Aula 03 - Revisão sobre métodos de otimização

Clodoaldo A. M. Lima

02 de setembro de 2021

Programa de Pós-Graduação em Sistemas de Informação
Mestrado acadêmico - EACH - USP
http://ppgsi.each.usp.br

Conteúdo

- Otimização Restrita
- Otimização Irrestrita

Preliminares

Tipos de problema de otimização

Problema de otimização irrestrita

$$(P_1) \quad \underset{x}{\text{minimize}} \quad f(x)$$
$$\text{s. t.} \quad x \in X.$$

onde $x = [x_i, \cdots, x_n] \in \mathcal{R}^n$, $f(x) : \mathcal{R}^n \to \mathcal{R}$, e X é um conjunto fechado (usualmente X$= \mathcal{R}^n$)

Problema de otimização restrita

$$(P_2) \quad \underset{x}{\text{minimize}} \quad f(x)$$
$$\text{subject to} \quad g_i(x) \leq 0, \ i = 1, \ldots, m.$$
$$h_i(x) = 0, \ i = 1, \ldots, l.$$
$$x \in X,$$

onde $g_1(x), \cdots, g_m(x)$, $h_1(x), \cdots, h_l(x) : \mathcal{R}^n \to \mathcal{R}$
Seja $g(x) = (g_1(x), \cdots, g_m(x)) : \mathcal{R}^n \to \mathcal{R}^m$, $h(x) = (h_1(x), \cdots, h_l(x)) : \mathcal{R}^n \to \mathcal{R}^l$

Preliminares

Então (P_2) pode ser escrito como

$$
\begin{aligned}
(P_2) \quad \underset{x}{\text{minimize}} \quad & f(x) \\
\text{subject to} \quad & g(x) \leq 0, \\
& h(x) = 0. \\
& x \in X,
\end{aligned}
$$

Nós dissemos que x é uma solução factível de (P_2), se $g(x) \leq 0$, $h(x) = 0$, e $x \in X$

Ótimo local, global

A bola centrada em $\bar{x}$ com raio ϵ é o conjunto:

$$B(\bar{x}, \epsilon) := \{x || x - \bar{x}|| \leq \epsilon\}$$

Considere o seguinte problema de otimização sobre o conjunto $\mathcal{F}$

$$(P_1) \quad \underset{x}{\text{minimize}} \quad f(x)$$
$$\text{s. t.} \quad x \in \mathcal{F}.$$

Nós temos a seguinte definição de mínimo/máximo, local/global, estrito/não estrito.

Definição 1

$x \in \mathcal{F}$ é um mínimo local de P_1 se lá existe $\epsilon > 0$ tal que $f(x) \leq f(y)$ para $\forall y \in B(x, \epsilon) \cap \mathcal{F}$

Ótimo local, global

Definição 2

$x \in \mathcal{F}$ é um mínimo global de P_1 se $f(x) \leq f(y)$ para $\forall y \in \mathcal{F}$

Definição 3

$x \in \mathcal{F}$ é um mínimo local estrito de P_1 se lá existe $\epsilon > 0$ tal que $f(x) < f(y)$ para $\forall y \in B(x, \epsilon) \cap \mathcal{F}, y \neq x$

Definição 4

$x \in \mathcal{F}$ é um mínimo global estrito de P_1 se $f(x) < f(y)$ para $\forall y \in \mathcal{F}, y \neq x$

Definição 5

$x \in \mathcal{F}$ é um máximo local de P_1 se lá existe $\epsilon > 0$ tal que $f(x) \geq f(y)$ para $\forall y \in B(x, \epsilon) \cap \mathcal{F}, y \neq x$

Definição 6

$x \in \mathcal{F}$ é um máximo global de P_1 se $f(x) \geq f(y)$ para $\forall y \in \mathcal{F}, y \neq x$

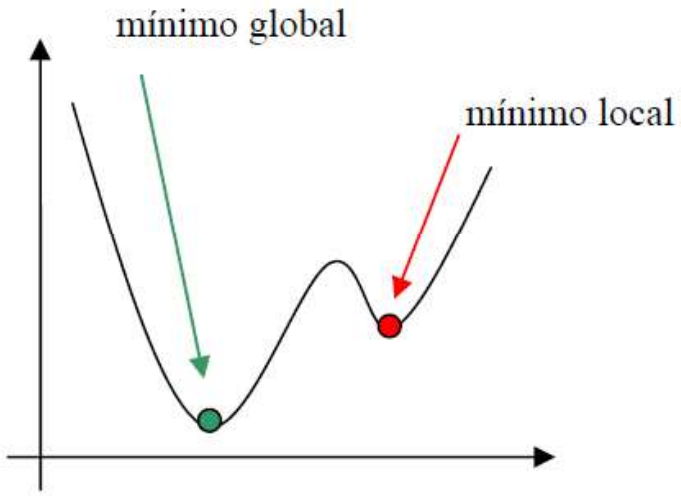
Ótimo local, global

Definição 7

$x \in \mathcal{F}$ é um máximo local estrito de P_1 se lá existe $\epsilon > 0$ tal que $f(x) > f(y)$ para $\forall y \in B(x, \epsilon) \cap \mathcal{F}, y \neq x$

Definição 8

$x \in \mathcal{F}$ é um máximo global estrito de P_1 se lá existe $\epsilon > 0$ tal que $f(x) > f(y)$ para $\forall y \in \mathcal{F}, y \neq x$



Gradiente e Hessiana

Seja $f(x) : X \to \mathcal{R}$, onde $X \subset \mathcal{R}^n$ é fechado.

$f(x)$ é diferenciável em $\bar{x} \in X$ se lá existe um vetor $\nabla f(\bar{x})$ (o gradiente de $f(x)$ em $\bar{x}$) tal que para cada $x \in X$

$$f(x) = f(\bar{x}) + \nabla f(\bar{x})^T (x - \bar{x}) + ||x - \bar{x}|| \alpha(\bar{x}, x - \bar{x}),$$

e $\lim_{y \to 0} \alpha(\bar{x}, y) = 0$.
$f(x)$ é diferencíavel $\forall \bar{x} \in X$. O vetor gradiente é o vetor de derivadas parciais:

$$\nabla f(\bar{x}) = \left[\frac{\partial f(\bar{x})}{\partial x_1}, \cdots, \frac{\partial f(\bar{x})}{\partial x_n} \right]^T$$

Exemplo 1

Seja $f(x) = 3x_1^2 x_2^3 + x_2^2 x_3^3$. Então

$$\nabla f(x) = \left[6x_1 x_2^3, 9x_1^2 x_2^2 + 2x_2 x_3^3, 3x_2^2 x_3^2 \right]^T$$

Gradiente e Hessiana

A derivada direcional de $f(x)$ em $\bar{x}$ na direção d é:

$$\lim_{\lambda \to 0} \frac{f(\bar{x} + \lambda) - f(\bar{x})}{\lambda} = \nabla f(\bar{x})^T d$$

A função f(x) é duas diferenciável em $\bar{x} \in X$ se existe um vetor $\nabla f(\bar{x})$ e uma matriz simétrica $n \times n$ $H(\bar{x})$ (a hessiana de $f(x)$ em $\bar{x}$) tal que para cada $x \in X$

$$f(x) = f(\bar{x}) + \nabla f(\bar{x})^T (x - \bar{x}) + \frac{1}{2}(x - \bar{x})^T H(\bar{x})(x - \bar{x}) + ||(x - \bar{x})||^2 \alpha(\bar{x}, x - \bar{x}),$$

e $\lim_{y \to 0} \alpha(\bar{x}, y) = 0$.

$f(x)$ é duas vezes diferenciável sobre X se $f(x)$ é duas vezes diferenciavel para todo $\bar{x} \in X$. A Hessiana é a matriz de derivada segunda parcial.

$$H(\bar{x})_{ij} = \frac{\partial^2 f(\bar{x})}{\partial x_i \partial x_j}$$

Exemplo 2

Seja $f(x) = 3x_1^2x_2^3 + x_2^2x_3^3$. Então

$$\nabla f(x) = \begin{bmatrix} 6x_1x_2^3 \\ 9x_1^2x_2^2 + 2x_2x_3^3 \\ 3x_2^2x_3^2 \end{bmatrix}.$$

$$H(x) = \begin{bmatrix} 6x_2^3 & 18x_1x_2^2 & 0 \\ 18x_1x_2^2 & 18x_1^2x_2 + 2x_3^3 & 6x_2x_3^2 \\ 0 & 6x_2x_3^2 & 6x_2^2x_3 \end{bmatrix}.$$

Matrizes Definida Positiva e Semidefinida Positiva

Um matriz M $n \times n$ é chamada

- Definida positiva se $x^T Mx > 0 \ \forall x \in \mathcal{R}^n, \ x \neq 0$
- Semidefinida positiva se $x^T Mx \geq 0 \ \forall x \in \mathcal{R}^n, \ x \neq 0$
- Definida negativa se $x^T Mx < 0 \ \forall x \in \mathcal{R}^n, \ x \neq 0$
- Semidefinida negativa se $x^T Mx \leq 0 \ \forall x \in \mathcal{R}^n, \ x \neq 0$
- Indefinida se este $x, y \in \mathcal{R}^n$ para o qual $x^T Mx > 0$ e $y^T My < 0$

Nós dissemos que M é *SPD* se M é simétrica e definida positiva. Similarmente, nós dissemos que M é *SPSD* se M é simétrica e semidefinida positiva.

Example 3

$$M = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}$$

é definida positiva

Exemplo 4

$$M = \begin{bmatrix} 8 & -1 \\ -1 & 1 \end{bmatrix}$$

é definida positivo. Para ver isto, observe que para $x \neq 0$,

$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$x^T M x = 8x_1^2 - 2x_1x_2 + x_2^2 = 7x_1^2 + (x_1 - x_2)^2 > 0$$

Existência de soluções ótimas

Muitos dos tópicos aqui estão concentrados com

- existência de soluções ótimas
- caracterização de solução ótimas
- algoritmos para computar a solução ótima

Para ilustrar a questão surgida no primeiro tópico, considere o seguinte problema de otimização:

$$(P_2) \quad \underset{x}{\text{minimize}} \quad \frac{1+x}{2x}$$
$$\text{subject to} \quad x \geq 1,$$

Aqui não há solução ótima por que a região factível não é limitada

Existência de soluções ótimas

$$(P_2) \quad \underset{x}{\text{minimize}} \quad \frac{1}{x}$$
$$\text{subject to} \quad 1 \geq x < 2,$$

Aqui não há solução ótima por que a região factível não é fechada ($x < 2$)

$$(P_2) \quad \underset{x}{\text{minimize}} \quad f(x)$$
$$\text{subject to} \quad 1 \leq x \leq 2,$$

$$f(x) = \begin{cases} 1/x, & \text{if} \quad x < 2 \\ 1, & \text{if} \quad x = 2 \end{cases}$$

Teorema

Teorema para Sequência - Weierstrass

Seja x_k, $k \to \infty$ é uma sequencia infinita de pontos no conjunto compacto (fechado e limitado) $\mathcal{F}$. Então alguma subsequência de pontos x_{kj} converge para um ponto contido em $\mathcal{F}$

Teorema para funções - Weierstrass

Seja $f(x)$ é uma função de valor real sobre conjunto compacto não vazio $\mathcal{F} \subset \mathcal{R}^n$. Então $\mathcal{F}$ contém um ponto que minimiza (maximiza) $f(x)$ sobre o conjunto $\mathcal{F}$

Condições de otimalidade para problemas irrestritos

$$(P_1) \quad \underset{x}{\text{minimize}} \quad f(x)$$
$$\text{s. t.} \quad x \in X.$$

onde $x = [x_i, \cdots, x_n] \in \mathcal{R}^n$, $f(x) : \mathcal{R}^n \to \mathcal{R}$, e X é um conjunto fechado (usualmente X$= \mathcal{R}^n$)

Definição

A direção $\bar{d}$ é chamada uma direção descendente de $f(x)$ em $x = \bar{x}$ se

$$f(\bar{x} + \epsilon \bar{d}) < f(\bar{x})$$

$\forall \epsilon > 0$ e suficientemente pequeno

Condições de otimalidade para problemas irrestritos

Teorema

Supondo que $f(x)$ seja diferenciavel em $\bar{x}$. Se há um vetor d tal que $\nabla f(\bar{x})^T d < 0$, então $\forall \lambda > 0$ e suficientemente pequeno, $f(\bar{x} + \lambda d) < f(\bar{x})$, e assim d é uma direção descente de $f(x)$ em $\bar{x}$.

Prova

$$f(\bar{x} + \lambda d) = f(\bar{x}) + \lambda \nabla f(\bar{x})^T d + \lambda ||d|| \alpha(\bar{x}, \lambda d),$$

onde $\alpha(\bar{x}, \lambda d) \to 0$ como $\lambda \to 0$. Rearranjando

$$\frac{f(\bar{x} + \lambda d) - f(\bar{x})}{\lambda} = \nabla f(\bar{x})^T d + ||d|| \alpha(\bar{x}, \lambda d),$$

Uma vez que $\nabla f(\bar{x})^T d < 0$ e $\alpha(\bar{x}, \lambda d) \to 0$ quando $\lambda \to 0$, $f(\bar{x} + \lambda d) - f(\bar{x}) < 0$ $\forall \lambda > 0$ suficientemente pequeno.

Condições de otimalidade para problemas irrestritos

Corolário

Supondo $f(x)$ seja diferenciável em $\bar{x}$. Se $\bar{x}$ é um mínimo local, então $\nabla f(\bar{x}) = 0$.

Prova

Se fosse verdade que $\nabla f(\bar{x}) \neq 0$, então $d = -\nabla f(\bar{x})$ seria uma direção descendente, ao passo que $\bar{x}$ não será um mínimo local.

Importante

O corolário acima é a condição necessária de primeira ordem para um problema minimização irrestrito. O teorema a seguir é uma condição de otimalidade de segunda ordem

Condições de otimalidade para problemas irrestritos

Teorema

Suponha que $f(x)$ seja duas vezes diferenciável em $\bar{x} \in X$. Se $\bar{x}$ é um mínimo local, então $\nabla f(\bar{x}) = 0$ e $H(\bar{x})$ é semidefinida positiva.

Prova

Da condição necessária de primeira ordem, $\nabla f(\bar{x}) = 0$. Suponha que $H(\bar{x})$ não é definida positiva. Então lá existe d tal que $d^T H d < 0$. Nós temos

$$f(\bar{x} + \lambda d) = f(\bar{x}) + \lambda \nabla f(x)^T d + \frac{1}{2}\lambda^2 d^T H(\bar{x}) d + \lambda^2 ||d||^2 \alpha(\bar{x}, \lambda d)$$

$$f(\bar{x} + \lambda d) = f(\bar{x}) + \frac{1}{2}\lambda^2 d^T H(\bar{x}) d + \lambda^2 ||d||^2 \alpha(\bar{x}, \lambda d)$$

onde $\alpha(\bar{x}, \lambda d) \to 0$, quando $\lambda \to 0$. Rearranjando,

$$\frac{f(\bar{x} + \lambda d) - f(\bar{x})}{\lambda^2} = \frac{1}{2} d^T H(\bar{x}) d + ||d||^2 \alpha(\bar{x}, \lambda d)$$

Uma vez que $d^T H(x) d < 0$ e $\alpha(\bar{x}, \lambda d) \to 0$, $f(\bar{x} + \lambda d) - f(\bar{x}) < 0 \ \forall \lambda > 0$ e suficientemente pequeno, produzindo a contradição desejada.

Exemplo

$f(x) = \frac{1}{2}x_1^2 + x_1x_2 + 2x_2^2 - 4x_1 - 4x_2 - x_2^3$

Então

$$\nabla f(x) = \left(x_1 + x_2 - 4, x_1 + 4x_2 - 4 - 3x_2^2\right)^T,$$

e

$$H(x)) = \begin{bmatrix} 1 & 1 \\ 1 & 4 - 6x_2 \end{bmatrix}$$

$\nabla f(x) = 0$ tem exatamente duas soluções: $\bar{x} = (4, 0)$ e $\tilde{x} = (3, 1)$. Mas

$$H(\tilde{x}) = \begin{bmatrix} 1 & 1 \\ 1 & -2 \end{bmatrix}$$

é indefinita, portanto, a única solução candidata para um mínimo local é $\bar{x} = (4, 0)$

Exemplo

Encontre os candidatos a mínimos e máximos locais da função

$$f(x) = (2x_1 - x_2)^2 + (3x_1 - x_3)^2 + (3x_2 - 2x_3)^2$$

Condições de otimalidade para problemas irrestritos

Teorema

Supondo que $f(x)$ é duas vezes diferenciavel em $\bar{x}$. Se $\nabla f(\bar{x}) = 0$ e $H(\bar{x})$ é definida positiva, então $\bar{x}$ é um mínimo local (estrito).

Prova

Exercício para casa

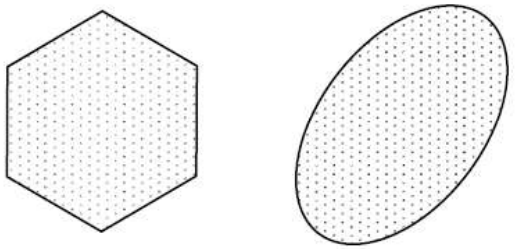
Observação

- Se $\nabla f(\bar{x}) = 0$ e $H(\bar{x})$ é definida negativa, então $\bar{x}$ é um máximo local
- Se $\nabla f(\bar{x}) = 0$ e $H(\bar{x})$ é semidefinida positiva, nós não podemos dizer com certeza que $\bar{x}$ é um mínimo local

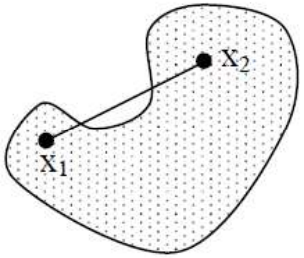
Convexidade e Minimização

Seja $x, y \in \mathcal{R}^n$. Pontos formados por $\lambda x + (1 - \lambda)y \ \forall \lambda \in [0, 1]$ são chamados de combinação convexa de x e y

Um conjunto $S \subset \mathcal{R}^n$ é chamada um conjunto convexo se $\forall x, y \in S$ e para $\lambda \in [0, 1]$, $\lambda x + (1 - \lambda)y \in S$



a) Convexo



b) Não-Convexo

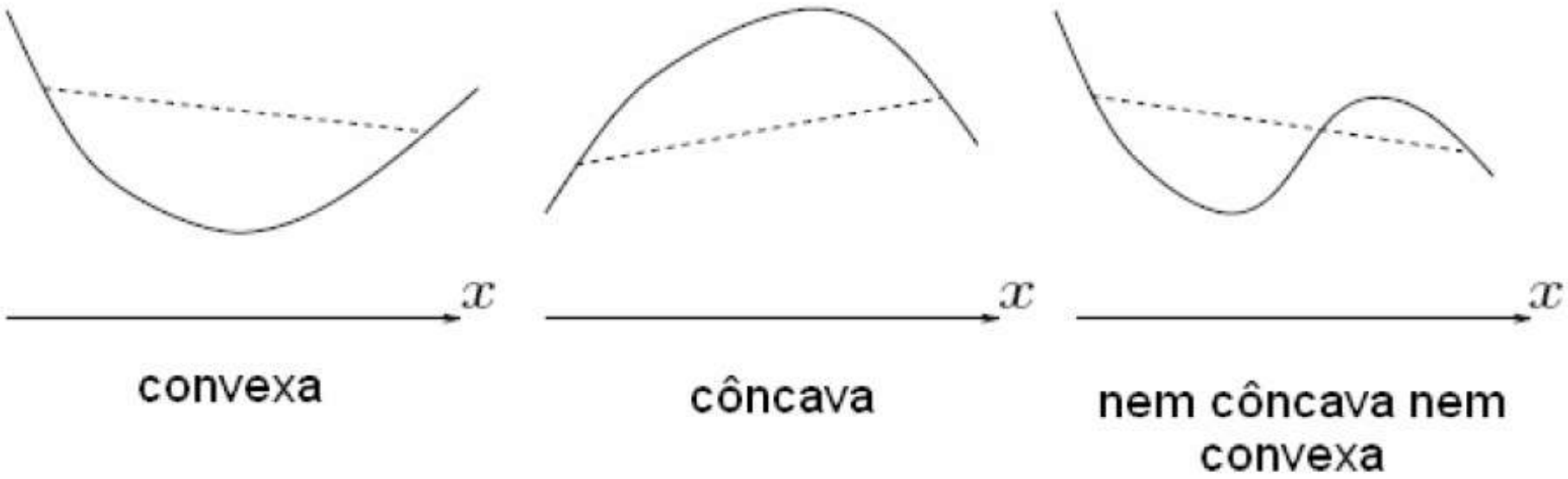
Convexidade e Minimização

Uma função $f(x) : S \to \mathcal{R}$, onde S é um conjunto convexo não vazio, é uma função convexa se

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y)$$

$\forall x, y, \in S$ e $\forall \lambda \in [0, 1]$

Uma função $f(x)$ como acima é chamada de uma função estritamente côncava se a desigualdade acima é estrita para $\forall x \neq y$ e $\lambda \in (0, 1)$



Convexidade e Minimização

$$(P_1) \quad \underset{x}{\text{minimize}} \quad f(x)$$
$$\text{s. t.} \quad x \in X.$$

Teorema

Supondo S é um conjunto de convexo, $f(x) : S \to \mathcal{R}$ é uma função convexa, e $\bar{x}$ é um mínimo local de (P_1). Então $\bar{x}$ é um mínimo global de $f(x)$ sobre S.

Teorema

Supondo S é conjunto convexo fechado não vazio, e $f(x) : S \to \mathcal{R}$ é diferenciavel. Então $f(x)$ é uma função convexa se e somente se $f(x)$ satisfaz a seguinte desiqualidade:

$$f(y) \geq f(x) + \nabla f(x)^T (y - x)$$

$\forall x, y \in S$

Prova

Exercício para casa

Convexidade e Minimização

Teorema

Supondo S é um conjunto convexo fechado não vazio, e $f(x) : S \to \mathcal{R}$ é duas vezes diferenciavel. Seja $H(x)$ denota a hessiana de $f(x)$. Então $f(x)$ é convexa se e somente se $H(x)$ é semidefinida positiva $\forall x \in S$

Teorema

Supondo $f(x) : X \to \mathcal{R}$ é convexa e diferenciavel sobre X. Então $\bar{x} \in X$ é um mínimo global se e somente se $\nabla f(\bar{x}) = 0$.

Gradiente descendente

O problema que nós

$$(P_1) \quad \underset{x}{\text{minimize}} \quad f(x)$$
$$\text{s. t.} \quad x \in \mathcal{R}^n.$$

onde $f(x)$ é diferenciável.

Se $x = \bar{x}$ é um dado ponto, $f(x)$ pode ser aproximado por sua expansão linear

$$f(\bar{x} + d) \approx f(\bar{x}) + \nabla f(\bar{x})^T d$$

Se d é pequeno, isto é, $||d||$ é pequeno. Observe que se a aproximação na equação acima é boa, então nós queremos escolher d tal que o produto interno $\nabla f(\bar{x})^T d$ é tão pequena quanto possível. Vamos normalizar d tal que $||d|| = 1$. Então entre todas as direções d com norma $||d|| = 1$, a direção

$$\tilde{d} = \frac{-\nabla f(\bar{x})}{||\nabla f(\bar{x})||}$$

torna o produto interno menor com o gradiente $\nabla f(\bar{x})$.

Gradiente descendente

Este fato segue da seguinte desigualdade:

$$\nabla f(x)^T d \geq -||\nabla f(\bar{x})||.||d|| = \nabla f(\bar{x})^T \left(\frac{-\nabla f(\bar{x})}{||\nabla f(\bar{x})||} \right) = \nabla f(\bar{x})^T \tilde{d}$$

Para esta razão a direção não normalizada:

$$\bar{d} = -\nabla f(\bar{x})$$

é chamada a direção de maior descida no ponto $\bar{x}$
Observe que $\bar{d} = -\nabla f(\bar{x})$ é uma direção descendente enquanto $\nabla f(\bar{x}) \neq 0$.
Para ver isto, simplesmente observe que $d^T \nabla f(\bar{x}) = -(\nabla f(\bar{x}))^T \nabla f(\bar{x}) < 0$ enquanto $\nabla f(\bar{x}) \neq 0$.

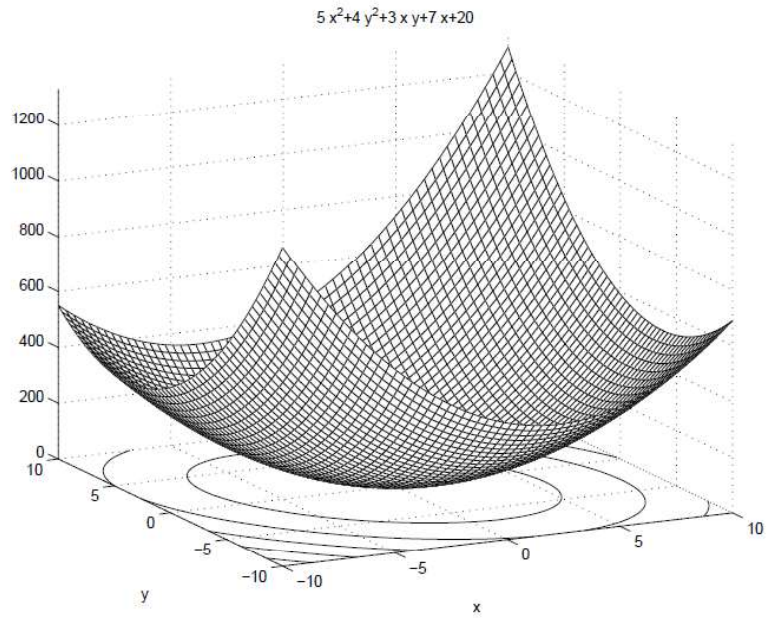
Gradiente descendente

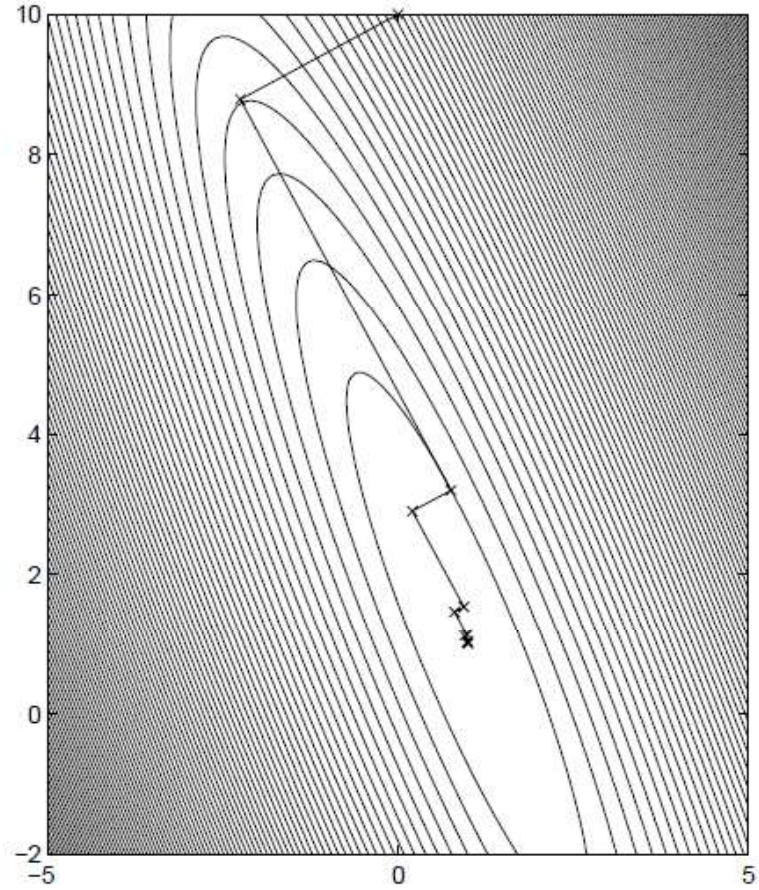
- Passo 0 - Dado x^0, faça $k := 0$
- Passo 1 - $d^k := -\nabla f(x^k)$. Se $d^k = 0$, então pare
- Passo 2 - Solucione $\min_\alpha f(x^k + \alpha d^k)$ para o tamanho do passo α^k, talvez escolhida por uma busca linear exata ou inexata
- Passo 3 - Faça $x^{k+1} \leftarrow x^k + \alpha^k d^k$, $k \leftarrow k + 1$. Vá para o passo 1.

Comportamento típico

$f(x_1, x_2) = 5x_1^2 + x_2^2 + 4x_1x_2 - 14x_1 - 6x_2 + 20$

Esta função tem sua solução ótima em $x^* = (x_1^*, x_2^*) = (1, 1)$ e $f(1, 1) = 10$





Um algoritmo de bisseção para uma busca linear de uma função convexa

Supondo que $f(x)$ é uma função convexa continuamente diferenciável e desejamos solucionar:

$$\bar{\alpha} := \underset{\alpha}{arg\min} \qquad f(\bar{x} + \alpha \bar{d}),$$

Gradiente Descendente

O problema

$$(P_1) \quad \underset{x}{\text{minimize}} \quad f(x)$$
$$\text{s. t.} \quad x \in X.$$

onde $\bar{x}$ é nossa iteração atual, $\bar{d}$ é a direção atual gerada por um algoritmo que busca minimizar $f(x)$

Suponha que $\bar{d}$ é uma direção descendente de $f(x)$ em $x = \bar{x}$, isto é

$$f(\bar{x} + \epsilon \bar{d}) < f(\bar{x})$$

$\forall \epsilon > 0$ e suficientemente pequena.

Algoritmo de bisseção

Seja

$$h(\alpha) := f(\bar{x} + \alpha \bar{d})$$

$h(\alpha)$ é uma função convexa na variável escala α, e nosso problema é solucionado por

$$\bar{\alpha} = arg \min_{\alpha} h(\alpha)$$

É elementar mostrar que

$$h'(\alpha) = \nabla f(\bar{x} + \alpha \bar{d})^T d$$

Proposição

$$h'(0) < 0$$

Como $h'(\alpha)$ é uma função crescente monotônica de α, nós temos que:

Proposição

$h'(\alpha)$ é uma função crescente monotônica de α

Algoritmo de bisseção

Como $h'(\alpha)$ é uma função crescente monotônica, nós podemos aproximadamente computar $\bar{\alpha}$, o ponto que satisfaz $h(\bar{\alpha}) = 0$, por um método de bisseção adequado. Supondo que nós sabemos um valor $\hat{\alpha}$ tal que $h'(\hat{\alpha}) > 0$. Desde que $h'(0) < 0$ e $h'(\hat{\alpha}) > 0$, o valor do meio $\tilde{\alpha} = \frac{0+\hat{\alpha}}{2}$ é um ponto de teste adequado. Observe o seguinte:

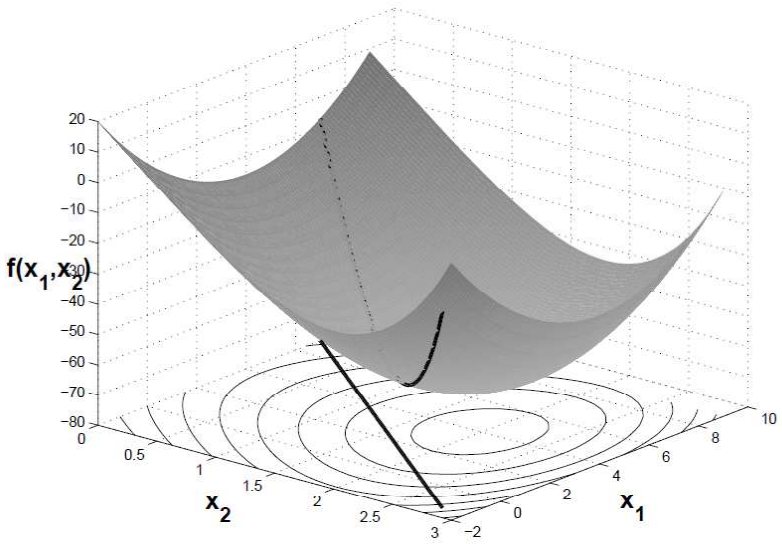
- se $h'(\tilde{\alpha}) = 0$, podemos parar
- se $h'(\tilde{\alpha}) > 0$, nós podemos colocar $\bar{\alpha}$ no intervalo $(0, \tilde{\alpha})$
- se $h'(\tilde{\alpha}) < 0$, nós podemos colocar $\bar{\alpha}$ no intervalo $(\tilde{\alpha}, \bar{\alpha})$

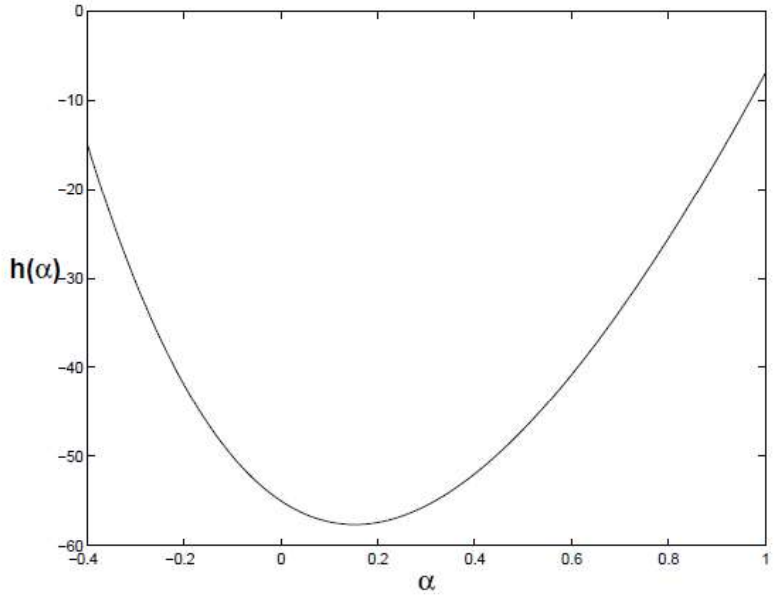
Este conduz para o seguinte algoritmo de bisseção para minimização de $h(\alpha) = f(\bar{x} + \alpha\bar{d})$ por solucionar a equação $h'(\alpha) \approx 0$

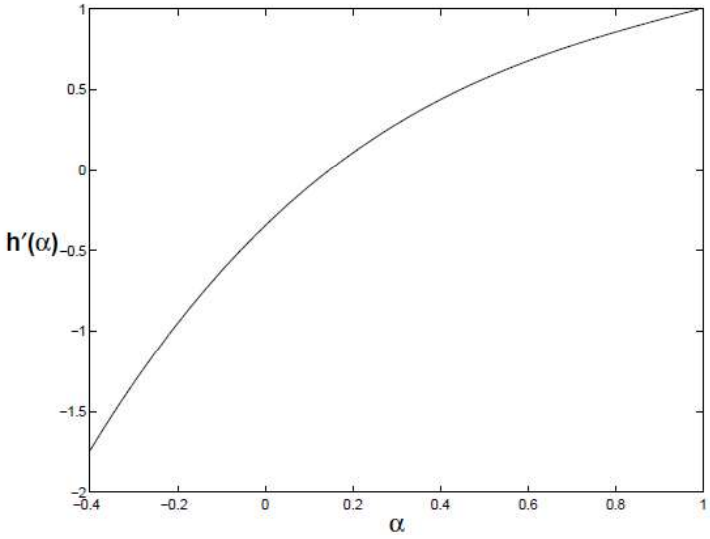
Algoritmo Bisseção

- Passo 0 - Faça $k = 0$, $\alpha_l := 0$, $\alpha_u := \hat{\alpha}$
- Passo 1 - Faça $\tilde{\alpha} = \frac{\alpha_u + \alpha_l}{2}$ e compute $h'(\tilde{\alpha})$
  Se $h'(\tilde{\alpha}) > 0$, faça $\alpha_u := \tilde{\alpha}$. Faça $k \leftarrow k + 1$
  Se $h'(\tilde{\alpha}) < 0$, faça $\alpha_l := \tilde{\alpha}$. Faça $k \leftarrow k + 1$
  Se $h'(\tilde{\alpha}) = 0$, pare

Algoritmo Bisseção







Algorítmo Bisseção

Proposição

Após toda iteração do algoritmo bisseção, o intervalo $[\alpha_l, \alpha_u]$ deve conter um ponto $\bar{\alpha}$ tal que $h'(\bar{\alpha}) = 0$

Proposição

Na k-ésima iteração do algoritmo bisseção, o comprimento do intervalo atual $[\alpha_l, \alpha_u]$ é

$$L = \left(\frac{1}{2}\right)^k (\hat{\alpha})$$

Proposição

O valor α tal que $|\alpha - \bar{\alpha}| \leq \epsilon$ pode ser encontrado em no máximo

$$\left\lceil \log_2 \left(\frac{\hat{\alpha}}{\epsilon}\right) \right\rceil$$

passos do algoritmo de bisseção.

Computando $\hat{\alpha}$ para o qual $h'(\hat{\alpha}) > 0$

Suponde que nós não temos um valor conveniente de $\hat{\alpha}$ para o qual $h'(\hat{\alpha}) > 0$.
Uma forma é pegar um candidato inicial de $\hat{\alpha}$ e calcular $h'(\bar{\alpha})$.
Se $h'(\alpha) > 0$, então proceda para o algoritmo de bisseção; se $h'(\alpha) \leq 0$, então faça $\hat{alpha} \leftarrow 2\hat{\alpha}$ e repita o processo.

Critério de parada para o algoritmo bisseção

Na prática, nós necessitamos executar o algoritmo bisseção com um critério de parada. Alguns critérios de parada relevante são:

- Pare após um numero fixo de iterações. Este é parado quando $k = \bar{K}$, onde $\bar{K}$ especificado pelo usuário
- Pare quando o intervalo torna-se pequeno. Isto é, pare quando $\alpha_u - \alpha_l \leq \epsilon$, onde ϵ é especificado pelo usuário
- Pare quando $|h'(\tilde{\alpha})|$ torna-se pequeno. Isto, pare quando $|h'(\tilde{\alpha})| \leq \epsilon$, onde ϵ é especificado pelo usuário

Este terceiro critério de parada tipicamente produz o melhor resultado na prática.

Método de Newton

Supondo que nós queremos resolver

$$(P_1) \quad \underset{x}{\text{minimize}} \quad f(x)$$
$$\text{s. t.} \quad x \in X.$$

Em $x = \bar{x}$, $f(x)$ pode ser aproximado por:

$$f(x) \approx h(x) := f(\bar{x}) + \nabla f(\bar{x})^T (x - \bar{x}) + \frac{1}{2}(x - \bar{x}) H(x) (x - \bar{x}),$$

a qual é uma expansão quadrática de Taylor de $f(x)$ em $x = \bar{x}$. Aqui $\nabla f(x)$ é o gradiente de $f(x)$ e $H(x)$ é a hessiana de $f(x)$
Observe que $h(x)$ é uma função quadrática, a qual é minimizada por solucionando $\nabla h(x) = 0$.

Método de Newton

Uma vez que o gradiente de $h(x)$ é:

$$\nabla h(x) = \nabla f(\bar{x}) + H(\bar{x})(x - \bar{x})$$

Nos portanto estamos motivado a solucionar:

$$\nabla f(\bar{x}) + H(\bar{x})(x - \bar{x}) = 0$$

que produz

$$(x - \bar{x}) = -H(\bar{x})^{-1}\nabla f(\bar{x})$$

A direção $-H(\bar{x})^{-1}\nabla f(\bar{x})$ é chamada de direção de Newton, ou passo de Newton em $x - \bar{x}$
Este conduz para o seguinte algoritmo para solução de (P_1)

Método de Newton

- Passo 0 - Dado x^0, faça $k \leftarrow 0$
- Passo 1 - $d^k = -H(x^k)^{-1}\nabla f(x^k)$. Se $d^k = 0$, então pare.
- Passo 2 - Escolha o tamanho do passo $\alpha^k = 1$
- Passo 3 - Faça $x^{k+1} \leftarrow x^k + \alpha^k d^k$, $k \leftarrow k + 1$. Vá para o passo 1.

Observe o seguinte

- O método assume $H(x^k)$ é não singular a cada iteração
- Não há garantia que $f(x^{k+1}) \leq f(x^k)$
- O passo 2 pode ser aumentado por uma busca linear de $f(x^k + \alpha d^k)$ para encontrar um valor ótimo do tamanho do passo, parametro α

Método de Newton

Proposição

Se $H(x)$ é simétrico e definido positivo e $d := -H(x)^{-1}\nabla f(x) \neq 0$, então d é uma direção descendente, isto é, $f(x + \alpha d) < f(x)$ para todos os valores de α suficientemente pequeno.

Método de Newton

- A matriz hessiana não precisa ser invertida explicitamente, em vez disso resolve-se o sistema linear.

$$\nabla f(\bar{x}) + H(\bar{x})(x - \bar{x}) = 0$$

$$H(\bar{x})(x - \bar{x}) = -\nabla f(\bar{x})$$

$$H(\bar{x})\delta = -\nabla f(\bar{x})$$

$$x^{k+1} \leftarrow x^k + \alpha^k \delta$$

- Convergência quadrática mas necessita de uma boa estimativa inicial da solução
- Necessita da segunda derivada da função

Exemplo

Usar o método de Newton para minimizar $f(x) = 0.5x_1^2 + 2.5x_2^2$
Gradiente e matriz hessiana são dados por

$$\nabla f(x) = \begin{bmatrix} x_1 \\ 5x_2 \end{bmatrix} \qquad H(x) = \begin{bmatrix} 1 & 0 \\ 0 & 5 \end{bmatrix}$$

Escolhendo $x^0 = \begin{bmatrix} 5 \\ 1 \end{bmatrix}$ obtemos $\nabla f(x) = \begin{bmatrix} 5 \\ 5 \end{bmatrix}$

Resolvendo o sistema linear $H(x^0)\delta_0 = -\nabla f(x^0)$, obtemos a direção $\delta_0 = \begin{bmatrix} -5 \\ -1 \end{bmatrix}$ e assumindo o passo igual a 1, temos

$$x^1 = x^0 + \alpha_0 * \delta_0$$

$$x^1 = \begin{bmatrix} 5 \\ 1 \end{bmatrix} + 1 * \begin{bmatrix} -5 \\ -1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

esta é uma solução exata do problema.

Método de Newton Modificado

Na forma como foi apresentado acima, o método de Newton ainda não apresenta garantia de convergência, pois nada pode ser afirmado sobre o sinal da matriz hessiana, que deve ser definida positiva por duas razões:

- para garantir que a aproximação quadrática tenha um mínimo; e
- para garantir que a inversa da matriz hessiana exista, que é a condição necessária para resolver a $x^{i+1} = x^k - H(x^k)^{-1}\nabla f(x^k)$ ou $x^{i+1} = x^k - \alpha_k H(x^k)^{-1}\nabla f(x^k)$

Portanto, é necessário testar a positividade de $H(x)$ a cada iteração, e na eventualidade de se obter $H(x) \leq 0$, deve-se aplicar um procedimento de positivação dessa matriz. Dada uma matriz H simétrica, conhecendo-se o menor autovalor λ_{min} desta matriz, é possível obter uma matriz M definida positiva a partir de H na forma:

- Se $\lambda_{min} > 0$, então $M = H$
- Se $\lambda_{min} \leq 0$, então $M = H + (\epsilon - \lambda_{min})I$, com $\epsilon > \lambda_{min}$

Método de Newton Modificado

A lei de ajuste do método de Newton torna-se

$$x^{k+1} = x^k - \alpha_k M_k^{-1} \nabla f(x_k)$$

onde

$$\begin{cases} M_k = H(x^k) & se \quad \lambda_{min}^k > 0 \\ M_k = H(x^k) + (\epsilon - \lambda_{min}^k) & se \quad \lambda_{min}^k \leq 0 \end{cases}$$

com λ_{mim}^k o autovalor mínimo de H_k. Ainda não existem resultados que conduzam à determinação automática de um valor ótimo para ϵ

- É importante mencionar ainda que em lugar de positivar H_k poder-se-ia, em princípio, utilizar qualquer outra matriz definida positiva de dimensões apropriadas.
- A razão para optar pelo processo de positivação da hessiana é a analogia com o tradicional método de Levenberg-Marquardt (que será visto a seguir) que pode ser interpretado como uma combinação entre a lei de ajuste do método do gradiente e a lei de ajuste do método de Newton.

Método de Levenberg-Marquardt - LM

Definição do Problema

O problema para o qual o algoritmo LM fornece uma solução é chamado Minimização dos Quadrados Minímos Não Linear.
Isto implica que a função a ser minimizada é da forma abaixo

$$f(x) = \frac{1}{2}\sum_{j=1}^{m} r_j^2(x)$$

onde $x = (x_1, x_2, \cdots, x_n)$ é um vetor, e cada r_j é uma função $\mathcal{R}^n \rightarrow \mathcal{R}$. Os r_j são referidos como residuos e é assumido que $m \geq n$

$f(x)$ pode ser reescrita como $f(x) = \frac{1}{2}||r(x)||^2$.

Método de Levenberg-Marquardt - LM

O gradiente de $f(x)$ pode ser escrito como:

$$\nabla f(x) = \begin{bmatrix} r_1(x)\frac{\partial r_1(x)}{\partial x_1} + r_2(x)\frac{\partial r_2(x)}{\partial x_1} + \cdots + r_m(x)\frac{\partial r_m(x)}{\partial x_1} \\ r_1(x)\frac{\partial r_1(x)}{\partial x_2} + r_2\frac{\partial r_2(x)}{\partial x_2} + \cdots + r_m(x)\frac{\partial r_m(x)}{\partial x_2} \\ \cdots \\ r_1(x)\frac{\partial r_1(x)}{\partial x_n} + r_2(x)\frac{\partial r_2(x)}{\partial x_n} + \cdots + r_m(x)\frac{\partial r_m(x)}{\partial x_n} \end{bmatrix}$$

$$r(x) = \begin{bmatrix} r_1(x) \\ r_2(x) \\ \cdots \\ r_m(x) \end{bmatrix}$$

$$\nabla r(x) = \begin{bmatrix} \nabla r_1(x)^T \\ \nabla r_2(x)^T \\ \cdots \\ \nabla r_m(x)^T \end{bmatrix} = \begin{bmatrix} \frac{\partial r_1(x)}{\partial x_1} & \frac{\partial r_1(x)}{\partial x_2} + & \cdots & \frac{\partial r_1(x)}{\partial x_n} \\ \frac{\partial r_2(x)}{\partial x_1} & \frac{\partial r_2(x)}{\partial x_2} + & \cdots & \frac{\partial r_2(x)}{\partial x_n} \\ \cdots & & & \\ \frac{\partial r_m(x)}{\partial x_1} & \frac{\partial r_m(x)}{\partial x_2} + & \cdots & \frac{\partial r_m(x)}{\partial x_n} \end{bmatrix}$$

$$\nabla f(x) = \nabla r(x)^T r$$

Método de Levenberg-Marquardt - LM

A hessiana $f(x)$ pode ser escrito como ($m = 2, n = 2$):

$$\nabla f(x) = \begin{bmatrix} r_1(x)\frac{\partial r_1(x)}{\partial x_1} + r_2(x)\frac{\partial r_2(x)}{\partial x_1} \\ r_1(x)\frac{\partial r_1(x)}{\partial x_2} + r_2\frac{\partial r_2(x)}{\partial x_2} \end{bmatrix}$$

$$r(x) = \begin{bmatrix} r_1(x) \\ r_2(x) \end{bmatrix}$$

$$\nabla r(x) = \begin{bmatrix} \nabla r_1(x)^T \\ \nabla r_2(x)^T \end{bmatrix} = \begin{bmatrix} \frac{\partial r_1(x)}{\partial x_1} & \frac{\partial r_1(x)}{\partial x_2} \\ \frac{\partial r_2(x)}{\partial x_1} & \frac{\partial r_2(x)}{\partial x_2} \end{bmatrix}$$

$$\nabla^2 r(x) = \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix}$$

$$h_{11} = \frac{\partial r_1(x)}{\partial x_1}\frac{\partial r_1(x)}{\partial x_1} + r_1(x)\frac{\partial r_1(x)^2}{\partial^2 x_1} + \frac{\partial r_2(x)}{\partial x_1}\frac{\partial r_2(x)}{\partial x_1} + r_2(x)\frac{\partial r_2(x)^2}{\partial^2 x_1}$$

$$h_{12} = \frac{\partial r_1(x)}{\partial x_2}\frac{\partial r_1(x)}{\partial x_1} + r_1(x)\frac{\partial r_1(x)}{\partial x_2}\frac{\partial r_1(x)}{\partial x_1} + \frac{\partial r_2(x)}{\partial x_2}\frac{\partial r_2(x)}{\partial x_1} + r_2(x)\frac{\partial r_2(x)}{\partial x_2}\frac{\partial r_2(x)}{\partial x_1}$$

Método de Levenberg-Marquardt - LM

O gradiente de $f(x)$ pode ser escrito como ($m = 2, n = 2$):

$$\nabla f(x) = \nabla r(x)^T r(x)$$

$$H(x) = \nabla r(x)^T \nabla r(x) + \sum_{k=1}^{2} r_k(x) \nabla^2 r_k(x)$$

Quando os erros residuais são suficientemente pequenos, a matriz hessiana pode ser aproximada pelo primeiro termo da equação acim

$$H(x) \approx \nabla r(x)^T \nabla r(x)$$

Esta aproximação geralmente é valida em um mínimo de $f(x)$ para a maioria dos próposito, e é a base para o método de Gaus-Newton. A lei de atualização torna-se então:

$$x^{k+1} = x^k - [\nabla r(x)^T \nabla r(x)]^{-1} \nabla r(x)^T r(x)$$

Método de Levenberg-Marquardt - LM

A modificação de Levenberg-Marquardt para o método de Gauss-Newton é:

$$x^{k+1} = x^k - [\nabla r(x)^T \nabla r(x) + \mu I]^{-1} \nabla r(x)^T r(x)$$

O efeito da matriz adicional μI é adicionar μ a cada autovalor de $J^T J$
Uma vez que a matriz J^J é semi-definida positiva e portanto mínimo possível é zero, qualquer valor positivo, pequeno, mas numericamente significativo, de μ será suficiente para restaurar a matriz aumentada e produzir uma direção descendente de busca.

Algoritmo Levenberg-Marquardt

1: Entrada: Atribua um valor inicial para $x \in \mathcal{R}^n$ para o vetor de parâmetros e um valor arbitrariamente pequeno para uma constante $\epsilon > 0$
2: Inicio
3: **while** condição de parada **do**
4: Construa a matriz $\nabla r(x)$, $f(x) = \sum_{k=1}^{m} r_k^2$,
5: Calcule o ajuste $\delta_i = [\nabla r(x) \nabla r(x) + \mu I]^{-1} \nabla r(x)^T r(x)$;
6: Utilize um procedimento de busca unidimensional para encontrar α_i;
7: Atualize: $x^{k+1} = x^k - \alpha_i \delta_i$;
8: **end while**

Algorithm 1: Pseudocódigo do Levenberg-Marquardt

Método de Davidon-Fletcher-Powell - DFD

- Este método, assim como o método BFGS, é classificado como método quase-Newton
- A idéia por trás dos métodos quase-Newton é fazer uma aproximação iterativa da inversa da matriz hessiana, de forma que:

$$\lim_{i \to \infty} M_i = [\nabla^2 f(x)]^{-1}$$

- São considerados os métodos teoricamente mais sofisticados na solução de problemas de otimização não-linear irrestrita e representam o ápice do desenvolvimento de algoritmos através de uma análise detalhada de problemas quadráticos.
- Para problemas quadráticos, gera as direções do método do gradiente conjugado (que será visto posteriormente) ao mesmo tempo que constrói a inversa da Hessiana.
- A cada passo a inversa da Hessiana é aproximada pela soma de duas matrizes simétricas de posto 1, procedimento que é geralmente chamado de correção de posto 2 (rank 2 correction procedure).

Método de Davidon-Fletcher-Powell - DFD

- Construção da inversa

$$M_{i+1} = M_i + \frac{p_i p_i^T}{p_i^T q_i} - \frac{M_i q_i q_i^T M_i}{q_i^T M_i q_i}, i = 0, 1, \cdots, n$$

onde $p_i = \alpha_i d_i$, e $q_i = g_{i+1} - g_i = \nabla^2 f(x) p_i$
Note que a avaliação em dois pontos fornece informações sobre a matriz hessiana $\nabla^2 f(x)$

Algoritmo DFP

1: Entrada: Atribua um valor inicial para $x \in \mathcal{R}^n$ para o vetor de parâmetros e um valor arbitrariamente pequeno para uma constante $\epsilon > 0$
2: Defina $d_0 = g_0$, $M_0 = I$, faça $i = 0$, $g_0 = -\nabla f(x)$
3: **while** condição de parada **do**
4: Determine a direção $d_i = M_i g_i$,
5: Se (i mod $P = 0$), faça $d_i = g_i$ e $M_i = I$;
6: Utilize um procedimento de busca unidimensional para encontrar α_i;
7: Atualize: $x^{k+1} = x^k - \alpha_i d_i$;
8: Calcule $p_i = \alpha_i d_i$, g_{i+1};
9: Faça $q_i = g_{i+1} - g_i$
10: Calcule M_{i+1}
11: Faça $i = i + 1$
12: **end while**

Algorithm 2: Pseudocódigo do DFB

Método de Broyden-Fletcher-Goldfarb-Shanno (BFGS)

$$M_{i+1} = M_i + \frac{p_i p_i^T}{p_i^T q_i}\left[1 + \frac{q_i^T M_i q_i}{p_i^T q_i}\right] - \frac{M_i q_i p_i^T + p_i q_i^T M_i}{p_i^T q_i}, i = 0, 1, \cdots, n$$

onde $p_i = \alpha_i d_i$, e $q_i = g_{i+1} - g_i = \nabla^2 f(x) p_i$

Algoritmo BFGS

1: Entrada: Atribua um valor inicial para $x \in \mathcal{R}^n$ para o vetor de parâmetros e um valor arbitrariamente pequeno para uma constante $\epsilon > 0$
2: Defina $d_0 = g_0$, $M_0 = I$, faça $i = 0$, $g_0 = -\nabla f(x)$
3: **while** condição de parada **do**
4: Determine a direção $d_i = M_i g_i$,
5: Se $(imodP = 0)$, faça $d_i = g_i$ e $M_i = I$;
6: Utilize um procedimento de busca unidimensional para encontrar α_i;
7: Atualize: $x^{k+1} = x^k - \alpha_i d_i$;
8: Calcule $p_i = \alpha_i d_i$, g_{i+1};
9: Faça $q_i = g_{i+1} - g_i$
10: Calcule M_{i+1}
11: Faça $i = i + 1$
12: **end while**

Algorithm 3: Pseudocódigo do BFGS

Método das Secantes de um Passo (OSS - One Step Secante)

- O termo método de secante provém do fato de que as derivadas são aproximadas por secantes avaliadas em dois pontos da função (neste caso a função é o gradiente).
- Uma vantagem deste método é que sua complexidade é de ordem $O(n)$, ou seja, é linear em relação ao número n de parâmetros, enquanto a complexidade dos métodos DFP e BFGS é de ordem $O(n^2)$
- A principal razão da redução do esforço computacional deste método em relação aos anteriores (DFP e BFGS), é que agora a direção de atualização é calculada somente a partir de vetores determinados pelos gradientes, e não há mais a armazenagem da aproximação da inversa da hessiana

- A nova direção d_{i+1} é obtida como segue:

$$d_{i+1} = -g_i + A_i s_i + B_i q_i$$

onde $s_i = x^{i+1} - x^i = p_i$, $q_i = g_{i+1} - g_i$, $p_i = \alpha_i d_i$

$$A_i = -\left[1 + \frac{q_i^T q_i}{s_i^T q_i}\right] \frac{s_i^T g_i}{s_i^T q_i} + \frac{q_i^T g_i}{s_i^T q_i}$$

$$B_i = \frac{s_i^T g_i}{s_i^T q_i}$$

Algoritmo Método das Secantes

1: Entrada: Atribua um valor inicial para $x \in \mathcal{R}^n$ para o vetor de parâmetros e um valor arbitrariamente pequeno para uma constante $\epsilon > 0$
2: Defina $d_0 = g_0$,$p_0 = \alpha_0 d_0$, faça $i = 0$, $g_0 = -\nabla f(x)$
3: **while** condição de parada **do**
4: Determine a direção $d_i = -g_i + A_i s_i + B_i q_i$,
5: Se (i mod $P = 0$), faça $d_i = g_i$;
6: Utilize um procedimento de busca unidimensional para encontrar α_i;
7: Atualize: $x^{k+1} = x^k - \alpha_i d_i$;
8: Calcule $p_i = \alpha_i d_i$, g_{i+1};
9: Faça $q_i = g_{i+1} - g_i$
10: Faça $i = i + 1$
11: **end while**

Algorithm 4: Pseudocódigo do Método das Secantes

Gradiente Conjugado: Polak-Ribiére (PR) e Fletcher-Reeves (FR)

- Existe um consenso geral da comunidade de análise numérica que a classe de métodos de otimização chamados métodos do gradiente conjugado, tratam de problemas de grande escala de maneira efetiva.
- Os métodos do gradiente conjugado possuem sua estratégia baseada no modelo geral de otimização apresentado no algoritmo padrão e do gradiente, mas escolhem a direção de busca d_i, o passo α_i e o coeficiente de momento β_i mais eficientemente utilizando informações de segunda ordem.
- É projetado para exigir menos cálculos que o método de Newton e apresentar taxas de convergência maiores que as do método do gradiente.
- É baseado no método das direções conjugadas proposto para tratar problemas quadráticos:

$$f(x) = \frac{1}{2}x^T Qx - b^T x$$

- A lei de ajuste do método das direções conjugadas é dada por: $x^{k+1} = x^k - \alpha_k d_k$, $\alpha_k = \frac{d_k^T \nabla f(x) d_k}{d_k^T Q d_k}$

Gradiente Conjugado: Polak-Ribiére (PR) e Fletcher-Reeves (FR)

- Antes de aplicarmos a lei de ajuste dada pela equação acima, é necessário obter as direções Q-conjugadas $d_i \in \mathcal{R}^n$, $i = 0, \cdots, n-1$. Uma maneira de determinar estas direções é tomá-las na forma

$$\begin{cases} d_0 = -\nabla f(x^0) \\ d_{i+1} = -\nabla f(x^{i+1}) + \beta_i d_i \quad i \geq 0 \end{cases}$$

com $\beta_i = \frac{\nabla f(x_{i+1})^T Q d_i}{d_i^T Q d_i}$

- Para problemas não-quadráticos a matriz Q deve ser aproximada pela matriz hessiana calculada no ponto x^i

Problemas não quadráticos - Método de Polak-Ribiére (PR)

- A aplicação destes algoritmos a problemas não quadráticos envolve um procedimento de busca unidimensional do passo de ajuste (taxa de aprendizagem) e a aproximação do parâmetro β utilizando informações de primeira ordem (gradientes).

Uma destas aproximações é dada pelo método de Polak-Ribière (PR)

Se $(i \bmod n \neq 0)$, faça $d_{i+1} = g_i + \beta_i d_i$, onde $\beta_i = \frac{g_{i+1}^T (g_{i+1} - g_i)}{g_i^T g_i}$
Senão, faça: $d_i = g_i$

Outra aproximação é dada pelo método de Fletcher-Reeves (FR)

Se $(i \bmod n \neq 0)$, faça $d_{i+1} = g_i + \beta_i d_i$, onde $\beta_i = \frac{\|g_{i+1}\|}{\|g_i\|}$
Senão, faça: $d_i = g_i$

Algoritmo Polak-Ribière

1: Atribua um valor incial $x^0 \in \mathcal{R}^n$ para o vetor de parâmetros e um valor arbitrariamente pequeno para a constante $\epsilon > 0$
2: Calcule $\nabla f(x^0)$, faça $d_0 = \nabla f(x^0)$ e $i = 0$
3: **while** Condição de parada não for satisfeita **do**
4: Utilize um procedimento de busca unidimensional para encontrar um $alpha_i$ que seja solução ótima do problema
5: Faça $x^{i+1} = x^i + \alpha_i d_i$
6: Calcule $g_i = -\nabla f(x_i)$
7: Se (i mod $n \neq 0$), faça $d_{i+1} = g_i + \beta_i d_i$, onde $\beta_i = \frac{g_{i+1}^T (g_{i+1} - g_i)}{g_i^T g_i}$
8: Senão, faça: $d_i = g_i$
9: $i = i + 1$
10: **end while**

Algorithm 5: Algoritmo Polak-Ribière

Algoritmo Fletcher-Reeves - FR

1: Atribua um valor incial $x^0 \in \mathcal{R}^n$ para o vetor de parâmetros e um valor arbitrariamente pequeno para a constante $\epsilon > 0$
2: Calcule $\nabla f(x^0)$, faça $d_0 = \nabla f(x^0)$ e $i = 0$
3: **while** Condição de parada não for satisfeita **do**
4: Utilize um procedimento de busca unidimensional para encontrar um $alpha_i$ que seja solução ótima do problema
5: Faça $x^{i+1} = x^i + \alpha_i d_i$
6: Calcule $g_i = -\nabla f(x_i)$
7: Se (i mod $n \neq 0$), faça $d_{i+1} = g_i + \beta_i d_i$, onde $\beta_i = \frac{\|g_{i+1}\|^2}{\|g_i\|^2}$
8: Senão, faça: $d_i = g_i$
9: $i = i + 1$
10: **end while**

Algorithm 6: Algoritmo Polak-Ribière

Exemplo

Usar o CG para minimizar $f(x) = 0.5x_1^2 + 2.5x_2^2$

Gradiente é igual a $\nabla f(x) = \begin{bmatrix} x_1 \\ 5x_2 \end{bmatrix}$

Escolhendo $x^0 = \begin{bmatrix} 5 \\ 1 \end{bmatrix}$ obtemos a direção de procura inicial (correspondente ao gradiente negativo)

$$d_0 = -g_0 = -\nabla f(x_0) = \begin{bmatrix} -5 \\ -5 \end{bmatrix}$$

O mínimo exato ao longo da linha de procura é $\alpha_0 = 1/3$, pelo que a próxima solução aproximada é $x_1 = \begin{bmatrix} 3.333 \\ -0.666 \end{bmatrix}$

Calculando o novo gradiente $g_1 = -\nabla f(x_1) = \begin{bmatrix} 3.333 \\ -3.333 \end{bmatrix}$

Exemplo - continuação

Neste ponto, em vez de procurar a direção do gradiente negativo

$$\beta_1 = \frac{g_1^T g_1}{g_0^T g_0} = 0.444$$

que origina a nova direção de procura

$$d_1 = g_1 + \beta_1 d_0 = \begin{bmatrix} 3.333 \\ -3.333 \end{bmatrix} + 0.444 \begin{bmatrix} -5 \\ -5 \end{bmatrix} = \begin{bmatrix} 5.556 \\ 1.111 \end{bmatrix}$$

O passo ao longo desta linha que minimiza a função é $\alpha_1 = 0.6$, que origina a solução exata $x_2 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$, tal como seria de esperar para um função quadrática.

Problema de Otimização Restrito

Relembre que um problema de otimização restrito é um problema da forma

$$(P_2) \quad \underset{x}{\text{minimize}} \quad f(x)$$
$$\text{subject to} \quad g_i(x) \leq 0, \ i = 1, \ldots, m.$$
$$h_i(x) = 0, \ i = 1, \ldots, l.$$
$$x \in X,$$

onde X é um conjunto fechado e $g(x) = (g_1(x), \cdots, g_m(x)) : \mathcal{R}^n \to \mathcal{R}^m$, $h(x) = (h_1(x), \cdots, h_l(x)) : \mathcal{R}^n \to \mathcal{R}^l$

Seja S a região de factibilidade de (P_2), isto é,

$$S := \{x \in X : g(x) \leq 0, h(x) = 0\}$$

Então o problema (P_2) pode ser escrito como

$$(P_1) \quad \underset{x \in S}{\text{minimize}} \quad f(x)$$

Problema de Otimização Restrito

Relembre que $\bar{x}$ é um mínimo local de (P_2) se lá existe $\epsilon > 0$ tal que $f(\bar{x}) \leq f(y)$ $\forall y \in S \cap B(\bar{x}, \epsilon)$

Mínimo e máximo local/global, estrito/não estrito são definidos analogamente.

Nós frequentemente usamos a seguinte notação

$$\nabla g(x) = \begin{bmatrix} \nabla g_1(x)^T \\ \dots \\ \nabla g_m(x)^T \end{bmatrix}. \qquad \nabla h(x) = \begin{bmatrix} \nabla h_1(x)^T \\ \dots \\ \nabla h_l(x)^T \end{bmatrix}.$$

isto é $\nabla g(x) \in \mathcal{R}^{m \times n}$ e $\nabla h(x) \in \mathcal{R}^{l \times n}$ são matrizes jacobianas, cuja i-ésima linha é a transposta do gradiente correspondente.

Problemas irrestritos

Condiçoes de otimalidade

Nos slides anteriores vimos as condições de otimalidade que nos permitem determinar se um dado ponto no espaço representa um ponto de ótimo para um problema irrestrito:

- $\nabla f(x^*) = 0$
- $H(x^*)$ é definida positiva

Agora examinaremos a definição das condições de otimalidade para o caso restrito.

Condições Analíticas

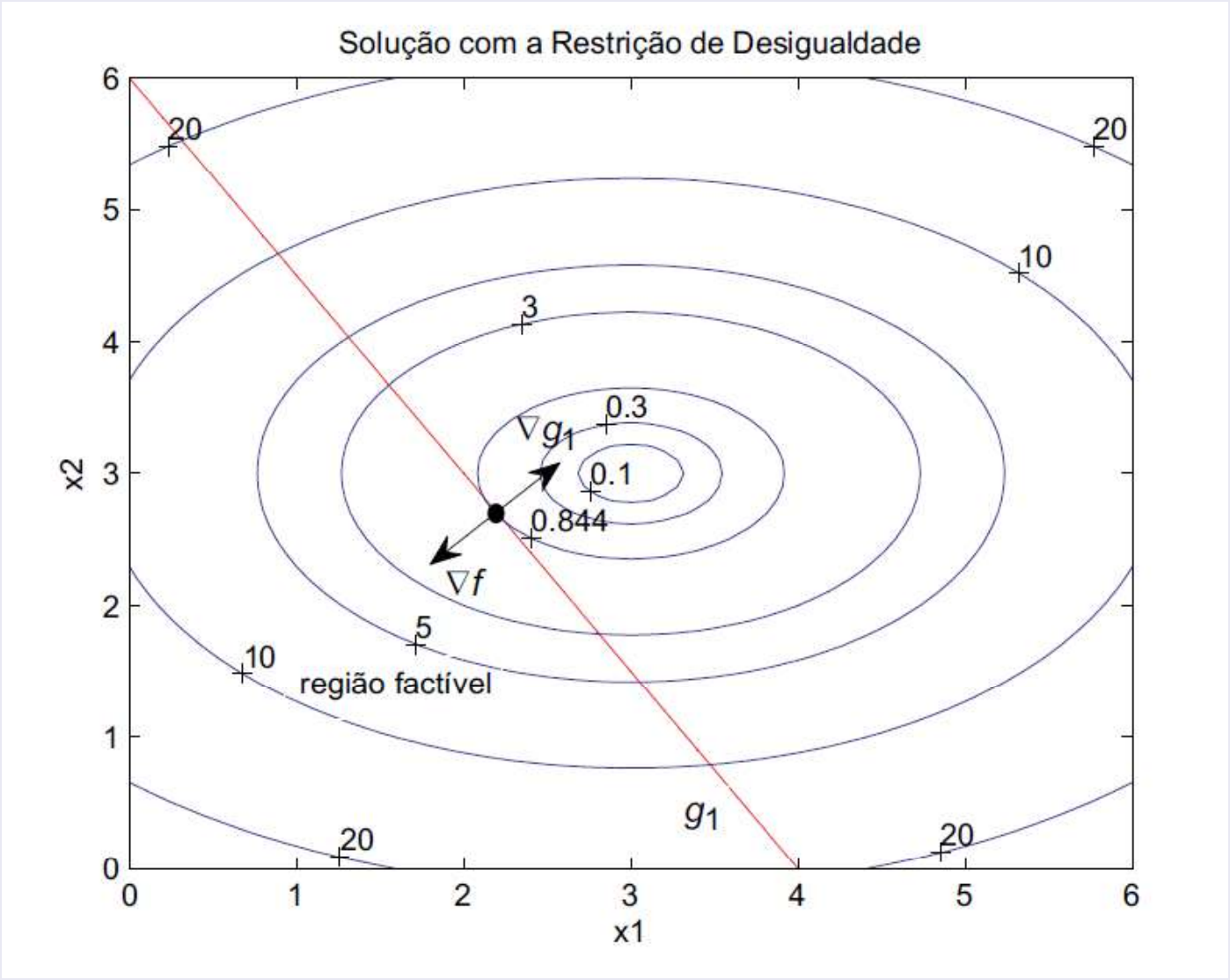
Problemas com restrição de desigualdade

O problema sujeito somente a uma restrição de desigualdade apenas pode ser definido como

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & f(x) = (x_1 - 3)^2 + 2(x_2 - 3)^2 \\ \text{subject to} \quad & g_1(x) = 3x_1 + 2x_2 - 12 \leq 0 \\ & x \in \mathcal{R} \end{aligned} \quad ,$$

Problemas com restrição de desigualdade

Solução gráfica



Condiões Analíticas

Problemas com restrição de desigualdade

Observa-se que no ponto solução, o vetor gradiente da função objetivo está na mesma direção e no sentido oposto do vetor gradiente da função restrição
Existe, portanto, uma relação proporcional entre $\nabla f(x)$ e $\nabla g_1(x)$ no ponto solução.
Representando a constante de proporcionalidade por $\beta_1 \geq 0$, pode-se expressar a relação entre os gradientes por:

$$\nabla f(x) = -\beta_1 \nabla g_1(x) \rightarrow \nabla f(x) + \beta_1 \nabla g_1(x) = 0$$

Condições Analíticas

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & f(x) = (x_1 - 3)^2 + 2(x_2 - 3)^2 \\ \text{subject to} \quad & g_1(x) = 3x_1 + 2x_2 - 12 \leq 0 \\ & x \in \mathcal{R} \end{aligned} \quad ,$$

Transformando o problema

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & f(x, \beta, z_1^2) = (x_1 - 3)^2 + 2(x_2 - 3)^2 + \beta_1[g_1(x) + z_1^2] \\ \text{subject to} \quad & g_1(x) + z_1^2 = 0 \\ & x \in \mathcal{R} \end{aligned} \quad ,$$

Condições Analíticas

Problemas com restrição de desigualdade

Considerando a função modificada $f(x, \beta_1, z_1^2)$ como uma função irrestrita, as condições de primeira ordem tem que ser satisfeitas no ponto de ótimo:

$$\frac{\partial f(x, \beta_1, z_1^2)}{\partial x_1} = \frac{\partial f(x)}{\partial x_1} + \beta_1 \frac{\partial g_1(x)}{\partial x_1} = 0$$

$$\frac{\partial f(x, \beta_1, z_1^2)}{\partial x_2} = \frac{\partial f(x)}{\partial x_2} + \beta_1 \frac{\partial g_1(x)}{\partial x_2} = 0$$

$$\frac{\partial f(x, \beta_1, z_1^2)}{\partial z_1} = 2\beta_1 z_1 = 0$$

$$\frac{\partial f(x, \beta_1, z_1^2)}{\partial \beta_1} = g_1(x) + z_1^2 = 0$$

Problemas com restrição de desigualdade

Podemos eliminar a variável de folga z_1^2

$$2\beta_1 z_1 = 0 \rightarrow (z_1)2\beta_1 z_1 = (z_1)0 \rightarrow 2\beta_1 z_1^2 = 0$$

$$g_1(x) + z_1^2 = 0 \rightarrow z_1^2 = -g(x)$$

$$\beta_1 g_1(x) = 0$$

O que fornece o sistema de equações

$$\left.\begin{array}{l} \frac{\partial f(x,\beta_1,z_1^2)}{\partial x_1} = \frac{\partial f(x)}{\partial x_1} + \beta_1 \frac{\partial g_1(x)}{\partial x_1} = 0 \\ \frac{\partial f(x,\beta_1,z_1^2)}{\partial x_2} = \frac{\partial f(x)}{\partial x_2} + \beta_1 \frac{\partial g_1(x)}{\partial x_2} = 0 \end{array}\right\} = \left[\begin{array}{c} \frac{\partial f(x)}{\partial x_1} \\ \frac{\partial f(x)}{\partial x_2} \end{array}\right] + \beta_1 \left[\begin{array}{c} \frac{\partial g_1(x)}{\partial x_1} \\ \frac{\partial g_1(x)}{\partial x_2} \end{array}\right] = \left[\begin{array}{c} 0 \\ 0 \end{array}\right]$$

$$\left\{\begin{array}{c} \nabla f(x) + \beta_1 \nabla g(x) = 0 \\ \beta_1 g_1(x) = 0 \end{array}\right.$$

Problemas com restrição de desigualdade

Assim, as condições de primeira ordem para um problema com uma restrição de desigualdade podem ser resumidas da seguinte maneira:

$$\begin{cases} \nabla f(x) + \beta_1 \nabla g(x) = 0 \\ \beta_1 g_1(x) = 0 \end{cases}$$

sendo que a segunda condição garante a factibilidade da solução.
A generalização para um número arbitrário de restrições de desigualdade será apresentada mais tarde.

Problema com restrição de desigualdade

Lembrando a definição do problema-exemplo:

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & f(x) = (x_1 - 3)^2 + 2(x^2 - 3)^2 \\ \text{subject to} \quad & g_1(x) = 3x_1 + 2x_2 - 12 \leq 0 \\ & x \in \mathcal{R} \end{aligned} \quad ,$$

Os gradientes podem ser obtidos analiticamente:

$$\nabla f(x) = [2(x_1 - 3), 4(x_2 - 3)]^T$$

$$\nabla g(x) = [3, 2]^T$$

Problema com restrição de desigualdade

$$\begin{bmatrix} 2(x_1 - 3) + 3\beta 1 \\ 4(x_2 - 3) + 2\beta \\ \beta_1(3x_1 + 2x_2 - 12) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Observe que para que estas condiçoes sejam satisfeitas é necessário que $\beta_1 = 0$ ou $g_1(x) = 0$

- Caso a: $\beta_1 = 0[g_1(x) < 0]$
- Caso b: $\beta_1 \neq 0[g_1(x) = 0]$

Problema com restrição de desigualdade

Resolvendo o sistema obtém-se

$$x^* = [2.18, 2.73]^T$$

$$\beta_1 = 0.55$$

O que resulta em $g_1(x^*) = 0$. Nestes caso, diz-se que a restrição de desigualdade está ativa no ponto solução

- Neste caso foi possível resolver o problema simplesmente resolvendo o sistema resultante das condições de otimalidade;
- Entretanto, na maioria dos casos isto não é possível - as condições são comumente utilizadas para verificação de pontos obtidos por métodos iterativos.

Condições Analíticas

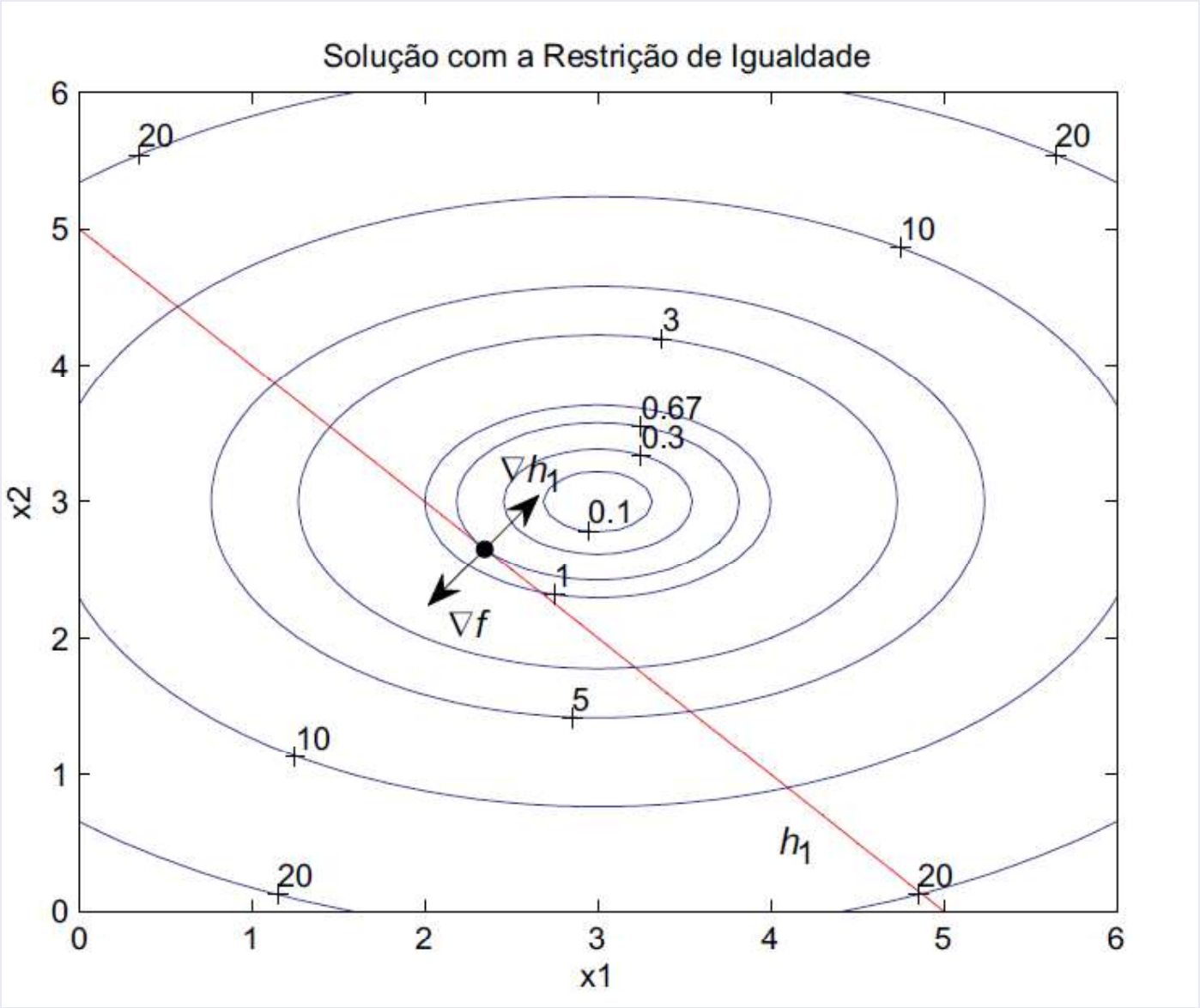
Problemas com restrição de igualdade

Um problema sujeito somente a uma restrição de igualdade pode ser definido como:

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & f(x) = (x_1 - 3)^2 + 2(x^2 - 3)^2 \\ \text{subject to} \quad & h_1(x) = x_1 + x_2 - 5 = 0 \\ & x \in \mathcal{R} \end{aligned} \quad ,$$

Problemas com restrição de igualdade

Solução gráfica



Problemas com restrição de igualdade

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & f(x, \lambda_1) = (x_1 - 3)^2 + 2(x^2 - 3)^2 \\ \text{subject to} \quad & h_1(x) = x_1 + x_2 - \leq 0 \\ & x \in \mathcal{R} \end{aligned} \quad ,$$

Transformando o problema

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & f(x, \lambda_1) = (x_1 - 3)^2 + 2(x^2 - 3)^2 + \lambda_1[h_1(x)] \\ \text{subject to} \quad & x \in \mathcal{R} \end{aligned} \quad ,$$

Obter condições de otimalidade a partir desta função

Problemas com restrição de igualdade

As condições necessárias de otimalidade são obtidas considerando $f(x, \lambda_1)$ como uma função irrestrita das variáveis x_1, x_2 e λ_1 , o que resulta nas seguintes características para o ponto de ótimo

$$\begin{cases} \frac{\partial f(x,\lambda_1)}{\partial x_1} = \frac{\partial f(x)}{\partial x_1} + \lambda_1 \frac{\partial h_1(x)}{\partial x_1} = 0 \\ \frac{\partial f(x,\lambda_1)}{\partial x_2} = \frac{\partial f(x)}{\partial x_2} + \lambda_1 \frac{\partial h_1(x)}{\partial x_2} = 0 \end{cases} = \begin{bmatrix} \frac{\partial f(x)}{\partial x_1} \\ \frac{\partial f(x)}{\partial x_2} \end{bmatrix} + \lambda_1 \begin{bmatrix} \frac{\partial h_2(x)}{\partial x_1} \\ \frac{\partial h_1(x)}{\partial x_2} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\begin{cases} \nabla f(x) + \lambda_1 \nabla h_1(x) = 0 \\ \frac{\partial f(x,\lambda_1)}{\partial \lambda_1} = h_1(x) = 0 \end{cases}$$

Problemas com restrição de igualdade

Em linhas gerais, as condições necessárias de otimalidade para um problema com uma restrição de desigualdade podem ser resumidas em:

$$\begin{cases} \nabla f(x) + \lambda_1 \nabla h_1(x) = 0 \\ \frac{\partial f(x,\lambda_1)}{\partial \lambda_1} = h_1(x) = 0 \end{cases}$$

A generalização para um número arbitrário de restrições de igualdade será apresentada mais tarde.

Problemas com restrição de igualdade

Resolver

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & f(x, \lambda_1) = (x_1 - 3)^2 + 2(x_2 - 3)^2 \\ \text{subject to} \quad & h_1(x) = x_1 + x_2 - 5 = 0 \\ & x \in \mathcal{R} \end{aligned} \quad ,$$

Problemas Geral de Otimização

Problema original

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & f(x) = (x_1 - 3)^2 + 2(x^2 - 3)^2 \\ \text{subject to} \quad & g_1(x) = 3x_1 + 2x_2 - 12 \leq 0 \\ & h_1(x) = x_1 + x_2 - 5 = 0 \\ & x \in \mathcal{R} \end{aligned} \quad ,$$

Problema Transformado

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & f(x, \beta_1, \lambda, , z_1^2) = (x_1 - 3)^2 + 2(x^2 - 3)^2 + \beta_1[g(x) + z_1^2] + \lambda_1 h_1(x) \\ \text{subject to} \quad & g_1(x) + z_1^2 = 0 \\ & h_1(x) = 0 \\ & x \in \mathcal{R} \end{aligned}$$

O problema geral de otimização

Considerando a função transformada $f(x, \beta_1, \lambda, z_1^2)$ como uma função irrestrita, as condições de primeira ordem para este caso podem ser obtidas por

$$\frac{\partial f(x, \beta_1, \lambda_1, z_1^2)}{\partial x_1} = \frac{\partial f(x)}{\partial x_1} + \beta_1 \frac{\partial g_1(x)}{\partial x_1} + \lambda_1 \frac{\partial h_1(x)}{\partial x_1} = 0$$

$$\frac{\partial f(x, \beta_1, \lambda_1, z_1^2)}{\partial x_2} = \frac{\partial f(x)}{\partial x_2} + \beta_1 \frac{\partial g_1(x)}{\partial x_2} + \lambda_1 \frac{\partial h_1(x)}{\partial x_2} = 0$$

$$\frac{\partial f(x, \beta_1, \lambda_1, z_1^2)}{\partial z_1} = 2\beta_1 z_1 = 0$$

$$\frac{\partial f(x, \beta_1, \lambda_1, z_1^2)}{\partial \lambda_1} = h_1(x) = 0$$

O problema de geral de otimização

As equações anteriores podem ser reduzidas a quatro expressões através da eliminação da variável de folga z_1^2

$$\frac{\partial f(x, \beta_1, \lambda_1, z_1^2)}{\partial x_1} = \frac{\partial f(x)}{\partial x_1} + \beta_1 \frac{\partial g_1(x)}{\partial x_1} + \lambda_1 \frac{\partial h_1(x)}{\partial x_1} = 0$$

$$\frac{\partial f(x, \beta_1, \lambda_1, z_1^2)}{\partial x_2} = \frac{\partial f(x)}{\partial x_2} + \beta_1 \frac{\partial g_1(x)}{\partial x_2} + \lambda_1 \frac{\partial h_1(x)}{\partial x_2} = 0$$

$$\beta_1 g_1(x) = 0$$

$$h_1(x) = 0$$

- Caso a: $\beta_1 = 0 [g_1(x) < 0]$
- Caso b: $\beta_1 \neq 0 [g_1(x) = 0]$

O problema geral de otimização

Em linhas gerais, as condições de primeira ordem podem ser resumidas da seguinte maneira:

$$\nabla f(x) + \beta_1 \nabla g_1(x) + \lambda_1 \nabla h_1(x) = 0$$

$$\beta_1 g(x) = 0$$

$$h_1(x) = 0$$

com $\beta_1 \geq 0$.

Problemas Geral de Otimização

Resolva

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & f(x) = (x_1 - 3)^2 + 2(x_2 - 3)^2 \\ \text{subject to} \quad & g_1(x) = 3x_1 + 2x_2 - 12 \leq 0 \\ & h_1(x) = x_1 + x_2 - 5 = 0 \\ & x \in \mathcal{R} \end{aligned} \quad ,$$

Condições de otimalidade necessária

Condições de Necessária Geométrica

Um conjunto $C \subseteq \mathcal{R}^n$ é um cone se $\forall x \in C$, $\alpha x \in C$ para qualquer $\alpha > 0$.
Um conjunto C é um cone convexo se C é um cone C é um conjunto convexo.

Supondo $\bar{x} \in S$. Nós temos as seguintes definições:

- $F_0 := \{d : \nabla f(\bar{x})^T d < 0\}$ é o cone de direções de melhoria de $f(x)$
- $I := \{i : g_i(\bar{x}) = 0\}$ é o conjunto de índices da restrição de desigualdade que satisfaz a desigualdade sem folga em $\bar{x}$
- $G_0 = d : \nabla g_i(x)^T d < 0 \forall i \in I$ é o cone das direções apontando para dentro para restrições limitadas em $\bar{x}$
- $H_0 = d : \nabla h_i(x)^T d < 0 \forall i \in I$ é o conjunto de direções tangente para a restrição de igualdade em $\bar{x}$

Condições Algébricas Necessária

Teorema - Condições necessárias John Fritz

Seja $\bar{x}$ ser uma solução factível de (P_2). Se $\bar{x}$ é um mínimo local de (P_2), então existe (u_0, u, v) tal que

$$u_0 \nabla f(\bar{x}) + \sum_{i=1}^{m} u_i \nabla g_i(\bar{x}) + \sum_{i=1}^{l} v_i \nabla h_i(\bar{x}) = 0,$$

$$u_0, u \geq 0, (u_0, u, v) \neq 0,$$

$$u_i g_i(\bar{x}) = 0, i = 1, \cdots, m.$$

Reescrevendo a equação acima

$$u_0 \nabla f(\bar{x}) + \nabla g_i(\bar{x})^T u + \nabla h_i(\bar{x})^T v = 0,$$

Condições Algébricas Necessária

Teorema - Condições necessárias Karush-Khun-Tucker (KKT)

Seja $\bar{x}$ ser uma solução factível de (P_2) e seja $I = \{i : g_i(\bar{x}) = 0\}$. Além disso, suponha que $\nabla h_i(\bar{x})$ para $i = 1, \cdots, l$ e $\nabla g_i(x)$ para $i \in I$ são linearmente independente. Se $\bar{x}$ é um minimo local, lá existe (u, v) tal que

$$\nabla f(\bar{x}) + \nabla g_i(\bar{x})^T u + \nabla h_i(\bar{x})^T v = 0,$$

$$u \geq 0$$

,

$$u_i g_i(\bar{x}), i = 1, \cdots, m$$

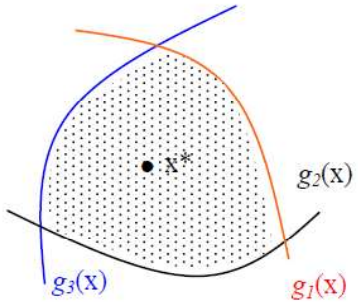


Fig. a)

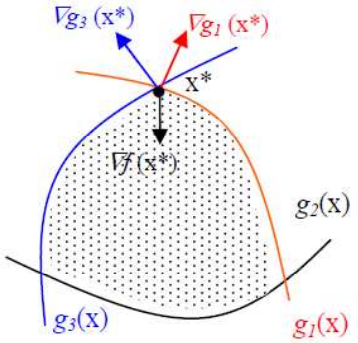


Fig. b)

Exemplo

$$\begin{aligned} \underset{(x_1,x_2)}{\text{minimize}} \quad & (x_1 - 12)^2 + (x_2 + 6)^2 \\ \text{subject to} \quad & x_1^2 + 3x_1 + x_2^2 - 4.5x_2 \leq 6.5 \\ & (x_1 - 9)^2 + x_2^2 \leq 64 \\ & 8x_1 + 4x_2 = 20 \end{aligned}$$

Neste problema, nós temos:

$$f(x) = (x_1 - 12)^2 + (x_2 + 6)^2$$

$$g_1(x) = x_1^2 + 3x_1 + x_2^2 - 4.5x_2 - 6.5$$

$$g_2(x) = (x_1 - 9)^2 + x_2^2 - 64$$

$$h_1(x) = 8x_1 + 4x_2 - 20$$

Vamos determinar se ou não os pontos $\bar{x} = (\bar{x_1}, \bar{x_2}) = (2, 1)$ é um candidato para ser uma solução ótima para este problema.

Exemplo

Vamos checar para factibilidade

$$g_1(\bar{x}) = 0 \leq 0$$

$$g_2(\bar{x}) = -14 < 0$$

$$h_1(\bar{x}) = 0$$

Exemplo

Para verificar a otimalidade, nós calculando todos os gradientes em $\bar{x}$

$$\nabla f(x) = \begin{bmatrix} 2(x_1 - 12) \\ 2(x_2 - 16) \end{bmatrix} = \begin{bmatrix} 20 \\ 14 \end{bmatrix} \qquad \nabla g_1(x) = \begin{bmatrix} 2x_1 + 3 \\ 2x_2 - 4.5 \end{bmatrix} = \begin{bmatrix} 7 \\ -2.5 \end{bmatrix}$$

$$\nabla g_2(x) = \begin{bmatrix} 2(x_1 - 9) \\ 2x_2 \end{bmatrix} = \begin{bmatrix} -14 \\ 2 \end{bmatrix} \qquad \nabla h(x) = \begin{bmatrix} 8 \\ 4 \end{bmatrix} = \begin{bmatrix} 8 \\ 4 \end{bmatrix}$$

Vamos checar, tentando solucionar para $u_1 \leq 0$, $u_2 = 0$, v_1 o seguinte sistema

$$\begin{bmatrix} 20 \\ 14 \end{bmatrix} + \begin{bmatrix} 7 \\ -2.5 \end{bmatrix} u_1 + \begin{bmatrix} -14 \\ 2 \end{bmatrix} u_2 + \begin{bmatrix} 8 \\ 4 \end{bmatrix} v = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Observe que $(\bar{u}, \bar{v}) = (\bar{u1}, \bar{u}_2, \bar{v}_1) = (4, 0, -1)$ soluciona este sistema e que $\bar{u} \leq 0$ e $\bar{u}_2 = 0$ Portanto, $\bar{x}$ é um candidato a ser uma solução ótima deste problema.

Generalização de convexidade

Supondo que X é um conjunto convexo em $\mathcal{R}^n$. A função $f(x) : X \to \mathcal{R}$ é uma função quase convexa se $\forall x, y \in X$ e $\forall \lambda \in [0, 1]$,

$$f(\lambda x + (1 - \lambda)y) \leq \max\{f(x), f(y)\}$$

$f(x)$ é quase côncava se $\forall x, y \in X$ e $\forall \lambda \in [0, 1]$

$$f(\lambda x + (1 - \lambda)y) \geq \min\{f(x), f(y)\}$$

Se $f(x) : X \to \mathcal{R}$, então o conjunto de nível de $f(x)$ são os conjuntos

$$S_\alpha = \{x \in X : f(x) \leq \alpha\}$$

para cada $\alpha \in \mathcal{R}$

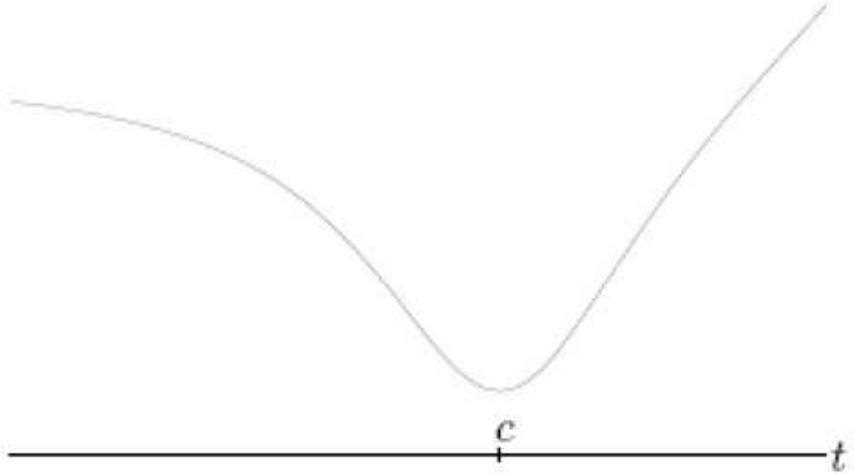
Generalização de convexidade

Proposição

Se $f(x)$ é convexa, então $f(x)$ é quase convexa

Uma função $f(x)$ é quase convexa sobre X se e somente se S_α é conjunto convexo para todo $\alpha \in \mathcal{R}$

Se $f(x)$ é uma função convexa, seus conjuntos de nível são conjuntos convexos.



Condições KKT

Condições Necessárias

$$\nabla f(x^*) + \sum_{j=1}^{p} \mu_j \nabla g_j(x^*) + \sum_{i=1}^{m} \lambda_i \nabla h_i(x^*) = 0$$

$$\mu_j \geq 0, \quad \lambda \quad \text{qualquer} \quad , \mu_i g_i(x) = 0$$

Condições Suficientes

$$\nabla f(x^*) + \sum_{j=1}^{p} \mu_j \nabla g_j(x^*) + \sum_{i=1}^{m} \lambda_i \nabla h_i(x^*) = 0$$

$$\mu_j \geq 0, \quad \lambda \quad \text{qualquer} \quad , \mu_i g_i(x) = 0$$

$\nabla^2 \mathcal{L}$é semi-definida positiva sobre M
$M = \{d : \nabla h_i(x^*)^T d = 0, i = 1, \cdots, m, \text{ e } \nabla g_j(x^*)^T d = 0 \forall j \in J\}$

Problema convexo

Condições necessária de KKT mínimo local $\rightarrow$ Condições suficientes para mínimo global

Métodos de Penalidade

Os métodos de penalidade transformam o problema restrito em um irrestrito adicionando uma função de penalidade a função objetivo.
Assim, o problema de otimização definido anteriormente passa a ser expresso pela minimização irrestrita de uma função pseudo-objetivo

$$(P_1) \quad \underset{x}{\text{minimize}} \quad f(x) + p(x)$$
$$\text{s. t.} \quad x \in X.$$

onde $p(x)$ é uma função de penalidade que incorpora as violações de igualdade e desigualdade

Penalidade Exterior

Para as restrições de igualdade, tem-se

$$p_h(x) = r^h \left(\sum_{j=1}^{m} h_j(x)^2 \right)$$

onde $x \in \mathcal{R}^n$ e $r^h > 0$. Qualquer violação da restrição de igualdade $h(x)$ implicará em uma penalização no valor da função pseudo-objetiva

Para as restrições de desigualdade, tem-se

$$p_g(x) = r^g \left(\sum_{i=1}^{m} \max\left[0, g_i(x)\right]^2 \right)$$

com $r^g > 0$. Se $g_i(x) \leq 0$, o ponto x_k encontra-se na região factível e a restrição de desigualdade é satisfeita; neste caso, $\max\left[0, g_i(x)\right] = 0$.
Caso contrário, se $g_i(x) > 0$ tem-se a penalização da função pseudo-objetiva.

Penalidade Exterior

Em geral, a função de penalidade é definida como:

$$p(x) = p_g(x) + p_h(x) = r^h \left(\sum_{j=1}^{m} h_j(x)^2 \right) + r^g \left(\max\left[0, g_i(x)\right]^2 \right)$$

Os multiplicadores r^g, r^h são usualmente atualizados usando-se um escalar, ou seja,

$$r^h_{k+1} \leftarrow r^h_k C^h$$

$$r^g_{k+1} \leftarrow r^g_k C^g$$

Com C^h, $C^g > 1$

Penalidade Exterior

O método da penalidade exterior apresenta as seguintes características:

- O processo de otimização pode se iniciar a partir de um ponto x_k tanto na região factível quanto na região não factível.
- No caso do processo se iniciar a partir de um ponto x_k fora da região factível, a função de penalidade $p(x)$ torna-se grande, fazendo com que os novos pontos gerados aproximem-se da região factível e minimizando a função-objetivo.
- Portanto, à medida que o processo iterativo for realizado e $r \to \infty$, a solução do problema penalizado convergirá para a solução do problema original

1: Entrada: $f, r_1^g, r_1^h, C^g, C^h, \epsilon, x^0$
2: Ignacio
3: $k \leftarrow 0$
4: Calcule a solução x_1^* do problema $[f(x) + g(x)]$ a partir de x_0;
5: **while** $p(x^{k+1} \geq \epsilon$ **do**
6: $\quad r_{k+1}^h \leftarrow r_k^h C^h$;
7: $\quad r_{k+1}^g \leftarrow r_k^g C^g$;
8: $\quad k \leftarrow k + 1$;
9: $\quad$ Calcule a solução ótima x_{k+1}^* do problema $[f(x) + p(x)]$ a partir de x_k;
10: **end while**

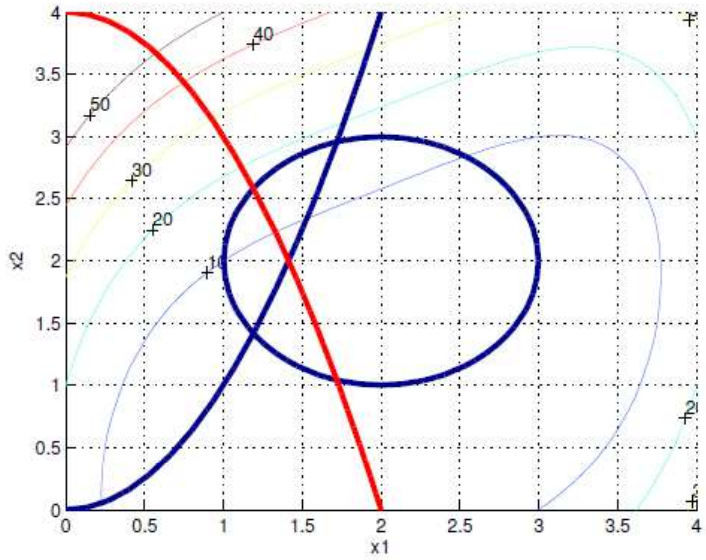
Algorithm 7: Pseudocódigo do Método da Penalidade

A solução dos problemas irrestritos é realizada utilizando-se os métodos usuais (gradiente, Newton)

Exemplo

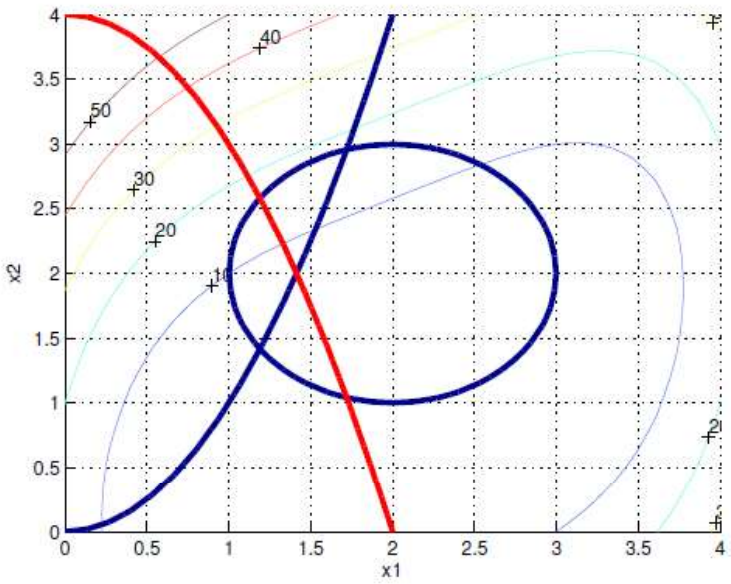
Seja o problema definido por:

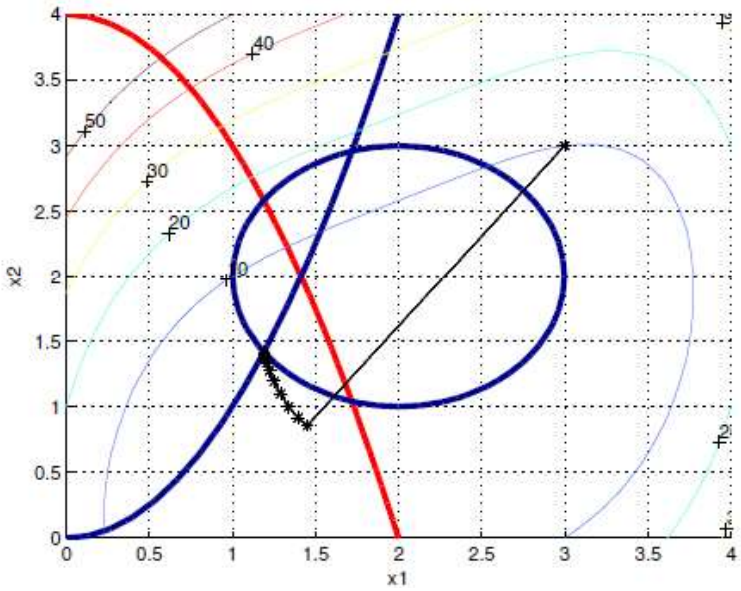
$$\begin{aligned}
\underset{x}{\text{minimize}} \quad & f(x) = (x_1 - 2)^4 + (x_1 - 2x_2)^2 \\
\text{subject to} \quad & h_1(x) = x_1^2 - x_2 = 0 \\
& h_2(x) = (x_1 - 2)^2 + (x_2 - 2)^2 - 1 = 0 \\
& g_1(x) = x_1^2 + x_2 - 4 \leq 0 \\
& 0 \leq x_1 \leq 4 \\
& 0 \leq x_2 \leq 4
\end{aligned}$$



Exemplo

Solução obtida a partir do ponto inicial $x_0 = (3, 3)^T$; $r_0^h = 1$, $C^h = 5$;





Método de Multiplicadores de Lagrange

Com o intuito de satisfazer as condições de Kuhn-Tucker no problema irrestrito, este método associa os multiplicadores de Lagrande às restrições de igualdade e desigualdade. A função de penalidade neste caso é dada por:

$$p(x) = r^h \sum_{k=1}^{m} \lambda_k \left[h_k(x)\right]^2 + r^g \sum_{j=1}^{m} \max\left[g_j(x), \frac{-\beta_j}{2r^g}\right] + \sum_{k=1}^{l} \lambda_k \left[h_k(x)\right] + \sum_{j=1}^{m} \max\left[g_j(x), \frac{-\beta_j}{2r^g}\right]$$

onde λ_k e β_j são os multiplicadores de Lagrande, e r^h e r^g são os multiplicadores de penalidade definidos de maneira similar ao metodo de penalidade.

Método de Multiplicadores de Lagrange Aumentado

Os multiplicadores de Lagrande são atualizados, em cada iteração, com informações a respeito das restrições de acordo com:

$$\lambda_{k+1} = \lambda_k + 2r^h h(x_k)$$

$$\beta_{k+1} = \beta_k + 2r^g \left(max \left[g(x_k), \frac{-\beta}{2r^g} \right] \right)$$

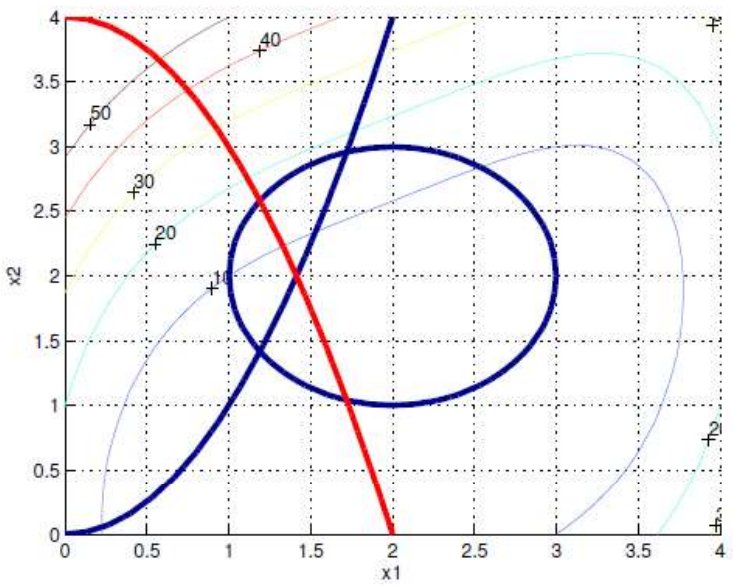
Método de Multiplicadores de Lagrange Aumentado

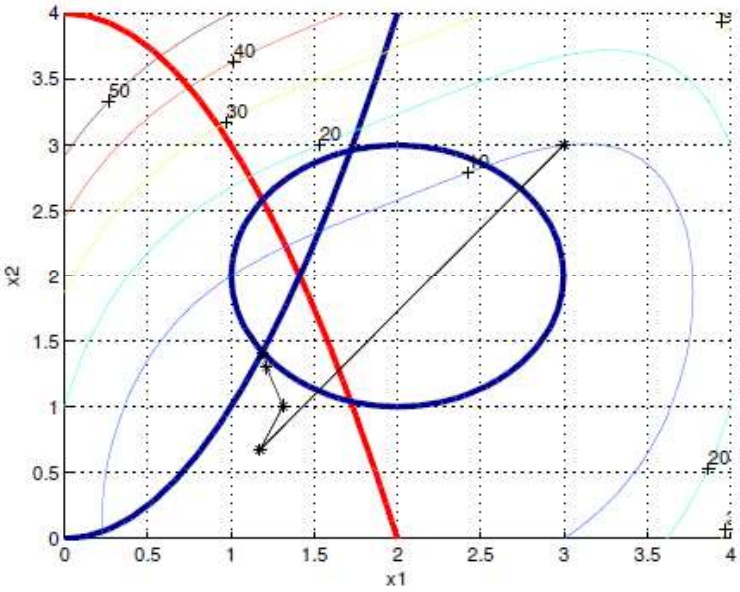
Características

- O ponto incial x_0 pode ser factível ou não;
- Insensível à escolha inicial de r^h e r^g;
- Não requer que $r \to \infty$;
- As restrições de igualdade podem ser satisfeitas com maior precisão;
- No ponto de convergência, os λ e β encontram seus valores ótimos, possibilitando a verificação das condições de KKT;
- Aumento no tempo de processamento.

Exemplo

Solução obtida a partir do ponto inicial $x_0 = (3, 3)^T$; $r_0^h = 1$, $C^h = 5$;





Método de Barreira

Similar para funções de penalidade, funções barreira são também usadas para transfomar um problema restrito em problema irrestrito ou em uma sequencia de problemas irrestrito.
Estas funções define uma barreira contra o abandono da região factível. Se a solução ótima ocorre na fronteira da região de factibilidade, o procedimento move do interior para a fronteira.
O problema primal e barreira são formulado abaixo.

Problema Primal

$$\begin{array}{ll} \underset{x}{\text{minimize}} & f(x) \\ \text{subject to} & g_i(x) \leq 0,\ i = 1, \ldots, m. \\ & x \in X \end{array}$$

Problema Barreira

$$\begin{array}{ll} \underset{x}{\text{minimize}} & f(x) + \mu B(x) \\ \text{subject to} & \mu \geq 0 \end{array}$$

,

Aqui B é uma função barreira que é não negativa e continua sobre a região $\{x : g(x) < 0\}$ e aproxima para ∞ quando a fronteira da região $\{x : g(x) \leq 0\}$ é aproximado do interior.

Função Barreira

Observe que restrições de igualdade, se presente, são acomodados dentro conjunto X.
Alternativamente, no caso de restrições de igualdade, nos podemos possivelmente elimina-las após solucionar alguns variáveis em termos de outras, reduzindo a dimensão do problema.
A razão por que este tratamento é necessário é que os métodos da função barreira requerem que o conjunto $\{x : g(x) < 0\}$ seja não vazio, o qual pode obviamente não ser possível se a restrição de igualdade $h(x) =$ fosse acomodada dentro o conjunto de desigualdade como $h(x) \geq 0$ e $h(x) \leq 0$

O parâmetro $\mu > 0$ é chamado de parâmetro de barreira
Pode-se mostrar que os minimizadores desta função se aproximam dos minimizadores do problema original quando μ se aproxima de zero.
Mais uma vez, a estratégia usual é encontrar minimizadores aproximados para a função barreira, usando uma sequencia de valores decrescentes de μ

Função Barreira

Mais especificamente, a função barreira B é definida por

$$B(x) = \sum_{i=1}^{m} \phi\left[g_i(x)\right]$$

onde ϕ é uma função univáriavel que é continua $\{y :< 0\}$ e satisfaz

$$\phi(y) \geq 0 \quad if \quad y < 0$$

$$\lim_{y \to 0^-} \phi(y) = \infty$$

Por exemplo, um função barreira tipica pode ser da forma

$$B(x) = \sum_{i=1}^{m} \frac{-1}{g_i(x)}$$

$$B(x) = -\sum_{i=1}^{m} \ln\left[\min\{1, -g_i(x)\}\right]$$

Função Barreira

Observe que a segunda função não é diferenciavel por causa do termo $min\{1, -g_i(x)\}$. Desde que as propriedades para ϕ é essencial na vizinhança $y = 0$, pode ser mostrado que a seguinte função função barreira popular, conhecida como função barreira logarimtica de Frisch,

$$B(x) = -\sum_{i=1}^{m} \ln\left[-g_i(x)\right]$$

também admite convergência.

Método Barreira

1: Seja $\epsilon > 0$ um escalar de termino
2: Escolha $x_1 \in X$ com $g(x_1) < 0$.
3: Seja $\mu_1 > 0, \beta \in (0, 1)$
4: $k \leftarrow 1$
5: Solucione o seguinte o problema

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & f(x) + \mu B(x) \\ \text{subject to} \quad & \mu_k \geq 0 \end{aligned}$$

Seja x_{k+1} ser uma solução ótima,
6: Se $\mu_k B(x_{k+1}) < \epsilon$, Pare. Caso contrário, $\mu_{k+1} = \beta \mu_k$, faça $k \leftarrow k + 1$ e vá para o passo 5.

Algorithm 8: Pseudocódigo Função Barreira

Métodos de Programação Quadrática Sequencial

Problema Primal

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & f(x) \\ \text{subject to} \quad & h_i(x) = 0, \ i = 1, \ldots, m. \\ & x \in X \end{aligned}$$

Função Lagrangeano

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & f(x) + \lambda h(x) \\ \text{subject to} \quad & \mu \neq 0 \end{aligned}$$

,

A hessiana do Lagrangeano, em relação a x, no ponto (x, λ) será denotada por

$$W(x, \lambda) = \nabla^2_{xx} L(x, \lambda)$$

Denotaremos a derivada, ou matriz de Jacobiana, de $h(x)$ num ponto $x \in \mathcal{R}^n$ por $A(x)$, isto é,

$$A^T(x) = \lfloor \nabla h_1(x), \nabla h_2(x), \cdots, \nabla h_m(x) \rfloor$$

Observe que a aproximação linear de $h(x)$ no ponto z em função de x e d dada por

$$h(z) = h(x + d) \approx h(x) + \nabla h(x) d = h(x) + A(x) d$$

Queremos $z = x + d$ mais perto do conjunto viável $\{x \in \mathcal{R}^n, h(x) = 0\}$. Para isto é razoável procurar d que satisfaça $h(x) + A(x)d = 0$

Métodos de Programação Quadrática Sequencial

A ideia de métodos de programação quadrática sequencial e, a cada iteração, modelar o problema como um subproblema quadrático.
Encontra-se, então, a solução deste modelo, que é usada como direção de busca.
Mais especificamente, para problemas com apenas restrições de igualdade, definimos a direção de busca na iteração (x_k, λ_k) como a solução do subproblema.

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & \frac{1}{2} p^T W_k p + \nabla f(x_k)^T p \\ \text{subject to} \quad & A_k p + c_k = 0 \end{aligned}$$

Métodos de Programação Quadrática Sequencial

- A função objetiva deste subproblema é uma aproximação da função Lagrangiana e as restrições são linearizações das restrições originais.
- A nova iteração é calculado fazendo-se uma busca na direção p_k até que uma função objetiva apresente decréscimo.
- Métodos de programação quadrática sequencial são muito eficientes na prática.
- Eles tipicamente fazem um número menor de iteração ao custo de resolver subproblemas relativamente complicados a cada iteração.

Função Lagrangeana

$$
\begin{aligned}
(P_2) \quad \underset{x}{\text{minimize}} \quad & f(x) \\
\text{subject to} \quad & g_i(x) \leq 0, \ i = 1, \ldots, m. \\
& h_i(x) = 0, \ i = 1, \ldots, l. \\
& x \in X,
\end{aligned}
$$

A função Lagrangeana $\mathcal{L}$ associada ao problema é dada por:

$$\mathcal{L}(x, \lambda, \mu) = f(x) + \sum_{j=1}^{m} \mu_j g_i(x) + \sum_{i=1}^{n} \lambda_i h_i(x)$$

onde μ_i e λ_i são os multiplicadores de Lagrange.

Função Lagrangeana

Utilizando esta definição, é possivel escrever as condições necessárias de KKT de forma mais compacta

$$\nabla_x \mathcal{L}(x, \mu, \lambda) = 0 \rightarrow \nabla f(x) + \sum \mu_j \nabla g_j(x) + \sum \lambda_i \nabla h_i(x) = 0$$

$$\nabla_\mu \mathcal{L}(x, \mu, \lambda) \leq 0 \rightarrow g_j(x) \leq 0, j = 1, \cdots, l$$

$$\nabla_\lambda \mathcal{L}(x, \mu, \lambda) = 0 \rightarrow h_i(x) = 0, i = 1, \cdots, m$$

$$\mu_i \leq 0, \lambda \text{qualquer}, \mu_i g_i = 0$$

Dualidade

Primeiramente, vamos considerar, por simplificação, que o problema em questão possui apenas restrições de desigualdade, uma vez que $h(x) = 0$ pode ser substituído por $h(x) \leq 0$ e $-h(x) \leq 0$

Problema Primal (P)

$$z^* = \underset{x}{\text{minimize}} \quad f(x)$$
$$\text{subject to} \quad g_i(x) \leq 0, \ i = 1, \ldots, m.$$
$$x \in X$$

,

Problema Dual (D)

$$v^* = \underset{\mu}{\text{maximize}} \quad \mathcal{L}(\mu)$$
$$\text{subject to} \quad \mu \geq 0$$

Sendo a função dual (Função Lagrangeano Dual) dada por:

$$\mathcal{L}(\mu) = \min \mathcal{L}(x, \mu) = \min f(x) + \sum_{i=1}^{m} \mu_i g_i(x)$$

Como o problema (D) consiste em maximizar o mínimo do Lagrangeano, este problema é muitas vezes referido como *problema max-min*.

Dualidade

Teorema

A função dual $\mathcal{L}(\mu)$ é sempre côncova, mesmo que $f(x)$ não seja convexa.

A função Lagrangeana incorpora as restrições do problema Primal. O Problema com restrições (P_1) pode ser resolvido pelo problema equivalente, o qual pode ser escrito como:

$$\mathcal{L}(\mu) = \min \mathcal{L}(x, \mu) = \min f(x) + \sum_{i=1}^{m} \mu_i g_i(x) + \sum_{j=1}^{l} \lambda_j h_j(x)$$

Onde os valores de λ e μ devem atender as condições de KKT:
Primal: $g_i(x) \leq 0, i = 1, \cdots, l$, $h_i(x) = 0, j = 1, \cdots, m$
Restrições do Dual: $\mu \geq 0$
Folga complementar: $\mu_j g_j(x) =, i = 1, \cdots, p$

Dualidade

Para problemas convexos, a minimização de

$$\mathcal{L}(\mu) = \min \mathcal{L}(x, \mu) = \min f(x) + \sum_{i=1}^{m} \mu_i g_i(x) + \sum_{j=1}^{l} \lambda_j h_j(x)$$

é equivalente a $\nabla f(x) + \sum_{i=1}^{m} \nabla g_i(x) + \sum_{j=1}^{l} \lambda_j h_j(x) = 0$
As equações abaixo são referidas como condições de KKT

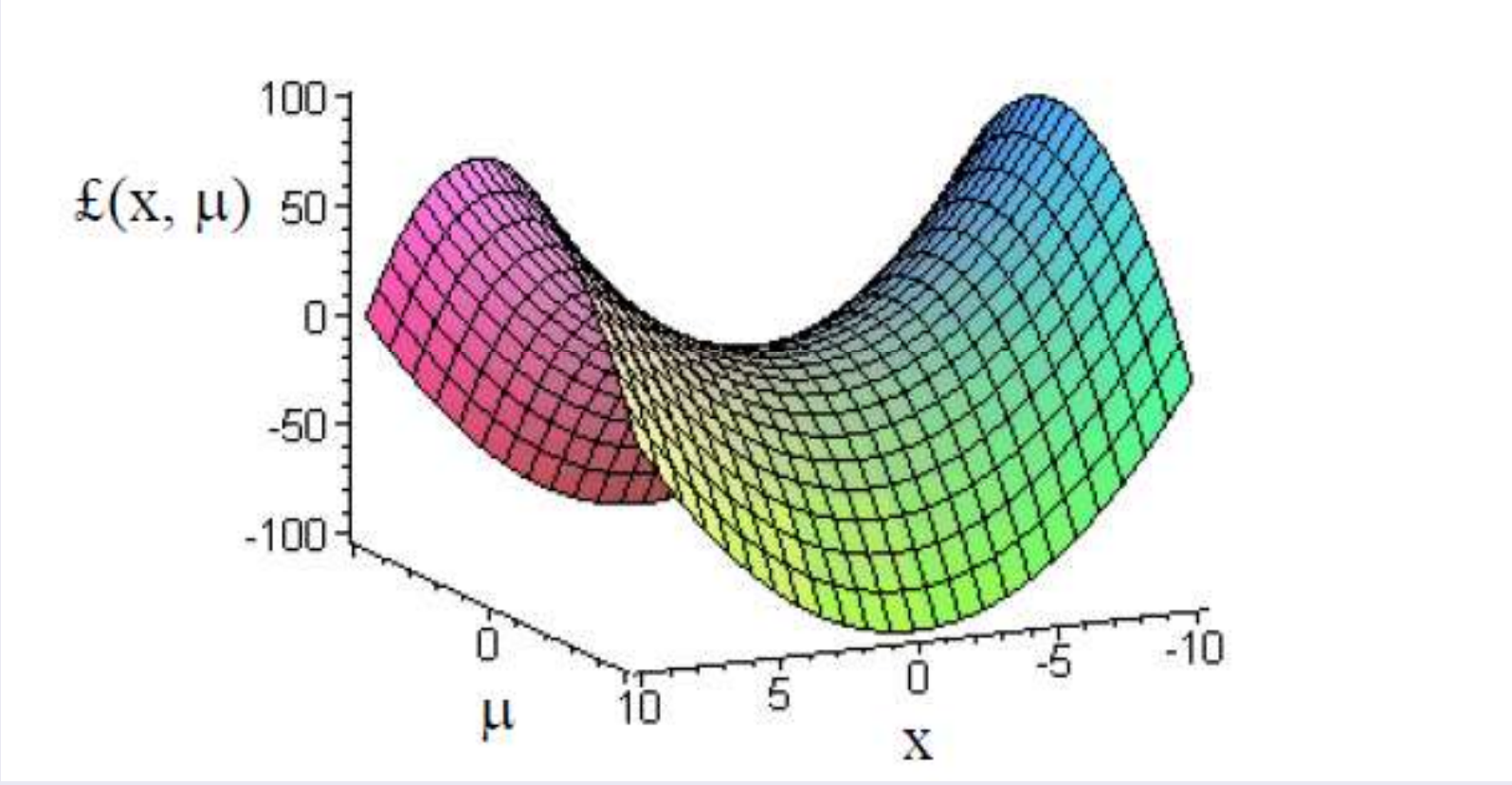
$$\mathcal{L}(\mu) = \min \mathcal{L}(x, \mu) = \min f(x) + \sum_{i=1}^{m} \mu_i g_i(x) + \sum_{j=1}^{l} \lambda_j h_j(x)$$

Primal: $g_i(x) \leq 0, i = 1, \cdots, l,\ h_i(x) = 0, j = 1, \cdots, m$
Restrições do Dual: $\mu \geq 0$
Folga complementar: $\mu_j g_j(x) = 0, i = 1, \cdots, p$

$$\nabla f(x) + \sum_{i=1}^{m} \nabla g_i(x) + \sum_{j=1}^{l} \lambda_j h_j(x) = 0$$

Ponto de Sela

Um ponto (x^0, μ^0) com $\mu^0 \geq 0 \in S$ é um ponto de sela $\mathcal{L}(x, \mu)$ se:
$\mathcal{L}(x^0, \mu^0) \leq \mathcal{L}(x, \mu^0)\ \forall x \in S$ (minimiza)
$\mathcal{L}(x^0, \mu^0) \geq \mathcal{L}(x^0, \mu)\ \forall \mu \geq 0S$ (maximiza)
Assim, o ponto de sela é sela é simultaneamente um mínimo na direção de x e um máximo na direção μ de $\mathcal{L}(x, \mu)$

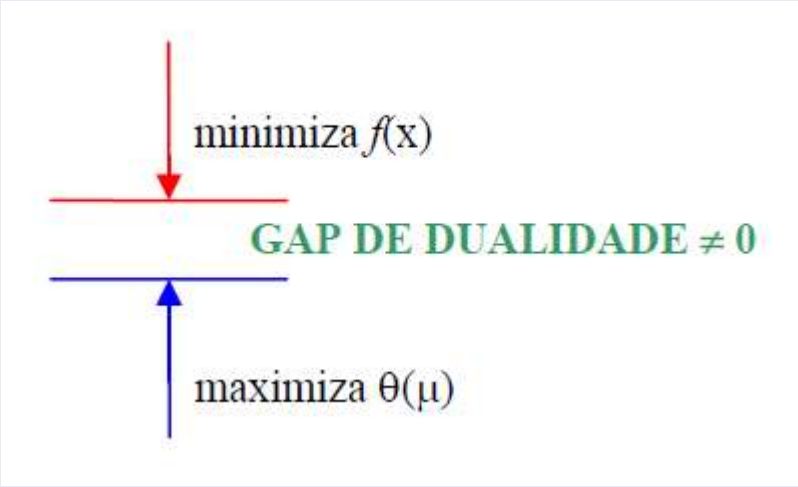


Dualidade

Seja x^0 e μ^0 soluções factíveis quaisquer do problema primal e dual. Então:

$$f(x^0) \leq \mathcal{L}(\mu^0)$$

Isto significa que para um certo (x^0, μ^0) a função primal será sempre maior ou igual do que a função dual.
Esta propriedade de dualidade fraca é sempre assegurada tanto para problemas convexos como para problemas não convexos.
Portanto, $\mathcal{L}(\mu^0)$ pode fornecer um limite inferior para o valor ótimo de $f(x)$. Esta propriedade é utilizada em problemas não convexos de dificil solução



Dualidade Fraca

Teorema da Dualidade Fraca

Se $\bar{x}$ é factível para Problema Primal e $\bar{\mu}$ é factível para o Dual, então

$$f(\bar{x}) \geq \mathcal{L}^*(\mu)$$

Corolário

Se $\bar{x}$ é factível para Problema Primal e $\bar{\mu} \leq 0$ é factível para o Dual, e $f(\bar{x}) = \mathcal{L}^*(\bar{\mu})$, então $\bar{x}$ e $\bar{u}$ são soluções ótimas do Primal e Dual, respectivalmente.

Corolário

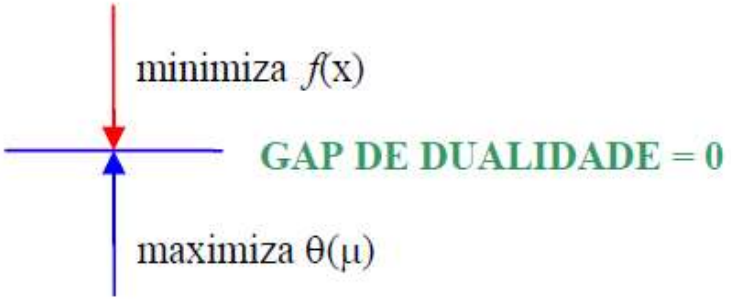
Se a solução do Primal é $z^* = -\infty$, então o Dual não tem solução factível.

Corolário

Se a solução do Dual é $v^* = +\infty$, então o Primal não tem solução factível.

Dualidade Forte

Se x^0 e μ^0 soluções factíveis do problema primal e dual tal que $\mathcal{L}(\mu) = f(x^0)$, dizemos que a dualidade é forte ($gap = 0$). O caso de dualiade forte não é assegurado para todos os problemas



Dualidade Forte

Teorema

Suponha que $f(x)$, $g(x)$ e $h(x)$ funções convexas e diferenciáveis (problema convexo). Suponha que exista (x^0, μ^0) que satisfaz KKT. Então (x^0, μ^0) são soluções ótimas do (P) e (D) com *gap* de dualidade nulo.

Teorema

Um ponto (x^0, μ^0) é ponto de sela de $\mathcal{L}(x, \mu)$ se e somente x^0 resolve (P), μ^0 resolve (D) para $\mu \geq 0$ e $\mathcal{L}(\mu^0) = f(x^0)$ (gap nulo).

Exemplo

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & f(x) = x_1^2 + x_2^2 \\ \text{subject to} \quad & -x_1 - x_2 + 4 \leq 0, \ i = 1, \ldots, m. \\ & x_1, x_2 \geq 0, \end{aligned}$$

Solução ótima: $x_1 = 2, x_2 = 2$
Problema Dual (D)

$$\begin{aligned} \underset{\mu}{\text{minimize}} \quad & \mathcal{L}(\mu) \\ \text{subject to} \quad & \mu \geq 0 \end{aligned}$$

$$\mathcal{L}(\mu) = \min(x, \mu) = x_1^2 + x_2^2 + \mu(-x_1 - x_2 + 4)$$

$\min \mathcal{L}(x, \mu) = x_1^2 + x_2^2 + \mu(-x_1 - x_2 + 4)$
Para um dado μ, fixar *mu* e minimizar em x :

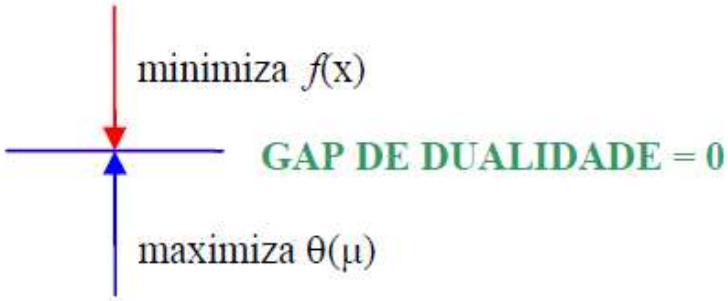
$$\begin{bmatrix} 2x_1 - \mu \\ 2x_2 - \mu \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$x_1 = x_2 = \mu/2$(Problema convexo, a condição necessária é suficiente)

Exemplo

$$\max \mathcal{L}(\mu) = \mu^2/4 + \mu^2/4 + \mu(-\mu/2 - \mu/2 + 4) = -\mu^2/2 + 4\mu$$

$$\frac{\partial \mu}{\partial \mu} = -\mu + 4 = 0 \rightarrow \mu = 4 \rightarrow x_1 = x_2 = 2 \rightarrow f(x) = 8$$



Decomposição Dual

Alguns tipos de problemas são separáveis

$$
\begin{aligned}
(P_2) \quad \underset{x}{\text{minimize}} \quad & \sum_{k=1}^{q} f_k(x_k) \\
\text{subject to} \quad & \sum_{k=1}^{q} g_k(x_k) \leq 0 \\
& \sum_{k=1}^{q} h_k(x_k) = 0 \\
& x \in X,
\end{aligned}
$$

Nesta formulação, o vetor x é dividido em q grupos distintos. Ambas as restrições e função objetiva são decompostos em somatório dos grupos individuais.
Este tipos de problema é muito atrativo para aplicação de métodos duas, pois a minimização irrestrita $\min \mathcal{L}(x, \lambda, \mu)$ pode ser decomposta em subproblemas menores. Este método é chamado de decomposição dual.

Decomposição Dual

$$\mathcal{L}(x, \lambda, \mu) = \sum_{k=1}^{q} \left[f_k(x_k) + \lambda^T h_k(x_k) + \mu^T g_k(x_k) \right]$$

$$\mathcal{L}(x, \lambda, \mu) = \sum_{k=1}^{q} \mathcal{L}_k(x, \lambda, \mu)$$

onde $mathcalL_k(x, \lambda, \mu)$ depende apenas x_k

$$\mathcal{L}_k(x, \lambda, \mu) = f_k(x_k) + \lambda^T h_k(x_k) + \mu g_k(x_k)$$

Assim, o problema primal é equivalente a

$$mathcalL_k(\lambda, \mu) = \min_x mathcalL_k(x, \lambda, \mu)$$

O qual pode ser decomposto $\min_{x_1, \cdots, x_q} \sum_{k=1}^{q} mathcalL_k(x, \lambda, \mu)$
E cada subproblema k pode ser resolvido separadamente

$$\min_{x_1, \cdots, x_q} \mathcal{L}_k(x, \lambda, \mu)$$

A solução destes problemas pode ser usualmente obtida de modo eficiente, uma vez que os subproblemas são de dimensão menor do que o problema original.