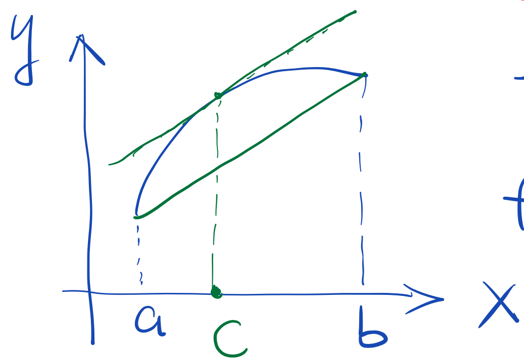
I. Calculus

$\exists$: there exists

① Mean Value Theorem:

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



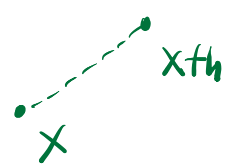
$$f'(c) = \frac{f(b) - f(a)}{b - a}$$

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

② Multi-Variable First Order Taylor Theorem

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

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. $g(1) - g(0) = g'(\theta) \cdot 1$

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

③ Second Order Mean Value Theorem

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

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

④ 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)

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

Theorem (Second Order Necessary Condition)

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

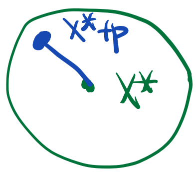
Theorem (Second order Sufficient Condition)

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

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.

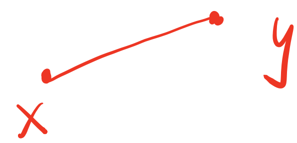
For any nonzero $p \in \mathbb{R}^n$ with $\|p\| < r$,

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

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

$$> f(x^*)$$

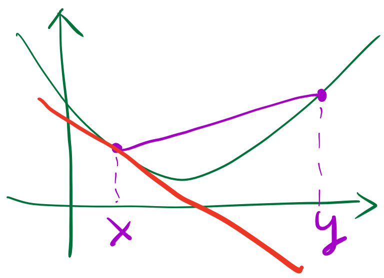
III. Convex Functions



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

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

Example :



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

Theorem ① $\nabla f(x)$ exists

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

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.

Def (Lipschitz Continuity)

$f(x)$ is Lipschitz continuous if

$$\exists L>0, \quad \forall x, y \in \mathbb{R}^n, \quad |f(x)-f(y)| \leq L \|x-y\|.$$

Example: ① If $f'(x)$ is bounded, then $f(x)$ is L-continuous

$|f'(x)| \leq M, \ \forall x$

Proof: $\left| \dfrac{f(x)-f(y)}{x-y} \right| = |f'(c)| \leq M$

$\Rightarrow |f(x)-f(y)| \leq M \cdot |x-y|.$

② $f(x) = |x|$

$$||x|-|y|| = \begin{cases} |x-y| & \text{if } x \cdot y > 0 \\ |x+y| < |x-y| & \text{if } x \cdot y \leq 0 \end{cases}$$

③ $\|\nabla f(x)\|$ is bounded $\quad f(y) = f(x) + \langle \nabla f(\cdot), y-x \rangle$

Proof: $g(t) = f(y + t(x-y))$

$$g(0) = f(y)$$
$$g(1) = f(x)$$
$$g'(t) = \nabla f(y + t(x-y))^T (x-y)$$

$$|f(x) - f(y)| = \left| \frac{g(0) - g(1)}{0 - 1} \right| = |g'(t)| = |\nabla f(z)^T (x-y)|$$
$$\leq \|\nabla f(z)\| \cdot \|x-y\|$$
$$\leq M \cdot \|x-y\|$$

Cauchy Schwartz Inequality

$$|a \cdot b| \leq \|a\| \cdot \|b\|$$

$$|a_1 b_1 + a_2 b_2 + \cdots + a_n b_n| \leq \sqrt{\sum_{i=1}^n a_i^2} \sqrt{\sum_{i=1}^n b_i^2}$$

Notation $A \in \mathbb{R}^{n \times n}$

① $\|A\|$ is defined as $\max_{x \in \mathbb{R}^n} \dfrac{\|Ax\|}{\|x\|}$

spectral norm

② The singular value of A, $\sigma_i(A) \geq 0$

is defined by

$\sigma_i^2(A)$ is the eigenvalue of AA^T or AA^T

③ $\sigma_1(A) \geq \sigma_2(A) \geq \cdots \geq \sigma_n(A) \geq 0$

④ $\|\lambda A + (1-\lambda) B\| \leq \lambda \|A\| + (1-\lambda) \|B\|$ $\lambda \in (0,1)$

Useful Facts spectral norm is convex

1) $\|A\| = \sigma_1$ spectral norm

2) $\|A\|_F \triangleq \sqrt{\sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij}^2} = \sqrt{\sum_{i=1}^{n} \sigma_i^2}$

$\downarrow$ Frobenius norm

3) Number of nonzero σ_i is $\operatorname{rank}(A)$

4) If A is real symmetric positive semi-definite, $\sigma_i(A) = \lambda_i(A)$

$\swarrow$ eigenvalue

5) If A is real symmetric, $\sigma_i(A) = |\lambda_i|$

6) $x \in \mathbb{R}^n$, $\|Ax\| \leq \|A\| \cdot \|x\|$

Theorem: $\|\nabla^2 f(x)\| \leq L$, $\forall x$

$\Rightarrow \nabla f(x)$ is Lipschitz continuous

Proof:

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

$\nabla f(x+h) = \nabla f(x) + \nabla^2 f(x+\theta h) h$

is wrong: no such vector-valued Taylor Thm.

$g(t) = \nabla f(x+th)$ is a vector-valued function

$g(1) = \nabla f(x+th)$

$g(0) = \nabla f(x)$

$$g'(t) = \nabla^2 f(x+th)\, h$$

Foundamental Theorem of Calculus

$$g(1) - g(0) = \int_0^1 g'(t)\, dt$$

$$\Rightarrow \nabla f(x+th) - \nabla f(x) = \int_0^1 \nabla^2 f(x+th)\, h \; dt$$

$$\Rightarrow \nabla f(x+h) - \nabla f(x) = \left[\int_0^1 \nabla^2 f(x+th)\, dt\right] h$$

$$\Rightarrow \|\nabla f(x+h) - \nabla f(x)\| = \left\|\left[\int_0^1 \nabla^2 f(x+th)\, dt\right] h\right\|$$

$$\leq \left\|\int_0^1 \nabla^2 f(x+th)\, dt\right\| \cdot \|h\|$$

Jensen's Inequality for convex function

$\| \int_0^1 A(t) dt \| \leq \int_0^1 \|A\| dt$

$$\leftarrow \leq \int_0^1 \|\nabla^2 f(x+th)\|\, dt \cdot \|h\|$$

$$\leq L \cdot \|h\|$$

$$\Rightarrow \|\nabla f(y) - \nabla f(x)\| \leq L \cdot \|y - x\|.$$

Example: For a smooth convex function $f(x)$,

we have $\nabla^2 f(x) \succeq 0$. $\|\nabla^2 f(x)\| \leq a$

If assume $\nabla^2 f(x) \preceq aI$, $\forall x$

$aI - \nabla^2 f(x) \succeq 0$ means its eigenvalues ≥ 0

then $\nabla f(x)$ is L-Continuous with $L = a$.

Descent Lemma Assume $\nabla f(x)$ is L-Lipschitz

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

Proof: Foundamental Theorem of Calculus

$$\Rightarrow f(y) - f(x) = \int_0^1 \langle \nabla f(\underbrace{x+t(y-x)}_{z}), y-x \rangle dt$$

$$g(t) = f(x+t(y-x))$$

$$g(0) = f(x), \quad g(1) = f(y)$$

$$g(1) - g(0) = \int_0^1 g'(t)\, dt$$

$$f(y) - f(x) = \langle \nabla f(x), y-x \rangle + \int_0^1 \langle \nabla f(z) - \nabla f(x), y-x \rangle dt$$

$$\left| f(y) - f(x) - \langle \nabla f(x), y-x \rangle \right| = \left| \int_0^1 \langle \nabla f(z) - \nabla f(x), y-x \rangle dt \right|$$

$$\le \int_0^1 \left| \langle \nabla f(z) - \nabla f(x), y-x \rangle \right| dt$$

$$\le \int_0^1 \|\nabla f(z) - \nabla f(x)\| \cdot \|y-x\|\, dt$$

$$= \int_0^1 \|\nabla f(x+t(y-x)) - \nabla f(x)\| \cdot \|y-x\|\, dt$$

$$\le \int_0^1 tL\|y-x\| \cdot \|y-x\|\, dt$$

$$= \frac{L}{2}\|y-x\|^2.$$

Remark: ① The proof also implies a lower bound

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

② Recall that if $f(x)$ is convex,

$$f(y) \geq f(x) + \nabla f(x)^T (y-x).$$