Otimização não linear
Otimização não diferenciável: Método Nelder-Mead

Ana Maria A. C. Rocha

Departamento de Produção e Sistemas

Universidade do Minho

arocha@dps.uminho.pt

Métodos numéricos de resolução

problema sem restrições

$$\min_{x \in \mathbb{R}^n} f(x) \quad (n > 1)$$

- Métodos de procura direta (que **não** usam derivadas);
- Métodos do gradiente.

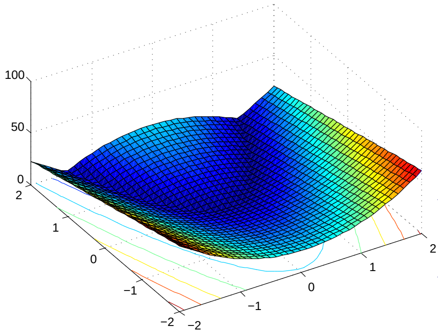
Métodos de procura direta:

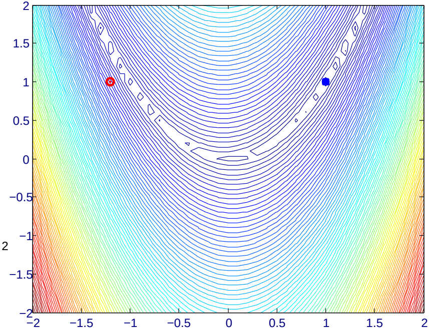
- só usam informação da função objetivo f;
- são apropriados para **problemas não diferenciáveis** (embora possam ser usados em problemas diferenciáveis);
- exemplo: método de Nelder-Mead (destina-se a problemas de otimização multidimensionais).

Problemas sem restrições, não diferenciáveis

Formulação geral do problema sem restrições: $\min_{x \in \mathbb{R}^n} f(x)$

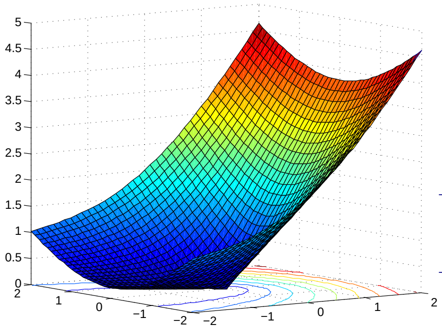
Exemplo não diferenciável: $\min_{x \in \mathbb{R}^2} f(x) \equiv |x_1 - 1| + 10|x_2 - x_1^2|$

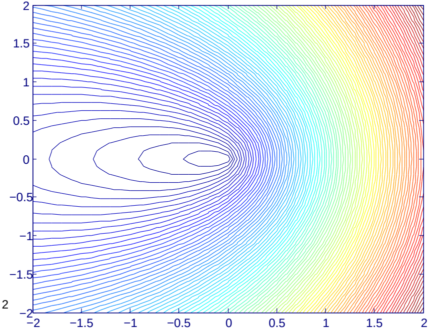




Problemas sem restrições, não diferenciáveis (cont.)

$$\min_{x \in \mathbb{R}^2} f(x) \equiv \begin{cases} 3(2|x_2| - x_1) + (0.9 + \sqrt{5}/2)x_1 & \text{se } \; x_1 > 2|x_2| \\ 0.9x_1 + \sqrt{x_1^2 + x_2^2} & \text{caso contrário} \end{cases}$$

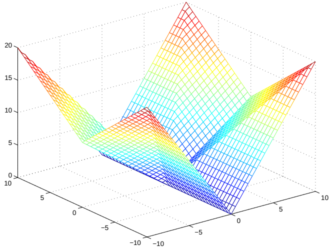




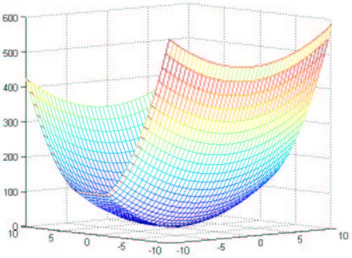
Outras funções não diferenciáveis

$$f(x) = |x| = \begin{cases} -x & \text{para } x < 0 \\ x & \text{para } x < 0 \end{cases} \qquad \text{em} \quad \mathbb{R}$$

$f(x) = \sum_{i=1}^{2} \min\{|x_i|, |x_1|\}$



$f(x) = \max\{(x_1 - 1)^2, x_1^2 + 4(x_2 - 1)^2\}$



Método do simplex de Nelder-Mead

O método de Nelder-Mead (NM)

- não usa derivadas da função f
- é um método iterativo
- define em cada iteração um **simplex** (que é um poliedro em $\mathbb{R}^n$).

Em $\mathbb{R}^n$, sejam

$$x_1, \; x_2, \; \ldots, \; x_n, \; x_{n+1}, \quad \text{sendo cada } \; x_i \in \mathbb{R}^n$$

os vértices do simplex de dimensão n (são $n+1$ vértices).

Por exemplo,

em $\mathbb{R}^2$, o simplex formado pelos $n+1(=3)$ pontos define um triângulo.
Em $\mathbb{R}^3$, um simplex é um tetraedro.

Nota: Um simplex diz-se regular se as suas arestas são iguais. Em $\mathbb{R}^2$, um simplex regular é um triângulo equilátero.

Método de Nelder-Mead

Notação:

$$S_k = \langle X_1, X_2, \ldots, X_n, X_{n+1} \rangle$$

representa o simplex, da iteração k, em que os vértices já estão ordenados por ordem crescente dos valores da função objetivo do problema, isto é,

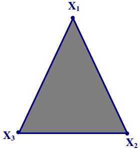
$$f(X_1) \leq f(X_2) \leq \cdots \leq f(X_n) \leq f(X_{n+1})$$

sendo

- X_1 - o melhor vértice
- X_n - o segundo pior vértice
- X_{n+1} - o pior vértice

Método de Nelder-Mead

Em $\mathbb{R}^2$, o simplex é um triângulo



Em cada iteração, definem-se pontos auxiliares - candidatos a vértices de um novo simplex - que serão aceites ou rejeitados comparando apenas os seus valores de f com

- $f(X_1)$ $f(X_n)$ $f(X_{n+1})$

Lista dos pontos auxiliares:

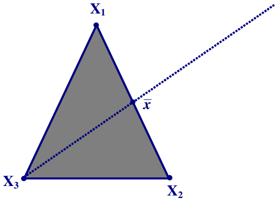
- vértice refletido;
- vértice expandido;
- vértices contraídos . . .;
- vértices de um simplex encolhido.

Método de Nelder-Mead

Seja S_1 o simplex inicial já ordenado ($k = 1$)

$$S_1 = \langle X_1, X_2, \ldots, X_{n+1} \rangle$$

Em cada iteração, começa-se por calcular o **centróide** do simplex, que é o ponto médio do hiperplano definido por $X_1, X_2, \ldots, X_n$ (n melhores vértices do simplex)

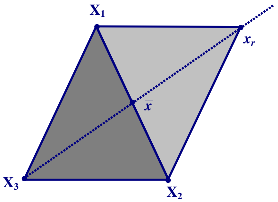


$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

Método de Nelder-Mead

De seguida, calcula-se o **vértice refletido** (com $\delta = 2$)

$$\boxed{x_r = X_{n+1} + \delta\,(\bar{x} - X_{n+1})}$$

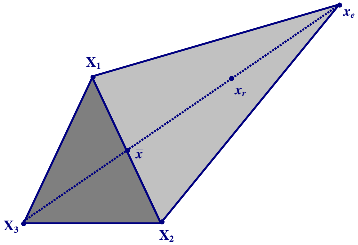


CASO 1: Se x_r for bom $(f(X_1) \leq f(x_r) < f(X_n))$ aceita-se x_r e $S_{k+1} = \langle X_1, X_2, \ldots, X_n, x_r \rangle$ é o simplex para a iteração seguinte.

Método de Nelder-Mead

CASO 2: Se x_r for muito bom ($f(x_r) < f(X_1)$) faz-se uma expansão do simplex:
- cálculo do **vértice expandido** (com $\delta = 3$)

$$\boxed{x_e = X_{n+1} + \delta\,(\bar{x} - X_{n+1})}$$

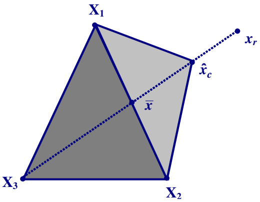


- Se x_e for muito bom ($f(x_e) < f(X_1)$) aceita-se x_e e $S_{k+1} = \langle X_1, X_2, x_e \rangle$
- Senão aceita-se x_r e $S_{k+1} = \langle X_1, X_2, x_r \rangle$

Método de Nelder-Mead

CASO 3: Se x_r for fraco $(f(X_n) \leq f(x_r) < f(X_{n+1}))$ faz-se uma contracção para o exterior:
- cálculo do **vértice contraído para o exterior** (com $\delta = 1.5$)

$$\boxed{\hat{x}_c = X_{n+1} + \delta\,(\bar{x} - X_{n+1})}$$

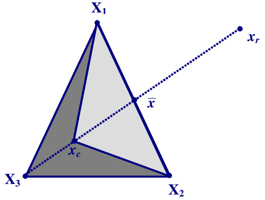


- Se $\hat{x}_c$ for bom $(f(\hat{x}_c) < f(X_n))$ aceita-se $\hat{x}_c$ e $S_{k+1} = \langle X_1, X_2, \ldots, X_n, \hat{x}_c \rangle$
- Senão encolhe-se o simplex $(\ldots)$

Método de Nelder-Mead

CASO 4: Se x_r for muito fraco ($f(x_r) \geq f(X_{n+1})$) faz-se uma **contracção** para o interior:
- cálculo do **vértice contraído para o interior** (com $\delta = 0.5$)

$$\boxed{x_c = X_{n+1} + \delta\,(\bar{x} - X_{n+1})}$$



- Se x_c for bom ($f(x_c) < f(X_n)$) aceita-se x_c e $S_{k+1} = \langle X_1, X_2, \ldots, X_n, x_c \rangle$
- Senão encolhe-se o simplex (...)

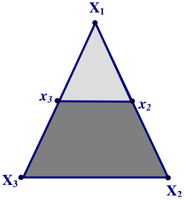
Método de Nelder-Mead

Encolher o simplex

Consiste em substituir cada um dos vértices X_i, $i = 2, ..., n+1$, pelo ponto médio do segmento que une esse X_i a X_1, i.e.,

$$\boxed{x_i = \frac{X_i + X_1}{2}}$$

e $S_{k+1} = \langle X_1, x_2, \ldots, x_n, x_{n+1} \rangle$.



Critério de paragem do NM

Nota:

Calculado o simplex para a iteração seguinte, é necessário ordenar o simplex para verificar o critério de paragem.

Critério de paragem

Consiste em verificar se o tamanho relativo do simplex já é inferior ou igual a uma quantidade pequena, $\varepsilon > 0(\approx 0)$, i.e.,

- **se**

$$\boxed{\frac{\max\limits_{2\leq i\leq n+1} \|X_i - X_1\|_2}{\max\{1, \|X_1\|_2\}} \leq \varepsilon,}$$

- **então** o processo iterativo termina (o vértice do simplex com menor valor da função objetivo, X_1, é considerado como a `melhor aproximação` calculada à solução);
- **senão**, o processo iterativo continua.

Exercício

Calcule o máximo da seguinte função não diferenciável

$$f(x_1, x_2) = -|x_1 x_2| - x_2^2$$

usando o método de Nelder-Mead. Inicie o processo iterativo com o seguinte simplex:

$$\left\langle \begin{pmatrix} -1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ -1 \end{pmatrix} \right\rangle$$

Para a paragem do processo iterativo use $\varepsilon = 1.2$ ou $n_{max} = 2$.

Resolução do Exercício

$$\min f(x_1,x_2) = |x_1 x_2| + x_2^2$$

$$\begin{cases} x_1 = (-1,1)^T \\ f(x_1) = 2 \end{cases} \qquad \begin{cases} x_2 = (1,0)^T \\ f(x_2) = 0 \end{cases} \qquad \begin{cases} x_3 = (-1,-1)^T \\ f(x_3) = 2 \end{cases}$$

1ª iteração $S_1 = \left\langle \underset{0}{\overset{X_1}{\begin{pmatrix} 1 \\ 0 \end{pmatrix}}}, \underset{2}{\overset{X_2}{\begin{pmatrix} -1 \\ 1 \end{pmatrix}}}, \underset{2}{\overset{X_3}{\begin{pmatrix} -1 \\ -1 \end{pmatrix}}} \right\rangle$

$\bar{x} = \frac{1}{2}(X_1 + X_2) = (0, 0.5)^T$ $\begin{cases} x_r = 2\bar{x} - X_3 = (1,2)^T \\ f(x_r) = 6 \end{cases}$

$f(x_r) \geq f(X_3) \Rightarrow x_r$ é muito fraco $\Rightarrow$ calcular x_c (contraído para o interior)

$$\begin{cases} x_c = 0.5X_3 + (1-0.5)\bar{x} = (-0.5, -0.25)^T \\ f(x_c) = 0.1875 \end{cases}$$

$f(x_c) < f(X_2) \Rightarrow x_c$ é bom $\Rightarrow$ aceitar x_c

Resolução do Exercício (cont.)

Novo simplex (já ordenado)

$$S_2 = \left\langle \begin{array}{ccc} X_1 & X_2 & X_3 \\ \left(\begin{array}{c} 1 \\ 0 \end{array}\right), & \left(\begin{array}{c} -0.5 \\ -0.25 \end{array}\right), & \left(\begin{array}{c} -1 \\ 1 \end{array}\right) \\ 0 & 0.1875 & 2 \end{array} \right\rangle$$

Testar critério de paragem

$\Delta = \max(1, \|X_1\|_2) = \max(1, 1) = 1$

$$\begin{aligned} \frac{1}{\Delta}\max(\|X_2 - X_1\|_2, \|X_3 - X_1\|_2) &= \frac{1}{1}\max(1.5207, 2.2361) \\ &= 2.2361 \leq 1.2 \text{ Falso, nova iteração} \end{aligned}$$

2^a iteração

$\bar{x} = \frac{(X_1+X_2)}{2} = (0.25, -0.125)^T$ $\quad \begin{cases} x_r = 2\bar{x} - X_3 = (1.5, -1, 25)^T \\ f(x_r) = 3.4375 \end{cases}$

$f(x_r) \geq f(X_3) \Rightarrow x_r$ é muito fraco $\Rightarrow$ calcular x_c (contraído para o interior)

Resolução do Exercício (cont.)

$$\begin{cases} x_c = 0.5X_3 + (1-0.5)\,\bar{x} = (-0.3750, 0.4375)^T \\ f(x_c) = 0.3555 \end{cases}$$

Como $f(x_c) \geq f(X_2) \Rightarrow$ encolher o simplex

$$x_2 = \frac{(X_1+X_2)}{2} = (0.25, -0.125)^T \quad \text{e} \quad f(x_2) = 0.0469$$
$$x_3 = \frac{(X_1+X_3)}{2} = (0, 0.5)^T \quad \text{e} \quad f(x_3) = 0.25$$

Novo simplex (já ordenado)

$$S_3 = \left\langle \underset{0}{\overset{X_1}{\begin{pmatrix} 1 \\ 0 \end{pmatrix}}}, \underset{0.0469}{\overset{X_2}{\begin{pmatrix} 0.25 \\ -0.125 \end{pmatrix}}}, \underset{0.25}{\overset{X_3}{\begin{pmatrix} 0 \\ 0.5 \end{pmatrix}}} \right\rangle$$

Testar critério de paragem:

$$\Delta = \max(1, \|X_1\|_2) = \max(1,1) = 1$$
$$\frac{1}{\Delta}\max(\|X_2 - X_1\|_2, \|X_3 - X_1\|_2) = \frac{1}{1}\max(0.7603, 1.118) = 1.118 \leq 1.2$$

Solução $\begin{cases} x^* \leftarrow (1,0)^T \\ f(x^*) \leftarrow 0 \end{cases}$