Linear Algebra Summary for Optimization Prerequisites

1 Vectors and Matrices

Overview

Vectors and matrices are fundamental in optimization. Vectors represent variables (e.g., $x \in \mathbb{R}^n$), and matrices define transformations or Hessians (e.g., $\nabla^2 f$). Key operations include addition, scalar multiplication, dot products, and matrix-vector multiplication.

Example

Given:

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

- Compute $x + y$:

$$x + y = \begin{bmatrix} 1 \\ 2 \end{bmatrix} + \begin{bmatrix} 3 \\ 4 \end{bmatrix} = \begin{bmatrix} 4 \\ 6 \end{bmatrix}$$

- Compute the dot product $x^T y$:

$$x^T y = 1 \cdot 3 + 2 \cdot 4 = 3 + 8 = 11$$

- Compute Ax:

$$Ax = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \cdot 1 + 2 \cdot 2 \\ 3 \cdot 1 + 4 \cdot 2 \end{bmatrix} = \begin{bmatrix} 5 \\ 11 \end{bmatrix}$$

2 Symmetric and Positive Definite Matrices

Overview

A matrix A is symmetric if $A = A^T$. It is positive definite if for all $x \neq 0$, $x^T Ax > 0$, or equivalently, all its eigenvalues are positive. In optimization, the Hessian $\nabla^2 f$ is often symmetric, and positive definiteness ensures convexity (a unique minimum).

Example

Consider:

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

- **Check symmetry:**

$$A^T = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} = A \quad \text{(Symmetric)}$$

- **Check positive definiteness:** Compute eigenvalues:

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

$$\lambda = 3, \quad \lambda = 1$$

  Both eigenvalues are positive, so A is positive definite.

3 Eigenvalues and Eigenvectors

Overview

For a matrix A, an eigenvalue λ and eigenvector $v \neq 0$ satisfy $Av = \lambda v$. Eigenvalues of the Hessian determine convexity: all positive eigenvalues imply positive definiteness. The condition number $\kappa(A) = \frac{\lambda_{\max}}{\lambda_{\min}}$ affects convergence rates.

Example

Consider:

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

- **Eigenvalues:**

$$\det(A - \lambda I) = (3-\lambda)^2 - 1 = 0 \implies \lambda = 4, \quad \lambda = 2$$

- **Eigenvectors:**
  - For $\lambda = 4$:

$$A - 4I = \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix} \implies v_1 = v_2 \implies v = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

  - For $\lambda = 2$:

$$A - 2I = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \implies v_1 = -v_2 \implies v = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

4 Matrix Decompositions

Overview

Matrix decompositions simplify computations in optimization, such as solving linear systems or analyzing Hessians.

a) Cholesky Decomposition ($A = LL^T$)

For symmetric positive definite matrices.

- **Example:**

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

$$LL^T = \begin{bmatrix} l_{11} & 0 \\ l_{21} & l_{22} \end{bmatrix} \begin{bmatrix} l_{11} & l_{21} \\ 0 & l_{22} \end{bmatrix} = \begin{bmatrix} l_{11}^2 & l_{11}l_{21} \\ l_{11}l_{21} & l_{21}^2 + l_{22}^2 \end{bmatrix} = \begin{bmatrix} 4 & 2 \\ 2 & 2 \end{bmatrix}$$

  - $l_{11}^2 = 4 \implies l_{11} = 2$
  - $l_{11}l_{21} = 2 \implies 2l_{21} = 2 \implies l_{21} = 1$
  - $l_{21}^2 + l_{22}^2 = 1 + l_{22}^2 = 2 \implies l_{22} = 1$

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

b) LU Decomposition ($A = LU$)

For general matrices.

- **Example:**

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

  - Eliminate below the diagonal:
  - Column 1: Subtract $2 \times$ row 1 from row 2: $[2, 5, 1] - 2[1, 2, 0] = [0, 1, 1]$
  - Column 2: Subtract row 2 from row 3: $[0, 1, 3] - [0, 1, 1] = [0, 0, 2]$

$$U = \begin{bmatrix} 1 & 2 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 2 \end{bmatrix}, \quad L = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}$$

c) Singular Value Decomposition (SVD) ($A = U\Sigma V^T$)

For any matrix.

- **Example:**

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

  - Compute $A^T A$:

$$A^T A = \begin{bmatrix} 5 & 0 \\ 0 & 1 \end{bmatrix} \implies \sigma_1 = \sqrt{5}, \quad \sigma_2 = 1$$

  - Eigenvectors of $A^T A$:

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

– Compute U:

$$u_1 = \frac{1}{\sqrt{5}} \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix}, \quad u_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad u_3 = \begin{bmatrix} \frac{1}{\sqrt{5}} \\ 0 \\ -\frac{2}{\sqrt{5}} \end{bmatrix}$$

$$U = \begin{bmatrix} \frac{2}{\sqrt{5}} & 0 & \frac{1}{\sqrt{5}} \\ 0 & 1 & 0 \\ \frac{1}{\sqrt{5}} & 0 & -\frac{2}{\sqrt{5}} \end{bmatrix}, \quad \Sigma = \begin{bmatrix} \sqrt{5} & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$$

d) Eigenvalue Decomposition ($A = Q\Lambda Q^T$)

For symmetric matrices.

- **Example:**

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

  – Eigenvalues: $\lambda = 3, \quad \lambda = 1$
  – Eigenvectors (normalized):

$$Q = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}, \quad \Lambda = \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix}$$

5 Norms and Condition Numbers

Overview

- **Euclidean Norm (Vector):** $\|x\|_2 = \sqrt{x^T x}$
- **Matrix Norm:** $\|A\|_2 =$ largest singular value of A
- **Condition Number:** For a symmetric matrix, $\kappa(A) = \frac{\lambda_{\max}}{\lambda_{\min}}$. A large κ indicates an ill-conditioned matrix, slowing convergence in optimization.

Example

Consider:

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

- **Euclidean Norm of x:**

$$\|x\|_2 = \sqrt{1^2 + 2^2 + (-1)^2} = \sqrt{6} \approx 2.449$$

- **Condition Number of A:** Eigenvalues are $\lambda = 4, \quad \lambda = 2$:

$$\kappa(A) = \frac{\lambda_{\max}}{\lambda_{\min}} = \frac{4}{2} = 2$$

6 Final Notes

These concepts are crucial for understanding optimization algorithms in *Numerical Optimization* by Nocedal and Wright:

- Vectors and matrices form the basis of computations.
- Symmetric positive definite matrices ensure convexity.
- Eigenvalues/eigenvectors and condition numbers predict convergence behavior.
- Decompositions simplify solving linear systems.
- Norms measure errors and distances.