

More on proximal operator $f(x)$ is convex

Def Moreau-Yosida Regularization of $f(x)$ is
a.k.a. Moreau Envelope $f_\eta(x) = \min_u \left[f(u) + \frac{1}{2\eta} \|u-x\|^2 \right]$ strong convexity
 $\Rightarrow$ existence of minimum

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

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

- Theorem

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

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

$\textcircled{2} \nabla f_\eta$ is L -continuous with $L = \frac{1}{\eta}$

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

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

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

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

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

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

$$\begin{aligned} &= f(u) + \frac{\|u - x\|^2}{2\eta} + \left\langle \frac{x - u}{\eta}, y - x \right\rangle + \frac{\|x - y\|^2}{2\eta} \\ &\quad + \left\langle \frac{u - v}{\eta}, y - x \right\rangle + \frac{\|u - v\|^2}{2\eta} \end{aligned}$$

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

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

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

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

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

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

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

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

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

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

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

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

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



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

Gradient Descent for $\min_x f_\eta(x)$

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

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

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

$$\min_x f_\eta(x) = \min_{x, u} \left[\underbrace{f(u)}_{f(x_*)} + \underbrace{\frac{1}{2\eta} \|u - x\|^2}_0 \right] = f_*$$

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

Convergence Rate for nonsmooth $f(x)$

① Subgradient Method

Convexity

$$O\left(\frac{1}{\sqrt{k}}\right)$$

Strong Convexity

$$O\left(\frac{1}{k}\right)$$

② Proximal Point Method

$$O\left(\frac{1}{k}\right)$$

$$O\left(\left(\frac{1}{1+2\eta\mu}\right)^k\right)$$

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 $k=1, 2, \dots$

Approximately solve $x_{k+1} = \arg\min_u \left[f(u) + \frac{1}{2\eta} \|u - x_k\|^2 \right]$

Use Subgradient (or Accelerated Gradient) for
 $\min_u \left[f(u) + \frac{1}{2\eta} \|u - x_k\|^2 \right]$

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

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

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

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

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

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

Proof:

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

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

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

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

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

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 $\mu > 0$

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

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

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

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

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

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

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

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

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

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

Convergence Rate for nonsmooth problems

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

$f(x_k) - f(x_*)$

