Convex Optimization

Part 2: Subgradient method (1/2)

Namhoon Lee

POSTECH

26 Sep 2022

Subgradient

g is a subgradient of a convex function f at x if

$$f(y) \geq f(x) + g^\top (y - x) \qquad \forall y$$

- If f is differentiable at x, then $g = \nabla f(x)$.
- If f is non-differentiable at x, then there could be multiple g.

Subdifferential

The subdifferential $\partial f(x)$ of f at x is the set of all subgradients

$$\partial f(x) = \{g \mid f(y) \geq f(x) + g^\top (y - x),\ \forall y\}$$

- $\partial f(x)$ is a closed convex set (by def. of convex set).
- When $x \in \operatorname{int} \operatorname{dom} f$, $\partial f(x)$ is nonempty (it could be empty for nonconvex) and bounded (see BV).
- If f is differentiable at x, then $\partial f(x) = \{\nabla f(x)\}$.

Examples

Absolute value

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

Subgradients

- For $x \neq 0$, $g = \text{sign}(x)$.
- For $x = 0$, g is any element in $[-1, 1]$ or $\partial f(x) = \{g \mid g \in [-1, 1]\}$.
  - Check if this satisfies the subgradient definition.

Examples

Euclidean norm (2-norm)

$$f(x) = \|x\|_2$$

Subgradients

- For $x \neq 0$, $g = \frac{1}{\|x\|_2} x$.
- For $x = 0$, $\partial f(x) = \{g \mid \|g\|_2 \leq 1\}$.
  - Why later.

Examples

Taxicab norm (1-norm)

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

Subgradients

- For $x_i \neq 0$, $g_i = \mathrm{sign}(x_i)$.
- For $x_i = 0$, i-th component g_i is any element in $[-1, 1]$.

Examples

Pointwise maximum of convex differentiable f_1, f_2

$$f(x) = \max\{f_1(x), f_2(x)\}$$

Subgradients

- For $f_1(x) > f_2(x)$, $g = \nabla f_1(x)$.
- For $f_1(x) < f_2(x)$, $g = \nabla f_2(x)$.
- For $f_1(x) = f_2(x)$, g is any point on line segment between $\nabla f_1(x)$ and $\nabla f_2(x)$ (*i.e.*, $t\nabla f_1(x) + (1-t)\nabla f_2(x)$ for any $t \in [0,1]$).

Examples

Indicator function over convex set

$$I_C(x) = \begin{cases} 0 & \text{if } x \in C \\ \infty & \text{if } x \notin C \end{cases}$$

Subgradients

- For $x \in C$, $\partial I_C(x) = \mathcal{N}_C(x)$
- Why? By the definitions of subgradient and normal cone

$$I_C(y) \geq I_C(x) + g^\top (y - x) \quad \forall y$$

$$\mathcal{N}_C(x) = \{g \mid g^\top (y - x) \leq 0, \ \forall y \in C\}$$

Examples

Piecewise-linear

$$f(x) = \max_{i=1,\ldots,m} (a_i^\top x + b_i)$$

Subgradients

- the subdifferential at x is a polyhedron

$$\partial f(x) = \operatorname{conv}\{a_i \mid i \in I(x)\}$$

with $I(x) = \{i \mid a_i^\top x + b_i = f(x)\}$

Subgradient calculus

Differentiable functions

If f is differentiable at x, then

$$\partial f(x) = \{\nabla f(x)\}$$

Nonnegative linear combination

If $f(x = \alpha_1 f_1(x) + \alpha_2 f_2(x))$ with $\alpha_1, \alpha_2 \geq 0$, then

$$\partial f(x) = \alpha_1 \partial f_1(x) + \alpha_2 \partial f_2(x)$$

(RHS is set sum)

Affine transformation of variables

if $g(x) = f(Ax + b)$, then

$$\partial g(x) = A^\top \partial f(Ax + b)$$

Pointwise maximum

If $f(x) = \max\{f_1(x), ..., f_m(x)\}$ and define $I(x) = \{i \mid f_i(x) = f(x)\}$, the "active" functions at x, then

$$\partial f(x) = \operatorname{conv} \bigcup_{i \in I(x)} \partial f_i(x)$$

- *i.e.*, the convex hull of the union of subdifferentials of active functions at x
- This extends to pointwise supremum (*i.e.*, m not finite) with some extra conditions.

Norms

If $f(x) = \|x\|_p$ and let q be such that $1/p + 1/q = 1$, then

$$\partial f(x) = \arg\max_{\|y\|_q \leq 1} y^\top x$$

- One way to understand this is via dual norm and from which the calculus rule for pointwise supremum
$$\|x\|_p = \max_{\|y\|_q \leq 1} y^\top x$$
- Check this pictorially as well for example when $p = 2$.

Optimality conditions for unconstrained optimization

For unconstrained optimization

$$\min_x f(x)$$

Optimality condition: x^* minimizes $f(x)$ if

$$0 \in \partial f(x^*)$$

Proof.

This follows directly from the definition of subgradient at x^*

$$f(y) \geq f(x^*) + 0^\top (y - x^*) \qquad \forall y$$

□

Optimality conditions for constrained optimization

For constrained optimization

$$\min_x f(x) \quad \text{subject to} \quad x \in C$$

Optimality condition: x^* minimizes $f(x)$ if

$$0 \in \partial f(x^*) + \mathcal{N}_C(x)$$

Proof.
The proof is done by converting the constraint into a penalty term and applying the optimality condition; or simply in a pictorial form. □

Any questions?