

数值线性代数

线性方程组的间接 (迭代) 解法

参考书目:

- 数值线性代数 (徐树方, 2013)
- Iterative Methods for Sparse Linear Systems (Saad, 2000)
- Accuracy and Stability of Numerical Algorithms (Higham, 2002)
- Fundamentals of Matrix Computations (Watkins, 2010)

本文主要介绍线性方程组的迭代解法, 分为以矩阵分割为基础的古典迭代法和以子空间投影为基础的现代迭代法 (Krylov 子空间方法) 两大部分。前者包括 Jacobi、Gauss-Seidel、SOR 等方法, 后者包括 CG、GMRES、BiCGSTAB 等方法。对于较完善的古典法将给出各方法的理论形式, 收敛性分析, 以及数值算法。而对于现代迭代法, 则主要介绍理论部分。

一般地, 线性方程组的迭代法是指采取逐次逼近的方法, 即从一个初始向量出发, 按照一定的计算格式, 构造一个向量序列, 该序列收敛到问题的真实解。对于大规模稀疏线性方程组, 迭代法往往比直接法更加高效。

1 古典迭代法——矩阵分割

古典迭代法是指以矩阵分割为基础的迭代法, 其中最重要的一类是单步线性定常 (*stationary*) 迭代法, 这种方法的基本形式是将要求解的线性方程组 $Ax = b$ 转化为 $Mx = Nx + b$, 令迭代格式为

$$Mx^{(k+1)} = Nx^{(k)} + b,$$

其中 $A = M - N$ 和 M 都要求是可逆矩阵, 并且为了保证在任意初始向量下得到的序列都收敛, 需要谱半径 $\rho(M^{-1}N) < 1$ 。如果在某种矩阵分割下, 线性方程组 $(I - M^{-1}N)x = M^{-1}b$ 与 $Ax = b$ 等价, 即存在非奇异矩阵 C 使得 $C(I - M^{-1}N) = A$, $CM^{-1}b = b$, 则称相应的迭代法与原方程组是相容的。当迭代法与原问题相容时, 原问题的解是迭代格式的一个不动点, 因此问题的求解就转化为了寻找不动点。常见的单步迭代法有 Jacobi 迭代法、Gauss-Seidel 迭代法、SOR 迭代法, 这三种方式基于 $A = D - L - U$ 的矩阵分割, 其中 D 是 A 的对角元素构成的对角矩阵, $-L$ 和 $-U$ 分别是 A 的严格下三角和严格上三角矩阵。

1.1 Jacobi 迭代法, Gauss-Seidel 迭代法, SOR 迭代法

首先, Jacobi 迭代法等价于取 $M = D$, $N = D - A = L + U$, 迭代格式为

$$Dx^{(k+1)} = (L + U)x^{(k)} + b,$$

该方法实际上是在要求下一次的迭代向量的第 i 各分量要使得 $x^{(k+i/n)} = (x_1^{(k)}, \dots, x_i^{(k+1)}, \dots, x_n^{(k)})$ 相应的残差向量的第 i 个分量为 0, 即

$$(b - Ax^{(k+i/n)})_i = 0,$$

上式写成分量形式即为

$$a_{ii}x_i^{(k+1)} = \sum_{j \neq i} a_{ij}x_j^{(k)} + b_i.$$

Gauss-Seidel 迭代则是取 $M = D - L$, $N = U$, 迭代格式为

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

因为 $D - L$ 是一个下三角矩阵, 因此可以使用前代法求解, 所以上述格式也称作前向 G-E 迭代。类似地, 后向 G-E 迭代形如

$$(D - U)x^{(k+1)} = Lx^{(k)} + b.$$

该方法与 Jacobi 迭代不同, 它要求的是 (当使用前向 G-E 时, 计算顺序从前到后) 下一次的迭代向量的第 i 各分量要使得 $x^{(k+i/n)} = (x_1^{(k+1)}, \dots, x_i^{(k+1)}, \dots, x_n^{(k)})$ 相应的残差向量的第 i 个分量为 0, 所以现在

$$(b - Ax^{(k+i/n)})_i = 0$$

的分量形式变为

$$a_{ii}x_i^{(k+1)} = \sum_{j < i} a_{ij}x_j^{(k+1)} + \sum_{j > i} a_{ij}x_j^{(k)} + b_i.$$

因为在计算 $x_i^{(k+1)}$ 时, $x_j^{(k+1)}$ 已经是新的值, 所以直觉上这种迭代法的收敛速度要比 Jacobi 迭代法快, 并且由于无需储存上一次迭代向量的前面的分量, 因此可以节省内存。

最后, 注意到通过引入参数 ω 有如下分割

$$\omega A = (D - \omega L) - [(1 - \omega)D + \omega U],$$

这样就得到了 *SOR*(连续超松弛, *successive over relaxation*) 迭代法: 取 $M = D - \omega L$, $N = (1 - \omega)D + \omega U$, 迭代格式为

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

与 G-E 迭代法类似, *SOR* 迭代法也有前向和后向两种形式, 上述形式为前向 *SOR* 迭代, 而后向 *SOR* 迭代为

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

除了这两种形式外, 还有一种混合形式, 即在每一步迭代中都先进行一次前向 *SOR* 迭代, 再进行一次后向 *SOR* 迭代, 称为对称 *SOR* 迭代 (SSOR), 即

$$\begin{aligned} (D - \omega L)x^{(k+1/2)} &= [(1 - \omega)D + \omega U]x^{(k)} + \omega b, \\ (D - \omega U)x^{(k+1)} &= [(1 - \omega)D + \omega L]x^{(k+1/2)} + \omega b. \end{aligned}$$

如果记 (前向)G-E 的各步迭代向量为 $x^{GE(k)}$, 从第 k 步到第 $k+1$ 步的修正项为 $\Delta x^{GE(k)}$, 即 $x^{GE(k+1)} = x^{GE(k)} + \Delta x^{GE(k)}$, 则 (前向)*SOR* 迭代法就相当于在修正项上添加了一个松弛因子 ω , 即 $x^{SOR(k+1)} = x^{SOR(k)} + \omega \Delta x^{GE(k)}$, 即

$$x^{(k+1)} = (1 - \omega)x^{(k)} + \omega D^{-1}(Lx^{(k+1)} + Ux^{(k)} + b),$$

相应的分量形式为

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

为了保证收敛性, ω 的取值范围为 $0 < \omega < 2$, 当 $\omega = 1$ 时, *SOR* 迭代法退化为 G-E 迭代法, $\omega > 1$ 时称为超松弛, $\omega < 1$ 时称为低松弛。

除了上面几种方法之外, 还有一种著名的迭代法是 *Richardson* 迭代法, 这种方法取 $M = \alpha I$, $N = \alpha I - A$, 迭代格式为

$$\alpha x^{(k+1)} = (\alpha I - A)x^{(k)} + b,$$

或者等价地

$$x^{(k+1)} = x^{(k)} + \alpha(b - Ax^{(k)}),$$

这种方法的优势在无需求解线性系统, 但是其收敛速度较慢, 因此一般不作为主要的迭代方法。

1.2 收敛性分析

上一小节中的所有方法 (单步线性定常迭代法) 都具有形如

$$x^{(k+1)} = Gx^{(k)} + f \tag{1.1}$$

的更新格式, 其中 $G = M^{-1}N$, 其中 G 称作相应方法的迭代矩阵, f 称作相应方法的常数项。本小节我们对这类方法的收敛性进行分析。

设 $e^{(k)} = x^{(k)} - x^*$ 是第 k 步迭代的误差向量, 其中 x^* 是真实解。如果某一种迭代法满足

$$\lim_{k \rightarrow \infty} e^{(k)} = 0,$$

则称该方法是收敛的。根据迭代格式(1.1), 有

$$e^{(k+1)} = Ge^{(k)},$$

进而

$$e^{(k)} = G^k e^{(0)},$$

所以对方法收敛性的分析转化为对迭代矩阵性质的分析, 于是不难证明如下的定理。

Theorem 1. 给定某种形如(1.1)的迭代法和起始向量, G 是迭代矩阵, 则迭代格式(1.1)收敛当且仅当

$$G^k \rightarrow 0 \quad \text{as } k \rightarrow \infty,$$

而这当且仅当 $\rho(G) < 1$, 所以对于单步线性定常迭代法, 收敛的充要条件为迭代矩阵的谱半径小于 1。

Hint: 要证明 $G^k \rightarrow 0 \iff \rho(G) < 1$, 充分性 (\Leftarrow) 可以考虑矩阵的 Jordan 分解中的各 Jordan 块矩阵的幂, 或者说明 $\|G^k\| \rightarrow 0$, 而这需要用到 $\rho(G) \leq \|G^k\|^{1/k} \leq \|G\|$ 对任意相容矩阵范数都成立, 并且对任意的 $\epsilon > 0$, 都存在算子范数使得 $\rho(G) \geq \|G\| - \epsilon$ (数值线性代数, Th 2.1.6); 而必要性 (\Rightarrow) 则使用零算子的定义, 将 G^k 作用在各特征向量上即可。

不过, 对于实际问题, 我们往往无法直接计算迭代矩阵的谱半径, 好在因为 $\rho(G)$ 可以被 $\|G\|$ 控制, 而矩阵有一些相容范数相对容易计算, 因此我们可以引入如下实用的判定定理。

Theorem 2. 给定某种形如(1.1)的迭代法, G 是迭代矩阵, 如果 $\|G\| = q < 1$ ($\|\cdot\|$ 是某相容的矩阵范数), 则

$$\|e^{(k)}\| \leq \frac{q^k}{1-q} \|x^{(1)} - x^{(0)}\|, \quad (1.2)$$

并且

$$\|e^{(k)}\| \leq \frac{q}{1-q} \|x^{(k)} - x^{(k-1)}\|. \quad (1.3)$$

Hint: 注意到 $e^{(k)} = G^k e^{(0)}$, 而

$$\|e^{(0)}\| \leq \|x^{(1)} - x^{(0)}\| + \|x^{(1)} - x^*\| = \|x^{(1)} - x^{(0)}\| + \|e^{(1)}\| \leq \|x^{(1)} - x^{(0)}\| + q\|e^{(0)}\|,$$

所以 $\|e^{(0)}\| \leq \|x^{(1)} - x^{(0)}\|/(1-q)$ 。类似地, $e^{(k)} = Ge^{(k-1)}$, 并且

$$\|e^{(k-1)}\| \leq \|x^{(k)} - x^{(k-1)}\| + \|x^{(k)} - x^*\| = \|x^{(k)} - x^{(k-1)}\| + \|e^{(k)}\| \leq \|x^{(k)} - x^{(k-1)}\| + q\|e^{(k-1)}\|,$$

因此定理的结论成立。

根据如上定理, 通过 q 和 $\|x^{(1)} - x^{(0)}\|$ 或者 $\|x^{(k)} - x^{(k-1)}\|$ 可以来估计误差向量, 从而判断迭代过程是否应该终止。接下来我们分别进一步分析这三种方法的收敛性。

1.2.1 Jacobi 迭代和 Gauss-Seidel 迭代的收敛性

对于 Jacobi 迭代而言, 迭代矩阵为 $G = D^{-1}(L+U)$, 由于其中的 D 是对角阵, 因此可以直接得到 D^{-1} 并计算得到矩阵 G , 但是对于 G-E 迭代而言, 迭代矩阵为 $G = (D-L)^{-1}U$, 需要计算一个上三角矩阵的逆, 因此直接计算 $\|(D-L)^{-1}U\|$ 也比较困难, 为了简化计算, 我们使用如下的定理来估计误差, 其中我们使用 Jacobi 迭代的迭代矩阵来控制 G-E 的迭代矩阵。

Theorem 3. 设 Jacobi 迭代和 G-E 迭代的迭代矩阵分别为 G^{JA} 和 G^{GE} , 给定某起始向量 $x^{(0)}$, 则由 G-E 迭代产生的近似解 $x^{(k)}$ 与真实解 x^* 的误差满足

- 使用矩阵无穷范数时, 如果 $\|G^{JA}\|_{\infty} \leq 1$, 则 $\|G^{GE}\|_{\infty} < 1$, 并且

$$\|x^{(k)} - x^*\|_{\infty} \leq \frac{\mu^k}{1-\mu} \|x^{(1)} - x^{(0)}\|_{\infty}, \quad (1.4)$$

其中

$$\mu = \max_i \frac{\sum_{j=i+1}^n |g_{ij}^{JA}|}{1 - \sum_{j=1}^i |g_{ij}^{JA}|} \leq \|G^{JA}\|_{\infty} < 1.$$

- 当使用矩阵 1 范数时, 如果 $\|G^{JA}\|_1 \leq 1$, 则 $\rho(G^{GE}) < 1$, 并且

$$\|x^{(k)} - x^*\|_1 \leq \frac{\tilde{\mu}^k}{(1 - \tilde{\mu})(1 - s)} \|x^{(1)} - x^{(0)}\|_1, \quad (1.5)$$

其中

$$s = \max_j \sum_{i=j+1}^n |g_{ij}^{JA}|, \quad \tilde{\mu} = \max_j \frac{\sum_{i=1}^{j-1} |g_{ij}^{JA}|}{1 - \sum_{i=j+1}^n |g_{ij}^{JA}|} \leq \|G^{JA}\|_1 < 1.$$

Hint: 当使用无穷范数时, 注意到存在范数为 1 的 x 使得 $\|G^{GE}\|_\infty = \|G^{GE}x\|_\infty$, 记 $y = G^{GE}x$, $|y_i| = \|G^{GE}\|$, 根据 G-E 的迭代格式 $G^{GE} = (D - L)^{-1}U$, 于是

$$(D - L)y = Ux \implies y = D^{-1}Ly + D^{-1}Ux,$$

注意到 $G^{JA} = D^{-1}(L + U)$, 所以上式的分量形式为

$$y_i = \frac{1}{a_{ii}} \left(\sum_{j < i} a_{ij} y_j + \sum_{j > i} a_{ij} x_j \right) = \sum_{j=1}^{i-1} g_{ij}^{JA} y_j + \sum_{j=i+1}^n g_{ij}^{JA} x_j,$$

进而对上式两侧取绝对值可得

$$\|G^{GE}\|_\infty = \|y\|_\infty = |y_i| \leq \frac{1}{a_{ii}} \left(\sum_{j < i} |a_{ij}| |y_j| + \sum_{j > i} |a_{ij}| |x_j| \right) \leq \|y\|_\infty \sum_{j=1}^{i-1} |g_{ij}^{JA}| + \sum_{j=i+1}^n |g_{ij}^{JA}|,$$

最后一个不等号中用到了 $\|x\|_\infty = 1$. 由上式可得

$$\|G^{GE}\|_\infty \leq \frac{\sum_{j=i+1}^n |g_{ij}^{JA}|}{1 - \sum_{j=1}^{i-1} |g_{ij}^{JA}|} \leq \mu.$$

又注意到对于任意 $a + b < 1$ ($a, b > 0$) 都有

$$\max \left(\frac{a}{1-b}, \frac{b}{1-a} \right) \leq a + b,$$

令 $a = \sum_{j=1}^{i-1} |g_{ij}^{JA}|$, $b = \sum_{j=i+1}^n |g_{ij}^{JA}|$ 可得 $\mu \leq \|G^{JA}\|_\infty < 1$. 最后, 将 $\|G^{GE}\|_\infty \leq \mu$ 带入定理2中即可得到误差估计。至于使用 1 范数的情况, 与上述证明类似, 需要证明 $\rho(G^{GE}) \leq \tilde{\mu} \leq \|G^{JA}\|_1 < 1$, 但是因为不再具有 $\|G^{GE}\|_1$ 的估计所以无法直接使用定理2, 这使得误差估计相较于无穷范数的情况则更加复杂, 见数值线性代数 Th 4.2.5。

自然地, 系数矩阵 A 的性质越好, 迭代法的收敛性质就应该越好, 下面我们来特别讨论一下系数矩阵是正定矩阵的情况, 有如下结论。

Theorem 4. 对于 $Ax = b$ 的迭代求解问题, 我们有如下结论:

1. 如果 $A \in S$ 并且各对角元 $a_{ii} > 0$, 则 Jacobi 迭代收敛当且仅当 $A, 2D - A \in S_{++}$;
2. 如果 $A \in S_{++}$, 则 G-E 迭代收敛。

一种容易判别的正定矩阵是对角占优矩阵。如果 n 阶矩阵 A 满足

$$|a_{ii}| \geq \sum_{j \neq i} |a_{ij}|, \quad i = 1, 2, \dots, n,$$

则称 A 是 (行) 对角占优矩阵, 如果上式中的不等号对于至少一个 i 是严格的, 则称 A 是弱严格 (行) 对角占优矩阵, 如果对于所有 i 都是严格的, 则称 A 是严格 (行) 对角占优矩阵。矩阵的另一种重要性质是可约性。如果关于矩阵 A 存在置换矩阵 P 使得

$$PAP^T = \begin{bmatrix} A_{11} & 0 \\ A_{21} & A_{22} \end{bmatrix},$$

其中 A_{11} 和 A_{22} 均是方阵, 则称 A 是可约矩阵或可分矩阵, 否则称 A 是不可约矩阵或不可分矩阵。该定义等价于记 $\mathcal{W} = \{1, 2, \dots, n\}$, 如果 \mathcal{W} 存在两个非空子集 $\mathcal{T}_1, \mathcal{T}_2$ 满足

$$\mathcal{T}_1 \cap \mathcal{T}_2 = \emptyset, \quad \mathcal{T}_1 \cup \mathcal{T}_2 = \mathcal{W},$$

使得

$$a_{ij} = 0, \quad i \in \mathcal{T}_1, j \in \mathcal{T}_2,$$

则称矩阵 A 是可约的, 通常会考察 $\mathcal{T}_1 = \{i : |x_i| = 1\}$ 和 $\mathcal{T}_2 = \{j : |x_j| < 1\}$, 其中 x 满足 $\|x\|_\infty = 1$ 是 A 的一个零向量。如果 A 是可约的, 则 $Ax = b$ 等价于两个较小的线性系统。如果矩阵同时是不可约和弱严格对角占优的, 则称 A 是不可约对角占优矩阵。使用反证法可以证明, 所有的严格对角占优和不可约对角占优矩阵都是可逆矩阵。这部分的结论用如下定理总结。

Theorem 5. 如下关于正定性、可约性、和对角占优性的结论成立:

1. 严格对角占优和不可约对角占优矩阵都非奇异;
2. 对角元均为正数的严格对角占优和不可约对角占优矩阵是正定矩阵;
3. 对于以严格对角占优和不可约对角占优矩阵为系数矩阵的线性方程组, *Jacobi* 迭代和 *G-E* 迭代收敛。

Hint: 第一条结论使用反证法, 在证明不可约对角占优矩阵非奇异时需要用到不可约的等价定义, 令 $\mathcal{T}_1 = \{i : |x_i| = 1\}$ 和 $\mathcal{T}_2 = \{j : |x_j| < 1\}$, 其中 x 满足 $\|x\|_\infty = 1$ 是 A 的一个零向量。第二条结论需要注意到任意 $\lambda \leq 0$, $A - \lambda I$ 都还是严格对角占优或者不可约对角占优矩阵, 因此根据第一条结论总是可逆的。第三条结论首先要根据严格对角占优和不可约对角占优矩阵的性质证明 D 可逆, 然后注意到

$$\lambda I - G^{JA} = D^{-1}(\lambda D - L - U),$$

而当 $|\lambda| \geq 1$ 时, $\lambda D - L - U$ 是严格对角占优或者不可约对角占优矩阵, 因此可逆, 进而 $\lambda I - G^{JA}$ 可逆, 所以 G^{JA} 没有绝对值大于 1 的特征值, 从而 *Jacobi* 迭代收敛。对于 *G-E* 迭代, 类似地有

$$\lambda I - G^{GE} = (D - L)^{-1}(\lambda D - \lambda L - U),$$

同样地, $\lambda D - \lambda L - U$ 是严格对角占优或者不可约对角占优矩阵, 所以 G^{GE} 没有绝对值大于 1 的特征值, 由此得到收敛性。

1.2.2 SOR 迭代的收敛性

SOR 的迭代矩阵为 $G^\omega = (D - \omega L)^{-1}[(1 - \omega)D + \omega U]$, 相比于 *G-E* 迭代, SOR 迭代的收敛性是类似的但是要更加复杂。和前两种方法相同, SOR 收敛的充要条件是 $\rho(G^\omega) < 1$, 为了满足该条件, 一个必要条件是松弛因子 ω 的取值范围为 $0 < \omega < 2$, 这是因为如果 $\rho(G^\omega) < 1$, 则

$$\det(G^\omega) = \prod_{i=1}^n |\lambda_i| < 1,$$

而

$$\begin{aligned} \det(G^\omega) &= \det((D - \omega L)^{-1}[(1 - \omega)D + \omega U]) \\ &= \det(I - D^{-1}\omega L)^{-1} \cdot \det[(1 - \omega)I + \omega D^{-1}U] \\ &= 1 \cdot (1 - \omega)^n \end{aligned}$$

因此 $(1 - \omega)^n < 1$, 所以 $0 < \omega < 2$ 。此外, 和前两种方法一样, SOR 迭代的收敛性还与系数矩阵的性质有关, 有如下定理。

Theorem 6. 对于 $Ax = b$ 的迭代求解问题, 我们有如下结论:

1. 如果系数矩阵 A 是严格对角占优或不可约对角占优的, 并且 $\omega \in (0, 1)$, 则 *SOR* 迭代收敛;
2. 如果 $A \in S_{++}$, 并且 $\omega \in (0, 2)$, 则 *SOR* 迭代收敛。

1.2.3 收敛速度

下面我们来考虑收敛的迭代法的收敛速度。一般评价一种算法的收敛速度有两个指标, 分别是平均收敛速度和渐进收敛速度。对于单步定常迭代法, 误差满足

$$e^{(k)} = G e^{(k-1)} = G^k e^{(0)},$$

所以

$$\|e^{(k)}\| \leq \|G^k\| \|e^{(0)}\|,$$

由于准确解 x^* 未知, 因此我们无法直接计算 $e^{(0)}$, 一般使用 $\|G^k\|$ 的大小来刻画迭代的速度, 定义

$$R_k(G) = \frac{-\ln \|G^k\|}{k} \quad (1.6)$$

为 k 次迭代的平均收敛速度。

1.3 误差分析

2 现代迭代法——空间投影

2.1 最速下降法

2.2 共轭梯度法

2.3 Krylov 方法