Last time:

1. Stationary points of $\inf_{x \in C} f(x)$

$$-\nabla f(\bar{x}) \in N_C(\bar{x}) \qquad \text{(OPT)}$$

2. Thm: $\bar{x}$ satisfies OPT if, and only if, $(\forall \gamma > 0)$ $\bar{x} = P_C(\bar{x} - \gamma \nabla f(\bar{x}))$.

3.

Algorithm: Projected gradient method for f with ∇f L-Lipschitz

Input: $x^0 \in C$, $0 < \gamma < \frac{2}{L}$

For $k = 0, 1, \ldots$, do

$$x^{k+1} = P_C(x^k - \gamma \nabla f(x^k))$$

4. Thm: (sufficient optimality) Let $f \in \mathcal{F}(\mathbb{R}^n)$ and let
continuously diff'ble, convex

$C \subseteq \mathbb{R}^n$ be a closed convex set.

Then $-\nabla f(x) \in N_C(x) \iff x \in \operatorname{argmin}_{y \in C} f(y)$.

5. Thm: Suppose, f and g are convex, f diff'ble, g continuous,

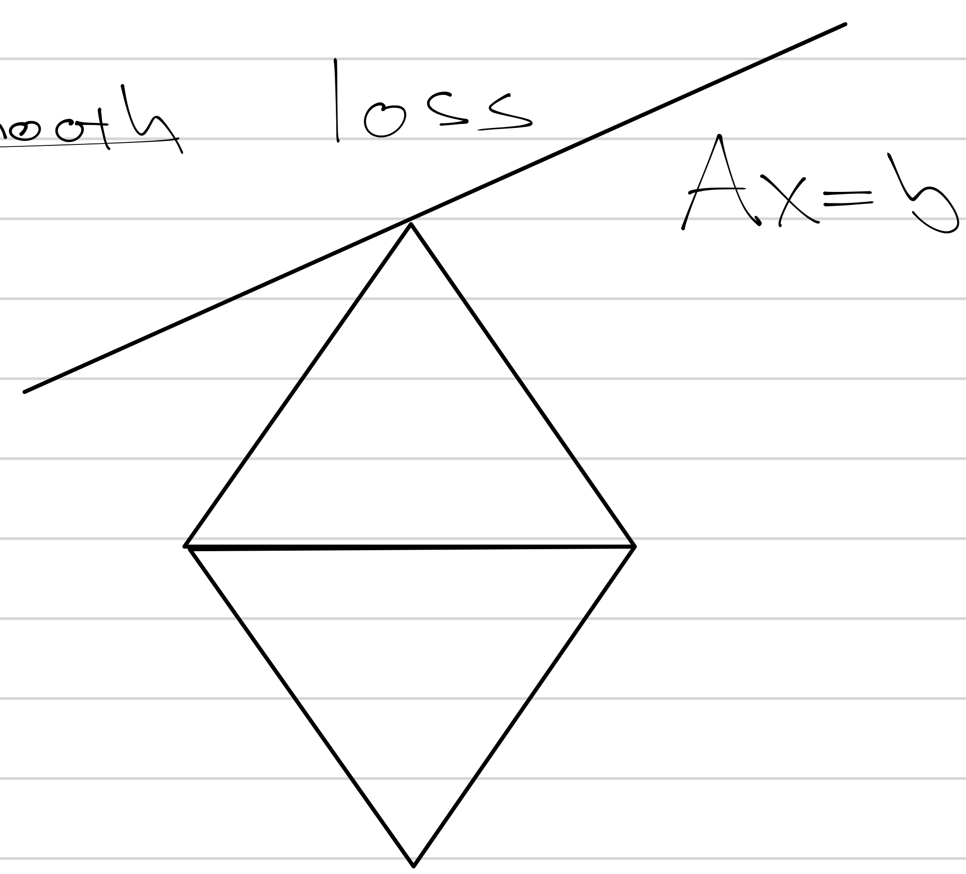
$\bar{x} \in \operatorname{argmin}\{f(x)+g(x)\} \iff 0 \in \nabla f(\bar{x}) + \partial g(\bar{x})$.

Today: Nonsmooth Convex functions

- Nonsmoothness is essential to accurate, expressive modeling in applied science.
- Example: Compressive Sensing

$$\min_{Ax=b} \|x\|_1$$

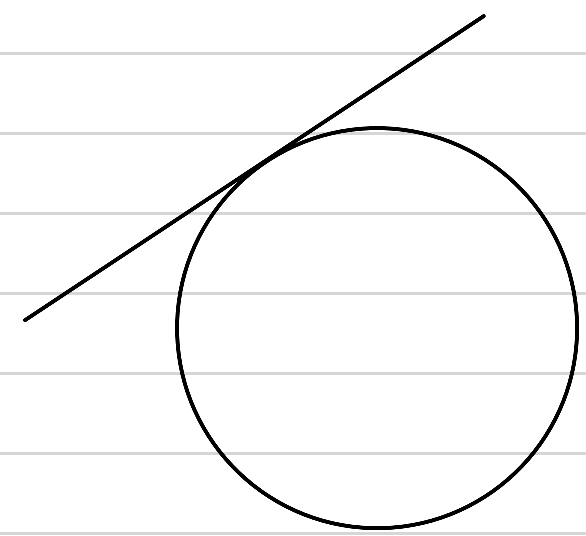
has a nonsmooth loss



Line much more likely to hit corner, than side

provides SPARSE solutions

In contrast



$$\min_{Ax=b} \|x\|_2$$

does not distinguish any type of point,

solutions never sparse.

Questions:

1. How do we differentiate nonsmooth problems? (Today)
2. How do we perform "projected gradient" descent on nonsmooth problems?

Def: A function $g: \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is called proper if $\exists\, x \in \mathbb{R}^n$ s.t. $g(x) < \infty$

Assumption: We will only deal with convex sets!

Def: (convexity) A function $g: \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is called convex if

$$(\forall \alpha \in [0,1]) (\forall x,y \in \mathbb{R}^n) \quad g((1-\alpha)x + \alpha y) \leq (1-\alpha) g(x) + \alpha g(y).$$

Def: (lower semicontinuity / closedness) A function $g: \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is called lower-semicontinuous, or closed, if

$$\operatorname{epi}(g) := \{ (x,t) \in \mathbb{R}^{n+1} \mid g(x) \leq t \}$$

is a closed set. We write

$$\operatorname{dom}(g) := \{ x \in \mathbb{R}^n \mid g(x) < \infty \}.$$

Examples:

1. $C \subseteq \mathbb{R}^n$ closed. Then

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

is closed. If C is convex, then ι_C is convex.

2. $g: \mathbb{R}^m \to \mathbb{R}$ closed, convex. Let $A \in \mathbb{R}^{m \times n}$. Then $g(Ax)$ is convex

3. Sum of convex functions convex.

4. $f:\mathbb{R}^n\times\mathbb{R}^m\to\mathbb{R}\cup\{\infty\}$, $f(\cdot,y)$ closed, convex $(\forall y)$ $\Rightarrow$ $g(x):=\sup_y f(x,y)$ is closed convex.

5. $\exp(x)$, $\log(1+\exp(x))$, $\frac{1}{2}\|Ax-b\|^2$, $\|x\|_p$ for $p\in[1,\infty]$,
$-\sqrt{x}$, $-\log(x)$, $\max\{0,1-x\}$ (hinge), ...

6. Convex but not closed: $g(x,y)=\begin{cases}0 & \text{if } x^2+y^2<1\\ \phi(x,y) & \text{if } x^2+y^2=1\end{cases}$ | ϕ arbitrary nonnegative!

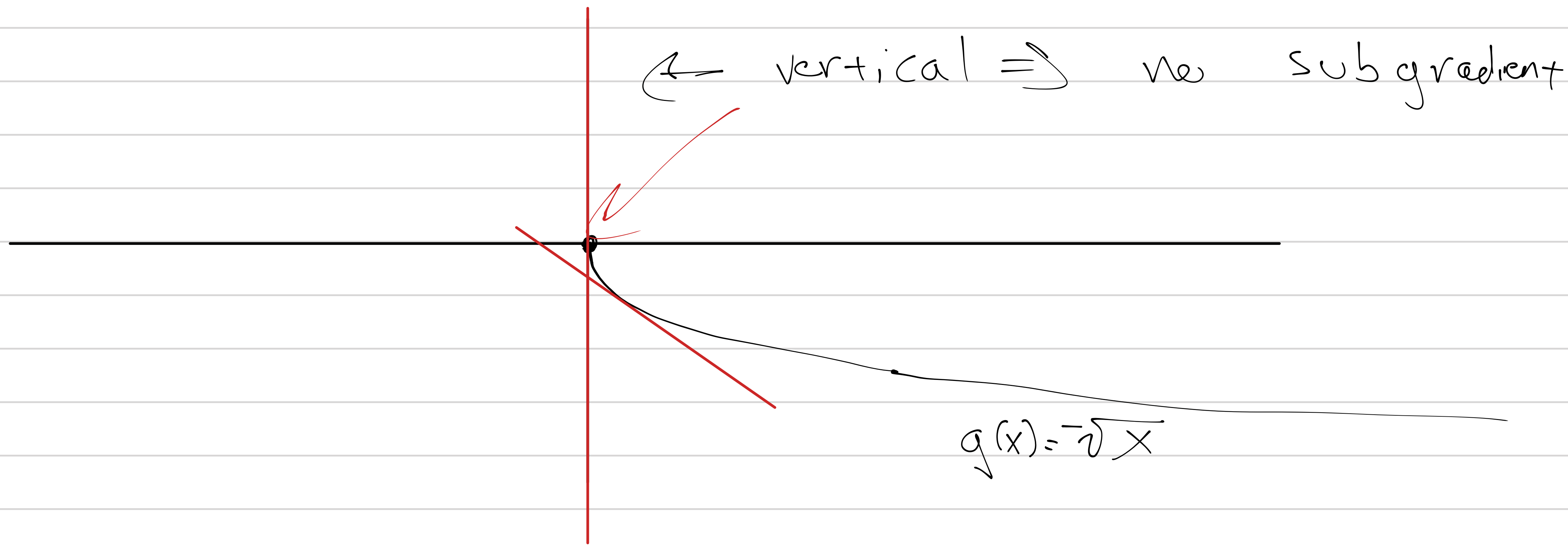
Differentiating nonsmooth ∞-valued functions

Def: Let $g:\mathbb{R}^n\to\mathbb{R}\cup\{\infty\}$ be a function. The set

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

is called the convex subdifferential operator of g.

$v\in\partial g(x)$ are called subgradients.



Example:

Let C be closed, convex. Then

$$\partial \iota_C(x) = N_C(x) \text{ !}$$

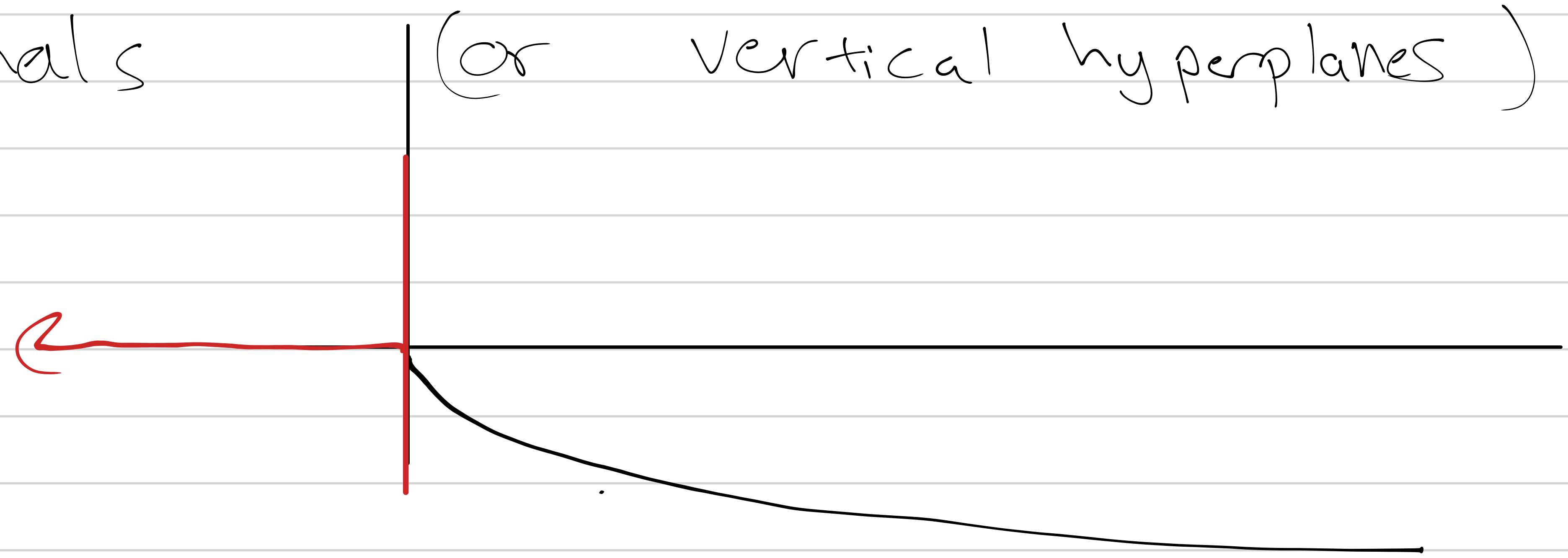
So we've been studying subgradients since week 3.

Conversely, you showed on homework 3, problem 3, that

$$\begin{bmatrix} v \\ -1 \end{bmatrix} \in N_{\mathrm{epi}(g)}(x) \iff v \in \partial g(x).$$

So subdifferential operators are deeply connected to normal cones of epigraphs.

When $\begin{bmatrix} v \\ 0 \end{bmatrix} \in N_{epi(g)}(x)$ we get horizontal normals (or vertical hyperplanes)



Def: For closed, convex $g: \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$, the <u>horizon subdifferential</u> is the set-valued operator

$$\partial^\infty g(x) = \left\{ v \in \mathbb{R}^n \;\middle|\; \begin{bmatrix} v \\ 0 \end{bmatrix} \in N_{epi(g)}(x, g(x)) \right\}$$

Exercise: $\partial^\infty g(x) = N_{dom(g)}(x)$.

Thm: Let $g: \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ be closed, convex.

Then $\forall\, x \in \operatorname{dom}(g)$

$$N_{\operatorname{epi}(g)}(x, g(x)) = \left\{ \lambda \begin{bmatrix} v \\ -1 \end{bmatrix} \;\middle|\; v \in \partial g(x),\ \lambda > 0 \right\} \cup \left\{ \begin{bmatrix} v \\ 0 \end{bmatrix} \;\middle|\; v \in \partial^{\infty} g(x) \right\}$$

Pf:

We only need to show that

$\neg \exists\, v \in \mathbb{R}^n$ s.t.

$$\begin{bmatrix} v \\ 1 \end{bmatrix} \in N_{\operatorname{epi}(g)}(x, g(x)).$$

Suppose for contradiction that such a v does exist.

Then

$$\begin{aligned} 0 &\geq \left\langle \begin{bmatrix} v \\ 1 \end{bmatrix}, (x, g(x)+1) - (x, g(x)) \right\rangle \\ &= (g(x)+1) - g(x) \\ &= 1. \end{aligned}$$

Which is a contradiction $\boxtimes$

Corollary: Let $g: \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ be closed, convex. Then

1. $(\forall x \in \mathrm{dom}(g))\ \partial g(x) \cup (\partial^\infty g(x) \setminus \{0\}) \neq \emptyset$
2. $(\forall x \in \mathrm{int}(\mathrm{dom}(g)))\ \partial g(x) \neq \emptyset$

Pf:

1. $(x, g(x)) \in \mathrm{Boundary}(\mathrm{epi}(f))$, so

   $N_{\mathrm{epi}(g)}(x, g(x)) \neq \{0\}$

2. $\partial^\infty g(x) = N_{\mathrm{dom}(g)}(x) = \{0\}$

   So by (1), $\partial g(x) \neq \emptyset$. $\boxtimes$

- So subgradients exist at any point in $\mathrm{dom}(g)$.
- Subgradients are either vertical (horizon) or non vertical.

Subdifferentials and directional derivatives

To compute with subdifferentials, it's extremely useful to relate subdifferentials to directional derivatives.

Def: Let $x \in \mathrm{dom}(g)$. We call g differentiable at x in the direction of p if

$$g'(x;p) := \lim_{\alpha \downarrow 0} \frac{g(x+\alpha p) - g(x)}{\alpha}$$

exists.

Lemma: Let g be closed, convex. Let $x \in \mathrm{int}(\mathrm{dom}(g))$. Then $g(x;p)$ exists $\forall p \in \mathbb{R}^n$.

Pf: Let $0 < \alpha \leq \beta < \infty$. Then let $z = x + \beta p$, $\lambda = \frac{\alpha}{\beta}$

$$r(\alpha) = \frac{1}{\alpha}\left[g(x+\alpha p) - g(x)\right] = \frac{1}{\alpha}\left[g(\lambda z + (1-\lambda)x) - g(x)\right] \leq \frac{1}{\alpha}\left[\lambda g(z) + (1-\lambda) g(x) - g(x)\right]$$

$$= \frac{1}{\alpha}\left[\lambda g(x+\beta p) - \lambda g(x)\right]$$

$$= r(\beta).$$

Further, $r(\alpha) \geq \frac{1}{\alpha}\left[g(x+\alpha p) - g(x)\right] \geq \langle p, v \rangle$ $(\forall v \in \partial g(x))$.

exist b/c $x \in \mathrm{int}(\mathrm{dom}(g))$

So limit must exist. $\boxtimes$

Proposition: Let g be closed, convex. Let $x \in \text{int}(\text{dom}(g))$.

Then 1. $g'(x;\cdot)$ is convex, homogeneous function of degree one

2. $(\forall y \in \mathbb{R}^n)$ $g(y) \geq g(x) + g'(x; y-x)$

Pf:

1. Exercise!

2. Define $y_\alpha = (1-\alpha)x + \alpha y$

$$g(y_\alpha) \leq (1-\alpha) g(x) + \alpha g(y)$$

$$\Rightarrow \frac{1}{\alpha}(1-\alpha)[g(y_\alpha) - g(x)] + g(y_\alpha) \leq g(y)$$

$$\Rightarrow g(y) \geq \lim_{\alpha \downarrow 0} \left\{ g(y_\alpha) + \frac{(1-\alpha)}{\alpha}[g(y_\alpha) - g(x)] \right\}$$

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

$\boxtimes$

Finally, we find the following exact relation between $g'(x;p)$ and $\partial g(x)$

Thm: (Max formula) Let $g: \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ be closed, convex and suppose $x \in \operatorname{int}(\operatorname{dom}(g))$. Then

$$(\forall p \in \mathbb{R}^n) \qquad g'(x;p) = \sup\{\langle v, p\rangle \mid v \in \partial g(x)\}$$

Pf: $\forall v \in \partial g(x)$,

$$g'(x;p) = \lim_{\alpha \downarrow 0} \frac{1}{\alpha}\left[g(x+\alpha p) - g(x)\right] \geq \frac{1}{\alpha}\langle v, \alpha p\rangle = \langle v, p\rangle = \langle v,p\rangle + g'(x;0) \qquad (*)$$

$\Rightarrow \; v \in \partial_p g'(x;0)$.

Moreover, $g(y) \geq g(x) + g'(x;y-x) \geq g(x) + \langle w, y-x\rangle$,

where $v \in \partial_p g'(x;0) \Rightarrow w \in \partial g(x)$. Thus, $\underline{\partial f(x) = \partial_p f'(x;0)}$.

We claim that $\partial_p g'(x;p) \subseteq \partial_p g'(x;0) \quad \forall p \in \mathbb{R}^n$.

Let $w \in \partial_p g'(x;p)$. Then $\forall \bar{p} \in \mathbb{R}^n$, $\tau > 0$, we have

$$\tau g'(x;\bar{p}) = g'(x;\tau\bar{p}) \geq g'(x;p) + \langle w, \tau\bar{p} - p\rangle$$

Take $\tau \to \infty$ to get $g'(x;\bar{p}) \geq \langle w, \tau\bar{p}\rangle \Rightarrow w \in \partial_p g'(x;0) = \partial g(x)$

Take $\tau \to 0$ + get $g'(x;p) \leq \langle w, p\rangle$

But $(*) \Rightarrow g'(x,p) \geq \langle w, p\rangle$ so $g'(x;p) = \langle w,p\rangle = \sup\{\langle v,p\rangle \mid v \in \partial g(x)\}$

$\boxtimes$

Next time we will use the max formula for developing subdifferential calculus.