REAL ANALYSIS

Part - 3

→ Derivatives

Consider a real valued function $f : X \to \mathbb{R}$ where $X \subseteq \mathbb{R}^n$. For a fixed $x \in X$ ($x \in \mathbb{R}^n$) consider the following expression:

$$\boxed{\lim_{\theta \to 0} \frac{f(x+\theta e_i) - f(x)}{\theta}}, \quad e_i = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \leftarrow i\text{th component is } 1$$

If the limit of the above expression exists, it is called "i^{th} partial derivative" of f at the point x → denoted as $\dfrac{\partial f(x)}{\partial x_i}$.

Ex :- $f(x) = x_1 + x_2^2 + 3 \to f : X \to \mathbb{R}$ when $X \subseteq \mathbb{R}^2$

- Assuming that all $\dfrac{\partial f(x)}{\partial x_i}$ for $i = 1, 2 \cdots n$, the gradient vector of f at x is :

$$\boxed{\nabla f(x) = \begin{bmatrix} \frac{\partial f(x)}{\partial x_1} \\ \frac{\partial f(x)}{\partial x_2} \\ \vdots \\ \frac{\partial f(x)}{\partial x_n} \end{bmatrix}}$$

- For any given $y \in \mathbb{R}^n$, the directional derivative of f at x in the direction of y is :

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

provided the limit exists.

- A funct^n $f : X \to \mathbb{R}$ with $X \subseteq \mathbb{R}^n$ is differentiable at x if

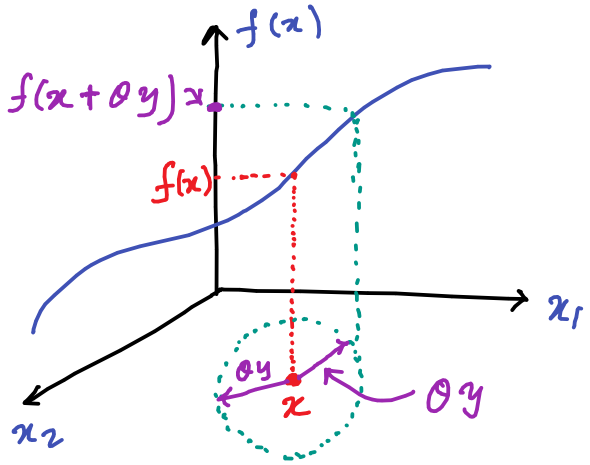
(i) $f'(x, y)$ exists in all directions 'y',

(ii) $f'(x, y)$ is a linear function of y.

→ Called Gateaux Differentiability

→ A function $f : X \to \mathbb{R}$ with $X \subseteq \mathbb{R}^n$ is differentiable at x if and only if $\nabla f(x)$ exists & following holds :

$$\boxed{\nabla f(x)^T y = f'(x, y), \; \forall \, y \in \mathbb{R}^n}$$



θ is a scaling factor of the vector y.

For differentiability $\nabla f(x)$ exists at x

$$f'(x, y) = \nabla f(x)^T y \quad \forall\, y \in \mathbb{R}^n.$$

- A function $f: X \longrightarrow \mathbb{R}$ with $X \subseteq \mathbb{R}^n$ is <u>differentiable</u> if it is differentiable at all points $x \in \mathbb{R}^n$.

- If $f: \mathbb{R}^n \to \mathbb{R}$ is <u>differentiable</u> over an <u>open</u> set $X \subset \mathbb{R}^n$ & its gradient $\nabla f(x)$ is <u>continuous</u> at <u>all $x \in X$</u>, then f is said to be <u>continuously differentiable</u> over X.

- If f is continuously differentiable over $\mathbb{R}^n$, then f is called a <u>smooth</u> function.

Example :

Let $f : \mathbb{R}^3 \to \mathbb{R}$ be $f(x) = \underbrace{\begin{bmatrix} 1 & 3 & 2 \end{bmatrix}}_{v^T} \underbrace{\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}}_{x}$

$= x_1 + 3x_2 + 2x_3$

The Gateaux derivative :

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

$$= \lim_{\theta \to 0} \frac{v^T(x+\theta y) - v^T x}{\theta}$$

$$= \lim_{\theta \to 0} \frac{v^T x + \theta v^T y - v^T x}{\theta}$$

$$= \lim_{\theta \to 0} \frac{\theta v^T y}{\theta} = v^T y$$

which is a linear function of y for all $y \in \mathbb{R}^n$.

Verify that for $f(x) = x^T A x$, when $A \in \mathbb{R}^{n \times n}$ be a given matrix,

$$f'(x, y) = x^T (A + A^T) y$$

→ A real valued function $f : X \to \mathbb{R}$ with $X \subseteq \mathbb{R}^n$ is "Frechet differentiable" at $x \in \mathbb{R}^n$ if there exists a vector $g \in \mathbb{R}^n$ s.t.

$$\boxed{\lim_{y \to 0} \frac{f(x+y) - f(x) - g^T y}{\|y\|} = 0 , \quad \forall x \in X .}$$

- If such a vector g exists, it can be shown that all the partial derivatives $\frac{\partial f(x)}{\partial x_i}$ for $i = 1, 2 \cdots n$ exists. In this case

$$\boxed{g = \nabla f(x)}$$

$$\Downarrow$$

→ A continuously differentiable function $f : X \to \mathbb{R}$ with $X \subseteq \mathbb{R}^n$, satisfies:

$$\boxed{\lim_{y \to 0} \frac{f(x+y) - f(x) - \nabla f(x)^T y}{\|y\|} = 0 , \quad \forall x \in X .}$$

$$\Downarrow$$

- A continuously differentiable function is both Gateaux & Frechet differentiable at all $x \in X$.

Example : let $f(x) = [1 \;\; 3 \;\; 2] \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = v^T x$

$= x_1 + 3x_2 + 2x_3$

$$\lim_{y \to 0} \frac{f(x+y) - f(x) - g^T y}{\|y\|}$$

$$= \lim_{y \to 0} \frac{v^T(x+y) - v^T x - g^T y}{\|y\|}$$

$$= \lim_{y \to 0} \frac{v^T x + v^T y - v^T x - g^T y}{\|y\|}$$

$$= \lim_{y \to 0} \frac{v^T y - g^T y}{\|y\|} = \lim_{y \to 0} \frac{(v-g)^T y}{\|y\|}$$

$$= 0 \quad \text{iff} \quad \boxed{v = g}$$

$\Downarrow$

$\exists \; g$ s.t.

$$\lim_{y \to 0} \frac{f(x+y) - f(x) - g^T y}{\|y\|} = 0$$

for all $x \in X$.

$\Downarrow$

f is Frechet differentiable at x.

Further $\nabla f(x) = v = \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix} = g$

- A vector valued function $f : X \to \mathbb{R}^m$ with $X \subseteq \mathbb{R}^n$ is called differentiable (and continuously differentiable) if each component of $f = \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_m \end{bmatrix}$ i.e. f_i is differentiable (continuously differentiable)

→ The Jacobian matrix J of f is:

$$\boxed{J = \nabla f(x)^T = \left[\nabla f_1(x) \quad \nabla f_2(x) \quad \cdots \quad \nabla f_m(x) \right]^T}$$

$$J = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \frac{\partial f_m}{\partial x_2} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix}_{m \times n}$$

For $f = \begin{bmatrix} x_1^2 \\ x_1 + x_2^2 \\ 2x_2 + 4x_1 \end{bmatrix} \to J = \begin{bmatrix} 2x_1 & 0 \\ 1 & 2x_2 \\ 4 & 2 \end{bmatrix}$

$f : \mathbb{R}^2 \to \mathbb{R}^3$

→ For a function $f: \mathbb{R}^n \to \mathbb{R}$, assume that all of its partial derivatives $\frac{\partial f}{\partial x_i}$, $i = 1, 2 \cdots n$ are smooth functions. Then, the

$$\boxed{i^{th} \text{ partial derivative of } \frac{\partial f}{\partial x_j} \longrightarrow \frac{\partial^2 f(x)}{\partial x_i \, \partial x_j}}$$

for $x \in \mathbb{R}^n$

$$f(x) = x_1 + 3x_2^3 + 2x_3^2 + x_1^2 + x_2 x_3^2$$

$$\frac{\partial f}{\partial x_1} = 1 + 2x_1 \qquad \frac{\partial f}{\partial x_2} = 9x_2^2 + x_3^2 \qquad \frac{\partial f}{\partial x_3} = 4x_3 + 2x_2 x_3$$

$$\frac{\partial f}{\partial x_1 \partial x_1} = \frac{\partial f}{\partial x_1^2} = 2 \qquad \frac{\partial f}{\partial x_1 \partial x_2} = 0 \qquad \frac{\partial f}{\partial x_1 \partial x_3} = 0$$

$$\frac{\partial f}{\partial x_2 \partial x_1} = 0 \qquad \frac{\partial f}{\partial x_2^2} = 18x_2 \qquad \frac{\partial f}{\partial x_2 \partial x_3} = 2x_3$$

$$\frac{\partial f}{\partial x_3 \partial x_1} = 0 \qquad \frac{\partial f}{\partial x_3 \partial x_2} = 2x_3 \qquad \frac{\partial f}{\partial x_3^2} = 4 + 2x_2$$

→ The Hessian matrix of $f: \mathbb{R}^n \to \mathbb{R}$ is:

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

- Note that

$$\boxed{\frac{\partial^2 f(x)}{\partial x_i \, \partial x_j} = \frac{\partial^2 f(x)}{\partial x_j \, \partial x_i}}$$

↳ Hence, Hessian matrix $\nabla^2 f(x)$ is symmetric.

→ To evaluate the Jacobian & Hessian matrices at some point $\bar{x} \in \mathbb{R}^n$, we need to evaluate J & $\underbrace{\nabla^2 f(x) \text{ at } \bar{x}}_{\text{real symmetric matrix}}$.

→ <u>Some Results on Differentiable functions:</u>

- Taylor Series Expansion:

Let a function $f: \mathbb{R} \to \mathbb{R}$. Assume that all of its higher order derivatives $(f^{(n)}(x))$ exist over an open interval I. Let $\bar{x} \in I$. Then, the function $f(x)$ in the neighborhood of $\bar{x}$, can be expressed as:

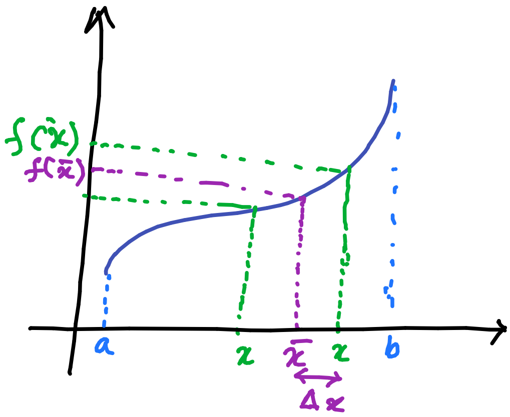
$$\boxed{f(x) = f(\bar{x}) + f'(\bar{x})(x-\bar{x}) + f''(\bar{x})\frac{(x-\bar{x})^2}{2!} + \cdots}$$

$$f'(x) = \lim_{\theta \to 0} \frac{f(x+\theta) - f(x)}{\theta}$$ ← Derivative of $f(x)$ at x.

$f'(\bar{x})$ is the derivative of $f(x)$, evaluated at $\bar{x}$.

Write a point $x \in I$ in the neighborhood of $\bar{x} \in I$ as:

$$x = \bar{x} + \Delta x \quad \Rightarrow \quad x - \bar{x} = \Delta x$$



- Then the first order approximation of $f(x)$ at point x, which is in the neighborhood of $\bar{x}$, i.e. there exists some Δx s.t. $\bar{x} + \Delta x = x$, is

$$\boxed{f(x) = f(x + \Delta x) \approx f(\bar{x}) + f'(\bar{x})\Delta x}$$

- The second order approximation:

$$\boxed{f(x) \approx f(\bar{x}) + f'(\bar{x})\Delta x + \tfrac{1}{2} f''(\bar{x})\Delta x^2}$$

$f(\bar{x} + \Delta x)$

- The k^{th} order approximation is obtained by truncating the power series ⊛ at some $(\Delta x)^k$.
  - Note that the error of the approximation goes to zero $\underbrace{\text{as fast as } (\Delta x)^k}_{\text{with the rate of } (\Delta x)^k}$ as $\Delta x \to 0$.

$\Downarrow$

The larger the k, better is the approximation.

→ Let $f : X \to \mathbb{R}$ with $X \subseteq \mathbb{R}^n$ be a real valued function, which is at least twice continuously differentiable over an open ball $\mathcal{B}(\bar{x}, \epsilon)$. Then the first order approximation of $f(x)$ at $x \in \mathcal{B}(\bar{x}, \epsilon)$ (in the neighborhood of $\bar{x}$) is

$$\boxed{f(x) \approx f(\bar{x}) + \nabla f(\bar{x})^T \Delta x}$$

$f(\bar{x} + \Delta x)$

Gradient of f, evaluated at $\bar{x}$.

- Second order approximation:

$$\boxed{f(x) \approx f(\bar{x}) + \nabla f(\bar{x})^T \Delta x + \frac{1}{2} \nabla x^T \nabla^2 f(\bar{x}) \Delta x}$$

Gradient of f evaluated at $\bar{x}$

Hessian of f evaluated at $\bar{x}$.

→ Mean Value Theorem

Let $f : \mathbb{R}^n \to \mathbb{R}$ be continuously differentiable over an open ball $\mathcal{B}(\bar{x}, \epsilon)$. Then for all $y \in \mathcal{B}(\bar{x}, \epsilon)$, there exists $\gamma \in [0, 1]$ such that

$$\boxed{f(\bar{x} + y) = f(\bar{x}) + \nabla f(\bar{x} + \gamma y)^T y}$$

- ## Second Order Expansions

Let $f : \mathbb{R}^n \to \mathbb{R}$ be twice continuously differentiable over an open ball $\mathcal{B}(\bar{x}, \epsilon)$.

small +ve no.

Then, for all $y \in \mathcal{B}(\bar{x}, \epsilon)$:

(i) there exists $r \in [0, 1]$ s.t.

$$\boxed{f(\bar{x}+y) = f(\bar{x}) + \nabla f(\bar{x})^T y + \frac{1}{2} y^T \nabla^2 f(\bar{x}+ry)\, y}$$

(ii) further, we have :

$$\boxed{f(\bar{x}+y) = f(\bar{x}) + \nabla f(\bar{x})^T y + \frac{1}{2} y^T \nabla^2 f(\bar{x})\, y + o(\|y\|^2)}$$

Gradient of f evaluated at $\bar{x}$

Hessian of f evaluated at $\bar{x}$

order of $\|y\|^2$

[If p is a +ve integer & $h : \mathbb{R}^n \to \mathbb{R}$, then

$$h(x) = o(\|x\|^p) \quad \text{iff}$$

$$\lim_{x_k \to 0} \frac{h(x_k)}{\|x_k\|^p} = 0$$

for all sequences $\{x_k\}$, with $x_k \neq 0 \ \forall k$, that converges to 0.]

The error of the truncated Taylor series is an example for this.

- ## Descent Lemma

Let $f : \mathbb{R}^n \to \mathbb{R}$ be continuously differentiable. Let x & y be two vectors in $\mathbb{R}^n$. Assume that

$$\boxed{\|\nabla f(x + ry) - \nabla f(x)\| \leq L\, r\|y\| \,, \ \forall\, r \in [0,1]}$$

where L is some scalar. Then,

$$\boxed{f(x+y) \leq f(x) + \nabla f(x)^T y + \frac{L}{2}\|y\|^2}$$

Lipschitz continuous condition (the gradient does not change arbitrarily fast).

Will be used in Gradient Descent Algorithm

→ <u>Contraction Mapping</u>

For an iterative algorithm, one can write $x_{k+1} = g(x_k) \quad k = 0, 1, \cdots$

where $g : X \to X$ where $X \subseteq \mathbb{R}^n$.

& has property:

$$\boxed{\|g(x) - g(y)\| \leq r\|x - y\| \quad , \forall\, x, y \in X}$$

Such mapping functions g are called "Contraction mapping".

r: is called "contraction modulus" of g.

- Any point $\bar{x} \in X$ which satisfies:

$$\boxed{g(\bar{x}) = \bar{x}}$$

is called "<u>Fixed Point</u>" of g.

For the iterative algorithm: $x_{k+1} = g(x_k)$, it is required to find such fixed point $\bar{x}$.

- Contraction Mapping Theorem

Let $g: X \to X$ be a contraction mapping with its contraction modulus $r \in [0, 1)$. Assume that $X \subseteq \mathbb{R}^n$ be closed. Then

(i) the mapping g has a unique fixed point $\bar{x} \in X$.

uniqueness & existence of fixed point

(ii) For every initial vector $x_0 \in X$, the sequence $\{x_k\}$, generated by:

$$x_{k+1} = g(x_k),$$

converges to $\bar{x}$. In particular,

$$\boxed{\|x_k - \bar{x}\| \leq r^k \|x_0 - \bar{x}\|, \quad \forall\, k \geq 0}$$

convergence property.