Review

① Convex Function : $\forall x, y \in \mathbb{R}^n$

$$f(\lambda x + (1-\lambda) y) \leq \lambda f(x) + (1-\lambda) f(y), \quad \forall \lambda \in (0,1)$$

Example: $A \in \mathbb{R}^{n\times n}$

$f(A) = \|A\|$ is convex w.r.t. A.

Convex function satisfies

1) $f(\underbrace{\sum_{i=1}^{n} a_i x_i}_{\text{Convex combination}}) \leq \sum_{i=1}^{n} a_i f(x_i), \quad \forall x_i \in \mathbb{R}^n$

Convex combination $\boxed{a_i \geq 0, \ \sum_{i=1}^{m} a_i = 1}$

2) If $g(\cdot)$ is an integrable function,

$$f\left(\frac{1}{b-a}\int_a^b g(t)\,dt\right) \leq \frac{1}{b-a}\int_a^b f[g(t)]\,dt$$

3) $\forall x, y, \quad f(y) \geq f(x) + \langle \nabla f(x), y - x\rangle$

if ∇f exists.

4) $\nabla^2 f(x) \geq 0$ if $\nabla^2 f$ exists

② Def (Lipschitz Continuity)

$f(x)$ is Lipschitz continuous if

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

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

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

③

Descent Lemma 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 $f(y) = f(x) + \nabla f(x)^T(y-x) + \frac{1}{2}(y-x)^T \underbrace{\nabla^2 f(z)}(y-x)$

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

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

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

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

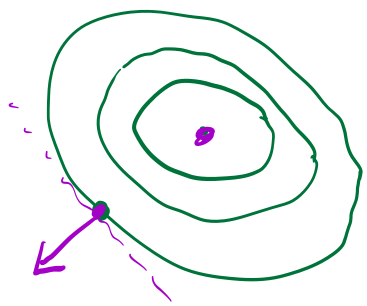
$$-\frac{1}{2}L\|y-x\|^2 \leq \frac{1}{2}(y-x)^T \nabla^2 f(z)(y-x) \leq \frac{1}{2}L\|y-x\|^2$$

$\boxed{\text{Sufficient Decrease Lemma}}$

Assume $\nabla f(x)$ is L-Lipschitz. Then $\forall x \in \mathbb{R}^n$, $\forall \eta > 0$

$$f(x) - f(x - \eta \nabla f(x)) \geq \eta \left(1 - \frac{L}{2}\eta\right) \|\nabla f(x)\|^2$$

Gradient Descent $x_{k+1} = x_k - \eta \nabla f(x_k)$, $\eta > 0$



Contour lines of

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

Proof: Descent Lemma $f(y) \leq f(x) + \langle \nabla f(x), y - x \rangle + \frac{L}{2}\|y - x\|^2$

$$\Rightarrow f(x - \eta \nabla f(x)) \leq f(x) + \langle \nabla f(x), -\eta \nabla f(x) \rangle + \frac{L}{2}\|\eta \nabla f(x)\|^2$$

$$= f(x) - \eta \left(1 - \frac{L}{2}\eta\right) \|\nabla f(x)\|^2.$$