$\boxed{\text{Convergence of } \{x_k\} \text{ for convex } f(x)}$

Theorem If ∇f is L-continuous and $f(x)$ is convex:

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

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

Proof: Pick any x_0, then for fixed x_0

define $\phi(x) = f(x) - \langle \nabla f(x_0), x \rangle$

So $\phi(x)$ is convex

$\nabla \phi(x) = \nabla f(x) - \nabla f(x_0) \Rightarrow \nabla \phi(x_0) = 0$

$\nabla \phi(y) = \nabla f(y) - \nabla f(x_0)$ $\Rightarrow x_0$ minimize $\phi(x)$

$\Rightarrow \|\nabla \phi(x) - \nabla \phi(y)\| = \|\nabla f(x) - \nabla f(y)\| \leq L\|x-y\|$

$\Rightarrow \nabla \phi(x)$ is L-continuous

Descent Lemma on $\phi(x)$:

$\phi(x) \leq \phi(y) + \langle \nabla \phi(y), y-x \rangle + \frac{L}{2}\|x-y\|^2$

$$\underbrace{\min_{x \in \mathbb{R}^n} \phi(x)}_{= \phi(x_0)} \leq \min_{x \in \mathbb{R}^n} \left[\phi(y) + \langle \nabla \phi(y), y-x \rangle + \frac{L}{2}\|x-y\|^2 \right]$$

$$\leq \min_{\text{some } x} \left[\phi(y) + \langle \nabla \phi(y), y-x \rangle + \frac{L}{2}\|x-y\|^2 \right]$$

$$x = y - r\frac{\nabla\phi(y)}{\|\nabla\phi(y)\|} \qquad (r = \|x-y\|)$$

$$= \min_{r\geq 0}\left[\phi(y) - \|\nabla\phi(y)\|r + \frac{L}{2}r^2\right]$$

$$\left(r = \frac{\|\nabla\phi(y)\|}{L}\right) = \phi(y) - \frac{1}{2L}\|\nabla\phi(y)\|^2$$

So $\phi(x_0) \leq \phi(y) - \frac{1}{2L}\|\nabla\phi(y)\|^2$

$$f(x_0) - \langle\nabla f(x_0), x_0\rangle \leq f(y) - \langle\nabla f(x_0), y\rangle - \frac{1}{2L}\|\nabla f(y) - \nabla f(x_0)\|^2$$

$$\underset{(a)}{\Rightarrow} f(x) + \langle\nabla f(x), y-x\rangle + \frac{1}{2L}\|\nabla f(x) - \nabla f(y)\|^2 \leq f(y) \quad ①$$

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

Add two inequalities $\Rightarrow$

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

<u>Theorem</u> Assume ∇f is L-continuous.

Assume $f(x) \geq f(x_*)$, $\forall x \in \mathbb{R}^n$

Assume $f(x)$ is convex.

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

where $\eta \in (0, \frac{2}{L})$ is a constant

$$f(x_k)-f(x_*) \le \frac{1}{\frac{1}{f(x_0)-f(x_*)}+k\omega\frac{1}{\|x_0-x_*\|^2}} < \frac{\|x_0-x_*\|^2}{k\omega}$$

$\omega = \eta(1-\frac{L}{2}\eta)$

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

$$r_{k+1}^2 = \|x_{k+1} - x_*\|^2$$

$$= \|x_k - \eta\nabla f(x_k) - x_*\|^2$$

$\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\|^2 \le L\langle \nabla f(\mathbf{x}) - \nabla f(\mathbf{y}), \mathbf{x}-\mathbf{y}\rangle.$

$$= \|x_k - x_* - \eta\nabla f(x_k)\|^2$$

$$= r_k^2 - 2\eta\langle \nabla f(x_k), x_k - x_*\rangle + \eta^2\|\nabla f(x_k)\|^2$$

2

$$= r_k^2 - 2\eta\langle \nabla f(x_k) \underline{-\nabla f(x_*)}, x_k - x_*\rangle + \eta^2\|\nabla f(x_k)\|^2$$

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

$$\le r_k^2 - \eta\left(\frac{2}{L} - \eta\right)\|\nabla f(x_k)\|^2$$

Let $R_k = f(x_k) - f(x_*)$

Convexity $\Rightarrow f(x) \ge f(x_k) + \langle \nabla f(x_k), x - x_k\rangle$

$$\Rightarrow f(x_*) \ge f(x_k) + \langle \nabla f(x_k), x_* - x_k\rangle$$

$$\Rightarrow f(x_k) - f(x_*) \le \langle \nabla f(x_k), x_k - x_*\rangle$$

$$\leq \|\nabla f(x_k)\| \cdot \|x_k - x_*\|$$

$$\Rightarrow \quad R_k \leq \|\nabla f(x_k)\| \cdot r_k$$

$$\Rightarrow \quad -\|\nabla f(x_k)\| \leq -\frac{R_k}{r_k}$$

Recall we have $\omega = \eta(1 - \frac{L}{2}\eta)$

$$f(x_{k+1}) \leq f(x_k) - \omega \|\nabla f(x_k)\|^2$$

$$\Rightarrow f(x_{k+1}) - f(x_*) \leq f(x_k) - f(x_*) - \omega \|\nabla f(x_k)\|^2$$

$$0 < R_{k+1} \leq R_k - \frac{\omega}{r_k^2} R_k^2$$

Multiply both sides by $\frac{1}{R_{k+1}} \frac{1}{R_k}$

$$\Rightarrow \quad \frac{1}{R_k} \leq \frac{1}{R_{k+1}} - \frac{\omega}{r_k^2} \frac{R_k}{R_{k+1}}$$

$$\Rightarrow \quad \frac{1}{R_{k+1}} \geq \frac{1}{R_k} + \frac{\omega}{r_k^2} \frac{R_k}{R_{k+1}}$$

$$\geq \frac{1}{R_k} + \frac{\omega}{r_k^2}$$

$r_k \leq r_0$

$\frac{1}{r_k} \geq \frac{1}{r_0}$

Summing up for $k = 0, 1, \ldots, N$

$$\Rightarrow \quad \frac{1}{R_{N+1}} \geq \frac{1}{R_0} + \frac{\omega}{r_0^2}(N+1)$$

$$\Rightarrow \quad R_{N+1} \leq \frac{1}{\frac{1}{R_0} + \frac{\omega}{r_0^2}(N+1)}$$

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

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

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

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 $r_k = \|x_k - x_*\|$

$$r_{k+1}^2 = \|x_{k+1} - x_*\|^2$$
$$= \|x_k - x^* - \eta \nabla f(x_k)\|^2$$

$$= r_k^2 + 2\langle -\eta \nabla f(x_k), x_k - x^* \rangle + \eta^2 \|\nabla f(x_k)\|^2$$

$$= r_k^2 - 2\eta \langle \nabla f(x_k) - \nabla f(x_*), x_k - x^* \rangle + \eta^2 \|\nabla f(x_k)\|^2$$

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

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

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

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

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

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

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

$$\Rightarrow r_k^2 \leq \left(1 - \frac{2\eta \mu L}{\mu + L}\right)^k r_0^2 \qquad \Leftrightarrow 4\mu L \leq (\mu + L)^2$$

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

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

$$\boxed{\begin{array}{l}\text{Descent Lemma} \\ f(y) \le f(x) + \nabla f(x)^T (y-x) + \frac{L}{2}\|x-y\|^2\end{array}}$$

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

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

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

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 $\mu > 0$, $\eta \in (0, \frac{2}{L+\mu}]$

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

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

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

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

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

1) Strong Convexity $\Rightarrow$ PL inequality

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

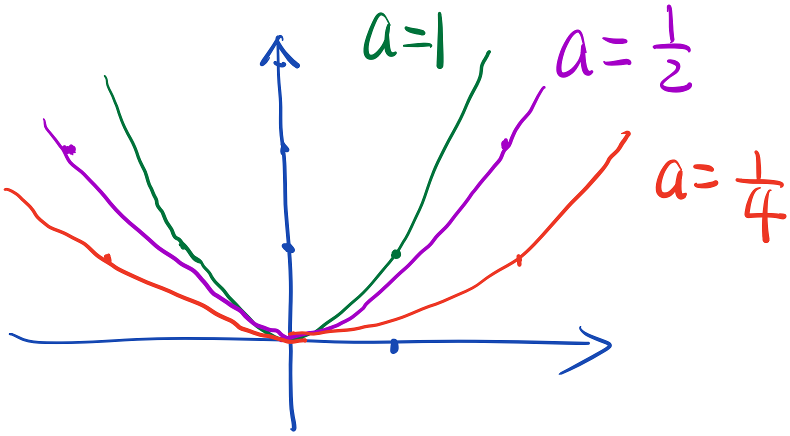
$$\nabla^2 f = A^T A = \square\,\square \succeq \mu I, \ \mu = 0$$

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

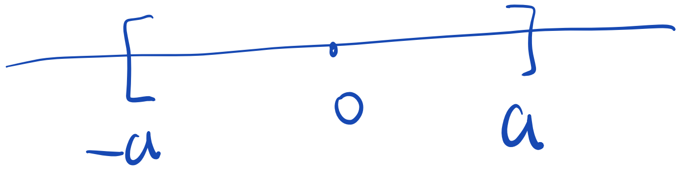
$f'(x) = 2x$ is L-cont. $\Leftarrow$ $f''(x) = 2$ is bounded

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

Convergence Theorems we proved may still apply

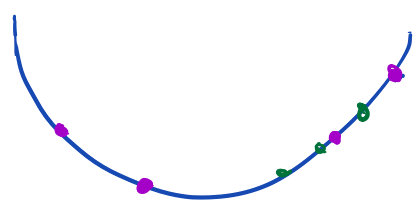
$f(x) = x^4$ is convex

$f'(x) = 4x^3$ 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: $f(x) = \frac{1}{2} x^T K x - x^T b$