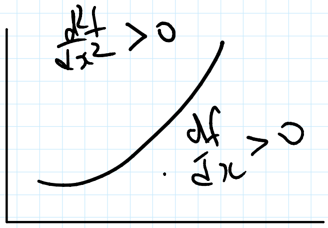
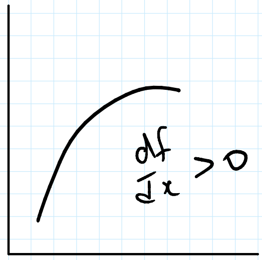
Hessian
15 March 2021 11:55

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

$n=1$ $\quad f(x+h) \approx f(x) + h\underbrace{f'(x)}_{\text{how fast } f^n \text{ changes}} + \frac{1}{2}h^2 \underbrace{f''(x)}_{\text{curvature}} + \dots$ higher order terms





Difference in second derivative $\frac{d^2f}{dx^2}$

$n \geq 1 \quad x \in \mathbb{R}^n \quad f: \mathbb{R}^n \to \mathbb{R} \quad f(x_1, x_2 \dots x_n) \in \mathbb{R}$

$$\nabla f = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix}$$

f^n is differentiable

∇^2 (Hessian) =

2^{nd} order differentiable

$$\begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \frac{\partial^2 f}{\partial x_1 \partial x_3} & \dots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_1 \partial x_2} & \frac{\partial^2 f}{\partial x_2^2} & & & \\ \vdots & & \ddots & & \\ & & & \ddots & \\ \frac{\partial^2 f}{\partial x_1 \partial x_n} & \frac{\partial^2 f}{\partial x_2 x_n} & \dots & & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$$

$n \times n$ symmetric matrix

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

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j x_i}$$

$f(x) = c_0$.

$\nabla f = 0$

$$\nabla^2 = \begin{bmatrix} 0 & \cdots & 0 \\ & \ddots & \\ & \cdots & 0 \end{bmatrix}_{n\times n}$$

$$f(x) = c_0 + \underset{\sim}{c}^T x \qquad c:\ n\times 1 \text{ vector} \qquad f(x) = c_0 + c_1 x$$

$$\nabla f = c \qquad \frac{df}{dx} = c_1$$

$$\nabla^2 = \begin{bmatrix} \frac{\partial f}{\partial x_1^2} \to 0 & \frac{\partial f}{\partial x_1 \partial x_2} \to 0 & \cdots \to 0 \\ & & \\ & & \end{bmatrix}_{n\times n} \qquad \frac{d^2 f}{dx^2} = 0$$

$$f(x) = c_0 + \underset{\sim}{c}^T x + \frac{1}{2} x^T \underset{-}{H} x \qquad f(x) = c_0 + c_1 x + \frac{1}{2} h x^2$$

$$\nabla f = c + Hx \qquad \frac{df}{dx} = c_1 + hx$$

$$\nabla^2 = ? \qquad \frac{d^2 f}{dx^2} = \underset{\cdot}{h}$$

$n = 2$.

$$f(x) = c_0 + c_1 x_1 + c_2 x_2 + \frac{1}{2}\left[h_{11} x_1^2 + h_{22} x_2^2 + 2 h_{12} x_1 x_2\right]$$

$$\frac{\partial f}{\partial x_1} = c_1 + h_{11} x_1 + h_{12} x_2$$

$$\frac{\partial f}{\partial x_2} = c_2 + h_{22} x_2 + h_{12} x_1$$

$$\nabla f = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} + \begin{bmatrix} h_{11} & h_{12} \\ h_{12} & h_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$c + \qquad Hx$$

$$\frac{\partial^2 f}{\partial x_1^2} = \frac{\partial}{\partial x_1}\left[\frac{\partial f}{\partial x_1}\right] = \frac{\partial}{\partial x_1}\left[\underbrace{c_1}_{0} + h_{11} x_1 + \underbrace{h_{12} x_2}_{0}\right] = h_{11}$$

$$\frac{\partial^2 f}{\partial x_2^2} = \frac{\partial}{\partial x_2}\left[\frac{\partial f}{\partial x_2}\right] = \frac{\partial}{\partial x_2}\left[\underbrace{c_2}_{0} + h_{22}x_2 + \underbrace{h_{12}x_1}_{0}\right] = h_{22}$$

$$\frac{\partial^2 f}{\partial x_1 \partial x_2} = \frac{\partial}{\partial x_1}\left[c_2 + h_{22}x_2 + h_{12}x_1\right] = h_{12}$$

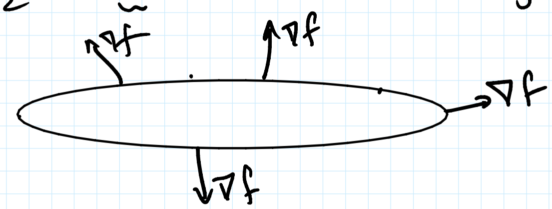
$$\frac{\partial^2 f}{\partial x_2 \partial x_1} = \frac{\partial}{\partial x_2}\left[c_1 + h_{11}x_1 + h_{12}x_2\right] = h_{12}$$

$$\nabla^2 = \begin{bmatrix} h_{11} & h_{12} \\ h_{12} & h_{22} \end{bmatrix} = H$$

$f(x) = c_0 + c^T x + \frac{1}{2} x^T H x$ $\quad$ (H: symmetric)

$\nabla f = c + Hx$

$\nabla^2 = H$



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

general nonlinear continuous & 2^{nd} order differentiable

$$\nabla f = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix} \qquad \nabla^2 = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & & \\ \frac{\partial^2 f}{\partial x_1 \partial x_2} & \ddots & \\ \vdots & & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$$

$n = 1$ Taylor series

$$f(x+h) = f(x) + h f'(x) + \frac{1}{2} h^2 \frac{d^2 f}{dx^2} + \dots \text{ higher order terms}$$

$n \geq 1$

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

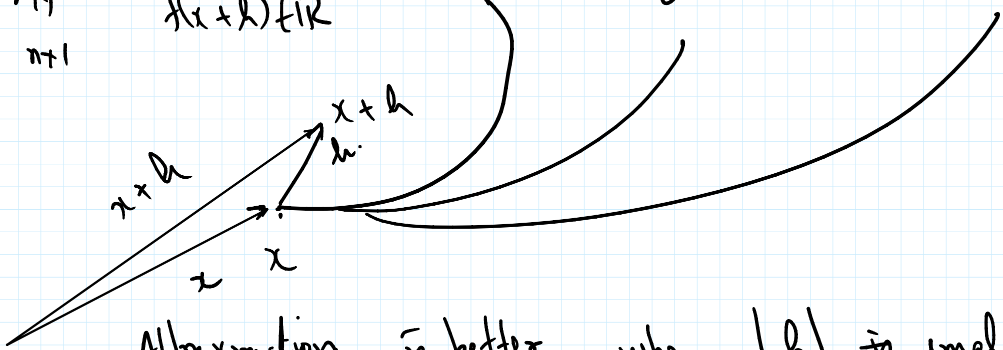
$n \geq 1$

$$f(\underset{\sim}{x}+h) \approx f(x) + \underbrace{\nabla f(x)^T}_{1\times n} \underbrace{h}_{n\times 1} + \frac{1}{2} \underbrace{h^T}_{1\times n} \underbrace{\nabla^2 (x)}_{n\times n} \cdot \underbrace{h}_{n\times 1} + \cdots$$

x: $n\times 1$

h: $n\times 1$

$f(x+h) \in \mathbb{R}$



Approximations is better when $|h|$ is small

$$f(x) = c_0 + c^T x + \frac{1}{2} x^T H x.$$

f: quadratic $\quad \nabla f = Hx + c$

$\nabla^2 = H$

$$f(x+h) = f(x) + \nabla f(x)^T h + \frac{1}{2} h^T \nabla^2 (x)\, h + \text{higher order term}$$

$$f(x+h) = f(x) + (Hx+c)^T h + \frac{1}{2} h^T H h. + \text{higher order term}$$

$$= c_0 + c^T (x+h) + \frac{1}{2} (x+h)^T H (x+h)$$

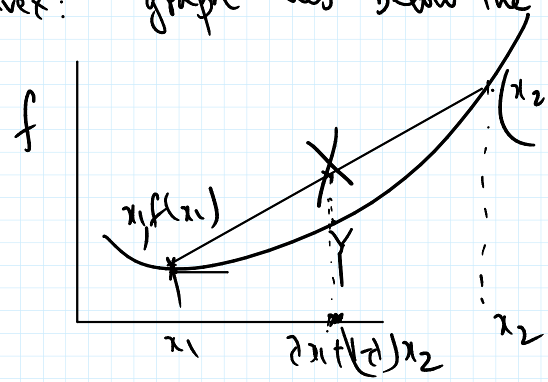
$$f(x+h) = f(x) + \underbrace{\nabla f^T}_{(Hx+c)} h + \frac{1}{2} h^T \underbrace{\nabla^2}_{(H)} h.$$

(exact when f is quadratic)

$$= c_0 + c^T (x+h) + \frac{1}{2} (x+h)^T H (x+h).$$

f: convex: graph lies below the line segment going $(x_1, f(x_1))$ $(x_2, f(x_2))$ $\forall x_1, x_2$ $(x_2\ f(x_2))$

$n = 1$



Any point on the line segment has co-ordinates (x)

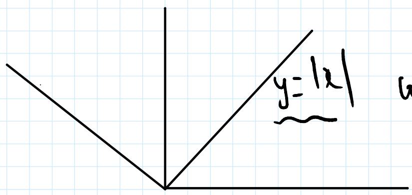
co-ordinates of X $\left(\lambda x_1 + (1-\lambda) x_2 , \quad \underbrace{\lambda f(x_1) + (1-\lambda)}_{\geq} f(x_2) \right)$ $\quad 0 \leq \lambda \leq 1$

co-ordinates of Y $\left(\lambda x_1 + (1-\lambda) x_2 , \quad f(\lambda x_1 + (1-\lambda) x_2) \right)$

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

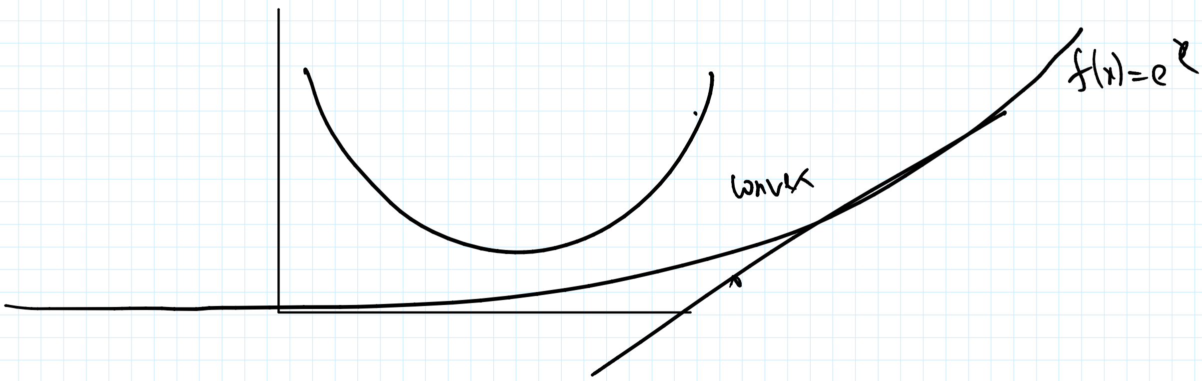
$0 \leq \lambda \leq 1$

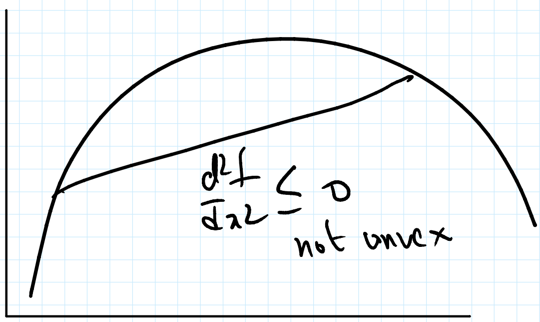
$\forall x_1 x_2$

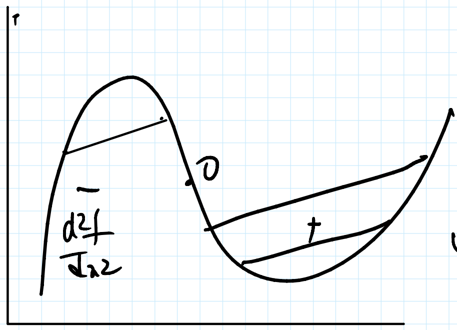


$y = |x|$ convex but not differentiable

fn. is convex & differentiable $\iff$ $\frac{d^2 f}{dx^2} \geq 0$







not convex!

$n \geq 1$ $\quad f(\lambda x_1 + ((1-\lambda) x_2)) \leq \lambda f(x_1) + (1-\lambda) f(x_2)$

$0 \leq \lambda \leq 1$

$0 \leq \lambda \leq 1$

$\forall x_1 \; x_2$

f is convex !

$n \geq 1 \quad f : \mathbb{R}^n \to \mathbb{R}$

$\nabla^2 \geq 0$ (f is convex & differentiable)

$(n \times n)$?

∇^2 is positive semi-definite (convexity)

∇^2 is positive definite (strictly convex)