

ALGORITMOS PARA MINIMIZAÇÃO IRRESTRITA.

"DADO UM x^k , COMO CALCULO x^{k+1} ?"

UM ALGORITMO ITERATIVO INTERESSANTE GERA UMA SEQ.
 $\{x^k\}$ TAL QUE $\lim_{k \rightarrow \infty} \nabla f(x^k) = 0$.

x^k : "ITERANDO".

PRÁTICA: CRITÉRIO DE PARADA É $\|\nabla f(x^k)\| \leq \varepsilon$,

PARA $\varepsilon > 0$ DADO.

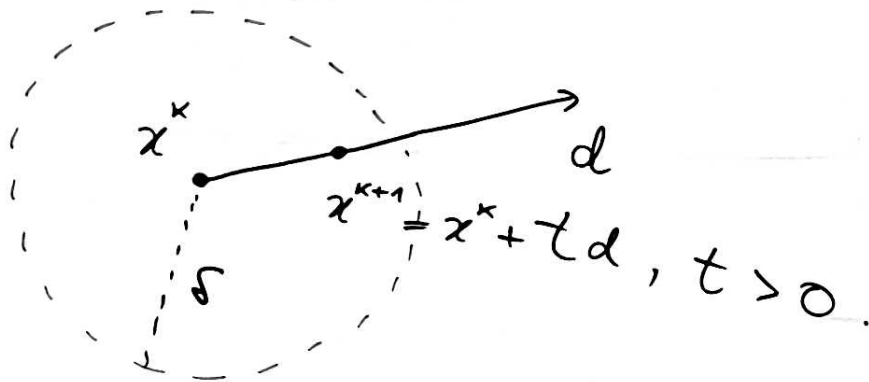
TEORIA: ACEITO QUE UM ALGORITMO "RODE INFINITAMENTE".

IDEIA BÁSICA P/ UM ALGORITMO:

A PARTIR DE x^k , OBTENHA x^{k+1} TAL QUE

$$f(x^{k+1}) < f(x^k).$$

ALGORITMOS DE DESCIDA



A DIREÇÃO d É
UMA DIREÇÃO DE DESCIDA:

f DECRESCER LOCALMENTE
NA DIREÇÃO d .

DEF.: d é uma DIREÇÃO DE DESCIDA PARA f A PARTIR DO PONTO x SE

$$f(x+td) < f(x), \quad \forall t \in (0, \delta],$$

PARA ALGUM $\delta > 0$.


ALGORITMO: DAR PASSOS EM DIREÇÕES DE DESCIDA, DIMINUINDO f !!!

COMO CARACTERIZAR d ?

TEO.: SEJA f DIFERENCIÁVEL. SE $\nabla f(x)^t d < 0$ ENTÃO d É DIREÇÃO DE DESCIDA PARA f A PARTIR DE x .

PROVA: TEMOS

$$0 > \nabla f(x)^T d = \lim_{t \rightarrow 0^+} \frac{f(x+td) - f(x)}{t}$$

Assim, $f(x+td) - f(x) < 0$, $\forall t$ PEQUENO, OU
SEJA, $f(x+td) < f(x)$, $\forall t$ PEQUENO. 

EXEMPLOS:

1) SE $\nabla f(x) \neq 0$, ENTÃO $d = -\nabla f(x)$ É DE DESCIDA.
(MÉTODO DO GRADIENTE)

2) SE $\nabla f(x) \neq 0$ E $\nabla^2 f(x) > 0$ ENTÃO $d = -(\nabla^2 f(x))^{-1} \nabla f(x)$
É DE DESCIDA. (MÉTODO DE NEWTON).

3) Se $\nabla f(x) \neq 0$ e B é uma matriz $n \times n$ DEF. POSITIVA,
então $d = -B \nabla f(x)$ é DE DESCIDA.
(MÉTODOS QUASE-NEWTON).

4) $f(x_1, x_2) = x_1^2 - x_2^2$, $\bar{x} = (1, 0)$, $\tilde{x} = (0, 0)$

$$\nabla f = \begin{bmatrix} 2x_1 \\ -2x_2 \end{bmatrix}.$$

$$\cdot \nabla f(\bar{x}) = \begin{bmatrix} 2 \\ 0 \end{bmatrix} \quad \bar{d} = (-1, -1).$$

$$\nabla f(\bar{x})^t \bar{d} = [2 \ 0] \begin{bmatrix} -1 \\ -1 \end{bmatrix} = -2 < 0.$$

\bar{d} é DE DESCIDA!

$$\cdot \nabla f(\tilde{x}) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}. \quad \tilde{d} = (0, 1).$$

$$\nabla f(\tilde{x})^t \tilde{d} = 0 \dots$$

$$f(\tilde{x} + t\tilde{d}) = f(0, t) = -t^2$$

$$< 0 = f(\tilde{x})$$

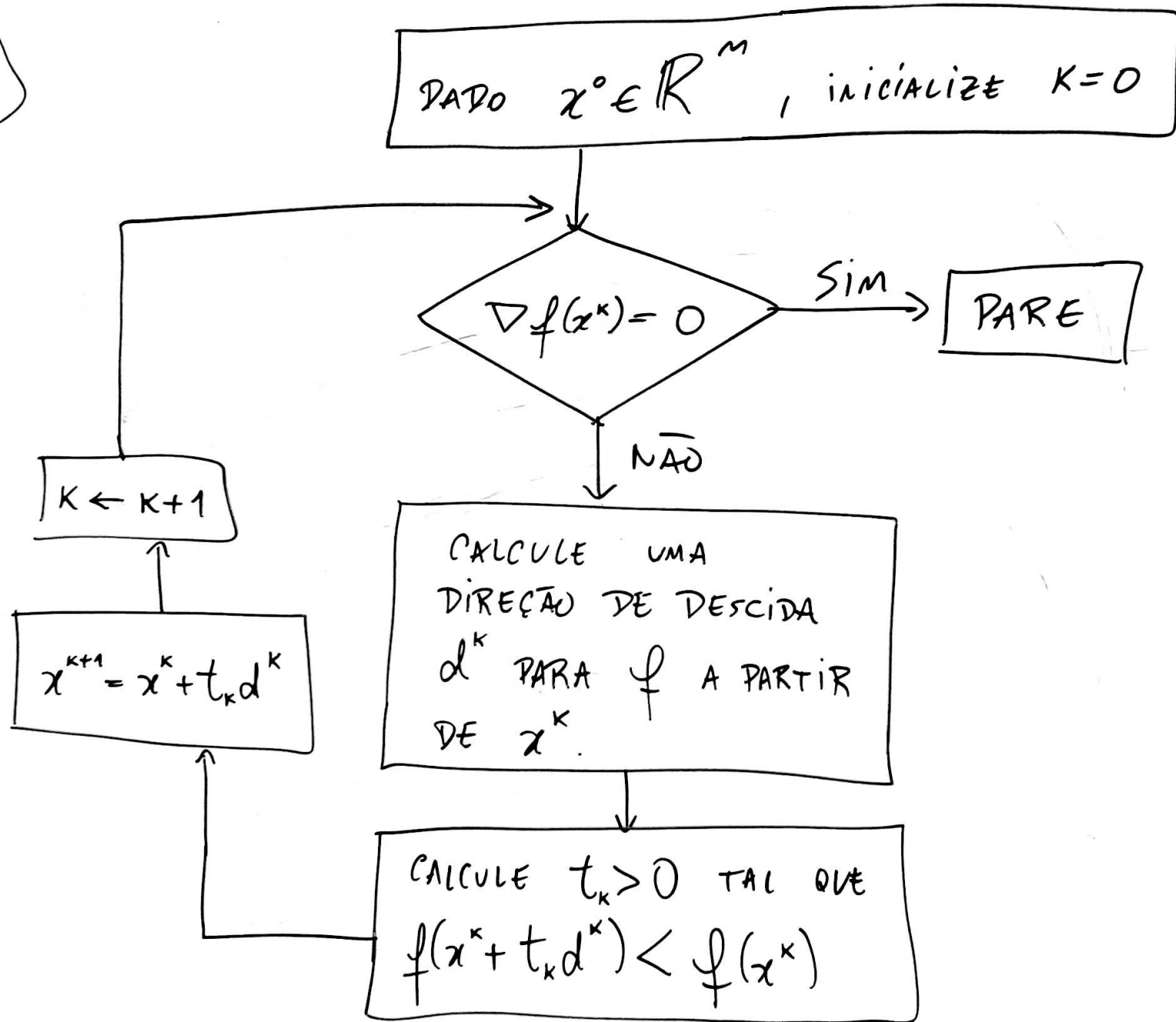
\Downarrow

\tilde{d} É DESCIDA \downarrow .

OU SEJA, A CONDIÇÃO $\nabla f(x)^t d < 0$ NÃO É
NECESSÁRIA PARA QUE d SEJA DE DESCIDA (NÃO VALE
A VOLTA DO TEOREMA).

ESQUEMA GENÉRICO DE UM ALGORITMO DE DESCIDA

$$\min_x f(x)$$



CRITÉRIOS DE PARADA PRÁTICOS.

• $\|\nabla f(x^k)\| \leq \varepsilon$, $\varepsilon > 0$ DADO (SUCESSO!)

• $k \geq k_{\max}$, $k_{\max} \geq 1$ (FRACASSO : ().

OBJETIVO: TORNAR O ESQUEMA ANTERIOR "PRÁTICO".

COMO CALCULAR O TAMANHO t_k DO PASSO?

(BUSCA LINEAR).

- BUSCA EXATA: CONSISTE EM RESOLVER

$$\min_{t \geq 0} f(x^k + t d^k). \quad \text{ESSE PROBLEMA NA VARIÁVEL } t$$

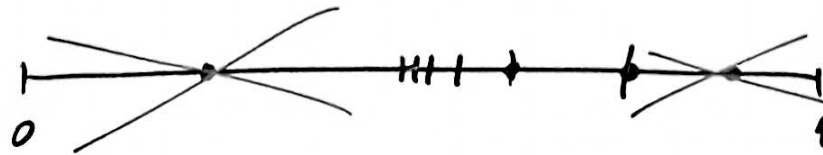
TEM SOLUÇÃO $t^* > 0$ POIS d^k É DE DESCIDA.

COMO RESOLVER?

* FÓRMULA FECHADA PARA t (DEPENDE DA SIMPLICIDADE DE

- POR EXEMPLO: $f(x) = \frac{1}{2} x^t A x + b^t x$, $A > 0$. f)

* UTILIZAR ALGUM ALGORITMO PARA MINIMIZAÇÃO EM 1 VARIÁVEL.
(MÉTODO DA SEÇÃO ÁUREA). (t_k APROXIMADO).



• BACKTRACKING

PASSO 1) COMECE COM $t = 1$.

PASSO 2) SE $f(x^k + t d^k) \geq f(x^k)$, DIMINUA
 t E TESTE NOVAMENTE.

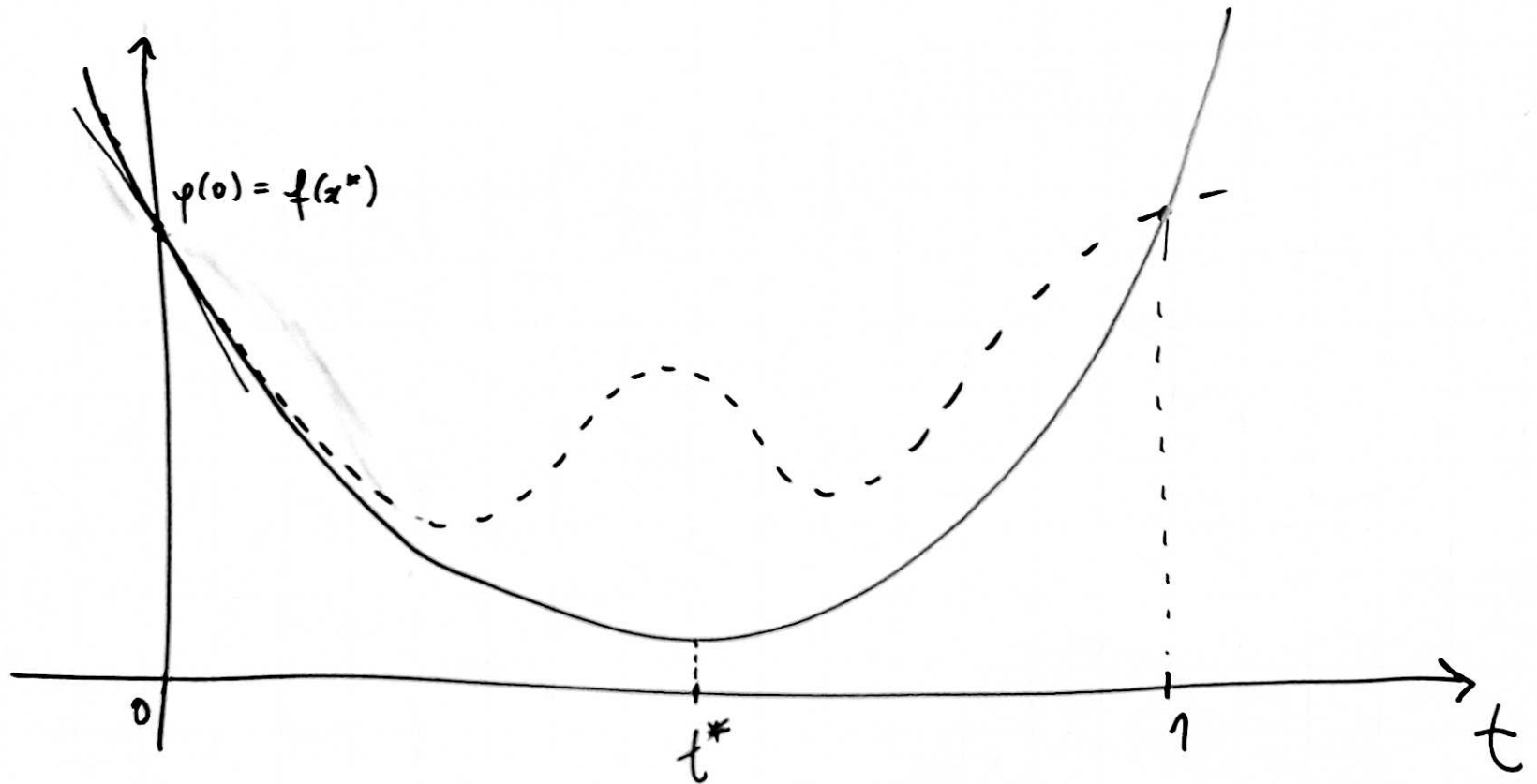
$$t \leftarrow \frac{t}{2}.$$

• INTERPOLAÇÃO QUADRÁTICA

• $\varphi(t) = f(x^k + t d^k)$. QUERO " $\min_{t \geq 0} \varphi(t)$ ".

• $\varphi'(t) = \nabla f(x^k + t d^k)^T d^k$.

• $\varphi'(0) = \nabla f(x^k)^T d^k < 0$ (d^k é DIR. DE DESCIDA, QU
VOU ESCOLHER ASSIM) .



EXISTE UMA PARÁBOLA $p(t) = at^2 + bt + c$ QUE PASSA POR $(0, \varphi(0))$, $(1, \varphi(1))$ E TEM INCLINAÇÃO $\varphi'(0)$ EM $t=0$.

- $p(0) = \varphi(0) \Rightarrow c = f(x^*)$.
- $p'(0) = \varphi'(0) \Rightarrow b = \nabla f(x^*)^T d^k$.
- $p(1) = \varphi(1) \Rightarrow a + b + c = f(x^* + d^k)$

$$\Rightarrow a = f(x^* + d^k) - f(x^*) - \nabla f(x^*)^T d^k.$$

OBS: $a > 0$ POIS A PARÁBOLA TEM CONCAVIDADE PARA CIMA.

RESOLVEMOS AGORA

$$\min_{t \geq 0} p(t) :$$

$$p'(t^*) = 2at^* + b = 0 \xRightarrow{a > 0} t^* = \frac{-b}{2a}$$

$$\Rightarrow t^* = - \frac{\nabla f(x^*)^T d^k}{2[f(x^k + d^k) - f(x^k) - \nabla f(x^k)^T d^k]}$$

OBSERVE QUE, DE FATO, $t^* > 0$ DADO QUE $\nabla f(x^k)^T d^k < 0$ E $a > 0$.

E SE ISSO DER ERRADO ($f(x^k + t^* d^k) \geq f(x^k)$) ?

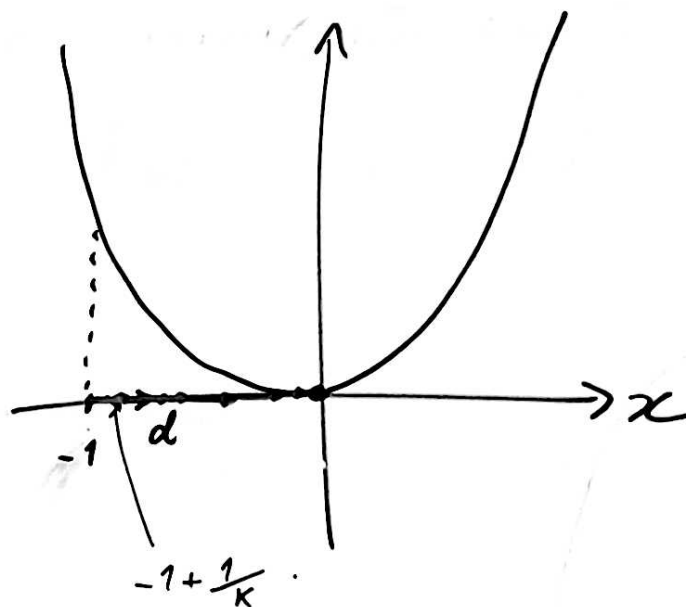
- TENTE OUTRA COISA... (BACKTRACKING).

EXEMPLO: $f(x) = x^2$.

- $x^k = -1$.

- $t_k = \frac{1}{k}$, $d^k = 1$.

$$f\left(x^k + \frac{1}{k}\right) < f(x^k).$$



$t_k = \frac{1}{k}$ É UM PASSO MUITO PEQUENO, O DECRÉSCIMO DE
 f É MUITO PEQUENO, E O ALGORITMO PODERIA
ESTACIONAR LONGE DO MINIMIZADOR!

- DECRÉSCIMO SIMPLES: $f(x^k + t_k d^k) < f(x^k)$ ~~X~~ RUIM.

- QUERO UM DECRÉSCIMO "GRANDE" (DECRÉSCIMO SUFICIENTE)

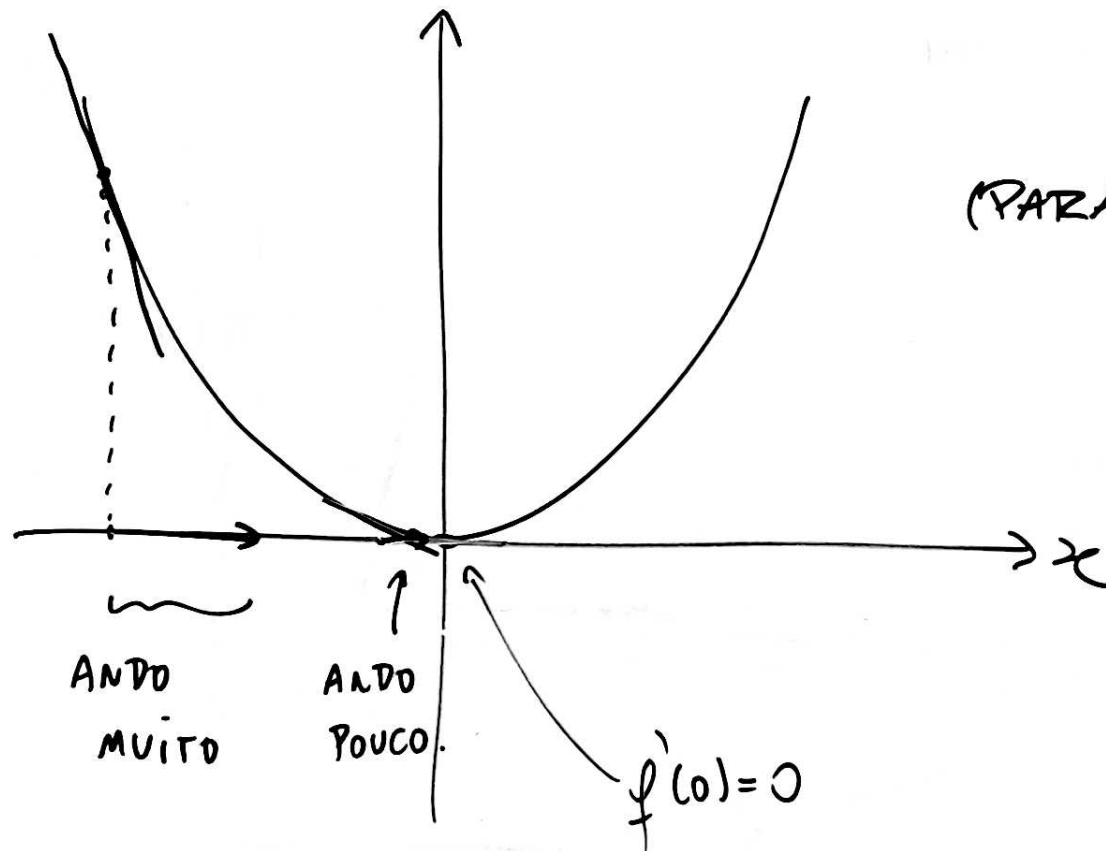
CONDIÇÃO DE ARMILHO.

$$f(x^k + t_k d^k) \leq f(x^k) + t_k \eta \nabla f(x^k)^T d^k,$$

ONDE $\eta \in (0, 1)$

FIXO.

(PARÂMETRO).



TEOREMA: SUPONHA QUE f SEJA DIFERENCIÁVEL EM x^k .

SEJAM d^k DIREÇÃO DE DESCIDA PARA f A PARTIR DE x^k E $\eta \in (0, 1)$. ENTÃO EXISTE $\delta > 0$ TAL QUE

$$f(x^k + t d^k) \leq f(x^k) + t \eta \nabla f(x^k)^T d^k, \quad \forall t \in [0, \delta].$$

PROVA: SE $\nabla f(x^*)^t d^k = 0$ ENTÃO O RESULTADO SEQUE DA
DEF. DE DIR. DE DESCIDA. SUPONHA $\nabla f(x^*)^t d^k < 0$.

TEMOS

$$\lim_{t \rightarrow 0^+} \frac{f(x^* + t d^k) - f(x^*)}{t} = \nabla f(x^*)^t d^k < \eta \nabla f(x^*)^t d^k$$

POIS $\eta < 1$ E $\nabla f(x^*)^t d^k < 0$. ASSIM, EXISTE
 $\delta > 0$ TAL, QUE

$$\frac{f(x^* + t d^k) - f(x^*)}{t} < \eta \nabla f(x^*)^t d^k, \quad \forall t \in (0, \delta)$$

$$\Rightarrow f(x^* + t d^k) \leq f(x^*) + t \eta \nabla f(x^*)^t d^k, \quad \forall t \in (0, \delta)$$

COMO CALCULAR UMA DIREÇÃO DE DESCIDA d^k ?

- DIREÇÕES d^k TAIS QUE $\nabla f(x^k)^T d^k < 0$.

* MÉTODO DO GRADIENTE

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

ESTA ESCOLHA VALE POIS $\nabla f(x^k)^T d^k = -\|\nabla f(x^k)\|^2 < 0$

($\nabla f(x^k) \neq 0$ POIS O MÉTODO NÃO PAROU).

* $d^k = -H^k \nabla f(x^k)$, ONDE H^k É DEFINIDA POSITIVA.

ESTA É VÁLIDA POIS

$$\nabla f(x^k)^T d^k = -\nabla f(x^k)^T H^k \nabla f(x^k) < 0.$$

1 POR EXEMPLO, $H^k = I$ RECAI NO MÉTODO DO GRADIENTE.

H^k É ATUALIZADA DURANTE O PROCESSO.

(MÉTODOS QUASE-NEWTON).

DUAS ESCOLHAS "FAMOSAS":

$$\odot H^0 = I,$$

$$H^{k+1} = H^k + \frac{(p_k p_k^T)}{p_k^T q_k} - \frac{H^k q_k q_k^T H^k}{q_k^T H^k q_k}$$

ONDE $p_k = x^{k+1} - x^k$, $q_k = \nabla f(x^{k+1}) - \nabla f(x^k)$.

(DFP — DAVIDON — FLETCHER — POWELL).

TEOREMA: SE H^k É DEF. POSITIVA, ENTÃO H^{k+1}
TAMBÉM É.

ASSIM, COMEÇANDO COM $H^0 = I$, TEREMOS UMA SEQUÊNCIA
DE MATRIZES $\{H^k\}$ DEFINIDAS POSITIVAS. OU SEJA,
AS DIREÇÕES $d^k = -H^k \nabla f(x^k)$ SERÃO DE DESCIDA.

$$\odot H^0 = I$$

$$H^{k+1} = H^k + \left(\frac{1 + q_k^T H^k q_k}{q_k^T p_k} \right) \frac{p_k p_k^T}{p_k^T q_k} - \frac{p_k q_k^T H^k + H^k q_k p_k^T}{q_k^T p_k}$$

ONDE p_k E q_k SÃO COMO ANTES.

) (BFGS — BROYDEN — FLETCHER — GOLDFARB — SHANNO)

O TEOREMA ANTERIOR VALE $(H^k > 0 \Rightarrow H^{k+1} > 0)$.

CARACTERÍSTICAS DOS MÉTODOS QUASE-NEWTON

- 1) TENTAM IMITAR NEWTON ("O MELHOR")
- 2) AS ITERAÇÕES SÃO BARATAS COMPUTACIONALMENTE
- 3) TENTAM SER MELHORES QUE O MÉTODO DO GRADIENTE.

SOBRE TEORIA: 'EQUAÇÕES SECANTE' (PROCURE!).

○ MÉTODO DE NEWTON

$$d^k = - \left(\nabla^2 f(x^k) \right)^{-1} \nabla f(x^k).$$

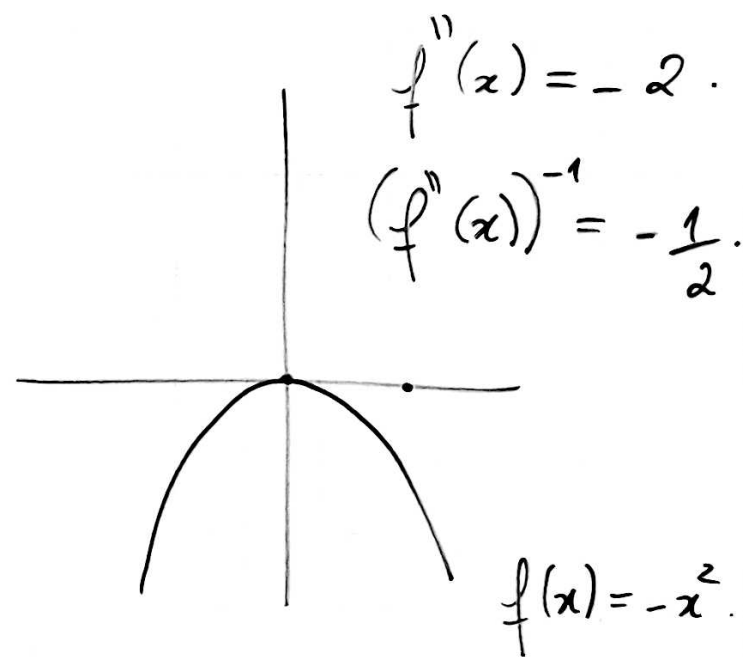
PROBLEMAS:

- 1) d^k PODE NÃO SER DIREÇÃO DE DESCIDA, NA MEDIDA EM QUE

$$\left(\nabla^2 f(x) \right)^{-1} \not\geq 0.$$

EXEMPLO: $f(x) = -x^2$.

$$f''(x) = -2 \Rightarrow \left(f''(x) \right)^{-1} = -\frac{1}{2}.$$



2) $\nabla^2 f(x)$ PODE SER SINGULAR (NÃO INVERSÍVEL).

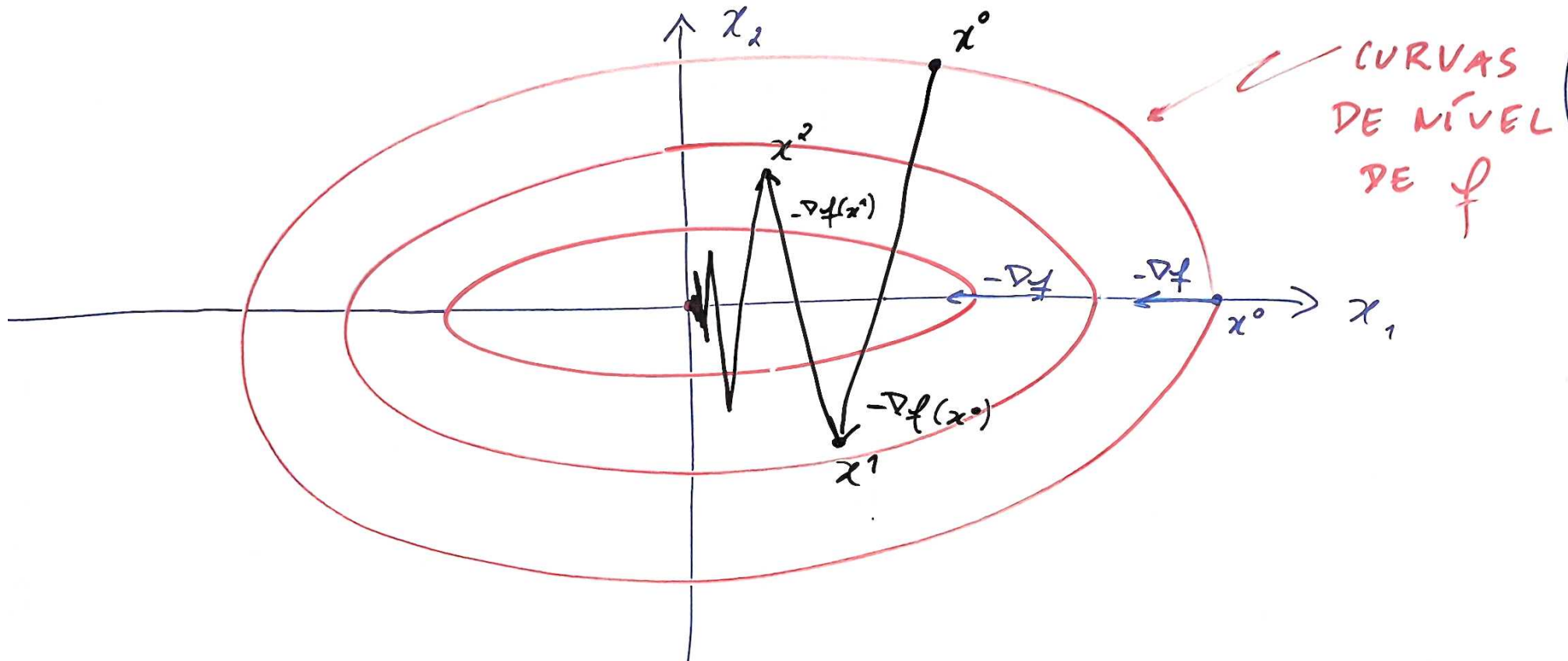
VANTAGEM

"QUANDA DÁ CERTO, DÁ MUITO CERTO!"

- O DECRÉSCIMO DE f É RÁPIDO PERTO DA SOLUÇÃO.
 - SOB CERTAS HIPÓTESES, A DIREÇÃO NEWTONIANA SEMPRE FUNCIONA PRÓXIMO À SOLUÇÃO.
-

MÉTODO DO GRADIENTE: $d^k = -\nabla f(x^k)$ É UMA BOA

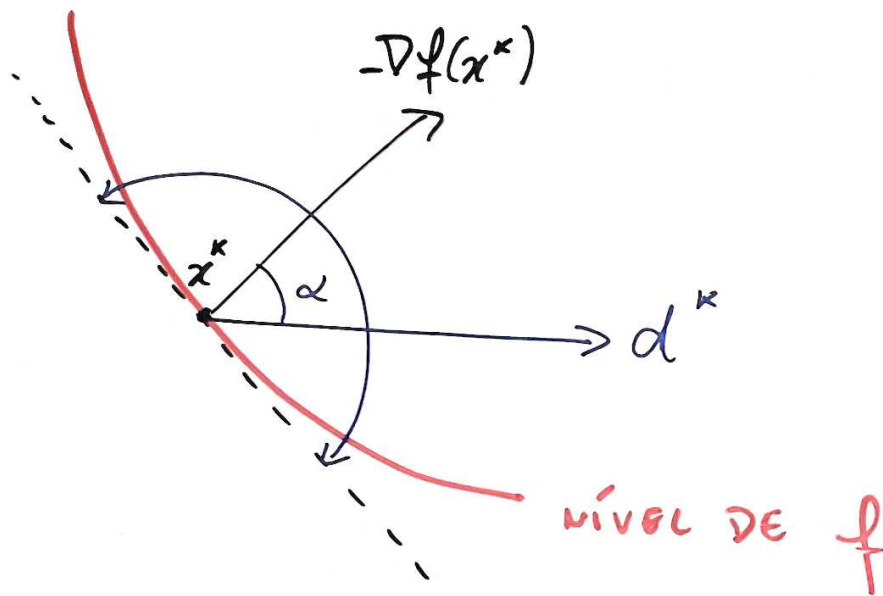
ESCOLHA SEMPRE ?



ZIG-ZAG (Ruim).

AJUSTANDO A DIREÇÃO:

• CONDIÇÃO DO ÂNGULO



Se $\alpha \approx 90^\circ$, a DIREÇÃO d^k NÃO DECRESCER MUITO f A PARTIR DE x^k . EXISTE QUE

$$\cos \alpha \geq \theta > 0, \quad \theta \in (0, 1) \text{ FIXADO.}$$

$$\Rightarrow \frac{-\nabla f(x^k)^T d^k}{\|\nabla f(x^k)\| \cdot \|d^k\|} \geq \Theta$$

$$\Rightarrow \boxed{\nabla f(x^k)^T d^k \leq -\Theta \cdot \|\nabla f(x^k)\| \cdot \|d^k\|} \quad \left(\text{CONDICÃO DO} \right. \\ \left. \text{ÂNGULO} \right)$$

ONDE $\Theta \in (0,1)$ É FIXO.

• CONDICÃO β : $\|d^k\| \geq \beta \|\nabla f(x^k)\|$, $\beta > 0$ FIXADO.

EVITA QUE LONGE DA SOLUÇÃO ($\|\nabla f(x^k)\| \gg 1$), TOMEMOS DIREÇÕES PEQUENAS.

ESSAS DUAS CONDIÇÕES SÃO SATISFEITAS POR $-\nabla f(x^*)$
POSSIVELMENTE MULTIPLICANDO $-\nabla f(x^*)$ POR β .

ESQUEMA DE DESCIDA.

DADOS $x^0 \in \mathbb{R}^m$, $\theta \in (0,1)$, $\beta > 0$, $\eta \in (0,1)$, $k=0$

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

Sim

PARA

NÃO

CALCULE d^k
SATISFAZENDO AS
CONDIÇÕES DO ÂNGULO
E β (P.EX. $d^k = -\nabla f(x^k)$
COM AJUSTE)

$t_k \leftarrow 1$

VALE ARMIZO ?

$f(x^k + t_k d^k) \leq f(x^k) + t_k \eta \nabla f(x^k)^T d^k$

NÃO

DIMINUA t_k ,
P.EX. $t_k \leftarrow \frac{t_k}{2}$

$x^{k+1} = x^k + t_k d^k$

$k \leftarrow k+1$

Sim

TEOREMA: O ESQUEMA ANTERIOR PARA COM ALGUM VALOR K
TAL QUE $\nabla f(x^k) = 0$ OU GERA UMA SEQUÊNCIA INFINITA
 $\{x^k\}$ TAL QUE QUALQUER PONTO DE ACUMULAÇÃO É UM
PONTO ESTACIONÁRIO DE f ($\nabla f(x^*) = 0$).

CONVERGÊNCIA É GLOBAL: O MÉTODO FUNCIONA
INICIANDO DE QUALQUER PONTO x^0 .

SOBRE A CONVERGÊNCIA GLOBAL DO MÉTODO DE DESCIDA:

* TAMANHO DO PASSO: AO INVÉS DE " $t_{\text{NOVO}} = \frac{t}{2}$ ", É SUFICIENTE EXIGIR $t_{\text{NOVO}} \in [0.1t, 0.9t]$ (SALVAGUARDA).

- $t_{\text{NOVO}} = \frac{t}{2}$ SATISFAZ A SALVAGUARDA.

- INTERPOLAÇÃO QUADRÁTICA PODE NÃO SATISFAZER...

SOLUÇÃO: FAÇA A INTERPOLAÇÃO, OBTENDO t^* .

SE $t^* \in [0.1t, 0.9t]$, $t_{\text{NOVO}} = t^*$. CASO CONTRÁRIO,

PROJETE t^* NO INTERVALO $[0.1t, 0.9t]$.

* SUPONHOS f LIMITADA INFERIORMENTE.

COM ISSO É POSSÍVEL PROVAR O TEOREMA DE CONVERGÊNCIA.