# Linear Algebra and Convexity Summary

## 0. Big Picture (One-Line Summary)

- Vector: numbers arranged in one line (point / direction / feature / parameter).
- Matrix: numbers arranged in a table (linear transformation).
- $\mathbb{R}^d$: space of real vectors with $d$ coordinates.
- Eigenvalue/eigenvector: direction preserved under transformation.
- Diagonalization: possible when enough eigenvectors exist.
- Orthogonal diagonalization: possible when matrix is symmetric.
- Hessian PSD $\Rightarrow$ convex function.

## 1 d-Dimensional Vector

Dimension = number of coordinates.

$$w = (w_1, w_2, \ldots, w_d)$$

Examples:

- $(3, 5)$: 2D vector
- $(1, 7, -2)$: 3D vector

## 2 $\mathbb{R}^d$

$$\mathbb{R}^d = \{(x_1, \ldots, x_d) \mid x_i \in \mathbb{R}\}$$

$$x \in \mathbb{R}^d$$

## 3 Matrix

$$A \in \mathbb{R}^{4\times 4}$$

Matrix represents a linear map:

$$y = Ax$$

## 4 Inner Product (Machine Learning Core)

$$\langle w, x\rangle = w^\top x = \sum_{i=1}^{d} w_i x_i$$

Neural network linear layer:

$$W \in \mathbb{R}^{m\times d}, \quad x \in \mathbb{R}^d$$

$$Wx \in \mathbb{R}^m$$

## 5 Eigenvalues and Eigenvectors

Definition:

$$Ax = \lambda x$$

Characteristic equation:

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

Nontrivial solution exists when:

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

## 6 Singular Matrix

Matrix $A$ is singular iff:

$$\det(A) = 0$$

Equivalent to:

- $A^{-1}$ does not exist
- 0 is an eigenvalue

# 7 Algebraic vs Geometric Multiplicity

- Algebraic multiplicity: multiplicity as root of characteristic polynomial
- Geometric multiplicity: dimension of eigenspace

$$1 \leq \text{geom mult} \leq \text{alg mult}$$

# 8 Diagonalization

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

Possible iff there exist $n$ linearly independent eigenvectors.

# 9 Orthogonal Diagonalization (Spectral Theorem)

If $A$ is symmetric:

$$A = QDQ^\top$$

$$Q^\top Q = I$$

# 10 PSD and PD

PSD:

$$x^\top Ax \geq 0 \quad \forall x$$

Equivalent:

$$\lambda_i \geq 0$$

PD:

$$x^\top Ax > 0 \quad \forall x \neq 0$$

Equivalent:

$$\lambda_i > 0$$

## 11 Matrix Square Root

If $A$ is symmetric and PSD:

$$A = Q\Lambda Q^\top$$

Define:

$$A^{1/2} = Q\Lambda^{1/2}Q^\top$$

where

$$\Lambda^{1/2} = \mathrm{diag}(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_n})$$

## 12 Convex Function

Definition:

$$f(\alpha u + (1-\alpha)v) \leq \alpha f(u) + (1-\alpha)f(v)$$

for all $\alpha \in [0, 1]$.

## 13 Convexity Tests

1D case:

$$f''(x) \geq 0 \Rightarrow f \text{ convex}$$

Multivariate case:

$$\nabla^2 f(w) \succeq 0 \Rightarrow f \text{ convex}$$

## 14 Why Hessian PSD Implies Convex

Let

$$g(t) = f(w + td)$$

Then:

$$g''(t) = d^\top \nabla^2 f(w + td)d$$

If Hessian is PSD:

$$d^\top \nabla^2 f d \geq 0$$

Thus:

$$g''(t) \geq 0$$

Therefore every 1D slice is convex $\Rightarrow$ $f$ is convex.

## 15 Optimization Meaning

- Hessian PSD $\Rightarrow$ convex (multiple minima possible)
- Hessian PD $\Rightarrow$ strictly convex (unique minimum)