

Matrix Calculus

Gradient

Def. Directional Derivative

The directional derivative of function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ in the direction u is

$$\nabla_u f(x) = \lim_{h \rightarrow 0} \frac{f(x + hu) - f(x)}{h}$$

- When u is the i -th standard unit vector e_i , then $\nabla_u f(x) = f'_i(x) = \frac{\partial f(x)}{\partial x_i}$.
- For any n -dimensional vector u , the directional derivative of f in the direction of u can be represented as $\nabla_u f(x) = \sum_{i=1}^n f'_i(x) \cdot u_i$.

$$\text{let } g(h) = f(x + hu)$$

$$\begin{aligned} \nabla_u f(x) &= g'(0) = \lim_{h \rightarrow 0} \frac{f(x + hu) - g(0)}{h} \\ \text{- Proof.} \Rightarrow \therefore g'(h) &= \sum_{i=1}^n f'_i(x) \frac{d}{dh}(x_i + hu_i) = \sum_{i=1}^n f'_i(x) u_i \\ \text{let } h &= 0 \therefore \nabla_u f(x) = \sum_{i=1}^n f'_i(x) u_i \end{aligned}$$

Def. Gradient

The gradient of f is a vector function $\nabla f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ defined by

$$\begin{aligned} \nabla f(x) &= \sum_{i=1}^n \frac{\partial f}{\partial x_i} e_i \\ \Rightarrow \nabla f(x) &= \left[\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \dots, \frac{\partial f}{\partial x_n} \right] \end{aligned}$$

- $\nabla_u f(x) = \nabla f(x) \cdot u = \|\nabla f(x)\| \cos a$ Where u is a unit vector.
- When $u = \nabla f(x)$ such that $a = 0$, we have the maximum directional derivative of f .

Matrix Derivatives

- The derivative of $f : \mathbb{R}^{m \times n} \rightarrow \mathbb{R}$ with respect to A is defined as:

$$\nabla f(A) = \begin{bmatrix} \frac{\partial f}{\partial A_{11}} & \dots & \frac{\partial f}{\partial A_{1n}} \\ \vdots & \ddots & \vdots \\ \frac{\partial f}{\partial A_{m1}} & \dots & \frac{\partial f}{\partial A_{mn}} \end{bmatrix}$$

trace

$$\text{Def. } \text{tr} A = \sum_{i=1}^n A_{ii}$$

- $\text{tr} ABCD = \text{tr} DABC = \text{tr} CDAB = \text{tr} BCDA$
- $\text{tr} A = \text{tr} A^T, \text{tr}(A + B) = \text{tr} A + \text{tr} B, \text{tr}(aA) = a \cdot \text{tr} A$
- $\nabla_A \text{tr} AB = B^T, \nabla_{A^T} f(A) = (\nabla_A f(A))^T$
- $\nabla_A \text{tr} ABA^T C = CAB + C^T AB^T, \nabla_A |A| = |A| (A^{-1})^T$
- Funky trace derivative $\nabla_{A^T} \text{tr} ABA^T C = B^T A^T C^T + BA^T C$

Jacobian Matrix

$$J = \begin{bmatrix} \frac{\partial f}{\partial x_1} & \cdots & \frac{\partial f}{\partial x_n} \end{bmatrix} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix}$$

Hesse Matrix

$$G(x_0) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}_{x_0}$$

- $H(f) = J(\nabla f)$