Linear Algebra

Lecture Notes

HAMZA BERAHMA

Student

December 17, 2025

For

[redacted],

my idea of beautiful

Contents

Chapter 1

Matrices and Systems of Equations

1.1 Systems of Linear Equations

A foundational problem in mathematics and its applications is the solving of systems of linear equations. These systems model a wide array of phenomena in fields ranging from engineering and physics to economics and biology.

Definition 1 (Linear System). A **linear equation** in n unknowns is an equation of the form:

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

where $a_1, \ldots, a_n$ are the coefficients, $x_1, \ldots, x_n$ are the variables, and b is the constant term.

A **system of linear equations** is a collection of one or more linear equations involving the same set of variables. An $m \times n$ system has m equations and n unknowns:

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

Definition 2 (Solution Set). A **solution** to an $m \times n$ system is an ordered n-tuple $(s_1, s_2, \ldots, s_n)$ that, when substituted for $(x_1, x_2, \ldots, x_n)$, satisfies every equation simultaneously. The set of all such solutions is the **solution set**.

Remark. A linear system can be classified by the nature of its solution set:

- **Inconsistent System:** A system with no solution. Its solution set is empty.
- **Consistent System:** A system with at least one solution.

Theorem 1 (Nature of Solutions). A consistent linear system has either:

1. exactly one solution, or

2. infinitely many solutions.

It is impossible for a consistent linear system to have a finite number of solutions greater than one.

Proof. If a linear system has at least two distinct solutions, say $\mathbf{x}_1$ and $\mathbf{x}_2$, then any point on the line connecting them, $\mathbf{x} = t\mathbf{x}_1 + (1-t)\mathbf{x}_2$ for $t \in \mathbb{R}$, is also a solution.

Therefore, having two solutions implies there are infinitely many. Hence, a consistent system can only have exactly one solution or infinitely many solutions. $\square$

1.1.1 Geometric Interpretation of 2×2 Systems

Each equation in a 2×2 system represents a line in the plane. The solution set corresponds to the intersection of these lines.

Remark (Geometric Possibilities). There are three possibilities for the intersection of two lines:

- **One Solution:** The lines intersect at a single point.
- **No Solution:** The lines are parallel and distinct.
- **Infinitely Many Solutions:** Both equations represent the same line.

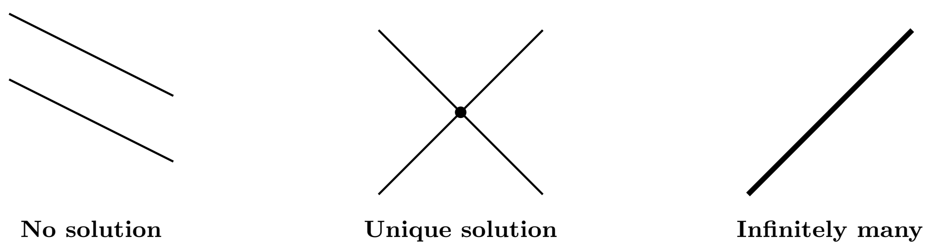


Figure 1.1: Geometric possibilities for a 2×2 linear system.

1.1.2 Equivalent Systems

The primary method for solving systems is to transform a complex system into a simpler, **equivalent** one (a system with the same solution set). This is achieved via three elementary operations.

Proposition 1 (Elementary Operations for Systems). The following operations produce an equivalent system:

I. **Interchange:** Swap two equations ($R_i \leftrightarrow R_j$).

II. **Scaling:** Multiply an equation by $c \neq 0$ ($R_i \to cR_i$).

III. **Replacement:** Add a multiple of one equation to another ($R_i \to R_i + cR_j$).

1.1.3 Augmented Matrix Representation

To simplify the solving process, we represent linear systems using matrices.

Definition 3 (Augmented Matrix). For a given linear system, the **coefficient matrix** contains the coefficients of the variables, and the **augmented matrix** is the coefficient matrix with an added column for the constant terms.

$$\text{System: } \begin{cases} x_1 + 2x_2 + \ \ x_3 = 3 \\ 3x_1 - \ \ x_2 - 3x_3 = -1 \end{cases} \quad \text{Augmented: } \left(\begin{array}{ccc|c} 1 & 2 & 1 & 3 \\ 3 & -1 & -3 & -1 \end{array}\right)$$

Remark. The elementary operations on systems correspond to **elementary row operations** on matrices:

I. Interchange two rows ($R_i \leftrightarrow R_j$).

II. Multiply a row by a non-zero scalar ($R_i \to cR_i$).

III. Replace a row by its sum with a multiple of another row ($R_i \to R_i + cR_j$).

Example (Basic Gaussian Elimination). Solve: $\begin{cases} 2x_1 + \ \ x_2 = 8 \\ 4x_1 - 3x_2 = 6 \end{cases}$.

Form augmented matrix and perform row operations:

$$\left(\begin{array}{cc|c} 2 & 1 & 8 \\ 4 & -3 & 6 \end{array}\right) \xrightarrow{R_2 \to R_2 + -2R_1} \left(\begin{array}{cc|c} 2 & 1 & 8 \\ 0 & -5 & -10 \end{array}\right)$$

This is REF. The system is $\begin{cases} 2x_1 + \ \ x_2 = 8 \\ \quad\ \ - 5x_2 = -10 \end{cases}$.

Back substitution: $-5x_2 = -10 \Rightarrow x_2 = 2 \Rightarrow 2x_1 + 2 = 8 \Rightarrow x_1 = 3$.

The unique solution is

$$\boxed{(\mathbf{3}, \mathbf{2})}$$

1.2 Row Echelon Form and Gaussian Elimination

1.2.1 Intuition and Core Strategy

Remark. The core idea behind solving a system $A\mathbf{x} = \mathbf{b}$ is to replace it with an equivalent, simpler system that has the same solution set. We achieve this by translating the system into an augmented matrix $(A|\mathbf{b})$ and using **elementary row operations** to isolate the variables one by one. This process yields a matrix in REF (Gaussian Elimination) or RREF (Gauss-Jordan Elimination), making **back substitution** trivial.

1.2.2 Row Echelon Forms

Definition 4 (Row Echelon Form (REF)). A matrix is in REF if:

1. All zero rows are at the bottom.
2. In each non-zero row, the first non-zero entry (the **pivot**) is in a column strictly to the right of the pivot in the row above it.
3. All entries below a pivot are zero.

Definition 5 (Reduced Row Echelon Form (RREF)). A matrix is in RREF if it satisfies all REF conditions, AND:

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

Remark (Variable Classification). Let $r = \text{rank}(A)$ be the number of pivots in the coefficient matrix A. Variables are classified as:

- **basic variables**: Variables corresponding to pivot columns. There are r basic variables.
- **free variables**: Variables corresponding to columns without pivots. There are $n-r$ free variables. If $n-r > 0$, the system has infinitely many solutions (provided it is consistent).

Example (Standard Gaussian Elimination). Solve the system:

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

Form augmented matrix: $\left(\begin{array}{ccc|c} 1 & 2 & 1 & 3 \\ 3 & -1 & -3 & -1 \\ 2 & 3 & 1 & 4 \end{array}\right)$

Step 1: Eliminate below R_1:

$$\left(\begin{array}{ccc|c} 1 & 2 & 1 & 3 \\ 3 & -1 & -3 & -1 \\ 2 & 3 & 1 & 4 \end{array}\right) \xrightarrow{R_2 \to R_2 + -3R_1, R_3 \to R_3 + -2R_1} \left(\begin{array}{ccc|c} 1 & 2 & 1 & 3 \\ 0 & -7 & -6 & -10 \\ 0 & -1 & -1 & -2 \end{array}\right)$$

Step 2: Swap $R_2 \leftrightarrow R_3$ for convenience, then eliminate below R_2:

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

Now in REF. System: $\begin{cases} x_1 + 2x_2 + x_3 = 3 \\ \quad - \ x_2 - x_3 = -2 \\ \qquad\qquad x_3 = 4 \end{cases}$

Back substitution: $x_3 = 4 \Rightarrow -x_2 - 4 = -2 \Rightarrow x_2 = -2 \Rightarrow x_1 + 2(-2) + 4 = 3 \Rightarrow x_1 = 3$

The unique solution is $\boxed{(\mathbf{3}, -\mathbf{2}, \mathbf{4})}$.

1.2.3 Solution Analysis and Consistency

The REF or RREF of $(A|\mathbf{b})$ reveals the solution structure. Let n be the number of variables and $r = \text{rank}(A)$.

Theorem 2 (Solution Characterization). For a linear system with augmented matrix $(A|\mathbf{b})$:

1. **Inconsistent** (no solution): Occurs if, after row reduction, there is a row of the form
$$[0 \ \ 0 \ \ \ldots \ \ 0 \mid b], \quad b \neq 0.$$
Such a row corresponds to the impossible equation $0 = b$.

2. **Consistent** : Occurs when no such row exists. The solution type depends on the number of leading variables compared to the number of unknowns n:
   - **Unique solution** : Every variable is a leading variable (no free variables).
   - **Infinitely many solutions** : At least one variable is free.

Example (Infinite Solutions). Consider the system in REF: $\left(\begin{array}{ccc|c} 1 & -2 & 1 & 4 \\ 0 & 1 & 2 & -3 \\ 0 & 0 & 0 & 0 \end{array}\right)$

System: $\begin{cases} x_1 - 2x_2 + \ x_3 = 4 \\ \qquad\quad x_2 + 2x_3 = -3 \end{cases}$

x_3 is a free variable. Let $x_3 = t$ where $t \in \mathbb{R}$.
Back substitution:
$$x_2 = -3 - 2x_3 \Rightarrow x_2 = -3 - 2t$$
$$x_1 = 4 + 2x_2 - x_3 \Rightarrow x_1 = 4 + 2(-3 - 2t) - t \Rightarrow x_1 = -2 - 5t$$

Solution set: $\boxed{\{(-\mathbf{2} - \mathbf{5t}, -\mathbf{3} - \mathbf{2t}, \mathbf{t}) \mid \mathbf{t} \in \mathbb{R}\}}$

Example (No Solution). Consider: $\left(\begin{array}{cc|c} 1 & 2 & 5 \\ 0 & 0 & 8 \end{array}\right)$

Row 2 gives $0x_1 + 0x_2 = 8$, or $0 = 8$ (contradiction).
The system is inconsistent and has no solution.

Example (Larger System). Solve: $\begin{cases} \quad\quad\; x_2 - \; x_3 + \; x_4 = 0 \\ x_1 + \; x_2 + \; x_3 + \; x_4 = 6 \\ 2x_1 + 4x_2 + \; x_3 - 2x_4 = -1 \\ 3x_1 + \; x_2 - 2x_3 + 2x_4 = 3 \end{cases}$

Augmented matrix: $\left(\begin{array}{cccc|c} 0 & -1 & -1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 6 \\ 2 & 4 & 1 & -2 & -1 \\ 3 & 1 & -2 & 2 & 3 \end{array}\right)$

Step 1: Swap to get leading 1:

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

Step 2: Eliminate below R_1:

$$\xrightarrow{R_3 \to R_3 + -2R_1, R_4 \to R_4 + -3R_1} \left(\begin{array}{cccc|c} 1 & 1 & 1 & 1 & 6 \\ 0 & -1 & -1 & 1 & 0 \\ 0 & 2 & -1 & -4 & -13 \\ 0 & -2 & -5 & -1 & -15 \end{array}\right)$$

Step 3: Eliminate below R_2:

$$\xrightarrow{R_3 \to R_3 + 2R_2, R_4 \to R_4 + -2R_2} \left(\begin{array}{cccc|c} 1 & 1 & 1 & 1 & 6 \\ 0 & -1 & -1 & 1 & 0 \\ 0 & 0 & -3 & -2 & -13 \\ 0 & 0 & -3 & -3 & -15 \end{array}\right)$$

Step 4: Final elimination:

$$\xrightarrow{R_4 \to R_4 + -1R_3} \left(\begin{array}{cccc|c} 1 & 1 & 1 & 1 & 6 \\ 0 & -1 & -1 & 1 & 0 \\ 0 & 0 & -3 & -2 & -13 \\ 0 & 0 & 0 & -1 & -2 \end{array}\right)$$

Matrix in REF. System: $\begin{cases} x_1 + x_2 + \; x_3 + \; x_4 = 6 \\ \quad - x_2 - \; x_3 + \; x_4 = 0 \\ \quad\quad\quad - 3x_3 - 2x_4 = -13 \\ \quad\quad\quad\quad\quad - \; x_4 = -2 \end{cases}$

Back substitution: $x_4 = 2 \Rightarrow -3x_3 - 4 = -13 \Rightarrow x_3 = 3 \Rightarrow -x_2 - 3 + 2 = 0 \Rightarrow x_2 = -1 \Rightarrow x_1 = 2$

The unique solution is $\boxed{(\mathbf{2}, \mathbf{-1}, \mathbf{3}, \mathbf{2})}$.

1.2.4 Homogeneous Systems

Definition 6 (Homogeneous System). A system of the form $A\mathbf{x} = \mathbf{0}$ (where $\mathbf{b} = \mathbf{0}$) is called **homogeneous**. Such systems are always consistent since $\mathbf{x} = \mathbf{0}$ is a solution (the **trivial solution**).

Theorem 3 (Homogeneous Systems). An $m \times n$ homogeneous system has a **nontrivial solution** (i.e., infinitely many solutions) if $n > m$.

Proof. The system is always consistent (trivial solution $\mathbf{x} = \mathbf{0}$ exists). The number of pivots r satisfies $r \leq m$. If $n > m$, then $r \leq m < n$, so $r < n$. Thus, there are $n - r \geq 1$ free variables, giving infinitely many solutions. $\square$

Example (Complete Reduction to RREF). Solve the homogeneous system using Gauss-Jordan elimination:

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

This is 3×4 with $n = 4 > m = 3$, so we expect infinitely many solutions.

Augmented matrix: $\left(\begin{array}{cccc|c} -1 & 1 & -1 & 3 & 0 \\ 3 & 1 & -1 & -1 & 0 \\ 2 & -1 & -2 & -1 & 0 \end{array}\right)$

Step 1: Reduce to REF:

$$\xrightarrow{R_2 \to R_2 + 3R_1, R_3 \to R_3 + 2R_1} \left(\begin{array}{cccc|c} -1 & 1 & -1 & 3 & 0 \\ 0 & 4 & -4 & 8 & 0 \\ 0 & 1 & -4 & 5 & 0 \end{array}\right)$$

$$\xrightarrow{R_2 \leftrightarrow R_3} \left(\begin{array}{cccc|c} -1 & 1 & -1 & 3 & 0 \\ 0 & 1 & -4 & 5 & 0 \\ 0 & 4 & -4 & 8 & 0 \end{array}\right) \xrightarrow{R_3 \to R_3 + -4R_2} \left(\begin{array}{cccc|c} -1 & 1 & -1 & 3 & 0 \\ 0 & 1 & -4 & 5 & 0 \\ 0 & 0 & 12 & -12 & 0 \end{array}\right)$$

$$\xrightarrow{R_1 \to -1R_1, R_3 \to 1/12 R_3} \left(\begin{array}{cccc|c} 1 & -1 & 1 & -3 & 0 \\ 0 & 1 & -4 & 5 & 0 \\ 0 & 0 & 1 & -1 & 0 \end{array}\right) \quad \text{(REF)}$$

Step 2: Continue to RREF (eliminate above pivots):

$$\xrightarrow{R_2 \to R_2 + 4R_3, R_1 \to R_1 + -1R_3} \left(\begin{array}{cccc|c} 1 & -1 & 0 & -2 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & -1 & 0 \end{array}\right)$$

$$\xrightarrow{R_1 \to R_1 + 1R_2} \left(\begin{array}{cccc|c} 1 & 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & -1 & 0 \end{array}\right) \quad \text{(RREF)}$$

Pivots in columns 1, 2, 3. Thus $r = 3$.

- **basic variables**: x_1, x_2, x_3
- **free variable**: x_4

Let $x_4 = t$ where $t \in \mathbb{R}$. From RREF:

$$\begin{cases} x_1 - t = 0 \\ x_2 + t = 0 \\ x_3 - t = 0 \end{cases} \iff x_1 = t \Rightarrow x_2 = -t \Rightarrow x_3 = t$$

Solution set as parameterized vector:

$$\boxed{\mathbb{S} = \left\{ \mathbf{t} \begin{bmatrix} 1 \\ -1 \\ 1 \\ 1 \end{bmatrix} \middle| \mathbf{t} \in \mathbb{R} \right\}}$$

Remark (Uniqueness of RREF). While a matrix can have many different REFs (depending on the sequence of row operations), the RREF of any given matrix is **unique**.

Remark (System Classification). The terms *overdetermined* ($m > n$) and *underdetermined* ($m < n$) refer to the matrix dimensions, not the solution type. An overdetermined system can be consistent, and an underdetermined system can be inconsistent.

1.3 Matrix Arithmetic

This section introduces the fundamental operations of matrix algebra: addition, scalar multiplication, matrix multiplication, and transposition. These operations provide a powerful framework for representing and solving systems of linear equations.

1.3.1 Matrix Notation and Terminology

Definition 7 (Matrix and Vectors). An $m \times n$ **matrix** A is a rectangular array of numbers with m rows and n columns. The entry in the i-th row and j-th column is denoted by a_{ij}.

$$A = (a_{ij}) = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}$$

The entries a_{ij} are called **scalars**.

- A **column vector** is an $m \times 1$ matrix. The set of all real $m \times 1$ vectors is denoted $\mathbb{R}^m$. We use bold lowercase letters for column vectors, e.g., $\mathbf{b}$.
- A **row vector** is a $1 \times n$ matrix.

1.3.2 Addition, Scalar Multiplication, and Equality

Definition 8 (Equality and Basic Operations). Let $A = (a_{ij})$ and $B = (b_{ij})$ be $m \times n$ matrices and let c be a scalar.

1. **Equality:** $A = B$ if and only if $a_{ij} = b_{ij}$ for all i, j.
2. **Matrix Addition:** The sum $A + B$ is the $m \times n$ matrix $C = (c_{ij})$ where $c_{ij} = a_{ij} + b_{ij}$.
3. **Scalar Multiplication:** The scalar product cA is the $m \times n$ matrix $D = (d_{ij})$ where $d_{ij} = c \cdot a_{ij}$.

Matrix subtraction is defined as $A - B = A + (-1)B$.

Remark (The Zero Matrix). The $m \times n$ **zero matrix**, denoted $\mathbf{0}$ or $\mathbf{0}_{m \times n}$, is a matrix whose entries are all zero. It serves as the additive identity: $A + \mathbf{0} = \mathbf{0} + A = A$.

Example (Basic Operations). Let $A = \begin{bmatrix} 4 & 0 \\ -1 & 5 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & -1 \\ 2 & 3 \end{bmatrix}$.

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

$$3A = \begin{bmatrix} 3 \cdot 4 & 3 \cdot 0 \\ 3 \cdot (-1) & 3 \cdot 5 \end{bmatrix} = \begin{bmatrix} 12 & 0 \\ -3 & 15 \end{bmatrix}$$

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

1.3.3 Matrix Multiplication

intuition. Matrix multiplication is defined in a way that is not as straightforward as addition, but is perfectly tailored to represent systems of linear equations. The equation $A\mathbf{x} = \mathbf{b}$ compactly represents an entire system.

The product of two matrices A and B is defined only if the number of columns in A equals the number of rows in B.

Definition 9 (Matrix Multiplication). If A is an $m \times n$ matrix and B is an $n \times p$ matrix, then the **product** $C = AB$ is an $m \times p$ matrix. The entry c_{ij} is the dot product of the i-th row of A and the j-th column of B:

$$c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}$$

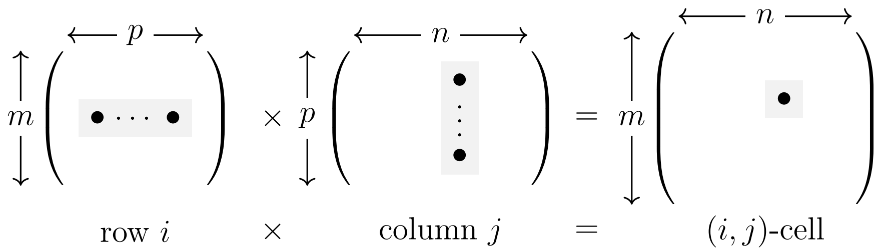


Figure 1.2: Matrix multiplication: entry c_{ij} is computed from row i of A and column j of B.

Example (Matrix Multiplication). Let

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

A is 2×2 and B is 2×3. The product AB is defined and has size 2×3:

$$AB = \begin{bmatrix} 2 & 1 \\ 3 & 0 \end{bmatrix} \begin{bmatrix} 1 & -1 & 4 \\ 5 & 2 & 0 \end{bmatrix} = \begin{bmatrix} 7 & 0 & 8 \\ 3 & -3 & 12 \end{bmatrix}.$$

The product BA is **not defined**, since the number of columns of B (3) does not equal the number of rows of A (2).

Remark (Non-Commutativity of Matrices). Matrix multiplication is **not commutative**: in general, $AB \neq BA$, even for square matrices.

$$A = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad AB \neq BA.$$

It *does* commute when A and B are diagonal, or more generally, when they belong to an abelian subgroup of $GL(n, \mathbb{R})$.

1.3.4 Matrix-Vector Product and Linear Systems

The product of a matrix A and a column vector $\mathbf{x}$ has two useful interpretations.

Proposition 2 (Interpretations of $A\mathbf{x}$). Let $A \in \mathbb{R}^{m \times n}$ with rows $\vec{a}_1^T, \dots, \vec{a}_m^T$ and columns $\mathbf{a}_1, \dots, \mathbf{a}_n$, and let $\mathbf{x} \in \mathbb{R}^n$. Then $A\mathbf{x}$ can be interpreted as:

1. **Row-Vector Product:** Each component of $A\mathbf{x}$ is the dot product of a row of A with $\mathbf{x}$:

$$A\mathbf{x} = \begin{bmatrix} \vec{a}_1^T \cdot \mathbf{x} \\ \vec{a}_2^T \cdot \mathbf{x} \\ \vdots \\ \vec{a}_m^T \cdot \mathbf{x} \end{bmatrix}.$$

2. **Linear Combination of Columns:** $A\mathbf{x}$ is a linear combination of the columns

of A, with entries of $\mathbf{x}$ as coefficients:

$$A\mathbf{x} = x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_n\mathbf{a}_n.$$

This second interpretation leads to a fundamental theorem about consistency.

Theorem 4 (Consistency Theorem). A linear system $A\mathbf{x} = \mathbf{b}$ is **consistent** if and only if the vector $\mathbf{b}$ can be written as a linear combination of the column vectors of A.

Proof. Suppose A is an $m \times n$ matrix with columns $\mathbf{a}_1, \dots, \mathbf{a}_n$, and let $\mathbf{x} = (x_1, \dots, x_n)^T$.

($\Rightarrow$) If the system is consistent: If $A\mathbf{x} = \mathbf{b}$ has a solution $\mathbf{x}$, then

$$\mathbf{b} = A\mathbf{x} = x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_n\mathbf{a}_n,$$

so $\mathbf{b}$ is a linear combination of the columns of A.

($\Leftarrow$) If b is a linear combination of the columns: If there exist scalars $x_1, \dots, x_n$ such that

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

then setting $\mathbf{x} = (x_1, \dots, x_n)^T$ gives a solution to $A\mathbf{x} = \mathbf{b}$. Hence, the system is consistent. $\square$

1.3.5 The Transpose of a Matrix

Definition 10 (Transpose). The **transpose** of an $m \times n$ matrix A, denoted A^{T} (or A^T), is the $n \times m$ matrix obtained by interchanging the rows and columns of A. That is, the (i, j)-entry of A^{T} is a_{ji}.

Example (Transpose). Let

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

Then its transpose is

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

A column vector becomes a row vector under transposition:

$$\mathbf{v} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \quad \Rightarrow \quad \mathbf{v}^T = \begin{bmatrix} 1 & 2 & 3 \end{bmatrix}.$$

Definition 11 (Symmetric Matrix). A square matrix A is **symmetric** if $A^{\mathsf{T}} = A$. This requires that $a_{ij} = a_{ji}$ for all i, j.

Example (Symmetric Matrix). The matrix

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

is symmetric because $A^\mathsf{T} = A$.

1.3.6 Properties of Matrix Operations

Proposition 3 (Algebraic Properties). Assuming the matrix dimensions are compatible for the operations:

(1) $A + B \;=\; B + A$ (Commutativity of Addition)

(2) $(A + B) + C \;=\; A + (B + C)$ (Associativity of Addition)

(3) $A(BC) \;=\; (AB)C$ (Associativity of Multiplication)

(4) $A(B + C) \;=\; AB + AC$ (Left Distributivity)

(5) $(B + C)A \;=\; BA + CA$ (Right Distributivity)

(6) $c(A + B) \;=\; cA + cB$ (Scalar Distributivity)

(7) $(c + d)A \;=\; cA + dA$ (Scalar Distributivity)

(8) $(cd)A \;=\; c(dA)$ (Scalar Associativity)

(9) $(A^\mathsf{T})^\mathsf{T} \;=\; A$ (Double Transpose)

(10) $(A + B)^\mathsf{T} \;=\; A^\mathsf{T} + B^\mathsf{T}$ (Transpose of Sum)

(11) $(AB)^\mathsf{T} \;=\; B^\mathsf{T} A^\mathsf{T}$ (Transpose of Product: **note order reversal**)

(12) $(cA)^\mathsf{T} \;=\; cA^\mathsf{T}$ (Scalar and Transpose)

Remark (Important Non-Properties). Be careful with these common mistakes:

- $AB \neq BA$ in general (multiplication is not commutative)
- $(A + B)^2 \neq A^2 + 2AB + B^2$ in general (requires commutativity)
- $AB = \mathbf{0}$ does **not** imply $A = \mathbf{0}$ or $B = \mathbf{0}$ (zero divisors exist)

DIY. Let $A \in \mathbb{R}^{3\times 7}$, $B \in \mathbb{R}^{7\times 5}$, $C \in \mathbb{R}^{5\times 8}$. For a product of an $m \times n$ and an $n \times p$ matrix, we require mnp multiplications and $mn(p - 1)$ additions.

$$\begin{aligned}
(AB)C: \quad & AB: (3\times 7\times 5) = 105 \text{ mult.},\ 3\times 5\times (7-1) = 90 \text{ add.}\\
& (AB)C: (3\times 5\times 8) = 120 \text{ mult.},\ 3\times 8\times (5-1) = 96 \text{ add.}\\
& \Rightarrow \text{Total: } 225 \text{ mult.},\ 186 \text{ add.}\\
A(BC): \quad & BC: (7\times 5\times 8) = 280 \text{ mult.},\ 7\times 8\times (5-1) = 224 \text{ add.}\\
& A(BC): (3\times 7\times 8) = 168 \text{ mult.},\ 3\times 8\times (7-1) = 144 \text{ add.}\\
& \Rightarrow \text{Total: } 448 \text{ mult.},\ 368 \text{ add.}
\end{aligned}$$

Hence, $(AB)C$ is optimal with 225 multiplications and 186 additions.

Example (11). Let A be a 5×3 matrix. If

$$b = a_1 + a_2 = a_2 + a_3,$$

then $a_1 = a_3$. Hence the columns of A are linearly dependent, and the system $Ax = b$ has infinitely many solutions.

Example (12). Let A be a 3×4 matrix. If

$$b = a_1 + a_2 + a_3 + a_4,$$

then b is a linear combination of the columns of A, so the system $Ax = b$ is consistent. Because A has more unknowns than equations (3×4), it is *under-determined* and has infinitely many solutions.

DIY. Let A and B be $n\times n$ matrices. If $AB = I_n$, then $BA = I_n$.

Proof. This proof relies on a key result from the Invertible Matrix Theorem: For a square matrix, the existence of a one-sided inverse (either right or left) is sufficient to guarantee the existence of a two-sided inverse.

(a) We are given that $AB = I_n$. This means that the matrix B is a **right inverse** of A.

(b) A fundamental theorem of linear algebra states that if a **square** matrix has a right inverse, then it is **invertible**. Therefore, A is invertible.

(c) Since A is invertible, it has a unique two-sided inverse, denoted A^{-1}, such that:

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

(d) Now, we can take the original equation $AB = I_n$ and left-multiply both sides by A^{-1}:

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

(e) Using associativity on the left side gives:

$$(A^{-1}A)B = A^{-1}$$

(f) Since $A^{-1}A = I_n$, this simplifies to:

$$I_n B = A^{-1} \Rightarrow B = A^{-1}$$

(g) This shows that B is the unique inverse of A. By the definition of an inverse, B must also work as a left inverse. Therefore:

$$BA = I_n$$

□

1.3.7 Finding the Inverse of a Matrix

A square matrix A is **invertible** if there exists a matrix A^{-1} such that $AA^{-1} = A^{-1}A = I$. The matrix A^{-1} is called the **inverse** of A. A matrix that is not invertible is called a **singular** matrix.

A common method to find the inverse of a matrix is by using Gauss-Jordan elimination. We form an **augmented matrix** of the form $[A|I]$, where I is the identity matrix of the same size as A. We then perform elementary row operations until the left side of the augmented matrix is in reduced row echelon form (RREF).

If the RREF of A is the identity matrix I, then A is invertible, and the right side of the augmented matrix will be A^{-1}. The transformation looks like:

$$[A|I] \xrightarrow{\text{row operations}} [I|A^{-1}]$$

Example (Finding an Inverse using RREF). Let's find the inverse of the matrix

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

We form the augmented matrix $[A|I]$ and apply row operations to transform the left side into the identity matrix.

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

The left side is now the identity matrix. Therefore, A is invertible and its inverse is the right side.

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

We can check our work: $AA^{-1} = \begin{bmatrix} 1 & 2 \\ 3 & 5 \end{bmatrix} \begin{bmatrix} -5 & 2 \\ 3 & -1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I$.

Example (Case of a Non-Invertible Matrix). Let's attempt to find the inverse of the singular matrix

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

We set up the augmented matrix $[B|I]$ and begin row reduction.

$$\left[\begin{array}{cc|cc} 1 & 2 & 1 & 0 \\ 2 & 4 & 0 & 1 \end{array}\right] \xrightarrow{R_2 \to R_2 - 2R_1} \left[\begin{array}{cc|cc} 1 & 2 & 1 & 0 \\ 0 & 0 & -2 & 1 \end{array}\right]$$

At this point, we have a row of zeros on the left side of the augmented matrix. It is impossible to continue the process to obtain the identity matrix on the left. Because the RREF of B is not the identity matrix, we conclude that the matrix B is singular and **has no inverse**.

Definition 12. An **elementary matrix** is a square matrix obtained by applying a single elementary row operation to the identity matrix I_n.

Example (Row swapping). Consider the 3×3 identity matrix

$$I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

If we swap the first and the last rows, we obtain

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

Now let

$$A = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}.$$

Then

$$EA = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} = \begin{bmatrix} g & h & i \\ d & e & f \\ a & b & c \end{bmatrix}.$$

Hence, EA is the matrix obtained from A by swapping its first and last rows.

Example (Row scaling). Next, consider multiplying the third row of I_3 by a nonzero constant, say 5. Then the corresponding elementary matrix is

$$E_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 5 \end{bmatrix}.$$

For the same matrix

$$A = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix},$$

we have

$$E_2 A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 5 \end{bmatrix} \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} = \begin{bmatrix} a & b & c \\ d & e & f \\ 5g & 5h & 5i \end{bmatrix}.$$

Hence, $E_2 A$ is the matrix obtained from A by multiplying its third row by 5.

Example (Row addition). Finally, suppose we add 2 times the first row of I_3 to the second row. Then the elementary matrix is

$$E_3 = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

For the same matrix A, we have

$$E_3 A = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} = \begin{bmatrix} a & b & c \\ 2a+d & 2b+e & 2c+f \\ g & h & i \end{bmatrix}.$$

Thus, $E_3 A$ is obtained from A by adding 2 times its first row to its second row.

Remark. Every elementary matrix is invertible, since each elementary row operation is reversible. Moreover, the inverse of an elementary matrix is also an elementary matrix of the same type:

- The inverse of a row-swap matrix is itself.
- The inverse of a row-scaling matrix (by $k \neq 0$) is obtained by scaling the same row by $\frac{1}{k}$.
- The inverse of a row-addition matrix (adding k times one row to another) is obtained by adding $-k$ times that row instead.

Hence, all elementary matrices are nonsingular, and a square matrix is invertible if and only if it can be expressed as a finite product of elementary matrices.

Remark. Let $A \in \mathbb{F}^{n\times n}$ be invertible. Then there exists a sequence of elementary matrices $E_1, \ldots, E_k$ such that

$$[A \mid I_n] \xrightarrow{E_1} [E_1 A \mid E_1] \xrightarrow{E_2} [E_2 E_1 A \mid E_2 E_1] \xrightarrow{\cdots} \xrightarrow{E_k} [E_k \cdots E_1 A \mid E_k \cdots E_1] = [I_n \mid A^{-1}],$$

where

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

This illustrates that performing a sequence of elementary row operations on $[A \mid I_n]$ transforms A into the identity and simultaneously produces A^{-1} on the right.

Definition 13 (Row Equivalence). Two matrices $A, B \in \mathbb{F}^{m\times n}$ are said to be *row equivalent* if there exists a finite sequence of elementary row operations that transforms A into B.

Equivalently, there exist elementary matrices $E_1, \ldots, E_k$ such that

$$B = E_k \cdots E_1 A.$$

Theorem 5 (Invertibility, Singularity, and Row Equivalence). Let $A \in \mathbb{F}^{n\times n}$. The following statements are equivalent:

1. A is invertible.
2. A is row equivalent to the identity matrix I_n; that is, there exists a finite sequence of elementary matrices $E_1, \ldots, E_k$ such that
$$E_k \cdots E_1 A = I_n.$$
3. The augmented matrix $[A \mid I_n]$ can be row reduced to $[I_n \mid A^{-1}]$.
4. $\det(A) \neq 0$.
5. The homogeneous system $A\mathbf{x} = \mathbf{0}$ has only the trivial solution $\mathbf{x} = \mathbf{0}$.
6. The linear system $A\mathbf{x} = \mathbf{b}$ has a unique solution for every $\mathbf{b} \in \mathbb{F}^n$.

In particular, if any of the above fail, A is singular.

Proof. We prove the equivalence of the statements using only row operations and linear system arguments.

(1) ⇒ (2): If A is invertible, then Gaussian elimination can reduce A to the identity I_n via a finite sequence of elementary row operations. Each operation corresponds to an elementary matrix E_i, so

$$E_k \cdots E_1 A = I_n.$$

(2) ⇒ (3): Applying the same sequence of operations to the augmented matrix $[A \mid I_n]$ gives

$$[A \mid I_n] \xrightarrow{E_1,\ldots,E_k} [I_n \mid E_k \cdots E_1] = [I_n \mid A^{-1}].$$

(3) ⇒ (1): If $[A \mid I_n]$ can be reduced to $[I_n \mid B]$, then $AB = I_n$, so A is invertible and $B = A^{-1}$.

(1) $\Rightarrow$ (5): If A is invertible, the homogeneous system $A\mathbf{x} = \mathbf{0}$ has the unique solution $\mathbf{x} = \mathbf{0}$.

(5) $\Rightarrow$ (1): If $A\mathbf{x} = \mathbf{0}$ has only the trivial solution, then the columns of A are linearly independent. For a square matrix, linear independence implies full rank, so A is invertible.

(1) $\Rightarrow$ (6): If A is invertible, then for any $\mathbf{b}$ the system $A\mathbf{x} = \mathbf{b}$ has the unique solution $\mathbf{x} = A^{-1}\mathbf{b}$.

(6) $\Rightarrow$ (1): If $A\mathbf{x} = \mathbf{b}$ has a unique solution for every $\mathbf{b}$, then in particular $A\mathbf{x} = \mathbf{0}$ has only the trivial solution, which implies A is invertible.

Therefore, all statements are equivalent. If any statement fails, A is singular. $\square$

Chapter 2

Determinants

The determinant is a special number associated with a square matrix. It tells us important things about the matrix, like whether it's invertible, and it shows up in many areas of mathematics. In this chapter, we'll learn different ways to compute determinants and compare how much work each method requires.

2.1 Introduction to Determinants

Definition 14 (Determinant of a 2×2 Matrix). For a 2×2 matrix

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix},$$

the **determinant** of A, denoted $\det(A)$ or $|A|$, is defined as:

$$\det(A) = ad - bc.$$

Example (Simple Determinant). Let $A = \begin{bmatrix} 3 & 2 \\ 1 & 4 \end{bmatrix}$. Then

$$\det(A) = (3)(4) - (2)(1) = 12 - 2 = 10.$$

Remark. A 2×2 matrix is invertible if and only if its determinant is nonzero. If $\det(A) = 0$, then the matrix is singular (not invertible).

2.2 Computing Determinants by Gaussian Elimination

One way to compute the determinant of a larger matrix is to use Gaussian elimination to reduce it to upper triangular form. The key idea is that the determinant of an upper triangular matrix is just the product of the diagonal entries.

Theorem 6 (Determinant of Triangular Matrix). If A is an upper or lower triangular matrix, then $\det(A)$ is the product of the diagonal entries:

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

Theorem 7 (Effect of Row Operations on Determinant). Let A be an $n \times n$ matrix.

1. If B is obtained from A by swapping two rows, then $\det(B) = -\det(A)$.
2. If B is obtained from A by multiplying a row by a scalar c, then $\det(B) = c \cdot \det(A)$.
3. If B is obtained from A by adding a multiple of one row to another, then $\det(B) = \det(A)$.

To compute $\det(A)$ using Gaussian elimination:

1. Reduce A to upper triangular form U using row operations, keeping track of:
   - The number of row swaps (each multiplies by -1)
   - Any row scalings (each multiplies by the scaling factor)
2. Compute $\det(U)$ as the product of diagonal entries.
3. Adjust for the row operations: $\det(A) = (-1)^k \cdot (\text{product of scaling factors})^{-1} \cdot \det(U)$.

Example (Determinant via Gaussian Elimination). Find $\det(A)$ where $A = \begin{bmatrix} 2 & 1 & 3 \\ 4 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix}$.

We perform Gaussian elimination:

$$\begin{aligned} A &= \begin{bmatrix} 2 & 1 & 3 \\ 4 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix} \\ &\xrightarrow{R_2 \to R_2 - 2R_1} \begin{bmatrix} 2 & 1 & 3 \\ 0 & 0 & -5 \\ 1 & 1 & 2 \end{bmatrix} \\ &\xrightarrow{R_3 \to R_3 - \frac{1}{2}R_1} \begin{bmatrix} 2 & 1 & 3 \\ 0 & 0 & -5 \\ 0 & \frac{1}{2} & \frac{1}{2} \end{bmatrix} \\ &\xrightarrow{R_2 \leftrightarrow R_3} \begin{bmatrix} 2 & 1 & 3 \\ 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & -5 \end{bmatrix} \end{aligned}$$

We performed one row swap, and the final matrix is upper triangular. The determinant of the triangular matrix is:

$$\det(U) = (2)\left(\frac{1}{2}\right)(-5) = -5.$$

Since we did one row swap, we multiply by -1:

$$\det(A) = (-1)^1 \cdot (-5) = 5.$$

2.2.1 Complexity Analysis: Gaussian Elimination Method

For an $n \times n$ matrix using Gaussian elimination:

- **Multiplications/Divisions:** Approximately $\frac{n^3}{3}$ operations to reduce to upper triangular form, plus n multiplications for the diagonal product. Total: roughly $\frac{n^3}{3}+n \approx \frac{n^3}{3}$.
- **Additions/Subtractions:** Approximately $\frac{n^3}{3}$ operations.

This method is quite efficient compared to other methods we'll see, especially for large matrices.

DIY. Derive the exact formulas for the number of operations in Gaussian elimination.

To eliminate the first column below the diagonal, we need to eliminate $n-1$ rows. For row i (where $i \geq 2$), we compute the multiplier $m_{i1} = a_{i1}/a_{11}$ (1 division), then update $n-1$ entries: $(a_{i2} - m_{i1}a_{12})$, $(a_{i3} - m_{i1}a_{13})$, $\dots$, $(a_{in} - m_{i1}a_{1n})$. Each update requires 1 multiplication and 1 subtraction. So row i needs $1 + (n-1) = n$ operations for elimination, giving $(n-1) \cdot n = n(n-1)$ operations total.

For the second column, we eliminate $n-2$ rows, each requiring $(n-2)$ updates, giving $(n-2)(n-1)$ operations. Continuing this pattern, the total number of multiplications/divisions is:

$$\sum_{k=1}^{n-1} k(k+1) = \sum_{k=1}^{n-1}(k^2+k) = \frac{(n-1)n(2n-1)}{6} + \frac{(n-1)n}{2} = \frac{n^3}{3} - \frac{n}{3}.$$

The number of additions/subtractions is the same. Verify this formula and compare it with the asymptotic estimate $\frac{n^3}{3}$.

2.3 Cofactor Expansion

Cofactor expansion (also called Laplace expansion) is another method to compute determinants. It works by breaking down the determinant into smaller parts.

Definition 15 (Minor and Cofactor). Let A be an $n \times n$ matrix. The **minor** M_{ij} is the determinant of the $(n-1) \times (n-1)$ submatrix obtained by deleting the i-th row and j-th column from A.

The **cofactor** C_{ij} is defined as:

$$C_{ij} = (-1)^{i+j} M_{ij}.$$

Theorem 8 (Cofactor Expansion). For any $n \times n$ matrix A, the determinant can be computed by expanding along any row or column:

$$\det(A) = \sum_{j=1}^{n} a_{ij} C_{ij} \quad \text{(expansion along row } i\text{)} \tag{2.1}$$

$$\det(A) = \sum_{i=1}^{n} a_{ij} C_{ij} \quad \text{(expansion along column } j\text{)} \tag{2.2}$$

Example (Cofactor Expansion Along First Row). Find $\det(A)$ where $A = \begin{bmatrix} 2 & 1 & 3 \\ 4 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix}$.

Expanding along the first row:

$$\det(A) = a_{11}C_{11} + a_{12}C_{12} + a_{13}C_{13} \tag{2.3}$$

$$= 2 \cdot C_{11} + 1 \cdot C_{12} + 3 \cdot C_{13}. \tag{2.4}$$

Now compute each cofactor:

$$C_{11} = (-1)^{1+1} \det \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} = 1 \cdot (4 - 1) = 3, \tag{2.5}$$

$$C_{12} = (-1)^{1+2} \det \begin{bmatrix} 4 & 1 \\ 1 & 2 \end{bmatrix} = -1 \cdot (8 - 1) = -7, \tag{2.6}$$

$$C_{13} = (-1)^{1+3} \det \begin{bmatrix} 4 & 2 \\ 1 & 1 \end{bmatrix} = 1 \cdot (4 - 2) = 2. \tag{2.7}$$

Therefore:

$$\det(A) = 2(3) + 1(-7) + 3(2) = 6 - 7 + 6 = 5.$$

2.3.1 Complexity Analysis: Cofactor Expansion

For an $n \times n$ matrix using cofactor expansion:

- We need to compute n cofactors, each requiring the determinant of an $(n-1) \times (n-1)$ matrix.
- This leads to a recursive process. The total number of operations grows roughly as $n!$ (factorial).
- **Multiplications:** Approximately $n!$ operations.
- **Additions:** Approximately $n!$ operations.

This is much slower than Gaussian elimination for large n because factorial grows very rapidly.

DIY. Derive the exact recurrence relation for the operation count in cofactor expansion.

Let $M(n)$ be the number of multiplications needed to compute the determinant of an $n \times n$ matrix using cofactor expansion. When expanding along a row, we compute n cofactors, each requiring $M(n-1)$ multiplications, plus n multiplications to multiply the entries by their cofactors. This gives:

$$M(n) = n \cdot M(n-1) + n, \quad M(1) = 0.$$

Show that this recurrence has the solution $M(n) = n! \sum_{k=1}^{n} \frac{1}{k!}$, which asymptotically approaches $e \cdot n!$ as n grows. What is the corresponding recurrence for additions?

2.4 The Permutation Method

The permutation method uses the definition of determinant directly in terms of permutations of the indices.

Definition 16 (Permutation). A **permutation** of $\{1, 2, \ldots, n\}$ is a rearrangement of these numbers. A permutation σ is called an **even permutation** if it can be obtained by an even number of swaps, and an **odd permutation** if it requires an odd number of swaps. The **sign** of a permutation, denoted $\operatorname{sgn}(\sigma)$, is $+1$ for even permutations and -1 for odd permutations.

Theorem 9 (Determinant via Permutations). For an $n \times n$ matrix $A = (a_{ij})$, the determinant is given by:

$$\det(A) = \sum_{\sigma} \operatorname{sgn}(\sigma) \cdot a_{1\sigma(1)} \cdot a_{2\sigma(2)} \cdot \cdots \cdot a_{n\sigma(n)},$$

where the sum is over all permutations σ of $\{1, 2, \ldots, n\}$.

Example (3×3 Determinant via Permutations). For a 3×3 matrix $A = (a_{ij})$, there are $3! = 6$ permutations:

Permutation σ	Sign	Product
$(1, 2, 3)$	$+1$	$a_{11}a_{22}a_{33}$
$(1, 3, 2)$	-1	$-a_{11}a_{23}a_{32}$
$(2, 1, 3)$	-1	$-a_{12}a_{21}a_{33}$
$(2, 3, 1)$	$+1$	$a_{12}a_{23}a_{31}$
$(3, 1, 2)$	$+1$	$a_{13}a_{21}a_{32}$
$(3, 2, 1)$	-1	$-a_{13}a_{22}a_{31}$

So:

$$\det(A) = a_{11}a_{22}a_{33} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31}.$$

This is the familiar "rule of Sarrus" for 3×3 matrices (though it doesn't extend to larger matrices).

2.4.1 Complexity Analysis: Permutation Method

For an $n \times n$ matrix using the permutation method:

- We need to consider all $n!$ permutations.
- For each permutation, we compute the sign (which takes at most n swaps to determine) and multiply n entries.
- **Multiplications:** $n! \cdot (n-1)$ operations (one product of n terms per permutation requires $n-1$ multiplications).
- **Additions:** $n! - 1$ additions (summing $n!$ terms).
- **Sign computations:** Up to $n! \cdot n$ operations in worst case.

This method has the worst complexity among the three, growing as $n!$, which becomes impractical very quickly.

> **DIY.** Derive the exact operation count formulas for the permutation method.
> For each of the $n!$ permutations σ of $\{1, 2, \ldots, n\}$, we need to:
>
> 1. Compute the sign $\mathrm{sgn}(\sigma)$. One way is to count inversions: pairs (i, j) with $i < j$ but $\sigma(i) > \sigma(j)$. This takes at most $\binom{n}{2} = \frac{n(n-1)}{2}$ comparisons.
> 2. Compute the product $a_{1\sigma(1)} \cdot a_{2\sigma(2)} \cdot \cdots \cdot a_{n\sigma(n)}$, which requires $n-1$ multiplications.
>
> Show that the total number of multiplications is exactly $n! \cdot (n-1)$, and the number of additions is exactly $n! - 1$. How does the sign computation cost compare? Can you optimize the sign computation?

2.5 Comparison of Methods

Let's compare the three methods as n grows. We'll count the approximate number of operations:

Method	Multiplications	Additions	Total Operations
Gaussian Elimination	$\approx \frac{n^3}{3}$	$\approx \frac{n^3}{3}$	$\approx \frac{2n^3}{3}$
Cofactor Expansion	$\approx n!$	$\approx n!$	$\approx 2n!$
Permutation Method	$\approx n! \cdot (n-1)$	$\approx n!$	$\approx n! \cdot n$

Table 2.1: Asymptotic complexity of determinant methods

Let's look at some specific values:

n	Gaussian (ops)	Cofactor (ops)	Permutation (ops)	Ratio (G:C:P)
2	≈ 5	≈ 4	≈ 6	1:1:1
3	≈ 18	≈ 12	≈ 24	1:1:1
4	≈ 43	≈ 96	≈ 600	1:2:14
5	≈ 84	≈ 960	$\approx 7,200$	1:11:86
10	≈ 667	$\approx 7.3 \times 10^6$	$\approx 4.0 \times 10^8$	1:10^4:6×10^5

Table 2.2: Approximate operation counts for different n

Remark. As we can see, Gaussian elimination is by far the most efficient method for larger matrices. The factorial growth of cofactor expansion and the permutation method makes them impractical for $n > 4$ or $n > 5$. However, cofactor expansion is sometimes useful for theoretical purposes or small matrices, and the permutation definition is important for understanding what the determinant really represents.

Example (Practical Example). To compute the determinant of a 10×10 matrix:

- **Gaussian elimination:** About 667 operations, very reasonable.
- **Cofactor expansion:** About 7.3 million operations, takes a long time.
- **Permutation method:** About 400 million operations, extremely slow, essentially impractical.

This is why Gaussian elimination is the method of choice in practice!

Chapter 3

Vector Spaces

The operations of addition and scalar multiplication are used in many diverse contexts in mathematics. A general theory of mathematical systems involving these operations is applicable to many areas, leading to the concept of vector spaces.

3.1 Definition and Examples

We begin by formalizing the definition of a vector space, generalizing the properties found in Euclidean space $\mathbb{R}^n$.

3.1.1 The Vector Space Axioms

Definition 17 (Vector Space). Let V be a set on which the operations of addition and scalar multiplication are defined. The set V, together with these operations, is a **vector space** if the following axioms are satisfied for all $\mathbf{x}, \mathbf{y}, \mathbf{z} \in V$ and scalars α, β:

1. $\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}$ (Commutativity)
2. $(\mathbf{x} + \mathbf{y}) + \mathbf{z} = \mathbf{x} + (\mathbf{y} + \mathbf{z})$ (Associativity)
3. There exists an element $\mathbf{0} \in V$ such that $\mathbf{x} + \mathbf{0} = \mathbf{x}$ for each $\mathbf{x} \in V$. (Additive Identity)
4. For each $\mathbf{x} \in V$, there exists an element $-\mathbf{x} \in V$ such that $\mathbf{x} + (-\mathbf{x}) = \mathbf{0}$. (Additive Inverse)
5. $\alpha(\mathbf{x} + \mathbf{y}) = \alpha\mathbf{x} + \alpha\mathbf{y}$ (Distributivity over Vector Addition)
6. $(\alpha + \beta)\mathbf{x} = \alpha\mathbf{x} + \beta\mathbf{x}$ (Distributivity over Scalar Addition)
7. $(\alpha\beta)\mathbf{x} = \alpha(\beta\mathbf{x})$ (Associativity of Scalar Multiplication)
8. $1\mathbf{x} = \mathbf{x}$ (Multiplicative Identity)

Remark (Closure Properties). Implicit in the definition are the closure properties, which are often checked first:

- **C1 (Closure under scalar multiplication):** If $\mathbf{x} \in V$ and α is a scalar, then $\alpha\mathbf{x} \in V$.

- **C2 (Closure under addition):** If $\mathbf{x}, \mathbf{y} \in V$, then $\mathbf{x} + \mathbf{y} \in V$.

Example (The Vector Space $\mathbb{R}^n$). The most familiar example is $\mathbb{R}^n$, the set of all n-tuples of real numbers, with component-wise addition and scalar multiplication:

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \end{bmatrix} x_n + \begin{bmatrix} y_1 \\ y_2 \\ \vdots \end{bmatrix} y_n = \begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \end{bmatrix} x_n + y_n, \quad c \begin{bmatrix} x_1 \\ x_2 \\ \vdots \end{bmatrix} x_n = \begin{bmatrix} cx_1 \\ cx_2 \\ \vdots \end{bmatrix} cx_n.$$

All eight axioms are satisfied (they follow from the properties of real numbers).

Example (The Vector Space of Polynomials P_n). Let P_n be the set of all polynomials of degree less than n:

$$P_n = \{a_0 + a_1 x + a_2 x^2 + \cdots + a_{n-1} x^{n-1} \mid a_0, a_1, \ldots, a_{n-1} \in \mathbb{R}\}.$$

Addition and scalar multiplication are defined as:

$$(a_0 + a_1 x + \cdots + a_{n-1} x^{n-1}) + (b_0 + b_1 x + \cdots + b_{n-1} x^{n-1}) = (a_0 + b_0) + (a_1 + b_1)x + \cdots + (a_{n-1} + b_{n-1}) \tag{3.1}$$

$$c(a_0 + a_1 x + \cdots + a_{n-1} x^{n-1}) = ca_0 + ca_1 x + \cdots + ca_{n-1} x^{n-1}. \tag{3.2}$$

The zero vector is the zero polynomial $0 + 0x + \cdots + 0x^{n-1}$.

Example (The Vector Space of Matrices $\mathbb{R}^{m\times n}$). The set of all $m \times n$ matrices with real entries forms a vector space with the usual matrix addition and scalar multiplication. For example, in $\mathbb{R}^{2\times 2}$:

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} + \begin{bmatrix} e & f \\ g & h \end{bmatrix} = \begin{bmatrix} a+e & b+f \\ c+g & d+h \end{bmatrix}, \quad k \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} ka & kb \\ kc & kd \end{bmatrix}.$$

The zero vector is the zero matrix $\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$.

Theorem 10 (Fundamental Properties). If V is a vector space and $\mathbf{x}$ is any element of V, then:

1. $0\mathbf{x} = \mathbf{0}$
2. $\mathbf{x} + \mathbf{y} = \mathbf{0}$ implies that $\mathbf{y} = -\mathbf{x}$ (Uniqueness of additive inverse).
3. $(-1)\mathbf{x} = -\mathbf{x}$

Proof. To prove (iii), note that:

$$\mathbf{x} + (-1)\mathbf{x} = 1\mathbf{x} + (-1)\mathbf{x} = (1 + (-1))\mathbf{x} = 0\mathbf{x} = \mathbf{0}$$

By the uniqueness of the additive inverse shown in (ii), it follows that $(-1)\mathbf{x} = -\mathbf{x}$. □

3.2 Subspaces

It is often possible to form a vector space by taking a subset of a larger vector space.

3.2.1 Definition and Tests

Definition 18 (Subspace). If S is a nonempty subset of a vector space V, then S is said to be a **subspace** of V if it satisfies the following conditions:

i. $\alpha\mathbf{x} \in S$ whenever $\mathbf{x} \in S$ for any scalar α.

ii. $\mathbf{x} + \mathbf{y} \in S$ whenever $\mathbf{x} \in S$ and $\mathbf{y} \in S$.

A subspace is a subset that is closed under the operations of the parent vector space.

Remark. Every subspace must contain the zero vector $\mathbf{0}$ (take $\alpha = 0$ in condition (i) with any $\mathbf{x} \in S$). The set $\{\mathbf{0}\}$ is the **zero subspace**.

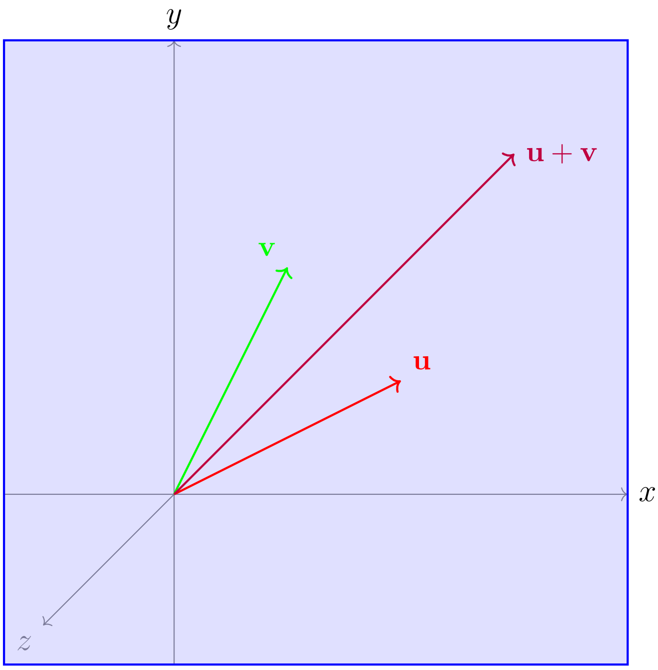
Example (Checking if a Set is a Subspace: Step by Step). Determine whether $S = \{(x, y, z) \in \mathbb{R}^3 \mid z = 0\}$ is a subspace of $\mathbb{R}^3$.

Step 1: Check if $\mathbf{0} \in S$: The zero vector is $(0, 0, 0)$. Since $z = 0$, we have $(0, 0, 0) \in S$. ✓

Step 2: Check closure under scalar multiplication: Let $\mathbf{u} = (x, y, 0) \in S$ and α be a scalar. Then $\alpha\mathbf{u} = (\alpha x, \alpha y, \alpha \cdot 0) = (\alpha x, \alpha y, 0)$. Since the third component is 0, we have $\alpha\mathbf{u} \in S$. ✓

Step 3: Check closure under addition: Let $\mathbf{u} = (x_1, y_1, 0) \in S$ and $\mathbf{v} = (x_2, y_2, 0) \in S$. Then $\mathbf{u} + \mathbf{v} = (x_1 + x_2, y_1 + y_2, 0 + 0) = (x_1 + x_2, y_1 + y_2, 0)$. Since the third component is 0, we have $\mathbf{u} + \mathbf{v} \in S$. ✓

Therefore, S is a subspace of $\mathbb{R}^3$. Geometrically, S is the xy-plane.



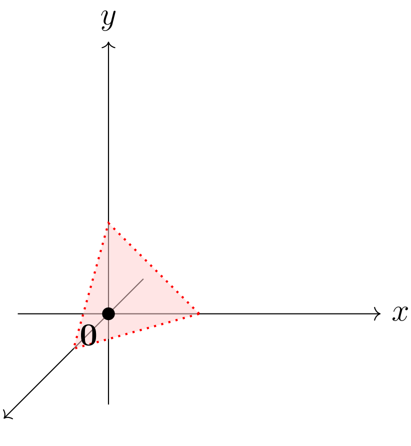
The xy-plane is a subspace of $\mathbb{R}^3$

Figure 3.1: The subspace $S = \{(x, y, z) \in \mathbb{R}^3 \mid z = 0\}$ is the xy-plane. Vectors in S remain in S under addition and scalar multiplication.

Example (A Set That is NOT a Subspace). Determine whether $T = \{(x, y, z) \in \mathbb{R}^3 \mid x + y + z = 1\}$ is a subspace of $\mathbb{R}^3$.

Step 1: Check if $\mathbf{0} \in T$: The zero vector is $(0, 0, 0)$. Checking: $0 + 0 + 0 = 0 \neq 1$. So $(0, 0, 0) \notin T$.

Since T does not contain the zero vector, T is *not* a subspace. (In fact, T is a plane that does not pass through the origin.)



The plane $x + y + z = 1$ is NOT a subspace (does not contain $\mathbf{0}$)

Figure 3.2: The set $T = \{(x, y, z) \in \mathbb{R}^3 \mid x + y + z = 1\}$ is not a subspace because it doesn't contain the zero vector.

Example (Matrix Subspaces: Skew-Symmetric Matrices). Let $S = \{A \in \mathbb{R}^{2\times 2} \mid a_{12} = -a_{21}\}$. Show that S is a subspace of $\mathbb{R}^{2\times 2}$.

A matrix $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is in S if $b = -c$, so $A = \begin{bmatrix} a & b \\ -b & d \end{bmatrix}$.

Step 1: Check if $\mathbf{0} \in S$: The zero matrix is $\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$. Since $0 = -0$, we have $\mathbf{0} \in S$. ✓

Step 2: Check closure under scalar multiplication: Let $A = \begin{bmatrix} a & b \\ -b & d \end{bmatrix} \in S$ and α be a scalar. Then:

$$\alpha A = \alpha \begin{bmatrix} a & b \\ -b & d \end{bmatrix} = \begin{bmatrix} \alpha a & \alpha b \\ -\alpha b & \alpha d \end{bmatrix}.$$

Since the $(1, 2)$ entry is αb and the $(2, 1)$ entry is $-\alpha b$, we have $(\alpha A)_{12} = -(\alpha A)_{21}$. So $\alpha A \in S$. ✓

Step 3: Check closure under addition: Let $A = \begin{bmatrix} a_1 & b_1 \\ -b_1 & d_1 \end{bmatrix} \in S$ and $B = \begin{bmatrix} a_2 & b_2 \\ -b_2 & d_2 \end{bmatrix} \in S$. Then:

$$A + B = \begin{bmatrix} a_1 + a_2 & b_1 + b_2 \\ -(b_1 + b_2) & d_1 + d_2 \end{bmatrix}.$$

Since $(A + B)_{12} = b_1 + b_2 = -(b_1 + b_2) = -(A + B)_{21}$, we have $A + B \in S$. ✓

Therefore, S is a subspace. (These are called skew-symmetric matrices.)

Example (Polynomial Subspace: Step by Step). Let $W = \{p(x) \in P_3 \mid p(0) = 0\}$. Show that W is a subspace of P_3.

Step 1: Check if $\mathbf{0} \in W$: The zero polynomial is $0(x) = 0$. Since $0(0) = 0$, we have $0(x) \in W$. ✓

Step 2: Check closure under scalar multiplication: Let $p(x) \in W$ (so $p(0) = 0$) and α be a scalar. Then $(\alpha p)(0) = \alpha \cdot p(0) = \alpha \cdot 0 = 0$. So $\alpha p \in W$. ✓

Step 3: Check closure under addition: Let $p(x), q(x) \in W$ (so $p(0) = 0$ and $q(0) = 0$). Then $(p + q)(0) = p(0) + q(0) = 0 + 0 = 0$. So $p + q \in W$. ✓

Therefore, W is a subspace of P_3.

In fact, if $p(x) = a_0 + a_1x + a_2x^2 + a_3x^3$, then $p(0) = a_0 = 0$, so $p(x) = a_1x + a_2x^2 + a_3x^3$. This subspace consists of all polynomials with no constant term.

DIY. If you have a linearly independent set $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ in a vector space V of dimension n where $k < n$, show that you can always extend it to a basis $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k, \mathbf{v}_{k+1}, \ldots, \mathbf{v}_n\}$ of V.

Hint: Start with a known basis $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ of V. Consider the set $\{\mathbf{v}_1, \ldots, \mathbf{v}_k, \mathbf{u}_1, \ldots, \mathbf{u}_n\}$. This set spans V but is linearly dependent. Show how to remove vectors from $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ while keeping the span of V until you have exactly n linearly independent vectors.

3.3 Linear Independence and Spanning

3.3.1 Linear Independence

Definition 19 (Linear Independence). A set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ in a vector space V is **linearly independent** if the only solution to the equation

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n = \mathbf{0}$$

is $c_1 = c_2 = \cdots = c_n = 0$. Otherwise, the set is **linearly dependent**.

Vectors are linearly independent if none of them can be written as a combination of the others.

Example (Testing Linear Independence in $\mathbb{R}^2$: Step by Step). Determine whether $\mathbf{v}_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ and $\mathbf{v}_2 = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$ are linearly independent.

We need to solve $c_1\mathbf{v}_1 + c_2\mathbf{v}_2 = \mathbf{0}$:

$$c_1 \begin{bmatrix} 1 \\ 2 \end{bmatrix} + c_2 \begin{bmatrix} 3 \\ 1 \end{bmatrix} = \begin{bmatrix} c_1 + 3c_2 \\ 2c_1 + c_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

This gives the system:

$$\begin{cases} c_1 + 3c_2 = 0 \\ 2c_1 + c_2 = 0 \end{cases}$$

From the first equation: $c_1 = -3c_2$. Substituting into the second: $2(-3c_2) + c_2 = -6c_2 + c_2 = -5c_2 = 0$, so $c_2 = 0$. Then $c_1 = 0$.

Since the only solution is $c_1 = c_2 = 0$, the vectors are linearly independent.

Example (Linear Dependence in $\mathbb{R}^2$: Detailed Example). Determine whether $\mathbf{v}_1 = \begin{bmatrix} 2 \\ 4 \end{bmatrix}$ and $\mathbf{v}_2 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ are linearly independent.

We solve $c_1\mathbf{v}_1 + c_2\mathbf{v}_2 = \mathbf{0}$:

$$c_1 \begin{bmatrix} 2 \\ 4 \end{bmatrix} + c_2 \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 2c_1 + c_2 \\ 4c_1 + 2c_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

This gives:

$$\begin{cases} 2c_1 + c_2 = 0 \\ 4c_1 + 2c_2 = 0 \end{cases}$$

From the first equation: $c_2 = -2c_1$. The second equation is $4c_1 + 2(-2c_1) = 4c_1 - 4c_1 = 0$, which is automatically satisfied.

So we can take $c_1 = 1$, then $c_2 = -2$. This gives a nontrivial solution: $1 \cdot \mathbf{v}_1 + (-2) \cdot \mathbf{v}_2 = \mathbf{0}$, or $\mathbf{v}_1 = 2\mathbf{v}_2$.

Therefore, the vectors are linearly dependent. In fact, $\mathbf{v}_1$ is a scalar multiple of $\mathbf{v}_2$.

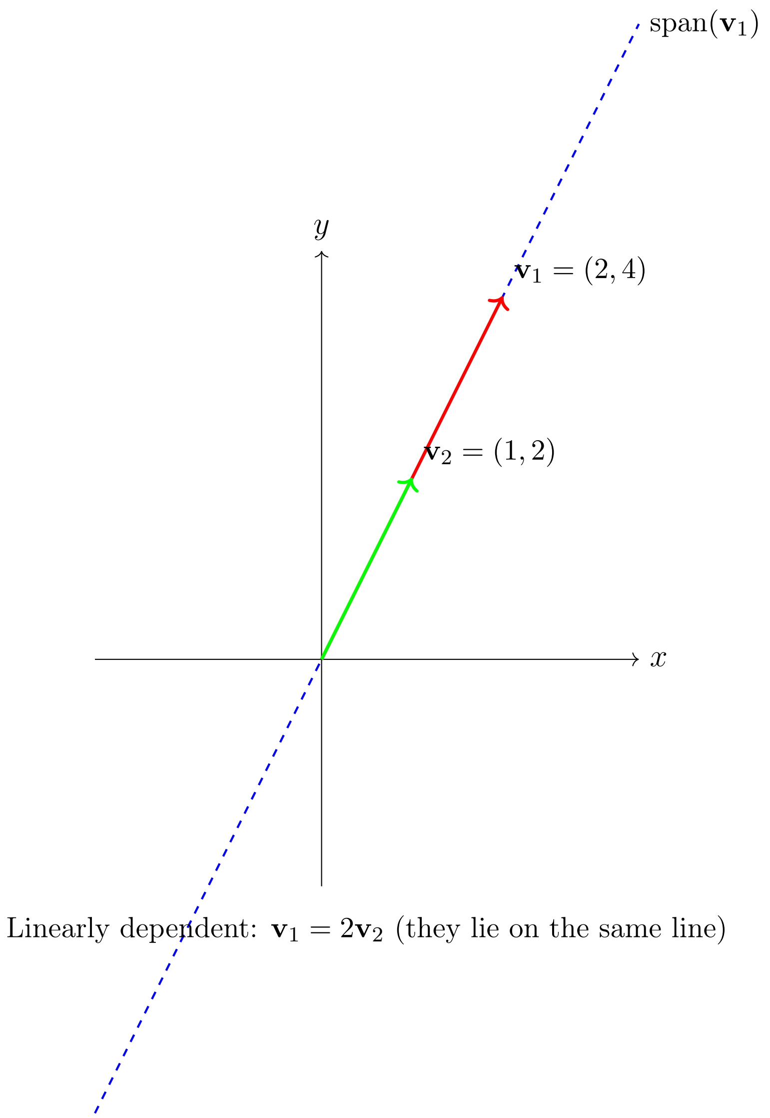


Figure 3.3: Linearly dependent vectors in $\mathbb{R}^2$. When vectors are multiples of each other, they lie on the same line through the origin.

Example (Linear Independence in $\mathbb{R}^3$: Using a Matrix). Determine whether the vectors $\mathbf{v}_1 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix}$, and $\mathbf{v}_3 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ are linearly independent.

We form the matrix with these vectors as columns and row reduce:

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

We want to solve $A\mathbf{c} = \mathbf{0}$ to check for linear independence. Row reducing:

$$\begin{bmatrix} 1 & 0 & 1 \\ 2 & 1 & 1 \\ 3 & 2 & 1 \end{bmatrix} \xrightarrow{R_2 \to R_2 - 2R_1} \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & -1 \\ 3 & 2 & 1 \end{bmatrix} \xrightarrow{R_3 \to R_3 - 3R_1} \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & -1 \\ 0 & 2 & -2 \end{bmatrix}$$

$$\xrightarrow{R_3 \to R_3 - 2R_2} \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{bmatrix}.$$

We have a row of zeros, so there is a free variable. The system has a nontrivial solution, meaning the vectors are linearly dependent.

In fact, from the reduced form, we can see that column 3 = column 1 - column 2, so $\mathbf{v}_3 = \mathbf{v}_1 - \mathbf{v}_2$.

3.3.2 Spanning Sets

Definition 20 (Spanning Set). A set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ in a vector space V **spans** V if every vector in V can be written as a linear combination of $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$. That is, for any $\mathbf{w} \in V$, there exist scalars $c_1, c_2, \ldots, c_n$ such that

$$\mathbf{w} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n.$$

We write $\text{span}(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n) = V$.

Example (Spanning $\mathbb{R}^2$: Detailed Example). Show that $\mathbf{v}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $\mathbf{v}_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$ span $\mathbb{R}^2$.

We need to show that any vector $\begin{bmatrix} a \\ b \end{bmatrix} \in \mathbb{R}^2$ can be written as $c_1\mathbf{v}_1 + c_2\mathbf{v}_2$ for some c_1, c_2.

That is, we need to solve:

$$c_1 \begin{bmatrix} 1 \\ 1 \end{bmatrix} + c_2 \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} c_1 + c_2 \\ c_1 - c_2 \end{bmatrix} = \begin{bmatrix} a \\ b \end{bmatrix}.$$

This gives the system:

$$\begin{cases} c_1 + c_2 = a \\ c_1 - c_2 = \quad b \end{cases}$$

Adding the two equations: $2c_1 = a + b$, so $c_1 = \frac{a+b}{2}$. Subtracting the second from the first: $2c_2 = a - b$, so $c_2 = \frac{a-b}{2}$.

Since we can always find such c_1 and c_2 for any a, b, the vectors span $\mathbb{R}^2$.

For example, to write $\begin{bmatrix} 3 \\ 1 \end{bmatrix}$ as a combination: $c_1 = \frac{3+1}{2} = 2$ and $c_2 = \frac{3-1}{2} = 1$, so $\begin{bmatrix} 3 \\ 1 \end{bmatrix} = 2\begin{bmatrix} 1 \\ 1 \end{bmatrix} + 1\begin{bmatrix} 1 \\ -1 \end{bmatrix}$.

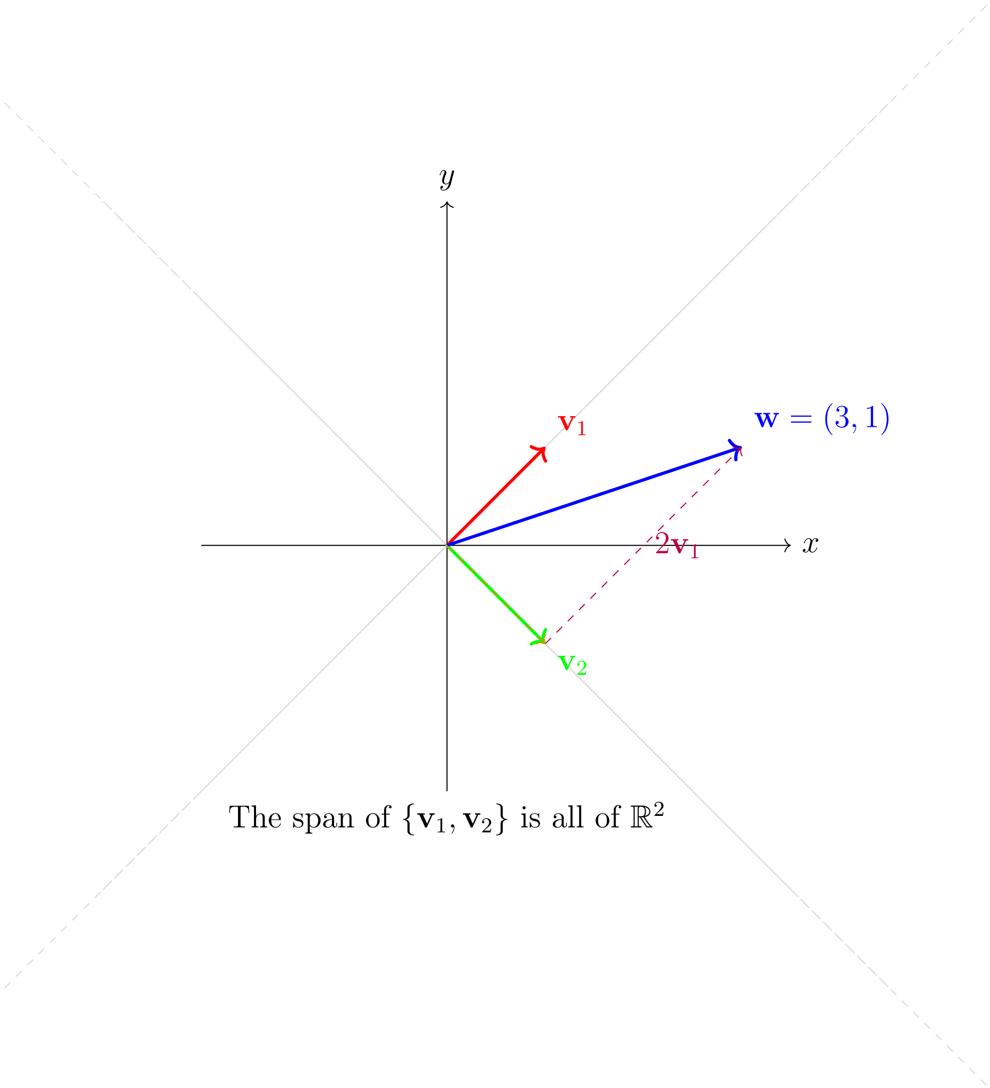


Figure 3.4: Two linearly independent vectors in $\mathbb{R}^2$ span the entire plane. The grid shows how any vector can be reached using combinations of $\mathbf{v}_1$ and $\mathbf{v}_2$.

Example (Spanning a Plane in $\mathbb{R}^3$). Let $\mathbf{v}_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$ and $\mathbf{v}_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$. Find $\text{span}(\mathbf{v}_1, \mathbf{v}_2)$.

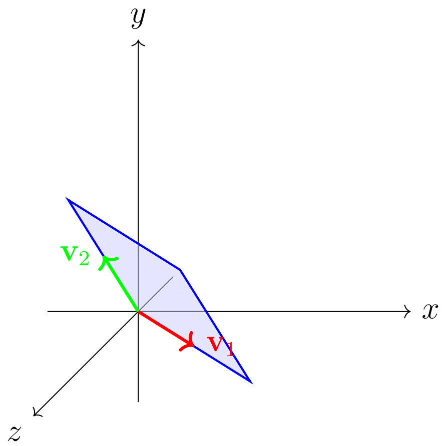
The span consists of all vectors of the form $c_1\mathbf{v}_1 + c_2\mathbf{v}_2$:

$$c_1 \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} + c_2 \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \\ c_1 + c_2 \end{bmatrix}.$$

This is a plane in $\mathbb{R}^3$ through the origin. To see this, note that if $\mathbf{w} = \begin{bmatrix} x \\ y \\ z \end{bmatrix}$ is in the span, then we need:

$$\begin{cases} c_1 & = x \\ & c_2 = & y \\ c_1 + c_2 = & & z \end{cases}$$

Substituting: $z = x + y$. So the span is the plane $\{(x, y, z) \in \mathbb{R}^3 \mid z = x + y\}$.



The span of $\{\mathbf{v}_1, \mathbf{v}_2\}$ is a plane through the origin

Figure 3.5: Two linearly independent vectors in $\mathbb{R}^3$ span a plane (not all of $\mathbb{R}^3$).

3.4 Basis and Dimension

Definition 21 (Basis). The vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ form a **basis** for a vector space V if and only if:

(i) $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ are linearly independent.

(ii) $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ span V.

A basis gives us the "right amount" of vectors: enough to span the space, but not so many that they're redundant.

Example (The Standard Basis for $\mathbb{R}^3$: Complete Verification). The vectors $\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$, $\mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$, $\mathbf{e}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$ form the **standard basis** for $\mathbb{R}^3$.

Linear Independence: If $c_1\mathbf{e}_1 + c_2\mathbf{e}_2 + c_3\mathbf{e}_3 = \mathbf{0}$, then:

$$c_1 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + c_2 \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + c_3 \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix},$$

which immediately gives $c_1 = c_2 = c_3 = 0$. So they are linearly independent.

Spanning: Any vector $\begin{bmatrix} x \\ y \\ z \end{bmatrix} \in \mathbb{R}^3$ can be written as:

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = x \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + y \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + z \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = x\mathbf{e}_1 + y\mathbf{e}_2 + z\mathbf{e}_3.$$

Therefore, $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ is a basis.

Example (Finding a Basis: Step by Step Example). Find a basis for the subspace $W = \{(x, y, z) \in \mathbb{R}^3 \mid x + y + z = 0\}$.

First, note that if $(x, y, z) \in W$, then $z = -x - y$. So any vector in W has the form:

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} x \\ y \\ -x - y \end{bmatrix} = x \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} + y \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}.$$

This shows that $W = \operatorname{span}\left(\begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix} \right)$.

Now check linear independence:

$$c_1 \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} + c_2 \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \\ -c_1 - c_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

gives $c_1 = 0$, $c_2 = 0$. So the vectors are linearly independent.

Therefore, $\left\{ \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix} \right\}$ is a basis for W.

Example (Basis for Polynomial Space: Detailed). Show that $\{1, x, x^2\}$ is a basis for P_2 (polynomials of degree at most 2).

Linear Independence: If $c_0 \cdot 1 + c_1 x + c_2 x^2 = 0$ (the zero polynomial), then this must hold for all x. Setting $x = 0$: $c_0 = 0$. Differentiating and setting $x = 0$: $c_1 = 0$. Differentiating again: $c_2 = 0$. So the only solution is $c_0 = c_1 = c_2 = 0$.

Spanning: Any polynomial in P_2 has the form $p(x) = a_0 + a_1 x + a_2 x^2$, which is exactly $a_0 \cdot 1 + a_1 x + a_2 x^2$.

Therefore, $\{1, x, x^2\}$ is a basis for P_2.

Definition 22 (Dimension). Let V be a vector space. If V has a basis consisting of n vectors, we say that V has **dimension** n, written $\dim(V) = n$. The subspace $\{\mathbf{0}\}$ has dimension 0.

A vector space V is **finite dimensional** if there is a finite set of vectors that spans V; otherwise, V is **infinite dimensional**.

Remark. A fundamental result (which we state without proof) is that all bases of a finite-dimensional vector space have the same number of elements. This ensures dimension is well-defined.

Example (Dimensions of Common Spaces).

- $\dim(\mathbb{R}^n) = n$ (standard basis has n vectors)
- $\dim(P_n) = n + 1$ (basis: $\{1, x, x^2, \ldots, x^n\}$)
- $\dim(\mathbb{R}^{m\times n}) = mn$ (standard basis: matrices with one 1 and zeros elsewhere)
- From the previous example, $\dim(W) = 2$ where $W = \{(x, y, z) \in \mathbb{R}^3 \mid x+y+z = 0\}$

Theorem 11 (Basis Test in Finite Dimensions). Let V be an n-dimensional vector space, and let $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ be a set of n vectors in V. The following are equivalent:

1. $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is a basis for V.
2. $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is linearly independent.
3. $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ spans V.

This theorem is very useful: if you have exactly $\dim(V)$ vectors, you only need to check one of linear independence or spanning (not both).

Example (Using the Basis Test Theorem). In $\mathbb{R}^3$ (dimension 3), consider the vectors:

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

To check if they form a basis, we only need to check linear independence (since we have 3 vectors and $\dim(\mathbb{R}^3) = 3$).

Form the matrix and row reduce:

$$\begin{bmatrix}1&1&0\\1&0&1\\0&1&1\end{bmatrix} \xrightarrow{R_2\to R_2-R_1} \begin{bmatrix}1&1&0\\0&-1&1\\0&1&1\end{bmatrix} \xrightarrow{R_3\to R_3+R_2} \begin{bmatrix}1&1&0\\0&-1&1\\0&0&2\end{bmatrix}.$$

Since we get three pivots, the vectors are linearly independent. By the Basis Test Theorem, they form a basis for $\mathbb{R}^3$.

Theorem 12 (Spanning Sets and Linear Dependence). If $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is a spanning set for a vector space V, then any collection of m vectors in V, where $m > n$, is linearly dependent.

Proof. Let $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ span V, and let $\{\mathbf{w}_1, \ldots, \mathbf{w}_m\}$ be any m vectors in V with $m > n$.

Since $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ spans V, each $\mathbf{w}_j$ can be written as:

$$\mathbf{w}_j = a_{1j}\mathbf{v}_1 + a_{2j}\mathbf{v}_2 + \cdots + a_{nj}\mathbf{v}_n.$$

Consider $c_1\mathbf{w}_1 + \cdots + c_m\mathbf{w}_m = \mathbf{0}$. Substituting the expressions above, we get a homogeneous system of n equations in m unknowns. Since $m > n$, there are more unknowns than equations, so there is a nontrivial solution. Therefore, $\{\mathbf{w}_1, \ldots, \mathbf{w}_m\}$ is linearly dependent. □

Corollary. If V is finite-dimensional, then:

- Any linearly independent set has at most $\dim(V)$ vectors.
- Any spanning set has at least $\dim(V)$ vectors.

Example (Applying the Corollary). In $\mathbb{R}^3$ (dimension 3):

- Any set of 4 or more vectors is linearly dependent.
- Any spanning set must have at least 3 vectors.
- A basis must have exactly 3 vectors.

DIY. Explore the relationship between the dimension of a vector space and its subspaces. If U and W are subspaces of a finite-dimensional vector space V, what can you say about $\dim(U + W)$ and $\dim(U \cap W)$?

Try to prove that $\dim(U+W) = \dim(U)+\dim(W)-\dim(U\cap W)$. Start by choosing a basis for $U \cap W$, extend it to a basis for U, extend it to a basis for W, and show that together these vectors span $U+W$. This result is sometimes called the dimension formula for subspaces.

Chapter 4

Solving Inconsistent Systems: The Least Squares Method

Sometimes we encounter systems of equations that don't have an exact solution: they're inconsistent. But even when we can't solve a system exactly, we can often find a "best" approximate solution. The least squares method does exactly this: it finds the solution that minimizes the sum of squared errors.

4.1 The Problem of Inconsistent Systems

Consider a system of equations $A\mathbf{x} = \mathbf{b}$ where A is an $m \times n$ matrix. If the system is inconsistent, there is no vector $\mathbf{x}$ such that $A\mathbf{x} = \mathbf{b}$ exactly. However, we can try to find a vector $\mathbf{x}$ that makes $A\mathbf{x}$ as close as possible to $\mathbf{b}$.

Definition 23 (Least Squares Solution). Given a system $A\mathbf{x} = \mathbf{b}$ (which may be inconsistent), a **least squares solution** is a vector $\hat{\mathbf{x}}$ that minimizes the quantity:

$$\|A\mathbf{x} - \mathbf{b}\|^2 = (A\mathbf{x} - \mathbf{b})^T(A\mathbf{x} - \mathbf{b}).$$

This is the sum of the squares of the differences between $A\mathbf{x}$ and $\mathbf{b}$.

Example (A Simple Example). Suppose we have two equations in one unknown:

$$\begin{cases} x = 2 \\ x = 3 \end{cases}$$

This system is clearly inconsistent: no single value of x can satisfy both equations. But we can ask: what value of x makes the errors as small as possible?

If we choose $x = 2$, the errors are $(2-2)^2 + (2-3)^2 = 0 + 1 = 1$. If we choose $x = 2.5$, the errors are $(2.5-2)^2 + (2.5-3)^2 = 0.25 + 0.25 = 0.5$. If we choose $x = 3$, the errors are $(3-2)^2 + (3-3)^2 = 1 + 0 = 1$.

So $x = 2.5$ minimizes the sum of squared errors. This is the least squares solution, the average of the two values!

4.2 Geometric Interpretation

There's a beautiful geometric way to understand least squares. The key idea is that we want to find the point in the column space of A that is closest to $\mathbf{b}$.

Remark (The Key Insight). The vector $A\mathbf{x}$ always lies in the column space of A (since it's a linear combination of the columns). When the system is inconsistent, $\mathbf{b}$ is not in the column space of A. The least squares solution finds the point $A\hat{\mathbf{x}}$ in the column space that is closest to $\mathbf{b}$.

This closest point is given by the **orthogonal projection** of $\mathbf{b}$ onto the column space of A.

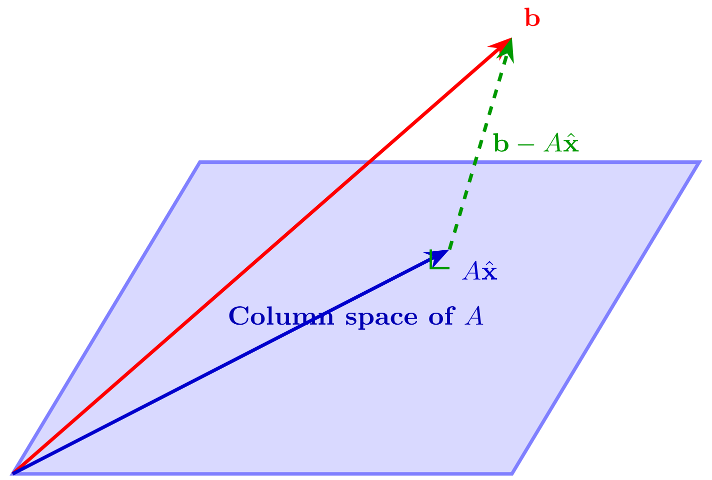


Figure 4.1: Geometric interpretation: $A\hat{\mathbf{x}}$ is the projection of $\mathbf{b}$ onto the column space of A. The error vector $\mathbf{b} - A\hat{\mathbf{x}}$ is perpendicular to the column space.

Example (Line Fitting). A very common application is fitting a line to data points. Suppose we have data points $(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)$ and we want to find a line $y = ax + b$ that best fits the data.

We want to solve:

$$\begin{cases} ax_1 + b = y_1 \\ ax_2 + b = y_2 \\ \quad\vdots \\ ax_m + b = y_m \end{cases}$$

In matrix form, this is $A\mathbf{x} = \mathbf{b}$ where:

$$A = \begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_m & 1 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} a \\ b \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \end{bmatrix} y_m.$$

Unless all points lie exactly on a line, this system is inconsistent. The least squares solution gives us the line that minimizes the sum of squared vertical distances from the points to the line.

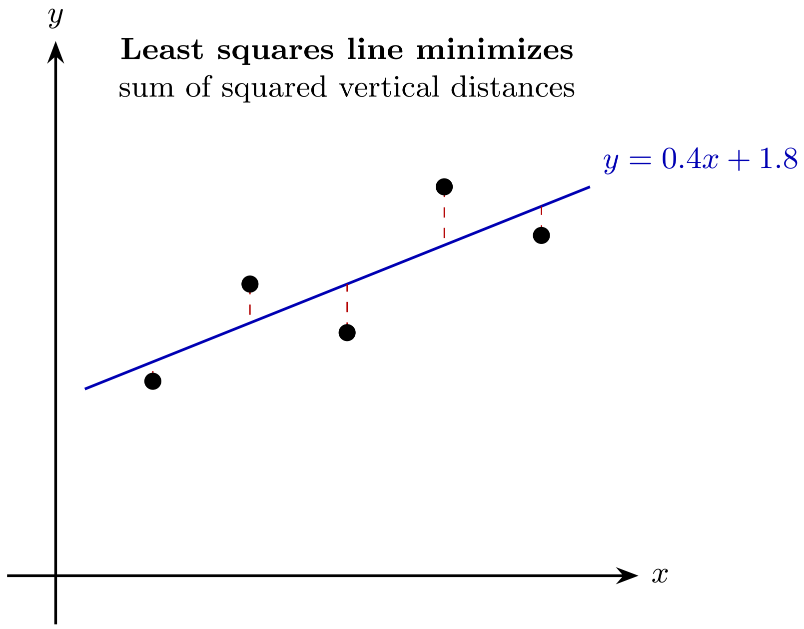


Figure 4.2: Fitting a line to data points using least squares. The vertical distances (red dashed lines) are minimized in the least squares sense.

4.3 Deriving the Normal Equations

Now let's derive the formula for finding the least squares solution. The key idea is that the error vector $\mathbf{b} - A\hat{\mathbf{x}}$ must be orthogonal (perpendicular) to the column space of A.

Theorem 13 (Normal Equations). The least squares solution $\hat{\mathbf{x}}$ to $A\mathbf{x} = \mathbf{b}$ satisfies:

$$A^T A\hat{\mathbf{x}} = A^T\mathbf{b}.$$

This system is called the **normal equations**.

Proof. We want to minimize $\|A\mathbf{x} - \mathbf{b}\|^2$. For the minimizer $\hat{\mathbf{x}}$, the vector $A\hat{\mathbf{x}}$ is the closest point in the column space of A to $\mathbf{b}$.

This means that the error vector $\mathbf{b} - A\hat{\mathbf{x}}$ must be orthogonal to every vector in the column space of A. In particular, it must be orthogonal to each column of A.

Let $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n$ be the columns of A. Then for each i:

$$\mathbf{a}_i^T(\mathbf{b} - A\hat{\mathbf{x}}) = 0.$$

Writing this in matrix form, we get:

$$A^T(\mathbf{b} - A\hat{\mathbf{x}}) = \mathbf{0},$$

which rearranges to:

$$A^T A\hat{\mathbf{x}} = A^T\mathbf{b}.$$

This is the normal equations. If $A^T A$ is invertible, we get:

$$\hat{\mathbf{x}} = (A^T A)^{-1}A^T\mathbf{b}.$$

□

Remark. The matrix $A^T A$ is always symmetric. If the columns of A are linearly independent, then $A^T A$ is invertible and the least squares solution is unique.

Example (Solving a Least Squares Problem). Find the least squares solution to:

$$\begin{cases} x = 1 \\ x = 2 \\ x = 4 \end{cases}$$

Here, $A = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} 1 \\ 2 \\ 4 \end{bmatrix}$.

First compute:

$$A^T A = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = 3,$$

$$A^T\mathbf{b} = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 4 \end{bmatrix} = 7.$$

The normal equation is $3x = 7$, so $x = \frac{7}{3} \approx 2.33$.

This is just the average of 1, 2, and 4! The least squares solution gives us the mean of the values.

Example (Fitting a Line). Find the line $y = ax + b$ that best fits the points $(1, 2)$, $(2, 3)$, and $(3, 3)$.

We want to solve:

$$\begin{cases} a(1) + b = 2 \\ a(2) + b = 3 \\ a(3) + b = 3 \end{cases}$$

In matrix form:

$$A = \begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} a \\ b \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 2 \\ 3 \\ 3 \end{bmatrix}.$$

Compute:

$$A^TA = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \end{bmatrix} = \begin{bmatrix} 14 & 6 \\ 6 & 3 \end{bmatrix},$$

$$A^T\mathbf{b} = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \\ 3 \end{bmatrix} = \begin{bmatrix} 17 \\ 8 \end{bmatrix}.$$

The normal equations are:

$$\begin{bmatrix} 14 & 6 \\ 6 & 3 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 17 \\ 8 \end{bmatrix}.$$

Solving this system (using Gaussian elimination or inverse):

$$\begin{aligned} \begin{bmatrix} 14 & 6 \\ 6 & 3 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} &= \begin{bmatrix} 17 \\ 8 \end{bmatrix} \\ \Rightarrow \quad \begin{bmatrix} a \\ b \end{bmatrix} &= \begin{bmatrix} \frac{1}{2} \\ \frac{3}{2} \end{bmatrix}. \end{aligned} \tag{4.1}$$

So the best-fit line is $y = \frac{1}{2}x + \frac{3}{2}$.

4.4 Minimizing the Norm

We've been minimizing $\|A\mathbf{x} - \mathbf{b}\|^2$, which is the square of the Euclidean norm. Let's see why this works.

Definition 24 (Euclidean Norm). For a vector $\mathbf{v} = (v_1, v_2, \ldots, v_m)^T$, the **Euclidean norm** (or 2-norm) is:

$$\|\mathbf{v}\| = \sqrt{v_1^2 + v_2^2 + \cdots + v_m^2}.$$

The square of the norm is:

$$\|\mathbf{v}\|^2 = v_1^2 + v_2^2 + \cdots + v_m^2.$$

Remark. Minimizing $\|A\mathbf{x} - \mathbf{b}\|^2$ is equivalent to minimizing $\|A\mathbf{x} - \mathbf{b}\|$ (since the square root is an increasing function). We work with the squared version because it's easier to differentiate and work with algebraically.

The squared error $\|A\mathbf{x} - \mathbf{b}\|^2$ expands as:

$$\begin{aligned} \|A\mathbf{x} - \mathbf{b}\|^2 &= (A\mathbf{x} - \mathbf{b})^T(A\mathbf{x} - \mathbf{b}) \\ &= \mathbf{x}^TA^TA\mathbf{x} - 2\mathbf{b}^TA\mathbf{x} + \mathbf{b}^T\mathbf{b}. \end{aligned} \tag{4.2}$$

To find the minimum, we can use calculus. Taking the gradient with respect to $\mathbf{x}$ and setting it to zero gives us the normal equations again. This confirms our geometric approach was correct.

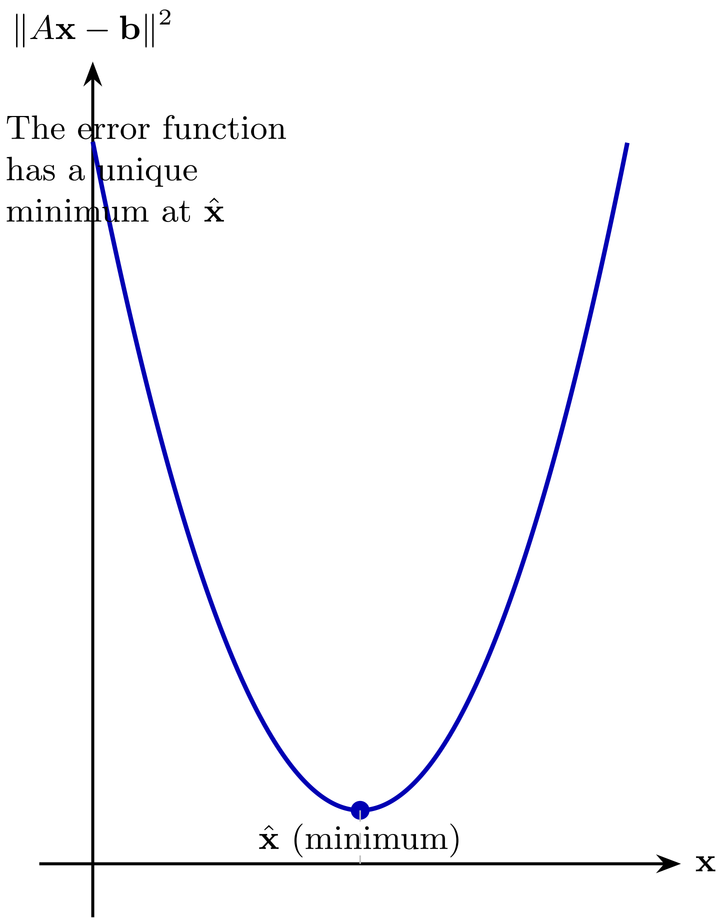


Figure 4.3: The error function $\|A\mathbf{x}-\mathbf{b}\|^2$ has a unique minimum at the least squares solution $\hat{\mathbf{x}}$.

4.5 Practical Considerations

In practice, solving the normal equations $A^TA\hat{\mathbf{x}} = A^T\mathbf{b}$ can sometimes be numerically unstable. A more stable approach uses the QR decomposition or singular value decomposition, but those are topics for more advanced courses.

DIY. Research the QR decomposition method for least squares. If $A = QR$ where Q is an orthogonal matrix (columns are orthonormal) and R is upper triangular, show that the normal equations simplify significantly.

The key insight is that $A^TA = R^TQ^TQR = R^TR$ since $Q^TQ = I$. Then the normal equations become $R^TR\hat{\mathbf{x}} = R^TQ^T\mathbf{b}$, which simplifies to $R\hat{\mathbf{x}} = Q^T\mathbf{b}$ (why?). Since R is upper triangular, this can be solved efficiently using back substitution. This method avoids computing A^TA explicitly, which can be numerically unstable. Explore when and why QR decomposition is preferred over the normal equations.

Remark. When the columns of A are linearly independent, A^TA is invertible and the least squares solution is unique. If the columns are linearly dependent, there are infinitely many least squares solutions, and we might want the one with smallest norm (called the minimum norm solution).

Example (When $A^T A$ is Not Invertible). Consider:

$$A = \begin{bmatrix} 1 & 2 \\ 2 & 4 \\ 3 & 6 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}.$$

Here, the second column is twice the first, so the columns are linearly dependent. Then:

$$A^T A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 2 & 4 \\ 3 & 6 \end{bmatrix} = \begin{bmatrix} 14 & 28 \\ 28 & 56 \end{bmatrix}.$$

Notice that the second row is twice the first, so $\det(A^T A) = 0$ and $A^T A$ is not invertible. In this case, there are infinitely many least squares solutions. The set of all solutions forms a line (or more generally, an affine subspace).

The least squares method is one of the most important techniques in applied mathematics, showing up everywhere from data fitting to signal processing to machine learning.

Chapter 5

Eigenvalues and Eigenvectors

Eigenvalues and eigenvectors are special numbers and vectors associated with a matrix. They show up in many applications, from understanding how systems evolve over time to analyzing networks and data. In this chapter, we'll learn what they are, how to find them, and see a practical application to Markov chains.

5.1 Definition and Basic Properties

Definition 25 (Eigenvalue and Eigenvector). Let A be an $n \times n$ matrix. A scalar λ is called an **eigenvalue** of A if there exists a nonzero vector $\mathbf{v}$ such that:

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

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

An eigenvector is a special vector that, when multiplied by the matrix A, gives back a scalar multiple of itself (the eigenvalue tells us by what factor it's scaled).

Example (A Simple Example). Let $A = \begin{bmatrix} 2 & 1 \\ 0 & 3 \end{bmatrix}$ and consider the vector $\mathbf{v} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$.

Then:

$$A\mathbf{v} = \begin{bmatrix} 2 & 1 \\ 0 & 3 \end{bmatrix}\begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \end{bmatrix} = 2\begin{bmatrix} 1 \\ 0 \end{bmatrix} = 2\mathbf{v}.$$

So $\mathbf{v} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ is an eigenvector with eigenvalue $\lambda = 2$.

Also, consider $\mathbf{w} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$:

$$A\mathbf{w} = \begin{bmatrix} 2 & 1 \\ 0 & 3 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ 3 \end{bmatrix} = 3\begin{bmatrix} 1 \\ 1 \end{bmatrix} = 3\mathbf{w}.$$

So $\mathbf{w} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ is an eigenvector with eigenvalue $\lambda = 3$.

Remark. Note that if $\mathbf{v}$ is an eigenvector, then any nonzero scalar multiple $c\mathbf{v}$ is also an eigenvector with the same eigenvalue. This is because $A(c\mathbf{v}) = c(A\mathbf{v}) = c(\lambda\mathbf{v}) = \lambda(c\mathbf{v})$.

5.2 Finding Eigenvalues

To find eigenvalues, we start with the equation $A\mathbf{v} = \lambda\mathbf{v}$ and rearrange it:

$$A\mathbf{v} - \lambda\mathbf{v} = \mathbf{0} \quad \Rightarrow \quad (A - \lambda I)\mathbf{v} = \mathbf{0}.$$

For this to have a nonzero solution $\mathbf{v}$, the matrix $A - \lambda I$ must be singular (not invertible). This happens exactly when its determinant is zero.

Definition 26 (Characteristic Polynomial). For an $n \times n$ matrix A, the **characteristic polynomial** is:

$$p(\lambda) = \det(A - \lambda I).$$

The eigenvalues of A are the roots of this polynomial (the values of λ for which $p(\lambda) = 0$).

Theorem 14 (Finding Eigenvalues). The eigenvalues of a matrix A are the solutions to the equation:

$$\det(A - \lambda I) = 0.$$

This is called the **characteristic equation**.

Example (Finding Eigenvalues of a 2×2 Matrix). Find the eigenvalues of $A = \begin{bmatrix} 2 & 1 \\ 0 & 3 \end{bmatrix}$.

First, form $A - \lambda I$:

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

Now compute the determinant:

$$\det(A - \lambda I) = (2-\lambda)(3-\lambda) - (1)(0) = (2-\lambda)(3-\lambda).$$

Set this equal to zero:

$$(2-\lambda)(3-\lambda) = 0.$$

So the eigenvalues are $\lambda = 2$ and $\lambda = 3$.

Example (Finding Eigenvalues of a 3×3 Matrix). Find the eigenvalues of $A = \begin{bmatrix} 1 & 2 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}$.

Form $A - \lambda I$:

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

Since this is upper triangular, the determinant is the product of diagonal entries:

$$\det(A - \lambda I) = (1-\lambda)(2-\lambda)(3-\lambda).$$

Setting this to zero gives eigenvalues $\lambda = 1$, $\lambda = 2$, and $\lambda = 3$.

Remark. For larger matrices, finding eigenvalues can be quite difficult. The characteristic polynomial is a degree n polynomial, and finding its roots exactly is not always possible. However, there are numerical methods (like the QR algorithm) that can approximate eigenvalues well. In this chapter, we focus on simple cases where we can find them exactly.

DIY. For a 3×3 or larger matrix, computing the characteristic polynomial $\det(A - \lambda I)$ can be computationally expensive. Research iterative methods like the power method or QR algorithm for finding eigenvalues of large matrices.

The power method is particularly simple: start with a random vector $\mathbf{v}_0$, and repeatedly compute $\mathbf{v}_{k+1} = A\mathbf{v}_k / \|A\mathbf{v}_k\|$. Under certain conditions, $\mathbf{v}_k$ converges to an eigenvector corresponding to the eigenvalue with largest absolute value. Why does this work? What are the limitations? When might this method fail?

5.3 Finding Eigenvectors

Once we know an eigenvalue λ, we can find the corresponding eigenvectors by solving the homogeneous system:

$$(A - \lambda I)\mathbf{v} = \mathbf{0}.$$

The eigenvectors are the nonzero solutions to this system. They form the null space (or kernel) of $A - \lambda I$.

Example (Finding Eigenvectors). Find the eigenvectors of $A = \begin{bmatrix} 2 & 1 \\ 0 & 3 \end{bmatrix}$.

We already found eigenvalues $\lambda = 2$ and $\lambda = 3$.

For $\lambda = 2$:

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

We solve $(A - 2I)\mathbf{v} = \mathbf{0}$:

$$\begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \Rightarrow \quad \begin{cases} v_2 = 0 \\ v_2 = 0 \end{cases}.$$

So $v_2 = 0$ and v_1 is free. Taking $v_1 = 1$, we get the eigenvector $\mathbf{v}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$.

For $\lambda = 3$:

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

We solve $(A - 3I)\mathbf{v} = \mathbf{0}$:

$$\begin{bmatrix} -1 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \Rightarrow \quad -v_1 + v_2 = 0.$$

So $v_1 = v_2$. Taking $v_1 = v_2 = 1$, we get the eigenvector $\mathbf{v}_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$.

Example (Finding All Eigenvectors). Find all eigenvalues and eigenvectors of $A = \begin{bmatrix} 4 & -2 \\ 1 & 1 \end{bmatrix}$.

First, find eigenvalues:

$$A - \lambda I = \begin{bmatrix} 4-\lambda & -2 \\ 1 & 1-\lambda \end{bmatrix},$$

$$\det(A - \lambda I) = (4-\lambda)(1-\lambda) - (-2)(1) = 4 - 4\lambda - \lambda + \lambda^2 + 2 = \lambda^2 - 5\lambda + 6.$$

Factoring: $\lambda^2 - 5\lambda + 6 = (\lambda - 2)(\lambda - 3) = 0$.
So the eigenvalues are $\lambda = 2$ and $\lambda = 3$.
For $\lambda = 2$:

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

The system is:

$$\begin{cases} 2v_1 - 2v_2 = 0 \\ v_1 - v_2 = 0 \end{cases} \quad \Rightarrow \quad v_1 = v_2.$$

Taking $v_1 = v_2 = 1$, we get $\mathbf{v}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$.
For $\lambda = 3$:

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

The system is:

$$\begin{cases} v_1 - 2v_2 = 0 \\ v_1 - 2v_2 = 0 \end{cases} \quad \Rightarrow \quad v_1 = 2v_2.$$

Taking $v_2 = 1$, so $v_1 = 2$, we get $\mathbf{v}_2 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$.

5.4 Application to Markov Chains

Markov chains model systems that change state randomly over time, where the probability of moving to the next state depends only on the current state. Eigenvalues and eigenvectors are incredibly useful for analyzing what happens to these systems over many steps.

Definition 27 (Markov Chain). A **Markov chain** is described by a **transition matrix** P where P_{ij} is the probability of moving from state i to state j in one step. The matrix P has the properties:

- All entries are between 0 and 1.
- Each row sums to 1 (since probabilities must sum to 1).

If $\mathbf{x}^{(0)}$ is the initial probability distribution (a vector whose i-th entry is the probability of being in state i at time 0), then after k steps, the distribution is:

$$\mathbf{x}^{(k)} = P^k \mathbf{x}^{(0)}.$$

Example (A Simple Weather Model). Suppose the weather can be either Sunny (S) or Rainy (R), and the transition probabilities are:

- If it's sunny today, there's a 90% chance it will be sunny tomorrow and 10% chance it will rain.
- If it's rainy today, there's a 50% chance it will be sunny tomorrow and 50% chance it will rain.

The transition matrix is:

$$P = \begin{bmatrix} 0.9 & 0.1 \\ 0.5 & 0.5 \end{bmatrix}.$$

The rows correspond to the current state (first row = sunny, second row = rainy), and the columns correspond to the next state (first column = sunny, second column = rainy).

Suppose today is sunny, so the initial distribution is $\mathbf{x}^{(0)} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ (100% sunny, 0% rainy).

After one day:

$$\mathbf{x}^{(1)} = P\mathbf{x}^{(0)} = \begin{bmatrix} 0.9 & 0.1 \\ 0.5 & 0.5 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0.9 \\ 0.1 \end{bmatrix}.$$

After two days:

$$\mathbf{x}^{(2)} = P\mathbf{x}^{(1)} = \begin{bmatrix} 0.9 & 0.1 \\ 0.5 & 0.5 \end{bmatrix} \begin{bmatrix} 0.9 \\ 0.1 \end{bmatrix} = \begin{bmatrix} 0.86 \\ 0.14 \end{bmatrix}.$$

We can continue this way, but it gets tedious. Instead, let's use eigenvalues and eigenvectors!

Example (Using Eigenvalues to Compute Powers). For the weather transition matrix $P = \begin{bmatrix} 0.9 & 0.1 \\ 0.5 & 0.5 \end{bmatrix}$, let's find the eigenvalues and eigenvectors.

First, find eigenvalues:

$$P - \lambda I = \begin{bmatrix} 0.9 - \lambda & 0.1 \\ 0.5 & 0.5 - \lambda \end{bmatrix},$$

$$\det(P-\lambda I) = (0.9-\lambda)(0.5-\lambda)-(0.1)(0.5) = 0.45-0.9\lambda-0.5\lambda+\lambda^2-0.05 = \lambda^2-1.4\lambda+0.4.$$

Solving $\lambda^2 - 1.4\lambda + 0.4 = 0$:

$$\lambda = \frac{1.4 \pm \sqrt{1.96 - 1.6}}{2} = \frac{1.4 \pm 0.6}{2}.$$

So $\lambda_1 = 1$ and $\lambda_2 = 0.4$.

For $\lambda_1 = 1$:

$$P - I = \begin{bmatrix} -0.1 & 0.1 \\ 0.5 & -0.5 \end{bmatrix}.$$

The system is $-0.1v_1 + 0.1v_2 = 0$, so $v_1 = v_2$. Taking $v_1 = v_2 = 1$, we get $\mathbf{v}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$.

For $\lambda_2 = 0.4$:

$$P - 0.4I = \begin{bmatrix} 0.5 & 0.1 \\ 0.5 & 0.1 \end{bmatrix}.$$

The system is $0.5v_1 + 0.1v_2 = 0$, so $v_2 = -5v_1$. Taking $v_1 = 1$, so $v_2 = -5$, we get $\mathbf{v}_2 = \begin{bmatrix} 1 \\ -5 \end{bmatrix}$.

To compute P^{30} (the weather probabilities after 30 days), we can use the eigenvalues. Since we found eigenvalues $\lambda_1 = 1$ and $\lambda_2 = 0.4$, and eigenvectors $\mathbf{v}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $\mathbf{v}_2 = \begin{bmatrix} 1 \\ -5 \end{bmatrix}$, we can write any initial distribution as a combination of these eigenvectors.

After 30 steps, the contribution from $\lambda_2 = 0.4$ becomes $(0.4)^{30}$, which is extremely small (approximately 1.15×10^{-12}), so it effectively vanishes.

After many steps, the distribution approaches the eigenvector corresponding to $\lambda = 1$ (normalized). Computing this, we find that the long-term distribution is approximately:

$$\mathbf{x}^{(\infty)} \approx \begin{bmatrix} 5/6 \\ 1/6 \end{bmatrix} = \begin{bmatrix} 0.833 \\ 0.167 \end{bmatrix}.$$

So in the long run, it will be sunny about 83.3% of the time and rainy about 16.7% of the time, regardless of the initial weather!

Remark (Steady State). For a Markov chain, if there is an eigenvalue equal to 1 (which is always the case for a transition matrix), the corresponding eigenvector (normalized so its entries sum to 1) gives the **steady state** or **stationary distribution**. This is the long-term probability distribution that the system approaches, regardless of where it starts.

Example (Finding the Steady State). For our weather example, the eigenvector corresponding to $\lambda = 1$ was $\mathbf{v}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$.

To normalize it (so the entries sum to 1), we divide by the sum:

$$\frac{1}{1+1}\begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix}.$$

But wait, this doesn't match our previous calculation! Let me recompute more carefully.

Actually, we need to be more careful. The steady state vector $\boldsymbol{\pi}$ satisfies $P\boldsymbol{\pi} = \boldsymbol{\pi}$ and $\boldsymbol{\pi}_1 + \boldsymbol{\pi}_2 = 1$.

From $P\boldsymbol{\pi} = \boldsymbol{\pi}$, we get:

$$\begin{cases} 0.9\pi_1 + 0.5\pi_2 = \pi_1 \\ 0.1\pi_1 + 0.5\pi_2 = \pi_2 \end{cases}$$

The first equation gives: $0.9\pi_1 + 0.5\pi_2 = \pi_1 \Rightarrow 0.5\pi_2 = 0.1\pi_1 \Rightarrow \pi_2 = 0.2\pi_1$.

Using $\pi_1 + \pi_2 = 1$: $\pi_1 + 0.2\pi_1 = 1 \Rightarrow 1.2\pi_1 = 1 \Rightarrow \pi_1 = \frac{5}{6}$.
Then $\pi_2 = \frac{1}{6}$.

So the steady state is $\boldsymbol{\pi} = \begin{bmatrix} 5/6 \\ 1/6 \end{bmatrix}$, which matches our earlier result!

DIY. Investigate the Cayley-Hamilton theorem, which states that every square matrix satisfies its own characteristic equation. That is, if $p(\lambda) = \det(A - \lambda I) = \lambda^n + c_{n-1}\lambda^{n-1} + \cdots + c_0$, then $p(A) = A^n + c_{n-1}A^{n-1} + \cdots + c_0 I = 0$.

Verify this for a 2×2 matrix $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ by computing the characteristic polynomial and then checking that $A^2 - (a + d)A + (ad - bc)I = 0$. What are the implications of this theorem? Can it be used to compute powers of matrices or matrix inverses?

Eigenvalues and eigenvectors are powerful tools that let us understand how systems evolve over time. They make it possible to analyze what happens to Markov chains and other systems after many iterations.

Keywords

This document covers the following key topics:

Matrix Operations
Matrix addition
Matrix multiplication
Matrix transpose
Matrix inverse
Gaussian elimination
Row echelon form
Reduced row echelon form

Determinants
Determinant computation
Cofactor expansion
Permutation method
Computational complexity

Vector Spaces
Vector space axioms
Subspaces
Linear independence
Spanning sets
Basis and dimension

Least Squares
Inconsistent systems
Normal equations
Orthogonal projection
QR decomposition
Line fitting

Eigenvalues
Characteristic polynomial
Eigenvectors
Markov chains
Steady state distribution

Linear Systems
Consistent systems
Inconsistent systems
Homogeneous systems
Solution sets
Free variables
Basic variables

Matrix Properties
Invertible matrices
Singular matrices
Rank and nullity
Column space
Row space
Null space