A (Very) Brief Intro to Math for ML

Author: Richard Brooks

Date: February, 2025

Version: 1.0

Contents

Chapter 1 Introduction to Linear Equations, Vectors, and Matrices

In 1949, Harvard Professor Wassily Leontief used one of the earliest computers, the Mark II, to solve a system of linear equations that modelled the U.S. economy. He divided the economy into 500 sectors, like coal, automotive, and communications, and described how each sector interacted using linear equations. Due to hardware limitations, he reduced the problem to 42 equations, which took the Mark II 56 hours to solve. This effort marked a milestone in computational applications, earning Leontief the 1973 Nobel Prize.

Leontief's work showcased how linear algebra, combined with computing, could solve large-scale problems — an idea that resonates even more today in the realm of software engineering. Linear algebra is fundamental to many modern technologies, particularly in Machine Learning (ML) and Artificial Intelligence (AI). From powering recommendation systems to optimising neural networks, linear algebra is at the core of algorithms that drive today's intelligent systems.

In ML and AI, large datasets are transformed into matrices, where linear algebra techniques like matrix factorisation, eigenvalue decomposition, and singular value decomposition (SVD) are used to extract insights and make predictions. Neural networks, the backbone of AI, rely on linear algebra to propagate data through layers and adjust weights during training. As a software engineer, mastering these concepts in linear algebra equips you to build scalable, intelligent systems capable of tackling today's most complex computational problems.

With the exponential growth in data and computing, the significance of linear algebra in software development, especially in AI and ML, continues to rise. It is a vital tool that bridges theoretical mathematics with real-world applications in tech.

1.1 Systems of Linear Equations

Linear equations and their systems form the foundation of linear algebra, a critical area in mathematics with extensive applications in software engineering. From computer graphics to machine learning algorithms, understanding how to model and solve linear systems is essential for developing efficient and effective software solutions. We begin with a definition:

Definition 1.1 (Linear Equations)

A linear equation in the variables $x_1, x_2, \ldots, x_n$ is an equation that can be written in the form

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

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

Linear equations represent straight lines, planes, and hyperplanes in various dimensions, making them useful for modeling relationships in data and algorithms.

Example 1.1 A line in two-dimensional space given by $y = mx + b$ is a linear equation. It can be rewritten to fit our standard form:

$$-mx + y = b$$

Here, $a_1 = -m$, $a_2 = 1$, and b is the constant term.

Example 1.2 The general equation of a plane in three-dimensional space is

$$ax + by + cz = d$$

where a, b, c, and d are constants. This equation is linear in the variables x, y, and z.

Definition 1.2 (Systems of Linear Equations)

A **system of linear equations** (also called a **linear system**) is a collection of one or more linear equations involving the same set of variables. ♣

Example 1.3 Consider the following system of linear equations in the variables x_1, x_2, x_3:

$$\begin{aligned} 2x_1 + 3x_2 + x_3 &= 3 \\ 7x_2 - 4x_3 &= 10 \\ x_3 &= 1 \end{aligned}$$

This system contains three equations with three unknowns.

To solve the system, we first note from the third equation that $x_3 = 1$. Substituting this into the second equation, we solve for x_2:

$$\begin{aligned} 7x_2 - 4(1) &= 10 \\ 7x_2 &= 14 \\ x_2 &= 2 \end{aligned}$$

Now, substituting $x_2 = 2$ and $x_3 = 1$ into the first equation, we solve for x_1:

$$\begin{aligned} 2x_1 + 3(2) + 1 &= 3 \\ 2x_1 &= -4 \\ x_1 &= -2 \end{aligned}$$

Thus, the solution to the system is $(x_1, x_2, x_3) = (-2, 2, 1)$.

Definition 1.3 (Solutions to a System of Linear Equations)

A **solution** of a linear system in the variables $x_1, x_2, \ldots, x_n$ is a list of numbers $(s_1, s_2, \ldots, s_n)$ that satisfies all equations of the system when substituted for the variables $x_1, x_2, \ldots, x_n$, respectively. The set of all possible solutions is called its **solution set**. Two linear systems are called **equivalent** if they have the same solution set. ♣

A system of linear equations has one of the following outcomes:

(i) **No solutions**, when the equations are inconsistent.
(ii) **Exactly one solution**, when there is a unique set of values satisfying all equations.
(iii) **Infinitely many solutions**, when there are multiple sets of values that satisfy the equations.

We say a system is **consistent** if it has either one or infinitely many solutions. We say a system is **inconsistent** if it has no solution. We formalise this later in this chapter.

Remark: Determining whether a system is consistent addresses the *existence* of solutions. If solutions exist, we may further explore the *uniqueness* of these solutions.

Representing Systems with Matrices

Matrices provide a compact and efficient way to represent and manipulate systems of linear equations, which is particularly beneficial in software applications involving large datasets.

Definition 1.4

A **matrix** is a rectangular array of numbers arranged in rows and columns. The **coefficient matrix** of a linear system contains only the coefficients of the variables, while the **augmented matrix** includes an additional column for the constants from the right-hand side of the equations. ♣

Example 1.4 For the linear system:

$$\begin{aligned} 2x_1 + 3x_2 + x_3 &= 3 \\ 7x_2 - 4x_3 &= 10 \\ x_3 &= 1 \end{aligned}$$

the coefficient matrix is:

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

and the augmented matrix is:

$$\left[\begin{array}{ccc|c} 2 & 3 & 1 & 3 \\ 0 & 7 & -4 & 10 \\ 0 & 0 & 1 & 1 \end{array}\right]$$

The vertical line between the coefficient part and the augmented part is optional and is used to visually separate the coefficients from the constants.

Definition 1.5

The **size** of a matrix is defined by the number of its rows and columns, expressed as *rows* $\times$ *columns*. For instance, the coefficient matrix above is of size 3×3, and the augmented matrix is of size 3×4. ♣

Solving Systems using Augmented Matrices

We now illustrate how to solve a system of linear equations using augmented matrix notation. This process will be formalised in the next section but here we offer an example. This method streamlines the process of solving systems by focusing on matrix manipulations.

Example 1.5 Let r_1 denote the first row, r_2 the second row, and r_3 the third row.

$$\begin{bmatrix} 2 & 3 & 1 & 3 \\ 0 & 7 & -4 & 10 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

To simplify the system, we perform *elementary row operations*. Specifically, we adjust r_1 and r_2 as follows:

$$r_1 \mapsto r_1 - r_3,$$
$$r_2 \mapsto r_2 + 4r_3,$$

which gives:

$$\begin{bmatrix} 2 & 3 & 0 & 2 \\ 0 & 7 & 0 & 14 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

Next, we scale r_2 by $\frac{1}{7}$:

$$r_2 \mapsto \frac{1}{7}r_2, \qquad \begin{bmatrix} 2 & 3 & 0 & 2 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

Now, we eliminate the 3 in the first row by performing:

$$r_1 \mapsto r_1 - 3r_2, \qquad \begin{bmatrix} 2 & 0 & 0 & -4 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

Finally, we scale r_1 by $\frac{1}{2}$ to get:

$$r_1 \mapsto \frac{1}{2}r_1, \qquad \begin{bmatrix} 1 & 0 & 0 & -2 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

We can now reinterpret the matrix as a linear system. From the augmented matrix, we find:

$$x_1 = -2,$$
$$x_2 = 2,$$
$$x_3 = 1.$$

This can be expressed compactly as:

$$(x_1, x_2, x_3) = (-2, 2, 1).$$

Thus, the solution is identical to the one we obtained through substitution.

1.2 Elementary Row Operations and Echelon Forms

In example 1.5, we performed a series of operations known as **elementary row operations**. These are fundamental transformations that simplify systems without changing their solution sets.

The three types of elementary row operations are:

- **Replacement:** Replace one row by the sum of itself and a multiple of another row.
- **Interchange:** Swap two rows.
- **Scaling:** Multiply all entries in a row by a nonzero constant.

Two matrices are called **row equivalent** if one can be transformed into the other through a sequence of elementary row operations.

Remark: Two linear systems are equivalent (i.e., they have the same solution set) if their augmented matrices are row equivalent. Performing elementary row operations on an augmented matrix does not change the solution set of the system.

As you may have noticed in the examples, the matrices resulting from these operations often have a special pattern — lots of zeros below certain entries. This isn't just a coincidence but a goal in the process. We call this the **echelon form** of a matrix. By organizing a matrix into echelon form, we:

- Make solving systems of linear equations easier.
- Systematically simplify the matrix, reducing the complexity of calculations.
- Reveal key properties about the system, like existence and number of solutions or whether certain equations are dependent.

Definition 1.6 (Echelon Forms)

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

1. All zero rows are at the bottom.
2. Each leading entry of a row is in a column to the right of the leading entry of the row above it.
3. All entries in a column below a leading entry are zeros.

In addition to echelon form, a matrix may also be in **reduced row echelon form** (RREF), which has the following properties:

4. The leading entry in each nonzero row is 1.
5. Each leading 1 is the only nonzero entry in its column.

♣

We say that an echelon matrix U is an echelon form of the matrix A if U is row equivalent to A. Similarly, we say that a reduced echelon matrix U is the reduced echelon form of the matrix A if U is row equivalent to A.

The significance of putting the augmented matrix of a linear system in echelon form is explained by the following theorem.

Theorem 1.1 (Existence Theorem)

A linear system is consistent if and only if an echelon form of the augmented matrix has no row of the form

$$\left[\begin{array}{cccc} 0 & \dots & 0 & b \end{array}\right]$$

where b is nonzero.

♡

An example of an inconsistent system if the echelon form of the augmented matrix has a row of the form $\left[\begin{array}{ccc|c} 0 & 0 & 0 & 1 \end{array}\right]$. This row indicates that the system has no solution.

Example 1.6 The augmented matrix of the linear system used as the main example in the preceding section,

$$\begin{bmatrix} 2 & 3 & 1 & 3 \\ 0 & 7 & -4 & 10 \\ 0 & 0 & 1 & 1 \end{bmatrix},$$

is already in echelon form. Since it has no row of the form mentioned in the theorem, we know immediately that this system is consistent. The leading entries are 2, 7, and 1, and all entries below them are zeros. This matrix is not in reduced row echelon form however because the leading entries are not all 1.

Recall that we performed a sequence of row operations on the preceding matrix to get

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

which is in reduced echelon form. This allowed us easily to see the solutions of this system, which is the main advantage of putting the matrix in this form. This brings us to the following important theorem.

Theorem 1.2 (Uniqueness of Reduced Echelon Form)

Each matrix is row equivalent to one and only one reduced echelon matrix ♡

A pivot position in a matrix A is a location in A that corresponds to a leading 1 in the reduced echelon form of A. A pivot column is a column of A that contains a pivot position. Notice that you can obtain a pivot by scaling the leading entry of a row to be 1. Therefore, the pivot column is the column of the leading entry.

Let A be the coefficient matrix of a linear system. The pivot columns in the matrix correspond to what we call **basic variables**. The nonpivot columns correspond to what we call **free variables**.

Example 1.7 Suppose the augmented matrix of a linear system has been reduced to the following form:

$$\begin{bmatrix} \blacksquare & * & * & * & * \\ 0 & \blacksquare & * & * & * \\ 0 & 0 & 0 & \blacksquare & * \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

where $\blacksquare$ represents any nonzero number, and $*$ represents any number (including 0). The basic variables of this system are x_1, x_2, and x_4. The only free variable is x_3.

Theorem 1.3 (Uniqueness Theorem)

If a linear system is consistent, then the solution set contains either

(i) a unique solution, when there are no free variables, or

(ii) infinitely many solutions, when there is at least one free variable. ♡

1.3 Vectors in $\mathbb{R}^2$ and $\mathbb{R}^n$

A **vector** is a mathematical object that has both **magnitude** (length) and **direction**. In two or three dimensions, a vector is often represented as an arrow pointing from one point to another. For example, the vector $\boldsymbol{v} = (x_1, x_2)$ in 2D space describes a movement from the origin $(0,0)$ to the point (x_1, x_2).

Vectors are used to represent quantities like velocity, force, and displacement, which require both magnitude and direction to fully describe them. Vectors can be added together, scaled by a number, and decomposed into components.

Matrices, on the other hand, are rectangular arrays of numbers arranged in rows and columns. They are used to represent and solve systems of linear equations, perform transformations, and encode relationships between sets of vectors.

In linear algebra, vectors are often represented as **column matrices**. For example, the vector $\boldsymbol{v} = (x_1, x_2)$ can be written as a column matrix:

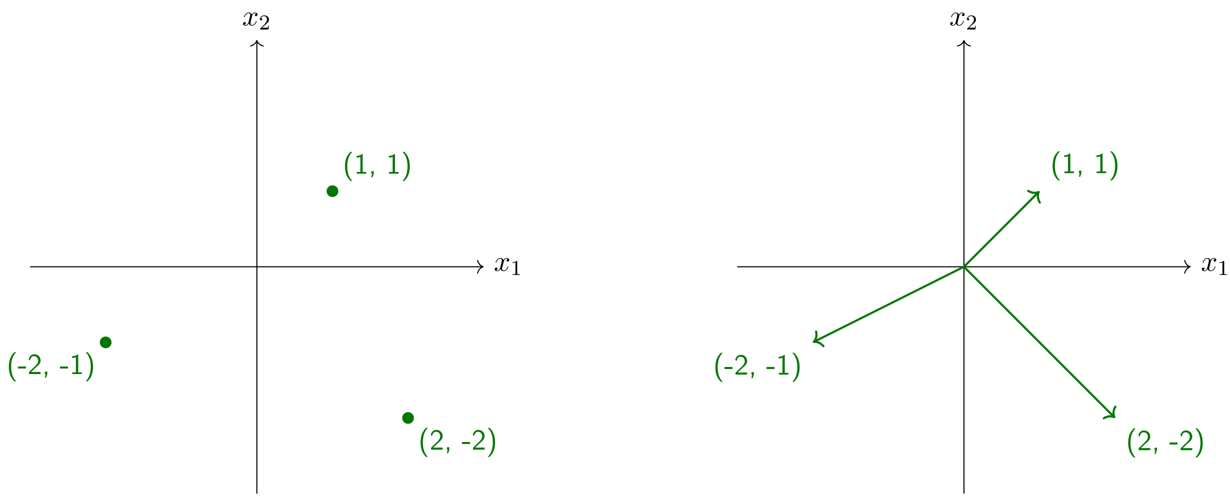
$$\boldsymbol{v} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

We say that two vectors are **equal** if and only if their corresponding entries are equal.

Example 1.8 The following are vectors in $\mathbb{R}^2$ (a.k.a. the plane consisting of ordered pairs of real numbers):

$$\boldsymbol{u} = \begin{bmatrix} 3 \\ 5 \end{bmatrix}, \quad \boldsymbol{v} = \begin{bmatrix} 5 \\ 3 \end{bmatrix}$$

They are not equal because their corresponding entries do not match.



(a) Points in the plane

(b) Vectors from the origin

Figure 1.1: Points and vectors in the plane

The sum of two vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ in $\mathbb{R}^2$, denoted $\boldsymbol{u} + \boldsymbol{v}$, is obtained by adding the corresponding entries of $\boldsymbol{u}$ and $\boldsymbol{v}$. Given a real number c, the scalar multiple of $\boldsymbol{u}$ by c, denoted $c\boldsymbol{u}$, is obtained by multiplying each entry in $\boldsymbol{u}$ by c.

Example 1.9 If $\boldsymbol{u}$ and $\boldsymbol{v}$ are as in the preceding example, then

$$\boldsymbol{u} + \boldsymbol{v} = \begin{bmatrix} 3 \\ 5 \end{bmatrix} + \begin{bmatrix} 5 \\ 3 \end{bmatrix} = \begin{bmatrix} 3+5 \\ 5+3 \end{bmatrix} = \begin{bmatrix} 8 \\ 8 \end{bmatrix}, \text{ and}$$

$$6\boldsymbol{u} = 6\begin{bmatrix} 3 \\ 5 \end{bmatrix} = \begin{bmatrix} 6 \cdot 3 \\ 6 \cdot 5 \end{bmatrix} = \begin{bmatrix} 18 \\ 30 \end{bmatrix}$$

Remark: It is often helpful to identify a vector $\begin{bmatrix} a \\ b \end{bmatrix} \in \mathbb{R}^2$ with a geometric point (a, b) in the plane in order to get a picture of what we are working with. Please see figure 1.1 for an example.

If $\boldsymbol{u}$ and $\boldsymbol{v}$ in $\mathbb{R}^2$ are thought of as points in the plane, then $\boldsymbol{u}+\boldsymbol{v}$ corresponds to the fourth vertex of the parallelogram whose other vertices are $\mathbf{0}, \boldsymbol{u}$, and $\boldsymbol{v}$. Note: By $\mathbf{0}$, we mean the zero vector, or the vector whose entries are all zero. In $\mathbb{R}^2$, we have $\mathbf{0} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$.

We call this the *parallelogram law of addition* and it can be seen in figure 1.2.

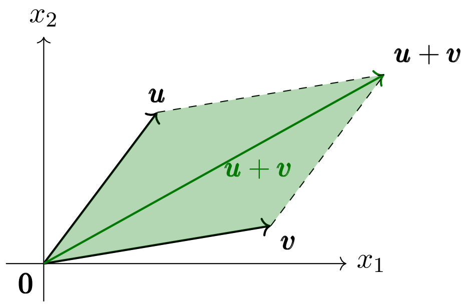


Figure 1.2: The parallelogram law of vector addition in $\mathbb{R}^2$.

These ideas generalise to higher-dimensional spaces. More specifically, we can define $\mathbb{R}^n$ as follows.

Definition 1.7

For each positive integer n, we let $\mathbb{R}^n$ denote the collection of ordered n-tuples with each entry in $\mathbb{R}$. We often write these elements as $n \times 1$ matrices. We define addition and scalar multiplication of vectors in $\mathbb{R}^n$ in the same way as we do for $\mathbb{R}^2$. That is, we go coordinate-by-coordinate. ♣

Example 1.10 If $u_1, u_2, \ldots, u_n \in \mathbb{R}$, then

$$\boldsymbol{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} \in \mathbb{R}^n.$$

Example 1.11 If $\boldsymbol{u}$ and $\boldsymbol{v}$ are in $\mathbb{R}^n$ (with entries denoted $u_1, \ldots, u_n$ and $v_1, \ldots, v_n$, respectively), and $c \in \mathbb{R}$, then

$$\boldsymbol{u}+\boldsymbol{v} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} + \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = \begin{bmatrix} u_1+v_1 \\ u_2+v_2 \\ \vdots \\ u_n+v_n \end{bmatrix}, \text{ and } \quad c\boldsymbol{u} = c\begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} = \begin{bmatrix} cu_1 \\ cu_2 \\ \vdots \\ cu_n \end{bmatrix}$$

Algebraic Properties of $\mathbb{R}^n$

Let $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w} \in \mathbb{R}^n$ and $c, d \in \mathbb{R}$. Then the following properties hold:

1. **Commutative Property of Addition:** $\boldsymbol{u} + \boldsymbol{v} = \boldsymbol{v} + \boldsymbol{u}$.
2. **Associative Property of Addition:** $(\boldsymbol{u} + \boldsymbol{v}) + \boldsymbol{w} = \boldsymbol{u} + (\boldsymbol{v} + \boldsymbol{w})$.
3. **Additive Identity:** There exists a vector $\mathbf{0} \in \mathbb{R}^n$ such that $\boldsymbol{u} + \mathbf{0} = \boldsymbol{u}$ for all $\boldsymbol{u} \in \mathbb{R}^n$.
4. **Additive Inverse:** For each $\boldsymbol{u} \in \mathbb{R}^n$, there exists a vector $-\boldsymbol{u} \in \mathbb{R}^n$ such that $\boldsymbol{u} + (-\boldsymbol{u}) = \mathbf{0}$.
5. **Distributive Property:** $c(\boldsymbol{u} + \boldsymbol{v}) = c\boldsymbol{u} + c\boldsymbol{v}$ and $(c + d)\boldsymbol{u} = c\boldsymbol{u} + d\boldsymbol{u}$.
6. **Associative Property of Scalar Multiplication:** $c(d\boldsymbol{u}) = (cd)\boldsymbol{u}$.
7. **Multiplicative Identity:** $1\boldsymbol{u} = \boldsymbol{u}$.

A **linear combination** is a way to combine vectors using scalar multiplication and addition. Given a set of vectors, we multiply each by a scalar and then sum the results. Linear combinations help us understand how vectors relate to each other and whether one vector can be expressed in terms of others.

Definition 1.8 (Linear Combinations)

Given a set of vectors $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_p \in \mathbb{R}^n$ and scalars $c_1, c_2, \ldots, c_p \in \mathbb{R}$, the vector $\boldsymbol{y}$ given by

$$\boldsymbol{y} = c_1\boldsymbol{v}_1 + \cdots + c_p\boldsymbol{v}_p$$

is called a linear combination of $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_p$ with weights $c_1, c_2, \ldots, c_p$. ♣

Example 1.12 Let $\boldsymbol{v}_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ and $\boldsymbol{v}_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$. Some linear combinations of $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ include

$$\begin{aligned} \mathbf{0} &= 0\boldsymbol{v}_1 + 0\boldsymbol{v}_2 \\ \begin{bmatrix} 3 \\ 0 \end{bmatrix} &= \boldsymbol{v}_1 + 2\boldsymbol{v}_2 \\ \begin{bmatrix} -5 \\ -1 \end{bmatrix} &= -2\boldsymbol{v}_1 - 3\boldsymbol{v}_2, \text{ and} \\ \begin{bmatrix} 5 \\ 1 \end{bmatrix} &= 2\boldsymbol{v}_1 + 3\boldsymbol{v}_2 \end{aligned}$$

Vector Equations and Linear Systems

It is often the case that we wish to know if some vector $\boldsymbol{b}$ can be formed as a linear combination of some other set of vectors $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$. The process for figuring this out is given by the following.

Using Matrices to Determine Linear Combinations

A vector equation

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

has the same solution set as the linear system whose augmented matrix is

$$\begin{bmatrix} \boldsymbol{a}_1 & \boldsymbol{a}_2 & \ldots & \boldsymbol{a}_n & \boldsymbol{b} \end{bmatrix}.$$

More specifically, if the $\boldsymbol{a}_i$'s are in $\mathbb{R}^m$ with

$$\boldsymbol{a}_1 = \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{bmatrix}, \quad \boldsymbol{a}_2 = \begin{bmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{bmatrix}, \ldots, \quad \boldsymbol{a}_n = \begin{bmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{bmatrix}, \quad \text{and} \quad \boldsymbol{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$$

then you would row reduce the matrix

$$\begin{array}{c} \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{bmatrix} \\ \begin{array}{ccccc} \uparrow & \uparrow & \cdots & \uparrow & \uparrow \\ \boldsymbol{a}_1 & \boldsymbol{a}_2 & \cdots & \boldsymbol{a}_n & \boldsymbol{b} \end{array} \end{array}$$

to determine if there is some set of weights $x_1, \ldots, x_n$ that work.

Example 1.13 Let

$$\boldsymbol{a}_1 = \begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix}, \quad \boldsymbol{a}_2 = \begin{bmatrix} 0 \\ 8 \\ -2 \end{bmatrix}, \quad \boldsymbol{a}_3 = \begin{bmatrix} 6 \\ 5 \\ 1 \end{bmatrix}, \quad \text{and } \boldsymbol{b} = \begin{bmatrix} 10 \\ 3 \\ 7 \end{bmatrix}$$

We determine if $\boldsymbol{b}$ is a linear combination of $\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3$, i.e. if there is some set of weights x_1, x_2, x_3 such that $x_1\boldsymbol{a}_1 + x_2\boldsymbol{a}_2 + x_3\boldsymbol{a}_3 = \boldsymbol{b}$. By the above, we translate this question to the matrix setting.

$$\begin{bmatrix} 2 & 0 & 6 & 10 \\ -1 & 8 & 5 & 3 \\ 1 & -2 & 1 & 7 \end{bmatrix} \quad r_1 \leftrightarrow r_3 \quad \begin{bmatrix} 1 & -2 & 1 & 7 \\ -1 & 8 & 5 & 3 \\ 2 & 0 & 6 & 10 \end{bmatrix}$$

$$\begin{array}{l} r_2 \mapsto r_2 + r_1 \\ r_3 \mapsto r_3 - 2r_1 \end{array} \quad \begin{bmatrix} 1 & -2 & 1 & 7 \\ 0 & 6 & 6 & 10 \\ 0 & 4 & 4 & -4 \end{bmatrix}$$

$$r_3 \mapsto r_3 - \frac{4}{6}r_2 \quad \begin{bmatrix} 1 & -2 & 1 & 7 \\ 0 & 6 & 6 & 10 \\ 0 & 0 & 0 & \blacksquare \end{bmatrix}.$$

Since the symbol $\blacksquare$ denotes something nonzero, we see that there's no solution, i.e. that $\boldsymbol{b}$ is not a linear combination of $\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3$.

Example 1.14 Let

$$\boldsymbol{a}_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \quad \boldsymbol{a}_2 = \begin{bmatrix} -4 \\ 6 \\ -4 \end{bmatrix}, \quad \boldsymbol{a}_3 = \begin{bmatrix} -6 \\ 7 \\ 5 \end{bmatrix}, \quad \text{and } \boldsymbol{b} = \begin{bmatrix} 11 \\ -5 \\ 9 \end{bmatrix}$$

We determine if $\boldsymbol{b}$ is a linear combination of $\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3$. We reduce the corresponding matrix to echelon form:

$$\begin{bmatrix} 1 & -4 & -6 & 11 \\ 0 & 6 & 7 & -5 \\ 1 & -4 & 5 & 9 \end{bmatrix} \quad r_3 \mapsto r_3 - r_1 \quad \begin{bmatrix} 1 & -4 & -6 & 11 \\ 0 & 6 & 7 & -5 \\ 0 & 0 & 11 & -2 \end{bmatrix}$$

and we see that there is a solution. Now we find what weights x_1, x_2, x_3 work by finding the reduced echelon form.

$$r_3 \mapsto \frac{1}{11} r_3 \qquad \begin{bmatrix} 1 & -4 & -6 & 11 \\ 0 & 6 & 7 & -5 \\ 0 & 0 & 1 & -\frac{2}{11} \end{bmatrix}$$

$$\begin{matrix} r_2 \mapsto r_2 - 7r_3 \\ r_1 \mapsto r_1 + 6r_3 \end{matrix} \qquad \begin{bmatrix} 1 & -4 & 0 & \frac{109}{11} \\ 0 & 6 & 0 & -\frac{41}{11} \\ 0 & 0 & 1 & -\frac{2}{11} \end{bmatrix}$$

$$r_2 \mapsto \frac{1}{6} r_2 \qquad \begin{bmatrix} 1 & -4 & 0 & \frac{109}{11} \\ 0 & 1 & 0 & -\frac{41}{66} \\ 0 & 0 & 1 & -\frac{2}{11} \end{bmatrix}$$

$$r_1 \mapsto r_1 + 4r_2 \qquad \begin{bmatrix} 1 & 0 & 0 & \frac{245}{38} \\ 0 & 1 & 0 & -\frac{41}{66} \\ 0 & 0 & 1 & -\frac{2}{11} \end{bmatrix},$$

so $(x_1, x_2, x_3) = \left(\frac{245}{33}, -\frac{41}{66}, -\frac{2}{11}\right)$. Since there are no free variables, this is the unique solution.

The **span** of a set of vectors is the collection of all possible linear combinations of those vectors. In other words, it's the set of all vectors you can reach by scaling and adding the given vectors. The span gives us insight into the "space" those vectors cover:

Definition 1.9 (Span of a Set of Vectors)

If $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_p$ are in $\mathbb{R}^n$, then the set of all linear combinations of $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_p$ is denoted by Span $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_p\}$ and is called the subset of $\mathbb{R}^n$ spanned by $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_p$. In other words, the span of $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_p$ is all vectors that can be written in the form

$$c_1\boldsymbol{v}_1 + \cdots + c_p\boldsymbol{v}_p$$

with $c_1, \ldots, c_p$ scalars.

♣

1.4 Matrix Equations

In the previous sections, we explored different ways of representing linear relationships. For instance, we looked at individual linear equations, such as

$$2x_1 + 3x_2 = 5,$$

and systems of linear equations, like

$$x_1 + 2x_2 = 4$$
$$3x_1 - x_2 = 2.$$

We also expressed these systems as vector equations, such as

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

Now, it's time to introduce a powerful new form: the **matrix equation**. This compact representation allows us to handle systems of linear equations efficiently using matrix notation. In this form, our system becomes

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

where A is a matrix, $\boldsymbol{x}$ is a vector of variables, and $\boldsymbol{b}$ is the result vector. Matrix equations give us a structured, algebraic approach to solving systems.

Definition 1.10 (Matrix Equation)

If A is an $m \times n$ matrix with columns $\boldsymbol{a}_1, \dots, \boldsymbol{a}_n$, and if $\boldsymbol{x} \in \mathbb{R}^n$, then the product of A and $\boldsymbol{x}$, denoted by $A\boldsymbol{x}$, is

$$A\boldsymbol{x} = \begin{bmatrix} \boldsymbol{a}_1 & \boldsymbol{a}_2 & \dots & \boldsymbol{a}_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = x_1\boldsymbol{a}_1 + x_2\boldsymbol{a}_2 + \cdots + x_n\boldsymbol{a}_n$$

♣

Example 1.15

$$\begin{aligned} \begin{bmatrix} 4 & 1 & 2 \\ 8 & 0 & 3 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \\ 4 \end{bmatrix} &= 2 \begin{bmatrix} 4 \\ 8 \end{bmatrix} + 1 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + 4 \begin{bmatrix} 2 \\ 3 \end{bmatrix} \\ &= \begin{bmatrix} 8 \\ 16 \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \begin{bmatrix} 8 \\ 12 \end{bmatrix} \\ &= \begin{bmatrix} 17 \\ 28 \end{bmatrix} \end{aligned}$$

Building a bit on the main result from the preceding section, we have the following theorem.

Theorem 1.4

If A is an $m \times n$ matrix, with columns $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n \in \mathbb{R}^m$ and if $\boldsymbol{b} \in \mathbb{R}^m$, the matrix equation

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

has the same solution set as the vector equation

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

which, in turn, has the same solution set as the system of linear equations with augmented matrix

$$\begin{bmatrix} \boldsymbol{a}_1 & \boldsymbol{a}_2 & \ldots & \boldsymbol{a}_n & \boldsymbol{b} \end{bmatrix}.$$

♡

This theorem is essential because it ties together the three forms of representing and solving systems of linear equations: matrix equations, vector equations, and systems of equations with augmented matrices. It shows that no matter which form we use, they all lead to the same solution set.

By stating that the matrix equation $A\boldsymbol{x} = \boldsymbol{b}$ is equivalent to the vector equation $x_1\boldsymbol{a}_1+x_2\boldsymbol{a}_2+\cdots+x_n\boldsymbol{a}_n = \boldsymbol{b}$, we can interpret solving a matrix equation as finding the right combination of the columns of A that yields $\boldsymbol{b}$. This connection allows us to visualize the problem in terms of vector spaces and linear combinations.

Moreover, the theorem shows that this same process is reflected in the augmented matrix, where row reduction reveals the solution through elementary row operations. So whether we approach the problem algebraically, geometrically, or algorithmically, we are dealing with the same underlying structure. This unifying perspective simplifies our approach to solving systems and provides flexibility in how we choose to represent and manipulate the problem.

Remark: The equation $A\boldsymbol{x} = \boldsymbol{b}$ has a solution if and only if $\boldsymbol{b}$ is a linear combination of the columns of A.

The following theorem tells us when the column vectors of an $m \times n$ matrix (the columns are vectors in $\mathbb{R}^m$, since there are m rows) can be used to generate all of $\mathbb{R}^m$. This basically summarises several things we have already seen in different contexts.

Theorem 1.5

Let A be an $m \times n$ matrix. Then the following statements are equivalent:

(a) For each $\boldsymbol{b} \in \mathbb{R}^m$, the equation $A\boldsymbol{x} = \boldsymbol{b}$ has a solution.
(b) Each $\boldsymbol{b} \in \mathbb{R}^m$ is a linear combination of the columns of A.
(c) The columns of A span $\mathbb{R}^m$.
(d) The matrix A has a pivot position in every row.

♡

This theorem shows that several seemingly different ideas are in fact equivalent. The existence of a solution to the matrix equation $A\boldsymbol{x} = \boldsymbol{b}$ (statement (a)) depends on whether the columns of A can form any vector in $\mathbb{R}^m$ (statement (b)). Geometrically, this means the columns span the entire space $\mathbb{R}^m$ (statement (c)).

Finally, the presence of a pivot position in every row (statement (d)) gives an algebraic condition that guarantees the span of the columns covers $\mathbb{R}^m$, ensuring that there is always a solution to $A\boldsymbol{x} = \boldsymbol{b}$.

This equivalence is a powerful tool, as it connects solutions, linear combinations, and geometric interpretations, while also providing a concrete method (checking for pivot positions) to verify these properties.

Row-Vector Rule for Computing $A\boldsymbol{x}$

Assuming the product $A\boldsymbol{x}$ is defined, the i-th entry in $A\boldsymbol{x}$ is the sum of the products of corresponding entries from row i of A and the vector $\boldsymbol{x}$. In other words, the i-th entry of $A\boldsymbol{x}$ is the dot product of the vector forming the i-th row of A and the vector $\boldsymbol{x}$.

This rule simplifies the matrix multiplication process by breaking it down into smaller, familiar operations — dot products — making it easier to compute and understand. It also highlights the connection between matrix multiplication and linear combinations, as the product $A\boldsymbol{x}$ is a linear combination of the rows of A with weights given by the entries of $\boldsymbol{x}$.

Example 1.16 Let

$$A = \begin{bmatrix} 1 & -1 & 4 \\ 5 & 0 & 2 \end{bmatrix}, \text{ and } \boldsymbol{x} = \begin{bmatrix} 2 \\ 1 \\ 5 \end{bmatrix}$$

Then

$$\begin{aligned} \begin{bmatrix} 1 & -1 & 4 \\ 5 & 0 & 2 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \\ 5 \end{bmatrix} &= \begin{bmatrix} 1 \cdot 2 + (-1) \cdot 1 + 4 \cdot 5 \\ 5 \cdot 2 + 0 \cdot 1 + 2 \cdot 5 \end{bmatrix} \\ &= \begin{bmatrix} 21 \\ 20 \end{bmatrix} \end{aligned}$$

Notice that the number of columns in A must match the number of rows in $\boldsymbol{x}$, for otherwise the dot product would not make sense!

The next theorem captures two important properties of matrix-vector multiplication: **distributivity** and **scalar multiplication**. These properties mirror the familiar rules of algebra but now apply in the context of matrices and vectors.

Theorem 1.6

If A is an $m \times n$ matrix, $\boldsymbol{u}$ and $\boldsymbol{v}$ are vectors in $\mathbb{R}^n$, and c is a scalar, then:

(a) $A(\boldsymbol{u} + \boldsymbol{v}) = A\boldsymbol{u} + A\boldsymbol{v}$

(b) $A(c\boldsymbol{u}) = c(A\boldsymbol{u})$

♡

Property (a) shows that matrix multiplication distributes over vector addition. It means that multiplying A by the sum of two vectors is the same as multiplying A by each vector separately and then adding the results. Property (b) illustrates that scalar multiplication commutes with matrix multiplication. Multiplying a vector by a scalar first, then applying the matrix, gives the same result as applying the matrix first and then multiplying the resulting vector by the scalar.

1.5 Linear Independence

Now we introduce one of the most important concepts in linear algebra: *linear independence*. This concept is fundamental to understanding the structure of vectors and their relationships in vector spaces.

Definition 1.11

An indexed set of vectors $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_p\} \subset \mathbb{R}^n$ is said to be **linearly independent** if the vector equation

$$x_1\boldsymbol{v}_1 + \cdots + x_p\boldsymbol{v}_p = \mathbf{0}$$

has only the trivial solution, i.e., if the only solution is $(x_1, \ldots, x_p) = (0, \ldots, 0)$. Likewise, the set $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_p\}$ is said to be **linearly dependent** if there exist weights $c_1, \ldots, c_p$, not all zero, such that

$$c_1\boldsymbol{v}_1 + \cdots + c_p\boldsymbol{v}_p = \mathbf{0}$$

We call such an equation a *linear dependence relation* when the weights are not all zero. ♣

Remark: A set of vectors cannot be both linearly independent and linearly dependent, but it must be one of them!

Determining if a set of vectors is linearly independent is tantamount to solving the matrix equation

$$A\boldsymbol{x} = 0,$$

where the columns of A are given by the vectors. The set of vectors is linearly independent if and only if the only solution is $\boldsymbol{x} = \mathbf{0}$. Otherwise, there is some linear dependence relation.

Example 1.17 Determine if the following vectors in $\mathbb{R}^3$ are linearly independent:

$$\boldsymbol{v}_1 = \begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix}, \quad \boldsymbol{v}_2 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \quad \boldsymbol{v}_3 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$$

Solution: We begin by reducing the corresponding augmented matrix to echelon form:

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

from which we see that there is a solution. Now we want to know if the solution is the trivial solution $(0, 0, 0)$. Normally, we would continue row operations until we reach reduced echelon form, but we can be smarter about this. Notice first of all that there are no bad rows (so there is at least one solution), which we expect since a homogeneous system always has at least the trivial solution. Since there are no free variables, we see there is exactly one solution. This tells us that the solution must be

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Hence the set of vectors $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3\}$ is linearly independent. ◀

In general, it is very useful to determine if some set of vectors is linearly independent, so it is good to have some theorems to handle this problem quickly in certain special cases.

Theorem 1.7

(a) If a set of vectors contains the zero vector, then the set is linearly dependent.

(b) If a set of vectors contains a scalar multiple of another vector, then the set is linearly dependent.

(c) If a set of vectors contains more vectors than there are entries in each vector, then the set is linearly dependent.

♡

The last part of the theorem is particularly useful, as it allows us to quickly determine if a set of vectors is linearly dependent by checking if one of the vectors is a linear combination of the others. We can formalise this idea in the following corollary.

Corollary 1.1

An indexed set $S = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_p\}$ of two or more vectors is linearly dependent if and only if at least one of the vectors in S is a linear combination of the others. If S is linearly dependent and $\boldsymbol{v}_1 \neq \mathbf{0}$, then some $\boldsymbol{v}_j$ (with $1 < j \leq p$) is a linear combination of the preceding vectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{j-1}$.

♡

Remark: The corollary tells us that if a set of vectors is linearly dependent, then at least one of the vectors is redundant, as it can be expressed as a linear combination of the others. This redundancy is what causes the linear dependence. This does not mean that every vector in S is a linear combination of the others, but only that *at least one* vector in a linearly dependent set is a linear combination of others.

Chapter 2 Matrix Algebra

This chapter covers the basic operations and properties of matrices, including addition, scalar multiplication, and matrix multiplication. These operations are fundamental to linear algebra and are used in many applications, including solving systems of linear equations and finding inverses of matrices.

2.1 Matrix Arithmetic

Matrix arithmetic like adding, subtracting, scaling, and multiplying let you combine and manipulate matrices in simple ways. These operations are the building blocks for more complex matrix manipulations, like solving systems of equations and finding inverses.

Addition and Scalar Multiplication

We begin with the definition of matrix addition.

Definition 2.1 (Matrix Addition)

Let A and B be $m \times n$ matrices. The **sum** of A and B, denoted $A+B$, is the $m \times n$ matrix whose entries are obtained by adding the corresponding entries of A and B.
Given matrices

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}, \quad B = \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ b_{m1} & b_{m2} & \cdots & b_{mn} \end{bmatrix}$$

both of the same dimension $m \times n$, the sum $A+B$ is thus defined as

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

♣

Remark: Similar considerations apply for substraction of matrices, although not mentioned here explicitly.

Next is the definition of scalar-matrix multiplication.

Definition 2.2 (Scalar-Matrix Multiplication)

Let A be an $m \times n$ matrix and c be a scalar. The **product** of c and A, denoted cA, is the $m \times n$ matrix whose entries are obtained by multiplying each entry of A by c, and is thus defined as

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

♣

Example 2.1 Given A and B below, find $3A - 2B$.

$$A = \begin{bmatrix} 1 & -2 & 5 \\ 0 & -3 & 9 \\ 4 & -6 & 7 \end{bmatrix}, B = \begin{bmatrix} 5 & 0 & -11 \\ 3 & -5 & 1 \\ -1 & -9 & 0 \end{bmatrix}$$

Solution: We compute:

$$\begin{aligned} 3A - 2B &= \begin{bmatrix} 3 & -6 & 15 \\ 0 & -9 & 27 \\ 12 & -18 & 21 \end{bmatrix} - \begin{bmatrix} 10 & 0 & -22 \\ 6 & -10 & 2 \\ -2 & -18 & 0 \end{bmatrix} \\ &= \begin{bmatrix} -7 & -6 & 37 \\ -6 & 1 & 25 \\ 14 & 0 & 21 \end{bmatrix} \end{aligned}$$

◀

Before moving on to matrix multiplication, we need to state some basic algebraic properties of matrix addition and scalar multiplication.

Theorem 2.1

Let A, B, C be matrices of the same size and let α, β be scalars. Then

(a) $A + B = B + A$
(b) $(A + B) + C = A + (B + C)$
(c) $A + 0 = A$
(d) $\alpha(A + B) = \alpha A + \alpha B$
(e) $(\alpha + \beta)A = \alpha A + \beta A$
(f) $\alpha(\beta A) = (\alpha\beta)A$

♡

Matrix Multiplication

Matrix multiplication is a bit more complex than addition and scalar multiplication. The product of two matrices is defined only when the number of columns in the first matrix is equal to the number of rows in the second matrix. The product of two matrices A and B is a new matrix C whose entries are determined by the dot product of the rows of A and the columns of B. Note that $\boldsymbol{x}$ is a column vector.

Let B be an $n \times p$ matrix and A be an $m \times n$ matrix. If $\boldsymbol{x} \in \mathbb{R}^p$, then multiplying B by $\boldsymbol{x}$ produces a new vector $B\boldsymbol{x}$ in $\mathbb{R}^n$. Once we have this result, we can further multiply it by A, giving us $A(B\boldsymbol{x})$, which is a vector in $\mathbb{R}^m$.

Thus, for any vector $\boldsymbol{x}$ in $\mathbb{R}^p$, this process produces a corresponding vector in $\mathbb{R}^m$. This two-step operation—first multiplying by B and then by A—is referred to as the composition of A and B, and is usually written as AB. Therefore, we have:

$$(AB)\boldsymbol{x} = A(B\boldsymbol{x})$$

To compute the matrix resulting from this composition, we multiply A by B directly, following the rules of matrix multiplication. The result is a matrix $C = AB$, where each entry of C is determined by the interactions between the rows of A and the columns of B.

In essence, C represents the matrix that captures the combined effect of applying both B and A to any vector $\boldsymbol{x} \in \mathbb{R}^p$, without needing to break it down into intermediate steps.

Definition 2.3

For $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{n\times p}$, with $B = \begin{bmatrix} \boldsymbol{b_1} & \boldsymbol{b_2} \cdots & \boldsymbol{b_p} \end{bmatrix}$, we define the product AB by the formula

$$AB = \begin{bmatrix} A\boldsymbol{b_1} & A\boldsymbol{b_2} & \cdots & A\boldsymbol{b_p} \end{bmatrix}$$

♣

The product AB is defined only when the number of columns of A equals the number of rows of B. The following diagram is useful for remembering this:

$$(m \times n) \cdot (n \times p) \to m \times p$$

The diagram shows that the number of columns in the first matrix must equal the number of rows in the second matrix (the blue arrow). The result is a new matrix with the number of rows from the first matrix and the number of columns from the second matrix (the red arrow).

Example 2.2 For A and B below compute AB and BA.

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

Solution: First $AB = \begin{bmatrix} A\boldsymbol{b}_1 & A\boldsymbol{b}_2 & A\boldsymbol{b}_3 & A\boldsymbol{b}_4 \end{bmatrix}$:

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

$$= \begin{bmatrix} \boxed{2} & \text{Row 1 of } A \text{ and Column 1 of } B \\ \boxed{7} & \text{Row 2 of } A \text{ and Column 1 of } B \end{bmatrix}$$

$$= \begin{bmatrix} 2 & \boxed{0} & \text{Row 1 of } A \text{ and Column 2 of } B \\ 7 & \boxed{9} & \text{Row 2 of } A \text{ and Column 2 of } B \end{bmatrix}$$

$$= \begin{bmatrix} 2 & 0 & \boxed{4} & \text{Row 1 of } A \text{ and Column 3 of } B \\ 7 & 9 & \boxed{10} & \text{Row 2 of } A \text{ and Column 3 of } B \end{bmatrix}$$

$$= \begin{bmatrix} 2 & 0 & 4 & \boxed{4} & \text{Row 1 of } A \text{ and Column 4 of } B \\ 7 & 9 & 10 & \boxed{2} & \text{Row 2 of } A \text{ and Column 4 of } B \end{bmatrix}$$

$$= \begin{bmatrix} 2 & 0 & 4 & 4 \\ 7 & 9 & 10 & 2 \end{bmatrix}$$

On the other hand, BA is not defined! B has 4 columns and A has 2 rows. Thus, the number of columns in B is not equal to the number of rows in A. ◀

Example 2.3 Example Compute the matrix AB, where

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

Solution: By the definition of multiplication given above,

$$AB = \left[A\begin{bmatrix} 2 \\ 3 \end{bmatrix} \; A\begin{bmatrix} -1 \\ 1 \end{bmatrix} \right]$$

$$= \left[\begin{bmatrix} 1 & 2 \\ 2 & 3 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \end{bmatrix} \; \begin{bmatrix} 1 & 2 \\ 2 & 3 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} -1 \\ 1 \end{bmatrix} \right]$$

$$= \left[\begin{bmatrix} 1\cdot 2 + 2\cdot 3 \\ 2\cdot 2 + 3\cdot 3 \\ 3\cdot 2 + 4\cdot 3 \end{bmatrix} \begin{bmatrix} 1\cdot(-1) + 2\cdot 1 \\ 2\cdot(-1) + 3\cdot 1 \\ 3\cdot(-1) + 4\cdot 1 \end{bmatrix} \right]$$

$$= \begin{bmatrix} 8 & 1 \\ 13 & 1 \\ 18 & 1 \end{bmatrix}$$

◀

Example 2.4 If A is 3×5 and B is 5×2, what are the sizes of AB and BA (assuming they are defined)?

Solution:

- The product AB is defined since the number of columns of A matches the number of rows of B (5). The resulting matrix AB is a 3×2 matrix.
- The product BA is not defined, since the number of columns of B (2) does not match the number of rows of A (3).

◀

The next example illustrate that even if both AB and BA are defined, they are not necessarily equal.

Example 2.5 For $A = \begin{bmatrix} -4 & 4 & 3 \\ 3 & -3 & -1 \\ -2 & -1 & 1 \end{bmatrix}$ and $B = \begin{bmatrix} -1 & -1 & 0 \\ -3 & 0 & -2 \\ -2 & 1 & -2 \end{bmatrix}$ compute AB and BA.

Solution: First AB:

$$\begin{aligned} AB &= \begin{bmatrix} -4 & 4 & 3 \\ 3 & -3 & -1 \\ -2 & -1 & 1 \end{bmatrix} \begin{bmatrix} -1 & -1 & 0 \\ -3 & 0 & -2 \\ -2 & 1 & -2 \end{bmatrix} \\ &= \left[\begin{matrix} -14 \\ 8 \\ 3 \end{matrix}\right. \\ &= \left[\begin{matrix} -14 & 7 \\ 8 & -4 \\ 3 & 3 \end{matrix}\right. \\ &= \begin{bmatrix} -14 & 7 & -14 \\ 8 & -4 & 8 \\ 3 & 3 & 0 \end{bmatrix} \end{aligned}$$

Next BA:

$$\begin{aligned} BA &= \begin{bmatrix} -1 & -1 & 0 \\ -3 & 0 & -2 \\ -2 & 1 & -2 \end{bmatrix} \begin{bmatrix} -4 & 4 & 3 \\ 3 & -3 & -1 \\ -2 & -1 & 1 \end{bmatrix} \\ &= \left[\begin{matrix} 1 \\ 16 \\ 15 \end{matrix}\right. \\ &= \left[\begin{matrix} 1 & -1 \\ 16 & -10 \\ 15 & -9 \end{matrix}\right. \\ &= \begin{bmatrix} 1 & -1 & -2 \\ 16 & -10 & -11 \\ 15 & -9 & -9 \end{bmatrix} \end{aligned}$$

We see that $AB \neq BA$. ◀

In regular arithmetic the multiplicative identity is 1. In matrix algebra, the multiplicative identity is the identity matrix, denoted by I. The identity matrix is a *square* matrix with 1s on the diagonal and 0s elsewhere. The size of the identity matrix is determined by the context, and is usually clear from the context. For example, I_2 is a 2×2 identity matrix, and I_3 is a 3×3 identity matrix and in general $I_n \in \mathbb{R}^{n \times n}$ is an $n \times n$ identity matrix:

$$I_n = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}$$

We can now state the following theorem.

Theorem 2.2

Let A, B, C be matrices, of appropriate dimensions, and let α be a scalar. Then

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

(b) $A(B + C) = AB + AC$

(c) $(B + C)A = BA + CA$

(d) $\alpha(AB) = (\alpha A)B = A(\alpha B)$

(e) $I_n A = A I_n = A$

♡

We conclude this section by looking at the kth power of a matrix.

Definition 2.4

Let A be a square matrix, i.e. $A \in \mathbb{R}^{n \times n}$. The kth power of A, denoted A^k, is defined as the product of A with itself k times. That is,

$$A^k = \underbrace{AAA \cdot \cdots \cdot A}_{k \text{ times}}$$

where A appears k times on the right-hand side.

♣

Example 2.6 Compute A^3 if

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

Solution:

Compute A^2 :

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

And then A^3 :

$$\begin{aligned} A^3 = A^2 A &= \begin{bmatrix} 7 & -6 \\ -2 & 3 \end{bmatrix} \begin{bmatrix} -2 & 3 \\ 1 & 0 \end{bmatrix} \\ &= \begin{bmatrix} -20 & 21 \\ 7 & -6 \end{bmatrix} \end{aligned}$$

We could also do:

$$A^3 = AA^2 = \begin{bmatrix} -2 & 3 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 7 & -6 \\ -2 & 3 \end{bmatrix} = \begin{bmatrix} -20 & 21 \\ 7 & -6 \end{bmatrix}$$

◀

2.2 Matrix Transpose

We begin with the definition of the transpose of a matrix.

Definition 2.5

Given a matrix $A \in \mathbb{R}^{m \times n}$, the transpose of A is the matrix A^T whose ith column is the ith row of A. ♣

If A is $m \times n$ then A^T is $n \times m$. For example, if

$$A = \begin{bmatrix} 0 & -1 & 8 & -7 & -4 \\ -4 & 6 & -10 & -9 & 6 \\ 9 & 5 & -2 & -3 & 5 \\ -8 & 8 & 4 & 7 & 7 \end{bmatrix}$$

then

$$A^T = \begin{bmatrix} 0 & -4 & 9 & -8 \\ -1 & 6 & 5 & 8 \\ 8 & -10 & -2 & 4 \\ -7 & -9 & -3 & 7 \\ -4 & 6 & 5 & 7 \end{bmatrix}$$

Example 2.7 Compute $(AB)^T$ and $B^T A^T$ if

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

Solution: First, compute (AB):

$$\begin{aligned} AB &= \begin{bmatrix} -2 & 1 & 0 \\ 3 & -1 & -3 \end{bmatrix} \begin{bmatrix} -2 & 1 & 2 \\ -1 & -2 & 0 \\ 0 & 0 & -1 \end{bmatrix} \\ &= \begin{bmatrix} 3 & -4 & -4 \\ -5 & 5 & 9 \end{bmatrix} \end{aligned}$$

and then $(AB)^T$:

$$\begin{aligned} (AB)^T &= \begin{bmatrix} 3 & -4 & -4 \\ -5 & 5 & 9 \end{bmatrix}^T \\ &= \begin{bmatrix} 3 & -5 \\ -4 & 5 \\ -4 & 9 \end{bmatrix} \end{aligned}$$

Next, compute B^TA^T:

$$B^TA^T = \begin{bmatrix} -2 & -1 & 0 \\ 1 & -2 & 0 \\ 2 & 0 & -1 \end{bmatrix} \begin{bmatrix} -2 & 3 \\ 1 & -1 \\ 0 & -3 \end{bmatrix}$$
$$= \begin{bmatrix} 3 & -5 \\ -4 & 5 \\ -4 & 9 \end{bmatrix}$$

We see that $(AB)^T = B^TA^T$. ◀

The following theorem summarises the properties of the transpose of a matrix.

Theorem 2.3

Let A and B be matrices of appropriate dimensions and let α be a scalar. Then

(a) $(A^T)^T = A$

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

(c) $(\alpha A)^T = \alpha A^T$

(d) $(AB)^T = B^TA^T$

♡

A consequence of property (4) is that

$$(A_1A_2\ldots A_k)^T = A_k^TA_{k-1}^T\cdots A_2^TA_1^T$$

and as a special case

$$\left(A^k\right)^T = \left(A^T\right)^k$$

2.3 Invertible Matrices

The inverse of a square matrix $A \in \mathbb{R}^{n\times n}$ extends the concept of the reciprocal for a nonzero real number $a \in \mathbb{R}$. More precisely, the inverse of a non-zero number $a \in \mathbb{R}$ is the unique number $c \in \mathbb{R}$ such that $ac = ca = 1$. The inverse of $a \neq 0$, typically written as $a^{-1} = \frac{1}{a}$, enables solving the equation $ax = b$:

$$ax = b \Rightarrow a^{-1}ax = a^{-1}b \Rightarrow x = a^{-1}b.$$

This concept extends to square matrices, where the inverse of a matrix $A \in \mathbb{R}^{n\times n}$ is a matrix $C \in \mathbb{R}^{n\times n}$ such that $AC = CA = I_n$, where I_n is the identity matrix of size $n \times n$. The inverse of a matrix is denoted by A^{-1}. We can now define the invertible matrix.

Definition 2.6

A square matrix $A \in \mathbb{R}^{n\times n}$ is invertible (or **nonsingular**) if there exists a matrix $C \in \mathbb{R}^{n\times n}$ such that $AC = CA = I_n$. The matrix C is called the inverse of A and is denoted by A^{-1}. Thus, $A^{-1}A = AA^{-1} = I_n$.

♣

Example 2.8 Given A and C below, show that C is the inverse of A.

$$A = \begin{bmatrix} 1 & -3 & 0 \\ -1 & 2 & -2 \\ -2 & 6 & 1 \end{bmatrix}, \quad C = \begin{bmatrix} -14 & -3 & -6 \\ -5 & -1 & -2 \\ 2 & 0 & 1 \end{bmatrix}$$

Solution: We need to show that $AC = CA = I_3$. First, compute AC:

$$AC = \begin{bmatrix} 1 & -3 & 0 \\ -1 & 2 & -2 \\ -2 & 6 & 1 \end{bmatrix} \begin{bmatrix} -14 & -3 & -6 \\ -5 & -1 & -2 \\ 2 & 0 & 1 \end{bmatrix}$$
$$= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = I_3$$

Next, compute CA:

$$CA = \begin{bmatrix} -14 & -3 & -6 \\ -5 & -1 & -2 \\ 2 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & -3 & 0 \\ -1 & 2 & -2 \\ -2 & 6 & 1 \end{bmatrix}$$
$$= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = I_3$$

We see that $C = A^{-1}$ is the inverse of A. ◀

The following theorem summarizes the relationship between the matrix inverse and matrix multiplication and matrix transpose.

Theorem 2.4

Let A and B be invertible $n \times n$ matrices. Then:

(a) A^{-1} is invertible, with
$$\left(A^{-1}\right)^{-1} = A$$

(b) The product AB is invertible, with
$$(AB)^{-1} = B^{-1}A^{-1}$$

(c) The transpose of A is also invertible, i.e. A^T is invertible, with
$$\left(A^T\right)^{-1} = \left(A^{-1}\right)^T$$

♡

2.4 The Invertible Matrix Theorem

The Invertible Matrix Theorem provides a comprehensive list of equivalent statements for a square matrix to be invertible and it enables us to determine the truth value of one statement by checking the truth value of another.

Theorem 2.5

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

(a) A is invertible.
(b) A is row equivalent to I_n.
(c) A has n pivot positions (i.e. one for each row and column).
(d) The equation $A\boldsymbol{x} = \overline{0}$ has only the trivial solution.
(e) The columns of A are linearly independent.
(f) The equation $A\boldsymbol{x} = \boldsymbol{b}$ has a unique solution for each $\boldsymbol{b} \in \mathbb{R}^n$.
(g) The columns of $A \operatorname{span} \mathbb{R}^n$.
(h) There is an $n \times n$ matrix C such that $CA = I$.
(i) There is an $n \times n$ matrix D such that $AD = I$.
(j) A^T is invertible.

♡

Note how the theorem connects the properties of a matrix to its invertibility, thus summarising the main elements of our discussion on linear systems, independence, vectors, and matrices.

2.5 Bases in Vector Spaces

A vector space is a collection of vectors that satisfies certain properties, such as closure under addition and scalar multiplication. A basis of a vector space provides a fundamental way to represent vectors, define coordinates, and understand transformations.

A **basis** of a vector space V is a set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ that satisfies two essential properties:

1. **Linear Independence**: No vector in the set can be written as a linear combination of the others.
2. **Spanning Property**: Any vector in V can be expressed as a linear combination of the basis vectors.

If V has dimension n, any basis of V consists of exactly n vectors.

Given a basis $\{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_n\}$ for an n-dimensional space, any vector $\mathbf{v} \in V$ can be uniquely expressed as:

$$\mathbf{v} = c_1\mathbf{b}_1 + c_2\mathbf{b}_2 + \cdots + c_n\mathbf{b}_n,$$

where the scalars $c_1, c_2, \ldots, c_n$ are called the **coordinates** of $\mathbf{v}$ in this basis.

The most common basis for $\mathbb{R}^n$ is the **standard basis**, given by:

$$\mathbf{e}_1 = (1, 0, \ldots, 0), \quad \mathbf{e}_2 = (0, 1, \ldots, 0), \quad \ldots, \quad \mathbf{e}_n = (0, 0, \ldots, 1).$$

Each vector in $\mathbb{R}^n$ can be directly written in terms of these basis vectors using its Cartesian coordinates.

A basis is **orthogonal** if its vectors are pairwise perpendicular, meaning $\mathbf{b}_i \cdot \mathbf{b}_j = 0$ for $i \neq j$. If, in addition, each basis vector has unit length ($\|\mathbf{b}_i\| = 1$), then the basis is **orthonormal**. Orthonormal bases simplify computations, as the coordinates of a vector can be found using dot products:

$$c_i = \mathbf{v} \cdot \mathbf{b}_i.$$

Change of Basis

A vector's representation depends on the chosen basis. Given two bases $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ and $\mathcal{C} = \{\mathbf{c}_1, \ldots, \mathbf{c}_n\}$, we can express any vector $\mathbf{v}$ in terms of either basis. The transformation between these representations is governed by a **change of basis matrix**, denoted as P, where:

$$[\mathbf{v}]_{\mathcal{C}} = P[\mathbf{v}]_{\mathcal{B}}.$$

The columns of P contain the coordinates of the new basis vectors relative to the old basis.

2.6 Norms and Inner Products

As we will see, the norm and the inner products are two fundamental concepts in linear algebra that play a crucial role in defining distances, angles, and orthogonality in vector spaces.

The Norm

A **norm**, i.e. the length of a vector, is a function that assigns a non-negative length or size to a vector. Given a vector $\mathbf{v} \in \mathbb{R}^n$, a norm is a function $\|\cdot\| : \mathbb{R}^n \to \mathbb{R}$ that satisfies the following properties:

1. **Non-negativity**: $\|\mathbf{v}\| \geq 0$, with equality if and only if $\mathbf{v} = \mathbf{0}$.
2. **Absolute Homogeneity**: $\|c\mathbf{v}\| = |c| \cdot \|\mathbf{v}\|$ for any scalar c.
3. **Triangle Inequality**: $\|\mathbf{u} + \mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|$.

Common norms include:

- **Euclidean norm** (or L_2 norm): $\|\mathbf{v}\|_2 = \sqrt{\sum_{i=1}^n v_i^2}$, which measures the standard distance from the origin.
- **Manhattan norm** (or L_1 norm): $\|\mathbf{v}\|_1 = \sum_{i=1}^n |v_i|$, used in sparse optimization problems.
- **Infinity norm** (L_∞ norm): $\|\mathbf{v}\|_\infty = \max |v_i|$, useful in worst-case scenarios and robust learning.

The norm of a vector is directly related to its length. Specifically, the **Euclidean norm** (or L_2 norm) of a vector $\mathbf{v} \in \mathbb{R}^n$ measures the standard distance of $\mathbf{v}$ from the origin in Euclidean space, making it a natural choice for measuring vector magnitudes.

Since norms quantify vector magnitudes, they play a crucial role in defining distances, optimization techniques, and regularization methods in machine learning. For example, the L_1 norm promotes sparsity in feature selection, while the L_2 norm is common in ridge regression.

Inner Products

The **inner product** (also called the dot product in Euclidean spaces) is a function that measures the similarity between two vectors. Given two vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$, the standard inner product is defined as:

$$\langle \mathbf{u}, \mathbf{v} \rangle = \sum_{i=1}^n u_i v_i.$$

The inner product satisfies the following properties:

1. **Commutativity**: $\langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{u} \rangle$.
2. **Linearity**: $\langle c\mathbf{u} + d\mathbf{w}, \mathbf{v} \rangle = c\langle \mathbf{u}, \mathbf{v} \rangle + d\langle \mathbf{w}, \mathbf{v} \rangle$.

3. **Positivity**: $\langle \mathbf{v}, \mathbf{v} \rangle \geq 0$, with equality if and only if $\mathbf{v} = \mathbf{0}$.

The inner product induces a norm, called the **Euclidean norm**, as:

$$\|\mathbf{v}\|_2 = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}.$$

Moreover, the inner product allows us to define angles between vectors, where the **cosine similarity** between two vectors is given by:

$$\cos\theta = \frac{\langle \mathbf{u}, \mathbf{v} \rangle}{\|\mathbf{u}\|\|\mathbf{v}\|}.$$

This measure is extensively used in machine learning for comparing text documents, clustering, and recommendation systems.

Norms and Inner Products in Machine Learning

Both norms and inner products play crucial roles in machine learning:

- **Distance Metrics**: Norms define distances in feature spaces, impacting clustering, nearest neighbor algorithms, and optimization.
- **Regularization**: L_1 and L_2 norms are used to constrain model complexity, preventing overfitting.
- **Gradient-Based Optimization**: Many learning algorithms, such as gradient descent, rely on inner products to compute updates efficiently.
- **Kernel Methods**: Inner products generalize to kernel functions, allowing the extension of linear models to non-linear feature spaces.

Norms and inner products are essential tools in linear algebra and machine learning. Norms provide a measure of size and distance, while inner products capture similarity and geometric relationships.

Chapter 3 Eigenvalues, -vectors, and Matrix Decomposition

Eigenvalues and eigenvectors are important concepts in linear algebra. They are used in many applications, including physics, engineering, and computer science. In this chapter, we will define eigenvalues and eigenvectors, and discuss their properties. We will also discuss matrix decomposition, which is a way to break down a matrix into simpler parts. This chapter will provide a foundation for understanding more advanced topics in Machine Learning.

3.1 Eigenvalues and Eigenvectors

Let us start by defining eigenvalues and eigenvectors.

Definition 3.1

Let A be an $n \times n$ matrix. A scalar λ is called an **eigenvalue** of A if there exists a non-zero vector $\mathbf{v}$ such that

$$A\mathbf{v} = \lambda\mathbf{v}.$$

The vector $\mathbf{v}$ is called an **eigenvector** of A corresponding to the eigenvalue λ.

♣

The **eigenspace** of an eigenvalue λ is the set of all eigenvectors corresponding to λ, along with the zero vector. The eigenspace of an eigenvalue λ is a subspace of $\mathbb{R}^n$. The dimension of the eigenspace of an eigenvalue λ is called the **geometric multiplicity** of λ. The geometric multiplicity of an eigenvalue λ is denoted by $g(\lambda)$. The geometric multiplicity of an eigenvalue λ is less than or equal to the algebraic multiplicity of λ. The algebraic multiplicity of an eigenvalue λ is the number of times λ appears as a root of the characteristic polynomial of A. The sum of the geometric multiplicities of all eigenvalues of A is equal to n.

Eigenvalues and eigenvectors have several important properties, including:

- The trace of a matrix A, given by $\text{tr}(A) = \sum_i a_{ii}$, is equal to the sum of its eigenvalues.
- If A is a triangular matrix (upper or lower triangular), its eigenvalues are simply the entries on its diagonal.
- A matrix is singular if and only if it has at least one eigenvalue equal to zero.
- The eigenvectors of a symmetric matrix are always orthogonal, and the matrix can be diagonalized using an orthonormal basis.

Eigenvalue Decomposition

If A has n linearly independent eigenvectors, it can be diagonalized as:

$$A = PDP^{-1},$$

where P is the matrix of eigenvectors and D is a diagonal matrix whose diagonal entries are the eigenvalues of A.

3.2 Matrix Decomposition - Singular Value Decomposition

Matrix decomposition is the process of breaking down a matrix into simpler parts that make it easier to work with. Decomposing matrices allows for efficient computation, better numerical stability, and deeper insights into their structure. There are several types of matrix decomposition, each with its own properties and applications. Some common types of matrix decomposition include:

- **LU Decomposition**: Factorizes a matrix as the product of a lower triangular matrix and an upper triangular matrix, useful for solving systems of linear equations.
- **QR Decomposition**: Expresses a matrix as the product of an orthogonal matrix and an upper triangular matrix, often used in solving least squares problems.
- **Singular Value Decomposition (SVD)**: Provides a decomposition that is applicable to any matrix, revealing fundamental geometric and algebraic properties.

We here focus on Singular Value Decomposition (SVD), which is among the most powerful and widely used decompositions.

Singular Value Decomposition (SVD)

Singular Value Decomposition (SVD) is a matrix factorization method that expresses any real $A \in \mathbb{R}^{m\times n}$ as:

$$A = U\Sigma V^T,$$

where:

- $U \in \mathbb{R}^{m\times m}$ is an orthogonal matrix whose columns are the left singular vectors of A.
- $V \in \mathbb{R}^{n\times n}$ is an orthogonal matrix whose columns are the right singular vectors of A.
- $\Sigma \in \mathbb{R}^{m\times n}$ is a diagonal matrix with non-negative real numbers on the diagonal, called the **singular values** of A.

The singular values are arranged in descending order, meaning:

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0,$$

where r is the rank of A and σ_i are the singular values.

Geometric Interpretation of SVD

SVD provides a powerful geometric interpretation of matrices. The transformation A maps an input space to an output space in a structured manner:

- The matrix V represents an orthonormal basis of the domain.
- The diagonal matrix Σ scales each basis vector by its corresponding singular value.
- The matrix U represents an orthonormal basis of the codomain.

Thus, SVD essentially decomposes a transformation into a rotation, a scaling, and another rotation.

Applications of SVD

SVD is widely used in various fields due to its ability to reveal fundamental properties of matrices and approximate them with lower-rank representations. Key applications include:

- **Dimensionality Reduction**: In Principal Component Analysis (PCA), SVD is used to find principal components by identifying directions of maximum variance.
- **Image Compression**: By keeping only the largest singular values and their corresponding singular vectors, images can be approximated with reduced storage while retaining most of their structure.
- **Latent Semantic Analysis (LSA)**: SVD is used in natural language processing to identify relationships between words and documents.
- **Solving Linear Systems**: The Moore-Penrose pseudoinverse, derived from SVD, is used to solve least-squares problems efficiently.
- **Machine Learning and Recommender Systems**: Many collaborative filtering algorithms leverage SVD to uncover hidden patterns in data.
- **Numerical Stability in Regression**: SVD is often used in Python implementations of linear regression to solve the normal equations with increased numerical stability compared to direct matrix inversion.

Low-Rank Approximation

One of the most powerful features of SVD is its ability to approximate a matrix with a lower-rank representation. Given:

$$A_k = U_k \Sigma_k V_k^T,$$

where U_k and V_k consist of only the first k singular vectors, and Σ_k contains only the top k singular values, A_k serves as the best rank-k approximation of A in terms of minimizing reconstruction error. This property is crucial for reducing complexity in high-dimensional data analysis.

Bibliography

Lay, D. (2003). *Linear algebra and its applications*. Pearson Education.