252R 기초 데이터 분석 및 실습

4. Mathematics (2)

October 1, 2025

정재준

Welcome To Industrial Engineering

고려대학교
KOREA UNIVERSITY

Contents

252R 기초 데이터 분석 및 실습

- Intro
- Inner Product
- Orthogonality
- Projection
- Partial Derivatives & Gradient
- Taylor Expansion

Intro

Mathematics (2)

- 내적(inner product)과 직교(orthogonality) 개념을 이해하고, **벡터 간 유사도·길이·각도를 수치로 표현**할 수 있다.
- **직교정규(orthonormal) 개념**을 이해하고, Gram-Schmidt 과정을 통해 임의의 기저를 직교정규 기저로 변환할 수 있다.
- 투영(projection)의 정의와 공식을 유도하고, **부분공간으로의 투영을 최소제곱 및 선형회귀 문제와 연결** 지을 수 있다.
- 편미분(partial derivative)과 그래디언트(gradient)를 정의하고, **기하학적 의미(가장 빠른 증가 방향)를 이해**할 수 있다.
- **테일러 전개(Taylor expansion)를 통해 다변수 함수를 국소적으로 근사하는 방법**을 이해하고, 이를 최적화 알고리즘(gradient descent, Newton's method)과 연결할 수 있다.

Inner Product

Mathematics (2)

▪ Definition

$$x = (x_1, x_2, \ldots, x_n), \quad y = (y_1, y_2, \ldots, y_n)$$

$$\boldsymbol{x}^\top \boldsymbol{y} = \sum_{i=1}^{n} x_i y_i \,.$$

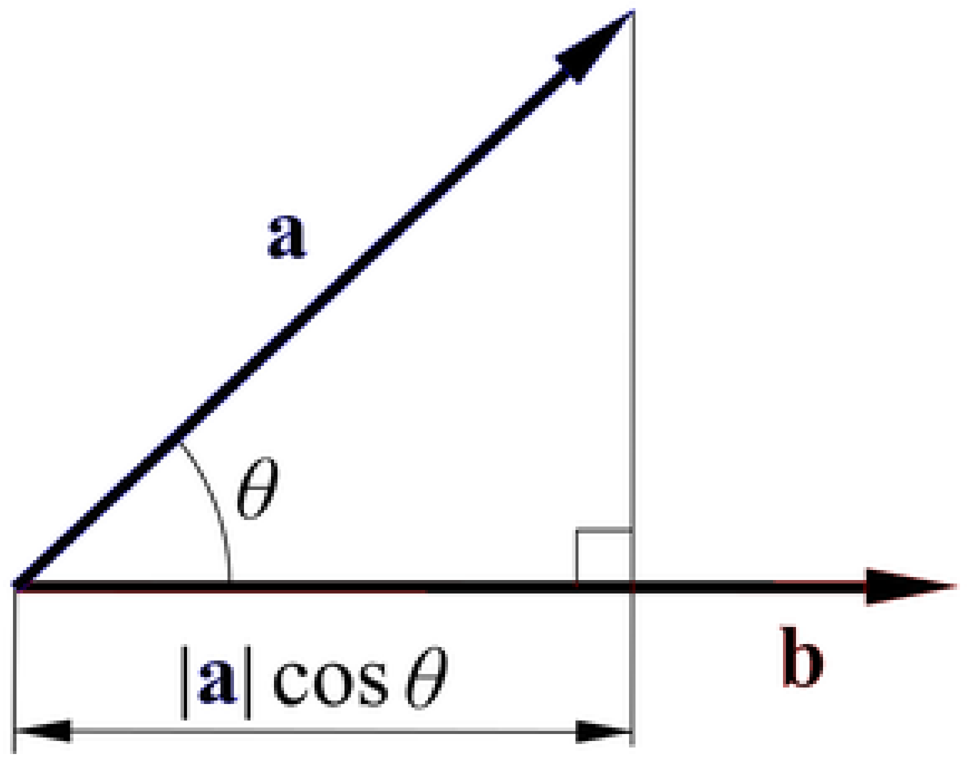
$$\|x\| = \sqrt{\langle x, x \rangle}$$

$x = (1, 2, 3), y = (4, 5, 6).$
$\langle x, y \rangle = 1 \cdot 4 + 2 \cdot 5 + 3 \cdot 6 = 32.$

Inner Product

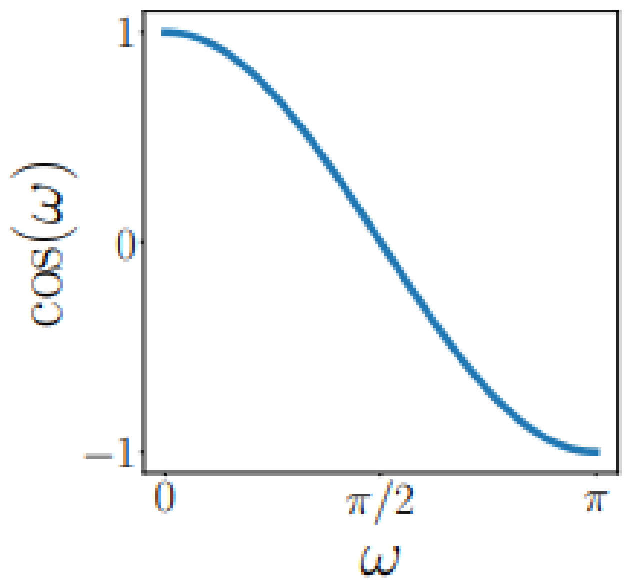
Mathematics (2)

- **Geometrically…**



$$\langle x, y \rangle = \|x\| \|y\| \cos\theta$$

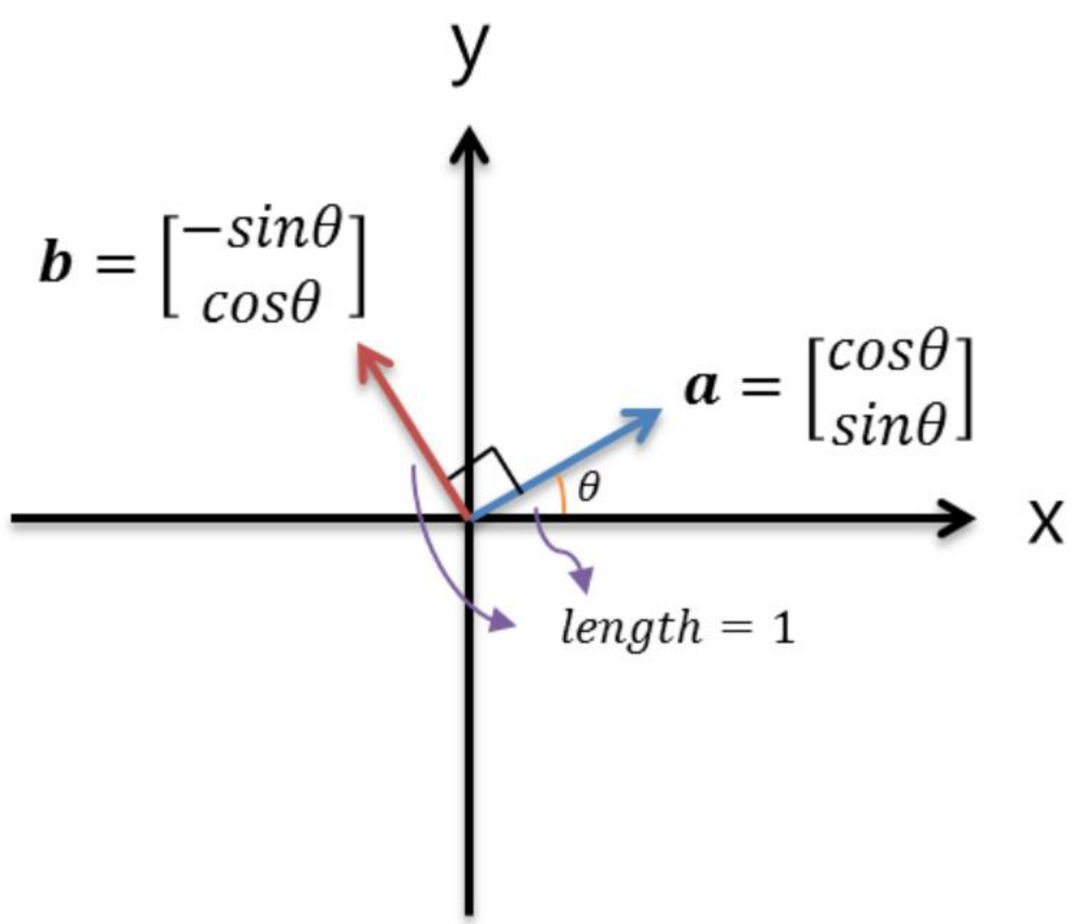
$$\cos\theta = \frac{\langle x, y \rangle}{\|x\| \|y\|}$$



▪ Orthogonal And Orthonomal

space V and a basis $\{\boldsymbol{b}_1, \dots, \boldsymbol{b}_n\}$ of V

$$\langle \boldsymbol{b}_i, \boldsymbol{b}_j \rangle = 0 \quad \text{for } i \neq j$$
$$\langle \boldsymbol{b}_i, \boldsymbol{b}_i \rangle = 1$$



▪ For Square Matrix A

$$\boldsymbol{A}\boldsymbol{A}^\top = \boldsymbol{I} = \boldsymbol{A}^\top\boldsymbol{A},$$

$$\boldsymbol{A}^{-1} = \boldsymbol{A}^\top,$$

$$\|\boldsymbol{A}\boldsymbol{x}\|^2 = (\boldsymbol{A}\boldsymbol{x})^\top(\boldsymbol{A}\boldsymbol{x}) = \boldsymbol{x}^\top\boldsymbol{A}^\top\boldsymbol{A}\boldsymbol{x} = \boldsymbol{x}^\top\boldsymbol{I}\boldsymbol{x} = \boldsymbol{x}^\top\boldsymbol{x} = \|\boldsymbol{x}\|^2 .$$

$$\cos\omega = \frac{(\boldsymbol{A}\boldsymbol{x})^\top(\boldsymbol{A}\boldsymbol{y})}{\|\boldsymbol{A}\boldsymbol{x}\|\,\|\boldsymbol{A}\boldsymbol{y}\|} = \frac{\boldsymbol{x}^\top\boldsymbol{A}^\top\boldsymbol{A}\boldsymbol{y}}{\sqrt{\boldsymbol{x}^\top\boldsymbol{A}^\top\boldsymbol{A}\boldsymbol{x}\boldsymbol{y}^\top\boldsymbol{A}^\top\boldsymbol{A}\boldsymbol{y}}} = \frac{\boldsymbol{x}^\top\boldsymbol{y}}{\|\boldsymbol{x}\|\,\|\boldsymbol{y}\|}$$

Projection

Why?

- Approximation
- Expression
- Modeling

Definition

$$\mathrm{proj}_u(x) = \langle x, u \rangle u$$

$x = (3, 4),\ \ u = \frac{1}{\sqrt{2}}(1, 1)$

$\mathrm{proj}_u(x) = (3.5, 3.5).$

Projection

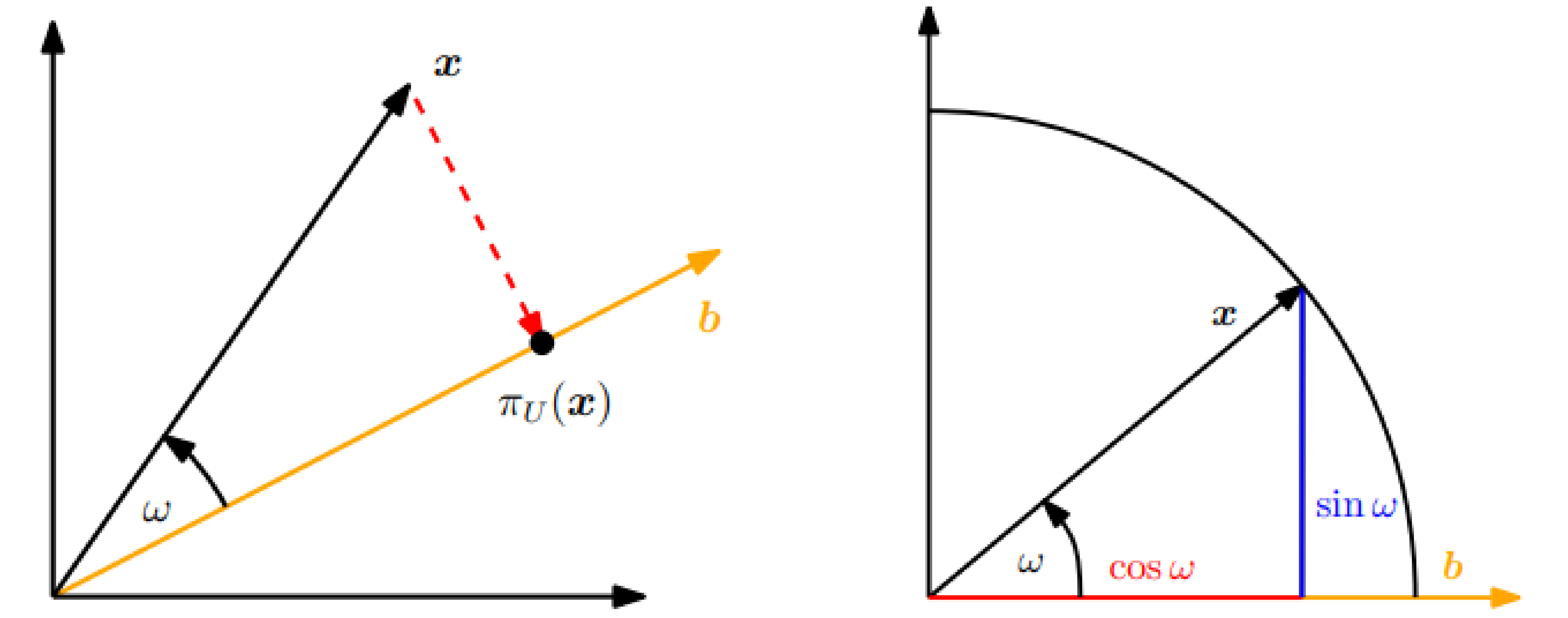
Mathematics (2)

▪ Find Closest Line of U to x

$$U = \mathrm{span}\{u_1, \dots, u_k\} \subset \mathbb{R}^n.$$

$$\min_c \|x - Uc\|^2.$$

Find c* to minimize(argmin problem)



(a) Projection of $x \in \mathbb{R}^2$ onto a subspace U with basis vector b.

(b) Projection of a two-dimensional vector x with $\|x\| = 1$ onto a one-dimensional subspace spanned by b.

$$x = \sum_{m=1}^{M} \lambda_m b_m + \cancel{\sum_{j=1}^{D-M} \psi_j b_j^{\perp}}, \quad \lambda_m, \psi_j \in \mathbb{R},$$

To Zero!

Projection

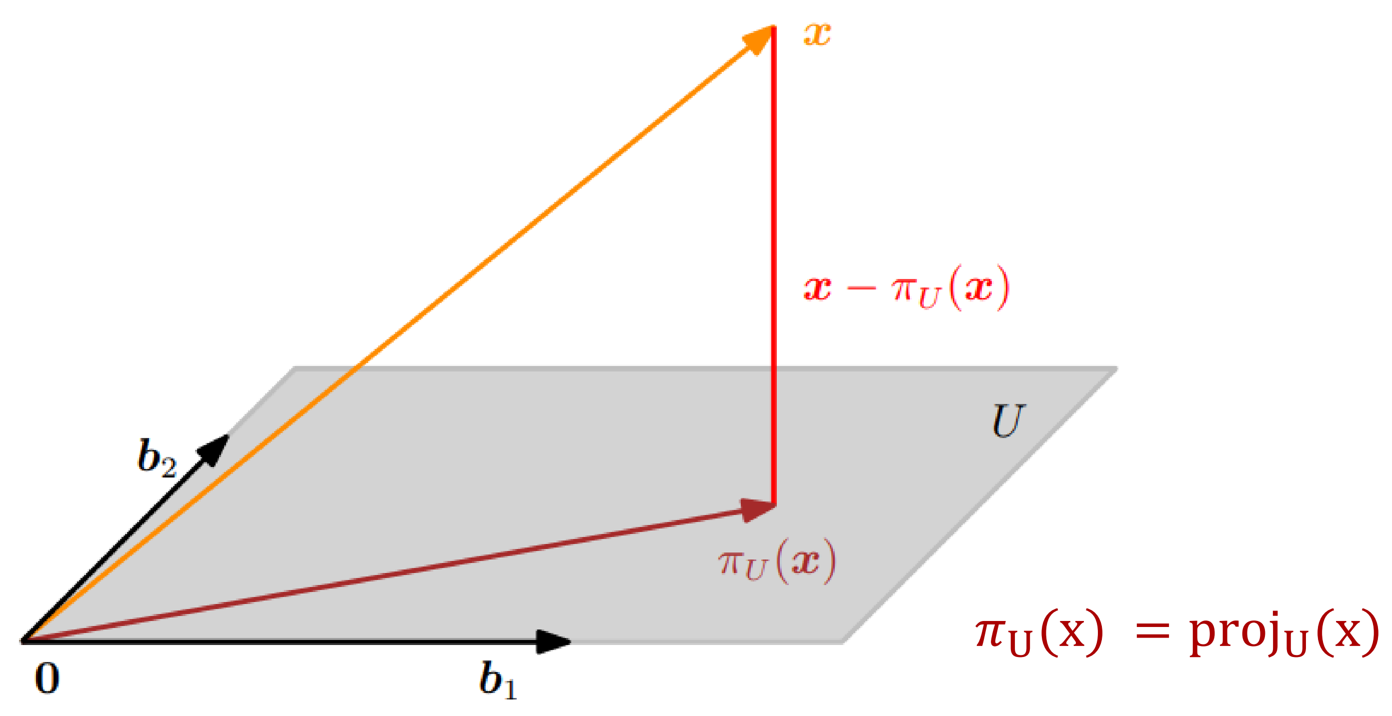
Mathematics (2)

- **Find Closest Line of U to x…**

$$U = \mathrm{span}\{u_1, \ldots, u_k\} \subset \mathbb{R}^n.$$

$$\min_c \|x - Uc\|^2.$$

Find c* to minimize(argmin problem)



$$\phi(c) = \|x - Uc\|^2 = (x - Uc)^\top (x - Uc).$$

$$\nabla_c \phi(c) = -2U^\top (x - Uc) = 0$$

$$\boxed{U^\top U c = U^\top x}$$

$$c^\star = (U^\top U)^{-1} U^\top x$$

$$\mathrm{proj}_U(x) = P_U x, \quad P_U = U(U^\top U)^{-1} U^\top.$$

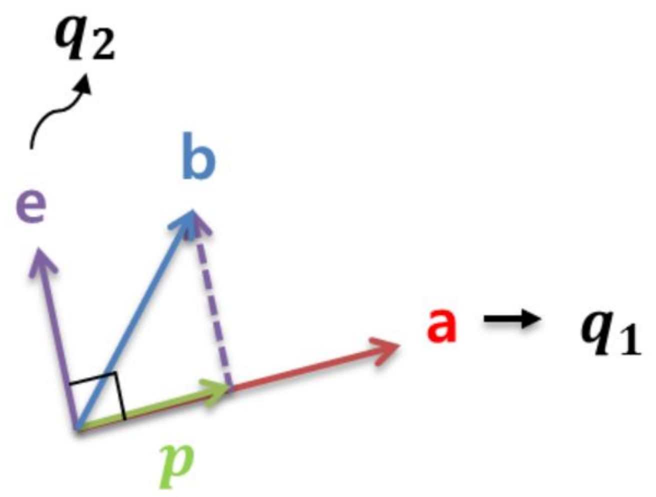
If Orthonormal?

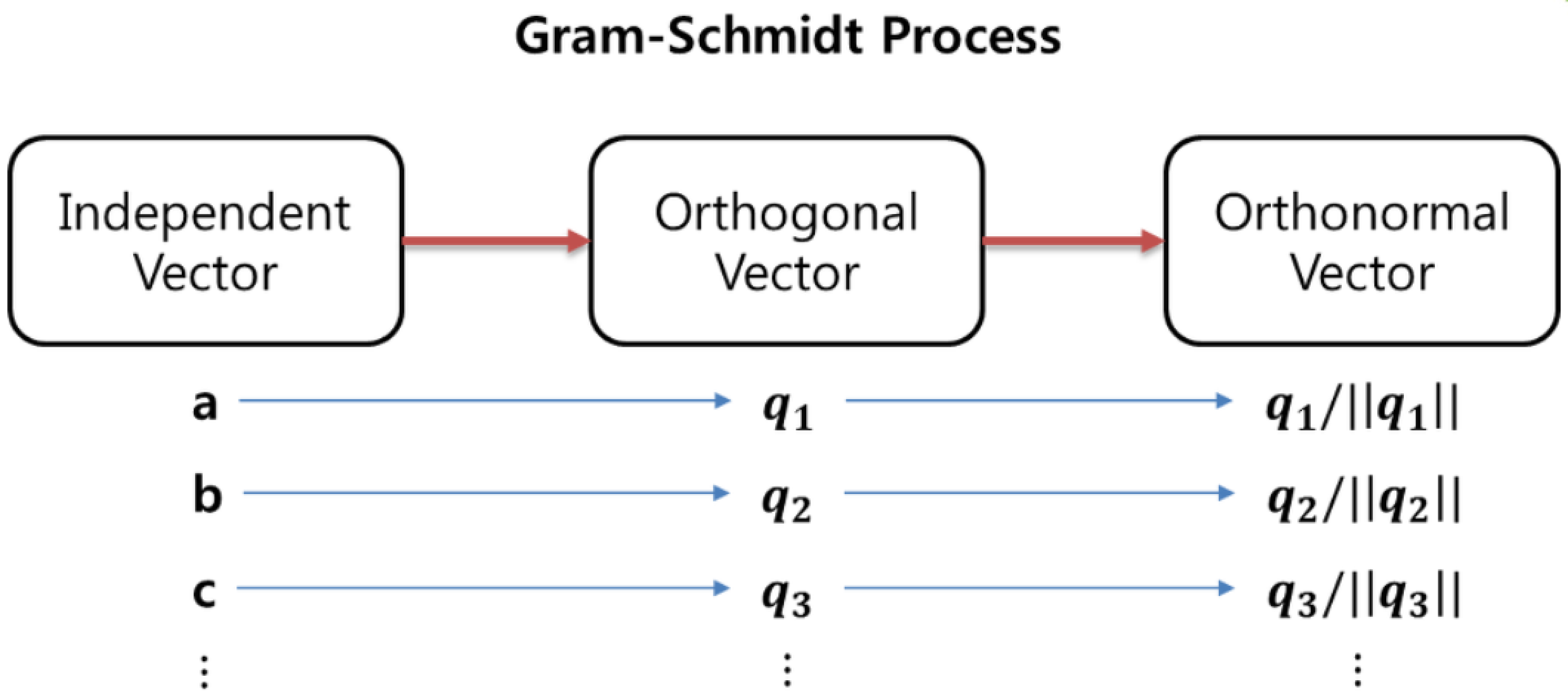
$$P_U = UU^\top, \quad \mathrm{proj}_U(x) = \sum_{i=1}^{k} \langle x, u_i \rangle u_i.$$

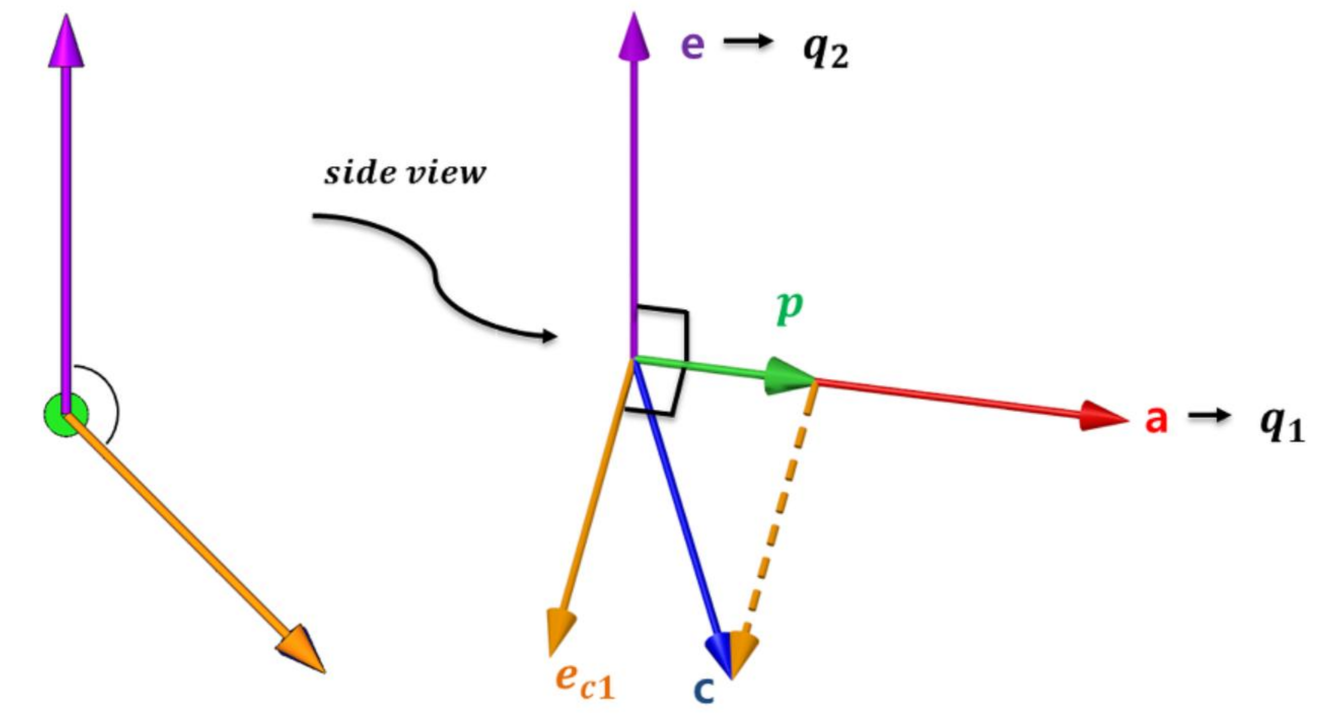
Gram-Schmidt Process

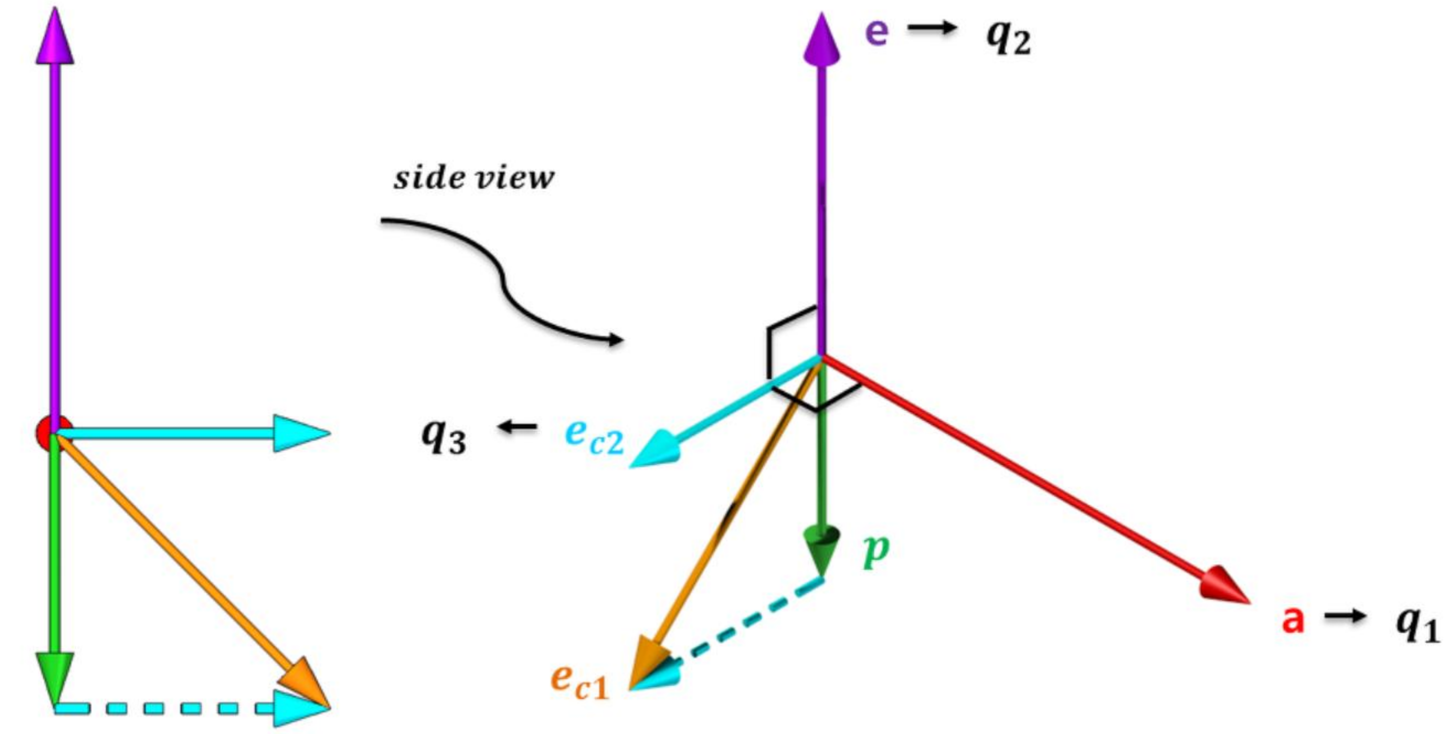
Mathematics (2)

▪ Make Bases Orthonormal









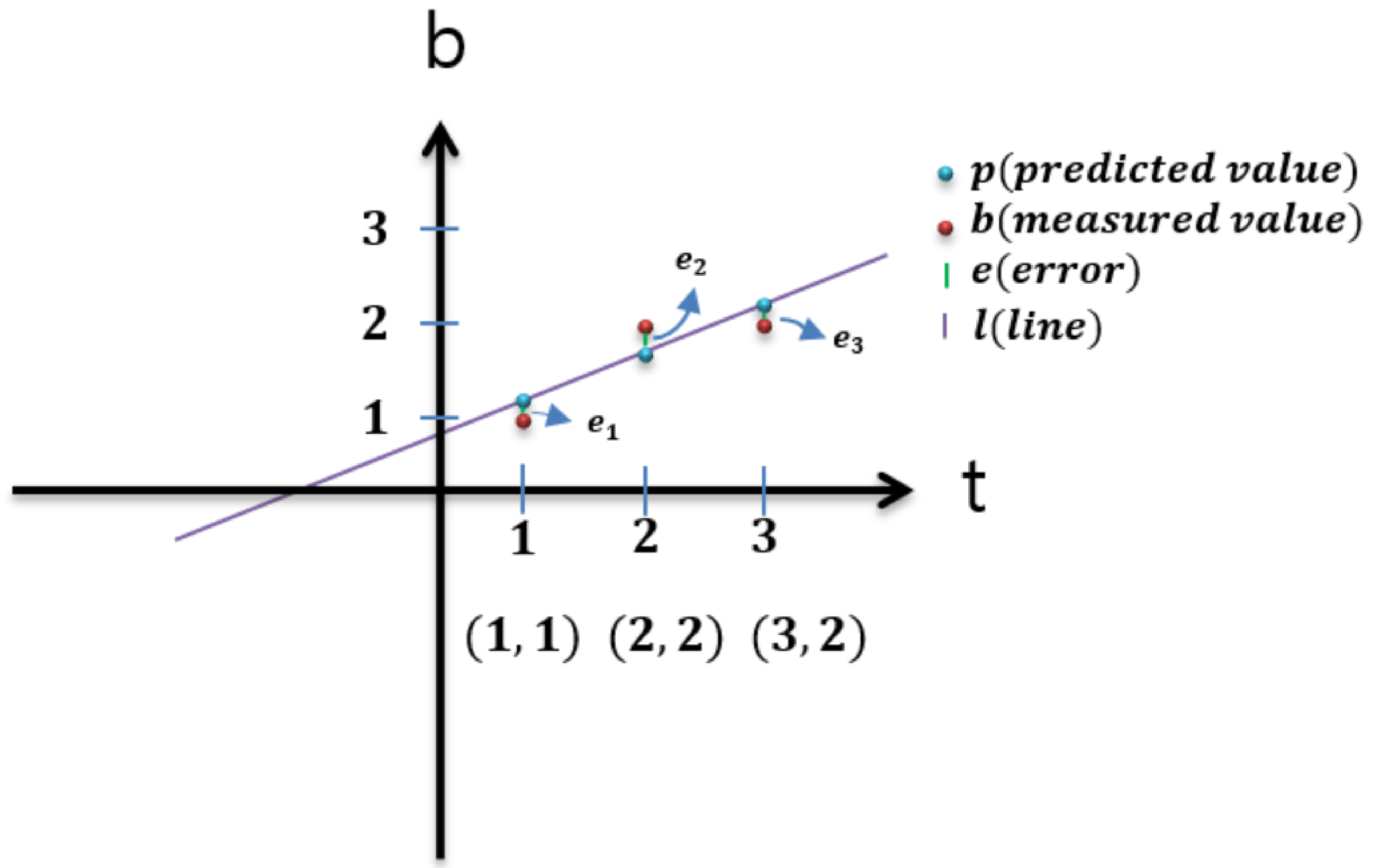
$a = (1, 1, 0), \quad b = (1, 0, 1), \quad c = (0, 1, 1).$

$e_1 = \frac{1}{\sqrt{2}}(1, 1, 0), \quad e_2 = \left(\frac{1}{\sqrt{6}}, -\frac{1}{\sqrt{6}}, \frac{2}{\sqrt{6}}\right), \quad e_3 = \left(-\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}\right).$

Least Squares

Mathematics (2)

- **Regression = Projection!**
  - For response y,
    Feature Matrix X,
    Parameter Set w



OLS Problem : $\min_{w} \|y - Xw\|^2.$ $X^\top X\hat{w} = X^\top y.$

Partial Derfferentiation

Mathematics (2)

▪ Derivative..

$$\frac{\mathrm{d}f}{\mathrm{d}x} := \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$$

▪ But, How to Differentiate Multivariate Function?

- E.g. f(x,y,z,w)
  - Partial Derivate!

  \+ Product, Sum, Chain Rule

$$\frac{\partial f}{\partial x_i}(x_1, \ldots, x_n) = \lim_{h \to 0} \frac{f(x_1, \ldots, x_i + h, \ldots, x_n) - f(x_1, \ldots, x_n)}{h}$$

$$f(x, y) = x^2 + xy + y^2$$

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

Gradient

Mathematics (2)

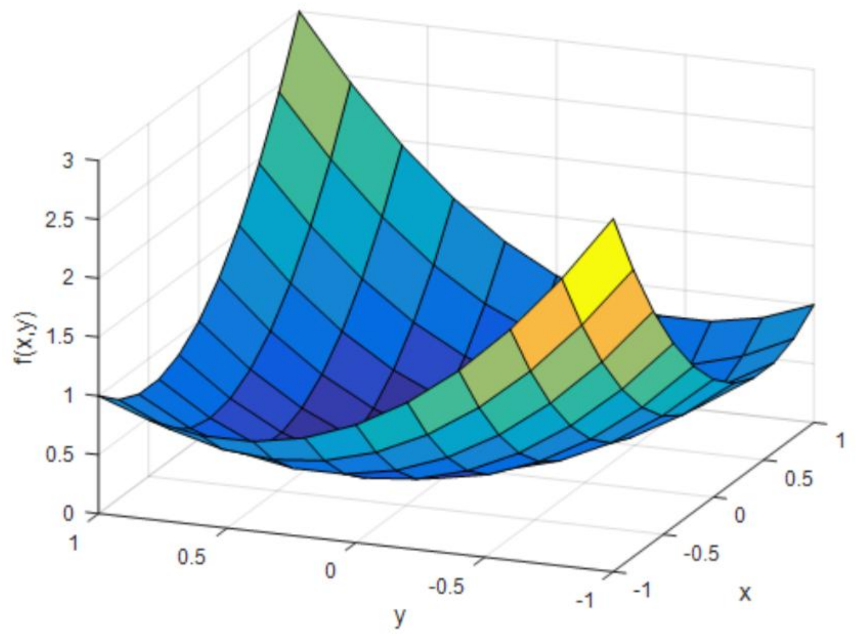
▪ Definition

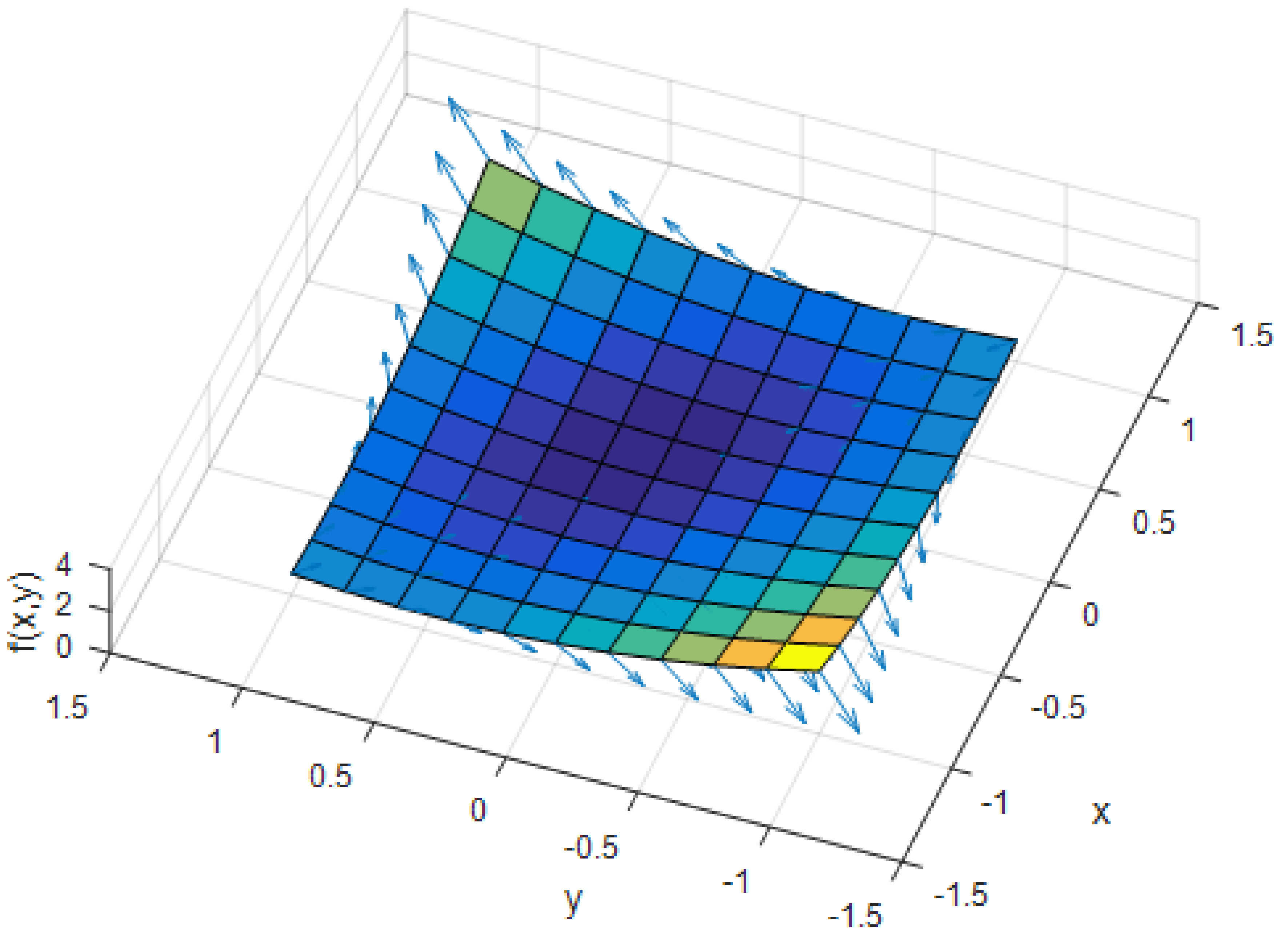
$$\nabla_{\boldsymbol{x}} f = \mathrm{grad} f = \frac{\mathrm{d}f}{\mathrm{d}\boldsymbol{x}} = \begin{bmatrix} \frac{\partial f(\boldsymbol{x})}{\partial x_1} & \frac{\partial f(\boldsymbol{x})}{\partial x_2} & \cdots & \frac{\partial f(\boldsymbol{x})}{\partial x_n} \end{bmatrix} \in \mathbb{R}^{1\times n}$$

▪ Geometrically…

$$f(x,y) = x^2 + xy + y^2$$

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





$$D_u f(x) = \nabla f(x) \cdot u$$

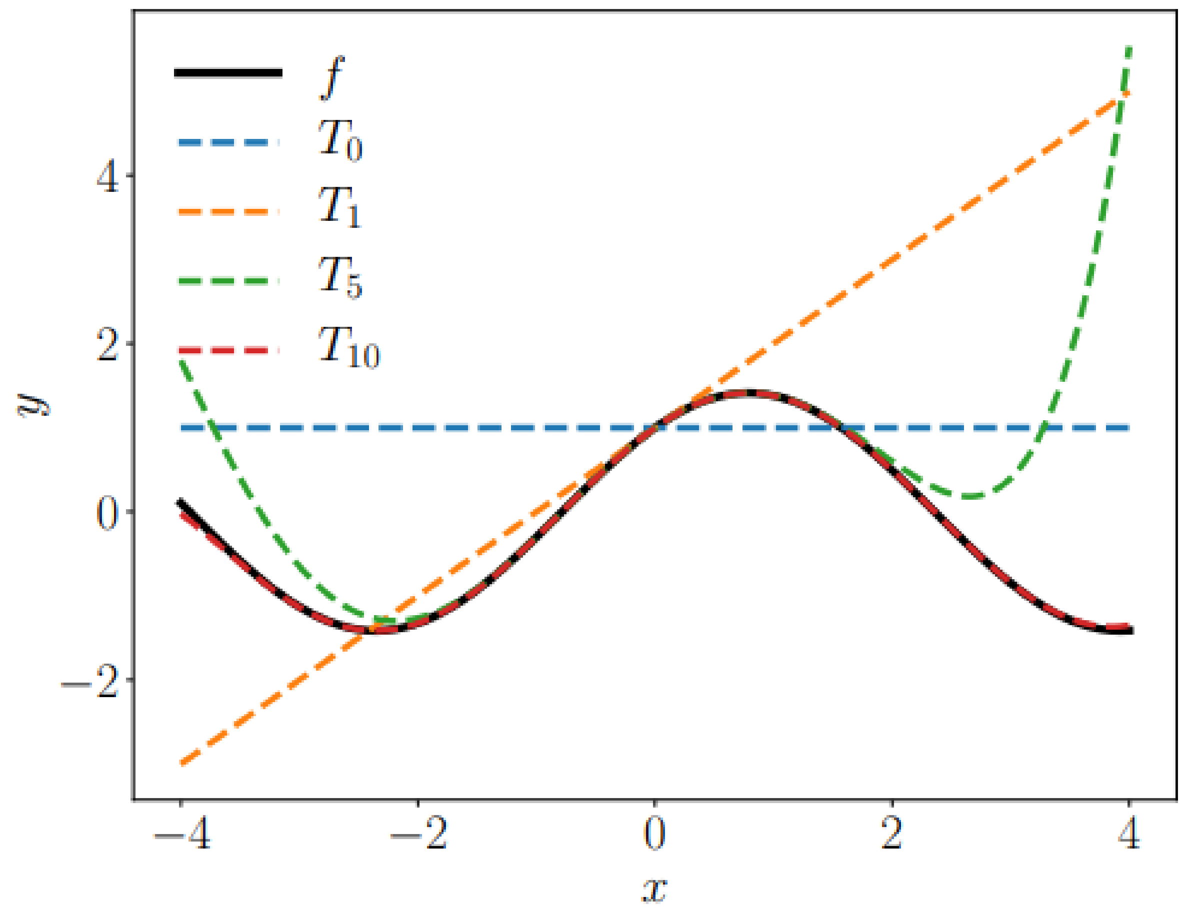
Taylor Expansion

Mathematics (2)

▪ Taylor Polynominal

$$T_n(x) := \sum_{k=0}^{n} \frac{f^{(k)}(x_0)}{k!}(x - x_0)^k \,,$$

$$f(x) \approx f(a) + f'(a)(x-a) + \frac{1}{2}f''(a)(x-a)^2 + \cdots$$



$$f(x) = e^x, \quad a = 0$$

$$f(x) \approx 1 + x + \tfrac{x^2}{2} + \tfrac{x^3}{6} + \cdots$$

$f : \mathbb{R}^n \to \mathbb{R}$

▪ For Multivariate Function

- Gradient Descent (1st)

$$f(x) \approx f(a) + \nabla f(a)^T (x - a)$$

$$f(x, y) = x^2 + y^2, \quad a = (1, 1)$$

$$\nabla f = (2x, 2y) \quad \Rightarrow \quad \nabla f(1, 1) = (2, 2)$$

$$f(x, y) \approx f(1, 1) + (2, 2) \cdot ((x, y) - (1, 1))$$

$$= 2 + 2(x - 1) + 2(y - 1)$$

- Newton's Method (2nd)

$$f(x) \approx f(a) + \nabla f(a)^T (x - a) + \frac{1}{2}(x - a)^T H_f(a)(x - a)$$

$$f(x, y) \approx 0 + 0 + \frac{1}{2}(x, y) \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = x^2 + y^2$$

QnA

References

- Mathematics for Machine Learning Book
- Tistory, Learn Again!
- Linear Algebra Book

감사합니다