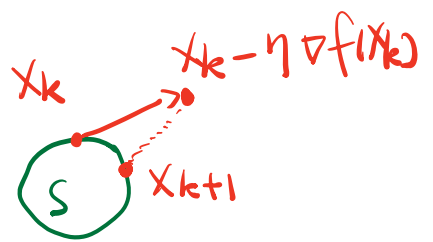


Constrained Minimization

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



$$\Leftrightarrow \min_x [f(x) + g(x)] \quad g(x) = i_S(x) = \begin{cases} 0, & x \in S \\ +\infty, & x \notin S \end{cases}$$

Projected Gradient Method

$$\begin{aligned} x_{k+1} &= \underline{P_S}(x_k - \eta \nabla f(x_k)) \\ &= \underline{(I + \eta \partial g)^{-1}}(x_k - \eta \nabla f(x_k)) \end{aligned}$$

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

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

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

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

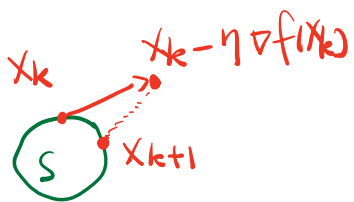
Example: ① Projected Gradient Method

$$\begin{cases} \min_x f(x) \\ \text{s.t. } x \in S \end{cases} \Leftrightarrow \min_x f(x) + g(x) \quad g(x) = i_S(x)$$

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

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

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



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

② Proximal Gradient for Composite Optimization

$$\min_x [f(x) + g(x)] \quad \begin{array}{l} f(x) \text{ is smooth} \\ g(x) \text{ is nonsmooth} \end{array}$$

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

$$\begin{aligned} x_{k+1} &= (I + \eta B)^{-1} (I - \eta A)(x_k) \\ &= (I + \eta \partial g)^{-1} (x_k - \eta \nabla f(x_k)) \end{aligned}$$

③ Implicit - Explicit (IMEX) for ODE

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

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

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

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

ISTA (Iterative Shrinkage-Thresholding Alg)

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

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

$$\begin{aligned} f(x) &= \frac{\lambda}{2} \|Ax - b\|^2 & g(x) &= \|x\| \\ \nabla f(x) &= \lambda A^T(Ax - b) \end{aligned}$$

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

$$\begin{aligned} x_{k+1} &= (I + \gamma B)^{-1} (I - \gamma A)(x_k) \\ &= T_\gamma [x_k - \gamma \lambda A^T(Ax_k - b)] \end{aligned}$$

Forward-Backward splitting

{ Projected Gradient
Proximal Gradient
IMEX
ISTA

$$\text{GD: } x_{k+1} = x_k - \eta \nabla f(x_k) \quad \min_x f(x)$$

$$\text{FE: } x_{k+1} = x_k - \Delta t \nabla f(x_k) \quad x'(t) = -\nabla f(x)$$

- Plan {
- More on convexity
 - Convergence of subgradient method

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

$$\Leftrightarrow \min_{x \in \mathbb{R}^n} \underbrace{\|x\|_1 + \frac{\mu}{2} \|x\|^2 + \mathcal{I}_{\{x: Ax=b\}}(x)}$$

This function is a proper closed μ -strongly convex function

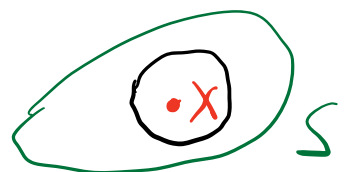
Read Section 4.3 in typed notes for details

4.3	Extended real-valued functions	71
4.3.1	Proper, convex, and closed functions	71
4.3.2	Existence and boundedness of subderivatives	72
4.3.3	Strong convexity	73

A function f is

- extended if $f: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{\pm\infty\}$
- proper if $\begin{cases} f: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\} \\ f(x) \in \mathbb{R} \text{ for at least one } x \end{cases}$
- closed if $\text{dom}(f) = \{x \in \mathbb{R}^n : f(x) \in \mathbb{R}\}$ 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 4.15. For an extended real-valued function $f: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{\pm\infty\}$, the following are equivalent:

1. f is a closed function.
2. The level set $\{x: f(x) \leq a\}$ is a closed set for any $a \in \mathbb{R}$.
3. f is **lower semicontinuous**: for any x , for any sequence $x_k \rightarrow x$,

$$f(x) \leq \liminf_{k \rightarrow \infty} f(x_k).$$

• Convexity of extended functions

Theorem 4.16. A proper (extended) function $f: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ is convex if and only if $\text{dom}(f)$ is convex and

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

Definition 4.17. A proper convex function $f: \mathbb{R}^n \rightarrow (-\infty, +\infty]$ is a function defined for any $x \in \mathbb{R}^n$, not identically $+\infty$, satisfying $\text{dom}(f)$ is convex and

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

Definition 4.21. An extended function $f: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ is μ -strongly convex with a constant $\mu > 0$ if $\text{dom}(f)$ is convex and the following holds for any $\lambda \in (0, 1)$:

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

Theorem 4.22. An extended function $f: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ is μ -strongly convex if and only if $f(x) - \frac{\mu}{2} \|x\|^2$ 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) S is closed $\Leftrightarrow i_S$ is closed $\Rightarrow i_S$ 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 $\partial f(x)$ denotes a set, for simplicity, we often abuse the notation by using it to denote any element in this set. Lemma 1.10 can be extended as (see [8, Theorem 5.24] for the proof):

Lemma 4.23. For a proper function $f(x) : \mathbb{R}^n \rightarrow (-\infty, +\infty]$, assume $\text{dom}(f)$ is convex, and denote $\text{dom}(\partial f) = \{x : \partial f(x) \neq \emptyset\}$ (the set of points where a subderivative exists), then

1. Assume a subderivative exists at every point of $\text{dom}(f)$, i.e., $\text{dom}(\partial f) = \text{dom}(f)$, then f is convex.
2. Assume f is convex, then $\text{int}(\text{dom}(f)) \subset \text{dom}(\partial f)$, but $\text{dom}(\partial f)$ can be smaller than $\text{dom}(f)$: for instance $f(x) = -\sqrt{x}$ for $x \geq 0$ and $f(x) = +\infty$ for $x < 0$ has $\partial f(0) = \emptyset$.

3. If f is convex, then

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

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

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

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

Subdifferential

Theorem 4.19. Consider a proper extended function $f : \mathbb{R}^n \rightarrow (-\infty, +\infty]$. Assume its domain $\text{dom}(f)$ is a convex set, then

1. Existence of subderivatives at any $\mathbf{x} \in \text{dom}(f)$ implies convexity of $f(\mathbf{x})$.
2. Convexity of f implies that subderivative exists at any $\mathbf{x}$ in the interior of $\text{dom}(f)$, denoted as $\text{int}(\text{dom}(f))$ and $\partial f(\mathbf{x})$ is bounded:

$$\forall \mathbf{v} \in \partial f(\mathbf{x}), \|\mathbf{v}\| \leq C \quad \text{for some } C.$$

3. If $U \subset \text{int}(\text{dom}(f))$ 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. If $\mathbf{x}$ is a boundary point of $\text{dom}(f)$, a subderivative may not exist, e.g., $f(x) = -\sqrt{x}$ for $x \geq 0$ and $f(x) = +\infty$ for $x < 0$ has $\partial f(0) = \emptyset$; and even if $\partial f(\mathbf{x})$ is nonempty it can be unbounded, e.g., $\partial \iota_{[0,1]}(0) = (-\infty, 0]$. If f is convex and $\text{dom}(f)$ has an empty interior (e.g., a line in $\mathbb{R}^2$), then subderivatives exist at any point in the relative interior of $\text{dom}(f)$ (the interior of $\text{dom}(f)$ regarded as a subset of the smallest affine set containing it), but $\partial f(\mathbf{x})$ is unbounded there.

Optimality

Theorem 4.25. For a proper convex function $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$, $\mathbf{x}_*$ minimizes $f(\mathbf{x})$ if and only if $\mathbf{0} \in \partial f(\mathbf{x}_*)$.

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

Theorem 4.26. For a proper closed function $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$, if it is μ -strongly convex with $\mu > 0$, then it has a unique minimizer $\mathbf{x}_*$.

Example: S is a convex set

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

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

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

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

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

For the indicator function ι_S , its subdifferential is given as

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

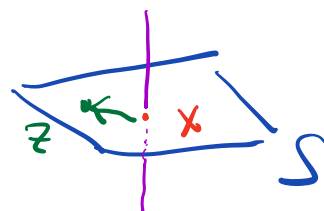
$\Updownarrow$
C.S.C.

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 $\mathbf{x} \notin S$.



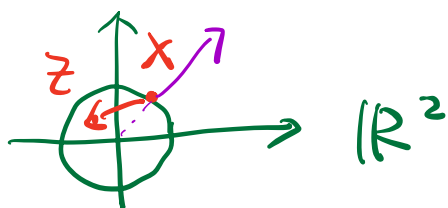
$$S = \{x \in \mathbb{R}^n : Ax = b\} \quad A \in \mathbb{R}^{m \times n}$$

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

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

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

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



5.1.1 Convergence of subgradient method

Theorem 5.1. *For the subgradient method*

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

which is (5.1) with the step size $\frac{\eta_k}{\|\mathbf{v}_k\|}$, written so that η_k is the length of each step, **assume the boundedness of subderivatives** $\|\mathbf{v}_k\| \leq M, \forall k$. Define

$$f(\bar{\mathbf{x}}_n) = \min_{0 \leq i \leq n} f(\mathbf{x}_i),$$

i.e., $\bar{\mathbf{x}}_n$ is the best iterate among $\mathbf{x}_0, \dots, \mathbf{x}_n$. For a proper convex function $f(\mathbf{x})$ with at least one 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 $f(\bar{\mathbf{x}}_n) - f(\mathbf{x}_*) \rightarrow 0$ as $n \rightarrow \infty$.

2. With a step size $\eta_k \equiv \frac{C}{\sqrt{n+1}}$ for $k = 0, 1, \dots, n$, then

$$f(\bar{\mathbf{x}}_n) - f(\mathbf{x}_*) \leq M \frac{\|\mathbf{x}_0 - \mathbf{x}_*\|^2 + C^2}{2C\sqrt{n+1}} = \mathcal{O}\left(\frac{1}{\sqrt{n}}\right).$$

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

$$\|\mathbf{x}_{k+1} - \mathbf{x}_*\| \leq \|\mathbf{x}_k - \mathbf{x}_*\|, \quad f(\bar{\mathbf{x}}_n) - f(\mathbf{x}_*) \leq \frac{M}{\sqrt{n+1}} \|\mathbf{x}_0 - \mathbf{x}_*\|.$$

If further assuming $f(\mathbf{x})$ is μ -strongly convex and **closed (or equivalently lower semicontinuous)**, then with the step size rule $\eta_k = \frac{2}{\mu(k+1)} \|\mathbf{v}_k\|$,

$$4. \quad f(\bar{\mathbf{x}}_n) - f(\mathbf{x}_*) \leq \frac{2M^2}{\mu} \frac{1}{n+1} \quad \text{and} \quad \|\bar{\mathbf{x}}_n - \mathbf{x}_*\| \leq \frac{2M}{\mu} \frac{1}{\sqrt{n+1}}.$$

Now assume strong convexity with $\mu > 0$

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

$$f(x) \geq f(x_*) + \underbrace{\langle \nabla f(x_*), x - x_* \rangle}_{\text{OE}} + \frac{\mu}{2} \|x - x_*\|^2$$