Vectors and Matrices

1. Definitions and Notation

Vector: A vector is a list of numbers in a row or column.
Example:

$$\mathbf{v} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

Matrix: A matrix is a rectangular array of numbers.
Example:

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

2. Vector and Matrix Operations

Addition and Subtraction

Rule: Add or subtract corresponding elements.
Example:

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

Scalar Multiplication

Rule: Multiply each element by the scalar.
Example:

$$2 \cdot \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} = \begin{bmatrix} 2 & 4 & 6 \\ 8 & 10 & 12 \\ 14 & 16 & 18 \end{bmatrix}$$

Dot Product (Vectors)

Formula: For vectors $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$:

$$\mathbf{a} \cdot \mathbf{b} = \sum_{i=1}^{n} a_i b_i$$

Example:

$$\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \cdot \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix} = 1 \cdot 4 + 2 \cdot 5 + 3 \cdot 6 = 32$$

Matrix Multiplication

Rule: Multiply rows of the first matrix by columns of the second.
Example:

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \cdot \begin{bmatrix} 9 & 8 & 7 \\ 6 & 5 & 4 \\ 3 & 2 & 1 \end{bmatrix} = \begin{bmatrix} 30 & 24 & 18 \\ 84 & 69 & 54 \\ 138 & 114 & 90 \end{bmatrix}$$

Transpose of a Matrix

Notation: A^T (flip rows and columns).
Example:

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

Identity Matrix and Diagonal Matrix

Identity Matrix (I):

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

Diagonal Matrix:

$$D = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 7 \end{bmatrix}$$

3. Inverse of a Matrix

Definition: For a matrix A, the inverse A^{-1} satisfies:

$$A \cdot A^{-1} = I$$

where I is the identity matrix.

Note: Not all matrices have an inverse. A matrix must be square (same number of rows and columns), and $\det(A) \neq 0$.

Example 1 (2x2 matrix): Let

$$A = \begin{bmatrix} 4 & 7 \\ 2 & 6 \end{bmatrix} \Rightarrow A^{-1} = \frac{1}{\det(A)} \begin{bmatrix} 6 & -7 \\ -2 & 4 \end{bmatrix}$$

$$\det(A) = (4)(6) - (7)(2) = 24 - 14 = 10 \Rightarrow A^{-1} = \frac{1}{10} \begin{bmatrix} 6 & -7 \\ -2 & 4 \end{bmatrix}$$

Example 2 (3x3 matrix): Let

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

Step 1: Find the determinant of A:

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

First, calculate the 2x2 determinants:

$$\begin{vmatrix} 1 & 4 \\ 6 & 0 \end{vmatrix} = (1)(0) - (4)(6) = -24$$

$$\begin{vmatrix} 0 & 4 \\ 5 & 0 \end{vmatrix} = (0)(0) - (4)(5) = -20$$

$$\begin{vmatrix} 0 & 1 \\ 5 & 6 \end{vmatrix} = (0)(6) - (1)(5) = -5$$

Now, substitute these values into the determinant calculation:

$$\det(A) = 1 \cdot (-24) - 2 \cdot (-20) + 3 \cdot (-5)$$

$$\det(A) = -24 + 40 - 15 = 1$$

Step 2: Find the adjugate matrix.

The adjugate matrix A^* is the transpose of the cofactor matrix of A. The cofactor matrix is computed by calculating the minor (determinant of 2x2 submatrices) and applying a checkerboard pattern of signs.

For each element, we calculate the minor and apply the sign pattern:

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

Substituting the minors calculated earlier:

$$A^* = \begin{bmatrix} -24 & 20 & -5 \\ -18 & 3 & 3 \\ 5 & -3 & 1 \end{bmatrix}$$

Step 3: Find the inverse.

Now that we have the adjugate matrix, we can find the inverse by dividing the adjugate matrix by the determinant:

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

Since $\det(A) = 1$, the inverse is simply:

$$A^{-1} = \begin{bmatrix} -24 & 20 & -5 \\ -18 & 3 & 3 \\ 5 & -3 & 1 \end{bmatrix}$$

4. Determinant of a 3×3 Matrix

Formula: For matrix

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

$$\det(A) = a(ei - fh) - b(di - fg) + c(dh - eg)$$

Example:

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 4 & 5 \\ 1 & 0 & 6 \end{bmatrix} \Rightarrow \det(A) = 1(4 \cdot 6 - 5 \cdot 0) - 2(0 \cdot 6 - 5 \cdot 1) + 3(0 \cdot 0 - 4 \cdot 1)$$

$$\det(A) = 1(24) - 2(-5) + 3(-4) = 24 + 10 - 12 = 22$$

5. Eigenvalues and Eigenvectors

Definition: Let A be an $n \times n$ matrix. A non-zero vector $\mathbf{x}$ is called an *eigenvector* of A if it satisfies the equation

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

for some scalar λ. The scalar λ is called the *eigenvalue* corresponding to the eigenvector $\mathbf{x}$.

To find eigenvalues: Solve the characteristic equation:

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

Example: Let

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

Step 1: Find the eigenvalues. Solve the characteristic equation:

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

First, compute $A - \lambda I$:

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

Now, compute the determinant:

$$\begin{aligned}
\det(A - \lambda I) &= \det \begin{bmatrix} 4-\lambda & 1 & 2 \\ 1 & 4-\lambda & 1 \\ 2 & 1 & 4-\lambda \end{bmatrix} \\
&= (4-\lambda)\left[(4-\lambda)(4-\lambda) - 1 \cdot 1\right] - 1\left[1(4-\lambda) - 1 \cdot 2\right] + 2\left[1 \cdot 1 - (4-\lambda) \cdot 2\right] \\
&= (4-\lambda)\left[(4-\lambda)^2 - 1\right] - 1\left[4 - \lambda - 2\right] + 2\left[1 - 2(4-\lambda)\right] \\
&= (4-\lambda)\left[\lambda^2 - 8\lambda + 15\right] - (2-\lambda) + 2\left[1 - 8 + 2\lambda\right]
\end{aligned}$$

Simplifying this determinant will give us a cubic equation in terms of λ. For brevity, let's assume we find the roots of this cubic equation (eigenvalues) to be:

$$\lambda_1 = 6, \quad \lambda_2 = 3, \quad \lambda_3 = 2$$

Step 2: Find the eigenvectors.
For $\lambda = 6$:

$$A - 6I = \begin{bmatrix} -2 & 1 & 2 \\ 1 & -2 & 1 \\ 2 & 1 & -2 \end{bmatrix}$$

Solve the system $(A - 6I)\mathbf{x} = 0$:

$$\begin{bmatrix} -2 & 1 & 2 \\ 1 & -2 & 1 \\ 2 & 1 & -2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Solving this system, we find the eigenvector:

$$\mathbf{x}_1 = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}$$

For $\lambda = 3$:

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

Solve the system $(A - 3I)\mathbf{x} = 0$:

$$\begin{bmatrix} 1 & 1 & 2 \\ 1 & 1 & 1 \\ 2 & 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Solving this system, we find the eigenvector:

$$\mathbf{x}_2 = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}$$

For $\lambda = 2$:

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

Solve the system $(A - 2I)\mathbf{x} = 0$:

$$\begin{bmatrix} 2 & 1 & 2 \\ 1 & 2 & 1 \\ 2 & 1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Solving this system, we find the eigenvector:

$$\mathbf{x}_3 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

Linear Transformations Using Matrices

1. Understanding Linear Transformations

Definition: A *linear transformation* is a function $T : \mathbb{R}^n \to \mathbb{R}^m$ that maps a vector $\mathbf{x} \in \mathbb{R}^n$ to another vector $T(\mathbf{x}) \in \mathbb{R}^m$ such that it satisfies the following two properties:

- **Additivity:** $T(\mathbf{a} + \mathbf{b}) = T(\mathbf{a}) + T(\mathbf{b})$ for all vectors $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$.
- **Homogeneity:** $T(c \cdot \mathbf{a}) = c \cdot T(\mathbf{a})$ for any scalar c and vector $\mathbf{a} \in \mathbb{R}^n$.

In matrix form: The linear transformation T can be represented as a matrix multiplication:

$$T(\mathbf{x}) = A\mathbf{x} \quad \text{where} \quad A \in \mathbb{R}^{m \times n} \text{ is a matrix.}$$

Example: Stretching along the X-axis

Consider the linear transformation defined by the matrix

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

Let the vector $\mathbf{x}$ be given by

$$\mathbf{x} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

The transformed vector $A\mathbf{x}$ is

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

This transformation stretches the vector $\mathbf{x}$ by a factor of 2 along the X-axis, while leaving the Y-coordinate unchanged.

2. Matrix as a Linear Transformation

Matrix multiplication transforms vectors geometrically. Different matrices cause different geometric effects on vectors in the plane.

$$\text{If } A \text{ is an } n \times n \text{ matrix, then } A\mathbf{x} \text{ transforms the vector } \mathbf{x} \text{ in } \mathbb{R}^n.$$

Example: Rotation (90° Counterclockwise)

The matrix for a counterclockwise rotation by 90° is given by:

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

Consider the vector:

$$\mathbf{v} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

The rotated vector $R\mathbf{v}$ is:

$$R\mathbf{v} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

3. Types of Linear Transformations (with Examples)

1. Scaling

Effect: Scaling stretches or shrinks a vector along the coordinate axes.

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

The transformed vector is:

$$A\mathbf{x} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}.$$

This transformation scales the vector by a factor of 2 along the X-axis and 3 along the Y-axis.

2. Rotation

Effect: A rotation matrix rotates vectors around the origin by an angle θ.

$$R_\theta = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

For a rotation by 90°, the matrix is:

$$R_{90^\circ} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}.$$

3. Reflection

Effect: A reflection matrix reflects a vector across a line, such as the X-axis or Y-axis.

Reflection across the X-axis:

The reflection matrix across the X-axis is:

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

This transformation flips the Y-component of any vector.

4. Shear

Effect: A shear transformation slants the shape of an object along one axis.
For a horizontal shear, the matrix is:

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

where k is the shear factor.

$$\text{For } \mathbf{x} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \quad k = 1, \quad A\mathbf{x} = \begin{bmatrix} 3 \\ 2 \end{bmatrix}.$$

This transformation shifts the X-coordinate by $k\times$ the Y-coordinate.

5. Projection

Effect: A projection matrix projects a vector onto a line or plane.

Projection onto the X-axis:

The matrix that projects vectors onto the X-axis is:

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

For the vector:

$$\mathbf{x} = \begin{bmatrix} 1 \\ 2 \end{bmatrix},$$

the projection onto the X-axis is:

$$A\mathbf{x} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

This transformation eliminates the Y-component and keeps only the X-component.

Systems of Linear Equations

1. Introduction

A **system of linear equations** is a collection of equations involving the same set of variables.

Example:

$$2x + y = 5$$
$$3x - y = 4$$

Goal: Find values of x and y that satisfy both equations.

2. Methods of Solving

(a) Substitution Method

1. Solve one equation for one variable. 2. Substitute into the other equation.
Example:
From $2x + y = 5$, solve for y:

$$y = 5 - 2x$$

Substitute into $3x - y = 4$:

$$3x - (5 - 2x) = 4 \Rightarrow 5x = 9 \Rightarrow x = \frac{9}{5}$$

Then, find y:

$$y = 5 - 2 \cdot \frac{9}{5} = \frac{7}{5}$$

(b) Elimination Method

1. Add or subtract equations to eliminate one variable.

$$2x + y = 5 \quad (1)$$
$$3x - y = 4 \quad (2)$$
$$\text{Add: } (1) + (2) \Rightarrow 5x = 9 \Rightarrow x = \frac{9}{5}$$

Substitute back:

$$2x + y = 5 \Rightarrow y = \frac{7}{5}$$

3. Matrix Form: $A\mathbf{x} = \mathbf{b}$

Any system can be written as:

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

Where:

$$A = \begin{bmatrix} 2 & 1 \\ 3 & -1 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} x \\ y \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 5 \\ 4 \end{bmatrix}$$

4. Types of Solutions (with Examples)

(a) Unique Solution (One Intersection Point)

$$x + y = 4$$
$$2x - y = 1$$

Solving:

$$\text{Add: } (x + y) + (2x - y) = 3x = 5 \Rightarrow x = \frac{5}{3}$$

$$y = 4 - x = \frac{7}{3}$$

One unique solution: $\left(\frac{5}{3}, \frac{7}{3}\right)$

(b) Infinite Solutions (Same Line)

$$\begin{aligned} x + y &= 2 \\ 2x + 2y &= 4 \end{aligned} \Rightarrow \text{Second equation is a multiple of the first}$$

All points on the line $x + y = 2$ are solutions.
Example: $(0, 2), (1, 1), (2, 0), \ldots$
Infinitely many solutions.

(c) No Solution (Parallel Lines)

$$\begin{aligned} x + y &= 2 \\ x + y &= 5 \end{aligned} \Rightarrow \text{Contradiction: } 2 \neq 5$$

These lines have the same slope but different intercepts — they never meet. **No solution exists.**

5. Role in Model Optimization

Systems of equations are central to optimization and machine learning:

- **Linear regression:** Solving $X\mathbf{w} = \mathbf{y}$ leads to optimal weights.
- **Least squares:** Used when no exact solution exists — minimize error.
- **Constraints:** Optimization often involves solving systems with constraints.

Example: Linear Regression Normal Equation

$$\mathbf{w} = (X^T X)^{-1} X^T \mathbf{y}$$

This equation minimizes squared error between predicted and actual values.

Determinants and Rank

1. Determinant

The **determinant** is a scalar value that can be computed from a square matrix. It tells us:

- Whether a matrix is invertible.
- The scaling factor of the transformation represented by the matrix.
- Whether rows/columns are linearly dependent.

Notation:

$$\text{For matrix } A, \quad \det(A) \text{ or } |A|$$

Example (2x2 matrix):

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}, \quad \det(A) = ad - bc$$

Example:

$$A = \begin{bmatrix} 2 & 3 \\ 1 & 4 \end{bmatrix} \Rightarrow \det(A) = (2)(4) - (3)(1) = 8 - 3 = 5$$

Since $\det(A) \neq 0$, the matrix is **invertible**.

Example (Singular Matrix):

$$B = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} \Rightarrow \det(B) = (1)(4) - (2)(2) = 4 - 4 = 0$$

Matrix B is **not invertible** — it is singular.

2. Determinant of a 3x3 Matrix

For a 3×3 matrix:

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

The determinant is calculated as:

$$\det(A) = a(ei - fh) - b(di - fg) + c(dh - eg)$$

Example:

Let

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

Then:

$$\det(A) = 2 \begin{vmatrix} 5 & 6 \\ 8 & 9 \end{vmatrix} - 3 \begin{vmatrix} 4 & 6 \\ 7 & 9 \end{vmatrix} + 1 \begin{vmatrix} 4 & 5 \\ 7 & 8 \end{vmatrix}$$

Now compute the 2x2 determinants:

$$\begin{vmatrix} 5 & 6 \\ 8 & 9 \end{vmatrix} = (5)(9) - (6)(8) = 45 - 48 = -3$$

$$\begin{vmatrix} 4 & 6 \\ 7 & 9 \end{vmatrix} = (4)(9) - (6)(7) = 36 - 42 = -6$$

$$\begin{vmatrix} 4 & 5 \\ 7 & 8 \end{vmatrix} = (4)(8) - (5)(7) = 32 - 35 = -3$$

Now substitute back:

$$\det(A) = 2(-3) - 3(-6) + 1(-3) = -6 + 18 - 3 = 9$$

Since $\det(A) = 9 \neq 0$, matrix A is **invertible**.

3. Rank of a Matrix

The **rank** of a matrix is the number of linearly independent rows or columns.

Key Points:

- Rank tells how much "information" the matrix holds.
- Rank $r \leq \min(m, n)$ for a matrix of size $m \times n$.
- A matrix with full rank has linearly independent rows and columns.

Example:

Let

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

Here, row 2 is a multiple of row 1:

$$\text{Row}_2 = 2 \times \text{Row}_1$$

So the rows are not all linearly independent. Rank = number of linearly independent rows = 2.

4. Full Rank vs. Rank-Deficient

Full Rank:

A matrix is **full rank** if:

- For square matrices: rank $= n$
- For rectangular matrices: rank $= \min(m, n)$

Example:

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \Rightarrow \text{rank} = 2 \text{ (full rank)}$$

Rank-Deficient:

A matrix is **rank-deficient** if some rows/columns are linearly dependent.
Example:

$$\begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} \Rightarrow \text{rank} = 1$$

5. Relationship Between Rank and Determinant

- A square matrix is invertible if and only if $\det(A) \neq 0$.
- If a matrix has full rank, its determinant is non-zero.
- If a matrix is rank-deficient, its determinant is zero.

Summary

- **Determinant** indicates invertibility and linear dependence.
- **Rank** counts the number of linearly independent rows/columns.
- **Full rank:** All rows/columns are independent.
- **Rank-deficient:** Some rows/columns are dependent.

Eigenvalues and Eigenvectors

1. Definition and Computation

An **eigenvalue** and its corresponding **eigenvector** are fundamental concepts in linear algebra, often used in various applications such as Principal Component Analysis (PCA).

Eigenvalue and Eigenvector Definition:

For a square matrix A, a non-zero vector $\mathbf{v}$ is called an **eigenvector** if it satisfies the equation:

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

where: - A is the square matrix, - $\mathbf{v}$ is the eigenvector, - λ is the corresponding eigenvalue.

In this equation, applying the matrix A to the vector $\mathbf{v}$ results in a scaled version of $\mathbf{v}$ by the factor λ.

Computing Eigenvalues and Eigenvectors:

The eigenvalue λ is found by solving the characteristic equation:

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

where: - det denotes the determinant, - I is the identity matrix of the same size as A, - λ are the eigenvalues.

Once the eigenvalues λ are found, the corresponding eigenvectors can be computed by solving:

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

Example:

Consider the matrix:

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

To find the eigenvalues, solve the characteristic equation:

$$\det(A - \lambda I) = \det \begin{bmatrix} 4-\lambda & 1 \\ 2 & 3-\lambda \end{bmatrix} = 0$$

$$(4-\lambda)(3-\lambda) - (1)(2) = 0$$

$$\lambda^2 - 7\lambda + 10 = 0$$

Solving the quadratic equation gives the eigenvalues:

$$\lambda_1 = 5, \quad \lambda_2 = 2$$

Next, substitute each eigenvalue into $(A - \lambda I)\mathbf{v} = 0$ to find the eigenvectors.

For $\lambda_1 = 5$:

$$\begin{bmatrix} 4-5 & 1 \\ 2 & 3-5 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = 0 \quad \Rightarrow \quad \begin{bmatrix} -1 & 1 \\ 2 & -2 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = 0$$

Solving this system gives the eigenvector corresponding to $\lambda_1 = 5$:

$$\mathbf{v_1} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

For $\lambda_2 = 2$:

$$\begin{bmatrix} 4-2 & 1 \\ 2 & 3-2 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = 0 \quad \Rightarrow \quad \begin{bmatrix} 2 & 1 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = 0$$

Solving this system gives the eigenvector corresponding to $\lambda_2 = 2$:

$$\mathbf{v_2} = \begin{bmatrix} -0.5 \\ 1 \end{bmatrix}$$

So, the eigenvalues of matrix A are $\lambda_1 = 5$ and $\lambda_2 = 2$, with corresponding eigenvectors $\mathbf{v_1} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $\mathbf{v_2} = \begin{bmatrix} -0.5 \\ 1 \end{bmatrix}$.

2. Spectral Decomposition

Spectral decomposition is a way to express a matrix as a sum of eigenvalues and eigenvectors.

For a square matrix A, if the matrix is diagonalizable, it can be expressed as:

$$A = V\Lambda V^{-1}$$

where: - V is the matrix of eigenvectors, - Λ is the diagonal matrix of eigenvalues, - V^{-1} is the inverse of the matrix V.

Example of Spectral Decomposition:

Given that matrix A has eigenvalues $\lambda_1 = 5$ and $\lambda_2 = 2$, and eigenvectors $\mathbf{v_1} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $\mathbf{v_2} = \begin{bmatrix} -0.5 \\ 1 \end{bmatrix}$, we can write the spectral decomposition as:

$$A = \begin{bmatrix} 1 & -0.5 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 5 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ -0.5 & 1 \end{bmatrix}$$

3. Importance in PCA (Principal Component Analysis)

Eigenvalues and eigenvectors are central to PCA, a technique used for dimensionality reduction. In PCA, the goal is to project data onto a lower-dimensional space while retaining as much variance as possible.

PCA Steps:

- Compute the covariance matrix C of the data.
- Find the eigenvalues and eigenvectors of C.
- The eigenvectors (principal components) define the directions of maximum variance in the data.
- The eigenvalues give the magnitude of the variance along each principal component.

The eigenvectors corresponding to the largest eigenvalues represent the principal components that capture the most significant variance in the data.

Example:

If we have a dataset represented as a covariance matrix, the eigenvalues will tell us how much variance each principal component accounts for. By selecting the principal components corresponding to the largest eigenvalues, we can reduce the dimensionality while retaining the most important features of the data.

4. Importance in Covariance Matrix Analysis

Eigenvalues and eigenvectors also play a key role in analyzing the **covariance matrix** of a dataset.

Covariance Matrix:

The covariance matrix is a square matrix that contains the covariances between pairs of elements in the dataset. The eigenvectors of the covariance matrix represent the directions of the maximum variance in the data, and the corresponding eigenvalues tell us how much variance is explained by each eigenvector.

Example:

If the covariance matrix C of a dataset is:

$$C = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

The eigenvalues and eigenvectors of C represent the directions and magnitudes of the most significant variances in the data.

Summary

- Eigenvalues and eigenvectors provide insight into the scaling and directions of transformation of a matrix.
- Spectral decomposition expresses a matrix as a sum of eigenvalues and eigenvectors, useful for diagonalization.
- In PCA, eigenvectors determine the principal components, and eigenvalues give the importance of each component.
- Eigenvalues and eigenvectors are crucial for analyzing the covariance matrix and understanding the directions of maximum variance in the data.

Singular Value Decomposition (SVD)

1. What is SVD?

Singular Value Decomposition (SVD) is a method that helps us break a matrix into simpler parts. It's like taking a complicated object and expressing it as a combination of basic building blocks.

If A is a matrix with m rows and n columns, then we can write it as:

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

Here's what each part means:

- A: The original matrix (size $m \times n$)
- U: An $m \times m$ matrix with special columns called **left singular vectors**
- Σ: A diagonal matrix (size $m \times n$) with non-negative numbers called **singular values**
- V^T: The transpose of an $n \times n$ matrix V, whose columns are called **right singular vectors**

The singular values in Σ are ordered from largest to smallest. They tell us how much each direction (from V) contributes to the original matrix A.

Why is SVD useful?

- It helps in data compression and noise reduction
- It's used in machine learning, image processing, and recommendation systems
- It gives us insight into the structure of a matrix

2. Example of SVD

Let's take a simple matrix:

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

We want to decompose it as:

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

Step 1: Compute A^TA and AA^T We first compute the Gram matrices A^TA and AA^T.

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

Now calculate A^TA:

$$A^TA = \begin{bmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} = \begin{bmatrix} 35 & 44 \\ 44 & 56 \end{bmatrix}$$

Next, calculate AA^T:

$$AA^T = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} \begin{bmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{bmatrix} = \begin{bmatrix} 5 & 11 & 17 \\ 11 & 25 & 39 \\ 17 & 39 & 61 \end{bmatrix}$$

Step 2: Find Eigenvalues and Eigenvectors We need to compute the eigenvalues and eigenvectors of A^TA and AA^T. **Eigenvalues of A^TA:** We solve $\det(A^TA - \lambda I) = 0$ to find the eigenvalues. Using numerical methods (or a computational tool), we find the eigenvalues of A^TA to be approximately:

$$\lambda_1 = 91.2270, \quad \lambda_2 = 0.7729$$

Eigenvectors of A^TA: The eigenvectors corresponding to these eigenvalues are:

$$v_1 = \begin{bmatrix} -0.6196 \\ -0.7849 \end{bmatrix}, \quad v_2 = \begin{bmatrix} 0.7849 \\ -0.6196 \end{bmatrix}$$

Thus, V is:

$$V = \begin{bmatrix} -0.6196 & 0.7849 \\ -0.7849 & -0.6196 \end{bmatrix}$$

Therefore, V^T is:

$$V^T = \begin{bmatrix} -0.6196 & -0.7849 \\ 0.7849 & -0.6196 \end{bmatrix}$$

Eigenvalues of AA^T: Similarly, we solve $\det(AA^T - \lambda I) = 0$ to find the eigenvalues of AA^T, which are approximately:

$$\lambda_1 = 91.2270, \quad \lambda_2 = 0.7729, \quad \lambda_3 = 0$$

Eigenvectors of AA^T: The eigenvectors corresponding to these eigenvalues are:

$$u_1 = \begin{bmatrix} -0.2298 \\ -0.5247 \\ -0.8196 \end{bmatrix}, \quad u_2 = \begin{bmatrix} 0.8835 \\ 0.2408 \\ -0.4019 \end{bmatrix}, \quad u_3 = \begin{bmatrix} 0.4082 \\ -0.8165 \\ 0.4082 \end{bmatrix}$$

Thus, U is:

$$U = \begin{bmatrix} -0.2298 & 0.8835 & 0.4082 \\ -0.5247 & 0.2408 & -0.8165 \\ -0.8196 & -0.4019 & 0.4082 \end{bmatrix}$$

Step 3: Construct the Σ Matrix The singular values are the square roots of the eigenvalues. Thus:

$$\sigma_1 = \sqrt{91.2270} \approx 9.5255, \quad \sigma_2 = \sqrt{0.7729} \approx 0.5143$$

So, the Σ matrix is:

$$\Sigma = \begin{bmatrix} 9.5255 & 0 \\ 0 & 0.5143 \\ 0 & 0 \end{bmatrix}$$

Step 4: Form the Decomposition Finally, we can express the original matrix A as the product of U, Σ, and V^T:

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

Substituting the matrices, we get:

$$A = \begin{bmatrix} -0.2298 & 0.8835 & 0.4082 \\ -0.5247 & 0.2408 & -0.8165 \\ -0.8196 & -0.4019 & 0.4082 \end{bmatrix} \begin{bmatrix} 9.5255 & 0 \\ 0 & 0.5143 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} -0.6196 & -0.7849 \\ 0.7849 & -0.6196 \end{bmatrix}$$

Multiplying these matrices will yield the original matrix A (or a very close approximation within numerical precision):

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

Summary: We performed Singular Value Decomposition (SVD) on the matrix A, resulting in the matrices U, Σ, and V^T, and verified that multiplying these matrices together reconstructs the original matrix A.

2. Applications of SVD

Singular Value Decomposition (SVD) is not just a mathematical concept — it's widely used in real-world applications. Below are some key areas where SVD plays an important role.

1. Dimensionality Reduction

SVD is a powerful tool for **reducing the number of features** in a dataset while keeping most of its important information. This is especially useful when dealing with high-dimensional data (lots of features).

In **Principal Component Analysis (PCA)**, SVD is applied to the data matrix to find directions (called principal components) that capture the most variation in the data. The singular values in the Σ matrix tell us how much "information" each principal component carries.

Why it matters:

- Improve computation speed
- Remove noise from data
- Visualize high-dimensional datasets

Example:

Suppose we have a dataset:

$$A = \begin{bmatrix} 4 & 2 & 3 \\ 2 & 6 & 8 \\ 3 & 7 & 9 \end{bmatrix}$$

We perform SVD to get:

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

By keeping only the top singular values (from Σ), we can build a lower-dimensional approximation of A. This reduced version captures the essence of the original data with fewer features.

2. Recommender Systems

SVD is widely used in **collaborative filtering**, which is the foundation of many recommendation engines (e.g., Netflix, Spotify, Amazon).

In a typical **user-item rating matrix**, rows represent users, columns represent items (like movies), and entries represent ratings. SVD helps us uncover hidden patterns (latent factors) — like genre preferences or user taste — by factorizing the matrix.

How it helps:

- Fills in missing ratings
- Suggests items a user is likely to enjoy
- Handles large, sparse datasets efficiently

Example:

Let R be a user-item matrix (users as rows, items as columns). We compute:

$$R \approx U\Sigma V^T$$

This approximation lets us estimate missing ratings and recommend items based on similar users or items — even if a user hasn't rated many things.

3. Latent Semantic Analysis (LSA)

In **Natural Language Processing (NLP)**, SVD is used in **Latent Semantic Analysis (LSA)** to find patterns in the relationships between terms and documents. This is done using a **document-term matrix**, where rows represent documents and columns represent words.

SVD helps reduce this matrix to reveal hidden topics and semantic relationships.

Applications of LSA:

- Information retrieval (search engines)
- Document classification
- Topic modeling

Example:

Let D be a document-term matrix. After SVD:

$$D = U\Sigma V^T$$

- U: captures document-topic relationships
- Σ: shows topic importance
- V^T: captures word-topic relationships

By keeping only the top k singular values and vectors, we reduce noise and keep only the most meaningful topics. This makes LSA effective in understanding the "theme" of documents, even if different words are used to describe the same idea.

These examples show how SVD isn't just a math tool — it's a bridge between raw data and deeper insights in areas like machine learning, data science, and language understanding.

3. Summary of SVD

SVD is a powerful and versatile tool in linear algebra, used in a wide range of applications such as dimensionality reduction, recommender systems, and latent semantic analysis. The key points are:

- SVD decomposes any matrix into three components: $A = U\Sigma V^T$.
- It is widely used in **dimensionality reduction**, where the top singular values and vectors are used to approximate the data with fewer dimensions.
- SVD is central to **recommender systems**, where it helps predict missing values in user-item matrices.
- **Latent Semantic Analysis (LSA)** uses SVD to uncover hidden semantic structures in text data, which is useful for document retrieval and topic modeling.

Orthogonality and Orthonormality

1. Orthogonal Vectors and Matrices

Orthogonal Vectors

Two vectors $\mathbf{u}$ and $\mathbf{v}$ are said to be **orthogonal** if their dot product is zero:

$$\mathbf{u} \cdot \mathbf{v} = 0$$

Example:

$$\mathbf{u} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \quad \mathbf{v} = \begin{bmatrix} -2 \\ 1 \end{bmatrix}$$

$$\mathbf{u} \cdot \mathbf{v} = 1(-2) + 2(1) = -2 + 2 = 0$$

Since the dot product is zero, the vectors are orthogonal.

Orthogonal Matrices

A matrix A is **orthogonal** if its columns (or rows) form an orthonormal set. This means:

$$A^T A = I \quad \text{or} \quad AA^T = I$$

where I is the identity matrix. An important property of orthogonal matrices is that the inverse is equal to the transpose:

$$A^{-1} = A^T$$

Example:

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

Verify orthogonality:

$$A^T A = I$$

Example of Orthogonal Matrix

Consider:

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

Check orthogonality:

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

Thus, A is an orthogonal matrix.

2. Orthogonal Projections

An **orthogonal projection** of a vector $\mathbf{v}$ onto a vector $\mathbf{u}$ is the component of $\mathbf{v}$ that lies in the direction of $\mathbf{u}$.

$$\text{Proj}_{\mathbf{u}}\mathbf{v} = \frac{\mathbf{v} \cdot \mathbf{u}}{\mathbf{u} \cdot \mathbf{u}}\mathbf{u}$$

Example:

Let:

$$\mathbf{v} = \begin{bmatrix} 3 \\ 4 \end{bmatrix}, \quad \mathbf{u} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

Compute the dot products:

$$\mathbf{v} \cdot \mathbf{u} = 3(1) + 4(2) = 11$$
$$\mathbf{u} \cdot \mathbf{u} = 1^2 + 2^2 = 5$$

Then:

$$\text{Proj}_{\mathbf{u}}\mathbf{v} = \frac{11}{5} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} \frac{11}{5} \\ \frac{22}{5} \end{bmatrix}$$

3. Gram-Schmidt Process

The **Gram-Schmidt process** transforms a set of linearly independent vectors into an orthonormal set.

Given $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$, the process produces $\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n\}$ where each $\mathbf{u}_i$ is orthogonal to the previous ones.

Steps:

1. Set $\mathbf{u}_1 = \mathbf{v}_1$
2. For $k = 2$ to n:
$$\mathbf{u}_k = \mathbf{v}_k - \sum_{i=1}^{k-1} \text{Proj}_{\mathbf{u}_i} \mathbf{v}_k$$
3. Normalize each $\mathbf{u}_k$:
$$\mathbf{u}_k = \frac{\mathbf{u}_k}{\|\mathbf{u}_k\|}$$

Example:

Given:

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

Step 1: $\mathbf{u}_1 = \mathbf{v}_1$
Step 2:

$$\text{Proj}_{\mathbf{u}_1} \mathbf{v}_2 = \frac{1}{2} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$\mathbf{u}_2 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} - \frac{1}{2} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} \\ -\frac{1}{2} \end{bmatrix}$$

Normalize:

$$\|\mathbf{u}_2\| = \frac{\sqrt{2}}{2}, \quad \mathbf{u}_2 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{bmatrix}$$

Summary

- **Orthogonal Vectors**: Two vectors are orthogonal if their dot product is zero.
- **Orthogonal Matrices**: A matrix is orthogonal if $A^T A = I$; it preserves lengths and angles.
- **Orthogonal Projections**: Projects a vector onto the direction of another vector.
- **Gram-Schmidt Process**: Turns a set of linearly independent vectors into an orthonormal set.

Projections and Subspaces

1. Column Space, Row Space, and Null Space

Column Space (Range of A)

The **column space** of a matrix A is the span of its columns. It represents all the possible linear combinations of the columns of A. The column space is a subspace of $\mathbb{R}^m$, where A is an $m \times n$ matrix.

For example, for a matrix:

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

The column space is the span of the vectors:

$$\mathbf{c}_1 = \begin{bmatrix} 1 \\ 3 \\ 5 \end{bmatrix}, \quad \mathbf{c}_2 = \begin{bmatrix} 2 \\ 4 \\ 6 \end{bmatrix}$$

These vectors span a subspace in $\mathbb{R}^3$.

Row Space

The **row space** of a matrix A is the span of the rows of A. The row space is a subspace of $\mathbb{R}^n$ (where A is $m \times n$).

For example, for the matrix A above, the row space is the span of the vectors:

$$\mathbf{r}_1 = \begin{bmatrix} 1 & 2 \end{bmatrix}, \quad \mathbf{r}_2 = \begin{bmatrix} 3 & 4 \end{bmatrix}, \quad \mathbf{r}_3 = \begin{bmatrix} 5 & 6 \end{bmatrix}$$

These vectors form the row space in $\mathbb{R}^2$.

Null Space

The **null space** of a matrix A consists of all the vectors $\mathbf{x}$ such that:

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

It represents all the solutions to the homogeneous equation $A\mathbf{x} = \mathbf{0}$. The null space is a subspace of $\mathbb{R}^n$, where A is $m \times n$.

For example, for the matrix:

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

The null space is the set of all vectors $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ that satisfy:

$$\begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \mathbf{0}$$

This gives the equation $x_1 + 2x_2 = 0$, so the null space is the set of all scalar multiples of the vector $\begin{bmatrix} -2 \\ 1 \end{bmatrix}$.

---

2. Projections onto Subspaces

A **projection** of a vector $\mathbf{v}$ onto a subspace W is the vector in W that is closest to $\mathbf{v}$. The projection is the orthogonal projection onto the subspace.

For example, the projection of $\mathbf{v}$ onto the subspace spanned by a vector $\mathbf{u}$ is given by:

$$\text{Proj}_{\mathbf{u}}\mathbf{v} = \frac{\mathbf{v} \cdot \mathbf{u}}{\mathbf{u} \cdot \mathbf{u}}\mathbf{u}$$

For a subspace spanned by multiple vectors, the projection is the sum of the projections onto each vector.

Example: Projection onto a Line

Let $\mathbf{v} = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$ and $\mathbf{u} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$. The projection of $\mathbf{v}$ onto the subspace spanned by $\mathbf{u}$ is:

$$\text{Proj}_{\mathbf{u}}\mathbf{v} = \frac{\mathbf{v} \cdot \mathbf{u}}{\mathbf{u} \cdot \mathbf{u}}\mathbf{u}$$

We already computed:

$$\mathbf{v} \cdot \mathbf{u} = 3(1) + 4(2) = 11$$

$$\mathbf{u} \cdot \mathbf{u} = 1^2 + 2^2 = 5$$

Thus:

$$\text{Proj}_{\mathbf{u}} \mathbf{v} = \frac{11}{5} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} \frac{11}{5} \\ \frac{22}{5} \end{bmatrix}$$

—

3. Applications in Linear Regression

In **linear regression**, the goal is to determine the best-fitting line (or hyperplane in higher dimensions) that minimizes the sum of squared differences between the observed values and the predicted values.

This solution can be viewed as a projection of the response vector $\mathbf{y}$ onto the column space of the feature matrix X. The projected vector gives the predicted values.

Mathematical Formulation

The linear regression model is given by:

$$\mathbf{y} = X\beta + \epsilon$$

Where:

- $\mathbf{y} \in \mathbb{R}^n$ is the vector of observed values,
- $X \in \mathbb{R}^{n \times p}$ is the matrix of input features,
- $\beta \in \mathbb{R}^p$ is the vector of coefficients,
- $\epsilon \in \mathbb{R}^n$ is the error term.

The least squares solution minimizes the squared error and is given by the projection:

$$\hat{\mathbf{y}} = X(X^T X)^{-1} X^T \mathbf{y}$$

Here, $\hat{\mathbf{y}}$ is the orthogonal projection of $\mathbf{y}$ onto the column space of X, and it represents the predicted values.

Example

Let:

$$X = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$$

Step-by-step:

$$X^T X = \begin{bmatrix} 3 & 6 \\ 6 & 14 \end{bmatrix}, \quad (X^T X)^{-1} = \begin{bmatrix} 7 & -3 \\ -3 & 1.5 \end{bmatrix}$$

Then:

$$X^T \mathbf{y} = \begin{bmatrix} 5 \\ 12 \end{bmatrix}, \quad \hat{\beta} = (X^T X)^{-1} X^T \mathbf{y} = \begin{bmatrix} 7 & -3 \\ -3 & 1.5 \end{bmatrix} \begin{bmatrix} 5 \\ 12 \end{bmatrix} = \begin{bmatrix} 1 \\ 0.5 \end{bmatrix}$$

Now compute:

$$\hat{\mathbf{y}} = X\hat{\beta} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 0.5 \end{bmatrix} = \begin{bmatrix} 1.5 \\ 2.0 \\ 2.5 \end{bmatrix}$$

So, the best-fit values $\hat{\mathbf{y}}$ are $[1.5, 2.0, 2.5]^T$, which represent the orthogonal projection of $\mathbf{y}$ onto the column space of X.

Summary

- Linear regression solves for the projection of $\mathbf{y}$ onto the column space of X.
- The solution minimizes the squared residuals $\|\mathbf{y} - X\beta\|^2$.
- This solution can be efficiently computed using matrix decompositions like QR or SVD.

Matrix Factorization Methods: LU and QR Decompositions

1. LU Decomposition

LU Decomposition is the factorization of a square matrix A into the product of a lower triangular matrix L and an upper triangular matrix U, such that:

$$A = LU$$

Where:

- L is a lower triangular matrix with ones on the diagonal.
- U is an upper triangular matrix.

Step-by-Step LU Decomposition:
1. Start with the matrix A:

$$A = \begin{bmatrix} 4 & -2 & 1 \\ 20 & -7 & 12 \\ -8 & 13 & 17 \end{bmatrix}$$

2. Gaussian elimination: Perform row operations to convert A into an upper triangular matrix U.

- First, use row 1 to eliminate the entries below the diagonal in the first column:

$$R_2 \to R_2 - 5R_1$$

$$R_3 \to R_3 + 2R_1$$

After applying the above operations, the matrix becomes:

$$U = \begin{bmatrix} 4 & -2 & 1 \\ 0 & 3 & 7 \\ 0 & 0 & 11 \end{bmatrix}$$

3. Construct the lower triangular matrix L: - The multipliers used during Gaussian elimination form the entries below the diagonal in L. - The diagonal of L is filled with ones.

From the row operations, we obtain:

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

4. Reconstruct the matrix A: Finally, the matrix A is the product of L and U:

$$A = L \cdot U = \begin{bmatrix} 1 & 0 & 0 \\ 5 & 1 & 0 \\ -2 & 3 & 1 \end{bmatrix} \begin{bmatrix} 4 & -2 & 1 \\ 0 & 3 & 7 \\ 0 & 0 & 11 \end{bmatrix}$$

Multiplying these matrices:

$$A = \begin{bmatrix} 4 & -2 & 1 \\ 20 & -7 & 12 \\ -8 & 13 & 17 \end{bmatrix}$$

So that $A = LU$, as expected.

Applications of LU Decomposition

- Solving linear systems $A\mathbf{x} = \mathbf{b}$
- Computing determinants: $\det(A) = \det(L)\det(U)$
- Matrix inversion (when A is invertible)

2. QR Decomposition

QR Decomposition expresses a matrix A as the product of an orthogonal matrix Q and an upper triangular matrix R:

$$A = QR$$

Where:

- Q is an orthogonal matrix ($Q^TQ = I$)
- R is an upper triangular matrix

Step-by-Step QR Decomposition:
1. Start with the matrix A:

$$A = \begin{bmatrix} 12 & -51 & 4 \\ 6 & 167 & -68 \\ -4 & 24 & -41 \end{bmatrix}$$

2. **Apply the Gram-Schmidt Process** to orthogonalize the columns of A to form Q.
- Let the columns of A be $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$.
- **Step 1:** Normalize $\mathbf{a}_1$ to form $\mathbf{q}_1$:

$$\mathbf{q}_1 = \frac{\mathbf{a}_1}{\|\mathbf{a}_1\|}$$

where $\|\mathbf{a}_1\|$ is the norm (magnitude) of the vector $\mathbf{a}_1$.
- **Step 2:** Compute $\mathbf{q}_2$ by removing the projection of $\mathbf{a}_2$ on $\mathbf{q}_1$:

$$\mathbf{q}_2 = \frac{\mathbf{a}_2 - (\mathbf{a}_2^T\mathbf{q}_1)\mathbf{q}_1}{\|\mathbf{a}_2 - (\mathbf{a}_2^T\mathbf{q}_1)\mathbf{q}_1\|}$$

- **Step 3:** Compute $\mathbf{q}_3$ by removing the projections of $\mathbf{a}_3$ on both $\mathbf{q}_1$ and $\mathbf{q}_2$:

$$\mathbf{q}_3 = \frac{\mathbf{a}_3 - (\mathbf{a}_3^T\mathbf{q}_1)\mathbf{q}_1 - (\mathbf{a}_3^T\mathbf{q}_2)\mathbf{q}_2}{\|\mathbf{a}_3 - (\mathbf{a}_3^T\mathbf{q}_1)\mathbf{q}_1 - (\mathbf{a}_3^T\mathbf{q}_2)\mathbf{q}_2\|}$$

After performing the Gram-Schmidt process, we obtain the matrix Q:

$$Q = \begin{bmatrix} -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \end{bmatrix}$$

3. **Compute the upper triangular matrix R:**
- R is computed by projecting the columns of A onto the columns of Q:

$$R = Q^TA$$

The result is the upper triangular matrix R:

$$R = \begin{bmatrix} -\sqrt{2} & \sqrt{2} & 0 \\ 0 & 12 & 2 \\ 0 & 0 & 5 \end{bmatrix}$$

4. **Reconstruct the matrix A:**

Finally, we verify the decomposition by multiplying Q and R to get back the original matrix A:

$$A = Q \cdot R = \begin{bmatrix} -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} -\sqrt{2} & \sqrt{2} & 0 \\ 0 & 12 & 2 \\ 0 & 0 & 5 \end{bmatrix}$$

Multiplying these matrices:

$$A = \begin{bmatrix} 12 & -51 & 4 \\ 6 & 167 & -68 \\ -4 & 24 & -41 \end{bmatrix}$$

So that $A = QR$, as expected.

Applications of QR Decomposition

- Solving least squares problems
- Eigenvalue computations (QR algorithm)
- Numerically stable matrix factorization

3. LU vs QR Decomposition

- **LU Decomposition** is efficient for solving systems when A is square and non-singular.
- **QR Decomposition** is better suited for rectangular matrices and least squares problems, and offers better numerical stability.

Summary

- **LU Decomposition:** $A = LU$, with L lower triangular and U upper triangular.
- **QR Decomposition:** $A = QR$, with Q orthogonal and R upper triangular.