Gradiente Projetado

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

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

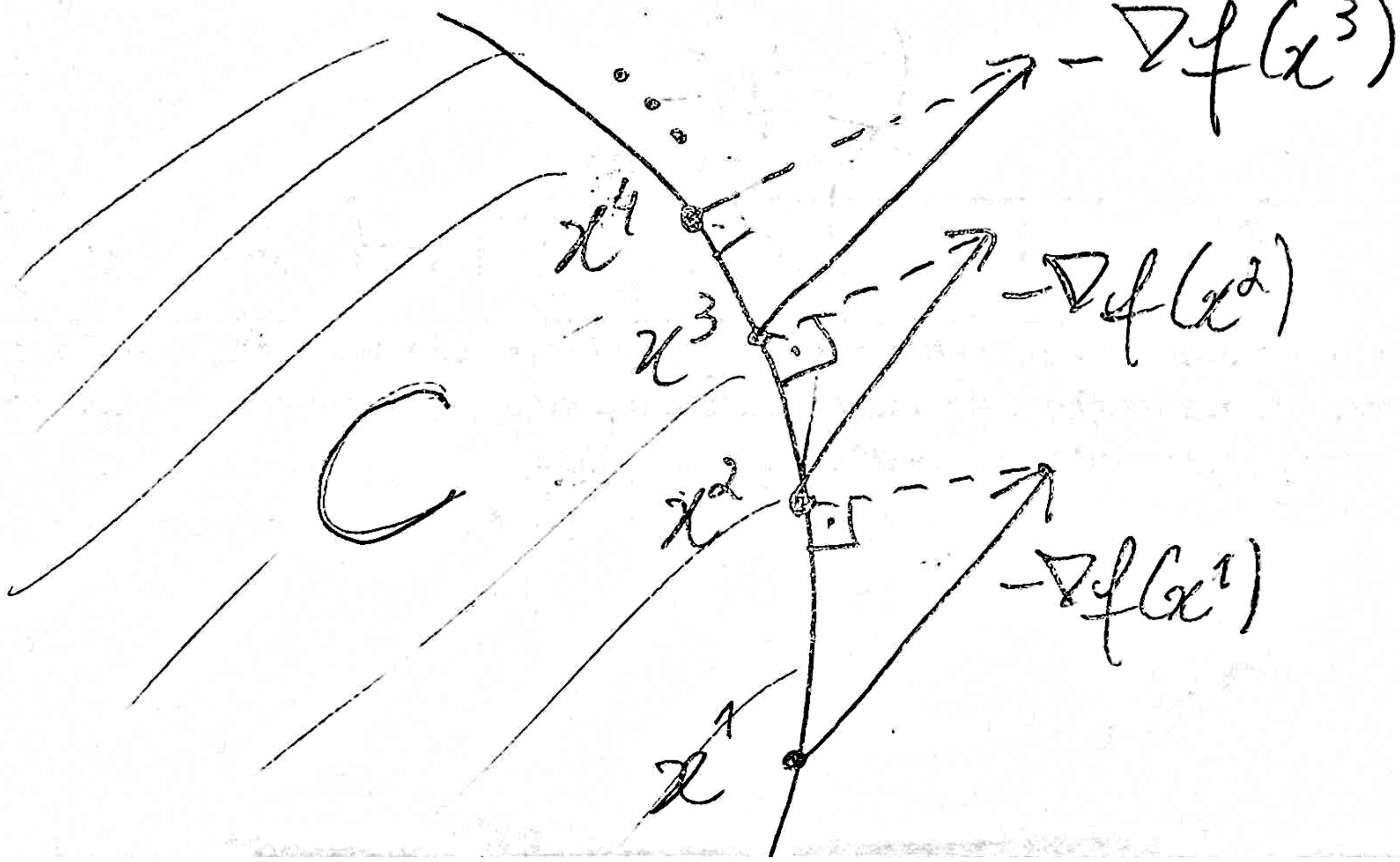
Ideia: Adaptar o método do gradiente

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

para garantir viabilidade:

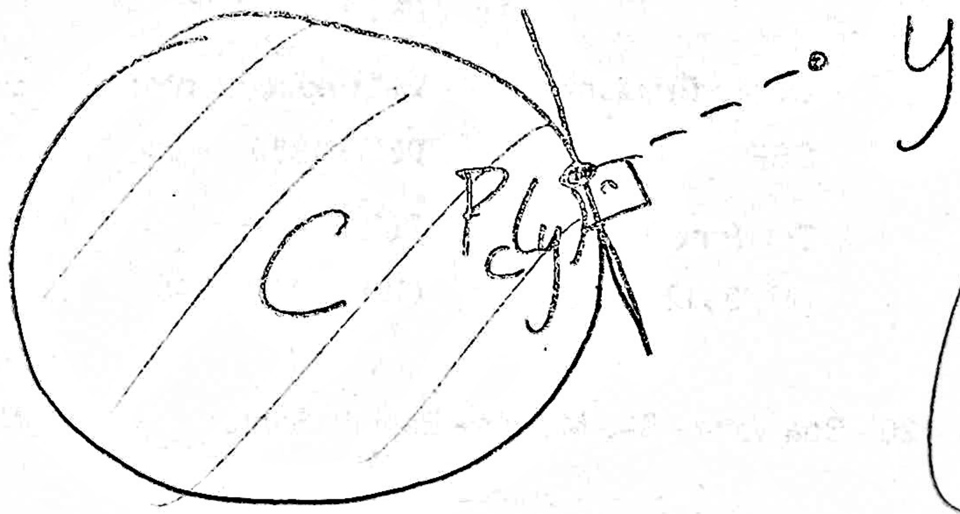
$$x^{k+1} \in C.$$

Como? 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 $P_C(y) \in C$ <u>mais próximo</u> à y.



$P_C(y)$ é a <u>projeção</u> de y sobre C.

Ou seja, $x^* = P_C(y)$ é a solução do problema 4

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

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

De fato, o problema de projeção 15

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

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

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

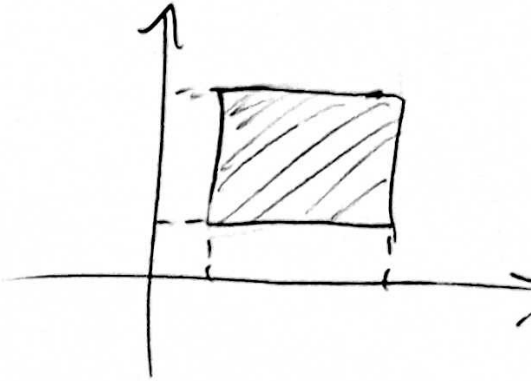
EXERCÍCIO: Use as condições KKT para mostrar que

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

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

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

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



Problema projeção:

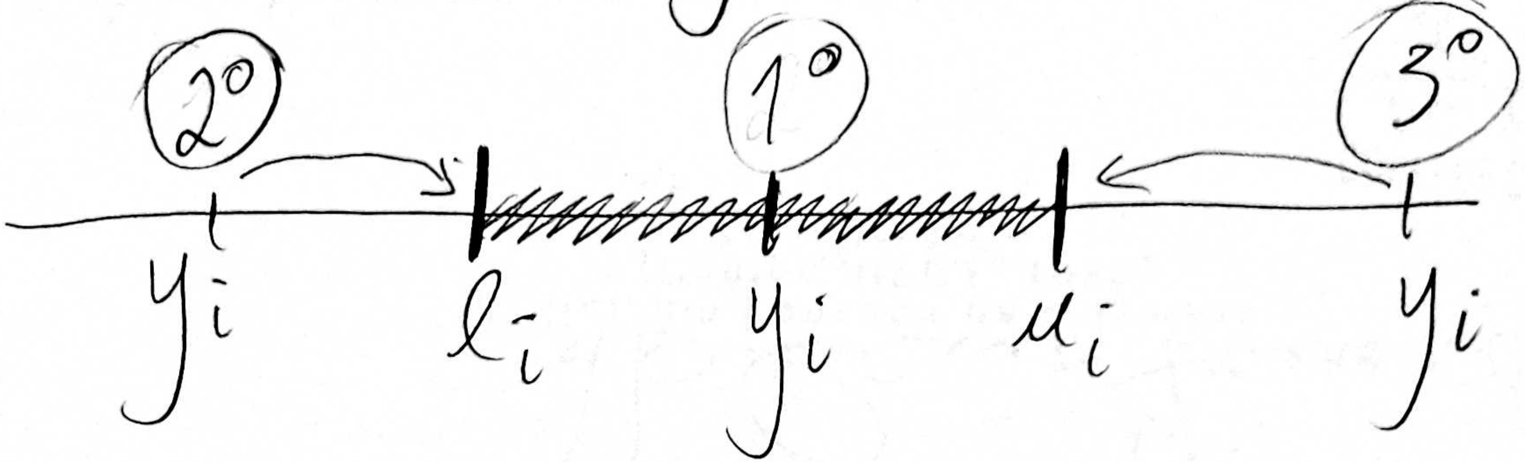
$$\min \frac{1}{2}\|x-y\|^2 \ \text{s.a.} \ -x+l \leq 0 , \ x-u \leq 0$$

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

$\mu^l \geq 0 , \ \mu^u \geq 0 , \ l \leq x^* \leq u$ (2)

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

3 casos para y_i:



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

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

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

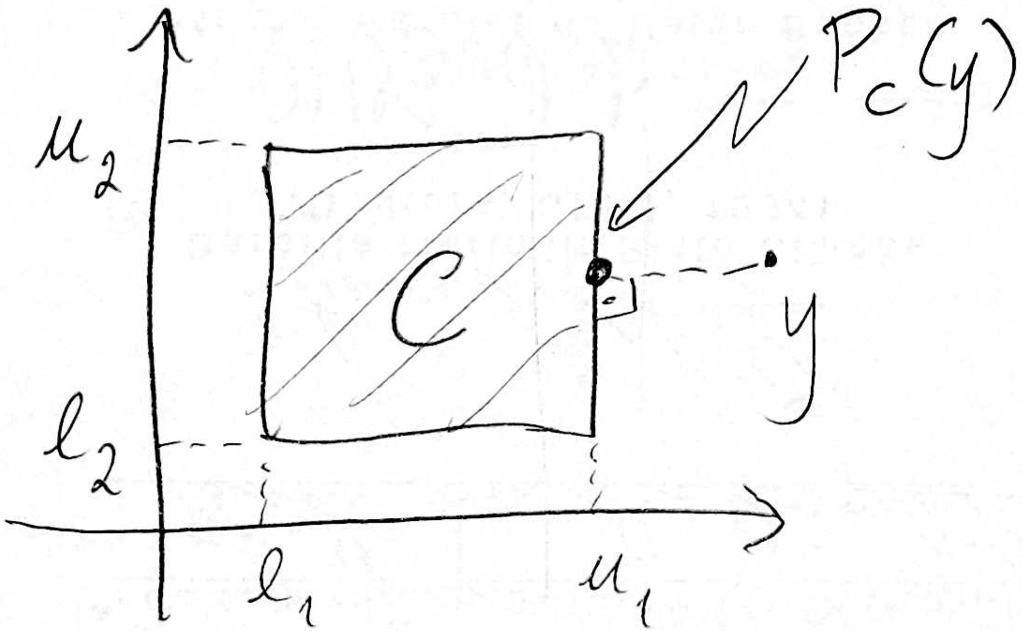
<u>Em resumo:</u>

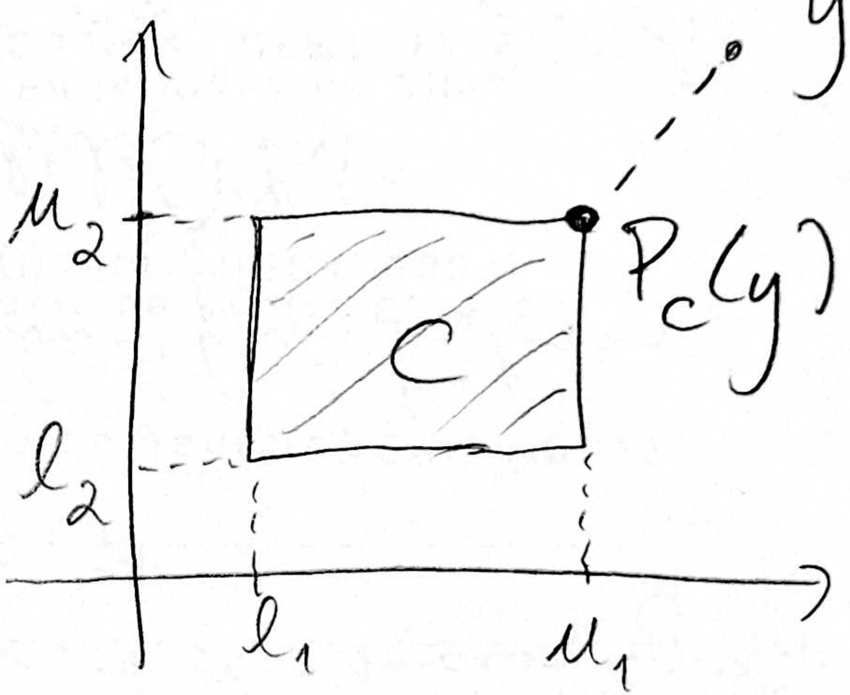
$$[P_C(y)]_i = \begin{cases} l_i & \text{se} \quad y_i \le l_i \\ y_i & \text{se} \quad l_i < y_i < u_i \\ u_i & \text{se} \quad y_i \ge u_i \end{cases}$$

Em implementações, fazemos, $\forall i$,

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

(verifique que essa é a projeção)





<u>Como parar?</u> Ou, o que é KKT para o problema original

$$\min f(x)$$

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

$\boxed{\text{Resposta}}$: 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.

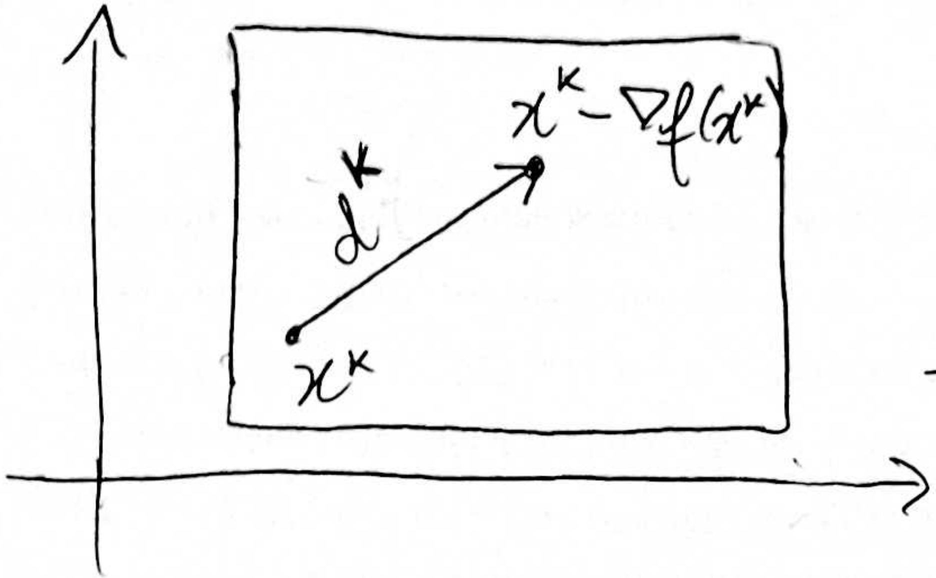
  $\hookrightarrow x^{k+1} = x^k - t_k \nabla f(x^k)\,, \; t_k > 0.$

- Restrições de caixa:

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

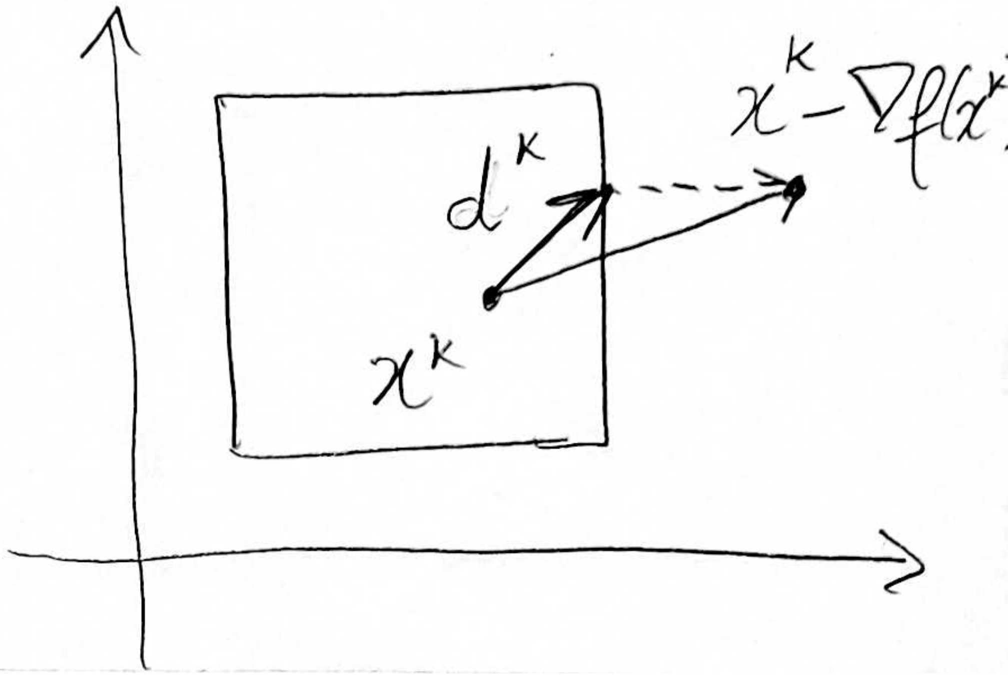
Observe que se $x^k - \nabla f(x^k) \in C$ então

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



(passo <u>dentro da caixa</u> = passo min. <u>sem restrições</u>).

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

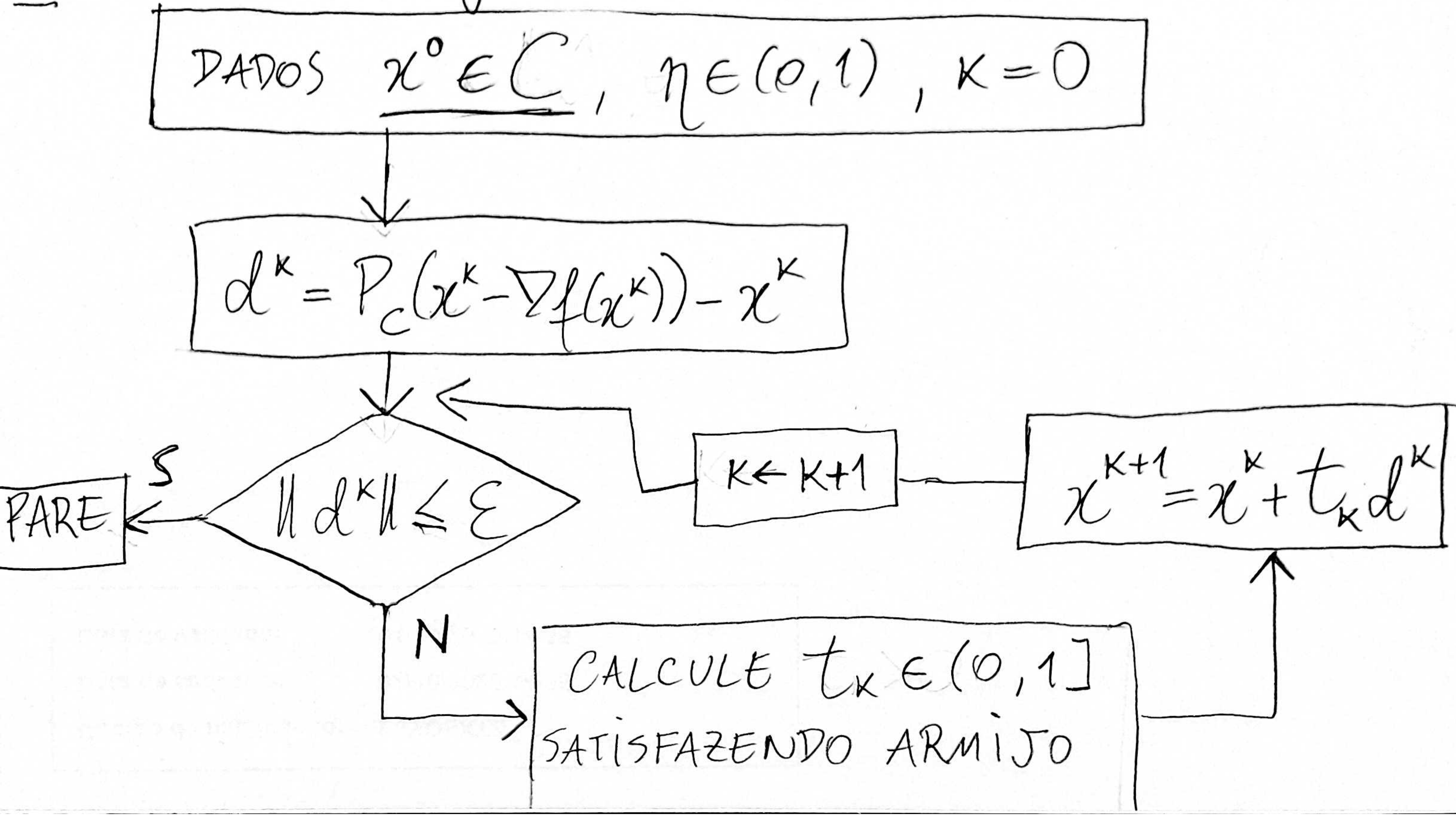


<u>EXERCÍCIO</u>: mostre que $x^k + d^k \in C$.

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

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

Método do gradiente projetado



Observações:

1) 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 :()

2) É exigido que o ponto inicial pertença à C $(x^0 \in C)$.

Na prática

- Escolhemos qualquer $x^0 \in \mathbb{R}^n$;
- Atualizamos $x^0 \leftarrow P_C(x^0)$.