Título: Métodos de Otimização Irrestritos

Autor: Tibério Bittencourt de Oliveira Martins

e-mail: tibas.b@uol.com.br

Unidade Acadêmica: Instituto de Matemática e Estatística

Palavras-chave: gradiente, Newton, Quase-Newton

1. Introdução

Neste trabalho, o problema de otimização a ser estudado é o seguinte:

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

onde f é uma função real. Vamos apresentar soluções através de métodos clássicos como método do Gradiente, método de Newton e métodos de direções conjugadas.

2. Metodologia

A metodologia adotada foi a seguinte:

- i) Levantamento bibliográfico sobre o assunto;
- ii) Reuniões semanais e seminários;
- iii) Investigação dos conceitos básicos da análise convexa, álgebra linear e análise numérica.

3. Resultados e discussões

Seja $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $f \in C^1$. Visto que o gradiente aponta na direção de maior crescimento da função no ponto, o método do gradiente procura em cada ponto caminhar na direção oposta ao gradiente, $-\nabla f(x)$. De um ponto x_k , efetuaremos uma busca linear pesquisando ao longo da direção do negativo do gradiente, $-\nabla f(x_k)$, pelo ponto mínimo nessa reta, esse mínimo será x_{k+1} , ou seja, dado o ponto x_k temos que:

$$x_{k+1} = x_k - \lambda_k \nabla f(x_k) \quad (2)$$

onde $\lambda_k \geq 0$ minimiza $f(x_k - \lambda \nabla f(x_k))$. Pelo teorema de convergência global, se a sequência $\{x_k\}$ é limitada, terá pontos de limite e cada um deles será uma solução. Basta investigarmos então se existe tal limite.

Teorema 1. *Seja $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $f \in C^2$. Seja $x^* \in \mathbb{R}^n$ um minimizador local de f , tal que a matriz $H(x^*)$ seja definida positiva. Se $\{x_k\}$ é uma sequência gerada pelo método do gradiente que converge para x^* , então a sequência $\{f(x_k)\}$ converge linearmente a $f(x^*)$ com taxa de convergência não superior a $\left(\frac{A-a}{A+a}\right)^2$. Onde a e A são o menor e o maior autovalores de $\nabla^2 f(x_k)$.*

O método do gradiente se baseia em aproximações lineares, ou de primeira ordem, de Taylor a f . O método de Newton estende a idéia aproveitando aproximações quadráticas a f . Perto de um ponto x_k , podemos aproximar f pela expansão truncada de Taylor por

$$f_Q(x) = f(x_k) + \nabla f(x_k)^T(x - x_k) + \frac{1}{2}(x - x_k)^T H(x_k)(x - x_k) \quad (3)$$

onde $H(x_k)$ é a matriz hessiana $\nabla^2 f(x_k)$.

Por ser uma condição suficiente, assumiremos que em um ponto de mínimo relativo x^* , a matriz hessiana, $H(x^*)$, é definida positiva e $\nabla f(x^*) = 0$. E então, as condições de ponto mínimo de (3) induzem o processo de iteração de Newton:

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

Teorema 2. *Seja $f \in C^3$ em $\mathbb{R}^n$, e assumamos que num ponto de mínimo local x^* , a hessiana $H(x^*)$ é definida positiva. Então se iniciado suficientemente perto de x^* , os pontos gerados pelo método de Newton convergem para x^* . A ordem de convergência é pelo menos dois.*

O método de Newton garante uma ótima convergência, no entanto é grande o número de cálculos e armazenagem de dados. Os métodos de direções conjugadas foram desenhados para acelerar o raio de convergência do método do gradiente e, ao mesmo tempo, diminuir o trabalho requerido no método de Newton. O método foi primeiramente desenvolvido para o problema quadrático e depois é estendido ao caso geral.

Definição 1. *Dada uma matriz simétrica Q , dois vetores d_1 e d_2 são ditos Q -ortogonais, ou Q -conjugados, se $d_1^T Q d_2 = 0$.*

Teorema 3. (de direções conjugadas para o caso quadrático)

Seja $\{d_i\}_{i=1}^n$ um conjunto de vetores Q -conjugados. Para qualquer $x_1 \in \mathbb{R}^n$, a seqüência $\{x_k\}$ que é gerada conforme

$$x_{k+1} = x_k + \lambda_k d_k, \quad k \geq 0 \quad (4)$$

$$\lambda_k = \frac{-g_k^T d_k}{d_k^T Q d_k}, \quad \text{onde } g_k = Qx_k - b \quad (5)$$

converge para solução única, x^ , de $Qx = b$ após n iterações, isto é, $x^* = x_n$.*

O próximo teorema se refere ao método dos gradientes conjugados (M.G.C.) para o caso quadrático, que consiste em construir as direções conjugadas a partir dos gradientes, aplicando o processo de ortogonalização de Gram-Schmidt.

Teorema 4. *As direções conjugadas geradas pelo M.G.C. são dadas por:*

$$d_0 = -g_0$$

$$d_k = -g_k + \mu_k d_k, \quad k = 1, \dots, n-1$$

$$\mu_k = \frac{g_k^T g_k}{g_{k-1}^T g_{k-1}}$$

Além disso, o método termina com uma solução ótima após n iterações.

No entanto, se a função objetivo não é quadrática, o método já não converge necessariamente em um número finito de passos. E para adaptá-lo ao caso geral, vamos reiniciá-lo a cada n iterações: Este é o método de Fletcher - Reeves, cujo algoritmo executa o método do gradiente de n em n passos, e por isso sua convergência é garantida.

4. Conclusão

O método do gradiente é muito mais simples de ser implementado, porém de convergência muito lenta. O método de Newton é de implementação muito mais complicada por envolver o gradiente e a matriz hessiana da função, porém de convergência quadrática. Surge daí, a idéia de métodos intermediários, como o método dos Gradientes conjugados que tenta melhorar a convergência do método do Gradiente e diminuir os cálculos de matrizes inversas do método de Newton. Estes métodos são baseados em propriedades da função quadrática e por isso convergem em um número finito de iterações para este caso. Como o método do gradiente não tem essa forte propriedade espera-se que quando estendido ao caso geral, o método dos gradientes conjugados também convirja mais rápido que o primeiro.

5. Bibliografia

- [1] Luenberger, D.G. Linear and nonlinear programming 2^a ed., Nova York, Addison - Wesley Publishing Company, 1986.
- [2] Friedlander, A.; Elementos de programação não-linear, Campinas, SP, Editora da UNICAMP, 1994.
- [3] Bazaraa, M.; Shetty, C.M. Nonlinear programming - theory and algorithms. Nova York, John Wiley and Sons, 1979.
- [4] Bertsekas, Dimitri P.; Nonlinear programming. Belmont, Mass., Athena Scientific, 1979.

6. Fonte de financiamento: CNPQ