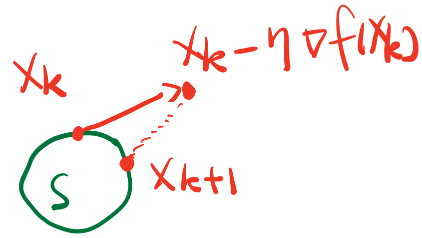
Constrained Minization



$$\begin{cases} \min\limits_{x} f(x) \\ \text{s.t. } x \in S \end{cases}$$

$$\iff \min_{x} \left[f(x) + g(x) \right] \qquad g(x) = i_S(x)$$

Projected Gradient Method

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

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

Forward-Backward Splitting :

Example

$$0 \in \partial f(x_*) + \partial g(x_*)$$

for solving $0 \in A(x) + B(x)$

where A & B are operators

$$x_{k+1} = x_k - \eta \left[A(x_k) + B(x_{k+1}) \right]$$

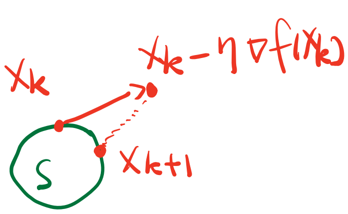
$$(I + \eta B)(x_{k+1}) = x_k - \eta A(x_k)$$

$$x_{k+1} = (I + \eta B)^{-1} (I - \eta A)(x_k)$$

Example : ① Projected Gradient Method

$$\begin{cases} \min\limits_{x} f(x) \\ \text{s.t. } x \in S \end{cases} \iff \min_{x} f(x) + g(x) \qquad g(x) = i_S(x)$$

$$0 \in \nabla f(x) + \partial g(x)$$



$$A(x) = \nabla f(x) , \quad B(x) = \partial g(x)$$

$$x_{k+1} = (I + \eta B)^{-1} (I - \eta A)(x_k)$$

$$= P_S(x_k - \eta \nabla f(x_k))$$

② Proximal Gradient for Composite Optimization

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

$f(x)$ is smooth

$g(x)$ is nonsmooth

$$A(x) = \nabla f(x), \quad B(x) = \partial g(x)$$

$$x_{k+1} = (I + \eta B)^{-1}(I - \eta A)(x_k)$$

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

③ Implicit - Explicit (IMEX) for ODE

$$\frac{d}{dt} u(t) = A(u) + B(u)$$

$$\frac{u^{n+1} - u^n}{\Delta t} = \underbrace{A(u^n)}_{\text{explicit}} + \underbrace{B(u^{n+1})}_{\text{implicit}}$$

Example: $u_t = u_x + u_{xx}$

④ LASSO: $\min_x \left(\frac{\lambda}{2} \|Ax - b\|^2 + \|x\|_1 \right)$

ISTA (Iterative Shrinkage-Thresholding Alg)

$$x_{k+1} = T_\gamma [x_k - \gamma \lambda A^T(Ax_k - b)]$$

$$T_\gamma(x)_i = \begin{cases} x_i - \gamma & , \text{ if } x_i > \gamma \\ x_i + \gamma & , \text{ if } x_i < -\gamma \\ 0 & , x_i \in [-\gamma, \gamma] \end{cases}$$

Shrinkage

$$f(x) = \frac{\lambda}{2} \|Ax - b\|^2 \qquad g(x) = \|x\|_1$$

$$\nabla f(x) = \lambda A^T (Ax - b)$$

$$A(x) = \nabla f(x), \quad B(x) = \partial g(x)$$

$$x_{k+1} = (I + \gamma B)^{-1} (I - \gamma A)(x_k)$$

$$= T_\gamma [x_k - \gamma \lambda A^T (Ax_k - b)]$$

Forward - Backward splitting

- Projected Gradient
- Proximal Gradient
- IMEX
- ISTA

Plan
- More on convexity
- Convergence of subgradient method

Example: $\min_{Ax=b} \|x\|_1 + \frac{\mu}{2}\|x\|^2$

$$\Leftrightarrow \min_{x\in\mathbb{R}^n} \underbrace{\|x\|_1 + \frac{\mu}{2}\|x\|^2 + i_{\{x:Ax=b\}}}$$

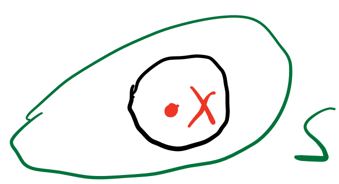
This function is a proper closed μ-strongly convex function

Read Section 3.3 in typed notes for details

A function $f(x)$ is

- extended if $f: \mathbb{R}^n \to \mathbb{R} \cup \{\pm\infty\}$
- proper if $\begin{cases} f: \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\} \\ f(x) \in \mathbb{R} \text{ for at least one } x \end{cases}$
- closed if $\operatorname{dom}(f) = \{x \in \mathbb{R}^n : f(x) \in \mathbb{R}\}$ is a closed set

an open set S means $\forall x \in S$, a ball centered at x is also in S



a closed set S means its complement is open

Ex: $\{x : Ax = b\}$ is closed.

Theorem 3.6. *For an extended real-valued function $f : \mathbb{R}^n \longrightarrow \mathbb{R} \cup \{\pm\infty\}$, the following are equivalent:*

1. *f is a closed function.*
2. *The level set $\{\mathbf{x} : f(\mathbf{x}) \le a\}$ is a closed set for any $a \in$.*
3. *f is* **lower semicontinuous**: *for any $\mathbf{x}$, for any sequence $\mathbf{x}_n \to \mathbf{x}$,*

$$f(\mathbf{x}) \le \liminf_{n\to\infty} f(\mathbf{x}_n).$$

- Convex

Theorem 3.7. *A proper (extended) function $f : \mathbb{R}^n \longrightarrow \mathbb{R} \cup \{+\infty\}$ is convex if and only if* $\mathrm{dom}(f)$ *is convex and*

$$f(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) \le \lambda f(\mathbf{x}) + (1-\lambda) f(\mathbf{y}), \quad \forall \mathbf{x}, \mathbf{y} \in \mathrm{dom}(f), \lambda \in (0,1).$$

- strongly convex

Definition 3.10. *An extended function $f : \mathbb{R}^n \longrightarrow \mathbb{R} \cup \{+\infty\}$ is μ-strongly convex if* $\mathrm{dom}(f)$ *is convex and the following holds for any $\lambda \in (0,1)$:*

$$f(\lambda\mathbf{x}+(1-\lambda)\mathbf{y}) \le \lambda f(\mathbf{x})+(1-\lambda) f(\mathbf{y}) - \frac{\mu}{2}\lambda(1-\lambda)\|\mathbf{x}-\mathbf{y}\|^2, \quad \forall \mathbf{x}, \mathbf{y} \in \mathrm{dom}(f).$$

Theorem 3.10. *An extended function $f : \mathbb{R}^n \longrightarrow \mathbb{R} \cup \{+\infty\}$ is μ-strongly convex if and only if $f(\mathbf{x}) - \frac{\mu}{2}\|\mathbf{x}\|^2$ is convex.*

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

1) i_S is convex if S is convex

2) i_S is closed $\Leftrightarrow$ S is closed

$\Updownarrow$

i_S is l.s.c. (lower semicontinuous)

3) $i_S(X) + \frac{\mu}{2}\|X\|^2$ is strongly convex

if S is a convex set

4) $\|X\|_1 + \frac{\mu}{2}\|X\|^2 + i_{\{X: AX=b\}}$ is strongly convex

Even though $\partial f(\mathbf{x})$ denotes a set, for simplicity, we often abuse the notation by using it to denote any element in this set. Lemma 1.1 can be extended as (see [3, Theorem 5.24] for the proof):

Lemma 3.1. *For a proper function $f(\mathbf{x}) : \mathbb{R}^n \to (-\infty, +\infty]$,* **assume** dom$(f)$ **is convex**, *then the following are equivalent:*

1. $f(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) \leq \lambda f(\mathbf{x}) + (1-\lambda) f(\mathbf{y}), \quad \forall \mathbf{x}, \mathbf{y} \in \text{dom}(f), \lambda \in (0, 1).$

2. $f(\mathbf{x}) \geq f(\mathbf{y}) + \langle \partial f(\mathbf{y}), \mathbf{x} - \mathbf{y} \rangle, \quad \forall \mathbf{x}, \mathbf{y} \in \text{dom}(f).$

3. $\langle \partial f(\mathbf{y}) - \partial f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle \geq 0, \quad \forall \mathbf{x}, \mathbf{y} \in \text{dom}(f).$

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

1. *f is μ-strongly convexity*

2. $f(\mathbf{x}) \geq f(\mathbf{y}) + \langle \partial f(\mathbf{y}), \mathbf{x} - \mathbf{y} \rangle + \frac{\mu}{2}\|\mathbf{x} - \mathbf{y}\|^2, \quad \forall \mathbf{x}, \mathbf{y} \in \text{dom}(f).$

3. $\langle \partial f(\mathbf{y}) - \partial f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle \geq \mu\|\mathbf{x} - \mathbf{y}\|^2, \quad \forall \mathbf{x}, \mathbf{y} \in \text{dom}(f).$

An example is that $\iota_S + \frac{\mu}{2}\|\mathbf{x}\|^2$ is μ-strongly convex for a convex set S.

Subdifferentials

Theorem 3.9. *Consider a proper extended function* $f : \mathbb{R}^n \longrightarrow (-\infty, +\infty]$. **Assume its domain** $\mathrm{dom}(f)$ **is a convex set**, *then*

1. *Existence of subderivatives at any* $\mathbf{x} \in \mathrm{dom}(f)$ *implies convexity of* $f(\mathbf{x})$.
2. *Convexity of* f *implies that subderivative exists at any* $\mathbf{x}$ *in the interior of* $\mathrm{dom}(f)$, *denoted as* $\mathrm{int}(\mathrm{dom}(f))$ *and* $\partial f(\mathbf{x})$ *is bounded:*
$$\forall \mathbf{v} \in \partial f(\mathbf{x}), \|\mathbf{v}\| \leq C \quad \textit{for some} \quad C.$$
3. *If* $U \subset \mathrm{int}(\mathrm{dom}(f))$ *is a nonempty compact set (bounded and closed set in* $\mathbb{R}^n$*), then convexity of* f *implies* $\bigcup_{\mathbf{x} \in U} \partial f(\mathbf{x})$ *is bounded (all subderivatives in* U *have a uniform bound).*
4. *For boundary points of* $\mathrm{dom}(f)$ *of a convex function, subderivatives exist at the relative interior of* $\mathrm{dom}(f)$ *but they can be unbounded.*

Optimality

Theorem 3.12. *For a proper convex function* $f : \mathbb{R}^n \longrightarrow \mathbb{R} \cup \{+\infty\}$, $\mathbf{x}_*$ *minimizes* $f(\mathbf{x})$ *if and only if* $\mathbf{0} \in \partial f(\mathbf{x}_*)$.

Proof. $f(\mathbf{x}) \geq f(\mathbf{x}_*) = f(\mathbf{x}_*) + \langle \mathbf{0}, \mathbf{x} - \mathbf{x}_* \rangle \Leftrightarrow \mathbf{0} \in \partial(\mathbf{x}_*), \quad \forall \mathbf{x} \in \mathrm{dom}(f)$. □

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

If S is a closed set, then $f(x)$ has a unique x_*,

Example: $f(x) = i_S(x) + \frac{1}{2}x^2$ is strongly convex

but not closed for $S = (0,1)$, thus no minimizer

$\|x\|_1 + \frac{\mu}{2}\|x\|^2 + i_{\{x: Ax=b\}}$ is strongly convex and closed

$\Updownarrow$

l.s.c.

For the indicator function ι_S, its subdifferential is given as

$$\partial\iota(\mathbf{x}) = N_S(\mathbf{x}), \quad \mathbf{x} \in S,$$

where N_S is the *normal cone* of S:

Definition 3.11. *For a set $S \subset \mathbb{R}^n$ and $\mathbf{x} \in S$, the normal cone of S at $\mathbf{x}$ is*

$$N_S(\mathbf{x}) = \{\mathbf{y} : \langle \mathbf{y}, \mathbf{z} - \mathbf{x} \rangle \leq 0, \forall \mathbf{z} \in S\},$$

and $N_S(\mathbf{x})$ is an empty set for $\mathbf{x} \notin S$.

$S = \{x \in \mathbb{R}^n : Ax = b\}$ $\quad A \in \mathbb{R}^{m\times n}$

$N_S = \{A^T y : y \in \mathbb{R}^m\}$ because

given $x \in S$, $\forall z \in S$ $\quad \langle A^T y, z - x \rangle$

$= \langle y, A(z-x) \rangle$

$= \langle y, b - b \rangle = 0$

3.5.3 Convergence of subgradient method

Theorem 3.14. *For the subgradient method*

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \eta_k \frac{\mathbf{v}_k}{\|\mathbf{v}_k\|}, \quad \textit{some} \quad \mathbf{v}_k \in \partial f(\mathbf{x}_k,)$$

assume $\|\mathbf{v}_k\| \leq M, \forall k$. *Define*

$$\bar{\mathbf{x}}_k = \operatorname*{argmin}_{1 \leq i \leq k} f(\mathbf{x}_k).$$

For a proper convex function $f(\mathbf{x})$ *with a global minimizer* $\mathbf{x}_*$, *the following holds:*

1. *If* $\sum_{k=0}^{\infty} \eta_k = +\infty$ *and* $\sum_{k=0}^{\infty} \eta_k^2 < +\infty$, *then* $f(\bar{\mathbf{x}}_k) - f(\mathbf{x}_*) \to 0$ *as* $k \to \infty$.

2. *If using a step size* $\eta_k \equiv \frac{C}{\sqrt{n+1}}$ *for* $k = 0, 1, \cdots, n$, *then*

$$f(\bar{\mathbf{x}}_k) - f(\mathbf{x}_*) = \mathcal{O}(\frac{1}{\sqrt{k}}).$$

3. *With the Polyak's step size* $\eta_k = \frac{f(\mathbf{x}_k) - f(\mathbf{x}_*)}{\|\mathbf{v}_k\|}$: $f(\bar{\mathbf{x}}_k) - f(\mathbf{x}_*) = \mathcal{O}(\frac{1}{\sqrt{k}})$.

If further assuming $f(\mathbf{x})$ *is strongly convex with* $\mu > 0$, *then with* $\eta_k = \frac{2}{\mu(k+1)}\|\mathbf{v}_k\|$,

4. $f(\bar{\mathbf{x}}_k) - f(\mathbf{x}_*) = \mathcal{O}(\frac{1}{k})$ *and* $\|\mathbf{x}_k - \mathbf{x}_*\| = \mathcal{O}(\frac{1}{\sqrt{k}})$.

Now assume strong convexity with $\mu > 0$

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

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

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

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

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

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

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

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

$$+ \frac{\|g_k\|}{2} \eta_k$$

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

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

$$+ \frac{1}{\mu(k+1)} \|g_k\|^2$$

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

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

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

$$\leq \frac{M^2 n}{\mu}$$

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

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

$$\frac{\mu}{2}\|\bar{x}_k - x_*\|^2 \leq f(\bar{x}_k)-f(x_*)$$

why?

$$\Rightarrow \quad \|\bar{x}_k - x_*\| \leq \frac{2M}{\mu\sqrt{k+1}}$$