Review

① A well defined function

$f(x): \mathbb{R}^n \to \mathbb{R}$ such as $f(x) = \|x\|_1$

- Subdifferential $\partial f(x)$ is set of all subderivatives

  Abuse notation:

  $$f(x) \geq f(x_0) + \langle \partial f(x_0), x - x_0 \rangle$$

- Convexity $\Rightarrow$ Continuity & subderivatives exist

- The proximal operator

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

  Well-defined due to strong convexity

  For $f(x) = |x|$, $\mathrm{Prox}_f^{\eta}(x) = \begin{cases} x - \eta, & x > \eta \\ x + \eta, & x < -\eta \\ 0, & x \in (-\eta, \eta) \end{cases}$

② A proper extended function

$f(x): \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$

indicator function

$$i_S(x) = \begin{cases} 0, & x \in S \\ +\infty, & x \notin S \end{cases}$$

- $\mathrm{dom}(f) = \{x \in \mathbb{R}^n : f(x) \in \mathbb{R}\}$

  $f: \mathrm{dom}(f) \to \mathbb{R}$

- f is closed : dom(f) is a closed set

  S is closed set ⇒ ι_S is a closed function

- Convexity means
  1) dom(f) is convex
  2)
  1. $f(\lambda\mathbf{x}+(1-\lambda)\mathbf{y}) \leq \lambda f(\mathbf{x})+(1-\lambda)f(\mathbf{y}), \quad \forall \mathbf{x},\mathbf{y} \in \mathrm{dom}(f), \lambda \in (0,1).$
  2. $f(\mathbf{x}) \geq f(\mathbf{y}) + \langle \partial f(\mathbf{y}), \mathbf{x}-\mathbf{y}\rangle, \quad \forall \mathbf{x},\mathbf{y} \in \mathrm{dom}(f).$
  3. $\langle \partial f(\mathbf{y}) - \partial f(\mathbf{x}), \mathbf{y}-\mathbf{x}\rangle \geq 0, \quad \forall \mathbf{x},\mathbf{y} \in \mathrm{dom}(f).$

- Convexity ⇒ { continuity in interior of dom(f)
  subderivatives exist in interior of dom(f)

- Strong convexity:

For **a proper closed and convex** *function, the following are equivalent:*

1. *f is μ-strongly convexity*
2. $f(\mathbf{x}) \geq f(\mathbf{y}) + \langle \partial f(\mathbf{y}), \mathbf{x}-\mathbf{y}\rangle + \frac{\mu}{2}\|\mathbf{x}-\mathbf{y}\|^2, \quad \forall \mathbf{x},\mathbf{y} \in \mathrm{dom}(f).$
3. $\langle \partial f(\mathbf{y}) - \partial f(\mathbf{x}), \mathbf{y}-\mathbf{x}\rangle \geq \mu\|\mathbf{x}-\mathbf{y}\|^2, \quad \forall \mathbf{x},\mathbf{y} \in \mathrm{dom}(f).$

- Optimality: x_* minimizes a proper convex function $f: \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$

  $\Longleftrightarrow \quad 0 \in \partial f(x_*)$

Theorem 3.13. *For a* **proper closed** *function $f : \mathbb{R}^n \longrightarrow \mathbb{R} \cup \{+\infty\}$, if it is μ-strongly convex, then it has a unique minimizer $\mathbf{x}_*$.*

Example: $f(x) = i_S + \frac{x^2}{2}$ is strongly convex

if S is any interval but

① $S = (0,1) \Rightarrow f(x)$ is not closed, no minimizer

② $S = [a,b] \Rightarrow f(x)$ is closed, unique minimizer.

- The proximal operator

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

is well-defined only if $f(x)$ is convex & closed

For simplicity, we focus on convex $f: \mathbb{R}^n \to \mathbb{R}$ in class e.g. $\|x\|_1$

But all results hold for convex closed $f: \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ e.g. i_S S is convex & closed

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

Convergence Rate of Proximal Gradient for

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

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

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

Example: $\min_{x \in S} g(x) \Longleftrightarrow \min_x i_S(x) + g(x)$

Projected Gradient

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

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 $u = \text{Prox}^{\gamma}f(x) = \underset{v}{\text{argmin}}\ f(v) + \frac{1}{2\gamma}\|v-x\|^2$

$\Updownarrow$

② $x-u \in \gamma\,\partial f(u)$ $\quad 0 \in \partial f(u) + \frac{1}{\gamma}(u-x)$

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

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

Proof: ① $\Leftrightarrow$ ②:

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

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

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

γ

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

② $\Longleftrightarrow$ ③ :

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

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

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

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

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

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

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

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

Descent Lemma for $g(x)$

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

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

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

$F(x) = f(x) + g(x)$ $\qquad \eta < \frac{2}{L}$

Remark: by setting $g(x) \equiv 0$, we get

Unconditional stability of Proximal Point Method

$$\forall \eta > 0, \quad \begin{cases} \min f(x) \\ x_{k+1} = \mathrm{Prox}_f^{\eta}(x_k) \\ f(x_k) - f(x_{k+1}) \geq \frac{1}{\eta}\|x_k - x_{k+1}\|^2 \end{cases}$$

Theorem [Prox-Grad Inequality]

Assume $\begin{cases} ① \ f(x) \text{ is convex} \\ ② \ \nabla g(x) \text{ is L-Continuous with } L \end{cases}$

Let $\bar{y} = \mathrm{Prox}_f^{\eta}(y - \eta \nabla g(y))$, and $\eta \leq \frac{1}{L}$

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

Proof:

$$\phi(u) = g(y) + \langle \nabla g(y), u-y \rangle + \underbrace{f(u) + \frac{1}{2\eta}\|u-y\|^2}_{h(u)}$$

$\nabla^2 h(u) = \frac{1}{\eta} I$

is strong convex w.r.t. u

$\Rightarrow$ $u^* = \arg\min_u \phi(u)$ satisfies

$$0 \in \nabla g(y) + \partial f(u^*) + \frac{1}{\eta}(u^* - y)$$

$$\Updownarrow$$

$$(I + \eta \partial f)(u^*) \ni y - \eta \nabla g(y)$$

$$\Updownarrow$$

$$u^* = (I + \eta \partial f)^{-1}(y - \eta \nabla g(y))$$

$\Rightarrow$ $\bar{y}$ minimizes $\phi(u)$

$$\Rightarrow \quad \phi(x) - \phi(\bar{y}) \geq \frac{1}{2\eta}\|x - \bar{y}\|^2$$

why?

$\phi(\cdot)$ is strongly convex

$$\Rightarrow \phi(x) \geq \phi(y) + \langle g, x-y \rangle + \frac{1}{2\eta}\|x-y\|^2$$

$$\forall g \in \partial\phi(y)$$

$$0 \in \partial\phi(\bar{y})$$

$$\boxed{F(x) - F(\bar{x}) \geq (\tfrac{1}{\eta} - \tfrac{L}{2})\|\bar{x} - x\|^2}$$

Descent Lemma

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

$$\phi(\bar{y}) = g(y) + \langle \nabla g(y), \bar{y} - y \rangle + \frac{1}{2\eta}\|\bar{y} - y\|^2 + f(\bar{y})$$

$$\geq g(\bar{y}) + f(\bar{y}) = F(\bar{y})$$

$$\phi(x) - \phi(\bar{y}) \geq \frac{1}{2\eta}\|x-\bar{y}\|^2$$

$$-F(\bar{y}) \geq -\phi(\bar{y})$$

$$\Rightarrow \quad \phi(x) - F(\bar{y}) \geq \frac{1}{2\eta}\|x-\bar{y}\|^2$$

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

$$\Rightarrow \quad F(x) - F(\bar{y}) \geq \frac{1}{2\eta}\|x-\bar{y}\|^2 - \frac{1}{2\eta}\|x-y\|^2 + g(x) - g(y) - \langle \nabla g(y), x-y \rangle$$

Remark: Let $x = y$, we get $\quad F = f + g$

$$F(y) - F(\bar{y}) \geq \frac{1}{2\eta}\|y-\bar{y}\|^2$$

Sufficient Decrease Lemma

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

For convenience, only consider $\eta = \frac{1}{L}$.

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

Proximal Gradient Method

$$\boxed{x_{k+1} = \mathrm{Prox}_f^{\eta}[x_k - \eta \nabla g(x_k)]}$$