Resolução de sistemas de equações lineares: Método dos Gradientes Conjugados

Marina Andretta/Franklina Toledo

ICMC-USP

21 de setembro de 2012

Baseado no livro Cálculo Numérico, de Neide B. Franco

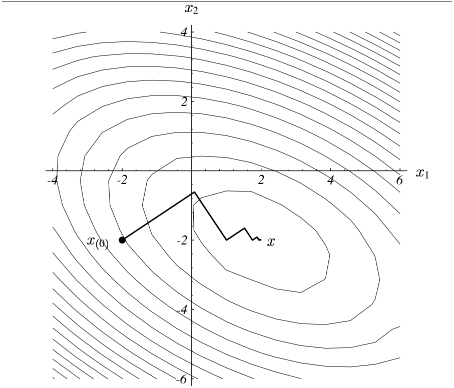
Estamos interessados em resolver o sistema linear

$$Ax = b,$$

com $A \in \boldsymbol{R}^{n \times n}$ e $b \in \boldsymbol{R}^n$ dados.

Quando a matriz A é esparsa, podemos não tê-la representada, mas dispor apenas de rotinas que calculem o produto de A por um dado vetor v. Neste caso, não é possível usar os métodos diretos ou os Métodos Jacobi-Richardson ou Gauss-Seidel.

Uma possibilidade, como já vimos, é utilizar o Método dos Gradientes. Uma desvantagem deste método é que sua convergência costuma ser muito lenta.



Fonte: Shewchuk, J R. *An Introduction to the Conjugate Gradient Method Without the Agonizing Pain* Edition 1. School of Computer Science Carnegie Mellon University Pittsburgh. PA, 1994.

Uma alternativa para resolver um sistema linear que veremos agora é o Método dos Gradientes Conjugados.

Este é um método iterativo, que parte da mesma ideia do Método dos Gradientes (minimizar uma função quadrática). No entanto, veremos que há um limitante para o número de iterações necessárias para que o Método dos Gradientes Conjugados convirja à solução do sistema linear.

Assim como no Método dos Gradientes, trocaremos o problema de encontrar uma solução para o sistema $Ax = b$ pelo problema equivalente de encontrar um minimizador de $\frac{1}{2}x^TAx - b^Tx$, com A simétrica e definida positiva.

Dada a função $f(x) = \frac{1}{2}x^TAx - b^Tx$, lembre que:

$$\nabla f(x) = Ax - b;$$

$$\nabla^2 f(x) = A.$$

Encontrar a solução do sistema linear $Ax = b$ é equivalente a encontrar o ponto x que satisfaz $\nabla f(x) = Ax - b = 0$, ou seja, o minimizador da função f.

Antes de descrever o Método dos Gradientes Conjugados, precisamos da seguinte definição:

Definição 1: *Dada uma aplicação linear A definida positiva, duas direções x e y são ditas conjugadas se*

$$(Ax)^T y = y^T Ax = 0.$$

A primeira iteração ($k = 1$) do Método dos Gradientes Conjugados será igual à primeira iteração do Método dos Gradientes.

Ou seja, calculamos o resíduo

$$r_0 = Ax_0 - b$$

e definimos a direção

$$p_1 = -r_0.$$

O novo ponto é dado por

$$x_1 = x_0 + tp_1,$$

com

$$t = q_1 = -\frac{r_0^T p_1}{(Ap_1)^T p_1} = \frac{r_0^T r_0}{(Ar_0)^T r_0}.$$

(igual ao método do gradiente)

Portanto,

$$x_1 = x_0 - \frac{r_0^T r_0}{(Ar_0)^T r_0} r_0. \tag{1}$$

Para definir a direção p_k, para cada iteração $k > 1$, tomamos uma direção que seja conjugada à direção p_{k-1}, isto é, queremos que

$$(Ap_k)^T p_{k-1} = p_k^T A p_{k-1} = 0.$$

Além disso, tomamos p_k como uma combinação linear de r_{k-1} e p_{k-1}.

Como o coeficiente de r_{k-1} nesta combinação linear não pode ser nulo, definimo-lo como -1. Assim,

$$p_k = -r_{k-1} + \alpha_{k-1} p_{k-1}. \tag{2}$$

Precisamos, então, definir o valor de α_{k-1}.

Note que

$$p_k^T A p_{k-1} = 0 \Rightarrow (-r_{k-1} + \alpha_{k-1} p_{k-1})^T A p_{k-1} = 0 \Rightarrow$$

$$-r_{k-1}^T A p_{k-1} + \alpha_{k-1} p_{k-1}^T A p_{k-1} = 0.$$

Ou seja,

$$\alpha_{k-1} = \frac{r_{k-1}^T A p_{k-1}}{p_{k-1}^T A p_{k-1}}, \tag{3}$$

para $k > 1$.

Calculada a direção p_k, precisamos definir o tamanho de passo q_k para calcular

$$x_k = x_{k-1} + q_k p_k. \tag{4}$$

O tamanho de passo q_k é definido como o minimizador da função $\frac{1}{2}x^T Ax - b^T x$ na direção p_k. Ou seja,

$$q_k = -\frac{r_{k-1}^T p_k}{(Ap_k)^T p_k}. \tag{5}$$

Note que α_{k-1} e q_k, definidos em (3) e (5), respectivamente, sempre são maiores do que zero.

Como $r_k = Ax_k - b$, temos que

$$r_k = A(x_{k-1} + q_k p_k) - b = Ax_{k-1} - b + q_k A p_k \Rightarrow$$

$$r_k = r_{k-1} + q_k A p_k. \tag{6}$$

O Método dos Gradientes Conjugados é definido pelas fórmulas (1) a (6).

No entanto, podemos usar algumas propriedades sobre o resíduo r_k para simplificar os cálculos de α_{k-1} e q_k.

Método dos Gradientes Conjugados

- O resíduo na iteração k (r_k) é ortogonal ao resíduo na iteração $k-1$ (r_{k-1}). Ou seja,

$$r_k^T r_{k-1} = 0.$$

- O resíduo na iteração k (r_k) é ortogonal à direção calculada na iteração k (p_k). Ou seja,

$$r_k^T p_k = 0.$$

- O resíduo na iteração k (r_k) é ortogonal à direção calculada na iteração $k-1$ (p_{k-1}). Ou seja,

$$r_k^T p_{k-1} = 0.$$

Método dos Gradientes Conjugados

Lembre-se que

$$q_k = -\frac{r_{k-1}^T p_k}{(Ap_k)^T p_k}.$$

Com estas propriedades, temos que

$$-(r_{k-1}^T p_k) = -r_{k-1}^T(-r_{k-1} + \alpha_{k-1} p_{k-1}) =$$

$$r_{k-1}^T r_{k-1} - \alpha_{k-1} r_{k-1}^T p_{k-1} = r_{k-1}^T r_{k-1}$$

Ou seja,

$$q_k = \frac{r_{k-1}^T r_{k-1}}{(Ap_k)^T p_k}.$$

O valor de α_{k-1} é dado por

$$\alpha_{k-1} = \frac{r_{k-1}^T A p_{k-1}}{p_{k-1}^T A p_{k-1}}.$$

Como

$$r_k = r_{k-1} + q_k A p_k,$$

temos que

$$A p_{k-1} = \frac{1}{q_{k-1}}(r_{k-1} - r_{k-2}).$$

Usando as propriedades mencionadas, temos que

$$r_{k-1}^T A p_{k-1} = r_{k-1}^T \left(\frac{1}{q_{k-1}} (r_{k-1} - r_{k-2}) \right) =$$

$$\frac{1}{q_{k-1}} r_{k-1}^T r_{k-1} - \frac{1}{q_{k-1}} r_{k-1}^T r_{k-2} \Rightarrow$$

$$r_{k-1}^T A p_{k-1} = \frac{1}{q_{k-1}} r_{k-1}^T r_{k-1}.$$

Além disso,

$$p_{k-1}^T A p_{k-1} = p_{k-1}^T \left(\frac{1}{q_{k-1}} (r_{k-1} - r_{k-2}) \right) =$$

$$\frac{1}{q_{k-1}} p_{k-1}^T r_{k-1} - \frac{1}{q_{k-1}} p_{k-1}^T r_{k-2} = -\frac{1}{q_{k-1}} p_{k-1}^T r_{k-2} =$$

$$-\frac{1}{q_{k-1}} (-r_{k-2} + \alpha_{k-2} p_{k-2})^T r_{k-2} = -\frac{1}{q_{k-1}} (-r_{k-2}^T r_{k-2} + \alpha_{k-2} p_{k-2}^T r_{k-2}) =$$

$$\frac{1}{q_{k-1}} r_{k-2}^T r_{k-2} - \frac{\alpha_{k-2}}{q_{k-1}} p_{k-2}^T r_{k-2} \Rightarrow$$

$$p_{k-1}^T A p_{k-1} = \frac{1}{q_{k-1}} r_{k-2}^T r_{k-2}.$$

Portanto,

$$\alpha_{k-1} = \frac{r_{k-1}^T A p_{k-1}}{p_{k-1}^T A p_{k-1}} \Rightarrow$$

$$\alpha_{k-1} = \frac{r_{k-1}^T r_{k-1}}{r_{k-2}^T r_{k-2}}.$$

Algoritmo

Método dos Gradientes Conjugados: dadas a dimensão n, uma matriz $A \in \mathbb{R}^{n \times n}$ simétrica definida positiva, um vetor $b \in \mathbb{R}^{n}$, uma aproximação inicial x_0, uma tolerância $TOL > 0$ e o número máximo de iterações $MAXIT$, devolve x_k uma aproximação da solução de $Ax = b$ ou emite uma mensagem de erro.

Passo 1: Faça $r_0 \leftarrow Ax_0 - b$.

Passo 2: Se $\|r_0\| < \epsilon$, então

devolva x_0 como solução e pare.

Passo 3: Faça $p_1 \leftarrow -r_0$ e $q_1 \leftarrow \frac{r_0^T r_0}{(Ar_0)^T r_0}$.

Passo 4: Faça $x_1 \leftarrow x_0 + q_1 p_1$.

Passo 5: Se $\|x_1 - x_0\| < \epsilon$ ou $\frac{\|x_1 - x_0\|}{\|x_1\|} < \epsilon$, então

devolva x_1 como solução e pare.

Passo 6: Faça $r_1 \leftarrow r_0 + q_1 A p_1$ e $k \leftarrow 2$.

Algoritmo - continuação

Passo 7: Enquanto $k \leq MAXIT$, execute os passos 8 a 14:

Passo 8: Se $\|r_{k-1}\| < \epsilon$, então

devolva x_{k-1} como solução e pare.

Passo 9: Faça $\alpha_{k-1} \leftarrow \frac{r_{k-1}^T r_{k-1}}{r_{k-2}^T r_{k-2}}$.

Passo 10: Faça $p_k \leftarrow -r_{k-1} + \alpha_{k-1} p_{k-1}$.

Passo 11: Faça $q_k \leftarrow \frac{r_{k-1}^T r_{k-1}}{(Ap_k)^T p_k}$.

Passo 12: Faça $x_k \leftarrow x_{k-1} + q_k p_k$.

Passo 13: Se $\|x_k - x_{k-1}\| < \epsilon$ ou $\frac{\|x_k - x_{k-1}\|}{\|x_k\|} < \epsilon$, então

devolva x_k como solução e pare.

Passo 14: Faça $r_k \leftarrow r_{k-1} + q_k A p_k$ e $k \leftarrow k + 1$.

Passo 15: Escreva "o método falhou após $MAXIT$ iterações" e pare.

Convergência

Teorema 1: *No Método dos Gradientes Conjugados, as direções p_k formam um sistema de direções conjugadas e os resíduos formam um sistema ortogonal. Isto é, para $i, j = 1, 2, ...,$ $i \neq j$,*

$$(Ap_i)^T p_j = 0$$

e

$$r_i^T r_j = 0.$$

Teorema 2: *Dado um sistema linear* $Ax = b$, $A \in \mathbb{R}^{n \times n}$, $b, x \in \mathbb{R}^n$, *o Método dos Gradientes Conjugados fornece a solução do sistema em, no máximo, n iterações.*

Exemplo

Vamos usar o Método dos Gradientes Conjugados para encontrar a solução do sistema linear

$$\begin{pmatrix} 10 & 1 & 0 \\ 1 & 10 & 1 \\ 0 & 1 & 10 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 11 \\ 11 \\ 1 \end{pmatrix}$$

com precisão 10^{-2}, usando o ponto inicial $x_0 = (0, 0, 0)^T$.

Exemplo

Na primeira iteração ($k = 1$), temos que

$$r_0 = \begin{pmatrix} -11 \\ -11 \\ -1 \end{pmatrix},$$

$p_1 = -r_0$, $q_1 = t_{\min} = 0.0902$,

$$x_1 = \begin{pmatrix} 0.9922 \\ 0.9922 \\ 0.0902 \end{pmatrix} \quad \text{e} \quad r_1 = \begin{pmatrix} -0.0858 \\ 0.0044 \\ 0.8942 \end{pmatrix}.$$

Exemplo

Para $k = 2$, temos

$$\alpha_1 = \frac{r_1^T r_1}{r_0^T r_0} = \frac{0.8070}{243} = 0.0033$$

e

$$p_2 = -r_1 + \alpha_1 p_1 = \begin{pmatrix} 0.0858 \\ -0.0044 \\ -0.8942 \end{pmatrix} + 0.0033 \begin{pmatrix} 11 \\ 11 \\ 1 \end{pmatrix} = \begin{pmatrix} 0.1221 \\ 0.0319 \\ -0.8909 \end{pmatrix}.$$

Exemplo

Assim,

$$Ap_2 = \begin{pmatrix} 10 & 1 & 0 \\ 1 & 10 & 1 \\ 0 & 1 & 10 \end{pmatrix} \begin{pmatrix} 0.1221 \\ 0.0319 \\ -0.8909 \end{pmatrix} = \begin{pmatrix} 1.2529 \\ -0.4498 \\ -8.8771 \end{pmatrix},$$

$$(Ap_2)^T p_2 = 0.1530 - 0.0143 + 7.9086 = 8.0473$$

e

$$q_2 = \frac{r_1^T r_1}{(Ap_2)^T p_2} = \frac{0.8070}{8.0473} = 0.1003.$$

Exemplo

Portanto,

$$x_2 = x_1 + q_2 p_2 = \begin{pmatrix} 0.9922 \\ 0.9922 \\ 0.0902 \end{pmatrix} + 0.1003 \begin{pmatrix} 0.1221 \\ 0.0319 \\ -0.8909 \end{pmatrix} = \begin{pmatrix} 1.0044 \\ 0.9954 \\ 0.0008 \end{pmatrix}$$

e

$$r_2 = r_1 + q_2 A p_2 = \begin{pmatrix} 0.0399 \\ -0.0407 \\ 0.0038 \end{pmatrix}.$$

Exemplo

Para $k = 3$, temos

$$\alpha_2 = \frac{r_2^T r_2}{r_1^T r_1} = \frac{0.0033}{0.8070} = 0.0041$$

e

$$p_3 = -r_2 + \alpha_2 p_2 = \begin{pmatrix} -0.0394 \\ 0.0408 \\ -0.0001 \end{pmatrix}.$$

Exemplo

Assim,

$$Ap_3 = \begin{pmatrix} -0.3532 \\ 0.3685 \\ 0.0398 \end{pmatrix},$$

$$(Ap_3)^T p_3 = 0.0139 + 0.0150 - 0.0000 = 0.0289$$

e

$$q_3 = \frac{r_2^T r_2}{(Ap_3)^T p_3} = \frac{0.0033}{0.0289} = 0.1142.$$

Portanto,

$$x_3 = x_2 + q_3 p_3 = \begin{pmatrix} 0.9999 \\ 1.0001 \\ 0.0008 \end{pmatrix}.$$

Note que

$$\frac{\|x_3 - x_2\|_\infty}{\|x_3\|_\infty} = \frac{0.0047}{1.0001} \approx 0.0047 < 10^{-2}.$$