Matrices

A matrix is a rectangular array of numbers (or elements) arranged in rows and columns. Matrices are fundamental in mathematics and are used extensively in various fields, including linear algebra, calculus, physics, and computer science.

General Form:

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

where A is an $m \times n$ matrix with elements a_{ij}, where i represents the row index and j represents the column index.

Vectors

A vector is a one-dimensional array of elements. It represents a quantity that has both magnitude and direction, such as force or velocity.

- **Column Vector:**

$$\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$$

  where $\mathbf{v}$ is an $n \times 1$ column vector with elements v_i.

- **Row Vector:**

$$\mathbf{w} = [w_1, w_2, \ldots, w_n]$$

  where $\mathbf{w}$ is a $1 \times n$ row vector.

Operations on Matrices

1. Addition and Subtraction

- **Definition**: Two matrices of the same dimensions can be added or subtracted element-wise.
- **Properties**:
  - Commutative: $A + B = B + A$
  - Associative: $(A + B) + C = A + (B + C)$
  - Identity: $A + 0 = A$

Operations on Matrices

1. Addition and Subtraction

- **Definition**: Two matrices of the same dimensions can be added or subtracted element-wise.
- **Properties**:
  - Commutative: $A + B = B + A$
  - Associative: $(A + B) + C = A + (B + C)$
  - Identity: $A + 0 = A$

2. Scalar Multiplication

- **Definition**: Multiplying each element of a matrix A by a scalar c.
- **Properties**:
  - Distributive: $c(A + B) = cA + cB$
  - Associative: $(cd)A = c(dA)$
  - Identity: $1A = A$

3. Matrix Multiplication

- **Definition**: The product of two matrices A (of size $m \times n$) and B (of size $n \times p$) is a matrix C (of size $m \times p$) where each element c_{ij} is given by the dot product of the i-th row of A and the j-th column of B.
- **Properties**:
  - Not Commutative: $AB \neq BA$ in general
  - Associative: $(AB)C = A(BC)$
  - Distributive: $A(B + C) = AB + AC$ and $(A + B)C = AC + BC$
  - Identity: $AI = IA = A$ where I is the identity matrix

4. Transpose

- **Definition**: The transpose of a matrix A (denoted A^T) is formed by swapping its rows and columns.

$$(A^T)_{ij} = A_{ji}$$

- **Properties**:
  - $(A^T)^T = A$
  - $(A + B)^T = A^T + B^T$
  - $(cA)^T = cA^T$
  - $(AB)^T = B^T A^T$

Trace

The **trace** of a square matrix A, denoted as $\mathrm{tr}(A)$, is the sum of the elements on its main diagonal (top-left to bottom-right).

Example:

For the matrix $A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$,

$\mathrm{tr}(A) = 1 + 4 = 5$

So, $\mathrm{tr}(A) = 5$.

Cofactor Formula

The **cofactor** of an element a_{ij} in a matrix A, denoted as C_{ij}, is given by $(-1)^{i+j}$ times the determinant of the matrix obtained by deleting the i-th row and j-th column from A.

Example:

Let's find the cofactor C_{12} for the matrix $A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$.

To find C_{12}:

1. Delete the 1st row and 2nd column from A, resulting in $\begin{pmatrix} 3 \end{pmatrix}$.
2. Compute $\det(\begin{pmatrix} 3 \end{pmatrix}) = 3$.

Since C_{12} corresponds to $a_{12} = 2$,
$C_{12} = (-1)^{1+2} \cdot 3 = -3$

So, $C_{12} = -3$.

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

To find the cofactor C_{ij} of an element a_{ij}, we follow these steps:

1. **Remove the i-th row and j-th column from A.
2. **Compute the determinant of the remaining 2×2 submatrix.
3. **Apply the sign $(-1)^{i+j}$ to the determinant.

Let's find the cofactors for each element in A.

The cofactor matrix C is:

$$C = \begin{bmatrix} 24 & 5 & -4 \\ -12 & 3 & 2 \\ -2 & -5 & 4 \end{bmatrix}$$

Summary Table:

Element a_{ij}	Submatrix	Determinant	Sign	Cofactor C_{ij}
$a_{11} = 1$	$\begin{bmatrix} 4 & 5 \\ 0 & 6 \end{bmatrix}$	24	$(+)$	24
$a_{12} = 2$	$\begin{bmatrix} 0 & 5 \\ 1 & 6 \end{bmatrix}$	-5	$(-)$	5
$a_{13} = 3$	$\begin{bmatrix} 0 & 4 \\ 1 & 0 \end{bmatrix}$	-4	$(+)$	-4
$a_{21} = 0$	$\begin{bmatrix} 2 & 3 \\ 0 & 6 \end{bmatrix}$	12	$(-)$	-12
$a_{22} = 4$	$\begin{bmatrix} 1 & 3 \\ 1 & 6 \end{bmatrix}$	3	$(+)$	3
$a_{23} = 5$	$\begin{bmatrix} 1 & 2 \\ 1 & 0 \end{bmatrix}$	-2	$(-)$	2
$a_{31} = 1$	$\begin{bmatrix} 2 & 3 \\ 4 & 5 \end{bmatrix}$	-2	$(+)$	-2
$a_{32} = 0$	$\begin{bmatrix} 1 & 3 \\ 0 & 5 \end{bmatrix}$	5	$(-)$	-5
$a_{33} = 6$	$\begin{bmatrix} 1 & 2 \\ 0 & 4 \end{bmatrix}$	4	$(+)$	4

Special Types of Matrices

1. Square Matrix

- **Definition**: A matrix with the same number of rows and columns ($n \times n$).
- **Properties**:
  - Determinant is defined.
  - Inverse may exist if the determinant is non-zero.

2. Diagonal Matrix

- **Definition**: A square matrix where all off-diagonal elements are zero.

$$D = \begin{bmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_n \end{bmatrix}$$

- **Properties**:
  - $D^T = D$
  - D^{-1} exists if all diagonal elements are non-zero.

3. Identity Matrix

- **Definition**: A diagonal matrix where all diagonal elements are 1.

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

- **Properties**:
  - $AI = IA = A$ for any matrix A.

4. Zero Matrix

- **Definition**: A matrix where all elements are zero.
- **Properties**:
  - $A + 0 = A$
  - $A0 = 0$

5. Symmetric Matrix

- **Definition**: A matrix that is equal to its transpose.

$$A = A^T$$

- **Properties**:
  - All eigenvalues are real.
  - Can be diagonalized by an orthogonal matrix.

6. Skew-Symmetric Matrix

- **Definition**: A square matrix that is equal to the negative of its transpose.

$$A = -A^T$$

- **Properties**:
  - Diagonal elements are zero.

7. Orthogonal Matrix

- **Definition**: A square matrix Q whose rows and columns are orthonormal vectors.

$$Q^T Q = QQ^T = I$$

- **Properties**:
  - $Q^{-1} = Q^T$
  - Preserves lengths and angles.

8. Invertible (Non-Singular) Matrix

- **Definition**: A square matrix A that has an inverse such that $AA^{-1} = A^{-1}A = I$.
- **Properties**:
  - A is invertible if and only if $\det(A) \neq 0$.

9. Singular Matrix

- **Definition**: A square matrix that does not have an inverse.
- **Properties**:
  - $\det(A) = 0$.

10. Row Matrix

- **Definition**: A matrix with only one row.

$$R = \begin{bmatrix} r_1 & r_2 & \cdots & r_n \end{bmatrix}$$

- **Properties**:
  - It has dimensions $1 \times n$.

11. Column Matrix

- **Definition**: A matrix with only one column.

$$C = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_m \end{bmatrix}$$

- **Properties**:
  - It has dimensions $m \times 1$.

12. Upper Triangular Matrix

- **Definition**: A square matrix where all elements below the main diagonal are zero.

$$U = \begin{bmatrix} u_{11} & u_{12} & \cdots & u_{1n} \\ 0 & u_{22} & \cdots & u_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & u_{nn} \end{bmatrix}$$

- **Properties**:
  - U is invertible if and only if all diagonal elements are non-zero.

13. Lower Triangular Matrix

- **Definition**: A square matrix where all elements above the main diagonal are zero.

$$L = \begin{bmatrix} l_{11} & 0 & \cdots & 0 \\ l_{21} & l_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ l_{n1} & l_{n2} & \cdots & l_{nn} \end{bmatrix}$$

- **Properties**:
  - L is invertible if and only if all diagonal elements are non-zero.

14. Diagonalizable Matrix

- **Definition**: A square matrix A is diagonalizable if it can be written as $A = PDP^{-1}$ where D is a diagonal matrix and P is an invertible matrix.
- **Properties**:
  - All eigenvalues of A are on the diagonal of D.
  - A is diagonalizable if it has n linearly independent eigenvectors.

15. Idempotent Matrix

- **Definition**: A matrix A such that $A^2 = A$.
- **Properties**:
  - If A is idempotent, then A has eigenvalues 0 or 1.

Positive Definite Matrix

Definition

A matrix A is positive definite if it is a symmetric matrix and all its eigenvalues are positive. Formally, a symmetric matrix A is positive definite if for any non-zero vector x,

$$x^T Ax > 0.$$

Properties

1. **Symmetry**: A must be symmetric.
2. **Positive Eigenvalues**: All eigenvalues of A are positive.
3. **Positive Diagonal Elements**: All the leading principal minors of A are positive.
4. **Cholesky Decomposition**: A can be decomposed as $A = LL^T$, where L is a lower triangular matrix with positive diagonal elements.
5. **Quadratic Form**: The quadratic form $x^T Ax$ is always positive for any non-zero vector x.

Positive Semi-Definite Matrix

Definition

A matrix A is positive semi-definite if it is a symmetric matrix and all its eigenvalues are non-negative. Formally, a symmetric matrix A is positive semi-definite if for any vector x,

$$x^T Ax \geq 0.$$

Properties

1. **Symmetry**: A must be symmetric.
2. **Non-Negative Eigenvalues**: All eigenvalues of A are non-negative.
3. **Non-Negative Quadratic Form**: The quadratic form $x^T Ax$ is always non-negative for any vector x.

Tests for Positive Definite and Positive Semi-Definite Matrices

1. **Eigenvalues**:
   - Compute the eigenvalues of the matrix. If all are positive, the matrix is positive definite. If all are non-negative, the matrix is positive semi-definite.
2. **Principal Minors**:
   - Check the leading principal minors (determinants of the upper-left $k \times k$ submatrices for $k = 1, 2, ..., n$).
     - For positive definite, all leading principal minors must be positive.
     - For positive semi-definite, all leading principal minors must be non-negative.
3. **Cholesky Decomposition**:
   - Attempt a Cholesky decomposition. If it succeeds and produces a lower triangular matrix with positive diagonal elements, the matrix is positive definite.

Theorems and Properties

1. Rank-Nullity Theorem

- **Statement**: For any $m \times n$ matrix A,

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

- **Explanation**: The rank of a matrix is the dimension of the column space, and the nullity is the dimension of the null space.

2. Inverse of a Matrix

- **Theorem**: If A is an invertible $n \times n$ matrix, then A^{-1} is also an $n \times n$ matrix, and $(A^{-1})^{-1} = A$.
- **Properties**:
  - $(AB)^{-1} = B^{-1}A^{-1}$
  - $(A^T)^{-1} = (A^{-1})^T$

3. Determinant Properties

- **Properties**:
  - $\det(A^T) = \det(A)$
  - $\det(AB) = \det(A)\det(B)$
  - If A is invertible, $\det(A^{-1}) = \frac{1}{\det(A)}$

Elementary Row Operations

Elementary row operations are operations that can be performed on the rows of a matrix without changing its row equivalence. There are three types of elementary row operations:

1. **Row Swapping**:
   - Interchanging two rows of a matrix.
   - Denoted as $R_i \leftrightarrow R_j$, where R_i and R_j are the i-th and j-th rows of the matrix, respectively.
2. **Row Multiplication**:
   - Multiplying a row by a non-zero scalar.
   - Denoted as kR_i, where k is a non-zero scalar and R_i is the i-th row.
3. **Row Addition**:
   - Adding a multiple of one row to another row.
   - Denoted as $R_i + kR_j \rightarrow R_i$, where k is a scalar, R_j is the j-th row, and R_i is the row being replaced.

Row Echelon Form (REF)

A matrix is in Row Echelon Form (REF) if it satisfies the following conditions:

1. **Non-Zero Rows**: All non-zero rows are above any rows of all zeros.
2. **Leading Entries**: The leading entry of each non-zero row (called the pivot) is 1.
3. **Pivot Position**: The pivot of each non-zero row is to the right of the pivot of the row above it.
4. **Zero Rows**: Any rows of all zeros are at the bottom of the matrix.

Example

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

Reduced Row Echelon Form (RREF)

A matrix is in Reduced Row Echelon Form (RREF) if it satisfies the following conditions in addition to those of REF:

1. **Leading 1s**: The leading entry in each non-zero row is 1.
2. **Zero Columns**: Each leading 1 is the only non-zero entry in its column.

Example

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

Steps to Convert a Matrix to Reduced Row Echelon Form (RREF)

1. **First Convert to REF**:
   - Follow the steps to convert the matrix to REF.
2. **Normalize Pivots**:
   - Ensure each pivot is 1 (if not already).
3. **Eliminate Entries Above Each Pivot**:
   - Use row addition to create zeros above each pivot.
4. **Continue for All Pivots**:
   - Repeat the process for each column until the entire matrix is in RREF.

Converting to REF and RREF

Steps to Convert a Matrix to Row Echelon Form (REF)

1. **Start with the First Column**:
    - Identify the first non-zero element in the first column (the pivot) and move the row containing it to the top (row swapping, if necessary).
    - Normalize the row by dividing it by the pivot (row multiplication).
2. **Eliminate Entries Below the Pivot**:
    - Use row addition to create zeros below the pivot.
3. **Move to the Next Column**:
    - Repeat the process for the submatrix that starts one row below and one column to the right.
4. **Continue Until All Pivots Are Identified**:
    - Repeat steps for all columns until the entire matrix is in REF.

Rank of a Matrix

The rank of a matrix is the maximum number of linearly independent rows or columns in the matrix. It can be found by transforming the matrix into its row echelon form (REF) or reduced row echelon form (RREF) and counting the number of non-zero rows.

Example: 3x3 Matrix

Given matrix:

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

To find the rank, we can reduce it to REF:

1. The first row is fine.

2. Eliminate the elements below the first pivot:

$$R2 \leftarrow R2 - 4R1 \quad \Rightarrow \quad \begin{bmatrix} 1 & 2 & 3 \\ 0 & -3 & -6 \\ 7 & 8 & 9 \end{bmatrix}$$

$$R3 \leftarrow R3 - 7R1 \quad \Rightarrow \quad \begin{bmatrix} 1 & 2 & 3 \\ 0 & -3 & -6 \\ 0 & -6 & -12 \end{bmatrix}$$

3. Eliminate the element below the second pivot:

$$R3 \leftarrow R3 + 2R2 \quad \Rightarrow \quad \begin{bmatrix} 1 & 2 & 3 \\ 0 & -3 & -6 \\ 0 & 0 & 0 \end{bmatrix}$$

The matrix in REF:

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

Since there are two non-zero rows, the rank of matrix B is 2.

Determinant

The determinant of a matrix is a scalar value that is a function of the entries of a square matrix. It provides important properties about the matrix, such as whether the matrix is invertible.

Example: 2x2 Matrix

Given matrix:

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

The determinant of A:

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

Example: 3x3 Matrix

Given matrix:

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

The determinant of B:

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

Minor

The minor of an element of a matrix is the determinant of the submatrix formed by deleting the row and column of that element.

Example: 3x3 Matrix

Given matrix:

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

To find the minor of the element c_{22} (element in the 2nd row, 2nd column, which is 4):

1. Remove the 2nd row and 2nd column:

$$\begin{bmatrix} 1 & 3 \\ 1 & 6 \end{bmatrix}$$

Adjoint Matrix

The adjoint (or adjugate) of a matrix A is the transpose of its cofactor matrix. The adjoint is used in finding the inverse of a matrix.

Steps to Find the Adjoint Matrix:

1. **Find the Cofactor Matrix**: Compute the cofactor for each element in the matrix.
2. **Transpose the Cofactor Matrix**: Transpose the matrix obtained in the first step to get the adjoint matrix.

Example for a 2×2 Matrix

Consider the matrix:

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

1. **Find the Cofactors**:
   - $C_{11} = d$
   - $C_{12} = -c$
   - $C_{21} = -b$
   - $C_{22} = a$

   Cofactor Matrix:

$$\text{Cof}(A) = \begin{bmatrix} d & -c \\ -b & a \end{bmatrix}$$

2. **Transpose the Cofactor Matrix**:

$$\text{Adj}(A) = \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

So, the adjoint of A is:

$$\text{Adj}(A) = \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

Example for a 3×3 Matrix

Consider the matrix:

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

1. **Find the Cofactors:**

   The cofactor matrix C is:

$$C = \begin{bmatrix} 24 & 5 & -4 \\ -12 & 3 & 2 \\ -2 & -5 & 4 \end{bmatrix}$$

2. **Transpose the Cofactor Matrix:**

$$\text{Adj}(A) = C^T = \begin{bmatrix} 24 & -12 & -2 \\ 5 & 3 & -5 \\ -4 & 2 & 4 \end{bmatrix}$$

So, the adjoint of A is:

$$\text{Adj}(A) = \begin{bmatrix} 24 & -12 & -2 \\ 5 & 3 & -5 \\ -4 & 2 & 4 \end{bmatrix}$$

Inverse of a Matrix

The inverse of a matrix is a fundamental concept in linear algebra, denoted as A^{-1}, which, when multiplied by the original matrix A, gives the identity matrix I. This concept is essential for solving systems of linear equations, computing solutions to optimization problems, and more. Let's explore how to calculate the inverse of a matrix and the different methods involved.

Given a square matrix A, the inverse A^{-1} exists if and only if A is non-singular (determinant $\det(A) \neq 0$). The inverse satisfies the following properties:

1. $AA^{-1} = A^{-1}A = I$, where I is the identity matrix of the same size as A.
2. $(A^{-1})^{-1} = A$.

Methods of Calculating the Inverse

1. Adjoint Method

For a 2×2 matrix $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, the inverse can be calculated using the formula:

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

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

2. Row Reduction (Gaussian Elimination)

For larger matrices, row reduction can be used to find the inverse. This method involves augmenting the original matrix with the identity matrix of the same size and then performing row operations until the original matrix is transformed into the identity matrix. The augmented identity matrix will then have the inverse of the original matrix on the right-hand side.

3. Matrix of Minors, Cofactor Method

For a 3×3 or larger matrix A, the inverse can be found using the matrix of minors and cofactors:

1. **Matrix of Minors**: Replace each element of A with its corresponding minor (determinant of the submatrix obtained by deleting the row and column of the element).
2. **Matrix of Cofactors**: Multiply each element of the matrix of minors by $(-1)^{i+j}$, where i and j are the row and column indices, respectively.
3. **Adjoint Matrix**: Transpose the matrix of cofactors to obtain the adjoint $\text{adj}(A)$.
4. **Inverse**: Multiply the adjoint matrix by $\frac{1}{\det(A)}$, where $\det(A)$ is the determinant of A.

$$A^{-1} = \frac{1}{\det(A)} \cdot \text{adj}(A)$$

4. Using Elementary Matrices

An elementary matrix is obtained by performing a single elementary row operation on an identity matrix. The inverse of A can be computed using the product of elementary matrices that transform A into I.

$$A^{-1} = E_k \cdot E_{k-1} \cdot \ldots \cdot E_2 \cdot E_1$$

Linear System

A linear system of equations is a collection of one or more linear equations involving the same set of variables. Generally, it can be represented in matrix form as:

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

where:

- A is an $m \times n$ matrix of coefficients,
- $\mathbf{x}$ is an $n \times 1$ vector of variables (unknowns),
- $\mathbf{b}$ is an $m \times 1$ vector of constants (right-hand side).

The goal is to find the vector $\mathbf{x}$ that satisfies all equations simultaneously.

Coefficient Matrix

The coefficient matrix A in a linear system contains only the coefficients of the variables from the left-hand side of the equations. For a system of m linear equations with n variables, the coefficient matrix A is an $m \times n$ matrix.

Augmented Matrix

An **augmented matrix** is a matrix that combines the coefficients of variables from the left-hand side of a system of linear equations with the constants from the right-hand side into a single matrix. This combined matrix allows for systematic manipulation and solution of the system using matrix operations. Let's explore the concept and usage of an augmented matrix with an example.

Gaussian Elimination

Gaussian elimination is a systematic method to transform a matrix into its **Row Echelon Form (REF)** or **Reduced Row Echelon Form (RREF)**. The primary goal is to simplify the augmented matrix of a system of linear equations to make it easier to read off solutions directly.

Steps of Gaussian Elimination:

1. **Augmented Matrix**: Start with the augmented matrix $[A \,|\mathbf{b}]$, where A is the coefficient matrix and $\mathbf{b}$ is the column vector of constants.
2. **Row Operations**: Apply elementary row operations to transform $[A \,|\mathbf{b}]$ into an upper triangular form (REF).
3. **Back Substitution**: Solve for variables starting from the last equation up to the first, using the simplified augmented matrix.
4. **Solution Interpretation**: Determine if the system has a unique solution, no solution, or infinite solutions based on the structure of the transformed matrix.

Gauss-Jordan Elimination

Gauss-Jordan elimination takes Gaussian elimination a step further by transforming the augmented matrix directly into its **Reduced Row Echelon Form (RREF)**. This method ensures that each variable is isolated in each equation, and all solutions are easily identified.

Steps of Gauss-Jordan Elimination:

1. **Augmented Matrix**: Start with the augmented matrix $[A \,|\mathbf{I}]$, where A is the original matrix and $\mathbf{I}$ is the identity matrix of the same order as A.
2. **Row Operations**: Apply elementary row operations to transform $[A \,|\mathbf{I}]$ into RREF.
3. **Solution Extraction**: The solution can be directly read off from the resulting matrix, where the right-hand side will show the inverse of A (if solving for inverse) or the solutions to the system of equations.

Key Differences:

- **Objective**: Gaussian elimination aims to simplify the matrix to solve a system of equations, while Gauss-Jordan elimination focuses on finding the inverse of a matrix or the solutions to the system directly.
- **Final Form**: Gaussian elimination typically results in REF, while Gauss-Jordan elimination results in RREF.
- **Application**: Gaussian elimination is often used for solving systems of equations, while Gauss-Jordan elimination is particularly useful for finding inverses and determining unique solutions efficiently.

Let's discuss the three possible cases that can arise when solving a system of linear equations, based on the behavior of the augmented matrix after performing Gaussian elimination or a similar method.

1. Unique Solution

A system of linear equations has a **unique solution** if and only if:

- Every variable (unknown) in the system can be expressed uniquely in terms of the other variables.
- This corresponds to the augmented matrix reducing to a diagonal form (Reduced Row Echelon Form, RREF), where each leading entry (pivot) is 1 and is the only non-zero entry in its column.

2. No Solution (Inconsistent System)

A system of linear equations has **no solution** (is inconsistent) if and only if:

- The augmented matrix leads to a contradiction, such as a row where the constants on the right-hand side do not balance after eliminating variables.
- This happens when the reduced matrix has a row where the left-hand side reduces to zero, but the corresponding right-hand side is non-zero.

3. Infinite Solutions

A system of linear equations has **infinitely many solutions** if and only if:

- The reduced row echelon form (RREF) of the augmented matrix has at least one row with all zeros on the left-hand side but a non-zero constant on the right-hand side.
- This indicates that one or more variables can be expressed in terms of one or more parameters, resulting in an infinite number of possible combinations.

Linear Dependence

Vectors are **linearly dependent** if one or more of the vectors in a set can be expressed as a linear combination of the others. In other words, at least one vector in the set can be written as a sum of scalar multiples of the remaining vectors.

Mathematical Definition:

Given vectors $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n$, they are linearly dependent if there exist scalars $c_1, c_2, ..., c_n$, not all zero, such that:

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

where $\mathbf{0}$ is the zero vector.

Example:

Consider vectors $\mathbf{v}_1 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$ and $\mathbf{v}_2 = \begin{pmatrix} 2 \\ 4 \end{pmatrix}$.

These vectors are linearly dependent because $\mathbf{v}_2$ can be expressed as $2\mathbf{v}_1$:

$$2\begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} 2 \\ 4 \end{pmatrix}$$

Linear Independence

Vectors are **linearly independent** if none of the vectors in a set can be expressed as a linear combination of the others except in the trivial way (where all coefficients are zero).

Mathematical Definition:

Given vectors $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n$, they are linearly independent if the equation:

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

implies that $c_1 = c_2 = ... = c_n = 0$.

Example:

Consider vectors $\mathbf{v}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $\mathbf{v}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$.

These vectors are linearly independent because the equation $c_1\begin{pmatrix} 1 \\ 0 \end{pmatrix} + c_2\begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ implies $c_1 = 0$ and $c_2 = 0$.

Importance in Linear Algebra

- **Basis**: Linearly independent vectors form the basis of a vector space. They span the space and are the minimum set of vectors required to represent every vector in that space.
- **Dimension**: The number of linearly independent vectors in a set determines the dimension of the vector space they span. For example, two linearly independent vectors span a two-dimensional space.
- **Applications**: Linear independence is fundamental in solving systems of linear equations, computing matrix rank, and understanding transformations in linear algebra applications such as computer graphics, data analysis, and physics.

Checking for Linear Dependence and Independence

1. **Construct the Augmented Matrix**: Form an augmented matrix with the vectors as columns and a zero vector (or column) on the right-hand side.

$$[\mathbf{v}_1 \quad \mathbf{v}_2 \quad \dots \quad \mathbf{v}_n \quad | \quad \mathbf{0}]$$

2. **Perform Gaussian Elimination**: Apply Gaussian elimination to reduce the augmented matrix to its row echelon form (REF) or reduced row echelon form (RREF).

Checking for Linear Dependence

A set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ is linearly dependent if there exist scalars $c_1, c_2, \dots, c_n$, not all zero, such that:

Identify Pivot Positions: If there is at least one column without a pivot position (leading 1) in the REF or RREF, then the corresponding vector is a linear combination of the others, indicating linear dependence.

Check Non-trivial Solutions: If you find non-trivial solutions (where not all coefficients are zero) to the system of equations derived from the augmented matrix, the vectors are linearly dependent.

Checking for Linear Independence

A set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ is linearly independent if the only solution to the equation:

3. **Identify Pivot Positions**: If every column in the REF or RREF has a pivot position (leading 1) and there are no free variables (non-pivot columns without leading 1), then the vectors are linearly independent.

4. **Check for Trivial Solutions**: The only solution to the system derived from the augmented matrix is $c_1 = c_2 = ... = c_n = 0$, indicating linear independence.

Vector Concepts

1. Vectors

Vectors are elements of a vector space and are represented as ordered lists of numbers. In Euclidean space, a vector $\mathbf{v}$ is typically represented as $\mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}$, where v_i are scalar components.

2. Vector Operations

- **Vector Addition**: Adding two vectors component-wise.

$$\mathbf{v} + \mathbf{w} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} + \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{pmatrix} = \begin{pmatrix} v_1 + w_1 \\ v_2 + w_2 \\ \vdots \\ v_n + w_n \end{pmatrix}$$

- **Scalar Multiplication**: Multiplying a vector by a scalar.

$$c\,\mathbf{v} = c\begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} = \begin{pmatrix} cv_1 \\ cv_2 \\ \vdots \\ cv_n \end{pmatrix}$$

3. Linear Combination

A linear combination of vectors $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n$ is a vector of the form $c_1\,\mathbf{v}_1 + c_2\,\mathbf{v}_2 + ... + c_n\,\mathbf{v}_n$, where c_i are scalars. Linear combinations play a crucial role in defining the span of vectors and vector spaces.

4. Span

The span of a set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n\}$ is the set of all possible linear combinations of these vectors. It forms a subspace of the vector space containing those vectors.

5. Linear Independence and Dependence

Vectors are linearly independent if no vector in the set can be expressed as a linear combination of the others. Conversely, vectors are linearly dependent if at least one vector in the set can be expressed as a linear combination of the others.

6. Basis

A basis for a vector space V is a set of linearly independent vectors that span V. It provides a minimal set of vectors necessary to describe every vector in V.

7. Dimension

The dimension of a vector space V is the number of vectors in any basis for V. It represents the maximum number of linearly independent vectors that can exist in V.

8. Inner Product (Dot Product)

An inner product $\langle \mathbf{v}, \mathbf{w} \rangle$ on vectors $\mathbf{v}$ and $\mathbf{w}$ in Euclidean space measures the angle between them and is defined as $\langle \mathbf{v}, \mathbf{w} \rangle = v_1 w_1 + v_2 w_2 + \ldots + v_n w_n$.

9. Orthogonality

Two vectors $\mathbf{v}$ and $\mathbf{w}$ are orthogonal if their inner product $\langle \mathbf{v}, \mathbf{w} \rangle = 0$. Orthogonal vectors play a key role in defining orthogonal bases and orthogonal projections.

10. Norm

The norm $\|\mathbf{v}\|$ of a vector $\mathbf{v}$ measures its magnitude and is defined as $\|\mathbf{v}\| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}$. It generalizes the notion of length or size of a vector.

11. Eigenvalues and Eigenvectors

In the context of matrices (which can be thought of as operators on vectors), eigenvalues λ and corresponding eigenvectors $\mathbf{v}$ satisfy $A\,\mathbf{v} = \lambda\,\mathbf{v}$, where A is a square matrix. They represent special directions (eigenvectors) and scaling factors (eigenvalues) under the action of the matrix A.

12. Orthogonalization

The process of orthogonalization converts a set of linearly independent vectors into a set of orthogonal (or orthonormal) vectors. This simplifies computations and is used in algorithms like Gram-Schmidt orthogonalization.

Vector Spaces

A **vector space** V over a field $\mathbb{F}$ (which is typically $\mathbb{R}$ or $\mathbb{C}$) is a set of elements called vectors, together with two operations:

Vector Addition: For any $\mathbf{v}, \mathbf{w} \in V$, there exists a unique vector $\mathbf{v} + \mathbf{w} \in V$ called the sum of $\mathbf{v}$ and $\mathbf{w}$.

- **Closure**: $\mathbf{v} + \mathbf{w} \in V$ for all $\mathbf{v}, \mathbf{w} \in V$.
- **Associativity**: $(\mathbf{v} + \mathbf{w}) + \mathbf{u} = \mathbf{v} + (\mathbf{w} + \mathbf{u})$ for all $\mathbf{v}, \mathbf{w}, \mathbf{u} \in V$.
- **Identity Element**: There exists a vector $\mathbf{0} \in V$ such that $\mathbf{v} + \mathbf{0} = \mathbf{v}$ for all $\mathbf{v} \in V$.
- **Inverse Element**: For each $\mathbf{v} \in V$, there exists a vector $-\mathbf{v} \in V$ such that $\mathbf{v} + (-\mathbf{v}) = \mathbf{0}$.

Scalar Multiplication: For any scalar $c \in \mathbb{F}$ and any $\mathbf{v} \in V$, there exists a unique vector $c\,\mathbf{v} \in V$.

- **Closure**: $c\,\mathbf{v} \in V$ for all $c \in \mathbb{F}$ and $\mathbf{v} \in V$.
- **Distributivity over Vector Addition**: $c(\mathbf{v} + \mathbf{w}) = c\,\mathbf{v} + c\,\mathbf{w}$ for all $c \in \mathbb{F}$ and $\mathbf{v}, \mathbf{w} \in V$.
- **Distributivity over Scalar Addition**: $(c + d)\mathbf{v} = c\,\mathbf{v} + d\,\mathbf{v}$ for all $c, d \in \mathbb{F}$ and $\mathbf{v} \in V$.
- **Compatibility with Field Multiplication**: $(cd)\mathbf{v} = c(d\,\mathbf{v})$ for all $c, d \in \mathbb{F}$ and $\mathbf{v} \in V$.
- **Identity Scalar**: $1\mathbf{v} = \mathbf{v}$ for all $\mathbf{v} \in V$.

Theorems and Proofs in Vector Spaces

1. **Existence of Identity Element**: There exists a unique vector $\mathbf{0} \in V$ such that $\mathbf{v} + \mathbf{0} = \mathbf{v}$ for all $\mathbf{v} \in V$.
   - **Proof**: Let $\mathbf{0} = 0\mathbf{v}$. Then $\mathbf{v} + \mathbf{0} = \mathbf{v} + 0\mathbf{v} = \mathbf{v}$.
2. **Uniqueness of the Identity Element**: The identity element $\mathbf{0}$ in a vector space V is unique.
   - **Proof**: Suppose $\mathbf{0}$ and $\mathbf{0}'$ are both identity elements. Then $\mathbf{0} = \mathbf{0} + \mathbf{0}' = \mathbf{0}'$.
3. **Existence of Inverse Elements**: For each $\mathbf{v} \in V$, there exists a unique $-\mathbf{v} \in V$ such that $\mathbf{v} + (-\mathbf{v}) = \mathbf{0}$.
   - **Proof**: Let $-\mathbf{v} = (-1)\mathbf{v}$. Then $\mathbf{v} + (-\mathbf{v}) = \mathbf{v} + (-1)\mathbf{v} = \mathbf{0}$.

Vector Subspaces

A **vector subspace** W of a vector space V over $\mathbb{F}$ is a subset of V that satisfies the following properties:

1. **Non-Empty**: $W \neq \emptyset$.
2. **Closed under Addition**: If $\mathbf{v}, \mathbf{w} \in W$, then $\mathbf{v} + \mathbf{w} \in W$.
3. **Closed under Scalar Multiplication**: If $\mathbf{v} \in W$ and $c \in \mathbb{F}$, then $c\,\mathbf{v} \in W$.

Theorems and Proofs in Vector Subspaces

1. **Intersection of Subspaces is a Subspace**: If W_1 and W_2 are subspaces of V, then $W_1 \cap W_2$ is also a subspace of V.

   - **Proof**:

     - **Non-Empty**: $\mathbf{0} \in W_1$ and $\mathbf{0} \in W_2$ imply $\mathbf{0} \in W_1 \cap W_2$.
     - **Closure under Addition**: If $\mathbf{v}, \mathbf{w} \in W_1 \cap W_2$, then $\mathbf{v}, \mathbf{w} \in W_1$ and $\mathbf{v}, \mathbf{w} \in W_2$, so $\mathbf{v} + \mathbf{w} \in W_1$ and $\mathbf{v} + \mathbf{w} \in W_2$, hence $\mathbf{v} + \mathbf{w} \in W_1 \cap W_2$.
     - **Closure under Scalar Multiplication**: If $\mathbf{v} \in W_1 \cap W_2$, then $\mathbf{v} \in W_1$ and $\mathbf{v} \in W_2$, so $c\,\mathbf{v} \in W_1$ and $c\,\mathbf{v} \in W_2$, hence $c\,\mathbf{v} \in W_1 \cap W_2$.

2. **Sum of Subspaces is a Subspace**: If W_1 and W_2 are subspaces of V, then $W_1 + W_2 = \{\mathbf{v}_1 + \mathbf{v}_2 \mid \mathbf{v}_1 \in W_1, \mathbf{v}_2 \in W_2\}$ is also a subspace of V.

   - **Proof**: Similar to the intersection proof, showing closure under addition and scalar multiplication.

Basis of a Vector Space

In linear algebra, a **basis** of a vector space V over a field $\mathbb{F}$ is a set of vectors that satisfies two main properties:

1. **Spanning Property**: Every vector in the vector space V can be expressed as a unique linear combination of the basis vectors.

   Mathematically, if $\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n\}$ is a set of vectors in V, then this set forms a basis if every vector $\mathbf{v} \in V$ can be uniquely expressed as:

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

   where $c_1, c_2, ..., c_n$ are scalars from the field $\mathbb{F}$.

2. **Linear Independence**: The basis vectors $\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n\}$ are linearly independent, meaning no vector in the set can be expressed as a linear combination of the others.

   Formally, this means the only solution to the equation $c_1\,\mathbf{v}_1 + c_2\,\mathbf{v}_2 + ... + c_n\,\mathbf{v}_n = \mathbf{0}$ is $c_1 = c_2 = ... = c_n = 0$.

Example

Consider a vector space $\mathbb{R}^3$ over $\mathbb{R}$.

- **Standard Basis**: The set $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ where $\mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$, $\mathbf{e}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$, and $\mathbf{e}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$ forms a basis for $\mathbb{R}^3$.

- **Explanation**:
  - **Spanning Property**: Any vector $\mathbf{v} = \begin{pmatrix} x \\ y \\ z \end{pmatrix} \in \mathbb{R}^3$ can be written as $x\,\mathbf{e}_1 + y\,\mathbf{e}_2 + z\,\mathbf{e}_3$.
  - **Linear Independence**: The vectors $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ are clearly linearly independent because the only solution to $c_1\,\mathbf{e}_1 + c_2\,\mathbf{e}_2 + c_3\,\mathbf{e}_3 = \mathbf{0}$ is $c_1 = c_2 = c_3 = 0$.

Theorems Related to Basis

1. **Existence of Basis**: Every vector space has a basis. This is a fundamental theorem in linear algebra and can be proven using techniques such as the exchange lemma or Zorn's lemma (for infinite-dimensional spaces).

2. **Dimension**: The number of vectors in any basis of a vector space V is called the dimension of V, denoted as $\dim(V)$. All bases of V have the same number of vectors.

3. **Extension Theorem**: Any linearly independent set of vectors in V can be extended to form a basis of V.

Bases allow us to represent vectors in a clear and structured manner and facilitate the study of transformations, systems of equations, and various mathematical structures. Understanding the properties and theorems related to bases is essential for mastering the deeper concepts of linear algebra.

Generating Set

A **generating set** for a vector space V over a field $\mathbb{F}$ is a set of vectors that, through linear combinations, can produce (or generate) every vector in V.

Formal Definition

Let $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ be a set of vectors in V. S is a generating set for V if every vector $\mathbf{v} \in V$ can be expressed as a linear combination of the vectors in S, that is,

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

for some scalars $c_1, c_2, \ldots, c_n \in \mathbb{F}$.

Example

Consider $\mathbb{R}^2$, the vector space of all 2-dimensional real vectors.

- **Generating Set**: The set $\{\mathbf{v}_1 = \begin{pmatrix}1\\0\end{pmatrix}, \mathbf{v}_2 = \begin{pmatrix}0\\1\end{pmatrix}\}$ is a generating set for $\mathbb{R}^2$ because any vector $\begin{pmatrix}x\\y\end{pmatrix} \in \mathbb{R}^2$ can be expressed as $x\,\mathbf{v}_1 + y\,\mathbf{v}_2$.

Relationship between Generating Sets and Basis

- Every basis is a generating set, but not every generating set is a basis.
- A generating set that is not linearly independent is called a **linearly dependent generating set**.
- The basis provides a unique way to represent each vector in the vector space, similar to how coordinates work in Euclidean spaces.
- The number of vectors in any basis of a vector space V is called the **dimension** of V, denoted as $\dim(V)$.

Generating sets allow us to understand how vectors can span a space, while bases provide a structured and unique way to represent vectors within that space

Linear Combination

In linear algebra, a **linear combination** of vectors is formed by scaling (multiplying) each vector by a scalar and then adding the results together. Formally, given vectors $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n$ in a vector space V over a field $\mathbb{F}$, a linear combination of these vectors is expressed as:

$$c_1\,\mathbf{v}_1 + c_2\,\mathbf{v}_2 + ... + c_n\,\mathbf{v}_n$$

where $c_1, c_2, ..., c_n$ are scalars (elements of $\mathbb{F}$).

Span of Vectors

The **span** of a set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n\}$ in a vector space V is the set of all possible linear combinations of these vectors. In other words, it is the set of all vectors that can be expressed as:

$$c_1\,\mathbf{v}_1 + c_2\,\mathbf{v}_2 + ... + c_n\,\mathbf{v}_n$$

where $c_1, c_2, ..., c_n$ are scalars from the field $\mathbb{F}$.

Example

Consider the vectors $\mathbf{v}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $\mathbf{v}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ in $\mathbb{R}^2$.

- **Linear Combination**: Any vector $\begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}^2$ can be written as $x\,\mathbf{v}_1 + y\,\mathbf{v}_2 = x\begin{pmatrix} 1 \\ 0 \end{pmatrix} + y\begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix}$.
- **Span**: The span of $\{\mathbf{v}_1, \mathbf{v}_2\}$ is $\mathbb{R}^2$. This means every vector in $\mathbb{R}^2$ can be expressed as a linear combination of $\mathbf{v}_1$ and $\mathbf{v}_2$.

Linear combinations allow us to explore how vectors can be combined, while span defines the set of all possible combinations that can be achieved with a given set of vectors

Inner Products

An **inner product** on a vector space V over a field $\mathbb{F}$ (typically $\mathbb{R}$ or $\mathbb{C}$) is a way to define a notion of angle and length (magnitude) between vectors in V. Formally, an inner product is a function that takes two vectors $\mathbf{u}, \mathbf{v} \in V$ and returns a scalar in $\mathbb{F}$, denoted as $\langle \mathbf{u}, \mathbf{v} \rangle$.

Properties of Inner Products

For an inner product $\langle \cdot, \cdot \rangle$ on V, the following properties must hold for all $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$ and $c \in \mathbb{F}$:

1. **Linearity in the First Argument**:
   $\langle c\,\mathbf{u} + \mathbf{v}, \mathbf{w} \rangle = c\langle \mathbf{u}, \mathbf{w} \rangle + \langle \mathbf{v}, \mathbf{w} \rangle$

2. **Conjugate Symmetry (Hermitian Symmetry)**:
   $\langle \mathbf{u}, \mathbf{v} \rangle = \overline{\langle \mathbf{v}, \mathbf{u} \rangle}$
   This property holds for inner products over $\mathbb{C}$, where $\overline{z}$ denotes the complex conjugate of z.

3. **Positive Definiteness**:
   $\langle \mathbf{u}, \mathbf{u} \rangle \geq 0$
   Equality holds if and only if $\mathbf{u} = \mathbf{0}$, the zero vector.

4. **Non-Degeneracy** (optional, depending on the context):
   If $\langle \mathbf{u}, \mathbf{v} \rangle = 0$ for all $\mathbf{v} \in V$, then $\mathbf{u} = \mathbf{0}$.

Examples of Inner Products

1. **Dot Product (Euclidean Inner Product)**:
   In $\mathbb{R}^n$, the dot product of vectors $\mathbf{u} = (u_1, u_2, ..., u_n)$ and $\mathbf{v} = (v_1, v_2, ..., v_n)$ is defined as:
   $\langle \mathbf{u}, \mathbf{v} \rangle = u_1 v_1 + u_2 v_2 + ... + u_n v_n$
   This inner product satisfies all the properties listed above.

2. **Standard Inner Product on $\mathbb{C}^n$**:
   For vectors $\mathbf{u} = (u_1, u_2, ..., u_n)$ and $\mathbf{v} = (v_1, v_2, ..., v_n)$ in $\mathbb{C}^n$, the standard inner product is:
   $\langle \mathbf{u}, \mathbf{v} \rangle = \sum_{i=1}^{n} \overline{u_i} v_i$
   This inner product also satisfies the properties of an inner product.

Consider two vectors $\mathbf{u} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$ and $\mathbf{v} = \begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix}$.

Calculation of Inner Product

The inner product (also known as dot product) of two vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^3$ is calculated as follows:

$\langle \mathbf{u}, \mathbf{v} \rangle = u_1 \cdot v_1 + u_2 \cdot v_2 + u_3 \cdot v_3$

Substituting the values from our example:

$\langle \mathbf{u}, \mathbf{v} \rangle = 1 \cdot 4 + 2 \cdot 5 + 3 \cdot 6$
$\langle \mathbf{u}, \mathbf{v} \rangle = 4 + 10 + 18$
$\langle \mathbf{u}, \mathbf{v} \rangle = 32$

Interpretation

The inner product $\langle \mathbf{u}, \mathbf{v} \rangle = 32$ represents a measure of how much the vectors $\mathbf{u}$ and $\mathbf{v}$ align with each other in $\mathbb{R}^3$. It can be thought of as the projection of $\mathbf{u}$ onto $\mathbf{v}$ (or vice versa) scaled by their magnitudes.

Geometric Interpretation

In a geometric sense, the inner product $\langle \mathbf{u}, \mathbf{v} \rangle$ tells us about the cosine of the angle θ between $\mathbf{u}$ and $\mathbf{v}$:
$\langle \mathbf{u}, \mathbf{v} \rangle = \|\mathbf{u}\| \|\mathbf{v}\| \cos(\theta)$

Norms

A **norm** on a vector space V is a function that assigns a non-negative scalar value to each vector $\mathbf{v} \in V$, denoted as $\|\mathbf{v}\|$. The norm measures the "length" or "magnitude" of the vector and satisfies the following properties:

1. **Non-negativity**: $\|\mathbf{v}\| \geq 0$, and $\|\mathbf{v}\| = 0$ if and only if $\mathbf{v} = \mathbf{0}$.
2. **Homogeneity**: $\|c\,\mathbf{v}\| = |c|\|\mathbf{v}\|$ for all scalars $c \in \mathbb{F}$.
3. **Triangle Inequality**: $\|\mathbf{u} + \mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|$ for all $\mathbf{u}, \mathbf{v} \in V$.

Types of Norms

1. **Euclidean Norm (or ℓ^2 Norm)**:
   In $\mathbb{R}^n$, the Euclidean norm of a vector $\mathbf{v} = (v_1, v_2, ..., v_n)$ is defined as:
   $\|\mathbf{v}\|_2 = \sqrt{v_1^2 + v_2^2 + ... + v_n^2}$
2. **Manhattan Norm (or ℓ^1 Norm)**:
   The Manhattan norm of $\mathbf{v}$ in $\mathbb{R}^n$ is:
   $\|\mathbf{v}\|_1 = |v_1| + |v_2| + ... + |v_n|$
3. **Infinity Norm (or ℓ^∞ Norm)**:
   The infinity norm of $\mathbf{v}$ in $\mathbb{R}^n$ is:
   $\|\mathbf{v}\|_\infty = \max(|v_1|, |v_2|, ..., |v_n|)$

Relationship between Inner Products and Norms

- An inner product $\langle \mathbf{v}, \mathbf{v} \rangle$ gives rise to a norm $\|\mathbf{v}\| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}$.
- The Cauchy-Schwarz inequality relates the inner product to the norm:
  $|\langle \mathbf{u}, \mathbf{v} \rangle| \leq \|\mathbf{u}\|\|\mathbf{v}\|$

Cauchy-Schwarz Inequality

The Cauchy-Schwarz inequality is a fundamental inequality in linear algebra that relates the inner product of vectors to their norms. It provides an upper bound on the absolute value of the inner product of two vectors, which also establishes a connection between the inner product space and the concept of norms.

Statement of the Cauchy-Schwarz Inequality

For any vectors $\mathbf{u}$ and $\mathbf{v}$ in an inner product space (such as $\mathbb{R}^n$ or $\mathbb{C}^n$), the Cauchy-Schwarz inequality states:

$$|\langle \mathbf{u}, \mathbf{v} \rangle| \leq \|\mathbf{u}\| \|\mathbf{v}\|$$

where:

- $\langle \mathbf{u}, \mathbf{v} \rangle$ denotes the inner product of $\mathbf{u}$ and $\mathbf{v}$.
- $\|\mathbf{u}\|$ and $\|\mathbf{v}\|$ denote the norms of $\mathbf{u}$ and $\mathbf{v}$, respectively.

Proof Outline

The Cauchy-Schwarz inequality can be proven using various methods, including geometric interpretations and algebraic manipulations. Here's an outline of a common proof:

1. **Bilinearity of the Inner Product**: Utilize the bilinearity of the inner product:
   $\langle \alpha \mathbf{u} + \beta \mathbf{v}, \alpha \mathbf{u} + \beta \mathbf{v} \rangle \geq 0$

2. **Simplify the Expression**: Expand and simplify the expression using the properties of the inner product.

3. **Apply the Properties of Inner Product and Norms**: Use the properties such as symmetry, positivity, and linearity to derive the inequality.

4. **Complete the Proof**: By rearranging terms and applying the properties correctly, arrive at the inequality $|\langle \mathbf{u}, \mathbf{v} \rangle| \leq \|\mathbf{u}\| \|\mathbf{v}\|$.

Interpretation and Significance

- **Geometric Interpretation**: The Cauchy-Schwarz inequality can be understood geometrically as bounding the absolute value of the cosine of the angle between two vectors by their normalized lengths.

- **Applications**: This inequality is widely used in various areas of mathematics, including analysis, probability theory, optimization, and signal processing. It forms the basis for proving other important results and inequalities in these fields.

- **Equality Condition**: Equality holds in the Cauchy-Schwarz inequality if and only if $\mathbf{u}$ and $\mathbf{v}$ are linearly dependent, meaning one vector is a scalar multiple of the other (i.e., $\mathbf{u} = \lambda \mathbf{v}$ or $\mathbf{v} = \mu \mathbf{u}$ for some scalars $\lambda, \mu \in \mathbb{F}$).

Example

Consider vectors $\mathbf{u} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$ and $\mathbf{v} = \begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix}$ in $\mathbb{R}^3$.

- **Calculate Inner Product:** $\langle \mathbf{u}, \mathbf{v} \rangle = 1 \cdot 4 + 2 \cdot 5 + 3 \cdot 6 = 4 + 10 + 18 = 32$.
- **Calculate Norms:** $\|\mathbf{u}\| = \sqrt{1^2 + 2^2 + 3^2} = \sqrt{14}$, $\|\mathbf{v}\| = \sqrt{4^2 + 5^2 + 6^2} = \sqrt{77}$.
- **Apply Cauchy-Schwarz Inequality:** $|\langle \mathbf{u}, \mathbf{v} \rangle| = |32|$.
- $\|\mathbf{u}\|\|\mathbf{v}\| = \sqrt{14} \cdot \sqrt{77} \approx 29.83$.

Since $|32| \leq 29.83$, the Cauchy-Schwarz inequality holds in this example.

The Cauchy-Schwarz inequality is a powerful tool that encapsulates the relationship between inner products and norms, providing insights into the geometry and structure of vector spaces.

Metric Space

A **metric space** is a fundamental concept in mathematics that provides a framework for studying distances and proximity between elements of a set. It is a mathematical structure that formalizes the intuitive notion of distance between points.

Definition

Formally, a metric space (X, d) consists of a set X (often called the underlying set or space) and a **distance function** $d: X \times X \to \mathbb{R}$ that satisfies the following properties for all $x, y, z \in X$:

1. **Non-negativity**: $d(x, y) \geq 0$, with equality $d(x, y) = 0$ if and only if $x = y$ (Identity of indiscernibles).
2. **Symmetry**: $d(x, y) = d(y, x)$ (Symmetry).
3. **Triangle inequality**: $d(x, z) \leq d(x, y) + d(y, z)$ (Triangle inequality).

Properties

- **Distance Function**: The function $d(x, y)$ measures the distance between points x and y in X.
- **Metric**: The function d is called a metric, and it determines the topology of the space X.

Examples

1. **Euclidean Space**: $\mathbb{R}^n$ with the Euclidean distance $d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|$, where $\|\cdot\|$ denotes the Euclidean norm.

2. **Discrete Metric**: For any set X, define $d(x, y) = 1$ if $x \neq y$ and $d(x, y) = 0$ if $x = y$.

3. **Taxicab (Manhattan) Metric**: $\mathbb{R}^n$ with the taxicab distance $d(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{n} |x_i - y_i|$.

Completeness

A metric space (X, d) is **complete** if every Cauchy sequence in X converges to a limit that is also in X. Completeness is an important property in metric spaces, especially in the study of convergence and continuity.

Metric spaces provide a rigorous mathematical framework for studying distances and proximity between elements of a set

Angles Between Vectors

In a vector space $\mathbb{R}^n$, the angle θ between two vectors $\mathbf{u}$ and $\mathbf{v}$ is defined using the dot product (inner product):

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

where $\langle \mathbf{u}, \mathbf{v} \rangle$ is the dot product of $\mathbf{u}$ and $\mathbf{v}$, and $\|\mathbf{u}\|$ and $\|\mathbf{v}\|$ are the norms (lengths) of $\mathbf{u}$ and $\mathbf{v}$, respectively.

Example 1: Angles in $\mathbb{R}^2$

Consider two vectors $\mathbf{u} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$ and $\mathbf{v} = \begin{pmatrix} -2 \\ 1 \end{pmatrix}$.

- **Calculate Dot Product**: $\langle \mathbf{u}, \mathbf{v} \rangle = 1 \cdot (-2) + 2 \cdot 1 = -2 + 2 = 0$.
- **Calculate Norms**: $\|\mathbf{u}\| = \sqrt{1^2 + 2^2} = \sqrt{5}$, $\|\mathbf{v}\| = \sqrt{(-2)^2 + 1^2} = \sqrt{5}$.
- **Compute Angle**: $\cos(\theta) = \frac{0}{\sqrt{5} \cdot \sqrt{5}} = 0$.
  - $\theta = \cos^{-1}(0) = \frac{\pi}{2}$ radians (or $90°$).

Numerical Interpretation

Angle Calculation: The angle between vectors measures how much they deviate from being parallel (small angle) or perpendicular (right angle).

Orthogonality

Two vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$ are **orthogonal** if their dot product $\langle \mathbf{u}, \mathbf{v} \rangle = 0$.

Example 2: Orthogonal Vectors in $\mathbb{R}^3$

Consider two vectors $\mathbf{u} = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}$ and $\mathbf{v} = \begin{pmatrix} 2 \\ 1 \\ 2 \end{pmatrix}$.

- **Calculate Dot Product**: $\langle \mathbf{u}, \mathbf{v} \rangle = 1 \cdot 2 + 0 \cdot 1 + (-1) \cdot 2 = 2 - 2 = 0$.

Since $\langle \mathbf{u}, \mathbf{v} \rangle = 0$, $\mathbf{u}$ and $\mathbf{v}$ are orthogonal.

Numerical Interpretation

Orthogonality: Orthogonal vectors are fundamental in vector spaces because they provide a basis for constructing perpendicular components and are essential in applications like coordinate systems and projections.

Orthonormal Matrix

An **orthonormal matrix** is a square matrix whose rows (and columns) are orthogonal unit vectors. This means that the matrix A satisfies the following conditions:

1. **Orthogonality**: The rows of A are orthogonal to each other, i.e., $A^\top A = I$, where I is the identity matrix.

2. **Normalization**: Each row (and column, due to orthogonality) of A has unit length, i.e., $\|\mathbf{a}_i\| = 1$ for each row vector $\mathbf{a}_i$.

Orthonormal matrices play a crucial role in various areas such as:

- **Coordinate Systems**: Rotations and reflections in $\mathbb{R}^n$ are represented by orthonormal matrices.

- **Linear Transformations**: They preserve distances and angles, making them useful in transformations like rotations in computer graphics and physics.

- **Signal Processing**: Orthonormal matrices are used in applications like Fourier transforms and wavelet transforms.

In summary, orthonormal matrices are fundamental in mathematics and its applications, providing a structured way to represent rotations, reflections, and other transformations while preserving the geometric properties of vectors and spaces

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

Orthonormal Basis

An **orthonormal basis** is a set of vectors in a vector space that is both orthogonal and normalized. This means that the vectors in the basis are mutually perpendicular (orthogonal) and each vector has a unit length (normalized).

Definition

Let V be a vector space over a field $\mathbb{F}$. An **orthonormal basis** for V is a set $\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n\}$ of vectors in V such that:

1. The vectors $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n$ are pairwise orthogonal, i.e., $\langle \mathbf{v}_i, \mathbf{v}_j \rangle = 0$ for all $i \neq j$.
2. Each vector $\mathbf{v}_i$ has unit length, i.e., $\|\mathbf{v}_i\| = 1$.

Example of an Orthonormal Basis

Let's construct an example of an orthonormal basis in $\mathbb{R}^3$:

$\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$

where:

- $\mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$
- $\mathbf{e}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$
- $\mathbf{e}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$

Verification:

1. **Orthogonality**:

   - $\langle \mathbf{e}_1, \mathbf{e}_2 \rangle = 1 \cdot 0 + 0 \cdot 1 + 0 \cdot 0 = 0$
   - $\langle \mathbf{e}_1, \mathbf{e}_3 \rangle = 1 \cdot 0 + 0 \cdot 0 + 0 \cdot 1 = 0$
   - $\langle \mathbf{e}_2, \mathbf{e}_3 \rangle = 0 \cdot 0 + 1 \cdot 0 + 0 \cdot 1 = 0$

   Each pair of vectors $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ are orthogonal to each other.

2. **Normalization**:

   - $\|\mathbf{e}_1\| = \sqrt{1^2 + 0^2 + 0^2} = 1$
   - $\|\mathbf{e}_2\| = \sqrt{0^2 + 1^2 + 0^2} = 1$
   - $\|\mathbf{e}_3\| = \sqrt{0^2 + 0^2 + 1^2} = 1$

   Each vector $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ has unit length.

Therefore, $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ forms an orthonormal basis for $\mathbb{R}^3$.

Matrix Decompositions

Matrix decomposition, also known as matrix factorization, is the process of breaking down a matrix into simpler, more manageable parts. Each type of matrix decomposition serves specific purposes in various mathematical and computational applications. Here's an overview of common types of matrix decompositions, their properties, and relevant theorems:

Types of Matrix Decompositions

1. **LU Decomposition (LU Factorization)**:

   - **Definition**: LU decomposition decomposes a square matrix A into the product of a lower triangular matrix L and an upper triangular matrix U.
   - **Form**: $A = LU$, where L is lower triangular ($l_{ii} = 1$) and U is upper triangular.

2. **QR Decomposition (QR Factorization)**:

   - **Definition**: QR decomposition decomposes a matrix A into the product of an orthogonal matrix Q and an upper triangular matrix R.
   - **Form**: $A = QR$, where Q is orthogonal ($Q^\top Q = I$) and R is upper triangular.

3. **Eigenvalue Decomposition (Eigendecomposition)**:
    - **Definition**: Eigendecomposition decomposes a square matrix A into the product of its eigenvectors and eigenvalues.
    - **Form**: $A = V \Lambda V^{-1}$, where V is a matrix whose columns are eigenvectors of A, and Λ is a diagonal matrix of eigenvalues.
4. **Singular Value Decomposition (SVD)**:
    - **Definition**: SVD decomposes any $m \times n$ matrix A into the product of three matrices: $A = U \Sigma V^{\top}$, where U and V are orthogonal matrices (or unitary matrices in the complex case), and Σ is a diagonal matrix of singular values.
5. **Cholesky Decomposition**:
    - **Definition**: Cholesky decomposition is applicable to positive definite matrices A and decomposes A into the product of a lower triangular matrix L and its transpose $L^{\top}$.
    - **Form**: $A = LL^{\top}$, where all diagonal elements of L are positive.

Properties and Theorems

1. **Existence and Uniqueness**: Some decompositions like LU and QR exist and are unique under certain conditions (e.g., non-singularity of matrices).
2. **Computational Efficiency**: Certain decompositions (e.g., LU, QR) are computationally efficient and stable compared to others.
3. **Applications**: Each decomposition has specific applications such as solving linear systems, calculating eigenvalues/eigenvectors, least squares solutions, and data compression (SVD).

Example: QR Decomposition

Let's illustrate QR decomposition with a numerical example:

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

- **QR Decomposition**:

  Compute Q (orthogonal matrix) and R (upper triangular matrix):

  $$Q = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ 0 & \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{3}} \end{pmatrix}$$

  $$R = \begin{pmatrix} \sqrt{2} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ 0 & \frac{\sqrt{6}}{2} & -\frac{\sqrt{6}}{2} \\ 0 & 0 & \frac{2}{\sqrt{3}} \end{pmatrix}$$

  Verify $A = QR$.

Eigenvalues and Eigenvectors

Eigenvalues:

For a square matrix A, an **eigenvalue** λ is a scalar such that there exists a non-zero vector $\mathbf{v}$ satisfying the equation:

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

Here, $\mathbf{v}$ is the eigenvector corresponding to λ.

Eigenvectors:

The **eigenvector** $\mathbf{v}$ corresponding to λ is a non-zero vector that satisfies the eigenvalue equation $A\mathbf{v} = \lambda\mathbf{v}$.

Characteristic Equation

The **characteristic equation** of a square matrix A is derived from the equation $A\mathbf{v} = \lambda\mathbf{v}$. It is given by:

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

Here, I is the identity matrix of the same size as A.

Example:

Let's find the eigenvalues and eigenvectors of the matrix $A = \begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix}$.

Characteristic Equation:

Compute $A - \lambda I$:

$$A - \lambda I = \begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix} - \lambda \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 3-\lambda & 1 \\ 1 & 3-\lambda \end{pmatrix}$$

Set the determinant of $A - \lambda I$ to zero:

$$\det\left(\begin{pmatrix} 3-\lambda & 1 \\ 1 & 3-\lambda \end{pmatrix}\right) = (3-\lambda)^2 - 1 = 0$$

Simplify and solve for λ:

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

$(3-\lambda)^2 = 1$

$3 - \lambda = \pm 1$

$\lambda = 2 \quad \text{or} \quad \lambda = 4$

So, the eigenvalues of A are $\lambda_1 = 2$ and $\lambda_2 = 4$.

Eigenvectors:

- For $\lambda_1 = 2$:
  Solve $(A - 2I)\mathbf{v} = 0$:
  $(A - 2I)\mathbf{v} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}\mathbf{v} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$
  A solution is $\mathbf{v}_1 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$.

- For $\lambda_2 = 4$:
  Solve $(A - 4I)\mathbf{v} = 0$:
  $(A - 4I)\mathbf{v} = \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}\mathbf{v} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$
  A solution is $\mathbf{v}_2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$.

Spectral Theorem

The **spectral theorem** is a fundamental result in linear algebra that provides a powerful connection between the properties of a symmetric (or more generally, a self-adjoint) matrix and its eigenvalues and eigenvectors. It states that for a symmetric matrix A, there exists an orthonormal basis of eigenvectors that diagonalizes A. Here's a detailed explanation

Spectral Theorem for Symmetric Matrices

1. **Symmetric Matrices**

Let A be an $n \times n$ real symmetric matrix. This means $A = A^{\mathsf{T}}$, where A^{T} denotes the transpose of A.

2. **Existence of Eigenvalues and Eigenvectors**

Since A is symmetric, it has n real eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$. Moreover, there exists an orthonormal set of eigenvectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ corresponding to these eigenvalues. Orthonormality means:

$$\langle \mathbf{v}_i, \mathbf{v}_j \rangle = \delta_{ij}$$

where δ_{ij} is the Kronecker delta (1 if $i = j$, 0 otherwise).

3. **Diagonalization**

The spectral theorem states that the matrix A can be diagonalized by an orthogonal matrix Q:

$$A = Q \Lambda Q^{\mathsf{T}}$$

where:

- $Q = [\mathbf{v}_1 \; \mathbf{v}_2 \cdots \mathbf{v}_n]$ is an orthogonal matrix whose columns are the eigenvectors of A.
- $\Lambda = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$ is a diagonal matrix containing the eigenvalues of A.

Example

Consider the symmetric matrix $A = \begin{pmatrix} 4 & 1 \\ 1 & 3 \end{pmatrix}$.

Eigenvalues: Solve $\det(A - \lambda I) = 0$:

$$\det \begin{pmatrix} 4-\lambda & 1 \\ 1 & 3-\lambda \end{pmatrix} = (4-\lambda)(3-\lambda) - 1 = \lambda^2 - 7\lambda + 11 = 0$$

Solving gives $\lambda_1 = 2$ and $\lambda_2 = 5$.

Eigenvectors: For $\lambda_1 = 2$:

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

$$(A - 2I)\mathbf{v} = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}\mathbf{v} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

A solution is $\mathbf{v}_1 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$.

For $\lambda_2 = 5$:

Solve $(A - 5I)\mathbf{v} = 0$:

$$(A - 5I)\mathbf{v} = \begin{pmatrix} -1 & 1 \\ 1 & -2 \end{pmatrix}\mathbf{v} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

A solution is $\mathbf{v}_2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$.

3. **Diagonalization**:

   Construct $Q = [\mathbf{v}_1\ \mathbf{v}_2] = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}$ (orthogonal matrix).

   Construct $\Lambda = \mathbf{diag}(2, 5)$.

   Verify $A = Q\Lambda Q^{\mathsf{T}}$:

   $$Q\Lambda Q^{\mathsf{T}} = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}\begin{pmatrix} 2 & 0 \\ 0 & 5 \end{pmatrix}\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 4 & 1 \\ 1 & 3 \end{pmatrix} = A$$

Therefore, A is diagonalized by Q with eigenvalues Λ.

Hermitian matrix

A **Hermitian matrix**, also known as a **self-adjoint matrix**, is a complex square matrix that is equal to its own conjugate transpose. Hermitian matrices generalize the concept of real symmetric matrices to the complex domain. Here's a detailed explanation:

Hermitian matrices play a role in quadratic programming and convex optimization problems.

Definition

Let A be an $n \times n$ matrix with complex entries. A is Hermitian if it satisfies:

$$A = A^*$$

where A^* denotes the conjugate transpose of A. Explicitly, for each element a_{ij} in A, this condition implies:

$$a_{ij} = \overline{a_{ji}}$$

where $\overline{a_{ji}}$ is the complex conjugate of a_{ji}.

Properties

1. **Real Eigenvalues**: Every eigenvalue λ of a Hermitian matrix A is real.
2. **Orthogonal Eigenvectors**: Corresponding eigenvectors of distinct eigenvalues are orthogonal.
3. **Diagonalizability**: A Hermitian matrix is diagonalizable by a unitary matrix U, meaning $A = U \Lambda U^*$, where Λ is a diagonal matrix of eigenvalues and U is a unitary matrix whose columns are eigenvectors of A.
4. **Positive Semi-Definiteness**: All eigenvalues of a Hermitian matrix are non-negative, making it positive semi-definite.
5. **Unique Representation**: Unlike general complex matrices, Hermitian matrices have a unique spectral decomposition.

Examples

Let's consider a simple example of a 2×2 Hermitian matrix:

$$A = \begin{pmatrix} 3 & 2i \\ -2i & 1 \end{pmatrix}$$

To verify if A is Hermitian:

- Compute A^*, the conjugate transpose of A:

  $$A^* = \begin{pmatrix} 3 & -2i \\ 2i & 1 \end{pmatrix}$$

- Check if $A = A^*$:

  $A = \begin{pmatrix} 3 & 2i \\ -2i & 1 \end{pmatrix}$ matches $A^* = \begin{pmatrix} 3 & -2i \\ 2i & 1 \end{pmatrix}$.

Since A equals its conjugate transpose, it is indeed a Hermitian matrix.

Cholesky Decomposition

Cholesky decomposition is a specialized form of matrix decomposition that applies specifically to symmetric positive definite matrices. It decomposes such a matrix into the product of a lower triangular matrix and its conjugate transpose. Here's how it works, along with a numerical example: For a symmetric positive definite matrix AAA, Cholesky decomposition finds a lower triangular matrix LLL such that:

A=LL*

where L* denotes the conjugate transpose of L

Procedure

1. **Matrix Form**:

   If A is an $n \times n$ symmetric positive definite matrix, then A can be expressed as:

   $$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{12} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{nn} \end{pmatrix}$$

2. **Cholesky Decomposition**:

   The Cholesky decomposition finds L such that:

   $$L = \begin{pmatrix} l_{11} & 0 & \cdots & 0 \\ l_{21} & l_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ l_{n1} & l_{n2} & \cdots & l_{nn} \end{pmatrix}$$

   Where L is lower triangular and satisfies $A = LL^*$.

Example for Cholesky Decomposition

Let's perform Cholesky decomposition on the matrix:

$$A = \begin{pmatrix} 4 & 2 \\ 2 & 5 \end{pmatrix}$$

Compute L:

- Compute l_{11}:

  $l_{11} = \sqrt{a_{11}} = \sqrt{4} = 2$

- Compute l_{21}:

  $l_{21} = \frac{a_{21}}{l_{11}} = \frac{2}{2} = 1$

- Compute l_{22}:

  $l_{22} = \sqrt{a_{22} - l_{21}^2} = \sqrt{5 - 1^2} = \sqrt{4} = 2$

Construct L:

$$L = \begin{pmatrix} 2 & 0 \\ 1 & 2 \end{pmatrix}$$

Verification:

- Compute LL^*:

  $$L \cdot L^* = \begin{pmatrix} 2 & 0 \\ 1 & 2 \end{pmatrix} \cdot \begin{pmatrix} 2 & 1 \\ 0 & 2 \end{pmatrix} = \begin{pmatrix} 4 & 2 \\ 2 & 5 \end{pmatrix} = A$$

The Cholesky decomposition of $A = \begin{pmatrix} 4 & 2 \\ 2 & 5 \end{pmatrix}$ is $L = \begin{pmatrix} 2 & 0 \\ 1 & 2 \end{pmatrix}$. This decomposition is computationally efficient and useful in various applications such as solving linear systems, generating correlated random variables, and optimizing quadratic forms.

Gram-Schmidt Process

The Gram-Schmidt process is a method for orthonormalizing a set of vectors in an inner product space, typically in Euclidean space with the standard inner product. It transforms a set of linearly independent vectors into an orthogonal (or orthonormal) set.

Given a set of linearly independent vectors $\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n\}$, the Gram-Schmidt process constructs an orthonormal set $\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_n\}$ where each vector $\mathbf{u}_i$ is orthogonal to all previous vectors and has unit length.

Steps:

1. **Initialization**:
   Let $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n$ be linearly independent vectors.

2. **Orthogonalization**:
   For each $i = 1, 2, ..., n$:

   - $\mathbf{u}_i = \mathbf{v}_i$ (initialize $\mathbf{u}_i$ with $\mathbf{v}_i$).
   - Subtract the projections of $\mathbf{v}_i$ onto $\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_{i-1}$:
     $\mathbf{u}_i = \mathbf{v}_i - \sum_{j=1}^{i-1} \frac{\langle \mathbf{v}_i, \mathbf{u}_j \rangle}{\langle \mathbf{u}_j, \mathbf{u}_j \rangle} \mathbf{u}_j$
   - Normalize $\mathbf{u}_i$:
     $\mathbf{u}_i = \frac{\mathbf{u}_i}{\|\mathbf{u}_i\|}$

3. **Output**:
   After applying the Gram-Schmidt process to $\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n\}$, the resulting set $\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_n\}$ is an orthonormal set of vectors.

Consider the vectors $\mathbf{v}_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$, $\mathbf{v}_2 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$, and $\mathbf{v}_3 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$.

Step-by-Step Calculation:

1. **Initialize**:
   Let $\mathbf{v}_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$, $\mathbf{v}_2 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$, and $\mathbf{v}_3 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$.

2. **Orthonormalize**:

   - **Step 1**: Compute $\mathbf{u}_1$:
     $$\mathbf{u}_1 = \mathbf{v}_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$$

   - **Step 2**: Compute $\mathbf{u}_2$:
     $$\mathbf{u}_2 = \mathbf{v}_2 - \frac{\langle \mathbf{v}_2, \mathbf{u}_1 \rangle}{\langle \mathbf{u}_1, \mathbf{u}_1 \rangle} \mathbf{u}_1$$
     $$\langle \mathbf{v}_2, \mathbf{u}_1 \rangle = 1 \cdot 1 + 0 \cdot 1 + 1 \cdot 0 = 1$$
     $$\langle \mathbf{u}_1, \mathbf{u}_1 \rangle = 1^2 + 1^2 + 0^2 = 2$$
     $$\mathbf{u}_2 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} - \frac{1}{2} \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} \frac{1}{2} \\ -\frac{1}{2} \\ 1 \end{pmatrix}$$

   - **Step 3**: Compute $\mathbf{u}_3$:
     $$\mathbf{u}_3 = \mathbf{v}_3 - \frac{\langle \mathbf{v}_3, \mathbf{u}_1 \rangle}{\langle \mathbf{u}_1, \mathbf{u}_1 \rangle} \mathbf{u}_1 - \frac{\langle \mathbf{v}_3, \mathbf{u}_2 \rangle}{\langle \mathbf{u}_2, \mathbf{u}_2 \rangle} \mathbf{u}_2$$
     $$\langle \mathbf{v}_3, \mathbf{u}_1 \rangle = 0 \cdot 1 + 1 \cdot 1 + 1 \cdot 0 = 1$$
     $$\langle \mathbf{v}_3, \mathbf{u}_2 \rangle = 0 \cdot \frac{1}{2} + 1 \cdot \left(-\frac{1}{2}\right) + 1 \cdot 1 = \frac{1}{2}$$
     $$\langle \mathbf{u}_2, \mathbf{u}_2 \rangle = \left(\frac{1}{2}\right)^2 + \left(-\frac{1}{2}\right)^2 + 1^2 = \frac{3}{2}$$
     $$\mathbf{u}_3 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} - \frac{1}{2} \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} - \frac{1}{3} \begin{pmatrix} \frac{1}{2} \\ -\frac{1}{2} \\ 1 \end{pmatrix} = \begin{pmatrix} -\frac{1}{6} \\ \frac{5}{6} \\ \frac{2}{3} \end{pmatrix}$$

3. **Normalization**:

   Normalize $\mathbf{u}_1, \mathbf{u}_2,$ and $\mathbf{u}_3$:
   $$\mathbf{u}_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \quad \mathbf{u}_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \quad \mathbf{u}_3 = \begin{pmatrix} -\frac{1}{\sqrt{6}} \\ \frac{5}{\sqrt{6}} \\ \frac{2}{\sqrt{6}} \end{pmatrix}$$

After applying the Gram-Schmidt process to $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$, we obtain an orthonormal set $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$. This process is essential in linear algebra and various applications where orthogonal or orthonormal bases are needed, such as in solving linear systems, computing projections, and in numerical algorithms.

Eigen Decomposition

Eigen decomposition is a process where a square matrix AAA is decomposed into a set of eigenvectors and eigenvalues. This is particularly useful in various applications, including principal component analysis (PCA) and solving differential equations

Eigen Decomposition

Given a square matrix A, the eigen decomposition expresses A as:

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

where:

- Q is a matrix whose columns are eigenvectors of A.
- Λ is a diagonal matrix containing the eigenvalues of A.

Steps for Eigen Decomposition

1. **Find Eigenvalues**: Solve the characteristic equation $\det(A - \lambda I) = 0$ to find the eigenvalues $\lambda_1, \lambda_2, ..., \lambda_n$.
2. **Find Eigenvectors**: For each eigenvalue λ_i, find the corresponding eigenvector $\mathbf{v}_i$ such that $A\,\mathbf{v}_i = \lambda_i\,\mathbf{v}_i$.
3. **Construct** Q: Form a matrix Q whose columns are the normalized eigenvectors $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n$.
4. **Construct** Λ: Form the diagonal matrix Λ with the eigenvalues $\lambda_1, \lambda_2, ..., \lambda_n$ along the diagonal.

Numerical Example

Let's find the eigen decomposition of the matrix:
$A = \begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix}$

Step-by-Step Calculation:

1. **Eigenvalues**: Solve $\det(A - \lambda I) = 0$.

   $$\det(A - \lambda I) = \det\left(\begin{pmatrix} 3-\lambda & 1 \\ 1 & 3-\lambda \end{pmatrix}\right) = (3-\lambda)^2 - 1 = \lambda^2 - 6\lambda + 8$$

   Solve $\lambda^2 - 6\lambda + 8 = 0$:
   $\lambda_1 = 2, \quad \lambda_2 = 4$

Find the Eigen Vectors, Eigen Values λ =2 and λ = 4

For $\lambda_1 = 2$:
Solve $(A - 2I)\mathbf{v}_1 = 0$:
$(A - 2I) = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$
$\mathbf{v}_1 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$

For $\lambda_2 = 4$:
Solve $(A - 4I)\mathbf{v}_2 = 0$:
$(A - 4I) = \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}$
$\mathbf{v}_2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$

3. **Construct** Q: Form the matrix Q with normalized eigenvectors.

$$Q = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix}$$

4. **Construct** Λ: Form the diagonal matrix Λ with eigenvalues.

$$\Lambda = \begin{pmatrix} 2 & 0 \\ 0 & 4 \end{pmatrix}$$

Verification

Verify $A = Q\Lambda Q^{-1}$:

$$Q\Lambda Q^{-1} = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix}\begin{pmatrix} 2 & 0 \\ 0 & 4 \end{pmatrix}\begin{pmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix} = \begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix} = A$$

Diagonalization of Matrices

Diagonalization of a matrix is a process of finding a similarity transformation that transforms the matrix into a diagonal matrix. Diagonalizable matrices are those for which such a transformation exists. This concept is crucial in various mathematical and applied contexts, including eigenanalysis, differential equations, and certain optimization problems.

Given a square matrix A, if there exists an invertible matrix P and a diagonal matrix D such that
$A = PDP^{-1}$
where D is diagonal, then A is said to be diagonalizable.

Steps for Diagonalization

1. **Find Eigenvalues and Eigenvectors**: Compute the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ of A by solving $\det(A - \lambda I) = 0$. For each eigenvalue λ_i, find the corresponding eigenvector $\mathbf{v}_i$ such that $A\,\mathbf{v}_i = \lambda_i\,\mathbf{v}_i$.

2. **Form Matrix** P: Construct the matrix P whose columns are the eigenvectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$.

3. **Form Diagonal Matrix** D: Construct the diagonal matrix D with eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ along the diagonal.

4. **Verify Diagonalization**: Check that $A = PDP^{-1}$.

Example

Consider the matrix:
$$A = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$$

Step-by-Step Calculation:

1. **Eigenvalues**: Solve $\det(A - \lambda I) = 0$.

   $$\det(A - \lambda I) = \det\left(\begin{pmatrix} 2-\lambda & 1 \\ 1 & 2-\lambda \end{pmatrix}\right) = (2-\lambda)^2 - 1 = \lambda^2 - 4\lambda + 3$$

   Solve $\lambda^2 - 4\lambda + 3 = 0$:

   $\lambda_1 = 1, \quad \lambda_2 = 3$

2. **Eigenvectors**: Find eigenvectors corresponding to $\lambda_1 = 1$ and $\lambda_2 = 3$.

- For $\lambda_1 = 1$:

  Solve $(A - I)\mathbf{v}_1 = 0$:

  $(A - I) = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$

  $\mathbf{v}_1 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$

- For $\lambda_2 = 3$:

  Solve $(A - 3I)\mathbf{v}_2 = 0$:

  $(A - 3I) = \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}$

  $\mathbf{v}_2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$

3. **Form Matrix** P: Construct P with normalized eigenvectors.

   $$P = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix}$$

4. **Form Diagonal Matrix** D: Construct D with eigenvalues.

   $$D = \begin{pmatrix} 1 & 0 \\ 0 & 3 \end{pmatrix}$$

5. **Verify Diagonalization**: Compute PDP^{-1} and check if it equals A.

   $$PDP^{-1} = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 3 \end{pmatrix}\begin{pmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} = A$$

The matrix $A = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$ is diagonalizable, as demonstrated by finding the eigenvalues, eigenvectors, constructing P, and verifying $A = PDP^{-1}$. Diagonalization allows us to simplify matrix operations and gain insights into the behavior of the matrix through its eigenvalues and eigenvectors, making it a fundamental concept in linear algebra and applications like machine learning.

Singular Value Decomposition (SVD)

Singular Value Decomposition (SVD) is a factorization of a real or complex matrix that generalizes the eigen decomposition of a square matrix to any m×n (Rectangular Matrix) matrix. SVD is widely used in signal processing, statistics, and machine learning for tasks such as dimensionality reduction, noise reduction, and data compression.

Given a matrix A of dimensions $m \times n$, SVD expresses A as:
$A = U \Sigma V^T$
where:

- U is an $m \times m$ orthogonal matrix.
- Σ is an $m \times n$ diagonal matrix with non-negative real numbers on the diagonal.
- V is an $n \times n$ orthogonal matrix.

The diagonal entries of Σ are known as the singular values of A, the columns of U are the left singular vectors, and the columns of V are the right singular vectors.

Properties of Each Component of SVD

1. **Matrix U (Left Singular Vectors)**

- **Orthogonality**: The columns of U are orthonormal vectors, meaning $U^T U = U U^T = I_m$ (the $m \times m$ identity matrix).
- **Left Singular Vectors**: The columns of U are called the left singular vectors of A. They span the column space of A.
- **Eigenvectors**: The left singular vectors are the eigenvectors of AA^T.

2. **Matrix Σ (Singular Values)**

- **Diagonal Matrix**: Σ is a diagonal matrix of the same dimensions as A (or $\min(m,n) \times \min(m,n)$ in its non-truncated form).
- **Non-negative Entries**: The diagonal entries of Σ, known as singular values, are non-negative and typically arranged in descending order: $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r \geq 0$, where $r = \min(m,n)$.
- **Rank**: The number of non-zero singular values equals the rank of the matrix A.
- **Spectral Norm**: The largest singular value σ_1 is the spectral norm of A.
- **Condition Number**: The condition number of A is given by $\frac{\sigma_1}{\sigma_r}$.

3. **Matrix V (Right Singular Vectors)**

- **Orthogonality**: The columns of V are orthonormal vectors, meaning $V^T V = V V^T = I_n$ (the $n \times n$ identity matrix).
- **Right Singular Vectors**: The columns of V are called the right singular vectors of A. They span the row space of A.
- **Eigenvectors**: The right singular vectors are the eigenvectors of $A^T A$.

Steps for SVD

1. **Compute** A^TA and AA^T.
2. **Find the eigenvalues and eigenvectors** of A^TA and AA^T.
3. **Construct** V from the eigenvectors of A^TA.
4. **Construct** U from the eigenvectors of AA^T.
5. **Construct** Σ from the square roots of the eigenvalues.

Example

Consider the matrix:

$$A = \begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix}$$

1. **Compute** A^TA **and** AA^T:

$$A^TA = \begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix}\begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix} = \begin{pmatrix} 10 & 6 \\ 6 & 10 \end{pmatrix}$$

$$AA^T = \begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix}\begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix} = \begin{pmatrix} 10 & 6 \\ 6 & 10 \end{pmatrix}$$

2. **Find the eigenvalues and eigenvectors of** A^TA:

$$\det(A^TA - \lambda I) = \begin{vmatrix} 10-\lambda & 6 \\ 6 & 10-\lambda \end{vmatrix} = (10-\lambda)^2 - 36 = 0$$

$$\lambda^2 - 20\lambda + 64 = 0$$

$$\lambda = 16, \quad \lambda = 4$$

Eigenvectors corresponding to $\lambda_1 = 16$ and $\lambda_2 = 4$:

- For $\lambda_1 = 16$:

$$(A^TA - 16I)\mathbf{v} = 0$$

$$\begin{pmatrix} -6 & 6 \\ 6 & -6 \end{pmatrix}\mathbf{v} = 0$$

$$\mathbf{v}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

- For $\lambda_2 = 4$:

$$(A^TA - 4I)\mathbf{v} = 0$$

$$\begin{pmatrix} 6 & 6 \\ 6 & 6 \end{pmatrix}\mathbf{v} = 0$$

$$\mathbf{v}_2 = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

3. **Construct** V from the normalized eigenvectors of A^TA:

$$V = \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}$$

4. **Construct** U from the normalized eigenvectors of AA^T:

$U = V$ (since AA^T and A^TA have the same eigenvectors)

5. **Construct** Σ from the square roots of the eigenvalues:

$$\Sigma = \begin{pmatrix} 4 & 0 \\ 0 & 2 \end{pmatrix}$$

Conclusion

The SVD of the matrix A is:

$$A = U\Sigma V^T = \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} 4 & 0 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}$$

This example demonstrates how SVD decomposes a matrix into its constituent parts, making it easier to understand and manipulate. SVD is a powerful tool in machine learning, especially for tasks such as data compression and noise reduction.

Matrix Approximation

Matrix approximation involves finding a matrix that is close to the original matrix but has a lower rank. One of the most common techniques for matrix approximation is using the truncated Singular Value Decomposition (SVD). This method is particularly useful in reducing the complexity of data while retaining most of its important features.

Matrix Approximation using Truncated SVD

Given a matrix A, the SVD of A is given by:

$A = U\Sigma V^T$

where:

- U is an $m \times m$ orthogonal matrix.
- Σ is an $m \times n$ diagonal matrix with non-negative real numbers on the diagonal.
- V is an $n \times n$ orthogonal matrix.

By truncating the SVD, we can create a lower-rank approximation of A.

Steps for Matrix Approximation using Truncated SVD

1. **Compute the SVD** of the matrix A.
2. **Select the top k singular values** and their corresponding singular vectors.
3. **Construct the truncated matrices** U_k, Σ_k, and V_k.
4. **Approximate the original matrix** A by multiplying the truncated matrices.

Example

Consider the matrix A:

$$A = \begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix}$$

Step-by-Step Calculation:

1. **Compute the SVD of A:**
   $A = U \Sigma V^T$

   The SVD decomposition gives us:
   $$U = \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}, \quad \Sigma = \begin{pmatrix} 4 & 0 \\ 0 & 2 \end{pmatrix}, \quad V = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}$$

2. **Select the top k singular values:**
   Let's select $k = 1$ for this example. This means we will keep only the largest singular value and its corresponding singular vectors.

3. **Construct the truncated matrices U_1, Σ_1, and V_1:**
   $$U_1 = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}, \quad \Sigma_1 = (4), \quad V_1 = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}$$

4. **Approximate the original matrix A** by multiplying the truncated matrices:
   $A_k = U_1 \Sigma_1 V_1^T$
   $$A_1 = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} (4) \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}$$
   $$A_1 = \begin{pmatrix} \frac{4}{\sqrt{2}} & \frac{4}{\sqrt{2}} \\ \frac{4}{\sqrt{2}} & \frac{4}{\sqrt{2}} \end{pmatrix}$$
   $$A_1 = \begin{pmatrix} 2 & 2 \\ 2 & 2 \end{pmatrix}$$

The original matrix A is approximated by A_1, which is a rank-1 matrix. This approximation retains the most significant feature (the largest singular value) while reducing the complexity of the matrix. Matrix approximation using truncated SVD is particularly useful in applications such as data compression, where it is essential to reduce the size of the data while retaining its essential features

Rank k Approximation

Rank k approximation is a technique used to approximate a matrix by another matrix of lower rank k, which is simpler and retains the most important features of the original matrix. This method is often used in data compression, noise reduction, and principal component analysis (PCA).

Steps for Rank k Approximation using SVD

Given a matrix A of dimensions $m \times n$, the Singular Value Decomposition (SVD) of A is:
$A = U \Sigma V^T$
where:

- U is an $m \times m$ orthogonal matrix.
- Σ is an $m \times n$ diagonal matrix with non-negative real numbers on the diagonal.
- V is an $n \times n$ orthogonal matrix.

To obtain a rank k approximation, we truncate the SVD by keeping only the top k singular values and their corresponding singular vectors.

Steps:

1. **Compute the SVD** of A.
2. **Select the top k singular values** and their corresponding singular vectors.
3. **Construct the truncated matrices** U_k, Σ_k, and V_k.
4. **Approximate the original matrix** A by multiplying the truncated matrices.

Properties of Rank k Approximation

- **Best Approximation**: The rank k approximation obtained by truncating the SVD provides the best approximation to A in terms of the Frobenius norm.
- **Error Bound**: The approximation error is minimized, and the sum of the squares of the discarded singular values gives the approximation error.
- **Dimensionality Reduction**: It reduces the dimensions of the data while preserving the most significant features.

Numerical Example

Consider the matrix A:

$$A = \begin{pmatrix} 3 & 1 & 1 \\ -1 & 3 & 1 \end{pmatrix}$$

Step-by-Step Calculation:

1. **Compute the SVD of A**:
   Let's assume the SVD of A is:
   $A = U \Sigma V^T$
   where:
   $$U = \begin{pmatrix} 0.707 & -0.707 \\ 0.707 & 0.707 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} 4 & 0 & 0 \\ 0 & 2 & 0 \end{pmatrix}, \quad V^T = \begin{pmatrix} 0.5 & -0.5 & 0.707 \\ -0.5 & -0.5 & 0.707 \\ 0.707 & 0.707 & 0 \end{pmatrix}$$

2. **Select the top $k = 1$ singular value**:
   For $k = 1$, we keep the largest singular value $\sigma_1 = 4$ and the corresponding singular vectors.

3. **Construct the truncated matrices U_1, Σ_1, and V_1:**

$$U_1 = \begin{pmatrix} 0.707 \\ 0.707 \end{pmatrix}, \quad \Sigma_1 = (4), \quad V_1^T = \begin{pmatrix} 0.5 & -0.5 & 0.707 \end{pmatrix}$$

4. **Approximate the original matrix A** by multiplying the truncated matrices:

$$A_1 = U_1 \Sigma_1 V_1^T$$
$$A_1 = \begin{pmatrix} 0.707 \\ 0.707 \end{pmatrix} (4) \begin{pmatrix} 0.5 & -0.5 & 0.707 \end{pmatrix}$$
$$A_1 = \begin{pmatrix} 0.707 \times 4 \times 0.5 & 0.707 \times 4 \times -0.5 & 0.707 \times 4 \times 0.707 \\ 0.707 \times 4 \times 0.5 & 0.707 \times 4 \times -0.5 & 0.707 \times 4 \times 0.707 \end{pmatrix}$$
$$A_1 = \begin{pmatrix} 1.414 & -1.414 & 2 \\ 1.414 & -1.414 & 2 \end{pmatrix}$$

Conclusion

The matrix A has been approximated by A_1, a rank-1 matrix. This approximation retains the most significant feature (the largest singular value) while reducing the complexity of the matrix. Rank k approximation using truncated SVD is a powerful technique for dimensionality reduction

Spectral Norm of a matrix

The spectral norm of a matrix is a measure of the matrix's "size" that is based on the largest singular value of the matrix. It is widely used in numerical analysis and linear algebra due to its properties and computational feasibility.

Definition

The spectral norm $\| A \|_2$ of a matrix A is defined as the largest singular value of A. If A has a Singular Value Decomposition (SVD) given by:

$A = U \Sigma V^T$

where Σ is a diagonal matrix with singular values $\sigma_1, \sigma_2, ..., \sigma_r$ (where r is the rank of A) on the diagonal in descending order, then the spectral norm is:

$\| A \|_2 = \sigma_1$

Properties

1. **Non-negativity**: The spectral norm is always non-negative, i.e., $\| A \|_2 \geq 0$.
2. **Sub-multiplicativity**: For any matrices A and B, $\| AB \|_2 \leq \| A \|_2 \| B \|_2$.
3. **Consistency with Euclidean Norm**: For any vector x, $\| Ax \|_2 \leq \| A \|_2 \| x \|_2$.
4. **Unitary Invariance**: For any unitary matrices U and V, $\| UAV \|_2 = \| A \|_2$.

Calculation

The spectral norm can be calculated using the largest singular value of the matrix, which can be found through SVD. For simpler cases like 2x2 matrices, the spectral norm can sometimes be computed directly from the matrix entries.

Example

Consider the matrix A:

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

Step-by-Step Calculation:

1. **Compute the SVD of A:**

   Let's assume the SVD of A is:

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

   where:

   $$U = \begin{pmatrix} -0.4046 & 0.9145 \\ -0.9145 & -0.4046 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} 5.4649 & 0 \\ 0 & 0.3659 \end{pmatrix}, \quad V = \begin{pmatrix} -0.5760 & -0.8174 \\ -0.8174 & 0.5760 \end{pmatrix}$$

2. **Extract the singular values:**

   The singular values are $\sigma_1 = 5.4649$ and $\sigma_2 = 0.3659$.

3. **Determine the spectral norm:**

   The spectral norm is the largest singular value:

   $$\| A \|_2 = 5.4649$$

Conclusion

The spectral norm provides a measure of the matrix's "size" that captures the maximum stretching effect the matrix has on any vector. It is especially useful in applications requiring stability analysis and error estimation, such as numerical solutions to differential equations and machine learning.

Vector Calculus

Vector calculus is a branch of mathematics that deals with vector fields and the differentiation and integration of vector functions. It extends the concepts of calculus to higher dimensions and is fundamental in physics and engineering.

Symbol	Reading	Meaning
∇	nabla	Gradient operator
f	f	Scalar function
$\mathbf{F}$	F (bold)	Vector field
x, y, z	x, y, z	Cartesian coordinates or scalar variables
∂	partial	Partial derivative
∇f	gradient of f	Vector of partial derivatives of f
$\frac{\partial f}{\partial x}$	partial derivative of f with respect to x	Rate of change of f in the x-direction
$\nabla \cdot \mathbf{F}$	divergence of F	Sum of the partial derivatives of the components of $\mathbf{F}$
$\nabla \times \mathbf{F}$	curl of F	Vector operation describing the rotation of $\mathbf{F}$
C	C	Curve or path
$\int$	integral	Integral sign
$\cdot$	dot	Dot product
$d\mathbf{r}$	d r (bold)	Differential vector
V	V	Volume
S	S	Surface
dV	d V	Differential volume
$d\mathbf{A}$	d A (bold)	Differential surface area vector
∂S	boundary of S	Boundary curve of surface S
P	P	Scalar function, often part of a vector field
Q	Q	Scalar function, often part of a vector field
dx	d x	Differential in the x-direction
dy	d y	Differential in the y-direction
$\mathbf{r}(a)$	r of a (bold)	Position vector at point a
$\mathbf{r}(b)$	r of b (bold)	Position vector at point b
$\nabla \phi$	gradient of phi	Gradient of scalar potential ϕ
$\nabla \times \mathbf{A}$	curl of A (bold)	Curl of vector potential $\mathbf{A}$
ϕ	phi	Scalar potential
$\mathbf{A}$	A (bold)	Vector potential
Δf	Laplacian of f	Divergence of the gradient of f, also written as $\nabla^2 f$

Basic Concepts

1. **Vectors and Vector Fields**

- A vector is an entity with both magnitude and direction.
- A vector field assigns a vector to every point in a subset of space.

2. **Gradient, Divergence, and Curl**

- **Gradient** (∇f): Measures the rate and direction of change in a scalar field.

$$\nabla f = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z} \right)$$

- **Divergence** ($\nabla \cdot \mathbf{F}$): Measures the magnitude of a source or sink at a given point in a vector field.

$$\nabla \cdot \mathbf{F} = \frac{\partial F_x}{\partial x} + \frac{\partial F_y}{\partial y} + \frac{\partial F_z}{\partial z}$$

- **Curl** ($\nabla \times \mathbf{F}$): Measures the rotation or the twisting force of a vector field.

$$\nabla \times \mathbf{F} = \left(\frac{\partial F_z}{\partial y} - \frac{\partial F_y}{\partial z}, \frac{\partial F_x}{\partial z} - \frac{\partial F_z}{\partial x}, \frac{\partial F_y}{\partial x} - \frac{\partial F_x}{\partial y} \right)$$

Theorems in Vector Calculus

1. **Gradient Theorem**

If $\mathbf{F} = \nabla f$, then:

$$\int_C \mathbf{F} \cdot d\,\mathbf{r} = f(\mathbf{r}(b)) - f(\mathbf{r}(a))$$

where C is a path from a to b.

2. **Divergence Theorem**

Relates the flux of a vector field through a surface to the divergence of the field in the volume enclosed by the surface.

$$\int_V (\nabla \cdot \mathbf{F})\, dV = \int_S \mathbf{F} \cdot d\,\mathbf{A}$$

where V is the volume and S is the boundary surface.

3. **Stokes' Theorem**

Relates a surface integral of the curl of a vector field to a line integral around the boundary of the surface.

$$\int_S (\nabla \times \mathbf{F}) \cdot d\,\mathbf{A} = \int_{\partial S} \mathbf{F} \cdot d\,\mathbf{r}$$

where S is a surface and ∂S is its boundary curve.

Properties and Relationships

1. **Linearity**

- The gradient, divergence, and curl are linear operations.

2. **Product Rules**

- Gradient: $\nabla(fg) = f\nabla g + g\nabla f$
- Divergence: $\nabla \cdot (f\,\mathbf{F}) = f(\nabla \cdot \mathbf{F}) + \mathbf{F} \cdot \nabla f$
- Curl: $\nabla \times (f\,\mathbf{F}) = f(\nabla \times \mathbf{F}) - \mathbf{F} \times (\nabla f)$

3. **Identities**

- $\nabla \cdot (\nabla \times \mathbf{F}) = 0$
- $\nabla \times (\nabla f) = 0$
- $\nabla \cdot (\nabla f) = \Delta f$ (where Δ is the Laplacian operator)

Differentiation of Univariate Functions

Differentiation of univariate functions involves finding the derivative of a function with respect to a single variable.

Example

Let's consider the function $f(x) = x^2 + 3x$.

Finding the Derivative $f'(x)$

To differentiate $f(x)$ with respect to x:

1. **Apply the Power Rule:**

   For x^2, the derivative is:
   $\frac{d}{dx}(x^2) = 2x$

2. **For $3x$, the derivative is:**
   $\frac{d}{dx}(3x) = 3$

3. **Combine the Results:**

   So, the derivative $f'(x)$ is:
   $f'(x) = \frac{d}{dx}(x^2 + 3x) = 2x + 3$

Numerical Example

Let's find $f'(2)$ using the derivative we found:

- $f(x) = x^2 + 3x$
- $f'(x) = 2x + 3$

Now, evaluate at $x = 2$:

$f'(2) = 2(2) + 3 = 4 + 3 = 7$

So, $f'(2) = 7$.

Interpretation

- **Derivative** $f'(x)$: Represents the rate of change of $f(x)$ with respect to x at any given point.
- **Application**: Useful in determining slopes of curves, finding maximum and minimum values, and understanding the behavior of functions.

Rules for Derivatives of Polynomial Functions

1. **Power Rule:**
   $\frac{d}{dx}[ax^n] = n \cdot ax^{n-1}$

   - a is a constant coefficient.
   - n is the exponent of x.

   Example:
   $\frac{d}{dx}[3x^4] = 4 \cdot 3x^{4-1} = 12x^3$

2. **Constant Rule:**
   $\frac{d}{dx}[c] = 0$

   - c is a constant (e.g., $c = 5$).

   Example:
   $\frac{d}{dx}[7] = 0$

3. **Sum and Difference Rule:**
   $\frac{d}{dx}[f(x) \pm g(x)] = \frac{d}{dx}[f(x)] \pm \frac{d}{dx}[g(x)]$

   - Applies to sums and differences of functions.

   Example:
   If $f(x) = 2x^2 + 3x$ and $g(x) = 4x - 1$,
   $\frac{d}{dx}[f(x) + g(x)] = \frac{d}{dx}[2x^2 + 3x + 4x - 1] = 4x + 7$

4. **Product Rule:**
   $\frac{d}{dx}[f(x) \cdot g(x)] = f(x) \cdot \frac{d}{dx}[g(x)] + g(x) \cdot \frac{d}{dx}[f(x)]$

   - Applies to the product of two functions.

   Example:
   If $f(x) = x^2$ and $g(x) = 3x$,
   $\frac{d}{dx}[x^2 \cdot 3x] = x^2 \cdot \frac{d}{dx}[3x] + 3x \cdot \frac{d}{dx}[x^2] = 3x^2 + 6x$

5. **Quotient Rule:**
   $\frac{d}{dx}\left[\frac{f(x)}{g(x)}\right] = \frac{g(x) \cdot \frac{d}{dx}[f(x)] - f(x) \cdot \frac{d}{dx}[g(x)]}{[g(x)]^2}$

   - Applies to the quotient of two functions.

   Example:
   If $f(x) = x^2$ and $g(x) = 3x$,
   $\frac{d}{dx}\left[\frac{x^2}{3x}\right] = \frac{3x \cdot \frac{d}{dx}[x^2] - x^2 \cdot \frac{d}{dx}[3x]}{(3x)^2} = \frac{3x^2 - 3x^2}{9x^2} = \frac{1}{3}$

Taylor Series

Taylor series and Taylor polynomials are essential concepts in calculus, particularly in approximating functions with polynomials around a specific point

The Taylor series of a function $f(x)$ around a point a is an infinite series that represents $f(x)$ as an infinite sum of terms involving the function's derivatives evaluated at a:

$f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!}(x - a)^n$

where:

- $f^{(n)}(a)$ denotes the n-th derivative of f evaluated at a.
- $n!$ denotes the factorial of n.

Taylor Polynomial

A Taylor polynomial of degree n for a function $f(x)$ around a point a is a polynomial of the form:

$$P_n(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \cdots + \frac{f^{(n)}(a)}{n!}(x-a)^n$$

It approximates $f(x)$ near a with increasing accuracy as n increases.

Chain Rule

The chain rule is a fundamental rule in calculus used to find the derivative of composite functions. It allows us to differentiate a function that is composed of other functions, where each function depends on the variable. The chain rule tells us how changes in the variable **x** affect the output **y** through the intermediate variable **u** It's crucial for differentiating complex functions that are composed of simpler functions. By breaking down the function into smaller parts and applying the chain rule iteratively, we can find derivatives of functions with multiple layers of composition.

Statement of the Chain Rule

Let $y = f(u)$ and $u = g(x)$ be two functions such that y depends on u, and u depends on x. The chain rule states:

$$\frac{dy}{dx} = \frac{dy}{du} \cdot \frac{du}{dx}$$

Explanation with an Example

Let's demonstrate the chain rule with a specific example:

Consider the functions:

- $u = g(x) = x^2 + 3x$
- $y = f(u) = \sin(u)$

We want to find $\frac{dy}{dx}$.

Step-by-Step Solution

1. **Find $\frac{du}{dx}$:**

   $$\frac{du}{dx} = \frac{d}{dx}[x^2 + 3x] = 2x + 3$$

2. **Find $\frac{dy}{du}$:**

   $$\frac{dy}{du} = \frac{d}{du}[\sin(u)] = \cos(u)$$

3. **Apply the Chain Rule:**

   Now, apply the chain rule using $\frac{dy}{du}$ and $\frac{du}{dx}$:

   $$\frac{dy}{dx} = \frac{dy}{du} \cdot \frac{du}{dx} = \cos(u) \cdot (2x + 3)$$

4. **Substitute $u = g(x)$:**

   $$\frac{dy}{dx} = \cos(x^2 + 3x) \cdot (2x + 3)$$

Partial Differentiation

Partial Differentiation refers to the process of finding the derivative of a function with respect to one of its variables, while holding all other variables constant. It extends the concept of differentiation from single-variable functions to functions with multiple variables.

For example, if $f(x, y)$ is a function of x and y, then $\frac{\partial f}{\partial x}$ represents the partial derivative of f with respect to x, treating y as constant.

Partial Derivatives

Partial Derivatives are the results of performing partial differentiation. They are the derivatives of functions of several variables with respect to one of those variables, keeping all other variables constant.

For instance, for a function $f(x, y)$, $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$ are partial derivatives with respect to x and y, respectively.

Summary

- **Partial Differentiation:** The process of finding partial derivatives of a function with respect to one of its variables while holding others constant.
- **Partial Derivatives:** The values obtained from partial differentiation, representing rates of change of the function with respect to each variable individually.

In essence, partial differentiation is the action of computing partial derivatives, which are the specific values resulting from that differentiation process. These concepts are fundamental in multivariable calculus and play a crucial role in various applications across science, engineering, economics, and

Partial Differentiation

Partial differentiation is the process of finding the derivative of a function with respect to one of its several variables, holding the others constant. It extends the idea of differentiation from single-variable functions to functions with multiple variables

Example

Consider a function $f(x, y) = x^2 + 3y$.

Finding Partial Derivatives

1. **Partial Derivative with Respect to x**

   To find $\frac{\partial f}{\partial x}$, treat y as a constant:
   $\frac{\partial f}{\partial x} = \frac{\partial}{\partial x}(x^2 + 3y) = 2x$

   Here, $\frac{\partial f}{\partial x}$ represents how f changes with respect to x, while y remains constant.

2. **Partial Derivative with Respect to y**

   To find $\frac{\partial f}{\partial y}$, treat x as a constant:
   $\frac{\partial f}{\partial y} = \frac{\partial}{\partial y}(x^2 + 3y) = 3$

   $\frac{\partial f}{\partial y}$ shows how f changes with respect to y, while x remains constant.

Partial Derivatives: Measure how the function changes in the direction of each variable, holding others constant.

Gradient

The gradient of a scalar-valued function $f(x, y)$, denoted by ∇f or $\frac{\partial f}{\partial x}\mathbf{i} + \frac{\partial f}{\partial y}\mathbf{j}$, is a vector that points in the direction of the greatest rate of increase of f. It consists of the partial derivatives of f with respect to each variable.

Example

Let's compute the gradient of $f(x, y) = x^2 + 3y$.

Finding the Gradient

1. **Compute $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$:**

   From earlier calculations:

   - $\frac{\partial f}{\partial x} = 2x$
   - $\frac{\partial f}{\partial y} = 3$

2. **Form the Gradient Vector ∇f:**

   $$\nabla f = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right) = (2x, 3)$$

   The gradient ∇f indicates the direction and rate of the steepest ascent of f at any point (x, y).

Gradient: Provides a vector pointing towards the direction of the greatest increase of the function f.

Basic Rules of Partial Differentiation

1. **Constant Rule:**
   $\frac{\partial}{\partial x}(c) = 0$
   where c is a constant.

2. **Power Rule:**
   $\frac{\partial}{\partial x}(x^n) = nx^{n-1}$
   This rule applies similarly to functions of y or any other variable.

3. **Sum and Difference Rule:**
   $\frac{\partial}{\partial x}[f(x, y) \pm g(x, y)] = \frac{\partial f}{\partial x} \pm \frac{\partial g}{\partial x}$

4. **Product Rule:**
   $\frac{\partial}{\partial x}[f(x, y) \cdot g(x, y)] = \frac{\partial f}{\partial x} \cdot g(x, y) + f(x, y) \cdot \frac{\partial g}{\partial x}$

5. **Quotient Rule:**
   $\frac{\partial}{\partial x}\left[\frac{f(x,y)}{g(x,y)}\right] = \frac{\frac{\partial f}{\partial x} \cdot g(x,y) - f(x,y) \cdot \frac{\partial g}{\partial x}}{(g(x,y))^2}$
   This rule assumes $g(x, y) \neq 0$.

6. **Chain Rule:**

   If $z = f(x, y)$ and $x = g(t)$, $y = h(t)$, then
   $\frac{\partial z}{\partial t} = \frac{\partial z}{\partial x} \cdot \frac{\partial x}{\partial t} + \frac{\partial z}{\partial y} \cdot \frac{\partial y}{\partial t}$

7. **Mixed Partial Derivatives:**

   If $f(x, y)$ has continuous second partial derivatives, then
   $\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right) = \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right)$
   This expresses the mixed partial derivative, which indicates the rate of change of the rate of change.

Gradient of Vector-Valued Functions

The gradient of a vector-valued function is a vector that contains the partial derivatives of each component function with respect to the variables.

Gradient of Vector-Valued Functions

Consider a vector-valued function $\mathbf{F}(x, y) = (f(x, y), g(x, y))$.

Definition of Gradient

The gradient of $\mathbf{F}$, denoted as $\nabla\mathbf{F}$, is given by:
$\nabla\mathbf{F} = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right)$

This represents a vector containing the partial derivatives of f and g with respect to x and y, respectively.

Example

Let's compute the gradient of the vector-valued function $\mathbf{F}(x, y) = (x^2 + 3y, 2x - y)$.

Finding the Gradient $\nabla\mathbf{F}$

1. **Compute $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$:**

   For $f(x, y) = x^2 + 3y$:

   - $\frac{\partial f}{\partial x} = 2x$
   - $\frac{\partial f}{\partial y} = 3$

2. **Compute $\frac{\partial g}{\partial x}$ and $\frac{\partial g}{\partial y}$:**

   For $g(x, y) = 2x - y$:

   - $\frac{\partial g}{\partial x} = 2$
   - $\frac{\partial g}{\partial y} = -1$

3. **Form the Gradient $\nabla \mathbf{F}$:**

$$\nabla \mathbf{F} = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right) = (2x, 3)$$
$$\nabla \mathbf{F} = \left(\frac{\partial g}{\partial x}, \frac{\partial g}{\partial y}\right) = (2, -1)$$

So, the gradient $\nabla \mathbf{F}$ of the vector-valued function $\mathbf{F}(x, y)$ is $(2x, 3)$ for the component $f(x, y)$, and $(2, -1)$ for the component $g(x, y)$.

Interpretation

- The gradient $\nabla \mathbf{F}$ points in the direction of the greatest rate of change of the vector-valued function $\mathbf{F}$.
- Each component of $\nabla \mathbf{F}$ corresponds to the rate of change of the corresponding component function f or g with respect to x and y.

Gradients of Vectors with Respect to Matrices

When we have a vector-valued function $\mathbf{g}(\mathbf{X})$, where $\mathbf{X}$ is an $m \times n$ matrix, the gradient $\nabla_{\mathbf{X}} \mathbf{g}(\mathbf{X})$ is an $m \times n$ matrix. Each element of this matrix represents the partial derivative of the corresponding element of $\mathbf{g}(\mathbf{X})$ with respect to each element of $\mathbf{X}$.

Example

Let $\mathbf{g}(\mathbf{X}) = \begin{bmatrix} x_{11}^2 + x_{21} \\ x_{12} + x_{22}^2 \end{bmatrix}$, where $\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix}$.

1. **Compute $\mathbf{g}(\mathbf{X})$:**
   $$\mathbf{g}(\mathbf{X}) = \begin{bmatrix} x_{11}^2 + x_{21} \\ x_{12} + x_{22}^2 \end{bmatrix}$$

Compute $\nabla_{\mathbf{X}} \mathbf{g}(\mathbf{X})$:

- For $g_1(\mathbf{X}) = x_{11}^2 + x_{21}$:
  $\frac{\partial g_1}{\partial x_{11}} = 2x_{11}, \quad \frac{\partial g_1}{\partial x_{21}} = 1$

- For $g_2(\mathbf{X}) = x_{12} + x_{22}^2$:
  $\frac{\partial g_2}{\partial x_{12}} = 1, \quad \frac{\partial g_2}{\partial x_{22}} = 2x_{22}$

Therefore,

$$\nabla_{\mathbf{X}}\mathbf{g}(\mathbf{X}) = \begin{bmatrix} 2x_{11} & 0 \\ 0 & 2x_{22} \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 1 \\ 0 & 2x_{22} \end{bmatrix} = \begin{bmatrix} 2x_{11} & 1 \\ 0 & 2x_{22} \end{bmatrix}$$

Gradients of Matrices with Respect to Matrices

When dealing with matrices, we are interested in how one matrix changes concerning another matrix. Suppose we have a matrix-valued function $\mathbf{G}(\mathbf{X})$, where $\mathbf{X}$ and $\mathbf{Y}$ are matrices of the same size. The gradient $\nabla_{\mathbf{X}}\mathbf{G}(\mathbf{X})$ with respect to $\mathbf{X}$ is also a matrix of the same size as $\mathbf{X}$. Each element of this matrix represents the partial derivative of the corresponding element of $\mathbf{G}(\mathbf{X})$ with respect to each element of $\mathbf{X}$.

Example

Let $\mathbf{G}(\mathbf{X}) = \mathbf{X}^T\mathbf{X}$, where $\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix}$.

1. **Compute $\mathbf{G}(\mathbf{X})$:**

$$\mathbf{G}(\mathbf{X}) = \mathbf{X}^T\mathbf{X} = \begin{bmatrix} x_{11} & x_{21} \\ x_{12} & x_{22} \end{bmatrix}^T \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix} = \begin{bmatrix} x_{11}^2 + x_{21}^2 & x_{11}x_{12} + x_{21}x_{22} \\ x_{12}x_{11} + x_{22}x_{21} & x_{12}^2 + x_{22}^2 \end{bmatrix}$$

2. **Compute $\nabla_{\mathbf{X}}\mathbf{G}(\mathbf{X})$:**

   - For $g_{11}(\mathbf{X}) = x_{11}^2 + x_{21}^2$:
     $\frac{\partial g_{11}}{\partial x_{11}} = 2x_{11}, \quad \frac{\partial g_{11}}{\partial x_{21}} = 2x_{21}$
     $\frac{\partial g_{11}}{\partial x_{12}} = 0, \quad \frac{\partial g_{11}}{\partial x_{22}} = 0$

   - For $g_{12}(\mathbf{X}) = x_{11}x_{12} + x_{21}x_{22}$:
     $\frac{\partial g_{12}}{\partial x_{11}} = x_{12}, \quad \frac{\partial g_{12}}{\partial x_{21}} = x_{22}$
     $\frac{\partial g_{12}}{\partial x_{12}} = x_{11}, \quad \frac{\partial g_{12}}{\partial x_{22}} = x_{21}$

   - And similarly for $g_{21}(\mathbf{X})$ and $g_{22}(\mathbf{X})$.

   Therefore,

$$\nabla_{\mathbf{X}}\mathbf{G}(\mathbf{X}) = \begin{bmatrix} 2x_{11} & 2x_{21} \\ 2x_{12} & 2x_{22} \end{bmatrix}$$

Gradients with respect to matrices: Provide information about how scalar or matrix-valued functions change concerning matrix variables.

Backpropagation and Automatic Differentiation (AD)

Backpropagation and Automatic Differentiation (AD) are fundamental techniques in machine learning and optimization, particularly for training neural networks

Backpropagation

Backpropagation is a method used to calculate the gradient of a loss function with respect to the weights of a neural network. It efficiently computes these gradients by leveraging the chain rule of calculus, propagating gradients backward through the network layers.

Calculates gradients of a neural network's loss function with respect to its weights using the chain rule.

Steps of Backpropagation:

1. **Forward Pass:**
    a. Compute the output of the neural network for a given input.
    b. Calculate the loss function that measures the difference between the predicted output and the actual output.
2. **Backward Pass (Backpropagation):**
    a. Compute the gradient of the loss function with respect to each weight in the network using the chain rule.
    b. Update the weights using an optimization algorithm like gradient descent to minimize the loss function.

Example:

Let's consider a simple neural network with one input layer, one hidden layer, and one output layer. Assume a Mean Squared Error (MSE) loss function.

- **Forward Pass:**
  Suppose the input $\mathbf{x} = [x_1, x_2]$ is fed into the network, and after passing through the hidden layer and output layer with weights $\mathbf{W}^{(1)}$ and $\mathbf{W}^{(2)}$, the predicted output $\hat{y}$ is computed.

  $\mathbf{h} = \sigma(\mathbf{W}^{(1)}\mathbf{x} + \mathbf{b}^{(1)})$
  $\hat{y} = \sigma(\mathbf{W}^{(2)}\mathbf{h} + b^{(2)})$

  Where σ is the activation function (e.g., sigmoid or ReLU).

- **Loss Function:**
  Compute the MSE loss:
  $\mathcal{L}(\hat{y}, y) = \frac{1}{2}(\hat{y} - y)^2$

Backward Pass:

- Compute gradients of the loss function with respect to the weights:

  For $\mathbf{W}^{(2)}$:

  $$\frac{\partial \mathcal{L}}{\partial \mathbf{W}^{(2)}} = (\hat{y} - y)\sigma'(\mathbf{W}^{(2)}\mathbf{h} + b^{(2)})\mathbf{h}^T$$

  For $\mathbf{W}^{(1)}$:

  $$\frac{\partial \mathcal{L}}{\partial \mathbf{W}^{(1)}} = (\hat{y} - y)\sigma'(\mathbf{W}^{(2)}\mathbf{h} + b^{(2)})(\mathbf{W}^{(2)})^T\sigma'(\mathbf{W}^{(1)}\mathbf{x} + \mathbf{b}^{(1)})\mathbf{x}^T$$

- Update the weights using gradient descent:

  $$\mathbf{W}^{(2)}: = \mathbf{W}^{(2)} - \eta \frac{\partial \mathcal{L}}{\partial \mathbf{W}^{(2)}}$$

  $$\mathbf{W}^{(1)}: = \mathbf{W}^{(1)} - \eta \frac{\partial \mathcal{L}}{\partial \mathbf{W}^{(1)}}$$

Automatic Differentiation (AD)

Automatic Differentiation is a computational technique that computes derivatives of functions numerically. It breaks down complex functions into elementary operations, computing derivatives efficiently and accurately.

Computes derivatives of functions numerically, useful for optimization and gradient-based methods in machine learning.

Types of AD:

1. **Forward Mode AD:**
   - Computes derivatives by propagating values and their derivatives from inputs to outputs.
   - Suitable for functions where the number of inputs is less than the number of outputs.
2. **Reverse Mode AD:**
   - Computes derivatives by propagating gradients backward from outputs to inputs.
   - Efficient for functions with many inputs and few outputs, as in neural networks.

Example:

Consider a simple function $f(x, y) = x^2 + \sin(y)$.

- **Forward Mode AD**:
  - Evaluate $f(x, y)$ and its derivatives with respect to x and y simultaneously.
- **Reverse Mode AD**:
  - Compute gradients by first evaluating $f(x, y)$ forward, then propagating gradients backward to x and y.

Formalization of Automatic Differentiation (AD)

Formalization of Automatic Differentiation (AD) involves understanding how to systematically compute derivatives of functions using elementary operations and the chain rule.

Automatic Differentiation (AD) is a technique for computing derivatives of functions that can be decomposed into a sequence of elementary arithmetic operations and elementary functions (such as exponential, logarithmic, trigonometric functions, etc.). The key idea is to evaluate derivatives efficiently, often by leveraging computational graphs that represent the function and its derivative computations.

Key Concepts:

1. **Computational Graph:**
    a. Represents the function as a directed acyclic graph (DAG), where nodes represent operations and edges represent dependencies.
    b. Each node in the graph computes a value and its derivative based on the chain rule.
2. **Dual Numbers:**
    a. Extend real numbers with an additional component (dual part) representing the derivative.
    b. ϵ is defined such that $\epsilon^2 = 0$
3. **Forward Mode and Reverse Mode:**
    a. **Forward Mode AD:** Computes derivatives by propagating values and their derivatives from inputs to outputs.
    b. **Reverse Mode AD:** Computes derivatives by propagating gradients backward from outputs to inputs.

Formalization of AD: Involves extending real numbers with a dual part representing derivatives, evaluating functions using dual numbers, and leveraging computational graphs.

Application: AD is widely used in machine learning for computing gradients efficiently, enabling optimization algorithms like gradient descent to train deep neural networks effectively.

By formalizing AD, we gain insights into how derivatives can be computed efficiently and accurately

Higher-Order Derivatives

Higher-order derivatives refer to the derivatives of a function beyond the first derivative

Higher-Order Derivatives

1. **Second-Order Derivative:**
   - The second-order derivative of a function f with respect to a variable x is denoted as $\frac{\partial^2 f}{\partial x^2}$ or $f''(x)$.
   - It represents how the rate of change of the function's slope changes with respect to x.
2. **Higher-Order Derivatives:**
   - Similarly, the n-th order derivative of f with respect to x is denoted as $\frac{\partial^n f}{\partial x^n}$ or $f^{(n)}(x)$.
   - It indicates the rate of change of the $(n-1)$-th order derivative of f with respect to x.

Notations for Higher-Order Partial Derivatives

In the context of multivariable calculus, where functions depend on multiple variables, higher-order partial derivatives extend the concept of differentiation to multiple dimensions.

1. **Second-Order Partial Derivatives:**
   - For a function $f(x, y)$, the second-order partial derivatives are:
     - $\frac{\partial^2 f}{\partial x^2}$: Second partial derivative of f with respect to x.
     - $\frac{\partial^2 f}{\partial y^2}$: Second partial derivative of f with respect to y.
     - $\frac{\partial^2 f}{\partial x \partial y}$ or $\frac{\partial^2 f}{\partial y \partial x}$: Mixed partial derivatives (derivative with respect to x, then y or vice versa).
2. **Higher-Order Partial Derivatives:**
   - The notation extends straightforwardly for higher-order partial derivatives:
     - $\frac{\partial^n f}{\partial x^n}$: n-th partial derivative with respect to x.
     - $\frac{\partial^n f}{\partial y^n}$: n-th partial derivative with respect to y.
     - $\frac{\partial^n f}{\partial x^{m_1} \partial y^{m_2} \ldots \partial z^{m_k}}$: Mixed partial derivatives involving multiple variables.

Example

Consider a function $f(x, y) = x^2 + y^3$.

- **First-Order Partial Derivatives:**
  $\frac{\partial f}{\partial x} = 2x$
  $\frac{\partial f}{\partial y} = 3y^2$

- **Second-Order Partial Derivatives:**
  $\frac{\partial^2 f}{\partial x^2} = 2$
  $\frac{\partial^2 f}{\partial y^2} = 6y$
  $\frac{\partial^2 f}{\partial x \partial y} = 0$ (since $\frac{\partial}{\partial x}(3y^2) = 0$)

- **Higher-Order Partial Derivatives:**
  $\frac{\partial^3 f}{\partial x^2 \partial y} = 0$
  $\frac{\partial^3 f}{\partial y^3} = 6$

Hessian matrix

The Hessian matrix is a square matrix of second-order partial derivatives of a scalar-valued function of several variables. It captures information about the curvature of the function's graph at a given point and is an essential tool in optimization and the study of critical points of functions

Definition

Given a scalar-valued function $f: \mathbb{R}^n \to \mathbb{R}$, the Hessian matrix $H(f)$ is an $n \times n$ matrix whose entries H_{ij} are the second partial derivatives of f:

$$H_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j}$$

Here, $x_1, x_2, ..., x_n$ are the variables in $\mathbb{R}^n$.

Properties

1. **Symmetry:**
   - The Hessian matrix $H(f)$ is symmetric, meaning $H_{ij} = H_{ji}$ for all i, j.
   - This symmetry arises because second partial derivatives commute (i.e., $\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i}$).

2. **Curvature:**

   - The Hessian matrix provides information about the local curvature of the function f.
   - Positive definiteness of the Hessian matrix at a point indicates a local minimum, negative definiteness indicates a local maximum, and indefinite Hessian suggests a saddle point.

3. **Eigenvalues:**

   - The eigenvalues of the Hessian matrix $H(f)$ are related to the principal curvatures of the graph of f at critical points.
   - Positive eigenvalues indicate directions of positive curvature, negative eigenvalues indicate directions of negative curvature, and zero eigenvalues indicate directions of no curvature.

Example

Consider the function $f(x, y) = x^2 + 2xy + y^2$.

- **Compute the Hessian Matrix:**

$$H(f) = \begin{bmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2} \end{bmatrix}$$

  Compute each partial derivative:

$$\frac{\partial^2 f}{\partial x^2} = 2, \quad \frac{\partial^2 f}{\partial y^2} = 2, \quad \frac{\partial^2 f}{\partial x \partial y} = 2, \quad \frac{\partial^2 f}{\partial y \partial x} = 2$$

  Therefore, the Hessian matrix $H(f)$ is:

$$H(f) = \begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix}$$

- **Interpretation:**

  - The Hessian matrix $H(f) = \begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix}$ indicates that the function $f(x, y) = x^2 + 2xy + y^2$ has positive eigenvalues (2 and 0) and is positive semi-definite.
  - This suggests that $(0, 0)$ is a local minimum point for the function f.

Applications

Optimization: The Hessian matrix is crucial in optimization algorithms like Newton's method and quasi-Newton methods, where it helps determine the direction and step size towards minimizing or maximizing the function.

Eigenvalue Analysis: By analysing the eigenvalues of the Hessian matrix, one can understand the nature of critical points (minima, maxima, saddle points) of the function.

In summary, the Hessian matrix is a powerful mathematical tool that encapsulates information about the curvature and critical points of a scalar-valued function of several variables, making it indispensable in optimization and theoretical analyses.

Linearization

Linearization and the Multivariate Taylor Series are methods to approximate a multivariable function around a point using its derivatives

Linearization is the process of approximating a nonlinear function near a point by a linear function. It involves using the first-order Taylor series expansion of the function.

For a function $f(x)$ of one variable, the linear approximation (or linearization) around a point $x = a$ is given by:

$$f(x) \approx f(a) + f'(a)(x - a)$$

where $f'(a)$ is the derivative of f at $x = a$.

Multivariate Taylor Series

The Multivariate Taylor Series expansion generalizes this concept to functions of multiple variables. For a function $f(\mathbf{x})$ of n variables $\mathbf{x} = (x_1, x_2, ..., x_n)$, the Multivariate Taylor Series expansion around a point $\mathbf{a} = (a_1, a_2, ..., a_n)$ is:

$$f(\mathbf{x}) \approx f(\mathbf{a}) + \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\mathbf{a})(x_i - a_i) + \frac{1}{2!}\sum_{i,j=1}^{n} \frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{a})(x_i - a_i)(x_j - a_j) + \cdots$$

where:

- $f(\mathbf{a}) = f(a_1, a_2, ..., a_n)$ is the value of f at $\mathbf{a}$.
- $\frac{\partial f}{\partial x_i}(\mathbf{a})$ are the partial derivatives of f at $\mathbf{a}$.
- $\frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{a})$ are the second partial derivatives of f at $\mathbf{a}$, and so on.

Example

Let's find the linearization and the Multivariate Taylor Series expansion of the function $f(x, y) = e^{x+y}$ around the point $(a, b) = (0, 0)$.

1. **Linearization:**

   Compute $f(0,0)$ and $\frac{\partial f}{\partial x}(0,0)$, $\frac{\partial f}{\partial y}(0,0)$:

   $f(0,0) = e^0 = 1$
   $\frac{\partial f}{\partial x} = e^{x+y}\Big|_{(0,0)} = e^0 = 1$
   $\frac{\partial f}{\partial y} = e^{x+y}\Big|_{(0,0)} = e^0 = 1$

   Therefore, the linearization of $f(x,y)$ around $(0,0)$ is:

   $f(x,y) \approx 1 + 1 \cdot x + 1 \cdot y = 1 + x + y$

2. **Multivariate Taylor Series Expansion:**

   Compute $\frac{\partial^2 f}{\partial x^2}$, $\frac{\partial^2 f}{\partial y^2}$, and $\frac{\partial^2 f}{\partial x \partial y}$ at $(0,0)$:

   $\frac{\partial^2 f}{\partial x^2} = \frac{\partial}{\partial x}(e^{x+y})\Big|_{(0,0)} = e^{x+y}\Big|_{(0,0)} = 1$
   $\frac{\partial^2 f}{\partial y^2} = \frac{\partial}{\partial y}(e^{x+y})\Big|_{(0,0)} = e^{x+y}\Big|_{(0,0)} = 1$
   $\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial}{\partial y}(e^{x+y})\Big|_{(0,0)} = e^{x+y}\Big|_{(0,0)} = 1$

   The Multivariate Taylor Series expansion around $(0,0)$ is:

   $f(x,y) \approx 1 + (x+y) + \frac{1}{2!}(x^2 + 2xy + y^2)$

   Simplifying, we get:

   $f(x,y) \approx 1 + x + y + \frac{1}{2}(x^2 + 2xy + y^2)$

Linearization: Approximates a function near a point using its first-order Taylor series.

Multivariate Taylor Series: Generalizes the concept to functions of multiple variables, incorporating higher-order partial derivatives to provide a more accurate approximation.

Taylor Polynomial

Given a function $f(x)$ that is infinitely differentiable at a point a, the n-th degree Taylor polynomial $T_n(x)$ of f around a is:

$$T_n(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \cdots + \frac{f^{(n)}(a)}{n!}(x-a)^n$$

where:

- $f(a)$ is the value of f at $x = a$,
- $f'(a)$ is the first derivative of f evaluated at $x = a$,
- $f''(a)$ is the second derivative of f evaluated at $x = a$,
- $f^{(n)}(a)$ is the n-th derivative of f evaluated at $x = a$.

Purpose

The Taylor polynomial $T_n(x)$ approximates $f(x)$ near $x = a$. As n increases, $T_n(x)$ becomes a more accurate approximation of $f(x)$ around a. The first-degree Taylor polynomial $T_1(x)$ is known as the linear approximation or linearization of f at a.

Example

Let's find the Taylor polynomials $T_1(x)$ and $T_2(x)$ for the function $f(x) = \sin(x)$ around the point $a = 0$.

1. **First-Degree Taylor Polynomial $T_1(x)$:**

   - $f(0) = \sin(0) = 0$
   - $f'(x) = \cos(x)$, so $f'(0) = \cos(0) = 1$

   Therefore,
   $T_1(x) = f(0) + f'(0)(x - 0) = 0 + 1 \cdot x = x$

   $T_1(x)$ approximates $\sin(x)$ near $x = 0$.

2. **Second-Degree Taylor Polynomial $T_2(x)$:**

   - $f''(x) = -\sin(x)$, so $f''(0) = -\sin(0) = 0$

   Therefore,
   $T_2(x) = T_1(x) + \frac{f''(0)}{2!}(x - 0)^2 = x + \frac{0}{2}(x)^2 = x$

   $T_2(x)$ also equals x since the second derivative of $\sin(x)$ at $x = 0$ is zero.

Approximation: Taylor polynomials are used to approximate functions in numerical analysis, physics, and engineering.

Analysis: They help analyze the behavior of functions around critical points and in optimization problems.

Higher-Order Approximations: Higher-order Taylor polynomials provide more accurate approximations for functions with more complex behavior.

In summary, the Taylor polynomial provides a powerful method to approximate functions locally using their derivatives,

Taylor series expansion for a function of two variables

The Taylor series expansion for a function of two variables f(x, y) around a point (a, b) generalizes the concept of the Taylor polynomial to multiple variables.

Taylor Series Expansion for $f(x, y)$

The Taylor series expansion for a function $f(x, y)$ around a point (a, b) up to the n-th degree is given by:

$$f(x,y) \approx f(a,b) + \frac{\partial f}{\partial x}(a,b)(x-a) + \frac{\partial f}{\partial y}(a,b)(y-b) + \frac{1}{2!}\left[\frac{\partial^2 f}{\partial x^2}(a,b)(x-a)^2 + 2\frac{\partial^2 f}{\partial x \partial y}(a,b)(x-a)(y-l\right.$$

where:

- $f(a, b)$ is the value of f at (a, b),
- $\frac{\partial f}{\partial x}(a, b)$ and $\frac{\partial f}{\partial y}(a, b)$ are the partial derivatives of f with respect to x and y evaluated at (a, b),
- $\frac{\partial^2 f}{\partial x^2}(a, b)$, $\frac{\partial^2 f}{\partial y^2}(a, b)$, and $\frac{\partial^2 f}{\partial x \partial y}(a, b)$ are the second partial derivatives of f evaluated at (a, b),
- Higher-order terms involve higher-order partial derivatives.

Example

Let's find the Taylor series expansion of $f(x, y) = e^{x+y}$ around the point $(a, b) = (0, 0)$ up to the second degree.

1. **Compute $f(a, b)$ and Partial Derivatives:**
   - $f(0, 0) = e^{0+0} = 1$
   - $\frac{\partial f}{\partial x} = e^{x+y}\Big|_{(0,0)} = 1$
   - $\frac{\partial f}{\partial y} = e^{x+y}\Big|_{(0,0)} = 1$
   - $\frac{\partial^2 f}{\partial x^2} = \frac{\partial}{\partial x}(e^{x+y})\Big|_{(0,0)} = e^{x+y}\Big|_{(0,0)} = 1$
   - $\frac{\partial^2 f}{\partial y^2} = \frac{\partial}{\partial y}(e^{x+y})\Big|_{(0,0)} = e^{x+y}\Big|_{(0,0)} = 1$
   - $\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial}{\partial y}(e^{x+y})\Big|_{(0,0)} = e^{x+y}\Big|_{(0,0)} = 1$

2. **Construct the Taylor Series Expansion:**

   Substitute these values into the Taylor series formula:

   $$f(x,y) \approx 1 + 1 \cdot x + 1 \cdot y + \frac{1}{2!}[1 \cdot (x-0)^2 + 2 \cdot 1 \cdot (x-0)(y-0) + 1 \cdot (y-0)^2]$$

   Simplify the expression:

   $$f(x,y) \approx 1 + x + y + \frac{1}{2}(x^2 + 2xy + y^2)$$

   So, the Taylor series expansion of e^{x+y} around $(0,0)$ up to the second degree is $1 + x + y + \frac{1}{2}(x^2 + 2xy + y^2)$.

Applications

Approximation: The Taylor series expansion provides a method to approximate functions with higher accuracy near a given point.

Analysis: It helps analyze the behavior of functions locally, which is useful in physics, engineering, and mathematical modeling.

Computational Methods: Higher-order terms in the series can be used for numerical methods to solve differential equations and optimization problems.

In summary, the Taylor series expansion for functions of two variables extends the concept of Taylor polynomials to multivariable functions, providing a powerful tool for local approximation and analysis in mathematical and scientific contexts.