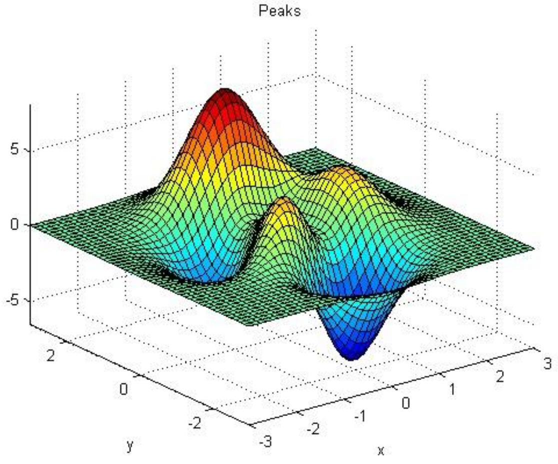
Aula 3 – Métodos de Otimização

André E. Lazzaretti

UTFPR/CPGEI



UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ

PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA E INFORMÁTICA INDUSTRIAL

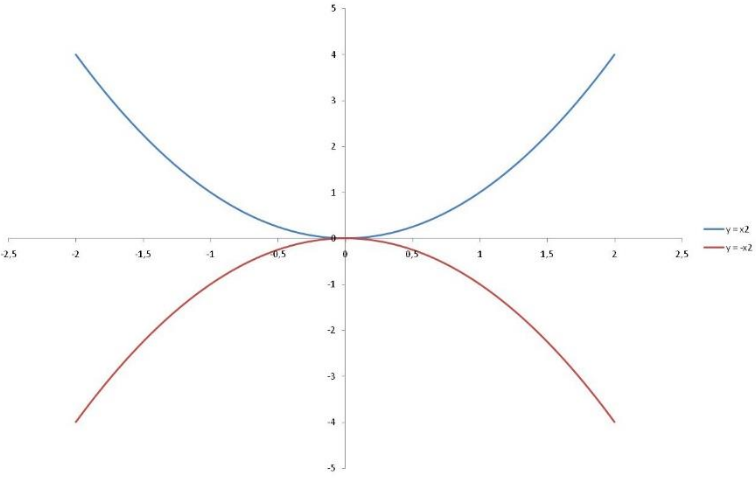
Otimização

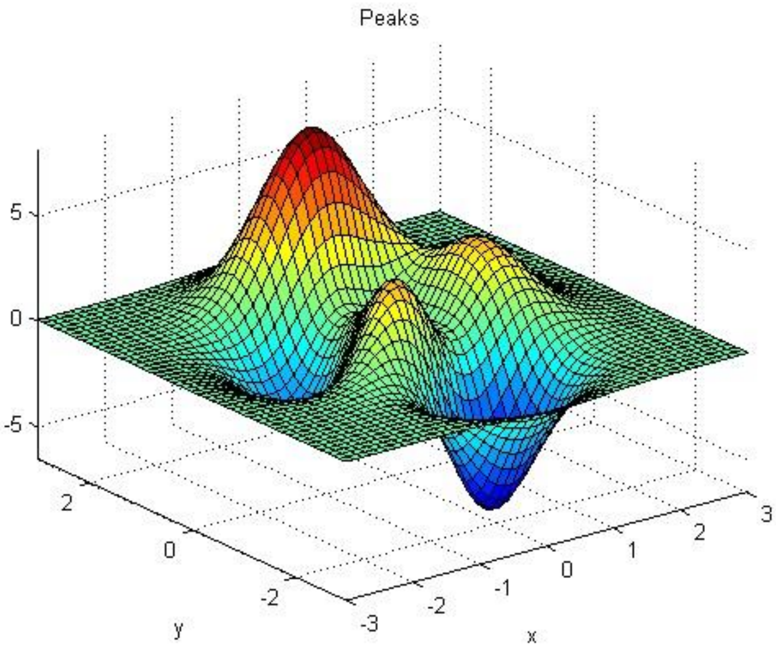
- Por que otimização?
  - Em reconhecimento de padrões, frequentemente busca-se otimizar (maximizar ou minimizar) uma determinada função matemática (função custo, função de perda, função objetivo).
- Por exemplo:
  - Log-likelihood (da aula passada) → maximizar
  - Erro de Classificação → minimizar
- Estrutura geral: monta-se a função e realiza-se a otimização
- Otimização (em geral não-linear):
  - Sem restrições
  - Com restrições de igualdade
  - Com restrições de igualdade e desigualdade

Otimização sem restrições

$$\min_{\mathbf{x}} g(\mathbf{x}) = \max_{\mathbf{x}} \left[-g(\mathbf{x}) \right]$$

- $g(\mathbf{x}) \rightarrow$ função objetivo
- $\mathbf{x} = [x_1, x_2, \ldots, x_n] \rightarrow$ variáveis a serem otimizadas

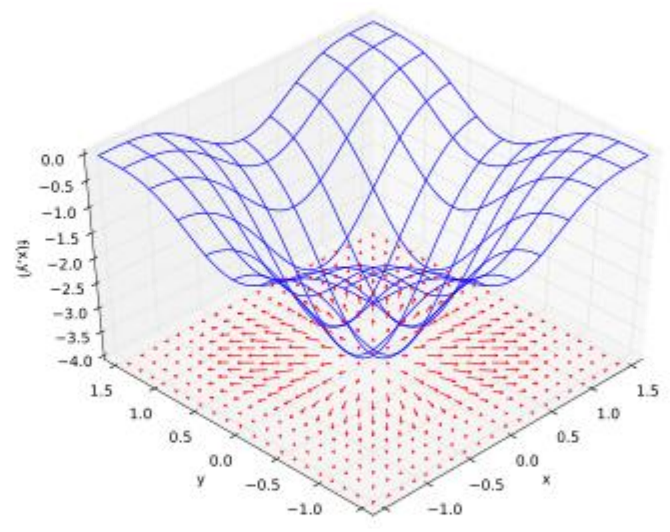




Otimização sem restrições

$$\nabla g(\mathbf{x}) = \left[\frac{\partial g(\mathbf{x})}{\partial x_1}, \frac{\partial g(\mathbf{x})}{\partial x_2}, \ldots, \frac{\partial g(\mathbf{x})}{\partial x_n} \right]^t$$

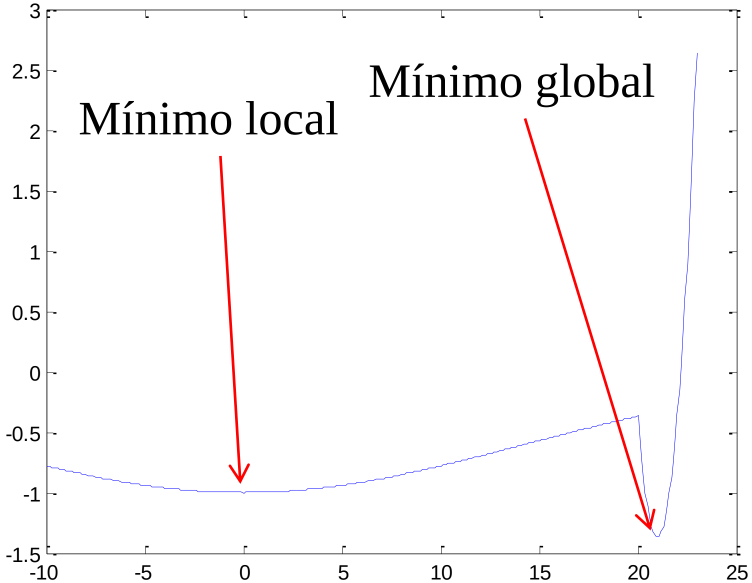
- $\nabla g(\mathbf{x}) \rightarrow$ gradiente de $g(\mathbf{x})$
  - Fornece a direção de maior inclinação (variação) de $g(\mathbf{x})$
  - Direção inversa de $\nabla g(\mathbf{x})$: Direção de máximo decrescimento

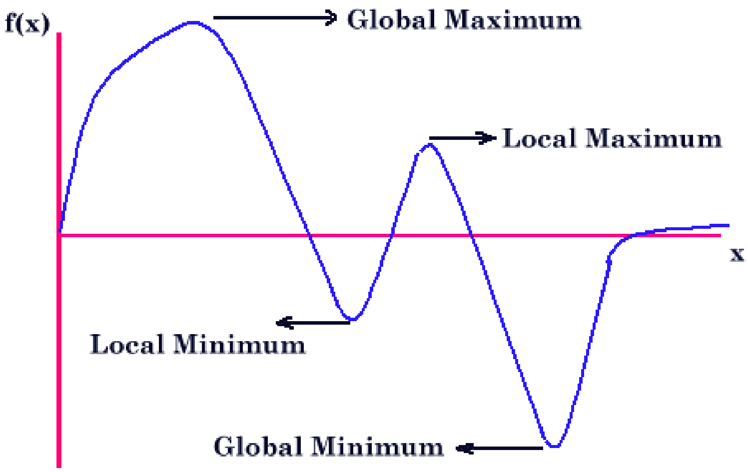


Otimização sem restrições

- Pontos de mínimo
  - Mínimo global
    - $\mathbf{x}^*$ é um mínimo global se

$$g\left(\mathbf{x}^*\right) \leq g\left(\mathbf{x}\right), \forall \mathbf{x} \in \Re^n$$



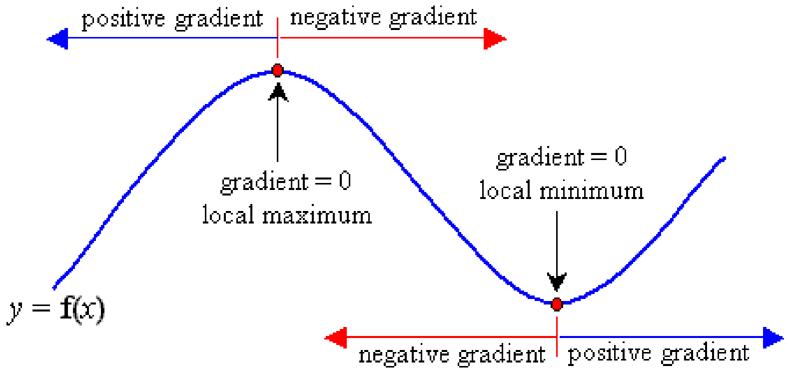


Otimização sem restrições

- Mínimo local: $g\left(\mathbf{x}^*\right) \le g\left(\mathbf{x}\right), \forall\, \mathbf{x} \in \Re^n \left\| \mathbf{x}^* - \mathbf{x} \right\| < \varepsilon$

- Condições de otimalidade:
  - **Condição necessária**: se $g(\mathbf{x})$ for continuamente diferenciável (condição de primeira ordem):

$$\nabla g\left(\mathbf{x}\right)\Big|_{\mathbf{x}=\mathbf{x}^*} = \begin{bmatrix} \partial g\left(\mathbf{x}\right)/\partial x_1 \\ \vdots \\ \partial g\left(\mathbf{x}\right)/\partial x_n \end{bmatrix}\Bigg|_{\mathbf{x}=\mathbf{x}^*} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}$$



Otimização sem restrições

- **Condição suficiente**: se cada elemento de $\nabla^2 g(\mathbf{x})$ for uma função contínua, $\nabla^2 g(\mathbf{x})$ calculada para $\mathbf{x} = \mathbf{x}^*$ deve ser positiva definida para que $\mathbf{x}^*$ seja um ponto de mínimo local (Condição de segunda ordem):

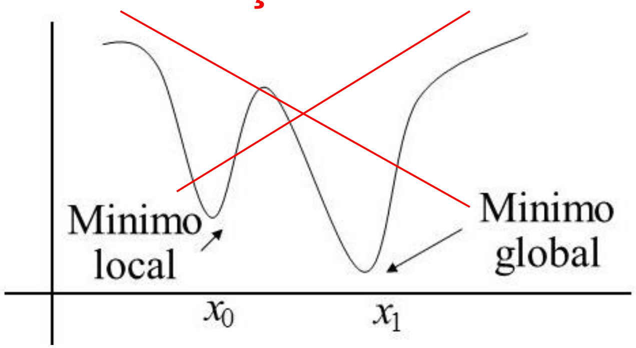
$$\nabla^2 g(\mathbf{x}) = \mathbf{H}(\mathbf{x}) = \begin{bmatrix} \partial^2 g(\mathbf{x}) / \partial x_1^2 & \cdots & \partial^2 g(\mathbf{x}) / \partial x_1 \partial x_n \\ \vdots & \ddots & \vdots \\ \partial^2 g(\mathbf{x}) / \partial x_1 \partial x_n & \cdots & \partial^2 g(\mathbf{x}) / \partial x_n^2 \end{bmatrix}$$

- Uma matriz $\mathbf{\Omega}$ é positiva definida se todos os autovalores λ_i, $i = 1, 2, \ldots, n$, de $\mathbf{\Omega}$ são positivos.
- Gradiente nulo: ponto de sela, máximo ou mínimo.
  - Negativa definida – ponto de máximo
  - Autovalores positivos e negativos: ponto de sela

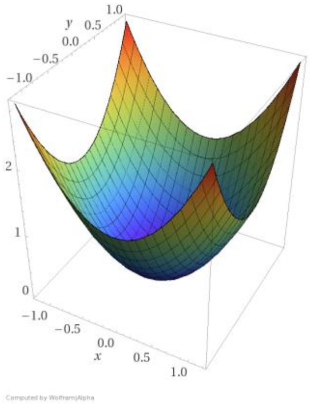
Otimização sem restrições

- Condições necessária e suficiente para que $\mathbf{x}^*$ seja um ponto de mínimo local
  1. $\nabla g(\mathbf{x})\big|_{\mathbf{x}=\mathbf{x}^*} = \mathbf{0}$
  2. $\mathbf{H}(\mathbf{x})$ calculada para $\mathbf{x} = \mathbf{x}^*$ deve ser positiva definida
- Se $g(\mathbf{x})$ for convexa e $\mathbf{x}^*$ satisfaz a condição de primeira ordem, então $\mathbf{x}^*$ é um ponto de mínimo global

Função não-convexa



Função convexa: todos os mínimos locais são mínimos globais(*).



Como Resolver Otimização sem Restrições

- Na prática utiliza-se: $\nabla g(\mathbf{x})=\mathbf{0}$

- Essa solução é viável quando $\nabla g(\mathbf{x})=\mathbf{0}$ forma um sistema linear de equações;

- Caso isso não seja possível (sistema não-linear de equações) → alternativas;

- Alternativa - Métodos iterativos:
  - Descida em gradiente;
  - Método de Newton.

Otimização sem restrições

- Descida em gradiente
  - Método de otimização numérica para minimização de $g(\mathbf{x})$
  - A partir de um estado inicial $\mathbf{x}^{0}$, busca novos estados na direção de máximo decrescimento da função:
    - Direção inversa ao gradiente

$$\mathbf{x}^{k} = \mathbf{x}^{k-1} - \alpha \nabla g(\mathbf{x})\Big|_{\mathbf{x}=\mathbf{x}^{k-1}}$$

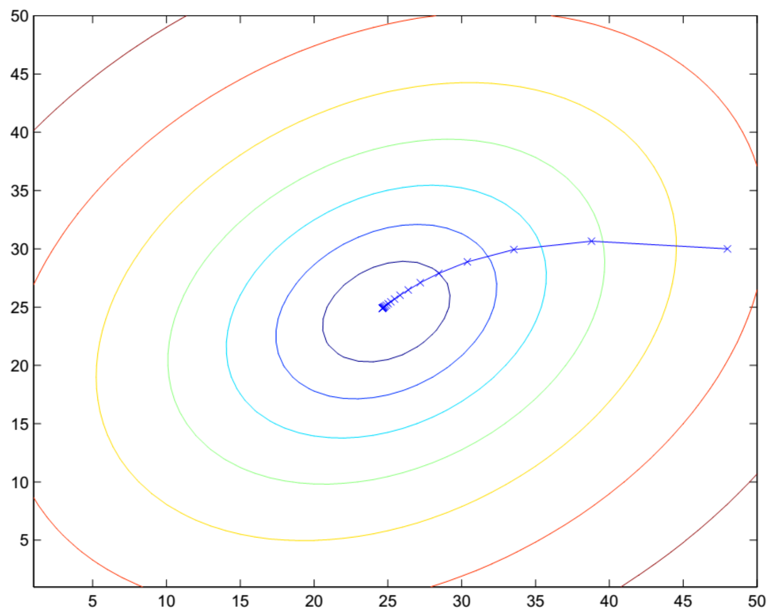
    - Subida em gradiente

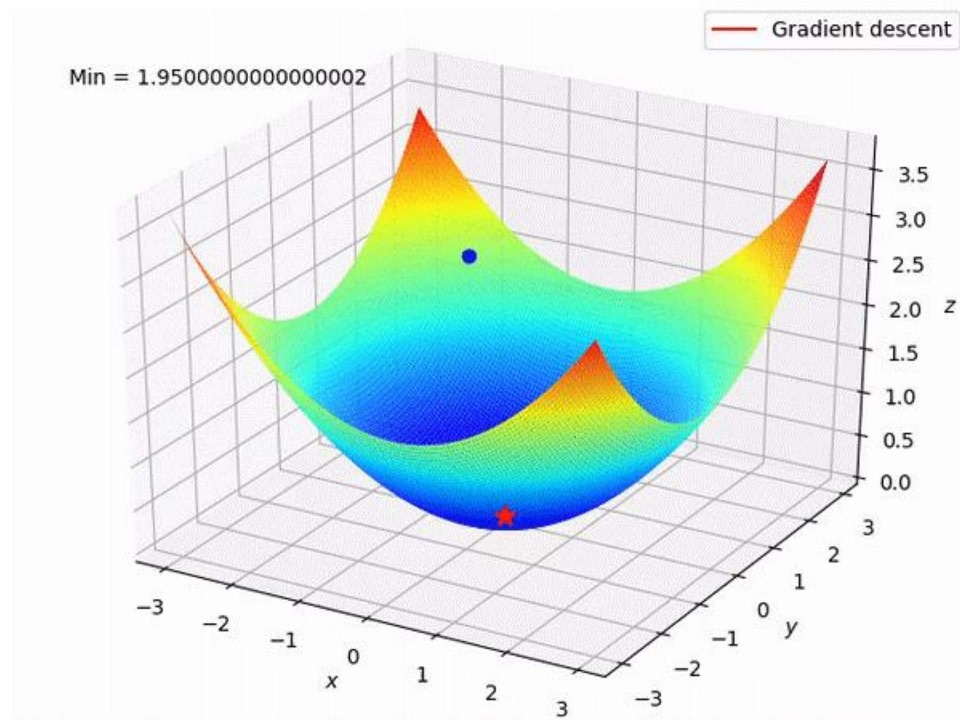
$$\mathbf{x}^{k} = \mathbf{x}^{k-1} + \alpha \nabla g(\mathbf{x})\Big|_{\mathbf{x}=\mathbf{x}^{k-1}}$$

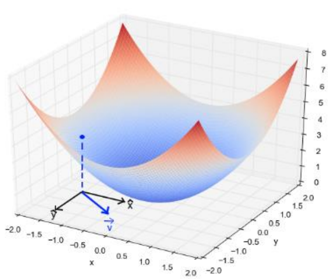
Descida em Gradiente

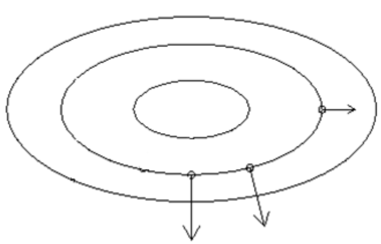
- Ideia Geral:

$$\mathbf{x}^{k} = \mathbf{x}^{k-1} - \alpha \nabla g\left(\mathbf{x}\right)\Big|_{\mathbf{x}=\mathbf{x}^{k-1}}$$







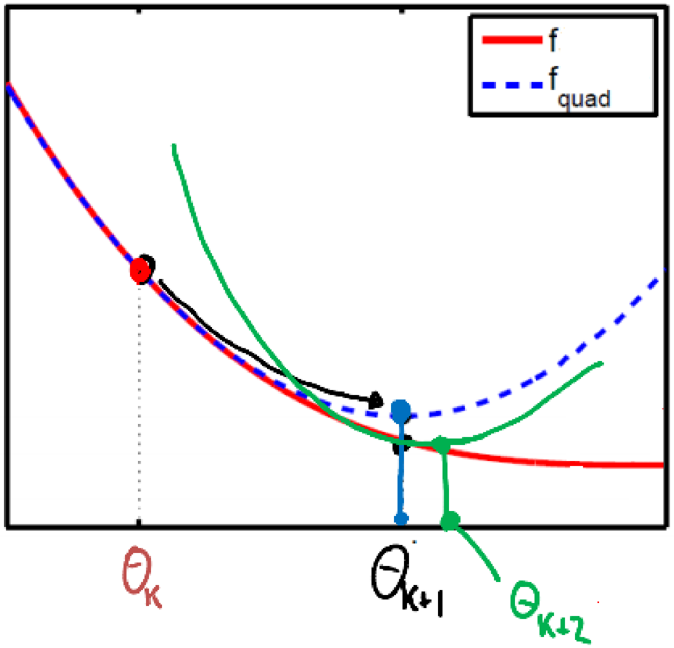


Método de Newton

- Altera-se o passo do algoritmo:

$$\mathbf{x}^k = \mathbf{x}^{k-1} - \left[\mathbf{H}(\mathbf{x})\Big|_{\mathbf{x}=\mathbf{x}^{k-1}} \right]^{-1} \nabla \mathbf{g}(\mathbf{x})\Big|_{\mathbf{x}=\mathbf{x}^{k-1}} \quad \text{ou} \quad \boldsymbol{\theta}_{k+1} = \boldsymbol{\theta}_k - \mathbf{H}_k^{-1}\mathbf{g}_k$$

- Esse método é derivado assumindo uma aproximação de segunda ordem (através da série de Taylor) da função no ponto em questão:



- **Exercício – verificar!**

Exemplo – fminunc.m

Nonlinear programming solver.

Finds the minimum of a problem specified by

$$\min_x f(x)$$

where $f(x)$ is a function that returns a scalar.

x is a vector or a matrix; see Matrix Arguments.

Syntax

```
x = fminunc(fun,x0)
x = fminunc(fun,x0,options)
```

Outra opção:

minimize.m - escrita por Carl E. Rasmussen (University of Cambridge)

http://learning.eng.cam.ac.uk/carl/code/minimize/

Minimize the function $f(x) = 3x_1^2 + 2x_1x_2 + x_2^2 - 4x_1 + 5x_2$.

Write an anonymous function that calculates the objective.

```
fun = @(x)3*x(1)^2 + 2*x(1)*x(2) + x(2)^2 - 4*x(1) + 5*x(2);
```

Call `fminunc` to find a minimum of `fun` near `[1,1]`.

```
x0 = [1,1];
[x,fval] = fminunc(fun,x0);
```

`fminunc` can be faster and more reliable when you provide derivatives.

$$f(x) = 100\left(x_2 - x_1^2\right)^2 + (1 - x_1)^2,$$

which has gradient

$$\nabla f(x) = \begin{bmatrix} -400\left(x_2 - x_1^2\right)x_1 - 2\left(1 - x_1\right) \\ 200\left(x_2 - x_1^2\right) \end{bmatrix}.$$

```
function [f,g] = rosenbrockwithgrad(x)
% Calculate objective f
f = 100*(x(2) - x(1)^2)^2 + (1-x(1))^2;

if nargout > 1 % gradient required
    g = [-400*(x(2)-x(1)^2)*x(1)-2*(1-x(1));
        200*(x(2)-x(1)^2)];
end
```

```
function [f,g] = rosenbrockwithgrad(x)
% Calculate objective f
f = 100*(x(2) - x(1)^2)^2 + (1-x(1))^2;

if nargout > 1 % gradient required
    g = [-400*(x(2)-x(1)^2)*x(1)-2*(1-x(1));
        200*(x(2)-x(1)^2)];
end
```

------------- ------------- ------------- ------------- ------------- ------------- ------------- ------------- -------------

```
options = optimoptions('fminunc','Algorithm','trust-region','SpecifyObjectiveGradient',true);

problem.options = options;
problem.x0 = [-1,2];
problem.objective = @rosenbrockwithgrad;
problem.solver = 'fminunc';

x = fminunc(problem);
```

Otimização com restrições

$$\min_{\mathbf{x}} f\left(\mathbf{x}\right)$$

$$\text{sujeito a (s.a.)}$$

$$\omega_i\left(\mathbf{x}\right) = 0, i = 1, 2, \ldots, N_i$$

$$g_i\left(\mathbf{x}\right) \leq 0, i = 1, 2, \ldots, N_d$$

- $\omega_i(\mathbf{x}) \rightarrow$ restrições de igualdade
  - $N_i \rightarrow$ número de restrições de igualdade
- $g_i(\mathbf{x}) \rightarrow$ restrições de desigualdade
  - $N_d \rightarrow$ número de restrições de desigualdade
- Região viável $\rightarrow$ definida pelas funções $\omega_i(\mathbf{x})$ e $g_i(\mathbf{x})$

Otimização com restrições

- Exemplo:

$$\min_{x_1, x_2} f\left(x_1, x_2\right) = 0,25x_1^2 + x_2^2$$

$$sa$$

$$x_1 + x_2 = 5$$

- Na forma padrão,

$$\min_{x_1, x_2} f\left(x_1, x_2\right) = 0,25x_1^2 + x_2^2$$

$$sa$$

$$\omega\left(\mathbf{x}\right) = 5 - x_1 - x_2 = 0$$

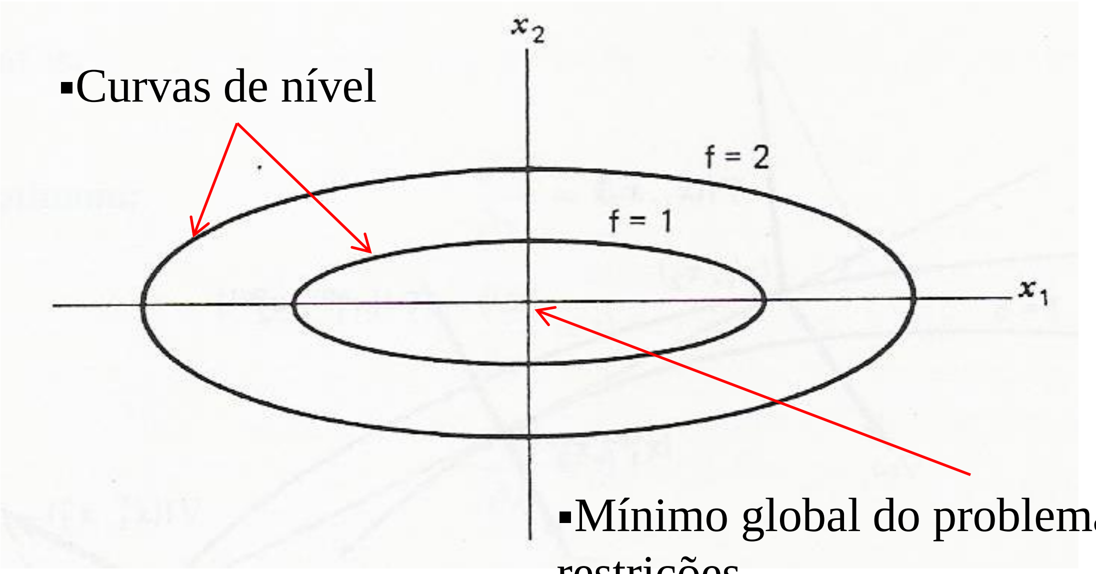
Otimização com restrições

- Exemplo:

$$\min_{x_1, x_2} f\left(x_1, x_2\right) = 0{,}25 x_1^2 + x_2^2$$

$$sa$$

$$\omega(\mathbf{x}) = 5 - x_1 - x_2 = 0$$



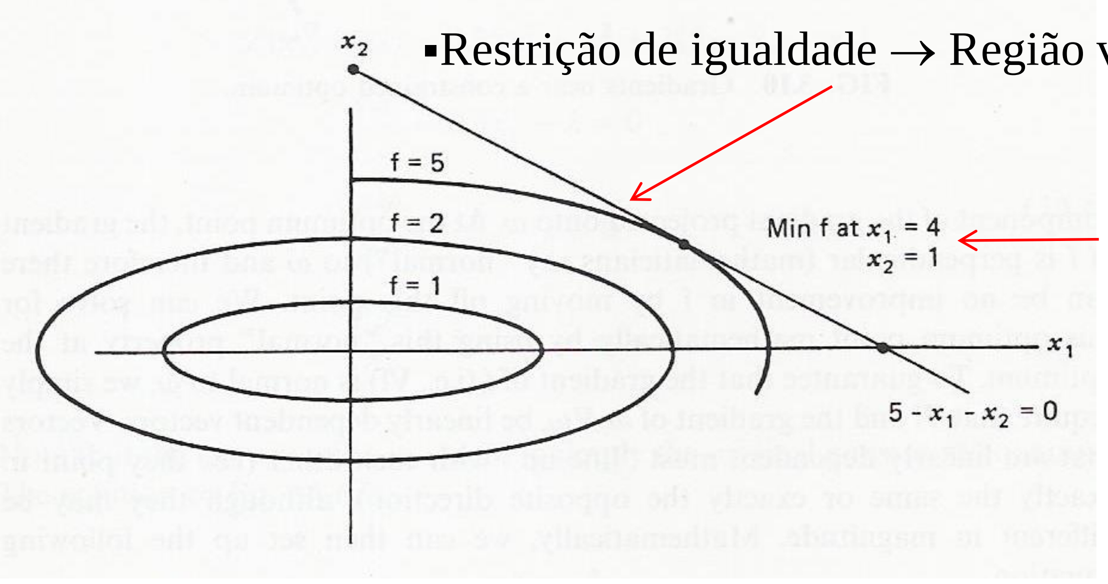
Otimização com restrições

- Exemplo:

$$\min_{x_1, x_2} f(x_1, x_2) = 0,25x_1^2 + x_2^2$$

$$sa$$

$$\omega(\mathbf{x}) = 5 - x_1 - x_2 = 0$$



- Restrição de igualdade → Região viável

- Tangente às curvas de nível

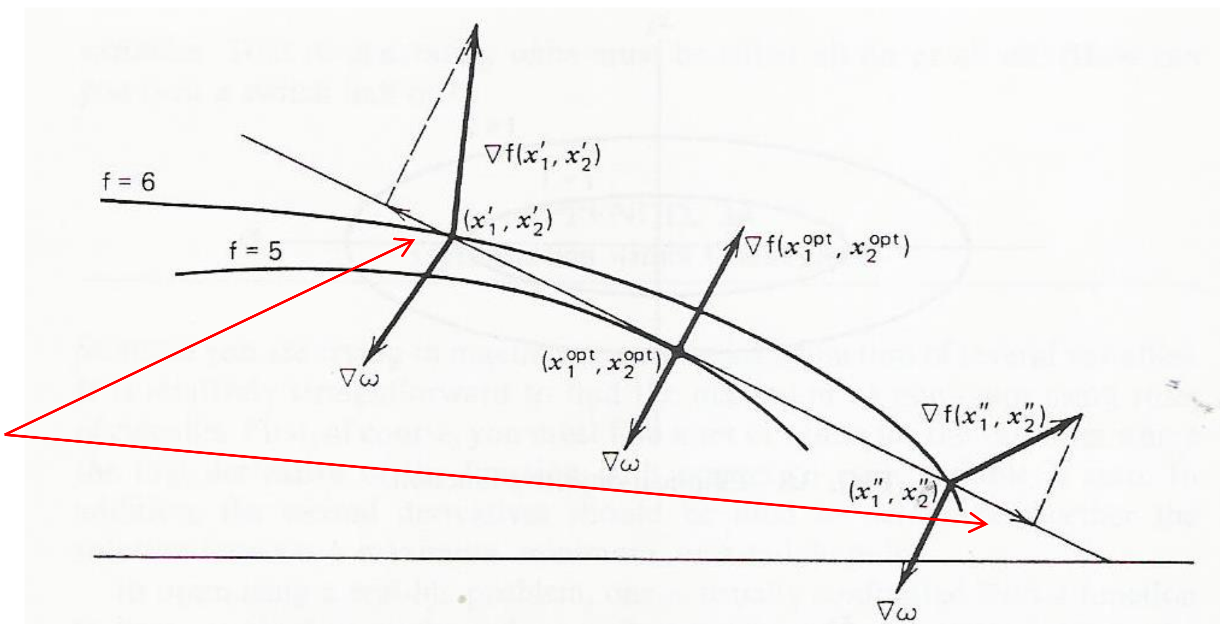
Otimização com restrições

- Exemplo:

$$\min_{x_1, x_2} f(x_1, x_2) = 0,25x_1^2 + x_2^2$$

$$sa$$

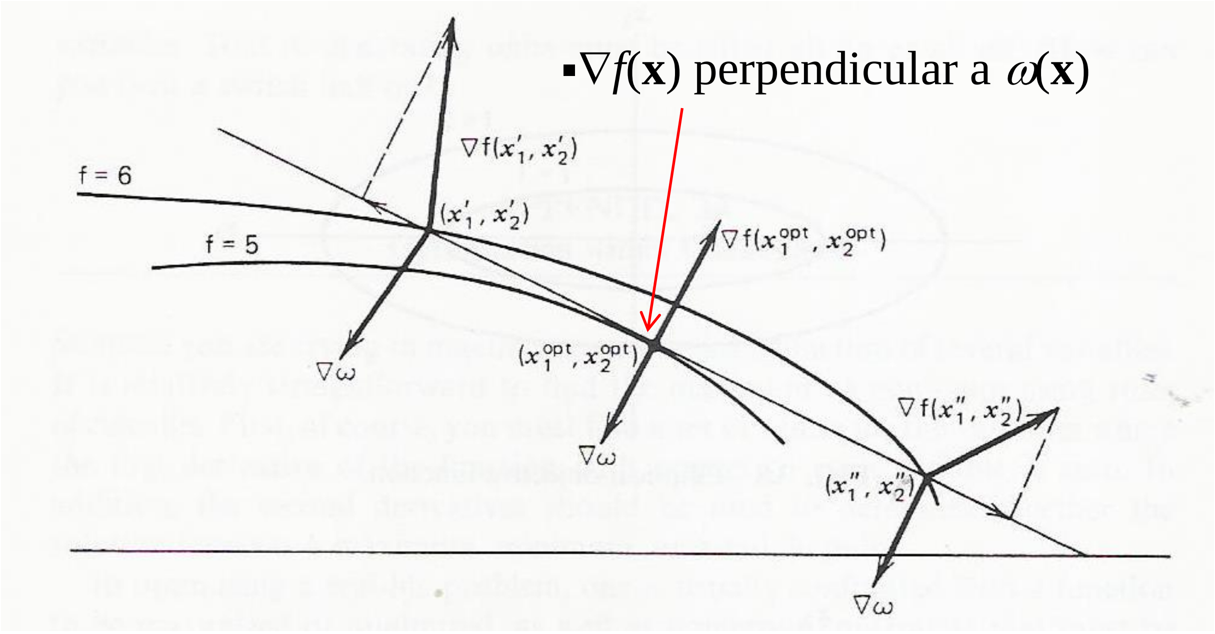
$$\omega(\mathbf{x}) = 5 - x_1 - x_2 = 0$$



- Componentes não-nulas de $\nabla f(\mathbf{x})$ na direção de $\omega(\mathbf{x})$

Otimização com restrições

- Exemplo: $$\min_{x_1,x_2} f\left(x_1,x_2\right)=0,25x_1^2+x_2^2$$
$$sa$$
$$\omega\left(\mathbf{x}\right)=5-x_1-x_2=0$$



Otimização com restrições

- Exemplo:

$$\min_{x_1,x_2} f\left(x_1,x_2\right) = 0,25x_1^2 + x_2^2$$

$$sa$$

$$\omega\left(\mathbf{x}\right) = 5 - x_1 - x_2 = 0$$

- No ponto ótimo, $\nabla f(\mathbf{x})$ é perpendicular a $\omega(\mathbf{x})$
  - Os gradientes de $f(\mathbf{x})$ e de $\omega(\mathbf{x})$ estão alinhados

$$\nabla f\left(\mathbf{x}\right) + \lambda \nabla \omega\left(\mathbf{x}\right) = \mathbf{0}$$

  - Princípio matemático do método dos multiplicadores de *Lagrange*

Otimização com restrições

- Exemplo:

$$\min_{x_1,x_2} f\left(x_1,x_2\right)=0,25x_1^2+x_2^2$$

$$sa$$

$$\omega\left(\mathbf{x}\right)=5-x_1-x_2=0$$

- Funcional *Lagrangeano*

$$L\left(x_1,x_2,\lambda\right)= f\left(x_1,x_2\right)+\lambda\omega\left(x_1,x_2\right)$$

- $\lambda \rightarrow$ multiplicador de *Lagrange*

Otimização com restrições

- Exemplo:

$$\min_{x_1,x_2} f\left(x_1,x_2\right)=0,25x_1^2+x_2^2$$

$$sa$$

$$\omega\left(\mathbf{x}\right)=5-x_1-x_2=0$$

  - Para o exemplo,

$$L\left(x_1,x_2,\lambda\right)=f\left(x_1,x_2\right)+\lambda\omega\left(x_1,x_2\right)$$

$$L\left(x_1,x_2,\lambda\right)=0,25x_1^2+x_2^2+\lambda\left(5-x_1-x_2\right)$$

Otimização com restrições

- Exemplo:
  - Para o exemplo,

$$L\left(x_1, x_2, \lambda\right) = 0,25x_1^2 + x_2^2 + \lambda\left(5 - x_1 - x_2\right)$$

$$\begin{cases} \dfrac{\partial}{\partial x_1} L\left(x_1, x_2, \lambda\right) = 0,5x_1 - \lambda = 0 \\ \dfrac{\partial}{\partial x_2} L\left(x_1, x_2, \lambda\right) = 2x_2 - \lambda = 0 \\ \dfrac{\partial}{\partial \lambda} L\left(x_1, x_2, \lambda\right) = 5 - x_1 - x_2 = 0 \end{cases} \Rightarrow \begin{cases} x_1^* = 4 \\ x_2^* = 1 \\ \lambda = 2 \end{cases}$$

Otimização com restrições

- Para problemas incluindo restrições de desigualdade, para que $\mathbf{x}^*$ seja um ponto de mínimo, além da condição de primeira ordem $\nabla L(\mathbf{x}^*,\lambda) = \mathbf{0}$, é necessário que $\mathbf{x}^*$ atenda as condições de *Karush-Kuhn-Tucker* (KKT)*:*

$$\min_{\mathbf{x}} f(\mathbf{x})$$

$$\text{sujeito a (s.a.)}$$

$$\omega_i(\mathbf{x}) = 0, i = 1, 2, \ldots, N_i$$

$$g_i(\mathbf{x}) \leq 0, i = 1, 2, \ldots, N_d$$

$$L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = f(\mathbf{x}) + \sum_{i=1}^{N_i} \lambda_i \omega_i(\mathbf{x}) + \sum_{i=1}^{N_d} \mu_i g_i(\mathbf{x})$$

Otimização com restrições

$$L(\mathbf{x},\boldsymbol{\lambda},\boldsymbol{\mu}) = f(\mathbf{x}) + \sum_{i=1}^{N_i} \lambda_i \omega_i(\mathbf{x}) + \sum_{i=1}^{N_d} \mu_i g_i(\mathbf{x})$$

- Condições KKT necessárias para que $\mathbf{x}^*$ seja um ponto de mínimo:

$$1)\ \frac{\partial}{\partial x_i} L\left(\mathbf{x}^*,\boldsymbol{\lambda}^*,\boldsymbol{\mu}^*\right) = 0, i = 1,2,\ldots,n$$

$$2)\ \omega_i\left(\mathbf{x}^*\right) = 0, i = 1,2,\ldots,N_i$$

$$3)\ g_i\left(\mathbf{x}^*\right) \leq 0, i = 1,2,\ldots,N_d$$

$$4)\ \begin{cases} \mu_i^* g_i\left(\mathbf{x}^*\right) = 0 \\ \mu_i^* \geq 0 \end{cases}, i = 1,2,\ldots,N_d$$

Lembrando que:

$$\min_{\mathbf{x}} f(\mathbf{x})$$

sujeito a (s.a.)

$$\omega_i(\mathbf{x}) = 0, i = 1,2,\ldots,N_i$$

$$g_i(\mathbf{x}) \leq 0, i = 1,2,\ldots,N_d$$

Otimização com restrições

- Exemplo

$$\min_{x_1,x_2} f\left(x_1,x_2\right)=0,25x_1^2+x_2^2$$

$$sa$$

$$x_1+x_2=5$$

$$x_1+0,2x_2\leq 3$$

- Na forma padrão,

$$\min_{x_1,x_2} f\left(x_1,x_2\right)=0,25x_1^2+x_2^2$$

$$sa$$

$$\omega\left(x_1,x_2\right)=5-x_1-x_2=0$$

$$g\left(x_1,x_2\right)=x_1+0,2x_2-3\leq 0$$

Otimização com restrições

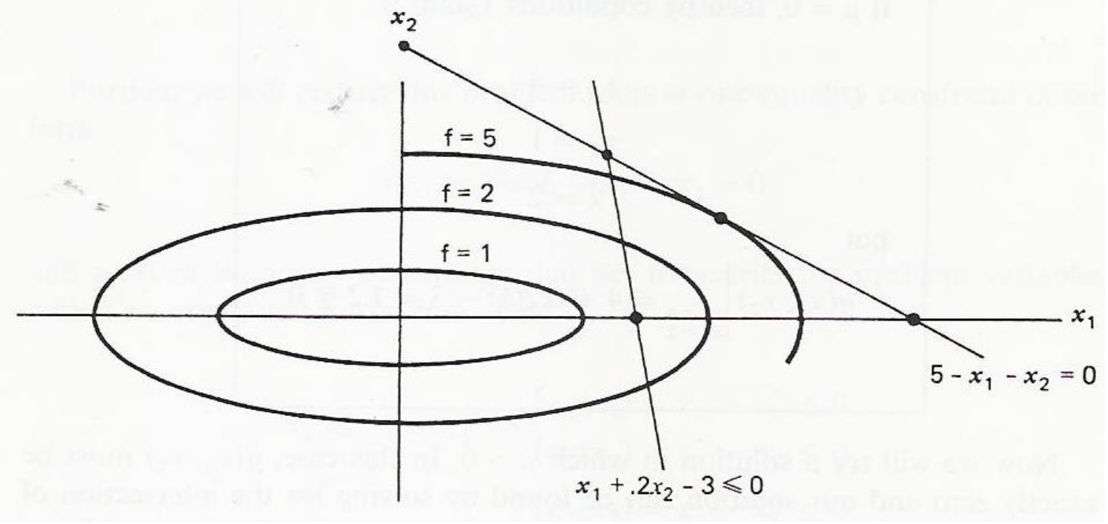
- Exemplo

$$\min_{x_1,x_2} f\left(x_1,x_2\right)=0,25x_1^2+x_2^2$$

$$sa$$

$$\omega\left(x_1,x_2\right)=5-x_1-x_2=0$$

$$g\left(x_1,x_2\right)=x_1+0,2x_2-3\leq 0$$



Otimização com restrições

- Exemplo
  - Calculando o *Lagrangeano*:

$$L(x_1, x_2, \lambda, \mu) = f(x_1, x_2) + \lambda\omega(x_1, x_2) + \mu g(x_1, x_2)$$

$$L(x_1, x_2, \lambda, \mu) = 0,25x_1^2 + x_2^2 + \lambda(5 - x_1 - x_2) + \mu(x_1 + 0,2x_2 - 3)$$

  - Condições KKT:

$$1)\begin{cases} \dfrac{\partial}{\partial x_1} L(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) = 0,5x_1 - \lambda + \mu = 0 \\ \dfrac{\partial}{\partial x_2} L(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) = 2x_2 - \lambda + 0,2\mu = 0 \end{cases}$$

Otimização com restrições

- Exemplo

$$L\left(x_1, x_2, \lambda, \mu\right) = f\left(x_1, x_2\right) + \lambda\omega\left(x_1, x_2\right) + \mu g\left(x_1, x_2\right)$$

$$L\left(x_1, x_2, \lambda, \mu\right) = 0,25x_1^2 + x_2^2 + \lambda\left(5 - x_1 - x_2\right) + \mu\left(x_1 + 0,2x_2 - 3\right)$$

  - Condições KKT:

$$2)\ \omega_1\left(\mathbf{x}^*\right) = 5 - x_1 - x_2 = 0$$

$$3)\ g_1\left(\mathbf{x}^*\right) = x_1 + 0,2x_2 - 3 \leq 0$$

$$4)\begin{cases} \mu^* g_1\left(\mathbf{x}^*\right) = \mu\left(x_1 + 0,2x_2 - 3\right) = 0 \\ \mu \geq 0 \end{cases}$$

Otimização com restrições

- Exemplo
  - Condições KKT:

$$1)\begin{cases}\dfrac{\partial}{\partial x_1}L\left(\mathbf{x}^*,\boldsymbol{\lambda}^*,\boldsymbol{\mu}^*\right)=0,5x_1-\lambda+\mu=0\\[2ex]\dfrac{\partial}{\partial x_2}L\left(\mathbf{x}^*,\boldsymbol{\lambda}^*,\boldsymbol{\mu}^*\right)=2x_2-\lambda+0,2\mu=0\end{cases}$$

$$2)\,\omega_1\left(\mathbf{x}^*\right)=5-x_1-x_2=0$$

$$3)\,g_1\left(\mathbf{x}^*\right)=x_1+0,2x_2-3\le 0$$

$$4)\begin{cases}\mu\ge 0\\ \mu^* g_1\left(\mathbf{x}^*\right)=\mu\left(x_1+0,2x_2-3\right)=0\end{cases}$$

- Equações lineares
- Inequações
- Equação não-linear

Otimização com restrições

- Exemplo
  - Condições KKT:
    - Se μ = 0:

$$\begin{cases} 0,5x_1 - \lambda = 0 \\ 2x_2 - \lambda = 0 \\ 5 - x_1 - x_2 = 0 \end{cases} \Rightarrow \begin{cases} \mu' = 0 \\ x_1' = 4 \\ x_2' = 1 \\ \lambda' = 2 \end{cases}$$

$$3)\, g_1(\mathbf{x}')\Big|_{x_1 = x_1';\, x_2 = x_2'} = x_1' + 0,2x_1' - 3 = 1,2 \not\leq 0$$

  - Violada a restrição de desigualdade
  - $\mathbf{x}'$ não é ponto de mínimo

Otimização com restrições

- Exemplo
  - Se μ > 0:

$$\begin{cases} \mu^* g_1\left(\mathbf{x}^*\right) = \mu\left(x_1 + 0,2x_2 - 3\right) = 0 \\ \mu \geq 0 \end{cases} \Rightarrow x_1 + 0,2x_2 - 3 = 0$$

$$\begin{cases} 0,5x_1 - \lambda + \mu = 0 \\ 2x_2 - \lambda + 0,2\mu = 0 \\ 5 - x_1 - x_2 = 0 \\ x_1 + 0,2x_2 - 3 = 0 \end{cases} \Rightarrow \begin{cases} \mu'' = 4,6875 \\ x_1'' = 2,5 \\ x_2'' = 2,5 \\ \lambda'' = 5,9375 \end{cases}$$

$$g_1\left(\mathbf{x}''\right)\Big|_{x_1 = x_1'';\, x_2 = x_2''} = x_1'' + 0,2x_1'' - 3 = 0 \leq 0$$

- Restrição de desigualdade atendida
  - Restrição ativa
- **x**″ é ponto de mínimo

Dificuldade: KKT não fornecem uma metodologia para obtenção de $\mathbf{x}^*$

Exemplo – fmincon.m

fmincon

Find minimum of constrained nonlinear multivariable function

Nonlinear programming solver.

Finds the minimum of a problem specified by

$$\min_x f(x) \text{ such that } \begin{cases} c(x) \leq 0 \\ ceq(x) = 0 \\ A \cdot x \leq b \\ Aeq \cdot x = beq \\ lb \leq x \leq ub, \end{cases}$$

Convex and nonconvex problems!

b and beq are vectors, A and Aeq are matrices, $c(x)$ and $ceq(x)$ are functions that return vectors, and $f(x)$ is a function that returns a scalar. $f(x)$, $c(x)$, and $ceq(x)$ can be nonlinear functions.

x, lb, and ub can be passed as vectors or matrices; see Matrix Arguments.

Syntax

```
x = fmincon(fun,x0,A,b)
x = fmincon(fun,x0,A,b,Aeq,beq)
x = fmincon(fun,x0,A,b,Aeq,beq,lb,ub)
x = fmincon(fun,x0,A,b,Aeq,beq,lb,ub,nonlcon)
x = fmincon(fun,x0,A,b,Aeq,beq,lb,ub,nonlcon,options)
```

Find the minimum value of Rosenbrock's function when there are both a linear inequality constraint and a linear equality constraint.

Set the objective function `fun` to be Rosenbrock's function.

```
fun = @(x)100*(x(2)-x(1)^2)^2 + (1-x(1))^2;
```

Find the minimum value starting from the point `[0.5,0]`, constrained to have $x(1) + 2x(2) \leq 1$ and $2x(1) + x(2) = 1$.

- Express the linear inequality constraint in the form `A*x <= b` by taking `A = [1,2]` and `b = 1`.
- Express the linear equality constraint in the form `Aeq*x <= beq` by taking `Aeq = [2,1]` and `beq = 1`.

```
x0 = [0.5,0];
A = [1,2];
b = 1;
Aeq = [2,1];
beq = 1;
x = fmincon(fun,x0,A,b,Aeq,beq)
```

Problema: Métodos geralmente custosos computacionalmente – para problemas de reconhecimento de padrões.

Problemas de Otimização Convexa

- Para problemas de otimização convexa, as condições KKT são necessárias e suficientes para caracterizar **x*** como um ponto de mínimo. Método mais “simples”: **DUALIDADE!**
- Problemas convexos:

$$\min_{\mathbf{x}} f\left(\mathbf{x}\right)$$

$$\text{sujeito a (s.a.)}$$

$$\omega_i\left(\mathbf{x}\right) = 0, i = 1, 2, \ldots, N_i$$

$$g_i\left(\mathbf{x}\right) \leq 0, i = 1, 2, \ldots, N_d$$

  - Função objetivo convexa $\rightarrow$ $f(\mathbf{x})$ **convexa**
  - Restrições de desigualdade **convexas** $\rightarrow$ $g_i(\mathbf{x})$ convexa
  - Restrições de igualdade são **funções afins** $\rightarrow$ $\omega_j(\mathbf{x})$
  - Funções afins: f(x) = ax+b

Dualidade

- Representação Dual de Wolfe:

Problema Original:

$$\min_{\mathbf{x}} f(\mathbf{x})$$

sujeito a (s.a.)

$$\omega_i(\mathbf{x}) = 0, i = 1, 2, \ldots, N_i$$

$$g_i(\mathbf{x}) \leq 0, i = 1, 2, \ldots, N_d$$

Valor ótimo **p*** (primal)

$$\equiv$$

$$\max_{\boldsymbol{\lambda},\boldsymbol{\mu}} \left[\min_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) \right]$$

$$s.t.: \boldsymbol{\lambda} \geq 0$$

$$\boldsymbol{\mu} \geq 0$$

$$\nabla_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = \mathbf{0}$$

Valor ótimo **d*** (dual)

Exemplo

- Problema:

$$\begin{aligned} \underset{x \in \mathbb{R}^2}{\text{minimize}} \quad & x_1^2 + x_2 \\ \text{subject to} \quad & 2x_1 + x_2 \geq 4 \\ & x_2 \geq 1. \end{aligned}$$

- Na forma padrão:

$$\begin{aligned} \underset{x \in \mathbb{R}^2}{\text{minimize}} \quad & x_1^2 + x_2 \\ \text{subject to} \quad & 4 - 2x_1 - x_2 \leq 0 \\ & 1 - x_2 \leq 0. \end{aligned}$$

Exemplo

- Lagrangeano:

$$\mathcal{L}(x,\alpha) = x_1^2 + x_2 + \alpha_1(4 - 2x_1 - x_2) + \alpha_2(1 - x_2)$$

- Representação de Wolfe:

$$\frac{\partial}{\partial x_1}\mathcal{L}(x,\alpha) = 2x_1 - 2\alpha_1 = 0 \quad \Longrightarrow \quad x_1 = \alpha_1$$

$$\frac{\partial}{\partial x_2}\mathcal{L}(x,\alpha) = 1 - \alpha_1 - \alpha_2 \ = 0$$

Exemplo

- Reescrevendo o Lagrangeano:

$$\mathcal{L}(x,\alpha) = x_1^2 + x_2 + \alpha_1(4 - 2x_1 - x_2) + \alpha_2(1 - x_2)$$

- Sendo: $x_1 = \alpha_1$

$$\left[\alpha_1^2 + x_2 + \alpha_1(4 - 2\alpha_1 - x_2) + \alpha_2(1 - x_2)\right]$$

$$\left[-\alpha_1^2 + 4\alpha_1 + \alpha_2 + x_2(1 - \alpha_1 - \alpha_2)\right]$$

- Sendo: $1 - \alpha_1 - \alpha_2 = 0$

$$\mathcal{L}(x,\alpha) = -\alpha_1^2 + 4\alpha_1 + \alpha_2$$

Exemplo

- Na representação de Wolfe:

$$\begin{aligned} \underset{\alpha \in \mathbb{R}^2}{\text{maximize}} \quad & -\alpha_1^2 + 4\alpha_1 + \alpha_2 \\ \text{subject to} \quad & \alpha_1 \geq 0 \\ & \alpha_2 \geq 0 \\ & 1 - \alpha_1 - \alpha_2 = 0 \end{aligned}$$

- Resulta em um problema de otimização quadrático com restrições mais simples (sempre convexo).

Exemplo – quadprog.m

Solve a simple quadratic programming problem:

find values of x that minimize $f(x) = \frac{1}{2}x_1^2 + x_2^2 - x_1x_2 - 2x_1 - 6x_2,$

subject to

$$x_1 + x_2 \le 2$$
$$-x_1 + 2x_2 \le 2$$
$$2x_1 + x_2 \le 3$$
$$0 \le x_1,\ 0 \le x_2.$$

In matrix notation this is $f(x) = \frac{1}{2}x^T H x + f^T x,$

where $H = \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix},\ f = \begin{bmatrix} -2 \\ -6 \end{bmatrix},\ x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}.$

$$A \cdot x \le b,$$
$$Aeq \cdot x = beq,$$
$$lb \le x \le ub.$$

1. Enter the coefficient matrices:

```
H = [1 -1; -1 2];
f = [-2; -6];
A = [1 1; -1 2; 2 1];
b = [2; 2; 3];
lb = zeros(2,1);
```

2. Set the options to use the `'interior-point-convex'` algorithm with no display:

```
options = optimoptions('quadprog',...
    'Algorithm','interior-point-convex','Display','off');
```

3. Call `quadprog`:

```
[x,fval,exitflag,output,lambda] = ...
   quadprog(H,f,A,b,[],[],lb,[],[],options);
```

4. Examine the final point, function value, and exit flag:

```
x,fval,exitflag

x =                 fval =            exitflag =
    0.6667             -8.2222             1
    1.3333
```

5. An exit flag of `1` means the result is a local minimum. Because `H` is a positive definite matrix, this problem is convex, so the minimum is a global minimum. You can see `H` is positive definite by noting all its eigenvalues are positive:

```
eig(H)
ans =
    0.3820
```

Resumo

- Condições de otimalidade:
  - Primeira ordem: gradiente nulo
  - Segunda ordem: autovalores da matriz Hessiana
- Otimização sem restrições (**fminuncon.m**):
  - Descida em Gradiente
  - Método de Newton (aproximação quadrática).
- Otimização com restrições de igualdade:
  - Lagrangeano → $\nabla L = 0$
- Otimização com restrições de igualdade e desigualdade:
  - Condições: KKT (necessárias e suficientes → convexo)
  - Possibilidade: **fmincon.m** → complexidade
  - Convexo (necessárias e suficientes → convexo) **quadprog.m**:
    - Dualidade de Wolfe → Programação Quadrática

Sequência “Típica”

- Definir o problema (classificação, regressão, clustering, etc);

$$J(\boldsymbol{w}) = \sum_{i=1}^{N} \left(y_i - \boldsymbol{x}_i^T \boldsymbol{w} \right)^2 \equiv \sum_{i=1}^{N} e_i^2$$

- Escrever o problema na forma de um problema de otimização, cujas variáveis são os parâmetros do classificador, por exemplo:

$$\hat{\boldsymbol{w}} = \arg\min_{\boldsymbol{w}} J(\boldsymbol{w})$$

- Escolher um método adequado para resolução do problema de otimização:

$$\frac{\partial J(\boldsymbol{w})}{\partial \boldsymbol{w}} = \boldsymbol{0} \longrightarrow \hat{\boldsymbol{w}} = (X^T X)^{-1} X^T \boldsymbol{y}$$

- Utilizar os parâmetros para realizar novas estimativas.

Referências

- Livro Convex Optimization (Boyd)
- Aulas Professor Nando de Freitas (UBC/Oxford)
- Aulas Professor Andrew Ng (Stanford)
- Aulas Professor Stephen Boyd (Stanford)
- Aulas Professor Vitor Hugo Ferreira (UFF)
- Livro Power Generation, Operation and Control – Allen Wood