Convergence Rate for Strongly Convex Function

Theorem Strong Convexity & L-conitinuity ∇f

$$\langle \nabla f(x) - \nabla f(y), x-y \rangle \geq \frac{\mu L}{\mu + L} \|x-y\|^2 + \frac{1}{\mu + L} \|\nabla f(x) - \nabla f(y)\|^2$$

Remark: Convexity ($\mu = 0$) & L-conitinuity

$$\Rightarrow \|\nabla f(x) - \nabla f(y)\|^2 \leq L \langle \nabla f(x) - \nabla f(y), x-y \rangle$$

Convexity $\Rightarrow$ $\langle \nabla f(x) - \nabla f(y), x-y \rangle \geq 0$

Strong Convexity $\Rightarrow$ $\langle \nabla f(x) - \nabla f(y), x-y \rangle \geq \mu \|x-y\|^2$

Theorem Assume ∇f is L-continuous.

Assume $f(x)$ is strongly convex with μ

Then for $x_{k+1} = x_k - \eta \nabla f(x_k)$

with any $\eta \in (0, \frac{2}{L+\mu}]$:

$$\|x_k - x^*\|^2 \leq \left(1 - \frac{2\eta \mu L}{\mu + L}\right)^k \|x_0 - x^*\|^2$$

If $\eta = \frac{2}{L+\mu}$, then

$$\|x_k - x^*\| \leq \left(\frac{L/\mu - 1}{L/\mu + 1} \right)^k \|x_0 - x^*\|$$

$$f(x_k) - f(x_*) \leq \frac{L}{2} \left(\frac{L/\mu - 1}{L/\mu + 1} \right)^{2k} \|x_0 - x^*\|^2$$

Remark: ① $\frac{L}{\mu}$ is the condition number of $f(x)$.

② $f(x_k) - f_* \to 0$ is faster than

$\|x_k - x_*\| \longrightarrow 0$

Proof: Let $r_k = \|x_k - x_*\|$

$$\begin{aligned} r_{k+1}^2 &= \|x_{k+1} - x_*\|^2 \\ &= \|x_k - x^* - \eta \nabla f(x_k)\|^2 \\ &= r_k^2 + 2\langle -\eta \nabla f(x_k), x_k - x^* \rangle + \eta^2 \|\nabla f(x_k)\|^2 \\ &= r_k^2 - 2\eta \langle \nabla f(x_k) - \nabla f(x_*), x_k - x^* \rangle \\ &\quad + \eta^2 \|\nabla f(x_k)\|^2 \end{aligned}$$

$$\langle \nabla f(x) - \nabla f(y), x - y \rangle \geq \frac{\mu L}{\mu + L} \|x - y\|^2 + \frac{1}{\mu + L} \|\nabla f(x) - \nabla f(y)\|^2$$

$$\langle \nabla f(x_k) - \nabla f(x_*), x_k - x_* \rangle \geq \frac{\mu L}{\mu+L} \|x_k - x_*\|^2 + \frac{1}{\mu+L} \|\nabla f(x_k)\|^2$$

$$\leq r_k^2 - 2\eta \frac{\mu L}{\mu+L} \|x_k - x_*\|^2 - 2\eta \frac{1}{\mu+L} \|\nabla f(x_k)\|^2 + \eta^2 \|\nabla f(x_k)\|^2$$

$$= \left(1 - 2\eta \frac{\mu L}{\mu+L}\right) r_k^2 + \eta\left(\eta - \frac{2}{\mu+L}\right) \|\nabla f(x_k)\|^2$$

So $\eta \in \left(0, \frac{2}{\mu+L}\right] \Rightarrow 0 \leq 1 - 2\eta \frac{\mu L}{\mu+L} < 1$

$$r_{k+1}^2 \leq \left(1 - 2\eta \frac{\mu L}{\mu+L}\right) r_k^2$$

$\Leftrightarrow 2\eta \frac{\mu L}{\mu+L} < 1$

$\frac{2}{\mu+L} \cdot \frac{2\mu L}{\mu+L} \leq 1$

$$\Rightarrow r_k^2 \leq \left(1 - \frac{2\eta \mu L}{\mu+L}\right)^k r_0^2$$

$\Leftrightarrow 4\mu L \leq (\mu+L)^2$

If $\eta = \frac{2}{\mu+L} \Rightarrow r_k^2 \leq \left[1 - \frac{4\mu L}{(\mu+L)^2}\right]^k r_0^2$

$$\Rightarrow r_k \leq \left[\frac{L-\mu}{L+\mu}\right]^k r_0$$

Descent Lemma

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

$$\Rightarrow f(x_k) \leq f(x_*) + \nabla f(x_*)^T (x_k - x_*) + \frac{L}{2} \|x_k - x_*\|^2$$

$$\Rightarrow \quad f(x_k) - f(x_*) \le \frac{L}{2}\|x_k - x_*\|^2$$

$$\le \frac{L}{2}\left[\frac{L-\mu}{L+\mu}\right]^{2k}\|x_0 - x_*\|^2 \quad \square$$

Convergence Rates of GD

Always assume $\nabla f(x)$ is Lip-cont. with L

$\omega = \eta\left(\frac{2}{L} - \eta\right)$

① No other assumptions, $\eta \in (0, \frac{2}{L})$

$$\min_{0 \le k \le N} \|\nabla f(x_k)\| \le \frac{1}{\sqrt{N+1}}\sqrt{\frac{1}{\omega}\left[f(x_0) - f(x_*)\right]} = O\left(\frac{1}{\sqrt{k}}\right)$$

② $f(x)$ is convex, $\eta \in (0, \frac{2}{L})$

$$f(x_k) - f(x_*) < \frac{\|x_0 - x_*\|^2}{\omega}\frac{1}{k} = O\left(\frac{1}{k}\right)$$

③ $f(x)$ is strongly convex with $\mu > 0$, $\eta \in (0, \frac{2}{L+\mu}]$

$$\|x_k - x^*\| \le c^k \|x_0 - x^*\| = O(c^k)$$

$$f(x_k) - f(x_*) \le c^{2k}\|x_0 - x^*\|^2 \quad c = \sqrt{1 - \frac{2\eta\mu L}{\mu + L}}$$

④ Polyak-Lojasiewicz (PL) inequality $\quad \eta = \frac{1}{L}$

$$\text{(PL)} \quad \frac{1}{2}\|\nabla f(\mathbf{x})\|^2 \ge \mu(f(\mathbf{x}) - f(\mathbf{x}_*)), \quad \mu > 0$$

$$f(x_k) - f(x_*) = \left(1 - \frac{\mu}{L}\right)^k [f(x_0) - f(x_*)]$$

1) Strong Convexity $\Rightarrow$ PL inequality

2) $f(x) = \frac{1}{2}\|Ax - b\|^2$ may not be strongly convex but it satisfies (PL)

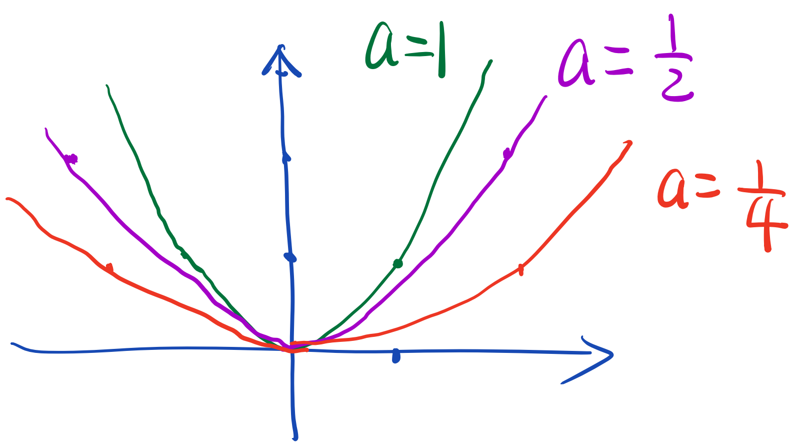
$$\nabla^2 f = A^T A = \square\ \square \succeq \mu I, \ \mu = 0$$

Remark: Larger μ $\Rightarrow$ smaller $c = \sqrt{1 - 2\eta \frac{L}{1 + L/\mu}}$

$\Rightarrow$ faster convergence

Example: $f(x) = ax^2 \quad x \in \mathbb{R}, \ a > 0$

$$\mu = 2a$$



Some Excercise :

① ∇f is L-Continuous $\not\Rightarrow$ $f(x)$ is L-Continuous

$f(x) = x^2$ is NOT L-Cont

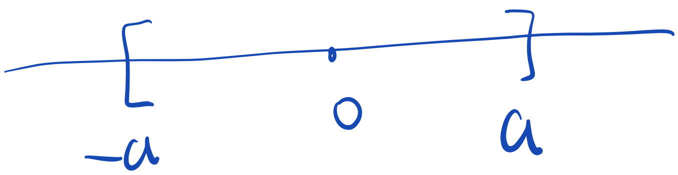
$f'(x) = 2x$ is L-Cont. $\Leftarrow$ $f''(x) = 2$ is bounded

② Even if $\nabla f(x)$ is not L-continuous on $\mathbb{R}$,

Convergence Theorems we proved may still apply

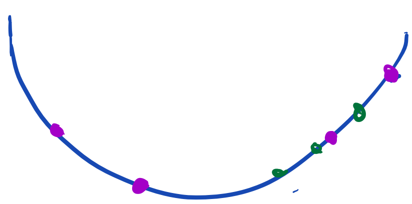
$f(x) = x^4$ is convex

$f'(x) = 4x^3$ is not L-Cont



$x_0 = a$

$x_{k+1} = x_k - \eta \nabla f(x_k)$



$f''(x)$ is bounded on $[-a, a]$

$\Rightarrow f'(x)$ is L-Cont. on $[-a, a]$

③ If ∇f is L-continuous with parameter $L > 0$

$f(x)$ is strongly convex with $\mu > 0$,

then $L \geq \mu$.

$$\langle \nabla f(x) - \nabla f(y), x-y \rangle \geq \mu \|x-y\|^2$$

$$\langle \nabla f(x) - \nabla f(y), x-y \rangle \leq \|\nabla f(x) - \nabla f(y)\| \cdot \|x-y\| \leq L \|x-y\|^2$$

$\frac{L}{\mu}$ is called condition number of $f(x)$

Example: $f(x) = \frac{1}{2} x^T K x - x^T b$

If ∇f is L-continuous with parameter $L > 0$

$f(x)$ is convex, then

$$\langle \nabla f(x) - \nabla f(y), x-y \rangle \geq \frac{1}{L} \|\nabla f(x) - \nabla f(y)\|^2 \quad ①$$

② is ① applied to $\phi(x) = f(x) - \frac{\mu}{2}\|x\|^2$

$$\langle \nabla f(x) - \nabla f(y), x-y \rangle \geq \frac{\mu L}{\mu + L} \|x-y\|^2 + \frac{1}{\mu + L} \|\nabla f(x) - \nabla f(y)\|^2 \quad ②$$

for a strongly convex function?

Case I: If $L = \mu$,

$$\langle \nabla f(x) - \nabla f(y), x-y \rangle \geq \frac{L}{2} \|x-y\|^2 + \frac{1}{2L} \|\nabla f(x) - \nabla f(y)\|^2$$

$\left.\begin{array}{l}\text{Convexity} \\ L\text{-cont. } \nabla f\end{array}\right\} \Rightarrow \langle \nabla f(x) - \nabla f(y), x-y \rangle \geq \frac{1}{L} \|\nabla f(x) - \nabla f(y)\|^2$

strong convexity $\Rightarrow \langle \nabla f(x) - \nabla f(y), x-y \rangle \geq \mu \|x-y\|^2$

Case II: If $L > \mu$, define $\phi(x) = f(x) - \frac{\mu}{2}\|x\|^2$.

Then $\left\{\begin{array}{l} 1) \ \phi(x) \text{ is convex} \\ 2) \ \nabla \phi(x) = \nabla f(x) - \mu x \\ \quad \text{is } L\text{-continuous with } (L-\mu). \end{array}\right.$

$$0 \leq \langle \nabla \phi(x) - \nabla \phi(y), x-y \rangle = \langle \nabla f(x) - \nabla f(y), x-y \rangle - \mu \|x-y\|^2 \leq (L-\mu)\|x-y\|^2$$

Lemma $f(x)$ is convex and $\nabla f(x)$ is L-cont. with L

$$\Leftrightarrow \quad 0 \leq \langle \nabla f(x) - \nabla f(y), x-y \rangle \leq L\|x-y\|^2$$

Proof: "$\Rightarrow$"

Convexity $\Leftrightarrow 0 \leq \langle \nabla f(x) - \nabla f(y), x-y \rangle$

C-S $\Rightarrow \langle \nabla f(x) - \nabla f(y), x-y \rangle \leq \|\nabla f(x) - \nabla f(y)\| \cdot \|x-y\|$