Gradiente Projetado

$$\min f(x) \quad \text{s.a.} \quad x \in C,$$

onde $C \subset \mathbb{R}^n$ é convexo e fechado.

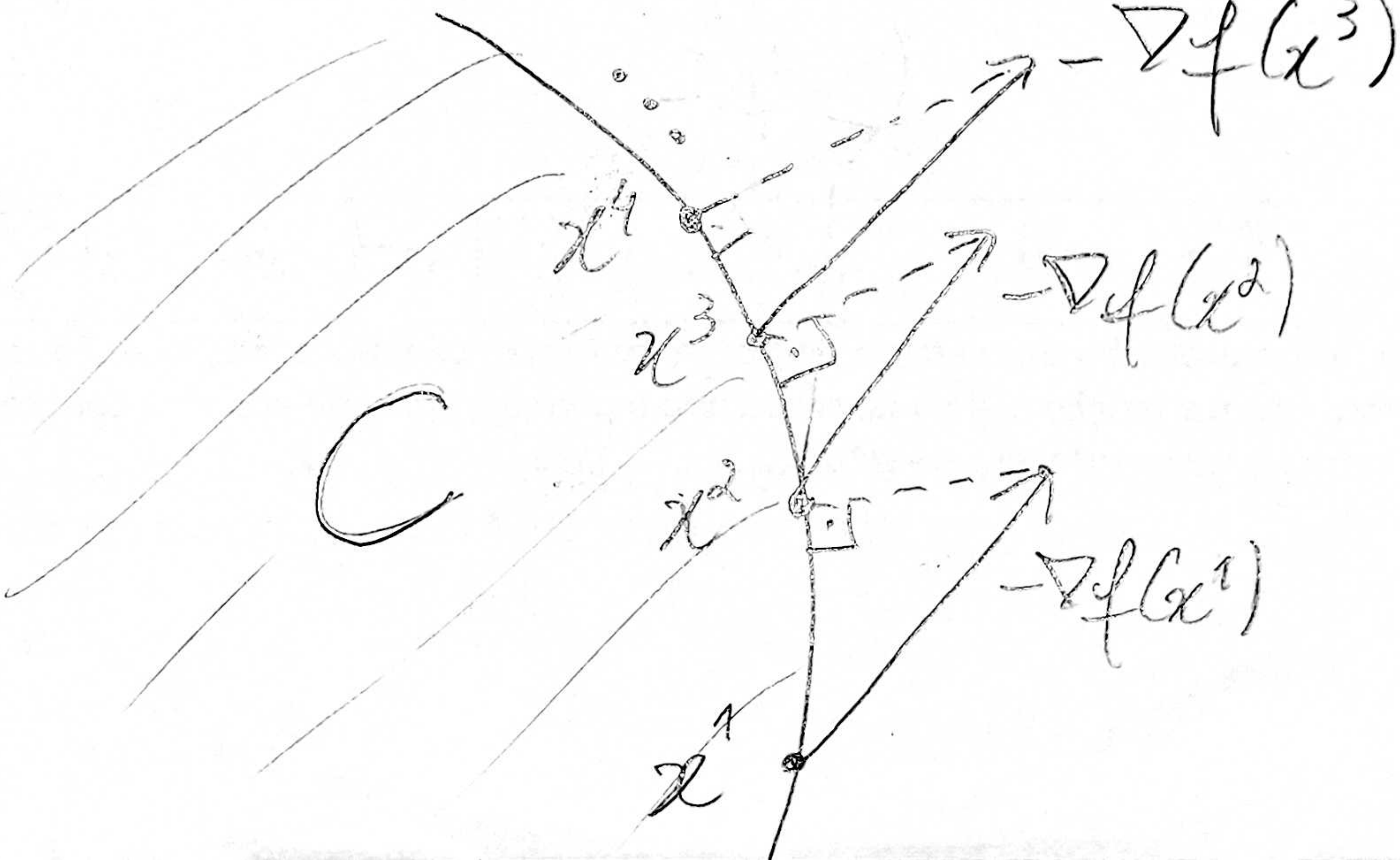
Ideia: Adaptar o método do gradiente

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

para garantir viabilidade:

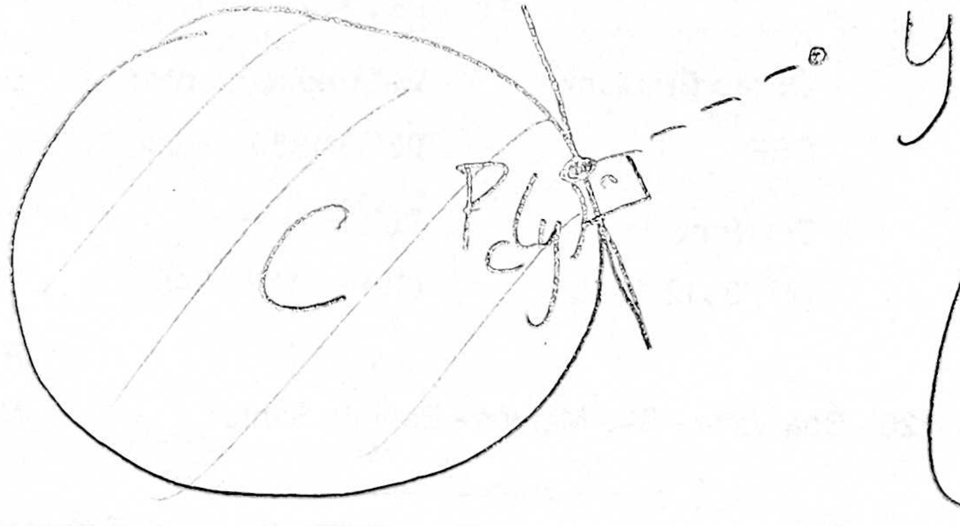
$$x^{k+1} \in C.$$

<u>Como</u>? Calcule x^{k+1} e "conserte" a viabilidade projetando no conjunto C.



O que é projetar sobre C?

- A projeção de um ponto $y \in \mathbb{R}^n$ em C é o ponto $P_C(y) \in C$ mais próximo à y.



$P_C(y)$ é a projeção de y sobre C.

Ou seja, $x^* = P_C(y)$ é a solução do problema

$$\min_x \frac{1}{2}\|x-y\|^2 \quad \text{s.a.} \quad x \in C.$$

<u>Teorema</u>: Se $C \neq \emptyset$ é convexo e fechado então $P_C(y)$ está bem definida (a projeção é <u>única</u> para cada $y \in \mathbb{R}^n$).

De fato, o problema de projeção

$$\min_{x} p(x) = \frac{1}{2}\|x - y\|^2 \quad s.a. \quad x \in C$$

é viável, dado que $C \neq \emptyset$. Agora,

$$p(x) = \frac{1}{2} x^t I x - y^t x + cte(y)$$

é uma quadrática estritamente convexa (sua Hessiana é $I \succ 0$). Logo este problema admite minimizador global.

Mas como este problema é convexo
(pois C é conjunto convexo), então
admite único minimizador global
x^*: se $\bar{x}$ e $\tilde{x}$ fossem minimizadores
distintos então $\frac{1}{2}\bar{x} + \frac{1}{2}\tilde{x} \in C$ e

$$P\left(\frac{1}{2}\bar{x} + \frac{1}{2}\tilde{x}\right) < \frac{1}{2}P(\bar{x}) + \frac{1}{2}P(\tilde{x}) = P^*,$$

uma contradição. ∎

$$P_C(y) = \underset{x \in C}{\arg\min} \ \tfrac{1}{2}\|x - y\|^2, \quad (*)$$ 7

$C \neq \emptyset$ convexo fechado.

Como calcular $P_C(y)$?

- Resolvendo $(*)$ via KKT, para conjuntos C específicos.

(lembre-se que KKT é suficiente para otimalidade em problemas convexos).

8

1) $C = \{ x \in \mathbb{R}^n ; Ax = b \}$,

A $m \times n$, posto $A = m$.

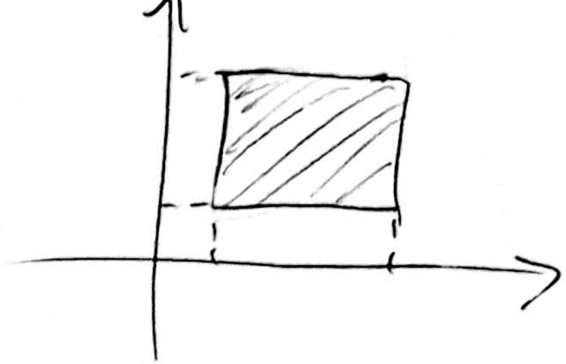
EXERCÍCIO: Use as condições KKT para mostrar que

$$P_C(y) = (I - A^t (AA^t)^{-1} A) y.$$

(Compare com a expressão (8.2) do livro de Cuna, página 57)

2) $C = \{x \in \mathbb{R}^n ; l_i \le x_i \le u_i , \forall i\}$

(caixa, $\underline{l_i < u_i}, \forall i$)



Problema projeção:

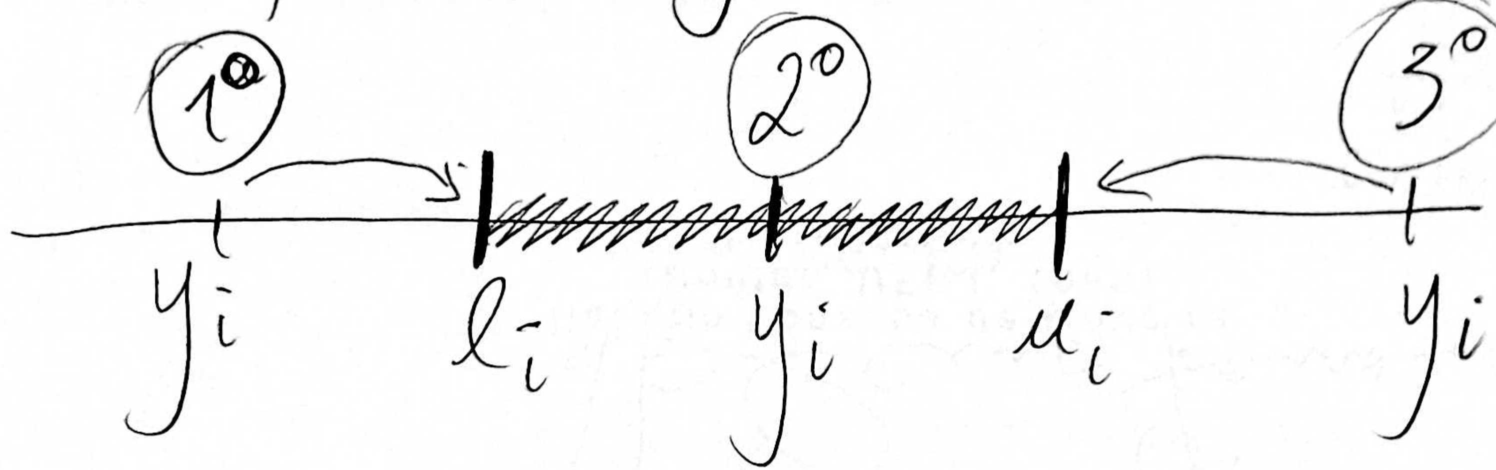
$$\min \frac{1}{2}\|x-y\|^2 \quad \text{s.a.} \quad -x+l \le 0, \; x-u \le 0$$

<u>KKT:</u> $(x^*-y) - \mu^l + \mu^u = 0 \qquad (1)$

$$\mu^l \ge 0, \; \mu^u \ge 0, \; l \le x^* \le u \qquad (2)$$

$$\mu_i^l(-x_i^* + l_i) = \mu_i^u(x_i^* - u_i) = 0, \; \forall i \qquad (3)$$

3 casos para y_i:



• $\underline{l_i < y_i < u_i}$. Com fim de satisfazer KKT, basta tomar $x_i^* = y_i$ e $\mu_i^l = \mu_i^u = 0$.

- $\underline{y_i \leq l_i}$: tomar $x_i^* = l_i$, $\mu_i^l = x_i^* - y_i \geq 0$ e $\mu_i^u = 0$.

- $\underline{y_i \geq u_i}$: tomar $x_i^* = u_i$, $\mu_i^l = 0$ e $\mu_i^u = y_i - x_i^* \geq 0$.

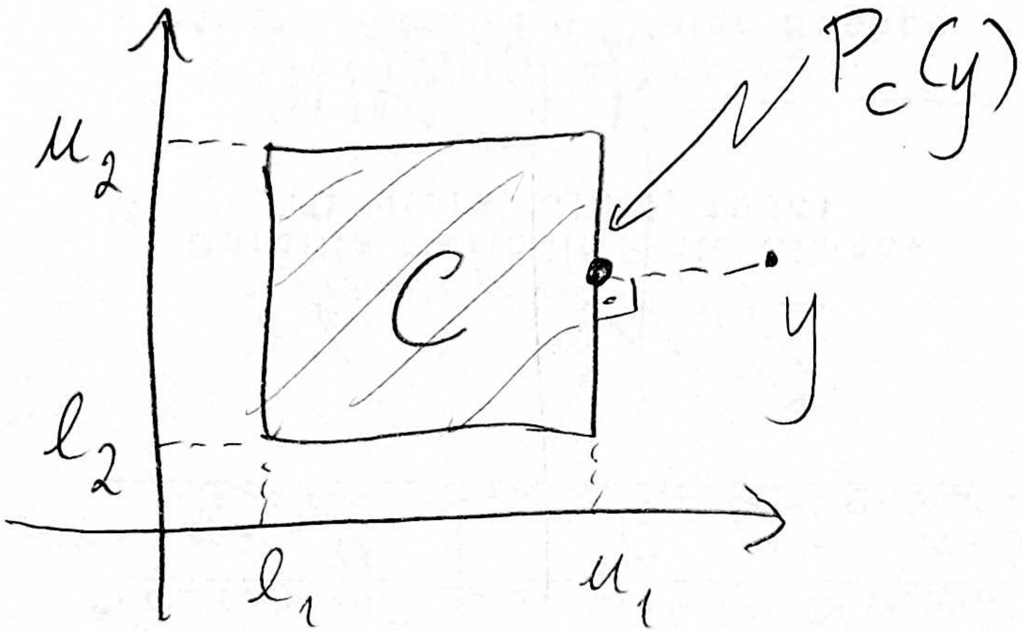
Em resumo:

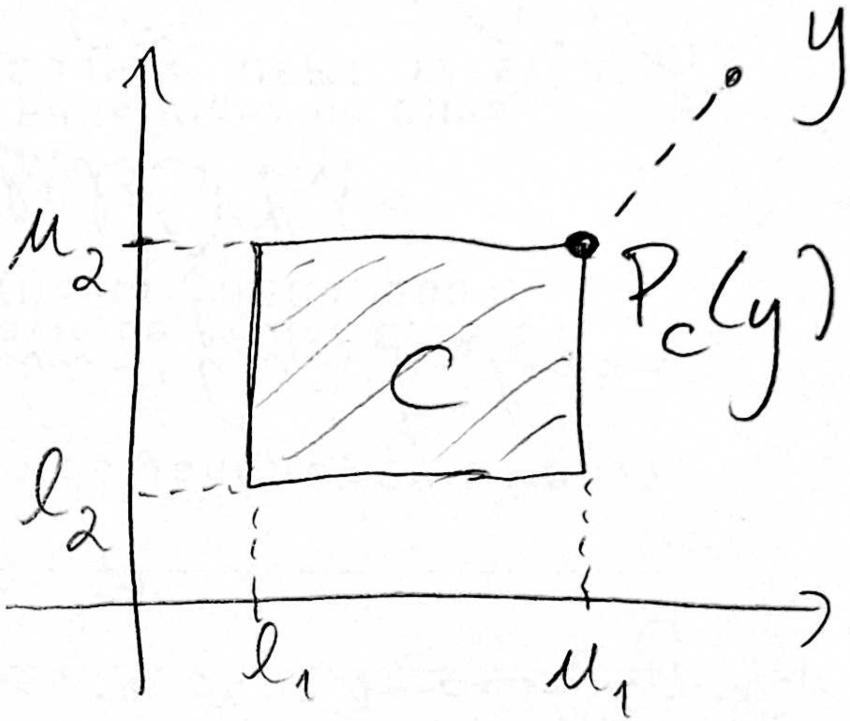
$$[P_C(y)]_i = \begin{cases} l_i & \text{se} \quad y_i \leq l_i \\ y_i & \text{se} \quad l_i < y_i < u_i \\ u_i & \text{se} \quad y_i \geq u_i \end{cases}$$

Em implementações, fazemos, $\forall i$,

$$[P_C(y)]_i = \min\{u_i, \max\{l_i, y_i\}\}$$

(verifique que essa é a projeção)





Como parar? Ou, o que é KKT para o problema original

$$\min f(x)$$

$$\text{s.a. } l_i \leq x_i \leq u_i, \forall i \text{ ?}$$

$\boxed{\text{Resposta}}$: x^* é KKT se, e somente se, $P_C(x^* - \nabla f(x^*)) - x^* = 0$.

(EXERCÍCIO)

Quais direções tomar?

- $d^k = -\nabla f(x^k) \longrightarrow$ min. sem restrições.

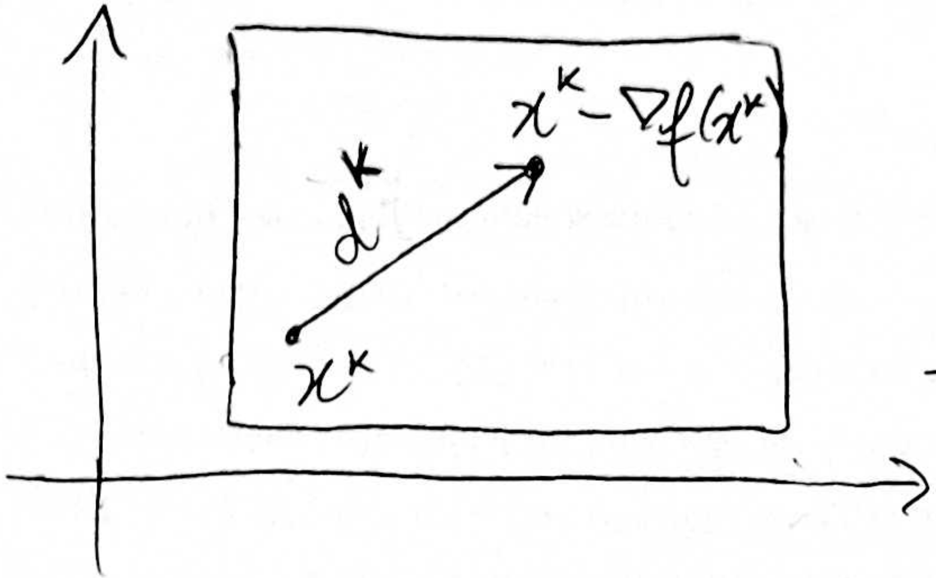
  $\hookrightarrow x^{k+1} = x^k - t_k \nabla f(x^k)$, $t_k > 0$.

- Restrições de caixa:

  $$d^k = P_C(x^k - \nabla f(x^k)) - x^k.$$

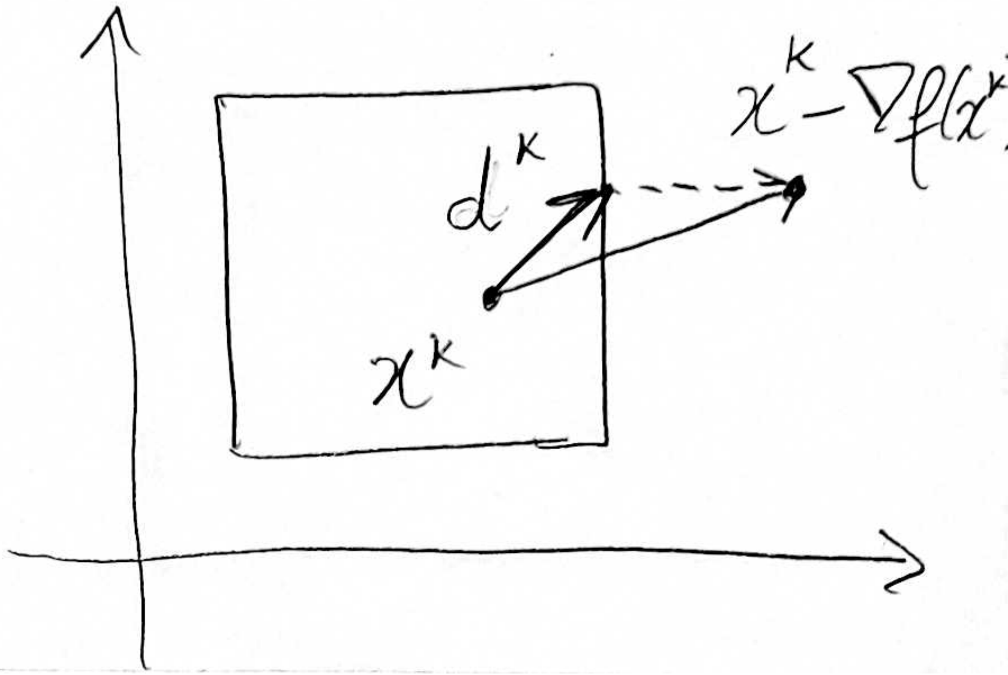
Observe que se $x^k - \nabla f(x^k) \in C$ então

$$d^k = (x^k - \nabla f(x^k)) - x^k = -\nabla f(x^k)$$



(passo dentro da caixa = passo min. sem restrições).

Agora, se $x^k - \nabla f(x^k) \notin C$, então $x^k + d^k \in C$.

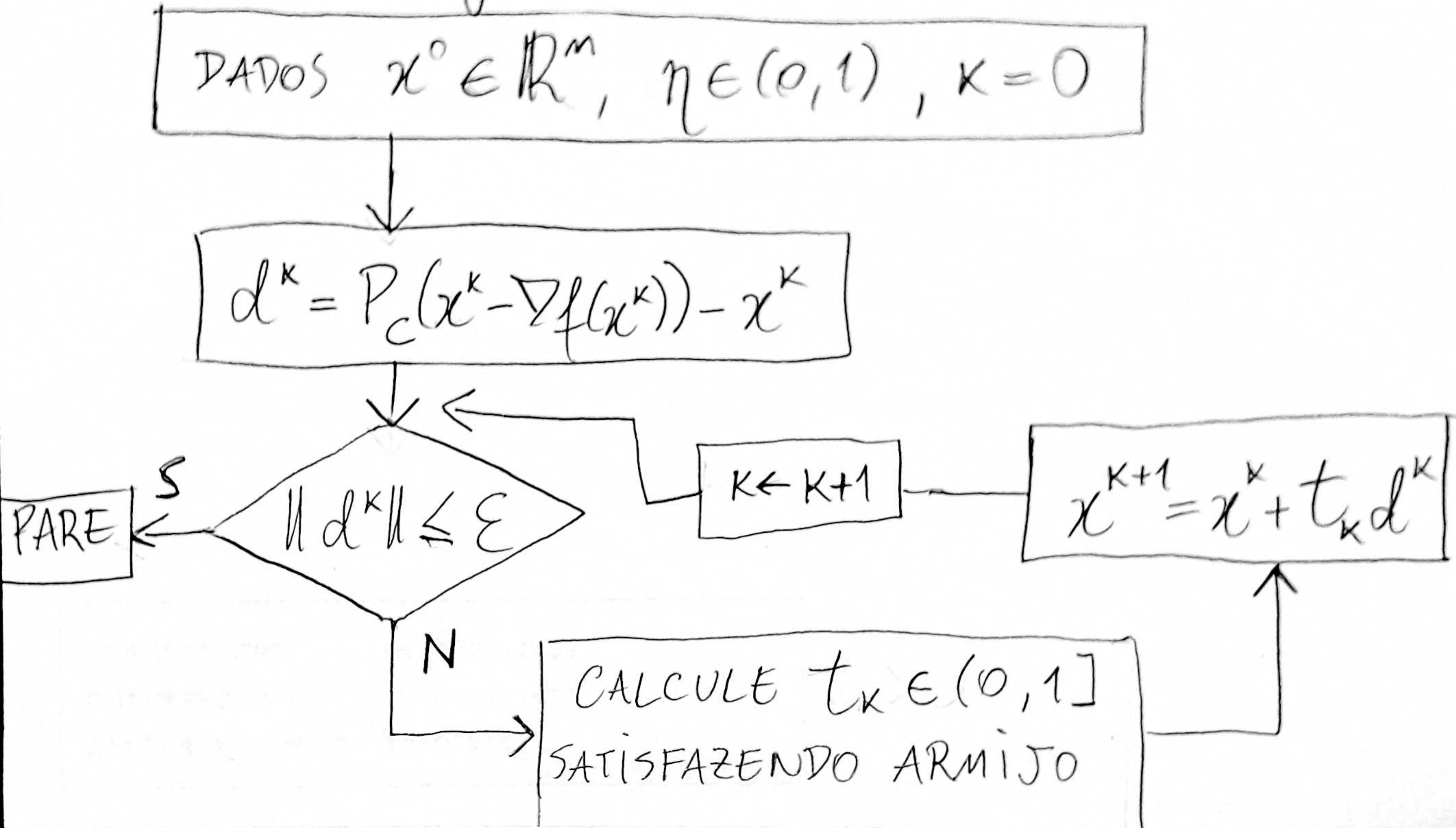


EXERCÍCIO: mostre que $x^k + d^k \in C$.

Conclusão: pela convexidade de C, temos $x^k + t_k d^k \in C$, $\forall t \in [0,1]$. Então podemos realizar uma busca linear em $t_k \in [0,1]$ (Armijo), e definir

$$x^{k+1} = x^k + t_k d^k \quad (\in C) \text{ !!!}$$

Método do gradiente projetado



Observação:

Calcular $t_k \in (0, 1]$ satisfazendo Armijo se faz da mesma maneira que no esquema geral de descida.

Aliás, $d^k = P_C(x^k - \nabla f(x^k)) - x^k$ é uma direção de descida para f a partir de x^k !!! (não vou provar :()