Optimization - Review of Calculus

↳ Derivative of a function

$$f'(x) = \lim_{y \to 0} \frac{f(x+y) - f(x)}{y}$$

$$\lim_{y \to 0} \left[\frac{f(x+y) - f(x) - y f'(x)}{y} \right] = 0$$

$x, y \in \mathbb{R}$

Go to $x, y \in \mathbb{R}^n$

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

$$\nabla f = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \partial f / \partial x_n \end{bmatrix}_{n \times 1}$$

Col vector.

↳ Define ∇f : $\lim_{\|y\| \to 0} \left[\frac{f(x+y) - f(x) - \nabla f^T y}{\|y\|} \right] = 0$

If ∇f exists s.t. ↑ holds the ∇f is called the Frechet derivative of f.

$\nabla f \longrightarrow \nabla_z f$

↳ If ∇f exists $\forall x \in$ dom f. $\rightarrow$ fn differentiable.

→ If ∇f is a continuous fn $\rightarrow$ f is continuosly differentiable.

$\hookrightarrow \quad f: \mathbb{R}^n \longrightarrow \mathbb{R}^{m}$ $\qquad\qquad f: \mathbb{R}^n \longrightarrow \mathbb{R}$

$\nabla f \rightarrow$ Jacobian: $\begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & & & \\ \frac{\partial f_m}{\partial x_1} & \cdots & & \frac{\partial f_m}{\partial x_n} \end{bmatrix}_{m \times n}$

$J \in \mathbb{R}^{m \times n}$ $\qquad J_{ij} = \frac{\partial f_i}{\partial x_j}$

$\hookrightarrow$ Chain rule $\qquad x(t)$ & $y(x)$ $\qquad \frac{dy}{dt} = \frac{d}{dt} y(x(t))$

Scalar case. $\qquad = \frac{dy}{dx} \cdot \frac{dx}{dt}$.

$\hookrightarrow \quad x \in \mathbb{R}^n \qquad h(t) = f(x(t))$

$$\nabla h(t) = \frac{\partial f}{\partial x_1}\frac{\partial x_1}{\partial t} + \frac{\partial f}{\partial x_2}\frac{\partial x_2}{\partial t} + \cdots + \frac{\partial f}{\partial x_n}\frac{\partial x_n}{\partial t}$$

$$= (\nabla f)^T \Delta x$$

e.g. $f : \mathbb{R}^3 \to \mathbb{R}$ $\quad f(u,v,w) = uv + vw - uw$

$g : \mathbb{R}^2 \to \mathbb{R}^3$ $\quad g(x,y) = (x+y,\ x+y^2,\ x^2+y)$

$\mathbb{R}^2 \to \mathbb{R}$ $\quad \nabla (f \circ g)(x,y)$? $\quad f(g(x,y))$

u, v, w

$$= f(u(x,y),\ v(x,y),\ w(x,y))$$

$$\nabla (f \circ g) = \begin{bmatrix} \partial/\partial x \\ \partial/\partial y \end{bmatrix}$$

$$\begin{bmatrix} \dfrac{\partial f\circ g}{\partial x} = \dfrac{\partial f}{\partial u}\dfrac{\partial u}{\partial x} + \dfrac{\partial f}{\partial v}\dfrac{\partial v}{\partial x} + \dfrac{\partial f}{\partial w}\dfrac{\partial w}{\partial x} \\ \dfrac{\partial f\circ g}{\partial y} = \dfrac{\partial f}{\partial u}\dfrac{\partial u}{\partial y} + \cdots \quad \cdots \end{bmatrix}$$

$$\frac{\partial f}{\partial u} = v - w$$

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

$$\frac{\partial u}{\partial x} = 1$$

Quadratic form $\phi(x) : \mathbb{R}^n \to \mathbb{R}$

$$\phi(x) = x^T A x + b^T x + c$$

Claim is $\nabla \phi(x) = (A^T + A)x + b$

$$\begin{bmatrix} \partial/\partial x_1 \\ \partial/\partial x_2 \\ \vdots \\ \partial/\partial x_n \end{bmatrix}$$

Proof:

$$\nabla(b^T x) \quad \nabla\left(\sum b_j x_j\right) \quad \begin{bmatrix} \uparrow \\ \frac{\partial}{\partial x_i}(\quad) \\ \downarrow \end{bmatrix} = \begin{bmatrix} b_1 \\ \vdots \\ b_i \\ \vdots \\ b_n \end{bmatrix} = b$$

$$\frac{d}{dx} f \cdot g = f'g + fg' \quad \text{prod rule.}$$

Now $x^T A x$

Aside: prod rule

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

$$f(x) = g^T(x)\,h(x) \quad , g, h \in \mathbb{R}^{n\times 1} \qquad g : \mathbb{R}^n \to \mathbb{R}^n$$

$$g = \begin{bmatrix} g_1(x) \\ g_2(x) \\ \vdots \\ g_n(x) \end{bmatrix}$$

$$\nabla g = \begin{bmatrix} \overline{} \\ \leftarrow \nabla g_i^T \rightarrow \\ - - - - \end{bmatrix}$$

$$\nabla f \to \frac{\partial f}{\partial x_i} = \frac{\partial}{\partial x_i}\left[\sum_j g_j(x)\,h_j(x)\right]$$

$$= \sum_j \left[\underbrace{\frac{\partial}{\partial x_i} g_j(x)}_{i^{th}\text{ col of } \nabla g} h_j(x) + \underbrace{g_j(x) \frac{\partial}{\partial x_i} h_j(x)}_{i^{th}\text{ col of } \nabla h} \right]$$

$$\frac{\partial f}{\partial x_i} = h^T (\nabla g)_{i^{th}\text{ col}} + g^T (\nabla h)_{i^{th}\text{ col}}$$

$$\left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \dots, \frac{\partial f}{\partial x_n}\right) = h^T \nabla g + g^T \nabla h = \nabla f^T$$

$$(AB)^T = B^T A^T$$

$$\boxed{\nabla f = \nabla g^T h + \nabla h^T g}$$

prod rule
Where $f = g^T h$

$$\nabla(\underbrace{x^T}_{g} \underbrace{Ax}_{h}) = \underbrace{(\nabla x)}_{I}^T Ax + (\nabla Ax)^T x = Ax + A^T x$$

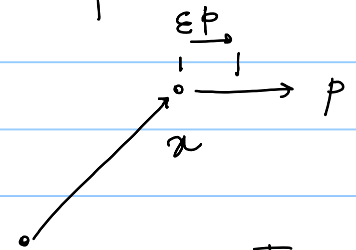
$$(Ax)_i = \sum_j A_{ij} x_j$$

$$\nabla(Ax) = A$$

$$\Rightarrow \nabla \phi(x) = (A + A^T)x + b$$

- Directional derivative: $\rightarrow$ direction '$\hat{p}$'

$$D(f(x), p) = \lim_{\varepsilon \to 0} \frac{f(x + \varepsilon p) - f(x)}{\varepsilon}$$



When f is Contns differentiable: $D(f(x), p) = \nabla f^T p$

- Defn: Hessian matrix $\quad f: \mathbb{R}^n \to \mathbb{R}$

$$\nabla^2 f(x) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \ddots \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ & & \ddots \end{bmatrix}_{n \times n}$$

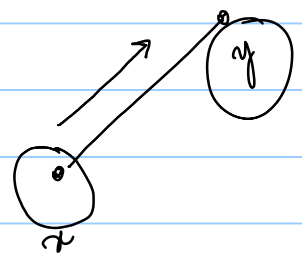
Mean value theorem : $f : \mathbb{R} \to \mathbb{R}$ and $x, y \in \mathbb{R}$ s.t. $x < y$

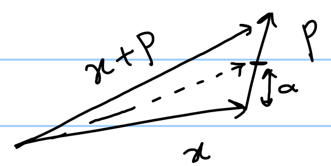
MVT

$$f(y) = f(x) + f'(z)(y - x) \qquad z \in (x, y)$$

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

$$f(x+p) = f(x) + \overbrace{\nabla f(x+\alpha p)^T p}$$





$\alpha \in (0,1)$

$x + \alpha p \to$ $\alpha = 0 \to x$

$\alpha = 1 \to x + p$