第三章　线性方程组的迭代解法

数值分析

南京理工大学数学统学院

3.1 迭代法的基本概念

设$\boldsymbol{A} \in \mathbb{R}^{n\times n}$, $\boldsymbol{b} \in \mathbb{R}^{n}$, 考虑线性方程组$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$,

设$\boldsymbol{A}$非奇异, 将$\boldsymbol{A}$分裂为$\boldsymbol{A} = \boldsymbol{M} - \boldsymbol{N}$, 其中$\boldsymbol{M}$为非奇异矩阵, 则有

$$\boldsymbol{x} = \boldsymbol{M}^{-1}\boldsymbol{N}\boldsymbol{x} + \boldsymbol{M}^{-1}\boldsymbol{b}.$$

令$\boldsymbol{B} = \boldsymbol{M}^{-1}\boldsymbol{N} = \boldsymbol{I} - \boldsymbol{M}^{-1}\boldsymbol{A}$, $\boldsymbol{f} = \boldsymbol{M}^{-1}\boldsymbol{b}$,

得到与$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$等价(同解)的方程组

$$\boldsymbol{x} = \boldsymbol{B}\boldsymbol{x} + \boldsymbol{f}. \tag{1.1}$$

由此可以构造方程组的单步定常迭代法:

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{B}\boldsymbol{x}^{(k)} + \boldsymbol{f}, \quad k = 0, 1, \cdots, \tag{1.2}$$

其中B为迭代矩阵, 给定初始向量$\boldsymbol{x}^{(0)} \in \mathbb{R}^n$, 按(1.2) 可产生向量序列$\{\boldsymbol{x}^{(k)}\}$.

其中B为迭代矩阵, 给定初始向量$\boldsymbol{x}^{(0)} \in \mathbb{R}^n$, 按(1.2) 可产生向量序列$\{\boldsymbol{x}^{(k)}\}$.

定义: 若存在向量$\boldsymbol{x}^\star \in \mathbb{R}^n$, 对任意$\boldsymbol{x}^{(0)} \in \mathbb{R}^n$, 由迭代法(1.2)产生的向量序列$\{\boldsymbol{x}^{(k)}\}$ 满足

$$\lim_{k\to\infty} \boldsymbol{x}^{(k)} = \boldsymbol{x}^\star,$$

则称迭代法(1.2)是收敛的.

若迭代法收敛, 则$\boldsymbol{x}^\star$是方程组(1.1)的解, 从而是方程组$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$的解.

记第k次迭代的误差向量为

$$\boldsymbol{e}^{(k)} = \boldsymbol{x}^{(k)} - \boldsymbol{x}^{\star}, \quad k = 0, 1, 2, \cdots, \tag{1.3}$$

于是有

$$\boldsymbol{e}^{(k)} = \boldsymbol{B}\boldsymbol{e}^{(k-1)} = \cdots = \boldsymbol{B}^{k}\boldsymbol{e}^{(0)}, \quad \boldsymbol{e}^{(0)} = \boldsymbol{x}^{(0)} - \boldsymbol{x}^{\star}.$$

记第k次迭代的误差向量为

$$\boldsymbol{e}^{(k)} = \boldsymbol{x}^{(k)} - \boldsymbol{x}^{\star}, \quad k = 0, 1, 2, \cdots, \tag{1.3}$$

于是有

$$\boldsymbol{e}^{(k)} = \boldsymbol{B}\boldsymbol{e}^{(k-1)} = \cdots = \boldsymbol{B}^{k}\boldsymbol{e}^{(0)}, \quad \boldsymbol{e}^{(0)} = \boldsymbol{x}^{(0)} - \boldsymbol{x}^{\star}.$$

定理: 迭代法$\boldsymbol{x}^{(k+1)} = \boldsymbol{B}\boldsymbol{x}^{(k)} + \boldsymbol{f}$, $k = 0, 1, \cdots$, 收敛的充要条件:

(i) $\lim\limits_{k\to\infty} \boldsymbol{B}^{k} = \mathbf{0}$,

(ii) $\rho(\boldsymbol{B}) < 1$,

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

定理(迭代法误差估计)：设 $\boldsymbol{x}^*$ 是方程 $\boldsymbol{x}=\boldsymbol{B}\boldsymbol{x}+\boldsymbol{f}$ 的唯一解，$\|\boldsymbol{B}\|<1$，则由 (1.2) 产生的向量序列 $\boldsymbol{x}^{(k)}$ 满足

$$\|\boldsymbol{x}^{(k)}-\boldsymbol{x}^*\| \leq \frac{\|\boldsymbol{B}\|}{1-\|\boldsymbol{B}\|}\|\boldsymbol{x}^{(k)}-\boldsymbol{x}^{(k-1)}\|; \tag{1.4}$$

$$\|\boldsymbol{x}^{(k)}-\boldsymbol{x}^*\| \leq \frac{\|\boldsymbol{B}\|^k}{1-\|\boldsymbol{B}\|}(\|\boldsymbol{f}\|+2\|\boldsymbol{x}^{(0)}\|). \tag{1.5}$$

注: 利用(1.5) 可估计迭代法需要的迭代次数, 若要 $\|\boldsymbol{x}^{(k)}-\boldsymbol{x}^*\|\leq\varepsilon$, 只要

$$\frac{\|\boldsymbol{B}\|^k}{1-\|\boldsymbol{B}\|}(\|\boldsymbol{f}\|+2\|\boldsymbol{x}^{(0)}\|)\leq\varepsilon,\ \text{即} k \geq \ln\left(\frac{\varepsilon(1-\|\boldsymbol{B}\|)}{\|\boldsymbol{f}\|+2\|\boldsymbol{x}^{(0)}\|}\right)\Big/\ln\|\boldsymbol{B}\|.$$

迭代法收敛速度:

设迭代法(1.2)收敛, 即$\rho(\boldsymbol{B}) < 1$, 第k次迭代的误差向量$\boldsymbol{e}^{(k)}$满足($\boldsymbol{e}^{(0)} \neq \boldsymbol{0}$)

$$\frac{||\boldsymbol{e}^{(k)}||}{||\boldsymbol{e}^{(0)}||} \leq ||\boldsymbol{B}^k||,$$

根据矩阵算子范数定义

$$||\boldsymbol{B}^k|| = \max_{\boldsymbol{e}^{(0)} \neq \boldsymbol{0}} \frac{||\boldsymbol{B}^k \boldsymbol{e}^{(0)}||}{||\boldsymbol{e}^{(0)}||} = \max_{\boldsymbol{e}^{(0)} \neq \boldsymbol{0}} \frac{||\boldsymbol{e}^{(k)}||}{||\boldsymbol{e}^{(0)}||}.$$

因此, 迭代k次后, 平均每次迭代误差向量范数的压缩率可以看成$||\boldsymbol{B}^k||^{\frac{1}{k}}$.

若要求迭代k次后有$\dfrac{||\boldsymbol{e}^{(k)}||}{||\boldsymbol{e}^{(0)}||} \le \epsilon,\quad \epsilon \le 10^{-s} \ll 1,$

只要满足$||\boldsymbol{B}^k|| \le \epsilon$即可，即$||\boldsymbol{B}^k||^{\frac{1}{k}} \le \epsilon^{\frac{1}{k}},$

取对数得$\ln ||\boldsymbol{B}^k||^{\frac{1}{k}} \le \dfrac{1}{k}\ln \epsilon$, 即 $k \ge \dfrac{-\ln \epsilon}{-\ln ||\boldsymbol{B}^k||^{\frac{1}{k}}}.$

定义(平均收敛率): $R_k(\boldsymbol{B}) = -\ln ||\boldsymbol{B}^k||^{\frac{1}{k}}$为迭代法(1.2)的平均收敛率.

定义(渐近收敛率): $R(\boldsymbol{B}) = -\ln \rho(\boldsymbol{B})$称为迭代法(1.2)的渐近收敛率或渐近收敛速度.

可用$k \ge \dfrac{-\ln \epsilon}{R(\boldsymbol{B})}$作为所需迭代次数的估计.

3.2 Jacobi 迭代法和 Gauss-Seidel 迭代法

考虑非奇异线性方程组$\boldsymbol{Ax}=\boldsymbol{b}$, $a_{ii}\neq 0$, 将 $\boldsymbol{A}$ 分裂为

$$\boldsymbol{A}=\boldsymbol{D}-\boldsymbol{L}-\boldsymbol{U},$$

其中 $\boldsymbol{D}=\mathrm{diag}(a_{11},a_{22},\cdots,a_{nn})$ 可逆,

$$\boldsymbol{L}=-\begin{pmatrix} 0 & & & & \\ a_{21} & 0 & & & \\ a_{31} & a_{32} & 0 & & \\ \vdots & \vdots & \ddots & \ddots & \\ a_{n1} & a_{n2} & \cdots & a_{n,n-1} & 0 \end{pmatrix},\ \boldsymbol{U}=-\begin{pmatrix} 0 & a_{12} & a_{13} & \cdots & a_{1n} \\ & 0 & a_{23} & \cdots & a_{2n} \\ & & \ddots & \ddots & \vdots \\ & & & 0 & a_{n-1,n} \\ & & & & 0 \end{pmatrix}$$

在迭代矩阵$\boldsymbol{B} = \boldsymbol{I} - \boldsymbol{M}^{-1}\boldsymbol{A}$中取$\boldsymbol{M} = \boldsymbol{D}$, 得迭代格式:

$$
\begin{aligned}
\boldsymbol{x}^{(k+1)} &= \left(\boldsymbol{I} - \boldsymbol{D}^{-1}\boldsymbol{A}\right)\boldsymbol{x}^{(k)} + \boldsymbol{D}^{-1}\boldsymbol{b}, \\
&= \boldsymbol{D}^{-1}\left(\boldsymbol{L} + \boldsymbol{U}\right)\boldsymbol{x}^{(k)} + \boldsymbol{D}^{-1}\boldsymbol{b}.
\end{aligned}
$$

$$
\boldsymbol{x}^{(k+1)} = \boldsymbol{B}_J\boldsymbol{x}^{(k)} + \boldsymbol{f}_J, \quad \boldsymbol{B}_J = \boldsymbol{D}^{-1}(\boldsymbol{L} + \boldsymbol{U}), \ \boldsymbol{f}_J = \boldsymbol{D}^{-1}\boldsymbol{b}. \tag{2.1}
$$

称为Jacobi迭代法, 简称J法.

在迭代矩阵$\boldsymbol{B}=\boldsymbol{I}-\boldsymbol{M}^{-1}\boldsymbol{A}$中取$\boldsymbol{M}=\boldsymbol{D}$, 得迭代格式:

$$\begin{aligned}\boldsymbol{x}^{(k+1)} &= \left(\boldsymbol{I}-\boldsymbol{D}^{-1}\boldsymbol{A}\right)\boldsymbol{x}^{(k)}+\boldsymbol{D}^{-1}\boldsymbol{b},\\ &= \boldsymbol{D}^{-1}\left(\boldsymbol{L}+\boldsymbol{U}\right)\boldsymbol{x}^{(k)}+\boldsymbol{D}^{-1}\boldsymbol{b}.\end{aligned}$$

$$\boldsymbol{x}^{(k+1)}=\boldsymbol{B}_J\boldsymbol{x}^{(k)}+\boldsymbol{f}_J,\quad \boldsymbol{B}_J=\boldsymbol{D}^{-1}(\boldsymbol{L}+\boldsymbol{U}),\ \boldsymbol{f}_J=\boldsymbol{D}^{-1}\boldsymbol{b}. \tag{2.1}$$

称为Jacobi迭代法, 简称J法.

Jacobi迭代的分量形式:

$$x_i^{(k+1)}=\frac{1}{a_{ii}}\left(b_i-\sum_{j=1}^{i-1}a_{ij}x_j^{(k)}-\sum_{j=i+1}^{n}a_{ij}x_j^{(k)}\right), \tag{2.2}$$

将(2.2)中右端$x_j^{(k)}, j = 1, \cdots, i-1$ 用$x_j^{(k+1)}$替换，可构造迭代格式:

$$x_i^{(k+1)} = \frac{1}{a_{ii}} \left(b_i - \sum_{j=1}^{i-1} a_{ij} x_j^{(k+1)} - \sum_{j=i+1}^{n} a_{ij} x_j^{(k)} \right), \tag{2.3}$$

将(2.2)中右端$x_j^{(k)}, j = 1, \cdots, i-1$ 用$x_j^{(k+1)}$替换，可构造迭代格式:

$$x_i^{(k+1)} = \frac{1}{a_{ii}}\left(b_i - \sum_{j=1}^{i-1} a_{ij}x_j^{(k+1)} - \sum_{j=i+1}^{n} a_{ij}x_j^{(k)}\right), \tag{2.3}$$

其矩阵形式:

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

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{B}_G\boldsymbol{x}^{(k)} + \boldsymbol{f}_G, \quad \boldsymbol{B}_G = (\boldsymbol{D} - \boldsymbol{L})^{-1}\boldsymbol{U}, \ \boldsymbol{f}_G = (\boldsymbol{D} - \boldsymbol{L})^{-1}\boldsymbol{b}. \tag{2.4}$$

称为Gauss-Seidel迭代法, 简称GS法.

例: 设初始向量 $\boldsymbol{x}^{(0)} = (0,0,0)^T$, 用 Jacobi 迭代法求解方程组

$$\begin{cases} x_1 + 2x_2 - 2x_3 &= 1 \\ x_1 + x_2 + x_3 &= 3 \\ 2x_1 + 2x_2 + x_3 &= 5 \end{cases}$$

解: Jacobi 迭代格式为

$$\begin{cases} x_1^{(k+1)} = 1 - 2x_2^{(k)} + 2x_3^{(k)}, \\ x_2^{(k+1)} = 3 - x_1^{(k)} - x_3^{(k)}, \\ x_3^{(k+1)} = 5 - 2x_1^{(k)} - 2x_2^{(k)}, \end{cases}$$

可得$\boldsymbol{x}^{(1)} = (1,3,5)^T$, $\boldsymbol{x}^{(2)} = (5,-3,-3)^T$, $\boldsymbol{x}^{(3)} = (1,1,1)^T$, $\boldsymbol{x}^{(4)} = (1,1,1)^T$, 所以方程组的解为 $\boldsymbol{x} = (1,1,1)^T$.

例: 设初始向量 $\boldsymbol{x}^{(0)}=(0,0,0)^T$, 用 Gauss-Seidel 迭代法求解方程组

$$\begin{cases} 9x_1-x_2-x_3 &= 7\\ -x_1+8x_2 &= 7\\ -x_1+9x_3 &= 8\end{cases}$$

解: Gauss-Seidel 迭代格式为

$$\begin{cases} x_1^{(k+1)} = \frac{1}{9}(7+x_2^{(k)}+x_3^{(k)})\\ x_2^{(k+1)} = \frac{1}{8}(7+x_1^{(k+1)})\\ x_3^{(k+1)} = \frac{1}{9}(8+x_1^{(k+1)})\end{cases}$$

可得，$\boldsymbol{x}^{(1)}=(0.7778,0.9722,0.9753)^T$,$\boldsymbol{x}^{(2)}=(0.9942,0.9993,0.9994)^T$, $\boldsymbol{x}^{(3)}=(0.9999,0.9999,0.9999)^T$, $\boldsymbol{x}^{(4)}=(1,1,1)^T$,$\boldsymbol{x}^{(5)}=(1,1,1)^T$。

所以方程组的解为 $\boldsymbol{x}=(1,1,1)^T$.

无论是 J 迭代法还是 G-S 迭代法, 它们收敛的充要条件都是迭代矩阵谱半径小于 1.

例: 分析求解 $\boldsymbol{A}\boldsymbol{x}=\boldsymbol{b}$ 的 J 迭代法和 G-S 迭代法收敛性, 其中

$$\boldsymbol{A}=\begin{pmatrix} 1 & -2 & 2 \\ -1 & 1 & -1 \\ -2 & -2 & 1 \end{pmatrix}.$$

解: 由 $\boldsymbol{B}_J = \boldsymbol{D}^{-1}(\boldsymbol{L}+\boldsymbol{U})$，所以 $\boldsymbol{B}_J$ 的特征方程为

$$\begin{aligned}\det(\lambda \boldsymbol{I} - \boldsymbol{B}_J) &= \det[\lambda \boldsymbol{I} - \boldsymbol{D}^{-1}(\boldsymbol{L}+\boldsymbol{U})] \\ &= \det(\boldsymbol{D}^{-1}) \det[\lambda \boldsymbol{D} - (\boldsymbol{L}+\boldsymbol{U})] = 0,\end{aligned}$$

等价于 $\det[\lambda \boldsymbol{D} - (\boldsymbol{L}+\boldsymbol{U})] = 0$，即

$$\begin{vmatrix} \lambda & -2 & 2 \\ -1 & \lambda & -1 \\ -2 & -2 & \lambda \end{vmatrix} = 0.$$

求解得到 $\lambda_1 = \lambda_2 = \lambda_3 = 0$，因此 $\rho(\boldsymbol{B}_J) = 0 < 1$，因此 J 迭代法收敛。

由 $\boldsymbol{B}_G = (\boldsymbol{D} - \boldsymbol{L})^{-1}\boldsymbol{U}$, 所以 $\boldsymbol{B}_G$ 的特征方程为

$$\begin{aligned}\det(\lambda \boldsymbol{I} - \boldsymbol{B}_G) &= \det[\lambda \boldsymbol{I} - (\boldsymbol{D} - \boldsymbol{L})^{-1}\boldsymbol{U}] \\ &= \det[(\boldsymbol{D} - \boldsymbol{L})^{-1}] \det[\lambda(\boldsymbol{D} - \boldsymbol{L}) - \boldsymbol{U}] = 0,\end{aligned}$$

等价于 $\det[\lambda(\boldsymbol{D} - \boldsymbol{L}) - \boldsymbol{U}] = 0$, 即

$$\begin{vmatrix} \lambda & -2 & 2 \\ -\lambda & \lambda & -1 \\ -2\lambda & -2\lambda & \lambda \end{vmatrix} = 0.$$

解得特征值为 $\lambda_1 = 0$, $\lambda_2 = 2(\sqrt{2} - 1)$, $\lambda_3 = 2(\sqrt{2} + 1)$, 因此谱半径 $\rho(\boldsymbol{B}_G) = 2(\sqrt{2} + 1) > 1$, 因此 G-S 迭代法发散.

需要注意的是, J 迭代法和 G-S 迭代法之间并无包含关系.

如设方程组的系数矩阵分别为

$$\boldsymbol{A}_1=\begin{pmatrix}1 & 2 & -2\\ 1 & 1 & 1\\ 2 & 2 & 1\end{pmatrix},\quad \boldsymbol{A}_2=\begin{pmatrix}2 & -1 & 1\\ 1 & 1 & 1\\ 1 & 1 & -2\end{pmatrix},$$

可验证: $\boldsymbol{A}_1$ 对 J 迭代法是收敛的, 但 G-S 迭代法不收敛; 而$\boldsymbol{A}_2$ 对 J 迭代法不收敛, 但 G-S 迭代法收敛.

定义: 设 $\boldsymbol{A}=(a_{ij})\in R^{n\times n}$, 若**不存在**置换矩阵 $\boldsymbol{P}$, 使得

$$\boldsymbol{P}^T\boldsymbol{A}\boldsymbol{P}=\begin{pmatrix} \widetilde{\boldsymbol{A}}_{11} & \widetilde{\boldsymbol{A}}_{12} \\ 0 & \widetilde{\boldsymbol{A}}_{22} \end{pmatrix} \tag{2.5}$$

其中 $\widetilde{\boldsymbol{A}}_{11}$ 与 $\widetilde{\boldsymbol{A}}_{22}$ 均为方阵, 则称 $\boldsymbol{A}$ 为**不可约矩阵**.

定理: 若 $\boldsymbol{A}=(a_{ij})$ 严格对角占优或不可约弱对角占优, 则对 $i=1,2,\cdots,n$, 有 $a_{ii}\neq 0$, 且 $\boldsymbol{A}$ 非奇异.

定理: 若系数矩阵 $\boldsymbol{A}$ 严格对角占优或不可约弱对角占优矩阵, 则 Jacobi 迭代法和 G-S 迭代法收敛.

证明: 若 $\boldsymbol{A}$ 严格对角占优，则 $|a_{ii}| > 0$, 因此 $\boldsymbol{D}$ 可逆. 现假设 Jacobi 迭代矩阵 $\boldsymbol{B}_J$ 的某个特征值$|\lambda| \geq 1$, 那么矩阵 $\lambda\boldsymbol{D} - \boldsymbol{L} - \boldsymbol{U}$ 也是严格对角占优的, 因此 $\lambda\boldsymbol{D} - \boldsymbol{L} - \boldsymbol{U}$ 非奇异.

再由

$$\lambda\boldsymbol{I} - \boldsymbol{B}_J = \lambda\boldsymbol{I} - \boldsymbol{D}^{-1}(\boldsymbol{L} + \boldsymbol{U}) = \boldsymbol{D}^{-1}(\lambda\boldsymbol{D} - \boldsymbol{L} - \boldsymbol{U})$$

可知 $\lambda\boldsymbol{I} - \boldsymbol{B}_J$ 非奇异, 这与 λ 是 $\boldsymbol{B}_J$ 的特征值矛盾.

故 $\boldsymbol{B}_J$ 的特征值的模均小于 1, 即 Jacobi 迭代法收敛.

定理: 设 $\boldsymbol{A}$ 具有正对角元且为对称矩阵, 则 Jacobi 方法收敛的充分必要条件是 $\boldsymbol{A}$ 和 $2\boldsymbol{D}-\boldsymbol{A}$ 同为正定矩阵.

定理: 若系数矩阵 $\boldsymbol{A}$ 是对称正定阵, 则 G-S 迭代法收敛.

定理: 设系数矩阵 $\boldsymbol{A}$ 对称、非奇异，且对角元素$a_{ii}>0$, 若方程组$\boldsymbol{A}x=\boldsymbol{b}$的 G-S 迭代法收敛，则$\boldsymbol{A}$正定.

定理: 设 $\boldsymbol{A}$ 具有正对角元且为对称矩阵, 则 Jacobi 方法收敛的充分必要条件是 $\boldsymbol{A}$ 和 $2\boldsymbol{D}-\boldsymbol{A}$ 同为正定矩阵.

定理: 若系数矩阵 $\boldsymbol{A}$ 是对称正定阵, 则 G-S 迭代法收敛.

定理: 设系数矩阵 $\boldsymbol{A}$ 对称、非奇异，且对角元素$a_{ii}>0$, 若方程组$\boldsymbol{A}x=\boldsymbol{b}$的 G-S 迭代法收敛，则$\boldsymbol{A}$正定.

例: 分析方程组 $\boldsymbol{A}\boldsymbol{x}=\boldsymbol{b}$ 的 J 迭代法和 G-S 迭代法收敛性, 其中

$$\boldsymbol{A}=\begin{pmatrix} 1 & a & a \\ a & 1 & a \\ a & a & 1 \end{pmatrix}.$$

3.3 超松驰迭代(SOR)法

在 G-S 迭代格式中

$$\widetilde{\boldsymbol{x}}^{(k+1)} = \boldsymbol{D}^{-1}\boldsymbol{L}\boldsymbol{x}^{(k+1)} + \boldsymbol{D}^{-1}\boldsymbol{U}\boldsymbol{x}^{(k)} + \boldsymbol{D}^{-1}\boldsymbol{b},$$

取

$$\boldsymbol{x}^{(k+1)} = (1-\omega)\boldsymbol{x}^{(k)} + \omega\widetilde{\boldsymbol{x}}^{(k+1)},$$

便得到逐次超松驰迭代法，简记为 SOR 法，其计算式

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} + \omega\boldsymbol{D}^{-1}[\boldsymbol{L}\boldsymbol{x}^{(k+1)} + (\boldsymbol{U}-\boldsymbol{D})\boldsymbol{x}^{(k)} + \boldsymbol{b}],$$

SOR法也是线性迭代（ω称为松弛因子）

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{B}_\omega\boldsymbol{x}^{(k)} + \boldsymbol{f}_\omega,$$

$$\boldsymbol{B}_\omega = (\boldsymbol{D}-\omega\boldsymbol{L})^{-1}[(1-\omega)\boldsymbol{D}+\omega\boldsymbol{U}], \quad \boldsymbol{f}_\omega = \omega(\boldsymbol{D}-\omega\boldsymbol{L})^{-1}\boldsymbol{b}.$$

SOR法的分量形式

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

定理: 在矩阵 $\boldsymbol{A}$ 是实对称正定矩阵的条件下, SOR 法收敛的充要条件是 $0 < \omega < 2$.

定理: 若矩阵 $\boldsymbol{A}$ 为严格对角占优或不可约弱对角占优矩阵，则当 $0 < \omega \leq 1$ 时, SOR 法收敛.

例: 用SOR 法求解方程组

$$\begin{cases} 4x_1 - x_2 &= 1 \\ -x_1 + 4x_2 - x_3 &= 4 \\ -x_2 + 4x_3 &= -3 \end{cases}$$

分别取 $\omega = 1.03$，$\omega = 1$，$\omega = 1.1$，并对每个 ω 确定迭代次数, 要求当 $\|\boldsymbol{x}^* - \boldsymbol{x}^{(k)}\|_\infty \leq 5 \times 10^{-6}$ 时迭代终止, 其中精确解 $\boldsymbol{x}^* = (\frac{1}{2}, 1, -\frac{1}{2})^T$.

解: SOR 迭代格式为

$$\begin{cases} x_1^{(k+1)} = (1-\omega)x_1^{(k)} + \frac{\omega}{4}(1 + x_2^{(k)}) \\ x_2^{(k+1)} = (1-\omega)x_2^{(k)} + \frac{\omega}{4}(4 + x_1^{(k+1)} + x_3^{(k)}) \\ x_3^{(k+1)} = (1-\omega)x_3^{(k)} + \frac{\omega}{4}(-3 + x_2^{(k+1)}) \end{cases}$$

取初始向量 $\boldsymbol{x}^{(0)} = (0,0,0)^T$，当$\omega = 1.03$ 时，迭代 5 次达到要求，

$$\boldsymbol{x}^{(5)} = (0.50000043, 1.0000001, -0.499999)^T,$$

当 $\omega = 1$ 时，迭代 6 次达到要求，

$$\boldsymbol{x}^{(6)} = (0.50000038, 1.0000002, -0.4999995)^T,$$

当 $\omega = 1.1$ 时，迭代 6 次达到要求，

$$\boldsymbol{x}^{(6)} = (0.50000035, 0.9999989, -0.5000003)^T.$$

3.4 求解线性方程组的极小化方法

3.4.1 变分原理与最速下降法

考虑方程组$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$,

其中$\boldsymbol{A}$ 为n阶对称正定阵, $\boldsymbol{b} \in R^n$ 为一给定向量.

定义二次函数: $\varphi(\boldsymbol{x}):\ R^n \to R$ 如下

$$\begin{aligned}\varphi(\boldsymbol{x}) &= \frac{1}{2}(\boldsymbol{A}\boldsymbol{x}, \boldsymbol{x}) - (\boldsymbol{b}, \boldsymbol{x}) \\ &= \frac{1}{2}\sum_{i=1}^{n}\sum_{i=1}^{n} a_{ij}x_i x_j - \sum_{j=1}^{n} b_j x_j.\end{aligned}$$

1. $\varphi(\boldsymbol{x})$ 的性质

(1). $\forall \boldsymbol{x} \in R^n$, $\nabla\varphi(\boldsymbol{x}) = \text{grad}\varphi(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}$.

(2). $\forall \boldsymbol{x}, \boldsymbol{y} \in R^n$, $\lambda \in R$, 有:

$$\begin{aligned}
\varphi(\boldsymbol{x}+\lambda\boldsymbol{y}) &= \frac{1}{2}\left(\boldsymbol{A}(\boldsymbol{x}+\lambda\boldsymbol{y}), \boldsymbol{x}+\lambda\boldsymbol{y}\right) - \left(\boldsymbol{b}, \boldsymbol{x}+\lambda\boldsymbol{y}\right) \\
&= \frac{1}{2}(\boldsymbol{A}\boldsymbol{x}, \boldsymbol{x}) + \lambda(\boldsymbol{A}\boldsymbol{x}, \boldsymbol{y}) + \frac{\lambda^2}{2}(\boldsymbol{A}\boldsymbol{y}, \boldsymbol{y}) - (\boldsymbol{b}, \boldsymbol{x}) - \lambda(\boldsymbol{b}, \boldsymbol{y}) \\
&= \frac{\lambda^2}{2}(\boldsymbol{A}\boldsymbol{y}, \boldsymbol{y}) + \lambda(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}, \boldsymbol{y}) + \frac{1}{2}(\boldsymbol{A}\boldsymbol{x}, \boldsymbol{x}) - (\boldsymbol{b}, \boldsymbol{x}) \\
&= \frac{\lambda^2}{2}(\boldsymbol{A}\boldsymbol{y}, \boldsymbol{y}) + \lambda(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}, \boldsymbol{y}) + \varphi(\boldsymbol{x}).
\end{aligned}$$

(3). 设$\boldsymbol{A}\boldsymbol{x}^* = \boldsymbol{b}$, 则$\varphi(\boldsymbol{x}^*) = \frac{1}{2}(\boldsymbol{A}\boldsymbol{x}^*, \boldsymbol{x}^*) - (\boldsymbol{b}, \boldsymbol{x}^*) = -\frac{1}{2}(\boldsymbol{A}\boldsymbol{x}^*, \boldsymbol{x}^*)$,

因此:

$$\begin{aligned}
\varphi(\boldsymbol{x}) - \varphi(\boldsymbol{x}^*) &= \frac{1}{2}(\boldsymbol{A}\boldsymbol{x}, \boldsymbol{x}) - (\boldsymbol{b}, \boldsymbol{x}) + \frac{1}{2}(\boldsymbol{A}\boldsymbol{x}^*, \boldsymbol{x}^*) \\
&= \frac{1}{2}(\boldsymbol{A}\boldsymbol{x}, \boldsymbol{x}) - (\boldsymbol{A}\boldsymbol{x}^*, \boldsymbol{x}) + \frac{1}{2}(\boldsymbol{A}\boldsymbol{x}^*, \boldsymbol{x}^*) \\
&= \frac{1}{2}\left(\boldsymbol{A}(\boldsymbol{x} - \boldsymbol{x}^*), \boldsymbol{x} - \boldsymbol{x}^*\right).
\end{aligned}$$

2. 求解线性方程组的变分原理

定理: 设$\boldsymbol{A} \in R^n$, $\boldsymbol{A}$对称正定, 则:

$$\boldsymbol{A}\boldsymbol{x}^* = \boldsymbol{b} \ \Leftrightarrow \ \varphi(\boldsymbol{x}^*) = \min_{\boldsymbol{x} \in R^n} \varphi(\boldsymbol{x}).$$

2. 求解线性方程组的变分原理

定理: 设$\boldsymbol{A} \in R^n$, $\boldsymbol{A}$对称正定, 则:

$$\boldsymbol{A}\boldsymbol{x}^* = \boldsymbol{b} \Leftrightarrow \varphi(\boldsymbol{x}^*) = \min_{\boldsymbol{x}\in R^n} \varphi(\boldsymbol{x}).$$

证明: $\forall \lambda \in R$, $\boldsymbol{x}, \boldsymbol{y} \in R^n$, 有:

$$\varphi(\boldsymbol{x}+\lambda\boldsymbol{y}) = \frac{\lambda^2}{2}(\boldsymbol{A}\boldsymbol{y}, \boldsymbol{y}) + \lambda(\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b}, \boldsymbol{y}) + \varphi(\boldsymbol{x})$$

$(\Rightarrow)$ 若$\boldsymbol{A}\boldsymbol{x}^* = \boldsymbol{b}$, 则

$$\begin{aligned}\varphi(\boldsymbol{x}) &= \varphi(\boldsymbol{x}^* + (\boldsymbol{x}-\boldsymbol{x}^*)) \\ &= \frac{1}{2}(\boldsymbol{A}\left(\boldsymbol{x}-\boldsymbol{x}^*\right), (\boldsymbol{x}-\boldsymbol{x}^*)) + \varphi(\boldsymbol{x}^*),\end{aligned}$$

由A的正定性和$\boldsymbol{x}$的任意性知, $\varphi(\boldsymbol{x}^*) = \min\limits_{\boldsymbol{x}\in R^n} \varphi(\boldsymbol{x})$.

$(\Leftarrow)$ 若$\varphi(\boldsymbol{x}^*) = \min\limits_{\boldsymbol{x}\in R^n} \varphi(\boldsymbol{x})$, 则对任意$\boldsymbol{y} \in R^n$, $\left.\dfrac{\mathrm{d}\varphi(\boldsymbol{x}^* + \lambda \boldsymbol{y})}{\mathrm{d}\lambda}\right|_{\lambda=0} = 0$,

即$(\boldsymbol{A}\boldsymbol{x}^* - \boldsymbol{b}, \boldsymbol{y}) = 0$, 由$\boldsymbol{y}$的任意性, 知$\boldsymbol{A}\boldsymbol{x}^* = \boldsymbol{b}$.

3. 求解$\varphi(\boldsymbol{x})$的极小值$\varphi(\boldsymbol{x}^*)$

"盲人下山":

首先，任意给定一个初始向量$\boldsymbol{x}^{(0)}$，确定一个下山方向$\boldsymbol{p}^{(0)}$，沿着经过点$\boldsymbol{x}^{(0)}$而方向为$\boldsymbol{p}^{(0)}$的直线$\boldsymbol{x}=\boldsymbol{x}^{(0)}+\alpha\boldsymbol{p}^{(0)}$，找一个点$\boldsymbol{x}^{(1)}=\boldsymbol{x}^{(0)}+\alpha_0\boldsymbol{p}^{(0)}$，使得:

$$\varphi(\boldsymbol{x}^{(1)})=\min_{\alpha\in R}\varphi(\boldsymbol{x}^{(0)}+\alpha\boldsymbol{p}^{(0)}).$$

然后，从$\boldsymbol{x}^{(1)}$出发，再确定一个下山方向$\boldsymbol{p}^{(1)}$，沿直线$\boldsymbol{x}=\boldsymbol{x}^{(1)}+\alpha\boldsymbol{p}^{(1)}$找一个点$\boldsymbol{x}^{(2)}=\boldsymbol{x}^{(1)}+\alpha_1\boldsymbol{p}^{(1)}$，使得:

$$\varphi(\boldsymbol{x}^{(2)})=\min_{\alpha\in R}\varphi(\boldsymbol{x}^{(1)}+\alpha\boldsymbol{p}^{(1)}).$$

如此继续下去, 得到一串:

$$\alpha_0, \alpha_1, \cdots, \alpha_k, \cdots \text{ 和 } \boldsymbol{p}^{(0)}, \boldsymbol{p}^{(1)}, \cdots, \boldsymbol{p}^{(k)}, \cdots,$$

称$\boldsymbol{p}^{(k)}$ 为搜索方向, α_k 为步长.

一般做法: 在$\boldsymbol{x}^{(k)}$点, 先找下山方向$\boldsymbol{p}^{(k)}$, 再确定步长α_k, 使得:

$$\varphi(\boldsymbol{x}^{(k)} + \alpha_k \boldsymbol{p}^{(k)}) = \min_{\alpha \in R} \varphi(\boldsymbol{x}^{(k)} + \alpha \boldsymbol{p}^{(k)}).$$

对于不同的选择下山(搜索)方向的方法, 就形成不同的算法.

(1). 如何确定步长α_k

设从$\boldsymbol{x}^{(k)}$ 出发，已选定下山方向$\boldsymbol{p}^{(k)}$，要在直线$\boldsymbol{x} = \boldsymbol{x}^{(k)} + \alpha \boldsymbol{p}^{(k)}$ 上确定最小值点.

(1). 如何确定步长α_k

设从$\boldsymbol{x}^{(k)}$ 出发，已选定下山方向$\boldsymbol{p}^{(k)}$，要在直线$\boldsymbol{x} = \boldsymbol{x}^{(k)} + \alpha \boldsymbol{p}^{(k)}$ 上确定最小值点.

令: $f(\alpha) = \varphi(\boldsymbol{x}^{(k)} + \alpha \boldsymbol{p}^{(k)})$，于是:

$$
\begin{aligned}
f(\alpha) &= \frac{1}{2}\left(\boldsymbol{A}(\boldsymbol{x}^{(k)} + \alpha \boldsymbol{p}^{(k)}), (\boldsymbol{x}^{(k)} + \alpha \boldsymbol{p}^{(k)})\right) - (\boldsymbol{b}, \boldsymbol{x}^{(k)} + \alpha \boldsymbol{p}^{(k)}) \\
&= \frac{1}{2}\alpha^2 \boldsymbol{p}^{(k)T} \boldsymbol{A} \boldsymbol{p}^{(k)} - \alpha(\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}^{(k)})^T \boldsymbol{p}^{(k)} + \varphi(\boldsymbol{x}^{(k)}) \\
&= \frac{1}{2}\alpha^2 \boldsymbol{p}^{(k)T} \boldsymbol{A} \boldsymbol{p}^{(k)} - \alpha \boldsymbol{r}^{(k)T} \boldsymbol{p}^{(k)} + \varphi(\boldsymbol{x}^{(k)}).
\end{aligned}
$$

其中 $\boldsymbol{r}^{(k)} = \boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}^{(k)}$ 为残量.

由极值必要条件: $\dfrac{\mathrm{d}f}{\mathrm{d}\alpha} = \alpha \boldsymbol{p}^{(k)T} \boldsymbol{A} \boldsymbol{p}^{(k)} - \boldsymbol{r}^{(k)T} \boldsymbol{p}^{(k)} = 0,$

得到: $\alpha_k = \dfrac{\boldsymbol{r}^{(k)T} \boldsymbol{p}^{(k)}}{\boldsymbol{p}^{(k)T} \boldsymbol{A} \boldsymbol{p}^{(k)}},$

又 $\dfrac{\mathrm{d}^2 f}{\mathrm{d}\alpha^2} = \boldsymbol{p}^{(k)T} \boldsymbol{A} \boldsymbol{p}^{(k)} > 0 \ (\boldsymbol{p}^{(k)} \neq 0),$

故α_k 为$f(\alpha)$ 的极小值点, 于是 $\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} + \alpha_k \boldsymbol{p}^{(k)}$.

$$\varphi(\boldsymbol{x}^{(k+1)}) - \varphi(\boldsymbol{x}^{(k)}) = -\frac{(\boldsymbol{r}^{(k)T} \boldsymbol{p}^{(k)})^2}{2\boldsymbol{p}^{(k)T} \boldsymbol{A} \boldsymbol{p}^{(k)}}.$$

$\varphi(\boldsymbol{x}^{(k+1)}) < \varphi(\boldsymbol{x}^{(k)}), (\boldsymbol{r}_k^T \boldsymbol{p}^{(k)} \neq 0).$

(2). 确定下山方向$\boldsymbol{p}^{(k)}$

若选取$\boldsymbol{p}^{(k)}$ 为负梯度方向, 即$\boldsymbol{p}^{(k)} = -(\boldsymbol{A}\boldsymbol{x}^{(k)} - \boldsymbol{b}) = \boldsymbol{r}^{(k)}$, 就得到最速下降法(梯度下降法, GD 法)

算法1:

选取初值$\boldsymbol{x}^{(0)}$,

对$k = 0, 1, 2, \cdots,$ $\boldsymbol{r}^{(k)} = \boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}^{(k)}$,

$$\alpha_k = \frac{(\boldsymbol{r}^{(k)}, \boldsymbol{r}^{(k)})}{(\boldsymbol{A}\boldsymbol{r}^{(k)}, \boldsymbol{r}^{(k)})},$$

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} + \alpha_k \boldsymbol{r}^{(k)}.$$

梯度下降法的性质:

- $(\boldsymbol{r}^{(k+1)}, \boldsymbol{r}^{(k)}) = 0$, 即$\boldsymbol{r}^{(k+1)}$与$\boldsymbol{r}^{(k)}$ 正交.
- $\{\varphi(\boldsymbol{x}^{(k)})\}$是一个单调下降序列, 且$\lim\limits_{k\to\infty} \boldsymbol{x}^{(k)} = \boldsymbol{x}^*$ $(\boldsymbol{A}\boldsymbol{x}^* = \boldsymbol{b})$. 并且有估计:

$$||\boldsymbol{x}^{(k)} - \boldsymbol{x}^*||_{\boldsymbol{A}} \le \left(\frac{\lambda_1 - \lambda_n}{\lambda_1 + \lambda_n}\right)^k ||\boldsymbol{x}^{(0)} - \boldsymbol{x}^*||_{\boldsymbol{A}},$$

其中: $||\boldsymbol{x}||_{\boldsymbol{A}} = \sqrt{\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}}$, $\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_n > 0$ 为$\boldsymbol{A}$的特征值.

梯度下降法的优缺点:

- 简单易用, 又可充分利用$\boldsymbol{A}$的稀疏性
- 当$\lambda_1 \gg \lambda_n$时收敛非常慢; 当$||\boldsymbol{r}^{(k)}||$很小时, 由于舍入误差, 计算不稳定.

例：用梯度下降法求解$\boldsymbol{Ax}=\boldsymbol{b}$, 其中$\boldsymbol{A}=\begin{pmatrix}1/18 & 0\\ 0 & 2\end{pmatrix}$, $\boldsymbol{b}=(0,0)^T$, $\boldsymbol{x}^{(0)}=(0.625,0.1)^T$.

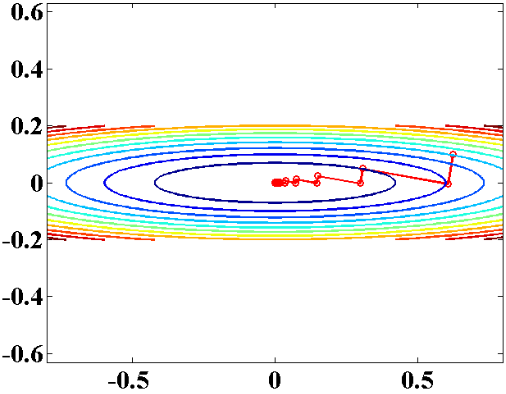

Figure: GD法迭代点图,迭代60次,误差$O(10^{-9})$.

3.4.2 共轭梯度法(CG法)

1. 计算过程:

给定初始向量$\boldsymbol{x}^{(0)}$, 第一步仍选负梯度方向为下山方向, 即$\boldsymbol{p}^{(0)}=\boldsymbol{r}^{(0)}$,

于是有: $\alpha_0=\dfrac{\boldsymbol{r}^{(0)T}\boldsymbol{r}^{(0)}}{\boldsymbol{r}^{(0)T}\boldsymbol{A}\boldsymbol{r}^{(0)}}$, $\boldsymbol{x}^{(1)}=\boldsymbol{x}^{(0)}+\alpha_0\boldsymbol{r}^{(0)}$, $\boldsymbol{r}^{(1)}=\boldsymbol{b}-\boldsymbol{A}\boldsymbol{x}^{(1)}$.

对于第$k+1(k\geq 1)$步, 下山方向不再取负梯度方向$\boldsymbol{r}^{(k)}$, 而是取为经过点$\boldsymbol{x}^{(k)}$ 由向量$\boldsymbol{r}^{(k)}$和$\boldsymbol{p}^{(k-1)}$ 所张成的二维平面

$$\pi_2=\{\boldsymbol{x}=\boldsymbol{x}^{(k)}+\xi\boldsymbol{r}^{(k)}+\eta\boldsymbol{p}^{(k-1)};\ \xi,\ \eta\in R\}.$$

内, 找出使函数φ下降最快的方向作为新的下山方向$\boldsymbol{p}^{(k)}$.

考虑φ在π_2上的限制:

$$\begin{aligned}\varphi(\xi,\eta) = \varphi(\boldsymbol{x}^{(k)} + \xi\boldsymbol{r}^{(k)} + \eta\boldsymbol{p}^{(k-1)}) \\ = \frac{1}{2}(\boldsymbol{x}^{(k)} + \xi\boldsymbol{r}^{(k)} + \eta\boldsymbol{p}^{(k-1)})^T\boldsymbol{A}(\boldsymbol{x}^{(k)} + \xi\boldsymbol{r}^{(k)} + \eta\boldsymbol{p}^{(k-1)}) \\ -\boldsymbol{b}^T(\boldsymbol{x}^{(k)} + \xi\boldsymbol{r}^{(k)} + \eta\boldsymbol{p}^{(k-1)}),\end{aligned}$$

根据函数取极值必要条件:

$$\begin{aligned}\frac{\partial\varphi}{\partial\xi} &= \xi\boldsymbol{r}^{(k)T}\boldsymbol{A}\boldsymbol{r}^{(k)} + \eta\boldsymbol{r}^{(k)T}\boldsymbol{A}\boldsymbol{p}^{(k-1)} - \boldsymbol{r}^{(k)T}\boldsymbol{r}^{(k)} = 0, \\ \frac{\partial\varphi}{\partial\eta} &= \xi\boldsymbol{r}^{(k)T}\boldsymbol{A}\boldsymbol{p}^{(k-1)} + \eta\boldsymbol{p}^{(k-1)T}\boldsymbol{A}\boldsymbol{p}^{(k-1)} = 0,\end{aligned}$$

这里用到了$\boldsymbol{r}^{(k)T}\boldsymbol{p}^{(k-1)} = 0$.

可得φ在π_2内有唯一极小点:

$\widetilde{\boldsymbol{x}}=\boldsymbol{x}^{(k)}+\xi_0\boldsymbol{r}^{(k)}+\eta_0\boldsymbol{p}^{(k-1)}$,

其中ξ_0, η_0满足:

$$\begin{cases}\xi_0\boldsymbol{r}^{(k)T}\boldsymbol{A}\boldsymbol{r}^{(k)}+\eta_0\boldsymbol{r}^{(k)T}\boldsymbol{A}\boldsymbol{p}^{(k-1)}=\boldsymbol{r}^{(k)T}\boldsymbol{r}^{(k)},\\ \xi_0\boldsymbol{r}^{(k)T}\boldsymbol{A}\boldsymbol{p}^{(k-1)}+\eta_0\boldsymbol{p}^{(k-1)T}\boldsymbol{A}\boldsymbol{p}^{(k-1)}=0.\end{cases}$$

于是, 可取 $\boldsymbol{p}^{(k)}=\frac{1}{\xi_0}(\widetilde{\boldsymbol{x}}-\boldsymbol{x}^{(k)})=\boldsymbol{r}^{(k)}+\frac{\eta_0}{\xi_0}\boldsymbol{p}^{(k-1)}$ 作为新的下山方向.

显然, 这是在平面π_2内找到的最佳下山方向.

令$\beta_{k-1}=\frac{\eta_0}{\xi_0}$, 则$\beta_{k-1}=-\frac{\boldsymbol{r}^{(k)T}\boldsymbol{A}\boldsymbol{p}^{(k-1)}}{\boldsymbol{p}^{(k-1)T}\boldsymbol{A}\boldsymbol{p}^{(k-1)}}$.

注意到, 这样确定的$\boldsymbol{p}^{(k)}$满足: $\boldsymbol{p}^{(k)T}\boldsymbol{A}\boldsymbol{p}^{(k-1)}=0$.

定义: $\boldsymbol{A}$对称正定, 如果R^n中向量组$\{\boldsymbol{p}^{(0)},\boldsymbol{p}^{(1)},\cdots,\boldsymbol{p}^{(l)}\}$满足

$$\left(\boldsymbol{A}\boldsymbol{p}^{(i)},\boldsymbol{p}^{(j)}\right)=0,\quad i\neq j.$$

则称它为R^n中的一个$\boldsymbol{A}$-共轭向量组.

即: $\boldsymbol{p}^{(k)}$与$\boldsymbol{p}^{(k-1)}$ 是$\boldsymbol{A}$共轭的.

$\boldsymbol{p}^{(k)}$, $\boldsymbol{p}^{(k-1)}$, $\boldsymbol{r}^{(k)}$的几何意义?

确定了$\boldsymbol{p}^{(k)}$后，可确定α_k，于是第$k+1$步的迭代计算过程如下：

计算搜索方向$\boldsymbol{p}^{(k)}$：

$$\boldsymbol{r}^{(k)} = \boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}^{(k)}, \quad \beta_{k-1} = -\frac{\boldsymbol{r}^{(k)T}\boldsymbol{A}\boldsymbol{p}^{(k-1)}}{\boldsymbol{p}^{(k-1)T}\boldsymbol{A}\boldsymbol{p}^{(k-1)}},$$

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

计算步长α_k:

$$\alpha_k = \frac{\boldsymbol{r}^{(k)T}\boldsymbol{p}^{(k)}}{\boldsymbol{p}^{(k)T}\boldsymbol{A}\boldsymbol{p}^{(k)}},$$

计算新的迭代点$\boldsymbol{x}^{(k+1)}$: $\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} + \alpha_k\boldsymbol{p}^{(k)}$;

简化计算公式:

(1). 简化$\boldsymbol{r}^{(k)}$ 的计算公式:

$$\boldsymbol{r}^{(k)} = \boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}^{(k)} = \boldsymbol{b} - \boldsymbol{A}(\boldsymbol{x}^{(k-1)} + \alpha_{k-1}\boldsymbol{p}^{(k-1)}) = \boldsymbol{r}^{(k-1)} - \alpha_{k-1}\boldsymbol{A}\boldsymbol{p}^{(k-1)}.$$

(2). 简化α_k, β_{k-1} 的计算公式:

注意到: $\boldsymbol{r}^{(k)T}\boldsymbol{r}^{(k-1)} = \boldsymbol{r}^{(k)T}\boldsymbol{p}^{(k-1)} = \boldsymbol{r}^{(k+1)T}\boldsymbol{p}^{(k)} = 0,\ k = 1, 2, \cdots$

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

$$\boldsymbol{r}^{(k)T}\boldsymbol{A}\boldsymbol{p}^{(k-1)} = \frac{1}{\alpha_{k-1}}\boldsymbol{r}^{(k)T}(\boldsymbol{r}^{(k-1)} - \boldsymbol{r}^{(k)}) = -\frac{1}{\alpha_{k-1}}\boldsymbol{r}^{(k)T}\boldsymbol{r}^{(k)},$$

$$\boldsymbol{p}^{(k)T}\boldsymbol{A}\boldsymbol{p}^{(k)} = \frac{1}{\alpha_k}\boldsymbol{p}^{(k)T}(\boldsymbol{r}^{(k)} - \boldsymbol{r}^{(k+1)}) = \frac{1}{\alpha_k}\boldsymbol{p}^{(k)T}\boldsymbol{r}^{(k)} = \frac{1}{\alpha_k}\boldsymbol{r}^{(k)T}\boldsymbol{r}^{(k)}.$$

于是:

$$\alpha_k = \frac{\boldsymbol{r}^{(k)T}\boldsymbol{r}^{(k)}}{\boldsymbol{p}^{(k)T}\boldsymbol{A}\boldsymbol{p}^{(k)}}, \quad \beta_{k-1} = \frac{\boldsymbol{r}^{(k)T}\boldsymbol{r}^{(k)}}{\boldsymbol{r}^{(k-1)T}\boldsymbol{r}^{(k-1)}}.$$

算法2(共轭梯度法):

选取初始向量$\boldsymbol{x}^{(0)}$, $\boldsymbol{r}^{(0)} = \boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}^{(0)}$, $\alpha_0 = \dfrac{\boldsymbol{r}^{(0)T}\boldsymbol{r}^{(0)}}{\boldsymbol{r}^{(0)T}\boldsymbol{A}\boldsymbol{r}^{(0)}}$,

$\boldsymbol{x}^{(1)} = \boldsymbol{x}^{(0)} + \alpha_0 \boldsymbol{r}^{(0)}$;

对于$k = 1, 2, \cdots$, $\boldsymbol{r}^{(k)} = \boldsymbol{r}^{(k-1)} - \alpha_{k-1}\boldsymbol{A}\boldsymbol{p}^{(k-1)}$,

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

$$\alpha_k = \frac{\boldsymbol{r}^{(k)T}\boldsymbol{r}^{(k)}}{\boldsymbol{p}^{(k)T}\boldsymbol{A}\boldsymbol{p}^{(k)}}, \qquad \boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} + \alpha_k \boldsymbol{p}^{(k)}.$$

例： 应用共轭梯度法求解方程组

$$\begin{pmatrix} 2 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 3 \\ 1 \\ 3 \end{pmatrix}$$

解：

(1) 当 $k=1$ 时，取$\boldsymbol{x}_0 = (0,0,0)^T$, 则

$$\boldsymbol{p}^{(0)} = \boldsymbol{r}^{(0)} = \boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}^{(0)} = (3,1,3)^T$$

计算得 $\boldsymbol{A}\boldsymbol{p}^{(0)} = (9,1,9)^T$， $\boldsymbol{p}^{(0)T}\boldsymbol{A}\boldsymbol{p}^{(0)} = 55$， 因此

$$\alpha_0 = \frac{(\boldsymbol{r}^{(0)}, \boldsymbol{p}^{(0)})}{(\boldsymbol{p}^{(0)}, \boldsymbol{A}\boldsymbol{p}^{(0)})} = \frac{19}{55}, \qquad \boldsymbol{x}^{(1)} = \boldsymbol{x}^{(0)} + \alpha_0 \boldsymbol{p}^{(0)} = \frac{19}{55}(3,1,3)^T,$$

(2) 当$k=2$ 时，$\boldsymbol{r}^{(1)}=\boldsymbol{b}-\boldsymbol{A}\boldsymbol{x}^{(1)}=\boldsymbol{r}^{(0)}-\alpha_0\boldsymbol{A}\boldsymbol{p}^{(0)}=\frac{6}{55}(-1,6,-1)^T$，

$\beta_0=-\frac{(\boldsymbol{r}^{(1)},\boldsymbol{r}^{(1)})}{(\boldsymbol{r}^{(0)},\boldsymbol{r}^{(0)})}=\frac{72}{55^2}$， $\boldsymbol{p}^{(1)}=\boldsymbol{r}^{(1)}+\beta_0\boldsymbol{p}^{(0)}=\frac{6\times19}{55^2}(-1,18,-1)^T$

计算得 $\boldsymbol{A}\boldsymbol{p}^{(1)}=\frac{6\times3\times19}{55^2}(-1,6,-1)^T$，$\boldsymbol{p}^{(1)T}\boldsymbol{A}\boldsymbol{p}^{(1)}=\frac{6^3\times19^2}{55^2}$

$\boldsymbol{r}^{(1)T}\boldsymbol{r}^{(1)}=\frac{6^2\times19\times2}{55^2}$, $\alpha_1=\frac{(\boldsymbol{r}^{(1)},\boldsymbol{r}^{(1)})}{(\boldsymbol{p}^{(1)},\boldsymbol{A}\boldsymbol{p}^{(1)})}=\frac{55}{3\times19}$

$\boldsymbol{x}^{(2)}=\boldsymbol{x}^{(1)}+\alpha_1\boldsymbol{p}^{(1)}=(1,1,1)^T$, $\boldsymbol{r}^{(2)}=\boldsymbol{b}-\boldsymbol{A}\boldsymbol{x}^{(2)}=\boldsymbol{0}$

故方程组的解为 $\boldsymbol{x}^{(2)}=(1,1,1)^T$。

2. CG法的相关概念及性质

(1). 共轭梯度法的基本性质:

定理: 由共轭梯度法得到的向量组$\{\boldsymbol{r}^{(i)}\}$和$\{\boldsymbol{p}^{(i)}\}$具有如下性质:

(i) $\boldsymbol{p}^{(i)T}\boldsymbol{r}^{(j)}=0,\ 0\leq i<j\leq k$

(ii) $\boldsymbol{r}^{(i)T}\boldsymbol{r}^{(j)}=0,\ i\neq j,\ 0\leq i,j\leq k$

(iii) $\boldsymbol{p}^{(i)T}\boldsymbol{A}\boldsymbol{p}^{(j)}=0,\ i\neq j,\ 0\leq i,j\leq k$

(iv) $\mathrm{span}\{\boldsymbol{r}^{(0)},\boldsymbol{r}^{(1)},\cdots,\boldsymbol{r}^{(k)}\}=\mathrm{span}\{\boldsymbol{p}^{(0)},\boldsymbol{p}^{(1)},\cdots,\boldsymbol{p}^{(k)}\}=\mathcal{K}(\boldsymbol{A},\boldsymbol{r}^{(0)},k+1)$, 其中:

$$\mathcal{K}(\boldsymbol{A},\boldsymbol{r}^{(0)},k+1)=\mathrm{span}\{\boldsymbol{r}^{(0)},\boldsymbol{A}\boldsymbol{r}^{(0)},\cdots,\boldsymbol{A}^k\boldsymbol{r}^{(0)}\}.$$

通常称为Krylov子空间.

定理: 共轭梯度法最多n步便可得到方程组的解$\boldsymbol{x}^*$.

(2). 定理: 用CG法计算得到的近似解$\boldsymbol{x}^{(k)}$满足:

$$\varphi(\boldsymbol{x}^{(k)}) = \min\{\varphi(\boldsymbol{x}) : \ \boldsymbol{x} \in \boldsymbol{x}^{(0)} + \mathcal{K}(\boldsymbol{A}, \boldsymbol{r}_0, k)\}.$$

或 $$||\boldsymbol{x}^{(k)} - \boldsymbol{x}^*||_{\boldsymbol{A}} = \min\left\{||\boldsymbol{x} - \boldsymbol{x}^*||_{\boldsymbol{A}} : \ \boldsymbol{x} \in \boldsymbol{x}^{(0)} + \mathcal{K}(\boldsymbol{A}, \boldsymbol{r}^{(0)}, k)\right\},$$

其中: $||\boldsymbol{x}||_{\boldsymbol{A}} = \sqrt{\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}}$, $\boldsymbol{x}^*$是方程组$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ 的解, $\mathcal{K}(\boldsymbol{A}, \boldsymbol{r}_0, k)$是由前面定理定义的Krylov子空间.

注: CG法第k步迭代得到的解实质就是由$\boldsymbol{x}^{(0)}$出发(经过$\boldsymbol{x}^{(0)}$), 由前面各步的搜索方向张成的子空间中的最优解.

(3). $\boldsymbol{A}$共轭向量组:

定理: $\boldsymbol{A}$-共轭向量组是线性无关的.

推论: 若R^n中一组$\boldsymbol{A}$-共轭向量组中向量个数为n, 则构成R^n 的一组基(称为$\boldsymbol{A}$-共轭基)

(4). 向量组的正交化与A-共轭化

Schmidt正交化

A-共轭化:

设$\boldsymbol{\alpha}^{(1)},\cdots,\boldsymbol{\alpha}^{(k-1)}$线性无关, 令$\boldsymbol{p}^{(1)}=\boldsymbol{\alpha}^{(1)}$, 且由$\boldsymbol{\alpha}^{(1)},\cdots,\boldsymbol{\alpha}^{(k-1)}$的组合已产生$A$-共轭向量组$\boldsymbol{p}^{(1)},\boldsymbol{p}^{(2)},\cdots,\boldsymbol{p}^{(k-1)}$, 令

$$\boldsymbol{p}^{(k)}=\boldsymbol{\alpha}^{(k)}+l_{k,1}\boldsymbol{p}^{(1)}+\cdots+l_{k,k-1}\boldsymbol{p}^{(k-1)},$$

由$(\boldsymbol{p}^{(k)},\boldsymbol{A}\boldsymbol{p}^{(j)})=(\boldsymbol{\alpha}^{(k)},\boldsymbol{A}\boldsymbol{p}^{(j)})+l_{k,j}(\boldsymbol{p}^{(j)},\boldsymbol{A}\boldsymbol{p}^{(j)})=0,\quad j=1,2,\cdots,k-1.$

于是:
$$l_{k,j}=-\frac{(\boldsymbol{\alpha}^{(k)},\boldsymbol{A}\boldsymbol{p}^{(j)})}{(\boldsymbol{A}\boldsymbol{p}^{(j)},\boldsymbol{p}^{(j)})}.$$

从而有:

$$\begin{cases} \boldsymbol{p}^{(1)} = \boldsymbol{\alpha}^{(1)}, \\ \boldsymbol{p}^{(k)} = \boldsymbol{\alpha}^{(k)} - \displaystyle\sum_{j=1}^{k-1} \frac{(\boldsymbol{\alpha}^{(k)}, \boldsymbol{A}\boldsymbol{p}^{(j)})}{(\boldsymbol{A}\boldsymbol{p}^{(j)}, \boldsymbol{p}^{(j)})} \boldsymbol{p}^{(j)} \\ \qquad\ = \boldsymbol{\alpha}^{(k)} - \displaystyle\sum_{j=1}^{k-1} \frac{(\boldsymbol{A}\boldsymbol{\alpha}^{(k)}, \boldsymbol{p}^{(j)})}{(\boldsymbol{A}\boldsymbol{p}^{(j)}, \boldsymbol{p}^{(j)})} \boldsymbol{p}^{(j)} \end{cases}$$

3. 实用CG法

应用CG法存在的问题:

◇ 理论上CG法至多经n步便得到方程组的精确解, 然而实际使用时, 由于舍入误差, 使得$\boldsymbol{r}^{(k)}$之间的正交性很快损失, 以致于其有限步终止性不再成立.

◇ 实际问题中, 由于n一般很大, 以致于迭代$O(n)$次耗时巨大, 因此实际上将CG法作为一种迭代算法使用. 迭代终止准则一般为: $||\boldsymbol{r}^{(k)}||$已经很小或迭代次数已达最大容许次数.

算法3(实用CG法):

设置ϵ, $k_{\max}$:

(1) 取初值$\boldsymbol{x}^{(0)} \in R^n$, 计算$\boldsymbol{r}^{(0)} = \boldsymbol{p}^{(0)} = \boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}^{(0)}$,

对$k = 0, 1, 2, \cdots$

(2) 计算$\alpha_k = \dfrac{\left(\boldsymbol{r}^{(k)}, \boldsymbol{r}^{(k)}\right)}{\left(\boldsymbol{A}\boldsymbol{p}^{(k)}, \boldsymbol{p}^{(k)}\right)}$,

(3) 计算$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} + \alpha_k \boldsymbol{p}^{(k)}$,

(4) 计算$\boldsymbol{r}^{(k+1)} = \boldsymbol{r}^{(k)} - \alpha_k \boldsymbol{A}\boldsymbol{p}^{(k)}$,

(5) 若$||\boldsymbol{r}^{(k+1)}||_2 < \epsilon$或$k > k_{\max}$, 输出$\boldsymbol{x}^{(k+1)}$, 迭代停止, 否则转下一步,

(6) 计算$\beta_k = \dfrac{\left(\boldsymbol{r}^{(k+1)}, \boldsymbol{r}^{(k+1)}\right)}{\left(\boldsymbol{r}^{(k)}, \boldsymbol{r}^{(k)}\right)}$,

(7) 计算$\boldsymbol{p}^{(k+1)} = \boldsymbol{r}^{(k+1)} + \beta_k \boldsymbol{p}^{(k)}$, $k = k + 1$, 转(2).

4. CG法的收敛性

定理: 用CG法求得的$\boldsymbol{x}^{(k)}$有如下误差估计:

$$||\boldsymbol{x}^{(k)}-\boldsymbol{x}^*||_{\boldsymbol{A}} \le 2\left(\frac{\sqrt{\kappa_2}-1}{\sqrt{\kappa_2}+1}\right)^k ||\boldsymbol{x}^{(0)}-\boldsymbol{x}^*||_{\boldsymbol{A}},$$

其中 $\kappa_2=\mathrm{cond}_2(\boldsymbol{A})=||\boldsymbol{A}||_2||\boldsymbol{A}^{-1}||_2=\dfrac{\lambda_1}{\lambda_n}$,
λ_1, λ_n分别为$\boldsymbol{A}$的最大特征值和最小特征值.

注: 只要线性方程组的系数矩阵是十分良态的($\mathrm{cond}_2(\boldsymbol{A})\approx 1$), 则CG法收敛很快.

5. CG法的优点:

- ◇ 算法中，系数矩阵A的作用仅仅用来由已知向量p 产生向量$w = AP$，可以充分利用A的稀疏性，且对某些提供A较困难但由p产生Ap十分方便的应用问题非常有益.
- ◇ 不需要预先估计任何参数.
- ◇ 每次迭代所需的计算，主要是向量之间的运算，便于并行化.

3.4.3 预处理共轭梯度法

1. 基本思想:

CG法在系数矩阵仅有少数几个互不相同的特征值或者非常良态时, 收敛非常快, 因此, 在应用CG法时, 首先应设法将原方程等价转化为一个系数矩阵仅有少数几个互不相同的特征值或非常良态的方程组, 然后再用CG 法求解转化后的方程组.

2. 算法:

设S为一可逆阵,

$$\boldsymbol{Ax}=\boldsymbol{b} \;\Leftrightarrow\; \boldsymbol{S}^{-1}\boldsymbol{A}\boldsymbol{S}^{-T}\boldsymbol{S}^{T}\boldsymbol{x}=\boldsymbol{S}^{-1}\boldsymbol{b} \;\Leftrightarrow\; \boldsymbol{H}\boldsymbol{u}=\boldsymbol{g},$$

其中: $\boldsymbol{H}=\boldsymbol{S}^{-1}\boldsymbol{A}\boldsymbol{S}^{-T}$, $\boldsymbol{u}=\boldsymbol{S}^{T}\boldsymbol{x}$, $\boldsymbol{g}=\boldsymbol{S}^{-1}\boldsymbol{b}$.

对$\boldsymbol{H}\boldsymbol{u}=\boldsymbol{g}$应用CG法, 记$\boldsymbol{M}=\boldsymbol{S}\boldsymbol{S}^{T}$, 整理得算法4(PCG 算法):

设置ϵ, $k_{\max}$:

(1) 取初值$\boldsymbol{x}^{(0)}\in R^{n}$, 计算$\boldsymbol{r}^{(0)}=\boldsymbol{b}-\boldsymbol{A}\boldsymbol{x}^{(0)}$, $\boldsymbol{z}^{(0)}=\boldsymbol{M}^{-1}\boldsymbol{r}^{(0)}$, $\boldsymbol{p}^{(0)}=\boldsymbol{z}^{(0)}$,

对$k=0,1,2,\cdots$

(2) 计算$\alpha_k=\dfrac{(\boldsymbol{z}^{(k)},\boldsymbol{r}^{(k)})}{(\boldsymbol{A}\boldsymbol{p}^{(k)},\boldsymbol{p}^{(k)})}$,

(3) 计算$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} + \alpha_k \boldsymbol{p}^{(k)}$,

(4) 计算$\boldsymbol{r}^{(k+1)} = \boldsymbol{r}^{(k)} - \alpha_k \boldsymbol{A}\boldsymbol{p}^{(k)}$,

(5) 若$||\boldsymbol{r}^{(k+1)}||_2 < \epsilon$或$k > k_{\max}$, 输出$\boldsymbol{x}^{(k+1)}$, 迭代停止, 否则转下一步,

(6) 计算$\boldsymbol{z}^{(k+1)} = \boldsymbol{M}^{-1}\boldsymbol{r}^{(k+1)}$, (或求解$\boldsymbol{M}\boldsymbol{z}^{(k+1)} = \boldsymbol{r}^{(k+1)}$),

(7) 计算$\beta_k = \dfrac{(\boldsymbol{z}^{(k+1)}, \boldsymbol{r}^{(k+1)})}{(\boldsymbol{z}^{(k)}, \boldsymbol{r}^{(k)})}$,

(8) 计算$\boldsymbol{p}^{(k+1)} = \boldsymbol{z}^{(k+1)} + \beta_k \boldsymbol{p}^{(k)}$, $k = k+1$, 转(2).

3. PCG方法的性质和误差估计:

(1). 由$(\widetilde{\boldsymbol{r}}^{(i)},\widetilde{\boldsymbol{r}}^{(j)})=0$, 有: $(\boldsymbol{M}^{-1}\boldsymbol{r}^{(i)},\boldsymbol{r}^{(j)})=0,\ i\neq j$;

由$(\boldsymbol{H}\widetilde{\boldsymbol{p}}^{(i)},\widetilde{\boldsymbol{p}}^{(j)})=0$, 有: $(\boldsymbol{A}\boldsymbol{p}^{(i)},\boldsymbol{p}^{(j)})=0,\ i\neq j$.

即$\{\boldsymbol{r}^{(k)}\}$是内积$(\cdot,\cdot)_{\boldsymbol{M}^{-1}}$意义下的正交组, $\{\boldsymbol{p}^{(k)}\}$是$\boldsymbol{A}$-共轭向量组.

(2). $$||\boldsymbol{x}^{(k)}-\boldsymbol{x}^*||_{\boldsymbol{A}}\leq 2\left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\right)^k||\boldsymbol{x}^{(0)}-\boldsymbol{x}^*||_{\boldsymbol{A}},$$

其中κ 是$\boldsymbol{M}^{-1}\boldsymbol{A}$的最大特征值和最小特征值之比.

4. 预处理矩阵M的选取

(1). M应满足的条件:

a. M对称正定
b. M是稀疏的
c. $M^{-1}A$的特征值分布“集中”，即条件数小
d. 形如$Mz = r$的方程容易求解(M具有分块或三角阵乘积等较好的性质)

5. 常用的预处理矩阵M的选取技巧

(1). 对角预处理阵:

若线性方程组的系数矩阵A的对角元素相差较大, 则可取

$M = \text{diag}(a_{11}, a_{22}, \cdots, a_{nn})$作为预处理阵, 常会使收敛速度大大提高.

推广: 若

$$\boldsymbol{A} = \begin{pmatrix} \boldsymbol{A}_{11} & \cdots & \boldsymbol{A}_{1k} \\ \vdots & \vdots & \vdots \\ \boldsymbol{A}_{k1} & \cdots & \boldsymbol{A}_{kk} \end{pmatrix}, \text{其中 } \boldsymbol{A}_{ii}\text{是易于求逆的方阵,}$$

则可取$\boldsymbol{M} = \text{diag}(\boldsymbol{A}_{11}, \cdots, \boldsymbol{A}_{kk})$.

(2). 不完全Cholesky分解法:

先求系数矩阵$\boldsymbol{A}$的不完全Cholesky分解:

$$\boldsymbol{A} = \boldsymbol{L}\boldsymbol{L}^T + \boldsymbol{R},$$

其中$\boldsymbol{L}$为单位下三角阵. 选取$\boldsymbol{M} = \boldsymbol{L}\boldsymbol{L}^T$作为预处理阵, 由于有$\boldsymbol{R}$可选择, 可调整$\boldsymbol{L}$具有与$\boldsymbol{A}$一样的稀疏性, 当然$\boldsymbol{L}\boldsymbol{L}^T$越接近$\boldsymbol{A}$越好.

缺点: 求$\boldsymbol{M}\boldsymbol{z} = \boldsymbol{r}$时需要求解两个三角方程组, 并行效率低.