# LECTURE 11

## NONSMOOTH ANALYSIS AND OPTIMIZATION

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

$f : \mathbb{R}^n \to \mathbb{R}$, not differentiable everywhere

But Lipschitz continuous, i.e., there exists $\gamma > 0$ such that

$$|f(x) - f(y)| \leq \gamma \|x-y\| \quad \forall x, y$$

Ex

① Total least error

$$\phi(t) = a_0 + a_1 t + \dots + a_k t^k$$

best approximates the data $(t_1, \phi_1), \dots, (t_N, \phi_N)$

$$\underset{a_0, \dots, a_k \in \mathbb{R}}{\text{minimize}} \sum_{j=1}^{N} |\phi_j - \phi(t_j)|$$

$$= \underset{a_0, \dots, a_k \in \mathbb{R}}{\text{minimize}} \left\| \begin{bmatrix} \phi_1 \\ \vdots \\ \phi_N \end{bmatrix} - \begin{bmatrix} 1 & t_1 & \cdots & t_1^k \\ \vdots & & & \\ 1 & t_N & \cdots & t_N^k \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_k \end{bmatrix} \right\|_1$$

② Minimizing the distance function

$C$ - given subset of $\mathbb{R}^n$

$$d_C(x) := \min \{ \|x-\tilde{x}\| \mid \tilde{x} \in C \}$$

$$\underset{x \in S}{\text{minimize}} \quad \underbrace{d_C(x)}_{\text{not differentiable}}$$

③ Minimizing the largest eigenvalue

$$A : \mathbb{R}^d \longrightarrow \mathbb{R}^{n \times n}$$

$A(x)$ is symmetric $\forall x$

$$\underset{x \in \mathbb{R}^d}{\text{minimize}} \quad \underbrace{\lambda_{max}(A(x))}_{\text{not differentiable}}$$

not differentiable even when $A(x)$ depends on $x$ smoothly

Background

Limit superior of a function

$$\phi_r(x) := \sup \{ f(\tilde{x}) \mid 0 < \|x-\tilde{x}\| < r \}$$

$$\limsup_{\tilde{x} \to x} f(\tilde{x}) := \inf \{ \phi_r(x) \mid r > 0 \}$$

equivalently

$$\limsup_{\tilde{x} \to x} f(\tilde{x}) := \lim_{r \to 0^+} \phi_r(x)$$

## Ex

$$f(x) = \begin{cases} \sin(1/x) & x \neq 0 \\ 0 & x = 0 \end{cases}$$

$$\limsup_{x \to 0} f(x) = \lim_{r \to 0^+} \underbrace{\phi_r(0)}_{1} = 1$$

Upper semi-continuity

$f$ is upper semi-continuous at $x$ if

$$\limsup_{\tilde{x} \to x} f(\tilde{x}) \leqslant f(x)$$

## Ex

$$f(x) = \begin{cases} \sin(1/x) & x \neq 0 \\ 1 & x = 0 \end{cases}$$

upper semi-continuous at $0$, i.e.,

$$\limsup_{x \to 0} f(x) = 1 = f(0)$$

Generalized directional derivative (in the direction $p$)

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

Well-defined under Lipschitz continuity assumption, since

$$\frac{f(y+tp) - f(y)}{t} \leqslant \frac{\gamma t \|p\|}{t} = \gamma \|p\|$$

<u>Ex</u>

$f(x) = |x|$,

$$\sup \left\{ \frac{|y+tp| - |y|}{t} \;\middle|\; 0 < \left\| \begin{bmatrix} y-0 \\ t-0 \end{bmatrix} \right\| < r \text{ and } t > 0 \right\}$$

$$= \begin{cases} p & \text{if } p > 0 \\ -p & \text{if } p < 0 \end{cases}$$

thus

$$f^{\circ}(0; p) = \begin{cases} p & \text{if } p > 0 \\ -p & \text{if } p < 0 \end{cases}$$

## THM

Let $f : \mathbb{R}^n \to \mathbb{R}$ be Lipschitz continuous with Lipschitz constant $\gamma$.

(i) $f^\circ(x;p) \leqslant \gamma \|p\| \quad \forall p \neq 0$

(ii) (Positive homogeneity)

$$f^\circ(x;tp) = t f^\circ(x;p) \quad \forall p \neq 0, \forall t > 0$$

(iii) ~~(Sublinearity)~~ (Subadditivity)

$$f^\circ(x;p+q) \leqslant f^\circ(x;p) + f^\circ(x;q)$$

(iv) (Upper semi-continuity)

$$\limsup_{\substack{y \to x \\ q \to p}} f^\circ(y;q) \leqslant f^\circ(x;p)$$

(v) (Lipschitz continuity)

$$| f^\circ(x;p) - f^\circ(x;q) | \leqslant \gamma \|p-q\|$$

(vi) $f^\circ(x;-p) = (-f)^\circ(x;p) \quad \forall p \neq 0$

## PROOF

(i) For every $r > 0$

$$\sup \left\{ \underbrace{\frac{f(y+tp)-f(y)}{t}}_{\leqslant \gamma \|p\|} \;\middle|\; 0 < \left\| \begin{bmatrix} y-x \\ t \end{bmatrix} \right\| < r \text{ and } t > 0 \right\}$$

$\leqslant \gamma \|p\|$

Taking the limit as $r \to 0^+$

$$f^\circ(x;p) \leqslant \gamma \|p\|$$

(ii)

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

$$= t \limsup_{\substack{y\to x\\ s\to 0^+}} \frac{f(y+stp)-f(y)}{st}$$

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

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

(iii)

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

$$= \limsup_{\substack{y\to x\\ s\to 0^+}} \frac{f(y+sp+sq)-f(y+sp)+f(y+sp)-f(y)}{s}$$

$$\leq \left[\limsup_{\substack{y\to x\\ s\to 0^+}} \frac{f(\overset{z}{y+sp}+sq)-f(\overset{z}{y+sp})}{s}\right] \to f^{\circ}(x;q) \quad z\to x$$

$$+$$

$$\left[\limsup_{\substack{y\to x\\ s\to 0^+}} \frac{f(y+sp)-f(y)}{s}\right] \to f^{\circ}(x;p)$$

(iv) Let $\{x^{(k)}\}$, $\{p^{(k)}\}$ be sequences such that $\lim_{k\to\infty} x^{(k)} = x$ and $\lim_{k\to\infty} p^{(k)} = p$.

For each $k$, there exist $y^{(k)}$ and $t^{(k)} > 0$ satisfying

$$f^\circ(x^{(k)}; p^{(k)}) \leq \frac{f(y^{(k)} + t^{(k)} p^{(k)}) - f(y^{(k)})}{t^{(k)}} + \frac{1}{k}$$

$$= \frac{f(y^{(k)} + t^{(k)} p) - f(y^{(k)})}{t^{(k)}} + \frac{f(y^{(k)} + t^{(k)} p^{(k)}) - f(y^{(k)} + t^{(k)} p)}{t^{(k)}} + \frac{1}{k}$$

and

$$\| y^{(k)} - x^{(k)} \| + t^{(k)} < \frac{1}{k}.$$

Taking the lim sup as $k \to \infty$ and exploiting Lipschitz continuity of $f$

$$\limsup_{k\to\infty} f^\circ(x^{(k)}; p^{(k)}) \leq \limsup_{k\to\infty} \frac{f(y^{(k)} + t^{(k)} p) - f(y^{(k)})}{t^{(k)}} + \lim_{k\to\infty} \gamma \| p^{(k)} - p \| + \lim_{k\to\infty} \frac{1}{k}$$

$$\leq \cancel{\limsup}\, f^\circ(x; p)$$

$$\Longrightarrow$$

$$\limsup_{\substack{y \to x \\ q \to p}} f^\circ(y; q) \leq f^\circ(x; p)$$

(v) Due to Lipschitz continuity, for $t>0$,

$$f(x+tp) - f(x) \leq f(x+tq) - f(x) + \gamma t\|p-q\|$$

$$\Longrightarrow$$

$$\frac{f(x+tp)-f(x)}{t} \leq \frac{f(x+tq)-f(x)}{t} + \gamma\|p-q\|$$

$$\Longrightarrow$$

$$f^{\circ}(x;p) \leq f^{\circ}(x;q) + \gamma\|p-q\|$$

Similarly,

$$f^{\circ}(x;q) \leq f^{\circ}(x;p) + \gamma\|p-q\|$$

implying the result.

(vi)

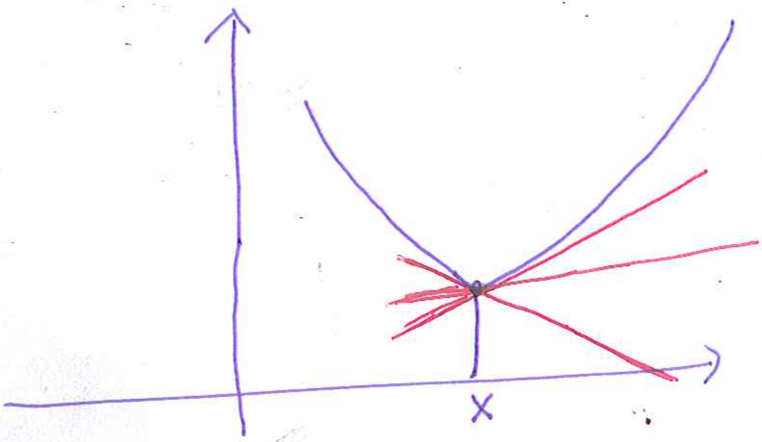
$$f^{\circ}(x;-p) = \limsup_{\substack{y\to x\\ t\to 0^+}} \frac{f(\overset{u}{y-tp})-f(y)}{t}$$

$$= \limsup_{\substack{u\to x\\ t\to 0^+}} \frac{-f(u+tp)+f(u)}{t}$$

$$= (-f)^{\circ}(x;p). \qquad \square$$

# Generalized gradient

$$\partial f(x) :\!= \{ \psi \in \mathbb{R}^n \mid f^\circ(x;v) \geq \psi^T v \quad \forall v \in \mathbb{R}^n \}$$




$\partial f(x) =$ slopes of all lines that locally lies underneath $f(x)$

Ex

$f(x) = |x|$

Case
$x > 0$

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

$$= p$$

~~Case~~

$$\partial f(x) = \{ \psi \in \mathbb{R} \mid p \geq \psi p \quad \forall p \in \mathbb{R} \}$$

$$= \{1\}.$$

## Case 2

$x < 0$

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

$$\partial f(x) = \{\psi \mid -p \geqslant \psi p \quad \forall p \in \mathbb{R}\}$$
$$= \{-1\}$$

## Case 3

$x = 0$

$$f^{\circ}(x;p) = \begin{cases} p & p > 0 \\ -p & p < 0 \end{cases}$$

$\psi \leq 1$

$$\partial f(x) = \{\psi \mid p \geqslant \psi p \quad \forall p > 0$$
$$-p \geqslant \psi p \quad \forall p < 0\}$$
$$= [-1, 1]$$

$\psi \geqslant -1$

## Proposition

(i) $\partial f(x)$ is non-empty.

(ii) $\partial f(x)$ is closed, convex and such that

$$\|\psi\| \leqslant \gamma \qquad \forall \psi \in \partial f(x).$$

(iii) $f^{\circ}(x;p) = \max\{\psi^T p \mid \psi \in \partial f(x)\}$

## Proof

(i) Let $\{q_1, \ldots, q_n\}$ be any basis for $\mathbb{R}^n$ such that $q_j \perp q_k$ for $j \neq k$. Define

$$\psi_1 := -f^\circ(x; -q_1) \frac{q_1}{\|q_1\|^2}$$

$$\psi_k := \psi_{k-1} + \alpha_k \frac{q_k}{\|q_k\|^2} \qquad k = 2, \ldots, n$$

$$= -f^\circ(x; -q_1) \frac{q_1}{\|q_1\|^2} + \sum_{\ell=2}^{k} \alpha_\ell \frac{q_\ell}{\|q_\ell\|^2}$$

where

$$\alpha_\ell := \sup_{v \in S_{\ell-1}} \psi_{\ell-1}^T v - f^\circ(x; v - q_\ell)$$

We claim that

$$f^\circ(x; v) \geq \psi_k^T v \quad \forall v \in S_k := \text{span}\{q_1, \ldots, q_k\}$$

for $k = 1, \ldots, n$. Thus $\psi_n \in \partial f(x)$. Proof is by induction.

Base case $(k = 1)$ Observe $f^\circ(x;v) \geq -f^\circ(x;-v)$

$$f^\circ(x; tq_1) \geq -t f^\circ(x; -q_1)$$

$$= \underbrace{\left(-f^\circ(x;-q_1) \frac{q_1}{\|q_1\|^2}\right)^T}_{\psi_1} (t q_1)$$

for all $t$.

(Note: $f^\circ(x; -q_1) = \psi_1^T(-q_1)$)

Inductive case

Suppose $f^\circ(x; v) \geq \psi_{k-1}^T v \quad \forall v \in S_{k-1}$.

For every $v, w \in S_{k-1}$

$$\psi_{k-1}^T (v+w) \leq f^\circ(x; v+w) = f^\circ(x; v - q_k + w + q_k)$$

$$\underset{\text{subadditivity}}{\leq} f^\circ(x; v-q_k) + f^\circ(x; w+q_k)$$

$$\Longrightarrow \underbrace{\psi_{k-1}^T v - f^\circ(x; v-q_k)}_{\sup\, = \alpha_k} \leq f^\circ(x; w+q_k) - \psi_{k-1}^T w$$

$$\Longrightarrow \begin{cases} (+)\ \psi_{k-1}^T v - \alpha_k \leq f^\circ(x; v-q_k) \\ (++)\ \psi_{k-1}^T w + \alpha_k \leq f^\circ(x; w+q_k) \end{cases}$$

Plug in $(t^{-1}v)$ for $v$ in $(+)$, and $(t^{-1}w)$ for $w$ in $(++)$ for any $t > 0$. Then multiply by $t$ leading to

$$\begin{cases} t\left(\psi_{k-1}^T (t^{-1}v)\right) - t\alpha_k \leq t f^\circ(x; t^{-1}v - q_k) \\ t\left(\psi_{k-1}^T (t^{-1}w)\right) + t\alpha_k \leq t f^\circ(x; t^{-1}w + q_k) \end{cases}$$

$$\Longrightarrow \begin{cases} (*)\ \psi_{k-1}^T v - t\alpha_k \leq f^\circ(x; v - tq_k) \\ (**)\ \psi_{k-1}^T w + t\alpha_k \leq f^\circ(x; w + tq_k) \end{cases}$$

Now take any $z \in S_k \setminus S_{k-1}$ if $z = w + tq_k$ for some $w \in S_{k-1}$ and $t > 0$, then

$$\psi_k^T z = \psi_{k-1}^T w + t\left(\alpha_k \frac{q_k^T}{\|q_k\|^2}\right) q_k$$

$$= \psi_{k-1}^T w + t\alpha_k \underset{(**)}{\leq} f^\circ(x; \underbrace{w + tq_k}_{z})$$

Otherwise $z = v - t q_k$ for some $v \in S_{k-1}$ and $t > 0$. Thus

$$\psi_k^T z = \psi_{k-1}^T v - t \left( \alpha_k \frac{q_k^T}{\|q_k\|^2} \right) q_k$$

$$= \psi_{k-1}^T v - t\alpha_k \underset{(*)}{\leqslant} f^\circ(x; \underbrace{v - t q_k}_{z})$$

On the other hand, if $\underset{\sim}{w}_Z \in S_{k-1}$

$$\psi_k^T \underset{\sim}{w}_Z = \psi_{k-1}^T \underset{\sim}{w}_Z \leqslant f^\circ(x; \underset{\sim}{w}_Z)$$

by the inductive hypothesis.

Consequently,

$$\psi_n^T v \leqslant f^\circ(x; v).$$

for all $v \in S_n$.

(ii) This is rather straightforward. Let us only prove convexity.

$$\psi_1, \psi_2 \in \partial f(x)$$

$$\Longleftrightarrow$$

$$f^\circ(x;v) \geqslant \psi_1^T v \quad \text{and} \quad f^\circ(x;v) \geqslant \psi_2^T v \quad \forall v$$

$$\Longleftrightarrow$$

$$\theta f^\circ(x;v) \geqslant \theta \psi_1^T v \quad \text{and} \quad (1-\theta) f^\circ(x;v) \geqslant (1-\theta)\psi_2^T v \quad \forall v,\ \forall \theta \in [0,1]$$

$$\Longrightarrow$$

$$f^\circ(x;v) \geqslant [\theta\psi_1 + (1-\theta)\psi_2]^T v \quad \forall v,\ \forall \theta \in [0,1]$$

$$\Longrightarrow$$

$$\theta\psi_1 + (1-\theta)\psi_2 \in \partial f(x) \quad \forall \theta \in [0,1]$$

(iii) By definition of $\partial f(x)$, for given $p$,

for each $\psi \in \partial f(x)$

$$f^{\circ}(x;p) \geqslant \psi^T p$$

so

$$f^{\circ}(x;p) \geqslant \max\{\psi^T p \mid \psi \in \partial f(x)\}$$

On the other hand, in part (i), using a basis $\{-p, q_2, \dots, q_n\}$ (where vectors are mutually orthogonal) the constructed $\psi_n \in \partial f(x)$ satisfies

$$\psi_n^T p = \psi_1^T p$$

$$= f^{\circ}(x;p)$$ (see the note at the end of base case)

establishing

$$f^{\circ}(x;p) \leqslant \max\{\psi^T p \mid \psi \in \partial f(x)\}.$$

First order necessary condition

THM

Let $f : \mathbb{R}^n \to \mathbb{R}$ be Lipschitz continuous. If $x_*$ is a local minimizer of $f$, then $0 \in \partial f(x_*)$.

PROOF

Suppose $0 \notin \partial f(x_*)$. There exists $p \in \mathbb{R}^n$ such that

$$f^\circ(x_*; p) < 0$$

(otherwise $f^\circ(x_*; p) \geq 0 \ \forall p \Longrightarrow 0 \in \partial f(x_*)$)

Without loss of generality, assume $\|p\| = 1$. (due to positive homogeneity.)

For each $\epsilon > 0$ small enough

$$0 > \sup \left\{ \frac{f(y+tp) - f(y)}{t} \;\middle|\; 0 < \left\| \begin{bmatrix} y - x_* \\ t \end{bmatrix} \right\| < \epsilon \text{ and } t > 0 \right\}$$

$$\geq \sup \left\{ \frac{f(x_* + tp) - f(x_*)}{t} \;\middle|\; 0 < t < \epsilon \right\}$$

$\Longrightarrow \exists (x_* + tp) \in B(x_*, \epsilon)$ s.t. $f(x_*) > f(x_* + tp)$

Thus $x_*$ is not a local minimizer. $\square$

Ex

$$f(x) = x^2 \cdot \sin^2\left(\frac{1}{x}\right) \quad x \neq 0, \qquad f(0) = 0$$

$f'(0) = 0$ differentiable but not continuously differentiable.

Smooth first order necessary conditions don't apply (they apply if $f$ is continuously differentiable.)

$$\phi_r(0) = \sup \left\{ \frac{(h+tp)^2 \sin^2\left(\frac{1}{h+tp}\right) - h^2 \cdot \sin^2\left(\frac{1}{h}\right)}{t} \;\middle|\; 0 < \left\| \begin{bmatrix} h \\ t \end{bmatrix} \right\|_2 < r \text{ and } t > 0 \right\}$$

$$\left[ (h+tp)^2 \sin^2\left(\frac{1}{h+tp}\right) - h^2 \cdot \sin^2\left(\frac{1}{h}\right) \right] / t =$$

$$\underbrace{\left(\frac{2htp + t^2p^2}{t}\right) \sin^2\left(\frac{1}{h+tp}\right)}_{\to 0 \text{ as } (h,t) \to 0} + \underbrace{\frac{h^2}{t}\left(\sin^2\left(\frac{1}{h+tp}\right) - \sin^2\left(\frac{1}{h}\right)\right)}_{\substack{\leq |p| \text{ as } t \to 0 \\ \geq |p| \text{ if } h \text{ is s.t. } \sin\left(\frac{1}{h}\right) = -\mathrm{sign}(p)}}$$

Consequently,

$$\cancel{\phi_r(0) =} \; f^\circ(0; p) = |p|$$

$$\partial f(0) = \Big\{ \psi \;\Big|\; \begin{array}{ll} p \geq \psi p \ \ \text{if } p \geq 0 & \rightarrow \psi \leq 1 \\ -p \geq \psi p \ \ \text{if } p < 0 & \rightarrow \psi \geq -1 \end{array} \Big\}$$

$$= [-1, 1].$$

Since, $0 \in \partial f(0)$, $x_* = 0$ satisfies first order necessary conditions.

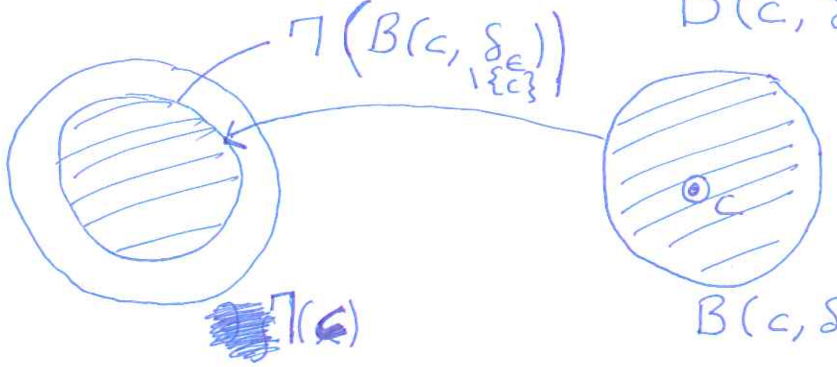
Upper semi-continuity

View $\partial f(x) : \mathbb{R}^n \to \mathbb{R}^n$ as a multi-valued function (i.e. from $\mathbb{R}^n$ to subsets of $\mathbb{R}^n$)

A multi-valued function $\Gamma : \mathbb{R}^n \to \mathbb{R}^m$ is called upper semi-continuous at $c \in \mathbb{R}^n$ if for each $\epsilon > 0$ there exist $\delta_\epsilon > 0$ such that

$$\Gamma(x) \subset \Gamma(c) + B(0, \epsilon)$$

for all $x \in$ ~~$c + B(0, \delta_\epsilon) \setminus \{c\}$~~ $B(c, \delta_\epsilon) \setminus \{c\}$



THM

(1) Let $\{x^{(k)}\}$ and $\{\psi^{(k)}\}$ be convergent sequences in $\mathbb{R}^n$ such that $\psi^{(k)} \in \partial f(x^{(k)})$. Then $\left[\lim_{k\to\infty} \psi^{(k)}\right] \in \partial f\left(\lim_{k\to\infty} x^{(k)}\right)$.

(2) $\partial f(x)$ is upper semi-continuous.

PROOF

(1) Let $\psi := \lim_{k\to\infty} \psi^{(k)}$ and $x := \lim_{k\to\infty} x^{(k)}$.

By definition of $\partial f(\cdot)$, for all $p \in \mathbb{R}^n$

$$f^\circ(x^{(k)}; p) \geq \left[\psi^{(k)}\right]^T p$$

$$\Longrightarrow \underbrace{\limsup_{k\to\infty} f^\circ(x^{(k)}; p)}_{f^\circ(x;p) \geq} \geq \lim_{k\to\infty} \left[\psi^{(k)}\right]^T p$$

$$\Longrightarrow f^\circ(x;p) \geq \psi^T p$$

Consequently, $\psi \in \partial f(x)$.

(2) Suppose $\partial f(x)$ is not upper semi-continuous. Then ~~for each~~ there exists $\epsilon > 0$ such that

$\forall k \in \mathbb{Z}^+$ $\cancel{\forall y \in \partial f(x)}$ $\exists x^{(k)} \in \mathbb{R}^n$

(i) $\| x^{(k)} - x \| \leq 1/k$ and

(ii) $\exists z^{(k)} \in \partial f(x^{(k)})$ s.t. $\| z^{(k)} - y \| > \epsilon$ $\forall y \in \partial f(x)$

This amounts to $\lim_{k \to \infty} x^{(k)} = x$,

$z^{(k)} \in \partial_f (x^{(k)})$, yet $\lim_{k \to \infty} z^{(k)} \notin \partial_f (x)$,

contradicting part (1). $\square$

## Gradients and generalized gradients

$f : \mathbb{R}^n \to \mathbb{R}$ is differentiable at $\tilde{x}$ is $\partial f(\tilde{x}) / \partial x_j$ exists and

$$\lim_{t \to 0^+} \frac{f(\tilde{x} + tp) - f(\tilde{x})}{t} = \nabla f(\tilde{x})^T p \quad \forall p.$$

THM

If $f$ is differentiable at $\tilde{x}$, then $\nabla f(\tilde{x}) \in \partial_f (\tilde{x})$.

PROOF

For every $p \in \mathbb{R}^n$

$$\nabla f(\tilde{x})^T p = \lim_{t \to 0^+} \frac{f(\tilde{x} + tp) - f(\tilde{x})}{t}$$

$$= \limsup_{t \to 0^+} \frac{f(\tilde{x} + tp) - f(\tilde{x})}{t}$$

$$\leq \limsup_{\substack{y \to \tilde{x} \\ t \to 0^+}} \frac{f(y+tp) - f(y)}{t}$$

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

Consequently, $\nabla f(\tilde{x}) \in \partial f(\tilde{x})$.

$\square$

Ex

$$f(x) = x^2 \sin^2(1/x)$$

Recall

$f$ is differentiable at $0$

but not continuously differentiable

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

THM

Suppose $f$ is differentiable on $B(\tilde{x}, \delta)$ for some $\delta > 0$. Then

$\partial f(\tilde{x}) = \{\nabla f(\tilde{x})\} \iff f$ is continuously differentiable at $\tilde{x}$

---

Recall

$f$ is continuously differentiable at $\tilde{x}$ if $\nabla f(x)$ is continuous at $\tilde{x}$.

PROOF

Suppose $f$ is continuously differentiable at $\tilde{x}$.
By the mean value thm

$$\phi_r(\tilde{x}) = \sup \left\{ \frac{f(y+tp) - f(y)}{t} \;\middle|\; 0 < \left\| \begin{bmatrix} y-\tilde{x} \\ t \end{bmatrix} \right\| < r, \ t>0 \right\}$$

$$= \sup \left\{ \nabla f(y+\tilde{t}p)^T p \;\middle|\; \forall y,t \ 0 < \left\| \begin{bmatrix} y-\tilde{x} \\ t \end{bmatrix} \right\| < r \text{ and } t>0, \ \exists \tilde{t} \in (0,t) \right\}.$$

Consequently

$$\underbrace{\inf\{\nabla f(y)^T p \mid 0<\|y-\tilde{x}\|<2r\}}_{L_{2r}(\tilde{x}) :=} \leq \phi_r(\tilde{x}) \leq \underbrace{\sup\{\nabla f(y)^T p \mid 0<\|y-\tilde{x}\|<2r\}}_{U_{2r}(\tilde{x}) :=}$$

By continuous differentiability at $\tilde{x}$

$$\lim_{r\to 0^+} L_{2r}(\tilde{x}) = \lim_{r\to 0^+} U_{2r}(\tilde{x}) = \nabla f(\tilde{x})^T p$$

so

$$f^\circ(\tilde{x};p) = \lim_{r\to 0^+} \phi_r(\tilde{x}) = \nabla f(\tilde{x})^T p.$$

This implies $\partial f(\tilde{x}) = \{\nabla f(\tilde{x})\}$.

Conversely, suppose $\partial f(\tilde{x}) = \{\nabla f(\tilde{x})\}$. We claim $f^\circ(\tilde{x};p) = \nabla f(\tilde{x})^T p \ \forall p$. For each $p$, there exists $\psi$ such that

$$f^\circ(\tilde{x};p) = \psi^T p \geq \nabla f(\tilde{x})^T p, \text{ and}$$

$$f^\circ(\tilde{x};q) \geq \psi^T q \quad \forall q \in \mathbb{R}^n.$$

(see the proof of proposition on page 10 part (i)). If $f^\circ(\tilde{x};p) > \nabla f(\tilde{x})^T p$, then $\psi \neq \nabla f(\tilde{x})$ and $\psi \in \partial f(\tilde{x})$. Thus $f^\circ(\tilde{x};p) = \nabla f(\tilde{x})^T p$ for all $p$. But then for all $p$

$$\limsup_{\substack{y \to \tilde{x} \\ t \to 0^+}} \frac{f(y+tp)-f(y)}{t}$$

$$= f^\circ(\tilde{x};p) = \nabla f(\tilde{x})^T p$$

$$= -\nabla f(\tilde{x})^T(-p) = -f^\circ(\tilde{x};-p)$$

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

$$= \liminf_{\substack{y \to \tilde{x} \\ t \to 0^+}} \frac{f(\overbrace{y}^{z+tp}) - f(\overbrace{y-tp}^{z})}{t}$$

$$= \liminf_{\substack{z \to \tilde{x} \\ t \to 0^+}} \frac{f(z+tp)-f(z)}{t}$$

Consequently, the limit

$$\lim_{\substack{y \to \tilde{x} \\ t \to 0^+}} \frac{f(y+tp)-f(y)}{t} = \nabla f(\tilde{x})^T p$$

exists. This in turn implies that $f$ is continuously differentiable at $\tilde{x}$.

To prove the last claim, for each $\epsilon > 0$ there exists $\delta_\epsilon > 0$ such that

$$\text{(a)} \quad \left| f(\tilde{y} + tp) - f(\tilde{y}) - \nabla f(\bar{x})^T (tp) \right| \leqslant t\epsilon$$

for all $t \leqslant \delta_\epsilon$ and all $\tilde{y} \in B(\bar{x}, \delta_\epsilon)$.

Pick any $y \in B(\bar{x}, \tilde{\delta})$ where $\tilde{\delta} := \min\left(\frac{\delta_\epsilon}{2}, \frac{\delta}{2}\right)$. Since $f$ is differentiable at $y$, there exists $\tilde{\delta}_\epsilon > 0$ such that

$$\text{(b)} \quad \left| f(y + tp) - f(y) - \nabla f(y)^T (tp) \right| \leqslant t\epsilon$$

for all $t \leqslant \tilde{\delta}_\epsilon$. Consequently, for every $t \leqslant \min(\delta_\epsilon, \tilde{\delta}_\epsilon)$,

$$\left| \nabla f(y)^T (tp) - \nabla f(\bar{x})^T (tp) \right| =$$

$$\Big| \underbrace{\left[ f(y+tp) - f(y) - \nabla f(\bar{x})^T (tp) \right]}_{\leqslant t\epsilon \text{ in abs. value by (a)}} - \underbrace{\left[ f(y+tp) - f(y) - \nabla f(y)^T (tp) \right]}_{\leqslant t\epsilon \text{ in abs. value by (b)}} \Big| \leqslant 2t\epsilon$$

$$\Longrightarrow \quad \left\| \nabla f(y) - \nabla f(\bar{x}) \right\| \leqslant \cancel{\sqrt{2}\epsilon}\ 2\epsilon \quad \forall y \in B(\bar{x}, \delta)$$

since $p$ is arbitrary

Consequently, $\lim_{y \to \bar{x}} \nabla f(y) = \nabla f(\bar{x})$, that is $f$ is continuously differentiable.

## Remark

In the second part of the proof ($\Longrightarrow$ direction), it is deduced that

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

$$\lim_{\substack{y \to \tilde{x} \\ t \to 0^+}} \frac{f(y+tp) - f(y)}{t} = \nabla f(\tilde{x})^T p \quad \forall p$$

$$\xRightarrow[\substack{\text{setting} \\ y = \tilde{x}}]{} f \text{ is differentiable at } \tilde{x}.$$

Thus

$$f \text{ is not differentiable at } x$$

$$\Longrightarrow$$

$$\partial f(\bar{x}) \neq \{\nabla f(\bar{x})\}.$$