

# Subgradients / subdiferencial

1

$$P: \min_x c^T x \text{ s.a. } Ax = b, Dx \leq e, x \in \mathbb{Z}_+^n.$$

$$P(u) : \underbrace{\min_x c^T x + u^T (Ax - b)}_{L(x, u)} \text{ s.a. } Dx \leq e, x \in \mathbb{Z}_+^n$$

$$L^*(u) = \min_x \{ c^T x + u^T (Ax - b) ; Dx \leq e, x \in \mathbb{Z}_+^n \}$$

Objetivo: resolver  $D: \max_u L^*(u)$ .

• Como  $L^*: \mathbb{R}^m \rightarrow \mathbb{R}$  é? ( $m = n^o$  de restrições dualizadas) (2)

Exemplo: P:  $\min_{\boldsymbol{x}} x_1 + 2x_2 + 3x_3$

s.a.  $x_1 + x_2 + x_3 = 1$

$$\boldsymbol{x} \in \{0, 1\}^3$$

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

$$P(\mu) = \min_{\boldsymbol{x}} (1+\mu)x_1 + (2+\mu)x_2 + (3+\mu)x_3 - \mu$$

s.a.  $\boldsymbol{x} \in \{0, 1\}^3$ .

Risolvendo  $P(u)$ :

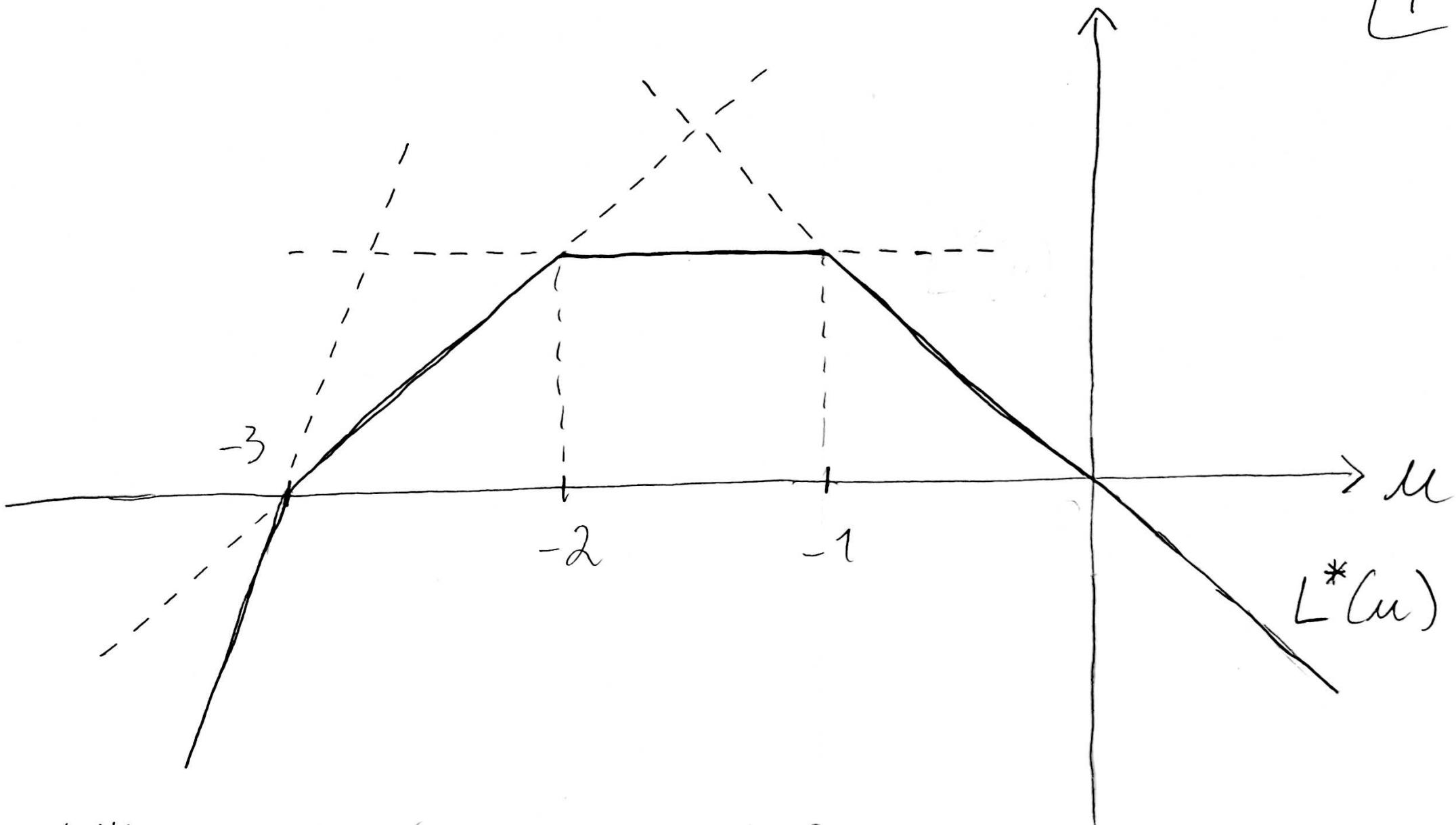
l3

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

Assim,

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

4



$L^*$  não é diferenciável em  $u = -1$ !



Em geral,  $L^*$  não é diferenciável, [5]  
o que impede o uso de métodos que  
dependem de gradientes para resolver  
 $D: \max_u L^*(u)$ .

Porém, nem tudo está perdido:

Teorema:  $L^*$  é côncava; isto é,  
 $-L^*$  é convexa.

Prova: Seja  $X = \{x ; Dx \leq e, x \in \mathbb{Z}_+^m\}$ . (6)

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

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

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

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

$$+ t \min_x \{ c^t x + \tilde{u}^t (Ax - b) ; x \in X \} ] \quad [7]$$

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

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  $-L^*$   $\rightarrow D: \min_u -L^*(u)$  (problema convexo).

• Como no exemplo, em geral  $L^*$  é o L<sup>\*</sup> mínimo entre funções afins.

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

Assim,  $X = \{x^1, x^2, \dots, x^q\}$  e logo

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

Neste caso, o problema D fica

$$D: \max_{\boldsymbol{x}} \min_u \left\{ c^T \boldsymbol{x}^j + u^T (A \boldsymbol{x}^j - b) \right\} \quad |9$$

que equivale a

$$D: \max_z z$$

s.a.  $z \leq c^T \boldsymbol{x}^j + u^T (A \boldsymbol{x}^j - b), \quad j=1, \dots, q$

$$u \in \mathbb{R}^m, \quad z \in \mathbb{R}.$$

Planejamento

Este PL sugere resolver D via geração<sup>10</sup> de linhas ou planos de corte, pois podemos 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. (planos de corte foi visto em "Optimização II").

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

↳ Vamos utilizar o método do gradiente!

Sabemos que para uma função convexa  
diferenciável  $f$ , a condição de optimi-  
dade de 1ª ordem é necessária e suficiente:

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

Porém,  $-L^*$  não é diferenciável ...

Como contornar?

→ subgradiêntes!

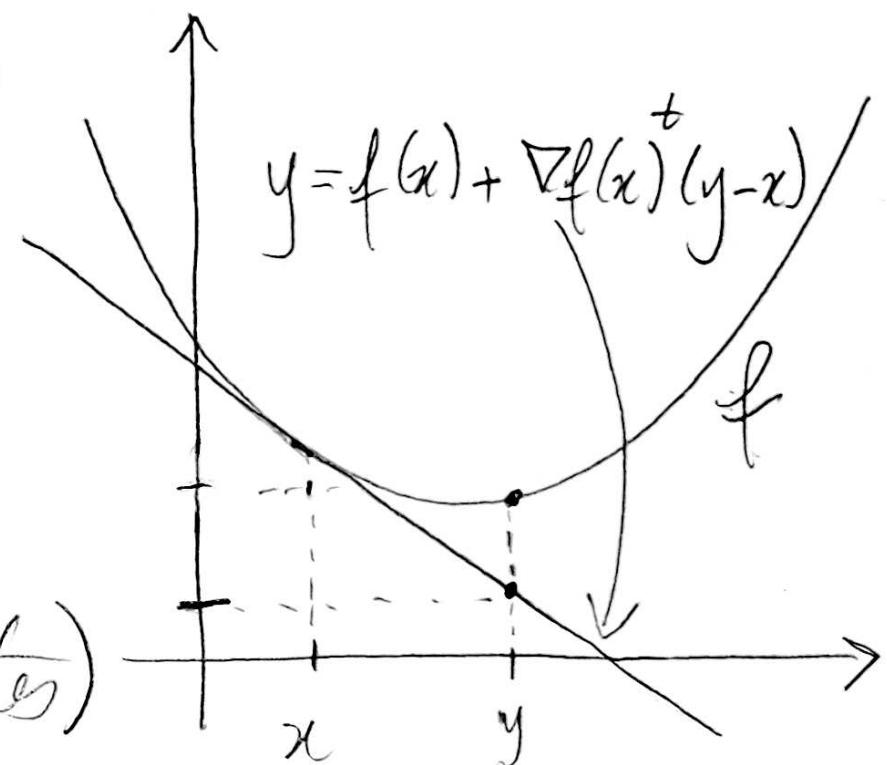
## Subgradiente (de funções convexas)

(12)

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

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

$$f(y) \geq f(x) + \nabla f(x)^T (y - x), \quad \forall x, y \text{ (grafia acima das tangentes)}$$



• O subgradiente inverte o papel de  $\nabla f$  [13] nessa desigualdade:

Definição: o vetor  $g \in \mathbb{R}^n$  é subgradiente de  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  no ponto  $x$  se

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

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

## Interpretacão geométrica

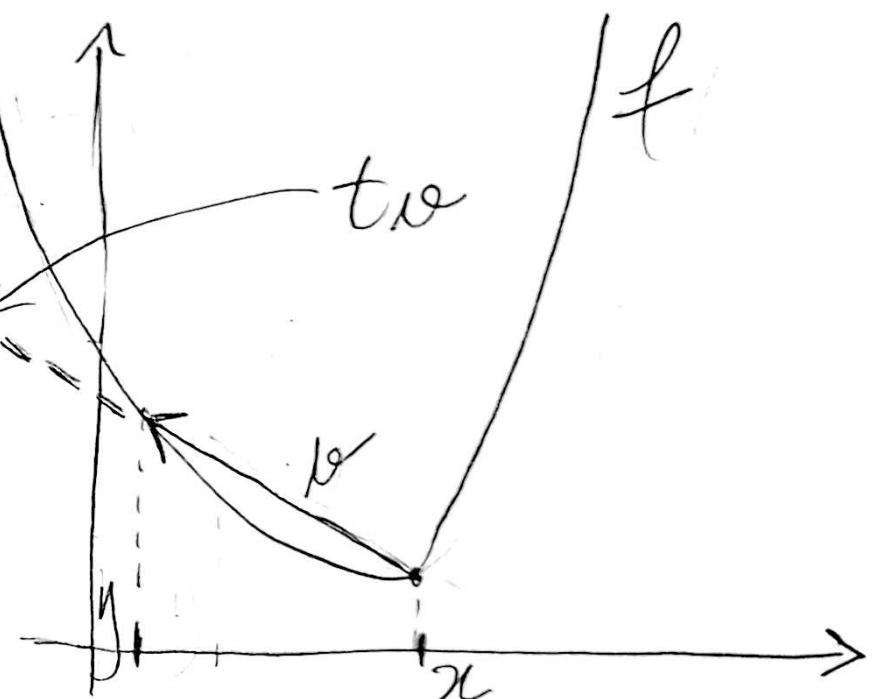
14

Tome  $x \in g \in \partial f(x)$  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  $t v$ .



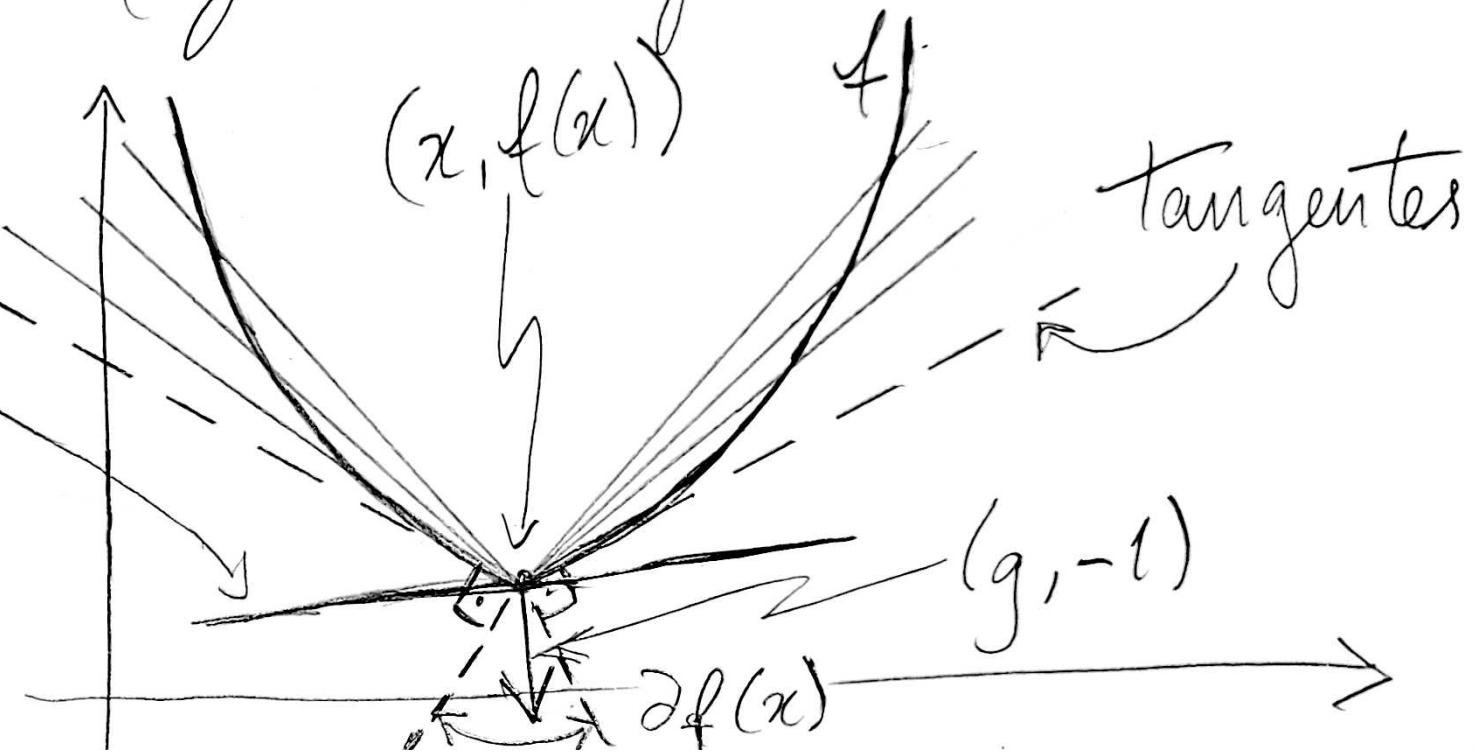
Xemos

[15]

$$\langle (g, -1); tv \rangle = t(g^t(y-x) - f(y) + f(x)) \leq 0$$

Qua seja,  $tv$  faz um ângulo  $\geq 90^\circ$  com  $g$ ,  $\forall t > 0$ .

reta com  
normal  $(g, -1)$



Assim, "g está associado à inclinação das retas que estão entre as tangentes à graf f". (16)

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 maquina" do gráfico de f.

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

# Propriedades dos subgradientes / subdiferenciais.

[17]

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  $\partial f(x) \neq \emptyset$  L8  
 $\forall x \in \mathbb{R}^n$ . (pense na interpretação geométrica  
para se convencer disso).

↳ Assim, um método que só use  
subgradientes estará bem definido para  
 $f$  convexa, mesmo  $f$  não sendo diferen-  
ciável.

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

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

4) Se  $f$  é convexa,  $\partial f(x)$  é conjunto [20  
convexo. De fato, se  $g, h \in \partial f(x)$   
e  $t \in [0,1]$ , então

$$\begin{aligned}
 & f(x) + [(1-t)g + th]^t(y-x) \\
 &= (1-t)[f(x) + g^t(y-x)] + t[f(x) + h^t(y-x)] \\
 &\leq (1-t)f(y) + tf(y) = f(y). \\
 \Rightarrow & (1-t)g + th \in \partial f(x).
 \end{aligned}$$

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  $x$ .

Como  $\lim_{k \rightarrow \infty} g_k = g^*$ . Dado  $y$ , temos  
 $f(y) \geq f(x) + g_k^t(y-x), \forall k$ .

Fazendo  $k \rightarrow \infty$  obtemos

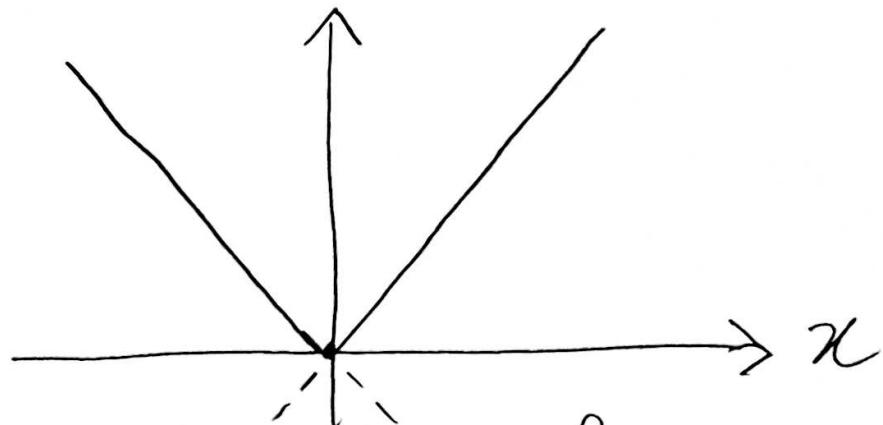
$$f(y) \geq f(x) + g_*^t(y-x), \forall y \Rightarrow g_* \in \partial f(x)$$

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

L22

Exemplos:

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



$f$  é diferenciável em  $x \neq 0$ . Logo

$$\partial f(x) = \begin{cases} f'(x) & , x \neq 0. \end{cases}$$

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

$\Leftrightarrow -1 \leq g \leq 1$ . Assim

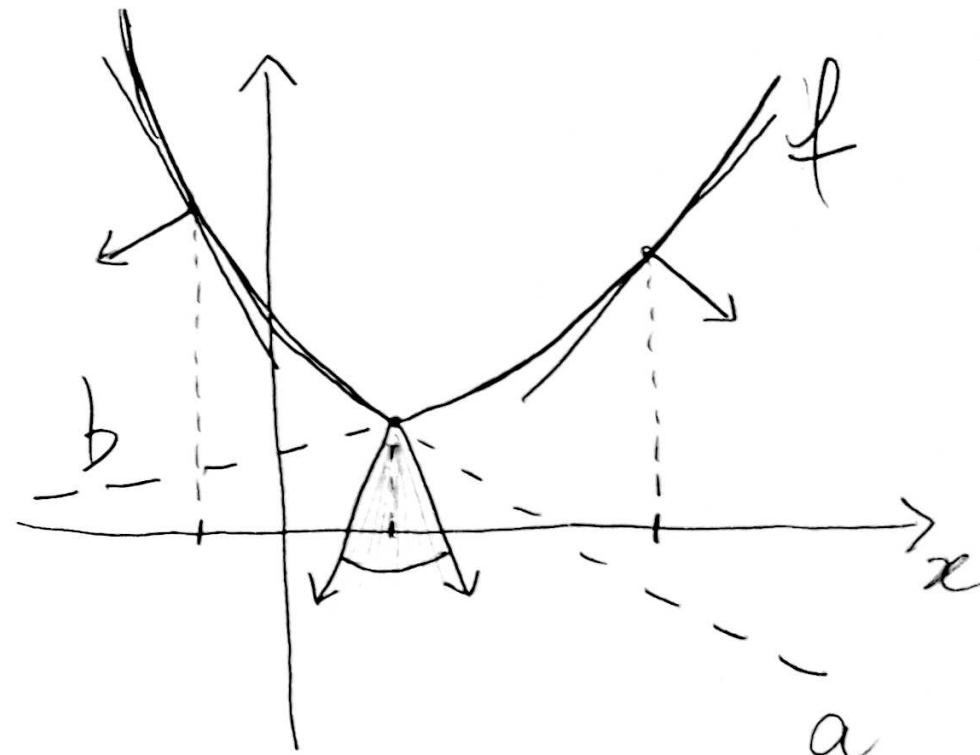
$$\partial|x| = \begin{cases} -1, & x < 0 \\ [-1, 1], & x = 0 \\ 1, & x > 0 // \end{cases}$$

2)  $a, b : \mathbb{R} \rightarrow \mathbb{R}$  convexas e diferenciáveis. (23)

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

Então

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



Note que  $|x| = \max\{|-x|, |x|\}$  (24)  
 (compare os dois exemplos anteriores). //

3) Em geral, se  $f_i : \mathbb{R}^m \rightarrow \mathbb{R}$ ,  $i = 1, \dots, p$ ,  
 são convexas e diferenciáveis, então  
 $f(x) = \max\{f_1(x), \dots, f_p(x)\}$   
 é convexa, e

$\partial f(x) = \text{conv}\{\nabla f_i(x); \text{ i tal que } f_i(x) = f(x)\}$ .  
 (fecho convexo dos gradientes das máximas  $f_i$ 's)

Lembre-se, supondo  $X = \{x; Dx \leq e\}$ , (25)

$x \in \mathbb{Z}_+^m \setminus \{x^1, \dots, x^q\}$  finito, temos

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

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

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