Convergence Rate for nonsmooth problems

	Convexity	Strong Convexity
① Subgradient Method	$O(\frac{1}{\sqrt{k}})$	$O(\frac{1}{k})$
② Proximal Point Method	$O(\frac{1}{k})$	$O((1-\mu)^k)$
③ Proximal Gradient	$O(\frac{1}{k})$	$O((1-\mu)^k)$

Proximal Gradient for

$$\min_x f(x) + g(x)$$

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

$\Rightarrow$

1) $g(x) \equiv 0$ is Proximal Point

2) $f(x) \equiv 0$ is Gradient Descent

Theorem $f(x): \mathbb{R}^n \to \mathbb{R}$ is convex $\Leftrightarrow$ $\partial f(x) \neq \phi$, $\forall x$

Remark: $\partial f(y) \neq \phi$, $\forall y$ $\Leftrightarrow$ $\forall x, y$, $\exists g \in \mathbb{R}^n$ s.t.

$$f(x) \geq f(y) + \langle g, x-y \rangle \qquad g \in \partial f(y)$$

Theorem $f(x): \mathbb{R}^n \to \mathbb{R}$ is strongly convex with $\mu > 0$

$\Leftrightarrow$ $\forall x, y$, $\forall g \in \partial f(y)$

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

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

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

$\Leftrightarrow$ $f_\mu(x) = f(x) - \frac{\mu}{2}\|x\|^2$ is convex

$\Leftrightarrow$ $\partial f_\mu(x) = \partial f(x) - \mu x$

$\Leftrightarrow$ $f_\mu(x) \geq f_\mu(y) + \langle g_\mu, x-y\rangle$, $g_\mu \in \partial f_\mu(y)$

$g_\mu = g - \mu y$, $g \in \partial f(y)$

$\Leftrightarrow$ $f_\mu(x) \geq f_\mu(y) + \langle g - \mu y, x-y\rangle$

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

$\mu\|y\|^2 - \mu\langle x, y\rangle$

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

Subgradient Method

$$x_{k+1} = x_k - \eta_k \frac{g_k}{\|g_k\|}, \quad g_k \in \partial f(x_k)$$

Now assume strong convexity with $\mu>0$

$$\|x_{k+1} - x_*\|^2 = \left\|x_k - x_* - \eta_k \frac{g_k}{\|g_k\|}\right\|^2$$

$$= \|x_k - x_*\|^2 - 2\eta_k \frac{1}{\|g_k\|}\langle x_k - x_*, g_k\rangle + \eta_k^2$$

$$f(x\ ) \geq f(x_k) + \langle g_k, x\ - x_k \rangle + \frac{\mu}{2}\|x\ - x_k\|^2$$

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

$$\Rightarrow -\langle g_k, x_k - x_* \rangle \leq f(x_*) - f(x_k) - \frac{\mu}{2}\|x_* - x_k\|^2$$

$$\leq \left(1 - 2\eta_k \frac{1}{\|g_k\|}\frac{\mu}{2}\right)\|x_k - x_*\|^2$$

$$- 2\eta_k \frac{1}{\|g_k\|}[f(x_k) - f(x_*)] + \eta_k^2$$

$$\Rightarrow f(x_k) - f(x_*) \leq \left(\frac{\|g_k\|}{2\eta_k} - \frac{\mu}{2}\right)\|x_k - x_*\|^2 - \frac{\|g_k\|}{2\eta_k}\|x_{k+1} - x_*\|^2 + \frac{\|g_k\|}{2}\eta_k$$

$$\eta_k = \frac{2}{\mu(k+1)} \cdot \|g_k\|$$

$$= \frac{\mu(k-1)}{4}\|x_k - x_*\|^2 - \frac{\mu(k+1)}{4}\|x_{k+1} - x_*\|^2 + \frac{1}{\mu(k+1)}\|g_k\|^2$$

$$k[f(x_k) - f(x_*)] \leq k\frac{\mu(k-1)}{4}\|x_k - x_*\|^2 - k\frac{\mu(k+1)}{4}\|x_{k+1} - x_*\|^2 + k\frac{1}{\mu(k+1)}\|g_k\|^2$$

$k \cdot (k-1) \cdot x_k$ $\quad (k+1) \cdot k \cdot x_{k+1}$

$$\|g_k\| \leq M := \max_{0 \leq k \leq n} \|g_k\|$$

Sum it for $k = 0, 1, \cdots, n$

$$\sum_{k=0}^{n} k[f(x_k) - f(x_*)] \leq -\frac{\mu}{4} n(n+1)\|x_{n+1} - x_*\|^2 + \frac{M^2}{\mu}\sum_{k=0}^{n}\frac{k}{k+1}$$

$$\leq \frac{M^2 n}{\mu} \qquad \bar{x}_n = \underset{0 \leq k \leq n}{\operatorname{argmin}} f(x_k)$$

$$\Rightarrow \Big(\underbrace{\sum_{k=0}^{n} k}_{\frac{n(n+1)}{2}}\Big)\left[f(\bar{x}_n) - f(x_*)\right] \leq \frac{M^2 n}{\mu}$$

$$\Rightarrow \quad f(\bar{x}_n) - f(x_*) \leq \frac{2M^2}{\mu(n+1)} \qquad O\left(\tfrac{1}{k}\right)$$

$$\frac{\mu}{2}\|\bar{x}_k - x_*\|^2 \underbrace{\leq}_{\text{why?}} f(\bar{x}_k) - f(x_*)$$

$$\|\bar{x}_k - x_*\| = O\left(\tfrac{1}{\sqrt{k}}\right)$$

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

$$g \in \partial f(y)$$

Pick $y = x_*$, Pick $g = 0 \in \partial f(x_*)$

$$f(x) \geq f(x_*) + \frac{\mu}{2}\|x - x_*\|^2$$

Convergence Rate of Proximal Gradient for

$$\min_x f(x) + g(x)$$

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

$$= \mathrm{Prox}_f^{\eta}\left[x_k - \nabla g(x_k)\right]$$

Assumptions :

① $f(x)$ is convex

② $g(x)$ is convex

③ $\nabla g(x)$ is L-continuous with L

Theorem (Properties of Prox)

The following are equivalent

① $u = \mathrm{Prox}^{\gamma} f(x)$

② $x - u \in \gamma\, \partial f(u)$

③ $\frac{1}{\gamma}\langle x-u, y-u\rangle \leq f(y) - f(u), \quad \forall y$

$f(y) \geq f(u) + \langle \frac{1}{\gamma}(x-u), y-u\rangle$

Proof: ① $\Leftrightarrow$ ② :

$$\mathrm{Prox}^{\gamma} f(x) = \underset{v}{\mathrm{argmin}}\left[f(v) + \frac{1}{2\gamma}\|v-x\|^2\right]$$

$$u = \mathrm{Prox}^{\gamma} f(x)$$

$\iff \quad 0 \in \partial f(u) + \frac{1}{\gamma}(u-x)$ $\qquad \gamma$

$\iff \quad x-u \in \gamma\, \partial f(u)$

② $\iff$ ③ :

$$g = \frac{1}{\gamma}[x-u] \in \partial f(u)$$

$$\iff \quad f(y) \geq f(u) + \langle g, y-u \rangle$$

$$\iff \quad \frac{1}{\gamma}\langle x-u, y-u \rangle \leq f(y) - f(u)$$

Sufficient Decrease Lemma

Assume
① $f(x)$ is convex
② $g(x)$ is convex
③ $\nabla g(x)$ is L-continuous with L

Let $F(x) = f(x) + g(x)$

and $\bar{x} = \mathrm{Prox}_f^{\eta}(x - \eta \nabla g(x))$, then

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

Proof: $\underbrace{\bar{x}}_{u} = \mathrm{Prox}_f^{\eta}(\underbrace{x - \eta \nabla g(x)}_{x})$

$$\Rightarrow \quad \frac{1}{\eta}\langle \underbrace{x - \eta \nabla g(x)}_{x} - \underbrace{\bar{x}}_{u}, \underbrace{x}_{y} - \underbrace{\bar{x}}_{u} \rangle \leq f(\underbrace{x}_{y}) - f(\underbrace{\bar{x}}_{u})$$

$$u = \mathrm{Prox}_f^{\gamma}(x) \iff \frac{1}{\gamma}\langle x-u, y-u \rangle \leq f(y) - f(u)$$

$$\Rightarrow \langle \nabla g(x), \bar{x} - x \rangle \leq -\frac{1}{\eta}\|x-\bar{x}\|^2 + f(x) - f(\bar{x})$$

$$\Rightarrow ① \; f(\bar{x}) \leq f(x) - \langle \nabla g(x), \bar{x}-x \rangle - \frac{1}{\eta}\|\bar{x}-x\|^2$$

Descent Lemma for $g(x)$

$$② \; g(\bar{x}) \leq g(x) + \langle \nabla g(x), \bar{x}-x \rangle + \frac{L}{2}\|\bar{x}-x\|^2$$

$$①+② \Rightarrow f(\bar{x}) + g(\bar{x}) \leq f(x) + g(x) + \left(\frac{L}{2} - \frac{1}{\eta}\right)\|\bar{x}-x\|^2$$

$$\Rightarrow F(x) - F(\bar{x}) \geq \left(\frac{1}{\eta} - \frac{L}{2}\right)\|\bar{x}-x\|^2$$

$F(x) = f(x) + g(x)$

$\eta < \frac{2}{L}$

$\min f(x)$

$x_{k+1} = \mathrm{Prox}_f^{\eta}(x_k)$