MATH 4501558
FALL 2015

LECTURE 12
NONSMOOTH ANALYSIS
SPECIAL CASES

Convex functions

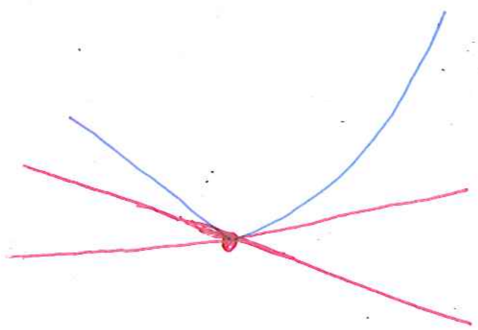
$f: \mathbb{R}^n \longrightarrow \mathbb{R}$ Lipschitz continuous, convex

$$f(\theta \bar{x} + (1-\theta)x) \leqslant \theta f(\bar{x}) + (1-\theta) f(x)$$

$\forall \theta \in [0,1]$ and $\forall x, \bar{x} \in \mathbb{R}^n$

Subdifferential at x

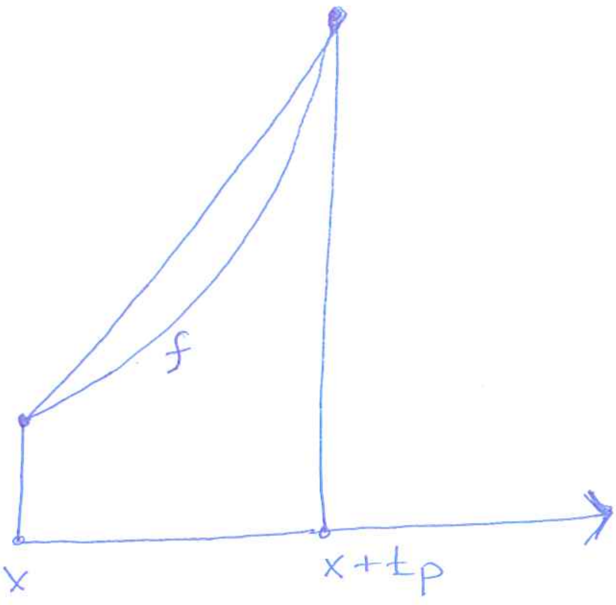
$$\partial f_s(x) := \{ \psi \in \mathbb{R}^n \mid f(\bar{x}) - f(x) \geqslant \psi^T(\bar{x} - x) \quad \forall \bar{x} \in \mathbb{R}^n \}$$



$\partial f_s(x)$ – set of normal vectors of hyperplanes tangent to f at x and "globally" lies underneath f.

Consider the map (for a fixed $p \in \mathbb{R}^n$)

$$s(t) = \frac{f(x+tp) - f(x)}{t}$$



Assume
$\|p\| = 1$

Proposition

$s(t)$ is monotonically nondecreasing.

Proof – exercise

Observe
$$\frac{f(x + t(\bar{x} - x)) - f(x)}{t} \leq f(\bar{x}) - f(x)$$

due to convexity of f. Exploit this.

Lemma

The directional derivative
$$f'(x; p) := \lim_{t \to 0^+} \frac{f(x+tp) - f(x)}{t}$$

exists.

Proof

The map $t \to \frac{f(x+tp) - f(x)}{t}$ is monotonically nonincreasing as $t \to 0^+$. Also, it is bounded below due to Lipschitzness, that is

$$\frac{f(x+tp)-f(x)}{t} \geqslant \frac{[f(x)-tr\|p\|]-f(x)}{t}$$

$$= -r\|p\|.$$

Consequently, the limit must exist. □

THM

$$f^{\circ}(x;p) = f'(x;p)$$

Proof

By definition for each $\delta > 0$

$$f^{\circ}(x;p) = \limsup_{\substack{t\to 0^+ \\ y \to x}} \frac{f(y+tp)-f(y)}{t}$$

$$= \lim_{r\to 0^+} \sup_{\|y-x\| \leqslant \delta r} \underbrace{\sup_{0<t<r} \frac{f(y+tp)-f(y)}{t}}_{\frac{f(y+rp)-f(y)}{r}}$$

(due to monotonicity

By Lipschitzness of f

$$\left| \frac{f(x+rp)-f(x)}{r} - \frac{f(y+rp)-f(y)}{r} \right|$$

$$= \left| \frac{f(x+rp)-f(y+rp)}{r} - \frac{f(x)-f(y)}{r} \right| \leqslant 2r\delta$$

$$\Longrightarrow \frac{f(y+rp)-f(y)}{r} \leqslant \frac{f(x+rp)-f(x)}{r} + 2r\delta$$

Consequently,

$$f^\circ(x;p) \leqslant \lim_{r\to 0^+} \frac{f(x+rp)-f(x)}{r} + 2\gamma\delta$$

$$= f'(x;p) + 2\gamma\delta$$

Since, this is true for all $\delta > 0$,

$$f^\circ(x;p) = f'(x;p).$$

$\square$

Corollary

$$\partial_f(x) = \partial_{f_s}(x)$$

Proof

First suppose $\psi \in \partial_f(x)$. Then

$$f'(x;p) = f^\circ(x;p) \geqslant \psi^T p \quad \forall p.$$

By monotonicity of $t \to \frac{f(x+tp)-f(x)}{t}$

$$f(x+p)-f(x) \geqslant \lim_{t\to 0^+} \frac{f(x+tp)-f(x)}{t}$$

$$= f'(x;p) \geqslant \psi^T p.$$

Now letting $\bar{x} := x+p$,

$$f(\bar{x}) - f(x) \geqslant \psi^T(\bar{x}-x) \quad \forall \bar{x},$$

so $\psi \in \partial_{f_s}(x)$.

Conversely, suppose $\psi \in \partial f_s(x)$. Then

$$f(\bar{x}) - f(x) \geq \psi^T(\bar{x} - x) \quad \forall \bar{x}.$$

Equivalently,

$$f(x+\alpha p) - f(x) \geq \psi^T(\alpha p) \quad \forall p, \forall \alpha > 0$$

$$\Longrightarrow \frac{f(x+\alpha p) - f(x)}{\alpha} \geq \psi^T p \quad \forall p, \forall \alpha > 0.$$

Taking the limit as $\alpha \to 0^+$ yields

$$\lim_{\alpha \to 0^+} \frac{f(x+\alpha p) - f(x)}{\alpha} = f'(x;p)$$

$$= f^\circ(x;p) \geq \psi^T p \quad \forall p,$$

so $\psi \in \partial f(x)$. $\square$

<u>THM</u> (Optimality Condition)

Let $f: \mathbb{R}^n \to \mathbb{R}$ be convex and Lipschitz continuous.

$$x_* \text{ is a global minimizer} \iff 0 \in \partial f_s(x_*)$$

<u>Proof</u>

Suppose $0 \in \partial f_s(x_*)$. By definition

$$f(\bar{x}) - f(x_*) \geq 0 \quad \forall \bar{x},$$

so x_* is a global minimizer.

Conversely, if x_* is a global minimizer, by first order necessary conditions

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

$\square$

Ex (Piece-wise linear functions)

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad f(x)$$

where

$$f(x) := \max \{ l_j(x) \mid j = 1, \ldots, K \}$$

$$l_j(x) := a_j^T x + b_j \qquad j = 1, \ldots, K$$

$a_j \in \mathbb{R}^n$, $b \in \mathbb{R}$ – given

such that $\{a_1, a_2, \ldots, a_K\}$ is orthonormal.

Subdifferential

$$f'(x;p) = \lim_{t \to 0^+} \frac{f(x+tp) - f(x)}{t}$$

$$= \lim_{t \to 0^+} \frac{\left\{ \max\limits_{j \in A(x)} l_j(x+tp) \right\} - f(x)}{t}$$

where

$$A(x) := \{ j \mid l_j(x) = f(x) \}.$$

Equivalently,

$$f'(x;p) = \lim_{t \to 0^+} \max_{j \in A(x)} \frac{l_j(x+tp) - l_j(x)}{t}$$

$$= \lim_{t \to 0^+} \max_{j \in A(x)} a_j^T p$$

$$= \max_{j \in A(x)} a_j^T p$$

Thus $\psi \in \partial f_s(x)$ if and only if

$$\max_{j \in A(x)} a_j^T p \geq \psi^T p \quad \forall p.$$

It is evident that $a_j \in \partial f_s(x)$ for each $j \in A(x)$.

Suppose $A(x) = \{j_1, \ldots, j_p\}$, since $\partial f_s(x)$ is convex, we have

$$\mathrm{Co}\{a_{j_1}, \ldots, a_{j_p}\} := \left\{ \lambda_1 a_{j_1} + \cdots + \lambda_p a_{j_p} \;\middle|\; \sum_{j=1}^{p} \lambda_j = 1,\ \lambda_j \geq 0,\ j = 1, \ldots, p \right\}$$

$$\subseteq \partial f_s(x).$$

We claim $\partial f_s(x) = \mathrm{Co}\{a_{j_1}, \ldots, a_{j_p}\}$

Recall for each $p \in \mathbb{R}^n$

$$\underbrace{\max_{j \in A(x)} a_j^T p}_{f^\circ(x;p)} = \max\{\psi^T p \mid \psi \in \partial f(x)\}$$

① Suppose $\psi \notin \mathrm{span}\{a_{j_1}, \ldots, a_{j_p}\}$

$\Longrightarrow$ $\exists q$ such that

(i) $a_{j_\ell}^T q = 0 \quad \ell = 1, \ldots, p$, and

(ii) $\psi^T q > 0$

$\Longrightarrow$ $\psi \notin \partial f_s(x)$.

② Suppose

$$\psi = \lambda_1 a_{j_1} + \cdots + \lambda_p a_{j_p} \quad \text{but some } \lambda_\ell < 0$$

$$\Longrightarrow$$

(i) $\max_{j \in A(x)} a_j^T (-a_{j_\ell}) = 0$, and

(ii) $\psi^T(-a_{j_\ell}) = -\lambda_\ell > 0$

$$\Longrightarrow$$

$$\psi \notin \partial f_s(x)$$

③ Now suppose

$$\psi = \lambda_1 a_{j_1} + \cdots + \lambda_p a_{j_p} \quad \lambda_j \geq 0 \quad \text{but } \sum_{j=1}^{p} \lambda_j \neq 1$$

if $\sum_{j=1}^{p} \lambda_j > 1$

(i) $\max_{j \in A(x)} a_j^T (a_{j_1} + \cdots + a_{j_p}) = 1$

(ii) $\psi^T (a_{j_1} + \cdots + a_{j_p}) = \sum_{j=1}^{p} \lambda_j > \cancel{0}\, 1$

$$\Longrightarrow \psi \notin \partial f(x)$$

if $\sum_{j=1}^{p} \lambda_j < 1$

(i) $\max_{j \in A(x)} a_j^T (-a_{j_1} + \cdots + -a_{j_p}) = -1$

(ii) $\psi^T (-a_{j_1} - \cdots - a_{j_p}) = -\sum_{j=1}^{p} \lambda_j > -1$

$$\Longrightarrow \psi \notin \partial f(x)$$

Consequently,

x is a global minimizer

$$\Longleftrightarrow$$

$$0 \in Co\{a_{j_1}, \ldots, a_{j_p}\}$$

$$\Longleftrightarrow$$

$\exists \lambda_1, \ldots, \lambda_p$ s.t.

(i) $\lambda_1, \ldots, \lambda_p \geq 0$,

(ii) $\lambda_1 + \cdots + \lambda_p = 1$, and

(iii) $\lambda_1 a_{j_1} + \cdots + \lambda_p a_{j_p} = 0$

Smooth characterization

$$\begin{array}{ll} \underset{x \in \mathbb{R}^n}{\text{minimize}} & \varepsilon \\ & l_j(x) \leq \varepsilon \quad j=1,\dots,K \end{array}$$

equivalently

$$\begin{array}{ll} \underset{x \in \mathbb{R}^n}{\text{minimize}} & \varepsilon \\ & \varepsilon - a_j^T x - b_j \geq 0 \quad j=1,\dots,K \end{array}$$

By the KKT conditions for the latter

x_* is a global minimizer

$\Longleftrightarrow$

$\exists\, \lambda_{j_1}, \dots, \lambda_{j_p}$ such that

(i) $\begin{bmatrix} 1 \\ 0 \end{bmatrix} = \lambda_{j_1} \begin{bmatrix} 1 \\ -a_{j_1} \end{bmatrix} + \dots + \lambda_{j_p} \begin{bmatrix} 1 \\ -a_{j_p} \end{bmatrix}$

(ii) $\lambda_{j_1}, \dots, \lambda_{j_p} \geq 0$

where $j_1, \dots, j_p$ are s.t. $A(x_*) = \{j_1, \dots, j_p\}$

$\Longleftrightarrow$

$\exists\, \lambda_{j_1}, \dots, \lambda_{j_p}$ such that

(i) $\lambda_{j_1}, \dots, \lambda_{j_p} \geq 0$,

(ii) $\lambda_{j_1} + \dots + \lambda_{j_p} = 1$, and

(iii) $\lambda_{j_1} a_{j_1} + \dots + \lambda_{j_p} a_{j_p} = 0$

where $j_1, \dots, j_p$ are s.t. $A(x_*) = \{j_1, \dots, j_p\}$.