Capítulo 5 - Optimização

Carlos Balsa

balsa@ipb.pt

Departamento de Matemática
Escola Superior de Tecnologia e Gestão de Bragança

Mestrados em Engenharia da Construção
Métodos de Aproximação em Engenharia
1º Semestre 2010/2011

INSTITUTO POLITÉCNICO DE BRAGANÇA
Escola Superior de Tecnologia e de Gestão

Sumário

Problemas de Optimização

Optimização Unidimensional

Optimização Multidimensional

Considerações Finais

Optimização

- Dada uma função $f : \mathbb{R}^n \to \mathbb{R}$ e um conjunto $\mathbf{S} \subseteq \mathbb{R}^n$, encontrar $\mathbf{x}^* \in \mathbf{S}$ tal que $f(\mathbf{x}^*) \leq f(\mathbf{x})$ para todo o $\mathbf{x} \in \mathbf{S}$.
- $\mathbf{x}^*$ é chamado *minimizador* ou *mínimo* de f
- Basta considerar a minimização, pois o máximo de f é igual ao mínimo de $-f$
- A *função objectivo* f é normalmente derivável, podendo ser linear ao não linear
- O conjunto restrição $\mathbf{S}$ é definido por um sistema de equações e inequações que pode ser linear ou não linear
- Os pontos $\mathbf{x} \in \mathbf{S}$ são chamados pontos *praticáveis*
- Se $\mathbf{S} = \mathbb{R}^n$ o problema não tem restrições

Problemas de Optimização

- Um problema genérico de optimização contínuo:

$$\min f(x) \;\; \text{sujeito a} \;\; g(x) = 0 \;\; \text{e} \;\; h(x) \leq 0$$

  em que $f : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}^m$ e $h : \mathbb{R}^n \to \mathbb{R}^p$
- *Programação linear*: f, g e h são todas lineares
- *Programação não linear*: pelo menos uma das funções f, g e h é não linear
- Vamos considerar apenas os problemas de programação não linear sem restrições

Método de Newton para Optimização Unidimensional

- *Método de Newton* para a resolução de $f(x) = 0$:

$$x_{k+1} = x_k - f(x_k)/f'(x_k), \quad para \quad k = 0, 1, \ldots$$

- Num problema de optimização o Método de Newton é usado para a resolução de $f'(x) = 0$:

$$x_{k+1} = x_k - f'(x_k)/f''(x_k), \quad \text{para} \quad k = 0, 1, \ldots$$

- Converge quadraticamente (duplica o número de dígitos correctos em cada iteração k) para o zero de $f'(x)$ desde que o ponto de partida esteja suficientemente próximo da solução

Exemplo 1: Método de Newton

- Usar o método de Newton para minimizar $f(x) = 0.5 - xe^{-x^2}$
- Primeira e a segunda derivada de f são dada por

$$f'(x) = (2x^2 - 1)e^{-x^2}$$

e

$$f''(x) = 2x(3 - 2x^2)e^{-x^2}$$

- Formula de recorrência do Método de Newton para encontrar o zero de $f'(x)$ é

$$x_{k+1} = x_k - (2x_k^2 - 1)/\left(2x_k\left(3 - 2x_k^2\right)\right), \quad \text{para} \quad k = 0, 1, \ldots$$

- Usando a estimativa inicial $x_0 = 1$, obtemos

k	x_k	$f(x_k)$
0	1.000	0.132
1	0.500	0.111
2	0.700	0.071
3	0.707	0.071

Pontos Críticos

- Para funções de uma única variável podemos encontrar o mínimo analisando os pontos em que a derivada é igual a zero
- Generalização a funções de n variáveis consiste em encontrar os pontos críticos, isto é, a solução do sistema não linear

$$\nabla f(x) = 0$$

  em que $\nabla f(x)$ é o vector *gradiente de f*, cuja $i$esima componente é $\partial f(x)/\partial x_i$
- Mas nem todos os pontos críticos são mínimos: podem ser também máximos ou pontos de sela

Pontos Críticos de Funções Multidimensionais

- Para uma função $f : \mathbb{R}^n \to \mathbb{R}$ duas vezes diferenciavel, podemos distinguir os pontos críticos através da matriz Hessiana $\mathbf{H}(\mathbf{x})$, matriz das segundas derivadas parciais, definida por

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

  que é simétrica

- Se no ponto x^*, $\mathbf{H}(x^*)$ é
  - Positiva definida, então x^* é um mínimo de f
  - Negativa definida, então x^* é um máximo de f
  - Indefinida, então x^* é um ponto de sela de f

Pontos Críticos de uma Função Bidimensional

Regras práticas para verificar se a matriz é positiva definida:

- Factorixação de Cholesky existe: $A = LL^T$
- Valores próprios de A são todos positivos
- Todos os determinantes superiores esquerdos de A são positivos

Regras práticas para verificar se a matriz A é negativa definida:

- Valores próprios são todos negativos
- Determinantes superiores esquerdos de ordem par são positivos, de ordem ímpar são negativos

Regras práticas para verificar se a matriz é indefinida:

- Valores próprios positivos e negativos

Pontos Críticos de uma Função Bidimensional

Para uma função $f(x_1, x_2) : \mathbb{R}^2 \rightarrow \mathbb{R}$ duas vezes diferenciavel, podemos distinguir os seguintes casos

- $|\mathbf{H}(x^*)| > 0$ e $\partial^2 f / \partial x_1^{*2} > 0$, então x^* é um mínimo de f
- $|\mathbf{H}(x^*)| > 0$ e $\partial^2 f / \partial x_1^{*2} < 0$, então x^* é um máximo de f
- $|\mathbf{H}(x^*)| < 0$, então x^* é um ponto sela f

Exercício 1: Determine os pontos críticos de cada uma das seguintes funções e caracterize cada um deles como mínimo, máximo ou ponto de sela:

- $f(x, y) = 2x^2 - 2xy + 2y^2$
- $f(x, y) = x^4 - 4xy + y^4$

Método de Newton para Optimização Multidimensional

- Método de Newton para minimizar uma função unidimensional, procura o zero de $f'(x)$ através da recorrência:

$$x_{k+1} = x_k - f''(x_k)^{-1} f'(x_k), \quad \text{para} \quad k = 0, 1, \ldots$$

- No caso multidimensional o método de Newton é usado para procurar o zero do gradiente da função, $\nabla f(\mathbf{x}) = 0$, através da formula de recorrência:

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \mathbf{H}_f^{-1}(\mathbf{x}_k) \nabla f(\mathbf{x}_k), \quad \text{para} \quad k = 0, 1, \ldots$$

em que $\mathbf{H}(\mathbf{x})$ é a matriz Hessiana das segundas derivadas parciais

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

Método de Newton, continuação

- A matriz Hessiana não é invertida explicitamente, em vez disso resolve-se o sistema linear

$$\mathbf{H}_f(\mathbf{x}_k)\delta_k = -\nabla f(\mathbf{x})$$

em ordem a δ_k, e depois actualiza-se a solução aproximada

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \delta_k$$

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

Exemplo 2: Método de Newton

- Usar o método de Newton para minimizar

$$f(\mathbf{x}) = 0.5x_1^2 + 2.5x_2^2$$

- Gradiente e matriz Hessiana são dados por

$$\nabla f(\mathbf{x}) = \begin{bmatrix} x_1 \\ 5x_2 \end{bmatrix} \quad \text{e} \quad \mathbf{H}_f(\mathbf{x}) = \begin{bmatrix} 1 & 0 \\ 0 & 5 \end{bmatrix}$$

- Escolhendo $\mathbf{x}_0 = \begin{bmatrix} 5 \\ 1 \end{bmatrix}$ obtemos $\nabla f(\mathbf{x}_0) = \begin{bmatrix} 5 \\ 5 \end{bmatrix}$
- Resolvendo o sistema linear $\mathbf{H}_f(\mathbf{x}_0)\delta_0 = -\nabla f(\mathbf{x_0})$, obtemos o passo $\delta_0 = \begin{bmatrix} -5 \\ -1 \end{bmatrix}$, e $\mathbf{x}_1 = \mathbf{x}_0 + \delta_0 = \begin{bmatrix} 5 \\ 1 \end{bmatrix} + \begin{bmatrix} -5 \\ -1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ que é a solução exacta deste problema

Método do Gradiente Conjugado

- Um outro método que não requer explicitamente as segundas derivadas é o método do *gradiente conjugado* (CG)
- CG gera uma sequência de direcções de procura conjugadas (ortogonais) entre elas
- Teoricamente, para funções objectivo quadráticas o CG converge para a solução exacta num numero máximo de n iterações, em que n é a dimensão do problema

Algoritmo do Gradiente Conjugado

$\mathbf{x}_0 =$ aproximação inicial

$\mathbf{g}_0 = \nabla f(\mathbf{x}_0)$

$\mathbf{s}_0 = -\mathbf{g}_0$

for $k = 0, 1, 2, \ldots$

Escolher o α_k que minimiza $f(\mathbf{x}_k + \alpha_k \mathbf{s}_k)$

$\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{s}_k$

$\mathbf{g}_{k+1} = \nabla f(\mathbf{x}_{k+1})$

$\beta_{k+1} = \left(\mathbf{g}_{k+1}^T \mathbf{g}_{k+1}\right) / \left(\mathbf{g}_k^T \mathbf{g}_k\right)$

$\mathbf{s}_{k+1} = -\mathbf{g}_{k+1} + \beta_{k+1} \mathbf{s}_k$

end

Exemplo 3: Método do Gradiente Conjugado

- Usar o CG para minimizar

$$f(\mathbf{x}) = 0.5x_1^2 + 2.5x_2^2$$

- Gradiente é igual a $\nabla f(\mathbf{x}) = \begin{bmatrix} x_1 \\ 5x_2 \end{bmatrix}$
- Escolhendo $\mathbf{x}_0 = \begin{bmatrix} 5 & 1 \end{bmatrix}$ obtemos a direcção de procura inicial (corresponde ao gradiente negativo)

$$\mathbf{s}_0 = -\mathbf{g}_0 = -\nabla f(\mathbf{x}_0) = \begin{bmatrix} -5 \\ -5 \end{bmatrix}$$

- O mínimo exacto ao longo da linha de procura é $\alpha_0 = 1/3$, pelo que a próxima solução aproximada é $\mathbf{x}_1 = \begin{bmatrix} 3.333 & -0.667 \end{bmatrix}^T$ com o qual calculamos o novo gradiente

$$\mathbf{g}_1 = \nabla f(\mathbf{x}_1) = \begin{bmatrix} 3.333 \\ -3.333 \end{bmatrix}$$

Exemplo, continuação

- Neste ponto, em vez de procurar na direcção do gradiente negativo calcula-se

$$\beta_1 = \left(\mathbf{g}_1^T \mathbf{g}_1\right) / \left(\mathbf{g}_0^T \mathbf{g}_0\right) = 0.444$$

que origina a nova direcção de procura

$$\mathbf{s}_1 = -\mathbf{g}_1 + \beta_1 \mathbf{s}_0 = \begin{bmatrix} -3.333 \\ 3.333 \end{bmatrix} + 0.444 \begin{bmatrix} -5 \\ -5 \end{bmatrix} = \begin{bmatrix} -5.556 \\ 1.111 \end{bmatrix}$$

- A deslocação ao longo desta linha que minimiza a função é $\alpha_1 = 0.6$, que origina a solução exacta $\mathbf{x}_2 = \begin{bmatrix} 0 & 0 \end{bmatrix}^T$, tal como seria de esperar para uma função quadrática

Métodos Disponíveis na NMLibforOctave

- Método de Newton: `[] = opt_newton()`
- Mét. Gradiente Conjugado: `[] = opt_cg()`

Bibliografia

Exposição baseada essencialmente no capítulo 6 de

- Michael T. Heath. "Scientific Computing an Introductory Survey". McGraw-Hill, 2002, New York.