Learning Linear Algebra

To my wife, Sally, and our children,

Courtney, Rebecca, and Benjamin,

whose support made this book possible

David A. Smith

Learning Linear Algebra

With Python

First Edition

Direct Knowledge

Typeset with LaTeX and written in the United States of America.

David A. Smith
B.S. Mathematics, M.S. Mathematics
https://directknowledge.com

Preface

Writing Philosophy

Writing mathematics with confidence and eloquence requires more than creativity; strong critical thinking skills are essential for constructing persuasive arguments to communicate ideas effectively. There must also be a mastery of the technical elements that give your writing its quality. To read about my exacting standards, please check out https://directknowledge.com/writing

Acknowledgements

A hearty debt of gratitude is owed to those who worked tirelessly and selflessly to see this book through its final version. Their contributions were immense, making the publication process a true success - I am extremely thankful for their efforts!

David A. Smith \ Fort Worth, Texas

Table of contents

Chapter 1

Systems of Linear Equations

Systems of Linear Equations (A Student's Guide) is a comprehensive guide written for students. This book provides in-depth explanations of how to solve systems of linear equations using various methods, including graphing, substitution, and elimination. It also includes examples and practice problems to help you master systems for linear equations.

A system of linear equations is a set of two or more linear equations that share the same set of unknowns. In other words, a system of linear equations is a collection of linear equations that you must solve together. There are three main types of systems of linear equations:

1. Consistent and independent: A consistent and independent system of linear equations has a unique solution. This means that there is only one way to solve the system, and the solution will not change if different values are used for the unknowns.

2. Consistent and dependent: A consistent and dependent system of linear equations has infinitely many solutions. This means that there is more than one way to solve the system, and the solution will change if different values are used for the unknowns.

3. Inconsistent: An inconsistent system of linear equations has no solution. This means that there is no way to solve the system.

The graphing method is a visual way to find the solution to a system of linear equations. To use the graphing method, graph both equations on the same coordinate plane. The point of intersection is the solution to the system.

To graph a linear equation, start by plotting the y-intercept. This is the

point where the line crosses the y-axis. To find the y-intercept, set x=0 and solve for y. Then, use the slope to find additional points on the line. The slope is the ratio of the change in y to the change in x. To find the slope, pick two points on the line and calculate the rise (the change in y) divided by the run (the change in x).

Once you have the y-intercept and the slope, you can plot additional points on the line and connect them with a straight line.

The substitution method is an algebraic way to find the solution to a system of linear equations. To use the substitution method, solve one of the equations for one of the unknowns. Then, substitute this value into the other equation. This will give you a new equation with one unknown. Solve this equation for the unknown. Now, substitute the value of the unknown back into one of the original equations. This will give you the solution to the system.

The substitution method is usually easiest when one of the equations is already solved for one of the unknowns. However, it is sometimes necessary to solve one of the equations for an unknown before you can substitute.

1.1   The Elimination Method

The elimination method is an algebraic way to find the solution to a system of linear equations. To use the elimination method, add or subtract the equations so that one of the unknowns cancels out. This will give you a new equation with one fewer unknown. Repeat this process until you are left with an equation that has only one unknown. Solve this equation for the unknown. Now, substitute the value of the unknown back into one of the original equations. This will give you the solution to the system.

The elimination method is usually easiest when the coefficients (the numbers in front of the unknowns) are already opposites. However, it is sometimes necessary to multiply one or both equations by a constant before you can eliminate an unknown.

There are many ways to solve a system of linear equations. The best way to choose a method is to look at the equations and decide which one will be the easiest to work with. In general, the substitution method is best when one equation is already solved for one unknown. The elimination method is best when the coefficients of one unknown are opposites. The graphing method is a good way to check your work, but it can be difficult to find the intersection of two lines.

1.2 Row Operations and Similar Matrices

In solving systems of linear equations, we will often need to use row operations. Row operations are mathematical operations that can be performed on the rows of a matrix. There are three types of row operations:

- Interchange two rows
- Multiply a row by a nonzero constant
- Add a multiple of one row to another row

Row operations do not change the solution set of a system of linear equations. This means that if we perform row operations on a matrix, the new matrix will have the same solution set as the original matrix. We can use row operations to transform a matrix into an equivalent matrix. Two matrices are equivalent if they have the same solution set.

We can use row operations to solve systems of linear equations. In general, we will use row operations to transform the coefficient matrix into an upper triangular matrix. An upper triangular matrix is a matrix where all the elements below the main diagonal are zero. Once we have transformed the coefficient matrix into an upper triangular matrix, we can use back substitution to solve the system.

We can use row operations to transform a matrix into any equivalent matrix. However, some matrices are more difficult to work with than others. In general, it is easier to work with matrices that are in reduced row echelon form. A matrix is in reduced row echelon form if it is upper triangular and if the leading entry in each nonzero row is a 1. The leading entry in a row is the first nonzero entry in the row.

We can use row operations to transform any matrix into an equivalent matrix that is in reduced row echelon form. Once we have transformed the matrix into reduced row echelon form, we can use back substitution to solve the system of linear equations.

1.3 Gaussian Elimination Method

The Gaussian elimination method is another way to solve a system of linear equations. To use the Gaussian elimination method, you must first arrange the equations into row form. This will make it easier to see how to eliminate an unknown. Next, multiply each equation by a constant so that one of the unknowns cancels out. Then, add or subtract these equations so that another unknown cancels out. Finally, solve the equation resulting from this process for the last unknown. Now, substitute this value back into one of the original equations and you have the solution to the system!

The Gaussian Elimination Method is a procedure used to solve systems of linear equations. The method is named after German mathematician

Carl Friedrich Gauss and Irish mathematician William Rowan Hamilton, who developed it independently in the early 19th century.

1.4 Gauss-Jordan Elimination Method

The Gauss-Jordan Elimination Method is similar to the traditional Gaussian Elimination Method, but it is more efficient and easier to use. The method works by first transforming the system of equations into an equivalent system with a triangular form. Then, the solution is obtained by a backward substitution.

This method is called Gauss-Jordan Elimination because it was developed by two mathematicians, Carl Friedrich Gauss and Wilhelm Jordan. Although it is not the quickest or most intuitive method, it is a robust way to solve linear equations. When done correctly, it will always give you the correct answer.

1.5 Conclusion

There are many ways to solve a system of linear equations. The best way to choose a method is to look at the equations and decide which one will be the easiest to work with. In general, the substitution method is best when one equation is already solved for one unknown. The elimination method is best when the coefficients of one unknown are opposites. The Gaussian elimination method is best when there are a lot of equations and unknowns. Lastly, the Gauss-Jordan elimination method is the most robust way to solve linear equations. It is not always the quickest or easiest method, but it will always give you the correct answer.

No matter which method you choose, make sure you understand the steps involved and why they work. This will help you not only solve the equations correctly but also to understand what is happening behind the scenes. With a little practice, you will be solving systems of linear equations like a pro!

A linear system of equations in two variables x and y has the form

$$\begin{cases} ax + by = c \\ dx + ey = f \end{cases}$$

where a, b, c, d, e, f are given numbers such as real numbers or complex numbers. Sometimes a linear system is also written using indices

$$\begin{cases} a_{11}x + a_{12}y = b_1 \\ a_{21}x + a_{22}y = b_2 \end{cases} \tag{1.1}$$

to help reduce on the number of letters used. For simplicity let's temporally assume that the coefficients are nonzero real numbers and ask the question: how many solutions can there be to a 2×2 system? The key idea is to realize that each linear equation in 1.1 represents a line in the Cartesian plane. If we consider the possible ways lines can intersect in the plane we come to the conclusion that there must either be no solutions, one unique solution, or infinitely many points (x, y) that solves the system in 1.1.

Example 1.1. Determine whether the linear system

$$\begin{cases} 2x + 3y = 0 \\ 4x + 5y = 0. \end{cases}$$

has no solutions, exactly one solution, or infinitely many solutions. If we multiple the first equation, namely $2x + 3y = 0$ by 2 and subtract from the second equation, $4x + 5y = 0$, we obtain $y = 0$. Therefore the solution is unique and is $(x, y) = (0, 0)$.

A 3×3 linear system has the form

$$\begin{cases} a_{11}x + a_{12}y + a_{13}z = b_1 \\ a_{21}x + a_{22}y + a_{23}z = b_2 \\ a_{31}x + a_{32}y + a_{33}z = b_3 \end{cases} \tag{1.2}$$

where a_{ij} and b_1, b_2, b_3 are numbers and x, y, z are variables. Geometrically, linear equations in three variables are just planes in three dimensions. So what are the different types of solution sets for the system in 1.2? By considering the possible ways three planes might intersect in three dimensions we come to the conclusion that there must either be no solutions, one unique solution, or infinitely many points (x, y, z) that solves the system in 1.2.

Example 1.2. Determine whether the linear system

$$\begin{cases} x + 2y + 3z = 0 \\ 4x + 5y + 6z = 3 \\ 7x + 8y + 9z = 0 \end{cases} \tag{1.3}$$

has no solutions, exactly one solution, or infinitely many solutions. Multiply the first equation by -2 and add to the second equation obtaining the equation $2x + y = 3$. Eliminating z from the first and third equations we obtain $4x + 2y = 0$ by multiplying the first equation by -3 and adding

to the third equation. The x and y that satisfies 1.3 must also satisfy the system

$$\begin{cases} 2x + y = 3 \\ 4x + 2y = 0. \end{cases} \tag{1.4}$$

Multiply the first equation of 1.4 by 2 yields $4x + 2y = 6$. Notice there are no x and y that satisfies both $4x + 2y = 6$ and $4x + 2y = 0$. Thus the system in 1.4 has no solutions; therefore the system in 1.3 also has no solutions.

Example 1.3. Given numbers a, b, and c, show that the linear system

$$\begin{cases} x + 2y + 3z = a \\ x + 3y + 8z = b \\ x + 2y + 2z = c \end{cases}$$

either has no solutions, exactly one solution, or infinitely many solutions. We choose to eliminate x first and we obtain the system

$$\begin{cases} -y - 5z = a - b \\ z = a - c. \end{cases}$$

Next we eliminate z obtaining $y = -6a + b + 5c$. Therefore the unique solution is $(x, y, z) = (10a - 2b - 7c, -6a + b + 5c, a - c)$.

Let's consider a set of linear equations that involves n unknown quantities represented by $x_1, x_2, \ldots, x_n$. Let a_{ij} represent the number that is the coefficient of x_j in the ith equation. Let $b_1, b_2, \ldots, b_m$ be given numbers. The linear system of equations

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \\ \quad\quad\quad\quad\quad\quad\quad\quad\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m \end{cases} \tag{1.5}$$

is called a **system of simultaneous linear algebraic equations**. A **solution** of this system is an ordered set of n numbers that satisfies each of the m statements in the system. A linear system of equations with no solution is called and a system with at least one solution is called **inconsistent** . The array

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ & & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{bmatrix}$$

is called the corresponding to the linear system in 1.5 **augmented matrix**.

For example, the linear system in 1.1 was shown to be consistent and the linear system in 1.2 was shown to be inconsistent. The augmented matrices for these systems are as follows.

$$\left[\begin{array}{cc|c} 2 & 3 & 0 \\ 4 & 5 & 0 \end{array} \right] \qquad \left[\begin{array}{ccc|c} 1 & 2 & 3 & 0 \\ 4 & 5 & 6 & 3 \\ 7 & 8 & 9 & 0 \end{array} \right]$$

Example 1.4. Find all solutions to the following linear system.

$$\begin{cases} -150x + 500y = z \\ 50x + 100y + z = 200 \end{cases}$$

The given system is equivalent to

$$\begin{cases} -150x + 500y - z = 0 \\ 50x + 100y + z = 200. \end{cases}$$

Adding these equations yields $-100x + 600y = 200$. Since we have one equation with two variables, one of the variables is free. We choose to let y be free. Let $y = t$ for an arbitrary number t. Then solving for x we obtain $-100x = 200y - 600y$, or equivalently $x = -2 + 6t$. By substituting into the original system, we find $z = -150(-2 + 6t) + 500t = 300 - 400t$. Therefore, there are an infinitely many solutions, which can be represented by the set

$$\{(x, y, z) \mid x = -2 + 6t, y = t, z = 300 - 400t \text{ where } t \in \mathbb{R}\}.$$

The augmented matrix for the system in 1.4 is

$$\left[\begin{array}{ccc|c} -150 & 500 & -1 & 0 \\ 50 & 100 & 1 & 200 \end{array} \right]$$

Do you think it is possible to decide, by inspection of an augmented matrix, whether or not the corresponding linear system will be consistent or inconsistent?

In our examples so far we have seen linear systems having no solutions, a unique solution, or perhaps an infinite number of solutions. These examples suggest the following definition.

Definition 1.1. Two linear systems are called **equivalent** if they have the same solution set.

Example 1.5. Find a system of linear equations with three unknowns x, y, z whose solutions are $x = 6 + 5t$, $y = 4 + 3t$, $z = 2 + t$ where t is arbitrary. We want to eliminate t. Solving for t yields $t = z - 2$. By substitution, $x = 6 + 5(z - 2)$ and $y = 4 + 3(z - 2)$. Thus we have a linear system

$$\begin{cases} x - 5z = -4 \\ y - 3z = -2 \end{cases}$$

which has an infinitely many solutions.

Theorem 1.1. *Linear systems of equations are equivalent if each can be obtained from the other by one or more of the following operations.*

- *Interchange the order of the equations.*
- *Multiply (or divide) one equation by a nonzero scalar.*
- *Add one multiple of one equation to another.*

Proof. If we consider the solution set as a geometric object, it should be very easy to understand that they way in which we write the equations that represents the object does not change the object. Thus it should be obvious that interchanging the order of the equations will not change the solutions to a linear system. Neither will multiplying (or dividing) both sides of an equation by a nonzero constant.

Let system A be an $m \times n$ system represented by

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\ \qquad\qquad\qquad\qquad\qquad \vdots \\ a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n = b_i \\ \qquad\qquad\qquad\qquad\qquad \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m \end{cases} \qquad (1.6)$$

Consider system B obtained from system A by adding k times equation i to equation j as follows:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\ \quad \vdots \qquad\qquad \vdots \\ (a_{j1} + ka_{i1})x_1 + (a_{j2} + ka_{i2})x_2 + \cdots + (a_{jn} + ka_{in})x_n = b_j + kb_i \\ \quad \vdots \qquad\qquad \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m \end{cases}$$

$$(1.7)$$

Let S_A and S_B be the solutions sets of systems A and B, respectively. We will show $S_A = S_B$. Let $x_0 = (x_1, x_2, \ldots, x_n)$ be a solution of system

A. Thus x_0 satisfies every linear equation in 1.6. So x_0 satisfies every equation in system B except possibly the jth equation. Working with the jth equation in system B we find

$$(a_{j1} + ka_{i1})x_1 + (a_{j2} + ka_{i2})x_2 + \cdots + (a_{jn} + ka_{in})x_n$$
$$= a_{j1}x_1 + \cdots + a_{jn}x_n + k(a_{i1}x_1 + \cdots a_{jn}x_n)$$
$$= b_j + kb_i.$$

This shows x_0 also satisfies the jth equation of system B. Since x_0 satisfies every equation in 1.7, x_0 is also a member of S_B. So far we have shown $S_A \subseteq S_B$. Conversely, assume x_0 is a solutions to every equation in 1.7. Working with the jth equation of system A we find,

$$a_{j1}x_1 + a_{j2}x_2 + \cdots + a_{jn}x_n = b_j$$
$$\Longleftrightarrow \quad a_{j1}x_1 + (ka_{i1} - ka_{i1})x_1 + \cdots + a_{jn} + (ka_{i1} - ka_{i1})x_n = b_j$$
$$\Longleftrightarrow \quad (a_{j1} + ka_{i1})x_1 + (a_{j2} + ka_{i2})x_2 + \cdots + (a_{jn} + ka_{in})x_n = b_j + kb_i.$$
$$(1.8)$$

Notice 1.8 holds since x_0 satisfies system B; and thus x_0 satisfies the jth equation of system A. Therefore $S_A = S_B$ as desired.

$\square$

While working through the details of the following two examples, notice how the operations in 1.1 are being used.

Example 1.6. Let a, b, and c be constants. Solve the linear system

$$\begin{cases} y + z = a \\ x + z = b \\ x + y = c. \end{cases}$$

Eliminating z from the first and second equations we obtain the system

$$\begin{cases} x - y = b - a \\ x + y = c \end{cases}$$

Solving for x yields $x = (c + b - a)/2$. Using the original system we find y and z to be

$$y = c - x = c - (c + b - a)/2 = (c - b + a)/2$$

$$z = a - y = a - (c - b + a)/2 = (a + b - c)/2$$

Therefore, the solution to the system is

$$(x, y, z) = \left(\frac{c + b - a}{2}, \frac{a + c - b}{2}, \frac{a + b - c}{2} \right).$$

Example 1.7. Find the smallest positive integer C such that x, y, z are integers and satisfies the linear system of equations

$$\begin{cases} 2x + y = C \\ 3y + z = C \\ x + 4z = C. \end{cases}$$

Multiply the third equation by -2 and add to the first equation, obtaining $y - 8z = -C$. Multiplying the second equation by 8 and adding to $y - 8z = -C$, yields $25y = 7C$. Solving for y we obtain $y = (7/25)C$. By substitution, $x = (9/25)C$ and $z = (4/25)C$. Therefore, 25 is the smallest integer C such that x, y, z are integers and solves the system.

We will solve a linear system of equations using elementary row operations on matrices using a procedure known as **Gaussian elimination** . The solution set will be a collection of vectors. Here is the overall strategy to solving linear systems using matrices.

$$\left\{ \begin{array}{c} \text{linear} \\ \text{system of} \\ \text{equations} \end{array} \right\} \rightarrow \left\{ \begin{array}{c} \text{matrix rep-} \\ \text{resentation} \\ \text{of system} \end{array} \right\} \rightarrow \left\{ \begin{array}{c} \text{row echelon} \\ \text{form of} \\ \text{system} \end{array} \right\} \rightarrow \left\{ \begin{array}{c} \text{solutions of} \\ \text{system as a} \\ \text{set of} \\ \text{vectors} \end{array} \right\}$$

Definition 1.2. The following operations are collectively known as **elementary row operations** .

- Interchange two rows.
- Multiply a row by a nonzero scalar.
- Add a multiple of a row from another row.

From 1.1 it is obvious that applying elementary row-operations to a linear system of equations leads to an equivalent system. Reducing a linear system of equations, while preserving the solutions set, is an extremely useful idea that we will develop extensively.

For example, in 1.8 (see below) we find that the linear systems

$$\begin{cases} x + 2y + z = 3 \\ 2x + 5y - z = -4 \\ 3x - 2y - z = 5 \end{cases} \qquad \begin{cases} x = 2 \\ y = -1 \\ z = 3 \end{cases} \qquad (1.9)$$

are equivalent. In 1.8 we will show how to apply elementary row-operations to obtain the system on the right. While the system on the left might be a given linear system of equations, the system on the right is solved.

Definition 1.3. A matrix is said to be in **row echelon form** if it satisfies all the following conditions.

- All rows with at least one nonzero coefficient are above any rows of all zeroes.
- The first nonzero number from the left, (called the **leading coefficient**) of a nonzero row is always strictly to the right of the leading coefficient of the row above it.
- All entries in a column below a leading coefficient are zeroes.

Further, a matrix is said to be in **reduced row echelon form** if it is in row echelon form and the additional condition holds.

- Every leading coefficient is 1 and is the only nonzero entry in its column.

For example, the augmented matrices for the linear systems in 1.9 are

$$
\left[\begin{array}{ccc|c}
1 & 2 & 1 & 3 \\
2 & 5 & -1 & -4 \\
3 & -2 & -1 & 5
\end{array}\right]
\qquad
\left[\begin{array}{ccc|c}
1 & 0 & 0 & 2 \\
0 & 1 & 0 & -1 \\
0 & 0 & 1 & 3
\end{array}\right].
$$

By checking conditions (i)-(iv) we can see that the matrix on the right is in reduced row echelon form. For more examples, consider the following matrices.

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

$$
D = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}
\qquad
E = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}
\qquad
F = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}
$$

Notice that matrices D and E are in reduced row echelon form and the others are not.

Example 1.8. Use Gauss-Jordan elimination to solve the system.

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

Using row operations on the augmented matrix we obtain the reduced

row echelon form.

$$\begin{bmatrix} 1 & 2 & 1 & 3 \\ 2 & 5 & -1 & -4 \\ 3 & -2 & -1 & 5 \end{bmatrix} \xrightarrow[\substack{-2R_1+R_2 \\ -3R_1+R_3}]{} \begin{bmatrix} 1 & 2 & 1 & 3 \\ 0 & 1 & -3 & -10 \\ 0 & -8 & -4 & -4 \end{bmatrix}$$

$$\xrightarrow[8R_2+R_3]{} \begin{bmatrix} 1 & 2 & 1 & 3 \\ 0 & 1 & -3 & -10 \\ 0 & 0 & -28 & -84 \end{bmatrix} \xrightarrow[-\frac{1}{28}R_3]{} \begin{bmatrix} 1 & 2 & 1 & 3 \\ 0 & 1 & -3 & -10 \\ 0 & 0 & 1 & 3 \end{bmatrix}$$

$$\xrightarrow[\substack{3R_3+R_2 \\ -R_3+R_1}]{} \begin{bmatrix} 1 & 2 & 0 & 0 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 3 \end{bmatrix} \xrightarrow[-2R_2+R_1]{} \begin{bmatrix} 1 & 0 & 0 & 2 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 3 \end{bmatrix}.$$

Therefore the unique solution is $x = 2$, $y = -1$, and $z = 3$.

Example 1.9. Use the Gauss-Jordan elimination method to solve the following linear system.

$$\begin{cases} x + y - 2z + 4t = 5 \\ 2x + 2y - 3z + t = 3 \\ 3x + 3y - 4z - 2t = 1 \end{cases} \tag{1.10}$$

Using row operations on the augmented matrix we obtain the reduced row echelon form.

$$\begin{bmatrix} 1 & 1 & -2 & 4 & 5 \\ 2 & 2 & -3 & 1 & 3 \\ 3 & 3 & -4 & -2 & 1 \end{bmatrix} \xrightarrow[\substack{-2R_1+R_2 \\ -3R_1+R_3}]{} \begin{bmatrix} 1 & 1 & -2 & 4 & 5 \\ 0 & 0 & 1 & -7 & -7 \\ 0 & 0 & 2 & -14 & -14 \end{bmatrix}$$

$$\xrightarrow[\substack{-2R_2+R_3 \\ 2R_2+R_1}]{} \begin{bmatrix} 1 & 1 & 0 & -10 & -9 \\ 0 & 0 & 1 & -7 & -7 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

The system in 1.10 is equivalent to

$$\begin{cases} x + y - 10t = -9 \\ z - 7t = -7 \end{cases} \quad \text{or more simply} \quad \begin{cases} x = -9 - y + 10t \\ z = -7 + 7t. \end{cases}$$

Therefore

$$\{(x, y, z, w) \in \mathbb{R}^4 \mid x = -9-y+10t, y = s, z = -7+7t, w = t, \text{ for } s, t \in \mathbb{R}\}.$$

is the solution set.

Example 1.10. Use the Gauss-Jordan elimination method to solve the following linear system.

$$\begin{cases} x_1 + x_2 - 2x_3 + 3x_4 = 4 \\ 2x_1 + 3x_2 + 3x_3 - x_4 = 3 \\ 5x_1 + 7x_2 + 4x_3 + x_4 = 5 \end{cases}$$

Using row operations on the augmented matrix we obtain the reduced row echelon form.

$$
\left[\begin{array}{ccc|c}
1 & 1 & -2 & 3 \\
2 & 3 & 3 & -1 \\
5 & 7 & 4 & 1
\end{array}\right]
\begin{array}{c} \\ 4 \\ 3 \\ 5 \end{array}
\xrightarrow[-5R_1+R_3]{-2R_1+R_2}
\left[\begin{array}{ccc|c}
1 & 1 & -2 & 3 \\
0 & 1 & 7 & -7 \\
0 & 2 & 14 & -14
\end{array}\right]
$$

$$
\xrightarrow{-2R_2+R_3}
\left[\begin{array}{cccc|c}
1 & 1 & -2 & 3 & 4 \\
0 & 1 & 7 & -7 & -5 \\
0 & 0 & 0 & 0 & 25
\end{array}\right]
$$

Notice the last row corresponds to the equation $0x_1+0x_2+0x_3+0x_4 = 25$. Therefore there are no solutions to this system. Actually using Gauss-Jordan elimination means reaching reduced row echelon form; however, once we encounter the row $[0\ 0\ 0\ 0\ 25]$ we immediately stop and reach the correct conclusion: no solutions to the system.

Example 1.11. Find the polynomial of degree 2 whose graph goes through the points $(1,-1)$, $(2,3)$, and $(3,13)$. Let $f(t) = a + bt + ct^2$ be such a polynomial. Using the given points we setup a system and solve.

$$
\left[\begin{array}{ccc|c}
a & b & c & -1 \\
a & 2b & 4c & 3 \\
a & 3b & 9c & 13
\end{array}\right]
\xrightarrow[-R_1+R_3]{-R_1+R_2}
\left[\begin{array}{ccc|c}
a & b & c & 1 \\
0 & b & 3c & 4 \\
0 & 2b & 8c & 14
\end{array}\right]
$$

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

$$
\xrightarrow{-3R_3+R_2}
\left[\begin{array}{ccc|c}
a & b & c & -1 \\
0 & b & 0 & -5 \\
0 & 0 & c & 3
\end{array}\right]
\xrightarrow{-R_2+R_1}
\left[\begin{array}{ccc|c}
a & 0 & c & 4 \\
0 & b & 0 & -5 \\
0 & 0 & c & 3
\end{array}\right]
$$

$$
\xrightarrow{-R_3+R_1}
\left[\begin{array}{ccc|c}
a & 0 & 0 & 1 \\
0 & b & 0 & -5 \\
0 & 0 & c & 3
\end{array}\right].
$$

Therefore the required polynomial is $f(t) = 3t^2 - 5t + 1$.

1.6 Exercises

Exercise 1.1. Verify that $(2, 3, -1)$ is a solution of the linear system.

- $\begin{cases} x + 2y + z = 7 \\ x - y = -1 \\ 4x + y + z = 10 \end{cases}$

- $$\begin{cases} x + y = 5 \\ x - z = 3 \\ y + z = 2 \end{cases}$$

Exercise 1.2. Verify that every triple of the form $(7 - 2k, 8 + 6k, k)$ is a solution of the linear system of equations

- $$\begin{cases} x_1 + 2x_3 = 7 \\ x_2 - 6x_3 = 8 \end{cases}$$

- $$\begin{cases} 2x_1 + 4x_3 = 14 \\ x_1 + 3x_2 - 16x_3 = 31 \end{cases}$$

Exercise 1.3. Do the following systems have no solutions, exactly one solution, or infinitely many solutions? Justify your answer.

- $$\begin{cases} x + y = 1 \\ x + 2y = 1 \end{cases}$$

- $$\begin{cases} 3x + y = 1 \\ y = -2 \end{cases}$$

- $$\begin{cases} x + y = 1 \\ 2x + 2y = 2 \end{cases}$$

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

Exercise 1.4. Sketch the graph of each equation of the linear system and decide whether it has no solutions, exactly one solution, or infinitely many solutions.

- $$\begin{cases} x + y = 2 \\ 2x + 3y = 0 \end{cases}$$

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

Exercise 1.5. Find the solutions, if any, of the following linear systems of equations without using matrices. If an equation has more than one solution, write the general solution.

- $$\begin{cases} 3x - 2y = 7 \\ 2x + y = 15 \end{cases}$$

- $$\begin{cases} x + y = 7 \\ 3x + 4y = 12 \end{cases}$$

$$\bullet \quad \begin{cases} 2x_1 - 3x_2 = 1 \\ 4x_1 - 6x_2 = -2 \\ x_1 + x_2 = 1 \end{cases}$$

$$\bullet \quad \begin{cases} 2x_1 - 5x_2 = 12 \end{cases}$$

$$\bullet \quad \begin{cases} ix_1 - 3ix_2 = 1 \\ (2+i)x_1 - x_2 = -1 \end{cases}$$

$$\bullet \quad \begin{cases} (1+i)x_1 - 2ix_2 = 2 \\ 2x_1 - 3ix_2 = 4 - 3i \end{cases}$$

Exercise 1.6. If possible, find the point(s) of intersection.

- $x - 4y = 11$ and $7x - 2y = 9$
- $x - 4y + 3z = 11$, $7x - 2y - z = -1$, $7x - 2y + z = -2$

Exercise 1.7. Let $u = (1, 1, 2, -1)$ and $v = (1, 1, 1, 0)$. For which scalars a and b is it true that $au + bv$ is a solution to the following system?

$$\bullet \quad \begin{cases} 4x - 2y - z - w = 1 \\ x + 3y - 2z - 2w = 2 \end{cases}$$

$$\bullet \quad \begin{cases} x - 4y - z - 2w = 4 \\ 7x + y - 5z - 2w = 12 \end{cases}$$

Exercise 1.8. Find all solutions, if any exist, for the following linear systems of equations.

$$\bullet \quad \begin{cases} x - y = 1 \\ 2x = 4 \end{cases}$$

$$\bullet \quad \begin{cases} 2x + 3y - z = 19 \\ 3x - 2y + 3z = 7 \end{cases}$$

$$\bullet \quad \begin{cases} x - 3y = 2 \\ -2x + 6y = -4 \end{cases}$$

$$\bullet \quad \begin{cases} x - y + 2z - 2w = 1 \\ 2x + y + 3w = 4 \\ 2x + y + 3w = 6 \end{cases}$$

Exercise 1.9. Find all solutions, if any exist, for the following linear systems of equations.

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

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

Exercise 1.10. Find all solutions, if any exist, for the following linear systems of equations.

- $\begin{cases} x_1 + 2x_2 - x_3 = 0 \\ 2x_1 + 5x_2 + 5x_3 = 0 \\ x_1 + 4x_2 + 7x_3 = 0 \\ x_1 + 3x_2 + 3x_3 = 0 \end{cases}$

- $\begin{cases} x + y + z + w = 1 \\ 2x - 2y + z + 2w = 3 \\ 2x + 6y + 3z + 2w = 1 \\ 5x - 3y + 3z + 5w = 8 \end{cases}$

Exercise 1.11. For what values of the constant k do the systems have no solution, exactly one solution, or infinity many solutions.

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

- $\begin{cases} x + ky = 1 \\ 2x - y = k \end{cases}$

Exercise 1.12. Find a, b, c such that

$$\frac{x^2 - x + 3}{(x^2 + 2)(2x - 1)} = \frac{ax + b}{x^2 + 2} + \frac{c}{2x - 1}.$$

Exercise 1.13. Let a and b be arbitrary constants. Find all solutions to the linear system.

- $\begin{cases} x + 2y = a \\ 3x + 5y = b \end{cases}$

- $\begin{cases} ax + 2y = 1 \\ 3x + by = 4 \end{cases}$

Exercise 1.14. Let a and b be arbitrary constants. Find all solutions to the linear system.

- $$\begin{cases} x + 2y + 3z = a \\ x + 3y + 8z = b \\ x + 2y + 2z = c \end{cases}$$

- $$\begin{cases} x - 2y + 4z = a \\ x - 3y + 5z = b \\ x - 2y + 6z = c \end{cases}$$

Exercise 1.15. A system of linear equations all of whose constant terms are zero is called a **homogenous system**.

- Show that a homogenous system always has at least one solution.
- Give examples to show that a homogenous system may have more than one solution or exactly one solution.

Exercise 1.16. Write the system corresponding to each augmented matrix.

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

Exercise 1.17. Which of the following matrices are in reduced row echelon form?

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

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

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

Exercise 1.18. Which of the following matrices are in reduced row echelon form?

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

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

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

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

Exercise 1.19. Find all solutions of the linear system using the Gauss-Jordan elimination method.

- $\begin{cases} x_2 + 2x_4 + 3x_5 = 0 \\ 4x_4 + 8x_5 = 0 \end{cases}$

- $\begin{cases} 3x + 4y = 0 \\ -2x + 7y = 0 \end{cases}$

Exercise 1.20. Find all solutions of the linear system using the Gauss-Jordan elimination method.

- $\begin{cases} x_4 + 2x_5 - x_6 = 2 \\ x_1 + 2x_2 + x_5 - x_6 = 0 \\ x_1 + 2x_2 + 2x_3 - x_5 + x_6 = 2 \end{cases}$

- $\begin{cases} 3x_1 + 3x_2 - 4x_3 + x_4 = 2 \\ x_1 + x_2 - x_3 - x_4 = 5 \end{cases}$

- $\begin{cases} 2x_1 + 3x_2 + 4x_3 = 6 - x_4 \\ 3x_1 - 2x_2 - x_4 = 1 + 4x_3 \\ 3x_1 + 3x_3 + x_4 = 4 - x_2 \\ x_2 + x_3 - 4x_4 = -3 - 4x_1 \end{cases}$

- $$\begin{cases} 4x_1 + 3x_2 + 2x_3 - x_4 = 4 \\ 5x_1 + 4x_2 + 3x_3 - x_4 = 4 \\ -2x_1 - 2x_2 - x_3 + 2x_4 = -3 \\ 11x_1 + 6x_2 + 4x_3 + x_4 = 11 \end{cases}$$

- $$\begin{cases} 2x_1 - 3x_2 + x_3 = 5 \\ x_1 + x_2 - x_3 = 3 \\ 4x_1 - x_2 - x_3 = 1 \end{cases}$$

- $$\begin{cases} x_1 - x_2 = 4 \\ 2x_1 + x_2 = 7 \\ 5x_1 - 2x_2 = 19 \end{cases}$$

Exercise 1.21. Find all 4×1 matrices in reduced row echelon form.

Exercise 1.22. How many types of 3×2 matrices in reduced row echelon form are there? - How many types of 2×3 matrices in reduced row echelon form are there?

Exercise 1.23.

- Describe the possible reduced row echelon forms for a matrix with two rows and two columns.
- Describe the possible reduced row echelon forms for a matrix with three rows and three columns.

Exercise 1.24. Find the polynomial of degree 3 whose graph passes through the points $(0, 1)$, $(1, 0)$, $(-1, 0)$, and $(2, -15)$. Sketch the graph of this cubic.

Exercise 1.25. Find the polynomial of degree 4 whose graph passes through the points $(1, 1)$, $(2, -1)$, $(3, -59)$, and $(-2, -29)$. Sketch the graph of this quartic.

Exercise 1.26. Find the values of k, if any, for which the system has

- only one solution,

- no solutions,

- an infinite number of solutions

- $\begin{cases} x_1 - x_2 = 3 \\ 3x_1 + kx_2 = 6 \end{cases}$

- $\begin{cases} x_1 + 2kx_2 = 7 \\ x_1 + (k^2 + 1)x_2 = 3 \end{cases}$

- $\begin{cases} kx_1 - x_2 = 1 \\ x_1 + kx_2 = i \end{cases}$

- $\begin{cases} kx + |k|y = 1 \\ |k|x - ky = -1 \end{cases}$

Exercise 1.27. Find values for a such that the system of linear equations

$$\begin{cases} x_1 - 3x_2 + x_3 = 1 \\ 2x_1 + x_2 - x_3 = -1 \\ 5x_1 - 8x_2 + (a^2 - 2)x_3 = a \end{cases}$$

will have

- infinitely many solutions,
- no solutions
- exactly one solution.

Exercise 1.28. Use Gauss-Jordan elimination to solve the linear system

$$\begin{cases} ax_1 + bx_2 & = r \quad (a \neq 0) \\ cx_1 + dx_2 & = s \quad (e \neq 0) \\ \qquad ex_3 + fx_4 & = t \\ \qquad gx_3 + hx_4 & = u \end{cases}$$

State conditions on a, b, c, d, e, f, g and h which guarantee a unique solution.

Exercise 1.29. Show that the system

$$\begin{cases} a_1x + b_1y + c_1z = 0 \\ a_2x + b_2y + c_2z = 0 \end{cases}$$

always has a solution other than $x = 0$ and $y = 0$.

Exercise 1.30. What can you say about the last row of the reduced row echelon form of the following matrix?

$$\begin{bmatrix} 1 & 2 & 3 & \cdots & n \\ n+1 & n+2 & n+3 & \cdots & 2n \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ n^2-n+1 & n^2-n+2 & n^2-n+3 & \cdots & n^2 \end{bmatrix}$$

Exercise 1.31. Show that the linear system of equations

$$\begin{cases} 3x_1 + 2x_2 - 4x_3 = a \\ -4x_1 + x_2 - x_3 = b \\ 7x_1 + 12x_2 - 22x_3 = c \end{cases}$$

will have infinitely many solutions if $c = 5a + 2b$.

Exercise 1.32. Find an equation of the form $x^2 + y^2 + ax + by + c = 0$ of the circle passing through the following points.

- $(-2, 1)$, $(5, 0)$, and $(4, 1)$
- $(1, 1)$, $(5, -3)$, and $(-3, -3)$

1.7   Matrices and Vectors

The entries of an $n \times m$ **matrix** A are denoted by $a_{i,j}$ where $1 \leq i \leq n$ and $1 \leq j \leq m$. For example, the matrix

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix}$$

is a 2×3 matrix with $i = 2$ rows and $j = 3$ columns. If $i = j$, then the matrix is called a **square matrix** .

Definition 1.4. Two matrices $A = [a_{ij}]$ and $B = [b_{ij}]$ of the same size are equal when $a_{ij} = b_{ij}$ for all i, j.

A square matrix $A = [a_{ij}]$ is called a **diagonal matrix** if $a_{ij} = 0$ whenever $i \neq j$. A square matrix $A = [a_{ij}]$ is called **upper triangular** (**lower triangular**) when $a_{ij} = 0$ whenever $i > j$ ($i < j$). The **zero matrix** has all entries zero and the **identity matrix** has $a_{ii} = 1$, $i = 1, 2, 3..., n$ and all other entries zero. We remark that when m and n are understood

from discussion they are usually left off.

$$0_{m\times n} = \begin{bmatrix} 0 & \cdots & 0 \\ 0 & \cdots & 0 \\ \vdots & \cdots & \vdots \\ 0 & \cdots & 0 \end{bmatrix}_{m\times n} \qquad I_n = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}_{n\times n}$$

Consider the following matrices.

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \quad B = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad C = \begin{bmatrix} 2 & 3 \\ 0 & 4 \end{bmatrix} \quad D = \begin{bmatrix} 5 & 0 & 0 \\ 4 & 0 & 0 \\ 3 & 2 & 1 \end{bmatrix}$$

Notice matrices B, C and D are square and A is not. Also notice the only diagonal matrix is B, matrices B and C are upper triangular, and the only lower triangular matrix is D.

We will see that matrix addition has many of the same properties that scalar addition has; for example, commutativity, associativity, additive inverses, and the additive identity property.

Definition 1.5. Let $A = [a_{ij}]$ and $B = [b_{ij}]$ be matrices of the same size. The **matrix sum** of A and B is the matrix C defined by

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

Theorem 1.2. *[Properties of Matrix Addition] Let A, B, and C denote arbitrary matrices such that the following operations are defined. Then the following hold.*

- $A + B = B + A$,
- $A + (B + C) = (A + B) + C$,
- *for each A, there exists 0 such that $0 + A = A$, and*
- *for each A, there exists a matrix $-A$ such that $A + (-A) = 0$.*

Proof.   • Let A and B denote matrices of the same size with entries A_{ij} and B_{ij}, respectively. Then

$$\begin{aligned} (A + B)_{ij} &= A_{ij} + B_{ij} &&\text{definition of matrix addition} \\ &= B_{ij} + A_{ij} &&\text{scalar commutativity} \\ &= (B + A)_{ij} &&\text{definition of matrix addition} \end{aligned}$$

Therefore, by the definition of matrix equality, $A + B = B + A$ follows.

- If matrix C is also of the same size as A and B with entries C_{ij} then,

$$
\begin{aligned}
(A + (B + C))_{ij} &= A_{ij} + (B + C)_{ij} && \text{definition of matrix addition} \\
&= A_{ij} + (B_{ij} + C_{ij}) && \text{definition of matrix addition} \\
&= (A_{ij} + B_{ij}) + C_{ij} && \text{scalar associativity} \\
&= (A + B)_{ij} + C_{ij} && \text{definition of matrix addition} \\
&= ((A + B) + C)_{ij} && \text{definition of matrix addition}
\end{aligned}
$$

  Therefore, by the definition of matrix equality, $A + (B + C) = (A + B) + C$ follows.

- Let A be an $m \times n$ matrix with entires A_{ij} and let $0_{m \times n}$ denote the $m \times n$ matrix with all entires equal to the scalar additive identity denoted by 0. Denoting the entries of $0_{m \times n}$ by 0_{ij} we find

$$
(0 + A)_{ij} = 0_{ij} + A_{ij} = A_{ij}.
$$

Therefore, by the definition of matrix equality, $0 + A = A$ follows. - The proof of this part is left for the reader as Exercise 1.66.

$\square$

The matrix properties listed in 1.2 are called the

- **commutative** ,
- **associative** ,
- **additive identity** , and

- **additive inverse**
  properties for matrix addition, respectively.

Just as matrix addition has several properties in common with the underlying field of scalars so too does scalar multiplication.

Definition 1.6. If A is any matrix and k is any scalar, the **scalar multiple** kA is the matrix obtained by multiplying each entry of A by k, that is if $A = [a_{ij}]$, then $kA = [ka_{ij}]$.

The following properties say that scalers and matrices distribute over each other and are associative with each other.

Theorem 1.3. *[Properties of Scalar Multiplication] Let A and B denote arbitrary matrices such that the following operations are defined. Let k and p denote arbitrary scalars. Then the following hold.*

- $k(A + B) = kA + kB$
- $(k + p)A = kA + pA$
- $(kp)A = k(pA)$
- $A(kB) = (kA)B$

Proof. • Let A and B be $n \times m$ matrices. Then, for $1 \leq i \leq n$ and $1 \leq j \leq m$ and an arbitrary scalar k we find,

$$
\begin{aligned}
(k(A + B))_{ij} &= k\left((A + B)_{ij}\right) & \text{definition of scalar multiplication} \\
&= k(A_{ij} + B_{ij}) & \text{definition matrix addition} \\
&= (kA)_{ij} + (kB)_{ij} & \text{distributive property of scalar multiplication} \\
&= (kA + kB)_{ij} & \text{definition of matrix addition}
\end{aligned}
$$

Therefore, by the definition of matrix equality, $k(A + B) = kA + kB$ follows.

Let A be an $n \times m$ matrix and let k and p be scalars. Then we find,

$$
\begin{aligned}
((kp)A)_{ij} &= (kp)A_{ij} & \text{definition of scalar multiplication} \\
&= k(pA_{ij}) & \text{associative property of scalars} \\
&= k(pA)_{ij} & \text{definition of scalar multiplication} \\
&= (k(pA))_{ij} & \text{definition of scalar multiplication}
\end{aligned}
$$

Therefore, by the definition of matrix equality, $(kp)A = k(pA)$ follows. The proof of the remaining parts are left for the reader as Exercise 1.67.

$\square$

Example 1.12. If $A = \begin{bmatrix} 3 & -1 & 4 \\ 2 & 0 & 6 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 2 & -1 \\ 0 & 3 & 2 \end{bmatrix}$, determine $5A$, $\frac{1}{2}B$, and $3A - 2B$. A quick computation reveals the following.

$$
5A = \begin{bmatrix} 15 & -5 & 20 \\ 10 & 0 & 30 \end{bmatrix} \quad \frac{1}{2}B = \begin{bmatrix} \frac{1}{2} & 1 & -\frac{1}{2} \\ 0 & \frac{3}{2} & 1 \end{bmatrix} \quad 3A - 2B = \begin{bmatrix} 7 & -7 & 14 \\ 6 & -6 & 14 \end{bmatrix}
$$

Example 1.13. Let A, B, and C be matrices of the same size. Simplify

$$
2(A + 3C) - 3(2C - B) - 3[2(2A + B - 4C) - 4(A - 2C)].
$$

We obtain $2A - 3B$.

Example 1.14. If $kA = 0$, show that either $k = 0$ or $A = 0$. Suppose $k \neq 0$ and $A = [a_{ij}]$, then $kA = [ka_{ij}] = 0$ and notice each entry ka_{ij} is zero if and only if $a_{ij} = 0$ for every i and j. Therefore, either $k = 0$ or every entry in A is zero.

Example 1.15. Find 1×3 matrices X and Y such that

$$\begin{cases} X + 2Y = \begin{bmatrix} 1 & 3 & -2 \end{bmatrix} \\ X + Y = \begin{bmatrix} 2 & 0 & 1 \end{bmatrix}. \end{cases}$$

Subtracting these equations together yields $Y = \begin{bmatrix} -1 & 3 & -3 \end{bmatrix}$. Then $X = \begin{bmatrix} 2 & 0 & 1 \end{bmatrix}$ and $-Y = \begin{bmatrix} 3 & -3 & 4 \end{bmatrix}$.

Let A be an $m \times n$ matrix and B a $n \times p$ matrix, their **matrix product** $C = AB$ is defined as the the the $m \times p$ matrix whose (i, j)th entry is the dot product of the ith row of A with the jth column of B, that is, the product of two matrices is the matrix C where $c_{ij} = \sum_{k=1}^{n} a_{ik}b_{kj}$. Matrix multiplication is illustrated by the following equation.

$$AB = \begin{bmatrix} a_{ij} \end{bmatrix} \begin{bmatrix} b_{ij} \end{bmatrix} = \begin{bmatrix} \sum_{k=1}^{n} a_{1k}b_{k1} & \cdots & \sum_{k=1}^{n} a_{1k}b_{kp} \\ \vdots & \sum_{k=1}^{n} a_{ik}b_{kj} & \vdots \\ \sum_{k=1}^{n} a_{mk}b_{k1} & \cdots & \sum_{k=1}^{n} a_{mk}b_{kn} \end{bmatrix} = \begin{bmatrix} c_{ij} \end{bmatrix} = C$$

Example 1.16. Let A and B be defined as follows.

$$A = \begin{bmatrix} 4 & -5 & 6 \\ 1 & 8 & -9 \end{bmatrix} \qquad B = \begin{bmatrix} 1 & 2 & -3 \\ -2 & -1 & 5 \\ -2 & 5 & -1 \end{bmatrix}$$

If possible, find AB and BA. Since A is a 2×3 matrix and B is a 3×3 matrix we see that AB is defined as a 2×3 matrix and BA is not defined.

$$AB = \begin{bmatrix} \begin{bmatrix} 4 & -5 & 6 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ -2 \\ -2 \end{bmatrix} & \begin{bmatrix} 4 & -5 & 6 \end{bmatrix} \cdot \begin{bmatrix} 2 \\ -1 \\ 5 \end{bmatrix} & \begin{bmatrix} 4 & -5 & 6 \end{bmatrix} \cdot \begin{bmatrix} -3 \\ 5 \\ -1 \end{bmatrix} \\ \begin{bmatrix} 1 & 8 & -9 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ -2 \\ -2 \end{bmatrix} & \begin{bmatrix} 1 & 8 & -9 \end{bmatrix} \cdot \begin{bmatrix} 2 \\ -1 \\ 5 \end{bmatrix} & \begin{bmatrix} 1 & 8 & -9 \end{bmatrix} \cdot \begin{bmatrix} -3 \\ 5 \\ -1 \end{bmatrix} \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$$

1.16 is a counterexample showing that matrix multiplication is not commutative. Even so, matrix multiplication has many useful properties such as associativity, multiplicative inverses, and distributive properties.

::: {#thm- } [Properties of Matrix Multiplication] Let A, B, and C denote arbitrary matrices such that the following operations are defined. Then the following hold.

- $A(BC) = (AB)C$
- $A(B + C) = AB + AC$
- $(B + C)A = BA + CA$
- $AI = A$
- $IA = A$ :::

Proof.   • Suppose B is an $n \times p$ matrix. Since BC is defined, C is a matrix with p rows, and BC has n rows. Because $A(BC)$ is defined we may assume A is an $m \times n$ matrix. Thus the product AB exists and it is an $m \times p$ matrix, from which it follows that the product $(AB)C$ exists. To show that $A(BC) = A(BC)$ we must show that

$$[A(BC)]_{ij} = [(AB)C]_{ij}. \tag{1.11}$$

for arbitrary i and j. By definition of matrix multiplication,

$$[A(BC)]_{ij} = \sum_t A_{it}(BC)_{tj} = \sum_t A_{it} \sum_s B_{ts}C_{sj}$$
$$= \sum_t \sum_s A_{it}B_{ts}C_{sj} = \sum_s \sum_t A_{it}B_{ts}C_{sj}$$
$$= \sum_s \left(\sum_t A_{it}B_{ts} \right) C_{sj} = \sum_s (AB_{is})C_{sj} = [(AB)C]_{ij}.$$

- Let A_{ij}, B_{ij}, and C_{ij} denote the entries of matrices A, B and C of sizes $m \times n$, $n \times p$, and $n \times p$, respectively. Then $B + C = [B_{ij} + C_{ij}]$, so the (i, j)th entry of $A(B + C)$ is

$$\sum_{k=1}^n A_{ik}(B_{kj} + C_{kj}) = \sum_{k=1}^n (A_{ik}B_{kj} + A_{ik}C_{kj}) = \sum_{k=1}^n A_{ik}B_{kj} + \sum_{k=1}^n A_{ik}C_{kj}$$

this is the (i, j)th entry of $AB + AC$ because the sums on the right are the (i, j)th entries of AB and AC, respectively. Hence $A(B + C) = AB + AC$.

Let $A = [a_{ij}]_{m \times n}$. Then

$$AI = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \ddots & \ddots & \vdots \\ a_{m_1} & a_{m2} & \cdots & a_{mm} \end{bmatrix}_{m \times n} \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}_{n \times n}.$$

and the (i, j)th entry of AI is the ith row of A times the jth column of I, that is

$$\begin{bmatrix} a_{i1} & a_{i2} & \cdots & a_{ij} & \cdots & a_{in} \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix} \quad \leftarrow \text{Position } j$$

So the (i, j)th entry of AI is a_{ij}, the (i, j)th entry of A. Thus, $AI = A$. The proof of the remaining parts are left for the reader as Exercise 1.68.

□

The matrix properties listed in 1.7 are called the

- **associative** ,
- **left distributive** ,
- **right distributive** ,
- **left identity** , and

- **right identity**
  property for matrix multiplication, respectively.

Example 1.17. Find all matrices that commute with $A = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}$. Let $B = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ be a matrix that commutes with A. Matrix multiplication yields

$$AB = \begin{bmatrix} a + 2c & b + 2d \\ c & d \end{bmatrix} \quad \text{and} \quad BA = \begin{bmatrix} a & 2a + b \\ c & 2c + d \end{bmatrix}.$$

Since $a + 2c = a$ it follows that $c = 0$. Since $b + 2d = 2a + b$ it follows that $a = d$. Thus all matrices that commute with A have the form $\begin{bmatrix} a & b \\ 0 & a \end{bmatrix}$.

Example 1.18. Write the matrix equation for the system

$$\begin{cases} 3x_1 + x_2 + 7x_3 + 2x_4 = 13 \\ 2x_1 - 4x_2 + 14x_3 - x_4 = -10 \\ 5x_1 + 11x_2 - 7x_3 + 8x_4 = 59 \\ 2x_1 + 5x_2 - 4x_3 - 3x_4 = 39. \end{cases}$$

Find the reduced row echelon form using Gauss-Jordan elimination. Solve the system. The matrix equation is

$$\begin{bmatrix} 3 & 1 & 7 & 2 \\ 2 & -4 & 14 & -1 \\ 5 & 11 & -7 & 8 \\ 2 & 5 & -4 & -3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 13 \\ -10 \\ 59 \\ 39 \end{bmatrix}.$$

The reduced row echelon form of the coefficient matrix and the solution set are as follows

$$\begin{bmatrix} 1 & 0 & 3 & 0 & 4 \\ 0 & 1 & -2 & 0 & 5 \\ 0 & 0 & 0 & 1 & -2 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \qquad \left\{ \begin{bmatrix} 4 - 3t \\ 5 + 2c \\ t \\ -2 \end{bmatrix} \,\middle|\, t \in \mathbb{R} \right\}$$

A matrix with a single column is called a **column vector** or just and a matrix with a single row is called a **row vector**.

Definition 1.7. A vector $\vec{b}$ is called a **linear combination** of the vectors $v_1, v_2, \ldots, v_m$ if there exists scalars $a_1, a_2, \ldots, a_m$ such that

$$\vec{b} = a_1 \vec{v}_1 + a_2 \vec{v}_2 + \cdots + a_m \vec{v}_m.$$

Example 1.19. Given the following vectors, is $\vec{u}$ a linear combination of the vectors $\vec{v}_1$ and $\vec{v}_2$?

$$\vec{u} = \begin{bmatrix} 7 \\ 8 \\ 9 \end{bmatrix} \qquad \vec{v}_1 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \qquad \vec{v}_2 = \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix}$$

We want to find scalars x_1 and x_2 such that $\vec{u} = x_1 \vec{v}_1 + x_2 \vec{v}_2$. We solve for x_1 and x_2 using an augmented matrix and row-operations as follows.

$$\begin{bmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{bmatrix} \begin{array}{c} \xrightarrow{} \\ -2R_1 + R_2 \\ \xrightarrow{} \\ -3R_1 + R_3 \end{array} \begin{bmatrix} 1 & 4 & 7 \\ 0 & -3 & -6 \\ 0 & -6 & -12 \end{bmatrix} \begin{array}{c} \xrightarrow{} \\ -\frac{1}{3}R_2 \\ \xrightarrow{} \\ -\frac{1}{6}R_3 \end{array} \begin{bmatrix} 1 & 4 & 7 \\ 0 & 1 & 2 \\ 0 & 1 & 2 \end{bmatrix}$$

$$\begin{array}{c} \xrightarrow{} \\ -R_2 + R_3 \end{array} \begin{bmatrix} 1 & 4 & 7 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{bmatrix} \begin{array}{c} \xrightarrow{} \\ -4R_2 + R_1 \end{array} \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{bmatrix}$$

Thus we find $x_1 = -1$ and $x_2 = 2$. Therefore, yes $\vec{u}$ is a linear combination of $\vec{v}_1$ and $\vec{v}_2$, namely $\vec{u} = 2\vec{v}_2 - \vec{v}_1$.

Example 1.20. Given the following vectors, determine whether $\vec{u}$ or $\vec{w}$ is a linear combination of the vectors $\vec{v}_1$ and $\vec{v}_2$.

$$\vec{u} = \begin{bmatrix} 1 \\ 1 \\ 4 \end{bmatrix} \qquad \vec{w} = \begin{bmatrix} 1 \\ 5 \\ 1 \end{bmatrix} \qquad \vec{v}_1 = \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix} \qquad \vec{v}_2 = \begin{bmatrix} 3 \\ 5 \\ 2 \end{bmatrix}$$

First, $\vec{u}$ is a linear combination of $\vec{v}_1$ and $\vec{v}_2$ since

$$\begin{bmatrix} 1 \\ 1 \\ 4 \end{bmatrix} = (-2) \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix} + (1) \begin{bmatrix} 3 \\ 5 \\ 2 \end{bmatrix}.$$

To determine whether $\vec{w}$ is a linear combination of $\vec{v}_1$ and $\vec{v}_2$ we consider the equation

$$\vec{w} = x \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix} + y \begin{bmatrix} 3 \\ 5 \\ 2 \end{bmatrix}$$

which leads to the system

$$\begin{cases} x + 3y = 1 \\ 2x + 5y = 5 \\ -x + 2y = 1. \end{cases}$$

This linear system can be shown to have an empty solution set; and thus $\vec{w}$ is not a linear combination of $\vec{v}_1$ and $\vec{v}_2$.

Example 1.21. For which values of the constant c is $\begin{bmatrix} 1 \\ c \\ c^2 \end{bmatrix}$ a linear com-

bination of $\begin{bmatrix} 1 \\ a \\ a^2 \end{bmatrix}$ and $\begin{bmatrix} 1 \\ b \\ b^2 \end{bmatrix}$, where a and b are arbitrary constants? We need to solve the following linear system

$$\begin{bmatrix} 1 \\ c \\ c^2 \end{bmatrix} = x \begin{bmatrix} 1 \\ a \\ a^2 \end{bmatrix} + y \begin{bmatrix} 1 \\ b \\ b^2 \end{bmatrix} \qquad \text{with augmented matrix} \qquad \left[\begin{array}{cc|c} 1 & 1 & 1 \\ a & b & c \\ a^2 & b^2 & c^2 \end{array} \right].$$

using row operations the augmented matrix reduces to

$$\left[\begin{array}{cc|c} 1 & 1 & 1 \\ 0 & b - a & c - a \\ 0 & 0 & (c-a)(c-b) \end{array} \right].$$

This system is consistent if and only if $c = a$ or $c = b$. Thus the vector is a linear combination if $c = a$ or $c = b$.

Example 1.22. Express the vector $\begin{bmatrix} 7 \\ 11 \end{bmatrix}$ as the sum of a vector on the line $y = 3x$ and a vector on the line $y = x/2$. We wish to find x_1 and x_2 such that

$$\begin{bmatrix} 7 \\ 11 \end{bmatrix} = \begin{bmatrix} x_1 \\ 3x_1 \end{bmatrix} + \begin{bmatrix} x_2 \\ \frac{1}{2}x_2 \end{bmatrix}.$$

We solve this linear system using row-operations as follows.

$$\begin{bmatrix} 1 & 1 & 7 \\ 3 & \frac{1}{2} & 11 \end{bmatrix} \xrightarrow{-3R_1 + R_2} \begin{bmatrix} 1 & 1 & 7 \\ 0 & \frac{-5}{2} & -10 \end{bmatrix}$$

$$\xrightarrow{(-2/5)R_2} \begin{bmatrix} 1 & 1 & 7 \\ 0 & 1 & 4 \end{bmatrix} \xrightarrow{-R_2 + R_1} \begin{bmatrix} 1 & 0 & 7 \\ 0 & 1 & 4 \end{bmatrix}$$

The solution is $x_1 = 3$ and $x_2 = 4$. Therefore the desired sum is

$$\begin{bmatrix} 7 \\ 11 \end{bmatrix} = \begin{bmatrix} 3 \\ 9 \end{bmatrix} + \begin{bmatrix} 4 \\ 1 \end{bmatrix}$$

since $\begin{bmatrix} 3 \\ 9 \end{bmatrix}$ is on the line $y = 3x$ and $\begin{bmatrix} 4 \\ 1 \end{bmatrix}$ is on the line $y = x/2$.

Theorem 1.4. *Let B be an $n \times p$ matrix and A and $p \times m$ matrix with columns $\vec{v}_1$, ..., $\vec{v}_m$. Then the product BA is*

$$BA = B \begin{bmatrix} \vec{v}_1 & \cdots & \vec{v}_m \end{bmatrix} = \begin{bmatrix} B\vec{v}_1 & \cdots & B\vec{v}_m \end{bmatrix}.$$

::: *{.proof } Notice that $\vec{v}_1$, ..., $\vec{v}_m$ are all $p \times 1$ column vectors consisting of the columns of A. So the product of B and each $\vec{v}_i$ $(0 \leq i \leq m)$ are the $n \times 1$ column vectors constituting BA.*

:::

Theorem 1.5. *If the column vectors of an $n \times m$ matrix A are $\vec{v}_1, \vec{v}_2, ..., \vec{v}_n$ and $\vec{x}$ is a vector with entries $x_1, x_2, ..., x_m$, then*

$$A\vec{x} = x_1\vec{v}_1 + x_2\vec{v}_2 + \cdots + x_m\vec{v}_m.$$

Proof. The proof follows from the following equation.

$$A\vec{x} = \begin{bmatrix} | & & | \\ \vec{v}_1 & \cdots & \vec{v}_m \\ | & & | \end{bmatrix} \vec{x} = \begin{bmatrix} v_{11} & \cdots & v_{1m} \\ \vdots & \ddots & \vdots \\ v_{n1} & \cdots & v_{nm} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix}$$

$$= \begin{bmatrix} v_{11}x_1 + v_{12}x_2 + \cdots + v_{1m}x_m \\ \vdots \\ v_{n1}x_1 + v_{n2}x_2 + \cdots + v_{nm}x_m \end{bmatrix}_{n \times 1} = x_1\vec{v}_1 + x_2\vec{v}_2 + \cdots + x_m\vec{v}_m$$

□

Corollary 1.1. *Let A be an $n \times m$ matrix, $\vec{x}$ and $\vec{y}$ be $m \times 1$ column vectors, and let k be a scalar. Then the following hold.*

- $A(\vec{x} + \vec{y}) = A(\vec{x}) + A(\vec{y})$
- $A(k\vec{x}) = kA(\vec{x})$

Proof. Since the column vectors $\vec{x}$ and $\vec{y}$ are matrices and the matrices $A(\vec{x}+\vec{y})$, $A(\vec{x})+A(\vec{y})$, $A(k\vec{x})$, and $kA(\vec{x})$ are defined we can apply 1.7 to obtain $A(\vec{x}+\vec{y}) = A(\vec{x})+A(\vec{y})$ and $A(k\vec{x}) = kA(\vec{x})$.

□

1.8 Exercises

Exercise 1.33. Let A be a matrix of size 4×5, let B be a matrix of size 6×4, let C be a matrix of size 4×5, and let D be a matrix of size 4×2. Which of the following are defined, and for those that are, what is their size?

- BA
- DA
- $B + A$
- $C + A$
- $C(AB)$
- $(BA)C$
- $C + DB$
- $D(C + A)$
- $A + CB$

Exercise 1.34. Let $A = \begin{bmatrix} 2 & 1 & 3 \\ 4 & 0 & -1 \\ -1 & 1 & 0 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & -1 & 1 \\ 0 & 2 & 0 \\ 4 & 3 & 2 \end{bmatrix}$. Where possible, find the following matrices.

- $A + B$
- $B - A$
- $3B$
- $-2A$
- $AB - BA$
- $-2BA$
- $2A - 4B$

- $B^2 - A^2$

Exercise 1.35. Find $a, b, c,$ and d that satisfies the equation.

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

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

Exercise 1.36. Let $A = \begin{bmatrix} 2 & 1 \\ 0 & -1 \end{bmatrix}$, $B = \begin{bmatrix} 3 & -1 & 2 \\ 0 & 1 & 4 \end{bmatrix}$, $C = \begin{bmatrix} 3 & -1 \\ 2 & 0 \end{bmatrix}$,

and $D = \begin{bmatrix} 1 & 3 \\ -1 & 0 \\ 1 & 4 \end{bmatrix}$. Where possible, find the following

- $3A - 2B$
- $5C$
- $B + D$
- $2B - 3D$
- $A - D$
- $A - 2C$

Exercise 1.37. Find A in terms of B.

- $A + B = 3A + 2B$
- $2A - B = 5(A + 2B)$

Exercise 1.38. Simpify the following expressions where A, B, and C are matrices.

- $2[9(A - B) + 7(2B - A)] - 2[3(2B + A) - 2(A + 3B)]$
- $5[3(A - B + 2C) - 2(3C - B) - A] + 2[3(3A - B + C) + 2(B - 2A) - 2C]$

Exercise 1.39. Show that if $Q + A = A$ holds for every $m \times n$ matrix A, then $Q = 0_{mn}$.

Exercise 1.40. Show that if A is an $m \times n$ matrix with $A + A' = 0_{mn}$, then $A' = -A$.

Exercise 1.41. Show that if A denotes an $m \times n$ matrix, then $A = -A$ if and only if $A = 0$.

Exercise 1.42. Given the following vectors, determine the values of a and b such that $\vec{u}$ is a linear combination of $\vec{v_1}$ and $\vec{v_2}$.

$$\vec{u} = \begin{bmatrix} 5 \\ 7 \\ a \\ b \end{bmatrix} \qquad \vec{v_1} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \qquad \vec{v_2} = \begin{bmatrix} 4 \\ 3 \\ 2 \\ 1 \end{bmatrix}$$

Exercise 1.43. Find all solutions x_1, x_2, x_3 of the equation $\vec{b} = x_1\vec{v_1} + x_2\vec{v_2} + x_3\vec{v_3}$ given the following vectors.

$$\vec{b} = \begin{bmatrix} -8 \\ -1 \\ 2 \\ 15 \end{bmatrix} \qquad \vec{v_1} = \begin{bmatrix} 1 \\ 4 \\ 7 \\ 5 \end{bmatrix} \qquad \vec{v_2} = \begin{bmatrix} 2 \\ 5 \\ 8 \\ 3 \end{bmatrix} \qquad \vec{v_3} = \begin{bmatrix} 4 \\ 6 \\ 9 \\ 1 \end{bmatrix}$$

Exercise 1.44. Determine the value of a such that $\vec{u}$ is a linear combination of $\vec{v_1}$ and $\vec{v_2}$ given the following vectors.

$$\vec{u} = \begin{bmatrix} 1 \\ a \\ a^2 \end{bmatrix} \qquad \vec{v_1} = \begin{bmatrix} 1 \\ 2 \\ 4 \end{bmatrix} \qquad \vec{v_2} = \begin{bmatrix} 1 \\ 3 \\ 9 \end{bmatrix}$$

Exercise 1.45. Find the product of the given matrices.

- $\begin{bmatrix} 1 & -1 & 2 \\ 2 & 0 & 4 \end{bmatrix}$
- $\begin{bmatrix} 2 & 3 & 1 \\ 1 & 9 & 7 \\ -1 & 0 & 2 \end{bmatrix}$
- $\begin{bmatrix} 1 & 3 & -3 \end{bmatrix}$
- $\begin{bmatrix} 3 & 0 \\ -2 & 1 \\ 0 & 6 \end{bmatrix}$

Exercise 1.46. Let $A = \begin{bmatrix} 2 & 3-i \\ 1 & i \end{bmatrix}$, $B = \begin{bmatrix} i & 1-1 \\ 0 & i \end{bmatrix}$, and $C = \begin{bmatrix} 2+1 & 1 \\ 3 & i+1 \end{bmatrix}$. Find each of the following.

- $A + B + C$
- $AB + AC$

- $A + BC$
- CB
- A^2C
- C^2A

Exercise 1.47. Find all matrices that commute with the given matrix.

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

Exercise 1.48. For what values of the constants a, b, c and d is $\vec{b}$ a linear combination of the vectors $\vec{v}_1$, $\vec{v}_2$, $\vec{v}_3$ given

$$\vec{b} = \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} \qquad \vec{v}_1 = \begin{bmatrix} 0 \\ 0 \\ 3 \\ 0 \end{bmatrix} \qquad \vec{v}_2 = \begin{bmatrix} 1 \\ 0 \\ 4 \\ 0 \end{bmatrix} \qquad \vec{v}_3 = \begin{bmatrix} 2 \\ 0 \\ 5 \\ 6 \end{bmatrix} ?$$

Exercise 1.49. If $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ where $a \neq 0$, show that A factors in the form

$$A = \begin{bmatrix} 1 & 0 \\ x & 1 \end{bmatrix} \begin{bmatrix} y & z \\ 0 & w \end{bmatrix}.$$

Exercise 1.50. Find all vectors in $\mathbb{R}^4$ that are perpendicular to the following three vectors.

$$u = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \qquad v = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix} \qquad w = \begin{bmatrix} 1 \\ 9 \\ 9 \\ 7 \end{bmatrix}$$

Exercise 1.51. Determine a scalar t such that $AX = tX$ where $A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$ and $X = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$.

Exercise 1.52. Show that if A and B commute with C, then so does $A + B$.

Exercise 1.53. Show that if A and B are $n \times n$ matrices, then $AB = BA$ if and only if $(A + B)^2 = A^2 + 2AB + B^2$.

Exercise 1.54. Let A and B be the 2×2 matrices

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

whose entries are themselves 2×2 matrices. Where possible, find each of the following.

- $A + B$
- $-2B$
- $3A$
- $2A - 3B$
- B^2
- AB
- BA
- $AB - BA$
- $A^2 - B^2$

Exercise 1.55. Let $A = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}$. Find an expression for

- A^2,
- A^3, and
- A^n where n is a positive integer.

Exercise 1.56. The Pauli spin matrices

$$
P_1 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad P_2 = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, \quad P_3 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}
$$

are used when studying electron spin. Find

- $P_1 P_2,$

- $P_2 P_1$,
- P_2^2,
- $P_1 P_3$,

- $P_3 P_1$, and

- $P_1 + iP_2$.

Exercise 1.57. Let A denote an arbitrary matrix and let k denote an arbitrary scalar. Prove the following properties hold.

- $A - A = 0$
- $0A = 0$
- $0A = 0$
- $kIA = kA$

Exercise 1.58. Find $A\vec{e}_1$, $A\vec{e}_2$, and $A\vec{e}_3$ given the following.

$$\vec{e}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \quad \vec{e}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \quad \vec{e}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \quad \text{and} \quad A = \begin{bmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{bmatrix}$$

Exercise 1.59. Find a 3×3 matrix A that satisfies all of the following.

$$A\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \quad A\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix} \quad \text{and} \quad A\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 7 \\ 8 \\ 9 \end{bmatrix}$$

Exercise 1.60. Find all vectors $\vec{x}$ such that $A\vec{x} = \vec{b}$ given the following.

$$A = \begin{bmatrix} 1 & 2 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \quad \text{and} \quad \vec{b} = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}.$$

Exercise 1.61. Write the system

$$\begin{cases} 2x_1 - 3x_2 + 5x_3 = 7 \\ 9x_1 + 4x_2 - 6x_3 = 8 \end{cases}$$

in matrix form and write $\begin{bmatrix} 7 \\ 8 \end{bmatrix}$ as a linear combination of the columns vectors of the coefficient matrix.

Exercise 1.62. Show that the sum of any two diagonal matrices is diagonal.

Exercise 1.63. Show that the sum of any two upper triangular matrices is upper triangular.

Exercise 1.64. Let A be an $m \times n$ matrix and let

$$B = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} \quad \text{and} \quad C = \begin{bmatrix} c_1 & c_2 & \cdots & c_m \end{bmatrix}$$

be a column vector and a row vector, respectively. Prove that $AB = \sum_{j=1}^{n} b_j A_j$ and $CA = \sum_{j=1}^{n} c_j A_j$.

Exercise 1.65. Let A and B denote $n \times n$ matrices. The **Jordan product**, $A \star B$ is defined by $A \star B = \frac{1}{2}(AB + BA)$. Determine whether this product is commutative, associative, and/or distributive.

Exercise 1.66. Complete the proof of 1.2.

Exercise 1.67. Complete the proof of 1.3.

Exercise 1.68. Complete the proof of 1.7.

1.9 On the Solutions of Linear Systems

Recall a system of linear equations is called consistent if it has at least one solution and is called inconsistent if it has no solutions. In this section we completely characterize when a linear system of equations has solutions; and we do so using the notion of *rank*. Basically, the rank of a linear system is the number of leading coefficients in the reduced row echelon form of the augmented matrix of the given linear system. Our first goal will be to show the notion of rank is well-defined; that is, we wish to show that every matrix has a unique reduced row echelon form. Thus the rank of a linear system will be a unique number.

Find a 2×3 linear system whose augmented matrix has two different row echelon forms.

Definition 1.8. If an matrix A can be obtained from another matrix B by a finite number of elementary row operations, then we say B is **row equivalent** to A.

Theorem 1.6. *Every matrix is row equivalent to a matrix in row echelon form.*

Proof. Let A be an $m \times n$ nonzero matrix, with entries a_{ij}, say

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

Either all entries in the first column of A are nonzero or not. If all entries are zero then the first column of A satisfies the conditions of row echelon form. Otherwise, assume i is the least such that a_i is nonzero. Interchanging rows 1 and i (if needed) we obtain a column where the first entry is nonzero. Now we have a matrix of the following form.

$$i\text{-th row} \rightarrow \begin{bmatrix} a_{i1} & a_{i2} & \cdots & a_{in} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ & & \vdots & \\ a_{11} & a_{12} & \cdots & a_{1n} \\ & & \vdots & \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

Next we multiply the first row by the multiplicative inverse of its nonzero entry, obtaining 1. For the remaining first column entries $a_{k,m}$ (where $k > 1$) multiply the row by the multiplicative inverse of $a_{k,m}$ and add to the first row replacing the k-th row. These steps lead to all zero entries in the first column below the leading coefficient. Our matrix is now in the form

$$\begin{bmatrix} 1 & a'_{i2} & \cdots & a'_{in} \\ 0 & a'_{22} & \cdots & a'_{2n} \\ & & \vdots & \\ 0 & a'_{m2} & \cdots & a'_{mn} \end{bmatrix}$$

where the a'_{ij} are the scalars obtain from completing the row operations. We repeat this process on the remaining columns taking into account that applying row operations will not change the fact that the previous columns will continue to satisfy the conditions of row echelon form.

□

Theorem 1.7. *Every matrix is row equivalent to a unique matrix in reduced row echelon form.*

Proof. Let A be an $m \times n$ matrix. We apply mathematical induction on n for an arbitrary m. Let $n = 1$. Now A is just a matrix with one column and is row equivalent to one of the following matrices.

$$\begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad \text{or} \quad \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

So the case for $n = 1$ is clear. Now assume $n > 1$ and let A and A' denote the following matrices.

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ & & \vdots & \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \quad \text{and} \quad A' = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1,n-1} \\ a_{21} & a_{22} & \cdots & a_{2,n-1} \\ & & \vdots & \\ a_{m1} & a_{m2} & \cdots & a_{m,n-1} \end{bmatrix}$$

Notice A' is just A with the n-th column deleted. By the way A' is defined, any sequence of elementary row operations that takes A into reduced row echelon form also takes A' into reduced row echelon form. By the induction hypothesis, any two matrices B and C that are reduced row echelon forms of A can only differ in the n column. Assume, for a contradiction, that B and C differ only in the n-th column. Then there exists an integer j such that the j-th row of B is not equal to the j-th row of C.

$\square$

Due to 1.7 the following definition is well-defined; meaning if a matrix A is reduced to the unique matrix in reduced row echelon form

Definition 1.9. The **rank** of a matrix A is the number of leading coefficients in $\mathrm{rref}(A)$.

Example 1.23. Let a, d, f be nonzero constants and let b, c, e be arbitrary constants. Find the rank of the following matrices.

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

Since a, f, and d are all nonzero,

$$\mathrm{rref}(A) = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \qquad \mathrm{rref}(B) = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}, \qquad \mathrm{rref}(C) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Thus the rank of A, B, and C is 1, 2, and 3, respectively.

Lemma 1.1. *Let $A\vec{x} = \vec{b}$ and $C\vec{x} = \vec{d}$ be two linear systems of the same size. If the augmented matrices $\begin{bmatrix} A \mid \vec{b} \end{bmatrix}$ and $\begin{bmatrix} C \mid \vec{d} \end{bmatrix}$ are row equivalent, then the linear systems are equivalent.*

Proof. The proof follows immediately from 1.1.

$\square$

Theorem 1.8. *[Fundamental Theorem of Linear Systems] Let A be the coefficient matrix of an $n \times m$ system. Then*

- *If $rank(A) = n$, then the system is consistent.*
- *If $rank(A) = m$, then the system has at most one solution.*
- *If $rank(A) < m$, then the system has either infinitely many solutions or none.*

Proof. First let's make two observations. By definition of the reduced row echelon form of a matrix, there is at most one leading 1 in each of the m columns and in each of the n columns. Let A be the coefficient matrix of an n by m system. Then

$$\text{rank}(A) \leq n \qquad \text{and} \qquad \text{rank}(A) \leq m.$$

A linear system is inconsistent if and only if the reduced row echelon form of its augmented matrix has a row of the form

$$\begin{bmatrix} 0 & 0 & \cdots & 0 \mid 1 \end{bmatrix}. \tag{1.12}$$

Moreover, for any linear system with m variables,

$$\begin{pmatrix} \text{number of} \\ \text{free} \\ \text{variables} \end{pmatrix} = \begin{pmatrix} \text{total} \\ \text{number of} \\ \text{variables} \end{pmatrix} - \begin{pmatrix} \text{number} \\ \text{of leading} \\ \text{variables} \end{pmatrix} = m - \text{rank}(A)$$

- If $\text{rank}A = n$ then each row must contain a leading 1. Thus there is no row of the form in 1.12 and so the system is consistent.
- If $\text{rank}(A) = m$ then there are no free variables. So the only possible choice is for there to be no solutions or exactly one solution.

- If $\text{rank}(A) = t < m$ then there $m - t$ free variables. So the only possible choice is for there to be no solutions or infinitely many solutions.

$\square$

The following two corollaries are immediate consequences of 1.8.

Corollary 1.2. *A linear system with fewer equations than unknowns has either no solutions or infinity many solutions.*

Proof. The proof is left for the reader as Exercise 1.79.

$\square$

Corollary 1.3. *A linear system of n equations in n variables has a unique solution if and only if the rank of its coefficients matrix A is n, and in this case $rref(A) = I_n$.*

Proof. The proof is left for the reader as Exercise 1.80.

$\square$

Example 1.24. Find the rank of the coefficient matrix and solve the linear system of equations

$$\begin{cases} x_1 - x_2 + x_3 = 4 \\ 3x_1 + 4x_2 - x_3 = 8 \\ 5x_1 + 9x_2 - 4x_3 = 13. \end{cases}$$

We use Gaussian elimination with the augmented matrix to find the rank of the coefficient matrix.

$$\begin{bmatrix} 1 & -1 & 2 & 4 \\ 3 & 4 & -1 & 8 \\ 5 & 9 & -4 & 13 \end{bmatrix} \xrightarrow[\substack{-3R_1 + R_2 \\ -5R_1 + R_3}]{} \begin{bmatrix} 1 & -1 & 2 & 4 \\ 0 & 7 & -7 & -4 \\ 0 & 14 & -14 & -7 \end{bmatrix}$$

$$\xrightarrow[\substack{-\frac{1}{7}R_2 \\ -14R_2 + R_3}]{} \begin{bmatrix} 1 & -1 & 2 & 4 \\ 0 & 1 & -1 & -\frac{4}{7} \\ 0 & 0 & 0 & 1 \end{bmatrix} \xrightarrow[\substack{\frac{4}{7}R_3 + R_2 \\ -4R_3 + R_2 \\ R_2 + R_1}]{} \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

The number of leading 1's is 2 and thus rank$(A) = 2$. By 1.8 the system either has no solutions or infinitely many solutions. The last row represents $0 = 1$, which means that the system has no solution.

Example 1.25. Consider the system

$$\begin{cases} y + 2kz = 0 \\ x + 2y + 6z = 2 \\ kx + 2z = 1 \end{cases}$$

where k is a constant.

- For which values of k does the system have a unique solution?
- When are there no solutions?
- When are there infinitely many solutions?

Using Gaussian elimination:

$$R_1 \leftrightarrow R_2 \begin{bmatrix} 1 & 2 & 6 & | & 2 \\ 0 & 1 & 2k & | & 0 \\ k & 0 & 2 & | & 1 \end{bmatrix} \xrightarrow{-kR_1 + R_2} \begin{bmatrix} 1 & 2 & 6 & | & 2 \\ 0 & 1 & 2k & | & 0 \\ 0 & -2k & -6k+2 & | & -2k+1 \end{bmatrix}$$

$$\xrightarrow[2kR_2 + R_3]{-2R_2 + R_1} \begin{bmatrix} 1 & 0 & -4k+6 & | & 2 \\ 0 & 1 & 2k & | & 0 \\ 0 & 0 & 4k^2 - 6k + 2 & | & -2k+1 \end{bmatrix}$$

Notice $ 4k^2-6k+2=-2(-2k+1)(k-1)=0 $ when $k = 1/2$ and $k = 1$. (a) When $k \neq 1/2$ and $k \neq 1$ there is a unique solution. (b) When $k \neq 1/2$ and $k = 1$, this system has no solutions. (c) When $k = 1/2$ this system has infinitely many solutions.

Example 1.26. Solve the linear system of equations

$$\begin{cases} (3 + i)x_1 + (1 + i)x_2 = 4 + 4i \\ x_1 - x_2 = 2i \end{cases}$$

where x_1 and x_2 are complex variables. We convert the system into a linear system with real variables. Let $x_1 = y_1 + iz_1$ and $x_2 = y_2 + iz_2$. Now substation into the original system leads to the system

$$\begin{cases} 3y_1 - z_1 + (y_1 + 3z_1)i + (y_2 - z_2) + (y_2 + z_2)i = 4 + 4i \\ (y_1 - y_2) + (z_1 - z_2)i = 0 + 2i \end{cases}$$

Equating real and imaginary parts leads to the system

$$\begin{cases} 3y_1 + y_2 - z_1 - z_2 = 4 \\ y_1 + y_2 + 3z_1 + z_2 = 4 \\ y_1 - y_2 = 0 \\ z_1 - z_2 = 2 \end{cases}$$

The solutions are $y_1 = 1$, $y_2 = 1$, $z_1 = 1$, and $z_2 = -1$. Thus the solutions to the original system are $x_1 = 1 + i$ and $x_2 = 1 - i$.

Definition 1.10. A linear system of the form $A\vec{x} = \vec{0}$ is called a **homogeneous** linear system; that is, if all of the constant terms in the linear system are zero.

In particular, if A and B are row equivalent $m \times n$ matrices, then the homogenous systems $A\vec{x} = \vec{0}$ and $B\vec{x} = \vec{0}$ are equivalent.

Lemma 1.2. *If A is an $n \times n$ matrix and the system $A\vec{x} = \vec{0}$ has no nontrivial solution, then A is row equivalent to I_n.*

Proof. If $A\vec{x} = \vec{0}$ has no nontrivial solutions, then the trivial solution is its unique solution. By 1.3, $\mathrm{rref}(A) = I_n$ and by 1.7, A is row equivalent to a unique matrix in reduced row echelon form, thus A is row equivalent to I_n.

$\square$

Lemma 1.3. *Let $A\vec{x} = \vec{0}$ be a linear homogeneous system.*

- *All homogeneous systems are consistent.*
- *A homogeneous system with fewer equations than unknowns has infinitely many solutions.*
- *If $\vec{x}_1$ and $\vec{x}_2$ are solutions, then $\vec{x}_1 + \vec{x}_2$ is also a solution.*
- *If $\vec{x}_1$ is a solution, then $k\vec{x}_1$ is also a solution.*

Proof. The proof is left for the reader as Exercise 1.90.

$\square$

There is a strong relationship between the solutions to a linear system $A\vec{x} = \vec{b}$ and the solutions to the corresponding homogeneous system, $A\vec{x} = \vec{0}$.

Theorem 1.9. *If $\vec{u}$ is any specific solution to the linear system $A\vec{x} = \vec{b}$, then the entire solution set of $A\vec{x} = \vec{b}$ can be described as*

$$\{\vec{u} + \vec{v} \mid \vec{v} \text{ is any solution to } A\vec{x} = \vec{0}\}.$$

Proof. Let $\vec{u}$ be any particular solution to the system $A\vec{x} = \vec{b}$ and let $\vec{v}$ be any solution to the system $A\vec{x} = \vec{0}$. Then

$$A(\vec{u} + \vec{v}) = A\vec{u} + A\vec{v} = \vec{b} + \vec{0} = \vec{b},$$

which shows that every vector of the form $\vec{u} + \vec{v}$ is a solution to the system $A\vec{x} = \vec{b}$. Conversely, let $\vec{w}$ be an arbitrary solution to the system $A\vec{x} = \vec{b}$. Notice

$$A(\vec{w} - \vec{u}) = A\vec{w} - A\vec{u} = \vec{b} - \vec{b} = \vec{0},$$

which shows $\vec{w} - \vec{u}$ is a solution to the system $A\vec{x} = \vec{0}$. Set $\vec{v} = \vec{w} - \vec{u}$, then $\vec{w} = \vec{u} + \vec{w} - \vec{u} = \vec{u} + \vec{v}$ where $\vec{v}$ is a solution to the system $A\vec{x} = \vec{0}$.

□

Corollary 1.4. *If A is an $n \times n$ and $A\vec{x} = \vec{0}$ has no nontrivial solutions, then the system $A\vec{x} = \vec{b}$ has a unique solution.*

Proof. The proof is left for the reader as Exercise 1.91.

□

1.10   Exercises

Exercise 1.69. Find an inconsistent system of two linear equations in three unknowns. Describe the situation geometrically.

Exercise 1.70. Find the rank of the matrix.

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

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

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

- $\begin{bmatrix} 2 & 2 & 2 \\ 2 & 2 & 2 \\ 2 & 2 & 2 \\ 2 & 2 & 2 \end{bmatrix}$

- $\begin{bmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{bmatrix}$

- $\begin{bmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ -3 & -6 & -9 \end{bmatrix}$

Exercise 1.71.

- If the rank of a 4×4 matrix A is 4, what is $\mathrm{rref}(A)$?
- If the rank of a 5×3 matrix A is 3, what is $\mathrm{rref}(A)$?
- If the rank of a 2×5 matrix A is 1, what is $\mathrm{rref}(A)$?

Exercise 1.72. Find the rank of the following matrices.

- $\begin{bmatrix} a & 0 & 0 \\ 0 & c & 0 \\ 0 & 0 & f \end{bmatrix}$

- $\begin{bmatrix} a & b & c \\ 0 & c & d \\ 0 & 0 & f \end{bmatrix}$

- $\begin{bmatrix} a & 0 & 0 \\ b & c & 0 \\ d & e & f \end{bmatrix}$

where a, d, f are nonzero and b, c, and e are arbitrary scalars.

Exercise 1.73. Find the rank of the system of equations.

- $\begin{cases} x + 2y + 3z = 0 \\ 2x + 3y + 4z = 0 \\ 3x + 4y + 6z = 0 \end{cases}$

- $\begin{cases} x + 2y + 3z = 2 \\ 2x + 3y + 4z = -2 \\ 3x + 4y + 6z = 2 \end{cases}$

- $\begin{cases} x + 2y + 3z = 0 \\ 2x + 3y + 4z = 1 \\ 3x + 4y + 6z = 3 \end{cases}$

- $\begin{cases} x + 2y + 3z = a \\ 2x + 3y + 4z = b \\ 3x + 4y + 6z = c \end{cases}$

Exercise 1.74. Find all solutions to the homogenous system.

- $\begin{cases} 4x_1 - x_2 = 0 \\ 7x_1 + 3x_2 = 0 \\ -8x_1 + 6x_2 = 0 \end{cases}$

- $$\begin{cases} x_1 - 2x_2 + x_3 = 0 \\ 3x_1 + 2x_3 + x_4 = 0 \\ 4x_2 - x_3 - x_4 = 0 \\ 5x_1 + 3x_3 - x_4 = 0 \end{cases}$$

- $$\begin{cases} x_1 - 3x_2 = 0 \\ -2x_1 + 6x_2 = 0 \\ 4x_1 - 12x_2 = 0 \end{cases}$$

- $$\begin{cases} x_1 + x_2 - x_3 = 0 \\ 4x_1 - x_2 + 5x_3 = 0 \\ 2x_1 - x_2 - 2x_3 = 0 \\ 3x_1 + 2x_2 - x_3 = 0 \end{cases}$$

Exercise 1.75. Show that the homogenous system of linear equations

$$\begin{cases} ax + by = 0 \\ cx + dy = 0 \end{cases}$$

has an infinite number of solutions if and only if $ad - bc = 0$.

Exercise 1.76. For any positive integer n, find a system of n equations in two variables that has infinitely many solutions.

Exercise 1.77. If possible, find a condition of the coefficients of the homogenous system of linear equations so that (i) the system only has the zero solution, and (ii) the system has infinitely many solutions.

- $$\begin{cases} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 = 0 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 = 0 \end{cases}$$

- $$\begin{cases} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 = 0 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 = 0 \\ a_{31}x_1 + a_{32}x_2 + a_{33}x_3 = 0 \end{cases}$$

Exercise 1.78. Show that if a system of linear equations is inconsistent, then its reduced row echelon form has a row of the form

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

Exercise 1.79. Prove 1.2.

Exercise 1.80. Prove 1.3.

Exercise 1.81. Determine the values of k for which the system has non-trivial solutions.

- $$\begin{cases} 2x_1 - 3x_2 + 5x_3 = 0 \\ -x_1 + 7x_2 - x_3 = 0 \\ 4x_1 - 11x_2 + kx_3 = 0 \end{cases}$$

- $$\begin{cases} x_1 - 2x_2 - 5x_3 = 0 \\ 2x_1 + 3x_2 + x_3 = 0 \\ x_1 - 7x_2 - kx_3 = 0 \end{cases}$$

Exercise 1.82. Show that if $AX = \vec{0}$ is a homogenous system with an infinite number of solutions and the system $AX = B$ has a solution, then $AX = B$ must have an infinite number of solutions.

Exercise 1.83. Find an example, where possible, for each of the following.

- A linear system of four equations in four unknowns that has a line as a solution set.
- A linear system of four equations in four unknowns that has a plane as a solution set.
- A linear system of four equations in three unknowns that has a line as a solution set.
- A linear system of four equations in two unknowns that has a plane as a solution set.

Exercise 1.84. Determine whether or not the following system is consistent.

$$\begin{bmatrix} 0 & i & 1-i \\ -i & 0 & i \\ 1-i & -i & 0 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$$

Exercise 1.85. Show that if $AX = 0$ for all vectors X, then $A = 0$.

Exercise 1.86. Under what conditions will k planes $a_j x + b_j y + c_j z = d_j$ for $j = 1, 2, ..., k$ intersect in exactly one point?

Exercise 1.87. Given that $AX = B$ is consistent and of rank r, for what sets of r unknowns can one solve?

Exercise 1.88. If possible, write the matrix A as a linear combination of the matrices

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}.$$

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

Exercise 1.89. Let $AZ = B$ be a given system of linear equations where $A_{n \times n}$ and $B_{n \times 1}$ are complex matrices. Let $Z = X + iY$. Show that the original complex $n \times n$ system is equivalent to the $2n \times 2n$ real system

$$\begin{cases} CX - DY = S \\ CX + DY = T \end{cases}$$

where $B = S + iT$.

Exercise 1.90. Prove 1.3.

Exercise 1.91. Prove 1.4.

Chapter 2

Vector Spaces

In mathematics, a linear space is a vector space over the real numbers or complex numbers in which every vector has a unique linear combination of basis vectors that satisfies certain properties. Linear spaces are very important in mathematics and physics because many physical problems can be reduced to problems about linear spaces. In this book, we will give a complete introduction to linear spaces.

Linear spaces, also called vector spaces, are mathematical structures that allow the addition of vectors and the multiplication of vectors by scalars. Linear spaces contain a set of vectors, called the basis, which spans the space. The dimension of a linear space is the number of vectors in the basis. Linear spaces are important in physics because they allow physicists to model physical phenomena such as electromagnetism and gravity.

Linear spaces also have applications in engineering and computer science. In engineering, linear spaces are used to design structural elements such as bridges and buildings. In computer science, linear spaces are used to represent data points in multidimensional space.

A linear space has the following properties:

1. Vector addition is commutative: For any two vectors u and v in the linear space, we have $u + v = v + u$.
2. Vector addition is associative: For any three vectors u, v, and w in the linear space, we have $(u + v) + w = u + (v + w)$.
3. There is a zero vector: There is a vector 0 in the linear space such that for any vector u in the linear space, we have $0 + u = u$.
4. Every vector has an additive inverse: For any vector u in the linear space, there is a vector $-u$ such that $u + (-u) = 0$.
5. Multiplication by a scalar is left distributive: For any scalar c and any two vectors u and v in the linear space, we have $c(u + v) =$

$cu + cv$.

6. Multiplication by a scalar is right distributive: For any scalars a and b and any vector u in the linear space, we have $(a + b)u = au + bu$.

7. Multiplication by a scalar is associative: For any three scalars $c, d, and e$, and any vector u in the linear space, we have $(cd)eu = c(de)u$.

8. There is a multiplicative identity: There is a scalar 1 such that for any vector u in the linear space, we have $1u = u$.

A basis for a linear space is a set of vectors that spans the space and is linearly independent. A set of vectors is linearly independent if no vector in the set can be written as a linear combination of the other vectors. The dimension of a linear space is the number of vectors in any basis for the space.

For example, consider the set of all real numbers. This set has infinite dimensions because it takes an infinite number of real numbers to describe any element in the set. On the other hand, the set of all two-dimensional vectors has dimension two because any two-dimensional vector can be written as a linear combination of the vectors (1,0) and (0,1).

The dimension of a linear space is also the number of degrees of freedom of the space. For example, a point in three-dimensional space has three degrees of freedom because it can be described by three coordinates. A point in four-dimensional space has four degrees of freedom because it can be described by four coordinates.

The dimension of a linear space is also the number of independent equations that are needed to describe the space. For example, consider the set of all real numbers. This set can be described by the equation $x + 0 = x$. Since this equation has only one unknown (x), it is said to have dimension one.

The dimension of a linear space is also the number of vectors in any basis for the space. So, in the previous example, the set of all real numbers has dimension one because any basis for this space must have only one vector.

The dimension of a linear space is also the number of dimensions in which the space can be embedded. For example, consider the set of all two-dimensional vectors. This space can be embedded in three-dimensional space by adding a third coordinate that is always zero. So, the dimension of the set of all two-dimensional vectors is three.

A linear map is a function that preserves the structure of a linear space. In other words, a linear map is a function that takes vectors in one linear space and maps them to vectors in another linear space such that the following properties are satisfied:

1. The map is commutative: For any two vectors u and v in the first linear space, we have $f(u) + f(v) = f(v) + f(u)$.

2. The map is associative: For any three vectors u, v, and w in the first linear space, we have $f((u + v) + w) = (f(u) + f(v)) + f(w)$.
3. The map is distributive: For any scalar c and any two vectors u and v in the first linear space, we have $f(cu + cv) = cf(u) + cf(v)$.
4. The map is additive: For any two vectors \$u and \$v in the first linear space, we have $f(u + v) = f(u) + f(v)$.
5. The map is homogeneous: For any scalar c and any vector u in the first linear space, we have $f(cu) = cf(u)$.

A linear map is also called a linear transformation or a linear operator if the function maps into itself. The set of all linear maps from one linear space to another is itself a linear space. This space is called the space of linear maps from the first space to the second space.

In this book, we'll introduce the concept of linear maps and explore some of their properties. We'll also discuss some of the ways in which they can be used.

Linear spaces are vector spaces with a basis. This means that every element in the space can be written as a linear combination of the elements in the basis. The coordinates of an element are the coefficients of this linear combination.

In other words, they tell us how much of each element in the basis we need to add together to get to our desired element. For example, consider the vector $v = (2, 3, 5)$. We can write this vector as a linear combination of the vectors $e_1 = (1, 0, 0)$, $e_2 = (0, 1, 0)$, and $e_3 = (0, 0, 1)$ like so: $v = 2e_1 + 3e_2 + 5e_3$. Therefore, the coordinates of v with respect to this basis are $(2, 3, 5)$.

Linear maps take vector spaces and transform them into (perhaps) new vector spaces. The matrix of a linear map is a way of representation that shows how the map transforms each basis vector.

The matrix of a linear map is a way of representing this transformation. It is a square array of numbers that define how the coordinates of one space are transformed into the coordinates of another space. By understanding how linear maps work, we can better understand the relationships between objects in space.

For example, let's say we have the linear map $f(x) = Ax$ where A is the matrix $[[1, -2], [3, 4]]$. This map takes vectors in $\mathbb{R}^2$ and transforms them into vectors in $\mathbb{R}^2$. The matrix A tells us how $f(x)$ does this transformation. It takes the first basis vector $(1, 0)$ and transforms it into $(-2, 3)$, which is (-2 times the first basis vector plus 3 times the second basis vector). It does the same thing to the second basis vector $(0, 1)$, transforming it into $(4, -1)$. So now we know that $f(x) = (-2, 3)x + (4, -1)y$ for all x and y in $\mathbb{R}^2$.

We can use matrices to represent any linear map, no matter how compli-
cated. This makes them a very powerful tool for mathematical analysis.

In this book, you'll learn about the concept of linear spaces and linear
maps. We've discussed how these maps can be represented by matrices,
and we've seen how they can be used to understand relationships between
objects in space.

This is just the beginning of our exploration into the world of linear al-
gebra. In future books, we'll delve deeper into the properties of linear
maps and matrices. We'll also see how they can be used in a variety of
applications.

2.1   Introduction to Vector Spaces

Definition 2.1. Let $\mathbb{F}$ be a field. The collection of vectors, as $n \times 1$
matrices over $\mathbb{F}$, is called a **vector space** over the field of scalars $\mathbb{F}$. We
denote a vector space by V where $\mathbb{F}$ is a positive integer n.

Since vectors are just $n \times 1$ matrices, the proof of the next two propositions
follow immediately from 1.2 and 1.3.

Theorem 2.1. *Let V be a vector space. Then the following hold.*

- *For all $\vec{u}, \vec{v} \in V$, $\vec{u} + \vec{v} = \vec{v} + \vec{u}$.*
- *For all $\vec{u}, \vec{v}, \vec{w} \in V$, $(\vec{u} + \vec{v}) + \vec{w} = \vec{u} + (\vec{v} + \vec{w})$.*
- *For all $\vec{v} \in V$, there exists $\vec{0} \in V$ such that $\vec{v} + \vec{0} = \vec{v}$.*
- *For every $\vec{v} \in V$, there exists $\vec{w} \in V$ such that $\vec{v} + \vec{w} = \vec{0}$.*

Proof. The proof is left for the reader as Exercise 2.53.

$\square$

Theorem 2.2. *Let V be a vector space. Then the following hold.*

- *For all $\vec{v} \in V$, $1\vec{v} = \vec{v}$.*
- *For all $a, b \in k$ and $\vec{u} \in V$, $(ab)\vec{v} = a(b\vec{v})$.*
- *For all $a \in k, u, v \in V$, $a(\vec{u} + \vec{v}) = a\vec{u} + a\vec{v}$.*

- *For all $a, b\ k$ and $\vec{u} \in V$, $(a + b)\vec{u} = a\vec{u} + b\vec{u}$.*

Proof. The proof is left for the reader as Exercise 2.54.

$\square$

Theorem 2.3. *Let V be a vector space. Then the following hold.*

- *There exists a unique additive identity (denoted by $\vec{0}$).*
- *Every $\vec{v} \in V$ has a unique additive inverse, (denoted by $-\vec{v}$).*

Proof.   • Let $\vec{u}_1$ and $\vec{u}_2$ be additive identities in V, then $\vec{v} + \vec{u}_1 = \vec{v}$ and $\vec{v} + \vec{u}_2 = \vec{v}$ for every $\vec{v} \in V$. Then,

$$\vec{u}_1 = \vec{u}_1 + \vec{u}_2 = \vec{u}_2 + \vec{u}_1 = \vec{u}_2$$

as desired.
- Let $\vec{v}_1$ and $\vec{v}_2$ be additive inverses of $\vec{w}$ in V, then $\vec{w} + \vec{v}_1 = \vec{0}$ and $\vec{w} + \vec{v}_2 = \vec{0}$. Then,

$$\vec{v}_1 = \vec{v}_1 + \vec{0} = \vec{v}_1 + (\vec{w} + \vec{v}_2) = (\vec{v}_1 + \vec{w}) + \vec{v}_2 = (\vec{w} + \vec{v}_1) + \vec{v}_2 = \vec{0} + \vec{v}_2 = \vec{v}_2$$

as desired.

$\square$

Theorem 2.4. *Let V be a vector space. Then the following hold.*

- *If $\vec{v} \in V$, then $0\,\vec{v} = \vec{0}$.*
- *If $a \in k$, then $a\,\vec{0} = \vec{0}$.*

Proof.   • Let $\vec{v} \in V$, then

$$\vec{v} = 1\vec{v} = (1 + 0)\vec{v} = 1\vec{v} + 0\vec{v} = \vec{v} + 0\vec{v}$$

which shows that $0\vec{v}$ is the additive identity of V, namely $0\vec{v} = \vec{0}$.
- Let $a \in k$, then

$$a\vec{0} = a(\vec{0} + \vec{0}) = a\vec{0} + a\vec{0}$$

which shows that $a\vec{0}$ is the additive identity of V, namely $a\vec{0} = \vec{0}$.

$\square$

Theorem 2.5. *Let V be a vector space. Then the following hold.*

- *If $\vec{v} \in V$, then $-(-\vec{v}) = \vec{v}$.*
- *If $\vec{v} \in V$, then $(-1)\,\vec{v} = -\vec{v}$.*

Proof.   • Let $\vec{v} \in V$, then

$$\vec{v} + (-1)\vec{v} = 1\vec{v} + (-1)\vec{v} = (1 + (-1))\vec{v} = 0\vec{v} = \vec{0}$$

which shows that $(-1)\vec{v}$ is the unique additive inverse of $\vec{v}$ namely, $(-1)\vec{v} = -\vec{v}$.

- Since $-\vec{v}$ is the unique additive inverse of $\vec{v}$, $\vec{v} + (-\vec{v}) = \vec{0}$. Then $(-\vec{v}) + \vec{v} = \vec{0}$ shows that $\vec{v}$ is the unique additive inverse of $-\vec{v}$, namely, $\vec{v} = -(-\vec{v})$ as desired.

$\square$

Theorem 2.6. *Let V be a vector space with $a \in k$ and $\vec{v} \in V$. If $a\vec{v} = \vec{0}$, then $a = 0$ or $\vec{v} = \vec{0}$.*

Proof. Suppose $a \neq 0$. If $av = \vec{0}$ then

$$\vec{v} = 1\vec{v} = (a^{-1}a)\vec{v} = a^{-1}(a\vec{v}) = a^{-1}\vec{0} = \vec{0}.$$

Otherwise $a = 0$ as desired.

$\square$

Exercises

Exercise 2.1. Determine whether the following collection of vectors in $\mathbb{R}^3$ are linearly independent or linearly dependent.

- $(0, 1, 1), (1, 2, 1), (0, 4, 6), (1, 0, -1)$
- $(0, 1, 0), (1, 2, 1), (0, -4, 6), (-1, 1, -1)$

Exercise 2.2. Determine whether the following collection of vectors in $\mathbb{R}^4$ are linearly independent or linearly dependent.

- $(0, 1, 1, 1), (1, 2, 1, 1), (0, 4, 6, 2), (1, 0, -1, 2)$
- $(0, 1, 0, 1), (1, 2, 1, 3), (0, -4, 6, -2), (-1, 1, -1, 2)$

Exercise 2.3. Show that the given vectors do not form a basis for the vector space V.

- $(21, -7), (-6, 1); V = \mathbb{R}^2$
- $(21, -7, 14), (-6, 1, -4), (1, 0, 0); V = \mathbb{R}^3$
- $(48, 24, 108, -72), (-24, -12, -54, 36), (1, 0, 0, 0), (1, 1, 0, 0); V = \mathbb{R}^4$

Exercise 2.4. Reduce the vectors to a basis of the vector space V.

- $(1, 0), (1, 2), (2, 4), V = \mathbb{R}^2$
- $(1, 2, 3), (-1, -10, 15), (1, 2, -3), (2, 0, 6), (1, -2, 3), V = \mathbb{R}^3$

Exercise 2.5. Which of the following collection of vectors in $\mathbb{R}^3$ are linearly dependent? For those that are express one vector as a linear combination of the rest.

- $(1, 1, 0), (0, 2, 3), (1, 2, 3)$
- $(1, 1, 0), (3, 4, 2), (0, 2, 3)$

Exercise 2.6. Prove 2.1.

Exercise 2.7. Prove 2.2.

Exercise 2.8. Let $S = \{v_1, v_2, ..., v_k\}$ be a set of vectors in a a vector space V. Prove that S is linearly dependent if and only if one of the vectors in S is a linear combination of all other vectors in S.

Exercise 2.9. Suppose that $S = \{v_1, v_2, v_3\}$ is a linearly independent set of vector in a vector space V. Prove that $T = \{u_1, u_2, u_3\}$ is also linearly independent where $u_1 = v_1$, $u_2 = v_1 + v_2$, and $u_3 = v_1 + v_2 + v_3$.

Exercise 2.10. Which of the following sets of vectors form a basis for the vector space V.   - $(1, 3), (1, -1)$; $V = \mathbb{R}^2$ - $(1, 3), (-2, 6)$; $V = \mathbb{R}^2$ - $(3, 2, 2), (-1, 2, 1), (0, 1, 0)$; $V = \mathbb{R}^3$ - $(3, 2, 2), (-1, 2, 0), (1, 1, 0)$; $V = \mathbb{R}^3$ - $(2, 2, 2, 2), (3, 3, 3, 2), (1, 0, 0, 0), (0, 1, 0, 0)$; $V = \mathbb{R}^4$ - $(1,1,2,0), (2,2,4,0), (1,2,3,1), (2,1,3,-1), (1,2,3,-1)$; $V = \mathbb{R}^4$

Exercise 2.11. Find a basis for the subspace of the vector space V.

- All vectors of the form (a, b, c) where $b = a + c$ where $V = \mathbb{R}^3$.
- All vectors of the form (a, b, c) where $b = a - c$ where $V = \mathbb{R}^3$.
- All vectors of the form $\begin{bmatrix} b - a \\ a + c \\ b + c \\ c \end{bmatrix}$ where $V = \mathbb{R}^4$.

Exercise 2.12. Let $\vec{v}_1 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$, $\vec{v}_2 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$ and $S = \text{span}(\vec{v}_1, \vec{v}_2)$.

- Is S a subspace of $\mathbb{R}^3$?
- Find a vector $\vec{u}$ in S other than $\vec{v}_1$, $\vec{v}_2$.
- Find scalars which verify that $3\vec{u}$ is in S.

- Find scalars which verify that $\vec{0}$ is in S.

Exercise 2.13. Let $\vec{u}_1 = \begin{bmatrix} 0 \\ 2 \\ 2 \end{bmatrix}$, $\vec{u}_2 = \begin{bmatrix} 2 \\ 0 \\ 0 \end{bmatrix}$ and $T = \text{span}(\vec{u}_1, \vec{u}_2)$. Show $S = T$ by showing $S \subseteq T$ and $T \subseteq S$ where S is defined in Exercise 2.12.

Exercise 2.14. Prove that the non-empty intersection of two subspaces of $\mathbb{R}^3$ is a subspace of $\mathbb{R}^3$.

Exercise 2.15. Let S and T be subspaces of $\mathbb{R}^3$ defined by

$$S = \text{span}\left(\begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}, \begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix} \right) \qquad \text{and} \qquad T = \text{span}\left(\begin{bmatrix} 2 \\ -2 \\ 3 \end{bmatrix}, \begin{bmatrix} 3 \\ -4 \\ 4 \end{bmatrix} \right).$$

Show they are the same subspace of $\mathbb{R}^3$.

Exercise 2.16. Let $\vec{v}_1, \vec{v}_2, \vec{v}_3$ be a linearly independent set of vectors. Show that if $\vec{v}_4$ is not a linear combination of $\vec{v}_1, \vec{v}_2, \vec{v}_3$, then $\vec{v}_1, \vec{v}_2, \vec{v}_3, \vec{v}_4$ is a linearly independent set of vectors.

Exercise 2.17. If
$\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$ is a linearly independent set of vectors in V, show that $\{\vec{v}_1, \vec{v}_1 + \vec{v}_2, \vec{v}_1 + \vec{v}_2 + \vec{v}_3\}$ is also a linearly independent set of vectors in V.

Exercise 2.18. If
$\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$ is a linearly independent set of vectors in V, show that $\{\vec{v}_1 + \vec{v}_2, \vec{v}_2 + \vec{v}_3, \vec{v}_3 + \vec{v}_1\}$ is also a linearly independent set of vectors in V.

Exercise 2.19. Let $\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$ be a linearly dependent set. Show that at least one of the $\vec{v}_i$ is a linear combination of the others.

Exercise 2.20. Prove or provide a counterexample to the following statement. If a set of vectors T spans the vector space V, then T is linearly independent.

Exercise 2.21. Which of the following are not a basis for $\mathbb{R}^3$?

- $\vec{v_1} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \vec{v_2} = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \vec{v_3} = \begin{bmatrix} 1 \\ -1 \\ -1 \end{bmatrix}$

- $\vec{u_1} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \vec{u_2} = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \vec{u_3} = \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix}$

Exercise 2.22. Let S be the space spanned by the vectors

$$\vec{v_1} = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 1 \end{bmatrix} \quad \vec{v_1} = \begin{bmatrix} -1 \\ -3 \\ 1 \\ 0 \end{bmatrix} \quad \vec{v_1} = \begin{bmatrix} 2 \\ 3 \\ 0 \\ 1 \end{bmatrix} \quad \vec{v_1} = \begin{bmatrix} 2 \\ 0 \\ 2 \\ 2 \end{bmatrix}$$

Find the dimension of S and a subset of T which could serve as a basis for S.

Exercise 2.23. Let $\{\vec{v_1}, \vec{v_2}, ..., \vec{v_n}\}$ be a basis for V, and suppose that $\vec{u} = a_1\vec{v_1} + a_2\vec{v_2} + \cdots + a_n\vec{v_n}$ with $a_1 \neq 0$. Prove that $\{\vec{u}, \vec{v_2}, ..., \vec{v_n}\}$ is also a basis for V.

Exercise 2.24. Let $S = \text{span}(\vec{v_1}, \vec{v_2}, \vec{v_3})$ and $T = \text{span}(\vec{u_1}, \vec{u_2}, \vec{u_3})$ where $\vec{v_i}$ and $\vec{u_i}$ are defined as follows.

$$\vec{v_1} = \begin{bmatrix} 1 \\ -1 \\ 2 \\ 0 \end{bmatrix} \quad \vec{v_2} = \begin{bmatrix} 2 \\ 1 \\ 1 \\ 1 \end{bmatrix} \quad \vec{v_3} = \begin{bmatrix} 3 \\ -1 \\ 2 \\ -1 \end{bmatrix} \quad \vec{u_1} = \begin{bmatrix} 3 \\ 0 \\ 3 \\ 1 \end{bmatrix} \quad \vec{u_2} = \begin{bmatrix} 1 \\ 2 \\ -1 \\ 1 \end{bmatrix} \quad \vec{u_3} = \begin{bmatrix} 4 \\ -1 \\ 5 \\ 1 \end{bmatrix}$$

Is one of these two subspaces strictly contained in the other or are they equal?

Exercise 2.25. Let $S = \text{span}(\vec{v_1}, \vec{v_2}, \vec{v_3})$ where $\vec{v_i}$ are defined as follows.

$$\vec{v_1} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 1 \end{bmatrix} \quad \vec{v_2} = \begin{bmatrix} 2 \\ -1 \\ 1 \\ -3 \end{bmatrix} \quad \vec{v_3} = \begin{bmatrix} 1 \\ 3 \\ 4 \\ 2 \end{bmatrix} \quad \text{and} \quad \vec{u} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 1 \end{bmatrix}$$

Is the vector $\vec{u}$ in S?

Exercise 2.26. If possible, find a value of a so that the vectors

$$\begin{bmatrix} 1 \\ 2 \\ a \end{bmatrix} \qquad \begin{bmatrix} 0 \\ 1 \\ a - 1 \end{bmatrix} \qquad \begin{bmatrix} 3 \\ 4 \\ 5 \end{bmatrix}$$

are linearly independent.

Exercise 2.27. Let $S = \text{span}(\vec{v}_1, \vec{v}_2, \vec{v}_3)$ where $\vec{v}_i$ are defined as follows.

$$\vec{v}_1 = \begin{bmatrix} 1 \\ -1 \\ 2 \\ 3 \end{bmatrix} \qquad \vec{v}_2 = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix} \qquad \vec{v}_3 = \begin{bmatrix} 3 \\ -2 \\ 5 \\ 7 \end{bmatrix} \qquad \text{and} \qquad \vec{u} = \begin{bmatrix} 1 \\ 1 \\ 0 \\ -1 \end{bmatrix}$$

Find a basis of S which includes the vector $\vec{u}$.

Exercise 2.28. Find a vector $\vec{u}$ in $\mathbb{R}^4$ such that $\vec{u}$ and the vectors

$$\vec{v}_1 = \begin{bmatrix} 1 \\ -1 \\ -1 \\ 1 \end{bmatrix} \qquad \vec{v}_2 = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 1 \end{bmatrix} \qquad \vec{v}_3 = \begin{bmatrix} 1 \\ 2 \\ 1 \\ 1 \end{bmatrix}$$

for a basis of $\mathbb{R}^4$.

Exercise 2.29. Show that every subspace of V has no more than n linearly independent vectors.

Exercise 2.30. Find two bases of $\mathbb{R}^4$ that have only the vectors $\vec{e}_3$ and $\vec{e}_4$ in common.

Exercise 2.31. Prove that if a list of vectors is linearly independent so is any sublist.

Exercise 2.32. Suppose $\vec{v}_1, \vec{v}_2, \vec{v}_3$ and $\vec{v}_1, \vec{v}_2, \vec{v}_4$ are two sets of linearly dependent vectors, and suppose that $\vec{v}_1$ and $\vec{v}_2$ are linearly independent. Prove that any set of three vectors chosen from $\vec{v}_1, \vec{v}_2, \vec{v}_3, \vec{v}_4$ is linearly dependent.

Exercise 2.33. If $\vec{u}$ and $\vec{v}$ are linearly independent vectors in V, prove that the vectors $a\vec{u} + b\vec{v}$ and $c\vec{u} + d\vec{v}$ are also linearly independent if and only if $ad - bc \neq 0$.

Exercise 2.34. Complete the proof of 2.12.

Exercise 2.35. Let U be the collection of vectors that satisfy the equations $x + y + z = 0$ and $x + 2y - z = 0$. Show U is a subspace of $\mathbb{R}^3$, find a basis for U, and find $\dim(U)$.

Exercise 2.36. Let U be the collection of vectors that satisfy the equations $x + y + z = 0$, $x + 2y - z = 0$, and $y - 2z = 0$. Show U is a subspace of $\mathbb{R}^3$, find a basis for U, and find $\dim(U)$.

Exercise 2.37. Show that the only subspaces of $\mathbb{R}$ are $\{\vec{0}\}$ and $\{\mathbb{R}\}$.

Exercise 2.38. Show that the only subspaces of $\mathbb{R}^2$ are $\{\vec{0}\}$, $\{\mathbb{R}^2\}$, and any set consisting of all scalar multiples of a nonzero vector. Describe these subspaces geometrically.

Exercise 2.39. Determine the various types of subspaces of $\mathbb{R}^3$ and describe them geometrically.

Exercise 2.40. For $\vec{b} \neq \vec{0}$, show that the set of solutions of the $n \times m$ linear system $A\vec{x} = \vec{b}$, is not a subspace of V.

Exercise 2.41. Suppose that $\vec{v}_1, \vec{v}_2, ..., \vec{v}_n$ are linearly independent in $\mathbb{R}^n$. Show that if A is an $n \times n$ matrix with $\text{rref}(A) = I_n$, then $A\vec{v}_1, A\vec{v}_2, ..., A\vec{v}_n$ are also linearly independent in $\mathbb{R}^n$.

Exercise 2.42. Let $S = \{v_1, v_2, ..., v_s\}$ and $T = \{u_1, u_2, ..., u_t\}$ be two sets of vectors in V where each $\vec{u}_i$, $(i = 1, 2, ..., t)$ is a linear combination of the vectors in S. Show that $\vec{w} = a_1\vec{u}_1 + a_2\vec{u}_2 + \cdots + a_m\vec{u}_t$ is a linear combination of the vectors in S.

Exercise 2.43. Let $S = \{v_1, v_2, \dots, v_m\}$ be a set of non-zero vectors in a vector space V such that every vector in V can be uniquely as a linear combination of the vectors in S. Prove that S is a basis for V.

Exercise 2.44. Find a basis for the solution space of the homogeneous system $(\lambda I_n - A)\vec{x} = \vec{0}$ for the given λ and A.

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

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

Exercise 2.45. Prove 2.9.

Exercise 2.46. Prove 2.1.

Exercise 2.47. Prove 2.13.

2.2 Subspaces

Definition 2.2. A subset U of a vector space V is called a **subspace** of V if it has the following three properties

- U contains the zero vector in V,
- U is closed under addition: if $\vec{u}$ and $\vec{v}$ are in U then so is $\vec{u} + \vec{v}$, and
- U is closed under scalar multiplication: if $\vec{v}$ is in U and a is any scalar, then $a\vec{v}$ is in U.

Example 2.1. Let U be the subset of $\mathbb{R}^5$ defined by

$$U = \{(x_1, x_2, x_3, x_4, x_5) \in \mathbb{R}^5 \mid x_1 = 3x_2 \text{ and } x_3 = 7x_4\}.$$

Show U is a subspace of $\mathbb{R}^5$. The zero vector $\vec{0} = (0, 0, 0, 0, 0, 0)$ is in U since $0 = 3(0)$ and $0 = 7(0)$. Let $\vec{u} = (u_1, u_2, u_3, u_4, u_5)$ and $\vec{v} = (v_1, v_2, v_3, v_4, v_5)$ be vectors in U and let a be a scalar. Then

$$\vec{u} + \vec{v} = (u_1 + v_1, u_2 + v_2, u_3 + v_3, u_4 + v_4, u_5 + v_5)$$

is in U since

$$u_1 = 3u_2 \text{ and } v_1 = 3v_2 \text{ imply } u_1 + v_1 = 3(u_2 + v_2)$$

and
$$u_3 = 7u_4 \text{ and } v_3 = 7v_4 \text{ imply } u_3 + v_3 = 7(u_4 + v_4).$$

Also, $a\vec{u}$ is in U since $u_1 = 3u_2$ implies $au_1 = 3(au_2)$. By 2.2, U is a subspace of $\mathbb{R}^5$.

Example 2.2. Give an example of a nonempty subset U of $\mathbb{R}^2$ such that U is closed under addition and under taking additive inverses, but U is not a subspace of $\mathbb{R}^2$. The subset $\mathbb{Z}^2$ of $\mathbb{R}^2$ is closed under additive inverses and addition, however $\mathbb{Z}^2$ is not a subspace of $\mathbb{R}^2$ since $\sqrt{2} \in \mathbb{R}, (1,1) \in \mathbb{Z}^2$ however $(\sqrt{2}, \sqrt{2}) \notin \mathbb{Z}^2$.

Example 2.3. Give an example of a nonempty subset U of $\mathbb{R}^2$ such that U is closed under scalar multiplication, but U is not a subspace of $\mathbb{R}^2$. The set $\{(x_1, x_2) \in \mathbb{R}^2 \mid x_1 x_2 = 0\} = M$ is closed under scalar multiplication because, if $\lambda \in \mathbb{R}$ and $(x_1, x_2) \in M$, then $(\lambda x_1, \lambda x_2) \in M$ holds since $\lambda x_1 \lambda x_2 = 0$. However, M is not a subspace because $(0,1) + (1,0) = (1,1) \notin M$ even though $(0,1), (1,0) \in M$.

Example 2.4. Show that the set of all solutions of an $m \times n$ homogenous linear system of equations is a subspace of V (called the **null space**). Let $A\vec{x} = \vec{0}$ be an $m \times n$ homogenous system of linear equations and let U be the set of solutions to this system. Of course $A\vec{0} = \vec{0}$ and so the zero vector is in U. Let $\vec{u}$ and $\vec{v}$ be in U and let a be a scalar. Then

$$A(\vec{u} + \vec{v}) = A\vec{u} + A\vec{v} = \vec{0} + \vec{0} = \vec{0}$$

and
$$A(a\vec{u}) = a(A\vec{u}) = a\vec{0} = \vec{0}$$

shows $\vec{u} + \vec{v}$ and $a\vec{u}$ are in U. By 2.2, U is a subspace of V.

Definition 2.3. Let $\vec{v}_1, \vec{v}_2, ..., \vec{v}_m$ be vectors in the vector space V. The set of all linear combinations

$$\text{span}(v_1, v_2, ..., v_m) = \{c_1\vec{v}_1 + c_2\vec{v}_2 + \cdots + c_m\vec{v}_m \mid c_1, c_2, ..., c_m \in k\}$$

is called the **spanning set** of the vectors $v_1, v_2, ..., v_m$.

Example 2.5. Show that the spanning set of the vectors $v_1, v_2, ..., v_m$ in V is a subspace of V.
Let $U = \text{span}(v_1, v_2, ..., v_m)$. Notice $\vec{0} \in U$ since $\vec{0} = 0_1\vec{v}_1 + 0_2\vec{v}_2 + \cdots + 0_m\vec{v}_m$ where $0 \in k$. Let $\vec{u}$ and $\vec{v}$ be vectors in U and let a be a scalar. By

2.3, there exists scalars $c_1, c_2, \ldots, c_m$ and scalars $d_1, d_2, \ldots, d_m$ such that

$$\vec{u} = c_1\vec{v}_1 + c_2\vec{v}_2 + \cdots + c_m\vec{v}_m \quad \text{and} \quad \vec{v} = d_1\vec{v}_1 + d_2\vec{v}_2 + \cdots + d_m\vec{v}_m$$

Then

$$\vec{u} + \vec{v} = \sum_{i=1}^{m} c_i\vec{v}_i + \sum_{i=1}^{m} d_i\vec{v}_i = \sum_{i=1}^{m}(c_i + d_i)\vec{v}_i$$

and

$$a\vec{u} = a\left(\sum_{i=1}^{m} c_i\vec{v}_i\right) = \sum_{i=1}^{m}(ac_i)\vec{v}_i$$

show $\vec{u} + \vec{v}$ and $a\vec{u}$ are in U; and thus U is a subspace of V.

We say that a nonzero $\vec{v}_i$ is in the list $\vec{v}_1, \ldots, \vec{v}_i, \ldots, \vec{v}_m$ if $\vec{v}_i$ can be written as a linear combination of the other nonzero vectors in the list. An equation of the form $a_1\vec{v}_1 + \cdots + a_m\vec{v}_m = \vec{0}$ is called a **linear relation** among the vectors $\vec{v}_1, \ldots, \vec{v}_m$; and is called a **nontrivial relation** if at least one of the a_i's is nonzero.

Definition 2.4. The vectors $v_1, v_2, \ldots, v_m$ in V are called **linear independent** if the only choice for

$$a_1\vec{v}_1 + a_2\vec{v}_2 + \cdots + a_m\vec{v}_m = \vec{0}$$

is $a_1 = a_2 = \cdots = a_m = 0$. Otherwise the vectors $v_1, v_2, \ldots, v_m$ are called **linear dependent**.

Lemma 2.1. *Show that any set $S = \{v_1, v_2, \ldots, v_m\}$ of vectors in V is a linearly dependent set of vectors if and only if at least one of the vectors in the set can be written as a linear combination of the others.*

Proof. Assume the vectors in the set S are linearly dependent. By 2.4, there exists scalars $c_1, c_2, \ldots, c_m$ (not all zero) such that $c_1\vec{v}_1 + c_2\vec{v}_2 + \cdots + c_m\vec{v}_m = \vec{0}$. Let i be the least index such that c_i is nonzero. Thus $c_1 = c_2 = \cdots = c_{i-1} = 0$. So

$$c_i\vec{v}_i = -c_{i+1}\vec{v}_{i+1} - \cdots - c_m\vec{v}_m$$

for some i. Since $c_i \neq 0$ and $c_i \in k$, c_i^{-1} exists and thus

$$\vec{v}_i = \left(\frac{-c_{i+1}}{c_i}\right)\vec{v}_{i+1} + \cdots + \left(\frac{-c_m}{c_i}\right)\vec{v}_m$$

which shows $\vec{v}_i$ is a linear combination of the others, via

$$\vec{v}_i = 0\vec{v}_1 + \cdots + 0\vec{v}_{i-1} + \left(\frac{-c_{i+1}}{c_i}\right)\vec{v}_{i+1} + \cdots + \left(\frac{-c_m}{c_i}\right)\vec{v}_m.$$

Now assume one of the vectors in the set S can be written as a linear combination of the others, say

$$\vec{v}_k = c_1\vec{v}_1 + \cdots + c_{k-1}\vec{v}_{k-1} + c_{k+1}\vec{v}_{k+1} + \cdots + c_m\vec{v}_m$$

where $c_1, c_2, \ldots, c_m$ are scalars. Thus,

$$\vec{0} = c_1\vec{v}_1 + \cdots + c_{k-1}\vec{v}_{k-1} + (-1)\vec{v}_k + c_{k+1}\vec{v}_{k+1} + \cdots + c_m\vec{v}_m$$

and so by 2.4, $\vec{v}_1, \ldots, \vec{v}_m$ are linearly dependent.

$\square$

For a list of vectors $v_1, v_2, \ldots, v_m$ in V the following equivalent statements follow from the appropriate definitions:

- vectors $v_1, v_2, \ldots, v_m$ are linearly independent,
- none of the vectors $v_1, v_2, \ldots, v_m$ are redundant,
- none of the vectors $v_1, v_2, \ldots, v_m$ can be written as a linear combination of the other vectors in the list,
- there is only the trivial relation among the vectors $v_1, v_2, \ldots, v_m$,
- the only solution to the equation $a_1\vec{v}_1 + a_2\vec{v}_2 + \cdots + a_m\vec{v}_m$ is $a_1 = a_2 = \cdots = a_m = 0$, and
- $\operatorname{rank}(A) = n$ where A is the $n \times m$ matrix whose columns are the vectors $v_1, v_2, \ldots, v_m$.

Example 2.6. Determine whether the following vectors $\vec{u}$, $\vec{v}$, and $\vec{w}$ are linearly independent.

$$\vec{u} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \qquad \vec{v} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix} \qquad \vec{w} = \begin{bmatrix} 1 \\ 4 \\ 7 \\ 10 \end{bmatrix}$$

Without interchanging rows, we use elementary row operations to find

$$\operatorname{rref} \begin{bmatrix} \vec{u} & \vec{v} & \vec{w} \end{bmatrix} = \begin{bmatrix} 1 & 0 & -2 \\ 0 & 1 & 3 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

From this we infer the nontrivial relation $\vec{0} = (-2)\vec{u} + (3)\vec{v} + (-1)\vec{w}$. Therefore the given vectors are linearly dependent.

Theorem 2.7. *Let*

$$\vec{v}_1 = \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{n1} \end{bmatrix} \qquad \vec{v}_2 = \begin{bmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{n2} \end{bmatrix} \qquad \cdots \qquad \vec{v}_m = \begin{bmatrix} a_{1s} \\ a_{2m} \\ \vdots \\ a_{nm} \end{bmatrix}$$

*be s vectors in V. These vectors are linearly dependent if and only if there
exists a solution to the system of linear equations*

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1m}x_m = 0 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2m}x_m = 0 \\ \quad\vdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nm}x_m = 0 \end{cases} \tag{2.1}$$

different from $x_1 = x_2 = \cdots = x_m = 0$.

Proof. Assume $v_1, v_2, \ldots, v_m$ are linear dependent. By 2.4, there exists
scalars $c_1, c_2, \ldots, c_m$ such that

$$c_1\vec{v}_1 + c_2\vec{v}_2 + \cdots + c_m\vec{v}_m = \vec{0} \tag{2.2}$$

and not all c_i's are zero. 2.2 yields a system

$$\begin{cases} a_{11}c_1 + a_{12}c_2 + \cdots + a_{1m}c_m = 0 \\ a_{21}c_1 + a_{22}c_2 + \cdots + a_{2m}c_m = 0 \\ \quad\vdots \\ a_{n1}c_1 + a_{n2}c_2 + \cdots + a_{nm}c_m = 0 \end{cases}$$

with solution $x_1 = c_1$, $x_2 = c_2$, ..., $x_m = c_m$. Since not all c_i's are zero
we have a solution different from $x_1 = x_2 = \cdots = x_m = 0$. Assume the
system in 2.1 has a solution $\vec{x}^*$ with $\vec{x}^* \neq \vec{0}$, say x_i^*. By 1.5 we can write

$$\vec{0} = A\vec{x} = x_1^*\vec{v}_1 + x_2^*\vec{v}_2 + \cdots + x_m^*\vec{v}_m \tag{2.3}$$

Isolating the term $x_i^*\vec{v}_i$ yields

$$x_i^*\vec{v}_i = -x_i^*\vec{v}_1 - \cdots - x_{i-1}\vec{v}_{i-1} - x_{i+1}\vec{v}_{i+1} - \cdots - x_m^*\vec{v}_m \tag{2.4}$$

and since $x_i^* \neq 0$, $(x_i^*)^{-1}$ exists. Therefore,

$$\vec{v}_1 = \left(-\frac{x_1^*}{x_i^*}\right)\vec{v}_1 - \cdots - \left(-\frac{x_{i-1}^*}{x_i^*}\right)\vec{v}_{i+1} - \cdots - \left(-\frac{x_m^*}{x_i^*}\right)\vec{v}_m \tag{2.5}$$

shows the vectors $v_1, v_2, \ldots, v_m$ are linearly dependent.

$\square$

Theorem 2.8. *The $n \times m$ linear system of equations $A\vec{x} = \vec{b}$ has a solution
if and only if the vector $\vec{b}$ is contained in the subspace of V generated by
the column vectors of A.*

Proof. Let $\vec{x}$ be a solution to $A\vec{x} = \vec{b}$ with $A = \begin{bmatrix} \vec{v}_1 & \vec{v}_2 & \cdots & \vec{v}_m \end{bmatrix}$. By 1.5, $\vec{b} = A\vec{x} = x_1\vec{v}_1 + x_2\vec{v}_2 + \cdots + x_m\vec{v}_m$ and thus $\vec{b} \in \mathrm{span}(v_1, v_2, \ldots, v_m)$ as needed. Conversely, assume $\vec{b}$ is in the subspace generated by the column vectors of A; that is assume $\vec{b} \in \mathrm{span}(v_1, v_2, \ldots, v_m)$. By 2.3, there exists scalars $c_1, c_2, \ldots, c_m$ such that $\vec{b} = c_1\vec{v}_1 + c_2\vec{v}_2 + \cdots + c_m\vec{v}_m$. By 2.3, $\vec{b} = c_1\vec{v}_1 + c_2\vec{v}_2 + \cdots + c_m\vec{v}_m = A\vec{c}$ where the components of $\vec{c}$ are the c_i's. Thus the system $A\vec{x} = \vec{b}$ has a solution, namely $\vec{c}$.

$\square$

Example 2.7. Let U and V be finite subsets of a vector space V with $U \subseteq V$.

- If U is linear dependent, then so is V.
- If V is linear independent, then so is U.

Let $U = \{u_1, u_2, \ldots, u_s\}$ $V = \{v_1, v_2, \ldots, v_t\}$.

- If U is linear dependent, then thee exists a vector, say $\vec{u}_k$ such that $\vec{u}_k$ is a linear combination of the other $\vec{u}_i$'s. Since $U \subseteq V$ all $\vec{u}_i$'s are in V. Thus we have a vector $\vec{u}_k$ in V that is a linear combination of other vectors in V. Therefore, by 2.1, V is linear dependent.
- Let $c_1, c_2, \ldots, c_s$ be scalars such that

$$c_1\vec{u}_1 + c_2\vec{u}_2 + \cdots + c_m\vec{u}_s = \vec{0}. \tag{2.6}$$

Since $U \subseteq V$, we know $u_i \in V$ for $1 \le i \le s$. Since V is linear independent, 2.6 implies $c_1 = c_2 = \cdots = c_m = 0$. By 2.4, U is linear independent as well.

Corollary 2.1. *Any vector in V, written as a column matrix, can be expressed (uniquely) as a linear combination of $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_m$ if and only if $A\vec{x} = \vec{v}$ has unique solution, where $A = \begin{bmatrix} \vec{v}_1 & \vec{v}_2 & \cdots & \vec{v}_m \end{bmatrix}$. When there is a solution, the components $x_1, x_2, \ldots, x_m$ of $\vec{x}$ give the coefficients for the linear combination.*

Proof. This proof is left for the reader as Exercise 2.93.

$\square$

Theorem 2.9. *The vectors $v_1, v_2, \ldots, v_n$ in V form a linearly independent set of vectors if and only if $\begin{bmatrix} \vec{v}_1 & \vec{v}_2 & \cdots & \vec{v}_n \end{bmatrix}$ is row equivalent to I_n.*

Proof. This proof is left for the reader as Exercise 2.92.

$\square$

Theorem 2.10. *Let V be a vector space and assume that the vectors $v_1, v_2, \dots, v_n$ are linearly independent and $\text{span}(s_1, s_2, \dots, s_m) = V$. Then $n \leq m$.*

Proof. We are given

$$\text{span}(s_1, s_2, \dots, s_m) = V \quad \text{and} \quad v_1, v_2, \dots, v_n \text{ are linearly independent.}$$

Since $\vec{v}_1$ is a linear combination of the vectors $\vec{s}_1, \vec{s}_2, \dots, \vec{s}_m$ we obtain

$$\text{span}(\vec{v}_1, \vec{s}_2, \dots, \vec{s}_m) = V \quad \text{and} \quad \vec{v}_2, \dots, \vec{v}_n \text{ are linearly independent,}$$

respectively. Since $\vec{v}_2$ is a linear combination of $\vec{v}_1, \vec{s}_2, \dots, \vec{s}_m$ we can obtain

$$\text{span}(\vec{v}_1, \vec{v}_2, \vec{s}_3, \dots, \vec{s}_m) = V \quad \text{and} \quad \vec{v}_3, \dots, \vec{v}_n \text{ are linearly independent,}$$

respectively. Now if $m < n$ then repeating this process will eventually exhaust the $\vec{s}_i$'s and lead to

$$\text{span}(\vec{v}_1, \vec{v}_2, \dots, \vec{v}_m) = V \quad \text{and} \quad \vec{v}_{m+1}, \dots, \vec{v}_n \text{ are linearly independent.}$$

This is a contradiction since $\vec{v}_n$ is not in $\text{span}(v_1, v_2, \dots, v_m)$; and whence $n \leq m$.

$\square$

Theorem 2.11. *A set $S = \{v_1, v_2, \dots, v_m\}$ of vectors in V is linearly independent if and only if for any vector $\vec{u}$, if $\vec{u} = u_1 \vec{v}_1 + u_2 \vec{v}_2 + \dots + u_m \vec{v}_m$, then this representation is unique.*

Proof. Assume the vectors in S are linearly independent and assume $\vec{u}$ is an arbitrary vector with

$$\vec{u} = a_1 \vec{v}_1 + a_2 \vec{v}_2 + \dots + a_m \vec{v}_m \qquad \text{and} \qquad \vec{u} = b_1 \vec{v}_1 + b_2 \vec{v}_2 + \dots + b_m \vec{v}_m \quad (2.7)$$

as both representations of $\vec{u}$ as linear combinations of the vectors in S. Then

$$\vec{0} = \vec{u} - \vec{u} = (a_1 - b_1)\vec{v}_1 + (a_2 - b_2)\vec{v}_2 + \dots + (a_m - b_m)\vec{v}_m.$$

Since S is linearly independent $a_1 - b_1 = a_2 - b_2 = \dots = a_m - b_m = 0$ and thus $a_1 = b_1$, $a_2 = b_2$, ..., $a_m = b_m$. Therefore, the representation of $\vec{u}$ as

a linear combination of the vectors in S is unique. Conversely, assume for nay vector $\vec{u}$ which can be written as a linear combination of the vectors in S, the representation is unique. If $c_1, c_2, \ldots, c_m$ are scalars such that $c_1\vec{v}_1 + c_2\vec{v}_2 + \cdots + c_m\vec{v}_m = \vec{0}$ then $c_1 = c_2 = \cdots = c_m = 0$ much hold since $0\vec{v}_1 + 0\vec{v}_2 + \cdots + 0\vec{v}_m = \vec{0}$ and this representation is unique. Therefore, the vectors in S are linearly independent.

$\square$

Definition 2.5. The vectors $v_1, v_2, \ldots, v_m$ in V are called a **basis** of a linear subspace V if they span V and are linearly independent.

Example 2.8. Find a basis for V for $n = 1, 2, 3, \ldots m$. For $n = 2$, the vectors $\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ form a basis for k^2. For $n = 3$, the vectors $\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix},$

$\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$ form a basis for k^3. In general, for a positive integer n, the following

n vectors of V form a basis (called the **standard basis**) of V.

$$\vec{e}_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \qquad \vec{e}_2 = \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix} \qquad \cdots \qquad \vec{e}_n = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix} \qquad (2.8)$$

The vectors in a standard basis are linearly independent by 2.9. Given any vector $\vec{v}$ in V with components v_i, we can write

$$\vec{v} = v_1\vec{e}_1 + v_2\vec{e}_2 + \cdots + v_m\vec{e}_n,$$

and thus $k^n = \text{span}(e_1, e_2, \ldots, e_n)$ which shows that any standard basis is in fact a basis.

Example 2.9. Show the following vectors $\vec{v}_1, \vec{v}_2, \vec{v}_3$, and $\vec{v}_4$ form a basis for $\mathbb{R}^4$.

$$\vec{v}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \qquad \vec{v}_2 = \begin{bmatrix} 1 \\ -1 \\ 1 \\ -1 \end{bmatrix} \qquad \vec{v}_3 = \begin{bmatrix} 1 \\ 2 \\ 4 \\ 8 \end{bmatrix} \qquad \vec{v}_4 = \begin{bmatrix} 1 \\ -2 \\ 4 \\ -8 \end{bmatrix}$$

We determine $\text{rref}(A) = I_4$ where A is the matrix with column vectors $\vec{v}_1, \vec{v}_2, \vec{v}_3, \vec{v}_4$. By 2.9, $\vec{v}_1, \vec{v}_2, \vec{v}_3, \vec{v}_4$ are linearly independent. Since $\vec{v}_1, \vec{v}_2, \vec{v}_3, \vec{v}_4$ also span $\mathbb{R}^4$, they form a basis of $\mathbb{R}^4$.

Example 2.10. Let U be the subspace of $\mathbb{R}^5$ defined by

$$U = \{(x_1, x_2, x_3, x_4, x_5) \in \mathbb{R}^5 \mid x_1 = 3x_2 \text{ and } x_3 = 7x_4\}.$$

Find a basis of U. The following vectors belong to U and are linearly independent in $\mathbb{R}^5$.

$$v_1 = \begin{bmatrix} 3 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \qquad v_2 = \begin{bmatrix} 0 \\ 0 \\ 7 \\ 1 \\ 0 \end{bmatrix} \qquad v_3 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

If $u \in U$, then the representation

$$u = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \\ u_5 \end{bmatrix} = \begin{bmatrix} 3u_2 \\ u_2 \\ 7u_4 \\ u_4 \\ u_5 \end{bmatrix} = u_2 \begin{bmatrix} 3 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + u_4 \begin{bmatrix} 0 \\ 0 \\ 7 \\ 1 \\ 0 \end{bmatrix} + u_5 \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

shows that they also span U, and thus form a basis of U by 2.5.

Theorem 2.12. *Let $S = \{v_1, v_2, \ldots, v_n\}$ be a set of vectors in a vector space V and let $W = \mathrm{span}(S)$. Then some subset of S is a basis for W.*

Proof. Assume $W = \mathrm{span}(S)$ and suppose S is a linearly independent set of vectors. Thus, in this case, S is a basis of W, by 2.5. So we can assume S is a linearly dependent set of vectors. By 2.1, there exists i such that $1 \leq i \leq m$ and $\vec{v}_i$ is a linear combination of the other vectors in S. It is left for Exercise 2.81 to show that

$$W = \mathrm{span}(S) = \mathrm{span}(S_1)$$

where $S_1 = S/\{\vec{v}_i\}$. If S_1 is linearly independent set of vectors, then S_1 is a basis of W. Otherwise, S_1 is a linear dependent set and we can delete a vector from S_1 that is a linear combination of the other vectors in S_1. We obtain another subset S_2 of S with

$$W = \mathrm{span}(S) = \mathrm{span}(S_1) = \mathrm{span}(S_2).$$

Since S is finite, if we continue, we find a linearly independent subset of S and thus a basis of W.

$\square$

Corollary 2.2. *All bases of a subspace U of a vector space V consists of the same number of vectors.*

Proof. Let $S = \{v_1, v_2, \ldots, v_n\}$ and $T = \{w_1, w_2, \ldots, w_m\}$ be bases of a subspace U. Then $\mathrm{span}(S) = V$ and T is a lineal indecent set of vectors. By 2.10, $m \leq n$. Similarly, since $\mathrm{span}(T) = V$ and S is a linearly independent set of vectors, $n \leq m$. Therefore, $m = n$ as desired.

$\square$

Corollary 2.3. *The vectors $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n$ form a basis of V if and only if the reduced row echelon form of the $n \times n$ matrix $\begin{bmatrix} \vec{v}_1 & \vec{v}_2 & \cdots & \vec{v}_n \end{bmatrix}$ is I_n.*

Proof. Suppose the vectors $v_1, v_2, \ldots, v_n$ form a basis of V and consider the $n \times n$ linear system

$$\begin{cases} v_{11}x_1 + v_{12}x_2 + \cdots + v_{1n}x_n = 0 \\ v_{21}x_1 + v_{22}x_2 + \cdots + v_{2n}x_n = 0 \\ \quad\vdots \\ v_{n1}x_1 + v_{n2}x_2 + \cdots + v_{nn}x_n = 0 \end{cases}$$

where the v_{ij}'s are the components of the $\vec{v}_j$'s. Since $\{v_1, v_2, \ldots, v_n\}$ is a basis, $\vec{v}_1 = \vec{v}_2 = \cdots = \vec{v}_n = \vec{0}$ and this linear system can not have another solution. By 1.3, $\mathrm{rref}(A) = I_n$ where $A = \begin{bmatrix} \vec{v}_1 & \vec{v}_2 & \cdots & \vec{v}_n \end{bmatrix}$.

$\square$

Definition 2.6. The number of vectors in a basis of a subspace U of V is called the **dimension** of U, and is denoted by $\dim U$.

Example 2.11. Find a basis of the subspace of $\mathbb{R}^4$ that consists of all vectors perpendicular to both of the following vectors $\vec{v}_1$ and $\vec{v}_2$.

$$\vec{v}_1 = \begin{bmatrix} 1 \\ 0 \\ -1 \\ 1 \end{bmatrix} \qquad \vec{v}_2 = \begin{bmatrix} 0 \\ 1 \\ 2 \\ 3 \end{bmatrix}$$

We need to find all vectors $\vec{x}$ in $\mathbb{R}^4$ such that $\vec{x} \cdot \vec{v}_1 = 0$ and $\vec{x} \cdot \vec{v}_2 = 0$. We solve both

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 0 \\ -1 \\ 1 \end{bmatrix} = 0 \qquad \text{and} \qquad \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 1 \\ 2 \\ 3 \end{bmatrix} = 0$$

which leads to the system and matrix

$$\begin{cases} x_1 - x_3 + x_4 = 0 \\ x_2 + 2x_3 + 3x_4 = 0 \end{cases} \quad \text{and} \quad A = \begin{bmatrix} 1 & 0 & -1 & 1 \\ 0 & 1 & 2 & 3 \end{bmatrix}.$$

All solutions are given by

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = t \begin{bmatrix} -1 \\ -3 \\ 0 \\ 1 \end{bmatrix} + u \begin{bmatrix} 1 \\ -2 \\ 1 \\ 0 \end{bmatrix} \quad \text{where } t, u \in \mathbb{R}.$$

It follows the vectors $\begin{bmatrix} -1 \\ -3 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ -2 \\ 1 \\ 0 \end{bmatrix}$ form a basis of the desired subspace.

Theorem 2.13. *Let U be a subspace of k^m with $\dim U = n$, then*

- *any list of linearly independent vectors contains n elements,*
- *any list of vectors that spans U contains at least n elements,*
- *if n vectors are linearly independent then they form a basis, and*
- *if n vectors span U, then they form a basis of U.*

Proof. The proof is left for the reader as Exercise 2.94

□

Example 2.12. Determine the values of a for which the following vectors $\vec{u}_1$, $\vec{u}_2$, $\vec{u}_3$, and $\vec{u}_4$ form a basis of $\mathbb{R}^4$.

$$\vec{u}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 4 \end{bmatrix} \qquad \vec{u}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 6 \end{bmatrix} \qquad \vec{u}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 8 \end{bmatrix} \qquad \vec{u}_4 = \begin{bmatrix} 4 \\ 5 \\ 6 \\ a \end{bmatrix}$$

Let $A = \begin{bmatrix} \vec{u}_1 & \vec{u}_2 & \vec{u}_3 & \vec{u}_4 \end{bmatrix}$. Using row operations we find the row-echelon form of A to be the following matrix.

$$\begin{bmatrix} 1 & 0 & 0 & 4 \\ 0 & 1 & 0 & 5 \\ 0 & 0 & 1 & 6 \\ 0 & 0 & 8 & a-94 \end{bmatrix}$$

Thus, $\text{rref}(A) = I_4$ if and only if $a = 95$. Therefore, by 2.3, $B = \{\vec{u}_1, \vec{u}_2, \vec{u}_3, \vec{u}_4\}$ is a basis if and only if $a = 95$.

Theorem 2.14. *The dimension of the row space of a matrix A is equal to the dimension of the column space of A.*

Proof. Gerber pg 226.

$\square$

Exercises

Exercise 2.48. Determine whether the following collection of vectors in $\mathbb{R}^3$ are linearly independent or linearly dependent. - $(0, 1, 1), (1, 2, 1), (0, 4, 6), (1, 0, -1)$ - $(0, 1, 0), (1, 2, 1), (0, -4, 6), (-1, 1, -1)$

Exercise 2.49. Determine whether the following collection of vectors in $\mathbb{R}^4$ are linearly independent or linearly dependent. - $(0, 1, 1, 1), (1, 2, 1, 1), (0, 4, 6, 2), (1, 0, -1, 2)$ - $(0, 1, 0, 1), (1, 2, 1, 3), (0, -4, 6, -2), (-$

Exercise 2.50. Show that the given vectors do not form a basis for the vector space V. - $(21, -7), (-6, 1)$; $V = \mathbb{R}^2$ - $(21, -7, 14), (-6, 1, -4), (1, 0, 0); V = \mathbb{R}^3$ - $(48, 24, 108, -72), (-24, -12, -54, 36), ($ $V = \mathbb{R}^4$

Exercise 2.51. Reduce the vectors to a basis of the vector space V. - $(1, 0), (1, 2), (2, 4), V = \mathbb{R}^2$ - $(1, 2, 3), (-1, -10, 15), (1, 2, -3), (2, 0, 6), (1, -2, 3),$ $V = \mathbb{R}^3$

Exercise 2.52. Which of the following collection of vectors in $\mathbb{R}^3$ are linearly dependent? For those that are express one vector as a linear combination of the rest.

- $(1, 1, 0), (0, 2, 3), (1, 2, 3)$
- $(1, 1, 0), (3, 4, 2), (0, 2, 3)$

Exercise 2.53. Prove 2.1.

Exercise 2.54. Prove 2.2.

Exercise 2.55. Let $S = \{v_1, v_2, ..., v_k\}$ be a set of vectors in a a vector space V. Prove that S is linearly dependent if and only if one of the vectors in S is a linear combination of all other vectors in S.

Exercise 2.56. Suppose that $S = \{v_1, v_2, v_3\}$ is a linearly independent set of vector in a vector space V. Prove that $T = \{u_1, u_2, u_3\}$ is also linearly independent where $u_1 = v_1$, $u_2 = v_1 + v_2$, and $u_3 = v_1 + v_2 + v_3$.

Exercise 2.57. Which of the following sets of vectors form a basis for the vector space V.   - $(1, 3), (1, -1)$; $V = \mathbb{R}^2$ - $(1, 3), (-2, 6)$; $V = \mathbb{R}^2$ - $(3, 2, 2), (-1, 2, 1), (0, 1, 0)$; $V = \mathbb{R}^3$ - $(3, 2, 2), (-1, 2, 0), (1, 1, 0)$; $V = \mathbb{R}^3$ - $(2, 2, 2, 2), (3, 3, 3, 2), (1, 0, 0, 0), (0, 1, 0, 0)$; $V = \mathbb{R}^4$ - $(1,1,2,0), (2,2,4,0), (1,2,3,1), (2,1,3,-1), (1,2,3,-1)$; $V = \mathbb{R}^4$

Exercise 2.58. Find a basis for the subspace of the vector space V.   - All vectors of the form (a, b, c) where $b = a + c$ where $V = \mathbb{R}^3$. - All vectors of the form (a, b, c) where $b = a - c$ where $V = \mathbb{R}^3$. - All vectors of the form $\begin{bmatrix} b - a \\ a + c \\ b + c \\ c \end{bmatrix}$ where $V = \mathbb{R}^4$.

Exercise 2.59. Let $\vec{v}_1 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$, $\vec{v}_2 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$ and $S = \text{span}(\vec{v}_1, \vec{v}_2)$.

- Is S a subspace of $\mathbb{R}^3$?
- Find a vector $\vec{u}$ in S other than $\vec{v}_1, \vec{v}_2$.
- Find scalars which verify that $3\vec{u}$ is in S.
- Find scalars which verify that $\vec{0}$ is in S.

Exercise 2.60. Let $\vec{u}_1 = \begin{bmatrix} 0 \\ 2 \\ 2 \end{bmatrix}$, $\vec{u}_2 = \begin{bmatrix} 2 \\ 0 \\ 0 \end{bmatrix}$ and $T = \text{span}(\vec{u}_1, \vec{u}_2)$. Show $S = T$ by showing $S \subseteq T$ and $T \subseteq S$ where S is defined in Exercise 2.12.

Exercise 2.61. Prove that the non-empty intersection of two subspaces of $\mathbb{R}^3$ is a subspace of $\mathbb{R}^3$.

Exercise 2.62. Let S and T be subspaces of $\mathbb{R}^3$ defined by

$$S = \text{span}\left(\begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}, \begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix}\right) \quad \text{and} \quad T = \text{span}\left(\begin{bmatrix} 2 \\ -2 \\ 3 \end{bmatrix}, \begin{bmatrix} 3 \\ -4 \\ 4 \end{bmatrix}\right).$$

Show they are the same subspace of $\mathbb{R}^3$.

Exercise 2.63. Let $\vec{v}_1, \vec{v}_2, \vec{v}_3$ be a linearly independent set of vectors. Show that if $\vec{v}_4$ is not a linear combination of $\vec{v}_1, \vec{v}_2, \vec{v}_3$, then $\vec{v}_1, \vec{v}_2, \vec{v}_3, \vec{v}_4$ is a linearly independent set of vectors.

Exercise 2.64. If
$\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$ is a linearly independent set of vectors in V, show that
$\{\vec{v}_1, \vec{v}_1 + \vec{v}_2, \vec{v}_1 + \vec{v}_2 + \vec{v}_3\}$ is also a linearly independent set of vectors in V.

Exercise 2.65. If
$\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$ is a linearly independent set of vectors in V, show that
$\{\vec{v}_1 + \vec{v}_2, \vec{v}_2 + \vec{v}_3, \vec{v}_3 + \vec{v}_1\}$ is also a linearly independent set of vectors in V.

Exercise 2.66. Let $\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$ be a linearly dependent set. Show that at least one of the $\vec{v}_i$ is a linear combination of the others.

Exercise 2.67. Prove or provide a counterexample to the following statement. If a set of vectors T spans the vector space V, then T is linearly independent.

Exercise 2.68. Which of the following are not a basis for $\mathbb{R}^3$?   -
$$\vec{v}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \vec{v}_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \vec{v}_3 = \begin{bmatrix} 1 \\ -1 \\ -1 \end{bmatrix} \text{ - } \vec{u}_1 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \vec{u}_2 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \vec{u}_3 = \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix}$$

Exercise 2.69. Let S be the space spanned by the vectors

$$\vec{v}_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 1 \end{bmatrix} \quad \vec{v}_1 = \begin{bmatrix} -1 \\ -3 \\ 1 \\ 0 \end{bmatrix} \quad \vec{v}_1 = \begin{bmatrix} 2 \\ 3 \\ 0 \\ 1 \end{bmatrix} \quad \vec{v}_1 = \begin{bmatrix} 2 \\ 0 \\ 2 \\ 2 \end{bmatrix}$$

Find the dimension of S and a subset of T which could serve as a basis
for S.

Exercise 2.70. Let $\{\vec{v}_1, \vec{v}_2, ..., \vec{v}_n\}$ be a basis for V, and suppose that
$\vec{u} = a_1\vec{v}_1 + a_2\vec{v}_2 + \cdots + a_n\vec{v}_n$ with $a_1 \neq 0$. Prove that $\{\vec{u}, \vec{v}_2, ..., \vec{v}_n\}$ is also
a basis for V.

Exercise 2.71. Let $S = \text{span}(\vec{v}_1, \vec{v}_2, \vec{v}_3)$ and $T = \text{span}(\vec{u}_1, \vec{u}_2, \vec{u}_3)$ where
$\vec{v}_i$ and $\vec{u}_i$ are defined as follows.

$$\vec{v}_1 = \begin{bmatrix} 1 \\ -1 \\ 2 \\ 0 \end{bmatrix} \quad \vec{v}_2 = \begin{bmatrix} 2 \\ 1 \\ 1 \\ 1 \end{bmatrix} \quad \vec{v}_3 = \begin{bmatrix} 3 \\ -1 \\ 2 \\ -1 \end{bmatrix} \quad \vec{u}_1 = \begin{bmatrix} 3 \\ 0 \\ 3 \\ 1 \end{bmatrix} \quad \vec{u}_2 = \begin{bmatrix} 1 \\ 2 \\ -1 \\ 1 \end{bmatrix} \quad \vec{u}_3 = \begin{bmatrix} 4 \\ -1 \\ 5 \\ 1 \end{bmatrix}$$

Is one of these two subspaces strictly contained in the other or are they
equal?

Exercise 2.72. Let $S = \text{span}(\vec{v}_1, \vec{v}_2, \vec{v}_3)$ where $\vec{v}_i$ are defined as follows.

$$\vec{v}_1 = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 1 \end{bmatrix} \quad \vec{v}_2 = \begin{bmatrix} 2 \\ -1 \\ 1 \\ -3 \end{bmatrix} \quad \vec{v}_3 = \begin{bmatrix} 1 \\ 3 \\ 4 \\ 2 \end{bmatrix} \quad \text{and} \quad \vec{u} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 1 \end{bmatrix}$$

Is the vector $\vec{u}$ in S?

Exercise 2.73. If possible, find a value of a so that the vectors

$$\begin{bmatrix} 1 \\ 2 \\ a \end{bmatrix} \quad \begin{bmatrix} 0 \\ 1 \\ a-1 \end{bmatrix} \quad \begin{bmatrix} 3 \\ 4 \\ 5 \end{bmatrix}$$

are linearly independent.

Exercise 2.74. Let $S = \text{span}(\vec{v}_1, \vec{v}_2, \vec{v}_3)$ where $\vec{v}_i$ are defined as follows.

$$\vec{v}_1 = \begin{bmatrix} 1 \\ -1 \\ 2 \\ 3 \end{bmatrix} \quad \vec{v}_2 = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix} \quad \vec{v}_3 = \begin{bmatrix} 3 \\ -2 \\ 5 \\ 7 \end{bmatrix} \quad \text{and} \quad \vec{u} = \begin{bmatrix} 1 \\ 1 \\ 0 \\ -1 \end{bmatrix}$$

Find a basis of S which includes the vector $\vec{u}$.

Exercise 2.75. Find a vector $\vec{u}$ in $\mathbb{R}^4$ such that $\vec{u}$ and the vectors

$$\vec{v}_1 = \begin{bmatrix} 1 \\ -1 \\ -1 \\ 1 \end{bmatrix} \qquad \vec{v}_2 = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 1 \end{bmatrix} \qquad \vec{v}_3 = \begin{bmatrix} 1 \\ 2 \\ 1 \\ 1 \end{bmatrix}$$

for a basis of $\mathbb{R}^4$.

Exercise 2.76. Show that every subspace of V has no more than n linearly independent vectors.

Exercise 2.77. Find two bases of $\mathbb{R}^4$ that have only the vectors $\vec{e}_3$ and $\vec{e}_4$ in common.

Exercise 2.78. Prove that if a list of vectors is linearly independent so is any sublist.

Exercise 2.79. Suppose $\vec{v}_1, \vec{v}_2, \vec{v}_3$ and $\vec{v}_1, \vec{v}_2, \vec{v}_4$ are two sets of linearly dependent vectors, and suppose that $\vec{v}_1$ and $\vec{v}_2$ are linearly independent. Prove that any set of three vectors chosen from $\vec{v}_1, \vec{v}_2, \vec{v}_3, \vec{v}_4$ is linearly dependent.

Exercise 2.80. If $\vec{u}$ and $\vec{v}$ are linearly independent vectors in V, prove that the vectors $a\vec{u} + b\vec{v}$ and $c\vec{u} + d\vec{v}$ are also linearly independent if and only if $ad - bc \neq 0$.

Exercise 2.81. Complete the proof of 2.12.

Exercise 2.82. Let U be the collection of vectors that satisfy the equations $x + y + z = 0$ and $x + 2y - z = 0$. Show U is a subspace of $\mathbb{R}^3$, find a basis for U, and find $\dim(U)$.

Exercise 2.83. Let U be the collection of vectors that satisfy the equations $x + y + z = 0$, $x + 2y - z = 0$, and $y - 2z = 0$. Show U is a subspace of $\mathbb{R}^3$, find a basis for U, and find $\dim(U)$.

Exercise 2.84. Show that the only subspaces of $\mathbb{R}$ are $\{\vec{0}\}$ and $\{\mathbb{R}\}$.

Exercise 2.85. Show that the only subspaces of $\mathbb{R}^2$ are $\{\vec{0}\}$, $\{\mathbb{R}^2\}$, and any set consisting of all scalar multiples of a nonzero vector. Describe these subspaces geometrically.

Exercise 2.86. Determine the various types of subspaces of $\mathbb{R}^3$ and describe them geometrically.

Exercise 2.87. For $\vec{b} \neq \vec{0}$, show that the set of solutions of the $n \times m$ linear system $A\vec{x} = \vec{b}$, is not a subspace of V.

Exercise 2.88. Suppose that $\vec{v}_1, \vec{v}_2, ..., \vec{v}_n$ are linearly independent in $\mathbb{R}^n$. Show that if A is an $n \times n$ matrix with $\mathrm{rref}(A) = I_n$, then $A\vec{v}_1, A\vec{v}_2, ..., A\vec{v}_n$ are also linearly independent in $\mathbb{R}^n$.

Exercise 2.89. Let $S = \{v_1, v_2, ..., v_s\}$ and $T = \{u_1, u_2, ..., u_t\}$ be two sets of vectors in V where each $\vec{u}_i$, $(i = 1, 2, ..., t)$ is a linear combination of the vectors in S. Show that $\vec{w} = a_1\vec{u}_1 + a_2\vec{u}_2 + \cdots + a_m\vec{u}_t$ is a linear combination of the vectors in S.

Exercise 2.90. Let $S = \{v_1, v_2, ..., v_m\}$ be a set of non-zero vectors in a vector space V such that every vector in V can be uniquely as a linear combination of the vectors in S. Prove that S is a basis for V.

Exercise 2.91. Find a basis for the solution space of the homogeneous system $(\lambda I_n - A)\vec{x} = \vec{0}$ for the given λ and A.

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

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

Exercise 2.92. Prove 2.9.

Exercise 2.93. Prove 2.1.

Exercise 2.94. Prove 2.13.

2.3 Introduction to Linear Spaces

Example 2.13. Show that the set of solution to a homogenous system form a linear space with standard operations.

Example 2.14. Show that the set of vectors for which a particular linear system has a solution is a linear space.

Definition 2.7. Let $\mathbb{F}$ be a field (whose elements are called **scalars**) and let V be a nonempty set (whose elements are called **vectors**) on which two operations, called **addition** and **scalar multiplication**, have been defined. The addition operation (denoted by $+$), assigns to each pair $(u, v) \in V \times V$, a unique vector $u + v$ in V. The scalar multiplication operation (denoted by juxtaposition), assigns to each pair $(a, v) \in \mathbb{F} \times V$ a unique vector av in V.
We call V a **linear space** if the following axioms (A1)-(A8) are also satisfied.

- For all $u, v \in V$, $u + v = v + u$.
- For all $u, v, w \in V$, $(u + v) + w = u + (v + w)$.
- There exists $0 \in V$ such that $v + 0 = v$ for all $v \in V$.
- For every $v \in V$, there exists $w \in V$ such that $v + w = 0$.
- For all $v \in V$, $1v = v$.
- For all $a, b \in \mathbb{F}$ and $u \in V$, $(ab)v = a(bv)$.
- For all $a \in \mathbb{F}$ and $u, v \in V$, $a(u + v) = au + av$.
- For all $a, b \in \mathbb{F}$ and $u \in V$, $(a + b)u = au + bu$.

If $\mathbb{F} = \mathbb{R}$ then V is a called a **real linear space** . If $\mathbb{F} = \mathbb{C}$ then V is called a **complex linear space** . We denote the zero vector (A3) as 0, to distinguish between the zero vector and the zero 0 in the field of scalars.

Example 2.15. Let $V = \{(x, y) \mid y = mx\}$, where m is a fixed real number and x is an arbitrary real number. Show that V is a linear space.

Example 2.16. Let $V = \{(x, y, x) \mid ax + by + cz = 0\}$ where a, b and c are fixed real numbers. Show that V is a linear space with the standard operations.

::: {#exm- } [Matrix Space] Show that the set $M_{m \times n}$ of all $m \times n$ matrices, with ordinary addition of matrices and scalar multiplication, forms a linear space.

:::

::: {#exm- } [Polynomial Space] Show that the set $P(t)$ of all polynomials with real coefficients, under the ordinary operations of addition of polynomials and multiplication of a polynomial by a scalar, forms a linear space. Show that the set of all polynomials with real coefficients of degree less than or equal to n, under the ordinary operations of addition of polynomials and multiplication of a polynomial by a scalar, forms a linear space.

:::

::: {#exm- } [Function Space] Show that the set $F(x)$ of all functions that map the real numbers into itself is a linear space. Show that the set $F[a, b]$ of all functions on the interval $[a, b]$ using the standard operations is a linear space.

:::

::: {#exm- } [The Space of Infinite Sequences] Show that the set of all infinite sequences of real numbers is a linear space, where addition and scale multiplication are defined term by term.

:::

::: {#exm- } [The Space of Linear Equations] Show that the set L_n of all linear equations with n variables, forms a linear space.

:::

Lemma 2.2. *Every linear space V has a unique additive identity (denoted by 0).*

Proof. Let u_1 and u_2 be additive identities in V, then $v + u_1 = v$ and $v + u_2 = v$ for every $v \in V$. Thus, $u_1 = u_1 + u_2 = u_2 + u_1 = u_2$ as desired.

$\square$

Lemma 2.3. *Every $v \in V$ has a unique additive inverse, (denoted by $-v$).*

Proof. Let v_1 and v_2 be additive inverses of w in V, then $w + v_1 = 0$ and $w + v_2 = 0$. Thus,

$$v_1 = v_1 + 0 = v_1 + (w + v_2) = (v_1 + w) + v_2 = (w + v_1) + v_2 = 0 + v_2 = v_2$$

as desired.

□

Lemma 2.4. *If $v \in V$, then $0\,v = 0$.*

Proof. Let $v \in V$, then $v = 1v = (1+0)v = 1v + 0v = v + 0v$ which shows that $0v$ is the additive identity of V, namely $0v = \mathbf{0}$.

□

Lemma 2.5. *If $a \in \mathbb{F}$, then $a\,0 = 0$.*

Proof. Let $a \in \mathbb{F}$, then

$$a\mathbf{0} = a(\mathbf{0} + \mathbf{0}) = a\mathbf{0} + a\mathbf{0}$$

which shows that $a\mathbf{0}$ is the additive identity of V, namely $a\mathbf{0} = \mathbf{0}$.

□

Lemma 2.6. *If $v \in V$, then $-(-v) = v$.*

Proof. Let $v \in V$, then

$$v + (-1)v = 1v + (-1)v = (1 + (-1))v = 0v = \mathbf{0}$$

which shows that $(-1)v$ is the unique additive inverse of v namely, $(-1)v = -v$.

□

Lemma 2.7. *If $v \in V$, then $(-1)\,v = -v$.*

Proof. Since $-v$ is the unique additive inverse of v, $v + (-v) = \mathbf{0}$. Then $(-v) + v = \mathbf{0}$ shows that v is the unique additive inverse of $-v$, namely, $v = -(-v)$ as desired.

□

Lemma 2.8. *If $a\,v = 0$, then $a = 0$ or $v = 0$.*

Proof. Suppose $a \neq 0$. If $av = \mathbf{0}$ then $v = 1v = (a^{-1}a)v = a^{-1}(av) = a^{-1}\mathbf{0} = \mathbf{0}$. Otherwise $a = 0$ as desired.

$\square$

Example 2.17. Let V be a linear space with $u \in V$ and let a and b be scalars. Prove that if $au = bu$ and $u \neq 0$, then $a = b$.

Let V be a linear space and U a nonempty subset of V. If U is a linear space with respect to the operations on V, then U is called a subspace of V.

Theorem 2.15. *A subset U of V is a **linear subspace** of V if and only if U has the following properties:*

- *U contains the zero vector of V,*
- *U is closed under the addition defined on V, and*
- *U is closed under the scalar multiplication defined on V.*

Proof. Koman pg 103.

$\square$

More generally, a subset U of V is called a **subspace** of V if U is also a vector space using the same addition and scalar multiplication as on V. Any vector space is a subspace of itself. The set containing just the 0 vector is also a subspace of any vector space. Given any vector space with a nonzero vector v, the scalar multiples of v is a vector subspace of V and is denoted by $\langle v \rangle$. Because any linear space V has V and 0 as subspaces, these subspaces are called the **trivial subspaces** of V. All other subspaces are called **proper subspaces** of V.

Example 2.18. Give an example of a real linear space V and a nonempty set S of V such that, whenever u and v are in S, $u + v$ is in S but S is not a subspace of V.

Example 2.19. Give an example of a real linear space V and a nonempty set S of V such that, whenever u and v are in S, cu is in S for every scalar c but S is not a subspace of V.

Example 2.20. Show that $P_n[0, 1]$ is a proper subspace of $C[0, 1]$.

Example 2.21. Show that $C'[0,1]$ (continuous first derivative) is a proper subspace of $C[0,1]$.

Example 2.22. Show that $R[0,1]$ (Riemann integrable) is a proper subspace of $C[0,1]$.

Example 2.23. Show that $D[0,1]$ (Differenable functions) is a proper subspace of $C[0,1]$.

Definition 2.8. A **linear combination** of a list of vectors $(v_1, v_2, \ldots, v_m)$ in V is a vector of the form $a_1\vec{v}_1 + a_2\vec{v}_2 + \cdots + a_m\vec{v}_m$ where $a_1, a_2, \ldots, a_m \in k$.

Lemma 2.9. *Let U be a nonempty subset of a vector space V. Then U is a subspace of V if and only if every linear combination of vectors in U is also in U.*

Proof. If U is a subspace of V, then U is a vector space and so is closed under linear combinations by definition of vector space. Conversely, suppose every linear combination of vectors in U is also in U. Thus for any $a, b \in k$, $au + bv \in U$ for every $u, v \in U$. In particular, when $a = b = 1$ then $u + v \in U$ and so U is closed with respect to addition. Notice when $b = 0$ and $a = -1$, then $-u \in U$ for every $u \in U$ and so U is closed under inverses. Notice when $u = v$, $a = 1$, and $b = -1$ then $u + (-u) = 0 \in U$ so U contains the identity element. The rest of the axioms in the definition of a vector space hold by containment.

$\square$

Definition 2.9. The intersection and union of subspaces is just the intersection and union of the subspaces as sets. The **sum of subspaces** $U_1, U_2, \ldots, U_m$ of a vector space U is defined by

$$U_1 + U_2 + \cdots + U_m = \{u_1 + u_2 + \cdots + u_m \mid u_i \in U_i \text{ for } 1 \le i \le m\}.$$

Lemma 2.10. *Let V be a linear space over a field k. The intersection of any collection of subspaces of V is a subspace of V.*

Proof. Let $\{U_i \mid i \in I\}$ be a collection of subspaces where I is some indexed set. Let $a, b \in k$ and $u, v \in \cap_{i \in I} U_i$. Since each U_i is a subspace of V,

$au + bv \in U_i$ for every $i \in I$. Thus $au + bv \in \cap_{i \in I} U_i$ and therefore $\cap_{i \in I} U_i$ is a subspace of V.

$\square$

Example 2.24. Show that the x-axis and the y-axis are subspaces on $\mathbb{R}^2$, yet the union of these axis is not.

Lemma 2.11. *Let V be a linear space over a field k. The union of two subspaces of V is a subspace of V if and only if one of the subspaces is contained in the other.*

Proof. Suppose U and W are subspaces of V with $U \subseteq W$. Then $U \cup W = W$ and so $U \cup W$ is also a subspace of V. Conversely, suppose $U, W, U \cup W$ are subspaces of V and suppose $u \in U$. If $u \in W$ then U is contained in W as desired. Thus we assume, $u \notin W$. If $w \in W$, then $u + w \in U \cup W$ and either $u + w \in U$ or $u + w \in W$. Notice $u + w \in W$ and $w \in W$ together yield $u \in W$ which is a contradiction. Thus $u + w \in U$ and so $w \in U$ which yields $W \subseteq U$ as desired.

$\square$

Lemma 2.12. *Let V be a linear space over a field k. The sum $U_1 + U_2 + \cdots + U_m$ is the smallest subspace containing each of the subspaces $U_1, U_2, \ldots, U_m$.*

Proof. The sum of two subspaces is a subspace since the sum of two subspaces is closed under linear combinations. Thus $U_1, U_2, \ldots, U_m$ is a subspace containing U_i for each $1 \leq i \leq m$. Let U be another subspace containing U_i for each $1 \leq i \leq m$. If $u \in U_1 + \cdots + U_m$, then u has the form $u = u_1 + \cdots + u_m$ where each $u_i \in U_i \subseteq U$. Since U is a subspace $u \in U$ and so $U_1 + U_2 + \cdots + U_m$ is the smallest such subspace.

$\square$

Definition 2.10. If $\{v_1, v_2, \ldots, v_m\}$ is a subset of a linear space V, then the subspace of all linear combinations of these vectors is called the **subspace generated (spanned)** by $v_1, v_2, \ldots, v_m$. The **spanning set** of the list of vectors $(v_1, v_2, \ldots, v_m)$ in V is denoted by

$$\operatorname{span}(v_1, v_2, \ldots, v_m) = \{a_1 \vec{v}_1 + a_2 \vec{v}_2 + \cdots + a_m \vec{v}_m \mid a_1, a_2, \ldots, a_m \in \mathbb{F}\}.$$

Lemma 2.13. *The **span** of a list of vectors in V is the smallest subspace of V containing all the vectors in the list.*

Proof. Let $(v_1, v_2, \dots, v_n)$ be a list of vectors in V and let S denote $\text{span}(v_1, v_2, \dots, v_n)$. Clearly, S contains v_i for each $1 \le i \le n$. Let $u, v \in S$ and $a, b \in k$. Then there exists $a_1, a_2, \dots, a_n$ in k and $b_1, b_2, \dots, b_n$ in k such that $u = a_1 v_1 + \cdots a_n v_n$ and $v = b_1 v_1 + \cdots b_n v_n$. Then
$$au + bv = (aa_1 + bb_1)v_1 + \cdots + (aa_n + bb_n)v_n$$
which shows $au + bv \in S$ since $aa_i + bb_i \in k$ for each $1 \le i \le n$. Thus S is a subspace containing each of the v_i. Let T be a subspace containing v_i for $1 \le i \le n$. If $s \in S$, then there exists $c_1, c_2, \dots, c_n \in k$ such that $s = c_1 v_1 + \cdots + c_n v_n$. Since $v_i \in T$ for each i and T is closed under linear combinations (since T is a subspace), $s \in T$. Meaning $S \subseteq T$, so indeed S is the smallest subspace of V containing all the vectors v_i.

$\square$

Definition 2.11. Let $\emptyset$ denote the empty set. Then $\text{span}(\emptyset) = \{0\}$.

Example 2.25. Let $A = 1 \quad 1 \, 0 \quad 0$. Show that $S = \{X \in M_{2 \times 2} \mid AX = XA\}$ is a subspace of $M_{2 \times 2}$ under the standard operations.

Example 2.26. Let $f_1 = x^2 + 1, f_2 = 3x - 1, f_3 = 2$. Determine the subspace generated by f_1, f_2, f_3 in P_4.

Example 2.27. Let $A_1 = \begin{bmatrix} 1 & 0 & 0 & 3 \\ 0 & 0 & 2 & 0 \end{bmatrix}$ and $A_2 = \begin{bmatrix} 0 & 2 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}$. Determine the subspace generated by A_1 and A_2 in $R_{2 \times 4}$.

Example 2.28. Describe $\text{span}(0)$.

Example 2.29. Consider the subset $S = \{x^3 - 2x^2 + x - 3, 2x^3 - 3x^2 + 2x + 5, 4x^3 - 7x^2 + 4x - 1, 4x^2 + x - 3\}$ of P. Show that $3x^3 - 8x^2 + 2x + 16$ is in $\text{span}(S)$ by expressing it as a linear combination of the elements of S.

Example 2.30. Determine if the matrices

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

span $M_{2\times 2}$.

2.4   Linear Independence and Bases

Let V be a linear save over a field $\mathbb{F}$. A subset S of V is said to be
linearly dependent if there exist distinct $v_1, v_2, \ldots, v_n$ in S and scalars
$a_1, a_2, \ldots, a_n$ in $\mathbb{F}$, not all zero, such that $a_1\vec{v}_1 + a_2\vec{v}_2 + \cdots + a_m\vec{v}\_0$. If a set
S cntains only finitely many v_i we sometimes say that $v_1, v_2, \ldots, v_n$ are
dependent.

Definition 2.12. A list of vectors $(v_1, v_2, \ldots, v_m)$ in V is called **linearly
independent** if the only choice of $a_1, a_2, \ldots, a_m \in k$ is $a_1 = \cdots = a_m =
0$. A list of vectors that are not linearly independent is called **linearly
dependent**.

Notice from the definition we can conclude that any set which contains
a linearly dependent set is linearly dependent. Any subset of a linearly
independent set is linearly independent. Any set which contains the zero
vector is linearly dependent. A set S of vectors is linearly independent if
and only if each finite subset of S is linearly independent. (Show!)

::: {#lem- } [Linear Dependence] If $(v_1, v_2, \ldots, v_m)$ is linearly dependent
in V and $v_1 \neq 0$ then there exists $j \in 2, \ldots, m$ such that the following hold:
$v_j \in \text{span}(v_1, \ldots, v_{j-1})$ and if the jth term is removed from $(v_1, v_2, \ldots, v_m)$,
the span of the remaining list equals $\text{span}(v_1, v_2, \ldots, v_m)$. :::

Proof. The proof is left for the reader as Exercise **??**.

$\square$

Example 2.31. Show that the following subset of $M_{2\times 2}$ is a linear de-
pendent set.

$$\left\{ \begin{bmatrix} 2 & 3 \\ -1 & 4 \end{bmatrix}, \begin{bmatrix} -11 & 3 \\ -2 & 2 \end{bmatrix}, \begin{bmatrix} 6 & -1 \\ 3 & 4 \end{bmatrix}, \begin{bmatrix} -1 & 0 \\ 2 & 2 \end{bmatrix} \right\} \tag{2.9}$$

Example 2.32. Suppose that S is the subset

$$S = \{2x^3 - x + 3, 3x^3 + 2x - 2, x^3 - 4x + 8, 4x^3 + 5x - 7\}$$

of P_3.

- Show that S is linear dependent.
- Show that every three element subset of S is linear dependent.
- Show that every two element subset of S is linear independent.

Example 2.33. Show that no linear independent set can contain zero.

::: {#lem- } [Linear Independence Lemma] Let S be a linearly indepen-
dent subset of a vector space V. Suppose v is a vector in V which is not
in the subspace spanned by S. Then the set obtained by adjoining v to
S is linearly independent. :::

Proof. Suppose $u_1, \dots, u_n$ are distinct vectors in S and that $a_1 u_1 + \cdots + a_n u_n + av = 0$. Then $a = 0$; for otherwise,

$$v = \left(-\frac{a_1}{a}\right) u_1 + \cdots + \left(-\frac{a_n}{a}\right) u_n$$

and v is in the subspace spanned by S. Thus $a_1 u_1 + \cdots + a_n u_n = 0$, and
since S is a linearly independent set each $a_i = 0$.

$\square$

Theorem 2.16. *If S is a set of vectors in a vector space V over a field
F then the following are equivalent.*

- *S is linearly independent and spans V*
- *For every vector $v \in V$, there is a unique set of vectors $v_1, \dots, v_n$ in S,
  along with a unique set of scalars in $\mathbb{F}$ for which $v = a_1 v_1 + \cdots + a_n v_n$*
- *S is a minimal spanning set in the sense that S spans V, and any
  proper subset of S does not span V*
- *S is a maximal linearly independent set in the sense that S is linearly
  independent, but any proper superset of S is not linearly independent.*

Proof. $(i) \Leftrightarrow (ii)$: Then S is a spanning set. If some proper subset S' of S
also spanned V, than any vector in $S-S$; would be a linear combination of
the vectors in S'. contradicting the fact that the vectors in S are linearly
independent. Conversely, if S is a minimal spanning set, then it must be
linearly independent. For if not, some vector $s \in S$ would be a linear of
the other vectors in S, and so $S - S'$ would be a proper spanning subset
of S, which is not possible.

$(i) \Leftrightarrow (iv)$: Then S is linearly independent. If S were not maximal,
there would be a vector $v \in V - S$ for which the set $S \cup \{v\}$ is linear
independent. But then V is not in the span of S, contradicting the fact
that S is a spanning set. Hence, S is a maximal linearly independent set,
and so (i) implies (iv). Conversely, if S is a maximal independent set,.

then it must span V, for if not, we could find a vector $v \in S - S'$ that is not a linear combination of the vectors in S. Hence, $S \cup \{v\}$ would be a contradiction.

□

Definition 2.13. A **basis** of V is a list of vectors in V that is linearly independent and spans V.

Example 2.34. Find a basis for the space of all 2×2 matrices S such that
$$\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} S = S.$$

Let $S = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$. Then
$$\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

meaning
$$\begin{bmatrix} a+c & b+d \\ a+c & b+d \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}.$$

So $a+c = a, b+d = b, a+c = c$, and $b+d = d$. These imply, respectively that $a = b = c = d = 0$.

Corollary 2.4. *A list $(v_1, \dots, v_m)$ of vectors in V is a basis of V if and only if every $v \in V$ can be written uniquely in the form $v = a_1 v_1 + \dots + a_m v_m$ where $a_1, a_2, \dots, a_m \in k$.*

Proof.   □

Corollary 2.5. *Any n linearly independent vectors in a linear space V of dimension n constitute a basis for V.*

Proof.   □

Example 2.35. Show that the space of all $m \times n$ matrices over a field $\mathbb{F}$ has dimension mn.

Corollary 2.6. *Any finitely generated linear space, generated by a asset of nonzero vectors, has a basis.*

Proof.   □

The zero linear space has no basis, because any subset contains the zero vector and must be linearly dependent.

Example 2.36. Show that

$$B = \{[2, 3, 0, -1], [-1, 1, 1, -1], [3, 2, -1, 0], [2, 3, 0, -1]\}$$

is a maximal linearly independent subset of

$$S = \{[1, 4, 1, -2], [-1, 1, 1, -1], [3, 2, -1, 0], [2, 3, 0, -1]\}.$$

Determine $\dim \text{span}(S)$ and determine whether or not $\text{span}(S) = P_3$.

2.5   Finite-dimensional Linear Spaces

Definition 2.14. The vector space V is a **direct sum** of subspaces $U_1, U_2, \ldots, U_m$ of V, written $V = U_1 \oplus \cdots \oplus U_m$, if each element of V can be written uniquely as a sum $u_1 + u_2 + \cdots + u_m$ where each $u_j \in U_j$.

Lemma 2.14. *If $U_1, U_2, \ldots, U_m$ are subspaces of V, then $V = U_1 \oplus \cdots \oplus U_m$ if and only if both conditions hold:*

- *$V = U_1 + \cdots + U_m$.*
- *the only way to write 0 as a sum $u_1 + u_2 + \cdots + u_m$ where each $u_j \in U_j$, is by taking all the u_j's equal to 0.*

Proof. If $V = U_1 \oplus U_2 \oplus \cdots \oplus U_m$, then every element in V can be written uniquely in the form $u_1 + u_2 + \cdots + u_m$ where each $u_j \in U_j$ for $1 \leq i \leq m$. Thus both conditions listed above are satisfied. Conversely, suppose both conditions hold and assume $u = u_1 + \cdots + u_m$ where $u_i \in U_i$ and $u = v_1 + \cdots + v_m$ where $v_i \in U_i$. Since

$$u - u = (u_1 + \cdots + u_m) - (v_1 + \cdots + v_m) = (u_1 - v_1) + \cdots + (u_m - v_m) = 0$$

it follows $u_i = v_i$ for $1 \leq i \leq m$; and so uniqueness is established.

□

Example 2.37. Let V be the linear space of all functions from $\mathbb{R}$ to $\mathbb{R}$ and let V_e and V_o be the set of all even functions $(f(-x) = f(x))$ and the set of odd functions $(f(-x) = -f(x))$, respectively.

- Prove that V_e and V_o are subspaces of V.
- Prove that $V_e + V_o = V$
- Prove that $V_e \cap V_o = \{0\}$.

Lemma 2.15. *If U and W are subspaces of V, then $V = U \oplus W$ if and only if $V = U + W$ and $U \cap W = \{0\}$.*

Proof. If $V = U \oplus W$ then $V = U + W$ and $\{0\} \subseteq U \cap W$ are immediate. Suppose $v \in U \cap W$. Since $v \in U \oplus W$, there exists unique $u \in U$ and $w \in W$ such that $v = u + w$. Assume for a contradiction, $u \neq 0$. Then $u = u + 0$ where $0 \in W$ and $u \in U$ shows that $u \notin W$ since $V = U \oplus W$. Also, $v = u + w \in U \cap W$ so $u + w \in W$ and $w \in W$ implies $u \in W$. This contradiction leads to $u = 0$. Similarly, $w = 0$ and so $v = u + w = 0$ which yields $U \cap W = \{0\}$. Conversely, suppose $V = U + W$ and $U \cap W = \{0\}$. If $0 = u + w$ then $u = -w \in W$ and together with $u \in U$ yields $u = 0$. Thus, $w = 0$ also. So the only way to write 0 as a sum $u + w$ is to have $u = w = 0$ therefore $V = U \oplus W$.

$\square$

Example 2.38. Prove or give a counterexample: if U_1, U_2, and W are subspaces of V such that $U_1 + W = U_2 + W$, then $U_1 = U_2$. False, here is a counterexample. Let $V = \mathbb{R}^2 = W$ with $U_1 = \{(x,0) \mid x \in \mathbb{R}\}$ and $U_2 = \{(0,y) \mid y \in \mathbb{R}\}$. Then $U_1 + W = U_2 + W = \mathbb{R}^2$. However, $U_1 \neq U_2$.

Example 2.39. Prove or give a counterexample: if U_1, U_2, W are subspaces of V such that $V = U_1 \oplus W$ and $V = U_2 \oplus W$, then $U_1 = U_2$. False, here is a counterexample. Let $V = \mathbb{R}^2$ with $W = \{(x,x) \mid x \in \mathbb{R}\}$, $U_1 = \{(x,0) \mid x \in \mathbb{R}\}$, and $U_2 = \{(0,y) \mid y \in \mathbb{R}\}$. All of these sets are subspaces. Let $u \in U_1 \cap W$, then $u = (x,0)$ and $u = (z,z)$ so $z = 0$ and $x = 0$ which implies $u = (0,0)$. In fact $U_1 \cap W = \{0\}$. Thus, $\mathbb{R}^2 = U_1 \oplus W$. Also $\mathbb{R}^2 = U_2 \oplus W$. However, $(1,0) \in U_1$ and $(1,0) \notin U_2$ showing $U_1 \neq U_2$.

Example 2.40. Suppose m is a positive integer. Is the set consisting of 0 and all polynomials with coefficients in F and with degree equal to m a subspace of $\mathcal{P}(\mathbb{F})$?

Definition 2.15. We call a linear space V a **finite-dimensional linear space** if there is a finite list of vectors $(v_1, \ldots, v_m)$ with $\text{span}(v_1, \ldots, v_m) = V$. If a linear space is not a finite-dimensional vector space it is called an **infinite-dimensional vector space**.

Lemma 2.16. *Let V be a finite-dimensional vector space. Then the length of every linearly independent list of vectors in V is less than or equal to the length of every spanning list of vectors in V.*

Proof. □

Lemma 2.17. *Let V be a finite-dimensional vector space. Then V only has finite-dimensional subspaces.*

Proof. □

Theorem 2.17. *Let V be a finite-dimensional vector space. Then every spanning list $(v_1, \ldots, v_m)$ of a vector space can be reduced to a basis of the vector space,*

Proof. □

Theorem 2.18. *Let V be a finite-dimensional vector space. Then if V is finite-dimensional then V has a basis.*

Proof. By definition, every finite-dimensional vector space has a finite spanning set. By **??**, every spanning set can be reduced to a basis, and so every finite-dimensional vector space does indeed have a basis.

□

Theorem 2.19. *Let V be a finite-dimensional vector space. Then if W is a subspace of V, every linearly independent subset of W is finite and is part of a finite basis for W.*

Proof. □

Theorem 2.20. *Let V be a finite-dimensional vector space. Then every linearly independent list of vectors in V can be extended to a basis of V,*

Proof.   □

Theorem 2.21. *Let V be a finite-dimensional vector space. Then if U is a subspace of V, then there is a subspace W of V such that $V = U \oplus W$, and*

Proof.   □

Theorem 2.22. *Let V be a finite-dimensional vector space. Then any two bases of V have the same length.*

Proof.   □

Definition 2.16. The **dimension** of a finite-dimensional vector space is defined to be the length of any basis of the vector space.

Theorem 2.23. *Let V be a finite-dimensional vector space. Then if U is a subspace of V then $dim U \leq dim V$.*

Proof.   □

Theorem 2.24. *Let V be a finite-dimensional vector space. Then every spanning list of vectors in V with length dim V is a basis of V.*

Proof.   □

Theorem 2.25. *Let V be a finite-dimensional vector space. Then every linearly independent list of vectors in V with length dim V is a basis of V.*

Proof.   □

Theorem 2.26. *Prove that if U_1 and U_2 are subspaces of a finite-dimensional vector space, then*

$$dim(U_1 + U_2) = dimU_1 + dimU_2 - dim(U_1 \cap U_2).$$

Proof. □

Theorem 2.27. *If V is finite-dimensional, and $U_1, \ldots, U_m$ are subspaces of V such that*

- $V = U_1 + \cdots + U_m$ *and*
- $dimV = dimU_1 + \cdots + dimU_m,$

then $V = U_1 \oplus \cdots \oplus U_m$.

Proof. Choose a basis for each U_j. Put these bases together in one list, forming a list that spans V and has length $\dim(V)$. Thus this list is a basis of V and in particular it is linearly independent. Now suppose that $u_j \in U_j$ for each j are such that $0 = u_1 + \cdots + u_m$. We can write each u_j as a linear combination of the basis vectors of U_j. Substituting these expressions above, we have written 0 as a linear combination of the basis vectors of V. Thus all scalars used in the linear combinations must be zero. Thus each $u_j = 0$ which proves $V = U_1 \oplus \cdots \oplus U_m$.

□

2.6 Infinite-dimensional Linear Spaces

Example 2.41. Prove that k^∞ is infinite-dimensional. Suppose $\mathbb{F}^\infty$ is finite-dimensional with dimension n. Let v_i be the vector in $\mathbb{F}^\infty$ consisting of all 0's except with 1 in the i-th position for $i = 1, \ldots, n$. The vectors $(v_1, \ldots, v_n)$ are linearly independent in $\mathbb{F}^\infty$ and so they must form a basis; however they do not span since the vector v_{n+1} consisting of all 0's except with 1 in the $(n+1)$-th position. Thus $\mathbb{F}^\infty$ can not be finite-dimensional.

Example 2.42. Prove that the real vector space consisting of all continuous real-valued functions on the interval $[0, 1]$ is infinite-dimensional. Notice that $f(x) = x^n$ is continuous on $[0, 1]$ for every positive integer n. If this vector space is finite-dimensional of say dimension n then the list of vectors $(1, x, x^2, \ldots, x^{n-1})$ must form a basis since they are linearly independent and there are n of them. However, x^n is not in the span of this list and so this list can not be a basis. This contradiction shows that this vector space must be infinite-dimensional.

Example 2.43. Prove that V is infinite-dimensional if and only if there is a sequence $v_1, v_2, \dots$ of vectors in V such that $(v_1, \dots, v_n)$ is linearly independent for every positive integer n. Suppose V is infinite-dimensional. Then V has no finite spanning list. Pick $v_1 \neq 0$. For each positive integer n choose $v_{n+1} \notin \mathrm{span}(v_1, \dots, v_n)$, by the linear independence lemma, and for each positive integer n, the list $(v_1, \dots, v_n)$ is linearly independent. Conversely, suppose there is a sequence $v_1, v_2, \dots$, of vectors in V such that $(v_1, \dots, v_n)$ is linearly independent for every positive integer n. If V is finite-dimensional, then it has a spanning list with M elements. By the previous theorem, every linearly independent list has no more than M elements. Therefore, V is infinite-dimensional.

Theorem 2.28. *Every vector space has a basis. Moreover, any two bases have the same carnality.*

Proof. Let V be a nonzero vector space and consider the collection A of all linearly independent subsets of V. This collection is nonempty, since any single nonzero vector forms a linearly independent set. Now, if $I_1 \subset I_2 \subset \cdots$ is a chain of linearly independent subsets of V, then the union U is also a linearly independent set. Hence, every chain in A has an upper bound in A, and according to Zorn's lemma, A must contain a maximal element, that is, V has a maximal linearly independent set, which is a basis for V.

We may assume that all bases for V are infinite sets, for if any basis is finite, then V has a finite spanning set and so is a finite-dimensional vector space. Let B be a basis for V. We may write $B = \{b_i \mid i \in I\}$ where I is some indexed set, used to index the vectors in B. Note that $|I| = |B|$. Now let C be another basis for V. Then any vector $c \in C$ can be written as a finite linear combination of the vectors in B, where all the coefficients are nonzero, say $c = \sum_{i \in U_c} r_i b_i$. Here U_c is a finite subset of the index set I. Now, because C is a basis for V, the union of all of the U_c's as C varies over C must be in I, in symbols, $\bigcup_{c \in C} U_c = I$. For if all vectors in the basis C can be expressed as a finite linear combination of the vectors $B - \{b_k\}$ spans V, which is not the case. Therefore, $|B| = |I| \leq |C|\alpha_0 = |C|$. But we may also reverse the roles of B and C we obtain the reverse inequality. Therefore, $|B| = |C|$ as desired.

$\square$

Chapter 3

Linear Transformations

This book is designed to help students from beginners to pro as they learn and grow in their understanding of linear transformations. Linear transformations are a powerful tool that can be used in a variety of settings, from solving equations to graphing functions.

In this book, students will learn the basics of linear transformations and how to apply them in various situations. By the end of this book, students will be well-prepared to tackle any problem that involves linear transformations.

Linear transformations are a critical tool in mathematics, allowing us to understand how certain objects change under certain conditions. In particular, they help us to understand how vector spaces change when we apply a linear transformation to them.

A vector space is a set of vectors that can be added together and scaled by real numbers. Linear transformations preserve the structure of vector spaces, meaning that they preserve the concept of addition and scalar multiplication.

In other words, if we have a vector space V and a linear transformation T, then the transformed vector space $T(V)$ will also be a vector space. This property is what makes linear transformations so powerful: they allow us to study how objects change without losing any important information about them.

The range of the transformation may be the same as the domain, and when that happens, the transformation is known as an endomorphism or, if invertible, an automorphism.

Vector spaces and linear transformations are important in physics because they provide a way to describe how physical quantities change under the

influence of external forces.

For example, the motion of a particle in a straight line can be described as a vector in a vector space, and the force that causes the particle to accelerate can be described as a linear transformation of that vector. Similarly, the electric field in an insulating material can be described as a vector, and the charge that produces that field can be described as a linear transformation of that vector.

Linear transformations are one of the most important tools in mathematical geometry, and they can be used to describe a wide variety of geometric operations. The best-known linear transformation is the translation, which simply moves a figure from one point to another. But there are many other possibilities, including scaling (which changes the size of a figure), rotation (which turns a figure around a fixed point), and reflection (which flips a figure over a line or plane).

Linear transformations can be represented by matrices, which makes them easy to work with in computer programs. And because they preserve straight lines and angles, they are often very useful for solving problems in physics and engineering. In short, linear transformations are a powerful tool for understanding and manipulating the shapes of objects in our world.

Invertible linear transformations are a special type of linear transformation that has an inverse function. This means that it can be undone and that it is reversible. Invertible linear transformations are important in many fields, such as engineering and physics. They are used to model systems that can be reversed, such as springs and magnets. Invertible linear transformations are also used in computer graphics and image processing, where they are used to rotate and resize images.

Linear transformations are a critical tool in mathematics, used to abstract and study problems in a wide variety of fields. In this book, you'll learn about the kernel and image of a linear transformation.

First, let's recall the definition of a linear transformation. A linear transformation is a function that satisfies the following two properties:

- It is additive: for any two vectors x and y in its domain, the function produces a new vector that is the sum of x and y. That is, $T(x+y) = T(x) + T(y)$.
- It is homogeneous: for any vector x in its domain and scalar c, the function produces a new vector that is equal to cx. That is, $T(cx) = cT(x)$.

Now let's turn our attention to the kernel of a linear transformation. The kernel of a linear transformation T is the set of all vectors x in the domain of T such that $T(x) = 0$. In other words, it is the set of all vectors that are mapped to 0 by T. It's important to note that the kernel is always

a subspace of the domain of T. This means that it must be closed under addition and scalar multiplication.

The image of a linear transformation T is the set of all vectors y in the codomain of T such that there exists some vector x in the domain of T such that $T(x) = y$. In other words, it is the set of all vectors that can be reached from some vector in the domain by applying T. Like the kernel, the image is always a subspace of the codomain of T.

The kernel and image of a linear transformation encode important information about how that transformation behaves. By understanding these concepts, we can better analyze and work with linear transformations.

In mathematics, a coordinate matrix is a matrix whose elements are the coordinates of a given vector with respect to a given basis. In other words, if v is a vector in $\mathbb{R}^n$ and B is an $n \times n$ matrix, then the product Bv is another vector in $\mathbb{R}^n$ whose coordinates with respect to the standard basis are precisely the entries of the matrix B.

Coordinate matrices are used to represent linear transformations. Given a linear transformation $T : V \to W$ between two finite-dimensional vector spaces V and W, there is a unique coordinate matrix A such that for any vector $v \in V$, we have $T(v) = Av$.

In particular, this means that the columns of A are precisely the images of the vectors in some basis for V. Similarly, the rows of A are the coordinates of the vectors in some basis for W with respect to the standard basis. Thus, coordinate matrices can be used to change between different bases.

For instance, if B is a basis for V and C is a basis for W, then the matrix CB is a coordinate matrix for T with respect to the bases B and C. Conversely, every matrix can be viewed as a coordinate matrix; simply choose any bases for V and W and view the matrix as a transformation between these spaces.

In short, coordinate matrices are useful when working with linear transformations because they make it easy to change between different bases.

In this book, we've learned about linear transformations and their properties. We've seen how to represent them using matrices, and how to use these matrices to change between different bases. We've also learned about the kernel and image of a linear transformation, and how these concepts can be used to better understand the behavior of linear transformations. With this knowledge in hand, we're now ready to tackle more advanced topics in linear algebra.

3.1   Introduction to Linear Transformations

A linear transformation is a function of the form $\vec{y} = A\vec{x}$ where A is an $n \times m$ matrix. More specifically, a linear transformation is a function that assigns to each $\vec{x} \in \mathbb{R}^m$, a unique $\vec{y} \in \mathbb{R}^n$ – and this assignment is defined by a matrix A. When A is the identity matrix and $T(\vec{x}) = A\vec{x}$ we call T the **identity transformation** .

Definition 3.1. A function T from $\mathbb{R}^m$ to $\mathbb{R}^n$ is called a **linear trans-formation** if there exists an $n \times m$ matrix A such that $T(\vec{x}) = A\vec{x}$, for all $\vec{x}$ in the vector space $\vec{R}^m$.

Lemma 3.1. *Let T be a linear transformation from $\mathbb{R}^m$ to $\mathbb{R}^n$, then the matrix of T is*

$$A = \begin{bmatrix} | & & | \\ T(\vec{e}_1) & \cdots & T(\vec{e}_m) \\ | & & | \end{bmatrix} \tag{3.1}$$

where $\vec{e}_i$ (for $0 \leq i \leq m$) are the standard vectors.

Proof. Suppose T is a linear transformation from $\mathbb{R}^m$ to $\mathbb{R}^n$, then there exists an $n \times m$ matrix A such that $T(\vec{x}) = A\vec{x}$ for all $\vec{x} \in \mathbb{R}^m$. Let $\vec{e}_1, ..., \vec{e}_m$ be the standard vectors of $\mathbb{R}^m$ and let $A = [a_{ij}]$, then

$$T(\vec{e}_1) = A\vec{e}_1 = \begin{bmatrix} a_{11} & \cdots & a_{1m} \\ \vdots & \cdots & \vdots \\ a_{n1} & \cdots & a_{nm} \end{bmatrix} \begin{bmatrix} 1 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} a_{11} \\ \vdots \\ a_{n1} \end{bmatrix}$$

$$\vdots$$

$$T(\vec{e}_m) = A\vec{e}_m = \begin{bmatrix} a_{11} & \cdots & a_{1m} \\ \vdots & \cdots & \vdots \\ a_{n1} & \cdots & a_{nm} \end{bmatrix} \begin{bmatrix} 0 \\ \vdots \\ 1 \end{bmatrix} = \begin{bmatrix} a_{1m} \\ \vdots \\ a_{nm} \end{bmatrix}$$

which are the columns of the matrix A.

$\square$

Example 3.1. Determine the linear transformation T given by the system of linear equations:

$$\begin{aligned} y_1 &= 7x_1 + 3x_2 - 9x_3 + 8x_4 \\ y_2 &= 6x_1 + 2x_2 - 8x_3 + 7x_4 \\ y_3 &= 8x_1 + 4x_2 + 7x_4 \end{aligned}$$

The matrix of the linear transformation is $A = \begin{bmatrix} 7 & 3 & -9 & 8 \\ 6 & 2 & -8 & 7 \\ 8 & 4 & 0 & 7 \end{bmatrix}$ since

$$T(\vec{e}_1) = \begin{bmatrix} 7 \\ 6 \\ 8 \end{bmatrix}, \qquad T(\vec{e}_2) = \begin{bmatrix} 3 \\ 2 \\ 4 \end{bmatrix}, \qquad T(\vec{e}_3) = \begin{bmatrix} -9 \\ -8 \\ 0 \end{bmatrix}, \qquad T(\vec{e}_4) = \begin{bmatrix} 8 \\ 7 \\ 7 \end{bmatrix}.$$

Notice T is a linear transformation from $\mathbb{R}^4$ to $\mathbb{R}^3$ and A is a 3×4 matrix.

Example 3.2. Is the transformation $T(\vec{x}) = \vec{v} \cdot \vec{x}$ from $\mathbb{R}^3$ to $\mathbb{R}$ a linear transformation? If so, find the matrix of T. Let $\vec{v} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}$. Then

$$T(\vec{x}) = \vec{v} \cdot \vec{x} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = v_1 x_1 + v_2 x_2 + v_3 x_3 = \begin{bmatrix} v_1 & v_2 & v_3 \end{bmatrix} \vec{x}.$$

Therefore, by 3.1, T is a linear transformation with matrix

$$\begin{bmatrix} v_1 & v_2 & v_3 \end{bmatrix}.$$

Example 3.3. Is the transformation $T(\vec{x}) = \vec{v} \times \vec{x}$ from $\mathbb{R}^3$ to $\mathbb{R}$ a linear transformation? If so, find the matrix of T. Let $\vec{v} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}$. Then

$$T(\vec{x}) = \vec{v} \times \vec{x} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} \times \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} v_2 x_3 - v_3 x_2 \\ v_3 x_1 - v_1 x_3 \\ v_1 x_2 - v_2 x_1 \end{bmatrix} = \begin{bmatrix} 0 & -v_3 & v_2 \\ v_3 & 0 & -v_1 \\ -v_2 & v_1 & 0 \end{bmatrix} \vec{x}$$

Therefore, by 3.1, T is a linear transformation with matrix

$$\begin{bmatrix} 0 & -v_3 & v_2 \\ v_3 & 0 & -v_1 \\ -v_2 & v_1 & 0 \end{bmatrix}.$$

Theorem 3.1. *A function T from $\mathbb{R}^m$ to $\mathbb{R}^n$ is a linear transformation if and only if both of the following hold:*

- *$T(\vec{v} + \vec{w}) = T(\vec{v}) + T(\vec{w})$ for all vectors $\vec{v}$ and $\vec{w}$ in $\mathbb{R}^m$, and*
- *$T(k\vec{v}) = kT(\vec{v})$ for all vectors $\vec{v}$ in $\mathbb{R}^m$ and all scalars k.*

Proof. Suppose T is a linear transformation from $\mathbb{R}^m$ to $\mathbb{R}^n$, then there exists an $n \times m$ matrix A such that $T(\vec{x}) = A\vec{x}$ for all $\vec{x} \in \mathbb{R}^m$. The proof of each part follows.

- Let $\vec{u}, \vec{v} \in \mathbb{R}^m$, then $T(\vec{v}+\vec{w}) = A(\vec{v}+\vec{w}) = A\vec{v}+A\vec{w} = T(\vec{v})+T(\vec{w})$.
- Let $\vec{v} \in \mathbb{R}^m$ and $k \in \mathbb{R}$. Then $T(k\vec{v}) = A(k\vec{v}) = (Ak)\vec{v} = k(A\vec{v}) = kT(\vec{v})$.

Now suppose both (i) and (ii) hold. We need to find a matrix A such that $Tx = A\vec{x}$ for all $\vec{x} \in \mathbb{R}^m$. We can use the standard vectors in $\mathbb{R}^m$. Then by 3.1,

$$T(\vec{x}) = T(x\vec{e}_1 + \cdots + x_m\vec{e}_m) = T(x_1\vec{e}_1) + \cdots + T(x_m\vec{e}_m)$$

$$= x_1 T(\vec{e}_1) + \cdots + x_m T(\vec{e}_m) = \begin{bmatrix} | & & | \\ T(\vec{e}_1) & \cdots & T(\vec{e}_m) \\ | & & | \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_m \end{bmatrix} = A\vec{x}$$

as desired.

$\square$

Example 3.4. Write $\begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}$ as a linear combination of $\begin{bmatrix} 3 \\ -1 \\ 2 \end{bmatrix}$ and $\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$.

Let $T : \mathbb{R}^3 \to \mathbb{R}$ be a linear transformation with $T\begin{bmatrix} 3 \\ -1 \\ 2 \end{bmatrix} = 5$ and $T\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} =$

2. Find $T\begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}$. Notice

$$\begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} = (-1)\begin{bmatrix} 3 \\ -1 \\ 2 \end{bmatrix} + 2\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}.$$

Therefore,

$$T\begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} = T\left((-1)\begin{bmatrix} 3 \\ -1 \\ 2 \end{bmatrix} + 2\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \right) = (-1)T\begin{bmatrix} 3 \\ -1 \\ 2 \end{bmatrix} + 2T\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} = -1.$$

Example 3.5. Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be defined by $T\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 2x_1 \\ x_2^2 \end{bmatrix}$. Is T a linear transformation? Let $\alpha = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ and $\beta = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$. Then

$$T(\alpha + \beta) = T\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \right) = T\begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \end{bmatrix} = \begin{bmatrix} 2(x_1 + y_1) \\ (x_2 + y_2)^2 \end{bmatrix}$$

On the other hand

$$T(\alpha) + T(\beta) = T\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + T\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 2x_1 \\ x_2^2 \end{bmatrix} + \begin{bmatrix} 2y_1 \\ y_2^2 \end{bmatrix} = \begin{bmatrix} 2(x_1 + y_1) \\ x_2^2 + y_2^2 \end{bmatrix}$$

Since $T(\alpha + \beta) \neq T(\alpha) + T(\beta)$, we use 3.1 to conclude that T is not linear transformation.

Theorem 3.2. *Let T be a linear transformation from $\mathbb{R}^m$ to $\mathbb{R}^n$.*

- *If $\vec{0}_m$ is the zero vector in $\mathbb{R}^m$, then $T(\vec{0}_m)$ is the zero vector in $\mathbb{R}^n$.*
- *For all $\vec{v}$ in $\mathbb{R}^m$, $T(-\vec{v}) = -T(\vec{v})$.*
- *For all $\vec{u}, \vec{v}$ in $\mathbb{R}^m$, $T(\vec{u} - \vec{v}) = T(\vec{u}) - T(\vec{v})$.*
- *For all $a_1, ..., a_n \in \mathbb{R}$ and for all $\vec{v}_1, ..., \vec{v}_n \in \mathbb{R}^m$,*

$$T(a_1\vec{v}_1 + a_2\vec{v}_2 + \cdots + a_n\vec{v}_n) = a_1T(\vec{v}_1) + a_2T(\vec{v}_2) + \cdots + a_nT(\vec{v}_n).$$

Proof. There proof is for the reader.

$\square$

3.2   Linear Transformations in Geometry

Next we give several examples of linear transformations from $\mathbb{R}^2$ to $\mathbb{R}^2$ that are commonly used in plane geometry.

Theorem 3.3. *Let T be a linear transformation from $\mathbb{R}^2$ to $\mathbb{R}^2$. If the matrix of T is of the form*

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

*then T is a **scaling** transformation. If $k > 1$ then the scaling is called a **dilation**, and is called a **contraction** when $k < 1$.*

Proof. The proof is left for the reader.

$\square$

Example 3.6. Is the linear transformation given by the system of linear equations

$$\begin{cases} y_1 = 7x_1 \\ y_2 = 7x_2 \end{cases}$$

from $\mathbb{R}^2$ to $\mathbb{R}^2$ a scaling? The answer is yes since the matrix of the linear transformation is $\begin{bmatrix} 7 & 0 \\ 0 & 7 \end{bmatrix}$, which by definition is a scaling. For example, we can write

$$T(\vec{x}) = \begin{bmatrix} 7 & 0 \\ 0 & 7 \end{bmatrix} \vec{x}.$$

We can use T to dilate the vector $\begin{bmatrix} 1 \\ 2 \end{bmatrix}$ by 7 to obtain

$$T\begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 7 & 0 \\ 0 & 7 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 7 \\ 14 \end{bmatrix}.$$

Theorem 3.4. *Let T be a linear transformation from $\mathbb{R}^2$ to $\mathbb{R}^2$. If the matrix of T is of the form*

$$\frac{1}{w_1^2 + w_2^2} \begin{bmatrix} w_1^2 & w_1 w_2 \\ w_1 w_2 & w_2^2 \end{bmatrix}$$

*then T is an **orthogonal projection** transformation onto the line L spanned by any nonzero vector $\vec{w} = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$ parallel to L.*

Proof. Suppose line L is spanned by $\vec{w}$. We can decompose any vector $\vec{x}$ as $\vec{x}^{\parallel} + \vec{x}^{\perp}$ as diagrammed: Notice $\vec{x}^{\perp}$ is the perpendicular component so

$$\vec{w} \cdot \vec{x}^{\perp} = 0 \qquad \text{or equivalently} \qquad \vec{w} \cdot (\vec{x} - \vec{x}^{\parallel}) = 0. \qquad (3.2)$$

To project $\vec{x}$ onto the line L we notice

$$\vec{x}^{\parallel} = k\vec{w} \qquad\qquad (3.3)$$

for some scalar k. By substitution, of 3.3 into 3.2 and solving for k we obtain

$$k = \frac{\vec{w} \cdot \vec{x}}{\vec{w} \cdot \vec{w}}. \qquad\qquad (3.4)$$

Using 3.3 and 3.4 we define the orthogonal projection of a vector $\vec{x}$ onto a given line L as

$$\text{proj}_L(\vec{x}) = \frac{\vec{w} \cdot \vec{x}}{||\vec{w}||^2} \vec{w}. \qquad\qquad (3.5)$$

We would like to have 3.5 in the form of a matrix. To do so let $\vec{w} = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$

and $\vec{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$. Then from 3.4 and 3.5 we find

$$\text{proj}_L(\vec{x}) = \frac{1}{||\vec{w}||^2} \left((x_1 w_1 + x_2 w_2) \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} \right)$$

$$= \frac{1}{||\vec{w}||^2} \left((x_1 w_1 \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} + x_2 w_2 \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} \right)$$

$$= \frac{1}{||\vec{w}||^2} \left(\begin{bmatrix} x_1 w_1^2 \\ x_1 w_1 w_2 \end{bmatrix} + \begin{bmatrix} x_2 w_1 w_2 \\ x_2 w_2^2 \end{bmatrix} \right)$$

$$= \frac{1}{||\vec{w}||^2} \begin{bmatrix} x_1 w_1^2 + x_2 w_1 w_2 \\ x_1 w_1 w_2 + x_2 w_2^2 \end{bmatrix} = \frac{1}{||\vec{w}||^2} \begin{bmatrix} w_1^2 & w_1 w_2 \\ w_1 w_2 & w_2^2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}.$$

$\square$

Example 3.7. Find the matrix A of the orthogonal projection onto the line L spanned by $\vec{w} = \begin{bmatrix} 4 \\ 3 \end{bmatrix}$ and project the vector $\vec{u} = \begin{bmatrix} 1 \\ 5 \end{bmatrix}$ onto the line L spanned by $\vec{w}$. By 3.4, the matrix is

$$A = \frac{1}{w_1^2 + w_2^2} \begin{bmatrix} w_1^2 & w_1 w_2 \\ w_1 w_2 & w_2^2 \end{bmatrix} = \frac{1}{25} \begin{bmatrix} 16 & 12 \\ 12 & 9 \end{bmatrix} = \begin{bmatrix} 16/25 & 12/25 \\ 12/25 & 9/25 \end{bmatrix}.$$

For example, we can project the vector $\vec{u}$ onto the line L spanned by $\vec{w}$. The matrix A defines this linear transformation T and so to project onto the line L is just matrix multiplication:

$$T \begin{bmatrix} 1 \\ 5 \end{bmatrix} = \frac{1}{25} \begin{bmatrix} 16 & 12 \\ 12 & 9 \end{bmatrix} \begin{bmatrix} 1 \\ 5 \end{bmatrix} = \begin{bmatrix} 76/25 \\ 57/25 \end{bmatrix}.$$

Theorem 3.5. *Let T be a linear transformation from $\mathbb{R}^2$ to $\mathbb{R}^2$. If the matrix of T is of the form*

$$\begin{bmatrix} 2u_1^2 - 1 & 2u_1 u_2 \\ 2u_1 u_2 & 2u_2^2 - 1 \end{bmatrix}$$

*then T defines a **reflection** transformation about the line L, where $\vec{u} = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$ is any unit vector lying on L.*

Proof. Suppose we want to reflect $\vec{x}$ through the line L as diagrammed: Then

$$\text{ref}_L(\vec{x}) = \vec{x}^{||} - \vec{x}^{\perp} \tag{3.6}$$

and

$$\text{proj}_L(\vec{x}) = \vec{x}^{\parallel}. \tag{3.7}$$

By subtracting 3.7 from 3.6, we obtain

$$\text{ref}_L(\vec{x}) = 2\text{proj}_L(\vec{x}) - \vec{x}. \tag{3.8}$$

For simplicity assume L is a any line that passes through the origin and let $\vec{u}$ be a unit vector $\vec{u} = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$ lying on L. Notice 3.5 in the special case of a unit vector $\vec{u}$ becomes $\text{proj}_L(\vec{x}) = (\vec{u} \cdot \vec{x})\vec{u}$. Then

$$\text{ref}_L(\vec{x}) = 2\text{proj}_L(\vec{x}) - \vec{x} = 2(\vec{u} \cdot \vec{x})\vec{u} - \vec{x}$$

$$= 2(u_1 x_1 + u_2 x_2) \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} - \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 2u_1 x_1 \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} + 2u_2 x_2 \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} - \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$= \begin{bmatrix} 2u_1^2 x_1 + 2u_1 u_2 x_2 - x_1 \\ 2u_1 u_2 x_1 + 2u_2^2 x_2 - x_2 \end{bmatrix} = \begin{bmatrix} 2u_1^2 - 1 & 2u_1 u_2 \\ 2u_1 u_2 & 2u_2^2 - 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}.$$

$\square$

Example 3.8. Find the matrix A of a reflection through the line through the origin spanned by $\vec{w} = \begin{bmatrix} 4 \\ 3 \end{bmatrix}$ and use it to reflect $\begin{bmatrix} 1 \\ 5 \end{bmatrix}$ about the line L. Let $\vec{u} = \begin{bmatrix} 4/5 \\ 3/5 \end{bmatrix}$. We notice $\vec{u}$ is a unit vector, since $||u|| = 1$. Then, by 3.5, the matrix we seek is

$$A = \begin{bmatrix} 7/25 & 24/25 \\ 24/25 & -7/25 \end{bmatrix}.$$

We can reflect the vector $\begin{bmatrix} 1 \\ 5 \end{bmatrix}$ about the line L using matrix multiplication

$$T\begin{bmatrix} 1 \\ 5 \end{bmatrix} = \frac{1}{25} \begin{bmatrix} 7 & 24 \\ 24 & -7 \end{bmatrix} \begin{bmatrix} 1 \\ 5 \end{bmatrix} = \begin{bmatrix} 127/25 \\ -11/25 \end{bmatrix}.$$

Theorem 3.6. *Let T be a linear transformation from $\mathbb{R}^2$ to $\mathbb{R}^2$. If the matrix of T is of the form*

$$\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

*then T is a (counterclockwise) **rotation** transformation through an angle θ.*

Proof. If a vector $\vec{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ is rotated through an angle of $\pi/2$, then a vector $\vec{y} = \begin{bmatrix} -x_2 \\ x_1 \end{bmatrix}$ is obtained, via $\vec{x} \cdot \vec{y} = \vec{0}$. More generally, if we rotate (counterclockwise) a given $\vec{x}$ through an angle θ we determine

$$T(\vec{x}) = (\cos\theta)\vec{x} + (\sin\theta)\vec{y} = (\cos\theta)\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + (\sin\theta)\begin{bmatrix} -x_2 \\ x_1 \end{bmatrix}$$

$$= \begin{bmatrix} (\cos\theta)x_1 - (\sin\theta)x_2 \\ (\sin\theta)x_1 - (\cos\theta)x_2 \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \vec{x}$$

as seen from the diagram.

□

Example 3.9. Find the matrix of the linear transformation that rotates the vector $\begin{bmatrix} 4 \\ 2 \end{bmatrix}$ by 30 degrees counterclockwise. By 3.6, the matrix of this transformation is $A = \begin{bmatrix} \sqrt{3}/2 & -1/2 \\ 1/2 & \sqrt{3}/2 \end{bmatrix}$. We will use matrix multiplication to perform the transformation

$$T\begin{bmatrix} 4 \\ 2 \end{bmatrix} = \begin{bmatrix} \frac{\sqrt{3}}{2} & \frac{-1}{2} \\ \frac{1}{2} & \frac{\sqrt{3}}{2} \end{bmatrix} \begin{bmatrix} 4 \\ 2 \end{bmatrix} = \cos 30° \begin{bmatrix} 4 \\ 2 \end{bmatrix} + \sin 30° \begin{bmatrix} 4 \\ 2 \end{bmatrix} = \begin{bmatrix} 2(\sqrt{3}+1) \\ \sqrt{3}+1 \end{bmatrix} \approx \begin{bmatrix} 2.46 \\ 3.73 \end{bmatrix}$$

Theorem 3.7. *Let T be a linear transformation from $\mathbb{R}^2$ to $\mathbb{R}^2$. If the matrix of T is of the form*

$$\begin{bmatrix} 1 & 0 \\ k & 1 \end{bmatrix} \quad or \quad \begin{bmatrix} 1 & k \\ 0 & 1 \end{bmatrix},$$

*where k is any constant, then T defines a **vertical shear** or **horizontal shear** transformation, respectively.*

Proof. The proof is left for the reader.

□

Example 3.10. Given the vector $\begin{bmatrix} 2 \\ 3 \end{bmatrix}$ in $\mathbb{R}^2$ show geometrically a vertical shear of 2 and a horizontal shear of $\frac{1}{2}$. By 3.7, we can apply these linear

transformations using matrix multiplication by using the matrices $\begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix}$ and $\begin{bmatrix} 1 & 1/2 \\ 0 & 1 \end{bmatrix}$.

$$\text{Vertical Shear:} \quad T\begin{bmatrix} 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix}\begin{bmatrix} 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 2 \\ 7 \end{bmatrix}$$

$$\text{Horizontal Shear:} \quad T\begin{bmatrix} 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 1 & 1/2 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 7/2 \\ 3 \end{bmatrix}$$

Example 3.11. Interpret the linear transformation $T(\vec{x}) = \begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix}\vec{x}$ geometrically. The transformation has a matrix of the form

$$\begin{bmatrix} 2u_1^2 - 1 & 2u_1 u_2 \\ 2u_1 u_2 & 2u_2^2 - 1 \end{bmatrix}$$

where $u_1 = \sqrt{2}/2$ and $u_2 = -\sqrt{2}/2$ since $2u_1^2 - 1 = 0$, $2u_2^2 - 1 = 0$, and $2u_1 u_2 = -1$. Since $||\vec{u}|| = 1$ and $\vec{u}$ lies on the line $y = x$, then matrix $\begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix}$ represents the linear transformation which is a reflection through the line $y = x$.

scaling $\begin{bmatrix} k & 0 \\ 0 & k \end{bmatrix}$ shears $\begin{bmatrix} 1 & 0 \\ k & 1 \end{bmatrix}$ or $\begin{bmatrix} 1 & k \\ 0 & 1 \end{bmatrix}$

reflection $\begin{bmatrix} 2u_1^2 - 1 & 2u_1 u_2 \\ 2u_1 u_2 & 2u_2^2 - 1 \end{bmatrix}$ rotation $\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$

orthogonal projection $\dfrac{1}{w_1^2 + w_2^2}\begin{bmatrix} w_1^2 & w_1 w_2 \\ w_1 w_2 & w_2^2 \end{bmatrix}$

Example 3.12. Interpret the linear transformation $T(\vec{x}) = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}\vec{x}$, geometrically. Explain. This is a rotation combined with a scaling. The transformation rotates 45 degrees counterclockwise and has a scaling factor of $\sqrt{2}$.

3.3   Introduction to Linear Maps

Definition 3.2. Let V and W be linear spaces. A function T from V to W is called a **linear map** if

$$T(f + g) = T(f) + T(g) \quad \text{and} \quad T(kf) = kT(f)$$

for all elements f and g of V and for all scalars k.

The collection of all linear maps from V to W is denoted by $\mathcal{L}(V, W)$ and if T is a linear map from V to W we denote this by $T \in \mathcal{L}(V, W)$.

Definition 3.3. Let $T \in \mathcal{L}(V, W)$.

- Then the **kernel** of T is the subset of V consisting of the vectors that T maps to 0; that is
  $\ker(T) = \{v \in V \mid Tv = 0\}$.
- Then the **image** of T is the subset of W consisting of the vectors of the form Tv for some $v \in V$; that is $\operatorname{im}(T) = \{w \in W \mid \text{there exists } v \in V \text{ such that } w = Tv\}$.

Example 3.13. The function from C^∞ to C^∞ defined by $T(f) = f''$ is a linear map. Find the kernel and image of T.

Example 3.14. The function from C^∞ to C^∞ defined by $T(f) = \int_0^1 f(x)\, dx$ is a linear map. Find the kernel and image of T.

Theorem 3.8. *If $T \in \mathcal{L}(V, W)$, then null T is a subspace of V.*

Proof. By definition, $\ker T = \{v \in V \mid Tv = 0\}$ and so $\ker T \subseteq W$. Let $u, v \in \ker T$. Then, $T(u + v) = Tu + Tv = 0 + 0 = 0$ which shows, $u + v \in \ker T$, for all $u, v \in \ker T$. Let k be a scalar and $u \in \ker T$. Then $T(ku) = kTu = k(0) = 0$, which shows, $ku \in \ker T$ for all scalars k and all $u \in \ker T$. Since $T(0) = T(0 + 0) = T(0) + T(0)$, $T(0) = 0$ in W which shows, $0 \in \ker T$. Therefore, $\ker T$ is a subspace of W.

$\square$

Definition 3.4. A linear map $T : V \to W$ is called **injective** whenever $u, v \in V$ and $Tu = Tv$, we have $u = v$.

Theorem 3.9. *Let $T \in \mathcal{L}(V, W)$, then T is injective if and only if null $T = \{0\}$.*

Proof. Suppose T is injective. Since $T(0) = 0$, $0 \in \ker T$ and so $\{0\} \subseteq \ker T$. Let $v \in \ker T$. Then $T(0) = 0 = T(v)$ yields $v = 0$ because T is injective. Thus, $\ker T \subseteq \{0\}$. Therefore, $\ker T = \{0\}$. Conversely, assume

$\ker T = \{0\}$. Let $u, v \in V$. If $Tu = Tv$, then $T(u - v) = Tu - Tv = 0$ which shows $u - v \in \ker T$. Thus $u = v$ and therefore, T is injective.

$\square$

Definition 3.5. For $T \in \mathcal{L}(V, W)$, the **image** of T, denoted by $\mathrm{im}(T)$, is the subset of W consisting of those vectors that are of the form Tv for some $v \in V$.

Definition 3.6. A linear map $T : V \to W$ is called **surjective** if its range equals W.

Theorem 3.10. *If $T \in \mathcal{L}(V, W)$, then $\mathrm{im}T$ is a subspace of W.*

Proof. By definition, $\mathrm{im}T = \{Tv \mid v \in V\} \subseteq W$. Let $w_1, w_2 \in W \in \mathrm{im}T$. Then there exists $v_1, v_2 \in V$ such that $w_1 = Tv_1$ and $w_2 = Tv_2$. By linearity of T, $w_1 + w_2 = Tv_1 + Tv_2 = T(v_1 + v_2)$ which shows $w_1 + w_2 \in \mathrm{im}T$ for all $w_1, w_2 \in \mathrm{im}T$. Let $w \in \mathrm{im}T$ and let k be a scalar. Then there exists $v \in V$ such that $w = Tv$. By linearity of T, $kw = k\,Tv = T(k\,v)$ which shows $kw \in \mathrm{im}T$ for all $w \in \mathrm{im}T$ and for all scalars k. Therefore, $\mathrm{im}T$ is a subspace of W.

$\square$

::: {#thm- } [Rank-Nullity Theorem] If V is finite-dimensional and $T \in \mathcal{L}(V, W)$, then range T is finite-dimensional and $\dim V = \dim(\mathrm{im}T) + \dim(\ker T)$. :::

Proof. Since V is a finite-dimensional and $\ker T$ is a subspace of V, $\ker T$ is finite-dimensional and so let $(u_1, \ldots, u_n)$ be a basis of $\ker T$. Since $(u_1, \ldots, u_n)$ is linearly independent in V, it can be extended to a basis of V, say $\mathcal{B} = (u_1, \ldots, u_n, v_1, \ldots, v_m)$. It suffices to show $(Tv_1, \ldots, Tv_m)$ is a basis of $\mathrm{im}T$, for then $\mathrm{im}T$ is finite-dimensional is proven and $\dim V = n + m = \dim \ker T + \dim \mathrm{im}T$ holds as well.

Let $w \in \mathrm{im}T$. Then there exists $v \in V$ and scalars $a_1, \ldots, a_n, b_1, \ldots, b_m$ such that

$$w = Tv = T(a_1 u_1 + \cdots + a_n u_n + b_1 v_1 + \cdots + b_m v_m) = b_1 Tv_1 + \cdots + b_m Tv_m.$$

Thus $(Tv_1, \ldots, Tv_m)$ spans $\mathrm{im}T$. Suppose $c_1, \ldots, c_m$ are scalars such that $cTv_1 + \cdots + c_m Tv_m = 0$. Then there exist scalars $d_1, \ldots, d_n$ such that

$$c_1 v_1 + \cdots + c_m v_m = d_1 u_1 + \cdots + d_n u_n.$$

Since $\mathcal{B}$ is a basis of V and $c_1 v_1 + \cdots c_m v_m + (-d_1)u_1 + \cdots + (-d_n)u_n = 0$ it follows $c_1 = \cdots = c_m = d_1 = \cdots = \cdots = d_n = 0$. In particular, $c_1 = \cdots = c_m = 0$ shows $Tv_1, \ldots, Tv_m$ are linearly independent. Therefore, $(Tv_1, \ldots, Tv_m)$ is a basis fo im$T$.

$\square$

Corollary 3.1. *If V and W are finite-dimensional vector spaces with $T \in \mathcal{L}(V, W)$, then*

- *$dimV > dimW \implies T$ is not injective, and*
- *$dimV < dimW \implies T$ is not surjective.*

Proof. The proof of each part follows.

- By the Rank-Nullity Theorem, $\dim V = \dim\ker T + \dim\text{im}T \leq \dim\ker T + \dim W$. Since $\dim V > \dim W$, it follows $0 < \dim V - \dim W \leq \dim\ker T$. Thus, $\ker T \neq \{0\}$ and so T is not injective.
- Again by the Rank-Nullity Theorem, $\dim V = \dim\ker T + \dim\text{im}T$ and so $\dim\text{im}T \leq \dim V$. Since $\dim V < \dim W$, $0 < \dim W - \dim V \leq \dim W - \dim\text{im}T$ and so there exists a nonzero element $w \in W$ with $w \in \text{im}T$. Therefore, T is not surjective.

$\square$

Example 3.15. Suppose that T is a linear map from V to $\mathbb{F}$. Prove that if $u \in V$ is not in $\ker T$, then $V = \ker T \oplus \{au : a \in F\}$. Let $U = \{au \mid a \in \mathbb{F}\}$. The following arguments show $V = \ker T + U$ and $\ker T \cap U = \{0\}$, respectively.

- Let $v \in V$ with $Tv = b$. Since $Tu \neq 0$ there is a $u_1 \in U$ such that $Tu_1 = b$. Then we can write $v = u_1 + (v - u_1)$, and $v - u_1 \in \ker T$. This gives $V = \ker T + U$.
- Let $v \in \ker T \cap U$, then there exists $a \in \mathbb{F}$ such that $v = au$ and so $T(v) = aTu = 0$ since $Tv \in \ker T$. Thus $a = 0$ and so $v = 0$ meaning $\ker T \cap U \subseteq \{0\}$. Since $0 \in \ker T \cap U$, it follows $\ker T \cap U = \{0\}$.

Example 3.16. Suppose that $T \in \mathcal{L}(U, W)$ is injective and $(v_1, \ldots, v_n)$ is linearly independent in V. Prove that $(Tv_1, \ldots, Tv_n)$ is linearly independent in W. Suppose $a_1 Tv_1 + \cdots a_n Tv_n = 0$ in W where $a_1, \ldots, a_n \in \mathbb{F}$. Then by linearity of T, $T(a_1 v_1 + \cdots + a_n v_n) = 0$. Since $T(0)$ and T is injective, $a_1 v_1 + \cdots + a_n v_n = 0$. Since $(v_1, \ldots, v_n)$ is linearly independent $a_1 = \cdots = a_n = 0$. Therefore, $(Tv_1, \ldots, Tv_n)$ is linearly independent.

Example 3.17. Show that every linear map from a one-dimensional vector space to itself is multiplication by some scalar. More precisely, prove that if $\dim V = 1$ and $T \in \mathcal{L}(V, V)$, then there exists $a \in \mathbb{F}$ such that $Tv = av$ for all $v \in V$. Let $\{w\}$ be a basis of V, and let $v \in V$. Then there exists $c \in \mathbb{F}$ such that $v = cw$. Applying T yields,

$$Tv = T(cw) = cTw = c(aw) = (ca)w = a(cw) = av$$

where $Tw = aw$ since $Tw \in V$ and $\{w\}$ is a basis of V.

Example 3.18. linear extension Suppose that V is finite-dimensional. Prove that any linear map on a subspace of V can be extended to a linear map on V. In other words, show that if U is a subspace of V and $S \in \mathcal{L}(U, W)$, then there exists $T \in \mathcal{L}(V, W)$ such that $Tu = Su$ for all $u \in U$. Let $(u_1, \dots, u_n)$ be a basis of U and extend this basis of U to a basis of V, say $(u_1, \dots, u_n, v_1, \dots, v_m)$. Define T as the linear extension of S, as follows

$$T(u_i) = S(u_i) \text{ for } 1 \le i \le n \qquad \text{and} \qquad T(v_j) = v_j \text{ for } 1 \le j \le m.$$

Then for all $u \in U$,

$$T(u) = T(a_1 u_1 + \dots + a_n u_n) = a_1 T u_1 + \dots a_n T u_n$$
$$= a_1 S u_1 + \dots + a_n S u_n = S(a_1 u_1 + \dots a_n u_n) = S(u)$$

where $a_1, \dots, a_n \in \mathbb{F}$. By definition of T, $T \in \mathcal{L}(V, W)$.

Example 3.19. Prove that if $S_1, \dots, S_n$ are injective linear maps such that $S_1 \cdots S_n$ makes sense, then $S_1 \cdots S_n$ is injective. Suppose v and w are any vectors and $(S_1 \cdots S_n)v = (S_1 \cdots S_n)w$. Then by definition of composition, $S_1(S_2 \cdots S_n)v = S_1(S_2 \cdots S_n)w$, and since S_1 is injective $S_2(S_3 \cdots S_n)v = S_2(S_3 \cdots S_n)w$. Since $S_2, \dots, S_n$ are all injective $v = w$ as desired showing $S_1 \cdots S_n$ is injective.

Example 3.20. Prove that if $(v_1, \dots, v_n)$ spans V and $T \in \mathcal{L}(V, W)$ is surjective, then $(Tv_1, \dots, Tv_n)$ spans W. Let $w \in W$. Then, since T is surjective, there exists $v \in V$ such that $T(v) = w$. Since $(v_1, \dots, v_n)$ spans V there exists scalars $a_1, \dots, a_n$ such that $v = a_1 v_1 + \dots a_n v_n$ and by linearity of T, $T(v) = a_1 T v_1 + \dots a_n T v_n = w$. Therefore, $(Tv_1, \dots, Tv_n)$ spans W because every element w in W is a linear combination of the Tv's.

Example 3.21. Suppose that V is finite-dimensional and that $T \in \mathcal{L}(V, W)$. Prove that there exists a subspace U of V such that $U \cap \ker T = \{0\}$ and range $T = \{Tu : u \in U\}$. Since V is finite-dimensional so is $\ker T$. Let $(v_1, \dots, v_n)$ be a basis of $\ker T$ and extend this basis to a basis of V, namely let $\mathcal{B} = (v_1, \dots, v_n, u_1, \dots, u_m)$ be a basis of V. Let $U = \text{span}(u_1, \dots, u_m)$, we will show $U \cap \ker T = \{0\}$ and range $T = \{Tu : u \in U\}$. Clearly, $\{0\} \subseteq U \cap \ker T$ since both U and $\ker T$ are subspaces. Let $v \in U \cap \ker T$. Then $v \in U$ implies there exists $a_1, \dots, a_m$ such that $v = a_1 u_1 + \cdots + a_m u_m$. Then $v \in \ker T$ implies there exists $b_1, \dots, b_n$ such that $v = b_1 v_1 + \cdots + b_n v_n$. Then $a_1 u_1 + \cdots + a_m u_m + (-b_1) v_1 + \cdots + (-b_n) v_n = 0$ and since $\mathcal{B}$ is a basis of V the u's and v's are linearly independent and so $a_1 = \cdots = a_m = b_1 = \cdots = b_n = 0$ meaning $v = 0$; and so $\ker T = \{0\}$. Let $v \in V$. Then $T(v) = a_1 T v_1 + \cdots + a_n T v_n + b_1 T u_1 + \cdots + b_m T u_m = b_1 T u_1 + \cdots + b_m T u_m$ showing $\{Tv : v \in T\} \subseteq \{Tv : v \in U\}$ conversely is obvious, since $U \subseteq V$.

Example 3.22. Prove that if there exists a linear map on V whose null space and range are both finite-dimensional, then V is finite-dimensional.

Example 3.23. Suppose that V and W are finite-dimensional and $T \in \mathcal{L}(V, W)$. Prove that there exists a surjective linear map from V onto W if and only if $\dim W \leq \dim V$. Suppose T is a linear map of V onto W, then $\dim V = \dim \ker T + \dim \text{im} T$. Since $\dim \ker T \geq 0$, $\dim V \geq \dim \text{im} T = \dim W$, since T is onto. Conversely, assume $m = \dim W \leq \dim V = n$ with bases of W say $(w_1, \dots, w_m)$ and of V say $(v_1, \dots, v_n)$. Define T to be the linear extension of:

$$\begin{cases} T(v_i) = w_i & \text{if } 1 \leq i \leq m \\ T(v_i) = 0 & \text{if } i > m \end{cases}$$

Then T is surjective since: if $w \in W$ then there exists $a_i \in \mathbb{F}$ such that $w = a_1 w_1 + \cdots + a_m w_m = a_1 T(v_1) + \cdots + a_m T(v_m) = T(a_1 v_1 + \cdots + a_m v_m)$ showing every element in W is in $\text{im} T$.

Example 3.24. Suppose that W is finite-dimensional and $T \in \mathcal{L}(V, W)$. Prove that T is injective if and only if there exists $S \in \mathcal{L}(W, V)$ such that ST is the identity map on V. Suppose $S \in \mathcal{L}(W, V)$ and ST is the identity on V. Then $\ker T \subseteq \ker ST = \{0\}$ and so $\ker T = \{0\}$. Therefore, T is injective. Conversely, suppose T is injective. So we can define $S_1 \in \mathcal{L}(W, V)$ by the following: $S_1(w) = v$ where $v = T^{-1}(w)$ (since T is injective). So $S_1 : \text{im} T \to V$ and since W is finite-dimensional, by **??**, S_1 can be extended to $S : W \to V$ and by definition of S_1, if $v \in V$, then $ST(v) = S(Tv) = T^{-1}(Tv) = v$.

Example 3.25. Suppose that V is finite-dimensional and $T \in \mathcal{L}(V, W)$. Prove that T is surjective if and only if there exists $S \in \mathcal{L}(W, V)$ such that TS is the identity map on W. We will present two proofs.

- Suppose $S \in \mathcal{L}(W, V)$ and $TS = I_W$. Let $w \in W$. Then $v = S(w) \in V$ is such that $T(v) = TS(w) = w$ and therefore T is surjective. Conversely, suppose T is surjective. Since V is a finite-dimensional vector space and T is a linear map, W is finite-dimensional. Let $(v_1, \dots, v_n)$ be a basis for V. Since T is surjective, $(Tv_1, \dots, Tv_n)$ spans W. Also since T is surjective $\dim W \leq \dim V = n$ Any spanning set reduced to a basis say $(Tv_1, \dots, Tv_m)$ is a basis of W where $m \leq n$. Define S as the linear extension of $S(Tv_i) = v_i$ for each $1 \leq i \leq m$. Then, for all $w \in W$, $S(w) = S(a_1 Tv_1 + \dots + a_m Tv_m) = a_1 STv_1 + \dots + a_m STv_m = a_1 v_1 + \dots + a_m v_m = w$ for scalars $a_1, \dots, a_m \in \mathbb{F}$, and so $TS = I_W$.
- Suppose T is surjective. Using **??**, there exists a subspace U of V such that $U \cap \ker T = \{0\}$ and $\mathrm{im} T = \{Tu : u \in U\}$. Define $T_1 : U \to W$ by $T_1 u = Tu$ for $u \in U$. Notice T_1 is injective and surjective and so T_1 has an inverse. Define $S = T_1^{-1}$ we have $TSw = T_1 T_1^{-1} w = w$ for all $w \in W$.

A linear map $T \in \mathcal{L}(V, W)$ is called **invertible** if there exists a linear map $S \in \mathcal{L}(W, V)$ such that ST equals the identity map on V and TS equal the identity map on W. Given $T \in \mathcal{L}(V, W)$. A linear map $S \in \mathcal{L}(W, V)$ satisfying $ST = I$ and $TS = I$ is called an **inverse** of T. Two vector spaces are called **isomorphic** if there is an invertible linear map from one vector space onto the other one.

Theorem 3.11. *If $\mathcal{B} = (f_1, \dots, f_n)$ is a basis of a linear space V, then the coordinate transformation*

$$L_{\mathcal{B}}(f) = [f]_{\mathcal{B}}$$

from V to $\mathbb{R}^n$ is an isomorphism. Thus any linear space V is isomorphic to $\mathbb{R}^n$.

Proof. The proof if left for the reader.

$\square$

An invertible linear transformation T is called an **isomorphism**. Recall the rank-nullity theorem states that if V and W are finite-dimensional vector spaces and T is a linear transformation from V to W then $\dim V = \dim \ker T + \dim \mathrm{im} T$ where $\dim \ker T$ is called the **nullity** and $\dim \mathrm{im} T$ is called the **rank**.

Theorem 3.12. *If V and W are finite-dimensional linear spaces, then*

- *T is an isomorphism from V to W if and only if $\ker(T) = \{0\}$ and $im(T) = W$,*
- *if V is isomorphic to W, then $dim(V) = dim(W)$,*
- *if $dim(V) = dim(W)$ and $\ker(T) = \{0\}$, then T is an isomorphism, and*
- *if $dim(V) = dim(W)$ and $im(T) = W$, then T is an isomorphism.*

Proof. The proof of each part follows.

- Suppose T is an isomorphism from V to W. This means there exists an invertible linear transformation T^{-1} from W to V such that $T(T^{-1}) = I_W$ and $T^{-1}(T) = I_V$, where I_W and I_V are the identity transformations on W and V respectively. We will use T^{-1} to show $\ker T = \{0\}$ and $imT = W$. Since $\ker T$ is a subspace, $\{0\} \subset \ker T$. Conversely, suppose $f \in \ker T$. Then $T(f) = 0$ and so $T^{-1}(T(f)) = f = 0$ so that $\ker T \subseteq \{0\}$. Thus $\ker T = \{0\}$. Since imT is a subspace of W, $imT \subseteq W$. To show conversely, let $w \in W$. Then $T^{-1}(w) \in V$ and so $T(T^{-1}w) = w$ shows $w \in imT$. Thus $imT = W$.

Conversely, suppose $\ker T = \{0\}$ and $imT = W$. We will show T is an isomorphism, which means we must show $T(f) = g$ has a unique solution f for each $g \in W$. Now since $imT = W$, $T(f) = g$ has at least one solution, say f. Suppose $T(f) = g$ has two solutions, say $T(f_1) = g = T(f_2)$. Then $0 = T(f_1) - T(f_2) = T(f_1 - f_2)$ shows $f_1 - f_2 \in \ker T$. Since $\ker T = \{0\}$, $f_1 = f_2$ must follow, which means $T(f) = g$ must have only one solution, so that T^{-1} exists, and therefore, T is an isomorphism.
- If V is isomorphic to W, then there exists a linear transformation T such that $\ker T = \{0\}$ and $imT = W$. By the rank-nullity theorem, $\dim V = \dim \ker T + \dim imT = 0 + \dim W$. - By the previous part, T is an isomorphism when $\ker = \{0\}$ and $imT = W$. We are assuming $\ker T = \{0\}$ so it remains to show $imT = W$. By the rank-nullity theorem $\dim V = \dim \ker T + \dim imT = \dim \ker T + \dim W = \dim W$. Thus, $\ker T = 0$ and so $\ker T = \{0\}$. Therefore, T is an isomorphism. - By the previous part, T is an isomorphism when $\ker = \{0\}$ and $imT = W$. We are assuming $imT = W$ so it remains to show $\ker T = \{o\}$. By the rank-nullity theorem $\dim V = \dim \ker T + \dim imT = \dim imT = \dim W$. Therefore, $imT = W$ and so T is an isomorphism.

$\square$

Example 3.26. Show that the transformation $T(A) = S^{-1}AS$, from $\mathbb{R}^{2\times 2}$ to $\mathbb{R}^{2\times 2}$ where $S = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$ is an isomorphism. Notice that $\dim\mathbb{R}^{2\times 2} = \dim\mathbb{R}^{2\times 2}$. Can we determine an invertible linear transformation T^{-1}? Checking the linearity conditions,

$$T(A_1 + A_2) = S^{-1}(A_1 + A_2)S = S^{-1}A_1 S + S^{-1}A_2 S = T(A_1) + T(A_2)$$

and

$$T(kA) = S^{-1}(kA)S = kS^{-1}AS = kT(A).$$

Now we know T is a linear transformation. Is it invertible? We need to solve the equation, $B = S^{-1}AS$ for input A. We can do this because S is an invertible matrix, so $SB = S(S^{-1}AS) = AS$ and multiplying by S^{-1} on the right $SBS^{-1} = A$ so that T is invertible, and the linear transformation is $T^{-1}(B) = SBS^{-1}$.

Example 3.27. Is the transformation $L(f) = \begin{bmatrix} f(1) \\ f(2) \\ f(3) \end{bmatrix}$ from $\mathcal{P}_3$ to $\mathbb{R}^3$ an isomorphism? Since $\dim\mathcal{P}_3 = 4$ and $\dim\mathbb{R}^3 = 3$, the spaces $\mathcal{P}_3$ and $\mathbb{R}^3$ fail to be isomorphic, so that L is not an isomorphism.

Example 3.28. Is the transformation $L(f) = \begin{bmatrix} f(1) \\ f(2) \\ f(3) \end{bmatrix}$ from $\mathcal{P}_2$ to $\mathbb{R}^3$ an isomorphism? Notice $\dim\mathcal{P}_2 = 3$ and $\dim\mathbb{R}^3 = 3$ so the dimensions of the domain and the target space have the same dimension. Checking the linearity conditions,

$$L(f_1 + f_2) = \begin{bmatrix} (f_1 + f_2)(1) \\ (f_1 + f_2)(2) \\ (f_1 + f_2)(3) \end{bmatrix} = \begin{bmatrix} f_1(1) \\ f_1(2) \\ f_1(3) \end{bmatrix} + \begin{bmatrix} f_2(1) \\ f_2(2) \\ f_2(3) \end{bmatrix} = L(f_1) + L(f_2)$$

and

$$T(kf) = \begin{bmatrix} (kf)(1) \\ (kf)(2) \\ (kf)(3) \end{bmatrix} = k \begin{bmatrix} f(1) \\ f(2) \\ f(3) \end{bmatrix} = kT(f).$$

The kernel of T consists of all polynomials $f(x)$ in $\mathcal{P}_2$ such that

$$T(f(x)) = \begin{bmatrix} f(1) \\ f(2) \\ f(3) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

Since a nonzero polynomial in $\mathcal{P}_2$ has at most two zeros, the zero polynomial is the only solution, so that $\ker T = \{0\}$. Therefore, T is an isomorphism.

Example 3.29. Determine whether the transformation

$$T(M) = M \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix} - \begin{bmatrix} 3 & 0 \\ 0 & 4 \end{bmatrix} M$$

from $\mathbb{R}^{2\times 2}$ to $\mathbb{R}^{2\times 2}$ is an isomorphism. Notice that $\dim\mathbb{R}^{2\times 2} = \dim\mathbb{R}^{2\times 2}$. Can we determine an invertible linear transformation T^{-1}? Checking the linearity conditions,

$$\begin{aligned} T(M + N) &= (M + N) \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix} - \begin{bmatrix} 3 & 0 \\ 0 & 4 \end{bmatrix} (M + N) \\ &= M \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix} - \begin{bmatrix} 3 & 0 \\ 0 & 4 \end{bmatrix} M + N \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix} - \begin{bmatrix} 3 & 0 \\ 0 & 4 \end{bmatrix} N = T(M) + T(N) \end{aligned}$$

and

$$\begin{aligned} T(kM) &= (kM) \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix} - \begin{bmatrix} 3 & 0 \\ 0 & 4 \end{bmatrix} (Mk) \\ &= k \left(M \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix} - \begin{bmatrix} 3 & 0 \\ 0 & 4 \end{bmatrix} M \right) k = kT(M) \end{aligned}$$

The kernel of T consists of $M = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ such that

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix} - \begin{bmatrix} 3 & 0 \\ 0 & 4 \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} -a & 0 \\ -2c & -d \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

Thus $a = c = d = 0$. However, when $M = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$ then $T(M) = 0$ so that $\ker T \neq \{0\}$, and so T is not an isomorphism.

Example 3.30. Determine whether the transformation $T(f(t)) = \begin{bmatrix} f(1) \\ f'(2) \\ f(3) \end{bmatrix}$ from $\mathcal{P}_2$ to $\mathbb{R}^3$ is a linear transformation. Then determine whether T is an isomorphism and if not then find its image and kernel. This transformation is linear since

$$T(f(t)+g(t)) = \begin{bmatrix} f(1) + g(1) \\ f'(2) + g'(2) \\ f(3) + g(3) \end{bmatrix} = \begin{bmatrix} f(1) \\ f'(2) \\ f(3) \end{bmatrix} + \begin{bmatrix} g(1) \\ g'(2) \\ g(3) \end{bmatrix} = T(f(t))+T(g(t))$$

and

$$T(kf(t)) = \begin{bmatrix} kf(1) \\ kf'(2) \\ kf(3) \end{bmatrix} = k \begin{bmatrix} f(1) \\ f'(2) \\ f(3) \end{bmatrix} = kT(f(t)).$$

Since $T(t-1)(t-3)) = T(t^2 - 4t + 3) = \vec{0}$ shows $\ker T \neq \{0\}$, T is not an isomorphism.

Theorem 3.13. *A linear map is invertible if and only if it is injective and surjective.*

Proof. Suppose $T \in \mathcal{L}(V, W)$ is invertible with inverse T. Let $u, v \in V$. If $Tu = Tv$ then

$$u = (T^{-1})u) = T^{-1}(Tu) = T^{-1}(Tv) = (T^{-1}T)(v) = v$$

and so T is injective. If $w \in W$, then $v = T^{-1}w \in V$ with $Tv = T(T^{-1}w) = w$ shows T is surjective. Assume T is injective and surjective. For each $w \in W$ assign $T(v) = w$. Such $S(w) = v$ exists because T is surjective and is unique since T is injective. Then $T(v) = w$ shows $ST(v) = S(w) = v$ so that ST is the identity on V. Also, $TS(w) = T(Sw) = Tv = w$ shows TS is the identity on W. Thus S and T are inverses.

Now S is linear since:

- if $w_1, w_2 \in W$, then there exists a unique v_1 and v_2 such that $Tv_1 = w_1$ and $Tv_2 = w_2$, $S(w_1) = v_1$, and $S(w_2) = v_2$. By linearity of T, $S(w_1 + w_2) = S(Tv_1 + Tv_2) = (ST)(v_1 + v_2) = v_1 + v_2 = S(w_1) + S(w_2)$.
- If $w \in W$ and $k \in \mathbb{F}$ then there exists a unique $v \in V$ such that $Tv = w$ and $S(w) = v$. By linearity of T, $S(kw) = S(kTv) = S(Tkv) = kv = kS(w)$.

Therefore S is linear and is the inverse of T.

$\square$

Consider the transformation

$$L\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix}$$

from $\mathbb{R}^{2\times2}$ to $\mathbb{R}^4$. Note that L is the coordinate transformation $L_{\mathcal{B}}$ with respect to the basis

$$\mathcal{B} = \left(\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}\right)$$

Being a coordinate transformation, L is both linear and invertible. Therefore L is an isomorphism.

Theorem 3.14. *Let T be the linear transformation from V to V and let B be the matrix of T with respect to a basis $\mathcal{B} = (f_1, \dots, f_n)$ of V. Then*

$$B = [[T(f_1)]_{\mathcal{B}} \quad \cdots \quad [T(f_n)]_{\mathcal{B}}]]$$

In other words, the columns of B are the $\mathcal{B}$-coordinate vectors of the transformation of the basis elements $f_1, \dots, f_n$ of V.

Proof. The proof if left for the reader.

$\square$

Theorem 3.15. *Two finite-dimensional vector spaces are isomorphic if and only if they have the same dimension.*

Proof. Suppose V and W are finite-dimensional vector spaces that are isomorphic. If $\dim V > \dim W$, then no linear map between T and W can be injective. Thus V and W are not isomorphic contrary to hypothesis. If $\dim V < \dim W$, then no linear map between V and W can be surjective. Thus V and W are not isomorphic contrary to hypothesis. Therefore, $\dim V = \dim W$.

Suppose V and W are vector spaces with $\dim V = \dim W = n$. Let $(v_1, \dots, v_n)$ ba a basis of V and let $(w_1, \dots, w_n)$ be a basis of W. Define a linear map T by $T(a_1 v_1 + \cdots + a_n v_n) = a_1 w_1 + \cdots + a_n w_n$. It's an easy exercise to show T is linear, injective, and surjective. Thus T is an isomorphism as required.

$\square$

Theorem 3.16. *If A and B are matrices of a linear transformation T with respect to different bases, then A and B are similar matrices.*

Proof. The proof is left for the reader.

$\square$

Theorem 3.17. *Suppose V and W are finite-dimensional vector spaces, then $\mathcal{M}$ is an invertible linear map between $\mathcal{L}(V, W)$ and $Mat(m, n, \mathbb{F})$*

Proof. Suppose $(v_1, \dots, v_n)$ and $w_1, \dots, w_m)$ are bases of V and W respectively. For each $T \in \mathcal{L}(V, W)$ we assign,

$$\mathcal{M}(T) = \begin{bmatrix} a_{11} & & a_{1n} \\ \vdots & \cdots & \vdots \\ a_{m1} & & a_{mn} \end{bmatrix}.$$

where $Tv_k = a_{1k}w_1 + \cdots + a_{mk}w_m$. Since $(w_1, \dots, w_m)$ is a basis of W, This assigment is well-defined. The following arguments show $\mathcal{M}$ is linear, injective, and surjective, respectively.

- If $S, T \in \mathcal{L}(V, W)$, then $\mathcal{M}(S + T) = \mathcal{M}(S) + \mathcal{M}(T)$ follows by the definition of matrix addition. If $S \in \mathcal{L}(V, W)$ and $k\mathbb{F}$, then $\mathcal{M}(kS) = k\mathcal{M}(S)$ follows by the definition of scalar multiplication of matrices.
- If $T \in \mathcal{L}(V, W)$ with $\mathcal{M}(T) = 0$, then for entries in $[a_{ij}]$ are zero showing $Tv_k = 0$ for each k. Since $(v_1, \dots, v_n)$ is a basis of V, T must be the zero map, and so ker$\mathcal{M} = \{0\}$. Therefore, $\mathcal{M}$ is injective.
- If $A \in Mat(m, n, \mathbb{F})$ then A is an $m \times n$ matrix with entries $[a_{ij}]$. We define $Tv_k = a_{1k}w_1 + \cdots a_{mk}w_m$ for each k. Then, by the definition of the matrix of a linear map, $\mathcal{M}(T) = A$ and so $\mathcal{M}$ is surjective.

$\square$

Theorem 3.18. *If V and W are finite-dimensional, then $\mathcal{L}(V, W)$ is finite-dimensional and $dim\mathcal{L}(V, W) = (dimV)(dimW)$.*

Proof. If V and W are finite-dimensional vector spaces then $\mathcal{L}(V, W)$ is isomorphic to $Mat(m, n, \mathbb{F})$ where $n = \dim V$ and $m = \dim W$. If follows

$$\dim\mathcal{L}(V, W) = \dim Mat(mn, \mathbb{F}) = mn = (\dim W)(\dim V).$$

since $\dim Mat(m, n, \mathbb{F}) = mn$.

$\square$

Theorem 3.19. *If V is finite-dimensional and $T \in \mathcal{L}(V)$, the the following are equivalent:*

- *T is invertible,*
- *T is injective, and*
- *T is surjective.*

Proof. The proof of each part follows.

- Suppose T is invertible, then by definition T is injective.

- Suppose T is injective. Thus $\ker T = \{0\}$. Since V is finite-dimensional, we can apply the Rank-Nullity Theorem so $\dim V = \dim \text{im} T$. Since $\text{im} T$ is a subspace of V with the same dimension as V, $\text{im} T = V$. Therefore, T is surjective.
- Suppose T is surjective. Then $T(V) = V$ and so $\dim V = \dim \text{im} T$. Since V is finite-dimensional, $\dim V = \dim \ker T + \dim \text{im T}$ by the Rank-Nullity Theorem. Thus, $\dim \ker T = 0$ and so $\ker T = \{0\}$. Therefore, T is injective, and since T is injective and surjective, T is invertible.

$\square$

3.4 Kernel and Image of a Linear Transformation

In this section we introduce the kernel and image of a linear transformation. We show how they can be realized as geometric objects and demonstrate how to find spanning sets for them.

Definition 3.7. Let T be a linear transformation from $\mathbb{R}^m$ to $\mathbb{R}^n$ with $n \times m$ matrix A.

- The **kernel** of T, denoted by $\ker(T)$, is the set of all vectors $\vec{x}$ in $\mathbb{R}^n$ such that $T(\vec{x}) = A\vec{x} = \vec{0}$.
- The **image** of T, denoted by $\text{im}(T)$, is the set of all vectors in $\mathbb{R}^n$ of the form $T(\vec{x}) = A\vec{x}$.

The kernel and image of a matrix A of T is defined as the kernel and image of T.

Now to adequately describe the kernel and image of a linear transformation we need the concept of the span of a collection of vectors. We will see that one of the best ways to describe the kernel and image of a linear transformation is to describe them in terms of collections of linear combinations of vectors.

The next three lemmas describe basic properties of kernel and image.

Lemma 3.2. *The kernel of any linear transformation T has the following properties:*

- $\vec{0} \in \ker(T)$,
- *if $\vec{v}, \vec{w} \in \ker(T)$, then $\vec{v} + \vec{w} \in \ker(T)$, and*
- *if $\vec{v} \in \ker(T)$ and $k \in \mathbb{R}$, then $k\vec{v} \in \ker(T)$.*

Proof. Let A be the matrix for T. Then $A\vec{0} = \vec{0}$ which shows $\vec{0} \in \ker(T)$. If $\vec{v}, \vec{w} \in \ker(T)$, then $A\vec{v} = \vec{0}$ and $A\vec{w} = \vec{0}$. Thus, $A\vec{v} + A\vec{w} = A(\vec{v} + \vec{w}) = \vec{0}$ implying $\vec{v} + \vec{w} \in \ker(T)$. If $\vec{v} \in \ker(T)$ and $k \in \mathbb{R}$, then $\vec{0} = A\vec{v}$ implying $A(k\vec{v}) = \vec{0}$. Thus, $k\vec{v} \in \ker(T)$.

$\square$

Lemma 3.3. *Let T be a linear transformation with matrix A.*

- *If A is an $n \times n$ matrix, then $\ker(A) = \{\vec{0}\}$ if and only if $rank(A) = n$.*

- *If A is an $n \times m$ matrix, then*

- $\ker(A) = \{\vec{0}\} \implies m \geq n$

- $m > n \implies \ker(A)$ *contains non-zero vectors*

- *If A is a square matrix, then $\ker(A) = \{\vec{0}\}$ if and only if A is invertible.*

Proof. If T is a linear transformation $T(\vec{x}) = A\vec{x}$ from $\mathbb{R}^m$ to $\mathbb{R}^n$ where $m > n$, then there will be free variables for the equation $T(\vec{x}) = A\vec{x} = \vec{0}$; that is the system will have infinitely many solutions. Therefore, the kernel of T will consists of infinitely many vectors. If $m = n$ and for an invertible $n \times n$ matrix A, how do we find $\ker(A)$? Since A is invertible, $A\vec{x} = \vec{0}$ can be solved by $A^{-1}(A\vec{x}) = A^{-1}\vec{0}$ showing $\vec{x} = \vec{0}$; that is the only solution to the system $A\vec{x} = \vec{0}$ is $\vec{0}$ so that $\ker(A) = \{\vec{0}\}$ whenever A is invertible.

$\square$

Lemma 3.4. *The image of any linear transformation T has the following properties:*

- $\vec{0} \in im(T)$,
- *if $\vec{v}, \vec{w} \in im(T)$, then $\vec{v} + \vec{w} \in im(T)$, and*
- *if $\vec{v} \in im(T)$ and $k \in \mathbb{R}$, then $k\vec{v} \in im(T)$.*

Proof. Let A be the matrix for T. Then $A\vec{0} = \vec{0}$ which shows $\vec{0} \in im(T)$. If $\vec{v}, \vec{w} \in im(T)$, then there exists $\vec{x}$ and $\vec{y}$ such that $A\vec{x} = \vec{v}$ and $A\vec{y} = \vec{w}$. Thus, $\vec{v} + \vec{w} = A\vec{x} + A\vec{y} = A(\vec{x} + \vec{y})$ implying $\vec{v} + \vec{w} \in im(T)$. If $\vec{v} \in im(T)$ and $k \in \mathbb{R}$, then there exists $\vec{u}$ such that $\vec{v} = A\vec{u}$ implying $A(k\vec{u}) = k\vec{v}$. Thus, $k\vec{v} \in im(T)$.

$\square$

Theorem 3.20. *Let $T : \mathbb{R}^m \to \mathbb{R}^n$ be a linear transformation with matrix A. Then $im(T) = span(\vec{v}_1, ..., \vec{v}_m)$ where $\vec{v}_1, ..., \vec{v}_m$ are the column vectors of A.*

Proof. Since $\vec{v}_1, ..., \vec{v}_m$ are the column vectors of A

$$T(\vec{x}) = A\vec{x} = \begin{bmatrix} \vec{v}_1 & \cdots & \vec{v}_m \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_m \end{bmatrix} = x_1 v_1 + \cdots + x_m v_m. \qquad (3.9)$$

If $\vec{u} \in span(\vec{v}_1, ..., \vec{v}_m)$ then there exists $x_1, ..., x_m$ such that $u = x_1 v_1 + \cdots + x_m v_m$. By 3.9, $u = T(\vec{x})$ for $\vec{x} \in \mathbb{R}^m$. Thus, $\vec{u} \in im(T)$. Conversely, assume $\vec{u} \in im(T)$. Then there exists $\vec{x}$ such that $\vec{u} = T(\vec{x})$. By 3.9, there exists $x_1, ..., x_m$ such that $u = x_1 v_1 + \cdots + x_m v_m$. Therefore, $\vec{u} \in span(\vec{v}_1, ..., \vec{v}_m)$ and so $im(T) = span(\vec{v}_1, ...\vec{v}_m)$ follows.::::

Example 3.31. Find vectors that span $ker(A)$ and $im(A)$ given

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

Describe $im(A)$ geometrically. To first find a spanning set of $ker(A)$ we solve the system $A\vec{x} = \vec{0}$. We use the augmented matrix and elementary row operations and find reduced row-echelon form

$$rref(A) = \begin{bmatrix} 1 & 0 & 2 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Thus the solution set is

$$\begin{bmatrix} x_1 \\ x_2 \\ x_2 \end{bmatrix} = t \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix} := t\vec{w}$$

where $t \in \mathbb{R}$. Therefore $ker(A) = span(\vec{w})$. Next we find $im(A)$. To do so let $\vec{v}_1, \vec{v}_2, \vec{v}_3$ be the column vectors of the matrix A. Since $im(A)$ is spanned by the columns of A and $\vec{v}_3 = 2\vec{v}_1 + (-1)\vec{v}_2$, we find $im(A) = span(\vec{v}_1, \vec{v}_2)$. Therefore the image of A is a plane in $\mathbb{R}^3$ that passes through the origin.

Example 3.32. Give an example of a matrix A such that $\text{im}(A)$ is the plane with normal vector $\vec{w} = \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix}$ in $\mathbb{R}^3$. Since $\vec{w}$ is a normal vector we let $A = \begin{bmatrix} 1 & 3 & 2 \end{bmatrix}$. First we find $\ker(A)$ because

$$\begin{bmatrix} 1 & 3 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

is the plane $x + 3y + 2z = 0$ where $\vec{w}$ is a normal vector. Let $z = t$ and $y = s$, then $x = -3s - 2t$ so the solutions to the system $A\vec{x} = \vec{0}$ are

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} -3s - 2t \\ s \\ t \end{bmatrix} = s \begin{bmatrix} 0 \\ 1 \\ -3 \end{bmatrix} + t \begin{bmatrix} 1 \\ 0 \\ -2 \end{bmatrix} := s\vec{u} + t\vec{v}$$

where s and t are real numbers; and thus $\ker(A) = \text{span}(\vec{u}, \vec{v})$. These vectors yield the image of A since $\text{im}(A)$ is the plane with normal vector $\vec{w}$ so A is one such matrix; and $\text{im}(A) = \text{span}(\vec{u}, \vec{v})$ is the plane in $\mathbb{R}^3$ with normal vector $\vec{w}$.

The kernel, being the most important subspace, has a special name for its dimension; namely, the dimension of $\ker A$ is called the **nullity** of A.

Example 3.33. Find the reduced row-echelon form of the matrix

$$A = \begin{bmatrix} 1 & 2 & 3 & 2 & 1 \\ 3 & 6 & 9 & 6 & 3 \\ 1 & 2 & 4 & 1 & 2 \\ 2 & 4 & 9 & 1 & 2 \end{bmatrix}.$$

Find a basis and state the dimension for the image and kernel of A.

Example 3.34. Find vectors that span $\ker(A)$ and $\text{im}(A)$ given

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

We will solve the system $A\vec{x} = \vec{0}$ to find $\ker(A)$. Using the augmented matrix and elementary row operations, we find

$$
\text{rref}(A) = \begin{bmatrix} 1 & -1 & 0 & 2 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.
$$

So the solutions to the system $A\vec{x} = \vec{0}$ are

$$
\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} s - 2t \\ s \\ -t \\ t \\ 0 \end{bmatrix} = s \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + t \begin{bmatrix} -2 \\ 0 \\ -1 \\ 1 \\ 0 \end{bmatrix} := s\vec{u} + t\vec{v}
$$

where $s, t \in \mathbb{R}$. Therefore $\ker(A) = \text{span}(\vec{u}, \vec{v})$. Since $\text{im}(A)$ is the span of the column vectors of A, we let $\vec{v}_1, \vec{v}_2, \vec{v}_3, \vec{v}_4, \vec{v}_5$ be the column vectors of A. By 3.20, $\text{im}(A) = \text{span}(\vec{v}_1, \vec{v}_2, \vec{v}_3, \vec{v}_4, \vec{v}_5)$. Using $\text{rref}(A)$ as a guide, we notice $\vec{v}_2 = (-1)\vec{v}_1$ so we eliminate $\vec{v}_2$. Also, $\vec{v}_4 = (2)\vec{v}_1 + \vec{v}_3$ so we also eliminate $\vec{v}_4$. Therefore, $\text{im}(A) = \text{span}(\vec{v}_1, \vec{v}_2, \vec{v}_5)$.

Example 3.35. Give an example of a linear transformation whose kernel is the line spanned by $\vec{w} = \begin{bmatrix} -1 \\ 1 \\ 2 \end{bmatrix}$. Considering the intersection of the planes $x + y = 0$ and $2x + z = 0$, we try to use the linear transformation,

$$
T(\vec{x}) = T \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} x + y \\ 2x + z \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 \\ 2 & 0 & 1 \end{bmatrix} \vec{x} := A\vec{x}.
$$

To find the kernel of T we solve $A\vec{x} = \vec{0}$. Since $\text{rref}(A) = \begin{bmatrix} 1 & 0 & 1/2 \\ 0 & 1 & -1/2 \end{bmatrix}$ the solutions are of the form $t\vec{w}$ where t is a real number. Therefore it suffices to let T be the requested linear transformation.

Example 3.36. Express the line L in $\mathbb{R}^3$ spanned by the vector $\vec{w} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ as the image of a matrix A and as the kernel of a matrix B. Let $A = \vec{w}$, then $L = \text{im}(A) = \text{span}(\vec{w})$. Therefore it suffices to let A be the requested

matrix. Considering the intersection of the planes $x = y$ and $y = z$, we try to find B using the linear transformation

$$T(\vec{x}) = T \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} x - y \\ y - z \end{bmatrix} = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{bmatrix} := B\vec{x}$$

To find the kernel of T we solve $B\vec{x} = \vec{0}$. Since $\operatorname{rref}(B) = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \end{bmatrix}$ the solutions are of the form $t\vec{w}$ where t is a real number. Therefore it suffices to let B be the other requested matrix.

Example 3.37. Find a basis for the kernel and the image of the linear transformations defined by

$$T_1 = \begin{cases} y_1 = & x_1 + x_2 + 3x_3 \\ y_2 = & 2x_1 + x_2 + 4x_3 \end{cases} \quad \text{with} \quad \operatorname{rref}(T_1) = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 2 \end{bmatrix}.$$

All solutions to the system $A\vec{x} = \vec{0}$ are $\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = t \begin{bmatrix} -1 \\ -2 \\ 1 \end{bmatrix}$ where $t \in \mathbb{R}$.

Since $\begin{bmatrix} -1 \\ -2 \\ 1 \end{bmatrix}$ is linearly independent and spans the kernel, the vector $\begin{bmatrix} -1 \\ -2 \\ 1 \end{bmatrix}$ forms a basis of $\ker(A)$. Since the columns of A spans the image of A and $\begin{bmatrix} 3 \\ 4 \end{bmatrix} = (1) \begin{bmatrix} 1 \\ 2 \end{bmatrix} + (2) \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ we find the vector $\begin{bmatrix} 3 \\ 4 \end{bmatrix}$ redundant. Since the remaining vectors are linearly independent and span $\operatorname{im}(A)$, they form a basis for $\operatorname{im}(A)$.

Example 3.38. Find a basis for the kernel and the image of the linear transformations defined by

$$T_2 = \begin{cases} y_1 = & x_1 + 3x_2 + 9x_3 \\ y_2 = & 4x_1 + 5x_2 + 8x_3 \\ y_3 = & 7x_1 + 6x_2 + 3x_3 \end{cases} \quad \text{with} \quad \operatorname{rref}(T_2) = \begin{bmatrix} 1 & 0 & -3 \\ 0 & 1 & 4 \\ 0 & 0 & 0 \end{bmatrix}.$$

To find a basis of the kernel we solve $A\vec{x} = \vec{0}$ where A is the matrix of the given transformation. Since

$$\operatorname{rref}(A) = \begin{bmatrix} 1 & 0 & -3 \\ 0 & 1 & 4 \\ 0 & 0 & 0 \end{bmatrix}$$

all solutions have the form $\begin{bmatrix} 3t \\ -4t \\ t \end{bmatrix}$. Therefore a basis of the kernel of A

is $\begin{bmatrix} 3 \\ -4 \\ 1 \end{bmatrix}$. In the original matrix A the third column is redundant since

$\begin{bmatrix} 9 \\ 8 \\ 3 \end{bmatrix} = (-3)\begin{bmatrix} 1 \\ 4 \\ 7 \end{bmatrix} + (4)\begin{bmatrix} 3 \\ 5 \\ 6 \end{bmatrix}$, and since the vectors $\begin{bmatrix} 1 \\ 4 \\ 7 \end{bmatrix}$ and $\begin{bmatrix} 3 \\ 5 \\ 6 \end{bmatrix}$ are linearly independent and span, they form a basis of $\text{im}(A)$.

Example 3.39. Find a basis for the kernel and the image of the linear transformations defined by

$$T_3 = \begin{cases} y_1 = & 4x_1 + 8x_2 + x_3 + x_4 + 6x_5 \\ y_2 = & 3x_1 + 6x_2 + x_3 + 2x_4 + 5x_5 \\ y_3 = & 2x_1 + 4x_2 + x_3 + 9x_4 + 10x_5 \\ y_4 = & x_1 + 2x_2 + 3x_3 + 2x_4 \end{cases} \quad \text{with} \quad \text{rref}(T_3) = \begin{bmatrix} 1 & 2 & 0 & 0 & \\ 0 & 0 & 1 & 0 & \\ 0 & 0 & 0 & 1 & \\ 0 & 0 & 0 & 0 & \end{bmatrix}$$

The solutions to the system $A\vec{x} = \vec{0}$ are

$$\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = t\begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} = t\vec{v} \qquad \text{where } t \in \mathbb{R}.$$

The vector $\vec{v}$ is linearly independent and spans $\ker(A)$; thus it forms a basis for $\ker(A)$. Since $\begin{bmatrix} 8 \\ 6 \\ 4 \\ 2 \end{bmatrix} = 2\begin{bmatrix} 4 \\ 3 \\ 2 \\ 1 \end{bmatrix}$ the vector $\begin{bmatrix} 8 \\ 6 \\ 4 \\ 2 \end{bmatrix}$ is redundant and since the column vectors of A are linearly independent and span $\text{im}(A)$, we have that the vectors $\begin{bmatrix} 4 \\ 3 \\ 2 \\ 1 \end{bmatrix}$, $\begin{bmatrix} 1 \\ 1 \\ 1 \\ 3 \end{bmatrix}$, $\begin{bmatrix} 1 \\ 2 \\ 9 \\ 2 \end{bmatrix}$, and $\begin{bmatrix} 6 \\ 5 \\ 10 \\ 0 \end{bmatrix}$ form a basis of $\text{im}(A)$.

Example 3.40. Show $\ker(A) \neq \ker(B)$ where

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

We solve the system $A\vec{x} = \vec{0}$, written out we find:

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{bmatrix} = s \begin{bmatrix} -4 \\ -5 \\ 0 \\ 6 \\ 1 \\ 0 \end{bmatrix} + t \begin{bmatrix} -2 \\ -3 \\ 1 \\ 0 \\ 0 \\ 00 \end{bmatrix} = s\vec{v}_1 + t\vec{v}_2 \quad \text{where } s, t \in \mathbb{R}.$$

Since v_1 and v_2 are linearly independent and span $\ker(A)$, they form a basis of the kernel of A. By noticing $B\vec{v}_1 \neq \vec{0}$ we conclude $\ker(A) \neq \ker(B)$.

Theorem 3.21. *For any matrix* A, $\dim(imA) = rank(A)$.

Proof. □

::: {#thm- } [Rank-Nullity] Let T be a linear transformation from $\mathbb{R}^m$ to $\mathbb{R}^n$ with $n \times m$ matrix A. Then

$$\text{rank}(A) + \text{nullity}(A) = \dim(imA) + \dim(\ker A) = m.$$

□

Proof. Let $\vec{y} = A\vec{x}$ be the corresponding system of linear equations. Recall,

$$\begin{pmatrix} \text{number of} \\ \text{free} \\ \text{variables} \end{pmatrix} = \begin{pmatrix} \text{total} \\ \text{number of} \\ \text{variables} \end{pmatrix} - \begin{pmatrix} \text{number of} \\ \text{leading} \\ \text{variables} \end{pmatrix} = \begin{array}{c} m - \\ \text{rank}(A) \end{array}$$

(3.10)

Apparently, the number of free variables is the dimension of the kernel of A. Thus 3.10 becomes $\text{nullity}(A) = m - \text{rank}(A)$. By 3.21 we arrive at the conclusion $\text{rank}(A) + \text{nullity}(A) = m$.

□

Example 3.41. If possible, find a 3×3 matrix such that $imA = \ker A$.

Example 3.42. If possible, find a 4×4 matrix such that $imA = \ker A$.

Example 3.43. Give an example of a 4×5 matrix A with $\dim(\ker A) = 3$.

Theorem 3.22. *The vectors $\vec{v}_1, \dots, \vec{v}_n$ in $\mathbb{R}^n$ form a basis of $\mathbb{R}^n$ if and only if the matrix whose columns consists of $\vec{v}_1, \dots, \vec{v}_n$ is invertible.*

Proof.   □

For example, consider the matrix

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

What is the smallest number of vectors needed to span the image of A? Of course we know,

$$\mathrm{im}(A) = \mathrm{span}\left(\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 2 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix} \right).$$

However, it is easy to show that $\begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix}$ and $\begin{bmatrix} 2 \\ 2 \\ 2 \end{bmatrix}$ are redundant; and that the remaining vectors are linearly independent. Thus,

$$\mathrm{im}(A) = \mathrm{span}\left(\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \right).$$

Clearly, the image of the linear transformation defined by A is more easily understood by having a spanning set of linearly independent vectors.

3.5   Invertible Linear Transformations

An $n \times n$ matrix A is called **invertible** if and only if there exists a matrix B such that $AB = I_n$ and $BA = I_n$. Using the inverse of a matrix we also define the inverse of a linear transformation.

Let $T(\vec{x}) = A\vec{x}$ be a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^n$. If the matrix A has inverse A^{-1}, then the linear transformation defined by $A^{-1}\vec{x}$ is called the **inverse transformation** of T and is denoted by $T^{-1}(\vec{x}) = A^{-1}$.

A function T from X to Y is called **invertible** if the equation $T(x) = y$ has a unique solution $x \in X$ for each $y \in Y$. A square matrix A is called **invertible** if the linear transformation $\vec{y} = T(\vec{x}) = A\vec{x}$ is invertible. In this case, then matrix of T^{-1} is denoted by A^{-1}. If the linear transformation is invertible, then its inverse is $\vec{x} = T^{-1}(\vec{y}) = A^{-1}$.

Example 3.44. Find the inverse transformation of the following linear transformation:

$$\begin{aligned} y_1 &= x_1 + 3x_2 + 3x_3 \\ y_2 &= x_1 + 4x_2 + 8x_3 \\ y_3 &= 2x_1 + 7x_2 + 12x_3 \end{aligned} \quad .$$

To find the inverse transformation we solve for x_1, x_2, x_3 in terms of $y_1, .y_2, y_3$. To do this we find the inverse matrix of

$$A = \begin{bmatrix} 1 & 3 & 3 \\ 1 & 4 & 8 \\ 2 & 7 & 12 \end{bmatrix}.$$

Applying elementary-row operations,

$$\begin{bmatrix} 1 & 3 & 3 & 1 & 0 & 0 \\ 1 & 4 & 8 & 0 & 1 & 0 \\ 2 & 7 & 12 & 0 & 0 & 1 \end{bmatrix} \quad \begin{matrix} R_2 - R_1 \\ \longrightarrow \\ -2R_1 + R_3 \end{matrix} \quad \begin{bmatrix} 1 & 3 & 3 & 1 & 0 & 0 \\ 0 & 1 & 5 & 1 & 1 & 0 \\ 0 & 1 & 6 & -2 & 0 & 1 \end{bmatrix}$$

$$\begin{matrix} \longrightarrow \\ -R_2 + R_3 \end{matrix} \begin{bmatrix} 1 & 3 & 3 & 1 & 0 & 0 \\ 0 & 1 & 5 & -1 & 1 & 0 \\ 0 & 0 & 1 & -1 & -1 & 1 \end{bmatrix} \begin{matrix} \longrightarrow \\ -5R_3 + R_2 \end{matrix} \begin{bmatrix} 1 & 3 & 3 & 1 & 0 & 0 \\ 0 & 1 & 0 & 4 & 6 & -5 \\ 0 & 0 & 1 & -1 & -1 & 1 \end{bmatrix}$$

$$\begin{matrix} \longrightarrow \\ -3R_3 + R_1 \end{matrix} \begin{bmatrix} 1 & 3 & 0 & 4 & -3 & 3 \\ 0 & 1 & 0 & 4 & 6 & -5 \\ 0 & 0 & 1 & -1 & -1 & 1 \end{bmatrix} \begin{matrix} \longrightarrow \\ -3R_2 + R_1 \end{matrix} \begin{bmatrix} 1 & 0 & 0 & -8 & -15 & 12 \\ 0 & 1 & 0 & 4 & 6 & -5 \\ 0 & 0 & 1 & -1 & -1 & 1 \end{bmatrix}$$

we find

$$A^{-1} = \begin{bmatrix} -8 & -15 & 12 \\ 4 & 6 & -5 \\ 1 & -1 & 1 \end{bmatrix}.$$

Therefore the requested linear transformation is

$$\begin{aligned} x_1 &= -8y_1 - 15y_2 + 12y_3 \\ x_2 &= 4y_1 + 6y_2 - 5y_3 \\ x_3 &= -y_1 - y_2 + y_3. \end{aligned}$$

Of course inverse transformations makes sense in terms of inverse functions; that is, if T^{-1} is the inverse transformation of T then $(T \circ T^{-1})(\vec{x}) = \vec{x}$ and $(T^{-1} \circ T)(\vec{x}) = \vec{x}$. For example, for T given in 3.11 we illustrate

$$(T^{-1} \circ T) \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = T^{-1} \begin{bmatrix} 2 \\ 4 \\ -5 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

as one can verify.

Example 3.45. Find the inverse of the linear transformation

$$y_1 = 3x_1 + 5x_2$$
$$y_2 = 3x_1 + 4x_2.$$

Reducing the system

$$\begin{bmatrix} 3x_1 + 5x_2 & = y_1 \\ 3x_1 + 4x_2 & = y_2 \end{bmatrix}$$

we obtain

$$\begin{bmatrix} x_1 & = -\frac{4}{3}y_1 + \frac{5}{3}y_2 \\ x_2 & = y_1 - y_2 \end{bmatrix}.$$

Theorem 3.23. *Let A be an $n \times n$ matrix. Then*

- *A is invertible if and only if $rref(A) = I_n$,*
- *A is invertible if and only if $rank(A) = n$, and*
- *A is invertible if and only if $A^{-1}A = I_n$ and $AA^{-1} = I_n$.*

Proof. The proof is left for the reader.

□

To find the inverse of an $n \times n$ matrix A, form the augmented matrix $[\,A\,|\,I_n\,]$ and compute $rref(, [, A , | , I_n ,] ,) $. If $rref(, [, A , | , I_n ,] ,) $ is of the form $rref(, [, I_n , | , B ,] ,) $, then A is invertible and $A^{-1} = B$. Otherwise A is not invertible. For example

$$= \mathrm{rref}\left(\begin{bmatrix} 1 & -1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ -1 & -2 & 0 & 0 & 0 & 1 \end{bmatrix} \right) = \begin{bmatrix} 1 & 0 & 0 & 2 & -2 & -1 \\ 0 & 1 & 0 & -1 & 1 & 0 \\ 0 & 0 & 1 & -2 & 3 & 1 \end{bmatrix} = [\,I_3\,|\,B\,]$$

(3.11)

shows $B = A^{-1}$ where

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

and

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

as one can verify, by showing $AB = I_3$ and $BA = I_3$.

Theorem 3.24. *Let A and B be $n \times n$ matrices. Then*

- *if A and B are invertible, then BA is invertible as well and*

$$(BA)^{-1} = A^{-1}B^{-1}$$

- *if $BA = I_n$, then A and B are both invertible,*

$$A^{-1} = B, \qquad B^{-1} = A, \qquad and \qquad AB = I_n.$$

Proof. The proof is left for the reader.

$\square$

Example 3.46. Find the inverse matrices of

$$A = \begin{bmatrix} 2 & 3 \\ 6 & 9 \end{bmatrix}$$

and

$$B = \begin{bmatrix} 1 & 2 \\ 3 & 9 \end{bmatrix}$$

Since $\operatorname{rref}(A) = \begin{bmatrix} 1/2 & 3/2 \\ 0 & 0 \end{bmatrix} \neq I_2$, A^{-1} does not exist. The inverse of B does exist and $B^{-1} = \begin{bmatrix} 3 & -2/3 \\ -1 & 1/3 \end{bmatrix}$ since $B^{-1}B = I_2$ and $BB^{-1} = I_2$.

Example 3.47. Show that $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is invertible if and only if $ad - bc \neq 0$ and when possible

$$A^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}. \qquad (3.12)$$

We proceed to find the inverse:

$$\begin{bmatrix} a & b & 1 & 0 \\ c & d & 0 & 1 \end{bmatrix} \xrightarrow[\frac{1}{c}R_2]{\frac{1}{a}R_1} \begin{bmatrix} 1 & \frac{b}{a} & \frac{1}{a} & 0 \\ 1 & \frac{d}{c} & 0 & \frac{1}{c} \end{bmatrix} \xrightarrow{-R_1 + R_2} \begin{bmatrix} 1 & \frac{b}{a} & \frac{1}{a} & 0 \\ 0 & \frac{ad-bc}{ac} & \frac{-1}{a} & \frac{1}{c} \end{bmatrix}$$

$$\xrightarrow{\frac{ac}{ad-bc}R_2} \begin{bmatrix} 1 & \frac{b}{a} & \frac{1}{a} & 0 \\ 0 & 1 & \frac{-c}{ad-bc} & \frac{a}{ad-bc} \end{bmatrix} \xrightarrow{\frac{-b}{a}R_2 + R_1} \begin{bmatrix} 1 & 0 & \frac{d}{ad-bc} & \frac{-b}{ad-bc} \\ 0 & 1 & \frac{-c}{ad-bc} & \frac{a}{ad-bc} \end{bmatrix}$$

Therefore, A is invertible if and only if $ad - bc \neq 0$ and 3.12 holds.

Example 3.48. For which values of constants a, b, c, is the matrix

$$A = \begin{bmatrix} 0 & a & b \\ -a & 0 & c \\ -b & -c & 0 \end{bmatrix}$$

invertible? Suppose $a \neq 0$. Applying row-operations

$$\begin{bmatrix} 0 & a & b & 1 & 0 & 0 \\ -a & 0 & c & 0 & 1 & 0 \\ -b & -c & 0 & 0 & 0 & 1 \end{bmatrix} \xrightarrow{R_2 \leftrightarrow R_1} \begin{bmatrix} -a & 0 & c & 0 & 1 & 0 \\ 0 & a & b & 1 & 0 & 0 \\ -b & -c & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\xrightarrow{-\frac{1}{a}R_1} \begin{bmatrix} 1 & 0 & -\frac{c}{a} & 0 & -\frac{1}{a} & 0 \\ 0 & a & b & 1 & 0 & 0 \\ -b & -c & 0 & 0 & 0 & 1 \end{bmatrix} \xrightarrow{bR_1 + R_3} \begin{bmatrix} 1 & 0 & -\frac{c}{a} & 0 & -\frac{1}{a} & 0 \\ 0 & a & b & 1 & 0 & 0 \\ 0 & -c & \frac{-bc}{a} & 0 & \frac{-b}{a} & 1 \end{bmatrix}$$

$$\xrightarrow{\frac{1}{a}R_2} \begin{bmatrix} 1 & 0 & -\frac{c}{a} & 0 & -\frac{1}{a} & 0 \\ 0 & 1 & \frac{b}{a} & \frac{1}{a} & 0 & 0 \\ 0 & -c & \frac{-bc}{a} & 0 & \frac{-b}{a} & 1 \end{bmatrix} \xrightarrow{cR_2 + R_3} \begin{bmatrix} 1 & 0 & -\frac{c}{a} & 0 & -\frac{1}{a} & 0 \\ 0 & 1 & \frac{b}{a} & \frac{1}{a} & 0 & 0 \\ 0 & 0 & 0 & \frac{c}{a} & \frac{-b}{a} & 1 \end{bmatrix}$$

Thus, if $a \neq 0$ then A is not invertible, since $\text{rref}(A) \neq I_3$. If $a = 0$, then clearly, $\text{rref}(A) \neq I_3$, and so A is not invertible in either case. Therefore, there are no constants a, b, c for which A is invertible.

Corollary 3.2. *Let A be an $n \times n$ matrix.*

- *Consider a vector $\vec{b}$ in $\mathbb{R}^n$. If A is invertible, then the system $A\vec{x} = \vec{b}$ has the unique solution $\vec{x} = A^{-1}b$. If A is non-invertible, then the system $A\vec{x} = \vec{b}$ has infinitely many solutions or none.*
- *The system $A\vec{x} = \vec{0}$ has $\vec{x} = \vec{0}$ as a solution. If A is invertible, then this is the only solution. If A is non-invertible, then the system $A\vec{x} = \vec{0}$ has infinitely many solutions.*

Proof. The proof is left for the reader.

$\square$

Example 3.49. Find all invertible matrices A such that $A^2 = A$. Since A is invertible we multiply by A^{-1} to obtain:

$$A = IA = (A^{-1}A)A = A^{-1}(A^2) = A^{-1}A = I_n$$

and therefore A must be the identity matrix.

Example 3.50. For which values of constants b and c is the matrix

$$B = \begin{bmatrix} 0 & 1 & b \\ -1 & 0 & c \\ -b & -c & 0 \end{bmatrix}$$

invertible? The matrix B is not invertible for any b and c since

$$\text{rref}(B) = \begin{bmatrix} 1 & 0 & -c \\ 0 & 1 & b \\ 0 & 0 & 0 \end{bmatrix} \neq I_3$$

for all b and c.

Example 3.51. Find the matrix A satisfying the equation

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

Let $B = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$ and $C = \begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix}$. Then

$$B^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \qquad \text{and} \qquad C^{-1} = \begin{bmatrix} 1/2 & 0 \\ 0 & -1/2 \end{bmatrix}.$$

Multiplying on the right by B^{-1} and on the left by C^{-1} we find

$$A = B^{-1} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} C^{-1} = \begin{bmatrix} 1/2 & -1/2 \\ -1/2 & 1/2 \end{bmatrix}.$$

Example 3.52. Suppose that A, B, and C are $n \times n$ matrices and that both A and B commute with C. Show that AB commutes with C. To show that AB commutes with C we need to show $(AB)C = C(AB)$. This is easy since

$$(AB)C = A(BC) = A(CB) = (AC)B = (CA)B = C(AB).$$

Can you justify each step?

Example 3.53. Show that $AB = BA$ if and only if $(A - B)(A + B) = A^2 - B^2$. Suppose $AB = BA$ we will show $(A - B)(A + B) = A^2 - B^2$. Starting with the left-hand side we obtain

$$(A - B)(A + B) = (A - B)A + (A - B)B = A^2 - BA + AB - B^2$$
$$= A^2 - BA + BA - B^2 = A^2 - B^2$$

Now suppose $(A - B)(A + B) = A^2 - B^2$, we will show $AB = BA$. This is easy since

$$(A - B)(A + B) = A^2 - BA + AB - B^2 = A^2 - B^2$$

implies $-BA + AB = 0$ as desired.

3.6   Coordinates

Definition 3.8. Let $\mathcal{B} = (v_1, ..., v_n)$ be a basis of a subspace V of $\mathbb{R}^n$. For any $\vec{x} \in V$ we can write $\vec{v} = c_1 \vec{v}_1 + \cdots + c_m \vec{v}_m$. The scalars $c_1, ..., c_m$ are called the $\mathcal{B}$-**coordinates** of $\vec{x}$ and the vector

$$[\vec{x}]_{\mathcal{B}} := \begin{bmatrix} c_1 \\ ... \\ c_m \end{bmatrix}$$

is called the $\mathcal{B}$-**coordinate vector** of $\vec{x}$

For example the coordinates of $\begin{bmatrix} -2 \\ 4 \end{bmatrix}$ with respect the standard basis $\mathcal{B} = (\vec{e}_1, \vec{e}_2)$ is $\begin{bmatrix} -2 \\ 4 \end{bmatrix}$ since $\begin{bmatrix} -2 \\ 4 \end{bmatrix} = -2\vec{e}_1 + 4\vec{e}_2$ which written as $\begin{bmatrix} -2 \\ 4 \end{bmatrix}_{\mathcal{B}} = \begin{bmatrix} -2 \\ 4 \end{bmatrix}$.

Notice the coordinates of $\begin{bmatrix} -2 \\ 4 \end{bmatrix}$ with respect the basis $\mathcal{B}' = \left(\begin{bmatrix} 2 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 2 \end{bmatrix} \right)$ is $\begin{bmatrix} -1 \\ 2 \end{bmatrix}$ since $\begin{bmatrix} -2 \\ 4 \end{bmatrix} = (-1)\begin{bmatrix} 2 \\ 0 \end{bmatrix} + 2\begin{bmatrix} 0 \\ 2 \end{bmatrix}$ which written as $\begin{bmatrix} -2 \\ 4 \end{bmatrix}_{\mathcal{B}'} = \begin{bmatrix} -1 \\ 2 \end{bmatrix}$.

Example 3.54. Consider the plane $2x_1 - 3x_2 + 4x_3 = 0$ with basis

$$\mathcal{B} = \left(\begin{bmatrix} 8 \\ 4 \\ -1 \end{bmatrix}, \begin{bmatrix} 5 \\ 2 \\ -1 \end{bmatrix} \right).$$

Let $[\vec{x}]_{\mathcal{B}} = \begin{bmatrix} 2 \\ -1 \end{bmatrix}$. Find $\vec{x}$. By definition of coordinates

$$\vec{x} = 2\begin{bmatrix} 8 \\ 4 \\ -1 \end{bmatrix} + (-1)\begin{bmatrix} 5 \\ 2 \\ -1 \end{bmatrix} = \begin{bmatrix} 11 \\ 6 \\ -1 \end{bmatrix}.$$

Lemma 3.5. *If $\mathcal{B} = (\vec{v}_1, ..., \vec{v}_n)$ is a basis of a subspace V of $\mathbb{R}^n$, then*

- $[\vec{x}]_{\mathcal{B}} + [\vec{y}]_{\mathcal{B}} = [\vec{x} + \vec{y}]_{\mathcal{B}}$, *for all vectors $\vec{x}, \vec{y}$*
- $[k\vec{x}]_{\mathcal{B}} = k[\vec{x}]_{\mathcal{B}}$, *for all vectors $\vec{x}$, and all scalars k.*

Proof. Let $[\vec{x}]_{\mathcal{B}} = c_1\vec{v}_1 + \cdots + c_n\vec{v}_n$
and $[\vec{y}]_{\mathcal{B}} = d_1\vec{v}_1 + \cdots + d_n\vec{v}_n$ be the representation of $\vec{x}$ and $\vec{y}$ with respect to $\mathcal{B}$. Then

$$[\vec{x}]_{\mathcal{B}} + [\vec{y}]_{\mathcal{B}} = \begin{bmatrix} c_1 \\ \vdots \\ c_3 \end{bmatrix} + \begin{bmatrix} d_1 \\ \vdots \\ d_3 \end{bmatrix} = \begin{bmatrix} c_1 + d_1 \\ \vdots \\ c_n + d_n \end{bmatrix} = [\vec{x} + y]_{\mathcal{B}}$$

where the last equality holds since

$$\vec{x} + \vec{y} = (c_1 + d_1)\vec{v}_1 + \cdots (c_n + d_n)\vec{v}_n.$$

We leave the proof of the second part for the reader.

$\square$

Example 3.55. Determine whether the vector $\vec{x} = \begin{bmatrix} 1 \\ -2 \\ -2 \end{bmatrix}$ is in span$V$ of

the vectors $\vec{v}_1 = \begin{bmatrix} 8 \\ 4 \\ -1 \end{bmatrix}$ and $\vec{v}_2 = \begin{bmatrix} 5 \\ 2 \\ -1 \end{bmatrix}$ and if so write the coordinates of

$\vec{x}$ with respect to this basis of V. We need to find scalars c_1 and c_2 such that $\vec{x} = c_1\vec{v}_1 + c_2\vec{v}_2$. Solving this system we find $c_1 = -3$ and $c_2 = 5$. Therefore we find, $[\vec{x}]_{\mathcal{B}} = \begin{bmatrix} -3 \\ 5 \end{bmatrix}$ where $\mathcal{B} = (\vec{v}_1, \vec{v}_2)$.

Example 3.56. Consider the plane $x + 2y + z = 0$. Find a basis of this plane. Find another basis $\mathcal{B}$ of this plane such that $[\vec{x}]_{\mathcal{B}} = \begin{bmatrix} 2 \\ -1 \end{bmatrix}$ for

$\vec{x} = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$. We find a basis by letting $z = t$ and $y = s$ be free variables.

Then $x = -t - 2s$. All solutions to the equation $x + 2y + z = 0$ are

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

So a basis for the plane is $\mathcal{B} = \left(\begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix} \right)$. Notice $\begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} =$

$(1)\begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} + (-1)\begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix}$ and thus $\begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}_{\mathcal{B}} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$. Thus this is not the

basis we seek. However, notice

$$\begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} = 2 \begin{bmatrix} x_1 \\ 0 \\ -x_1 \end{bmatrix} + (-1) \begin{bmatrix} -2y_2 \\ y_2 \\ 0 \end{bmatrix}$$

holds when $y_2 = 1$ and $x_1 = -1/2$. Also notice the vectors $\begin{bmatrix} -1/2 \\ 0 \\ 1/2 \end{bmatrix}$ and

$\begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix}$ span the plane and are linearly independent. Therefore we have

the basis we seek, namely

$$\mathcal{B} = \left(\begin{bmatrix} -1/2 \\ 0 \\ 1/2 \end{bmatrix}, \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix} \right).$$

Lemma 3.6. *Let T be a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^n$ and $\mathcal{B} = (\vec{v}_1, ..., \vec{v}_n)$ a basis of $\mathbb{R}^n$. The $n \times n$ matrix B that transforms $[\vec{x}]_\mathcal{B}$ into $[T(\vec{x})]_\mathcal{B}$ is called the $\mathcal{B}$-**matrix** of T, written as $[T(\vec{x})]_\mathcal{B} = B[\vec{x}]_\mathcal{B}$ for all $\vec{x}$ in $\mathbb{R}^n$ and*

$$B = \left[[T(\vec{v}_1)]_\mathcal{B} \quad \cdots \quad [T(\vec{v}_n)]_\mathcal{B} \right].$$

Proof. Since $\mathcal{B}$ is a basis of $\mathbb{R}^n$, there exists scalars $c_1, ..., c_n$ such that $\vec{x} = c_1\vec{v}_1 + c_2\vec{v}_2 + \cdots + c_n\vec{v}_n$. Using the linearity of T we find

$$T(\vec{x}) = c_1 T(\vec{v}_1) + c_2 T(\vec{v}_2) + \cdots + c_n T(\vec{v}_n).$$

By 3.5 we find

$$[T(\vec{x})]_\mathcal{B} = c_1 [T(\vec{v}_1)]_\mathcal{B} + c_2 [T(\vec{v}_2)]_\mathcal{B} + \cdots + c_n [T(\vec{v}_n)]_\mathcal{B}$$
$$= \left[[T(\vec{v}_1)]_\mathcal{B} \quad \cdots \quad [T(\vec{v}_n)]_\mathcal{B} \right] [\vec{x}]_\mathcal{B}$$

as desired.

$\square$

Example 3.57. Find the matrix $\mathcal{B}$ of the linear transformation $T(\vec{x}) = A\vec{x}$ where

$$A = \begin{bmatrix} 5 & -4 & -2 \\ -4 & 5 & -2 \\ -2 & -2 & 8 \end{bmatrix}$$

with respect to the basis $\mathcal{B} = \left(\begin{bmatrix} 2 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ -2 \end{bmatrix} \right)$. By 3.6, T with

respect to $\mathcal{B}$ has the following matrix B.

$$\begin{bmatrix} \left[T \begin{bmatrix} 2 \\ 2 \\ 1 \end{bmatrix} \right]_{\mathcal{B}} & \left[T \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} \right]_{\mathcal{B}} & \left[T \begin{bmatrix} 0 \\ 1 \\ -2 \end{bmatrix} \right]_{\mathcal{B}} \end{bmatrix} = \begin{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}_{\mathcal{B}} & \begin{bmatrix} 9 \\ -9 \\ 0 \end{bmatrix}_{\mathcal{B}} & \begin{bmatrix} 0 \\ 9 \\ -18 \end{bmatrix}_{\mathcal{B}} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 9 & 0 \\ 0 & 0 & 9 \end{bmatrix}.$$

Theorem 3.25. *Let T be a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^n$ with standard matrix A and let B be the $\mathcal{B}$-matrix of T where $\mathcal{B} = (\vec{v}_1, ..., \vec{v}_n)$ then $AS = SB$ where*

$$S = \begin{bmatrix} \vec{v}_1 & \cdots & \vec{v}_n \end{bmatrix}.$$

Proof. The proof is left for the reader.

$\square$

Definition 3.9. Two $n \times n$ matrices A and B are called **similar** if there exists an invertible matrix S such that $AS = SB$.

Example 3.58. Determine whether the following two matrices are similar.

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

We are looking for a matrix

$$S = \begin{bmatrix} x & y \\ z & t \end{bmatrix}$$

such that $AS = SB$. Writing out we find

$$\begin{bmatrix} 1 & 2 \\ 4 & 3 \end{bmatrix} \begin{bmatrix} x & y \\ z & t \end{bmatrix} = \begin{bmatrix} x & y \\ z & t \end{bmatrix} \begin{bmatrix} 5 & 0 \\ 0 & -1 \end{bmatrix} \implies \begin{bmatrix} x + 2z & y + 2t \\ 4x + 3z & 4y + 3t \end{bmatrix} = \begin{bmatrix} 5x & -y \\ 5z & -t \end{bmatrix}.$$

This leads to $z = 2x$ and $t = -y$ so that S is any invertible matrix of the form

$$\begin{bmatrix} x & y \\ 2x & -y \end{bmatrix}. \tag{3.13}$$

Since $S = \begin{bmatrix} 1 & 1 \\ 2 & -1 \end{bmatrix}$ is invertible and of the form in (3.13), we can say, yes A is similar to B.

Theorem 3.26. *If A is similar to B then A^k is similar to B^k for any positive integer k.*

Proof. If A is similar to B then there exists an invertible matrix S such that $B = S^{-1}AS$. Then

$$B^k = (S^{-1}AS)(S^{-1}AS)\cdots(S^{-1}AS) = S^{-1}A^kS$$

which shows that B^k is similar to A^k.

□

Theorem 3.27. *Two $n \times n$ matrices A, A' are similar if and only if they are matrices of the same linear transformation T from $\mathbf{R}^n$ to $\mathbf{R}^n$ with respect to two bases for $\mathbf{R}^n$*

Proof. The proof is left for the reader.

□

Theorem 3.28. *Similar matrices have the same rank.*

Proof. Let A and B be similar matrices. By 3.27 A and B represent the same linear transformation. Thus they must have the same rank.

□

Example 3.59. Are the following similiar matrices?

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

Since $\text{rank}(A) = 3$ and $\text{rank}(B) = 2$, these are not similar matrices.

Theorem 3.29. *Similarity of matrices is an equivalence relation.*

Proof. To show transitivity, assume A is similar to B and B is similar to C, then there exists invertible matrices E and F such that $AE = EB$ and $BF = FC$. Then

$$A = EBE^{-1} = E(FCF^{-1})E^{-1} = (EF)C(F^{-1}E^{-1}) = (EF)C(EF)^{-1}$$

which shows that A is similar to C. The proof of the reflexive property and the symmetric property are left for the reader.

□

By 3.29, a linear transformation represents a whole class of (similar) matrices. That is the collection of $n \times n$ matrices is partitioned into non overlapping sets of matrices, with all matrices in any such set being similar and representing the same linear transformation from $\mathbf{R}^n$ to $\mathbf{R}^n$.

3.7   Coordinates and the Matrix of a Linear Map

Let V and W be finite dimensional linear spaces.

Definition 3.10 (The Matrix of a Linear Map). Let $T \in \mathcal{L}(V, W)$ and let $b_1 = \{v_1, \dots, v_n\}$ be a basis for V and $b_2 = \{w_1, \dots, w_m\}$ be a base for W. Then the matrix of T with respect to the bases b_1 and b_2 is

$$\begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \cdots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix}$$

where the $a_{ij} \in \mathbb{F}$ are determined by $Tv_k = a_{1k}w_1 + \cdots + a_{mk}w_m$ for each $k = 1, \dots, n$.

Example 3.60. Consider the linear transformation $T(f) = f' + f''$ from $\mathcal{P}_2$ to $\mathcal{P}_2$. Since $\mathcal{P}_2$ is isomorphic to $\mathbb{R}^3$ with isomorphism given by a 3×3 matrix B, how do we find this matrix B? Let $a + bx + cx^2$, then we write T as

$$T(a + bx + cx^2) = (a + bx + cx^2)' + (a + bx + cx^2)''$$
$$= b + 2cx + 2c = (b + 2c) + 2cx.$$

Next let's write the input $f(x) = a + bx + cx^2$ and the output $T(f(x)) = (b + 2c) + 2cx$ in coordinates with respect to the standard bases $\mathcal{B} = (1, x, x^2)$ of $\mathcal{P}_2$ Written in $\mathcal{P}_2$ coordinates, transformation T takes $[f(x)]_\mathcal{B}$ to

$$[T(f(x))]_\mathcal{B} = \begin{bmatrix} b + 2c \\ 2c \\ 0 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 2 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} 0 & 1 & 2 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{bmatrix} [f(x)]_\mathcal{B}$$

The matrix

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

is called the $\mathcal{B}$-matrix of the transformation T.

Example 3.61. Find the B-matrix for the linear transformation given by

$$T(M) = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} M - M \begin{bmatrix} 5 & 0 \\ 0 & -1 \end{bmatrix} M$$

from $\mathbb{R}^{2\times 2}$ to $\mathbb{R}^{2\times 2}$. Determine whether T is an isomorphism and if not find kernel, image, nullity, and the rank of T. We will use the standard basis of $\mathbb{R}^{2\times 2}$:

$$\mathcal{B} = \left(\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right)$$

and we will construct the $\mathcal{B}$-matrix column-by-column:

$$B = \left[\left[T \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \right]_{\mathcal{B}}, \left[T \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \right]_{\mathcal{B}}, \left[T \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \right]_{\mathcal{B}}, \left[T \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right]_{\mathcal{B}} \right]$$

$$= \left[\begin{bmatrix} -4 & 0 \\ 4 & 0 \end{bmatrix}_{\mathcal{B}}, \begin{bmatrix} 0 & 2 \\ 0 & 4 \end{bmatrix}_{\mathcal{B}}, \begin{bmatrix} 2 & 0 \\ -2 & 0 \end{bmatrix}_{\mathcal{B}}, \begin{bmatrix} 0 & 2 \\ 0 & 4 \end{bmatrix}_{\mathcal{B}} \right] = \begin{bmatrix} -4 & 0 & 2 & 0 \\ 0 & 2 & 0 & 2 \\ 4 & 0 & -2 & 0 \\ 0 & 4 & 0 & 4 \end{bmatrix}$$

with

$$\text{rref}(B) = \begin{bmatrix} 1 & 0 & -\frac{1}{2} & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

After eliminating redundant columns from B we find a basis of $\text{im}\, T$ is

$$\left(\begin{bmatrix} -4 \\ 0 \\ 4 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 2 \\ 0 \\ 4 \end{bmatrix} \right).$$

To find $\ker T$ we solve $B\vec{x} = \vec{0}$ and using the rref(B) we find

$$\left(\begin{bmatrix} -\frac{1}{2} \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \\ 0 \\ 1 \end{bmatrix} \right)$$

to be a basis for $\ker T$. Notice since the rank of T is $2 = \text{dimim}\, T$, T is not an isomorphism. Therefore, since $\ker T \neq \{0\}$, T is not an isomorphism.

Example 3.62. Find the matrix of the linear transformation $T(f(t)) = f(3)$ from $\mathcal{P}_2$ to $\mathcal{P}_2$ with respect to the basis $(1, t-3, (t-3)^2)$. Determine whether the transformation is an isomorphism, it it isn't an isomorphism then determine the kernel and image of T, and also determine the nullity and rank of T. The matrix of T is

$$B = [[T(1)]_{\mathcal{B}} \quad [T(x-3)]_{\mathcal{B}} \quad [T((x-3)^2)]_{\mathcal{B}}] = [[1]_{\mathcal{B}} \quad [0]_{\mathcal{B}} \quad [0]_{\mathcal{B}}] = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Notice the vectors $\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$, $\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$ form a basis of the kernel of B and $\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$ is a basis of the image of B. Therefore, the rank is 1 and the nullity is 2; and therefore, T is not an isomorphism.

Definition 3.11 (The Matrix of a Vector). Let $b = \{v_1, \dots, v_n\}$ be a basis for V and let $v \in V$. We define the matrix of v, denoted by $\mathcal{M}(v)$, to be the n-by-1 matrix $\begin{bmatrix} b_1 & \cdots & b_n \end{bmatrix}^T$.

Theorem 3.30. If $T \in \mathcal{L}(V,W)$, then $\mathcal{M}(Tv) = \mathcal{M}(T)\mathcal{M}(v)$ for all $v \in V$.

Proof. Let $(v_1, \dots, v_n)$ be a basis of V and $(w_1, \dots, w_m)$ be a basis of W. If $v \in V$, then there exists $b_1, \dots, b_n \in \mathbb{F}$ such that $v = b_1 v_1 + \cdots + b_n v_n$ so that $\mathcal{M}(v) = \begin{bmatrix} b_1 & \cdots & b_n \end{bmatrix}^T$. For each k, $1 \le k \le n$ we write $Tv_k = a_{1k}w_1 + \cdots + a_{mk}w_m$ and so by definition of the matrix of a linear map T:

$$\mathcal{M} = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & & a_{mn} \end{bmatrix}.$$

By linearity of T:

$$\begin{aligned} Tv &= b_1 Tv_1 + \cdots b_n Tv_n \\ &= b_1 \left(\sum_{j=1}^m a_{j1} w_j \right) + \cdots + b_n \left(\sum_{j=1}^m a_{jn} w_j \right) \\ &= w_1(a_{11}b_1 + \cdots + a_{1n}b_n) + \cdots + w_m(a_{m1}b_1 + \cdots + a_{mn}b_n). \end{aligned}$$

Therefore,

$$\mathcal{M}(Tv) = \begin{bmatrix} a_{11}b_1 + \cdots + a_{1n}b_n \\ \cdots \\ a_{m1}b_1 + \cdots + a_{mn}b_n \end{bmatrix} = \mathcal{M}(T)\mathcal{M}(v).$$

where the last equality holds by definition of matrix multiplication.

$\square$

Let $\overline{u} = (u_1, \dots, u_p)$ be a basis of U, let $\overline{v} = (v_1, \dots, v_n)$ be a basis of V, and let $\overline{w} = (w_1, \dots, w_m)$ be a basis of W. If $T \in \mathcal{L}(U, V)$ and $S \in \mathcal{L}(V, W)$, then $ST \in \mathcal{L}(U, W)$ and by the definition of matrix multiplication,

$$\mathcal{M}(ST, \overline{u}, \overline{w}) = \mathcal{M}(S, \overline{v}, \overline{w})\mathcal{M}(T, \overline{u}, \overline{v}). \tag{3.14}$$

Theorem 3.31. *If $\overline{u} = (u_1, \dots, u_p)$ and $\overline{v} = (v_1, \dots, v_n)$ are bases of V, then $\mathcal{M}(I, \overline{u}, \overline{v})$ is invertible and*

$$\mathcal{M}(I, \overline{u}, \overline{v})^{-1} = \mathcal{M}(I, \overline{v}, \overline{u}).$$

Proof. In 3.14, replace U and W with V, replace w_j with u_j, and replace S and T with I, getting

$$I = \mathcal{M}(I, (\overline{v}, \overline{u}))\mathcal{M}(I, \overline{u}, \overline{v}).$$

Now interchange the roles of the u's and v's, getting

$$I = \mathcal{M}(I, (\overline{u}, \overline{v}))\mathcal{M}(I, \overline{v}, \overline{u}).$$

These equations give the desired result.

$\square$

For example, obviously,

$$\mathcal{M}\left(I, \left(\begin{bmatrix} 4 \\ 2 \end{bmatrix}, \begin{bmatrix} 5 \\ 3 \end{bmatrix}\right), \left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}\right)\right) = \begin{bmatrix} 4 & 5 \\ 2 & 3 \end{bmatrix}.$$

The inverse of the matrix above is $\begin{bmatrix} 3/2 & -5/2 \\ -1 & 2 \end{bmatrix}$. Thus,

$$\mathcal{M}\left(I, \left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}\right), \left(\begin{bmatrix} 4 \\ 2 \end{bmatrix}, \begin{bmatrix} 5 \\ 3 \end{bmatrix}\right)\right) = \begin{bmatrix} 3/2 & -5/2 \\ -1 & 2 \end{bmatrix}.$$

Theorem 3.32. *Suppose $T \in \mathcal{L}(V)$. Let $\overline{u} = (u_1, \dots, u_n)$ and $(v_1, \dots, v_n)$ be bases of V. Let $A = \mathcal{M}(I, \overline{u}, \overline{v})$. Then*

$$\mathcal{M}(T, \overline{u}) = A^{-1}\mathcal{M}(T, \overline{v})A. \tag{3.15}$$

Proof. In 3.14, replace U and W with V, replace w_j with v_j, replace T with I, and replace S with T, getting

$$\mathcal{M}(T, \overline{u}, \overline{v}) = \mathcal{M}(T, \overline{v})A. \tag{3.16}$$

In 3.14, replace U and W with V, replace w_j with u_j, replace S with I, and replace S with T, getting

$$\mathcal{M}(T, \overline{u}) = A^{-1}\mathcal{M}(T, \overline{u}, \overline{v}) \qquad (3.17)$$

by **??**. Substitution of 3.16 into 3.17 yields 3.15.

□

Example 3.63. Prove that every linear map from $\mathrm{Mat}(n, 1, F)$ to $\mathrm{Mat}(m, 1, F)$ is given by matrix multiplication. In other words, prove that if T is a linear transformation from $\mathrm{Mat}(n, 1, F)$ to $\mathrm{Mat}(m, 1, F))$, then there exists an m-by-n matrix A such that $TB = AB$ for every $B \in \mathrm{Mat}(n, 1, F)$. Let $(e_1, \dots, e_n)$ be a basis for $Mat(n, 1, \mathbb{F})$ and let $(v_1, \dots, v_m)$ be a basis for $Mat(m, 1, \mathbb{F})$. For each k, there exists $a_{1k}, \dots, a_{mk} \in \mathbb{F}$ such that $Te_k = a_{1k}v_1 + \cdots a_{mk}v_m$. Define the $m \times n$ matrix A as follows:
$$A = \begin{bmatrix} Te_1 & \cdots & Te_n \end{bmatrix}.$$

If $B \in Mat(n, 1, \mathbb{F})$ there exists $b_1, \dots, b_n \in \mathbb{F}$ such that $B = b_1 e_1 + \cdots + b_n e_n$, and thus

$$TB = T(b_1 e_1 + \cdots + b_n e_n) = b_1 Te_1 + \cdots + b_n Te_n = BA$$

as desired. Notice the word "the" follows since $(v_1, \dots, v_m)$ is a basis. In other words one bases have been chosen, the matrix A is unique.

Example 3.64. Suppose that V is finite-dimensional and $S, T \in \mathcal{L}(V)$. Prove that ST is invertible if and only if both S and T are invertible. Suppose both S and T are invertible. Then both S and T are injective and surjective. Thus, ST is both injective and surjective showing ST is invertible. Conversely, suppose ST is invertible. Since $\ker T \subseteq \ker ST = \{0\}$ because ST is injective, T is also injective. Thus, T is invertible. Since ST is surjective, if $w \in W$, then there exists $v \in V$ such that $(ST)v = w$. Rewriting $S(Tv) = w$, showing S is surjective. Thus, S is also invertible.

Example 3.65. Suppose that V is finite-dimensional and $S, T \in \mathcal{L}(V)$. Prove that $ST = I$ if and only if $TS = I$. Without loss of generality, we will show $ST = I impliesTS = I$. Suppose $ST = I$. Since I is invertible, the previous exercise implies S and T are both invertible. Then $ST = I impliesS^{-1}(ST) = S^{-1}I impliesT = S^{-1}$. Therefore, $TS = S^{-1}S = I$.

Example 3.66. Suppose that V is finite-dimensional and $T \in \mathcal{L}(V)$. Prove that T is a scalar multiple of the identity if and only if $ST = TS$ for every $S \in \mathcal{L}(V)$. If T is a scalar multiple of the identity, say $T = \alpha I$, then for all $v \in V$,

$$STv = S\alpha v = \alpha Sv = TSv.$$

Conversely, suppose $ST = TS$ for every $S \in \mathcal{L}(V)$. Pick a basis $v_1, \ldots, v_N$) for V. For $m = 1, \ldots, N$, define linear maps $S + m \in \mathcal{L}(V)$ by

$$S_m = v_n = \begin{cases} v_m & \text{if } m = n \\ 0 & \text{if } m \neq n \end{cases}$$

Now if $v = \sum \alpha_n v_n$, then

$$S_m \sum \alpha_n v_n = \alpha_m v_m$$

Thus the only vectors satisfying $S_m v = v$ are $v = \alpha v_m$ for some $\alpha \in \mathbb{F}$. The condition $S_m T = T S_m$ gives

$$S_m T v_m = T S_m v_m = T v_m$$

and by the above observation $T v_m = \alpha_m v_m$. Now consider another collection of linear maps $A(m, n)$ defined by

$$A(m,n)v_m = v_n, \qquad A(m,n)v_n = v_m, \qquad A(m,n)v_k = 0, \text{ when } k \neq m, n.$$

The condition $A(m,n)T v_n = T A(m,n)v_n$ gives

$$A(m,n)T v_n = T A(m,n)v_n = T v_m = \alpha_m v_m$$

and

$$A(m,n)\alpha_n v_n = A(m,n)\alpha_n v_n = \alpha_n A(m,n)v_n = \alpha_n v_m.$$

Whence $\alpha_m v_m = \alpha_n v_m$ that is, $\alpha_m = \alpha_n$ for $m, n = 1, \ldots, N$; and thus T is a scalar multiple of the identity.

Example 3.67. Prove that if V is finite-dimensional with $\dim V > 1$, then the set of non-invertible operators on V is not a subspace of $\mathcal{L}(V)$. Suppose $(v_1, \ldots, v_n)$ is a basis of V, with $n \geq 2$. Define the linear maps S and T by

$$Sv_1 = v_1, \qquad Sv_k = 0, \text{ when } k \geq 2$$
$$Tv_1 = 0, \qquad Tv_k = v_k, \text{ when } k \geq 2.$$

Since S and T have nontrivial null spaces, they are not invertible. However, $(S + T)v_k = v_k$, for $k = 1, \ldots, n$, so $S + T = I$, which is invertible. Thus the set of noninvertible operators on V with $n \geq 2$ is not closed under addition.

Chapter 4

Inner Products Spaces

Write a content brief for the book "Orthogonality (A Lively Introduction with Proofs)." The book covers orthonormal bases, orthogonal projections, Gram-Schmidt process, QR factorization, orthogonal transformations, and orthogonal matrices. It also includes inner-products.

Orthogonality is a vital topic in linear algebra and mathematics as a whole. It is often studied in the context of inner product spaces, where it plays a key role in many important results. In this book, we will give a lively introduction to orthogonality, with an emphasis on proofs and intuition. We will also discuss inner-products and their relationship to orthogonality. This book is aimed at students who are studying linear algebra or related topics. It will be a valuable resource for those who want to deepen their understanding of orthogonality and its many applications.

Orthonormal bases and orthogonal projections are two of the most important concepts in linear algebra. An orthonormal basis is a set of vectors that are all perpendicular to each other, and an orthogonal projection is a way of computing the projection of a vector onto an orthonormal basis.

In mathematics, an orthonormal basis is a special type of basis for a vector space. It consists of mutually orthogonal unit vectors that are also normalized, meaning that they have a length of 1. An orthogonal projection is a transformation of a vector onto another vector that is perpendicular to it. In other words, it is the process of projecting one vector onto another vector that is at right angles to it. Orthonormal bases and orthogonal projections are important in many areas of mathematics, including linear algebra and geometry.

Together, these concepts allow us to decompose any vector into a series of simple, linearly independent components. This can be helpful for many purposes, including data compression, error correction, and signal pro-

cessing. In this book, we will discuss these concepts in more detail and show how they can be used in practice.

Orthonormal bases and orthogonal projections are closely related concepts, and understanding both can be helpful in a variety of contexts.

The Gram-Schmidt process is a method for orthogonalizing a set of vectors. It is commonly used in numerical analysis and in particular, in the QR factorization of a matrix.

The Gram-Schmidt process begins with a set of linearly independent vectors $u_1, u_2, ..., u_n$ and produces a set of orthonormal vectors $v_1, v_2, ..., v_n$ that span the same space as the original vectors.

The v's are constructed iteratively as follows: For $k = 1, 2, ..., n$, let v_k be the vector $u_k - \text{proj}(v_1) - \text{proj}(v_2) - ... - \text{proj}(v_k) - 1$ where $\text{proj}(v_j)$ is the projection of u_k onto v_j.

Once the v's have been computed, they can be used to obtain the QR factorization of any matrix A as follows: $A = QR$ where Q is an $n \times n$ matrix whose columns are the v's and R is an upper triangular matrix.

The Gram-Schmidt process is sometimes also called the QR algorithm. The name comes from its use in computing the QR factorization of a matrix. However, it should be noted that the Gram-Schmidt process can be used to orthogonalize any set of vectors, not just those that arise from a matrix.

In general, the Gram-Schmidt process is a reliable QR factorization method when used with care.

In mathematics, an orthogonal transformation is a linear transformation that preserves the inner product of vectors. In other words, it preserves angles between vectors. If a transformation is both orthogonal and preserves length, then it is also an isometry. Orthogonal transformations are also known as rigid motions or rotations.

An orthogonal matrix is a square matrix whose columns and rows are orthogonal unit vectors. That is, the transpose of the matrix is equal to its inverse. Every orthogonal matrix has determinant $+1$ or -1, since the determinant of a matrix is equal to the product of its eigenvalues.

Orthogonal transformations arise naturally in many areas of mathematics and physics. For example, they are used to define rotations in space, and they can be used to diagonalize Hermitian matrices. In addition, they are used in many numerical algorithms, such as the QR factorization algorithm for solving linear least squares problems.

Orthogonal matrices have many interesting properties, including the fact that their eigenvectors are always orthogonal to one another. This makes

them particularly useful for many applications involving signal processing and image compression.

An inner-product is a mathematical operation that takes two vectors and produces a scalar result. The most common type of inner-product is the dot-product, which simply multiplies the corresponding components of the two vectors and sums the results.

The dot-product can be used to calculate the magnitude of a vector, as well as the angle between two vectors. More generally, an inner-product can be any symmetric bilinear map from a vector space to its underlying field.

Inner-products are often used in machine learning and statistics, where they can be used to measure similarity between data points.

In physics, inner-products are used to define Hermitian operators, which are important in quantum mechanics.

Orthogonality is a powerful mathematical tool that has many applications in physics, engineering, and numerical analysis. In this book, you'll learn how the Gram-Schmidt process can be used to orthogonalize a set of vectors, and how orthogonal matrices can be used to define rotations in space. We have also seen how inner-products can be used to measure similarity between vectors, and to define important operators in physics. With this book as your guide, you should now have a good understanding of the basics of orthogonality and its many uses.

4.1   Orthonormal Bases and Orthogonal Projections

The **norm** of a vector $\vec{v}$ in $\mathbb{R}^n$ is $||\vec{v}|| = \sqrt{\vec{v} \cdot \vec{v}}$. A vector $\vec{u}$ in $\mathbb{R}^n$ is called a **unit vector** if $||\vec{u}|| = 1$.

Example 4.1. If $\vec{v} \in \mathbb{R}^n$ and k is a scalar, then $||k\vec{v}|| = |k| \, ||\vec{v}||$, and if v is nonzero then $\vec{u} = \frac{1}{||\vec{v}||}\vec{v}$ is a unit vector. Since $||k\vec{v}||^2 = (k\vec{v}) \cdot (k\vec{v}) = k^2(\vec{v} \cdot \vec{v}) = k^2 \, ||\vec{v}||^2$, taking square roots provides $||k\vec{v}|| = |k| \, ||\vec{v}||$. If v is nonzero, then $\frac{1}{||\vec{v}||}$ is defined and so $||\vec{u}|| = \frac{1}{||\vec{v}||} \, ||\vec{v}|| = 1$ which follows by the first part.

Two vectors $\vec{v}$ and $\vec{w}$ in $\mathbb{R}^n$ are called **perpendicular** or **orthogonal** if $\vec{v} \cdot \vec{w} = 0$. The vectors $\vec{u}_1, \dots, \vec{u}_m$ in $\mathbb{R}^n$ are called **orthonormal** if they are all unit vectors and orthogonal to one another. A basis of $\mathbb{R}^n$ consisting only of orthonormal vectors is called an **orthonormal basis**.

Theorem 4.1 (Orthonormal Vectors). *Orthonormal vectors are linearly independent, and thus orthonormal vectors $\vec{u}_1, \dots, \vec{u}_n \in \mathbb{R}^n$ form a basis for $\mathbb{R}^n$.*

Proof. Suppose $(\vec{u}_1, \dots, \vec{u}_m)$ are orthonormal vectors in $\mathbb{R}^n$. To show linear independence suppose,

$$c_1\vec{u}_1 + \dots + c_m\vec{u}_m = \vec{0}$$

for some scalars $c_1, \dots, c_m$ in $\mathbb{R}$. Applying the dot product with $\vec{u}_i$,

$$(c_1\vec{u}_1 + \dots + c_m\vec{u}_m) \cdot \vec{u}_i = \vec{0} \cdot \vec{u}_i = 0.$$

Because the dot product is distributive, $c_1(\vec{u}_1 \cdot \vec{u}_i) + \dots + c_m(\vec{u}_m \cdot \vec{u}_i) = 0$. We know that $\vec{u}_i \cdot \vec{u}_i = 1$ and all other dot products are zero. Therefore, $c_i = 0$. Since this holds for all $i = 1, \dots, m$, it follows that $c_1 = \dots = c_m = 0$, and therefore, $(\vec{u}_1, \dots, \vec{u}_m)$ are linearly independent. The second part follows since n linearly independent vectors in $\mathbb{R}^n$ always forms a basis.

$\square$

Example 4.2. Find three examples of an orthonormal basis for a subspace they span. The vectors $\vec{e}_1, \dots, \vec{e}_m$ in $\mathbb{R}^n$ form an orthonormal basis of the subspace they span. For any scalar θ, the vectors $\begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix}$, $\begin{bmatrix} -\sin\theta \\ \cos\theta \end{bmatrix}$ form an orthonormal basis of $\mathbb{R}^2$. The vectors

$$\vec{u}_1 = \begin{bmatrix} 1/2 \\ 1/2 \\ 1/2 \\ 1/2 \end{bmatrix}, \quad \vec{u}_2 = \begin{bmatrix} 1/2 \\ 1/2 \\ -1/2 \\ -1/2 \end{bmatrix}, \quad \vec{u}_3 = \begin{bmatrix} 1/2 \\ -1/2 \\ 1/2 \\ -1/2 \end{bmatrix}$$

in $\mathbb{R}^4$ form an orthonormal basis of the subspace they span.

Let V be a subspace of $\mathbb{R}^n$. The **orthogonal complement** $V^\perp$ of V is the set of those vectors $\vec{x}$ in $\mathbb{R}^n$ that are orthogonal to all vectors in V namely,

$$V^\perp = \{\vec{x} \in \mathbb{R}^n \mid \vec{v} \cdot \vec{x} = 0, \text{ for all } \vec{v} \text{ in } V\}. \tag{4.1}$$

It is easy to verify that $V^\perp$ is always a subspace and that $(\mathbb{R}^n)^\perp = \{0\}$ and $\{0\}^\perp = \mathbb{R}^n$. Also notice that if $U_1 \subseteq U_2$ then $U_2^\perp \subseteq U_1^\perp$. If $\vec{x} \in V^\perp$ then $\vec{x}$ is said to be **perpendicular** to V. The vector $\vec{x}^\parallel$ in the following theorem is called the **orthogonal projection** of $\vec{x}$ on a subspace V of $\mathbb{R}^n$ and is denoted by $\text{proj}_V(\vec{x})$.

Theorem 4.2 (Orthogonal Projection).

- *If V is a subspace of $\mathbb{R}^n$ and $\vec{x} \in \mathbb{R}^n$, then $\vec{x} = \vec{x}^{\parallel} + \vec{x}^{\perp}$ where $\vec{x}^{\perp}$ is perpendicular to V, and this representation is unique.*
- *If V is a subspace of $\mathbb{R}^n$ with an orthonormal basis $\vec{u}_1, \dots, \vec{u}_m$, then*

$$\text{proj}_V(\vec{x}) := \vec{x}^{\parallel} = (\vec{u}_1 \cdot \vec{x})\vec{u}_1 + \cdots + (\vec{u}_m \cdot \vec{x})\vec{u}_m \qquad (4.2)$$

  for all $\vec{x}$ in $\mathbb{R}^n$.
- *Let $\vec{u}_1, \dots, \vec{u}_n$ be an orthonormal basis in $\mathbb{R}^n$, then*

$$\vec{x} = (\vec{u}_1 \cdot \vec{x})\vec{u}_1 + \cdots + (\vec{u}_n \cdot \vec{x})\vec{u}_n \qquad (4.3)$$

  for all $\vec{x}$ in $\mathbb{R}^n$.

Proof. The proof is left for the reader.

$\square$

Example 4.3. Find the orthogonal projection of $\begin{bmatrix} 49 \\ 49 \\ 49 \end{bmatrix}$ onto the subspace

of $\mathbb{R}^3$ spanned by $\begin{bmatrix} 2 \\ 3 \\ 6 \end{bmatrix}$ and $\begin{bmatrix} 3 \\ -6 \\ 2 \end{bmatrix}$. The two given vectors spanning the

subspace are orthogonal since $2(3) + 3(-6) + 6(2) = 0$, but they are not unit vectors since both have length 7. To obtain an orthonormal basis

$\vec{u}_1, \vec{u}_2$ of the subspace, we divide by 7: $\vec{u}_1 = \frac{1}{7}\begin{bmatrix} 2 \\ 3 \\ 6 \end{bmatrix}$ and $\vec{u}_2 = \frac{1}{7}\begin{bmatrix} 3 \\ -6 \\ 2 \end{bmatrix}$.

Now we can use (**??**) with $\vec{x} = \begin{bmatrix} 49 \\ 49 \\ 49 \end{bmatrix}$. Then

$$\text{proj}_V(\vec{x}) = (\vec{u}_1 \cdot \vec{x})\vec{u}_1 + (u_2 \cdot \vec{x})\vec{u}_2 = 11\begin{bmatrix} 2 \\ 3 \\ 6 \end{bmatrix} + (-1)\begin{bmatrix} 3 \\ -6 \\ 2 \end{bmatrix} = \begin{bmatrix} 19 \\ 39 \\ 64 \end{bmatrix}.$$

Example 4.4. Find the coordinates of the vector $\vec{x} = \begin{bmatrix} 4 \\ 5 \\ 6 \\ 7 \end{bmatrix}$ with respect

to the orthonormal basis

$$\vec{u}_1 = \begin{bmatrix} 1/2 \\ 1/2 \\ 1/2 \\ 1/2 \end{bmatrix}, \quad \vec{u}_2 = \begin{bmatrix} 1/2 \\ 1/2 \\ -1/2 \\ -1/2 \end{bmatrix}, \quad \vec{u}_3 = \begin{bmatrix} 1/2 \\ -1/2 \\ 1/2 \\ -1/2 \end{bmatrix}, \quad \vec{u}_4 = \begin{bmatrix} 1/2 \\ -1/2 \\ -1/2 \\ 1/2 \end{bmatrix}.$$

Normally to find the coordinates of $\vec{x}$ we would solve the system

$$\begin{bmatrix} 4 \\ 5 \\ 6 \\ 7 \end{bmatrix} = c_1 \begin{bmatrix} 1/2 \\ 1/2 \\ 1/2 \\ 1/2 \end{bmatrix} + c_2 \begin{bmatrix} 1/2 \\ 1/2 \\ -1/2 \\ -1/2 \end{bmatrix} + c_3 \begin{bmatrix} 1/2 \\ -1/2 \\ 1/2 \\ -1/2 \end{bmatrix} + c_4 \begin{bmatrix} 1/2 \\ -1/2 \\ -1/2 \\ 1/2 \end{bmatrix}$$

for c_1, c_2, c_3, c_4. However we can use (**??**) instead:

$$c_1 = u_1 \cdot \vec{x} = \begin{bmatrix} 1/2 \\ 1/2 \\ 1/2 \\ 1/2 \end{bmatrix} \cdot \begin{bmatrix} 4 \\ 5 \\ 6 \\ 7 \end{bmatrix} = 11,$$

$$c_2 = u_2 \cdot \vec{x} = \begin{bmatrix} 1/2 \\ 1/2 \\ -1/2 \\ -1/2 \end{bmatrix} \cdot \begin{bmatrix} 4 \\ 5 \\ 6 \\ 7 \end{bmatrix} = -2,$$

$$c_3 = u_3 \cdot \vec{x} = \begin{bmatrix} 1/2 \\ -1/2 \\ 1/2 \\ -1/2 \end{bmatrix} \cdot \begin{bmatrix} 4 \\ 5 \\ 6 \\ 7 \end{bmatrix} = -1,$$

$$c_4 = u_4 \cdot \vec{x} = \begin{bmatrix} 1/2 \\ -1/2 \\ -1/2 \\ 1/2 \end{bmatrix} \cdot \begin{bmatrix} 4 \\ 5 \\ 6 \\ 7 \end{bmatrix} = 0.$$

Therefore the $\mathcal{B}$-coordinate vector of $\vec{x}$ is $\begin{bmatrix} 11 \\ -2 \\ -1 \\ 0 \end{bmatrix}$.

Theorem 4.3 (Properties of the Orthogonal Complement). *Let V be a subspace of $\mathbb{R}^n$, then*

- $\text{proj}_V(\vec{x})$ *is a linear transformation* $\mathbb{R}^n \to V$ *with kernel* $V^\perp$,
- $V \cap V^\perp = \{\vec{0}\}$,
- $\dim V + \dim V^\perp = n$, *and*
- $(V^\perp)^\perp = V$.

Proof. The proof of each part follows.

- To prove the linearity of $T(x) := \text{proj}_V(\vec{x})$ we will use the definition of a projection: $T(\vec{x})$ is in V, and $\vec{x} - T(\vec{x})$ is in $V^\perp$. To show $T(\vec{x} + \vec{y}) = T(\vec{x}) + T(\vec{y})$, notice $T(\vec{x}) + T(\vec{y})$ is in V (since V is a subspace), and $\vec{x} + \vec{y} - (T(\vec{x}) + T(\vec{y})) = (\vec{x} - T(\vec{x})) + (\vec{y} - T(\vec{y}))$ is in $V^\perp$

(since $V^\perp$ is a subspace). To show that $T(k\vec{x}) = kT(\vec{x})$, note that $kT(\vec{x})$ is in V (since V is a subspace) and $k\vec{x} - kT(\vec{x}) = k(\vec{x} - T(\vec{x}))$ is in $V^\perp$ (since $V^\perp$ is a subspace).

- Since $\{\vec{0}\} \subseteq V$ and $\{\vec{0}\} \subseteq V^\perp$, $\{\vec{0}\} \subseteq V \cap V^\perp$. If a vector $\vec{x}$ is in V as well as in $V^\perp$, then $\vec{x}$ is orthogonal to itself: $\vec{x} \cdot \vec{x} = ||x||^2 = 0$, so that $\vec{x}$ must equal $\vec{0}$ which shows $V \subseteq \{\vec{0}\}$. Therefore, $V \cap V^\perp = \{\vec{0}\}$.

- Apply the Rank-Nullity Theorem to the linear transformation $T(\vec{x}) = \text{proj}_V(\vec{x})$ yielding $n = \dim\mathbb{R}^n = \dim\text{image}T + \dim\ker T = \dim V + \dim V^\perp$.

- Let $\vec{v} \in V$. Then $\vec{v} \cdot \vec{x} = 0$ for all $\vec{x}$ in $V^\perp$. Since $(V^\perp)^\perp$ contains all vectors $\vec{y}$ such that $\vec{y} \cdot \vec{x} = 0$, $\vec{v}$ is in $(V^\perp)^\perp$. So V is a subspace of $(V^\perp)^\perp$. Using (iii) with V ($n = \dim V + \dim V^\perp$) and again with $V^\perp$ ($\dim V^\perp + \dim(V^\perp)^\perp$) yielding $\dim V = \dim(V^\perp)^\perp$; and since V is a subspace of $(V^\perp)^\perp$ it follows that $V = (V^\perp)^\perp$.

□

Example 4.5. Find a basis for $W^\perp$, where $W = \text{span}\left(\begin{bmatrix}1\\2\\3\\4\end{bmatrix}, \begin{bmatrix}5\\6\\7\\8\end{bmatrix}\right)$. The orthogonal complement $W^\perp$ of W consists of the vectors $\vec{x}$ in $\mathbb{R}^4$ such that

$$\begin{bmatrix}x_1\\x_2\\x_3\\x_4\end{bmatrix} \cdot \begin{bmatrix}1\\2\\3\\4\end{bmatrix} = 0 \qquad \text{and} \qquad \begin{bmatrix}x_1\\x_2\\x_3\\x_4\end{bmatrix} \cdot \begin{bmatrix}5\\6\\7\\8\end{bmatrix} = 0.$$

Finding these vectors amounts to solving the system

$$\begin{cases} x_1 + 2x_2 + 3x_3 + 4x_4 &= 0 \\ 5x_1 + 6x_2 + 7x_3 + 8x_4 &= 0 \end{cases}$$

The solutions are of the form

$$\begin{bmatrix}x_1\\x_2\\x_3\\x_4\end{bmatrix} = \begin{bmatrix}s + 2t\\-2s - 3t\\s\\t\end{bmatrix} = s\begin{bmatrix}1\\-2\\1\\0\end{bmatrix} + t\begin{bmatrix}2\\-3\\0\\1\end{bmatrix}.$$

The two vectors on the right form a basis of $W^\perp$.

Theorem 4.4. *Let $\vec{x}, \vec{y} \in \mathbb{R}^n$. Then*

- (**Pythagorean Theorem**)

$$||\vec{x} + \vec{y}||^2 = ||\vec{x}||^2 + ||\vec{y}||^2$$

  holds if and only if $\vec{x} \perp \vec{y}$,
- (**Cauchy-Schwarz**)

$$|\vec{x} \cdot \vec{y}| \leq ||\vec{x}|| \, ||\vec{y}||$$

  where equality holds if and only if $\vec{x}$ and $\vec{y}$ are parallel,
- (**Law of Cosines**) the angle θ between $\vec{x}$ and $\vec{y}$ is defined as

$$\theta = \arccos \frac{\vec{x} \cdot \vec{y}}{||\vec{x}|| \, ||\vec{y}||},$$

- (**Triangular Inequality**)

$$||\vec{x} + \vec{y}|| \leq ||\vec{x}|| + ||\vec{y}||.$$

Proof. The proof of each part follows.

- The verification is straightforward:

$$||\vec{x} + \vec{y}||^2 = (\vec{x} + \vec{y}) \cdot (\vec{x} + \vec{y}) = \vec{x} \cdot \vec{x} + 2(\vec{x} \cdot \vec{y}) + \vec{y} \cdot \vec{y}$$
$$= ||\vec{x}||^2 + 2(\vec{x} \cdot \vec{y}) + ||\vec{y}||^2 = ||\vec{x}||^2 + ||\vec{y}||^2$$

  where the last equality holds if and only if $\vec{x} \cdot \vec{y} = 0$.
- Let V be a one-dimensional subspace of $\mathbb{R}^n$ spanned by a nonzero vector $\vec{y}$. Let $\vec{u} = \frac{1}{||\vec{y}||} \vec{y}$. Then

$$||\vec{x}|| \geq \left|\left|\mathrm{proj}_V(\vec{x})\right|\right| = ||(\vec{x} \cdot \vec{u})\vec{u}|| = |\vec{x} \cdot \vec{u}| = \left|\vec{x} \cdot \left(\frac{1}{||y||}\vec{y}\right)\right| = \frac{1}{||y||}|\vec{x} \cdot \vec{y}|$$

  multiplying by $||\vec{y}||$, yields $|\vec{x} \cdot \vec{y}| \leq ||\vec{x}|| \, ||\vec{y}||$. Notice that $||\vec{x}|| \geq \left|\left|\mathrm{proj}_V(\vec{x})\right|\right|$ holds by applying the Pythagorean theorem to $\vec{x} = \vec{x}^{\parallel} + \vec{x}^{\perp}$ with $\vec{x}^{\perp} \cdot \vec{x}^{\parallel} = 0$ so that $||\vec{x}||^2 = \left|\left|\mathrm{proj}_V(\vec{x})\right|\right|^2 + \left|\left|\vec{x}^{\perp}\right|\right|^2$ which leads to $\left|\left|\mathrm{proj}_V \vec{x}\right|\right| \leq ||\vec{x}||$.
- We have to make sure that (**??**) is defined, that is θ is between -1 and 1, or equivalently,

$$\left|\frac{\vec{x} \cdot \vec{y}}{||x|| \, ||y||}\right| \leq 1.$$

  But this follows from the Cauchy-Schwarz inequality.
- Using the Cauchy-Schwarz inequality, the verification is straightforward,

$$||\vec{x} + \vec{y}||^2 = (\vec{x} + \vec{y}) \cdot (\vec{x} + \vec{y}) = ||\vec{x}||^2 + ||\vec{y}||^2 + 2(\vec{x} \cdot \vec{y})$$
$$\leq ||\vec{x}||^2 + ||\vec{y}||^2 + 2\,||\vec{x}|| \, ||\vec{y}|| = (||\vec{x}|| + ||\vec{y}||)^2$$

  Taking the square root of both sides yields $||\vec{x} + \vec{y}|| \leq ||\vec{x}|| + ||\vec{y}||$.

□

Example 4.6. Determine whether the angle between the vectors $\vec{u} = \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix}$, $\vec{v} = \begin{bmatrix} 2 \\ -8 \\ 5 \end{bmatrix}$ is a right angle using the Pythagorean Theorem. Since

$||\vec{u}|| = \sqrt{2^2 + 3^2 + 4^2} = \sqrt{29}$ and $||\vec{v}|| = \sqrt{2^2 + (-8)^2 + 5^2} = \sqrt{93}$. Then

$$||\vec{u} + \vec{v}||^2 = \left|\left| \begin{bmatrix} 4 \\ -5 \\ 9 \end{bmatrix} \right|\right|^2 = 122 = 29 + 93 = ||\vec{u}||^2 + ||\vec{v}||^2$$

shows $\vec{u} \perp \vec{v}$.

Example 4.7. Consider the vectors $\vec{u} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$ and $\vec{v} = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$ in $\mathbb{R}^n$. For

$n = 2, 3, 4$, find the angle θ between $\vec{u}$ and $\vec{v}$. Then find the limit of θ as n approaches infinity. For any possible value of n,

$$\theta_n = \arccos \frac{\vec{u} \cdot \vec{v}}{||\vec{u}|| \, ||\vec{v}||} = \arccos \frac{1}{\sqrt{n}}.$$

Then

$$\left[\theta_2 = \arccos \tfrac{1}{\sqrt{2}} = \tfrac{\pi}{4}, \qquad \theta_3 = \arccos \tfrac{1}{\sqrt{3}} = \tfrac{\pi}{4} \sim 0.955 \text{ rads}, \qquad \theta_4 = \arccos \tfrac{1}{\sqrt{4}} = \tfrac{\pi}{3}. \right]$$

Since $y = \arccos(x)$ is a continuous function,

$$\lim_{x \mapsto \infty} \theta_n = \lim_{x \mapsto \infty} \arccos \left(\frac{1}{\sqrt{n}} \right) = \arccos \left(\lim_{x \mapsto \infty} \frac{1}{\sqrt{n}} \right) = \arccos(0) = \frac{\pi}{2}.$$

4.2 Gram-Schmidt Process and QR Factorization

The Gram-Schmidt process represents a change of basis from a basis $\mathcal{B} = (\vec{v}_1, \vec{v}_2, ..., \vec{v}_m)$ of a subspace V of $\mathbb{R}^n$ to an orthonormal basis $\mathcal{U} = (\vec{u}_1, \vec{u}_2, ..., \vec{u}_m)$ of V; it is most sufficiently described in terms of the change of basis matrix R from $\mathcal{B}$ to $\mathcal{U}$ via,

$$M := \begin{bmatrix} \vec{v}_1 & \vec{v}_2 & \cdots & \vec{v}_m \end{bmatrix} = \begin{bmatrix} \vec{u}_1 & \vec{u}_2 & \cdots & \vec{u}_m \end{bmatrix} R =: QR \qquad (4.4)$$

The QR factorization is an effective way to organize and record the work performed in the Gram-Schmidt process; it is also useful for many computational and theoretical purposes.

Theorem 4.5 (Gram-Schmidt Process). *Let* $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_m$ *be a basis of a subspace* V *of* $\mathbb{R}^n$. *Then*

$$\vec{u}_1 = \frac{1}{||\vec{v}_1||}\vec{v}_1, \quad \vec{u}_2 = \frac{1}{||\vec{v}_2^{\perp}||}\vec{v}_2^{\perp}, \quad \ldots, \quad \vec{u}_m = \frac{1}{||\vec{v}_m^{\perp}||}\vec{v}_m^{\perp} \tag{4.5}$$

is an orthonormal basis of V *where*

$$\vec{v}_j^{\perp} = \vec{v}_j - \vec{v}_j^{\parallel} = \vec{v}_j - (\vec{u}_1 \cdot \vec{v}_j)\vec{u}_1 - (\vec{u}_2 \cdot \vec{v}_j)\vec{u}_2 - \cdots - (\vec{u}_{j-1} \cdot \vec{v}_j)\vec{u}_{j-1}. \tag{4.6}$$

Proof. For each j, we resolve the vectors $\vec{v}_j$ into its components parallel and perpendicular to the span of the preceding vectors $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_{j-1}$:

$$\vec{v}_j = \vec{v}_j^{\parallel} + \vec{v}_j^{\perp} \qquad \text{with respect to span}(\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_{j-1}).$$

Then use **??**.

$\square$

Example 4.8. Perform the Gram-Schmidt process on the vectors

$$\vec{v}_1 = \begin{bmatrix} 4 \\ 0 \\ 3 \end{bmatrix}, \qquad \vec{v}_2 = \begin{bmatrix} 25 \\ 0 \\ -25 \end{bmatrix}, \qquad \vec{v}_3 = \begin{bmatrix} 0 \\ -2 \\ 0 \end{bmatrix}.$$

By Theorem (**??**), we determine $\vec{u}_1, \vec{u}_2, \vec{u}_3$ as follows:

$$\vec{u}_1 = \begin{bmatrix} 4/5 \\ 0 \\ 3/5 \end{bmatrix}, \qquad \vec{u}_2 = \frac{\vec{v}_2^{\perp}}{||\vec{v}_2^{\perp}||} = \frac{\vec{v}_2 - (\vec{u}_1 \cdot \vec{v}_2)\,\vec{u}_1}{||\vec{v}_2 - (\vec{u}_1 \cdot \vec{v}_2)\,\vec{u}_1||} = \begin{bmatrix} 3/5 \\ 0 \\ -4/5 \end{bmatrix},$$

and

$$\vec{u}_3 = \frac{v_3^{\perp}}{||v_3^{\perp}||} = \frac{\vec{v}_3 - (\vec{u}_1 \cdot \vec{v}_3)\,\vec{u}_1 - (\vec{u}_2 \cdot \vec{v}_3)\,\vec{u}_2}{||\vec{v}_3 - (\vec{u}_1 \cdot \vec{v}_3)\,\vec{u}_1 - (\vec{u}_2 \cdot \vec{v}_3)\,\vec{u}_2||} = \begin{bmatrix} 0 \\ -1 \\ 0 \end{bmatrix}.$$

Therefore

$$\left(\begin{bmatrix} 4/5 \\ 0 \\ 3/5 \end{bmatrix}, \begin{bmatrix} 3/5 \\ 0 \\ -4/5 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \\ 0 \end{bmatrix} \right).$$

is an orthonormal basis for span$(\vec{v}_1, \vec{v}_2, \vec{v}_3)$.

::: {#thm- } [QR Factorization] Let M be an $n \times m$ matrix with linearly independent columns $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_m$. Then there exists an $n \times m$ matrix Q whose columns $\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_m$ are orthonormal and an upper triangular matrix R with positive diagonal entries such that $M = QR$; and this representation is unique. Furthermore, the entries r_{ij} of R are given by

- $r_{11} = ||\vec{v}_1||$,
- $r_{jj} = ||\vec{v}_j^{\perp}||$ (for $j = 2, \ldots, m$), and
- $r_{ij} = \vec{u}_i \cdot \vec{v}_j$ (for $i < j$). ::::

Proof. The proof is left for the reader.

$\square$

Example 4.9. Find the QR factorization of the matrix and display the commutative diagram.

$$M = \begin{bmatrix} 4 & 25 & 0 \\ 0 & 0 & -2 \\ 3 & -25 & 0 \end{bmatrix}$$

Let

$$\vec{v}_1 = \begin{bmatrix} 4 \\ 0 \\ 3 \end{bmatrix}, \quad \vec{v}_2 = \begin{bmatrix} 25 \\ 0 \\ -25 \end{bmatrix}, \quad \vec{v}_3 = \begin{bmatrix} 0 \\ -2 \\ 0 \end{bmatrix}$$

then (as determined above) an orthonormal basis for the column vectors of M is

$$(\vec{u}_1, \vec{u}_2, \vec{u}_3) = \left(\begin{bmatrix} 4/5 \\ 0 \\ 3/5 \end{bmatrix}, \begin{bmatrix} 3/5 \\ 0 \\ -4/5 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \\ 0 \end{bmatrix} \right).$$

Determining the entries of R (also as determined above):

$$r_{11} = ||\vec{v}_1|| = 5, \quad r_{22} = ||\vec{v}_2^{\perp}|| = 35, \quad r_{33} = ||\vec{v}_3^{\perp}|| = 2$$

$$r_{12} = \vec{u}_1 \cdot \vec{v}_2 = 5, \quad r_{13} = \vec{u}_1 \cdot \vec{v}_3 = 0, \quad r_{23} = \vec{u}_2 \cdot \vec{v}_3 = 0$$

and therefore, the QR-factorization of M is

$$M = \begin{bmatrix} 4 & 25 & 0 \\ 0 & 0 & -2 \\ 3 & -25 & 0 \end{bmatrix} = \begin{bmatrix} 4/5 & 3/5 & 0 \\ 0 & 0 & -1 \\ 3/5 & -4/5 & 0 \end{bmatrix} \begin{bmatrix} 5 & 5 & 0 \\ 0 & 35 & 0 \\ 0 & 0 & 2 \end{bmatrix} = QR.$$

Example 4.10.

(a) Perform the Gram-Schmidt process on the vectors

$$\vec{v}_1 = \begin{bmatrix} 1 \\ 7 \\ 1 \\ 7 \end{bmatrix}, \quad \vec{v}_2 = \begin{bmatrix} 0 \\ 7 \\ 2 \\ 7 \end{bmatrix}, \quad \vec{v}_3 = \begin{bmatrix} 1 \\ 8 \\ 1 \\ 6 \end{bmatrix}.$$

(b) Find the QR factorization of the matrix $M = \begin{bmatrix} \vec{v}_1 & \vec{v}_2 & \vec{v}_3 \end{bmatrix}$.

By Theorem (??), we determine $\vec{u}_1, \vec{u}_2, \vec{u}_3$ as follows:

$$\vec{u}_1 = \begin{bmatrix} 1/10 \\ 7/10 \\ 1/10 \\ 7/10 \end{bmatrix}, \vec{u}_2 = \frac{\vec{v}_2^{\perp}}{||\vec{v}_2^{\perp}||} = \frac{\vec{v}_2 - (\vec{u}_1 \cdot \vec{v}_2)\,\vec{u}_1}{||\vec{v}_2 - (\vec{u}_1 \cdot \vec{v}_2)\,\vec{u}_1||} = \begin{bmatrix} -1/\sqrt{2} \\ 0 \\ 1/\sqrt{2} \\ 0 \end{bmatrix},$$

and

$$\vec{u}_3 = \frac{\vec{v}_3^{\perp}}{||\vec{v}_3^{\perp}||} = \frac{\vec{v}_3 - (\vec{u}_1 \cdot \vec{v}_3)\,\vec{u}_1 - (\vec{u}_2 \cdot \vec{v}_3)\,\vec{u}_2}{||\vec{v}_3 - (\vec{u}_1 \cdot \vec{v}_3)\,\vec{u}_1 - (\vec{u}_2 \cdot \vec{v}_3)\,\vec{u}_2||} = \begin{bmatrix} 0 \\ 1/\sqrt{2} \\ 0 \\ -1/\sqrt{2} \end{bmatrix}.$$

Therefore an orthonormal basis for $\mathrm{span}(\vec{v}_1, \vec{v}_2, \vec{v}_3)$ is

$$\left(\begin{bmatrix} 1/10 \\ 7/10 \\ 1/10 \\ 7/10 \end{bmatrix}, \begin{bmatrix} -1/\sqrt{2} \\ 0 \\ 1/\sqrt{2} \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1/\sqrt{2} \\ 0 \\ -1/\sqrt{2} \end{bmatrix} \right).$$

Determining the entries of R (also as determined above):

$$r_{11} = ||\vec{v}_1|| = 10, \quad r_{22} = ||\vec{v}_2^{\perp}|| = \sqrt{2}, \quad r_{33} = ||\vec{v}_3^{\perp}|| = \sqrt{2}$$

$$r_{12} = \vec{u}_1 \cdot \vec{v}_2 = 10, \quad r_{13} = \vec{u}_1 \cdot \vec{v}_3 = 10, \quad r_{23} = \vec{u}_2 \cdot \vec{v}_3 = 0$$

and therefore, the QR-factorization of M is

$$M = \begin{bmatrix} 1 & 0 & 1 \\ 7 & 7 & 8 \\ 1 & 2 & 1 \\ 7 & 7 & 6 \end{bmatrix} = \begin{bmatrix} 1/10 & -1/\sqrt{2} & 0 \\ 7/10 & 0 & 1/\sqrt{2} \\ 1/10 & 1/\sqrt{2} & 0 \\ 7/10 & 0 & -1/\sqrt{2} \end{bmatrix} \begin{bmatrix} 10 & 10 & 10 \\ 0 & \sqrt{2} & 0 \\ 0 & 0 & \sqrt{2} \end{bmatrix} = QR.$$

4.3 Orthogonal Transformations and Orthogonal Matrices

A linear transformation T from $\mathbb{R}^n$ to $\mathbb{R}^n$ is called an **orthogonal transformation** if it preserves the length of vectors: $||T(\vec{x})|| = ||x||$ for all $\vec{x} \in \mathbb{R}^n$. If $T(\vec{x}) = A\vec{x}$ is an orthogonal transformation, we say A is an **orthogonal matrix** .

Lemma 4.1. *Let T be an orthogonal transformation from $\mathbb{R}^n$ to $\mathbb{R}^n$. If $\vec{v}, \vec{w} \in \mathbb{R}^n$ are orthogonal, then $T(\vec{v}), T(\vec{w}) \in \mathbb{R}^n$ are orthogonal.*

Proof. We want to show $T(\vec{v}), T(\vec{w})$ are orthogonal, and by the Pythagorean theorem, we have to show

$$||T(\vec{v}) + T(\vec{w})||^2 = ||T(\vec{v})||^2 + ||T(\vec{w})||^2 .$$

This equality follows

$$||T(\vec{v}) + T(\vec{w})||^2 = ||T(\vec{v} + \vec{w})||^2 = ||\vec{v} + \vec{w}||^2$$
$$= ||\vec{v}||^2 + ||\vec{w}||^2 = ||T(\vec{v})||^2 + ||T(\vec{w})||^2$$

since T is linear, orthogonal and that $\vec{v}, \vec{w}$ are orthogonal, respectively.

$\square$

Theorem 4.6. *A linear transformation T from $\mathbb{R}^n$ to $\mathbb{R}^n$ is orthogonal if and only if the vectors $T(\vec{e}_1)$, ..., $T(\vec{e}_n)$ form an orthonormal basis.*

Proof. If T is an orthogonal transformation, then by definition, the $T(\vec{e}_i)$ are unit vectors, and also, by **??** they are orthogonal. Therefore, $T(\vec{e}_1)$, ..., $T(\vec{e}_n)$ form an orthonormal basis.

Conversely, suppose $T(\vec{e}_1)$, ..., $T(\vec{e}_n)$ form an orthonormal basis. Consider a vector $\vec{x} = x_1\vec{e}_1 + \cdots + x_n\vec{e}_n$. Then

$$||T(\vec{x})||^2 = ||T(x_1\vec{e}_1 + \cdots + x_n\vec{e}_n)||^2 = ||x_1T(\vec{e}_1) + \cdots + x_nT(\vec{e}_n)||^2$$
$$= ||x_1T(\vec{e}_1)||^2 + \cdots + ||x_nT(\vec{e}_n)||^2 = x_1^2 + \cdots + x_n^2 = ||x||^2 .$$

Taking the square root of both sides shows that T preserves lengths and therefore, T is an orthogonal transformation.

$\square$

Corollary 4.1. *An $n \times n$ matrix A is orthogonal if and only if its columns form an orthonormal basis.*

Proof. The proof is left for the reader.

$\square$

The **transpose** A^T of an $n \times n$ matrix A is the $n \times n$ matrix whose ij-th entry is the ji-th entry of A. We say that a square matrix A is **symmetric** if $A^T = A$, and A is called **skew-symmetric** if $A^T = -A$.

Theorem 4.7 (Orthogonal and Transpose Properties)**.**

- *The product of two orthogonal $n \times n$ matrices is orthogonal.*
- *The inverse of an orthogonal matrix is orthogonal.*
- *If the products $(AB)^T$ and $B^T A^T$ are defined then they are equal.*
- *If A is invertible then so is A^T, and $(A^T)^{-1} = (A^{-1})^T$.*
- *For any matrix A, $rank(A) = rank(A^T)$.*
- *If $\vec{v}$ and $\vec{w}$ are two column vectors in $\mathbb{R}^n$, then $\vec{v} \cdot \vec{w} = \vec{v}^T \vec{w}$.*
- *The $n \times n$ matrix A is orthogonal if and only if $A^{-1} = A^T$.*

Proof. The proof of each part follows.

- Suppose A and B are orthogonal matrices, then AB is an orthogonal matrix since $T(\vec{x}) = AB\vec{x}$ preserves length because $||T(\vec{x})|| = ||AB\vec{x}|| = ||A(B\vec{x})|| = ||B\vec{x}|| = ||\vec{x}||$.
- Suppose A is an orthogonal matrix, then A^{-1} is orthogonal an matrix since $T(\vec{x}) = A^{-1}\vec{x}$ preserves length because $||A^{-1}\vec{x}|| = ||A(A^{-1}\vec{x})|| = ||\vec{x}||$.
- We will compare entries in the matrices $(AB)^T$ and $B^T A^T$ as follows:

$$\begin{aligned} ij\text{-th entry of } (AB)^T &= ji\text{-th entry of } AB \\ &= (j\text{-th row of } A) \cdot (i\text{-th column of } B) \end{aligned}$$

$$\begin{aligned} ij\text{-th entry of } B^T A^T &= (i\text{-th row of } B^T) \cdot (j\text{th column of } A^T) \\ &= (i\text{-th column of } B) \cdot (j\text{-th row of } A) \\ &= (j\text{-th row of } A) \cdot (i\text{-th column of } B). \end{aligned}$$

Therefore, the ij-th entry of $(AB)^T$ is the same of the ij-th entry of $B^T A^T$.

- Suppose A is invertible, then $AA^{-1} = I_n$. Taking the transpose of both sides along with (iii) it yields, $(AA^{-1})^T = (A^{-1})^T A^T = I_n$. Thus A^T is invertible and since inverses are unique, it follows $(A^T)^{-1} = (A^{-1})^T$.
- Exercise.
- If $\vec{v} = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}$ and $\vec{w} = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}$, then

$$\vec{v} \cdot \vec{w} = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} \cdot \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix} = a_1 b_1 + \cdots + a_n b_n = \begin{bmatrix} a_1 & \cdots & a_n \end{bmatrix} \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix} = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}^T \vec{w} = \vec{v}^T \vec{u}$$

- Let's write A in terms of its columns: $A = \begin{bmatrix} \vec{v}_1 & \cdots & \vec{v}_n \end{bmatrix}$. Then

$$A^T A = \begin{bmatrix} \vec{v}_1^T \\ \vdots \\ \vec{v}_n^T \end{bmatrix} \begin{bmatrix} \vec{v}_1 & \cdots & \vec{v}_n \end{bmatrix} = \begin{bmatrix} \vec{v}_1 \cdot \vec{v}_1 & & \vec{v}_1 \cdot \vec{v}_n \\ \vdots & \cdots & \vdots \\ \vec{v}_n \cdot \vec{v}_1 & & \vec{v}_n \cdot \vec{v}_n \end{bmatrix}. \qquad (4.7)$$

Now A is orthogonal, by **??**, if and only if A has orthonormal columns, meaning A is orthogonal if and only if $A^T A = I_n$ by 4.7. Therefore, A is orthogonal if and only if $A^{-1} = A^T$.

□

Theorem 4.8 (Orthogonal Projection Matrix)**.**

- *Let V be a subspace of $\mathbb{R}^n$ with orthonormal basis $\vec{u}_1$, ..., $\vec{u}_m$. The matrix of the orthogonal projection onto V is QQ^T where $Q = \begin{bmatrix} \vec{u}_1 & \cdots & \vec{u}_m \end{bmatrix}$.*
- *Let V be a subspace of $\mathbb{R}^n$ with basis $\vec{v}_1, \ldots, \vec{v}_m$ and let $A = \begin{bmatrix} \vec{v}_1 & \cdots \vec{v}_m \end{bmatrix}$, then the orthogonal projection matrix onto V is $A(A^T A)^{-1} A^T$.*

Proof. The proof of each part follows.

- Since $\vec{u}_1$, ..., $\vec{u}_m$ is an orthonormal basis of V we can, by **??**, write,

$$\text{proj}_V(\vec{x}) = (\vec{u}_1 \cdot \vec{x})\vec{u}_1 + \cdots + (\vec{u}_m \cdot \vec{x})\vec{u}_m = \vec{u}_1 \vec{u}_1^T \vec{x} + \cdots + \vec{u}_m \vec{u}_m^T \vec{x}$$

$$= (\vec{u}_1 \vec{u}_1^T + \cdots + \vec{u}_m \vec{u}_m^T)\vec{x} = \begin{bmatrix} \vec{u}_1 & \cdots & \vec{u}_m \end{bmatrix} \begin{bmatrix} \vec{u}_1^T \\ \vdots \\ \vec{u}_m^T \end{bmatrix} \vec{x} = QQ^T \vec{x}.$$

- Since $\vec{v}_1, \ldots, \vec{v}_m$ form a basis of V, there exists unique scalars $c_1, \ldots, c_m$ such that $\text{proj}_V(\vec{x}) = c_1 \vec{v}_1 + \cdots + c_m \vec{v}_m$. Since $A = \begin{bmatrix} \vec{v}_1 & \cdots & \vec{v}_m \end{bmatrix}$ we can write $\text{proj}_V(\vec{x}) = A\vec{c}$. Consider the system $A^T A \vec{c} = A^T \vec{x}$ where $A^T A$ is the coefficient matrix and $\vec{c}$ is the unknown. Since $\vec{c}$ is the coordinate vector of $\text{proj}_V(\vec{x})$ with respect to the basis $(v_1, \ldots, v_m)$, the system has a unique solution. Thus, $A^T A$ must be invertible, and so we can solve for $\vec{c}$, namely $\vec{c} = (A^T A)^{-1} A^T \vec{x}$. Therefore, $\text{proj}_V(\vec{x}) = A\vec{c} = A(A^T A)^{-1} A^T$ as desired. Notice it suffices to consider the system $A^T A \vec{c} = A^T \vec{x}$, or equivalently $A^T(\vec{x} - A\vec{c}) = \vec{0}$, because

$$A^T(\vec{x} - A\vec{c}) = A^T(\vec{x} - c_1\vec{v}_1 - \cdots - c_m\vec{v}_m)$$

is the vector whose ith component is

$$(\vec{v}_i)^T(\vec{x} - c_1\vec{v}_1 - \cdots - c_m\vec{v}_m) = \vec{v}_i \cdot (\vec{x} - c_1\vec{v}_1 - \cdots - c_m\vec{v}_m)$$

which we know to be zero since $\vec{x} - \text{proj}_V(\vec{x})$ is orthogonal to V.

□

Example 4.11. Is there an orthogonal transformation T from $\mathbb{R}^3$ to $\mathbb{R}^3$ such that

$$T \begin{bmatrix} 2 \\ 3 \\ 0 \end{bmatrix} = \begin{bmatrix} 3 \\ 0 \\ 2 \end{bmatrix} \qquad \text{and} \qquad T \begin{bmatrix} -3 \\ 2 \\ 0 \end{bmatrix} = \begin{bmatrix} 2 \\ -3 \\ 0 \end{bmatrix} ?$$

No, since the vectors $\begin{bmatrix} 2 \\ 3 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} -3 \\ 2 \\ 0 \end{bmatrix}$ are orthogonal, whereas $\begin{bmatrix} 3 \\ 0 \\ 2 \end{bmatrix}$ and

$\begin{bmatrix} 2 \\ -3 \\ 0 \end{bmatrix}$ are not, by **??**.

Example 4.12. Find an orthogonal transformation T from $\mathbb{R}^3$ to $\mathbb{R}^3$ such that

$$T \begin{bmatrix} 2/3 \\ 2/3 \\ 1/3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

Let's think about the inverse of T first. The inverse of T, if it exists,

must satisfy $T^{-1}(\vec{e}_3) = \begin{bmatrix} 2/3 \\ 2/3 \\ 1/3 \end{bmatrix} = \vec{v}_3$. Furthermore, the vectors $\vec{v}_1, \vec{v}_2, \vec{v}_3$

must form an orthonormal basis of $\mathbb{R}^3$ where $T^{-1}\vec{x} = \begin{bmatrix} \vec{v}_1 & \vec{v}_2 & \vec{v}_3 \end{bmatrix}\vec{x}$. We require a vector $\vec{v}_1$ with $\vec{v}_1 \cdot \vec{v}_3 = 0$ and $||\vec{v}_1|| = 1$. By inspection, we find

$\vec{v}_1 = \begin{bmatrix} -2/3 \\ 1/3 \\ 2/3 \end{bmatrix}$. Then

$$\vec{v}_2 = \vec{v}_1 \times \vec{v}_3 = \begin{bmatrix} -2/3 \\ 1/3 \\ 2/3 \end{bmatrix} \times \begin{bmatrix} 2/3 \\ 2/3 \\ 1/3 \end{bmatrix} = \begin{bmatrix} 1/9 - 4/9 \\ -(-2/9 - 4/9) \\ -4/9 - 2/9 \end{bmatrix} = \begin{bmatrix} -1/3 \\ 2/3 \\ -2/3 \end{bmatrix}$$

does the job since $||v_1|| = ||v_2|| = ||v_3|| = 1$ and $\vec{v}_1 \cdot \vec{v}_2 = \vec{v}_1 \cdot \vec{v}_3 = \vec{v}_2 \cdot \vec{v}_3 = 0$. In summary

$$T^{-1} = \begin{bmatrix} -2/3 & -1/3 & 2/3 \\ 1/3 & 2/3 & 2/3 \\ 2/3 & -2/3 & 1/3 \end{bmatrix} \vec{x}.$$

By **??** the matrix of T^{-1} is orthogonal and the matrix $T = (T^{-1})^{-1}$ is the transpose of the matrix of T^{-1}. Therefore, it suffices to use

$$T = \begin{bmatrix} -2/3 & -1/3 & 2/3 \\ 1/3 & 2/3 & 2/3 \\ 2/3 & -2/3 & 1/3 \end{bmatrix}^T \vec{x} = \begin{bmatrix} -2/3 & 1/3 & 2/3 \\ -1/3 & 2/3 & -2/3 \\ 2/3 & 2/3 & 1/3 \end{bmatrix} \vec{x}.$$

Example 4.13. Show that a matrix with orthogonal columns need not be an orthogonal matrix. For example $A = \begin{bmatrix} 4 & -3 \\ 3 & 4 \end{bmatrix}$ is not an orthogonal matrix $T\vec{x} = A\vec{x}$ does not preserve length by comparing the lengths of $\vec{x}$ and $T\vec{x}$ with $\begin{bmatrix} -3 \\ 4 \end{bmatrix}$.

Example 4.14. Find all orthogonal 2×2 matrices. Write $A = \begin{bmatrix} \vec{v}_1 & \vec{v}_2 \end{bmatrix}$. The unit vector $\vec{v}_1$ can be expressed as $\vec{v}_1 = \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix}$, for some θ. Then v_2 will be one of the two unit vectors orthogonal to $\vec{v}_1$, namely $\vec{v}_2 = \begin{bmatrix} -\sin\theta \\ \cos\theta \end{bmatrix}$ or $\vec{v}_2 = \begin{bmatrix} \sin\theta \\ -\cos\theta \end{bmatrix}$. Therefore, an orthogonal 2×2 matrix is either of the form

$$A = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \qquad \text{or} \qquad A = \begin{bmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{bmatrix}$$

representing a rotation or a reflection, respectively.

Example 4.15. Given $n \times n$ matrices A and B which of the following must be symmetric?

- BB^T
- $A^T B^T BA$
- $B(A + A^T)B^T$

The solution to each part follows.

- **??**, BB^T is symmetric because

$$(BB^T)^T = (B^T)^T B^T = BB^T.$$

- **??**, $A^T B^T BA$ is symmetric because

$$(A^T B^T BA)^T = A^T B^T (B^T)^T (A^T)^T = A^T B^T BA.$$

- **??**, $B(A + A^T)B^T$ is symmetric because

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

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

Example 4.16. If the $n \times n$ matrices A and B are symmetric which of the following must be symmetric as well?

- $2I_n + 3A - 4A^2$,
- $AB^2 A$.

The solution to each part follows.

- First note that $(A^2)^T = (A^T)^2 = A^2$ for a symmetric matrix A. Now we can use the linearity of the transpose,

$$(2I_n + 3A - 4A^2)^T = 2I_n^T + 3A^T - 4(A^2)^T = 2I_n + 3A - 4A^2$$

  showing that the matrix $2I_n + 3A - 4A^2$ is symmetric.
- The matrix AB^2A is symmetric since,

$$(AB^2A)^T = (ABBA)^T = (BA)^T(AB)^T = A^TB^TB^TA^T = AB^2A.$$

Example 4.17. Use **??** to find the matrix A of the orthogonal projection onto

$$W = \text{span}\left(\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 9 \\ -5 \\ 3 \end{bmatrix} \right).$$

Then find the matrix of the orthogonal projection onto the subspace of $\mathbb{R}^4$ spanned by the vectors $\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}$. First we apply **??**, the Gram-Schmidt process, to $W = \text{span}(\vec{v}_1, \vec{v}_2)$, to find that the vectors

$$\vec{u}_1 = \frac{\vec{v}_1}{||\vec{v}_1||} = \begin{bmatrix} 1/2 \\ 1/2 \\ 1/2 \\ 1/2 \end{bmatrix}, \qquad \vec{u}_2 = \frac{\vec{v}_2^\perp}{||\vec{v}_2^\perp||} = \frac{\vec{v}_2 - (\vec{u}_1 \cdot \vec{v}_2)\vec{u}_1}{||\vec{v}_2 - (\vec{u}_1 \cdot \vec{v}_2)\vec{u}_1||} = \begin{bmatrix} -1/10 \\ 7/10 \\ -7/10 \\ 1/10 \end{bmatrix}$$

form an orthonormal basis of W. By **??**, the matrix of the projection onto W is $A = QQ^T$ where $Q = \begin{bmatrix} \vec{u}_1 & \vec{u}_2 \end{bmatrix}$. Therefore the orthogonal projection onto W is

$$A = \begin{bmatrix} 1/2 & -1/10 \\ 1/2 & 7/10 \\ 1/2 & -7/10 \\ 1/2 & 1/10 \end{bmatrix} \begin{bmatrix} 1/2 & 1/2 & 1/2 & 1/2 \\ -1/10 & 7/10 & -7/10 & 1/10 \end{bmatrix} = \frac{1}{100} \begin{bmatrix} 26 & 18 & 32 & 24 \\ 18 & 74 & -24 & 32 \\ 32 & -24 & 74 & 18 \\ 24 & 32 & 18 & 26 \end{bmatrix}$$

Let $A = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix}$ and then the orthogonal projection matrix is

$$A(A^TA)^{-1}A^T = \frac{1}{10} \begin{bmatrix} 7 & 4 & 1 & -2 \\ 4 & 3 & 2 & 1 \\ 1 & 2 & 3 & 4 \\ -2 & 1 & 4 & 7 \end{bmatrix}.$$

4.4   Inner Products

Recall that the norm of $x \in \mathbb{R}^n$ defined by $||x|| = \sqrt{x_1^2 + x_2^2}$ is not linear. To injective linearity into the discussion we introduce the dot product: for $x, y \in \mathbb{R}^n$ the **dot product** of x and y is defined as $x \cdot y = x_1 y_1 + \cdots + x_n y_n$. Obviously $x \cdot x = ||x||^2$, and with the dot product being so useful, so we generalize the dot product into an inner product on a vector space V.

An **inner product** on V is a function that takes each ordered pair (u, v) of elements of V to a number $\langle u, v \rangle \in \mathbb{F}$ and has the following properties:

- () **positivity** $\langle v, v \rangle \geq 0$ for all $v \in V$;
- () **definiteness** $\langle v, v \rangle = 0$ if and only if $v = 0$;
- (**additivity** in the first slot) $\langle u + v, w \rangle = \langle u, w \rangle + \langle v, w \rangle$ for all $u, v, w \in V$;
- (**homogeneity** in the first slot) $\langle av, w \rangle = a \langle v, w \rangle$ for all $a \in \mathbb{F}$ and all $v, w, \in V$;
- (**conjugate symmetry**) $\langle v, w \rangle = \overline{\langle w, v \rangle}$ for all $v, w \in V$.

Recall that for $z \in \mathbb{C}^n$, we define the norm of z by

$$||z|| = \sqrt{|z_1|^2 + \cdots + |z_n|^2}$$

where the absolute values are needed because we want $||z||$ to be a non-negative number. Then

$$||z||^2 = z_1 \overline{z_1} + \cdots + z_n \overline{z_n} \tag{4.8}$$

because every $\lambda \in \mathbb{C}$ satisfies $|\lambda|^2 = \lambda \overline{\lambda}$. Since $||z^2||$ is the inner-product of z with itself, as in $\mathbb{R}^n$, the Equation (4.8) suggests that the inner product of $w \in \mathbb{C}^n$ with z should equal

$$w_1 \overline{z_1} + \cdots + w_n \overline{z_n}.$$

We should expect that the inner product of w with z equals the complex conjugate of the inner product of z with w, thus motivating the definition of conjugate symmetry.

An **inner-product space** is a vector space V along with an inner product on V. The inner product defined on $\mathbb{F}^n$ by

$$\langle (w_1, \ldots, w_n), (z_1, \ldots, z_n) \rangle = w_1 \overline{z_1} + \cdots + w_n \overline{z_n}$$

is called the **Euclidean inner product**.

Continue to let V denote a complex or real vector space. In this section we develop the basic theorems for norms. For $v \in V$, the **norm** of v is defined by $||v|| = \sqrt{\langle v, v \rangle}$. Two vectors $u, v \in V$ are **orthogonal** if $\langle u, v \rangle = 0$.

Theorem 4.9. *[Pythagorean Theorem] If u and v are orthogonal vectors in V, then*

$$||u + v||^2 = ||u||^2 + ||y||^2.$$

Proof. Suppose that u, v are orthogonal vectors in V. Then

$$||u + v||^2 = \langle u + v, u + v \rangle = ||u||^2 + ||v||^2 + \langle u, v \rangle + \langle v, u \rangle = ||u||^2 + ||v||^2,$$

as desired.

$\square$

::: {#thm- } [Orthogonal Decomposition] If v is a nonzero vector in V, then u can be written as a scalar multiple of v plus a vector orthogonal to v. :::

Proof. Let $a \in \mathbb{F}$. Then

$$u = av + (u - av).$$

Thus we need to choose a so that v is orthogonal to $(u - av)$. In other words, we want

$$0 = \langle u - av, v \rangle = \langle u, v \rangle - a \langle v, v \rangle = \langle u, v \rangle - a ||v||^2.$$

The equation above shows that we should choose a to be $\langle u, v \rangle / ||v||^2$ (assume that $v \neq 0$ to avoid division by 0). Making this choice of a, we can write

$$u = \frac{\langle u, v \rangle}{||v||^2} v + \left(u - \frac{\langle u, v \rangle}{||v||^2} v \right).$$

Thus, if $v \neq 0$ then the equation above writes u as a scalar multiple of v plus a vector orthogonal to v.

$\square$

::: {#thm- } [Cauchy-Schwarz] If $u, v \in V$, then

$$|\langle u, v \rangle| \leq ||u|| \, ||v||$$

where equality holds if and only if one of u, v is a scalar multiple of the other. :::

Proof. Let $u, v \in V$. If $v = 0$, then both sides and the desired inequality holds. Thus we can assume that $v \neq 0$. Consider the orthogonal decomposition

$$u = \frac{\langle u, v \rangle}{||v||^2} v + w$$

where w is orthogonal to v. By the Pythagorean theorem,

$$\begin{aligned}
||u||^2 &= \left\|\frac{\langle u, v\rangle}{||v||^2}v\right\|^2 + ||w||^2 \\
&= \frac{|\langle u, v\rangle|^2}{||v||^2} + ||w||^2 \\
&\geq \frac{|\langle u, v\rangle|^2}{||v||^2}
\end{aligned}$$

Multiplying both sides by $||v||^2$ and then taking square roots gives the Cauchy-Schwarz inequality. Notice that there is equality if and only if $w = 0$, that is, if and only if u is a multiple of v.

□

::: {#thm- } [Triangular Inequality] If $u, v \in V$, then

$$||u + v|| \leq ||u|| + ||v||$$

where equality holds if and only if one of u, v is a nonnegative multiple of the other. :::

Proof. Let $u, v \in V$. Then

$$\begin{aligned}
||u + v||^2 &= \langle u + v, u + v\rangle \\
&= \langle u, u\rangle + \langle v, v\rangle + \langle u, v\rangle + \langle v, u\rangle \\
&= \langle u, u\rangle + \langle v, v\rangle + \langle u, v\rangle + \overline{\langle u, v\rangle} \\
&= ||u, u||^2 + ||v, v||^2 + 2\text{remark}\,\langle u, v\rangle \\
&\leq ||u, u||^2 + ||v, v||^2 + 2|\langle u, v\rangle| \qquad (4.9) \\
&\leq ||u, u||^2 + ||v, v||^2 + 2\,||u||\,||v|| \qquad (4.10) \\
&= (||u|| + ||v||)^2
\end{aligned}$$

and so by taking square root of both sides yields the triangular inequality. This proof shows that the triangle inequality is an equality if and only if we have equality in (4.9) and (4.10). Thus we have equality in the triangular inequality if and only if

$$\langle u, v\rangle = ||u||\,||v||. \qquad (4.11)$$

If one of u, v is a nonnegative multiple of the other, then (4.11) holds. Conversely, suppose (4.11) holds. The the condition for equality in the Cauchy-Schwarz inequality implies that one of u, v must be a scalar multiple of the other. Clearly, then (4.11) forces the scalar in question to be nonnegative, as desired.

□

::: {#thm- } [Parallelogram Equality] If $u, v \in V$, then

$$||u + v||^2 + ||u - v||^2 = 2 \left(||u||^2 + ||v||^2 \right).$$

:::

Proof. If $u, v \in V$, then

$$
\begin{aligned}
||u + v||^2 + ||u - v||^2 &= \langle u + v, u + v \rangle + \langle u - v, u - v \rangle \\
&= ||u||^2 + ||v||^2 + \langle u, v \rangle + \langle v, u \rangle + ||u||^2 + ||v||^2 - \langle u, v \rangle - \langle v, \\
&= 2 \left(||u||^2 + ||v||^2 \right)
\end{aligned}
$$

as desired.

$\square$

A list of vectors is called **orthonormal** if the vectors in it are pairwise orthogonal and each vector has norm 1.

Theorem 4.10. *If $(e_1, \ldots, e_m)$ is an orthonormal list of vectors in V, then*

$$||a_1 e_1 + \cdots + a_m e_m||^2 = |a_1|^2 + \cdots + |a_n|^2$$

for all $a_1, \ldots, a_m \in \mathbb{F}$.

Proof. Because each e_j has norm 1, this follows easily from repeated by application of the Pythagorean theorem.

$\square$

::: {#thm- } Every orthonormal list of vectors is linearly independent. :::

Proof. Suppose $(e_1, \ldots, e_n)$ is an orthonormal list of vectors in V and $a_1, \ldots, a_n \in \mathbb{F}$ are such that $a_1 e_1 + \cdots + a_n e_n = 0$. Then $|a_1|^2 + \cdots + |a_n|^2 = 0$, which means that all the a_j's are 0, as desired.

$\square$

An **orthonormal basis** of V is an orthonormal list of vectors in V that is also a basis of V.

The importance of orthonormal bases stems mainly from the following proposition.

::: {#thm- } Suppose $(e_1, \ldots, e_n)$ is an orthonormal basis of V. Then

$$v = \langle v, e_1 \rangle e_1 + \cdots + \langle v, e_n \rangle e_n \tag{4.12}$$

and
$$||v|| = |\langle v, e_1 \rangle|^2 + \cdots + |\langle v, e_n \rangle|^2 \qquad (4.13)$$
for every $v \in V$. :::

Proof. Let $v \in V$. Because $(e_1, \ldots, e_n)$ is a basis of V, there exist scalars $a_1, \ldots, a_n$ such that $v = a_1 e_1 + \cdots + a_n e_n$. Take the inner product of both sides of this equation with e_j, getting $\langle v, e_j \rangle = a_j$. Thus (4.12) holds. Clearly (4.13) holds by (4.12) and (**??**).

$\square$

::: {#thm- } [Gram-Schmidt] If $(v_1, \ldots, v_m)$ is a linearly independent list of vectors in V, then there exists an orthonormal list $(e_1, \ldots, e_m)$ of vectors in V such that
$$\text{span}(v_1, \ldots, v_j) = \text{span}(e_1, \ldots, e_j) \qquad (4.14)$$
for $j = 1, \ldots, m$ :::

Proof. Suppose $(v_1, \ldots, v_m)$ is a linearly independent list of vectors in V. To construct the e's, start by setting $e_1 = \frac{v_1}{||v_1||}$. This satisfies (4.14) for $j = 1$. We will choose $e_2, \ldots e_m$ inductively, as follows. Suppose $j > 1$ and an orthonormal list $(e_1, \ldots, e_{j-1})$ has been chosen so that
$$\text{span}(v_1, \ldots, v_{j-1}) = \text{span}(e_1, \ldots, e_{j-1}). \qquad (4.15)$$

Let
$$e_j = \frac{v_j - \langle v_j, e_1 \rangle e_1 - \cdots - \langle v_j, e_{j-1} \rangle e_{j-1}}{||v_j - \langle v_j, e_1 \rangle e_1 - \cdots - \langle v_j, e_{j-1} \rangle e_{j-1}||}. \qquad (4.16)$$
Note that $v_j \notin \text{span}(v_1, \ldots, v_{j-1})$ (because $(v_1, \ldots, v_m)$ is linearly independent) and thus $v_j \notin \text{span}(e_1, \ldots, e_{j-1})$. Hence we are not dividing by 0 in the equation above, and so e_j is well-defined. Dividing a vector by its norm produces a new vector with norm 1; thus $||e_j|| = 1$.

Let $1 \leq k < j$. Then
$$\langle e_j, e_k \rangle = \left\langle \frac{v_j - \langle v_j, e_1 \rangle e_1 - \cdots - \langle v_j, e_{j-1} \rangle e_{j-1}}{||v_j - \langle v_j, e_1 \rangle e_1 - \cdots - \langle v_j, e_{j-1} \rangle e_{j-1}||}, e_k \right\rangle$$
$$= \frac{\langle v_j, e_k \rangle - \langle v_j, e_k \rangle \langle e_k, e_k \rangle}{||v_j - \langle v_j, e_1 \rangle e_1 - \cdots - \langle v_j, e_{j-1} \rangle e_{j-1}||}$$
$$= 0.$$

Thus $(e_1, \ldots, e_j)$ is an orthonormal list.

From (4.16), we see that $v_j \in \text{span}(e_1, \ldots, e_j)$. Combining this information with (4.15) shows that
$$\text{span}(v_1, \ldots, v_{j-1}) \subset \text{span}(e_1, \ldots, e_j).$$

Both lists above are linearly independent (the v's by hypothesis, the e's by orthonormality and (4.4). Thus both subspaces above have dimension j, and hence must be equal, completing the proof.

$\square$

Theorem 4.11. *Every finite-dimensional inner-product space has an orthonormal basis.*

Proof. Choose a basis of V. Apply the Gram-Schmidt procedure to it, producing an orthonormal list. This list is linearly independent and it spans V. Thus it is an orthonormal basis.

$\square$

Theorem 4.12. *Every orthonormal list of vectors in V can be extended to an orthonormal basis of V.*

Proof. Suppose $(e_1, \dots, e_m)$ is an orthonormal list of vectors in V. Then $(e_1, \dots, e_m)$ is linearly independent and so can be extended to a basis

$$\mathcal{B} = (e_1, \dots, e_m, v_1, \dots, v_n)$$

of V. Now apply the Gram-Schmidt procedure to $\mathcal{B}$ producing an orthonormal list $(e_1, \dots, e_m, f_1, \dots, f_n)$; here the Gram-Schmidt procedure leaves the first m vectors unchanged because they are already orthonormal. Clearly $\mathcal{B}$ is an orthonormal basis of V because it is linearly independent and its span equals V. Hence we have our extension of $(e_1, \dots, e_m)$ to an orthonormal basis of V.

$\square$

Recall that if V is a complex vector space, then for each operator on V there is a basis with respect to which the matrix of the operator is upper-triangular. Now for inner-product spaces we would like to know the same question.

::: {#thm- } Suppose $T \in \mathcal{L}(V)$. If T has an upper-triangular matrix with respect to some basis of V, then T has an upper-triangular matrix with respect to some orthonormal basis of V. :::

Proof. Suppose T has upper-triangular matrix with respect to some basis $(v_1, \dots, v_n)$ of V. Thus $\mathrm{span}(v_1, \dots, v_j)$ is invariant under T for each $j =$

$1, \dots, n$. Apply the Gram-Schmidt procedure to $(v_1, \dots, v_n)$, producing an orthonormal basis $(e_1, \dots, e_n)$ of V. Because

$$\text{span}(e_1, \dots, e_j) = \text{span}(v_1, \dots, v_j)$$

for each j, we conclude that $\text{span}(e_1, \dots, e_j)$ is invariant under T for each $j = 1, \dots, n$. Thus, by (??), T has an upper-triangular matrix with respect to the orthonormal basis $(e_1, \dots, e_n)$.

□

Theorem 4.13. *Suppose V is a complex vector space and $T \in \mathcal{L}(V)$. Then T has an upper-triangular matrix with respect to some orthonormal basis of V.*

Proof. This follows immediately from (??) and (4.4).

□

If U is a subset of an inner-product space V, then the **orthogonal complement** of U is defined as $U^{\perp} = \{v \in V : \langle v, u \rangle = 0 \text{ for all } u \in U\}$.

::: {#thm- } [Orthogonal Decomposition] If U is a subspace of an inner-product space V, then $v = U \oplus U^{\perp}$. :::

Proof. Suppose that U is a subspace of V. First we will show that

$$V = U + U^{\perp}. \tag{4.17}$$

To do this, suppose $v \in V$. Let $(e_1, \dots, e_m)$ be an orthonormal basis of U. Obviously,

$$v = \underbrace{\langle v, e_1 \rangle\, e_1 + \dots + \langle v, e_m \rangle\, e_m}_{u} + \underbrace{v - \langle v, e_1 \rangle\, e_1 - \dots - \langle v, e_m \rangle\, e_m}_{w}.$$

Clearly, $u \in U$. Because $(e_1, \dots, e_m)$ is an orthonormal list, for each j we have

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

Thus w is orthogonal to every vector in $\text{span}(e_1, \dots, e_m)$. In other words, $w \in U^{\perp}$, completing the proof of (4.17).

If $v \in U \cap U^{\perp}$, then v (which is in U) is orthogonal to every vector in U (including v itself), which implies that $\langle v, v \rangle = 0$, which implies that $v = 0$. Thus

$$U \cap U^{\perp} = \{0\}. \tag{4.18}$$

Now (4.17) and (4.18) imply that $U \oplus U^{\perp}$.

□

Theorem 4.14. *If U is a subspace of an inner-product space V, then $U = (U^\perp)^\perp$.*

Proof. Suppose that U is a subspace of V. First we will show that

$$U \subseteq (U^\perp)^\perp. \tag{4.19}$$

To do this, suppose that $u \in U$. Then $\langle u, v \rangle = 0$ for every $v \in U^\perp$ (by definition of $U^\perp$). Because u is orthogonal to every vector in $U^\perp$, we have $u \in (U^\perp)^\perp$, completing the proof of (4.19).

To prove the inclusion in the other direction, suppose $v \in (U^\perp)^\perp$. By (4.4), we can write $v = u + w$, where $u \in U$ and $w \in U^\perp$. We have $v - u = w \in U^\perp$. Because $v \in (U^\perp)^\perp$ and $u \in (U^\perp)^\perp$ (from (4.19)), we have $v - u \in (U^\perp)^\perp$. Thus $v - u \in U^\perp \cap (U^\perp)^\perp$, which implies that $v = u$, which implies that $v \in U$. Thus $(U^\perp)^\perp \subseteq U$, which along with (4.19) completes the proof.

$\square$

Let $V = U \oplus U^\perp$ and for $v \in V$ let $v = u + w$ where $w \in U^\perp$. Then u is called the **orthogonal projection** of V onto U and is denoted by $P_U v$.

Theorem 4.15. *If U is a subspace of an inner-product space V and $v \in V$. Then $||v - P_U v|| \leq ||v - u||$ for every $u \in U$. Furthermore, if $u \in U$ and the inequality above is an equality; then $u = P_U v$.*

Proof. Suppose $u \in U$. Then

$$||v - P_U v||^2 \leq ||v - P_U v|| + ||P_U v - u||^2 \tag{4.20}$$

$$= ||v - P_U v + P_U v - u||^2 = ||v - u||^2 \tag{4.21}$$

where (4.21) comes from the Pythagorean theorem, which applies because $v - P_U v \in U^\perp$ and $P_U v - u \in U$. Taking the square root gives the desired inequality. The inequality is an equality if and only if (4.20) is an equality, which happens if and only if $||P_U v - u|| = 0$, which happens if and only if $u = P_u v$.

$\square$

Example 4.18. Show that if $c_1, \dots, c_n$ are positive numbers, then

$$\langle (w_1, \dots, w_n), (z_1, \dots, z_n) \rangle = c_1 w_1 \overline{z_1} + \dots + c_n w_n \overline{z_n}$$

defines an inner product on $\mathbb{F}^n$.

Example 4.19. Show that if $p, q \in \mathcal{P}_m(\mathbb{F})$, then

$$\langle p, q \rangle = \int_0^1 p(x)\overline{q(x)}dx$$

is an inner product on the vector space $\mathcal{P}_m(\mathbb{F})$.

Example 4.20. Show that every inner product is a linear map in the first slot, as well as a linear map in the second slot.

Example 4.21. If $v \in V$ and $a \in \mathbb{F}$, then $||av|| = |a|\, ||v||$, and if v is nonzero then $u = \frac{1}{||v||}v$ is a unit vector. Since $||av||^2 = \langle av, av \rangle = a\bar{a}\langle v, v \rangle = |a|^2\, ||v||^2$, taking square roots provides $||av|| = |a|\, ||v||$.

Example 4.22. Prove that if x, y are nonzero vectors in $\mathbb{R}^2$, then

$$\langle x, y \rangle = ||x||\, ||y|| \cos \theta,$$

where θ is the angle between x and y. The law of cosines gives

$$||x - y||^2 = ||x||^2 + ||y||^2 - 2\, ||x||\, ||y|| \cos \theta.$$

The left hand side of this equation is

$$||x - y||^2 = (x - y) \cdot (x - y) = ||x||^2 - 2(x \cdot y) + ||y||^2$$

so

$$x \cdot y = ||x||\, ||y|| \cos \theta.$$

Example 4.23. Suppose $u, v \in V$. Prove that $\langle u, v \rangle = 0$ if and only if $||u|| \leq ||u + av||$ for all $a \in \mathbb{F}$. If $\langle u, v \rangle = 0$, then by the Pythagorean theorem

$$||u + \alpha v||^2 = ||u||^2 + ||\alpha v||^2 \geq ||u||.$$

Conversely, we will prove the contrapositive, that is we will prove: if $\langle u, v \rangle \neq 0$ then there exists $a \in \mathbb{F}$ such that $||u|| > ||u + av||$. Suppose $\langle u, v \rangle \neq 0$ then u and v are both nonzero vectors. By the orthogonal decomposition, we can write

$$u = \alpha v + w \tag{4.22}$$

for some $\alpha \in \mathbb{F}$ and where $\langle w, v \rangle = 0$. Notice $\alpha \neq 0$ since $\langle u, v \rangle \neq 0$. Since v and w are orthogonal

$$||u||^2 = |\alpha|^2 ||v||^2 + ||w||^2$$

Let $a = -\alpha$. Then by equation (4.22)

$$||u + av||^2 = ||w||^2$$

and so

$$||u||^2 = |a|^2 ||v||^2 + ||u + av||^2 > ||u + av||^2$$

which implies

$$||u|| > ||u + av||$$

as desired.

Example 4.24. Prove that

$$\left(\sum_{k=1}^{n} a_k b_k \right)^2 \leq \left(\sum_{k=1}^{n} k a_k^2 \right) \left(\sum_{k=1}^{n} \frac{b_k^2}{k} \right)$$

for all real numbers $a_1, \ldots, a_n$ and $b_1, \ldots, b_n$. This is a simple trick.

$$\left(\sum_{k=1}^{n} a_k b_k \right)^2 = \left(\sum_{k=1}^{n} \sqrt{k} a_k \frac{b_k}{\sqrt{k}} \right)^2 \leq \left(\sum_{k=1}^{n} k a_k^2 \right) \left(\sum_{k=1}^{n} \frac{b_k^2}{k} \right)$$

where the last inequality is from the Cauchy-Schwarz inequality.

Example 4.25. Suppose $u, v \in V$ are such that $||u|| = 3$, $||u + v|| = 4$, and $||u - v|| = 6$. What number must $||v||$ equal? Using the parallelogram equality

$$||u + v||^2 + ||u - v||^2 = 2 \left(||u||^2 + ||v||^2 \right)$$

to get

$$16 + 36 = 2(9 + ||v||^2), \qquad ||v|| = \sqrt{17}.$$

Example 4.26. Prove or disprove: there is an inner product on $\mathbb{R}^2$ such that the associated norm is given by $||(x_1, x_2)|| = |x_1| + |x_2|$ for all $(x_1, x_2) \in \mathbb{R}^2$. There is no such inner product. Take for instance,

$$u = (1/4, 0), \qquad v = (0, 3/4), \qquad u + v = (1/4, 3/4).$$

Then we have equality in the triangular inequality

$$1 = ||u + v|| \leq ||u|| + ||v|| = 1/4 + 3/4.$$

By the triangular inequality, we must have $u = av$ or $v = av$, with $a \geq 0$. But clearly no such $a \in \mathbb{F}$ exists.

Example 4.27. Prove that if V is a real inner-product space, then

$$\langle u, v \rangle = \frac{||u + v||^2 - ||u - v||^2}{4}$$

for all $u, v \in V$. Expressing the norms as inner products

$$
\begin{aligned}
\frac{||u + v||^2 - ||u - v||^2}{4} &= \frac{\langle u + v, u + v \rangle - \langle u - v, u - v \rangle}{4} \\
&= \frac{\langle u, u \rangle + \langle v, v \rangle + \langle u, v \rangle + \langle v, u \rangle - \langle u, u \rangle - \langle v, v \rangle + \langle u, v \rangle + \langle v, u \rangle}{4} \\
&= \frac{2 \langle u, v \rangle + 2 \langle v, u \rangle}{4} \\
&= \frac{2 \langle u, v \rangle + 2 \langle u, v \rangle}{4} \quad \text{(because V is a real inner product space)} \\
&= \langle u, v \rangle
\end{aligned}
$$

as desired.

Example 4.28. Prove that if V is a complex inner-product space, then

$$\langle u, v \rangle$$

is

$$\frac{||u + v||^2 - ||u - v||^2 + ||u + iv||^2 i - ||u - iv||^2 i}{4}$$

for all $u, v \in V$.

Example 4.29. A norm on a vector space U is a function $|| \ || : U \to [0, \infty)$ such that $||u|| = 0$ if and only if $u = 0$, $||\alpha u|| = |\alpha| ||u||$ for all $\alpha \in F$ and all $u \in U$, and $||u + v|| \leq ||u|| + ||v||$ for all $u, v \in U$. Prove that a norm satisfying the parallelogram equality comes from an inner product (in other words, show that if $|| \ ||$ is a norm on U satisfying the parallelogram equality, then there is an inner product $\langle \ , \ \rangle$ on U such that $||u|| = \langle u, u \rangle^{1/2}$ for all $u \in U$).

Example 4.30. Suppose n is a positive integer. Prove that

$$\left(\frac{1}{\sqrt{2\pi}}, \frac{\sin x}{\sqrt{\pi}}, \frac{\sin 2x}{\sqrt{\pi}}, \dots, \frac{\sin nx}{\sqrt{\pi}}, \frac{\cos x}{\sqrt{\pi}}, \frac{\cos 2x}{\sqrt{\pi}}, \dots, \frac{\cos nx}{\sqrt{\pi}} \right)$$

is an orthonormal list of vectors in $\mathcal{C}[-\pi, \pi]$, the vector space of continuous real-valued functions on $[-\pi, \pi]$ with inner product

$$\langle f, g \rangle = \int_{-\pi}^{\pi} f(x) g(x) \, dx.$$

Computation of these integrals is based on the product-to-sum formulas from trigonometry:

$$\sin(A)\sin(B) = \frac{1}{2}\cos(A - B) - \frac{1}{2}\cos(A + B)$$

$$\cos(A)\cos(B) = \frac{1}{2}\cos(A - B) + \frac{1}{2}\cos(A + B)$$

$$\sin(A)\cos(B) = \frac{1}{2}\sin(A - B) + \frac{1}{2}\sin(A + B).$$

Here is a sample computation, valid for $m, n = 1, 2, 3, , \ldots$ when $m \neq n$.

$$\langle \sin(mx), \cos(nx) \rangle = \int_{-\pi}^{\pi} \sin(mx)\cos(nx)\, dx$$

$$= \int_{-\pi}^{\pi} \left(\frac{1}{2}\sin((m - n)x) + \frac{1}{2}\sin((m + n)x) \right)\, dx$$

$$= \frac{-1}{2(m - n)}\cos\left((m - n)x\right) + \frac{-1}{2(m + n)}\cos\left((m + n)x\right)\Big|_{-\pi}^{\pi},$$

$$= \frac{-1}{2(m - n)}[-1 - (-1)] + \frac{-1}{2(m - n)}[-1 - (-1)]$$

$$= 0.$$

Example 4.31. On $\mathcal{P}_2(\mathbb{R})$, consider the inner product given by

$$\langle p, q \rangle = \int_{o}^{1} p(x)q(x)\, dx.$$

Apply the Gram-Schmidt procedure to the basis $(1, x, x^2)$ to produce an orthonormal basis of $\mathcal{P}_2(\mathbb{R})$. Computing e_1: A calculation gives

$$\langle 1, 1 \rangle = \int_{0}^{1} 1\, dx = 1,$$

so $e_1 = 1$. Computing e_2: A calculation gives

$$\langle x, 1 \rangle = \int_{0}^{1} x\, dx = \frac{1}{2}.$$

Let

$$f_2 = x - 1/2.$$

Then

$$\|f_2\|^2 = \langle x - 1/2, x - 1/2 \rangle = \int_{0}^{1} \left(x^2 - x + \frac{1}{4} \right)\, dx = \frac{1}{12},$$

so

$$e_2 = \frac{f_2}{||f_2||} = 2\sqrt{3}\left(x - \frac{1}{2}\right).$$

Computing e_3: A calculation gives

$$\langle x^2, 1\rangle = \int_0^1 x^2\,dx = \frac{1}{3},$$

and

$$\left\langle x^2, 2\sqrt{3}\left(x - \frac{1}{2}\right)\right\rangle = 2\sqrt{3}\int_0^1 \left(x^3 - \frac{x^2}{2}\right)dx = \frac{1}{2\sqrt{3}}.$$

Let

$$f_3 = x^2 - \frac{1}{2\sqrt{3}}\left(x - \frac{1}{2}\right) - \frac{1}{3} = x^2 - x + \frac{1}{6}.$$

Then

$$||f_3||^2 = \int_0^1 \left(x^2 - x + \frac{1}{6}\right)^2 dx$$

and

$$e_3 = \frac{f_3}{||f_3||}.$$

Example 4.32. What happens if the Gram-Schmidt procedure is applied to a list of vectors that is not linearly independent? By examining the proof, notice that the numerator in the Gram-Schmidt formula is the difference between v_j and the orthogonal projection P_u of v_j onto the subspace

$$U = \text{span}(v_1, \dots, v_{j-1}) = \text{span}(e_1, \dots, e_{j-1}).$$

If $v_j \in U$, then $v_j - P_U v_j = 0$, so the numerator has norm 0 and division by the denominator is not defined. The algorithm can be adapted to handle this case by testing for 0 in the denominator. If 0 is found, throw v_j out of the list and continue. The result will be an orthonormal basis for $\text{span}(v_1, \dots, v_N)$.

Example 4.33. Suppose V is a real inner-product space and $(v_1, \dots, v_m)$ is a linearly independent list of vectors in V. Prove that there exist exactly 2^m orthonormal lists $(e_1, \dots, e_m)$ of vectors in V such that span $(v_1, \dots, v_j) = \text{span}(e_1, \dots, e_j)$ for all $j \in \{1, \dots, m\}$.

Example 4.34. Suppose $(e_1, \dots, e_M)$ is an orthonormal list of vectors in V. Let $v \in V$. Prove that

$$||v||^2 = \sum_{n=1}^{M} |\langle v, e_n\rangle|^2$$

if and only if v span $(e_1, \dots, e_M)$. Extend $(e_1, \dots, e_M)$ to an orthonormal basis for V. Then

$$\sum_{n=1}^{N} \langle v, e_n \rangle e_n$$

and

$$||v||^2 = \sum_{n=1}^{N} |\langle v, e_n \rangle|^2 .$$

If $v \in \text{span}(e_1, \dots, e_M)$ then $\langle v, e_n \rangle = 0$ for $n > M$, and

$$||v||^2 = \sum_{m=1}^{M} |\langle v, e_n \rangle|^2 .$$

If $v \notin \text{span}(e_1, \dots, e_M)$ then for some $n > M$ we have $\langle v, e_n \rangle \neq 0$. This gives

$$||v||^2 = \sum_{n=1}^{N} |\langle v, e_n \rangle|^2 > \sum_{m=1}^{M} |v, e_n|^2 .$$

Example 4.35. Find an orthonormal basis of $\mathcal{P}_2(\mathbb{R})$, such that the differentiation operator on $\mathcal{P}_2(\mathbb{R})$ has an upper-triangular matrix with respect to this basis.

Example 4.36. Suppose U is a subspace of V. Prove that $\dim U^\perp = \dim V - \dim U$.

Example 4.37. Suppose U is a subspace of V. Prove that $U^\perp = \{0\}$ if and only if $U = V$. If $U = V$ and $v \in U^\perp$ then $v \in U \cap U^\perp$, and $\langle v, v \rangle = 0$, so $v = 0$. Therefore, $V = U \oplus U^\perp$. If $U^\perp = \{0\}$, then $U = V$.

Example 4.38. Prove that if $P \in \mathcal{L}(V)$ is such that $P^2 = P$ and every vector in null P is orthogonal to every vector in range P, then P is an orthogonal projection.

Example 4.39. Prove that if $P \in \mathcal{L}(V)$ is such that $P^2 = P$ and $||Pv|| \leq ||v||$ for every $v \in V$, then P is an orthogonal projection.

Example 4.40. Suppose $T \in \mathcal{L}(V)$ and U is a subspace of V. Prove that U is invariant under T if and only if $P_U T = T P_U$.

Example 4.41. Suppose $T \in \mathcal{L}(V)$ and U is a subspace of V. Prove that U and $U^{\perp}$ are both invariant under T if and only if $P_U T = T P_U$.

Example 4.42. In $\mathbb{R}^4$, let $U = \text{span}\,((1,1,0,0),(1,1,1,2))$. Find $u \in U$ such that $||u - (1,2,3,4)||$ is as small as possible. in $\mathbb{R}^4$ let $U = \text{span}((1,1,0,0),(1,1,1,2))$. Find $u \in U$ such that $||u - (1,2,3,4)||$ is as small as possible. We want the orthogonal projection $P_U(1,2,3,4)$. Notice that $U = \text{span}((1,1,0,0),(0,0,1,2))$. An orthonormal basis for U is

$$\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0, 0 \right), \left(0, 0, \frac{1}{\sqrt{5}}, \frac{2}{\sqrt{5}} \right)$$

Thus the desired vector is

$$P_U(1,2,3,4) = \left(\frac{3}{2}, \frac{3}{2}, 0, 0 \right) + \left(0, 0, \frac{11}{5}, \frac{2}{5} \right).$$

Example 4.43. Find a polynomial $p \in \mathcal{P}_3(\mathbb{R})$ such that $p(0) = 0$, $p'(0) = 0$, and

$$\int_0^1 |2 + 3x - p(x)|^2 \, dx$$

is as small as possible.

Example 4.44. Find a polynomial $p \in \mathcal{P}_5(\mathbb{R})$ that makes

$$\int_{-\pi}^{\pi} |\sin x - p(x)|^2 \, dx$$

is as small as possible.

Example 4.45. Find a polynomial $p \in \mathcal{P}_2(\mathbb{R})$ such that

$$\phi(p) = p\left(\frac{1}{2} \right) = \int_0^1 p(x)\, q(x)\, dx$$

for every $p \in \mathcal{P}_2(\mathbb{R})$. Here is the direct approach. Every $q \in \mathcal{P}_2(\mathbb{R})$ can be expressed as

$$\alpha + \beta(x - 1/2) + \gamma(x - 1/2)^2.$$

The desired polynomial q must satisfy

$$p(x) = 1: \qquad \phi(1) = p(1/2) = 1 = \int_0^1 \left(\alpha + \beta(x - 1/2) + \gamma(x - 1/2)^2 \right) \, dx = \alpha + \gamma \frac{1}{12}.$$

Moving to $p(x) = x - 1/2$ we find

$$p(x) = x - 1/2 : \qquad \phi(x - 1/2) = p(1/2) = 0$$

$$= \int_0^1 (x - 1/2)[\alpha + \beta(x - 1/2) + \gamma(x - 1/2)^2] \, dx = \beta \frac{1}{12},$$

so $\beta = 0$. Finally

$$p(x) = (x - 1/2)^2 : \qquad \phi((x - 1/2)^2) = p(1/2) = 0$$

$$= \int_0^1 (x - 1/2)^2 [\alpha + \beta(x - 1/2) + \gamma(x - 1/2)^2] \, dx = \alpha \frac{1}{12} + \gamma \frac{1}{80}.$$

Solving gives

$$\alpha = \frac{27}{12}, \qquad \beta = 0, \qquad \gamma = -15.$$

Thus

$$q(x) = \frac{27}{12} - 15(x - 1/2)^2.$$

Example 4.46. Find a polynomial $q \in \mathcal{P}_2(\mathbb{R})$ such that

$$\phi(p) = \int_0^1 p(x) \, (\cos \pi x) \, dx = \int_0^1 p(x) \, q(x) \, dx$$

for every $p \in \mathcal{P}_2(\mathbb{R})$. Taking the same approach as in the previous example. We compute

$$p(x) = 1 : \qquad \phi(1) = \int_0^1 \cos(\pi x) \, dx = 0$$

$$= \int_0^1 [\alpha + \beta(x - 1/2) + \gamma(x - 1/2)^2] \, dx = \alpha + \gamma \frac{1}{12}.$$

Moving to $p(x) = x - 1/2$ we find

$$p(x) = x - 1/2 : \qquad \phi(x - 1/2) = \int_0^1 (x - 1/2) \cos(\pi x) \, dx = \int_0^1 x \cos(\pi x) \, dx$$

$$= -\frac{2}{\pi^2} = \int_0^1 (x - 1/2)[\alpha + \beta(x - 1/2) + \gamma(x - 1/2)^2] \, dx = \beta \frac{1}{12},$$

so $\beta = \frac{-24}{\pi^2}$. Finally, since $\cos(\pi x)$ is odd about $x = 1/2$,

$$p(x) = (x - 1/2)^2 : \qquad \phi((x - 1/2)^2) = \int_0^1 (x - 1/2)^2 \cos(\pi x) \, dx = 0$$

$$= \int_0^1 (x - 1/2)^2 [\alpha + \beta(x - 1/2) + \gamma(x - 1/2)^2] = \alpha\frac{1}{12} + \gamma\frac{1}{80}.$$

Solving gives

$$\alpha = \gamma = 0, \qquad \beta = \frac{-24}{\pi^2}.$$

Thus

$$q(x) = \frac{-24}{\pi^2}(x - 1/2).$$

Example 4.47. Give an example of a real vector space V and $T \in \mathcal{L}(V)$ such that $\text{trace}(T^2) < 0$.

Example 4.48. Suppose V is a real vector space, $T \in \mathcal{L}(V)$, and V has a basis consisting of eignvectors of T. Prove that $\text{trace}(T^2) \geq 0$.

Example 4.49. Suppose V is an inner-product space and $v, w \in V$. Define $T \in \mathcal{L}(V)$ by $Tu = \langle u, v \rangle w$. Find a formula for trace T.

Example 4.50. Prove that if $P \in \mathcal{L}(V)$ satisfies $P^2 = P$, then trace P is a nonnegative integer.

Example 4.51. Prove that if V is an inner-product space and $T \in \mathcal{L}(V)$, then trace $T^* = \overline{\text{trace}T}$.

Example 4.52. Suppose V is an inner-product space and $T \in \mathcal{L}(V)$ is a positive operator with trace $T = 0$, then $T = 0$.

Example 4.53. Suppose $T \in \mathcal{L}(\mathbb{C}^3)$ is the operator whose matrix is

$$\begin{bmatrix} 51 & -12 & -21 \\ 60 & -40 & -28 \\ 57 & -68 & 1 \end{bmatrix}.$$

If -48 and 24 are eigenvalues of T, find the third eigenvalue of T.

Example 4.54. Prove or give a counterexample: if $T \in \mathcal{L}(V)$ and $c \in F$, then $\text{trace}(cT) = c$ trace T.

Example 4.55. Prove or give a counterexample,: if $S, T \in \mathcal{L}(V)$, then trace $(ST) = (\mathrm{trace} S)(\mathrm{trace} T)$.

Example 4.56. Suppose $T \in \mathcal{L}(V)$. Prove that if $\mathrm{trace}(ST) = 0$ for all $S \in \mathcal{L}(V)$, then $T = 0$.

Example 4.57. Suppose V is an inner-product space and $T \in \mathcal{L}(V)$. Prove that if $(e_1, \ldots., e_n)$ is an orthonormal basis of V, then

$$\mathrm{trace}(T^*T) = ||Te_1||^2 + \cdots + ||Te_n||^2.$$

Conclude that the right side of the equation above is independent of which orthonormal basis $(e_1, \ldots, e_n)$ is chosen for V.

Example 4.58. Suppose V is a complex inner-product space and $T \in \mathcal{L}(V)$. Let $\lambda_1, \ldots., \lambda_n$ be the eigenvalues of T, repeated according to multiplicity.

Example 4.59. Suppose

$$\begin{bmatrix} a_{1,1} & \cdots & a_{1,n} \\ \vdots & & \vdots \\ a_{n,1} & \cdots & a_{n,n} \end{bmatrix}$$

is the matrix of T with respect to some orthonormal basis of V. Prove that

$$|\lambda_1|^2 + \cdots + |\lambda_n|^2 \le \sum_{k=1}^{n} \sum_{j=1}^{n} |a_{j,k}|^2.$$

Example 4.60. Suppose V is an inner-product space. Prove that $\langle S, T \rangle = \mathrm{trace}(ST^*)$ defines an inner-product on $\mathcal{L}(V)$.

Example 4.61. Suppose V is an inner-product space and $T \in \mathcal{L}(V)$. Prove that if $||T^*v|| \le ||Tv||$ for every $v \in V$, then T is normal.

Example 4.62. Prove or give a counterexample: if $T \in \mathcal{L}(V)$ and $c \in F$, then $\det(cT) = c^{\dim V} \det T$.

Example 4.63. Prove or give a counterexample: if $T \in \mathcal{L}(V)$, then $\det(S + T) = \det S + \det T$.

Example 4.64. Suppose A is a block upper-triangular matrix

$$
A = \begin{bmatrix} A_1 & & * \\ & \ddots & \\ 0 & & A_m \end{bmatrix},
$$

where each A_j along the diagonal is a square matrix. Prove that

$$
\det A = (\det A_1) \cdots (\det A_m).
$$

Chapter 5

Determinants

5.1 What are determinants and what do they do

Determinants are mathematical objects that affect the size, shape, or position of something else. In other words, they determine how something looks or behaves. The most common type of determinant is a matrix, which is used to determine the size, shape, and position of a two-dimensional object.

For example, a matrix can be used to determine the position of a point on a graph or the angle of a line. Determinants can also be used to solve problems in physics and engineering.

For example, they can be used to calculate the forces acting on an object in order to determine its motion. Determinants are also used in economics to predict changes in prices and wages. In short, determinants are powerful tools that can be used to solve a wide variety of problems.

A determinant is a value that can be computed from the elements of a square matrix. The determinant of a matrix A is denoted by det(A), det A, or $|A|$.

Geometrically, it can be viewed as the volume of the n-dimensional parallelepiped that has the column vectors of A as its edges. If A is an invertible matrix, then the determinant is nonzero and A can be undone by multiplying it on either side by its inverse matrix. In particular, for a 2×2 matrix, the determinant is simply the product of the diagonal entries (top left times bottom right, minus top right times bottom left).

5.2   How to find the determinant of a matrix

A determinant is a scalar value that can be computed from the elements of a square matrix. It is a useful tool in linear algebra for solving systems of linear equations and for computing the inverse of a matrix.

The determinant of a 2×2 matrix is: $|A| = ad - bc$ where $ad - bc$ is the signed determinant (the "discriminant") because its value changes sign when two rows or columns are interchanged.

For matrices of larger size, the determinant can be computed using any of several methods, such as cofactor expansion or Laplace expansion. When computing the determinant, it is often convenient to change the order of the rows or columns in order to simplify the calculation.

Another method is to use Gauss-Jordan elimination. This method involves using row operations to transform the matrix into an upper triangular matrix. The determinant of the matrix can then be found by taking the product of the diagonal elements.

5.3   Determinant Properties

The determinant has several important properties:

1. If two rows (or columns) of a matrix are identical, then the determinant is zero.

2. If two rows (or columns) are interchanged, then the sign of the determinant changes.

3. If each element of a row (or column) is multiplied by a nonzero constant, then the determinant is multiplied by that constant.

5.4   Expansion by Cofactors

The determinant of a matrix A is equal to the sum of the products of the elements in any one row or column, with each element being multiplied by a corresponding cofactor. Cofactors are special numbers that can be calculated from the minors of the matrix, which are the determinants of the matrices that result from removing one row and one column from A.

The sign of each cofactor is determined by its position in the matrix: odd-numbered rows and columns have positive signs, while even-numbered rows and columns have negative signs. The determinant of a matrix can be used to solve systems of linear equations, and it also provides information about the nature of the matrix itself.

For example, a matrix with a zero determinant is known as singular, meaning that it cannot be inverted. This makes it an important tool

for solving mathematical problems. Determinants can be calculated by hand for small matrices, but larger matrices require the use of specialized software or an electronic calculator.

5.5 Cramer's Rule

Cramer's Rule is a method for solving systems of linear equations. In order to use Cramer's Rule, the system must be expressed in matrix form. The determinant of the matrix is then calculated. This determinant is used to calculate the value of each variable in the system. Cramer's Rule can be used to solve systems with any number of variables, but the matrices involved become increasingly complex as the number of variables increases.

Despite this complexity, Cramer's Rule provides a convenient way to solve systems of linear equations.

5.6 How I teach in this book

How I teach in this book, is by showing examples that students can learn from and understand. I also offer an in-depth look into linear algebra. My approach in this book is to first provide a gentle introduction to the topic at hand. Next, I guide the reader through the necessary calculations required to fully grasp the concepts being presented. Finally, I show how these ideas can be applied in practical scenarios.

Throughout the book, I include worked examples and exercises for readers to test their understanding as they progress. With this book as your companion, you will soon be confident in manipulating determinants and matrices.

Determinants are a key concept in Linear Algebra, and understanding them is crucial for being able to do many matrix operations. This book will take you from the very basics of what determinants are, all the way up to using them in advanced topics like Cramer's Rule. By the end, you will be a pro at using determinants in all sorts of situations.

The book will focus on the fundamentals of eigenvalues and eigenvectors, and how they can be applied in various settings. It is perfect for readers who want to learn more about these topics, or for those who need to brush up on their skills.

The book begins with a review of linear algebra, before moving on to discuss eigenvalues and eigenvectors. This is followed by a discussion of applications of these concepts, including in physics and engineering. The book concludes with a set of exercises for readers to test their understanding.

Eigenvalues and eigenvectors are fundamental concepts in linear algebra, with many applications in physics and engineering. In this section, we will discuss how to find eigenvalues and eigenvectors for a given matrix.

Suppose we have a square matrix A. To find the eigenvalues of A, we need to solve the equation $Ax = \lambda x$ for λ and x. Here, λ is the eigenvalue and x is the corresponding eigenvector.

To solve this equation, we can use the characteristic polynomial of A. This is a polynomial equation whose roots are the eigenvalues of A. To find the characteristic polynomial, we need to take the determinant of $A - \lambda I$, where the matrix I is the identity matrix.

In other words, an eigenvector is a vector that does not change direction when multiplied by a matrix. Eigenvectors are often used to diagonalize matrices, which means that they can be used to simplify matrices that do not have a simple structure.

To find the eigenvalues and eigenvectors of a matrix, one can use the characteristic polynomial of the matrix. The characteristic polynomial is a polynomial equation whose roots are the eigenvalues of the matrix. To find the eigenvectors of the matrix, one can use the inverse of the matrix.

The inverse of the matrix is a matrix that satisfies the equation: $A * A - 1 = I$. The inverse of a matrix exists if and only if the determinant of the matrix is not equal to zero. If the determinant of the matrix is equal to zero, then the matrix is singular and does not have an inverse. Once you have found the inverse of the matrix, you can multiply it by the original matrix to find the eigenvectors of the original matrix.

Linear algebra is the study of mathematical objects that can be described by linear equations. These objects include vectors, matrices, and linear transformations. Linear algebra is a powerful tool that can be used to solve many problems in physics and engineering.

One of the most important techniques in linear algebra is diagonalization. This technique can be used to simplify complicated systems of linear equations. It is also a powerful tool for solving eigenvalue problems. In general, diagonalization is an essential tool for anyone who needs to work with linear equations.

Eigenvectors and eigenvalues are often introduced in the context of the diagonalization of matrices. In this context, an n-by-n matrix A is diagonalizable if and only if there exists n linearly independent eigenvectors v_i of A, $i = 1, ..., n$ such that $Av_i = \lambda_i v_i$ for some scalars λ_i which are called eigenvalues. If these conditions hold, then we say that A has n distinct eigenvectors and n distinct eigenvalues.

Diagonalization is the process of diagonalizing a matrix, which means finding a matrix that is equivalent to the original matrix but with the

eigenvalues on the diagonal. This can be done by solving for the eigenvectors and then putting them into a matrix.

The eigenvectors are the vectors that stay the same when multiplied by the matrix. They are also called the characteristic vectors or natural modes. The eigenvalues are the scalars that you get when you multiply the eigenvector by the matrix. They tell you how much the vector changes.

In order to find them, you need to set up an equation and solve for them. Once you have them, you can put them into a diagonal matrix and then diagonalize it. This is a very important tool in mathematics and physics.

Diagonalization is not always possible. For example, if the null space of a matrix is not spanned by eigenvectors, then the matrix cannot be diagonalized. However, when diagonalization is possible, it is often very useful. Diagonalization can be used to find solutions to systems of linear equations, simplify calculations, and understand the structure of matrices.

Eigenvalues and eigenvectors are important concepts in linear algebra. They can be used to diagonalize matrices, which is a powerful tool for solving problems. Eigenvalues and eigenvectors can also be used to understand the structure of matrices. In general, they are useful tools for anyone who needs to work with linear equations.

This book includes eigenvectors and eigenvalues and diagonalization of matrices. The teaching in this book is clear and concise, with plenty of opportunity for practice. The finding eigenvectors and eigenvalues chapter is particularly well explained, and the examples are helpful.

The diagonalization of matrices is also explained well, with plenty of opportunity for practice. The new exercises have been added to help reinforce the concepts learned in the book. Overall, this is an excellent resource for those who want to learn more about linear algebra.

Write a content brief for a book review on canonical forms from linear algebra. In mathematical terms, canonical forms are a way of representing a matrix in a simplified way. This book provides an in-depth introduction to the topic, starting with the basics and progressing to more complex concepts.

It is aimed at readers who have a basic understanding of linear algebra, and it covers topics such as invariant subspaces, the Jordan form, and the singular value decomposition. The book also includes worked examples and exercises to help readers develop their understanding of the material.

Invariant subspaces are key to understanding canonical forms. In short, they are mathematical objects that do not change under certain transformations. This book starts with an introduction to invariant subspaces. It explains what they are and how they can be used to simplify calculations. Invariant subspaces are a powerful tool for understanding and manipu-

lating linear transformations. With this book, you will learn how to use them to your advantage.

In mathematics, an invariant subspace of a linear mapping $T : V \rightarrow V$, i.e. from some vector space V to itself, is a subspace W of V that is preserved by T; that is, $T(W) \subseteq W$. In other words, an invariant subspace is a subspace that is in some sense "left unchanged" by the transformation. Intuitively, this means that if we take any vector in the subspace and apply the transformation to it, the result will still be in the subspace.

Invariant subspaces arise naturally in many areas of mathematics, and they can be used to great effect in studying linear transformations. For instance, the study of invariant subspaces can help us to understand the structure of a linear transformation and to decompose it into simpler parts.

In addition, the existence of an invariant subspace can often be used to simplify computations involving the transformation. Consequently, the study of invariant subspaces is a fundamental tool in many areas of mathematics.

The dimension of an invariant subspace is called the multiplicity of the eigenvalue associated with that subspace. If the multiplicity of an eigenvalue is equal to its algebraic multiplicity (i.e. if it has the same number of eigenvectors associated with it as there are linearly independent solutions to the equation), then the eigenvalue is said to be geometrically simple.

In contrast, if the multiplicity of an eigenvalue is less than its algebraic multiplicity, then the eigenvalue is said to be geometrically multiple. Geometrically multiple eigenvalues may or may not have distinct eigenspaces; if they do, then each such eigenspace is called an invariant subspace of T.

Geometrically simple eigenvalues always have distinct eigenspaces; in this case, each such eigenspace is also an invariant subspace of T. The study of invariant subspaces plays an important role in many areas of mathematics, including functional analysis, differential equations, and numerical linear algebra.

Jordan normal form is a specific type of upper triangular matrix that is used to represent linear operators on a finite-dimensional vector space. This matrix is made up of Jordan blocks, which are submatrices with identical eigenvalues that have each non-zero off-diagonal entry equal to 1.

Jordan normal form is useful because it allows for the efficient computation of matrix powers and roots. In addition, this form can be used to transform one matrix into another that is similar, meaning that they have the same eigenvalues. Jordan normal form is an important tool in linear algebra and has many applications in physics and engineering.

Jordan Canonical Form, also called Jordan normal form, is a block diagonal matrix consisting of Jordan blocks. This is not entirely unique as the order of the Jordan blocks are not fixed. The Jordan blocks are grouped together by their eigenvalues, but there is no ordering imposed among the eigenvalues or Jordan blocks. Jordan normal form can be applied to any square matrix as long as the field of coefficients contains all the eigenvalues of the matrix. This normal form is named after mathematician Camille Jordan.

Today, Jordan blocks are used in many different areas of mathematics and physics.

In this book, I provide a comprehensive introduction to invariant subspaces and Jordan canonical forms. This will include a review of the necessary linear algebra, as well as numerous examples and applications. In addition, I will discuss how these concepts can be used to simplify computations involving linear transformations.

My approach is to first introduce the necessary theory, followed by worked examples that illustrate how the theory can be applied in practice. I believe that this approach will provide readers with a strong understanding of the material covered in this book.

My approach is based on two fundamental principles: first, that students should be given a firm foundation in the core material; and second, that they should be given ample opportunity to work with the material in order to develop their understanding.

The study of invariant subspaces and Jordan canonical forms is a fundamental tool in many areas of mathematics. These concepts are important in understanding the structure of linear transformations and in simplifying computations. In addition, the existence of an invariant subspace can often be used to simplify computations involving the transformation. Consequently, the study of invariant subspaces and Jordan canonical forms is a fundamental tool in many areas of mathematics.

The determinant function can be defined by essentially two different methods. The advantage of the first definition, one which uses permutations, is that it provides an actual formula for $\det(A)$, a fact of theoretical importance. The disadvantage is that, quite frankly, computing a determinant by this method can be cumbersome.

- A **pattern** in an $n \times n$ matrix is a way to choose n entries of the matrix so that there is one chosen entry in each row and in each column of A.
- With a pattern P we associate the **product** of all its entries, denoted by prod P.
- Two entries in a pattern are said to be **inverted** if one of them is located to the right and above the other in the matrix.

- The **signature** of a pattern P is defined as $\text{sgn} P = (-1)^{(\text{number of inversions in } P)}$.

Definition 5.1. The **determinant** of an $n \times n$ matrix A is defined as

$$\det(A) = \sum (\text{sgn } P)(\text{prod } P)$$

where the sum is taken over all $n!$ patterns P.

Example 5.1. Use the definition of an $n \times n$ determinant to prove the formulas for determinants for 2×2 matrices and 3×3 matrices. We find

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21}$$

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

Example 5.2. Find the determinant of the following matrix using the definition of determinant

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

There are only two patterns with a nonzero product. Therefore

$$\det M = (-1)^8 (2 \cdot 3 \cdot 3 \cdot 2 \cdot 2) + (-1)^5 (2 \cdot 3 \cdot 1 \cdot 2 \cdot 4) = 24.$$

Lemma 5.1. *The determinant of an (upper or lower) triangular matrix A is the product of the diagonal entries of the matrix.*

Proof. Suppose

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ 0 & a_{22} & \ddots & \vdots \\ \vdots & \ddots & \ddots & a_{n-1,n} \\ 0 & \cdots & 0 & a_{nn} \end{bmatrix} \quad \text{or} \quad A = \begin{bmatrix} a_{11} & 0 & \cdots & 0 \\ a_{21} & a_{22} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ a_{n,1} & \cdots & a_{n,n-1} & a_{nn} \end{bmatrix}.$$

To have a nonzero product a pattern must contain the first component of the first column, then the second component of the second column, and

so on. Thus, only the diagonal pattern P makes a nonzero contribution. We conclude that

$$\det A = (\text{sgn } P)(\text{prod } P) = (-1)^0 a_{11} \cdots a_{nn} = a_{11} \cdots a_{nn}.$$

$\square$

Theorem 5.1. *Prove that if A and C are square matrices (not necessarily of the same size), then*

$$\det \begin{bmatrix} A & B \\ 0 & C \end{bmatrix} = \det \begin{bmatrix} A & 0 \\ B & C \end{bmatrix} = (\det A)(\det C).$$

Proof. The proof is left for the reader.

$\square$

Example 5.3. Find the determinant of the matrix

$$M = \begin{bmatrix} a_{11} & a_{12} & b_{11} & b_{12} \\ a_{21} & a_{22} & b_{21} & b_{22} \\ 0 & 0 & c_{11} & c_{12} \\ 0 & 0 & c_{21} & c_{22} \end{bmatrix}.$$

Let

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

and

$$C = \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix},$$

then

$$\det M = (\det A)(\det B) = (a_{11}a_{22} - a_{21}a_{12})(c_{11}c_{22} - c_{21}c_{12}).$$

Example 5.4. Find 2×2 matrices A, B, C, D such that

$$\det \begin{bmatrix} A & B \\ C & D \end{bmatrix} \neq (\det A)(\det D) - (\det B)(\det C).$$

The standard basis for $\mathbb{R}^{2 \times 2}$ works.

The **trace** of an $n \times n$ matrix A is defined as the sum of the diagonal entries of A and is denoted by $\text{trace}(A)$.

Theorem 5.2. *If A is a square matrix then $\det(A^T) = \det(A)$ and* $\text{trace}(A^T) = \text{trace}(A)$.

Proof. For each pattern P in A we can consider the corresponding (transposed) pattern P^T in A^T. The two patterns P and P^T involve the same numbers, and they contain the same number of inversions, but the role of the two numbers in each inversion is reversed. Therefore, the two patterns make the same contribution to the respective determinant and so, $(\text{sgn}P)(\text{prod}P) = (\text{sgn}P^T)(\text{prod}P^T)$. Therefore, we conclude that $\det(A) = \det(A^T)$.

□

Example 5.5. Find the determinant of the matrices M, N, M^T, and N^T where

$$M = \begin{bmatrix} 4 & 3 & 2 & 1 \\ 0 & 5 & 6 & 7 \\ 0 & 0 & 3 & 2 \\ 0 & 0 & 0 & 4 \end{bmatrix} \quad \text{and} \quad N = \begin{bmatrix} 0 & 0 & 0 & 8 \\ 0 & 0 & 2 & 3 \\ 0 & 7 & 6 & 5 \\ 1 & 2 & 3 & 4 \end{bmatrix}.$$

Since M and N are upper triangular and lower triangular, respectively,

$$\det M = 4 \cdot 5 \cdot 3 \cdot 4 = 240 \quad \text{and} \quad \det N = 1 \cdot 7 \cdot 2 \cdot 8 = 112.$$

Since

$$M^T = \begin{bmatrix} 4 & 0 & 0 & 0 \\ 3 & 5 & 0 & 0 \\ 2 & 6 & 3 & 0 \\ 1 & 7 & 2 & 4 \end{bmatrix} \quad \text{and} \quad N^T = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 7 & 2 \\ 0 & 2 & 6 & 3 \\ 8 & 3 & 5 & 4 \end{bmatrix}$$

and these matrices are lower triangular and upper triangular, respectively,

$$\det M = \det M^T \quad \text{and} \quad \det N = \det N^T.$$

Lemma 5.2. *Suppose A and B are $n \times n$ matrices.*

- *If B is obtained from A by dividing a row of A by a scalar k, then* $\det B = (1/k) \det(A)$.
- *If B is obtained from A by a row swap, then $\det(B) = -\det(A)$.*
- *If B is obtained from A by adding a multiple of a row of A to another row, then $\det(B) = \det(A)$.*
- *Suppose you swap rows s times as you transform A into B, and you divide various rows by the scalars $k_1, \dots, k_r$, then*

$$\det(A) = (-1)^s k_1 \cdots k_r \det(B).$$

- If s is the number of row swaps and $k_1, \ldots, k_r$ are the scalars used to divide rows to obtain pivots in producing $rref(A)$ when performing Gauss-Jordan elimination, then

$$\det(A) = (-1)^s k_1 \cdots k_r \det(rref(A)). \tag{5.1}$$

Proof. The proof is left for the reader.

$\square$

Example 5.6. Use Gauss-Jordan elimination to find the determinant of the matrix

$$M = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 & 5 \\ 1 & 3 & 6 & 10 & 15 \\ 1 & 4 & 10 & 20 & 35 \\ 1 & 5 & 15 & 35 & 70 \end{bmatrix}.$$

Using elementary row operations $R_2 \to R_2 - R_1$, $R_3 \to R_3 - R_1$, $R_4 \to R_4 - R_1$, and $R_5 \to R_5 - R_1$,

$$M \to M_1 = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 & 4 \\ 0 & 2 & 5 & 9 & 14 \\ 0 & 3 & 9 & 19 & 34 \\ 0 & 4 & 14 & 34 & 69 \end{bmatrix} \qquad \text{with} \qquad \det(M) = \det(M_1).$$

Using elementary row operations $R_3 \to -2R2 + R_3$, $R_4 \to -3R2 + R_4$, and $R_5 \to -4R2 + R_5$,

$$M_1 \to M_2 = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 & 4 \\ 0 & 0 & 1 & 3 & 6 \\ 0 & 0 & 3 & 10 & 22 \\ 0 & 0 & 6 & 22 & 53 \end{bmatrix} \qquad \text{with} \qquad \det(M_1) = \det(M_2).$$

Using elementary row operations $R_4 \to -3R3 + R_4$, $R_5 \to -6R3 + R_5$,

$$M_2 \to M_3 = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 & 4 \\ 0 & 0 & 1 & 3 & 6 \\ 0 & 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 4 & 17 \end{bmatrix} \qquad \text{with} \qquad \det(M_2) = \det(M_3).$$

Using the elementary row operation $R_5 \to -4R4 + R_5$,

$$M_3 \to M_4 = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 & 4 \\ 0 & 0 & 1 & 3 & 6 \\ 0 & 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \qquad \text{with} \qquad \det(M_3) = \det(M_4).$$

Therefore $\det(M) = \det(M_4) = 1$.

Theorem 5.3. *A square matrix A is invertible if and only if $\det A \neq 0$.*

Proof. If A is invertible, then $\mathrm{rref}(A) = I_n$, so that by 5.1 $\det(A) \neq 0$. Conversely, by contrapositive, if A is noninvertible, then the last row of $\mathrm{rref}(A)$ contains all zeros, so that $\det(\mathrm{rref}(A)) = 0$. Therefore, 5.1 shows $\det(A) = 0$.

$\square$

Example 5.7. Consider a 4×4 matrix A with rows $\vec{v}_1, \vec{v}_2, \vec{v}_3, \vec{v}_4$. If $\det(A) = 8$ find the determinant of the matrix

$$\begin{bmatrix} \vec{v}_4 \\ \vec{v}_2 + 9\vec{v}_4 \\ -\vec{v}_3 \\ \vec{v}_1 \end{bmatrix}.$$

Switching rows 1 and 4, then multiplying row 3 by -1, then adding $-9R_4$ to R_2 yields,

$$\det \begin{bmatrix} \vec{v}_4 \\ \vec{v}_2 + 9\vec{v}_4 \\ -\vec{v}_3 \\ \vec{v}_1 \end{bmatrix} = -\det \begin{bmatrix} \vec{v}_1 \\ \vec{v}_2 + 9\vec{v}_4 \\ -\vec{v}_3 \\ \vec{v}_4 \end{bmatrix} = \det \begin{bmatrix} \vec{v}_1 \\ \vec{v}_2 + 9\vec{v}_4 \\ \vec{v}_3 \\ \vec{v}_4 \end{bmatrix} = \det \begin{bmatrix} \vec{v}_1 \\ \vec{v}_2 \\ \vec{v}_3 \\ \vec{v}_4 \end{bmatrix} = 8.$$

Example 5.8. Determine whether the linear transformation given by the following linear equations is an isomorphism. If so, find the inverse transformation.

$$\begin{cases} y_1 &= x_1 + 2x_2 + 3x_3 \\ y_2 &= 3x_2 + 4x_3 \\ y_3 &= x_1 + 6x_2 + 5x_3. \end{cases}$$

The coefficient matrix is

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

We will form the matrix $[A|I_3]$ and apply row operations, thus completing two tasks at once. This process will convert the matrix A into an upper-triangular matrix and thus be able to determine the determinant of A (of thus of the linear transformation) before we actually find the inverse matrix. If at any time we see the determinant will be 0 we will stop

and decide that the linear transformation is not an isomorphism. Otherwise we continue applying row operations until we determine A^{-1}. After completing these steps we find

$$A^{-1} = \begin{bmatrix} 9/10 & -4/5 & 1/10 \\ -2/5 & -1/5 & 2/5 \\ 3/10 & 2/5 & -3/10 \end{bmatrix}.$$

Therefore the linear transformation is an isomorphism and the inverse transformation is

$$\begin{cases} x_1 &= \frac{9}{10}y_1 - \frac{4}{5}y_2 + \frac{1}{10}y_3 \\ x_2 &= -\frac{2}{5}y_1 - \frac{1}{5}y_2 + \frac{2}{5}y_3 \\ x_3 &= \frac{3}{10}y_1 + \frac{2}{5}y_2 - \frac{3}{10}y_3. \end{cases}$$

Example 5.9. The matrix defined by

$$V_n = \begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^{n-1} \\ 1 & x_2 & x_2^2 & \cdots & x_2^{n-1} \\ 1 & x_3 & x_3^2 & \cdots & x_3^{n-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^{n-1} \end{bmatrix} \tag{5.2}$$

is called the n-th **Vandermonde matrix** . Prove

$$\det(V_n) = \prod_{1 \le i < j \le n} (x_j - x_i). \tag{5.3}$$

By 5.2-(5.2), we can perform the row operations $R_i \to R_i - R_1$ for $2 \le i \le n$ leaving the value of the determinant unchanged, so

$$\det(V_n) = \begin{vmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^{n-1} \\ 1 & x_2 & x_2^2 & \cdots & x_2^{n-1} \\ 1 & x_3 & x_3^2 & \cdots & x_3^{n-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^{n-1} \end{vmatrix} = \begin{vmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^{n-1} \\ 0 & x_2 - x_1 & x_2^2 - x_1^2 & \cdots & x_2^{n-1} - x_1^{n-1} \\ 0 & x_3 - x_1 & x_3^2 - x_1^2 & \cdots & x_3^{n-1} - x_1^{n-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & x_n - x_1 & x_n^2 - x_1^2 & \cdots & x_n^{n-1} - x_1^{n-1} \end{vmatrix}$$
$$\tag{5.4}$$

By 5.2-(5.2) and 5.6-(**??**), we can perform the column operations $C_i \to C_i - x_1 C_{i-1}$ for $2 < i \le n$ leaving the value of the determinant unchanged, so

$$= \begin{vmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & x_2 - x_1 & (x_2 - x_1)x_2 & \cdots & (x_2 - x_1)x_2^{n-2} \\ 0 & x_3 - x_1 & (x_3 - x_1)x_3 & \cdots & (x_3 - x_1)x_3^{n-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & x_n - x_1 & (x_n - x_1)x_n & \cdots & (x_n - x_1)x_n^{n-2} \end{vmatrix} = \prod_{k=2}^{n} (x_k - x_1) \det(V_{n-1}).$$
$$\tag{5.5}$$

This process ends with $\det(V_2) = x_n - x_{n-1}$ and so 5.3 follows.

Example 5.10. Show the transformation

$$T(\vec{x}) = \det \begin{bmatrix} \vec{v}_1 & \cdots & \vec{v}_{i-1} & \vec{x} & \vec{v}_{i+1} & \cdots & \vec{v}_n \end{bmatrix}$$

from $\mathbb{R}^{n \times 1}$ to $\mathbb{R}$ is a linear transformation. Similarly, show the transformation

$$T(\vec{x}) = \det \begin{bmatrix} \vec{v}_1 & \cdots & \vec{v}_{i-1} & \vec{x} & \vec{v}_{i+1} & \cdots & \vec{v}_n \end{bmatrix}^T$$

from $\mathbb{R}^{1 \times n}$ to $\mathbb{R}$ is a linear transformation.

Example 5.11. Find the image and kernel of the linear transformation

$$T(\vec{x}) = \det \begin{bmatrix} \vec{x} & \vec{u} & \vec{v} \end{bmatrix}$$

from $\mathbb{R}^3$ to $\mathbb{R}$, where $\vec{u} = \begin{bmatrix} 2 \\ 0 \\ 0 \end{bmatrix}^T$ and $\vec{v} = \begin{bmatrix} 0 \\ 1 \\ 3 \end{bmatrix}^T$. Notice that

$$T(\vec{e}_1) = \det \begin{bmatrix} 1 & 2 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 3 \end{bmatrix} = 0,$$

$$T(\vec{e}_2) = \det \begin{bmatrix} 0 & 2 & 0 \\ 1 & 0 & 1 \\ 0 & 0 & 3 \end{bmatrix} = -\det \begin{bmatrix} 1 & 0 & 1 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix} = -6, \quad \text{and}$$

$$T(\vec{e}_3) = \det \begin{bmatrix} 0 & 2 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 3 \end{bmatrix} = \det \begin{bmatrix} 1 & 0 & 3 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix} = 2.$$

Therefore, the matrix of T with respect to the standard basis $\vec{e}_1, \vec{e}_2, \vec{e}_3$ is $B = \begin{bmatrix} 0 & -6 & 2 \end{bmatrix}$, and so $T(\vec{x}) = B\vec{x}$. To find the kernel we solve the system

$$\begin{bmatrix} 0 & -6 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}^T = 0.$$

Thus the kernel has basis

$$\left(\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}^T, \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}^T \right)$$

which means the dimension of the kernel of T is 2, and so by the rank nullity theorem the dimension of the image of T is 1.

Example 5.12. Consider the linear transformation

$$T(\vec{x}) = \det \begin{bmatrix} \vec{x} & \vec{v}_2 & \vec{v}_3 & \cdots & v_n \end{bmatrix}$$

from $\mathbb{R}^n$ to $\mathbb{R}$, where $\vec{v}_2, \vec{v}_3, \ldots, \vec{v}_n$ are linearly independent vectors in $\mathbb{R}^n$. Describe the image and kernel of this transformation, and determine their dimensions. Since $\vec{v}_1, \ldots, \vec{v}_n$ are linearly independent, $T(\vec{x}) = 0$ only if $\vec{x}$ is a linear combination of the $\vec{v}_i$'s, (otherwise the matrix $\begin{bmatrix} \vec{x} & \vec{v}_1 & \cdots & \vec{v}_n \end{bmatrix}$ is invertible, and $T(\vec{x}) \neq 0$). Hence, the kernel of T is the span of $\vec{v}_2, \ldots, \vec{v}_n$, an $(n-1)$-dimensional subspace of $\mathbb{R}^n$. The image of T is the real line $\mathbb{R}$ (since it must be 1-dimensional).

Example 5.13. For a fixed positive integer n, let D be a function which assigns to any $n \times n$ matrix A a number $D(A)$ such that

- D is linear in the rows,
- $D(B) = -D(A)$ if B is obtained from A by a row swap, and
- $D(I_n) = 1$.

Show that $D(A) = \det(A)$ for all $n \times n$ matrices A. First we comment that if a square matrix A has two equal rows, then $D(A) = 0$. Indeed, if we swap two equal rows and call the resulting matrix B, then $B = A$, so that $D(A) = D(B) = -D(A)$, by property (ii), and $D(A) = 0$ as claimed.

Next we need to understand how the elementary row operations affect D. Properties (i) and (ii) tell us about row multiplication and row swaps, but we still need to think about row additions. We will show that if B is obtained from A by adding k times the ith row to the jth, then $D(B) = D(A)$, Let's label the row vectors of A by $\vec{v}_1, ., ., \vec{v}_n$. By linearity of D in the jth row (i) we have

$$D(B) = D\left(\begin{bmatrix} \vdots \\ \vec{v}_i \\ \vdots \\ \vec{v}_j + k\vec{v}_i \\ \vdots \end{bmatrix}\right) = D(A) + kD\left(\begin{bmatrix} \vdots \\ \vec{v}_i \\ \vdots \\ \vec{v}_i \\ \vdots \end{bmatrix}\right) = D(A).$$

Note that in the last step we have used the remark made at the beginning. Now we can write

$$D(A) = (-1)^s k_1 \cdots k_r D(\mathrm{rref}(A))$$

where s is the number of row swaps and $k_1, \ldots, k_r$ are scalars used to obtain the pivots in the Gaussian reduction.

Next we observe that if $D(\mathrm{rref}(A)) = \det(\mathrm{rref}(A))$ for all square matrices A. Indeed, if A is invertible, then $\mathrm{rref}(A) = I_n$, and $D(I_n) = 1 = \det(I_n)$ by property (iii). If A fails to be invertible then

$$D(\mathrm{rref}(A)) = 0 = \det(\mathrm{rref}(A)$$

by linearity in the last row.

It follows that

$$D(A) = (-1)^s k_1 \cdots k_r D(\mathrm{rref}(A)) = (-1)^s k_1 \cdots k_r \det(\mathrm{rref}(A)) = \det(A)$$

as desired.

::: {#thm- } [Properties of Determinant and Trace]

- If A and B are square matrices, then $\det(AB) = (\det A)(\det B)$.
- If A and B are square matrices of the same size, then $\mathrm{trace}(AB) = \mathrm{trace}(BA)$ and $\mathrm{trace}(A + B) = \mathrm{trace}(A) + \mathrm{trace}(B)$.
- If A and B are similar matrices, then $\det(A) = \det(B)$ and $\mathrm{trace}(A) = \mathrm{trace}(B)$.
- If A is an invertible matrix, then $\det(A^{-1}) = (\det A)^{-1}$. :::

Proof. The proof of each part follows.

- Suppose $A = \begin{bmatrix} a_{ij} \end{bmatrix}$ and $B = \begin{bmatrix} b_{ij} \end{bmatrix}$. The ith term on the diagonal of $A + B$ is $a_{i,i} + b_{i,i}$, thus

$$\mathrm{trace}(A+B) = \sum_{i=1}^{n}(a_{i,i}+b_{i,i}) = \sum_{i=1}^{n} a_{i,i}+\sum_{i=1}^{n} b_{i,i} = \mathrm{trace}(A)+\mathrm{trace}(B).$$

  The ith term on the diagonal of AB is $\sum_{j=1}^{n} a_{i,j}b_{j,i}$ and the ith term on the diagonal of BA is $\sum_{i=1}^{n} b_{j,i}a_{i,j}$. Therefore,

$$\mathrm{trace}(AB) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{i,j}b_{j,i} = \sum_{i=1}^{n}\sum_{j=1}^{n} b_{j,i}a_{i,j} = \mathrm{trace}(BA)$$

  as desired.
- If matrices A and B are similar there exists an invertible matrix P such that $B = P^{-1}AB$. Then,

$$\mathrm{trace}(P^{-1}AP) = \mathrm{trace}(P^{-1}(AP)) = \mathrm{trace}((AP)P^{-1}) = \mathrm{trace}(A).$$

$\square$

Example 5.14. Consider an $n \times n$ matrix A such that both A and A^{-1} have integer entries. What are the possible values of $\det(A)$? Applying the determinant to the equation $AA^{-1} = I_n$, it follows $\det(A)\det(A^{-1}) = 1$. The only way the product of the two integers $\det(A)$ and $\det(A^{-1})$ can be 1 is when they are both 1 or both -1. Therefore, $\det(A) = \pm 1$.

Consider the $(n-1) \times (n-1)$ matrix $M_{i,j}$ obtained from an $n \times n$ matrix A by deleting the i-th row and j-th column. Then the determinant of $M_{i,j}$ is called the (i, j)-th **minor** of A.

Theorem 5.4 (Laplace Expansion). *The determinant of an $n \times n$ matrix $A = [a_{i,j}]$ can be evaluated by using either one of the following equations:*

$$\det(A) = (-1)^{i+1} a_{i,1} \det(M_{i,1}) + \cdots + (-1)^{i+n} a_{i,n} \det(M_{i,n}) \qquad (5.6)$$

where i is any integer such that $1 \le i \le n$, or

$$\det(A) = (-1)^{1+j} a_{1,j} \det(M_{1,j}) + \cdots + (-1)^{n+j} a_{n,j} \det(M_{n,j}) \qquad (5.7)$$

where j is any integer such that $1 \le j \le n$.

Proof. The proof is left for the reader.

$\square$

Example 5.15. Evaluate the determinant of the matrix

$$A = \begin{bmatrix} 3 & 2 & 6 & 4 \\ 1 & -2 & -3 & 1 \\ 0 & 2 & 3 & 8 \\ 4 & -1 & 7 & 2 \end{bmatrix}$$

by using expansion of minors along the second row. Expanding along the second row, by using 5.6 with $i = 2$ and $n = 4$

$$\det(A) = -\det(M_{2,1}) - 2\det(M_{2,2}) + 3\det(M_{2,3}) + \det(M_{2,4})$$
$$= -(-104) - 2(-6) + 3(68) + 27 = 347.$$

The trace and determinant share come common properties, for example, they both are invariants of a linear transformation. To be more explicit, let T be a linear transformation from a linear space V to V, where V is a finite-dimensional linear space. If $\mathcal{B}$ is a basis of V and B is the $\mathcal{B}$-matrix of T, then we define the **determinant** and **trace** of the linear transformation T as $\det(T) = \det(B)$ and $\text{trace}(T) = \text{trace}(B)$, respectively.

Theorem 5.5. *The determinant and trace of a linear transformation are independent of the basis chosen.*

Example 5.16. Compute the determinant of the linear transformation $T(f(t)) = f(3t - 2)$ from $\mathcal{P}_2$ to $\mathcal{P}_2$ and use it to determine whether the linear transformation is an isomorphism. We choose to use the standard basis of $\mathcal{P}_2$, namely $(1, t, t^2)$. With respect to this basis the matrix of T is $A = \begin{bmatrix} 1 & -2 & 4 \\ 0 & 3 & -12 \\ 0 & 0 & 9 \end{bmatrix}$, so that $\det(T) = \det(A) = 27$. Therefore, the given transformation is an isomorphism.

Example 5.17. Find the determinant of the linear transformation

$$T(M) = \begin{bmatrix} 1 & 2 \\ 2 & 3 \end{bmatrix} M + M \begin{bmatrix} 1 & 2 \\ 2 & 3 \end{bmatrix}$$

from the space V of symmetric 2×2 matrices to V and use it to determine whether the linear transformation is an isomorphism. A basis for the space of symmetric 2×2 matrices is

$$\left(\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right).$$

The matrix of T with respect to this basis is

$$A = \begin{bmatrix} 2 & 4 & 0 \\ 2 & 4 & 2 \\ 0 & 4 & 6 \end{bmatrix}$$

so that $\det(T) = \det(A) = -16$. Therefore, the given transformation is an isomorphism.

::: {#thm- } Cramer

Cramer's rule can be used to prove the Cayley Hamilton theorem of linear algebra, as well as Nakayama's lemma, which is fundamental in commutative ring theory.

5.7   Cramer

Consider the system of equations $A\vec{x} = \vec{b}$ where A is an $n \times n$ invertible matrix. Then the unique solution to the system is given by

$$x_1 = \frac{\det(A_1)}{\det(A)}, \quad x_2 = \frac{\det(A_2)}{\det(A)}, \cdots, \quad x_n = \frac{\det(A_n)}{\det(A)}$$

where the matrix A_i is the matrix A with the i-th column replaced by $\vec{b}$.
:::

Proof. The proof is left for the reader.

$\square$

Example 5.18. Solve the system using Cramer's Rule.

$$\begin{cases} 2x_1 + 3x_2 - x_3 & = 1 \\ 4x_1 + x_2 + 2x_3 & = 5 \\ x_1 - x_2 + x_3 & = 2 \end{cases}$$

First we compute the determinant of the coefficient matrix A, and find $\det(A) = 5$ Since $\det(A) \neq 0$, we can apply Cramer's rule, so we compute the required determinants and obtain, according to Cramer's rule, the solution to the system is

$$x_1 = \frac{\det(A_1)}{\det(A)} = \frac{7}{5}, \quad x_2 = \frac{\det(A_2)}{\det(A)} = -\frac{3}{5}, \quad x_3 = \frac{\det(A_n)}{\det(A)} = 0.$$

Example 5.19. Use paper, pencil, and Cramer's rule to solve the system $A\vec{x} = \vec{b}$ where A and $\vec{b}$ are the following matrices

$$A = \begin{bmatrix} 1 & 2 & 3 & 4 \\ -5 & -6 & -7 & -8 \\ 9 & -10 & 11 & -12 \\ -13 & 14 & -15 & 16 \end{bmatrix} \quad \text{and} \quad \vec{b} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}.$$

First we find $\det(A) = -256$ and then we find the solution vector

$$\vec{x} = \begin{bmatrix} -51/8 \\ 5 \\ 41/8 \\ -9/2 \end{bmatrix}.$$

Chapter 6

Eigenvalues and Eigenvectors

Let $\mathbb{F}$ be either the real numbers or the complex numbers. A nonzero vector $\vec{v}$ in $\mathbb{F}^n$ is called an of an $n \times n$ matrix A if $A\vec{v}$ is a scalar multiple of $\vec{v}$, that is $A\vec{v} = \lambda\vec{v}$ for some scalar λ. Note that this scalar λ may be zero. The scalar λ is called the **eigenvalue** associated with the eigenvector $\vec{v}$. Even though, $A\vec{0} = \lambda\vec{0}$ we do not call $\vec{0}$ an eigenvector. Of course a matrix need not have any eigenvalues or eigenvectors, but notice if $\vec{v}$ is an eigenvector of matrix A, then $\vec{v}$ is an eigenvector of matrices A^2, A^3, ...as well, with $A^t\vec{v} = \lambda^t\vec{v}$, for all positive integers t. If $\mathbb{F} = \mathbb{C}$, then counting multiplicities, every $n \times n$ matrix has exactly n eigenvalues.

If $\vec{v}$ is an eigenvector of the $n \times n$ matrix A with associated eigenvalue λ, what can you say about $\ker(A - \lambda I_n)$? Is the matrix $A - \lambda I_n$ invertible? We know $A\vec{v} = \lambda\vec{v}$ so $(A - \lambda I_n)\vec{v} = A\vec{v} - \lambda I_n\vec{v} = \lambda\vec{v} - \lambda\vec{v} = 0$. Thus a nonzero vector $\vec{v}$ is in the kernel of $(A - \lambda I_n)$. Therefore, $\ker(A - \lambda I_n) \neq \{\vec{0}\}$ and so $A - \lambda I_n$ is not invertible.

Lemma 6.1. *Let A be an $n \times n$ matrix A and λ a scalar. Then λ is an eigenvalue of A if and only if $\det(A - \lambda I_n) = 0$.*

Proof. The proof follows from the chain of equivalent statements:

- λ is an eigenvalue of A,
- there exists a nonzero vector $\vec{v}$ such that $(A - \lambda I_n)\vec{v} = 0$,
- $\ker(A - \lambda I_n) \neq \{\vec{0}\}$,
- matrix $A - \lambda I_n$ fails to be invertible, and
- $\det(A - \lambda I_n) = 0$.

□

Example 6.1. Find all eigenvectors and eigenvalues of the identity matrix I_n. Since $I_n \vec{v} = \lambda \vec{v} = 1\vec{v}$ for all $\vec{v} \in \mathbb{R}^n$, all nonzero vectors in $\mathbb{R}^n$ are eigenvectors of I_n, with eigenvalues $\lambda = 1$.

Lemma 6.2. *The eigenvalues of a triangular matrix are its diagonal entries.*

Proof. Let A be a triangular matrix. Then $A - \lambda I_n$ is also a triangular matrix, and so $\det(A - \lambda I_n)$ is the product of its diagonal entries. Let a_{ii} be any diagonal entry of A. Then $a_{ii} - \lambda$ is the corresponding diagonal entry of $A - \lambda I_n$. Thus λ is an eigenvalue of A if and only if $a_{ii} - \lambda = 0$ by 6.1.

□

Example 6.2. Find a basis of the linear space V of all 2×2 matrices for which $\vec{e}_1$ is an eigenvector. For an arbitrary 2×2 matrix we want

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} a \\ c \end{bmatrix} = \begin{bmatrix} \lambda \\ 0 \end{bmatrix} = \lambda \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

for any λ. Hence a, b, d are free and $c = 0$; thus a desired basis of V is

$$\left(\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right).$$

Example 6.3. Find a basis of the linear space V of all 4×4 matrices for which $\vec{e}_2$ is an eigenvector. We want to find all 4×4 matrices A such that $A\vec{e}_2 = \lambda e_2$. Thus the second column of an arbitrary 4×4 matrix A must be of the form $\begin{bmatrix} 0 \\ \lambda \\ 0 \\ 0 \end{bmatrix}^T$, so

$$A = \begin{bmatrix} a & 0 & c & d \\ e & \lambda & f & g \\ h & 0 & i & j \\ k & 0 & l & m \end{bmatrix}.$$

Let E_{ij} denote the 4×4 matrix with all entries zero except for a 1 in the i-th row and j-th column. Then a basis for V is

$$(E_{11}, E_{21}, E_{31}, E_{41}, E_{22}, E_{13}, E_{23}, E_{33}, E_{34}, E_{41}, E_{42}, E_{43}, E_{44})$$

and so the dimension of V is 13.

Example 6.4. Find the eigenvalues and find a basis for each eigenspace given $A = \begin{bmatrix} 1 & 0 & 0 \\ -5 & 0 & 2 \\ 0 & 0 & 1 \end{bmatrix}$. Find an eigenbasis for A. The eigenvalues are

$\lambda_1 = 0$ and $\lambda_2 = \lambda_3 = 1$. A basis for E_0 is $\left(\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right)$. A basis for E_1 is

$\left(\begin{bmatrix} 1 \\ -5 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix} \right)$. An eigenbasis for A is $\left(\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ -5 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix} \right)$.

Example 6.5. Find a basis of the linear space V of all 2×2 matrices A for which $\begin{bmatrix} 1 \\ -3 \end{bmatrix}$ is an eigenvector. For an arbitrary 2×2 matrix we want

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} 1 \\ -3 \end{bmatrix} = \lambda \begin{bmatrix} 1 \\ -3 \end{bmatrix} = \begin{bmatrix} \lambda \\ -3\lambda \end{bmatrix}.$$

Thus $a - 3b = \lambda$, $c - 3d = -3\lambda$ and so $c = -3a + 9b + 3d$. Thus A must be of the form

$$\begin{bmatrix} a & b \\ -3a + 9b + 3d & d \end{bmatrix} = a \begin{bmatrix} 1 & 0 \\ -3 & 0 \end{bmatrix} + b \begin{bmatrix} 0 & 1 \\ 9 & 0 \end{bmatrix} + d \begin{bmatrix} 0 & 0 \\ 3 & 1 \end{bmatrix}.$$

Thus a basis of V is

$$\left(\begin{bmatrix} 1 & 0 \\ -3 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 9 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 3 & 1 \end{bmatrix} \right)$$

and so the dimension of V is 3.

Example 6.6. Find a basis of the linear space V of all 3×3 matrices A for which both $\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}^T$ and $\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}^T$ are eigenvectors. Since $A \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}^T$ is simply the first column of A, the first column must be a multiple of $\vec{e}_1$. Similarly, the third column must be a multiple of $\vec{e}_3$. There are no other restrictions on the form of A, meaning it can be any matrix of the form

$$\begin{bmatrix} a & b & 0 \\ 0 & c & 0 \\ 0 & d & e \end{bmatrix}$$

Thus a basis of V is

$$\left(\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \right)$$

and so the dimension of V is 5.

Theorem 6.1. *If A is an $n \times n$ matrix, then $\det(A - \lambda I_n)$ is a polynomial of degree n, of the form*

$$f_A(\lambda) = (-\lambda)^n + trace(A)(-\lambda)^{n-1} + \cdots + \det(A). \qquad (6.1)$$

Proof. This proof is left for the reader.

$\square$

The equation $\det(A - \lambda I_n) = 0$ is called the **characteristic equation** of A. The polynomial in 6.1 is called the **characteristic polynomial** and is denoted by $f_A(\lambda)$. We say that an eigenvalue λ_0 of a square matrix A has **algebraic multiplicity** k if λ_0 is a root of multiplicity k of the characteristic polynomial $f_A(\lambda)$ meaning that we can write

$$f_A(\lambda) = (\lambda_0 - \lambda)^k g(\lambda)$$

for some polynomial $g(\lambda)$ with $g(\lambda_0) \neq 0$.

Example 6.7. Find the characteristic equation for a 2×2 matrix A. The characteristic equation of $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is

$$f_A(\lambda) = \det \begin{bmatrix} a - \lambda & b \\ c & d - \lambda \end{bmatrix} = \lambda^2 - (a + d)\lambda + (ad - bc) = 0.$$

Example 6.8. Use the characteristic polynomial $f_A(\lambda)$ to determine the eigenvalues and their multiplicities of

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

The characteristic equation is $f_A(\lambda) = -\lambda^2(\lambda + 3)$. So $\lambda_1 = 0$ with algebraic multiplicity of 2 and $\lambda_2 = -3$ with algebraic multiplicity of 1 are the eigenvalues of A.

Example 6.9. Consider the matrix $A = \begin{bmatrix} a & b \\ b & c \end{bmatrix}$, where a, b, c are nonzero constants. For which values of a, b, c does A have two distinct eigenvalues? The characteristic equation is $f_A(\lambda) = \lambda^2 - (a + c)\lambda + (ac - b^2)$. The discriminant of this quadratic equation is

$$(a + c)^2 - 4(ac - b^2) = a^2 + 2ac + c^2 - 4ac + 4b^2 = (a - c)^2 + 4b^2.$$

The discriminant is always positive since $b \neq 0$. Thus, the matrix A there will always have two distinct real eigenvalues.

Example 6.10. In terms of eigenvalues of A, which 2×2 matrices A does there exist an invertible matrix S such that $AS = SD$, where $D = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}$? If we let $S = \begin{bmatrix} \vec{v}_1 & \vec{v}_2 \end{bmatrix}$, then $AS = \begin{bmatrix} A\vec{v}_1 & A\vec{v}_2 \end{bmatrix}$ and $SD = \begin{bmatrix} 2\vec{v}_1 & 3\vec{v}_2 \end{bmatrix}$. So that $\vec{v}_1$ must be an eigenvector of A with eigenvalue 2, and $\vec{v}_2$ must be an eigenvector of A with eigenvalue 3. Thus, the matrix S will exist and will have first column has an eigenvector of A with eigenvalue 2, and have second column is an eigenvector of A with eigenvalue of 3. Therefore, A can be any matrix satisfying these requirements.

Example 6.11. Let A be a matrix with eigenvalues $\lambda_1, \dots, \lambda_k$.

- Show the eigenvalues of A^T are $\lambda_1, \dots, \lambda_k$.
- Show the eigenvalues of αA are $\alpha\lambda_1, \dots, \alpha\lambda_k$.

Show A^{-1} exists if and only if $\lambda_1 \cdots \lambda_k \neq 0$. - Also, show that if A^{-1} exists then its eigenvalues are $1/\lambda_1, \dots, 1/\lambda_k$.

Example 6.12. Let A be a matrix with eigenvalues $\lambda_1, \dots, \lambda_k$ and let m be a positive integer. Show that the eigenvalues of A^m are $\lambda_1^m, \dots, \lambda_k^m$.

Example 6.13. By using the matrix

$$\begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ \frac{-a_n}{a_0} & \frac{-a_{n-1}}{a_0} & \frac{-a_{n-2}}{a_0} & \cdots & \frac{-a_1}{a_0} \end{bmatrix}$$

Show that any given polynomial $a_0\lambda^n + a_1\lambda^{n-1} + \cdots + a_{n-1}\lambda + a_n$ where $a_0 \neq 0$, of degree n may be regarded as the characteristic polynomial of a matrix of order n. This matrix is called the **companion matrix** of the given polynomial.

Example 6.14. Let A and B be $n \times n$ matrices. Show that AB and BA have the same eigenvalues.

Example 6.15. Let A and B be real $n \times n$ matrices with distinct eigenvalues. Prove that $AB = BA$ if and only if A and B have the same eigenvectors.

Example 6.16. Prove that the characteristic polynomial of the block-triangular matrix $A = \begin{bmatrix} B & C \\ 0 & D \end{bmatrix}$ is the product of the characteristic polynomials of B and D.

Example 6.17. Suppose that A is an invertible $n \times n$ matrix. Prove that

$$f_{A^{-1}}(x) = (-x)^n \det(A^{-1}) f_A \left(\frac{1}{x} \right). \tag{6.2}$$

Example 6.18. Let A be an $n \times n$ matrix. Prove that A and A^T have the same characteristic polynomial and hence the same eigenvalues.

If λ is an eigenvalue of an $n \times n$ matrix A, then the kernel of the matrix $A - \lambda I_n$ is called the **eigenspace** associated with λ and is denoted by E_λ. The dimension of the eigenspace is called the **geometric multiplicity** of eigenvalue λ. In other words, the geometric multiplicity is the nullity of the matrix $A - \lambda I_n$.

Theorem 6.2. *Let A be an $n \times n$ matrix. If $\lambda_1, \dots, \lambda_k$ are distinct eigenvalues of A, and $\vec{v}_1, \dots, \vec{v}_k$ are any nonzero eigenvectors associated with these eigenvalues respectively, then $\vec{v}_1, \dots, \vec{v}_k$ are linearly independent.*

Proof. Suppose there exists constants $c_1, \dots, c_k$ such that

$$c_1 \vec{v}_1 + \cdots + c_k \vec{v}_k = 0 \tag{6.3}$$

Using the fact that $A\vec{v}_i = \lambda_i \vec{v}_i$ we multiply 6.3 by A to obtain

$$c_1 \lambda_1 \vec{v}_1 + \cdots c_k \lambda_k \vec{v}_k = 0. \tag{6.4}$$

Repeating this again we obtain

$$c_1 \lambda_1^2 \vec{v}_1 + \cdots + c_k \lambda_k^2 \vec{v}_k = 0. \tag{6.5}$$

Repeating, we are lead to the system in the vector unknowns $\vec{v}_1, \dots, \vec{v}_k$

$$\begin{bmatrix} c_1\vec{v}_1 & \cdots & c_k\vec{v}_k \end{bmatrix}_{n \times k} \begin{bmatrix} 1 & \lambda_1 & \lambda_1^2 & \cdots \lambda_1^{k-1} \\ 1 & \lambda_2 & \lambda_2^2 & \cdots \lambda_2^{k-1} \\ 1 & \lambda_3 & \lambda_3^2 & \cdots \lambda_3^{k-1} \\ \vdots & \vdots & \vdots & \ddots \vdots \\ 1 & \lambda_k & \lambda_k^2 & \cdots \lambda_k^{k-1} \end{bmatrix}_{k \times k} = 0_{n \times k}. \qquad (6.6)$$

Since the eigenvalues are distinct, the coefficient matrix is an invertible Vandermonde matrix. Multiplying on the right by its inverse shows that

$$\begin{bmatrix} c_1\vec{v}_1 & \cdots & c_k\vec{v}_k \end{bmatrix} = 0_{n \times k}.$$

It follows that every c_i must be zero. Hence $\vec{v}_1, \dots, \vec{v}_k$ are linearly independent.

□

A basis of $\mathbb{F}^n$ consisting of eigenvectors of A is called an **eigenbasis** for A.

In particular, if an $n \times n$ matrix A has n distinct eigenvalues, then there exists an eigenbasis for A, namely, construct an eigenbasis by finding an eigenvector for each eigenvalue.

Example 6.19. Find the characteristic equation, the eigenvalues, and a basis for the eigenspace.

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

The eigenvalues are $\lambda_1 = 8$ (with algebraic multiplicity 1) and $\lambda_2 = -1$ (with algebraic multiplicity 2) since

$$\det(A - \lambda I) = \begin{vmatrix} 3 - \lambda & 2 & 4 \\ 2 & -\lambda & 2 \\ 4 & 2 & 3 - \lambda \end{vmatrix} = -\lambda^3 + 6\lambda^2 + 15\lambda + 8 = 0.$$

For $\lambda_1 = 8$ we obtain

$$\begin{bmatrix} -5 & 2 & 4 \\ 2 & -8 & 2 \\ 4 & 2 & -5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

and we obtain the eigenvector $\vec{v}_1 = \begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix}^T$ with $E_8 = \text{span}(\vec{v}_1)$. Therefore

the geometric multiplicity of $\lambda_1 = 8$ is 1. For $\lambda_1 = -1$ we obtain

$$\begin{bmatrix} 4 & 2 & 4 \\ 2 & 1 & 2 \\ 4 & 2 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

and we obtain the eigenvectors $\vec{v}_2 = \begin{bmatrix} 1 \\ -2 \\ 0 \end{bmatrix}^T$ and $\vec{v}_3 = \begin{bmatrix} 0 \\ -2 \\ 1 \end{bmatrix}^T$ with

$E_{-1} = \mathrm{span}(\vec{v}_2, \vec{v}_3)$. Therefore the geometric multiplicity of $\lambda_2 = -1$ is 2.

Example 6.20. Show that for each of the following matrices, $\lambda = 3$ is an eigenvalue of algebraic multiplicity 4. In each case, compute the geometric multiplicity of λ.

$$\begin{bmatrix} 3 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix} \quad \begin{bmatrix} 3 & 1 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix} \quad \begin{bmatrix} 3 & 1 & 0 & 0 \\ 0 & 3 & 1 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix} \quad \begin{bmatrix} 3 & 1 & 0 & 0 \\ 0 & 3 & 1 & 0 \\ 0 & 0 & 3 & 1 \\ 0 & 0 & 0 & 3 \end{bmatrix}$$

Theorem 6.3. *Similar matrices A and B have the same determinant, trace, characteristic polynomial, rank, nullity, and the same eigenvalues with the same algebraic multiplicities.*

Proof. The case for the determinant and trace are proven in 5.6. Since A and B are similar, there exists an invertible matrix P such that $B = P^{-1}AP$. Using 5.6 we find

$$\det(B - \lambda I) = \det(P^{-1}AP - \lambda I) = \det(P^{-1}AP - P^{-1}\lambda IP) = \det(P^{-1}(A - \lambda I)P)$$
$$= \det(P^{-1})\det(A - \lambda I)\det(P) = \det(P^{-1})\det(P)\det(A - \lambda I)$$
$$= \det(P^{-1}P)\det(A - \lambda I) = \det(A - \lambda I)$$

Thus A and B have the same characteristic equation. Therefore also the same eigenvalues with the same algebraic multiplicities according to 6.1 and 6.1.

$\square$

In light of 6.3, if T is a linear transformation from V to V then a scalar λ is called an **eigenvalue** of T if there exists a nonzero element $\vec{v}$ in V such that $T(\vec{v}) = \lambda\vec{v}$. Assuming V is finite-dimensional then a basis $\mathcal{D}$ of V consisting of eigenvectors of T is called an **eigenbasis** for T.

Theorem 6.4. *Let T be a linear transformation on a finite-dimensional vector space V, and let λ be an eigenvalue of T. The geometric multiplicity of λ is less than or equal to the algebraic multiplicity of λ.*

Proof. Let k represent the geometric multiplicity of λ and assume $\dim V = n$. First notice, by definition, the eigenspace E_λ must contain at least one nonzero vector, and thus $k = \dim E_\lambda \geq 1$. Choose a basis $\vec{v}_1, \ldots, \vec{v}_k$ for E_λ and, by 2.17, extend it to a basis $\mathcal{B} = (\vec{v}_1, \ldots, \vec{v}_k, \vec{v}_{k+1}, \ldots, \vec{v}_n)$ of V. For $1 \leq i \leq k$, notice

$$[T(\vec{v}_i)]_\mathcal{B} = [\lambda \vec{v}_i]_\mathcal{B} = \lambda[\vec{v}_i]_\mathcal{B} = \lambda \vec{e}_i.$$

Thus the matrix representation for T with respect to $\mathcal{B}$ has the form

$$B = \begin{bmatrix} \lambda I_k & C \\ 0 & D \end{bmatrix} \tag{6.7}$$

where C is a $k \times (n-k)$ submatrix, O is an $(n-k) \times k$ zero submatrix, and D is an $(n-k) \times (n-k)$ submatrix.

Using 5.1 we determine the characteristic polynomial of T

$$f_T(x) = |xI_n - B| = \left| xI_n - \begin{bmatrix} \lambda I_k & C \\ 0 & D \end{bmatrix} \right| = \left| \begin{matrix} (x-\lambda)I_k & C \\ 0 & xI_{n-k} - D \end{matrix} \right| = (x-\lambda)^k f_D(x).$$

It follows that $f_T(x) = (x - \lambda)^{k+m} g(x)$ where $g(\lambda) \neq 0$ and m is the number of factors of $x - \lambda$ in $f_D(x)$. Hence $k \leq k + m$ leading to the desired conclusion.

$\square$

Example 6.21. Let $T(M) = M - M^T$ be a linear transformation from $\mathbb{R}^{2\times2}$ to $\mathbb{R}^{2\times2}$. For each eigenvalue find a basis for the eigenspace and state the geometric multiplicity. Since $A = A^T$ for every symmetric matrix, we notice $T(M) = M - M^T = M - M = 0$ whenever M is a symmetric matrix. Thus the nonzero symmetric matrices are eigenmatrices with eigenvalue 0.

Also notice the nonzero skew-symmetric matrices have eigenvalue 2 since $L(M) = M - M^T = M + M = 2M$. For eigenvlaue $\lambda = 0$ we have eigenspace E_0 with basis

$$\left(\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right).$$

This follow from the condition $A = A^T$ in $\mathbb{R}^{2\times2}$. Therefore the geometric multiplicity of $\lambda = 0$ is 3. For eigenvalue $\lambda = 2$ we have eigenspace E_2 with basis

$$\left(\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \right).$$

which follows from the condition $A = -A^T$ in $\mathbb{R}^{2\times 2}$. Therefore the geometric multiplicity of $\lambda = 2$ is 1. By 6.2 we have an eigenbasis

$$\left(\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \right).$$

for T.

6.1   Diagonalization

An $n \times n$ matrix A is called **diagonalizable** if A is similar to some diagonal matrix D. If the matrix of a linear transformation T with respect to some basis is diagonal then we call T **diagonalizable** .

Theorem 6.5. *An $n \times n$ matrix A is diagonalizable if and only if it has n linearly independent eigenvectors. In that case, the diagonal matrix D is similar to A and is given by*

$$D = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix} \tag{6.8}$$

where $\lambda_1, ..., \lambda_n$ are the eigenvalues of A. If C is a matrix whose columns are linearly independent eigenvectors of A, then $D = C^{-1}AC$.

Proof. The proof is left for the reader.

$\square$

Corollary 6.1. *Let T be a linear transformation given by $T(\vec{x}) = A\vec{x}$ where A is a square matrix. If $\mathcal{D} = (\vec{v}_1, ..., \vec{v}_n)$ is an eigenbasis for T, with $A\vec{v}_i = \lambda_i \vec{v}_i$, then the $\mathcal{D}$-matrix D of T given in 6.8 is $D = [\vec{v}_1, ..., \vec{v}_n]^{-1} A [\vec{v}_1, ..., \vec{v}_n]$.*

Proof. The proof follows from 6.5 and 6.2.

$\square$

Corollary 6.2. *A matrix A is diagonalizable if andy only if there exists an eigenbasis for A. In particular, if an $n \times n$ matrix A has n distinct eigenvalues, then A is diagonalizable.*

Proof. The proof follows from 6.5 and 6.2.

$\square$

Example 6.22. Let $T : P_2 \to P_2$ be the linear transformation defined by

$$T(a_0 + a_1 x + a_2 x^2) = (a_0 + a_1 + a_2) + (a_1 + a_2)x + a_2 x^2.$$

Show that T is not diagonalizable. The matrix of T with respect to the usual basis $(, x, x^2)$ for P_2 is easily seen to be

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

The characteristic polynomial is $f_A(x) = -(x - 1)^3$ since A is upper triangular. So T has only one (repeated) eigenvalue $\lambda = 1$. A nonzero polynomial g with $g(x) = a_0 + a_1 x + a_2 x^2$ is an eigenvector if and only if

$$\begin{bmatrix} 0 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

Thus $a_1 = 0$ and $a_2 = 0$, so there is only one linearly independent eigenvector for $\lambda = 1$. Thus T is not diagonalizable by 6.2.

Example 6.23. Let $T : P_2 \to P_2$ be the linear transformation defined by

$$T(f(x)) = x^2 f''(x) + (3x - 2)f'(x) + 5f(x). \tag{6.9}$$

Find a basis for P_2 such that the matrix representation of T with respect to B is diagonal. Since $T(x^2) = 13x^2 - 4x$, $T(x) = 8x - 2$, and $T(1) = 5$ the matrix representation of T with respect to the basis $B = (x^2, x, 1)$ is

$$A = \begin{bmatrix} 13 & 0 & 0 \\ -4 & 8 & 0 \\ 0 & -2 & 5 \end{bmatrix}.$$

Hence

$$f_T(x) = f_A(x) = \begin{vmatrix} x - 13 & 0 & 0 \\ 4 & x - 8 & 0 \\ 0 & 2 & x - 5 \end{vmatrix} = (x - 13)(x - 8)(x - 5). \tag{6.10}$$

The eigenvalues of T are $\lambda_1 = 13$, $\lambda_2 = 8$ and $\lambda_3 = 5$. Solving each of the homogenous systems $(A - 13I_3)\vec{x} = \vec{0}$, $(A - 8I_3)\vec{x} = \vec{0}$, and $(A - 5I_3)\vec{x} = \vec{0}$ yields the eigenvectors $\vec{v}_1 = 5x^2 - 4x + 1$, $\vec{v}_2 = 3x - 2$, and

$\vec{v}_3 = 1$, respectively. Notice $\vec{v}_1, \vec{v}_2, \vec{v}_3$ are 3 linearly independent vectors, so by 6.5, T is diagonalizable. We let $\mathcal{D} = (\vec{v}_1, \vec{v}_2, \vec{v}_3)$ and since $T(\vec{v}_1) = 13\vec{v}_1, T(\vec{v}_2) = 8\vec{v}_2$, and $T(\vec{v}_3) = 5\vec{v}_3$, the matrix representation of T with respect to $\mathcal{D}$ is the diagonal matrix

$$\begin{bmatrix} 13 & 0 & 0 \\ 0 & 8 & 0 \\ 0 & 0 & 5 \end{bmatrix}$$

according to 6.5.

Example 6.24. Let $T : \mathcal{P}_3 \to \mathcal{P}_3$ be the linear transformation defined by

$$T(f(x)) = xf'(x) + f(x+1). \tag{6.11}$$

Find a basis for $\mathcal{P}_3$ such that the matrix representation of T with respect to $\mathcal{B}$ is diagonal. Since $T(x^3) = 4x^3 + 3x^2 + 3x + 1$, $T(x^2) = 3x^2 + 2x + 1$, $T(x) = 2x + 1$, and $T(1) = 1$ the matrix representation of T with respect to the basis $\mathcal{B} = (x^3, x^2, x, 1)$ is

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

Since A is lower triangular, $f_T(x) = f_A(x) = (x-4)(x-3)(x-2)(x-1)$; and so the eigenvalues are $\lambda_1 = 4$, $\lambda_2 = 3$, $\lambda_3 = 2$, and $\lambda_4 = 1$. Solving for a basis for each eigenspace of A yields

$$E_{\lambda_1} = \left(\begin{bmatrix} 6 \\ 18 \\ 27 \\ 17 \end{bmatrix} \right), \quad E_{\lambda_2} = \left(\begin{bmatrix} 0 \\ 2 \\ 4 \\ 3 \end{bmatrix} \right), \quad E_{\lambda_3} = \left(\begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix} \right), \quad E_{\lambda_4} = \left(\begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \right).$$

By taking the polynomial corresponding to the basis vectors, we let $\mathcal{D} = (\vec{v}_1, \vec{v}_2, \vec{v}_3, \vec{v}_4)$ where $\vec{v}_1 = 6x^3 + 18x^2 + 27x + 17$, $\vec{v}_2 = 2x^2 + 4x + 3$, $\vec{v}_3 = x + 1$, and $\vec{v}_4 = 1$. The diagonal matrix

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

is the matrix representation of T in $\mathcal{D}$-coordinates and has the eigenvalues of T on its main diagonal. The transition matrix P from $\mathcal{B}$-coordinates to $\mathcal{D}$-coordinates is

$$P = \begin{bmatrix} 6 & 0 & 0 & 0 \\ 18 & 2 & 0 & 0 \\ 27 & 4 & 1 & 0 \\ 17 & 3 & 1 & 1 \end{bmatrix}.$$

and satisfies the required relation $D = P^{-1}AP$ as can be verified.

Example 6.25. If A is similar to B, show that A^n is similar to B^n, for any positive integer n.

Example 6.26. Suppose that $C^{-1}AC = D$. Show that for any integer n, $A^n = CD^nC^{-1}$.

Example 6.27. Let a and b be real numbers. By diagonalizing

$$M = \begin{bmatrix} a & b - a \\ 0 & b \end{bmatrix},$$

prove that

$$M^n = \begin{bmatrix} a^n & b^n - a^n \\ 0 & b^n \end{bmatrix}$$

for all positive integers n. We need a basis of $\mathbb{R}^2$ consisting of eigenvectors of M. One such basis is $\vec{v}_1 = \vec{e}_1$ and $\vec{v}_2 = \vec{e}_1 + \vec{e}_2$ where a and b are eigenvalues for corresponding to these eigenvectors, respectively. Let $P = [\vec{v}_1 \quad \vec{v}_2]$, then by 6.1, the diagonalization is

$$D = [\vec{v}_1 \quad \vec{v}_2]^{-1} M [\vec{v}_1 \quad \vec{v}_2] = \begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix}.$$

Therefore

$$M^n = (PDP^{-1})^n = \underbrace{(PDP^{-1}) \cdots (PDP^{-1})}_{n\text{-times}} = PD^nP^{-1} = \begin{bmatrix} a^n & b^n - a^n \\ 0 & b^n \end{bmatrix}.$$

Chapter 7

Canonical Forms

7.1 Invariant Subspaces

In this section we let V and W denote real or complex vector spaces.

Suppose $T \in \mathcal{L}(V)$ and U a subspace of V, then we say U is an **invariant subspace** of T if $u \in U$ implies $Tu \in U$.

Lemma 7.1. *Suppose $T \in \mathcal{L}(V)$ and U a subspace of V. Then all of the following hold*

- *U is invariant under T if and only if $T|_U$ is an operator on U,*
- *$\ker T$ is invariant under T, and*
- *$\mathrm{im} T$ is invariant under T.*

Proof. The proof of each part follows.

- By definition, invariant subspace U is invariant under T if and only if $u \in U \implies Tu \in U$, which, by definition of operator, is the same as $T|_U$ being an operator.
- If $u \ker T$ then $Tu = 0$, and hence $Tu \in \ker T$.
- If $u \mathrm{im} T$, then by the definition of range $Tu \in \mathrm{im} T$.

$\square$

Theorem 7.1. *Suppose $T \in \mathcal{L}(V)$. Let $\lambda_1, \dots, \lambda_m$ denote distinct eigenvalues of T. The the following are equivalent:*

- *T has a diagonal matrix with respect to some basis of V;*
- *V has a basis consisting of eigenvectors of T;*

- *there exist one-dimensional T-invariant subspaces $U_1, \ldots, U_n$ of V such that $V = U_1 \oplus \cdots \oplus U_n$;*
- *$V = (T - \lambda_1 I) \oplus \cdots \oplus (T - \lambda_m I)$;*
- *$\dim V = \dim(T - \lambda_1 I) + \cdots + \dim(T - \lambda_m I)$.*

Proof.   (i) $\Longleftrightarrow$ (ii): Exercise.

(ii) $\Longleftrightarrow$ (iii): Suppose (ii) holds; thus suppose V has a basis consisting of eigenvectors of T. For each j, let $U_j = \mathrm{span}(v_j)$. Obviously each U_j is a one-dimensional subspace of V that is invariant under T (because each v_j is an eigenvector of T). Because $(v_1, \ldots, v_n)$ is a basis of V, each vector in V can be written uniquely as a linear combination of $(v_1, \ldots, v_n)$. In other words, each vector in V can be written uniquely as a sum $u_1 + \cdots + u_n$, where each $u_j \in U_j$. Thus $V = U_1 \oplus \cdots \oplus U_n$. Hence (ii) implies (iii). Conversely, suppose now that (iii) holds; thus there are one-dimensional subspaces $U_1, \ldots, U_n$ of V, each invariant under T, such that $V = U_1 \oplus \cdots \oplus U_n$. For each j, let v_j be a nonzero vector in U_j. Then each v_j is an eigenvector of T. Because each vector in V can be written uniquely as a sum $u_1 + \cdots + u_n$, where each $u_j \in U_j$ (so each u_j is a scalar multiple of v_j), we see that $(v_1, \ldots, v_n)$ is a basis of V. Thus (iii) implies (ii).

(iii) $\Longrightarrow$ (iv): Suppose (ii) holds; thus thus suppose V has a basis consisting of eigenvectors of T. Thus every vector in V is a linear combination of eigenvectors of T. Hence $V = (T - \lambda_1)I + \cdots + (T - \lambda_m)I$. To show that the sum above is direct, suppose that $0 = u_1 + \cdots + u_m$, where each $u_j \in (T - \lambda_j I)$. Because nonzero eigenvectors correspond to distinct eigenvalues are linearly independent, this implies that each u_j equals 0. This implies that the sum is a direct sum, completing the proof that

(iv) implies (iii).

(v) $\Longrightarrow$ (v): Exercise.

(vi) $\Longrightarrow$ (ii): Suppose (v) holds; thus $\dim V = \dim(T - \lambda_1 I) + \cdots + \dim(T - \lambda_m I)$. Choose a basis of each $(T - \lambda_j I)$; put all these bases together to form a list $(v_1, \ldots, v_n)$ of eigenvectors of T, where $n = \dim V$. To show that this list is linearly independent, suppose $a_1 v_1 + \cdots + a_n v_n = 0$, where $a_1, \ldots, a_n$ are scalars. For each $j = 1, \ldots, m$, let u_j denote the sum of all the terms $a_k v_k$ such that $v_k \in (T - \lambda_j I)$. Thus each u_j is an eigenvector of T with eigenvalue λ_j, and $u_1 + \cdots + u_m = 0$. Because nonzero eigenvectors corresponding to distinct eigenvalues are linearly independent, this implies that each u_j equals to 0. Because each u_j is a sum of terms $a_k v_k$ where the v_k's where chosen to be a basis of $(T - \lambda_j I)$, this implies that all the a_k's equal to 0. Thus $(v_1, \ldots, v_n)$ is linearly independent and hence a basis of V. Thus (v) implies (ii).

$\square$

A vector v is called a **generalized eigenvector** of T corresponding to λ, where λ is an eigenvalue of T, if $(T - \lambda I)^j v = 0$ for some positive integer j.

Lemma 7.2. *If $T \in \mathcal{L}(V)$. Then, if m is a nonnegative integer such that $ker T^m = ker T^{m+1}$, then*

- $\{0\} = ker T^0 \subseteq ker T^1 \subseteq \cdots \subseteq ker T^m = ker T^{m+1} = ker T^{m+2} = \cdots$
- $ker T^{dimV} = ker T^{dimV+1} = ker T^{dimV+2} \cdots$
- $V = im T^0 \supseteq im T^1 \supseteq \cdots \supseteq im T^k \supseteq im T^{k+1} \supseteq \cdots$
- $im T^{dimV} = im T^{dimV+1} = im T^{dimV+2} \cdots$

Lemma 7.3. *Suppose $T \in \mathcal{L}(V)$ and λ is an eigenvalue of T. Then the set of generalized eigenvalues of T corresponding to λ equals $ker(T - \lambda I)^{dimV}$.*

Proof. The proof is left for the reader.

$\square$

An operator is called **nilpotent** if some power of it equal 0.

Lemma 7.4. *Suppose $N \in \mathcal{L}(V)$ is nilpotent, then $N^{dimV} = 0$.*

Proof. The proof is left for the reader.

$\square$

Theorem 7.2. *Let $T \in \mathcal{L}(V)$ and $\lambda \in \mathbb{F}$. Then for every basis of V with respect to which T has an upper-triangular matrix, λ appears on the diagonal of the matrix of T precisely $dim(T - \lambda I)^{dimV}$ times.*

Proof. The proof is left for the reader.

$\square$

The **multiplicity** of an eigenvalue λ of T is defined to be the dimension of the subspace of generalized eigenvectors corresponding to λ, that is the multiplicity of λ is equal to $\dim ker(T - \lambda I)^{\dim V}$.

Theorem 7.3. *If V is a complex vector space and $T \in \mathcal{L}(V)$, then the sum of the multiplicities of all the eigenvalues of T equals $\dim V$.*

Proof. The proof is left for the reader.

$\square$

Let d_j denote the multiplicity of λ_j as an eigenvalue of T, the polynomial

$$(z - \lambda_1)^{d_1} \cdots (z - \lambda_m)^{d_m}$$

is called the **characteristic polynomial** of T.

Theorem 7.4. *Suppose that V is a complex vector space and $T \in \mathcal{L}(V)$. Let q denote the characteristic polynomial of T. Then $q(T) = 0$.*

Proof. The proof is left for the reader.

$\square$

Theorem 7.5. *If $T \in \mathcal{L}(V)$ and $p \in \mathcal{P}(\mathbb{F})$, then $\ker p(T)$ is invariant under T.*

Proof. The proof is left for the reader.

$\square$

Theorem 7.6. *Suppose V is a complex vector space and $T \in \mathcal{L}(V)$. Let $\lambda_1, \dots, \lambda_m$ be the distinct eigenvalues of T, and let $U_1, \dots, U_m$ be the corresponding subspaces of generalized eigenvectors. Then*

- *$V = U_1 \oplus \cdots \oplus U_m$;*
- *each U_J is invariant under T;*
- *each $(T - \lambda_j I)\big|_{U_j}$ is nilpotent.*

Proof. The proof is left for the reader.

$\square$

Theorem 7.7. *Suppose V is a complex vector space and $T \in \mathcal{L}(V)$. Then there is a basis of V consisting of generalized eigenvectors of T.*

Proof. The proof is left for the reader.

□

Theorem 7.8. *Suppose N is a nilpotent operator on V. Then there is a basis of V with respect to which the matrix of N has the form*

$$\begin{bmatrix} 0 & & * \\ & \ddots & \\ 0 & & 0 \end{bmatrix};$$

here all entries on and below the diagonal are 0's.

Proof. The proof is left for the reader.

□

Theorem 7.9. *Suppose V is a complex vector space and $T \in \mathcal{L}(V)$. Let $\lambda_1, \dots, \lambda_m$ be the distinct eigenvalues of T. Then there is a basis of V with respect to which T has block diagonal matrix of the form*

$$\begin{bmatrix} A_1 & & 0 \\ & \ddots & \\ 0 & & A_m \end{bmatrix},$$

where each A_j is an upper-triangular matrix of the form

$$\begin{bmatrix} \lambda_1 & & * \\ & \ddots & \\ 0 & & \lambda_j \end{bmatrix}.$$

Proof. The proof is left for the reader.

□

Theorem 7.10. *Suppose $N \in \mathcal{L}(V)$ is nilpotent. Then $I + N$ has a square root.*

Proof. The proof is left for the reader.

□

On real vector spaces there exist invertible operators that have no square roots. For example, the operator of multplication by -1 on $\mathbb{R}$ has no square root because no real number has its square equal to -1.

Theorem 7.11. *Suppose V is a complex vector space. If $T \in \mathcal{L}(V)$ is invertible, then T has a square root.*

Proof. The proof is left for the reader.

$\square$

The **minimal polynomial** of T is the monic polynomial $p \in \mathcal{P}(\mathbb{F}$ of smallest degree such that $p(T) = 0$.

Theorem 7.12. *Let $T \in \mathcal{L}(V)$ and let $q \in \mathcal{P}(\mathbb{F})$. Then $q(T) = 0$ if and only if the minimal polynomial of T divided q.*

Proof. The proof is left for the reader.

$\square$

Theorem 7.13. *Let $T \in \mathcal{L}(V)$. Then the roots of the minimal polynomial of T are precisely the eigenvalues of T.*

Proof. The proof is left for the reader.

$\square$

Every $T \in \mathcal{L}(V)$ where V is a complex vector space, there is a basis of V with respect to which T has a nice upper-triangular matrix. We can do even better. There is a basis of V with respect to which the matrix of T contains zeros everywhere except possibly on the diagonal and the line directly above the diagonal.

Suppose $N \in \mathcal{L}(V)$ is nilpotent. For each nonzero vector $v \in V$, let $m(v)$ denote the largest nonnegative integer such that $N^{m(v)} \neq 0$.

Theorem 7.14. *If $N \in \mathcal{L}(V)$ is nilpotent, then there exist vectors $v_1, \ldots, v_k \in V$ such that*

- *$\left(v_1, Nv_1, \ldots, N^{m(v_1)}, \ldots, v_k, Nv_k, \ldots, N^{m(v_k)}v_k\right)$ is a basis of V;*
- *$\left(N^{m(v_1)}v_1, \ldots, N^{m(v_k)}v_k\right)$ is a basis of $\ker N$.*

Proof. The proof is left for the reader.

$\square$

Example 7.1. Suppose $T \in \mathcal{L}(V)$. Prove that if $U_1, \ldots, U_m$ are subspaces of V invariant under T, then $U_1 + \cdots + U_m$ is invariant under T. Suppose $v \in U_1 + \cdots + U_m$. Then there exists $u_1, \ldots, u_m$ such that $v = u_1 + \cdots + u_m$ with $u_j \in U_j$. Then $Tv = Tu_1 + \cdots + Tu_m$. Since each U_j is invariant under T, $Tu_j \in U_j$, so $Tv \in U_1 + \cdots + U_m$.

Example 7.2. Suppose $T \in \mathcal{L}(V)$. Prove that the intersection of any collection of subspaces of V invariant under T is invariant under T. Suppose we have subspaces $\{U_j\}$ with each U_j invariant under T. Let $v \in \cap_j U_j$. Then $Tv \in U_j$ for each j, and so $\cap_j U_j$ is invariant under T.

Example 7.3. Prove or give a counterexample: if U is a subspace of V that is invariant under every operator on V, then $U = \{0\}$ or $U = V$. We will prove the contrapositive: if U is a subspace of V and $U \neq \{0\}$ and $U \neq V$, then there exists an operator T on V such that U is not invariant under T. Let $(u_1, \ldots, u_m)$ be a basis for U, which we extend to a basis $(u_1, \ldots, u_m, v_1, \ldots, v_n)$ of V. The assumption $U \neq \{0\}$ and $U \neq V$ means that $m \geq 1$ and $n \geq 1$. Define a linear map T by $Tu_1 = v_1$ and for $j > 1$, $Tu_j = 0$. Since $v_1 \notin U$, the subspace U is not invariant under the operator T.

Example 7.4. Suppose that $S, T \in \mathcal{L}(V)$ are such that $ST = TS$. Prove that null $(T - \lambda I)$ is invariant under S for every $\lambda \in F$. Suppose $v \in \ker(T - \lambda I)$. Then $Tv = \lambda v$ and using $TS = ST$, Sv satisfies $T(Sv) = S(Tv) = S(\lambda v) = \lambda(Sv)$. Thus $Sv \in \ker(T - \lambda I)$ and so $\ker(T - \lambda I)$ is invariant under S.

Example 7.5. Define $T \in \mathcal{L}(F^2)$ by $T(w, z) = (z, w)$. Find all eigenvalues and eigenvectors of T. Suppose $(w, z) \neq (0, 0)$ and $T(w, z) = (z, w) = \lambda(w, z)$. Then $z = \lambda w$ and $w = \lambda z$. Of course this leads to $w = \lambda z = \lambda^2 w$, $z = \lambda w = \lambda^2 z$. Since $w \neq 0$ or $z \neq 0$, we see that $\lambda^2 = 1$ so that $\lambda = \pm 1$. A basis of eigenvectors is $(w_1, z_1) = (1, 1)$, $(w_2, z_2) = (-1, 1)$ and they have eigenvalues 1 and -1 respectively.

Example 7.6. Define $T \in \mathcal{L}(F^3)$ by $T(z_1, z_2, z_3) = (2z_2, 0, 5z_3)$. Find all eigenvalues and eigenvectors of T. Suppose $(z_1, z_2, z_3) \neq (0, 0, 0)$ and

$$T(z_1, z_2, z_3) = (2z_2, 0, 5z_3) = \lambda(z_1, z_2, z_3).$$

If $\lambda = 0$ then $z_2 = z_3 = 0$, and one checks that $v_1 = (0, 0, 0)$ is an eigenvector with eigenvalue 0. If $\lambda \neq 0$ then $z_2 = 0$, $2z_2 = \lambda z_1 = 0$, $5z_3 = \lambda z_3$, so $z_1 = 0$ and $\lambda = 5$. The eigenvector for $\lambda = 5$ is $v_2 = (0, 0, 1)$. These are the only eigenvalues and each eigenspace is one dimensional.

Example 7.7. Suppose n is a positive integer and $T \in \mathcal{L}(\mathbb{F}^n)$ is defined by

$$T(x_1, \ldots, x_n) = (x_1 + \cdots + x_n, \ldots, x_1 + \cdots + x_n).$$

Find all eigenvalues and eigenvectors of T. First, any vector of the form $v_1 = (\alpha, \ldots, \alpha)$, for $\alpha \in \mathbb{F}$, is an eigenvector with eigenvalue n. If v_2 is any vector $v\_2 = (x\_1, \ldots, x\_n),$ such that $x_1 + \cdots + x_n = 0$ then v_2 is an eigenvector with eigenvalue 0. Here are the independent eigenvectors: $v_1 = (1, 1, \ldots, 1)$, and $v_n = (1, 0, \ldots, 0) - E_n$, for $n \geq 2$ where E_n denoted the n-th standard basis vector.

Example 7.8. Suppose $T \in \mathcal{L}(V)$ is invertible and $0 \neq \lambda \in F$. Prove that λ is an eigenvalue of T if and only if $\frac{1}{\lambda}$ is an eigenvalue of T^{-1}. Suppose $v \neq 0$ and $Tv = \lambda v$. Then $v = T^{-1}Tv = \lambda T^{-1}v$, or $T^{-1}v = \frac{1}{\lambda}v$, and the other direction is similar.

Example 7.9. Suppose $S, T \in \mathcal{L}(V)$. Prove that ST and TS have the same eigenvalues. Suppose $v \neq 0$ and $STv = \lambda v$. Multiply by T to get $TS(Tv) = \lambda Tv$. Thus if $t \neq 0$ then λ is also an eigenvalue of TS, with nonzero=o eigenvector Tv. On the other hand, if $Tv = 0$, then $\lambda = 0$ is an eigenvalue of ST. But if T is not invertible, then $\mathrm{im}TS \subset \mathrm{im}T$ is not equal to V, so TS has a nontrivial null space, hence 0 is an eigenvalue of TS.

Example 7.10. Suppose $T \in \mathcal{L}(V)$ is such that every vector in V is an eigenvector of T. Prove that T is a scalar multiple of the identity operator. Pick a basis $(v_1, \ldots, v_N)$ for V. By assumption, $Tv_n = \lambda_n v_n$. Pick any two distance indices, m, n. We also have $T(v_m + v_n) = \lambda(v_m + v_n) = \lambda(v_m + v_n) = \lambda_m v_m + \lambda_n v_n$. Write this as $0 = (\lambda - \lambda_m)v_m + (\lambda - \lambda_n)v_n$. Since v_m and v_n are independent, $\lambda = \lambda_m = \lambda_n$, and all the λ_n are equal.

Example 7.11. Suppose $S, T \in \mathcal{L}(V)$ and S is invertible. Prove that if $p \in \mathcal{P}(\mathbb{F})$ is a polynomial, then $p(STS^{-1}) = Sp(T)S^{-1}$. First let's show that for positive integers n, $(STS^{-1})^n = ST^nS^{-1}$. We may do this by induction, with nothing to show if $n = 1$. Assume it's true for $n = k$, and consider

$$(STS^{-1})^{k+1} = (STS^{-1})^k(STS^{-1}) = ST^kS^{-1}STS^{-1} = ST^{k+1}S^{-1}.$$

Now suppose $P(z) = a_n z^N + \cdots + a_1 z + a_0$. Then

$$p(STS^{-1}) = \sum_{n=0}^{N} a_n(STS^{-1})^n = \sum_{n=0}^{N} a_n ST^n S^{-1}$$

$$= S \left(\sum_{n=0}^{N} a_n T^n \right) S^{-1} = Sp(T)S^{-1}.$$

Example 7.12. Suppose $F = \mathbb{C}$, $T \in \mathcal{L}(V)$, $p \in \mathcal{P}(\mathbb{C})$, and $a \in \mathbb{C}$. Prove that a is an eigenvalue of $p(T)$ if and only if $a = p(\lambda)$ for some eigenvalue $\$\$$ of T. Suppose first that $v \neq 0$ is an eigenvalue of T with eigenvalue λ; that is $Tv = \lambda v$. Then for positive integers n, $T^n v = T^{n-1}\lambda v = \cdots \lambda^n v$, and so $p(T)v = p(\lambda)v$. That is $\alpha = p(\lambda)$ is an eigenvalue of $p(T)$ if λ is an eigenvalue of T. Conversely, suppose now that α is a eigenvalue of $p(T)$, so there is a $v \neq 0$ with $p(T)v = \alpha v$, or $(p(T) - \alpha I)v = 0$. Since $\mathbb{F} = \mathbb{C}$, we may factor the polynomial $\$p(T)- I \$$ into linear factors

$$0 = (p(T) - \alpha I)v = \prod(T - \lambda_n I)v.$$

At least one of the factors is not invertible, so at least one of the λ_n, say λ_1, is an eigenvalue of T. Let $w \neq 0$ be an eigenvector for T with eigenvalue λ_1. Then

$$0 = (T - \lambda_N I) \cdots (T - \lambda_1 I)w = (p(T) - \alpha I)w,$$

so w is an eigenvalue for $p(T)$ with eigenvalue α. But by the first part of the argument, $p(T)w = p(\lambda_1)w = \alpha w$ and $\alpha = p(\lambda_1)$.

Example 7.13. Show that the previous exercise does not hold with $F = \mathbb{R}$. Take $T : \mathbb{R}^2 \to \mathbb{R}^2$ given by $T(x, y) = (-y, x)$. We've seen perviously that T has no real eigenvalues. On the other hand, $T^2(x, y) = (-x, -y) = -1(x, y)$.

Example 7.14. Suppose V is a complex vector space and $T \in \mathcal{L}(V)$. Prove that T has an invariant subspace of dimension j for each $j = 1, \ldots, \dim V$. Let $(v_1, \ldots, v_N)$ be a basis with respect to which T has an upper triangular matrix. Then by a previous proposition, $T : \text{span}(v_1, \ldots, v_j) \to \text{span}(v_1, \ldots, v_j)$.

Example 7.15. Give an example of an operator whose matrix with respect to some basis contains only 0's on the diagonal, but the operator is invertible. Consider $T = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$.

Example 7.16. Give an example of an operator whose matrix with respect to some basis contains only nonzero numbers on the diagonal, but the operator is not invertible. Taking $T = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$. If $v = [1, -1]$, then $Tv = 0$.

Example 7.17. Give an example of an operator on $\mathbb{C}^4$ whose characteristic and minimal polynomials both equal $z(z-1)^2(z-3)$.

Example 7.18. Give an example of an operator on $\mathbb{C}^4$ whose characteristic polynomial equals $z(z-1)^2(z-3)$ and whose minimal polynomial equals $z(z-1)(z-3)$.

Example 7.19. Suppose $a_0, \ldots, a_{n-1} \in \mathbb{C}$. Find the minimal and characteristic polynomials of the operator on $\mathbb{C}^n$ whose matrix is

$$
\begin{bmatrix}
0 & & & & & -a_0 \\
1 & 0 & & & & -a_1 \\
& 1 & \ddots & & & -a_2 \\
& & \ddots & & & \vdots \\
& & 1 & & 0 & -a_{n-2} \\
& & & & 1 & -a_{n-1}
\end{bmatrix}
$$

with respect to the standard bases.

7.2   Jordan Canonical Form

A basis of V is called a **Jordan basis** for T if with respect to this basis T has block diagonal matrix

$$
\begin{bmatrix}
A_1 & & 0 \\
& \ddots & \\
0 & & A_m
\end{bmatrix}
$$

where each A_j is an upper triangular matrix of the form

$$
A_j = \begin{bmatrix} \lambda_j & 1 & & 0 \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ 0 & & & \lambda_j \end{bmatrix}
$$

where the diagonal is filled with some eigenvalue λ_j of T.

Because there exist operators on real vector spaces that have no eigenvalues, there exist operators on real vector spaces for which there is no corresponding Jordan basis.

Theorem 7.15. *Suppose V is a complex vector space. If $T \in \mathcal{L}(V)$, then there is a basis of V that is a Jordan basis for T.*

Proof. The proof is left for the reader.

$\square$

A basis of V is called a **Jordan basis** for T if with respect to this basis T has block diagonal matrix

$$
\begin{bmatrix} A_1 & & 0 \\ & \ddots & \\ 0 & & A_m \end{bmatrix}
$$

where each A_j is an upper triangular matrix of the form

$$
A_j = \begin{bmatrix} \lambda_j & 1 & & 0 \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ 0 & & & \lambda_j \end{bmatrix}
$$

where the diagonal is filled with some eigenvalue λ_j of T.

An operator T can be put into Jordan canonical form if its characteristic and minimal polynomials factor into linear polynomials. this is always true if the vector space is complex.

Theorem 7.16. *Let $T \in \mathcal{L}(V)$ whose characteristic and minimal polynomials are, respectively,*

$$
c(t) = (t-\lambda_1)^{n_1} \cdots (t-\lambda_r)^{n_r} \qquad and \qquad m(t) = (t-\lambda_1)^{m_1} \cdots (t-\lambda_r)^{m_r}
$$

where the λ_i are distinct scalars. Then T has block diagonal matrix representation J whose diagonal entries are of the form

$$
J_{ij} = \begin{bmatrix}
\lambda_i & 1 & 0 & \cdots & 0 & 0 \\
0 & \lambda_i & 1 & \cdots & 0 & 0 \\
\vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\
0 & 0 & 0 & \cdots & \lambda_i & 1 \\
0 & 0 & 0 & \cdots & 0 & \lambda_i
\end{bmatrix}
$$

For each λ_i the corresponding blocks have the following properties:

- There is at least one J_{ij} of order m_i; all other J_{ij} are of order $\leq m_i$
- The sum of the orders of the J_{ij} is n_i.
- The number of J_{ij} equals the geometric multiplicity of λ_i.
- The number of J_{ij} of each possible order is uniquely determined by T.

Proof. The proof is left for the reader.

$\square$

The matrix J in the above proposition is called the **Jordan canonical form** of the operator T. A diagonal block J_{ij} is called a **Jordan block** belongng to the eigenvalue λ_i. Observe that

$$
\begin{bmatrix}
\lambda_i & 1 & 0 & \cdots & 0 & 0 \\
0 & \lambda_i & 1 & \cdots & 0 & 0 \\
\vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\
0 & 0 & 0 & \cdots & \lambda_i & 1 \\
0 & 0 & 0 & \cdots & 0 & \lambda_i
\end{bmatrix}
=
\begin{bmatrix}
\lambda_i & 0 & 0 & \cdots & 0 & 0 \\
0 & \lambda_i & 0 & \cdots & 0 & 0 \\
\vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\
0 & 0 & 0 & \cdots & \lambda_i & 0 \\
0 & 0 & 0 & \cdots & 0 & \lambda_i
\end{bmatrix}
+
\begin{bmatrix}
0 & 1 & 0 & \cdots & 0 & 0 \\
0 & 0 & 1 & \cdots & 0 & 0 \\
\vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\
0 & 0 & 0 & \cdots & 0 & 1 \\
0 & 0 & 0 & \cdots & 0 & 0
\end{bmatrix}
$$

That is, $J_{ij} = \lambda_i I + N$ where N is the nilpotent block.

Example 7.20. Suppose the characteristic and minimum polynomials of an operator T are, respectively,

$$c(t) = (t-2)^4(t-3)^3 \qquad \text{and} \qquad m(t) = (t-2)^2(t-3)^2.$$

Then the Jordan canonical form of T is one of the following matrices:

$$
\begin{bmatrix}
2 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 2 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 2 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 2 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 3 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 3 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 3
\end{bmatrix}
\quad \text{or} \quad
\begin{bmatrix}
2 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 2 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 2 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 2 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 3 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 3 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 3
\end{bmatrix}
$$

The first matrix occurs if T has two independent eigenvectors belonging to the eigenvalue 2; and the second matrix occurs if T has three independent eigenvectors belonging to 2.

Example 7.21. Suppose $N \in \mathcal{L}(V)$ is nilpotent. Prove that the minimal polynomial of N is z^{m+1}, where m is the length of the longest consecutive string of $1's$ that appears on the line directly above the diagonal in the matrix of N with respect to any Jordan basis for N.

Example 7.22. Suppose V is a complex vector space and $T \in \mathcal{L}(V)$. Prove that there does not exist a direct sum decomposition of V into two proper subspaces invariant under T if and only if the minimal polynomial of T is of the form $(z - \lambda)^{\dim V}$ for some $\lambda \in \mathbb{C}$.

Example 7.23. Suppose $T \in \mathcal{L}(V)$ and $(v_1, \dots, v_n)$ is a basis of V that is a Jordan basis for T. Describe the matrix of T with respect to the basis $(v_n, \dots, v_1)$ obtained by reversing the order of the v's.

Example 7.24. Consider a 2-by-2 matrix of real numbers

$$\begin{bmatrix} a & c \\ b & d \end{bmatrix}.$$

Example 7.25. Suppose A is a block diagonal matrix

$$A = \begin{bmatrix} A_1 & & 0 \\ & \ddots & \\ 0 & & A_m \end{bmatrix},$$

where each A_j is a square matrix. Prove that the set of eigenvalues of A equals the union of the eigenvalues of $A_1, \dots, A_m$.

Example 7.26. Suppose A is a block upper-triangular matrix

$$A = \begin{bmatrix} A_1 & & * \\ & \ddots & \\ 0 & & A_m \end{bmatrix},$$

where each A_j is a square matrix. Prove that the set of eigenvalues of A equals the union of the eigenvalues of $A_1, \dots, A_m$.

Example 7.27. Suppose V is a real vector space and $T \in \mathcal{L}(V)$. Suppose $\alpha, \beta \in \mathbb{R}$ are such that $T^2 + \alpha T + \beta I = 0$. Prove that T has an eigenvalue if and only if $\alpha^2 \geq 4\beta$.

Example 7.28. Suppose V is a real inner-product space and $T \in \mathcal{L}(V)$. Prove that there is an orthonormal basis of V with respect to which T has a block upper-triangular matrix

$$\begin{bmatrix} A_1 & & * \\ & \ddots & \\ 0 & & A_m \end{bmatrix}.$$

where each A_j is a 1-by-1 matrix or a 2-by-2 matrix with no eigenvalues.

Example 7.29. Prove that if $T \in \mathcal{L}(V)$ and j is a positive integer such that $j \leq \dim V$, then T has an invariant subspace whose dimension equals $j - 1$ or j.

Example 7.30. Prove that there does not exist an operator $T \in \mathcal{L}(\mathbb{R}^7)$ such that $T^2 + T + I$ is nilpotent.

Example 7.31. Give an example of an operator $T \in \mathcal{L}(\mathbb{C}^7)$ such that $T^2 + T + I$ is nilpotent.

Example 7.32. Suppose V is a real vector space and $T \in \mathcal{L}(V)$. Suppose $\alpha, \beta \in \mathbb{R}$ are such that $\alpha^2 < 4\beta$. Prove that $\text{null}\, (T^2 + \alpha T + \beta I)^k$ has even dimension for every positive integer k.

Example 7.33. Suppose V is a real vector space and $T \in \mathcal{L}(V)$. Suppose $\alpha, \beta \in \mathbb{R}$ are such that $\alpha^2 < 4\beta$ and $T^2 + \alpha T + \beta I$ is nilpotent. Prove that $\dim V$ is even and $(T^2 + \alpha T + \beta I)^{\dim V/2} = 0$.

Example 7.34. Prove that if $T \in \mathcal{L}(\mathbb{R}^3)$ and 5, 7 are eigenvalues of T, then T has no eigenpairs.

Example 7.35. Suppose V is a real vector space with $\dim V = n$ and $T \in \mathcal{L}(V)$ is such that null T^{n-2} null T^{n-1}. Prove that if T has at most two distinct eigenvalues and that T has no eigenpairs.

Example 7.36. Prove that 1 is an eigenvalue of every square matrix with the property that the sum of the entries in each row equals 1.

Example 7.37. Suppose V is a real vector space with $\dim V = 2$. Prove that if

$$\begin{bmatrix} a & c \\ b & d \end{bmatrix}$$

is the matrix of T with respect to some basis of V, then the characteristic polynomial of T equals $(z - a)(z - d) - bc$.

Example 7.38. Suppose V is a real inner-product space and $S \in \mathcal{L}(V)$ is an isometry. Prove that if (α, β) is an eigenpair of S, then $\beta = 1$.

References

Index