Subgradientes / subdiferencial

P: $\min\limits_{x} c^t x$ s.a. $Ax = b$, $Dx \le e$, $x \in \mathbb{Z}^n_+$.

P(u): $\min\limits_{x} \underbrace{c^t x + u^t (Ax - b)}_{L(x,u)}$ s.a. $Dx \le e$, $x \in \mathbb{Z}^n_+$

$$L^*(u) = \min_{x} \{ c^t x + u^t (Ax - b) \; ; \; Dx \le e, \; x \in \mathbb{Z}^n_+ \}$$

<u>Objetivo</u>: resolver D: $\max\limits_{u} L^*(u)$.

Como $L^*: \mathbb{R}^m \to \mathbb{R}$ é? (m = nº de restrições dualizadas) (2)

Exemplo: $P: \min_x \; x_1 + 2x_2 + 3x_3$

$$\text{s.a.} \quad x_1 + x_2 + x_3 = 1$$
$$x \in \{0,1\}^3$$

Dualizando $\sum_1^3 x_i = 1$, obtemos

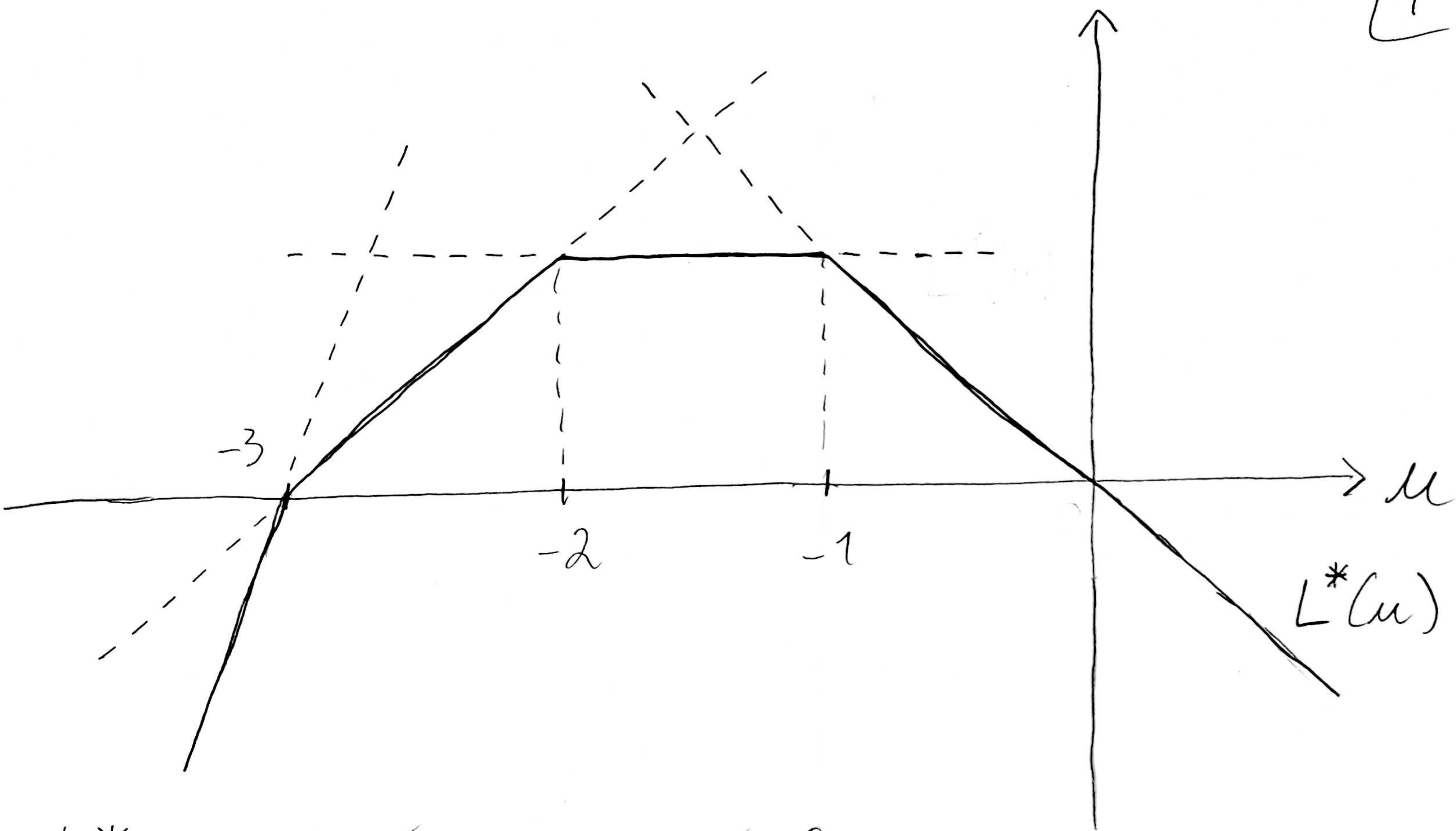
$$P(u) = \min_x \; (1+u)x_1 + (2+u)x_2 + (3+u)x_3 - u$$
$$\text{s.a.} \quad x \in \{0,1\}^3.$$

Resolvendo $P(u)$:

- $u \leq -3 \Rightarrow x_1 = x_2 = x_3 = 1$
- $-3 < u \leq -2 \Rightarrow x_1 = x_2 = 1, \quad x_3 = 0$
- $-2 < u \leq -1 \Rightarrow x_1 = 1, \quad x_2 = x_3 = 0$
- $u > -1 \Rightarrow x_1 = x_2 = x_3 = 0.$

Assim,

$$L^*(u) = \begin{cases} 6 + 2u, & u \leq -3 \\ 3 + u, & -3 < u \leq -2 \\ 1, & -2 < u \leq -1 \\ -u, & u > -1 \end{cases}$$



L^* não é diferenciável $\forall u$!

Em geral, L^* não é diferenciável, o que impede o uso de métodos que dependam de gradientes para resolver $D: \max_u L^*(u)$.

Porém, nem tudo está perdido:

<u>Teorema</u>: L^* é côncava, isto é, $-L^*$ é convexa.

Prova: Seja $X = \{ x \; ; \; Dx \leq e, \; x \in \mathbb{Z}_+^n \}$.

Dados $u, \tilde{u} \in \mathbb{R}^m$ e $t \in [0,1]$ temos

$$-L^*((1-t)u + t\tilde{u}) = -\min_x \{ c^t x + [(1-t)u + t\tilde{u}]^t (Ax-b) \; ; \; x \in X \}$$

$$= -\min_x \{ (1-t)\left[c^t x + u^t(Ax-b)\right] + t\left[c^t x + \tilde{u}^t(Ax-b)\right] \; ; \; x \in X \}$$

$$\leq -\Big[(1-t) \min_x \{ c^t x + u^t(Ax-b) \; ; \; x \in X \}$$

$$+ t \min_x \{ c^t x + \tilde{u}^t (Ax - b) ; x \in X \}]$$

$$= (1-t)(-L^*(u)) + t(-L^*(\tilde{u})) \quad \blacksquare$$

Assim, $D: \max_u L^*(u)$ é um problema de maximização de uma função côncava.

Para manter o padrão, note que isso é o mesmo que minimizar a função convexa $-L^*$ $\longrightarrow$ $D: -\min_u -L^*(u)$ (problema convexo).

· Como no exemplo, em geral L^* é o LS mínimo entre funções afins.

Para ver isto, suponha por simplicidade que $X = \{x ; Dx \leq e, x \in \mathbb{Z}_+^n\}$ seja finito. Assim, $X = \{x^1, x^2, \ldots, x^q\}$ e logo

$$L^*(u) = \min_{j=1,\ldots,q} \{c^t x^j + u^t(Ax^j - b)\}.$$

Neste caso, o problema D fica

$$D: \max_{u} \min_{j=1,\dots,q} \{ c^t x^j + u^t (Ax^j - b) \}$$

que equivale à

$$\begin{aligned} D: \max_{z} \ & z \\ \text{s.a.} \ & z \leq c^t x^j + u^t (Ax^j - b), \quad j = 1, \dots, q \\ & u \in \mathbb{R}^m, \quad z \in \mathbb{R}. \end{aligned}$$

Este PL sugere resolver D por uma estratégia de planos de corte, pois poderíamos resolver uma sequência de PL's inserindo restrições do tipo $z \leq c^t \tilde{x} + u^t(A\tilde{x} - b)$, com $\tilde{x}$ previamente calculado. (esta estratégia foi vista em "Otimização II").

Porém não queremos (e não precisamos) acumular restrições em grandes PL's...

↳ Vamos imitar o método do gradiente!

Sabemos que para uma função convexa diferenciável f, a condição de otimalidade de 1ª ordem é necessária e suficiente:

$$\nabla f(x^*) = 0 \iff x^* \text{ minimiza } f.$$

Porém, $-L^*$ não é diferenciável...

Como contornar?

$\hookrightarrow$ subgradientes!

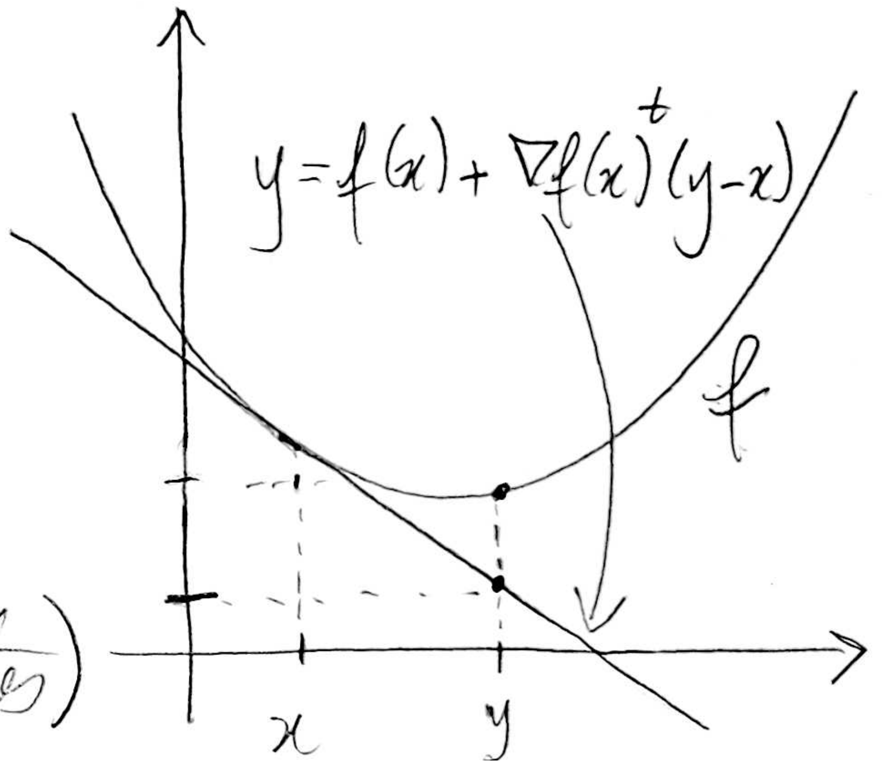
Subgradiente (de funções convexas)

Considere $f : \mathbb{R}^n \to \mathbb{R}$ convexa, não necessariamente diferenciável.

Sabemos ("Otimização I") que f diferenciável é convexa se, e somente se,

$$f(y) \geq f(x) + \nabla f(x)^t (y-x),$$

$\forall x, y$ (graf f acima das tangentes)



· O subgradiente imita o papel de ∇f nessa desigualdade:

<u>Definição</u>: o vetor $g \in \mathbb{R}^n$ é <u>subgradiente</u> de $f: \mathbb{R}^n \to \mathbb{R}$ no ponto x se

$$f(y) \geq f(x) + g^t(y-x), \quad \forall y.$$

Chamamos o conjunto dos subgradientes de f em x de <u>subdiferencial</u>, e o denotamos por $\partial f(x)$.

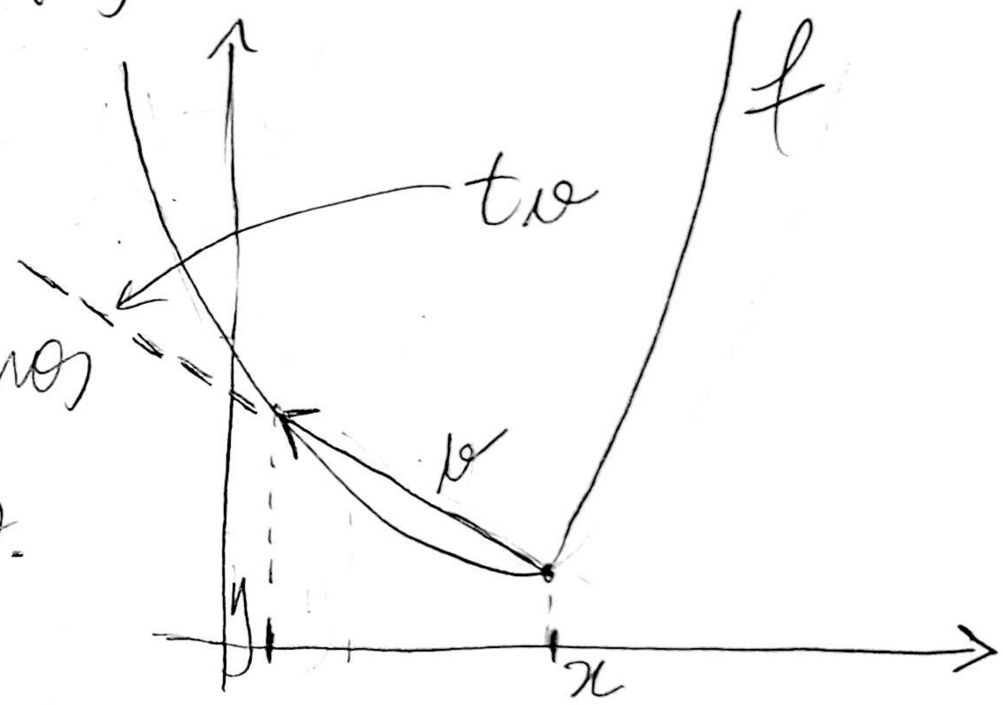
Interpretação geométrica

Tome x e $g \in \partial f(x)$ subgradiente.

Dado $y \neq x$, considere o vetor v que liga $(x, f(x))$ à $(y, f(y))$.

$$v = (y - x, f(y) - f(x)).$$

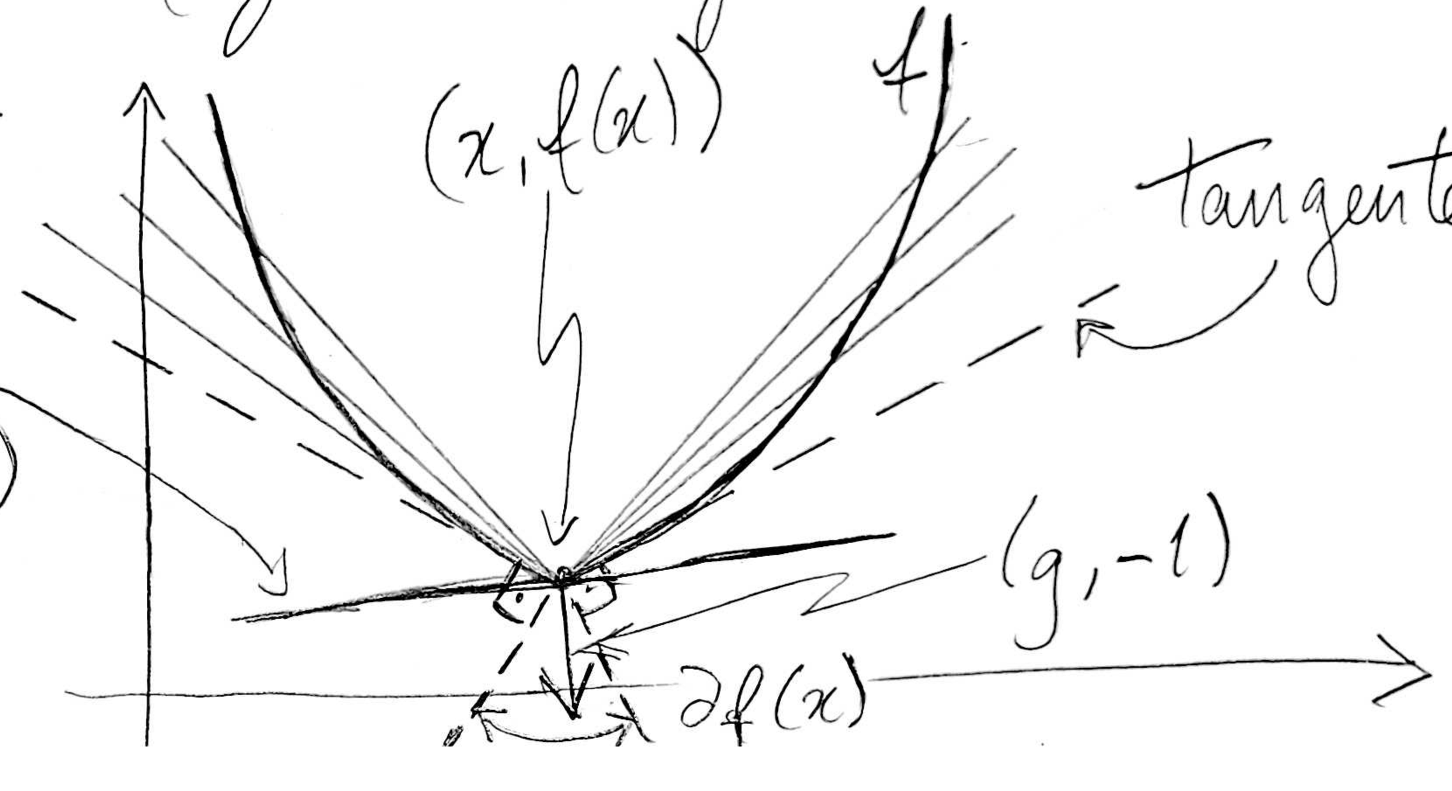
Dado $t > 0$, consideramos ainda a semireta tv.



Temos

$$\langle (g,-1); tv \rangle = t\underbrace{\left(g^t(y-x) - f(y) + f(x)\right)}_{\leq 0} \leq 0$$

Ou seja, tv faz um ângulo $\geq 90°$ com g, ~~$\forall t > 0$~~.



Assim, "g está associado à inclinação das retas que estão <u>entre</u> as tangentes à graf f".

Em outras palavras, graf f está acima das retas $y(x) = f(x) + g^t(y - x)$.

Na figura, $\partial f(x)$ está relacionado com o cone na "quina" do gráfico de f.

A interpretação geométrica usual é via epígrafo de f (ver livro Bertsekas).

Propriedades dos subgradientes / subdiferenciais.

1) Se f é diferenciável em x então

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

(subgradiente generaliza gradiente).

De fato, é claro que $g = \nabla f(x)$ é um subgradiente. E ele é único, pois neste só há uma tangente ao gráfico de f em x, aquela com normal $(\nabla f(x), -1)$.

2) Se f é convexa, então $\partial f(x) \neq \emptyset$ $\forall x \in \mathbb{R}^n$ (pense na interpretação geométrica para se convencer disso).

↳ Assim, um método que <u>só</u> use subgradientes estará bem definido para f convexa, <u>mesmo f não sendo diferenciável</u>.

3) Se f é côncava, também $\partial f(x) \neq \emptyset$, $\forall x \in \mathbb{R}^n$. Assim, L^* admite subgradientes $\forall u$!!!

Exercício: mostre que $\partial f(x) \neq \emptyset$ quando f é concava, lembrando que neste caso $-f$ é convexa.

4) Se f é convexa, $\partial f(x)$ é conjunto convexo. De fato, se $g, h \in \partial f(x)$ e $t \in [0,1]$, então

$$f(x) + [(1-t)g + th]^t (y-x)$$

$$= (1-t)\left[f(x) + g^t(y-x)\right] + t\left[f(x) + h^t(y-x)\right]$$

$$\leq (1-t) f(y) + t f(y) = f(y).$$

$$\Longrightarrow (1-t)g + th \in \partial f(x).$$

5) Se f é convexa, $\partial f(x)$ é conjunto fechado. De fato, seja $\{g_k\} \subset \partial f(x)$ uma sequência de subgradientes em x com $\lim_{k \to \infty} g_k = g_*$. Dado y, temos

$$f(y) \geq f(x) + g_k^t (y - x), \quad \forall k.$$

Fazendo $k \to \infty$ obtemos

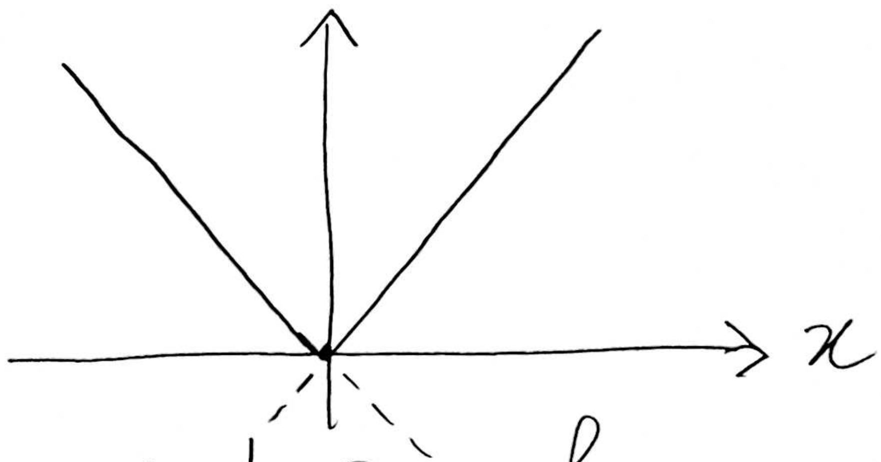
$$f(y) \geq f(x) + g_*^t (y - x), \quad \forall y \Rightarrow g_* \in \partial f(x)$$

$\blacksquare$

Obs: mais, $\partial f(x)$ é compacto.

Exemplos:

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



f é diferenciável em $x \neq 0$. Logo
$\partial f(x) = \{ f'(x) \}$, $x \neq 0$.

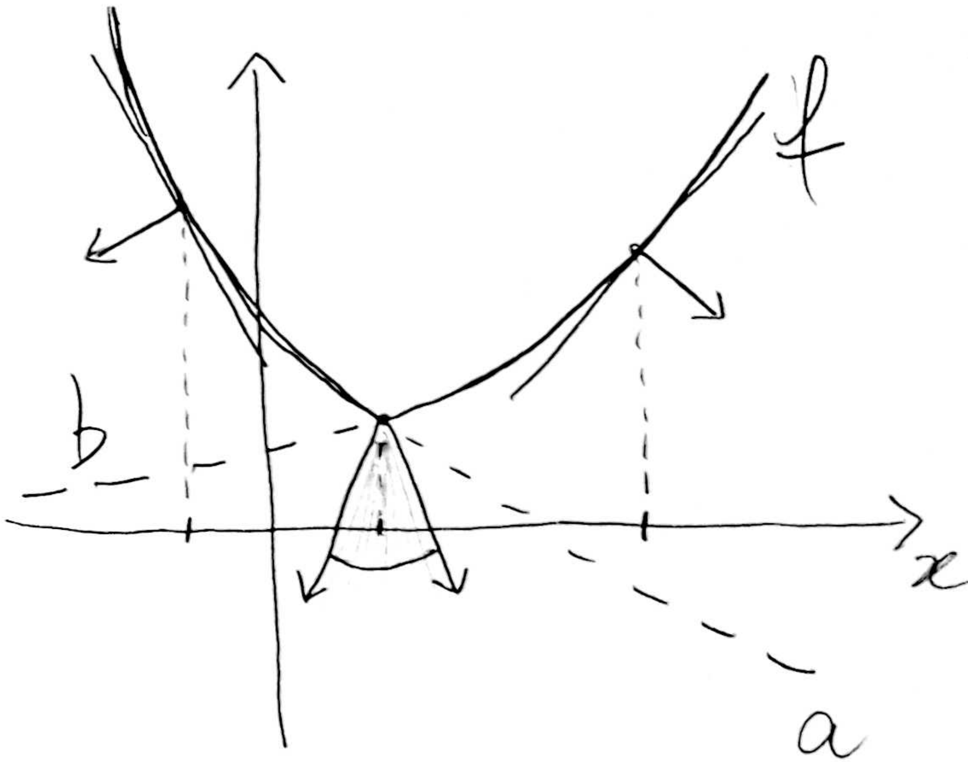
$\partial f(0)$: $f(y) \geq f(0) + g(y-0), \forall y \Leftrightarrow |y| \geq gy, \forall y$
$\Leftrightarrow -1 \leq g \leq 1$. Assim

$$\partial |x| = \begin{cases} -1, & x < 0 \\ [-1,1], & x = 0 \\ 1, & x > 0 \end{cases}$$

//

2) $a, b : \mathbb{R} \to \mathbb{R}$ convexas e diferenciáveis.

$f(x) = \max\{a(x), b(x)\}$ é convexa (verifique).



Então

- $b(x) < a(x) = f(x)$
  $\Rightarrow \partial f(x) = \{a'(x)\}$
- $a(x) < b(x) = f(x)$
  $\Rightarrow \partial f(x) = \{b'(x)\}$
- $a(x) = b(x) = f(x) \Rightarrow \partial f(x) = \{(1-t)a'(x) + t\,b'(x);\ t \in [0,1]\}$

Note que $|x| = \max\{-x, x\}$

(compare os dois exemplos anteriores). //

---

3) Em geral, se $f_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, \dots, p$, são convexas e diferenciáveis, então

$$f(x) = \max\{f_1(x), \dots, f_p(x)\}$$

é convexa, e

$$\partial f(x) = \operatorname{conv}\{\nabla f_i(x) \,;\ i \text{ tal que } f_i(x) = f(x)\}.$$

(fecho convexo dos gradientes das máximas f_i's)

Lembre-se, supondo $X = \{x ;\ Dx \leq e,\ x \in \mathbb{Z}^n_+\} = \{x^1, \dots, x^q\}$ finito, temos

$$L^*(u) = \min_{j=1,\dots,q} \{c^t x^j + u^t (A x^j - b)\}.$$

Logo, subgradientes de L^* têm relação com os gradientes das funções afins, dado que

$$-L^*(u) = \max_{j=1,\dots,q} \{-c^t x^j - u^t (A x^j - b)\}.$$