EEL 510247 - Métodos Numéricos de Otimização I

Métodos de Otimização Irrestrita

Agosto, 2018

0.1 Métodos de Direções Conjugadas

Esses métodos podem ser vistos como intermediários entre o método do gradiente e o método de Newton. Eles possuem uma melhor convergência do que o método do gradiente porém não requerem a montagem e a inversão de uma matriz como acontece com o método de Newton. Atualmente tais métodos vêm sendo usados com sucesso na resolução de sistemas lineares de grande porte.

Seja o problema puramente quadrático expresso como

$$Min \quad \frac{1}{2}\mathbf{x}^T\mathbf{Q}\mathbf{x} - \mathbf{b}^T\mathbf{x} \tag{1}$$

onde $\mathbf{Q}$ é uma matriz $n \times n$ simétrica positiva definida.

Os conceitos apresentados preliminarmente para este problema serão posteriormente estendidos a problemas não lineares genéricos.

Dada uma matriz simétrica $\mathbf{Q}$, dois vetores $\mathbf{d}_1$ e $\mathbf{d}_2$ são ditos $\mathbf{Q}$-ortogonais, ou conjugados em relação a $\mathbf{Q}$, se $\mathbf{d}_1^T\mathbf{Q}\mathbf{d}_2 = 0$.

Observação 1 *: Se* $\mathbf{Q}$ *é positiva definida, e o conjunto de vetores não-nulos* $\mathbf{d}_1, \mathbf{d}_2, ..., \mathbf{d}_k$ *são* $\mathbf{Q}$*-ortogonais, então estes vetores são linearmente independentes.*

Considere o problema de minimização 1. Das condições necessárias de otimalidade e das características do problema quadrático, pode ser inferido que a única solução para este problema é também a solução para a equação linear

$$\mathbf{Q}\mathbf{x} = \mathbf{b} \tag{2}$$

e, portanto, o problema de minimização representado pela equação 1 é equivalente ao problema linear mostrado na equação 2.

É possível associar à matriz $\mathbf{Q}$ $n \times n$ os vetores $\mathbf{Q}$-ortogonais $\mathbf{d}_0, \mathbf{d}_1, ..., \mathbf{d}_{n-1}$. Uma vez que estes vetores são linearmente independentes, a solução do problema expresso em 1 (ou em 2), denotada $\mathbf{x}^*$, pode ser expressa como

$$\mathbf{x}^* = \alpha_0\mathbf{d}_0 + \alpha_1\mathbf{d}_1 + ... + \alpha_{n-1}\mathbf{d}_{n-1}$$

para um dado conjunto de $\alpha_i's$.

Os valores dos $\alpha_i's$ podem ser obtidos multiplicando-se a expressão anterior por $\mathbf{Q}$ e fazendo o produto escalar com $\mathbf{d}_i$; isto é,

$$\mathbf{d}_i^T\mathbf{Q}\mathbf{x}^* = \alpha_0\mathbf{d}_i^T\mathbf{Q}\mathbf{d}_0 + \alpha_1\mathbf{d}_i^T\mathbf{Q}\mathbf{d}_1 + ... + \alpha_{n-1}\mathbf{d}_i^T\mathbf{Q}\mathbf{d}_{n-1}$$

ou ainda

$$\alpha_i = \frac{\mathbf{d}_i^T \mathbf{Q} \mathbf{x}^*}{\mathbf{d}_i^T \mathbf{Q} \mathbf{d}_i} = \frac{\mathbf{d}_i^T \mathbf{b}}{\mathbf{d}_i^T \mathbf{Q} \mathbf{d}_i}$$

A expansão para $\mathbf{x}^*$ pode ser considerada como sendo o resultado de um processo iterativo de n passos onde no passo i é adicionado $\alpha^i \mathbf{d}^i$. Vendo o procedimento desta forma, e permitindo um ponto inicial para as iterações, obtém-se o algoritmo básico de direções conjugadas.

Teorema 1 : *Seja $\{\mathbf{d}^i\}_{i=0}^{n-1}$ uma seqüência gerada pelos vetores $\mathbf{Q}$-ortogonais. Então para todo $\mathbf{x}_0 \in E^n$, a seqüência $\{\mathbf{x}^k\}$ gerada de acordo com*

$$\mathbf{x}^{k+1} = \mathbf{x}^k + \alpha^k \mathbf{d}^k$$

com

$$\alpha^k = -\frac{\left(\mathbf{g}^k\right)^T \mathbf{d}^k}{\left(\mathbf{d}^k\right)^T \mathbf{Q} \mathbf{d}^k}$$

e

$$\mathbf{g}^k = \mathbf{Q}\mathbf{x}^k - \mathbf{b}$$

converge para a solução única, $\mathbf{x}^{ast}$, de $\mathbf{Q}\mathbf{x} = \mathbf{b}$ após n iterações, ou seja, $\mathbf{x}^n = \mathbf{x}^{ast}$.

Até este ponto o método do gradiente conjugado foi derivado essencialmente pela observação de que resolver o problema 1 é equivalente a resolver o sistema 2. Este ponto de vista, embora importante, ignora alguns aspectos do algoritmo que são especialmente importantes quando o mesmo é estendido a problemas não quadráticos. A seguir são colocadas algumas propriedades do método:

1. Seja $\mathcal{B}^k$ o subespaço de E^n gerado por $\{\mathbf{d}^0, \mathbf{d}^1, \ldots, \mathbf{d}^{k-1}\}$. A medida em que o método de direções conjugadas progride, cada $\mathbf{x}^k$ minimiza a função objetivo de 1 na *variedade linear* $\mathbf{x_0} + \mathcal{B}^k$.

2. No método das direções conjugadas, os gradientes $\mathbf{g}^k,\ k = 1, \ldots, n$ satisfazem a equação $\mathbf{g}^{kt}\mathbf{d}^i = 0$ *para* $i < k$.

Definição 1 *Um conjunto V em E^n é uma* variedade linear *se, dados quaisquer $\mathbf{x}^1, \mathbf{x}^2 \in V$, tem-se $\lambda\mathbf{x}^1 + (1-\lambda)\mathbf{x}^2 \in V$ para todos os números reais λ.*

O método do gradiente conjugado é obtido selecionando-se as sucessivas direções de minimização como uma versão conjungada dos sucessivos gradientes obtidos na medida em que a busca da solução ótima progride. No passo k avalia-se o negativo do gradiente e adiciona-se a este vetor uma combinação das direções previamente usadas para se obter uma nova direção conjugada ao longo da qual a busca prossegue. Existem vantagens para se adotar esse método de seleção de direção de minimização. Em primeiro lugar, a menos que a solução seja atingida em menos do que n iterações, o gradiente é sempre não nulo e ortogonal ao subespaço gerado pelas direções nanteriores (isto é, $\mathbf{g}^k$ é ortogonal a $\mathcal{B}^k$ gerado por $\{\mathbf{d}^0, \mathbf{d}^1, \ldots, \mathbf{d}^{k-1}\}$. Se a solução é atingida em menos do que n iterações, o gradiente é nulo e o processo termina. A segunda vantagem é que a fórmula usada para determinar o novo vetor de direção é relativamente simples. Por fim, devido ao fato das direções serem baseadas nos gradientes, o método progride uniformemente para a solução em cada iteração.

Os passos para a aplicação do Método das Direções Conjugadas são descritos a seguir.

1. Passo Inicial: Defina a tolerância ε. A partir de qualquer $\mathbf{x}^0 \in E^n$ faça

$$\mathbf{d}^0 = -\mathbf{g}^0 = \mathbf{b} - \mathbf{Q}\mathbf{x}^0$$

   com $k = 0$ e prossiga ao Passo Principal;

2. Passo Principal: Enquanto $\left|\mathbf{g}^k\right|_\infty \geq \varepsilon$, faça

$$\begin{aligned} \mathbf{x}^{k+1} &= \mathbf{x}^k + \alpha^k \mathbf{d}^k \quad \textit{com} \quad \alpha^k = -\frac{\left(\mathbf{g}^k\right)^T \mathbf{d}^k}{\left(\mathbf{d}^k\right)^T \mathbf{Q}\mathbf{d}^k} \\ \mathbf{d}^{k+1} &= -\mathbf{g}^{k+1} + \beta^k \mathbf{d}^k \quad \textit{com} \quad \beta^k = -\frac{\left(\mathbf{g}^{k+1}\right)^T \mathbf{Q}\mathbf{d}^k}{\left(\mathbf{d}^k\right)^T \mathbf{Q}\mathbf{d}^k} \\ \mathbf{g}^{k+1} &= \mathbf{Q}\mathbf{x}^{k+1} - \mathbf{b} \end{aligned}$$

3. Faça $k = k + 1$ e repita o passo principal.

No algoritmo o primeiro passo é identico a um passo do método do gradiente; cada passo posterior é feito numa direção que é uma combinação linear do gradiente corrente e o vetor de direção da iteração anterior.

Problemas genéricos de otimização, com a forma **??**podem ser resolvidos fazendo-se pequenas alterações no algoritmo descrito acima. O método descrito a seguir é conhecido por *Gradiente Conjugado de Fletcher-Reeves*. O algoritmo pode ser resumido nos seguintes passos:

Passo 1. Dado $\mathbf{x}^0$, calcule $\mathbf{g}^0 = \nabla f(\mathbf{x}^0)$ e faça $\mathbf{d}^0 = -\mathbf{g}^0$.

Passo 2. Para $k = 0, 1, \ldots, n-1$:

   1. Faça $\mathbf{x}^{k+1} = \mathbf{x}^k + \alpha^k \mathbf{d}^k$ onde α^k minimiza $f(\mathbf{x}^k + \alpha^k \mathbf{d}^k)$.
   2. Calcule $\mathbf{g}^{k+1} = \nabla f(\mathbf{x}^{k+1})$.
   3. Caso $k \neq n-1$, faça $\mathbf{d}^{k+1} = -\mathbf{g}^{k+1} + \beta^k \mathbf{d}^k$ onde

$$\beta^k = \frac{(\mathbf{g}^{k+1})^t \mathbf{g}^{k+1}}{(\mathbf{g}^k)^t \mathbf{g}^k}.$$

Passo 3. Substitua $\mathbf{x}^0$ por $\mathbf{x}^n$ e retorne ao Passo 1.

Exemplo 1 *Dê uma iteração com o método do gradiente conjugdo de Fletcher-Reeves para minimizar a função*

$$f(\mathbf{x}) = \frac{5}{2}{x_1}^2 + 2{x_2}^2 + 4x_1x_2 + 3x_1 + 2x_2$$

a partir do ponto $\mathbf{x}^0 = (3, 3)$.