EA 044 Planejamento e Análise
de Sistemas de Produção

Programação
Não Linear Irrestrita

©DCA-FEEC-Unicamp

Tópicos

1-Introdução

2-Busca unidimensional

3-Condições de otimalidade

4-Convexidade e otimalidade global

5-Algoritmos

©DCA-FEEC-Unicamp

1-Introdução

- Modelos de programação não linear (PNL)

$$\begin{aligned} &\max\ (\min) && f(x) \\ &\text{sa} && x \in D \subset R^n \end{aligned}$$

restrito

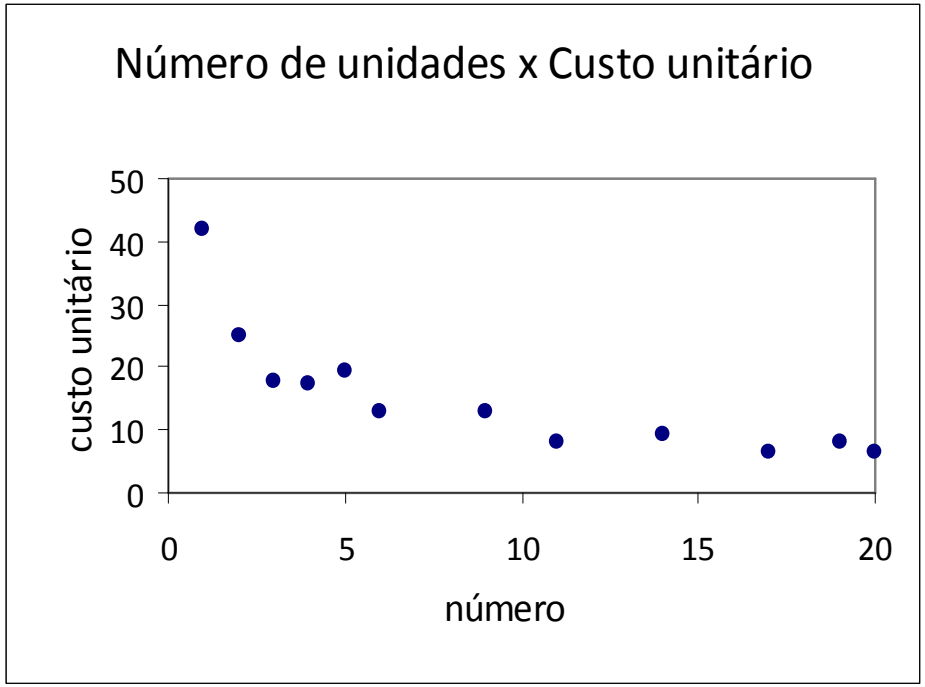
$$\begin{aligned} &\max\ (\min) && f(x) \\ &\text{sa} && x \in D = R^n \end{aligned}$$

irrestrito

©DCA-FEEC-Unicamp

Exemplo: regressão não linear

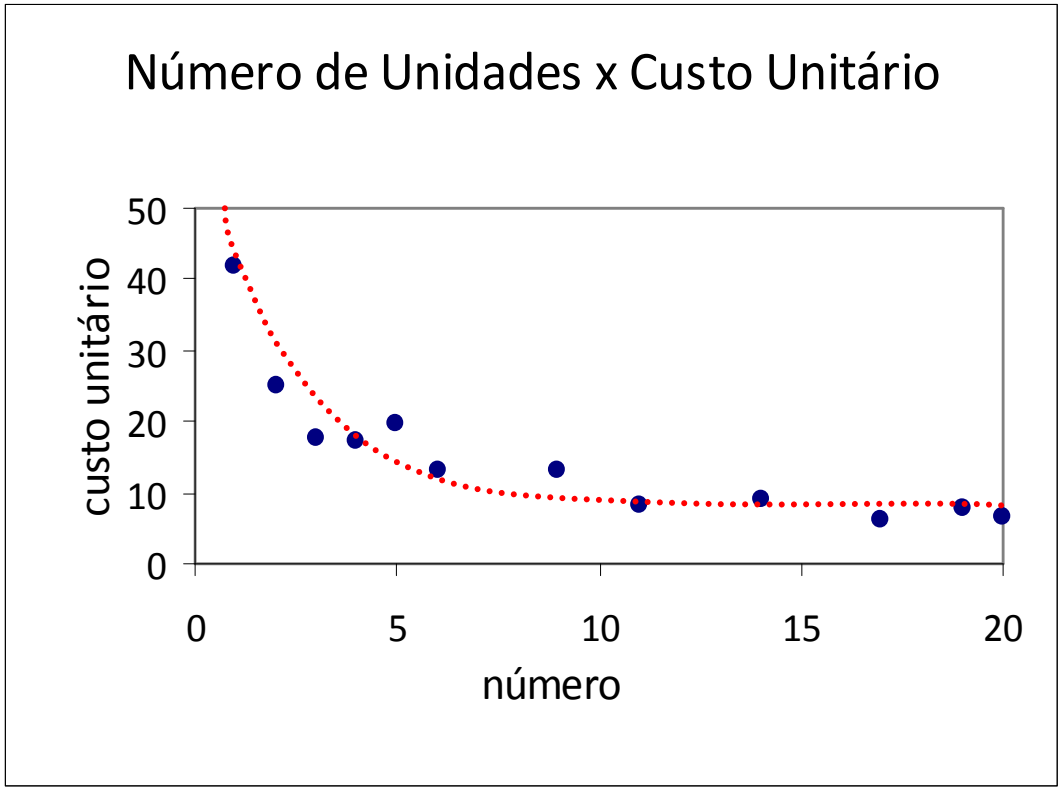
	número	custo		número	custo		número	custo
i	p_i	q_i	i	p_i	q_i	i	p_i	q_i
1	19	7.9	5	5	19.5	9	14	9.2
2	2	25	6	6	13	10	17	6.3
3	9	13.1	7	3	17.8	11	1	42.0
4	4	17.4	8	11	8.0	12	20	6.6



©DCA-FEEC-Unicamp

Hipótese: $q = r(p) = x_1(p)^{x_2}$

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



$$m = 12$$

$$\downarrow$$

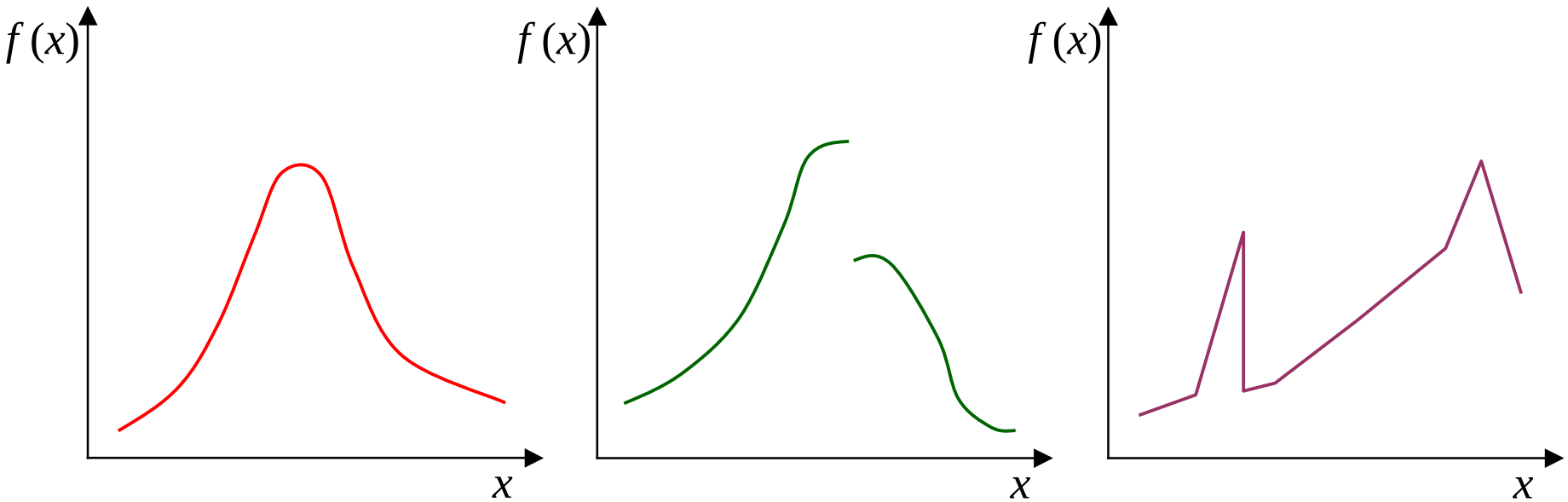
$$x_1^* = 40.69; \; x_2^* = -0.6024$$

$$\downarrow$$

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

©DCA-FEEC-Unicamp

Funções suaves e derivadas



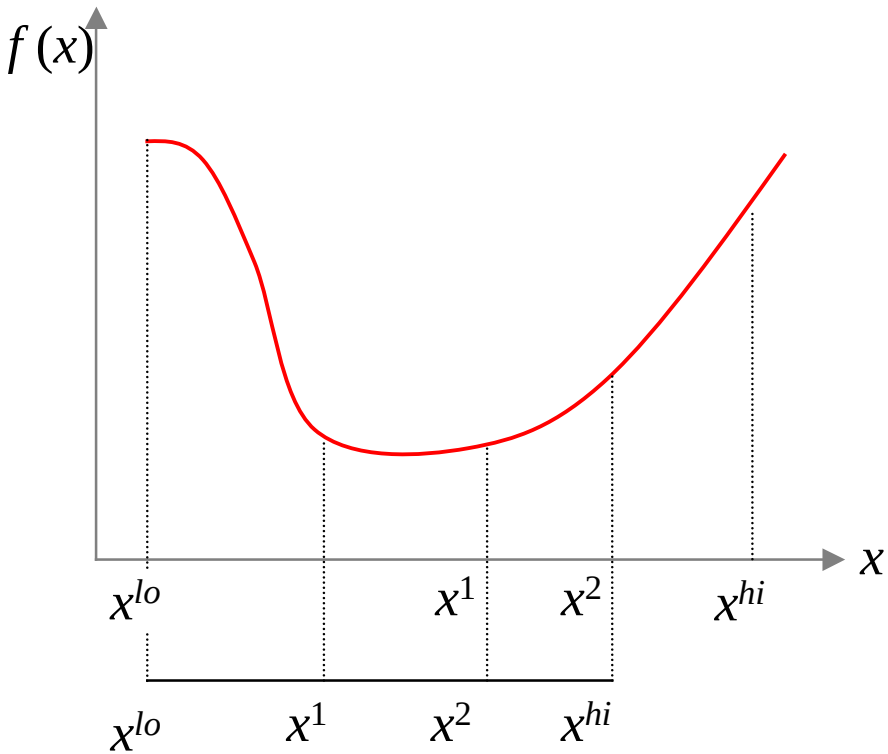
suave | não contínua | não diferenciável

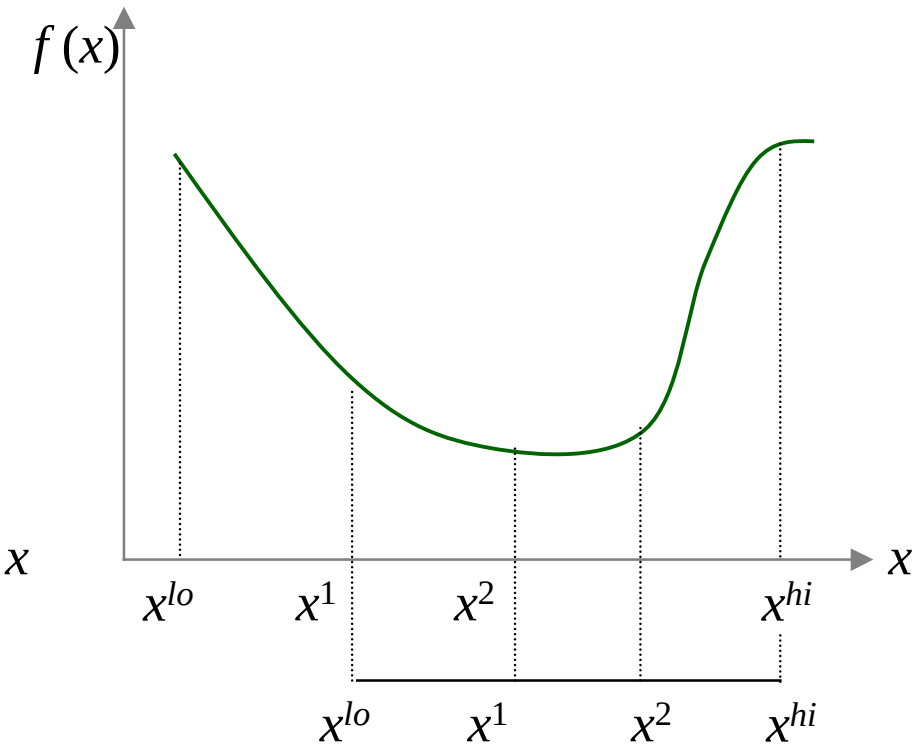
©DCA-FEEC-Unicamp

- $f(x)$ suave = contínua e diferenciável no domímio $\mathrm{D} \subseteq \mathrm{R}^n$
- Modelos PNL com funções suaves são mais tratáveis
- Funções suaves: possuem derivadas $\rightarrow$ busca mais eficiente
- Derivada: forma analítica pode ser difícil/impossível de ser obtida

©DCA-FEEC-Unicamp

2-Busca unidimensional





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

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

$$\alpha = 0.618$$

$f(x)$ unimodal

$[x^{hi},\ x^{lo}]$ contém x^*

número de ouro

©DCA-FEEC-Unicamp

Busca unidimensional com número de ouro

Passo 0 Inicialização: escolher x^{lo}, x^{hi} e tolerância $\varepsilon > 0$; $\alpha = 0.618$; determinar

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

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

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;

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

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

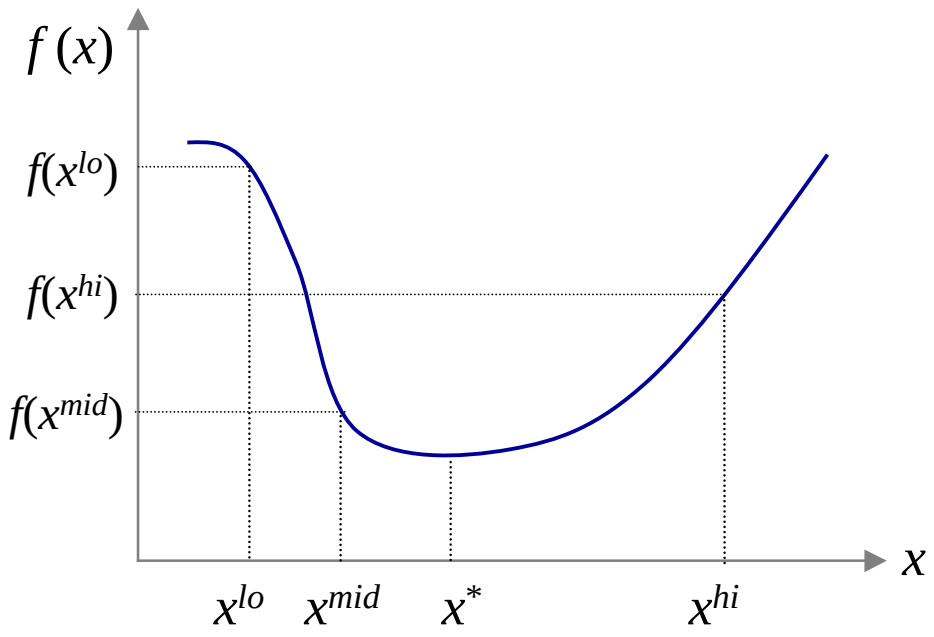
avaliar $f(x^2)$; $t \leftarrow t + 1$; ir para Passo 1;

©DCA-FEEC-Unicamp

Intervalo inicial para busca unidimensional

- Intervalo inicial deve ser tal que $x^* \in [\, x^{lo}, x^{hi} \,] \rightarrow$ 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 $f(x^{lo})$

$$\{x^{lo}, x^{mid}, x^{hi}\} \rightarrow x^* \in [\, x^{hi}, x^{lo} \,]$$



©DCA-FEEC-Unicamp

Algoritmo padrão três pontos

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;

©DCA-FEEC-Unicamp

3-Condições de otimalidade

- Vetor gradiente

$\mathrm{x} = (x_1, x_2, \ldots, x_n)$

f: $\mathrm{R}^n \rightarrow \mathrm{R}$ 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

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

©DCA-FEEC-Unicamp

Exemplo: regressão não linear

$$x = (x_1, x_2)$$

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

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

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

©DCA-FEEC-Unicamp

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

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

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

©DCA-FEEC-Unicamp

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

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

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

©DCA-FEEC-Unicamp

Aproximação via série 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

©DCA-FEEC-Unicamp

Gradientes e ótimos locais

- Ponto estacionário de f em x^t: $\nabla f(\mathrm{x}^t) = 0$
- Ponto estacionário: ótimo local de uma função objetivo suave
- Condição necessária de 1ª ordem

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

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

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

$$\downarrow$$

função objetivo melhora (a não ser que $\nabla f(\mathrm{x}^t) = 0$)

©DCA-FEEC-Unicamp

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 que melhora em x^t

- Condição 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

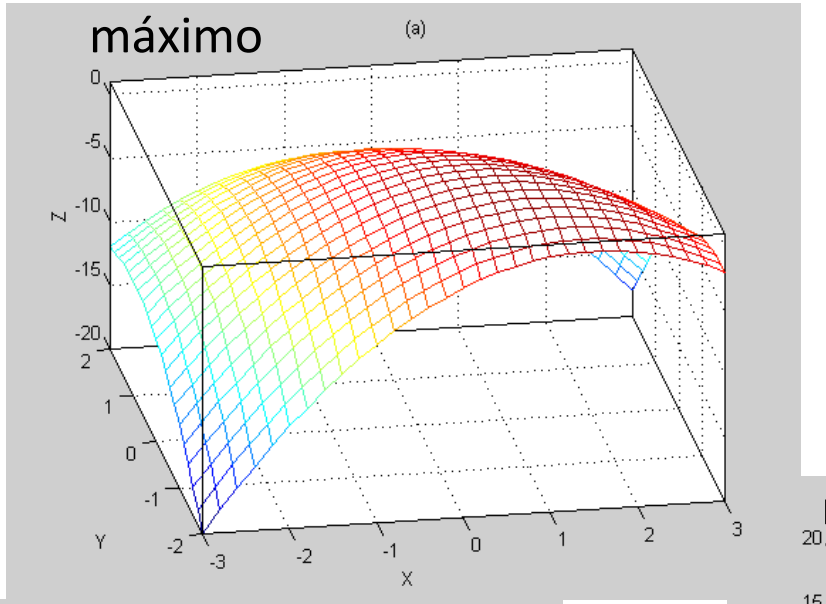
- Condição suficiente

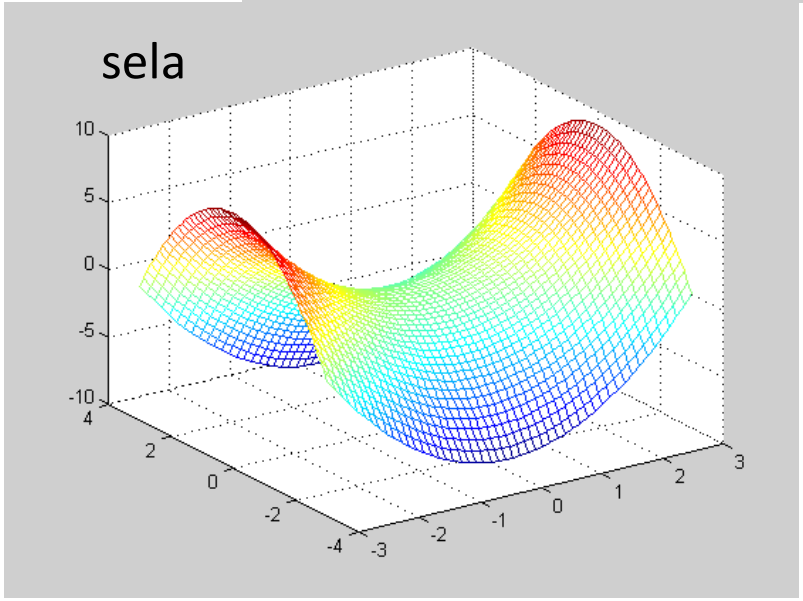
x^t ponto estacionário de f e $H(x^t)$ positiva definida $\rightarrow$ x^t mínimo local de f

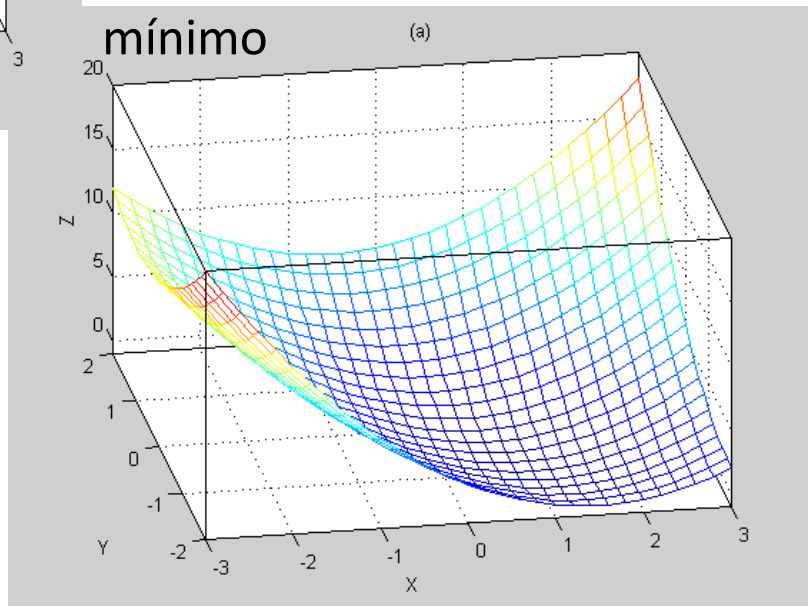
x^t ponto estacionário de f e $H(x^t)$ negativa definida $\rightarrow$ x^t máximo local de f

©DCA-FEEC-Unicamp

Exemplos de pontos estacionários



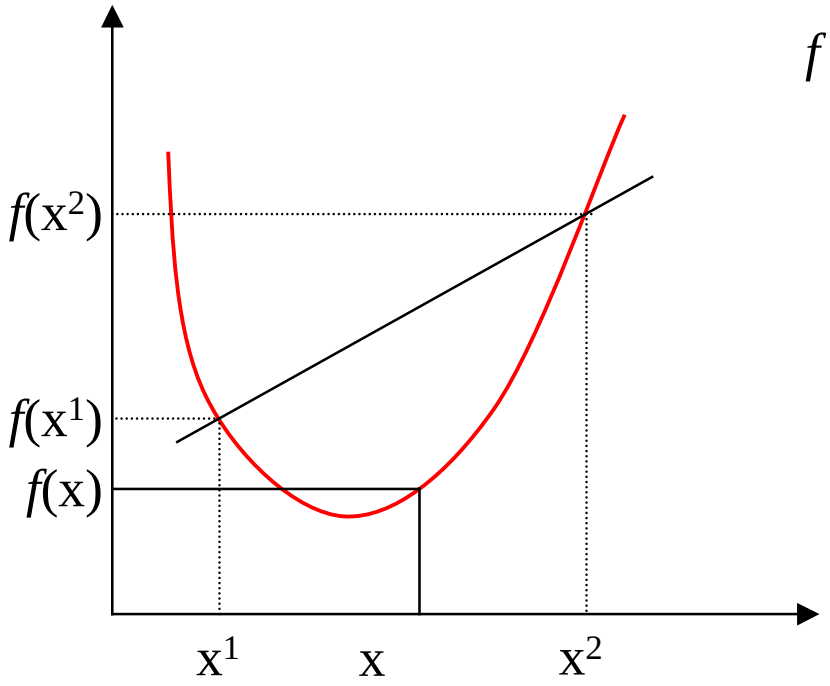


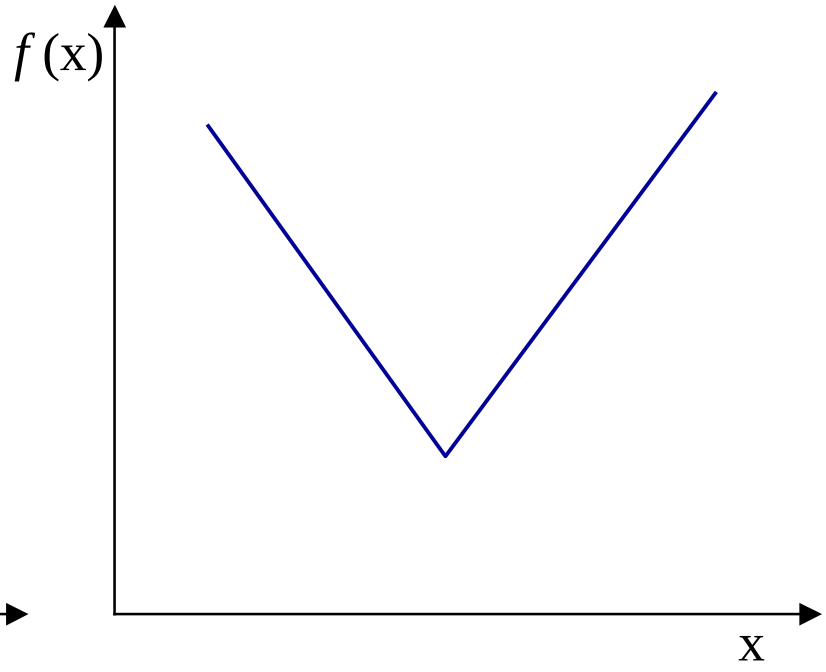


©DCA-FEEC-Unicamp

4-Convexidade e otimalidade global

- Funções convexas

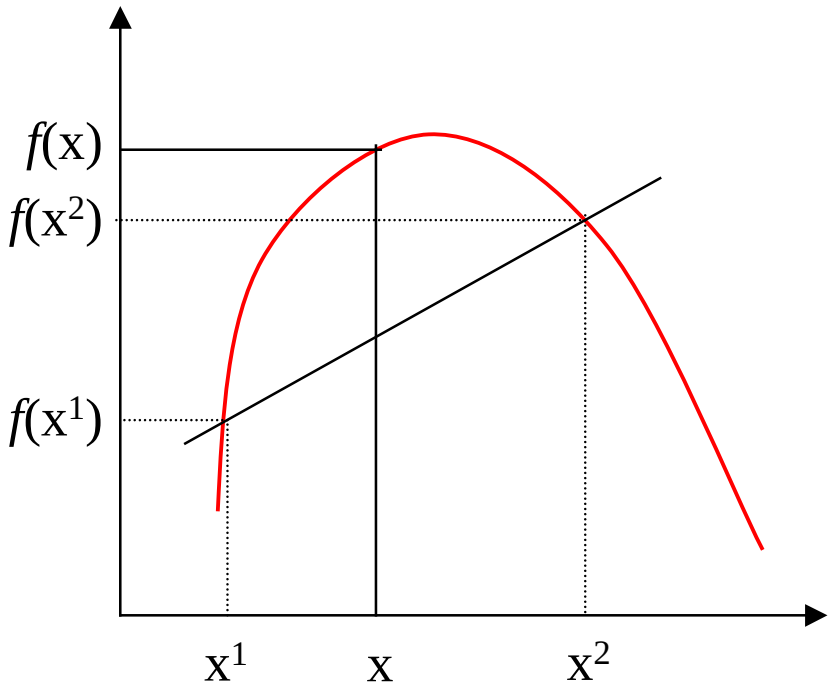


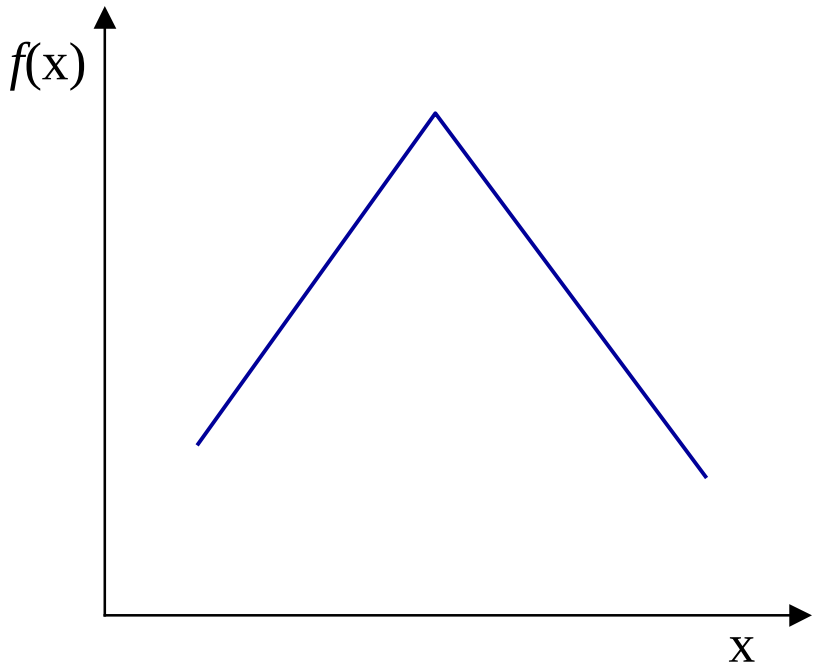


$$f(x^1+\lambda(x^2-x^1)) \le f(x^1)+\lambda[f(x^2)-f(x^1)]; \quad x^1, x^2 \in D; \quad \lambda \in [0,1]$$

©DCA-FEEC-Unicamp

- Funções côncavas





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

©DCA-FEEC-Unicamp

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 $H(\mathrm{x})$ é semi-positiva definida em um domínio convexo (aberto); $f(\mathrm{x})$ é côncava se e somente se $H(\mathrm{x})$ 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 $\forall\ \mathrm{x} \mid f(\mathrm{x}) > 0$

se $f(\mathrm{x})$ é convexa, $g(\mathrm{x}) = 1/\, f(\mathrm{x})$ é côncava $\forall\ \mathrm{x} \mid f(\mathrm{x}) < 0$

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

©DCA-FEEC-Unicamp

6- $f(\mathrm{x})$ é convexa se, para $\alpha_i \geq 0$ 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)

©DCA-FEEC-Unicamp

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

Considerando o 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 que melhora em $\mathrm{x}^1, \forall \mathrm{x}^1 \in$ D

$\downarrow$

x^* ótimo global

©DCA-FEEC-Unicamp

- 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:

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

definição convexidade

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

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

©DCA-FEEC-Unicamp

5-Algoritmos

Algoritmo 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

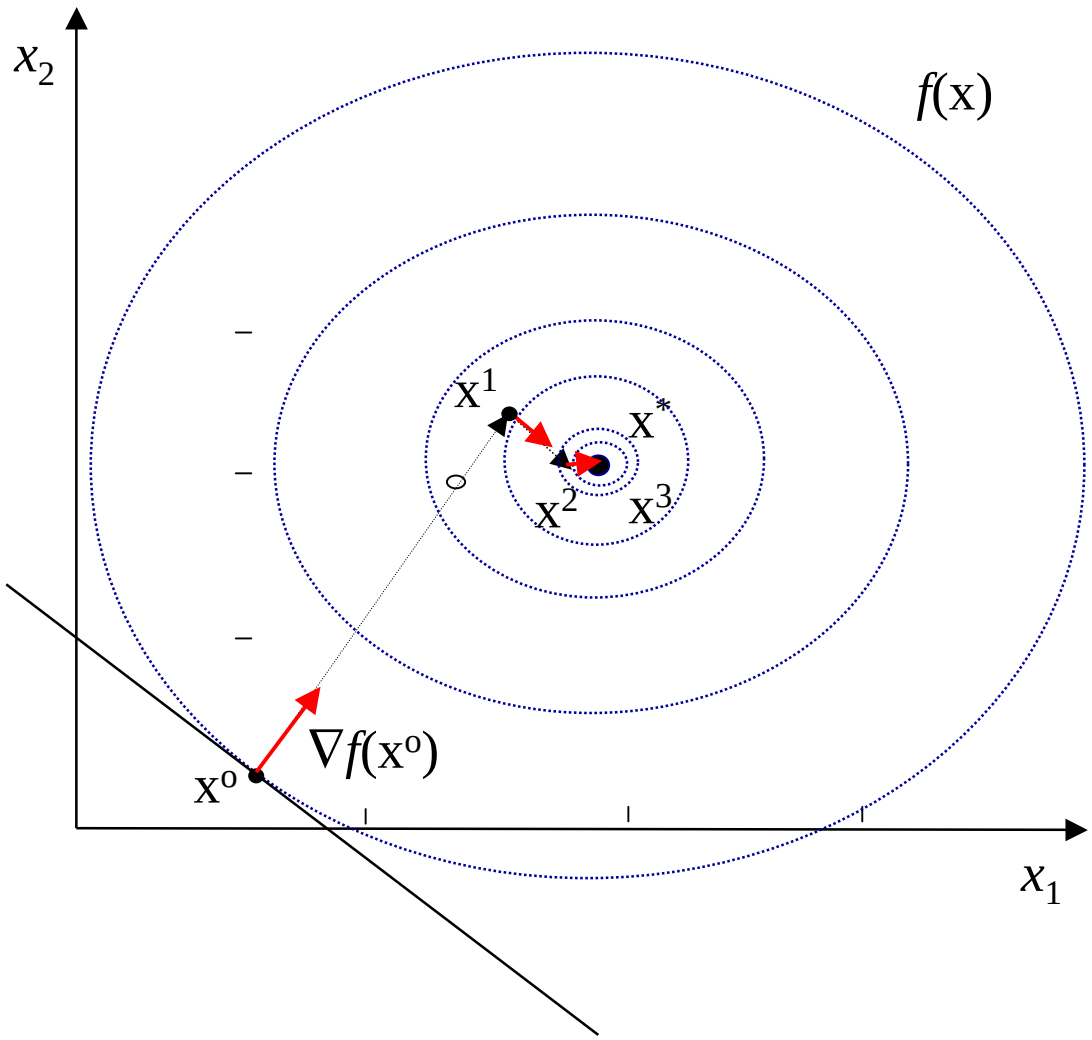
$$\max_\lambda (\min)\ f(x^t + \lambda\, \Delta x^{t+1});$$

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

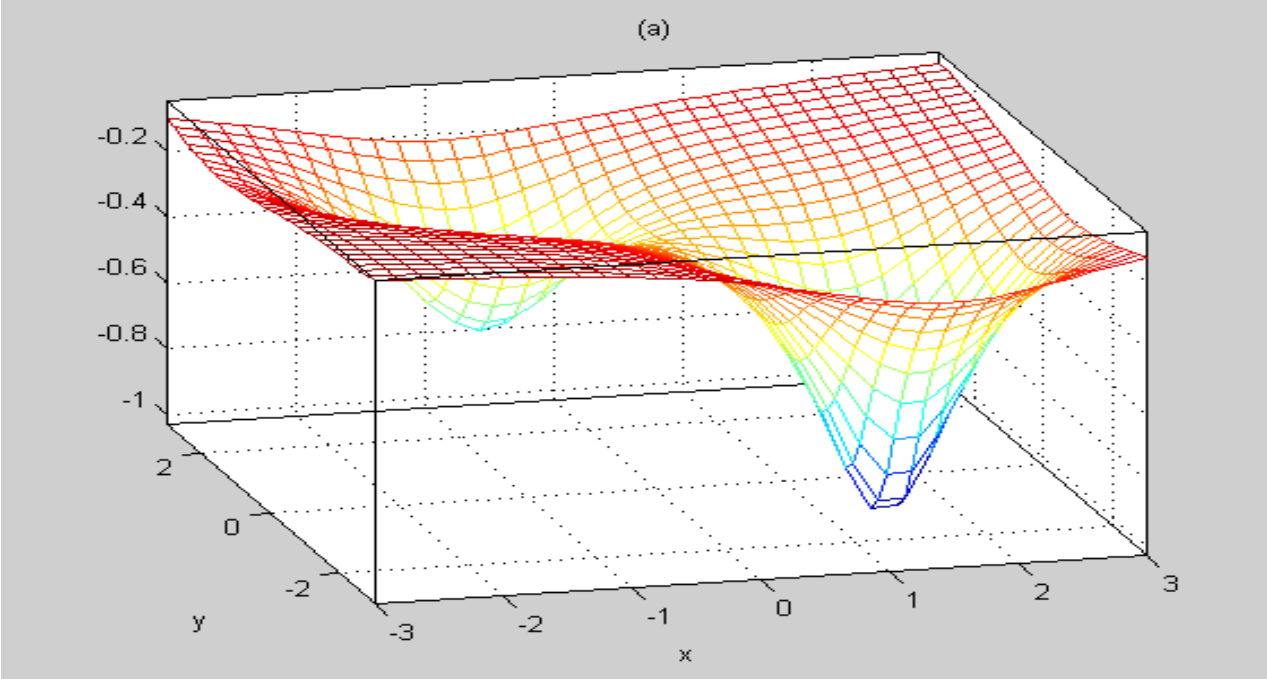
Passo 6 Incrementa: $t = t + 1$; ir para Passo 1;

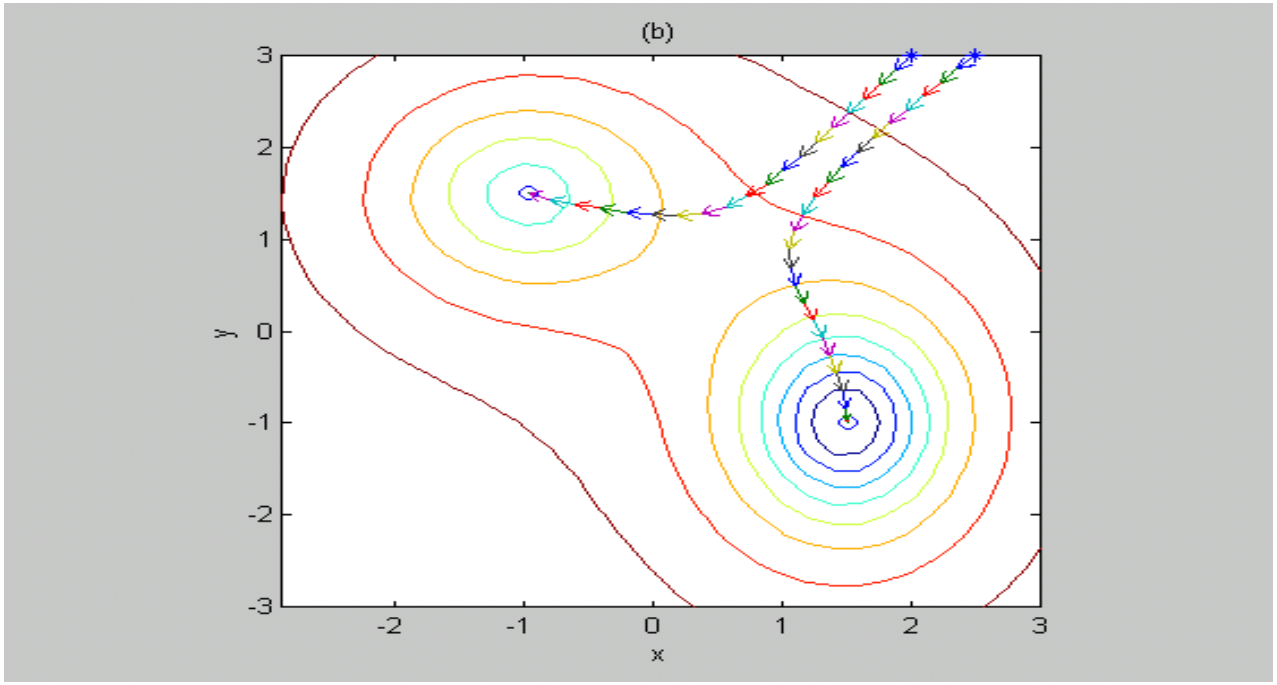
©DCA-FEEC-Unicamp

Algoritmo do gradiente

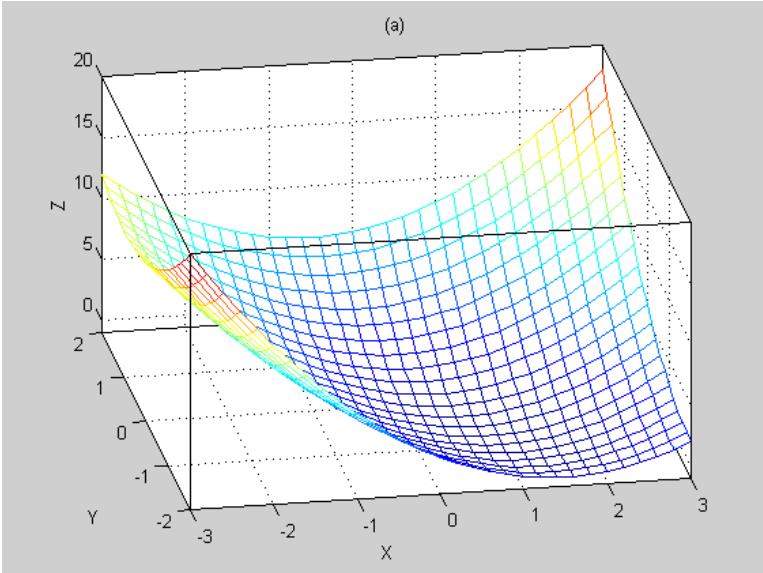


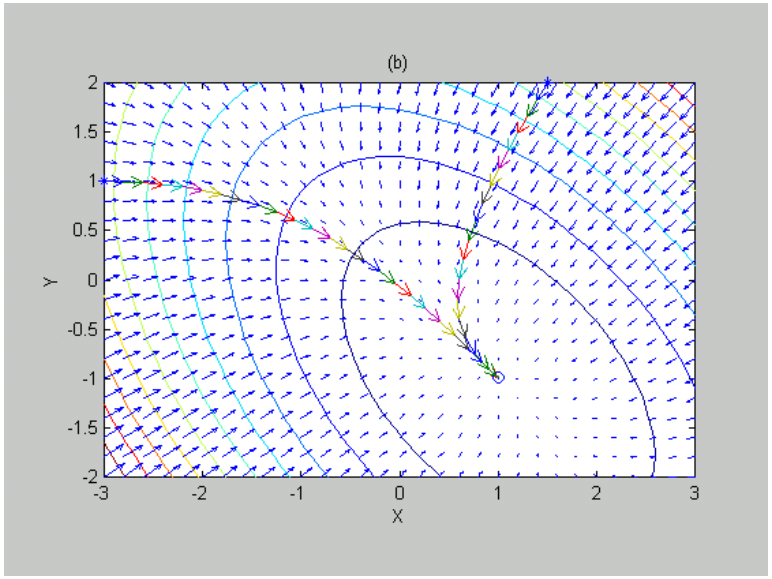
©DCA-FEEC-Unicamp

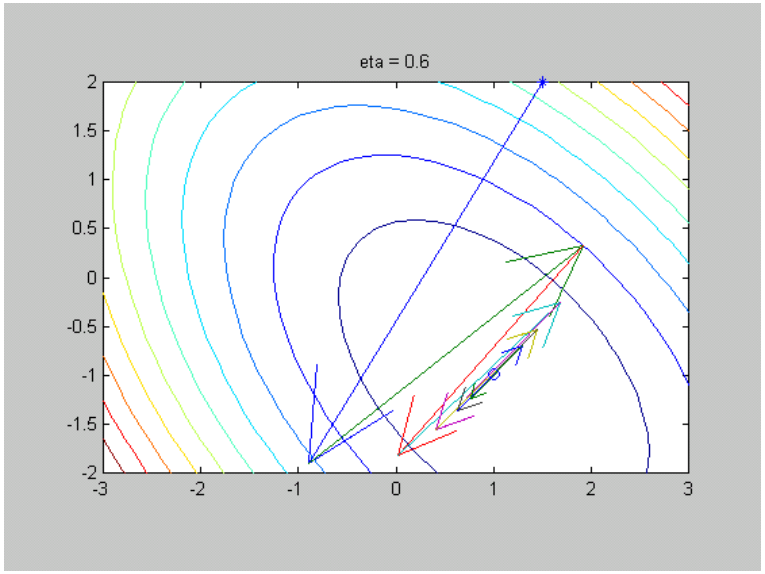


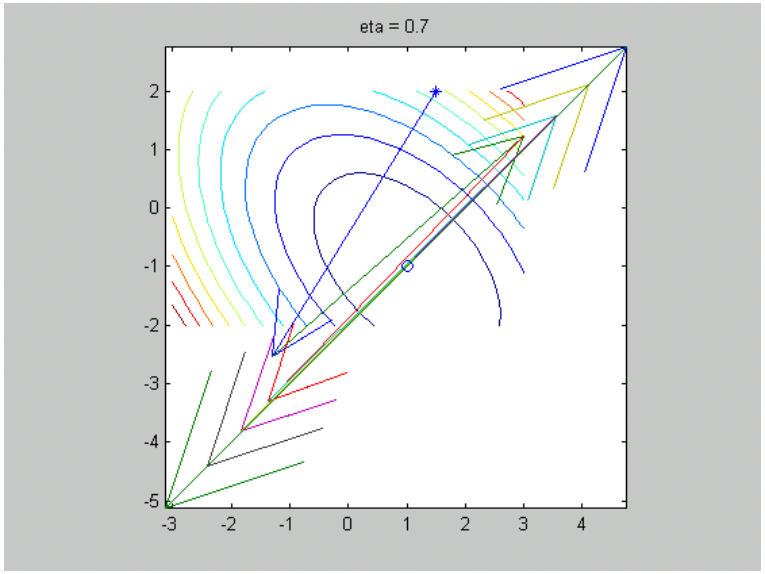


©DCA-FEEC-Unicamp









©DCA-FEEC-Unicamp

Algoritmo 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 $\lambda = 1$ e derivando com relação à Δx_i

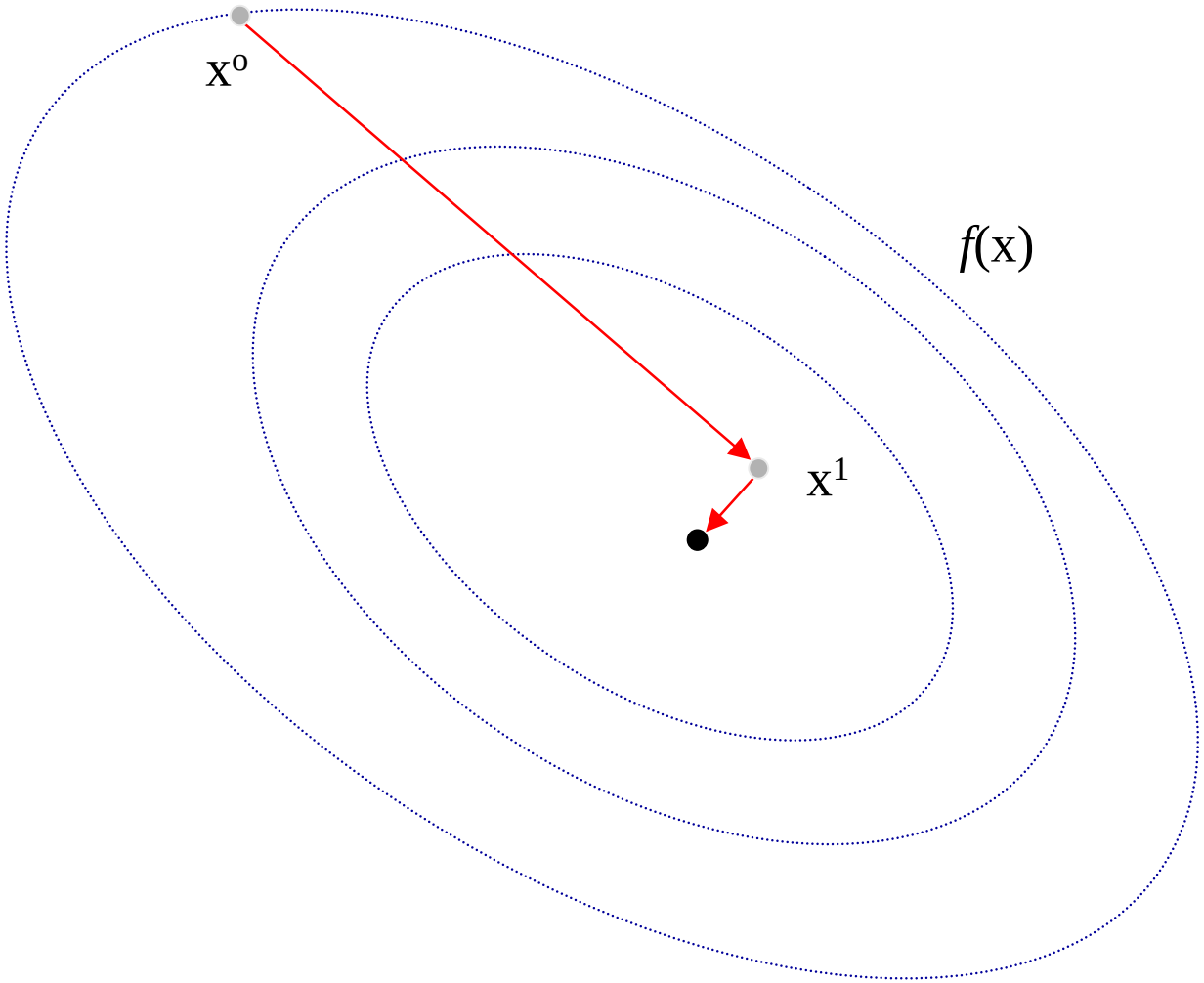
$$\frac{\partial f_2}{\partial \Delta x_i} = \frac{\partial f}{\partial x_i} + \sum_{j=1}^{n}\frac{\partial^2 f}{\partial x_i \partial x_j}\Delta x_j \ , \ i = 1, \cdots, n$$

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

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

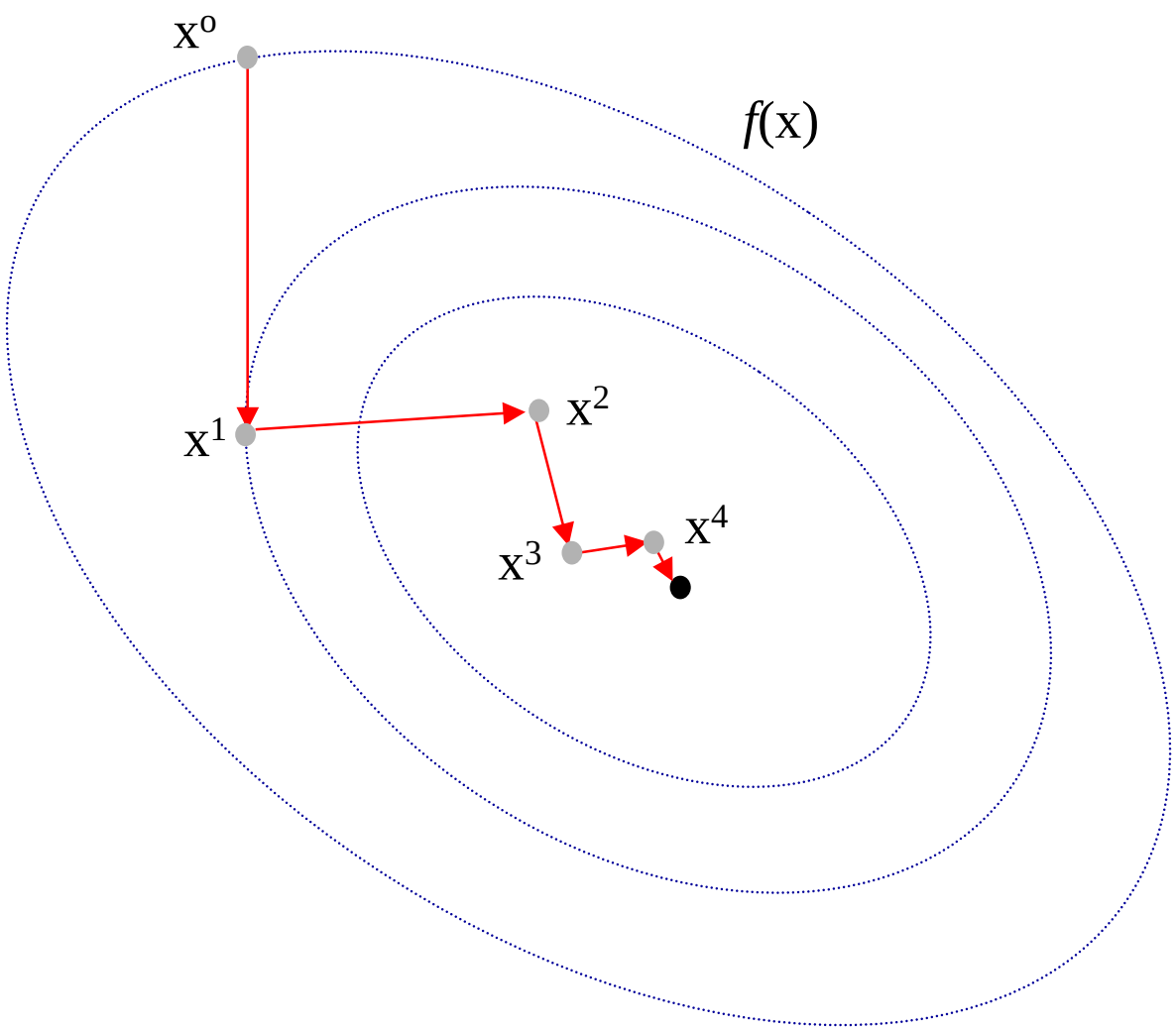
©DCA-FEEC-Unicamp

Newton



©DCA-FEEC-Unicamp

Gradiente



©DCA-FEEC-Unicamp

- 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
- Ideia: combinar gradiente + Newton

$$\downarrow$$

- Métodos quase Newtonianos: $\Delta \mathrm{x}^{t+1} = -D_t \nabla f(\mathrm{x}^t)$
- Matriz D aproxima da inversa da Hessiana H^{-1} durante a busca
- Hessiana: relacionada com a variação do gradiente:

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

©DCA-FEEC-Unicamp

Broyden, Fletcher, Goldfarb, Shanno: BFGS

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

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

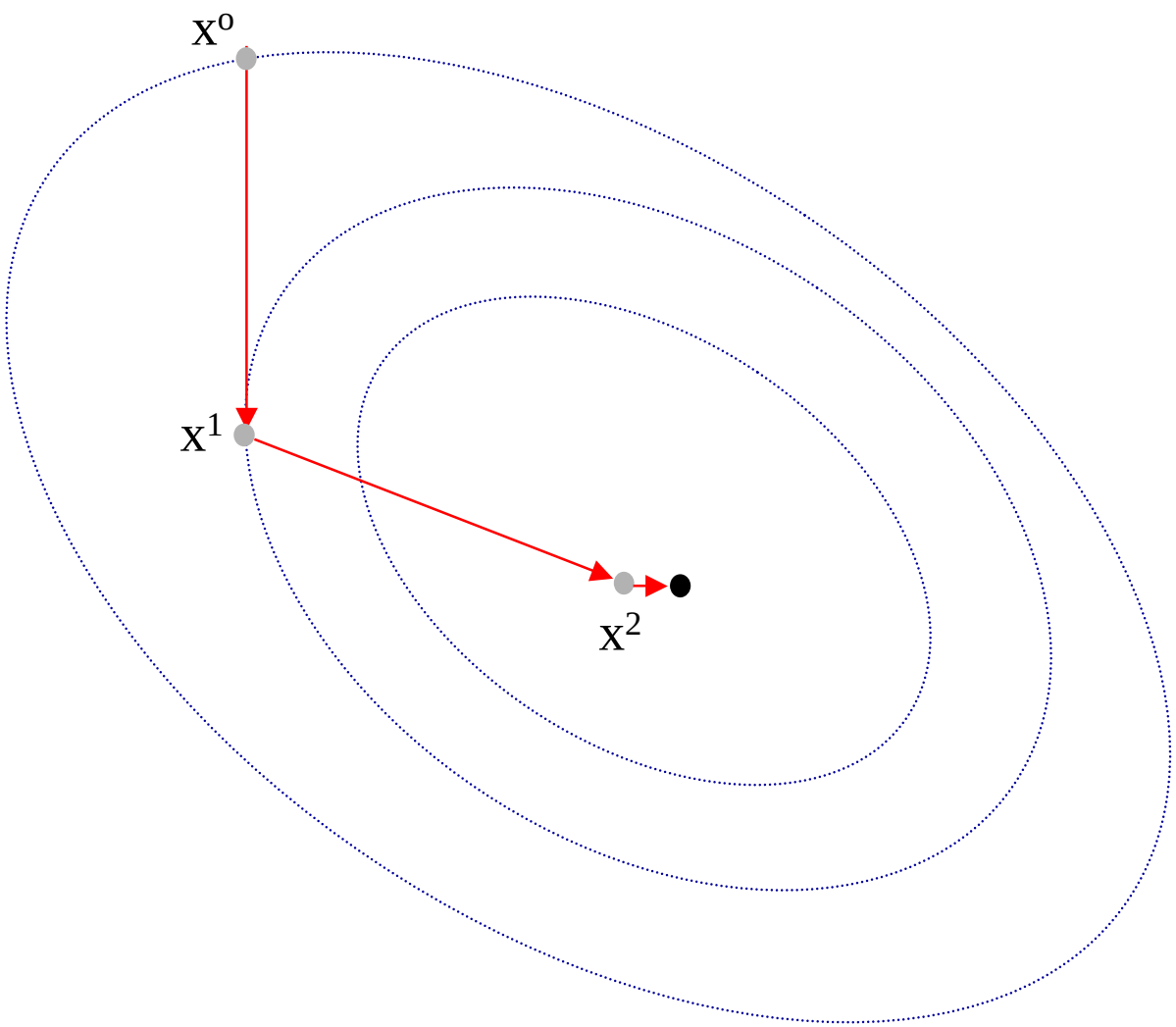
$$D_{t+1} \leftarrow D_t + \left(1 + \frac{\mathrm{g}^T D_t \mathrm{g}}{\mathrm{d}^T \mathrm{g}}\right) \frac{\mathrm{d}\mathrm{d}^T}{\mathrm{d}^T \mathrm{g}} - \frac{D_t \mathrm{g}\mathrm{d}^T + \mathrm{d}\mathrm{g}^T D_t}{\mathrm{d}^T \mathrm{g}}$$

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

$(-\ \min,\ +\ \max)$

©DCA-FEEC-Unicamp

BFGS



©DCA-FEEC-Unicamp

Algoritmo de Nelder-Mead

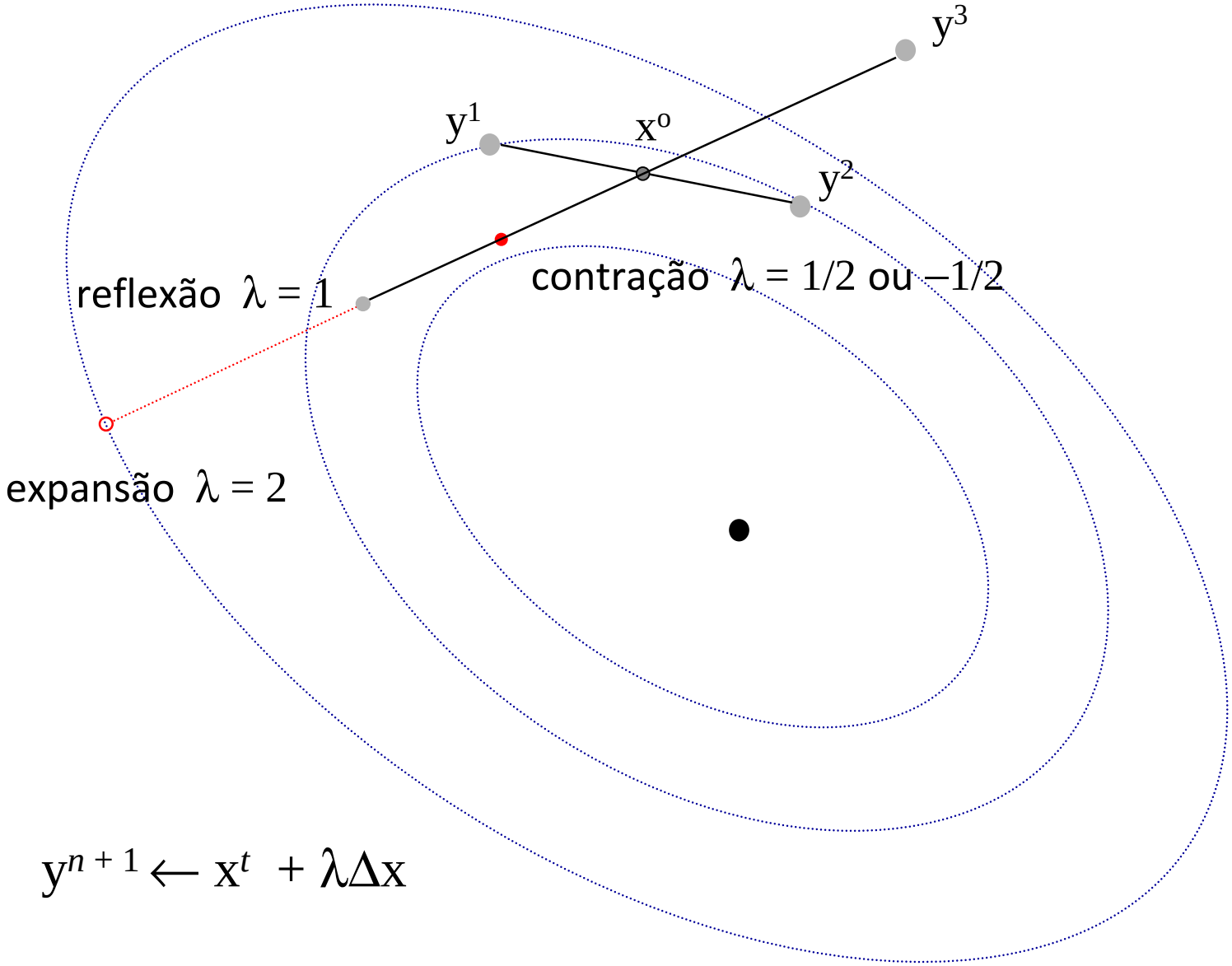
- Não utiliza derivadas
- Mantém $(n+1)$ soluções candidatas
- Baseia-se nos conceitos de:
  - reflexão
  - expansão
  - contração
  - encolhimento

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

$\mathrm{x}^t \equiv \dfrac{1}{n}\sum_{i=1}^{n} \mathrm{y}^i$ $\qquad$ centróide dos n melhores

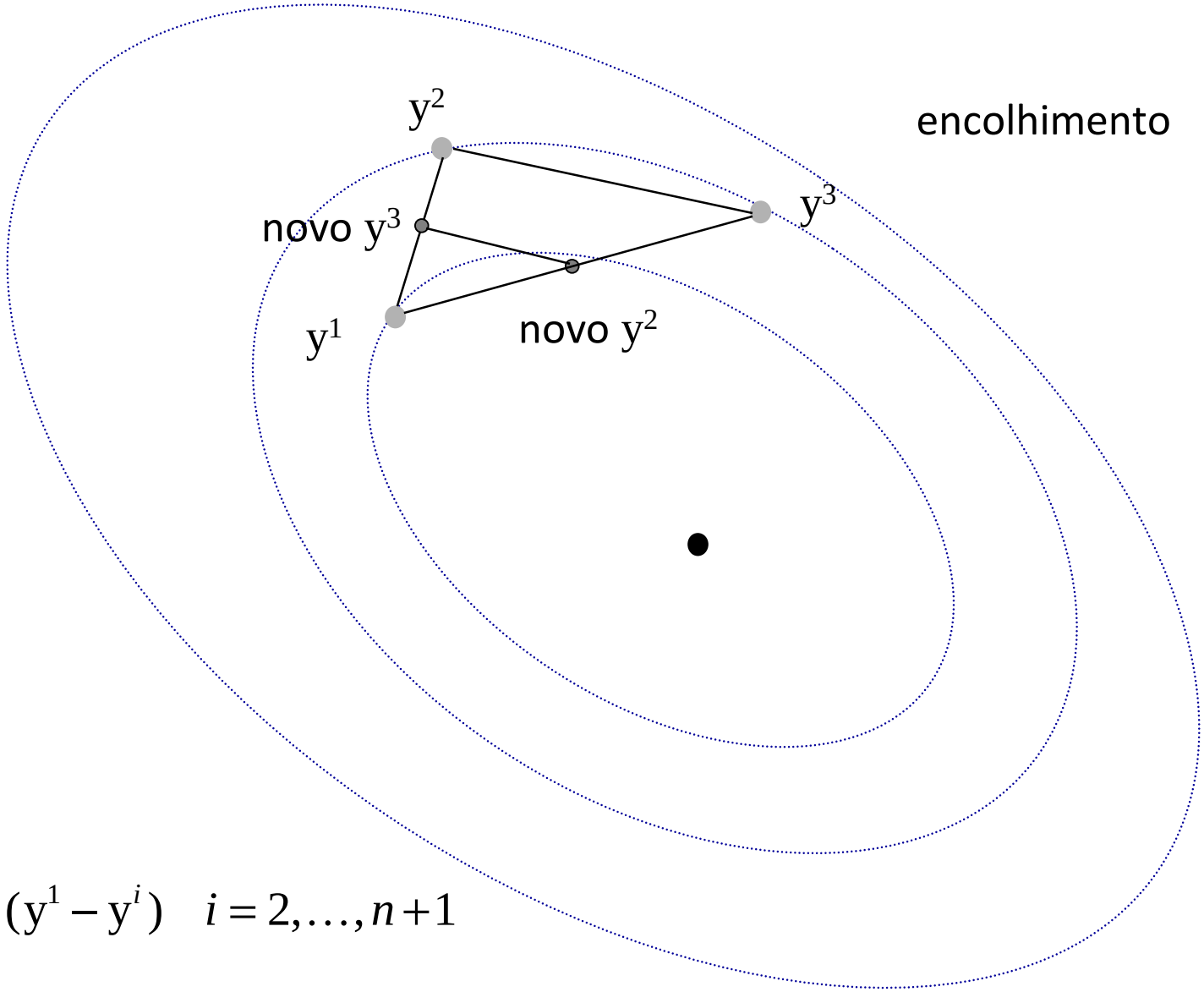
$\mathrm{y}^i = \dfrac{1}{2}(\mathrm{y}^1 - \mathrm{y}^i) \quad i = 2,\ldots,n+1$ $\qquad$ encolhimento

©DCA-FEEC-Unicamp



$$y^{n+1} \leftarrow x^t + \lambda \Delta x$$

©DCA-FEEC-Unicamp



$$\mathrm{y}^i = \frac{1}{2}(\mathrm{y}^1 - \mathrm{y}^i) \quad i = 2, \ldots, n+1$$

©DCA-FEEC-Unicamp

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.

©DCA-FEEC-Unicamp