More on proximal operator

$f(x)$ is convex

Def Moreau-Yosida Regularization of $f(x)$ is

$$f_\eta(x) = \min_u \left[f(u) + \frac{1}{2\eta}\|u-x\|^2\right]$$

strong convexity $\Rightarrow$ existence of minimum

$$\mathrm{Prox}_f^\eta(x) = (I+\partial f)^{-1}(x) = \arg\min_u \left[f(u) + \frac{1}{2\eta}\|u-x\|^2\right]$$

- It can be proven that $f_\eta(x)$ is convex & differentiable
- Theorem

① $$\nabla f_\eta(x) = \frac{x - \mathrm{Prox}_f^\eta(x)}{\eta}$$

$$\mathrm{Prox}_f^\eta(x) = x - \eta \nabla f_\eta(x)$$

② ∇f_η is L-continuous with $L = \frac{1}{\eta}$

Proof: ① Let $u = \mathrm{Prox}_f^\eta(x)$ $\quad f_\eta(x) = \min_u \left[f(u) + \frac{\|u-x\|^2}{2\eta}\right]$

$v = \mathrm{Prox}_f^\eta(y)$

$$f_\eta(y) = f(v) + \frac{\|v-y\|^2}{2\eta}$$

$$= f(v) + \frac{\|v-x\|^2}{2\eta} + \left\langle \frac{x-v}{\eta}, y-x \right\rangle + \frac{\|x-y\|^2}{2\eta}$$

$$\geq f(u) + \frac{\|u-x\|^2}{2\eta} + \frac{\|v-u\|^2}{2\eta} + \left\langle \frac{x-v}{\eta}, y-x \right\rangle + \frac{\|x-y\|^2}{2\eta}$$

u is minimizer of $g(w) = f(w) + \frac{1}{2\eta}\|w - x\|^2$ (strongly convex)

$$\Rightarrow \quad g(v) \geq g(u) + \langle 0, v-u \rangle + \frac{1}{2\eta}\|v-u\|^2$$

$$= f(u) + \frac{\|u-x\|^2}{2\eta} + \langle \frac{x-u}{\eta}, y-x \rangle + \frac{\|x-y\|^2}{2\eta}$$

$$+ \langle \frac{u-v}{\eta}, y-x \rangle + \frac{\|u-v\|^2}{2\eta}$$

$$= f_\eta(x) + \langle \frac{x-u}{\eta}, y-x \rangle + \frac{\eta}{2}\left\| \frac{x-y}{\eta} - \frac{u-v}{\eta} \right\|^2$$

$$f_\eta(y) \geq \quad = f_\eta(x) + \langle \frac{x-u}{\eta}, y-x \rangle + \frac{\eta}{2}\left\| \frac{x-u}{\eta} - \frac{y-v}{\eta} \right\|^2$$

$$\Rightarrow \quad f_\eta(y) \geq f_\eta(x) + \langle \frac{x-u}{\eta}, y-x \rangle, \quad \forall x, y$$

$\Rightarrow \frac{x-u}{\eta}$ is a subgradient of $f_\eta(x)$ at x.

$$\frac{x - \mathrm{Prox}_f^\eta(x)}{\eta} \in \partial f_\eta(x) \Rightarrow \nabla f_\eta(x) = \frac{x - \mathrm{Prox}_f^\eta(x)}{\eta}$$

② $$f_\eta(y) \geq f_\eta(x) + \langle \frac{x-u}{\eta}, y-x \rangle + \frac{\eta}{2}\left\| \frac{x-u}{\eta} - \frac{y-v}{\eta} \right\|^2$$

$$f_\eta(x) \geq f_\eta(y) + \langle \frac{y-v}{\eta}, x-y \rangle + \frac{\eta}{2}\left\| \frac{x-u}{\eta} - \frac{y-v}{\eta} \right\|^2$$

$$\Rightarrow \quad \langle \frac{x-u}{\eta} - \frac{y-v}{\eta}, x-y \rangle \geq \eta \left\| \frac{x-u}{\eta} - \frac{y-v}{\eta} \right\|^2$$

Let $G(x) = \dfrac{x - \mathrm{Prox}_f^{\eta}(x)}{\eta} = \nabla f_\eta(x)$

$\|G(x)-G(y)\| \cdot \|x-y\| \geq \langle G(x) - G(y), x-y \rangle \geq \eta \|G(x)-G(y)\|^2$

$\Rightarrow \|G(x) - G(y)\| \leq \frac{1}{\eta}\|x-y\|$

$\Rightarrow G(x)$ is a Lip-continuous function with $L = \frac{1}{\eta}$

Proximal Point Method for $\min_x f(x)$

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

$\Updownarrow$

$$\boxed{\begin{array}{c} x_{k+1} = x_k - \eta \nabla f_\eta(x_k) \\ \text{Gradient Descent for } \min_x f_\eta(x) \end{array}}$$

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

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

$f_\eta(x) = \min_u \left[f(u) + \frac{1}{2\eta}\|u-x\|^2 \right]$

$\min_x f_\eta(x) = \min_{x,u} \left[\underbrace{f(u)}_{f(x_*)} + \underbrace{\frac{1}{2\eta}\|u-x\|^2}_{0} \right] = f_*$

$f_\eta(x)$ has the same minimizer as $f(x)$

Convergence Rate for nonsmooth $f(x)$

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

Convergence Rate for smoother $f_\eta(x)$

$\nabla f_\eta(x)$ is Lip-cont. with $L = \frac{1}{\eta}$

	Convexity	Strong Convexity
Gradient Descent	$O(\frac{1}{k})$	$O\left[\left(1 - \frac{2\eta\mu L}{L+\mu}\right)^k\right]$

Implementing Proximal Point Method without formula for Prox:

for $k = 1, 2, \ldots$

$\boxed{\begin{array}{l}\text{Approximately solve } x_{k+1} = \underset{u}{\operatorname{argmin}}\left[f(u) + \frac{1}{2\eta}\|u - x_k\|^2\right] \\ \text{Use Subgradient (or Accelerated Gradient) for} \\ \quad \underset{u}{\min}\left[f(u) + \frac{1}{2\eta}\|u - x_k\|^2\right]\end{array}}$

for $j = 1, 2, \ldots N$

$$u_{j+1} = u_j - \tau\left[\partial f(u_j) + \frac{1}{\eta}(u_j - x_k)\right]$$

end

$x_{k+1} \approx u_N$

end

Is this better than subgradient method for $\min_x f(x)$?

Theorem [Prox is firmly nonexpansive] $f(x)$ is convex

$$\|\mathrm{Prox}_f^\eta(x) - \mathrm{Prox}_f^\eta(y)\|^2 \leq \langle \mathrm{Prox}_f^\eta(x) - \mathrm{Prox}_f^\eta(y), x - y \rangle$$

It implies $\|\mathrm{Prox}_f^\eta(x) - \mathrm{Prox}_f^\eta(y)\| \leq \|x-y\|$ (nonexpansive)

$\hookrightarrow$ L-cont. with $L=1$

Proof:

$$u = \mathrm{Prox}_f^\eta(x) \Leftrightarrow u = (I+\eta\partial f)^{-1}(x) \Leftrightarrow \frac{x-u}{\eta} \in \partial f(u)$$

$$v = \mathrm{Prox}_f^\eta(y) \Leftrightarrow v = (I+\eta\partial f)^{-1}(y) \Leftrightarrow \frac{y-v}{\eta} \in \partial f(v)$$

$$\left.\begin{array}{l} f(u) \geq f(v) + \langle \partial f(v), u-v\rangle \\ f(v) \geq f(u) + \langle \partial f(u), v-u\rangle \end{array}\right\} \Rightarrow \langle \partial f(u) - \partial f(v), u-v\rangle \geq 0$$

$$\Rightarrow \left\langle \frac{x-u}{\eta} - \frac{y-v}{\eta}, u-v\right\rangle \geq 0$$

$$\Rightarrow \langle x-y, u-v\rangle \geq \|u-v\|^2$$

Def An operator T is

1) Firmly nonexpansive if $\|Tx-Ty\|^2 \leq \langle Tx-Ty, x-y\rangle$

2) Nonexpansive if $\|Tx-Ty\| \leq \|x-y\|$

3) Contraction if $\|Tx-Ty\| < \|x-y\|$

Example: $T(x)$ is L-cont. with $L<1$ is a contraction.

Theorem [Prox is a contraction for strongly convex function]

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

$$(1+\eta\mu)\|\mathrm{Prox}_f^\eta(x) - \mathrm{Prox}_f^\eta(y)\|^2 \leq \langle \mathrm{Prox}_f^\eta(x) - \mathrm{Prox}_f^\eta(y), x-y\rangle$$

It implies $\|\mathrm{Prox}_f^\eta(x) - \mathrm{Prox}_f^\eta(y)\| \leq \frac{1}{1+\eta\mu}\|x-y\|$

$$\Rightarrow \|x_{k+1} - x_*\| \le \frac{1}{1+\eta\mu}\|x_k - x_*\|$$

$$\left(\frac{1}{1+\eta\mu}\right)^2 < \frac{1}{1+2\eta\mu} \Leftrightarrow 1+2\eta\mu < 1+2\eta\mu+\eta^2\mu^2$$

Proof: $u = \mathrm{Prox}_f^\eta(x) \Leftrightarrow u = (I+\eta\partial f)^{-1}(x) \Leftrightarrow \frac{x-u}{\eta} \in \partial f(u)$

$v = \mathrm{Prox}_f^\eta(y) \Leftrightarrow v = (I+\eta\partial f)^{-1}(y) \Leftrightarrow \frac{y-v}{\eta} \in \partial f(v)$

$$f(u) \ge f(v) + \langle \partial f(v), u-v \rangle + \frac{\mu}{2}\|u-v\|^2$$

$$f(v) \ge f(u) + \langle \partial f(u), v-u \rangle + \frac{\mu}{2}\|u-v\|^2$$

$$\Rightarrow \langle \partial f(u) - \partial f(v), u-v \rangle \ge \mu\|u-v\|^2$$

$$\Rightarrow \left\langle \frac{x-u}{\eta} - \frac{y-v}{\eta}, u-v \right\rangle \ge \mu\|u-v\|^2$$

$$\Rightarrow \langle x-y, u-v \rangle \ge [1+\eta\mu]\|u-v\|^2$$

Convergence Rate for nonsmooth problems

	Convexity	Strong Convexity
① Subgradient Method	$O(\frac{1}{\sqrt{k}})$	$O(\frac{1}{k})$
② Proximal Point Method	$O(\frac{1}{k})$	$O\left(\frac{1}{[1+2\eta\mu]^2}\right)^k$

$\boxed{\text{Fixed Point Iteration}}$

$T(x_*) = x_*$

Theorem (Browder-Göhde-Kirk)

$T: \mathbb{R}^n \to \mathbb{R}^n$ is nonexpansive $\Rightarrow$ T has at least one fixed point.

$T(x_*) = x_*$.

If only assume T is nonexpansive, $x_{k+1} = T(x_k)$ may NOT converge.

Example: $T(x) = -x$, then $\|Tx - Ty\| = \|x - y\|$

Want to show convergence of $\boxed{x_{k+1} = \theta x_k + (1-\theta) T(x_k), \quad \theta \in (0,1)}$

Example: $x_{k+1} = \theta x_k + (1-\theta) \operatorname{Prox}_f^{\eta}(x_k)$, $\theta \in (0,1)$

$\boxed{\text{Theorem}}$ Assume $T: \mathbb{R}^n \to \mathbb{R}^n$ is nonexpansive,

then T has at least one fixed point.

$x_{k+1} = \theta x_k + (1-\theta) T(x_k)$, $\theta \in (0,1)$ satisfies

1) $\{x_k\}$ converges to one fixed point y of T

2) $\|x_{k+1} - x_k\|^2 \leq \frac{1}{(k+1)} \left(\frac{1}{\theta} - 1\right) \|x_0 - y\|^2$

This is not error!

Proof: ① $x_{k+1} - x_* = \theta[x_k - x_*] + (1-\theta)[T(x_k) - x_*]$

$$\|x_{k+1} - x_*\|^2 = \|\theta[x_k - x_*] + (1-\theta)[T(x_k) - x_*]\|^2$$

$$\|\theta a + (1-\theta) b\|^2 = \theta\|a\|^2 + (1-\theta)\|b\|^2 - \theta(1-\theta)\|a-b\|^2$$

$$= \theta\|x_k - x_*\|^2 + (1-\theta)\|T(x_k) - x_*\|^2$$

$$- \theta(1-\theta)\|T(x_k) - x_k\|^2$$

$$\|T(x_k) - x_*\| = \|T(x_k) - T(x_*)\| \leq \|x_k - x_*\|$$

$$\leq \theta\|x_k - x_*\|^2 + (1-\theta)\|x_k - x_*\|^2$$

$$- \theta(1-\theta)\|T(x_k) - x_k\|^2$$

$$= \|x_k - x_*\|^2 - \theta(1-\theta)\|T(x_k) - x_k\|^2$$

$$x_{k+1} = S(x_k) := \theta x_k + (1-\theta)T(x_k)$$

$$S(x) - x = (1-\theta)[T(x) - x]$$

$$\|x_{k+1} - x_*\|^2 \leq \;\; = \|x_k - x_*\|^2 - \frac{\theta}{1-\theta}\|S(x_k) - x_k\|^2$$

② We first get $\|x_{k+1} - x_*\|^2 \leq \|x_k - x_*\|^2$

Sum it up

$$\sum_{k=0}^{n} \|S(x_n) - x_n\|^2 \leq \frac{1-\theta}{\theta}\left[\|x_0 - x_*\|^2 - \|x_{n+1} - x_*\|^2\right]$$

$$\sum_{k=0}^{n} \|x_{n+1} - x_n\|^2 \leq \frac{1-\theta}{\theta}\left[\|x_0 - x_*\|^2 - \|x_{n+1} - x_*\|^2\right]$$

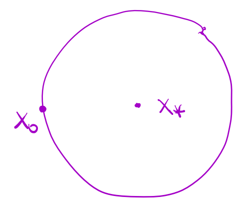
$$\|x_{n+1} - x_n\| = \|S(x_n) - S(x_{n-1})\|$$
$$\leq \theta \|x_n - x_{n-1}\| + (1-\theta)\|Tx_n - Tx_{n-1}\|$$
$$\leq \|x_n - x_{n-1}\|$$

$$\Rightarrow \quad (n+1)\, \|x_{n+1} - x_n\|^2 \leq \frac{1-\theta}{\theta} \|x_0 - x_*\|^2$$

③ $$\|S(x) - x_*\| = \|S(x) - S(x_*)\| \leq \|x - x_*\|$$

$$\Rightarrow \quad \|x_{k+1} - x_*\|^2 \leq \|x_k - x_*\|^2$$

$\Rightarrow$ $\{x_k\}$ is in the ball centered at x_* with radius $\|x_0 - x_*\|$



$$\boxed{\text{Real Analysis} \quad \text{Bounded Sequence in } \mathbb{R}^n \text{ has a convergent subsequence}}$$

$$\Rightarrow \quad x_{k_j} \to y_*, \quad j \to \infty$$

$$\|x_{k+1} - x_k\|^2 \leq \frac{1}{k+1} \frac{1-\theta}{\theta} \|x_0 - x_*\|^2$$

$$\Rightarrow \|x_{k+1} - x_k\|^2 \to 0$$

$$\Rightarrow \|S(x_k) - x_k\| \to 0$$

$$\Rightarrow (1-\theta)\|T(x_k) - x_k\| \to 0$$

$$\Rightarrow \|T(x_{k_j}) - x_{k_j}\| \to 0$$

$T(x)-x$ is continuous because $\|T(x)-x-T(y)+y\| \leq 2\|x-y\|$

$\Rightarrow \|T(y_*)-y_*\| = 0$

$\Rightarrow y_* = T(y_*) \Rightarrow y_* = S(y_*)$

$$\boxed{\begin{aligned} &\|S(x)-x_*\| = \|S(x)-S(x_*)\| \leq \|x-x_*\| \\ \Rightarrow\ &\|x_{k+1}-x_*\|^2 \leq \|x_k-x_*\|^2 \end{aligned}}$$

$$\Downarrow$$

$$\begin{cases} \|x_{k+1}-y_*\|^2 \leq \|x_k-y_*\|^2 \\ x_{k_j} \to y_* \end{cases}$$

$\Rightarrow \{x_k\} \to y_*$.

Theorem $T: \mathbb{R}^n \to \mathbb{R}^n$ is an operator

$I: \mathbb{R}^n \to \mathbb{R}^n$ is the identity operator

The following are equivalent:

① T is firmly nonexpansive

② $I-T$ is firmly nonexpansive

③ $2T - I$ is nonexpansive

④ $\|Tx - Ty\|^2 + \|(I-T)x - (I-T)y\|^2 \leq \|x-y\|^2$

Proof: ① ⟺ 2 : $\|T(x) - T(y)\|^2 \leq \langle T(x) - T(y), x-y \rangle$

$$\|[x - T(x)] - [y - T(y)]\|^2$$

$$= \|x-y\|^2 + \|T(x) - T(y)\|^2 - 2\langle T(x) - T(y), x-y\rangle$$

$$\leq \|x-y\|^2 - \langle T(x) - T(y), x-y\rangle$$

$$= \langle [x - T(x)] - [y - T(y)], x-y \rangle$$

① ⟺ ③ $\quad R = 2T - I$

$$\|T(x) - T(y)\|^2 \leq \langle T(x) - T(y), x-y\rangle$$

$$\Updownarrow$$

$$\|R(x) - R(y)\|^2 = \|2(T(x) - T(y)) - (x-y)\|^2$$

$$= 4\|T(x) - T(y)\|^2 + \|x-y\|^2 - 4\langle T(x) - T(y), x-y\rangle$$

③ $\leq \|x-y\|^2$

① ⟺ ④

$$\|Tx - Ty\|^2 + \|(I-T)x - (I-T)y\|^2 \leq \|x-y\|^2$$

$$\Longleftrightarrow \quad 2\|Tx - Ty\|^2 + \|x-y\|^2 - 2\langle x-y, Tx - Ty\rangle \leq \|x-y\|^2$$

$\iff \|T(x) - T(y)\|^2 \leq \langle T(x) - T(y), x - y \rangle$

Exercise: What other operator is also firmly nonexpansive?

2.2.3 Convergence for convex functions

Theorem 2.8. *Assume $\nabla f(\mathbf{x})$ is Lipschitz-continuous with Lipschitz constant L and $f(\mathbf{x}) : \mathbb{R}^n \longrightarrow \mathbb{R}$ is convex. Then for any $\mathbf{x}, \mathbf{y}$:*

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