Math 273a: Optimization
Subgradients of convex functions

Made by: Damek Davis
Edited by Wotao Yin

Department of Mathematics, UCLA

Fall 2015

online discussions on piazza.com

Subgradients

Assumptions and Notation

1. $f : \mathbb{R}^n \to \mathbb{R}^n \cup \{\infty\}$ is a closed proper convex function, i.e.

$$\mathbf{epi} f \quad = \quad \{(x,t) \in \mathbb{R}^n \times \mathbb{R} | f(x) \leq t\}$$

   is closed and convex.

2. The effective domain of f is

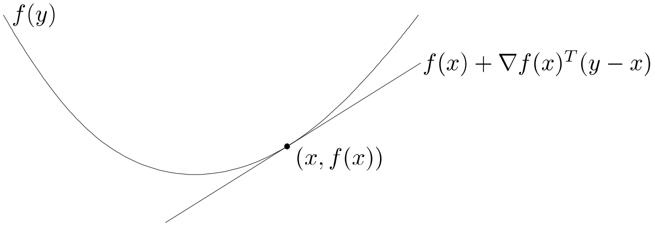
$$\mathrm{dom} f = \{x \in \mathbb{R}^n : f(x) < \infty\}$$

3. The function f is **proper**, i.e. $\mathrm{dom} f \neq \emptyset$.
4. A raised $*$ (e.g., x^*) means **global minimum** of some function.

C^1 convex function

Recall: a convex function of C^1 obeys

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



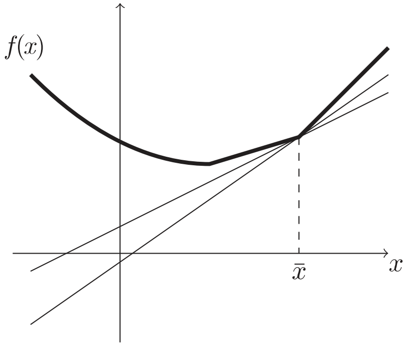
[1]

[1] figure taken from Boyd and Vandenberghe, Convex Optimization.

Non-C^1 convex functions

For all each $\bar{x} \in \mathbb{R}^n$,

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



- **subgradient** is defined via a global property, not by taking limits.

- in contrast, the set of **regular subgradients** is defined as

$$\hat{\partial} f(x) = \{g \mid \exists \delta > 0 \text{ such that } f(y) \geq f(x) + \langle g, y - x \rangle + o(\|y - x\|),\ \forall y \in B_\delta(x)\}$$

- g is a **general subgradient** of f at x if:

there is a sequence $x^i \to x$ and $g^i \in \hat{\partial} f(x^i)$ with $g^i \to g$.

- Reference: Variational analysis by Rockafellar and Wets. Definition 8.3.

Which functions have subgradients?

- If $f \in C^1$, then $\nabla f(x) \in \partial f(x)$.
- In fact, $\partial f(x) = \{\nabla f(x)\}$.

Proof: if g is a subgradient, then for $y \in \mathbb{R}^n$

$$\begin{aligned} \langle \nabla f(x), y \rangle &= \lim_{t \to 0} \frac{f(x+ty) - f(x)}{t} \quad \text{(by } \nabla f \text{ def.)} \\ &\geq \lim_{t \to 0} \frac{\langle g, x+ty-x \rangle}{t} \quad \text{(by subgrad def.)} \\ &= \langle g, y \rangle. \end{aligned}$$

Change y to $-y$, and the inequality still holds. $\implies \langle \nabla f(x), y \rangle = \langle g, y \rangle$.

Plugging in standard basis vectors $\implies \nabla f(x) = g$.

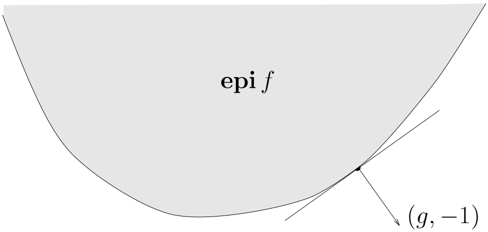
- Next, the general case.

Which functions have subgradients?

Theorem (Nesterov'03 Thm 3.1.13)

*Let f be a **closed convex function and** $x_0 \in \mathrm{int}(\mathrm{dom}(f))$. Then $\partial f(x_0)$ is a nonempty bounded set.*

- Proof uses supporting hyperplanes of epigraph to show existence, and local Lipschitz continuity of convex functions to show boundedness.



[2]

[2] figure taken from Boyd and Vandenberghe,
http://see.stanford.edu/materials/lsocoee364b/01-subgradients_notes.pdf.

The converse

Lemma (Nesterov'03 Lm 3.1.6)

If $\partial f(x) \neq \emptyset$ for all $x \in \mathrm{dom}(f)$, then f is convex.

Proof.

- $x, y \in \mathrm{dom}(f)$, $\alpha \in [0, 1]$, $y_\alpha = (1-\alpha)x + \alpha y = x + \alpha(y - x)$, $g \in \partial f(y_\alpha)$.

$$f(y) \geq f(y_\alpha) + \langle g, y - y_\alpha \rangle = f(y_\alpha) + (1-\alpha)\langle g, y - x \rangle \tag{1}$$

$$f(x) \geq f(y_\alpha) + \langle g, x - y_\alpha \rangle = f(y_\alpha) - \alpha\langle g, y - x \rangle \tag{2}$$

- Multiply equation (1) by α and equation (2) by $(1-\alpha)$.
- Add them together to get

$$\alpha f(y) + (1-\alpha) f(x) \geq f(y_\alpha)$$

□

Technicality: $\mathrm{int}(\mathrm{dom}(f))$

We cannot relax the assumption $x \in \mathrm{int}(\mathrm{dom}(f))$ to $x \in \mathrm{dom}(f)$.

Example:

$$f : [0, +\infty) \to \mathbb{R}. \quad f(x) = -\sqrt{x}.$$

$\mathrm{dom}(f) = [0, +\infty)$ but

$$\partial f(0) = \emptyset.$$

Compute subgradients: general rules

- **Smooth functions:** $\partial f(x) = \{\nabla f(x)\}$.
- **Composition with affine mapping:** $\phi(x) = f(A(x) + b)$

$$\partial \phi(x) = A^T \partial f(A(x) + b).$$

- **Positive sums:** $\alpha, \beta > 0$, $f(x) = \alpha f_1(x) + \beta f_2(x)$.

$$\partial f(x) = \alpha \partial f_1(x) + \beta \partial f_2(x)$$

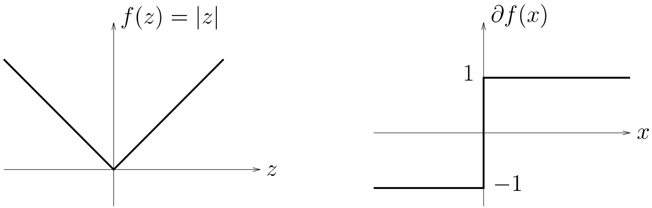
- **Maximums:** $f(x) = \max_{i \in \{1, \cdots, n\}} \{f_i(x)\}$

$$\partial f(x) = \mathrm{conv}\{\partial f_i(x) | f_i(x) = f(x)\}$$

Examples

- $f(x) = |x|$.

$$\partial f(x) \quad = \quad \begin{cases} \{\text{sign}(x)\} & x \neq 0; \\ [-1, 1] & \text{otherwise} \end{cases}$$



[3]

- as seen, ∂f is a *relation* or a *point-to-set* mapping

[3] figure taken from Boyd and Vandenberghe,
http://see.stanford.edu/materials/lsocoee364b/01-subgradients_notes.pdf.

Examples

- $f(x) = \sum_{i=1}^{n} |\langle a_i, x\rangle - b_i|$. Define

$$I_-(x) = \{i | \langle a_i, x\rangle - b_i < 0\}$$
$$I_+(x) = \{i | \langle a_i, x\rangle - b_i > 0\}$$
$$I_0(x) = \{i | \langle a_i, x\rangle - b_i = 0\}.$$

Then

$$\partial f(x) = \sum_{i \in I_+(x)} a_i - \sum_{i \in I_-(x)} a_i + \sum_{i \in I_0(x)} [-a_i, -a_i]$$

(the last sum is the Minkowski sum)

- $\text{minimize}_x \|Ax - b\|_1$ is known as robust fitting, which is more robust to the outliers than the least-squares problem.

- $f(x) = \max_{i \in \{1, \cdots, n\}} x_i$. Then

$$\partial f(x) = \mathrm{conv}\{e_i | x_i = f(x)\}$$
$$\partial f(0) = \mathrm{conv}\Big\{e_i | i \in \{1, \cdots, n\}\Big\}$$

Note: conv denotes the *convex hull*:

$$\mathrm{conv}\{x_i\} := \Big\{ \sum_i \alpha_i x_i : a_i \geq 0, \sum_i \alpha_i = 1 \Big\}$$

Examples

- $f(x) = \|x\|_2$. f is differential away from 0, so:

$$\partial f(x) = \frac{x}{\|x\|_2}, \quad \text{where } x \neq 0.$$

At 0, go back to subgradient equation:

$$\|y\|_2 \quad \geq \quad 0 + \langle g, y - 0 \rangle$$

Thus, $g \in \partial f(0)$, if, and only if, $\frac{\langle g,y \rangle}{\|y\|_2} \leq 1$ for all $y \neq 0$. Thus, g is in the dual ball $B_2^*(0,1) = B_2(0,1)$.

- This is a common pattern!

How to compute subgradients: Examples

- $f(x) = \|x\|_\infty = \max_{i\in\{1,\cdots,n\}} |x^{(i)}|$.

$$\partial f(x) = \mathrm{conv}\{g^i : g^i \in \partial|x^{(i)}|, |x^{(i)}| = f(x)\}, \quad \text{where } x \neq 0.$$

Going back to subgradient equation

$$\|y\|_\infty \quad \geq \quad 0 + \langle g, y\rangle$$

Thus, $g \in \partial f(0)$, if, and only if, $\frac{\langle g,y\rangle}{\|y\|_\infty} \leq 1$ for all $y \neq 0$. Thus, $\partial f(0)$ is the dual ball to the l_∞ norm: $B_1(0,1)$.

Examples

- $f(x) = \|x\|_1 = \sum_{i=1}^{n} |x_i|$. Then

$$\partial f(x) \quad = \quad \sum_{x_i > 0} e_i - \sum_{x_i < 0} e_i + \sum_{x_i = 0} [-e_i, e_i]$$

for all x. Then

$$\partial f(0) = \sum_{i=1}^{n} [-e_i, e_i] = B_\infty(0, 1).$$

Semi-continuity

Definition (upper semi-continuity)

A point-to-set mapping M is upper semi-continuous at x' if any convergent sequence $(x^k, s^k) \to (x', s')$ satisfying $s^k \in M(x^k)$ for each k also obeys

$$s' \in M(x').$$

Interpretation: if (i) $x^k \to x'$ and (ii) for each x^k you can find $s^k \in M(x^k)$ so that $s^k \to s'$, then $s' \in M(x')$.

Lower semi-continuity is essentially the opposite.

Definition (lower semi-continuity, lsc)

A point-to-set mapping M is lower semi-continuous if any convergent sequence $x^k \to x'$ and $s' \in M(x')$, there exists sequence $s^i \in M(x^{k_i})$ such that

$$s^i \to s'.$$

Lemma

Let f be a proper convex function. ∂f is upper semi-continuous, and $\partial f(x)$ is a convex set.

Proof: Take the limit of

$$f(y) \geq f(x^k) + \langle s^k, y - x^k \rangle, \quad s^k \in \partial f(x^k).$$

The second part is a direct result of linearity of $\langle \cdot, y - x \rangle$. □

However, if $f(x) = |x|$, the ∂f is not lower semi-continuous at $x = 0$.

Directional derivative versus Subgradient

Assume that $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is a proper, closed (thus lsc), and convex. Then

1. the directional derivative is well defined for every $d \in \mathbb{R}^n$:

$$f'(x; d) = \lim_{\alpha \downarrow 0} \frac{f(x + \alpha d) - f(x)}{\alpha}.$$

2. the directional derivatives at x bound all the subgradient projections

$$\partial f(x) = \{p \in \mathbb{R}^n : f'(x; d) \geq \langle p, d \rangle, \ \forall \ d \in \mathbb{R}^n\}.$$

3. directional derivatives are *extreme* subgradients:

$$f'(x; d) = \max\{\langle p, d \rangle : p \in \partial f(x)\}.$$

0 subgradient $\implies$ minimum

- Suppose that $0 \in \partial f(x)$.
- If f is smooth and convex, $0 \in \partial f(x) = \{\nabla f(x)\} \implies \nabla f(x) = 0$.
- In general: If $0 \in \partial f(x)$, then

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

for all $y \in \mathbb{R}^n$.

- $\implies x$ is a minimum!
- Converse also true: $f(y) \geq f(x^*) + 0 = f(x^*) + \langle 0, y - x^* \rangle$.