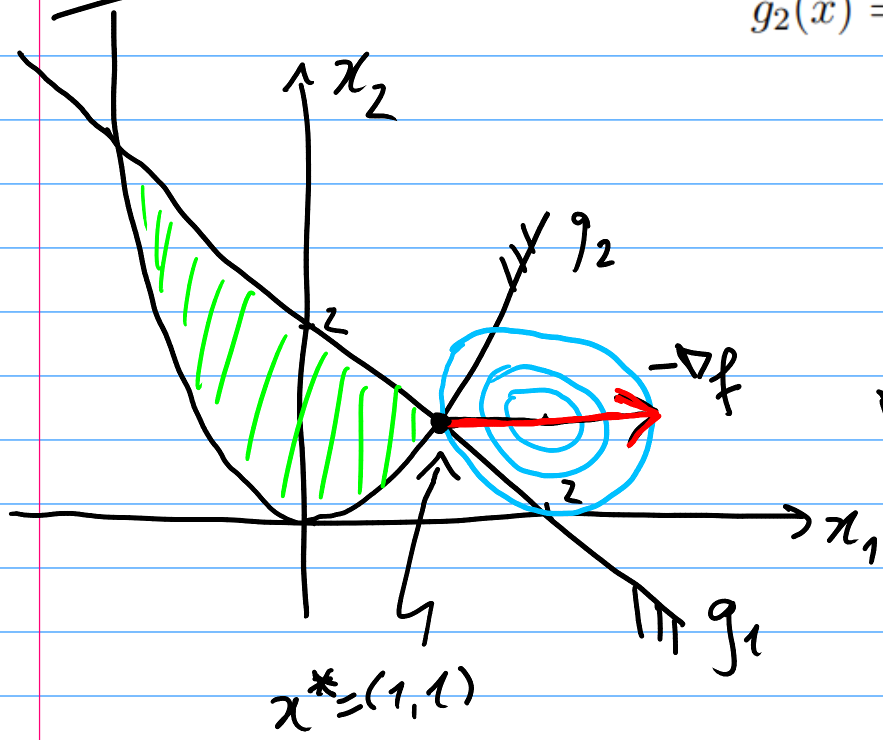
Condições Karush-Kuhn-Tucker (KKT)

Exemplo:

$$\begin{aligned} \text{minimizar} \quad & f(x) = (x_1-2)^2 + (x_2-1)^2 \\ \text{sujeito a} \quad & g_1(x) = x_1 + x_2 - 2 \leq 0 \quad \checkmark \\ & g_2(x) = x_1^2 - x_2 \leq 0. \quad \checkmark \end{aligned}$$



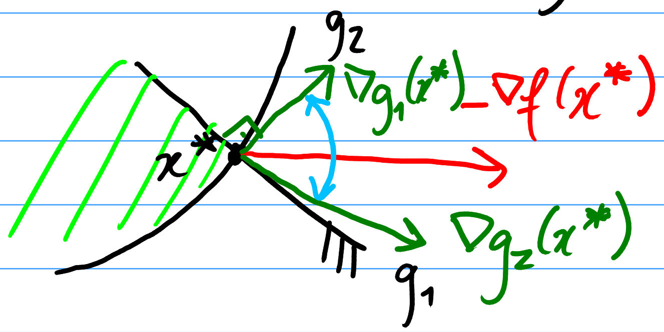
$g_1(0,0) = -2 \leq 0.$

$g_2(0,2) = 0^2 - 2 \leq 0$

$$\nabla f(x_1, x_2) = \begin{bmatrix} 2(x_1-2) \\ 2(x_2-1) \end{bmatrix}$$

$$\nabla f(1,1) = \begin{bmatrix} -2 \\ 0 \end{bmatrix}$$

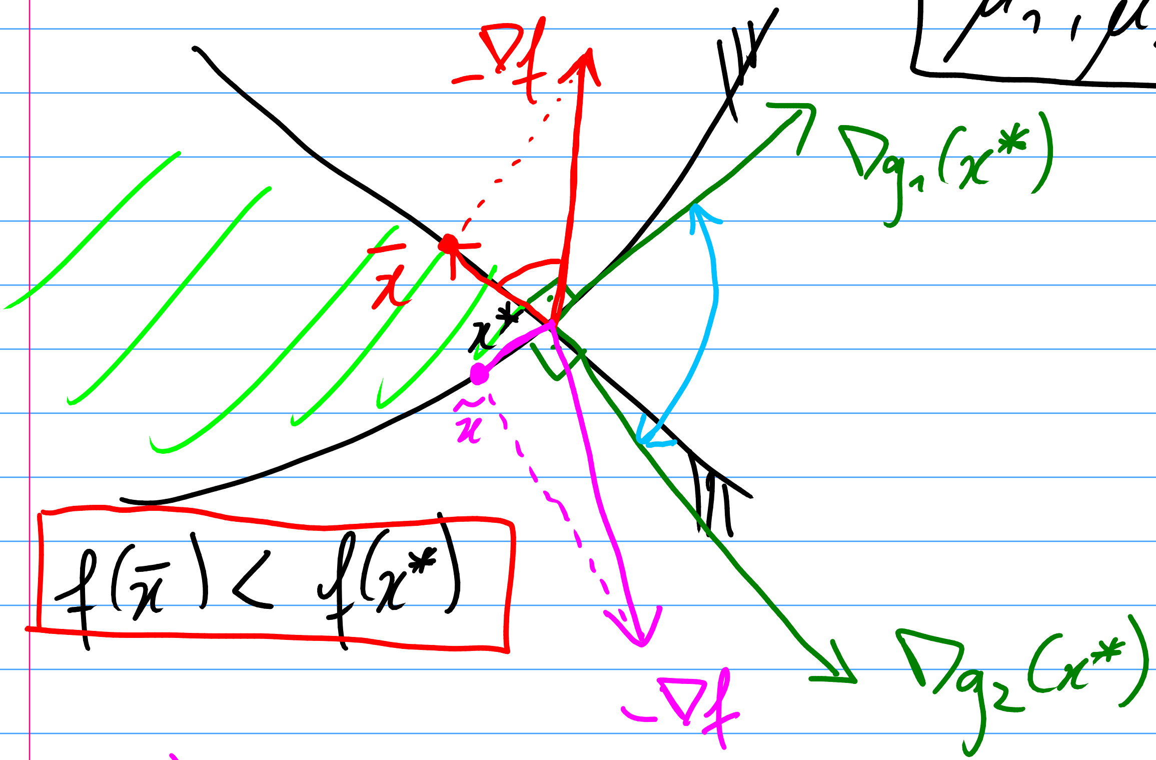
Obs: ao tentar caminhar na direção $-\nabla f(1,1)$ a partir de $(1,1)$, perdemos viabilidade.
($(1,1)$ é minimizador global deste problema)



$$-\nabla f(x^*) = \mu_1 \nabla g_1(x^*) + \mu_2 \nabla g_2(x^*) ,$$

$$\mu_1, \mu_2 \geq 0$$

$$(*) \quad \boxed{\nabla f(x^*) + \mu_1 \nabla g_1(x^*) + \mu_2 \nabla g_2(x^*) = 0, \quad \mu_1, \mu_2 \geq 0}$$



$\boxed{f(\bar{x}) < f(x^*)}$

$f(\tilde{x}) < f(x^*)$

($\bar{x}$ e $\tilde{x}$ são viáveis
$\Downarrow$
x^* não é minimizador local)

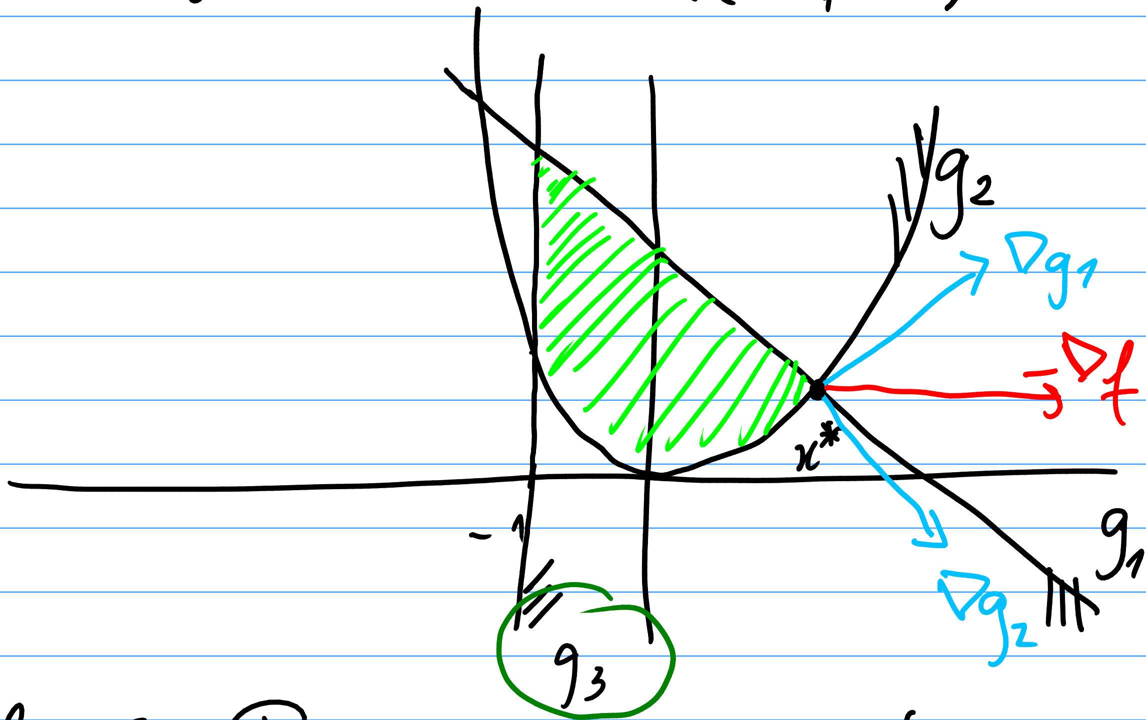
Obs: se não há restrições, (*) resume-se à

$$\nabla f(x^*) = 0$$

(coerente com minimização sem restrições).

$$\begin{aligned} \text{minimizar} \quad & f(x) = (x_1 - 2)^2 + (x_2 - 1)^2 \\ \text{sujeito a} \quad & g_1(x) = x_1 + x_2 - 2 \leq 0 \\ & g_2(x) = x_1^2 - x_2 \leq 0. \end{aligned}$$

$g_3(x) = -x_1 \leq 1 \quad (\Leftrightarrow x_1 \geq -1)$



Definição: Dizemos que uma restrição $g(x) \leq 0$ é ativa no ponto x^* se $g(x^*) = 0$. Caso $g(x^*) < 0$, então g é inativa em x^*.

Como entra g_3 (inativa em x^*) na expressão $(*)$?

$$\nabla f(x^*) + \mu_1 \nabla g_1(x^*) + \mu_2 \nabla g_2(x^*) + \mu_3 \nabla g_3(x^*) = 0$$

$$\mu_1, \mu_2, \mu_3 \geq 0$$

complementaridade $\left(\mu_i \, g_i(x^*) = 0, \quad i = 1, 2, 3 \right.$

$\rightarrow \mu_i = 0$ caso $g_i(x^*) < 0$ (inativa)

Se $g_i(x^*) = 0$ (ativa), μ_i é livre!

Agregando restrições de igualdade

$$\min_x f(x)$$
$$\text{s.a. } h(x) = 0, \quad g(x) \leq 0.$$

Karush-Kuhn-Tucker (KKT)

Def.: x^* é ponto KKT se é viável (i.é., $h(x^*) = 0$ e $g(x^*) \leq 0$) e existem $\lambda_1, \dots, \lambda_m$ e $\mu_1, \dots, \mu_p$ tais que

$$\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i \nabla h_i(x^*) + \sum_{j=1}^{p} \mu_j \nabla g_j(x^*) = 0,$$

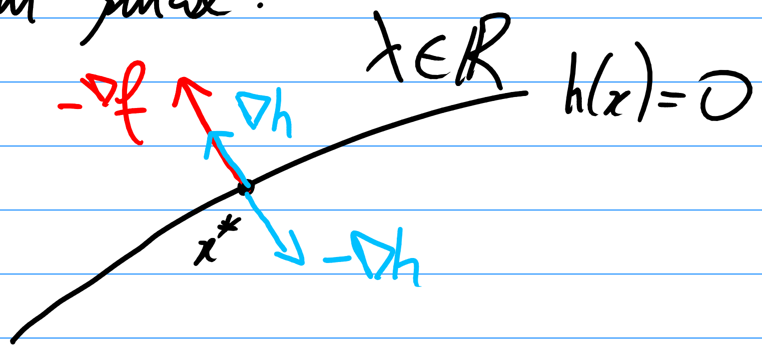
$$\mu_1, \dots, \mu_p \geq 0$$

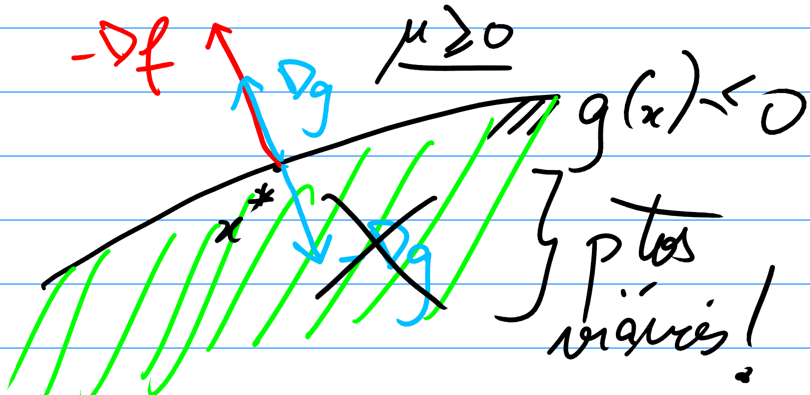
$$\mu_j\, g_j(x^*) = 0, \quad \forall j.$$

"Todo ponto KKT é candidato à minimizador".

Obs: • se $h \equiv 0$ e $g \equiv 0$ então KKT resume-se à $\nabla f(x^*) = 0$. ✓

• λ_i's não tem sinal.





• Nem todo minimizador é ponto KKT!

$$\min_x \; x \qquad \text{Claro que } x^* = 0.$$
$$\text{s.a. } x^2 \leq 0$$

KKT (para x^*): $\boxed{f'(x^*) + \mu\, g'(x^*) = 0}$,

$\mu \geq 0$

$\mu\, g(x^*) = 0$

$1 + \mu\, \underline{(2.0)} = 0$ ✗

falha $\forall \mu \geq 0$.

<u>Problema</u>: $g'(x^*) = 0$!!!

(ou $\exists\, \nabla g_j(x^*)$; g_j ativa } é L.D.).

• <u>Def.</u>: x^* é <u>regular</u> se

$\nabla h_i(x^*)$, $i = 1, \dots, m$,

$\nabla g_j(x^*)$, j tq $g_j(x^*) = 0$ (ativas)

formam um conjunto linearmente independente.

• <u>Teorema</u>: se x^* é minimizador local e regular então x^* é KKT.

(condição necessária de otimalidade - CNO - de 1ª ordem)

<u>Prova</u>: depois, usando método de penalidade externa.

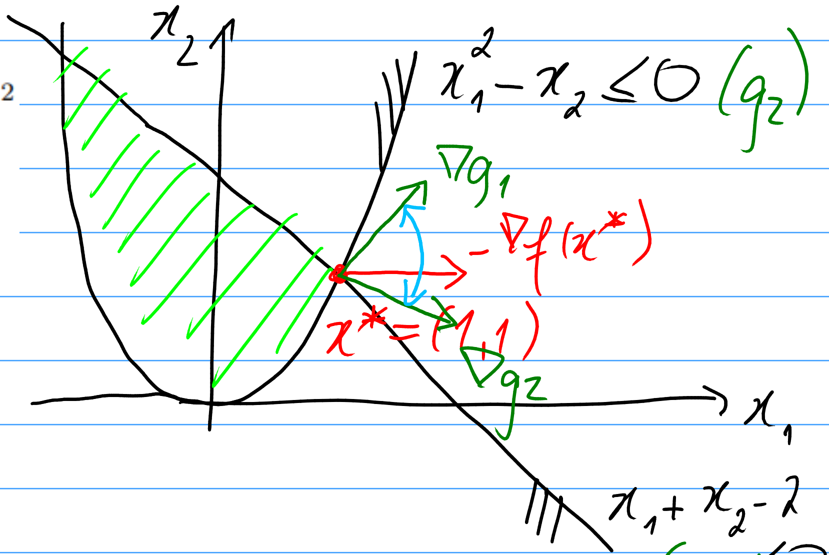
- λ e μ são chamados de <u>multiplicadores de Lagrange</u>

- Em <u>problemas convexos</u> (a definir), vale que: se x^* é KKT então x^* é minimizador global.

Em problemas não convexos, não vale em geral!

↳ KKT apenas fornece candidatos a minimizador (tal como $\nabla f = 0$ na otimização irrestrita)

$$\begin{array}{ll} \text{minimizar} & f(x) = (x_1 - 2)^2 + (x_2 - 1)^2 \\ \text{sujeito a} & g_1(x) = x_1 + x_2 - 2 \leq 0 \\ & g_2(x) = x_1^2 - x_2 \leq 0. \end{array}$$



<u>KKT:</u>

$$\nabla f(x) + \mu_1 \nabla g_1(x) + \mu_2 \nabla g_2(x) = 0$$

x viável!

$$\mu_1, \mu_2 \geq 0$$

$$\mu_1 g_1(x) = 0, \quad \mu_2 g_2(x) = 0.$$

$$(*)\quad \begin{bmatrix} 2(x_1-2) \\ 2(x_2-1) \end{bmatrix} + \mu_1 \begin{bmatrix} 1 \\ 1 \end{bmatrix} + \mu_2 \begin{bmatrix} 2x_1 \\ -1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad x \text{ viável.}$$

$$\mu_1, \mu_2 \geq 0$$

$$\mu_1(x_1+x_2-2)=0$$

$$\mu_2(x_1^2-x_2)=0.$$

Caso 1: $\mu_1=\mu_2=0$.

Neste caso, $2(x_1-2)=2(x_2-1)=0 \Rightarrow (x_1,x_2)=(2,1)$. Veja que $g_1(2,1)=1 \not\leq 0$, ou seja, $(2,1)$ não é viável. Logo não pode ser $\mu_1=\mu_2=0$.

Caso 2: $\mu_1>0$ e $\mu_2=0$

Não é possível (exercício)

Caso 3: $\mu_1=0$ e $\mu_2>0$

Não é possível (exercício)

Caso 4: $\mu_1>0$ e $\mu_2>0$

Neste caso, a condição de complementaridade dizem que ambas restrições $g_1(x)\leq 0$ e $g_2(x)\leq 0$ são ativas, isto é,

$$g_1(x)=x_1+x_2-2=0 \quad \text{e} \quad g_2(x)=x_1^2-x_2=0$$

Assim, resolvemos o sistema

$$\begin{cases} x_1+x_2-2=0 & (1) \\ x_1^2-x_2=0 & (2) \end{cases}$$

Substituindo (2) em (1), obtemos $x_1^2+x_1-2=0$. Daí $x_1=1$ ou $x_1=-2$. As possibilidades de pontos KKT são portanto $(1,1)$ e $(-2,4)$.

Agora, verificamos se cada ponto é de fato KKT. Eles são viáveis. Falta verificar (*):

$$\begin{bmatrix} 2(x_1-2) \\ 2(x_2-1) \end{bmatrix} + \mu_1 \begin{bmatrix} 1 \\ 1 \end{bmatrix} + \mu_2 \begin{bmatrix} 2x_1 \\ -1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$\checkmark$ $\underline{(1,1):}$ $\begin{bmatrix} 2(1-2) \\ 2(1-1) \end{bmatrix} + \mu_1 \begin{bmatrix} 1 \\ 1 \end{bmatrix} + \mu_2 \begin{bmatrix} 2 \\ -1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

$$\Rightarrow \begin{cases} \mu_1+2\mu_2=2 \\ \mu_1-\mu_2=0 \end{cases} \Rightarrow \mu_1=\mu_2=\tfrac{2}{3}>0.$$

$\underline{(-2,4):}$ $\begin{bmatrix} 2(-2-2) \\ 2(4-1) \end{bmatrix} + \mu_1 \begin{bmatrix} 1 \\ 1 \end{bmatrix} + \mu_2 \begin{bmatrix} -4 \\ -1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

$\Rightarrow \begin{cases} \mu_1-4\mu_2=8 \\ \mu_1-\mu_2=-6 \end{cases}$. Daí $-3\mu_2=14 \Rightarrow \mu_2=-\frac{14}{3}$,

o que é proibido visto que $\mu_2 \geq 0$.
Logo $(-2,4)$ não é KKT.
O único ponto KKT, e minimizador, é $(1,1)$.

Relembrando...

- KKT fornece candidatos à minimizador
- Nem todo minimizador é KKT. Por exemplo,

$$\min x \text{ s.a. } x^2 \leq 0.$$

não é regular

$x^* = 0$ é o (único) minimizador, porém KKT não vale: $1 + \mu(\underbrace{2x^*}_{0}) = 0$ é impossível para qualquer $\mu \geq 0$!

Veja que o gradiente de g se anula em x^*. $\longrightarrow$ os gradientes das restrições ativas são L.D.

- x^* ponto viável é regular se

$\nabla h_i(x^*)$, $\forall i$, $\nabla g_j(x^*)$, $j: g_j(x^*) = 0$

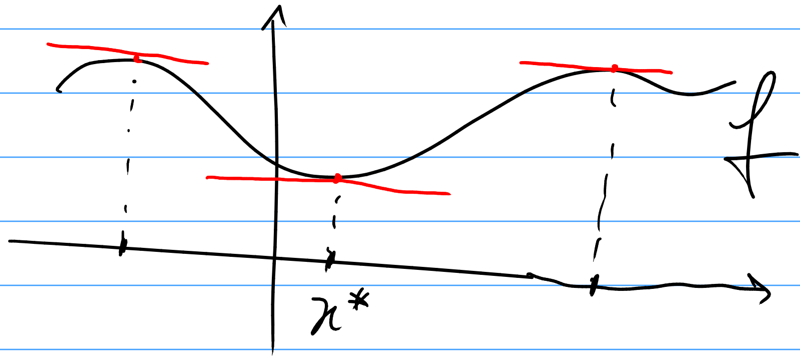
formam um conjunto L.I.

hip. sobre as restrições

- **TEO:** Se x^* é minimizador regular então x^* é KKT.

Otimização irrestrita: x^* é min. $\Rightarrow \nabla f(x^*) = 0$

→ prova através do método de penalidade externa.

- 

Algoritmo busca anular $\nabla f(x)$. (otim. irrestrita)

- Otimização restrita:

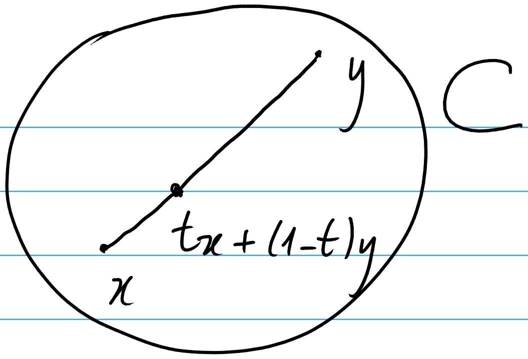
Algoritmos buscam satisfazer KKT.

Problemas de Otimização Convexos

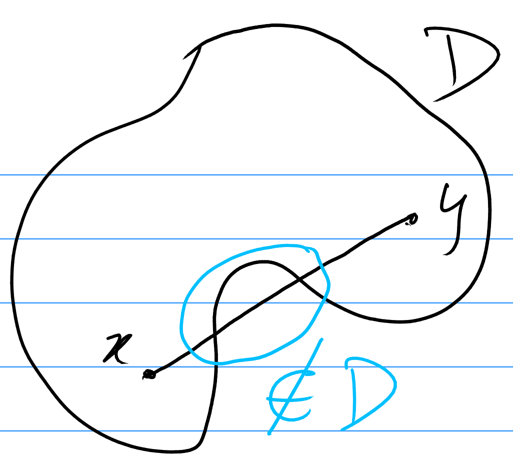
Def: Um conjunto $C \subset \mathbb{R}^n$ é convexo se dados $x, y \in C$ e $t \in [0,1]$, temos

$$tx + (1-t)y \in C.$$

"o segmento de x à y está contido em C"



Convexo



<u>não</u> convexo.

<u>Exemplos:</u>

1) $\mathbb{R}^n$ é convexo.

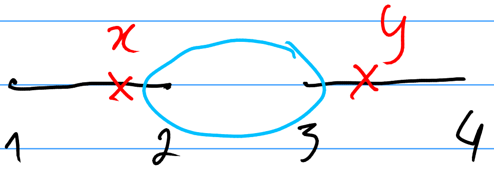
2) $[a,b]$, $[a,b]\times[c,d]$, $\bigtimes_{i=1}^{m}[a_i,b_i]$

são convexos.

3) $\{x\in\mathbb{R}^n ; Ax=b,\ Cx\leq d\}$ é convexo (conjunto viável de PL's)

4) $C_1,\dots,C_p$ convexos $\Longrightarrow \bigcap_{i=1}^{p} C_i$ é convexo. (prove!)

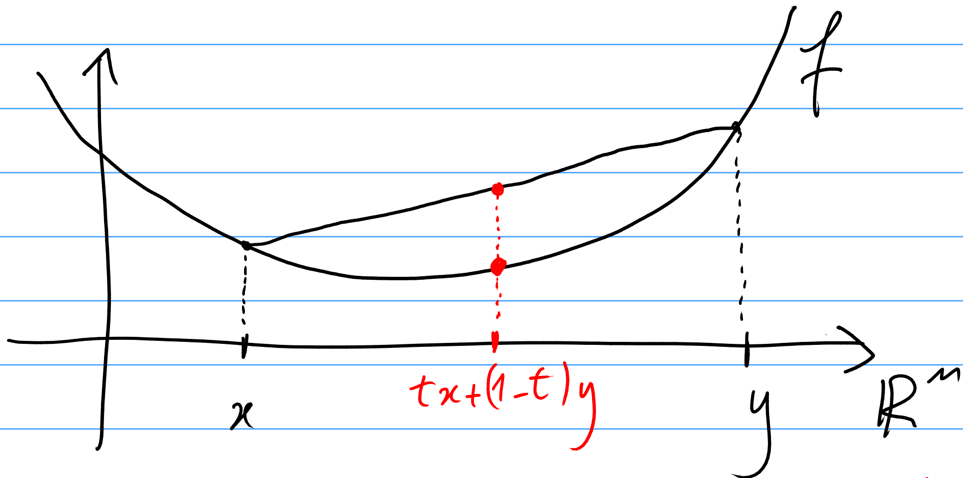
5) $[1,2]$ e $[3,4]$. Então $[1,2]\cup[3,4]$ <u>não</u> é convexo



Def.: Seja $C \subset \mathbb{R}^n$ conjunto convexo. A função $f: C \to \mathbb{R}$ é <u>convexa</u> se

$$f(tx+(1-t)y) \leq t f(x) + (1-t) f(y),$$

$\forall x, y \in C$ e $\forall t \in [0,1]$



"O gráfico de f está abaixo das cordas que ligam pontos de seu gráfico".

Exemplos:

1) $f(x) = c$, c constante é convexa.

2) $f(x) = c^t x$ é convexa (F.O. dos PL's)

3) $f(x) = \frac{1}{2} x^t A x + b^t x + c$, onde A é simétrica de definida positiva (função quadrática).

$$f: \mathbb{R} \to \mathbb{R} \quad , \quad f(x) = \frac{1}{2}x^2 + x \,.$$

$$f: \mathbb{R}^2 \to \mathbb{R} \quad , \quad f(x) = \frac{1}{2}[x_1 \; x_2] \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$\to f(x) = \frac{1}{2}[2x_1 \;\; 3x_2] \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$f(x) = \frac{1}{2}(2x_1^2 + 3x_2^2)$$

Problemas convexos

$$\min f(x) \quad s.a. \quad h(x) = 0 \,,\; g(x) \leq 0$$

é um problema convexo se:

(i) f é convexa

(ii) g_j são convexas, $\forall j$

(iii) h_i são afins, $\forall i$

($\Leftrightarrow$ h_i e $-h_i$ são convexas)

Interesse em convexidade:

Em problemas convexos,

(1) x^* minimizador local $\Rightarrow$ x^* é min global

$\checkmark$(2) x^* é KKT $\Rightarrow$ x^* é minimizador global.

Condições de Karush-Kuhn-Tucker (KKT)

(problemas convexos)

P: $\min f(x)$ s.a. $h(x)=0$, $g(x)\leq 0$

$f: \mathbb{R}^n \to \mathbb{R}$ convexa

$g_j: \mathbb{R}^n \to \mathbb{R}$ convexas , $\forall j = 1, \dots, p$

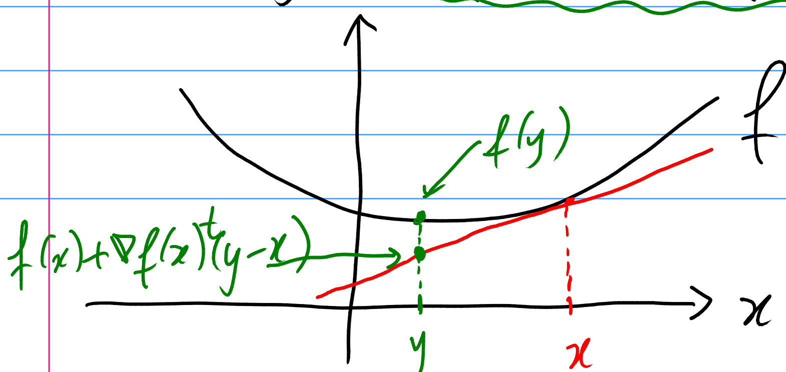
$h_i: \mathbb{R}^n \to \mathbb{R}$ afim , $\forall i = 1, \dots, m$

$\hookrightarrow h(x) = Ax - b$

→ KKT é suficiente para otimalidade em problemas convexos.

Teorema: Seja $C \subset \mathbb{R}^n$ um conjunto convexo e $f: C \to \mathbb{R}$ uma função diferenciável. Então f é convexa se, e somente se,

$$f(y) \geq f(x) + \nabla f(x)^t (y-x) \ , \forall x, y \in C.$$



Teorema: Seja x^* um ponto KKT do problema convexo P. Então x^* é minimizador global de P.

Prova: Como x^* é KKT, existem $\lambda \in \mathbb{R}^m$ e $\mu \in \mathbb{R}^p$ tais que

- x^* é viável $(h(x^*)=0, g(x^*)\leq 0)$
- $\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i \nabla h_i(x^*) + \sum_{j=1}^{p} \mu_j \nabla g_j(x^*) = 0$
- $\mu_j \geq 0, \forall j$
- $\mu_j g_j(x^*) = 0, \forall j$.

Dado qualquer y viável, multiplicamos o item por $(y-x^*)$, obtendo

$$(1)\ \nabla f(x^*)^t(y-x^*) + \sum_{i=1}^{m} \lambda_i \nabla h_i(x^*)^t(y-x^*) + \sum_{j=1}^{p} \mu_j \nabla g_j(x^*)^t(y-x^*) = 0$$

Usando o teorema anterior com $x=x^*$, temos

$$(1) \leq f(y) - f(x^*) + \sum_{i=1}^{m} \lambda_i (h_i(y) - h_i(x^*)) + \sum_{j=1}^{p} \mu_j (g_j(y) - g_j(x^*))$$

Como $h_i(x) = a_i^t x - b_i$, note que

$$\underbrace{Dh_i(x^*)^t(y - x^*)}\ = h_i(y) - h_i(x^*).$$

De fato, $a_i^t(y - x^*) = a_i^t y - b_i - (a_i^t x^* - b_i)$

$$= h_i(y) - h_i(x^*).$$

Como x^* e y são viáveis, temos

$$(1) \leq f(y) - f(x^*) + \sum_{j=1}^{p} \underbrace{\mu_j g_j(y)}_{\leq 0} \qquad (\mu_j g_j(x^*) = 0 \ \forall j)$$

$$\Rightarrow 0 \leq f(y) - f(x^*) \Rightarrow f(y) \geq f(x^*).$$

Como y é arbitrário ($f(y) \geq f(x^*)$ $\forall y$ viável), concluímos que x^* é minimizador global de f. ∎