CIMAT

Gradiente biconjugado

Miguel Vargas-Félix

miguelvargas@cimat.mx
http://www.cimat.mx/~miguelvargas

Gradiente biconjugado

El método de gradiente biconjugado se basa en el método de gradiente conjugado y al igual que éste sirve para resolver systemas de ecuaciones lineales

$$\mathbf{A}\,\mathbf{x}=\mathbf{b},$$

con la diferencia que $\mathbf{A}\in\mathbb{R}^{m\times m}$ no tiene que ser simétrica.

Este método requiere calcular un pseudo-gradiente $\tilde{\mathbf{r}}_k$ y una pseudo-dirección de descenso $\tilde{\mathbf{p}}_k$.

El método se construye de tal forma que que los pseudo-gradientes $\tilde{\mathbf{r}}_k$ sean ortogonales a los gradientes $\mathbf{r}_k$ y que las pseudo-direcciónes de descenso $\tilde{\mathbf{p}}_k$ sean $\mathbf{A}$-ortogonales al las direcciones de descenso $\mathbf{p}_k$ [Meie94 pp6-7].

Si la matriz $\mathbf{A}$ es simétrica, entonces el método es equivalente al gradiente conjugado.

No garantiza convergencia en m pasos como lo hace el gradiente conjugado.

Requerimos ahora hacer dos multiplicaciones matrix-vector. El algoritmo sería [Meie94]...

$\mathbf{x}_0$, coordenada inicial
$\mathbf{r}_0 \leftarrow \mathbf{b} - \mathbf{A}\,\mathbf{x}_0$, gradiente inicial
$\tilde{\mathbf{r}}_0 \leftarrow \mathbf{r}_0$, pseudo-gradiente inicial
$\mathbf{p}_0 \leftarrow \mathbf{r}_0$, dirección de descenso inicial
$\tilde{\mathbf{p}}_0 \leftarrow \mathbf{r}_0$, pseudo-dirección de descenso inicial
ε, tolerancia.
$k \leftarrow 0$
mientras $\|\mathbf{r}_k\| > \varepsilon$

$$\mathbf{w} \leftarrow \mathbf{A}\,\mathbf{p}_k$$
$$\tilde{\mathbf{w}} \leftarrow \mathbf{A}^{\mathrm{T}}\tilde{\mathbf{p}}_k$$
$$\alpha_k \leftarrow \frac{\tilde{\mathbf{r}}_k^{\mathrm{T}}\mathbf{r}_k}{\tilde{\mathbf{p}}_k^{\mathrm{T}}\mathbf{w}}$$
$$\mathbf{x}_{k+1} \leftarrow \mathbf{x}_k + \alpha_k\,\mathbf{p}_k$$
$$\mathbf{r}_{k+1} \leftarrow \mathbf{r}_k - \alpha\,\mathbf{w}$$
$$\tilde{\mathbf{r}}_{k+1} \leftarrow \tilde{\mathbf{r}}_k - \alpha\,\tilde{\mathbf{w}}$$
$$\beta_k \leftarrow \frac{\tilde{\mathbf{r}}_{k+1}^{\mathrm{T}}\mathbf{r}_{k+1}}{\tilde{\mathbf{r}}_k^{\mathrm{T}}\mathbf{r}_k}$$
$$\mathbf{p}_{k+1} \leftarrow \mathbf{r}_{k+1} + \beta_{k+1}\,\mathbf{p}_k$$
$$\tilde{\mathbf{p}}_{k+1} \leftarrow \tilde{\mathbf{r}}_{k+1} + \beta_{k+1}\,\tilde{\mathbf{p}}_k$$
$$k \leftarrow k+1$$

Gradiente biconjugado precondicionado

Sigue la misma idea que el gradiente conjugado precondicionado, utiliza un precondicionador para mejorar la convergencia del algoritmo. Hay que hacer dos multiplicaciones matrix-vector. Hay que resolver dos sistemas de ecuaciones en cada paso.

$\mathbf{x}_0$, coordenada inicial

$\mathbf{r}_0 \leftarrow \mathbf{b} - \mathbf{A}\,\mathbf{x}_0$, gradiente inicial

$\tilde{\mathbf{r}}_0 \leftarrow \mathbf{r}_0$, pseudo-gradiente inicial

$\mathbf{q}_0 \leftarrow \mathbf{M}^{-1}\mathbf{r}_0$

$\tilde{\mathbf{q}}_0 \leftarrow \left(\mathbf{M}^{-1}\right)^{\mathrm{T}}\tilde{\mathbf{r}}_0$

$\mathbf{p}_0 \leftarrow \mathbf{q}_0$, dirección de descenso inicial

$\tilde{\mathbf{p}}_0 \leftarrow \tilde{\mathbf{q}}_0$, pseudo-dirección de descenso

ε, tolerancia

$k \leftarrow 0$

mientras $\|\mathbf{r}_k\| > \varepsilon$

$$\mathbf{w} \leftarrow \mathbf{A}\,\mathbf{p}_k$$
$$\tilde{\mathbf{w}} \leftarrow \mathbf{A}^{\mathrm{T}}\tilde{\mathbf{p}}_k$$
$$\alpha_k \leftarrow \frac{\tilde{\mathbf{r}}_k^{\mathrm{T}}\mathbf{q}_k}{\tilde{\mathbf{p}}_k^{\mathrm{T}}\mathbf{w}}$$
$$\mathbf{x}_{k+1} \leftarrow \mathbf{x}_k + \alpha_k\,\mathbf{p}_k$$
$$\mathbf{r}_{k+1} \leftarrow \mathbf{r}_k - \alpha_k\,\mathbf{w}$$
$$\tilde{\mathbf{r}}_{k+1} \leftarrow \tilde{\mathbf{r}}_k - \overline{\alpha_k}\,\tilde{\mathbf{w}}$$
$$\mathbf{q}_{k+1} \leftarrow \mathbf{M}^{-1}\mathbf{r}_{k+1}$$
$$\tilde{\mathbf{q}}_{k+1} \leftarrow \left(\mathbf{M}^{-1}\right)^{\mathrm{T}}\tilde{\mathbf{r}}_{k+1}$$
$$\beta_k \leftarrow \frac{\tilde{\mathbf{r}}_{k+1}^{\mathrm{T}}\mathbf{q}_{k+1}}{\tilde{\mathbf{r}}_k^{\mathrm{T}}\mathbf{q}_k}$$
$$\mathbf{p}_{k+1} \leftarrow \mathbf{q}_{k+1} + \beta_{k+1}\,\mathbf{p}_k$$
$$\tilde{\mathbf{p}}_{k+1} \leftarrow \tilde{\mathbf{q}}_{k+1} + \beta_{k+1}\,\tilde{\mathbf{p}}_k$$
$$k \leftarrow k+1$$

Precondicionador LU incompleto

Es análogo al precondicionador Cholesky incompleto, pero ahora el precondicionador está formado por dos matrices

$$\mathbf{M} = \mathbf{G}_t \mathbf{H}_t,$$

dónde $\mathbf{G}_t$ es una matriz triangular inferior rala y $\mathbf{H}_t$ una matriz triangular superior rala que tienen estructuras similares a la factorización $\mathbf{L}\,\mathbf{U}$ de $\mathbf{A} \in \mathbb{R}^{m \times m}$.

La estructura de $\mathbf{G}_0$ será igual a la estructura de la parte triangular inferior de $\mathbf{A}$.
La estructura de $\mathbf{G}_m$ será igual a la estructura de $\mathbf{L}$ (factorización LU de $\mathbf{A}$.)
La estructura de $\mathbf{H}_0$ será igual a la estructura de la parte triangular superior de $\mathbf{A}$.
La estructura de $\mathbf{H}_m$ será igual a la estructura de $\mathbf{U}$ (factorización LU de $\mathbf{A}$.)

Los valores de $\mathbf{G}_t$ y $\mathbf{H}_t$ se llenarán utilizando la fórmulas de la factorización $\mathbf{L}\,\mathbf{U}$, así

$$H_{ij} = A_{ij} - \sum_{k=1}^{j-1} G_{ik} H_{kj} \ \text{ para } i > j,$$

$$H_{ii} = A_{ii} - \sum_{k=1}^{i-1} G_{ik} H_{ki},$$

$$G_{ji} = \frac{1}{H_{ii}} \left(A_{ji} - \sum_{k=1}^{i-1} G_{jk} H_{ki} \right) \text{ para } i > j,$$

$$G_{ii} = 1.$$

Factorización Cholesky simbólica truncada

Para una matriz rala $\mathbf{A}$, definamos

$$\boldsymbol{a}_i \stackrel{\text{def}}{=} \{k>i \mid A_{ik}\neq 0\},\ i=1\ldots n,$$

como el conjunto de los índices de los elementos no nulos del renglón i de la parte estrictamente triangular superior de $\mathbf{A}$.

De forma análoga definimos para la matriz $\mathbf{H}_t$, los conjuntos

$$\boldsymbol{h}_i \stackrel{\text{def}}{=} \{k>i \mid H_{ik}\neq 0\},\ i=1\ldots n.$$

Para la matriz ejemplo siguiente, el conjuto $\boldsymbol{a}_2$ contendrá $\{3,4\}$; el conjunto $\boldsymbol{h}_2$ contendrá $\{3,4,6\}$.

$$\mathbf{A}=\begin{pmatrix} A_{11} & A_{12} & & & & A_{16} \\ A_{21} & A_{22} & \boxed{A_{23} & A_{24} & &} \\ & A_{32} & A_{33} & & A_{35} & \\ & A_{42} & & A_{44} & & \\ & & A_{53} & & A_{55} & A_{56} \\ A_{61} & & & & A_{65} & A_{66} \end{pmatrix} \qquad \mathbf{H}_t^{\mathrm{T}}=\begin{pmatrix} H_{11} & H_{12} & & & & H_{16} \\ & H_{22} & H_{23} & H_{24} & & H_{26} \\ & & H_{33} & H_{34} & H_{35} & H_{36} \\ & & & H_{44} & H_{45} & H_{46} \\ & & & & H_{55} & H_{56} \\ & & & & & H_{66} \end{pmatrix}$$

$$\boldsymbol{a}_2=\{3,4\} \qquad \boldsymbol{h}_2=\{3,4,6\}$$

Requeriremos de conjuntos $\boldsymbol{r}_i$ que serán usados para registrar las columnas de $\mathbf{H}_t$ cuyas estructuras afectarán al renglón i de $\mathbf{H}_t$.

$$
\begin{array}{l}
\text{for } i \leftarrow 1,2,\ldots,n \\
\cdot \quad \boldsymbol{r}_i \leftarrow \emptyset \\
\\
\text{for } i \leftarrow 1,2,\ldots,n \\
\cdot \quad \boldsymbol{h}_i \leftarrow \boldsymbol{a}_i \\
\cdot \quad \text{for each } j \in \boldsymbol{r}_i \\
\cdot \quad \cdot \quad \boldsymbol{h}_i \leftarrow \boldsymbol{h}_i \cup \{x \in \boldsymbol{h}_j \mid x - i < t\} \setminus \{i\} \\
\cdot \quad p \leftarrow \begin{cases} \min\{j \in \boldsymbol{h}_i\} & \text{if } \boldsymbol{h}_i \neq \emptyset \\ i & \text{other case} \end{cases} \\
\cdot \quad \boldsymbol{r}_p \leftarrow \boldsymbol{r}_p \cup \{i\}
\end{array}
$$

Esta algoritmo de factorización simbólica es muy eficiente, la complejidad en tiempo y espacio es de orden $\mathrm{O}(\eta(\mathbf{H}_t))$.

El algoritmo del gradiente biconjugado con precondicionador LU incompleto (LUI) es:

$\mathbf{x}_0$, coordenada inicial

$\mathbf{G}_t, \mathbf{H}_t \leftarrow \mathrm{LUI}_t(\mathbf{A})$

$\mathbf{r}_0 \leftarrow \mathbf{b} - \mathbf{A}\,\mathbf{x}_0$, gradiente inicial

$\tilde{\mathbf{r}}_0 \leftarrow \mathbf{r}_0$, pseudo-gradiente inicial

Resolver $\mathbf{G}_t \mathbf{H}_t \mathbf{q}_0 = \mathbf{r}_0$

Resolver $\mathbf{H}_t^{\mathrm{T}} \mathbf{G}_t^{\mathrm{T}} \tilde{\mathbf{q}}_0 = \tilde{\mathbf{r}}_0$

$\mathbf{p}_0 \leftarrow \mathbf{q}_0$, dirección de descenso inicial

$\tilde{\mathbf{p}}_0 \leftarrow \tilde{\mathbf{q}}_0$, pseudo-dirección de descenso

ε, tolerancia

$k \leftarrow 0$

mientras $\|\mathbf{r}_k\| > \varepsilon$

$$\mathbf{w} \leftarrow \mathbf{A}\,\mathbf{p}_k$$

$$\tilde{\mathbf{w}} \leftarrow \mathbf{A}^{\mathrm{T}} \tilde{\mathbf{p}}_k$$

$$\alpha_k \leftarrow \frac{\tilde{\mathbf{r}}_k^{\mathrm{T}} \mathbf{q}_k}{\tilde{\mathbf{p}}_k^{\mathrm{T}} \mathbf{w}}$$

$$\mathbf{x}_{k+1} \leftarrow \mathbf{x}_k + \alpha_k \mathbf{p}_k$$

$$\mathbf{r}_{k+1} \leftarrow \mathbf{r}_k - \alpha_k \mathbf{w}$$

$$\tilde{\mathbf{r}}_{k+1} \leftarrow \tilde{\mathbf{r}}_k - \overline{\alpha_k}\, \tilde{\mathbf{w}}$$

Resolver $\mathbf{G}_t \mathbf{H}_t \mathbf{q}_{k+1} = \mathbf{r}_{k+1}$

Resolver $\mathbf{H}_t^{\mathrm{T}} \mathbf{G}_t^{\mathrm{T}} \tilde{\mathbf{q}}_{k+1} = \tilde{\mathbf{r}}_{k+1}$

$$\beta_k \leftarrow \frac{\tilde{\mathbf{r}}_{k+1}^{\mathrm{T}} \mathbf{q}_{k+1}}{\tilde{\mathbf{r}}_k^{\mathrm{T}} \mathbf{q}_k}$$

$$\mathbf{p}_{k+1} \leftarrow \mathbf{q}_{k+1} + \beta_{k+1} \mathbf{p}_k$$

$$\tilde{\mathbf{p}}_{k+1} \leftarrow \tilde{\mathbf{q}}_{k+1} + \beta_{k+1} \tilde{\mathbf{p}}_k$$

$$k \leftarrow k+1$$

¿Preguntas?

miguelvargas@cimat.mx

Referencias

[Meie94] U. Meier-Yang. Preconditioned conjugate gradient-like methods for nonsymmetric linear systems. University of Illinois. 1992

https://en.wikipedia.org/wiki/Biconjugate_gradient_method