Properties of Continuous Functions

Let f and g be two real functions, continuous at $x = a$. Let α be a real number. Then,

1. $f \pm g$ is continuous at $x = a$.
2. αf is continuous at $x = a$.
3. fg is continuous at $x = a$.
4. $\dfrac{f}{g}$ is continuous at $x = a$, provided $g(a) \neq 0$.

Differentiability of a function at a point

Let f be a real valued function defined on an open interval (a, b) and let $c \in (a, b)$.

Then, f is said to be differentiable at $x = c$ if and only if $\lim\limits_{x \to c} \dfrac{f(x) - f(c)}{x - c}$ exists.

It is denoted by $f'(c)$ or $\left(\dfrac{d}{dx} f(x)\right)_{x=c}$.

Now, $\lim\limits_{x \to c} \dfrac{f(x) - f(c)}{x - c}$ exists $\implies \lim\limits_{x \to c^-} \dfrac{f(x) - f(c)}{x - c} = \lim\limits_{x \to c^+} \dfrac{f(x) - f(c)}{x - c}$

$$\implies \lim_{h \to 0} \frac{f(c - h) - f(c)}{-h} = \lim_{h \to 0} \frac{f(c + h) - f(c)}{h}$$

Geometrically, if f is differentiable at a point P, then there exists a unique tangent at P.

Basic Rules of Differentiation

(1) **Scalar Product:** $\frac{d}{dx}(\alpha f(x)) = \alpha \frac{d}{dx} f(x)$, for any $\alpha \in \mathbb{R}$.

(2) **Sum and Difference Rule:** $\frac{d}{dx}(f(x) \pm g(x)) = \frac{d}{dx} f(x) \pm \frac{d}{dx} g(x)$

(3) **Product Rule:** $\frac{d}{dx}(f(x)g(x)) = f(x)g'(x) + g(x)f'(x)$

(4) **Quotient Rule:** $\frac{d}{dx}\left(\frac{f(x)}{g(x)}\right) = \frac{f'(x)g(x) - f(x)g'(x)}{(g(x))^2}, \; g(x) \neq 0$

provided $f(x), g(x)$ both are differentiable functions.

Partial Derivatives

Consider a function of n variables, $z = f(x_1, x_2, \ldots, x_n)$. Then, the partial derivative of f with respect to an independent variable x_i for any $i = 1, 2, \ldots, n$, denoted as f_{x_i} or $\dfrac{\partial f}{\partial x_i}$ and is given as

$$f_{x_i} = \frac{\partial f}{\partial x_i} = \lim_{\Delta x_i \to 0} \frac{f(x_1, x_2, \ldots, x_i + \Delta x_i, \ldots, x_n) - f(x_1, x_2, \ldots, x_n)}{\Delta x_i}$$

provided the limit exists.

$f : \mathbb{R}^2 \to \mathbb{R}, \qquad f(x,y) = x^4 - x^2y^2 + y^4$

Find $f_x,\ f_y,\ f_{xx},\ f_{yy},\ f_{xy}$ at $(-1,1)$?

$$f = x^4 - x^2y^2 + y^4$$

$$f_x = \frac{\partial f}{\partial x} = 4x^3 - 2xy^2 \qquad (f_x)_{(-1,1)} = -4 + 2 = -2$$

$$f_y = \frac{\partial f}{\partial y} = -2x^2y + 4y^3 \qquad (f_y)_{(-1,1)} = -2 + 4 = 2$$

$$f_{xx} = \frac{\partial^2 f}{\partial x^2} = \frac{\partial}{\partial x}(f_x) = 12x^2 - 2y^2 \qquad f_{xx}$$

The single most important concept from calculus in the context of machine learning is the **gradient.** Gradients generalize derivatives to scalar functions of several variables.

Gradient

The gradient of a function $f : \mathbb{R}^n \to \mathbb{R}$, denoted by ∇f, is given as

$$\nabla f = \begin{bmatrix} \frac{\partial f}{\partial x_1} & \frac{\partial f}{\partial x_2} & \cdots & \frac{\partial f}{\partial x_n} \end{bmatrix}^T$$

$$\text{i.e. } [\nabla f]_i = \frac{\partial f}{\partial x_i}$$

Directional Derivatives

The rate of change of the function $f(x_1, x_2)$ of two variables in the direction of unit vector $\overrightarrow{u} = < a_1, a_2 >$ is called the **directional derivative** of f in the direction of $\overrightarrow{u}$, denoted by $D_{\overrightarrow{u}} f(x_1, x_2)$

$$D_{\overrightarrow{u}} f(x_1, x_2) = \lim_{h \to 0} \frac{f(x_1 + a_1 h, x_2 + a_2 h) - f(x_1, x_2)}{h}$$

provided the limit exists.
The above can be generalized for a function of n variables as

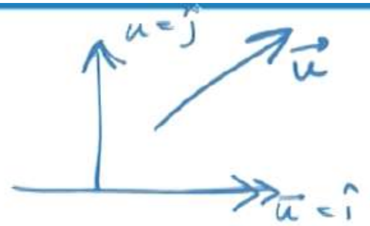
$$D_{\overrightarrow{u}} f(x_1, x_2, \ldots, x_n) = \lim_{h \to 0} \frac{f(x_1 + a_1 h, \ldots, x_n + a_n h) - f(x_1, x_2, \ldots, x_n)}{h}$$

provided the limit exists.

Directional derivatives:

$w = f(x, y)$

$\frac{\partial f}{\partial x} , \frac{\partial f}{\partial y}$



$$\left(\frac{df}{ds}\right)_{P_0, \vec{u}} = \lim_{s \to 0} \frac{f(x_0 + su_1, y_0 + su_2) - f(x_0, y_0)}{s}$$

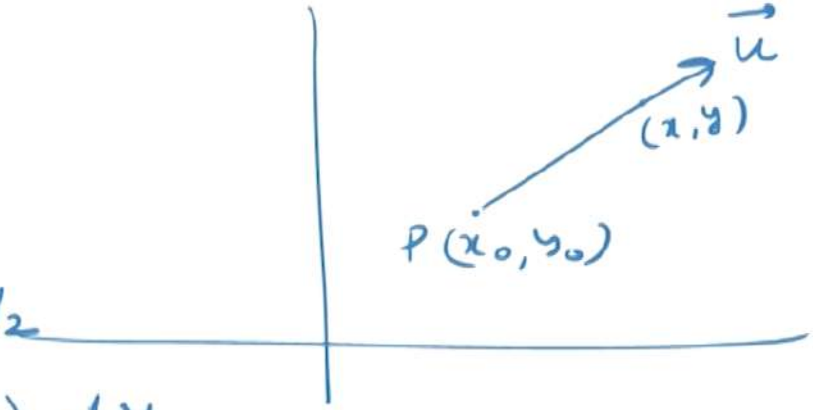
provided the limit exists

$P_0 = (x_0, y_0)$

$\vec{u} = u_1 \hat{i} + u_2 \hat{j}$

$|\vec{u}| = 1$

$$\frac{x - x_0}{u_1} = \frac{y - y_0}{u_2} = s$$



$$\Rightarrow \quad x = x_0 + s u_1, \quad y = y_0 + s u_2$$

$$\vec{u} = u_1 \hat{i} + u_2 \hat{j}$$
$$|\vec{u}| = 1$$

$$\left(\frac{df}{ds}\right)_{P_0, \vec{u}} = \left(\frac{\partial f}{\partial x}\right)_{P_0} \frac{dx}{ds} + \left(\frac{\partial f}{\partial y}\right)_{P_0} \frac{dy}{ds}$$

$$= \left(\frac{\partial f}{\partial x}\right)_{P_0} u_1 + \left(\frac{\partial f}{\partial y}\right)_{P_0} u_2$$

$$= \left(\left(\frac{\partial f}{\partial x}\right)_{P_0} \hat{i} + \left(\frac{\partial f}{\partial y}\right)_{P_0} \hat{j}\right) \cdot \left(u_1 \hat{i} + u_2 \hat{j}\right)$$

$$= (\nabla f)_{P_0} \cdot \vec{u}$$

$$\left(\frac{df}{ds}\right)_{P_0,\vec{u}} = (\nabla f)_{P_0} \cdot \vec{u} \quad , \quad \begin{array}{l} \vec{u} = u_1\hat{i} + u_2\hat{j} \\ |\vec{u}| = 1 \end{array}$$

$$= |(\nabla f)_{P_0}|\,|\vec{u}| \cos\theta$$

$$= |(\nabla f)_{P_0}| \cos\theta$$

$\left(\frac{df}{ds}\right)_{P_0,\vec{u}} \rightarrow$ maximum if $\cos\theta = 1$

$\Rightarrow \theta = 0^\circ$

$\vec{u} \parallel (\nabla f)_{P_0}$

$\left(\frac{df}{ds}\right)_{P_0,\vec{u}} \rightarrow$ minimum if $\cos\theta = -1$

$\theta = \pi$

$\Rightarrow \vec{u} \parallel -(\nabla f)_{i}$

$f : \mathbb{R}^3 \to \mathbb{R}$; $f = x^2 y + y^2 z + z^2 x$ $\quad P_0 \ (1, 0, 1)$

$$(\nabla f)_{P_0} = \begin{pmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} & \frac{\partial f}{\partial z} \end{pmatrix}^T_{P_0} = \begin{pmatrix} 2xy + z^2 \\ x^2 + 2yz \\ y^2 + 2zx \end{pmatrix}_{(1,0,1)}$$

$$= \begin{pmatrix} 1 & 1 & 2 \end{pmatrix}^T$$

$\longrightarrow$ direction where f increases most rapidly.

$$\longrightarrow \frac{\hat{i} + \hat{j} + 2\hat{k}}{\sqrt{1+1+4}} = \frac{\hat{i} + \hat{j} + 2\hat{k}}{\sqrt{6}}$$

The direction where f decreases most rapidly is

$$-(\nabla f)_{P_0} = \frac{-\hat{i} - \hat{j} - 2\hat{k}}{\sqrt{6}}.$$

It is alternatively also expressed as

$$D_{\overrightarrow{u}} f(x_1, x_2, \ldots, x_n) = \sum_{i=1}^{n} f_{x_i}(x_1, x_2, \ldots, x_n) a_i = \nabla f \cdot \overrightarrow{u}$$

Important Property of Gradient:

$\nabla f(\mathbf{x})$ points in the direction of **steepest ascent** from $\mathbf{x}$. Similarly, $-\nabla f(\mathbf{x})$ points in the direction of **steepest descent** from $\mathbf{x}$.

The Jacobian

The Jacobian of $f : \mathbb{R}^n \to \mathbb{R}^m$ is a matrix of first-order partial derivatives, given as

$$\mathbf{J_f} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\ . & . & \cdots & . \\ . & . & \cdots & . \\ \frac{\partial f_m}{\partial x_1} & \frac{\partial f_m}{\partial x_2} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix} \quad \text{i.e. } [\mathbf{J_f}]_{i,j} = \frac{\partial f_i}{\partial x_j}$$

Note: For $m = 1$, we get $\mathbf{J_f}^T = \nabla f$

$$f : \mathbb{R}^2 \longrightarrow \mathbb{R}^2, \quad f(r,\theta) = (r\cos\theta,\ r\sin\theta)$$

$$= (f_1,\ f_2)$$

$$J_f = \begin{pmatrix} \frac{\partial f_1}{\partial r} & \frac{\partial f_1}{\partial \theta} \\ \frac{\partial f_2}{\partial r} & \frac{\partial f_2}{\partial \theta} \end{pmatrix} = \begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix}_{2\times 2}$$

$$f : \mathbb{R}^3 \to \mathbb{R}^2, \qquad f(x,y,z) = \Big(\underbrace{x^2+y^2}_{f_1}, \underbrace{y^2+z^2}_{f_2} \Big)$$

$$J_f = \begin{pmatrix} \dfrac{\partial f_1}{\partial x} & \dfrac{\partial f_1}{\partial y} & \dfrac{\partial f_1}{\partial z} \\[2ex] \dfrac{\partial f_2}{\partial x} & \dfrac{\partial f_2}{\partial y} & \dfrac{\partial f_2}{\partial z} \end{pmatrix}$$

$$= \begin{pmatrix} 2x & 2y & 0 \\ 0 & 2y & 2z \end{pmatrix}_{2\times 3}$$

The Hessian

The Hessian matrix of $f : \mathbb{R}^n \to \mathbb{R}$ is a matrix of second-order partial derivatives, given as:

$$H = \nabla^2 f = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$$

$$\text{i.e. } [\nabla^2 f]_{i,j} = \frac{\partial^2 f}{\partial x_i \partial x_j}$$

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

$$H_f = \nabla^2 f = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & & & \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{pmatrix}_{n \times n}$$

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i} \quad \forall\ i \,\&\, j, \; i \neq j$$

Remark:

If all the partial derivatives, f_x, f_y, f_{xy}, f_{yx} all exist and are all continuous, then by Euler's theorem, the order of differentiation is interchangeable, i.e.,

$$f_{xy} = f_{yx} \; \forall x, y$$

In such a case, the Hessian matrix becomes a symmetric matrix.

$H_f \longrightarrow$ symmetric real matrix.

$\hookrightarrow$ all the eigen-values of H_f are

rea

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

$$H_f = \begin{pmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} & \frac{\partial^2 f}{\partial x \partial z} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2} & \frac{\partial^2 f}{\partial y \partial z} \\ \frac{\partial^2 f}{\partial z \partial x} & \frac{\partial^2 f}{\partial z \partial y} & \frac{\partial^2 f}{\partial z^2} \end{pmatrix} = \left(\begin{matrix} 6x & 3z & 3y+2z \\ 3z & 2 & 3x \\ 3y+2z & 3 & \end{matrix} \right.$$

$$f_x = 3x^2 + 3yz + z^2$$

$$f_y = 3xz + 2y$$

$$f_z = 3xy + 2zx$$

Vector and matrix gradients

$$\nabla_{\mathbf{x}}(a^T\mathbf{x}) = a$$

$$\nabla_{\mathbf{x}}(\mathbf{x}^\mathsf{T}\mathbf{A}\mathbf{x}) = (\mathbf{A} + \mathbf{A}^\mathsf{T})\mathbf{x}$$

where **A** is a square matrix.

$$a = (a_1 \quad a_2 \quad \cdots \quad a_n)^T, \qquad x = (x_1, x_2, \cdots x_n)^T$$

$$a^T x = (a_1 \quad a_2 \quad \cdots \quad a_n) \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = a_1 x_1 + a_2 x_2 + \cdots + a_n x_n$$

$$\nabla (a^T x) = \left(\frac{\partial (a^T x)}{\partial x_1} \quad \frac{\partial (a^T x)}{\partial x_2} \quad \cdots \quad \frac{\partial (a^T x)}{\partial x_n} \right)^T$$

$$= (a_1 \quad a_2 \quad \cdots \quad a_n)^T = a$$

Let $A = \begin{pmatrix} 1 & -1 \\ 0 & 2 \end{pmatrix}_{2\times 2}$ $\quad x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$

$$x^T A x = \begin{pmatrix} x_1 & x_2 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

$$= \begin{pmatrix} x_1 & x_2 \end{pmatrix} \begin{pmatrix} x_1 - x_2 \\ 2x_2 \end{pmatrix} = x_1^2 - x_1 x_2 + 2x_2^2$$

$$\nabla(x^T A x) = \begin{pmatrix} 2x_1 - x_2 & -x_1 + 4x_2 \end{pmatrix}^T = \begin{pmatrix} 2x_1 - x_2 \\ -x_1 + 4x_2 \end{pmatrix}$$

$$= \begin{pmatrix} 2 & -1 \\ -1 & 4 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

$$= \left(\underbrace{\begin{pmatrix} 1 & -1 \\ 0 & 2 \end{pmatrix}}_{A} + \underbrace{\begin{pmatrix} 1 & 0 \\ -1 & 2 \end{pmatrix}}_{A^T} \right) \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

For single-variable function

$$(f \circ g)'(x) = f'(g(x))g'(x)$$

where $\circ$ denotes function composition.

For multi-variate functions

Suppose $f : \mathbb{R}^m \to \mathbb{R}$ and $g : \mathbb{R}^n \to \mathbb{R}^m$. Then, $f \circ g : \mathbb{R}^n \to \mathbb{R}$ and

$$\nabla(f \circ g)(\mathbf{x}) = \mathbf{J}_g(\mathbf{x})^T \nabla f(g(\mathbf{x}))$$

The above can further be generalized for $f : \mathbb{R}^m \to \mathbb{R}^k$ as

$$\mathbf{J}_{f \circ g}(\mathbf{x}) = \mathbf{J}_f(g(\mathbf{x}))\mathbf{J}_g(\mathbf{x})$$

$$w = x^2 + y^2, \quad x = u^2 + v^2, \quad y = uv$$

$$\frac{\partial w}{\partial u} = \left(\frac{\partial w}{\partial x}\right)\left(\frac{\partial x}{\partial u}\right) + \left(\frac{\partial w}{\partial y}\right)\left(\frac{\partial y}{\partial u}\right)$$

$$= (2x)(2u) + (2y)(v) = 4xu + 2yv$$

$$\frac{\partial w}{\partial v} = \left(\frac{\partial w}{\partial x}\right)\left(\frac{\partial x}{\partial v}\right) + \left(\frac{\partial w}{\partial y}\right)\left(\frac{\partial y}{\partial v}\right)$$

$$= (2x)(2v) + (2y)(u) = 4xv + 2yu.$$

Chain Rule when dependent variables are single variable function

Case I: Single Variable Function

Suppose $z = f(y)$ and $y = \phi(x)$, then

$$\frac{dz}{dx} = \frac{dz}{dy}\frac{dy}{dx}$$

Case II: Function of Two Variables

Let $z = f(x, y)$, $x = \phi(t)$ and $y = \xi(t)$, then

$$\frac{dz}{dt} = \frac{\partial z}{\partial x}\frac{dx}{dt} + \frac{\partial z}{\partial y}\frac{dy}{dt}$$

Case III: Function of Three Variables

Let $z = f(x, y, w)$, $x = x(t)$, $y = y(t)$, and $w = w(t)$, then

$$\frac{dz}{dt} = \frac{\partial z}{\partial x}\frac{dx}{dt} + \frac{\partial z}{\partial y}\frac{dy}{dt} + \frac{\partial z}{\partial w}\frac{dw}{dt}$$

Case IV: Function of n variables

Let $z = f(x_1, x_2, \ldots, x_n)$, where $x_i = x_i(t), \forall i$, then

$$\frac{dz}{dt} = \frac{\partial z}{\partial x_1}\frac{dx_1}{dt} + \frac{\partial z}{\partial x_2}\frac{dx_2}{dt} + \cdots + \frac{\partial z}{\partial x_n}\frac{dx_n}{dt} = \sum_{i=1}^{n}\frac{\partial z}{\partial x_i}\frac{dx_i}{dt}$$

$$w = x^2 y^3 z \quad , \quad x = \cos t, \quad y = \sin t, \quad z = t^2 - 1$$

$$\frac{dw}{dt} = ?$$

$$\frac{dw}{dt} = \frac{\partial w}{\partial x} \cdot \frac{dx}{dt} + \frac{\partial w}{\partial y} \cdot \frac{dy}{dt} + \frac{\partial w}{\partial z} \cdot \frac{dz}{dt}$$

$$= (2xy^3 z)(-\sin t) + (x^2 (3y^2) z)(\cos t) + (x^2 y^3)(2t)$$

$$= \_\_\_\_$$

$$z = f(x, y), \quad x = r\cos\theta, \quad y = r\sin\theta,$$

$$z_r^2 + \frac{1}{r^2} z_\theta^2 = ?$$

$$\frac{\partial z}{\partial r} = \frac{\partial f}{\partial x} \cdot \frac{\partial x}{\partial r} + \frac{\partial f}{\partial y} \cdot \frac{\partial y}{\partial r} = f_x \cos\theta + f_y \sin\theta \quad \text{—— (1)}$$

$$\frac{\partial z}{\partial \theta} = \frac{\partial f}{\partial x} \cdot \frac{\partial x}{\partial \theta} + \frac{\partial f}{\partial y} \cdot \frac{\partial y}{\partial \theta} = f_x \cdot (-r\sin\theta) + f_y (r\cos\theta)$$

$$\Rightarrow \quad \frac{1}{r} z_\theta = f_x (-\sin\theta) + f_y (\cos\theta) \quad \text{—— (2)}$$

$$\boxed{z_r^2 + \frac{1}{r^2} z_\theta^2 = f_x^2 + f_y^2}$$

- While dealing with the algorithms related to Machine Learning, "Convexity" plays a vital role.
- Several results have been developed in optimization theory based on the concept of convexity.

Convex Sets

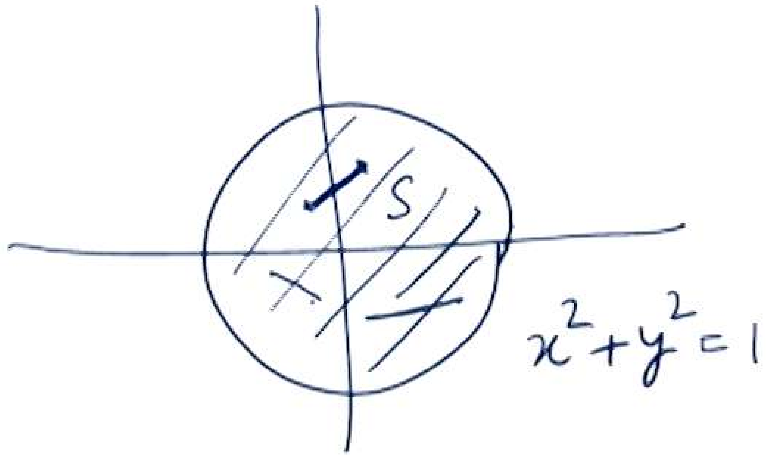
Convex Set

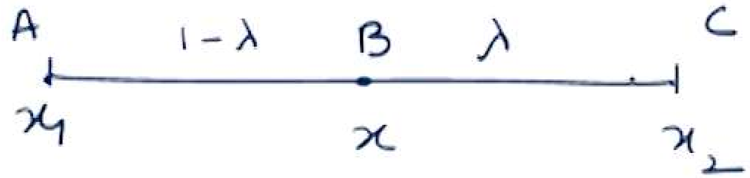
A set $S \subseteq \mathbb{R}^n$ is called convex if for all $x_1, x_2 \in S$ and $\lambda \in [0, 1]$,

$$\lambda x_1 + (1 - \lambda) x_2 \in S.$$

Geometrically, this means that a set is convex if the line segment joining any two points of S is also in S.

$$S = \{ (x, y) \mid x^2 + y^2 \leq 1 \}$$





$$x = \lambda x_1 + (1-\lambda) x_2$$

$$AB : BC = (1-\lambda) : \lambda$$

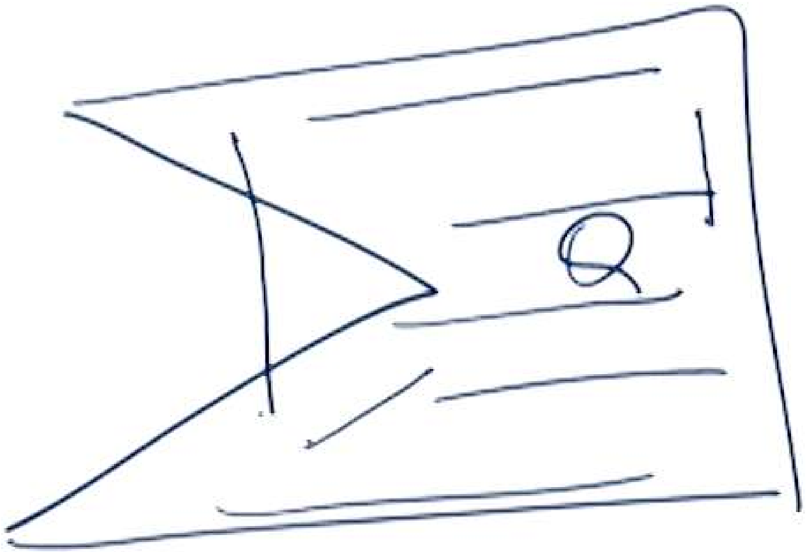
$$\begin{cases} \lambda \geq 0, & 1-\lambda \geq 0 \\ & \Rightarrow \lambda \leq 1 \end{cases}$$

$$\underline{0 \leq \lambda \leq 1}$$

$$\left.\begin{array}{l} \forall\ x_1,\ x_2 \in S \\ 0 \le \lambda \le 1 \end{array}\right] \Rightarrow \lambda x_1 + (1-\lambda) x_2 \in S$$

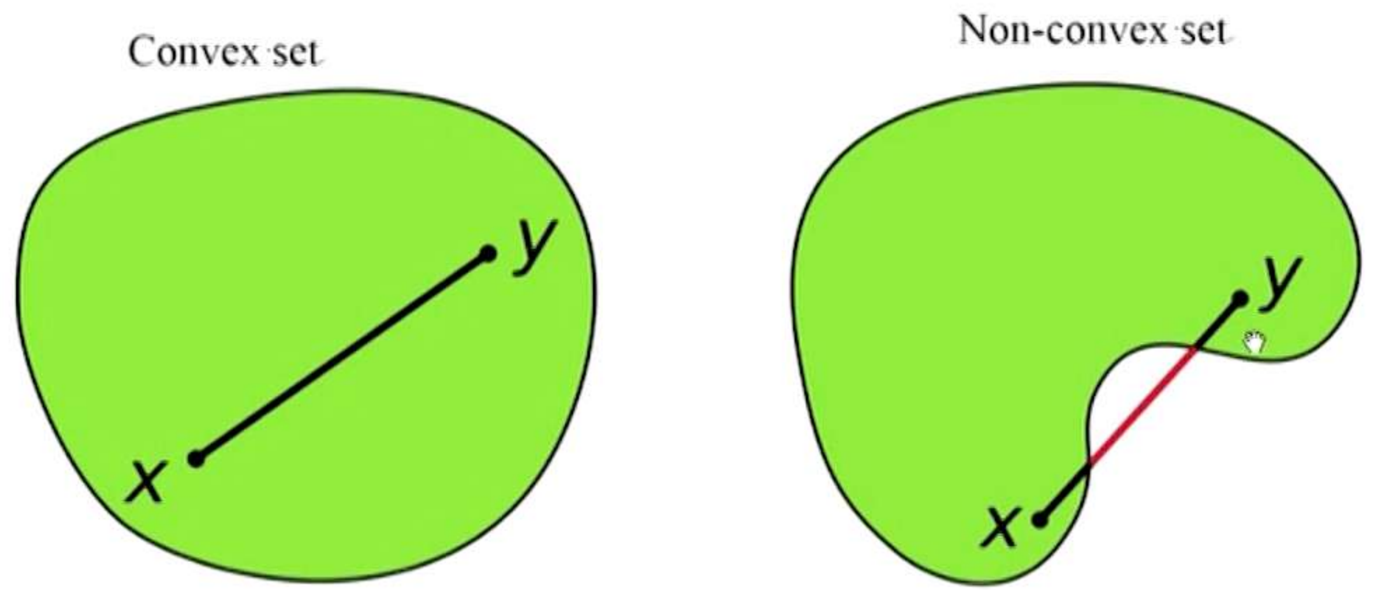
$\Rightarrow$ S is a convex-set

in !



Not convex

Following depicts what convex sets look like:



Examples

- $S = \{(x, y) \in \mathbb{R}^2 \ : \ x^2 + 2y^2 \leq 4\}$ is a convex set.
- $S = \{(x, y) \in \mathbb{R}^2\} \ : \ y^2 \geq 4x\}$ is not a convex set.

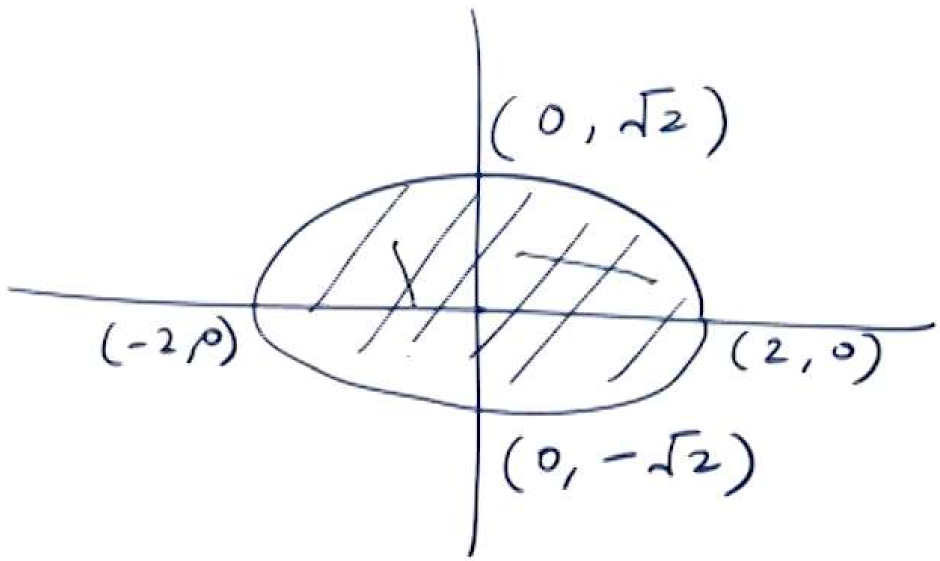
Problem

Check whether the set $S = \{(x, y) \in \mathbb{R}^2 \ : \ x + 3y \geq 6\}$ is convex or not?

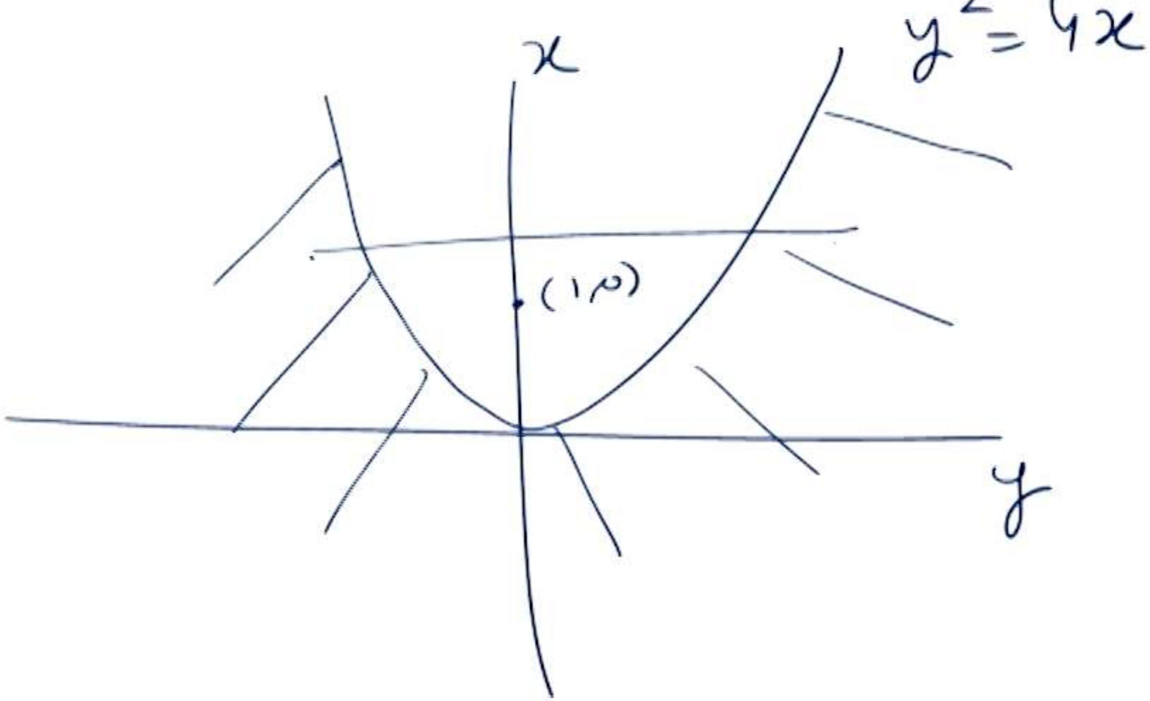
$$S = \{ (x, y) : x^2 + 2y^2 \le 4 \}$$

$$x^2 + 2y^2 = 4$$

$$\Rightarrow \frac{x^2}{4} + \frac{y^2}{2} = 1$$



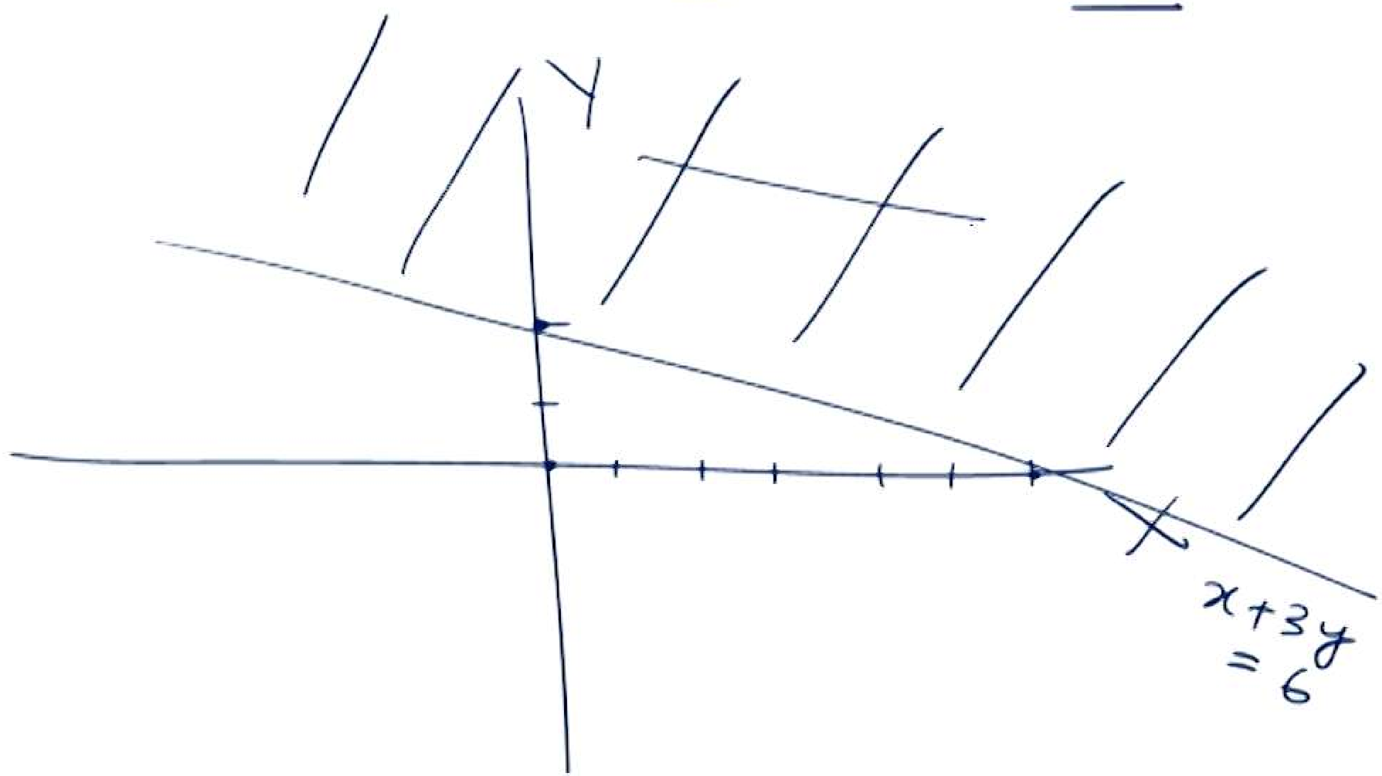
$S = \{ (x, y) : y^2 \geq 4x \}$



Not a convex set

$S = \{ (x,y) : x + 3y \geq 6 \} \longrightarrow$ convex set.

$x + 3y = 6$



$$S = \{ (x, y) : \quad x + 3y \geq 6 \}$$

let (x_1, y_1), $(x_2, y_2) \in S$

$$\Rightarrow x_1 + 3y_1 \geq 6, \quad x_2 + 3y_2 \geq 6,$$

$$0 \leq \lambda \leq 1$$

$$(z_1, z_2) = \lambda (x_1, y_1) + (1-\lambda)(x_2, y_2)$$

$$= (\lambda x_1 + (1-\lambda) x_2, \ \lambda y_1 + (1-\lambda) y_2)$$

$$\Rightarrow \quad z_1 = \lambda x_1 + (1-\lambda) x_2, \quad z_2 = \lambda y_1 + (1-\lambda) y_2$$

$$z_1 + 3z_2 = \underline{\underline{\lambda}} x_1 + (1-\lambda)x_2 + 3(\underline{\underline{\lambda}} y_1 + (1-\lambda)y_2)$$

$$= \lambda(x_1 + 3y_1) + (1-\lambda)(x_2 + 3y_2)$$

$$\geq \lambda \times 6 + (1-\lambda) \times 6 = 6$$

$$\Rightarrow (z_1, z_2) \in S$$

$\Rightarrow$ S is a conv.

Properties of Convex sets

1. The intersection of any arbitrary collection of convex sets is convex.
2. Union of two convex sets need not be convex.
3. The vector sum $C_1 + C_2$ of two convex sets C_1 and C_2 is convex.
4. The set αC is convex for any convex set C and scalar α.

C_i , $i \in I$, I = Index-set $\longrightarrow$ Convex sets in $\mathbb{R}^n$.

$\bigcap_{i \in I} C_i$ is also convex $\longrightarrow$ To-show.

Let $x, y \in \bigcap_{i \in I} C_i \Rightarrow x, y \in C_i \ \forall i$

$\Rightarrow \lambda x + (1-\lambda) y \in C_i \ \forall i$

($\because$ C_i are convex-sets for all i)

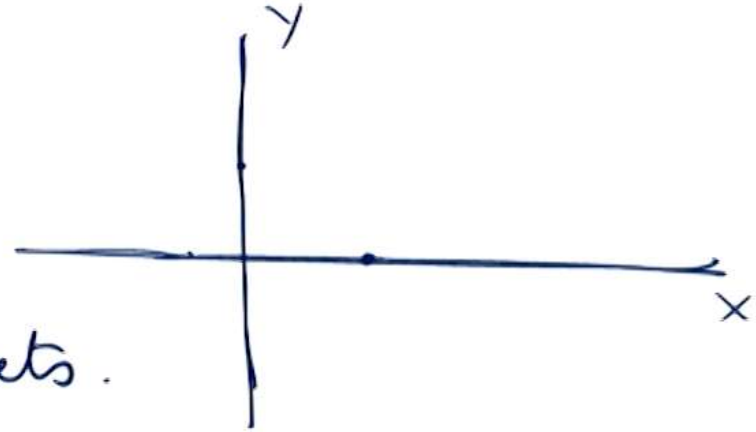
$\Rightarrow \lambda x + (1-\lambda) y \in \bigcap_{i \in I} C_i$

$\Rightarrow \bigcap_{i \in I} C_i$ is also convex.

$S_1 = \{ (x, 0) : x \in \mathbb{R} \}$

$S_2 = \{ (0, y) : y \in \mathbb{R} \}$

S_1 & S_2 are convex-sets in $\mathbb{R}^2$.



$\underline{\underline{S_1 \cup S_2}}$:

$(1, 0) \in S_1 \cup S_2$, $(0, 1) \in S_1 \cup S_2$

But their mid point :

$\left(\frac{1}{2}, \frac{1}{2} \right) \notin S_1 \cup S_2$

$\Rightarrow S_1 \cup S_2$ is not a convex set.

$$\alpha C = \{ \alpha c \; : \; c \in C \}, \quad \alpha \text{ is a scalar.}$$

C is a convex set in $\mathbb{R}^n$.

Let $x, y \in \alpha C$

$\Rightarrow \exists \; c_1, c_2 \in C$ such that

$$x = \alpha c_1, \quad y = \alpha c_2$$

$\lambda x + (1-\lambda) y \quad (\text{for } \lambda \in [0,1])$

$= \lambda \alpha c_1 + (1-\lambda) \alpha c_2$

$\Rightarrow \alpha [\underbrace{\lambda c_1 + (1-\lambda) c_2}]$

$\in \alpha C \quad \{ \because \lambda c_1 + (1-\lambda) c_2 \in C \}$

Convex function

Let $S \subseteq \mathbb{R}^n$ be a convex set. A function $f : S \to \mathbb{R}$ is said to be **convex** over S if for all $x_1, x_2 \in S$, and for all λ with $0 \leq \lambda \leq 1$,

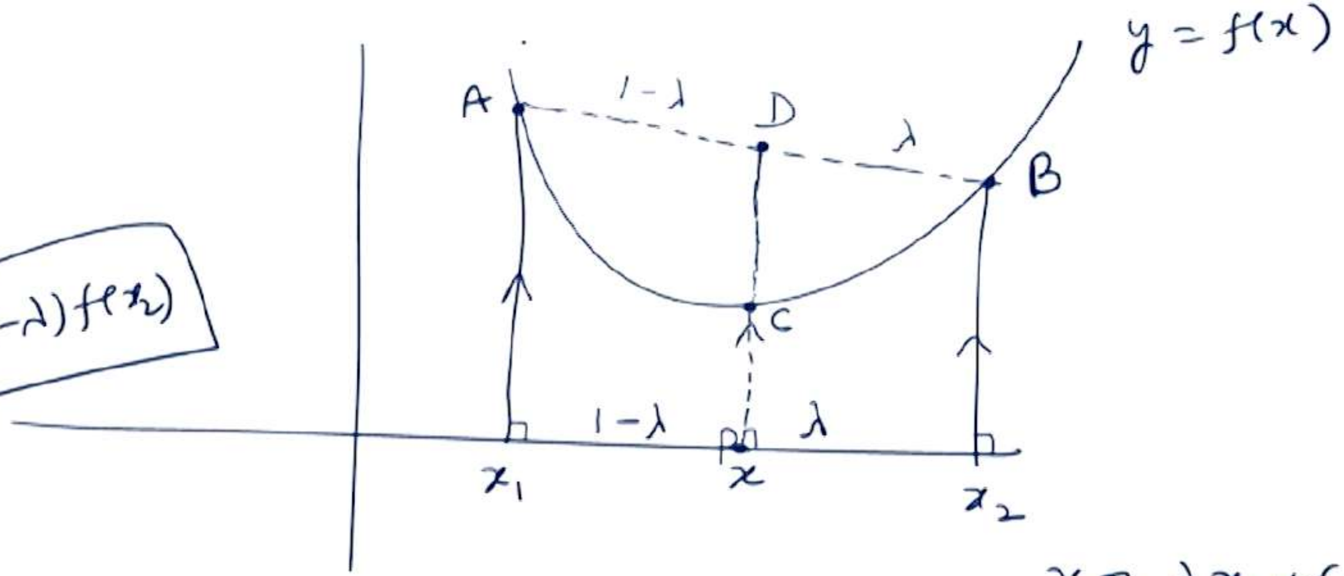
$$\lambda f(x_1) + (1 - \lambda) f(x_2) \geq f(\lambda x_1 + (1 - \lambda) x_2).$$

If the above inequality holds as strict inequality then the function f is called **strictly convex** function on S.

$PC \le PD$

$$\boxed{f(x) \le \lambda f(x_1) + (1-\lambda) f(x_2)}$$

$$f(\lambda x_1 + (1-\lambda) x_2) \le \lambda f(x_1) + (1-\lambda) f(x_2)$$



$x = \lambda x_1 + (1-\lambda) x_2$

$0 \le \lambda \le 1.$

$A(x_1, f(x_1))$

$B(x_2, f(x_2))$

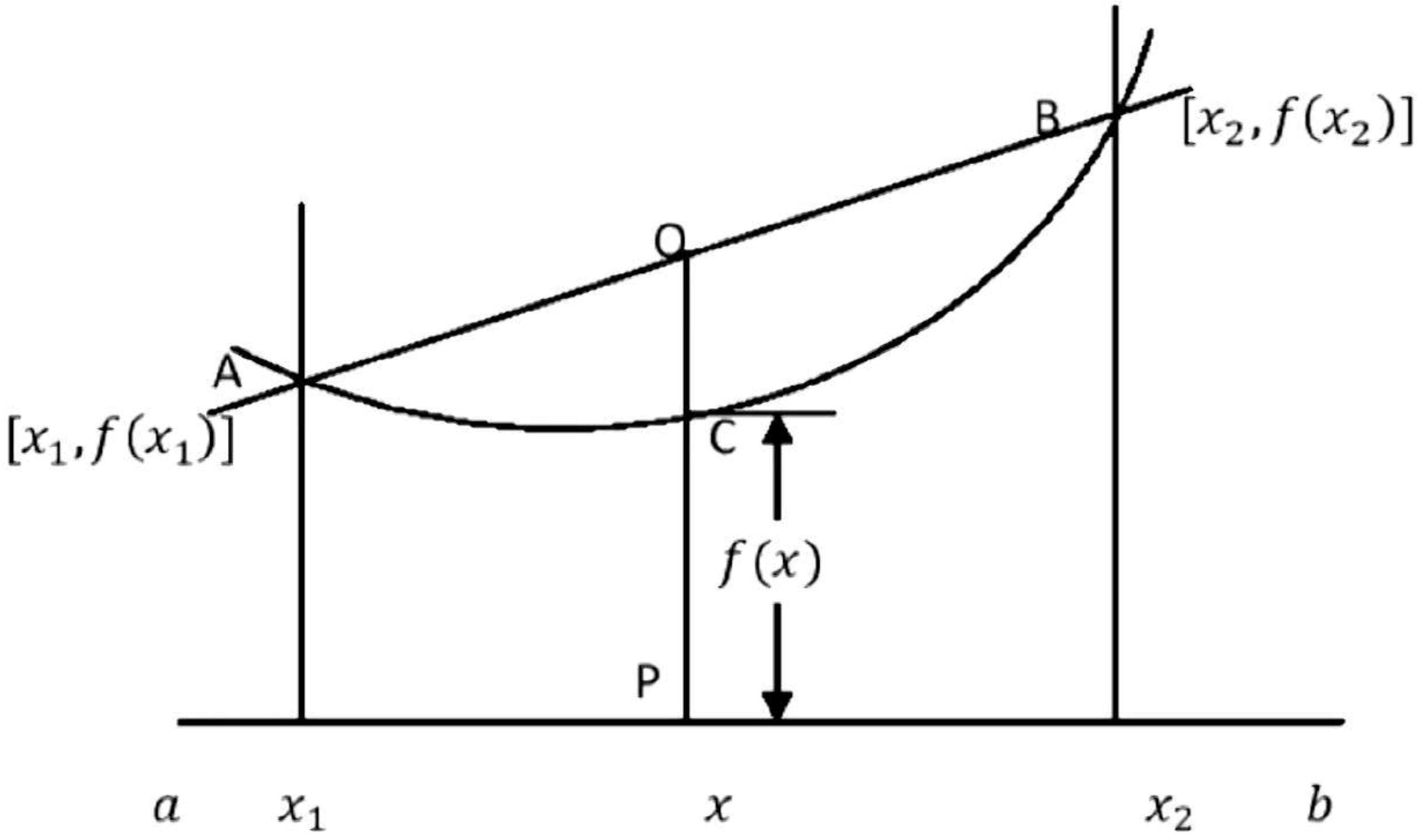
$C(x, f(x))$

$D(x, \underbrace{\lambda f(x_1) + (1-\lambda) f(x_2)})$

$P(x, 0)$

Geometrical Interpretation of Convex function

Let x_1 and x_2 be two distinct points in the domain of f and consider the point $\lambda x_1 + (1-\lambda)x_2$, with $\lambda \in (0,1)$. Note that $\lambda f(x_1) + (1-\lambda)f(x_2)$ gives the weighted average of $f(x_1)$ and $f(x_2)$, while $f[\lambda x_1 + (1-\lambda)x_2]$ gives the value of f at the point $\lambda x_1 + (1-\lambda)x_2$, so, for a convex function f, the value of f at the points on the line segment $\lambda x_1 + (1-\lambda)x_2$ is less then or equal to the height of the chord joining the points $[x_1, f(x_1)]$ and $[x_2, f(x_2)]$ (See figure for illustration)


B
$[x_2, f(x_2)]$
O
A
$[x_1, f(x_1)]$
C
$f(x)$
P
a
x_1
x
x_2
b

Concave function

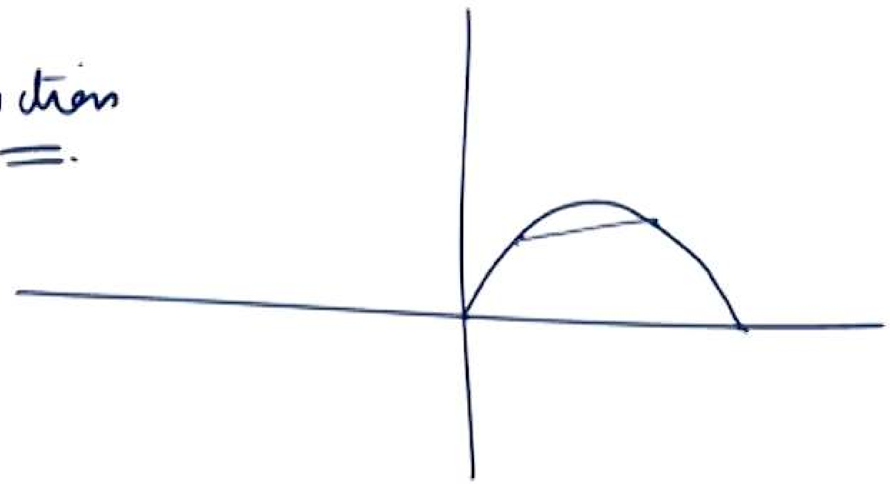
Let $S \subseteq \mathbb{R}^n$ be a convex set. A function $f : S \longrightarrow \mathbb{R}$ is said to be concave over S if for all $x_1, x_2 \in S$, and for all λ with $0 \leq \lambda \leq 1$,

$$\lambda f(x_1) + (1 - \lambda) f(x_2) \leq f(\lambda x_1 + (1 - \lambda) x_2).$$

Obviously, a function f is a concave function if and only if $-f$ is a convex function.

$f : [0, \pi] \to \mathbb{R}, \qquad f(x) = \sin x$

$\longrightarrow$ concave function



Examples

1. $f : \mathbb{R} \to \mathbb{R}$, $f(x) = |x|$ (convex function)
2. $f : \mathbb{R} \to \mathbb{R}$, $f(x) = -x^4$ (concave function)
3. $f : \mathbb{R} \to \mathbb{R}$, $f(x) = 4x + 1$ (both convex and concave function)
4. $f : \mathbb{R} \to \mathbb{R}$, $f(x) = x^3$ (neither convex nor concave function)

Properties of Convex functions

1. The sum of two convex functions is a convex function.
   If $f : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R}^n \to \mathbb{R}$ are convex functions, then
   $f + g : \mathbb{R}^n \to \mathbb{R}, (f + g)(x) = f(x) + g(x)$ is also convex.
2. Positive scalar multiple of a convex function is convex.
   If $f : \mathbb{R}^n \to \mathbb{R}$ is a convex functions, then
   - $\alpha f : \mathbb{R}^n \to \mathbb{R}$ is convex for $\alpha > 0$
   - $\alpha f : \mathbb{R}^n \to \mathbb{R}$ is concave for $\alpha < 0$

Properties of Convex functions

Theorems

1. Let $S \subseteq \mathbb{R}^n$ be a convex set. If $f : S \to \mathbb{R}$ is convex, then any local minimum of f in S is a global minimum on S.
2. Let $S \subseteq \mathbb{R}^n$ be a convex set and $f : S \to \mathbb{R}$ be strictly convex. Then, there is a unique minimizing point of f over S.

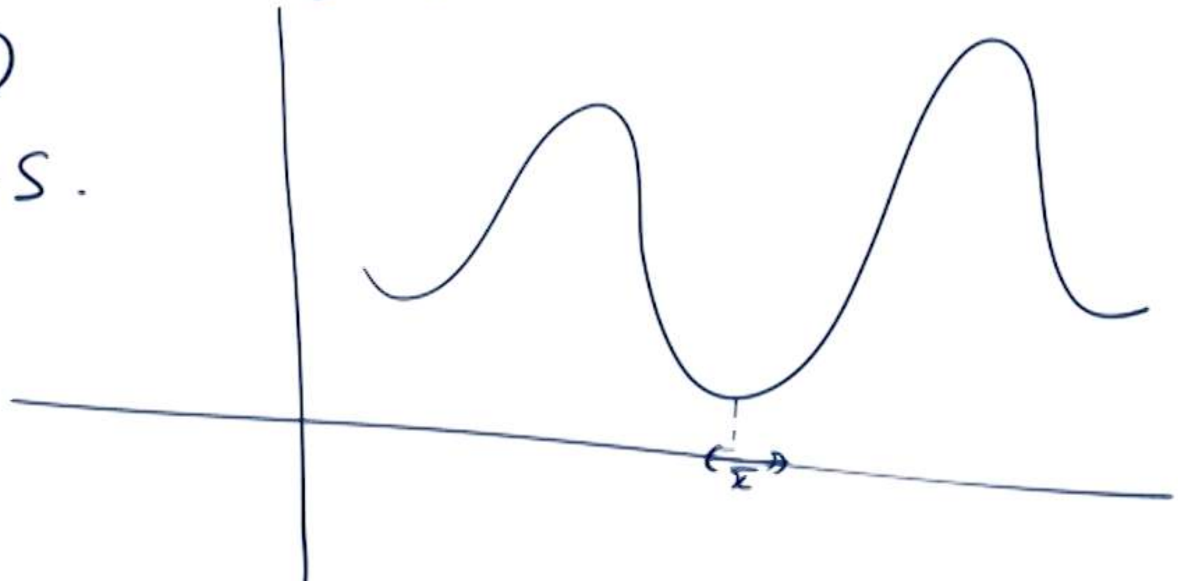
$\bar{x}$ is a local minimum of f on S

$\Rightarrow f(\bar{x}) \le f(x), \ \forall\, x \in N_\delta(\bar{x}) \cap S$, for some $\delta > 0$

$\bar{x}$ is a global minimum of f on S

$\Rightarrow f(\bar{x}) \le f(x)$

$\forall\, x \in S$.



<u>Proof</u>: Let f be convex function on S. Let $\bar{x}$ be a point of local minimum of f.

$\Rightarrow \exists\ \delta > 0$ such that

$$f(\bar{x}) \leq f(x), \quad \forall\ x \in N_\delta(\bar{x}) \cap S \quad \text{——— (1)}$$

Suppose $\bar{x}$ is not a point of global minimum.

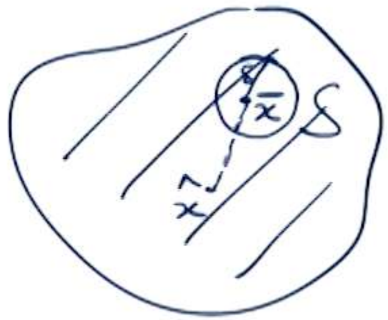
$\Rightarrow \exists\ \hat{x} \in S$ such that

$$f(\hat{x}) < f(\bar{x}) \quad \text{——— (2)}$$

Let $x = \lambda \hat{x} + (1-\lambda)\bar{x}, \quad 0 < \lambda < 1$

$\exists\ \bar{\lambda}, \quad 0 < \bar{\lambda} < 1, \quad$ such that

$$\tilde{x} = \bar{\lambda}\hat{x} + (1-\bar{\lambda})\bar{x} \in N_\delta(\bar{x}) \cap S$$



$$\begin{aligned} f(\tilde{x}) &= f(\lambda\hat{x} + (1-\lambda)\bar{x}) \\ &\le \lambda f(\hat{x}) + (1-\lambda) f(\bar{x}) \\ &< \lambda f(\bar{x}) + (1-\lambda) f(\bar{x}) = f(\bar{x}) \end{aligned}$$

This contradicts (1). Hence, $\bar{x}$ is a point of global minimum.

f is a strictly convex function on S.

Suppose $\bar{x} \in S$ is a global minimum point of f.

let $\bar{x}$ be not a unique point.

$\Rightarrow \exists \; \hat{x} \in S, \quad \bar{x} \neq \hat{x}, \quad f(\bar{x}) = f(\hat{x})$.

$$z = \lambda \bar{x} + (1-\lambda)\hat{x}, \quad 0 < \lambda < 1, \quad z \in S$$

$$\begin{aligned} f(z) &= f(\lambda \bar{x} + (1-\lambda)\hat{x}) \\ &< \lambda f(\bar{x}) + (1-\lambda) f(\hat{x}) \\ &= \lambda f(\hat{x}) + (1-\lambda) f(\hat{x}) \\ &= f(\hat{x}) \qquad \Rightarrow \quad f(z) < f(\hat{x}) \end{aligned}$$

$\bar{x}$ is a unique global minimum point of f on S.

Differentiable functions

1. Let $f : S \longrightarrow \mathbb{R}$ be **differentiable** at $\bar{x} \in S$, where S is an open subset of $\mathbb{R}^n$. Then for $x + \bar{x} \in S$,

$$f(x+\bar{x}) = f(\bar{x}) + x^T(\nabla f(\bar{x})) + \alpha(\bar{x}, x)\,||x||$$

$$\text{where} \quad \lim_{x\to 0} \alpha(\bar{x}, x) = 0.$$

2. Let $f : S \longrightarrow \mathbb{R}$ be **twice differentiable** at $\bar{x} \in S$, where S is an open subset of $\mathbb{R}^n$. Then for $x + \bar{x} \in S$,

$$f(x+\bar{x}) = f(\bar{x}) + x^T(\nabla f(\bar{x})) + \tfrac{1}{2}x^T\nabla^2 f(\bar{x})x + \beta(\bar{x}, x)\,||x||^2$$

$$\text{where} \quad \lim_{x\to 0} \beta(\bar{x}, x) = 0.$$

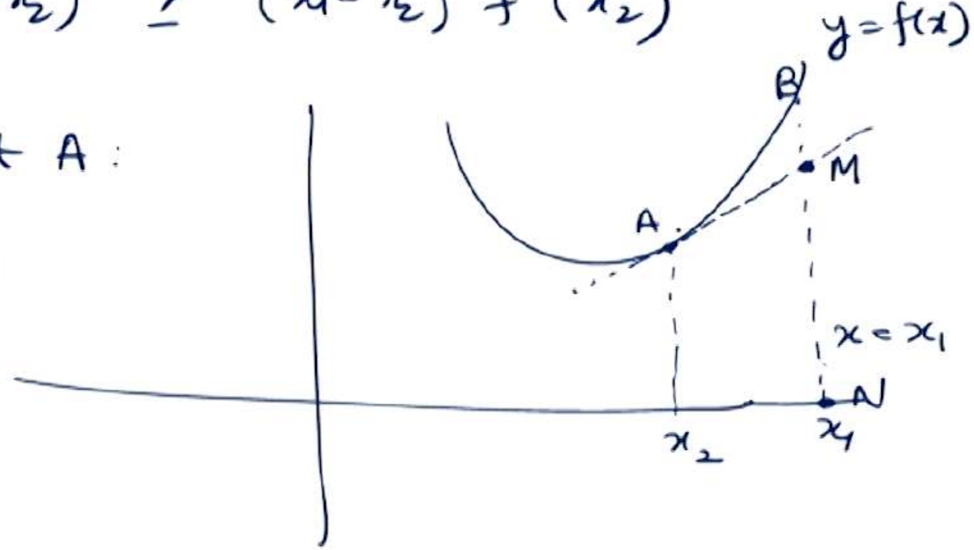
Theorem

Let $f : S \longrightarrow \mathbb{R}$ be differentiable function on an open convex subset S of $\mathbb{R}^n$. Then f is a convex function if and only if

$$f(x_1) - f(x_2) \geq (x_1 - x_2)^T \nabla f(x_2) \qquad \forall x_1, x_2 \in S.$$

$f: S \to \mathbb{R}, \quad S \subseteq \mathbb{R}.$

$\forall x_1, x_2 \in S, \quad f(x_1) - f(x_2) \geq (x_1 - x_2) f'(x_2)$



Equation of Tangent of $y = f(x)$ at A:

$$y - f(x_2) = f'(x_2)(x - x_2)$$

$M\left(x_1,\ f(x_2) + f'(x_2)(x_1 - x_2)\right)$

$MN \leq BN$

$f(x_2) + f'(x_2)(x_1 - x_2) \leq f(x_1)$

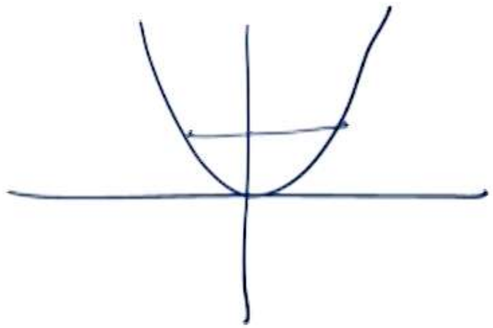
$\Rightarrow f'(x_2)(x_1 - x_2) \leq f(x_1) - f(x_2)$

$A(x_2, f(x_2))$

$B(x_1, f(x_1))$.

$N(x_1, 0)$

$f : \underline{\underline{\mathbb{R}}} \to \mathbb{R}, \quad \underline{\underline{f(x) =}}\ x^2$



let $x_1, x_2 \in \mathbb{R}$

$f(x_1) - f(x_2) - (x_1 - x_2) f'(x_2)$

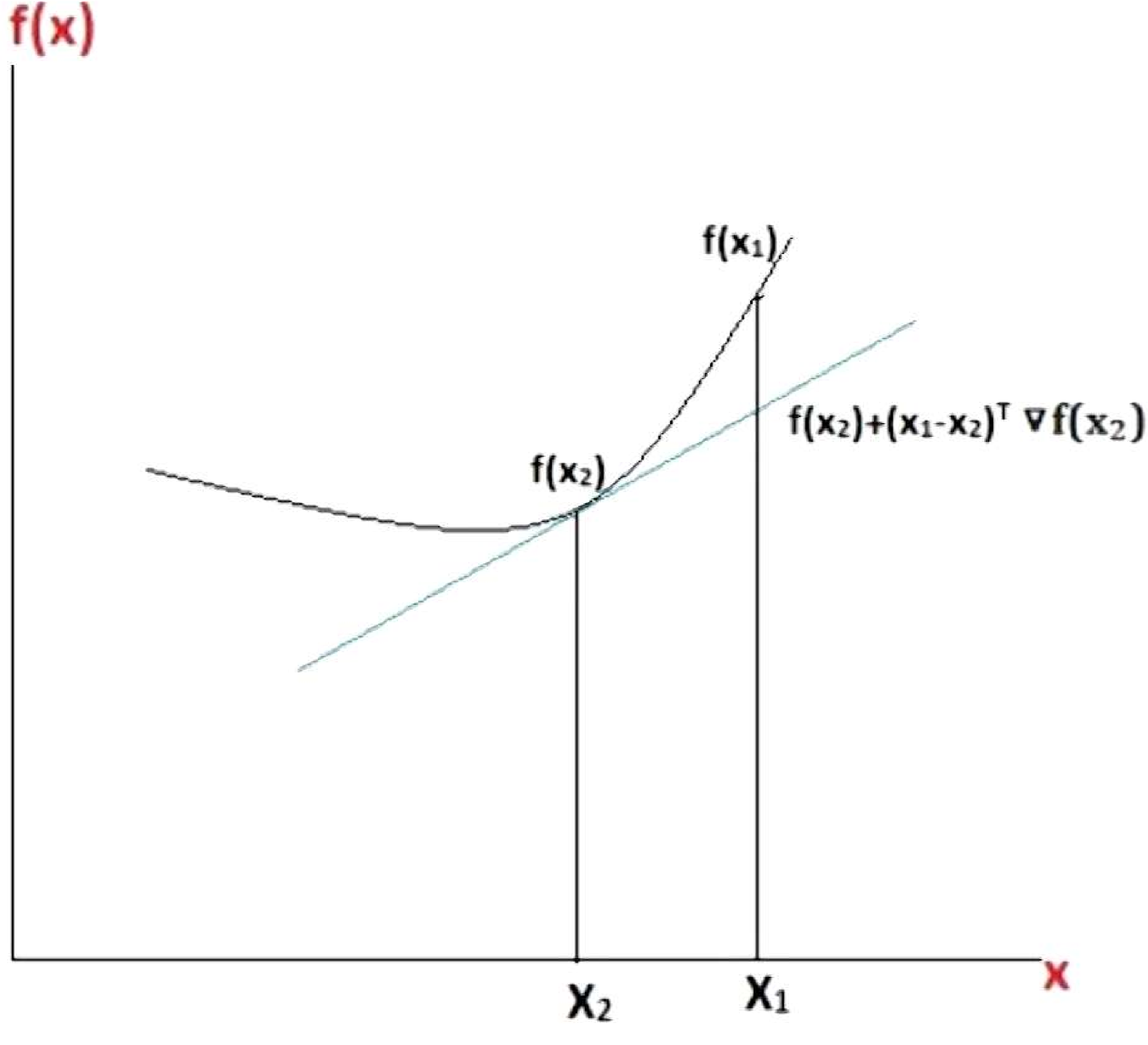
$f'(x) = 2x$

$= x_1^2 - x_2^2 - (x_1 - x_2)\, 2x_2$

$= x_1^2 - x_2^2 - 2x_1x_2 + 2x_2^2$

$= x_1^2 + x_2^2 - 2x_1x_2$

$= (x_1 - x_2)^2 \geq 0$


f(x)
f(x₁)
f(x₂)+(x₁-x₂)ᵀ ∇f(x₂)
f(x₂)
X₂
X₁
x

Definiteness of a matrix

A symmetric matrix H of order $n \times n$ is said to be-

- positive semi-definite if $x^T Hx \geq 0$ for all $x \in \mathbb{R}^n$.
- negative semi-definite if $x^T Hx \leq 0$ for all $x \in \mathbb{R}^n$.
- positive definite if $x^T Hx > 0$ for all $x \in \mathbb{R}^n$, $x \neq 0$.
- negative definite if $x^T Hx < 0$ for all $x \in \mathbb{R}^n$, $x \neq 0$.

Remark

A symmetric matrix H is negative semi-definite (negative definite) if and only if $-H$ is positive semi-definite (positive definite).

$$A = \begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix}_{2\times 2} \qquad A = A^T$$

$$X^T A X = (x_1 \quad x_2) \begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \quad X = (x_1 \quad x_2)^T$$

$$= (x_1 \quad x_2) \begin{pmatrix} 3x_1 + x_2 \\ x_1 + 3x_2 \end{pmatrix}$$

$$= 3x_1^2 + x_1 x_2 + x_2 x_1 + 3x_2^2$$

$$= 3x_1^2 + 2x_1 x_2 + 3x_2^2$$

$$= 2(x_1^2 + x_2^2) + (x_1^2 + x_2^2 + 2x_1 x_2)$$

$$= 2(x_1^2 + x_2^2) + (x_1 + x_2)^2 \geq 0 \quad \forall\, x \in \mathbb{R}^2$$

$$M = \begin{pmatrix} 2 & -2 \\ -2 & 2 \end{pmatrix}_{2\times 2} \qquad M = M^T$$

$$X^T M X = \begin{pmatrix} x_1 & x_2 \end{pmatrix} \begin{pmatrix} 2 & -2 \\ -2 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

$$= \begin{pmatrix} x_1 & x_2 \end{pmatrix} \begin{pmatrix} 2x_1 - 2x_2 \\ -2x_1 + 2x_2 \end{pmatrix} = 2x_1^2 - 4x_1x_2 + 2x_2^2$$

$$= 2\,(x_1 - x_2)^2 \geq 0$$

$$\forall\ (x_1, x_2) \in \mathbb{R}^2$$

$\Rightarrow$ M is positive semi-definite.

Tests for Definiteness of a Matrix

Test 1: Eigenvalue Test

Let A be a real symmetric matrix of order n. Then, A is

- **positive definite** if and only if all its eigenvalues are positive.
- **positive semi-definite** if and only if all its eigenvalues are non-negative.
- **negative definite** if and only if all its eigenvalues are negative.
- **negative semi-definite** if and only if all its eigenvalues are non-positive.
- **indefinite** if and only if there is atleast one positive eigenvalue and atleast one negative eigenvalue of A.

Principle minor

A principal minor D_k of a matrix A of order k is the determinant of the matrix formed by deleting any $(n-k)$ rows and $(n-k)$ columns with the same number.

Test 2: Principal Minor Test

The necessary and sufficient condition for a symmetric matrix to be **positive semi-definite** is that all the possible principal minors should be non-negative.

$$M = \begin{pmatrix} 4 & 2 & 3 \\ 2 & 3 & 2 \\ 3 & 2 & 4 \end{pmatrix}_{3\times 3} \qquad M = M^T$$

Minors of order 1×1 : $4, 4, 3$

Minors of order 2×2 : $\begin{vmatrix} 3 & 2 \\ 2 & 4 \end{vmatrix} = 8$, $\begin{vmatrix} 4 & 3 \\ 3 & 4 \end{vmatrix} = 7$, $\begin{vmatrix} 4 & 2 \\ 2 & 3 \end{vmatrix} = 8$

Minor of order 3×3 : $\begin{vmatrix} 4 & 2 & 3 \\ 2 & 3 & 2 \\ 3 & 2 & 4 \end{vmatrix} = 4(12-4) - 2(8-6) + 3(4-9)$

$= 32 - 4 - 15$

> 0

M is positive definite.

Theorem

Let S be a non-empty open convex subset of $\mathbb{R}^n$ and $f : S \longrightarrow \mathbb{R}$ be twice differentiable on S. Then f is a convex-function on S *iff* the Hessian matrix $\nabla^2 f(x)$ is positive semi-definite $\forall x \in S$.

$f : \mathbb{R}^3 \to \mathbb{R}$, $\quad f(x,y,z) = x^2 + 4y^2 + z^2 + 4xy + 4yz + 2xz$

$$\nabla^2 f = \begin{pmatrix} 2 & 4 & 2 \\ 4 & 8 & 4 \\ 2 & 4 & 2 \end{pmatrix}$$

$f_x = 2x + 4y + 2z$

$f_y = 8y + 4x + 4z$

$f_z = 2z + 4y + 2x$

Minors of order 1×1 : 2, 8, 2

Minors of order 2×2 :

$\begin{vmatrix} 2 & 4 \\ 4 & 8 \end{vmatrix} = 0 \qquad \begin{vmatrix} 8 & 4 \\ 4 & 2 \end{vmatrix} = 0$

$\begin{vmatrix} 2 & 2 \\ 2 & 2 \end{vmatrix} = 0$.

Minors of order 3×3 : $|\nabla^2 f|$

$= 2(0) - 4(0) + 2(0)$

$= 0$

$\nabla^2 f$ is positive semi-definite $\forall\ (x, y, z) \in \mathbb{R}^3$

$\Rightarrow$ f is convex function on $\mathbb{R}^3$.

Area of the greatest rectangle that can be inscribed in the ellipse $\frac{x^2}{a^2}+\frac{y^2}{b^2}=1$ is :

AIEEE 2005

$2a\cos\frac{\pi}{4} = \sqrt{2}a, \quad B = \sqrt{2}b$

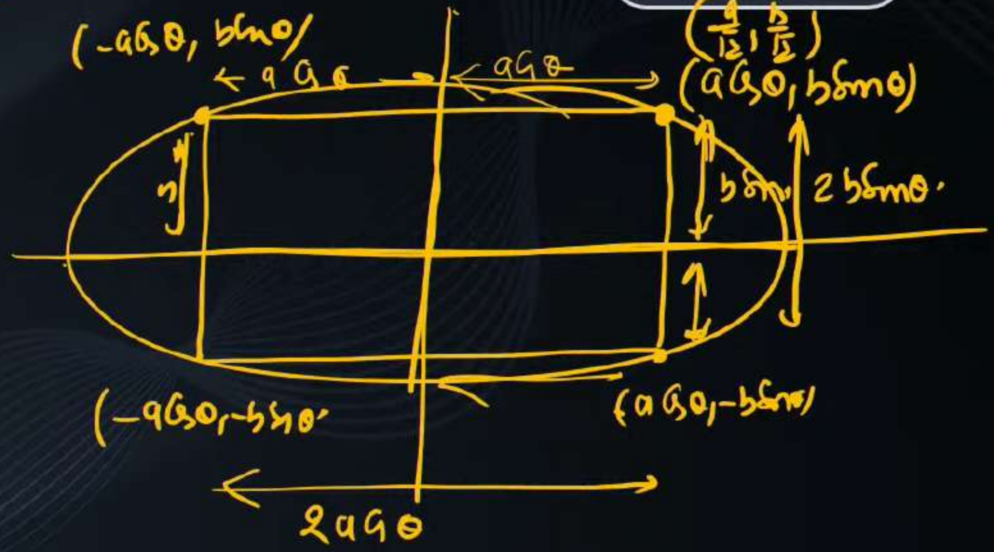
$\frac{x^2}{a^2}+\frac{y^2}{b^2}=1$

a) $2ab$

b) ab

c) $\sqrt{ab}$

d) $\frac{a}{b}$



$\sin 2\theta = 1$

$2\theta = \frac{\pi}{2}$

$\theta = \frac{\pi}{4}$

$A = 2a\cos\theta \times 2b\sin\theta$

$A_{Max} = 2ab\,\boxed{\sin 2\theta}_{Max}$

$A_{Max} = 2ab$

Lagrange's Mean Value Thm.

If $f(x)$ in Interval $[a,b]$ satisfy following 2 cond^n

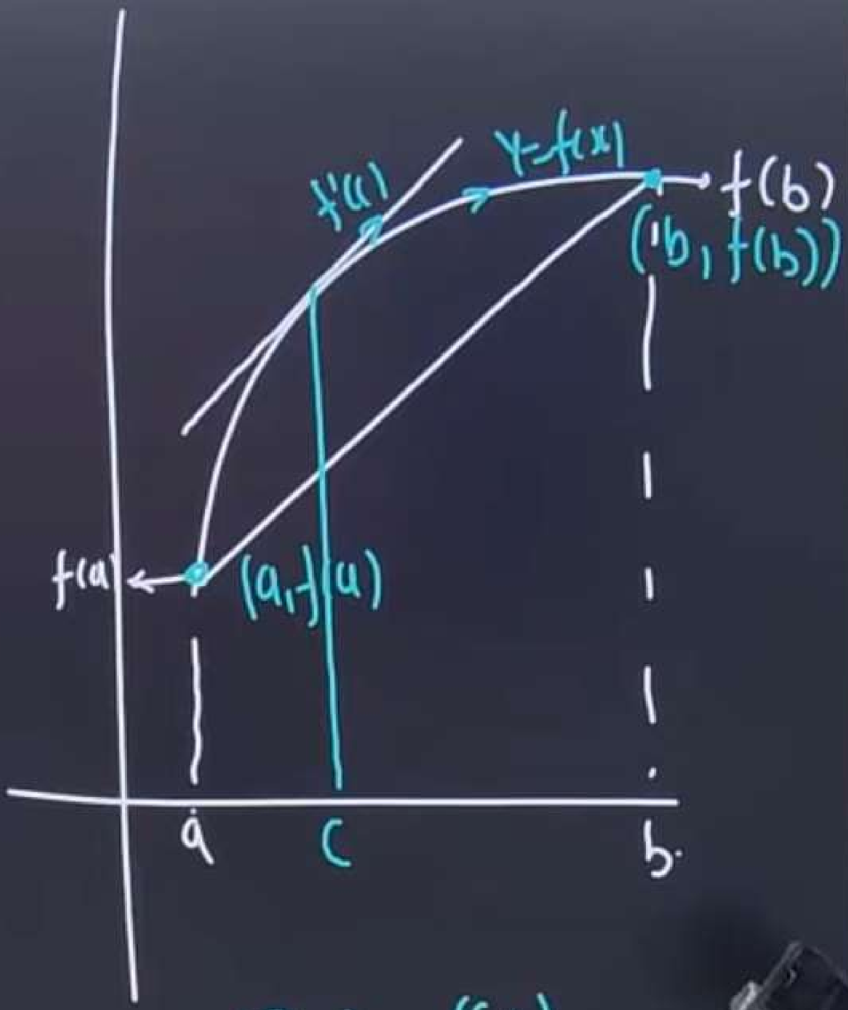
① $f(x)$ is conts in $[a,b]$

② $f(x)$ is diff^{ble} in (a,b)

then acc to LMVT $\exists$ atleast one Pt. $x=c$ where

$$f'(c)=\frac{f(b)-f(a)}{b-a}$$

When $f(a)=f(b)$ then $f'(c)=0$ [R MVT]



$$(Sl)_T=(Sl)_{Chord}$$

$$f'(c)=\frac{f(b)-f(a)}{b-a}$$

Q $f(x)$ is Concave downward & $f'(x) > 0$

$x_1 \neq x_2$ Which is grtr

$$f'\left(\frac{x_1+x_2}{2}\right) \text{ or } \frac{f'(x_1)+f'(x_2)}{2}$$

Rolle's Thm. [Mean Value Thm]

If a fxn $y = f(x)$ in Interval $[a, b]$

Satisfies

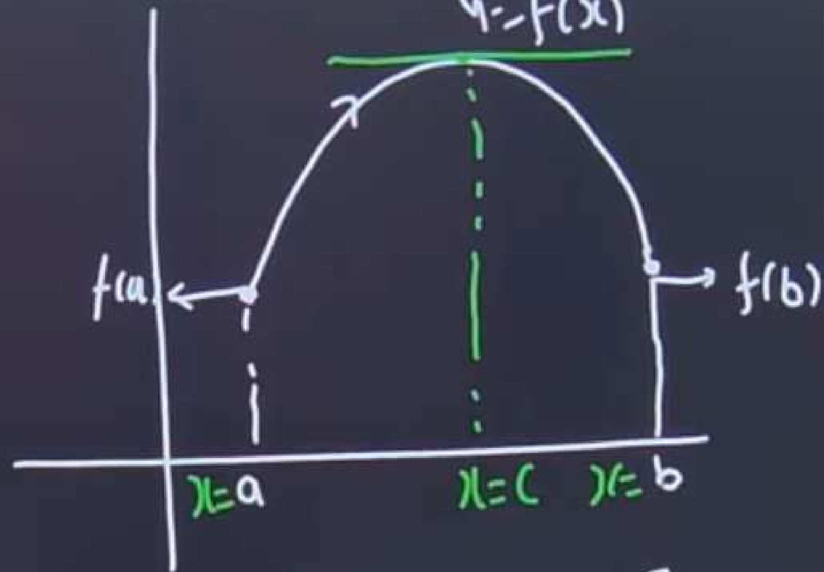
A) $f(x)$ is Conts in $[a, b]$ ✓

B) $f(x)$ is diff in (a, b) ✓

C) $f(a) = f(b)$

then acc. to RMVT $\exists$ at least at Pt $x = c$ such that $f'(c) = 0$

(B) Geometrical Meaning of RMVT.



Bet $x = a$ & $x = b$ $\exists$ at least one Pt $x = c$ where tangent is $\parallel^{r}$ to X Axis

Optimization Problem

The basic form of an optimization problem is as follows:

$$\begin{array}{cc} (P) & \text{Min } f(x) \\ & \text{subject to } x \in C, \end{array}$$

where $f : R^n \longrightarrow R$, and $C \subseteq R^n$.

- The problem (P) is also called basic mathematical programming problem.
- The function f is called the objective function and the set C is called the **constraint set or feasible set**.

continued...

- A point $\bar{x} \in C$ is called a **feasible point**. The feasible point where the above problem attains maxima or minima is called **optimal solution or optimal point**. If $C = \phi$, then the problem (P) is called infeasible.

- If $C = R^n$ then the problem (P) is called **unconstrained optimization** problem, otherwise we call it **constrained optimization** problem.

Optimality conditions

First order optimality conditions

Let $f : U \to \mathbb{R}$ be a function defined on a set $U \subseteq \mathbb{R}^n$. Suppose that $\bar{x} \in \text{int}(U)$ is a local optimal point and all partial derivatives of f exist at $\bar{x}$. Then, $\nabla f(\bar{x}) = 0$.

Stationary Point

Let $f : U \to \mathbb{R}$ be a function defined on a set $U \subseteq \mathbb{R}^n$. Suppose $\bar{x} \in \text{int}(U)$ and f be differentiable over some neighbourhood of $\bar{x}$. Then, $\bar{x}$ is called a stationary point of f if $\nabla f(\bar{x}) = 0$.

Second order optimality conditions

Necessary conditions

Let $f : U \to \mathbb{R}$ be a twice differentiable function defined on an open set $U \subseteq \mathbb{R}^n$. Let $\bar{x} \in int(U)$ be a local minimum of f. Then, $\nabla^2 f(\bar{x})$ is positive semi-definite.

Sufficient conditions

Let $f : U \to \mathbb{R}$ be a twice differentiable function defined on an open set $U \subseteq \mathbb{R}^n$ and $\bar{x} \in int(U)$ be a stationary point. If $\nabla^2 f(\bar{x})$ is positive definite then $\bar{x}$ is a strict local minimum point of f.

$$f = xy - x^2 - y^2 - 2x - 2y + 4$$

$$\nabla f = 0 \Rightarrow \frac{\partial f}{\partial x} = 0, \quad \frac{\partial f}{\partial y} = 0$$

$$\Rightarrow y - 2x - 2 = 0, \quad x - 2y - 2 = 0$$

$$\Rightarrow x = y = -2$$

$(-2, -2)$ is only stationary point.

local maxima of f

$$H_f = \begin{pmatrix} -2 & 1 \\ 1 & -2 \end{pmatrix}$$

Minors of order 1×1: $-2, -2$

" " " 2×2: $\begin{vmatrix} -2 & 1 \\ 1 & -2 \end{vmatrix} = 4 - 1 = \underline{\underline{3}}$