Método iterativo para solução de sistemas lineares Gradientes e Gradientes Conjugados

Silvia Maria Pereira Grandi dos Santos

USP - São Carlos/SP

Outubro 2008

Roteiro

- Motivação;
- Processos de Relaxação;
- Estudo do caso $f(x) = \frac{1}{2}x^T Ax - b^T x + x$;
- Método dos Gradientes;
- Direção Conjugada;
- Método dos Gradientes Conjugados;
- Exemplo resolvido.

Motivação

Considere o sistema linear $Ax = b$. Deseja-se encontrar a solução x desse sistema.

Muitas vezes, usar Métodos Exatos para encontrar a solução desse sistema é tão eficiente quanto usar Métodos Iterativos, mas a situação pode se complicar quando usamos Métodos Exatos em sistemas cuja matriz dos coeficientes, A, é uma matriz esparsa (cheia de zeros).

O Método dos Gradientes é utilizado quando a matriz A é simétrica definida positiva.

Processos de Relaxação

- Dado um sistema $Ax = b$ e um ponto para a aproximação inicial $x^{(0)}$ da solução do sistema, queremos reduzir o resíduo dado por
  $r^{(0)} = b - Ax^{(0)}$
  até que este resíduo seja nulo.

- Para que o resíduo diminua, tomamos uma direção v e corrigimos $x^{(0)}$ nesta direção, ou seja,
  $x^{(1)} = x^{(0)} + \lambda\vec{v}$ de forma que $r^{(1)} = b - Ax^{(1)} < r^{(0)}$.

- Dessa forma, construimos uma sequência $\{x^{(k)}\}$ que converge para a solução do sistema $Ax = b$, pois estamos considerando que o resíduo diminua em cada passo do processo iterativo, ou seja,
  $r^{(k+1)} < r^{(k)}$.

- Quando $k \to \infty$ temos $r^{(k)} \to 0$ e, dessa forma, $Ax^{(k)} \to b$ e $k^{(k)}$ é uma aproximação para a solução do sitema dado $Ax = b$.

Estudo de um caso específico

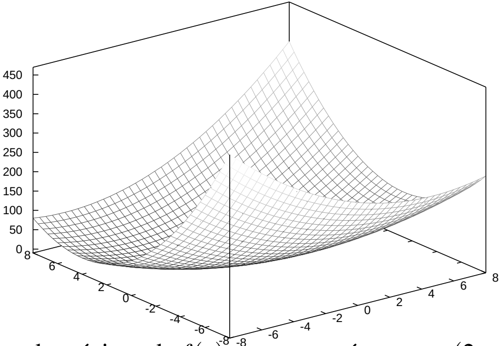
Para o estudo da função $f(x) = \frac{1}{2}x^T Ax - b^T x + c$ consideremos as seguintes notações:

- $x^T y$ representa o produto escalar de x com y no espaço $\mathbb{R}^n$, assim
$$x^T y = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \sum_{i=1}^{n} x_i y_i$$
- Se x e y são vetores ortogonais, então $x^T y = 0$
- A é uma matriz definida positiva se, para todo vetor x não nulo, tem-se $x^T Ax > 0$
- Lembre-se que $(AB)^T = B^T A^T$.

Estudo de um caso específico

Considere $f : \mathbb{R}^n \to \mathbb{R}$ em que $f(x) = \dfrac{1}{2}x^T A x - b^T x + c$, A é uma matrix $n \times n$, b e x são vetores $n \times 1$ e c uma constante.

Sejam $A = \begin{bmatrix} 3 & 2 \\ 2 & 6 \end{bmatrix}$, $b = \begin{bmatrix} 2 \\ -8 \end{bmatrix}$ e $c = 0$. Nestas condições, o gráfico da função f pode ser visto na Figura 6.



O ponto de mínimo de $f(x)$ neste caso é o ponto $(2, -2)$.

Estudo de um caso específico

É possível mostrar que o ponto de mínimo de $f(x)$ é a solução do sistema $Ax = b$. De fato,
sabemos que o ponto de mínimo de $f(x)$ é um ponto crítico, ou seja, $f^{'}(x) = \nabla f(x) = 0$.

Mas $f : \mathbb{R}^n \to \mathbb{R}$; $f(x) = \dfrac{1}{2}\displaystyle\sum_{i,j=1}^{n} a_{ij}x_i x_j - \sum_{i=1}^{n} b_i x_i + c$. Assim

$$\nabla f(x) = \begin{bmatrix} \frac{\partial}{\partial x_1} f(x) \\ \frac{\partial}{\partial x_2} f(x) \\ \vdots \\ \frac{\partial}{\partial x_n} f(x) \end{bmatrix} = \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \cdots a_{1n}x_n - b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots a_{2n}x_n - b_2 \\ \vdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots a_{nn}x_n - b_n \end{bmatrix} = Ax - b$$

Fazendo $\nabla f(x) = 0$ temos $Ax - b = 0 \quad \Rightarrow Ax = b$, ou seja,
o ponto crítico de $f(x)$ é o ponto x tal que $Ax = b$. Nos resta saber se $x; Ax = b$ é de mínimo ou de máximo.

Estudo de um caso específico

Sabemos que um ponto x é ponto de mínimo de $f : \mathbb{R}^n \to \mathbb{R}$ se $J(f)$ é positiva definida.

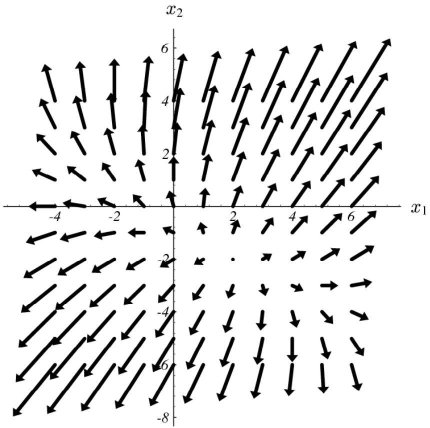
Como $J = \begin{bmatrix} \dfrac{\partial^2 f(x)}{\partial x_1^2} & \dfrac{\partial^2 f(x)}{\partial x_1 \partial x_2} & \cdots & \dfrac{\partial^2 f(x)}{\partial x_1 \partial x_n} \\ \dfrac{\partial^2 f(x)}{\partial x_2 \partial x_1} & \dfrac{\partial^2 f(x)}{\partial x_2^2} & \cdots & \dfrac{\partial^2 f(x)}{\partial x_2 \partial x_n} \\ \vdots & & & \\ \dfrac{\partial^2 f(x)}{\partial x_n \partial x_1} & \dfrac{\partial^2 f(x)}{\partial x_n \partial x_2} & \cdots & \dfrac{\partial^2 f(x)}{\partial x_n^2} \end{bmatrix}$

$$= \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & & & \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} = A$$

e A é uma matriz positiva definida por hipótese, segue **o ponto** x **tal que** $Ax = b$ **é ponto de mínimo de** $f(x)$. Dessa forma, se encontrarmos o ponto de mínimo de $f(x)$ encontramos a solução do sistema $Ax = b$.

Estudo de um caso específico

Sabemos que $\nabla f(x)$, num dado ponto x, aponta a direção de maior crescimento da função f.



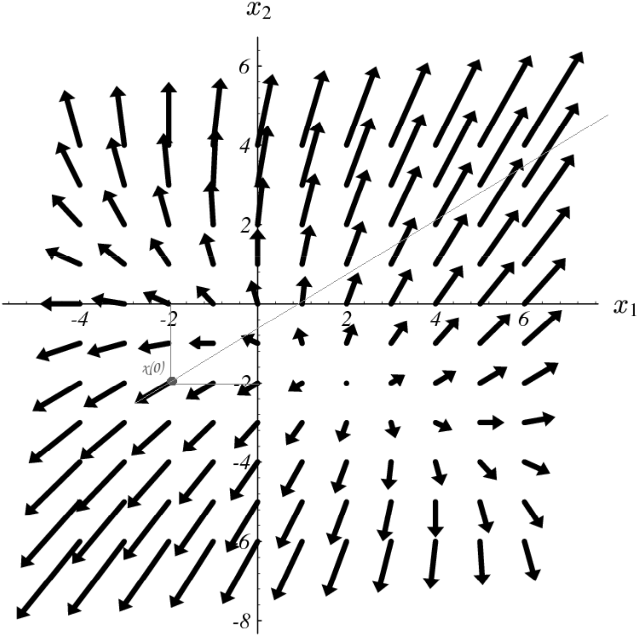
Método dos Gradientes

Dada uma aproximação inicial $x^{(0)}$ para a solução do sistema $Ax = b$, é seria natural pensarmos em tomar a direção oposta a $\nabla f(x)$ para a correção de $x^{(0)}$, pois $-\nabla f(x)$ aponta na direção que f decresce mais rapidamente (o ponto de mínimo de f é a solução do sistema $Ax = b$).

Pelos cálculos feitos anteriormente, sabemos que $-\nabla f(x^{(i)}) = b - Ax^{(i)} = r^{(i)}$.

Dessa forma, tomando $x^{(0)} = (-2, -2)$, sabemos em qual direção caminhar, mas não sabemos o quanto caminhar.

Método dos Gradientes



Método dos Gradientes

Como $x^{(1)} = x^{(0)} + \lambda r^{(0)}$ e λ minimiza $f(x^{(1)})$ quando a derivada direcional $\dfrac{\partial f(x^{(1)})}{\partial \lambda} = 0.$

Mas $\dfrac{\partial f(x^{(1)})}{\partial \lambda} = \left(f^{'}(x^{(1)})\right)^T \dfrac{\partial x^{(1)}}{\partial \lambda} = \left(f^{'}(x^{(1)})\right)^T r^{(0)}.$

Igualando essa expressão a zero temos $\left(f^{'}(x^{(1)})\right)^T r^{(0)} = 0$ que nos sugere que λ será escolhido de forma que $r^{(0)}$ e $f^{'}(x^{(1)})$ sejam ortogonais.

Para determinar λ usamos o fato de $f^{'}(x^{(1)} = -r^{(1)}$. Dessa forma, teremos

Método dos Gradientes

$\left(r^{(1)}\right)^T r^{(0)} = 0$ Substituindo $r^{(1)}$ nessa expressão

$\left(b - Ax^{(1)}\right)^T r^{(0)} = 0$ Como $x^{(1)} = x^{(0)} + \lambda r^{(0)}$ temos

$$\left(b - A\left(x^{(0)} + \lambda r^{(0)}\right)\right)^T r^{(0)} = 0$$

$$\left(b - Ax^{(0)}\right)^T r^{(0)} - \lambda\left(Ar^{(0)}\right)^T r^{(0)} = 0$$

$$\left(b - Ax^{(0)}\right)^T r^{(0)} = \lambda\left(Ar^{(0)}\right)^T r^{(0)}$$

$$\left(r^{(0)}\right)^T r^{(0)} = \lambda\left(r^{(0)}\right)^T Ar^{(0)}$$

Logo

$$\lambda = \frac{\left(r^{(0)}\right)^T r^{(0)}}{\left(r^{(0)}\right)^T Ar^{(0)}}$$

Método dos Gradientes

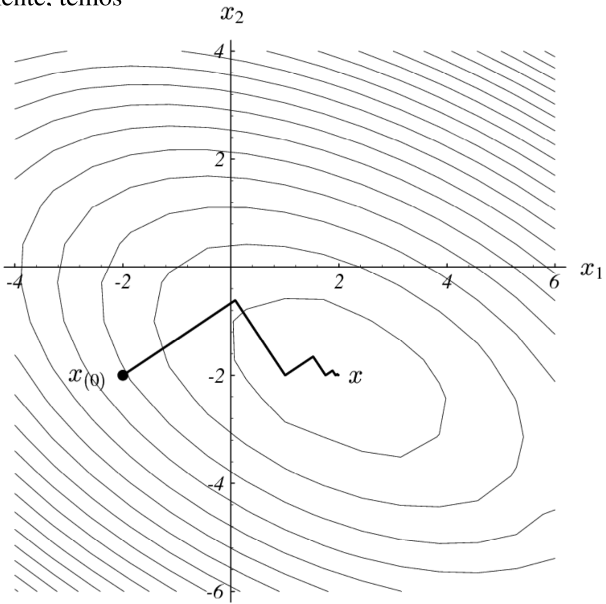
Colocando todas essas informações juntas, temos

dada uma aproximação inicial $x^{(0)}$ para a solução do sistema $Ax = b$, o processo iterativo, conhecido como **Método dos Gradientes**, é dado por

- $r^{(i)} = b - Ax^{(i)}$
- $\lambda^{(i)} = \dfrac{\left(r^{(0)}\right)^T r^{(0)}}{\left(r^{(0)}\right)^T A r^{(0)}}$
- $x^{(i+1)} = x^{(i)} + \lambda^{(i)} r^{(i)}$

Método dos Gradientes

Intuitivamente, temos



Método dos Gradientes

Exemplo: Usando o Método dos Gradientes, obtenha uma solução aproximada para $\begin{bmatrix} 4 & 1 \\ 1 & 3 \end{bmatrix} x = \begin{bmatrix} 5 \\ 4 \end{bmatrix}$ com $x^{(0)} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ e $\epsilon = 10^{-1}$.

Abrir arquivo Exemplo1.xls e arquivo Exemplo1.mws

Método dos Gradientes Conjugados

Método dos Gradientes Conjugados

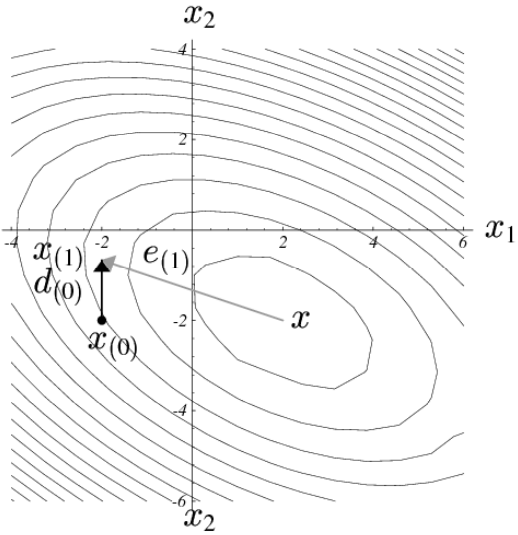
É possível, no Método dos Gradientes, uma direção que está sendo adotada na iteração i ter sido usada em iterações anteriores, como pode ser visto na Figura 15.

Para evitar que se tome várias vezes uma mesma direção $r^{(i)}$ para a correção de $x^{(i)}$, o Método dos Gradientes Conjugados propõe uma modificação no Método dos Gradientes.

O Método dos Gradientes Conjugados sugere que, dada uma aproximação inicial $x^{(0)}$ para o sistema $n \times n,\ Ax = b$, tomemos um conjunto de direções conjugadas $\{d_0, d_1, \cdots, d_{n-1}\}$ e em até n iterações, teremos encontrado uma aproximação satisfatória para a solução do sistema.

Dois vetores x e y são conjugados se $x^T A y = y^T A x = 0$.

Método dos Gradientes Conjugados



Método dos Gradientes Conjugados

Dado $x^{(0)}$, faça $d^0 = r^0 = b - Ax$ e $x^{(1)}$ será obtido fazendo $x^{(1)} = x^{(0)} + \lambda^{(0)} d^{(0)}$.

Para obter o tamanho do passo $\lambda^{(0)}$, procedemos como no caso do Método dos Gradientes:

$$\frac{d\left(f(x^{(1)})\right)}{d\lambda} = f'\left(x^{(1)}\right)^T \frac{dx^{(1)}}{d\lambda} = -r^{(1)T} d^{(0)}.$$

Como queremos que λ minimize $f\left(x^{(1)}\right)$, fazemos

$$r^{(1)T} d^{(0)} = 0 \quad \rightarrow \quad d^{(0)T} r^{(1)} = 0 \quad \rightarrow \quad d^{(0)T}\left(b - Ax^{(1)}\right) = 0$$

Substituindo $x^{(1)}$ nessa expressão concluimos que

$$\lambda^{(0)} = \frac{r^{(0)T} r^{(0)}}{d^{(0)T} A d^{(0)}}$$

Método dos Gradientes Conjugados

Para que $r^{(1)} = r^{(0)} - \lambda^{(0)} A d^{(0)}$ seja ortogonal à nova direção $d^{(1)} = r^{(1)} + \beta^{(1)} d^{(0)}$ é preciso escolher um valor para $\beta^{(1)}$ conveniente. Assim

$$r^{(1)T} d^{(1)} = 0 \quad \rightarrow \quad \left(r^{(0)} - \frac{r^{(0)T} r^{(0)}}{d^{(0)T} A d^{(0)}} A d^{(0)} \right)^T \left(r^{(1)} + \beta^{(1)} d^{(0)} \right) = 0$$

$$r^{(0)T} r^{(1)} + \beta^{(1)} r^{(0)T} d^{(0)} - d^{(0)T} A \frac{r^{(0)T} r^{(0)}}{d^{(0)T} A d^{(0)}} r^{(1)} - \beta^{(1)} d^{(0)T} A \frac{r^{(0)T} r^{(0)}}{d^{(0)T} A d^{(0)}} d^{(0)} = 0$$

Simplificando temos $\beta^{(1)} = -\dfrac{d^{(0)T} A r^1}{d^{(0)T} A d^{(0)}} = -\dfrac{r^{(1)T} A d^0}{d^{(0)T} A d^{(0)}}$

Assim a nova direção é dada por $d^{(1)} = r^{(1)} + \beta^{(1)} d^{(0)}$

Método dos Gradientes Conjugados

Podemos resumir o Método dos Gradientes Conjugados por, dado $x^{(0)}$ inicial

- $d^{(0)} = r^{(0)} = b - Ax^{(0)}$
- Para $i = 0, 1, \ldots$
  - $\lambda^{(i)} = \dfrac{r^{(i)T} r^{(i)}}{d^{(i)T} A d^{(i)}}$
  - $x^{(i+1)} = x^{(i)} + \lambda^{(i)} d^{(i)}$
  - $r^{(i+1)} = r^{(i)} - \lambda^{(i)} A d^{(i)}$
  - $\beta^{(i+1)} = -\dfrac{r^{(i+1)T} A d^{(i)}}{d^{(i)T} A d^{(i)}}$
  - $d^{(i+1)} = r^{(i+1)} + \beta^{(i+1)} d^{(i)}$

Método dos Gradientes Conjugados

Podemos simplificar um pouco mais a expressão de $\beta^{(i+1)}$ considerando $r^{(i+1)} = r^{(i)} - \lambda^{(i)} A d^{(i)}$, pois dessa forma temos

$$Ad^{(i)} = -\frac{1}{\lambda^{(i)}} \left(r^{(i+1)} - r^{(i)} \right)$$

Fazendo $r^{(i+1)} A d^{(i)}$ e $d^{(i)T} A d^{(i)}$ temos
$$r^{(i+1)T} A d^{(i)} = -\frac{1}{\lambda^{(i)}} \left(r^{(i+1)T} r^{(i+1)} - r^{(i+1)T} r^{(i)} \right) = -\frac{1}{\lambda^{(i)}} \left(r^{(i+1)T} r^{(i+1)} \right)$$

e
$$d^{(i)T} A d^{(i)} = -\frac{1}{\lambda^{(i)}} \left(d^{(i)T} r^{(i+1)} - d^{(i)T} r^{(i)} \right) = -\frac{1}{\lambda^{(i)}} \left(-d^{(i)T} r^{(i)} \right) = \frac{1}{\lambda^{(i)}} d^{(i)T} r^{(i)}$$

Método dos Gradientes Conjugados

Substituindo $\quad r^{(i+1)T}Ad^{(i)} = -\frac{1}{\lambda^{(i)}}\left(r^{(i+1)T}r^{(i+1)}\right) \quad$ e $d^{(i)T}Ad^{(i)} = \frac{1}{\lambda^{(i)}}d^{(i)T}r^{(i)} \quad$ em $\quad \beta^{(i+1)}$ temos

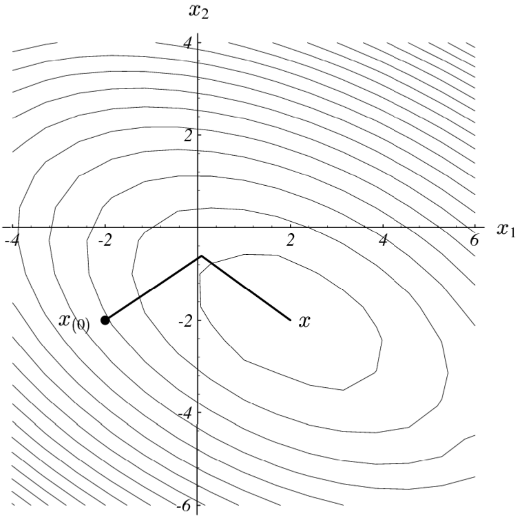
$$\beta^{(i+1)} = -\frac{-\frac{1}{\lambda^{(i)}}\left(r^{(i+1)T}r^{(i+1)}\right)}{\frac{1}{\lambda^{(i)}}d^{(i)T}r^{(i)}} = \frac{r^{(i+1)T}r^{(i+1)}}{d^{(i)T}r^{(i)}} = \frac{r^{(i+1)T}r^{(i+1)}}{r^{(i)T}r^{(i)}}$$

Método dos Gradientes Conjugados

Dessa forma, dado $x^{(0)}$ uma aproximação inicial do sistema $Ax = b$, o Método dos Gradientes Conjugados é dado por

- $d^{(0)} = r^{(0)} = b - Ax^{(0)}$
- Para $i = 0, 1, ...$
  - $\lambda^{(i)} = \dfrac{r^{(i)T} r^{(i)}}{d^{(i)T} A d^{(i)}}$
  - $x^{(i+1)} = x^{(i)} + \lambda^{(i)} d^{(i)}$
  - $r^{(i+1)} = r^{(i)} - \lambda^{(i)} A d^{(i)}$
  - $\beta^{(i+1)} = \dfrac{r^{(i+1)T} r^{(i+1)}}{r^{(i)T} r^{(i)}}$
  - $d^{(i+1)} = r^{(i+1)} + \beta^{(i+1)} d^{(i)}$

Método dos Gradientes Conjugados



Método dos Gradientes Conjugados

Exemplo: Obtenha uma solução aproximada para o sistema
$\begin{bmatrix} 4 & 1 \\ 1 & 3 \end{bmatrix} x = \begin{bmatrix} 5 \\ 4 \end{bmatrix}$ com $x^{(0)} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ e $\epsilon = 10^{-1}$.

Abrir arquivo Exemplo2.xls e arquivo Exemplo2.mws

Referências

- Shewchuk, J R. *An Introduction to the Conjugate Gradient Method Without the Agonizing Pain* Edition $1\frac{1}{4}$. School of Computer Science Carnegie Mellon University Pittsburgh. PA, 1994.

- Franco, N B. Cálculo Numérico. Editora: Pearson / Prentice Hal. 2006.

- Todas as figuras (exceto a primeira) foram extraídas de Shewchuk, 1994.