Método do subgradiente (para convexas)

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

caso diferenciável	caso não diferenciável
x^* min f (global) $\iff \nabla f(x^*) = 0$	x^* min. f (global) $\iff 0 \in \partial f(x^*)$
método gradiente: $x^{k+1} = x^k - t_k \nabla f(x^k)$, $t_k > 0$	método subgradiente: $x^{k+1} = x^k - t_k g^k$, $g^k \in \partial f(x^k)$, $t_k > 0$

Teorema (Condição de otimalidade)

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

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

é suficiente escolher $g = 0$, dado que $f(y) \geq f(x^*)$, $\forall y$. Isto é, $0 \in \partial f(x^*)$ $\blacksquare$

Observações:

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

2) $0 \in \partial f(x^*)$ <u>não</u> significa que $\partial f(x^*)$ <u>só</u> 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 subgradientes não nulos atrapalha estabelecermos um critério de parada para algoritmos: um algoritmo poderia obter x^* <u>sem</u> que

pudéssemos decidir parar declarando "minimizador encontrado", 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 $x_k = \frac{1}{k} \xrightarrow[k\to\infty]{} x^* = 0$. Apesar de convergir ao minimizador, $\partial f(x_k) = \{1\}$.

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

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

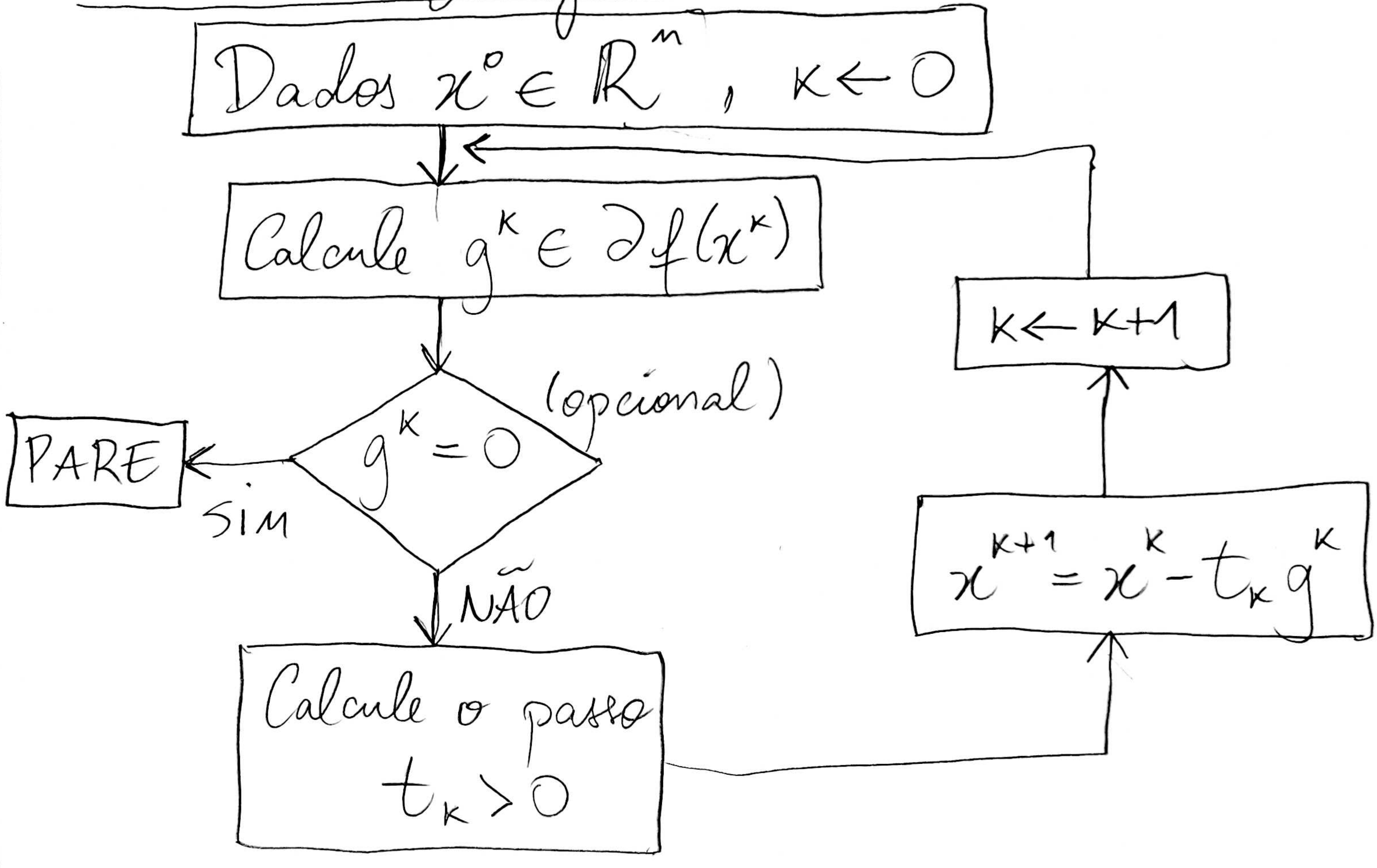
(geralmente é isso que acontece!).

↳ 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 atingido" poderá ser acionado mesmo que já estejamos indo à solução. 7

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 $\|g\| \leq \varepsilon$ pois é barato.

Método do subgradiente



Convergência do método do subgradiente para funções convexas

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

Um resultado fundamental:

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

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

Prova: $\|x^{k+1}-y\|^2 = \|x^k - t_k g^k - y\|^2$

$$= \|x^k-y\|^2 - 2t_k (g^k)^t(x^k-y) + t_k^2\|g^k\|^2$$

$$\leq \|x^k-y\|^2 - 2t_k(f(x^k)-f(y)) + t_k^2\|g^k\|^2,$$

onde a última desigualdade segue do fato $g^k \in \partial f(x^k)$ $\blacksquare$

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

1) passo constante: $t_k = t > 0$, $\forall k$

2) passo decrescente: $\{t_k\}$ tal que $t_k > 0$, $\forall k$, $t_k \to 0$ e $\sum_{k=0}^{\infty} t_k = \infty$.

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

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

Hipótese comum:

H1) Existe uma constante $c \geq \sup_{k=0,1,\dots} \|g^k\|$.

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

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

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

(i) Se $f^* = \inf f(x) = -\infty$ então

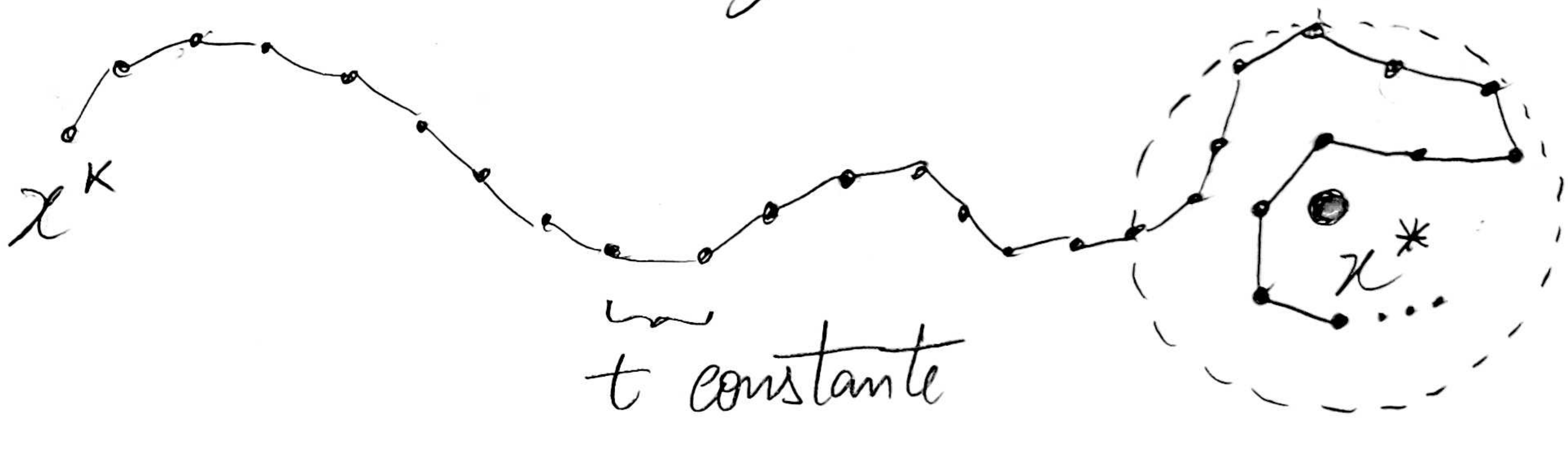
$$f_\infty = \liminf_{k \to \infty} f(x^k) = f^*$$

(ii) Se $f^* > -\infty$ então

$$f_\infty \leq f^* + \frac{t c^2}{2}.$$

Antes de demonstrar o teorema, note que (ii) <u>não</u> garante que o método atinja o valor ótimo f^*:

Isso 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 $\varepsilon > 0$ constante tal que

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

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

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

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

(2)

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

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

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

$\{\inf_{k \geq \bar{k}} f(x^k)\} \xrightarrow[\bar{k}\to\infty]{} f_\infty$). Somando (1) e (2)

obtemos, para um certo k_0,

$$f(x^k) - f(y) \geq \frac{tc^2}{2} + \varepsilon, \quad \forall k \geq k_0 .$$

Do lema e da hipótese H1, segue que

$$\|x^{k+1}-y\|^2 \leq \|x^k-y\|^2 - 2t\left(f(x^k)-f(y)\right) + t^2\|g^k\|^2$$

$$\leq \|x^k-y\|^2 - 2t\left(\frac{tc^2}{2}+\varepsilon\right) + t^2c^2$$

$$= \|x^k-y\|^2 - 2t\varepsilon\,, \quad \forall k \geq k_0$$

Aplicando essa desigualdade sucessivamente,

$$\|x^{k+1}-y\|^2 \leq \|x^k-y\|^2 - 2t\varepsilon$$

$$\leq \|x^{k-1}-y\|^2 - 4t\varepsilon$$

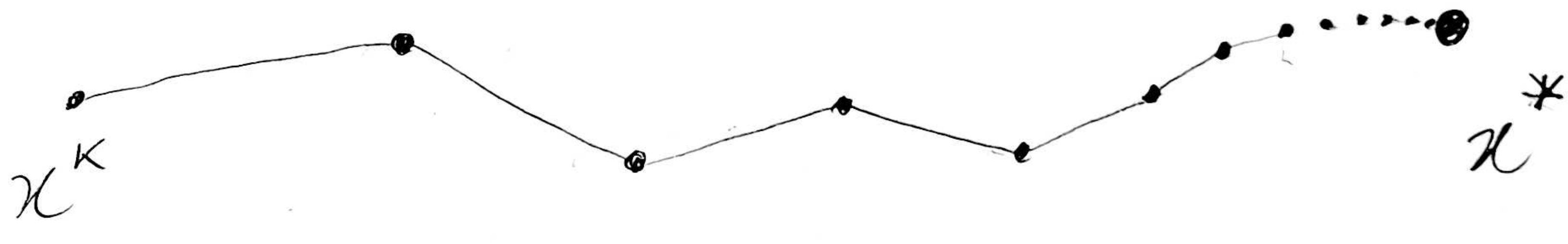
$$\cdots \leq \|x^{k_0}-y\|^2 - 2(k+1-k_0)t\varepsilon.$$

Tomando $k \gg k_0$ obtemos uma contradição pois $\|x^{k_0} - y\|^2 - 2(k+1-k_0)t\varepsilon \xrightarrow[k\to\infty]{} -\infty$.

Isso completa a demonstração. ■

2) Convergência com passo decrescente

$$t_k \to 0^+, \quad \sum_{k=0}^{\infty} t_k = \infty \qquad x^k \to x^*$$



<u>Teorema</u>: Suponha válida H1 e considere $\{t_k\}$ tal que $t_k > 0$, $\forall k$,

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

Então

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

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

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

$$f_\infty - 2\varepsilon > f^*.$$

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

$$f_\infty - 2\varepsilon > f(y). \qquad (1)$$

Para todo $k \gg 1$, temos

$$f(x^k) \geq f_\infty - \varepsilon. \qquad (2)$$

Assim, somando (1) e (2) obtemos

$f(x^k) - f(y) > \varepsilon$, $\forall k \gg 1$. Do lema e da hipótese H1, segue que

$$\|x^{k+1} - y\|^2 \leq \|x^k - y\|^2 - 2t_k\left(f(x^k) - f(y)\right) + t_k^2\|g^k\|^2$$

$$\leq \|x^k - y\|^2 - 2t_k\varepsilon + t_k^2 c^2$$

$$= \|x^k - y\|^2 - t_k\left(2\varepsilon - t_k c^2\right) , \forall k \gg 1.$$

Como $t_k \to 0$, 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

$$\|x^{k+1} - y\|^2 \leq \|x^k - y\|^2 - t_k \varepsilon \quad , \quad \forall k \geq k_0 .$$

Aplicando essa desigualdade sucessivamente,

$$\|x^{k+1} - y\|^2 \leq \|x^k - y\|^2 - t_k \varepsilon \leq \|x^{k-1} - y\|^2 - \varepsilon(t_{k-1} + t_k)$$

$$\leq \cdots \leq \|x^{k_0} - y\|^2 - \varepsilon \sum_{j=k_0}^{k} t_j .$$

Fazendo $k \to \infty$, 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 sequências geradas pelo método. A atualização de t_k é feita passo a passo (passo dinâmico).

A inspiração vem do seguinte resultado: 24

Lema 2: Se $f(y) < f(x^k)$ então

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

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) $\blacksquare$

Tomando $y = x^*$, o lema 2 sugere 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 aproximamos da solução x^* a cada passo. Lembre-se também que se $g^k = 0$, podemos parar. Logo t_k está bem definido.

Problema óbvio: não conhecemos f^* !!!

Possíveis saídas:

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

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

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

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

É possível mostrar convergência desta estratégia:

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

(veja Proposição 3.2.8 de Bertsekas).

2) $t_k = \dfrac{f(x^k) - f_k}{\max\{\rho, \|g^k\|^2\}}$, f_k o mesmo, $\rho > 0$ parâmetro

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

---

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

Ideia: a mesma da escolha anterior.

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

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

(i) $\lambda_0 = 2$

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