高级工程数学

Chaomin Shen

Dept of Computer Science, East China Normal University, Shanghai, China

2021-11-4

Outline

- Convex functions
  - Definition
  - Properties
- Subgradient and subdifferential
  - Definition
  - Properties
  - Examples
  - Applications of subgradient/subdifferential
    - Shrinkage operator

Convex functions

dom

$$\text{dom}\, f := \{\boldsymbol{x} \in \mathbb{R}^n : |f(\boldsymbol{x})| < \infty\}$$

Convex set

$C \in \mathbb{R}^n$ is called convex, if $\alpha \boldsymbol{x} + (1-\alpha)\boldsymbol{y} \in C$ for $\forall \boldsymbol{x}, \boldsymbol{y} \in C$, $\forall \alpha \in [0,1]$

Convex function

$f : \mathbb{R}^n \to \mathbb{R}$ is convex, if dom f is convex and

$$f(\alpha \boldsymbol{x} + (1-\alpha)\boldsymbol{y}) \leq \alpha f(\boldsymbol{x}) + (1-\alpha) f(\boldsymbol{y})$$

for $\forall \boldsymbol{x}, \boldsymbol{y} \in$ dom f, $\forall \alpha \in [0,1]$.

f is concave, if $-f$ is convex.

Properties of convex functions

1. Jensen inequality
   For $\forall \boldsymbol{x}_1, \cdots, \boldsymbol{x}_m \in \operatorname{dom} f$, $\alpha_1, \cdots, \alpha_m > 0$ s.t. $\sum_{i=1}^m \alpha_i = 1$, then

$$f(\sum_{i=1}^m \alpha_i \boldsymbol{x}_i) \leq \sum_{i=1}^m \alpha_i f(\boldsymbol{x}_i)$$

2. f is convex, iff its epigraph is convex
3. f is closed if epi f is a closed set.
4. f is lower semi continuous (l.s.c) at $\boldsymbol{x}$, if

$$f(\boldsymbol{x}) \leq \lim \inf_{k \to \infty} f(\boldsymbol{x}_k) \quad \text{for every } \boldsymbol{x}_k \to \boldsymbol{x}$$

5. If $f_1, \cdots, f_m$ are convex (closed), then $\sum_{i=1}^m \lambda_i f_i$ $(\lambda_i > 0)$ is convex (closed)

Properties of convex functions

6. If f is convex (closed), A is an $m \times n$ matrix, then $f(A\boldsymbol{x})$ is convex (closed)
7. f_i $(i = 1, \cdots, n)$ are convex (closed), then $\sup_i f_i(\boldsymbol{x})$ is convex (closed) on $\prod_{i=1}^{m} \operatorname{dom} f_i$.

Properties of convex functions

8 First order condition for a convex function
Assume that f is diff. dom f is convex. Then f is convex iff

$$f(\boldsymbol{y}) \geq f(\boldsymbol{x}) + \langle \nabla f(\boldsymbol{x}), \boldsymbol{y} - \boldsymbol{x} \rangle \quad \text{for } \forall \boldsymbol{x}, y \in \text{dom } f$$

9 Second order condition
Assume that f is twice diff. dom f is convex. Then, f is convex iff

$$\left(\frac{\partial^2 f}{\partial x_i \partial x_j}\right)_{i,j=1,\cdots,n} \geq 0$$

Sketch of Proof for Property 8

$$f(ty + (1-t)x) \le tf(y) + (1-t)f(x)$$

$$\Longrightarrow \; f(x + t(y-x)) \le f(x) + t(f(y) - f(x))$$

$$\Longrightarrow \; f(x + t(y-x)) - f(x) \le t(f(y) - f(x))$$

$$\Longrightarrow \; \frac{f(x + t(y-x)) - f(x)}{t} \le f(y) - f(x)$$

$$\Longrightarrow \; f(y) \ge f(x) + \frac{f(x + t(y-x)) - f(x)}{t} \qquad (2)$$

Now, let

$$g(t) = f(x + t(y-x))$$

We now express Eq.(2) in terms of g(t), as shown below:

$$f(y) \ge f(x) + \frac{g(t) - g(0)}{t} \qquad (3)$$

Properties of convex functions

10 A continuously diff. function f is convex, iff

$$\langle \nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y}), \boldsymbol{x} - \boldsymbol{y} \rangle \geq 0$$

Proof.

$\rightarrow$

$$f(\boldsymbol{x}) \geq f(\boldsymbol{y}) + \langle \nabla f(\boldsymbol{y}), \boldsymbol{x} - \boldsymbol{y} \rangle$$
$$f(\boldsymbol{y}) \geq f(\boldsymbol{x}) + \langle \nabla f(\boldsymbol{x}), \boldsymbol{y} - \boldsymbol{x} \rangle$$

相加即得

$\leftarrow$

....

Subgradient and subdifferential

Subgradient

Let C be a convex open set in $\mathbb{R}^N$, and $f : C \to \mathbb{R}$ is convex. Then $\boldsymbol{g} \in \mathbb{R}^N$ is called *subgradient* of f at $\boldsymbol{x}_0 \in C$, if

$$f(\boldsymbol{x}) \geq f(\boldsymbol{x}_0) + \langle \boldsymbol{g}, \boldsymbol{x} - \boldsymbol{x}_0 \rangle \qquad \forall \boldsymbol{x} \in C$$

回忆

Convex and differentiable

$$f(\boldsymbol{x}) \geq f(\boldsymbol{x}_0) + \langle \nabla f, \boldsymbol{x} - \boldsymbol{x}_0 \rangle$$

Subgradient and subdifferential (Cont'd)

Subgradient is not unique.

Subdifferential

The set of all subgradients of f at $\boldsymbol{x}_0$ is called subdifferential of f at $\boldsymbol{x}_0$, denoted by $\partial f(\boldsymbol{x}_0)$, i.e.,

$$\partial f(\boldsymbol{x}_0) = \{\boldsymbol{g} | f(\boldsymbol{x}) \geq f(\boldsymbol{x}_0) + \langle \boldsymbol{g}, \boldsymbol{x} - \boldsymbol{x}_0 \rangle, \quad \forall \boldsymbol{x} \in C = \mathrm{dom}\, f\}$$

Subdifferential

Properties

- $\partial f(\boldsymbol{x})$ is a closed convex set
- $\langle \boldsymbol{u}-\boldsymbol{v}, \boldsymbol{x}-\boldsymbol{y}\rangle \geq 0 \quad \forall \boldsymbol{u} \in \partial f(\boldsymbol{x}), \boldsymbol{v} \in \partial f(\boldsymbol{y})$
- $\partial f(\boldsymbol{x}) = \{\nabla f(\boldsymbol{x})\}$ if f is diff. at $\boldsymbol{x}$.
- $\boldsymbol{x}^* = \arg\min_{\boldsymbol{x}} f(\boldsymbol{x})$ iff $\mathbf{0} \in \partial f(\boldsymbol{x}^*)$

Examples

a) $f(x) = |x|$, $x \in \mathbb{R}$

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

Examples

b) $f(\boldsymbol{x}) = \|\boldsymbol{x}\|_1 := \sum_{i=1}^n |x_i|, \qquad \boldsymbol{x} \in \mathbb{R}^n$

$$\partial_{x_i} f(\boldsymbol{x}) = \begin{cases} \frac{x_i}{|x_i|} & x_i \neq 0 \\ \{g_i : |g_i| \leq 1\} & x_i = 0 \end{cases}$$

c) $f(\boldsymbol{x}) = \|\boldsymbol{x}\|_2 := (\sum_{i=1}^n x_i^2)^{1/2}$ Euclidean norm

$$\partial f(\boldsymbol{x}) = \begin{cases} \boldsymbol{x}/\|\boldsymbol{x}\|_2 & \boldsymbol{x} \neq \mathbf{0} \\ \{\boldsymbol{g} : \|\boldsymbol{g}\|_2 \leq 1\} & \boldsymbol{x} = \mathbf{0} \end{cases}$$

(因为 $f(\boldsymbol{x}) = \|\boldsymbol{x}\|_2 \geq \langle \boldsymbol{g}, \boldsymbol{x} - \mathbf{0} \rangle$ if $\|\boldsymbol{g}\|_2 \leq 1$)

Shrinkage Operator

a) For given $\boldsymbol{b} \in \mathbb{R}^n$ and $\lambda > 0$, the solution to

$$\min_{\boldsymbol{u} \in \mathbb{R}^n} \lambda \|\boldsymbol{u}\|_2 + \frac{1}{2}\|\boldsymbol{u} - \boldsymbol{b}\|^2$$

is

$$\boldsymbol{u}^* = \max\{\|\boldsymbol{b}\|_2 - \lambda, 0\} \frac{\boldsymbol{b}}{\|\boldsymbol{b}\|_2} := \text{shrink}\,(\boldsymbol{b}, \lambda)$$

Proof.

i) For $\boldsymbol{u} \neq \boldsymbol{0}$, 可导, 得$\lambda \frac{\boldsymbol{u}}{\|\boldsymbol{u}\|_2} + \boldsymbol{u} - \boldsymbol{b} = \boldsymbol{0}$, 整理得

$$\left(\frac{\lambda}{\|\boldsymbol{u}\|_2} + 1\right)\boldsymbol{u} = \boldsymbol{b} \qquad (\boldsymbol{u} \parallel \boldsymbol{b})$$

两边取norm, 得 $\lambda + \|\boldsymbol{u}\|_2 = \|\boldsymbol{b}\|_2$, i.e., $\|\boldsymbol{u}\|_2 = \|\boldsymbol{b}\|_2 - \lambda$, i.e.
$\boldsymbol{u} = (\|\boldsymbol{b}\|_2 - \lambda)\frac{\boldsymbol{b}}{\|\boldsymbol{b}\|_2}$.

ii) $u = 0$, 原式分成2部分, 重点是第一部分

$$\lambda \boldsymbol{g} + \boldsymbol{u} - \boldsymbol{b} = \boldsymbol{0} \qquad \text{for } \|\boldsymbol{g}\| \leq 1 \quad \text{and } \boldsymbol{u} = 0$$

i.e. $\boldsymbol{b} = \lambda \boldsymbol{g}$, thus $\|\boldsymbol{b}\|_2 \leq \lambda$. 所以 $\boldsymbol{u} = \max\{\|\boldsymbol{b}\|_2 - \lambda, 0\}\frac{\boldsymbol{b}}{\|\boldsymbol{b}\|_2}$ $\square$

Shrinkage Operator

b) the solution to

$$\min_{\boldsymbol{u}\in\mathbb{R}^n} \lambda\|\boldsymbol{u}\|_1 + \frac{1}{2}\|\boldsymbol{u}-\boldsymbol{b}\|^2$$

is

$$\boldsymbol{u}^* = \max\{|b_i|-\lambda, 0\}\text{sign}\,(b_i)$$

Proof.

原式化为 $\min_{u_1,\cdots,u_n} \sum_i (|u_i| + \frac{1}{2}|u_i - b_i|^2)$

$$u_i = \max\{|b_i|-\lambda, 0\}\text{sign}\, b_i$$

□