

Vietnam National University, Hanoi

Hanoi University of Science



ENGLISH FOR CIS - K68A3 CIS

Final Report: Linear Algebra

Report conducted by:

Full Name	Student ID
Nguyen Tran Phuc Chau	23001835
Bui Quang Chien	23001837
Nguyen Le Ngoc Bao	23001832

Hanoi, December, 2024

Introduction

The purpose of this report is to summarize the work of Group 5, which consists of three members who are the authors of this report. Due to the limited time available and the practical focus of the project, this report is not intended to serve as an educational resource or as an academically rigorous document. Rather, it provides sufficient knowledge and explanations, though there may still be some gaps. We welcome feedback from readers to help improve the quality of this report.

The authors thank the readers for taking the time to read this report.

Sincerely

Signed
Chau
Nguyen Tran Phuc Chau

Contents

1	System of Linear Equation: Algebra	4
1.1	System of Linear Equation	4
1.1.1	Line, Plane, Space, Etc	4
1.1.2	Pictures of Solution Sets	8
1.1.3	Parametric Description of Solution Sets	8
1.2	Row Reduction	9
1.2.1	The elimination method - Gaussian elimination	9
1.2.2	Echelon Forms	10
1.2.3	The Row Reduction Algorithm	11
1.3	Parametric Form	13
1.3.1	Free Variables	13
1.3.2	Number of Solutions	13
2	System of Linear Equation: Geometry	15
2.1	Vectors	15
2.2	Vector Addition and Scalar Multiplication	15
2.2.1	Vector Addition	15
2.2.2	Scalar Multiplication	16
2.2.3	Vector Subtraction	16
2.3	Vectors Equations and Spans	16
2.3.1	Linear Combinations of Vectors	16
2.3.2	Vector Equations	18
2.3.3	Spans	18
2.3.4	Geometric Interpretation of Spans	19
2.3.5	Pictures of Spans	20
2.4	Matrix Equations:	23
2.4.1	The Matrix Equation: $Ax = b$	23
2.4.2	How to write a Linear System?	24
2.4.3	Spans and Consistency	24
2.5	Linear Independence	25
2.5.1	Pictures of Linear Independence:	26
2.5.2	Linear Dependence and Free Variables	28
2.6	Subspaces	29
2.6.1	Common Types of Subspaces	29
2.6.2	Column Space and Null Space	31
2.6.3	Recipe: Compute a Spanning Set for a Null Space	31
2.7	Basis and Dimension	32
2.7.1	Dimension	32
2.7.2	Computing a Basis for a Subspace	33
2.7.3	A basis for the Null space	35
2.7.4	Finding a Basis for a General Subspace	35
2.8	The Rank Theorem	35
2.8.1	Definition of Rank and Nullity	36
2.8.2	The Rank Theorem:	36
2.8.3	Implications of the Rank Theorem:	36
3	Linear Transformations and Matrix Algebra	38

3.1	Matrix Transformation	38
3.1.1	Matrices as Function	38
3.1.2	Transformations	40
3.1.3	Matrix Transformations	42
3.2	One-to-one and Onto Transformations	45
3.2.1	One-to-one Transformations	45
3.2.2	Onto Transformation	47
3.2.3	Comparison between onto & one-to-one	48
3.3	Linear Transformation	48
3.3.1	The Standard Coordinate Vectors	49
3.3.2	The Matrix of a Linear Transformation	51
3.4	Matrix Multiplication	53
3.4.1	Composition of linear transformations	53
3.4.2	Matrix multiplication	55
3.4.3	Composition and Matrix Multiplication	58
3.4.4	The algebra of transformations and matrices	59
3.5	Matrix Inverses	61
3.5.1	Invertible Matrices	62
3.5.2	Solving Linear Systems using Inverses	64
3.5.3	Invertible Linear Transformations	64
3.5.4	The Invertible Matrix Theorem	65
4	Determinant	67
4.1	Determinants: Definition	67
4.2	Cofactor Expansions	71
4.2.1	Cofactor Expansions	71
4.2.2	Cramer's Rule and Matrix Inverses	72
4.2.3	Summary: Methods for Computing Determinants	74
4.3	Determinants and Volumes	75
4.3.1	Parallelograms and Parallelepipeds	75
4.3.2	Determinants and Volumes	76
4.3.3	Volumes of Regions	77
5	Eigenvalue and Eigenvector	82
5.1	The definitions of Eigenvalue and Eigenvector	82
5.1.1	Eigenvalues and Eigenvector	82
5.1.2	Eigenspaces	83
5.1.3	The Inverible Matrix Theorem: Addenda	86
5.2	The Characteristic Polynomial	86
5.3	Similarity	88
5.3.1	Similar Matrices	88
5.3.2	Eigenvalues of Similar Matrices	89
5.4	Diagonalization	90
5.4.1	Algebraic and Geometric Multiplicity	95
5.4.2	Similarity and multiplicity	98

1 System of Linear Equation: Algebra

The objective of this content is to introduce the definitions of a linear equation system, to solve a linear equation system and to present the solutions using the **parametric form**.

1.1 System of Linear Equation

To begin, it is essential to develop a clear understanding of what a **system of linear equations** entails and the principles underlying its structure. Furthermore, we must dive into the various methods and techniques used to solve such systems effectively, ranging from traditional substitution and elimination methods to more advanced approaches like matrix operations and **Gaussian elimination** (This will be covered later on in section 1.2.1). This foundational knowledge serves as the foundation for the tackling of more complex problems and applications in mathematics and beyond. An **equation** in the unknown x, y, z, \dots is called **linear** if both sides of the equation are the sum of (constant) multiples of x, y, z, \dots plus an optional constant.

Example 1: $2x + y = 12, 3x + 6y = 2z, \dots$

A **System of Linear Equation** is a collection of some several linear equations.

Example 2:

$$\begin{cases} x + y = 3, \\ 2x - y = 4. \end{cases} \quad (1)$$

The **Solution:** of the system of linear equation is the values of x, y, z, \dots that simultaneously satisfy all the equations in the system. The **solution set** is the collection of all solutions.

Solving the system of linear equation means finding the solution set of it.

We can easily prove that the solution set of a system of linear equation can have 1, infinite or 0 solution. The system which has no solution is called **inconsistent**, and be called **consistent** otherwise.

1.1.1 Line, Plane, Space, Etc

We use \mathbf{R} to denote the set of all real number, i.e, the number line.

Definition

Let n be a positive whole number. We define:

$$\mathbb{R}^n = \text{all ordered } n\text{-tuples of real numbers } (x_1, x_2, x_3, \dots, x_n)$$

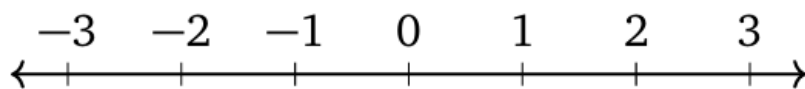


Figure 1: The shape of \mathbb{R}^1 .

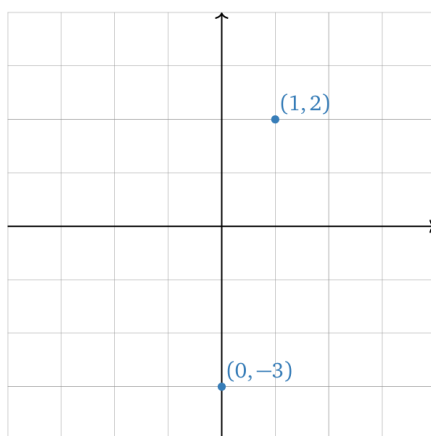


Figure 2: The shape of \mathbb{R}^2 .

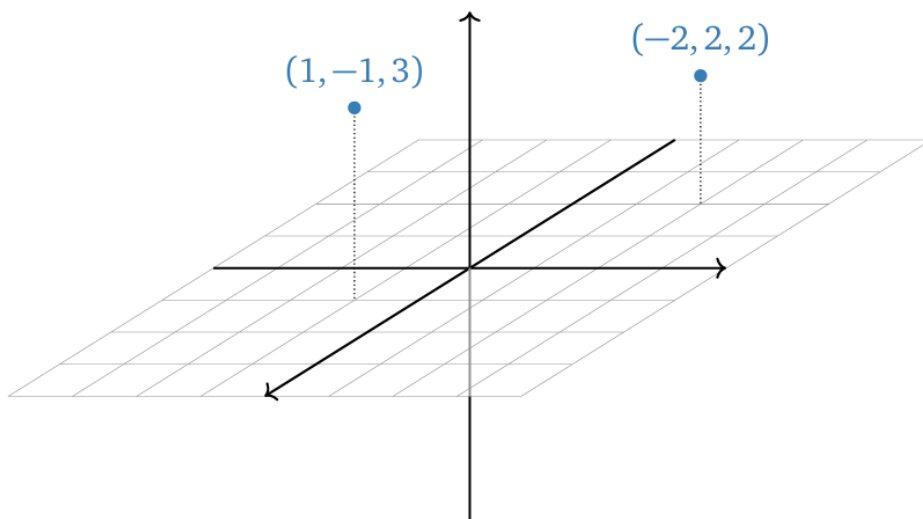


Figure 3: The shape of \mathbb{R}^3 .

An n -tuples of real numbers is a **point** of \mathbb{R}^n . We have been introduced to \mathbb{R}^1 , \mathbb{R}^2 , and \mathbb{R}^3 , where \mathbb{R}^1 takes the form of a straight line with points on it, \mathbb{R}^2 represents a coordinate plane, and \mathbb{R}^3 corresponds to the $Oxyz$ coordinate system. With \mathbb{R}^4 , \mathbb{R}^5 and \mathbb{R}^n , just harder to visualize, because the world we live in is a three-dimensional world, or \mathbb{R}^3 . So we need to go back to the definition: \mathbb{R}^n is the set of all ordered n -tuples of real numbers.

They're still "geometric" space, just like \mathbb{R}^2 and \mathbb{R}^3 extend to \mathbb{R}^n . Thus, we can represent \mathbb{R}^4 and \mathbb{R}^5 using sequences of 4 or 5 numbers, but it's hard to visualize them.

To make things clear, we have some examples:

Example 3: Imagine a complex traffic intersection with four main roads labeled x , y , z , and w . These road are simply used by a variety of vehicles, including cars, trucks, buses, bikes, etc. Each road can be though as an axis in a four-dimensional space, analogous to the \mathbb{R}^4 -space in linear algebra.

If there are 100, 200, 300, 400 cars per hour passing through roads x, y, z, w respectively, then this can be recorded by the point $(100, 200, 300, 400)$ in \mathbb{R}^4 .

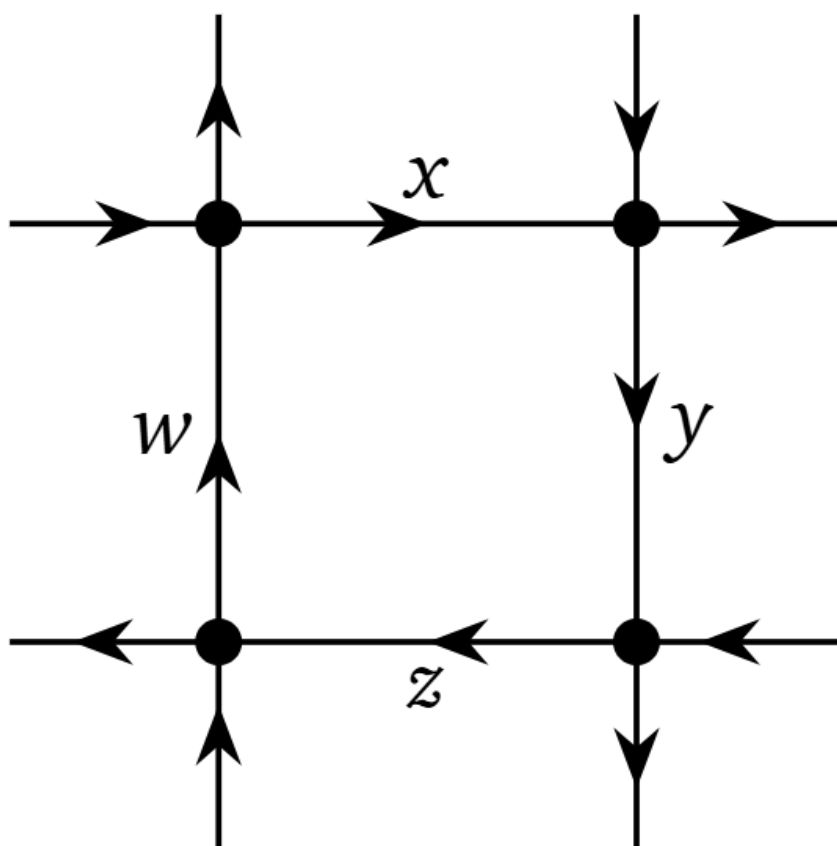


Figure 4: Traffic Flow - The shape of \mathbb{R}^4 .

This is a really interesting point from a psychological perspective. Instead of having to juggle four separate numbers, we can now focus on a single, unified piece of data. This simplification

can significantly reduce cognitive load, making it easier to understand, remember, and process the information. It's like taking a complex puzzle and reducing it to a single, clear picture.

Example 4: A QR code is basically a fancy barcode. It's made up of a grid of black and white squares that computers can easily read. Picture a 29x29 grid. If you read each line left-to-right, then top-to-bottom (like a book), you can think of it as a long string of 841 digits, either 0 or 1. Each digit represents a black or white square.



Figure 5: QR Code - a 29 x 29 array of white/black squares.

So, a whole QR code can be seen as a single point in a super high-dimensional space, \mathbb{R}^{841} . This might sound super complex, but it's a really cool way to think about it. From a psychological standpoint, it's much easier to think of a QR code as one big thing rather than a bunch of tiny squares. It simplifies things in our minds.

From now on, the complexity and confusion is no more. We can see that it is useful and easy to save a point in \mathbb{R}^n instead of saving about 800 numbers in QR code, or a list of 4 numbers (in traffic flow).

1.1.2 Pictures of Solution Sets

Before discussing about how to solve a system of linear equation, it's helpful to understand what these solution sets look like geometrically.

Consider the linear equation:

$$x + y = 1 \quad (2)$$

We can change it to

$$y = 1 - x \quad (3)$$

That is a line in space, which have the slope is 1, and x-intercept is -1.

Line's Definition

The **Line** is a ray that is straight and infinite in both directions.

Another linear equation is:

$$x + y + z = 1 \quad (4)$$

We know that it is the implicit equation of a plane in space.

Plane's Definition

The **Plane** is a sheet that is infinite in all directions.

Remake: The equation: $x + y + z + w = 1$ is the equation of a space in \mathbb{R}^4 . And more geometrically, the equation in \mathbb{R}^n defines an $(n - 1)$ -spaces in n -space.

1.1.3 Parametric Description of Solution Sets

We had the definition of solution sets. In section 1.1.2, we have proved that an equation in \mathbb{R}^n defines an $(n - 1)$ -spaces. So, we can defines that equation with n-1 variable.

Consider the equation(3): by taking a new variable name t, we can define:

$$(x, y) = (t, 1 - t) \text{ for any } x \in R.$$

Or with equation(4): We can define:

$$(x, y, z) = (1 - t - u, t, u) \text{ for any } t, u \in R.$$

We call t or u is **parameter**, as it *Parameterizes* the point we need to define.

Note that parameter allow us to use \mathbb{R}^n to *label* the point in n -planes, but that n -planes is not like \mathbb{R}^n , because every point on that n -planes have $n+1$ coordinates.

1.2 Row Reduction

Now, we will discuss about how to solve a system of linear equation.

1.2.1 The elimination method - Gaussian elimination

Solve the following system of equations:

$$\begin{aligned} x + y + z &= 6 \\ 2x + 3y + 3z &= 14 \\ 3x + 2y + 4z &= 13 \end{aligned}$$

That's how we solve that:

Transformation 1: $R_2 \leftarrow R_2 - 2R_1$

$$\begin{aligned} x + y + z &= 6 \\ (2x + 3y + 3z) - 2(x + y + z) &= 14 - 2(6) \\ 3x + 2y + 4z - 2x - 2y - 2z &= 13 - 12 \\ x + y + z &= 6 \\ \Rightarrow 0x + y + z &= 2 \\ 3x + 2y + 4z &= 13 \end{aligned}$$

Transformation 2: $R_3 \leftarrow R_3 - 3R_1$

$$\begin{aligned} x + y + z &= 6 \\ (3x + 2y + 4z) - 3(x + y + z) &= 13 - 3(6) \\ 3x + 2y + 4z - 3x - 3y - 3z &= 13 - 18 \\ x + y + z &= 6 \\ \Rightarrow 0x + y + z &= 2 \\ 0x - y + z &= -5 \end{aligned}$$

Transformation 3: $R_3 \leftarrow R_3 + R_2$

$$\begin{aligned} x + y + z &= 6 \\ y + z &= 2 \\ (-y + z) + (y + z) &= -5 + 2 \\ x + y + z &= 6 \\ \Rightarrow y + z &= 2 \\ 2z &= -3 \end{aligned}$$

Transformation 4: $R_3 \leftarrow \frac{R_3}{2}$

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

(*) From the equation $z = \frac{-3}{2}$, substitute into $y + z = 2$:

$$y + \frac{-3}{2} = 2 \Rightarrow y = \frac{7}{2}$$

Substitute the values of y and z into the equation $x + y + z = 6$:

$$x + \frac{7}{2} + \frac{-3}{2} = 6 \Rightarrow x + 2 = 6 \Rightarrow x = 4$$

The solution of the system is:

$$x = 4, \quad y = \frac{7}{2}, \quad z = \frac{-3}{2}$$

Instead of solve a system of linear of equation, we can use a **matrix** and solve it in the same way.

$$\begin{aligned}\left(\begin{array}{ccc|c}1 & 1 & 1 & 6 \\2 & 3 & 3 & 14 \\3 & 2 & 4 & 13\end{array}\right) &\xrightarrow{R_2 - 2R_1 \rightarrow R_2} \left(\begin{array}{ccc|c}1 & 1 & 1 & 6 \\0 & 1 & 1 & 2 \\0 & -1 & 1 & -3\end{array}\right) \\&\longrightarrow \left(\begin{array}{ccc|c}1 & 1 & 1 & 6 \\0 & 1 & 1 & 2 \\0 & 0 & 2 & -3\end{array}\right) \\&\longrightarrow \left(\begin{array}{ccc|c}1 & 1 & 1 & 6 \\0 & 1 & 1 & 2 \\0 & 0 & 1 & -\frac{3}{2}\end{array}\right) \\&\longrightarrow \left(\begin{array}{ccc|c}1 & 0 & 0 & 4 \\0 & 1 & 0 & \frac{7}{2} \\0 & 0 & 1 & -\frac{3}{2}\end{array}\right)\end{aligned}$$

Attention!

When we wrote our row operations above, we used expressions like $R_2 - 2 \times R_1 \rightarrow R_2$. It is important to note that this does not imply the second row is mathematically equal to itself minus twice the first row. Rather, it indicates that the second row is being updated or replaced by the result of this operation. This style of notation is commonly used in computer programming, where a similar syntax is employed to signify the reassignment of a variable's value. Understanding this distinction is key to correctly interpreting the operation.

1.2.2 Echelon Forms

Imagine a neatly organized staircase of numbers, where each step moves further to the right as you descend. At the top, the structure is solid, with leading numbers boldly defining their position. Below these leaders, an empty void stretches downward—zeros filling the space, like a structured march towards simplicity. This visual hierarchy captures the essence of Echelon Form, a mathematical framework that brings clarity to systems of equations, organizing chaos into order and paving the way for solutions to emerge.

Definition

A matrix is in **row echelon form** if it satisfies the following conditions:

1. All nonzero rows are above any rows of all zeros.
2. The leading (left most) entry of each nonzero row (called the pivot) is strictly to the right of the leading entry in the row above it.
3. All entries below a pivot are zeros.

$$\begin{pmatrix} \boxed{\star} & \star & \star & \star & \star \\ 0 & \boxed{\star} & \star & \star & \star \\ 0 & 0 & 0 & \boxed{\star} & \star \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \quad \begin{array}{l} \star = \text{any number} \\ \boxed{\star} = \text{any nonzero number} \end{array}$$

Figure 6: Echelon Form

A matrix is in **reduced row echelon form** if it is in row echelon form and in addition:

5. Each pivot is equal to 1.
6. Each pivot is the only nonzero entry in its column.

$$\begin{pmatrix} \mathbf{1} & 0 & \star & 0 & \star \\ 0 & \mathbf{1} & \star & 0 & \star \\ 0 & 0 & 0 & \mathbf{1} & \star \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \quad \begin{array}{l} \star = \text{any number} \\ \mathbf{1} = \text{pivot} \end{array}$$

Figure 7: Reduced row echelon form

If a matrix is in reduced row echelon form, we can say that it is *solved*. We will prove it in next part.

1.2.3 The Row Reduction Algorithm

At first, we have a theorem: *Every matrix is row equivalent to one and only one matrix in reduced echelon form*

We will give an algorithm, called **row reduction** or **Gaussian elimination**.

With this, we can easily prove that: how you row reduce a matrix does not affect the matrix in reduced echelon form you get.

Gaussian algorithm: **Step 1:** Swap the topmost unprocessed row with a lower one so a leftmost nonzero entry is in the 1st row (if necessary).

Step 2: Scale the topmost unprocessed row so that its first nonzero entry is equal to 1.

Step 3: Use row replacements so all entries below this 1 are 0.

Repeat these 3 steps for each subsequent row until the last row.

Example : Row reduce this matrix:

$$\left(\begin{array}{ccc|c} 0 & 7 & 4 & 2 \\ 2 & 4 & 6 & 12 \\ 3 & 1 & 1 & 2 \end{array} \right)$$

Solution:

$$\xrightarrow{R_1 \leftrightarrow R_2} \left(\begin{array}{ccc|c} 2 & 4 & 6 & 12 \\ 0 & -7 & -4 & 2 \\ 3 & 1 & -1 & -2 \end{array} \right)$$

$$\xrightarrow{\frac{1}{2} \times R_1 \rightarrow R_1} \left(\begin{array}{ccc|c} 1 & 2 & 3 & 6 \\ 0 & -7 & -4 & 2 \\ 3 & 1 & -1 & -2 \end{array} \right)$$

$$\xrightarrow{R_3 + (-3) \times R_1 \rightarrow R_3} \left(\begin{array}{ccc|c} 1 & 2 & 3 & 6 \\ 0 & -7 & -4 & 2 \\ 0 & -5 & -10 & -20 \end{array} \right)$$

$$\xrightarrow{R_2 \leftrightarrow R_3} \left(\begin{array}{ccc|c} 1 & 2 & 3 & 6 \\ 0 & -5 & -10 & -20 \\ 0 & -7 & -4 & 2 \end{array} \right)$$

$$\xrightarrow{R_2 \times \frac{-1}{5} \rightarrow R_2} \left(\begin{array}{ccc|c} 1 & 2 & 3 & 6 \\ 0 & 1 & 2 & 4 \\ 0 & -7 & -4 & 2 \end{array} \right)$$

$$\xrightarrow{R_3 + 7 \times R_2 \rightarrow R_3} \left(\begin{array}{ccc|c} 1 & 2 & 3 & 6 \\ 0 & 1 & 2 & 4 \\ 0 & 0 & 10 & 30 \end{array} \right)$$

$$\xrightarrow{\frac{1}{10} \times R_3 \rightarrow R_3} \left(\begin{array}{ccc|c} 1 & 2 & 3 & 6 \\ 0 & 1 & 2 & 4 \\ 0 & 0 & 1 & 3 \end{array} \right)$$

$$\xrightarrow{R_2 + (-2) \times R_3 \rightarrow R_2} \left(\begin{array}{ccc|c} 1 & 2 & 3 & 6 \\ 0 & 1 & 0 & -2 \\ 0 & 0 & 1 & 3 \end{array} \right)$$

$$\xrightarrow{R_1 + (-2) \times R_2 + (-3) \times R_3 \rightarrow R_1} \left(\begin{array}{ccc|c} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & -2 \\ 0 & 0 & 1 & 3 \end{array} \right)$$

So we can have the reduced row form is:

$$\left(\begin{array}{ccc|c} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & -2 \\ 0 & 0 & 1 & 3 \end{array} \right) \rightarrow \begin{cases} x & = & 1 \\ y & = & -2 \\ z & = & 3 \end{cases}$$

Or the matrix have only 1 solution is

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

Definition

A pivot position of a matrix is an entry that is a pivot of a row echelon form of that matrix.

When working with a matrix that has only one solution, it is easy to use reduced row echelon form (RREF) to solve the matrix. However, with a matrix that has infinite solutions or is inconsistent, RREF may not be as effective. We will discuss this in the next part.

1.3 Parametric Form

1.3.1 Free Variables

We will see in an example:

$$\begin{aligned} \left(\begin{array}{ccc|c} 2 & 1 & 12 & 1 \\ 1 & 2 & 9 & -1 \end{array} \right) &\xrightarrow{R_1 \leftrightarrow R_2} \left(\begin{array}{ccc|c} 1 & 2 & 9 & -1 \\ 2 & 1 & 12 & 1 \end{array} \right) \xrightarrow{R_2 = R_2 - 2 \times R_1} \left(\begin{array}{ccc|c} 1 & 2 & 9 & -1 \\ 0 & -3 & -6 & 3 \end{array} \right) \\ &\xrightarrow{R_2 = R_2 \times \frac{-1}{3}} \left(\begin{array}{ccc|c} 1 & 2 & 9 & -1 \\ 0 & 1 & 2 & -1 \end{array} \right) \xrightarrow{R_1 = R_1 - 2 \times R_2} \left(\begin{array}{ccc|c} 1 & 0 & 5 & 1 \\ 0 & 1 & 2 & -1 \end{array} \right) \rightarrow \begin{cases} x + 5z & = & 1 \\ y + 2z & = & -1 \end{cases} \end{aligned}$$

In what sense is that matrix solved? We rewrite as:

$$\begin{cases} x & = & 1 - 5z \\ y & = & -1 - 2z \end{cases} \Rightarrow \begin{cases} x & = & 1 - 5z \\ y & = & -1 - 2z \\ z & = & z \end{cases} \quad z \text{ any real number}$$

This is called the *parametric form* for the solution to the linear system, and variable z is called a free variable. Also

$$(x, y, z) = (1 - 5z, -1 - 2z, z)$$

This called a **parameterized equation** for the same line.

1.3.2 Number of Solutions

There are *three possibilities* for the reduced row echelon form of the augmented matrix of a linear system.

1. **The last column is a pivot column.** In this case, the system is inconsistent. There are zero solutions, i.e., the solution set is empty. For example, the matrix:

$$\left(\begin{array}{ccc|c} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{array} \right)$$

comes from a linear system with no solutions.

2. **Every column except the last column is a pivot column.** In this case, the system has a unique solution. For example, the matrix:

$$\left(\begin{array}{ccc|c} 1 & 0 & 0 & a \\ 0 & 1 & 0 & b \\ 0 & 0 & 1 & c \end{array} \right)$$

tells us that the unique solution is $(x, y, z) = (a, b, c)$.

3. **The last column is not a pivot column, and some other column is not a pivot column either.** In this case, the system has infinitely many solutions, corresponding to the infinitely many possible values of the free variable(s). For example, in the system corresponding to the matrix:

$$\left(\begin{array}{cccc|c} 1 & \textcolor{red}{2} & 0 & \textcolor{blue}{3} & 1 \\ 0 & \textcolor{red}{0} & 1 & \textcolor{blue}{4} & 1 \end{array} \right)$$

any values for $\textcolor{red}{x}_2$ and $\textcolor{blue}{x}_4$ yield a solution to the system of equations.

2 System of Linear Equation: Geometry

First and foremost, it's important to grasp the concept of a system of linear equations and the various methods available to solve it effectively.

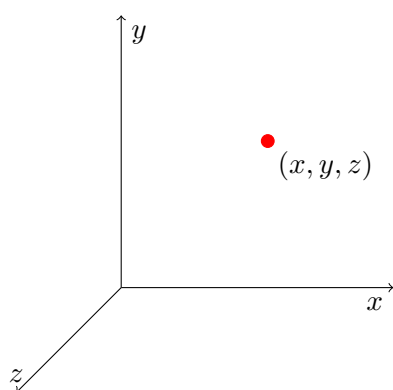
2.1 Vectors

Definition

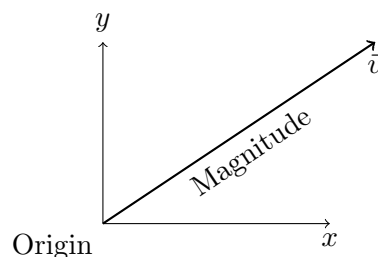
A vector is a list of numbers. There are (at least) two ways to understand what this list of numbers means:

- One way is to think of the vector as being a point in a space. Then this list of numbers is a way of identifying that point in space, where each number represents the vector's component in that dimension.
- Another way to think of a vector is as a magnitude and a direction, e.g., a quantity like velocity ("the fighter jet's velocity is 250 mph north-by-northwest"). In this way, a vector is a directed arrow pointing from the origin to the endpoint given by the list of numbers.

Vector as a Point



Vector as Magnitude and Direction



2.2 Vector Addition and Scalar Multiplication

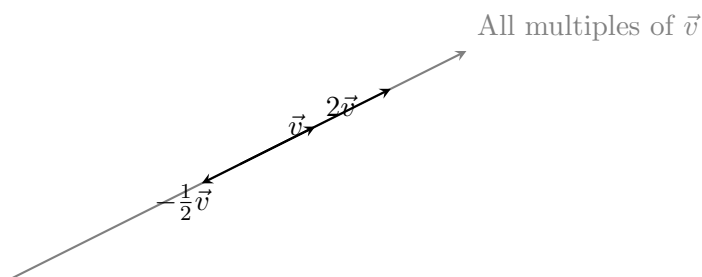
In this section, we will learn how to add vectors and how to multiply vectors by numbers, both algebraically and geometrically.

2.2.1 Vector Addition

To add two vectors together algebraically, simply add their corresponding components. For example:

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

Geometrically, vector addition follows the parallelogram law: place the tail of the second vector at the head of the first vector. The resulting vector is the diagonal of the parallelogram formed by these two vectors. For example:



2.2.2 Scalar Multiplication

To multiply a vector by a scalar, multiply each component of the vector by the scalar. For example:

$$c \cdot \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} c \cdot x \\ c \cdot y \\ c \cdot z \end{pmatrix}.$$

We call c a scalar to distinguish it from a vector. If \vec{v} is a vector and c is a scalar, then $c\vec{v}$ is called a *scalar multiple* of \vec{v} .

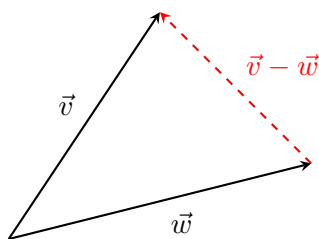
Example

$$\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} + \begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix} = \begin{pmatrix} 5 \\ 7 \\ 9 \end{pmatrix}, \quad \text{and} \quad -2 \cdot \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = \begin{pmatrix} -2 \\ -4 \\ -6 \end{pmatrix}.$$

2.2.3 Vector Subtraction

Geometrically, vector subtraction involves placing the tails of both vectors at the same point. The vector $\vec{v} - \vec{w}$ points from the head of \vec{w} to the head of \vec{v} . For example:

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



2.3 Vectors Equations and Spans

2.3.1 Linear Combinations of Vectors

A very fundamental operation in vector algebra is to construct *linear combinations* of vectors.

Definition

A linear combination of vectors $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p$ is any expression of the form:

$$\mathbf{y} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \dots + c_p\mathbf{v}_p,$$

where c_1, c_2, \dots, c_p are scalars (real numbers) called *weights*.

Key Properties:

- The weights c_i can be any real number, including zero or negative values.
- If all weights are zero, the result is the *zero vector*:

$$\mathbf{y} = \mathbf{0} = 0\mathbf{v}_1 + 0\mathbf{v}_2 + \cdots + 0\mathbf{v}_p.$$

- Linear combinations allow us to represent any point in a subspace spanned by the vectors $\mathbf{v}_1, \dots, \mathbf{v}_p$.

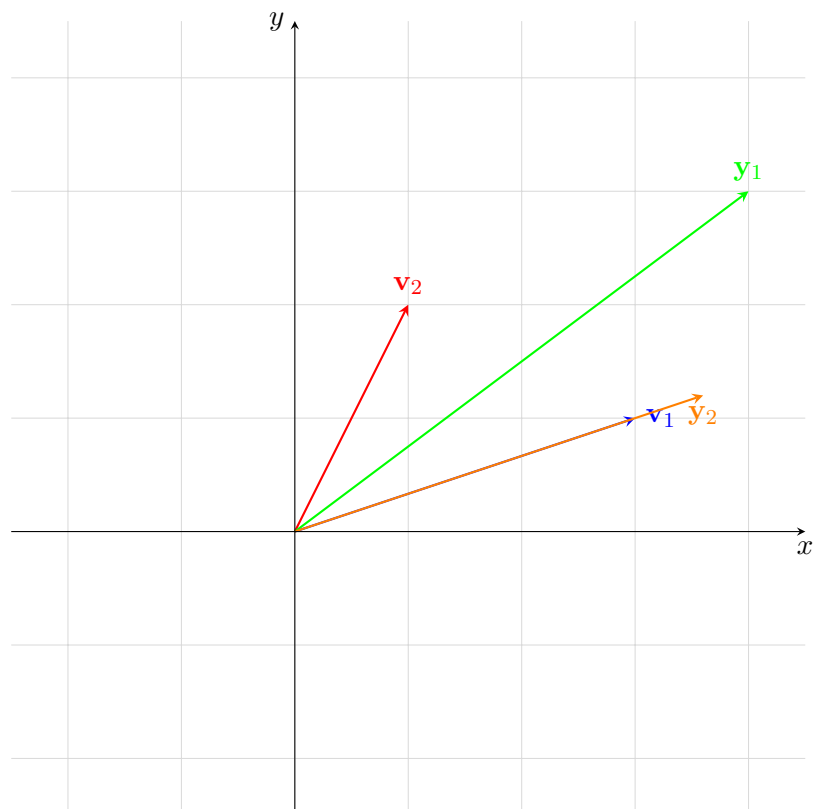
Examples of Linear Combinations:

$$\mathbf{y}_1 = \sqrt{3}\mathbf{v}_1 + \mathbf{v}_2,$$

$$\mathbf{y}_2 = 12\mathbf{v}_1 + 0\mathbf{v}_2 \quad (\text{equivalent to } 12\mathbf{v}_1),$$

$$\mathbf{y}_3 = \mathbf{0} \quad (\text{equivalent to the zero vector } 0\mathbf{v}_1 + 0\mathbf{v}_2).$$

Geometric Interpretation: Linear combinations geometrically represent all possible ways to "scale and add" vectors to form new vectors. Below is an illustration of linear combinations of two vectors \mathbf{v}_1 and \mathbf{v}_2 :



2.3.2 Vector Equations

An equation involving vectors with n coordinates is the same as n equations involving only numbers. For example, the equation:

$$x \begin{pmatrix} 2 \\ 6 \\ 1 \end{pmatrix} + y \begin{pmatrix} -1 \\ -2 \\ -1 \end{pmatrix} = \begin{pmatrix} 8 \\ 16 \\ 3 \end{pmatrix}$$

simplifies to:

$$\begin{pmatrix} 2x - y \\ 6x - 2y \\ x - y \end{pmatrix} = \begin{pmatrix} 8 \\ 16 \\ 3 \end{pmatrix}.$$

For two vectors to be equal, all of their coordinates must be equal, so this is just the system of linear equations:

$$\begin{aligned} 2x - y &= 8, \\ 6x - 2y &= 16, \\ x - y &= 3. \end{aligned}$$

Definition

A *vector equation* is an equation involving a linear combination of vectors with possibly unknown coefficients.

This is a powerful concept, as it relates a single equation involving the columns of vectors to a set of equations corresponding to their rows. Specifically, a vector equation:

$$x_1 \mathbf{v}_1 + x_2 \mathbf{v}_2 + \cdots + x_n \mathbf{v}_n = \mathbf{b}$$

is equivalent to the system of linear equations derived from the rows of the vectors $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ and \mathbf{b} .

2.3.3 Spans

If a vector equation is equivalent to a linear system, then it must be possible for a vector equation to be inconsistent as well. How can we understand what it means – in terms of vectors – for a vector equation to be inconsistent? The answer involves a new concept: the span of a set of vectors.

Definition

Let $S = \{v_1, v_2, \dots, v_k\}$ be a set of vectors in vector space V . The **span** of S , denoted **Span**(S), is the set of all linear combinations of the vectors in S .

$$\text{Span}(S) = \{c_1 \vec{v}_1 + c_2 \vec{v}_2 + \cdots + c_k \vec{v}_k \mid c_1, c_2, \dots, c_k \in \mathbb{R} \text{ (or } \mathbb{C})\}.$$

c_1, c_2, c_n are **real** (or **complex**) scalars. If S contains a single vector \vec{v} , then **Span**(S) represents the **line through the origin** and \vec{v} .

We also say that $\text{Span}(S)$ is the subset spanned by or generated by the vectors v_1, v_2, \dots, v_k .

Three characterizations of consistency:

1. A vector b is in the span of v_1, v_2, \dots, v_k .
2. The vector equation:

$$x_1v_1 + x_2v_2 + \dots + x_kv_k = b$$

has a solution.

3. The linear system with augmented matrix

$$\left(\begin{array}{ccc|c} | & | & & | \\ v_1 & v_2 & \dots & v_k \\ | & | & & | \end{array} \middle| \begin{array}{c} | \\ b \\ | \end{array} \right) \quad (5)$$

is consistent.

2.3.4 Geometric Interpretation of Spans

If S is a set then:

- *1 vector*: $\text{Span}(S)$ is the line passing through the origin and the vector.
- *2 linearly independent vectors*: $\text{Span}(S)$ is a plane passing through the origin.
- *3 linearly independent vectors*: $\text{Span}(S)$ is the entire \mathbb{R}^3 space;

If the vectors in S are not linearly independent, $\text{Span}(S)$ is still a subspace, but its dimension is smaller than the number of vectors in S .

1. $\text{Span}(S)$ is a **subspace** of V .
2. $\text{Span}(S)$ always includes the zero vector $\vec{0}$ and all linear combinations of the vectors in S .
3. If S generates the entire vector space V , then S is called a **spanning set** of V :

$$\text{Span}(S) = V$$

Example 1: Let $S = (1,0), (0,1)$ in \mathbb{R}^2 :

- Any linear combination is the form:

$$c_1(1,0) + c_2(0,1) = (c_1, c_2).$$

- $\text{Span}(S) = \mathbb{R}^2$, since any vector in \mathbb{R}^2 can be written as (c_1, c_2) .

Example 2: Let $S = (1,0,0), (0,1,0), (0,0,1)$ in \mathbb{R}^3 :

- Any linear combination is the form:

$$c_1(1, 0, 0) + c_2(0, 1, 0) + c_3(0, 0, 1) = (c_1, c_2, c_3).$$

- $\text{Span}(S) = \mathbb{R}^3$, since any vector in \mathbb{R}^3 can be written as (c_1, c_2, c_3) .

Equivalent means that, for any given list of vectors v_1, v_2, \dots, v_k, b , either all three statements are true, or all three statements are false. In other words, the truth of one statement guarantees the truth of the other two, and if one is false, the others must also be false.

To break it down simply, imagine you're working with these vectors and trying to determine certain properties, such as whether a solution exists to a system of equations involving them. If one statement about these vectors is true (for example, the system has a solution), this ensures the other two statements must also be true—they are inseparably connected. Conversely, if one statement is false (such as no solution exists), it immediately tells you that the other two statements are also false.

2.3.5 Pictures of Spans

Drawing a picture of $\text{Span}(S)$ is essentially the same as visualizing all possible linear combinations of the vectors v_1, v_2, \dots, v_k . In linear algebra, the span of a set of vectors refers to the collection of all vectors that can be formed by taking scalar multiples of the given vectors and then adding them together. This means that if you have a set of vectors $\{v_1, v_2, \dots, v_k\}$, the span represents every possible point that these vectors can "reach" through scaling and addition.

For example, if v_1 and v_2 are linearly independent vectors in \mathbb{R}^2 , their span will form a plane passing through the origin. If they are scalar multiples of each other, their span will instead form a line. When we draw the span, we are effectively illustrating the subspace these vectors generate. This geometric representation is not only helpful in understanding the relationships between the vectors but also gives us insight into the structure of the underlying vector space.

In practice, drawing the span helps us visualize how the given vectors interact and the "space" they cover. Whether it's a line, a plane, or a higher-dimensional space, the span provides a complete picture of the solutions and possibilities within that vector space. Here's some example of it:

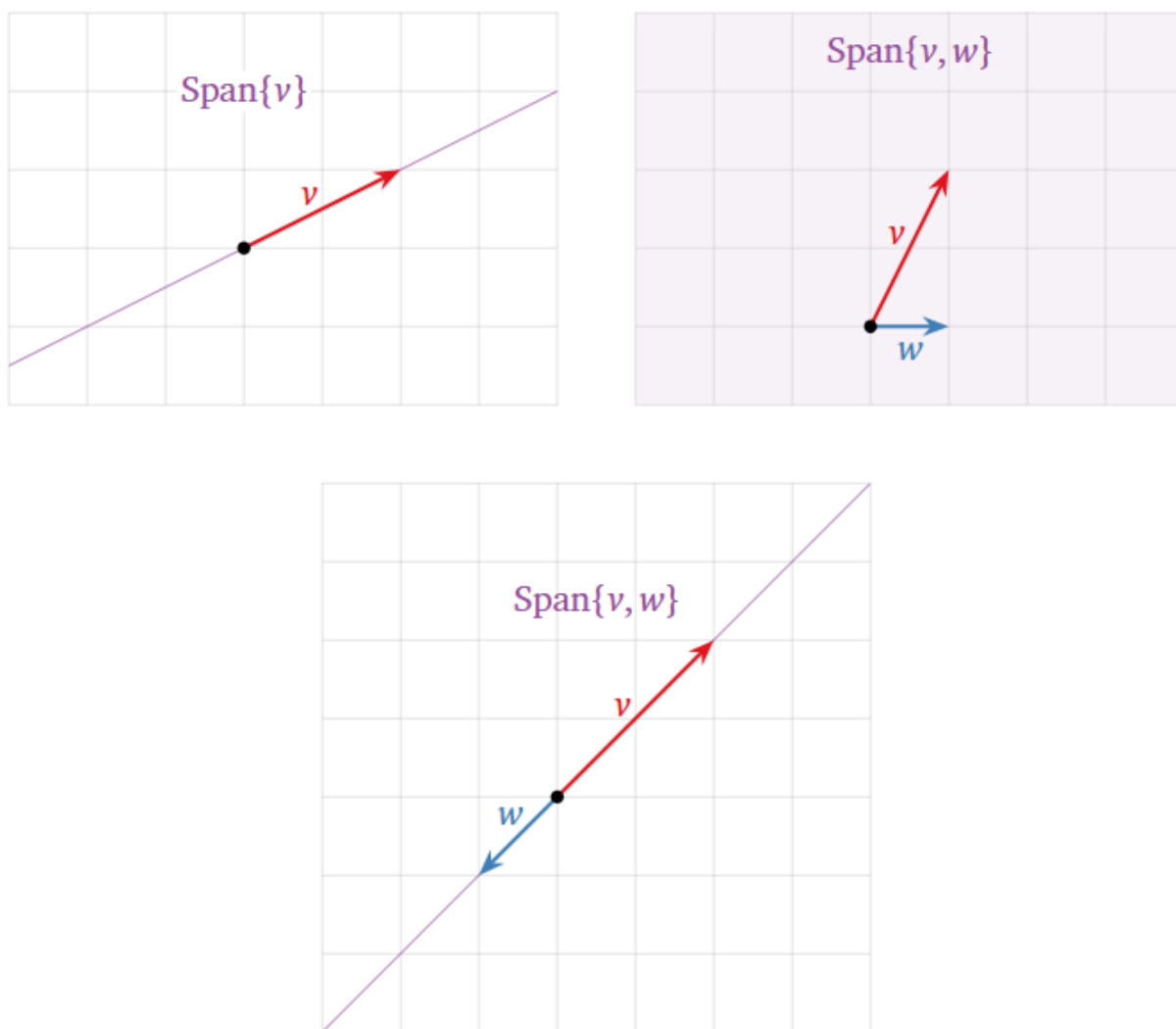


Figure 8: Spans in \mathbb{R}^2

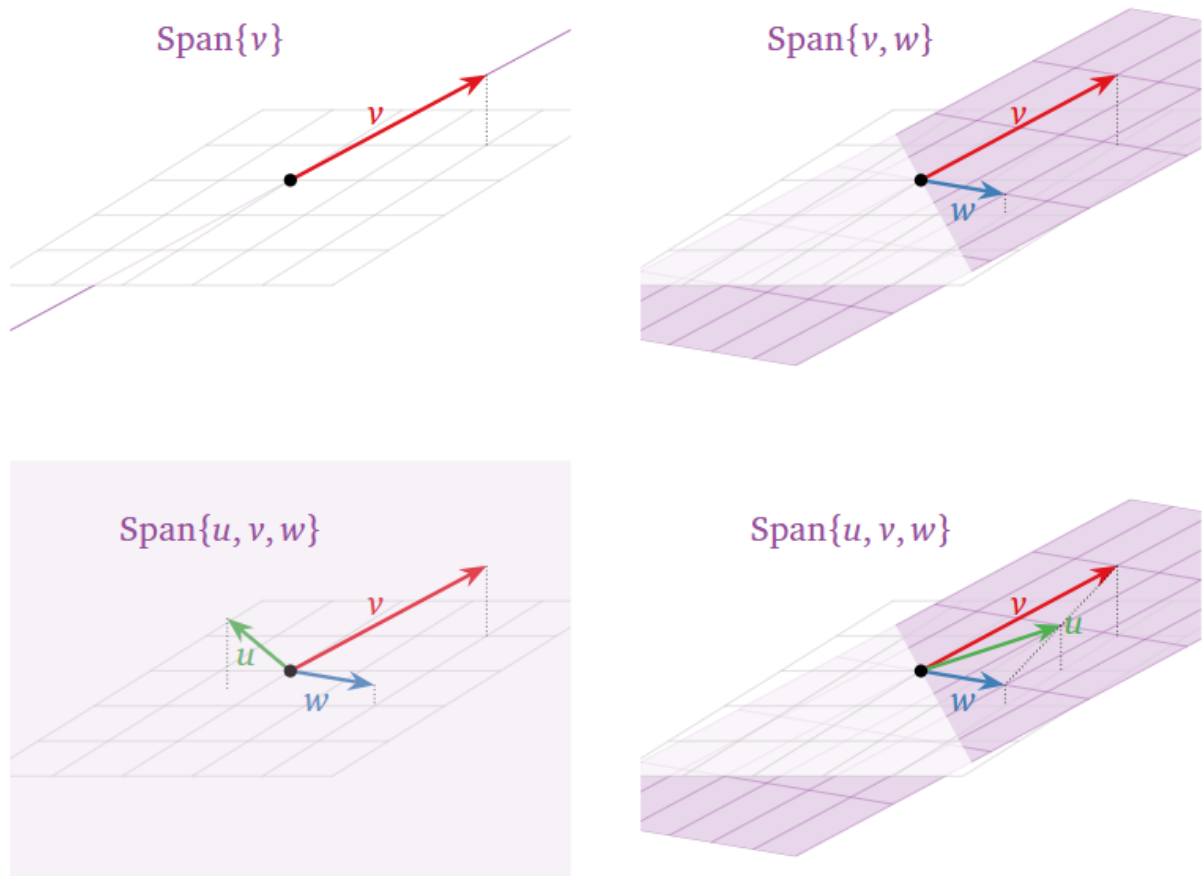


Figure 9: Spans in \mathbb{R}^3

The span of two non-collinear vectors forms a plane through the origin and their heads.
 Similarly, three coplanar (but not collinear) vectors also span a plane, while two
 collinear vectors only span a line.

2.4 Matrix Equations:

2.4.1 The Matrix Equation: $Ax = b$

In this section, we introduce a concise way to represent a system of linear equations: $Ax = b$. Here, A is a matrix, and x and b are vectors, typically of different sizes. Before diving into this concept, we first need to explain how to multiply a matrix by a vector.

When we say that A is an $m \times n$ matrix, we mean that A consists of m rows and n columns. This structure is fundamental to understanding the multiplication process and the relationships between the equations in the system.

Definition

A **matrix equation** is an equation of the form $Ax = b$, where A is an $m \times n$ matrix, b is a vector in \mathbb{R}^m , and x is a vector whose coefficients x_1, x_2, \dots, x_n are unknown. Let A be an $m \times n$ matrix with columns v_1, v_2, \dots, v_n :

$$A = \begin{pmatrix} | & | & \dots & | \\ v_1 & v_2 & \dots & v_n \\ | & | & \dots & | \end{pmatrix} \quad (6)$$

The **product** of A with a vector x in \mathbb{R}^n is a linear combination.

$$Ax = \begin{pmatrix} | & | & \dots & | \\ v_1 & v_2 & \dots & v_n \\ | & | & \dots & | \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = x_1 v_1 + x_2 v_2 + \dots + x_n v_n.$$

This result is a vector in \mathbb{R}^m .

Example:

$$\begin{pmatrix} 1 & 2 & 3 \end{pmatrix} \begin{pmatrix} 7 \\ 8 \\ 9 \end{pmatrix} = 1 \begin{pmatrix} 7 \\ 8 \\ 9 \end{pmatrix} + 2 \begin{pmatrix} 7 \\ 8 \\ 9 \end{pmatrix} + 3 \begin{pmatrix} 7 \\ 8 \\ 9 \end{pmatrix} = \begin{pmatrix} 50 \end{pmatrix}$$

In order for Ax to be well defined and make sense, the number of entries in the vector x must be equal to the number of columns in the matrix A . This is because the entries of x serve as the coefficients for the columns of A when forming a linear combination. Each entry in x determines the contribution of the corresponding column in A to the resulting vector.

The resulting vector Ax will have the same number of entries as the number of rows in A . This is because each column of A consists of entries corresponding to the rows of the matrix. As a result, the linear combination of these columns produces a vector with dimensions that match the number of rows in A . Understanding this relationship is fundamental to understanding the mechanics of matrix-vector multiplication and its role in representing systems of linear equations.

If A is an $m \times n$ matrix, then Ax makes sense when x has n entries. The product Ax has m entries.

Properties of the Matrix-Vector Product: Let A be a matrix $m \times n$, let u, v be vectors in \mathbb{R}^n , and let c be a scalar. Then:

- $A(u + v) = Au + Av$

- $A(cu) = cAu$

2.4.2 How to write a Linear System?

The simplest way to represent a linear system is by using **matrix notation**, which compactly writes the entire system as:

$$Ax = b$$

1. **Matrix A:** Contains the coefficients of the variables in the system. Each row corresponds to one equation, and each column corresponds to one variable.

2. **Vector x:** Represents the unknown variables in the system. It is a column vector where each entry corresponds to a variable.

3. **Vector b:** Represents the constants on the right-hand side of the equations. It is also a column vector.

For the system of equations:

$$2x + 3y = 54x + 6y = 10$$

We can write it in matrix form as:

$$\begin{pmatrix} 2 & 3 \\ 4 & 6 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 5 \\ 10 \end{pmatrix}$$

There are many ways to write or represent a linear system, but this is the simplest and most straightforward method to do it.

2.4.3 Spans and Consistency

The matrix equation $Ax = b$ has a solution if and only if b is in the span of the columns of A .

Definition: Let A be a matrix with columns v_1, v_2, \dots, v_n :

$$A = \begin{pmatrix} | & | & \cdots & | \\ v_1 & v_2 & \cdots & v_n \\ | & | & \cdots & | \end{pmatrix}.$$

Then

$$Ax = b \text{ has a solution} \iff \text{there exist } x_1, x_2, \dots, x_n \text{ such that } A \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = b$$

$$\iff \text{there exist } x_1, x_2, \dots, x_n \text{ such that } x_1v_1 + x_2v_2 + \cdots + x_nv_n = b$$

$$\iff b \text{ is a linear combination of } v_1, v_2, \dots, v_n \iff b \text{ is in the span of the columns of } A.$$

To make this clear, we have some examples:

Example: Let matrix $A = \begin{pmatrix} 2 & 1 \\ -1 & 0 \\ 1 & -1 \end{pmatrix}$. Does the equation $Ax = \begin{pmatrix} 1 \\ -1 \\ 2 \end{pmatrix}$ have a solution?

Solution: The columns of A are: $v_1 = \begin{pmatrix} 2 \\ -1 \\ 1 \end{pmatrix}$ and $v_2 = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}$. We simply solve this equation by using row reduction (Gaussian elimination)

$$\left(\begin{array}{cc|c} 2 & 1 & 1 \\ -1 & 0 & -1 \\ 1 & -1 & 2 \end{array} \right) \xrightarrow{\text{Gaussian elimination}} \left(\begin{array}{cc|c} 1 & 0 & 1 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{array} \right)$$

The solution for vector x is: $(x_1, x_2)^t = (1, -1)^t$.

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

When solution always exist:

Let A be an $m \times n$ matrix. The following are equivalent:

1. $Ax = b$ has a solution for all b in \mathbb{R}^m .
2. The span of the columns of A is all of \mathbb{R}^m .
3. A has a pivot position in every row.

2.5 Linear Independence

In linear algebra, the span of a set of vectors is sometimes smaller than expected, meaning one or more vectors are redundant. This is called linear independence.

A set is linearly independent if no vector can be made from the others. If one can, the set is linearly dependent.

Linear independence helps us understand vector spaces, showing if vectors are adding unique information or just repeating others.

Definition

A set of vectors $\{v_1, v_2, \dots, v_k\}$ is **linearly independent** if the equation

$$x_1v_1 + x_2v_2 + \dots + x_kv_k = 0$$

has only the trivial solution $x_1 = x_2 = \dots = x_k = 0$. Otherwise, the set is **linearly dependent**, meaning there exist scalars x_1, x_2, \dots, x_k , not all zero, such that

$$x_1v_1 + x_2v_2 + \dots + x_kv_k = 0.$$

Checking Linear Independence: The set $\{v_1, v_2, \dots, v_k\}$ is linearly independent if the matrix equation $Ax = 0$, where A has columns $\{v_1, v_2, \dots, v_k\}$, has only the trivial solution.

Facts About Linear Independence

1. **Two vectors are linearly dependent if and only if they are collinear.** This means that one vector is a scalar multiple of the other. In other words, if you can write one vector as $c \cdot v_1 = v_2$ for some scalar c , then the two vectors are not independent.
2. **Any set that includes the zero vector is always linearly dependent.** The zero vector contributes no meaningful direction to the set because any scalar multiplied by the zero vector still results in the zero vector. Therefore, the presence of a zero vector guarantees that the set cannot be independent.
3. **If a subset of $\{v_1, v_2, \dots, v_k\}$ is linearly dependent, then the entire set $\{v_1, v_2, \dots, v_k\}$ is also linearly dependent.** This is because the dependence within the subset propagates to the larger set. Adding additional vectors cannot eliminate the dependency that already exists in the subset.

These facts provide a quick way to determine whether a set of vectors is independent or dependent without solving the entire system

Proof of these facts: If $v_1 = cv_2$, then $v_1 - cv_2 = 0$, so $\{v_1, v_2\}$ is linearly dependent. In the other direction, if

$$x_1v_1 + x_2v_2 = 0$$

with $x_1 \neq 0$ (say), then

$$v_1 = -\frac{x_2}{x_1}v_2.$$

It is easy to produce a linear dependence relation if one vector is the zero vector: for instance, if $v_1 = 0$, then

$$1 \cdot v_1 + 0 \cdot v_2 + \dots + 0 \cdot v_k = 0.$$

After reordering, we may suppose that $\{v_1, v_2, \dots, v_r\}$ is linearly dependent, with $r < p$. This means there is an equation of linear dependence:

$$x_1v_1 + x_2v_2 + \dots + x_rv_r = 0,$$

where at least one of x_1, x_2, \dots, x_r is nonzero. This is also an equation of linear dependence among $\{v_1, v_2, \dots, v_k\}$, since we can take the coefficients of v_{r+1}, \dots, v_k to all be zero.

→ ***What needs to be proven.***

2.5.1 Pictures of Linear Independence:

A set containing a single vector, denoted as $\{\mathbf{v}\}$, is linearly independent if and only if the vector \mathbf{v} is not the zero vector.

In this case, the set $\{\mathbf{v}\}$ contains just one vector. The only possible linear combination is

$$x\mathbf{v} = \mathbf{0}$$

where x is a scalar. For this equation to hold true, we must have either $x = 0$ or $\mathbf{v} = \mathbf{0}$. If \mathbf{v} is the zero vector, the set is not linearly independent because x could take any value. However, if $\mathbf{v} \neq \mathbf{0}$, the only way $x\mathbf{v} = \mathbf{0}$ is for $x = 0$.

Therefore, the set $\{\mathbf{v}\}$ is linearly independent if and only if $\mathbf{v} \neq \mathbf{0}$.

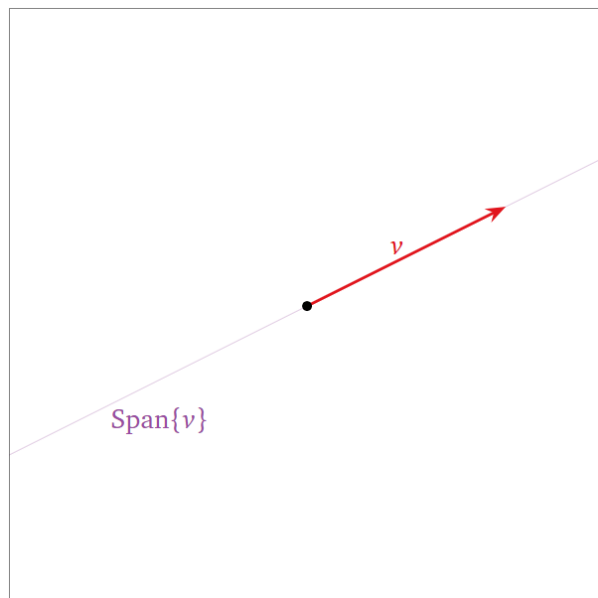


Figure 10: A set containing one vector v is linearly independent when $v \neq 0$.

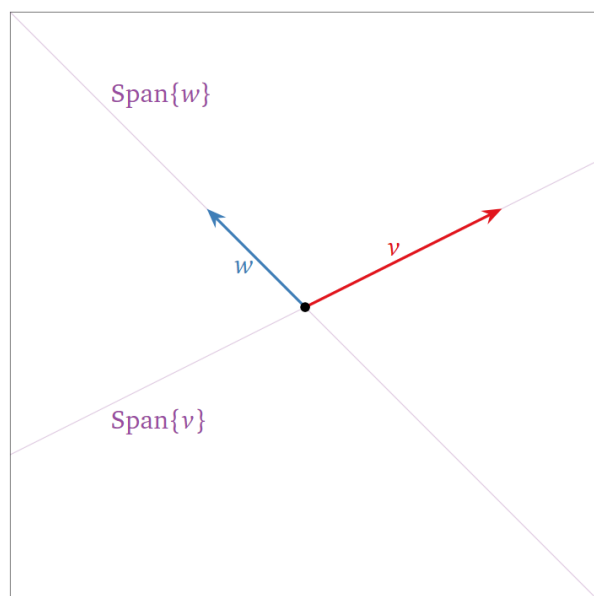


Figure 11: A set of two noncollinear vectors v, w is linearly independent.

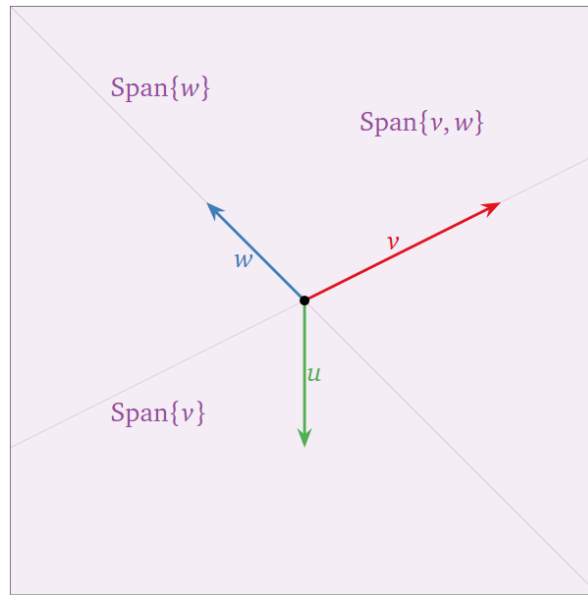


Figure 12: The set of three vector v, w, u above is linearly independent.

2.5.2 Linear Dependence and Free Variables

Theorem

Let v_1, v_2, \dots, v_k be vectors in \mathbb{R}^n , and consider the matrix

$$A = \begin{pmatrix} | & | & \cdots & | \\ v_1 & v_2 & \cdots & v_k \\ | & | & \cdots & | \end{pmatrix}.$$

→ We can delete the columns of A without the pivots, without changing $\text{Span}\{v_1, v_2, \dots, v_k\}$

Proof:

Row Reduction to Row Echelon Form: We apply Gaussian elimination to the matrix A . This matrix will have the properties that we just listed in section 1.2.2

Linear Dependence: The non-pivot columns can be written as linear combinations of the pivot columns. This means that when we remove the non-pivot columns, we are not changing the span of the set of vectors. The span of the original columns is the same as the span of just the pivot columns.

By removing the non-pivot columns, we are only taking away columns that don't add any new directions to the space, since they can already be written as combinations of the pivot columns. The pivot columns, on the other hand, are linearly independent, and removing them would change the span. But as long as we keep the pivot columns, the span remains unchanged.

2.6 Subspaces

In this section, we'll dive into the concept of subspaces in \mathbb{R}^n . A subspace is essentially a span, but we don't always have a specific set of vectors in mind to generate it. This new way of thinking makes it easier to recognize subspaces, even when they aren't clearly expressed as spans. For instance, consider the solution set to the equation $x + 3y + z = 0$. This is a subspace, although we'd need to find its parametric vector form to express it as a span. Throughout this section, we'll learn how to check if a subset is a subspace, as well as explore key examples like column spaces and null spaces.

Definition

A **subset** of \mathbb{R}^n is any collection of points of \mathbb{R}^n . For instance, the unit circle

$$C = \{(x, y) \text{ in } \mathbb{R}^2 \mid x^2 + y^2 = 1\}$$

is a subset of \mathbb{R}^2 .

A **subspace** of \mathbb{R}^n is a subset V of \mathbb{R}^n satisfying:

1. **Non-emptiness:** The zero vector is in V .
2. **Closure under addition:** If u and v are in V , then $u + v$ is also in V .
3. **Closure under scalar multiplication:** If v is in V and c is in \mathbb{R} , then cv is also in V .

As a consequence of these properties, we see:

- If v is a vector in V , then all scalar multiples of v are in V by the third property.
- If u, v are vectors in V and c, d are scalars, then cu, dv are also in V .
→ All of $\text{Span}\{u, v\}$ is contained in V .
- If v_1, v_2, \dots, v_n are all in V , then $\text{Span}\{u, v\}$ is contained in V .

Attention!

Suppose V is a non-empty subset of \mathbb{R}^n that satisfies properties 2 and 3. Let v be any vector in V . By property 3, $0v = 0$ is in V , so V also satisfies property 1. Therefore, the only subset of \mathbb{R}^n that satisfies properties 2 and 3 but not property 1 is the empty set \emptyset . This is why property 1 is called "Non-emptiness".

2.6.1 Common Types of Subspaces

Theorem

"Spans are Subspaces and Subspaces are Spans."

Let v_1, v_2, \dots, v_p be any vectors in \mathbb{R}^n . The span of the set of vectors $\{v_1, v_2, \dots, v_p\}$ forms a subspace of \mathbb{R}^n . Moreover, any subspace of \mathbb{R}^n can be expressed as the span of a set of linearly independent vectors from \mathbb{R}^n , where the number of such vectors is at most n .

If v_1, v_2, \dots, v_k are any vectors in \mathbb{R}^n , then $\text{Span}\{u, v\}$ is a subspace of \mathbb{R}^n . Moreover, any subspace of \mathbb{R}^n can be written as a span of a set of p linearly independent vectors in \mathbb{R}^n for $p \leq n$.

Verifying That Spans Are Subspaces

To prove that $\text{Span}\{v_1, v_2, \dots, v_p\}$ is a subspace, we need to verify the three defining properties of a subspace. These properties are:

- The set contains the zero vector.
- The set is closed under vector addition.
- The set is closed under scalar multiplication.

Property 1 - Zero Vector: The zero vector can always be expressed as a linear combination of the given vectors v_1, v_2, \dots, v_p . Specifically, we have:

$$0 = 0v_1 + 0v_2 + \dots + 0v_p.$$

This guarantees that the zero vector is always in the span.

Property 2 - Closure under Addition: Suppose u and v are two vectors in $\text{Span}\{v_1, v_2, \dots, v_p\}$. This means they can be written as:

$$u = a_1v_1 + a_2v_2 + \dots + a_pv_p, \quad v = b_1v_1 + b_2v_2 + \dots + b_pv_p.$$

The sum of these two vectors is:

$$u + v = (a_1 + b_1)v_1 + (a_2 + b_2)v_2 + \dots + (a_p + b_p)v_p,$$

which is still a linear combination of v_1, v_2, \dots, v_p . Hence, $u + v$ is also in the span.

Property 3 - Closure under Scalar Multiplication: If v is in $\text{Span}\{v_1, v_2, \dots, v_p\}$, then v can be written as:

$$v = a_1v_1 + a_2v_2 + \dots + a_pv_p.$$

If c is a scalar, then the scalar multiple cv becomes:

$$cv = c(a_1v_1 + a_2v_2 + \dots + a_pv_p) = (ca_1)v_1 + (ca_2)v_2 + \dots + (ca_p)v_p,$$

which is also in $\text{Span}\{v_1, v_2, \dots, v_p\}$.

Since $\text{Span}\{v_1, v_2, \dots, v_p\}$ satisfies all three properties, it is indeed a **subspace**.

Subspaces Are Spans

Let V be a subspace of \mathbb{R}^n . If V is the zero subspace, then it can be written as the span of the empty set \emptyset . For nonzero subspaces, we can construct a spanning set as follows:

1. Start with any nonzero vector v_1 in V .
2. If $V = \text{Span}\{v_1\}$, we are done. Otherwise, find another vector v_2 in V that is not in $\text{Span}\{v_1\}$.
3. Continue this process, adding linearly independent vectors v_3, v_4, \dots until $V = \text{Span}\{v_1, v_2, \dots, v_p\}$.
4. The process stops after at most n steps because \mathbb{R}^n can have at most n linearly independent vectors.

Thus, any subspace V can be written as the span of a set of linearly independent vectors, called a spanning set for V .

2.6.2 Column Space and Null Space

Every matrix naturally gives rise to two fundamental subspaces: the column space and the null space.

Definition (Column Space)

Let A be an $m \times n$ matrix. The column space of A , denoted $\text{Col}(A)$, is the subspace of \mathbb{R}^m spanned by the columns of A .

Definition (Null Space)

The null space of A , denoted $\text{Nul}(A)$, is the subspace of \mathbb{R}^n consisting of all solutions to the equation $Ax = 0$. Formally:

$$\text{Nul}(A) = \{x \in \mathbb{R}^n \mid Ax = 0\}.$$

2.6.3 Recipe: Compute a Spanning Set for a Null Space

To find a spanning set for $\text{Nul}(A)$:

1. Solve the homogeneous equation $Ax = 0$ using row reduction.
2. Write the solution in parametric vector form.
3. The vectors corresponding to the free variables form a spanning set for $\text{Nul}(A)$.

Example 1: Let

$$\begin{aligned} V &= \left\{ \begin{pmatrix} a \\ b \end{pmatrix} \text{ in } \mathbb{R}^2 \mid 2a = 3b \right\} \\ &\rightarrow V = \text{Nul} \begin{pmatrix} 2 & -3 \end{pmatrix} \\ \text{Span}(V) &= \left\{ \begin{pmatrix} \frac{3}{2} \\ 1 \end{pmatrix} \right\} \end{aligned}$$

Example 2: Let V be the plane in \mathbb{R}^3 defined by

$$V = \left\{ \begin{pmatrix} 2x + y \\ x - y \\ 3x - 2y \end{pmatrix} \mid x, y \text{ are in } \mathbb{R} \right\}$$

Question: Write V as the column space of matrix?

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

2.7 Basis and Dimension

The goal of this section is to understand the concept of a *basis* for a subspace, the **Basis Theorem**, and methods for finding bases for different types of subspaces, such as column spaces, null spaces, and spans.

In \mathbb{R}^2 or \mathbb{R}^3 , a basis gives us a clear geometric understanding of a subspace. As we saw in section 2.6, a subspace can be described as a *span*, which is the set of all linear combinations of some given vectors. However, the specific vectors used to form the span are not unique; there are infinitely many spanning sets for any nonzero subspace.

To make things simpler and avoid redundancy, we aim to find a spanning set with the smallest possible number of vectors. Such a set is called a **basis**. A basis not only fully describes the subspace but also ensures simplicity and efficiency in representing it.

Definition

Let V be a subspace of \mathbb{R}^n . A **basis** of V is a set of vectors: $\{v_1, v_2, \dots, v_m\}$ in V such that:

1. $V = \text{Span}\{v_1, v_2, \dots, v_m\}$ is linearly independent.
2. The set $\{v_1, v_2, \dots, v_m\}$ is linearly independent.

2.7.1 Dimension

Definition

Let V be a subspace of \mathbb{R}^n . The number of vectors in any basis of V is called the **dimension** of V → denoted as: $\dim(V)$

To ensure that we have a comprehensive and clear understanding of the concept of a basis for a subspace, we provide several illustrative examples below:

Example 1: Find a basis of \mathbb{R}^2 ?

Solution: To find two vectors in \mathbb{R}^2 that span \mathbb{R}^2 and are linearly independent, consider the basis

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

These vectors span \mathbb{R}^2 because any vector $\begin{pmatrix} a \\ b \end{pmatrix}$ can be written as a linear combination:

$$\begin{pmatrix} a \\ b \end{pmatrix} = a \begin{pmatrix} 1 \\ 0 \end{pmatrix} + b \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

They are linearly independent because if

$$x \begin{pmatrix} 1 \\ 0 \end{pmatrix} + y \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

then $x = 0$ and $y = 0$. This confirms that \mathbb{R}^2 has dimension 2.

Example 2: Let

$$V = \left\{ \begin{pmatrix} x \\ y \\ z \end{pmatrix} \in \mathbb{R}^3 \mid x + 3y + z = 0 \right\}, B = \left\{ \begin{pmatrix} -3 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ -3 \end{pmatrix} \right\}.$$

We will verify that V is a subspace and show directly that B is a basis for V .

Solution:

First, observe that V is the solution set of the homogeneous equation $x + 3y + z = 0$, so it is a subspace.

To show that B is a basis, we need to verify three things:

1. Both vectors are in V : - For $\begin{pmatrix} -3 \\ 1 \\ 0 \end{pmatrix}$, we have $(-3) + 3(1) + 0 = 0$. - For $\begin{pmatrix} 0 \\ 1 \\ -3 \end{pmatrix}$, we have $0 + 3(1) + (-3) = 0$.

2. Span: Suppose that $\begin{pmatrix} x \\ y \\ z \end{pmatrix} \in V$. Since $x + 3y + z = 0$, we have $y = -\frac{1}{3}(x + z)$. Thus,

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x \\ -\frac{1}{3}(x + z) \\ z \end{pmatrix} = -\frac{x}{3} \begin{pmatrix} -3 \\ 1 \\ 0 \end{pmatrix} - \frac{z}{3} \begin{pmatrix} 0 \\ 1 \\ -3 \end{pmatrix}.$$

Hence, B spans V .

3. Linear independence: Suppose that

$$c_1 \begin{pmatrix} -3 \\ 1 \\ 0 \end{pmatrix} + c_2 \begin{pmatrix} 0 \\ 1 \\ -3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

This gives the system:

$$\begin{pmatrix} -3c_1 \\ c_1 + c_2 \\ -3c_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix},$$

which leads to the equations:

$$-3c_1 = 0, \quad c_1 + c_2 = 0, \quad -3c_2 = 0.$$

From $-3c_1 = 0$ and $-3c_2 = 0$, we obtain $c_1 = 0$ and $c_2 = 0$. Hence, the vectors are linearly independent. Alternatively, we can observe that the two vectors are not collinear. Since V has a basis with two vectors, it has dimension 2, meaning that V is a plane.

2.7.2 Computing a Basis for a Subspace

Theorem: *The pivot columns of a matrix A form a basis for $\text{Col}(A)$.*

Corollary: *The dimension of $\text{Col}(A)$ is the number of pivots of A .*

A basis of a span: Computing a basis for a span is the same as computing a basis for a column space. Indeed, the span of finitely many vectors v_1, v_2, \dots, v_m is the column space of a matrix, namely, the matrix A whose columns are v_1, v_2, \dots, v_m :

$$A = \begin{pmatrix} | & | & \dots & | \\ v_1 & v_2 & \dots & v_m \\ | & | & \dots & | \end{pmatrix}.$$

Example 1: Find a basis of the subspace

$$V = \text{Span} \left\{ \begin{pmatrix} 1 \\ -2 \\ 2 \end{pmatrix}, \begin{pmatrix} 2 \\ -3 \\ 4 \end{pmatrix}, \begin{pmatrix} 0 \\ 4 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ 5 \\ -2 \end{pmatrix} \right\}.$$

Solution: The subspace V is the column space of the matrix

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

The reduced row echelon form of this matrix is

$$A = \begin{pmatrix} 1 & 0 & -8 & -7 \\ 0 & 1 & 4 & 3 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

The first two columns are pivot columns, so a basis for V is

$$\left\{ \begin{pmatrix} 1 \\ -2 \\ 2 \end{pmatrix}, \begin{pmatrix} 2 \\ -3 \\ 4 \end{pmatrix} \right\}.$$

Example 2: Find a basis of the subspace

$$V = \text{Span} \left\{ \begin{pmatrix} 1 \\ -2 \\ 2 \end{pmatrix}, \begin{pmatrix} 2 \\ -3 \\ 4 \end{pmatrix}, \begin{pmatrix} 0 \\ 4 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ 5 \\ -2 \end{pmatrix} \right\},$$

which does not consist of the first two vectors, as in the previous example.

Solution: The point of this example is that the above gives one basis for V ; as always, there are infinitely more.

Reordering the vectors, we can express V as the column space of

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

The reduced row echelon form of this matrix is

$$\begin{pmatrix} 1 & 0 & 3/4 & 7/4 \\ 0 & 1 & -1 & -2 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

The first two columns are pivot columns, so a basis for V is

$$\left\{ \begin{pmatrix} 0 \\ 4 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ 5 \\ -2 \end{pmatrix} \right\}.$$

These are the last two vectors in the given spanning set.

2.7.3 A basis for the Null space

Theorem

The vectors associated with the free variables in the parametric vector form of the solution set of

$$A\mathbf{x} = \mathbf{0}$$

form a basis for $\text{Nul}(A)$.

Proof: The proof of this theorem involves two main parts:

Part 1: Every solution lies in the span of the given vectors. This part is relatively straightforward. When we solve the equation $A\mathbf{x} = \mathbf{0}$, we express the solution set in terms of free variables. These free variables correspond to independent vectors that span the entire solution set. Essentially, the vectors are specifically chosen so that every solution can be written as a linear combination of these vectors. In other words, the span of these vectors completely captures all solutions to $A\mathbf{x} = \mathbf{0}$.

Part 2: The vectors are linearly independent. This part requires a bit more thought. By construction, the vectors corresponding to the free variables in the parametric form of the solution are linearly independent. This is because each free variable introduces a new independent direction in the null space. As a result, the set of vectors that span $\text{Nul}(A)$ is not only sufficient but also minimal. There are no redundant vectors, ensuring linear independence.

Thus, the vectors corresponding to the free variables form a basis for the null space.

2.7.4 Finding a Basis for a General Subspace

As mentioned earlier, when a subspace is given in a different form (not directly as a column space or null space), the best approach to compute a basis is usually to rewrite it as either the column space or the null space of a matrix.

Steps to Find a Basis for the Null Space

1. Write the homogeneous equation $A\mathbf{x} = \mathbf{0}$.
2. Perform Gaussian elimination (or row reduce the matrix A) to find the general solution of the system.
3. Express the solution in parametric vector form:

$$\mathbf{x} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_k\mathbf{v}_k,$$

where c_1, c_2, \dots, c_k are free parameters and $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k$ are vectors.

4. The set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k\}$ forms a basis for $\text{Nul}(A)$.

2.8 The Rank Theorem

The rank theorem provides a powerful and elegant relationship between two fundamental concepts in linear algebra: the column space and the null space of a matrix. This theorem represents the culmination of much of the work we have done so far, tying together ideas about dimensions and solutions of linear systems.

2.8.1 Definition of Rank and Nullity

The **rank** of a matrix A , written $\text{rank}(A)$, is defined as the dimension of the column space $\text{Col}(A)$, which represents the number of linearly independent columns in A . The **nullity** of a matrix A , written $\text{nullity}(A)$, is the dimension of the null space $\text{Nul}(A)$, which corresponds to the number of free variables in the solutions of $A\mathbf{x} = \mathbf{0}$. Together, these two quantities are related by the rank theorem.

Examples of Column Space and Null Space Dimensions: To build intuition, let us consider examples of matrices of various sizes:

- For a 3×2 matrix A , the column space $\text{Col}(A)$ is a line in \mathbb{R}^2 , and the null space $\text{Nul}(A)$ is a line in \mathbb{R}^3 .

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

- For a 2×3 matrix A , the null space $\text{Nul}(A)$ is a plane in \mathbb{R}^3 , while the column space $\text{Col}(A)$ is a line in \mathbb{R}^2 .

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

- For a 3×3 matrix A , the null space $\text{Nul}(A)$ is a line in \mathbb{R}^3 , and the column space $\text{Col}(A)$ is a plane in \mathbb{R}^3 .

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

In each case, the sum of the dimensions of $\text{Col}(A)$ and $\text{Nul}(A)$ equals the number of columns of A .

2.8.2 The Rank Theorem:

Theorem

If A is a matrix with n columns, then:

$$\text{rank}(A) + \text{nullity}(A) = n.$$

In simpler terms, the dimension of the column space plus the dimension of the null space equals the total number of columns in the matrix.

2.8.3 Implications of the Rank Theorem:

Definition

The rank theorem provides critical insights into the solutions of the equation $A\mathbf{x} = \mathbf{b}$:

- The rank of A , or $\dim(\text{Col}(A))$, represents the number of linearly independent columns and determines the dimension of the set of vectors \mathbf{b} for which $A\mathbf{x} = \mathbf{b}$ is consistent.
- The nullity of A , or $\dim(\text{Nul}(A))$, tells us how many free variables there are in the solution of $A\mathbf{x} = \mathbf{0}$.

Thus, the rank theorem highlights a fundamental trade-off: the more freedom we have in choosing \mathbf{x} (i.e., the larger the nullity), the less freedom we have in choosing \mathbf{b} (i.e., the smaller the rank), and vice versa.

3 Linear Transformations and Matrix Algebra

Linear transformations are a fascinating area of study in linear algebra, connecting abstract mathematical concepts with real-world applications. At their core, transformations are functions: they take an input and produce an output. Often, the questions we ask about transformations revolve around their geometry—how they change, move, or manipulate spaces. Linear algebra provides powerful tools to answer such questions, but only if the transformation can be expressed using a matrix.

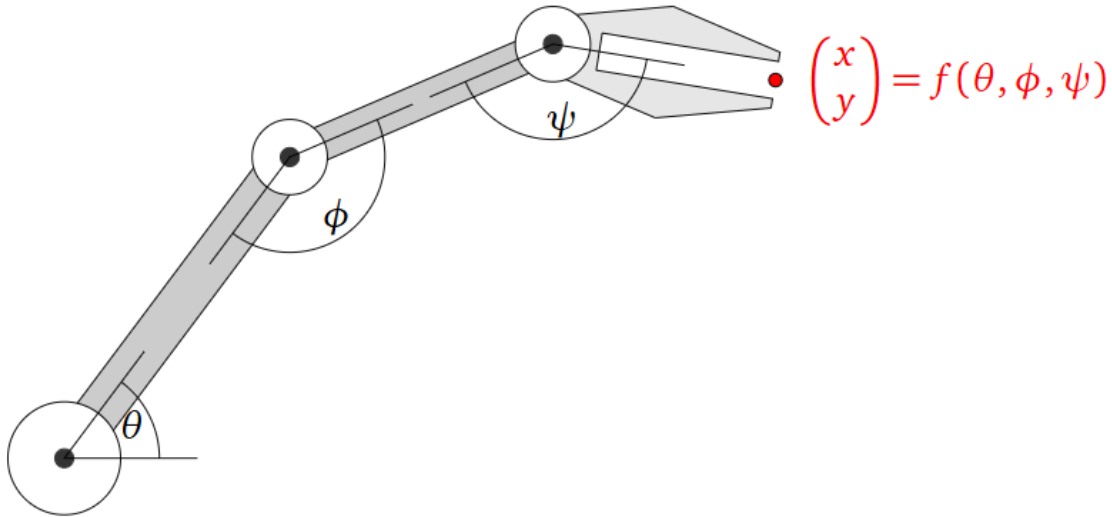


Figure 13: A robotic arm.

Imagine designing a robot arm with three joints. These joints can rotate by angles θ , ϕ , and ψ , allowing the arm to position its hand anywhere on a plane. To determine where the hand will be for given joint angles, we define a transformation $f(\theta, \phi, \psi)$. This function calculates the hand's (x, y) position on the plane based on the angles of the joints.

While this transformation is undoubtedly useful, it cannot be described by a matrix. It is nonlinear, and such transformations fall outside the realm of linear algebra. Instead, their analysis belongs to a more complex field known as *inverse kinematics*.

This example highlights a crucial distinction: not all transformations can be captured by matrices. The focus of this chapter, however, is on those transformations that can. These are the *linear transformations*, which not only simplify computations but also reveal deep insights into the structure of the spaces they act on. In this chapter, we dive into the relationship between matrices and transformations. We start by visualizing the equation

$$b = Ax,$$

where A is a matrix, x is the input, and b is the output. This provides an intuitive way to interpret linear transformations geometrically.

3.1 Matrix Transformation

3.1.1 Matrices as Function

Informally, a *function* is a rule that accepts *input* and produces *output*. For instance, with $f(x) = x^2$ is a function; it accept one input x , and produce one output $f(x)$: $f(2) = 4$. In this

subsection, we interpret matrices as functions.

Definition

Let A be a matrix with m columns and n rows. Consider a matrix equation: $b = Ax$. If we vary x , b will also vary. In this way, we think of A as a function with independent variable x and dependent variable b .

- The independent variable(input) is x , which is a vector in \mathbb{R}^n
- The dependent variable(output) is b , which is a vector in \mathbb{R}^m

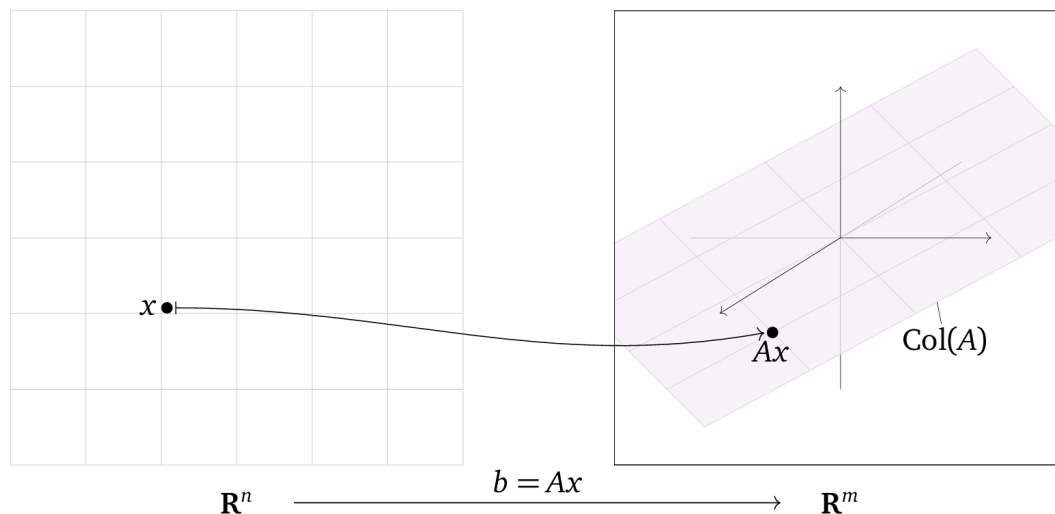


Figure 14: Function $b = Ax$

So what's the difference between normal function and matrix as function? Let see some examples.

Example 1. Let:

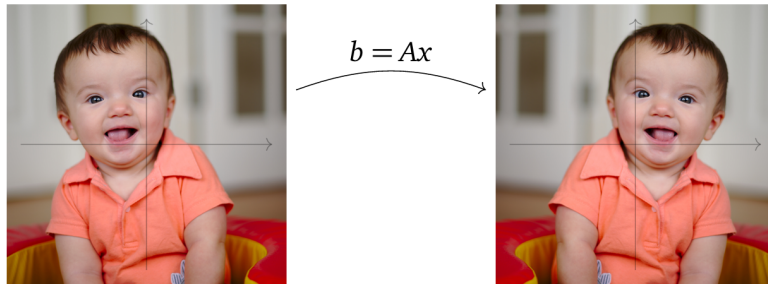
$$A = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$$

Describe the function $b = Ax$ geometrically.

Solutions: In the equation $Ax = b$, the input vector x and the output vector b both in \mathbb{R}^2 . First we multiply A by a vector to see what it does:

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

Multiplication with A negates the coordinate x : it *reflects on the y -axis*



Example 2. Let:

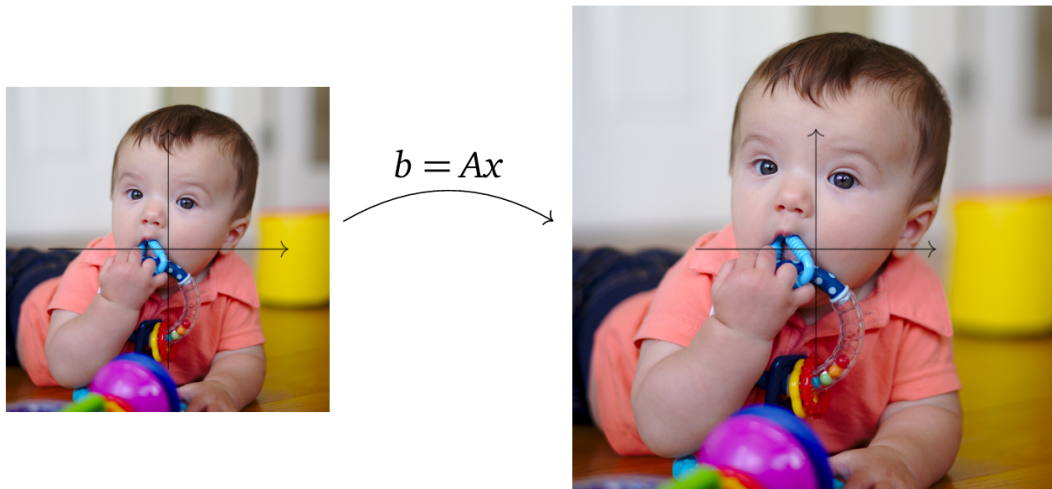
$$A = \begin{pmatrix} 1.5 & 0 \\ 0 & 1.5 \end{pmatrix}$$

Describe the function $b = Ax$ geometrically.

Solutions: In the equation $Ax = b$, the input vector x and the output vector b both in \mathbb{R}^2 . First we multiply A by a vector to see what it does:

$$A \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1.5 & 0 \\ 0 & 1.5 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1.5x \\ 1.5y \end{pmatrix} = 1.5 \begin{pmatrix} x \\ y \end{pmatrix}$$

Multiplication by A is the same as scalar multiplication by 1.5: it *scales* or *dilates the plane* by a factor of 1.5.



With the examples above, we have seen function matrices and how they work. In fact, function matrices are used in many research fields, and we will continue to explore them in the following sections.

3.1.2 Transformations

Definition

A **transformation** from \mathbb{R}^n to \mathbb{R}^m is a rule T that assigns to each vector x in \mathbb{R}^n a vector $T(x)$ in \mathbb{R}^m .

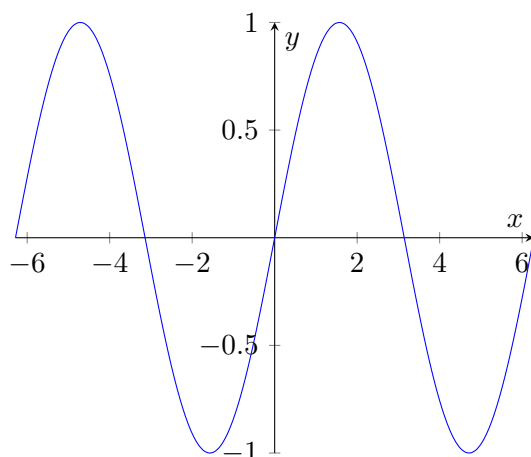
Example 1: Most of the functions that you may have seen previously have a domain and a co-domain equal to \mathbb{R} . For example,

$$f : \mathbb{R} \rightarrow \mathbb{R},$$

$$x \mapsto \sin(x)$$

Notice that we have defined \sin by a rule: a function is defined by specifying what the output of the function is for any possible input.

You may be used to thinking of such functions in terms of their graphs:



In this case, the horizontal axis is the domain, and the vertical axis is the codomain. This is useful when the domain and codomain are \mathbb{R} , but it is hard to do when, for instance, the domain is \mathbb{R}^2 and the codomain is \mathbb{R}^3 . The graph of such a function is a subset of \mathbb{R}^5 , which is difficult to visualize. For this reason, we will rarely graph a transformation.

Example 2: The definition of *transformation* and its associated vocabulary may seem quite abstract, but transformations are extremely common in real life. Here is an example from the fields of robotics and computer graphics.

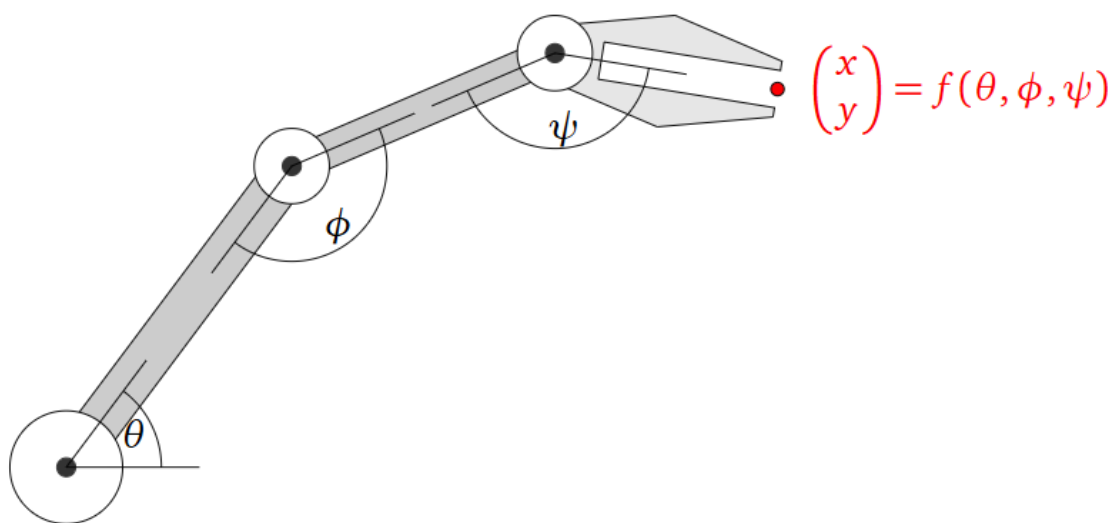


Figure 15: $f : \mathbb{R}^3 \rightarrow \mathbb{R}^2 \quad f(\theta, \phi, \psi) \mapsto (x, y)$

Suppose you are building a robot arm with three joints that can move its hand around a plane, as in the following picture.

3.1.3 Matrix Transformations

Definition

Let A be an $m \times n$ matrix. We have:

$$T: \mathbb{R}^n \rightarrow \mathbb{R}^m \text{ defined by } T(x) = Ax.$$

If A has n columns, then it only makes sense to multiply A by vectors with n entries. This is why the domain of $T(x) = Ax$ is \mathbb{R}^n . If A has m rows, then Ax has m entries for any vector x in \mathbb{R}^n ; this is why the codomain of $T(x) = Ax$ is \mathbb{R}^m .

Suppose that A has columns $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$. If we multiply A by a general vector \mathbf{x} , we get

$$A\mathbf{x} = A \begin{pmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \cdots + x_n\mathbf{v}_n.$$

This is just a general linear combination of $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$. The outputs of $T(\mathbf{x}) = A\mathbf{x}$ are exactly the linear combinations of the columns of A : the range of T is the column space of A .

Properties: Let A be an $m \times n$ matrix, and let $T(x) = Ax$ be the associated matrix transformation.

- The domain of T is \mathbb{R}^n , where n is the number of columns of A .
- The codomain of T is \mathbb{R}^m , where m is the number of rows of A .
- The range of T is the column space of A .

Example 1: Let

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \text{ and let } T(x) = Ax.$$

Domain, Codomain, and Range of T :

- **Domain:** The domain of T is \mathbb{R}^3 because the input vector x has three entries.
- **Codomain:** The codomain of T is also \mathbb{R}^3 because the transformation maps vectors in \mathbb{R}^3 to vectors in \mathbb{R}^3 .
- **Range:** The range of T consists of all vectors that lie on the xy -plane, because the transformation projects vectors directly onto the xy -plane. Every point on the xy -plane is an output of T , with the last entry always being zero. Hence, the range of T is the xy -plane in \mathbb{R}^3 .

Geometrical Interpretation:

The transformation T geometrically projects any vector in \mathbb{R}^3 onto the xy -plane by setting the z -coordinate to zero. This means:

$$T \left(\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \right) = \begin{pmatrix} x_1 \\ x_2 \\ 0 \end{pmatrix}.$$

Thus, the outputs of T are always vectors of the form $\begin{pmatrix} x_1 \\ x_2 \\ 0 \end{pmatrix}$, where x_1, x_2 are arbitrary real numbers, and the third component is always zero.

Example 2: Let

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

and let $T(x) = Ax$, so $T : \mathbb{R}^2 \rightarrow \mathbb{R}^3$ is a matrix transformation.

Evaluate $T(u)$ for $u = \begin{pmatrix} 3 \\ 4 \end{pmatrix}$:

We evaluate $T(u)$ by performing matrix multiplication:

$$T \left(\begin{pmatrix} 3 \\ 4 \end{pmatrix} \right) = A \begin{pmatrix} 3 \\ 4 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 3 \\ 4 \end{pmatrix} = \begin{pmatrix} 7 \\ 4 \\ 7 \end{pmatrix}.$$

Find a vector v such that $T(v) = b$, where $b = \begin{pmatrix} 7 \\ 5 \\ 7 \end{pmatrix}$:

We want to solve the equation $Av = b$. This gives us the system of equations:

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 7 \\ 5 \\ 7 \end{pmatrix}.$$

This translates to the system:

$$x + y = 7,$$

$$y = 5,$$

$$x + y = 7.$$

Substitute $y = 5$ into the first and third equations:

$$x + 5 = 7 \quad \Rightarrow \quad x = 2.$$

Thus, $v = \begin{pmatrix} 2 \\ 5 \end{pmatrix}$.

Is there more than one solution?

Since the system of equations has a unique solution, there is only one vector v such that $T(v) = b$.

Does there exist a vector w in \mathbb{R}^3 such that the solution set of $Av = w$ has more than one vector in it?

The solution set of $Av = w$, if non-empty, is a translate of the solution set of $Av = b$, which has one vector in it. Thus, the solution set of $Av = w$ can have at most one vector.

Find a vector w such that the matrix equation $Av = w$ is not consistent.

Notice that if we take

$$w = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix},$$

then the matrix equation $Av = w$ translates into the system of equations:

$$x + y = 1,$$

$$y = 2,$$

$$x + y = 3,$$

which is clearly inconsistent.

Example 3: Define a transformation $T : \mathbb{R}^2 \rightarrow \mathbb{R}^3$ by the formula

$$T \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \ln(x) \\ \cos(y) \\ \ln(x) \end{pmatrix}.$$

Solution: evaluate $T(u)$ for $u = \begin{pmatrix} 1 \\ \pi \end{pmatrix}$:

We evaluate $T(u)$ using the defining formula:

$$T \begin{pmatrix} 1 \\ \pi \end{pmatrix} = \begin{pmatrix} \ln(1) \\ \cos(\pi) \\ \ln(1) \end{pmatrix} = \begin{pmatrix} 0 \\ -1 \\ 0 \end{pmatrix}.$$

Let $b = \begin{pmatrix} 7 \\ 17 \\ 7 \end{pmatrix}$. Find a vector v in \mathbb{R}^2 such that $T(v) = b$. Is there more than one?

We want to find $v = \begin{pmatrix} x \\ y \end{pmatrix}$ such that:

$$T \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \ln(x) \\ \cos(y) \\ \ln(x) \end{pmatrix} = \begin{pmatrix} 7 \\ 17 \\ 7 \end{pmatrix}.$$

This gives the system:

$$\ln(x) = 7,$$

$$\cos(y) = 17,$$

$$\ln(x) = 7.$$

Since $\cos(y) = 17$ is not possible (because $\cos(y)$ must be between -1 and 1), there is no solution.

Find a vector w in \mathbb{R}^3 which is not in the range of T .

Since $\cos(y)$ is always between -1 and 1, any vector $w = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$ where $y \notin [-1, 1]$ is not in the range of T . For example:

$$w = \begin{pmatrix} 0 \\ 2 \\ 0 \end{pmatrix}$$

is not in the range of T .

3.2 One-to-one and Onto Transformations

When studying transformations in linear algebra, two key questions we ask are whether the transformation is **one-to-one** and whether it is **onto**. These questions help us understand how the transformation works and how it maps elements between spaces.

To make these ideas clearer, think of a transformation as a process that takes inputs from one set (often a vector space) and produces outputs in another. A transformation is one-to-one (or injective) if each output corresponds to a unique input, meaning no two inputs map to the same output. On the other hand, a transformation is onto (or surjective) if every output has at least one corresponding input, meaning the transformation covers the entire output space.

3.2.1 One-to-one Transformations

A *one-to-one transformation* (or *injection*) means that every input maps to a unique output—no two different inputs share the same output.

Imagine assigning people to seats at a concert. If it's one-to-one, each person sits in their own seat, and no two people share one.

For a matrix transformation $T(x) = Ax$, it's one-to-one if different inputs always lead to different outputs, meaning the matrix A has linearly independent columns.

In short, a one-to-one transformation ensures that each input has its own unique output.

Definition (One-to-one transformations)

A transformation:

$$T: \mathbb{R}^n \rightarrow \mathbb{R}^m$$

is **one-to-one** if, for every vector b in \mathbb{R}^m , the equation $T(x) = b$ has at most one solution x in \mathbb{R}^n .

Or in Calculus, we say:

$$\forall x_1, x_2 \in X; x_1 \neq x_2 \Rightarrow T(x_1) \neq T(x_2)$$

Here are some equivalent ways of saying that T is **NOT** one-to-one:

- For every vector b in \mathbb{R}^m , the equation $T(x) = b$ has zero or one solution x in \mathbb{R}^n .
- There are two different inputs of T with the same output.
- There exist vectors u, v such that $u \neq v$, but $T(u) = T(v)$.

Theorem (One-to-one transformations)

Let A be an $m \times n$ matrix and $T(x) = Ax$ the associated transformation. The following are equivalent:

1. T is one-to-one.
2. For every $b \in \mathbb{R}^m$, the equation $T(x) = b$ has at most one solution.
3. For every $b \in \mathbb{R}^m$, $Ax = b$ has either a unique solution or no solution.
4. The homogeneous system $Ax = 0$ has only the trivial solution.
5. The columns of A are linearly independent.
6. A has a pivot in every column.
7. The range of T has dimension n .

Example 1: A matrix transformation that is one-to-one Let A be the matrix

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

and define $T : \mathbb{R}^2 \rightarrow \mathbb{R}^3$ by $T(x) = Ax$.

Is T one-to-one?

Solution: The reduced row echelon form of A is

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

Hence, A has a pivot in every column, so T is one-to-one.

Example 2: A matrix transformation that is not one-to-one Let A be the matrix

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

and define $T : \mathbb{R}^3 \rightarrow \mathbb{R}^2$ by $T(x) = Ax$.

Is T one-to-one?

Solution: The reduced row echelon form of A is

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

There is not a pivot in every column, so T is not one-to-one. Therefore, we know from the fact that $Ax = 0$ has nontrivial solutions. If v is a nontrivial (i.e., nonzero) solution of $Av = 0$, then $T(v) = Av = 0 = A0 = T(0)$, so 0 and v are different vectors with the same output.

In order to find a nontrivial solution, we find the parametric form of the solutions of $Ax = 0$ using the reduced matrix above:

$$x - z = 0$$

$$y + z = 0$$

$$\Rightarrow \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} z \\ -z \\ z \end{pmatrix}$$

The free variable is z . Taking $z = 1$ gives the nontrivial solution:

$$\begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}$$

$$\text{Thus, } T\left(\begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}\right) = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} = T\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}\right).$$

3.2.2 Onto Transformation

An *onto transformation* (or *surjection*) means that every point in the target space has at least one corresponding point in the domain. In simpler terms, every possible output is covered by some input.

Imagine you're assigning people to seats at a concert. If it's "onto," every seat gets filled by at least one person, even if some people take multiple seats. No seat is left empty.

For a matrix transformation $T(x) = Ax$, it's onto if every point in the target space can be reached from the domain. This usually means that the matrix A has enough coverage to map to the entire target space.

In short, an onto transformation ensures every possible output has a matching input.

Definition (Onto Transformation)

A transformation $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is onto if for every vector b in the target space \mathbb{R}^m , there exists at least one vector x in the domain \mathbb{R}^n such that

$$T(x) = b.$$

Or in calculus we say:

$$\forall y \in Y, \exists x \in X : f(x) = y.$$

Here are some equivalent ways of saying that T is onto:

- The range of T is equal to the codomain of T .
- Every vector in the codomain is the output of some input vector.

Here are some equivalent ways of saying that T is not onto:

- The range of T is smaller than the codomain of T .
- There exists a vector b in \mathbb{R}^m such that the equation $T(x) = b$ does not have a solution.
- There is a vector in the codomain that is not the output of any input vector.

Suppose $T(x) = Ax$ is a matrix transformation that is not onto. This means the range of T , or $\text{Col}(A)$, is a subspace of \mathbb{R}^m with dimension less than m . The range is much smaller than the codomain, like a line in 2D or 3D space, or a plane in 3D.

To find a vector not in the range of T , you can pick any random nonzero vector $b \in \mathbb{R}^m$. It's very unlikely that b will be in the range of T . To check if b is in the range, solve the matrix equation $Ax = b$ to see if it's consistent.

Attention!

Tall matrices do not have onto transformations. If $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is an onto matrix transformation, what can we say about the relative sizes of n and m ?

The matrix associated with T has n columns and m rows. Each row and each column can only contain one pivot, so in order for A to have a pivot in every row, it must have at least as many columns as rows: $m \leq n$.

This says that, for instance, \mathbb{R}^2 is “too small” to admit an onto linear transformation to \mathbb{R}^3 .

Note that there exist wide matrices that are not onto.

For example,

$$G = \begin{pmatrix} 10 & 2 & 3 \\ 1 & -1 & 4 \end{pmatrix}$$

does not have a pivot in every row.

3.2.3 Comparison between onto & one-to-one

We’ve provided a table here to help you compare two important types of transformations: one-to-one (injective) and onto (surjective). These concepts are key in understanding functions, and knowing the difference between them is essential. The table below makes it easier to see how these two types of transformations work.

T is one-to-one	T is onto
$T(x) = b$ has at most one solution for every b .	$T(x) = b$ has at least one solution for every b .
The columns of A are linearly independent.	The columns of A span \mathbb{R}^m .
A has a pivot in every column.	A has a pivot in every row.
The range of T has dimension n .	The range of T has dimension m .

By taking a closer look, you’ll be able to clearly tell the difference: a one-to-one function ensures that every element in the domain is paired with a unique element in the codomain, while an onto function guarantees that every element in the codomain has at least one corresponding element in the domain.

3.3 Linear Transformation

In this section, we change our approach. Instead of thinking of matrices as functions, we focus on a transformation we want to understand. If we can show that the transformation is a matrix transformation, we can apply linear algebra to study it. This brings up two key questions:

1. How can we determine if a transformation is a matrix transformation?
2. If it is, how can we find the matrix that represents it?

For example, consider the matrix transformation:

$$T : \mathbb{R}^2 \rightarrow \mathbb{R}^2, \quad T(x) = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} x$$

This represents a counterclockwise rotation of the plane by 90° . However, we could describe T in a more abstract way:

$$T : \mathbb{R}^2 \rightarrow \mathbb{R}^2, \quad T(x) = \text{the counterclockwise rotation of } x \text{ by } 90^\circ.$$

In this form, it's not immediately clear that T is a matrix transformation, or what matrix it corresponds to.

Definition

A **linear transformation** is a transformation $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$ satisfying:

- $T(u + v) = T(u) + T(v)$. ("additivity")
- $T(cu) = cT(u)$. ("homogeneity")

for all vectors u, v in \mathbb{R}^n and all scalars c .

Let $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$ be a matrix transformation $T(x) = Ax$ for an $m \times n$ matrix A . By this proposition in Section 2.4.1, we have:

$$T(u + v) = A(u + v) = Au + Av = T(u) + T(v) \quad T(cu) = A(cu) = cAu = cT(u)$$

Since both properties hold, we can conclude that matrix transformations are indeed linear transformations.

Key Insight In the next section, we'll see that every linear transformation is, in fact, a matrix transformation. This means that for any linear transformation, there is always a matrix that represents it. We might not know the matrix right away, but it's always possible to find one.

Important Facts about Linear Transformations Let $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$ be a linear transformation. Then the following facts hold:

1. Zero Vector Mapping:

$$T(0) = 0.$$

This means that a linear transformation always maps the zero vector to the zero vector. This is a fundamental property of linear transformations.

2. Superposition Principle: For any vectors v_1, v_2, \dots, v_k in \mathbb{R}^n and scalars c_1, c_2, \dots, c_k , we have:

$$T(c_1v_1 + c_2v_2 + \dots + c_kv_k) = c_1T(v_1) + c_2T(v_2) + \dots + c_kT(v_k).$$

This is known as the *superposition principle* and reflects how linear transformations distribute over vector addition and scalar multiplication. For example, if we have vectors u and v and scalars c and d , we can apply the transformation T as follows:

$$T(cu + dv) = cT(u) + dT(v).$$

This ensures that linear transformations behave in a predictable and additive manner.

3.3.1 The Standard Coordinate Vectors

Standard coordinate vectors The standard coordinate vectors in \mathbb{R}^n are the n vectors

$$\mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad \mathbf{e}_2 = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \quad \dots, \quad \mathbf{e}_{n-1} = \begin{pmatrix} 0 \\ \vdots \\ 1 \\ 0 \end{pmatrix}, \quad \mathbf{e}_n = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}.$$

The i th entry of \mathbf{e}_i is equal to 1, and the other entries are zero.

From now on, for the rest of the book, we will use the symbols $\mathbf{e}_1, \mathbf{e}_2, \dots$ to denote the standard coordinate vectors.

There is an inherent ambiguity in the notation of the standard coordinate vectors. Specifically, when we write \mathbf{e}_1 , we must rely on the context to understand how many entries are in the vector. This means that the vector

$$\mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

could be interpreted as a two-dimensional vector in \mathbb{R}^2 , but it could also refer to

$$\mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

in the case of \mathbb{R}^3 , depending on the space we are working in. Both vectors are denoted by the same symbol, \mathbf{e}_1 , but the number of entries—i.e., the dimension of the vector—differs depending on whether we are considering vectors in \mathbb{R}^2 or \mathbb{R}^3 . This can lead to confusion unless the dimension of the space is explicitly stated or can be inferred from the context.

For example, in two-dimensional space, the standard coordinate vectors are:

$$\mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \mathbf{e}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

These vectors have two entries each, corresponding to the x and y axes in the 2D plane. On the other hand, in three-dimensional space, the standard coordinate vectors are:

$$\mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad \mathbf{e}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad \mathbf{e}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

Here, each vector has three entries, corresponding to the x , y , and z axes in 3D space. Notice that the notation \mathbf{e}_1 in both of these cases refers to different vectors—one with two entries and one with three. Therefore, it's essential to keep in mind the dimensionality of the space when using or interpreting standard coordinate vectors. This will help avoid any confusion about the number of entries the vectors have.

The standard coordinate vectors in \mathbb{R}^2 and \mathbb{R}^3 are pictured below.

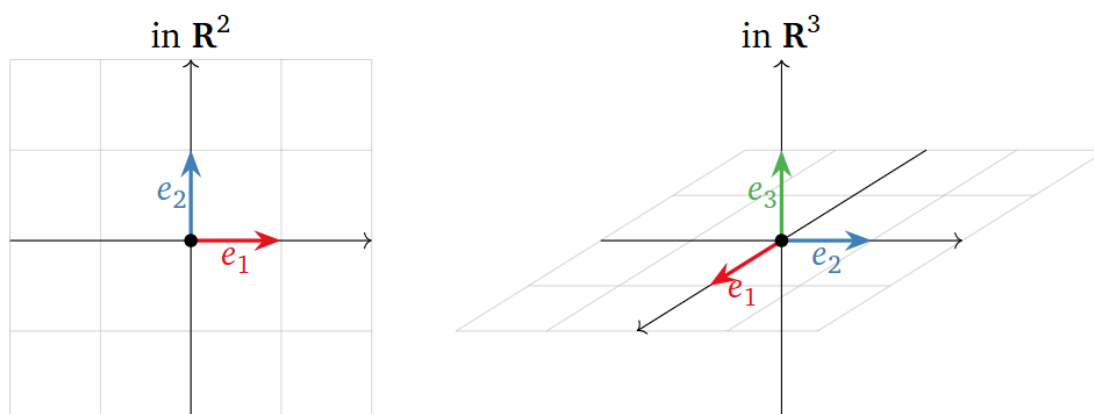


Figure 16: These are the vectors of length 1 that point in the positive directions of each of the axes.

Multiplying a matrix by the standard coordinate vectors If A is an $m \times n$ matrix with columns v_1, v_2, \dots, v_n , then $Ae_i = v_i$ for each $i = 1, 2, \dots, n$:

$$\begin{pmatrix} | & | & \dots & | \\ v_1 & v_2 & \dots & v_n \\ | & | & \dots & | \end{pmatrix} e_i = v_i$$

Definition

The $n \times n$ **identity matrix** is the matrix I_n whose columns are the n standard coordinate vectors in \mathbb{R}^n

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

3.3.2 The Matrix of a Linear Transformation

Theorem (The matrix of a linear transformation)

Let $T: \mathbb{R}^n \rightarrow \mathbb{R}^m$ be a linear transformation. Let A be the $m \times n$ matrix.

$$A = \begin{pmatrix} | & | & \dots & | \\ T(e_1) & T(e_2) & \dots & T(e_n) \\ | & | & \dots & | \end{pmatrix}.$$

\Rightarrow Then T is the **matrix transformation** associated with A : that is, $T(x) = Ax$.

Proof: We suppose for simplicity that T is a transformation from \mathbb{R}^3 to \mathbb{R}^2 . Let A be the matrix given in the statement of the theorem. Then

$$\begin{aligned}
T \begin{pmatrix} x \\ y \\ z \end{pmatrix} &= T \begin{pmatrix} x \\ 0 \\ 0 \end{pmatrix} + T \begin{pmatrix} 0 \\ y \\ 0 \end{pmatrix} + T \begin{pmatrix} 0 \\ 0 \\ z \end{pmatrix} \\
&= T(xe_1 + ye_2 + ze_3) = xT(e_1) + yT(e_2) + zT(e_3) = \begin{pmatrix} T(e_1) & T(e_2) & T(e_3) \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = A \begin{pmatrix} x \\ y \\ z \end{pmatrix}.
\end{aligned}$$

The matrix A described in the above theorem is known as the *standard matrix* for the linear transformation T . Essentially, this matrix A represents the way T acts on vectors in \mathbb{R}^n . The columns of A are the images of the standard coordinate vectors under the transformation T . In other words, each column of the matrix corresponds to the vector that results from applying the transformation T to one of the standard basis vectors in \mathbb{R}^n .

To summarize a key part of the theorem, we observe the following fundamental idea:

Matrix transformations and linear transformations are, in fact, the same thing.

This idea forms the core of the connection between linear transformations and matrices. We can think of a linear transformation as a specific type of matrix transformation, and every matrix represents a linear transformation. This relationship is summarized in a useful correspondence, which can be outlined in the following way:

$$T : \mathbb{R}^n \rightarrow \mathbb{R}^m \quad (\text{Linear transformation}) \quad \longrightarrow \quad \text{an } m \times n \text{ matrix}$$

Here, the linear transformation T takes a vector from \mathbb{R}^n and maps it to \mathbb{R}^m , and this transformation can be fully captured by an $m \times n$ matrix.

Now, let's consider the matrix A . This matrix is formed by taking the images of the standard basis vectors of \mathbb{R}^n under the transformation T . Specifically, the columns of A are the vectors $T(e_1), T(e_2), \dots, T(e_n)$, where e_1, e_2, \dots, e_n are the standard basis vectors in \mathbb{R}^n . Thus, the matrix A is given by:

$$A = \begin{pmatrix} T(e_1) & T(e_2) & \cdots & T(e_n) \end{pmatrix}$$

Finally, we can express the action of the linear transformation T in terms of matrix multiplication. Given a vector $x \in \mathbb{R}^n$, the transformation T acts on x by multiplying it with the matrix A , as follows:

$$T : \mathbb{R}^n \rightarrow \mathbb{R}^m \quad T(x) = Ax \longleftarrow \text{This is the matrix corresponding to linear transformation } A.$$

Every linear transformation can be represented by a matrix, and every matrix transformation is a linear transformation. The matrix captures all the information about the action of the transformation on the vector space.

Question: Verify that the identity transformation $\text{Id}_{\mathbb{R}^n} : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is linear, and compute its standard matrix.

Solution: To verify that the identity transformation $\text{Id}_{\mathbb{R}^n} : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is linear, we need to check the two properties of linearity:

1. Additivity: For any two vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$, we have:

$$\text{Id}_{\mathbb{R}^n}(\mathbf{u} + \mathbf{v}) = \mathbf{u} + \mathbf{v} = \text{Id}_{\mathbb{R}^n}(\mathbf{u}) + \text{Id}_{\mathbb{R}^n}(\mathbf{v}).$$

So, additivity holds.

2. Homogeneity: For any scalar c and vector $\mathbf{u} \in \mathbb{R}^n$, we get:

$$\text{Id}_{\mathbb{R}^n}(c\mathbf{u}) = c\mathbf{u} = c \text{Id}_{\mathbb{R}^n}(\mathbf{u}).$$

Therefore, homogeneity also holds.

Since both properties are satisfied, $\text{Id}_{\mathbb{R}^n}$ is a linear transformation.

The standard matrix of a linear transformation is determined by how it acts on the standard basis vectors. The identity transformation maps each basis vector e_1, e_2, \dots, e_n to itself:

$$\text{Id}_{\mathbb{R}^n}(e_1) = e_1, \quad \text{Id}_{\mathbb{R}^n}(e_2) = e_2, \quad \dots, \quad \text{Id}_{\mathbb{R}^n}(e_n) = e_n.$$

Thus, the standard matrix of $\text{Id}_{\mathbb{R}^n}$ is the identity matrix:

$$\Rightarrow A = I_n.$$

3.4 Matrix Multiplication

In this section, we explore compositions of transformations. Composition allows us to chain multiple transformations together, where the result of one transformation becomes the input for the next. Interestingly, the composition of matrix transformations is equivalent to multiplying the corresponding matrices.

We will also discuss how transformations and matrices can be added together and how they interact with scalar multiplication.

3.4.1 Composition of linear transformations

Composition in linear algebra follows the same idea as in calculus. The definition is as follows:

Definition

Let $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$ and $U : \mathbb{R}^p \rightarrow \mathbb{R}^n$ be transformations. Their composition is the transformation $T \circ U : \mathbb{R}^p \rightarrow \mathbb{R}^m$, defined by:

$$(T \circ U)(x) = T(U(x)).$$

Composing two transformations means chaining them together. The composition $T \circ U$ is the transformation that first applies U , and then applies T (note the order of operations). More precisely, to evaluate $T \circ U$ on an input vector \mathbf{x} , you first apply $U(\mathbf{x})$, then take this output vector of U and use it as the input for T :

$$(T \circ U)(\mathbf{x}) = T(U(\mathbf{x})).$$

This composition is valid only when the outputs of U are valid inputs for T , meaning the range of U must be contained within the domain of T .

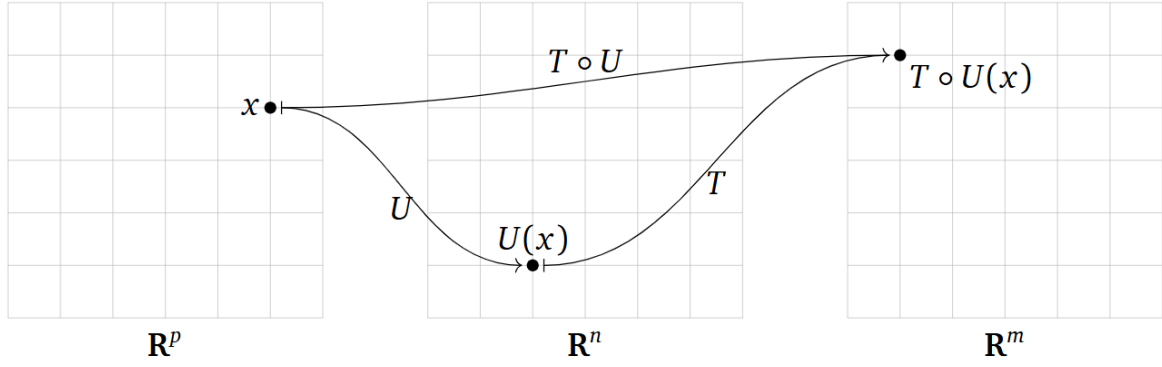


Figure 17: Composition $T \circ U$

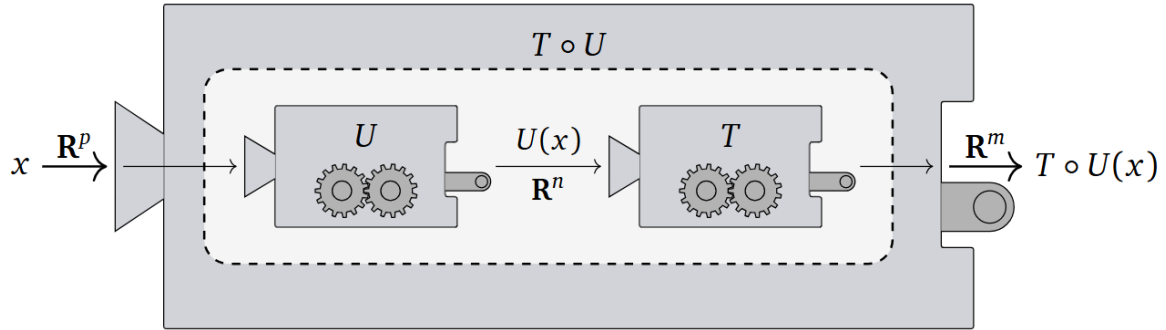


Figure 18: Composition $T \circ U$

Domain and codomain of a composition

- In order for $T \circ U$ to be defined, the codomain of U must equal to the domain of T .
- The domain of $T \circ U$ is the domain of U .
- The codomain of $T \circ U$ is the codomain of T .

Example 1: Define $f : \mathbb{R} \rightarrow \mathbb{R}$ by $f(x) = x^2$ and $g : \mathbb{R} \rightarrow \mathbb{R}$ by $g(x) = x^3$. The composition $f \circ g : \mathbb{R} \rightarrow \mathbb{R}$ is the transformation defined by the rule:

$$(f \circ g)(x) = f(g(x)) = f(x^3) = (x^3)^2 = x^6.$$

Example 2: Define $T : \mathbb{R}^3 \rightarrow \mathbb{R}^2$ and $U : \mathbb{R}^2 \rightarrow \mathbb{R}^3$ by

$$T(x) = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \end{pmatrix} x \quad \text{and} \quad U(x) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \end{pmatrix} x.$$

Their composition is a transformation $T \circ U : \mathbb{R}^2 \rightarrow \mathbb{R}^2 \Rightarrow$ The matrix transformation associated to the matrix $\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$.

3.4.2 Matrix multiplication

Matrix multiplication corresponds precisely to the **composition** of the corresponding linear transformations.

Terminology Let A be an $m \times n$ matrix. We will generally use a_{ij} to denote the entry in the i -th row and the j -th column of the matrix. This is referred to as the *entry* in the i, j position of the matrix.

$$\begin{pmatrix}
 a_{11} & \cdots & a_{1j} & \cdots & a_{1n} \\
 \vdots & & \vdots & & \vdots \\
 a_{i1} & \cdots & a_{ij} & \cdots & a_{in} \\
 \vdots & & \vdots & & \vdots \\
 a_{m1} & \cdots & a_{mj} & \cdots & a_{mn}
 \end{pmatrix}$$

jth column

ith row

Here, a_{11} refers to the element in the first row and first column, a_{12} refers to the element in the first row and second column, and so on.

Definition (Matrix multiplication)

Consider two matrices: A , which is an $m \times n$ matrix, and B , which is an $n \times p$ matrix. Notice that the number of columns in A must match the number of rows in B for the multiplication to be defined. The matrix product AB will result in a new matrix, which we'll call C , and C will have dimensions $m \times p$ (i.e., the number of rows from A and the number of columns from B).

To understand this, think of the matrix B as being made up of column vectors, like this:

$$B = \begin{pmatrix} | & | & | & | \\ v_1 & v_2 & \cdots & v_p \\ | & | & | & | \end{pmatrix}$$

Here, each v_j represents the j -th column of matrix B . Now, the product AB is calculated by multiplying matrix A with each of the column vectors v_1, v_2, \dots, v_p , which results in a new matrix:

$$AB = (Av_1 \quad Av_2 \quad \cdots \quad Av_p)$$

In simpler terms, matrix multiplication is computed column-by-column. Each column of the resulting matrix AB is obtained by multiplying A with the corresponding column from matrix B .

Matrix Dimensions and Multiplication Rules One important thing to keep in mind is the dimensional compatibility required for matrix multiplication. To perform the multiplication AB , the number of columns of matrix A must equal the number of rows of matrix B . Specifically:

- If A is an $m \times n$ matrix and B is an $n \times p$ matrix, the resulting matrix AB will have dimensions $m \times p$. - If matrix B has only one column, then the resulting matrix AB will also have one column. This is similar to a matrix-vector multiplication where matrix A acts on a vector (i.e., a column matrix).

Furthermore, if A is a square matrix (i.e., the number of rows equals the number of columns), then it's possible to multiply it by itself, and we define its powers as:

$$A^2 = A \cdot A, \quad A^3 = A \cdot A \cdot A, \quad \text{and so on.}$$

Row-Column Rule for Matrix Multiplication To make matrix multiplication more concrete, let's look at the row-column rule, which describes how to compute each entry of the resulting matrix.

Recall that the product of a row vector and a column vector is simply a scalar (a number).

For instance, the product of the row vector $(a_1 \quad a_2 \quad \cdots \quad a_n)$ and the column vector $\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$ results in the scalar:

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n$$

In matrix multiplication, the rule is similar: to find the entry c_{ij} of the product matrix C , we compute the dot product of the i -th row of A and the j -th column of B . This is expressed as:

$$c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{in}b_{nj}$$

This procedure is known as the *row-column rule*, and it gives us an easy way to compute matrix products one entry at a time.

$$C = AB = \begin{pmatrix} c_{11} & c_{12} & \cdots & c_{1p} \\ c_{21} & c_{22} & \cdots & c_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ c_{m1} & c_{m2} & \cdots & c_{mp} \end{pmatrix}$$

For example, let's say you want to compute the first entry c_{11} of the resulting matrix C . You would multiply the first row of A by the first column of B and sum the products:

$$c_{11} = a_{11}b_{11} + a_{12}b_{21} + \cdots + a_{1n}b_{n1}$$

This process is repeated for each entry in the matrix.

Proof:

$$Ax = \begin{pmatrix} r_1 \\ r_2 \\ \vdots \\ r_m \end{pmatrix} x = \begin{pmatrix} r_1 x \\ r_2 x \\ \vdots \\ r_m x \end{pmatrix}. \quad (7)$$

The matrix multiplication is:

$$A \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_p \end{pmatrix} = \begin{pmatrix} Ac_1 \\ Ac_2 \\ \vdots \\ Ac_p \end{pmatrix}.$$

It follows that:

$$\begin{pmatrix} r_1 \\ r_2 \\ \vdots \\ r_m \end{pmatrix} (c_1 \ c_2 \ \cdots \ c_p) = \begin{pmatrix} r_1 c_1 & r_1 c_2 & \cdots & r_1 c_p \\ r_2 c_1 & r_2 c_2 & \cdots & r_2 c_p \\ \vdots & \vdots & \ddots & \vdots \\ r_m c_1 & r_m c_2 & \cdots & r_m c_p \end{pmatrix}.$$

Properties and Caveats of Matrix Multiplication Matrix multiplication behaves in many ways as one would expect from regular multiplication, but there are some important caveats to keep in mind:

- **Non-commutativity:** In general, matrix multiplication is *not* commutative. This means that $AB \neq BA$ in most cases, even when both products are defined. For instance, multiplying matrix A by matrix B may not yield the same result as multiplying B by A .
- **Cancellation law does not hold:** The cancellation law, which states that if $AB = AC$ then $B = C$, does not always hold. This is an important property to be aware of when working with matrix equations.
- **Zero product:** It is possible for $AB = 0$ even if neither A nor B is the zero matrix. This can happen due to the structure of the matrices involved.

Although matrix multiplication is not commutative in general, there are some special cases where $AB = BA$. For example, if A is the zero matrix or if $A = B$, the multiplication will be commutative.

Matrix Multiplication as Composition of Transformations Now, let's relate matrix multiplication to linear transformations. Let $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$ and $U : \mathbb{R}^p \rightarrow \mathbb{R}^n$ be linear transformations, and let A and B represent their corresponding standard matrices, respectively. Recall that the composition of these transformations, $T \circ U$, applies U to a vector x and then applies T to the result:

$$(T \circ U)(x) = T(U(x))$$

On the matrix side, the standard matrix for the composition $T \circ U$ is the matrix product AB , so:

$$(T \circ U)(x) = (AB)x$$

By associativity of matrix multiplication, we can rewrite this as:

$$(AB)x = A(Bx)$$

This shows that to compute the result of the composition of two transformations, we first apply B to x , and then apply A to the result.

Thus, matrix multiplication directly mirrors the process of composing linear transformations. It is important to note that matrices and transformations are written in opposite order: the matrix for $T \circ U$ is AB , but the transformations are applied in the order U first, then T .

3.4.3 Composition and Matrix Multiplication

The main idea of this subsection is to demonstrate that matrix multiplication is directly related to the composition of linear transformations. In simpler terms, the standard matrix for $T \circ U$ (the composition of transformations T and U) is simply the product of the standard matrices for T and U . At first, it may seem hard to believe that the seemingly complex formula for matrix multiplication actually represents something as intuitive as "chaining two transformations together." But once we understand this connection, it becomes a powerful and insightful concept!

Theorem

Let $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$ and $U : \mathbb{R}^p \rightarrow \mathbb{R}^n$ be linear transformations, and let A and B be their standard matrices, respectively, so A is an $m \times n$ matrix and B is an $n \times p$ matrix. Then $T \circ U : \mathbb{R}^p \rightarrow \mathbb{R}^m$ is a linear transformation, and its standard matrix is the product AB .

Proof: First, we verify that $T \circ U$ is linear. Let u, v be vectors in \mathbb{R}^p . Then

$$(T \circ U)(u + v) = T(U(u + v)) = T(U(u) + U(v)) = T(U(u)) + T(U(v)) = (T \circ U)(u) + T \circ U(v).$$

If c is a scalar, then

$$(T \circ U)(cv) = T(U(cv)) = T(cU(v)) = cT(U(v)) = c(T \circ U)(v).$$

Since $T \circ U$ satisfies the two properties, it is a linear transformation.

Now that we know that $T \circ U$ is linear, it makes sense to compute its standard matrix. Let C be the standard matrix of $T \circ U$, so $T(x) = Ax$, $U(x) = Bx$, and $T \circ U(x) = Cx$. By this definition, the first column of C is Ce_1 , and the first column of B is Be_1 .

We have

$$(T \circ U)(e_1) = T(U(e_1)) = T(Be_1) = A(Be_1).$$

By definition, the first column of the product AB is the product of A with the first column of B , which is Be_1 , so

$$Ce_1 = (T \circ U)(e_1) = A(Be_1) = (AB)e_1.$$

It follows that C has the same first column as AB . The same argument, applied to the i -th standard coordinate vector e_i , shows that C and AB have the same i -th column; since they have the same columns, they are the same matrix.

Products and compositions: The matrix of the composition of two linear transformations is the product of the matrices of the transformations.

3.4.4 The algebra of transformations and matrices

In this subsection, we explore two additional operations that can be performed on transformations: *addition* and *scalar multiplication*. We then translate these operations into the language of matrices, providing a direct analogy to how we handled the composition of linear transformations earlier. However, these operations are somewhat less subtle than composition.

Definition

Consider two transformations $T, U : \mathbb{R}^n \rightarrow \mathbb{R}^m$. The *sum* of these two transformations, denoted $T + U$, is another transformation that also maps from \mathbb{R}^n to \mathbb{R}^m . For any vector $x \in \mathbb{R}^n$, the value of the transformation $(T + U)(x)$ is simply the sum of the values of $T(x)$ and $U(x)$:

$$(T + U)(x) = T(x) + U(x).$$

This operation of addition is only valid when both transformations T and U have the same domain and codomain, meaning they map from the same input space and to the same output space.

Scalar Multiplication of a Transformation Next, we define the operation of scalar multiplication for a transformation. Let $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$ be a transformation, and let c be a scalar. The *scalar product* of c with T is a new transformation, denoted cT , which also maps from \mathbb{R}^n to \mathbb{R}^m . For any vector $x \in \mathbb{R}^n$, the value of the transformation $(cT)(x)$ is simply the scalar multiple of $T(x)$:

$$(cT)(x) = c \cdot T(x).$$

Thus, multiplying a transformation by a scalar scales the output of the transformation by that scalar.

Properties of Addition and Scalar Multiplication for Transformations Now that we have defined these operations, let us explore some of their properties. Let $T, U, S : \mathbb{R}^n \rightarrow \mathbb{R}^m$ be transformations, and let c, d be scalars. The following properties of addition and scalar multiplication hold for transformations:

1. **Commutativity of Addition:** The sum of two transformations is commutative, meaning that:

$$T + U = U + T.$$

This property states that the order in which we add transformations does not matter.

2. **Associativity of Addition:** Addition of transformations is associative, which means:

$$(T + U) + S = T + (U + S).$$

In other words, when adding more than two transformations, the grouping of terms does not affect the result.

3. **Distributivity of Scalar Multiplication over Addition:** Scalar multiplication distributes over the addition of transformations:

$$c(T + U) = cT + cU.$$

This property tells us that multiplying the sum of two transformations by a scalar is the same as multiplying each transformation by the scalar and then adding the results.

4. **Distributivity of Scalar Addition over Scalar Multiplication:** Scalar addition distributes over scalar multiplication of transformations:

$$(c + d)T = cT + dT.$$

This means that the sum of two scalars, when multiplied by a transformation, is the same as multiplying each scalar by the transformation and adding the results.

5. **Multiplying by a Scalar:** If we multiply a transformation T by a scalar d , this can be written as:

$$dT = (cd)T.$$

In other words, multiplying the transformation T by a scalar d is the same as multiplying the transformation by the scalar product of c and d .

6. **Identity Element:** The zero transformation, denoted 0 , is the additive identity for transformations. That is, for any transformation T , we have:

$$T + 0 = T.$$

The zero transformation maps every vector x to the zero vector, i.e., $0(x) = 0$ for all $x \in \mathbb{R}^n$.

Matrix Analog of Addition and Scalar Multiplication Next, we translate the operations of addition and scalar multiplication into the language of matrices. If we have two matrices A and B of size $m \times n$, their sum is the matrix obtained by adding their corresponding entries:

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

In other words, the (i, j) -entry of $A + B$ is the sum of the (i, j) -entries of A and B .

Similarly, the scalar product of a scalar c with a matrix A is obtained by multiplying each entry of A by c :

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

The scalar c is multiplied by each entry of the matrix A , scaling the entire matrix by c .

Properties of Addition and Scalar Multiplication for Matrices For matrices, we have a similar set of properties for addition and scalar multiplication. Let A, B, C be $m \times n$ matrices, and let c, d be scalars. Then, the following properties hold for matrices:

1. **Commutativity of Matrix Addition:**

$$A + B = B + A.$$

2. **Associativity of Matrix Addition:**

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

3. **Distributivity of Scalar Multiplication over Matrix Addition:**

$$c(A + B) = cA + cB.$$

4. **Distributivity of Scalar Addition over Scalar Multiplication:**

$$(c + d)A = cA + dA.$$

5. **Multiplying by a Scalar:**

$$dA = (cd)A.$$

6. **Identity Element:** The zero matrix, denoted 0 , is the additive identity for matrices. That is,

$$A + 0 = A.$$

The zero matrix is the matrix whose entries are all zero.

Matrix Multiplication and Its Properties Finally, we can combine these operations with matrix multiplication. Matrix multiplication corresponds to the composition of linear transformations. Thus, the following properties of matrix multiplication can be derived from the corresponding properties of transformation composition:

For matrices A, B, C and scalar c , the following identities hold:

1. **Distributivity of Matrix Multiplication over Addition:**

$$C(A + B) = CA + CB.$$

2. **Distributivity of Scalar Multiplication over Matrix Multiplication:**

$$c(AB) = (cA)B = A(cB).$$

3. **Associativity of Matrix Multiplication:**

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

4. **Multiplying by the Identity Matrix:**

$$AI_n = A, \quad AI_m = A.$$

5. **Matrix Multiplication is Not Commutative:** In general, matrix multiplication is not commutative. That is,

$$AB \neq BA.$$

The associativity property $(AB)C = A(BC)$ is generally easier to prove using the composition of transformations, as composition of transformations is associative. However, the commutativity of matrix multiplication does not hold in general, and this is something to keep in mind when working with matrices.

3.5 Matrix Inverses

In section 3.1, we learned how to multiply matrices. Now, in this section, we explore how to "divide" by a matrix. This concept allows us to solve matrix equations more elegantly. For example, given the equation

$$A\mathbf{x} = \mathbf{b},$$

we can solve for \mathbf{x} using the matrix inverse:

$$\mathbf{x} = A^{-1}\mathbf{b}.$$

However, it's important to remember that not every matrix has an inverse, and the order of multiplication must be carefully considered when dealing with matrices.

3.5.1 Invertible Matrices

The reciprocal or inverse of a nonzero number a is the number b which is characterized by the property that $ab = 1$. For instance, the inverse of 7 is $\frac{1}{7}$. We use this formulation to define the inverse of a matrix.

Definition

Let A be an $n \times n$ (square) matrix. We say that A is invertible if there exists an $n \times n$ matrix B such that

$$AB = I_n \quad \text{and} \quad BA = I_n.$$

In this case, B is called the inverse of A , and we write $B = A^{-1}$.

Facts about Invertible Matrices: Let A and B be invertible $n \times n$ matrices.

- A^{-1} is invertible, and its inverse is $(A^{-1})^{-1} = A$.
- AB is invertible, and its inverse is $(AB)^{-1} = B^{-1}A^{-1}$ (note the order).

Proof: The equations $AA^{-1} = I_n$ and $A^{-1}A = I_n$ show that A^{-1} is the inverse of A , and A is the inverse of A^{-1} .

We compute:

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

This shows that $B^{-1}A^{-1}$ is the inverse of AB .

Why is $(AB)^{-1} \neq A^{-1}B^{-1}$?

If it were, then:

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

However, there's no reason for $ABA^{-1}B^{-1}$ to equal the identity matrix, because the order of multiplication matters. In fact, $(AB)^{-1} = A^{-1}B^{-1}$ if and only if $AB = BA$.

In general, the inverse of a product of several invertible matrices is the product of the inverses, in the opposite order. For example:

$$(ABC)^{-1} = C^{-1}B^{-1}A^{-1}.$$

Definition

The determinant of a 2×2 matrix is:

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

Proposition: Let $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. If $\det(A) \neq 0$, then A is invertible, and:

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

If $\det(A) = 0$, then A is not invertible.

Proof: If $\det(A) \neq 0$, define:

$$B = \frac{1}{\det(A)} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

We compute:

$$AB = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} = \frac{1}{\det(A)} \begin{pmatrix} ad - bc & -ab + ab \\ -cd + cd & ac - bd \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = I_2.$$

Thus, $AB = I_2$, and similarly, $BA = I_2$, so $B = A^{-1}$.

Computing the Inverse of Larger Matrices The formula for the inverse of a 2×2 matrix can be extended to larger matrices, but it is more complex and computationally intensive. One way to compute the inverse of an $n \times n$ matrix is by augmenting the matrix A with the identity matrix and performing row reduction to obtain the inverse.

Theorem

Let A be an $n \times n$ matrix. Form the augmented matrix $(A|I_n)$. If the reduced row echelon form of $(A|I_n)$ is $(I_n|B)$, then A is invertible, and $B = A^{-1}$. If the reduced row echelon form does not have this form, A is not invertible.

Example Let:

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

We augment A with the identity matrix:

$$(A|I_3) = \begin{pmatrix} 1 & 0 & 4 & 1 & 0 & 0 \\ 0 & 1 & 2 & 0 & 1 & 0 \\ -3 & -4 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

After performing row reduction, we obtain:

$$(I_3|A^{-1}) = \begin{pmatrix} 1 & 0 & 0 & \frac{2}{5} & \frac{-4}{5} & \frac{-1}{5} \\ 0 & 1 & 0 & \frac{-3}{10} & \frac{3}{5} & \frac{-1}{10} \\ 0 & 0 & 1 & \frac{3}{20} & \frac{1}{5} & \frac{1}{20} \end{pmatrix}.$$

Example of a Non-Invertible Matrix: Let:

$$A = \begin{pmatrix} 5 & 0 & 9 \\ 0 & 5 & 0 \\ 0 & -4 & 0 \end{pmatrix}.$$

We augment A with the identity matrix:

$$(A|I_3) = \begin{pmatrix} 5 & 0 & 9 & 1 & 0 & 0 \\ 0 & 5 & 0 & 0 & 1 & 0 \\ 0 & -4 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

After row reduction, we see that the reduced row echelon form will have a row of zeros, indicating that A is not invertible. We also see that:

$$\det(A) = 0$$

3.5.2 Solving Linear Systems using Inverses

Theorem

Let A be an invertible $n \times n$ matrix, and let b be a vector in \mathbb{R}^n . Then the matrix equation $Ax = b$ has exactly one solution:

$$x = A^{-1}b.$$

Proof: We start with the equation $Ax = b$. Multiply both sides by A^{-1} :

$$A^{-1}(Ax) = A^{-1}b.$$

By associativity of matrix multiplication, this becomes:

$$(A^{-1}A)x = A^{-1}b.$$

Since $A^{-1}A = I_n$, the identity matrix, we have:

$$I_n x = A^{-1}b.$$

Finally, $I_n x = x$ for any vector x , so:

$$x = A^{-1}b.$$

Thus, the matrix equation $Ax = b$ has the unique solution $x = A^{-1}b$.

3.5.3 Invertible Linear Transformations

Definition

A transformation $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is invertible if there exists a transformation $U : \mathbb{R}^n \rightarrow \mathbb{R}^n$ such that

$$(T \circ U) = \text{Id}_{\mathbb{R}^n} \quad \text{and} \quad (U \circ T) = \text{Id}_{\mathbb{R}^n}.$$

In this case, the transformation U is called the inverse of T , and we write $U = T^{-1}$. The inverse U of T "undoes" whatever T did. We have:

$$(T \circ U)(x) = x \quad \text{and} \quad (U \circ T)(x) = x \quad \text{for all vectors } x.$$

This means that if you apply T to x , and then apply U , you get the vector x back, and likewise in the other order.

3.5.4 The Invertible Matrix Theorem

Definition

Let A be an $n \times n$ matrix, and let $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$ be the matrix transformation $T(x) = Ax$. The following statements are equivalent:

1. A is invertible.
2. A has n pivots.
3. $\text{Nul}(A) = \{0\}$.
4. The columns of A are linearly independent.
5. The columns of A span \mathbb{R}^n .
6. $Ax = b$ has a unique solution for each $b \in \mathbb{R}^n$.
7. T is invertible.
8. T is one-to-one.
9. T is onto.

Proof: We will prove that the following statements are equivalent:

1. A is invertible.
2. A has n pivots.
3. $\text{Nul}(A) = \{0\}$.
4. The columns of A are linearly independent.
5. The columns of A span \mathbb{R}^n .
6. $Ax = b$ has a unique solution for each $b \in \mathbb{R}^n$.
7. T is invertible.
8. T is one-to-one.
9. T is onto.

We prove this by showing that each statement implies the next in the list:

1. **(1) implies (2):** If A is invertible, then it has full rank, meaning that it has n pivots in its row echelon form.
2. **(2) implies (3):** If A has n pivots, then the homogeneous system $Ax = 0$ only has the trivial solution, meaning that $\text{Nul}(A) = \{0\}$.
3. **(3) implies (4):** If $\text{Nul}(A) = \{0\}$, then the columns of A are linearly independent because there is no nontrivial solution to $Ax = 0$.
4. **(4) implies (5):** If the columns of A are linearly independent, they span \mathbb{R}^n , because a set of n linearly independent vectors in \mathbb{R}^n must span the entire space.

5. **(5) implies (6):** If the columns of A span \mathbb{R}^n , then for each $b \in \mathbb{R}^n$, the system $Ax = b$ has at least one solution. Since the columns are linearly independent (as shown in step (4)), the solution must be unique.
6. **(6) implies (7):** If $Ax = b$ has a unique solution for each $b \in \mathbb{R}^n$, then the transformation T is invertible because for every x , there is a unique $T(x)$, and the inverse transformation exists.
7. **(7) implies (8):** If T is invertible, it is one-to-one, because if $T(x_1) = T(x_2)$, we can apply the inverse to get $x_1 = x_2$.
8. **(8) implies (9):** If T is one-to-one, then it is onto because for every $b \in \mathbb{R}^n$, there is a unique $x \in \mathbb{R}^n$ such that $T(x) = b$, implying that the transformation covers the entire space.
9. **(9) implies (1):** If T is onto, then every vector in \mathbb{R}^n is in the image of T , meaning that the matrix A has full column rank and is invertible.

Thus, we have shown that all the statements are equivalent.

4 Determinant

4.1 Determinants: Definition

4.1.1 The Definition of the Determinant

The determinant of a square matrix A is a real number $\det(A)$. It is defined via its behavior with respect to row operations; this means we can use row reduction to compute it. We will give a recursive formula for the determinant in **Section 4.2**. We will also show in this subsection that the determinant is related to invertibility, and in **Section 4.3** that it is related to volumes.

Definition

The determinant is a function

$$\det : \{\text{square matrices}\} \rightarrow \mathbb{R}$$

satisfying the following properties:

1. Doing a row replacement on A does not change $\det(A)$.
2. Scaling a row of A by a scalar c multiplies the determinant by c .
3. Swapping two rows of a matrix multiplies the determinant by -1 .
4. The determinant of the identity matrix I_n is equal to 1.

In other words, to every square matrix A we assign a number $\det(A)$ in a way that satisfies the above properties.

In each of the first three cases, doing a row operation on a matrix scales the determinant by a **nonzero number**. (Multiplying a row by zero is not a row operation.) Therefore, doing row operations on a square matrix A does not change whether or not the determinant is zero.

The main motivation behind using these particular defining properties is geometric; see **Section 4.3**. Another motivation for this definition is that it tells us how to compute the determinant: we row reduce and keep track of the changes.

Example 1:

Let us compute $\det \left(\begin{pmatrix} 2 & 1 \\ 1 & 4 \end{pmatrix} \right)$. First we row reduce, then we compute the determinant in the opposite order:

$$\begin{pmatrix} 2 & 1 \\ 1 & 4 \end{pmatrix} \quad \det = 7$$

$$\xrightarrow{R_2 \rightarrow R_2 - \frac{1}{2}R_1} \begin{pmatrix} 2 & 1 \\ 0 & \frac{7}{2} \end{pmatrix} \quad \det = 7$$

$$\xrightarrow{R_2 \rightarrow \frac{2}{7}R_2} \begin{pmatrix} 2 & 1 \\ 0 & 1 \end{pmatrix} \quad \det = 2$$

$$\xrightarrow{R_1 \rightarrow R_1 - R_2} \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix} \quad \det = 2$$

$$\xrightarrow{R_1 \rightarrow \frac{1}{2}R_1} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \det = 1.$$

Note that our answer agrees with this **definition** of the determinant.

Example 2:

Compute $\det \left(\begin{pmatrix} 1 & 0 \\ 0 & 3 \end{pmatrix} \right)$.

Solution. Let $A = \begin{pmatrix} 1 & 0 \\ 0 & 3 \end{pmatrix}$. Since A is obtained from the identity matrix I_2 by multiplying the second row by the constant 3, the determinant is:

$$\det(A) = 3 \det(I_2) = 3 \cdot 1 = 3.$$

Thus, the determinant of A is 3, which aligns with the definition of the determinant.

Note: Our answer agrees with the properties of the determinant.

Example 3: Compute $\det \left(\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 5 & 1 & 0 \end{pmatrix} \right)$.

Solution. To find the determinant, we row reduce the matrix step by step and compute the determinant in reverse order:

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 5 & 1 & 0 \end{pmatrix} \quad \det = -1$$

$$\xrightarrow{R_3 \rightarrow R_3 - 5R_1} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \quad \det = 1$$

$$\xrightarrow{R_2 \leftrightarrow R_3} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \det = 1.$$

After performing the row operations, the reduced row echelon form of the matrix is I_3 , which has a determinant of 1. Working backward:

1. Swapping rows introduces a factor of -1 . 2. The remaining steps involve row replacements, which do not change the determinant value.

Thus, the determinant of the original matrix is -1 , confirming our calculations.

General Method for Determinants via Row Reduction

Here is a general method for computing determinants through row reduction:

Recipe: Computing Determinants by Row Reducing

Let A be a square matrix. Suppose you perform a series of row operations to reduce A to a row echelon form B . Then:

$$\det(A) = (-1)^r \cdot (\text{product of diagonal entries of } B) \cdot (\text{product of scaling factors}),$$

where r is the number of row swaps.

This method systematically reduces any square matrix and provides a straightforward way to compute the determinant. Row swaps, scaling, and row replacements are all factored into the calculation.

Diagonal Entries and Triangular Matrices

Definition

- The **diagonal** entries of a matrix A are the entries a_{11}, a_{22}, \dots :

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{pmatrix} \quad \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \quad \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}.$$

A square matrix is called **upper-triangular** if its nonzero entries all lie above the diagonal, and it is called **lower-triangular** if its nonzero entries all lie below the diagonal. It is called **diagonal** if all of its nonzero entries lie on the diagonal, i.e., if it is both upper-triangular and lower-triangular.

upper-triangular

lower-triangular

diagonal

$$\begin{pmatrix} * & * & * & * \\ 0 & * & * & * \\ 0 & 0 & * & * \\ 0 & 0 & 0 & * \end{pmatrix} \quad \begin{pmatrix} * & 0 & 0 & 0 \\ * & * & 0 & 0 \\ * & * & * & 0 \\ * & * & * & * \end{pmatrix} \quad \begin{pmatrix} * & 0 & 0 & 0 \\ 0 & * & 0 & 0 \\ 0 & 0 & * & 0 \\ 0 & 0 & 0 & * \end{pmatrix}.$$

Proposition.

Let A be an $n \times n$ matrix.

- If A has a zero row or column, then $\det(A) = 0$.
- If A is upper-triangular or lower-triangular, then $\det(A)$ is the product of its diagonal entries.

Example. Compute the determinants of these matrices:

$$\begin{pmatrix} 1 & 2 & 3 \\ 0 & 4 & 5 \\ 0 & 0 & 6 \end{pmatrix} \quad \begin{pmatrix} -20 & 0 & 0 \\ \pi & 0 & 0 \\ 0 & 0 & 6 \end{pmatrix} \quad \begin{pmatrix} 17 & -3 & 4 \\ 0 & 0 & 0 \\ 11/2 & 1 & 0 \end{pmatrix}.$$

Solution. The first matrix is upper-triangular, the second is lower-triangular, and the third has a zero row:

$$\begin{aligned} \det \begin{pmatrix} 1 & 2 & 3 \\ 0 & 4 & 5 \\ 0 & 0 & 6 \end{pmatrix} &= 1 \cdot 4 \cdot 6 = 24 \\ \det \begin{pmatrix} -20 & 0 & 0 \\ \pi & 0 & 0 \\ 0 & 0 & -7 \end{pmatrix} &= -20 \cdot 0 \cdot -7 = 0 \\ \det \begin{pmatrix} 17 & -3 & 4 \\ 0 & 0 & 0 \\ 11/2 & 1 & e \end{pmatrix} &= 0. \end{aligned}$$

A matrix can always be transformed into row echelon form by a series of row operations, and a matrix in row echelon form is upper-triangular. Therefore, we have completely justified the recipe for computing the determinant.

The determinant is characterized by its defining properties, since we can compute the determinant of any matrix using row reduction, as in the above recipe. However, we have not yet proved the existence of a function satisfying the defining properties! Row reducing will compute the determinant *if it exists*, but we cannot use row reduction to prove existence, because we do not yet know that you compute the same number by row reducing in two different ways.

Theorem (Existence of the determinant). *There exists one and only one function from the set of square matrices to the real numbers, that satisfies the four defining properties.*

We will prove the existence theorem in Section 4.2, by exhibiting a recursive formula for the determinant. Again, the real content of the existence theorem is:

No matter which row operations you do, you will always compute the same value for the determinant.

4.1.2 Magical Properties of the Determinant

The determinant of a square matrix possesses several elegant and powerful properties, which are summarized as follows:

1. **Existence and Uniqueness:** There exists one and only one function:

$$\det : \{n \times n \text{ matrices}\} \rightarrow \mathbb{R}$$

that satisfies the four defining properties of the determinant.

2. **Determinant of Triangular Matrices:** For an upper-triangular, lower-triangular, or diagonal matrix A , the determinant is the product of its diagonal entries:

$$\det(A) = a_{11} \cdot a_{22} \cdot \cdots \cdot a_{nn}.$$

3. **Invertibility Condition:** A square matrix A is invertible if and only if $\det(A) \neq 0$. In that case, the determinant of the inverse matrix A^{-1} is given by:

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

4. **Multiplicative Property:** For any two square matrices A and B of size $n \times n$:

$$\det(AB) = \det(A) \cdot \det(B).$$

5. **Determinant of the Transpose:** The determinant of a square matrix A is equal to the determinant of its transpose:

$$\det(A^T) = \det(A).$$

6. **Row and Column Operations:** The determinant can be computed by performing row or column operations. These operations affect the determinant as follows:

- Swapping two rows (or columns) changes the sign of the determinant.
- Multiplying a row (or column) by a scalar k multiplies the determinant by k .
- Adding a multiple of one row (or column) to another leaves the determinant unchanged.

These properties demonstrate the remarkable utility of the determinant in linear algebra. Not only does the determinant provide key insights into matrix invertibility and transformations, but it also serves as a powerful computational tool for solving systems of linear equations and analyzing geometric transformations.

4.2 Cofactor Expansions

In this subsection, we will learn about the cofactor expansions (Laplace expansion) and Cramer's rule to help us solving problems.

4.2.1 Cofactor Expansions

A recursive formula must have a starting point. For example:

$$\det(a) = a.$$

Definition

Let A be an $n \times n$ matrix. The (i, j) minor, denoted A_{ij} , is the $(n - 1) \times (n - 1)$ matrix obtained from A by deleting the i -th row and the j -th column. The (i, j) cofactor C_{ij} is defined in terms of the minor by:

$$C_{ij} = (-1)^{i+j} \det(A_{ij}).$$

Example: For $A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}$, compute A_{23} and C_{23} .

Solution: The minor A_{23} is obtained by deleting the 2nd row and 3rd column from A :

$$A_{23} = \begin{pmatrix} 1 & 2 \\ 7 & 8 \end{pmatrix}.$$

Next, the cofactor C_{23} is given by:

$$C_{23} = (-1)^{2+3} \det \begin{pmatrix} 1 & 2 \\ 7 & 8 \end{pmatrix}.$$

Calculating the determinant:

$$\det \begin{pmatrix} 1 & 2 \\ 7 & 8 \end{pmatrix} = (1)(8) - (2)(7) = 8 - 14 = -6.$$

Thus,

$$C_{23} = (-1)^5(-6) = 6.$$

Theorem (Cofactor expansion): Let A be an $n \times n$ matrix with entries a_{ij} . For any $i = 1, 2, \dots, n$, we have

$$\det(A) = \sum_{j=1}^n a_{ij}C_{ij} = a_{i1}C_{i1} + a_{i2}C_{i2} + \cdots + a_{in}C_{in}.$$

This is called cofactor expansion along the i -th row. For any $j = 1, 2, \dots, n$, we have

$$\det(A) = \sum_{i=1}^n a_{ij}C_{ij} = a_{1j}C_{1j} + a_{2j}C_{2j} + \cdots + a_{nj}C_{nj}.$$

This is called cofactor expansion along the j -th column.

Recipe: Computing the Determinant of a 3×3 Matrix

To compute the determinant of a 3×3 matrix, first draw a larger matrix with the first two columns repeated on the right. Then add the products of the downward diagonals together, and subtract the products of the upward diagonals:

$$\det \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} = \begin{matrix} a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} \\ -a_{13}a_{22}a_{31} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} \end{matrix}$$

4.2.2 Cramer's Rule and Matrix Inverses

Recall that one can compute the determinant of a 2×2 matrix using the rule:

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \Rightarrow A^{-1} = \frac{1}{\det(A)} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

We computed the cofactors of a 2×2 matrix as follows: $C_{11} = d, C_{12} = -c, C_{21} = -b, C_{22} = a$. We can rewrite the formula as:

$$A^{-1} = \frac{1}{\det(A)} \begin{pmatrix} C_{11} & C_{21} \\ C_{12} & C_{22} \end{pmatrix}.$$

It turns out that this formula generalizes to $n \times n$ matrices.

Theorem: Let A be an invertible $n \times n$ matrix, with cofactors C_{ij} . Then

$$A^{-1} = \frac{1}{\det(A)} \begin{pmatrix} C_{11} & C_{21} & \cdots & C_{n-1,1} & C_{n1} \\ C_{12} & C_{22} & \cdots & C_{n-1,2} & C_{n2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ C_{1,n-1} & C_{2,n-1} & \cdots & C_{n-1,n-1} & C_{n,n-1} \\ C_{1n} & C_{2n} & \cdots & C_{n-1,n} & C_{nn} \end{pmatrix}.$$

The matrix of cofactors is sometimes called the adjugate matrix of A , and is denoted $\text{adj}(A)$:

$$\text{adj}(A) = \begin{pmatrix} C_{11} & C_{21} & \cdots & C_{n-1,1} & C_{n1} \\ C_{12} & C_{22} & \cdots & C_{n-1,2} & C_{n2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ C_{1,n-1} & C_{2,n-1} & \cdots & C_{n-1,n-1} & C_{n,n-1} \\ C_{1n} & C_{2n} & \cdots & C_{n-1,n} & C_{nn} \end{pmatrix}.$$

Note that the (i, j) -cofactor C_{ij} goes in the (j, i) -entry of the adjugate matrix, not the (i, j) -entry. Thus, the adjugate matrix is the transpose of the cofactor matrix.

In fact, one always has:

$$A \cdot \text{adj}(A) = \text{adj}(A) \cdot A = \det(A)I_n,$$

whether or not A is invertible.

Example 1: Compute determinant of A^{-1} ?

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

The minors are:

$$A_{11} = \det \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} = -1, \quad A_{12} = \det \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = -1, \quad A_{13} = \det \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} = -1,$$

$$A_{21} = \det \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = -1, \quad A_{22} = \det \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} = -1, \quad A_{23} = \det \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} = -1,$$

$$A_{31} = \det \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} = -1, \quad A_{32} = \det \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} = -1, \quad A_{33} = \det \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} = -1.$$

The cofactors are:

$$C_{11} = -1, \quad C_{12} = 1, \quad C_{13} = -1, \quad C_{21} = 1, \quad C_{22} = -1, \quad C_{23} = -1,$$

$$C_{31} = -1, \quad C_{32} = -1, \quad C_{33} = 1.$$

Expanding along the first row, we compute the determinant:

$$\det(A) = 1 \cdot C_{11} + 0 \cdot C_{12} + 1 \cdot C_{13} = -2.$$

Therefore, the inverse is:

$$A^{-1} = \frac{1}{\det(A)} \begin{pmatrix} C_{11} & C_{21} & C_{31} \\ C_{12} & C_{22} & C_{32} \\ C_{13} & C_{23} & C_{33} \end{pmatrix} = \frac{-1}{2} \begin{pmatrix} -1 & 1 & -1 \\ 1 & -1 & -1 \\ -1 & -1 & 1 \end{pmatrix}.$$

Thus, the inverse is:

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

Theorem: Cramer's rule

Let $\mathbf{x} = (x_1, x_2, \dots, x_n)$ be the solution of $A\mathbf{x} = \mathbf{b}$, where A is an invertible $n \times n$ matrix and \mathbf{b} is a vector in \mathbb{R}^n . Let A_i be the matrix obtained from A by replacing the i -th column by \mathbf{b} . Then

$$x_i = \frac{\det(A_i)}{\det(A)}.$$

4.2.3 Summary: Methods for Computing Determinants

There are several methods to compute determinants:

- Special formulas for 2×2 and 3×3 matrices. This is best for small matrices, except when a 3×3 matrix has many zeros.
- Cofactor expansion. Most efficient when a row or column has zeros, or if the matrix has unknown entries.
- Row and column operations. Fastest for large matrices without many zeros.

Any combination of these methods can be used. Cofactor expansion is recursive, but the minors can be computed using any convenient method. Row and column operations can also simplify matrices before expanding cofactors.

Attention

\Rightarrow Remember, all methods give the same result for the determinant.

4.3 Determinants and Volumes

In this section we give a geometric interpretation of determinants, in terms of volumes.

4.3.1 Parallelograms and Parallelepipeds

The determinant computes the volume of the following kind of geometric object.

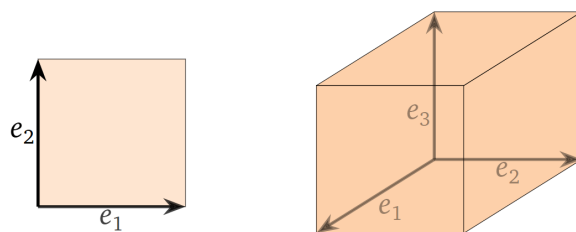
Definition

The parallelepiped determined by n vectors v_1, v_2, \dots, v_n in \mathbb{R}^n is the subset

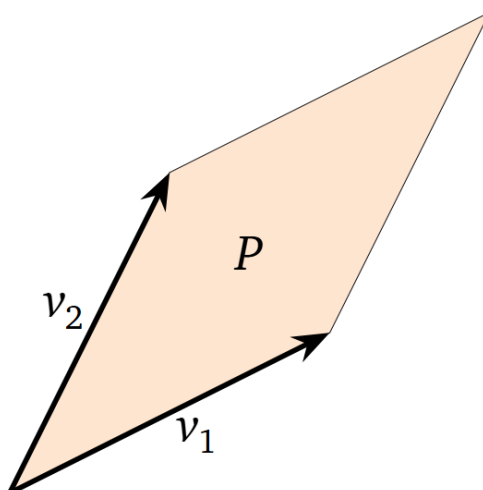
$$P = \{a_1 v_1 + a_2 v_2 + \dots + a_n v_n \mid 0 \leq a_1, a_2, \dots, a_n \leq 1\}.$$

In other words, a parallelepiped is the set of all linear combinations of n vectors with coefficients in $[0, 1]$. We can draw parallelepipeds using the parallelogram law for vector addition.

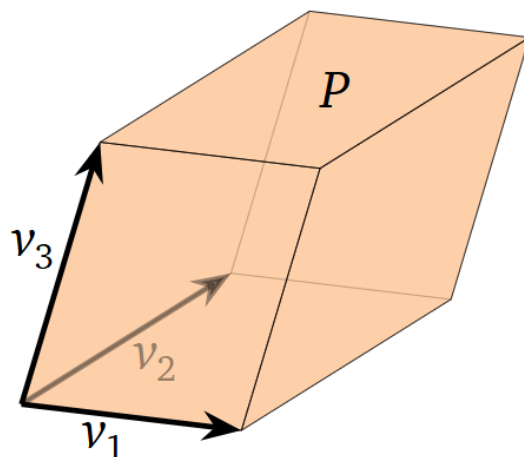
Example 1 (The unit cube): The parallelepiped determined by the standard coordinate vectors e_1, e_2, \dots, e_n is the unit n -dimensional cube.



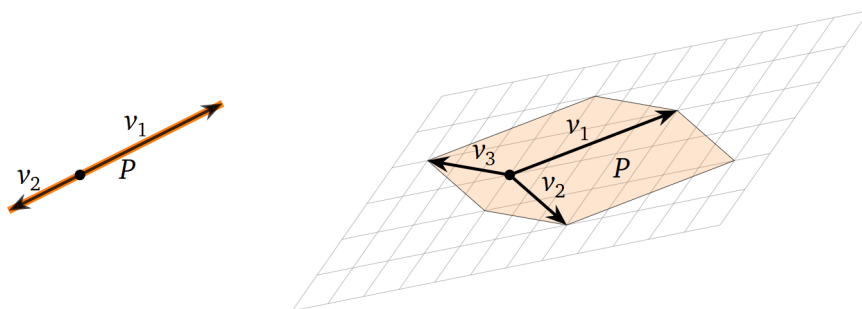
Example 2 (Parallelograms): When $n = 2$, a parallelepiped is just a parallelogram in \mathbb{R}^2 . Note that the edges come in parallel pairs.



Example 3 (Parallelepiped): When $n = 3$, a parallelepiped is a kind of skewed cube. Note that the faces come in parallel pairs.



When does a parallelepiped have zero volume? This can happen only if the parallelepiped is flat, i.e., it is squashed into a lower dimension.



This means exactly that $\{v_1, v_2, \dots, v_n\}$ is linearly dependent, which by this means that the matrix with rows v_1, v_2, \dots, v_n has determinant zero. To summarize:

Key Observation

The parallelepiped defined by v_1, v_2, \dots, v_n has zero volume if and only if the matrix with rows v_1, v_2, \dots, v_n has zero determinant.

4.3.2 Determinants and Volumes

The key observation above is only the beginning of the story: the volume of a parallelepiped is always a determinant.

Theorem(Determinants and volumes)

Let v_1, v_2, \dots, v_n be vectors in \mathbb{R}^n , let P be the parallelepiped determined by these vectors, and let A be the matrix with rows v_1, v_2, \dots, v_n . Then the absolute value of the determinant of A is the volume of P :

$$|\det(A)| = \text{vol}(P)$$

Proof: Since the properties of 4.1 characterize the determinant, they also characterize the absolute value of the determinant. Explicitly, $|\det|$ is a function on square matrices that satisfies the following properties:

- Doing a row replacement on A does not change $|\det(A)|$.
- Scaling a row of A by a scalar c multiplies $|\det(A)|$ by $|c|$.
- Swapping two rows of a matrix does not change $|\det(A)|$.
- The determinant of the identity matrix I_n is equal to 1.

The absolute value of the determinant is the only such function: indeed, by this 4.1, if you do some number of row operations on A to obtain a matrix B in row echelon form, then

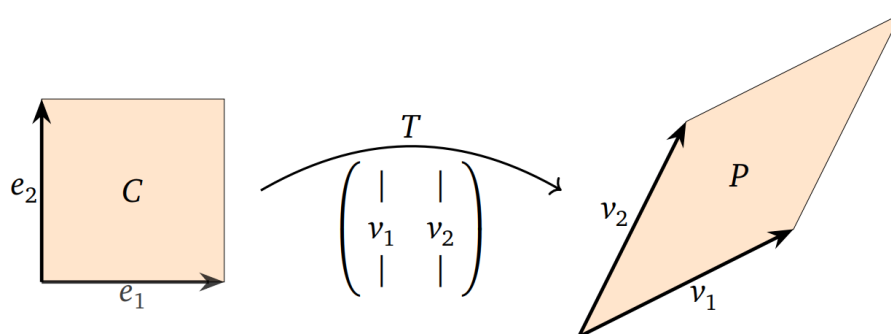
$$|\det(A)| = (\text{product of the diagonal entries of } B) \times (\text{product of the scaling factors used}).$$

For a square matrix A , we abuse notation and let $\text{vol}(A)$ denote the volume of the parallelepiped determined by the rows of A . Then we can regard vol as a function from the set of square matrices to the real numbers. We will show that vol also satisfies the above four properties.

4.3.3 Volumes of Regions

Definition

Let A be an $n \times n$ matrix with columns v_1, v_2, \dots, v_n , and let $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$ be the associated matrix transformation $T(x) = Ax$. Then $T(e_1) = v_1$ and $T(e_2) = v_2$, so T takes the unit cube C to the parallelepiped P determined by v_1, v_2, \dots, v_n .



Since the unit cube has volume 1 and its image has volume $|\det(A)|$, the transformation T scaled the volume of the cube by a factor of $|\det(A)|$. To rephrase:

If A is an $n \times n$ matrix with corresponding matrix transformation $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$, and if C is the unit cube in \mathbb{R}^n , then the volume of $T(C)$ is $|\det(A)|$.

The notation $T(S)$ means the image of the region S under the transformation T . In set builder notation, this is the subset

$$T(S) = \{T(x) \mid x \in S\}.$$

In fact, T scales the volume of *any* region in \mathbb{R}^n by the same factor, even for curvy regions.

Theorem

Let A be an $n \times n$ matrix, and let $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$ be the associated matrix transformation $T(x) = Ax$. If S is any region in \mathbb{R}^n , then

$$\text{vol}(T(S)) = |\det(A)| \cdot \text{vol}(S).$$

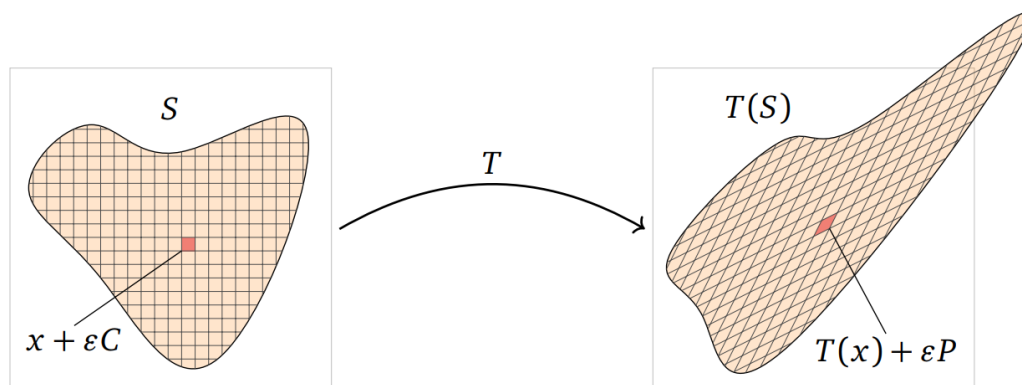
Proof. Let C be the unit cube, let v_1, v_2, \dots, v_n be the columns of A , and let P be the parallelepiped determined by these vectors, so $T(C) = P$ and $\text{vol}(P) = |\det(A)|$. For $\varepsilon > 0$, let εC be the cube with side lengths ε , i.e., the parallelepiped determined by the vectors $\varepsilon e_1, \varepsilon e_2, \dots, \varepsilon e_n$, and we define εP similarly. By the second derivative property, T takes εC to εP . The volume of εS (we've scaled each of the n standard vectors by a factor of ε) and the volume of εP is $|\det(A)| \cdot \varepsilon^n$ (for the same reason), so we have shown that T scales the volume of εC by $|\det(A)|$.

By the first *defining property*, the image of a translation of εC is simply a translation of εP :

$$T(x + \varepsilon C) = T(x) + \varepsilon T(C) = T(x) + \varepsilon P.$$

Since translations don't affect volume, this shows that T scales the volume of εC by $|\det(A)|$.

At this point, we rely on multivariable calculus for the rest of the argument. The idea is that any region S can be closely approximated by small cubes of the form $x + \varepsilon C$. The image $T(S)$ can then be approximated by the collection of images of these cubes, each of which takes the form $T(x) + \varepsilon P$.



The volume of S is roughly the sum of the volumes of these cubes. When ε approaches zero, this approximation becomes exact. Therefore, the volume of $T(S)$ can be calculated as the sum of the volumes of the parallelepipeds, taking the limit as $\varepsilon \rightarrow 0$.

The key point here is that the volume of each small cube gets scaled by $|\det(A)|$. So, the total volume of the parallelepipeds is $|\det(A)|$ times the total volume of the cubes. This leads to the conclusion:

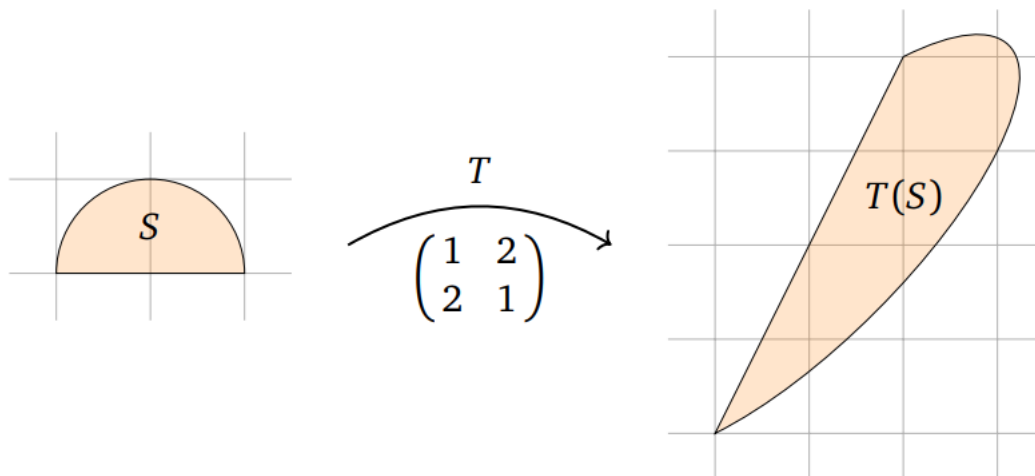
$$\text{vol}(T(S)) = |\det(A)| \cdot \text{vol}(S).$$

Example (Area of an ellipse): Find the area of the interior E of the ellipse defined by the equation

Example. Let S be a half-circle of radius 1, and let

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

We define a linear transformation $T : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ by $T(x) = Ax$. The question is: what is the area of $T(S)$?



Solution. The area of the original unit half-circle S is $\pi/2$. The transformation T scales the area by a factor of $|\det(A)|$. Since

$$\det(A) = \det \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix} = 1 \cdot 1 - 2 \cdot 2 = -3,$$

we have $|\det(A)| = 3$. Therefore, the area of $T(S)$ is:

$$\text{vol}(T(S)) = 3 \cdot \frac{\pi}{2} = \frac{3\pi}{2}.$$

Example (Area of an ellipse). Find the area of the interior E of the ellipse defined by the equation:

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

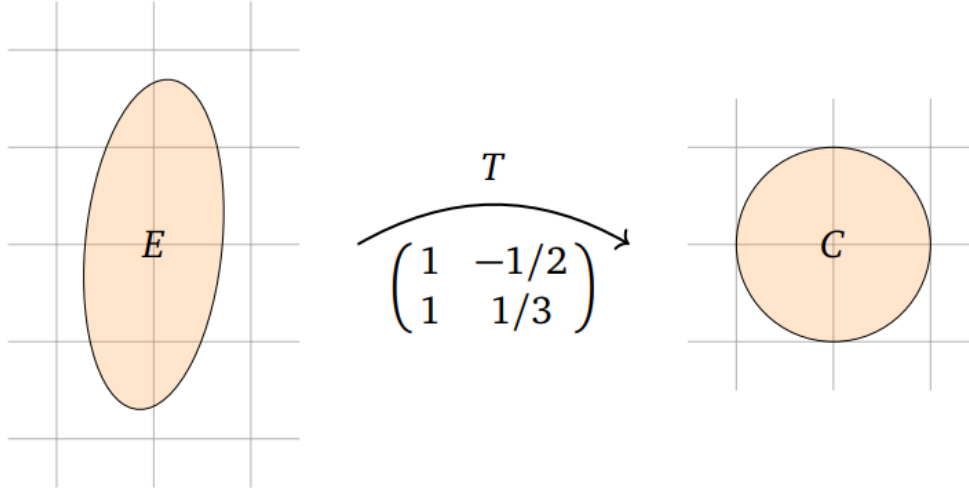
Solution. This ellipse can be obtained from the unit circle $x^2 + y^2 = 1$ by applying the following linear transformation:

$$A = \frac{2x-y}{2}, \quad B = \frac{y+3x}{3}.$$

Alternatively, in matrix form, we can express this transformation as:

$$T \left(\begin{pmatrix} A \\ B \end{pmatrix} \right) = \begin{pmatrix} \frac{2x-y}{2} \\ \frac{y+3x}{3} \end{pmatrix}.$$

This means $T(z, y)$ lies on the unit circle if and only if (z, y) lies on the ellipse E .



The transformation T scales the area by $|\det(A)|$, where A is the corresponding matrix. This gives us the area of the ellipse E as the scaled area of the unit circle.

Alternatively, in matrix form, we express the transformation T as:

$$T\left(\begin{pmatrix} x \\ y \end{pmatrix}\right) = \begin{pmatrix} \frac{2x-y}{2} \\ \frac{x+3y}{3} \end{pmatrix}.$$

This means $T(A, B)$ lies on the unit circle if and only if (A, B) lies on the ellipse E .

The transformation T scales the area by a factor of $|\det(A)|$, where A is the corresponding matrix for T . We write A explicitly as:

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

The determinant of A is:

$$\det(A) = \begin{vmatrix} 1 & -\frac{1}{2} \\ 1 & \frac{1}{3} \end{vmatrix} = \frac{5}{6}.$$

Since the area of the unit circle is π , the area of the ellipse E is scaled by $|\det(A)|$:

$$\text{Area of } E = |\det(A)| \cdot \text{Area of unit circle} = \frac{6}{5} \cdot \pi = \frac{6\pi}{5}.$$

Thus, the area of the ellipse E is:

$$\boxed{\frac{6\pi}{5}}.$$

Remark (Multiplicativity of $|\det|$). The above result highlights an important property of determinants: their multiplicative nature. Suppose A and B are $n \times n$ matrices, and $T, U : \mathbb{R}^n \rightarrow \mathbb{R}^n$ are the corresponding linear transformations. If C represents the unit cube, then the volume transformation follows this logic:

$$\text{vol}(T \circ U(C)) = \text{vol}(T(U(C))) = |\det(A)| \cdot \text{vol}(U(C)).$$

We know that $\text{vol}(U(C))$ itself scales by $|\det(B)|$, so:

$$\text{vol}(T \circ U(C)) = |\det(A)| \cdot |\det(B)| \cdot \text{vol}(C).$$

On the other hand, the matrix of the composition $T \circ U$ is precisely the product AB , which also scales the volume. Thus:

$$\text{vol}(T \circ U(C)) = |\det(AB)| \cdot \text{vol}(C).$$

Combining these results, we obtain the key property:

$$|\det(AB)| = |\det(A)| \cdot |\det(B)|.$$

This property ensures that determinants are multiplicative, reinforcing their fundamental role in linear transformations and volume scaling.

5 Eigenvalue and Eigenvector

5.1 The definitions of Eigenvalue and Eigenvector

5.1.1 Eigenvalues and Eigenvector

At this part, we will answer the question: What is **eigenvalue** and **eigenvector**?

Firstly, we need to remember that definition:

Definition

Let A be an $n \times n$ matrix:

- An **eigenvalue** of A is a *nonzero* scalar λ such that $Av = \lambda v$, for some vector v in R^n .
- An **eigenvector** of A is a vector v such that the equation $Ax = \lambda x$ has *nontrivial* solution.

If $Av = \lambda v$, $v \neq 0$, we say that λ is the **eigenvalue** for v , and v is an **eigenvector** for λ .

Note: With the equation $Av = \lambda v$

- A is an $n \times n$ matrix
- v is an $1 \times n$ matrix
- λ is a scalar

For the multiplication Av to make sense, A must be a square matrix, as it maps the vector v from space R^n back to same space R^n .

An eigenvector is always nonzero, but an eigenvalue can be zero. In the equation $A\mathbf{v} = \lambda\mathbf{v}$, λ can be zero for some nonzero \mathbf{v} . However, if $\mathbf{v} = 0$, any $\lambda \in \mathbb{R}$ would satisfy the equation, making the case trivial and thus excluded.

So, how to find the eigenvalue with eigenvector v ? The solution is so simple, just multiply the matrix A with vector v , and check that answer is a multiplication of v or not. If it's true, we can find λ easily. Let see it in example.

Example Find the eigenvalue with:

$$A = \begin{pmatrix} 2 & 2 \\ -4 & 8 \end{pmatrix}; v = \begin{pmatrix} 1 \\ 1 \end{pmatrix}; w = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$$

What are eigenvectors? What are their eigenvalues?

Solution We have:

$$Av = \begin{pmatrix} 2 & 2 \\ -4 & 8 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 4 \\ 4 \end{pmatrix} = 4v$$

Hence, v is an eigenvector of A , with eigenvalue $\lambda = 4$. On the other hand,

$$Aw = \begin{pmatrix} 2 & 2 \\ -4 & 8 \end{pmatrix} \begin{pmatrix} 2 \\ 1 \end{pmatrix} = \begin{pmatrix} 6 \\ 0 \end{pmatrix}$$

which is not a scalar multiple of w . Hence, w is not an eigenvector of A .

Geometrically, we can understand that Eigenvectors are vectors that, when transformed by A , remain on the same straight line through the origin. On the other hand, Eigenvalues are scalars that represent the degree of stretching or rotation of the Eigenvectors. This is particularly useful for certain fixed vectors, as instead of storing the complex matrix A , we only need to store the simple λ , which saves memory.

Properties

- Eigenvectors corresponding to different eigenvalues are always *linearly independent*.

Let A be an $n \times n$ matrix, $v_1, v_2, v_3, \dots, v_n$ linear combination of eigenvectors $v_1, v_2, v_3, \dots, v_n$ and coefficients $c_1, c_2, c_3, \dots, c_n$ that:

$$v_1 c_1 + v_2 c_2 + \dots + v_n c_n = 0$$

Multiply both sides by matrix A :

$$\iff A(v_1 c_1 + v_2 c_2 + \dots + v_n c_n) = 0$$

$Av_i = \lambda_i v_i$ with λ_i is the eigenvalue with eigenvector v_i

$$\iff \lambda_1 v_1 c_1 + \lambda_2 v_2 c_2 + \dots + \lambda_n v_n c_n = 0$$

Because the eigenvectors are distinct, the coefficients c_1, c_2, \dots, c_n must be all zero, proving that the eigenvectors are linearly independent.

- A matrix of size $n \times n$ can have at most n eigenvalues.

For a matrix belonging to the space \mathbb{R}^n , we can think of that matrix as being similar to a vector in the same space. The eigenvectors in this space can be seen as the axes of a coordinate system, and we can represent the matrix in a manner similar to representing a vector in a coordinate system using eigenvalues. Since the space is \mathbb{R}^n , we can have at most n eigenvalues.

5.1.2 Eigenspaces

We already know how to determine the eigenvalue from a given eigenvector. Our next goal is, given a matrix A and an eigenvalue λ , to find the eigenvector that satisfies the equation $Ax = \lambda x$.

Let's see the equation:

$$Ax = \lambda v$$

$$\iff Ax - \lambda v = 0$$

$$\iff Ax - \lambda I_n v = 0$$

$$\iff (A - \lambda I_n)v = 0$$

Therefore, the eigenvector with eigenvalue λ , if any, are the nontrivial solution of the equation $(A - \lambda I_n)v = 0$. If the equation has no nontrivial solution, λ is not an eigenvalue of A .

Definition

Let A be an $n \times n$ matrix, and let λ be an eigenvalue of A . The eigenspace of A is the solution set of $(A - \lambda I)v = 0$, i.e., the subspace $\text{Nul}(A - \lambda I)$.

Example . (Computing eigenspaces):

For each of the numbers $\lambda = 2, 1, 3$, decide if λ is an eigenvalue of the matrix

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

and if so, compute a basis for the λ -eigenspace.

Solution :

The number 3 is an eigenvalue of A if and only if $\text{Nul}(A - 3I_2)$ is nonzero. Hence, we have to solve the matrix equation $(A - 3I_2)v = 0$. We compute $A - 3I_2$:

$$A - 3I_2 = \begin{pmatrix} 2 & -4 \\ -1 & -1 \end{pmatrix} - \begin{pmatrix} 3 & 0 \\ 0 & 3 \end{pmatrix} = \begin{pmatrix} -1 & -4 \\ -1 & -4 \end{pmatrix}$$

Next, we reduce this matrix to row echelon form:

$$\begin{pmatrix} 1 & 4 \\ 0 & 0 \end{pmatrix}$$

From this, we can write the system of equations:

$$x = -4y, \quad y = y$$

In parametric vector form, we have:

$$\begin{pmatrix} x \\ y \end{pmatrix} = y \begin{pmatrix} -4 \\ 1 \end{pmatrix}$$

Since y is a free variable, the null space of $A - 3I_2$ is nonzero, so 3 is an eigenvalue. A basis for the 3-eigenspace is:

$$\left\{ \begin{pmatrix} -4 \\ 1 \end{pmatrix} \right\}$$

Concretely, we have shown that the eigenvectors of A with eigenvalue 3 are exactly the nonzero multiples of $\begin{pmatrix} -4 \\ 1 \end{pmatrix}$. In particular, $\begin{pmatrix} -4 \\ 1 \end{pmatrix}$ is an eigenvector, which we can verify:

$$A \begin{pmatrix} -4 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 & -4 \\ -1 & -1 \end{pmatrix} \begin{pmatrix} -4 \\ 1 \end{pmatrix} = \begin{pmatrix} -12 \\ 3 \end{pmatrix} = 3 \begin{pmatrix} -4 \\ 1 \end{pmatrix}$$

The number 1 is an eigenvalue of A if and only if $\text{Nul}(A - I_2)$ is nonzero. Hence, we have to solve the matrix equation $(A - I_2)v = 0$. We compute $A - I_2$:

$$A - I_2 = \begin{pmatrix} 2 & -4 \\ -1 & -1 \end{pmatrix} - \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & -4 \\ -1 & -1 \end{pmatrix}$$

The determinant of this matrix is 6, so it is invertible. By the invertible matrix theorem, we have $\text{Nul}(A - I_2) = \{0\}$, so 1 is not an eigenvalue.

The eigenvectors of A with eigenvalue 2, if any, are the nonzero solutions of the matrix equation $(A - 2I_2)v = 0$. We have:

$$A - 2I_2 = \begin{pmatrix} 2 & 4 \\ 1 & 1 \end{pmatrix} - \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} = \begin{pmatrix} 0 & 4 \\ 1 & -1 \end{pmatrix}$$

We reduce this matrix to row echelon form:

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

In parametric form, we have:

$$x = y, \quad y = y$$

In parametric vector form:

$$\begin{pmatrix} x \\ y \end{pmatrix} = y \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

Hence, there exist eigenvectors with eigenvalue 2, namely, any nonzero multiple of $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$. A basis for the 2-eigenspace is:

$$\left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\}$$

Properties

- If λ is an eigenvalue, its eigenspace contains infinitely many eigenvectors (every vector is a scalar multiple of the original eigenvector).
- The dimension of the eigenspace is equal to the number of free variables in the system of equations $(A - \lambda I_n)v = 0$. (Easily prove by the properties of eigenvector).

Recipe Let A be an $n \times n$ matrix and let λ be a number.

1. λ is an eigenvalue of A if and only if $(A - \lambda I_n)v = 0$ has a nontrivial solution, which is equivalent to $\text{Nul}(A - \lambda I_n) \neq \{0\}$.
2. In this case, finding a basis for the λ -eigenspace of A means finding a basis for $\text{Nul}(A - \lambda I_n)$. This can be done by solving $(A - \lambda I_n)v = 0$ and expressing the solutions in parametric vector form.
3. The dimension of the λ -eigenspace of A is equal to the number of free variables in the system of equations $(A - \lambda I_n)v = 0$, which corresponds to the number of columns of $A - \lambda I_n$ without pivots.
4. The eigenvectors with eigenvalue λ are the nonzero vectors in $\text{Nul}(A - \lambda I_n)$, or equivalently, the nontrivial solutions of $(A - \lambda I_n)v = 0$.

5.1.3 The Inverible Matrix Theorem: Addenda

We now have two new ways of saying that a matrix is invertible, so we add them to the invertible matrix theorem.

Invertible Matrix Theorem. Let A be an $n \times n$ matrix, and let $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$ be the matrix transformation $T(x) = Ax$. The following statements are equivalent:

1. A is invertible.
2. A has n pivots.
3. $\text{Nul}(A) = \{0\}$.
4. The columns of A are linearly independent.
5. The columns of A span \mathbb{R}^n .
6. $Ax = b$ has a unique solution for each b in \mathbb{R}^n .
7. T is invertible.
8. T is one-to-one.
9. T is onto.
10. $\det(A) \neq 0$.
11. 0 is not an eigenvalue of A .

5.2 The Characteristic Polynomial

In the previous subsection, we learned how to find the eigenvector for a given eigenvalue. Now, the question is: how do we find the eigenvalues?

Definition

Let A be an $n \times n$ matrix.

1. The **characteristic polynomial** of A is the function $f(\lambda)$ defined as:

$$f(\lambda) = \det(A - \lambda I_n)$$

2. A number λ is called an *eigenvalue* of A if $f(\lambda) = 0$.

Now, we prove it.

$$\begin{aligned} \lambda \text{ is an eigenvalue of } A &\implies Ax = \lambda x \text{ has a nontrivial solution,} \\ &\implies (A - \lambda I_n)x = 0 \text{ has a nontrivial solution.} \\ &\implies A - \lambda I_n \text{ is not invertible.} \\ &\implies \det(A - \lambda I_n) = 0. \\ &\implies f(\lambda) = 0. \end{aligned}$$

Example: Finding Eigenvalues and Eigenvectors Let

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

Solution

Characteristic Polynomial

The characteristic polynomial is:

$$f(\lambda) = \det(A - \lambda I_2) = \det \begin{pmatrix} 3 - \lambda & 2 \\ 4 & 1 - \lambda \end{pmatrix}.$$

Expanding the determinant:

$$f(\lambda) = (3 - \lambda)(1 - \lambda) - (4)(2),$$

$$f(\lambda) = \lambda^2 - 4\lambda - 5.$$

Finding Eigenvalues

To find the eigenvalues, solve $f(\lambda) = 0$:

$$\lambda^2 - 4\lambda - 5 = (\lambda - 5)(\lambda + 1) = 0.$$

Thus, the eigenvalues are:

$$\lambda_1 = 5 \quad \text{and} \quad \lambda_2 = -1.$$

Finding Eigenvectors

Solve $(A - \lambda I_2)v = 0$ for each eigenvalue.

For $\lambda_1 = 5$:

$$A - 5I_2 = \begin{pmatrix} 3 - 5 & 2 \\ 4 & 1 - 5 \end{pmatrix} = \begin{pmatrix} -2 & 2 \\ 4 & -4 \end{pmatrix}.$$

The system of equations is:

$$-2x + 2y = 0, \quad 4x - 4y = 0.$$

Simplifying:

$$x = y.$$

Thus, the eigenvector corresponding to $\lambda_1 = 5$ is:

$$v_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

For $\lambda_2 = -1$:

$$A - (-1)I_2 = A + I_2 = \begin{pmatrix} 3 + 1 & 2 \\ 4 & 1 + 1 \end{pmatrix} = \begin{pmatrix} 4 & 2 \\ 4 & 2 \end{pmatrix}.$$

The system of equations is:

$$4x + 2y = 0, \quad 4x + 2y = 0.$$

Simplifying:

$$2x + y = 0 \quad \Rightarrow \quad y = -2x.$$

Thus, the eigenvector corresponding to $\lambda_2 = -1$ is:

$$v_2 = \begin{pmatrix} 1 \\ -2 \end{pmatrix}.$$

Final Results

- **Eigenvalues:** $\lambda_1 = 5$, $\lambda_2 = -1$.
- **Eigenvectors:**

- For $\lambda_1 = 5$: $v_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$.
- For $\lambda_2 = -1$: $v_2 = \begin{pmatrix} 1 \\ -2 \end{pmatrix}$.

5.3 Similarity

Some matrices are easy to understand. For instance, a diagonal matrix

$$D = \begin{pmatrix} 2 & 0 \\ 0 & \frac{1}{2} \end{pmatrix}$$

just scales the coordinates of a vector:

$$D \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 2x \\ \frac{y}{2} \end{pmatrix}.$$

The purpose of most of the rest of this chapter is to understand complicated-looking matrices by analyzing to what extent they “behave like” simple matrices. For instance, the matrix

$$A = \begin{pmatrix} 1 & 1 \\ 6 & 9 \end{pmatrix}$$

has eigenvalues 2 and $\frac{1}{2}$, with corresponding eigenvectors

$$v_1 = \begin{pmatrix} \frac{2}{3} \\ 1 \end{pmatrix}, \quad v_2 = \begin{pmatrix} -1 \\ 1 \end{pmatrix}.$$

Notice that

$$D(xe_1 + ye_2) = xDe_1 + yDe_2 = \begin{pmatrix} 2x \\ -\frac{1}{2}y \end{pmatrix},$$

and

$$A(xv_1 + yv_2) = xAv_1 + yAv_2 = \begin{pmatrix} 2x \\ -\frac{1}{2}y \end{pmatrix}.$$

Using v_1, v_2 instead of the usual coordinates makes A “behave” like a diagonal matrix.

In this section, we study in detail the situation when two matrices behave similarly with respect to different coordinate systems.

5.3.1 Similar Matrices

Definition

Two $n \times n$ matrices A and B are *similar* if there exists an invertible $n \times n$ matrix C such that

$$A = CBC^{-1}.$$

Proposition: Let A , B , and C be $n \times n$ matrices. The following properties hold:

Proof: Taking $C = I_n = I_n^{-1}$, we have

$$A = I_n A I_n^{-1}.$$

Suppose that $A = CBC^{-1}$. Multiplying both sides on the left by C^{-1} and on the right by C gives

$$C^{-1}AC = C^{-1}(CBC^{-1})C = B.$$

Since $(C^{-1})^{-1} = C$, we have

$$B = C^{-1}A(C^{-1})^{-1},$$

so that B is similar to A .

Suppose that $A = DBD^{-1}$ and $B = ECE^{-1}$. Substituting for B and remembering that $(DE)^{-1} = E^{-1}D^{-1}$, we have

$$A = D(ECE^{-1})D^{-1} = (DE)C(DE)^{-1},$$

which shows that A is similar to C .

Fact: Let $A = CBC^{-1}$. Then, for any $n \geq 1$, we have

$$A^n = CB^nC^{-1}.$$

Proof: First, note that

$$A^2 = AA = (CBC^{-1})(CBC^{-1}) = CB(C^{-1}C)BC^{-1} = CBI_nBC^{-1} = CB^2C^{-1}.$$

Next, we have

$$A^3 = A^2A = (CB^2C^{-1})(CBC^{-1}) = CB^2(C^{-1}C)BC^{-1} = CB^3C^{-1}.$$

The pattern is clear.

5.3.2 Eigenvalues of Similar Matrices

Since similar matrices behave in the same way with respect to different coordinate systems, we should expect their eigenvalues and eigenvectors to be closely related.

Fact

Similar matrices have the same characteristic polynomial.

Proof: Suppose that $A = CBC^{-1}$, where A , B , and C are $n \times n$ matrices. We calculate

$$A - \lambda I_n = CBC^{-1} - \lambda I_n = C(B - \lambda I_n)C^{-1}.$$

Therefore,

$$\det(A - \lambda I_n) = \det(C(B - \lambda I_n)C^{-1}) = \det(C) \det(B - \lambda I_n) \det(C^{-1}) = \det(B - \lambda I_n).$$

5.4 Diagonalization

Diagonal matrices are the easiest kind of matrices to understand: they just scale the coordinate directions by their diagonal entries. In 5.3, we saw that similar matrices behave in the same way, with respect to different coordinate systems. Therefore, if a matrix is similar to a diagonal matrix, it is also relatively easy to understand. This section is devoted to the question: “When is a matrix similar to a diagonal matrix?”

Definition

An $n \times n$ matrix A is *diagonalizable* if it is similar to a diagonal matrix: that is, if there exists an invertible $n \times n$ matrix C and a diagonal matrix D such that

$$A = CDC^{-1}.$$

Powers of diagonalizable matrices Multiplying diagonal matrices together just multiplies their diagonal entries:

$$\begin{pmatrix} x_1 & 0 & 0 \\ 0 & x_2 & 0 \\ 0 & 0 & x_3 \end{pmatrix} \begin{pmatrix} y_1 & 0 & 0 \\ 0 & y_2 & 0 \\ 0 & 0 & y_3 \end{pmatrix} = \begin{pmatrix} x_1 y_1 & 0 & 0 \\ 0 & x_2 y_2 & 0 \\ 0 & 0 & x_3 y_3 \end{pmatrix}$$

Therefore, it is easy to take powers of a diagonal matrix:

$$\begin{pmatrix} x & 0 & 0 \\ 0 & y & 0 \\ 0 & 0 & z \end{pmatrix}^n = \begin{pmatrix} x^n & 0 & 0 \\ 0 & y^n & 0 \\ 0 & 0 & z^n \end{pmatrix}$$

Recipe: Compute powers of a diagonalizable matrix

If $A = CDC^{-1}$, where D is a diagonal matrix, then

$$A^n = CD^nC^{-1}.$$

Specifically, for the matrix A ,

$$A = C \begin{pmatrix} x & 0 & 0 \\ 0 & y & 0 \\ 0 & 0 & z \end{pmatrix} C^{-1} \Rightarrow A^n = C \begin{pmatrix} x^n & 0 & 0 \\ 0 & y^n & 0 \\ 0 & 0 & z^n \end{pmatrix} C^{-1}.$$

A fundamental question about a matrix is whether or not it is diagonalizable. The following is the primary criterion for diagonalizability. It shows that diagonalizability is an eigenvalue problem.

Diagonalization Theorem

An $n \times n$ matrix A is diagonalizable if and only if A has n linearly independent eigenvectors. In this case,

$$A = CDC^{-1}$$

for

$$C = \begin{pmatrix} | & | & & | \\ v_1 & v_2 & \cdots & v_n \\ | & | & & | \end{pmatrix}, \quad D = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix},$$

where v_1, v_2, \dots, v_n are linearly independent eigenvectors, and $\lambda_1, \lambda_2, \dots, \lambda_n$ are the corresponding eigenvalues, in the same order.

Non-Uniqueness of Diagonalization: We saw in the above example that changing the order of the eigenvalues and eigenvectors produces a different diagonalization of the same matrix. There are generally many different ways to diagonalize a matrix, corresponding to different orderings of the eigenvalues of that matrix. The important thing is that the eigenvalues and eigenvectors have to be listed in the same order.

For example:

$$A = C \begin{pmatrix} | & | & | \\ v_1 & v_2 & v_3 \\ | & | & | \end{pmatrix} DC^{-1} = C \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix} C^{-1} = C \begin{pmatrix} | & | & | \\ v_1 & v_2 & v_3 \\ | & | & | \end{pmatrix} DC^{-1},$$

but we can also have a different ordering of the eigenvectors and eigenvalues:

$$A = C \begin{pmatrix} | & | & | \\ v_3 & v_2 & v_1 \\ | & | & | \end{pmatrix} DC^{-1} = C \begin{pmatrix} \lambda_3 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_1 \end{pmatrix} C^{-1} = C \begin{pmatrix} | & | & | \\ v_3 & v_2 & v_1 \\ | & | & | \end{pmatrix} DC^{-1}.$$

There are other ways of finding different diagonalizations of the same matrix. For instance, you can scale one of the eigenvectors by a constant c :

$$A = C \begin{pmatrix} | & | & | \\ v_1 & v_2 & v_3 \\ | & | & | \end{pmatrix} DC^{-1} = C \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix} C^{-1} = C \begin{pmatrix} | & | & | \\ v_1 & v_2 & v_3 \\ | & | & | \end{pmatrix} DC^{-1},$$

but we can also scale the first eigenvector by a constant c :

$$A = C \begin{pmatrix} | & | & | \\ cv_1 & v_2 & v_3 \\ | & | & | \end{pmatrix} DC^{-1} = C \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix} C^{-1}.$$

You can also find a different basis entirely for an eigenspace of dimension at least 2, etc.

Example 1: Diagonalize the matrix:

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

Solution: We need to find the eigenvalues and eigenvectors of A . First, we compute the characteristic polynomial:

$$f(\lambda) = \lambda^2 - \text{Tr}(A)\lambda + \det(A) = \lambda^2 - \lambda - 2 = (\lambda + 1)(\lambda - 2).$$

Therefore, the eigenvalues are $\lambda_1 = -1$ and $\lambda_2 = 2$.

We need to compute eigenvectors for each eigenvalue. We start with $\lambda_1 = -1$:

$$(A + I_2)v = 0 \quad \Rightarrow \quad \begin{pmatrix} \frac{3}{2} & \frac{3}{2} \\ \frac{3}{2} & \frac{3}{2} \end{pmatrix} v = 0.$$

Performing row reduction (RREF):

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

The parametric form is $x = -y$, so $v_1 = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$ is an eigenvector with eigenvalue $\lambda_1 = -1$.

Now we find an eigenvector with eigenvalue $\lambda_2 = 2$:

$$(A - 2I_2)v = 0 \quad \Rightarrow \quad \begin{pmatrix} -\frac{3}{2} & \frac{3}{2} \\ \frac{3}{2} & -\frac{3}{2} \end{pmatrix} v = 0.$$

Performing row reduction (RREF):

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

The parametric form is $x = y$, so $v_2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ is an eigenvector with eigenvalue $\lambda_2 = 2$.

The eigenvectors v_1 and v_2 are linearly independent, so the diagonalization theorem says that

$$A = CDC^{-1} \quad \text{for} \quad C = \begin{pmatrix} -1 & 1 \\ 1 & 1 \end{pmatrix}, \quad D = \begin{pmatrix} -1 & 0 \\ 0 & 2 \end{pmatrix}.$$

Alternatively, if we choose 2 as our first eigenvalue, then

$$A = C_A D_A C_A^{-1} \quad \text{for} \quad C_A = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}, \quad D_A = \begin{pmatrix} 2 & 0 \\ 0 & -1 \end{pmatrix}.$$

Example 2: Diagonalize the matrix:

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

Solution: We need to find the eigenvalues and eigenvectors of A .

First, we compute the characteristic polynomial:

$$f(\lambda) = \lambda^2 - \text{Tr}(A)\lambda + \det(A) = \lambda^2 - 2\lambda = \lambda(\lambda - 2).$$

Therefore, the eigenvalues are $\lambda_1 = 0$ and $\lambda_2 = 2$.

We need to compute eigenvectors for each eigenvalue. We start with $\lambda_1 = 0$:

$$(A - 0I_2)v = 0 \quad \Rightarrow \quad \begin{pmatrix} \frac{2}{3} & -\frac{4}{3} \\ -\frac{2}{3} & \frac{4}{3} \end{pmatrix} v = 0.$$

Performing row reduction (RREF):

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

The parametric form is $x = 2y$, so $v_1 = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$ is an eigenvector with eigenvalue $\lambda_1 = 0$.

Now we find an eigenvector with eigenvalue $\lambda_2 = 2$:

$$(A - 2I_2)v = 0 \quad \Rightarrow \quad \begin{pmatrix} -\frac{4}{3} & -\frac{4}{3} \\ -\frac{3}{3} & -\frac{3}{3} \end{pmatrix} v = 0.$$

Performing row reduction (RREF or Gaussian elimination):

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

The parametric form is $x = -y$, so $v_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$ is an eigenvector with eigenvalue $\lambda_2 = 2$.

The eigenvectors v_1 and v_2 are linearly independent, so the diagonalization theorem says that

$$A = CDC^{-1} \quad \text{for} \quad C = \begin{pmatrix} 2 & 1 \\ 1 & -1 \end{pmatrix}, \quad D = \begin{pmatrix} 0 & 0 \\ 0 & 2 \end{pmatrix}.$$

Alternatively, if we choose 2 as our first eigenvalue, then

$$A = C_A D_A C_A^{-1} \quad \text{for} \quad C_A = \begin{pmatrix} 1 & 2 \\ -1 & 1 \end{pmatrix}, \quad D_A = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}.$$

Example 3: Diagonalize the matrix:

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

Solution: We need to find the eigenvalues and eigenvectors of A .

First, we compute the characteristic polynomial by expanding cofactors along the third column:

$$f(\lambda) = \det(A - \lambda I_3) = (1 - \lambda) \det \left(\begin{pmatrix} 4 & -3 & -1 \\ -2 & -1 & 1 \\ -1 & -1 & 1 \end{pmatrix} - \lambda I_2 \right) = (1 - \lambda)(\lambda^2 - 3\lambda + 2) = -(\lambda - 1)^2(\lambda - 2).$$

Therefore, the eigenvalues are $\lambda_1 = 1$ and $\lambda_2 = 2$.

We need to compute eigenvectors for each eigenvalue. We start with $\lambda_1 = 1$:

$$(A - I_3)v = 0 \quad \Rightarrow \quad \begin{pmatrix} 3 & -3 & 2 \\ -2 & -2 & 1 \\ -1 & -1 & 0 \end{pmatrix} v = 0.$$

Performing row reduction (RREF):

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

The parametric vector form is:

$$x = y, \quad z = z \quad \Rightarrow \quad v = y \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} + z \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

Hence, a basis for the 1-eigenspace is:

$$B_1 = \left\{ \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\}.$$

Now, we compute the eigenspace for $\lambda_2 = 2$:

$$(A - 2I_3)v = 0 \quad \Rightarrow \quad \begin{pmatrix} 2 & -3 & 2 \\ -2 & -3 & 1 \\ -1 & -1 & -1 \end{pmatrix} v = 0.$$

Performing row reduction (RREF):

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

The parametric form is:

$$x = 3z, \quad y = 2z, \quad \Rightarrow \quad v = z \begin{pmatrix} 3 \\ 2 \\ 1 \end{pmatrix}.$$

So, an eigenvector with eigenvalue 2 is:

$$v_3 = \begin{pmatrix} 3 \\ 2 \\ 1 \end{pmatrix}.$$

The eigenvectors v_1, v_2, v_3 are linearly independent: v_1 and v_2 form a basis for the 1-eigenspace, and v_3 is not contained in the 1-eigenspace because its eigenvalue is 2. Therefore, the diagonalization theorem says that

$$A = CDC^{-1}, \quad \text{for } C = \begin{pmatrix} 1 & 0 & 3 \\ 1 & 0 & 2 \\ 0 & 1 & 1 \end{pmatrix}, \quad D = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix}.$$

Recipe: Diagonalization

Let A be an $n \times n$ matrix. To diagonalize A :

1. Find the eigenvalues of A using the characteristic polynomial.
2. For each eigenvalue λ of A , compute a basis B_λ for the λ -eigenspace.
3. If there are fewer than n total vectors in all of the eigenspace bases B_λ , then the matrix is not diagonalizable.
4. Otherwise, the n vectors v_1, v_2, \dots, v_n in the eigenspace bases are linearly independent, and

$$A = CDC^{-1}$$

where

$$C = (v_1 \ v_2 \ \dots \ v_n) \quad \text{and} \quad D = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{pmatrix},$$

where λ_i is the eigenvalue for v_i .

5.4.1 Algebraic and Geometric Multiplicity

In algebra, the multiplicity of a root λ_0 of a polynomial $f(\lambda)$ refers to the number of times the factor $\lambda - \lambda_0$ appears in the factorization of $f(\lambda)$.

For example, consider the polynomial

$$f(\lambda) = -\lambda^3 + 4\lambda^2 - 5\lambda + 2 = -(\lambda - 1)^2(\lambda - 2),$$

where the root $\lambda_0 = 2$ has multiplicity 1, and the root $\lambda_0 = 1$ has multiplicity 2.

Definition

Let A be an $n \times n$ matrix, and let λ be an eigenvalue of A .

- The *algebraic multiplicity* of λ is its multiplicity as a root of the characteristic polynomial of A .
- The *geometric multiplicity* of λ is the dimension of the λ -eigenspace.

Since the λ -eigenspace of A is $\text{Nul}(A - \lambda I_n)$, its dimension is the number of free variables in the system of equations

$$(A - \lambda I_n)x = 0,$$

i.e., the number of columns without pivots in the matrix $A - \lambda I_n$.

Theorem (Algebraic and Geometric Multiplicity):

Let A be a square matrix and let λ be an eigenvalue of A . Then,

$$1 \leq (\text{the geometric multiplicity of } \lambda) \leq (\text{the algebraic multiplicity of } \lambda).$$

Diagonalization Theorem-Variant

Let A be an $n \times n$ matrix. The following are equivalent:

1. A is diagonalizable.
2. The sum of the geometric multiplicities of the eigenvalues of A is equal to n .
3. The sum of the algebraic multiplicities of the eigenvalues of A is equal to n , and for each eigenvalue, the geometric multiplicity equals the algebraic multiplicity.

The first part of the third statement simply says that the characteristic polynomial of A factors completely into linear polynomials over the real numbers: in other words, there are no complex (non-real) roots. The second part of the third statement says, in particular, that for any diagonalizable matrix, the algebraic and geometric multiplicities coincide.

Proof: We will show $1 \Rightarrow 2 \Rightarrow 3 \Rightarrow 1$.

First suppose that A is diagonalizable. Then A has n linearly independent eigenvectors v_1, v_2, \dots, v_n . This implies that the sum of the geometric multiplicities is at least n : for instance, if v_1, v_2, v_3 have the same eigenvalue λ , then the geometric multiplicity of λ is at least 3 (as the λ -eigenspace contains three linearly independent vectors), and so on. But the sum of the algebraic multiplicities is greater than or equal to the sum of the geometric multiplicities by the *Algebraic and Geometric Multiplicity Theorem*, and the sum of the algebraic multiplicities is at most n because the characteristic polynomial has degree n . Therefore, the sum of the geometric multiplicities equals n .

Now suppose that the sum of the geometric multiplicities equals n . As above, this forces the sum of the algebraic multiplicities to equal n as well. As the algebraic multiplicities are all greater than or equal to the geometric multiplicities in any case, this implies that they are in fact equal.

Finally, suppose that the third condition is satisfied. Then the sum of the geometric multiplicities equals n . Suppose that the distinct eigenvalues are $\lambda_1, \lambda_2, \dots, \lambda_k$, and that B_i is a basis for the λ_i -eigenspace, which we call V_i . We claim that the collection $B = \{v_1, v_2, \dots, v_n\}$ of all vectors in all of the eigenspace bases B_i is linearly independent. Consider the vector equation

$$0 = c_1 v_1 + c_2 v_2 + \dots + c_n v_n.$$

Grouping the eigenvectors with the same eigenvalues, this sum has the form

$$0 = (\text{something in } V_1) + (\text{something in } V_2) + \dots + (\text{something in } V_k).$$

Since each "something in V_i " is equal to zero. But this implies that all coefficients c_1, c_2, \dots, c_n are equal to zero, since the vectors in each B_i are linearly independent. Therefore, A has n linearly independent eigenvectors, so it is diagonalizable.

Important consequence Let A be a square matrix and let λ be an eigenvalue of A . If the algebraic multiplicity of λ does not equal the geometric multiplicity, then A is not diagonalizable.

The examples at the beginning of this subsection illustrate the theorem. Here we give some general consequences for diagonalizability of 2×2 and 3×3 matrices.

Example 1: Diagonalizability of 2×2 Matrices

Let A be a 2×2 matrix. There are four cases:

1. A has two different eigenvalues. In this case, each eigenvalue has algebraic and geometric multiplicity equal to one. This implies A is diagonalizable. For example:

$$A = \begin{pmatrix} 1 & 7 \\ 0 & 2 \end{pmatrix}.$$

2. A has one eigenvalue λ of algebraic and geometric multiplicity 2. To say that the geometric multiplicity is 2 means that $\text{Nul}(A - \lambda I_2) = \mathbb{R}^2$, i.e., that every vector in \mathbb{R}^2 is in the null space of $A - \lambda I_2$. This implies that $A - \lambda I_2$ is the zero matrix, so that A is the diagonal matrix λI_2 . In particular, A is diagonalizable. For example:

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

3. A has one eigenvalue λ of algebraic multiplicity 2 and geometric multiplicity 1. In this case, A is not diagonalizable, by part 3 of the Diagonalization Theorem. For example:

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

4. A has no eigenvalues. This happens when the characteristic polynomial has no real roots. In particular, A is not diagonalizable. For example:

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

Example 2: Diagonalizability of 3×3 Matrices

Let A be a 3×3 matrix. We can analyze the diagonalizability of A on a case-by-case basis, as in the previous cases.

1. A has three different eigenvalues. In this case, each eigenvalue has algebraic and geometric multiplicity equal to one. This implies A is diagonalizable. For example:

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

2. A has two distinct eigenvalues λ_1, λ_2 . In this case, one has algebraic multiplicity one and the other has algebraic multiplicity two; after reordering, we can assume λ_1 has multiplicity 1 and λ_2 has multiplicity 2. This implies that λ_1 has geometric multiplicity 1, so A is diagonalizable if and only if the λ_2 -eigenspace is a plane. For example:

$$A = \begin{pmatrix} 1 & 7 & 4 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix}.$$

On the other hand, if the geometric multiplicity of λ_2 is 1, then A is not diagonalizable. For example:

$$A = \begin{pmatrix} 1 & 7 & 4 \\ 0 & 2 & 0 \\ 1 & 0 & 2 \end{pmatrix}.$$

3. A has only one eigenvalue λ . If the algebraic multiplicity of λ is 1, then A is not diagonalizable. This happens when the characteristic polynomial has two complex (non-real) roots. For example:

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

Otherwise, the algebraic multiplicity of λ is equal to 3. In this case, if the geometric multiplicity is 1:

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

or 2:

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

then A is not diagonalizable. If the geometric multiplicity is 3, then $\text{Nul}(A - \lambda I_3) = \mathbb{R}^3$, so that $A - \lambda I_3$ is the zero matrix, and hence $A = \lambda I_3$. Therefore, in this case, A is necessarily diagonal, as in:

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

5.4.2 Similarity and multiplicity

Theorem

Theorem: Let A and B be similar $n \times n$ matrices, and let λ be an eigenvalue of both A and B . Then:

1. The algebraic multiplicity of λ is the same for A and B .
2. The geometric multiplicity of λ is the same for A and B .

Proof: Since A and B have the same characteristic polynomial, the multiplicity of λ as a root of the characteristic polynomial is the same for both matrices, which proves the first statement.

For the second statement, suppose that $A = CBC^{-1}$ for an invertible matrix C . Let $\{v_1, v_2, \dots, v_k\}$ be a basis of the λ -eigenspace of B . We claim that $\{Cv_1, Cv_2, \dots, Cv_k\}$ is linearly independent. Suppose that

$$c_1 Cv_1 + c_2 Cv_2 + \dots + c_k Cv_k = 0.$$

Regrouping, this means

$$C(a_1 v_1 + a_2 v_2 + \dots + a_k v_k) = 0.$$

By the invertibility of C , this implies

$$a_1 v_1 + a_2 v_2 + \dots + a_k v_k = 0.$$

Since v_1, v_2, \dots, v_k are linearly independent, we get $c_1 = c_2 = \dots = c_k = 0$, as desired.

By the previous paragraph, the dimension of the λ -eigenspace of A is greater than or equal to the dimension of the λ -eigenspace of B . By symmetry (since B is similar to A as well), the dimensions are equal, so the geometric multiplicities of λ for A and B are the same.

Final Words

This concludes the sections on linear algebra that we have gathered, studied, and compiled into this report. There might still be numerous issues and errors, and there are many topics we could not deliver in time, such as *Orthogonality* and complex numbers. However, these topics will be included in the second volume.

[Click here to access the LaTeX file](#)

Thank you very much for reading this far.

References

- [1] [Interactive Linear Algebra](#), June 3, 2019.
- [2] [Giao trình Đại số tuyến tính - Pho Duc Tai](#), 2013.