Programação Não Linear

Modelo programação não linear (PNL)

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

restrito

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

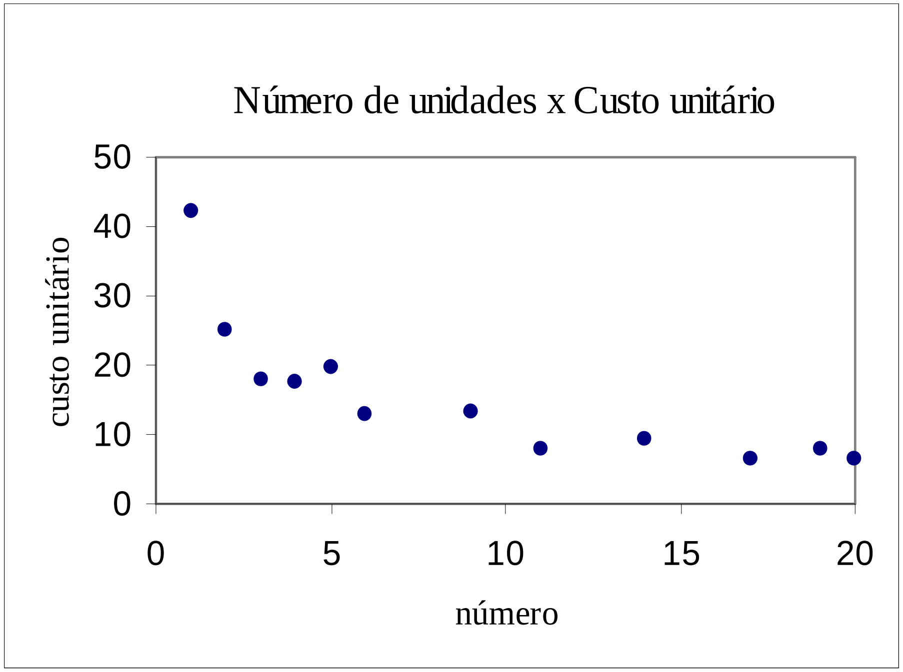
irrestrito

I- programação não linear irrestrita

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

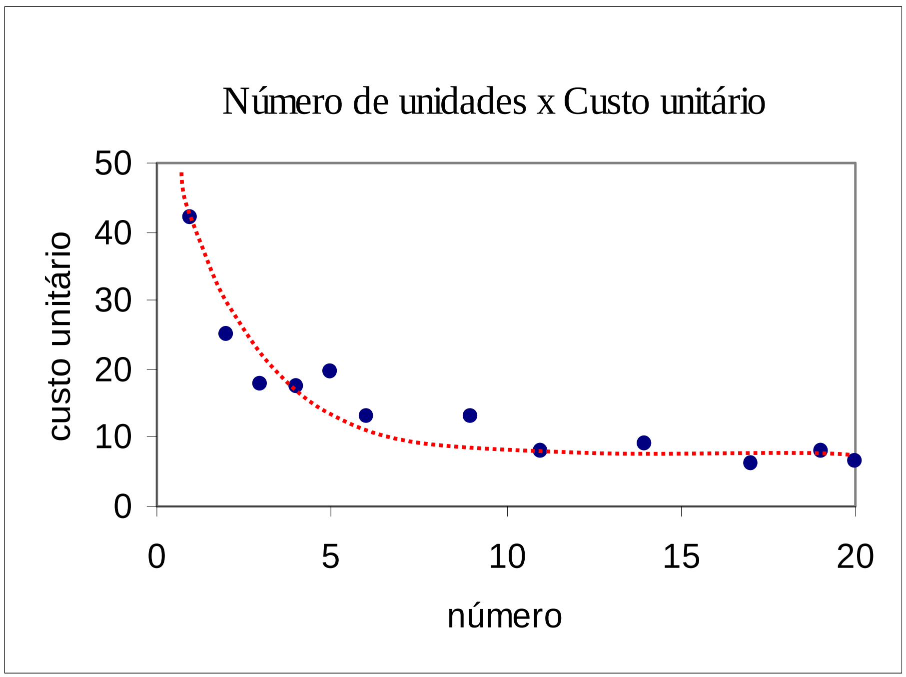
Exemplo: regressão não linear

	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



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

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



$$m = 12$$

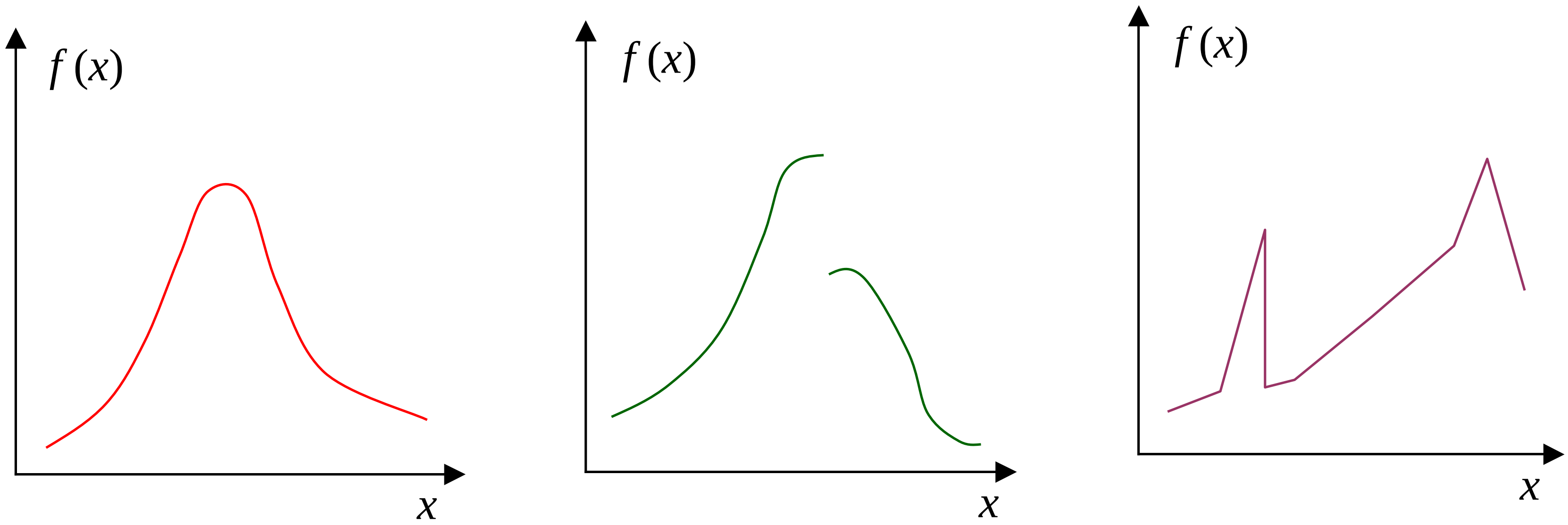
$$\downarrow$$

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

$$\downarrow$$

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

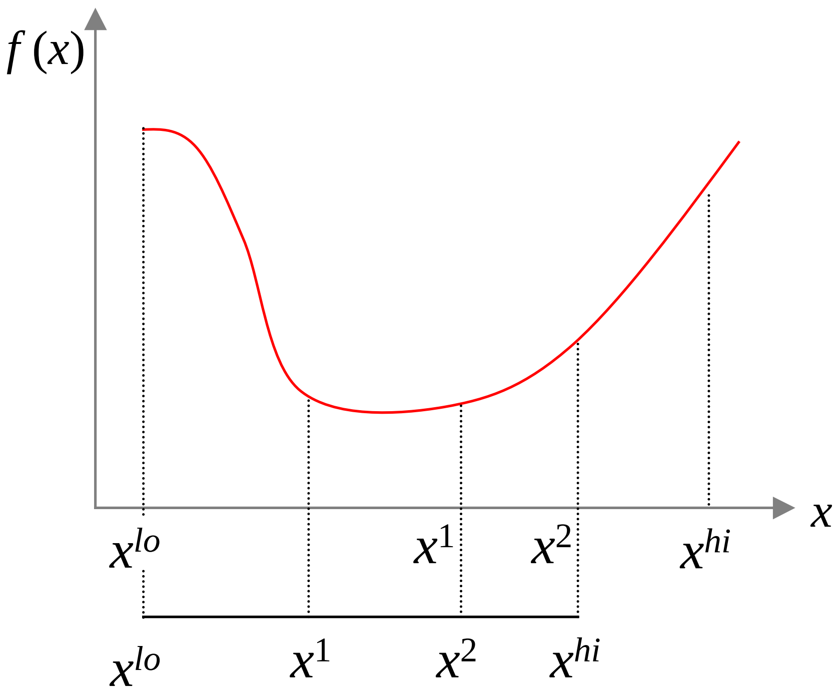
Funções suaves e derivadas

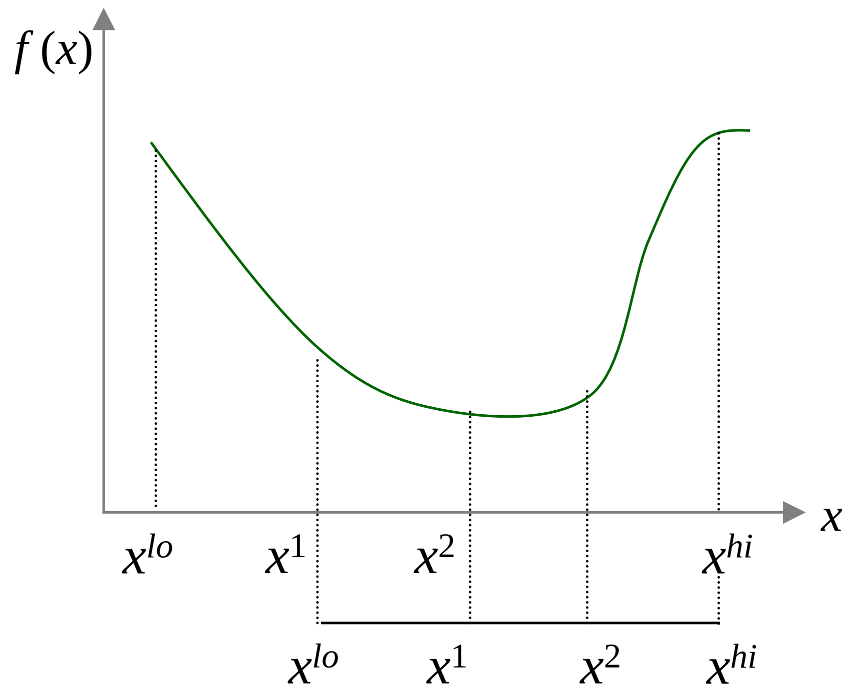


suave | não contínua | não diferenciável

- 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 : forma 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

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

Busca unidimensional (número de ouro)

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

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

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

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 $f\,(x^1) > f\,(x^2)$; caso contrário ir para o Passo 3;

Passo 2 Esquerdo: estreitar lado esquerdo do intervalo;

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

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

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

Passo 3 Direito: estreitar lado direito do intervalo;

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

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

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

Busca unidimensional: intervalo inicial

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

Intervalo inicial: algoritmo

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 sem restrições

- Vetor gradiente

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

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

- Matriz Hessiana

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

Exemplo: regressão não linear

$$f(x_1, x_2) = \sum_{i=1}^{m} [q_i - x_1 (p_i)^{x2}]^2$$

$$\frac{\partial f}{\partial x_1} = -2\sum_{i=1}^{m} (q_i - x_1 p_i^{x2}) p_i^{x2}$$

$$\frac{\partial f}{\partial x_2} = -2\sum_{i=1}^{m} (q_i - x_1 p_i^{x2}) x_1 p_i^{x2} \ln(p_i)$$

$$\frac{\partial f}{\partial x_1^2} = 2\sum_{i=1}^{m} p_i^{x2}$$

$$\frac{\partial f}{\partial x_2^2} = -2\sum_{i=1}^{m} \ln^2(p_i)[(q_i - x_1 p_i^{x2})(x_1 p_i^{x2}) - (x_1 p_i^{x2})^2]$$

$$\frac{\partial f}{\partial x_1 \partial x_2} = \frac{\partial f}{\partial x_2 \partial x_1} = -2\sum_{i=1}^{m} [(q_i - x_1 p_i^{x2})(p_i^{x2}) \ln(p_i) - (p_i^{x2})(x_1 p_i^{x2}) \ln(p_i)]$$

$$x = (x_1, x_2)$$

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

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

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

Aproximações usando série de Taylor

$$f_1(x^t + \lambda\,\Delta\, x) = f\,(x^t) + \lambda \nabla f\,(x^t)\Delta x$$

1ª ordem

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

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

2ª ordem

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

Gradiente e ótimo local

- Condição necessária (de 1ª ordem)
  - x^t é um ponto estacionário de f se $\nabla f(x^t) = 0$
  - ótimo local de uma função suave: ponto estacionário

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

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

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

$$\Downarrow$$

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

Hessianas e ótimos locais

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

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

$$\Downarrow$$

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

Δx direção que melhora f em x^t

- Condições de 2ª ordem

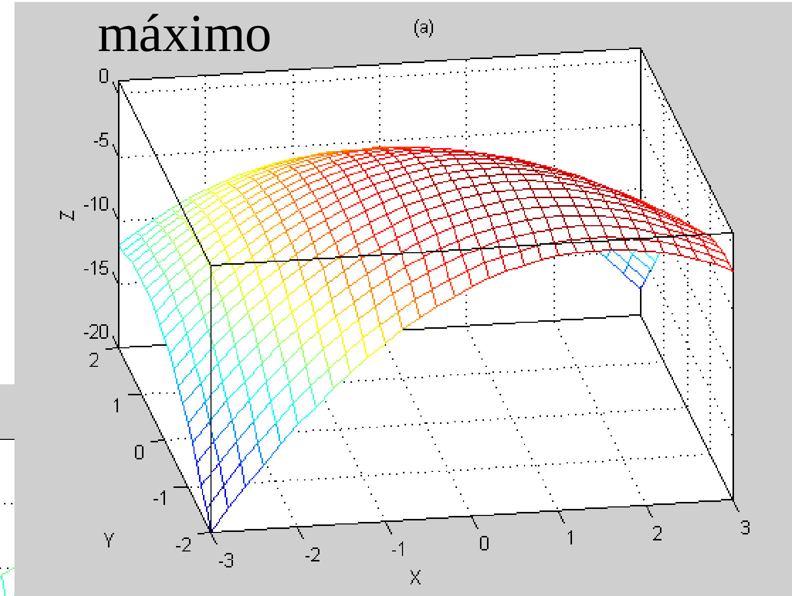
  – necessárias

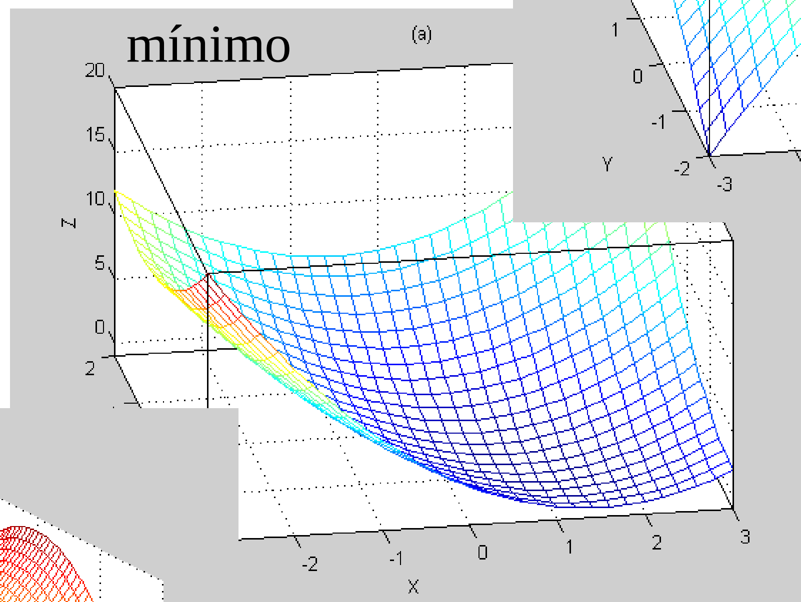
    - x^t mínimo local de f $\rightarrow$ $H(x^t)$ semi-positiva definida
    - x^t máximo local de f $\rightarrow$ $H(x^t)$ semi-negativa definida

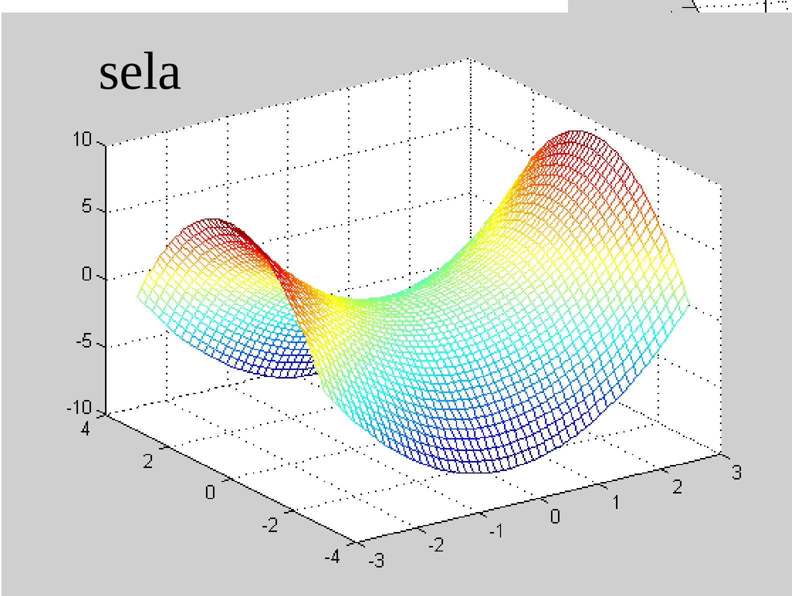
  – suficientes

    - x^t ponto estacionário f , $H(x^t)$ positiva definida $\rightarrow$ x^t mínimo local de f
    - x^t ponto estacionário f , $H(x^t)$ negativa definida $\rightarrow$ x^t máximo local de f

Pontos estacionários

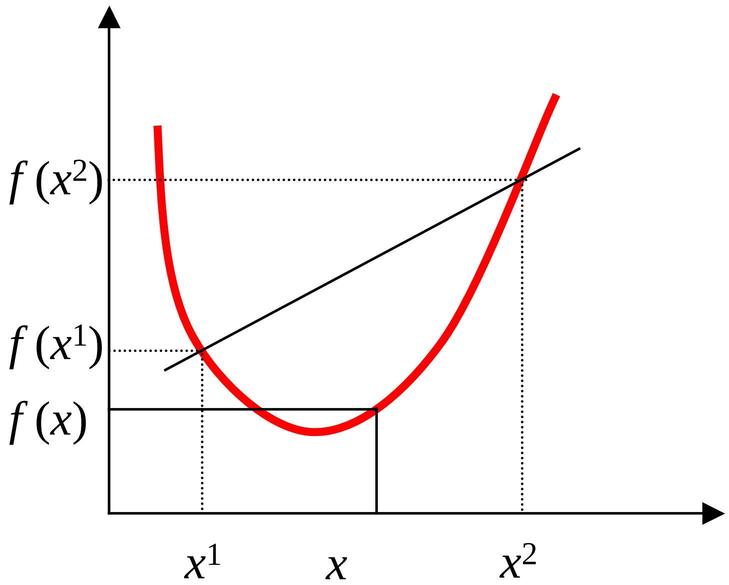


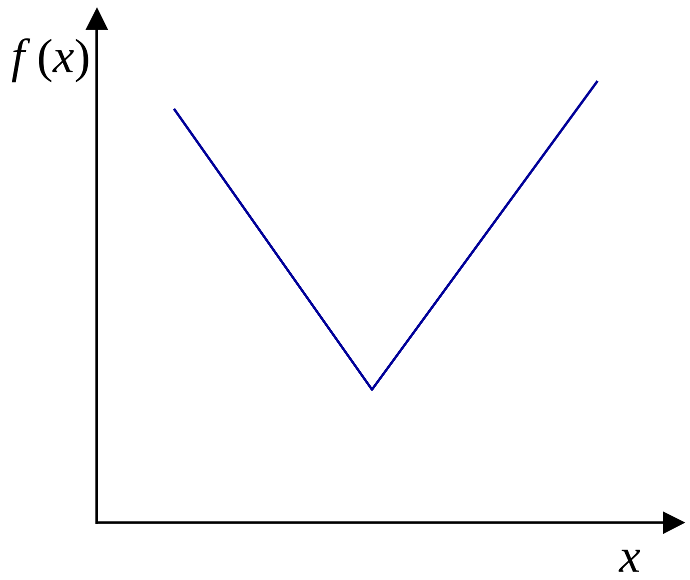




Convexidade e otimalidade global

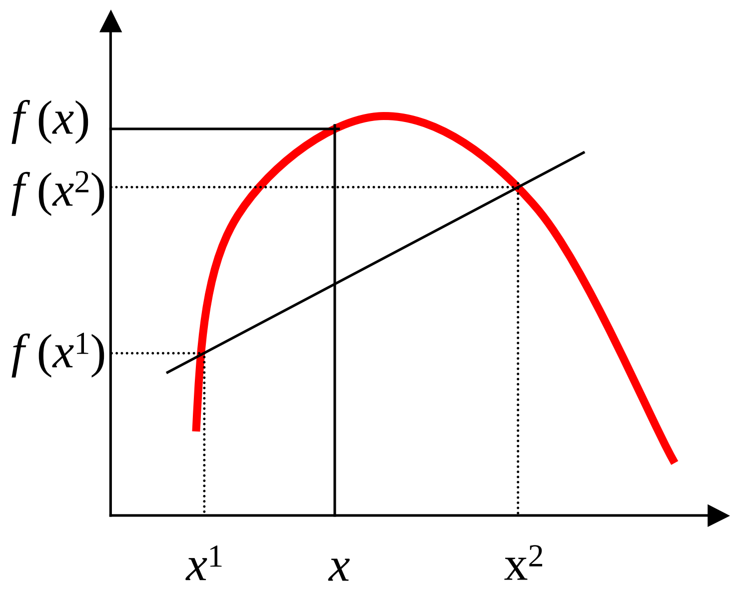
- Função convexa

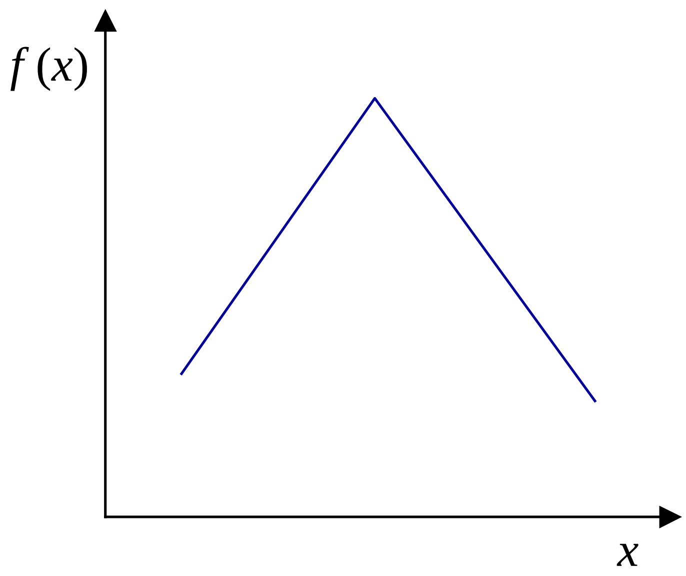




$$f(x^1 + \lambda(x^2 - x^1)) \leq f(x^1) + \lambda[f(x^2) - f(x^1)]; \quad x^1, x^2 \in \mathrm{D}; \quad \lambda \in [0,1]$$

- Função côncava





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

Funções convexas e côncavas

1- Se $f(x)$ é convexa então $-f(x)$ é côncava

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

3- Funções lineares são convexas e côncavas

4- Se $f(x)$ é côncava, $g(x) = 1/f(x)$ é convexa $\forall x \mid f(x) > 0$
se $f(x)$ é convexa, $g(x) = 1/f(x)$ é côncava $\forall x \mid f(x) < 0$

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

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

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

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

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

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

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

Condições suficientes: otimalidade global

- Se $f(x)$ é uma função convexa, então todo mínimo local é mínimo global
- Se $f(x)$ é uma função côncava, então todo máximo local é máximo global

No caso de mínimo: seja x^* mínimo global e $x^1 \neq x^*$

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

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

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

$$\downarrow$$

x^* mínimo global

- Ponto estacionário de uma função convexa suave é um mínimo global

- Ponto estacionário de uma função côncava suave é um máximo global

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

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

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

$$f(x) - f(x^*) \geq \nabla f(x^*)(x - x^*)$$

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

$$\downarrow$$

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];

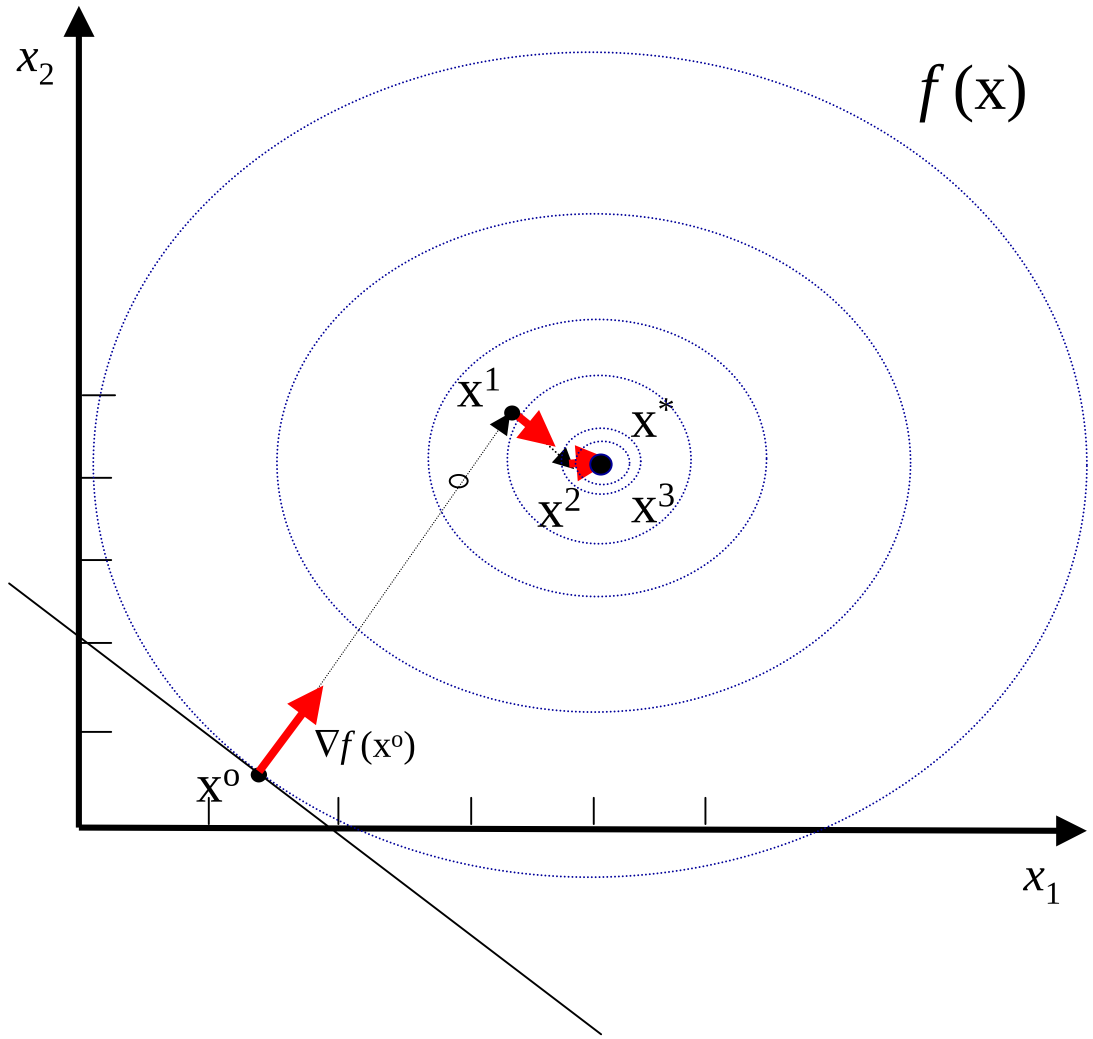
Passo 4 Busca Unidimensional: determinar λ_{t+1} resolvendo

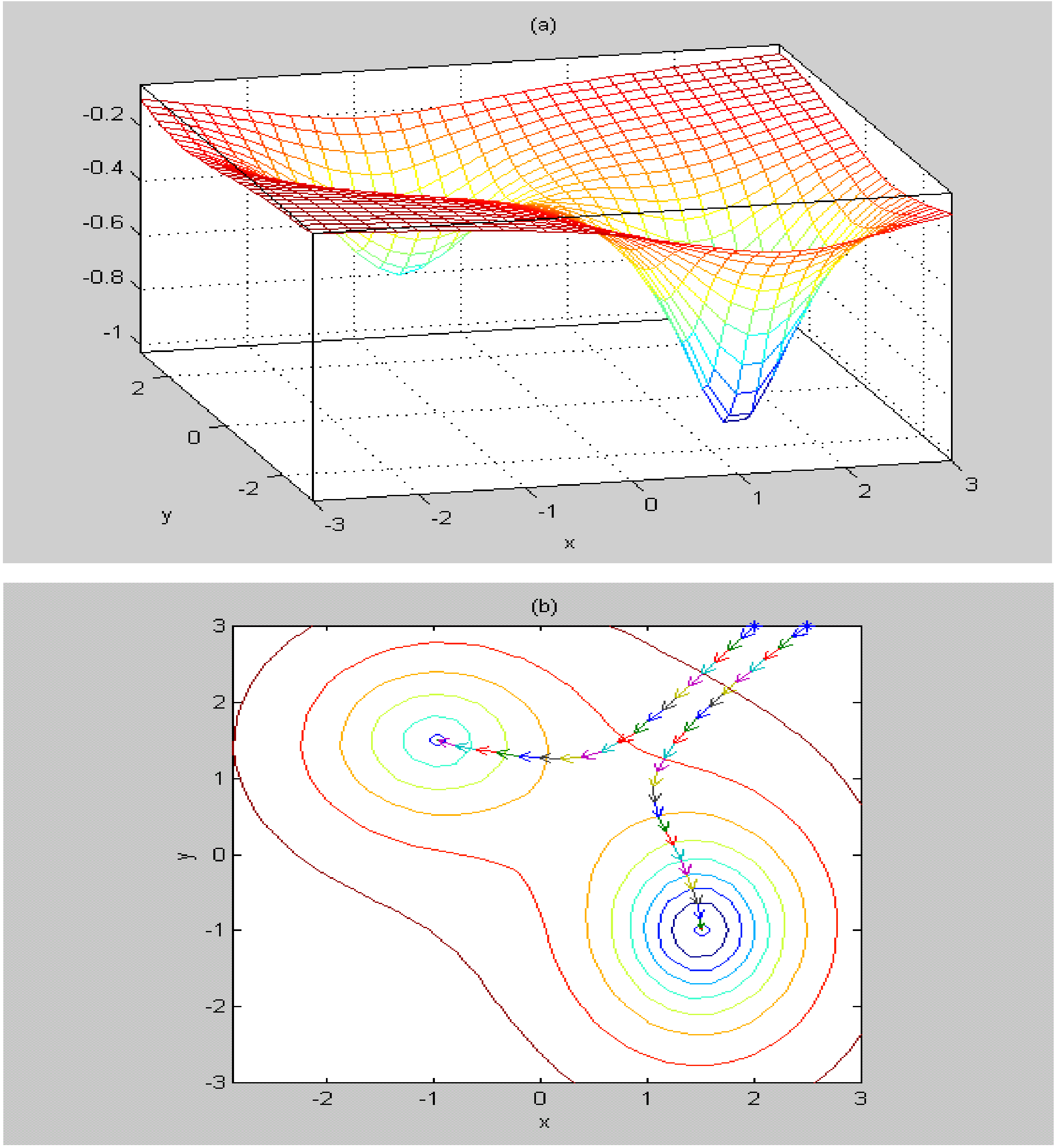
$$\max\ (\min)\ f(x^t + \lambda\, \Delta x^{t+1})\ ;$$

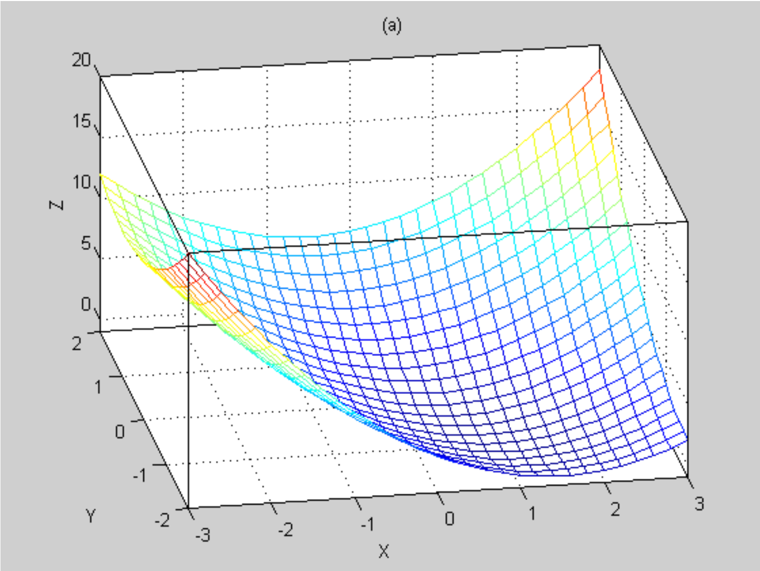
Passo 5 Atualizar: $x^{t+1} = x^t + \lambda\, \Delta x^{t+1}$;

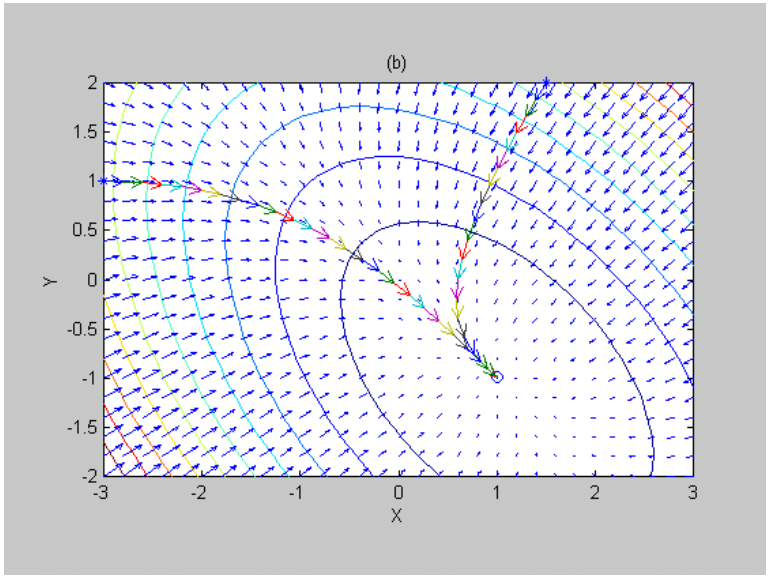
Passo 6 Incrementa: $t = t + 1$; ir para Passo 1;

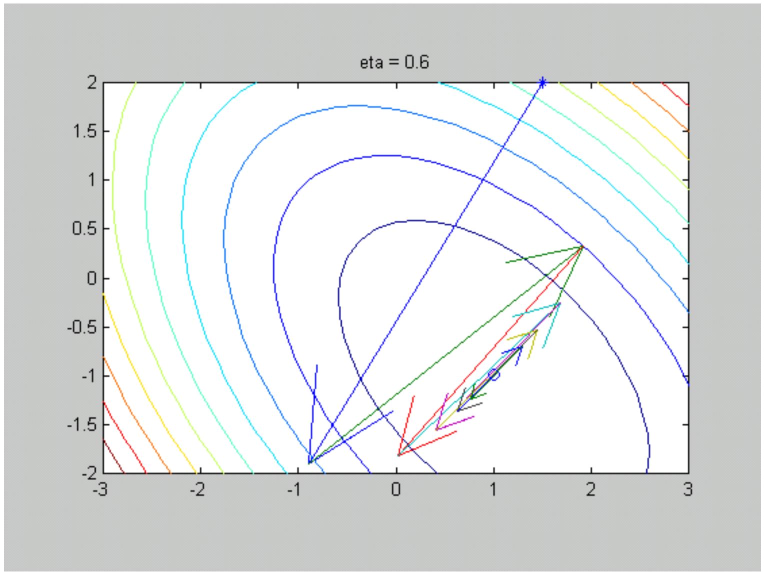
Algoritmo do gradiente

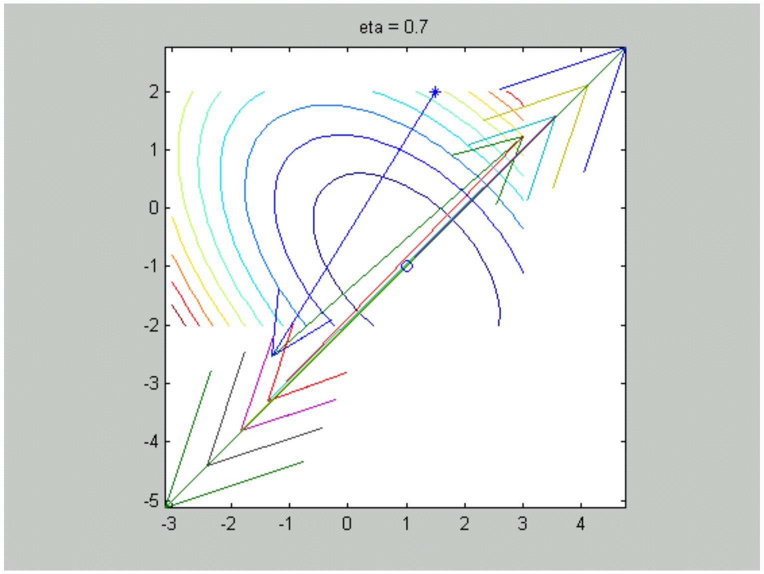



(a)
-0.2
-0.4
-0.6
-0.8
-1
2
0
-2
y
-3
-2
-1
0
1
2
3
x
(b)
3
2
1
0
-1
-2
-3
y
-2
-1
0
1
2
3
x


(a)
Z
Y
X


(b)
Y
X


eta = 0.6


eta = 0.7

Método de Newton

- Utiliza informação de segunda ordem

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

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

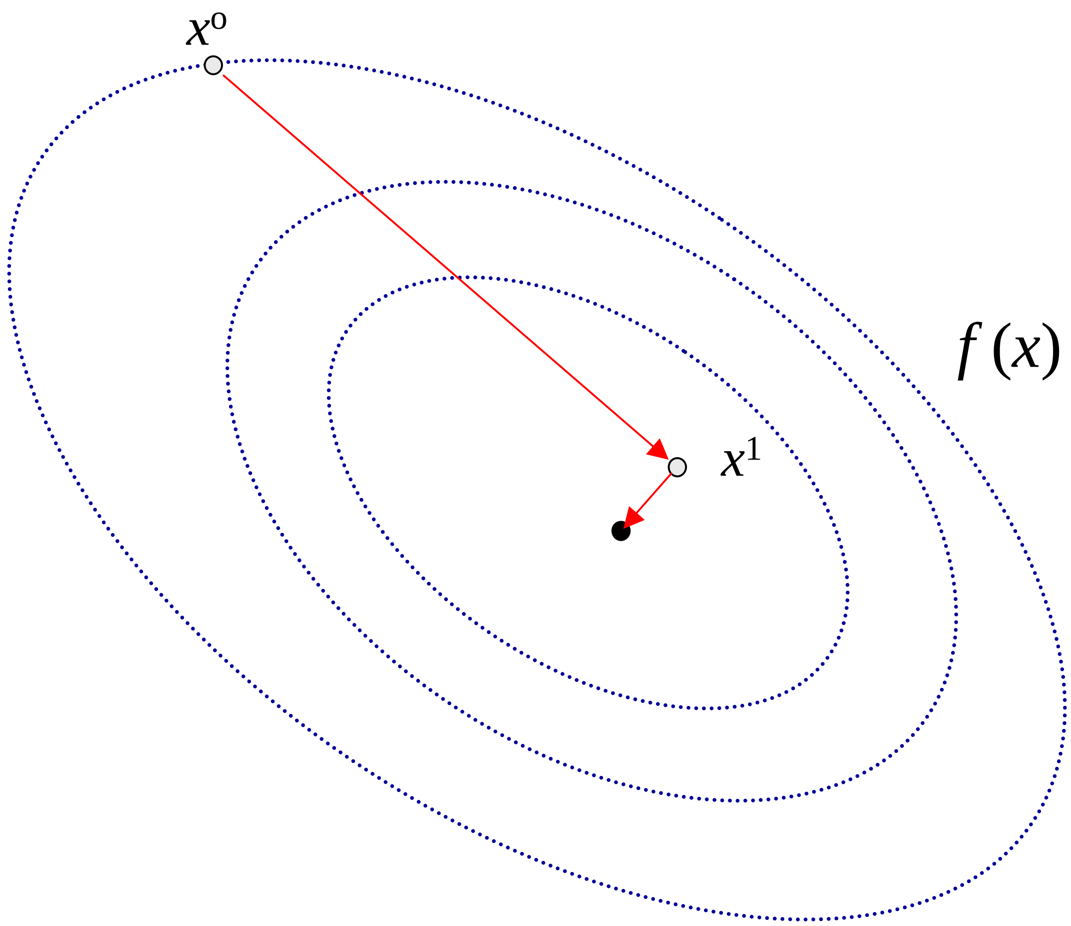
- Fazendo $\lambda = 1$ e derivando com relação à Δx_i

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

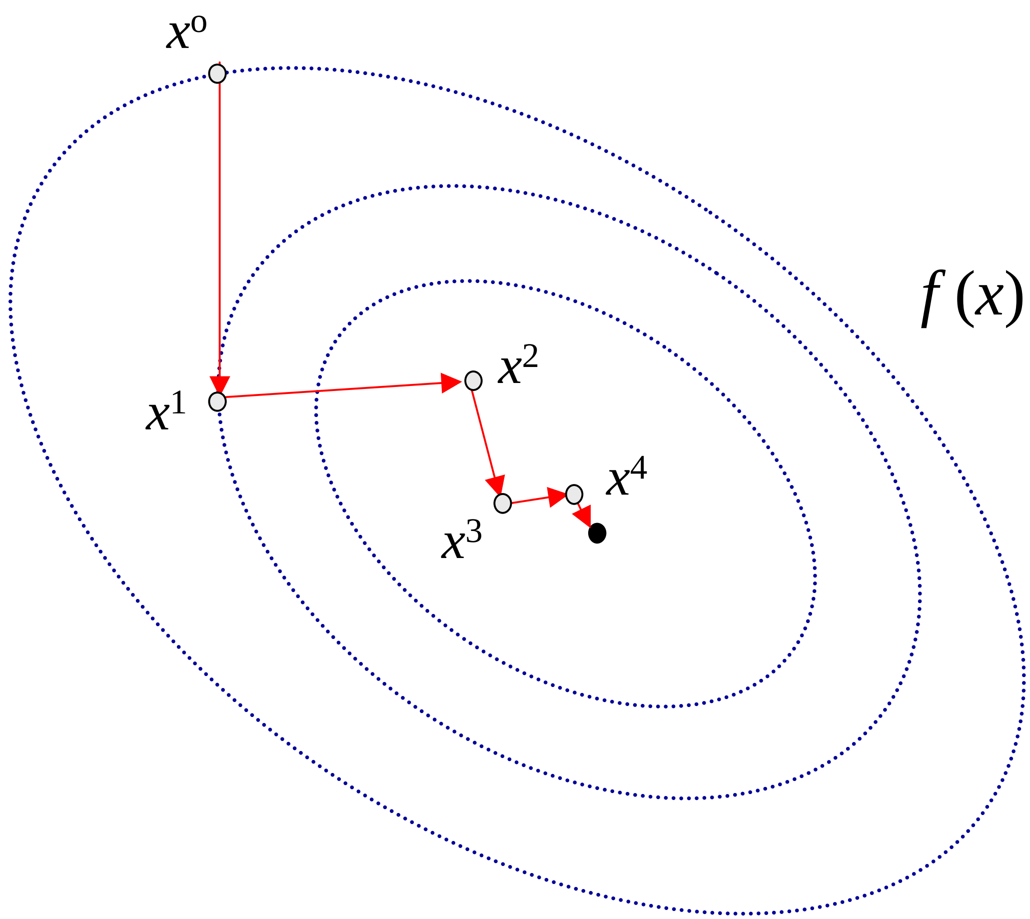
$$\nabla f_2(\Delta x) = \nabla f(x^t) + H(x^t)\Delta x$$

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

Newton



Gradiente



– algoritmo Newton 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: $\Delta x^{t+1} = - D_t \nabla f(x^t)$

– 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

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

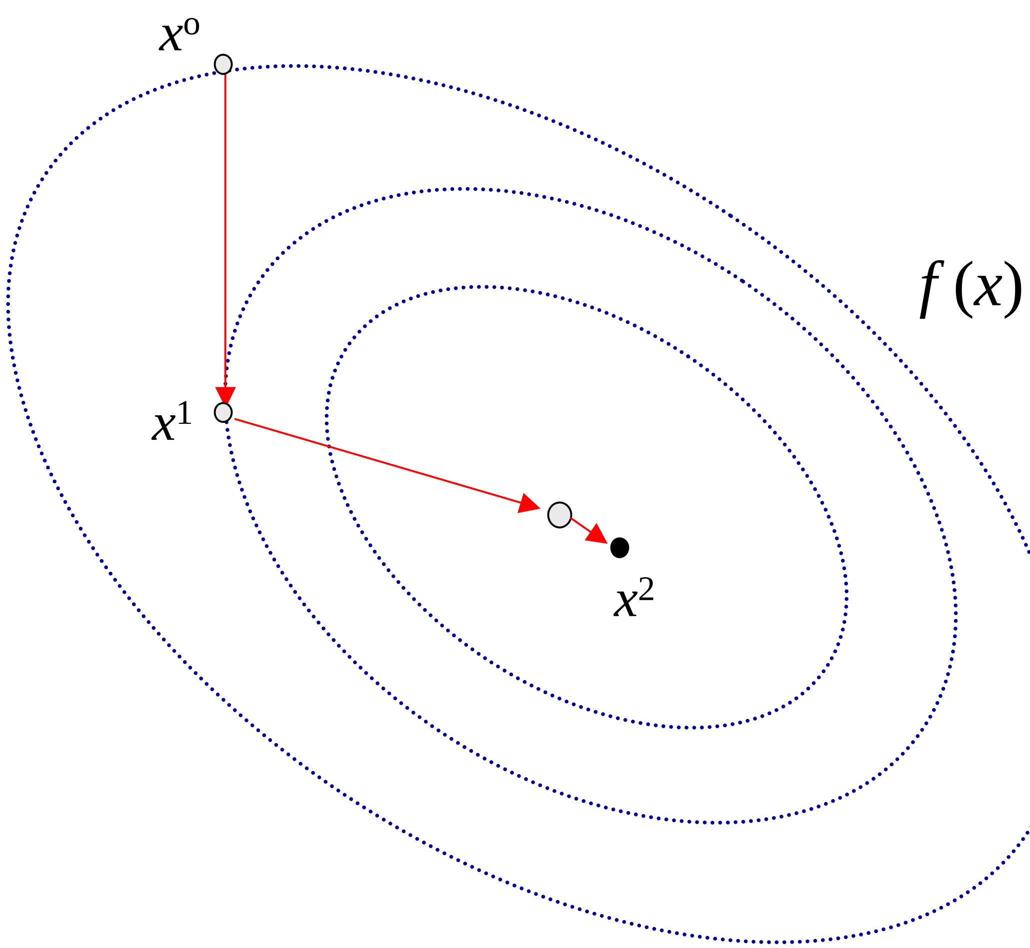
$$x^{t+1} \leftarrow x^t + \lambda_{t+1} \Delta x^{t+1}$$

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

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

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

BFGS



Algoritmo de Nelder-Mead

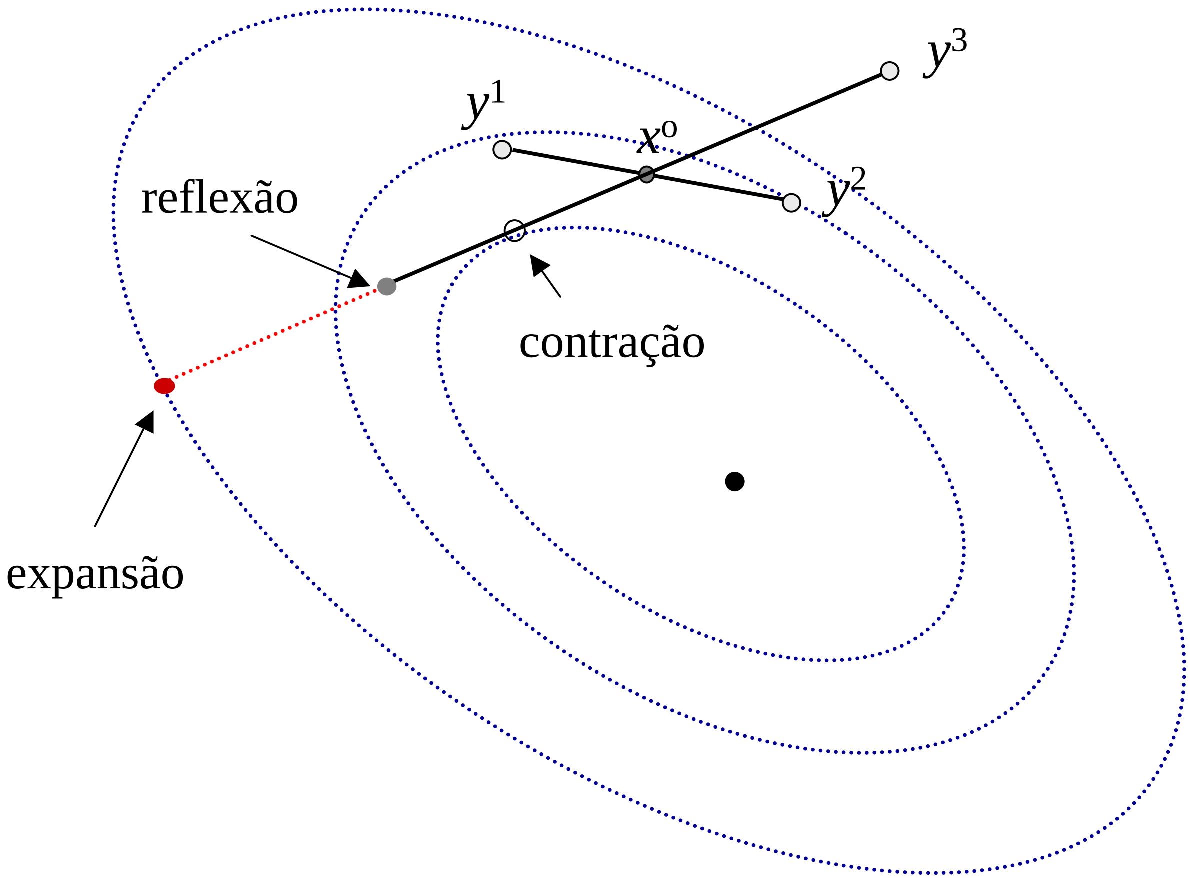
- Não utiliza derivadas

- Mantém (n + 1) soluções candidatas

- Baseia-se nos conceitos de:
  - reflexão
  - expansão
  - contração
  - encolhimento

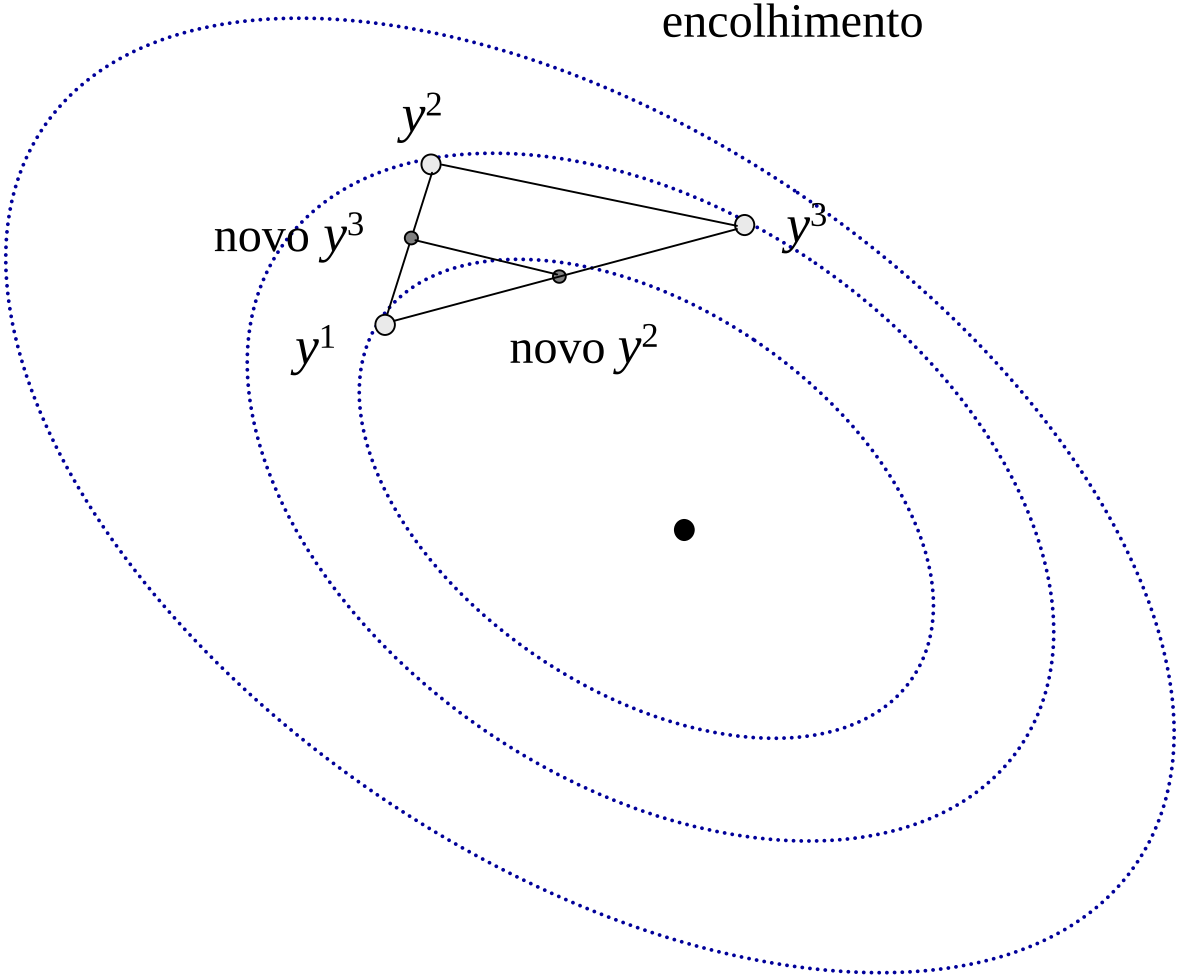
$\Delta x \equiv x^t - y^{n+1}$ $\quad$ $y^{n+1} \leftarrow$ pior solução entre as $(n+1)$ candidatas

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



y^3
y^1
x^o
y^2
reflexão
contração
expansão

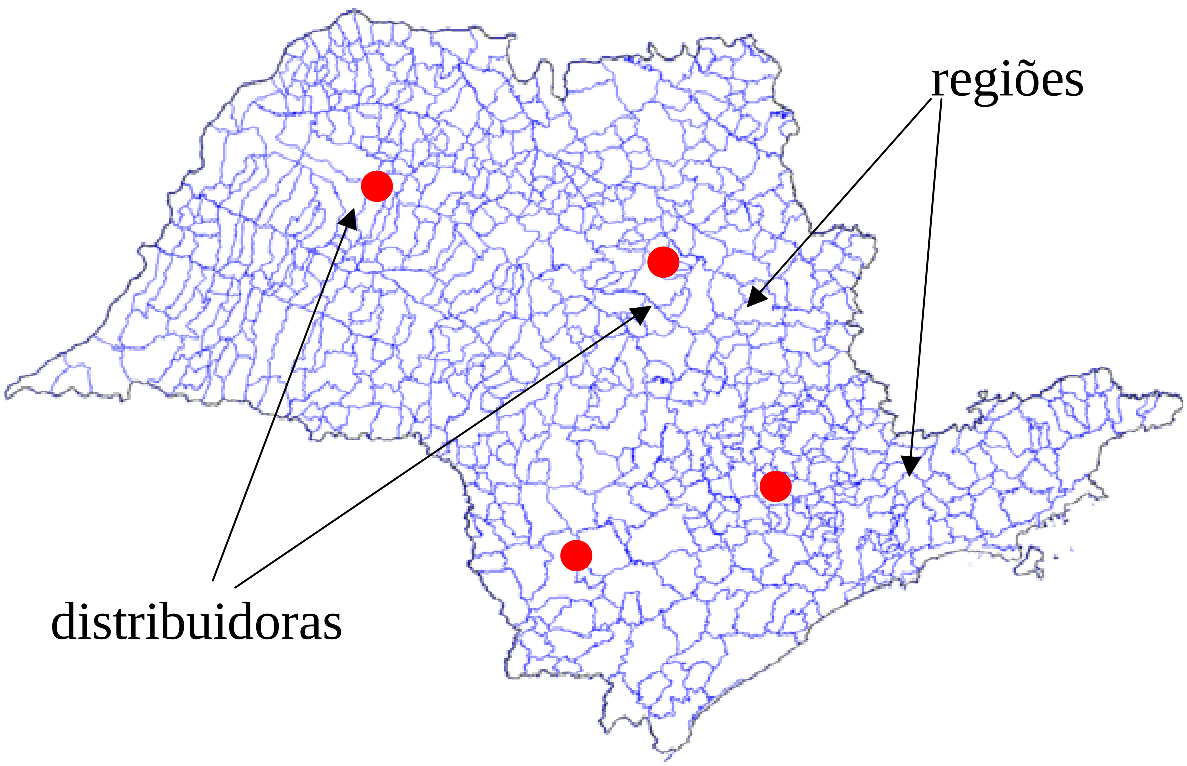
encolhimento



II- programação não linear restrita

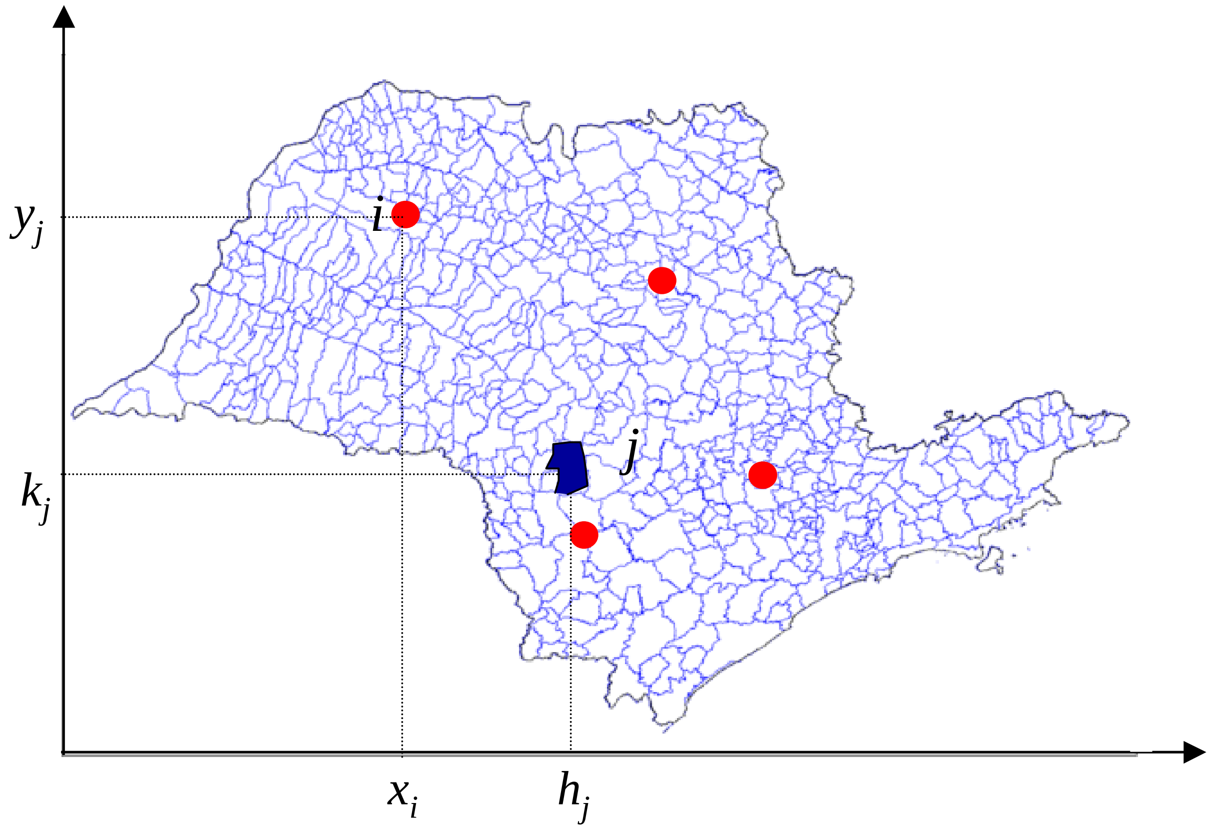
$$\begin{array}{ll} \min & f(x) \\ \text{sa} & x \in \mathrm{D} \subset \mathrm{R}^n \end{array}$$

Exemplo: localização e alocação



Distribuidores: 17
Regiões: 650 agrupando 24.000 clientes

i = distribuidor, i = 1,....,17
j = região, j = 1,...,650



d_j = número de viagens/ano para entregas à j-ésima região

- Minimizar custo de transporte
  - minimizar número de viagens (w_{ij}) de $i \rightarrow j$
  - minimizar distância entre i e j

- Modelo PNL

$$\min \sum_{i=1}^{17} \sum_{j=1}^{650} w_{ij} \sqrt{(x_i - h_j)^2 + (y_i - h_j)^2}$$

$$\text{sa} \quad \sum_{i=1}^{17} w_{ij} = d_j, \qquad j = 1, \ldots, 650$$

$$w_{ij} \geq 0, \; i = 1, \ldots, 17; \; j = 1, \ldots, 650$$

Modelos convexos

$$\begin{array}{ll} \min & f(x) \\ \text{sa} & g_i(x) \geq a_i,\ i = 1, \ldots, I \\ & h_j(x) \leq b_j,\ j = 1, \ldots, J \\ & l_k(x) = c_k,\ \ k = 1, \ldots, K \end{array}$$

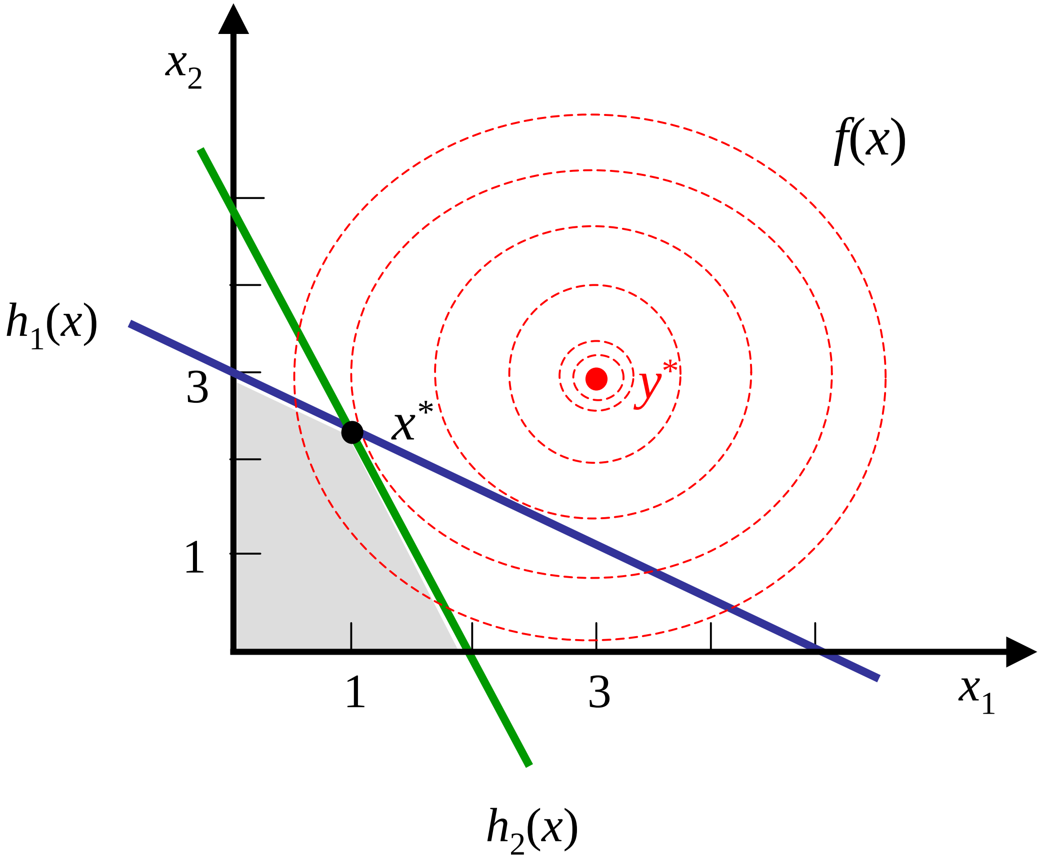
$f(x)$ convexa

$g_i(x)$ côncava $\forall i$

$h_j(x)$ convexa $\forall j$

$l_k(x)$ linear $\forall k$

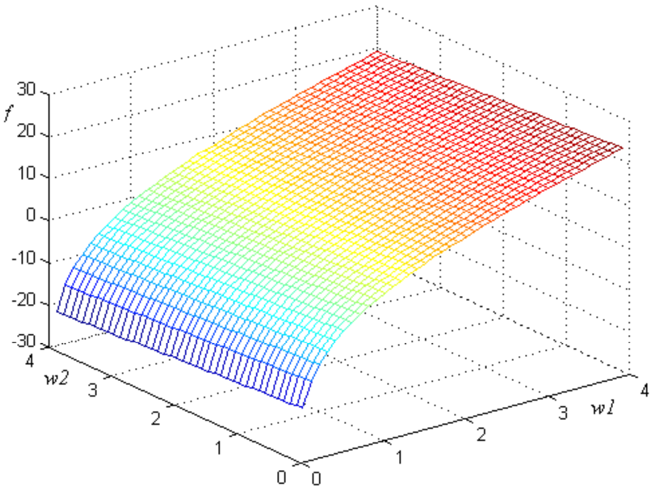
$$\begin{array}{ll} \min & (x_1 - 3)^2 + (x_2 - 3)^2 \\ \text{sa} & 3x_1 + 5x_2 \le 15 \\ & 5x_1 + 2x_2 \le 10 \\ & x_1, x_2 \ge 0 \end{array}$$

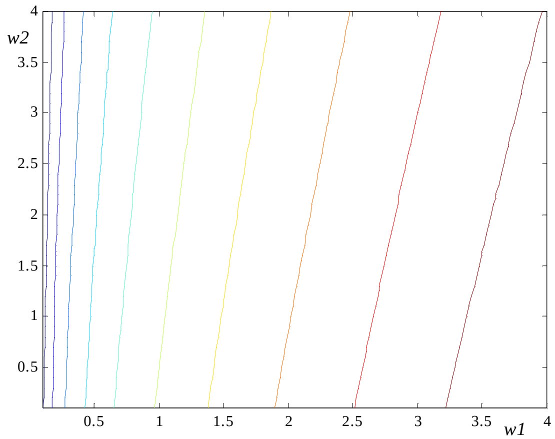


$$\begin{array}{ll} \max & 3w_1 - w_2 + 8\ln w_1 \\ \text{sa} & 4w_1^2 - w_1 w_2 + w_2^2 \leq 100 \\ & w_1 + w_2 = 4 \\ & w_1, w_2 \geq 0 \end{array}$$

- Função objetivo

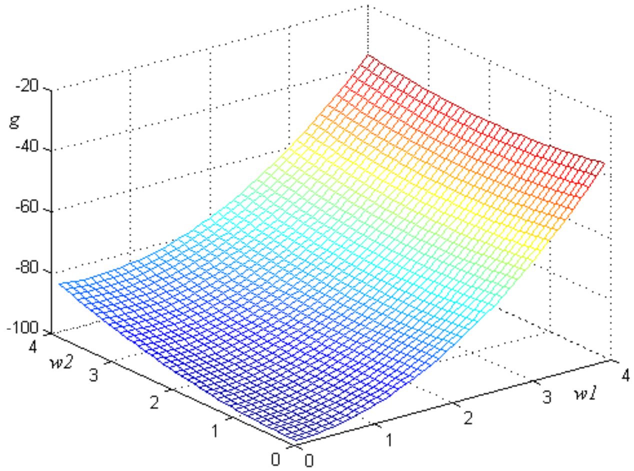
$$f(w_1, w_2) = 3w_1 - w_2 + 8\ln w_1$$

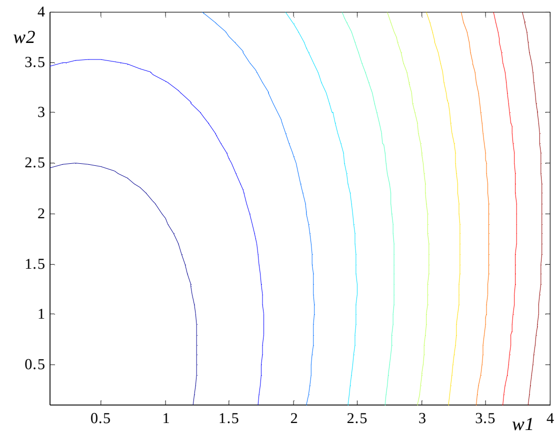




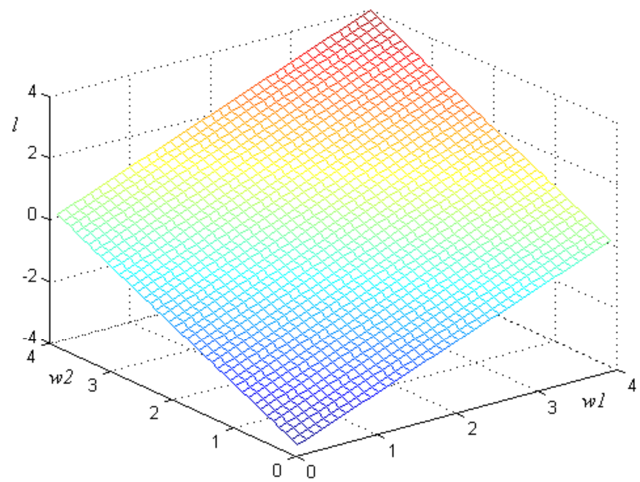
- Restrições principais

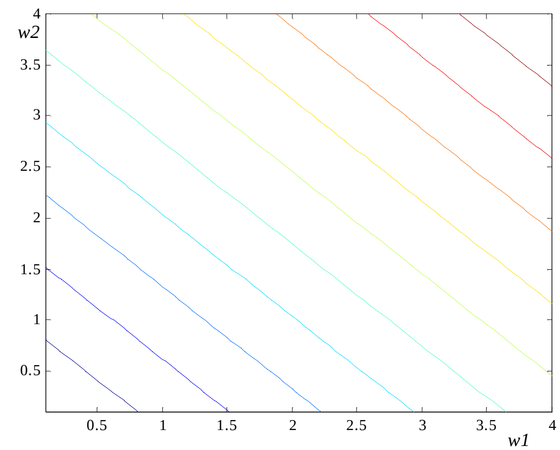
$$h(w_1, w_2) = 4\,w_1^2 - w_1\,w_2 + w_2^2 \quad \leq \quad 100$$



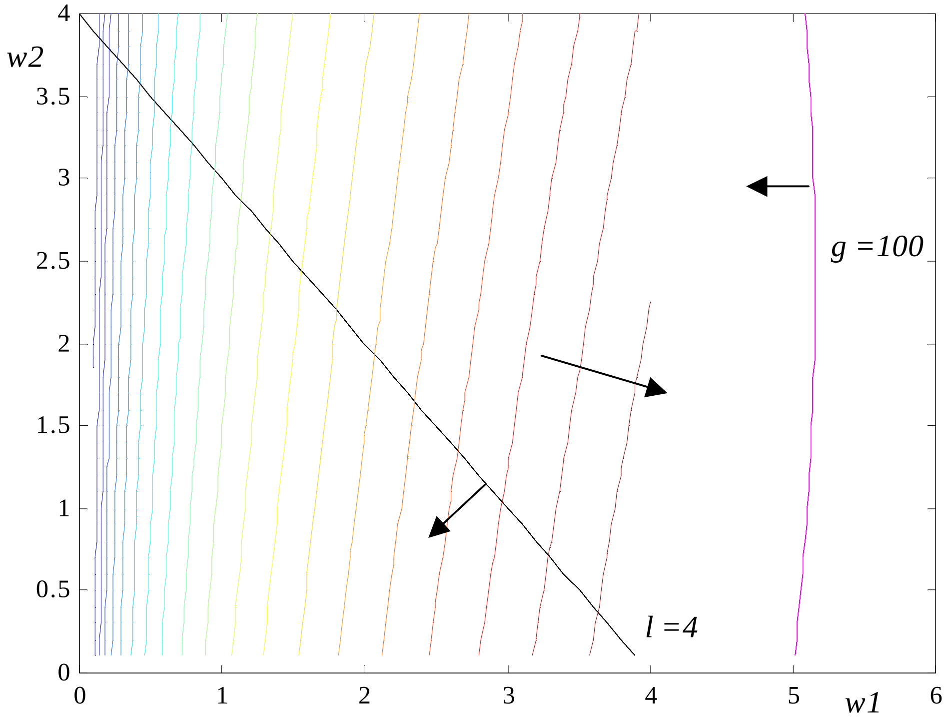


$$l(w_1, w_2) = w_1 + w_2 = 4$$





$$
\begin{array}{ll}
\max & 3w_1 - w_2 + 8\ln w_1 \\
\text{sa} & 4w_1^2 - w_1 w_2 + w_2^2 \le 100 \\
 & w_1 + w_2 = 4 \\
 & w_1, w_2 \ge 0
\end{array}
$$



Propriedade de modelos convexos

Se a função objetivo do modelo é convexa e as restrições produzem um conjunto convexo factível, então toda solução ótima local é uma solução ótima global.

$$\left.\begin{array}{ll} g_i(\mathrm{x}) \text{ côncava} & \forall i \\ h_j(\mathrm{x}) \text{ convexa} & \forall j \\ l_k(\mathrm{x}) \text{ linear} & \forall k \end{array}\right\} \Rightarrow \quad \text{Conjunto convexo}$$

$$f \text{ convexa (côncava)} \quad \Rightarrow \quad \text{Função unimodal}$$

Método dos multiplicadores de Lagrange

$$\min(\max)\ f(x)$$

$$\text{sa} \qquad g_i(x) = b_i \qquad \text{Restrições igualdade}$$

$$i = 1, \ldots, m$$

$$L(x, \nu) = f(x) + \sum_{i=1}^{m} \nu_i [b_i - g_i(x)] \qquad \text{Função Lagrangeana}$$

v_i = multiplicador de Lagrange associado à restrição *i*

Pontos estacionários da função Lagrangeana

(x^*, v^*) é um ponto estacionário de $L(x,v)$ se $\nabla L(x^*,v^*) = 0$:

$$\nabla f(x^*) = \sum_{i=1}^{m} \nu_i^* \nabla g_i(x^*)] \quad \text{ou} \quad \sum_{i=1}^{m} \frac{\partial g_i}{\partial x_j} v_i = \frac{\partial f}{\partial x_j}, \ \forall j$$

$$g_i(x^*) = b_i, \ \forall i$$

Se (x^*, v^*) é um ponto estacionário de $L(x,v)$ e x^* é uma solução ótima irrestrita de $L(x,v)$, então x^* é uma solução ótima do modelo original com restrições de igualdade.

Exemplo

$$
\begin{aligned}
\min \quad & 6x_1^2 + 4x_2^2 + x_3^2 \\
\text{sa} \quad & 24x_1 + 24x_2 = 360 \\
& x_3 = 1
\end{aligned}
$$

$$L(x,v) = 6x_1^2 + 4x_2^2 + x_3^2 + v_1[360 - 24x_1 - 24x_2] + v_2[1 - x_3]$$

$$\nabla f(x) = (12x_1, 8x_2, 2x_3)$$

$$\nabla g_1(x) = (24, 24, 0)$$

$$\nabla g_2(x) = (0, 0, 1)$$

$$
\begin{aligned}
24v_1 \qquad\qquad &= 12x_1 \\
24v_1 \qquad\qquad &= 8x_2 \\
1v_2 &= 2x_3 \\
24x_1 + 24x_2 &= 360 \\
1x_3 \qquad\qquad &= 1
\end{aligned}
$$

$$L(x,v) = 6x_1^2 + 4x_2^2 + x_3^2 - 72x_1 - 72x_2 - 2x_3 + 1082 \qquad \text{convexa}$$

$$\Downarrow$$

$(x_1^*, x_2^*, x_3^*, v_1^*, v_2^*) = (6, 9, 1, 3, 2) \Rightarrow$ mínimo de $L(x,v)$

Interpretação multiplicadores de Lagrange

- v_i = variação valor ótimo f quando b_i aumenta de 1 unidade
- v_i = variável dual

$$v_i^* = \frac{\partial L}{\partial b_i}$$

- Limitações
  - condições de estacionariedade difíceis de resolver
  - restrições de desigualdade aumenta a complexidade
  - soluções ótimas globais somente para modelos tratáveis

Caso geral: condições de Karush-Kuhn-Tucker

- Forma geral PNL diferenciáveis

$$\begin{array}{ll} \min(\max) & f(x) \\ \text{sa} & g_i(x) \geq b_i, \ \forall i \in G \\ & g_i(x) \leq b_i, \ \forall i \in L \\ & g_i(x) = b_i, \ \forall i \in E \end{array}$$

f, g_i : funções diferenciáveis

G, L, E : conjuntos de índices

Folgas complementares

$$v_i[b_i - g_i(x)] = 0$$ para todas desigualdades i

- Sinais dos multiplicadores

objetivo	i é $\leq$	i é $\geq$	i é $=$
min	$v_i \leq 0$	$v_i \geq 0$	irrestrito
max	$v_i \geq 0$	$v_i \leq 0$	irrestrito

Condições de Karush-Kuhn-Tucker

- Solução (x, v) satisfaz as condições de KKT se:
  - f, g_i são funções diferenciáveis
  - condições das folgas complementares
  - restrições no sinais dos multiplicadores
  - restrições primais
  - equação dos gradientes

$$\sum_i \nabla g_i(x) v_i = \nabla f(x)$$

- Ponto de KKT
  - qualquer x para o qual $\exists\, v$ que satisfaça as condições KKT

Característica da direção de busca Δx

$$f(x+\Delta x) \approx f(x) + \nabla f(x)\Delta x$$

objetivo	$\nabla f(x)\Delta x > 0$	$\nabla f(x)\Delta x < 0$
min	não melhora	melhora
max	melhora	não melhora

(14.23)

$\nabla f(\mathrm{x})\Delta \mathrm{x} = 0 \quad \Rightarrow$ necessitamos de mais informação

Factibilidade da direção de busca Δx

$$g_i(x+\Delta x) \approx g_i(x) + \nabla g_i(x)\Delta x$$

$$\nabla g_i(x)\Delta x \begin{cases} \geq 0 & \text{restrições} \geq \textit{ativas} \\ \leq 0 & \text{restrições} \leq \textit{ativas} \\ = 0 & \text{todas restrições} \; = \end{cases} \qquad (14.24)$$

Condições suficientes de KKT

- As condições de KKT fornecem um teste de primeira ordem para determinar a ausência de direções factíveis que melhoram o valor da função objetivo

- Isto é, x é um ponto de KKT se e somente se não existe uma direção Δx em x que satisfaz as condições de primeira ordem (14.23) e (14.24)

$$\Downarrow$$

- Condições KKT são suficientes para ótimo local, isto é:
  - se x é um ponto de KKT, então x é ótimo local

- Se modelo de PNL é convexo:
  - condições de KKT são suficientes para ótimo global

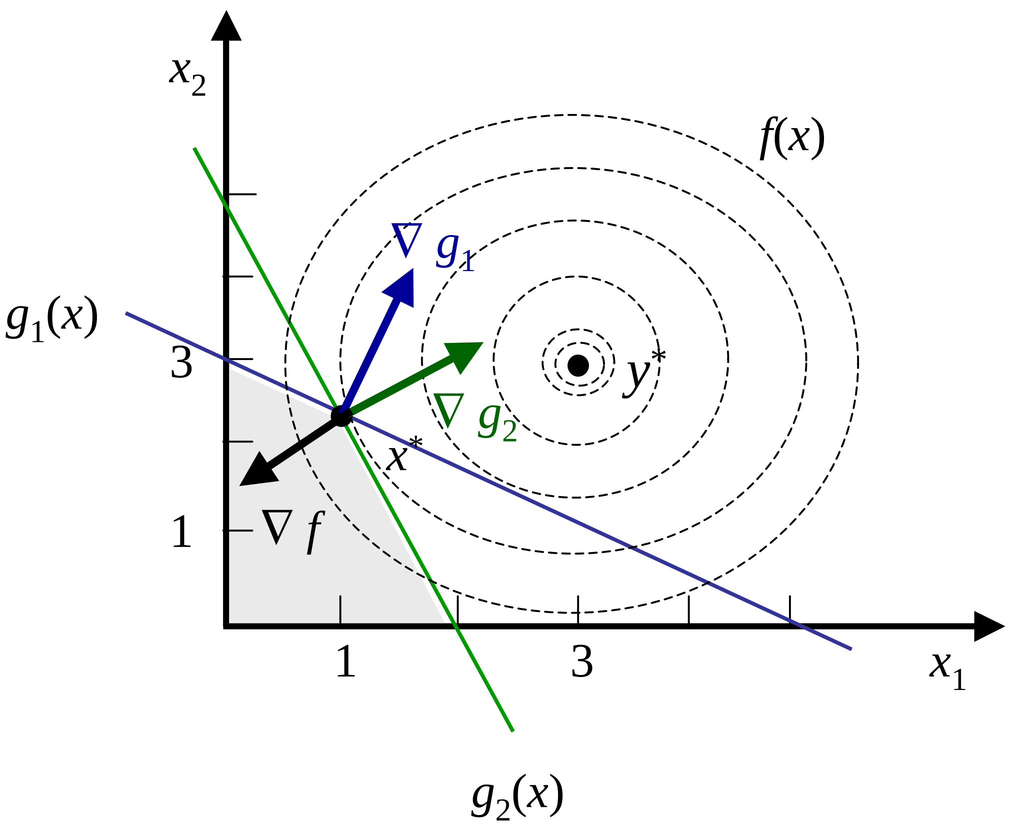
Condições necessárias de KKT

- Necessidade:
  - se *x* é uma solução ótima, então *x* satisfaz KKT

- Solução ótima local é um ponto de KKT se:
  - todas as restrições são lineares
  - gradientes das restrições ativas no ponto ótimo local são LI (qualificação de restrição)

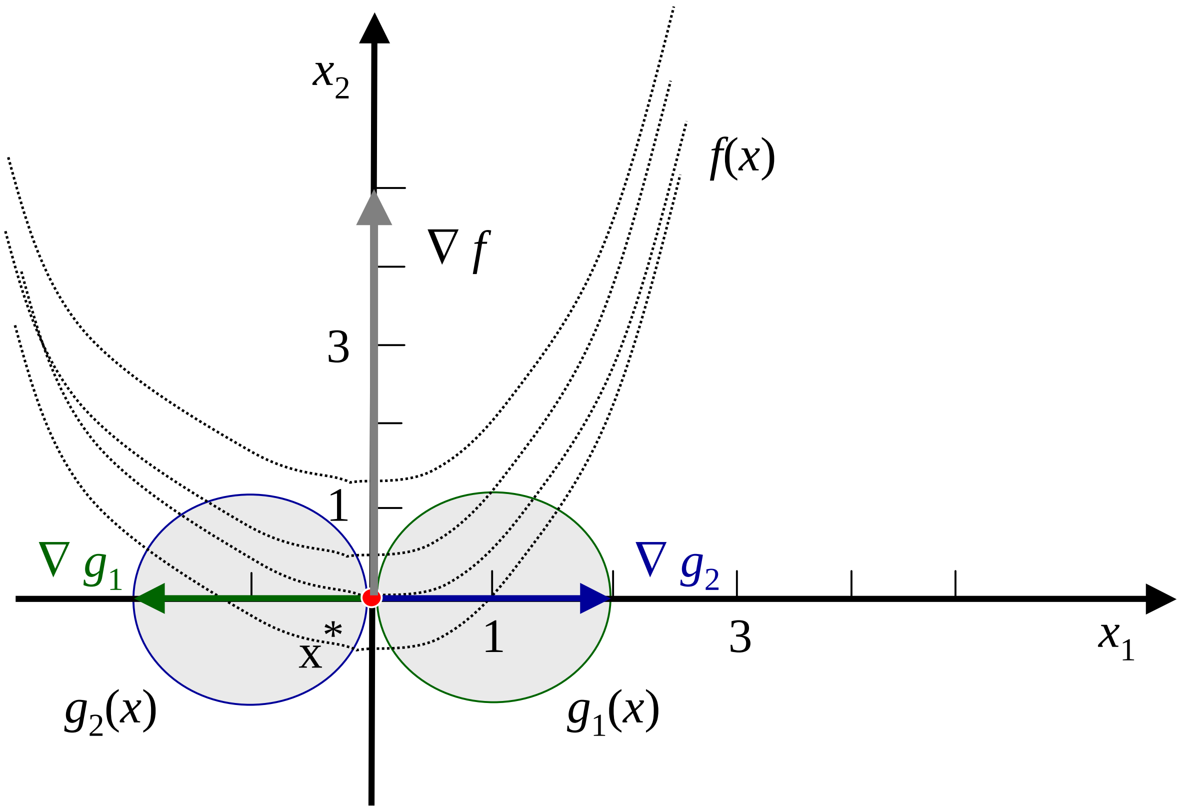
Exemplo

$$
\begin{array}{ll}
\min & (x_1 - 3)^2 + (x_2 - 3)^2 \\
\text{sa} & 3x_1 + 5x_2 \le 15 \\
 & 5x_1 + 2x_2 \le 10 \\
 & x_1, x_2 \ge 0
\end{array}
$$



Exemplo

$$\min \quad x_1^{\,2} + 4x_2$$
$$\text{sa} \quad (x_1 - 1)^2 + x_2^{\,2} \leq 1$$
$$(x_1 + 1)^2 + x_2^{\,2} \leq 1$$



Observação

Este material refere-se às notas de aula do curso CT 820 Teoria de Sistemas e Otimização Fuzzy: Introdução e Aplicações da Faculdade de Engenharia Elétrica e de Computação da Unicamp e do Centro Federal de Educação Tecnológica do Estado de Minas Gerais. 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.