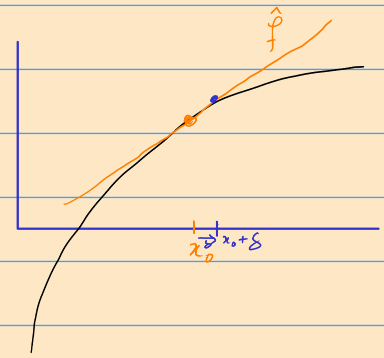
1. What
2. How
   - Basics
   - Product
   - Chain

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

$f(x) = \log x$

$\hat{f}(x_0+\delta) = f(x_0) + \left(\frac{df}{dx}\Big|_{x_0}\right)\delta$



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

$f(x) = x + 2x_2 - 5x_3$, $\quad x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$

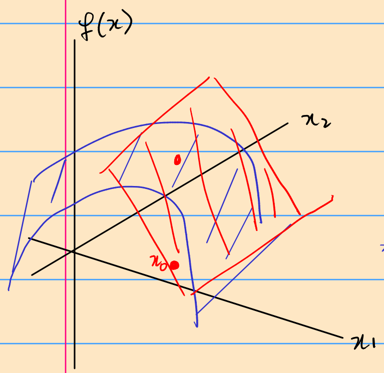
$\partial f/\partial x_2 = 2$, $\partial f/\partial x_1 = 1$

$\frac{\partial f}{\partial x_3} = \frac{\partial}{\partial x_3}(k - 5x_3) = -5$

$\frac{\partial f}{\partial x} = [1 \quad 2 \quad -5]$

$f(x) = w^T x$, $w \in \mathbb{R}^n$

$= \sum_{i=1}^{n} w_i x_i$



$f(x) = \log(w^T x) = \log\left(\sum_{i=1}^{n} w_i x_i\right)$

$f(x) = \|x\|_2^2 = x_1^2 + x_2^2 + \cdots + x_n^2$

$x^T A x$, $A \in \mathbb{R}^{n \times n}$

$\frac{\partial f}{\partial x_1} = 2x_1, \ldots, \frac{\partial f}{\partial x_i} = 2x_i \qquad \frac{\partial}{\partial x}\|x\|_2^2 = 2x^T$

$n = 2$

"derivative"

$\hat{f}(\underset{\in \mathbb{R}^n}{x_0 + \Delta}) = f(x_0) + \boxed{\frac{\partial f}{\partial x}} \begin{bmatrix} \Delta_1 \\ \Delta_2 \\ \vdots \\ \Delta_n \end{bmatrix}$

$[\qquad] \in \mathbb{R}^{1 \times n}$

$\mathbb{R}^n$

$\boxed{1(b)}$ $\nabla_x w^T x$

$\frac{\partial w^T x}{\partial x} = \frac{\partial}{\partial x}\left(\sum w_i x_i\right)$

$\frac{\partial}{\partial x_i} \sum w_i x_i = w_i$

$(w^T)^T = w \qquad \frac{\partial}{\partial x} = [w_1 \ w_2 \ \cdots \ w_n] = w^T$

$= \underset{\mathbb{R}}{f(\underset{\in \mathbb{R}^n}{x_0})} + \sum_{i=1}^{n} \Delta_i \frac{\partial f}{\partial x_i}$

$\nabla_x f = \left(\frac{\partial f}{\partial x}\right)^T$

$w, x \in \mathbb{R}^n$

$\sum w_i x_i \in \mathbb{R}$

$f(x) = w^T x \qquad w^T x \in \mathbb{R}$

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

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

$f(x) = Ax$ 1(a)

$f(x) = I_{n\times n} x \quad : \mathbb{R}^n \to \mathbb{R}^n$

$f(x) = \log x = \begin{bmatrix} \log x_1 \\ \log x_2 \\ \vdots \\ \log x_n \end{bmatrix}$

$\hat{f}(x_0 + \Delta) = f(x_0) + \boxed{\frac{\partial f}{\partial x}} \Delta$

"Jacobian"

"derivative"

$i \to \begin{bmatrix} \frac{\partial f_i}{\partial x_j} \end{bmatrix}$

j

$f = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \quad \frac{\partial f_1}{\partial x_1} = 1 \quad \frac{\partial f_1}{\partial x_3} = 0$

$\begin{bmatrix} \cancel{\frac{\partial f_1}{\partial x_1}}^{1} & \cancel{\frac{\partial f_1}{\partial x_2}}^{0} \\ \cancel{\frac{\partial f_2}{\partial x_1}}^{0} & \cancel{\frac{\partial f_2}{\partial x_2}}^{1} \end{bmatrix} = I$

$\frac{\partial}{\partial x} Ax = A$

$f(x) = \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_m \end{bmatrix}$

1(d) $a^T x x^T b$

$f_j = \sum_{i=1}^{n} A_{ji} x_i \qquad \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ & & & \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$

$\frac{\partial f_j}{\partial x_i} = A_{ji}$

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

$f(A) = \mathrm{trace}(A)$

$f(A) = x^T A w$ const

$f(A) = \mathrm{trace}(XA)$ $\in \mathbb{R}^{n\times m}$

$\mathrm{trace}\left(\begin{bmatrix} 1 & 17 & 0 \\ 14 & 1 & 0 \\ 0 & -7 & 5 \end{bmatrix}\right) = 1 + 1 + 5 = 7$

$\hat{f}(A_0 + \Delta) = f(A_0) + \mathrm{trace}\left(\boxed{\frac{\partial f}{\partial A}} \Delta\right)$ $\in \mathbb{R}^{n\times m}$

$f(A) = \sum_{i=1}^{n} A_{ii}$

$\langle X, A \rangle_F = \mathrm{trace}(X^T A)$

$X, A \in \mathbb{R}^{m\times n}$

$\frac{\partial f}{\partial A_{ij}} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$

$\nabla_A f = \left(\frac{\partial f}{\partial A}\right)^T \in \mathbb{R}^{m\times n}$

$= I_{n\times n}$

$\frac{\partial f}{\partial A} = X$

$f(A) = \sum_{ij} w_i A_{ij} x_j$

$\frac{\partial f}{\partial A_{ij}} = w_i x_j \qquad \frac{\partial f}{\partial A} = w x^T$

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

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

$\nabla_x^2 f = \frac{\partial}{\partial x}(\nabla_x f) \in \mathbb{R}^{n\times n}$