Def (Lipschitz Continuity)

$f(x)$ is Lipschitz continuous if

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

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 $f(x) = |x|$

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

③ $\|\nabla f(x)\|$ is bounded

$f(y) = f(x) + \langle \nabla f(\cdot), y-x \rangle$

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

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

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

Cauchy-Schwartz Inequality $\leqslant M\cdot\|x-y\|$

$$|a\cdot b| \leqslant \|a\|\cdot\|b\|$$

$$|a_1b_1+a_2b_2+\cdots+a_nb_n| \leqslant \sqrt{\sum_{i=1}^{n}a_i^2}\sqrt{\sum_{i=1}^{n}b_i^2}$$

Notation $A\in\mathbb{R}^{n\times n}$ $\quad \sigma(A)=\sqrt{\lambda(A^TA)}=\sqrt{\lambda(AA^T)}$

① $\|A\|$ is defined as $\max\limits_{x\in\mathbb{R}^n}\dfrac{\|Ax\|}{\|x\|}=\max\limits_i \sigma_i(A)$

spectral norm

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

is defined by $\quad x^TA^TAx=(Ax)^TAx\geqslant 0$

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

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

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

spectral norm is convex

Useful Facts

1) $\|A\|=\sigma_1(A)=\sqrt{\lambda_1(A^TA)}$ 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 rank(A)

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

eigenvalue

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

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

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

Theorem: $\|\nabla^2 f(x)\| \leq L$, $\forall x \in \mathbb{R}^n$

$\Rightarrow \nabla f(x)$ 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

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

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

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

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

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

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

Jensen's Inequality for convex function
$\| \int_0^1 A(t)\,dt \| \leq \int_0^1 \|A\|\,dt$

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

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

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 $L = a$.

<u>Descent Lemma</u> Assume $\nabla f(x)$ is L-Lipschitz

① $f(y) \leq f(x) + \nabla f(x)^T (y-x) + \frac{L}{2}\|x-y\|^2$

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

Taylor Theorem $f(y) = f(x) + \nabla f(x)^T (y-x) + \frac{1}{2}(y-x)^T \nabla^2 f(z)(y-x)$

$$z = x + \theta(y-x)$$

Proof: Foundamental Theorem of Calculus

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

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

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

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

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

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

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

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

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

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

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

Remark: ① The proof also implies a lower bound

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

② Recall that if $f(x)$ is convex,

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

Courant-Fischer-Weyl Minmax Principle:

$A \in \mathbb{R}^{n\times n}$ is real symmetric with eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$, then

$$\lambda_n \leq \frac{x^T A x}{x^T x} \leq \lambda_1 \quad , \quad \forall x \in \mathbb{R}^n$$

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

$$\Rightarrow \quad |\lambda_i(\nabla^2 f)| = \sigma_i(\nabla^2 f) \leq \sigma_1 \leq L \, , \ \forall x$$

$$\Rightarrow \quad \left| \frac{h^T \nabla^2 f(z) h}{h^T h} \right| \leq \max\{|\lambda_1|, |\lambda_n|\} = \sigma_1 \leq L$$

$$-\frac{1}{2}L\|y-x\|^2 \leq \frac{1}{2}(y-x)^T \nabla^2 f(z)(y-x) \leq \frac{1}{2}L\|y-x\|^2$$