

Método do subgradiente (para convexas) (1)

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

caso diferenciável | caso não diferenciável

$$x^* \text{ min } f \Leftrightarrow \nabla f(x^*) = 0 \\ (\text{global})$$

método gradiente:

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

$$t_k > 0$$

$$x^* \text{ min. } f \Leftrightarrow 0 \in \partial f(x^*) \\ (\text{global})$$

método subgradiente:

$$x^{k+1} = x^k - t_k g^k, \quad g^k \in \partial f(x^k)$$

$$t_k > 0$$

Teorema (condição de optimidade)

(2)

Seja $f: \mathbb{R}^n \rightarrow \mathbb{R}$ convexa. Então x^* é minimizador de f se, e somente se,
 $0 \in \partial f(x^*)$.

Prova: \leftarrow) $0 \in \partial f(x^*) \Rightarrow f(x^*) = f(x^*) + 0^t(y - x^*) \leq f(y), \forall y \Rightarrow x^*$ é minimizador global de f .

\Rightarrow) Se x^* é minimizador de f então

a fim de satisfazer

(3)

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

é suficiente escolher $g = 0$, dado que

$$f(y) \geq f(x^*) , \forall y . \text{ Isto é, } 0 \in \partial f(x^*) \blacksquare$$

Observações:

- 1) Quando f é diferenciável em x^* , o resultado diz que " $x^* \text{ min} \Leftrightarrow \nabla f(x^*) = 0$ ", o que recai no que já conhecíamos.

2) $0 \in \partial f(x^*)$ não significa que 14
 $\partial f(x^*)$ só possua 0 como subgradiente.
Por exemplo, $x^* = 0$ é minimizador
de $f(x) = |x|$ e $\partial f(0) = [-1, 1] \ni 0$.

3) O fato de $\partial f(x^*)$ poder conter subgradi-
entes não nulos atrapalha estabelecermos
um critério de parada para algoritmos:
um algoritmo poderia obter x^* sem que

pudéssemos decidir para declarando 15
"minimizadores encontrados", simplesmente
pelo fato que em geral não calculamos
todo o conjunto $\partial f(x^*)$. Em geral apenas
1 subgradiente é computado.

Por exemplo, considere $f(x) = |x|$ e a
sequência $x_k = \frac{1}{k} \rightarrow x^* = 0$. Apesar de
convergir ao minimizador, $\partial f(x_k) = \{-1\}$.

Vejá que mesmo se $x^* = 0$ for alcançado, um algoritmo poderia computar

$$1 \in \partial f(x^*) = [-1, 1].$$

(geralmente é isso que acontece!).

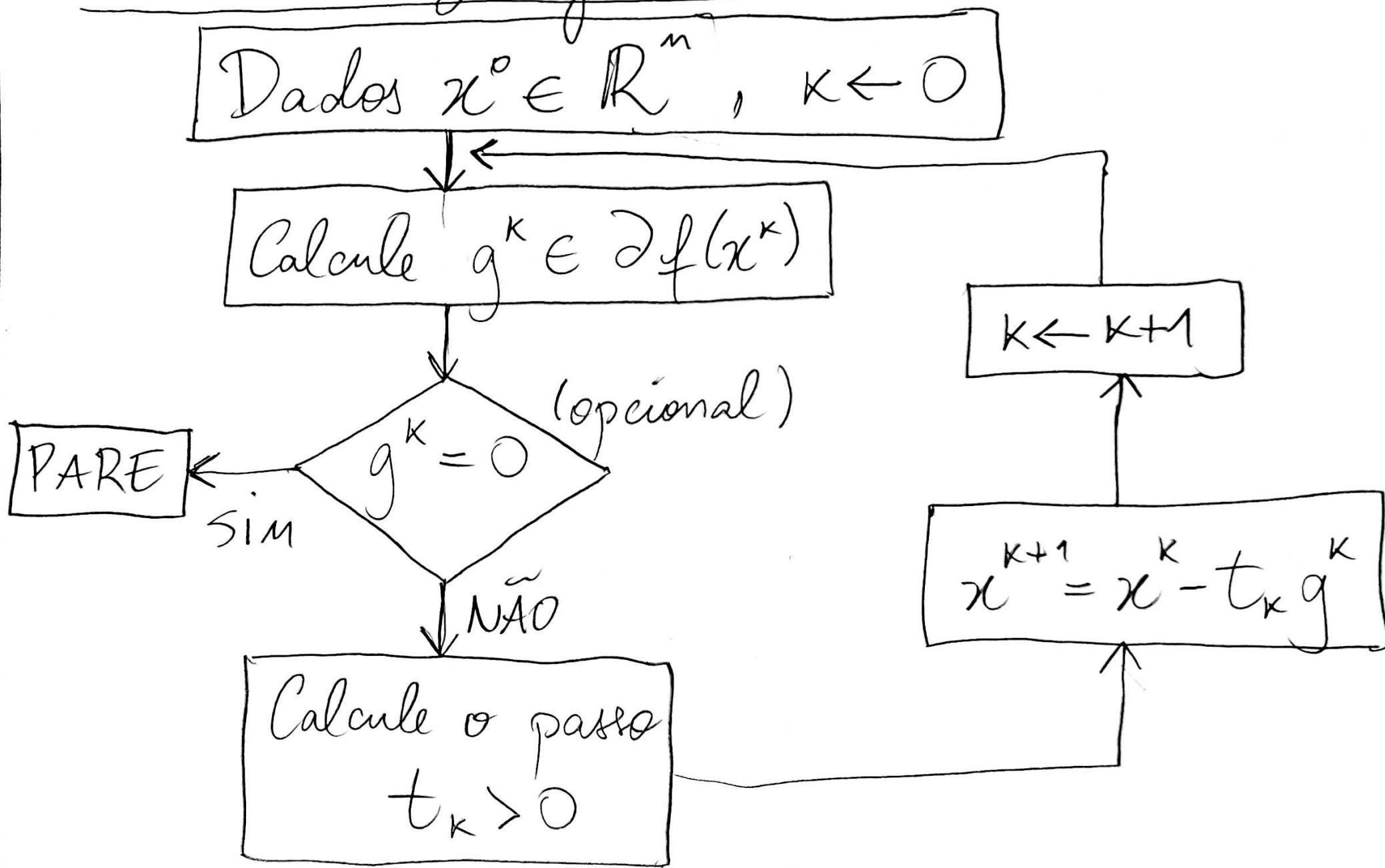
→ há formas de tentar contornar isso, p. ex., calculando gradientes randomicamente ao redor de x^* e tomando uma média, ou utilizando subdiferenciais aproximados... (não precisaremos disso!)

4) Portanto o critério por "máximo de iterações atingida" poderá ser acionado mesmo que já estejamos indo à solução.

5) Agora, se dermos "sorte" de calcular um subgradiente $g \approx 0$, podemos parar! O teorema anterior garante um minimizador. Vamos fazer o teste de parada $\|Ng\| \leq \epsilon$ pois é barato.

Método do subgradiente

18



Convergência do método do subgradiente 19 para funções convexas

- $f: \mathbb{R}^n \rightarrow \mathbb{R}$ convexa (\Rightarrow método bem definido)
- $\{x^k\}$ sequência gerada pelo método.

Um resultado fundamental:

Lema: Para todo $y \in \mathbb{R}^n$ temos

$$\|x^{k+1} - y\|^2 \leq \|x^k - y\|^2 - 2t_k (f(x^k) - f(y)) + t_k^2 \|g_k\|^2.$$

$$\begin{aligned}
 \text{Prova: } & \|x^{k+1} - y\|^2 = \|x^k - t_k g^k - y\|^2 \\
 &= \|x^k - y\|^2 - 2t_k (g^k)^t (x^k - y) + t_k^2 \|g^k\|^2 \\
 &\leq \|x^k - y\|^2 - 2t_k (f(x^k) - f(y)) + t_k^2 \|g^k\|^2,
 \end{aligned}
 \tag{10}$$

onde a ultima desigualdade segue do

facto $g^k \in \partial f(x^k)$



Para f convexa, é comum estudar a convergência nos seguintes casos:

- 1) passo constante: $t_k = t > 0, \forall k$
- 2) passo decrescente: $\exists t_k \in \mathbb{R}$ tal que $t_k > 0, \forall k, t_k \rightarrow 0$ e $\sum_{k=0}^{\infty} t_k = \infty$.

A ideia é refinar a busca próximo à solução ($t_k \rightarrow 0$) sem dar passos muito pequenos ($\sum_0^\infty t_k = \infty$). Por exemplo, $t_k = \frac{1}{k}$.

3) t_k escolhido "dinamicamente", tendo [12] em vista limitantes para o valor ótimo f^* (costuma funcionar melhor na prática).

Hipótese comum:

H1) Existe uma constante $c \geq \sup_{k=0, \dots} \|g^k\|$.

Obs: a compactidade de $\partial f(x^k)$ não implica H1, pois c independe de k .

1) Convergência com passo constante $t_k = t$. 13

Teorema: Suponha válida H1 e $t_k = t > 0, \forall k$.

(i) Se $f^* = \inf f(x) = -\infty$ então
 $f_\infty = \liminf_{k \rightarrow \infty} f(x^k) = f^*$

(ii) Se $f^* > -\infty$ (então)
 $f_\infty \leq f^* + \frac{t c^2}{2}$

Antes de demonstrar o teorema, note que

14

(ii) não garante que o método atinja o valor ótimo f^* .

Isto é coerente, pois com passo constante, o máximo de garantia é convergir à uma vizinhança do minimizador!



Prova: Vamos mostrar (i) e (ii) simultaneamente. Suponha que o resultado não valha. Então existe $\epsilon > 0$ constante tal que

$$f_\infty > f^* + \frac{tc^2}{2} + 2\epsilon$$

(vale para $f^* = -\infty$ e $f^* > -\infty$). Existe $y \in \mathbb{R}^n$ tal que

$$f_\infty \geq f(y) + \frac{tc^2}{2} + 2\epsilon. \quad (1)$$

Também, para todo $k \geq 1$, temos $f(x^k) \geq f_\infty - \epsilon$

pois $f_{\infty} = \liminf_{k \rightarrow \infty} f(x^k)$ (de fato, se 16

$f_{\infty} = -\infty$ então $f(x^k) \geq f_{\infty}$ e se $f_{\infty} > -\infty$

então $\{f(x^k)\}$ é limitada inferiormente, e logo

$\{\inf_{k \geq \bar{k}} f(x^k)\}_{\bar{k} \rightarrow \infty} \rightarrow f_{\infty}$). Somando (1) e (2)

obtemos, para um certo K_0 ,

$$f(x^k) - f(y) \geq \frac{tc^2}{2} + \epsilon, \quad \forall k \geq K_0.$$

Do lema e da hipótese H1, segue que

$$\begin{aligned}
 \|x^{k+1} - y\|^2 &\leq \|x^k - y\|^2 - 2t(f(x^k) - f(y)) + t^2 \|g^k\|^2 \quad [17] \\
 &\leq \|x^k - y\|^2 - 2t\left(\frac{tc^2}{2} + \epsilon\right) + t^2 c^2 \\
 &= \|x^k - y\|^2 - 2t\epsilon, \quad \forall k \geq k_0
 \end{aligned}$$

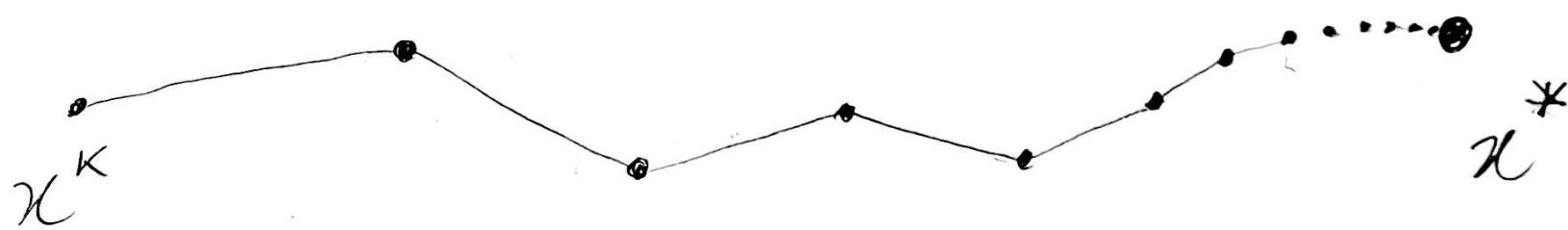
Aplicando essa desigualdade sucessivamente,

$$\begin{aligned}
 \|x^{k+1} - y\|^2 &\leq \|x^k - y\|^2 - 2t\epsilon \\
 &\leq \|x^{k-1} - y\|^2 - 4t\epsilon \\
 \dots &\leq \|x^{k_0} - y\|^2 - 2(k+1-k_0)t\epsilon.
 \end{aligned}$$

Tomando $K \gg K_0$ obtemos uma contradição [18].
 São pois $\|x^{K_0} - y\|^2 - 2(K+1-K_0)t\epsilon \xrightarrow[k \rightarrow \infty]{} -\infty$.
 Isso completa a demonstração. 

2) Convergência com passo decrescente

$$t_k \rightarrow 0^+, \sum_{k=0}^{\infty} t_k = \infty \quad x^k \xrightarrow{\sim} x^*$$



Teorema: Suponha válida H1 e considere L19

$\{t_k\}$ tal que $t_k > 0, \forall k,$

$$t_k \rightarrow 0 \quad \text{e} \quad \sum_{k=0}^{\infty} t_k = \infty.$$

Então

$$f_\infty = \liminf_{k \rightarrow \infty} f(x^k) = f^*,$$

onde $f^* = \inf f(x)$ (que pode ser $= -\infty$ ou $> -\infty$).

Prova: Suponha por contradição que existe $\epsilon > 0$ tal que

(20)

$$f_\infty - 2\epsilon > f^*$$

Existe então $y \in \mathbb{R}^n$ tal que

$$f_\infty - 2\epsilon > f(y). \quad (1)$$

Para todo $k \gg 1$, temos

$$f(x_k) \geq f_\infty - \epsilon. \quad (2)$$

Assim, somando (1) e (2) obtemos

$f(x^k) - f(y) > \varepsilon$, $\forall k \gg 1$. Do Lema 21.

e da hipótese H1, segue que

$$\begin{aligned}\|x^{k+1} - y\|^2 &\leq \|x^k - y\|^2 - 2t_k(f(x^k) - f(y)) + t_k^2 \|g^k\|^2 \\ &\leq \|x^k - y\|^2 - 2t_k \varepsilon + t_k^2 c^2 \\ &= \|x^k - y\|^2 - t_k(2\varepsilon - t_k c^2), \quad \forall k \gg 1.\end{aligned}$$

Como $t_k \rightarrow 0$, temos $2\varepsilon - t_k c^2 \geq \varepsilon$, $\forall k \gg 1$.

Agora, como todas as desigualdades anteriores valem $\forall k \gg 1$, podemos supor que isso

Ocorra $\forall k \geq K_0$. Daí

(22)

$$\|x^{k+1} - y\|^2 \leq \|x^k - y\|^2 - t_k \epsilon, \quad \forall k \geq K_0.$$

Aplicando essa desigualdade sucessivamente,

$$\begin{aligned} \|x^{k+1} - y\|^2 &\leq \|x^k - y\|^2 - t_k \epsilon \leq \|x^{k-1} - y\|^2 - \epsilon(t_{k-1} + t_k) \\ &\leq \dots \leq \|x^{K_0} - y\|^2 - \epsilon \sum_{j=K_0}^k t_j. \end{aligned}$$

Fazendo $k \rightarrow \infty$, obtemos uma contradição

pois $\sum_{j=K_0}^{\infty} t_j = \infty$. Isso conclui a demonstração



Obs: diferentemente do passo constante, com passo decrescente atingimos o valor ótimo f^* .

3) Passo dinâmico.

Objetivo: ajustar t_k levando em conta as frequências geradas pelo método. A atualização de t_k é feita passo a passo (passo dinâmico).

Li inspiração vem do seguinte resultado: (24)

Lema 2: Se $f(y) < f(x^k)$ então

$$\|x^{k+1} - y\| \leq (\leq) \|x^k - y\|$$

para todos os passos t_k tais que

$$0 < t_k \leq \frac{2(f(x^k) - f(y))}{\|g^k\|^2}.$$

Prova: Segue diretamente do Lema anterior
(exercício) 

Tomando $y = x^*$, o Lema 2 sugere .25

tomar

$$t_k = \frac{2(f(x^k) - f^*)}{\|g^k\|^2}.$$

Como $f^* < f(x^k)$, o Lema 2 garante que
 $\|x^{k+1} - x^*\| \leq \|x^k - x^*\|$, isto é, nos approxima-
mos da solução x^* a cada passo. Lembrar-se
também que se $g^k = 0$, podemos parar. Logo
 t_k está bem definido.

Problema óbvio: não conhecemos f^* !!! L26

Possíveis saídas:

$$1) t_k = \frac{f(x^k) - f_k}{\|g^k\|^2}, \quad f_k = \min_{0 \leq j \leq k} f(x^j) - s_k,$$

$$s_{k+1} = \begin{cases} \theta s_k & \text{se } f(x^{k+1}) \leq f_k \\ (\max\{\beta s_k, \delta\} & \text{se } f(x^{k+1}) > f_k \end{cases}$$

Ideia: tenta aproximar f^* pelos valores $f(x^j)$ dos pontos $\bar{x}^0, \dots, \bar{x}^k$ já computados.

Aqui, $\theta > 1$, $\beta \in (0, 1)$ e $\delta > 0$ são parâmetros. [27]

É possível mostrar convergência desta estratégia:

$$\bullet f^* = -\infty \Rightarrow \inf_{k \geq 0} f(x^k) = f^*$$

$$\bullet f^* > -\infty \Rightarrow \inf_{k \geq 0} f(x^k) \leq f^* + \delta$$

(veja Proposição 3.2.8 de Bertsekas).

$$2) t_k = \frac{f(x^*) - f_k}{\max\{p, \|g^k\|^2\}}, \quad f_k \text{ o mesmo, } \underline{L28}$$

$p > 0$ parâmetro

Ideia: a mesma, tomando o cuidado de não dividir por $\|g^k\|^2 \approx 0$, ($\Rightarrow t_k \gg 1$)

$$3) t_k = \min\left\{p, \frac{f(x^*) - f_k}{\|g^k\|^2}\right\}$$

Ideia: a mesma da escolha anterior.

$$4) t_k = \frac{\lambda_k (f(x^k) - f_{\text{low}})}{\|g^k\|^2},$$

29

onde f_{low} é um limite inferior conhecido para f^* e $0 < \lambda_k \leq 2$ é escolhido da seguinte forma:

$$(i) \lambda_0 = 2$$

$$(ii) \lambda_{k+1} = \begin{cases} \lambda_k & \text{se ocorreu redução de } f \text{ nas últimas } K > 0 \\ \lambda_k/2 & \text{caso contrário} \end{cases}$$