Notation:

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

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

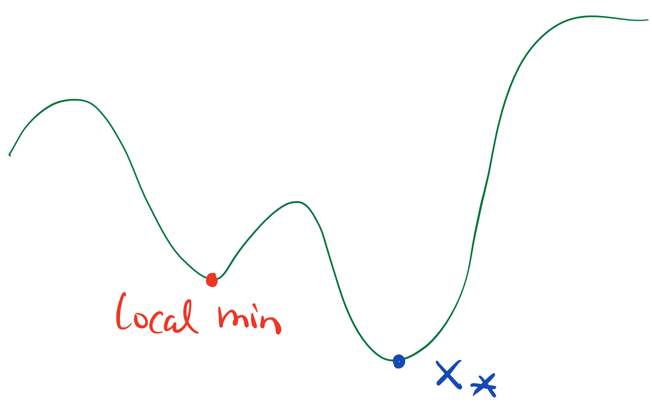
Unconstrained Minimization (Part I):

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

① Global minimizer x_* means

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

② Local mimimizer



I. Review of Calculus:

① Taylor Theorem for single variable:

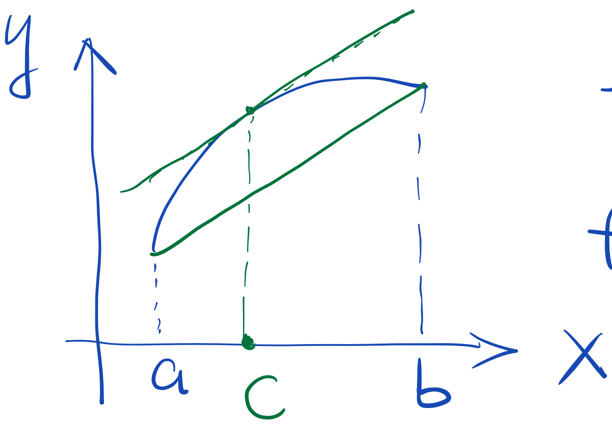
$$f(x+\Delta x) = f(x) + f'(x)\Delta x + \frac{1}{2} f''(y) \Delta x^2$$

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

② $\exists$: there exists

Mean Value Theorem:

For a single variable function, $\exists\, c \in (a, b)$ s.t.



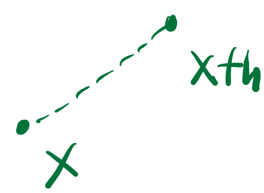
$$f'(c) = \frac{f(b)-f(a)}{b-a}$$

$$f(b)-f(a) = f'(c)(b-a)$$

③ Multi-Variable First Order Taylor Theorem $\quad x \in \mathbb{R}^n$

$$f(x+h) = f(x) + \langle \nabla f(x+\theta h), h \rangle, \quad \theta \in (0,1)$$

Proof: $g(t) = f(x+th)$



$$\begin{cases} g(0) = f(x) \\ g(1) = f(x+h) \\ g'(t) = \langle \nabla f(x+th), h \rangle \end{cases}$$

$\exists\, \theta \in (0,1)$ s.t. $g(1) - g(0) = g'(\theta) \cdot 1$

$$\Rightarrow f(x+h) - f(x) = \langle \nabla f(x+\theta h), h \rangle$$

④ Second Order Mean Value Theorem

$f(x)$ is single variable : $\exists\, \theta \in (0,1)$

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

⑤ Multi-Variable 2nd-Order Taylor Theorem

$\forall\, x, h \in \mathbb{R}^n$, $\exists\, \theta \in (0,1)$

$$f(x+h) = f(x) + \langle \nabla f(x), h \rangle + \frac{1}{2} h^T \nabla^2 f(x+\theta h) h .$$

Proof: $g(t) = f(x+th)$, $t \in [0,1]$

II.

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

Theorem (First Order Necessary Condition)

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

Theorem (Second Order Necessary Condition)

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

Hessian is positive semi-definite

all its eigenvalues ≥ 0

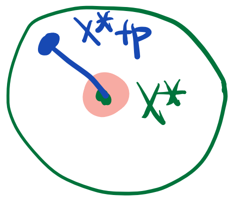
Theorem (Second order Sufficient Condition)

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

Proof: $\left.\begin{array}{l} \nabla^2 f(x) \text{ is continuous} \\ \nabla^2 f(x^*) > 0 \end{array}\right\} \Rightarrow$ There is an open ball

$B = \{ y \in \mathbb{R}^n : \|y - x^*\| < r \}$

s.t. $\nabla^2 f(x) > 0$, $\forall x \in B$.



Entries of $\nabla^2 f(x)$ are continuous w.r.t. x

Eigenvalues of a matrix are continuous w.r.t. entries.

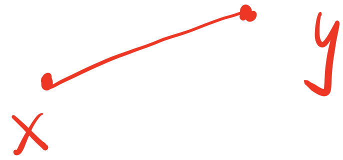
For any nonzero $p \in \mathbb{R}^n$ with $\|p\| < r$,

we have $x^* + p \in B$

$$\Rightarrow f(x^*+p) = f(x^*) + p^T \underbrace{\nabla f(x^*)}_{=0} + \frac{1}{2} \underbrace{p^T \nabla^2 f(x+\theta p) p}_{>0}$$

$$> f(x^*)$$

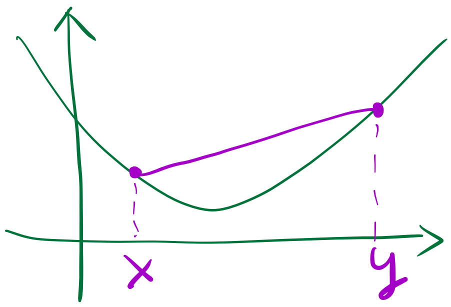
III. Convex Functions



Def $f(x)$ is convex if $\forall x, y \in \mathbb{R}^n$, $\forall \lambda \in (0,1)$,

$$f(\lambda x + (1-\lambda) y) \leq \lambda f(x) + (1-\lambda) f(y)$$

Example:



$f(x) = (x-2)^2 + 1$

Theorem ① $\nabla f(x)$ exists

$f(x)$ is convex $\Longleftrightarrow$ $f(y) \geq f(x) + \nabla f(x)^T (y-x), \forall x, y$

② If $\nabla^2 f(x)$ exists,

$f(x)$ is convex $\Longleftrightarrow$ $\nabla^2 f(x) \succeq 0$, $\forall x$.

Theorem Assume $f(x)$ is convex. Then

① $\nabla f(x_*) = 0 \Longleftrightarrow x_*$ is a global minimizer

② Any local minimizer is a global one

Proof:

Assume $\nabla f(x_*)=0$

$$f(y) \geq f(x_*) + \nabla f(x_*)^T(y - x_*) = f(x_*)$$

$\Rightarrow$ x_* is a global minimizer.