线性方程组的迭代解法

包承龙

丘成桐数学科学中心

本章研究对象及目标

假设$\mathbf{A} \in \mathbb{R}^{n\times n}$, 非奇异，求解线性方程组$\mathbf{A}\mathbf{x} = \mathbf{b}$

- 直接解法必须运行完所有的步骤才能得到准确解。
  （当n很大时，运行速度较慢。）
- 迭代法：从初始解$\mathbf{x}^{(0)}$出发，得到$\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(k)}, \ldots$
  - 多步迭代法：

$$\mathbf{x}^{(k+1)} = \mathbf{F}_k(\mathbf{x}^{(k)}, \cdots, \mathbf{x}^{(k-m)})$$

  - 单步迭代法：

$$\mathbf{x}^{(k+1)} = \mathbf{F}_k(\mathbf{x}^{(k)}) \Rightarrow \mathbf{x}^{(k+1)} = \mathbf{B}_k\mathbf{x}^{(k)} + \mathbf{f}_k \Rightarrow \mathbf{x}^{(k+1)} = \mathbf{B}\mathbf{x}^{(k)} + \mathbf{f}_k$$

  单步定常线性迭代，其中$\mathbf{B}$称为迭代矩阵。

目录

向量极限

设$\|\cdot\|$为向量空间或者矩阵空间中的范数。
向量：$\lim_{k\to\infty} \mathbf{x}^{(k)} = \mathbf{x}$，如果 $\lim_{k\to\infty} \|\mathbf{x}^{(k)} - \mathbf{x}\| = 0$

- 设$\mathbf{x}^{(k)} = (x_1^{(k)}, \ldots, x_n^{(k)})$，$\mathbf{x} = (x_1, \ldots, x_n)$，则

$$\begin{aligned} \lim_{k\to\infty} \mathbf{x}_k = \mathbf{x} \quad & \Leftrightarrow \quad \lim_{k\to\infty} \max_{1\le i\le n} |x_i^{(k)} - x_i| = 0 \\ & \Leftrightarrow \quad \lim_{k\to\infty} |x_i^{(k)} - x_i| = 0,\ \forall i = 1, \ldots, n \\ & \Leftrightarrow \quad \lim_{k\to\infty} x_i^{(k)} = x_i,\ \forall i = 1, \ldots, n. \end{aligned}$$

- Cauchy序列：$\{x_k\}$是一个收敛的序列等价于对于任给的$\epsilon > 0$，存在$N \in \mathbb{Z}$，使得

$$|x_n - x_m| \le \epsilon,\ \forall n, m > N$$

矩阵极限

矩阵： $\lim_{k\to\infty}\mathbf{A}^{(k)}=\mathbf{A}$, 如果 $\lim_{k\to\infty}\|\mathbf{A}^{(k)}-\mathbf{A}\|=0\Leftrightarrow\lim_{k\to\infty}a_{ij}^{(k)}=a_{ij},\ \forall i,j$

定理

以下命题等价：

- $\lim_{k\to\infty}\mathbf{A}^{(k)}=\mathbf{0}$
- $\lim_{k\to\infty}a_{ij}^{(k)}=0,\ \forall i,j=1,\ldots,n$
- $\lim_{k\to\infty}\mathbf{A}^{(k)}\mathbf{x}=\mathbf{0},\ \forall\mathbf{x}\in\mathbb{R}^n$
- $\rho(\mathbf{A})<1$
- 至少存在一种矩阵从属范数$\|\cdot\|$,使$\|\mathbf{A}\|<1$

定理

设$\mathbf{B} \in \mathbb{R}^{n\times n}$, $\|\cdot\|$为任何一种矩阵范数，则

$$\lim_{k\to\infty} \|\mathbf{B}^k\|^{\frac{1}{k}} = \rho(\mathbf{B})$$

由于

$$\rho(\mathbf{B}) = [\rho(\mathbf{B}^k)]^{\frac{1}{k}} \leq \|\mathbf{B}^k\|^{\frac{1}{k}}$$

$\rho(A^k) = \rho(A)^k$

对任意$\epsilon > 0$, 记矩阵

$$\mathbf{B}_\epsilon = [\rho(\mathbf{B}) + \epsilon]^{-1}\mathbf{B} \quad \Rightarrow \quad \rho(\mathbf{B}_\epsilon) < 1 \quad \Rightarrow \lim_{k\to\infty} \mathbf{B}_\epsilon^k = \mathbf{0}$$

存在N, 使得$k > N$有

$$\rho(\mathbf{B}) \leq \|\mathbf{B}^k\|^{\frac{1}{k}} < \rho(\mathbf{B}) + \epsilon$$

迭代公式的构造

$$\mathbf{A}\mathbf{x} = \mathbf{b} \Leftrightarrow \mathbf{x} = \underbrace{\mathbf{M}^{-1}\mathbf{N}}_{\mathbf{B}}\mathbf{x} + \underbrace{\mathbf{M}^{-1}\mathbf{b}}_{\mathbf{f}}, \quad \mathbf{A} = \mathbf{M} - \mathbf{N}$$

且有$\mathbf{B} = \mathbf{I} - \mathbf{M}^{-1}\mathbf{A}$

单步定常迭代算法:

$$\mathbf{x}^{(k+1)} = \mathbf{B}\mathbf{x}^{(k)} + \mathbf{f}, \quad k = 0, 1, \ldots \tag{2}$$

$\mathbf{x}^{(k)} \to \mathbf{x}^*$, $k \to +\infty$, 则称迭代法(2)是收敛的

收敛性分析

定义$\mathbf{e}^{(k)} = \mathbf{x}^{(k)} - \mathbf{x}^*$, 迭代法收敛等价于 $\lim_{k\to\infty} \mathbf{e}^{(k)} = \mathbf{0}$, $\forall \mathbf{e}^{(0)} \in \mathbb{R}^n$

不难发现，$\mathbf{e}^{(k)} = \mathbf{B}(\mathbf{x}^{(k-1)} - \mathbf{x}^*) = \mathbf{B}\mathbf{e}^{(k-1)} = \mathbf{B}^k\mathbf{e}^{(0)}$

定理

迭代法$\mathbf{x}^{(k+1)} = \mathbf{B}\mathbf{x}^{(k)} + \mathbf{f}$收敛的两个充分必要条件分别是

- $\rho(\mathbf{B}) < 1$
- 至少存在一种矩阵从属范数$\|\cdot\|$, 使$\|\mathbf{B}\| < 1$

若 $\|\mathbf{B}\| = q < 1$, 则有

$$\|\mathbf{x}^{(k)} - \mathbf{x}^*\| \leq \frac{q}{1-q}\|x^{(k)} - x^{(k-1)}\|$$

$$\|\mathbf{x}^{(k)} - \mathbf{x}^*\| \leq \frac{q^k}{1-q}\|x^{(1)} - x^{(0)}\|$$

收敛速率

由$\mathbf{e}^{(k)} = \mathbf{B}^k \mathbf{e}^{(0)}$可知

$$\frac{\|\mathbf{e}^{(k)}\|}{\|\mathbf{e}^{(0)}\|} \le \|\mathbf{B}^k\|$$

且最大值可达。若需要满足$\frac{\|\mathbf{e}^{(k)}\|}{\|\mathbf{e}^{(0)}\|} \le \epsilon$, 则可用

$$\|\mathbf{B}^k\| \le \epsilon \quad \Leftrightarrow \quad \|\mathbf{B}^k\|^{\frac{1}{k}} \le \epsilon^{\frac{1}{k}} \quad \Leftrightarrow \quad \ln \|\mathbf{B}^k\|^{\frac{1}{k}} \le \ln \epsilon^{\frac{1}{k}}$$

$$\Leftrightarrow \quad k \ge \frac{-\ln \epsilon}{\underbrace{-\ln \|\mathbf{B}^k\|^{\frac{1}{k}}}_{\text{平均收敛率：}R_k(\mathbf{B})}}$$

得迭代次数近似

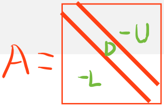
$\to -\ln \rho(\mathbf{B})$

$R_k(\mathbf{B})$依赖于向量从属范数的定义。

渐近收敛率：$R(\mathbf{B}) = -\ln \rho(\mathbf{B})$, 由于$\lim\limits_{k\to\infty} \|\mathbf{B}^k\|^{\frac{1}{k}} = \rho(\mathbf{B})$

目录

Jacobi迭代法(J法)



设$\mathbf{A} = \mathbf{D} - \mathbf{L} - \mathbf{U}$, 其中$\mathbf{D}$为对角部分, $\mathbf{L}$, $\mathbf{U}$分别为严格下、上三角部分
若$\mathbf{D}$非奇异，即$a_{ii} \neq 0$

$$\mathbf{A}\mathbf{x} = \mathbf{b} \quad \Leftrightarrow \quad \mathbf{x} = \mathbf{B}_J\mathbf{x} + \mathbf{f}_J \quad \Rightarrow \quad \mathbf{x}^{(k+1)} = \mathbf{B}_J\mathbf{x}^{(k)} + \mathbf{f}_J$$

其中，$\mathbf{B}_J = \mathbf{D}^{-1}(\mathbf{L}+\mathbf{U}) = \mathbf{I} - \mathbf{D}^{-1}\mathbf{A}$, $\mathbf{f}_J = \mathbf{D}^{-1}\mathbf{b}$

分量形式:

$$x_i^{(k+1)} = \frac{1}{a_{ii}}\left(b_i - \sum_{j=1}^{i-1} a_{ij}x_j^{(k)} - \sum_{j=i+1}^{n} a_{ij}x_j^{(k)}\right), \ i = 1, 2, \ldots, n$$

Gauss-Seidel迭代法(GS法)

$$A = \begin{bmatrix} & -U \\ D-L & \end{bmatrix}$$

$$\mathbf{Ax} = \mathbf{b} \quad \Leftrightarrow \quad \mathbf{x} = \mathbf{B}_G\mathbf{x} + \mathbf{f}_G \quad \Rightarrow \quad \mathbf{x}^{(k+1)} = \mathbf{B}_G\mathbf{x}^{(k)} + \mathbf{f}_G$$

其中，$\mathbf{B}_G = (\mathbf{D}-\mathbf{L})^{-1}\mathbf{U} = \mathbf{I} - (\mathbf{D}-\mathbf{L})^{-1}\mathbf{A}$, $\mathbf{f}_J = (\mathbf{D}-\mathbf{L})^{-1}\mathbf{b}$

分量形式:

$$x_i^{(k+1)} = \frac{1}{a_{ii}}\left(b_i - \sum_{j=1}^{i-1} a_{ij}x_j^{(k+1)} - \sum_{j=i+1}^{n} a_{ij}x_j^{(k)}\right), \ i = 1, 2, \ldots, n$$

J法与GS法收敛性

定理

设$\mathbf{A}$为严格对角占优矩阵，或为不可约的弱对角占优矩阵，则*J*法与*GS*法均收敛。

对$\mathbf{A}$为不可约矩阵进行证明。首先可知$\mathbf{A}$非奇异，且$a_{ii} \neq 0, \forall i$.
若存在$\lambda \in \sigma(\mathbf{B}_G), \mathbf{B}_G = (\mathbf{D}-\mathbf{L})^{-1}\mathbf{U}$满足$|\lambda| \geq 1$, 则有

$$0 = \det(\lambda \mathbf{I} - (\mathbf{D}-\mathbf{L})^{-1}\mathbf{U}) = \det((\mathbf{D}-\mathbf{L})^{-1}) \det(\mathbf{D}-\mathbf{L}-\lambda^{-1}\mathbf{U})$$

由$a_{ii} \neq 0$可知$\det(\mathbf{D}-\mathbf{L}) \neq 0$.
此外，$\mathbf{D}-\mathbf{L}-\mathbf{U}$ 与 $\mathbf{D}-\mathbf{L}-\lambda^{-1}\mathbf{U}$非零元素位置一致，
则$\mathbf{D}-\mathbf{L}-\lambda^{-1}\mathbf{U}$为不可约矩阵，由$|\lambda| \geq 1$可知 $\mathbf{D}-\mathbf{L}-\lambda^{-1}\mathbf{U}$为~~不可约~~
矩阵，从而$\det(\mathbf{D}-\mathbf{L}-\lambda^{-1}\mathbf{U}) \neq 0$,矛盾！ 且弱对角占优

定理

设$\mathbf{A}$对称，且对角元素$a_{ii} > 0$, $i = 1, 2, \ldots, n$, 则方程组J法收敛的充要条件是$\mathbf{A}$和$2\mathbf{D} - \mathbf{A}$均正定，其中$\mathbf{D} = \mathrm{diag}(a_{11}, \ldots, a_{nn})$.

$$\mathbf{B}_J = \mathbf{I} - \mathbf{D}^{-1}\mathbf{A} = \mathbf{D}^{-\frac{1}{2}}(\mathbf{I} - \mathbf{D}^{-\frac{1}{2}}\mathbf{A}\mathbf{D}^{-\frac{1}{2}})\mathbf{D}^{\frac{1}{2}}$$

由$\mathbf{A}$ 对称可知，$\mathbf{D}^{-\frac{1}{2}}\mathbf{A}\mathbf{D}^{-\frac{1}{2}}$, $\mathbf{I} - \mathbf{D}^{-\frac{1}{2}}\mathbf{A}\mathbf{D}^{-\frac{1}{2}}$ 和 $2\mathbf{I} - \mathbf{D}^{-\frac{1}{2}}\mathbf{A}\mathbf{D}^{-\frac{1}{2}}$均对称，且$\mathbf{B}_J$与$\mathbf{I} - \mathbf{D}^{-\frac{1}{2}}\mathbf{A}\mathbf{D}^{-\frac{1}{2}}$相似

必要性。若J法收敛，则$\rho(\mathbf{B}_J) < 1$。对任意$\mu \in \sigma(\mathbf{D}^{-\frac{1}{2}}\mathbf{A}\mathbf{D}^{-\frac{1}{2}})$, 有$1 - \mu \in \sigma(\mathbf{B}_J)$, 从而$|1 - \mu| < 1 \Leftrightarrow \mu \in (0, 2)$。由此，可知$\mathbf{D}^{-\frac{1}{2}}\mathbf{A}\mathbf{D}^{-\frac{1}{2}}$ 正定
$\mathbf{A}$和$2\mathbf{D} - \mathbf{A} = \mathbf{D}^{\frac{1}{2}}(2\mathbf{I} - \mathbf{D}^{-\frac{1}{2}}\mathbf{A}\mathbf{D}^{-\frac{1}{2}})\mathbf{D}^{\frac{1}{2}}$正定。

充分性可以类似证明。

定理

设$\mathbf{A}$对称正定，则方程组*GS*法收敛。

定理

设$\mathbf{A}$对称，非奇异，且对角元素$a_{ii} > 0$，若*GS*法收敛，则$\mathbf{A}$正定。

例：分析方程组$\mathbf{Ax} = \mathbf{b}$的J法和GS法的收敛性，其中

$$\mathbf{A} = \begin{bmatrix} 1 & a & a \\ a & 1 & a \\ a & a & 1 \end{bmatrix}$$

在一般情况下，有可能J法和GS法都收敛或者都不收敛，也可能一者收敛而另一者不收敛。

目录

超松弛迭代公式（SOR）

回忆GS法的分量形式：假设$x_1^{(k+1)},\ldots,x_{i-1}^{(k-1)}$已经算好

$$\bar{x}_i^{(k+1)} = \frac{1}{a_{ii}}\left(b_i - \sum_{j=1}^{i-1} a_{ij}x_j^{(k+1)} - \sum_{j=i+1}^{n} a_{ij}x_j^{(k)}\right)$$

对$\bar{x}_i^{(k+1)}$ 与 $x_i^{(k)}$作加权平均：

$$\begin{aligned} x_i^{(k+1)} &= w\bar{x}_i^{(k+1)} + (1-w)x_i^{(k)} \\ &= (1-w)x_i^{(k)} + \frac{w}{a_{ii}}\left(b_i - \sum_{j=1}^{i-1} a_{ij}x_j^{(k+1)} - \sum_{j=i+1}^{n} a_{ij}x_j^{(k)}\right) \end{aligned}$$

注意与GS法的区别：若$w=1$, SOR⇒GS

设$\mathbf{A}=\mathbf{D}-\mathbf{L}-\mathbf{U}$, 则SOR的迭代格式为

$Ax=b$

$$\mathbf{x}^{(k+1)}=(1-w)\mathbf{x}^{(k)}+w\mathbf{D}^{-1}(\mathbf{b}+\mathbf{L}\mathbf{x}^{(k+1)}+\mathbf{U}\mathbf{x}^{(k)})$$

$$\Rightarrow \mathbf{x}^{(k+1)}=\mathcal{L}_w\mathbf{x}^{(k)}+w(\mathbf{D}-w\mathbf{L})^{-1}\mathbf{b},\ \mathcal{L}_w=(\mathbf{D}-w\mathbf{L})^{-1}[(1-w)\mathbf{D}+w\mathbf{U}]$$

定理 写 $A=D-L-U=\frac{1}{w}(D-wL)-\frac{1}{w}((1-w)D+wU)$

设$\mathbf{A}$非奇异，且所有对角元$a_{ii}\neq 0$，则对所有实数w，有

$$\rho(\mathcal{L}_w)\geq|1-w|.$$

设A有特征值$\lambda_1,\lambda_2,\cdots,\lambda_n$, 则

$$\mathcal{L}_w\ \lambda_1\lambda_2\cdots\lambda=\det(\mathcal{L}_w)=\det((\mathbf{D}-w\mathbf{L})^{-1})\det((1-w)\mathbf{D}+w\mathbf{U})$$

$$=\det(\mathbf{D}^{-1})\det((1-w)\mathbf{D})=(1-w)^n$$

$$\Rightarrow\quad \rho(\mathcal{L}_w)=\max_{1\leq i\leq n}|\lambda_i|\geq|\lambda_1\cdots\lambda_n|^{\frac{1}{n}}=|1-w|$$

若SOR收敛，则$|1-w|<1$,即$w\in(0,2)$

定理

若$\mathbf{A}$对称正定，且$w \in (0,2)$，则*SOR*法收敛。

设$\lambda \in \sigma(\mathcal{L}_w)$ 且$\mathbf{x} \neq \mathbf{0}$ 为对应特征向量，则有

$$\mathcal{L}_w \mathbf{x} = \lambda \mathbf{x} \Rightarrow [(1-w)\mathbf{D} + w\mathbf{U}]\mathbf{x} = \lambda(\mathbf{D} - w\mathbf{L})\mathbf{x}$$

由于$\mathbf{A} = \mathbf{A}^\top$，则有$\mathbf{L}^\top = \mathbf{U}$，上式两边同时对x取内积，

$$(1-w)(\mathbf{Dx},\mathbf{x}) + w(\mathbf{Ux},\mathbf{x}) = \lambda[(\mathbf{Dx},\mathbf{x}) - w(\mathbf{Lx},\mathbf{x})] \tag{3}$$

由于$\mathbf{A}$正定，则$\mathbf{D}$也正定，记$p = (\mathbf{Dx},\mathbf{x}) > 0$与$(\mathbf{Lx},\mathbf{x}) = \alpha + i\beta$，有

$$(\mathbf{Ux},\mathbf{x}) = (\mathbf{L}^\top\mathbf{x},\mathbf{x}) = (\mathbf{x},\mathbf{Lx}) = \overline{(\mathbf{Lx},\mathbf{x})} = \alpha - i\beta$$

由(3)可知，

$$\lambda = \frac{(1-w)p + w\alpha - iw\beta}{p - w\alpha - iw\beta} \Rightarrow |\lambda|^2 = \frac{[p - w(p-\alpha)]^2 + w^2\beta^2}{(p - w\alpha)^2 + w^2\beta^2}$$

分子减去分母，

$$[p - w(p - \alpha)]^2 - (p - w\alpha)^2 = pw(2 - w)(2\alpha - p) \tag{4}$$

由$\mathbf{A}$正定可知，

$$(\mathbf{Ax}, \mathbf{x}) = (\mathbf{Dx}, \mathbf{x}) - (\mathbf{Lx}, \mathbf{x}) - (\mathbf{Ux}, \mathbf{x}) = p - 2\alpha > 0$$

由于$w \in (0, 2)$，可知(4)大于0，则有$|\lambda|^2 < 1$，故$\rho(\mathcal{L}_w) < 1$.

定理

若$\mathbf{A}$对称，非奇异，且对角元素$a_{ii} > 0$，若*SOR*法收敛，则$\mathbf{A}$正定，且$w \in (0, 2)$

$w \in (0,2)$

最优松弛因子

选取w_b，使得

$$\rho(\mathcal{L}_{w_b}) = \min_{w\in(0,2)} \rho(\mathcal{L}_w)$$

定理

设$\mathbf{A}$是对称正定的的三对角或者块三对角矩阵，$\mathbf{B}_J$，$\mathbf{B}_G$和$\mathcal{L}_w$是*J*法，*GS*法和*SOR*法的迭代矩阵，则有

$$\rho(\mathbf{B}_G) = \rho(\mathbf{B}_J)^2 < 1,$$
$$w_b = \frac{2}{1+\sqrt{1-[\rho(\mathbf{B}_J)]^2}}, \quad \rho(\mathcal{L}_b) = w_b - 1$$

$\Rightarrow \quad \rho(\mathcal{L}_b) \le \rho(B_G) \le \rho(B_J) < 1$

对称超松弛迭代（SSOR）

$$
\begin{aligned}
(\mathbf{D}-w\mathbf{L})x^{(k+\frac{1}{2})} &= [(1-w)\mathbf{D}+w\mathbf{U}]\mathbf{x}^{(k)}+w\mathbf{b} \\
(\mathbf{D}-w\mathbf{U})\mathbf{x}^{(k+1)} &= [(1-w)\mathbf{D}+w\mathbf{L}]\mathbf{x}^{(k+\frac{1}{2})}+w\mathbf{b}
\end{aligned} \tag{5}
$$

迭代(5)等价于：

$$
\mathbf{x}^{(k+1)}=\mathbf{B}_w\mathbf{x}^{(k)}+\mathbf{f}_w
$$

其中，

$$
\begin{aligned}
\mathbf{B}_w &= (\mathbf{D}-w\mathbf{U})^{-1}[(1-w)\mathbf{D}+w\mathbf{L}](\mathbf{D}-w\mathbf{L})^{-1}[(1-w)\mathbf{D}+w\mathbf{U}] \\
\mathbf{I}-\mathbf{B}_w &= \left[\frac{1}{w(2-w)}(\mathbf{D}-w\mathbf{L})^{-1}\mathbf{D}^{-1}(\mathbf{D}-w\mathbf{L})\right]^{-1}\mathbf{A}
\end{aligned}
$$

可以类似地推出（SGS）方法。

目录

等价问题

设$\mathbf{A}$为实对称正定矩阵，定义$\varphi(\mathbf{x}) = \frac{1}{2}(\mathbf{A}\mathbf{x}, \mathbf{x}) - (\mathbf{x}, \mathbf{b})$

定理

$\mathbf{x}^*$满足$\mathbf{A}\mathbf{x}^* = \mathbf{b}$当且仅当$\varphi(\mathbf{x}^*) = \min_{\mathbf{x}} \varphi(\mathbf{x})$

必要性。$\mathbf{x}^* = \mathbf{A}^{-1}\mathbf{b}$, 且对任意$\mathbf{x}$有

$$\begin{aligned}\varphi(\mathbf{x}) - \varphi(\mathbf{x}^*) &= \frac{1}{2}\mathbf{x}^\top\mathbf{A}\mathbf{x} - \mathbf{x}^\top\mathbf{b} - \frac{1}{2}\mathbf{x}^{*\top}\mathbf{A}\mathbf{x}^* + \mathbf{x}^{*\top}\mathbf{b} \\ &= \frac{1}{2}(\mathbf{x} - \mathbf{x}^{*\top})\mathbf{A}(\mathbf{x} - \mathbf{x}^{*\top}) \geq 0\end{aligned}$$

充分性。由最优性条件可知，$\nabla\varphi(\mathbf{x}^*) = \mathbf{A}\mathbf{x}^* - \mathbf{b} = \mathbf{0}$.

将方程组问题转化为求$\varphi(\mathbf{x})$的最小值问题。

构造序列$\{\mathbf{x}^{(k)}\}$，逐渐极小化$\varphi(\mathbf{x})$，定义$\mathbf{r}^{(k)} = \mathbf{b} - \mathbf{A}\mathbf{x}^{(k)} = -\nabla\varphi(\mathbf{x}^{(k)})$

一般想法： 找到$\mathbf{p}^{(k)}$和α_k满足：

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \alpha_k \mathbf{p}^{(k)}$$

最速下降法： 设$\mathbf{p}^{(k)} = -\nabla\varphi^{(k)} = \mathbf{r}^{(k)}$，取$\alpha_k$满足

$$\alpha_k = \arg\min_{\alpha>0} \varphi(\mathbf{x}^{(k)} + \alpha\mathbf{p}^{(k)}) \Rightarrow \frac{\mathrm{d}}{\mathrm{d}\alpha}\varphi(\mathbf{x}^{(k)} + \alpha_k\mathbf{p}^{(k)}) = 0$$

$$\Leftrightarrow \alpha_k = \frac{(\mathbf{r}^{(k)}, \mathbf{p}^{(k)})}{(\mathbf{A}\mathbf{p}^{(k)}, \mathbf{p}^{(k)})}$$

性质： $(\mathbf{r}^{(k+1)}, \mathbf{r}^{(k)}) = 0$, $\{\varphi(\mathbf{x}^{(k)})\}$单调下降且下有界，满足$\mathbf{x}^{(k)} \to \mathbf{x}^*$.

假设$\lambda_1 \geq \cdots \geq \lambda_n > 0$, 则有 $\|\mathbf{x}^{(k)} - \mathbf{x}^*\|_{\mathbf{A}} \leq \left(\frac{\lambda_1-\lambda_n}{\lambda_1+\lambda_n}\right)^k \|\mathbf{x}^{(0)} - \mathbf{x}^*\|_{\mathbf{A}}$

共轭梯度法

利用递归公式，

$$\begin{aligned}\mathbf{x}^{(k+1)} &= \mathbf{x}^{(k)} + \alpha_k \mathbf{p}^{(k)} = \mathbf{x}^{(0)} + \alpha_0 \mathbf{p}^{(0)} + \cdots + \alpha_k \mathbf{p}^{(k)} \\ &\in \mathbf{x}^{(0)} + \mathrm{Span}\{\mathbf{p}^{(0)}, \mathbf{p}^{(1)}, \ldots, \mathbf{p}^{(k)}\}\end{aligned}$$

能否找到$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \alpha_k \mathbf{p}^{(k)}$满足

$$\mathbf{x}^{(k+1)} = \arg\min\{\varphi(\mathbf{x}) | \mathbf{x} \in \mathbf{x}^{(0)} + \mathrm{Span}\{\mathbf{p}^{(0)}, \mathbf{p}^{(1)}, \ldots, \mathbf{p}^{(k)}\}\} \tag{6}$$

设$\mathbf{x} = \mathbf{y} + \alpha \mathbf{p}^{(k)}$，其中$\mathbf{y} \in \mathbf{x}^{(0)} + \mathrm{Span}\{\mathbf{p}^{(0)}, \ldots, \mathbf{p}^{(k-1)}\}$，且有

$$\begin{aligned}\varphi(\mathbf{x}) &= \varphi(\mathbf{y}) + \alpha(\mathbf{A}\mathbf{y}, \mathbf{p}^{(k)}) - \alpha(\mathbf{b}, \mathbf{p}^{(k)}) + \frac{\alpha^2}{2}(\mathbf{A}\mathbf{p}^{(k)}, \mathbf{p}^{(k)}) \\ &= \varphi(\mathbf{y}) + \alpha(\mathbf{A}\mathbf{q}, \mathbf{p}^{(k)}) - \alpha(\mathbf{b} - \mathbf{A}\mathbf{x}^{(0)}, \mathbf{p}^{(k)}) + \frac{\alpha^2}{2}(\mathbf{A}\mathbf{p}^{(k)}, \mathbf{p}^{(k)})\end{aligned}$$

其中， $\mathbf{q} \in \mathrm{Span}\{\mathbf{p}^{(0)}, \ldots, \mathbf{p}^{(k-1)}\}$

定义： A-共轭正交向量， $(\mathbf{A}\mathbf{p}^{(i)}, \mathbf{p}^{(j)}) = 0, \forall i \neq j$

若$(\mathbf{p}^{(0)}, \ldots, \mathbf{p}^{(k)})$为A-共轭正交向量，则极小化问题(6)等价于

$$\min_{\mathbf{y}\in\mathbf{x}^{(0)}+\mathrm{Span}\{\mathbf{p}^{(0)},\ldots,\mathbf{p}^{(k-1)}\}} \varphi(\mathbf{y}) + \underbrace{\min_{\alpha}\left\{-\alpha(\mathbf{r}^{(0)}, \mathbf{p}^{(k)}) + \frac{\alpha^2}{2}(\mathbf{A}\mathbf{p}^{(k)}, \mathbf{p}^{(k)})\right\}}_{\alpha_k = \frac{(\mathbf{r}^{(0)},\mathbf{p}^{(k)})}{(\mathbf{A}\mathbf{p}^{(k)},\mathbf{p}^{(k)})}}$$

由于$\mathbf{x}^{(k)} \in \mathbf{x}^{(0)} + \mathrm{Span}\{\mathbf{p}^{(0)}, \ldots, \mathbf{p}^{(k-1)}\}$, 则 $(\mathbf{A}(\mathbf{x}^{(k)} - \mathbf{x}^{(0)}), \mathbf{p}^{(k)}) = 0$

$$(\mathbf{r}^{(0)}, \mathbf{p}^{(k)}) = (\mathbf{r}^{(0)} - (\mathbf{A}(\mathbf{x}^{(k)} - \mathbf{x}^{(0)})), \mathbf{p}^{(k)}) = (\mathbf{r}^{(k)}, \mathbf{p}^{(k)})$$

选取$\mathbf{p}^{(k)}$, 满足

$$\mathbf{p}^{(k)} = \mathbf{r}^{(k)} + \beta_{k-1}\mathbf{p}^{(k-1)}$$

第k步残差

由$(\mathbf{p}^{(k)}, \mathbf{A}\mathbf{p}^{(k-1)}) = 0$可知，

$$\beta_{k-1} = -\frac{(\mathbf{r}^{(k)}, \mathbf{A}\mathbf{p}^{(k-1)})}{(\mathbf{p}^{(k-1)}, \mathbf{A}\mathbf{p}^{(k-1)})}$$

注意： 需要验证$(\mathbf{p}^{(k)}, \mathbf{A}\mathbf{p}^{(j)}) = 0, \forall j \leq k-2$

CG法的计算步骤：

- 初始化：$\mathbf{x}^{(0)} \in \mathbb{R}^n$，$\mathbf{r}^{(0)} = \mathbf{b} - \mathbf{A}\mathbf{x}^{(0)}$，$\mathbf{p}^{(0)} = \mathbf{r}^{(0)}$。
- 迭代：对于$k = 0, \ldots,$

$$
\begin{aligned}
\alpha_k &= \frac{(\mathbf{r}^{(k)}, \mathbf{p}^{(k)})}{(\mathbf{A}\mathbf{p}^{(k)}, \mathbf{p}^{(k)})} \Rightarrow \alpha_k = \frac{(\mathbf{r}^{(k)}, \mathbf{r}^{(k)})}{(\mathbf{p}^{(k)}, \mathbf{A}\mathbf{p}^{(k)})} \\
\mathbf{x}^{(k+1)} &= \mathbf{x}^{(k)} + \alpha_k \mathbf{p}^{(k)} \\
\mathbf{r}^{(k+1)} &= \mathbf{b} - \mathbf{A}\mathbf{x}^{(k+1)} \\
\beta_k &= -\frac{(\mathbf{r}^{(k+1)}, \mathbf{A}\mathbf{p}^{(k)})}{(\mathbf{p}^{(k)}, \mathbf{A}\mathbf{p}^{(k)})} \Rightarrow \beta_k = \frac{(\mathbf{r}^{(k+1)}, \mathbf{r}^{(k+1)})}{(\mathbf{r}^{(k)}, \mathbf{r}^{(k)})} \\
\mathbf{p}^{(k+1)} &= \mathbf{r}^{(k+1)} + \beta_k \mathbf{p}^{(k)}
\end{aligned}
\tag{7}
$$

由$(\mathbf{r}^{(k+1)}, \mathbf{p}^{(k)}) = (\mathbf{r}^{(k)}, \mathbf{p}^{(k)}) - \alpha_k(\mathbf{A}\mathbf{p}^{(k)}, \mathbf{p}^{(k)}) = 0$ 可知

$$(\mathbf{r}^{(k)}, \mathbf{p}^{(k)}) = (\mathbf{r}^{(k)}, \mathbf{r}^{(k)} + \beta_{k-1}\mathbf{p}^{(k-1)}) = (\mathbf{r}^{(k)}, \mathbf{r}^{(k)})$$

从而公式(7)成立

同时，我们有

$$\mathbf{r}^{(k+1)} = \mathbf{b} - \mathbf{A}\mathbf{x}^{(k+1)} = \mathbf{b} - \mathbf{A}(\mathbf{x}^{(k)} + \alpha_k\mathbf{p}^{(k)}) = \mathbf{r}^{(k)} - \alpha_k\mathbf{A}\mathbf{p}^{(k)}$$

$$(\mathbf{r}^{(k+1)}, \mathbf{r}^{(k)}) = (\mathbf{r}^{(k)}, \mathbf{r}^{(k)}) - \alpha_k(\underbrace{\mathbf{r}^{(k)}}_{\mathbf{p}^{(k)} - \beta_{k-1}\mathbf{p}^{(k-1)}}, \mathbf{A}\mathbf{p}^{(k)}) = 0, \qquad (8)$$

且有

$$\beta_k = -\frac{(\mathbf{r}^{(k+1)}, \mathbf{A}\mathbf{p}^{(k)})}{(\mathbf{p}^{(k)}, \mathbf{A}\mathbf{p}^{(k)})} = -\frac{(\mathbf{r}^{(k+1)}, \alpha_k^{-1}(\mathbf{r}^{(k)} - \mathbf{r}^{(k+1)}))}{(\mathbf{p}^{(k)}, \mathbf{A}\mathbf{p}^{(k)})}$$
$$= \frac{(\mathbf{r}^{(k+1)}, \mathbf{r}^{(k+1)})}{\alpha_k(\mathbf{p}^{(k)}, \mathbf{A}\mathbf{p}^{(k)})} = \frac{(\mathbf{r}^{(k+1)}, \mathbf{r}^{(k+1)})}{(\mathbf{r}^{(k)}, \mathbf{r}^{(k)})}$$

由CG法产生的迭代满足以下性质：

- $(\mathbf{r}^{(i)}, \mathbf{r}^{(j)}) = 0, \forall i \neq j$, 即剩余向量构成一个正交向量组
- $(\mathbf{A}\mathbf{p}^{(i)}, \mathbf{p}^{(j)}) = 0, \forall i \neq j$, 即$\{\mathbf{p}^{(k)}\}$ 满足$\mathbf{A}$-共轭。

用归纳法证明。 由α_0与β_0的构造可知：

$$(\mathbf{r}^{(1)}, \mathbf{r}^{(0)}) = (\mathbf{r}^{(0)}, \mathbf{r}^{(0)}) - \alpha_0(\mathbf{A}\mathbf{r}^{(0)}, \mathbf{r}^{(0)}) = 0$$
$$(\mathbf{p}^{(1)}, \mathbf{A}\mathbf{p}^{(0)}) = (\mathbf{r}^{(1)}, \mathbf{A}\mathbf{p}^{(0)}) + \beta_0(\mathbf{p}^{(0)}, \mathbf{A}\mathbf{p}^{(0)}) = 0$$

假设$\mathbf{r}^{(0)}, \cdots, \mathbf{r}^{(k)}$相互正交，$\mathbf{p}^{(0)}, \ldots, \mathbf{p}^{(k)}$相互$\mathbf{A}$-共轭，对$k+1$, 我们有

$$(\mathbf{r}^{(k+1)}, \mathbf{r}^{(j)}) = (\mathbf{r}^{(k)}, \mathbf{r}^{(j)}) - \alpha_k(\mathbf{A}\mathbf{p}^{(k)}, \mathbf{r}^{(j)}) \tag{9}$$

若$j = k$, 由(8)可得。若$j < k$, 则由(9)可知

$$(\mathbf{r}^{(k+1)}, \mathbf{r}^{(j)}) = -\alpha_k(\mathbf{A}\mathbf{p}^{(k)}, \mathbf{p}^{(j)} - \beta_{j-1}\mathbf{p}^{(j-1)}) = 0$$

再看$\mathbf{p}^{(k+1)}$，显然$(\mathbf{p}^{(k+1)}, \mathbf{A}\mathbf{p}^{(k)}) = 0$，对于所有$j = 0, 1, \ldots, k-1$，有

$$(\mathbf{p}^{(k+1)}, \mathbf{A}\mathbf{p}^{(j)}) = (\mathbf{r}^{(k+1)}, \underbrace{\mathbf{A}\mathbf{p}^{(j)}}_{\alpha_j^{-1}(\mathbf{r}^{(j)}-\mathbf{r}^{(j+1)})}) + \beta_k \underbrace{(\mathbf{p}^{(k)}, \mathbf{A}\mathbf{p}^{(j)})}_{=0}$$

备注： 由于$\{\mathbf{r}^{(k)}\}$相互正交，则CG法之多n步收敛

若$\mathbf{A}$对称正定，

- 如果$\mathbf{A} = \mathbf{I} + \mathbf{B}$，且$\mathrm{rank}(\mathbf{B}) = r$，则CG法至多$r+1$步收敛。
- 记$K = \mathrm{cond}(\mathbf{A})_2$，$\|\mathbf{u}\|_{\mathbf{A}}^2 = (\mathbf{A}\mathbf{x}, \mathbf{x})$，则

$$\|\mathbf{x}^{(k)} - \mathbf{x}^*\|_{\mathbf{A}}^2 \le 2\left[\frac{\sqrt{K}-1}{\sqrt{K}+1}\right]^k \|\mathbf{x}^{(0)} - \mathbf{x}^*\|_{\mathbf{A}}$$

条件数越小 收敛速度越大

预处理方法

设$\mathbf{A}$对称正定，$\mathbf{S}$可逆，且$\mathbf{M}=\mathbf{S}\mathbf{S}^\top$

$$\mathbf{A}\mathbf{x}=\mathbf{b}\Rightarrow \underbrace{\mathbf{S}^{-1}\mathbf{A}\mathbf{S}^{-\top}\mathbf{u}=\mathbf{S}^{-1}\mathbf{b}}_{\mathbf{F}\mathbf{u}=\mathbf{g}},\ \mathbf{x}=\mathbf{S}^{-\top}\mathbf{u}$$

其中$\mathbf{F}$具有较好的条件数，CG法具有较快的收敛速度。

预条件S的选取

- 设$\mathbf{S}=\mathbf{L}$, 满足$\mathbf{L}\mathbf{L}^\top\approx\mathbf{A}$
- $\mathbf{S}=(\mathrm{Diag}(a_{11},a_{22},\cdots,a_{nn}))^{\frac{1}{2}}$
- $\mathbf{S}=[w(2-w)]^{-\frac{1}{2}}(\mathbf{D}-w\mathbf{L})\mathbf{D}^{-\frac{1}{2}}$, 有$\mathrm{cond}(\mathbf{F})\approx\sqrt{\mathrm{cond}(\mathbf{A})}$