 0$ and, because the column must sum to 1, $p_{31} = \frac{1}{2}$. Column 2 describes what happens if it starts in R_2: $p_{22} = \frac{1}{2}$ and p_{12} and p_{32} are equal (by habit 3), so $p_{12} = p_{32} = \frac{1}{2}$ because the column sum must equal 1. Column 3 is filled in a similar way. Now let Monday be the initial stage. Then $\mathbf{s}_0 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$ because the pack hunts in R_1 on that day. Then $\mathbf{s}_1$, $\mathbf{s}_2$, and $\mathbf{s}_3$ describe Tuesday, Wednesday, and Thursday, respectively, and we compute them using Theorem **??**.

$$\mathbf{s}_1 = P\mathbf{s}_0 = \begin{bmatrix} \frac{1}{2} \\ 0 \\ \frac{1}{2} \end{bmatrix} \quad \mathbf{s}_2 = P\mathbf{s}_1 = \begin{bmatrix} \frac{3}{8} \\ \frac{1}{8} \\ \frac{4}{8} \end{bmatrix} \quad \mathbf{s}_3 = P\mathbf{s}_2 = \begin{bmatrix} \frac{11}{32} \\ \frac{6}{32} \\ \frac{15}{32} \end{bmatrix}$$

Hence, the probability that the pack hunts in Region R_1 on Thursday is $\frac{11}{32}$.

Steady State Vector

Another phenomenon that was observed in Example **??** can be expressed in general terms. The state vectors $\mathbf{s}_0, \mathbf{s}_1, \mathbf{s}_2, \ldots$ were calculated in that example and were found to "approach" $\mathbf{s} = \begin{bmatrix} 0.2 \\ 0.8 \end{bmatrix}$. This means that the first component of $\mathbf{s}_m$ becomes and remains very close to 0.2 as m becomes large, whereas the second component gets close to 0.8 as m increases. When this is the case, we say that $\mathbf{s}_m$ *converges* to $\mathbf{s}$. For large m, then, there is very little error in taking $\mathbf{s}_m = \mathbf{s}$, so the long-term probability that the system is in state 1 is 0.2, whereas the probability that it is in state 2 is 0.8. In Example **??**, enough state vectors were computed for the limiting vector $\mathbf{s}$ to be apparent. However, there is a better way to do this that works in most cases.

Suppose P is the transition matrix of a Markov chain, and assume that the state vectors $\mathbf{s}_m$ converge to a limiting vector $\mathbf{s}$. Then $\mathbf{s}_m$ is very close to $\mathbf{s}$ for sufficiently large m, so $\mathbf{s}_{m+1}$ is also very close to $\mathbf{s}$. Thus, the equation $\mathbf{s}_{m+1} = P\mathbf{s}_m$ from Theorem **??** is closely approximated by

$$\mathbf{s} = P\mathbf{s}$$

so it is not surprising that $\mathbf{s}$ should be a solution to this matrix equation. Moreover, it is easily solved because it can be written as a system of homogeneous linear equations

$$(I - P)\mathbf{s} = \mathbf{0}$$

with the entries of $\mathbf{s}$ as variables.

In Example **??**, where $P = \begin{bmatrix} 0 & 0.25 \\ 1 & 0.75 \end{bmatrix}$, the general solution to $(I - P)\mathbf{s} = \mathbf{0}$ is $\mathbf{s} = \begin{bmatrix} t \\ 4t \end{bmatrix}$, where t is a parameter. But if we insist that the entries of S sum to 1 (as must be true of all state vectors), we find $t = 0.2$ and so $\mathbf{s} = \begin{bmatrix} 0.2 \\ 0.8 \end{bmatrix}$ as before.

All this is predicated on the existence of a limiting vector for the sequence of state vectors of the Markov chain, and such a vector may not always exist. However, it does exist in one commonly occurring situation. A stochastic matrix P is called *regular* if some power P^m of P has every entry greater than zero. The matrix $P = \begin{bmatrix} 0 & 0.25 \\ 1 & 0.75 \end{bmatrix}$ of Example **??** is regular (in this case, each entry of P^2 is positive), and the general theorem is as follows:

Theorem 129. Let P be the transition matrix of a Markov chain and assume that P is regular. Then there is a unique column matrix $\mathbf{s}$ satisfying the following conditions:

1. $P\mathbf{s} = \mathbf{s}$.

2. The entries of **s** are positive and sum to 1.

Moreover, condition 1 can be written as

$$(I - P)\mathbf{s} = \mathbf{0}$$

and so gives a homogeneous system of linear equations for **s**. Finally, the sequence of state vectors $\mathbf{s}_0, \mathbf{s}_1, \mathbf{s}_2, \ldots$ converges to **s** in the sense that if m is large enough, each entry of $\mathbf{s}_m$ is closely approximated by the corresponding entry of **s**.

This theorem will not be proved here. The interested reader can find an elementary proof in J. Kemeny, H. Mirkil, J. Snell, and G. Thompson, *Finite Mathematical Structures* (Englewood Cliffs, N.J.: Prentice-Hall, 1958).

If P is the regular transition matrix of a Markov chain, the column **s** satisfying conditions 1 and 2 of Theorem **??** is called the *steady-state vector*for the Markov chain. The entries of **s** are the long-term probabilities that the chain will be in each of the various states.

Example 211. A man eats one of three soups—beef, chicken, and vegetable—each day. He never eats the same soup two days in a row. If he eats beef soup on a certain day, he is equally likely to eat each of the others the next day; if he does not eat beef soup, he is twice as likely to eat it the next day as the alternative.

1. If he has beef soup one day, what is the probability that he has it again two days later?
2. What are the long-run probabilities that he eats each of the three soups?

Explanation. The states here are B, C, and V, the three soups. The transition matrix P is given in the table. (Recall that, for each state, the corresponding column lists the probabilities for the next state.) If he has beef soup initially, then the initial state vector is

$$\mathbf{s}_0 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

Then two days later the state vector is $\mathbf{s}_2$. If P is the transition matrix, then

$$\mathbf{s}_1 = P\mathbf{s}_0 = \frac{1}{2}\begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \quad \mathbf{s}_2 = P\mathbf{s}_1 = \frac{1}{6}\begin{bmatrix} 4 \\ 1 \\ 1 \end{bmatrix}$$

so he eats beef soup two days later with probability $\frac{2}{3}$. This answers (a.) and also shows that he eats chicken and vegetable soup each with probability $\frac{1}{6}$.

To find the long-run probabilities, we must find the steady-state vector **s**. Theorem **??** applies because P is regular (P^2 has positive entries), so **s** satisfies $P\mathbf{s} = \mathbf{s}$. That is, $(I - P)\mathbf{s} = \mathbf{0}$ where

$$I - P = \frac{1}{6}\begin{bmatrix} 6 & -4 & -4 \\ -3 & 6 & -2 \\ -3 & -2 & 6 \end{bmatrix}$$

The solution is $\mathbf{s} = \begin{bmatrix} 4t \\ 3t \\ 3t \end{bmatrix}$, where t is a parameter, and we use $\mathbf{s} = \begin{bmatrix} 0.4 \\ 0.3 \\ 0.3 \end{bmatrix}$ because the entries of **s** must sum to 1. Hence, in the long run, he eats beef soup 40% of the time and eats chicken soup and vegetable soup each 30% of the time.

Practice Problems

Problems **??**-**??**

Which of the following stochastic matrices is regular?

Problem 547 $\begin{bmatrix} 0 & 0 & \frac{1}{2} \\ 1 & 0 & \frac{1}{2} \\ 0 & 1 & 0 \end{bmatrix}$

Multiple Choice:

1. *Regular*
2. *Not regular*

Problem 548 $\begin{bmatrix} \frac{1}{2} & 0 & \frac{1}{3} \\ \frac{1}{4} & 1 & \frac{1}{3} \\ \frac{1}{4} & 0 & \frac{1}{3} \end{bmatrix}$

Multiple Choice:

1. *Regular*
2. *Not regular*

Problems **??-??**

In each case find the steady-state vector and, assuming that it starts in state 1, find the probability that it is in state 2 after 3 transitions.

Problem 549

$$\begin{bmatrix} 0.5 & 0.3 \\ 0.5 & 0.7 \end{bmatrix}$$

Steady state vector = $\begin{bmatrix} \ldots \\ \ldots \end{bmatrix}$, *probability* = ...

Problem 550

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

Problem 551

$$\begin{bmatrix} 0 & \frac{1}{2} & \frac{1}{4} \\ 1 & 0 & \frac{1}{4} \\ 0 & \frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

Steady state vector = $\begin{bmatrix} \ldots \\ \ldots \\ \ldots \end{bmatrix}$, *probability* = ...

Problem 552

$$\begin{bmatrix} 0.4 & 0.1 & 0.5 \\ 0.2 & 0.6 & 0.2 \\ 0.4 & 0.3 & 0.3 \end{bmatrix}$$

Problem 553

$$\begin{bmatrix} 0.8 & 0.0 & 0.2 \\ 0.1 & 0.6 & 0.1 \\ 0.1 & 0.4 & 0.7 \end{bmatrix}$$

Steady state vector = $\begin{bmatrix} \ldots & \ldots \\ \ldots & \ldots \\ \ldots & \ldots \end{bmatrix}$, *probability* = ...

Problem 554

$$\begin{bmatrix} 0.1 & 0.3 & 0.3 \\ 0.3 & 0.1 & 0.6 \\ 0.6 & 0.6 & 0.1 \end{bmatrix}$$

Problems **??-??**

A fox hunts in three territories *A*, *B*, and *C*. He never hunts in the same territory on two successive days. If he hunts in *A*, then he hunts in *C* the next day. If he hunts in *B* or *C*, he is twice as likely to hunt in *A* the next day as in the other territory.

Problem 555 *What proportion of his time does he spend in A, in B, and in C?*

Problem 556 *If he hunts in A on Monday (C on Monday), what is the probability that he will hunt in B on Thursday?*

Problems **??-??**

Assume that there are three social classes—upper, middle, and lower—and that social mobility behaves as follows:

1. Of the children of upper-class parents, 70% remain upper-class, whereas 10% become middle-class and 20% become lower-class.
2. Of the children of middle-class parents, 80% remain middle-class, whereas the others are evenly split between the upper class and the lower class.
3. For the children of lower-class parents, 60% remain lower-class, whereas 30% become middle-class and 10% upper-class.

Problem 557 *Find the probability that the grandchild of lower-class parents becomes upper-class.*

Problem 558 *Find the long-term breakdown of society into classes. ... ...% middle, ...% upper, ...% lower*

Problem 559 *The prime minister says she will call an election. This gossip is passed from person to person with a probability* $p \neq 0$ *that the information is passed incorrectly at any stage. Assume that when a person hears the gossip he or she passes it to one person who does not know. Find the long-term probability that a person will hear that there is going to be an election.*

Problem 560 *John makes it to work on time one Monday out of four. On other work days his behaviour is as follows: If he is late one day, he is twice as likely to come to work on time the next day as to be late. If he is on time one day, he is as likely to be late as not the next day. Find the probability of his being late and that of his being on time Wednesdays. Probability of being late = ..., Probability of being on time = ...*

Problem 561 *Suppose you have* 1 *cent and match coins with a friend. At each match you either win or lose* 1 *cent with equal probability. If you go broke or ever get* 4 *cents, you quit. Assume your friend never quits. If the states are 0, 1, 2, 3, and 4 representing your wealth, show that the corresponding transition matrix P is not regular. Find the probability that you will go broke after* 3 *matches.*

Problems **??**-**??**

A mouse is put into a maze of compartments, as in the diagram. Assume that he always leaves any compartment he enters and that he is equally likely to take any tunnel entry.

Problem 562 *If he starts in compartment 1, find the probability that he is in compartment 1 again after* 3 *moves. ... ...*

Problem 563 *Find the compartment in which he spends most of his time if he is left for a long time. He spends most of his time in compartment ... . The steady state vector is* $\begin{bmatrix} ... \\ ... \\ ... \\ ... \\ ... \end{bmatrix}$

Problem 564 *If a stochastic matrix has a* 1 *on its main diagonal, show that it cannot be regular. Assume it is not* 1×1*.*

Problem 565 *If* $\mathbf{s}_m$ *is the stage-m state vector for a Markov chain, show that* $\mathbf{s}_{m+k} = P^k \mathbf{s}_m$ *holds for all* $m \geq 1$ *and* $k \geq 1$ *(where P is the transition matrix).*

Problem 566 *A stochastic matrix is doubly stochastic if all the row sums also equal* 1*. Find the steady-state vector for a doubly stochastic matrix.*

Problem 567 *Consider the* 2×2 *stochastic matrix* $P = \begin{bmatrix} 1-p & q \\ p & 1-q \end{bmatrix}$*, where* $0 < p < 1$ *and* $0 < q < 1$*.*

1. *Show that* $\frac{1}{p+q}\begin{bmatrix} q \\ p \end{bmatrix}$ *is the steady-state vector for* P*.*
2. *Show that* P^m *converges to the matrix* $\frac{1}{p+q}\begin{bmatrix} q & q \\ p & p \end{bmatrix}$ *by first verifying inductively that* $P^m = \frac{1}{p+q}\begin{bmatrix} q & q \\ p & p \end{bmatrix} + \frac{(1-p-q)^m}{p+q}\begin{bmatrix} p & -q \\ -p & q \end{bmatrix}$ *for* $m = 1, 2, \ldots$*. (It can be shown that the sequence of powers* $P, P^2, P^3, \ldots$ *of any regular transition matrix converges to the matrix each of whose columns equals the steady-state vector for* P*.)*

Text Source

This application was adapted from Section 2.9 of Keith Nicholson's *Linear Algebra with Applications*. (CC-BY-NC-SA)

W. Keith Nicholson, *Linear Algebra with Applications*, Lyryx 2018, Open Edition, p. 134

83 Curve Fitting

We know that two points determine a line. Do you know how many points determine a quadratic function of the form $f(x) = ax^2 + bx + c$? Given any number of points in the plane, is it always possible to find a polynomial function whose graph contains every one of the given points? To address these questions we will start with an alternative way of finding an equation of a line.

Exploration 568 *Consider two points $A(-3, 2)$ and $B(1, -2)$. We will find a function f whose graph is a line that passes through these points. We know that $f(x) = ax + b$ for some constants a and b. Because the graph of f passes through A and B, we must have the following:*

$$\begin{array}{rcrcrcr} f(-3) & = & a(-3) & + & b & = & 2 \\ f(1) & = & a(1) & + & b & = & -2 \end{array}$$

To solve for a and b, we need to solve the following matrix equation:

$$\begin{bmatrix} \ldots\ldots & \ldots\ldots \\ \ldots\ldots & \ldots\ldots \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} \ldots\ldots \\ \ldots\ldots \end{bmatrix}$$

Solving the equation, we find that $a = -1$ and $b = -1$. This gives us:

$$f(x) = -x - 1$$

The GeoGebra interactive below shows two points A and B, together with the matrix equation that produces function coefficients for the function whose graph passes through A and B. Drag the points around the plane to see how the matrix equation changes.

Access GeoGebra interactives through the online version of this text at

https://ximera.osu.edu/oerlinalg.

From a purely formal standpoint, we observe that the matrix equation has the form:

$$\begin{array}{l} \textit{point } A \to \\ \textit{point } B \to \end{array} \begin{bmatrix} x\textit{-coordinate} & 1 \\ x\textit{-coordinate} & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} y\textit{-coordinate} \\ y\textit{-coordinate} \end{bmatrix}$$

where each row corresponds to one point.

Now we are ready to move to quadratic, and higher degree polynomial functions.

Linear function in Exploration **??** had two unknown coefficients that we needed to find in order to determine the function. Two points gave us a system of two equations and two unknowns.

A quadratic polynomial function, whose graph is a parabola, is given by:

$$f(x) = ax^2 + bx + c$$

Three unknown coefficients will require three points to determine them.

Exploration 569 *We will find a quadratic function of the form $f(x) = ax^2 + bx + c$ whose graph passes through $A(-2, 2)$, $B(0, -1)$ and $C(1, 5)$. To do this, we need to find coefficients a, b and c such that*

$$\begin{array}{rcrcrcrcr} f(-2) & = & a(-2)^2 & + & b(-2) & + & c & = & 2 \\ f(0) & = & a(0)^2 & + & b(0) & + & c & = & -1 \\ f(1) & = & a(1)^2 & + & b(1) & + & c & = & 5 \end{array}$$

The following GeoGebra interactive shows points A, B, and C, together with the matrix equation, and its solution.

Access GeoGebra interactives through the online version of this text at

https://ximera.osu.edu/oerlinalg.

Drag the points around the plane to observe changes in the coefficient matrix. Think geometrically to find locations of A, B and C such that

- *$a = b = 0$; $c \neq 0$.*
- *$a = 0$; $b, c \neq 0$.*

Observe the structure of the matrix equation.

$$\begin{array}{l} \textit{point } A \to \\ \textit{point } B \to \\ \textit{point } C \to \end{array} \begin{bmatrix} (x\textit{-coordinate})^2 & x\textit{-coordinate} & 1 \\ (x\textit{-coordinate})^2 & x\textit{-coordinate} & 1 \\ (x\textit{-coordinate})^2 & x\textit{-coordinate} & 1 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} y\textit{-coordinate} \\ y\textit{-coordinate} \\ y\textit{-coordinate} \end{bmatrix}$$

In general (provided that no one point lies directly above another), given n points, we can always find an $(n-1)$-degree polynomial function whose graph contains every one of the given points. To find such a polynomial function, given by $f(x) = a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$, we need to solve a system of n equations with n unknowns which translates into the following matrix equation.

$$\begin{bmatrix} \vdots & \vdots & \vdots & \vdots \\ (x\text{-coordinate})^{n-1} & \ldots & x\text{-coordinate} & 1 \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix} \begin{bmatrix} a_{n-1} \\ \vdots \\ a_1 \\ a_0 \end{bmatrix} = \begin{bmatrix} \vdots \\ y\text{-coordinate} \\ \vdots \end{bmatrix} \tag{115}$$

In Practice Problems **??** and **??** you will show that the matrix equation in (**??**) has a unique solution if and only if no two of the given points share an x-coordinate.

Using Technology

Throughout this section we have omitted the tedious process of solving matrix equations. It is useful to practice solving smaller matrix equations by hand, but for larger matrices, we can use technology. Below is the Octave code that can be used to find the solution to Exploration **??**. You will be able to modify this code to solve some of the practice problems.

```
% Define the coefficient matrix
A=[4 -2 1;
0 0 1;
1 1 1];

% Define vector b
b=[2;-1;5];

% Find the solution
x=inv(A)*b
```

Link to code.

For more information on using Octave, please refer to the Octave Tutorial section of this text.

On the Dangers of Overfitting

It is exciting to know that we can fit a function to a set of data points, but before we get carried away fitting a 299-degree polynomial function to 300 points, let's consider the following situation.

In the GeoGebra interactive below, you can see that points A - F form a somewhat linear pattern. A linear model can be used to describe these points. Click on the "Display linear model" check-box to see the trend line. (You will learn how to find such models in Least-Squares Approximation). You can see that even though the line does not pass through any of the given points, it fits the overall pattern of the points and can be used to estimate the y-coordinates of other points whose x-coordinates fall within the limits of the scatter plot.

It might be tempting to think that we can find a better model by finding a 5^{th}-degree polynomial function whose graph contains every one of the six points. Click on the "Display 5th degree poly model" check-box to see the alternative model. Can this model be successfully used to make predictions?

Try moving individual points around to see how their placement affects the line and the curve.

Access GeoGebra interactives through the online version of this text at

https://ximera.osu.edu/linearalgebradzv3/.

Any modeling process which insists on fitting the existing data points exactly, at the risk of failing to predict future observations, is referred to as *overfitting*. While sometimes it is beneficial to have a curve that passes through specific points, more often it is the trend, not the individual instances, that we try to capture. We will return to this topic in Least-Squares Approximation.

Practice Problems

Problems **??-??**

Find a polynomial function of an appropriate degree that passes through the given points.

Problem 570

$$(-3, 0), (2, -2), (-1, 1)$$

$$f(x) = \ldots\ldots$$

Plot the graph of f in the Desmos window below.

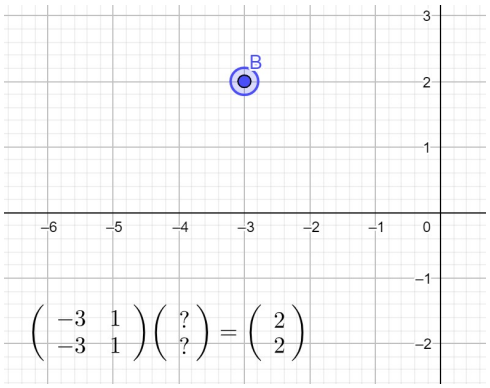
Problem 571

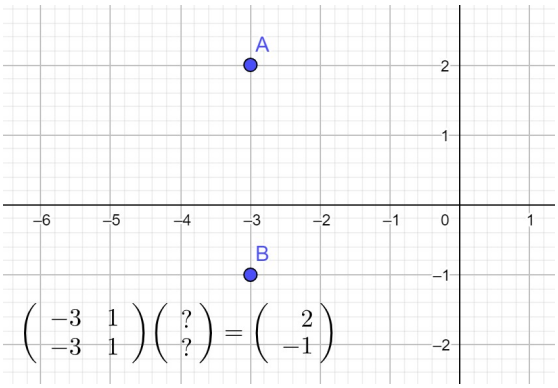
$$(0, 4), (-2, 2), (-1, 1), (0.5, 1)$$

$$g(x) = \ldots\ldots$$

Plot the graph of g in the Desmos window below.

Problem 572 *Two GeoGebra screenshots are shown below:*





In the first screenshot, points A and B coincide. In the second screenshot, point A is located directly above point B. In both cases, GeoGebra failed to produce a linear function whose graph passes through A and B.

Based on what you know about functions and geometry, explain why the process fails for these two examples. How do your observations correspond to what happens from an algebraic standpoint?

Multiple Choice:

1. *Both systems are inconsistent.*
2. *The first system is inconsistent, the second has infinitely many solutions.*
3. *Both systems have infinitely many solutions.*
4. *The first system has infinitely many solutions, the second system is inconsistent.*

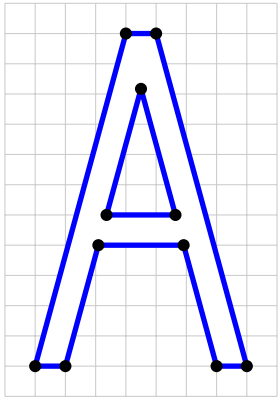
Problem 573 *Prove that equation (**??**) has a unique solution if and only if no two given points share an x-coordinate. Under what circumstances is a solution not unique? Under what circumstances does a solution not exist?*

84 Application to Computer Graphics

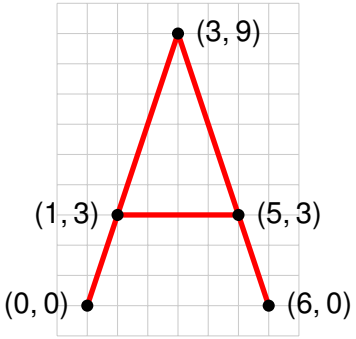
Computer graphics deals with images displayed on a computer screen, and so arises in a variety of applications, ranging from word processors, to *Star Wars* animations, to video games, to wire-frame images of an airplane. These images consist of a number of points on the screen, together with instructions on how to fill in areas bounded by lines and curves. Often curves are approximated by a set of short straight-line segments, so that the curve is specified by a series of points on the screen at the end of these segments. Matrix transformations are important here because matrix images of straight line segments are again line segments.

Note that a color image requires that three images are sent, one to each of the red, green, and blue phosphorus dots on the screen, in varying intensities.

Consider displaying the letter *A*. In reality, it is depicted on the screen, as in the figure below, by specifying the coordinates of the 11 corners and filling in the interior.



For simplicity, we will disregard the thickness of the letter, so we require only five coordinates as in the figure below.



This simplified letter can then be stored as a data matrix

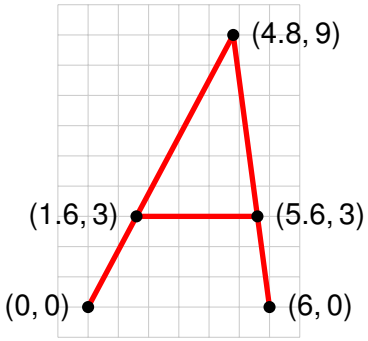
$$D = \begin{bmatrix} 0 & 6 & 5 & 1 & 3 \\ 0 & 0 & 3 & 3 & 9 \end{bmatrix}$$

where the columns are the coordinates of the vertices. Then if we want to transform the letter by a 2×2 matrix A, we left-multiply this data matrix by A (the effect is to multiply each column by A and so transform each vertex).

For example, we can slant the letter to the right by multiplying by a horizontal shear matrix $A = \begin{bmatrix} 1 & 0.2 \\ 0 & 1 \end{bmatrix}$ (See Formula **??**.) The result is the letter with data matrix

$$A = \begin{bmatrix} 1 & 0.2 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 6 & 5 & 1 & 3 \\ 0 & 0 & 3 & 3 & 9 \end{bmatrix} = \begin{bmatrix} 0 & 6 & 5.6 & 1.6 & 4.8 \\ 0 & 0 & 3 & 3 & 9 \end{bmatrix}$$

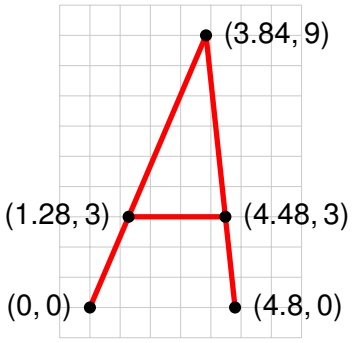
which is shown in the figure below.



If we want to make this slanted matrix narrower, we can now apply a scaling matrix $B = \begin{bmatrix} 0.8 & 0 \\ 0 & 1 \end{bmatrix}$ that shrinks the x-coordinate by 0.8. (See Formula **??**.) The result is the composite transformation

$$\begin{aligned} BAD &= \begin{bmatrix} 0.8 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0.2 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 6 & 5 & 1 & 3 \\ 0 & 0 & 3 & 3 & 9 \end{bmatrix} \\ &= \begin{bmatrix} 0 & 4.8 & 4.48 & 1.28 & 3.84 \\ 0 & 0 & 3 & 3 & 9 \end{bmatrix} \end{aligned}$$

which is drawn in the figure below.



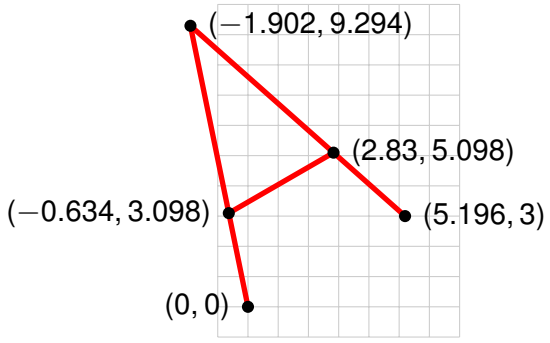
On the other hand, we can rotate the letter about the origin through $\frac{\pi}{6}$ (or 30°) by multiplying by the rotation matrix

$$R_{\frac{\pi}{2}} = \begin{bmatrix} \cos(\frac{\pi}{6}) & -\sin(\frac{\pi}{6}) \\ \sin(\frac{\pi}{6}) & \cos(\frac{\pi}{6}) \end{bmatrix} = \begin{bmatrix} 0.866 & -0.5 \\ 0.5 & 0.866 \end{bmatrix}$$

(See Formula **??**.) This gives

$$\begin{aligned} R_{\frac{\pi}{2}} &= \begin{bmatrix} 0.866 & -0.5 \\ 0.5 & 0.866 \end{bmatrix} \begin{bmatrix} 0 & 6 & 5 & 1 & 3 \\ 0 & 0 & 3 & 3 & 9 \end{bmatrix} \\ &= \begin{bmatrix} 0 & 5.196 & 2.83 & -0.634 & -1.902 \\ 0 & 3 & 5.098 & 3.098 & 9.294 \end{bmatrix} \end{aligned}$$

and is plotted in the figure below.



This poses a problem: How do we rotate about a point other than the origin? It turns out that we can do this when we have solved another more basic problem. It is clearly important to be able to translate a screen image by a fixed vector **w**, that is apply the transformation $T_w : \mathbb{R}^2 \to \mathbb{R}^2$ given by $T_w(\mathbf{v}) = \mathbf{v} + \mathbf{w}$ for all $\mathbf{v}$ in $\mathbb{R}^2$. The problem is that these translations are not matrix

transformations $\mathbb{R}^2 \to \mathbb{R}^2$ because they do not carry $\mathbf{0}$ to $\mathbf{0}$ (unless $\mathbf{w} = \mathbf{0}$). (See Practice Problem **??**.) However, there is a clever way around this.

The idea is to represent a point $\mathbf{v} = \begin{bmatrix} x \\ y \end{bmatrix}$ as a 3×1 column $\begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$, called the *homogeneous coordinates* of $\mathbf{v}$. Then translation by $\mathbf{w} = \begin{bmatrix} p \\ q \end{bmatrix}$ can be achieved by multiplying by a 3×3 matrix:

$$\begin{bmatrix} 1 & 0 & p \\ 0 & 1 & q \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} x+p \\ y+q \\ 1 \end{bmatrix} = \begin{bmatrix} T_{\mathbf{w}}(\mathbf{v}) \\ 1 \end{bmatrix}$$

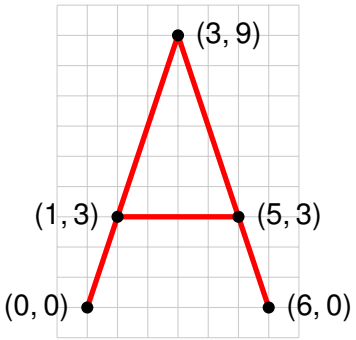
Thus, by using homogeneous coordinates we can implement the translation T_w in the top two coordinates. At the same time, matrix transformations we performed earlier can still be accomplished using homogeneous coordinates. For instance, the matrix transformation induced by $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is given by a 3×3 matrix:

$$\begin{bmatrix} a & b & 0 \\ c & d & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} ax+by \\ cx+dy \\ 1 \end{bmatrix} = \begin{bmatrix} A\mathbf{v} \\ 1 \end{bmatrix}$$

So everything can be accomplished at the expense of using 3×3 matrices and homogeneous coordinates.

It is interesting to examine the geometric implications of homogeneous coordinates. All points of the form $\begin{bmatrix} x \\ y \end{bmatrix}$, originally located in the xy-plane in $\mathbb{R}^2$, now have the form $\begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$ and are located in a horizontal plane $z = 1$ in $\mathbb{R}^3$. What we perceive as a translation in the horizontal plane $z = 1$, is actually a shear in $\mathbb{R}^3$. This is similar to how horizontal shears in $\mathbb{R}^2$ affect points along the line $y = 1$: all such points move a fixed distance to the left or to the right.

Example 212. Rotate the letter A through $\frac{\pi}{6}$ about the point $\begin{bmatrix} 4 \\ 5 \end{bmatrix}$.



Explanation. Using homogenous coordinates for the vertices of the letter results in a data matrix with three rows:

$$K_d = \begin{bmatrix} 0 & 6 & 5 & 1 & 3 \\ 0 & 0 & 3 & 3 & 9 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

If we write $\mathbf{w} = \begin{bmatrix} 4 \\ 5 \end{bmatrix}$, the idea is to use a composite of transformations: First translate the letter by $-\mathbf{w}$ so that the point $\mathbf{w}$ moves to the origin, then rotate this translated letter, and then translate it by $\mathbf{w}$ back to its original position. The following GeoGebra interactive illustrates this sequence.

Access GeoGebra interactives through the online version of this text at

https://ximera.osu.edu/oerlinalg.

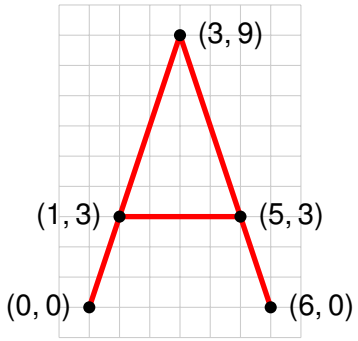
The matrix arithmetic is as follows (remember the order of composition!):

$$\begin{bmatrix} 1 & 0 & 4 \\ 0 & 1 & 5 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0.866 & -0.5 & 0 \\ 0.5 & 0.866 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & -4 \\ 0 & 1 & -5 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 6 & 5 & 1 & 3 \\ 0 & 0 & 3 & 3 & 9 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$
$$= \begin{bmatrix} 3.036 & 8.232 & 5.866 & 2.402 & 1.134 \\ -1.33 & 1.67 & 3.768 & 1.768 & 7.964 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

This discussion merely touches the surface of computer graphics, and the reader is referred to specialized books on the subject. Realistic graphic rendering requires an enormous number of matrix calculations. In fact, matrix multiplication algorithms are now embedded in microchip circuits, and can perform over 100 million matrix multiplications per second. This is particularly important in the field of three-dimensional graphics where the homogeneous coordinates have four components and 4×4 matrices are required.

Practice Problems

Problem 574 *Consider the letter A below.*



Find the data matrix for the letter obtained by:

1. *Rotating the letter through* $\frac{\pi}{4}$ *about the origin.*
2. *Rotating the letter through* $\frac{\pi}{4}$ *about the point* $\begin{bmatrix} 1 \\ 2 \end{bmatrix}$.

   Click the arrow to see the answer. $\frac{1}{2}\begin{bmatrix} \sqrt{2}+2 & 7\sqrt{2}+2 & 3\sqrt{2}+2 & -\sqrt{2}+2 & -5\sqrt{2}+2 \\ -3\sqrt{2}+4 & 3\sqrt{2}+4 & 5\sqrt{2}+4 & \sqrt{2}+4 & 9\sqrt{2}+4 \\ 2 & 2 & 2 & 2 & 2 \end{bmatrix}$

Problem 575 *Find the* 3×3 *matrix for reflecting in the line* $y = mx + b$. *Use* $\begin{bmatrix} 1 \\ m \end{bmatrix}$ *as direction vector for the line.*

Problem 576 *Find the* 3×3 *matrix for rotating through the angle* θ *about the point* $P(a, b)$.

Problem 577 *Find the reflection of the point* P *in the line* $y = 1 + 2x$ *in* $\mathbb{R}^2$ *if:*

1. $P = P(1, 1)$
2. $P = P(1, 4)$

   Click the arrow to see the answer. $P(\frac{9}{5}, \frac{18}{5})$
3. *What about* $P = P(1, 3)$*? Explain.*

85 Octave Exercises

86 Using Octave

Why Octave?

In writing *Linear Algebra, An Interactive Introduction*, we drew upon many years of experience teaching elementary linear algebra. We have found that when students experiment with scientific computing software such as MATLAB, the subject comes alive! Students can quickly recognize patterns, form conjectures, and test their theories.

We recognize that not every learning institution has the funding to purchase software for student use. For this reason we chose to use Octave as our computational tool. GNU Octave is a programming language compatible with MATLAB. GNU Octave can be downloaded for free at https://www.octave.org/download or executed in SageCells online. Online SageCells and a desktop version of Octave can be equally effective in enhancing the experience in linear algebra.

The purpose of this Octave tutorial is to gather commands and examples useful for this linear algebra course into one place. Students and instructors may find this to be a useful reference as they proceed through the course. Please refer to the Octave manual for additional support.

How to Use SageCells

We present code in two different ways. First, the code appears as text within the narrative, like this:

```
disp('Hello, world!');
```

This code can be copied and pasted into MATLAB or Octave. Second, we provide permalinks to our code in SageCells, like this:

Link to code.

Users should feel free to modify our code inside SageCells. To save their work, users should generate their own permalink. Permalinks should be stored to access the code at a later time and

to share the code with others. Note that every time the code is modified a new permalink must be generated to reflect the changes.

When writing Octave code in SageCells from scratch, make sure to choose *Octave* as your language from the drop down menu below the cell.

This 3:00 video explains how to use permalinks. The key step is just past the 2:15 mark.

YouTube link: `https://www.youtube.com/watch?v=NhujjT-p2D0`

Getting Help

The following link provides documentation for linear algebra functions used in Octave. Link to linear algebra functions documentation.

To get immediate help inside of a SageCell or in Octave Desktop, you can use the help command, as follows:

```
help eig
```

87 Octave for Chapter 1

Octave for Chapter 1

The templates in this section provide sample Octave code for vector operations. You can access our code through the link at the bottom of each template. Feel free to modify the code and experiment to learn more!

You can write your own code using Octave software or online Octave cells. To access Octave cells online, go to the Sage Math Cell Webpage, select OCTAVE as the language, enter your code, and press EVALUATE.

To ”save” or share your online code, click on the *Share* button, select *Permalink*, then copy the address directly from the browser window. You can store this link to access your work later or share this link with others. You will need to get a new Permalink every time you modify the code.

Octave Tutorial

Basic Operations

Template 1. Here are some basic vector operations.

```
% Define vectors v and w
v=[3;1;1;2;4];
w=[-2;1;0;3;-1];

% Find v+w
sum=v+w

% Find 5v
five_v=5*v

% Find the dot product of v and w
dot_v_w=dot(v,w)

% Find the length of v
length_v=norm(v)
```

Link to code.

Remark 23.
- In Octave, we use % to add comments to our code. Octave does not execute anything that appears after % on the same line. It is always a good idea to make notes about your code for yourself and others.
- Note that lines 2 and 3 of our code have a semi-column at the end while lines 6, 9, 12, and 15 do not. Remove the semi-column from lines 2 and 3, add semi-columns to the other lines. Run the code. What happens?

Template 2. We can find the cross-product of two vectors in $\mathbb{R}^3$ as follows.

```
% Define vectors v and w
v=[3;1;1];
w=[-2;1;0];

% Find the cross product v x w
cross_v_w=cross(v,w)
```

Link to code.

Example 213. Find the angle between vectors $\mathbf{v} = \begin{bmatrix} 3 \\ 1 \\ 1 \\ 2 \\ 4 \end{bmatrix}$ and $\mathbf{w} = \begin{bmatrix} -2 \\ 1 \\ 0 \\ 3 \\ -1 \end{bmatrix}$.

Explanation.
```
% Define vectors v and w
v=[3;1;1;2;4];
w=[-2;1;0;3;-1];

% Find the angle between v and w in radians and degrees.
% To do so, use the arccosine function acos()
angle_radians = acos(dot(v,w)/(norm(v)*norm(w)))
angle_degrees = rad2deg(angle_radians)
```

Link to code.

Sometimes it is useful to reference a single component of a vector.

Example 214. Given vector $\mathbf{v} = \begin{bmatrix} 1 \\ -2 \\ -5 \end{bmatrix}$, find the third component of $\mathbf{v}$.

Explanation.
```
% Define vector v
v=[1;-2;-5];

% Find the third component of v
z=v(3)
```

Link to code.

Loops

In this section we will introduce one of the most fundamental concepts in programming, a *loop*. Loops are used to repeat a procedure multiple times. We will start with a simple *for* loop. Every time you perform the steps inside the loop, you come back to the beginning, and increase the index *i* by 1 until you reach the desired number of iterations.

Example 215. Here is an example of a loop. Try to guess the outcome of the loop, then run the code to verify your guess.

```
% Enter the starting value
n=10;

% At every iteration, n gets replaced with n+1
for i=1:10
    n=n+1
end
```

Link to code.

Example 216. Here is another loop. Try to guess the outcome of the loop, then run the code to verify your guess.

```
% Enter the first component of vector v
v(1)=1;

% Enter the second component of vector v
```

```
v(2)=1;

% we continue to assign values to vector components
for i=1:10
    v(i+2)=v(i)+v(i+1);
end
v
```

Link to code.

Do you recognize this famous sequence?

Octave Exercises

Problem 578 *Use Octave to find the angle (in degrees and radians) between vectors*

$$\mathbf{v} = \begin{bmatrix} 2 \\ -4 \\ 10 \\ -25 \end{bmatrix}, \quad \mathbf{w} = \begin{bmatrix} 5 \\ 3 \\ -20 \\ 11 \end{bmatrix}$$

Problem 579 *Write Octave code to show that the points $A(1, -2, 3, 5)$, $B(-2, 0, 1, 1)$, $C(2, 4, -3, 2)$ are the vertices of a right triangle in $\mathbb{R}^4$.*

Problem 580 *Use vectors* **v** *and* **w** *from Template* **??** *to illustrate that the cross product of two vectors is orthogonal to both vectors. Pick another pair of vectors* **v** *and* **w** *and repeat the demonstration.*

Problem 581 *Write an Octave routine that allows the user to input three points and returns a normal vector to the plane containing these points. Your code might start as follows:*

```
% Enter coordinates of point A
A=[1; 4; 6];

% Enter coordinates of point B
B=[-2; 3; 7]
```

```
% Enter coordinates of point C
C=[1; -2; 4]

% Normal vector to the plane containing A, B, C is given by
```

What type of user inputs (other than all zeros) would result in the computed normal vector of $\mathbf{n} = \mathbf{0}$*? Explain this phenomenon geometrically.*

Use your code to find an equation of a plane containing points $(-2, 4, 0)$*,* $(-1, 8, 1)$*,* $(3, -1, 2)$*.*

Problem 582 *In this problem you will use the following definition of the dot product.*

Definition 103. (**??**) Let $\mathbf{u}$ and $\mathbf{v}$ be vectors in $\mathbb{R}^n$. The *dot product* of $\mathbf{u}$ and $\mathbf{v}$, denoted by $\mathbf{u} \cdot \mathbf{v}$, is given by

$$\mathbf{u} \cdot \mathbf{v} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} \cdot \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = u_1 v_1 + u_2 v_2 + \ldots + u_n v_n.$$

Question 582.1 *Due to a false fire alarm, a programmer left line 10 of the following code unfinished. Finish the code so that it computes the dot product (dp) of two vectors. Note that this code uses Definition* **??** *directly, instead of relying on the built-in* dot *function.*

```
% Define vectors v and w
v=[3; 6; -1; 7; 12];
w=[-2; 4; -3; 5; 11];

% Initialize the dot product (dp)
dp=0;

% Find the dot product v*w directly from the definition
for i=1:5
    dp=dp+;
end

% Print your answer
dp
```

```
% Check your answer
correct_answer=dot(v,w)
```

Link to code.

Why did we have to initialize dp*? What would happen if we didn't?*

88 Octave for Chapter 2

Octave for Chapter 2

The templates in this section provide sample Octave code for solving systems of equations. You can access our code through the link at the bottom of each template. Feel free to modify the code and experiment to learn more!

You can write your own code using Octave software or online Octave cells. To access Octave cells online, go to the Sage Math Cell Webpage, select OCTAVE as the language, enter your code, and press EVALUATE.

To ”save” or share your online code, click on the *Share* button, select *Permalink*, then copy the address directly from the browser window. You can store this link to access your work later or share this link with others. You will need to get a new Permalink every time you modify the code.

Octave Tutorial

When solving systems by hand, you utilized Gauss-Jordan elimination to find rref. Having the reduced row echelon form available made it easy to write out the solution(s) or determine that the system is inconsistent. We can find rref and the rank of a matrix using Octave as follows.

Template 3. Consider the system

$$\begin{array}{rcrcrcrcl} x & - & y & & & & & = & 0 \\ 2x & - & 2y & + & z & + & 2w & = & 4 \\ & & y & & & + & w & = & 0 \\ & & & & 2z & + & w & = & 5 \end{array} \tag{116}$$

(You may remember this system from Augmented Matrix Notation and Elementary Row Operations.) We will solve this system by finding the reduced row echelon form of the corresponding augmented matrix in Octave.

```
% Define the augmented matrix
% corresponding to the coefficient matrix A and vector b
A_b = [1 -1 0 0 0;
```

```
     2 -2 1 2 4;
     0 1 0 1 0;
     0 0 2 1 5];

% Find rref of A_b
rref(A_b)

% Find the rank of A_b
rank(A_b)
```

Link to code.

If you start with a matrix A and a separate vector $\mathbf{b}$, you can create an augmented matrix without re-typing as follows:

```
% Define the coefficient matrix A
A = [1 -1 0 0;
     2 -2 1 2;
     0 1 0 1;
     0 0 2 1];

% Define vector b
b = [0;4;0;5];

% Create an augmented matrix
A_b=[A b]
```

Link to code.

Template 4. We can access individual rows, columns, and entries of a matrix. Let

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

```
% Define a matrix
A=[2 -1 3 -1; 1 0 2 1; 1 -1 1 -2]

% Third row of A
```

```
row_3=A(3,:)

% Second column of A
col_2=A(:,2)

% Second row third entry of A
a_23=A(2,3)
```

Link to code.

Example 217. Use Octave code to show one step of Gaussian elimination by subtracting half of the first row from the second row of the given matrix *A*.

Explanation.
```
% Define matrix A
A=[2 -1 3 -1; 1 0 2 1; 1 -1 1 -2]

% Replace the second row of A with second row minus half the first row.
A(2,:)=-1/2*A(1,:)+A(2,:)
```

Link to code.

Octave Exercises

Problem 583 *Given the code below, write out the system of equations and the corresponding solution(s).*

```
% Coefficient matrix A
A=[2 -4 1 1 3;
0 2 -1 4 -1;
2 -2 1 1 4];

% Column vector b
b=[-2; 1; 4];

% Augmented matrix A_b
A_b=[A b];

% Reduced row-echelon form
rref_A_b=rref(A_b)
```

Link to code.

Problem 584 *Solve the following system using Octave.*

$$\begin{array}{rcrcrcrcrcr} x_1 & - & x_2 & + & 2x_3 & & & + & x_5 & = & -1 \\ -x_1 & & & - & x_3 & + & x_4 & & & = & 1 \\ 2x_1 & + & 2x_2 & & & - & x_4 & + & 2x_5 & = & 1 \\ x_1 & & & & & + & 2x_4 & - & x_5 & = & 0 \end{array} \tag{117}$$

Problem 585 *Solve the following system using Octave.*

$$\begin{array}{rcrcrcrcrcr} 3x_1 & + & x_2 & + & 4x_3 & - & x_4 & + & 2x_5 & = & 10 \\ x_1 & & & + & 2x_3 & - & x_4 & + & x_5 & = & 3 \\ & & x_2 & + & x_3 & + & x_4 & + & 3x_5 & = & 2 \\ 2x_1 & - & x_2 & - & x_3 & & & + & 3x_5 & = & -2 \\ 4x_1 & + & x_2 & - & 5x_3 & + & 2x_4 & + & x_5 & = & 1 \end{array} \tag{118}$$

Problem 586 *Suppose the graph of function f of the form $f(x) = ax^3 + bx^2 + cx + d$ passes through points $(-2, 1)$, $(0, 4)$, $(1, -2)$ and $(4, 5)$. Set up a system of linear equations to find the coefficients a, b, c and d. Update the coefficients in the GeoGebra interactive below to sketch the graph.*

GeoGebra link: https://www.geogebra.org/m/xry6un6h

Problem 587 *The following code generates a random 20×3 matrix A whose entries are integers -1, 0, and 1, and a random 20×1 vector $\mathbf{b}$ whose components are integers -1, 0, and 1.*

```
% A is a randomly generated 20 by 3 matrix whose entries are -1, 0, 1.
A=randi([-1,1],20,3);

% b is a randomly generated column vector whose entries are -1, 0, 1.
b=randi([-1,1],20,1);

% Augmented matrix A_b
A_b=[A b];

% Reduced row-echelon form
```

```
rref_A_b=rref(A_b)

% Rank of A_b
rank(A_b)
```

Link to code.

Run the code several times to generate several random combinations of A and **b** *and find the rank and the reduced row-echelon form of* [*A*|**b**]*. Are you getting the answers you were expecting? Answer the following questions based on your* theoretical *knowledge.*

What are the possibilities for the rank of such [*A*|**b**]*?*

Multiple Choice:

1. *The rank can be any positive integer less than or equal to* 20*.*
2. *The rank can be any positive integer less than or equal to* 3*.*
3. *The rank is always equal to* 4*.*
4. *The rank can be any non-negative integer less than or equal to* 4*.*

What are the possibilities for the number of solutions of such [*A*|**b**]*?*

Multiple Choice:

1. *This system always has a unique solution.*
2. *This system may have no solutions, infinitely many solutions or a unique solution.*
3. *This system will always have infinitely many solutions.*
4. *This system will always be infeasible.*

There is a good chance that your experiments have led you to believe that rank[*A*|**b**] = 4*. Can you manually construct* [*A*|**b**]*, satisfying the condition of the problem, with a smaller rank? How many solutions does your system have? Why do you think it is so rare to randomly generate a system* [*A*|**b**] *of rank less than* 4*?*

Further Considerations

Issues related to computational aspects of linear algebra are outside the scope of this text. However, we want the reader to be aware of some potential problems that may arise.

Warning 4. The *rref* function utilizes a modified version of the Gauss-Jordan elimination algorithm (Reference Link). For some matrices, implementing this algorithm directly leads to round-off errors and other problems that are outside of the scope of this text. We illustrate what can happen in the next example.

Example 218. We will use a 12×12 Hilbert matrix to illustrate how the *rref* function can fail. You can learn more about Hilbert matrices here.

```
% Use a 12 by 12 Hilbert matrix
A=hilb(12);

% Theoretical rank of A is 12, but the following computation shows a rank of 11
rank(A)

% rref of A should be the identity matrix, but the following result contains a row o
rref(A)
```

Link to code.

Acknowledgements

The inspiration for Example **??** came from this *Stack Overflow* comment.

89 Octave for Chapter 3

Octave for Chapter 3

Examples in this section provide sample Octave code for answering questions about linear combinations, span, and linear independence. You can access our code through the link at the bottom of each template. Feel free to modify the code and experiment to learn more!

You can write your own code using Octave software or online Octave cells. To access Octave cells online, go to the Sage Math Cell Webpage, select OCTAVE as the language, enter your code, and press EVALUATE.

To "save" or share your online code, click on the *Share* button, select *Permalink*, then copy the address directly from the browser window. You can store this link to access your work later or share this link with others. You will need to get a new Permalink every time you modify the code.

Octave Tutorial

Example 219. If possible, express each of $\mathbf{b}_1$ and $\mathbf{b}_2$ as a linear combination of vectors in W.

$$\mathbf{b}_1 = \begin{bmatrix} 0 \\ 4 \\ 1 \\ -1 \\ 7 \end{bmatrix}, \quad \mathbf{b}_2 = \begin{bmatrix} -4 \\ 26 \\ -11 \\ -41 \\ 10 \end{bmatrix}, \quad W = \left\{ \begin{bmatrix} 3 \\ 6 \\ -1 \\ 0 \\ 4 \end{bmatrix}, \begin{bmatrix} 1 \\ -2 \\ 3 \\ 5 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 4 \\ -2 \\ 3 \end{bmatrix} \right\}$$

Explanation. For each of $\mathbf{b}_1$ and $\mathbf{b}_2$ we will attempt to solve the system

$$a\mathbf{w}_1 + b\mathbf{w}_2 + c\mathbf{w}_3 = \mathbf{b}_i$$

where $\mathbf{w}_1$, $\mathbf{w}_2$, $\mathbf{w}_3$ are vectors of W.

We will use the following code.

```
% Let A_b1 and A_b2 be augmented matrices
% whose columns are the vectors of W
% together with vectors b1 and b2, respectively.
```

```
w1=[3; 6;-1; 0; 4];
w2=[1; -2; 3; 5; 1];
w3=[-1; 0; 4; -2; 3];

b1=[0; 4; 1; -1; 7];
b2=[-4; 26; -11; -41; 10];

A_b1=[w1 w2 w3 b1];
A_b2=[w1 w2 w3 b2];

% Reduced row-echelon form
rref_A_b1=rref(A_b1)
rref_A_b2=rref(A_b2)
```

Link to code.

The reduced row-echelon form of $[A|\mathbf{b}_1]$ looks like this:

```
rref_A_b1 =

1 0 0 0
0 1 0 0
0 0 1 0
0 0 0 1
0 0 0 0
```

The row $[000|1]$ indicates that the system is infeasible, so $\mathbf{b}_1$ cannot be written as a linear combination of vectors in W.

The reduced row-echelon form of $[A|\mathbf{b}_2]$ looks like this:

```
rref_A_b2 =

1 0 0 2
0 1 0 -7
0 0 1 3
0 0 0 0
0 0 0 0
```

This shows that the system of equations

$$a\mathbf{w}_1 + b\mathbf{w}_2 + c\mathbf{w}_3 = \mathbf{b}_2$$

has a solution: $a = 2$, $b = -7$, $c = 3$. We write,

$$\mathbf{b}_2 = 2\mathbf{w}_1 - 7\mathbf{w}_2 + 3\mathbf{w}_3$$

Example 220. Determine whether the given vectors are linearly independent.

$$\mathbf{w}_1 = \begin{bmatrix} 38 \\ 15 \\ -11 \end{bmatrix}, \quad \mathbf{w}_2 = \begin{bmatrix} 64 \\ 82 \\ 14 \end{bmatrix}, \quad \mathbf{w}_3 = \begin{bmatrix} -3 \\ 13 \\ 9 \end{bmatrix}$$

Explanation. We will have to determine whether the equation

$$a\mathbf{w}_1 + b\mathbf{w}_2 + c\mathbf{w}_3 = \mathbf{0} \tag{119}$$

has a non-trivial solution. To do this we will consider an augmented matrix $[\mathbf{w}_1\mathbf{w}_2\mathbf{w}_3|\mathbf{0}]$. In the following code, we will skip the zero vector.

```
% Define the vectors individually

w1=[38; 15; -11];
w2=[64; 82; 14];
w3=[-3; 13; 9];

A=[w1 w2 w3];

% Reduced row-echelon form
rref_A=rref(A)
```

Link to code.

We get the following output

```
rref_A =

1 0 -0.5
0 1 0.25
0 0 0
```

We can see that there are infinitely many solutions $a = 0.5t$, $b = -0.25t$, $c = t$. Letting $t = 4$, we write,

$$2\mathbf{w}_1 - \mathbf{w}_2 + 4\mathbf{w}_3 = \mathbf{0}$$

Since (**??**) has a non-trivial solution, we conclude that the vectors are linearly dependent.

Example 221. From the set of vectors below, construct a linearly independent set that has the same span as W.

$$W = \left\{ \begin{bmatrix} 12 \\ 10 \\ -20 \\ 5 \end{bmatrix}, \begin{bmatrix} 3 \\ -14 \\ 21 \\ 11 \end{bmatrix}, \begin{bmatrix} -6 \\ -38 \\ 62 \\ 17 \end{bmatrix}, \begin{bmatrix} 4 \\ -4 \\ 23 \\ 19 \end{bmatrix}, \begin{bmatrix} -6 \\ 6 \\ 47 \\ 30 \end{bmatrix}, \begin{bmatrix} 2 \\ 11 \\ 17 \\ 8 \end{bmatrix} \right\}$$

Explanation. We will refer to vectors of W as $\mathbf{w}_1 \ldots \mathbf{w}_6$, in order of appearance. We will build a linearly independent set from these vectors by adding one vector to our set at a time, and checking for linear independence every time a vector is added. If, after adding one vector, the resulting collection is linearly independent, we will keep the newly added vector, if the collection is not linearly independent we will discard the new vector and move on to the next. The following code accomplishes this.

```
% Define the vectors individually

w1=[12; 10; -20; 5];
w2=[3; -14; 21; 11];
w3=[-6; -38; 62; 17];
w4=[4; -4; 23; 19];
w5=[-6; 6; 47; 30];
w6=[2; 11; 17; 8];

% Clearly {w1} is linearly independent
% Let's add w2 to the set and check.

rref([w1 w2]) % Looks like w1 and w2 are linearly independent

rref([w1 w2 w3]) % w3 is redundant - this set is linearly dependent. We won't ad

rref([w1 w2 w4]) % these are linearly independent

rref([w1 w2 w4 w5]) % w5 is redundant; don't add
```

```
rref([w1 w2 w4 w6])
```

Link to code.

Question 588 *Explain why*

$$span(\mathbf{w}_1 \ldots \mathbf{w}_6) = span(\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_4, \mathbf{w}_6)$$

See discussion below Definition **??** *for guidance.*

By starting with $\mathbf{w}_1$, adding vectors to our set one at a time, and checking for linear independence, we determined that $\mathbf{w}_3$ and $\mathbf{w}_5$ are redundant and can be removed from the set without changing the span.

Octave Exercises

Problem 589 *If possible, express vector* $\mathbf{b}$ *as a linear combination of the vectors in* W*. If you find a linear combination, verify its correctness by hand.*

1.

$$\mathbf{b} = \begin{bmatrix} 2 \\ -1 \\ 0 \\ 4 \end{bmatrix}, \quad W = \left\{ \begin{bmatrix} 1 \\ 0 \\ 1 \\ 2 \end{bmatrix}, \begin{bmatrix} -2 \\ 4 \\ -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \\ 4 \\ -1 \end{bmatrix} \right\}$$

Multiple Choice:

(i) $\mathbf{b}$ *is in* $span(W)$.

(ii) $\mathbf{b}$ *is NOT in* $span(W)$.

2.

$$\mathbf{b} = \begin{bmatrix} 4 \\ -6.5 \\ 4.5 \\ 1.5 \end{bmatrix}, \quad W = \left\{ \begin{bmatrix} 1 \\ 0 \\ 1 \\ 2 \end{bmatrix}, \begin{bmatrix} -2 \\ 4 \\ -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \\ 4 \\ -1 \end{bmatrix} \right\}$$

Multiple Choice:

(i) **b** *is in span*(W).

(ii) **b** *is NOT in span*(W).

Problem 590 *Determine whether each of the following sets of vectors is linearly independent.*

1.

$$\left\{ \begin{bmatrix} -4 \\ -2 \\ 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 8 \\ 2 \\ 3 \\ 5 \\ 7 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 1 \\ 5 \\ -2 \end{bmatrix}, \begin{bmatrix} 5 \\ 2 \\ 1 \\ 3 \\ 2 \end{bmatrix} \right\}$$

2.

$$\left\{ \begin{bmatrix} -4 \\ -2 \\ 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 1 \\ 5 \\ -2 \end{bmatrix}, \begin{bmatrix} 5 \\ 2 \\ 1 \\ 3 \\ 2 \end{bmatrix}, \begin{bmatrix} -1 \\ 6 \\ 1 \\ 0 \\ 4 \end{bmatrix} \right\}$$

Problem 591 *Find the number of redundant vectors in the following set.*

$$\left\{ \begin{bmatrix} 2 \\ 6 \\ -1 \\ 4 \\ 3 \\ 5 \end{bmatrix}, \begin{bmatrix} -1 \\ 6 \\ -11 \\ 9 \\ 4 \\ 4 \end{bmatrix}, \begin{bmatrix} -1 \\ 4 \\ 1 \\ 1 \\ -2 \\ 8 \end{bmatrix}, \begin{bmatrix} 3 \\ 1 \\ 4 \\ -1 \\ 2 \\ -1 \end{bmatrix}, \begin{bmatrix} -1 \\ 32 \\ 3 \\ 13 \\ -4 \\ 50 \end{bmatrix}, \begin{bmatrix} -1 \\ 3 \\ 7 \\ -3 \\ -5 \\ 10 \end{bmatrix} \right\}$$

There are ... ... redundant vectors in the set.

Problem 592 *In Example* **??***, we constructed a linearly independent set from vectors of W by starting with* $\mathbf{w}_1$ *and adding one vector at a time to our set moving from left to right and checking for linear independence at every step. Our resulting set* $\{\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_4, \mathbf{w}_6\}$ *has the same span as the original set W. What if we start this process with* $\mathbf{w}_6$ *and move from right to left to identify and remove redundant vectors? Would we end up with the same final set? Modify the code of Example* **??** *to find out. Does our final set have the same span as W? Prove your claim.*

90 Octave for Chapter 4

Octave for Chapter 4

The templates in this section provide sample Octave code for matrix operations and *LU*-factorization. You can access our code through the link at the bottom of each template. Feel free to modify the code and experiment to learn more!

You can write your own code using Octave software or online Octave cells. To access Octave cells online, go to the Sage Math Cell Webpage, select OCTAVE as the language, enter your code, and press EVALUATE.

To ”save” or share your online code, click on the *Share* button, select *Permalink*, then copy the address directly from the browser window. You can store this link to access your work later or share this link with others. You will need to get a new Permalink every time you modify the code.

Octave Tutorial

Basic Operations

Template 5. Here are some matrix operations.

```
% Define matrices A, B, and C
A = [1 -1 0 0;
    2 -2 1 2;
    0 1 0 1];

B = [2 3 -1 4;
    1 -1 2 -2;
    1 -2 0 3];

C = [1 -2 1 4;
    2 -2 0 1;
    -2 1 1 0;
    1 3 -2 -1];
```

```
% Find A+B
sum=A+B

% Find the transpose of A
A_trans=transpose(A)

% Find AC
product=A*C

% Find the inverse of C
inv_C=inv(C)
```

Link to code.

Template 6. Sometimes it is useful to generate a commonly used matrix or a vector (e.g. identity matrix, zero vector).

```
% 5 by 5 identity matrix
I_5=eye(5)

% 5 by 5 zero matrix
O_5=zeros(5)

% The zero vector in R^5
O_col=zeros(5,1)

% Row of zeros
O_row=zeros(1,5)

% Non-square zero matrix
O_block=zeros(2,3)

% Matrix filled with 1s
One_block=ones(2,3)
```

Link to code.

Using Loops

We first introduced loops in Octave for Chapter 1. Examples in this section and the corresponding exercises will allow you to practice using loops while also encouraging you to delve deeper into various ways of thinking about matrix multiplication.

Example 222. In Matrix Multiplication we gave the followng definition of matrix-vector multiplication.

$$A\mathbf{x} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & \dots & \dots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n \end{bmatrix}$$

Write a routine that uses this definition to compute the product $A\mathbf{x}$ of a user-defined matrix A and vector $\mathbf{x}$.

Explanation. We will use the followng outline for our code.

- The user will enter matrix A and vector $\mathbf{x}$ at the top.
- We will use the *size* function determine the size of the user-defined matrix A.

```
% Find the size of matrix A
[m, n] = size(A);  % m is the number of rows, and n is the number of column
```

- We will utilize the *length* function to find the length of $\mathbf{x}$.

```
% Find the length of vector x
L=length(x);
```

- We will check that the dimensions of A and $\mathbf{x}$ are compatable. If the number of columns of A doesn't match the number of entries in $\mathbf{x}$, we will issue a warning.

```
if n ~= L
    warning("Dimensions of A and x are incompatable.");
else
    % insert product code here ...
end
```

Here is the complete code.

```
% User-defined matrix A
A=[4 -1 6 10 -2;
3 11 8 -3 4;
-5 2 14 6 -1;
3 2 -3 0 7];

% User-defined vector x
x=[-2; 4; 8; 2; 3];

% Find the size of matrix A
[m,n]=size(A); % m is the number of rows, and n is the number of columns of A

% Find the length of vector x
L=length(x);

if n ~= L
    warning("Dimensions of A and x are incompatable.");
else
    % if the dimensions are compatable, proceed with the multiplication

    % Initialize the product matrix
    A_x=zeros(m,1);

    % Find the product Ax
    for i=1:m
        for j=1:n
            A_x(i)=A_x(i)+A(i,j)*x(j);
        end
    end

    % Print our answer
    A_x

    % Check our answer
    correct_answer=A*x
end
```

Link to code.

Example 223. In Matrix Multiplication we noted that a matrix-vector product can also be computed by dotting the rows of the matrix with the vector as follows:

$$A\mathbf{x} = \begin{bmatrix} - & \mathbf{r}_1 & - \\ - & \mathbf{r}_2 & - \\ & \vdots & \\ - & \mathbf{r}_n & - \end{bmatrix} \mathbf{x} = \begin{bmatrix} \mathbf{r}_1 \cdot \mathbf{x} \\ \mathbf{r}_2 \cdot \mathbf{x} \\ \vdots \\ \mathbf{r}_n \cdot \mathbf{x} \end{bmatrix}$$

Demonstrate this method for matrix-vector multiplication in Octave.

Explanation. We will use the following syntax to reference the entire row of a matrix.

```
A(2,:) % this is row 2 of matrix A
```

Alternatively, if we ever need to reference a single column of a matrix, we would use the following:

```
A(:,5) % this is column 5 of matrix A
```

Here is our code that utilizes the dot product to perform matrix-vector multiplication.

```
% Define matrix A
A=[4 -1 6 10 -2;
3 11 8 -3 4;
-5 2 14 6 -1;
3 2 -3 0 7];

% Define vector x
x=[-2; 4; 8; 2; 3];

% Find the product Ax using the dot product method
for i=1:4
    A_x(i,1)=dot(A(i,:),x);
end

% Print our answer
A_x
```

```
% Check our answer
correct_answer=A*x
```

Link to code.

LU-factorization

Before we proceed with the code, recall that if the elementary matrices used to reduce the given matrix to row-echelon form are all lower triangular, then we can find an *LU* factorization. In general, the following theorem applies.

Theorem 130 (**??**)**.** Suppose an $m \times n$ matrix A is transformed to a row-echelon matrix U using Gaussian elimination. Let $P_1, P_2, \ldots, P_s$ be the elementary matrices corresponding (in order) to the row interchanges used, and write $P = P_s \cdots P_2 P_1$. (If no interchanges are used take $P = I$.) Then:

1. PA is the matrix obtained from A by doing these interchanges (in order) to A.
2. PA has an LU factorization.

Template 7. We can find *LU*-factorization of a matrix as follows.

```
% Define matrix A
A=[2 2 -1 2;
   -1 2 2 1;
   1 2 0 -2;
   1 1 -2 -4];

% Find LU factorization for A
[L, U, P] = lu (A)
% where L is a lower triangular matrix,
% U is an upper triangular matrix, and
% P is an appropriate permutation matrix

% Verify that this factorization is correct
productLU=L*U
productPA=P*A
```

Link to code.

Remark 24.
- Observe that the permutation matrix P is not the identity matrix. What elementary row operation does multiplication by P induce?
- With the last two lines of our code we checked the correctness of our factorization. Is it true that $PA = LU$? Interpret the value $-2.77556e-17$ in the product LU. What do you think caused this value to occur?

Octave Exercises

Problem 593 *Let*

$$A = \begin{bmatrix} 2 & 6 & -4 & 1 & 1 \\ -1 & 0 & 2 & 2 & 4 \\ -2 & 1 & 3 & -2 & 1 \\ 1 & 1 & 0 & 0 & 6 \\ -3 & 4 & 1 & -2 & 0 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} -12 \\ 6 \\ -2 \\ -7 \\ -8 \end{bmatrix}$$

Solve $A\mathbf{x} = \mathbf{b}$ *in two different ways. First, use the reduced row-echelon form of* $[A|\mathbf{b}]$*, second multiply both sides of the equation by* A^{-1}*. Verify that both methods produce the same answer.*

Problem 594 *Let* $A = \begin{bmatrix} 0 & 1 \\ 5 & 3 \end{bmatrix}$*. Find* A^{-1} *in two different ways. First, use the* inv *function. Second, find rref*$[A|I]$*.*

Problem 595 *Generate a* 12×14 *matrix filled with* 3*'s.*

Problem 596 *Write a routine that multiplies an* $m \times n$ *matrix by an* $n \times 1$ *vector by computing the linear combination of the columns of the matrix, as in the following definition from Matrix Multiplication:*

$$A\mathbf{x} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & \dots & \dots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = x_1 \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{bmatrix} + x_2 \begin{bmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{bmatrix} + \cdots + x_n \begin{bmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{bmatrix}$$

Problem 597 *Write a routine that multiplies an* $m \times n$ *matrix by an* $n \times k$ *matrix.*

Problem 598 *A Vandermonde matrix of order n is a square matrix of the form*

$$V = \begin{bmatrix} 1 & x_1 & x_1^2 & \dots & x_1^{n-1} \\ 1 & x_2 & x_2^2 & \dots & x_2^{n-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & x_n^2 & \dots & x_n^{n-1} \end{bmatrix}$$

If you studied section Curve Fitting this matrix may look familiar to you.

Use loops to create a 5×5 *Vandermonde matrix for* $x_1 = 2, x_2 = -1, x_3 = 5, x_4 = -2, x_5 = 3$.

Problem 599 *Use LU-factorization to solve the system*

$$\begin{bmatrix} 2 & -2 & 0 & 2 \\ 1 & -2 & 2 & 2 \\ -1 & 2 & 0 & 0 \\ 3 & -3 & 1 & 4 \end{bmatrix} \mathbf{x} = \begin{bmatrix} 4 \\ -6 \\ 4 \\ 5 \end{bmatrix}$$

Text Source

Problem **??** was adapted from Problem 2.7.3(d) of Keith Nicholson's *Linear Algebra with Applications*. (CC-BY-NC-SA)

W. Keith Nicholson, *Linear Algebra with Applications*, Lyryx 2018, Open Edition, p. 127

91 Octave for Chapters 5-6

Octave for Chapters 5-6

The templates in this section provide sample Octave code for finding subspaces associated with matrices as well as the kernel and image of a linear transformation. You can access our code through the link at the bottom of each template. Feel free to modify the code and experiment to learn more!

You can write your own code using Octave software or online Octave cells. To access Octave cells online, go to the Sage Math Cell Webpage, select OCTAVE as the language, enter your code, and press EVALUATE.

To ”save” or share your online code, click on the *Share* button, select *Permalink*, then copy the address directly from the browser window. You can store this link to access your work later or share this link with others. You will need to get a new Permalink every time you modify the code.

Octave Tutorial

Column Space of a Matrix (Image of a Linear Transformation)

Recall that the column space of a matrix is the span of the columns of the matrix. In Subspaces of $\mathbb{R}^n$ Associated with Matrices we introduced the following procedure for finding the column space.

Procedure 3 (**??**)**.** Given a matrix B, a basis for col(B) can be found as follows:

1. Find rref(B) (or any row-echelon form B' of B.)
2. Identify the pivot columns of rref(B) (or B').
3. The columns of B corresponding to the pivot columns of rref(B) (or B') form a basis for col(B).

We will implement this procedure in Octave.

Template 8. Let

$$A = \begin{bmatrix} 2 & -1 & 1 & -4 & 1 \\ 1 & 0 & 3 & 3 & 0 \\ -2 & 1 & -1 & 5 & 2 \\ 4 & -1 & 7 & 2 & 1 \end{bmatrix}$$

In Example **??** we found that columns 1, 2 and 4 of A constitute a basis of col(A). We will now verify this using Octave.

We can find the reduced row-echelon form of A using the *rref* function, and identify the pivot columns visually. We can also take advantage of the expanded version of the *rref* function to identify the pivot columns automatically, as shown below.

```
% Define A
A = [2 -1 1 -4 1;
    1 0 3 3 0;
    -2 1 -1 5 2;
    4 -1 7 2 1];

% Compute rref of A
[rref_A, pivot_cols] = rref(A);

% the pivot column indices are...
pivot_cols

% Extract the pivot columns from A, using pivot column indices
colSpace_basis = A(:, pivot_cols)
```

Link to code.

When you study linear transformations, you will learn that the column space of a matrix is the image of the linear transformation induced by the matrix. So, Template **??** can be used to find the image of a linear transformation.

Null Space of a Matrix (Kernel of a Linear Transformation)

Recall the following definition of the null space of a matrix from Subspaces of $\mathbb{R}^n$ Associated with Matrices.

Definition 104 (**??**)**.** Let A be an $m \times n$ matrix. The *null space* of A, denoted by null(A), is the set of all vectors $\mathbf{x}$ in $\mathbb{R}^n$ such that $A\mathbf{x} = \mathbf{0}$.

We can use the reduced row-echelon form (*rref* function) to find the null space, as shown in Example **??** of Subspaces of $\mathbb{R}^n$ Associated with Matrices. We will omit the details here.

We can also use the *null* function to find a basis for the null space of a matrix. This function finds an *orthonormal* basis for the null space. An orthonormal set of vectors consists of unit vectors that are pairwise orthogonal. We demonstrate the *null* function below.

Template 9. We will use the matrix from Example **??**. Let

$$A = \begin{bmatrix} 2 & -1 & 1 & -4 & 1 \\ 1 & 0 & 3 & 3 & 0 \\ -2 & 1 & -1 & 5 & 2 \\ 4 & -1 & 7 & 2 & 1 \end{bmatrix}$$

Find null(A).

```
% Define A
A = [2 -1 1 -4 1;
    1 0 3 3 0;
    -2 1 -1 5 2;
    4 -1 7 2 1];

% Basis for null(A)
% Octave constructs an orthonormal basis, meaning that the basis elements are pairwise o
B=null(A)

% To find the dimension of the null space, find the number of columns in B
[rows, cols]=size(B)
```

Link to code.

The basis for the null space we found in Template **??** looks different from what you would have found by hand using the *rref* function. In Problem **??** you will verify that B in Template **??** spans the null space of A.

If you are familiar with linear transformations, you may recall that the *kernel* of a linear transformation induced by a matrix is the null space of the matrix. The *null* function gives us a way to compute the kernel.

Octave Exercises

For Problems **??** and **??**, use Octave to find col(A), null(A), and their respective dimensions. Verify that the Rank-Nullity Theorem (**??**) holds for A.

Problem 600 *Let*

$$A = \begin{bmatrix} 2 & -1 & 5 & -1 \\ 1 & 0 & 2 & 1 \\ 1 & -1 & 3 & -2 \\ 1 & -1 & 3 & -2 \end{bmatrix}$$

Problem 601 *Let*

$$A = \begin{bmatrix} -2 & -9 & -33 & -22 \\ 0 & 4 & 15 & 10 \\ 1 & 3 & 12 & 10 \\ -3 & -9 & -36 & -24 \end{bmatrix}$$

Problem 602 *Let*

$$H = span\left\{ \begin{bmatrix} 2 \\ 3 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 8 \\ 15 \\ 6 \\ 3 \end{bmatrix}, \begin{bmatrix} 3 \\ 6 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 4 \\ 6 \\ 6 \\ 3 \end{bmatrix}, \begin{bmatrix} 8 \\ 15 \\ 6 \\ 3 \end{bmatrix} \right\}$$

Find the dimension of H and determine a basis.

Problem 603 *Find the rank of the following matrix. Also find a basis for the row and column spaces.*

$$\begin{bmatrix} 1 & 3 & 0 & -2 & 0 & 3 \\ 3 & 9 & 1 & -7 & 0 & 8 \\ 1 & 3 & 1 & -3 & 1 & -1 \\ 1 & 3 & -1 & -1 & -2 & 10 \end{bmatrix}$$

Problem 604 *Find the rank of the following matrix. Also find a basis for the row and column spaces.*

$$\begin{bmatrix} 1 & 3 & 0 & -2 & 7 & 3 \\ 3 & 9 & 1 & -7 & 23 & 8 \\ 1 & 3 & 1 & -3 & 9 & 2 \\ 1 & 3 & -1 & -1 & 5 & 4 \end{bmatrix}$$

Problem 605 *Find the rank of the following matrix. Also find a basis for the row and column spaces.*

$$\begin{bmatrix} 1 & 0 & 3 & 0 & 7 & 0 \\ 3 & 1 & 10 & 0 & 23 & 0 \\ 1 & 1 & 4 & 1 & 7 & 0 \\ 1 & -1 & 2 & -2 & 9 & 1 \end{bmatrix}$$

Problem 606 *In Template* **??** *we used the* null *function to find a basis B for the null space of matrix*

$$A = \begin{bmatrix} 2 & -1 & 1 & -4 & 1 \\ 1 & 0 & 3 & 3 & 0 \\ -2 & 1 & -1 & 5 & 2 \\ 4 & -1 & 7 & 2 & 1 \end{bmatrix}$$

1. *Verify that the product of A with of each element of the basis B is* **0**.
2. *Use rref*(*A*) *to find another basis for the null space.*
3. *The basis we found using the* null *function and the basis you found in part (b) look very different. How do we know that they span the same subspace?*

Problem 607 *Use some of our earlier matrices, together with their null spaces, for the following experiment.*

- *Given a matrix, find the product of the matrix with some vector in its null space. What do you observe? Why does this make sense?*
- *Formulate a conjecture about the relationship between the rows of a matrix and the elements of its null space.*
- *Prove your conjecture.*

Bibliography

Matrices used in Problems **??**-**??** come from the end of Chapter 4 of Ken Kuttler's *A First Course in Linear Algebra*. (CC-BY)

Ken Kuttler, *A First Course in Linear Algebra*, Lyryx 2017, Open Edition, pp. 227–232.

92 Octave for Chapter 7

Octave for Chapter 7

Examples in this section provide sample Octave code for finding determinants and illustrating their uses and properties. You can access our code through the link at the bottom of each template. Feel free to modify the code and experiment to learn more!

You can write your own code using Octave software or online Octave cells. To access Octave cells online, go to the Sage Math Cell Webpage, select OCTAVE as the language, enter your code, and press EVALUATE.

To "save" or share your online code, click on the *Share* button, select *Permalink*, then copy the address directly from the browser window. You can store this link to access your work later or share this link with others. You will need to get a new Permalink every time you modify the code.

Octave Tutorial

Warning 5. In practice, computing determinants using Laplace Expansion is computationally expensive. *LU*-factorization can be used to improve efficiency, but may be numerically unstable. As a result, determinants are rarely used in computational practice. Examples and exercises in this section are designed to illustrate theoretical properties and computational pitfalls associated with determinants and should not be interpreted as templates for large-scale computational problems.

We will use the *det* function to find determinants in Octave.

Example 224. Compute the determinant of each of the following matrices. If a matrix is singular, say so.

$$A = \begin{bmatrix} -1 & 5 & -1 \\ 7 & -4 & -1 \\ 5 & 6 & -3 \end{bmatrix}, \quad B = \begin{bmatrix} 6 & -1 & 2 \\ 12 & -3 & 1 \\ -1 & 4 & 11 \end{bmatrix}$$

Explanation. We find the determinant of each matrix.

```
% Define matrices A and B
A=[-1  5  -1;
7  -4  -1;
```

```
5  6 -3];

B=[6  -1  2;
12  -3  1;
-1  4  11];

% Find det(A) and det(B)
det_A=det(A)
det_B=det(B)
```

Link to code.

We see that $\det(A) = 0$ and $\det(B) \neq 0$. We conclude that A is singular and B is not.

The following example illustrates a common technique for computing determinants numerically.

Example 225. In Exploration **??**, we found that

$$A = \begin{bmatrix} 2 & -1 & 2 \\ 6 & 1 & 5 \\ -4 & 10 & -3 \end{bmatrix} = LU = \begin{bmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ -2 & 2 & 1 \end{bmatrix} \begin{bmatrix} 2 & -1 & 2 \\ 0 & 4 & -1 \\ 0 & 0 & 3 \end{bmatrix}$$

Having a lower triangular/upper triangular factorization makes it easy to compute the determinant. Find the determinant of A by sight.

Explanation.

$$\det(A) = \det(LU) = \det(L)\det(U) = (1)(24) = 24$$

Unfortunately, LU-factorization is prone to round-off error, which may lead to erroneous or ambiguous outcomes, as you will see Problem **??**.

Recall the following properties of determinants.

Theorem 131 (**??**)**.** Let $A = \left[a_{ij}\right]$ be an $n \times n$ matrix.

1. If B is obtained from A by interchanging two different rows, then

$$\det B = -\det A$$

2. If B is obtained from A by multiplying one of the rows of A by a non-zero constant k. Then

$$\det B = k \det A$$

3. If B is obtained from A by adding a multiple of one row of A to another row, then

$$\det B = \det A$$

Example 226. Let

$$A = \begin{bmatrix} 2 & -1 & 5 \\ 0 & 3 & -4 \\ -2 & 1 & -1 \end{bmatrix}$$

Illustrate part (b) of the above theorem using the second row of matrix A, and constant $k = 4$.

Explanation. We will use the following code to illustrate the property.

```
% Define rows of matrix A
r_1=[2  -1  5];
r_2=[0  3  -4];
r_3=[-2  1  -1];

% Combine the rows to form matrix A
A=[r_1; r_2; r_3];

% Compute the determinant of A
det_A=det(A)

% Form matrix B by multiplying the second row of A by k=4
B=[r_1; 4*r_2; r_3];

% Compute the determinant of B
det_B=det(B)

% Compute 4*det(A)
det_A_times4=4*det(A)
```

Link to code.

For our next example, recall the following theorem, known as Cramer's Rule.

Theorem 132 (**??**)**.** Let A be a nonsingular $n \times n$ matrix, and let $\mathbf{b}$ be an $n \times 1$ vector. Then the components of the solution vector $\mathbf{x}$ of $A\mathbf{x} = \mathbf{b}$ are given by

$$x_i = \frac{\det A_i(\mathbf{b})}{\det A}$$

where $A_i(\mathbf{b})$ is matrix A with its i^{th} column replaced by $\mathbf{b}$.

Example 227. Use Octave to solve the following system for x_1 using Cramer's rule.

$$\begin{bmatrix} 1 & 2 & -1 \\ -1 & 1 & 1 \\ 0 & 3 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -2 \\ 3 \\ 1 \end{bmatrix}$$

Explanation. We will use the following code.

```
% Define columns of matrix A
c_1=[1; -1; 0];
c_2=[2; 1;  3];
c_3=[-1; 1; 1];

% Combine the columns to form matrix A
A=[c_1 c_2 c_3];

% Define vector b
b=[-2; 3; 1];

% Form A_1 by replacing the first column of A with b
A_1=[b c_2 c_3];

% Find x_1
x_1=det(A_1)/det(A)
```

Link to code.

Octave Exercises

Problem 608 *Let*

$$A = \begin{bmatrix} 1 & -2 & 1 & 6 & 1 \\ 0 & -1 & 3 & 5 & -1 \\ 2 & -1 & 8 & 4 & 2 \\ 0 & 14 & -11 & -7 & 2 \\ 3 & -3 & 9 & 10 & 3 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & -2 & 1 & 6 & 1 \\ 0 & -1 & 3 & 5 & 0 \\ 2 & -1 & 8 & 4 & 1 \\ 3 & 12 & -2 & 3 & 0 \\ 2 & 0 & -11 & -9 & 5 \end{bmatrix}$$

Determine whether each of the above matrices is singular in two different ways.

- *Use the* det *function*
- *Use an alternative approach*

Are the two methods used above in agreement? Click the arrow on the right to learn more.

Octave uses LU-decomposition to compute the determinant. This method is susceptible to rounding errors (Reference). Because of this, using the det *function to determine singularity is generally not a good idea.*

If A is non-singular, find A^{-1}*, and* $\det(A^{-1})$*. Illustrate the relationship between* $\det(A)$ *and* $\det(A^{-1})$ *computationally.*

Problem 609 *Modify Example* **??** *to illustrate parts (a) and (c) of Theorem* **??***. (Use rows and coefficients of your choice.)*

Problem 610 *Modify the code of Example* **??** *to find* x_2 *and* x_3*.*

93 Octave for Chapter 8

Octave for Chapter 8

Examples in this section provide sample Octave code for computing eigenvalues. You can access our code through the link at the bottom of each template. Feel free to modify the code and experiment to learn more!

You can write your own code using Octave software or online Octave cells. To access Octave cells online, go to the Sage Math Cell Webpage, select OCTAVE as the language, enter your code, and press EVALUATE.

To ”save” or share your online code, click on the *Share* button, select *Permalink*, then copy the address directly from the browser window. You can store this link to access your work later or share this link with others. You will need to get a new Permalink every time you modify the code.

Octave Tutorial

Template 10. Find eigenvalues and the corresponding eigenvectors of the given matrix *A*.

```
% Define A
A=[5 0 0;
1 2 -1;
1 3 -2];

% Find eigenvalues of A
eigen_A=eig(A)

% Find eigenvalues and eigenvectors of A
[V,D]=eig(A)
% D is a diagonal matrix whose diagonal entries are the eigenvalues of A
% Columns of V are the eigenvectors of A
```

Link to code.

Template 11. We can find the characteristic polynomial of the given matrix *A* as follows.

```
% Define A
A=[5 0 0;
    1 2 -1;
    1 3 -2];

% Find the coefficients of the char poly of A
char_poly = poly(A)
```

The output looks like this:

```
char_poly =

 1 -5 -1 5
```

Link to code.

The output is a vector of coefficients, starting with the leading coefficient. In this case, we see that the characteristic polynomial is $\lambda^3 - 5\lambda^2 - \lambda + 5$.

Warning 6. In this text, we find the characteristic polynomial by computing $\det(A - \lambda I)$. Octave, as well as some other textbooks, finds the characteristic polynomial based on $\det(\lambda I - A)$. Note that setting these expressions equal to zero to find eigenvalues produces the same result. However, the coefficients of the characteristic polynomials from the two approaches are negatives of each other. If you were to do this problem by hand, using the method in this text, you would get $-\lambda^3 + 5\lambda^2 + \lambda - 5$.

Example 228. In Example **??**, you found that the rotation matrix $M = \begin{bmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{bmatrix}$ has complex eigenvalues. Use the *eig* command in Octave to compute the eigenvalues and eigenvectors of M.

Explanation. Here is the input code:

```
% Define A
A=[sqrt(2)/2  -sqrt(2)/2;
sqrt(2)/2  sqrt(2)/2];
```

```
% Find eigenvalues of A
eigen=eig(A)

% Find eigenvalues and eigenvectors of A
[V,D]=eig(A)
% D is a diagonal matrix whose diagonal entries are the eigenvalues of A
% Columns of V are the eigenvectors of A
```

This is what we get for the output:

```
eigen =

 (0.707107,0.707107)
 (0.707107,-0.707107)

V =

 (0,0.707107) (0,-0.707107)
 (0.707107,0) (0.707107,-0)

D =

 (0.707107,0.707107) (0,0)
 (0,0) (0.707107,-0.707107)
```

Link to code.

Each complex number in the output is presented using its real and imaginary parts as x and y coordinates of a point. For example, the second eigenvalue, written as $(0.707107, -0.707107)$, is the complex number $\frac{\sqrt{2}}{2} - \frac{\sqrt{2}}{2}i$. Compare this to the eigenvalue in Example **??**. The corresponding eigenvector is the second column of the matrix V: $\begin{bmatrix}(0, -0.707107)\\(0.707107, -0)\end{bmatrix}$. This should be interpreted as $\begin{bmatrix}(-\sqrt{2}/2)i\\ \sqrt{2}/2\end{bmatrix}$. Compare this to the eigenvector in Example **??**. There is a discrepancy! Did we make a mistake? How do we reconcile the two eigenvectors?

Octave Exercises

Problem 611 *Let* $A = \begin{bmatrix} -148 & 16 & -528 \\ 71 & -28 & 292 \\ 39 & -4 & 140 \end{bmatrix}$.

1. *Use Octave to find the eigenvalues of A.*
2. *Find the eigenvalues of* $4A$*. How do they compare?*
3. *Formulate and prove a conjecture about the relationship between the eigenvalues of A and the eigenvalues of kA, where k is a constant.*
4. *How do the eigenvectors of A compare to the eigenvectors of kA? Prove your claim.*

Problem 612 *Let*

$$A = \begin{bmatrix} 2 & -20 & 36 \\ -1 & 8 & -14 \\ 0 & 5 & -7 \end{bmatrix}$$

1. *Find the eigenvalues of A using Octave. List the eigenvalues below in increasing order.*

$$\lambda_1 = \ldots\ldots, \quad \lambda_2 = \ldots\ldots, \quad \lambda_3 = \ldots\ldots$$

2. *By hand, find the corresponding eigenvectors of A. List them below.*

$$\mathbf{x}_{\lambda_1} = \begin{bmatrix} \ldots\ldots \\ \ldots\ldots \\ 1 \end{bmatrix}, \quad \mathbf{x}_{\lambda_2} = \begin{bmatrix} \ldots\ldots \\ \ldots\ldots \\ 5 \end{bmatrix}, \quad \mathbf{x}_{\lambda_3} = \begin{bmatrix} \ldots\ldots \\ \ldots\ldots \\ 1 \end{bmatrix}$$

3. *Use Octave to find the eigenvectors of A. Reconcile your answers from the previous part with the answers you got from Octave.*

Problem 613 *Use Octave to find the eigenvalues of* $A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$*. Interpret the numbers you are seeing. Interpret your results geometrically.*

Problem 614 *Let*

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$$

When we use Octave to find the eigenvalues of A, we get the following printout.

```
ans =

16.1168
-1.11684
-1.30368e-15
```

Should we interpret the last eigenvalue as 0*, or as a very small non-zero number? How can you be sure of your answer in this particular case?*

Problem 615 *Modify the code for the Power Method given in The Power Method and the Dominant Eigenvalue to approximate the dominant eigenvalue of* $A = \begin{bmatrix} -4 & 64 \\ 16 & 44 \end{bmatrix}$*. Experiment with the number of iterations. Compare your answer to the answer you get when using the* eig *function.*

94 Octave for Chapter 9

Octave for Chapter 9

Examples and Templates in this section provide sample Octave code for least squares, and *QR*-factorization. You can access our code through the link at the bottom of each template. Feel free to modify the code and experiment to learn more!

You can write your own code using Octave software or online Octave cells. To access Octave cells online, go to the Sage Math Cell Webpage, select OCTAVE as the language, enter your code, and press EVALUATE.

To "save" or share your online code, click on the *Share* button, select *Permalink*, then copy the address directly from the browser window. You can store this link to access your work later or share this link with others. You will need to get a new Permalink every time you modify the code.

Octave Tutorial

Least Squares

The following theorem establishes one way to implement least squares.

Theorem 133 (**??**)**.** Let A be an $m \times n$ matrix, let $\mathbf{b}$ be a column vector in $\mathbb{R}^m$. Consider the matrix equation

$$A\mathbf{x} = \mathbf{b}$$

1. Any solution $\mathbf{z}$ to the normal equations

   $$\left(A^T A\right)\mathbf{z} = A^T\mathbf{b}$$

   is a best approximation to a solution to $A\mathbf{x} = \mathbf{b}$ in the sense that $\|\mathbf{b} - A\mathbf{z}\|$ is minimized.

2. If the columns of A are linearly independent, then $A^T A$ is invertible and $\mathbf{z}$ is given uniquely by

   $$\mathbf{z} = \left(A^T A\right)^{-1} A^T\mathbf{b}$$

We can implement this process as follows.

Template 12. We show how to find a least squares solution to $A\mathbf{x} = \mathbf{b}$.

```
% Define matrix A
A=[9 -3 3;
    1 -1 1;
    4 0 1;
    1 1 1;
    9 -1 2];

% Define vector b
b=[3;1;1;2;4];

% Now we solve for z
z=inv(transpose(A)*A)*transpose(A)*b
```

Link to code.

Warning 7. Finding matrix inverses is both computationally expensive and unstable.

To do least squares in Octave, we use the backslash operator (\). Depending on the type of matrix, backslash uses different techniques to compute the least squares solution more efficiently. For more information about *inv* and the backslash operator (\) see Reference.

The following two examples illustrate the two implementations of least squares side-by-side.

Example 229. We will compare the results of the two implementations of least squares for a 7×3 coefficient matrix.

```
% Define matrix A
A=[4 -1 0;
    2 -1 1;
    6 3 1;
    -1 1 2;
    3 1 -2;
    -2 8 7;
    1 1 -5];

% Define vector b
b=[-2;4;1;2;-5;1;-3];
```

```
% Implement least squares using our original method
z1=inv(transpose(A)*A)*transpose(A)*b

% Implement least squares using the backslash operator
z2=A\b
```

Link to code.

Example 230. We will compare the performance of two the implementations of least squares for a 2000×20 coefficient matrix.

```
% Define matrix A
A=rand(2000,20); % rand(m,n) creates an m by n matrix whose entries are between 0

% Define vector b
b=rand(2000,1);

% Start the timer
tic

% Implement least squares using our original method
z1=inv(transpose(A)*A)*transpose(A)*b;

% Stop the timer
timeINV=toc

% Start the timer
tic

% Implement least squares using the backslash operator
z2=A\b;

% Stop the timer
timeBSlash=toc
```

Link to code.

Each time you run the algorithm, you will get slightly different run times for each implementation. Most of the time, you will see that the backslash operator outperforms our initial method.

The backslash operator (\) can be used to solve systems of equations when a solution exists. Here is an example.

Example 231. Consider the system

$$\begin{array}{rcrcrcrcl} x & - & y & & & & & = & 0 \\ 2x & - & 2y & + & z & + & 2w & = & 4 \\ & & y & & & + & w & = & 0 \\ & & & & 2z & + & w & = & 5 \end{array} \tag{120}$$

(You may remember this system from Augmented Matrix Notation and Elementary Row Operations.) We will solve this system using the backslash operator (\).

```
% Define the coefficient matrix A
A = [1 -1 0 0;
     2 -2 1 2;
     0 1 0 1;
     0 0 2 1];

% Define vector b
b = [0;4;0;5];

% Solve the system
x = A \ b

% Compare the solution above to the solution obtained
% using rref
A_b=[A b];
rref(A_b)
```

Link to code.

Warning 8. While the backslash (\) operator can be used to solve feasible systems, we urge you to be very cautious because

- If a system has infinitely many solutions, the backslash (\) operator will find one particular solution.
- If a system has no solutions, the backslash (\) operator will find an approximate solution using least-squares.

The way the answer is presented, you cannot tell the difference!

QR-factorization

Recall the definition of *QR*-factorization.

Definition 105 (**??**)**.** Let A be an $m \times n$ matrix with independent columns. A *QR-factorization* of A expresses it as $A = QR$ where Q is $m \times n$ with orthonormal columns and R is an invertible and upper triangular matrix with positive diagonal entries.

Example 232. Find the *QR*-factorization of

$$A = \begin{bmatrix} 2 & -1 & 4 & 0 \\ 3 & 1 & 1 & -2 \\ 1 & 5 & -1 & 6 \\ -3 & 4 & 0 & 7 \end{bmatrix}$$

and demonstrate that matrix Q is orthogonal (columns are orthonormal).

Explanation.

```
% Define A
A=[2 -1 4 0;
    3 1 1 -2;
    1 5 -1 6;
    -3 4 0 7];

% Find the QR-factorization of A
[Q,R]=qr(A)

% Verify that columns of Q are orthonormal
% You may get numbers that are close to zero but not zero
% due to round-off error
transpose(Q)*Q

% why does this verify that the columns are orthonormal?
```

Link to code.

Question 616 *In the code above, we used the product Q^TQ to verify that the columns of Q are orthonormal. What do you expect this product to look like if the columns are orthonormal? What will the product look like if they are not? You will be asked to explain your reasoning in Problem* **??***.*

Recall the following algorithm for using *QR*-factorization to approximate eigenvalues from *QR*-Factorization.

Algorithm 4 (**??**)**.** Let A be an invertible matrix.

Step 1: Define $A_1 = A$ and factor it as $A_1 = Q_1R_1$.

Step 2: Define $A_2 = R_1Q_1$ and factor it as $A_2 = Q_2R_2$.

Step 3: Define $A_3 = R_2Q_2$ and factor it as $A_3 = Q_3R_3$.

$\vdots$

Step $k+1$: Define $A_{k+1} = R_kQ_k$ (where $A_k = Q_kR_k$).

Note that A_{k+1} is similar to A_k (in fact, $A_{k+1} = R_kQ_k = (Q_k^{-1}A_k)Q_k$), and hence each A_k has the same eigenvalues as A. If the eigenvalues of A are real and have distinct absolute values, the sequence of matrices $A_1, A_2, A_3, ...$ converges to an upper triangular matrix with these eigenvalues on the main diagonal.

Compare the following example to Example **??**.

Example 233. Use the *QR* algorithm to approximate the eigenvalues of $A = \begin{bmatrix} 1 & 1 \\ 2 & 0 \end{bmatrix}$.

Explanation. Here is a very simple implementation of the algorithm.

```
% Define A
A=[1 1;
    2 0];

% Set the desired number of iterations
iter=20;

% Implement the algorithm. Note that we are over-writing the previous A with every itera
for i=1:iter
    [Q,R]=qr(A);
    A=R*Q
end
```

Link to code.

Observe that the main diagonal entries are approaching the eigenvalues of A.

Octave Exercises

Problem 617 *Attempt to solve each of the following systems of equations using the backslash (\) operator, and the* rref *function. Interpret your results.*

1.

$$\begin{array}{rcrcrcl} 3x & - & y & + & 4z & = & 2 \\ & & y & + & z & = & 1 \\ -2x & & & + & 3z & = & 1 \end{array} \tag{121}$$

What is true about this system?

Multiple Choice:

(i) *The system is consistent. Using the backslash operator produces the same result as using the* rref *function.*

(ii) *The system is inconsistent. The backslash operator gives the least squares approximation.*

(iii) *The system is consistent but has infinitely many solutions. The backslash operator gives one particular solution.*

2.

$$\begin{array}{rcrcrcr} 3x & - & y & + & 4z & = & 2 \\ x & - & y & + & 2z & = & 4 \\ -2x & + & y & - & 3z & = & -3 \end{array} \tag{122}$$

What is true about this system?

Multiple Choice:

(i) *The system is consistent. Using the backslash operator produces the same result as using the* rref *function.*

(ii) *The system is inconsistent. The backslash operator gives the least squares approximation.*

(iii) *The system is consistent but has infinitely many solutions. The backslash operator gives one particular solution.*

3.

$$\begin{array}{rcrcrcl} 3x & - & y & + & 4z & = & 2 \\ x & - & y & + & 2z & = & 1 \\ -2x & + & y & - & 3z & = & 1 \end{array} \tag{123}$$

What is true about this system?

Multiple Choice:

(i) *The system is consistent. Using the backslash operator produces the same result as using the* rref *function.*

(ii) *The system is inconsistent. The backslash operator gives the least squares approximation.*

(iii) *The system is consistent but has infinitely many solutions. The backslash operator gives one particular solution.*

Problem 618 *Use least squares to find a line of best fit for the points shown in the GeoGebra interactive below. Enter the equation of the line into the GeoGebra interactive to view your result.*

GeoGebra link: https://www.geogebra.org/m/kmt5ukhc

1. *Find the residual, $\|\mathbf{b} - A\mathbf{z}\|$, for your line of best fit.*
2. *The original line, $y = x + 1$ initially shown in the interactive, visually appeared to be a pretty good fit. Find the residual for the line $y = x + 1$ and compare it to the residual you got for your line. Is your line a better fit?*

Problem 619 *Explain how to interpret the product Q^TQ to determine whether matrix Q has orthonormal columns. Prove your claims.*

Problem 620 *Modify the code in Example* **??** *to approximate the eigenvalues of* $A = \begin{bmatrix} 9 & 2 & 8 \\ 2 & -6 & -2 \\ -8 & 2 & -5 \end{bmatrix}$. *Compare your answers to the answers you got for Problem* **??**.

Problem 621 *Modify the code in Example* **??** *for the matrix* $A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$. *What are you observing? Are we getting close to finding the eigenvalues of A? Explain what is happening and why.*

95 Appendix

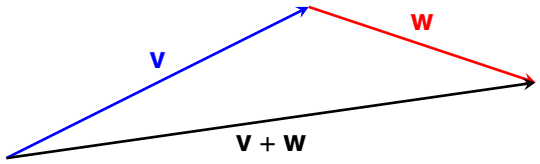
96 Triangle Inequality

The *Triangle Inequality* is a simple, yet powerful result used widely in analysis and topology as well as other branches of mathematics. The triangle inequality has its roots in geometry. It initially appeared as a proposition in the *Elements* - a treatise comprised of thirteen books covering plane and solid geometry, and number theory - written by Euclid of Alexandria around 300 B.C.

The geometric version of the triangle inequality states that the sum of the lengths of any two sides of a triangle is greater than the length the third side.

Here we are interested in the vector version of this result. Given vectors **v** and **w**, we have

$$\|\mathbf{v} + \mathbf{w}\| \leq \|\mathbf{v}\| + \|\mathbf{w}\|$$



Intuitively, we observe that equality occurs when either **v** or **w** (or both) are zero, or when non-zero vectors **v** and **w** point in the same direction, otherwise the inequality is strict.

Proving the triangle inequality requires some preliminary results.

Lemma 13. For any vectors **v** and **w** in $\mathbb{R}^n$,

$$\|\mathbf{v} + \mathbf{w}\|^2 = \|\mathbf{v}\|^2 + 2(\mathbf{v} \cdot \mathbf{w}) + \|\mathbf{w}\|^2 \tag{124}$$

Proof By Theorem **??** we have,

$$\begin{aligned} \|\mathbf{v} + \mathbf{w}\|^2 &= (\mathbf{v} + \mathbf{w}) \cdot (\mathbf{v} + \mathbf{w}) \\ &= \mathbf{v} \cdot \mathbf{v} + \mathbf{v} \cdot \mathbf{w} + \mathbf{w} \cdot \mathbf{v} + \mathbf{w} \cdot \mathbf{w} \\ &= \|\mathbf{v}\|^2 + 2(\mathbf{v} \cdot \mathbf{w}) + \|\mathbf{w}\|^2 \end{aligned}$$

■

Theorem 134 (Cauchy-Schwarz Inequality)**.** For any vectors **v** and **w** in $\mathbb{R}^n$,

$$|\mathbf{v} \cdot \mathbf{w}| \leq \|\mathbf{v}\| \, \|\mathbf{w}\| \tag{125}$$

Proof Recall that by Theorem **??**, $\mathbf{v} \cdot \mathbf{w} = \|\mathbf{v}\| \, \|\mathbf{w}\| \cos\theta$, where θ is the included angle. Our result follows from the fact that $|\cos\theta| \leq 1$. ■

Theorem 135 (Triangle Inequality)**.** For any vectors **v** and **w** in $\mathbb{R}^n$,

$$\|\mathbf{v} + \mathbf{w}\| \leq \|\mathbf{v}\| + \|\mathbf{w}\| \tag{126}$$

Proof We will use Lemma **??** and Theorem **??**.

$$\begin{aligned} \|\mathbf{v} + \mathbf{w}\|^2 &= \|\mathbf{v}\|^2 + 2(\mathbf{v} \cdot \mathbf{w}) + \|\mathbf{w}\|^2 \\ &\leq \|\mathbf{v}\|^2 + 2\,\|\mathbf{v}\| \, \|\mathbf{w}\| + \|\mathbf{w}\|^2 \\ &= \left(\|\mathbf{v}\| + \|\mathbf{w}\|\right)^2 \end{aligned}$$

Taking the square root of both sides yields the desired result.

■

97 Complex Numbers

The complex numbers, denoted $\mathbb{C}$, have the wonderful property that *every* polynomial with complex coefficients has a (complex) root. This fact is known as the *Fundamental Theorem of Algebra*.

One pleasant aspect of the complex numbers is that, whereas describing the real numbers in terms of the rationals is a rather complicated business, the complex numbers are quite easy to describe in terms of real numbers. Every *complex number* has the form

$$a + bi$$

where a and b are real numbers, and i is a root of the polynomial $x^2 + 1$. Here a and b are called the *real part* and the *imaginary part* of the complex number, respectively. The real numbers are now regarded as special complex numbers of the form $a + 0i = a$, with zero imaginary part. The complex numbers of the form $0 + bi = bi$ with zero real part are called *pure imaginary* numbers. The complex number i itself is called the *imaginary unit* and is distinguished by the fact that

$$i^2 = -1$$

As the terms *complex* and *imaginary* suggest, these numbers met with some resistance when they were first used. This has changed; now they are essential in science and engineering as well as mathematics, and they are used extensively. The names persist, however, and continue to be a bit misleading: These numbers are no more "*complex*" than the real numbers, and the number i is no more "*imaginary*" than -1.

Much as for polynomials, two complex numbers are declared to be *equal* if and only if they have the same real parts and the same imaginary parts. In symbols,

$$a + bi = a' + b'i \quad \text{if and only if } a = a' \text{ and } b = b'$$

The addition and subtraction of complex numbers is accomplished by adding and subtracting real and imaginary parts:

$$(a + bi) + (a' + b'i) = (a + a') + (b + b')i$$
$$(a + bi) - (a' + b'i) = (a - a') + (b - b')i$$

This is analogous to these operations for linear polynomials $a + bx$ and $a' + b'x$, and the multiplication of complex numbers is also analogous with one difference: $i^2 = -1$. The definition is

$$(a + bi)(a' + b'i) = (aa' - bb') + (ab' + ba')i$$

With these definitions of equality, addition, and multiplication, the complex numbers *satisfy all the basic arithmetical axioms adhered to by the real numbers* (the verifications are omitted). One consequence of this is that they can be manipulated in the obvious fashion, except that i^2 is replaced by -1 wherever it occurs, and the rule for equality must be observed.

Example 234. If $z = 2 - 3i$ and $w = -1 + i$, write each of the following in the form $a + bi$: $z + w$, $z - w$, zw, $\frac{1}{3}z$, and z^2.

Explanation.

$$\begin{aligned} z + w &= (2 - 3i) + (-1 + i) = (2 - 1) + (-3 + 1)i = 1 - 2i \\ z - w &= (2 - 3i) - (-1 + i) = (2 + 1) + (-3 - 1)i = 3 - 4i \\ zw &= (2 - 3i)(-1 + i) = (-2 - 3i^2) + (2 + 3)i = 1 + 5i \\ \frac{1}{3}z &= \frac{1}{3}(2 - 3i) = \frac{2}{3} - i \\ z^2 &= (2 - 3i)(2 - 3i) = (4 + 9i^2) + (-6 - 6)i = -5 - 12i \end{aligned}$$

Example 235. Find all complex numbers z such as that $z^2 = i$.

Explanation. Write $z = a + bi$; we must determine a and b. Now $z^2 = (a^2 - b^2) + (2ab)i$, so the condition $z^2 = i$ becomes

$$(a^2 - b^2) + (2ab)i = 0 + i$$

Equating real and imaginary parts, we find that $a^2 = b^2$ and $2ab = 1$. The solution is $a = b = \pm\frac{1}{\sqrt{2}}$, so the complex numbers required are $z = \frac{1}{\sqrt{2}} + \frac{1}{\sqrt{2}}i$ and $z = -\frac{1}{\sqrt{2}} - \frac{1}{\sqrt{2}}i$.

As for real numbers, it is possible to divide by every nonzero complex number z. That is, there exists a complex number w such that $wz = 1$. As in the real case, this number w is called the *inverse* of z and is denoted by z^{-1} or $\frac{1}{z}$. Moreover, if $z = a + bi$, the fact that $z \neq 0$ means that $a \neq 0$ or $b \neq 0$. Hence $a^2 + b^2 \neq 0$, and an explicit formula for the inverse is

$$\frac{1}{z} = \frac{a}{a^2 + b^2} - \frac{b}{a^2 + b^2}i$$

In actual calculations, the work is facilitated by two useful notions: the conjugate and the absolute value of a complex number. The next example illustrates the technique.

Example 236. Write $\frac{3 + 2i}{2 + 5i}$ in the form $a + bi$.

Explanation. Multiply top and bottom by the complex number $2 - 5i$ (obtained from the denominator by negating the imaginary part). The result is

$$\frac{3 + 2i}{2 + 5i} = \frac{(2 - 5i)(3 + 2i)}{(2 - 5i)(2 + 5i)} = \frac{(6 + 10) + (4 - 15)i}{2^2 - (5i)^2} = \frac{16}{29} - \frac{11}{29}i$$

Hence the simplified form is $\frac{16}{29} - \frac{11}{29}i$, as required.

The key to this technique is that the product $(2 - 5i)(2 + 5i) = 29$ in the denominator turned out to be a *real* number. The situation in general leads to the following notation: If $z = a + bi$ is a complex number, the *conjugate* of z is the complex number, denoted $\overline{z}$, given by

$$\overline{z} = a - bi \quad \text{where } z = a + bi$$

Hence $\overline{z}$ is obtained from z by negating the imaginary part. Thus $\overline{(2 + 3i)} = 2 - 3i$ and $\overline{(1 - i)} = 1 + i$. If we multiply $z = a + bi$ by $\overline{z}$, we obtain

$$z\overline{z} = a^2 + b^2 \quad \text{where } z = a + bi$$

The real number $a^2 + b^2$ is always nonnegative, so we can state the following definition: The *absolute value* or *modulus* of a complex number $z = a + bi$, denoted by $|z|$, is the positive square root $\sqrt{a^2 + b^2}$; that is,

$$|z| = \sqrt{a^2 + b^2} \quad \text{where } z = a + bi$$

For example, $|2 - 3i| = \sqrt{2^2 + (-3)^2} = \sqrt{13}$ and $|1 + i| = \sqrt{1^2 + 1^2} = \sqrt{2}$.

Note that if a real number a is viewed as the complex number $a + 0i$, its absolute value (as a complex number) is $|a| = \sqrt{a^2}$, which agrees with its absolute value as a *real* number.

With these notions in hand, we can describe the technique applied in Example **??** as follows: When converting a quotient $\frac{z}{w}$ of complex numbers to the form $a + bi$, multiply top and bottom by the conjugate $\overline{w}$ of the denominator.

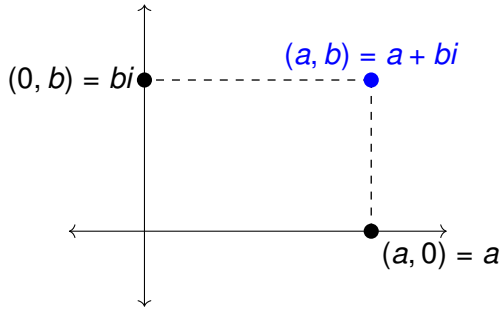
The following list contains the most important properties of conjugates and absolute values. Throughout, z and w denote complex numbers.

*C*1.	$\overline{z \pm w} = \overline{z} \pm \overline{w}$	*C*7.	$\frac{1}{z} = \frac{1}{\lvert z\rvert^2}\overline{z}$
*C*2.	$\overline{zw} = \overline{z}\,\overline{w}$	*C*8.	$\lvert z\rvert \geq 0$ for all complex numbers z
*C*3.	$\overline{\left(\frac{z}{w}\right)} = \frac{\overline{z}}{\overline{w}}$	*C*9.	$\lvert z\rvert = 0$ if and only if $z = 0$
*C*4.	$\overline{(\overline{z})} = z$	*C*10.	$\lvert zw\rvert = \lvert z\rvert\lvert w\rvert$
*C*5.	z is real if and only if $\overline{z} = z$	*C*11.	$\lvert\frac{z}{w}\rvert = \frac{\lvert z\rvert}{\lvert w\rvert}$
*C*6.	$z\overline{z} = \lvert z\rvert^2$	*C*12.	$\lvert z + w\rvert \leq \lvert z\rvert + \lvert w\rvert$(the Triangle Inequality)

All these properties (except property C12) can (and should) be verified by the reader for arbitrary complex numbers $z = a + bi$ and $w = c + di$. They are not independent; for example, property C10 follows from properties C2 and C6.

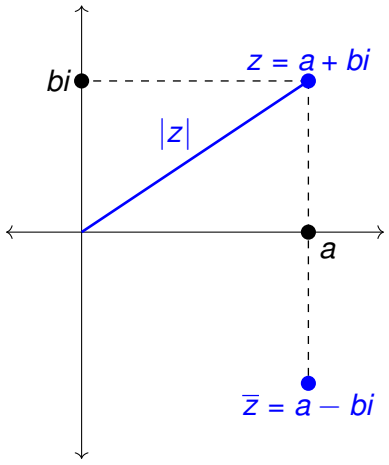
The triangle inequality, as its name suggests, comes from a geometric representation of the complex numbers analogous to identification of the real numbers with the points of a line. The representation is achieved as follows:

Introduce a rectangular coordinate system in the plane, and identify the complex number $a + bi$ with the point (a, b), as shown in the figure below.



When this is done, the plane is called the *complex plane*. Note that the point $(a, 0)$ on the x axis now represents the *real* number $a = a + 0i$, and for this reason, the x axis is called the *real axis*. Similarly, the y axis is called the *imaginary axis*. The identification $(a, b) = a + bi$ of the geometric point (a, b) and the complex number $a + bi$ will be used in what follows without comment. For example, the origin will be referred to as 0.

This representation of the complex numbers in the complex plane gives a useful way of describing the absolute value and conjugate of a complex number $z = a + bi$. The absolute value $|z| = \sqrt{a^2 + b^2}$ is just the distance from z to the origin. This makes properties C8 and C9 quite obvious. The conjugate $\overline{z} = a - bi$ of z is just the reflection of z in the real axis (x axis), a fact that makes properties C4 and C5 clear.



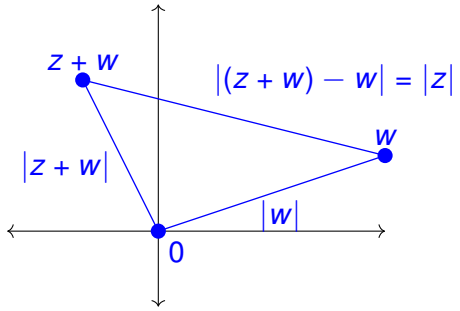
Given two complex numbers $z_1 = a_1 + b_1 i = (a_1, b_1)$ and $z_2 = a_2 + b_2 i = (a_2, b_2)$, the absolute value of their difference

$$|z_1 - z_2| = \sqrt{(a_1 - a_2)^2 + (b_1 - b_2)^2} \tag{127}$$

is just the distance between them. This gives the *complex distance formula*:

$$|z_1 - z_2| \text{ is the distance between } z_1 \text{ and } z_2$$

This useful fact yields a simple verification of the triangle inequality, property C12. Suppose z and w are given complex numbers. Consider the triangle in the figure below whose vertices are 0, w, and $z + w$.



The three sides have lengths $|z|$, $|w|$, and $|z + w|$ by the complex distance formula, so the inequality

$$|z + w| \leq |z| + |w|$$

expresses the obvious geometric fact that the sum of the lengths of two sides of a triangle is at least as great as the length of the third side.

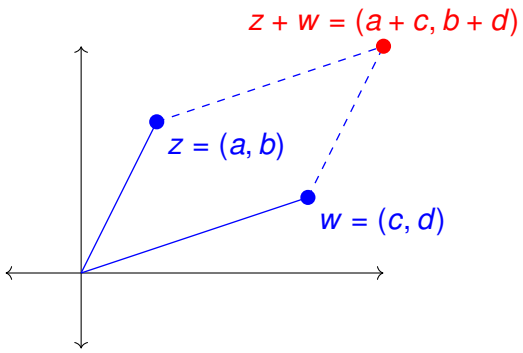
The representation of complex numbers as points in the complex plane has another very useful property: It enables us to give a geometric description of the sum and product of two complex numbers. To obtain the description for the sum, let

$$z = a + bi = (a, b)$$
$$w = c + di = (c, d)$$

denote two complex numbers. We claim that the four points 0, z, w, and $z + w$ form the vertices of a parallelogram. In fact, in the figure below, the lines from 0 to z and from w to $z + w$ have slopes

$$\frac{b - 0}{a - 0} = \frac{b}{a} \quad \text{and} \quad \frac{(b + d) - d}{(a + c) - c} = \frac{b}{a}$$

respectively, so these lines are parallel. (If it happens that $a = 0$, then both these lines are vertical.)



Similarly, the lines from z to $z + w$ and from 0 to w are also parallel, so the figure with vertices 0, z, w, and $z + w$ is indeed a parallelogram. Hence, the complex number $z + w$ can be obtained geometrically from z and w by *completing* the parallelogram. This is sometimes called the *parallelogram law* of complex addition. Readers who have studied mechanics will recall that velocities and accelerations add in the same way; in fact, these are all special cases of *vector* addition.

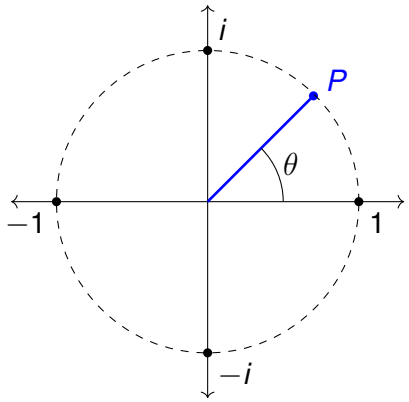
Exploration 622 *In this exploration we return to the complex numbers of Example* **??***, and we analyze them from a geometric point of view. Let* $z = 2 - 3i$ *and* $w = -1 + i$*.*

1. *Plot* z*,* w*, and* $z + w$ *in the on the same plane to see an example of the parallelogram law of complex addition.*

2. *What happens geometrically when you multiply a complex number by* i*? What if you multiply by* i *a second time? Verify that this is the same as multiplying by* i^2*.*

Polar Form

The geometric description of what happens when two complex numbers are multiplied is at least as elegant as the parallelogram law of addition, but it requires that the complex numbers be represented in polar form. Before discussing this, we pause to recall the general definition of the trigonometric functions sine and cosine. An angle θ in the complex plane is in *standard position* if it is measured counterclockwise from the positive real axis as indicated in the figure below.



Rather than using degrees to measure angles, it is more natural to use radian measure. This is defined as follows: The circle with its centre at the origin and radius 1 (called the *unit circle*) is drawn as in the above figure. It has circumference 2π, and the *radian measure* of θ is the length of the arc on the unit circle counterclockwise from 1 to the point P on the unit circle determined by θ. Hence $90^\circ = \frac{\pi}{2}$, $45^\circ = \frac{\pi}{4}$, $180^\circ = \pi$, and a full circle has the angle $360^\circ = 2\pi$. Angles measured clockwise from 1 are negative; for example, $-i$ corresponds to $-\frac{\pi}{2}$ (or to $\frac{3\pi}{2}$).

Consider an angle θ in the range $0 \leq \theta \leq \frac{\pi}{2}$. If θ is plotted in standard position as in the above figure, it determines a unique point P on the unit circle, and P has coordinates $(\cos\theta, \sin\theta)$ by elementary trigonometry. However, *any* angle θ (acute or not) determines a unique point on the unit circle, so we *define* the *cosine* and *sine* of θ (written $\cos\theta$ and $\sin\theta$) to be the x and y coordinates of this point. For example, the points

$$1 = (1,0) \quad i = (0,1) \quad -1 = (-1,0) \quad -i = (0,-1)$$

plotted in the figure are determined by the angles 0, $\frac{\pi}{2}$, π, $\frac{3\pi}{2}$, respectively. Hence

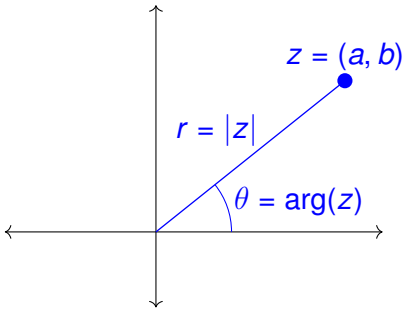
$$\begin{array}{llll} \cos 0 = 1 & \cos\frac{\pi}{2} = 0 & \cos\pi = -1 & \cos\frac{3\pi}{2} = 0 \\ \sin 0 = 0 & \sin\frac{\pi}{2} = 1 & \sin\pi = 0 & \sin\frac{3\pi}{2} = -1 \end{array}$$

Now we can describe the polar form of a complex number. Let $z = a + bi$ be a complex number, and write the absolute value of z as

$$r = |z| = \sqrt{a^2 + b^2}$$

If $z \neq 0$, the angle θ shown in the figure below is called an *argument* of z and is denoted

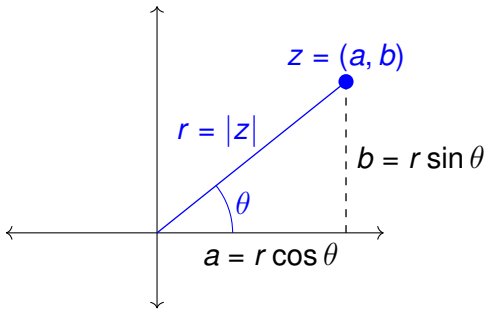
$$\theta = \arg z$$



This angle is not unique ($\theta + 2\pi k$ would do as well for any $k = 0, \pm 1, \pm 2, \ldots$). However, there is only one argument θ in the range $-\pi < \theta \leq \pi$, and this is sometimes called the *principal argument* of z.

Referring to the figure below, we find that the real and imaginary parts a and b of z are related to r and θ by

$$a = r\cos\theta$$
$$b = r\sin\theta$$



Hence the complex number $z = a + bi$ has the form

$$z = r(\cos\theta + i\sin\theta) \quad \text{where} \quad r = |z| \quad \text{and} \quad \theta = \arg(z)$$

The combination $\cos\theta + i\sin\theta$ is so important that a special notation is used:

$$e^{i\theta} = \cos\theta + i\sin\theta$$

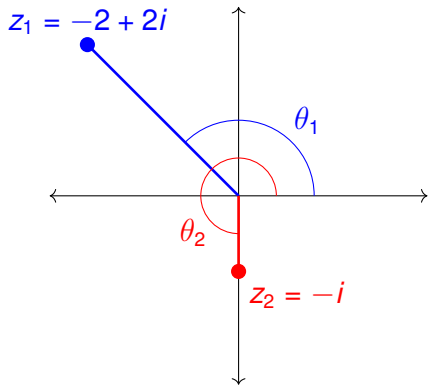
is called *Euler's formula* after the great Swiss mathematician Leonhard Euler (1707–1783). With this notation, z is written

$$z = re^{i\theta} \quad \text{where} \quad r = |z| \quad \text{and} \quad \theta = \arg(z)$$

This is a *polar form* of the complex number z. Of course it is not unique, because the argument can be changed by adding a multiple of 2π.

Example 237. Write $z_1 = -2 + 2i$ and $z_2 = -i$ in polar form.

Explanation. The two numbers are plotted in the complex plane, as shown in the figure below.



The absolute values are

$$r_1 = |-2 + 2i| = \sqrt{(-2)^2 + 2^2} = 2\sqrt{2}$$
$$r_2 = |-i| = \sqrt{0^2 + (-1)^2} = 1$$

By inspection, arguments of z_1 and z_2 are

$$\theta_1 = \arg(-2 + 2i) = \frac{3\pi}{4}$$
$$\theta_2 = \arg(-i) = \frac{3\pi}{2}$$

The corresponding polar forms are $z_1 = -2 + 2i = 2\sqrt{2}e^{3\pi i/4}$ and $z_2 = -i = e^{3\pi i/2}$. Of course, we could have taken the argument $-\dfrac{\pi}{2}$ for z_2 and obtained the polar form $z_2 = e^{-\pi i/2}$.

In Euler's formula $e^{i\theta} = \cos\theta + i\sin\theta$, the number e is the familiar constant $e = 2.71828\ldots$ from calculus. The reason for using e will not be given here; the reason why $\cos\theta + i\sin\theta$ is written as an *exponential* function of θ is that the *law of exponents* holds:

$$e^{i\theta} \cdot e^{i\phi} = e^{i(\theta+\phi)}$$

where θ and ϕ are any two angles. In fact, this is an immediate consequence of the addition identities for $\sin(\theta + \phi)$ and $\cos(\theta + \phi)$:

$$\begin{aligned} e^{i\theta}e^{i\phi} &= (\cos\theta + i\sin\theta)(\cos\phi + i\sin\phi) \\ &= (\cos\theta\cos\phi - \sin\theta\sin\phi) + i(\cos\theta\sin\phi + \sin\theta\cos\phi) \\ &= \cos(\theta+\phi) + i\sin(\theta+\phi) \\ &= e^{i(\theta+\phi)} \end{aligned}$$

This is analogous to the rule $e^a e^b = e^{a+b}$, which holds for real numbers a and b, so it is not unnatural to use the exponential notation $e^{i\theta}$ for the expression $\cos\theta + i\sin\theta$. In fact, a whole theory exists wherein functions such as e^z, $\sin z$, and $\cos z$ are studied, where z is a *complex* variable. Many deep and beautiful theorems can be proved in this theory, one of which is the so-called fundamental theorem of algebra mentioned later (Theorem **??**). We shall not pursue this here.

The geometric description of the multiplication of two complex numbers follows from the law of exponents.

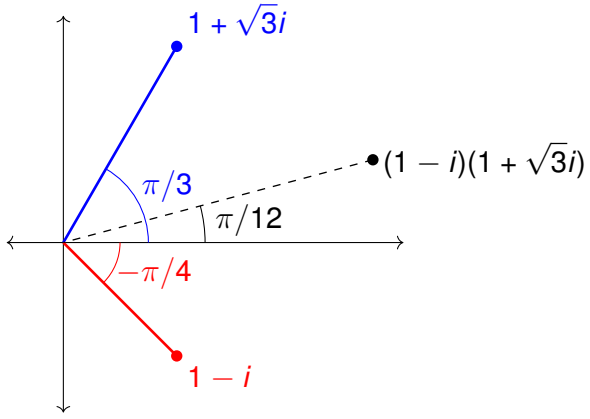
Theorem 136. If $z_1 = r_1 e^{i\theta_1}$ and $z_2 = r_2 e^{i\theta_2}$ are complex numbers in polar form, then

$$z_1 z_2 = r_1 r_2 e^{i(\theta_1+\theta_2)}$$

In other words, to multiply two complex numbers, simply multiply the absolute values and add the arguments. This simplifies calculations considerably, particularly when we observe that it is valid for *any* arguments θ_1 and θ_2.

Example 238. Multiply $(1-i)(1+\sqrt{3}i)$ in two ways.

Explanation. We have $|1-i| = \sqrt{2}$ and $|1+\sqrt{3}i| = 2$. Consider the figure below.



We have,

$$1 - i = \sqrt{2}e^{-i\pi/4}$$
$$1 + \sqrt{3}i = 2e^{i\pi/3}$$

Hence, by Theorem **??**,

$$\begin{aligned}(1 - i)(1 + \sqrt{3}i) &= (\sqrt{2}e^{-i\pi/4})(2e^{i\pi/3}) \\ &= 2\sqrt{2}e^{i(-\pi/4+\pi/3)} \\ &= 2\sqrt{2}e^{i\pi/12}\end{aligned}$$

This gives the required product in polar form. Of course, direct multiplication gives $(1 - i)(1 + \sqrt{3}i) = (\sqrt{3} + 1) + (\sqrt{3} - 1)i$. Hence, equating real and imaginary parts gives the formulas $\cos(\frac{\pi}{12}) = \frac{\sqrt{3}+1}{2\sqrt{2}}$ and $\sin(\frac{\pi}{12}) = \frac{\sqrt{3}-1}{2\sqrt{2}}$.

Roots of Unity

If a complex number $z = re^{i\theta}$ is given in polar form, the powers assume a particularly simple form. In fact, $z^2 = (re^{i\theta})(re^{i\theta}) = r^2e^{2i\theta}$, $z^3 = z^2 \cdot z = (r^2e^{2i\theta})(re^{i\theta}) = r^3e^{3i\theta}$, and so on. Continuing in this way, it follows by induction that the following theorem holds for any positive integer n. The name honors Abraham De Moivre (1667–1754).

Theorem 137 (De Moivre's Theorem)**.** If θ is any angle, then $(e^{i\theta})^n = e^{in\theta}$ holds for all integers n.

Proof The case $n > 0$ has been discussed, and the reader can verify the result for $n = 0$. To derive it for $n < 0$, first observe that

$$\text{if} \quad z = re^{i\theta} \neq 0 \quad \text{then} \quad z^{-1} = \frac{1}{r}\, e^{-i\theta}$$

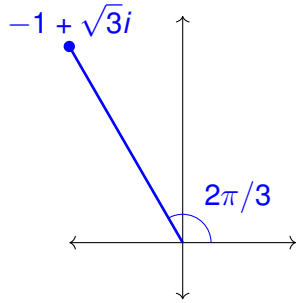
In fact, $(re^{i\theta})(\frac{1}{r}e^{-i\theta}) = 1e^{i0} = 1$ by the multiplication rule. Now assume that n is negative and write it as $n = -m$, $m > 0$. Then

$$(re^{i\theta})^n = [(re^{i\theta})^{-1}]^m = (\frac{1}{r}\, e^{-i\theta})^m = r^{-m}e^{i(-m\theta)} = r^n e^{in\theta}$$

If $r = 1$, this is De Moivre's theorem for negative n. ■

Example 239. Verify that $(-1 + \sqrt{3}i)^3 = 8$.

Explanation. We have $|-1+\sqrt{3}i| = 2$, so $-1+\sqrt{3}i = 2e^{2\pi i/3}$ (see the figure below).



De Moivre's theorem gives

$$(-1+\sqrt{3}i)^3 = (2e^{2\pi i/3})^3 = 8e^{3(2\pi i/3)} = 8e^{2\pi i} = 8$$

De Moivre's theorem can be used to find nth roots of complex numbers where n is positive. The next example illustrates this technique.

Example 240. Find the cube roots of unity; that is, find all complex numbers z such that $z^3 = 1$.

Explanation. First write $z = re^{i\theta}$ and $1 = 1e^{i0}$ in polar form. We must use the condition $z^3 = 1$ to determine r and θ. Because $z^3 = r^3e^{3i\theta}$ by De Moivre's theorem, this requirement becomes

$$r^3e^{3i\theta} = 1e^{0i}$$

These two complex numbers are equal, so their absolute values must be equal and the arguments must either be equal or differ by an integral multiple of 2π:

$$r^3 = 1$$
$$3\theta = 0 + 2k\pi, \quad k \text{ some integer}$$

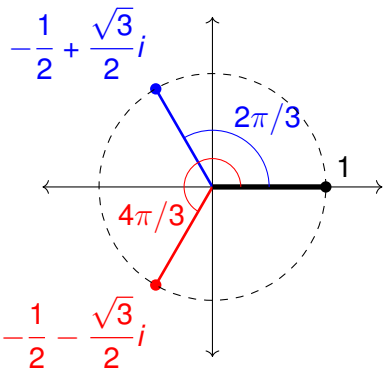
Because r is real and positive, the condition $r^3 = 1$ implies that $r = 1$. However,

$$\theta = \frac{2k\pi}{3}, \quad k \text{ some integer}$$

seems at first glance to yield infinitely many different angles for z. However, choosing $k = 0, 1, 2$ gives three possible arguments θ (where $0 \leq \theta < 2\pi$), and the corresponding roots are

$$1e^{0i} = 1$$
$$1e^{2\pi i/3} = -\frac{1}{2} + \frac{\sqrt{3}}{2}i$$
$$1e^{4\pi i/3} = -\frac{1}{2} - \frac{\sqrt{3}}{2}i$$

These are displayed in the figure below.



All other values of k yield values of θ that differ from one of these by a multiple of 2π—and so do not give new roots. Hence we have found all the roots.

The same type of calculation gives all complex *nth roots of unity*; that is, all complex numbers z such that $z^n = 1$. As before, write $1 = 1e^{0i}$ and

$$z = re^{i\theta}$$

in polar form. Then $z^n = 1$ takes the form

$$r^n e^{ni\theta} = 1e^{0i}$$

using De Moivre's theorem. Comparing absolute values and arguments yields

$$\begin{aligned} r^n &= 1 \\ n\theta &= 0 + 2k\pi, \quad k \text{ some integer} \end{aligned}$$

Hence $r = 1$, and the n values

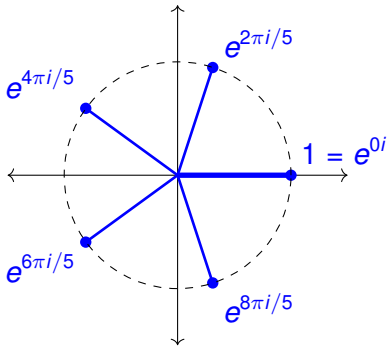
$$\theta = \frac{2k\pi}{n}, \quad k = 0, 1, 2, \ldots, n-1$$

of θ all lie in the range $0 \leq \theta < 2\pi$. As in Example **??**, *every* choice of k yields a value of θ that differs from one of these by a multiple of 2π, so these give the arguments of *all* the possible roots.

Theorem 138 (nth Roots of Unity)**.** If $n \geq 1$ is an integer, the nth roots of unity (that is, the solutions to $z^n = 1$) are given by

$$z = e^{2\pi ki/n}, \quad k = 0, 1, 2, \ldots, n-1$$

The nth roots of unity can be found geometrically as the points on the unit circle that cut the circle into n equal sectors, starting at 1. The case $n = 5$ is shown in the figure below, where the five fifth roots of unity are plotted.



The method just used to find the nth roots of unity works equally well to find the nth roots of any complex number in polar form. We give one example.

Example 241. Find the fourth roots of $\sqrt{2} + \sqrt{2}i$.

Explanation. First write $\sqrt{2} + \sqrt{2}i = 2e^{\pi i/4}$ in polar form. If $z = re^{i\theta}$ satisfies $z^4 = \sqrt{2} + \sqrt{2}i$, then De Moivre's theorem gives

$$r^4 e^{i(4\theta)} = 2e^{\pi i/4}$$

Hence $r^4 = 2$ and $4\theta = \dfrac{\pi}{4} + 2k\pi$, k an integer. We obtain four distinct roots (and hence all) by

$$r = \sqrt[4]{2}, \quad \theta = \frac{\pi}{16} = \frac{2k\pi}{16}, k = 0, 1, 2, 3$$

Thus the four roots are

$$\sqrt[4]{2}e^{\pi i/16} \quad \sqrt[4]{2}e^{9\pi i/16} \quad \sqrt[4]{2}e^{17\pi i/16} \quad \sqrt[4]{2}e^{25\pi i/16}$$

Of course, reducing these roots to the form $a + bi$ would require the computation of $\sqrt[4]{2}$ and the sine and cosine of the various angles.

An expression of the form $ax^2 + bx + c$, where the coefficients $a \neq 0$, b, and c are real numbers, is called a *real quadratic*. A complex number u is called a *root* of the quadratic if $au^2 + bu + c = 0$. The roots are given by the famous *quadratic formula*:

$$u = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

The quantity $d = b^2 - 4ac$ is called the *discriminant* of the quadratic $ax^2 + bx + c$, and there is no real root if and only if $d < 0$. In this case the quadratic is said to be *irreducible*. Moreover, the fact that $d < 0$ means that $\sqrt{d} = i\sqrt{|d|}$, so the two (complex) roots are conjugates of each other:

$$u = \frac{1}{2a}(-b + i\sqrt{|d|}) \quad \text{and} \quad \overline{u} = \frac{1}{2a}(-b - i\sqrt{|d|})$$

The converse of this is true too: Given any nonreal complex number u, then u and $\overline{u}$ are the roots of some real irreducible quadratic. Indeed, the quadratic

$$x^2 - (u + \overline{u})x + u\overline{u} = (x - u)(x - \overline{u})$$

has real coefficients ($u\overline{u} = |u|^2$ and $u + \overline{u}$ is twice the real part of u) and so is irreducible because its roots u and $\overline{u}$ are not real.

Example 242. Find a real irreducible quadratic with $u = 3 - 4i$ as a root.

Explanation. We have $u + \overline{u} = 6$ and $|u|^2 = 25$, so $x^2 - 6x + 25$ is irreducible with u and $\overline{u} = 3 + 4i$ as roots.

Fundamental Theorem of Algebra

As we mentioned earlier, the complex numbers are the culmination of a long search by mathematicians to find a set of numbers large enough to contain a root of every polynomial. The fact that the complex numbers have this property was first proved by Gauss in 1797 when he was 20 years old. The proof is omitted.

Theorem 139 (Fundamental Theorem of Algebra)**.** Every polynomial of positive degree with complex coefficients has a complex root.

If $f(x)$ is a polynomial with complex coefficients, and if u_1 is a root, then the Factor Theorem (see Precalculus by Stitz-Zeager, for instance) asserts that

$$f(x) = (x - u_1)g(x)$$

where $g(x)$ is a polynomial with complex coefficients and with degree one less than the degree of $f(x)$. Suppose that u_2 is a root of $g(x)$, again by the fundamental theorem. Then $g(x) = (x - u_2)h(x)$, so

$$f(x) = (x - u_1)(x - u_2)h(x)$$

This process continues until the last polynomial to appear is linear. Thus $f(x)$ has been expressed as a product of linear factors. The last of these factors can be written in the form $u(x - u_n)$, where u and u_n are complex (verify this), so the fundamental theorem takes the following form.

Theorem 140. Every complex polynomial $f(x)$ of degree $n \geq 1$ has the form

$$f(x) = u(x - u_1)(x - u_2) \cdots (x - u_n)$$

where $u, u_1, \ldots, u_n$ are complex numbers and $u \neq 0$. The numbers $u_1, u_2, \ldots, u_n$ are the roots of $f(x)$ (and need not all be distinct), and u is the coefficient of x^n.

This form of the fundamental theorem, when applied to a polynomial $f(x)$ with *real* coefficients, can be used to deduce the following result.

Theorem 141. Every polynomial $f(x)$ of positive degree with real coefficients can be factored as a product of linear and irreducible quadratic factors.

In fact, suppose $f(x)$ has the form

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

where the coefficients a_i are real. If u is a complex root of $f(x)$, then we claim first that $\overline{u}$ is also a root. In fact, we have $f(u) = 0$, so

$$\begin{aligned} 0 = \overline{0} = \overline{f(u)} &= \overline{a_n u^n + a_{n-1}u^{n-1} + ... + a_1 u + a_0} \\ &= \overline{a_n u^n} + \overline{a_{n-1}u^{n-1}} + ... + \overline{a_1 u} + \overline{a_0} \\ &= \overline{a}_n \overline{u}^n + \overline{a}_{n-1}\overline{u}^{n-1} + ... + \overline{a}_1 \overline{u} + \overline{a}_0 \\ &= a_n \overline{u}^n + a_{n-1}\overline{u}^{n-1} + ... + a_1 \overline{u} + a_0 \\ &= f(\overline{u}) \end{aligned}$$

where $\overline{a}_i = a_i$ for each i because the coefficients a_i are real. Thus if u is a root of $f(x)$, so is its conjugate $\overline{u}$. Of course some of the roots of $f(x)$ may be real (and so equal their conjugates), but the nonreal roots come in pairs, u and $\overline{u}$. By Theorem **??**, we can thus write $f(x)$ as a product:

$$f(x) = a_n(x - r_1)\cdots(x - r_k)(x - u_1)(x - \overline{u}_1)\cdots(x - u_m)(x - \overline{u}_m) \tag{128}$$

where a_n is the coefficient of x^n in $f(x)$; $r_1, r_2, ..., r_k$ are the real roots; and $u_1, \overline{u}_1, u_2, \overline{u}_2, ..., u_m, \overline{u}_m$ are the nonreal roots. But the product

$$(x - u_j)(x - \overline{u}_j) = x^2 - (u_j + \overline{u}_j)x + (u_j\overline{u}_j)$$

is a real irreducible quadratic for each j (see the discussion preceding Example **??**). Hence (**??**) shows that $f(x)$ is a product of linear and irreducible quadratic factors, each with real coefficients. This is the conclusion in Theorem **??**.

Practice Problems

Problem 623 *Solve each of the following for the real number x.*

1. $x - 4i = (2 - i)^2$

$x =$ ... ...

2. $(2 + xi)(3 - 2i) = 12 + 5i$

3. $(2 + xi)^2 = 4$

$x =$ ... ... *or* $x =$ ... ...

4. $(2 + xi)(2 - xi) = 5$

Problem 624 *Convert each of the following to the form* $a + bi$.

1. *(challenge problem)* $(2 - 3i) - 2(2 - 3i) + 9$

2. *(challenge problem)* $(3 - 2i)(1 + i) + |3 + 4i|$

... ... $+ i$

3. $\dfrac{1 + i}{2 - 3i} + \dfrac{1 - i}{-2 + 3i}$

4. $\dfrac{3 - 2i}{1 - i} + \dfrac{3 - 7i}{2 - 3i}$

$\dfrac{\text{... ...}}{\text{... ...}} + \dfrac{\text{... ...}}{\text{... ...}}i$

5. i^{131}

6. $(2 - i)^3$

... ... + ... i

7. $(1 + i)^4$

... ... + ... i

8. $(1 - i)^2(2 + i)^2$

9. *(challenge problem)* $\dfrac{3\sqrt{3} - i}{\sqrt{3} + i} + \dfrac{\sqrt{3} + 7i}{\sqrt{3} - i}$

Problem 625 *In each case, find the complex number* z.

1. $iz - (1 + i)^2 = 3 - i$

2. *(challenge problem)* $(i + z) - 3i(2 - z) = iz + 1$

$\dfrac{\text{... ...}}{\text{... ...}} + \dfrac{\text{... ...}}{\text{... ...}}i$

3. $z^2 = -i$
4. $z^2 = 3 - 4i$

   $\pm(... ... - i)$
5. $z(1 + i) = \overline{z} + (3 + 2i)$
6. *(challenge problem)* $z(2 - i) = (\overline{z} + 1)(1 + i)$

   Click the arrow to see the answer.

   $1 + i$

Problem 626 *In each case, find the roots of the real quadratic equation.*

1. $x^2 - 2x + 3 = 0$
2. $x^2 - x + 1 = 0$

   Click the arrow to see the answer.

   $\frac{1}{2} \pm \frac{\sqrt{3}}{2} i$
3. $3x^2 - 4x + 2 = 0$
4. $2x^2 - 5x + 2 = 0$

   Click the arrow to see the answer.

   $2, \frac{1}{2}$

Problem 627 *Find all numbers x in each case.*

1. $x^3 = 8$
2. $x^3 = -8$

   Click the arrow to see the answer.

   $-2, 1 \pm \sqrt{3}i$
3. $x^4 = 16$
4. $x^4 = 64$

   Click the arrow to see the answer.

   $\pm 2\sqrt{2}, \pm 2\sqrt{i}$

Problem 628 *In each case, find a real quadratic with u as a root, and find the other root.*

1. $u = 1 + i$
2. $u = 2 - 3i$

   Click the arrow to see the answer.

   $$x^2 - 4x + 13; 2 + 3i$$

3. $u = -i$
4. $u = 3 - 4i$

   Click the arrow to see the answer.

   $$x^2 - 6x + 25; 3 + 4i$$

Problem 629 *Find the roots of* $x^2 - 2\cos\theta x + 1 = 0$, θ *any angle.*

Problem 630 *Find a real polynomial of degree* 4 *with* $2 - i$ *and* $3 - 2i$ *as roots.*

Click the arrow to see the answer.

$$x^4 - 10x^3 + 42x^2 - 82x + 65$$

Problem 631 *Let re z and im z denote, respectively, the real and imaginary parts of z. Show that:*

1. $im(iz) = re\ z$
2. $re(iz) = -im\ z$
3. $z + \overline{z} = 2rez$
4. $z - \overline{z} = 2iimz$
5. *(challenge problem)* $re(z + w) = re\ z + re\ w$, *and* $re(tz) = t \cdot re\ z$ *if t is real*
6. *(challenge problem)* $im(z + w) = im\ z + im\ w$, *and* $im(tz) = t \cdot im\ z$ *if t is real*

Problem 632 *In each case, show that u is a root of the quadratic equation, and find the other root.*

1. *(challenge problem)* $x^2 - 3ix + (-3 + i) = 0$*;* $u = 1 + i$
2. *(challenge problem)* $x^2 + ix - (4 - 2i) = 0$*;* $u = -2$
3. *(challenge problem)* $x^2 - (3 - 2i)x + (5 - i) = 0$*;* $u = 2 - 3i$
4. *(challenge problem)* $x^2 + 3(1 - i)x - 5i = 0$*;* $u = -2 + i$

Problem 633 *Find the roots of each of the following complex quadratic equations.*

1. $x^2 + 2x + (1 + i) = 0$
2. $x^2 - x + (1 - i) = 0$

   Click the arrow to see the answer.

   $-i, 1 + i$
3. *(challenge problem)* $x^2 - (2 - i)x + (3 - i) = 0$
4. *(challenge problem)* $x^2 - 3(1 - i)x - 5i = 0$

   Click the arrow to see the answer.

   $2 - i, 1 - 2i$

Problem 634 *In each case, describe the graph of the equation (where z denotes a complex number).*

1. $|z| = 1$
2. $|z - 1| = 2$

   Click the arrow to see the answer. Circle, centre at 1*, radius* 2
3. $z = i\overline{z}$
4. $z = -\overline{z}$

   Click the arrow to see the answer. Imaginary axis
5. $z = |z|$
6. *im* $z = m \cdot$ *re* z*, m a real number*

   Click the arrow to see the answer. Line $y = mx$

Problem 635 *a. Verify* $|zw| = |z||w|$ *directly for* $z = a + bi$ *and* $w = c + di$*.*

b. Deduce (a) from properties C2 and C6.

Problem 636 *Prove that* $|z+w| = |z|^2 + |w|^2 + w\overline{z} + \overline{w}z$ *for all complex numbers* w *and* z*.*

Problem 637 *If* zw *is real and* $z \neq 0$*, show that* $w = a\overline{z}$ *for some real number* a*.*

Problem 638 *If* $zw = \overline{z}v$ *and* $z \neq 0$*, show that* $w = uv$ *for some* u *in* $\mathbb{C}$ *with* $|u| = 1$*.*

Problem 639 *Show that* $(1+i)^n + (1-i)^n$ *is real for all* n*, using property C5.*

Problem 640 *Express each of the following in polar form (use the principal argument).*

1. $3 - 3i$
2. $-4i$

   Click the arrow to see the answer.

   $4e^{-\pi i/2}$
3. $-\sqrt{3} + i$
4. $-4 + 4\sqrt{3}i$

   Click the arrow to see the answer.

   $8e^{2\pi i/3}$
5. $-7i$
6. $-6 + 6i$

   Click the arrow to see the answer.

   $6\sqrt{2}e^{3\pi i/4}$

Problem 641 *Express each of the following in the form* $a + bi$*.*

1. $3e^{\pi i}$
2. $e^{7\pi i/3}$

   Click the arrow to see the answer.

   $\frac{1}{2} + \frac{\sqrt{3}}{2}i$

3. $2e^{3\pi i/4}$
4. $\sqrt{2}e^{-\pi i/4}$

   Click the arrow to see the answer.

   $1 - i$

5. $e^{5\pi i/4}$
6. $2\sqrt{3}e^{-2\pi i/6}$

   Click the arrow to see the answer.

   $\sqrt{3} - 3i$

Problem 642 *Express each of the following in the form $a + bi$.*

1. $(-1 + \sqrt{3}i)^2$
2. $(1 + \sqrt{3}i)^{-4}$

   Click the arrow to see the answer.

   $$-\frac{1}{32} + \frac{\sqrt{3}}{32}i$$

3. $(1 + i)^8$
4. $(1 - i)^{10}$

   Click the arrow to see the answer.

   $-32i$

5. $(1 - i)^6(\sqrt{3} + i)^3$
6. $(\sqrt{3} - i)^9(2 - 2i)^5$

   Click the arrow to see the answer.

   $-2^{16}(1 + i)$

Problem 643 *Use De Moivre's theorem to show that:*

a. $\cos 2\theta = \cos^2\theta - \sin^2\theta$; $\sin 2\theta = 2\cos\theta\sin\theta$

b. $\cos 3\theta = \cos^3\theta - 3\cos\theta\sin^2\theta$;
   $\sin 3\theta = 3\cos^2\theta\sin\theta - \sin^3\theta$

Problem 644 *a. Find the fourth roots of unity.*

b. Find the sixth roots of unity.

Problem 645 *Find all complex numbers z such that:*

1. $z^4 = -1$
2. $z^4 = 2(\sqrt{3}i - 1)$

   Click the arrow to see the answer.

   $$\pm\frac{\sqrt{2}}{2}(\sqrt{3} + i), \pm\frac{\sqrt{2}}{2}(-1 + \sqrt{3}i)$$

3. $z^3 = -27i$
4. $z^6 = -64$

   Click the arrow to see the answer.

   $$\pm 2i, \pm(\sqrt{3} + i), \pm(\sqrt{3} - i)$$

Problem 646 *If $z = re^{i\theta}$ in polar form, show that:*

1. $\overline{z} = re^{-i\theta}$
2. $z^{-1} = \frac{1}{r}e^{-i\theta}$ *if* $z \neq 0$

Problem 647 *Show that the sum of the nth roots of unity is zero.*

Problem 648 *1. Let z_1, z_2, z_3, z_4, and z_5 be equally spaced around the unit circle. Show that $z_1 + z_2 + z_3 + z_4 + z_5 = 0$.*

2. Repeat (a) for any $n \geq 2$ points equally spaced around the unit circle.

3. If $|w| = 1$, show that the sum of the roots of $z^n = w$ is zero.

Problem 649 *If z^n is real, $n \geq 1$, show that $(\overline{z})^n$ is real.*

Problem 650 *If $\overline{z}^2 = z^2$, show that z is real or pure imaginary.*

Problem 651 *If a and b are rational numbers, let p and q denote numbers of the form $a+b\sqrt{2}$. If $p=a+b\sqrt{2}$, define $\tilde{p}=a-b\sqrt{2}$ and $[p]=a^2-2b^2$. Show that each of the following holds.*

1. *(challenge problem) $a+b\sqrt{2}=a_1+b_1\sqrt{2}$ only if $a=a_1$ and $b=b_1$*
2. *$\widetilde{p\pm q}=\tilde{p}\pm\tilde{q}$*
3. *$\widetilde{pq}=\tilde{p}\tilde{q}$*
4. *$[p]=p\tilde{p}$*
5. *$[pq]=[p][q]$*
6. *(challenge problem) If $f(x)$ is a polynomial with rational coefficients and $p=a+b\sqrt{2}$ is a root of $f(x)$, then $\tilde{p}$ is also a root of $f(x)$.*

Text Source

This section was adapted from Appendix A of Keith Nicholson's *Linear Algebra with Applications*. (CC-BY-NC-SA)

W. Keith Nicholson, *Linear Algebra with Applications*, Lyryx 2018, Open Edition, pp. 581–594.

98 Complex Matrices

Nearly everything we have studied in this book would remain true if the phrase *real number* were replaced by *complex number* wherever it occurs. Then we would deal with matrices with complex entries, systems of linear equations with complex coefficients (and complex solutions), determinants of complex matrices, and vector spaces with scalar multiplication by any complex number allowed. Moreover, the proofs of most theorems about (the real version of) these concepts extend easily to the complex case. It is not our intention here to give a full treatment of complex linear algebra. However, we will carry the theory far enough to give another proof of the Real Spectral Theorem (**??**).

The set of complex numbers is denoted $\mathbb{C}$. We will use only the most basic properties of these numbers (mainly conjugation and absolute values), and the reader can find this material in Complex Numbers.

If $n \geq 1$, we denote the set of all n-tuples of complex numbers by $\mathbb{C}^n$. As with $\mathbb{R}^n$, these n-tuples will be written either as row or column matrices and will be referred to as *vectors*. We define vector operations on $\mathbb{C}^n$ as follows:

$$[v_1, v_2, \ldots, v_n] + [w_1, w_2, \ldots, w_n] = [v_1 + w_1, v_2 + w_2, \ldots, v_n + w_n]$$
$$u[v_1, v_2, \ldots, v_n] = [uv_1, uv_2, \ldots, uv_n] \quad \text{for } u \text{ in } \mathbb{C}$$

With these definitions, $\mathbb{C}^n$ satisfies the axioms for a vector space (with complex scalars) given in Abstract Vector Spaces. Thus we can speak of spanning sets for $\mathbb{C}^n$, of linearly independent subsets, and of bases. In all cases, the definitions are identical to the real case, except that the scalars are allowed to be complex numbers. In particular, the standard basis of $\mathbb{R}^n$ remains a basis of $\mathbb{C}^n$, called the *standard basis* of $\mathbb{C}^n$.

A Generalization of the Dot Product for Complex Vectors

There is a generalization to $\mathbb{C}^n$ of the dot product in $\mathbb{R}^n$.

Definition 106. Given $\mathbf{z} = [z_1, z_2, \ldots, z_n]$ and $\mathbf{w} = [w_1, w_2, \ldots, w_n]$ in $\mathbb{C}^n$, define their *standard inner product* $\langle \mathbf{z}, \mathbf{w} \rangle$ by

$$\langle \mathbf{z}, \mathbf{w} \rangle = z_1\overline{w}_1 + z_2\overline{w}_2 + \ldots + z_n\overline{w}_n$$

where $\overline{w}$ is the conjugate of the complex number w.

Clearly, if $\mathbf{z}$ and $\mathbf{w}$ actually lie in $\mathbb{R}^n$, then $\langle \mathbf{z}, \mathbf{w} \rangle = \mathbf{z} \cdot \mathbf{w}$ is the usual dot product.

Example 243. If $\mathbf{z} = [2, 1-i, 2i, 3-i]$ and $\mathbf{w} = [1-i, -1, -i, 3+2i]$, then

$$\langle \mathbf{z}, \mathbf{w} \rangle = 2(1+i) + (1-i)(-1) + (2i)(i) + (3-i)(3-2i) = 6 - 6i$$
$$\langle \mathbf{z}, \mathbf{z} \rangle = 2 \cdot 2 + (1-i)(1+i) + (2i)(-2i) + (3-i)(3+i) = 20$$

Note that $\langle \mathbf{z}, \mathbf{w} \rangle$ is a complex number in general, as opposed to requiring an inner product to be real as we do in Inner Product Spaces. However, if $\mathbf{w} = \mathbf{z} = [z_1, z_2, \ldots, z_n]$, the definition gives $\langle \mathbf{z}, \mathbf{z} \rangle = |z_1|^2 + \ldots + |z_n|^2$ which is a nonnegative real number, equal to 0 if and only if $\mathbf{z} = \mathbf{0}$. This explains the conjugation in the definition of $\langle \mathbf{z}, \mathbf{w} \rangle$, and it gives **??** of the following theorem.

Theorem 142. Let $\mathbf{z}$, $\mathbf{z}_1$, $\mathbf{w}$, and $\mathbf{w}_1$ denote vectors in $\mathbb{C}^n$, and let λ denote a complex number.

1. $\langle \mathbf{z} + \mathbf{z}_1, \mathbf{w} \rangle = \langle \mathbf{z}, \mathbf{w} \rangle + \langle \mathbf{z}_1, \mathbf{w} \rangle$ and $\langle \mathbf{z}, \mathbf{w} + \mathbf{w}_1 \rangle = \langle \mathbf{z}, \mathbf{w} \rangle + \langle \mathbf{z}, \mathbf{w}_1 \rangle$.
2. $\langle \lambda \mathbf{z}, \mathbf{w} \rangle = \lambda \langle \mathbf{z}, \mathbf{w} \rangle$ and $\langle \mathbf{z}, \lambda \mathbf{w} \rangle = \overline{\lambda} \langle \mathbf{z}, \mathbf{w} \rangle$.
3. $\langle \mathbf{z}, \mathbf{w} \rangle = \overline{\langle \mathbf{w}, \mathbf{z} \rangle}$.
4. $\langle \mathbf{z}, \mathbf{z} \rangle \geq 0$, and $\langle \mathbf{z}, \mathbf{z} \rangle = 0$ if and only if $\mathbf{z} = \mathbf{0}$.

Proof We leave **??** and **??** to the reader (Practice Problem **??**), and **??** has already been proved. To prove **??**, write $\mathbf{z} = [z_1, z_2, \ldots, z_n]$ and $\mathbf{w} = [w_1, w_2, \ldots, w_n]$. Then

$$\begin{aligned} \overline{\langle \mathbf{w}, \mathbf{z} \rangle} = (\overline{w_1 \overline{z}_1 + \ldots + w_n \overline{z}_n}) &= \overline{w}_1 \overline{\overline{z}}_1 + \ldots + \overline{w}_n \overline{\overline{z}}_n \\ &= z_1 \overline{w}_1 + \ldots + z_n \overline{w}_n = \langle \mathbf{z}, \mathbf{w} \rangle \end{aligned}$$

■

Definition 107. As for the dot product on $\mathbb{R}^n$, property **??** enables us to define the *norm* or *length* $\|\mathbf{z}\|$ of a vector $\mathbf{z} = (z_1, z_2, \ldots, z_n)$ in $\mathbb{C}^n$:

$$\|\mathbf{z}\| = \sqrt{\langle \mathbf{z}, \mathbf{z} \rangle} = \sqrt{|z_1|^2 + |z_2|^2 + \ldots + |z_n|^2}$$

The only properties of the norm function we will need are the following (the proofs are left to the reader):

Theorem 143. If $\mathbf{z}$ is any vector in $\mathbb{C}^n$, then

1. $\|\mathbf{z}\| \geq 0$ and $\|\mathbf{z}\| = 0$ if and only if $\mathbf{z} = \mathbf{0}$.
2. $\|\lambda \mathbf{z}\| = |\lambda| \, \|\mathbf{z}\|$ for all complex numbers λ.

A vector $\mathbf{u}$ in $\mathbb{C}^n$ is called a *unit vector* if $\|\mathbf{u}\| = 1$. Property **??** in Theorem **??** then shows that if $\mathbf{z} \neq \mathbf{0}$ is any nonzero vector in $\mathbb{C}^n$, then $\mathbf{u} = \dfrac{1}{\|\mathbf{z}\|}\mathbf{z}$ is a unit vector.

Example 244. In $\mathbb{C}^4$, find a unit vector $\mathbf{u}$ that is a positive real multiple of $\mathbf{z} = \begin{bmatrix} 1-i \\ i \\ 2 \\ 3+4i \end{bmatrix}$.

Explanation. $\|\mathbf{z}\| = \sqrt{2+1+4+25} = \sqrt{32} = 4\sqrt{2}$, so take $\mathbf{u} = \dfrac{1}{4\sqrt{2}}\mathbf{z}$.

A matrix $A = \begin{bmatrix} a_{ij} \end{bmatrix}$ is called a *complex matrix* if every entry a_{ij} is a complex number. The notion of conjugation for complex numbers extends to matrices as follows: Define the *conjugate* of $A = \begin{bmatrix} a_{ij} \end{bmatrix}$ to be the matrix

$$\overline{A} = \begin{bmatrix} \overline{a}_{ij} \end{bmatrix}$$

obtained from A by conjugating every entry. Then (using Appendix **??**)

$$\overline{A+B} = \overline{A} + \overline{B} \quad \text{and} \quad \overline{AB} = \overline{A}\,\overline{B}$$

holds for all (complex) matrices of appropriate size.

Transposition of complex matrices is defined just as in the real case, and the following notion is fundamental.

Definition 108. The *conjugate transpose* A^H of a complex matrix A is defined by

$$A^H = (\overline{A})^T = \overline{(A^T)}$$

Observe that $A^H = A^T$ when A is real.

Remark 25. Other notations for A^H are A^* and $A^\dagger$.

Example 245.

$$\begin{bmatrix} 3 & 1-i & 2+i \\ 2i & 5+2i & -i \end{bmatrix}^H = \begin{bmatrix} 3 & -2i \\ 1+i & 5-2i \\ 2-i & i \end{bmatrix}$$

The following properties of A^H follow easily from the rules for transposition of real matrices and extend these rules to complex matrices. Note the conjugate in property **??**.

Theorem 144. Let A and B denote complex matrices, and let λ be a complex number.

1. $(A^H)^H = A$.
2. $(A + B)^H = A^H + B^H$.
3. $(\lambda A)^H = \overline{\lambda} A^H$.
4. $(AB)^H = B^H A^H$.

Hermitian and Unitary Matrices

If A is a real symmetric matrix, it is clear that $A^H = A$. The complex matrices that satisfy this condition turn out to be the most natural generalization of the real symmetric matrices:

Definition 109. A square complex matrix A is called *Hermitian* if $A^H = A$, equivalently $\overline{A} = A^T$.

Remark 26. The name Hermitian honours Charles Hermite (1822–1901), a French mathematician who worked primarily in analysis and is remembered as the first to show that the number e from calculus is transcendental—that is, e is not a root of any polynomial with integer coefficients.

Hermitian matrices are easy to recognize because the entries on the main diagonal must be real, and the "reflection" of each off-diagonal entry in the main diagonal must be the conjugate of that entry.

Example 246. $\begin{bmatrix} 3 & i & 2+i \\ -i & -2 & -7 \\ 2-i & -7 & 1 \end{bmatrix}$ is Hermitian, whereas $\begin{bmatrix} 1 & i \\ i & -2 \end{bmatrix}$ and $\begin{bmatrix} 1 & i \\ -i & i \end{bmatrix}$ are not.

The following theorem extends Theorem **??**, and gives a very useful characterization of Hermitian matrices in terms of the standard inner product in $\mathbb{C}^n$.

Theorem 145. An $n \times n$ complex matrix A is Hermitian if and only if

$$\langle A\mathbf{z}, \mathbf{w} \rangle = \langle \mathbf{z}, A\mathbf{w} \rangle$$

for all n-tuples $\mathbf{z}$ and $\mathbf{w}$ in $\mathbb{C}^n$.

Proof If A is Hermitian, we have $A^T = \overline{A}$. If $\mathbf{z}$ and $\mathbf{w}$ are columns in $\mathbb{C}^n$, then $\langle \mathbf{z}, \mathbf{w} \rangle = \mathbf{z}^T \overline{\mathbf{w}}$, so

$$\langle A\mathbf{z}, \mathbf{w} \rangle = (A\mathbf{z})^T \overline{\mathbf{w}} = \mathbf{z}^T A^T \overline{\mathbf{w}} = \mathbf{z}^T \overline{A} \overline{\mathbf{w}} = \mathbf{z}^T (\overline{A\mathbf{w}}) = \langle \mathbf{z}, A\mathbf{w} \rangle$$

To prove the converse, let $\mathbf{e}_j$ denote column j of the identity matrix. If $A = \left[a_{ij} \right]$, the condition gives

$$\overline{a}_{ij} = \langle \mathbf{e}_i, A\mathbf{e}_j \rangle = \langle A\mathbf{e}_i, \mathbf{e}_j \rangle = a_{ij}$$

Hence $\overline{A} = A^T$, so A is Hermitian. ■

Let A be an $n \times n$ complex matrix. As in the real case, a complex number λ is called an *eigenvalue* of A if $A\mathbf{x} = \lambda\mathbf{x}$ holds for some column $\mathbf{x} \neq \mathbf{0}$ in $\mathbb{C}^n$. In this case $\mathbf{x}$ is called an *eigenvector* of A corresponding to λ.

Definition 110. Let A be an $n \times n$ complex matrix. Then

$$c_A(z) = \det(zI - A)$$

is a degree n polynomial (in z) called the *characteristic polynomial* of A.

Remark 27. Some books define the characteristic polynomial to be $\det(A - zI)$ (the left side of the characteristic equation **??**). Our definition ensures that the characteristic polynomial is a monic polynomial (i.e., its leading coefficient is 1). The two polynomials will have the same roots, which are the eigenvalues of the matrix A.

If A is an $n \times n$ matrix, the characteristic polynomial $c_A(z)$ is a polynomial of degree n and the eigenvalues of A are just the roots of $c_A(z)$. In most of our examples these roots have been *real* numbers (in fact, the examples have been carefully chosen so this will be the case!); but it need not happen, even when the characteristic polynomial has real coefficients. For example, if $A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$ then $c_A(z) = z^2 + 1$ has roots i and $-i$, where i is a complex number satisfying $i^2 = -1$. Therefore, we have to deal with the possibility that the eigenvalues of a (real) square matrix might be complex numbers.

For a complex matrix, $c_A(z)$ has complex coefficients (possibly nonreal). However, an argument like that given in Exploration **??** of The Characteristic Equation still works to show that the eigenvalues of A are the roots (possibly complex) of $c_A(z)$.

It is at this point that the advantage of working with complex numbers becomes apparent. The real numbers are incomplete in the sense that the characteristic polynomial of a real matrix may fail to have all its roots real. However, this difficulty does not occur for the complex numbers. The so-called fundamental theorem of algebra ensures that *every* polynomial of positive degree with complex coefficients has a complex root. Hence every square complex matrix A has a (complex) eigenvalue. Indeed (see **??**), $c_A(z)$ factors completely as follows:

$$c_A(z) = (z - \lambda_1)(z - \lambda_2) \cdots (z - \lambda_n)$$

where $\lambda_1, \lambda_2, \ldots, \lambda_n$ are the eigenvalues of A (with possible repetitions due to multiple roots).

The next result extends Theorem **??**, which asserts that eigenvectors of a symmetric real matrix corresponding to distinct eigenvalues are orthogonal. In the complex context, two n-tuples $\mathbf{z}$ and $\mathbf{w}$ in $\mathbb{C}^n$ are said to be *orthogonal* if $\langle \mathbf{z}, \mathbf{w} \rangle = 0$.

Theorem 146. Suppose A is a Hermitian matrix.

1. The eigenvalues of A are real.

2. Eigenvectors of A corresponding to distinct eigenvalues are orthogonal.

Proof Let λ and μ be eigenvalues of A with (nonzero) eigenvectors $\mathbf{z}$ and $\mathbf{w}$. Then $A\mathbf{z} = \lambda\mathbf{z}$ and $A\mathbf{w} = \mu\mathbf{w}$, so Theorem **??** gives

$$\lambda\langle \mathbf{z}, \mathbf{w}\rangle = \langle \lambda\mathbf{z}, \mathbf{w}\rangle = \langle A\mathbf{z}, \mathbf{w}\rangle = \langle \mathbf{z}, A\mathbf{w}\rangle = \langle \mathbf{z}, \mu\mathbf{w}\rangle = \overline{\mu}\langle \mathbf{z}, \mathbf{w}\rangle \tag{129}$$

If $\mu = \lambda$ and $\mathbf{w} = \mathbf{z}$, this becomes $\lambda\langle \mathbf{z}, \mathbf{z}\rangle = \overline{\lambda}\langle \mathbf{z}, \mathbf{z}\rangle$. Because $\langle \mathbf{z}, \mathbf{z}\rangle = \|\mathbf{z}\|^2 \neq 0$, this implies $\lambda = \overline{\lambda}$. Thus λ is real, proving (1). Similarly, μ is real, so equation (**??**) gives $\lambda\langle \mathbf{z}, \mathbf{w}\rangle = \mu\langle \mathbf{z}, \mathbf{w}\rangle$. If $\lambda \neq \mu$, this implies $\langle \mathbf{z}, \mathbf{w}\rangle = 0$, proving (2). ■

Corollary 16. Let A be a real symmetric matrix. Then the eigenvalues of A are real.

Proof Symmetric real matrices are Hermitian, and so the result follows immediately from Theorem **??**. ■

The Real Spectral Theorem (**??**) asserts that every real symmetric matrix A is orthogonally diagonalizable—that is U^TAU is diagonal where U is an orthogonal matrix ($U^{-1} = U^T$). The next theorem identifies the complex analogs of these orthogonal real matrices.

Definition 111. As in the real case, a set of nonzero vectors $\{\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_m\}$ in $\mathbb{C}^n$ is called *orthogonal* if $\langle \mathbf{z}_i, \mathbf{z}_j\rangle = 0$ whenever $i \neq j$, and it is *orthonormal* if, in addition, $\|\mathbf{z}_i\| = 1$ for each i.

Theorem 147. The following are equivalent for an $n \times n$ complex matrix A.

1. A is invertible and $A^{-1} = A^H$.

2. The rows of A are an orthonormal set in $\mathbb{C}^n$.

3. The columns of A are an orthonormal set in $\mathbb{C}^n$.

Proof If $A = \begin{bmatrix} | & | & & | \\ \mathbf{c}_1 & \mathbf{c}_2 & \cdots & \mathbf{c}_n \\ | & | & & | \end{bmatrix}$ is a complex matrix with jth column $\mathbf{c}_j$, then $A^T\overline{A} = \left[\langle \mathbf{c}_i, \mathbf{c}_j\rangle\right]$, as in Theorem **??**. Now **??** $\Leftrightarrow$ **??** follows, and **??** $\Leftrightarrow$ **??** is proved in the same way. ■

Definition 112. A square complex matrix U is called *unitary* if $U^{-1} = U^H$.

Thus a real matrix is unitary if and only if it is orthogonal.

Example 247. The matrix $A = \begin{bmatrix} 1+i & 1 \\ 1-i & i \end{bmatrix}$ has orthogonal columns, but the rows are not orthogonal. Normalizing the columns gives the unitary matrix $\frac{1}{2}\begin{bmatrix} 1+i & \sqrt{2} \\ 1-i & \sqrt{2}i \end{bmatrix}$.

Given a real symmetric matrix A, we saw in Orthogonal Matrices and Symmetric Matrices a procedure for finding an orthogonal matrix Q such that Q^TAQ is diagonal (see Example **??**). The following example illustrates Theorem **??** and shows that the technique works for complex matrices.

Example 248. Consider the Hermitian matrix $A = \begin{bmatrix} 3 & 2+i \\ 2-i & 7 \end{bmatrix}$. Find the eigenvalues of A, find two orthonormal eigenvectors, and so find a unitary matrix U such that U^HAU is diagonal.

Explanation. The characteristic polynomial of A is

$$c_A(z) = \det(zI - A) = \det \begin{bmatrix} x-3 & -2-i \\ -2+i & x-7 \end{bmatrix} = (z-2)(z-8)$$

Hence the eigenvalues are 2 and 8 (both real as expected), and corresponding eigenvectors are $\begin{bmatrix} 2+i \\ -1 \end{bmatrix}$ and $\begin{bmatrix} 1 \\ 2-i \end{bmatrix}$ (orthogonal as expected). Each has length $\sqrt{6}$, so let $U = \frac{1}{\sqrt{6}}\begin{bmatrix} 2+i & 1 \\ -1 & 2-i \end{bmatrix}$ be the unitary matrix with the normalized eigenvectors as columns.

Then $U^HAU = \begin{bmatrix} 2 & 0 \\ 0 & 8 \end{bmatrix}$ is diagonal.

Unitary Diagonalization

An $n \times n$ complex matrix A is called *unitarily diagonalizable* if U^HAU is diagonal for some unitary matrix U. As Example **??** suggests, we are going to prove that every Hermitian matrix is unitarily diagonalizable. However, with only a little extra effort, we can get a very important theorem that has this result as an easy consequence.

A complex matrix is called *upper triangular* if every entry below the main diagonal is zero. We owe the following theorem to Issai Schur.

Remark 28. Issai Schur (1875–1941) was a German mathematician who did fundamental work in the theory of representations of groups as matrices.

Theorem 148 (Schur's Theorem)**.** If A is any $n \times n$ complex matrix, there exists a unitary matrix U such that

$$U^H A U = T$$

is upper triangular. Moreover, the entries on the main diagonal of T are the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ of A (including multiplicities).

Proof We use induction on n, mirroring the form of the proof of **??**. If $n = 1$, A is already upper triangular. If $n > 1$, assume the theorem is valid for $(n-1) \times (n-1)$ complex matrices. Let λ_1 be an eigenvalue of A, and let $\mathbf{y}_1$ be an eigenvector with $\|\mathbf{y}_1\| = 1$. Next, the (complex analog of the) Gram-Schmidt process provides $\mathbf{y}_2, \ldots, \mathbf{y}_n$ such that $\{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n\}$ is an orthonormal basis of $\mathbb{C}^n$. If $U_1 = \begin{bmatrix} | & | & & | \\ \mathbf{y}_1 & \mathbf{y}_2 & \cdots & \mathbf{y}_n \\ | & | & & | \end{bmatrix}$ is the matrix with these vectors as its columns, then

$$U_1^H A U_1 = \begin{bmatrix} \lambda_1 & X_1 \\ 0 & A_1 \end{bmatrix}$$

in block form. Now apply induction to find a unitary $(n-1) \times (n-1)$ matrix W_1 such that $W_1^H A_1 W_1 = T_1$ is upper triangular. Then $U_2 = \begin{bmatrix} 1 & 0 \\ 0 & W_1 \end{bmatrix}$ is a unitary $n \times n$ matrix. Hence $U = U_1 U_2$ is unitary (using Theorem **??**), and

$$\begin{aligned} U^H A U &= U_2^H (U_1^H A U_1) U_2 \\ &= \begin{bmatrix} 1 & 0 \\ 0 & W_1^H \end{bmatrix} \begin{bmatrix} \lambda_1 & X_1 \\ 0 & A_1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & W_1 \end{bmatrix} = \begin{bmatrix} \lambda_1 & X_1 W_1 \\ 0 & T_1 \end{bmatrix} \end{aligned}$$

is upper triangular. Finally, A and $U^H A U = T$ have the same eigenvalues by (the complex version of) Theorem **??**, and they are the diagonal entries of T because T is upper triangular. ■

The fact that similar matrices have the same traces and determinants gives the following consequence of Schur's theorem.

Corollary 17. Let A be an $n \times n$ complex matrix, and let $\lambda_1, \lambda_2, \ldots, \lambda_n$ denote the eigenvalues of A, including multiplicities. Then

$$\det A = \lambda_1 \lambda_2 \cdots \lambda_n \quad \text{and} \quad \operatorname{tr} A = \lambda_1 + \lambda_2 + \cdots + \lambda_n$$

Schur's theorem asserts that every complex matrix can be "unitarily triangularized." However, we cannot substitute "unitarily diagonalized" here. In fact, if $A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$, there is no invertible complex matrix U at all such that $U^{-1}AU$ is diagonal. However, the situation is much better for Hermitian matrices.

Theorem 149. If A is Hermitian, there is a unitary matrix U such that U^HAU is diagonal.

Proof By Schur's theorem, let $U^HAU = T$ be upper triangular where U is unitary. Since A is Hermitian, this gives

$$T^H = (U^HAU)^H = U^HA^HU^{HH} = U^HAU = T$$

This means that T is both upper and lower triangular. Hence T is actually diagonal. ■

The Real Spectral Theorem asserts that a real matrix A is symmetric if and only if it is orthogonally diagonalizable (that is, P^TAP is diagonal for some real orthogonal matrix P). Theorem **??** is the complex analog of half of this result. However, the converse is false for complex matrices: There exist unitarily diagonalizable matrices that are not Hermitian.

Example 249. Show that the non-Hermitian matrix $A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$ is unitarily diagonalizable.

Explanation. The characteristic polynomial is $c_A(z) = z^2 + 1$. Hence the eigenvalues are i and $-i$, and it is easy to verify that $\begin{bmatrix} i \\ -1 \end{bmatrix}$ and $\begin{bmatrix} -1 \\ i \end{bmatrix}$ are corresponding eigenvectors. Moreover, these eigenvectors are orthogonal and both have length $\sqrt{2}$, so $U = \frac{1}{\sqrt{2}} \begin{bmatrix} i & -1 \\ -1 & i \end{bmatrix}$ is a unitary matrix such that $U^HAU = \begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix}$ is diagonal.

There is a very simple way to characterize those complex matrices that are unitarily diagonalizable. To this end, an $n \times n$ complex matrix N is called *normal* if $NN^H = N^HN$. It is clear that every Hermitian or unitary matrix is normal, as is the matrix $\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$ in Example **??**. In fact we have the following result.

Theorem 150. An $n \times n$ complex matrix A is unitarily diagonalizable if and only if A is normal.

Proof Assume first that $U^HAU = D$, where U is unitary and D is diagonal. Then $DD^H = D^HD$ as is easily verified. Because $DD^H = U^H(AA^H)U$ and $D^HD = U^H(A^HA)U$, it follows by cancellation that $AA^H = A^HA$.

Conversely, assume A is normal—that is, $AA^H = A^HA$. By Schur's theorem, let $U^HAU = T$, where T is upper triangular and U is unitary. Then T is normal too:

$$TT^H = U^H(AA^H)U = U^H(A^HA)U = T^HT$$

Hence it suffices to show that a normal $n \times n$ upper triangular matrix T must be diagonal. We induct on n; it is clear if $n = 1$. If $n > 1$ and $T = \begin{bmatrix} t_{ij} \end{bmatrix}$, then equating $(1,1)$-entries in TT^H and T^HT gives

$$|t_{11}|^2 + |t_{12}|^2 + ... + |t_{1n}|^2 = |t_{11}|^2$$

This implies $t_{12} = t_{13} = ... = t_{1n} = 0$, so $T = \begin{bmatrix} t_{11} & 0 \\ 0 & T_1 \end{bmatrix}$ in block form. Hence $T = \begin{bmatrix} \overline{t}_{11} & 0 \\ 0 & T_1^H \end{bmatrix}$ so $TT^H = T^HT$ implies $T_1T_1^H = T_1T_1^H$. Thus T_1 is diagonal by induction, and the proof is complete. ■

We conclude this section by using Schur's theorem (Theorem **??**) to prove a famous theorem about matrices. Recall that the characteristic polynomial of a square matrix A is defined by $c_A(z) = \det(zI - A)$, and that the eigenvalues of A are just the roots of $c_A(z)$.

Theorem 151 (Cayley-Hamilton Theorem)**.** If A is an $n \times n$ complex matrix, then $c_A(A) = 0$; that is, A is a root of its characteristic polynomial.

Remark 29. Named after the English mathematician Arthur Cayley (1821–1895) and William Rowan Hamilton (1805–1865), an Irish mathematician famous for his work on physical dynamics.

Proof If $p(z)$ is any polynomial with complex coefficients, then $p(P^{-1}AP) = P^{-1}p(A)P$ for any invertible complex matrix P. Hence, by Schur's theorem, we may assume that A is upper triangular. Then the eigenvalues $\lambda_1, \lambda_2, ..., \lambda_n$ of A appear along the main diagonal, so

$$c_A(z) = (z - \lambda_1)(z - \lambda_2)(z - \lambda_3) \cdots (z - \lambda_n)$$

Thus

$$c_A(A) = (A - \lambda_1 I)(A - \lambda_2 I)(A - \lambda_3 I) \cdots (A - \lambda_n I)$$

Note that each matrix $A - \lambda_i I$ is upper triangular. Now observe:

1. $A - \lambda_1 I$ has zero first column because column 1 of A is $(\lambda_1, 0, 0, ..., 0)^T$.

2. Then $(A - \lambda_1 I)(A - \lambda_2 I)$ has the first two columns zero because the second column of $(A - \lambda_2 I)$ is $(b, 0, 0, ..., 0)^T$ for some constant b.

3. Next $(A - \lambda_1 I)(A - \lambda_2 I)(A - \lambda_3 I)$ has the first three columns zero because column 3 of $(A - \lambda_3 I)$ is $(c, d, 0, ..., 0)^T$ for some constants c and d.

Continuing in this way we see that $(A-\lambda_1 I)(A-\lambda_2 I)(A-\lambda_3 I)\cdots(A-\lambda_n I)$ has all n columns zero; that is, $c_A(A) = 0$. ■

Practice Problems

Problem 652 *In each case, compute the norm of the complex vector.*

1. $(1, 1-i, -2, i)$

... ...

2. $(1-i, 1+i, 1, -1)$

3. $(2+i, 1-i, 2, 0, -i)$

... ...

4. $(-2, -i, 1+i, 1-i, 2i)$

Problem 653 *In each case, determine whether the two vectors are orthogonal.*

1. $\begin{bmatrix} 4 \\ -3i \\ 2+i \end{bmatrix}, \begin{bmatrix} i \\ 2 \\ 2-4i \end{bmatrix}$ *(Yes/ No)*

2. $\begin{bmatrix} i \\ -i \\ 2+i \end{bmatrix}, \begin{bmatrix} i \\ i \\ 2-i \end{bmatrix}$ *(Yes/ No)*

3. $\begin{bmatrix} 1 \\ 1 \\ i \\ i \end{bmatrix}, \begin{bmatrix} 1 \\ i \\ -i \\ 1 \end{bmatrix}$ *(Yes/ No)*

4. $\begin{bmatrix} 4+4i \\ 2+i \\ 2i \end{bmatrix}, \begin{bmatrix} -1+i \\ 2 \\ 3-2i \end{bmatrix}$ *(Yes/ No)*

Problem 654 *A subset U of $\mathbb{C}^n$ is called a complex subspace of $\mathbb{C}^n$ if it contains* 0 *and if, given* $\mathbf{v}$ *and* $\mathbf{w}$ *in U, both* $\mathbf{v}+\mathbf{w}$ *and* $z\mathbf{v}$ *lie in U (z any complex number). In each case, determine whether U is a complex subspace of $\mathbb{C}^3$.*

1. $U = \{(w, \overline{w}, 0) \mid w \text{ in } \mathbb{C}\}$
2. $U = \{(w, 2w, a) \mid w \text{ in } \mathbb{C}, a \text{ in } \mathbb{R}\}$
3. $U = \mathbb{R}^3$ *(Subspace/ Not a Subspace)*
4. $U = \{(v+w, v-2w, v) \mid v, w \text{ in } \mathbb{C}\}$

Problem 655 *In each case, find a basis over $\mathbb{C}$, and determine the dimension of the complex subspace U of $\mathbb{C}^3$ (see the previous exercise).*

1. $U = \{(w, v+w, v-iw) \mid v, w \text{ in } \mathbb{C}\}$
2. $U = \{(iv+w, 0, 2v-w) \mid v, w \text{ in } \mathbb{C}\}$
3. $U = \{(u, v, w) \mid iu - 3v + (1-i)w = 0;$
   $u, v, w \text{ in } \mathbb{C}\}$
4. $U = \{(u, v, w) \mid 2u + (1+i)v - iw = 0;$
   $u, v, w \text{ in } \mathbb{C}\}$

Problem 656 *In each case, determine whether the given matrix is Hermitian, unitary, or normal.*

1. $\begin{bmatrix} 1 & -i \\ i & i \end{bmatrix}$
2. $\begin{bmatrix} 2 & 3 \\ -3 & 2 \end{bmatrix}$
3. $\begin{bmatrix} 1 & i \\ -i & 2 \end{bmatrix}$

4. $\begin{bmatrix} 1 & -i \\ i & -1 \end{bmatrix}$

5. $\frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$

6. $\begin{bmatrix} 1 & 1+i \\ 1+i & i \end{bmatrix}$

7. $\begin{bmatrix} 1+i & 1 \\ -i & -1+i \end{bmatrix}$

8. $\frac{1}{\sqrt{2}|z|} \begin{bmatrix} z & z \\ \overline{z} & -\overline{z} \end{bmatrix}$, $z \neq 0$

Problem 657 *Show that a matrix N is normal if and only if $\overline{N}N^T = N^T\overline{N}$.*

Problem 658 *Let $A = \begin{bmatrix} z & \overline{v} \\ v & w \end{bmatrix}$ where v, w, and z are complex numbers. Characterize in terms of v, w, and z when A is*

1. *Hermitian*
2. *unitary*
3. *normal.*

Problem 659 *In each case, find a unitary matrix U such that $U^H AU$ is diagonal.*

1. $A = \begin{bmatrix} 1 & i \\ -i & 1 \end{bmatrix}$
2. $A = \begin{bmatrix} 4 & 3-i \\ 3+i & 1 \end{bmatrix}$
3. $A = \begin{bmatrix} a & b \\ -b & a \end{bmatrix}$; *a, b, real*

4. $A = \begin{bmatrix} 2 & 1+i \\ 1-i & 3 \end{bmatrix}$

5. $A = \begin{bmatrix} 1 & 0 & 1+i \\ 0 & 2 & 0 \\ 1-i & 0 & 0 \end{bmatrix}$

6. $A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1+i \\ 0 & 1-i & 2 \end{bmatrix}$

Problem 660 *Show that* $\langle A\mathbf{x}, \mathbf{y}\rangle = \langle \mathbf{x}, A^H\mathbf{y}\rangle$ *holds for all* $n \times n$ *matrices* A *and for all* n*-tuples* $\mathbf{x}$ *and* $\mathbf{y}$ *in* $\mathbb{C}^n$.

Problem 661 1. *Prove* **??** *and* **??** *of Theorem* **??**.

2. *Prove Theorem* **??**.

3. *Prove Theorem* **??**.

Problem 662 a. *Show that* A *is Hermitian if and only if* $\overline{A} = A^T$.

b. *Show that the diagonal entries of any Hermitian matrix are real.*

Problem 663 1. *Show that every complex matrix* Z *can be written uniquely in the form* $Z = A + iB$*, where* A *and* B *are real matrices.*

2. *If* $Z = A + iB$ *as in (a), show that* Z *is Hermitian if and only if* A *is symmetric, and* B *is skew-symmetric (that is,* $B^T = -B$*).*

Problem 664 *If* Z *is any complex* $n \times n$ *matrix, show that* ZZ^H *and* $Z + Z^H$ *are Hermitian.*

Problem 665 *A complex matrix* B *is called skew-Hermitian if* $B^H = -B$.

a. *Show that* $Z - Z^H$ *is skew-Hermitian for any square complex matrix* Z.

b. *If* B *is skew-Hermitian, show that* B^2 *and* iB *are Hermitian.*

c. *If* B *is skew-Hermitian, show that the eigenvalues of* B *are pure imaginary (*$i\lambda$ *for real* λ*).*

d. *Show that every $n \times n$ complex matrix Z can be written uniquely as $Z = A + B$, where A is Hermitian and B is skew-Hermitian.*

Problem 666 *Let U be a unitary matrix. Show that:*

1. $\|U\mathbf{x}\| = \|\mathbf{x}\|$ *for all columns* $\mathbf{x}$ *in* $\mathbb{C}^n$.
2. $|\lambda| = 1$ *for every eigenvalue* λ *of* U.

Problem 667 1. *If Z is an invertible complex matrix, show that Z^H is invertible and that $(Z^H)^{-1} = (Z^{-1})^H$.*

2. *Show that the inverse of a unitary matrix is again unitary.*
3. *If U is unitary, show that U^H is unitary.*

Problem 668 *Let Z be an $m \times n$ matrix such that $Z^H Z = I_n$ (for example, Z is a unit column in $\mathbb{C}^n$).*

1. *Show that $V = ZZ^H$ is Hermitian and satisfies $V^2 = V$.*
2. *Show that $U = I - 2ZZ^H$ is both unitary and Hermitian (so $U^{-1} = U^H = U$).*

Problem 669 1. *If N is normal, show that zN is also normal for all complex numbers z.*

2. *Show that (a) fails if normal is replaced by Hermitian.*

Problem 670 *Show that a real 2×2 normal matrix is either symmetric or has the form* $\begin{bmatrix} a & b \\ -b & a \end{bmatrix}$.

Problem 671 *If A is Hermitian, show that all the coefficients of $c_A(z)$ are real numbers.*

Problem 672 1. *If* $A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$, *show that $U^{-1}AU$ is not diagonal for any invertible complex matrix U.*

2. *If* $A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$, *show that $U^{-1}AU$ is not upper triangular for any real invertible matrix U.*

Problem 673 *If A is any $n \times n$ matrix, show that $U^H AU$ is lower triangular for some unitary matrix U.*

Problem 674 *If A is a 3×3 matrix, show that $A^2 = 0$ if and only if there exists a unitary matrix U such that $U^H AU$ has the form $\begin{bmatrix} 0 & 0 & u \\ 0 & 0 & v \\ 0 & 0 & 0 \end{bmatrix}$ or the form $\begin{bmatrix} 0 & u & v \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$.*

Problem 675 *If $A^2 = A$, show that rank A = tr$A$. [Hint: Use Schur's theorem.]*

Text Source

This section was adapted from Section 8.6 of Keith Nicholson's *Linear Algebra with Applications*. (CC-BY-NC-SA)

W. Keith Nicholson, *Linear Algebra with Applications*, Lyryx 2018, Open Edition, pp. 445–456.

99 Inner Product Spaces

We have used the dot product in $\mathbb{R}^n$ to compute the length of vectors (Corollary **??**) and also the angle between vectors. The goal of this section is to define an *inner product* on an arbitrary vector space V over the real numbers. The dot product is an example an inner product for $\mathbb{R}^n$.

Definition 113. An *inner product* on a real vector space V is a function that assigns a real number $\langle \mathbf{v}, \mathbf{w} \rangle$ to every pair $\mathbf{v}$, $\mathbf{w}$ of vectors in V in such a way that the following properties are satisfied.

1. $\langle \mathbf{v}, \mathbf{w} \rangle$ *is a real number for all* $\mathbf{v}$ *and* $\mathbf{w}$ *in* V.
2. $\langle \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{w}, \mathbf{v} \rangle$ *for all* $\mathbf{v}$ *and* $\mathbf{w}$ *in* V.
3. $\langle \mathbf{v} + \mathbf{w}, \mathbf{u} \rangle = \langle \mathbf{v}, \mathbf{u} \rangle + \langle \mathbf{w}, \mathbf{u} \rangle$ *for all* $\mathbf{u}$, $\mathbf{v}$, *and* $\mathbf{w}$ *in* V.
4. $\langle r\mathbf{v}, \mathbf{w} \rangle = r\langle \mathbf{v}, \mathbf{w} \rangle$ *for all* $\mathbf{v}$ *and* $\mathbf{w}$ *in* V *and all* r *in* $\mathbb{R}$.
5. $\langle \mathbf{v}, \mathbf{v} \rangle > 0$ *for all* $\mathbf{v} \neq \mathbf{0}$ *in* V.

A real vector space V with an inner product $\langle\ ,\ \rangle$ will be called an *inner product space*. Note that every subspace of an inner product space is again an inner product space using the same inner product.

Example 250. $\mathbb{R}^n$ is an inner product space with the dot product as inner product:

$$\langle \mathbf{v}, \mathbf{w} \rangle = \mathbf{v} \cdot \mathbf{w} \quad \text{for all } \mathbf{v}, \mathbf{w} \in \mathbb{R}^n$$

See Theorem **??**. This is also called the *euclidean* inner product, and $\mathbb{R}^n$, equipped with the dot product, is called *euclidean n-space*.

Example 251. If A and B are $m \times n$ matrices, define $\langle A, B \rangle = \text{tr}(AB^T)$ where $\text{tr}(X)$ is the trace of the square matrix X. Show that $\langle\ ,\ \rangle$ is an inner product in $\mathbb{M}_{mn}$.

Explanation. Property **??** is clear. Since $\text{tr}(P) = \text{tr}(P^T)$ for every square matrix P, we have Property **??**:

$$\langle A, B \rangle = \text{tr}(AB^T) = \text{tr}[(AB^T)^T] = \text{tr}(BA^T) = \langle B, A \rangle$$

Next, Property **??** and Property **??** follow because trace is a linear transformation $\mathbb{M}_{mn} \to \mathbb{R}$ (Practice Problem **??**). Turning to Property **??**, let $\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_m$ denote the rows of the matrix A. Then the (i, j)-entry of AA^T is $\mathbf{r}_i \cdot \mathbf{r}_j$, so

$$\langle A, A \rangle = \text{tr}(AA^T) = \mathbf{r}_1 \cdot \mathbf{r}_1 + \mathbf{r}_2 \cdot \mathbf{r}_2 + \cdots + \mathbf{r}_m \cdot \mathbf{r}_m$$

But $\mathbf{r}_j \cdot \mathbf{r}_j$ is the sum of the squares of the entries of $\mathbf{r}_j$, so this shows that $\langle A, A \rangle$ is the sum of the squares of all nm entries of A. Property **??** follows.

The next example is important in analysis.

Example 252. Let $\mathcal{F}[a, b]$ be the set of all functions $f : [a, b] \to \mathbb{R}$. Observe that $\mathcal{F}[a, b]$ is a vector space. Let $\mathcal{C}[a, b]$ be a subset of $\mathcal{F}[a, b]$ consisting of all continuous functions. Why is $\mathcal{C}[a, b]$ a subspace of $\mathcal{F}[a, b]$? Show that

$$\langle f, g \rangle = \int_a^b f(x)g(x)dx$$

defines an inner product on $\mathcal{C}[a, b]$.

Remark 30. This example (and others later that refer to it) can be omitted with no loss of continuity by students with no calculus background.

Explanation. Property **??** and Property **??** are clear. As to Property **??**,

$$\langle rf, g \rangle = \int_a^b rf(x)g(x)dx = r \int_a^b f(x)g(x)dx = r\langle f, g \rangle$$

Property **??** is similar. Finally, theorems of calculus show that $\langle f, f \rangle = \int_a^b f(x)^2 dx \geq 0$ and, if f is continuous, that this is zero if and only if f is the zero function. This gives Property **??**.

If $\mathbf{v}$ is any vector, then, using Property **??** of Definition **??**, we get

$$\langle \mathbf{0}, \mathbf{v} \rangle = \langle \mathbf{0} + \mathbf{0}, \mathbf{v} \rangle = \langle \mathbf{0}, \mathbf{v} \rangle + \langle \mathbf{0}, \mathbf{v} \rangle$$

and it follows that the number $\langle \mathbf{0}, \mathbf{v} \rangle$ must be zero. This observation is recorded for reference in the following theorem, along with several other properties of inner products. The other proofs are left as Practice Problem **??**.

Theorem 152. Let $\langle\ ,\ \rangle$ be an inner product on a space V; let $\mathbf{v}$, $\mathbf{u}$, and $\mathbf{w}$ denote vectors in V; and let r denote a real number.

1. $\langle \mathbf{u}, \mathbf{v} + \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{v} \rangle + \langle \mathbf{u}, \mathbf{w} \rangle$
2. $\langle \mathbf{v}, r\mathbf{w} \rangle = r\langle \mathbf{v}, \mathbf{w} \rangle = \langle r\mathbf{v}, \mathbf{w} \rangle$
3. $\langle \mathbf{v}, \mathbf{0} \rangle = 0 = \langle \mathbf{0}, \mathbf{v} \rangle$
4. $\langle \mathbf{v}, \mathbf{v} \rangle = 0$ if and only if $\mathbf{v} = \mathbf{0}$

Remark 31. If r is a complex number, then **??** must be modified. See Theorem **??** in Complex Matrices.

If $\langle\ ,\ \rangle$ is an inner product on a space V, then, given $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ in V,

$$\langle r\mathbf{u} + s\mathbf{v}, \mathbf{w}\rangle = \langle r\mathbf{u}, \mathbf{w}\rangle + \langle s\mathbf{v}, \mathbf{w}\rangle = r\langle \mathbf{u}, \mathbf{w}\rangle + s\langle \mathbf{v}, \mathbf{w}\rangle$$

for all r and s in $\mathbb{R}$ by Property **??** and Property **??** of Definition **??**. Moreover, there is nothing special about the fact that there are two terms in the linear combination or that it is in the first component:

$$\langle r_1\mathbf{v}_1 + r_2\mathbf{v}_2 + \cdots + r_n\mathbf{v}_n, \mathbf{w}\rangle = r_1\langle \mathbf{v}_1, \mathbf{w}\rangle + r_2\langle \mathbf{v}_2, \mathbf{w}\rangle + \cdots + r_n\langle \mathbf{v}_n, \mathbf{w}\rangle$$

and

$$\langle \mathbf{v}, s_1\mathbf{w}_1 + s_2\mathbf{w}_2 + \cdots + s_m\mathbf{w}_m\rangle = s_1\langle \mathbf{v}, \mathbf{w}_1\rangle + s_2\langle \mathbf{v}, \mathbf{w}_2\rangle + \cdots + s_m\langle \mathbf{v}, \mathbf{w}_m\rangle$$

hold for all r_i and s_i in $\mathbb{R}$ and all $\mathbf{v}$, $\mathbf{w}$, $\mathbf{v}_i$, and $\mathbf{w}_j$ in V. These results are described by saying that inner products "preserve" linear combinations. For example,

$$\begin{aligned}\langle 2\mathbf{u} - \mathbf{v}, 3\mathbf{u} + 2\mathbf{v}\rangle &= \langle 2\mathbf{u}, 3\mathbf{u}\rangle + \langle 2\mathbf{u}, 2\mathbf{v}\rangle + \langle -\mathbf{v}, 3\mathbf{u}\rangle + \langle -\mathbf{v}, 2\mathbf{v}\rangle \\ &= 6\langle \mathbf{u}, \mathbf{u}\rangle + 4\langle \mathbf{u}, \mathbf{v}\rangle - 3\langle \mathbf{v}, \mathbf{u}\rangle - 2\langle \mathbf{v}, \mathbf{v}\rangle \\ &= 6\langle \mathbf{u}, \mathbf{u}\rangle + \langle \mathbf{u}, \mathbf{v}\rangle - 2\langle \mathbf{v}, \mathbf{v}\rangle\end{aligned}$$

If A is a symmetric $n \times n$ matrix and $\mathbf{x}$ and $\mathbf{y}$ are columns in $\mathbb{R}^n$, we regard the 1×1 matrix $\mathbf{x}^T A\mathbf{y}$ as a number. If we write

$$\langle \mathbf{x}, \mathbf{y}\rangle = \mathbf{x}^T A\mathbf{y} \quad \text{for all columns } \mathbf{x}, \mathbf{y} \text{ in } \mathbb{R}^n$$

then Properties **??** -**??** of Definition **??** follow from matrix arithmetic (only Property **??** of Definition **??** requires that A is symmetric). Property **??** of Definition **??** reads

$$\mathbf{x}^T A\mathbf{x} > 0 \quad \text{for all columns } \mathbf{x} \neq \mathbf{0} \text{ in } \mathbb{R}^n$$

and this condition characterizes the positive definite matrices (Theorem **??**). This proves the first assertion in the next theorem.

Theorem 153. If A is any $n \times n$ positive definite matrix, then

$$\langle \mathbf{x}, \mathbf{y}\rangle = \mathbf{x}^T A\mathbf{y} \text{ for all columns } \mathbf{x}, \mathbf{y} \text{ in } \mathbb{R}^n$$

defines an inner product on $\mathbb{R}^n$, and every inner product on $\mathbb{R}^n$ arises in this way.

Proof Given an inner product $\langle\ ,\ \rangle$ on $\mathbb{R}^n$, let $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$ be the standard basis of $\mathbb{R}^n$. If $\mathbf{x} = \sum_{i=1}^{n} x_i\mathbf{e}_i$ and $\mathbf{y} = \sum_{j=1}^{n} y_j\mathbf{e}_j$ are two vectors in $\mathbb{R}^n$, compute $\langle \mathbf{x}, \mathbf{y}\rangle$ by adding the inner product of each term $x_i\mathbf{e}_i$ to each term $y_j\mathbf{e}_j$. The result is a double sum.

$$\langle \mathbf{x}, \mathbf{y}\rangle = \sum_{i=1}^{n}\sum_{j=1}^{n}\langle x_i\mathbf{e}_i, y_j\mathbf{e}_j\rangle = \sum_{i=1}^{n}\sum_{j=1}^{n} x_i\langle \mathbf{e}_i, \mathbf{e}_j\rangle y_j$$

As the reader can verify, this is a matrix product:

$$\langle \mathbf{x}, \mathbf{y} \rangle = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} \langle \mathbf{e}_1, \mathbf{e}_1 \rangle & \langle \mathbf{e}_1, \mathbf{e}_2 \rangle & \cdots & \langle \mathbf{e}_1, \mathbf{e}_n \rangle \\ \langle \mathbf{e}_2, \mathbf{e}_1 \rangle & \langle \mathbf{e}_2, \mathbf{e}_2 \rangle & \cdots & \langle \mathbf{e}_2, \mathbf{e}_n \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle \mathbf{e}_n, \mathbf{e}_1 \rangle & \langle \mathbf{e}_n, \mathbf{e}_2 \rangle & \cdots & \langle \mathbf{e}_n, \mathbf{e}_n \rangle \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

Hence $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^T A \mathbf{y}$, where A is the $n \times n$ matrix whose (i, j)-entry is $\langle \mathbf{e}_i, \mathbf{e}_j \rangle$. The fact that

$$\langle \mathbf{e}_i, \mathbf{e}_j \rangle = \langle \mathbf{e}_j, \mathbf{e}_i \rangle$$

shows that A is symmetric. Finally, A is positive definite by Theorem **??**. ■

Thus, just as every linear operator $\mathbb{R}^n \to \mathbb{R}^n$ corresponds to an $n \times n$ matrix, every inner product on $\mathbb{R}^n$ corresponds to a positive definite $n \times n$ matrix. In particular, the dot product corresponds to the identity matrix I_n.

Remark 32. If we refer to the inner product space $\mathbb{R}^n$ without specifying the inner product, we mean that the dot product is to be used.

Example 253. Let the inner product $\langle\ ,\ \rangle$ be defined on $\mathbb{R}^2$ by

$$\left\langle \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}, \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} \right\rangle = 2v_1w_1 - v_1w_2 - v_2w_1 + v_2w_2$$

Find a symmetric 2×2 matrix A such that $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^T A \mathbf{y}$ for all $\mathbf{x}$, $\mathbf{y}$ in $\mathbb{R}^2$.

Explanation. The (i, j)-entry of the matrix A is the coefficient of $v_i w_j$ in the expression, so $A = \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix}$. Incidentally, if $\mathbf{x} = \begin{bmatrix} x \\ y \end{bmatrix}$, then

$$\langle \mathbf{x}, \mathbf{x} \rangle = 2x^2 - 2xy + y^2 = x^2 + (x - y)^2 \geq 0$$

for all $\mathbf{x}$, so $\langle \mathbf{x}, \mathbf{x} \rangle = 0$ implies $\mathbf{x} = \mathbf{0}$. Hence $\langle\ ,\ \rangle$ is indeed an inner product, so A is positive definite.

Let $\langle\ ,\ \rangle$ be an inner product on $\mathbb{R}^n$ given as in Theorem **??** by a positive definite matrix A. If $\mathbf{x} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}^T$, then $\langle \mathbf{x}, \mathbf{x} \rangle = \mathbf{x}^T A \mathbf{x}$ is an expression in the variables $x_1, x_2, \ldots, x_n$ called a *quadratic form*. For more on quadratic forms, see Section 8.8 of [Nicholson], pp. 472–482.

Norm and Distance

Definition 114. As in $\mathbb{R}^n$, if $\langle\ ,\ \rangle$ is an inner product on a space V, the *norm* $\|\mathbf{v}\|$ of a vector $\mathbf{v}$ in V is defined by

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

We define the *distance* between vectors $\mathbf{v}$ and $\mathbf{w}$ in an inner product space V to be

$$\mathrm{d}(\mathbf{v}, \mathbf{w}) = \|\mathbf{v} - \mathbf{w}\|$$

Remark 33. If the dot product is used in $\mathbb{R}^n$, the norm $\|\mathbf{x}\|$ of a vector $\mathbf{x}$ is usually called the *length* of $\mathbf{x}$.

Note that Property **??** of Definition **??** guarantees that $\langle \mathbf{v}, \mathbf{v} \rangle \geq 0$, so $\|\mathbf{v}\|$ is a real number.

Example 254. The norm of a continuous function $f = f(x)$ in $\mathcal{C}[a, b]$ (with the inner product from Example **??**) is given by

$$\|f\| = \sqrt{\int_a^b f(x)^2 dx}$$

Hence $\|f\|^2$ is the area beneath the graph of $y = f(x)^2$ between $x = a$ and $x = b$.

Example 255. Show that $\langle \mathbf{u} + \mathbf{v}, \mathbf{u} - \mathbf{v} \rangle = \|\mathbf{u}\|^2 - \|\mathbf{v}\|^2$ in any inner product space.

Explanation.

$$\begin{aligned} \langle \mathbf{u} + \mathbf{v}, \mathbf{u} - \mathbf{v} \rangle &= \langle \mathbf{u}, \mathbf{u} \rangle - \langle \mathbf{u}, \mathbf{v} \rangle + \langle \mathbf{v}, \mathbf{u} \rangle - \langle \mathbf{v}, \mathbf{v} \rangle \\ &= \|\mathbf{u}\|^2 - \langle \mathbf{u}, \mathbf{v} \rangle + \langle \mathbf{u}, \mathbf{v} \rangle - \|\mathbf{v}\|^2 \\ &= \|\mathbf{u}\|^2 - \|\mathbf{v}\|^2 \end{aligned}$$

A vector $\mathbf{v}$ in an inner product space V is called a *unit vector* if $\|\mathbf{v}\| = 1$. The set of all unit vectors in V is called the *unit ball* in V. For example, if $V = \mathbb{R}^2$ (with the dot product) and $\mathbf{v} = (x, y)$, then

$$\|\mathbf{v}\|^2 = 1 \quad \text{if and only if} \quad x^2 + y^2 = 1$$

Hence the unit ball in $\mathbb{R}^2$ is the *unit circle* $x^2 + y^2 = 1$ with centre at the origin and radius 1. However, the shape of the unit ball varies with the choice of inner product.

Example 256. Let $a > 0$ and $b > 0$. If $\mathbf{v} = (x, y)$ and $\mathbf{w} = (x_1, y_1)$, define an inner product on $\mathbb{R}^2$ by

$$\langle \mathbf{v}, \mathbf{w} \rangle = \frac{x x_1}{a^2} + \frac{y y_1}{b^2}$$

The reader can verify (Practice Problem **??**) that this is indeed an inner product. In this case

$$\|\mathbf{v}\|^2 = 1 \quad \text{if and only if} \quad \frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$$

so the unit ball is the ellipse shown in the diagram.

Example **??** graphically illustrates the fact that norms and distances in an inner product space V vary with the choice of inner product in V.

Theorem 154. If $\mathbf{v} \neq \mathbf{0}$ is any vector in an inner product space V, then $\dfrac{1}{\|\mathbf{v}\|}\mathbf{v}$ is the unique unit vector that is a positive multiple of $\mathbf{v}$.

The next theorem reveals an important and useful fact about the relationship between norms and inner products, extending the Cauchy inequality for $\mathbb{R}^n$ (Theorem **??**).

Theorem 155 (Cauchy-Schwarz Inequality)**.** If $\mathbf{v}$ and $\mathbf{w}$ are two vectors in an inner product space V, then

$$\langle \mathbf{v}, \mathbf{w} \rangle^2 \leq \|\mathbf{v}\|^2 \|\mathbf{w}\|^2$$

Moreover, equality occurs if and only if one of $\mathbf{v}$ and $\mathbf{w}$ is a scalar multiple of the other.

Remark 34. Hermann Amandus Schwarz (1843–1921) was a German mathematician at the University of Berlin. He had strong geometric intuition, which he applied with great ingenuity to particular problems. A version of the inequality appeared in 1885.

Proof of Cauchy-Schwarz Inequality Write $\|\mathbf{v}\| = a$ and $\|\mathbf{w}\| = b$. Using Theorem **??** we compute:

$$\begin{aligned} \|b\mathbf{v} - a\mathbf{w}\|^2 &= b^2\|\mathbf{v}\|^2 - 2ab\langle \mathbf{v}, \mathbf{w} \rangle + a^2\|\mathbf{w}\|^2 = 2ab(ab - \langle \mathbf{v}, \mathbf{w} \rangle) \\ \|b\mathbf{v} + a\mathbf{w}\|^2 &= b^2\|\mathbf{v}\|^2 + 2ab\langle \mathbf{v}, \mathbf{w} \rangle + a^2\|\mathbf{w}\|^2 = 2ab(ab + \langle \mathbf{v}, \mathbf{w} \rangle) \end{aligned} \tag{130}$$

It follows that $ab - \langle \mathbf{v}, \mathbf{w} \rangle \geq 0$ and $ab + \langle \mathbf{v}, \mathbf{w} \rangle \geq 0$, and hence that $-ab \leq \langle \mathbf{v}, \mathbf{w} \rangle \leq ab$. But then $|\langle \mathbf{v}, \mathbf{w} \rangle| \leq ab = \|\mathbf{v}\| \|\mathbf{w}\|$, as desired.

Conversely, if $|\langle \mathbf{v}, \mathbf{w} \rangle| = \|\mathbf{v}\| \|\mathbf{w}\| = ab$ then $\langle \mathbf{v}, \mathbf{w} \rangle = \pm ab$. Hence (**??**) shows that $b\mathbf{v} - a\mathbf{w} = \mathbf{0}$ or $b\mathbf{v} + a\mathbf{w} = \mathbf{0}$. It follows that one of $\mathbf{v}$ and $\mathbf{w}$ is a scalar multiple of the other, even if $a = 0$ or $b = 0$. ■

Example 257. If f and g are continuous functions on the interval $[a, b]$, then (see Example **??**)

$$\left(\int_a^b f(x)g(x)dx \right)^2 \leq \int_a^b f(x)^2 dx \int_a^b g(x)^2 dx$$

Another famous inequality, the so-called *triangle inequality* (See Triangle Inequality in the Appendix), also comes from the Cauchy-Schwarz inequality. It is included in the following list of basic properties of the norm of a vector.

Theorem 156. If V is an inner product space, the norm $\|\cdot\|$ has the following properties.

1. $\|\mathbf{v}\| \geq 0$ for every vector $\mathbf{v}$ in V.
2. $\|\mathbf{v}\| = 0$ if and only if $\mathbf{v} = \mathbf{0}$.
3. $\|r\mathbf{v}\| = |r| \, \|\mathbf{v}\|$ for every $\mathbf{v}$ in V and every r in $\mathbb{R}$.
4. $\|\mathbf{v} + \mathbf{w}\| \leq \|\mathbf{v}\| + \|\mathbf{w}\|$ for all $\mathbf{v}$ and $\mathbf{w}$ in V (*triangle inequality*).

Proof Because $\|\mathbf{v}\| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}$, properties **??** and **??** follow immediately from **??** and **??** of Theorem **??**. As to **??**, compute

$$\|r\mathbf{v}\|^2 = \langle r\mathbf{v}, r\mathbf{v} \rangle = r^2 \langle \mathbf{v}, \mathbf{v} \rangle = r^2 \|\mathbf{v}\|^2$$

Hence **??** follows by taking positive square roots. Finally, the fact that $\langle \mathbf{v}, \mathbf{w} \rangle \leq \|\mathbf{v}\| \, \|\mathbf{w}\|$ by the Cauchy-Schwarz inequality gives

$$\begin{aligned} \|\mathbf{v} + \mathbf{w}\|^2 = \langle \mathbf{v} + \mathbf{w}, \mathbf{v} + \mathbf{w} \rangle &= \|\mathbf{v}\|^2 + 2\langle \mathbf{v}, \mathbf{w} \rangle + \|\mathbf{w}\|^2 \\ &\leq \|\mathbf{v}\|^2 + 2 \|\mathbf{v}\| \, \|\mathbf{w}\| + \|\mathbf{w}\|^2 \\ &= (\|\mathbf{v}\| + \|\mathbf{w}\|)^2 \end{aligned}$$

Hence **??** follows by taking positive square roots. ■

It is worth noting that the usual triangle inequality for absolute values,

$$|r + s| \leq |r| + |s| \text{ for all real numbers } r \text{ and } s$$

is a special case of **??** where $V = \mathbb{R} = \mathbb{R}^1$ and the dot product $\langle r, s \rangle = rs$ is used.

In many calculations in an inner product space, it is required to show that some vector $\mathbf{v}$ is zero. This is often accomplished most easily by showing that its norm $\|\mathbf{v}\|$ is zero. Here is an example.

Example 258. Let $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ be a spanning set for an inner product space V. If $\mathbf{v}$ in V satisfies $\langle \mathbf{v}, \mathbf{v}_i \rangle = 0$ for each $i = 1, 2, \ldots, n$, show that $\mathbf{v} = \mathbf{0}$.

Explanation. Write $\mathbf{v} = r_1\mathbf{v}_1 + \cdots + r_n\mathbf{v}_n$, r_i in $\mathbb{R}$. To show that $\mathbf{v} = \mathbf{0}$, we show that $\|\mathbf{v}\|^2 = \langle \mathbf{v}, \mathbf{v} \rangle = 0$. Compute:

$$\langle \mathbf{v}, \mathbf{v} \rangle = \langle \mathbf{v}, r_1\mathbf{v}_1 + \cdots + r_n\mathbf{v}_n \rangle = r_1 \langle \mathbf{v}, \mathbf{v}_1 \rangle + \cdots + r_n \langle \mathbf{v}, \mathbf{v}_n \rangle = 0$$

by hypothesis, and the result follows.

The norm properties in Theorem **??** translate to the following properties of distance familiar from geometry. The proof is Practice Problem **??**.

Theorem 157. Let V be an inner product space.

1. $d(\mathbf{v}, \mathbf{w}) \geq 0$ for all $\mathbf{v}$, $\mathbf{w}$ in V.
2. $d(\mathbf{v}, \mathbf{w}) = 0$ if and only if $\mathbf{v} = \mathbf{w}$.
3. $d(\mathbf{v}, \mathbf{w}) = d(\mathbf{w}, \mathbf{v})$ for all $\mathbf{v}$ and $\mathbf{w}$ in V.
4. $d(\mathbf{v}, \mathbf{w}) \leq d(\mathbf{v}, \mathbf{u}) + d(\mathbf{u}, \mathbf{w})$ for all $\mathbf{v}$, $\mathbf{u}$, and $\mathbf{w}$ in V.

Practice Problems

Problem 676 *In each case, determine which of Property* **??***–Property* **??** *in Definition* **??** *fail to hold.*

1. $V = \mathbb{R}^2$, $\left\langle \begin{bmatrix} x_1 \\ y_1 \end{bmatrix}, \begin{bmatrix} x_2 \\ y_2 \end{bmatrix} \right\rangle = x_1 y_1 x_2 y_2$
2. $V = \mathbb{R}^3$, $\left\langle \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} \right\rangle = x_1 y_1 - x_2 y_2 + x_3 y_3$

   Click the arrow to see the answer. Property **??** *fails.*
3. $V = \mathbb{C}$, $\langle z, w \rangle = z\overline{w}$, *where* $\overline{w}$ *is complex conjugation*

   Click the arrow to see the answer. Property **??** *fails, as sometimes we get a complex number. However, we will return to this definition of* $\langle\ ,\ \rangle$ *in Complex Matrices – see Definition* **??**
4. $V = \mathbb{P}^3$, $\langle p(x), q(x) \rangle = p(1)q(1)$

   Click the arrow to see the answer. Property **??** *fails.*
5. $V = \mathbb{M}_{22}$, $\langle A, B \rangle = det(AB)$
6. $V = \mathcal{F}[0, 1]$, $\langle f, g \rangle = f(1)g(0) + f(0)g(1)$

   Click the arrow to see the answer. Property **??** *fails.*

Problem 677 *Let V be an inner product space. If $U \subseteq V$ is a subspace, show that U is an inner product space using the same inner product.*

Problem 678 *In each case, find a scalar multiple of $\mathbf{v}$ that is a unit vector.*

1. $\mathbf{v} = f$ *in* $\mathcal{C}[0,1]$ *where* $f(x) = x^2$
   $\langle f, g \rangle \int_0^1 f(x)g(x)dx$
2. $\mathbf{v} = f$ *in* $\mathcal{C}[-\pi, \pi]$ *where* $f(x) = \cos x$
   $\langle f, g \rangle \int_{-\pi}^{\pi} f(x)g(x)dx$

   Click the arrow to see the answer. $\dfrac{1}{\sqrt{\pi}} f$
3. $\mathbf{v} = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$ *in* $\mathbb{R}^2$ *where* $\langle \mathbf{v}, \mathbf{w} \rangle = \mathbf{v}^T \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} \mathbf{w}$
4. $\mathbf{v} = \begin{bmatrix} 3 \\ -1 \end{bmatrix}$ *in* $\mathbb{R}^2$, $\langle \mathbf{v}, \mathbf{w} \rangle = \mathbf{v}^T \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} \mathbf{w}$

   Click the arrow to see the answer. $\dfrac{1}{\sqrt{17}} \begin{bmatrix} 3 \\ -1 \end{bmatrix}$

Problem 679 *In each case, find the distance between $\mathbf{u}$ and $\mathbf{v}$.*

1. $\mathbf{u} = \begin{bmatrix} 3 \\ -1 \\ 2 \\ 0 \end{bmatrix}$, $\mathbf{v} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 3 \end{bmatrix}$; $\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{u} \cdot \mathbf{v}$
2. $\mathbf{u} = \begin{bmatrix} 1 \\ 2 \\ -1 \\ 2 \end{bmatrix}$, $\mathbf{v} = \begin{bmatrix} 2 \\ 1 \\ -1 \\ 3 \end{bmatrix}$; $\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{u} \cdot \mathbf{v}$

   Click the arrow to see the answer. $\sqrt{3}$
3. $\mathbf{u} = f$, $\mathbf{v} = g$ *in* $\mathcal{C}[0,1]$ *where* $f(x) = x^2$ *and* $g(x) = 1 - x$; $\langle f, g \rangle = \int_0^1 f(x)g(x)dx$

4. $\mathbf{u} = f$, $\mathbf{v} = g$ *in* $\mathcal{C}[-\pi, \pi]$ *where* $f(x) = 1$ *and* $g(x) = \cos x$; $\langle f, g \rangle = \displaystyle\int_{-\pi}^{\pi} f(x)g(x)dx$

   Click the arrow to see the answer. $\sqrt{3\pi}$

Problem 680 *Let* $a_1, a_2, \ldots, a_n$ *be positive numbers. Given* $\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$ *and* $\mathbf{w} = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix}$, *define* $\langle \mathbf{v}, \mathbf{w} \rangle = a_1 v_1 w_1 + \cdots + a_n v_n w_n$. *Show that this is an inner product on* $\mathbb{R}^n$.

Problem 681 *If* $\{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ *is a basis of* V *and if* $\mathbf{v} = v_1\mathbf{b}_1 + \cdots + v_n\mathbf{b}_n$ *and* $\mathbf{w} = w_1\mathbf{b}_1 + \cdots + w_n\mathbf{b}_n$ *are vectors in* V, *define*

$$\langle \mathbf{v}, \mathbf{w} \rangle = v_1 w_1 + \cdots + v_n w_n.$$

Show that this is an inner product on V.

Problem 682 *Let* $re(z)$ *denote the real part of the complex number* z. *Show that* $\langle\ ,\ \rangle$ *is an inner product on* $\mathbb{C}$ *if* $\langle \mathbf{z}, \mathbf{w} \rangle = re(z\overline{w})$.

Problem 683 *If* $T : V \to V$ *is an isomorphism of the inner product space* V, *show that*

$$\langle \mathbf{v}, \mathbf{w} \rangle_1 = \langle T(\mathbf{v}), T(\mathbf{w}) \rangle$$

defines a new inner product $\langle\ ,\ \rangle_1$ *on* V.

Problem 684 *Show that every inner product* $\langle\ ,\ \rangle$ *on* $\mathbb{R}^n$ *has the form* $\langle \mathbf{x}, \mathbf{y} \rangle = (U\mathbf{x}) \cdot (U\mathbf{y})$ *for some upper triangular matrix* U *with positive diagonal entries.*

Problem 685 *In each case, show that* $\langle \mathbf{v}, \mathbf{w} \rangle = \mathbf{v}^T A\mathbf{w}$ *defines an inner product on* $\mathbb{R}^2$ *and hence show that* A *is positive definite.*

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

2. $A = \begin{bmatrix} 5 & -3 \\ -3 & 2 \end{bmatrix}$

   Click the arrow to see the answer. $\langle \mathbf{v}, \mathbf{v} \rangle = 5v_1^2 - 6v_1v_2 + 2v_2^2 = \dfrac{1}{5}[(5v_1 - 3v_2)^2 + v_2^2]$

3. $A = \begin{bmatrix} 3 & 2 \\ 2 & 3 \end{bmatrix}$

4. $A = \begin{bmatrix} 3 & 4 \\ 4 & 6 \end{bmatrix}$

Click the arrow to see the answer. $\langle \mathbf{v}, \mathbf{v} \rangle = 3v_1^2 + 8v_1v_2 + 6v_2^2 = \frac{1}{3}[(3v_1 + 4v_2)^2 + 2v_2^2]$

Problem 686 *In each case, find a symmetric matrix A such that* $\langle \mathbf{v}, \mathbf{w} \rangle = \mathbf{v}^T A \mathbf{w}$.

1. $\left\langle \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}, \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} \right\rangle = v_1w_1 + 2v_1w_2 + 2v_2w_1 + 5v_2w_2$

2. $\left\langle \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}, \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} \right\rangle = v_1w_1 - v_1w_2 - v_2w_1 + 2v_2w_2$

   Click the arrow to see the answer. $\begin{bmatrix} 1 & -2 \\ -2 & 1 \end{bmatrix}$

3. $\left\langle \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}, \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} \right\rangle = 2v_1w_1 + v_2w_2 + v_3w_3 - v_1w_2$
   $- v_2w_1 + v_2w_3 + v_3w_2$

4. $\left\langle \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}, \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} \right\rangle = v_1w_1 + 2v_2w_2 + 5v_3w_3$
   $- 2v_1w_3 - 2v_3w_1$

   Click the arrow to see the answer. $\begin{bmatrix} 1 & 0 & -2 \\ 0 & 2 & 0 \\ -2 & 0 & 5 \end{bmatrix}$

Problem 687 *If A is symmetric and* $\mathbf{x}^T A \mathbf{x} = 0$ *for all columns* $\mathbf{x}$ *in* $\mathbb{R}^n$*, show that* $A = 0$.

Click the arrow to see the answer. By the condition, $\langle \mathbf{x}, \mathbf{y} \rangle = \frac{1}{2}\langle \mathbf{x} + \mathbf{y}, \mathbf{x} + \mathbf{y} \rangle = 0$ *for all* $\mathbf{x}$, $\mathbf{y}$. *Let* $\mathbf{e}_i$ *denote column i of I. If* $A = \begin{bmatrix} a_{ij} \end{bmatrix}$, *then* $a_{ij} = \mathbf{e}_i^T A \mathbf{e}_j = \{\mathbf{e}_i, \mathbf{e}_j\} = 0$ *for all i and j.*

Problem 688 *Show that the sum of two inner products on V is again an inner product.*

Problem 689 *Let $\|\mathbf{u}\| = 1$, $\|\mathbf{v}\| = 2$, $\|\mathbf{w}\| = \sqrt{3}$, $\langle \mathbf{u}, \mathbf{v} \rangle = -1$, $\langle \mathbf{u}, \mathbf{w} \rangle = 0$ and $\langle \mathbf{v}, \mathbf{w} \rangle = 3$. Compute:*

1. $\langle \mathbf{v} + \mathbf{w}, 2\mathbf{u} - \mathbf{v} \rangle$
2. $\langle \mathbf{u} - 2\mathbf{v} - \mathbf{w}, 3\mathbf{w} - \mathbf{v} \rangle$

   ... ...

Problem 690 *Given the data in Practice Problem* **??**, *show that* $\mathbf{u} + \mathbf{v} = \mathbf{w}$.

Problem 691 *Show that no vectors exist such that $\|\mathbf{u}\| = 1$, $\|\mathbf{v}\| = 2$, and $\langle \mathbf{u}, \mathbf{v} \rangle = -3$.*

Problem 692 *Complete Example* **??**.

Problem 693 *Prove Theorem* **??**.

Problem 694 *Prove Theorem* **??**.

Problem 695 *Let* $\mathbf{u}$ *and* $\mathbf{v}$ *be vectors in an inner product space V.*

1. *Expand* $\langle 2\mathbf{u} - 7\mathbf{v}, 3\mathbf{u} + 5\mathbf{v} \rangle$.
2. *Expand* $\langle 3\mathbf{u} - 4\mathbf{v}, 5\mathbf{u} + \mathbf{v} \rangle$.
   Click the arrow to see the answer. $15\|\mathbf{u}\|^2 - 17\langle \mathbf{u}, \mathbf{v} \rangle - 4\|\mathbf{v}\|^2$
3. *Show that* $\|\mathbf{u} + \mathbf{v}\|^2 = \|\mathbf{u}\|^2 + 2\langle \mathbf{u}, \mathbf{v} \rangle + \|\mathbf{v}\|^2$.
4. *Show that* $\|\mathbf{u} - \mathbf{v}\|^2 = \|\mathbf{u}\|^2 - 2\langle \mathbf{u}, \mathbf{v} \rangle + \|\mathbf{v}\|^2$.
   Click the arrow to see the answer. $\|\mathbf{u} + \mathbf{v}\|^2 = \langle \mathbf{u} + \mathbf{v}, \mathbf{u} + \mathbf{v} \rangle = \|\mathbf{u}\|^2 + 2\langle \mathbf{u}, \mathbf{v} \rangle + \|\mathbf{v}\|^2$

Problem 696 *Show that*

$$\|\mathbf{v}\|^2 + \|\mathbf{w}\|^2 = \frac{1}{2}\{\|\mathbf{v} + \mathbf{w}\|^2 + \|\mathbf{v} - \mathbf{w}\|^2\}$$

for any $\mathbf{v}$ *and* $\mathbf{w}$ *in an inner product space.*

Problem 697 *Let $\langle\ ,\rangle$ be an inner product on a vector space V. Show that the corresponding distance function is translation invariant. That is, show that $d(\mathbf{v},\mathbf{w}) = d(\mathbf{v}+\mathbf{u},\mathbf{w}+\mathbf{u})$ for all $\mathbf{v}$, $\mathbf{w}$, and $\mathbf{u}$ in V.*

Problem 698

1. *Show that $\langle \mathbf{u},\mathbf{v}\rangle = \frac{1}{4}[\|\mathbf{u}+\mathbf{v}\|^2 - \|\mathbf{u}-\mathbf{v}\|^2]$ for all $\mathbf{u}$, $\mathbf{v}$ in an inner product space V.*
2. *If $\langle\ ,\rangle$ and $\langle\ ,\rangle'$ are two inner products on V that have equal associated norm functions, show that $\langle \mathbf{u},\mathbf{v}\rangle = \langle \mathbf{u},\mathbf{v}\rangle'$ holds for all $\mathbf{u}$ and $\mathbf{v}$.*

Problem 699 *Let $\mathbf{v}$ denote a vector in an inner product space V.*

1. *Show that $W = \{\mathbf{w} \mid \mathbf{w} \text{ in } V, \langle \mathbf{v},\mathbf{w} = 0\}$ is a subspace of V.*
2. *Let W be as in (a). If $V = \mathbb{R}^3$ with the dot product, and if $\mathbf{v} = \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}$, find a basis for W.*

   Click the arrow to see the answer. $\left\{ \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix} \right\}$

Problem 700 *Given vectors $\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_n$ and $\mathbf{v}$, assume that $\langle \mathbf{v},\mathbf{w}_i\rangle = 0$ for each i. Show that $\langle \mathbf{v},\mathbf{w}\rangle = 0$ for all $\mathbf{w}$ in span$\{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_n\}$.*

Problem 701 *If $V = \text{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ and $\langle \mathbf{v},\mathbf{v}_i\rangle = \langle \mathbf{w},\mathbf{v}_i\rangle$ holds for each i. Show that $\mathbf{v} = \mathbf{w}$.*

Problem 702 *Use the Cauchy-Schwarz inequality in an inner product space to show that:*

1. *If $\|\mathbf{u}\| \leq 1$, then $\langle \mathbf{u},\mathbf{v}\rangle^2 \leq \|\mathbf{v}\|^2$ for all $\mathbf{v}$ in V.*
2. *$(x\cos\theta + y\sin\theta)^2 \leq x^2 + y^2$ for all real x, y, and θ.*
3. *$\|r_1\mathbf{v}_1 + \cdots + r_n\mathbf{v}_n\|^2 \leq [r_1\|\mathbf{v}_1\| + \cdots + r_n\|\mathbf{v}_n\|]^2$ for all vectors $\mathbf{v}_i$, and all $r_i > 0$ in $\mathbb{R}$.*

Problem 703 *If A is a $2 \times n$ matrix, let $\mathbf{u}$ and $\mathbf{v}$ denote the rows of A.*

1. *Show that $AA^T = \begin{bmatrix} \|\mathbf{u}\|^2 & \mathbf{u}\cdot\mathbf{v} \\ \mathbf{u}\cdot\mathbf{v} & \|\mathbf{v}\|^2 \end{bmatrix}$.*

2. *Show that* $\det(AA^T) \geq 0$.

Problem 704 1. *If* $\mathbf{v}$ *and* $\mathbf{w}$ *are nonzero vectors in an inner product space* V*, show that* $-1 \leq \dfrac{\langle \mathbf{v}, \mathbf{w} \rangle}{\|\mathbf{v}\| \, \|\mathbf{w}\|} \leq 1$*, and hence that a unique angle* θ *exists such that* $\dfrac{\langle \mathbf{v}, \mathbf{w} \rangle}{\|\mathbf{v}\| \, \|\mathbf{w}\|} = \cos\theta$ *and* $0 \leq \theta \leq \pi$*. This angle* θ *is called the angle between* $\mathbf{v}$ *and* $\mathbf{w}$.

2. *Find the angle between* $\mathbf{v} = \begin{bmatrix} 1 \\ 2 \\ -1 \\ 1 \\ 3 \end{bmatrix}$ *and* $\mathbf{w} = \begin{bmatrix} 2 \\ 1 \\ 0 \\ 2 \\ 0 \end{bmatrix}$ *in* $\mathbb{R}^5$ *with the dot product.*

3. *If* θ *is the angle between* $\mathbf{v}$ *and* $\mathbf{w}$*, show that the law of cosines is valid:*

$$\|\mathbf{v} - \mathbf{w}\| = \|\mathbf{v}\|^2 + \|\mathbf{w}\|^2 - 2\|\mathbf{v}\| \, \|\mathbf{w}\| \cos\theta.$$

Problem 705 *If* $V = \mathbb{R}^2$*, define* $\left\| \begin{bmatrix} x \\ y \end{bmatrix} \right\| = |x| + |y|$.

1. *Show that* $\|\cdot\|$ *satisfies the conditions in Theorem* **??**.

2. *Show that* $\|\cdot\|$ *does not arise from an inner product on* $\mathbb{R}^2$ *given by a matrix* A.

Text Source

This section was adapted from Section 10.1 of Keith Nicholson's *Linear Algebra with Applications*. (CC-BY-NC-SA)

W. Keith Nicholson, *Linear Algebra with Applications*, Lyryx 2018, Open Edition, pp. 521–530.