

Review

Def $f(x) : \mathbb{R}^n \rightarrow \mathbb{R}$ is convex if

$$f[(1-a)x + ay] \leq (1-a)f(x) + af(y), \quad 0 < a < 1$$

Jensen's Inequality

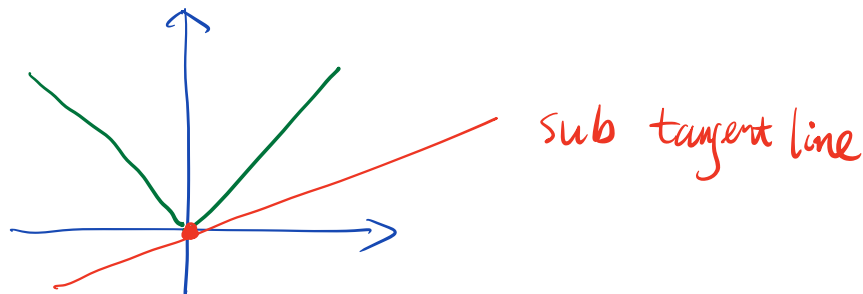
Theorem If $f(x) : \mathbb{R}^n \rightarrow \mathbb{R}$ is convex,
then $f(x)$ is continuous on $\mathbb{R}^n$

Subgradients

Definition: For $f : \mathbb{R}^n \rightarrow \mathbb{R}$, a vector $g \in \mathbb{R}^n$ is
a subgradient of $f(x)$ at x if

$$f(y) \geq f(x) + \langle g, y - x \rangle, \quad \forall y \in \mathbb{R}^n$$

Example: $f(x) = |x|$
 $g \in [-1, 1]$ at $x = 0$



Def: The set of all subgradients of $f(x)$ at x
is called the subdifferential, denoted by $\partial f(x)$.

$$\partial f(x) = \{ g \in \mathbb{R}^n : f(y) \geq f(x) + \langle g, y - x \rangle, \quad \forall y \}$$

Theorem $f(x) : \mathbb{R}^n \rightarrow \mathbb{R}$

$f(x)$ is convex $\Leftrightarrow \partial f(x)$ is not empty at any x

Def Indicator function of a closed set S

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

S is convex
 $\Rightarrow i_S$ is convex

Not a function from $\mathbb{R}^n$ to $\mathbb{R}$

$\begin{cases} \textcircled{1} i_S(x) \text{ is a proper convex function } \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\} \\ \textcircled{2} \text{ domain } S \text{ is convex \& closed} \end{cases}$

$\Rightarrow$ existence of subderivative $i'_S(x)$ on $x \in S$

Example: $\begin{cases} \min_{x \in \mathbb{R}^n} \|x\|_1 \\ \text{s.t. } Ax = b \end{cases} (*)$

$$S = \{x \in \mathbb{R}^n : Ax = b\}$$

or $S = \{x : Ax = b\}$

$(*)$ is equivalent to

$$\min_{x \in \mathbb{R}^n} \|x\|_1 + i_{\{x : Ax = b\}}(x) (**)$$

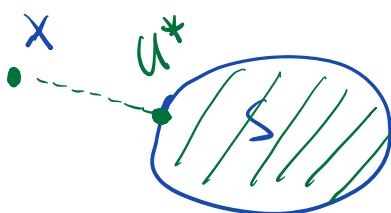
$$\begin{matrix} n \\ \boxed{A} \end{matrix} \begin{matrix} m \\ \boxed{x} \end{matrix} = \begin{matrix} m \\ \boxed{b} \end{matrix}$$

Theorem The proximal operator of $i_S(x)$ is the projection to S

Proof: $f(x) = i_S(x)$

$$\begin{aligned} \text{Prox}_f^{\gamma}(x) &= \arg\min_u \left[f(u) + \frac{1}{2\gamma} \|u - x\|^2 \right] \\ &= \arg\min_u \left[i_S(u) + \frac{1}{2\gamma} \|u - x\|^2 \right] \end{aligned}$$

The minimizer is projection of x onto S



$P_S(x)$

Assume $f: \mathbb{R}^n \rightarrow \mathbb{R}$ is convex

Then convexity $\Rightarrow \begin{cases} \textcircled{1} f(x) \text{ is continuous} \\ \textcircled{2} \partial f(x) \text{ is nonempty at any } x. \end{cases}$

① Consider $f: \mathbb{R} \rightarrow \mathbb{R}$.

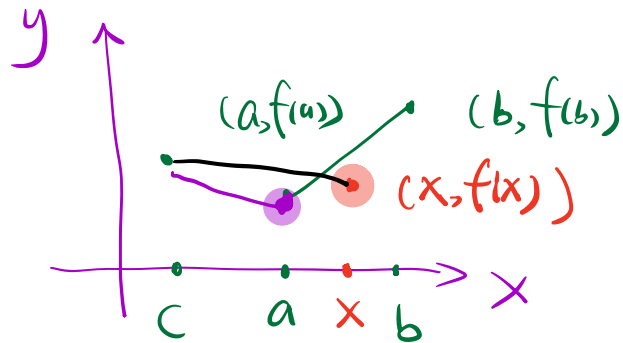
Show convexity $\Rightarrow$ continuity

Proof: For any a , want to show

$$|f(x) - f(a)| \rightarrow 0 \text{ as } x \rightarrow a$$

Two cases:

1) $x > a$



$f(x)$ is below green line segment

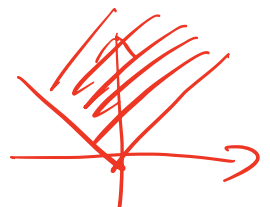
$f(a)$ is below black line segment

$$\Rightarrow \begin{cases} \frac{f(x) - f(a)}{x - a} \leq \frac{f(b) - f(a)}{b - a} \\ \frac{f(x) - f(a)}{x - a} \geq \frac{f(c) - f(a)}{c - a} \end{cases}$$

$$(x-a) \frac{f(c) - f(a)}{c - a} \leq f(x) - f(a) \leq (x-a) \frac{f(b) - f(a)}{b - a}$$

$$x-a \rightarrow 0 \Rightarrow f(x) \rightarrow f(a).$$

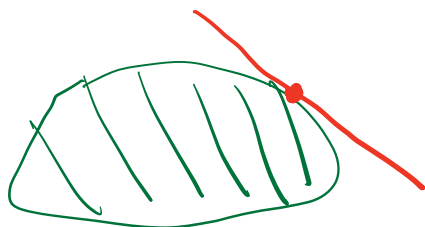
2) $x < a$ is similar



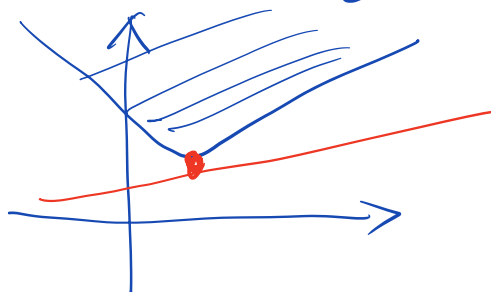
② Show convexity $\Rightarrow$ subgradient exists at any x .

Sketchy proof: 1) The epigraph of $f(x)$ is a convex set

2) Any convex set has a supporting plane



3) The supporting plane of the epigraph is a sub-tangent line



Now consider $\min_{x \in \mathbb{R}^n} f(x)$

where $\begin{cases} f(x) \text{ is convex on } \mathbb{R}^n \\ f(x) \geq f(x_*) \end{cases}$

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

Two simple algorithms

① Subgradient method

$$x_{k+1} = x_k - \eta_k g_k, \quad g_k \in \partial f(x_k)$$

② Proximal Point Method

$$x_{k+1} = x_k - \eta_k g_{k+1}, \quad g_{k+1} \in \partial f(x_{k+1})$$

$$\Leftrightarrow (I + \eta_k \partial f)(x_{k+1}) = x_k$$

$$\Leftrightarrow x_{k+1} = (I + \eta_k \partial f)^{-1}(x_k)$$

$$\Leftrightarrow x_{k+1} = \text{Prox}_f^{\eta_k}(x_k)$$

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

No closed formula for Proximal Operator in general

Example: $f(x) = \|x\|_1$

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

③ Proximal Gradient for Composite Optimization

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

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

This is also called $\min_x \frac{\lambda}{2} \|Ax - b\|^2 + \|x\|_1$

1) Forward-Backward splitting

2) Proximal Gradient

3) For ODE, this is Implicit-Explicit (IMEX) method

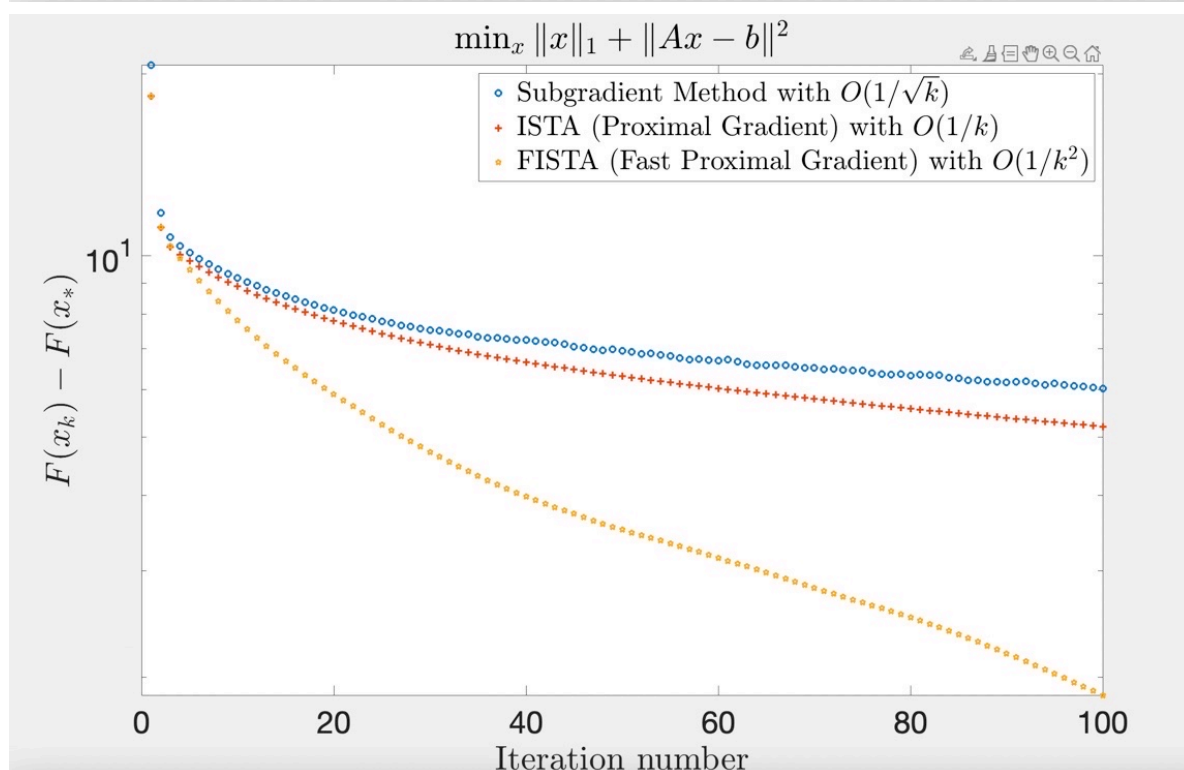
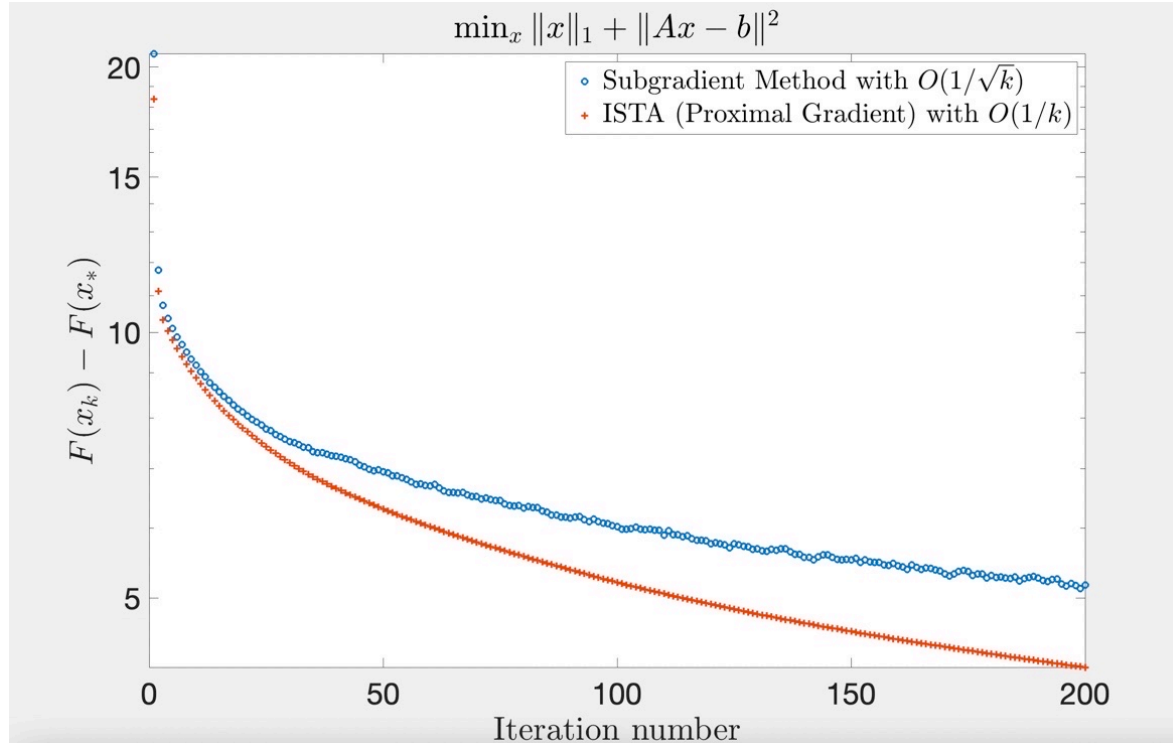
Convergence Rate for nonsmooth problems

	Convexity	Strong Convexity
Subgradient Method	$O(\frac{1}{\sqrt{k}})$	$O(\frac{1}{k})$
Proximal Point Method	$O(\frac{1}{k})$	$O((1-\mu)^k)$
Proximal Gradient	$O(\frac{1}{k})$	$O((1-\mu)^k)$
Fast/Accelerated Proximal Gradient	$O(\frac{1}{k^2})$	$O((1-\sqrt{\mu})^k)$

Convergence Rate for Smooth Convex Problems

$\nabla f(x)$ is L -cont.

	Convex	Strongly Convex
Gradient Descent	$O(\frac{1}{k})$	$\frac{L}{2} \left(\frac{L/\mu - 1}{L/\mu + 1} \right)^k$
Accelerated Gradient Method	$O(\frac{1}{k^2})$	



① Subgradient Method

$$x_{k+1} = x_k - \eta_k \frac{g_k}{\|g_k\|}, \quad g_k \in \partial f(x_k)$$

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

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

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

$$f(x) \geq f(X_k) + \langle g_k, x - X_k \rangle$$

Assume $\|g_k\| \leq M$, let $f(\bar{X}_k) = \min_{1 \leq i \leq k} f(X_k)$

Sum it for $1, 2, \dots, n$

$$\Rightarrow \|X_{n+1} - X_*\|^2 \leq \|X_0 - X_*\|^2 - \frac{2}{M} \left(\sum_{k=0}^n \eta_k \right) [f(\bar{X}_n) - f(X_*)] + \sum_{k=0}^n \eta_k^2$$

$$\Rightarrow f(\bar{X}_n) - f(X_*) \leq M \frac{\sum_{k=0}^n \eta_k^2 + \|X_0 - X_*\|^2}{2 \sum_{k=0}^n \eta_k}$$

1) So for convergence, we can use $\begin{cases} \sum_{k=0}^{\infty} \eta_k = +\infty \\ \sum_{k=0}^{\infty} \eta_k^2 < +\infty \end{cases} \quad \eta_k = \frac{1}{k}$

2) Optimal strategy: $\eta_k = \frac{C}{\sqrt{n+1}}$ for $k=0, \dots, n$

$$\Rightarrow f(\bar{X}_n) - f(x_*) \leq M \frac{C \sum_{k=0}^n \frac{1}{n+1} + \|x_0 - x_*\|^2}{2C \sum_{k=0}^n \frac{1}{\sqrt{n+1}}}$$

$$M = \max_{0 \leq k \leq n} \|g_k\| = M \frac{C + \|x_0 - x_*\|^2}{2C \sqrt{n+1}}$$

$$3) \text{ Polyak's step size } \eta_k = \frac{f(x_k) - f(x_*)}{\|g_k\|^2}$$

$$\begin{aligned} \|x_{k+1} - x_*\|^2 &\leq \|x_k - x_*\|^2 - 2\eta_k \frac{1}{\|g_k\|} [f(x_k) - f(x_*)] + \eta_k^2 \\ &= \|x_k - x_*\|^2 - \frac{|f(x_k) - f(x_*)|^2}{\|g_k\|^2} \\ &\leq \|x_k - x_*\|^2 - \frac{|f(x_k) - f(x_*)|^2}{M^2} \end{aligned}$$

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

$$\begin{aligned} \frac{1}{n^2} \sum_{k=0}^n |f(x_k) - f(x_*)|^2 &\leq \|x_0 - x_*\|^2 - \|x_{n+1} - x_*\|^2 \\ &\leq \|x_0 - x_*\|^2 \end{aligned}$$

$$(n+1) |f(\bar{x}_n) - f(x_*)|^2 \leq \sum_{k=0}^n |f(x_k) - f(x_*)|^2$$

$$\Rightarrow f(\bar{x}_n) - f(x_*) \leq \frac{M}{\sqrt{n+1}} \|x_0 - x_*\|$$

$$M = \max_{0 \leq k \leq n} \|g_k\|$$

Now assume strong convexity with $\mu > 0$

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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