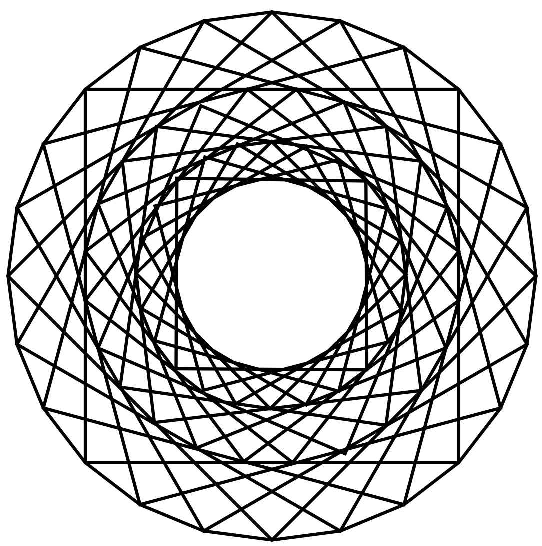
CAMPION COLLEGE
AT THE UNIVERSITY OF REGINA

LUTHER COLLEGE
UNIVERSITY OF REGINA

Math 122
Linear Algebra I



by Robert G. Petry and Fotini Labropulu

published by Campion College and Luther College

History

- 1st Edition produced in 2017 entitled "Linear Algebra I" written by principal authors Robert G. Petry and Fotini Labropulu. Typeset in part by Michael Thatcher. Published by Campion College and Luther College.

The source of this document (i.e. a transparent copy) is available via
`http://campioncollege.ca/contact-us/faculty-listing/dr-robert-petry`

About the cover: The cover line art drawing is in the public domain and available from `http://openclipart.org`.

Contents

Chapter 1: Linear Equations

1.1 Equations

Definition: An **equation** is two mathematical **expressions** joined with an equal sign.

> **Example 1-1**
>
> The equation
> $$y^2 = 1 + x$$
> is composed of the expressions y^2 and $1 + x$.

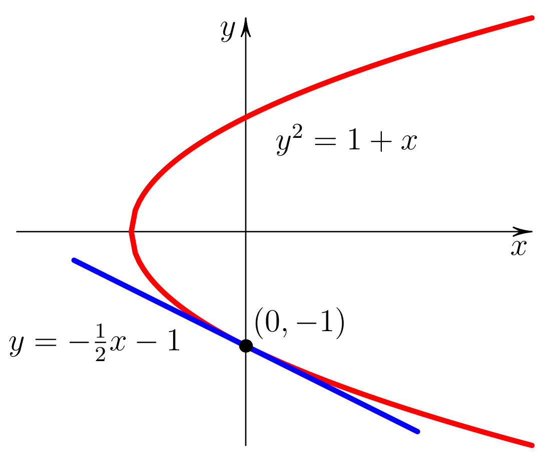
When an equation involves one or more variables (like x and y here) equality will usually hold for only certain values of the variables. These particular values are called the **solutions** of the equation. As an example the pair of values $x = 0$, $y = -1$, or, more simply $(x, y) = (0, -1)$, is one solution to the previous equation. **Solving** an equation is the act of finding the solutions of it. If we plotted points for all the solutions to the previous equation in the Cartesian coordinate system we would see these form a curve (a parabola).

We are perhaps more familiar with an equation that represent a (straight) line, like

$$y = -\frac{1}{2}x - 1$$

Recalling the slope-intercept general form for a line, $y = mx+b$, we see that the constants are $m = -1/2$ and $b = -1$ for the slope and intercept of the line respectively. This latter equation is an example of a **linear equation** which we will define precisely shortly.

If we graphed the curve $y^2 = x + 1$, the point $(0, -1)$, and the line $y = -\frac{1}{2}x - 1$, we would see that the line is actually a **tangent line** to the curve at the point $(0, -1)$ as it just touches the curve at $(0, -1)$ without crossing it. One observes that if we were only interested in the parabola at points very close to the point $(0, -1)$, the straight line determined by the linear equation would approximate the curve quite well.

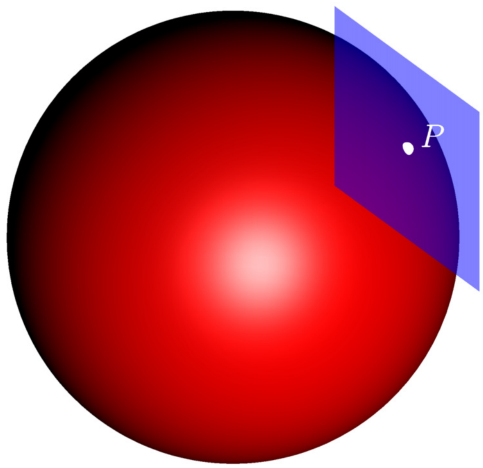


As such, near a point on a curve that might arise from some arbitrary equation involving two variables there is a line which approximates it well and that line is represented by a linear equation.

If we would introduce a further variable (like z) into our original equation the triplets (x, y, z) that satisfy it would, in general, generate a **surface** in three dimensions. For example the points of a sphere of radius $r = 5$ centred on the origin satisfy the equation

$$x^2 + y^2 + z^2 = 25\,.$$

Near one of the points on that surface we could approximate the surface by a **tangent plane**. As an example the spherical Earth locally at a point P is approximately a flat plane.



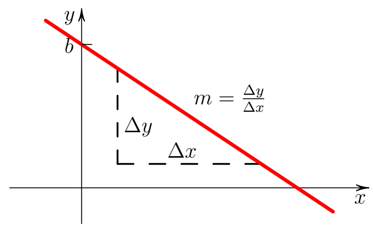
A plane, in turn, we will see can be represented algebraically by the solutions of a linear equation in three variables. As such, an understanding of these linear algebraic structures, and, to the extent we can visualize them, their graphs, will provide us with useful insights and approximations for more general equations.

1.2 Equations of a Line

A straight line in the x-y plane can be characterized in several ways. The following are common forms of a linear equation in two variables. Which of the "recipes" to use depends on what information one has regarding the line.

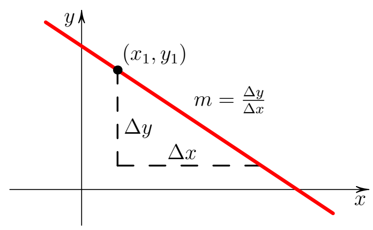
The **slope–intercept form** of a line is used when the y–intercept b is known as well as its direction, characterized by the slope m.[1] It has the familiar form:

$$\boxed{y = mx + b}$$



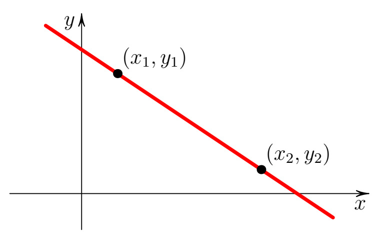
The **point–slope form** of an equation is used in the more general case when one still knows the slope, m, but now an arbitrary point (x_1, y_1) on the line. It is given by

$$\boxed{y - y_1 = m(x - x_1)}$$



In the event two points (x_1, y_1) and (x_2, y_2) on the line are known rather than the slope one can use the **two point form** for a line:

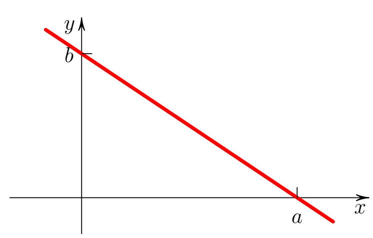
$$\boxed{y - y_1 = \frac{y_2 - y_1}{x_2 - x_1}(x - x_1)}$$



which arises from the point–slope form by noting that $m = \frac{\Delta y}{\Delta x} = \frac{y_2 - y_1}{x_2 - x_1}$. None of the previous three line equations can represent vertical lines.

If one knows the x-intercept a of a line in addition to its y-intercept b the **two intercept** form is just

$$\boxed{\frac{x}{a} + \frac{y}{b} = 1}$$



[1] One prefers the y–intercept b over the x–intercept a (where the line crosses the x–axis), because any line that can be written as a function $y = f(x)$ can always be written in the form $y = mx + b$. The horizontal line $y = 3$, which is a valid function, has no x-intercept.

since clearly when $x = 0$ one has $y = b$ and when $y = 0$ one has $x = a$. This two intercept form cannot, however, represent lines that lack an intercept (either x or y), namely horizontal and vertical lines.

The **standard form** of a line is given by

$$\boxed{a_1x + a_2y = b}$$

where a_1, a_2, and b are constants with a_1 and a_2 not both zero. While geometrically the constants do not have an immediate meaning like the previous equations, the standard form is able to represent all possible lines in the plane.[2] One can always rearrange it into one of the other forms to interpret it geometrically. Alternatively substitute any two values of x and evaluate (solve for) their y-coordinates to establish two points on the line for graphing.

Example 1-2

Find the standard form of the equation of the line going through points $(1, 3)$ and $(-2, 9)$.

Solution:
Since we are given two points, use the two point form with $(x_1, y_1) = (1, 3)$ and $(x_2, y_2) = (-2, 9)$ to get

$$y - 3 = \frac{9-3}{-2-1}(x-1)$$

which simplifies to

$$y - 3 = -2(x-1)$$

Expanding the right hand side to get $y - 3 = -2x + 2$ and isolating the variables on the left gives the standard form

$$2x + y = 5\,.$$

One can check that the points $(1, 3)$ and $(-2, 9)$ satisfy the equation.

Example 1-3

Convert the line with standard form $6x - y = 2$ into two intercept form and sketch the line.

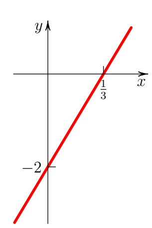
Solution:
The two intercept form $x/a + y/b = 1$ requires a 1 on the right hand side so dividing $6x - y = 2$ by two gives

$$3x - \frac{y}{2} = 1\,.$$

Noting that multiplying by 3 in the first term is the same as dividing by its reciprocal $1/3$ and bringing the -1 into the denominator of the second term gives the two intercept form

$$\frac{x}{1/3} + \frac{y}{-2} = 1\,.$$

with x-intercept $a = 1/3$ and y-intercept $b = -2$. Plotting the points $(1/3, 0)$ and $(0, -2)$ and joining them with the straight line gives the following graph.



[2]It will be shown later that constants a_1 and a_2 can be interpreted in terms of the *normal* direction to the line.

1.3 Systems of Linear Equations

The equation of a line, $a_1x + a_2y = b$ is a linear equation for 2 variables, x and y. More generally we have the following.

Definition: A **linear equation** in n variables x_1, x_2, $\dots$, x_n is an equation that can be written in the form

$$\boxed{a_1x_1 + a_2x_2 + \ldots + a_nx_n = b}$$

where the a_i are the (constant) **coefficients** and b is the **constant term**.

Note that one or more of the constants in a linear equation may be zero.

Definition: A **solution** of the linear equation $a_1x_1 + a_2x_2 + \ldots + a_nx_n = b$ is a sequence of numbers $(t_1, t_2, \ldots, t_n)$ such that the substitution $x_1 = t_1$, $x_2 = t_2$, $\ldots x_n = t_n$ into the equation makes it true.

Example 1-4

1. For $3x_1 + 2x_2 = 5$, we have $x_1 = 3$, $x_2 = -2$ as a solution and $x_1 = -5$ and $x_2 = 10$ as another solution as can be confirmed by direct substitution into the equation.
2. For $2x + 3y - 2z = 2$ we have $x = 2$, $y = 2$, and $z = 4$ as a solution.

Definition: Linear equation $a_1x_1 + a_2x_2 + \ldots + a_nx_n = b$ is called **homogeneous** if $b = 0$ and **non-homogeneous** otherwise.

Example 1-5

1. $2x - 3y = 0$ is homogeneous
2. $-2x_1 + 3x_2 - x_3 = 5$ is non-homogeneous.

Definition: A **system of linear equations** (or **linear system**) is a finite set of two or more linear equations involving the same set of variables called the **unknowns**.

Definition: A **solution** of a linear system involving n unknowns is a sequence of numbers $(t_1, t_2, \ldots, t_n)$ that is simultaneously a solution of every linear equation in the system.

Example 1-6

The equations

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

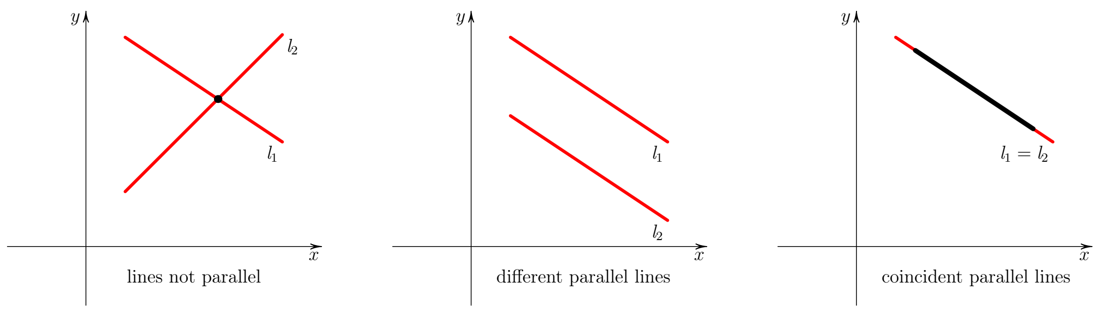
form a linear system of two equations in two unknowns. $x = 1$, $y = -2$ is a solution to the system. Equivalently we may write the solution as $(x, y) = (1, -2)$.

1.3.1 Geometrical Interpretation of the Solution

The solution of a linear system with two equations in two dimensions has a convenient geometrical interpretation. Suppose we have such a system of two equations in two unknowns:

$$\begin{aligned} a_1x + b_1y &= c_1 \\ a_2x + b_2y &= c_2 \end{aligned}$$

Since each linear equation separately represents a line in the plane, a solution x, y of the system must be any point (x, y) that geometrically lies on both lines. Thus a unique solution occurs if the lines are not parallel since they then intersect at a single point. The system has no solution if the lines are parallel (equal slopes) but do not overlap. The system has infinitely many solutions if the two lines are parallel and do overlap (i.e. are coincident).



In Example 1-6 the solution is unique as the lines are not parallel and intersect only at the point $(1, -2)$.

1.3.2 Consistent and Homogeneous Linear Systems

Definition: A linear system which has at least one solution is called **consistent**. This means that the system will have **at least one solution**. A system which has no solution is called **inconsistent**.

Example 1-7

1. The system $\begin{cases} x - 3y = 2 \\ 2x - 6y = 2 \end{cases}$ is inconsistent because there are no solutions to this system. To see this algebraically note the left side of the second equation is double the left side of the first but the right sides are equal. If, for given x and y, the left side is zero then this is not a solution. If, for given x and y, the left sides are non-zero then both must be so with one being double the other which contradicts them being equal as required by the right sides. Geometrically, writing the equations in point-slope form shows they both have slope $m = 1/3$ so they are parallel but a solution to the first, such as $(x, y) = (0, -2/3)$, is not a solution of the second, so the lines do not overlap.

2. The system $\begin{cases} y - x = 1 \\ 2y + x = 1 \end{cases}$ is consistent because it has at least one solution, $x = -1/3$, $y = 2/3$ which can be verified by substitution. Geometrically, writing the lines in point-slope form shows they have slopes $m_1 = 1$ and $m_2 = -1/2$ respectively which are not equal so the lines are not parallel. Since they lie in the plane they intersect at a point, in fact $(-1/3, 2/3)$, which is a solution of the system.

Definition: A **homogeneous system** is a linear system in which every linear equation is homogeneous (i.e. right-hand side constant terms are all zero). A **non-homogeneous system** is a system in which at least one linear equation is non-homogeneous.

Example 1-8

The system $\begin{cases} x + y = 0 \\ 2x - 4y = 0 \end{cases}$ is homogeneous.

Every homogeneous system is consistent since setting all unknowns equal zero ($x = 0$, $y = 0$ in the previous example) will be a solution to the linear system.

Definition: If the number of equations in a linear system equals the number of unknowns the system is called **determined**. If the number of equations is less than the number of unknowns the system is **underdetermined**. If the number of equations exceeds the number of unknowns the system is **overdetermined**.

Example 1-9

Describe the following systems of linear equations:

1.

$$\begin{aligned} 3x + 2z &= 7 \\ x - 4y - 4z &= 3 \\ 3x + 3y + 8z &= 0 \end{aligned}$$

There are 3 equations with 3 unknowns, thus the system is determined. It is non-homogeneous since the right hand side is not identically zero.

2.

$$\begin{aligned} -x + 2z + w &= 0 \\ 2x + 3y - 5z + w &= 0 \\ z - 2w &= 0 \end{aligned}$$

There are 3 equations with 4 unknowns, therefore the system is underdetermined. It is homogeneous since the right hand side is identically zero and so also a consistent system as $(x, y, z, w) = (0, 0, 0, 0)$ will be at least one solution.

3.

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

There are 3 equations with 2 unknowns, therefore the system is overdetermined. It is also non-homogeneous.

1.4 Parameterization of Solutions

When a linear system has an infinite number of solutions, these can be conveniently described by the introduction of one or more **parameters**. These are independent variables in the solution description which can be assigned any numerical values to produce one of the solutions to the linear system.

Example 1-10

Show that the following parametric form

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

where parameter s is any real number, is a solution to the underdetermined linear system

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

Solution:
We must show that $x = 2s + 1$, $y = 1 - s$, $z = s$ satisfies each linear equation in the system for any value of s. Direct substitution in the left-hand side of the first equation gives

$$\begin{aligned} 2x + 3y - z &= 2(2s + 1) + 3(1 - s) - s \\ &= 4s + 2 + 3 - 3s - s \\ &= 5 \end{aligned}$$

as required. Substitution into the second linear equation gives

$$\begin{aligned} -x + 2z &= -(2s + 1) + 2s \\ &= -2s - 1 + 2s \\ &= -1 \end{aligned}$$

showing that it is also satisfied. Choosing a particular value of s in the parametric form (such as $s = 2$) would provide $(x, y, z) = (5, -1, 2)$, one of the infinitely many solutions of the system.

We will see that any consistent linear system with more than one solution can have its **general solution** described, using enough parameters, in a **parametric form**. Such a parametric solution is not unique and can be written in several ways. The reader may verify that

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

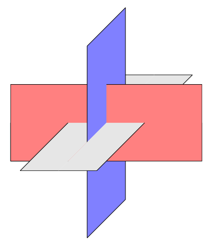
for parameter t is also a parametric solution to the linear system in Example 1-10 .

We can interpret our parametric solution in the previous example geometrically. A linear equation in three unknowns has the form

$$a_1x + a_2y + a_3z = b\ .$$

If at least one of the coefficients a_i is non-zero we will see that its solutions (x, y, z) constitute a plane in three-dimensional space. A solution of a linear system involving two equations in three unknowns requires geometrically that the solutions lie simultaneously on two planes. The intersection of two planes that are not parallel is a line. Our first parametric solution, by finding the points generated by evaluating it for all possible values of s, describes this line in three-dimensional space.

If we had a third equation in our linear system we would have had a determined system. We would have expected typically a single solution since three planes (assuming each equation represents a plane) typically intersect at a point. The line intersection of the first two planes will be cut at a single point by the third plane.



If we had added a fourth equation forming an overdetermined system we typically would expect no solution since if each equation represents a plane, the point that is the intersection of the first three planes will not, in general, lie on the fourth.

So generally one expects that underdetermined systems have an infinite number of solutions, determined systems have one solution and overdetermined systems have no solution. There are exceptions, however, due to the fact that (as in a three dimensional example such as ours) some of the planes could be parallel and, if so, either overlap or not. If they overlap then one could really have removed one of them without affecting the system. If they don't overlap then the system must be inconsistent. Also a linear equation in our three dimensional example may not represent a plane, for instance, if all the coefficients $a_i = 0$. In that case if $b = 0$ such an equation could be removed as it produces no restriction on the variables $(0 = 0)$. If $b \neq 0$ then that equation has no solution $(0 \neq b)$ and hence the linear system would be inconsistent. Other exceptions are also possible. As such we will have to come up with a more systematic analysis of our linear systems. That said, if, for instance, the constants in our linear system were generated completely at random our expectation on the nature of the solution set would be based on a consideration of the number of equations and the number of unknowns.

As a final comment, the behaviour of linear equations gives us insight into the behaviour of arbitrary equations. For instance, if we had a system of two non-linear equations in three unknowns (three dimensions) each equation would represent more generally a surface, not a plane. However if a solution exists (the surfaces intersect) then in the neighbourhood around that point the surfaces could typically be approximated by their tangent planes which, as we have discussed, would intersect in a line. Now for our surfaces as we move further away from our initial point we do not expect a straight line, but rather a curve for the solution. So, for instance, two intersecting spherical surfaces will typically intersect in a circle.[3] In the case of having three equations in three unknowns we can again consider that in the vicinity of a solution the surfaces will behave like three intersecting tangent planes. As such we expect the solution to be typically isolated (a point) since the intersecting planes, generally speaking, would have a unique solution. As an example, a third spherical surface would intersect the circular intersection of the first two at isolated points. As such our expectation for general systems of equations follows from the behaviour of linear ones. If the number of equations equals the number of unknowns we expect, if solution exist, that they are isolated solutions. Having fewer equations than unknowns we expect that if solutions exist that there will be an infinite number of them. Finally if we have more equations than unknowns we expect no solutions.

[3]They could intersect at a point if they just touched each other. Note that the tangent planes in this case are also exceptional as they are parallel and coincident.

1.5 Elementary Methods for Solving Linear Systems

So far we have only checked that certain values are solutions to linear equations and systems. The obvious question is how does one find these solutions. To solve a linear system involving few variables there are several strategies that involve eliminating variables.

Example 1-11

Solve the linear system

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

Solution:
One method to eliminate a variable is to add a suitable multiple of one equation to the other. In this example simply adding both sides of the equation eliminates the y variable:

$$\begin{aligned} 5x + y + 2x - y &= 3 + 4 \\ 7x &= 7 \\ x &= 1 \end{aligned}$$

Substitution of $x = 1$ into the second original equation then gives

$$\begin{aligned} 2(1) - y &= 4 \\ 2 - y &= 4 \\ -y &= 2 \\ y &= -2 \end{aligned}$$

Thus $x = 1$ and $y = -2$ is a solution to the system.

A second method to solve the system is the **method of substitution**. Solving the first equation for y gives $y = 3 - 5x$. Substitution of $3 - 5x$ for y in the second equation then gives

$$\begin{aligned} 2x - (3 - 5x) &= 4 \\ 7x - 3 &= 4 \\ 7x &= 7 \\ x &= 1 \end{aligned}$$

Then substitution of $x = 1$ into $y = 3 - 5x$ gives $y = 3 - 5(1) = -2$ as before.

One can easily check that $(x, y) = (1, -2)$ solves the original system.

For a linear system with more variables and many non-zero constants like

$$\begin{aligned} x + y + 2z &= 9 \\ 2x + 4y - 3z &= 1 \\ 3x + 6y - 5z &= 0 \end{aligned}$$

we cannot easily eliminate the variables to obtain a solution. Thus we need to develop systematic techniques to solve linear systems.

Chapter 2: Matrices

2.1 Definition of a Matrix

Definition: A **matrix** is a rectangular array of numbers each of which is called an **entry** of the matrix.

Example 2-1

$$A = \begin{bmatrix} 2 & 1 \\ -1 & 2 \\ 0 & 5 \end{bmatrix} \qquad B = \begin{bmatrix} 3 & 2 & -1 \\ 0 & 5 & 1 \\ 1 & 1 & -2 \end{bmatrix} \qquad C = \begin{bmatrix} 5 & 1 & 0 \end{bmatrix} \qquad D = \begin{bmatrix} 2 \\ 2 \\ 4 \\ 8 \end{bmatrix}$$

We can assign a name (A, B, etc.) to the matrix as shown. Each horizontal line of numbers is called a **row** of the matrix; each vertical line of numbers is called a **column**. Any matrix consists of m rows and n columns. Matrix A above has 3 rows and 2 columns. It is an example of a 3×2 ("three by two") matrix. These are the **dimensions** of the matrix. Matrix B is an example of a **square matrix**, a matrix in which the number of columns equals the number of rows. We can refer to an entry of a matrix by subscripts, row first then column. So for matrix A above, $a_{21} = -1$. We will use upper case letters to represent matrices as a whole and lower case letters for their entries. For a matrix A with entries a_{ij} it is convenient to use the notation $A = [a_{ij}]$.

Definition: A **row matrix** is a matrix of dimension $1 \times n$ and a **column matrix** is a matrix of dimension $m \times 1$.

In Example 2-1 matrix C is a row matrix and matrix D is a column matrix.

2.2 Linear Systems and Matrices

The **general form** (or **standard form**) of a linear system of m equations in n unknowns is:

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

Note that here two indices are required to keep track of the coefficients; the first index indicates to which equation the coefficient belongs, while the second index indicates of which variable it is the coefficient.

The constants and unknowns can be organized into the following matrices:

$$A = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{bmatrix}, \ X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \ B = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}.$$

The $m \times n$ matrix A is called the **coefficient matrix** of the system, X is the matrix of unknowns and B is the right-hand side of the system.[1]

Definition: The matrix which is made up of the coefficient matrix A and the right-hand side B is called the **augmented matrix** of the system and is denoted by $[A|B]$. For a linear system of m equations in n unknowns the augmented matrix is the $m \times (n+1)$ matrix:

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

Note that each equation of a linear system corresponds to a row of the augmented matrix and *vice versa*.

Example 2-2

Derive the augmented matrix for the following linear system.

$$\begin{aligned} 3x + 2z &= 7 \\ x + 4y - 4z &= 3 \\ 2x + 2y + 8z &= 1 \end{aligned}$$

Solution:
We have the coefficient, unknown, and right-hand side matrices:

$$A = \begin{bmatrix} 3 & 0 & 2 \\ 1 & 4 & -4 \\ 3 & 3 & 8 \end{bmatrix}, \quad X = \begin{bmatrix} x \\ y \\ z \end{bmatrix}, \quad B = \begin{bmatrix} 7 \\ 3 \\ 1 \end{bmatrix},$$

[1] Later we will show that a matrix multiplication operation can be introduced so that the linear system is reducible to the matrix equation $AX = B$.

so the augmented matrix is:

$$[A|B] = \left[\begin{array}{ccc|c} 3 & 0 & 2 & 7 \\ 1 & 4 & -4 & 3 \\ 3 & 3 & 8 & 1 \end{array}\right] .$$

Example 2-3

Write down the system of linear equations for the following augmented matrix.

$$\left[\begin{array}{cccc|c} -1 & 0 & 2 & 1 & -10 \\ 2 & 3 & -5 & 4 & 8 \\ 0 & 0 & 1 & -2 & 5 \end{array}\right]$$

Solution:
Identifying the coefficient matrix A and the unknown matrix B we see we have the linear system of three equations in four unknowns is given by

$$\begin{aligned} -x_1 + 2x_3 + x_4 &= -10 \\ 2x_1 + 3x_2 - 5x_3 + 4x_4 &= 8 \\ x_3 - 2x_4 &= 5 \end{aligned}$$

2.3 Row Echelon Form

Consider the linear system

$$\begin{aligned} 2x + 3y - 2z &= -7 \\ 3y + 2z &= 3 \\ 5z &= 15 \end{aligned}$$

Many of the coefficients are zero. It is, equivalently

$$\begin{aligned} 2x + 3y - 2z &= -7 \\ 0x + 3y + 2z &= 3 \\ 0x + 0y + 5z &= 15 \end{aligned}$$

and so its corresponding augmented matrix is therefore:

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

A system of this type is easy to solve by a process called **back-substitution**. In back-substitution we start by solving the last equation first for the final unknown. That result can then be used in the second last equation to solve for the second last unknown, etc.

Example 2-4

Solve the linear system:

$$\begin{aligned} 2x + 3y - 2z &= -7 \\ 3y + 2z &= 3 \\ 5z &= 15 \end{aligned}$$

Solution:
Solving the last equation first and using back-substitution gives:

- $5z = 15 \implies \boxed{z = 3}$
- $3y + 2z = 3 \implies 3y + 2(3) = 3 \implies 3y = -3 \implies \boxed{y = -1}$
- $2x + 3y - 2z = -7 \implies 2x + 3(-1) - 2(3) = -7 \implies 2x = 2 \implies \boxed{x = 1}$

The solution is therefore

$$x = 1,\ y = -1,\ z = 3\,.$$

An even simpler linear system to solve would have been

$$\begin{aligned} 1x + \frac{3}{2}y - z &= -\frac{7}{2} \\ 0x + 1y + \frac{2}{3}z &= 1 \\ 0x + 0y + 1z &= 3 \end{aligned} \qquad \implies \qquad \left[\begin{array}{ccc|c} 1 & \frac{3}{2} & -1 & -\frac{7}{2} \\ 0 & 1 & \frac{2}{3} & 1 \\ 0 & 0 & 1 & 3 \end{array}\right]$$

since, upon back-substitution there would be no division required. Finally an even easier linear system would have the form:

$$\begin{aligned} 1x + 0y + 0z &= 1 \\ 0x + 1y + 0z &= -1 \\ 0x + 0y + 1z &= 3 \end{aligned} \qquad \Longrightarrow \qquad \left[\begin{array}{ccc|c} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 3 \end{array}\right]$$

Here finding the solution requires no calculation at all! Note that these last three linear systems are equivalent in the sense that they all have the same solution. The facility with which a linear system in these forms may be solved suggests that it would be desirable if for *any* linear system we could find an equivalent linear system that had a similar arrangement of zeros and ones in its augmented matrix. We now show how that may be done, commencing with suitable definitions.

Definition: Two linear systems with m equations in n unknowns are **equivalent** to each other if they have the same solutions.

Definition: The leftmost, nonzero entry in a row of a matrix is called the **leading entry** or **pivot**.

Definition: A matrix is in **row echelon form** (**REF**) if:

1. All zero rows (i.e. rows consisting entirely of zeros) are at the bottom of the matrix.
2. All elements below a leading entry (pivot) are zero.
3. Each leading entry is to the right of the leading entries of all rows above it.
4. Each leading entry is equal to 1. Such an entry is called the **leading 1**.

Definition: A matrix is in **reduced row echelon form** (**RREF**) if it is in row echelon form and the leading one is the only nonzero entry in its column.

Example 2-5

Determine if the matrices are in REF, RREF, or neither.

1. $$A = \begin{bmatrix} 1 & 2 & -4 \\ 0 & 1 & -5 \\ 0 & 0 & 0 \end{bmatrix} \Leftarrow REF$$
2. $$B = \begin{bmatrix} 1 & 0 & 3 \\ 0 & 1 & 4 \\ 0 & 0 & 0 \end{bmatrix} \Leftarrow RREF$$
3. $$C = \begin{bmatrix} 2 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \end{bmatrix} \Leftarrow \text{Neither form}$$
4. $$D = \begin{bmatrix} 1 & 0 & 0 & -2 \\ 0 & 1 & 0 & 5 \\ 0 & 0 & 1 & -3 \end{bmatrix} \Leftarrow RREF$$
5. $$E = \begin{bmatrix} 1 & -7 & 5 & 5 \\ 0 & 1 & 3 & 2 \end{bmatrix} \Leftarrow REF$$
6. $$F = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \Leftarrow \text{Neither form}$$

2.4 Elementary Row Operations

Performing one of the following three operations, called **elementary row operations** on a given linear system gives an equivalent linear system.

1. Interchange two equations (or rows), $R_i \leftrightarrow R_j$.
2. Multiply an equation (row) by a nonzero number, $R_i \to cR_i$.
3. Add a multiple of one equation (row) to another equation (row), $R_i \to R_i + cR_j$.

Here R denotes a row of an augmented matrix and c represents a number.

Note: Every augmented matrix can be reduced to its row echelon form using elementary row operations. This process is called **Gaussian elimination**. Every augmented matrix can be reduced to its reduced row echelon form by a process called **Gauss-Jordan elimination**.

2.4.1 Steps in Gaussian Elimination

1. Locate the leftmost nonzero column in the augmented matrix. If the top entry of the column is zero, interchange the top row with another so the top entry (call it a) is nonzero.
2. If a is not a leading 1 make it so by multiplying the row by $1/a$.
3. Make all other entries in the column below the leading 1 equal zero by adding suitable multiples of the first row to the remaining rows.
4. Consider the remaining matrix produced by ignoring the top row. Repeat steps 1–4 on that matrix. If the remaining matrix has no rows the original matrix is now in row echelon form.

Note that once the matrix is in row echelon form the system may be solved as shown in the following examples.

Example 2-6

Solve each linear system using Gaussian elimination.

1. From the end of the last chapter (unsolved):

$$\begin{aligned} x + y + 2z &= 9 \\ 2x + 4y - 3z &= 1 \\ 3x + 6y - 5z &= 0 \end{aligned}$$

Solution:
The augmented matrix for the system is

$$[A|B] = \left[\begin{array}{ccr|r} 1 & 1 & 2 & 9 \\ 2 & 4 & -3 & 1 \\ 3 & 6 & -5 & 0 \end{array}\right]$$

Since the top left entry is already a leading 1, zero the 2 below it by adding -2 times row 1 to row 2. (This is the same as subtracting 2 times row 1 from row 2.)

$$R_2 \to R_2 + (-2)R_1 \left[\begin{array}{ccr|r} 1 & 1 & 2 & 9 \\ 0 & 2 & -7 & -17 \\ 3 & 6 & -5 & 0 \end{array}\right]$$

Next zero the first entry in the third row by adding -3 times row 1 to row 3.

$$R_3 \to R_3 - 3R_1 \left[\begin{array}{ccc|c} 1 & 1 & 2 & 9 \\ 0 & 2 & -7 & -17 \\ 0 & 3 & -11 & -27 \end{array}\right]$$

To get a leading 1 in the second row, multiply the row by 1/2, or equivalently, divide the row by 2.

$$R_2 \to \tfrac{1}{2}R_2 \left[\begin{array}{ccc|c} 1 & 1 & 2 & 9 \\ 0 & 1 & -\frac{7}{2} & -\frac{17}{2} \\ 0 & 3 & -11 & -27 \end{array}\right]$$

Zero the second entry in the third row by adding -3 times row 2 to row 3.

$$R_3 \to R_3 - 3R_2 \left[\begin{array}{ccc|c} 1 & 1 & 2 & 9 \\ 0 & 1 & -\frac{7}{2} & -\frac{17}{2} \\ 0 & 0 & -\frac{1}{2} & -\frac{3}{2} \end{array}\right]$$

Get a leading 1 in the third row by dividing the row by -1/2, or, equivalently, multiplying it by -2.

$$R_3 \to -2R_3 \left[\begin{array}{ccc|c} 1 & 1 & 2 & 9 \\ 0 & 1 & -\frac{7}{2} & -\frac{17}{2} \\ 0 & 0 & 1 & 3 \end{array}\right]$$

The linear system represented by the last augmented matrix is

$$\begin{aligned} x + y + 2z &= 9 \\ y - \frac{7}{2}z &= -\frac{17}{2} \\ z &= 3 \end{aligned}$$

so back-substitution gives the following.

- $\boxed{z = 3}$
- $y - \frac{7}{2}z = -\frac{17}{2} \implies y - \frac{7}{2}(3) = -\frac{17}{2} \implies y = -\frac{17}{2} + \frac{21}{2} \implies \boxed{y = 2}$
- $x + y + 2z = 9 \implies x + 2 + 2(3) = 9 \implies \boxed{x = 1}$

The solution is therefore $x = 1,\ y = 2,\ z = 3$ which is easily checked in the original system.

2.

$$\begin{aligned} x_1 + x_2 &= 1 \\ 4x_1 - x_2 &= -6 \\ 2x_1 - 3x_2 &= 8 \end{aligned}$$

Solution:
The augmented matrix is

$$[A|B] = \left[\begin{array}{cc|c} 1 & 1 & 1 \\ 4 & -1 & -6 \\ 2 & -3 & 8 \end{array}\right]$$

which may be put into row echelon form as follows:

$$\begin{array}{l} R_2 \to R_2 - 4R_1 \\ R_3 \to R_3 - 2R_1 \end{array} \left[\begin{array}{cc|c} 1 & 1 & 1 \\ 0 & -5 & -10 \\ 0 & -5 & 6 \end{array}\right] \Rightarrow R_2 \to -\tfrac{1}{5}R_2 \left[\begin{array}{cc|c} 1 & 1 & 1 \\ 0 & 1 & 2 \\ 0 & -5 & 6 \end{array}\right] \Rightarrow R_3 \to R_3 + 5R_2 \left[\begin{array}{cc|c} 1 & 1 & 1 \\ 0 & -5 & -10 \\ 0 & 0 & 16 \end{array}\right]$$

The last augmented matrix presents a contradiction as its last equation is $0x_1 + 0x_2 = 16$ or simply $0 = 16$ which is never true. There is therefore no solution to this system.

3.

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

Solution:

$$[A|B] = \left[\begin{array}{rrrr|r} 1 & -1 & 2 & -1 & -1 \\ 2 & 1 & -2 & -2 & -2 \\ -1 & 2 & -4 & 1 & 1 \\ 3 & 0 & 0 & -3 & -3 \end{array}\right]$$

$$\Downarrow$$

$$\begin{array}{r} \\ R_2 \to R_2 - 2R_1 \\ R_3 \to R_3 + R_1 \\ R_4 \to R_4 - 3R_1 \end{array} \left[\begin{array}{rrrr|r} 1 & -1 & 2 & -1 & -1 \\ 0 & 3 & -6 & 0 & 0 \\ 0 & 1 & -2 & 0 & 0 \\ 0 & 3 & -6 & 0 & 0 \end{array}\right]$$

$$\Downarrow$$

$$\begin{array}{r} \\ R_2 \to \frac{1}{3}R_2 \\ \\ \\ \end{array} \left[\begin{array}{rrrr|r} 1 & -1 & 2 & -1 & -1 \\ 0 & 1 & -2 & 0 & 0 \\ 0 & 1 & -2 & 0 & 0 \\ 0 & 3 & -6 & 0 & 0 \end{array}\right]$$

$$\Downarrow$$

$$\begin{array}{r} \\ \\ R_3 \to R_3 - R_2 \\ R_4 \to R_4 - 3R_2 \end{array} \left[\begin{array}{rrrr|r} 1 & -1 & 2 & -1 & -1 \\ 0 & 1 & -2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

The linear system corresponding to the row echelon form is then

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

This system will have an infinite number of solutions. To characterize them we will introduce parameters.

Definition: The variables that correspond to the leading entries of the row echelon form of an augmented matrix are called the **leading variables** or **dependent variables**. The remaining variables are called the **free variables** or **independent variables**.

To solve a system of linear equations, set the free variables equal to parameters and use the row echelon form with back-substitution to solve for the leading variables.

Example 2-7

Complete the solution to the Question 3 of Example 2-6 .

Solution:
The row echelon form found and corresponding linear system are

$$\left[\begin{array}{cccc|c} 1 & -1 & 2 & -1 & -1 \\ 0 & 1 & -2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right] \Leftrightarrow \begin{aligned} w - x + 2y - z &= -1 \\ x - 2y &= 0 \\ 0 &= 0 \\ 0 &= 0 \end{aligned} .$$

The leading entries are circled and are found in the w and x variable columns. Thus w and x are the dependent variables and y and z are the independent variables. So introduce two parameters for the independent variables letting $\boxed{y = s}$ and $\boxed{z = t}$. Next solve for the dependent variables in terms of the parameters using back-substitution.

- $x - 2y = 0 \implies x = 2y \implies \boxed{x = 2s}$
- $w - x + 2y - z = -1 \implies w - 2s + 2s - t = -1 \implies \boxed{w = -1 + t}$

We can write the solution to the system as

$$w = -1 + t,\ x = 2s,\ y = s,\ z = t\,,$$

where s and r are parameters taking on any numbers. We can also write the solution in the matrix form:

$$\begin{bmatrix} w \\ x \\ y \\ z \end{bmatrix} = \begin{bmatrix} -1 + t \\ 2s \\ s \\ t \end{bmatrix} .$$

2.4.2 Gaussian Elimination in Practice

Gaussian Elimination, as has been presented, is an algorithmic method for finding solutions to linear systems and can be easily encoded into a computer program. However, as we saw in Question 1 of Example 2-6 one can often get fractions in later row entries as one produces a leading 1. If working by hand it is often easier to use the elementary row operations more liberally so that this may be avoided. Such strategies include

- Swapping rows if there is a leading 1 already in a column, even if the top row leading entry is nonzero.
- Zeroing leading entries in lower rows that are a multiple of a top row leading entry by subtracting the appropriate multiple, before making the top row leading entry equal 1 .

- Converting leading entries to 1 *after* the matrix is otherwise reduced, or never converting them at all.

Example 2-8

Solve the linear system.

$$\begin{aligned} 2x_1 - x_2 - x_3 &= 3 \\ -6x_1 + 6x_2 + 5x_3 &= -3 \\ 4x_1 + 4x_2 + 7x_3 &= 3 \end{aligned}$$

Solution:
The augmented matrix for the system is:

$$\left[\begin{array}{rrr|r} 2 & -1 & -1 & 3 \\ -6 & 6 & 5 & -3 \\ 4 & 4 & 7 & 3 \end{array}\right]$$

Perform the row operations:

$$\left[\begin{array}{rrr|r} 2 & -1 & -1 & 3 \\ -6 & 6 & 5 & -3 \\ 4 & 4 & 7 & 3 \end{array}\right] \Rightarrow \begin{array}{r} R_2 \to R_2 + 3R_1 \\ R_3 \to R_3 - 2R_1 \end{array} \left[\begin{array}{rrr|r} 2 & -1 & -1 & 3 \\ 0 & 3 & 2 & 6 \\ 0 & 6 & 9 & -3 \end{array}\right] \Rightarrow \begin{array}{r} \\ R_3 \to R_3 - 2R_2 \end{array} \left[\begin{array}{rrr|r} 2 & -1 & -1 & 3 \\ 0 & 3 & 2 & 6 \\ 0 & 0 & 5 & -15 \end{array}\right]$$

At this stage one could get row echelon form by multiplying the rows by 1/3, 1/2 and 1/5 respectively to get:

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

but it is simpler just to do back-substitution on the previous matrix:

- $5x_3 = -15 \implies \boxed{x_3 = -3}$
- $3x_2 + 2x_3 = 6 \implies 3x_2 + 2(-3) = 6 \implies \boxed{x_2 = 4}$
- $2x_1 - x_2 - x_3 = 3 \implies 2x_1 - 4 + 3 = 3 \implies \boxed{x_1 = 2}$

So the solution is

$$x_1 = 2,\ x_2 = 4,\ x_3 = -3\ .$$

2.4.3 Gauss-Jordan Elimination

Gauss-Jordan elimination takes Gaussian elimination one further step to produce an augmented matrix in reduced row echelon form. One does the following steps:

1. Perform Gaussian elimination to put the augmented matrix in row echelon form (REF).
2. Add suitable multiples of the last nonzero row to the rows above it to introduce zeros into them above the leading 1 of this row.

3. Consider the remaining matrix produced by ignoring the last nonzero row and any zero rows beneath it, if any. Repeat steps 2–3 on that matrix. If the remaining matrix has no rows the original matrix is now in reduced row echelon form (RREF).

Example 2-9

Solve the linear system using Gauss-Jordan elimination:

$$\begin{aligned} -y + 5z &= 9 \\ x + y + 2z &= 8 \\ 3x - 7y + 4z &= 10 \end{aligned}$$

Solution:

$$[A|B] = \left[\begin{array}{ccc|c} 0 & -1 & 5 & 9 \\ 1 & 1 & 2 & 8 \\ 3 & -7 & 4 & 10 \end{array}\right]$$

First put the augmented matrix in row echelon form:

$$R_1 \leftrightarrow R_2 \quad \left[\begin{array}{ccc|c} 1 & 1 & 2 & 8 \\ 0 & -1 & 5 & 9 \\ 3 & -7 & 4 & 10 \end{array}\right]$$

$$\Downarrow$$

$$R_3 \to R_3 - 3R_1 \quad \left[\begin{array}{ccc|c} 1 & 1 & 2 & 8 \\ 0 & -1 & 5 & 9 \\ 0 & -10 & -2 & -14 \end{array}\right]$$

$$\Downarrow$$

$$R_3 \to R_3 - 10R_2 \quad \left[\begin{array}{ccc|c} 1 & 1 & 2 & 8 \\ 0 & -1 & 5 & 9 \\ 0 & 0 & -52 & -104 \end{array}\right]$$

$$\Downarrow$$

$$\begin{array}{r} R_2 \to -R_2 \\ R_3 \to -\frac{1}{52}R_3 \end{array} \quad \left[\begin{array}{ccc|c} 1 & 1 & 2 & 8 \\ 0 & 1 & -5 & -9 \\ 0 & 0 & 1 & 2 \end{array}\right] \quad \text{(REF)}$$

To achieve reduced row echelon form work from the bottom of the matrix upward, to get zeros above any leading one.

$$\Downarrow$$

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

$$\Downarrow$$

$$R_1 \to R_1 - R_2 \quad \left[\begin{array}{ccc|c} 1 & 0 & 0 & 3 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 2 \end{array}\right] \quad \text{(RREF)}$$

The unique solution to the system is therefore

$$x = 3,\ y = 1,\ z = 2\,.$$

2.4.4 Interpreting the Reduced Row Echelon Form

Once the augmented matrix has been put in RREF by Gauss-Jordan elimination one proceeds as follows:

1. If the last nonzero row of the matrix is of the form
$$\left[\begin{array}{cccc|c} 0 & 0 & \dots & 0 & 1 \end{array}\right]$$
then the linear system has **no solution**.[2]

2. If not, assign parameters to any free (independent) variables and solve for the leading (dependent) variables using the nonzero rows. If no such parameters are needed there is a **unique solution**, otherwise one has **infinitely many solutions**.

Example 2-10

Find values of m such that the linear system

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

has:

1. No solution
2. Infinitely many solutions
3. A unique solution

Solution:
The augmented matrix is:

$$[A|B] = \left[\begin{array}{ccc|c} 1 & 1 & 1 & 1 \\ 2 & 1 & 4 & 3 \\ 2 & 2 & 2m^2 & 2m \end{array}\right]$$

The following matrix is equivalent:

$$\begin{array}{r} \\ R_2 \to R_2 - 2R_1 \\ R_3 \to R_3 - 2R_1 \end{array} \left[\begin{array}{ccc|c} 1 & 1 & 1 & 1 \\ 0 & -1 & 2 & 1 \\ 0 & 0 & 2m^2-2 & 2m-2 \end{array}\right]$$

1. There is no solution if you have the contradiction $2m^2 - 2 = 0$ and $2m - 2 \neq 0$ in the last row. Solving gives
$$2m^2 - 2 = 0 \implies m^2 - 1 = 0 \implies (m-1)(m+1) = 0 \implies m = \pm 1$$
and
$$2m - 2 \neq 0 \implies m - 1 \neq 0 \implies m \neq 1$$
Thus $m = \pm 1$ and $m \neq 1$ for no solution. This implies $m = -1$ gives no solution.

2. There are an infinite number of solutions when the last row is 0=0 since then the number of leading terms is less than the number of unknowns. We saw $2m^2 - 2 = 0 \implies m = \pm 1$. Similarly $2m - 2 = 0 \implies m = 1$. Thus we get an infinite number of solutions when $m = 1$.

[2] Note one can stop at REF if its last nonzero row indicates no solution.

3. Finally a unique solution occurs the rest of the time, logically when $m \neq 1$ and $m \neq -1$. We can see this directly by noting that when $2m^2 - 2 \neq 0$ we would be able to divide row three in the augmented matrix to get a leading 1, since $m \neq \pm 1 \implies 2m^2 - 2 \neq 0$.

Note:

1. For a given linear system, the row echelon form generated by Gaussian elimination is not unique. The reduced row echelon form generated by Gauss-Jordan elimination is however unique.
2. On a computer, Gaussian Elimination (finding REF and using back-substitution) is more efficient, in general, than Gauss-Jordan elimination.
3. Terminology is not universal. Some consider matrices for which the leading entry is not equal to 1 to be in row echelon form. Some call Gaussian elimination what we have called Gauss-Jordan elimination.

2.4.5 Rank of a Matrix

Characterization of the solutions of a linear system is simplified by the introduction of the rank of a matrix. It can be shown that any REF and the RREF of a matrix A always have the same number of nonzero rows allowing for the following definition.

Definition: The **rank** of a matrix A, rank(A), is the number of nonzero rows in the row echelon or reduced row echelon form of A.

Example 2-11

Find the rank of $A = \begin{bmatrix} 1 & 2 & -5 & 2 \\ 2 & -3 & 4 & 4 \\ 4 & 1 & -6 & 8 \end{bmatrix}$.

Solution:
Put the matrix A in REF form:

$$\begin{bmatrix} 1 & 2 & -5 & 2 \\ 2 & -3 & 4 & 4 \\ 4 & 1 & -6 & 8 \end{bmatrix} \Rightarrow \begin{matrix} R_2 \to R_2 - 2R_1 \\ R_3 \to R_3 - 4R_1 \end{matrix} \begin{bmatrix} 1 & 2 & -5 & 2 \\ 0 & -7 & 14 & 0 \\ 0 & -7 & 14 & 0 \end{bmatrix} \Rightarrow \begin{matrix} R_2 \to -\frac{1}{7}R_2 \\ R_3 \to R_3 - R_2 \end{matrix} \begin{bmatrix} 1 & 2 & -5 & 2 \\ 0 & 1 & -2 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Therefore rank$(A) = 2$.

Theorem 2-1: Consider a linear system of m equations in n unknowns, with coefficient matrix A and right-hand side matrix B. If p is the rank of A and q is the rank of $[A|B]$. The linear system has:

1. No solution if $p < q$.
2. A unique solution if $p = q = n$.
3. Infinitely many solutions if $p = q$ and $p < n$.

Example 2-12

Suppose the augmented matrix of a linear system is given by

$$[A|B] = \left[\begin{array}{ccc|c} 1 & 0 & 0 & 2 \\ 0 & 1 & 0 & 3 \\ 0 & 0 & x & y \end{array}\right] .$$

For what values of x and y give does the system have
1. No solution? 2. Exactly one solution? 3. Infinitely many solutions?

Solution:
The original system has $n = 3$ unknowns. We consider the various ranks which occur of A and $[A|B]$ when either x or y or both are zero.

1. No solution $\Rightarrow x = 0, y \neq 0$ since then $p = \text{rank}(A) = 2 < 3 = \text{rank}([A|B]) = q$
2. Exactly one solution $\Rightarrow x \neq 0$ since then $p = \text{rank}(A) = 3 = \text{rank}([A|B]) = q = n$.
3. Infinitely many solutions $\Rightarrow x = 0, y = 0$ since then
   $p = \text{rank}(A) = 2 = \text{rank}([A|B]) = q < 3 = n$.

Example 2-13

Consider the linear system $AX = B$, where:

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}, B = \begin{bmatrix} e \\ f \end{bmatrix}, a \neq 0$$

Determine conditions on the constants a, b, c, d, e, f so that:

1. $\text{rank}(A) = 2$.
2. $\text{rank}(A) = 1$ but $\text{rank}([A|B]) = 2$.
3. $\text{rank}(A)$ and $\text{rank}([A|B])$ are both 1 .

Solution:
The augmented matrix is

$$[A|B] = \left[\begin{array}{cc|c} a & b & e \\ c & d & f \end{array}\right] .$$

Since $a \neq 0$ perform the row operation $R_2 \to R_2 - \frac{c}{a} R_1$:

$$\left[\begin{array}{cc|c} a & b & e \\ 0 & d - \frac{bc}{a} & f - \frac{ce}{a} \end{array}\right]$$

Then:

1. $\text{rank}(A) = 2 \implies d - \frac{bc}{a} \neq 0 \implies ad - bc \neq 0$
2. $\text{rank}(A) = 1 \implies d - \frac{bc}{a} = 0 \implies ad - bc = 0$ and
   $\text{rank}([A|B]) = 2 \implies f - \frac{ce}{a} \neq 0 \implies af - ce \neq 0$
3. $\text{rank}(A) = 1 \implies ad - bc = 0$ and
   $\text{rank}([A|B]) = 1 \implies f - \frac{ce}{a} = 0 \implies af - ce = 0$

> Note: The student may wish to consider the implications if $a = 0$ but some other element of A (i.e. b, c, or d) is taken to be nonzero.

Theorem 2-1 has some useful corollaries. For a homogeneous system ($B = 0$) the last column of $[A|B]$ is zero so $p = q$ and further the trivial solution ($X = 0$) is always a solution to the system. This gives the corollary:

Corollary 1: A homogeneous linear system with $m \times n$ coefficient matrix A has:

1. A unique solution (trivial solution $X = 0$) if $\text{rank}(A) = n$.
2. Infinitely many solutions if $\text{rank}(A) < n$.

For coefficient matrix A we have $p = \text{rank}(A)$ necessarily less than or equal to its number of rows, which equals the number of equations m. Since for an underdetermined system (fewer equations m than unknowns n), it follows then that $p < n$, and one has the following corollaries:

Corollary 2: An underdetermined linear system has no solution or infinitely many solutions.

Corollary 3: An underdetermined homogeneous linear system has infinitely many solutions.

Example 2-14

Without solving the system what can you say about the number of solution for the following systems?

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

   Solution:
   This is a homogeneous system with 3 equations in 4 unknowns and so underdetermined. Therefore there are infinitely many solutions.

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

   Solution:
   This is a non-homogeneous system with 3 equations in 4 unknowns and so underdetermined. Therefore there are either no solutions or infinitely many solutions.

2.5 Matrix Equality, Addition, and Subtraction

Definition: Two matrices are **equal** if they have the same dimensions and their corresponding entries are equal.

Therefore $A = B$ implies $a_{ij} = b_{ij}$ for all indices i and j.

Example 2-15

Given the matrices:

$$A = \begin{bmatrix} 2 & 0 \\ 1 & 3 \end{bmatrix} \qquad B = \begin{bmatrix} \frac{8}{4} & 0 \\ 1 & \frac{12}{4} \end{bmatrix}$$

We see that $A = B$ since they are both 2×2 matrices with equal corresponding entries.

Definition: Let $A = [a_{ij}]$ and $B = [b_{ij}]$ be matrices with the same dimensions. The **sum** of A and B, written $\boldsymbol{A + B}$, is the matrix obtained by adding corresponding entries of A and B. The **difference** of A and B, written $\boldsymbol{A - B}$, is obtained by subtracting corresponding entries from entry A by entry B. In symbols:

$$A + B = [a_{ij}] + [b_{ij}] = [a_{ij} + b_{ij}]$$

$$A - B = [a_{ij}] - [b_{ij}] = [a_{ij} - b_{ij}]$$

Example 2-16

If

$$A = \begin{bmatrix} 1 & 0 & -1 \\ 2 & 3 & 8 \end{bmatrix} \quad B = \begin{bmatrix} 2 & 5 & -1 \\ 0 & 1 & 3 \end{bmatrix},$$

then $A + B$ and $A - B$ are, respectively,

$$A + B = \begin{bmatrix} (1+2) & (0+5) & (-1-1) \\ (2+0) & (3+1) & (8+3) \end{bmatrix} = \begin{bmatrix} 3 & 5 & -2 \\ 2 & 4 & 11 \end{bmatrix}$$

$$A - B = \begin{bmatrix} (1-2) & (0-5) & (-1+1) \\ (2-0) & (3-1) & (8-3) \end{bmatrix} = \begin{bmatrix} -1 & -5 & 0 \\ 2 & 2 & 5 \end{bmatrix}.$$

Note that $A + B$ and $A - B$ will not be defined if the matrices do not have the same dimension.

Example 2-17

For the matrices

$$A = \begin{bmatrix} 2 & 5 \\ 3 & 1 \end{bmatrix} \qquad B = \begin{bmatrix} 1 & -1 \\ 0 & 5 \\ 2 & 3 \end{bmatrix}$$

the sum $A + B$ and difference $A - B$ are not possible since A is a 2×2 matrix and B is a 3×2 matrix.

Note $A + B = B + A$ when the sum exists.

Definition: A **zero matrix**, denoted by $\mathit{0}$, is an $m \times n$ matrix where all entries are zero.

One may write $\mathit{0}_{mn}$ to make the dimension explicit. Clearly $A + \mathit{0} = A$ for the zero matrix with same dimension as A.

2.6 Scalar Multiplication

Definition: Let $A = [a_{ij}]$ be a matrix and c be a number (a *scalar*) then the **scalar product** of c times A, written **cA**, is the matrix obtained by multiplying each entry of A by c. In symbols one has

$$cA = [ca_{ij}]\,.$$

Example 2-18

If

$$A = \begin{bmatrix} 2 & 1 \\ -1 & 5 \\ 3 & 0 \end{bmatrix},\ c = 3$$

then the scalar product is

$$cA = 3A = \begin{bmatrix} (3)(2) & (3)(1) \\ (3)(-1) & (3)(5) \\ (3)(3) & (3)(0) \end{bmatrix} = \begin{bmatrix} 6 & 3 \\ -3 & 15 \\ 9 & 0 \end{bmatrix}.$$

Definition: The **negative** of matrix $A = [a_{ij}]$, written **-A**, is defined to be $(-1)A$.

Clearly $-A = [-a_{ij}]$ and $-A + A = \mathit{0}$.

One can combine scalar multiplication and matrix addition and subtraction to form new matrices.

Example 2-19

Let

$$A = \begin{bmatrix} -1 & 3 & 2 \\ 0 & 1 & -1 \\ 2 & 1 & 4 \end{bmatrix} \qquad B = \begin{bmatrix} 1 & -1 & 0 \\ 2 & 5 & 1 \\ -1 & 1 & -2 \end{bmatrix}$$

Find $A + 2B$ and $3A - B$.

Solution:

$$A + 2B = \begin{bmatrix} -1 & 3 & 2 \\ 0 & 1 & -1 \\ 2 & 1 & 4 \end{bmatrix} + \begin{bmatrix} 2 & -2 & 0 \\ 4 & 10 & 2 \\ -2 & 2 & -4 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 2 \\ 4 & 11 & 1 \\ 0 & 3 & 0 \end{bmatrix}$$

$$3A - B = \begin{bmatrix} -3 & 9 & 6 \\ 0 & 3 & -3 \\ 6 & 3 & 12 \end{bmatrix} - \begin{bmatrix} 1 & -1 & 0 \\ 2 & 5 & 1 \\ -1 & 1 & -2 \end{bmatrix} = \begin{bmatrix} -4 & 10 & 6 \\ -2 & -2 & -4 \\ 7 & 2 & 14 \end{bmatrix}$$

2.7 Solutions of Homogeneous Linear Systems

One can use scalar multiplication and addition to represent solutions to a linear system.

Example 2-20

Find the solution of the homogeneous linear system

$$\begin{aligned} x_3 + 2x_4 - x_5 &= 0 \\ x_4 - x_5 &= 0 \\ x_3 + 3x_4 - 2x_5 &= 0 \\ 2x_1 + 4x_2 + x_3 + 7x_4 &= 0 \end{aligned}$$

if its augmented matrix reduces to

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

and write your solution using matrix addition and the scalar product.

Solution:
The RREF augmented matrix and corresponding equivalent linear system are

$$\left[\begin{array}{ccccc|c} 1 & 2 & 0 & 0 & 3 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right] \quad \Leftrightarrow \quad \begin{aligned} x_1 + 2x_2 + 3x_5 &= 0 \\ x_3 + x_5 &= 0 \\ x_4 - x_5 &= 0 \\ 0 &= 0 \end{aligned}$$

The leading ones (circled) are in the variable x_1, x_3, and x_4 columns. These are the leading (dependent) variables and the remaining variables, x_2 and x_5, are the free (independent) variables. Assigning parameters to the latter we have $\boxed{x_2 = s}$ and $\boxed{x_5 = t}$. Next solve the dependent variables using back-substitution.

- $x_4 - x_5 = 0 \implies x_4 - t = 0 \implies \boxed{x_4 = t}$
- $x_3 + x_5 = 0 \implies x_3 + t = 0 \implies \boxed{x_3 = -t}$
- $x_1 + 2x_2 + 3x_5 = 0 \implies x_1 + 2s + 3t = 0 \implies \boxed{x_1 = -2s - 3t}$

Writing the solution as a column matrix we have

$$X = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} -2s - 3t \\ s \\ -t \\ t \\ t \end{bmatrix}.$$

Using matrix addition we can break the solution matrix into a column matrix for each parameter which we then factor out using scalar multiplication.

$$X = \begin{bmatrix} -2s \\ s \\ 0 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} -3t \\ 0 \\ -t \\ t \\ t \end{bmatrix} = s\begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + t\begin{bmatrix} -3 \\ 0 \\ -1 \\ 1 \\ 1 \end{bmatrix}.$$

The general solution to the homogeneous linear system in the last example is therefore $X = sX_1 + tX_2$ where

$$X_1 = \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad X_2 = \begin{bmatrix} -3 \\ 0 \\ -1 \\ 1 \\ 1 \end{bmatrix}$$

Both X_1 and X_2 are themselves solutions to the system by $s = 1$, $t = 0$ and $s = 0$, $t = 1$ respectively. These solutions are called **basic solutions** of the homogeneous linear system and can be systematically found by reducing the system to RREF using Gauss-Jordan elimination. Basic solutions are not unique since we can always replace the multiplicative parameter s by, say, $2\tilde{s}$ in the general solution X and then absorb the 2 into the basic solution using scalar multiplication.[3] However up to such a scalar multiple they are unique. The number of basic solutions will correspond to the number of free parameters.

The solution of the last homogeneous linear system, $X = sX_1 + tX_2$ suggests the following general definition.

Definition: Let X_1, X_2, $\ldots$, X_n be matrices of the same dimension and c_1, c_2, $\ldots c_n$ be numbers, then

$$c_1X_1 + c_2X_2 + \ldots + c_nX_n$$

is a **linear combination** of X_1, X_2, $\ldots$, X_n .

With this definition we can now characterize solutions to homogeneous linear systems.

Theorem 2-2: Let A be the coefficient matrix of a homogeneous linear system of m equations in n unknowns. Then the system has $n - \text{rank}(A)$ basic solutions and every solution to the system is a linear combination of these basic solutions and *vice versa*. (If the system has no basic solutions it has only the trivial solution $X = 0$.)

Let $Y = a_1X_1 + \ldots + a_kX_k$ and $Z = b_1X_1 + \ldots + b_kX_k$ be any two solutions to a homogeneous linear system written in terms of the basic solutions X_1, $\ldots$, X_k. Then their sum can be written

$$Y + Z = (a_1 + b_1)X_1 + \ldots + (a_k + b_k)X_k = c_1X_1 + \ldots c_kX_k,$$

where we have defined $c_i = a_i + b_i$ $(i = 1, \ldots, k)$, and hence the sum itself is a solution to the system as it is a linear combination of the basic solutions. Similarly the scalar product cY can be written

$$cY = (ca_1)X_1 + \ldots + (ca_k)X_k = d_1X_1 + \ldots d_kX_k,$$

where we have defined $d_i = ca_i$ $(i = 1, \ldots, k)$, and hence the scalar product is also a solution to the system as it is a linear combination of the basic solutions.[4]

More generally we have the following result.

Theorem 2-3: A linear combination of any solutions of a homogeneous linear system is also a solution to the system.

Note that the system has to be homogeneous for this property to hold.

[3] This can be a useful step to remove fractions from a basic solution.

[4] These two results can also be shown directly by considering two solutions $Y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$ and $Z = \begin{bmatrix} z_1 \\ \vdots \\ z_n \end{bmatrix}$ of a homogeneous linear system and plugging $Y + Z$ and kY into each homogeneous equation to see that they still hold.

We will often be interested in whether a matrix, particularly a column or row matrix, can be written as a linear combination of other matrices. This amounts to solving a linear system.

Example 2-21

Let $X = \begin{bmatrix} -1 \\ 2 \\ 3 \end{bmatrix}$ and $Y = \begin{bmatrix} 1 \\ 4 \\ 2 \end{bmatrix}$.

Write $V = \begin{bmatrix} -3 \\ 0 \\ 4 \end{bmatrix}$ as a linear combination of X and Y or show such a combination does not exist.

Solution:
For V to be a linear combination of X and Y we must find values for s and t such that

$$V = sX + tY .$$

This implies

$$\begin{bmatrix} -3 \\ 0 \\ 4 \end{bmatrix} = s \begin{bmatrix} -1 \\ 2 \\ 3 \end{bmatrix} + t \begin{bmatrix} 1 \\ 4 \\ 2 \end{bmatrix} = \begin{bmatrix} -s+t \\ 2s+4t \\ 3s+2t \end{bmatrix} .$$

Matrix equality of the first and last matrix implies we must solve the linear system:

$$\begin{aligned} -s + t &= -3 \\ 2s + 4t &= 0 \\ 3s + 2t &= 4 \end{aligned}$$

Reducing the corresponding augmented matrix gives

$$\left[\begin{array}{cc|c} -1 & 1 & -3 \\ 2 & 4 & 0 \\ 3 & 2 & 4 \end{array}\right] \Rightarrow \begin{array}{r} R_2 \to R_2 + 2R_1 \\ R_3 \to R_3 + 3R_1 \end{array} \left[\begin{array}{cc|c} -1 & 1 & -3 \\ 0 & 6 & -6 \\ 0 & 5 & -5 \end{array}\right] \Rightarrow \begin{array}{r} R_1 \to -R_1 \\ R_2 \to \frac{1}{6}R_2 \\ R_3 \to \frac{1}{5}R_3 \end{array} \left[\begin{array}{cc|c} 1 & -1 & 3 \\ 0 & 1 & -1 \\ 0 & 1 & -1 \end{array}\right] \Rightarrow \begin{array}{r} R_1 \to R_1 + R_2 \\ \\ R_3 \to R_3 - R_2 \end{array} \left[\begin{array}{cc|c} 1 & 0 & 2 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{array}\right]$$

The solution corresponding to the RREF is

$$s = 2,\ t = -1 ,$$

so $\boxed{V = 2X - 1Y}$. This is easily checked:

$$\begin{bmatrix} -3 \\ 0 \\ 4 \end{bmatrix} = 2 \begin{bmatrix} -1 \\ 2 \\ 3 \end{bmatrix} - 1 \begin{bmatrix} 1 \\ 4 \\ 2 \end{bmatrix} .$$

2.8 Matrix Multiplication

We now formally introduce how to multiply two matrices the motivation for which will be seen later.

Definition: Let $A = [a_{ij}]$ and $B = [b_{ij}]$ be two matrices with the number of columns of A being equal to the number of rows of B. Suppose A is an $m \times n$ matrix and B is an $n \times q$ matrix. Then the **matrix product** AB is an $m \times q$ matrix C, denoted $[c_{ij}]$ where the entry c_{ij} in i^{th} row and j^{th} column of the product, is found by multiplying each element in the i^{th} row of A with the corresponding element in the j^{th} column of B and then adding the products. In symbols

$$c_{ij} = [i^{\text{th}}\text{row of } A]\begin{bmatrix} j^{\text{th}} \\ \text{column} \\ \text{of} \\ B \end{bmatrix} = a_{i1}b_{1j} + a_{i2}b_{2j} + \ldots + a_{in}b_{nj}$$

Example 2-22

Find AB and BA if possible for the following pairs of matrices:

1.
$$A = \begin{bmatrix} 2 & 1 \\ -1 & 3 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 0 & 2 \\ 4 & -1 & 3 \end{bmatrix}$$

Solution:
A is 2×2 and B is 2×3 . So the product

$$A_{2\times \underset{\uparrow}{2}} B_{\underset{\uparrow}{2}\times 3} = C_{2\times 3}$$

is defined and is a 2×3 matrix. (The inner dimensions, indicated by $\uparrow$, are equal so the multiplication is possible and the outer dimensions are the dimensions of the new matrix.) Direct calculation of AB gives

$$\begin{aligned}
AB &= \begin{bmatrix} 2 & 1 \\ -1 & 3 \end{bmatrix}\begin{bmatrix} 1 & 0 & 2 \\ 4 & -1 & 3 \end{bmatrix} \\
&= \begin{bmatrix} \begin{bmatrix} 2 & 1 \end{bmatrix}\begin{bmatrix} 1 \\ 4 \end{bmatrix} & \begin{bmatrix} 2 & 1 \end{bmatrix}\begin{bmatrix} 0 \\ -1 \end{bmatrix} & \begin{bmatrix} 2 & 1 \end{bmatrix}\begin{bmatrix} 2 \\ 3 \end{bmatrix} \\ \\ \begin{bmatrix} -1 & 3 \end{bmatrix}\begin{bmatrix} 1 \\ 4 \end{bmatrix} & \begin{bmatrix} -1 & 3 \end{bmatrix}\begin{bmatrix} 0 \\ -1 \end{bmatrix} & \begin{bmatrix} -1 & 3 \end{bmatrix}\begin{bmatrix} 2 \\ 3 \end{bmatrix} \end{bmatrix} \\
&= \begin{bmatrix} (2)(1) + (1)(4) & (2)(0) + (1)(-1) & (2)(2) + (1)(3) \\ (-1)(1) + (3)(4) & (-1)(0) + (3)(-1) & (-1)(2) + (3)(3) \end{bmatrix} \\
&= \begin{bmatrix} 2 + 4 & 0 + (-1) & 4 + 3 \\ -1 + 12 & 0 + (-3) & -2 + 9 \end{bmatrix} \\
&= \begin{bmatrix} 6 & -1 & 7 \\ 11 & -3 & 7 \end{bmatrix} \quad \leftarrow (2 \times 3 \text{ matrix})
\end{aligned}$$

Notice the pattern in the second step where the i^{th} row of the first matrix multiplies the j^{th} of the second matrix term by term and these results are added together.

The product BA on the other hand does not exist

$$B_{2\times \underset{\uparrow}{3}} A_{\underset{\uparrow}{2}\times 2}$$

since the inner dimensions are not the same. (The i^{th} row of the first matrix has three entries which cannot multiply term by term the j^{th} column of the second matrix which has only two entries.)

2.

$$A = \begin{bmatrix} 1 & 2 & 4 \\ 2 & 6 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 4 & 1 & 4 & 3 \\ 0 & -1 & 3 & 1 \\ 2 & 7 & 5 & 2 \end{bmatrix}$$

Solution:
A is a 2×3 matrix and B is a 3×4 matrix and so the product AB exists and is a 2×4 matrix:

$$A_{2\times \underset{\uparrow}{3}} B_{\underset{\uparrow}{3}\times 4} = C_{2\times 4}$$

Matrix multiplication gives

$$\begin{aligned} AB &= \begin{bmatrix} 1 & 2 & 4 \\ 2 & 6 & 0 \end{bmatrix} \begin{bmatrix} 4 & 1 & 4 & 3 \\ 0 & -1 & 3 & 1 \\ 2 & 7 & 5 & 2 \end{bmatrix} \\ &= \begin{bmatrix} 4+0+8 & 1-2+28 & 4+6+20 & 3+2+8 \\ 8+0+0 & 2-6+0 & 8+18+0 & 6+6+0 \end{bmatrix} \\ &= \begin{bmatrix} 12 & 27 & 30 & 13 \\ 8 & -4 & 26 & 12 \end{bmatrix} \end{aligned}$$

The matrix product BA is not defined,

$$B_{3\times \underset{\uparrow}{4}} A_{\underset{\uparrow}{2}\times 3}$$

since the number of columns of B does not equal the number of rows of A.

3.

$$A = \begin{bmatrix} -3 & 0 \\ -1 & 2 \\ 1 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 6 & 1 & 3 \\ -1 & 1 & 4 \\ 4 & 1 & 0 \end{bmatrix}$$

Solution:
AB is not defined because the number of columns in A (2) does not equal the number of rows in B (3),

$$A_{3\times \underset{\uparrow}{2}} B_{\underset{\uparrow}{3}\times 3}\,.$$

The product BA is defined because the number of columns in B (3) equals the number of rows in A.

$$BA = B_{3\times \underset{\uparrow}{3}} A_{\underset{\uparrow}{3}\times 2} = \begin{bmatrix} 6 & 1 & 3 \\ -1 & 1 & 4 \\ 4 & 1 & 0 \end{bmatrix} \begin{bmatrix} -3 & 0 \\ -1 & 2 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} -16 & 5 \\ 6 & 6 \\ -13 & 2 \end{bmatrix}$$

The product BA has dimensions 3×2 as expected.

Note: When multiplying matrices it is helpful to proceed systematically by multiplying the first row times each of the columns to get the first row of the product, followed by the second row times each of the columns, to get the second row, etc.

2.9 Diagonal Matrices

Recall a square matrix is a matrix where the number of rows equals the number of columns, i.e. an $n \times n$ matrix.

Definition: The entries a_{ij} of a square matrix A for which $i = j$ form the **main diagonal** of A.

Example 2-23

$$A = \begin{bmatrix} 5 & 6 & -7 \\ -1 & \text{-2} & 3 \\ 0 & 4 & \text{-1} \end{bmatrix}$$

The main diagonal of A consists of the entries 5, -2, and -1 .

Definition: A square matrix in which every element not on the main diagonal is zero is called a **diagonal matrix**. A special type of diagonal matrix is the **identity matrix**, denoted by I, in which every entry on the diagonal is 1 .

Example 2-24

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

is a diagonal matrix, while

$$I = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

is the 3×3 identity matrix.

Note that if we wish to denote a particular identity matrix we can write I_m to represent the $m \times m$ identity matrix. So in Example 2-24 the identity matrix is I_3 .

2.10 Properties of Matrix Operations

We now collect some of the basic properties of matrices involving matrix addition, scalar multiplication, and matrix multiplication.

Theorem 2-4: Let A, B, and C be matrices and let a and b be scalars. Let I be an identity matrix and $\mathit{0}$ a zero matrix. Assuming that the dimensions of each of the matrices is such that the following operations is defined we have the following:

(1)	$A + B = B + A$	(commutative law for addition)
(2)	$(A + B) + C = A + (B + C)$	(associative law for addition)
(3)	$A + \mathit{0} = A$	
(4)	$A + (-A) = \mathit{0}$	
(5)	$A(BC) = (AB)C$	(associative law for multiplication)
(6)	$a(AB) = (aA)B = A(aB)$	
(7)	$(ab)C = a(bC) = b(aC)$	
(8)	$A(B + C) = AB + AC$	(left distributive law)
(9)	$(A + B)C = AC + BC$	(right distributive law)
(10)	$a(B + C) = aB + aC$	(scalar distributive law)
(11)	$(a + b)C = aC + bC$	(scalar distributive law)
(12)	$AI = A$	
(13)	$IB = B$	
(14)	$1A = A$	
(15)	$A\mathit{0} = \mathit{0}$	
(16)	$0B = \mathit{0}$	
(17)	$a\mathit{0} = \mathit{0}$	

As with regular numbers the associative laws of matrix addition and multiplication ensure it is meaningful to write $A + B + C$ and ABC without using parentheses.

The properties of Theorem 2-4 are analogous to the properties of real numbers. However **not all** real number properties correspond to matrix properties. We note the following

1. It is possible for AB to equal zero even if $A \neq \mathit{0}$ and $B \neq \mathit{0}$:

   Example 2-25

   If
   $$A = \begin{bmatrix} -1 & 1 \\ 2 & -2 \end{bmatrix}, \quad B = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$
   then
   $$AB = \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \mathit{0}\,,$$
   yet $A \neq \mathit{0}$ and $B \neq \mathit{0}$.

2. Even if $AB = AC$ with $A \neq \mathit{0}$, it may occur that B may not equal C.

Example 2-26

If

$$A = \begin{bmatrix} 1 & 2 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 1 \\ -1 & 0 \end{bmatrix}, \; C = \begin{bmatrix} 3 & -1 \\ -2 & 1 \end{bmatrix},$$

then

$$AB = \begin{bmatrix} -1 & 1 \end{bmatrix}, \quad AC = \begin{bmatrix} -1 & 1 \end{bmatrix}.$$

Therefore $AB = AC$ but $B \neq C$.

3. In general, even when it is defined, $AB \neq BA$. Matrix multiplication is *not commutative.*

Example 2-27

If

$$A = \begin{bmatrix} 1 & 1 \\ -1 & 0 \end{bmatrix}, \; B = \begin{bmatrix} 3 & 1 \\ 2 & 1 \end{bmatrix},$$

then

$$AB = \begin{bmatrix} 5 & 2 \\ -3 & -1 \end{bmatrix}, \quad BA = \begin{bmatrix} 2 & 3 \\ 1 & 2 \end{bmatrix}.$$

Therefore $AB \neq BA$.

2.10.1 Commutative Matrices

For the exceptional case that one can commute the product of two matrices and still get the same result one has the following definition.

Definition: A matrix A **commutes** with a matrix B if $AB = BA$.

Example 2-28

For matrices A and B defined by

$$A = \begin{bmatrix} -2 & 0 \\ 0 & 3 \end{bmatrix} \quad B = \begin{bmatrix} 4 & 0 \\ 0 & -1 \end{bmatrix},$$

show that A and B commute.

Solution:

$$AB = \begin{bmatrix} -2 & 0 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} 4 & 0 \\ 0 & -1 \end{bmatrix} = \begin{bmatrix} -8 & 0 \\ 0 & -3 \end{bmatrix}$$

$$BA = \begin{bmatrix} 4 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} -2 & 0 \\ 0 & 3 \end{bmatrix} = \begin{bmatrix} -8 & 0 \\ 0 & -3 \end{bmatrix}$$

Since $AB = BA$ the matrices A and B commute.

2.11 Matrix Equations

So far we have used equality involving matrices when assigning a variable to a matrix, or similarly with other identities when we convert a matrix to an equivalent matrix. In regular algebra one often creates equations involving variables and then seeks to find values of those variables that make the equation true (i.e. solve the equation). This can also be done with matrices where some or often all entries may be unknown. Matrix equality requires corresponding entries on both sides of a matrix equation must be equal; thus each entry generates an equation. These equations could then be solved for the unknowns. However it is usually quicker if we do operations on a matrix as a whole to solve for the unknowns.

Example 2-29

Suppose X is a matrix of unknowns and C and D are constant matrices defined by

$$X = \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix} \quad C = \begin{bmatrix} 1 & -2 \\ -1 & 3 \end{bmatrix} \quad D = \begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix}.$$

If the matrices satisfy the equation

$$2X - 4C = D\,,$$

find X.

Solution:
One *could* evaluate the left and right-hand sides directly as follows:

$$\begin{aligned} 2\begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix} - 4\begin{bmatrix} 1 & -2 \\ -1 & 3 \end{bmatrix} &= \begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix} \\ \begin{bmatrix} 2x_{11}-4 & 2x_{12}+8 \\ 2x_{21}+4 & 2x_{22}-12 \end{bmatrix} &= \begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix}. \end{aligned}$$

Then equating corresponding terms on both sides gives:

$$\begin{aligned} 2x_{11} - 4 &= 2 & \qquad 2x_{12} + 8 &= 0 \\ 2x_{21} + 4 &= 0 & \qquad 2x_{22} - 12 &= -2\,. \end{aligned}$$

Solving the equations one has:

$$\begin{aligned} x_{11} &= 3 & \qquad x_{12} &= -4 \\ x_{21} &= -2 & \qquad x_{22} &= 5\,, \end{aligned}$$

So $X = \begin{bmatrix} 3 & -4 \\ -2 & 5 \end{bmatrix}$. However it is more useful just to work with the matrices as a whole as we would in a regular equation, using inverse operations to isolate the matrix variable:

$$\begin{aligned} 2X - 4C &= D \\ 2X &= D + 4C \\ X &= \frac{1}{2}(D + 4C) \\ X &= \frac{1}{2}D + 2C \\ X &= \frac{1}{2}\begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix} + 2\begin{bmatrix} 1 & -2 \\ -1 & 3 \end{bmatrix} \\ X &= \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} + \begin{bmatrix} 2 & -4 \\ -2 & 6 \end{bmatrix} \\ X &= \begin{bmatrix} 3 & -4 \\ -2 & 5 \end{bmatrix}. \end{aligned}$$

2.12 Transpose of a Matrix

Definition: The **transpose** of an $m \times n$ matrix A, denoted by A^T, is the $n \times m$ matrix whose j^{th} column is the j^{th} row of A. In symbols

$$(A^T)_{ij} = A_{ji} \ .$$

Note: To find A^T one interchanges the rows and columns of A.

Example 2-30

Find the transpose of the following matrices:

$$A = \begin{bmatrix} -1 & 2 \\ 3 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 2 & 1 & 3 \\ -1 & 0 & 5 \end{bmatrix} .$$

Solution:
Exchanging rows and columns on has:

$$A^T = \begin{bmatrix} -1 & 3 \\ 2 & 1 \end{bmatrix}, \quad B^T = \begin{bmatrix} 2 & -1 \\ 1 & 0 \\ 3 & 5 \end{bmatrix} .$$

2.12.1 Properties of the Transpose

Theorem 2-5: Let A and B be matrices of dimensions such that the following operations are defined and b a scalar, then the transpose has the following properties:

(1) $$(A + B)^T = A^T + B^T$$

(2) $$(AB)^T = B^T A^T$$

(3) $$(A^T)^T = A$$

(4) $$(bA)^T = bA^T$$

Example 2-31

Given the matrices

$$A = \begin{bmatrix} 1 & 2 & 0 \\ 3 & 0 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 2 \\ 0 & 1 \\ -2 & 4 \end{bmatrix}$$

compute AB , $(AB)^T$, $A^T B^T$ and $B^T A^T$.

Solution:

$$AB = \begin{bmatrix} 1 & 2 & 0 \\ 3 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 0 & 1 \\ -2 & 4 \end{bmatrix} = \begin{bmatrix} 1 & 4 \\ 1 & 10 \end{bmatrix}$$

$$(AB)^T = \begin{bmatrix} 1 & 1 \\ 4 & 10 \end{bmatrix}$$

$$A^T B^T = \begin{bmatrix} 1 & 3 \\ 2 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & -2 \\ 2 & 1 & 4 \end{bmatrix} = \begin{bmatrix} 7 & 3 & 10 \\ 2 & 0 & -4 \\ 2 & 1 & 4 \end{bmatrix}$$

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

So $(AB)^T = B^T A^T$ as predicted by the theorem. It does not equal $A^T B^T$.

The transpose property of a product can be generalized to more than two matrices.

Example 2-32

Prove that $(ABC)^T = C^T B^T A^T$

Proof:

$$(ABC)^T = ((AB)C)^T = C^T (AB)^T = C^T (B^T A^T) = C^T B^T A^T$$

Generalizing to k matrices we have the following result.

Theorem 2-6: Let A_1, A_2, …, A_k be matrices for which the product $A_1 A_2 \cdots A_k$ is defined, then

$$(A_1 A_2 \cdots A_k)^T = A_k^T \cdots A_2^T A_1^T .$$

2.12.2 Symmetric Matrices

Definition: A square matrix , $A = [a_{ij}]$, is called **symmetric** if $a_{ij} = a_{ji}$ for all i and j. A square matrix is called **skew symmetric** if $a_{ij} = -a_{ji}$ for all i and j.

Example 2-33

The following matrices are symmetric:

$$A = \begin{bmatrix} 1 & 2 \\ 2 & 3 \end{bmatrix}, \quad B = \begin{bmatrix} 2 & -1 & 3 \\ -1 & 5 & 0 \\ 3 & 0 & -7 \end{bmatrix}.$$

The following matrices are skew symmetric:

$$C = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 0 & -1 & 2 \\ 1 & 0 & 1 \\ -2 & -1 & 0 \end{bmatrix}.$$

Theorem 2-7: A square matrix $A = [a_{ij}]$ is symmetric if and only if $A^T = A$. It is skew symmetric if and only if $A^T = -A$.

2.13 Power of a Matrix

Definition: Let A be a square matrix, then the n^{th} power of A, denoted A^n, is the product of n factors of A, i.e.

$$\begin{aligned} A^1 &= A \\ A^2 &= AA \\ A^3 &= AAA \\ &\vdots \\ A^n &= \underbrace{AA\cdots A}_{n \text{ times}} \end{aligned}$$

Example 2-34

If

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

find A^2.

Solution:

$$A^2 = \begin{bmatrix} 1 & -1 & 3 \\ 0 & 1 & -2 \\ -1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & -1 & 3 \\ 0 & 1 & -2 \\ -1 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 1+0-3 & -1-1+3 & 3+2+3 \\ 0+0+2 & 0+1-2 & 0-2-2 \\ -1+0-1 & 1+1+1 & -3-2+1 \end{bmatrix}$$

$$A^2 = \begin{bmatrix} -2 & 1 & 8 \\ 2 & -1 & -4 \\ -2 & 3 & -4 \end{bmatrix}$$

Having defined the power of a matrix, it is now possible to create meaningful polynomial functions of a square matrix such as

$$p(A) = c_0 I + c_1 A + c_2 A^2 + \ldots + c_n A^n$$

for some positive integer n and the identity matrix I of the same dimension of A, and scalars c_i.[5] In more advanced courses we could similarly consider a power series in A where we let $n \to \infty$. Then questions of the meaning of the convergence of such a series needs to be addressed, just as with power series in terms of real variable x.

2.13.1 Idempotent Matrices

Definition: Matrix A is called **idempotent** if $A^2 = A$.

Example 2-35

Show that the matrix A defined by

$$A = \begin{bmatrix} 4 & -1 \\ 12 & -3 \end{bmatrix}$$

is idempotent.

[5]Note some texts will define $A^0 = I$ (analogous to $x^0 = 1$) so one may write $p(A) = c_0 A^0 + c_1 A^1 + c_2 A^2 + \ldots + c_n A^n$.

Solution:

$$A^2 = AA = \begin{bmatrix} 4 & -1 \\ 12 & -3 \end{bmatrix} \begin{bmatrix} 4 & -1 \\ 12 & -3 \end{bmatrix} = \begin{bmatrix} 4 & -1 \\ 12 & -3 \end{bmatrix} = A$$

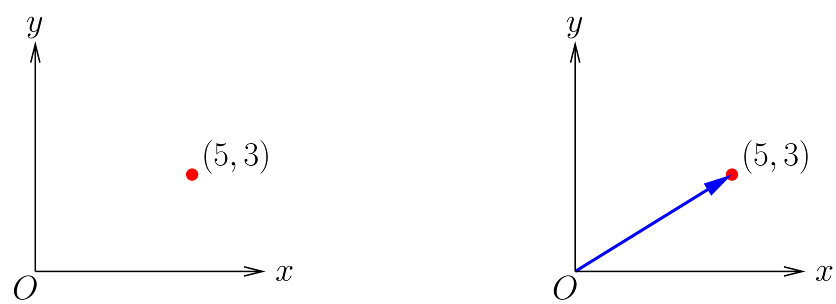
2.14 Ordered n-tuples

Definition: The ordered sequence $(x_1, x_2, \ldots, x_n)$ where x_i are real numbers and n is a positive integer is called an **ordered n-tuple**. The set of all ordered n-tuples is called **n-space** and is denoted by $\mathbb{R}^n$.

Example 2-36

- $(5, 3)$ is an ordered 2-tuple in $\mathbb{R}^2$. Note that it is different from $(3, 5)$ (order matters).
- $(1, -2, 4)$ is an ordered 3-tuple in $\mathbb{R}^3$.
- $(-3, 2.1, 7, 4, -9.5)$ is an ordered 5-tuple in $\mathbb{R}^5$.

As the last example suggests, an obvious geometric interpretation presents itself. We can think of the ordered 2-tuple $(5, 3)$ as representing a **point** in the two-dimensional coordinate plane. Alternatively we can consider it as representing a directed line segment (an arrow), called a **vector**, originating at the origin of the coordinate system and terminating at the point.



Similarly $(1, -2, 4)$ could be considered a representation of a point or vector in three-dimensional space, while an ordered n-tuple of higher dimension can be thought of as a generalized point or n-vector in some higher-dimensional space. In future we will tend to use a vector interpretation and typically refer to ordered n-tuples as vectors. However vectors, as will be discussed further in Chapter 4, are constructions that are conceptually independent of a particular set of coordinates (n-tuple) used to represent them.

2.14.1 Notation

We will use lower case boldface letters such as $\boldsymbol{x}$ to represent the ordered n-tuple $(x_1, x_2, \ldots, x_n)$, or vector. For example, $\boldsymbol{x} = (5, 3)$. When hand-written it is more common to put an arrow or bar over the letter, such as $\vec{x}$, or $\bar{x}$. When speaking of the ordered n-tuple containing all zeros we will write $\mathbf{0}$. It is convenient, when representing an ordered n-tuple (vector) by a matrix, to use an **$n \times 1$ column matrix**

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix},$$

We can now write $A\boldsymbol{x}$, where previously we would have written AX, provided A is an $m \times n$ matrix so the multiplication is meaningful. If we wish to represent a vector within a sentence as a matrix we can conveniently use the transpose, since then $\boldsymbol{x} = [x_1\, x_2\, \ldots\, x_n]^T$.

The motivation for considering a vector to be represented by a column matrix arises, in part, from linear systems. Recall the linear system of m equations in n unknowns is:

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

By assigning the coefficient constants to A as before, the unknowns to vector $\boldsymbol{x}$ and the right-hand side constants to vector $\boldsymbol{b}$ as follows

$$A = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{bmatrix}, \; \boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \; \boldsymbol{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix},$$

we now see that we can represent the linear system in terms of matrix multiplication in terms of the simple matrix equation:

$$A\boldsymbol{x} = \boldsymbol{b} \; .$$

This follows since multiplying the left hand side out gives precisely the $m \times 1$ column matrix

$$A\boldsymbol{x} = \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n \end{bmatrix}.$$

The matrix equality of this with the right-hand side matrix $\boldsymbol{b} = [b_1, b_2, \ldots b_m]^T$ gives back our original linear system of equations. This shows a clear advantage for defining matrix multiplication the way that we have done.

Example 2-37

The linear system

$$\begin{aligned} 3x_1 + 2x_3 &= 7 \\ x_1 + 4x_2 - 4x_3 &= 3 \\ 2x_1 + 2x_2 + 8x_3 &= 1 \end{aligned}$$

can be represented by the matrix equation $A\boldsymbol{x} = \boldsymbol{b}$ where

$$A = \begin{bmatrix} 3 & 0 & 2 \\ 1 & 4 & -4 \\ 3 & 3 & 8 \end{bmatrix}, \quad \boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, \text{ and } \boldsymbol{b} = \begin{bmatrix} 7 \\ 3 \\ 1 \end{bmatrix}.$$

2.14.2 Matrix Multiplying a Vector

We can consider the $m \times n$ matrix A to be composed of n columns of vectors in m-space labelled $\boldsymbol{a}_1$, $\boldsymbol{a}_2$, ..., $\boldsymbol{a}_n$. These are the **column vectors** of A. We write

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} = [\boldsymbol{a}_1 \, \boldsymbol{a}_2 \, \cdots \, \boldsymbol{a}_n] \, .$$

Cast in this form we see that our previous product $A\boldsymbol{x}$ can be written as a sum of $m \times 1$ matrices (vectors), namely

$$A\boldsymbol{x} = \begin{bmatrix} a_{11}x_1 \\ a_{21}x_1 \\ \vdots \\ a_{m1}x_1 \end{bmatrix} + \begin{bmatrix} a_{12}x_2 \\ a_{22}x_2 \\ \vdots \\ a_{m2}x_2 \end{bmatrix} + \ldots + \begin{bmatrix} a_{1n}x_n \\ a_{2n}x_n \\ \vdots \\ a_{mn}x_n \end{bmatrix} = x_1\boldsymbol{a}_1 + x_2\boldsymbol{a}_2 + \ldots + x_n\boldsymbol{a}_n \, .$$

We have the following result.

Theorem 2-8: If $A = [\boldsymbol{a}_1 \, \boldsymbol{a}_2 \, \ldots \, \boldsymbol{a}_n]$ is an $m \times n$ matrix written in terms of its column vectors $\boldsymbol{a}_i$, then the matrix product of A times the vector $\boldsymbol{x} = [x_1 \, x_2 \, \ldots \, x_n]^T$ in $\mathbb{R}^n$ can be written

$$A\boldsymbol{x} = x_1\boldsymbol{a}_1 + x_2\boldsymbol{a}_2 + \ldots + x_n\boldsymbol{a}_n \, .$$

The matrix equation $A\boldsymbol{x} = \boldsymbol{b}$ takes the new vector equation form

$$x_1\boldsymbol{a}_1 + x_2\boldsymbol{a}_2 + \ldots + x_n\boldsymbol{a}_n = \boldsymbol{b} \, .$$

From this follows the following theorem.

Theorem 2-9: The linear system $A\boldsymbol{x} = \boldsymbol{b}$ is consistent if and only if $\boldsymbol{b}$ can be written as a linear combination of the columns of matrix A.

Example 2-38

Find a vector equation involving only constant vectors that is equivalent to the linear system

$$\begin{aligned} 3x_1 + 2x_3 &= 7 \\ x_1 + 4x_2 - 4x_3 &= 3 \\ 2x_1 + 2x_2 + 8x_3 &= 1 \end{aligned}$$

Solution:
The above linear system is equivalent to the vector equation

$$\begin{bmatrix} 3x_1 + 2x_3 \\ x_1 + 4x_2 - 4x_3 \\ 2x_1 + 2x_2 + 8x_3 \end{bmatrix} = \begin{bmatrix} 7 \\ 3 \\ 1 \end{bmatrix} \, .$$

An equivalent vector equation is

$$\begin{bmatrix} 3x_1 \\ x_1 \\ 2x_1 \end{bmatrix} + \begin{bmatrix} 0x_2 \\ 4x_2 \\ 2x_2 \end{bmatrix} + \begin{bmatrix} 2x_3 \\ -4x_3 \\ 8x_3 \end{bmatrix} = \begin{bmatrix} 7 \\ 3 \\ 1 \end{bmatrix} \, .$$

Finally factoring out the variables one has:

$$x_1 \begin{bmatrix} 3 \\ 1 \\ 2 \end{bmatrix} + x_2 \begin{bmatrix} 0 \\ 4 \\ 2 \end{bmatrix} + x_3 \begin{bmatrix} 2 \\ -4 \\ 8 \end{bmatrix} = \begin{bmatrix} 7 \\ 3 \\ 1 \end{bmatrix} .$$

This result also follows directly from our previous formula

$$x_1 \boldsymbol{a}_1 + x_2 \boldsymbol{a}_2 + \ldots + x_n \boldsymbol{a}_n = \boldsymbol{b},$$

since

$$A = [\boldsymbol{a}_1 \boldsymbol{a}_2 \boldsymbol{a}_3] = \begin{bmatrix} 3 & 0 & 2 \\ 1 & 4 & -4 \\ 2 & 2 & 8 \end{bmatrix} \text{ and } \boldsymbol{b} = \begin{bmatrix} 7 \\ 3 \\ 1 \end{bmatrix} .$$

2.14.3 General Solution of a Linear System

Consider a general linear system $A\boldsymbol{x} = \boldsymbol{b}$. If the system is consistent then there exists at least one particular solution $\boldsymbol{x}_p$ of the system. If $\boldsymbol{x}$ is any other solution of the system then the vector difference $\boldsymbol{x}_0 = \boldsymbol{x} - \boldsymbol{x}_p$ is a solution of the **associated homogeneous system** $A\boldsymbol{x} = \boldsymbol{0}$ since

$$A\boldsymbol{x}_0 = A(\boldsymbol{x} - \boldsymbol{x}_p) = A\boldsymbol{x} - A\boldsymbol{x}_p = \boldsymbol{b} - \boldsymbol{b} = \boldsymbol{0}.$$

Furthermore any vector $\boldsymbol{x} = \boldsymbol{x}_p + \boldsymbol{x}_0$ where $\boldsymbol{x}_0$ is any solution of the associated homogeneous system is a solution of the original linear system since

$$A\boldsymbol{x} = A(\boldsymbol{x}_p + \boldsymbol{x}_0) = A\boldsymbol{x}_p + A\boldsymbol{x}_0 = \boldsymbol{b} + \boldsymbol{0} = \boldsymbol{b}.$$

We have the following result

Theorem 2-10: For any consistent linear system $A\boldsymbol{x} = \boldsymbol{b}$ the general solution can be written in the form

$$\boldsymbol{x} = \boldsymbol{x}_p + \boldsymbol{x}_0$$

where $\boldsymbol{x}_p$ is a particular solution of the linear system and $\boldsymbol{x}_0$ is any solution of the associated homogeneous system, i.e. $A\boldsymbol{x}_0 = \boldsymbol{0}$.

As such one approach to solving a non-homogeneous linear system is to find a particular solution and then add to it the general solution of the associated homogeneous system. This is a pattern that arises in other contexts such as solving differential equations. In practice for us, the separation of the solution of a consistent linear system into its particular solution plus a general solution to the homogeneous system (which may involve parameters) arises straight from Gauss-Jordan elimination.

Example 2-39

Express all solutions of the following system as a sum of a particular solution plus a solution of the associated homogeneous system.

$$\begin{aligned} x_3 + 2x_4 - x_5 &= 4 \\ x_4 - x_5 &= 3 \\ x_3 + 3x_4 - 2x_5 &= 7 \\ 2x_1 + 4x_2 + x_3 + 7x_4 &= 7 \end{aligned}$$

Solution:
The augmented matrix

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

reduces via Gauss-Jordan elimination to the RREF and corresponding equivalent linear system

$$\left[\begin{array}{ccccc|c} 1 & 2 & 0 & 0 & 3 & -6 \\ 0 & 0 & 1 & 0 & 1 & -2 \\ 0 & 0 & 0 & 1 & -1 & 3 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right] \Leftrightarrow \begin{aligned} x_1 + 2x_2 + 3x_5 &= -6 \\ x_3 + x_5 &= -2 \\ x_4 - x_5 &= 3 \\ 0 &= 0 \end{aligned}$$

The general solution is found by setting free variables x_2 and x_5 to parameters, $\boxed{x_2 = s}$ and $\boxed{x_5 = t}$, and solving by back-substitution the remaining (leading) variables:

- $x_4 - x_5 = 3 \implies x_4 - t = 3 \implies \boxed{x_4 = 3 + t}$
- $x_3 + x_5 = -2 \implies x_3 + t = -2 \implies \boxed{x_3 = -2 - t}$
- $x_1 + 2x_2 + 3x_5 = -6 \implies x_1 + 2s + 3t = -6 \implies \boxed{x_1 = -6 - 2s - 3t}$.

Written as a vector the solution is

$$\boldsymbol{x} = \begin{bmatrix} -6-2s-3t \\ s \\ -2-t \\ 3+t \\ t \end{bmatrix} = \begin{bmatrix} -6 \\ 0 \\ -2 \\ 3 \\ 0 \end{bmatrix} + \begin{bmatrix} -2s \\ s \\ 0 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} -3t \\ 0 \\ -t \\ t \\ t \end{bmatrix} = \underbrace{\begin{bmatrix} -6 \\ 0 \\ -2 \\ 3 \\ 0 \end{bmatrix}}_{\boldsymbol{x}_p} + \underbrace{s\begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + t\begin{bmatrix} -3 \\ 0 \\ -1 \\ 1 \\ 1 \end{bmatrix}}_{\boldsymbol{x}_0} = \boldsymbol{x}_p + \boldsymbol{x}_0$$

As seen in Example 2-20 the vector $\boldsymbol{x}_0$ which involves the parameters is indeed the general solution of the homogeneous system.

Note that the particular solution $\boldsymbol{x}_p$ is not unique when there are parameters. In this problem we could introduce new parameters $\tilde{s}$ and $\tilde{t}$ by the substitutions $s = \tilde{s} + 1$ and $t = \tilde{t} - 2$ and the new particular solution would be, after collecting all the constants in one vector, $\boldsymbol{x}_p = [-2\ 1\ 0\ 1\ -2]^T$. The homogeneous solution with that substitution would have the same form with s and t replaced by $\tilde{s}$ and $\tilde{t}$.

For linear systems with a unique solution (no parameters) $\boldsymbol{x}_p$ will be that unique solution and the solution to the associated homogeneous system will just be $\boldsymbol{x}_0 = \mathbf{0}$.

2.15 Matrix Inversion

We have seen we can write a linear system $A\boldsymbol{x} = \boldsymbol{b}$. If we had the algebraic problem $ax = b$ where a and b were just real constants and x were a variable we would just divide by a (assuming it was non-zero), or equivalently multiply by a^{-1} on both sides to get $(a)^{-1}ax = (a)^{-1}b$, or just $x = (a)^{-1}b$. This raises the question of whether a *matrix* multiplicative inverse can be found that we could similarly left-multiply to solve our matrix equation. Even the algebra problem, however, suggests this may not always be possible since $a = 0$ has no multiplicative inverse.

Definition: Let A be an $n \times n$ matrix. The **inverse of $\boldsymbol{A}$** is an $n \times n$ matrix denoted A^{-1} satisfying:

$$AA^{-1} = A^{-1}A = I$$

where I is the $n \times n$ identity matrix.

Definition: If A^{-1} exists, we say A is **invertible** (or **non-singular**). If A does not have an inverse it is said to **noninvertible** or **singular**.

Example 2-40

If $A = \begin{bmatrix} 7 & 2 \\ 3 & 1 \end{bmatrix}$ show $A^{-1} = \begin{bmatrix} 1 & -2 \\ -3 & 7 \end{bmatrix}$ is an inverse of A.

Solution:
We have by direct calculation:

$$AA^{-1} = \begin{bmatrix} 7 & 2 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} 1 & -2 \\ -3 & 7 \end{bmatrix} = \begin{bmatrix} 7-6 & -14+14 \\ 3-3 & -6+7 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I\,,$$

$$A^{-1}A = \begin{bmatrix} 1 & -2 \\ -3 & 7 \end{bmatrix} \begin{bmatrix} 7 & 2 \\ 3 & 1 \end{bmatrix} = \begin{bmatrix} 7-6 & 2-2 \\ -21+21 & -6+7 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I\,.$$

Thus A^{-1} is an inverse of invertible matrix A.

Properties of the Inverse

Theorem 2-11: If an $n \times n$ matrix A has an inverse then that inverse is unique.

Proof:
Suppose matrix A has inverse A^{-1} and that B is another inverse of A then, by definition,

$$AA^{-1} = A^{-1}A = I$$

and

$$AB = BA = I\,.$$

Using these properties and the associativity of matrix multiplication (Theorem 2-4) we have:

$$B = BI = B(AA^{-1}) = (BA)(A^{-1}) = IA^{-1} = A^{-1}\,.$$

Therefore $B = A^{-1}$ and the inverse is unique.

Theorem 2-12: If an $n \times n$ matrix A is invertible then A^{-1} is invertible and the inverse of A^{-1} is A. In symbols:

$$(A^{-1})^{-1} = A\,.$$

Proof:
The defining property $AA^{-1} = A^{-1}A = I$ implies that left and right multiplying A^{-1} by A gives the identity I.

Theorem 2-13: If A is an $n \times n$ invertible matrix, and $c \neq 0$ is a scalar then

$$(cA)^{-1} = \frac{1}{c}A^{-1}$$

Proof:
We must show that

$$(cA)\left(\frac{1}{c}A^{-1}\right) = \left(\frac{1}{c}A^{-1}\right)(cA) = I$$

Using the properties of scalar and matrix multiplication (Theorem 2-4) we have:

$$(cA)\left(\frac{1}{c}A^{-1}\right) = c\left[A\left(\frac{1}{c}A^{-1}\right)\right] = c\left[\frac{1}{c}\left(AA^{-1}\right)\right] = c\left(\frac{1}{c}I\right) = \left(c\frac{1}{c}\right)I = 1I = I\,,$$

Similarly:

$$\left(\frac{1}{c}A^{-1}\right)(cA) = \frac{1}{c}\left[A^{-1}(cA)\right] = \frac{1}{c}\left[c\left(A^{-1}A\right)\right] = \frac{1}{c}(cI) = \left(\frac{1}{c}c\right)I = 1I = I\,.$$

Thus $(cA)^{-1} = \dfrac{1}{c}A^{-1}$.

Theorem 2-14: If A and B are $n \times n$ invertible matrices, then the product AB is also invertible with

$$(AB)^{-1} = B^{-1}A^{-1}\,.$$

Proof:
A and B invertible imply A^{-1} and B^{-1} exist satisfying

$$AA^{-1} = A^{-1}A = I\,,$$

$$BB^{-1} = B^{-1}B = I\,.$$

We must show that:

$$(AB)(B^{-1}A^{-1}) = (B^{-1}A^{-1})(AB) = I\,.$$

We have:

$$(AB)(B^{-1}A^{-1}) = A(BB^{-1})A^{-1} = AIA^{-1} = AA^{-1} = I\,.$$

Similarly

$$(B^{-1}A^{-1})(AB) = B^{-1}(A^{-1}A)B = B^{-1}IB = B^{-1}B = I\,.$$

Thus $(AB)^{-1} = B^{-1}A^{-1}$.

The previous theorem generalizes to a product of k matrices.

Theorem 2-15: If $A_1, A_2, \ldots, A_k$ are $n \times n$ invertible matrices then the product $A_1A_2 \cdots A_k$ is also invertible with

$$(A_1A_2\cdots A_k)^{-1} = A_k^{-1}\cdots A_2^{-1}A_1^{-1}\,.$$

Theorem 2-16: If A is a $n \times n$ invertible matrix, then A^T is also invertible with

$$(A^T)^{-1} = (A^{-1})^T .$$

Proof:
A is invertible so A^{-1} exists satisfying

$$AA^{-1} = A^{-1}A = I .$$

We must show that:

$$A^T(A^{-1})^T = (A^{-1})^T A^T = I$$

We have:

$$\begin{aligned} A^T(A^{-1})^T &= (A^{-1}A)^T = I^T = I , \\ (A^{-1})^T A^T &= (AA^{-1})^T = I^T = I . \end{aligned}$$

Therefore $(A^T)^{-1} = (A^{-1})^T$.

Example 2-41

Simplify $(AB)^{-1}(AB^{-1})(BA^T)(DA^{-1})^T$.

Solution:

$$\begin{aligned} (AB)^{-1}(AB^{-1})(BA^T)(DA^{-1})^T &= (B^{-1}A^{-1})(AB^{-1})(BA^T)((A^{-1})^T D^T) \\ &= B^{-1}(A^{-1}A)(B^{-1}B)(A^T(A^T)^{-1})D^T \\ &= B^{-1}(I)(I)(I)D^T \\ &= B^{-1}D^T \end{aligned}$$

2.15.1 Orthogonal Matrices

A special class of square matrices is defined in terms of the properties of its inverse.

Definition: A square matrix A is **orthogonal** if $\boxed{A^{-1} = A^T}$.

Example 2-42

The square matrix $A = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$ is orthogonal since $A^T = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$ and

$$AA^T = \left(\frac{1}{\sqrt{2}}\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}\right)\left(\frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}\right) = \frac{1}{2}\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I .$$

Similarly $A^T A = I$ and thus $A^{-1} = A^T$.

2.15.2 Finding the Inverse

The following example illustrates how to find the inverse using row operations.

Example 2-43

Find the inverse of $A = \begin{bmatrix} -1 & 2 \\ -3 & 5 \end{bmatrix}$.

Solution:
This solution needs expansion (as per the in-class example) and show the row operations.

If A^{-1} exists, then $AA^{-1} = A^{-1}A = I$
Assume that A^{-1} exists, given by $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$

$AA^{-1} = I$ implies

$$\begin{bmatrix} -1 & 2 \\ -3 & 5 \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$\begin{bmatrix} -a + 2c & -b + 2d \\ -3a + 5c & -3b + 5d \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Matrix equality gives four equations in the four unknowns (a, b, c, and d). However the two equations arising from the first column involve only a and c and the two equations arising from the second column involve only b and d. Therefore we really need to solve two linear systems involving two unknowns, one per column. The first column system solution is as follows.

$$\begin{aligned} -a + 2c &= 1 \\ -3a + 5c &= 0 \end{aligned} \quad \Leftrightarrow \quad \left[\begin{array}{cc|c} -1 & 2 & 1 \\ -3 & 5 & 0 \end{array}\right]$$

$$\left[\begin{array}{cc|c} -1 & 2 & 1 \\ -3 & 5 & 0 \end{array}\right] \Rightarrow \begin{array}{r} R_1 \to -R_1 \\ R_2 \to R_2 - 3R_1 \end{array} \left[\begin{array}{cc|c} 1 & -2 & -1 \\ 0 & -1 & -3 \end{array}\right] \Rightarrow \begin{array}{r} R_1 \to R_1 - 2R_2 \\ R_2 \to -R_2 \end{array} \left[\begin{array}{cc|c} 1 & 0 & 5 \\ 0 & 1 & 3 \end{array}\right] \Rightarrow \begin{array}{c} a = 5 \\ c = 3 \end{array}$$

Similarly the second column system and solution is

$$\begin{aligned} -b + 2d &= 0 \\ -3b + 5d &= 1 \end{aligned} \quad \Leftrightarrow \quad \left[\begin{array}{cc|c} -1 & 2 & 0 \\ -3 & 5 & 1 \end{array}\right]$$

$$\left[\begin{array}{cc|c} -1 & 2 & 0 \\ -3 & 5 & 1 \end{array}\right] \Rightarrow \begin{array}{r} R_1 \to -R_1 \\ R_2 \to R_2 - 3R_1 \end{array} \left[\begin{array}{cc|c} 1 & -2 & 0 \\ 0 & -1 & 1 \end{array}\right] \Rightarrow \begin{array}{r} R_1 \to R_1 - 2R_2 \\ R_2 \to -R_2 \end{array} \left[\begin{array}{cc|c} 1 & 0 & -2 \\ 0 & 1 & -1 \end{array}\right] \Rightarrow \begin{array}{c} b = -2 \\ d = -1 \end{array}$$

Therefore $A^{-1} = \begin{bmatrix} 5 & -2 \\ 3 & -1 \end{bmatrix}$. Comparison of the solutions for the two systems shows that the row operations to solve both systems will depend, in the event there is a unique solution, entirely on the coefficient matrix. This suggests solving both systems simultaneously using the augmented matrix $[A|I]$ where I is the identity matrix:

$$[A|I] = \left[\begin{array}{cc|cc} -1 & 2 & 1 & 0 \\ -3 & 5 & 0 & 1 \end{array}\right]$$

Reducing to RREF gives

$$\left[\begin{array}{cc|cc} -1 & 2 & 1 & 0 \\ -3 & 5 & 0 & 1 \end{array}\right] \Rightarrow \begin{array}{r} R_1 \to -R_1 \\ R_2 \to R_2 - 3R_1 \end{array} \left[\begin{array}{cc|cc} 1 & -2 & -1 & 0 \\ 0 & -1 & -3 & 1 \end{array}\right] \Rightarrow \begin{array}{r} R_1 \to R_1 - 2R_2 \\ R_2 \to -R_2 \end{array} \left[\begin{array}{cc|cc} 1 & 0 & 5 & -2 \\ 0 & 1 & 3 & -1 \end{array}\right]$$

We see the inverse, after the operations have been completed, is the right-hand side of the final augmented matrix which has the form $[I|A^{-1}]$. Thus

$$A^{-1} = \begin{bmatrix} 5 & -2 \\ 3 & -1 \end{bmatrix}$$

In general we see that to find the inverse we do the reduction $[A|I] \Rightarrow [I|A^{-1}]$. If the left hand side cannot reduce to I, then no inverse exists.

Steps for Finding the Inverse of a Matrix:

1. Write the identity matrix (of the same dimension as A) adjacent to the matrix A to form an augmented matrix $[A|I]$.
2. Perform row operations on this augmented matrix until the matrix that was A is reduced to the identity matrix if possible.
$$[I|A^{-1}]$$
3. The matrix in the position of the original identity matrix is A's inverse A^{-1}.

In summary $[A|I] \Rightarrow [I|A^{-1}]$. If the procedure cannot be successfully completed then A is noninvertible.

The previous steps can be applied once and for all to the 2×2 matrix $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ to find the following inverse which can then be confirmed by direct matrix multiplication.

Theorem 2-17: If $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is a 2×2 matrix with $ad - bc \neq 0$, then A^{-1} exists and

$$A^{-1} = \frac{1}{(ad-bc)} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

Example 2-44

Find A^{-1} if $A = \begin{bmatrix} 2 & 5 \\ -3 & 7 \end{bmatrix}$.

Solution:
Since $a = 2$, $b = 5$, $c = -3$, and $d = 7$ we have

$$ad - bc = 2(7) - 5(-3) = 29$$

which is nonzero so the inverse matrix A^{-1} exists and equals

$$A^{-1} = \frac{1}{(ad-bc)} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} = \frac{1}{29} \begin{bmatrix} 7 & -5 \\ 3 & 2 \end{bmatrix} = \begin{bmatrix} \frac{7}{29} & -\frac{5}{29} \\ \frac{3}{29} & \frac{2}{29} \end{bmatrix}$$

Example 2-45

Find the inverse of $A = \begin{bmatrix} 2 & 1 & 0 \\ -4 & -1 & -3 \\ 3 & 1 & 2 \end{bmatrix}$.

Solution:
We reduce $[A|I]$ to RREF:

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

$$\Downarrow$$

$$\begin{array}{r} R_2 \to R_2 + 2R_1 \\ R_3 \to R_3 - \frac{3}{2}R_1 \end{array} \left[\begin{array}{rrr|rrr} 2 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & -3 & 2 & 1 & 0 \\ 0 & -\frac{1}{2} & 2 & -\frac{3}{2} & 0 & 1 \end{array}\right]$$

$$\Downarrow$$

$$R_3 \to R_3 + \tfrac{1}{2}R_2 \left[\begin{array}{rrr|rrr} 2 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & -3 & 2 & 1 & 0 \\ 0 & 0 & \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} & 1 \end{array}\right]$$

$$\Downarrow$$

$$\begin{array}{r} R_1 \to \frac{1}{2}R_1 \\ \\ R_3 \to 2R_3 \end{array} \left[\begin{array}{rrr|rrr} 1 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\ 0 & 1 & -3 & 2 & 1 & 0 \\ 0 & 0 & 1 & -1 & 1 & 2 \end{array}\right]$$

$$\Downarrow$$

$$R_2 \to R_2 + 3R_3 \left[\begin{array}{rrr|rrr} 1 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\ 0 & 1 & 0 & -1 & 4 & 6 \\ 0 & 0 & 1 & -1 & 1 & 2 \end{array}\right]$$

$$\Downarrow$$

$$R_1 \to R_1 - \tfrac{1}{2}R_2 \left[\begin{array}{rrr|rrr} 1 & 0 & 0 & 1 & -2 & -3 \\ 0 & 1 & 0 & -1 & 4 & 6 \\ 0 & 0 & 1 & -1 & 1 & 2 \end{array}\right] = [I|A^{-1}]$$

Therefore A is invertible with

$$A^{-1} = \begin{bmatrix} 1 & -2 & -3 \\ -1 & 4 & 6 \\ -1 & 1 & 2 \end{bmatrix}.$$

Example 2-46

Find the inverse of $A = \begin{bmatrix} 1 & 0 & -3 \\ 0 & 1 & 2 \\ 1 & 2 & 1 \end{bmatrix}$.

Solution:
Reducing $[A|I]$ gives

$$[A|I] = \left[\begin{array}{ccc|ccc} 1 & 0 & -3 & 1 & 0 & 0 \\ 0 & 1 & 2 & 0 & 1 & 0 \\ 1 & 2 & 1 & 0 & 0 & 1 \end{array}\right] \Rightarrow \underset{R_3 \to R_3 - R_1}{} \left[\begin{array}{ccc|ccc} 1 & 0 & -3 & 1 & 0 & 0 \\ 0 & 1 & 2 & 0 & 1 & 0 \\ 0 & 2 & 4 & -1 & 0 & 1 \end{array}\right]$$
$$\Rightarrow \underset{R_3 \to R_3 - 2R_2}{} \left[\begin{array}{ccc|ccc} 1 & 0 & -3 & 1 & 0 & 0 \\ 0 & 1 & 2 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 & -2 & 1 \end{array}\right]$$

Here the bottom row coefficients are all zero so no identity occurs and therefore A is noninvertible.

2.15.3 Solving Linear Systems Using Matrix Inversion

If A is a $n \times n$ invertible matrix (ie. A^{-1} exists), then the linear system

$$A\boldsymbol{x} = \boldsymbol{b}$$

can be solved using A^{-1} as follows. Left-multiplying both sides by A^{-1} gives

$$\begin{aligned} A^{-1}A\boldsymbol{x} &= A^{-1}\boldsymbol{b} \\ I\boldsymbol{x} &= A^{-1}\boldsymbol{b} \\ \boldsymbol{x} &= A^{-1}\boldsymbol{b} \end{aligned}$$

Thus $\boxed{\boldsymbol{x} = A^{-1}\boldsymbol{b}}$. This process is known as **method of inverses**. Note that this method cannot be used to solve a linear system $A\boldsymbol{x} = \boldsymbol{b}$ if A is not square or if A is noninvertible.

Example 2-47

Solve the linear system by the method of inverses:

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

Solution:

$$A = \begin{bmatrix} 2 & 1 & 0 \\ -4 & -1 & -3 \\ 3 & 1 & 2 \end{bmatrix}, \boldsymbol{x} = \begin{bmatrix} x \\ y \\ z \end{bmatrix}, \boldsymbol{b} = \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix}$$

In Example 2-45 we found

$$A^{-1} = \begin{bmatrix} 1 & -2 & -3 \\ -1 & 4 & 6 \\ -1 & 1 & 2 \end{bmatrix}$$

Therefore, using the method of inverses:

$$\boldsymbol{x} = A^{-1}\boldsymbol{b} = \begin{bmatrix} 1 & -2 & -3 \\ -1 & 4 & 6 \\ -1 & 1 & 2 \end{bmatrix} \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix} = \begin{bmatrix} -1-4-3 \\ 1+8+6 \\ 1+2+2 \end{bmatrix} = \begin{bmatrix} -8 \\ 15 \\ 5 \end{bmatrix}$$

So $x = -8$, $y = 15$, and $z = 5$.

We now state some conditions relating the invertibility of a square matrix A, to other matrix and linear system properties we have studied thus far.

Theorem 2-18: If A is an $n \times n$ matrix, the follow statements are equivalent:

1. A is invertible.
2. $A\boldsymbol{x} = \boldsymbol{b}$ has a unique solution.
3. $A\boldsymbol{x} = \mathbf{0}$ has the unique trivial solution $\boldsymbol{x} = \mathbf{0}$.
4. $\operatorname{rank}(A) = n$.
5. The reduced row echelon form of A is I, the identity matrix.

Using our ability to find inverses as well as the basic properties of matrices (Theorem 2-4) allows further methods for solving matrix equations. Note that the properties $(A^T)^T = A$ for transposes (Theorem 2-5) and the similar property $(A^{-1})^{-1} = A$ for inverses (Theorem 2-12) can be useful for isolating a matrix variable.

Example 2-48

Solve the following matrix equation for the matrix A.

$$(5A^T)^{-1} = \begin{bmatrix} 2 & 8 \\ 1 & 5 \end{bmatrix}$$

Solution:
Inverting both sides of the equation gives

$$\left((5A^T)^{-1}\right)^{-1} = \begin{bmatrix} 2 & 8 \\ 1 & 5 \end{bmatrix}^{-1}.$$

But $(B^{-1})^{-1} = B$ so the left-hand side simplifies to

$$5A^T = \begin{bmatrix} 2 & 8 \\ 1 & 5 \end{bmatrix}^{-1}.$$

Using Theorem 2-17 the right-hand side may be directly evaluated.

$$5A^T = \frac{1}{(2)(5) - (8)(1)} \begin{bmatrix} 5 & -8 \\ -1 & 2 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 5 & -8 \\ -1 & 2 \end{bmatrix}$$

Multiply both sides by 1/5 to get

$$A^T = \frac{1}{10} \begin{bmatrix} 5 & -8 \\ -1 & 2 \end{bmatrix}.$$

Taking the transpose of both sides gives

$$(A^T)^T = \left(\frac{1}{10} \begin{bmatrix} 5 & -8 \\ -1 & 2 \end{bmatrix}\right)^T$$

Using Theorem 2-5 the left-hand side simplifies to A and one can pull the constant 1/10 out of the transpose on the right-hand side to get

$$A = \frac{1}{10} \begin{bmatrix} 5 & -8 \\ -1 & 2 \end{bmatrix}^T.$$

Evaluating the transpose gives

$$A = \frac{1}{10}\begin{bmatrix} 5 & -1 \\ -8 & 2 \end{bmatrix},$$

and one can, if desired, bring the scalar constant into the matrix:

$$A = \begin{bmatrix} \frac{5}{10} & -\frac{1}{10} \\ -\frac{8}{10} & \frac{2}{10} \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & -\frac{1}{10} \\ -\frac{4}{5} & \frac{1}{5} \end{bmatrix}.$$

This solution can be checked in the original matrix equation.

An Application to Geometry

As an interesting application of solving linear systems consider the following problem. A common tool in computer drawing programs (such as the free software program *xfig*) is a tool for drawing circular arcs. The user enters three points and the program will then find a circular arc through those points. Let us explore how this is calculated. The equation for a circle of radius 2 centred at the origin $(0,0)$ of a Cartesian Coordinate system is, using the Pythagorean Theorem,

$$x^2 + y^2 = 2^2 .$$

If the centre of the circle is at the point $(3,4)$ instead of the origin one would have the equation

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

In general, a circle with radius r centred at the point (h,k) has the equation

$$(x-h)^2 + (y-k)^2 = r^2 .$$

If this equation is expanded one gets

$$x^2 - 2xh + h^2 + y^2 - 2yk + k^2 = r^2 .$$

If one rearranges this one has

$$x^2 + y^2 + (-2h)x + (-2k)y + (h^2 + k^2 - r^2) = 0$$

By introducing three new constants, $a = -2h$, $b = -2k$ and $c = h^2 + k^2 - r^2$ one can replace h, k, and r to get a new circle equation:

$$x^2 + y^2 + ax + by + c = 0 .$$

If we can therefore figure out a, b, and c for this equation we could solve to get h, k, and r and thereby find our circle.

Suppose we know a point $(x,y) = (-1,-3)$ sits on the desired circle. For this to be true it must satisfy the circle equation and we have

$$(-1)^2 + (-3)^2 + a(-1) + b(-3) + c = 0 ,$$

which can be rewritten

$$-a - 3b + c = -10 .$$

Now despite the equation of the circle being quadratic in the variables x and y, this equation in terms of the unknown constants a, b, and c, is linear! Having knowledge of two more points on the circle produces two more equations involving the unknown constants, thereby creating a determined linear system which we can solve.

Example 2-49

Find the circle that goes through the three points $(-1,-3)$, $(5,5)$, and $(-2,4)$.

Solution:
Inserting the the (x,y) values of each point into the equation

$$x^2+y^2+ax+by+c=0$$

produces the following determined linear system in constants a, b, and c:

$$\begin{aligned}
(-1,-3):(-1)^2+(-3)^2+a(-1)+b(-3)+c=0 \quad &\implies \quad -a-3b+c=-10\\
(5,5):(5)^2+(5)^2+a(5)+b(5)+c=0 \quad &\implies \quad 5a+5b+c=-50\\
(-2,4):(-2)^2+(4)^2+a(-2)+b(4)+c=0 \quad &\implies \quad -2a+4b+c=-20
\end{aligned}$$

The system can be represented by $A\boldsymbol{x}=\boldsymbol{b}$ with matrices defined by

$$A=\begin{bmatrix}-1 & -3 & 1\\ 5 & 5 & 1\\ -2 & 4 & 1\end{bmatrix} \qquad \boldsymbol{x}=\begin{bmatrix}a\\ b\\ c\end{bmatrix} \qquad \boldsymbol{b}=\begin{bmatrix}-10\\ -50\\ -20\end{bmatrix}.$$

The quickest way to find $\boldsymbol{x}$ is to reduce $[A|\boldsymbol{b}]$ and back-substitute. A longer method is to solve for the inverse A^{-1} by reducing $[A|I]$ to $[I|A^{-1}]$ to get (show this!)

$$A^{-1}=\begin{bmatrix}\frac{1}{50} & \frac{7}{50} & -\frac{4}{25}\\ -\frac{7}{50} & \frac{1}{50} & \frac{3}{25}\\ \frac{3}{5} & \frac{1}{5} & \frac{1}{5}\end{bmatrix},$$

and then use the method of inverses to find

$$\boldsymbol{x}=\begin{bmatrix}a\\ b\\ c\end{bmatrix}=A^{-1}\boldsymbol{b}=\begin{bmatrix}\frac{1}{50} & \frac{7}{50} & -\frac{4}{25}\\ -\frac{7}{50} & \frac{1}{50} & \frac{3}{25}\\ \frac{3}{5} & \frac{1}{5} & \frac{1}{5}\end{bmatrix}\begin{bmatrix}-10\\ -50\\ -20\end{bmatrix}=\begin{bmatrix}-\frac{1}{5}-7+\frac{80}{25}\\ \frac{7}{5}-1-\frac{60}{25}\\ -6-10-4\end{bmatrix}=\begin{bmatrix}-4\\ -2\\ -20\end{bmatrix}.$$

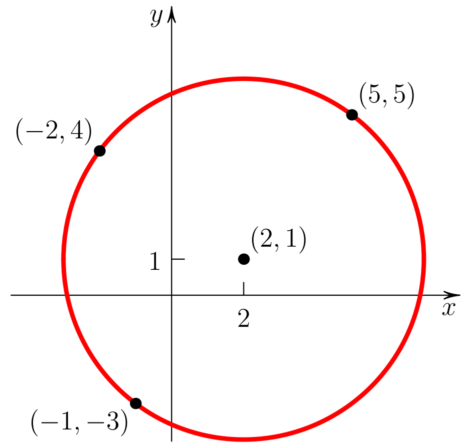
So $a=-4$, $b=-2$, and $c=-20$. Next solve for constants h, k, and r using the formulas from our previous discussion:

- $a=-4=-2h \implies \boxed{h=2}$
- $b=-2=-2k \implies \boxed{k=1}$
- $c=-20=h^2+k^2-r^2 \implies -20=(2)^2+(1)^2-r^2 \implies r^2=25 \implies \boxed{r=5}$

So the desired circle has centre $(h,k)=(2,1)$ and radius $r=5$. Inserting these constants into $(x-h)^2+(y-k)^2=r^2$ gives the equation for the circle

$$(x-2)^2+(y-1)^2=25.$$

A plot verifies the circle is correct



If three distinct points are collinear, like $(-1, 0)$, $(0, 0)$, and $(0, 1)$, the linear system generated will be inconsistent. (Show this!)

2.16 Elementary Matrices

Definition: A square matrix is called an **elementary matrix** if it can be obtained from the identity matrix of the same dimension by performing a single elementary row operation.

Since only a single row operation is allowed, this means an elementary matrix arises in one of three ways (and can be classified by this):

1. Multiplication of a row by a nonzero scalar.
2. Addition of a multiple of one row to a different row.
3. Interchanging of two rows.

Example 2-50

Determine the elementary matrices for each of the following row operations for the square matrix of given size.

1. $R_1 \to 2R_1, \quad 2 \times 2$

   Solution:
$$I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad E = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$$

2. $R_2 \to R_2 - 3R_1, \quad 4 \times 4$

   Solution:
$$I = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad E = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -3 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

3. $R_1 \leftrightarrow R_3, \quad 3 \times 3$

   Solution:
$$I = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad E = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

Example 2-51

Determine whether the given matrices are elementary matrices. If they are write down the corresponding row operation.

1. $E_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -2 & 1 \end{bmatrix}$

   Solution: Elementary, $R_3 \to R_3 - 2R_2$

2. $E_2 = \begin{bmatrix} 1 & 0 & 3 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

   Solution: Elementary, $R_1 \to R_1 + 3R_3$

3. $E_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

   Solution: Elementary, $R_1 \to (1)R_1$

4. $E_4 = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & -2 \\ 0 & 0 & 0 & 1 \end{bmatrix}$

   Solution: Not an elementary matrix

Elementary Matrix Notation and Inverse

We can write a particular elementary matrix with the following notation for given $n \times n$ dimension.

1. $E_{ii}(c)$ is obtained from I by multiplying $c \neq 0$ times row i .
2. $E_{ij}(c), i \neq j$ is obtained from I by adding c times row j to row i .
3. P_{ij} is obtained from I by interchanging rows i and j .

Consideration of how to undo the row operation underlying a given elementary matrix results in the following theorem.

Theorem 2-19: Every elementary matrix is invertible where the inverse is an elementary matrix given by:

$$\begin{aligned} (E_{ii}(c))^{-1} &= E_{ii}\left(\frac{1}{c}\right) \\ (E_{ij}(c))^{-1} &= E_{ij}(-c) \qquad (i \neq j) \\ (P_{ij})^{-1} &= P_{ij}\ . \end{aligned}$$

Example 2-52

Find the inverse of the given elementary matrix.

1. $E = \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix}$

   Solution:
   E arises from $R_2 \to R_2 + 2R_1$ therefore $E = E_{21}(2)$.

   Thus $E^{-1} = E_{21}(-2)$ and $E^{-1} = \begin{bmatrix} 1 & 0 \\ -2 & 1 \end{bmatrix}$.

2. $E = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$

   Solution:
   E arises from $R_2 \leftrightarrow R_3$ therefore $E = P_{23}$.

   Thus $E^{-1} = P_{23}$ and $E^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$.

3. $E = \begin{bmatrix} 1 & 0 \\ 0 & -2 \end{bmatrix}$

   Solution:
   E arises from $R_2 \to (-2)R_2$ therefore $E = E_{22}(-2)$.

   Thus $E^{-1} = E_{22}\left(-\frac{1}{2}\right) = \begin{bmatrix} 1 & 0 \\ 0 & -\frac{1}{2} \end{bmatrix}$.

Row Operations by Matrix Multiplication

The following theorem demonstrates the utility of elementary matrices. They allow us to represent row operations using matrix multiplication.

Theorem 2-20: If the elementary matrix E results from performing a certain elementary row operation on I_m (the $m \times m$ identity matrix) and if A is an $m \times n$ matrix, then the product EA is the matrix that results when the same row operation is performed on A .

Example 2-53

Given the elementary matrix E and A, identify the row operation corresponding to E and find the product EA directly to verify the row operation is indeed performed on A .

1. $E = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 3 & 0 & 1 \end{bmatrix} \quad A = \begin{bmatrix} 1 & 0 & 2 & 3 \\ 2 & -1 & 3 & 6 \\ 1 & 4 & 4 & 0 \end{bmatrix}$

   Solution:
   E corresponds to $R_3 \to R_3 + 3R_1$

   $EA = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 3 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 2 & 3 \\ 2 & -1 & 3 & 6 \\ 1 & 4 & 4 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 2 & 3 \\ 2 & -1 & 3 & 6 \\ 4 & 4 & 10 & 9 \end{bmatrix}$ which is $R_3 \to R_3 + 3R_1$ acting on A .

2. $E = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad A = \begin{bmatrix} 1 & 0 & 1 & 5 & 0 \\ 1 & 2 & 0 & 1 & 0 \\ -1 & -3 & -1 & 3 & 1 \\ 2 & 1 & 1 & -2 & 0 \end{bmatrix}$

   Solution:
   E corresponds to $R_3 \to 2R_3$

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

   which is $R_3 \to 2R_3$ acting on A .

3. $E = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \quad A = \begin{bmatrix} 2 & -1 & 3 \\ -1 & 0 & 2 \\ 1 & 1 & 5 \end{bmatrix}$

   Solution:
   E corresponds to $R_2 \leftrightarrow R_3$
   $EA = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 2 & -1 & 3 \\ -1 & 0 & 2 \\ 1 & 1 & 5 \end{bmatrix} = \begin{bmatrix} 2 & -1 & 3 \\ 1 & 1 & 5 \\ -1 & 0 & 2 \end{bmatrix}$ which is $R_2 \leftrightarrow R_3$ acting on A.

Example 2-54

For each pair of matrices, find an elementary matrix such that $B = EA$.

1. $A = \begin{bmatrix} 1 & -1 \\ 2 & 1 \end{bmatrix} \quad B = \begin{bmatrix} 1 & -1 \\ 4 & 2 \end{bmatrix}$

   Solution:
   Since B arises from A by the row operation $R_2 \to 2R_2$,
   $E = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$
   Direct multiplication confirms $EA = B$.

2. $A = \begin{bmatrix} 1 & 1 & -1 \\ 2 & 1 & 2 \\ -1 & 1 & 0 \end{bmatrix} \quad B = \begin{bmatrix} 1 & 1 & -1 \\ 2 & 1 & 2 \\ 1 & 2 & 2 \end{bmatrix}$

   Solution:
   Since B arises from A by the row operation $R_3 \to R_3 + R_2$,
   $E = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}$
   Direct multiplication confirms $EA = B$.

Elementary Matrix Decomposition

Theorem 2-21: A matrix A is invertible if and only if it can be written as a product of elementary matrices of the same dimension as A.

To find the matrix product, recall the reduced row echelon form of A is I. Thus one can reduce A to I keeping track of the k row operations required. One then has

$$E_k E_{k-1} \cdots E_1 A = I\,.$$

for some k elementary matrices E_i. Then, considering $B = E_k E_{k-1} \cdots E_1$ one can multiply both sides by B^{-1} to get:

$$A = (E_k E_{k-1} \cdots E_1)^{-1} I = (E_k E_{k-1} \cdots E_1)^{-1} = (E_1)^{-1} \cdots (E_{k-1})^{-1} (E_k)^{-1}\,.$$

Since each inverted elementary matrix is itself an elementary matrix one has the required product.

Example 2-55

If possible, write $A = \begin{bmatrix} 1 & -1 & 0 \\ 1 & -1 & 1 \\ 0 & 1 & 1 \end{bmatrix}$ as a product of elementary matrices.

Solution:

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

$$\Downarrow$$

$$R_2 \to R_2 - R_1 \quad \begin{bmatrix} 1 & -1 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix} = E_{21}(-1)A$$

$$\Downarrow$$

$$R_2 \leftrightarrow R_3 \quad \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} = P_{23}E_{21}(-1)A$$

$$\Downarrow$$

$$R_2 \to R_2 - R_3 \quad \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = E_{23}(-1)P_{23}E_{21}(-1)A$$

$$\Downarrow$$

$$R_1 \to R_1 + R_2 \quad \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = I = E_{12}(1)E_{23}(-1)P_{23}E_{21}(-1)A$$

Having reduced A to RREF, the matrix is invertible and the process of reduction can be written as follows:

$$E_{12}(1)E_{23}(-1)P_{23}E_{21}(-1)A = I$$

Remember that order is important here with the first row operation needing to act on A first. Then we have

$$\begin{aligned} A &= (E_{12}(1)E_{23}(-1)P_{23}E_{21}(-1))^{-1}\, I \\ &= (E_{21}(-1))^{-1}(P_{23})^{-1}(E_{23}(-1))^{-1}(E_{12}(1))^{-1} \\ &= E_{21}(1)P_{23}E_{23}(1)E_{12}(-1)\,. \end{aligned}$$

Example 2-56

If possible, write $A = \begin{bmatrix} 0 & 1 & -2 \\ 1 & 0 & 4 \\ 0 & 0 & 3 \end{bmatrix}$ as a product of elementary matrices.

Solution:

$$A = \begin{bmatrix} 0 & 1 & -2 \\ 1 & 0 & 4 \\ 0 & 0 & 3 \end{bmatrix}$$

$$\Downarrow$$

$$\begin{matrix} R_1 \leftrightarrow R_2 \\ \\ R_3 \to \frac{1}{3}R_3 \end{matrix} \begin{bmatrix} 1 & 0 & 4 \\ 0 & 1 & -2 \\ 0 & 0 & 1 \end{bmatrix} = P_{12}E_{33}\left(\tfrac{1}{3}\right) A$$

$$\Downarrow$$

$$\begin{matrix} R_1 \to R_1 - 4R_3 \\ R_2 \to R_2 + 2R_3 \\ \\ \end{matrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = I = E_{13}(-4)E_{23}(2)P_{12}E_{33}\left(\tfrac{1}{3}\right) A$$

Then A is invertible and one has

$$E_{13}(-4)E_{23}(2)E_{33}(1/3)P_{12}A = I\,.$$

Therefore:

$$\begin{aligned} A &= (E_{13}(-4)E_{23}(2)E_{33}(1/3)P_{12})^{-1}I \\ &= (P_{12})^{-1}(E_{33}(1/3))^{-1}(E_{23}(2))^{-1}(E_{13}(-4))^{-1} \\ &= P_{12}E_{33}(3)E_{23}(-2)E_{13}(4)\,. \end{aligned}$$

Note that while the order of the matrices in general matters, the order of

$$P_{12} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad E_{33}(1/3) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1/3 \end{bmatrix}$$

could have been swapped as both were introduced at the same initial step. That this is acceptable is because one can verify $P_{12}E_{33}(1/3) = E_{33}(1/3)P_{12}$, i.e. the two matrices commute. Similarly $E_{13}(-4)$ and $E_{23}(2)$ could have been introduced in a swapped order at the second step.

Example 2-57

If possible, write $A = \begin{bmatrix} 1 & -1 & 1 \\ -2 & 2 & 1 \\ 0 & 0 & 1 \end{bmatrix}$ as a product of elementary matrices.

Solution:

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

$$\Downarrow$$

$$R_2 \to R_2 + 2R_1 \begin{bmatrix} 1 & -1 & 1 \\ 0 & 0 & 3 \\ 0 & 0 & 1 \end{bmatrix} = E_{21}(2)A$$

$$\Downarrow$$

$$R_2 \leftrightarrow R_3 \begin{bmatrix} 1 & -1 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 3 \end{bmatrix} = P_{23}E_{21}(2)A$$

$$\Downarrow$$

$$R_3 \to R_3 - 3R_2 \begin{bmatrix} 1 & -1 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} = E_{32}(-3)P_{23}E_{21}(2)A$$

The matrix A cannot be reduced to the identity matrix, therefore A is noninvertible and A cannot be written as a product of elementary matrices.

We found, for an invertible matrix A, that we could reduce it by elementary row matrices to identity matrix I by

$$\underbrace{E_k E_{k-1} \cdots E_1}_{=B} A = I\,.$$

Since $BA = I$ it follows that $B = A^{-1}$ and the elementary matrix expansion of A^{-1}, if desired, is therefore

$$A^{-1} = E_k E_{k-1} \cdots E_1\,.$$

In practice this mechanism is a complicated way to find the inverse due to the multiplication required. As was already seen, augmenting A to $[A|I]$ and reducing to $[I|A^{-1}]$ is an algorithmically superior solution.

Consider the more general problem of reducing an arbitrary $m \times n$ matrix A to its reduced row echelon form. Since this can be done by elementary row operations which can be represented by multiplication by elementary matrices we have

$$E_k E_{k-1} \cdots E_1 A = R$$

where R is the RREF of matrix A and the E_i are $m \times m$ elementary matrices. Defining $B = E_k E_{k-1} \cdots E_1$ we see that for any matrix A we can find an invertible matrix B, written as a product of elementary matrices such that

$$BA = R\,.$$

In other words, we can reduce any matrix A to its $RREF$ by matrix multiplication. Now since B is a product of elementary matrices which in turn are found by reducing A to R we could find B by matrix multiplication. However our experience with inverting matrices suggests that a more efficient mechanism to find B is to augment the $m \times n$ matrix A by the identity matrix I_m and reduce as follows

$$[A|I] \Rightarrow [R|B]\,.$$

The desired product $B = E_k E_{k-1} \cdots E_1$ will then be found quickly. The following example illustrates the procedure.

Example 2-58

Given the matrix

$$A = \begin{bmatrix} 1 & 2 & 0 \\ 3 & 5 & -2 \end{bmatrix}$$

find an invertible matrix B such that $BA = R$, where R is the reduced row echelon form of A, and express B as a product of elementary matrices.

Solution:
Augment A by I_2, the 2×2 identity matrix, and reduce to RREF keeping track of the row operations.

$$[A|I] = \left[\begin{array}{ccc|cc} 1 & 2 & 0 & 1 & 0 \\ 3 & 5 & -2 & 0 & 1 \end{array}\right]$$

$$\Downarrow$$

$$R_2 \to R_2 - 3R_1 \left[\begin{array}{ccc|cc} 1 & 2 & 0 & 1 & 0 \\ 0 & -1 & -2 & -3 & 1 \end{array}\right] = [E_1A|E_1]$$

$$\Downarrow$$

$$R_2 \to -R_2 \left[\begin{array}{ccc|cc} 1 & 2 & 0 & 1 & 0 \\ 0 & 1 & 2 & 3 & -1 \end{array}\right] = [E_2E_1A|E_2E_1]$$

$$\Downarrow$$

$$R_1 \to R_1 - 2R_2 \left[\begin{array}{ccc|cc} 1 & 0 & -4 & -5 & 2 \\ 0 & 1 & 2 & 3 & -1 \end{array}\right] = [E_3E_2E_1A|E_3E_2E_1] = [R|B]$$

Thus the RREF of A is

$$R = \begin{bmatrix} 1 & 0 & -4 \\ 0 & 1 & 2 \end{bmatrix},$$

and the matrix B which reduces A, so that $BA = R$, is given by

$$B = \begin{bmatrix} -5 & 2 \\ 3 & -1 \end{bmatrix}.$$

As a product of elementary matrices we have $B = E_3E_2E_1$, where, looking back at our row operations we have, calculating the elementary matrices by operating on the identity matrix,

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \Rightarrow R_2 \to R_2 - 3R_1 \begin{bmatrix} 1 & 0 \\ -3 & 1 \end{bmatrix} = E_1 = E_{21}(-3)$$

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \Rightarrow R_2 \to -R_2 \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} = E_2 = E_{22}(-1)$$

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \Rightarrow R_1 \to R_1 - 2R_2 \begin{bmatrix} 1 & -2 \\ 0 & 1 \end{bmatrix} = E_3 = E_{12}(-2)$$

Thus as a product of elementary matrices we have

$$B = \begin{bmatrix} -5 & 2 \\ 3 & -1 \end{bmatrix} = E_3E_2E_1 = E_{12}(-2)E_{22}(-1)E_{21}(-3) = \begin{bmatrix} 1 & -2 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -3 & 1 \end{bmatrix}.$$

One can confirm by direct matrix multiplication that $BA = R$ and that $B = E_3E_2E_1$ for these matrices.

Chapter 3: Determinants

3.1 Defining the Determinant

In Theorem 2-17 the inverse of the 2×2 matrix $\left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right]$ was found to be

$$\frac{1}{(ad-bc)}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

Clearly for this inverse formula to work the number $ad-bc$ must be nonzero. This single number $ad-bc$ which depends on all the entries of the matrix is called its determinant and the goal of this chapter is to generalize the formula to square matrices of arbitrary dimension, determine ways to calculate it, find out its properties and its useful applications.

Let A be an $n \times n$ square matrix. Associated with A there is a number called the **determinant** of A denoted by $\det A$, $\det(A)$, or $|A|$. Note that the vertical bars here *do not* refer to absolute value.

If A is a 1×1 matrix, so $A = [a_{11}]$, then $\det A = a_{11}$.

If A is a 2×2 matrix, so $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$, then $\det A = a_{11}a_{22} - a_{12}a_{21}$.

Example 3-1

If $A = \begin{bmatrix} -9 \end{bmatrix}$, then $\det A = |A| = -9$

If $B = \begin{bmatrix} 1 & 2 \\ 2 & -3 \end{bmatrix}$, then $\det B = |B| = (1)(-3) - (2)(2) = -3 - 4 = -7$

To define the determinant of square matrices of arbitrary dimension we will do so recursively in terms of the determinants of the smaller matrices they contain. This requires the following definition.

Definition: Let A be an $m \times n$ matrix, then the **submatrix** A_{ij} is obtained from A by deleting the i^{th} row and j^{th} column.

Example 3-2

If $A = \begin{bmatrix} 2 & -1 & 1 \\ 0 & 1 & 2 \\ 3 & 5 & -4 \end{bmatrix}$ then $A_{12} = \begin{bmatrix} 0 & 2 \\ 3 & -4 \end{bmatrix}$ and $A_{33} = \begin{bmatrix} 2 & -1 \\ 0 & 1 \end{bmatrix}$.

With this notation we note that the determinant of a 2×2 matrix $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$ can be written in terms of the determinants of its 1×1 submatrices as

$$\det A = a_{11}a_{22} - a_{12}a_{21} = a_{11} \det A_{11} - a_{12} \det A_{12}.$$

This suggests that we can generalize the determinant to a 3×3 matrix

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

by adding and subtracting alternately the product of each entry in its first row times the determinant of its corresponding submatrix:

$$\det A = a_{11} \det A_{11} - a_{12} \det A_{12} + a_{13} \det A_{13}.$$

Since these submatrices are 2×2 matrices, this is well-defined. Proceeding recursively we can now define the determinant for square matrices of any dimension.

Definition: Let $A = [a_{ij}]$ be a square matrix of dimension $n \times n$. Associated with A there is a number called the **determinant of $\boldsymbol{A}$** denoted by $\det A$, $\det(A)$, or $|A|$. If $n = 1$ then define $\det A = a_{11}$.
For $n > 1$ define

$$\det A = a_{11}\det A_{11} - a_{12}\det A_{12} + \ldots + (-1)^{n+1} a_{1n}\det A_{1n}\,,$$

where A_{ij} is the submatrix obtained from A by deleting the i^{th} row and j^{th} column from A.

Example 3-3

Compute the determinant of A for the given matrices.

1. $A = \begin{bmatrix} 6 & -3 \\ 5 & 9 \end{bmatrix}$

   Solution:

$$\det A = (6)(9) - (-3)(5) = 54 + 15 = 69$$

2. $A = \begin{bmatrix} 2 & 4 & 7 \\ 6 & 0 & 3 \\ 1 & 5 & 3 \end{bmatrix}$

   Solution:

$$\begin{aligned} \det A &= a_{11}\det A_{11} - a_{12}\det A_{12} + a_{13}\det A_{13} \\ &= 2\begin{vmatrix} 0 & 3 \\ 5 & 3 \end{vmatrix} - 4\begin{vmatrix} 6 & 3 \\ 1 & 3 \end{vmatrix} + 7\begin{vmatrix} 6 & 0 \\ 1 & 5 \end{vmatrix} \\ &= 2(0-15) - 4(18-3) + 7(30-0) \\ &= 2(-15) - 4(15) + 7(30) \\ &= -30 - 60 + 210 \\ &= 120 \end{aligned}$$

3. $A = \begin{bmatrix} 1 & 3 & 0 \\ -5 & 2 & 3 \\ 1 & 0 & -1 \end{bmatrix}$

   Solution:

$$\begin{aligned} \det A &= a_{11}\det A_{11} - a_{12}\det A_{12} + a_{13}\det A_{13} \\ &= 1\begin{vmatrix} 2 & 3 \\ 0 & -1 \end{vmatrix} - 3\begin{vmatrix} -5 & 3 \\ 1 & -1 \end{vmatrix} + \cancel{(0)\det A_{13}} \\ &= 1(-2-0) - 3(5-3) \\ &= 1(-2) - 3(2) \\ &= -2 - 6 \\ &= -8 \end{aligned}$$

4. $A = \begin{bmatrix} -1 & 2 & -3 \\ 1 & 1 & 2 \\ 0 & 2 & -5 \end{bmatrix}$

Solution:

$$\begin{aligned}\det A &= (-1)\begin{vmatrix}1 & 2\\ 2 & -5\end{vmatrix} - 2\begin{vmatrix}1 & 2\\ 0 & -5\end{vmatrix} + (-3)\begin{vmatrix}1 & 1\\ 0 & 2\end{vmatrix}\\ &= (-1)(-5-4) - 2(-5-0) - 3(2-0)\\ &= 9 + 10 - 6\\ &= 13\end{aligned}$$

The Rule of Sarrus for 3×3 Matrix Determinants

For a 2×2 matrix the determinant $\begin{vmatrix}a & b\\ c & d\end{vmatrix} = ad - bc$ is easily remembered as the product of the diagonal from left to right minus the product of the diagonal from right to left. For a 3×3 matrix a similar pattern emerges if one appends the first two columns of the matrix on the right of the matrix to form a 3×5 matrix.

$$\left|\begin{array}{ccc|cc} a_{11} & a_{12} & a_{13} & a_{11} & a_{12}\\ a_{21} & a_{22} & a_{23} & a_{21} & a_{22}\\ a_{31} & a_{32} & a_{33} & a_{31} & a_{32}\end{array}\right.$$

The determinant of the 3×3 matrix is then the sum of the three diagonal products from left to right minus the sum of the three diagonal products from right to left. This is known as the **Rule of Sarrus**. That this is true can be seen by applying the definition of the determinant to a general 3×3 matrix to get:

$$\begin{aligned}\begin{vmatrix}a_{11} & a_{12} & a_{13}\\ a_{21} & a_{22} & a_{23}\\ a_{31} & a_{32} & a_{33}\end{vmatrix} &= a_{11}\begin{vmatrix}a_{22} & a_{23}\\ a_{32} & a_{33}\end{vmatrix} - a_{12}\begin{vmatrix}a_{21} & a_{23}\\ a_{31} & a_{33}\end{vmatrix} + a_{13}\begin{vmatrix}a_{21} & a_{22}\\ a_{31} & a_{32}\end{vmatrix}\\ &= a_{11}(a_{22}a_{33} - a_{23}a_{32}) - a_{12}(a_{21}a_{33} - a_{23}a_{31}) + a_{13}(a_{21}a_{32} - a_{22}a_{31})\\ &= a_{11}a_{22}a_{33} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31}\\ &= a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33}\end{aligned}$$

Note that the Rule of Sarrus applies only to determinants of 3×3 matrices. The pattern fails for determinants of matrices of order greater than three.

Example 3-4

Find the determinant of the 3×3 matrix using the Rule of Sarrus.

1. $A = \begin{bmatrix}-1 & 2 & -3\\ 1 & 1 & 2\\ 0 & 2 & -5\end{bmatrix}$

   Solution:

$$\left|\begin{array}{ccc|cc} -1 & 2 & -3 & -1 & 2\\ 1 & 1 & 2 & 1 & 1\\ 0 & 2 & -5 & 0 & 2\end{array}\right.$$

$$\begin{aligned}\det A &= (-1)(1)(-5) + (2)(2)(0) + (-3)(1)(2) - (-3)(1)(0) - (-1)(2)(2) - (2)(1)(-5)\\ &= 5 + 0 - 6 + 0 + 4 + 10\\ &= 13\end{aligned}$$

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

   Solution:

$$\left|\begin{array}{ccc} 0 & 2 & 3\\ 1 & 4 & 0\\ 5 & 0 & 6\end{array}\right|\begin{array}{cc} 0 & 2\\ 1 & 4\\ 5 & 0\end{array}$$

$$\begin{aligned}\det A &= 0 + 0 + 0 - (3)(4)(5) - 0 - (2)(1)(6)\\ &= -60 - 12\\ &= -72\end{aligned}$$

3.1.1 Cofactor Expansion

Consider finding the determinant of the matrix

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

We have

$$\det A = 5\det A_{11} - 1\det A_{12} + 2\det A_{13} - 4\det A_{14}$$

and the determinants of the submatrices of dimension 3×3 would then need to be evaluated. To compute this requires a fair amount of work. We wish to study other methods of evaluating determinants.

Definition: Let A be an $n\times n$ square matrix (a matrix of order n) and let A_{ij} denote the $(n-1)\times(n-1)$ submatrix obtained from A by deleting the i^{th} row and j^{th} column, then:

- The ***i*, *j*-minor** of A, denoted by $\boldsymbol{m_{ij}}$, is given by $\boxed{m_{ij} = \det A_{ij}}$.
- The ***i*, *j*-cofactor** of A, denoted by $\boldsymbol{c_{ij}}$, is given by $\boxed{c_{ij} = (-1)^{i+j} m_{ij} = (-1)^{i+j}\det A_{ij}}$.
- The **cofactor matrix of *A*** is the $n\times n$ matrix $C = [c_{ij}]$.

Note that since $i+j$ is even if i and j are both even or both odd and $i+j$ is odd if i is even and j is odd, or *vice versa*, it follows that -1^{i+j} has the pattern

$$\begin{bmatrix} + & - & + & - & \cdots\\ - & + & - & + & \cdots\\ + & - & + & - & \cdots\\ - & + & - & + & \cdots\\ \vdots & \vdots & \vdots & \vdots & \end{bmatrix}.$$

Example 3-5

Find the minors, cofactors, and cofactor matrix of the given matrices.

1. $A = \begin{bmatrix} 2 & -3 & 1 \\ 4 & 0 & -2 \\ 3 & -1 & -3 \end{bmatrix}$

   Solution:
   The matrix has 9 elements, therefore it has 9 minors and 9 cofactors. The minors are:

$$m_{11} = \begin{vmatrix} 0 & -2 \\ -1 & -3 \end{vmatrix} = 0 - 2 = -2 \quad m_{12} = \begin{vmatrix} 4 & -2 \\ 3 & -3 \end{vmatrix} = -12 + 6 = -6 \quad m_{13} = \begin{vmatrix} 4 & 0 \\ 3 & -1 \end{vmatrix} = -4 + 0 = -4$$

$$m_{21} = \begin{vmatrix} -3 & 1 \\ -1 & -3 \end{vmatrix} = 9 + 1 = 10 \quad m_{22} = \begin{vmatrix} 2 & 1 \\ 3 & -3 \end{vmatrix} = -6 - 3 = -9 \quad m_{23} = \begin{vmatrix} 2 & -3 \\ 3 & -1 \end{vmatrix} = -2 + 9 = 7$$

$$m_{31} = \begin{vmatrix} -3 & 1 \\ 0 & -2 \end{vmatrix} = 6 - 0 = 6 \quad m_{32} = \begin{vmatrix} 2 & 1 \\ 4 & -2 \end{vmatrix} = -4 - 4 = -8 \quad m_{33} = \begin{vmatrix} 2 & -3 \\ 4 & 0 \end{vmatrix} = 0 + 12 = 12$$

   The cofactors additionally have the multiplicative sign factor -1^{i+j} to get

$$\begin{array}{lll} c_{11} = +m_{11} = -2 & c_{12} = -m_{12} = 6 & c_{13} = +m_{13} = -4 \\ c_{21} = -m_{21} = -10 & c_{22} = +m_{22} = -9 & c_{23} = -m_{23} = -7 \\ c_{31} = +m_{31} = 6 & c_{32} = -m_{32} = 8 & c_{33} = +m_{33} = 12 \end{array}$$

   The cofactor matrix of A is therefore

$$C = \begin{bmatrix} -2 & 6 & -4 \\ -10 & -9 & -7 \\ 6 & 8 & 12 \end{bmatrix}.$$

2. $A = \begin{bmatrix} 1 & -3 & 2 \\ 0 & 1 & 3 \\ -1 & -2 & 5 \end{bmatrix}$

   Solution:
   The minors of A are:

$$m_{11} = \begin{vmatrix} 1 & 3 \\ -2 & 5 \end{vmatrix} = 11 \quad m_{12} = \begin{vmatrix} 0 & 3 \\ -1 & 5 \end{vmatrix} = 3 \quad m_{13} = \begin{vmatrix} 0 & 1 \\ -1 & -2 \end{vmatrix} = 1$$

$$m_{21} = \begin{vmatrix} -3 & 2 \\ -2 & 5 \end{vmatrix} = -11 \quad m_{22} = \begin{vmatrix} 1 & 2 \\ -1 & 5 \end{vmatrix} = 7 \quad m_{23} = \begin{vmatrix} 1 & -3 \\ -1 & -2 \end{vmatrix} = -5$$

$$m_{31} = \begin{vmatrix} -3 & 2 \\ 1 & 3 \end{vmatrix} = -11 \quad m_{32} = \begin{vmatrix} 1 & 2 \\ 0 & 3 \end{vmatrix} = 3 \quad m_{33} = \begin{vmatrix} 1 & -3 \\ 0 & 1 \end{vmatrix} = 1$$

   The cofactors of A are:

$$\begin{array}{lll} c_{11} = +m_{11} = 11 & c_{12} = -m_{12} = -3 & c_{13} = +m_{13} = 1 \\ c_{21} = -m_{21} = 11 & c_{22} = +m_{22} = 7 & c_{23} = -m_{23} = 5 \\ c_{31} = +m_{31} = -11 & c_{32} = -m_{32} = -3 & c_{33} = +m_{33} = 1 \end{array}$$

   The cofactor matrix of A is therefore:

$$C = \begin{bmatrix} 11 & -3 & 1 \\ 11 & 7 & 5 \\ -11 & -3 & 1 \end{bmatrix}$$

Theorem 3-1: If A is an $n \times n$ matrix then $\det A$ can be evaluated by a **cofactor expansion** along any row or any column as follows:
Along the i^{th} row:

$$\boxed{\det A = a_{i1}c_{i1} + a_{i2}c_{i2} + \ldots + a_{in}c_{in}}\,.$$

Along the j^{th} column:

$$\boxed{\det A = a_{1j}c_{1j} + a_{2j}c_{2j} + \ldots + a_{nj}c_{nj}}\,.$$

In addition to illustrating a profound property of the determinant, cofactor expansion has practical utility. When evaluating a determinant we can choose to expand a row or column containing one or more zeros to simplify calculation.

Since a matrix with a zero row or column may always be expanded along it when finding the determinant we have the following corollary of the cofactor expansion theorem.

Corollary: If square matrix A has a zero row or column then $\det A = 0$.

Example 3-6

Evaluate the determinant of the given matrix.

1. $A = \begin{bmatrix} 2 & -3 & 1 \\ 4 & 0 & -2 \\ 3 & -1 & -3 \end{bmatrix}$

   Solution:
   Expanding along the second row we have:

$$\begin{aligned}
\det A &= a_{21}c_{21} + \cancel{a_{22}c_{22}} + a_{23}c_{23} \\
&= 4(-1)^{2+1}m_{21} + (-2)(-1)^{2+3}m_{23} \\
&= -4m_{21} + 2m_{23} \\
&= -4\begin{vmatrix} -3 & 1 \\ -1 & -3 \end{vmatrix} + 2\begin{vmatrix} 2 & -3 \\ 3 & -1 \end{vmatrix} \\
&= -4[(-3)(-3) - (1)(-1)] + 2[(2)(-1) - (-3)(3)] \\
&= -40 + 14 \\
&= -26
\end{aligned}$$

2. $A = \begin{bmatrix} 1 & 0 & 1 & -7 \\ 9 & 0 & 3 & -1 \\ 2 & 0 & 1 & 5 \\ 0 & 0 & 4 & 0 \end{bmatrix}$

   Solution:
   Since its second column is all zeros, $\det A = 0$.

3. $A = \begin{bmatrix} 1 & -3 & 2 \\ 0 & 1 & 3 \\ -1 & -2 & 5 \end{bmatrix}$

Solution:
Expanding along the first column:

$$\begin{aligned}
\det A &= a_{11}c_{11} + \cancel{a_{21}c_{21}} + a_{31}c_{31} \\
&= 1(-1)^{1+1}m_{11} + (-1)(-1)^{3+1}m_{31} \\
&= m_{11} - m_{31} \\
&= \begin{vmatrix} 1 & 3 \\ -2 & 5 \end{vmatrix} - \begin{vmatrix} -3 & 2 \\ 1 & 3 \end{vmatrix} \\
&= 5 - (-6) - [(-9) - 2] \\
&= 11 + 11 \\
&= 22
\end{aligned}$$

4. $A = \begin{bmatrix} 5 & 1 & 2 & 4 \\ -1 & 0 & 2 & 3 \\ 1 & 1 & 6 & 1 \\ 1 & 0 & 0 & -4 \end{bmatrix}$ (Our 4×4 matrix from before.)

Solution:
Expanding along the fourth row containing two zeros:

$$\begin{aligned}
\det A &= (1)(-1) \underbrace{\begin{vmatrix} 1 & 2 & 4 \\ 0 & 2 & 3 \\ 1 & 6 & 1 \end{vmatrix}}_{\text{expand } 2^{\text{nd}} \text{ row}} + 0 + 0 + (-4)(+1) \underbrace{\begin{vmatrix} 5 & 1 & 2 \\ -1 & 0 & 2 \\ 1 & 1 & 6 \end{vmatrix}}_{2^{\text{nd}} \text{ row}} \\
&= -\left[2(+1)\begin{vmatrix} 1 & 4 \\ 1 & 1 \end{vmatrix} + 3(-1)\begin{vmatrix} 1 & 2 \\ 1 & 6 \end{vmatrix}\right] - 4\left[(-1)(-1)\begin{vmatrix} 1 & 2 \\ 1 & 6 \end{vmatrix} + 2(-1)\begin{vmatrix} 5 & 1 \\ 1 & 1 \end{vmatrix}\right] \\
&= -[2(1-4) - 3(6-2)] - 4[(1)(6-2) - 2(5-1)] \\
&= -(-6-12) - 4(4-8) \\
&= 18 + 16 \\
&= 34
\end{aligned}$$

Determinants of Triangular Matrices

Definition: A square matrix is called **upper triangular** if all the entries below the main diagonal are zero and **lower triangular** if all entries above the main diagonal are zero.

Clearly a diagonal matrix is both upper and lower triangular.

Example 3-7

1. $U = \begin{bmatrix} 1 & 3 & 5 \\ 0 & 1 & 2 \\ 0 & 0 & 10 \end{bmatrix}$ is upper triangular.

2. $L = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 2 & 0 & 0 & 0 \\ -2 & 1 & 5 & 0 \\ 1 & 0 & 1 & 1 \end{bmatrix}$ is lower triangular.

3. $D = \begin{bmatrix} -2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 10 \end{bmatrix}$ is a diagonal matrix (and upper and lower triangular).

Theorem 3-2: If an $n \times n$ matrix A is upper triangular, lower triangular, or diagonal, then its determinant is the product of its entries on the main diagonal,

$$\boxed{\det A = a_{11}a_{22}\cdots a_{nn}}.$$

Proof:
Suppose A is upper triangular given by

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ 0 & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & a_{nn} \end{bmatrix}.$$

Expanding along its first column we have

$$\begin{aligned} \det A &= a_{11}c_{11} \\ &= a_{11}(-1)^{1+1}m_{11} \\ &= a_{11}\begin{vmatrix} a_{22} & \dots & a_{2n} \\ \vdots & \ddots & \vdots \\ 0 & \dots & a_{nn} \end{vmatrix} \end{aligned}$$

The resulting submatrix is still upper triangular and we can expand along its first column :

$$\det A = a_{11}a_{22}\begin{vmatrix} a_{33} & \dots & a_{3n} \\ \vdots & \ddots & \vdots \\ 0 & \dots & a_{nn} \end{vmatrix}$$

Continuing this process we have $\det A = a_{11}a_{22}\cdots a_{nn}$. A similar argument follows for lower triangular matrices using row expansion. Finally diagonal matrices are triangular so the theorem follows for them as well.[1]

Since an $n \times n$ identity matrix I is diagonal with all main diagonal entries equal to one we have that the product of those entries is also one and we have the following result.

Corollary: If I is an identity matrix, then $\boxed{\det I = 1}$.

Example 3-8

Compute the determinant of the given matrix.

1. $A = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & -2 \end{bmatrix}$ will have a $\det A = (-1)(3)(-2) = 6$ as it is a diagonal matrix.

[1]A rigorous proof of this theorem using mathematical induction is as follows. Let $P(n)$ be the proposition that the determinant of an order n upper diagonal matrix is the product of its diagonal elements. Then $P(1)$ is true since then $A = [a_{11}]$ and $\det A = a_{11}$. Next suppose $P(n)$ is true. Then an $(n+1) \times (n+1)$ upper diagonal matrix A can be expanded along its top row to get $\det A = a_{11} \det A_{11}$. But submatrix A_{11} is of dimension n and is upper triangular so its determinant, since $P(n)$ is true, is the product of its diagonal elements, $\det A_{11} = a_{22}\cdots a_{n+1,n+1}$. Therefore $\det A = a_{11}a_{22}\cdots a_{n+1,n+1}$ and $P(n+1)$ is true. By mathematical induction the theorem is therefore true for all n .

2. $A = \begin{bmatrix} -1 & 0 & 0 & 0 \\ 3 & 5 & 0 & 0 \\ 2 & 1 & 8 & 0 \\ -1 & 1 & 3 & 2 \end{bmatrix}$ is lower triangular therefore $\det A = (-1)(5)(8)(2) = -80$

3.1.2 Determinant of a Matrix Product

Determinants have the following remarkable property.

Theorem 3-3: Let A and B be $n \times n$ matrices then the determinant of their product is the product of their determinants. In symbols,

$$\boxed{\det(AB) = \det(A)\det(B)}\,.$$

Note, however that the result is not true of the sum of two matrices. In general

$$\det(A+B) \neq \det A + \det B\,.$$

Example 3-9

Let $A = \begin{bmatrix} -1 & 2 \\ 3 & 1 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 4 \\ -2 & 1 \end{bmatrix}$. Find $\det A$, $\det B$, $\det(AB)$ and $\det(A+B)$.

Solution:

$$\begin{aligned} \det A &= \begin{vmatrix} -1 & 2 \\ 3 & 1 \end{vmatrix} = -1 - 6 = -7 \\ \det B &= \begin{vmatrix} 1 & 4 \\ -2 & 1 \end{vmatrix} = 1 - (-8) = 9 \\ \det(AB) &= \begin{vmatrix} -5 & -2 \\ 1 & 13 \end{vmatrix} = -65 - (-2) = -63 \\ \det(A+B) &= \begin{vmatrix} 0 & 6 \\ 1 & 2 \end{vmatrix} = 0 - 6 = -6 \end{aligned}$$

Note that $\det(AB) = -63$ equals $(\det A)(\det B) = (-7)(9) = -63$ as predicted by Theorem 3-3, but that $\det(A+B) = -6$ does not equal $\det A + \det B = -7 + 9 = 2$.

For three square matrices of same order we have, by the determinant product theorem:

$$\det(ABC) = \det[(AB)C] = \det(AB)\det(C) = \det(A)\det(B)\det(C)\ .$$

In general we have the following for any finite number of such matrices.

Corollary 1: Let A_1,A_2,..., A_k be k matrices of dimension $n \times n$. Then

$$\boxed{\det(A_1A_2\cdots A_k) = \det(A_1)\det(A_2)\cdots\det(A_k)}\,.$$

Setting $A_i = A$ for all i in the previous corollary gives the following for the determinant of a power of A.

Corollary 2: Let A be a square matrix and k a positive integer. The determinant of the k^{th} power of A is the k^{th} power of its determinant. In symbols

$$\boxed{\det\left(A^k\right) = (\det A)^k}.$$

We have seen that matrix multiplication does not, in general, commute so $AB \neq BA$. However, because multiplication of numbers commutes and determinants are just numbers we have $\boxed{\det(BA) = \det(AB)}$ since

$$\det(BA) = \det(B)\det(A) = \det(A)\det(B) = \det(AB)\ .$$

Generalizing to the product of k matrices gives the following final corollary.

Corollary 3: Let A_1,A_2,..., A_k be k matrices of dimension $n \times n$. Then the determinant of the product $A_1A_2 \cdots A_k$ equals the determinant of the product of the k matrices evaluated in any order.

3.1.3 Determinant of a Transpose

Theorem 3-4: If A is a square matrix then the determinant of its transpose equals the determinant of A,

$$\boxed{\det\left(A^T\right) = \det A}.$$

Proof:
Consider the case where A is a 2×2 matrix. Then

$$A = \begin{bmatrix} a_{11}a_{12} \\ a_{21}a_{22} \end{bmatrix} \quad \text{and therefore} \quad A^T = \begin{bmatrix} a_{11}a_{21} \\ a_{12}a_{22} \end{bmatrix}\ .$$

By direct computation one has:

$$\det A = a_{11}a_{22} - a_{12}a_{21} = \det\left(A^T\right)\ .$$

For A of higher dimension one may proceed by induction by doing cofactor expansion along the first row of A and the first column of A^T. The submatrices generated will be the transposes of each other and of smaller dimension than A so their determinants will be equal by assumption of the truth of the n^{th} step.

3.1.4 Determinants of Orthogonal Matrices

Orthogonal matrices have some unique properties that distinguish them from other square matrices, one particular property is related to their determinants.

Theorem 3-5: If A is an orthogonal matrix then $\boxed{\det(A) = 1 \text{ or } \det(A) = -1}$.

Proof:

Using that $A^{-1} = A^T$ for an orthogonal matrix we have:

$$1 = \det(I) = \det\left(A^{-1}A\right) = \det\left(A^TA\right) = \det\left(A^T\right)\det(A) = \det(A)\det(A) = [\det(A)]^2\,.$$

Since $1 = (\det A)^2$ the result follows.

3.1.5 Determinants of Elementary Matrices

We have seen that an invertible matrix can be decomposed into a product of elementary matrices. Since the determinant of such a product is just the product of the determinants, knowledge of the determinants of elementary matrices will be useful.

Example 3-10

Find the determinant of the given elementary matrix.

1. $E = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 7 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$

   The matrix is diagonal with determinant $\det E = E_{22}(7) = (1)(7)(1)(1) = 7$

2. $E = \begin{bmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

   The matrix is lower diagonal with 1's along the main diagonal, therefore $\det E = E_{21}(3) = (1)(1)(1) = 1$.

3. $E = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$

   Expanding along the first row gives $\det E = \det P_{23} = (1)(+1) \begin{vmatrix} 0 & 1 \\ 1 & 0 \end{vmatrix} = 1(0-1) = -1$

The results of the previous example may be generalized to arbitrary elementary matrix examples of the three types as given in the following theorem.

Theorem 3-6: Let E be an elementary matrix.

1. If E results from multiplying a row by a nonzero scalar c then $\det E = \det E_{ii}(c) = c$.
2. If E results from the addition of a multiple of one row to a different row then $\det E = \det E_{ij}(c) = 1$.
3. If E results from interchanging two rows then $\det E = \det P_{ij} = -1$.

Example 3-11

In Example 2-56 we found the decomposition

$$A = \begin{bmatrix} 0 & 1 & -2 \\ 1 & 0 & 4 \\ 0 & 0 & 3 \end{bmatrix} = (P_{12})^{-1}(E_{33}(1/3))^{-1}(E_{13}(-4))^{-1}(E_{23}(2))^{-1} = P_{12}E_{33}(3)E_{13}(4)E_{23}(-2)\,.$$

It follows that

$$\det A = \det P_{12} \det E_{33}(3) \det E_{13}(4) \det E_{23}(-2) = (-1)(3)(1)(1) = -3\,.$$

Note that, as this last example recalls, we must invert the matrices that implement the row reduction to find the decomposition of A. From Theorem 2-19 and the last theorem it follows that $\det\left(E^{-1}\right) = 1/\det E$. (We will see shortly this is true of any invertible matrix.) As such the determinant of A is the product of the reciprocals of the determinants of the elementary matrices that implement the actual row reduction, or, more simply, the reciprocal of the product of those determinants.

3.1.6 Effect of Row Operations on Determinants

In evaluating determinants we have seen the value of doing cofactor expansion along a row or column dominated by zeros. We have seen previously how Gauss-Jordan elimination was able to produce zeros in a matrix. We now explore what effect an elementary row operation has on the determinant of a matrix, with an eye to using such knowledge to simplify determinant calculations. Since row operations can be implemented by elementary matrices, knowledge of their determinants yields the following useful theorem.

Theorem 3-7: (**Effect of Row/Column Operations on Determinants**)

Let A be a square matrix.

1. If one row (column) of A is multiplied by a nonzero scalar c then the determinant changes by a factor of c.
2. If a scalar multiple of one row (column) is added to another row (column), then the determinant is unchanged.
3. If two rows (columns) of A are interchanged, then the determinant changes by a factor of -1.

Proof:
Suppose matrix A' is created by such a row operation on matrix A. Then there exists an elementary matrix E such that $A' = EA$. Then $\det A' = \det(EA) = \det(E)\det(A)$ and the stated result follows from Theorem 3-6 by consideration of the type of row operation. Next suppose A' is created by such an operation on a column of A. Then said operation can be represented by a row operation on the transpose of A so that $A' = (EA^T)^T$. Then $\det A' = \det\left((EA^T)^T\right) = \det\left(EA^T\right) = \det(E)\det\left(A^T\right) = \det E \det A$ and the result follows again from Theorem 3-6.

The theorem has some useful corollaries.

Corollary 1: If A is an $n \times n$ matrix and c is a scalar, then $\boxed{\det(cA) = c^n \det A}$.

This follows by 1. since multiplying A by c is equivalent to multiplying each of the n rows by c.

Corollary 2: If square matrix A has a row (column) that is a scalar multiple of another row (column) then $\det A = 0$.

This follows by 2. since if c is the scalar multiple then one can add $-c$ of the one row (column) to the second to produce a new matrix with a row (column) of all zeros and the same determinant as the original matrix.

Corollary 3: If square matrix A has two equal rows (columns) then $\det A = 0$.

This follows by 3. for if we switch the identical rows (columns) then the determinant of the new matrix equals the negative of the determinant of the original matrix. However the new matrix is just the original matrix, so $\det A = -\det A$ which implies $\det A = 0$. This corollary also follows as a special case of Corollary 2 with scalar equal to one.

We can now use these procedures to simplify the calculation of the determinant as illustrated in the following examples.

Example 3-12

Compute the determinant of the given matrix.

1. $A = \begin{bmatrix} -1 & 0 & 4 & 0 \\ 0 & 1 & 3 & 1 \\ 2 & 2 & 0 & 2 \\ -3 & 1 & 1 & 1 \end{bmatrix}$

   Solution:
   $\det A = 0$ since A has two identical columns.

2. $A = \begin{bmatrix} 3 & -2 & -1 \\ -6 & -4 & 2 \\ -3 & -2 & 4 \end{bmatrix}$

   Solution:
   We perform the following row operations to simplify the matrix and consider the effect on the determinant.

   $R_2 \to R_2 + 2R_1$ (does not change determinant)
   $R_3 \to R_3 + R_1$ (no change)

$$\det A = \begin{vmatrix} 3 & -2 & -1 \\ 0 & -8 & 0 \\ 0 & -4 & 3 \end{vmatrix}$$

   $R_2 \leftrightarrow R_3$ (new determinant differs by minus sign so introduce one to preserve equality)

$$= -\begin{vmatrix} 3 & -2 & -1 \\ 0 & -4 & 3 \\ 0 & -8 & 0 \end{vmatrix}$$

   $R_3 \to R_3 - 2R_2$ (no change)

$$= -\begin{vmatrix} 3 & -2 & -1 \\ 0 & -4 & 3 \\ 0 & 0 & -6 \end{vmatrix}$$

   Matrix is now upper diagonal so $\det A = -(3)(-4)(-6) = -72$.

3. $A = \begin{bmatrix} 2 & 3 & 1 & -3 \\ 3 & 0 & -1 & 4 \\ -1 & 0 & 1 & -2 \\ 13 & 13 & 0 & -13 \end{bmatrix}$

   Solution:
   $R_4 \to \frac{1}{13} R_4$
   Note here that multiplying row 4 by $c = 1/13$ multiplies the original determinant by $1/13$ so we must multiply by $1/c = 13$ to compensate this. Effectively this looks like “factoring out 13” from the row.

$$\det A = 13 \begin{vmatrix} 2 & 3 & 1 & -3 \\ 3 & 0 & -1 & 4 \\ -1 & 0 & 1 & -2 \\ 1 & 1 & 0 & -1 \end{vmatrix}$$

$R_1 \to R_1 - 3R_4$

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

Expanding along the second column:

$$= 13(1)(+1) \begin{vmatrix} -1 & 1 & 0 \\ 3 & -1 & 4 \\ -1 & 1 & -2 \end{vmatrix}$$

$R_2 \to R_2 + 2R_3$

$$= 13 \begin{vmatrix} -1 & 1 & 0 \\ 1 & 1 & 0 \\ -1 & 1 & -2 \end{vmatrix}$$

Expand along the last column:

$$= 13(-2)(+1) \begin{vmatrix} -1 & 1 \\ 1 & 1 \end{vmatrix}$$
$$= -26(-1-1) = 52$$

4. $A = \begin{bmatrix} 0 & 1 & 2 & 0 \\ 4 & 0 & 1 & 0 \\ 0 & 3 & 2 & 1 \\ -2 & 0 & 1 & 3 \end{bmatrix}$

Solution:
We will find the determinant by placing the matrix in upper diagonal form.
$R_1 \leftrightarrow R_4$

$$\det A = - \begin{vmatrix} -2 & 0 & 1 & 3 \\ 4 & 0 & 1 & 0 \\ 0 & 3 & 2 & 1 \\ 0 & 1 & 2 & 0 \end{vmatrix}$$

$R_2 \to R_2 + 2R_1$

$$= - \begin{vmatrix} -2 & 0 & 1 & 3 \\ 0 & 0 & 3 & 6 \\ 0 & 3 & 2 & 1 \\ 0 & 1 & 2 & 0 \end{vmatrix}$$

$R_2 \leftrightarrow R_4$

$$= \begin{vmatrix} -2 & 0 & 1 & 3 \\ 0 & 1 & 2 & 0 \\ 0 & 3 & 2 & 1 \\ 0 & 0 & 3 & 6 \end{vmatrix}$$

$R_3 \to R_3 - 3R_2$

$$= \begin{vmatrix} -2 & 0 & 1 & 3 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & -4 & 1 \\ 0 & 0 & 3 & 6 \end{vmatrix}$$

$R_4 \to \frac{1}{3}R_4$

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

$R_3 \leftrightarrow R_4$

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

$R_4 \rightarrow R_4 + 4R_3$

$$= -3 \begin{vmatrix} -2 & 0 & 1 & 3 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 9 \end{vmatrix} \quad \Leftarrow \text{Upper Diagonal Matrix}$$

$$= -3(-2)(1)(1)(9) = 54$$

3.2 Adjugate of a Matrix

For any $n \times n$ matrix A, recall the i,j-cofactor of A, denoted by c_{ij} is given by:

$$c_{ij} = (-1)^{i+j} m_{ij} = (-1)^{i+j} \det A_{ij}$$

Then A has the cofactor matrix $C = [c_{ij}]$:

$$C = \begin{bmatrix} c_{11} & c_{12} & \dots & c_{1n} \\ c_{21} & c_{22} & \dots & c_{2n} \\ \vdots & \vdots & & \vdots \\ c_{n1} & c_{n2} & \dots & c_{nn} \end{bmatrix}.$$

Definition: If A is an $n \times n$ matrix, then the **adjugate** of A denoted by $\operatorname{adj} A$ is the transpose of the cofactor matrix of A.

$$\operatorname{adj} A = C^T = \begin{bmatrix} c_{11} & c_{21} & \dots & c_{n1} \\ c_{12} & c_{22} & \dots & c_{n2} \\ \vdots & \vdots & & \vdots \\ c_{1n} & c_{2n} & \dots & c_{nn} \end{bmatrix}$$

The adjugate is also known as the **adjunct** or the **classical adjoint** however the word adjoint finds different usage in linear algebra and, as such, should be avoided.

Example 3-13

Find the adjugate of A if $A = \begin{bmatrix} 2 & -3 & 1 \\ 4 & 0 & -2 \\ 3 & -1 & -3 \end{bmatrix}$.

Solution:
Recall from Example 3-5 we found the cofactor matrix to be

$$C = \begin{bmatrix} +\begin{vmatrix} 0 & -2 \\ -1 & -3 \end{vmatrix} & -\begin{vmatrix} 4 & -2 \\ 3 & -3 \end{vmatrix} & +\begin{vmatrix} 4 & 0 \\ 3 & -1 \end{vmatrix} \\ -\begin{vmatrix} -3 & 1 \\ -1 & -3 \end{vmatrix} & +\begin{vmatrix} 2 & 1 \\ 3 & -3 \end{vmatrix} & -\begin{vmatrix} 2 & -3 \\ 3 & -1 \end{vmatrix} \\ +\begin{vmatrix} -3 & 1 \\ 0 & -2 \end{vmatrix} & -\begin{vmatrix} 2 & 1 \\ 4 & -2 \end{vmatrix} & +\begin{vmatrix} 2 & -3 \\ 4 & 0 \end{vmatrix} \end{bmatrix} = \begin{bmatrix} -2 & 6 & -4 \\ -10 & -9 & -7 \\ 6 & 8 & 12 \end{bmatrix}$$

Therefore, taking the transpose, one has

$$\operatorname{adj} A = C^T = \begin{bmatrix} -2 & -10 & 6 \\ 6 & -9 & 8 \\ -4 & -7 & 12 \end{bmatrix}$$

The adjugate of a matrix has the following important property.

Theorem 3-8: If A is an $n \times n$ matrix, then:

$$\boxed{A(\operatorname{adj} A) = (\operatorname{adj} A)A = (\det A)I},$$

where I is the $n \times n$ identity matrix.

Proof:
To show $A(\operatorname{adj} A) = (\det A)I$ we note that the latter matrix is just the diagonal matrix with $\det A$ along the diagonal. The i^{th} diagonal entry of the product $A(\operatorname{adj} A)$ involves the i^{th} row of A and the i^{th} column of $\operatorname{adj} A$. But, since $\operatorname{adj} A = C^T$ this is just the i^{th} row of C and we have[2]

$$[A(\operatorname{adj} A)]_{ii} = a_{i1}c_{i1} + a_{i2}c_{i2} + \ldots + a_{in}c_{in}\,,$$

which we recognize to be $\det A$ evaluated along the i^{th} row of A. For an off-diagonal entry of the product, $i \neq j$, one would get

$$[A(\operatorname{adj} A)]_{ij} = a_{i1}c_{j1} + a_{i2}c_{j2} + \ldots + a_{in}c_{jn}\,.$$

One observes that none of the cofactors on the right hand side involve the j^{th} row of A since we delete the row and column when working out the cofactor and all the entries are in the j^{th} row. The entries a_{i1} to a_{in} which multiply the cofactors all sit in a row different from the j^{th} by assumption as well. Consider then, a new matrix B that is identical to A except in the j^{th} row which is just made to be a copy of A's i^{th} row. Then since B has two rows equal we have $\det B = 0$. However cofactor expanding along the j^{th} row of B gives

$$b_{j1}c_{j1} + b_{j2}c_{j2} + \ldots + b_{jn}c_{jn} = \det B = 0\,,$$

where the cofactor entries are identical to those of A by construction. However since the j^{th} row of B equals the i^{th} row of A we then have

$$a_{i1}c_{j1} + a_{i2}c_{j2} + \ldots + a_{in}c_{jn} = 0$$

thereby proving $[A(\operatorname{adj} A)]_{ij} = 0$ for $i \neq j$, thereby completing the proof. Similarly one may argue $(\operatorname{adj} A)A = (\det A)I$ by considering determinant column cofactor expansions.

Theorem 3-9: If A is a square matrix with $\det A \neq 0$, then A is invertible with

$$\boxed{A^{-1} = \frac{1}{\det A}(\operatorname{adj} A)}\,.$$

Proof:
Let A be square with $\det A \neq 0$. From Theorem 3-8 we have

$$A(\operatorname{adj} A) = (\operatorname{adj} A)A = (\det A)I\,.$$

Since $\det A \neq 0$ multiply each term by scalar $1/\det A$ to get

$$\frac{1}{\det A}\left[A(\operatorname{adj} A)\right] = \frac{1}{\det A}\left[(\operatorname{adj} A)A\right] = \frac{1}{\det A}\left[(\det A)I\right]\,.$$

Using the properties of scalar multiplication we have

$$A\left[\frac{1}{\det A}(\operatorname{adj} A)\right] = \left[\frac{1}{\det A}(\operatorname{adj} A)\right]A = \left[\frac{1}{\det A}(\det A)\right]I\,.$$

Since the last term simplifies to I we have $A^{-1} = \dfrac{1}{\det A}(\operatorname{adj} A)$ by definition of the inverse.

[2]Here we have introduced, for convenience, notation for the i-j^{th} entry in a matrix, namely $[A]_{ij} = a_{ij}$.

Example 3-14

Find the inverse of $A = \begin{bmatrix} 2 & -3 & 1 \\ 4 & 0 & -2 \\ 3 & -1 & -3 \end{bmatrix}$ using the adjugate.

Solution:
Evaluating the determinant of A along the second column of A gives

$$\begin{aligned} \det A &= -3(-1)\begin{vmatrix} 4 & -2 \\ 3 & -3 \end{vmatrix} + 0 + (-1)(-1)\begin{vmatrix} 2 & 1 \\ 4 & -2 \end{vmatrix} \\ &= 3(-12+6)] + 1(-4-4) \\ &= -18 - 8 \\ &= -26 \end{aligned}$$

which is nonzero so the inverse exists. From Example 3-13 we found

$$\operatorname{adj} A = \begin{bmatrix} -2 & -10 & 6 \\ 6 & -9 & 8 \\ -4 & -7 & 12 \end{bmatrix}$$

Therefore

$$A^{-1} = \frac{1}{\det A}(\operatorname{adj} A) = -\frac{1}{26}\begin{bmatrix} -2 & -10 & 6 \\ 6 & -9 & 8 \\ -4 & -7 & 12 \end{bmatrix}.$$

Example 3-15

Find the inverse of $A = \begin{bmatrix} 3 & 2 & -1 \\ 1 & 6 & 3 \\ 2 & -4 & 0 \end{bmatrix}$ using the adjugate.

Solution:
Evaluating the determinant with cofactor expansion along the third row gives

$$\begin{aligned} \det A &= 2(+1)\begin{vmatrix} 2 & -1 \\ 6 & 3 \end{vmatrix} + 0 + 4(+1)\begin{vmatrix} 3 & -1 \\ 1 & 3 \end{vmatrix} \\ &= 2(6+6) + 4(9+1) \\ &= 24 + 40 \\ &= 64 \end{aligned}$$

which is nonzero so A^{-1} exists with

$$A^{-1} = \frac{1}{\det A}\operatorname{adj} A = \frac{1}{\det A} C^T$$

$$= \frac{1}{64}\begin{bmatrix} +\begin{vmatrix} 6 & 3 \\ -4 & 0 \end{vmatrix} & -\begin{vmatrix} 1 & 3 \\ 2 & 0 \end{vmatrix} & +\begin{vmatrix} 1 & 6 \\ 2 & -4 \end{vmatrix} \\ -\begin{vmatrix} 2 & -1 \\ -4 & 0 \end{vmatrix} & +\begin{vmatrix} 3 & -1 \\ 2 & 0 \end{vmatrix} & -\begin{vmatrix} 3 & 2 \\ 2 & -4 \end{vmatrix} \\ +\begin{vmatrix} 2 & -1 \\ 6 & 3 \end{vmatrix} & -\begin{vmatrix} 3 & -1 \\ 1 & 3 \end{vmatrix} & +\begin{vmatrix} 3 & 2 \\ 1 & 6 \end{vmatrix} \end{bmatrix}^T$$

$$= \frac{1}{64}\begin{bmatrix} 12 & 6 & -16 \\ 4 & 2 & 16 \\ 12 & -10 & 16 \end{bmatrix}^T = \frac{1}{64}\begin{bmatrix} 12 & 4 & 12 \\ 6 & 2 & -10 \\ -16 & 16 & 16 \end{bmatrix} = \frac{1}{32}\begin{bmatrix} 6 & 2 & 6 \\ 3 & 1 & -5 \\ -8 & 8 & 8 \end{bmatrix}$$

Example 3-16

Find the inverse of the 2×2 matrix $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ using the adjugate.

Solution: Assuming $\det A = ad - bc \neq 0$ the inverse is

$$A^{-1} = \frac{1}{\det A}\operatorname{adj} A = \frac{1}{\det A} C^T = \frac{1}{ad - bc}\begin{bmatrix} (+1)d & (-1)c \\ (-1)b & (+1)a \end{bmatrix}^T = \frac{1}{ad - bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

as we found before.

The converse of Theorem 3-9 also holds with the result:

Theorem 3-10: A square matrix A is invertible if and only if $\det A \neq 0$. If A is invertible then

$$\boxed{\det(A^{-1}) = \frac{1}{\det A} = (\det A)^{-1}}\,.$$

Proof:
By Theorem 3-9 $\det A \neq 0$ implies A is invertible. For the converse let A be an invertible matrix. Then there exists A^{-1} satisfying

$$AA^{-1} = I\,.$$

Taking the determinant of both sides gives

$$\det(AA^{-1}) = \det I\,.$$

But the determinant of a product is the product of the determinants and the determinant of an identity matrix is one, therefore:

$$(\det A)\det\left(A^{-1}\right) = 1\,.$$

If $\det A = 0$ we have a contradiction to the last statement as the left hand side would be zero. Therefore $\det A \neq 0$. In this case we can divide both sides of the last equation by $\det A$ to get $\det\left(A^{-1}\right) = \dfrac{1}{\det A}$.

Example 3-17

Determine whether the given matrix is invertible.

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

   Solution:
   $\det A = -2 - 3 = -5 \neq 0$ therefore A is invertible.

2. $A = \begin{bmatrix} 2 & 6 \\ 1 & 3 \end{bmatrix}$

   Solution:
   $\det A = 6 - 6 = 0$ therefore A is noninvertible.

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

   Solution:
   Using the Rule of Sarrus,

$$\left|\begin{array}{ccc} -1 & 1 & -2 \\ 0 & 1 & -3 \\ 1 & 1 & -2 \end{array}\right| \begin{array}{cc} -1 & 1 \\ 0 & 1 \\ 1 & 1 \end{array}$$

$$\begin{aligned} \det A &= (-1)(1)(-2) + (1)(-3)(1) + \cancel{(-2)(0)(1)} - (-2)(1)(1) - (-1)(-3)(1) - \cancel{(1)(0)(-2)} \\ &= 2 - 3 + 2 - 3 \\ &= -2\,, \end{aligned}$$

which is nonzero therefore A is invertible.

Example 3-18

Find the values of m so that the given matrix is invertible.

$$A = \begin{bmatrix} 2 & -3 & m \\ 2 & 0 & -2 \\ 3 & -m & -3 \end{bmatrix}$$

Solution:
A is invertible if and only if $\det A \neq 0$. Evaluating $\det A$ along the second row we have

$$\begin{aligned} \det A &= \begin{vmatrix} 2 & -3 & m \\ 2 & 0 & -2 \\ 3 & -m & -3 \end{vmatrix} \\ &= 2(-1)\begin{vmatrix} -3 & m \\ -m & -3 \end{vmatrix} + 0 + (-2)(-1)\begin{vmatrix} 2 & -3 \\ 3 & -m \end{vmatrix} \\ &= -2(9 + m^2) + 2(-2m + 9) \\ &= -18 - 2m^2 - 4m + 18 \\ &= -2m(m + 2) \end{aligned}$$

Then $\det A = 0$ implies $m = 0$ or $m = -2$. Therefore A is invertible ($\det A \neq 0$) if $m \neq 0$ and $m \neq -2$.

The adjugate of a matrix is itself a matrix with determinant and potentially an inverse. This is explored in the following theorems.

Theorem 3-11: Let A be an $n \times n$ matrix with $n > 1$, then:

$$\boxed{\det(\operatorname{adj} A) = (\det A)^{n-1}}\,.$$

Proof:
By Theorem 3-8

$$(\operatorname{adj} A)A = (\det A)(I)\,.$$

Taking the determinant of both sides gives

$$\det[(\operatorname{adj} A)(A)] = \det[(\det A)(I)]\,.$$

Since $\det(AB) = \det A \det B$, $\det(cA) = c^n \det A$, and $\det I = 1$ we have

$$\det(\operatorname{adj} A) \det A = (\det A)^n \det I = (\det A)^n\,.$$

Case I: $\det A \neq 0$. Dividing through the previous equation by $\det A$ gives

$$\det(\operatorname{adj} A) = \frac{(\det A)^n}{\det A} = (\det A)^{n-1}\,,$$

and the theorem holds.

Case II: $\det A = 0$. Then $0^{n-1} = 0$ so we need to prove $\det(\operatorname{adj} A) = 0$. We will use proof by contradiction by supposing $\det(\operatorname{adj} A) \neq 0$. Then $(\operatorname{adj} A)A = (\det A)(I)$ implies

$$(\operatorname{adj} A)A = 0\,.$$

Since $\det(\operatorname{adj} A) \neq 0$ Theorem 3-9 implies the matrix $\operatorname{adj} A$ itself is invertible with inverse $(\operatorname{adj} A)^{-1}$. Left-multiplying both sides of $(\operatorname{adj} A)A = 0$ by this inverse implies $A = 0$. Then $\operatorname{adj} A = C^T = 0$ since all the cofactors of A vanish if $A = 0$. But then $\operatorname{adj} A = 0$ implies $\det(\operatorname{adj} A) = 0$ and we have a contradiction to the original supposition. Hence the opposite of the supposition must be true and $\det(\operatorname{adj} A) = 0$.

Theorem 3-12: If A is an $n \times n$ invertible matrix, then $\operatorname{adj} A$ is invertible with

$$\boxed{(\operatorname{adj} A)^{-1} = \frac{1}{\det A} A}\,.$$

Proof:
Let A be an $n \times n$ invertible matrix. Then $\det A \neq 0$ by Theorem 3-10. From Theorem 3-8 we have

$$A(\operatorname{adj} A) = (\operatorname{adj} A)A = (\det A)I\,.$$

Since $\det A \neq 0$ multiply each term by scalar $1/\det A$ to get

$$\frac{1}{\det A}\left[A(\operatorname{adj} A)\right] = \frac{1}{\det A}\left[(\operatorname{adj} A)A\right] = \frac{1}{\det A}\left[(\det A)I\right]\,.$$

Using the properties of scalar multiplication we have

$$\left[\frac{1}{\det A}A\right](\operatorname{adj} A) = (\operatorname{adj} A)\left[\frac{1}{\det A}A\right] = \left[\frac{1}{\det A}(\det A)\right] I\,.$$

Since the last term simplifies to I we have $(\operatorname{adj} A)^{-1} = \dfrac{1}{\det A} A$ by definition of the inverse.

Example 3-19

Find the determinant and inverse of the cofactor matrix C of matrix A.

Solution:
Since $\operatorname{adj} A = C^T$ we have (taking the transpose of both sides) $C = (\operatorname{adj} A)^T$, therefore

$$\det C = \det\left[(\operatorname{adj} A)^T\right] = \det(\operatorname{adj} A) = (\det A)^{n-1}\,.$$

Also since the transpose is invertible if and only if the original matrix is and $\operatorname{adj} A$ is invertible if A is, then C is invertible when A is invertible and

$$C^{-1} = \left[(\operatorname{adj} A)^T\right]^{-1} = \left[(\operatorname{adj} A)^{-1}\right]^T = \left[\frac{1}{\det A}A\right]^T = \frac{1}{\det A}A^T\,.$$

3.3 Summary of Properties of Determinants

Let A and B be $n \times n$ matrices and let c be a scalar, then:

1. $\det(AB) = (\det A)(\det B)$ (generalizable to product of k matrices)
2. $\det(AB) = \det(BA)$ (generalizable to permutation of k matrices)
3. $\det\left(A^k\right) = (\det A)^k$
4. $\det\left(A^T\right) = \det A$
5. $\det(cA) = c^n \det A$
6. $\det\left(A^{-1}\right) = \dfrac{1}{\det A} = (\det A)^{-1}$
7. $\det(\operatorname{adj} A) = (\det A)^{n-1}$

Also we have the two important inverse relations $A^{-1} = \dfrac{1}{\det A}(\operatorname{adj} A)$ and $(\operatorname{adj} A)^{-1} = \dfrac{1}{\det A} A$.

More abstract problems involving matrix determinants may be simplified using these determinant properties.

Example 3-20

Let A and B be 3×3 invertible matrices such that $\det A = 2$ and $\det B = -1$.
Compute 1. $\det\left(AB^T\right)$ 2. $\det\left(2ABA^T\right)$ 3. $\det\left[-(\operatorname{adj} A)^2 B^4\right]$ 4. $\det\left(B^{-1}A^{-1}BA^3B^{-1}\right)$.

Solution:

1. $\det\left(AB^T\right) = (\det A)(\det B^T) = (\det A)(\det B) = (2)(-1) = -2$
2. $\det\left(2ABA^T\right) = 2^3(\det A)(\det B)(\det A^T) = 8(2)(-1)(2) = -32$
3. $\det\left[-(\operatorname{adj} A)^2 B^4\right] = (-1)^3[\det(\operatorname{adj} A)]^2(\det B)^4 = -(2^{3-1})^2(-1)^4 = -16$
4. $\det\left(B^{-1}A^{-1}BA^3B^{-1}\right) = \det\left(A^{-1}AA^2B^{-1}BB^{-1}\right) = \det\left(IA^2B^{-1}I\right) = (\det A)^2(\det B)^{-1}$
$= 2^2\dfrac{1}{(-1)} = -4$

Note in question 4. how effectively one could just add the exponents of the matrices, treating the inverses as exponents of -1, to get the simplified answer.

Example 3-21

If $\det\begin{bmatrix} a & b \\ c & d \end{bmatrix} = 3$, compute $\det\left(\frac{1}{2}A^{-1}\right)$ if $A = \begin{bmatrix} a-b & 5b \\ c-d & 5d \end{bmatrix}$.

Solution:
We have $\det\left(\dfrac{1}{2}A^{-1}\right) = \left(\dfrac{1}{2}\right)^2 \det\left(A^{-1}\right) = \dfrac{1}{4\det A}$ so we need to evaluate $\det A$.

$$\det A = \begin{vmatrix} a-b & 5b \\ c-d & 5d \end{vmatrix}$$

Do the following column operations:
$C_2 \to \frac{1}{5}C_2$

$\det A = 5\begin{vmatrix} a-b & b \\ c-d & d \end{vmatrix}$

$C_1 \to C_1 + C_2$

$\det A = 5\begin{vmatrix} a & b \\ c & d \end{vmatrix} = (5)(3) = 15$

Therefore

$$\det\left(\frac{1}{2}A^{-1}\right) = \frac{1}{4\det A} = \frac{1}{(4)(15)} = \frac{1}{60}\,.$$

3.3.1 Determinants and Linear Systems

We have seen that in determined linear systems for which the number of unknowns equals the number of equations results in a square coefficient matrix A. Such matrices have determinants and, as such, our knowledge of determinants can directly inform this restricted class of linear systems. Despite this limitation to linear systems of n equations in n unknowns, it should be noted that many applications naturally result in such systems.

Since matrix A is invertible if and only if its determinant is nonzero we have the following result regarding linear systems which follows from Theorem 2-18.

Theorem 3-13: Let A be a square matrix. Then the determined linear system $A\boldsymbol{x} = \boldsymbol{b}$ has a unique solution if and only if $\det A \neq 0$.

Example 3-22

For what values of m does the following linear system have:

1. A unique solution?
2. No solution?
3. Infinitely many solutions?

$$\begin{aligned} x - my + 4z &= 2 \\ mx - 2m^2y + (2m^3 + 4m)z &= 2m \\ -2x + (1+2m)y - 9z &= -3 \end{aligned}$$

Solution:

The coefficient matrix is $A = \begin{bmatrix} 1 & -m & 4 \\ m & -2m^2 & 2m^3+4m \\ -2 & 1+2m & -9 \end{bmatrix}$.

Expanding the determinant along the first row gives:

$$\begin{aligned} \det A &= (1)(+1)\begin{vmatrix} -2m^2 & 2m^2+4m \\ 1+2m & -9 \end{vmatrix} + (-m)(-1)\begin{vmatrix} m & 2m^3+4m \\ -2 & -9 \end{vmatrix} + 4(+1)\begin{vmatrix} m & -2m^2 \\ -2 & 1+2m \end{vmatrix} \\ &= 18m^2 - (2m^3+4m)(1+2m) + m[-9m + 2(2m^3+4m)] + 4[m(1+2m) - 4m^2] \\ &= 18m^2 - 2m^3 - 4m^4 - 4m - 8m^2 - 9m^2 + 4m^4 + 8m^2 + 4m + 8m^2 - 16m^2 \\ &= -2m^3 + m^2 \end{aligned}$$

Then $0 = \det A = -2m^3 + m^2 = m^2(-2m+1)$ if and only if $m = 0$ or $m = 1/2$.

Since a unique solution occurs when $\det A \neq 0$ this is when $m \neq 0$ and $m \neq 1/2$.

When $m = 0$ the augmented matrix for the system is

$$[A|B] = \left[\begin{array}{rrr|r} 1 & 0 & 4 & 2 \\ 0 & 0 & 0 & 0 \\ -2 & 1 & -9 & -3 \end{array}\right]$$

$$\Downarrow$$

$$R_2 \leftrightarrow R_3 \left[\begin{array}{rrr|r} 1 & 0 & 4 & 2 \\ -2 & 1 & -9 & -3 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

$$\Downarrow$$

$$R_2 \to R_2 + 2R_1 \left[\begin{array}{rrr|r} 1 & 0 & 4 & 2 \\ 0 & 1 & -1 & 1 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

Then $\text{rank}(A) = \text{rank}([A|B]) = 2$ is less than the number of unknowns (3). Therefore $m = 0$ gives infinitely many solutions.

When $m = 1/2$ the augmented matrix for the system is

$$[A|B] = \left[\begin{array}{rrr|r} 1 & -\frac{1}{2} & 4 & 2 \\ \frac{1}{2} & -\frac{1}{2} & \frac{9}{4} & 1 \\ -2 & 2 & -9 & -3 \end{array}\right]$$

$$\Downarrow$$

$$\begin{array}{l} R_1 \to 2R_1 \\ R_2 \to 4R_2 \end{array} \left[\begin{array}{rrr|r} 2 & -1 & 8 & 4 \\ 2 & -2 & 9 & 4 \\ -2 & 2 & -9 & -3 \end{array}\right]$$

$$\Downarrow$$

$$\begin{array}{l} R_2 \to R_2 - R_1 \\ R_3 \to R_3 + R_1 \end{array} \left[\begin{array}{rrr|r} 2 & -1 & 8 & 4 \\ 0 & -1 & 1 & 0 \\ 0 & 1 & -1 & 1 \end{array}\right]$$

$$\Downarrow$$

$$R_3 \to R_3 + R_2 \left[\begin{array}{rrr|r} 2 & -1 & 8 & 4 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right]$$

Here $2 = \text{rank}(A) < \text{rank}([A|B]) = 3$ or, equivalently, the bottom row yields a contradiction. Therefore $m = \frac{1}{2}$ gives no solution.

Note that this example shows that for almost any value of m the determined system produces a unique solution. The events of the determined system having no solution or an infinite number of solutions are exceptional. In linear systems arising from physical problems one should suspect that the problem has some symmetry or constraint when such exceptional cases arise.

3.4 Cramer's Rule

Let $A\boldsymbol{x} = \boldsymbol{b}$ be a determined linear system of n equations in n unknowns where $\det A \neq 0$ and:

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix} \qquad \boldsymbol{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}.$$

Define $A(1)$ to be the matrix given by

$$A(1) = \begin{bmatrix} b_1 & a_{12} & \dots & a_{1n} \\ b_2 & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots \\ b_n & a_{n2} & \dots & a_{nn} \end{bmatrix},$$

i.e. Replace the first column of A by the right hand side $\boldsymbol{b}$. If we denote $A = [\boldsymbol{a}_1 \boldsymbol{a}_2 \cdots \boldsymbol{a}_n]$ where $\boldsymbol{a}_i$ is a column vector of A, then

$$A(1) = [\boldsymbol{b}\boldsymbol{a_2} \cdots \boldsymbol{a_n}].$$

In general let $A(i)$ denote the matrix given by replacing the i^{th} column of A, $\boldsymbol{a}_i$, by $\boldsymbol{b}$:

$$A(i) = \begin{bmatrix} a_{11} & \dots & a_{1,i-1} & b_1 & a_{1,i+1} & \dots & a_{1n} \\ a_{21} & \dots & a_{2,i-1} & b_2 & a_{2,i+1} & \dots & a_{2n} \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ a_{n1} & \dots & a_{n,i-1} & b_n & a_{n,i+1} & \dots & a_{nn} \end{bmatrix} = [\boldsymbol{a}_1 \cdots \boldsymbol{a}_{i-1} \boldsymbol{b} \boldsymbol{a}_{i+1} \cdots \boldsymbol{a}_n]$$

Theorem 3-14: Let $A\boldsymbol{x} = \boldsymbol{b}$ be a determined linear system of n equations in n unknowns where $\det A \neq 0$. The unique solution $\boldsymbol{x} = [x_1, \dots, x_n]^T$ to the system is then given by

$$x_1 = \frac{\det A(1)}{\det A} \qquad \dots \qquad x_i = \frac{\det A(i)}{\det A} \qquad \dots \qquad x_n = \frac{\det A(n)}{\det A}.$$

This is called **Cramer's Rule**.

Proof:
Since $\det A \neq 0$ A is invertible and $A\boldsymbol{x} = \boldsymbol{b}$ has a unique solution given by $\boldsymbol{x} = A^{-1}\boldsymbol{b}$. Evaluating the latter gives:

$$\begin{aligned} \boldsymbol{x} &= A^{-1}\boldsymbol{b} \\ &= \frac{1}{\det A}(\operatorname{adj} A)\boldsymbol{b} \\ &= \frac{1}{\det A} C^T \boldsymbol{b} \\ &= \frac{1}{\det A} \begin{bmatrix} c_{11} & c_{21} & \dots & c_{n1} \\ c_{12} & c_{22} & \dots & c_{n2} \\ \vdots & \vdots & & \vdots \\ c_{1n} & c_{2n} & \dots & c_{nn} \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} \end{aligned}$$

Multiplying out the right hand side gives

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \frac{1}{\det A} \begin{bmatrix} b_1 c_{11} + b_2 c_{21} + \ldots + b_n c_{n1} \\ b_1 c_{12} + b_2 c_{22} + \ldots + b_n c_{n2} \\ \vdots \\ b_1 c_{1n} + b_2 c_{2n} + \ldots + b_n c_{nn} \end{bmatrix}$$

Equating the i^{th} entry on each side gives

$$x_i = \frac{1}{\det A}(b_1c_{1i} + b_2c_{2i} + \ldots + b_nc_{ni}) \qquad (i = 1, \ldots, n)\,.$$

But evaluating the determinant of $A(i)$ along the i^{th} column gives

$$\det A(i) = b_1c_{1i} + b_2c_{2i} + \ldots + b_nc_{ni}\,.$$

Therefore

$$x_i = \frac{\det A(i)}{\det A} \qquad (i = 1, \ldots, n)\,.$$

Note Cramer's Rule is not an efficient method for solving a linear system compared to Gaussian elimination. It does have theoretical utility, however, as it gives a closed form for the solution of a determined system having a unique solution.

Example 3-23

Solve the linear system using Cramer's Rule if possible.

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

Solution:

$$A = \begin{bmatrix} 3 & 2 & 3 \\ -2 & -4 & 2 \\ 2 & 0 & 3 \end{bmatrix} \qquad \det A = 2(+1)(4+12) + 3(+1)(-12+4) = 32 - 24 = 8$$

$\det A \neq 0$ so Cramer's Rule will work to find the unique solution.

$$A(1) = \begin{bmatrix} 4 & 2 & 3 \\ -12 & -4 & 2 \\ 0 & 0 & 3 \end{bmatrix} \qquad \det A(1) = 3(+1)(-16+24) = 24$$

$$A(2) = \begin{bmatrix} 3 & 4 & 3 \\ -2 & -12 & 2 \\ 2 & 0 & 3 \end{bmatrix} \qquad \det A(2) = 2(+1)(8+36) + 3(+1)(-36+8) = 88 - 84 = 4$$

$$A(3) = \begin{bmatrix} 3 & 2 & 4 \\ -2 & -4 & -12 \\ 2 & 0 & 0 \end{bmatrix} \qquad \det A(3) = 2(+1)(-24+16) = -16$$

Using Cramer's Rule we have the solution:

$$x = \frac{\det A(1)}{\det A} = \frac{24}{8} = 3 \qquad y = \frac{\det A(2)}{\det A} = \frac{4}{8} = \frac{1}{2} \qquad z = \frac{\det A(3)}{\det A} = \frac{-16}{8} = -2\,.$$

Example 3-24

Solve the linear system using Cramer's Rule if possible.

$$\begin{aligned} 3x_1 + x_2 &= -1 \\ -2x_1 - 4x_2 + 3x_3 &= 2 \\ -x_1 - 7x_2 + 6x_3 &= 3 \end{aligned}$$

Solution:

$$A = \begin{bmatrix} 3 & 1 & 0 \\ -2 & -4 & 3 \\ -1 & -7 & 6 \end{bmatrix} \qquad \det A = 3(+1)(-24+21) + (1)(-1)(-12+3) = -9+9 = 0$$

The determined system does not have a unique solution and Cramer's Rule cannot be applied. Reducing the augmented matrix gives

$$[A|B] = \left[\begin{array}{ccc|c} 3 & 1 & 0 & -1 \\ -2 & -4 & 3 & 2 \\ -1 & -7 & 6 & 3 \end{array}\right]$$

$$\Downarrow$$

$$R_1 \leftrightarrow R_3 \left[\begin{array}{ccc|c} -1 & -7 & 6 & 3 \\ -2 & -4 & 3 & 2 \\ 3 & 1 & 0 & -1 \end{array}\right]$$

$$\Downarrow$$

$$\begin{array}{r} R_1 \to -R_1 \\ R_2 \to R_2 - 2R_1 \\ R_3 \to R_3 + 3R_1 \end{array} \left[\begin{array}{ccc|c} 1 & 7 & -6 & -3 \\ 0 & 10 & -9 & -4 \\ 0 & -20 & 18 & 8 \end{array}\right]$$

$$\Downarrow$$

$$\begin{array}{r} \\ R_2 \to \frac{1}{10}R_2 \\ R_3 \to R_3 + 2R_2 \end{array} \left[\begin{array}{ccc|c} 1 & 7 & -6 & -3 \\ 0 & 1 & -\frac{9}{10} & -\frac{4}{10} \\ 0 & 0 & 0 & 0 \end{array}\right]$$

$$\Downarrow$$

$$R_1 \to R_1 - 7R_2 \left[\begin{array}{ccc|c} 1 & 0 & \frac{3}{10} & -\frac{2}{10} \\ 0 & 1 & -\frac{9}{10} & -\frac{4}{10} \\ 0 & 0 & 0 & 0 \end{array}\right]$$

The system has infinitely many solutions. The final augmented matrix corresponds to the equivalent linear system:

$$\begin{aligned} x_1 + \frac{3}{10}x_3 &= -\frac{2}{10} \\ x_2 - \frac{9}{10}x_3 &= -\frac{4}{10} \\ 0 &= 0 \end{aligned}$$

Setting free (independent) variable $\boxed{x_3 = s}$ we then solve for the leading variables by back-substitution:

- $x_2 - \frac{9}{10}x_3 = -\frac{4}{10} \implies x_2 - \frac{9}{10}s = -\frac{4}{10} \implies \boxed{x_2 = -\frac{4}{10} + \frac{9}{10}s}$
- $x_1 + \frac{3}{10}x_3 = -\frac{2}{10} \implies x_1 + \frac{3}{10}s = -\frac{2}{10} \implies \boxed{x_1 = -\frac{2}{10} - \frac{3}{10}s}$

The general solution is therefore

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -\frac{2}{10} \\ -\frac{4}{10} \\ 0 \end{bmatrix} + s \begin{bmatrix} -\frac{3}{10} \\ \frac{9}{10} \\ 1 \end{bmatrix},$$

where s is a parameter.

Example 3-25

Solve the linear system using Cramer's Rule if possible.

$$\begin{aligned} x_1 + 2x_3 &= 6 \\ -3x_1 + 4x_2 + 6x_3 &= 30 \\ -x_1 - 2x_2 + 3x_3 &= 8 \end{aligned}$$

Solution:

$$A = \begin{bmatrix} 1 & 0 & 2 \\ -3 & 4 & 6 \\ -1 & -2 & 3 \end{bmatrix} \qquad \det A = (1)(+1)(12+12) + 2(+1)(6+4) = 24 + 20 = 44 \neq 0$$

$$A(1) = \begin{bmatrix} 6 & 0 & 2 \\ 30 & 4 & 6 \\ 8 & -2 & 3 \end{bmatrix} \qquad \det A(1) = 6(+1)(12+12) + 2(+1)(-60-32) = 144 - 184 = -40$$

$$A(2) = \begin{bmatrix} 1 & 6 & 2 \\ -3 & 30 & 6 \\ -1 & 8 & 3 \end{bmatrix} \qquad \det A(2) = 1(+1)(90-48) + 6(-1)(-9+6) + 2(+1)(-24+30)$$

$$= 42 + 18 + 12 = 72$$

$$A(3) = \begin{bmatrix} 1 & 0 & 6 \\ -3 & 4 & 30 \\ -1 & -2 & 8 \end{bmatrix} \qquad \det A(3) = 1(+1)(32+60) + 6(+1)(6+4) = 92 + 60 = 152$$

Using Cramer's Rule we have the solution:

$$x_1 = \frac{\det A(1)}{\det A} = -\frac{40}{44} = -\frac{10}{11} \qquad x_2 = \frac{\det A(2)}{\det A} = \frac{72}{44} = \frac{18}{11} \qquad x_3 = \frac{\det A(3)}{\det A} = \frac{152}{44} = \frac{38}{11}.$$

Example 3-26

In the following system find the value(s) of a such that a solution exists with $y = 0$.

$$\begin{aligned} x - 3z &= 1 \\ ax + 2y &= 0 \\ y + z &= a \end{aligned}$$

Solution:
We are solving only for y so we need only consider A and $A(2)$.

$$A = \begin{bmatrix} 1 & 0 & -3 \\ a & 2 & 0 \\ 0 & 1 & 1 \end{bmatrix} \qquad \det A = 1(+1)(2-0) - 3(+1)(a-0) = 2 - 3a$$

A unique solution exists if $2 - 3a \neq 0$, and so $a \neq \frac{2}{3}$.

$$A(2) = \begin{bmatrix} 1 & 1 & -3 \\ a & 0 & 0 \\ 0 & a & 1 \end{bmatrix} \qquad \det A(2) = a(-1)(1+3a) = -a(1+3a)$$

Using Cramer's Rule:

$$0 = y = \frac{\det A(2)}{\det A} = \frac{-a(1+3a)}{2-3a}.$$

The solution requires the numerator vanish and the denominator be non-zero:

$$-a(1+3a) = 0 \text{ and } 2 - 3a \neq 0$$

Thus ($a = 0$ or $1 + 3a = 0$) and $a \neq 2/3$, or equivalently ($a = 0$ or $a = -1/3$) and $a \neq 2/3$ which reduces logically to just the solution $a = 0$ or $a = -1/3$. In the case $a = 2/3$ there are potentially solutions with $y = 0$ as this corresponds to the case where $\det A = 0$ and Cramer's Rule could not be applied. Putting that value of a into the original system creates a system with augmented matrix

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

The reader may verify this is an inconsistent system (no solution).

Chapter 4: Vectors

4.1 Vectors in $\mathbb{R}^n$

Recall when we introduced the ordered n-tuple $(x_1, x_2, \ldots, x_n)$ in $\mathbb{R}^n$ we labelled it $\boldsymbol{x}$ and represented it by the column matrix

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = [x_1\, x_2\, \ldots\, x_n]^T$$

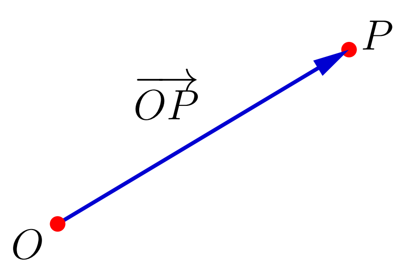
At that time we mentioned it could be used to represent a vector. We now make more precise the concept of a vector.

Definition: A **vector** is a quantity having both a **magnitude** and **direction**.

Physical examples of vectors are displacement which has both a magnitude (distance travelled) as well as a direction (like east) associated with it. Other important vectors include velocity, acceleration, momentum, and force. Notationally vectors will be identified by boldfaced lower case letters in this text such as $\boldsymbol{u}$. In handwritten form it is common to put an arrow on top of the letter or a bar above or below the letter, i.e. $\vec{u}$, $\bar{u}$, or $\underline{u}$.

Geometrically a vector quantity is an arrow having length and direction. To characterize them mathematically we introduce the related concept of a directed line segment. We visualize the following in two or three dimensions but we can generalize to n dimensions.

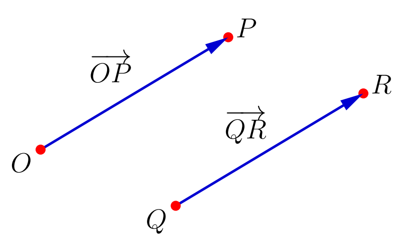
Definition: Let O and P be two points. The line segment from point O (called the **tail** or **initial point**) to the point P (called the **tip** or **terminal point**) is called the **directed line segment** from O to P and is denoted by the symbol $\overrightarrow{OP}$.



A directed line segment is almost the same as a vector except a vector is independent of its position in space. With that in mind we define the equivalence of two directed line segments as follows.

Definition: Two directed line segments $\overrightarrow{OP}$ and $\overrightarrow{QR}$ are said to be **equivalent** if they have the same direction and length.

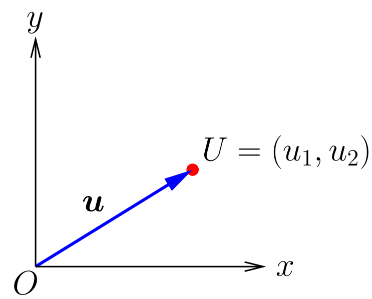
In the following diagram $\overrightarrow{OP}$ is equivalent to $\overrightarrow{QR}$.

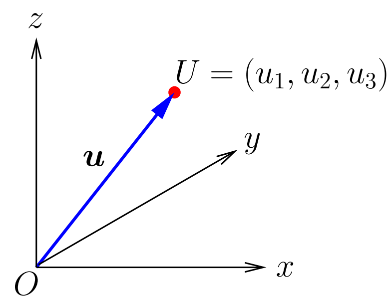


If we fix a point in space, O, call it the **origin**, then every directed line segment will be equivalent to some directed line segment whose initial point is O. Any quantity having magnitude and direction can also be represented by one such directed line segment. Hence an alternate working definition for a vector is as follows:[1]

Definition: A **vector** is a directed line segment from the origin O to a point P.

The connection to ordered n-tuples now comes about by introducing a Cartesian coordinate system. A vector can be completely characterized by the coordinates of the terminal (tip) point P. In two and three dimensions, now using the same letter for the tip point U as the vector $\boldsymbol{u}$, this becomes:





In summary:

- In $\mathbb{R}^2$ a vector is written as $\boldsymbol{u} = (u_1, u_2) = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$.
- In $\mathbb{R}^3$ a vector is written as $\boldsymbol{u} = (u_1, u_2, u_3) = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix}$.
- In $\mathbb{R}^n$ a vector is written as $\boldsymbol{u} = (u_1, u_2, \ldots, u_n) = [u_1 \, u_2 \, \ldots \, u_n]^T$.

The numbers $u_1, u_2, \ldots, u_n$ are called the **components** of the vector.

Vector equality can now be defined in terms of components.

Definition: Two vectors, $\boldsymbol{u}$ and $\boldsymbol{v}$ in $\mathbb{R}^n$ are equal if their corresponding components are equal, i.e.

$$u_i = v_i \quad \text{for } i = 1, \ldots, n\,.$$

Definition: The **zero vector** in $\mathbb{R}^n$ has all components equal to zero.

- In $\mathbb{R}^2$: $\mathbf{0} = (0, 0)$.
- In $\mathbb{R}^3$: $\mathbf{0} = (0, 0, 0)$.
- In $\mathbb{R}^n$: $\mathbf{0} = (\underbrace{0, 0, \ldots, 0}_{n \text{ times}})$.

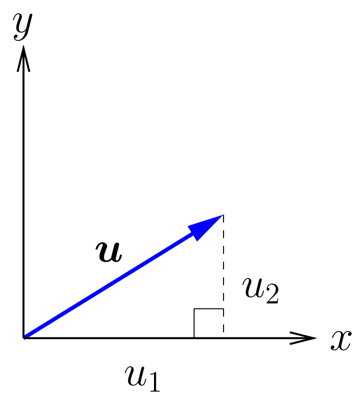
[1]To be even more precise we could characterize a vector as the unique representative of the equivalence class of directed line segments whose tail is at the origin O.

4.1.1 Vector Length

Definition: The **length** (or **norm** or **magnitude**) of a vector $\boldsymbol{u}$ is the distance from the origin to the terminal point of $\boldsymbol{u}$ and is denoted by $\|\boldsymbol{u}\|$. In terms of the components of $\boldsymbol{u}$,

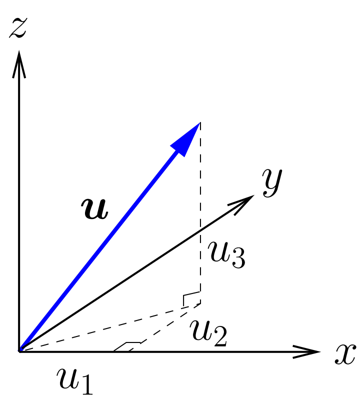
- In $\mathbb{R}^2$: $\|\boldsymbol{u}\| = \sqrt{u_1^2 + u_2^2}$.
- In $\mathbb{R}^3$: $\|\boldsymbol{u}\| = \sqrt{u_1^2 + u_2^2 + u_3^2}$.
- In $\mathbb{R}^n$: $\|\boldsymbol{u}\| = \sqrt{u_1^2 + u_2^2 + \ldots + u_n^2}$.

Here the length in $\mathbb{R}^2$ is as expected from the Pythagorean Theorem.



In $\mathbb{R}^3$ the length $\|\boldsymbol{u}\|$ is the length of the hypotenuse of a triangle the base of which lies in the x-y plane having length $\sqrt{u_1^2 + u_2^2}$ and whose height is $|u_3|$. It follows, again by the Pythagorean Theorem, that

$$\|\boldsymbol{u}\| = \sqrt{\left(\sqrt{u_1^2 + u_2^2}\right)^2 + |u_3|^2} = \sqrt{u_1^2 + u_2^2 + u_3^2}.$$



Example 4-1

Find the length $\|\boldsymbol{u}\|$ of the given vector.

1. $\boldsymbol{u} = (-1, 1)$

   Solution:
   $\|\boldsymbol{u}\| = \sqrt{(-1)^2 + 1^2} = \sqrt{2}$

2. $\boldsymbol{u} = \begin{bmatrix} 2 \\ -1 \\ -3 \end{bmatrix}$

   Solution:
   $\|\boldsymbol{u}\| = \sqrt{2^2 + (-1)^2 + (-3)^2} = \sqrt{14}$

Only the zero vector has a vanishing length.

Theorem 4-1: $\|\boldsymbol{u}\| = 0$ if and only if $\boldsymbol{u} = \mathbf{0}$.

Definition: If the length of a vector is 1, $\|\boldsymbol{u}\| = 1$, then the vector $\boldsymbol{u}$ is called a **unit vector**.

Example 4-2

Show the vector $\boldsymbol{u} = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0\right)$ in $\mathbb{R}^3$ is a unit vector.

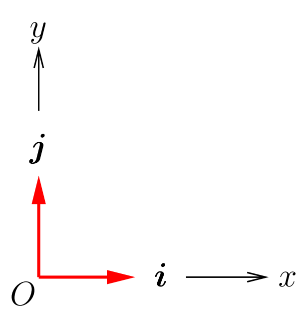
Solution:
Finding the length of $\boldsymbol{u}$,

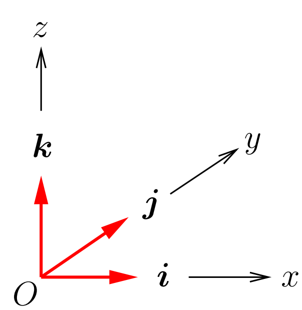
$$\|\boldsymbol{u}\| = \sqrt{u_1^2 + u_2^2 + u_3^2} = \sqrt{\left(\frac{1}{\sqrt{2}}\right)^2 + \left(\frac{1}{\sqrt{2}}\right)^2 + 0^2} = \sqrt{\frac{1}{2} + \frac{1}{2} + 0} = \sqrt{1} = 1\,,$$

shows $\boldsymbol{u}$ has unit length and hence is a unit vector.

It is convenient to define unit vectors along the direction of the positive coordinate axes as follows.

- In $\mathbb{R}^2$:
  - The unit vector along the positive x-axis is $\boldsymbol{i} = (1, 0)$.
  - The unit vector along the positive y-axis is $\boldsymbol{j} = (0, 1)$.
- In $\mathbb{R}^3$:
  - The unit vector along the positive x-axis is $\boldsymbol{i} = (1, 0, 0)$.
  - The unit vector along the positive y-axis is $\boldsymbol{j} = (0, 1, 0)$.
  - The unit vector along the positive z-axis is $\boldsymbol{k} = (0, 0, 1)$.





For $\mathbb{R}^n$ it is convenient to generalize this approach.

Definition: An **elementary vector** in $\mathbb{R}^n$ is a vector that has one component equal to 1 and all other components equal to 0. If the 1 occurs as the i^{th} component, then the elementary vector is denoted by $\boldsymbol{e}_i$.

- In $\mathbb{R}^2$:
  - $\boldsymbol{e}_1 = (1,0) = \boldsymbol{i}$.
  - $\boldsymbol{e}_2 = (0,1) = \boldsymbol{j}$.
- In $\mathbb{R}^3$:
  - $\boldsymbol{e}_1 = (1,0,0) = \boldsymbol{i}$.
  - $\boldsymbol{e}_2 = (0,1,0) = \boldsymbol{j}$.
  - $\boldsymbol{e}_3 = (0,0,1) = \boldsymbol{k}$.

Note that when writing unit vectors by hand a common convention is to write $\hat{\imath}$, $\hat{\jmath}$, $\hat{k}$, $\hat{e}_1$, etc. to indicate the vector has unit length.

4.1.2 Scalar Multiplication

Recall we have used the term **scalar** to refer to an element of the set of real numbers $\mathbb{R}$. Scalars physically represent quantities such as temperature or length which do not have a direction.[2] We can define multiplication of a vector by a scalar.

Definition: Let a be a scalar and $\boldsymbol{u}$ be a vector. The scalar multiple of $\boldsymbol{u}$ by a is a vector given by:

- In $\mathbb{R}^2$: $a\boldsymbol{u} = (au_1, au_2)$
- In $\mathbb{R}^3$: $a\boldsymbol{u} = (au_1, au_2, au_3)$
- In $\mathbb{R}^n$: $a\boldsymbol{u} = (au_1, au_2, \ldots, au_n)$

Note since a vector written as a column matrix is just a matrix, this is consistent with our previous definition of scalar multiplication of a matrix.

Example 4-3

Find $a\boldsymbol{u}$ for the given scalar and vector.

1. $a = -2$, $\boldsymbol{u} = (-1, 2)$

   Solution:

   $a\boldsymbol{u} = -2\boldsymbol{u} = ((-2)(-1), (-2)(2)) = (2, -4)$

2. $a = 3$, $\boldsymbol{u} = \begin{bmatrix} 2 \\ 3 \\ -1 \end{bmatrix}$

   Solution:

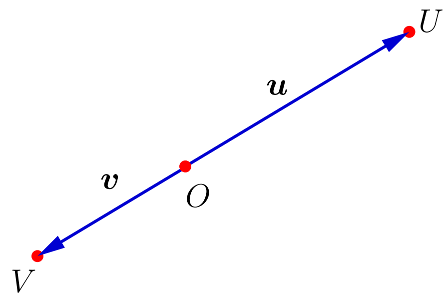
   $a\boldsymbol{u} = 3\boldsymbol{u} = \begin{bmatrix} 6 \\ 9 \\ -3 \end{bmatrix}$

[2] At a more technical level, scalars represent quantities that are numbers that are independent of the choice of coordinate system. So, for instance, the value of the first coordinate of a vector, while being a number, would not be a scalar as a different choice of axes could be made which would make the first coordinate have a different value. The length of a vector however would be independent of such a choice and is a proper scalar.

3. $a = 2$, $\boldsymbol{u} = (-1, 1, 3, 5)$

   Solution:
   $a\boldsymbol{u} = 2\boldsymbol{u} = (-2, 2, 6, 10)$

Definition: Two nonzero vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ are **collinear** if the lines they determine are one and the same. i.e. if their terminal points U and V along with the origin O are collinear.



Theorem 4-2: Let $\boldsymbol{u}$ be a nonzero vector and let a be a scalar. Then $\boldsymbol{u}$ and $a\boldsymbol{u}$ are collinear and:

1. If $a > 0$, then $\boldsymbol{u}$ and $a\boldsymbol{u}$ have the same direction (parallel).
2. If $a < 0$, then $\boldsymbol{u}$ and $a\boldsymbol{u}$ have opposite direction (antiparallel).
3. $\|a\boldsymbol{u}\| = |a| \, \|\boldsymbol{u}\|$

If nonzero vector $\boldsymbol{u}$ is not a unit vector we can find a unit vector along the same direction as $\boldsymbol{u}$ by multiplying it by the scalar that is the reciprocal of its length.

Theorem 4-3: Let $\boldsymbol{u}$ be a nonzero vector then a **unit vector** in the same direction as $\boldsymbol{u}$ is given by:

$$\frac{1}{\|\boldsymbol{u}\|}\boldsymbol{u}\,.$$

Proof:
Since $\boldsymbol{u} \neq \boldsymbol{0}$, Theorem 4-1 shows $\|\boldsymbol{u}\| \neq 0$. Then using Theorem 4-2 with positive scalar $a = 1/\|\boldsymbol{u}\|$ shows $a\boldsymbol{u}$ is directed along $\boldsymbol{u}$ and has length

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

Example 4-4

Find a unit vector parallel to the given vector.

1. $\boldsymbol{u} = (1, -1, 2)$

   Solution:
   $\|\boldsymbol{u}\| = \sqrt{1^2 + (-1)^2 + 2^2} = \sqrt{6}$
   The unit vector is $\dfrac{1}{\|\boldsymbol{u}\|}\boldsymbol{u} = \dfrac{1}{\sqrt{6}}(1, -1, 2)\,.$

2. $\boldsymbol{u} = (-1, 0, -2, 2)$

   Solution:
   $\|\boldsymbol{u}\| = \sqrt{(-1)^2 + 0 + (-2)^2 + 2^2} = \sqrt{9} = 3$
   The unit vector is $\dfrac{1}{\|\boldsymbol{u}\|}\boldsymbol{u} = \dfrac{1}{3}(-1, 0, -2, 2)$.

4.1.3 Vector Addition

Definition: Let $\boldsymbol{u}$ and $\boldsymbol{v}$ be nonzero vectors in $\mathbb{R}^2$, $\mathbb{R}^3$ or $\mathbb{R}^n$. The sum of $\boldsymbol{u}$ and $\boldsymbol{v}$ is the **resultant vector** $\boldsymbol{u} + \boldsymbol{v}$ given by:

- In $\mathbb{R}^2$: $\boldsymbol{u} + \boldsymbol{v} = (u_1 + v_1, u_2 + v_2)$.
- In $\mathbb{R}^3$: $\boldsymbol{u} + \boldsymbol{v} = (u_1 + v_1, u_2 + v_2, u_3 + v_3)$.
- In $\mathbb{R}^n$: $\boldsymbol{u} + \boldsymbol{v} = (u_1 + v_1, u_2 + v_2, \ldots, u_n + v_n)$.

Thus vector addition is done componentwise. This is consistent with matrix addition and our identification of vectors with column matrices.

Example 4-5

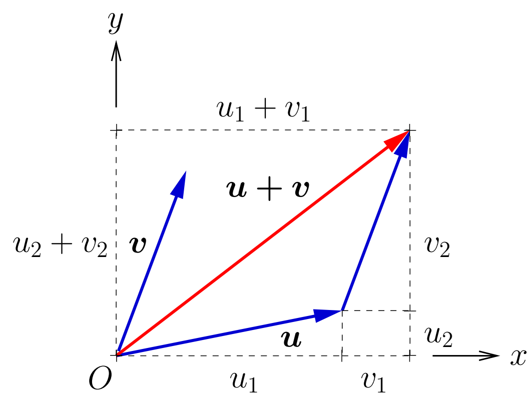
Find the sum of the given vectors.

1. $\boldsymbol{u} = (1, -1, 2)$, $\boldsymbol{v} = (-5, 4, 3)$
   Solution:
   $\boldsymbol{u} + \boldsymbol{v} = (1 - 5, -1 + 4, 2 + 3) = (-4, 3, 5)$

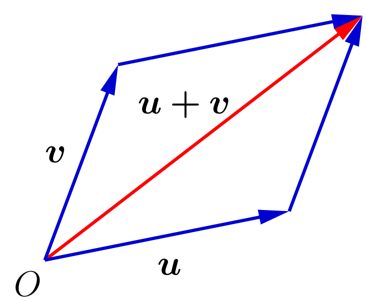
2. $\boldsymbol{u} = \begin{bmatrix} 3 \\ -1 \\ 3 \\ 2 \end{bmatrix}$, $\boldsymbol{v} = \begin{bmatrix} -1 \\ 2 \\ -2 \\ 3 \end{bmatrix}$

   Solution:
   $$\boldsymbol{u} + \boldsymbol{v} = \begin{bmatrix} 3 + (-1) \\ -1 + 2 \\ 3 + (-2) \\ 2 + 3 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \\ 1 \\ 5 \end{bmatrix}$$

The componentwise addition of vectors has the following geometric interpretation in two and three dimensions. The resultant vector arising from adding two (or more) vectors is the vector formed by joining the vectors successively from tip to tail as shown in the following diagram.



Here the vector $\boldsymbol{v}$ is shown with its tail translated from the origin O to the tip of $\boldsymbol{u}$. If we remove the coordinate system scaffolding we see that vector addition is a geometrical property independent of coordinates. Noting that our above definition implies that $\boldsymbol{v} + \boldsymbol{u} = \boldsymbol{u} + \boldsymbol{v}$ (vector addition commutes) the resultant can be found by placing the vector $\boldsymbol{v}$ first as well and one has the following diagram:



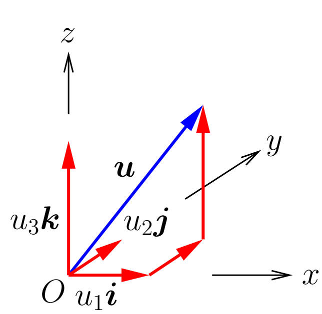
As the diagram shows the two vectors with tails placed at the origin form a parallelogram, the diagonal of which is the resultant vector with tail placed at the origin. The is known as the **parallelogram law** of vector addition. In three dimensions the two vectors determine a plane in which the parallelogram lies.

4.1.4 Vector Components Along Coordinate Axes

Consider the vector $\boldsymbol{u} = (u_1, u_2, u_3) = (3, 2, 4)$. Then

$$\boldsymbol{u} = \begin{bmatrix} 3 \\ 2 \\ 4 \end{bmatrix} = \begin{bmatrix} 3 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 2 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 4 \end{bmatrix} = 3 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + 2 \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + 4 \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = 3\boldsymbol{i} + 2\boldsymbol{j} + 4\boldsymbol{k}\,.$$

Here $3\boldsymbol{i}$ is called the **vector component** of $\boldsymbol{u}$ along $\boldsymbol{i}$ and geometrically the vector $\boldsymbol{u}$ is the sum of these vectors directed along the coordinate axes.



Generalizing, we have the following decomposition of vectors into their vector components along coordinate axes.

- In $\mathbb{R}^2$: $\boldsymbol{u} = (u_1, u_2) = u_1\boldsymbol{i} + u_2\boldsymbol{j}$.
- In $\mathbb{R}^3$: $\boldsymbol{u} = (u_1, u_2, u_3) = u_1\boldsymbol{i} + u_2\boldsymbol{j} + u_3\boldsymbol{k}$.
- In $\mathbb{R}^n$: $\boldsymbol{u} = (u_1, u_2, \ldots, u_n) = u_1\boldsymbol{e}_1 + u_2\boldsymbol{e}_2 + \cdots + u_n\boldsymbol{e}_n$.

Example 4-6

One can go back and forth between vector components as shown in the following examples:

1. $\boldsymbol{u} = (-1, 2) = -\boldsymbol{i} + 2\boldsymbol{j}$
2. $\boldsymbol{u} = \begin{bmatrix} 3 \\ -4 \\ 5 \\ 2 \end{bmatrix} = 3\boldsymbol{e}_1 - 4\boldsymbol{e}_2 + 5\boldsymbol{e}_3 + 2\boldsymbol{e}_4$
3. $\boldsymbol{u} = 2\boldsymbol{i} + 3\boldsymbol{j} - \boldsymbol{k} = (2, 3, -1)$

In physical problems writing vectors in terms of elementary vectors can often simplify calculations involving the vectors.

4.1.5 Vector Subtraction

Definition: Let $\boldsymbol{u}$ be a vector. The **negative of** $\boldsymbol{u}$, denoted $-\boldsymbol{u}$, is defined to be $(-1)\boldsymbol{u}$.

- In $\mathbb{R}^2$: $-\boldsymbol{u} = (-u_1, -u_2)$
- In $\mathbb{R}^3$: $-\boldsymbol{u} = (-u_1, -u_2, -u_3)$
- In $\mathbb{R}^n$: $-\boldsymbol{u} = (-u_1, -u_2, \ldots, -u_n)$

Theorem 4-2 implies that $-\boldsymbol{u}$ has the same length as $\boldsymbol{u}$ but is directed in the opposite direction (assuming $\boldsymbol{u} \neq \boldsymbol{0}$). Vector subtraction can now be defined in terms of vector addition by *adding the negative* as follows.

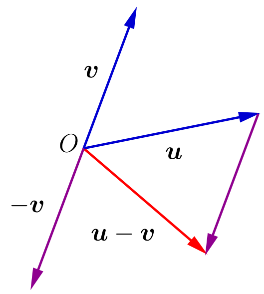
Definition: Let $\boldsymbol{u}$ and $\boldsymbol{v}$ be nonzero vectors in $\mathbb{R}^2$, $\mathbb{R}^3$ or $\mathbb{R}^n$. The **difference** of $\boldsymbol{u}$ and $\boldsymbol{v}$ is the vector $\boldsymbol{u} - \boldsymbol{v}$ given by

$$\boldsymbol{u} - \boldsymbol{v} = \boldsymbol{u} + (-\boldsymbol{v}) .$$

In terms of components it is given by

- In $\mathbb{R}^2$: $\boldsymbol{u} - \boldsymbol{v} = (u_1 - v_1, u_2 - v_2)$.
- In $\mathbb{R}^3$: $\boldsymbol{u} - \boldsymbol{v} = (u_1 - v_1, u_2 - v_2, u_3 - v_3)$.
- In $\mathbb{R}^n$: $\boldsymbol{u} - \boldsymbol{v} = (u_1 - v_1, u_2 - v_2, \ldots, u_n - v_n)$.

Once again, the definition is consistent with the interpretation of vectors as column matrices and matrix subtraction. Geometrically we add the negative vector $-\boldsymbol{v}$ to $\boldsymbol{u}$ tip-to-tail to get the difference $\boldsymbol{u} - \boldsymbol{v}$.



Example 4-7

Find the difference $\boldsymbol{u}-\boldsymbol{v}$ of the given vectors.

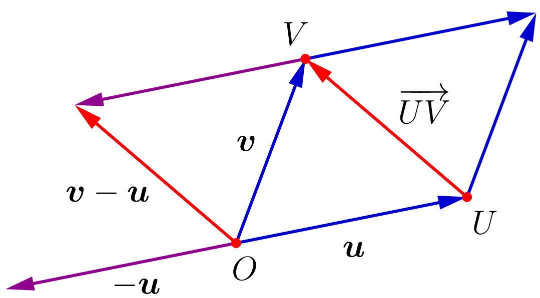
1. $\boldsymbol{u}=(1,-1,2)$, $\boldsymbol{v}=(-5,4,3)$

   Solution:
   $\boldsymbol{u}-\boldsymbol{v}=(1-(-5),-1-4,2-3)=(6,-5,-1)$

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

   Solution:
   $$\boldsymbol{u}-\boldsymbol{v}=\begin{bmatrix}3-(-1)\\-1-2\\3-(-2)\\2-3\end{bmatrix}=\begin{bmatrix}4\\-3\\5\\-1\end{bmatrix}$$

The difference of vectors has a convenient geometric interpretation in terms of the parallelogram induced by $\boldsymbol{u}$ and $\boldsymbol{v}$ which we introduced with vector addition. If we draw the other diagonal as the directed line segment from the point U at the tip of $\boldsymbol{u}$ to the point V at the tip of $\boldsymbol{v}$ and we draw the vector $\boldsymbol{v}-\boldsymbol{u}$ we see that the two will be equivalent.



In other words, the vector $\boldsymbol{v}-\boldsymbol{u}$ is essentially just the other diagonal between the tips of the two vectors. The correct direction is easily remembered since we must have $\boldsymbol{u}+(\boldsymbol{v}-\boldsymbol{u})=\boldsymbol{v}$ when added tip-to-tail. We summarize the discussion with the following theorem.

Theorem 4-4: Let U and V be distinct points in $\mathbb{R}^n$ with associated vectors $\boldsymbol{u}$ and $\boldsymbol{v}$. Then the vector $\boldsymbol{v}-\boldsymbol{u}$ is equivalent to the directed line segment $\overrightarrow{UV}$ from U to V. In other words, $\overrightarrow{UV}$ is parallel to the vector $\boldsymbol{v}-\boldsymbol{u}$ and the length of the segment (the distance between U and V) equals the length of that difference, $\|\boldsymbol{v}-\boldsymbol{u}\|$.

Example 4-8

Find the vector that is equivalent to the directed line segment from $P(1,5,7)$ to $Q(2,1,0)$ and use it to find the distance between P and Q.

Solution:
The vectors associated with the points are $\boldsymbol{p} = (1,5,7)$ and $\boldsymbol{q} = (2,1,0)$. The directed line segment $\overrightarrow{PQ}$ is then equivalent to the difference

$$\begin{aligned}\boldsymbol{q} - \boldsymbol{p} &= (2,1,0) - (1,5,7) \\ &= (2-1, 1-5, 0-7) \\ &= (1,-4,-7)\,.\end{aligned}$$

The distance between P and Q equals

$$\|\boldsymbol{q} - \boldsymbol{p}\| = \sqrt{1^2 + (-4)^2 + (-7)^2} = \sqrt{1+16+49} = \sqrt{66}\,.$$

The previous discussion suggests the following definition for the distance between two vectors in $\mathbb{R}^n$ as the distance between their terminal points (tips).

Definition: The **distance** $d(\boldsymbol{u},\boldsymbol{v})$ between two vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ is defined to be

$$d(\boldsymbol{u},\boldsymbol{v}) = \|\boldsymbol{v} - \boldsymbol{u}\|\ .$$

Specifically one has:

- In $\mathbb{R}^2$: $d(\boldsymbol{u},\boldsymbol{v}) = \sqrt{(v_1-u_1)^2 + (v_2-u_2)^2}\,.$
- In $\mathbb{R}^3$: $d(\boldsymbol{u},\boldsymbol{v}) = \sqrt{(v_1-u_1)^2 + (v_2-u_2)^2 + (v_3-u_3)^2}\,.$
- In $\mathbb{R}^n$: $d(\boldsymbol{u},\boldsymbol{v}) = \sqrt{(v_1-u_1)^2 + (v_2-u_2)^2 + \ldots + (v_n-u_n)^2}\,.$

Example 4-9

Find the distance between the given vectors.

1. $\boldsymbol{u} = (-1,1)$, $\boldsymbol{v} = (2,5)$

   Solution:

$$\begin{aligned}d(\boldsymbol{u},\boldsymbol{v}) &= \|\boldsymbol{v} - \boldsymbol{u}\| \\ &= \sqrt{(2-(-1))^2 + (5-1)^2} \\ &= \sqrt{9+16} = \sqrt{25} = 5\end{aligned}$$

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

   Solution:

$$\begin{aligned}d(\boldsymbol{u},\boldsymbol{v}) &= \|\boldsymbol{v} - \boldsymbol{u}\| \\ &= \sqrt{(-1-2)^2 + (4-(-1))^2 + (2-(-3))^2} \\ &= \sqrt{9+25+25} = \sqrt{59}\end{aligned}$$

4.1.6 Properties of Vector Operations

The following properties of vector addition and scalar multiplication are analogous to the more general matrix properties of Theorem 2-4 and indeed follow from that with the identification of vectors with column matrices.

Theorem 4-5: Let $\boldsymbol{u}$, $\boldsymbol{v}$, and $\boldsymbol{w}$ be vectors in $\mathbb{R}^n$ and let a and b be scalars. The following are true:

(1) $\boldsymbol{u}+\boldsymbol{v}=\boldsymbol{v}+\boldsymbol{u}$ (commutative law for addition)
(2) $(\boldsymbol{u}+\boldsymbol{v})+\boldsymbol{w}=\boldsymbol{u}+(\boldsymbol{v}+\boldsymbol{w})$ (associative law for addition)
(3) $\boldsymbol{u}+\mathbf{0}=\boldsymbol{u}$
(4) $\boldsymbol{u}+(-\boldsymbol{u})=\mathbf{0}$
(5) $(ab)\boldsymbol{u}=a(b\boldsymbol{u})=b(a\boldsymbol{u})$
(6) $a(\boldsymbol{u}+\boldsymbol{v})=a\boldsymbol{u}+a\boldsymbol{v}$ (scalar distributive law)
(7) $(a+b)\boldsymbol{u}=a\boldsymbol{u}+b\boldsymbol{u}$ (scalar distributive law)
(8) $1\boldsymbol{u}=\boldsymbol{u}$
(9) $0\boldsymbol{u}=\mathbf{0}$

Proof:
Selected proofs for the two-dimensional case ($n=2$) are as follows. The more general case is analogous. Let $\boldsymbol{u}=(u_1,u_2)$, $\boldsymbol{v}=(v_1,v_2)$, $\boldsymbol{w}=(w_1,w_2)$ be vectors and a and b scalars.

(1) commutative law:

$$\begin{aligned}\boldsymbol{u}+\boldsymbol{v}&=(u_1+v_1,u_2+v_2)\\&=(v_1+u_1,v_2+u_2)\\&=(v_1,v_2)+(u_1,u_2)\\&=\boldsymbol{v}+\boldsymbol{u}\end{aligned}$$

(2) associative law:

$$\begin{aligned}(\boldsymbol{u}+\boldsymbol{v})+\boldsymbol{w}&=(u_1+v_1,u_2+v_2)+(w_1,w_2)\\&=(u_1+v_1+w_1,u_2+v_2+w_2)\\&=(u_1+(v_1+w_1),u_2+(v_2+w_2))\\&=(u_1,u_2)+(v_1+w_1,v_2+w_2)\\&=\boldsymbol{u}+(\boldsymbol{v}+\boldsymbol{w})\end{aligned}$$

(7) scalar distributive law:

$$\begin{aligned}(a+b)\boldsymbol{u}&=(a+b)(u_1,u_2)\\&=((a+b)u_1,(a+b)u_2)\\&=(au_1+bu_1,au_2+bu_2)\\&=(au_1,au_2)+(bu_1,bu_2)\\&=a(u_1,u_2)+b(u_1,u_2)\\&=a\boldsymbol{u}+b\boldsymbol{u}\end{aligned}$$

4.2 Dot Product

Definition: Let $\boldsymbol{u}$ and $\boldsymbol{v}$ be two vectors in $\mathbb{R}^2$, $\mathbb{R}^3$ or $\mathbb{R}^n$. Then the **dot product** (or the **inner product** or **scalar product**) of $\boldsymbol{u}$ with $\boldsymbol{v}$ is a scalar denoted $\boldsymbol{u} \cdot \boldsymbol{v}$ given by:

- In $\mathbb{R}^2$: $\boldsymbol{u} \cdot \boldsymbol{v} = u_1 v_1 + u_2 v_2$.
- In $\mathbb{R}^3$: $\boldsymbol{u} \cdot \boldsymbol{v} = u_1 v_1 + u_2 v_2 + u_3 v_3$.
- In $\mathbb{R}^n$: $\boldsymbol{u} \cdot \boldsymbol{v} = u_1 v_1 + u_2 v_2 + \cdots + u_n v_n$.

Writing the vectors as column matrices the dot product effectively equals

$$\boldsymbol{u} \cdot \boldsymbol{v} = \boldsymbol{u}^T \boldsymbol{v} ,$$

provided we interpret the latter as the entry of the resulting 1×1 matrix.

Example 4-10

Compute the dot product of the given vectors.

1. $\boldsymbol{u} = (1, -2, -1)$, $\boldsymbol{v} = (3, -1, 1)$

   Solution:
   $\boldsymbol{u} \cdot \boldsymbol{v} = 1(3) + (-2)(-1) + (-1)(1) = 3 + 2 - 1 = 4$

2. $\boldsymbol{u} = \begin{bmatrix} -1 \\ 1 \\ 2 \\ 3 \end{bmatrix}$, $\boldsymbol{v} = \begin{bmatrix} 5 \\ 2 \\ 3 \\ -2 \end{bmatrix}$

   Solution:

   $$\boldsymbol{u} \cdot \boldsymbol{v} = \boldsymbol{u}^T \boldsymbol{v} = \begin{bmatrix} -1 & 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 5 \\ 2 \\ 3 \\ -2 \end{bmatrix} = [(-1)(5) + 1(2) + 2(3) + 3(-2)] = [(-5) + 2 + 6 - 6] = [-3] ,$$

   which we interpret as the scalar -3 .

4.2.1 Properties of the Dot Product

The following properties follow from the definition of the dot product.

Theorem 4-6: Let $\boldsymbol{u}$, $\boldsymbol{v}$ and $\boldsymbol{w}$ be vectors in $\mathbb{R}^n$ and a be a scalar. The following are true:

(1) $\boldsymbol{u} \cdot \boldsymbol{v} = \boldsymbol{v} \cdot \boldsymbol{u}$ (commutative law)

(2) $a(\boldsymbol{u} \cdot \boldsymbol{v}) = (a\boldsymbol{u}) \cdot \boldsymbol{v} = \boldsymbol{u} \cdot (a\boldsymbol{v})$

(3) $\boldsymbol{u} \cdot (\boldsymbol{v} + \boldsymbol{w}) = \boldsymbol{u} \cdot \boldsymbol{v} + \boldsymbol{u} \cdot \boldsymbol{w}$ (distributive law)

(4) $\boldsymbol{u} \cdot \boldsymbol{0} = 0$

(5) $\boldsymbol{u} \cdot \boldsymbol{u} = \|\boldsymbol{u}\|^2$

Proof:
Selected proofs for the two-dimensional case ($n = 2$) are as follows. The more general case is analogous. Let $\boldsymbol{u} = (u_1, u_2)$, $\boldsymbol{v} = (v_1, v_2)$, $\boldsymbol{w} = (w_1, w_2)$ be vectors and a a scalar.

(1) commutative law:

$$\begin{aligned} \boldsymbol{u} \cdot \boldsymbol{v} &= u_1 v_1 + u_2 v_2 \\ &= v_1 u_1 + v_2 u_2 \\ &= \boldsymbol{v} \cdot \boldsymbol{u} \end{aligned}$$

(3) distributive law:

$$\begin{aligned} \boldsymbol{u} \cdot (\boldsymbol{v} + \boldsymbol{w}) &= (u_1, u_2) \cdot (v_1 + w_1, v_2 + w_2) \\ &= u_1(v_1 + w_1) + u_2(v_2 + w_2) \\ &= u_1 v_1 + u_1 w_1 + u_2 v_2 + u_2 w_2 \\ &= (u_1 v_1 + u_2 v_2) + (u_1 w_1 + u_2 w_2) \\ &= \boldsymbol{u} \cdot \boldsymbol{v} + \boldsymbol{u} \cdot \boldsymbol{w} \end{aligned}$$

(5) $\boldsymbol{u} \cdot \boldsymbol{u} = \|\boldsymbol{u}\|^2$**:**

$$\begin{aligned} \boldsymbol{u} \cdot \boldsymbol{u} &= (u_1, u_2) \cdot (u_1, u_2) \\ &= u_1 u_1 + u_2 u_2 \\ &= u_1^2 + u_2^2 \\ &= \left(\sqrt{u_1^2 + u_2^2}\right)^2 \\ &= \|\boldsymbol{u}\|^2 \end{aligned}$$

The relationship between vector length and the dot product can be exploited to obtain further results.

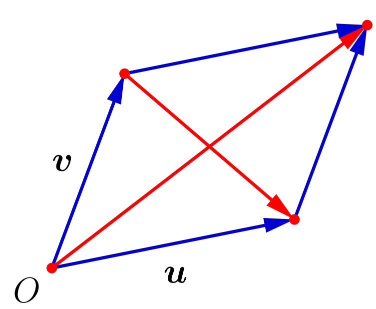
Theorem 4-7: The lengths of the sum and difference of two vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ satisfy

$$\|\boldsymbol{u} + \boldsymbol{v}\|^2 + \|\boldsymbol{u} - \boldsymbol{v}\|^2 = 2\|\boldsymbol{u}\|^2 + 2\|\boldsymbol{v}\|^2 \ .$$

Proof:

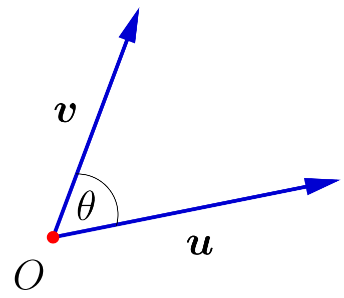
$$\begin{aligned} \|\boldsymbol{u} + \boldsymbol{v}\|^2 + \|\boldsymbol{u} - \boldsymbol{v}\|^2 &= (\boldsymbol{u} + \boldsymbol{v}) \cdot (\boldsymbol{u} + \boldsymbol{v}) + (\boldsymbol{u} - \boldsymbol{v}) \cdot (\boldsymbol{u} - \boldsymbol{v}) \\ &= (\boldsymbol{u} \cdot \boldsymbol{u} + \cancel{\boldsymbol{u}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

In terms of the relationship of $\boldsymbol{u} + \boldsymbol{v}$ and $\boldsymbol{u} - \boldsymbol{v}$ to the diagonals of the parallelogram induced by vectors $\boldsymbol{u}$ and $\boldsymbol{v}$, the previous theorem has a geometrical interpretation in two and three dimensions. It proves that the sum of the squares of the diagonal lengths of a parallelogram equals the sum of the squares of its side lengths, since two sides have length $\|\boldsymbol{u}\|$ and two have length $\|\boldsymbol{v}\|$.



4.2.2 Angle between Vectors

Two nonzero vectors in two or three dimensions will lie in a plane with an angle θ between them formed at the origin O between 0 and 180 degrees (π radians).



The dot product allows us to calculate this angle and conversely the angle can be used to evaluate the dot product.

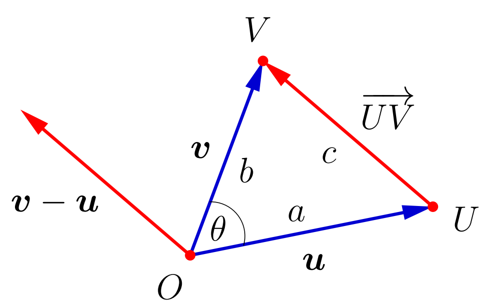
Theorem 4-8: Let $\boldsymbol{u}$ and $\boldsymbol{v}$ be vectors in $\mathbb{R}^2$ or $\mathbb{R}^3$ and let θ be the angle between $\boldsymbol{u}$ and $\boldsymbol{v}$ with $0 \le \theta \le \pi$. Then:

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

If $\boldsymbol{u} \neq \boldsymbol{0}$ and $\boldsymbol{v} \neq \boldsymbol{0}$ then

$$\cos\theta = \frac{\boldsymbol{u} \cdot \boldsymbol{v}}{\|\boldsymbol{u}\| \, \|\boldsymbol{v}\|} \, .$$

Proof:
The result $\boldsymbol{u} \cdot \boldsymbol{v} = \|\boldsymbol{u}\| \, \|\boldsymbol{v}\| \cos\theta$ holds trivially if either $\boldsymbol{u}$ or $\boldsymbol{v}$ (or both) are zero vectors, so assume neither are zero vectors. Let OUV be the triangle determined by the vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ with U and V being the terminal points of their respective vectors.



Let $a = \|\boldsymbol{u}\|$ and $b = \|\boldsymbol{v}\|$. If c is the length of directed line segment $\overrightarrow{UV}$ then by Theorem 4-4 $c = \|\boldsymbol{v} - \boldsymbol{u}\|$. Applying the law of cosines one has

$$c^2 = a^2 + b^2 - 2ab\cos\theta \, .$$

This implies

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

Expanding the square lengths using the dot product and simplifying gives:

$$\begin{aligned}
(\boldsymbol{v}-\boldsymbol{u})\cdot(\boldsymbol{v}-\boldsymbol{u}) &= (\boldsymbol{u}\cdot\boldsymbol{u})+(\boldsymbol{v}\cdot\boldsymbol{v})-2\,\|\boldsymbol{u}\|\,\|\boldsymbol{v}\|\cos\theta \\
\cancel{\boldsymbol{v}\cdot\boldsymbol{v}}-\boldsymbol{v}\cdot\boldsymbol{u}-\boldsymbol{u}\cdot\boldsymbol{v}+\cancel{\boldsymbol{u}\cdot\boldsymbol{u}} &= \cancel{\boldsymbol{u}\cdot\boldsymbol{u}}+\cancel{\boldsymbol{v}\cdot\boldsymbol{v}}-2\,\|\boldsymbol{u}\|\,\|\boldsymbol{v}\|\cos\theta \\
-\boldsymbol{u}\cdot\boldsymbol{v}-\boldsymbol{u}\cdot\boldsymbol{v} &= -2\,\|\boldsymbol{u}\|\,\|\boldsymbol{v}\|\cos\theta \\
-2\boldsymbol{u}\cdot\boldsymbol{v} &= -2\,\|\boldsymbol{u}\|\,\|\boldsymbol{v}\|\cos\theta \\
\boldsymbol{u}\cdot\boldsymbol{v} &= \|\boldsymbol{u}\|\,\|\boldsymbol{v}\|\cos\theta
\end{aligned}$$

Since $\boldsymbol{u}$ and $\boldsymbol{v}$ are nonzero their lengths are nonzero and it follows that

$$\cos\theta = \frac{\boldsymbol{u}\cdot\boldsymbol{v}}{\|\boldsymbol{u}\|\,\|\boldsymbol{v}\|}\,.$$

Example 4-11

Find the angle between each pair of vectors.

1. $\boldsymbol{u}=(-1,2,1)\ ,\ \boldsymbol{v}=(2,1,1)$
   Solution:

$$\begin{aligned}
\boldsymbol{u}\cdot\boldsymbol{v} &= (-1,2,1)\cdot(2,1,1) = (-1)2+2(1)+1(1) = 1 \\
\|\boldsymbol{u}\| &= \sqrt{(-1)^2+2^2+1^2} = \sqrt{6} \\
\|\boldsymbol{v}\| &= \sqrt{2^2+1^2+1^2} = \sqrt{6}
\end{aligned}$$

   Therefore:

$$\cos\theta = \frac{\boldsymbol{u}\cdot\boldsymbol{v}}{\|\boldsymbol{u}\|\,\|\boldsymbol{v}\|} = \frac{1}{\sqrt{6}\cdot\sqrt{6}} = \frac{1}{6}$$

$$\theta = \cos^{-1}(1/6) \approx 80.4^\circ = 1.40\ (\text{radians})\,.$$

2. $\boldsymbol{u}=(2,1,-1)\ ,\ \boldsymbol{v}=(1,-1,1)$
   Solution:

$$\boldsymbol{u}\cdot\boldsymbol{v} = (2,1,-1)\cdot(1,-1,1) = 2(1)+1(-1)+(-1)(1) = 0$$

   Therefore:

$$\cos\theta = \frac{\boldsymbol{u}\cdot\boldsymbol{v}}{\|\boldsymbol{u}\|\,\|\boldsymbol{v}\|} = \frac{0}{\|\boldsymbol{u}\|\,\|\boldsymbol{v}\|} = 0$$

   Since $\cos\theta = 0$, we have $\theta = \cos^{-1}(0) = \dfrac{\pi}{2}$.

Definition: Two vectors, $\boldsymbol{u}$ and $\boldsymbol{v}$ in $\mathbb{R}^n$ are **orthogonal** if $\boldsymbol{u}\cdot\boldsymbol{v} = 0$.

Orthogonality can therefore happen if the angle θ between the vectors is $\pi/2$ (as in the last example) or either $\boldsymbol{u}$ or $\boldsymbol{v}$ is a zero vector.

Example 4-12

The elementary vectors $\boldsymbol{e}_i$ are all mutually orthogonal. For example, in $\mathbb{R}^3$

$$\boldsymbol{e}_1\cdot\boldsymbol{e}_2 = \boldsymbol{i}\cdot\boldsymbol{j} = (1,0,0)\cdot(0,1,0) = (1)(0)+(0)(1)+(0)(0) = 0\,.$$

The sign of the dot product provides useful information. Since the lengths $\|u\|$ and $\|v\|$ are both positive for nonzero vectors, the sign of $\cos\theta$ will be determined by that of the numerator $\boldsymbol{u}\cdot\boldsymbol{v}$ in our formula and we have the following result.

Theorem 4-9: Let $\boldsymbol{u}$ and $\boldsymbol{v}$ be nonzero vectors in $\mathbb{R}^2$ or $\mathbb{R}^3$ and let θ be the angle between them. Then θ satisfies

1. $0 \leq \theta < \pi/2$ if $\boldsymbol{u} \cdot \boldsymbol{v} > 0$ (θ is 0 or acute).
2. $\theta = \pi/2$ if $\boldsymbol{u} \cdot \boldsymbol{v} = 0$ (θ is a right angle).
3. $\pi/2 < \theta \leq \pi$ if $\boldsymbol{u} \cdot \boldsymbol{v} < 0$ (θ is obtuse or π).

To distinguish the two exceptional cases of $\theta = 0$ (parallel vectors) or $\theta = \pi$ (antiparallel vectors) one needs to evaluate the formula for $\cos\theta$ to see if it actually is $+1$ or -1 respectively.

Example 4-13

Use the dot product to determine the range of the angle formed by the given vectors.

1. $\boldsymbol{u} = (3, 2, -2)$, $\boldsymbol{v} = (2, 1, 2)$

   Solution:
$$\boldsymbol{u} \cdot \boldsymbol{v} = (3)(2) + (2)(1) + (-2)(2) = 4 > 0 \implies 0 \leq \theta < \frac{\pi}{2}.$$

2. $\boldsymbol{u} = (2, 1, -1)$, $\boldsymbol{v} = (1, -1, 1)$

   Solution:
$$\boldsymbol{u} \cdot \boldsymbol{v} = (2)(1) + (1)(-1) + (-1)(1) = 0 \implies \theta = \frac{\pi}{2} \text{ (right angle).}$$

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

   Solution:
$$\boldsymbol{u} \cdot \boldsymbol{v} = \boldsymbol{u}^T \boldsymbol{v} = \begin{bmatrix} 4 & -3 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = (4)(1) + (-3)(2) = -2 < 0 \implies \frac{\pi}{2} < \theta \leq \pi.$$

Angles in $\mathbb{R}^n$

In higher dimensions ($n > 3$) we cannot resort to geometry to evaluate angles. However we can *define* the angle θ between two nonzero vectors in $\mathbb{R}^n$ to be that value $0 \leq \theta \leq \pi$ satisfying
$$\cos\theta = \frac{\boldsymbol{u} \cdot \boldsymbol{v}}{\|\boldsymbol{u}\| \, \|\boldsymbol{v}\|} \, .$$
Since cosine ranges between -1 and 1 that this angle is well-defined is not obvious. That the definition works follows from the **Cauchy-Schwarz inequality**:

Theorem 4-10: If $\boldsymbol{u}$ and $\boldsymbol{v}$ are vectors in $\mathbb{R}^n$ then
$$|\boldsymbol{u} \cdot \boldsymbol{v}| \leq \|\boldsymbol{u}\| \, \|\boldsymbol{v}\| \, .$$

For nonzero vectors we can divide both sides by the positive quantity $\|\boldsymbol{u}\| \, \|\boldsymbol{v}\|$ to get
$$\left| \frac{\boldsymbol{u} \cdot \boldsymbol{v}}{\|\boldsymbol{u}\| \, \|\boldsymbol{v}\|} \right| \leq 1 \, ,$$
and our angle θ is well-defined. Furthermore, one can show that when $\theta = 0$ the vectors are parallel ($\boldsymbol{u} = a\boldsymbol{v}$ for some positive scalar a) and when $\theta = \pi$ the vectors are antiparallel ($\boldsymbol{u} = a\boldsymbol{v}$ for some negative scalar a) as one expects.

4.2.3 Projection Theorem

We have seen that an arbitrary vector in $\mathbb{R}^n$ can be written

$$\boldsymbol{u} = (u_1, u_2, \ldots, u_n) = u_1\boldsymbol{e}_1 + u_2\boldsymbol{e}_2 + u_n\boldsymbol{e}_n\,.$$

Here u_1 is the first component and $u_1\boldsymbol{e_1}$ is the vector component directed along the first axis. If we take the dot product of both sides with $\boldsymbol{e}_1$ we get

$$\begin{aligned}
\boldsymbol{u}\cdot\boldsymbol{e}_1 &= (u_1\boldsymbol{e}_1 + u_2\boldsymbol{e}_2 + \cdots + u_n\boldsymbol{e}_n)\cdot\boldsymbol{e}_1\\
&= (u_1\boldsymbol{e}_1)\cdot\boldsymbol{e}_1 + (u_2\boldsymbol{e}_2)\cdot\boldsymbol{e}_1 + \cdots + (u_n\boldsymbol{e}_n)\cdot\boldsymbol{e}_1\\
&= u_1(\boldsymbol{e}_1\cdot\boldsymbol{e}_1) + u_2(\boldsymbol{e}_1\cdot\boldsymbol{e}_2) + \cdots + u_n(\boldsymbol{e}_1\cdot\boldsymbol{e}_n)\\
&= u_1\left\|\boldsymbol{e}_1\right\|^2 + u_2(0) + \cdots + u_n(0)\\
&= u_1(1) + 0 + \cdots + 0\\
&= u_1
\end{aligned}$$

and in general $\boldsymbol{u}\cdot\boldsymbol{e}_i = u_i$. In other words the dot product can be used to find the i^{th} component of the vector and it follows that the vector component along that direction is just

$$u_i\boldsymbol{e}_i = (\boldsymbol{u}\cdot\boldsymbol{e}_i)\boldsymbol{e}_i\,.$$

Moreover our above calculation (in the $i = 1$ case) shows that

$$\boldsymbol{u} = \boldsymbol{u_1} + \boldsymbol{u_2}$$

where

$$\boldsymbol{u}_1 = u_1\boldsymbol{e}_1 = (\boldsymbol{u}\cdot\boldsymbol{e}_1)\boldsymbol{e}_1$$

is directed along the direction $\boldsymbol{e_1}$ and

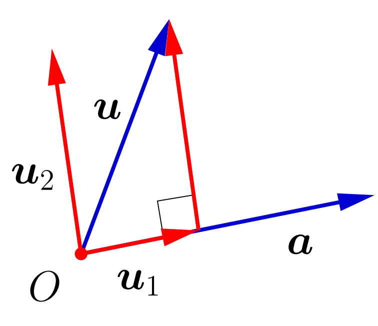
$$\boldsymbol{u}_2 = u_2\boldsymbol{e}_2 + \cdots + u_n\boldsymbol{e}_n = \boldsymbol{u} - u_1\boldsymbol{e_1} = \boldsymbol{u} - (\boldsymbol{u}\cdot\boldsymbol{e}_1)\boldsymbol{e}_1$$

is orthogonal to it, $\boldsymbol{u}_2\cdot\boldsymbol{e}_1 = 0\,.$

We can generalize this decomposition to arbitrary directions, not simply coordinate axis directions. We often want to decompose one vector into a sum of two other vectors such that

$$\boldsymbol{u} = \boldsymbol{u}_1 + \boldsymbol{u}_2$$

where $\boldsymbol{u}_1$ is in the same direction as $\boldsymbol{a}$ (ie. $\boldsymbol{u}_1$ is a scalar multiple of $\boldsymbol{a}$) and $\boldsymbol{u}_2$ is orthogonal to $\boldsymbol{a}$.



If we define $\boldsymbol{e}$ to be the unit vector along the direction $\boldsymbol{a}$, so $\boldsymbol{e} = \frac{1}{\|a\|}\boldsymbol{a}$, then our previous discussion suggests that the vector component of $\boldsymbol{u}$ along the direction of $\boldsymbol{a}$ should be

$$\boldsymbol{u}_1 = (\boldsymbol{u}\cdot\boldsymbol{e})\boldsymbol{e} = \left(\boldsymbol{u}\cdot\frac{1}{\|a\|}\boldsymbol{a}\right)\frac{1}{\|a\|}\boldsymbol{a} = \frac{\boldsymbol{u}\cdot\boldsymbol{a}}{\|\boldsymbol{a}\|^2}\boldsymbol{a}\,.$$

This is correct and the result is summarized in the following projection theorem.

Theorem 4-11: Let $\boldsymbol{u}$ and $\boldsymbol{a}$ be vectors in $\mathbb{R}^n$ with $\boldsymbol{a} \neq \mathbf{0}$. Then $\boldsymbol{u}$ has a unique decomposition into a vector projection along the direction of $\boldsymbol{a}$ and one perpendicular to it:

$$\boldsymbol{u} = \boldsymbol{u}_1 + \boldsymbol{u}_2 \,,$$

where the vector $\boldsymbol{u}_1$, denoted by $\text{proj}_{\boldsymbol{a}}\boldsymbol{u}$, is given by

$$\boldsymbol{u}_1 = \text{proj}_{\boldsymbol{a}}\boldsymbol{u} = \frac{\boldsymbol{u} \cdot \boldsymbol{a}}{\|\boldsymbol{a}\|^2}\boldsymbol{a}$$

and is called the **vector component of $\boldsymbol{u}$ along $\boldsymbol{a}$** or the **orthogonal projection of $\boldsymbol{u}$ on $\boldsymbol{a}$** and

$$\boldsymbol{u}_2 = \boldsymbol{u} - \frac{\boldsymbol{u} \cdot \boldsymbol{a}}{\|\boldsymbol{a}\|^2}\boldsymbol{a}$$

is the **component of $\boldsymbol{u}$ orthogonal to $\boldsymbol{a}$**.

Proof: Let $\boldsymbol{u}$, $\boldsymbol{a} \neq \mathbf{0}$ be vectors in $\mathbb{R}^n$ and suppose $\boldsymbol{u} = \boldsymbol{u}_1 + \boldsymbol{u}_2$ where $\boldsymbol{u}_1 = c\boldsymbol{a}$ is directed along $\boldsymbol{a}$ and $\boldsymbol{u}_2$ is orthogonal to it. Then

$$\begin{aligned}
\boldsymbol{u} \cdot \boldsymbol{a} &= (\boldsymbol{u}_1 + \boldsymbol{u}_2) \cdot \boldsymbol{a} \\
&= \boldsymbol{u}_1 \cdot \boldsymbol{a} + \boldsymbol{u}_2 \cdot \boldsymbol{a} \\
&= c\boldsymbol{a} \cdot \boldsymbol{a} + 0 \\
&= c\,\|\boldsymbol{a}\|^2
\end{aligned}$$

Since $\|\boldsymbol{a}\| \neq 0$, $c = \dfrac{\boldsymbol{u} \cdot \boldsymbol{a}}{\|\boldsymbol{a}\|^2}$ and $\boldsymbol{u}_1 = \text{proj}_{\boldsymbol{a}}\boldsymbol{u}$. Then $\boldsymbol{u_2} = \boldsymbol{u} - \boldsymbol{u}_1 = \boldsymbol{u} - \text{proj}_{\boldsymbol{a}}\boldsymbol{u}$. Furthermore $\boldsymbol{u}_2$ is orthogonal to $\boldsymbol{a}$ since

$$\begin{aligned}
\boldsymbol{u}_2 \cdot \boldsymbol{a} &= (\boldsymbol{u} - \text{proj}_{\boldsymbol{a}}\boldsymbol{u}) \cdot \boldsymbol{a} \\
&= \left(\boldsymbol{u} - \frac{\boldsymbol{u} \cdot \boldsymbol{a}}{\|\boldsymbol{a}\|^2}\boldsymbol{a}\right) \cdot \boldsymbol{a} \\
&= \boldsymbol{u} \cdot \boldsymbol{a} - \frac{\boldsymbol{u} \cdot \boldsymbol{a}}{\|\boldsymbol{a}\|^2}(\boldsymbol{a} \cdot \boldsymbol{a}) \\
&= \boldsymbol{u} \cdot \boldsymbol{a} - \frac{\boldsymbol{u} \cdot \boldsymbol{a}}{\|\boldsymbol{a}\|^2}\,\|\boldsymbol{a}\|^2 \\
&= \boldsymbol{u} \cdot \boldsymbol{a} - \boldsymbol{u} \cdot \boldsymbol{a} \\
&= 0
\end{aligned}$$

Thus any such decomposition is unique. Since clearly $\boldsymbol{u}_1 = \text{proj}_{\boldsymbol{a}}\boldsymbol{u}$ and $\boldsymbol{u}_2 = \boldsymbol{u} - \text{proj}_{\boldsymbol{a}}\boldsymbol{u}$ exist as vectors the decomposition exists.

Example 4-14

Let $\boldsymbol{u} = (2, -1, 3)$, $\boldsymbol{a} = (4, -1, 2)$. Find the vector component ($\boldsymbol{u_1}$) of $\boldsymbol{u}$ along $\boldsymbol{a}$ and the vector component ($\boldsymbol{u_2}$) of $\boldsymbol{u}$ orthogonal to $\boldsymbol{a}$. Solution:

$$\begin{aligned}
&\boldsymbol{u} \cdot \boldsymbol{a} = (2)(4) + (-1)(1) + (3)(2) = 15 \\
&\|\boldsymbol{a}\|^2 = 4^2 + (-1)^2 + 2^2 = 21
\end{aligned}$$

Therefore:

$$\begin{aligned}
\boldsymbol{u_1} &= \operatorname{proj}_{\boldsymbol{a}}\boldsymbol{u} = \frac{\boldsymbol{u}\cdot\boldsymbol{a}}{\|\boldsymbol{a}\|^2}\boldsymbol{a} = \frac{15}{21}(4,-1,2) = \frac{5}{7}(4,-1,2) \\
\boldsymbol{u_2} &= \boldsymbol{u} - \operatorname{proj}_{\boldsymbol{a}}\boldsymbol{u} = (2,-1,3) - \frac{5}{7}(4,-1,2) \\
&= \left(2-\frac{20}{7}, -1+\frac{5}{7}, 3-\frac{10}{7}\right) \\
&= \left(-\frac{6}{7}, -\frac{2}{7}, \frac{11}{7}\right)
\end{aligned}$$

Example 4-15

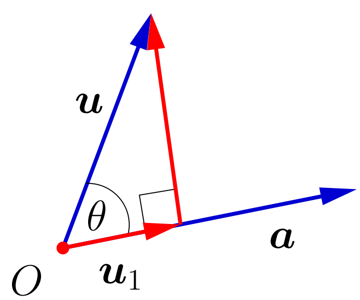
Find the projection of $\boldsymbol{u} = (2,0,1)$ on $\boldsymbol{a} = (1,2,3)$.
Solution:

$$\begin{aligned}
\boldsymbol{u}\cdot\boldsymbol{a} &= 2(1) + (0)(2) + 1(3) = 5 \\
\|\boldsymbol{a}\|^2 &= 1^2 + 2^2 + 3^2 = 14
\end{aligned}$$

Therefore

$$\operatorname{proj}_{\boldsymbol{a}}\boldsymbol{u} = \frac{\boldsymbol{u}\cdot\boldsymbol{a}}{\|\boldsymbol{a}\|^2}\boldsymbol{a} = \frac{5}{14}(1,2,3).$$

The length of the projection of $\boldsymbol{u}$ along $\boldsymbol{a}$ can be written in terms of the angle between them.



Theorem 4-12: If $\boldsymbol{u}$ and $\boldsymbol{a} \neq \boldsymbol{0}$ are vectors in $\mathbb{R}^n$ then the length of the projection of $\boldsymbol{u}$ on $\boldsymbol{a}$ satisfies:

$$\|\operatorname{proj}_{\boldsymbol{a}}\boldsymbol{u}\| = \|\boldsymbol{u}\|\,|\cos\theta|\,,$$

where θ is the angle between $\boldsymbol{u}$ and $\boldsymbol{a}$.

Proof:
Noting that $\dfrac{\boldsymbol{u}\cdot\boldsymbol{a}}{\|\boldsymbol{a}\|^2}$ is a scalar that can be pulled out of $\|\cdot\|$, we have

$$\|\operatorname{proj}_{\boldsymbol{a}}\boldsymbol{u}\| = \left\|\frac{\boldsymbol{u}\cdot\boldsymbol{a}}{\|\boldsymbol{a}\|^2}\boldsymbol{a}\right\| = \left|\frac{\boldsymbol{u}\cdot\boldsymbol{a}}{||\boldsymbol{a}||^2}\right|\,||\boldsymbol{a}|| = \frac{|\boldsymbol{u}\cdot\boldsymbol{a}|}{\|\boldsymbol{a}\|^2}\,\|\boldsymbol{a}\| = \frac{|\boldsymbol{u}\cdot\boldsymbol{a}|}{\|\boldsymbol{a}\|} = \frac{\|\boldsymbol{u}\|\,\|\boldsymbol{a}\|\,|\cos\theta|}{\|\boldsymbol{a}\|} = \|\boldsymbol{u}\|\,|\cos\theta|\,.$$

4.3 Cross Product

It is possible in three dimensions ($\mathbb{R}^3$) to define a useful multiplication between two vectors that produces a *vector*.[3]

Definition: Let $\boldsymbol{u} = (u_1, u_2, u_3)$ and $\boldsymbol{v} = (v_1, v_2, v_3)$ be vectors in $\mathbb{R}^3$. The **cross product** of $\boldsymbol{u}$ with $\boldsymbol{v}$ is a vector in $\mathbb{R}^3$ given by:

$$\boldsymbol{u} \times \boldsymbol{v} = (u_2v_3 - u_3v_2,\ u_3v_1 - u_1v_3,\ u_1v_2 - u_2v_1)\,.$$

The cross product formula is easily remembered if we formally allow unit vectors into a 3×3 determinant

$$\boldsymbol{u} \times \boldsymbol{v} = \begin{vmatrix} \boldsymbol{i} & \boldsymbol{j} & \boldsymbol{k} \\ u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{vmatrix} = \begin{vmatrix} \boldsymbol{i} & u_1 & v_1 \\ \boldsymbol{j} & u_2 & v_2 \\ \boldsymbol{k} & u_3 & v_3 \end{vmatrix}\,,$$

and cofactor expand along the first row or column respectively. For example, expanding along the first row of the first determinant,

$$\begin{aligned} \boldsymbol{u} \times \boldsymbol{v} = \begin{vmatrix} \boldsymbol{i} & \boldsymbol{j} & \boldsymbol{k} \\ u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{vmatrix} &= \boldsymbol{i}\begin{vmatrix} u_2 & u_3 \\ v_2 & v_3 \end{vmatrix} - \boldsymbol{j}\begin{vmatrix} u_1 & u_3 \\ v_1 & v_3 \end{vmatrix} + \boldsymbol{k}\begin{vmatrix} u_1 & u_2 \\ v_1 & v_2 \end{vmatrix} \\ &= \boldsymbol{i}(u_2v_3 - u_3v_2) - \boldsymbol{j}(u_1v_3 - u_3v_1) + \boldsymbol{k}(u_1v_2 - u_2v_1) \\ &= (u_2v_3 - u_3v_2,\ u_3v_1 - u_1v_3,\ u_1v_2 - u_2v_1)\,. \end{aligned}$$

Example 4-16

Find the cross product of the given vectors.

1. $\boldsymbol{u} = (-1, 2, 1)$, $\boldsymbol{v} = (1, 1, 0)$

   Solution:

$$\begin{aligned} \boldsymbol{u} \times \boldsymbol{v} &= \begin{vmatrix} \boldsymbol{i} & \boldsymbol{j} & \boldsymbol{k} \\ -1 & 2 & 1 \\ 1 & 1 & 0 \end{vmatrix} \qquad \leftarrow \text{Cofactor expansion along the first row.} \\ &= \boldsymbol{i}\begin{vmatrix} 2 & 1 \\ 1 & 0 \end{vmatrix} - \boldsymbol{j}\begin{vmatrix} -1 & 1 \\ 1 & 0 \end{vmatrix} + \boldsymbol{k}\begin{vmatrix} -1 & 2 \\ 1 & 1 \end{vmatrix} \\ &= \boldsymbol{i}(0-1) - \boldsymbol{j}(0-1) + \boldsymbol{k}(-1-2) \\ &= -\boldsymbol{i} + \boldsymbol{j} - 3\boldsymbol{k} \\ &= (-1, 1, -3) \end{aligned}$$

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

[3] Unlike the dot product which is defined for all $\mathbb{R}^n$ a cross product with the properties to be outlined later, cannot be defined in most dimensions. It is possible to define a cross product in $\mathbb{R}^7$. Later we will introduce complex numbers. These can be generalized to quaternions and octonions with four and eight real components respectively. The vector product in $\mathbb{R}^7$ can be related to the vector part of octonion multiplication just as the $\mathbb{R}^3$ cross product is related to the vector part of quaternion multiplication.

Solution:

$$\begin{aligned}
\boldsymbol{u}\times\boldsymbol{v} &= \begin{vmatrix} \boldsymbol{i} & 3 & -1 \\ \boldsymbol{j} & -4 & 1 \\ \boldsymbol{k} & 1 & 1 \end{vmatrix} && \leftarrow \text{Cofactor expansion along the first column.} \\
&= \boldsymbol{i}\begin{vmatrix} -4 & 1 \\ 1 & 1 \end{vmatrix} - \boldsymbol{j}\begin{vmatrix} 3 & -1 \\ 1 & 1 \end{vmatrix} + \boldsymbol{k}\begin{vmatrix} 3 & -1 \\ -4 & 1 \end{vmatrix} \\
&= \boldsymbol{i}(-4-1) - \boldsymbol{j}(3+1) + \boldsymbol{k}(3-4) \\
&= -5\boldsymbol{i} - 4\boldsymbol{j} - \boldsymbol{k} \\
&= (-5,-4,-1)
\end{aligned}$$

Example 4-17

Given $\boldsymbol{u} = (2,-1,1)$ and $\boldsymbol{v} = (3,2,1)$, evaluate the following, if possible.

1. $\boldsymbol{u}\times(\boldsymbol{u}\cdot\boldsymbol{v})$

   Solution:
   The dot product $\boldsymbol{u}\cdot\boldsymbol{v}$ is a scalar but the cross product acts on two vectors so $\boldsymbol{u}\times(\boldsymbol{u}\cdot\boldsymbol{v})$ is not defined.

2. $\boldsymbol{u}\cdot(\boldsymbol{u}\times\boldsymbol{v})$

   Solution:

$$\begin{aligned}
\boldsymbol{u}\times\boldsymbol{v} &= \begin{vmatrix} \boldsymbol{i} & \boldsymbol{j} & \boldsymbol{k} \\ 2 & -1 & 1 \\ 3 & 2 & 1 \end{vmatrix} \\
&= \boldsymbol{i}\begin{vmatrix} -1 & 1 \\ 2 & 1 \end{vmatrix} - \boldsymbol{j}\begin{vmatrix} 2 & 1 \\ 3 & 1 \end{vmatrix} + \boldsymbol{k}\begin{vmatrix} 2 & -1 \\ 3 & 2 \end{vmatrix} \\
&= -3\boldsymbol{i} + \boldsymbol{j} + 7\boldsymbol{k} \\
&= (-3,1,7)
\end{aligned}$$

   Therefore:

$$\boldsymbol{u}\cdot(\boldsymbol{u}\times\boldsymbol{v}) = (2,-1,1)\cdot(-3,1,7) = -6-1+7 = 0\,.$$

In the last example $\boldsymbol{u}\times\boldsymbol{v}$ was found to be orthogonal to $\boldsymbol{u}$. This is true in general.

Theorem 4-13: The vector $\boldsymbol{u}\times\boldsymbol{v}$ is orthogonal to both $\boldsymbol{u}$ and $\boldsymbol{v}$.

Proof:

Let $\boldsymbol{u} = (u_1,u_2,u_3)$ and $\boldsymbol{v} = (v_1,v_2,v_3)$. Then $\boldsymbol{u}\times\boldsymbol{v} = \begin{vmatrix} \boldsymbol{i} & \boldsymbol{j} & \boldsymbol{k} \\ u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{vmatrix}$ and

$$\begin{aligned}
\boldsymbol{u}\cdot(\boldsymbol{u}\times\boldsymbol{v}) &= (u_1,u_2,u_3)\cdot(u_2v_3-u_3v_2,\, u_3v_1-u_1v_3,\, u_1v_3-u_3v_1) \\
&= u_1(u_2v_3-u_3v_2) + u_2(u_3v_1-u_1v_3) + u_3(u_1v_2-u_2v_1) \\
&= u_1u_2v_3 - u_1u_3v_2 + u_2u_3v_1 - u_2u_1v_3 + u_3u_1v_2 - u_3u_2v_1 \\
&= 0
\end{aligned}$$

Similarly one can show that $\boldsymbol{v}\cdot(\boldsymbol{u}\times\boldsymbol{v}) = 0\,.$

Example 4-18

Find a vector $\boldsymbol{w}$ that is orthogonal to both vectors.

1. $\boldsymbol{u} = (0, 1, -2)$, $\boldsymbol{v} = (1, -1, 3)$

   Solution:

$$\begin{aligned} \boldsymbol{w} &= \boldsymbol{u} \times \boldsymbol{v} \\ &= \begin{vmatrix} \boldsymbol{i} & \boldsymbol{j} & \boldsymbol{k} \\ 0 & 1 & -2 \\ 1 & -1 & 3 \end{vmatrix} \\ &= \boldsymbol{i} \begin{vmatrix} 1 & -2 \\ -1 & 3 \end{vmatrix} - \boldsymbol{j} \begin{vmatrix} 0 & -2 \\ 1 & 3 \end{vmatrix} + \boldsymbol{k} \begin{vmatrix} 0 & 1 \\ 1 & -1 \end{vmatrix} \\ &= \boldsymbol{i} - 2\boldsymbol{j} - \boldsymbol{k} \\ &= (1, -2, -1) \end{aligned}$$

   $\boldsymbol{w} = (1, -2, -1)$ is orthogonal to both $\boldsymbol{u}$ and $\boldsymbol{v}$.

   Check:

$$\begin{aligned} \boldsymbol{u} \cdot \boldsymbol{w} &= 0(1) + 1(-2) - 2(-1) \\ &= 0 - 2 + 2 = 0 \\ \boldsymbol{v} \cdot \boldsymbol{w} &= 1(1) + (-1)(-2) + 3(-1) \\ &= 1 + 2 - 3 = 0 \end{aligned}$$

2. $\boldsymbol{u} = \begin{bmatrix} -2 \\ 5 \\ -1 \end{bmatrix}$, $\boldsymbol{v} = \begin{bmatrix} -3 \\ 0 \\ 1 \end{bmatrix}$

   Solution:

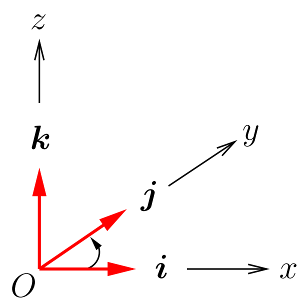
$$\boldsymbol{w} = \boldsymbol{u} \times \boldsymbol{v} = \begin{vmatrix} \boldsymbol{i} & -2 & -3 \\ \boldsymbol{j} & 5 & 0 \\ \boldsymbol{k} & -1 & 1 \end{vmatrix} = \boldsymbol{i}(5 - 0) - \boldsymbol{j}(-2 - 3) + \boldsymbol{k}(0 + 15) = 5\boldsymbol{i} + 5\boldsymbol{j} + 15\boldsymbol{k} = \begin{bmatrix} 5 \\ 5 \\ 15 \end{bmatrix}$$

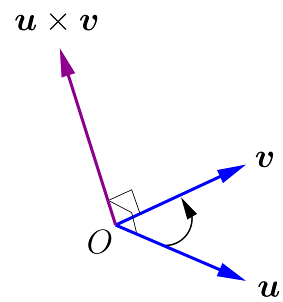
4.3.1 Right-Hand Rule

Since $\boldsymbol{u} \times \boldsymbol{v}$ is orthogonal to both $\boldsymbol{u}$ and $\boldsymbol{v}$ it is directed out of the plane determined by those two vectors and it is natural to ask in which of the two possible directions this is. One can readily verify that the unit vectors in $\mathbb{R}^3$ satisfy

$$\boldsymbol{i} \times \boldsymbol{j} = \boldsymbol{k} \qquad \boldsymbol{j} \times \boldsymbol{k} = \boldsymbol{i} \qquad \boldsymbol{k} \times \boldsymbol{i} = \boldsymbol{j}\,.$$

Assume we choose, as has been done so far in this text, a coordinate system that is **right-handed**. This means that if you straighten your right hand and point it in the x direction $\boldsymbol{i}$ and then curl your fingers in the y direction $\boldsymbol{j}$, your thumb will point in the z direction $\boldsymbol{k}$. Provided such a coordinate system is used then the direction of the cross product is similarly determined by the **right-hand rule**. Directing your straightened right hand along the direction of $\boldsymbol{u}$ and curling your fingers in the direction $\boldsymbol{v}$, the cross product $\boldsymbol{u} \times \boldsymbol{v}$ points in the direction of your thumb.





If left-handed coordinate systems are used then cross products will follow a left-hand rule, but such coordinate systems will be avoided in this text.[4]

4.3.2 Properties of the Cross Product

Theorem 4-14: Let $\boldsymbol{u}$, $\boldsymbol{v}$, and $\boldsymbol{w}$ be vectors in $\mathbb{R}^3$ and let a be a scalar. The following are true:

(1) $\boldsymbol{u} \times \boldsymbol{v} = -\boldsymbol{v} \times \boldsymbol{u}$ (anticommutative law)

(2) $\boldsymbol{u} \times (\boldsymbol{v} + \boldsymbol{w}) = \boldsymbol{u} \times \boldsymbol{v} + \boldsymbol{u} \times \boldsymbol{w}$ (left distributive law)

(3) $(\boldsymbol{u} + \boldsymbol{v}) \times \boldsymbol{w} = \boldsymbol{u} \times \boldsymbol{w} + \boldsymbol{v} \times \boldsymbol{w}$ (right distributive law)

(4) $a(\boldsymbol{u} \times \boldsymbol{v}) = (a\boldsymbol{u}) \times \boldsymbol{v} = \boldsymbol{u} \times (a\boldsymbol{v})$

(5) $\boldsymbol{u} \times \boldsymbol{0} = \boldsymbol{0} \times \boldsymbol{u} = \boldsymbol{0}$

(6) $\boldsymbol{u} \times \boldsymbol{u} = \boldsymbol{0}$

(7) $\|\boldsymbol{u} \times \boldsymbol{v}\| = \|\boldsymbol{u}\| \, \|\boldsymbol{v}\| \sin\theta = \sqrt{\|\boldsymbol{u}\|^2 \, \|\boldsymbol{v}\|^2 - (\boldsymbol{u} \cdot \boldsymbol{v})^2}$ (cross product length)

Here θ is the angle determined by $\boldsymbol{u}$ and $\boldsymbol{v}$, $0 \le \theta \le \pi$.

Proof:
Let $\boldsymbol{u} = (u_1, u_2, u_3)$, $\boldsymbol{v} = (v_1, v_2, v_3)$, and $\boldsymbol{w} = (w_1, w_2, w_3)$ be vectors in $\mathbb{R}^3$ and a a scalar. Selected proofs of the properties follow below.

(1)anticommutative law:

$$
\begin{aligned}
\boldsymbol{u} \times \boldsymbol{v} &= \begin{vmatrix} \boldsymbol{i} & \boldsymbol{j} & \boldsymbol{k} \\ u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{vmatrix} \\
&= (u_2v_3 - u_3v_2, u_3v_1 - u_1v_3, u_1v_2 - u_2v_1) \\
&= -1(v_2u_3 - v_3u_2, v_3u_1 - v_1u_3, v_1u_2 - v_2u_1) \\
&= -\begin{vmatrix} \boldsymbol{i} & \boldsymbol{j} & \boldsymbol{k} \\ v_1 & v_2 & v_3 \\ u_1 & u_2 & u_3 \end{vmatrix} \\
&= -\boldsymbol{v} \times \boldsymbol{u}
\end{aligned}
$$

(Exchanging rows in the determinant flips its sign as expected.)

[4]It may be wondered how the cross product, in our formulation, can represent physical quantities if its direction depends on the choice of a right-handed or left-handed coordinate system. In fact the cross product is known as a **pseudovector** or **axial vector**. If one tries to avoid appealing to coordinates by *defining* the cross product in terms of the right-hand rule the cross product still behaves unvector-like under improper rotations such as reflections, where it flips sign (direction) when $\boldsymbol{u}$ and $\boldsymbol{v}$ are reflected across a plane. This said, the cross product is invariant under proper rotations and finds many useful physical applications with angular momentum and torque being among them.

(5) $\boldsymbol{u} \times \boldsymbol{u} = \mathbf{0}$:

$$\begin{aligned}
\boldsymbol{u} \times \boldsymbol{u} &= \begin{vmatrix} \boldsymbol{i} & \boldsymbol{j} & \boldsymbol{k} \\ u_1 & u_2 & u_3 \\ u_1 & u_2 & u_3 \end{vmatrix} \\
&= (u_2u_3 - u_3u_2, u_3u_1 - u_1u_3, u_1u_2 - u_2u_1) \\
&= (0, 0, 0) \\
&= \mathbf{0}
\end{aligned}$$

(Determinant of matrix with two equal rows vanishes as expected.)

(7) cross product length:

$$\begin{aligned}
\|\boldsymbol{u} \times \boldsymbol{v}\|^2 &= \|(u_2v_3 - u_3v_2, u_3v_1 - u_1v_3, u_1v_2 - u_2v_1)\|^2 \\
&= (u_2v_3 - u_3v_2)^2 + (u_3v_1 - u_1v_3)^2 + (u_1v_2 - u_2v_1)^2 \\
&= u_2^2v_3^2 - 2u_2u_3v_2v_3 + u_3^2v_2^2 + u_3^2v_1^2 - 2u_1u_3v_1v_3 + u_1^2v_3^2 \\
&\quad + u_1^2v_3^2 + u_1^2v_2^2 - 2u_1u_2v_1v_2 + u_2^2v_1^2 \\
\|\boldsymbol{u}\|^2 \|\boldsymbol{v}\|^2 - (\boldsymbol{u} \cdot \boldsymbol{v})^2 &= (u_1^2 + u_2^2 + u_3^2)(v_1^2 + v_2^2 + v_3^2) - (u_1v_1 + u_2v_2 + u_3v_3)^2 \\
&= \cancel{u_1^2v_1^2} + u_1^2v_2^2 + u_1^2v_3^2 + u_2^2v_1^2 + \cancel{u_2^2v_2^2} + u_2^2v_3^2 + u_3^2v_1^2 + u_3^2v_2^2 + \cancel{u_3^2v_3^2} \\
&\quad - \cancel{u_1^2v_1^2} - \cancel{u_2^2v_2^2} - \cancel{u_3^2v_3^2} - 2u_1v_1u_2v_2 - 2u_1v_1u_3v_3 - 2u_2v_2u_3v_3 \\
&= u_2^2v_3^2 - 2u_2u_3v_2v_3 + u_3^2v_2^2 + u_3^2v_2^2 + u_3^2v_1^2 - 2u_1u_3v_1v_3 + u_1^2v_3^2 \\
&\quad + u_1^2v_2^2 - 2u_1u_2v_1v_2 + u_2^2v_1^2
\end{aligned}$$

Therefore

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

Furthermore,

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

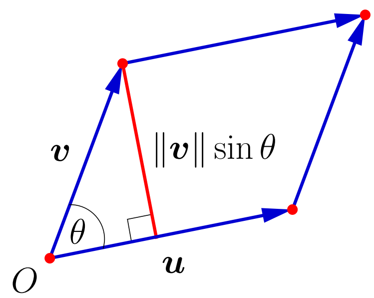
Thus

$$\begin{aligned}
\sqrt{\|\boldsymbol{u} \times \boldsymbol{v}\|^2} &= \sqrt{\|\boldsymbol{u}\|^2 \|\boldsymbol{v}\|^2 - (\boldsymbol{u} \cdot \boldsymbol{v})^2} = \sqrt{\|\boldsymbol{u}\|^2 \|\boldsymbol{v}\|^2 \sin^2\theta} \\
\|\boldsymbol{u} \times \boldsymbol{v}\| &= \sqrt{\|\boldsymbol{u}\|^2 \|\boldsymbol{v}\|^2 - (\boldsymbol{u} \cdot \boldsymbol{v})^2} = \|\boldsymbol{u}\| \|\boldsymbol{v}\| \sin\theta
\end{aligned}$$

Note here that $\sqrt{\sin^2\theta} = |\sin\theta| = \sin\theta$ since $0 \le \theta \le \pi$.

4.3.3 Area of a Parallelogram

The area of a parallelogram equals its base times its height. If one considers the parallelogram determined by $\boldsymbol{u}$ and $\boldsymbol{v}$, one has a convenient interpretation of the length of the cross product as the area of the parallelogram.



If θ is the angle between $\boldsymbol{u}$ and $\boldsymbol{v}$ the height of the parallelogram shown is $\|\boldsymbol{v}\| \sin\theta$ and we have that parallelogram area equals its base times its height,

$$\boxed{\text{Area Parallelogram} = \|\boldsymbol{u}\| \, \|\boldsymbol{v}\| \sin\theta = \|\boldsymbol{u} \times \boldsymbol{v}\|},$$

where the last equality follows by Theorem 4-14 .

Example 4-19

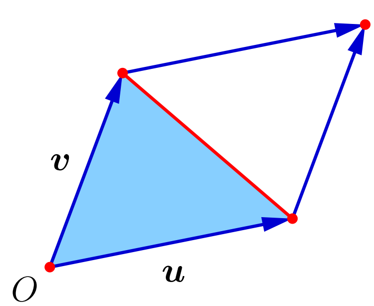
Find the area of the parallelogram determined by the vectors $\boldsymbol{u} = (1, 4, 4)$ and $\boldsymbol{v} = (0, 3, 2)$.
Solution:

$$\begin{aligned}
\boldsymbol{u} \times \boldsymbol{v} &= \begin{vmatrix} \boldsymbol{i} & \boldsymbol{j} & \boldsymbol{k} \\ 1 & 4 & 4 \\ 0 & 3 & 2 \end{vmatrix} \\
&= (8 - 12, 0 - 2, 3 - 0) \\
&= (-4, -2, 3) \\
A = ||\boldsymbol{u} \times \boldsymbol{v}|| &= \sqrt{(-4)^2 + (-2)^2 + (3)^2} \\
&= \sqrt{29}
\end{aligned}$$

The area of the parallelogram is $\sqrt{29}$ square units.

4.3.4 Area of a Triangle

The triangle determined by the terminal points of $\boldsymbol{u}$ and $\boldsymbol{v}$ and the origin O is just half the parallelogram determined by $\boldsymbol{u}$ and $\boldsymbol{v}$.



Therefore the area of the triangle determined by $\boldsymbol{u}$ and $\boldsymbol{v}$ is given by

$$\boxed{\text{Area Triangle} = \frac{1}{2} \|\boldsymbol{u} \times \boldsymbol{v}\|}.$$

Example 4-20

Find the area of the triangle determined by the points $P_1(2,2,0)$, $P_2(-1,0,2)$, and $P_3(0,4,3)$.
Solution:
Let $\boldsymbol{p}_1$, $\boldsymbol{p}_2$ and $\boldsymbol{p}_3$ have terminal points P_1, P_2 and P_3 respectively. Then $\overrightarrow{P_1P_2}$ is equivalent to

$$\boldsymbol{u} = \boldsymbol{p}_2 - \boldsymbol{p}_1 = (-1,0,2) - (2,2,0) = (-3,-2,2)\,,$$

and $\overrightarrow{P_1P_3}$ is equivalent to

$$\boldsymbol{v} = \boldsymbol{p}_3 - \boldsymbol{p}_2 = (0,4,3) - (2,2,0) = (-2,2,3)\,.$$

The triangle determined by P_1, P_2, and P_3 is symmetric to that determined by $\boldsymbol{u}$ and $\boldsymbol{v}$ at the origin. Their cross product is

$$\begin{aligned}\boldsymbol{u}\times\boldsymbol{v} &= \begin{vmatrix} \boldsymbol{i} & \boldsymbol{j} & \boldsymbol{k} \\ -3 & -2 & 2 \\ -2 & 2 & 3 \end{vmatrix} \\ &= (-6-4, -4+9, -6-4) \\ &= (-10, 5, -10)\,. \end{aligned}$$

So the area of the symmetric triangles equal

$$\begin{aligned}\text{Area Triangle} &= \frac{1}{2}\left\|\boldsymbol{u}\times\boldsymbol{v}\right\| \\ &= \frac{1}{2}\sqrt{(-10)^2 + 5^2 + (-10)^2} \\ &= \frac{1}{2}\sqrt{225} \\ &= \frac{1}{2}(15) \\ &= \frac{15}{2}\ \text{square units.}\end{aligned}$$

4.4 Scalar Triple Product

A scalar can be formed from three vectors as follows.[5]

Definition: If $\boldsymbol{u}$, $\boldsymbol{v}$, and $\boldsymbol{w}$ are vectors in $\mathbb{R}^3$, then $\boldsymbol{u}\cdot(\boldsymbol{v}\times\boldsymbol{w})$ is called the **scalar triple product** of $\boldsymbol{u}, \boldsymbol{v}$ and $\boldsymbol{w}$.

Theorem 4-15: The scalar triple product satisfies

$$\boldsymbol{u}\cdot(\boldsymbol{v}\times\boldsymbol{w}) = \begin{vmatrix} u_1 & v_1 & w_1 \\ u_2 & v_2 & w_2 \\ u_3 & v_3 & w_3 \end{vmatrix}.$$

From this determinant formula for the scalar triple product it follows that

$$\boldsymbol{u}\cdot(\boldsymbol{v}\times\boldsymbol{w}) = \boldsymbol{v}\cdot(\boldsymbol{w}\times\boldsymbol{u}) = \boldsymbol{w}\cdot(\boldsymbol{u}\times\boldsymbol{v}),$$

since these require two column exchanges to accomplish.

Example 4-21

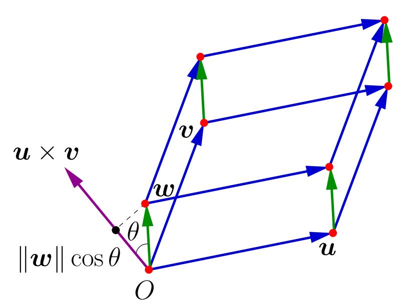
Calculate the scalar triple product of $\boldsymbol{u} = (3, -2, -5)$, $\boldsymbol{v} = (1, 4, -4)$, and $\boldsymbol{w} = (0, 3, 2)$.

Solution: Using the determinant formula evaluated along the last row gives

$$\begin{aligned}\boldsymbol{u}\cdot(\boldsymbol{v}\times\boldsymbol{w}) &= \begin{vmatrix} 3 & 1 & 0 \\ -2 & 4 & 3 \\ -5 & -4 & 2 \end{vmatrix} \\ &= 3(+1)(8-(-12)) + (1)(-1)(-4-(-15)) \\ &= 3(20) - (1)(11) \\ &= 60 - 11 \\ &= 49\end{aligned}$$

4.4.1 Volume of a Parallelepiped

The scalar triple product has a useful geometrical application. Three vectors $\boldsymbol{u}$, $\boldsymbol{v}$, and $\boldsymbol{w}$ in $\mathbb{R}^3$ that are not coplanar will determine a **parallelepiped**.



[5] Technically the scalar triple product is a pseudoscalar since the presence of the cross product in its definition causes it to change sign under improper transformations such as reflections.

The volume will be the area of the parallelogram of its base, $\|\boldsymbol{u} \times \boldsymbol{v}\|$, times its height. Since the vector $\boldsymbol{u} \times \boldsymbol{v}$ is orthogonal to the plane of the parallelogram the height is just the absolute value of the projection of $\boldsymbol{w}$ onto $\boldsymbol{u} \times \boldsymbol{v}$, namely

$$\|\boldsymbol{w}\| \, |\cos\theta| \, ,$$

where θ is the angle between $\boldsymbol{w}$ and $\boldsymbol{u} \times \boldsymbol{v}$. The volume is therefore

$$(\text{height})(\text{base area}) = \|\boldsymbol{w}\| \, |\cos\theta| \, \|\boldsymbol{u} \times \boldsymbol{v}\| = | \, \|\boldsymbol{w}\| \, \|\boldsymbol{u} \times \boldsymbol{v}\| \cos\theta|$$

with the result that the volume of is just the absolute value of one of the forms of the scalar triple product of the three vectors, $|\boldsymbol{w} \cdot (\boldsymbol{u} \times \boldsymbol{v})|$. Using the form of the scalar triple product we originally introduced gives the formula:

$$\boxed{\text{Volume Parallelepiped} = |\boldsymbol{u} \cdot (\boldsymbol{v} \times \boldsymbol{w})|} \, .$$

One notes that the three formulas for the scalar triple product reflect the fact that any one of the three sides can be considered the base of the parallelepiped.

Example 4-22

Find the volume of the parallelepiped generated by the vectors $\boldsymbol{u} = (3, -2, -5)$, $\boldsymbol{v} = (1, 4, -4)$, and $\boldsymbol{w} = (0, 3, 2)$.

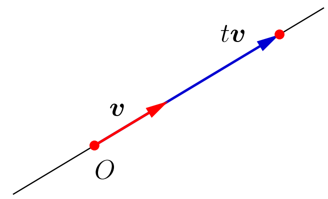
Solution:
Using the result from Example 4-21 we have

$$\text{Volume} = |\boldsymbol{u} \cdot (\boldsymbol{v} \times \boldsymbol{w})| = |49| = 49 \ (\text{units}^3) \, .$$

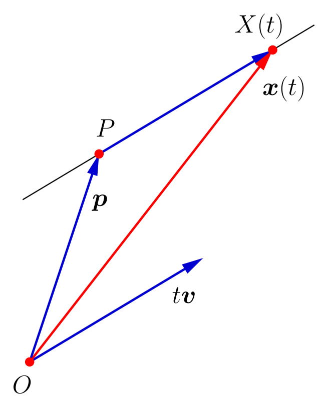
Chapter 5: Lines and Planes

5.1 Point-Parallel Form of a Line

Consider a vector $\boldsymbol{v} \neq \mathbf{0}$. The terminal points of the scalar multiples of $\boldsymbol{v}$, given by $t\boldsymbol{v}$ for t any scalar, will be collinear, by Theorem 4-2. Since $t = 0$ implies $\mathbf{0}$ is one of these points, the line goes through the origin O.



An arbitrary line going through a point P with direction $\boldsymbol{v}$ can be formed by adding the vector $t\boldsymbol{v}$ to the vector $\boldsymbol{p}$ with terminal point P.



We summarize these observations with the following definition.

Definition: The equation of a line that passes through a given point P and is parallel to a given vector $\boldsymbol{v} \neq \mathbf{0}$ is given by:

$$\boxed{\boldsymbol{x}(t) = \boldsymbol{p} + t\boldsymbol{v}},$$

where t is a scalar parameter. Here P is the terminal point of vector $\boldsymbol{p}$. This is called the **point-parallel form** of a line.

Example 5-1

Find a point-parallel form for the line in $\mathbb{R}^3$ that passes through the point $P(2, 1, -3)$ and is parallel to the vector $\boldsymbol{v} = (1, 2, 2)$.

Solution:
Letting $\boldsymbol{p} = (2, 1, -3)$ be the vector with terminal point P we have by the previous formula.

$$\boldsymbol{x}(t) = \boldsymbol{p} + t\boldsymbol{v} = (2, 1, -3) + t(1, 2, 2)\,,$$

where t is a scalar.

We note that while the point-parallel form of a line can be visualized in two and three dimensions, the formula can be used to characterize lines more generally in $\mathbb{R}^n$.

5.1.1 Parametric Equations of a Line

Consider the vector equation $\boldsymbol{x}(t) = \boldsymbol{p} + t\boldsymbol{v}$. In $\mathbb{R}^3$ we can write $\boldsymbol{x} = (x, y, z)$, $\boldsymbol{p} = (x_0, y_0, z_0)$, and $\boldsymbol{v} = (v_1, v_2, v_3)$. Then:

$$\begin{aligned} \boldsymbol{x}(t) &= \boldsymbol{p} + t\boldsymbol{v} \\ (x, y, z) &= (x_0, y_0, z_0) + t(v_1, v_2, v_3) \\ &\Rightarrow \begin{cases} x &= x_0 + tv_1 \\ y &= y_0 + tv_2 \\ z &= z_0 + tv_3 \end{cases} \end{aligned}$$

These are called **parametric equations** for the line.

Parametric equations for a line in $\mathbb{R}^2$ are similarly $\begin{cases} x = x_0 + tv_1 \\ y = y_0 + tv_2 \end{cases}$.

Example 5-2

Find parametric equations for the line passing through the point $P(2, 1, -1)$ that is parallel to the vector $\boldsymbol{v} = (-1, 1, 3)$ and determine if the point $Q(0, 5, 5)$ is on the line.

Solution:

$$\begin{aligned} \boldsymbol{x}(t) &= \boldsymbol{p} + t\boldsymbol{v} \\ (x, y, z) &= (2, 1, -1) + t(-1, 1, 3) \\ (x, y, z) &= (2 - t, 1 + t, -1 + 3t) \end{aligned}$$

Parametric equations for the line are therefore

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

The point $Q(1, 5, 5)$ is on the line if the overdetermined linear system

$$\begin{aligned} 0 &= 2 - t \\ 5 &= 1 + t \\ 5 &= -1 + 3t \end{aligned}$$

has a solution for t. The first equation implies $t = 2$ but while this satisfies the third equation it fails to satisfy the second. Therefore the point Q is not on the line.

5.2 Two-Point Form of a Line

Let P and Q be distinct points. If vectors $\boldsymbol{p}$ and $\boldsymbol{q}$ have these terminal points then the vector $\boldsymbol{v} = \boldsymbol{q} - \boldsymbol{p}$ will be parallel to the line determined by the points. Inserting this into the point-parallel equation $\boldsymbol{x}(t) = \boldsymbol{p} + t\boldsymbol{v}$ gives the following.

Definition: The **two-point form** of a line determined by points P and Q is:

$$\boxed{\boldsymbol{x}(t) = (1-t)\boldsymbol{p} + t\boldsymbol{q}},$$

where P and Q are terminal points of vectors $\boldsymbol{p}$ and $\boldsymbol{q}$ and t is a scalar parameter.

One observes that $\boldsymbol{x} = \boldsymbol{p}$ when $t = 0$ (so point P) and $\boldsymbol{x} = \boldsymbol{q}$ when $t = 1$ (so point Q) with this parameterization of the line.

Example 5-3

Describe the line that passes through the points $P(1, -1, 3)$ and $Q(1, 2, -4)$ in both two-point form and parametric form.

$$\begin{aligned} \boldsymbol{x}(t) &= (1-t)\boldsymbol{p} + \boldsymbol{q} \\ \boldsymbol{x}(t) &= (1-t)(1, -1, 3) + t(1, 2, -4) \qquad \text{(two-point form)} \end{aligned}$$

Expanding gives the parametric form:

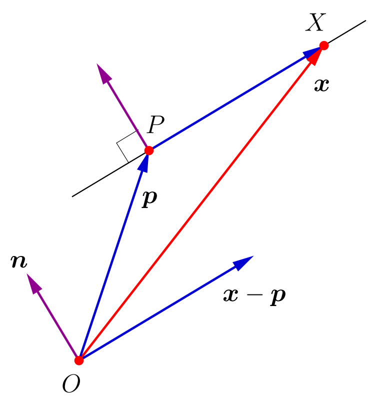
$$\begin{aligned} (x, y, z) &= (1-t, -1+t, 3-3t) + (t, 2t, -4t) \\ &= (1, -1+3t, 3-7t) \\ &\Rightarrow \begin{cases} x = 1 \\ y = -1 + 3t \\ z = 3 - 7t \end{cases} \qquad \text{(parametric form)} \end{aligned}$$

5.3 Point-Normal Form of a Line

If we consider two-dimensional space $\mathbb{R}^2$ a **normal vector** $\boldsymbol{n}$ to a line in $\mathbb{R}^2$ is a nonzero vector that is perpendicular to the direction $\boldsymbol{v}$ of the line, $\boldsymbol{n} \cdot \boldsymbol{v} = 0$. Assuming a point-parallel description of a line $\boldsymbol{x} = \boldsymbol{p} + t\boldsymbol{v}$ it follows that $\boldsymbol{x} - \boldsymbol{p} = t\boldsymbol{v}$ and we have

$$\boldsymbol{n} \cdot (\boldsymbol{x} - \boldsymbol{p}) = \boldsymbol{n} \cdot (t\boldsymbol{v}) = t(\boldsymbol{n} \cdot \boldsymbol{v}) = t(0) = 0\,,$$

for any point X on the line.



This suggests the following alternate description of a line in $\mathbb{R}^2$ in terms of a normal vector.

Definition: The **point-normal form** of a line in $\mathbb{R}^2$ passing through a given point $P(x_0, y_0)$ that is normal (perpendicular) to a given vector $\boldsymbol{n} = (a, b)$ is given by the equation

$$\boxed{\boldsymbol{n} \cdot (\boldsymbol{x} - \boldsymbol{p}) = 0}\,,$$

where $X(x, y)$ is any point on the line. Here P and X are the terminal points of vectors $\boldsymbol{p}$ and $\boldsymbol{x}$ respectively.

Example 5-4

Find the equation of a line passing through $P(1, -1)$ with normal $\boldsymbol{n} = (2, -1)$.

Solution:
Since we are given a point and normal for the line we use the point-normal form recipe with $\boldsymbol{x} = (x, y)$ and $\boldsymbol{p} = (1, -1)$ to get:

$$\begin{aligned} \boldsymbol{n} \cdot (\boldsymbol{x} - \boldsymbol{p}) &= 0 \\ (2, -1) \cdot (x - 1, y + 1) &= 0 \\ 2(x - 1) - (y + 1) &= 0 && \text{(point-normal form)} \end{aligned}$$

Further expansion gives the standard form:

$$\begin{aligned} 2x - 2 - y - 1 &= 0 \\ 2x - y &= 3 && \text{(standard form)} \end{aligned}$$

If we expand the point-normal equation in terms of the vector components we can recover the standard form of the line as follows.

$$\begin{aligned}
\boldsymbol{n}\cdot(\boldsymbol{x}-\boldsymbol{p}) &= 0\\
(a,b)\cdot[(x,y)-(x_0,y_0)] &= 0\\
(a,b)\cdot(x-x_0,y-y_0) &= 0\\
a(x-x_0)+b(y-y_0) &= 0\\
ax-ax_0+by-by_0 &= 0\\
ax+by &= \underbrace{ax_0+by_0}_{=c}
\end{aligned}$$

Consideration of the last line shows that the coefficients of x and y in the standard form of a line in $\mathbb{R}^2$ have the geometrical interpretation as the components (a,b) of a normal $\boldsymbol{n}$ to the line. If one desires a vector $\boldsymbol{v}$ that is parallel to the line one observes that

$$(a,b)\cdot(b,-a) = ab-ba = 0,$$

which shows $\boxed{\boldsymbol{v}=(b,-a)}$ will be a nonzero vector perpendicular to $\boldsymbol{n}$ which in two dimensions implies it is parallel to the line. Finally a point P on the line can be found by choosing an arbitrary value of x and solving for y (or vice versa if coefficient $b=0$).

Example 5-5

Find a point-parallel form for the line in $\mathbb{R}^2$ given by the equation $2x+3y=1$.

Solution:
We are given the standard form $(ax+by=c)$ of $2x+3y=1$. We need to find a point on the line and a vector parallel to it.

When $x=0$:

$$\begin{aligned}
2x+3y &= 1\\
2(0)+3y &= 1\\
y &= \frac{1}{3}\,.
\end{aligned}$$

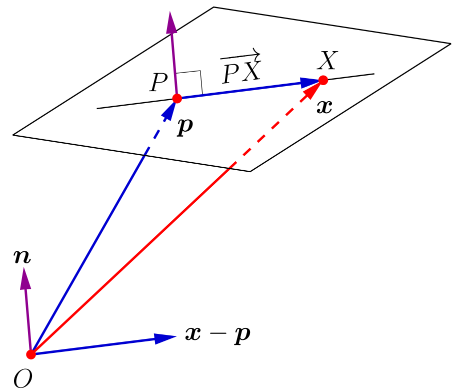
Therefore the point $P\left(0,\frac{1}{3}\right)$ is on the line.

The vector $\boldsymbol{n}=(a,b)=(2,3)$ is normal to the line. Then $\boldsymbol{v}=(b,-a)=(3,-2)$ is therefore orthogonal to it. Using the point-parallel form recipe gives

$$\begin{aligned}
\boldsymbol{x}(t) &= \boldsymbol{p}+t\boldsymbol{v}\\
\boldsymbol{x}(t) &= \left(0,\frac{1}{3}\right)+t(3,-2) \qquad \text{(point-parallel form)}
\end{aligned}$$

5.4 Point-Normal Form of a Plane

In three dimensions a plane may be determined by a point P through which it passes and a normal vector $\boldsymbol{n}$ which is perpendicular to the plane. We want to find the equation of the plane passing through a given point $P(x_0, y_0, z_0)$ and perpendicular to a given vector $\boldsymbol{n} = (a, b, c)$.



Let $X(x, y, z)$ be any other point on the plane. P and X determine a line in the plane. The directed line segment $\overrightarrow{PX}$ is equivalent to the vector $\boldsymbol{x} - \boldsymbol{p}$ which must therefore be orthogonal to $\boldsymbol{n}$:

$$\boxed{\boldsymbol{n} \cdot (\boldsymbol{x} - \boldsymbol{p}) = 0},$$

where as usual X and P are the terminal points of $\boldsymbol{x}$ and $\boldsymbol{p}$ respectively. Any point X on the plane has to satisfy this equation. With $\boldsymbol{n} = (a, b, c)$ and $\boldsymbol{x} - \boldsymbol{p} = (x - x_0, y - y_0, z - z_0)$, then the equation of the plane becomes:

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

These equations are the **point-normal form** for the plane.

We can rewrite the point-normal form as follows:

$$\begin{aligned} \boldsymbol{n} \cdot (\boldsymbol{x} - \boldsymbol{p}) &= 0 \\ \boldsymbol{n} \cdot \boldsymbol{x} - \boldsymbol{n} \cdot \boldsymbol{p} &= 0 \\ \boldsymbol{n} \cdot \boldsymbol{x} &= \boldsymbol{n} \cdot \boldsymbol{p} \end{aligned}$$

With $\boldsymbol{n} = (a, b, c)$ and setting the constant $d = \boldsymbol{n} \cdot \boldsymbol{p} = ax_0 + by_0 + cz_0$ this becomes

$$\boxed{ax + by + cz = d}.$$

This is called the **standard form** of the equation of the plane. In other words a linear equation in three dimensions written in this form will represent a plane with normal $\boldsymbol{n} = (a, b, c)$ provided one of a, b, or c is nonzero. This is analogous to the normal to a line appearing in the standard form of a linear equation in $\mathbb{R}^2$. Note that the standard form of a plane is not unique since one can multiply the equation by a nonzero scalar to get an equivalent equation. Geometrically this is just scaling the normal vector by that amount to produce a new normal to the plane.

Example 5-6

Find a point-normal and standard form of the equation of the plane that passes through the given point P which is perpendicular to the given vector $\boldsymbol{n}$.

1. $P(1,3,-2)\,, \boldsymbol{n} = (-2,1,-1)$

   Solution:
   Let $X(x,y,z)$ be any other point on the plane, then $\overrightarrow{PX}$ in the plane is equivalent to the vector $\boldsymbol{x} - \boldsymbol{p} = (x-1, y-3, z+2)$ and a point normal form for the plane is

$$\begin{aligned} \boldsymbol{n} \cdot (\boldsymbol{x} - \boldsymbol{p}) &= 0 \\ (-2,1,-1) \cdot (x-1, y-3, z+2) &= 0 \\ -2(x-1) + 1(y-3) + (-1)(z+2) &= 0 \qquad \text{(point-normal form)} \end{aligned}$$

   Expanding gives the standard form:

$$\begin{aligned} -2x + 2 + y - 3 - z - 2 &= 0 \\ -2x + y - z &= 3 \qquad \text{(standard form)} \end{aligned}$$

2. $P(1,1,4)\,, \boldsymbol{n} = (1,9,8)$

   Solution:
   Let $X(x,y,z)$ be any other point on the plane, then $\overrightarrow{PX}$ in the plane is equivalent to the vector $\boldsymbol{x} - \boldsymbol{p} = (x-1, y-1, z-4)$ and the point-normal form of the plane is

$$\begin{aligned} \boldsymbol{n} \cdot (\boldsymbol{x} - \boldsymbol{p}) &= 0 \\ (1,9,8) \cdot (x-1, y-1, z-4) &= 0 \\ 1(x-1) + 9(y-1) + 8(z-4) &= 0 \qquad \text{(point-normal form)} \end{aligned}$$

   Expanding gives:

$$\begin{aligned} x - 1 + 9y - 9 + 8z - 32 &= 0 \\ x + 9y + 8z &= 42 \qquad \text{(standard form)} \end{aligned}$$

Example 5-7

Find the point of intersection of the line $\boldsymbol{x}(t) = (2,1,1) + t(-1,0,4)$ and the plane $x - 3y - z = 1$.

Solution:
The point on the line will be determined by the value of t in the point-parallel form of the line. The parametric form of the line is given by:

$$\begin{aligned} \boldsymbol{x}(t) &= (2,1,1) + t(-1,0,4) \\ \boldsymbol{x}(t) &= (2-t, 1, 1+4t) \\ &\Rightarrow \begin{cases} x = 2 - t \\ y = 1 \\ z = 1 + 4t \end{cases} \end{aligned}$$

To also sit on the plane (x,y,z) must additionally satisfy the planar equation. Inserting the parametric form into that equation gives:

$$\begin{aligned} x - 3y - z &= 1 \\ (2-t) - 3(1) - (1+4t) &= 1 \\ 2 - t - 3 - 1 - 4t &= 1 \\ -5t &= 3 \\ t &= -\frac{3}{5} \end{aligned}$$

The point of intersection has coordinates:

$$\begin{aligned} &x = 2 - \left(-\frac{3}{5}\right) = \frac{13}{5} \\ &y = 1 \\ &z = 1 + 4\left(-\frac{3}{5}\right) = -\frac{7}{5} \end{aligned}$$

Therefore the point is $P\left(\frac{13}{5}, 1, -\frac{7}{5}\right)$.

The standard form of a plane in $\mathbb{R}^3$ is just a linear equation. Thus a linear system of m equations in three unknowns has the geometrical interpretation of the intersection of these m planes. We require planar intersection because a solution to the linear system must satisfy all its equations.

Example 5-8

The equations given below represent planes in $\mathbb{R}^3$. Describe geometrically the given solution set of these equations

1. $$\begin{aligned} x + y + z &= -4 \\ x + 2y &= 1 \\ 2y + 3z &= -2 \end{aligned} \quad , \quad \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} -\frac{21}{5} \\ \frac{13}{5} \\ -\frac{12}{5} \end{bmatrix}$$

   Solution:

   The solution represents the single intersection point $P\left(-\frac{21}{5}, \frac{13}{5}, -\frac{12}{5}\right)$ of the 3 planes.

2. $$\begin{aligned} x - y + z &= 0 \\ y - 2z &= 1 \\ 2x - y &= 1 \end{aligned} \quad , \quad \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 1+t \\ 1+2t \\ t \end{bmatrix}$$

   Solution:

   Since
   $$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + t \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix},$$
   this is a line that passes through $P(1,1,0)$ and is parallel to the vector $\boldsymbol{v} = (1,2,1)$.

3.
$$\begin{aligned} x + 4y - 5z &= 0 \\ 2x - y + 8z &= 9 \end{aligned} \qquad , \qquad \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 4 - 3t \\ -1 + 2t \\ t \end{bmatrix}$$

Solution:

Since
$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 4 \\ -1 \\ 0 \end{bmatrix} + t \begin{bmatrix} -3 \\ 2 \\ 1 \end{bmatrix},$$
this is a line that passes through $P(4, -1, 0)$ and is parallel to the vector $\boldsymbol{v} = (-3, 2, 1)$.

5.5 Plane through Three Points

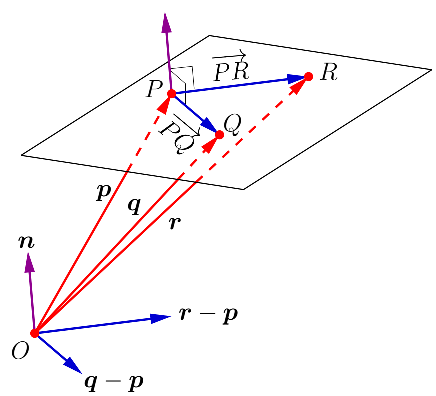
Three noncollinear points in $\mathbb{R}^3$ will determine a plane. To find the equation of the plane that passes through the three noncollinear points P, Q, R we consider the directed line segments $\overrightarrow{PQ}$ and $\overrightarrow{PR}$. These lie in the plane and are equivalent to the vectors $\boldsymbol{q}-\boldsymbol{p}$ and $\boldsymbol{r}-\boldsymbol{p}$ respectively. Recall the cross product of two vectors is orthogonal to both vectors. Therefore a normal vector $\boldsymbol{n}$ to the plane is given by

$$\boxed{\boldsymbol{n} = (\boldsymbol{q}-\boldsymbol{p}) \times (\boldsymbol{r}-\boldsymbol{p})},$$

which can then be inserted into the point-normal equation,

$$\boldsymbol{n} \cdot (\boldsymbol{x}-\boldsymbol{p}) = 0\,,$$

to find an equation for the plane.



Example 5-9

Find a point-normal form and standard form for the plane passing through the points $P(-1,1,3)$, $Q(0,3,1)$, and $R(2,1,-1)$.

Solution:
First find a normal $\boldsymbol{n}$. $\overrightarrow{PQ}$ and $\overrightarrow{PR}$ are equivalent to $\boldsymbol{q}-\boldsymbol{p}$ and $\boldsymbol{r}-\boldsymbol{p}$ where

$$\begin{aligned} \boldsymbol{q}-\boldsymbol{p} &= (0,3,1)-(-1,1,3) = (1,2,-2) \\ \boldsymbol{r}-\boldsymbol{p} &= (2,1,-1)-(-1,1,3) = (3,0,-4) \end{aligned}$$

Taking the cross product gives

$$\begin{aligned} \boldsymbol{n} = (\boldsymbol{q}-\boldsymbol{p}) \times (\boldsymbol{r}-\boldsymbol{p}) &= \begin{vmatrix} \boldsymbol{i} & \boldsymbol{j} & \boldsymbol{k} \\ 1 & 2 & -2 \\ 3 & 0 & -4 \end{vmatrix} \\ &= (-8+0, -6+4, 0-6) \\ &= (-8,-2,-6) \end{aligned}$$

Substitute into the point-normal equation:

$$\begin{aligned} \boldsymbol{n} \cdot (\boldsymbol{x}-\boldsymbol{p}) &= 0 \\ (-8,-2,-6) \cdot (x+1, y-1, z-3) &= 0 \\ -8(x+1)+(-2)(y-1)+(-6)(z-3) &= 0 \qquad \text{(point-normal form)} \end{aligned}$$

Expanding gives the standard form:

$$\begin{aligned}
-8x - 8 - 2y + 2 - 6z + 18 &= 0 \\
-8x - 2y - 6z + 12 &= 0 \\
-8x - 2y - 6z &= -12 \\
4x + y + 3z &= 6 \qquad \text{(standard form)}
\end{aligned}$$

Here we simplified the equation by dividing both sides by -2 .

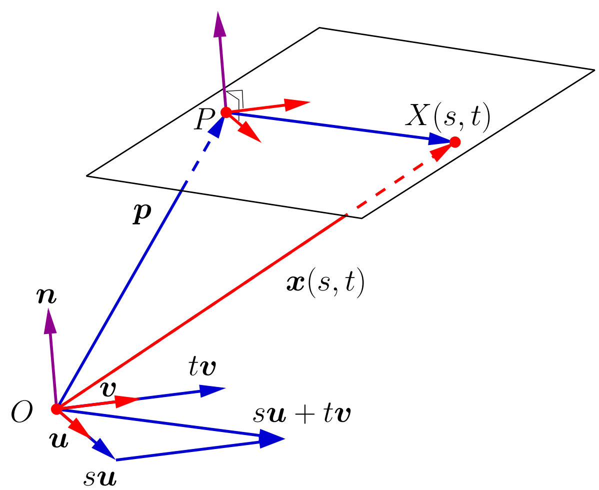
5.6 Point-Parallel Form of a Plane

If $P(x_0, y_0, z_0)$ is a point and $\boldsymbol{u} = (u_1, u_2, u_3)$ and $\boldsymbol{v} = (v_1, v_2, v_3)$ are nonzero, noncollinear vectors then the lines $\boldsymbol{x}(s) = \boldsymbol{p} + s\boldsymbol{u}$ and $\boldsymbol{x}(t) = \boldsymbol{p} + t\boldsymbol{v}$ with parameters s and t will intersect at P and will determine a plane.

Definition: The **point-parallel form** of a plane in $\mathbb{R}^3$ through point P that is parallel to noncollinear vectors $\boldsymbol{u} \neq \boldsymbol{0}$ and $\boldsymbol{v} \neq \boldsymbol{0}$ is given by

$$\boxed{\boldsymbol{x}(s,t) = \boldsymbol{p} + s\boldsymbol{u} + t\boldsymbol{v}},$$

where s and t are scalar parameters and P is the terminal point of vector $\boldsymbol{p}$.



That the lines $\boldsymbol{x}(s) = \boldsymbol{p} + s\boldsymbol{u}$ and $\boldsymbol{x}(t) = \boldsymbol{p} + t\boldsymbol{v}$ lie in the surface generated by $\boldsymbol{x}(s,t) = \boldsymbol{p} + s\boldsymbol{u} + t\boldsymbol{v}$ is clear by setting $t = 0$ or $s = 0$ respectively in the latter equation. To see that the points X given by $\boldsymbol{x}(s,t)$ really do lie on a plane, note that a normal to the plane, orthogonal to both $\boldsymbol{u}$ and $\boldsymbol{v}$, is $\boldsymbol{n} = \boldsymbol{u} \times \boldsymbol{v}$ and we have

$$\begin{aligned}
\boldsymbol{n} \cdot (\boldsymbol{x} - \boldsymbol{p}) &= \boldsymbol{n} \cdot (\boldsymbol{p} + s\boldsymbol{u} + t\boldsymbol{v} - \boldsymbol{p}) \\
&= \boldsymbol{n} \cdot (s\boldsymbol{u} + t\boldsymbol{v}) \\
&= \boldsymbol{n} \cdot (s\boldsymbol{u}) + \boldsymbol{n} \cdot (t\boldsymbol{v}) \\
&= s\boldsymbol{n} \cdot \boldsymbol{u} + t\boldsymbol{n} \cdot \boldsymbol{v} \\
&= s(0) + t(0) \\
&= 0
\end{aligned}$$

as required.

5.6.1 Parametric Equations of a Plane

Writing out the vector equation $\boldsymbol{x} = \boldsymbol{p} + s\boldsymbol{u} + t\boldsymbol{v}$ in terms of components with $\boldsymbol{x} = (x, y, z)$, $\boldsymbol{p} = (x_0, y_0, z_0)$, $\boldsymbol{u} = (u_1, u_2, u_3)$, and $\boldsymbol{v} = (v_1, v_2, v_3)$ we get

$$\begin{aligned}
\boldsymbol{x} &= \boldsymbol{p} + s\boldsymbol{u} + t\boldsymbol{v} \\
(x, y, z) &= (x_0, y_0, z_0) + s(u_1, u_2, u_3) + t(v_1, v_2, v_3) \\
(x, y, z) &= (x_0 + su_1 + tv_1, y_0 + su_2 + tv_2, z_0 + su_3 + tv_3)
\end{aligned}$$

Therefore

$$\begin{cases} x &= x_0 + su_1 + tv_1 \\ y &= y_0 + su_2 + tv_2 \\ z &= z_0 + su_3 + tv_3 \end{cases} .$$

These are called **parametric equations** for the plane.

Example 5-10

Find the point-parallel form and parametric equations of the plane passing through the point $P(-2,1,3)$ and parallel to the vectors $\boldsymbol{u} = (1,1,-1)$ and $\boldsymbol{v} = (-1,2,0)$.

Solution:

$$\begin{aligned} \boldsymbol{x} &= \boldsymbol{p} + s\boldsymbol{u} + t\boldsymbol{v} \\ (x,y,z) &= (-2,1,3) + s(1,1,-1) + t(-1,2,0) && \text{(point-parallel form)} \\ (x,y,z) &= (-2+s-t, 1+s+2t, 3-s) \\ &\Rightarrow \begin{cases} x = -2+s-t \\ y = 1+s+2t \\ z = 3-s \end{cases} && \text{(parametric equations)} \end{aligned}$$

Example 5-11

Given the plane in standard form

$$2x + 2y - 4z = 10\,,$$

find a point-parallel form and parametric equations for the plane.

Solution:
To find the point-parallel form we solve the linear system containing the single equation

$$2x + 2y - 4z = 10\,.$$

The corresponding augmented matrix is reduced to RREF:

$$\left[\begin{array}{ccc|c} 2 & 2 & -4 & 10 \end{array}\right] \Rightarrow R_1 \to \tfrac{1}{2}R_1 \left[\begin{array}{ccc|c} 1 & 1 & -2 & 5 \end{array}\right] \Leftrightarrow x + y - 2z = 5\,.$$

Assign the free (independent) variables y and z to parameters so that $\boxed{y = s}$ and $\boxed{z = t}$. Solving for the leading (dependent) variable x gives

- $x + y - 2z = 5 \implies x + s - 2t = 5 \implies \boxed{x = 5 - s + 2t}$.

Writing the solution in terms of vectors gives

$$\boldsymbol{x}(s,t) = \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 5-s+2t \\ s \\ t \end{bmatrix} = \underbrace{\begin{bmatrix} 5 \\ 0 \\ 0 \end{bmatrix}}_{\boldsymbol{p}} + s\underbrace{\begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}}_{\boldsymbol{u}} + t\underbrace{\begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix}}_{\boldsymbol{v}} . \qquad \text{(point-parallel form)}$$

One sees that the particular solution here is geometrically a point $\boldsymbol{p}$ on the plane, while the homogeneous solution, $s\boldsymbol{u} + t\boldsymbol{v}$, gives the offset vector from that point.

The first vector equation implies

$$\begin{cases} x = 5 - s + 2t \\ y = s \\ z = t \end{cases} . \qquad \text{(parametric equations)}$$

5.7 Distance to Lines and Planes

A common geometrical problem is to find the distance d between a line (in $\mathbb{R}^2$ or $\mathbb{R}^3$) or plane (in $\mathbb{R}^3$) and a point P off the line or plane. The problem may be solved by finding the closest point Q on the line or plane to P. The distance is then $\|\boldsymbol{q}-\boldsymbol{p}\|$. Summary of our general approach is as follows:

1. Find a point-parallel equation for a line on which Q lies.
2. Substitute $\boldsymbol{q}(t)$ into a geometrical constraint equation to get an equation involving only parameter t.
3. Solve this for t to find $\boldsymbol{q}$ (and hence Q).
4. Calculate the distance $d = d(\boldsymbol{p},\boldsymbol{q}) = \|\boldsymbol{q}-\boldsymbol{p}\|$.

5.7.1 Point-Parallel Line

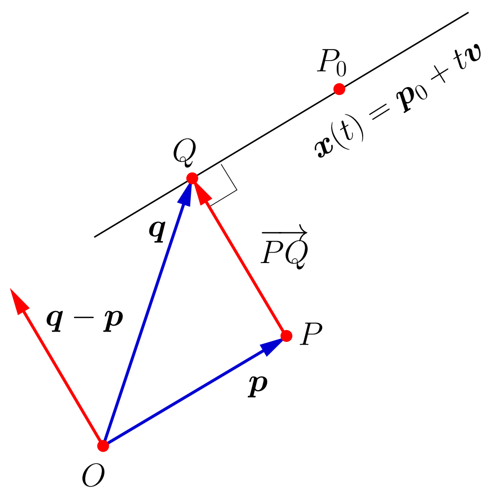
As an example, suppose one wishes to find the closest point and distance from P to a line, in either $\mathbb{R}^2$ or $\mathbb{R}^3$, where the line is described by the point-parallel form $\boldsymbol{x}(t) = \boldsymbol{p}_0 + t\boldsymbol{v}$. Then Q clearly lies on this line and thus satisfies

$$\boxed{\boldsymbol{q}(t) = \boldsymbol{p}_0 + t\boldsymbol{v}}$$

for some particular value of the parameter t to be determined. The directed line segment $\overrightarrow{PQ}$ must be perpendicular to the line and it follows that the constraint required to find t is just

$$\boxed{(\boldsymbol{q}-\boldsymbol{p})\cdot\boldsymbol{v} = 0},$$

into which we substitute $\boldsymbol{q}(t)$ from above and solve.



> **Example 5-12**
>
> Find the distance between the point $P(2,5)$ and the line $\boldsymbol{x}(t) = (6,3) + t(4,3)$. Also find the closest point Q on the line.
>
> Solution:
> The point Q, the tip of vector $\boldsymbol{q}$, must lie on the line so we have
>
> $$\boldsymbol{q} = (6,3) + t(4,3)$$

for some *particular* value of t which we must find. Now $\overrightarrow{PQ}$, which is equivalent to $\boldsymbol{q}-\boldsymbol{p}$ where

$$\boldsymbol{q}-\boldsymbol{p}=(6,3)+t(4,3)-(2,5)=(6+4t-2,3+3t-5)=(4+4t,-2+3t)\,,$$

must be orthogonal to the direction of the line, $\boldsymbol{v}=(4,3)$. This gives the necessary constraint to find t:

$$\begin{aligned}(\boldsymbol{q}-\boldsymbol{p})\cdot\boldsymbol{v}&=0\\(4+4t,-2+3t)\cdot(4,3)&=0\\(4+4t)(4)+(-2+3t)(3)&=0\\16+16t-6+9t&=0\\25t&=-10\\t&=-\frac{10}{25}=-\frac{2}{5}\end{aligned}$$

This implies

$$\begin{aligned}\boldsymbol{q}&=(6,3)+\left(-\frac{2}{5}\right)(4,3)\\&=(6,3)-\left(\frac{8}{5},\frac{6}{5}\right)\\&=\left(6-\frac{8}{5},3-\frac{6}{5}\right)\\&=\left(\frac{30-8}{5},\frac{15-6}{5}\right)\\&=\left(\frac{22}{5},\frac{9}{5}\right)\,,\end{aligned}$$

so the closest point on the line is $Q=(22/5,9/5)$. The distance from the line to P is therefore the length of

$$\boldsymbol{q}-\boldsymbol{p}=\left(\frac{22}{5},\frac{9}{5}\right)-(2,5)=\left(\frac{22}{5}-\frac{10}{5},\frac{9}{5}-\frac{25}{5}\right)=\left(\frac{12}{5},-\frac{16}{5}\right)=\frac{1}{5}(12,-16)$$

which is

$$d=\|\boldsymbol{q}-\boldsymbol{p}\|=\left\|\frac{1}{5}(12,-16)\right\|=\frac{1}{5}\,\|(12,-16)\|=\frac{1}{5}\sqrt{(12)^2+(-16)^2}=\frac{1}{5}\sqrt{400}=\frac{20}{5}\,.$$

5.7.2 Distance Given Normal

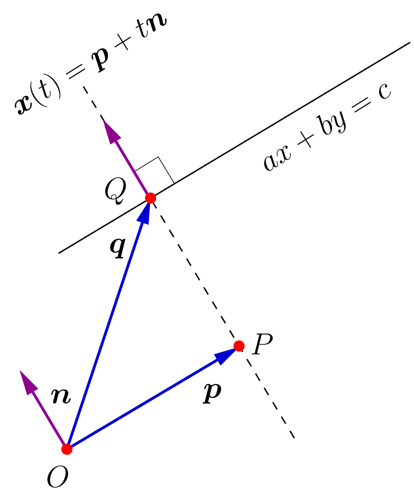
If an equation of a line in $\mathbb{R}^2$ or a plane in $\mathbb{R}^3$ is given in the standard form

$$ax+by=c\quad\text{or}\quad ax+by+c=d$$

respectively, then the normal direction to the line $\boldsymbol{n}=(a,b)$ or to the plane $\boldsymbol{n}=(a,b,c)$ is known. A point-normal equation similarly gives the normal $\boldsymbol{n}$. If we want the point Q that lies on the line or plane that is closest to the point P off of it, then Q must sit on the line through P that in the direction of the normal, namely the line $\boldsymbol{x}(t) = \boldsymbol{p} + t\boldsymbol{n}$. So for some particular value of t we have

$$\boxed{\boldsymbol{q}(t) = \boldsymbol{p} + t\boldsymbol{n}}\,.$$

Setting $\boldsymbol{q} = (x, y, z)$ one has parametric equations for $x(t)$, $y(t)$, and $z(t)$ which can then be inserted into the original line or plane equation to find t, since Q is constrained to lie there.



Example 5-13

Find the distance between the point $P(5, 1, 15)$ and the plane $2x - 3y + 6z = -1$. Also find the closest point Q in the plane.

Solution:
The plane equation gives the normal to the plane to be

$$\boldsymbol{n} = (a, b, c) = (2, -3, 6)\,.$$

The point Q, tip of the vector $\boldsymbol{q} = (x, y, z)$, lies on the line through P with direction $\boldsymbol{n}$ and so

$$\begin{aligned} \boldsymbol{q} &= \boldsymbol{p} + t\boldsymbol{n} \\ \begin{bmatrix} x \\ y \\ z \end{bmatrix} &= \begin{bmatrix} 5 \\ 1 \\ 15 \end{bmatrix} + t \begin{bmatrix} 2 \\ -3 \\ 6 \end{bmatrix} \end{aligned}$$

for some value of the parameter t to be determined. The parametric equations are

$$\begin{cases} x = 5 + 2t \\ y = 1 - 3t \\ z = 15 + 6t \end{cases}\,.$$

The constraint to find t comes from the fact that $\boldsymbol{q}$ must lie on the plane and hence satisfy the plane equation. Inserting the parametric equations into the latter gives:

$$\begin{aligned} 2x - 3y + 6z &= -1 \\ 2(5 + 2t) - 3(1 - 3t) + 6(15 + 6t) &= -1 \\ 10 + 4t - 3 + 9t + 90 + 36t &= -1 \\ 49t &= -98 \\ t &= -2 \end{aligned}$$

The closest point in the plane is therefore given by

$$\boldsymbol{q} = \begin{bmatrix} 5 \\ 1 \\ 15 \end{bmatrix} + (-2) \begin{bmatrix} 2 \\ -3 \\ 6 \end{bmatrix} = \begin{bmatrix} 5-4 \\ 1+6 \\ 15-12 \end{bmatrix} = \begin{bmatrix} 1 \\ 7 \\ 3 \end{bmatrix},$$

so $Q(1, 7, 3)$, and the distance between P and the plane is the length of

$$\boldsymbol{q} - \boldsymbol{p} = \begin{bmatrix} 1 \\ 7 \\ 3 \end{bmatrix} - \begin{bmatrix} 5 \\ 1 \\ 15 \end{bmatrix} = \begin{bmatrix} -4 \\ 6 \\ -12 \end{bmatrix}$$

which is

$$d = \|\boldsymbol{q} - \boldsymbol{p}\| = \|(-4, 6, -12)\| = \sqrt{(-4)^2 + 6^2 + (-12)^2} = \sqrt{196} = 14\,.$$

The approach can be used to solve, once and for all for the distance between P and the line or plane and one has the following theorems.

Theorem 5-1: In $\mathbb{R}^2$ the distance d between the point $P(x_0, y_0)$ and the line $ax + by = c$ is given by

$$d = \frac{|ax_0 + by_0 - c|}{\sqrt{a^2 + b^2}}\,.$$

Theorem 5-2: In $\mathbb{R}^3$ the distance d between the point $P(x_0, y_0, z_0)$ and the plane $ax + by + cz = d$ is given by

$$d = \frac{|ax_0 + by_0 + cz_0 - d|}{\sqrt{a^2 + b^2 + c^2}}\,.$$

Example 5-14

Find the distance between the point $P(5, 1, 15)$ and the plane $2x - 3y + 6z = -1$ using the distance formula.

Solution:

$$\begin{aligned} d &= \frac{|ax_0 + by_0 + cz_0 - d|}{\sqrt{a^2 + b^2 + c^2}} \\ &= \frac{|2(5) - 3(1) + 6(15) - (-1)|}{\sqrt{(2)^2 + (-3)^2 + 6^2}} \\ &= \frac{|98|}{\sqrt{49}} \\ &= \frac{98}{7} \\ &= 14 \end{aligned}$$

This is the same as we found in Example 5-13.

In three dimensions two planes will intersect (distance between them $d = 0$) unless they are parallel. If they are parallel we can find the distance between them by finding a point on one of the planes and then calculating the distance from that point to the other plane.

Example 5-15

Find the distance between the two planes $x + 2y - 2z = 3$ and $2x + 4y - 4z = 7$.

Solution:
The normal of $x + 2y - 2z = 3$ is $\boldsymbol{n_1} = (1, 2, -2)$.
The normal of $2x + 4y - 4z = 7$ is $\boldsymbol{n_2} = (2, 4, -4) = 2(1, 2, -2) = 2\boldsymbol{n_1}$.
So $\boldsymbol{n_1}$ is parallel to $\boldsymbol{n_2}$ and therefore the planes are parallel. To find a point in the first plane set $y = z = 0$ to get $x = 3$ so $P(3, 0, 0)$ lies on the first plane. Now find the distance between $P(3, 0, 0)$ and the plane $2x + 4y - 4z = 7$ using the distance formula:

$$\begin{aligned} d &= \frac{|ax_0 + by_0 + cz_0 - d|}{\sqrt{a^2 + b^2 + c^2}} \\ &= \frac{|2(3) + 0 + 0 - 7|}{\sqrt{2^2 + 4^2 + (-4)^2}} \\ &= \frac{|-1|}{\sqrt{36}} \\ &= \frac{1}{6} \end{aligned}$$

Chapter 6: Linear Transformations

6.1 A Survey of Linear Transformations

The student will have encountered the idea of real-valued functions of real variables, such as $f(x) = \sin(x)$. Similarly one can define real-valued functions of vectors in $\mathbb{R}^n$. The length of a vector,

$$f(\boldsymbol{u}) = \|\boldsymbol{u}\| ,$$

would be such a function. In this chapter we will go one step further and look at functions of vectors whose result is a vector, that is mappings from $\mathbb{R}^n \to \mathbb{R}^m$. Such a function can be thought of transforming a vector into a new vector.

Definition: A **transformation** from $\mathbb{R}^n$ to $\mathbb{R}^m$, written $T : \mathbb{R}^n \to \mathbb{R}^m$, is a mapping that assigns a unique vector $\boldsymbol{v} = T(\boldsymbol{u})$ in $\mathbb{R}^m$ to each vector $\boldsymbol{u}$ in $\mathbb{R}^n$.

The important class of transformations under consideration here will be linear transformations. We will see that these are ultimately representable in terms of matrices and their action on vectors with matrix multiplication.

Definition: A **linear transformation** is a transformation $L(\boldsymbol{u})$ from $\mathbb{R}^n$ to $\mathbb{R}^m$ satisfying

$$L(\boldsymbol{u}+\boldsymbol{v}) = L(\boldsymbol{u}) + L(\boldsymbol{v}) \tag{1}$$
$$L(c\boldsymbol{u}) = cL(\boldsymbol{u}) \tag{2}$$

for any vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ in $\mathbb{R}^n$ and scalar c.

Example 6-1

The **identity transformation**, which will be denoted by $\mathbb{1}$, takes a vector $\boldsymbol{u}$ in $\mathbb{R}^n$ to itself and is thus defined by $\boxed{\mathbb{1}(\boldsymbol{u}) = \boldsymbol{u}}$. It is a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^n$ since

$$\begin{aligned}\mathbb{1}(\boldsymbol{u}+\boldsymbol{v}) &= \boldsymbol{u}+\boldsymbol{v} = \mathbb{1}(\boldsymbol{u}) + \mathbb{1}(\boldsymbol{v})\\ \mathbb{1}(c\boldsymbol{u}) &= c\boldsymbol{u} = c\mathbb{1}(\boldsymbol{u}) .\end{aligned}$$

To more specifically identify the identity transformation from $\mathbb{R}^n$ to $\mathbb{R}^n$ we may write $\mathbb{1}_n$.

Example 6-2

The **zero transformation**, which will be denoted by $\mathbb{O}$, takes any vector $\boldsymbol{u}$ in $\mathbb{R}^n$ to the zero vector in $\mathbb{R}^m$, and is thus defined by $\boxed{\mathbb{O}(\boldsymbol{u}) = \mathbf{0}}$. It is a linear transformation since

$$\begin{aligned}\mathbb{O}(\boldsymbol{u}+\boldsymbol{v}) &= \mathbf{0} = \mathbf{0}+\mathbf{0} = \mathbb{O}(\boldsymbol{u}) + \mathbb{O}(\boldsymbol{v})\\ \mathbb{O}(c\boldsymbol{u}) &= \mathbf{0} = c\mathbf{0} = c\,\mathbb{O}(\boldsymbol{u}) .\end{aligned}$$

To be more specific we may write $\mathbb{O}_{mn}$ to identify the zero transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$.

A class of non-trivial linear transformations is as follows.

Example 6-3

Suppose k is a scalar constant. Define the **scalar transformation** C_k from $\mathbb{R}^n$ to $\mathbb{R}^n$ by

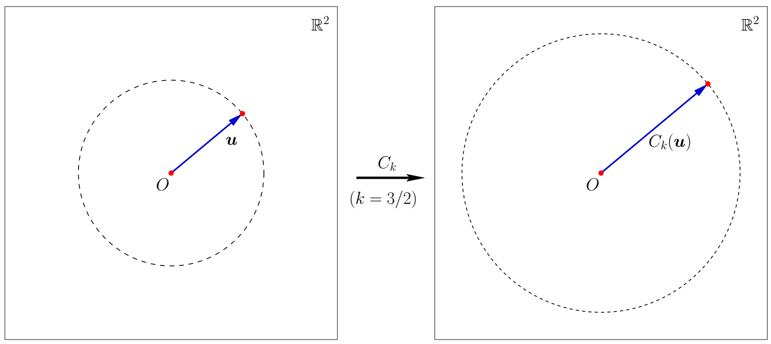
$$\boxed{C_k(\boldsymbol{u}) = k\boldsymbol{u}}.$$

Then C_k is a linear transformation since

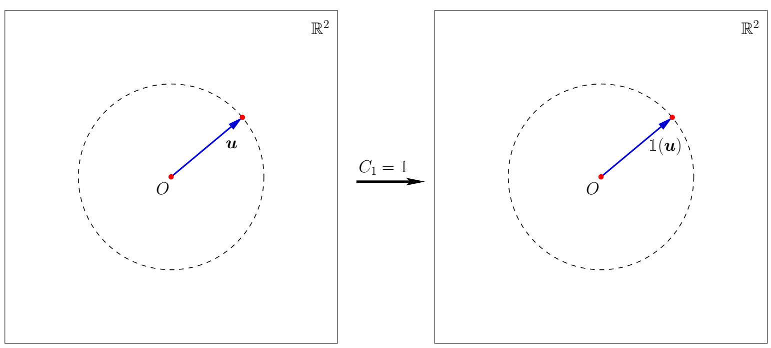
$$\begin{aligned} C_k(\boldsymbol{u}+\boldsymbol{v}) &= k(\boldsymbol{u}+\boldsymbol{v}) = k\boldsymbol{u}+k\boldsymbol{v} = C_k(\boldsymbol{u})+C_k(\boldsymbol{v}) \\ C_k(c\boldsymbol{u}) &= k(c\boldsymbol{u}) = c(k\boldsymbol{u}) = cC_k(\boldsymbol{u}) \end{aligned}$$

For real-valued k geometrically the effect of the transformation is as follows:

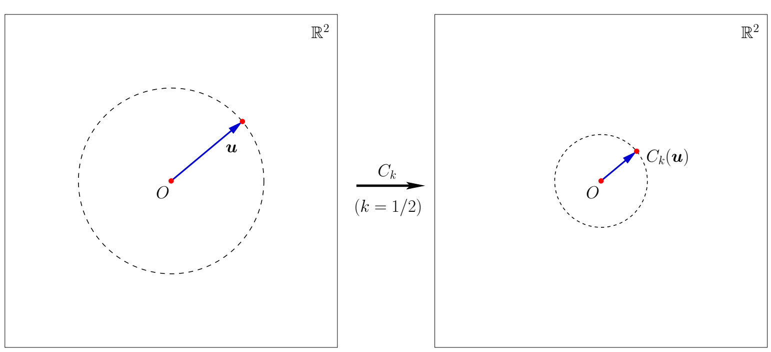
Case $1 < k$: $C_k(\boldsymbol{u})$ is a **dilation**. The vector retains its direction but is lengthened by a factor of k.



Case $k = 1$: $C_1(\boldsymbol{u}) = 1\boldsymbol{u} = \boldsymbol{u}$ and we recover the identity transformation $\mathbb{1}_n$.

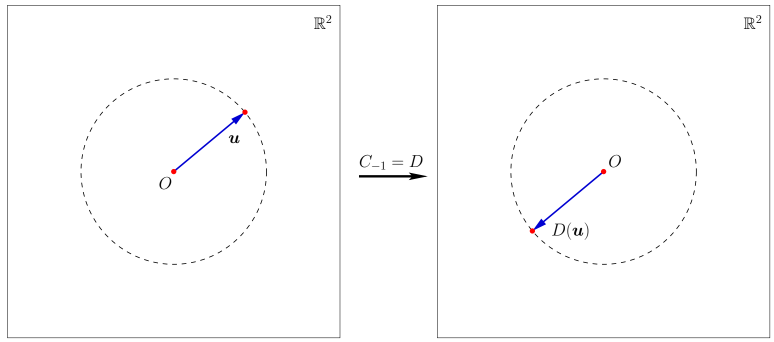


Case $0 < k < 1$: $C_k(\boldsymbol{u})$ is a **contraction**. The vector retains its direction but is shortened by a factor of k.



Case $\boldsymbol{k = 0}$: $C_0(\boldsymbol{u}) = 0\boldsymbol{u} = \boldsymbol{0}$ and we recover the zero transformation $\mathbb{O}_{nn}$.

Case $\boldsymbol{k = -1}$: $C_{-1}(\boldsymbol{u}) = -\boldsymbol{u}$ is **inversion through the origin $\boldsymbol{O}$**. The vector retains its length but points in the opposite direction. We will denote inversion by D so $\boxed{D(\boldsymbol{u}) = -\boldsymbol{u}}$.



Inversion through the original is sometimes called **reflection through the origin** though we will avoid that terminology.

Case $\boldsymbol{-1 < k < 0}$: A combination of inversion through the origin and contraction by $|k|$.

Case $\boldsymbol{k < -1}$: A combination of inversion through the origin and dilation by $|k|$.

One observes that the previous linear transformations written in terms of scalar multiplication are independent of choice of coordinate system (except for the choice of origin O upon which they do depend). Secondly one often considers the transformation as acting on the terminal points of the vectors themselves.

Example 6-4

Let $\boldsymbol{n}$ be a unit vector in $\mathbb{R}^n$. The **projection** onto $\boldsymbol{n}$, denoted by $P_{\boldsymbol{n}}$, and given by

$$\boxed{P_{\boldsymbol{n}}(\boldsymbol{u}) = \operatorname{proj}_{\boldsymbol{n}}\boldsymbol{u} = \frac{\boldsymbol{u}\cdot\boldsymbol{n}}{\|\boldsymbol{n}\|^2}\boldsymbol{n} = (\boldsymbol{u}\cdot\boldsymbol{n})\boldsymbol{n}}$$

is a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^n$. It is linear due to the linearity of the dot product:

$$\begin{aligned}
P_{\boldsymbol{n}}(\boldsymbol{u}+\boldsymbol{v}) &= \operatorname{proj}_{\boldsymbol{n}}(\boldsymbol{u}+\boldsymbol{v}) = [(\boldsymbol{u}+\boldsymbol{v})\cdot\boldsymbol{n}]\boldsymbol{n} = (\boldsymbol{u}\cdot\boldsymbol{n}+\boldsymbol{v}\cdot\boldsymbol{n})\boldsymbol{n}\\
&= (\boldsymbol{u}\cdot\boldsymbol{n})\boldsymbol{n} + (\boldsymbol{v}\cdot\boldsymbol{n})\boldsymbol{n} = \operatorname{proj}_{\boldsymbol{n}}\boldsymbol{u} + \operatorname{proj}_{\boldsymbol{n}}\boldsymbol{v}\\
&= P_{\boldsymbol{n}}(\boldsymbol{u}) + P_{\boldsymbol{n}}(\boldsymbol{v})\\
P_{\boldsymbol{n}}(c\boldsymbol{u}) &= \operatorname{proj}_{\boldsymbol{n}}(c\boldsymbol{u}) = [(c\boldsymbol{u})\cdot\boldsymbol{n}]\boldsymbol{n} = c(\boldsymbol{u}\cdot\boldsymbol{n})\boldsymbol{n} = c\operatorname{proj}_{\boldsymbol{n}}\boldsymbol{u}\\
&= cP_{\boldsymbol{n}}(\boldsymbol{u})
\end{aligned}$$

6.1.1 Sum and Scalar Product of Transformations

Just as we define the sum of two real-valued functions $f+g$ to be the result when we add the action of each function separately, i.e. $(f+g)(x) = f(x)+g(x)$ we can define the sum of two transformations $S(\boldsymbol{u})$ and $T(\boldsymbol{u})$ by

$$(S+T)(\boldsymbol{u}) = S(\boldsymbol{u}) + T(\boldsymbol{u})\,.$$

Similarly a scalar multiple c times a transformation T can be defined by multiplying the scalar times the result of acting T on the vector,

$$(cT)(\boldsymbol{u}) = cT(\boldsymbol{u})\,.$$

Theorem 6-1: Let K and L be linear transformations from $\mathbb{R}^n$ to $\mathbb{R}^m$ and c a scalar then the transformations $K + L$ and cL are themselves linear.

More generally one can prove inductively that an arbitrary linear combination of a finite number of such linear transformations will be linear.

Example 6-5

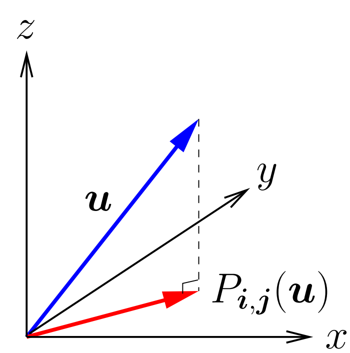
Suppose $S = \{\boldsymbol{n}_1,\ \boldsymbol{n}_2,\ \ldots,\ \boldsymbol{n}_k\}$ is a set of unit vectors in $\mathbb{R}^n$ that are mutually orthogonal. The **orthogonal projection** onto the span of these vectors will be denoted by $P_{\boldsymbol{n}_1,\boldsymbol{n}_2,\ldots,\boldsymbol{n}_k}$ and is given by

$$\boxed{P_{\boldsymbol{n}_1,\boldsymbol{n}_2,\ldots,\boldsymbol{n}_k}(\boldsymbol{u}) = \mathrm{proj}_{\boldsymbol{n}_1}\boldsymbol{u} + \mathrm{proj}_{\boldsymbol{n}_2}\boldsymbol{u} + \cdots + \mathrm{proj}_{\boldsymbol{n}_k}\boldsymbol{u}}\,.$$

It is a linear transformation since it is a sum of linear transformations,

$$P_{\boldsymbol{n}_1,\boldsymbol{n}_2,\ldots,\boldsymbol{n}_k} = P_{\boldsymbol{n}_1} + P_{\boldsymbol{n}_2} + \cdots + P_{\boldsymbol{n}_k}\,.$$

As a specific example, $P_{\boldsymbol{i},\boldsymbol{j}}$ in $\mathbb{R}^3$ is the projection of $\boldsymbol{u}$ onto the x-y plane.



Example 6-6

We have seen how a vector $\boldsymbol{u}$ can be broken into a vector component parallel to a given direction $\boldsymbol{n}$ and a component orthogonal to it by

$$\boldsymbol{u} = \boldsymbol{u}_1 + \boldsymbol{u}_2 = \mathrm{proj}_{\boldsymbol{n}}\boldsymbol{u} + (\boldsymbol{u} - \mathrm{proj}_{\boldsymbol{n}}\boldsymbol{u})$$

If we transform the vector by multiplying only the parallel component $\boldsymbol{u}_1$ by a constant scalar k to get $k\boldsymbol{u}_1 + \boldsymbol{u}_2$ we have the transformation from $\mathbb{R}^n$ to $\mathbb{R}^n$ given by

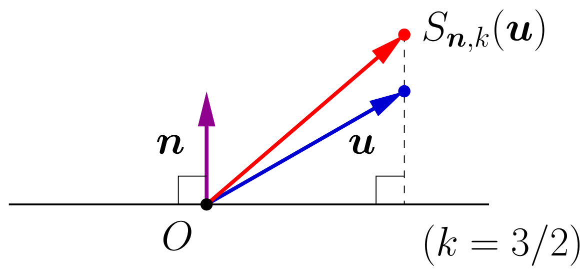
$$\boxed{S_{\boldsymbol{n},k}(\boldsymbol{u}) = k\,\mathrm{proj}_{\boldsymbol{n}}\boldsymbol{u} + (\boldsymbol{u} - \mathrm{proj}_{\boldsymbol{n}}\boldsymbol{u}) = \boldsymbol{u} + (k-1)\,\mathrm{proj}_{\boldsymbol{n}}\boldsymbol{u}}$$

The transformation is linear since it can be written as a linear combination of linear transformations:

$$S_{\boldsymbol{n},k} = \mathbb{1} + (k-1)P_{\boldsymbol{n}}\,.$$

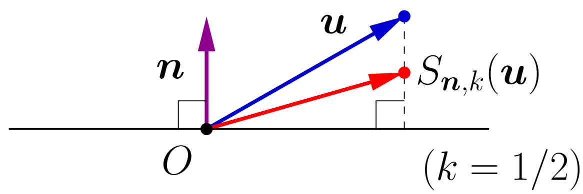
Important special cases are as follows.

Case $\boldsymbol{k > 1}$: The component along $\boldsymbol{n}$ expands by a factor of k and we call $S_{\boldsymbol{n},k}$ an **expansion**.



Case $\boldsymbol{k = 1}$: We see $S_{\boldsymbol{n},1} = \mathbb{1}$, the identity transformation.

Case $\boldsymbol{0 < k < 1}$: The component along $\boldsymbol{n}$ is compressed by the factor k and we call $S_{\boldsymbol{n},k}$ a **compression**.



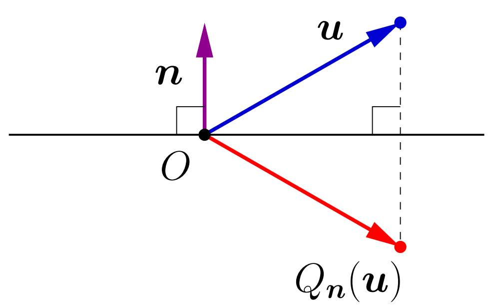
If we imagine an image made up of points considered to be terminal points of their respective vectors from the origin then for k positive $S_{\boldsymbol{n},k}$ will stretch ($k > 1$) or squash ($0 < k < 1$) the entire image along the direction of $\boldsymbol{n}$ measured from the line (in $\mathbb{R}^2$) or plane (in $\mathbb{R}^3$) going through the origin orthogonal to $\boldsymbol{n}$.

Case $\boldsymbol{k = 0}$: We see $S_{\boldsymbol{n},0} = \mathbb{1} - P_{\boldsymbol{n}}$ which returns the component of $\boldsymbol{u}$ orthogonal to $\boldsymbol{n}$, i.e. $\boldsymbol{u}_2$.

Case $\boldsymbol{k = -1}$: The linear transformation $S_{\boldsymbol{n},-1}$ which will be denoted $Q_{\boldsymbol{n}}$, is given by

$$\boxed{Q_{\boldsymbol{n}}(\boldsymbol{u}) = -\mathrm{proj}_{\boldsymbol{n}}\boldsymbol{u} + (\boldsymbol{u} - \mathrm{proj}_{\boldsymbol{n}}\boldsymbol{u}) = \boldsymbol{u} - 2\,\mathrm{proj}_{\boldsymbol{n}}\boldsymbol{u}}$$

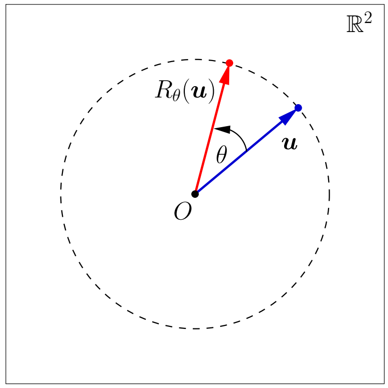
and is called a **reflection**. In $\mathbb{R}^2$ it represents a **reflection about the line** through the **origin** with normal $\boldsymbol{n}$. In $\mathbb{R}^3$ it is a **reflection across the plane** through the **origin** with normal $\boldsymbol{n}$. It is called a reflection because a set of points (tips of their respective vectors) will be transformed to their mirror images across the line or plane.



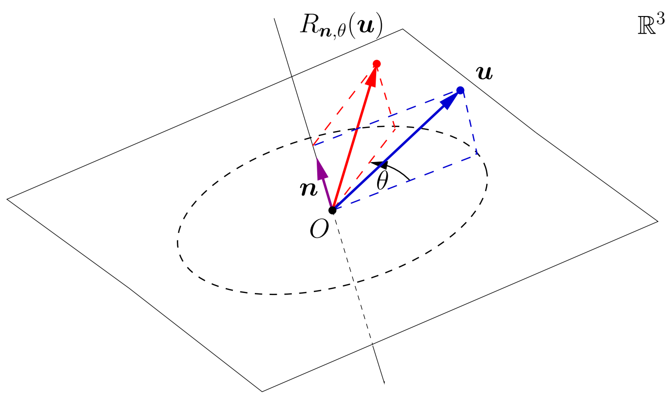
Note that expansion/compression should be compared with dilation/contraction seen earlier and reflection should be compared with inversion about the origin. In dilation/contraction and inversion the entire vector is modified, not simply a projection of it.

Example 6-7

As an important example we consider rotation in the plane $\mathbb{R}^2$ and more generally in $\mathbb{R}^3$. In the plane a counterclockwise rotation is determined by an angle θ. Such a transformation will be written $R_\theta(\boldsymbol{u})$.



In three dimensions a rotation is determined by a direction, given by unit vector $\boldsymbol{n}$, which determines the **axis of rotation** through the origin, as well as the angle of rotation, θ about that axis. The right-hand rule with thumb pointing in the direction of $\boldsymbol{n}$ is used to determine the direction of rotation. The projection of $\boldsymbol{u}$ in the direction of $\boldsymbol{n}$ is unaffected by the rotation. The orthogonal component which lies in the plane through the origin perpendicular to $\boldsymbol{n}$ is rotated.



If we denote the three-dimensional rotation by $R_{\boldsymbol{n},\theta}$ then it can be given explicitly in $\mathbb{R}^3$ by **Rodrigues' rotation formula**:

$$\boxed{R_{\boldsymbol{n},\theta}(\boldsymbol{u}) = (\cos\theta)\boldsymbol{u} + (\sin\theta)(\boldsymbol{n}\times\boldsymbol{u}) + (1-\cos\theta)(\boldsymbol{n}\cdot\boldsymbol{u})\boldsymbol{n}}\,.$$

Recalling the linear properties of the cross and dot product (Theorems 4-14 and 4-6 respectively):

$$\begin{aligned} \boldsymbol{n}\times(\boldsymbol{u}+\boldsymbol{v}) &= \boldsymbol{n}\times\boldsymbol{u}+\boldsymbol{n}\times\boldsymbol{v} \qquad & \boldsymbol{n}\cdot(\boldsymbol{u}+\boldsymbol{v}) &= \boldsymbol{n}\cdot\boldsymbol{u}+\boldsymbol{n}\cdot\boldsymbol{v} \\ \boldsymbol{n}\times(c\boldsymbol{u}) &= c(\boldsymbol{n}\times\boldsymbol{u}) & \boldsymbol{n}\cdot(c\boldsymbol{u}) &= c(\boldsymbol{n}\cdot\boldsymbol{u}) \end{aligned}$$

it follows that $R_{\boldsymbol{n},\theta}$ is a linear transformation. Rotations in higher dimensions can similarly be determined by using a normal direction $\boldsymbol{n}$ and a rotation in its orthogonal plane. The absence of the cross product for general $\mathbb{R}^n$ does not hinder this.

Example 6-8

Given the vectors $\boldsymbol{u} = (2, 0, 3)$ and $\boldsymbol{v} = (1, 3, 5)$ in $\mathbb{R}^3$ apply the following linear transformations. Describe each transformation.

1. $\mathbb{1}(\boldsymbol{u})$
2. $\mathbb{O}_{3,3}(\boldsymbol{u})$
3. $C_2(\boldsymbol{u})$
4. $C_{\frac{1}{2}}(\boldsymbol{u})$
5. $D(\boldsymbol{u})$
6. $P_{\boldsymbol{i}}(\boldsymbol{u})$
7. $P_{\boldsymbol{i},\boldsymbol{k}}(\boldsymbol{v})$
8. $S_{\boldsymbol{k},2}(\boldsymbol{u})$
9. $S_{\boldsymbol{i},\frac{1}{2}}(\boldsymbol{u})$
10. $Q_{\boldsymbol{i}}(\boldsymbol{u})$
11. $R_{\boldsymbol{k},\frac{\pi}{2}}(\boldsymbol{u})$

Solution:

1. $\mathbb{1}(\boldsymbol{u}) = \boldsymbol{u} = (2, 0, 3)$, the identity transformation.
2. $\mathbb{O}_{3,3}(\boldsymbol{u}) = \boldsymbol{0} = (0, 0, 0)$, the zero transformation.
3. $C_2(\boldsymbol{u}) = 2\boldsymbol{u} = (4, 0, 6)$, dilation by a factor of 2.
4. $C_{\frac{1}{2}}(\boldsymbol{u}) = \frac{1}{2}\boldsymbol{u} = \left(1, 0, \frac{3}{2}\right)$, contraction by a factor $1/2$.
5. $D(\boldsymbol{u}) = -\boldsymbol{u} = (-2, 0, -3)$, inversion about the origin O.
6. $P_{\boldsymbol{i}}(\boldsymbol{u}) = (\boldsymbol{u} \cdot \boldsymbol{i})\boldsymbol{i} = (2, 0, 3) \cdot (1, 0, 0)\boldsymbol{i} = (2 + 0 + 0)\boldsymbol{i} = 2\boldsymbol{i} = (2, 0, 0)$, projection along the direction of $\boldsymbol{i}$, the x-axis.
7. $P_{\boldsymbol{i},\boldsymbol{k}}(\boldsymbol{v}) = P_{\boldsymbol{i}}(\boldsymbol{v}) + P_{\boldsymbol{k}}(\boldsymbol{v}) = [(1, 3, 5) \cdot \boldsymbol{i}]\boldsymbol{i} + [(1, 3, 5) \cdot \boldsymbol{k}]\boldsymbol{k} = 1\boldsymbol{i} + 5\boldsymbol{k} = (1, 0, 5)$, projection onto the x-z plane.
8. $S_{\boldsymbol{k},2}(\boldsymbol{u}) = \boldsymbol{u} + (2 - 1)\text{proj}_{\boldsymbol{k}}\boldsymbol{u} = (2, 0, 3) + 1[(2, 0, 3) \cdot \boldsymbol{k}]\boldsymbol{k} = (2, 0, 3) + 3\boldsymbol{k} = (2, 0, 6)$, expansion in z direction by a factor of 3.
9. $S_{\boldsymbol{i},\frac{1}{2}}(\boldsymbol{u}) = \boldsymbol{u} + (1/2 - 1)\text{proj}_{\boldsymbol{i}}\boldsymbol{u} = (2, 0, 3) - \frac{1}{2}\underbrace{[(2, 0, 3) \cdot \boldsymbol{i}]}_{=2}\boldsymbol{i} = (2, 0, 3) - \boldsymbol{i} = (1, 0, 3)$, compression in x direction by a factor of $1/2$.
10. $Q_{\boldsymbol{i}}(\boldsymbol{u}) = \boldsymbol{u} - 2\text{proj}_{\boldsymbol{i}}\boldsymbol{u} = (2, 0, 3) - 2[(2, 0, 3) \cdot \boldsymbol{i}]\boldsymbol{i} = (2, 0, 3) - 4\boldsymbol{i} = (-2, 0, 3)$, reflection across plane with normal $\boldsymbol{i}$ (the y-z plane).
11. Noting that $\boldsymbol{k} \cdot \boldsymbol{u} = \boldsymbol{k} \cdot (2, 0, 3) = 3$ and
$$\boldsymbol{k} \times \boldsymbol{u} = \begin{vmatrix} \boldsymbol{i} & \boldsymbol{j} & \boldsymbol{k} \\ 0 & 0 & 1 \\ 2 & 0 & 3 \end{vmatrix} = \boldsymbol{i}(0 - 0) - \boldsymbol{j}(0 - 2) + \boldsymbol{k}(0 - 0) = 2\boldsymbol{j}$$
we have
$$\begin{aligned} R_{\boldsymbol{k},\frac{\pi}{2}}(\boldsymbol{u}) &= \left(\cos\frac{\pi}{2}\right)\boldsymbol{u} + \left(\sin\frac{\pi}{2}\right)\boldsymbol{k} \times \boldsymbol{u} + \left(1 - \cos\frac{\pi}{2}\right)(\boldsymbol{k} \cdot \boldsymbol{u})\boldsymbol{k} \\ &= (0)\boldsymbol{u} + (1)(2\boldsymbol{j}) + (1 - 0)(3)\boldsymbol{k} \\ &= 2\boldsymbol{j} + 3\boldsymbol{k} = (0, 2, 3), \end{aligned}$$
a rotation of $90°$ about the positive z-axis.

The student is encouraged to plot and label all the points (vector tips) that lie in the x-z plane (i.e. with y-component equal to zero) by using x as the horizontal axis and z as the vertical axis, to help visualize the effect of the transformations.

In Example 6-8 we transformed a single vector, or equivalently a single terminal point. In general we are often interested in transforming a set of points which constitute an image or object in two or three dimensions. It is worth considering the effect of such transformations on multiple points whereby their meaning and application becomes clearer.

Rodrigues' rotation formula can be used to determine R_θ in $\mathbb{R}^2$ by associating the vector (x, y) in $\mathbb{R}^2$ with the vector $(x, y, 0)$ in the x-y plane of $\mathbb{R}^3$ and setting $\boldsymbol{n} = \boldsymbol{k}$.

Example 6-9

Let $\boldsymbol{x} = (x, y)$ in $\mathbb{R}^2$. Derive the formula for the linear transformation $R_\theta(\boldsymbol{x})$ by considering the effect of rotating the vector $\boldsymbol{u} = (x, y, 0)$ in the x-y plane of $\mathbb{R}^3$ by the angle θ about the z-axis, i.e. about the direction $\boldsymbol{n} = \boldsymbol{k}$. Show, by inspection, that the result can be written as the left-multiplication of $\boldsymbol{x} = \left[\begin{smallmatrix} x \\ y \end{smallmatrix}\right]$ by an appropriate 2×2 matrix.

Solution:
Since

$$\boldsymbol{k} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \quad \text{and} \quad \boldsymbol{u} = \begin{bmatrix} x \\ y \\ 0 \end{bmatrix} = x\boldsymbol{i} + y\boldsymbol{j}\,.$$

this implies

$$\boldsymbol{k} \cdot \boldsymbol{u} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \\ 0 \end{bmatrix} = 0 + 0 + 0 = 0, \text{ and}$$

$$\boldsymbol{k} \times \boldsymbol{u} = \begin{vmatrix} \boldsymbol{i} & 0 & x \\ \boldsymbol{j} & 0 & y \\ \boldsymbol{k} & 1 & 0 \end{vmatrix} = \boldsymbol{i}(0 - y) - \boldsymbol{j}(0 - x) + \boldsymbol{k}(0 - 0) = -y\boldsymbol{i} + x\boldsymbol{j}\,.$$

The rotation formula gives

$$\begin{aligned} R_{\boldsymbol{k},\theta}(\boldsymbol{u}) &= (\cos\theta)\boldsymbol{u} + (\sin\theta)(\boldsymbol{k} \times \boldsymbol{u}) + (1 - \cos\theta)(\boldsymbol{k} \cdot \boldsymbol{u})\boldsymbol{k} \\ &= \cos\theta(x\boldsymbol{i} + y\boldsymbol{j}) + \sin\theta(-y\boldsymbol{i} + x\boldsymbol{j}) + 0 \\ &= ((\cos\theta)x - (\sin\theta)y)\,\boldsymbol{i} + ((\sin\theta)x + (\cos\theta)y)\,\boldsymbol{j} \end{aligned}$$

Next by identifying $\boldsymbol{x} = \left[\begin{smallmatrix} x \\ y \end{smallmatrix}\right]$, $\boldsymbol{i} = \left[\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right]$, and $\boldsymbol{j} = \left[\begin{smallmatrix} 0 \\ 1 \end{smallmatrix}\right]$ in $\mathbb{R}^2$, the two dimensional result is

$$\begin{aligned} R_\theta(\boldsymbol{x}) &= ((\cos\theta)x - (\sin\theta)y)\,\boldsymbol{i} + ((\sin\theta)x + (\cos\theta)y)\,\boldsymbol{j} \\ &= ((\cos\theta)x - (\sin\theta)y)\begin{bmatrix} 1 \\ 0 \end{bmatrix} + ((\sin\theta)x + (\cos\theta)y)\begin{bmatrix} 0 \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} (\cos\theta)x - (\sin\theta)y \\ (\sin\theta)x + (\cos\theta)y \end{bmatrix} \qquad (\leftarrow \text{a vector in } \mathbb{R}^2) \end{aligned}$$

The last expression can be rewritten using matrix multiplication as

$$R_\theta(\boldsymbol{x}) = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}\,.$$

Example 6-9 raises two interesting questions. We were able to represent a linear transformation by a matrix multiplying the vector. Does multiplying a vector by some other matrix produce a linear transformation? Secondly, can other linear transformations be similarly represented by multiplication by a matrix?

6.2 Matrix Transformations

Turning to our first question raised at the end of the last section we have the following definition.

Definition: Let A be an $m \times n$ matrix. A **matrix transformation** is the transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$ given by the matrix multiplication of vector $\boldsymbol{u}$ in $\mathbb{R}^n$ (written as a column matrix) on the left by A:

$$\boxed{L_A(\boldsymbol{u}) = A\boldsymbol{u}}\,.$$

That a matrix transformation is a linear transformation follows from the linearity of matrix multiplication.

Theorem 6-2: The matrix transformation $L_A(\boldsymbol{u})$ from $\mathbb{R}^n$ to $\mathbb{R}^m$ is a linear transformation.

Proof:
Let L_A be a matrix transformation with A the $m \times n$ matrix. Suppose $\boldsymbol{u}$ and $\boldsymbol{v}$ are vectors in $\mathbb{R}^n$ and c is a scalar, then we have the following:

$$\begin{aligned} L_A(\boldsymbol{u}+\boldsymbol{v}) &= A(\boldsymbol{u}+\boldsymbol{v}) = A\boldsymbol{u} + A\boldsymbol{v} = L_A(\boldsymbol{u}) + L_A(\boldsymbol{v}) \\ L_A(c\boldsymbol{u}) &= A(c\boldsymbol{u}) = c(A\boldsymbol{u}) = cL_A(\boldsymbol{u}) \end{aligned}$$

Example 6-10

Let $A = \begin{bmatrix} 1 & 3 \\ 0 & 1 \end{bmatrix}$ and let $\boldsymbol{u} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $\boldsymbol{v} = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$ be vectors. Find

1. $L_A(\boldsymbol{u})$
2. $L_A(\boldsymbol{v})$
3. $L_A(\boldsymbol{u}+\boldsymbol{v})$
4. $L_A(5\boldsymbol{u})$

Solution:

1. $L_A(\boldsymbol{u}) = A\boldsymbol{u} = \begin{bmatrix} 1 & 3 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1+3 \\ 0+1 \end{bmatrix} = \begin{bmatrix} 4 \\ 1 \end{bmatrix}$
2. $L_A(\boldsymbol{v}) = A\boldsymbol{v} = \begin{bmatrix} 1 & 3 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 2 \end{bmatrix} = \begin{bmatrix} 2+6 \\ 0+2 \end{bmatrix} = \begin{bmatrix} 8 \\ 2 \end{bmatrix}$
3. $L_A(\boldsymbol{u}+\boldsymbol{v}) = A(\boldsymbol{u}+\boldsymbol{v}) = \begin{bmatrix} 1 & 3 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1+2 \\ 1+2 \end{bmatrix} = \begin{bmatrix} 1 & 3 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 3 \\ 3 \end{bmatrix} = \begin{bmatrix} 3+9 \\ 0+3 \end{bmatrix} = \begin{bmatrix} 12 \\ 3 \end{bmatrix}$

   which equals $L_A(\boldsymbol{u}) + L_A(\boldsymbol{v})$ as expected.
4. $L_A(5\boldsymbol{u}) = A(5\boldsymbol{u}) = \begin{bmatrix} 1 & 3 \\ 0 & 1 \end{bmatrix} \left(5 \begin{bmatrix} 1 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} 1 & 3 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 5 \\ 5 \end{bmatrix} = \begin{bmatrix} 5+15 \\ 0+5 \end{bmatrix} = \begin{bmatrix} 20 \\ 5 \end{bmatrix}$

   which equals $5L_A(\boldsymbol{u})$ as expected.

Next let us consider the question of which linear transformations may be represented by matrix transformations. It turns out that all of them are. To see why observe, the following important property of linear transformations. First we need our definition of linear combination with matrices applied to vectors.

Definition: Let $\boldsymbol{v}_1$, $\boldsymbol{v}_2$, ..., $\boldsymbol{v}_k$ be vectors in $\mathbb{R}^n$ and c_1, c_2, ..., c_k be scalars. Then the vector

$$c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \cdots + c_k\boldsymbol{v}_k$$

is a **linear combination** of $\boldsymbol{v}_1$, $\boldsymbol{v}_2$, ..., $\boldsymbol{v}_k$.

Theorem 6-3: Let $L(\boldsymbol{u})$ be a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$. If $\boldsymbol{u}$ is a linear combination of vectors $\boldsymbol{v}_1$, $\boldsymbol{v}_2$,... $\boldsymbol{v}_k$ in $\mathbb{R}^n$, so that

$$\boldsymbol{u} = c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \cdots + c_k\boldsymbol{v}_k\,,$$

then

$$\boxed{L(\boldsymbol{u}) = c_1L(\boldsymbol{v}_1) + c_2L(\boldsymbol{v}_2) + \cdots + c_kL(\boldsymbol{v}_k)}\,.$$

The implication of the theorem is that if you can decompose a vector into a linear combination of other vectors, then knowledge of how the linear transformation acts on those vectors is sufficient to determine how it acts on the original vector.

Example 6-11

Suppose $\boldsymbol{u} = \begin{bmatrix}1\\0\\3\end{bmatrix}$, $\boldsymbol{v} = \begin{bmatrix}2\\2\\0\end{bmatrix}$ and $\boldsymbol{w} = \begin{bmatrix}0\\-2\\6\end{bmatrix}$ and let $L(\boldsymbol{u})$ be a linear transformation from $\mathbb{R}^3$ to $\mathbb{R}^2$. If $L(\boldsymbol{u}) = \begin{bmatrix}1\\1\end{bmatrix}$ and $L(\boldsymbol{v}) = \begin{bmatrix}-2\\3\end{bmatrix}$, find $L(\boldsymbol{w})$.

Solution:
Our strategy is to write $\boldsymbol{w}$ as a linear combination of $\boldsymbol{u}$ and $\boldsymbol{v}$ (if possible) and then apply the linearity of L to the result. If $\boldsymbol{w} = a\boldsymbol{u} + b\boldsymbol{v}$ then constants a and b must satisfy

$$a\boldsymbol{u} + b\boldsymbol{v} = a\begin{bmatrix}1\\0\\3\end{bmatrix} + b\begin{bmatrix}2\\2\\0\end{bmatrix} = \begin{bmatrix}a\\0\\3a\end{bmatrix} + \begin{bmatrix}2b\\2b\\0\end{bmatrix} = \begin{bmatrix}a+2b\\2b\\3a\end{bmatrix} = \boldsymbol{w} = \begin{bmatrix}0\\-2\\6\end{bmatrix}$$

Equality of the vectors implies solving the system of equations

$$\begin{aligned} a + 2b &= 0 \\ 2b &= -2 \\ 3a &= 6 \end{aligned}$$

The second equation implies $\boxed{b = -1}$ and the third implies $\boxed{a = 2}$. These values satisfy the first equation, $2 + 2(-1) = 0$, so this is a solution to the system and we have that $\boldsymbol{w} = 2\boldsymbol{u} - 1\boldsymbol{v}$, which is easily checked. The linearity of L implies

$$L(\boldsymbol{w}) = L(2\boldsymbol{u} - 1\boldsymbol{v}) = L(2\boldsymbol{u}) + L(-1\boldsymbol{v}) = 2L(\boldsymbol{u}) + (-1)L(\boldsymbol{v}) = 2\begin{bmatrix}1\\1\end{bmatrix} - 1\begin{bmatrix}-2\\3\end{bmatrix} = \begin{bmatrix}2+2\\2-3\end{bmatrix} = \begin{bmatrix}4\\-1\end{bmatrix}.$$

Recall that $\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n$ are the elementary vectors in $\mathbb{R}^n$ defined by $\boldsymbol{e}_1 = (1, 0, \ldots, 0)$, $\boldsymbol{e}_2 = (0, 1, \ldots, 0)$, $\ldots$, $\boldsymbol{e}_n = (0, 0, \ldots, 1)$. Then $\{\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n\}$ is called the **standard basis** of $\mathbb{R}^n$. With respect to this basis we saw that any vector $\boldsymbol{u}$ has the unique expansion[1]

$$\boldsymbol{u} = (u_1, u_2, \ldots, u_n) = u_1\boldsymbol{e}_1 + u_2\boldsymbol{e}_2 + \cdots u_n\boldsymbol{e}_n \,.$$

Now consider acting on $\boldsymbol{u}$ by a linear transformation L that takes vectors from $\mathbb{R}^n$ to $\mathbb{R}^m$.

$$\begin{aligned} L(\boldsymbol{u}) &= L(u_1\boldsymbol{e}_1 + u_2\boldsymbol{e}_2 + \cdots + u_n\boldsymbol{e}_n) \\ &= u_1L(\boldsymbol{e}_1) + u_2L(\boldsymbol{e}_2) + \cdots + u_nL(\boldsymbol{e}_n) \end{aligned}$$

Defining vectors $\boldsymbol{a}_i = L(\boldsymbol{e}_i)$ in $\mathbb{R}^m$ this implies

$$L(\boldsymbol{u}) = u_1\boldsymbol{a_1} + u_2\boldsymbol{a}_2 + \cdots + u_n\boldsymbol{a}_n \,.$$

However, when discussing matrix multiplication of a vector (Section 2.14.2) we noted that this right hand side equals $A\boldsymbol{u}$ if $A = [\boldsymbol{a}_1\boldsymbol{a}_2 \cdots \boldsymbol{a}_n]$ where $\boldsymbol{a}_i$ is the i^{th} column of A. Hence

$$L(\boldsymbol{u}) = A\boldsymbol{u}$$

where

$$A = [\boldsymbol{a}_1\boldsymbol{a}_2 \cdots \boldsymbol{a}_n] = [L(\boldsymbol{e}_1)L(\boldsymbol{e}_2) \cdots L(\boldsymbol{e}_n)]$$

is an $m \times n$ matrix whose i^{th} column is the vector $L(\boldsymbol{e}_i)$. We summarize the result in the following theorem.

Theorem 6-4: Let L be a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$ and let $\{\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n\}$ be the standard basis of $\mathbb{R}^n$. Then L equals the matrix transformation L_A where A is the unique $m \times n$ matrix

$$\boxed{A = [L(\boldsymbol{e}_1)L(\boldsymbol{e}_2) \cdots L(\boldsymbol{e}_n)]}\,.$$

Here $L(\boldsymbol{e}_i)$ is a vector in $\mathbb{R}^m$ written as a column matrix.

Proof:
The remaining item to show is that the matrix A is unique. We have seen that $L(\boldsymbol{e}_i) = \boldsymbol{a}_i$. Suppose transformation L equalled a second matrix transformation L_B. Then

$$L(\boldsymbol{e}_i) = L_B(\boldsymbol{e}_i) = B\boldsymbol{e}_i = \boldsymbol{b}_i \,.$$

where $\boldsymbol{b}_i$ is the i^{th} column of B. The final equality follows due to the components of $\boldsymbol{e}_i$. Hence $\boldsymbol{b}_i = \boldsymbol{a}_i$. Since the choice of i was arbitrary, $B = A$.

Since we have shown that every matrix transformation is linear and now that every linear transformation equals a matrix transformation we have the following result.

Corollary: A transformation T is linear if and only if it equals a matrix transformation.

Theorem 6-4 gives a prescription for finding a matrix to represent any linear transformation L in terms of the action of L on the elementary vectors $\boldsymbol{e}_i$. The latter can be found using the explicit forms of L of common transformations found earlier, or one can deduce them directly from the transformation the operator presents as the following examples show.

[1] Uniqueness follows for if $\boldsymbol{u}$ had some other expansion $\boldsymbol{u} = v_1\boldsymbol{e}_1 + v_2\boldsymbol{e}_2 + \cdots v_n\boldsymbol{e}_n$ we could take the dot product of each expansion with respect to $\boldsymbol{e}_i$ to get $\boldsymbol{u} \cdot \boldsymbol{e}_i = u_i$ for the first and $\boldsymbol{u} \cdot \boldsymbol{e}_i = v_i$ for the second, showing $u_i = v_i$. Since i was arbitrary the expansion is unique.

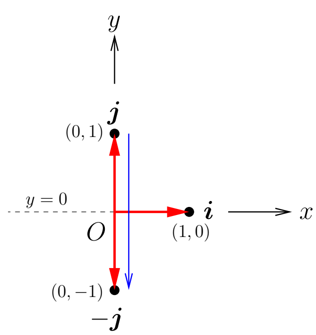
Example 6-12

Find a matrix transformation equal to the linear transformation.

1. The reflection $Q_{\boldsymbol{j}}$ in $\mathbb{R}^2$.
2. The dilation C_2 in $\mathbb{R}^2$.

Solution:

1. The reflection $Q_{\boldsymbol{j}}$ in $\mathbb{R}^2$ is a reflection across the line with normal $\boldsymbol{j}$. Since $\boldsymbol{j}$ points along the y-axis, the reflection is across the x-axis (i.e. the line $y = 0$). Such a reflection leaves the vector $\boldsymbol{i}$ unchanged and takes $\boldsymbol{j}$ to $-\boldsymbol{j}$. (This can be seen by thinking of the action of the transformation on the endpoints $(1, 0)$ and $(0, 1)$ of $\boldsymbol{i}$ and $\boldsymbol{j}$ respectively.) In symbols

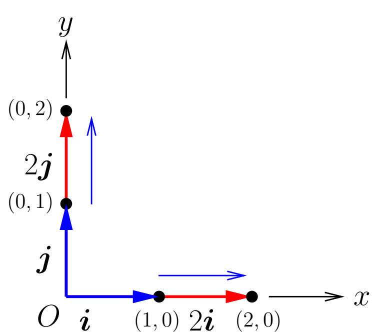


$$L(\boldsymbol{e}_1) = L(\boldsymbol{i}) = \boldsymbol{i} = \begin{bmatrix}1\\0\end{bmatrix}$$

$$L(\boldsymbol{e}_2) = L(\boldsymbol{j}) = -\boldsymbol{j} = -1\begin{bmatrix}0\\1\end{bmatrix} = \begin{bmatrix}0\\-1\end{bmatrix}$$

Thus $Q_{\boldsymbol{j}} = L_A$ where $A = [L(\boldsymbol{e}_1)L(\boldsymbol{e}_2)] = \begin{bmatrix}1 & 0\\0 & -1\end{bmatrix}$.

2. The dilation C_2 in $\mathbb{R}^2$ stretches all vectors by a factor of two, so, in particular the elementary vectors transforms as



$$L(\boldsymbol{e}_1) = L(\boldsymbol{i}) = 2\boldsymbol{i} = 2\begin{bmatrix}1\\0\end{bmatrix} = \begin{bmatrix}2\\0\end{bmatrix}$$

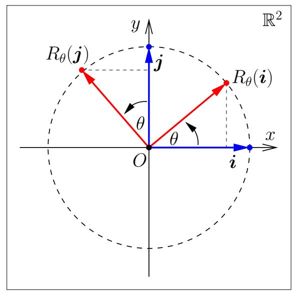
$$L(\boldsymbol{e}_2) = L(\boldsymbol{j}) = 2\boldsymbol{j} = 2\begin{bmatrix}0\\1\end{bmatrix} = \begin{bmatrix}0\\2\end{bmatrix}$$

Thus $C_2 = L_A$ where $A = [L(\boldsymbol{e}_1)L(\boldsymbol{e}_2)] = \begin{bmatrix}2 & 0\\0 & 2\end{bmatrix}$.

Example 6-13

Rederive the matrix transformation for R_θ in $\mathbb{R}^2$ using Theorem 6-4. Also find the transformation for $R_{\boldsymbol{k},\theta}$ in $\mathbb{R}^3$.

Solution:
The following diagram shows the effect of rotating the elementary vectors $\boldsymbol{i}$ and $\boldsymbol{j}$ by an arbitrary counterclockwise rotation θ in $\mathbb{R}^2$.



Consideration of the triangles shown and noting the hypotenuse of each is 1 since they sit on the unit circle, gives the following components for $R_\theta(\boldsymbol{i})$ and $R_\theta(\boldsymbol{j})$ respectively:

$$R_\theta(\boldsymbol{i}) = \begin{bmatrix}\cos\theta\\ \sin\theta\end{bmatrix} \qquad\qquad R_\theta(\boldsymbol{j}) = \begin{bmatrix}-\sin\theta\\ \cos\theta\end{bmatrix}.$$

These are the columns of the equivalent matrix transformation for R_θ and we have that R_θ equals the matrix transformation L_A where

$$A = \begin{bmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{bmatrix}.$$

This is the matrix that was found in Example 6-9.

In $\mathbb{R}^3$ if we rotate by θ about the z-axis the elementary vectors $\boldsymbol{i}$ and $\boldsymbol{j}$ stay in the plane while $\boldsymbol{k}$ is unchanged and we have

$$R_{\boldsymbol{k},\theta}(\boldsymbol{i}) = \begin{bmatrix}\cos\theta\\ \sin\theta\\ 0\end{bmatrix} \qquad R_{\boldsymbol{k},\theta}(\boldsymbol{j}) = \begin{bmatrix}-\sin\theta\\ \cos\theta\\ 0\end{bmatrix} \qquad R_{\boldsymbol{k},\theta}(\boldsymbol{k}) = \begin{bmatrix}0\\ 0\\ 1\end{bmatrix}.$$

We see that the linear transformation $R_{\boldsymbol{k},\theta}$ equals the matrix transformation L_B with matrix

$$B = \begin{bmatrix}\cos\theta & -\sin\theta & 0\\ \sin\theta & \cos\theta & 0\\ 0 & 0 & 1\end{bmatrix}.$$

By describing our basic linear transformations by vectors and their operations we were able to show that they were linear by taking advantage of the linear properties of scalar, dot, and cross products. An additional advantage of this approach is that we observe that a linear transformation is independent of choice of Cartesian coordinates (aside from the fixed point origin upon which they all depended) just as vectors are.

When considering matrix transformations we have defined them with respect to the standard basis $\{\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_{\boldsymbol{n}}\}$ of $\mathbb{R}^n$. When using $\mathbb{R}^3$, say, to represent a physical problem then the components of a vector will depend on the orientation of the coordinate axes chosen in physical space which, in turn, determine the directions the elementary vectors $\boldsymbol{e}_i$ represent. The matrix A representing a linear transformation $L(\boldsymbol{u})$ such as a rotation of θ about a particular direction in physical space will itself depend on this orientation of axes. Someone else choosing the same origin but a different orientation of coordinate axes would find different components and matrix to represent the same physical vector and linear transformation. It is a standard problem in linear algebra to consider how the components of the same vector in different coordinate systems are related. Similarly one considers how the matrix representations of the same operator in different systems are related.

Having chosen a particular orientation of coordinate system in space one can *define* a linear transformation by a matrix transformation acting on the standard basis of that system. However the matrix representation of that transformation in other coordinate systems will be, in general, a different matrix. With further linear algebra, that matrix can be determined knowing the orientation of the other coordinate system to the initial one.

6.3 Composition of Linear Transformations

If $f(x) = \sin x$ and $g(x) = x^2$, then the composition of functions $f \circ g$ is defined by $(f \circ g)(x) = f(g(x)) = \sin(x^2)$, with g applied first and f applied to that result. Order in application of the functions typically matters, as it does in this example where $g \circ f \neq f \circ g$ since $(g \circ f)(x) = g(f(x)) = (\sin x)^2$ which is a different function. We have seen that linear combinations of transformations from $\mathbb{R}^n$ to $\mathbb{R}^m$ produced new transformations and that if all the transformations were linear the new transformation was also linear. To this method of producing new transformation we can add transformation composition.

Definition: If S is a transformation from $\mathbb{R}^m$ to $\mathbb{R}^l$ and T is a transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$ then the composition of transformations $S \circ T$ defined by

$$\boxed{(S \circ T)(\boldsymbol{u}) = S(T(\boldsymbol{u}))},$$

is a transformation from $\mathbb{R}^n$ to $\mathbb{R}^l$.

Composition of more that two transformations is similarly accomplished. In the event the transformations are linear we have the following result.

Theorem 6-5: Let K be a linear transformation from $\mathbb{R}^m$ to $\mathbb{R}^l$ and let L be a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$ equal to matrix transformations L_A and L_B respectively where A is an $l \times m$ matrix and B is an $m \times n$ matrix. Then the composition transformation $K \circ L$ defined by

$$\boxed{(K \circ L)(\boldsymbol{u}) = K(L(\boldsymbol{u}))},$$

is a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^l$ and equals the matrix transformation given by L_{AB} where AB is the $l \times n$ matrix product of A and B.

Proof:

$$(K \circ L)(\boldsymbol{u}) = K[L(\boldsymbol{u})] = K(B\boldsymbol{u}) = A(B\boldsymbol{u}) = (AB)\boldsymbol{u} = L_{AB}(\boldsymbol{u})\,.$$

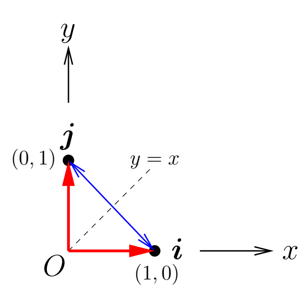
Note the following:

1. The matrix of the transformation applied first is placed furthest to the right.
2. Composition of more than one linear transformation is possible (assuming the dimensions of $\mathbb{R}$ align appropriately) and the result can be shown to be linear by induction.
3. The fact that function composition typically depends on the order of application of the functions is mirrored by the fact that matrix multiplication does not, in general, commute ($AB \neq BA$).

Example 6-14

If K is the reflection about the line $y = x$ and L is the rotation $R_{\frac{\pi}{2}}$ in $\mathbb{R}^2$, find a matrix transformation equal to the composition $K \circ L$. What, geometrically, is the new linear transformation?

Solution:
The line $y = x$ is the diagonal with slope $m = 1$. Reflection of $\boldsymbol{i}$ across it goes to $\boldsymbol{j}$ and similarly $\boldsymbol{j}$ becomes $\boldsymbol{i}$. In other words, the endpoints $(0,1)$ and $(1,0)$ exchange locations. Thus

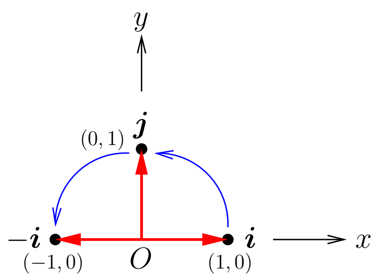


$$K(\boldsymbol{i}) = \boldsymbol{j} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

$$K(\boldsymbol{j}) = \boldsymbol{i} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

so $K = L_A$ where $A = [K(\boldsymbol{i})K(\boldsymbol{j})] = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$.

L is a positive rotation of $\frac{\pi}{2} = 90°$ which rotates $\boldsymbol{i}$ to $\boldsymbol{j}$, i.e. point $(1, 0)$ goes to $(0, 1)$. The vector $\boldsymbol{j}$ is rotated to $-\boldsymbol{i}$, i.e. point $(0, 1)$ goes to $(-1, 0)$ under the rotation. Thus



$$L(\boldsymbol{i}) = \boldsymbol{j} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

$$L(\boldsymbol{j}) = -\boldsymbol{i} = -\begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} -1 \\ 0 \end{bmatrix}$$

so $L = L_B$ where $B = [L(\boldsymbol{i})L(\boldsymbol{j})] = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$.

The composition transformation $K \circ L$ is equivalent to the matrix transformation L_{AB} where AB is the product of the matrices

$$AB = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

The effect of the transformation composition on a vector is thus

$$K \circ L(\boldsymbol{x}) = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x \\ -y \end{bmatrix}.$$

We see that under $K \circ L$ the x-component is unchanged but the y-component is flipped. This a reflection about the x-axis. The normal to that axis is the vector $\boldsymbol{j}$ and so we have the identification

$$K \circ L = Q_{\boldsymbol{j}}.$$

Example 6-15

In addition to writing an arbitrary rotation in three dimensions by $R = R_{\boldsymbol{n},\theta}$ it is possible to write it as a composition of three angular rotations about the axes of a fixed coordinate system with basis $\{\boldsymbol{i},\boldsymbol{j},\boldsymbol{k}\}$, such as

$$R_{\alpha,\beta,\gamma} = R_{\boldsymbol{k},\gamma} \circ R_{\boldsymbol{j},\beta} \circ R_{\boldsymbol{i},\alpha}.$$

The matrix transformation is then the product of the three corresponding matrices. The angles (α, β, γ) are referred to as **Euler Angles**. There are many conventions for such angles. This formulation of rotation has utility when describing the position of the axes of one coordinate system with respect to another.

6.4 Linear Operators

In most of the examples we have done so far our linear transformations have been from $\mathbb{R}^n$ back to itself. This important class of transformations has its own name.

Definition: A **linear operator** is a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^n$.

So dilations, contractions, inversions, expansions, compressions, reflections, and rotations are all linear operators. The matrix of a linear operator from $\mathbb{R}^n$ to $\mathbb{R}^n$ will be a square matrix of order n and we can classify our operators according to the type of matrix transformation they equal.

6.4.1 Symmetric Operators

Recall a symmetric matrix satisfies $A^T = A$. Consideration of our general transformation $C_k(\boldsymbol{u}) = k\boldsymbol{u}$, of which the identity, zero, dilation, contraction, inversion operators are special cases, show they all give rise to symmetric matrix representations. For instance in $\mathbb{R}^3$ we have, by consideration of $C_k(\boldsymbol{i})$, etc., that C_k is equal to L_A where

$$A = \begin{bmatrix} k & 0 & 0 \\ 0 & k & 0 \\ 0 & 0 & k \end{bmatrix}$$

which is clearly symmetric.

As another example, the projection operator $P_{\boldsymbol{n}}$, where $\boldsymbol{n} = (n_x, n_y, n_z)$ is a unit vector, has a symmetric matrix representation. In $\mathbb{R}^3$ we find, by evaluating $P_{\boldsymbol{n}}(\boldsymbol{i})$, etc., that $P_{\boldsymbol{n}}$ is equal to L_A where

$$A = \begin{bmatrix} n_x n_x & n_y n_x & n_z n_x \\ n_x n_y & n_y n_y & n_z n_y \\ n_x n_z & n_y n_z & n_z n_z \end{bmatrix} ,$$

which is symmetric. Since $(A + B)^T = A^T + B^T = A + B$ for symmetric matrices A, B, it follows that the sum of symmetric matrices is also symmetric. Therefore the more general orthogonal projection $P_{\boldsymbol{n}_1,\boldsymbol{n}_2,\ldots,\boldsymbol{n}_k}$ will also induce a symmetric matrix.

As yet a further example consider the linear operator $S_{\boldsymbol{n},k}$ of which compression, expansion, and reflection are special cases. We saw that $S_{\boldsymbol{n},k}$ could be written

$$S_{\boldsymbol{n},k} = \mathbb{1} + (k-1)P_{\boldsymbol{n}} .$$

This will also be symmetric since $(k-1)$ is a scalar and $(cA)^T = cA^T = cA$ if A is symmetric. This implies that cA is symmetric if A is and adding this to symmetric $\mathbb{1}$ will still be symmetric. Explicitly in $\mathbb{R}^3$ we have that $S_{\boldsymbol{n},k}$ is equal to L_A where

$$A = \begin{bmatrix} 1 + (k-1)n_x n_x & (k-1)n_y n_x & (k-1)n_z n_x \\ (k-1)n_x n_y & 1 + (k-1)n_y n_y & (k-1)n_z n_y \\ (k-1)n_x n_z & (k-1)n_y n_z & 1 + (k-1)n_z n_z \end{bmatrix} ,$$

which explicitly is symmetric.

Definition: A **linear operator** on $\mathbb{R}^n$ equipped with the usual dot product is called **symmetric** if it equals a matrix transformation L_A where A is a symmetric matrix.

The zero, identity, dilations, contraction, inversion, expansion, compression, and reflection operators are therefore symmetric. The product of symmetric operators is also symmetric if A and B commute since $(AB)^T = B^T A^T = BA = AB$ for two commuting symmetric matrices. As such we expect an arbitrary symmetric matrix may be decomposable in terms of these types of matrices.

Theorem 6-6: A symmetric linear operator L on $\mathbb{R}^n$ satisfies

$$\boxed{\boldsymbol{u} \cdot L(\boldsymbol{v}) = L(\boldsymbol{u}) \cdot \boldsymbol{v}},$$

for all $\boldsymbol{u}$ and $\boldsymbol{v}$ in $\mathbb{R}^n$.

Proof:
L equals L_A where A is a symmetric matrix. Then for any $\boldsymbol{u}$, $\boldsymbol{v}$ in $\mathbb{R}^n$ we have

$$\boldsymbol{u} \cdot L(\boldsymbol{v}) = \boldsymbol{u} \cdot (A\boldsymbol{v}) = (\boldsymbol{u}^T)A\boldsymbol{v} = (\boldsymbol{u}^T)A^T\boldsymbol{v} = (A\boldsymbol{u})^T\boldsymbol{v} = L(\boldsymbol{u}) \cdot \boldsymbol{v}\,.$$

Note that this latter property can also be considered the definition of a symmetric linear operator since it can be shown to imply a symmetric matrix representation.

6.4.2 Idempotent Operators

Recall an idempotent matrix satisfies $A^2 = A$. It can be shown that the zero, identity, and projection operators $P_{\boldsymbol{n}}$ and $P_{\boldsymbol{n}_1,\boldsymbol{n},\ldots,\boldsymbol{n}_k}$ induce idempotent matrix transformations. This can be shown explicitly using our $\mathbb{R}^3$ version of $P_{\boldsymbol{n}}$ shown previously.

Definition: A **linear operator** on $\mathbb{R}^n$ equipped with the usual dot product is called **idempotent** if it equals a matrix transformation L_A where A is an idempotent matrix.

The zero, identity, and projection operators $P_{\boldsymbol{n}}$ and $P_{\boldsymbol{n}_1,\boldsymbol{n},\ldots,\boldsymbol{n}_k}$ are therefore idempotent.

Theorem 6-7: An idempotent linear operator L satisfies $L \circ L = L$. That is, for any vector $\boldsymbol{u}$ in $\mathbb{R}^n$

$$\boxed{L(L(\boldsymbol{u})) = L(\boldsymbol{u})}.$$

Proof: L equals L_A where A is an idempotent matrix. Then for any $\boldsymbol{u}$ in $\mathbb{R}^n$ we have

$$L(L(\boldsymbol{u})) = L(A\boldsymbol{u}) = A(A\boldsymbol{u}) = (A^2)\boldsymbol{u} = A\boldsymbol{u} = L(\boldsymbol{u})\,.$$

Note that this latter property can also be considered the definition of an idempotent linear operator since it can be shown to imply an idempotent matrix representation.

6.4.3 Orthogonal Operators

Inversion, reflections, and rotations fall into an important group of linear operators which we now explore. Recall an orthogonal matrix satisfies $A^{-1} = A^T$. The identity operator's matrix representation is just I, an orthogonal matrix. The inversion operator has matrix representation $-I$ which is also orthogonal. The reflection operator $Q_{\boldsymbol{n}}$ also equals a matrix transformation where A is orthogonal. Finally rotations have orthogonal matrix representations.

Example 6-16

The R_θ matrix given by $A = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$ is orthogonal since $A^T = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}$ and

$$\begin{aligned} AA^T &= \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \\ &= \begin{bmatrix} \cos^2\theta + \sin^2\theta & \cos\theta\sin\theta - \sin\theta\cos\theta \\ \sin\theta\cos\theta - \cos\theta\sin\theta & \cos^2\theta + \sin^2\theta \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I\,. \end{aligned}$$

Similarly $A^T A = I$ and thus $A^{-1} = A^T$.

Using Rodrigues' rotation formula to evaluate $R_{\boldsymbol{n},\theta}(\boldsymbol{i})$, etc. one can show that the rotation operator equals L_A where matrix A is:

$$A = \begin{bmatrix} \cos\theta + n_x^2\,(1-\cos\theta) & n_x n_y\,(1-\cos\theta) - n_z\sin\theta & n_x n_z\,(1-\cos\theta) + n_y\sin\theta \\ n_y n_x\,(1-\cos\theta) + n_z\sin\theta & \cos\theta + n_y^2\,(1-\cos\theta) & n_y n_z\,(1-\cos\theta) - n_x\sin\theta \\ n_z n_x\,(1-\cos\theta) - n_y\sin\theta & n_z n_y\,(1-\cos\theta) + n_x\sin\theta & \cos\theta + n_z^2\,(1-\cos\theta) \end{bmatrix}\,.$$

This can be shown to be orthogonal as well.

Definition: A **linear operator** on $\mathbb{R}^n$ equipped with the usual dot product is called **orthogonal** if it equals a matrix transformation L_A where A is an orthogonal matrix.

The identity, inverse, reflection and rotation operators are therefore orthogonal. Orthogonal linear operators have the following important property.

Theorem 6-8: If L is an orthogonal linear operator on $\mathbb{R}^n$ equipped with the usual dot product then it preserves the dot product. For any vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ in $\mathbb{R}^n$ we have

$$\boxed{L(\boldsymbol{u}) \cdot L(\boldsymbol{v}) = \boldsymbol{u} \cdot \boldsymbol{v}}\,.$$

Proof:
L equals L_A where A is an orthogonal matrix. Then for any vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ in $\mathbb{R}^n$ we have

$$\begin{aligned} L(\boldsymbol{u}) \cdot L(\boldsymbol{v}) = (A\boldsymbol{u}) \cdot (A\boldsymbol{v}) &= (A\boldsymbol{u})^T (A\boldsymbol{v}) = (\boldsymbol{u}^T A^T)(A\boldsymbol{v}) \\ &= \boldsymbol{u}^T (A^T A)\boldsymbol{v} = \boldsymbol{u}^T (A^{-1}A)\boldsymbol{v} = \boldsymbol{u}^T I \boldsymbol{v} = \boldsymbol{u}^T \boldsymbol{v} = \boldsymbol{u} \cdot \boldsymbol{v}\,. \end{aligned}$$

Setting $\boldsymbol{u} = \boldsymbol{v}$ shows $\|L(\boldsymbol{u})\|^2 = L(\boldsymbol{u}) \cdot L(\boldsymbol{u}) = \boldsymbol{u} \cdot \boldsymbol{u} = \|\boldsymbol{u}\|^2$ from which the following corollary is implied.

Corollary: An orthogonal linear operator L on $\mathbb{R}^n$ preserves the length of a vector, $\boxed{\|L(\boldsymbol{u})\| = \|\boldsymbol{u}\|}$.

Note that property is clear for rotations, reflections, and inversions. Since angle between vectors can be written in terms of the dot product and lengths of two vectors it follows that an orthogonal linear operator also preserves angles between vectors.

Note that either the preservation of dot product or of length can be used to define an orthogonal linear operator as they can be shown to be equivalent and they imply an orthogonal matrix representation.

Finally a product of orthogonal matrices is orthogonal since $(AB)^{-1} = B^{-1}A^{-1} = B^T A^T = (AB)^T$ for orthogonal matrices A and B. So we expect a general orthogonal matrix to be decomposable in terms of rotations, reflections, and inversions.

6.4.4 Polar Decomposition of Operators

To conclude this chapter we note the following theorem which shows that every matrix A can be decomposed into a product of a symmetric and an orthogonal matrix.

Theorem 6-9: Let A be a square matrix with real entries. Then A can be written as the product of a symmetric matrix S and an orthogonal matrix O as[2]

$$\boxed{A = SO}\,.$$

This is called the **polar decomposition**[3] of A.

Since any linear operator on $\mathbb{R}^n$ equals a matrix operator L_A we have the following corollary.

Corollary: If L is a linear operator on $\mathbb{R}^n$ then L can be written as the composition of a symmetric operator S and an orthogonal operator O as

$$\boxed{L = S \circ O}\,.$$

These theorems are useful as they characterize all linear operators quite generally as being decomposable into two operations. The first is a length-preserving orthogonal operation which will consist of a composition of rotations, reflections, and inversions. This will then be followed by a symmetric operator which will consist of a composition of operations that will typically scale and project the vector along various directions.

[2] In fact the theorem is stronger than stated. The symmetric matrix found can be restricted to those that are positive semi-definite (i.e. no reflections) and in the event that A is invertible this decomposition is unique.

[3] The term polar decomposition is used as it is analogous to the decomposition of a complex number as $re^{i\theta}$ which, upon multiplication of another complex number will scale it by a factor of r and rotate in the plane by an angle θ. Complex numbers will be discussed in Chapter 9 .

Chapter 7: Subspaces of $\mathbb{R}^n$

7.1 Subspaces of $\mathbb{R}^n$

As the name suggests, a subspace S of $\mathbb{R}^n$ is a set of vectors contained within $\mathbb{R}^n$. It inherits all the properties of $\mathbb{R}^n$. Additionally we require it be closed.

Definition: A non-empty subset S of vectors from $\mathbb{R}^n$ that is **closed** under vector addition and scalar multiplication is called a **subspace** of $\mathbb{R}^n$. Closure means that the vectors $\boldsymbol{u}+\boldsymbol{v}$ and $c\boldsymbol{u}$ are also in S for any vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ in S and scalar c.

Theorem 7-1: The zero vector $\mathbf{0}_n$ is an element of S for any subspace S of $\mathbb{R}^n$.

Proof: Since S is non-empty let $\boldsymbol{u}$ be a vector in S. Then $(-1)\boldsymbol{u} = -\boldsymbol{u}$ is also in S due to scalar closure and so $\boldsymbol{u}+(-\boldsymbol{u}) = \boldsymbol{u}-\boldsymbol{u} = \mathbf{0}$ lies in S due to closure under addition.

Note that $S = \{\mathbf{0}_n\}$ is a subspace of $\mathbb{R}^n$. Also $S = \mathbb{R}^n$ itself is a subspace of $\mathbb{R}^n$.

Definition: A set of vectors $B = \{\boldsymbol{u_1}, \boldsymbol{u_2}, \ldots, \boldsymbol{u_k}\}$ in subspace S is said to **span** S if every vector $\boldsymbol{v}$ in S can be written as a linear combination of vectors from B. That is, there exist scalars $c_1, \ldots, c_k$ such that

$$\boldsymbol{v} = c_1\boldsymbol{u}_1 + c_2\boldsymbol{u}_2 + \ldots + \boldsymbol{u}_k\,,$$

for every $\boldsymbol{v}$ in S.

Example 7-1

The standard basis $B = \{\boldsymbol{e}_1, \ldots, \boldsymbol{e_n}\}$ spans $\mathbb{R}^n$.

Conversely we can define the span of a set of vectors.

Definition: Let $B = \{\boldsymbol{u_1}, \boldsymbol{u_2}, \ldots, \boldsymbol{u_k}\}$ be a set of k vectors from $\mathbb{R}^n$. Then the **span of the set**, denoted by $\operatorname{span}(B) = \operatorname{span}\{\boldsymbol{u_1}, \boldsymbol{u_2}, \ldots, \boldsymbol{u_k}\}$, is the set of all linear combinations of the vectors in the set, i.e. all vectors

$$\boldsymbol{v} = c_1\boldsymbol{u_1} + c_2\boldsymbol{u_2} + \cdots + c_k\boldsymbol{u}_k$$

where c_i are scalars.

Since clearly the sum of two such linear combinations will still be a linear combination and the scalar product of such a linear combination will also be so, we have the following non-trivial subspaces.

Theorem 7-2: The span of a set of vectors $B = \{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_n\}$ from $\mathbb{R}^n$, $\operatorname{span}(B)$, is a subspace of $\mathbb{R}^n$.

Example 7-2

Let $\boldsymbol{v}$ be a nonzero vector in $\mathbb{R}^2$ or $\mathbb{R}^3$ and let t be a parameter. Then the line through the origin given by the set

$$L = \{\boldsymbol{x}(t) = t\boldsymbol{v} \text{ such that } t \text{ is in } \mathbb{R}\}$$

is a subspace of $\mathbb{R}^2$ or $\mathbb{R}^3$ respectively.

Example 7-3

Let $\boldsymbol{u}$ and $\boldsymbol{v}$ be two nonzero, noncollinear vectors in $\mathbb{R}^3$ and let s and t be parameters. Then the plane through the origin given by the set

$$P = \{\boldsymbol{x}(s,t) = s\boldsymbol{u} + t\boldsymbol{v} \text{ such that } s \text{ and } t \text{ are in } \mathbb{R}\}$$

is a subspace of $\mathbb{R}^3$.

So in the previous examples the sum of two vectors or the product of one of them by a scalar will remain in the subspace, line or plane, respectively.

As a final non-trivial subspace example consider the following.

Definition: If L is a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$ then the **null space** or **kernel** of L is the set of all vectors in $\mathbb{R}^n$ satisfying $L(\boldsymbol{u}) = \mathbf{0}_m$.

If L equals matrix transformation L_A for matrix A then it follows that the null space of L is the set of solutions of the homogeneous linear system

$$A\boldsymbol{x} = \mathbf{0}\,.$$

Theorem 7-3: The null space of a linear transformation L from $\mathbb{R}^n$ to $\mathbb{R}^m$ is a subspace of $\mathbb{R}^n$.

Proof: Let $\boldsymbol{u}$ and $\boldsymbol{v}$ be vectors of $\mathbb{R}^n$ in the null space of linear transformation L and let c be a scalar. Then

$$\begin{aligned} L(\boldsymbol{u}+\boldsymbol{v}) &= L(\boldsymbol{u}) + L(\boldsymbol{v}) = \mathbf{0} + \mathbf{0} = \mathbf{0} \\ L(c\boldsymbol{u}) &= cL(\boldsymbol{u}) = c\mathbf{0} = \mathbf{0} \end{aligned}$$

Thus $\boldsymbol{u}+\boldsymbol{v}$ and $c\boldsymbol{u}$ will also lie in the null space of L and it is therefore closed under addition and scalar multiplication and hence a subspace of $\mathbb{R}^n$.

7.2 Linear Independence

We are often interested in finding the fewest number of vectors required to span a subspace S. To that end we define the following concept.

Definition: A set $B = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n\}$ of vectors from $\mathbb{R}^m$ is **linearly independent** if the vector equation

$$c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \ldots + c_n\boldsymbol{v}_n = \mathbf{0}$$

has only the solution $c_i = 0$ for $i = 1, \ldots, n$. Otherwise the set B is called **linearly dependent**.

Note the following:

- A set of two or more vectors is linearly dependent if and only if one of them can be expressed as a linear combination of the others.
- A linearly independent set of vectors cannot contain the zero vector as, assuming it is the first vector, $\boldsymbol{v}_1 = \mathbf{0}$, then c_1 could be set to anything and all other scalars could be set to 0 thereby providing a solution to the equation which was not identically zero.
- If we let A be the matrix whose i^{th} column is the vector $\boldsymbol{v}_i$, so

  $$A = [\boldsymbol{v}_1\boldsymbol{v}_2 \cdots \boldsymbol{v}_n]$$

  then the previous vector equation is equivalent to a homogeneous system given by:

  $$\begin{bmatrix}\boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_n\end{bmatrix}\begin{bmatrix}c_1\\c_2\\\vdots\\c_n\end{bmatrix} = \begin{bmatrix}0\\0\\\vdots\\0\end{bmatrix}$$

  or

  $$A\boldsymbol{c} = \mathbf{0}\,.$$

  If the system has the unique (trivial) solution $\boldsymbol{c} = \mathbf{0}$, then the set $B = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n\}$ is linearly independent. If this system has infinitely many solutions, then the set B is linearly dependent.
- If the matrix A is square this is further simplified. If $\det A \neq 0$ then we have the unique (trivial) solution and the set is independent. If $\det A = 0$ then there are infinitely many solutions and the set is dependent.

Example 7-4

Determine whether the given set of vectors is linearly dependent or independent.

1. $B = \{\boldsymbol{v}_1, \boldsymbol{v}_2\}$ where $\boldsymbol{v}_1 = (1, 1)$, $\boldsymbol{v}_2 = (-1, 2)\,.$

   Solution:

$$\begin{aligned} c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 &= \mathbf{0} \\ \iff \begin{bmatrix}1 & -1\\1 & 2\end{bmatrix}\begin{bmatrix}c_1\\c_2\end{bmatrix} &= \begin{bmatrix}0\\0\end{bmatrix} \\ A\boldsymbol{c} &= \mathbf{0}\end{aligned}$$

Solving the system $[A|\mathbf{0}]$ directly we have:

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

$$\Downarrow$$

$$R_2 \to R_2 - R_1 \quad \left[\begin{array}{rr|r} 1 & -1 & 0 \\ 0 & 3 & 0 \end{array}\right] \Leftrightarrow \begin{aligned} c_1 - c_2 &= 0 \\ 3c_2 &= 0 \end{aligned}$$

Using back-substitution:

- $3c_2 = 0 \implies \boxed{c_2 = 0}$
- $c_1 - c_2 = 0 \implies c_1 - 0 = 0 \implies \boxed{c_1 = 0}$

So $\boldsymbol{c} = (0, 0)$ is the only solution and therefore $B = \{\boldsymbol{v}_1, \boldsymbol{v}_2\}$ is a linearly independent set. Alternatively, since A is square we can use the determinant:

$$\det A = \begin{vmatrix} 1 & -3 \\ 0 & 3 \end{vmatrix} = 1(2) - (-1)(1) = 3 \neq 0$$

which implies B is a linearly independent set.

2. $B = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3\}$ where $\boldsymbol{v}_1 = (1, 1, -2)$, $\boldsymbol{v}_2 = (2, 5, -1)$, and $\boldsymbol{v}_3 = (0, 1, 1)$.

   Solution:
   We must consider the solutions $\boldsymbol{c} = (c_1, c_2, c_3)$ for the following system:

$$\begin{gathered} c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + c_3\boldsymbol{v}_3 = \mathbf{0} \\ \iff \begin{bmatrix} 1 & 2 & 0 \\ 1 & 5 & 1 \\ -2 & -1 & 1 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \\ A\boldsymbol{c} = 0 \end{gathered}$$

   Expanding coefficient matrix A along the first row, we have

$$\det A = \begin{vmatrix} 1 & 2 & 0 \\ 1 & 5 & 1 \\ -2 & -1 & 1 \end{vmatrix} = (1)(+1)(5+1) + 2(-1)(1+2) + 0 = 6 - 2(3) = 0$$

   Therefore there are infinitely many solutions for $\boldsymbol{c} = (c_1, c_2, c_3)$ and set $B = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3\}$ is linearly dependent.

3. $B = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3\}$ where $\boldsymbol{v}_1 = (0, 0, 2, 2)$, $\boldsymbol{v}_2 = (3, 3, 0, 0)$, and $\boldsymbol{v}_3 = (1, 1, 0, -1)$.

   Solution:

$$\begin{gathered} c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + c_3\boldsymbol{v}_3 = \mathbf{0} \\ \iff \begin{bmatrix} 0 & 3 & 1 \\ 0 & 3 & 1 \\ 2 & 0 & 0 \\ 2 & 0 & -1 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \\ A\boldsymbol{c} = \mathbf{0} \end{gathered}$$

Since A is not square so solve the system using $[A|\mathbf{0}]$:

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

$$\Downarrow$$

$$R_1 \leftrightarrow R_3 \left[\begin{array}{ccc|c} 2 & 0 & 0 & 0 \\ 0 & 3 & 1 & 0 \\ 0 & 3 & 1 & 0 \\ 2 & 0 & -1 & 0 \end{array}\right]$$

$$\Downarrow$$

$$\begin{array}{r} \\ \\ R_3 \to R_3 - R_2 \\ R_4 \to R_4 - R_1 \end{array} \left[\begin{array}{ccc|c} 2 & 0 & 0 & 0 \\ 0 & 3 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 \end{array}\right]$$

$$\Downarrow$$

$$R_3 \leftrightarrow R_4 \left[\begin{array}{ccc|c} 2 & 0 & 0 & 0 \\ 0 & 3 & 1 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right] \Leftrightarrow \begin{aligned} 2c_1 &= 0 \\ 3c_2 + c_3 &= 0 \\ -c_3 &= 0 \\ 0 &= 0 \end{aligned}$$

Using back-substitution:

- $-c_3 = 0 \implies \boxed{c_3 = 0}$
- $3c_2 + c_3 = 0 \implies 3c_2 + 0 = 0 \implies \boxed{c_2 = 0}$
- $2c_1 = 0 \implies \boxed{c_1 = 0}$

Therefore the only solution is $(c_1, c_2, c_3) = \mathbf{0}$ and $B = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3\}$ is linearly independent.

Definition: Let S be a subspace of $\mathbb{R}^n$ then if $B = \{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots \boldsymbol{u}_k\}$ is a linearly independent set of vectors that spans S, then B is called a **basis** for S.

Example 7-5

The standard basis $B = \{\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n\}$ is a basis for $\mathbb{R}^n$.

Example 7-6

The set $B = \{\boldsymbol{v}\}$ where $\boldsymbol{v}$ is a nonzero vector is a basis for the line L in $\mathbb{R}^2$ or $\mathbb{R}^3$ given in Example 7-2 .

Example 7-7

The set $B = \{\boldsymbol{u}, \boldsymbol{v}\}$ where $\boldsymbol{u}$ and $\boldsymbol{v}$ are nonzero, noncollinear vectors is a basis for the plane P in $\mathbb{R}^3$ given in Example 7-3.

Note the following:

- A basis for S is made up of the smallest number of vectors that will span S.
- While the vectors that make up a basis for S are not unique, it can be shown that every basis must have the same number of elements. This is called the **dimension** of S. So the line L in $\mathbb{R}^2$ or $\mathbb{R}^3$ has dimension of 1, and the plane P in $\mathbb{R}^3$ has dimension of 2 as expected.
- If $\mathbb{R}^n$ is equipped with the usual dot product, note that the vectors in the basis for a subspace S need not be unit vectors nor do they need to be mutually orthogonal. However it can be shown that given such a basis one can always construct a new basis that has these properties. One such procedure is called the **Gram-Schmidt process**. Such a basis is called an **orthonormal basis**.

Chapter 8: Eigenvalues and Eigenvectors

8.1 Eigenvalues and Eigenvectors

Definition: Let A be a square $n \times n$ matrix. A scalar λ is said to be an **eigenvalue** or a **characteristic value** of A if there exists a nonzero vector $\boldsymbol{x}$ in $\mathbb{R}^n$ such that:

$$\boxed{A\boldsymbol{x} = \lambda\boldsymbol{x}}\,.$$

The nonzero vector $\boldsymbol{x}$ is called an **eigenvector** of A corresponding to the eigenvalue λ.

Example 8-1

Show $\boldsymbol{x} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$ is an eigenvector of $A = \begin{bmatrix} 4 & -2 \\ 1 & 1 \end{bmatrix}$ and find its corresponding eigenvalue.

Solution:

$$A\boldsymbol{x} = \begin{bmatrix} 4 & -2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 8-2 \\ 2+1 \end{bmatrix} = \begin{bmatrix} 6 \\ 3 \end{bmatrix} = 3 \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

Thus $A\boldsymbol{x} = 3\boldsymbol{x}$ and therefore $\boldsymbol{x} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$ is an eigenvector of A corresponding to eigenvalue $\lambda = 3$.

Example 8-2

Show $\boldsymbol{x} = (-2, 1, 1)$ is an eigenvector of $A = \begin{bmatrix} 0 & 0 & -2 \\ 1 & 2 & 1 \\ 1 & 0 & 3 \end{bmatrix}$ and find its corresponding eigenvalue.

Solution:

$$A\boldsymbol{x} = \begin{bmatrix} 0 & 0 & -2 \\ 1 & 2 & 1 \\ 1 & 0 & 3 \end{bmatrix} \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0+0-2 \\ -2+2+1 \\ -2+0+3 \end{bmatrix} = \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix} = (1)\boldsymbol{x}$$

Thus $A\boldsymbol{x} = 1\boldsymbol{x}$ and $\boldsymbol{x} = (-2, 1, 1)$ is an eigenvector of A corresponding to eigenvalue $\lambda = 1$.

A nonzero scalar multiple of an eigenvector of A also is an eigenvector of A.

Example 8-3

In Example 8-2 it was shown $\boldsymbol{x} = (-2, 1, 1)$ was an eigenvector of A corresponding to eigenvalue $\lambda = 1$. Then $2\boldsymbol{x} = (-4, 2, 2)$ is also an eigenvector corresponding to $\lambda = 1$ since

$$A(2\boldsymbol{x}) = \begin{bmatrix} 0 & 0 & -2 \\ 1 & 2 & 1 \\ 1 & 0 & 3 \end{bmatrix} \begin{bmatrix} -4 \\ 2 \\ 2 \end{bmatrix} = \begin{bmatrix} 0+0-4 \\ -4+4+2 \\ -4+0+6 \end{bmatrix} = \begin{bmatrix} -4 \\ 2 \\ 2 \end{bmatrix} = (1) \begin{bmatrix} -4 \\ 2 \\ 2 \end{bmatrix} = (1)(2\boldsymbol{x})$$

Therefore $2\boldsymbol{x}$ is also an eigenvector of A with the same eigenvalue of $\lambda = 1$.

Additionally if two eigenvectors correspond to the same eigenvalue λ of A their sum will also be an eigenvector of A. We summarize these results in the following theorem.

Theorem 8-1: If $\boldsymbol{u}$ and $\boldsymbol{v}$ are eigenvectors of A associated with the same eigenvalue λ, and c is a scalar then $\boldsymbol{u} + \boldsymbol{v}$ and $c\boldsymbol{u}$ are also eigenvectors of A associated with λ provided they do not equal the zero vector. (i.e. $\boldsymbol{v} \neq -\boldsymbol{u}$ and $c \neq 0$.)

Proof:
Let $\boldsymbol{u}$, $\boldsymbol{v}$ be eigenvectors of A corresponding to eigenvalue λ and let c be a scalar. Then

$$\begin{aligned} A(\boldsymbol{u}+\boldsymbol{v}) &= A\boldsymbol{u}+A\boldsymbol{v} = \lambda\boldsymbol{u}+\lambda\boldsymbol{v} = \lambda(\boldsymbol{u}+\boldsymbol{v}) \\ A(c\boldsymbol{u}) &= c(A\boldsymbol{u}) = c(\lambda\boldsymbol{u}) = \lambda(c\boldsymbol{u})\,. \end{aligned}$$

It follows from Theorem 8-1 that the eigenvectors corresponding to a particular eigenvalue λ of A combined with the zero vector $\mathbf{0}$ form a subspace of $\mathbb{R}^n$.

Definition: The subspace of $\mathbb{R}^n$ consisting of all eigenvectors $\boldsymbol{x}$ associated with a particular eigenvalue λ of A and the zero vector is called the **eigenspace** of A corresponding to λ.

8.2 Finding Eigenvalues and Eigenvectors

Recall that if B is a square matrix of order n then the homogeneous system $B\boldsymbol{x} = 0$ has the unique trivial solution $\boldsymbol{x} = \mathbf{0}$ if $\det B \neq 0$ and it have infinitely many nontrivial solutions if $\det B = 0$. We can cast the problem of finding eigenvalues and eigenvectors of the $n \times n$ square matrix A in this form by solving the eigenvalue equation for $\boldsymbol{x}$.

$$\begin{aligned} A\boldsymbol{x} &= \lambda\boldsymbol{x} \\ \Rightarrow A\boldsymbol{x} - \lambda\boldsymbol{x} &= 0 \\ \Rightarrow A\boldsymbol{x} - \lambda I\boldsymbol{x} &= 0 \\ \Rightarrow (A - \lambda I)\boldsymbol{x} &= 0 \end{aligned}$$

Setting $B = A - \lambda I$ it follows that to have non-trivial ($\boldsymbol{x} \neq \mathbf{0}$) solutions to this system of n equations in n unknowns we require

$$\boxed{\det(A - \lambda I) = 0}.$$

Theorem 8-2: Let A be an $n \times n$ matrix. The number λ is an eigenvalue of A if and only if $\det(A - \lambda I) = 0$.

Consideration of the form of the determinant shows the left hand side of the equation is a polynomial in λ with coefficients determined by A. With that in mind we have the following definition.

Definition: The polynomial $P_A(\lambda) = \det(A - \lambda I)$ is called the **characteristic polynomial** of A and the equation $\det(A - \lambda I) = 0$ is called its **characteristic equation**.

To find the eigenvalues of a matrix A and their associated eigenvectors now becomes a two step process:

1. Solve the characteristic equation for the eigenvalues of A.
2. For each eigenvalue λ solve the linear system $\boxed{(A - \lambda I)\boldsymbol{x} = \mathbf{0}}$ for $\boldsymbol{x}$ to find its associated eigenvectors.

Example 8-4

Find the eigenvalues and eigenvectors of the given matrix.

$$A = \begin{bmatrix} 2 & 3 \\ 4 & 3 \end{bmatrix}$$

Solution:

$$\begin{aligned} \det(A - \lambda I) = \begin{vmatrix} 2-\lambda & 3 \\ 4 & 3-\lambda \end{vmatrix} &= 0 \\ \Rightarrow (2-\lambda)(3-\lambda) - 12 &= 0 \\ \Rightarrow 6 - 2\lambda - 3\lambda + \lambda^2 - 12 &= 0 \\ \Rightarrow \lambda^2 - 5\lambda - 6 &= 0 \\ \Rightarrow (\lambda - 6)(\lambda + 1) &= 0 \\ \Rightarrow \begin{cases} \lambda_1 = 6 \\ \lambda_2 = -1 \end{cases} \end{aligned}$$

$\lambda_1 = 6$:

To find the eigenvalues $\boldsymbol{x}_1 = (x_1, x_2)$ corresponding to λ_1 we need to solve:

$$(A - 6I)\boldsymbol{x}_1 = \mathbf{0}$$

Since

$$A - 6I = \begin{bmatrix} 2 & 3 \\ 4 & 3 \end{bmatrix} - 6\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} -4 & 3 \\ 4 & -3 \end{bmatrix}$$

this can be done using the augmented matrix as follows

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

$$\Downarrow$$

$$R_2 \to R_2 + R_1 \left[\begin{array}{cc|c} -4 & 3 & 0 \\ 0 & 0 & 0 \end{array}\right] \quad \Leftrightarrow \quad \begin{aligned} -4x_1 + 3x_2 &= 0 \\ 0 &= 0 \end{aligned}$$

Here we have circled the leading entry. Back-substitution in the linear system gives

- $\boxed{x_2 = t}$
- $-4x_1 + 3x_2 = 0 \implies -4x_1 + 3t = 0 \implies \boxed{x_1 = \frac{3}{4}t}$

Thus the eigenvectors are $\boldsymbol{x}_1 = \begin{bmatrix} \frac{3}{4}t \\ t \end{bmatrix} = t\begin{bmatrix} \frac{3}{4} \\ 1 \end{bmatrix} = \frac{1}{4}t\begin{bmatrix} 3 \\ 4 \end{bmatrix}$ and $\{(3, 4)\}$ is a basis for the $\lambda = 6$ eigenspace.

$\lambda_2 = -1$:

$$\begin{aligned} (A - (-1)I)\boldsymbol{x}_2 &= \mathbf{0} \\ (A + I)\boldsymbol{x}_2 &= \mathbf{0} \end{aligned}$$

$$\left[\begin{array}{cc|c} 3 & 3 & 0 \\ 4 & 4 & 0 \end{array}\right]$$

$$\Downarrow$$

$$\begin{array}{l} R_1 \to \frac{1}{3}R_1 \\ R_2 \to \frac{1}{4}R_2 \end{array} \left[\begin{array}{cc|c} 1 & 1 & 0 \\ 1 & 1 & 0 \end{array}\right]$$

$$\Downarrow$$

$$R_2 \to R_2 - R_1 \left[\begin{array}{cc|c} 1 & 1 & 0 \\ 0 & 0 & 0 \end{array}\right]$$

- $\boxed{x_2 = t}$
- $x_1 + x_2 = 0 \implies x_1 + t = 0 \implies \boxed{x_1 = -t}$

The eigenvectors are $\boldsymbol{x}_2 = \begin{bmatrix} -t \\ t \end{bmatrix} = t\begin{bmatrix} -1 \\ 1 \end{bmatrix}$ and $\{(-1, 2)\}$ is a basis for the $\lambda = -1$ eigenspace.

Note:

- Since the eigenvalues of a square matrix A are the solutions of the corresponding characteristic equation (i.e. roots or zeroes of the characteristic polynomial), it follows by the Fundamental Theorem of Algebra that an $n \times n$ matrix A has at least one eigenvalue and at most n numerically different eigenvalues. (So a 3×3 matrix has 1, 2, or 3 eigenvalues). However these eigenvalues may not all be real-valued.
- Since the eigenvectors corresponding to λ are solutions of
$$(A - \lambda I)\boldsymbol{x} = \mathbf{0}$$
the eigenspace of A corresponding to eigenvalue λ is the null space of the matrix $B = A - \lambda I$. i.e. the solutions of $B\boldsymbol{x} = \mathbf{0}$. The basic solution eigenvectors will span the eigenspace. So in Example 8-4 $\{(3,4)\}$ is a basis for the $\lambda = 6$ eigenspace and $\{(-1,1)\}$ is a basis for the $\lambda = -1$ eigenspace.

Example 8-5

Find the eigenvalues and bases for the eigenspaces of the given matrix.

$$A = \begin{bmatrix} 5 & -7 & 7 \\ 4 & -3 & 4 \\ 4 & -1 & 2 \end{bmatrix}$$

Solution:
First find the eigenvalues:

$\det(A - \lambda I) = 0$

$$\begin{aligned}
&\Rightarrow \begin{vmatrix} 5-\lambda & -7 & 7 \\ 4 & -3-\lambda & 4 \\ 4 & -1 & 2-\lambda \end{vmatrix} = 0 \\
&\Rightarrow (5-\lambda)[(-3-\lambda)(2-\lambda)+4] + 7[4(2-\lambda)-16] + 7[-4-4(-3-\lambda)] = 0 \\
&\Rightarrow (5-\lambda)[\lambda^2+\lambda-2] + 7[-4\lambda-8] + 7[4\lambda+8] = 0 \\
&\Rightarrow 5\lambda^2 + 5\lambda - 10 - \lambda^3 - \lambda^2 + 2\lambda = 0 \ \text{ (Simplify and multiply both sides by } -1\text{.)} \\
&\Rightarrow \lambda^3 - 4\lambda^2 - 7\lambda + 10 = 0 \ \text{ (Note } \lambda = 1 \text{ solves this. Group to get } (\lambda - 1) \text{ factor.)} \\
&\Rightarrow \lambda^3 - 4\lambda^2 - 7\lambda + 7 + 4 - 1 = 0 \\
&\Rightarrow (\lambda^3 - 1) - 4(\lambda^2 - 1) - 7(\lambda - 1) = 0 \ \text{ (Use } a^3 - b^3 = (a-b)(a^2+ab+b^2)\text{.)} \\
&\Rightarrow (\lambda-1)(\lambda^2+\lambda+1) - 4(\lambda+1)(\lambda-1) - 7(\lambda-1) = 0 \\
&\Rightarrow (\lambda-1)[\lambda^2+\lambda+1-4(\lambda+1)-7] = 0 \\
&\Rightarrow (\lambda-1)(\lambda^2-3\lambda-10) = 0 \\
&\Rightarrow (\lambda-1)(\lambda-5)(\lambda+2) = 0 \\
&\Rightarrow \begin{cases} \lambda_1 = 1 \\ \lambda_2 = 5 \\ \lambda_3 = -2 \end{cases}
\end{aligned}$$

$\boldsymbol{\lambda_1 = 1}$**:**

$$(A - (1)I)\boldsymbol{x}_1 = \mathbf{0}$$

$$\left[\begin{array}{ccc|c} 4 & -7 & 7 & 0 \\ 4 & -4 & 4 & 0 \\ 4 & -1 & 1 & 0 \end{array}\right]$$

$$\Downarrow$$

$$\begin{array}{r} \\ R_2 \to R_2 - R_1 \\ R_3 \to R_3 - R_1 \end{array} \left[\begin{array}{rrr|r} 4 & -7 & 7 & 0 \\ 0 & 3 & -3 & 0 \\ 0 & 6 & -6 & 0 \end{array}\right]$$

$$\Downarrow$$

$$R_3 \to R_3 - 2R_2 \left[\begin{array}{rrr|r} 4 & -7 & 7 & 0 \\ 0 & 3 & -3 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

- $\boxed{x_3 = t}$
- $3x_2 - 3x_3 = 0 \implies 3x_2 - 3t = 0 \implies \boxed{x_2 = t}$
- $4x_1 - 7x_2 + 7x_3 = 0 \implies 4x_1 - 7t + 7t = 0 \implies \boxed{x_1 = 0}$

Therefore $\boldsymbol{x}_1 = \begin{bmatrix} 0 \\ t \\ t \end{bmatrix} = t \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$ and so $\{(0, 1, 1)\}$ is a basis for the $\lambda = 1$ eigenspace.

$\boldsymbol{\lambda_2 = 5:}$

$$(A - 5I)\boldsymbol{x}_2 = \mathbf{0}$$

$$\left[\begin{array}{rrr|r} 0 & -7 & 7 & 0 \\ 4 & -8 & 4 & 0 \\ 4 & -1 & -3 & 0 \end{array}\right]$$

$$\Downarrow$$

$$R_1 \leftrightarrow R_3 \left[\begin{array}{rrr|r} 4 & -1 & -3 & 0 \\ 4 & -8 & 4 & 0 \\ 0 & -7 & 7 & 0 \end{array}\right]$$

$$\Downarrow$$

$$R_2 \to R_2 - R_1 \left[\begin{array}{rrr|r} 4 & -1 & -3 & 0 \\ 0 & -7 & 7 & 0 \\ 0 & -7 & 7 & 0 \end{array}\right]$$

$$\Downarrow$$

$$R_3 \to R_3 - R_2 \left[\begin{array}{rrr|r} 4 & -1 & -3 & 0 \\ 0 & -7 & 7 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

- $\boxed{x_3 = t}$
- $-7x_2 + 7x_3 = 0 \implies -7x_2 + 7t = 0 \implies \boxed{x_2 = t}$
- $4x_1 - x_2 - 3x_3 = 0 \implies 4x_1 - t - 3t = 0 \implies \boxed{x_1 = t}$

Therefore $\boldsymbol{x}_2 = \begin{bmatrix} t \\ t \\ t \end{bmatrix} = t \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ and so $\{(1, 1, 1)\}$ is a basis for the $\lambda = 5$ eigenspace.

$\boldsymbol{\lambda_3 = -2:}$

$$(A - (-2)I)\boldsymbol{x}_3 = \mathbf{0}$$
$$(A + 2I)\boldsymbol{x}_3 = \mathbf{0}$$

$$\left[\begin{array}{ccc|c} 7 & -7 & 7 & 0 \\ 4 & -1 & 4 & 0 \\ 4 & -1 & 4 & 0 \end{array}\right]$$

$$\Downarrow$$

$$\begin{array}{r} R_1 \to \frac{1}{7}R_1 \\ \\ R_3 \to R_3 - R_2 \end{array} \left[\begin{array}{ccc|c} 1 & -1 & 1 & 0 \\ 4 & -1 & 4 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

$$\Downarrow$$

$$R_2 \to R_2 - 4R_1 \left[\begin{array}{ccc|c} 1 & -1 & 1 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

- $\boxed{x_3 = t}$
- $3x_2 = 0 \implies \boxed{x_2 = 0}$
- $x_1 - x_2 + x_3 = 0 \implies x_1 - 0 + t = 0 \implies \boxed{x_1 = -t}$

Therefore $\boldsymbol{x}_3 = \begin{bmatrix} -t \\ 0 \\ t \end{bmatrix} = t \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$ and so $\{(-1, 0, 1)\}$ is a basis for the $\lambda = -2$ eigenspace.

It can occur that the characteristic polynomial $P_A(\lambda)$ contains a particular eigenvalue λ more than once as a root. Additionally the corresponding eigenspace of λ may have a dimension greater than one.

Definition: The **algebraic multiplicity** M of an eigenvalue λ of square matrix A is the number of times it appears as a root of the characteristic polynomial $P_A(\lambda)$. The **geometric multiplicity** m of eigenvalue λ of A is the number of linearly independent eigenvectors corresponding to λ, i.e. the dimension of the eigenspace of λ.

The two multiplicities are related as follows.

Theorem 8-3: If λ is an eigenvalue of A with algebraic multiplicity M then the number m of linearly independent eigenvectors associated with λ (its geometric multiplicity) satisfies

$$1 \leq m \leq M\,.$$

Example 8-6

Find the eigenvalues and eigenvectors of the given matrix. Identify the algebraic and geometric multiplicities of each eigenvalue.

$$A = \begin{bmatrix} 2 & -5 & 5 \\ 0 & 3 & -1 \\ 0 & -1 & 3 \end{bmatrix}$$

Solution:

$$\begin{aligned}
&\det(A - \lambda I) = 0 \\
&\Rightarrow \begin{vmatrix} 2-\lambda & -5 & 5 \\ 0 & 3-\lambda & -1 \\ 0 & -1 & 3-\lambda \end{vmatrix} = 0 \\
&\Rightarrow (2-\lambda)\left[(3-\lambda)^2 - 1\right] = 0 \\
&\Rightarrow (2-\lambda)(9 - 6\lambda + \lambda^2 - 1) = 0 \\
&\Rightarrow (2-\lambda)(\lambda^2 - 6\lambda + 8) = 0 \quad \text{(Multiply both sides by -1 and factor.)} \\
&\Rightarrow (\lambda - 4)(\lambda - 2)(\lambda - 2) = 0 \\
&\Rightarrow \begin{cases} \lambda_1 = 4, \; M_1 = 1 \\ \lambda_2 = 2, \; M_2 = 2 \end{cases}
\end{aligned}$$

Next find the corresponding eigenvectors.

$\boldsymbol{\lambda_1 = 4:}$

$$(A - 4I)\boldsymbol{x}_1 = \mathbf{0}$$

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

$$\Downarrow$$

$$R_3 \to R_3 - R_2 \quad \left[\begin{array}{ccc|c} -2 & -5 & 5 & 0 \\ 0 & -1 & -1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

- $\boxed{x_3 = t}$
- $-x_2 - x_3 = 0 \implies -x_2 - t = 0 \implies \boxed{x_2 = -t}$
- $-2x_1 - 5x_2 + 5x_3 = 0 \implies -2x_1 - 5(-t) + 5t = 0 \implies \boxed{x_1 = 5t}$

Therefore $\boldsymbol{x}_1 = \begin{bmatrix} 5t \\ -t \\ t \end{bmatrix} = t \begin{bmatrix} 5 \\ -1 \\ 1 \end{bmatrix}$, so $\{(5, -1, 1)\}$ is a basis for the $\lambda = 4$ eigenspace and the geometric multiplicity of $\lambda_1 = 4$ is $m_1 = 1$.

$\boldsymbol{\lambda_2 = 2:}$

$$(A - 2I)\boldsymbol{x}_2 = \mathbf{0}$$

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

$$\Downarrow$$

$$R_1 \to -\tfrac{1}{5}R_1 \quad \left[\begin{array}{ccc|c} 0 & 1 & -1 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & -1 & 1 & 0 \end{array}\right]$$

$$\Downarrow$$

$$\begin{array}{r} \\ R_2 \to R_2 - R_1 \\ R_3 \to R_3 + R_1 \end{array} \left[\begin{array}{ccc|c} 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

- $\boxed{x_1 = t}$
- $\boxed{x_3 = s}$
- $x_2 - x_3 = 0 \implies x_2 - s = 0 \implies \boxed{x_2 = s}$

Therefore $\boldsymbol{x}_2 = \begin{bmatrix} t \\ s \\ s \end{bmatrix} = t \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + s \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$, so $\{(1,0,0),(0,1,1)\}$ is a basis (check linear independence) for the $\lambda = 2$ eigenspace and the geometric multiplicity of $\lambda_2 = 2$ is $m_2 = 2$.

In this example the geometric multiplicity equalled the algebraic multiplicity for both eigenvalues.

8.3 Linear Independence of Eigenvectors

The following can be proved by induction.

Theorem 8-4: If $B = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_k\}$ is a set of eigenvectors of $n \times n$ matrix A corresponding to distinct eigenvalues (so $\lambda_i \neq \lambda_j$ for $i \neq j$) then B is linearly independent.

As a special case when all the eigenvalues of A are distinct we have the following.

Corollary: Let A be a $n \times n$ matrix. If A has n distinct eigenvalues, then A has a set of n linearly independent eigenvectors.

A more general result than the corollary is the following, proved similarly to the original theorem.

Theorem 8-5: Let A be an $n \times n$ matrix. If the geometric multiplicity m of each eigenvalue of A equals its algebraic multiplicity M then A has a set of n linearly independent eigenvectors.

Example 8-7

In Example 8-5 we found the matrix

$$A = \begin{bmatrix} 5 & -7 & 7 \\ 4 & -3 & 4 \\ 4 & -1 & 2 \end{bmatrix}$$

had distinct eigenvalues $\lambda_1 = 1$, $\lambda_2 = 5$, and $\lambda_3 = -1$ with corresponding eigenvectors $\boldsymbol{v}_1 = (0, 1, 1)$, $\boldsymbol{v}_2 = (1, 1, 1)$, and $\boldsymbol{v}_3 = (-1, 0, 1)$. Evaluating the following determinant along the first row

$$\begin{vmatrix} 0 & 1 & -1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{vmatrix} = 0 + 1(-1)(1-0) + (-1)(+1)(1-1) = -1 \neq 0$$

shows the set $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3\}$ is linearly independent as expected from the corollary.

8.4 Diagonalization

8.4.1 Properties of Diagonal Matrices

Recall that an $n \times n$ matrix $D = [d_{ij}]$ is called diagonal if every entry not on the main diagonal is zero, i.e. if $d_{ij} = 0$ whenever $i \neq j$. In general an $n \times n$ diagonal matrix has the form:

$$D = \begin{bmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_n \end{bmatrix}.$$

Note that some of the $d_1, d_2, \cdots, d_n$ values may be equal to zero.

Example 8-8

The following matrices are diagonal.

$$A = \begin{bmatrix} 2 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 5 \end{bmatrix} \qquad D = \begin{bmatrix} 4 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 5 \end{bmatrix}$$

Diagonal matrices have convenient properties.

Theorem 8-6: Let $D = \begin{bmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_n \end{bmatrix}$ and $W = \begin{bmatrix} w_1 & 0 & \cdots & 0 \\ 0 & w_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & w_n \end{bmatrix}$ be $n \times n$ diagonal matrices.

Then:

1. $DW = WD = \begin{bmatrix} d_1 w_1 & 0 & \cdots & 0 \\ 0 & d_2 w_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_n w_n \end{bmatrix}$
2. $\det D = (d_1)(d_2) \cdots (d_n)$
3. D is nonsingular (invertible) if and only if each entry on the main diagonal is nonzero ($d_i \neq 0$) in which case
$$D^{-1} = \begin{bmatrix} \frac{1}{d_1} & 0 & \cdots & 0 \\ 0 & \frac{1}{d_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{d_n} \end{bmatrix}.$$
4. The eigenvalues of D are its main diagonal entries: $d_1, d_2, \ldots, d_n$.

Example 8-9

Find the inverse of the given matrix.

1. $D = \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & -3 & 0 & 0 \\ 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix}$

   Solution:
   Since $d_i \neq 0$ for $i = 1, 2, 3, 4$ then D^{-1} exists with

$$D^{-1} = \begin{bmatrix} \frac{1}{2} & 0 & 0 & 0 \\ 0 & -\frac{1}{3} & 0 & 0 \\ 0 & 0 & \frac{1}{4} & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix}$$

2. $D = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 0 \end{bmatrix}$

   Solution:
   $d_3 = 0$ so D is noninvertible.

8.4.2 Diagonalizable Matrices

Most matrices are not diagonal, but some are related to diagonal matrices.

Definition: If A and B are $n \times n$ matrices then A is **similar** to B if there exits an invertible matrix P satisfying

$$\boxed{P^{-1}AP = B}.$$

A matrix A will be diagonalizable if it is similar to a diagonal matrix.

Definition: Let A be a $n \times n$ matrix. Then A is **diagonizable** if there exists an $n \times n$ invertible matrix P such that $P^{-1}AP$ is a diagonal matrix. When such a P exists we say that P **diagonalizes** A.

Theorem 8-7: (Conditions for diagonalizability)
Let A be an $n \times n$ matrix with eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$. Let $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n$ be eigenvectors of A with $\boldsymbol{v}_j$ associated with λ_j. Suppose that these eigenvectors are linearly independent and let P be the $n \times n$ matrix having $\boldsymbol{v}_j$ as its j^{th} column, so $\boxed{P = [\boldsymbol{v}_1 \boldsymbol{v}_2 \cdots \boldsymbol{v}_n]}$. Then P is nonsingular (invertible) and

$$\boxed{P^{-1}AP = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix} = D}$$

The diagonal matrix D has the eigenvalues of A along its main diagonal in the same order as the eigenvectors are listed as columns of P.

Remarks:

1. P^{-1} exists because $\det P \neq 0$ since the eigenvectors are linearly independent.
2. The eigenvalues of A need not be distinct. So $\lambda_1, \lambda_2, \ldots, \lambda_n$ in the theorem includes multiplicity.
3. The necessary condition for A to be diagonizable is that it have n linearly independent eigenvectors.
4. If A has n distinct eigenvalues or all the eigenvalues have equal algebraic and geometric multiplicity then A will automatically have n linearly independent eigenvectors and hence be diagonalizable.

Example 8-10

Diagonalize the given matrix if possible.

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

Solution:
From Example 8-4 matrix A had two distinct eigenvalues $\lambda_1 = 6$, $\lambda_2 = -1$ with corresponding (linearly independent) eigenvectors $\boldsymbol{v}_1 = (3, 4)$ and $\boldsymbol{v}_2 = (-1, 1)$. So A is diagonalizable by

$$P = [\boldsymbol{v}_1 \boldsymbol{v}_2] = \begin{bmatrix} 3 & -1 \\ 4 & 1 \end{bmatrix}.$$

Check:

$$\begin{aligned} P^{-1}AP &= \frac{1}{7}\begin{bmatrix} 1 & 1 \\ -4 & 3 \end{bmatrix}\begin{bmatrix} 2 & 3 \\ 4 & 3 \end{bmatrix}\begin{bmatrix} 3 & -1 \\ 4 & 1 \end{bmatrix} = \frac{1}{7}\begin{bmatrix} 1 & 1 \\ -4 & 3 \end{bmatrix}\begin{bmatrix} 18 & 1 \\ 24 & -1 \end{bmatrix} = \frac{1}{7}\begin{bmatrix} 42 & 0 \\ 0 & -7 \end{bmatrix} \\ &= \begin{bmatrix} 6 & 0 \\ 0 & -1 \end{bmatrix} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} = D \end{aligned}$$

2.
$$A = \begin{bmatrix} 5 & -7 & 7 \\ 4 & -3 & 4 \\ 4 & -1 & 2 \end{bmatrix}$$

Solution:
From Example 8-5 matrix A had three distinct eigenvalues $\lambda_1 = 1$, $\lambda_2 = 5$, and $\lambda_3 = -2$ with corresponding (linearly independent) eigenvectors $\boldsymbol{v}_1 = (0, 1, 1)$, $\boldsymbol{v}_2 = (1, 1, 1)$, and $\boldsymbol{v}_3 = (-1, 0, 1)$. So A is diagonalizable by

$$P = [\boldsymbol{v}_1 \boldsymbol{v}_2 \boldsymbol{v}_3] = \begin{bmatrix} 0 & 1 & -1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}.$$

Check:

$$P^{-1}AP = \begin{bmatrix} -1 & 2 & -1 \\ 1 & -1 & 1 \\ 0 & -1 & 1 \end{bmatrix}\begin{bmatrix} 5 & -7 & 7 \\ 4 & -3 & 4 \\ 4 & -1 & 2 \end{bmatrix}\begin{bmatrix} 0 & 1 & -1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & -2 \end{bmatrix} = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix} = D$$

3.
$$A = \begin{bmatrix} 2 & -5 & 5 \\ 0 & 3 & -1 \\ 0 & -1 & 3 \end{bmatrix}$$

Solution:
From Example 8-6 matrix A had two eigenvalues but each had algebraic multiplicity equalling geometric multiplicity. Then $\lambda_1 = 4$, $\lambda_2 = 2$, and $\lambda_3 = 2$ with corresponding (linearly independent) eigenvectors $\boldsymbol{v}_1 = (5, -1, 1)$, $\boldsymbol{v}_2 = (1, 0, 0)$, and $\boldsymbol{v}_3 = (0, 1, 1)$. So A is diagonalizable by

$$P = [\boldsymbol{v}_1 \boldsymbol{v}_2 \boldsymbol{v}_3] = \begin{bmatrix} 5 & 1 & 0 \\ -1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix}.$$

Check:

$$P^{-1}AP = \frac{1}{2}\begin{bmatrix} 0 & -1 & 1 \\ 2 & 5 & -5 \\ 0 & 1 & 1 \end{bmatrix}\begin{bmatrix} 2 & -5 & 5 \\ 0 & 3 & -1 \\ 0 & -1 & 3 \end{bmatrix}\begin{bmatrix} 5 & 1 & 0 \\ -1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix} = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix} = D$$

Not every matrix is diagonalizable as shown in the next example.

Example 8-11

Diagonalize the given matrix if possible.

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

Solution:

$$\begin{aligned} \det(A - \lambda I) &= 0 \\ &\Rightarrow \begin{vmatrix} -2-\lambda & 0 & 1 \\ 1 & 1-\lambda & 0 \\ 0 & 0 & -2-\lambda \end{vmatrix} = 0 \\ &\Rightarrow (-2-\lambda)(1-\lambda)(-2-\lambda) = 0 \\ &\Rightarrow (1-\lambda)(\lambda+2)^2 = 0 \\ &\Rightarrow \begin{cases} \lambda_1 = 1, \ M_1 = 1 \\ \lambda_2 = -2, \ M_2 = 2 \end{cases} \end{aligned}$$

$\boldsymbol{\lambda_1 = 1}$**:**

$$(A - (1)I)\boldsymbol{x}_1 = \mathbf{0}$$

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

$$\Downarrow$$

$$R_1 \leftrightarrow R_2 \quad \left[\begin{array}{ccc|c} 1 & 0 & 0 & 0 \\ -3 & 0 & 1 & 0 \\ 0 & 0 & -3 & 0 \end{array}\right]$$

$$\Downarrow$$

$$R_2 \to R_2 + 3R_1 \left[\begin{array}{ccc|c} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & -3 & 0 \end{array}\right]$$

$$\Downarrow$$

$$R_3 \to R_3 + 3R_2 \left[\begin{array}{ccc|c} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

- $\boxed{x_2 = t}$
- $\boxed{x_3 = 0}$
- $\boxed{x_1 = 0}$

Therefore $\boldsymbol{x}_1 = \begin{bmatrix} 0 \\ t \\ 0 \end{bmatrix} = t \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$ and $\{(0,1,0)\}$ is a basis for the $\lambda = 1$ eigenspace.

$\boldsymbol{\lambda_2 = -2}$:

$$(A - (-2)I)\boldsymbol{x}_2 = \mathbf{0}$$
$$(A + 2I)\boldsymbol{x}_2 = \mathbf{0}$$

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

$$\Downarrow$$

$$R_1 \leftrightarrow R_2 \left[\begin{array}{ccc|c} 1 & 3 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

- $\boxed{x_2 = t}$
- $\boxed{x_3 = 0}$
- $x_1 + 3x_2 = 0 \implies x_1 + 3t = 0 \implies \boxed{x_1 = -3t}$

Therefore $\boldsymbol{x}_2 = \begin{bmatrix} -3t \\ t \\ 0 \end{bmatrix} = t \begin{bmatrix} -3 \\ 1 \\ 0 \end{bmatrix}$ and $\{(-3,1,0)\}$ is a basis for the $\lambda = -2$ eigenspace.

So geometric multiplicity is $m_2 = 1$. Since $m_2 = 1 < M_2 = 2$, A is not diagonalizable.

8.4.3 Applications of Diagonalization

If an $n \times n$ matrix A is diagonalizable, then:

$$P^{-1}AP = D$$

where D is a diagonal matrix. We can solve for A in terms of D by premultiplying by P and postmultiplying by P^{-1} on both sides to get:

$$\begin{aligned} P^{-1}AP &= D \\ P(P^{-1}AP)P^{-1} &= PDP^{-1} \\ (PP^{-1})A(PP^{-1}) &= PDP^{-1} \\ IAI &= PDP^{-1} \\ A &= PDP^{-1} \end{aligned}$$

Recalling that the m^{th} power of square matrix A is just the product of A multiplied m times we have

$$\begin{aligned} A^m &= \underbrace{AA\cdots A}_{m \text{ times}} \\ &= (PDP^{-1})(PDP^{-1})\cdots(PDP^{-1}) \\ &= PD(P^{-1}P)D(P^{-1}P)\cdots(P^{-1}P)DP^{-1} \\ &= PDIDI\cdots DP^{-1} \\ &= P\,\underbrace{DD\cdots D}_{m \text{ times}}\,P^{-1} \\ &= PD^mP^{-1} \end{aligned}$$

where

$$D^m = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}^m = \begin{bmatrix} \lambda_1^m & 0 & \cdots & 0 \\ 0 & \lambda_2^m & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n^m \end{bmatrix}.$$

We observe that the same P that diagonalizes A also diagonalizes A^m and furthermore that the eigenvalues of A^m appearing in its diagonal matrix are λ_i^m. This is the case since if $\boldsymbol{x}$ is an eigenvector of A corresponding to eigenvalue λ we have that $\boldsymbol{x}$ is also an eigenvector of A^m corresponding to eigenvalue λ^m since

$$A^m\boldsymbol{x} = A^{m-1}A\boldsymbol{x} = A^{m-1}\lambda\boldsymbol{x} = \lambda A^{m-1}\boldsymbol{x} = \ldots = \lambda^m\boldsymbol{x}\,.$$

More generally yet, the matrix P will diagonalize any polynomial function of diagonalizable matrix A, $p(A)$. It will have the same eigenvectors as A and its eigenvalues will be $p(\lambda_i)$.

A direct application of the result $A^m = PD^mP^{-1}$ is the simplicity of taking large powers of a diagonalizable matrix.

Example 8-12

Find A^{10} if $A = \begin{bmatrix} 3 & -2 \\ -4 & 1 \end{bmatrix}$.

Solution:

$$\begin{aligned} \det(A - \lambda I) &= 0 \\ \begin{vmatrix} 3-\lambda & -2 \\ -4 & 1-\lambda \end{vmatrix} &= 0 \\ (3-\lambda)(1-\lambda) - 8 &= 0 \\ 3 - 3\lambda - \lambda + \lambda^2 - 8 &= 0 \\ \lambda^2 - 4\lambda - 5 &= 0 \\ (\lambda - 5)(\lambda + 1) &= 0 \\ \begin{cases} \lambda_1 &= 5 \\ \lambda_2 &= -1 \end{cases} \end{aligned}$$

$\boldsymbol{\lambda_1 = 5}$**:**

$$(A - 5I)\boldsymbol{x}_1 = \mathbf{0}$$

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

$$\Downarrow$$

$$\begin{array}{r} R_1 \to -\frac{1}{2}R_1 \\ R_2 \to R_2 - 2R_1 \end{array} \left[\begin{array}{cc|c} 1 & 1 & 0 \\ 0 & 0 & 0 \end{array}\right]$$

- $\boxed{x_2 = t}$
- $x_1 + x_2 = 0 \implies x_1 + t = 0 \implies \boxed{x_1 = -t}$

Therefore $\boldsymbol{x}_1 = \begin{bmatrix} -t \\ t \end{bmatrix} = t \begin{bmatrix} -1 \\ 1 \end{bmatrix}$ so $\{\boldsymbol{v}_1\} = \{(-1, 1)\}$ is a basis for the $\lambda = 5$ eigenspace.

$\boldsymbol{\lambda_2 = -1}$**:**

$$(A + I)\boldsymbol{x}_2 = \mathbf{0}$$

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

$$\Downarrow$$

$$\begin{array}{r} R_1 \to \frac{1}{4}R_1 \\ R_2 \to R_2 + R_1 \end{array} \left[\begin{array}{cc|c} 1 & -\frac{1}{2} & 0 \\ 0 & 0 & 0 \end{array}\right]$$

- $\boxed{x_2 = t}$
- $x_1 - \frac{1}{2}x_2 = 0 \implies x_1 - \frac{1}{2}t = 0 \implies \boxed{x_1 = \dfrac{1}{2}t}$

Therefore $\boldsymbol{x}_2 = \begin{bmatrix} \frac{1}{2}t \\ t \end{bmatrix} = \frac{1}{2}t \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ so $\{\boldsymbol{v}_2\} = \{(1, 2)\}$ is a basis for the $\lambda = -1$ eigenspace.

Then $P = [\boldsymbol{v}_1 \boldsymbol{v}_2] = \begin{bmatrix} -1 & 1 \\ 1 & 2 \end{bmatrix}$ and $D = \begin{bmatrix} 5 & 0 \\ 0 & -1 \end{bmatrix}$.

We find $P^{-1} = \frac{1}{3} \begin{bmatrix} -2 & 1 \\ 1 & 1 \end{bmatrix}$.

Finally we have

$$\begin{aligned} A^m &= PD^m P^{-1} \\ A^{10} &= PD^{10} P^{-1} \\ &= \begin{bmatrix} -1 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 5^{10} & 0 \\ 0 & (-1)^{10} \end{bmatrix} \frac{1}{3} \begin{bmatrix} -2 & 1 \\ 1 & 1 \end{bmatrix} \\ &= \frac{1}{3} \begin{bmatrix} -(5)^{10} & 1 \\ 5^{10} & 2 \end{bmatrix} \begin{bmatrix} -2 & 1 \\ 1 & 1 \end{bmatrix} \\ &= \frac{1}{3} \begin{bmatrix} 2(5^{10}) + 1 & -5^{10} + 1 \\ -2(5^{10}) + 2 & 5^{10} + 2 \end{bmatrix} \end{aligned}$$

8.5 Properties of Eigenvalues

Definition: Let $A = [a_{ij}]$ be an $n \times n$ matrix. Then the **trace of matrix $\boldsymbol{A}$**, denoted by $\mathbf{tr}\,(\boldsymbol{A})$ is the sum of the diagonal entries of A,

$$\boxed{\operatorname{tr}(A) = a_{11} + a_{22} + \ldots + a_{nn}}.$$

Example 8-13

Find the trace of $A = \begin{bmatrix} 5 & 4 & -3 \\ 2 & 1 & 0 \\ -1 & 4 & -2 \end{bmatrix}$

Solution:

$$\operatorname{tr}(A) = 5 + 1 + (-2) = 4$$

Theorem 8-8: Let $A_1,\ A_2,\ \ldots A_k$ be square matrices of order n. Then the trace of the product, $\operatorname{tr}(A_1 A_2 \cdots A_k)$ is invariant under cyclic permutations,

$$\operatorname{tr}(A_1 A_2 \cdots A_k) = \operatorname{tr}(A_2 A_3 \cdots A_k A_1) = \operatorname{tr}(A_3 A_4 \cdots A_k A_1 A_2) = \ldots$$

Theorem 8-9: Let A be an $n \times n$ matrix with eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$, where here we are allowing for multiplicity. Then:

1. $\det A = \lambda_1 \lambda_2 \cdots \lambda_n$
2. $\operatorname{tr}(A) = \lambda_1 + \lambda_2 + \cdots + \lambda_n$

Proof:
We prove the theorem in the event that matrix A is diagonalizable. It is more generally true.[1] Let A be diagonalized by matrix P so $P^{-1}AP = D$ where D is diagonal with eigenvalues λ_i on the diagonal. Then $A = PDP^{-1}$ and

1. $\det A = \det\left(PDP^{-1}\right) = \det\left(PP^{-1}D\right) = \det(ID) = \det D = \lambda_1 \lambda_2 \cdots \lambda_n$
2. $\operatorname{tr}(A) = \operatorname{tr}\left(PDP^{-1}\right) = \operatorname{tr}\left(DP^{-1}P\right) = \operatorname{tr}(DI) = \operatorname{tr}(D) = \lambda_1 + \lambda_2 + \cdots + \lambda_n$

where here we used that we can commute matrices under a determinant as well as the invariance of the trace under cyclic permutation of the matrices.

Example 8-14

Verify the properties of Theorem 8-9 for the following matrix.

$$A = \begin{bmatrix} 2 & -5 & 5 \\ 0 & 3 & -1 \\ 0 & -1 & 3 \end{bmatrix}$$

[1] In the more general case it can be shown that any square matrix A is similar to an upper triangular matrix with the eigenvalues of A along the diagonal. This is called the **Jordan normal form** of A. Because the latter matrix has the same determinant and trace as the special case diagonal matrix, the proof above is essentially the same.

Solution:
From Example 8-6 we had the following eigenvalues with the given algebraic multiplicities.

$$\lambda_1 = 4,\ M_1 = 1$$
$$\lambda_2 = 2,\ M_2 = 2\,.$$

So $\lambda_1 = 4$, $\lambda_2 = 2$, $\lambda_3 = 2$ allowing for multiplicities.

Then using Theorem 8-9

$$\det A = \lambda_1\lambda_2\lambda_3 = (4)(2)(2) = 16$$
$$\operatorname{tr}(A) = \lambda_1 + \lambda_2 + \lambda_3 = 4 + 2 + 2 = 8\,.$$

Direct calculation using the matrix A gives

$$\det A = 2(+1)(9-1) + 0 + 0 = 16$$
$$\operatorname{tr}(A) = 2 + 3 + 3 = 8\,.$$

in agreement with the theorem.

Theorem 8-9 can be used in reverse to find eigenvalues in limited cases.

Example 8-15

Find the eigenvalues of a 2×2 matrix A with $\det A = -4$ and $\operatorname{tr}(A) = -3$.

Solution:
Let λ_1 and λ_2 be the eigenvalues of the matrix A.
We know that:

$$\lambda_1 + \lambda_2 = \operatorname{tr}(A) = -3$$
$$\lambda_1\lambda_2 = \det A = -4$$

Therefore

$$\begin{aligned} \lambda_2 &= -3 - \lambda_1 \\ \Rightarrow \lambda_1(-3 - \lambda_1) &= -4 \\ -3\lambda_1 - \lambda_1^2 &= -4 \\ \lambda_1^2 + 3\lambda_1 - 4 &= 0 \\ (\lambda_1 + 4)(\lambda_1 - 1) &= 0\,. \end{aligned}$$

So $\lambda_1 = -4$ (in which case $\lambda_2 = -3 - (-4) = 1$) or $\lambda_1 = 1$ (in which case $\lambda_2 = -3 - 1 = -4$).
Therefore the eigenvalues of A are $\lambda = -4$ and $\lambda = 1$.

8.6 Interpreting Eigenvalues and Eigenvectors

We have seen in Section 6.4 that a linear operator $L(\boldsymbol{x})$ on $\mathbb{R}^n$ has a square matrix representation so that $L(\boldsymbol{x}) = L_A(\boldsymbol{x}) = A\boldsymbol{x}$. We can cast the discussion of eigenvalues and eigenvectors into linear operator terminology as follows.

Definition: Let L be a linear operator on $\mathbb{R}^n$. A scalar λ is said to be an **eigenvalue** of L if there exists a nonzero vector $\boldsymbol{x}$ in $\mathbb{R}^n$ such that:

$$\boxed{L(\boldsymbol{x}) = \lambda\boldsymbol{x}}\,.$$

The nonzero vector $\boldsymbol{x}$ is called an **eigenvector** of L corresponding to the eigenvalue λ.

Now consider a linear operator such as $S_{\boldsymbol{i},1/2}$ on $\mathbb{R}^3$. This is a compression by a factor of one half along the direction of the x-axis. What can we say about the eigenvalues and eigenvectors of this operator? The eigenvalue problem geometrically asks us to find those nonzero vectors (eigenvectors) that, when operated upon by L, return a scalar multiple (the eigenvalue) times themselves. Most vectors in $\mathbb{R}^3$ will not be eigenvectors of $S_{\boldsymbol{i},1/2}$ since their x-components will be compressed by a factor $1/2$ while their orthogonal components will remain unchanged. However if one considers a vector that lies only along the x-axis, such as $\boldsymbol{u} = u_1\boldsymbol{i}$ then $S_{\boldsymbol{i},1/2}(\boldsymbol{u}) = \frac{1}{2}u_1\boldsymbol{i} = \frac{1}{2}\boldsymbol{u}$. We thus conclude that such vectors are eigenvectors of $S_{\boldsymbol{i},1/2}$ with eigenvalue $\lambda = 1/2$. Indeed the eigenspace corresponding to $\lambda = 1/2$ will be the vectors directed along the x-axis (i.e. the line) with basis $\{\boldsymbol{i}\}$.

Are there other eigenvalues and eigenvectors of $S_{\boldsymbol{i},1/2}$? Suppose a vector has zero component along the x-axis, so $\boldsymbol{u} = u_2\boldsymbol{j} + u_3\boldsymbol{k}$. Then this vector would be its own orthogonal projection and it would be unaffected by the compression. i.e. $S_{\boldsymbol{i},1/2}(\boldsymbol{u}) = u_2\boldsymbol{j} + u_3\boldsymbol{k} = \boldsymbol{u} = 1\boldsymbol{u}$. In other words such a vector will be an eigenvector corresponding to the eigenvalue $\lambda = 1$. The eigenspace corresponding to $\lambda = 1$ will just be these vectors lying in the y-z plane, a basis of which is just $\{\boldsymbol{j}, \boldsymbol{k}\}$. As such we see that the eigenvalues and eigenvectors of a linear operator have important physical interpretations. Moreover, cast in this linear operator form, we realize that, quite generally, they will be independent of choice of coordinates. Any compression $S_{\boldsymbol{n},k}$ will have an eigenvalue $\lambda = k$ corresponding to eigenspace with basis $\{\boldsymbol{n}\}$ and eigenvalue $\lambda = 1$ corresponding to the subspace of dimension $n-1$ of vectors in $\mathbb{R}^n$ orthogonal to $\boldsymbol{n}$.

Example 8-16

The matrix transformation L_A on $\mathbb{R}^3$ has

$$A = \frac{1}{5}\begin{bmatrix} 14 & 12 & 0 \\ 12 & 21 & 0 \\ 0 & 0 & 5 \end{bmatrix}\,.$$

If A has eigenvalue $\lambda_1 = 1$ with eigenspace basis $\{(4, -3, 0), (0, 0, 1)\}$ and eigenvalue $\lambda_2 = 6$ with eigenspace basis $\{(3, 4, 0)\}$, describe the linear operator L_A.

Solution:
Vectors parallel to $(3, 4, 0)$ are scaled by a factor of $\lambda_2 = 6$. The eigenvector $(3, 4, 0)$ is orthogonal to vectors in the eigenspace of $\lambda_1 = 1$ since $(3, 4, 0) \cdot (4, -3, 0) = 0$ and $(3, 4, 0) \cdot (0, 0, 1) = 0$. Vectors in the eigenspace of λ_1 remain unchanged by L_A since the eigenvalue is 1. We conclude that L_A is an expansion by factor $k = 6$ along the direction of $(3, 4, 0)$. A unit vector in that direction is

$$\boldsymbol{n} = \frac{1}{\sqrt{3^2 + 4^2 + 0^2}}(3, 4, 0) = (3/5, 4/5, 0)\,.$$

Thus $L_A = S_{\boldsymbol{n},6}$.

Example 8-17

What eigenvalues and eigenvectors does the rotation operator $R_{\boldsymbol{n},\theta}$ in $\mathbb{R}^3$ have?

Solution:
The only vector that remains unchanged, short of becoming a scalar multiple of itself, is a vector directed along the axis of rotation, namely $\boldsymbol{u} = u\boldsymbol{n}$. For such a vector we have

$$R_{\boldsymbol{n},\theta}(\boldsymbol{u}) = u\boldsymbol{n} = \boldsymbol{u} = 1\boldsymbol{u}\,.$$

It follows that $\lambda = 1$ is the eigenvalue corresponding to this eigenvector and that the eigenspace corresponding to $\lambda = 1$ is the line directed along the axis of rotation with basis $\{\boldsymbol{n}\}$. A rotation matrix has additional complex-valued eigenvalues and eigenvectors. While the latter do not represent physical vectors as they are not real-valued, they nevertheless provide useful information about the rotation as one may determine the plane of rotation from them. The complex eigenvalues similarly determine the angle of rotation. Complex eigenvalues and eigenvectors will be discussed in Section 9.6.

We note that for matrices arising from linear operators in physical problems, the eigenvalues and corresponding eigenvectors typically have physical meaning as in our geometrical examples. In $\mathbb{R}^3$ two observers working with the same linear operator but potentially in coordinate systems rotated with respect to each other would find the same eigenvalues for the operator despite the fact their matrix representations would be different. Similarly they would find that the corresponding eigenvectors despite having different coordinates in their respective systems would represent the same vectors in physical space. Due to this coordinate system independence, eigenvalues of operators arising from physical problems can represent observable properties and we expect them to show up in analysis of such problems. Since by Theorem 8-9 the determinant and trace of a matrix depends only on its eigenvalues, these too are properties of the operator that are independent of the matrix representation used by each observer and thus typically have a physical meaning.

8.7 Interpreting Diagonalization

Continuing the discussion from the previous section we saw that the compression $S_{\boldsymbol{i},1/2}$ had eigenvalue $\lambda_1 = 1/2$ with corresponding eigenvector $\boldsymbol{i}$ and eigenvalue $\lambda_2 = 1$ with corresponding eigenvectors $\boldsymbol{j}$ and $\boldsymbol{k}$. Since $S_{\boldsymbol{i},1/2}(\boldsymbol{i}) = \frac{1}{2}\boldsymbol{i}$ and $S_{\boldsymbol{i},1/2}(\boldsymbol{j}) = \boldsymbol{j}$ and $S_{\boldsymbol{i},1/2}(\boldsymbol{k}) = \boldsymbol{k}$ it follows that the operator equals L_A where

$$A = \begin{bmatrix} \frac{1}{2} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix}.$$

If we had the general operator $S_{\boldsymbol{n},k}$ and we chose our coordinate axes so that $\boldsymbol{n} = \boldsymbol{i}$ then the matrix A would also be diagonal. For a general choice of axes, as we saw in Section 6.4.1, the matrix representation of $S_{\boldsymbol{n},k}$ would be symmetric but not necessarily diagonal.

So if we are working with a linear operator L represented by matrix A in $\mathbb{R}^3$ (or more generally $\mathbb{R}^n$) one might wonder if there is some set of coordinate axes that we might have chosen in which the operator's matrix representation would be diagonal. Questions of coordinates in physical space imply a notion of distance which, in $\mathbb{R}^n$, implies the use of an inner product which we take to be our usual Euclidean dot product. In that case the answer to the question is that such a system of coordinate axes exists if the matrix is orthogonally diagonalizable.

Definition: Square matrix A is **orthogonally diagonalizable** if there exists an orthogonal matrix P such that

$$P^{-1}AP = D\,,$$

where D is a diagonal matrix.

Recall that $P^{-1} = P^T$ for an orthogonal matrix so $P^TAP = D$. The orthogonal matrix P provides the information required to relate the coordinate axes in which the matrix representation is diagonal to the original axes. As an orthogonal matrix it will be composed of reflections, rotations, and inversions. The following theorem characterizes all orthogonally diagonalizable matrices.

Theorem 8-10: An $n \times n$ matrix A is orthogonally diagonalizable if and only if it is symmetric.

Many useful operators on $\mathbb{R}^n$ are symmetric as we have seen and therefore are orthogonally diagonalizable. Working in a coordinate system in which the operator is diagonal often simplifies calculations. Such coordinate system transformations requires a more general discussion of vector spaces and their bases than will be provided at this juncture.[2]

[2] We note that the more general criteria of a matrix being merely diagonalizable only guarantees that new coordinate axes can be chosen in which the matrix representation is diagonal but these axes are no longer necessarily mutually orthogonal. Lengths of vectors transformed into those coordinates would not, in general, be preserved.

Chapter 9: Complex Numbers

9.1 Origin of Complex Numbers

The idea of a complex number arose out of the problem of finding solutions to the equation

$$p(z) = 0$$

where $p(z) = a_n z^n + a_{n-1} z^{n-1} + \cdots + a_2 z^2 + a_1 z + a_0$ is a polynomial of order n (so $a_n \neq 0$). The solutions of this equation, known as the roots or zeros of the polynomial $p(z)$, are useful for factoring and vice versa. If c is a solution of this equation (so $p(c) = 0$) this implies $(z - c)$ is a factor of $p(z)$ and we can write $p(z) = (z - c)q(z)$ where $q(z)$ is a simpler polynomial of order $n - 1$. However not every polynomial equation has a real solution. So a quadratic equation like

$$z^2 - 2z + 5 = 0$$

has no solution since, using the quadratic formula,

$$z = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} = \frac{2 \pm \sqrt{4 - 20}}{2} = 1 \pm \frac{1}{2}\sqrt{-16}$$

and no real number squares to get -16. Assuming square root obeys the usual properties, namely $\sqrt{rs} = \sqrt{r}\sqrt{s}$, we could write $\sqrt{-16} = \sqrt{16}\sqrt{-1} = 4\sqrt{-1}$ and the (non-real) solutions simplify to

$$z = 1 \pm 4\sqrt{-1}\,.$$

Thus the lack of solution for all similar equations reduces to the lack of a square root to -1. While solutions involving $\sqrt{-1}$ cannot represent anything physical, it was found, when seeking a general formula for the root of a cubic equation

$$az^3 + bz^2 + cz + d = 0\,,$$

that if one pretended that $\sqrt{-1}$ behaved like any other number it worked, as an intermediary, to finding real-valued solutions to the cubic equation. In this way $\sqrt{-1}$ had bookkeeping utility in generating actual (real) solutions to problems. With the development of mathematics axiomatically and the proven utility of treating $\sqrt{-1}$ as a number, complex numbers were born.[1]

Definition: A number of the form $z = x + iy$ where x and y are real numbers and $i = \sqrt{-1}$ is called a **complex number**. Here

- x is called the **real part** of z and is denoted by $x = \mathrm{Re}(z)$
- y is called the **imaginary part** of z and is denoted by $y = \mathrm{Im}(z)$

Example 9-1

The solutions to the previous quadratic are the complex numbers

$$z_1 = 1 + 4i, \qquad z_2 = 1 - 4i\,,$$

with real and imaginary parts

$$\mathrm{Re}(z_1) = 1 \qquad \mathrm{Im}(z_1) = 4 \qquad \mathrm{Re}(z_2) = 1 \qquad \mathrm{Im}(z_2) = -4\,.$$

[1] It is to be noted that complex numbers are now at the heart of our physical theories. In quantum mechanics, the physical theory describing atoms and other microscopic phenomenon, the wave function Ψ is a *complex* scalar field. One manipulates this field of complex numbers to extract real numbers that describe actual physical measurements.

9.2 Complex Conjugate

Every complex number z has a complex conjugate.

Definition: The **complex conjugate** of complex number $z = x + iy$ is denoted $\overline{z}$ and is given by

$$\boxed{\overline{z} = x - iy}\,.$$

Example 9-2

Find the complex conjugates of

1. $z = 4 + 3i$
2. $z = -3 + 5i$
3. $z = 2i$

Solution:

1. $z = 4 + 3i \implies \overline{z} = 4 - 3i$
2. $z = -3 + 5i \implies \overline{z} = -3 - 5i$
3. $z = 2i \implies \overline{z} = -2i$

9.3 Operations on Complex Numbers

Let $z_1 = x_1 + iy_1$ and $z_2 = x_2 + iy_2$ be complex numbers. For addition and subtraction the i behaves just like a variable or constant:

Addition: $\boxed{z_1 + z_2 = (x_1 + x_2) + i(y_1 + y_2)}$

Subtraction: $\boxed{z_1 - z_2 = (x_1 - x_2) + i(y_1 - y_2)}$

For multiplication of two complex numbers one remembers that $i = \sqrt{-1}$ simplifies powers of i:

Example 9-3

1. $i = \sqrt{-1}$
2. $i^2 = -1$ (definition of square root)
3. $i^3 = i^2 i = -i$
4. $i^4 = i^2 i^2 = (-1)(-1) = 1$

With this in mind we have

Multiplication: $\boxed{z_1 z_2 = (x_1 x_2 - y_1 y_2) + i(x_1 y_2 + x_2 y_1)}$

Since

$$\begin{aligned} z_1 z_2 &= (x_1 + iy_1)(x_2 + iy_2) \\ &= x_1 x_2 + ix_1 y_2 + iy_1 x_2 + i^2 y_1 y_2 \\ &= x_1 x_2 + i(x_1 y_2 + y_1 x_2) - y_1 y_2 \\ &= (x_1 x_2 - y_1 y_2) + i(x_1 y_2 + y_1 x_2)\,. \end{aligned}$$

Note that complex multiplication is commutative, $\boxed{z_1 z_2 = z_2 z_1}$ just as for real numbers.

Theorem 9-1: The complex conjugate of a product is the product of the complex conjugates,

$$\boxed{\overline{z_1 z_2} = \overline{z_1}\,\overline{z_2}}\,.$$

Evaluation of the quotient of two complex numbers, z_1/z_2 can be resolved by multiplying the fraction by $1 = \overline{z_2}/\overline{z_2}$ as shown below.

Division: $\boxed{\dfrac{z_1}{z_2} = \dfrac{x_1 x_2 + y_1 y_2}{x_2^2 + y_2^2} + i\dfrac{x_2 y_1 - x_1 y_2}{x_2^2 + y_2^2}}$

Since:

$$\begin{aligned}\frac{z_1}{z_2} &= \frac{x_1+iy_1}{x_2+iy_2}\\ &= \frac{(x_1+iy_1)(x_2-iy_2)}{(x_2+iy_2)(x_2-iy_2)}\\ &= \frac{x_1x_2-ix_1y_2+ix_2y_1-i^2y_1y_2}{x_2^2-ix_2y_2+ix_2y_2-i^2y_2^2}\\ &= \frac{x_1x_2+i(x_2y_1-x_1y_2)+y_1y_2}{x_2^2+y_2^2}\\ &= \frac{(x_1x_2+y_1y_2)+i(x_2y_1-x_1y_2)}{x_2^2+y_2^2}\\ &= \frac{x_1x_2+y_1y_2}{x_2^2+y_2^2}+i\frac{x_2y_1-x_1y_2}{x_2^2+y_2^2}\,.\end{aligned}$$

Note that the form of the quotient shows every complex number z has a multiplicative inverse

$$\boxed{z^{-1}=\frac{1}{z}=\frac{\overline{z}}{z\overline{z}}=\frac{x}{x^2+y^2}-i\frac{y}{x^2+y^2}}$$

provided $z=x+iy\neq 0$ since then $x^2+y^2\neq 0$, just like for real numbers.[2]

With this discussion in mind complex arithmetic is readily performed. Note that the above formulas for multiplication and division need not be memorized. One need only remember that $i^2=-1$ for multiplication and to multiply by $1=\overline{z}/\overline{z}$ to evaluate division by z.

Example 9-4

If $z_1=3-2i$ and $z_2=2+i$, evaluate:

1. z_1+z_2
2. $z_1-\overline{z_2}$
3. z_1z_2
4. $\overline{z_1}\,\overline{z_2}$
5. $\dfrac{z_1}{z_2}$

Solution:

1. $z_1+z_2=(3-2i)+(2+i)=5-i$
2. $z_1-\overline{z_2}=(3-2i)-(2-i)=1-i$
3. $z_1z_2=(3-2i)(2+i)=6+3i-4i-2i^2=6-i-2(-1)=6-i-2=8-i$
4. $\overline{z_1}\,\overline{z_2}=(3+2i)(2-i)=6-3i+4i-2i^2=6+i-2(-1)=6+i+2=8+i=\overline{z_1z_2}$
5. $\dfrac{z_1}{z_2}=\dfrac{3-2i}{2+i}=\dfrac{(3-2i)(2-i)}{(2+i)(2-i)}=\dfrac{6-3i-4i+2i^2}{4-2i+2i-i^2}=\dfrac{6-7i+2(-1)}{4-(-1)}=\dfrac{6-7i-2}{4+1}$
   $=\dfrac{4-7i}{5}=\dfrac{4}{5}-\dfrac{7}{5}i$

[2]Note this, in part, explains why we introduced i such that $i^2=-1$ as opposed to something else. If $i^2=1$ then there would be nonzero numbers having no multiplicative inverse.

9.4 Solving Complex Equations

Equations involving complex variables and numbers can be solved using the usual algebraic manipulations.

Example 9-5

Solve the following equations for z.

1. $2z + 3 + 2i = (2+i)^2$
2. $3z + i - 2(2z - i) = 2 - i$
3. $\dfrac{1}{z} = 2 + 3i$
4. $2z - iz = 2 + i$
5. $z^2 = 4i$

Solution:

1. $2z + 3 + 2i = (2+i)^2$

$$\begin{aligned}
2z &= (2+i)^2 - 3 - 2i \\
2z &= 4 + 4i + i^2 - 3 - 2i \\
2z &= 1 + 2i + (-1) \\
2z &= 2i \\
z &= i
\end{aligned}$$

2. $3z + i - 2(2z - i) = 2 - i$

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

3. $2z - iz = 2 + i$

$$\begin{aligned}
(2-i)z &= 2 + i \\
z &= \frac{2+i}{2-i} \\
z &= \frac{(2+i)(2+i)}{(2-i)(2+i)} \\
z &= \frac{4 + 4i + i^2}{4 + 2i - 2i - i^2} \\
z &= \frac{4 + 4i - 1}{4 + 1} \\
z &= \frac{3 + 4i}{5} \\
z &= \frac{3}{5} + \frac{4}{5}i
\end{aligned}$$

4. $\dfrac{1}{z} = 2 + 3i$

$$\begin{aligned}
z &= \frac{1}{2+3i} \\
z &= \frac{2-3i}{(2+3i)(2-3i)} \\
z &= \frac{2-3i}{4-6i+6i-9i^2} \\
z &= \frac{2-3i}{4-9(-1)} \\
z &= \frac{2-3i}{13} \\
z &= \frac{2}{13} - \frac{3}{13}i
\end{aligned}$$

5. $z^2 = 4i$
   Write $z = x + iy$ where x and y are real variables. Then

$$\begin{aligned}
z^2 &= 4i \\
(x+iy)^2 &= 4i \\
x^2 + 2ixy + i^2y^2 &= 4i \\
x^2 + i(2xy) - y^2 &= 4i \\
(x^2 - y^2) + i(2xy) &= 4i \\
\begin{cases} x^2 - y^2 = 0 \\ 2xy = 4 \end{cases} &\Rightarrow \begin{cases} x^2 - y^2 = 0 \\ xy = 2 \end{cases} \\
x^2 - y^2 &= 0 \\
x^2 &= y^2 \\
x &= \pm y
\end{aligned}$$

   If $x = -y$, then substitution into $xy = 2$ gives

$$\begin{aligned}
(-y)(y) &= 2 \\
-y^2 &= 2 \\
y^2 &= -2
\end{aligned}$$

   No solution since y is a real number.
   If $x = y$, then

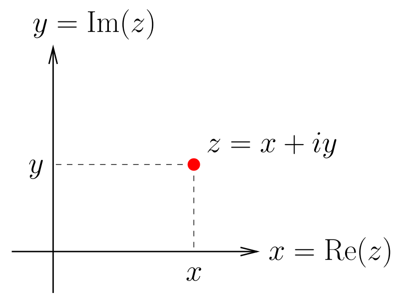
$$\begin{aligned}
(y)(y) &= 2 \\
y^2 &= 2 \\
y &= \pm\sqrt{2}
\end{aligned}$$

   Since $x = y$, then solutions are $z = \sqrt{2} + i\sqrt{2}$ or $z = -\sqrt{2} - i\sqrt{2}$.

As with any equations the solutions can be checked in the original equation.

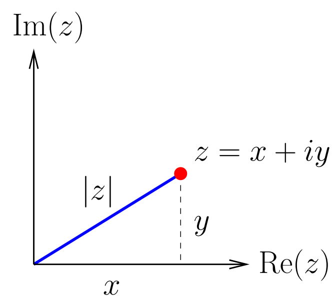
9.5 The Complex Plane

Since a complex number $z = x + iy$ is composed of two real numbers x and y constituting its real and imaginary parts, we can consider z to be represented by the ordered pair (x, y). With that in mind we consider complex numbers to live in the **complex plane**.



A real number x, which is a special case of a complex number z where $\mathrm{Im}(z) = 0$, lives on the real axis line of the complex plane.

The distance from the origin of the complex plane to the point z is its magnitude.



Definition: The **magnitude** or **absolute value** or **modulus** of a complex number $z = x + iy$, denoted by $|z|$, is defined to be

$$|z| = \sqrt{x^2 + y^2}\,.$$

Note that when z is real then $|z| = |x + 0i| = \sqrt{x^2} = |x|$ reduces to the real absolute value.

Example 9-6

Find the magnitude of the given complex number.

1. $z = 4 + 2i$

$$\begin{aligned} x &= 4\ ,\ \ y = 2 \\ |z| &= \sqrt{4^2 + 2^2} = \sqrt{20} \\ |z| &= 2\sqrt{5} \end{aligned}$$

2. $z = -5 + 3i$

$$\begin{aligned} x &= -5\ ,\ \ y = 3 \\ |z| &= \sqrt{(-5)^2 + 3^2} \\ |z| &= \sqrt{34} \end{aligned}$$

3. $z = -2$

$$x = -2\ ,\ \ y = 0$$
$$|z| = \sqrt{(-2)^2 + 0^2}$$
$$|z| = \sqrt{4} = 2$$

Direct calculation shows the following.

Theorem 9-2: If z is a complex number then $|z|^2 = z\overline{z}$ and $|\overline{z}| = |z|$.

Interpreting complex numbers as points in the plane we have the following.

Definition: The **distance** between two complex numbers $z_1 = x_1 + iy_1$ and $z_2 = x_2 + iy_2$ is given by:

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

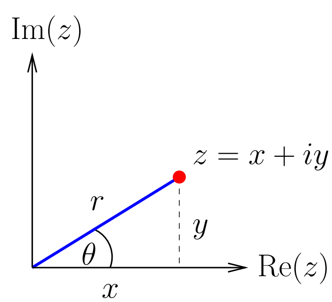
Example 9-7

Find the distance between the complex numbers $z_1 = 2 - 3i$ and $z_2 = 3 + i$.

Solution:

$$|z_1 - z_2| = |(2 - 3i) - (3 + i)| = |-1 - 4i| = \sqrt{(-1)^2 + (-4)^2} = \sqrt{17}\,.$$

9.5.1 Polar Representation



A point in polar coordinates (r, θ) as shown above, satisfies

$$\cos\theta = \frac{x}{r} \quad \Rightarrow \quad \boxed{x = r\cos\theta}$$
$$\sin\theta = \frac{y}{r} \quad \Rightarrow \quad \boxed{y = r\sin\theta}$$

To find r and θ we solve the previous equations.

$$x^2 + y^2 = r^2\cos^2\theta + r^2\sin^2\theta = r^2(\cos^2\theta + \sin^2\theta) = r^2(1) = r^2 \quad \Rightarrow \quad \boxed{r = \sqrt{x^2 + y^2} = |z|}$$
$$\frac{y}{x} = \frac{r\sin\theta}{r\cos\theta} = \frac{\sin\theta}{\cos\theta} = \tan\theta \quad \Rightarrow \quad \boxed{\tan\theta = \frac{y}{x}}$$

Note the quadrant for angle θ which solves this last equation is determined by the position of (x, y) in the complex plane.

A complex number $z = x + iy$ can therefore be written

$$z = x + iy = r\cos\theta + ir\sin\theta$$

where $r = |z|$ is the magnitude of z and θ is called the **argument of** z and is denoted $\mathbf{arg}\,(\boldsymbol{z})$. Note the argument of z is not unique. (It can be replaced by $\theta + 2n\pi$ where $n = 0, \pm 1, \pm 2 \ldots$). However there exists only one argument θ is in the range $-\pi \leq \theta \leq \pi$ and this is called the **principle argument**.

Since Euler's formula states[3]

$$\boxed{e^{i\theta} = \cos\theta + i\sin\theta}$$

we can simplify the form of z in terms of r and θ.

Definition: The **polar form** of a complex number $z = x + iy$ is

$$\boxed{z = r(\cos\theta + i\sin\theta) = re^{i\theta}},$$

where $r = |z|$ and $\theta = \arg(z)$.

Here $e = 2.7182\ldots$ is the natural constant (Euler's number).

Example 9-8

Find the polar form of the complex number.

1. $z = 2 + 2i$
   Solution:

$$\begin{aligned} x &= 2\ ,\ \ y = 2 \\ r &= \sqrt{x^2 + y^2} = \sqrt{2^2 + 2^2} = \sqrt{8} = 2\sqrt{2} \\ \tan\theta &= \frac{y}{x} = \frac{2}{2} = 1 \ \Rightarrow\ \theta = \frac{\pi}{4} \quad \text{(Since } (2, 2) \text{ is in Quadrant I.)} \end{aligned}$$

   Therefore $z = 2 + 2i = 2\sqrt{2}e^{i\frac{\pi}{4}}$.

2. $z = 4i$
   Solution:

$$\begin{aligned} x &= 0\ ,\ \ y = 4 \\ r &= \sqrt{x^2 + y^2} = \sqrt{0^2 + 4^2} = 4 \\ \sin\theta &= 4 > 0,\ \cos\theta = 0 \ \Rightarrow\ \theta = \frac{\pi}{2} \end{aligned}$$

   Therefore $z = 4i = 4e^{i\frac{\pi}{2}}$.

3. $z = -1 + \sqrt{3}i$

   Solution:

$$\begin{aligned} x &= -1\ ,\ \ y = \sqrt{3} \\ r &= \sqrt{x^2 + y^2} = \sqrt{(-1)^2 + \sqrt{3}^2} = \sqrt{4} = 2 \\ \tan\theta &= \frac{y}{x} = \frac{\sqrt{3}}{-1} = -\sqrt{3} \ \Rightarrow\ \theta = \pi - \frac{\pi}{3} = \frac{2\pi}{3} \ \text{(Since } (-1, \sqrt{3}) \text{ is in Quadrant II.)} \end{aligned}$$

   Therefore $z = -1 + \sqrt{3}i = 2e^{i\frac{2\pi}{3}}$.

[3]For readers who have studied series, Euler's Formula can be proven by plugging $i\theta$ for x in the Maclaurin series for e^x, simplifying the powers of i using $i^2 = -1$, and then breaking the resulting terms into the real and imaginary ones. The Maclaurin series of cosine and sine will be recognized in these pieces.

The law of exponents,

$$e^w e^z = e^{w+z}$$

which holds for complex numbers w and z show the value of the polar form of a complex number for multiplication, division, and powers.

Theorem 9-3: If $z_1 = r_1 e^{i\theta_1}$ and $z_2 = r_2 e^{i\theta_2}$, then:

1. $\boxed{z_1 z_2 = r_1 r_2 e^{i(\theta_1+\theta_2)}}$
2. $\boxed{\dfrac{z_1}{z_2} = \dfrac{r_1}{r_2} e^{i(\theta_1-\theta_2)}}$

Proof:

$$z_1 z_2 = r_1 e^{i\theta_1} r_2 e^{i\theta_2} = r_1 r_2 e^{i\theta_1} e^{i\theta_2} = r_1 r_2 e^{i\theta_1 + i\theta_2} = r_1 r_2 e^{i(\theta_1+\theta_2)}$$

Note that for $z = re^{i\theta}$ we have $\boxed{\overline{z} = re^{-i\theta}}$ since

$$\overline{z} = \overline{r(\cos\theta + i\sin\theta)} = r(\cos\theta - i\sin\theta) = r[\cos(-\theta) + i\sin(-\theta)] = re^{-i\theta}$$

where we used that cosine is an even function and sine is an odd function. Then

$$\frac{z_1}{z_2} = \frac{z_1\overline{z_2}}{z_2\overline{z_2}} = \frac{r_1 e^{i\theta_1} r_2 e^{-i\theta_2}}{r_2 e^{i\theta_2} r_2 e^{-i\theta_2}} = \frac{r_1 e^{i\theta_1 - i\theta_2}}{r_2 e^{i\theta_2 - i\theta_2}} = \frac{r_1 e^{i(\theta_1-\theta_2)}}{r_2 e^0} = \frac{r_1}{r_2} e^{i(\theta_1-\theta_2)},$$

where we used $e^0 = 1$.

Note:
The result for multiplication shows that the effect of multiplying a complex number by $z = re^{i\theta}$ is to scale the magnitude by r and rotate the number counterclockwise by an angle θ.

Theorem 9-4: If $z = re^{i\theta}$, then:

$$\boxed{z^n = r^n e^{in\theta}}.$$

Proof:
Using the usual rules for powers which hold for complex numbers we have:

$$z^n = (re^{i\theta})^n = r^n (e^{i\theta})^n = r^n e^{in\theta}$$

Letting $z = e^{i\theta}$ in the previous theorem gives the following corollary which can be used to prove many trigonometric identities.

Corollary: (De Moivre's Theorem)
For any integer n one has $\left(e^{i\theta}\right)^n = e^{in\theta}$, which implies

$$\boxed{(\cos\theta + i\sin\theta)^n = \cos(n\theta) + i\sin(n\theta)}.$$

Example 9-9

If $z_1 = 2+2i = 2\sqrt{2}e^{i\frac{\pi}{4}}$, $z_2 = 4i = 4e^{i\frac{\pi}{2}}$, and $z_3 = -1+\sqrt{3}i = 2e^{i\frac{2\pi}{3}}$, evaluate the indicated power.

1. $(z_1)^4$

   Solution:

   $$(z_1)^4 = (2+2i)^4 = (2\sqrt{2}e^{i\frac{\pi}{4}})^4 = 2^4(\sqrt{2})^4(e^{i\frac{\pi}{4}})^4 = 16(4)(e^{i\pi}) = 64e^{i\pi} = 64(-1) = -64$$

2. $(z_2)^{10}$

   Solution:

   $$(z_2)^{10} = (4i)^{10} = (4)^{10}(e^{i\frac{\pi}{2}})^{10} = (4)^{10}e^{i5\pi} = (4)^{10}(e^{i\pi})^5 = (4)^{10}(-1)^5 = -(4)^{10}$$

3. $(z_3)^6$

   Solution:

   $$(z_3)^6 = (2e^{i\frac{2\pi}{3}})^6 = 2^6(e^{i\frac{2\pi}{3}})^6 = 64e^{i4\pi} = 64(e^{i2\pi})^2 = 64(1)^2 = 64$$

To conclude this section it is easy to confuse complex numbers with other mathematical constructions that are similar but conceptually different. For instance, it is easy to confuse a vector in $\mathbb{R}^2$ with a complex number since they have two components which add in the same way. However complex numbers are numbers in that they share the same axioms as real numbers and form a mathematical **field**. In particular they have a commutative multiplication that is not possessed by a vector in $\mathbb{R}^2$. As such we should think of complex numbers in linear algebra as being used in the same way as are real numbers, namely as entries in a matrix (or potentially a vector) or as scalars that multiply a matrix (or vector).

A second confusion that is easily made is to think of the polar representation of a complex number as the same as polar coordinates. While it is true that one derives r and θ in $z = re^{i\theta}$ from x and y in the same way one transforms a function $f(x,y)$ of Cartesian coordinates into a function $\tilde{f}(r,\theta)$ of polar coordinates, the underlying coordinate system does not carry with it any of the structure of complex numbers; there is no i, etc. So transforming from Cartesian to polar coordinates can be generalized in three dimensions to a transformation from Cartesian coordinates (x,y,z) to spherical-polar coordinates (r,θ,ϕ), while no analogue of complex numbers even exists in three dimensions.[4] We can be interested in complex (or real) functions of complex numbers, $f(z)$, and this is the basis of a course in complex analysis. Here the multiplicative structure of the complex number imbues such functions with rich properties with wide application. Our rudimentary treatment of complex numbers here is simply to solve some basic problems that arise in linear algebra as will be shown in the next section.

[4] Analogues of complex numbers can be created in four dimensions (called **quaternions**) and eight dimensions (called **octonions**). The latter numbers do not satisfy all the field axioms that real and complex numbers do, in particular their multiplication does not commute. However they too have their useful applications. Historically quaternions were developed before vectors, and vectors in arbitrary n dimensions arose in part by realizing that the multiplicative structure of quaternions was not needed for many physical problems.

9.6 Complex Eigenvalues and Eigenvectors

The **Fundamental Theorem of Algebra** implies that every polynomial $p(x)$ of order n has n complex roots (allowing for algebraic multiplicity). This shows that, in general we will find n eigenvalues for a square matrix A of order n, though some of those values may be complex. A polynomial with real coefficients having a complex root z will also have $\overline{z}$ as a root. So complex roots show up in pairs. As an example, our polynomial from Section 9.1 with real coefficients, $p(z) = z^2 - 2z + 5$, had two complex roots, namely $z = 1 + 4i$ and its complex conjugate $z = 1 - 4i$. So if $p(z)$ were the characteristic polynomial for a 2×2 matrix, these roots would have been its (complex) eigenvalues. The following theorem shows that certain matrices are guaranteed to have real eigenvalues.

Theorem 9-5: The eigenvalues of a symmetric matrix with real entries are real.

We can solve for eigenvalues and eigenvectors involving complex eigenvalues using our usual techniques of solving linear systems, now applied to matrices with complex entries.

Example 9-10

Find the eigenvalues and eigenvectors of the given matrix.

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

Solution:

$$\begin{aligned} &\det(A - 2I) = 0 \\ &\qquad \Rightarrow \begin{vmatrix} -\lambda & -1 \\ 1 & -\lambda \end{vmatrix} = 0 \\ &\qquad \Rightarrow \lambda^2 + 1 = 0 \\ &\qquad \Rightarrow \begin{cases} \lambda_1 = i \\ \lambda_2 = -i \end{cases} \qquad \text{(complex eigenvalues despite real matrix!)} \end{aligned}$$

Next find the eigenvectors by manipulating the matrices with complex numbers in the same way we did the real matrices.

$\boldsymbol{\lambda_1 = i}$**:**

$$(A - iI)\boldsymbol{x}_1 = 0$$

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

$$\Downarrow$$

$$R_1 \leftrightarrow R_2 \left[\begin{array}{cc|c} 1 & -i & 0 \\ -i & -1 & 0 \end{array}\right]$$

$$\Downarrow$$

$$R_2 \to R_2 + iR_1 \left[\begin{array}{cc|c} 1 & -i & 0 \\ 0 & -1 - i^2 & 0 \end{array}\right] \quad \text{(But } i^2 = -1.\text{)}$$

$$\Downarrow$$

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

- $\boxed{x_2 = t}$
- $x_1 - ix_2 = 0 \implies x_1 - it = 0 \implies \boxed{x_1 = it}$

Therefore $\boldsymbol{x}_1 = \begin{bmatrix} it \\ t \end{bmatrix} = t \begin{bmatrix} i \\ 1 \end{bmatrix}$ and so $\{(i, 1)\}$ is a basis for the $\lambda = i$ eigenspace.

$\boldsymbol{\lambda_2 = -i}$**:**

$$(A - (-i)I)\boldsymbol{x}_2 = 0$$
$$(A + iI)\boldsymbol{x}_2 = 0$$

$$\left[\begin{array}{cc|c} i & -1 & 0 \\ 1 & i & 0 \end{array}\right]$$
$$\Downarrow$$
$$R_1 \leftrightarrow R_2 \left[\begin{array}{cc|c} 1 & i & 0 \\ i & -1 & 0 \end{array}\right]$$
$$\Downarrow$$
$$R_2 \to R_2 - iR_1 \left[\begin{array}{cc|c} 1 & i & 0 \\ 0 & -1 - i^2 & 0 \end{array}\right]$$
$$\Downarrow$$
$$\left[\begin{array}{cc|c} 1 & i & 0 \\ 0 & 0 & 0 \end{array}\right]$$

- $\boxed{x_2 = t}$
- $x_1 + ix_2 = 0 \implies x_1 + it = 0 \implies \boxed{x_1 = -it}$

Therefore $\boldsymbol{x}_2 = \begin{bmatrix} -it \\ t \end{bmatrix} = t \begin{bmatrix} -i \\ 1 \end{bmatrix}$ and so $\{(-i, 1)\}$ is a basis for the $\lambda = -i$ eigenspace.

GNU Free Documentation License

Preamble

The purpose of this License is to make a manual, textbook, or other functional and useful document "free" in the sense of freedom: to assure everyone the effective freedom to copy and redistribute it, with or without modifying it, either commercially or noncommercially. Secondarily, this License preserves for the author and publisher a way to get credit for their work, while not being considered responsible for modifications made by others.

This License is a kind of "copyleft", which means that derivative works of the document must themselves be free in the same sense. It complements the GNU General Public License, which is a copyleft license designed for free software.

We have designed this License in order to use it for manuals for free software, because free software needs free documentation: a free program should come with manuals providing the same freedoms that the software does. But this License is not limited to software manuals; it can be used for any textual work, regardless of subject matter or whether it is published as a printed book. We recommend this License principally for works whose purpose is instruction or reference.

1. APPLICABILITY AND DEFINITIONS

This License applies to any manual or other work, in any medium, that contains a notice placed by the copyright holder saying it can be distributed under the terms of this License. Such a notice grants a world-wide, royalty-free license, unlimited in duration, to use that work under the conditions stated herein. The "**Document**", below, refers to any such manual or work. Any member of the public is a licensee, and is addressed as "**you**". You accept the license if you copy, modify or distribute the work in a way requiring permission under copyright law.

A "**Modified Version**" of the Document means any work containing the Document or a portion of it, either copied verbatim, or with modifications and/or translated into another language.

A "**Secondary Section**" is a named appendix or a front-matter section of the Document that deals exclusively with the relationship of the publishers or authors of the Document to the Document's overall subject (or to related matters) and contains nothing that could fall directly within that overall subject. (Thus, if the Document is in part a textbook of mathematics, a Secondary Section may not explain any mathematics.) The relationship could be a matter of historical connection with the subject or with related matters, or of legal, commercial, philosophical, ethical or political position regarding them.

The "**Invariant Sections**" are certain Secondary Sections whose titles are designated, as being those of Invariant Sections, in the notice that says that the Document is released under this License. If a section does not fit the above definition of Secondary then it is not allowed to be designated as Invariant. The Document may contain zero Invariant Sections. If the Document does not identify any Invariant Sections then there are none.

The "**Cover Texts**" are certain short passages of text that are listed, as Front-Cover Texts or Back-Cover Texts, in the notice that says that the Document is released under this License. A Front-Cover Text may be at most 5 words, and a Back-Cover Text may be at most 25 words.

A "**Transparent**" copy of the Document means a machine-readable copy, represented in a format whose specification is available to the general public, that is suitable for revising the document straightforwardly with generic text editors or (for images composed of pixels) generic paint programs or (for drawings) some widely available drawing editor, and that is suitable for input to text formatters or for automatic translation to a variety of formats suitable for input to text formatters. A copy made in an otherwise Transparent file format whose markup, or absence of markup, has been arranged to thwart or discourage subsequent modification by readers is not Transparent. An image format is not Transparent if used for any substantial amount of text. A copy that is not "Transparent" is called "**Opaque**".

Examples of suitable formats for Transparent copies include plain ASCII without markup, Texinfo input format, LaTeX input format, SGML or XML using a publicly available DTD, and standard-conforming simple HTML, PostScript or PDF designed for human modification. Examples of transparent image formats include PNG, XCF and JPG. Opaque formats include proprietary formats that can be read and edited only by proprietary word processors, SGML or XML for which the DTD and/or processing tools are not generally available, and the machine-generated HTML, PostScript or PDF produced by some word processors for output purposes only.

The "**Title Page**" means, for a printed book, the title page itself, plus such following pages as are needed to hold, legibly, the material this License requires to appear in the title page. For works in formats which do not have any title page as such, "Title Page" means the text near the most prominent appearance of the work's title, preceding the beginning of the body of the text.

The "**publisher**" means any person or entity that distributes copies of the Document to the public.

A section "**Entitled XYZ**" means a named subunit of the Document whose title either is precisely XYZ or contains XYZ in parentheses following text that translates XYZ in another language. (Here XYZ stands for a specific section name mentioned below, such as "**Acknowledgements**", "**Dedications**", "**Endorsements**", or "**History**".) To "**Preserve the Title**" of such a section when you modify the Document means that it remains a section "Entitled XYZ" according to this definition.

The Document may include Warranty Disclaimers next to the notice which states that this License applies to the Document. These Warranty Disclaimers are considered to be included by reference in this License, but only as regards disclaiming warranties: any other implication that these Warranty Disclaimers may have is void and has no effect on the meaning of this License.

2. VERBATIM COPYING

You may copy and distribute the Document in any medium, either commercially or noncommercially, provided that this License, the copyright notices, and the license notice saying this License applies to the Document are reproduced in all copies, and that you add no other conditions whatsoever to those of this License. You may not use technical measures to obstruct or control the reading or further copying of the copies you make or distribute. However, you may accept compensation in exchange for copies. If you distribute a large enough number of copies you must also follow the conditions in section 3.

You may also lend copies, under the same conditions stated above, and you may publicly display copies.

3. COPYING IN QUANTITY

If you publish printed copies (or copies in media that commonly have printed covers) of the Document, numbering more than 100, and the Document's license notice requires Cover Texts, you must enclose the copies in covers that carry, clearly and legibly, all these Cover Texts: Front-Cover Texts on the front cover, and Back-Cover Texts on the back cover. Both covers must also clearly and legibly identify you as the publisher of these copies. The front cover must present the full title with all words of the title equally prominent and visible. You may add other material on the covers in addition. Copying with changes limited to the covers, as long as they preserve the title of the Document and satisfy these conditions, can be treated as verbatim copying in other respects.

If the required texts for either cover are too voluminous to fit legibly, you should put the first ones listed (as many as fit reasonably) on the actual cover, and continue the rest onto adjacent pages.

If you publish or distribute Opaque copies of the Document numbering more than 100, you must either include a machine-readable Transparent copy along with each Opaque copy, or state in or with each Opaque copy a computer-network location from which the general network-using public has access to download using public-standard network protocols a complete Transparent copy of the Document, free of added material. If you use the latter option, you must take reasonably prudent steps, when you begin distribution of Opaque copies in quantity, to ensure that this Transparent copy will remain thus accessible at the stated location until at least one year after the last time you distribute an Opaque copy (directly or through your agents or retailers) of that edition to the public.

It is requested, but not required, that you contact the authors of the Document well before redistributing any large number of copies, to give them a chance to provide you with an updated version of the Document.

4. MODIFICATIONS

You may copy and distribute a Modified Version of the Document under the conditions of sections 2 and 3 above, provided that you release the Modified Version under precisely this License, with the Modified Version filling the role of the Document, thus licensing distribution and modification of the Modified Version to whoever possesses a copy of it. In addition, you must do these things in the Modified Version:

A. Use in the Title Page (and on the covers, if any) a title distinct from that of the Document, and from those of previous versions (which should, if there were any, be listed in the History section of the Document). You may use the same title as a previous version if the original publisher of that version gives permission.

B. List on the Title Page, as authors, one or more persons or entities responsible for authorship of the modifications in the Modified Version, together with at least five of the principal authors of the Document (all of its principal authors, if it has fewer than five), unless they release you from this requirement.

C. State on the Title page the name of the publisher of the Modified Version, as the publisher.

D. Preserve all the copyright notices of the Document.

E. Add an appropriate copyright notice for your modifications adjacent to the other copyright notices.

F. Include, immediately after the copyright notices, a license notice giving the public permission to use the Modified Version under the terms of this License, in the form shown in the Addendum below.

G. Preserve in that license notice the full lists of Invariant Sections and required Cover Texts given in the Document's license notice.

H. Include an unaltered copy of this License.

I. Preserve the section Entitled "History", Preserve its Title, and add to it an item stating at least the title, year, new authors, and publisher of the Modified Version as given on the Title Page. If there is no section Entitled "History" in the Document, create one stating the title, year, authors, and publisher of the Document as given on its Title Page, then add an item describing the Modified Version as stated in the previous sentence.

J. Preserve the network location, if any, given in the Document for public access to a Transparent copy of the Document, and likewise the network locations given in the Document for previous versions it was based on. These may be placed in the "History" section. You may omit a network location for a work that was published at least four years before the Document itself, or if the original publisher of the version it refers to gives permission.

K. For any section Entitled "Acknowledgements" or "Dedications", Preserve the Title of the section, and preserve in the section all the substance and tone of each of the contributor acknowledgements and/or dedications given therein.

L. Preserve all the Invariant Sections of the Document, unaltered in their text and in their titles. Section numbers or the equivalent are not considered part of the section titles.

M. Delete any section Entitled "Endorsements". Such a section may not be included in the Modified Version.

N. Do not retitle any existing section to be Entitled "Endorsements" or to conflict in title with any Invariant Section.

O. Preserve any Warranty Disclaimers.

If the Modified Version includes new front-matter sections or appendices that qualify as Secondary Sections and contain no material copied from the Document, you may at your option designate some or all of these sections as invariant. To do this, add their titles to the list of Invariant Sections in the Modified Version's license notice. These titles must be distinct from any other section titles.

You may add a section Entitled "Endorsements", provided it contains nothing but endorsements of your Modified Version by various parties—for example, statements of peer review or that the text has been approved by an organization as the authoritative definition of a standard.

You may add a passage of up to five words as a Front-Cover Text, and a passage of up to 25 words as a Back-Cover Text, to the end of the list of Cover Texts in the Modified Version. Only one passage of Front-Cover Text and one of Back-Cover Text may be added by (or through arrangements made by) any one entity. If the Document already includes a cover text for the same cover, previously added by you or by arrangement made by the same entity you are acting on behalf of, you may not add another; but you may replace the old one, on explicit permission from the previous publisher that added the old one.

The author(s) and publisher(s) of the Document do not by this License give permission to use their names for publicity for or to assert or imply endorsement of any Modified Version.

5. COMBINING DOCUMENTS

You may combine the Document with other documents released under this License, under the terms defined in section 4 above for modified versions, provided that you include in the combination all of the Invariant Sections of all of the original documents, unmodified, and list them all as Invariant Sections of your combined work in its license notice, and that you preserve all their Warranty Disclaimers.

The combined work need only contain one copy of this License, and multiple identical Invariant Sections may be replaced with a single copy. If there are multiple Invariant Sections with the same name but different contents, make the title of each such section unique by adding at the end of it, in parentheses, the name of the original author or publisher of that section if known, or else a unique number. Make the same adjustment to the section titles in the list of Invariant Sections in the license notice of the combined work.

In the combination, you must combine any sections Entitled "History" in the various original documents, forming one section Entitled "History"; likewise combine any sections Entitled "Acknowledgements", and any sections Entitled "Dedications". You must delete all sections Entitled "Endorsements".

6. COLLECTIONS OF DOCUMENTS

You may make a collection consisting of the Document and other documents released under this License, and replace the individual copies of this License in the various documents with a single copy that is included in the collection, provided that you follow the rules of this License for verbatim copying of each of the documents in all other respects.

You may extract a single document from such a collection, and distribute it individually under this License, provided you insert a copy of this License into the extracted document, and follow this License in all other respects regarding verbatim copying of that document.

7. AGGREGATION WITH INDEPENDENT WORKS

A compilation of the Document or its derivatives with other separate and independent documents or works, in or on a volume of a storage or distribution medium, is called an "aggregate" if the copyright resulting from the compilation is not used to limit the legal rights of the compilation's users beyond what the individual works permit. When the Document is included in an aggregate, this License does not apply to the other works in the aggregate which are not themselves derivative works of the Document.

If the Cover Text requirement of section 3 is applicable to these copies of the Document, then if the Document is less than one half of the entire aggregate, the Document's Cover Texts may be placed on covers that bracket the Document within the aggregate, or the electronic equivalent of covers if the Document is in electronic form. Otherwise they must appear on printed covers that bracket the whole aggregate.

8. TRANSLATION

Translation is considered a kind of modification, so you may distribute translations of the Document under the terms of section 4. Replacing Invariant Sections with translations requires special permission from their copyright holders, but you may include translations of some or all Invariant Sections in addition to the original versions of these Invariant Sections. You may include a translation of this License, and all the license notices in the Document, and any Warranty Disclaimers, provided that you also include the original English version of this License and the original versions of those notices and disclaimers. In case of a disagreement between the translation and the original version of this License or a notice or disclaimer, the original version will prevail.

If a section in the Document is Entitled "Acknowledgements", "Dedications", or "History", the requirement (section 4) to Preserve its Title (section 1) will typically require changing the actual title.

9. TERMINATION

You may not copy, modify, sublicense, or distribute the Document except as expressly provided under this License. Any attempt otherwise to copy, modify, sublicense, or distribute it is void, and will automatically terminate your rights under this License.

However, if you cease all violation of this License, then your license from a particular copyright holder is reinstated (a) provisionally, unless and until the copyright holder explicitly and finally terminates your license, and (b) permanently, if the copyright holder fails to notify you of the violation by some reasonable means prior to 60 days after the cessation.

Moreover, your license from a particular copyright holder is reinstated permanently if the copyright holder notifies you of the violation by some reasonable means, this is the first time you have received notice of violation of this License (for any work) from that copyright holder, and you cure the violation prior to 30 days after your receipt of the notice.

Termination of your rights under this section does not terminate the licenses of parties who have received copies or rights from you under this License. If your rights have been terminated and not permanently reinstated, receipt of a copy of some or all of the same material does not give you any rights to use it.

10. FUTURE REVISIONS OF THIS LICENSE

The Free Software Foundation may publish new, revised versions of the GNU Free Documentation License from time to time. Such new versions will be similar in spirit to the present version, but may differ in detail to address new problems or concerns. See http://www.gnu.org/copyleft/.

Each version of the License is given a distinguishing version number. If the Document specifies that a particular numbered version of this License "or any later version" applies to it, you have the option of following the terms and conditions either of that specified version or of any later version that has been published (not as a draft) by the Free Software Foundation. If the Document does not specify a version number of this License, you may choose any version ever published (not as a draft) by the Free Software Foundation. If the Document specifies that a proxy can decide which future versions of this License can be used, that proxy's public statement of acceptance of a version permanently authorizes you to choose that version for the Document.

11. RELICENSING

"Massive Multiauthor Collaboration Site" (or "MMC Site") means any World Wide Web server that publishes copyrightable works and also provides prominent facilities for anybody to edit those works. A public wiki that anybody can edit is an example of such a server. A "Massive Multiauthor Collaboration" (or "MMC") contained in the site means any set of copyrightable works thus published on the MMC site.

"CC-BY-SA" means the Creative Commons Attribution-Share Alike 3.0 license published by Creative Commons Corporation, a not-for-profit corporation with a principal place of business in San Francisco, California, as well as future copyleft versions of that license published by that same organization.

"Incorporate" means to publish or republish a Document, in whole or in part, as part of another Document.

An MMC is "eligible for relicensing" if it is licensed under this License, and if all works that were first published under this License somewhere other than this MMC, and subsequently incorporated in whole or in part into the MMC, (1) had no cover texts or invariant sections, and (2) were thus incorporated prior to November 1, 2008.

The operator of an MMC Site may republish an MMC contained in the site under CC-BY-SA on the same site at any time before August 1, 2009, provided the MMC is eligible for relicensing.

ADDENDUM: How to use this License for your documents

To use this License in a document you have written, include a copy of the License in the document and put the following copyright and license notices just after the title page:

> Copyright © YEAR YOUR NAME. Permission is granted to copy, distribute and/or modify this document under the terms of the GNU Free Documentation License, Version 1.3 or any later version published by the Free Software Foundation; with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts. A copy of the license is included in the section entitled "GNU Free Documentation License".

If you have Invariant Sections, Front-Cover Texts and Back-Cover Texts, replace the "with ... Texts." line with this:

> with the Invariant Sections being LIST THEIR TITLES, with the Front-Cover Texts being LIST, and with the Back-Cover Texts being LIST.

If you have Invariant Sections without Cover Texts, or some other combination of the three, merge those two alternatives to suit the situation.

If your document contains nontrivial examples of program code, we recommend releasing these examples in parallel under your choice of free software license, such as the GNU General Public License, to permit their use in free software.