

Part II: Nonsmooth Convex Optimization

A quick preview of what we plan to cover:

① The main examples:

1) "Lasso" (Tibshirani 1996)

Least absolute shrinkage & selection operator

$$\min_{x \in \mathbb{R}^n} \frac{\lambda}{2} \|Ax - b\|^2 + \|x\|_1$$

2) "Basis Pursuit" (Chen, Donoho, Saunders, 1998)

$$\min_{x \in \mathbb{R}^n} \|x\|_1$$

such that $Ax = b$

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

minimizing $\|x\|_1$ gives
a sparse x
(less nonzero entries of x)

② Subgradient

$f(x) = |x|$ does not have a gradient

but it has a subgradient $\partial f(x)$

$$\partial f(x) = \begin{cases} 1 & , x > 0 \\ -1 & , x < 0 \\ [-1, 1] & , x = 0 \end{cases}$$

③ Algorithms & Convergence

Subgradient descent

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

④ Composite optimization for $\min_x f(x) + g(x)$

1) Forward-Backward Splitting

option for reading: F-B-F splitting

2) ADMM

3) Douglas-Rachford

More on Convex Functions

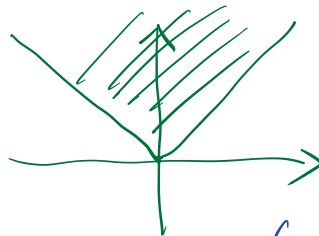
Def $f(x) : \mathbb{R}^n \rightarrow \mathbb{R}$ is convex if

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

Def Epigraph of $f(x)$ is

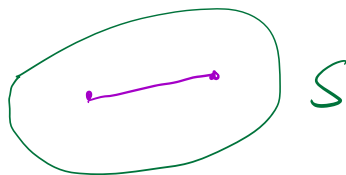
$$\{(x, y) \in \mathbb{R}^n \times \mathbb{R} : f(x) \leq y\}$$

$$f(x) = |x|$$



Def A set S is convex if

$$\forall x, y \in S, (1-a)x + ay \in S, \quad \forall a \in (0, 1)$$



Theorem For $f: \mathbb{R}^n \rightarrow \mathbb{R}$,

$f(x)$ is convex \Leftrightarrow Epigraph of $f(x)$ is convex.

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

and $f_i(x): \mathbb{R}^n \rightarrow \mathbb{R}$ are convex on \mathbb{R}^n ,
then $i=1, 2, \dots, N$

① $g(x) = f(Ax+b)$ is convex

② $g(x) = \sum_{i=1}^N a_i f_i(x)$ is convex, $a_i \geq 0$

③ $g(x) = \max_i f_i(x)$ is convex

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

then $f(x)$ is $\left\{ \begin{array}{l} \text{locally Lip-continuous} \\ \text{continuous on } \mathbb{R}^n \end{array} \right.$

$\forall x_0 \in \mathbb{R}^n$, there is a ball centered at x_0
with a radius $\delta > 0$

$$B_\delta(x_0) = \{x \in \mathbb{R}^n, \|x - x_0\| < \delta\}$$

s.t. $\forall x \in B_\delta(x_0), \exists L$ s.t.

$$|f(x) - f(x_0)| \leq L \|x - x_0\|$$



Subgradients

Definition: For $f: \mathbb{R}^n \rightarrow \mathbb{R}$, a vector $g \in \mathbb{R}^n$ is

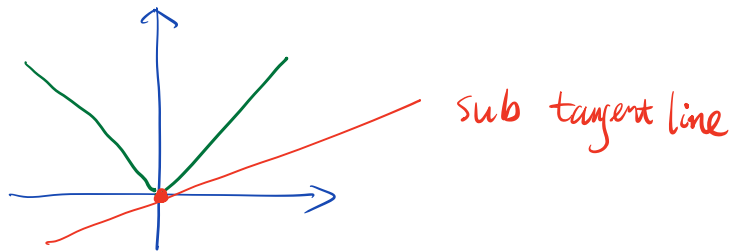
a subgradient of $f(x)$ at x if

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

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

at $x=0$

$$g \in [-1, 1]$$



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

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

$$f(x) = |x| \quad \partial f(x) = \{ 1 \}, \quad x > 0$$

$$\partial f(x) = \{ -1 \}, \quad x < 0$$

$$\partial f(0) = [-1, 1]$$

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

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

Proof: " \Leftarrow " $z = (1-a)x + ay, \quad a \in (0, 1)$

$$g \in \partial f(z)$$

$$f(x) \geq f(z) + \langle g, x-z \rangle = f(z) - a \langle g, y-x \rangle$$

$$f(y) \geq f(z) + \langle g, y-z \rangle = f(z) + (1-a) \langle g, y-x \rangle$$

$$\begin{aligned}
 (1-a)f(x) + af(y) &\geq (1-a)f(z) + af(z) \\
 &= f(z) \\
 &= f((1-a)x + ay)
 \end{aligned}$$

" \Rightarrow " skipped.

Example: $f(x) = \|x\|_1 = \sum_{i=1}^n |x_i|$

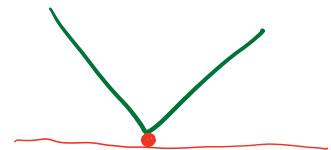
$$g \in \partial f(x) \Leftrightarrow g_i \begin{cases} = 1 & x_i > 0 \\ = -1 & x_i < 0 \\ \in [-1, 1] & x_i = 0 \end{cases}$$

$x \in \mathbb{R}^n$
 $g \in \mathbb{R}^n$

Theorem

$$x_* \text{ minimizes } f(x) \Leftrightarrow \vec{0} \in \partial f(x_*)$$

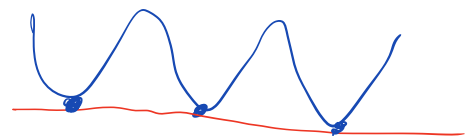
Proof: x_* is a minimizer



$$\begin{aligned}
 f(x) &\geq f(x_*) \\
 &= f(x_*) + \langle \vec{0}, x - x_* \rangle
 \end{aligned}$$



$$\vec{0} \in \partial f(x_*)$$



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

① $\nabla f(x)$ exists $\Rightarrow \partial f(x) = \{\nabla f(x)\}$

② $\partial f(x)$ has only one element $\Rightarrow \nabla f(x)$ exists

Theorem: ① $\partial [af](x) = a[\partial f(x)]$

② $\partial [f+g] = \partial f + \partial g$

Two simple Algorithms for $\min_x f(x)$ $x_{k+1} = x_k - \eta \nabla f(x_k)$

$0 \in \partial f(x_*)$ with a convex non-differentiable $f(x)$:

① Subgradient Method (Forward Euler for ODE)

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

② Proximal Point Method (Backward Euler)

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

Need to solve:

$$x_{k+1} + \eta g_{k+1} = x_k$$

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

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

Proximal Operator

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

Claim $\text{Prox}_{\gamma f}(x) = (I + \gamma \partial f)^{-1}(x)$

Proof: $\begin{cases} \frac{1}{2} \|u-x\|^2 \text{ is convex w.r.t. } u \\ \eta f(u) \text{ is convex} \end{cases}$

$\Rightarrow \eta f(u) + \frac{1}{2} \|u-x\|^2 \text{ is convex}$

Let $u_* = \underset{u}{\text{argmin}} \left[\eta f(u) + \frac{1}{2} \|u-x\|^2 \right]$

Then $0 \in \partial[\eta f](u_*) + (u_* - x)$

$\Rightarrow x \in u_* + \eta \cdot \partial f(u_*)$

$\Rightarrow u_* = (I + \eta \partial f)^{-1}(x)$

If there is another y s.t.

$x \in y + \eta \partial f(y)$

then $0 \in \partial[\eta f](y) + (y-x)$

$\Rightarrow y$ minimizes $g(u) = \eta f(u) + \frac{1}{2} \|u-x\|^2$

$g(u)$ is strongly convex (because $\frac{1}{2} \|u-x\|^2$ is)

\Rightarrow unique minimizer $\Rightarrow y = u^*$.

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

$$(I + \delta \partial f)^{-1}(x) = \begin{cases} x - \delta, & \text{if } x > \delta \\ x + \delta, & \text{if } x < -\delta \\ 0, & \text{if } x \in [-\delta, \delta] \end{cases}$$

Shrinkage operator