Método dos Gradientes Conjugados

Departamento de Matemática
Universidade do Minho

O método do gradiente conjugado linear foi proposto por Hestenes e Stiefel em 1950. É um método iterativo para resolver sistemas lineares em que a matriz dos coeficientes é definida positiva. Este método é uma alternativa ao método de Gauss por permitir resolver sistemas de grandes dimensões.

O primeiro método do gradiente conjugado não linear foi introduzido por Fletcher e Reeves em 1960. Foi uma das primeiras técnicas para resolver problemas não lineares de grande dimensão, e algumas são largamente usadas na prática. As principais vantagens deste algoritmo são que não requer o armazenamento da matriz e é mais rápido que o método do gradiente.

Método do Gradiente Conjugado

Pretende-se resolver um sistema linear

$$Aw = -b \tag{1}$$

onde A é uma matriz definida positiva e simétrica. Este sistema linear pode ser reformulado como o seguinte problema de minimização:

$$\underset{w \in \mathbb{R}^d}{\text{minimizar}}\ F(w) = \frac{1}{2} w^T A w + b^T w, \tag{2}$$

ou seja, (1) e (2) têm a mesma solução única.
Note-se que :

$$\nabla F(w) = Aw + b.$$

Método do Gradiente Conjugado

Definição 1

Um conjunto de vetores não nulos $\{s^{(0)}, s^{(1)}, \ldots, s^{(l)}\}$ é definido como conjunto A - conjugado com respeito à matriz simétrica definida positiva A se

$$s^{(i)^T} A s^{(j)} = 0, \quad \forall i \neq j, \quad i, j \leq d$$

Proposição 1

Seja A uma matriz simétrica definida positiva. Se $\{s^{(0)}, s^{(1)}, \ldots, s^{(l)}\}$ é o conjunto conjugado relativamente a A, então estes vetores são linearmente independentes.

Demonstração.

Vamos assumir que $\{s^{(0)}, s^{(1}, \ldots, s^{(l)}\}$ é um conjunto linearmente dependente. Então sem perda de generalidade, existem $\alpha_1, \alpha_2, \ldots, \alpha_{k-1}$ tal que

$$s^{(k)} = \alpha_1 s^{(0)} + \alpha_2 s^{(1)} + \ldots + \alpha_{k-1} s^{(k-1)}.$$

Demonstração.

Isto leva à seguinte contradição

$$0 < s^{(k)^T} A s^{(k)} = s^{(k)^T} A \left(\sum_{i=1}^{k-1} \alpha_i s^{(i)} \right) = \sum_{i=0}^{k-1} \alpha_i s^{(k)^T} A s^{(i)} = 0,$$

pela definição de conjunto conjugado. $\square$

Proposição 2

Para vetores A - conjugados $s^{(0)}, s^{(1)}, \ldots, s^{(d-1)}$ *e para qualquer ponto inicial* $w^{(0)} \in \mathbb{R}^d$*, seja* $w^{(1)}, \ldots w^{(d)}$ *definida recursivamente por:*

$$w^{(k+1)} = w^{(k)} + \eta_k s^{(k)}, \; \textit{com } \eta_k = -\frac{\nabla F(w^{(k)})^T s^{(k)}}{s^{(k)^T} A s^{(k)}}.$$

Então, $F(w^{(k+1)}) = \min_{\eta \in \mathbb{R}} F(w^{(k)} + \eta s^{(k)})$, $\nabla F(w^{(k+1)})^T s^{(k)} = 0$ *e*

$$F(w^{(d)}) = \min_{w \in \mathbb{R}^d} F(w).$$

Demonstração.

Observe que:

$$0 = \frac{d}{d\eta} F(w^{(k)} + \eta s^{(k)})_{\eta=\eta_k} = \nabla F(w^{(k+1)})^T s^{(k)}.$$

Os vetores $s^{(0)}, s^{(1)}, \ldots, s^{(d-1)}$ são linearmente independentes uma vez que se assumiu que são A - conjugados. Logo, eles forma uma base de $\mathbb{R}^d$. De

$$w^{(d)} = w^{(d-1)} + \eta_{d-1} s^{(d-1)} = \ldots = w^{(j+1)} + \sum_{i=j+1}^{d-1} \eta_i s^{(i)},$$

para $0 \leq j \leq d-1$ tem-se que:

$$\nabla F(w^{(d)}) = \nabla F(w^{(j+1)}) + \sum_{i=j+1}^{d-1} \eta_i A s^{(i)}$$

e

$$\nabla F(w^{(d)})^T s^{(j)} = \nabla F(w^{(j+1)})^T s^{(j)} + \sum_{i=j+1}^{d-1} \eta_i {s^{(i)}}^T A s^{(j)} = 0.$$

Logo, $\nabla F(w^{(d)}) = 0$. Donde se conclui que $F(w^{(d)}) = \min_{w \in \mathbb{R}^d} F(w)$. $\square$

Algoritmo: CG de Hestenes-Stiefel

1. Dar: $w^{(0)}$
2. Determinar: $\nabla F(w^{(0)})$ e $s^{(0)} = -\nabla F(w^{(0)})$
3. Fazer: $k = 0$
4. **Enquanto** $\|\nabla F(w^{(k)})\| \neq 0$
5. Fazer $\eta_k = -\frac{\nabla F(w^{(k)})^T s^{(k)}}{s^{(k)T} A s^{(k)}}$
6. Fazer $w^{(k+1)} = w^{(k)} + \eta_k s^{(k)}$
7. Fazer $\beta_{k+1} = \frac{\|\nabla F(w^{(k+1)})\|^2}{\|\nabla F(w^{(k)})\|^2}$
8. Fazer $s^{(k+1)} = -\nabla F(w^{(k+1)}) + \beta_{k+1} s^{(k)}$
9. Fazer $k = k + 1$
10. **Fim enquanto**

Exercício: Use o algortimo de CG de Hestenes-Stiefel para determinar a solução do problema $\underset{w \in \mathbb{R}^2}{\text{minimizar}} \frac{1}{2}(w_1^2 + 9w_2^2)$ com $w^{(0)} = (9, 1)^T$ e tolerância $\epsilon = 10^{-8}$.

Método do Gradiente Conjugado no caso não quadrático

A primeira versão apareceu em 1964 com a FLETCHER/REEVES. Uma variante foi apresentada por POLAK/RIBIÉRE em 1971, propuseram considerar para β_{k+1}:

$$\frac{\nabla F(w^{(k+1)})^T \left(\nabla F(w^{(k+1)}) - \nabla F(w^{(k)})\right)}{\nabla F(w^{(k)})^T \nabla F(w^{(k})}$$

em vez de

$$\frac{\nabla F(w^{(k+1)})^T \nabla F(w^{(k+1)})}{\nabla F(w^{(k)})^T \nabla F(w^{(k})}.$$

Algoritmo: **CG de FLETCHER - REEVES com Recomeço**

1. Dar: $w^{(0)}$
2. Determinar: $\nabla F(w^{(0)})$ e $s^{(0)} = -\nabla F(w^{(0)})$
3. Fazer: $k = 1$
4. **Enquanto** $\nabla F(w^{(k)}) \neq 0$
5. Determinar usando a procura exata ou não exata de $\eta_k > 0$

$$F(w^{(k)} + \eta_k s^{(k)}) \approx \min\{F(w^{(k)} + \eta s^{(k)}) : \eta > 0\}$$

6. Fazer $w^{(k+1)} = w^{(k)} + \eta_k s^{(k)}$
7. Fazer $\beta_{k+1} =$

$$\begin{cases} 0 & \text{se } k+1 \equiv 0 \ mod \ n \text{ (Recomeço)} \\ \dfrac{\nabla F(w^{(k+1)})^T \left(\nabla F(w^{(k+1)}) - \nabla F(w^{(k)})\right)}{\nabla F(w^{(k)})^T \nabla F(w^{(k)})} & \text{caso contrário.} \end{cases}$$

8. Fazer $s^{(k+1)} = -\nabla F(w^{(k+1)}) + \beta_{k+1} s^{(k)}$
9. Fazer $k = k + 1$
10. **Fim enquanto**

Exercício: Considere-se o problema

$$\underset{w\in\mathbb{R}^2}{\text{minimizar}} \sqrt{w_1^2+1+} + \sqrt{w_2^2+1}$$

utilize o algoritmo CG de FLETCHER - REEVES com recomeço para determinar a solução com $w^0 = (0.5, 0.5)^T$ com tolerância $\epsilon = 10^{-8}$.

Os apontamentos foram baseados na seguinte bibliografia: [1] e [2] .

W. Forst and D. Hoffmann.
Optimization—theory and practice.
Springer Science & Business Media, 2010.

J. Nocedal and S. J. Wright.
Numerical optimization.
Springer, 1999.