Programação
Não Linear Irrestrita

Modelo Programação Não Linear: PNL

$$\max\ (\min) \quad f(\mathrm{x})$$

$$\text{s. a.} \quad \mathrm{x} \in D \subset R^n$$

restrito

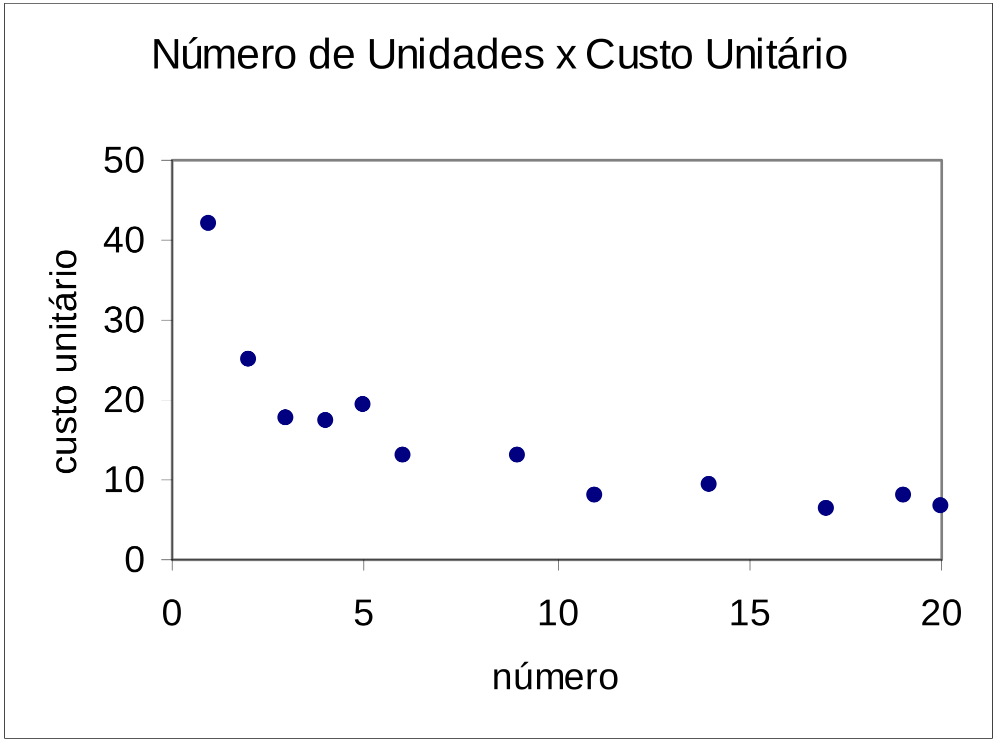
$$\max\ (\min) \quad f(\mathrm{x})$$

$$\text{s. a.} \quad \mathrm{x} \in D = R^n$$

irrestrito

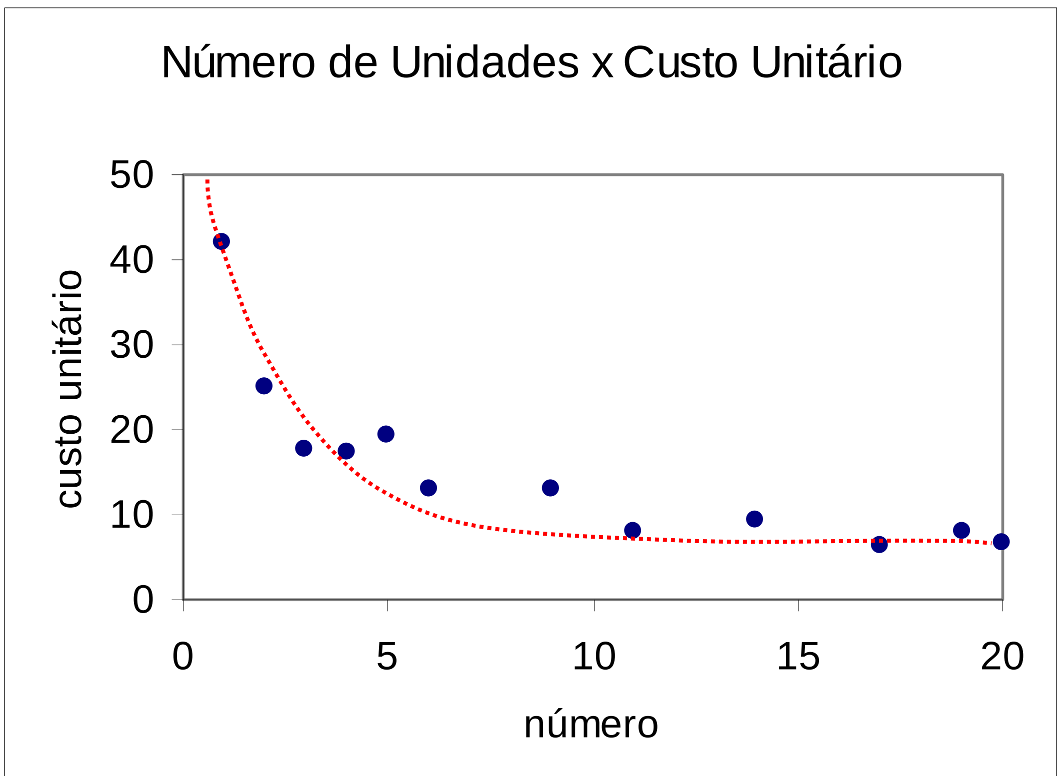
Exemplo: regressão não linear

	número	custo		número	custo		número	custo
i	p_i	q_i	i	p_i	q_i	i	p_i	q_i
1	19	7.9	5	5	19.5	9	14	9.2
2	2	25	6	6	13	10	17	6.3
3	9	13.1	7	3	17.8	11	1	42.0
4	4	17.4	8	11	8.0	12	20	6.6



$$q = r(p) = x_1(p)^{x_2}$$

$$\min \; f(x_1, x_2) = \sum_{i=1}^{m} [q_i - x_1(p_i)^{x_2}]^2$$



$$m = 12$$

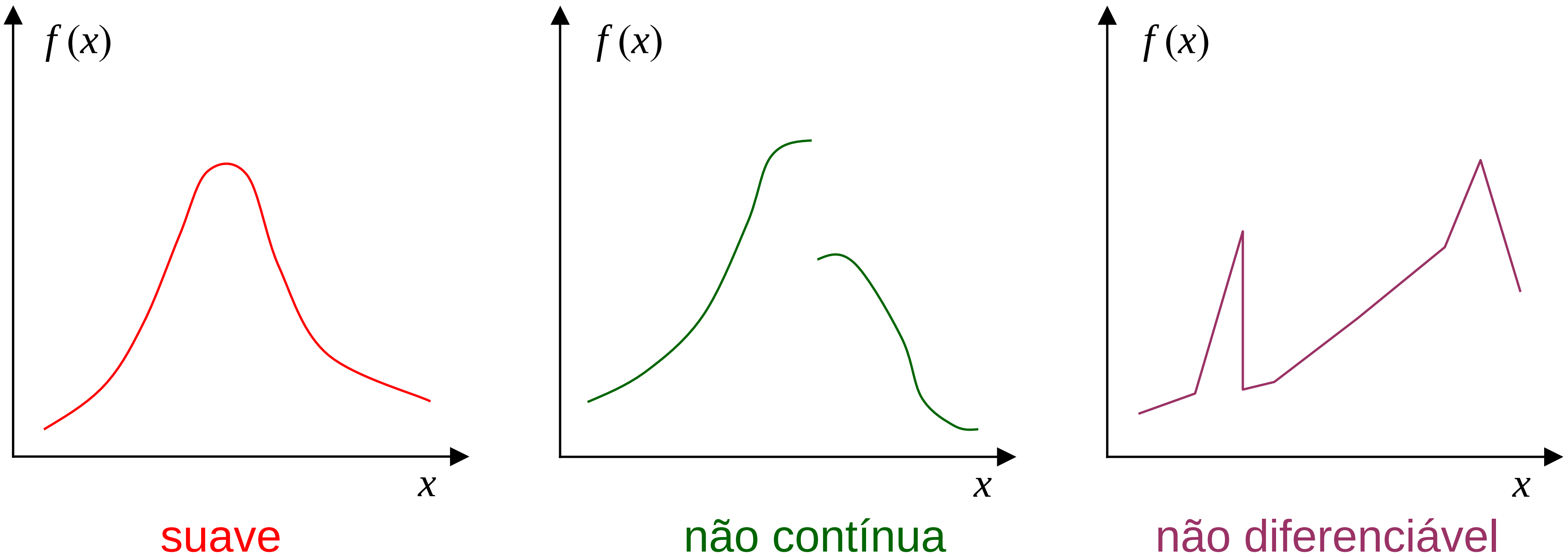
$$\downarrow$$

$$x_1^* = 40.69; \;\; x_2^* = -0.6024$$

$$\downarrow$$

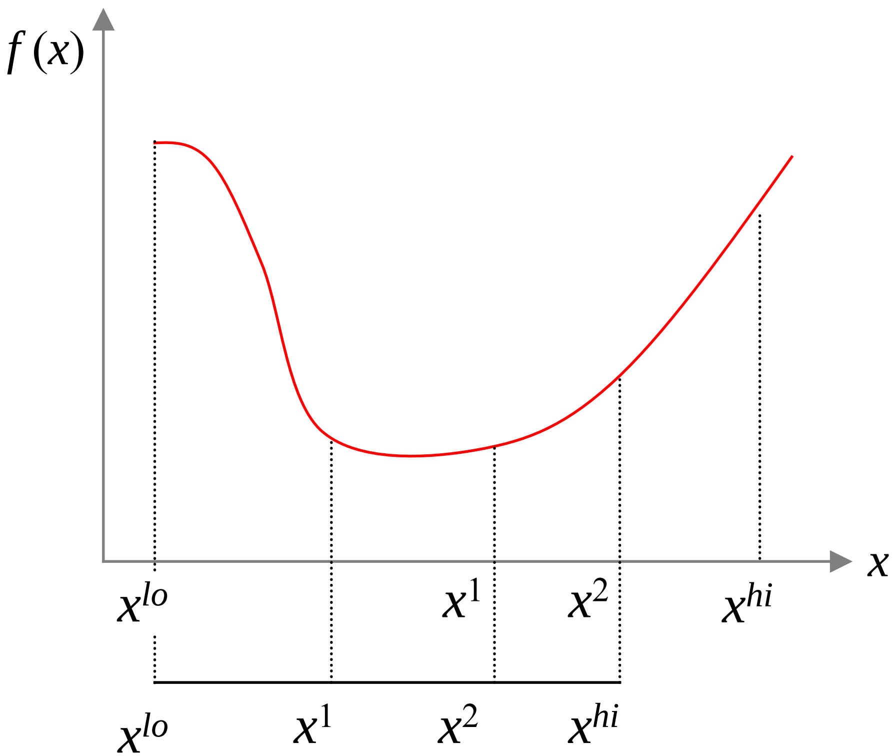
$$q = r(p) = 40.69\, p^{-0.6024}$$

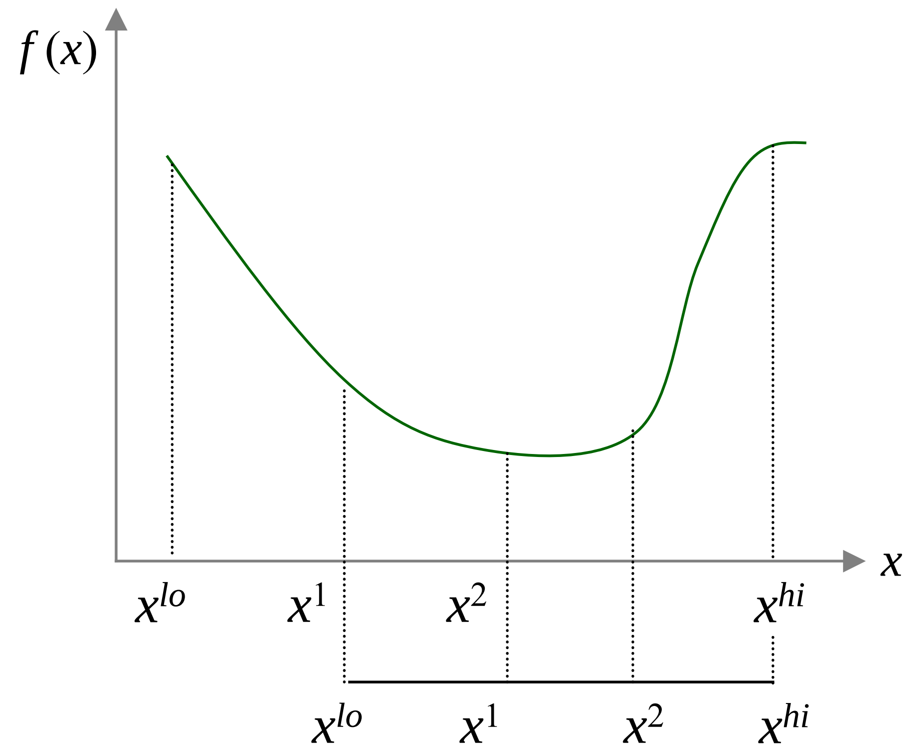
Funções Suaves e Derivadas



- função $f(x)$ suave: contínua e diferenciável no domímio de interesse $D \subseteq R^n$
- modelos de PNL com funções suaves são, geralmente, mais tratáveis
- funções suaves: possuem derivadas $\rightarrow$ possibilitam busca mais eficiente
- derivada : foma analítica pode ser difícil / impossível de ser obtida

Busca Unidimensional





$$x^1 = x^{hi} - \alpha(x^{hi} - x^{lo})$$

$$x^2 = x^{lo} + \alpha(x^{hi} - x^{lo})$$

$$\alpha = 0.618 \ \ (\text{número de ouro})$$

- $f(x)$ unimodal
- [x^{hi} , x^{lo}] contém x^*

Busca Unidimensional com Número de Ouro

Passo 0 Inicialização: escolher x^{lo} , x^{hi} e tolerância $\varepsilon > 0$. Computar ($\alpha = 0.618$)

$$x^1 \leftarrow x^{hi} - \alpha\,(x^{hi} - x^{lo})$$

$$x^2 \leftarrow x^{hi} + \alpha\,(x^{hi} - x^{lo})$$

calcular valor da função $f(x)$ para os quatro pontos; $t \leftarrow 0$;

Passo 1 Parada: se $(x^{hi} - x^{lo}) \leq \varepsilon$, parar: solução ótima aproximada $x^* = 1/2\,(x^{hi} - x^{lo})$;
senão ir para Passo 2 se $f(x^1)$ é superior a $f(x^2)$; caso contrário ir para o Passo 3;

Passo 2 Esquerdo: estreitar lado esquerdo do intervalo;

$$x^{hi} \leftarrow x^2 \;; \quad x^2 \leftarrow x^1$$

$$x^1 \leftarrow x^{hi} - \alpha\,(x^{hi} - x^{lo})$$

avaliar $f(x^1)$; $t \leftarrow t + 1$; ir para Passo 1;

Passo 3 Direito: estreitar lado direito do intervalo;

$$x^{lo} \leftarrow x^1 \;; \quad x^1 \leftarrow x^2$$

$$x^2 \leftarrow x^{lo} + \alpha\,(x^{hi} - x^{lo})$$

avaliar $f(x^2)$; $t \leftarrow t + 1$; ir para Passo 1;

Intervalo Inicial para Busca Unidimensional

- intervalo inicial deve ser tal que $x^* \in [\, x^{hi} ,\; x^{lo} \,] \rightarrow$ padrão 3 pontos
- padrão 3 pontos $\{x^{lo}, x^{mid}, x^{hi}\}$, $x^{lo} < x^{mid} < x^{hi}$, $f(x^{mid})$ melhor que $f(x^{hi})$ e $f(x^{lo})$
- padrão 3 pontos $\{x^{lo}, x^{mid}, x^{hi}\} \rightarrow x^* \in [\, x^{hi} ,\; x^{lo} \,]$

Passo 0 Inicializa: escolher limitante inferior x^{lo} para x^* e passo $\delta > 0$;

Passo 1 Esquerda ou Direita: se $f(x^{lo} + \delta)$ é superior à $f(x^{lo})$ então $x^{mid} \leftarrow x^{lo} + \delta$; ir para o Passo 2 para busca à direita; caso contrário ótimo está à esquerda; fazer $x^{hi} \leftarrow x^{lo} + \delta$; ir para Passo 3;

Passo 2 Expande: amentar $\delta \leftarrow 2\,\delta$; se $f(x^{mid})$ é superior à $f(x^{mid} + \delta)$ então $x^{hi} \leftarrow x^{mid} + \delta$ e parar; $\{x^{lo}, x^{mid}, x^{hi}\}$ fornece padrão 3 pontos; senão $x^{lo} \leftarrow x^{mid}$; $x^{mid} \leftarrow x^{mid} + \delta$; repetir Passo 2;

Passo 3 Reduz: diminuir $\delta \leftarrow \delta/2$; se $f(x^{lo} + \delta)$ é superior à $f(x^{lo})$ então $x^{mid} \leftarrow x^{lo} + \delta$ e parar; $\{x^{lo}, x^{mid}, x^{hi}\}$ fornece padrão 3 pontos; senão $x^{hi} \leftarrow x^{lo} + \delta$; repetir Passo 3;

Condições de Otimalidade

Vetor Gradiente

$$\mathrm{x} = (x_1, x_2, \cdots, x_n);\ f : R^n \to R \text{ diferenciável em x :}$$

$$\nabla f(\mathrm{x}) = (\partial f/\partial x_1 \ldots \partial f/\partial x_j \ldots \partial f/\partial x_n)$$

Matriz Hessiana

$$H(\mathrm{x}) = \begin{bmatrix} \dfrac{\partial^2 f}{\partial x_1^2} & \cdots & \dfrac{\partial^2 f}{\partial x_1 \partial x_n} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial^2 f}{\partial x_n \partial x_1} & \cdots & \dfrac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$$

$$x=(x_1,x_2); \quad f(x_1,x_2)=\sum_{i=1}^{m}[q_i - x_1(p_i)^{x_2}]^2$$

Exemplo
regressão não linear

$$\frac{\partial f}{\partial x_1} = -2\sum_{i=1}^{m}(q_i - x_1 p_i^{x_2})\,p_i^{x_2}$$

$$\frac{\partial f}{\partial x_2} = -2\sum_{i=1}^{m}(q_i - x_1 p_i^{x_2})x_1 p_i^{x_2}\,\ln(p_i)$$

$$\frac{\partial f}{\partial x_1^2} = 2\sum_{i=1}^{m} p_i^{x_2}$$

$$\frac{\partial f}{\partial x_2^2} = -2\sum_{i=1}^{m}\ln^2(p_i)[(q_i - x_1 p_i^{x_2})(x_1 p_i^{x_2}) - (x_1 p_i^{x_2})^2]$$

$$\frac{\partial f}{\partial x_1 \partial x_2} = \frac{\partial f}{\partial x_2 \partial x_1} = -2\sum_{i=1}^{m}[(q_i - x_1 p_i^{x_2})(p_i^{x_2})\ln(p_i) - (p_i^{x_2})(x_1 p_i^{x_2})\ln(p_i)]$$

$$\hat{\mathrm{x}} = (x_1, x_2) = (33, -0.5)$$

$$\nabla f(33,\ \text{-}\,0.5) = (\partial f/\partial x_1, \partial f/\partial x_2)_{\hat{\mathrm{x}}} = (-23.07,\ -174.23)$$

$$H(33,\ \text{-}\,0.5) = \begin{bmatrix} \dfrac{\partial^2 f}{\partial x_1^2} & \dfrac{\partial^2 f}{\partial x_1 \partial x_2} \\ \dfrac{\partial^2 f}{\partial x_2 \partial x_1} & \dfrac{\partial^2 f}{\partial x_2^2} \end{bmatrix}_{\hat{\mathrm{x}}} = \begin{bmatrix} 5.77 & 179.65 \\ 179.65 & 11003.12 \end{bmatrix}$$

Aproximação via séries de Taylor

$$f_1(\mathrm{x}^t + \lambda \Delta \mathrm{x}) = f(\mathrm{x}^t) + \lambda \nabla f(\mathrm{x}^t) \Delta \mathrm{x}$$

$$f_1(\mathrm{x}^t + \lambda \Delta \mathrm{x}) = f(\mathrm{x}^t) + \lambda \sum_{j=1}^{n} \left(\frac{\partial f}{\partial x_j} \right) \Delta x_j$$

1a ordem

$$f_2(\mathrm{x}^t + \lambda \Delta \mathrm{x}) = f(\mathrm{x}^t) + \lambda \nabla f(\mathrm{x}^t) \Delta \mathrm{x} + \frac{\lambda^2}{2} \Delta \mathrm{x} H(\mathrm{x}^t) \Delta \mathrm{x}$$

$$f_2(\mathrm{x}^t + \lambda \Delta \mathrm{x}) = f(\mathrm{x}^t) + \lambda \sum_{j=1}^{n} \left(\frac{\partial f}{\partial x_j} \right) \Delta x_j + \frac{\lambda^2}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \left(\frac{\partial f}{\partial x_i \partial x_j} \right) \Delta x_i \Delta x_j$$

2a ordem

Gradientes e ótimos locais

condição necessária de 1a ordem

x^t é um ponto estacionário de f se $\nabla f(x^t) = 0$

ótimo local de uma função objetivo suave é um ponto estacionário

$$f(x^t + \lambda \Delta x) = f(x^t) + \lambda \nabla f(x^t) \Delta x$$

$$\Delta x = \pm \nabla f(x^t) \qquad [+ \max, - \min]$$

$$f(x^t + \lambda \Delta x) = f(x^t) \pm \lambda \nabla f(x^t) \nabla f(x^t)$$

$$\Downarrow$$

função objetivo melhora a menos que $\nabla f(x^t) = 0$

Hessianas e ótimos locais

condições de 2a ordem

se x^t é um ponto estacionário de f então $\nabla f(x^t) = 0$; logo

$$f(x^t + \lambda \Delta x) \approx f(x^t) + \lambda \nabla f(x^t)\Delta x + \frac{\lambda^2}{2}\Delta x H(x^t)\Delta x$$

$$\approx f(x^t) + \quad 0 \quad + \frac{\lambda^2}{2}\Delta x H(x^t)\Delta x$$

Δx direção de melhora em x^t

$$\Downarrow$$

necessária

x^t mínimo local de f $\rightarrow$ $H(x^t)$ semi-positiva definida

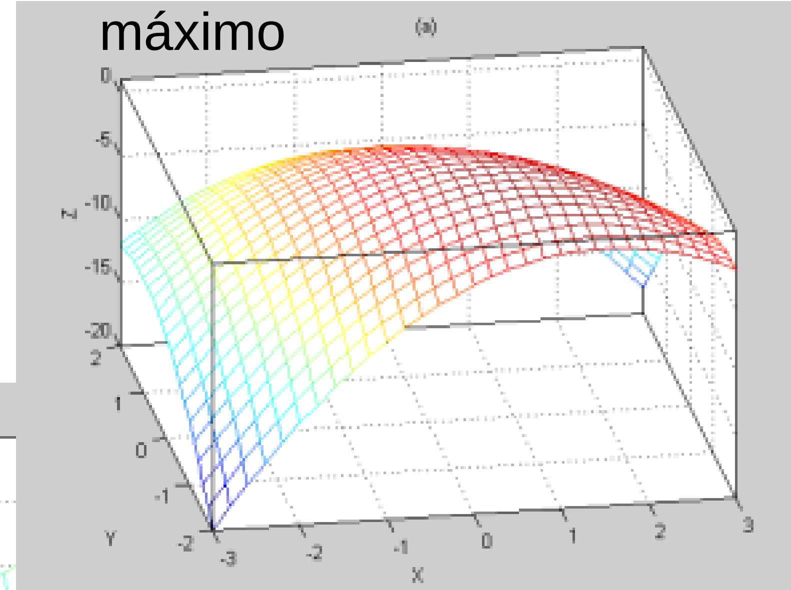
x^t máximo local de f $\rightarrow$ $H(x^t)$ semi-negativa definida

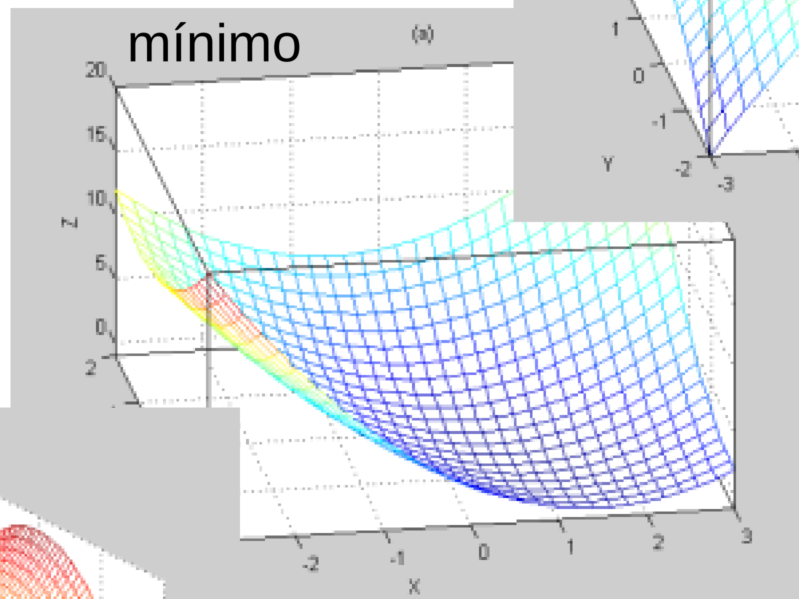
suficiente

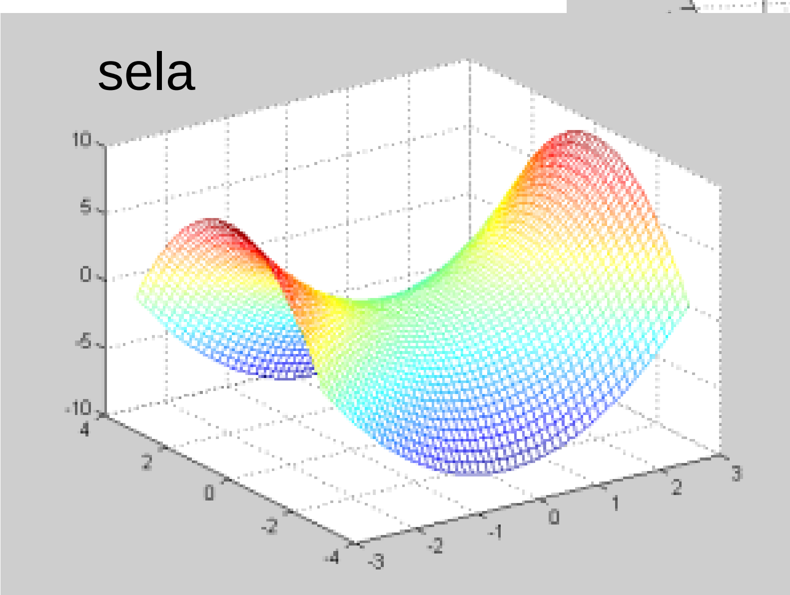
x^t ponto estacionário de f e $H(x^t)$ positiva definida $\rightarrow$ x^t mínimo local de f

x^t ponto estacionário de f e $H(x^t)$ negativa definida $\rightarrow$ x^t máximo local de f

pontos estacionários
exemplos

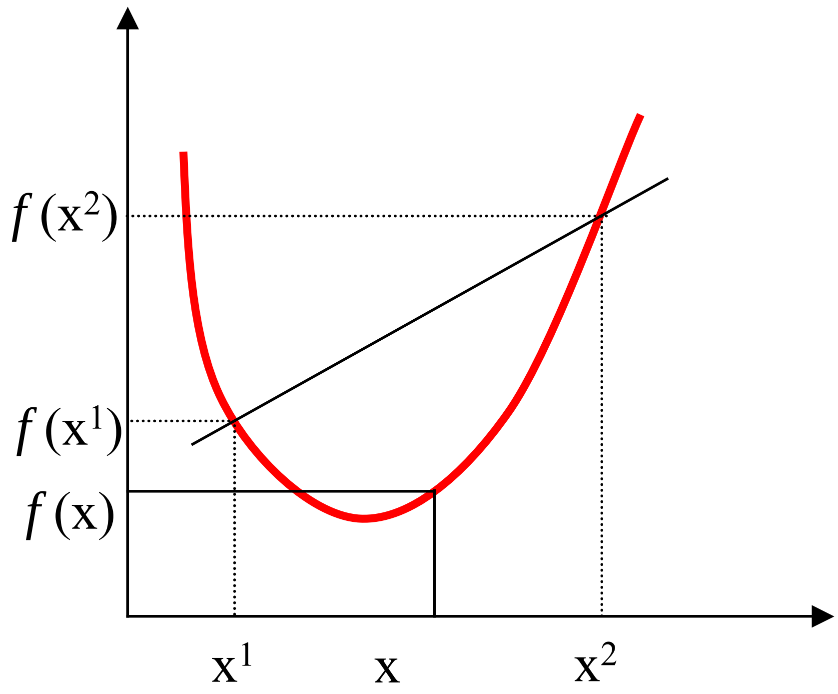


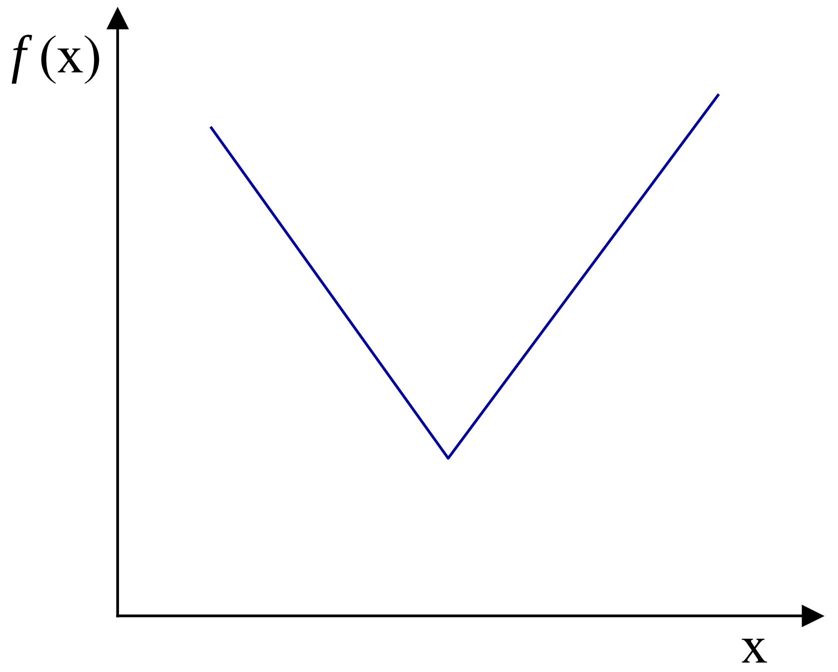




Convexidade e Otimalidade Global

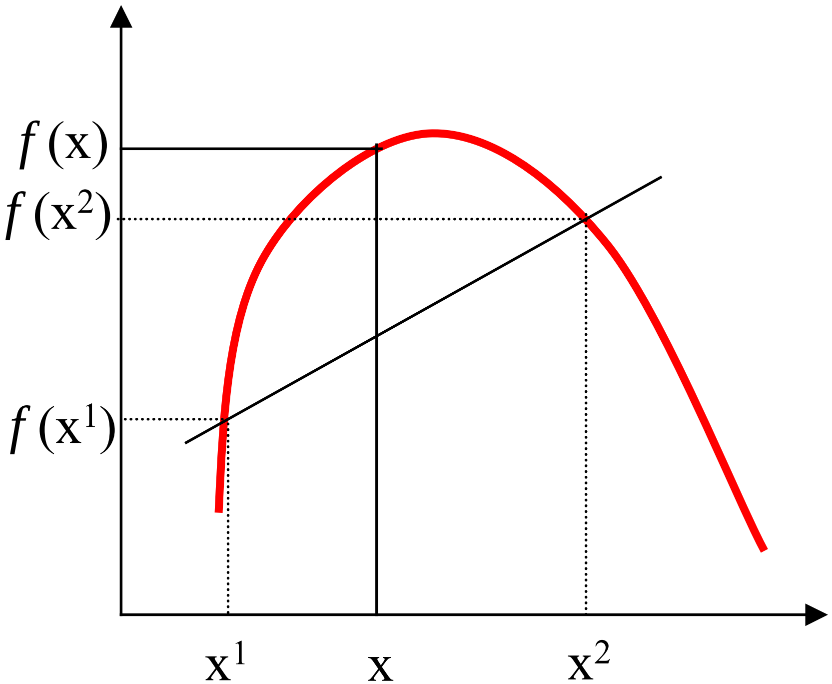
funções convexas

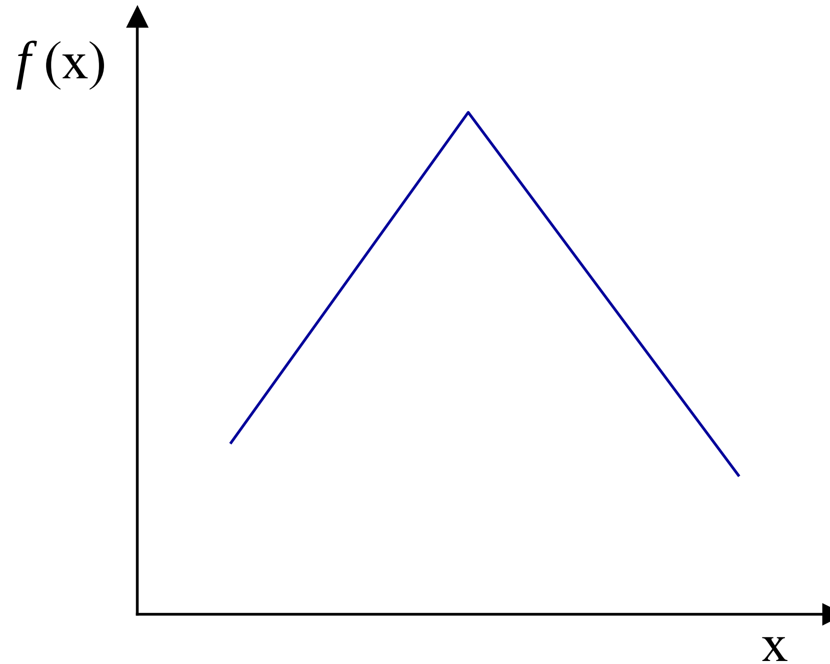




$$f(x^1+\lambda(x^2-x^1)\le f(x^1)+\lambda[f(x^2)-f(x^1)];\quad x^1,x^2\in D;\quad \lambda\in[0,1]$$

funções côncavas





$$f(x^1 + \lambda(x^2 - x^1)) \geq f(x^1) + \lambda[f(x^2) - f(x^1)]; \quad x^1, x^2 \in D; \quad \lambda \in [0,1]$$

Funções Convexas e Côncavas

1- se $f(\mathrm{x})$ é convexa então $-f(\mathrm{x})$ é côncava

2- $f(\mathrm{x})$ com segundas derivadas continuas é convexa se e somente se a matriz Hessiana $\mathrm{H}(\mathrm{x})$ é semi-positiva definida em um domínio convexo (aberto)'; $f(\mathrm{x})$ é côncava se e somente se $\mathrm{H}(\mathrm{x})$ é semi-negativa definida.

3- funções lineares são convexas e côncavas

4- se $f(\mathrm{x})$ é côncava, $g(\mathrm{x}) = 1/f(\mathrm{x})$ é convexa $\forall$ x | $f(\mathrm{x}) > 0$
se $f(\mathrm{x})$ é convexa, $g(\mathrm{x}) = 1/f(\mathrm{x})$ é c^ncava $\forall$ x | $f(\mathrm{x}) < 0$

5- se $g(y)$ é uma função convexa não decrescente e $h(\mathrm{x})$ é convexa, então $f(\mathrm{x}) = g(h(\mathrm{x}))$ é convexa; se $g(y)$ é uma função côncava não decrescente e $h(\mathrm{x})$ é côncava, então $f(\mathrm{x}) = g(h(\mathrm{x}))$ é côncava

6- $f(\mathrm{x})$ é convexa se, para $\alpha_{\mathrm{i}} \geq 0$ e $g_i(\mathrm{x})$ convexa, $i = 1, .., k$

$$f(\mathrm{x}) = \sum_{i=1}^{k} \alpha_i g_i(\mathrm{x})$$

7- $f(\mathrm{x})$ formada a partir de máximos de funções convexas é convexa
$f(\mathrm{x})$ formada a partir do mínimo de funções côncavas é côncava

$$f(\mathrm{x}) = \max_{\mathrm{x}} \{g_i(\mathrm{x}); i = 1, \cdots, k\}$$

$$f(\mathrm{x}) = \min_{\mathrm{x}} \{g_i(\mathrm{x}); i = 1, \cdots, k\}$$

8- funções convexas (côncavas) são unimodais (o contrário não)

Otimalidade Global: Condições Suficientes

- se $f\,(\mathrm{x})$ é uma função convexa, então todo mínimo local é mínimo global
- se $f\,(\mathrm{x})$ é uma função côncava, então todo máximo local é máximo global

por exemplo, no caso de mínimo: seja x^* mínimo global e $\mathrm{x}^1 \neq \mathrm{x}^*$

$$f(\mathrm{x}^*) < f(\mathrm{x}^1) \;\Rightarrow\; \lambda[f(\mathrm{x}^*) - f(\mathrm{x}^1)] < 0; \quad \forall \lambda > 0$$

$$f[\mathrm{x}^1 + \lambda(\mathrm{x}^* - \mathrm{x}^1)] \leq f(\mathrm{x}^1) + \lambda[f(\mathrm{x}^*) - f(\mathrm{x}^1)] < f(\mathrm{x}^1); \quad \lambda \in [0,1]$$

$\Delta\mathrm{x} = (\mathrm{x}^1 - \mathrm{x}^*)$ direção melhora em $\mathrm{x}^1, \forall\, \mathrm{x}^1 \in$ D

$$\Downarrow$$

x^* ótimo global

• todo ponto estacionário de uma função convexa suave é um mínimo global

• todo ponto estacionário de uma função côncava suave é um máximo global

por exemplo, se $f(\mathrm{x})$ é convexa, então:

$$f[\mathrm{x}^* + \lambda(\mathrm{x} - \mathrm{x}^*)] \leq f(\mathrm{x}^*) + \lambda[f(\mathrm{x}) - f(\mathrm{x}^*)] \qquad \lambda \in (0,1] \qquad \text{definição convexidade}$$

$$f[\mathrm{x}^* + \lambda(\mathrm{x} - \mathrm{x}^*)] \approx f(\mathrm{x}^*) + \lambda \nabla f(\mathrm{x})(\mathrm{x} - \mathrm{x}^*) \qquad \text{Taylor}$$

$$f(\mathrm{x}) - f(\mathrm{x}^*) \geq \nabla f(\mathrm{x}^*)(\mathrm{x} - \mathrm{x}^*)$$

$$\nabla f(\mathrm{x}^*) = 0 \quad \Rightarrow \quad f(\mathrm{x}) - f(\mathrm{x}^*) \geq 0 \quad \forall \mathrm{x}$$

$$\downarrow$$

x* é mínimo global

Algoritmo (Busca) do Gradiente

Passo 0 Inicialização: com solução inicial x^0; tolerância $\varepsilon > 0$; $t \leftarrow 0$;

Passo 1 Gradiente: calcular $\nabla f\,(x^t)$ em x^t;

Passo 2 Ponto Estacionário: se $\|\nabla f\,(x^t)\| \leq \varepsilon$ então parar; x^t é ótimo;

Passo 3 Direção: $\Delta x^{t+1} \leftarrow \pm\nabla f\,(x^t)$ [+ para max, - para min];

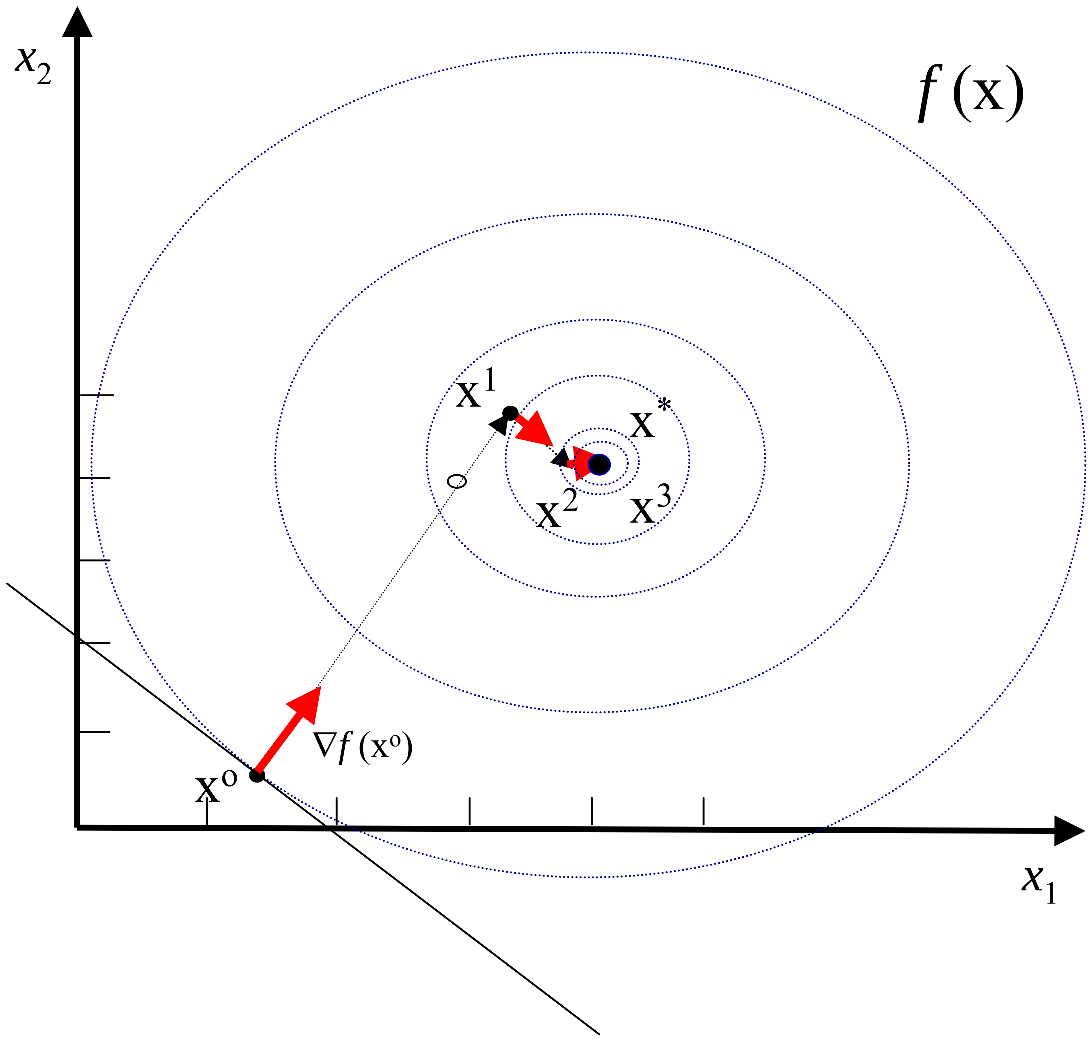
Passo 4 Busca Unidimensional: determinar λ_{t+1} resolvendo

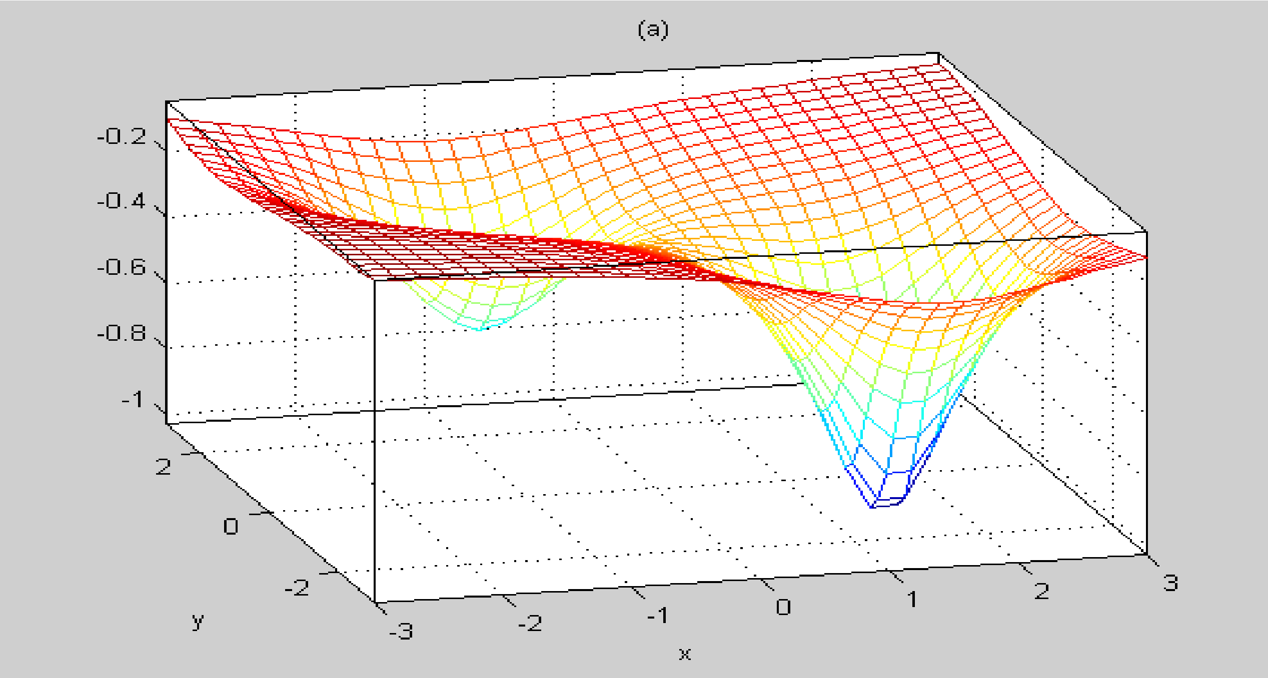
$$\max\ (\min)\ f\,(x^t + \lambda\, \Delta x^{t+1})\ ;$$

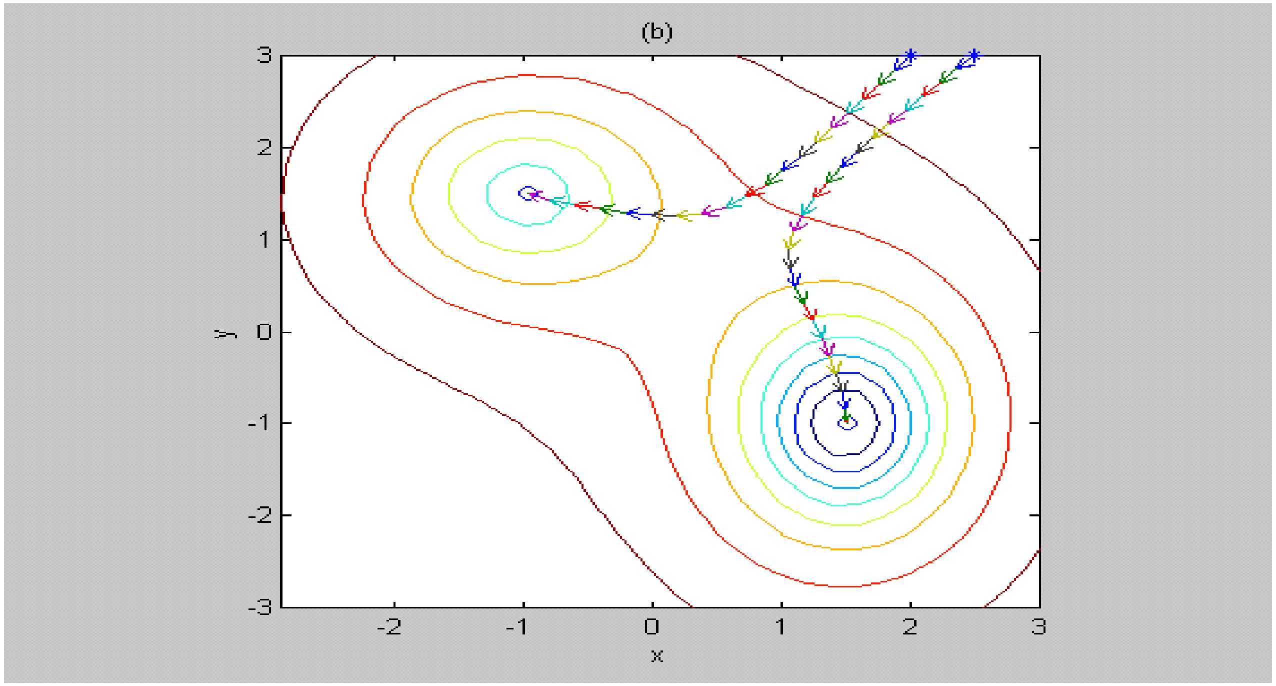
Passo 5 Atualizar: $x^{t+1} = x^t + \lambda\, \Delta x^{t+1}$;

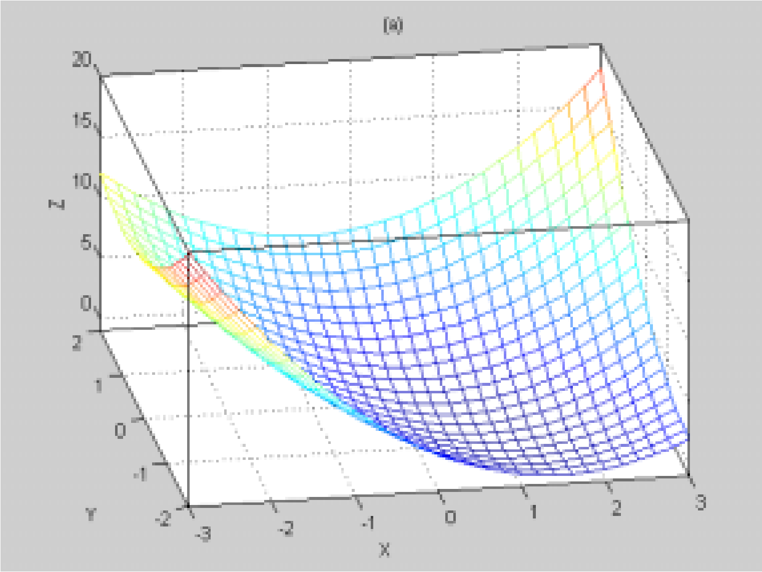
Passo 6 Incrementa: $t = t + 1$; ir para Passo 1;

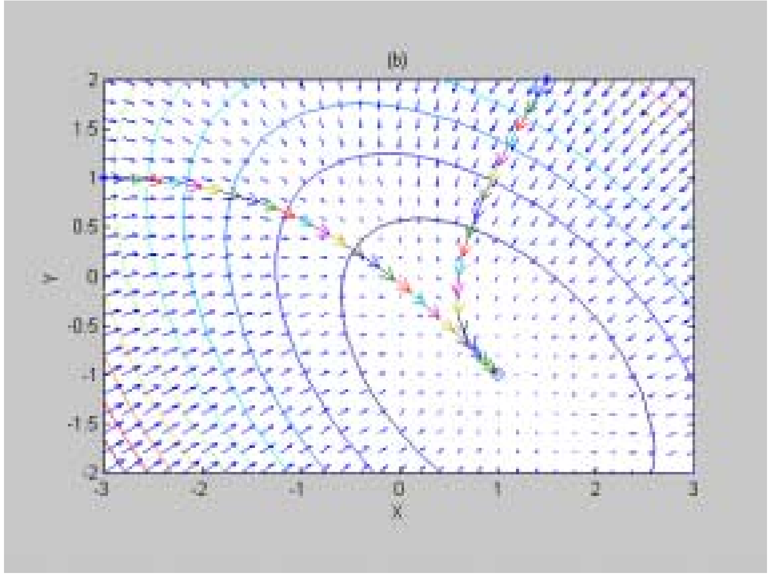
Algoritmo do Gradiente

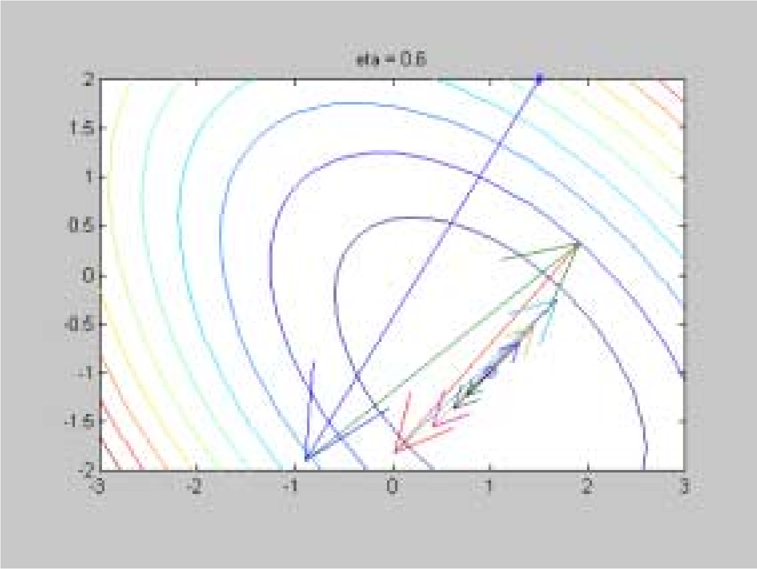


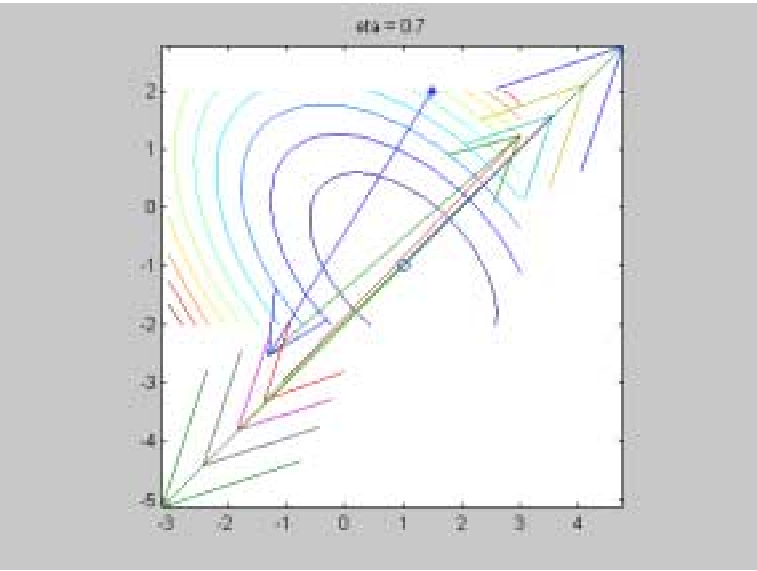

(a)
-0.2
-0.4
-0.6
-0.8
-1
2
0
-2
y
-3
-2
-1
0
1
2
3
x


(b)
3
2
1
0
-1
-2
-3
y
-2
-1
0
1
2
3
x


(a)
Z
X
Y


(b)
X
Y


eta = 0.6


eta = 0.7

Método de Newton

Utiliza informação de segunda ordem

$$f_2(\mathrm{x}^t + \lambda\Delta\,\mathrm{x}) = f(\mathrm{x}^t) + \lambda\nabla f(\mathrm{x}^t)\Delta\mathrm{x} + \frac{\lambda^2}{2}\Delta\mathrm{x}H(\mathrm{x}^t)\Delta\mathrm{x}$$

$$f_2(\mathrm{x}^t + \lambda\Delta\,\mathrm{x}) = f(\mathrm{x}^t) + \lambda\sum_{j=1}^{n}\left(\frac{\partial f}{\partial x_j}\right)\Delta x_j + \frac{\lambda^2}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(\frac{\partial^2 f}{\partial x_i \partial x_j}\right)\Delta x_i \Delta x_j$$

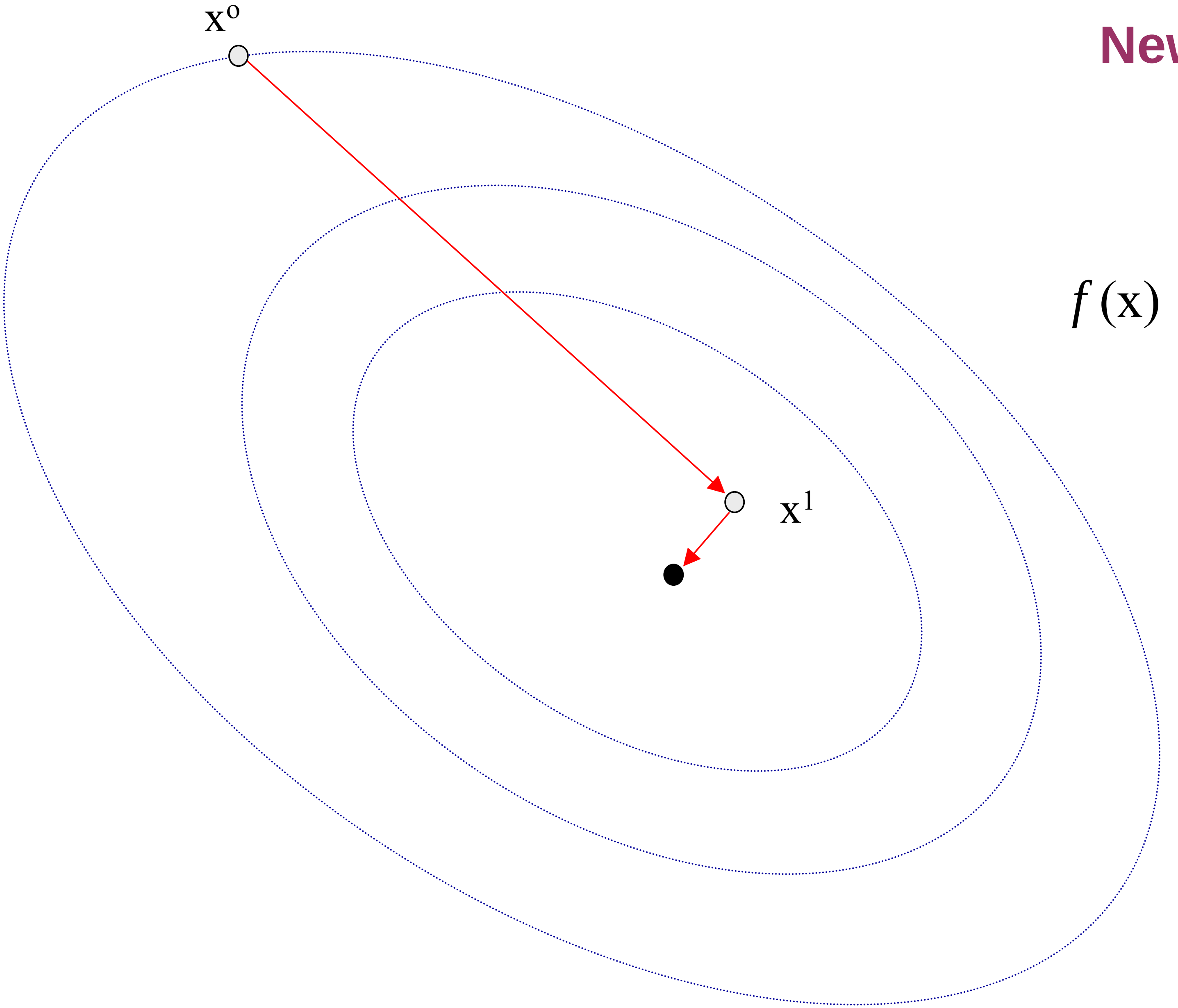
fazendo λ = 1 e derivando com relação à Δx_i

$$\frac{\partial f_2}{\partial \Delta x_i} = \frac{\partial f}{\partial x_i} + \sum_{j=1}^{n}\frac{\partial^2 f}{\partial x_i \partial x_j}\Delta x_j \ , \ \ i = 1, \cdots, n$$

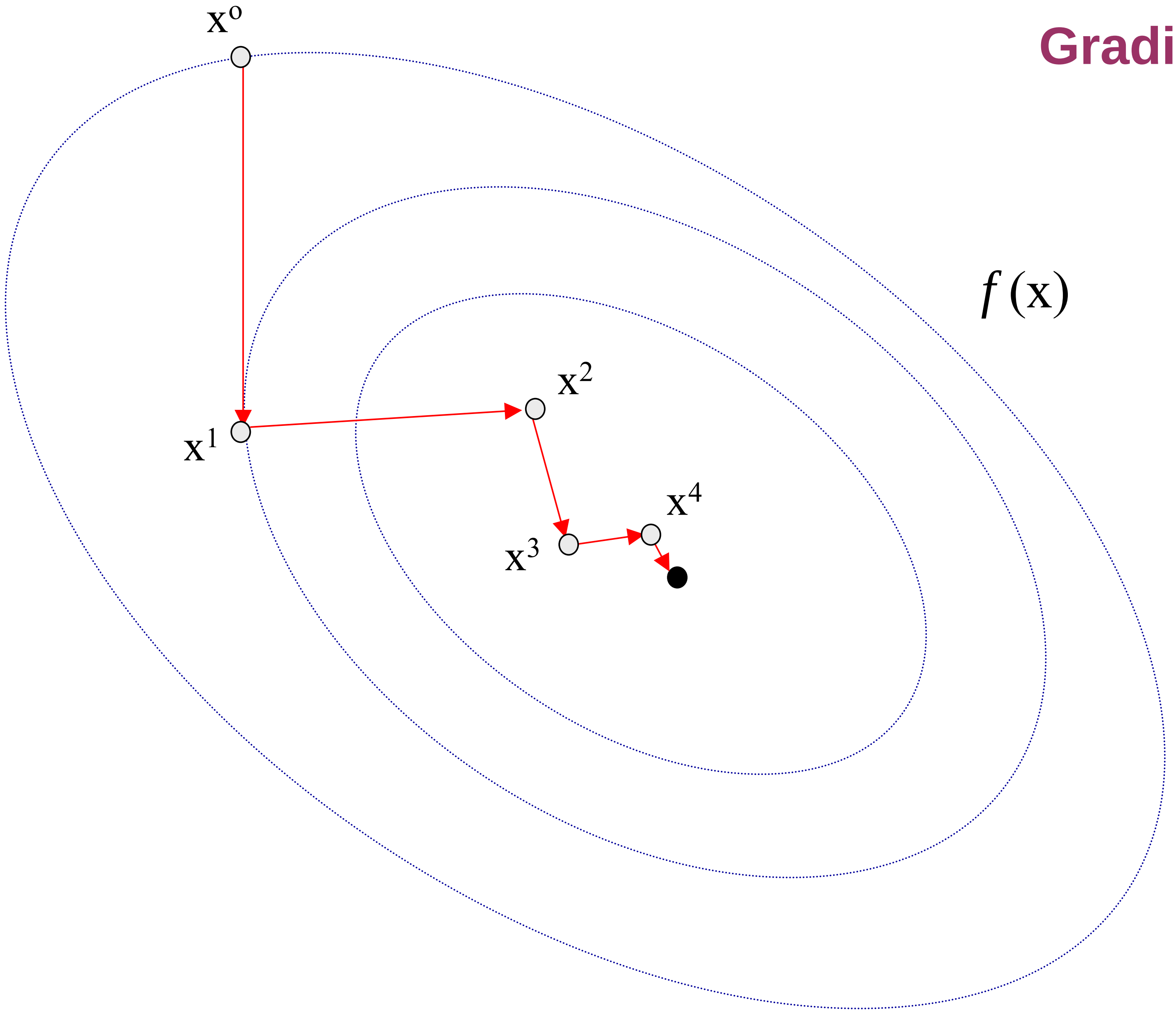
$$\nabla f_2(\Delta\,\mathrm{x}) = \nabla f(\mathrm{x}^t) + H(\mathrm{x}^t)\Delta\,\mathrm{x}$$

$$\nabla f_2(\Delta\,\mathrm{x}) = 0 \ \rightarrow \ H(\mathrm{x}^t)\Delta\,\mathrm{x} = -\nabla f(\mathrm{x}^t)$$

Newton



Gradiente



- algoritmo Newtoniano converge para ótimo local se inicialização é suficientemente próxima do ótimo local

- não há garantia de que a matriz Hessiana seja não singular em todo domínio de interesse

- idéia: combinar gradiente + Newton

$$\downarrow$$

- métodos Quase Newtonianos: $\Delta x = - D_t \nabla f(x_t)$

- matriz D aproxima da inversa da Hessiana H^{-1} ao longo da busca

- Hessiana: relacionada com a variação do gradiente:

$$\downarrow$$

$$\nabla f(x^{t+1}) - \nabla f(x^t) \approx H(x^t)[x^{t+1} - x^t]$$

Broyden, Fletcher, Goldfarb, Shanno: Método BFGS

$$\Delta \mathrm{x}^{t+1} \leftarrow -D_t \nabla f(\mathrm{x}^t)$$

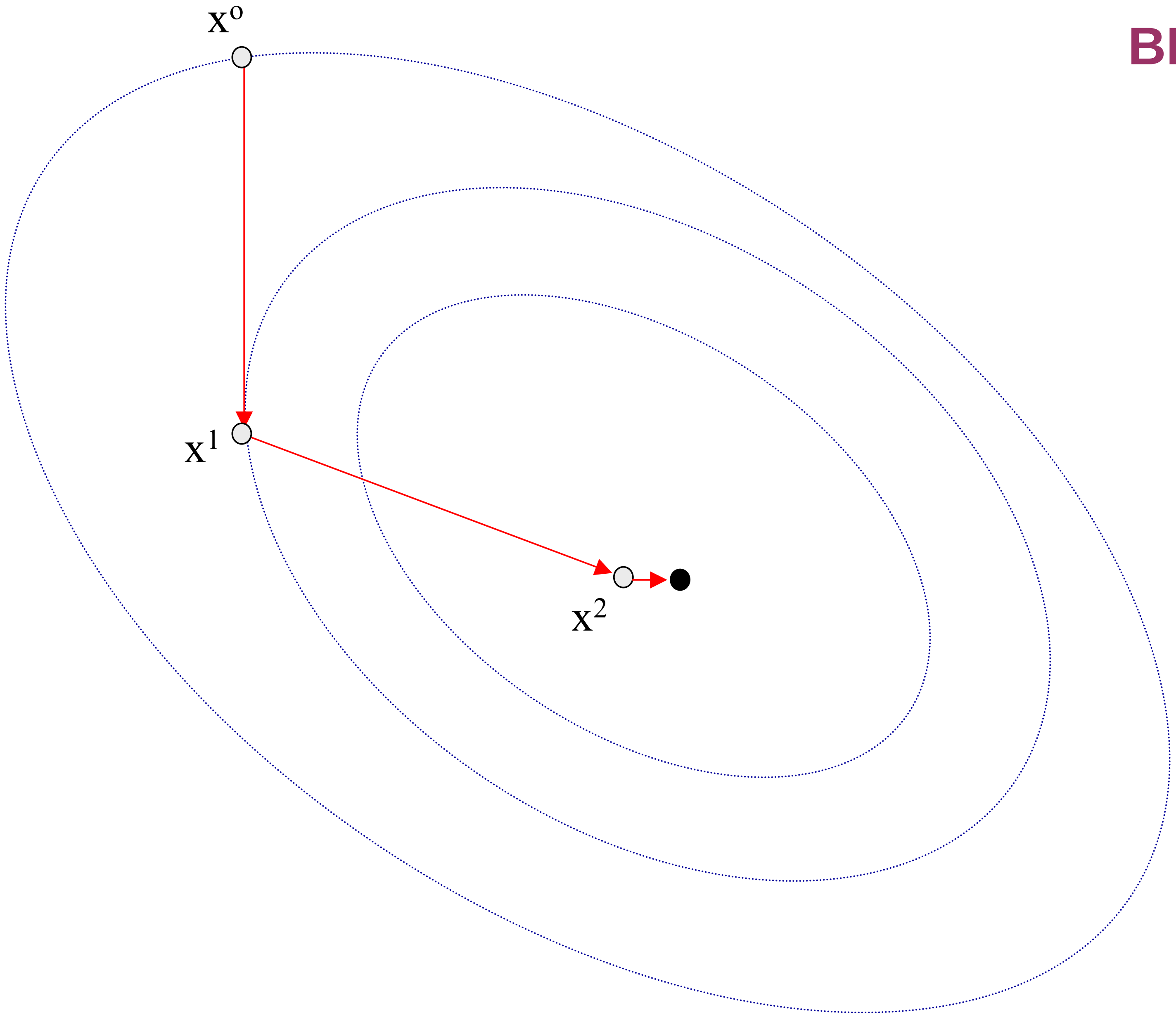
$$\mathrm{x}^{t+1} \leftarrow \mathrm{x}^t + \lambda_{t+1} \Delta \mathrm{x}^{t+1}$$

$$D_{t+1} \leftarrow D_t + \left(1 + \frac{\mathrm{g}^T D_t \mathrm{g}}{\mathrm{d}^T \mathrm{g}}\right) \frac{\mathrm{d}\mathrm{d}^T}{\mathrm{d}^T \mathrm{g}} - \frac{D_t \mathrm{g}\mathrm{d}^T + \mathrm{d}\mathrm{g}^T D_t}{\mathrm{d}^T \mathrm{g}}$$

$$\mathrm{d} = \mathrm{x}^{t+1} - \mathrm{x}^t \qquad \mathrm{g} = \nabla f(\mathrm{x}^{t+1}) - \nabla f(\mathrm{x}^t) \qquad D_0 = \pm I$$

$(-\ \text{min},\ +\ \text{max})$

BFGS



Algoritmo de Nelder-Mead

- não utiliza derivadas

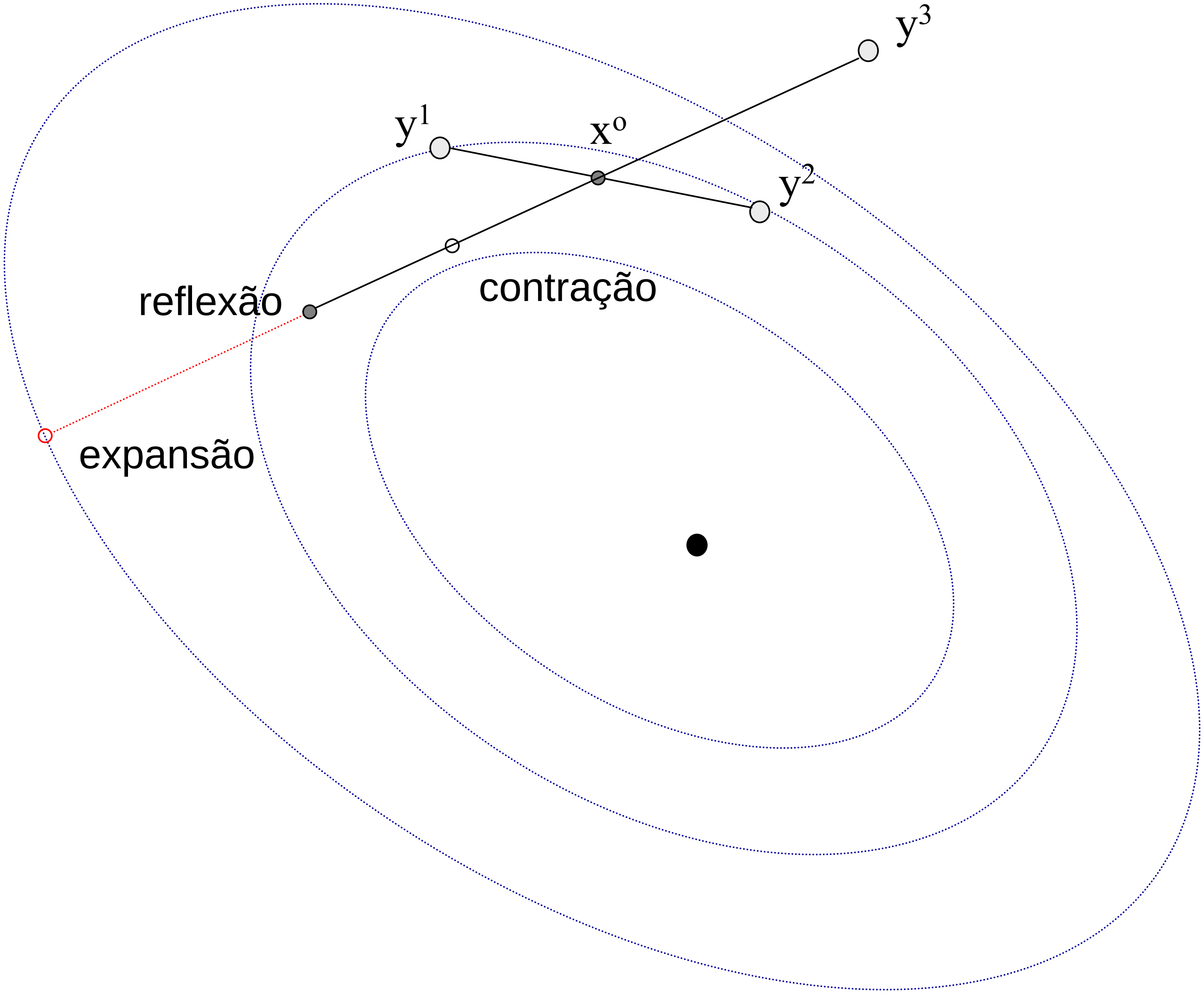
- mantém ($n+1$) soluções candidatas

- baseia-se nos conceitos de:

  - reflexão
  - expansão
  - contração
  - encolhimento

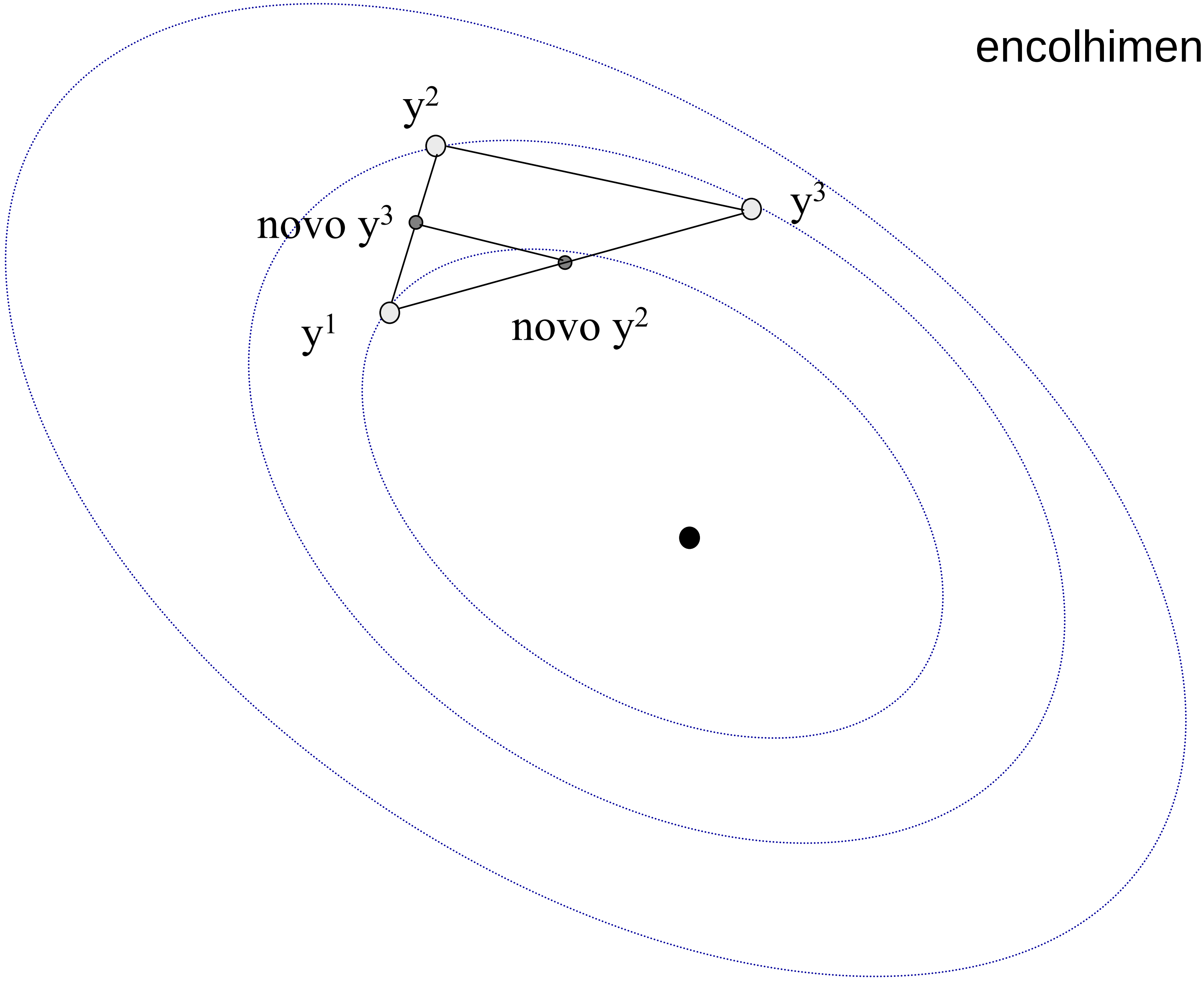
$\Delta x \equiv x^t - y^{n+1}$ $\qquad$ $y^{n+1} \leftarrow$ pior solução entre as $(n+1)$ candidatas

$x^t \equiv \frac{1}{n} \sum_{i=1}^{n} y^i$ $\qquad$ $n-$ésimo centróide



y3
y1
xo
y2
reflexão
contração
expansão

encolhimento



Observação

Este material refere-se às notas de aula do curso EA 044 Planejamento e Análise de Sistemas de Produção da Faculdade de Engenharia Elétrica e de Computação da Unicamp. Não substitui o livro texto, as referências recomendadas e nem as aulas expositivas. Este material não pode ser reproduzido sem autorização prévia dos autores. Quando autorizado, seu uso é exclusivo para atividades de ensino e pesquisa em instituições sem fins lucrativos.