PCC175/BCC465 - TÉCNICAS DE OTIMIZAÇÃO MULTI-OBJETIVO

Otimização Não-linear

Gladston Juliano Prates Moreira
email: gladston@ufop.edu.br

CSILab, Departamento de Computação
Universidade Federal de Ouro Preto

26 de agosto de 2024

Apresentação

Otimização Não-Linear
- Definições gerais
- Condições de Otimalidade
- Método de Penalidade

Métodos Determinísticos
- Métodos de direção de busca

Apresentação

Otimização Não-Linear
- Definições gerais
- Condições de Otimalidade
- Método de Penalidade

Métodos Determinísticos
- Métodos de direção de busca

Otimização Não-Linear

Problema de Otimização

Um problema de otimização, pode descrito por um modelo geral como segue:

$$\min \ f(x) \in \mathbb{R}$$

sujeito a:

$$\begin{cases} g_i(x) \leq 0 \ \forall i = 1, ..., m \\ h_j(x) = 0 \ \forall j = m+1, ..., l \\ x \in \mathcal{X} \end{cases}$$

Problema de Otimização

- A escolha da técnica de otimização depende da natureza das funções $f(x), g(x), h(x)$.
- Não há uma técnica de otimização que seja universal.
- Como fazer a escolha certa?

Conjuntos Convexos

Um conjunto $\mathcal{X} \subset \mathbb{R}^n$ é dito convexo se, $\forall x_1, x_2 \in \mathcal{X}$ e $0 \leq \lambda \leq 1$, tem-se que

$$\lambda x_1 + (1 - \lambda) x_2 \in \mathcal{X}$$

1. $\mathbb{C} = \left\{ (x_1; x_2) : x_1^2 + x_2^2 \leq 4 \right\} \subset \mathbb{R}^2$
2. $\mathbb{C} = \{ \mathrm{x} : A\mathrm{x} \leq \mathrm{b} \}$

Conjuntos Convexos

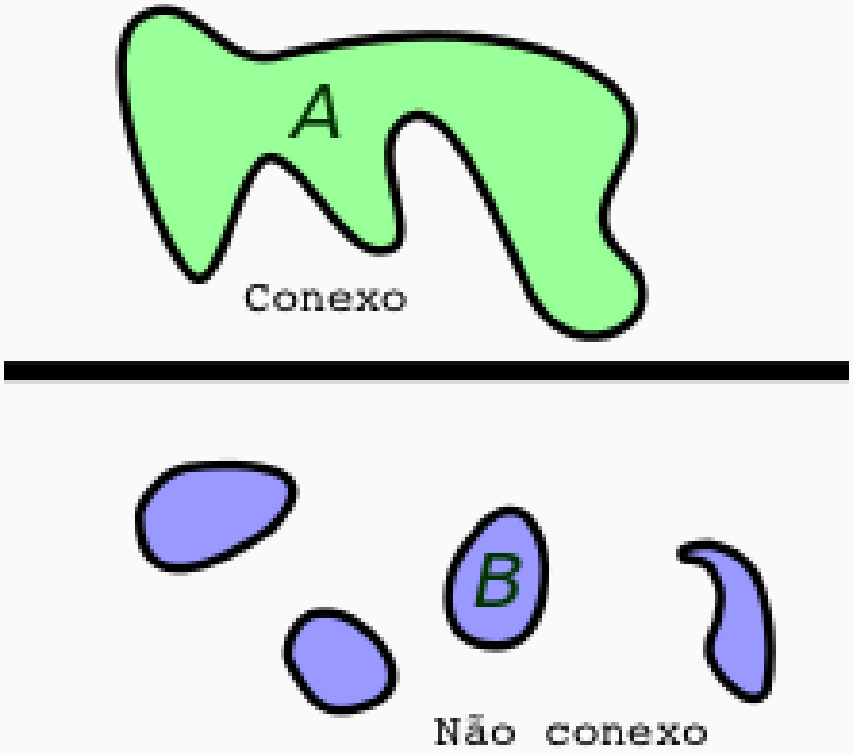
Um conjunto $\mathcal{X} \subset \mathbb{R}^n$ é dito convexo se, $\forall x_1, x_2 \in \mathcal{X}$ e $0 \leq \lambda \leq 1$, tem-se que

$$\lambda x_1 + (1 - \lambda) x_2 \in \mathcal{X}$$

1. $\mathbb{C} = \left\{ (x_1; x_2) : x_1^2 + x_2^2 \leq 4 \right\} \subset \mathbb{R}^2$
2. $\mathbb{C} = \{ \mathrm{x} : A\mathrm{x} \leq \mathrm{b} \}$

Conjuntos Conexos

Seja o conjunto $\mathcal{X} \subset \mathbb{R}^n$. Esse conjunto é dito conexo se, dados quaisquer dois pontos $x_1, x_2 \in \mathcal{X}$, existe uma curva contínua finita q tal que $x_1 \in q$, $x_2 \in q$, e $q \subset \mathcal{X}$.



Fundamentos

Uma função é uma relação que associa a um elemento de um conjunto A a um único elemento de um conjunto B, ou seja,

$$a \in A \rightarrow b = f(a) \in B$$

- Se $f(.) : \mathcal{X} \subset \mathbb{R}^n \rightarrow \mathbb{R}$, dizemos que $f(.)$ é um funcional.

Uma função $f(.) : \mathcal{X} \rightarrow Y$ é dita uma função convexa sobre $\mathcal{X} \subset \mathbb{R}^n$ se, $\forall x_1, x_2 \in \mathcal{X}$ e $0 \leq \lambda \leq 1$, tem-se que

$$f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda) f(x_2)$$

Fundamentos

Uma função é uma relação que associa a um elemento de um conjunto A a um único elemento de um conjunto B, ou seja,

$$a \in A \rightarrow b = f(a) \in B$$

- Se $f(.) : \mathcal{X} \subset \mathbb{R}^n \rightarrow \mathbb{R}$, dizemos que $f(.)$ é um funcional.

Uma função $f(.) : \mathcal{X} \rightarrow Y$ é dita uma função convexa sobre $\mathcal{X} \subset \mathbb{R}^n$ se, $\forall x_1, x_2 \in \mathcal{X}$ e $0 \leq \lambda \leq 1$, tem-se que

$$f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda) f(x_2)$$

Fundamentos

Uma função é uma relação que associa a um elemento de um conjunto A a um único elemento de um conjunto B, ou seja,

$$a \in A \rightarrow b = f(a) \in B$$

- Se $f(.) : \mathcal{X} \subset \mathbb{R}^n \rightarrow \mathbb{R}$, dizemos que $f(.)$ é um funcional.

Uma função $f(.) : \mathcal{X} \rightarrow Y$ é dita uma função convexa sobre $\mathcal{X} \subset \mathbb{R}^n$ se, $\forall x_1, x_2 \in \mathcal{X}$ e $0 \leq \lambda \leq 1$, tem-se que

$$f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda) f(x_2)$$

Superfície de nível e região subnível

Dada uma função $f(.) : \mathcal{X} \subset \mathbb{R}^n \to \mathbb{R}$, o conjunto $\mathcal{R}(f, \alpha) = \{x \in \mathcal{X} : f(x) \leq \alpha\}$ é denominada região de subnível.

- Se $f(.)$ é convexa então $\mathcal{R}(f, \alpha)$ é um conjunto convexo.

A superfície de nível (ou curva de nível) é definida como $\mathcal{S}(f, \alpha) = \{x \in \mathcal{X} : f(x) = \alpha\}$

Superfície de nível e região subnível

Dada uma função $f(.) : \mathcal{X} \subset \mathbb{R}^n \to \mathbb{R}$, o conjunto $\mathcal{R}(f, \alpha) = \{x \in \mathcal{X} : f(x) \leq \alpha\}$ é denominada região de subnível.

- Se $f(.)$ é convexa então $\mathcal{R}(f, \alpha)$ é um conjunto convexo.

A superfície de nível (ou curva de nível) é definida como $\mathcal{S}(f, \alpha) = \{x \in \mathcal{X} : f(x) = \alpha\}$

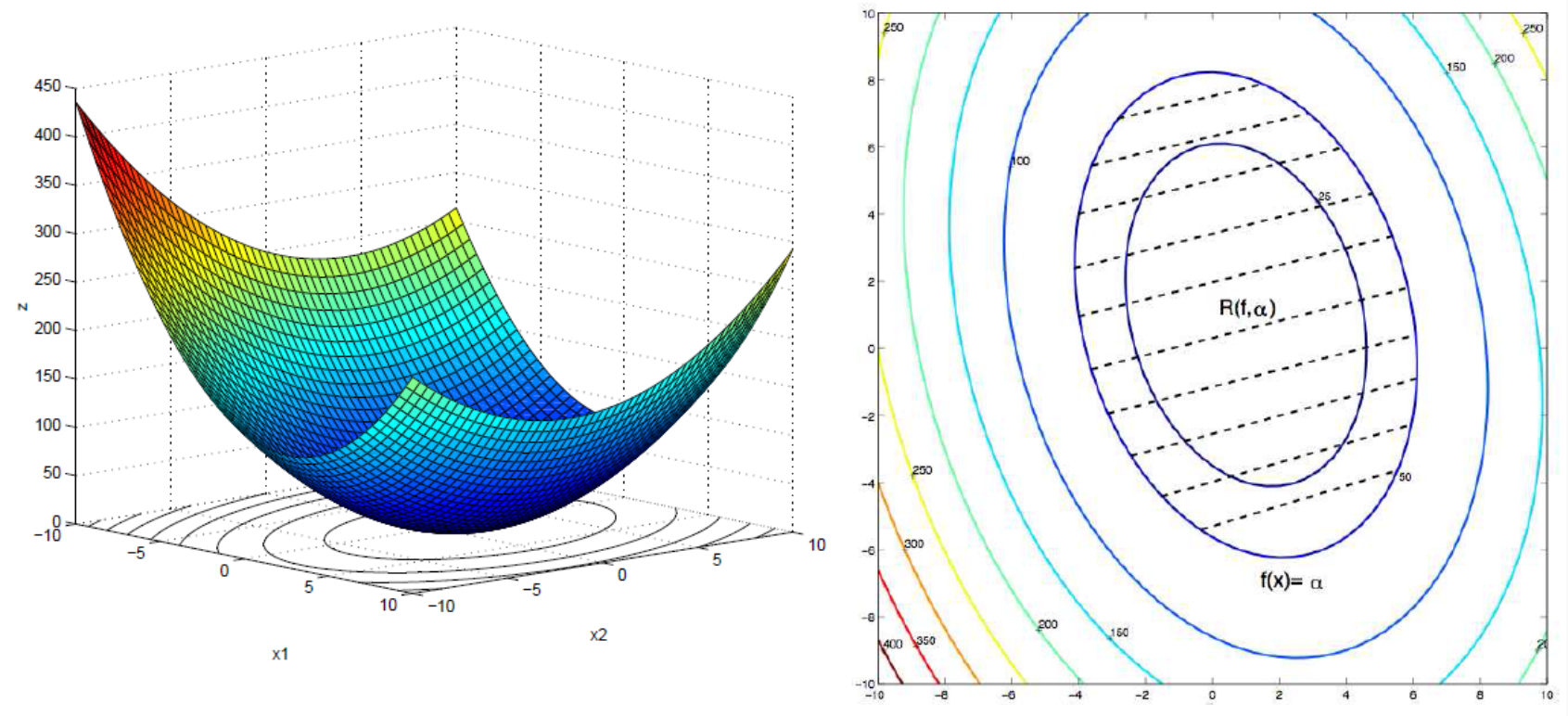
Superfície de nível e região subnível

Dada uma função $f(.) : \mathcal{X} \subset \mathbb{R}^n \to \mathbb{R}$, o conjunto $\mathcal{R}(f, \alpha) = \{x \in \mathcal{X} : f(x) \leq \alpha\}$ é denominada região de subnível.

- Se $f(.)$ é convexa então $\mathcal{R}(f, \alpha)$ é um conjunto convexo.

A superfície de nível (ou curva de nível) é definida como $\mathcal{S}(f, \alpha) = \{x \in \mathcal{X} : f(x) = \alpha\}$

Superfície de nível e região subnível



Modalidades de Funções

Uma função $f(.) : \mathcal{X} \subset \mathbb{R}^n \to \mathbb{R}$ é dita unimodal, se $\mathcal{R}(f, \alpha)$ é um conjunto conexo para todo $\alpha \in \mathbb{R}$.

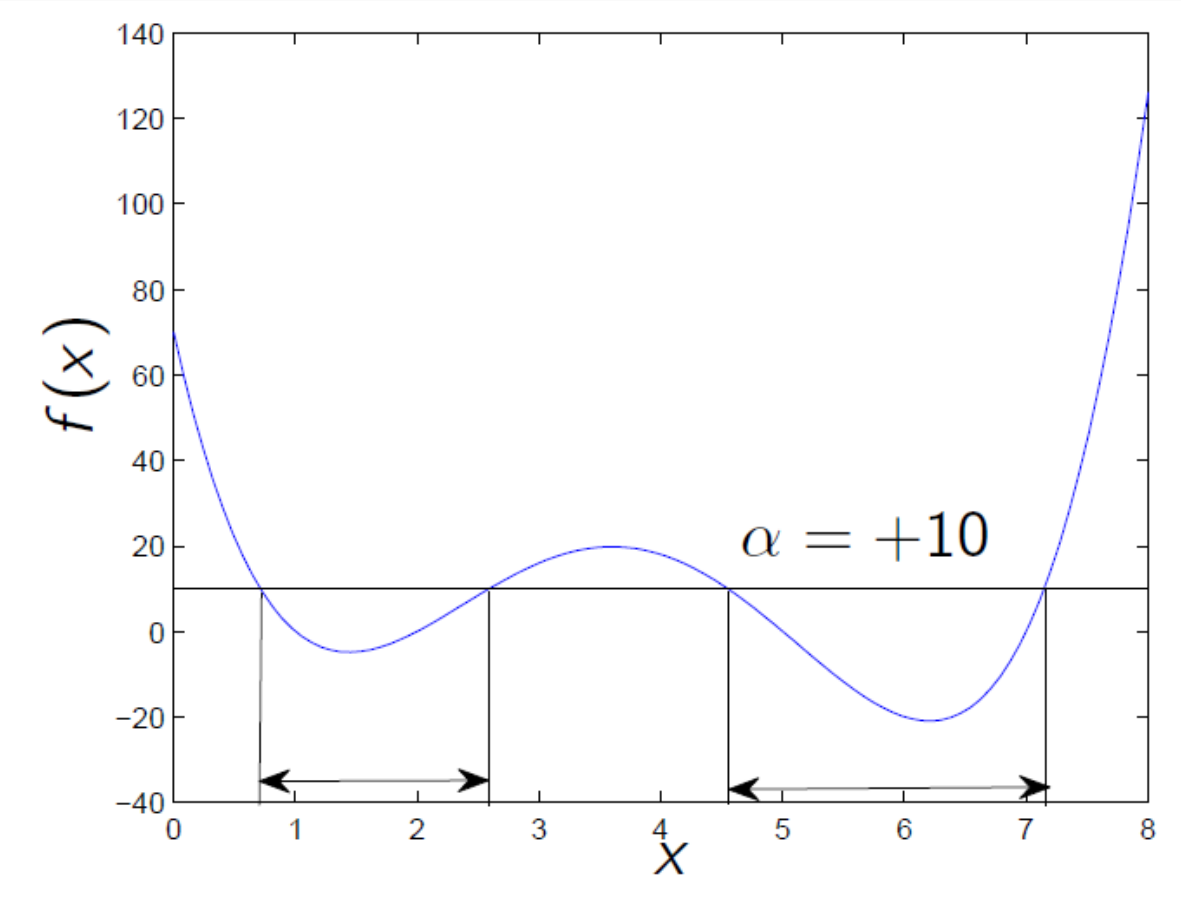
Uma função $f(.) : \mathcal{X} \subset \mathbb{R}^n \to \mathbb{R}$ é dita multimodal, se $\mathcal{R}(f, \alpha)$ é um conjunto desconexo para algum $\alpha \in \mathbb{R}$.

Modalidades de Funções

Uma função $f(.) : \mathcal{X} \subset \mathbb{R}^n \to \mathbb{R}$ é dita unimodal, se $\mathcal{R}(f, \alpha)$ é um conjunto conexo para todo $\alpha \in \mathbb{R}$.

Uma função $f(.) : \mathcal{X} \subset \mathbb{R}^n \to \mathbb{R}$ é dita multimodal, se $\mathcal{R}(f, \alpha)$ é um conjunto desconexo para algum $\alpha \in \mathbb{R}$.

Modalidades de Funções



Um exemplo de multimodalidade.

Modalidades de Funções

Note que uma função unimodal pode possuir múltiplos mínimos, desde que o conjunto destes forme um conjunto conexo (segundo esta definição).
A função

$$f(x) = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

os mínimos ocorrem sobre o eixo $x_1 = 0$.

Bacias de atração

Sejam $f(.) : \mathcal{X} \subset \mathbb{R}^n \to \mathbb{R}$ e $x^* \in \mathcal{X}$ um mínimo local de $f(.)$. Seja $\mathcal{R}(f, \alpha, x)$ o maior subconjunto conexo de $\mathcal{R}(f, \alpha)$ que contém x. A bacia de atração de x^*, $\mathcal{B}(x^*)$ é definida como a maior região conexa de subnível $\mathcal{R}(f, \alpha)$ associada x^*, sendo α^* o nível correspondente, tal que a função restrita a essa região:

$$f(.) : \mathcal{R}(f, \alpha^*, x^*) \to \mathbb{R}, \quad \text{é unimodal.}$$

Modalidades de Funções

Note que uma função unimodal pode possuir múltiplos mínimos, desde que o conjunto destes forme um conjunto conexo (segundo esta definição).
A função

$$f(x) = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

os mínimos ocorrem sobre o eixo $x_1 = 0$.

Bacias de atração

Sejam $f(.) : \mathcal{X} \subset \mathbb{R}^n \to \mathbb{R}$ e $x^* \in \mathcal{X}$ um mínimo local de $f(.)$.
Seja $\mathcal{R}(f, \alpha, x)$ o maior subconjunto conexo de $\mathcal{R}(f, \alpha)$ que contém x. A bacia de atração de x^*, $\mathcal{B}(x^*)$ é definida como a maior região conexa de subnível $\mathcal{R}(f, \alpha)$ associada x^*, sendo α^* o nível correspondente, tal que a função restrita a essa região:

$$f(.) : \mathcal{R}(f, \alpha^*, x^*) \to \mathbb{R}, \quad \text{é unimodal.}$$

A convexidade de funções pode ser relacionada com as regiões de sub-nível, superfícies de nível e bacias de atração.

1. Todas as regiões de sub-nível de uma função convexa num domínio convexo são conjuntos convexos.
2. Uma função convexa em um domínio convexo possui uma única bacia de atração, a qual é um conjunto convexo.

Continuidade e Diferenciabilidade

Uma função $f(.) : \mathcal{X} \subset \mathbb{R}^n \to \mathbb{R}$ é dita contínua se $\forall x_0 \in \mathcal{X}$:

- $f(x_0)$ está definido;
- $\lim_{x \to x_0} = f(x_0)$

- Uma função contínua é aquela para a qual uma pequena variação no domínio gera uma pequena variação na imagem (no resultado da função).

Continuidade e Diferenciabilidade

Uma função $f(.) : \mathcal{X} \subset \mathbb{R}^n \to \mathbb{R}$ é dita contínua se $\forall x_0 \in \mathcal{X}$:

- $f(x_0)$ está definido;
- $\lim\limits_{x \to x_0} = f(x_0)$

- Uma função contínua é aquela para a qual uma pequena variação no domínio gera uma pequena variação na imagem (no resultado da função).

Continuidade e Diferenciabilidade

- Uma função $f(.) : \mathcal{X} \subset \mathbb{R} \to \mathbb{R}$ é dita diferenciável em $x \in \mathcal{X}$ se existe o limite:

$$f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$$

- Uma função $f(.) : \mathcal{X} \subset \mathbb{R}^n \to \mathbb{R}$ é dita diferenciável de segunda ordem em $x \in \mathcal{X}$ se existe $f''(x)$.

Continuidade e Diferenciabilidade

- Uma função $f(.) : \mathcal{X} \subset \mathbb{R}^n \to \mathbb{R}$ é dita diferenciável em $x \in \mathcal{X}$ se existe o vetor gradiente:

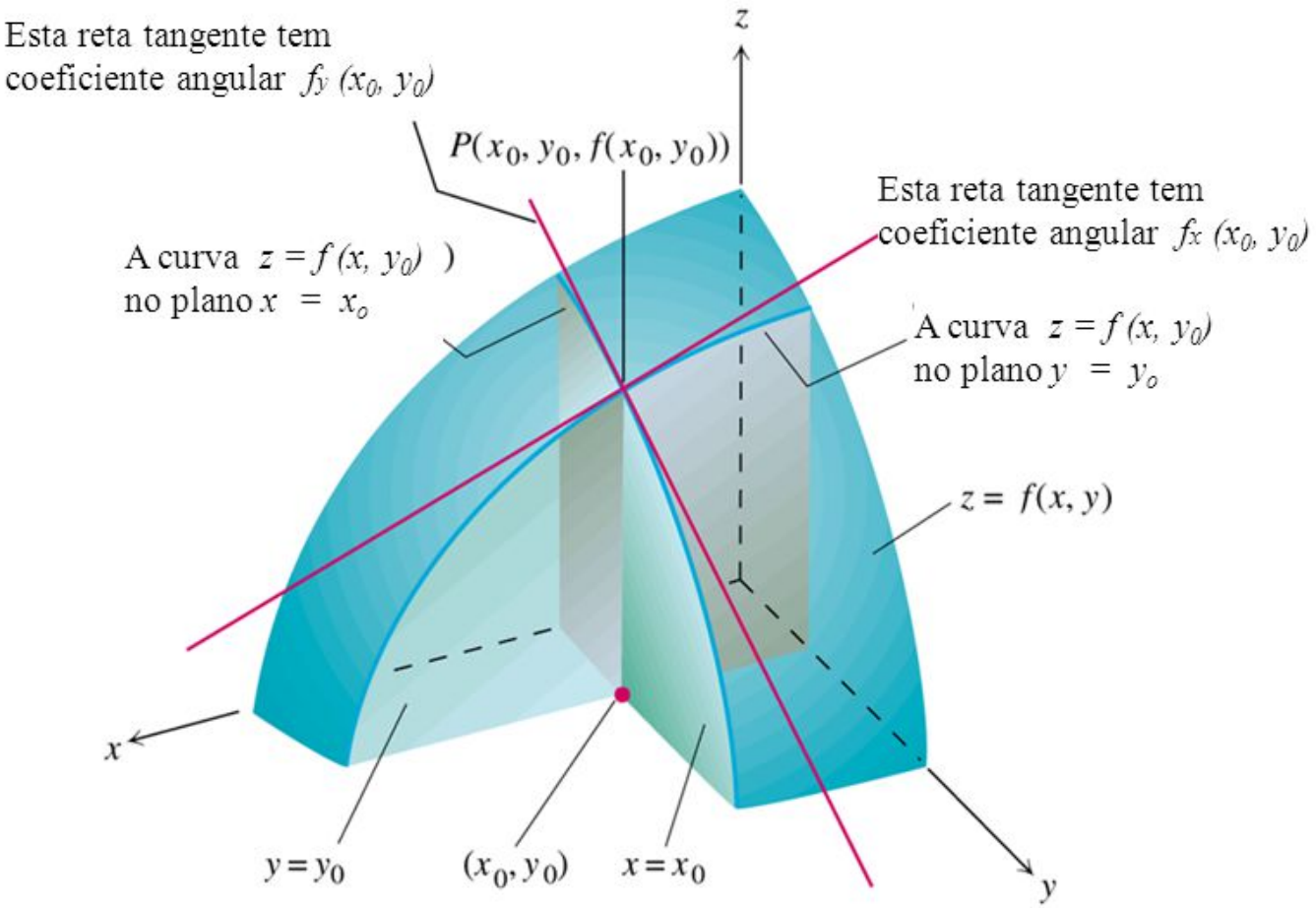
$$\nabla f(x) = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \ldots, \frac{\partial f}{\partial x_n} \right),$$

  e além disso as derivadas parciais $\frac{\partial f}{\partial x_i}, \ i = 1, \ldots, n$ são contínuas.

- Uma função $f(.) : \mathcal{X} \subset \mathbb{R}^n \to \mathbb{R}$ é dita diferenciável de segunda ordem em $x \in \mathcal{X}$ se existe a matriz Hessiana:

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

Significado geométrico



Direções factíveis e direções minimizantes

Diz-se que $d \in \mathbb{R}^n$ é uma direção factível a partir de um ponto $x_0 \in \mathcal{X}$, se existe um $\overline{\alpha} > 0$, tal que $(x + \alpha d) \in \mathcal{X}$, $\forall \alpha \in [0, \overline{\alpha}]$.

Sejam $f(.) : \mathcal{X} \subset \mathbb{R}^n \to \mathbb{R}$ diferenciável e $\nabla f(x)$ o gradiente de $f(.)$ em x. Seja $d \in \mathbb{R}^n$, se

$$d.\nabla f(x) < 0$$

então existe $\epsilon > 0$ tal que $f(x + \epsilon d) < f(x)$.
Diz-se que d é uma direção minimizante de f(x) a partir de um ponto x.

Direções factíveis e direções minimizantes

Diz-se que $d \in \mathbb{R}^n$ é uma direção factível a partir de um ponto $x_0 \in \mathcal{X}$, se existe um $\overline{\alpha} > 0$, tal que $(x + \alpha d) \in \mathcal{X}$, $\forall \alpha \in [0, \overline{\alpha}]$.

Sejam $f(.) : \mathcal{X} \subset \mathbb{R}^n \to \mathbb{R}$ diferenciável e $\nabla f(x)$ o gradiente de $f(.)$ em x. Seja $d \in \mathbb{R}^n$, se

$$d.\nabla f(x) < 0$$

então existe $\epsilon > 0$ tal que $f(x + \epsilon d) < f(x)$.
Diz-se que d é uma direção minimizante de f(x) a partir de um ponto x.

Direções factíveis e direções minimizantes

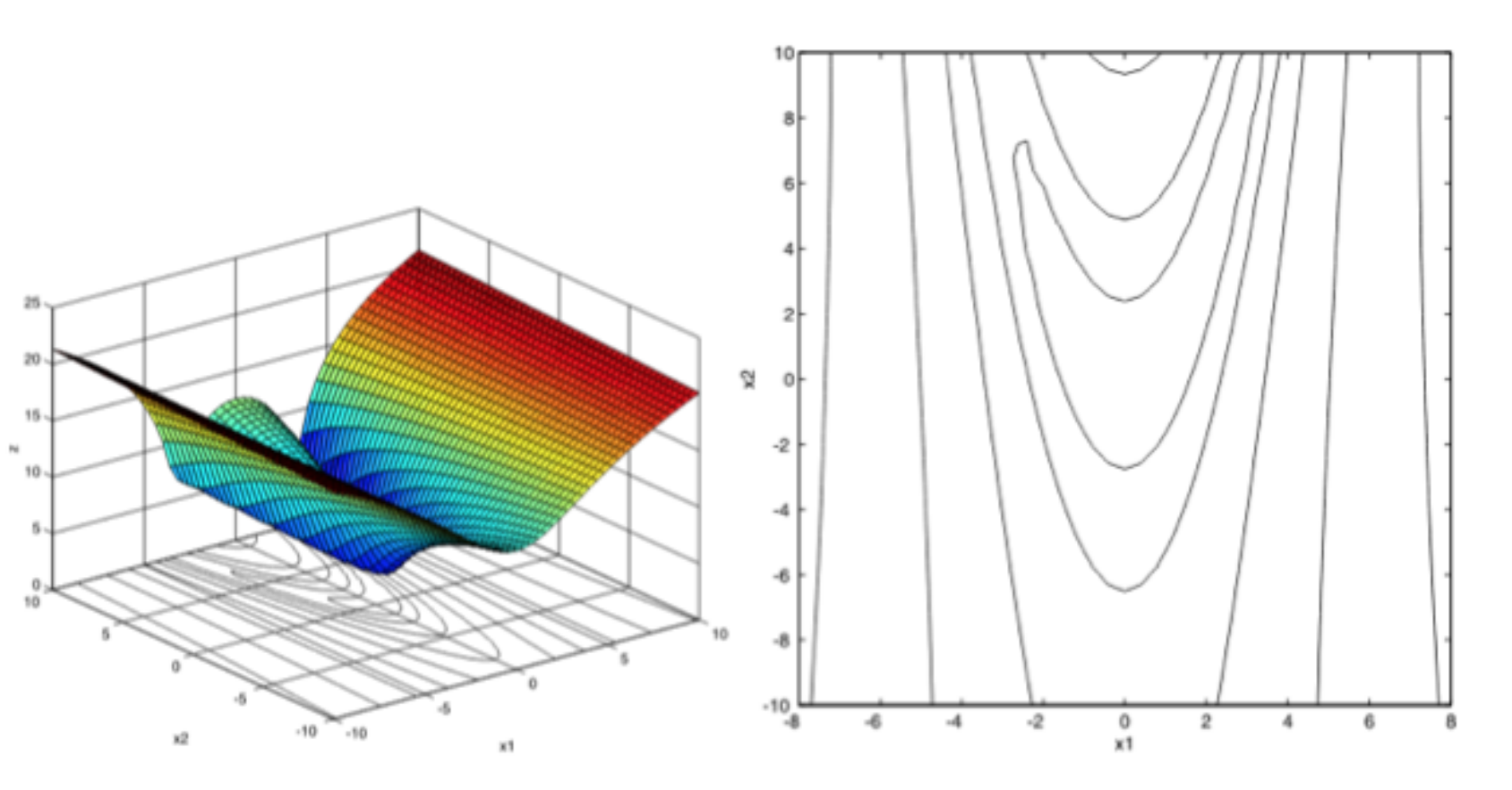
Diz-se que $d \in \mathbb{R}^n$ é uma direção factível a partir de um ponto $x_0 \in \mathcal{X}$, se existe um $\overline{\alpha} > 0$, tal que $(x + \alpha d) \in \mathcal{X}$, $\forall \alpha \in [0, \overline{\alpha}]$.

Sejam $f(.) : \mathcal{X} \subset \mathbb{R}^n \to \mathbb{R}$ diferenciável e $\nabla f(x)$ o gradiente de $f(.)$ em x. Seja $d \in \mathbb{R}^n$, se

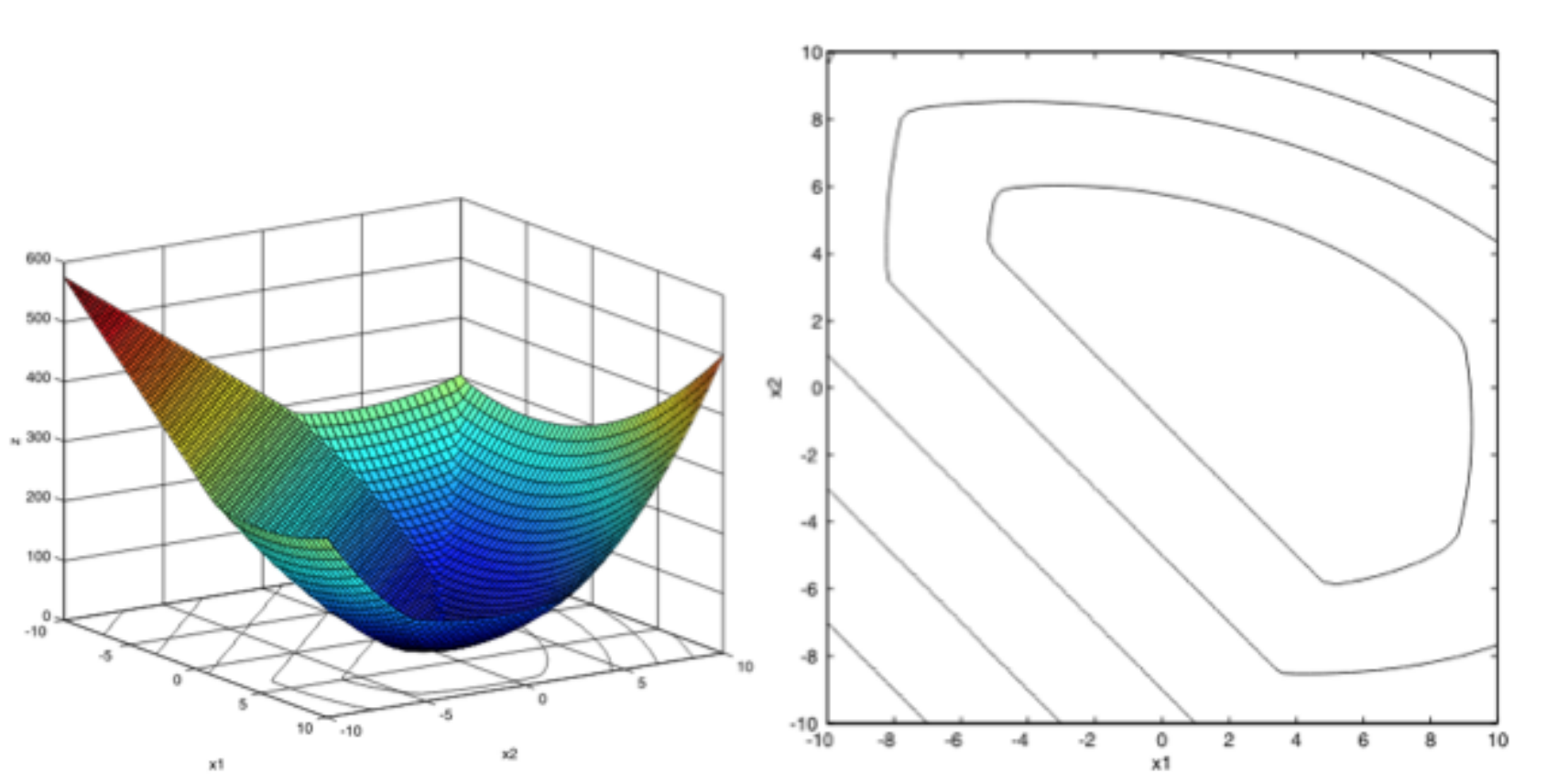
$$d.\nabla f(x) < 0$$

então existe $\epsilon > 0$ tal que $f(x + \epsilon d) < f(x)$.
Diz-se que d é uma direção minimizante de f(x) a partir de um ponto x.

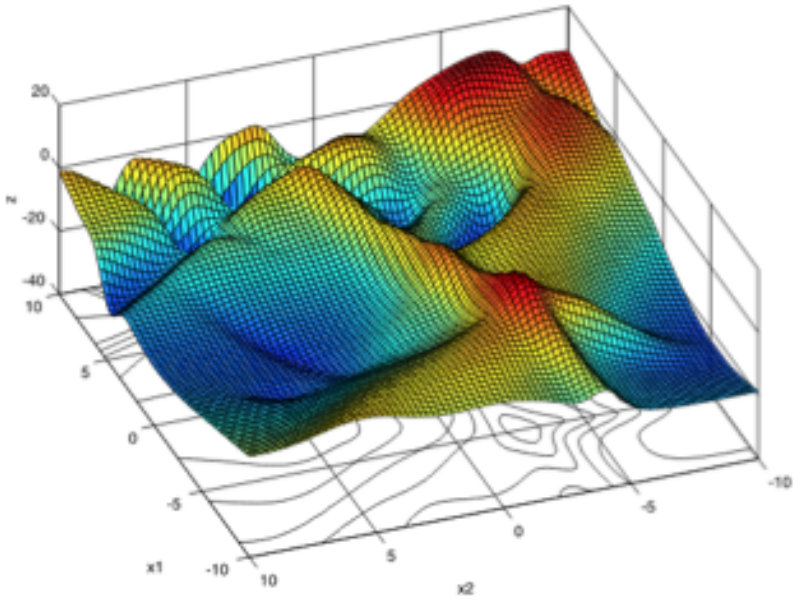
Exemplos

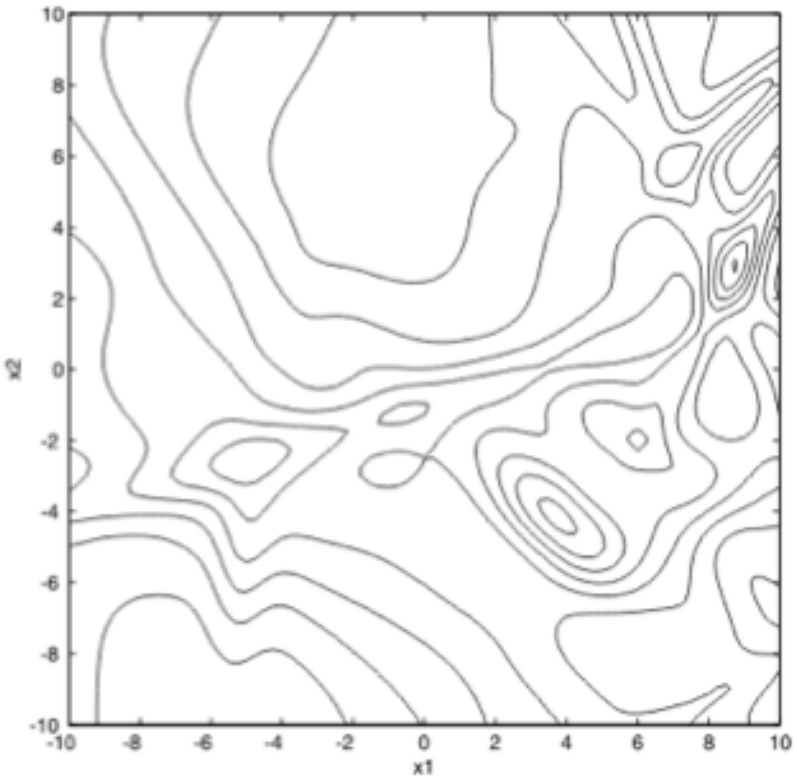


Exemplos

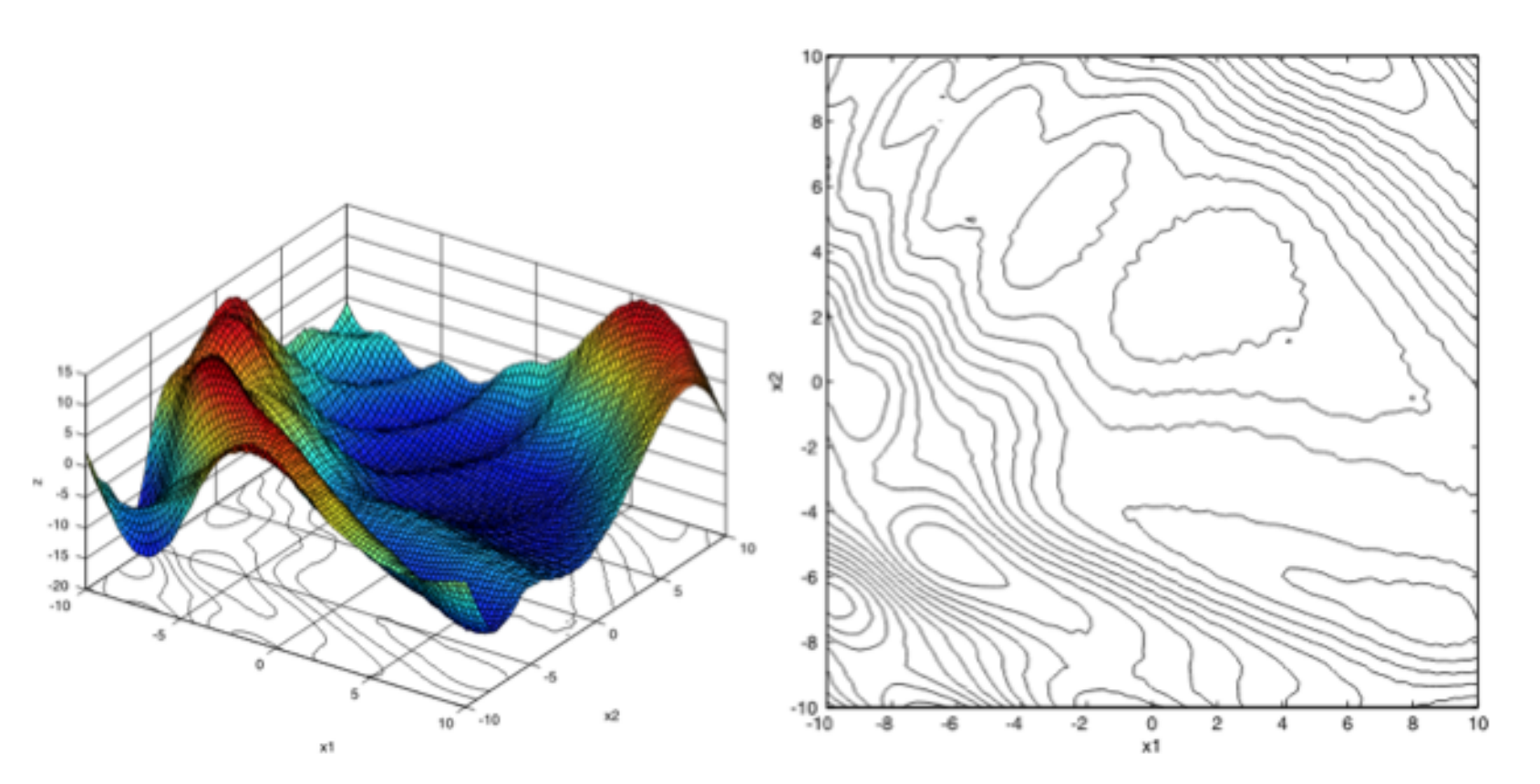

600
500
400
300
200
100
0
z
-10
-5
0
5
10
x1
-10
-5
0
5
10
x2
10
8
6
4
2
0
-2
-4
-6
-8
-10
x2
-10
-8
-6
-4
-2
0
2
4
6
8
10
x1

Exemplos


20
0
-20
-40
z
10
5
0
-5
-10
x1
10
5
0
-5
-10
x2


10
8
6
4
2
0
-2
-4
-6
-8
-10
x2
-10
-8
-6
-4
-2
0
2
4
6
8
10
x1

Exemplos



Otimalidade para Problemas Irrestritos

Um problema de otimização irrestrito, pode descrito por um modelo geral como segue:

$$\min \ f(x) \in \mathbb{R}, \ x \in \mathcal{X} \subseteq \mathbb{R}^n$$

Condições de Otimalidade

(Condição necessária de 1ª Ordem)
Seja $x^* \in \mathcal{X}$ um mínimo local da função $f(.) : \mathcal{X} \subset \mathbb{R}^n \to \mathbb{R}$. O vetor gradiente em x^* é nulo.

(Condição necessária de 2ª ordem)
Seja $x^* \in \mathcal{X}$ um mínimo local da função $f(.) : \mathcal{X} \subset \mathbb{R}^n \to \mathbb{R}$. A matriz Hessiana em x^* é semi-definida positiva.

Condições de Otimalidade

(Condição necessária de 1ª Ordem)
Seja $x^* \in \mathcal{X}$ um mínimo local da função $f(.) : \mathcal{X} \subset \mathbb{R}^n \to \mathbb{R}$. O vetor gradiente em x^* é nulo.

(Condição necessária de 2ª ordem)
Seja $x^* \in \mathcal{X}$ um mínimo local da função $f(.) : \mathcal{X} \subset \mathbb{R}^n \to \mathbb{R}$. A matriz Hessiana em x^* é semi-definida positiva.

Condições de Otimalidade

(Condições suficientes)
Seja $f(.) : \mathcal{X} \subset \mathbb{R}^n \to \mathbb{R}$ duas vezes diferenciável em $x^* \in \mathcal{X}$. Se $\nabla f(x^*) = 0$ e $H(x^*) > 0$ (definida positiva), então x^* é um mínimo local de $f(.)$

(Observação:)
As condições necessárias devem ser verdadeiras para todo ótimo local. Entretanto, um ponto que satisfaça estas condições não precisa ser um ótimo.

Condições de Otimalidade

(Condições suficientes)
Seja $f(.) : \mathcal{X} \subset \mathbb{R}^n \to \mathbb{R}$ duas vezes diferenciável em $x^* \in \mathcal{X}$. Se $\nabla f(x^*) = 0$ e $H(x^*) > 0$ (definida positiva), então x^* é um mínimo local de $f(.)$

(Observação:)
As condições necessárias devem ser verdadeiras para todo ótimo local. Entretanto, um ponto que satisfaça estas condições não precisa ser um ótimo.

De forma geral, uma matriz H é:

- positiva definida $\Leftrightarrow$ todos os autovalores são positivos
- positiva semi-definida $\Leftrightarrow$ todos os autovalores são não-negativos
- negativa definida $\Leftrightarrow$ todos os autovalores são negativos
- negativa semi-definida $\Leftrightarrow$ todos os autovalores são não-posititos
- indefinida $\Leftrightarrow$ possui autovalores positivos e negativos

Otimalidade para Problemas Irrestritos

Seja o problema restrito a seguir:

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

- Ótimo em $(0, 0)$
- $\nabla f(x) = (2x_1, 2x_2)$
- $H(x) = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$

Otimalidade para Problemas Irrestritos

Seja o problema restrito a seguir:

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

- Ótimo em $(0, 0)$
- $\nabla f(x) = (2x_1, 2x_2)$
- $H(x) = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$

Otimalidade para Problemas Restritos

Um problema de otimização, pode descrito por um modelo geral como segue:

$$\min \ f(x) \in \mathbb{R}$$

sujeito a:

$$\begin{cases} g_i(x) \leq 0 \ \forall i = 1, ..., m \\ h_j(x) = 0 \ \forall j = m+1, ..., l \\ x \in \mathcal{X} \subseteq \mathbb{R}^n \end{cases}$$

Otimalidade para Problemas Restritos

Seja o problema restrito a seguir:

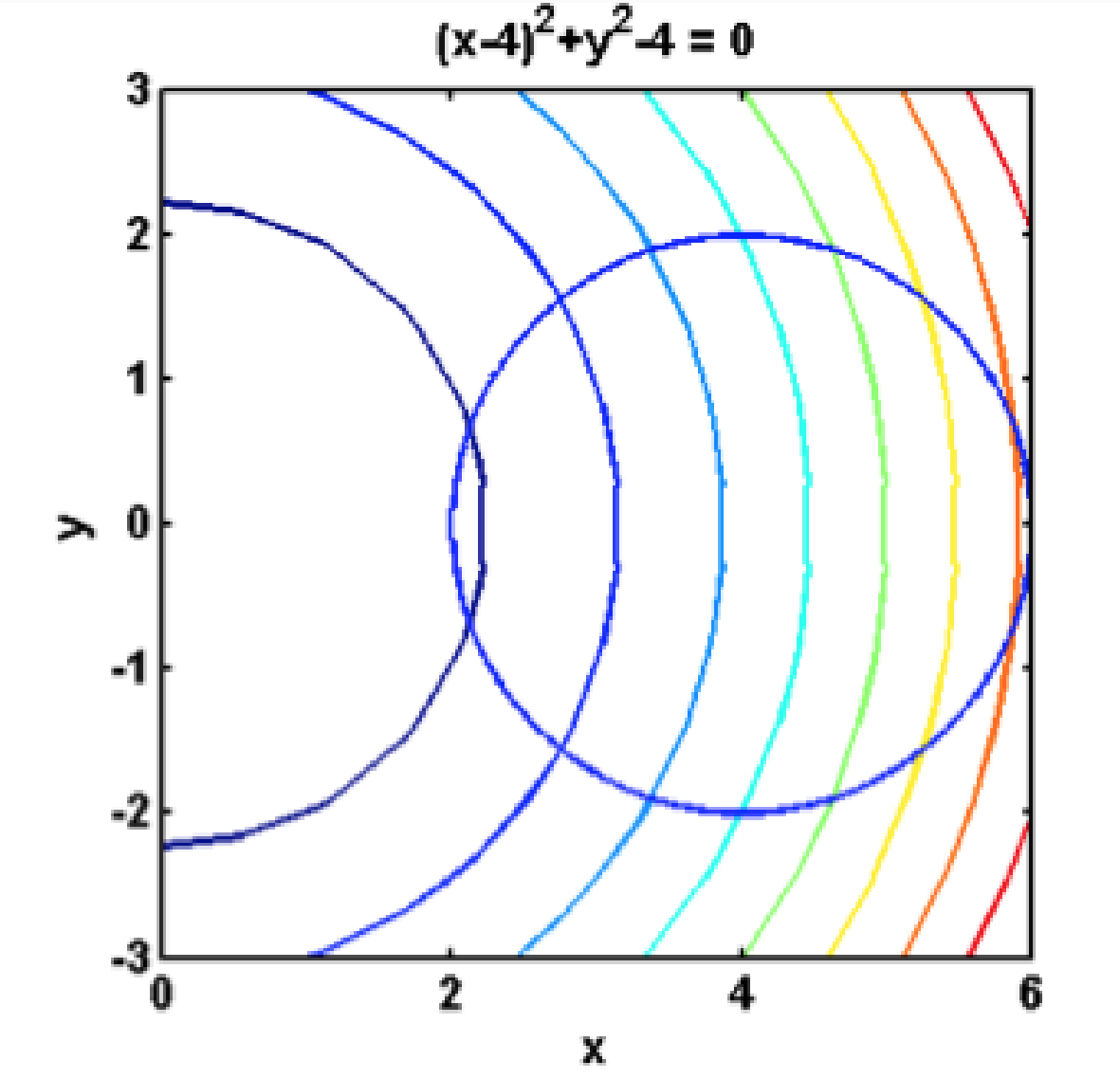
$$\min f(x) = x_1^2 + x_2^2$$

sujeito a:

$$h(x) = (x_1 - 4)^2 + x_2^2 - 4 = 0$$

- Como determinar as condições de otimalidade?

Otimalidade para Problemas Restritos



Otimalidade para Problemas Restritos

Graficamente:

- O mínimo global de $f(x)$ sobre a curva $h(x) = 0$ é $x^* = (2, 0)$.
- Os vetores ∇f e ∇h satisfazem:

$$\nabla f(x^*) = -\lambda^* \nabla h(x^*)$$

$\nabla[f(x) + \lambda^* h(x)]_{x=x^*} = 0$
$\nabla_x L(x^*, \lambda^*) = 0$
Fazendo $L(x, \lambda) = f(x) + \lambda h(x)$.

Otimalidade para Problemas Restritos

Graficamente:

- O mínimo global de $f(x)$ sobre a curva $h(x) = 0$ é $x^* = (2, 0)$.
- Os vetores ∇f e ∇h satisfazem:

$$\nabla f(x^*) = -\lambda^* \nabla h(x^*)$$

$\nabla[f(x) + \lambda^* h(x)]_{x=x^*} = 0$
$\nabla_x L(x^*, \lambda^*) = 0$
Fazendo $L(x, \lambda) = f(x) + \lambda h(x)$.

Otimalidade para Problemas Restritos

A solução do problema restrito

$$\min f(x) \text{ sujeito a: } h(x) = 0$$

é o ponto critico da função

$$L(x, \lambda) = f(x) + \lambda h(x)$$

- Implicando em:

$$\begin{bmatrix} \nabla_x L \\ \nabla_\lambda L \end{bmatrix} = 0$$

Otimalidade para Problemas Restritos

Com relação a uma restrição de desigualdade, podemos converter $g(x) \leq 0$ em $g(x) + z^2 = 0$, sendo z uma variável de folga:

$$\min f(x) \text{ sujeito a: } \{g(x) \leq 0 \text{ ou } g(x) + z^2 = 0\}$$

- A função Lagrangiana é dada por:

$$L(x, z, \mu) = f(x) + \mu[g(x) + z^2]$$

- O ponto crítico:

$$\begin{bmatrix} \nabla_x L \\ \nabla_z L \\ \nabla_\mu L \end{bmatrix} = 0 \Rightarrow \begin{cases} \nabla f(x^*) + \mu^* \nabla g(x^*) = 0 \\ \mu^* g(x^*) = 0 \\ g(x^*) \leq 0 \end{cases}$$

Otimalidade Karush-Kuhn-Tucker

Dado o problema de otimização:

$$\min \ f(x) \in \mathbb{R}$$

sujeito a:

$$\begin{cases} g_i(x) \leq 0 \ \forall i = 1, ..., m \\ h_j(x) = 0 \ \forall j = m+1, ..., l \\ x \in \mathcal{X} \end{cases}$$

x^* é ótimo se existem multiplicadores de Lagrange $\mu^* \geq 0$ e λ^* tais que:

$$\begin{cases} \nabla f(x^*) + \sum_{i=1}^{m} \mu_i^* \nabla g_i(x^*) + \sum_{j=1}^{l} \lambda_j^* \nabla h_j(x^*) = 0 \\ \mu_i^* g_i(x^*)^* = 0 \\ g_i(x^*) \leq 0 \\ h_j(x^*) = 0 \end{cases}$$

Problema Exemplo

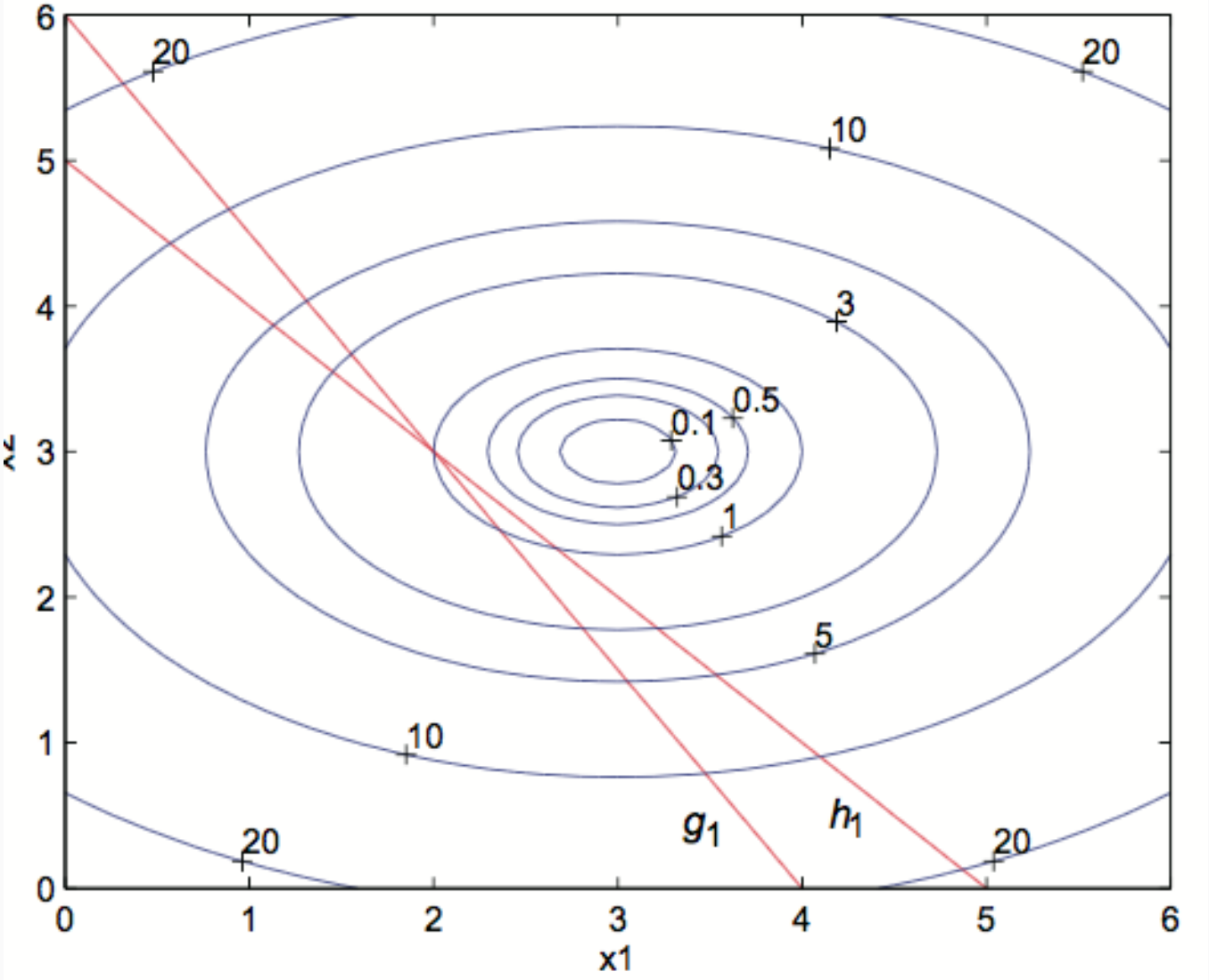
Considere o problema:

$$\min f(x) = (x_1 - 3)^2 + 2(x_2 - 3)^2$$

sujeito a:

$$\begin{cases} h(x) : x_1 + x_2 - 5 = 0 \\ g(x) : 3x_1 + 2x_2 \leq 12 \\ 0 \leq x_1 \leq 6; 0 \leq x_2 \leq 6 \end{cases}$$

Problema Exemplo



Método de Penalidade

Infelizmente, o Método de Lagrange não é muito prático, pelas seguintes razões:

- A construção do sistema de equações pressupõe acesso às expressões analíticas de $f(x), g_i(x), h_j(x)$;
- A resolução do problema de otimização não linear original requer a solução de um sistema de equações não lineares;

Método de Penalidade

Infelizmente, o Método de Lagrange não é muito prático, pelas seguintes razões:

- A construção do sistema de equações pressupõe acesso às expressões analíticas de $f(x), g_i(x), h_j(x)$;
- A resolução do problema de otimização não linear original requer a solução de um sistema de equações não lineares;

Método de Penalidade

- A abordagem dos Métodos de Penalidades consiste na transformação do problema restrito original num problema irrestrito equivalente (dois problemas são ditos equivalentes se possuem a mesma solução);
- Métodos indiretos para otimização restrita.

Método de Penalidade

- A abordagem dos Métodos de Penalidades consiste na transformação do problema restrito original num problema irrestrito equivalente (dois problemas são ditos equivalentes se possuem a mesma solução);
- Métodos indiretos para otimização restrita.

Métodos de Penalidade

Nos métodos de Penalidade, penalidades são adicionadas à função-objetivo:

- **Métodos de penalidade interior ou métodos Barreira**: pontos gerados devem ser sempre viáveis e qualquer tentativa de sair da região factível é penalizada.
- **Métodos de penalidade exterior ou métodos de Penalidade**: qualquer violação de alguma restrição é penalizada no valor da função-objetivo.

Métodos de Penalidade

Nos métodos de Penalidade, penalidades são adicionadas à função-objetivo:

- **Métodos de penalidade interior ou métodos Barreira**: pontos gerados devem ser sempre viáveis e qualquer tentativa de sair da região factível é penalizada.
- **Métodos de penalidade exterior ou métodos de Penalidade**: qualquer violação de alguma restrição é penalizada no valor da função-objetivo.

Métodos de Penalidade

Em geral, em problemas restritos a solução reside na fronteira da região factível:

- Nos métodos de penalidade interior, a solução ótima é aproximada internamente por uma sequência de soluções do problema irrestrito transformado;
- Nos métodos de penalidade exterior, a solução ótima é aproximada externamente por uma sequência de soluções do problema irrestrito transformado;

Métodos de Penalidade

Em geral, em problemas restritos a solução reside na fronteira da região factível:

- Nos métodos de penalidade interior, a solução ótima é aproximada internamente por uma sequência de soluções do problema irrestrito transformado;
- Nos métodos de penalidade exterior, a solução ótima é aproximada externamente por uma sequência de soluções do problema irrestrito transformado;

Método de Penalidade Exterior

Função de Penalidade

Uma função de penalidade deve atender as seguintes condições:

$$\begin{cases} p(x) > 0, & \text{se } x \notin \mathcal{X} \\ p(x) = 0, & \text{se } x \in \mathcal{X} \end{cases}$$

Método de Penalidade Exterior

Função de Penalidade

1. No caso de um problema com uma restrição de igualdade:

$$\begin{aligned} &\min f(x) \\ &\text{s.a } h(x) = 0 \qquad \Longrightarrow \min f(x) + u[h(x)]^2 \end{aligned}$$

2. No caso de um problema com uma restrição de desigualdade:

$$\begin{aligned} &\min f(x) \\ &\text{s.a } g(x) \leq 0 \qquad \Longrightarrow \min f(x) + u \max[0, g(x)]^2 \end{aligned}$$

Método de Penalidade Exterior

Função de Penalidade

1. No caso de um problema com uma restrição de igualdade:

$$\begin{aligned} &\min f(x) \\ &\text{s.a } h(x) = 0 \qquad \Longrightarrow \min f(x) + u[h(x)]^2 \end{aligned}$$

2. No caso de um problema com uma restrição de desigualdade:

$$\begin{aligned} &\min f(x) \\ &\text{s.a } g(x) \leq 0 \qquad \Longrightarrow \min f(x) + u \max[0, g(x)]^2 \end{aligned}$$

Função de Penalidade

De forma geral temos::

$$p(x,u) = u\left\{\sum_{i=1}^{p}[h_i(x)]^2 + \sum_{j=1}^{q}\max[0, g_j(x)]^2\right\}$$

Exemplo

Seja o problema min $f(x)$, com $f(x) = x$, sujeito a $g(x) = -x + 3 \leq 0$.

Solução

Usando $p(x) = \max[0, g(x)]^2$, tem-se:

$$p(x) = \begin{cases} 0, & x \geq 3 \\ (3-x)^2, & x < 3 \end{cases}$$

Para $x < 3$, $f(x) + p(x, u) = x + u(3-x)^2$. Assim:

$$\frac{df}{dx} = 1 - 2u(3-x) = 0 \;\rightarrow\; x^* = 3 - \frac{1}{2u}$$

Com $u \rightarrow \infty$, temos $x^* \rightarrow 3^-$.

Método de Penalidade Exterior

Algoritmo 1: Método de Penalidade Exterior

Input: $\mathbf{x}_0 \in \mathcal{X}$, $u_0 > 0$, função-objetivo $f(\cdot)$ e restrições $\mathbf{g}(\cdot)$ e $\mathbf{h}(\cdot)$

$k \leftarrow 0$;
while $\neg$ *critério de parada* **do**
  Começando de $\mathbf{x}_k$, encontre $\mathbf{x}_{k+1} \leftarrow \arg\min_{\mathbf{x}} f(\mathbf{x}) + p(\mathbf{x}, u_k)$;
  $u_{k+1} \leftarrow \alpha u_k$, com $\alpha > 1$;
  $k \leftarrow k + 1$;
end

Método de Penalidade Interior

No método de barreira, deve-se construir uma função que penaliza tentativas de sair da região factível.

Função de Barreira

Uma função de barreira deve atender as seguintes condições:

$$\begin{cases} b(x) \to \infty, & \text{se } x \to \partial\mathcal{X} \\ b(x) \to 0, & \text{se } x \in \mathcal{X} \end{cases}$$

Método de Penalidade Interior

Função de Barreira

1. Uma possível função barreira é:

$$\begin{aligned} &\min f(x) \\ &\text{s.a } g(x) \leq 0 \qquad \Longrightarrow \min f(x) + (-u)\log[-g(x)] \end{aligned}$$

2. Outra opção:

$$\begin{aligned} &\min f(x) \\ &\text{s.a } g(x) \leq 0 \qquad \Longrightarrow \min f(x) + (-u)\frac{1}{[g(x)]} \end{aligned}$$

Método de Penalidade Interior

Função de Barreira

1. Uma possível função barreira é:

$$\begin{aligned} &\min f(x) \\ &\text{s.a } g(x) \leq 0 \qquad \Longrightarrow \min f(x) + (-u)\log[-g(x)] \end{aligned}$$

2. Outra opção:

$$\begin{aligned} &\min f(x) \\ &\text{s.a } g(x) \leq 0 \qquad \Longrightarrow \min f(x) + (-u)\frac{1}{[g(x)]} \end{aligned}$$

Método de Penalidade Interior

- $b(x)$ é uma função barreira não negativa e contínua em $\mathcal{F}$ e tende a infinito à medida que aproximamos da fronteira.
- Formulando o problema como a minimização de $f(x) + b(x, u)$, com $u \to 0^+$, gera-se uma sequência de soluções no interior de $\mathcal{X}$ que converge para x^*.

Método de Penalidade Interior

- $b(x)$ é uma função barreira não negativa e contínua em $\mathcal{F}$ e tende a infinito à medida que aproximamos da fronteira.
- Formulando o problema como a minimização de $f(x) + b(x, u)$, com $u \to 0^+$, gera-se uma sequência de soluções no interior de $\mathcal{X}$ que converge para x^*.

Função de Barreira

De forma geral temos::

$$b(x,u) = -u\left\{\sum_{j=1}^{q}\frac{1}{g_j(x)}\right\}$$

Observe que não é possível definir uma função barreira para restrições de igualdade.

Exemplo

Seja o problema min $f(x)$, com $f(x) = x$, sujeito a $g(x) = -x + 3 \leq 0$.

Solução

Usando a função barreira, tem-se:

$$f(x) + b(x, u) = x + u\frac{1}{x-3}$$

Assim:

$$\frac{d}{dx}\left(x + u\frac{1}{x-3}\right) = 1 - u\frac{1}{(x-3)^2} = 0 \;\rightarrow\; x^* = 3 + \sqrt{u}$$

Com $u \rightarrow 0^+$, temos $x^* \rightarrow 3^+$.

Método de Penalidade Interior

Algoritmo 2: Método de Penalidade Interior

Input: $\mathbf{x}_0 \in \mathcal{X}$, $u_0 > 0$, função-objetivo $f(\cdot)$ e restrições $\mathbf{g}(\cdot)$ e $\mathbf{h}(\cdot)$

$k \leftarrow 0$;
while $\neg$ *critério de parada* **do**
  Começando de $\mathbf{x}_k$, encontre $\mathbf{x}_{k+1} \leftarrow \arg\min_{\mathbf{x}} f(\mathbf{x}) + b(\mathbf{x}, u_k)$;
  $u_{k+1} \leftarrow \alpha u_k$, com $0 < \alpha < 1$;
  $k \leftarrow k + 1$;
end

Considerações

A teoria prevê que a solução converge para a solução ótima quando $u_k \to \infty$ (ou $u_k \to 0$),porém:

- A Hessiana da função-objetivo penalizada depende de u_k e pode apresentar ma-condicionamento numérico;
- A variação de u_k deve ser gradual;
- A precisão final fica bastante limitada.
- Usar termos de penalidade distintos para cada restrição evitando problemas de escala;
- Penalizar apenas a restrição mais violada.

Considerações

A teoria prevê que a solução converge para a solução ótima quando $u_k \to \infty$ (ou $u_k \to 0$),porém:

- A Hessiana da função-objetivo penalizada depende de u_k e pode apresentar ma-condicionamento numérico;
- A variação de u_k deve ser gradual;
- A precisão final fica bastante limitada.
- Usar termos de penalidade distintos para cada restrição evitando problemas de escala;
- Penalizar apenas a restrição mais violada.

Considerações

A teoria prevê que a solução converge para a solução ótima quando $u_k \to \infty$ (ou $u_k \to 0$),porém:

- A Hessiana da função-objetivo penalizada depende de u_k e pode apresentar ma-condicionamento numérico;
- A variação de u_k deve ser gradual;
- A precisão final fica bastante limitada.
- Usar termos de penalidade distintos para cada restrição evitando problemas de escala;
- Penalizar apenas a restrição mais violada.

Considerações

A teoria prevê que a solução converge para a solução ótima quando $u_k \to \infty$ (ou $u_k \to 0$),porém:

- A Hessiana da função-objetivo penalizada depende de u_k e pode apresentar ma-condicionamento numérico;
- A variação de u_k deve ser gradual;
- A precisão final fica bastante limitada.
- Usar termos de penalidade distintos para cada restrição evitando problemas de escala;
- Penalizar apenas a restrição mais violada.

Considerações

A teoria prevê que a solução converge para a solução ótima quando $u_k \to \infty$ (ou $u_k \to 0$),porém:

- A Hessiana da função-objetivo penalizada depende de u_k e pode apresentar ma-condicionamento numérico;
- A variação de u_k deve ser gradual;
- A precisão final fica bastante limitada.
- Usar termos de penalidade distintos para cada restrição evitando problemas de escala;
- Penalizar apenas a restrição mais violada.

Métodos Determinísticos

Métodos Determinísticos

Geram uma sequência determinística de possíveis soluções.

- Métodos baseados em derivadas:
  - Método do Gradiente;
  - Método de Newton;
  - Métodos Quasi-Newton;
  - Métodos de Gradientes Conjugados;
- Métodos sem derivadas:
  - Método Nelder-Mead Simplex;
  - Método Hooke-Jeeves.

Métodos Determinísticos

Geram uma sequência determinística de possíveis soluções.

- Métodos baseados em derivadas:
  - Método do Gradiente;
  - Método de Newton;
  - Métodos Quasi-Newton;
  - Métodos de Gradientes Conjugados;
- Métodos sem derivadas:
  - Método Nelder-Mead Simplex;
  - Método Hooke-Jeeves.

Métodos de direção de busca

Estrutura geral

A estrutura geral de métodos baseados em direção de busca obedece a seguinte forma:

$$x_{k+1} \leftarrow x_k + \alpha_k d_k$$

- Os métodos variam na forma como o passo α_k e a direção d_k são escolhidos.

Métodos de direção de busca

Estrutura geral

A estrutura geral de métodos baseados em direção de busca obedece a seguinte forma:

$$x_{k+1} \leftarrow x_k + \alpha_k d_k$$

- Os métodos variam na forma como o passo α_k e a direção d_k são escolhidos.

Método do Gradiente

O Método do Gradiente ou Método da Descida Mais Íngreme *(Steepest Descent Method)* é o método mais simples entre os métodos de direção de busca.

$$d_k = -\nabla f(x_k)$$

Método do Gradiente

- O algoritmo gera uma sequência monotônica $\{x_k, f(x_k)\}$ tal que $\nabla f(x_k) \to 0$ quando $k \to \infty$.
- O passo α_k é um escalar não negativo que minimiza $f(x_k + \alpha d_k)$, ou seja, representa um passo dado na direção minimizante d_k.

Método do Gradiente

Avaliação numérica do gradiente

Para avaliar o vetor gradiente numericamente, podemos usar a fórmula de diferenças finitas:

$$\left.\frac{\partial f}{\partial x_i}\right|_x \approx \frac{f(x+\delta_i e_i) - f(x)}{\delta_i}, \;\; i = 1, \ldots, n$$

Uma fórmula mais precisa é a diferença finita central:

$$\left.\frac{\partial f}{\partial x_i}\right|_x \approx \frac{f(x+\delta_i e_i) - f(x-\delta_i e_i)}{2\delta_i}, \;\; i = 1, \ldots, n$$

Método do Gradiente

Algoritmo 2: Método do Gradiente

Input: $\mathbf{x}_0 \in \mathcal{X}$, função-objetivo $f(\cdot)$

$k \leftarrow 0$;
while $\neg$ *critério de parada* **do**
  Calcule $\nabla f(\mathbf{x}_k)$;
  $\mathbf{d}_k \leftarrow -\nabla f(\mathbf{x}_k)$;
  $\alpha_k \leftarrow \arg\min_\alpha f(\mathbf{x}_k + \alpha \mathbf{d}_k)$;
  $\mathbf{x}_{k+1} \leftarrow \mathbf{x}_k + \alpha_k \mathbf{d}_k$;
  $k \leftarrow k + 1$;
end

Na prática, α_k é obtido com um método de minimização unidirecional.

Método do Gradiente

Algoritmo 2: Método do Gradiente

Input: $\mathbf{x}_0 \in \mathcal{X}$, função-objetivo $f(\cdot)$

$k \leftarrow 0$;
while $\neg$ *critério de parada* **do**
  Calcule $\nabla f(\mathbf{x}_k)$;
  $\mathbf{d}_k \leftarrow -\nabla f(\mathbf{x}_k)$;
  $\alpha_k \leftarrow \arg\min_\alpha f(\mathbf{x}_k + \alpha \mathbf{d}_k)$;
  $\mathbf{x}_{k+1} \leftarrow \mathbf{x}_k + \alpha_k \mathbf{d}_k$;
  $k \leftarrow k + 1$;
end

Na prática, α_k é obtido com um método de minimização unidirecional.

Método do Gradiente

Método de busca unidirecional

- Busca irrestrita;
- Busca exaustiva;
- Busca por bisseção;
- Busca dicotômica;
- Método de Fibonacci;
- Método da Seção Áurea;
- ...

Métodos baseados em reduções sucessivas de intervalos, exigem funções unimodais, porém não exigem diferenciabilidade

Método de busca unidirecional

Definição

$$\alpha^* = \min \theta(\alpha) \in \mathbb{R},\ \alpha \in [0, \infty]$$

$$\theta(\alpha) = f(x_k + \alpha d),\ x_k, d \in \mathbb{R}^n$$

Exemplo

Determinar z_1 que minimiza $f(x) = 2x_1^2 + x_2^2$ partindo de $z_0 = [1, 1]$ na direção $d = -\nabla f(x_0)$.

Busca por Bisseção

- Exclui metade do intervalo de busca a cada iteração.
- Especifica três pontos, x_1, m e x_2, igualmente espaçados no intervalo inicial $(a; b)$;
- Assumindo unimodalidade, tem-se:
  - Se $\theta_{x_1} < \theta_m < \theta_{x_2}$, deleta $(m; b)$, e faz-se $b = m$, e $m = x_1$;
  - Se $\theta_{x_1} > \theta_m > \theta_{x_2}$, deleta $(a; m)$, e faz-se $a = m$, e $m = x_2$;
  - Se $\theta_{x_1} > \theta_m$ e $\theta_{x_2} > \theta_m$, deleta $(a; x_1)$ e $(x_2; b)$, e faz-se $a = x_1$, e $b = x_2$;
- Especifica novos pontos x_1 e x_2, e continua o processo até $b_k - a_k \leq \epsilon$.

Busca dicotômica

- Representa uma busca sequencial onde os testes realizados influenciam na escolha dos testes subsequentes.
- Assume que o intervalo $(a; b)$ que cerca u^* seja conhecido.
- Escolhe dois pontos próximos ao centro do intervalo

$$x_1 = \frac{b+a}{2} - \frac{\delta}{2}, \quad x_2 = \frac{b+a}{2} + \frac{\delta}{2}, \quad \delta > 0$$

- Baseado na avaliação de $\theta(.)$ nestes dois pontos, exclui-se quase metade do intervalo.
- O processo se repete até atingir a precisão desejada, ou seja, $b_k - a_k \leq \epsilon$.

Método da Fibonacci ou Seção Áurea

- Assume unimodalidade de θ. e o conhecimento do intervalo $[a; b]$ que contém o ótimo.
- Define dois pontos, x_1 e x_2, no intervalo inicial $[a; b]$;
  - Se $\theta_{x_1} < \theta_{x_2}$, mínimo está em $[a; x_2]$;
  - Se $\theta_{x_1} > \theta_{x_2}$, mínimo está em $[x_1; b]$;
- Apenas um novo ponto precisará ser especificado nas iterações subsequentes.
- O número de avaliações de θ (ou a precisão desejada) deve ser especificado.

Método da Seção Áurea

- Para uma iteração i qualquer ($i = 0; 1; \ldots$), tem-se:

$$x_1^i = a_i + \lambda(b_i - a_i), \quad x_2^i = b_i - \lambda(b_i - a_i)$$

- Onde o fator λ é dado pela sequência de seção áurea;
- Se $\lambda = (3 - \sqrt(5))/2 = 0,382$, temos:
  - O comprimento do intervalo após k iterações é:

$$b_k - a_k = 0,618^k(b_0 - a_0)$$

- Os métodos de Fibonacci e seção áurea são os mais eficientes, porém o segundo é mais prático

Método de Newton

Se $f(x)$ é duas vezes diferenciável, usando a expansão em séries de Taylor em torno de x_k, temos:

$$f(x) = f(x_k) + \nabla f(x_k)'(x - x_k) + \frac{1}{2}(x - x_k)'H(x_k)(x - x_k) + \mathcal{O}$$

Assumindo a aproximação de 2a ordem, derivando e igualando a zero, obtemos:

$$\nabla f(x_k) + H(x_k)(x_{k+1} - x_k) = 0$$

$$x_{k+1} = x_k - H^{-1}(x_k)\nabla f(x_k)$$

Método de Newton

- Se $f(.)$ for quadrática, o método de Newton determina a solução ótima em um passo.
- A matriz $H^{-1}(x_k)$ pode ser interpretada como uma "correção" aplicada à direção $\nabla f(x_k)$, levando em conta a curvatura da função.
- Para casos gerais, com funções não quadráticas, deve-se determinar o passo ótimo α_k: $x_{k+1} \leftarrow x_k + \alpha_k d_k$, $d_k = -H^{-1}(x_k)\nabla f(x_k)$

Método de Newton

Algoritmo 3: Método de Newton

Input: $\mathbf{x}_0 \in \mathcal{X}$, função-objetivo $f(\cdot)$

$k \leftarrow 0$;
while $\neg$ *critério de parada* **do**
  Calcule $\nabla f(\mathbf{x}_k)$;
  Calcule $\mathbf{H}(\mathbf{x}_k)$;
  $\mathbf{d}_k \leftarrow -\mathbf{H}^{-1}(\mathbf{x}_k)\nabla f(\mathbf{x}_k)$;
  $\alpha_k \leftarrow \arg\min_\alpha f(\mathbf{x}_k + \alpha \mathbf{d}_k)$;
  $\mathbf{x}_{k+1} \leftarrow \mathbf{x}_k + \alpha_k \mathbf{d}_k$;
  $k \leftarrow k + 1$;
end

Método de Newton

- O método apresenta convergência quadrática;
- A convergência do método de Newton é atendida sob duas premissas:
  1. que $H(x_k)$ seja não singular, ou seja, possua inversa;
  2. que $H(x_k)$ seja definida positiva, para garantir que $d_k = -H^{-1}(x_k)\nabla f(x_k)$ seja uma direção minimizante.

Método de Newton

- Dificuldades:
  - Necessita do cálculo da matriz inversa;
  - Mal-condicionamento numérico da matriz Hessiana dificulta o cálculo de sua inversa;
  - Derivadas numéricas: maiores erros numéricos e muitas avaliações de função para a aproximação por diferenças finitas.

Métodos Determinísticos

- Métodos baseados em derivadas:
  - Método do Gradiente;
  - Método de Newton;
  - Métodos Quasi-Newton;
  - Métodos de Gradientes Conjugados;
- Métodos sem derivadas:
  - Método Nelder-Mead Simplex;
  - Método Hooke-Jeeves.

Referências:

- Takahashi, R.H.C.; Notas de Aula: Otimização Escalar e Vetorial. http://www.decom.ufop.br/moreira/site_media/uploads/arquivos/Archive.zip
- J. A. Ramírez, F. Campelo, F. G. Guimarães, Lucas S. Batista, Ricardo H. C. Takahashi, Notas de Aula de Otimização, 2010. http://www.decom.ufop.br/moreira/site_media/uploads/arquivos/nota1.pdf
- Izmailov, A, Solodov, V.M , Otimização Vol.1 Condições de Otimalidade, Elementos de Análise Convexa e de Dualidade. Editora IMPA, 2012.
- Izmailov, A, Solodov, V.M , Otimização Vol. 2. Métodos Computacionais. Editora IMPA, 2012.