3.1 Definition & Example.

Def 3.1 Subgradient

Let $f: E \to (-\infty, \infty]$ be a proper function, $x \in \mathrm{dom}(f)$.

A vector $g \in E^*$ is called a subgradient of f at x if:

$$f(y) \geq f(x) + \langle g, y-x \rangle \quad \text{for all } y \in E. \qquad (3.1)$$

(3.1) ... subgradient inequality.

Def 3.2 Subdifferential.

The set of all subgradients of f at x is called subdifferential of f at x and is denoted by $\partial f(x)$:

$$\partial f(x) \equiv \{ g \in E^* : f(y) \geq f(x) + \langle g, y-x \rangle \text{ for all } y \in E \}.$$

when $x \notin \mathrm{dom}(f)$, define $\partial f(x) = \phi$.

Eg.s.

1. (3.3) Subdifferential of norms at 0.

$f: E \to \mathbb{R}$, $f(x) = \|x\|$.

$\partial f(0) = B_{\|\cdot\|_*}[0,1] = \{ g \in E^* : \|g\|_* \leq 1 \}$. → dual norm unit ball.

proof: $\forall y \in E$, $f(y) \geq f(0) + \langle g, y-0 \rangle = \langle g, y \rangle$.

$\Rightarrow \|y\| \geq \langle g, y \rangle$ ※ We will prove that ※ is equivalent to $\|g\|_* \le 1$.

($\Leftarrow$) If $\|g\|_* \le 1$, then by Cauchy-Schwarz,

$$\langle g, y\rangle \le \|g\|_* \|y\| \le \|y\|. \quad \text{Done.}$$

($\Rightarrow$) If $\|y\| \ge \langle g, y\rangle$, then

$$\|g\|_* = \max_{\|y\|=1} \langle g, y\rangle \le \max_{\|y\|=1} \|y\| = 1 \quad \text{Done.}$$

2. (3.4). Subdifferential of l_1-norm at 0. (Special case of eg 1)

$f: \mathbb{R}^n \to \mathbb{R}$. $\quad f(x) = \|x\|_1$.

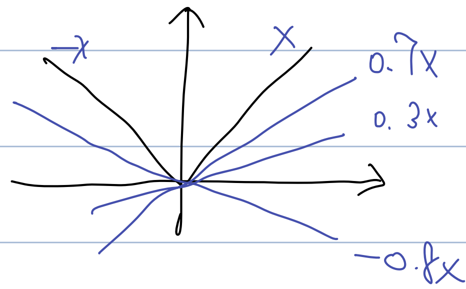
$$\partial f(0) = B_{\|\cdot\|_\infty}[0,1] = [-1,1]^n.$$

(l-∞ norm is the dual of l_1-norm).

specifically, when $n=1$ $\quad \partial f(0) = [-1,1]$.

$\Downarrow$

$f(x) = |x|$.



3. (3.5) Subdifferential of indicator functions.

$S \subseteq E$, $S = \phi$, Consider δ_S

For $x \in S$,

$$y \in \partial \delta_S(x) \text{ iff } \forall z \in S, \quad \underline{\delta_S(z) \ge \delta_S(x) + \langle y, z-x\rangle} \quad *$$

$x, z \in S \Rightarrow \delta_S(z) = \delta_S(x)$

$* \iff \langle y, z-x\rangle \le 0, \quad \forall z \in S.$

We call it the normal cone,

$$N_S(x) = \{y \in E^* : \langle y, z-x \rangle \leq 0, \ \forall z \in S\}.$$

$$\partial \delta_S(x) = N_S(x),$$ closed and convex.

4. (3.6) Subdifferential of the indicator function of the unit ball. (Special case of eg 3).

Let $S = B[0,1] = \{x \in E : \|x\| \leq 1\}$.

$$\partial \delta_S(x) = N_S(x) = \{y \in E^* : \langle y, z-x \rangle \leq 0 \text{ for } \forall z \in S\}.$$

explicitly, since $\|x\| \leq 1$,

$y \in N_S(x)$ iff $\langle y, z-x \rangle \leq 0$ for z s.t. $\|z\| \leq 1$

That is, $\max_{\|z\| \leq 1} \langle y, z \rangle \leq \langle y, x \rangle$.

$$\max_{\|z\| \leq 1} \langle y, z \rangle = \|y\|_*$$

$$\|y\|_* \leq \langle y, x \rangle$$

Therefore,

$$\partial \delta_{B[0,1]}(x) = N_{B[0,1]}(x) = \begin{cases} \{y \in E^* : \|y\|_* \leq \langle y, x \rangle\}. & \|x\| \leq 1 \\ \phi & \|x\| > 1 \end{cases}$$

5. (3.7) Subgradient of dual function.

For minimization problem, $\min\{f(x) : g(x) \leq 0, x \in X\}$.

$f : E \to \mathbb{R}$, $g : E \to \mathbb{R}^m$.

The Lagrangian objective function:

$$q(\lambda) = \min_{x \in X} \{L(x;\lambda) \equiv f(x) + \lambda^T g(x)\}.$$

6. (3.8) subgradient of the maximum eigenvalue function.

3.2 Properties of Subdifferential Set.

Thm 3.9 Subdifferential sets are closed & convex.

$f: E \to (-\infty, \infty]$, proper function. Then $\partial f(x)$ is closed & convex for $\forall x \in E$.

proof: $\partial f(x) = \bigcap_{y \in E} H_y$. $H_y := \{g \in E^*: f(y) \geq f(x) + \langle g, y-x \rangle\}$

H_y are half-spaces $\Rightarrow$ closed and convex.

$\Rightarrow \partial f(x)$ is closed and convex.

linear trans.

x, y are fixed $\langle g, y-x \rangle \leq f(y) - f(x)$

Def 3.10 Subdifferentiability.

$\partial f(x) \neq \phi \Rightarrow$ Call f is subdifferentiable at x.

$f \in (-\infty, \infty]$, $x \in dom(f)$.

Collection of points of subdifferentiability

$\mathrm{dom}(\partial f) = \{x \in E : \partial f(x) \neq \phi\}$.

Lemma 3.11 Nonemptiness of subdifferential sets $\Rightarrow$ convexity.

f定义域为凸集，若f处处 subdifferentiable，则 f为凸函数.

Let $f: E \to (-\infty, \infty]$ be proper, assume $\mathrm{dom}(f)$ is convex.

If $\forall x \in \mathrm{dom}(f)$, $\partial f(x) \neq \phi$, then f is convex.

proof: $x, y \in \mathrm{dom}(f)$. Let $\alpha \in [0,1]$. Then $z_\alpha = \alpha x + (1-\alpha) y \in \mathrm{dom}(f)$

For $g \in \partial f(z_\alpha)$, it should satisfy:

$$f(x) \geq f(z_\alpha) + \langle g, x - z_\alpha \rangle = f(z_\alpha) + (1-\alpha)\langle g, x-y \rangle \quad ①$$

$$f(y) \geq f(z_\alpha) + \langle g, y - z_\alpha \rangle = f(z_\alpha) + \alpha \langle g, y-x \rangle \quad ②$$

$$① \Rightarrow \alpha f(z_\alpha) \leq \alpha f(x) + \alpha(1-\alpha)\langle g, y-x \rangle \quad ③$$

$$② \Rightarrow (1-\alpha) f(z_\alpha) \leq (1-\alpha) f(y) - \alpha(1-\alpha)\langle g, y-x \rangle \quad ④$$

$③+④ \Rightarrow f(\alpha x + (1-\alpha) y) = f(z_\alpha) \leq \alpha f(x) + (1-\alpha) f(y)$. f is convex.

Note: Reverse Not True!

Convex Function f $\not\Longrightarrow$ Subdifferentiable at every $x \in \mathrm{dom}(f)$.

eg. $f(x) = \begin{cases} -\sqrt{x} & x \geq 0 \\ \infty & \text{else.} \end{cases}$ not subdifferentiable at $x=0$. with $\mathrm{dom}(f) = \mathbb{R}_+$.

Explanation: Suppose for contradiction, $\exists g$ s.t.

$\forall y \in \mathbb{R}^+$, $f(y) \geq f(0) + \langle g, y-0 \rangle$

$-\sqrt{y} \geq gy$. Let $y=1$ then $-1 \geq g \Rightarrow g<0$

When $y = \frac{1}{2g^2}$ $\quad \frac{1}{2g} \leq -\sqrt{\frac{1}{2g^2}}$

$\overset{\uparrow}{\underset{g<0}{\Rightarrow}} \frac{1}{4g^2} \geq \frac{1}{2g^2}$ Impossible!

Thm 3.13 Supporting hyperplane theorem

Let $C \neq \emptyset$, $C \subseteq E$. C is a convex set. Then $\exists\, p \neq 0$, $p \in E^*$ s.t.

$y \notin \text{int}(C)$ $\langle p, x \rangle \leq \langle p, y \rangle$ for $\forall x \in C$.

Thm 3.14. Nonemptiness and boundness of the subdifferential set at interior points of the domain.

凸函数f在其定义域 dom(f) 中的内点一定可次微分.

Let $f: E \to (-\infty, \infty]$ be proper, convex function, assume $\tilde{x} \in \text{int}(\text{dom}(f))$. Then $\partial f(\tilde{x}) \neq \emptyset$ and $\partial f(\tilde{x})$ is bounded.

proof:

$(\tilde{x}, f(\tilde{x}))$ is on boundary of $\text{epi}(f) \subseteq E \times \mathbb{R}$.

By Thm 3.13, there exists separating hyperplane between $(\tilde{x}, f(\tilde{x}))$ and $\text{epi}(f)$ $\Rightarrow \exists\, (p, -\alpha) \in E^* \times \mathbb{R}$ s.t.

$$\langle (p, -\alpha), (\tilde{x}, f(\tilde{x})) \rangle \geq \langle (p, -\alpha), (x, t) \rangle \quad \text{for } \forall (x,t) \in \text{epi}(f)$$

$$\Rightarrow \underline{\langle p, \tilde{x} \rangle - \alpha f(\tilde{x}) \geq \langle p, x \rangle - \alpha t.} \quad *$$

let $t = f(\tilde{x}) + 1$ $\quad (\tilde{x}, f(\tilde{x}) + 1)$ obviously $\in \text{epi}(f)$.

$$\Big|\, x = \tilde{x}$$

$\Rightarrow \quad \alpha \geq 0.$

$\tilde{x} \in \text{int}(\text{dom}(f))$, by Thm 2.21 (Local lipschitz continuity)
$\exists\, \varepsilon > 0,\ L > 0$ s.t. $B_{\|\cdot\|}[\tilde{x}, \varepsilon] \subseteq \text{dom}(f)$
and $|f(x) - f(\tilde{x})| \leq L\|x - \tilde{x}\|$ for $x \in B_{\|\cdot\|}[\tilde{x}, \varepsilon]$.

For $\tilde{x}$, for $x \in B_{\|\cdot\|}[\tilde{x}, \varepsilon] \subseteq \text{dom}(f)$, $(x, f(x)) \in \text{epi}(f)$.
Let $t = f(x)$, we have.

$$\langle p, \tilde{x} - x\rangle \geq \alpha(f(\tilde{x}) - f(x))$$
$$\langle p, x - \tilde{x}\rangle \leq \alpha(f(x) - f(\tilde{x})) \leq \alpha L\|x - \tilde{x}\|$$

Take $p^{\dagger} \in E$ s.t. $\langle p, p^{\dagger}\rangle = \|p\|_{*}$ and $\|p^{\dagger}\| = 1$.
$\tilde{x} + \varepsilon p^{\dagger} \in B_{\|\cdot\|}[\tilde{x}, \varepsilon]$. Plugs in $x = \tilde{x} + \varepsilon p^{\dagger}$,
we have

$$\langle p, \varepsilon p^{\dagger}\rangle = \varepsilon\langle p, p^{\dagger}\rangle \leq \alpha L \varepsilon \|p^{\dagger}\| = \alpha L \varepsilon.$$

α must be $\alpha > 0$, or $p = 0$, contradicts with Thm 3.13.
Taking $t = f(x)$ in 3.8, divided by α,

$$f(x) \geq f(\tilde{x}) + \langle g, x - \tilde{x}\rangle \text{ for all } x \in \text{dom}(f). \quad g = p/\alpha.$$

$\Rightarrow g \in \partial f(\tilde{x})$. So $\partial f(\tilde{x}) \neq \emptyset$.

Take g^+ s.t. $\langle g, g^+ \rangle = \|g\|_*$ and $\|g^+\| = 1$.

Plugs in $x = \tilde{x} + \varepsilon g^+$

$\Rightarrow \langle g, \varepsilon g^+ \rangle = \varepsilon \|g\|_* \leq f(x) - f(\tilde{x}) \leq L\|x - \tilde{x}\| \leq L\varepsilon.$

$\Rightarrow \|g\|_* \leq L$, implying $\partial f(\tilde{x})$ is bounded

Thm 3.14 In a word: $\mathrm{int}(\mathrm{dom}(f)) \subseteq \mathrm{dom}(\partial f)$.

Collary 3.15.

Let $f: E \to \mathbb{R}$ be convex function. Then f is subdifferentiable over E.

Thm 3.16. 次梯度在紧集上有界.

Let $f: E \to (-\infty, \infty]$ be proper, convex, $X \subseteq \mathrm{int}(\mathrm{dom}(f))$, $X \neq \phi$, X is compact, Then $Y = \bigcup_{x \in X} \partial f(x)$ satisfies $\begin{cases} Y \neq \phi \\ Y \text{ is bounded.} \end{cases}$

proof: $\forall x \in X$, $\partial f(x) \neq \phi$ by Thm 3.14.

$\Rightarrow Y = \bigcup_{x \in X} \partial f(x) \neq \phi$.

Now we prove Y is bounded. Assume by contradiction, $\exists \{x_k\}_{k \geq 1} \subseteq X$ and $g_k \in \partial f(x_k)$ s.t. $\|g_k\|_* \to \infty$ as $k \to \infty$.

Let g_k^+ s.t. $\begin{cases} \langle g_k^+, g_k \rangle = \|g_k\|_* \\ \|g_k^+\| = 1 \end{cases}$

内点集为开集.

$[\text{int}(\text{dom}(f))]^c$ is closed

and $X \cap [\text{int}(\text{dom}(f))]^c = \phi$

$(\text{since } X \subset \text{int}(\text{dom}(f)))$

$\Rightarrow \exists\, \varepsilon > 0$ s.t.

$\|x-y\| \geq \varepsilon$ for $\forall x \in X,\ y \notin \text{int}(\text{dom} f)$ $\ast$

Then $g_k \in \partial f(x_k)$ implies.

$$f(x_k + \tfrac{\varepsilon}{2} g_k^{+}) - f(x_k) \geq \langle \tfrac{\varepsilon}{2} g_k^{+}, g_k \rangle$$
$$= \tfrac{\varepsilon}{2} \|g_k\|_*$$

$x_k + \frac{\varepsilon}{2} g_k^{+} \in \text{int}(\text{dom}(f))$ by $\ast$. In order to show $\|g_k\|_*$ is bounded, we show the left side above is bounded.

Suppose by contradiction, $f(x_k + \frac{\varepsilon}{2} g_k^{+}) - f(x_k)$ is not bounded,

$\exists$ subseq $\{x_{k_i}\}$, $\{g_{k_i}^{+}\}$ s.t.

$$f(x_{k_i} + \tfrac{\varepsilon}{2} g_{k_i}^{+}) - f(x_{k_i}) \to \infty \text{ as } k, i \to \infty.$$

$x \in \text{int}(S) \Rightarrow N_r(x) \subset \text{int}(S)$.

$\exists\, N_r(x) \subset S$.

Consider $\boxed{y \in N_r(x)}$, $y \neq x$.

$N_r(x)$ is open

$\Rightarrow \exists\, r'$ s.t. $N_{r'}(y) \subset N_r(x) \subset$

$\Rightarrow y \in \text{int}(S)$

$\Rightarrow N_r(x) \subset \text{int}(S)$.

Thm 3.17 非空凸集的相对内点非空。

Let $C \subseteq E$, $C \neq \phi$, C is convex. Then $ri(C) \neq \phi$.

[108, Thm 6.2].

Thm 3.18. 凸函数在其定义域中相对内点上永远次可微。

Let $f: E \to (-\infty, \infty]$ be proper convex function. Let $\tilde{x} \in ri(dom(f))$.

Then $\partial f(\tilde{x}) \neq \phi$. [108, Thm 23.4].

$ri(dom(f)) \subseteq dom(\partial f)$.

Collary 3.19.

Let $f: E \to (-\infty, \infty]$ be a proper convex function. Then

$\exists x \in dom(f)$ s.t. $\partial f(x) \neq \phi$.

Thm 3.20. 在 $dim(dom(f)) < dim(E)$ 时，次微分集无界。

Let $f: E \to (-\infty, \infty]$ be a proper convex function. Suppose

that $dim(dom(f)) < dim(E)$, let $x \in dom(f)$.

If $\partial f(x) \neq \phi$, then $\partial f(x)$ is unbounded.

Proof: Let $y \in \partial f(x)$, y is arbitrary.

Define $V \equiv aff(dom(f)) - \{x\}$.

V is a vector space.

$dim(V) = dim(dom(f)) < dim(E)$

$\Rightarrow$ $\exists v \neq 0$, $v \in E$ s.t. $\langle v, w \rangle = 0$ for $\forall w \in V$. *

Take any $\beta \in \mathbb{R}$, for $\forall y \in \text{dom}(f)$ $y - x \in V$. $\Rightarrow$ ※ holds for $y-x$

$\Rightarrow f(y) \geq f(x) + \langle \eta, y-x \rangle = f(x) + \langle \eta + \beta v, y-x \rangle$

In that case, $\eta + \beta v \in \partial f(x)$, $\forall \beta$

$\Rightarrow$ Unboundness of $\partial f(x)$.

3.3. Directional Derivatives. 方向导数.

3.3.1. Def, properties.

Def: $f: E \to (-\infty, \infty]$ be proper, $x \in \text{int}(\text{dom}(f))$, The directional derivative of f at x given direction $d \in E$ is defined by:

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

note: No restriction on $\|d\|$: As a result of $\alpha \to 0^+$.

Thm 3.2.1 凸函数在定义域的内点上,任意方向上的方向导数均存在. Let $f: E \to (-\infty, \infty]$ be proper convex function. $x \in \text{int}(\text{dom}(f))$. Then for $\forall d \in E$, the directional derivative $f'(x; d)$ exists. [08, 23.1].

下面讨论 $d \mapsto f'(x; d)$

Lemma 3.22.

Let $f: E \to (-\infty, \infty]$ be a proper convex function and let $x \in \text{int}(\text{dom}(f))$. Then

(a) Then function $d \mapsto f'(x, d)$ is convex.

(b) $\forall \lambda \geq 0$, $d \in E$, $f'(x; \lambda d) = \lambda f'(x; d)$ 一阶齐次.

Proof: (a)

Take $d_1, d_2 \in E$, $\lambda \in [0, 1]$.

$$f'(x, \lambda d_1 + (1-\lambda) d_2)$$

$$= \lim_{\alpha \to 0^+} \frac{f(x + \alpha(\lambda d_1 + (1-\lambda) d_2)) - f(x)}{\alpha}$$

$$= \lim_{\alpha \to 0^+} \frac{f(\lambda(x + \alpha d_1) + (1-\lambda)(x + \alpha d_2)) - f(x)}{\alpha}$$

f is convex

$$\leq \lim_{\alpha \to 0^+} \frac{\lambda(f(x + \alpha d_1) - f(x)) + (1-\lambda)(f(x + \alpha d_2) - f(x))}{\alpha}$$

$$= \lambda \lim_{\alpha \to 0^+} \frac{f(x + \alpha d_1) - f(x)}{\alpha} + (1-\lambda) \lim_{\alpha \to 0^+} \frac{f(x + \alpha d_2) - f(x)}{\alpha}$$

$$= \lambda f'(x, d_1) + (1-\lambda) f'(x, d_2) \Rightarrow d \mapsto f'(x; d) \text{ is convex.}$$

(b). $f'(x; \lambda d)$

$$= \lim_{\alpha\to 0^+} \frac{f(x+\alpha\lambda d) - f(x)}{\alpha}$$

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

Lemma 3.23. 凸函数 函数值 与 方向导数 关系.

Let $f: E \to (-\infty, \infty]$ be a proper convex function, let $x \in \mathrm{int}(\mathrm{dom} f)$. Then

$$f(y) \geq f(x) + f'(x; y-x), \quad \forall y \in \mathrm{dom}(f).$$

proof: $f'(x; y-x) = \lim_{\alpha\to 0^+} \frac{f(x+\alpha(y-x)) - f(x)}{\alpha}$ $\quad \Downarrow$ f is convex

$$\leq \lim_{\alpha\to 0^+} \frac{(1-\alpha)f(x) + \alpha f(y) - f(x)}{\alpha}$$

$$= \lim_{\alpha\to 0^+} \frac{\alpha(f(y) - f(x))}{\alpha} = f(y) - f(x).$$

$$\Rightarrow f(y) \geq f(x) + f'(x; y-x).$$

Thm 3.24. 无需凸假设, 最大函数的方向导数.

$f(x) = \max\{f_1(x), \dots f_m(x)\}$, $f_1 \sim f_m: E \to (-\infty, \infty]$ are proper. Let $x \in \bigcap_{i=1}^m \mathrm{int}(\mathrm{dom}(f_i))$ and $d \in E$. Assume $f'(x_i; d)$ exists for $\forall i \in 1 \sim m$. Then

$$f'(x; d) = \max_{i \in I(x)} f_i'(x, d)$$

where $I(x) = \{i : f_i(x) = f(x)\}$.

proof: $\lim_{t\to 0^+} f_i(x+td) = \lim_{t\to 0^+} \left(f_i(x) + t \cdot \frac{f_i(x+td) - f_i(x)}{t}\right)$

$= f_i(x) + 0 \cdot f_i'(x) = f_i(x). \quad \forall i \in [m].$

For $i \in I(x)$, $j \notin I(x)$, by def of $f(x)$ & $I(x)$,

$$f_i(x) > f_j(x)$$

$$\Rightarrow \lim_{t\to 0}[f_i(x+td) - f_j(x+td)] > 0.$$

$\exists\, \varepsilon > 0$ s.t. $t \in (0, \varepsilon]$ implies $f_i(x+td) > f_j(x+td)$

Then, for $\forall t \in (0, \varepsilon]$,

we have: $f(x+td) = \max_{i:[m]} f_i(x+td) = \max_{i\in I(x)} f_i(x+td)$

$$\frac{f(x+td) - f(x)}{t} = \max_{i\in I(x)} \frac{f_i(x+td) - f_i(x)}{t}$$

$$\Rightarrow f'(x; d) = \lim_{t\to 0^+} \frac{f(x+td) - f(x)}{t}$$

$$= \max_{i\in I(x)} \lim_{t\to 0^+} \frac{f_i(x+td) - f_i(x)}{t}$$

$$= \max_{i\in I(x)} f_i'(x; d).$$

Note: the assumption of Thm 3.24 : $f_i'(x; d)$ exists.

$\Uparrow$

Obviously holds when f_i are convex.

Collary 3.25.

$f(x) = \max\{f_1(x), \cdots, f_m(x)\}$. $f_1, f_2, \cdots f_m : E \to (-\infty, \infty]$ are proper convex functions. Let $x \in \bigcap_{i=1}^{m} \mathrm{int}(\mathrm{dom}(f_i))$, $d \in E$.

Then $$f'(x;d) = \max_{i \in I(x)} f_i'(x;d)$$

where $I(x) = \{i : f_i(x) = f(x)\}$.

3.3.2 The Max Formula

— Connects $\begin{cases} \text{subgradients} \\ \text{directional derivatives.} \end{cases}$

Thm 3.26 Max Formula. 写成 support function notation: $f'(x;d) = \sigma_{\partial f(x)}(d)$.

$f : E \to (-\infty, \infty]$ is a proper convex function. Then for $\forall x \in \mathrm{int}(\mathrm{dom}(f))$ and $d \in E$,

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

proof: Let $x \in \mathrm{int}(\mathrm{dom}(f))$. By subgradient inequality,

$\forall d$, $f'(x;d) = \lim_{\alpha \to 0^+} \frac{1}{\alpha}(f(x+\alpha d) - f(x)) \geq \lim_{\alpha \to 0^+} \frac{1}{\alpha} \cdot \langle g, \alpha d\rangle = \langle g, d\rangle$

$\Rightarrow$ Consequently, $f'(x;d) \geq \max\{\langle g, d\rangle : g \in \partial f(x)\}$. ①

Then we only need to show the reverse inequality.

Consider $h(w) \equiv f'(x;w)$, By Lemma 3.21(a),

h is convex, thus subdifferential over E (Corollary 3.15)

Let $\tilde{g} \in \partial h(d)$, for $\forall v \in E$ and $\alpha \geq 0$, using Lemma 3.22(b), we have

$$\alpha f'(x;v) = f'(x;\alpha v) = h(\alpha v) \geq \underset{\overset{\|}{f'(x;d)}}{h(d)} + \langle \tilde{g}, \alpha v - d \rangle$$

$$\Rightarrow \alpha(f'(x;v) - \langle \tilde{g}, v \rangle) \geq f'(x;d) - \langle \tilde{g}, d \rangle \quad *$$

Then $f'(x;v) - \langle \tilde{g}, v \rangle \geq 0$, $\forall v$. (To guarantee the above inequality holds).

$$f'(x;v) \geq \langle \tilde{g}, v \rangle$$

Then by Lemma 3.23, $\forall y \in \mathrm{dom}(f)$,

$$f(y) \geq f(x) + f'(x; y-x) \geq f(x) + \langle \tilde{g}, y-x \rangle.$$

$$\Rightarrow \tilde{g} \in \partial f(x).$$

Take $\alpha = 0$ in $*$, we have

$$f'(x;d) \leq \langle \tilde{g}, d \rangle \leq \max\{\langle g, d \rangle : g \in \partial f(x)\}. \quad ②$$

Combine ① and ②,

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

3.3.3. Differentiability.

Def 3.28.

$f: E \to (-\infty, \infty]$, $x \in \mathrm{int}(\mathrm{dom}(f))$.

f is <u>differentiable</u> at x if

$\exists\, g \in E^*$ s.t. $\lim\limits_{h \to 0} \dfrac{f(x+h) - f(x) - \langle g, h \rangle}{\|h\|} = 0.$

g is called the gradient of f at x, denoted by $\nabla f(x)$.

Note: g is unique.

(Suppose g_1, g_2. $\Rightarrow \lim_{h\to 0} \frac{\langle g_1 - g_2, h\rangle}{\|h\|} = 0 \Rightarrow g_1 = g_2$)

Thm 3.28 方向导数与梯度.

$f: E \to [-\infty, \infty]$ be proper, and suppose f is differentiable at $x \in \operatorname{int}(\operatorname{dom}(f))$. Then for $\forall d \in E$,

$$f'(x; d) = \langle \nabla f(x), d\rangle.$$

proof: $d = 0 \Rightarrow$ Obviously correct.

$d \neq 0$. Since f is differentiable at x,

$$\lim_{\alpha\to 0^+} \frac{f(x+\alpha d) - f(x) - \langle \nabla f(x), \alpha d\rangle}{\|\alpha d\|} = 0.$$

$$\frac{1}{\|d\|} f'(x; d) = \frac{1}{\|d\|} \lim_{\alpha\to 0^+} \frac{\langle \nabla f(x), \alpha d\rangle}{\|\alpha\|}$$

$$f'(x; d) = \langle \nabla f(x), d\rangle.$$

Eg 1. (Thm 3.24 进一步)

$f(x) = \max_{i=1\sim m} f_i(x) \Rightarrow \forall d \in E,\ f'(x; d) = \max_{i\in I(x)} f_i'(x; d)$

$$= \max_{i\in I(x)} \langle \nabla f_i(x), d\rangle.$$

$I(x) = \{i : f_i(x) = f(x)\}$.

Eg 2. Gradient of $\frac{1}{2} d_C^2(\cdot)$.

E: Euclidean space. $C \subseteq E$, $C \neq \emptyset$. C is $\begin{cases} \text{closed} \\ \& \\ \text{convex.} \end{cases}$

Consider $\varphi_C : E \to \mathbb{R}$ given by $\varphi_C(x) = \frac{1}{2} d_C^2(x) = \frac{1}{2}\|x - P_C(x)\|^2$

P_C : orthogonal projection,
defined by $P_C(x) = \underset{y \in C}{\arg\min} \|y - x\|$.

Actually, P_C is well-defined (exists & unique)
when $C \neq \phi$, closed and convex. We show that $\forall x \in E$,

$$\nabla \varphi_C(x) = x - P_C(x).$$

proof: Fix x, define $g_x(d) = \varphi_C(x+d) - \varphi_C(x) - \langle d, z_x \rangle$

$z_x = x - P_C(x)$.

Then we only need to show $\lim_{d \to 0} \frac{g_x(d)}{\|d\|} = 0$

That is, $\frac{g_x(d)}{\|d\|} \to 0$ as $d \to 0$.

By def of $P_C(x)$, we have, $\forall d \in E$

$$\|x+d - P_C(x+d)\|^2 \leq \|x+d - P_C(x)\|^2$$

Then $g_x(d) = \frac{1}{2}\|x+d - P_C(x+d)\|^2 - \frac{1}{2}\|x - P_C(x)\|^2 - \langle d, z_x \rangle$

$$\leq \frac{1}{2}\|x+d - P_C(x)\|^2 - \frac{1}{2}\|x - P_C(x)\|^2 - \langle d, z_x \rangle$$

$$= \frac{1}{2}\|x - P_C(x)\|^2 + \underbrace{\langle x - P_C(x), d \rangle} + \frac{1}{2}\|d\|^2 - \frac{1}{2}\|x - P_C(x)\|^2 \underbrace{- \langle d, z_x \rangle}$$

$$= \frac{1}{2}\|d\|^2 \quad ①$$

$$\frac{g_x(d)}{\|d\|} \leq \frac{1}{2}\|d\|$$

In particular, we have $g_x(-d) \leq \frac{1}{2}\|d\|^2$

φ_C is convex $\Rightarrow g_x$ is convex.

$$0 = g_x\left(\tfrac{1}{2}(-d) + \tfrac{1}{2}d\right) \le \tfrac{1}{2}g_x(-d) + \tfrac{1}{2}g_x(d)$$

$$\Rightarrow g_x(d) \ge -\tfrac{1}{2}g_x(-d) \ge -\tfrac{1}{2}\|d\|^2 \qquad ②$$

Combine ①, ②,

$$\|g_x(d)\| \le \tfrac{1}{2}\|d\|^2 \quad \Rightarrow \quad \frac{|g_x(d)|}{\|d\|} \le \tfrac{1}{2}\|d\| \to 0$$

as $d \to 0$.

Done.

Remark 3.32. What is the gradient?

The gradient depends on the choice of the inner product in the underlying space.

When f is differentiable at x, then

$$\underbrace{(\nabla f(x))_i}_{\text{ith component of } \nabla f(x)} = \langle \nabla f(x), e_i \rangle = \underbrace{f'(x; e_i)}_{\frac{\partial f}{\partial x_i}(x)}. \qquad x_i = \langle x, e_i \rangle.$$

$$\Rightarrow \quad \nabla f(x) = D_f(x) = \begin{pmatrix} \frac{\partial f}{\partial x_1}(x) \\ \frac{\partial f}{\partial x_2}(x) \\ \vdots \\ \frac{\partial f}{\partial x_n}(x) \end{pmatrix}$$

$$\Rightarrow \underline{f'(x; d) = \langle \nabla f(x), d \rangle = D_f^T(x)\, d = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(x)\, d_i.}$$

This holds for any choice of inner product in the space.

But, there might be $\nabla f(x) \neq Df(x)$ when $\langle , \rangle$ is not dot product.

e.g. $\langle x, y \rangle = x^T H y$. H $n \times n$, $H \succeq 0$.

$$(\nabla f(x))_i = \nabla f(x)^T e_i = \nabla f(x)^T H (H^{-1} e_i)$$
$$= \langle \nabla f(x), H^{-1} e_i \rangle$$
$$= f'(x; H^{-1} e_i)$$
$$= Df(x)^T H^{-1} e_i$$

$\Rightarrow$ $\nabla f(x) = H^{-1} Df(x)$, a "weighted" gradient.

Now, consider $E = \mathbb{R}^{m \times n}$.

$\langle X, Y \rangle = \mathrm{Tr}(X^T Y)$ for $\forall X, Y \in \mathbb{R}^{m \times n}$.

Given a proper $f: \mathbb{R}^{m \times n} \to (-\infty, \infty]$. $X \in \mathrm{int}(\mathrm{dom}(f))$.

Then $\nabla f(x) = Df(x) = \left(\frac{\partial f}{\partial x_{ij}} \right)_{ij}$

If $\langle X, Y \rangle = \mathrm{Tr}(X^T H Y)$,

then $\nabla f(x) = H^{-1} Df(x)$.

Thm 3.33. 凸函数可导处的次微分集：单元素集合 $\{\nabla f(x)\}$.

Let $f: E \to (-\infty, \infty]$ be proper, convex function. $x \in \mathrm{int}(\mathrm{dom}(f))$.

If f is differentiable at x, then $\partial f(x) = \{\nabla f(x)\}$.

If f has a unique subgradient at x, then it is differentiable at x and $\partial f(x) = \{\nabla f(x)\}$.

Proof: Let $g \in \partial f(x)$, we show that $g = \nabla f(x)$.

By Max Formula, $\forall d \in E$,

$$f'(x;d) = \langle \nabla f(x), d \rangle \geq \langle g, d \rangle$$

$$\Rightarrow \langle g - \nabla f(x), d \rangle \leq 0, \quad \forall g \in \partial f(x)$$

$$\Rightarrow \|g - \nabla f(x)\|_* = \max_{\|d\| \leq 1} \langle g - \nabla f(x), d \rangle \leq 0 \quad \Rightarrow \|g - \nabla f(x)\|_* = 0$$

$g = \nabla f(x)$. $\Rightarrow \partial f(x) = \{\nabla f(x)\}$.

For the reverse, suppose the unique subgradient of f at x is g. we show that $\nabla f(x) = g$. By def of differentiability, to show

$$\lim_{u \to 0} \frac{f(x+u) - f(x) - \langle g, u \rangle}{\|u\|} = 0.$$

Let $h(u) = f(x+u) - f(x) - \langle g, u \rangle$. $\Rightarrow$ To show $\lim_{u \to 0} \frac{h(u)}{\|u\|} = 0$

$f(x+u) \geq f(x) + \langle g, u \rangle$

0 is the unique subgradient of h at 0 and $0 \in \operatorname{int}(\operatorname{dom}(f))$.

By Max Formula, $\forall d \in E$, $h'(0;d) = 0 = \lim_{\alpha \to 0^+} \frac{h(\alpha d) - h(0)}{\alpha}$

$$= \lim_{\alpha \to 0^+} \frac{h(\alpha d)}{\alpha}$$

Let $\{v_1, v_2, \cdots v_k\}$ be an orthonormal basis of E,

$0 \in \operatorname{int}(\operatorname{dom}(h)) \Rightarrow \exists \varepsilon \in (0,1)$ s.t. $\varepsilon v_i, -\varepsilon v_j \in \operatorname{dom}(h)$.

Since $\operatorname{dom}(h)$ is convex, the set

$D = \operatorname{conv}(\{\pm \varepsilon v_i\}_{i=1}^{k})$. satisfied $D \subseteq \operatorname{dom}(h)$.

$B_{\|\cdot\|}[0, \nu] \subseteq D$, where $\nu = \frac{\varepsilon}{K}$.

Let $w \in B_{\|\cdot\|}[0, \nu]$,

then we have $w = \sum_{i=1}^{k} \langle w, V_i \rangle V_i$

and $\|w\|^2 = \sum_{i=1}^{k} \langle w, V_i \rangle^2 \leq \nu^2$

Then $|\langle w, V_i \rangle| \leq \nu$ must hold.

Hence

$$w = \sum_{i=1}^{k} \langle w, V_i \rangle V_i$$

Eg. subdifferential of the ℓ_2-norm.

$f: \mathbb{R}^n \to \mathbb{R}$ given by $f(x) = \|x\|_2$.

$x=0 \Rightarrow$ Given by eg 3.3. $g \in \partial f(0) \Leftrightarrow g \in \{g \in (\mathbb{R}^n)^*: \|g\|_* \le 1\}$

单位闭球.

$x \neq 0$. $\nabla f(x) = \frac{x}{\|x\|_2}$, f is obviously differentiable.

so, $\partial f(x) = \begin{cases} B_{\|\cdot\|}[0,1] & x=0. \\ \left\{\frac{x}{\|x\|_2}\right\} & x \neq 0. \end{cases}$

Specifically, consider $n=1$, $f(x) = |x|$.

$\partial f(x) = \begin{cases} [-1, 1] & x=0. \\ \{\operatorname{sgn}(x)\} & x \neq 0. \end{cases}$