Notation :

① $x \in \mathbb{R}^n$ $\quad x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$, $x^T = (x_1, \cdots, x_n)$

② $f(x)$ is real valued and assume it smooth for now

③ Gradient $\nabla f(x) \in \mathbb{R}^n$ $\quad \nabla f = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{pmatrix}$

④ Hessian $\nabla^2 f(x) \in \mathbb{R}^{n \times n}$ $\quad [\nabla^2 f]_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j}$

⑤ $y^T x = \langle x, y \rangle = \sum_{i=1}^{n} x_i y_i$

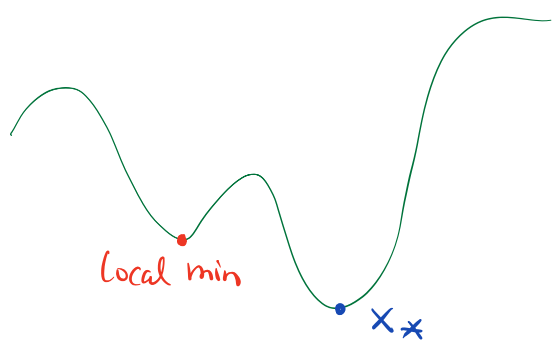
Unconstrained Minimization (Part I) :

$$\min_{x \in \mathbb{R}^n} f(x)$$

① Global minimizer x_* means

$$f(x_*) \le f(x) \ , \ \forall x$$

② Local mimimizer



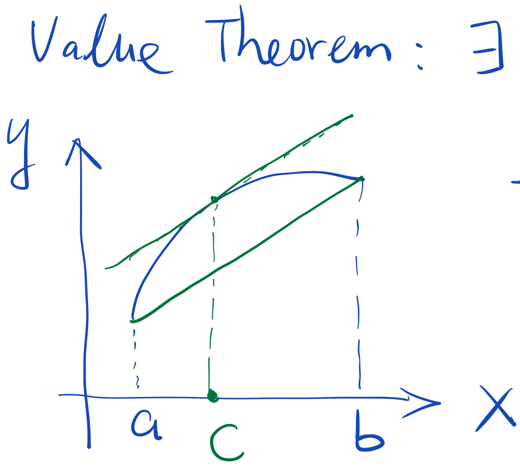
Review of Calculus :

① Taylor Theorem for single variable :

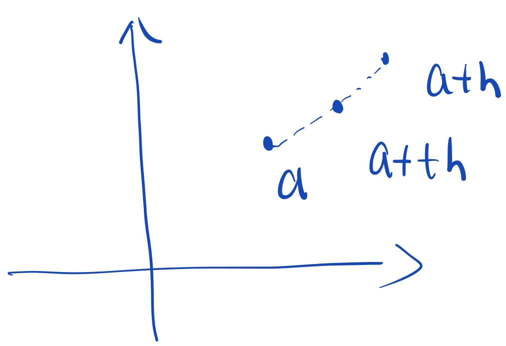
$$f(x + \Delta x) = f(x) + f'(x)\Delta x + \frac{1}{2} f''(y) \Delta x^2$$

for some $y \in (x, x + \Delta x)$

② Mean Value Theorem: $\exists\, c \in (a, b)$ s.t.



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



Theorem B.1 (Multivariate Quadratic Taylor's Theorem)**.** *Suppose that $S \subset \mathbb{R}^n$ is an open set and that $f : S \longrightarrow \mathbb{R}$ is a function of class C^2 on S. Then for $\mathbf{a} \in S$ and $\mathbf{h} \in \mathbb{R}^n$ such that the line segment connecting $\mathbf{a}$ and $\mathbf{a} + \mathbf{h}$ is contained in S, there exists $\theta \in (0, 1)$ such that*

$$f(\mathbf{a} + \mathbf{h}) = f(\mathbf{a}) + \nabla f(\mathbf{a}) \cdot \mathbf{h} + \frac{1}{2}\mathbf{h}^T \nabla^2 f(\mathbf{a} + \theta\mathbf{h})\mathbf{h}.$$

Proof. Define $g(t) = f(\mathbf{a} + t\mathbf{h})$. By Lemma B.1 on $g(t)$, there is $\theta \in (0, 1)$ s.t.

$$g(1) = g(0) + g'(0) + \frac{1}{2}g''(\theta).$$

By chain rule, we have $g'(0) = \nabla f(\mathbf{a}) \cdot \mathbf{h}$ and $g''(\theta) = h^T \nabla^2 f(\mathbf{a} + \theta\mathbf{h})\mathbf{h}$, which complete the proof. □

Lemma B.1 (Second-order Mean Value Theorem)**.** *Suppose that $I \subset \mathbb{R}$ is an open interval and that $f(x)$ is a function of class C^2 ($f''(x)$ exists and is continuous) on I. For $a \in I$ and h such that $a + h \in I$, there exists some $\theta \in (0, 1)$ such that*

$$f(a + h) = f(a) + hf'(a) + \frac{h^2}{2}f''(a + \theta h).$$

Proof. Consider $g_1(x) = f(x) - f(a) - (x - a)f'(a)$ then $g_1(a) = g_1'(a) = 0$. Define

$$g(x) = g_1(x) - \left(\frac{x - a}{h}\right)^2 g_1(a + h),$$

then $g(a) = g'(a) = g(a + h) = 0$. By Mean Value Theorem , we have

$$g(a) = g(a + h) = 0 \Longrightarrow g'(a + \alpha h) = 0,$$

for some $\alpha \in (0, 1)$. Use Mean Value Theorem again on $g'(a) = g'(a+\alpha h) = 0$, we get $g''(a+\theta h) = 0$ for some $\theta \in (0, \alpha)$. Since $g''(x) = f''(x) - \frac{2}{h^2}g_1(a+h)$, $g''(a + \theta h) = 0$ implies that we get the explicit remainder for the second order Taylor expansion as $g_1(a + h) = \frac{h^2}{2}f''(a + \theta h)$. □

$g'(a) = 0$

$g'(a+\alpha h) = 0$

Theorem (First Order Necessary Condition)

$$\left.\begin{array}{l} x_* \text{ is a local minimizer} \\ f(x) \text{ is smooth} \end{array}\right\} \Rightarrow \nabla f(x_*) = 0$$

Proof: Assume $\nabla f(x_*) \neq 0$.

Let $p = -\nabla f(x_*)$, then $p^T \nabla f(x_*) = -\|\nabla f(x_*)\|^2 < 0$

$\nabla f(x)$ is continuous $\Rightarrow p^T \nabla f(x^* + tp) < 0$, $\forall t \in [0, T]$ for some T. ↳ for any

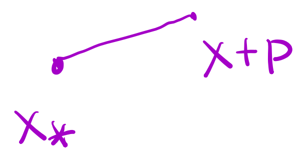
$h(t) = p^T \nabla f(x^* + tp)$ is continuous and $h(0) < 0$

$\forall t \in (0, T]$, by Taylor's Theorem,

$$f(x_* + tp) = f(x_*) + t p^T \nabla f(x_* + \theta p) \quad \text{for some } \theta \in (0, t)$$

$$\Rightarrow f(x_* + tp) < f(x_*).$$

Contradiction



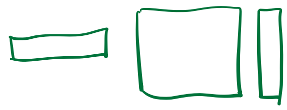
Stationary / Critical Point $\nabla f(x) = 0$

Theorem (Second Order Necessary Condition)

$$\left.\begin{array}{l} x_* \text{ is a local minimizer} \\ f(x) \text{ is smooth} \end{array}\right\} \Rightarrow \nabla f(x_*) = 0,\ \nabla^2 f(x_*) \geq 0$$

$\nabla^2 f(x_*) \geq 0$ means the Hessian Matrix is positive semi-definite, meaning ① all its eigenvalues are non-negative.

② $y^T \nabla^2 f(x_*) y \geq 0$ for any $y \in \mathbb{R}^n$



Proof: We already proved $\nabla f(x_*) = 0$

Assume $\nabla^2 f(x_*)$ is NOT PSD.

$\Rightarrow \exists\ p \in \mathbb{R}^n$ s.t. $p^T \nabla^2 f(x_*) p < 0$

$\nabla^2 f(x)$ is continuous $\Rightarrow p^T \nabla^2 f(x_* + tp) p < 0$, $\forall t \in [0, T]$
for some T

Taylor's Thm $\Rightarrow \forall t \in (0, T]$, $\exists\ \theta \in (0, t)$ s.t.

$$f(x_* + tp) = f(x_*) + t p^T \underbrace{\nabla f(x_*)}_{=0} + \frac{1}{2} t^2 \underbrace{p^T \nabla^2 f(x^* + \theta p) p^T}_{<0}$$

$< f(x_*)$

Contradiction.