第三讲 定常迭代法

目录

3.1 定常迭代法

3.2 收敛性分析

3.3 正则分裂

3.4 交替方向迭代法

3.5 迭代法的加速

线性方程组的数值求解

- **直接法**
  - LU 分解, Cholesky 分解, ...
- **迭代法**
  - 定常 (经典, 基本) 迭代法: Jacobi, Gauss-Seidel, SOR, SSOR, ...
  - 现代 (Krylov 子空间) 迭代法: CG, MINRES, GMRES, ...
- **快速算法** (基于特殊结构和性质)
  - 基于各类快速变换, 如 FFT, DCT, DST, ...
  - 代数多重网格法 (Algebraic multigrid)
  - 快速多极子算法 (Fast multipole)
  - Hierarchical Matrices, ...

!

- 有些方法可能只是对某类方程有效, 比如快速算法.
- 在实际应用中, 这些方法经常结合使用, 如混合算法, 预处理算法等.

本讲主要介绍定常 (经典, 基本) 迭代法

更多迭代方法可参见:

Templates for the Solution of Linear Systems: Building Blocks for Iterative Methods, 1994.

3-1 定常迭代法

3.1 定常迭代法

3.1.1 矩阵分裂与迭代法

3.1.2 经典迭代法 (Jacobi, G-S, SOR, SSOR, AOR)

3.1.3 Richardson 迭代法

3.1.4 分块迭代法

什么是迭代法

当直接求解方程组 $Ax = b$ 较困难时, 我们可以求解一个近似方程组

$$\boxed{Mx = b}$$

其中 M 是 A 的某个近似, 且该方程组较容易求解.

设其解为 $x^{(1)}$. 易知它与真解之间的误差满足

$$\boxed{A(x_* - x^{(1)}) = b - Ax^{(1)}}$$

如果 $x^{(1)}$ 已经满足精度要求, 则停止计算, 否则需要通过修正来满足精度要求.

怎么修正近似解

设修正量为 Δx. 显然最佳的修正量 Δx 应该满足方程 $\boxed{A\Delta x = b - Ax^{(1)}}$.
但由于直接求解该方程比较困难, 因此我们还是求解近似方程组

$$M\Delta x = b - Ax^{(1)}.$$

于是得到修正后的近似解

$$\boxed{x^{(2)} = x^{(1)} + \Delta x = x^{(1)} + M^{-1}(b - Ax^{(1)})}$$

若 $x^{(2)}$ 已经满足精度要求, 则停止计算, 否则继续按以上的方式进行修正.

定常迭代法

不断重复以上步骤，于是，我们就得到一个序列

$$x^{(1)}, x^{(2)}, \ldots, , x^{(k)}, \ldots.$$

满足以下递推关系

$$\boxed{x^{(k+1)} = x^{(k)} + M^{-1}(b - Ax^{(k)})}, \quad k = 1, 2, \ldots.$$

由于每次迭代的格式是一样的，因此称为 定常迭代法 .

构造定常迭代法的基本准则

构造一个好的定常迭代法通常需要考虑以下两点:

(1) 以 M 为系数矩阵的线性方程组必须比原线性方程组更容易求解;

(2) M 应该是 A 的一个很好的近似, 或者迭代序列 $\{x^{(k)}\}$ 要收敛 (越快越好).

本讲主要介绍以下定常迭代法 (基于 **矩阵分裂**):

- Jacobi, Gauss-Seidel, SOR (Successive Over-Relaxation)

  SSOR (Symmetric SOR), AOR (Accelerated over-relaxation), SAOR
- Richardson 算法
- 分块迭代算法
- 交替方向算法

迭代法基本思想

给定一个迭代初始值 $x^{(0)}$, 通过一定的**迭代格式**生成一个迭代序列

$$x^{(1)}, x^{(2)}, x^{(3)}, \ldots, x^{(k)}, \ldots$$

使得

$$\boxed{\lim_{k\to\infty} x^{(k)} = x_* \triangleq A^{-1}b}$$

迭代法一般格式

一般来说, 迭代格式可表示为

$$x^{(k+1)} = \boldsymbol{\varphi_k}\left(x^{(k)}, x^{(k-1)}, \ldots, x^{(1)}, x^{(0)}, A, b\right), \quad k = 0, 1, 2, \ldots,$$

其中 $x^{(0)} = \varphi_0(A, b)$, 或者 $x^{(0)}$ 直接给出. 这里 φ_k 就称为**迭代函数**.

- 若 φ_k 都是线性的, 则称为**线性迭代法**, 否则为**非线性迭代法**
- 若存在正整数 ℓ, 当 $k \geq \ell$ 时, φ_k 与 k 无关, 则称为**定常迭代法**, 否则是**非定常**的
- 对于定常迭代, 有 $\varphi_\ell = \varphi_{\ell+1} = \cdots \triangleq \varphi$, 此时 $x^{(k+1)}$ 只与 $x^{(k)}, x^{(k-1)}, \ldots, x^{(k-\ell+1)}$ 有关. 若 $x^{(k+1)}$ 只与 $x^{(k)}$ 有关, 则称为**单步迭代**, 否则就称为**多步迭代**.

本讲主要关注基于矩阵分裂的单步线性迭代方法 (简称矩阵分裂迭代法)

3-1-2 矩阵分裂与迭代法

定义 (矩阵分裂 Matrix splitting) 设 $A \in \mathbb{R}^{n\times n}$ 非奇异, 称

$$A = M - N$$

为 A 的一个矩阵分裂, 其中 M 非奇异.

原方程组等价于 $Mx = Nx + b$. 于是我们就可以构造迭代格式

$$\boxed{x^{(k+1)} = M^{-1}Nx^{(k)} + M^{-1}b \triangleq Gx^{(k)} + g}, \quad k = 0, 1, \ldots,$$

其中 $G = M^{-1}N$ 称为该迭代格式的**迭代矩阵**.

迭代矩阵对算法的收敛性和收敛速度起着决定性作用.

3-1-3 经典迭代法

将矩阵 A 分裂为

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

其中 D 为 A 的对角线部分, $-L$ 和 $-U$ 分别为 A 的严格下三角和严格上三角部分.

$$D = \begin{bmatrix} a_{11} & & & \\ & a_{22} & & \\ & & \ddots & \\ & & & a_{nn} \end{bmatrix}, \quad L = -\begin{bmatrix} 0 & & & \\ a_{21} & \ddots & & \\ \vdots & \ddots & \ddots & \\ a_{n1} & \cdots & a_{n,n-1} & 0 \end{bmatrix}, \quad U = -\begin{bmatrix} 0 & a_{12} & \cdots & a_{1n} \\ & \ddots & \ddots & \vdots \\ & & \ddots & a_{n-1,n} \\ & & & 0 \end{bmatrix}.$$

本小节介绍的经典迭代法包括: Jacobi, G-S, SOR, SSOR, AOR, SAOR

Jacobi 迭代法

- 取 $M=D,\ N=L+U$, 则可得 **Jacobi 迭代法**

$$\boxed{x^{(k+1)} = D^{-1}(L+U)x^{(k)} + D^{-1}b}\,, \qquad k=0,1,2,\ldots.$$

迭代矩阵: $\boxed{G_{\mathrm{J}} = D^{-1}(L+U)}$

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

算法 1 求解线性方程组的 Jacobi 迭代法

Choose an initial guess $x^{(0)}$
while not converge **do**
  for $i = 1$ to n **do**
    $x_i^{(k+1)} = \left(b_i - \sum\limits_{j=1, j\neq i}^{n} a_{ij} x_j^{(k)} \right) \Big/ a_{ii}$
  end for
end while

由于 Jacobi 迭代中 $x_i^{(k+1)}$ 的更新顺序与 i 无关, 因此非常适合并行计算.

Gauss-Seidel 迭代法

- 取 $M = D - L$, $N = U$, 即可得 **Gauss-Seidel (G-S) 迭代法**

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

迭代矩阵: $\boxed{G_{\mathrm{GS}} = (D-L)^{-1} U}$

将 G-S 迭代改写为 $Dx^{(k+1)} = Lx^{(k)+1} + Ux^{(k)} + b$, 可得分量形式

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

G-S 算法的主要优点是充分利用了已经获得的最新数据

算法 2 求解线性方程组的 G-S 迭代法

```
Choose an initial guess x^(0)
while not converge do
    for i = 1 to n do
        x_i^(k+1) = (1/a_ii) ( b_i - Σ_{j=1}^{i-1} a_ij x_j^(k+1) - Σ_{j=i+1}^{n} a_ij x_j^(k) )
    end for
end while
```

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

> G-S 算法中未知量的更新是按自然顺序进行的，不适合并行计算

SOR 迭代法

在 G-S 的基础上, 通过引入一个松弛参数 ω 来加快收敛速度:

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

这就是 **SOR 迭代法**, 整理后即为

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

其中 ω 为**松弛参数**.

(1) 当 $\omega = 1$ 时, SOR 即为 G-S;

(2) 当 $\omega < 1$ 时, 称为**低松弛 (under relaxation)**方法;

(3) 当 $\omega > 1$ 时, 称为**超松弛 (over relaxation)**方法.

SOR 的迭代矩阵: $$\boxed{G_{\mathrm{SOR}} = (D-\omega L)^{-1}\big((1-\omega)D+\omega U\big),}$$

对应的矩阵分裂: $M = \dfrac{1}{\omega}D - L, \quad N = \dfrac{1-\omega}{\omega}D + U.$

分量形式:

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

$$i = 1, 2, \ldots, n.$$

算法 3 求解线性方程组的 SOR 迭代法

Choose an initial guess $x^{(0)}$ and parameter ω
while not converge **do**
  for $i = 1$ to n **do**
    $$x_i^{(k+1)} = (1-\omega)x_i^{(k)} + \frac{\omega}{a_{ii}}\left(b_i - \sum_{j=1}^{i-1} a_{ij}x_j^{(k+1)} - \sum_{j=i+1}^{n} a_{ij}x_j^{(k)}\right)$$
  end for
end while

SOR 最大的优点是引入了 ω, 通过选取适当的 ω 可以大大提高收敛速度

但 SOR 最大的难点就是如何选取最优参数 ω_{opt}

SOR 迭代曾经在很长一段时间内是科学计算中求解线性方程组的首选方法

SSOR 迭代法

将 SOR 算法中的 L 和 U 相交换, 即可得迭代格式

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

将其与 SOR 相结合, 交替迭代, 得到下面的两步迭代法

$$\begin{cases} x^{(k+\frac{1}{2})} = (D-\omega L)^{-1}\big((1-\omega)D+\omega U\big)x^{(k)} + \omega(D-\omega L)^{-1}b \\ x^{(k+1)} = (D-\omega U)^{-1}\big((1-\omega)D+\omega L\big)x^{(k+\frac{1}{2})} + \omega(D-\omega U)^{-1}b \end{cases}$$

这就是 **SSOR 迭代法**, 相当于将 L 与 U 同等看待, 交替做两次 SOR 迭代.

消去中间迭代向量 $x^{(k+\frac{1}{2})}$, 可得

$$x^{(k+1)} = G_{\mathrm{SSOR}} x^{(k)} + g,$$

其中迭代矩阵

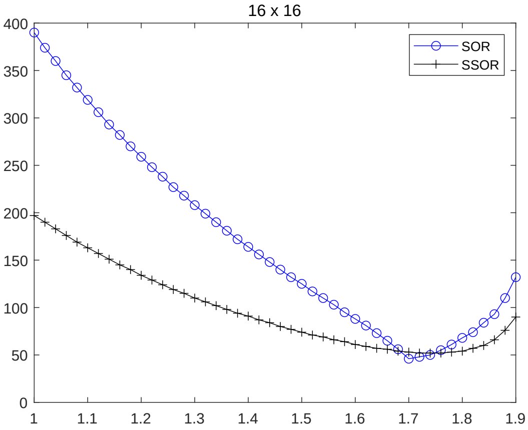
$$\boxed{G_{\mathrm{SSOR}} = (D - \omega U)^{-1} \big[(1-\omega)D + \omega L\big] (D - \omega L)^{-1} \big[(1-\omega)D + \omega U\big]}$$

对应的矩阵分裂:

$$\begin{aligned} M &= \frac{1}{\omega(2-\omega)} \big(D - \omega(L+U) + \omega^2 L D^{-1} U\big) \\ &= \frac{1}{\omega(2-\omega)} (D - \omega L) D^{-1} (D - \omega U), \\ N &= \frac{1}{\omega(2-\omega)} \big((1-\omega)D + \omega L\big) D^{-1} \big((1-\omega)D + \omega U\big). \end{aligned}$$

关于 SSOR 的几点说明

- 对于某些特殊问题, SOR 算法不收敛, 但仍然可能构造出收敛的 SSOR 算法.
- 一般来说, SOR 算法的渐进收敛速度对参数 ω 更加敏感. (Poisson_SOR_SSOR.m)



AOR 迭代法

AOR (Accelerated over-relaxation) 由 Hadjidimos 于 1978 年提出, 迭代矩阵为

$$\boxed{G_{\mathrm{AOR}} = (D-\gamma L)^{-1}\big((1-\omega)D + (\omega-\gamma)L + \omega U\big)},$$

其中 γ 和 ω 为松弛参数. 对应的矩阵分解为

$$M = \frac{1}{\omega}(D-\gamma L), \quad N = \frac{1}{\omega}\big((1-\omega)D + (\omega-\gamma)L + \omega U\big).$$

(1) 当 $\gamma = \omega$ 时, AOR 即为 SOR;

(2) 当 $\gamma = \omega = 1$ 时, AOR 即为 G-S;

(3) 当 $\gamma = 0$, $\omega = 1$ 时, AOR 即为 Jacobi.

与 SSOR 类似, 我们也可以定义 SAOR 迭代法.

3-1-4 Richardson 迭代法

Richardson 迭代法是一类形式非常简单的算法, 其迭代格式为

$$\boxed{x^{(k+1)} = x^{(k)} + \omega(b - Ax^{(k)})}, \quad k = 0, 1, 2, \ldots.$$

对应的矩阵分裂和迭代矩阵分别为

$$M = \frac{1}{\omega} I, \quad N = \frac{1}{\omega} I - A, \quad G_{\mathrm{R}} = I - \omega A.$$

如果在每次迭代时取不同的参数, 即

$$x^{(k+1)} = x^{(k)} + \omega_k(b - Ax^{(k)}), \quad k = 0, 1, 2, \ldots,$$

则称为 nonstationary Richardson 算法.

定理 设 $A \in \mathbb{R}^{n\times n}$ 是对称正定矩阵, λ_1 和 λ_n 分别是 A 的最大和最小特征值, 则 Richardson 算法收敛当且仅当

$$0 < \omega < \lambda_1^{-1}.$$

最优参数为

$$\omega_* = \arg\min_{\omega} \rho(G_{\mathrm{R}}) = \frac{2}{\lambda_1 + \lambda_n},$$

即当 $\omega = \omega_*$ 时, 迭代矩阵的谱半径达到最小, 且有

$$\rho(G_{\mathrm{R}}) = \begin{cases} 1 - \omega\lambda_n & \text{if} \quad \omega \le \omega_* \\ \dfrac{\lambda_1 - \lambda_n}{\lambda_1 + \lambda_n} = \dfrac{\kappa(A) - 1}{\kappa(A) + 1} & \text{if} \quad \omega = \omega_* \\ \omega\lambda_1 - 1 & \text{if} \quad \omega \ge \omega_*. \end{cases}$$

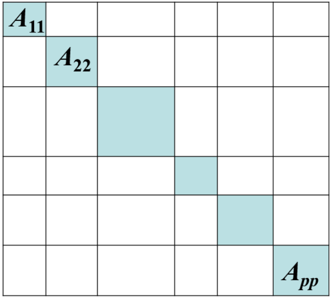
(留作课外自习)

3-1-5 分块迭代法

在实际使用中, 为了**扩大定常迭代法的使用范围**, 同时**提升算法的计算效率**, 通常会采用分块形式进行迭代.

将 A 写成如下的分块形式:

$$A = \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1p} \\ A_{21} & A_{22} & \cdots & A_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ A_{p1} & A_{p2} & \cdots & A_{pp} \end{bmatrix}.$$



分块 Jacobi 迭代

$$A_{ii}\boldsymbol{x}_i^{(k+1)} = \boldsymbol{b}_i - \sum_{j=1, j\neq i}^{p} A_{ij}\boldsymbol{x}_j^{(k)}, \quad i = 1, 2, \ldots, p.$$

分块 Gauss-seidel 迭代

$$A_{ii}\boldsymbol{x}_i^{(k+1)} = \boldsymbol{b}_i - \sum_{j=1}^{i-1} A_{ij}\boldsymbol{x}_j^{(k+1)} - \sum_{j=i+1}^{p} A_{ij}\boldsymbol{x}_j^{(k)}, \quad i = 1, 2, \ldots, p.$$

分块 SOR 迭代

$$\boldsymbol{x}_i^{(k+1)} = \boldsymbol{x}_i^{(k)} + \omega A_{ii}^{-1}\left(\boldsymbol{b}_i - \sum_{j=1}^{i} A_{ij}\boldsymbol{x}_j^{(k+1)} - \sum_{j=i+1}^{p} A_{ij}\boldsymbol{x}_j^{(k)}\right),$$

$$i = 1, 2, \ldots, p.$$

第三讲 定常迭代法

目录

3.1 定常迭代法

3.2 收敛性分析

3.3 正则分裂

3.4 交替方向迭代法

3.5 加速技巧

3-2 收敛性分析

3.2 收敛性分析

3.2.1 向量与矩阵序列收敛基本概念

3.2.2 定常迭代法的收敛性

3.2.3 迭代矩阵非负情形

3.2.4 不可约对角占优矩阵

3.2.5 对称正定矩阵

3.2.6 相容次序矩阵

3-2-1 向量与矩阵序列收敛基本概念

设 $\{x^{(k)}\}_{k=1}^{\infty}$ 是 $\mathbb{R}^n$ 中的一个向量序列, 如果存在 $x = [x_1, x_2, \ldots, x_n]^\intercal \in \mathbb{R}^n$, 使得

$$\lim_{k\to\infty} x_i^{(k)} = x_i, \quad i = 1, 2, \ldots, n,$$

则称 $x^{(k)}$ **(按分量) 收敛到** x, 即 x 为 $x^{(k)}$ 的极限, 记为 $\boxed{\lim_{k\to\infty} x^{(k)} = x}$.

定理 (向量序列收敛性 基本定理) 设 $\|\cdot\|$ 是 $\mathbb{R}^n$ 上的任意一个向量范数, 则 $\lim_{k\to\infty} x^{(k)} = x$ 的充要条件是

$$\lim_{k\to\infty} \|x^{(k)} - x\| = 0.$$

(留作课外自习, 根据范数的等价性, 只需证明对无穷范数成立即可)

矩阵序列的收敛

设 $\{A^{(k)}\}_{k=0}^{\infty}$ 是 $\mathbb{R}^{m\times n}$ 中的一个矩阵序列, 如果存在矩阵 $A=[a_{ij}]\in\mathbb{R}^{m\times n}$ 使得

$$\lim_{k\to\infty} a_{ij}^{(k)} = a_{ij}, \quad i=1,2,\ldots,m,\ j=1,2,\ldots,n,$$

则称 $A^{(k)}$ **(按分量) 收敛到** A, 即 A 是 $A^{(k)}$ 的极限, 记为 $\boxed{\lim_{k\to\infty} A^{(k)} = A}$.

定理 (矩阵序列收敛性 基本定理) 设矩阵序列 $\{A^{(k)}\}_{k=0}^{\infty}\subset\mathbb{R}^{m\times n}$, 则 $\lim_{k\to\infty} A^{(k)} = A$ 的充要条件是

$$\lim_{k\to\infty} \|A^{(k)}-A\| = 0,$$

其中 $\|\cdot\|$ 为任一矩阵范数. (留作课外自习)

特例: $\lim\limits_{k\to\infty} A^{(k)} = 0$

- $\lim\limits_{k\to\infty} A^{(k)} = 0$ 当且仅当 $\lim\limits_{k\to\infty} \|A^{(k)}\| = 0$

- $\lim\limits_{k\to\infty} A^{(k)} = 0$ 当且仅当 $\lim\limits_{k\to\infty} A^{(k)}x = 0, \forall\, x \in \mathbb{R}^n$

引理 (谱半径与范数的关系) 设 $A \in \mathbb{R}^{n\times n}$, 则

(1) 对任意矩阵范数, 有 $\rho(A) \leq \|A\|$;

(2) 反之, 对任意 $\varepsilon > 0$, 都存在一个算子范数 $\|\cdot\|_\varepsilon$, 使得 $\|A\|_\varepsilon \leq \rho(G) + \varepsilon$, 所以, 若 $\rho(A) < 1$, 则存在算子范数 $\|\cdot\|_\varepsilon$, 使得 $\|A\|_\varepsilon < 1$;

(板书)

特例: $\lim\limits_{k\to\infty} A^k = 0$

- $\lim\limits_{k\to\infty} A^k = 0$ 当且仅当 $\lim\limits_{k\to\infty} \|A^k\| = 0$
- $\lim\limits_{k\to\infty} A^k = 0$ 当且仅当 $\rho(A) < 1.$
- $\lim\limits_{k\to\infty} A^k = 0$ 充要条件是存在某个矩阵范数 $\|\cdot\|$ 使得 $\|A\| < 1$

(板书)

下面的结论是谱半径与算子范数之间的一个非常重要的性质.

引理 设 $G \in \mathbb{R}^{n\times n}$, 则对任意算子范数 $\|\cdot\|$, 有

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

(板书)

3-2-2 定常迭代法的收敛性

考虑迭代法

$$x^{(k+1)} = M^{-1}Nx^{(k)} + M^{-1}b \triangleq Gx^{(k)} + g \tag{3.1}$$

定义 (迭代法的收敛性) 设 $\{x^{(k)}\}_{k=1}^{\infty}$ 是由定常迭代生成的向量序列, 如果对任意的初始向量 $x^{(0)}$, 都有

$$\lim_{k\to\infty} x^{(k)} \to x_*,$$

则称迭代格式 (3.1) 是收敛的, 否则就称其为发散的.

收敛性判断

首先给出一个迭代法收敛的充分条件.

引理 若存在算子范数 $\|\cdot\|$, 使得 $\|G\| < 1$, 则迭代格式 (3.1) 收敛. (板书)

定理 (收敛性定理) 对任意初始向量 $x^{(0)}$, 迭代格式 (3.1) 收敛的充要条件是 $\rho(G) < 1$.

(板书)

收敛速度

定义 设 G 是迭代矩阵, 则迭代格式 (3.1) 的平均收敛速度定义为

$$R_k(G) \triangleq -\ln \|G^k\|^{\frac{1}{k}},$$

渐进收敛速度定义为

$$R(G) \triangleq \lim_{k\to\infty} R_k(G) = -\ln \rho(G).$$

平均收敛速度与迭代步数和范数有关, 但渐进收敛速度只依赖于迭代矩阵谱半径

收敛估计

定理 考虑算法 3.1. 如果存在某个算子范数 $\|\cdot\|$ 使得 $\|G\| = q < 1$, 则

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

(2) $\|x^{(k)} - x_*\| \leq \dfrac{q}{1-q}\|x^{(k)} - x^{(k-1)}\|$;

(3) $\|x^{(k)} - x_*\| \leq \dfrac{q^k}{1-q}\|x^{(1)} - x^{(0)}\|$.

(板书)

✍ 由此可见, 一个好的迭代格式 (基于矩阵分裂的单步定常迭代法) 应该满足:

(1) 以 M 为系数矩阵的线性方程组比较容易求解;

(2) $\rho(G)$ 尽可能小.

3-2-3 迭代矩阵非负情形

我们分情形讨论经典迭代法 (Jacobi, G-S 和 SOR) 的收敛性.

记

$$\tilde{L} = D^{-1}L,\ \tilde{U} = D^{-1}U$$

- Jacobi 迭代矩阵可写为

$$G_{\mathrm{J}} = D^{-1}(L+U) = \tilde{L} + \tilde{U},$$

其对角线全部为零.

本小节考虑 G_{J} 非负 (即 $\tilde{L}$ 和 $\tilde{U}$ 都非负) 时 Jacobi 迭代法和 G-S 迭代法的收敛性, 以及两者之间收敛速度的比较.

设 $\alpha > 0$, 定义

$$\tilde{\rho}(\alpha) = \rho(\alpha\tilde{L} + \tilde{U}), \quad \hat{\rho}(\alpha) = \rho(\tilde{L} + \alpha\tilde{U})$$

则有

$$\tilde{\rho}(0) = \hat{\rho}(0) = 0, \quad \tilde{\rho}(1) = \hat{\rho}(1) = \rho(G_{\mathrm{J}})$$

引理 设 $G_{\mathrm{J}} = \tilde{L} + \tilde{U} \geq 0$, $\alpha > 0$.

(1) 如果 $\rho(G_{\mathrm{J}}) = 0$, 则有 $\tilde{\rho}(\alpha) = \hat{\rho}(\alpha) = 0$.

(2) 如果 $\rho(G_{\mathrm{J}}) > 0$, 则 $\tilde{\rho}(\alpha)$ 和 $\hat{\rho}(\alpha)$ 都是关于 α 的严格递增函数.

(板书)

设 G_{J} **非负不可约**, 考虑 G-S 迭代矩阵 $\boxed{G_{\mathrm{GS}} = (D-L)^{-1}U = (I-\tilde{L})^{-1}\tilde{U}}$.

由于 $\tilde{L}$ 是严格下三角矩阵, 所以 $\rho(\tilde{L}) = 0 < 1$, $\tilde{L}^n = 0$, 因此

$$G_{\mathrm{GS}} = (I-\tilde{L})^{-1}\tilde{U} = \left(\sum_{k=0}^{n-1}\tilde{L}^k\right)\tilde{U} \geq 0.$$

所以 $\lambda = \rho(G_{\mathrm{GS}})$ 是 G_{GS} 的特征值, 且存在对应的非负特征向量 $x \gneq 0$, 即

$$G_{\mathrm{GS}}x = \lambda x \quad \text{或} \quad (I-\tilde{L})^{-1}\tilde{U}x = \lambda x.$$

由于 G_{J} 不可约, 可以证明 $\lambda > 0$ 且 $x > 0$ (留作练习). 于是上式可以改写为

$$(\lambda\tilde{L}+\tilde{U})x = \lambda x \quad \text{或} \quad \left(\tilde{L}+\frac{1}{\lambda}\tilde{U}\right)x = x.$$

又 $\lambda\tilde{L}+\tilde{U}$ 和 $\tilde{L}+\dfrac{1}{\lambda}\tilde{U}$ 都非负不可约且 $x > 0$, 所以

$$\tilde{\rho}(\lambda) = \rho(\lambda\tilde{L}+\tilde{U}) = \lambda, \quad \hat{\rho}(1/\lambda) = \rho\left(\tilde{L}+\frac{1}{\lambda}\tilde{U}\right) = 1.$$

$\rho(G_{\rm J})$ 与 $\rho(G_{\rm GS})$ 的关系

- 若 $\boxed{\rho(G_{\rm J})=1}$, 则 $\hat\rho(1)=\rho(G_{\rm J})=1$. 由 $\hat\rho(\alpha)$ 的严格单调性可知, $1/\lambda=1$, 即 $\lambda=1$, 也即 $\boxed{\rho(G_{\rm GS})=1}$.

- 若 $\boxed{0<\rho(G_{\rm J})<1}$, 则 $\hat\rho(1)=\rho(G_{\rm J})<1$. 根据 $\hat\rho(\alpha)$ 的严格单调性可知 $1/\lambda>1$, 即 $0<\lambda<1$. 再根据 $\tilde\rho(\alpha)$ 的严格单调性和 $\tilde\rho(1)=\rho(G_{\rm J})$, 以及 $\tilde\rho(\lambda)=\lambda$, 可得 $\lambda<\rho(G_{\rm J})$. 因此有

$$0<\lambda<\rho(G_{\rm J})<1, \quad 即 \quad \boxed{0<\rho(G_{\rm GS})<\rho(G_{\rm J})<1}.$$

- 若 $\boxed{\rho(G_{\rm J})>1}$, 我们可以类似地证明

$$\boxed{1<\rho(G_{\rm J})<\rho(G_{\rm GS})}.$$

如果 $G_{\rm J}$ 是非负可约的, 通过考察 $G_{\rm J}$ 的约化规范型, 我们可以得到类似的结论, 唯一的区别在于 $\rho(G_{\rm J})$ 可能为 0. 可以证明, 若 $\rho(G_{\rm J})=0$, 则 $\rho(G_{\rm GS})=0$.

综合以上讨论, 我们有下面的 **Stein-Rosenberg 定理**.

定理 设 $A\in\mathbb{R}^{n\times n}$, 若 $G_{\rm J}\geq 0$, 则下面四个结论有且仅有一个成立:

(1) $\rho(G_{\rm GS})=\rho(G_{\rm J})=0$,

(2) $0<\rho(G_{\rm GS})<\rho(G_{\rm J})<1$,

(3) $\rho(G_{\rm GS})=\rho(G_{\rm J})=1$,

(4) $\rho(G_{\rm GS})>\rho(G_{\rm J})>1$.

这表明, Jacobi 迭代法和 G-S 迭代法此时具有相同的收敛性. 而且当两者都收敛时, G-S 迭代法的收敛速度快于 Jacobi 迭代法.

注记

- 若 A 是对角线元素均为正的 Z-矩阵, 则 $G_{\mathrm{J}} \geq 0$.
- 该定理结论对一般矩阵并不成立: 对某些矩阵, Jacobi 迭代法收敛, 但 G-S 迭代法却不一定收敛.

例 设 $A = \begin{bmatrix} 1 & 2 & -2 \\ 1 & 1 & 1 \\ 2 & 2 & 1 \end{bmatrix}$, 则 Jacobi 和 G-S 的迭代矩阵分别为

$$G_{\mathrm{J}} = \begin{bmatrix} 0 & -2 & 2 \\ -1 & 0 & -1 \\ -2 & -2 & 0 \end{bmatrix}, \quad G_{\mathrm{GS}} = \begin{bmatrix} 0 & -2 & 2 \\ 0 & 2 & -3 \\ 0 & 0 & 2 \end{bmatrix}.$$

直接计算可得

$$\det(\lambda I - G_{\mathrm{J}}) = \lambda^3, \quad \det(\lambda I - G_{\mathrm{GS}}) = \lambda(\lambda - 2)^2.$$

所以 $\rho(G_{\mathrm{J}}) = 0 < 1$, 但 $\rho(G_{\mathrm{GS}}) = 2 > 1$.

3-2-4 不可约对角占优情形

定理 设 $A \in \mathbb{R}^{n\times n}$, 若 A 严格对角占优, 则 Jacobi 迭代法和 G-S 迭代法都收敛, 且

$$\|G_{\mathrm{GS}}\|_\infty \le \|G_{\mathrm{J}}\|_\infty < 1.$$

(留作课外自习)

✍ 当 A 是严格列对角占优时, 有类似的结论.

✍ 需要指出的是, $\|G_{\mathrm{GS}}\|_\infty \le \|G_{\mathrm{J}}\|_\infty$ 并不意味着 $\rho(G_{\mathrm{GS}}) \le \rho(G_{\mathrm{J}})$.

不可约对角占优

定理 设 $A \in \mathbb{R}^{n\times n}$, 若 A 是不可约对角占优, 则 Jacobi 迭代法和 G-S 迭代法都收敛.

(留作课外自习)

SOR 收敛的必要条件

定理 对于 SOR 迭代法, 有 $\rho(G_{\mathrm{SOR}}) \geq |1-\omega|$, 故 SOR 迭代法收敛的必要条件是

$$0 < \omega < 2.$$

(留作课外自习)

SOR 收敛的充分条件

定理 设 $A \in \mathbb{R}^{n\times n}$, 若 A 严格对角占优且 $0 < \omega \leq 1$, 则 SOR 迭代法收敛.

(留作课外自习)

定理 设 $A \in \mathbb{R}^{n\times n}$, 若 A 是不可约对角占优的, 且 $0 < \omega \leq 1$, 则 SOR 迭代法收敛.

(留作课外自习)

3-2-5 对称正定矩阵情形

Jacobi 迭代法的收敛性

定理 设 $A \in \mathbb{R}^{n\times n}$ 对称且对角线均为正, 则 Jacobi 迭代法收敛的充要条件是 A 和 $2D-A$ 都正定.

(留作课外自习)

G-S 迭代法的收敛性

定理 设 $A \in \mathbb{R}^{n\times n}$ 对称且对角线元素均为正, 则 G-S 迭代法收敛的充要条件是 A 正定.

SOR 迭代法的收敛性

定理 考虑 SOR 迭代法.

(1) 设 $A \in \mathbb{R}^{n\times n}$ 对称正定, 则 SOR 迭代法收敛的充要条件是 $0 < \omega < 2$.

(2) 设 $A \in \mathbb{R}^{n\times n}$ 对称且对角线元素均为正, 若存在 ω 使得 SOR 迭代法收敛, 则 A 正定.

推论 设 $A \in \mathbb{R}^{n\times n}$ 对称且对角线元素均为正, 则 SOR 迭代法收敛的充要条件是 A 正定且 $0 < \omega < 2$.

例 考虑线性方程组 $Ax = b$, 其中 $A = \begin{bmatrix} 1 & a & a \\ a & 1 & a \\ a & a & 1 \end{bmatrix}$. 试给出 Jacobi, G-S 和 SOR 收敛的充要条件.

解. A 对称且对角线均为正, 因此 Jacobi 收敛的充要条件是 A 和 $2D - A$ 都正定. 通过计算 A 的各阶顺序主子式可知, A 正定的充要条件是

$$\Delta_1 = 1 > 0, \quad \Delta_2 = 1 - a^2 > 0, \quad \Delta_3 = (a-1)^2(2a+1) > 0.$$

即 $-\frac{1}{2} < a < 1$. 类似地, 可知 $2D - A$ 正定的充要条件是 $-1 < a < \frac{1}{2}$.

所以 Jacobi 收敛的充要条件是 $-\frac{1}{2} < a < \frac{1}{2}$.

G-S 收敛的充要条件是 $-\frac{1}{2} < a < 1$.

SOR 收敛的充要条件是 $-\frac{1}{2} < a < 1$, $0 < \omega < 2$. □

3-2-6 相容次序矩阵

性质 A

定义 设 $A \in \mathbb{C}^{n\times n}$, 如果存在一个置换矩阵 P, 使得

$$PAP^{\intercal} = \begin{bmatrix} D_1 & F \\ E & D_2 \end{bmatrix},$$

其中 D_1, D_2 为对角矩阵, 则称 A 具有性质 **A**.

引理 设 $B \in \mathbb{C}^{n\times n}$ 具有下面的结构

$$B = \begin{bmatrix} 0 & B_{12} \\ B_{21} & 0 \end{bmatrix},$$

令 B_L 和 B_U 分别表示 B 的下三角和上三角部分, 则

(1) 若 μ 是 B 的特征值, 则 $-\mu$ 也是 B 的特征值;

(2) $B(\alpha)$ 的特征值与 α 无关, 其中

$$B(\alpha) = \alpha B_L + \frac{1}{\alpha} B_U, \quad 0 \neq \alpha \in \mathbb{C}.$$

(留作课外自习)

相容次序

定义 设 $A \in \mathbb{C}^{n\times n}$ 的对角线元素全不为零, 记

$$\tilde{L} = D^{-1}L, \quad \tilde{U} = D^{-1}U.$$

若矩阵 $G(\alpha) = \alpha\tilde{L} + \dfrac{1}{\alpha}\tilde{U}$ 的特征值与 α 无关, 则称 A 具有相容次序.

定理 设 A 的对角线元素全不为零, 若 A 具有性质 A, 则存在置换矩阵 P, 使得 $PAP^{\intercal}$ 具有相容次序.

Jacobi 与 SOR

定理 设 $A \in \mathbb{C}^{n\times n}$ 具有相容次序且 $\omega \neq 0$, 则下列命题成立

(1) $G_{\rm J}$ 的特征值正负成对出现;

(2) 若 μ 是 $G_{\rm J}$ 的特征值且 λ 满足

$$(\lambda+\omega-1)^2 = \lambda\omega^2\mu^2, \tag{3.2}$$

则 λ 是 $G_{\rm SOR}$ 的特征值;

(3) 反之, 若 $\lambda \neq 0$ 是 $G_{\rm SOR}$ 的特征值且 μ 满足 (3.2), 则 μ 是 $G_{\rm J}$ 的特征值.

(留作课外自习)

推论 设 A 具有相容次序且对角线均非零. 若 $G_{\rm J}$ 的特征值全部为实数, 则 SOR 迭代法收敛的充要条件是 $0 < \omega < 2$ 且 $\rho(G_{\rm J}) < 1$.

(留作练习)

推论 若 A 具有相容次序, 则 $\rho(G_{\rm GS}) = \rho(G_{\rm J})^2$, 即当 Jacobi 算法收敛时, G-S 算法比 Jacobi 算法快一倍.

SOR 最优参数选取

定理 设 A 具有相容次序，G_{J} 的特征值全部为实数，且 $\rho_{\mathrm{J}} = \rho(G_{\mathrm{J}}) < 1$，则 SOR 算法的最优参数和对应的谱半径分别为：

$$\omega_{opt} = \frac{2}{1+\sqrt{1-\rho_{\mathrm{J}}^2}}, \quad \rho(G_{\mathrm{SOR}}) = \omega_{opt} - 1 = \frac{\rho_{\mathrm{J}}^2}{\left(1+\sqrt{1-\rho_{\mathrm{J}}^2}\right)^2}.$$

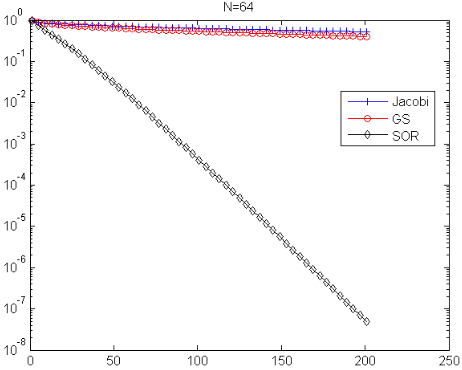
进一步，有

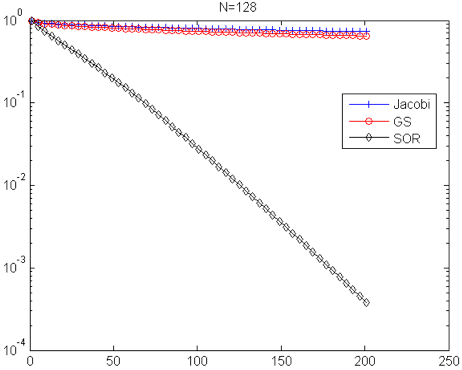
$$\rho(G_{\mathrm{SOR}}) = \begin{cases} \omega - 1, & \omega_{opt} \leq \omega \leq 2 \\ 1-\omega+\frac{1}{2}\omega^2\rho_{\mathrm{J}}^2 + \omega\rho_{\mathrm{J}}\sqrt{1-\omega+\frac{1}{4}\omega^2\rho_{\mathrm{J}}^2}\,, & 0 < \omega \leq \omega_{opt} \end{cases}$$

（留作课外自习）

例 采用红黑排序, 二维离散 Poisson 方程, 系数矩阵具有相容次序, G_J 对称且 $\rho(G_J) < 1$

(Poisson_Jacobi_GS_SOR.m)





第三讲 定常迭代法

目录

3.1 定常迭代法

3.2 收敛性分析

3.3 正则分裂

3.4 交替方向迭代法

3.5 迭代法的加速

3-3 正则分裂

3.3 正则分裂

3.3.1　正则分裂

3.3.2　正则分裂迭代法的收敛性

3.3.3　P-正则分裂

https://math.ecnu.edu.cn/~jypan/Teaching/MatrixIter

3-3-1 正则分裂

定义 设 $A \in \mathbb{R}^{n\times n}$, 矩阵分裂 $A = M - N$.

(1) 如果 $M^{-1} \geq 0$, $N \geq 0$, 则称该分裂为**正则分裂**.

(2) 如果 $M^{-1} \geq 0$, $M^{-1}N \geq 0$, 则称该分裂为**弱正则分裂**.

(3) 如果 $M^{-1}N \geq 0$, 则称该分裂为**非负分裂**.

✍ 显然, 正则分裂一定是弱正则分裂, 而弱正则分裂一定是非负分裂.

引理 设 $A = M - N$, 则 τ 是 $A^{-1}N$ 的特征值当且仅当 $\mu = \tau/(1+\tau)$ 是 $M^{-1}N$ 的特征值. 并且, $A^{-1}N$ 和 $M^{-1}N$ 具有相同的特征向量. (板书)

正则分裂：$M^{-1}N$ 的谱半径

定理 设 $A = M - N$ 是 $A \in \mathbb{R}^{n\times n}$ 的一个正则分裂，则 $A^{-1} \geq 0$ 当且仅当

$$\rho(M^{-1}N) = \frac{\rho(A^{-1}N)}{1+\rho(A^{-1}N)} < 1.$$

(板书)

M-矩阵情形

推论 设 $A \in \mathbb{R}^{n\times n}$ 是 M-矩阵, 且 $A = M - N$ 是一个正则分裂, 则 $\rho(M^{-1}N) < 1$.

设 A 是 M-矩阵, 下面的结论给出了构造 A 的正则分裂的一种方法.

推论 设 $A \in \mathbb{R}^{n\times n}$ 是 M-矩阵. 现将 A 的某些非对角元素设置为 0, 得到的新矩阵记为 M, 则 $A = M - N$ 是正则分裂, 故 $\rho(M^{-1}N) < 1$.

对称正定 M-矩阵情形

如果 A 还是对称正定的, 则我们有下面的结论.

定理 设 A 是 M-矩阵, $A = M - N$ 是正则分裂. 若 A 对称正定且 N 对称, 则

$$\rho(M^{-1}N) \leq \frac{\rho(A^{-1})\rho(N)}{1 + \rho(A^{-1})\rho(N)} < 1.$$

(板书)

比较定理

下面的定理给出了两个正则分裂所对应迭代矩阵的谱半径之间的关系.

定理 设 $A = M_1 - N_1 = M_2 - N_2$ 是 $A \in \mathbb{R}^{n\times n}$ 的两个正则分裂.

(1) 若 $A^{-1} \geq 0$ 且 $N_2 \geq N_1 \geq 0$, 则

$$0 \leq \rho(M_1^{-1}N_1) \leq \rho(M_2^{-1}N_2) < 1.$$

(2) 若 $A^{-1} > 0$ 且 $N_2 \gneq N_1 \gneq 0$, 则

$$0 < \rho(M_1^{-1}N_1) < \rho(M_2^{-1}N_2) < 1.$$

(板书)

不可约 M-矩阵情形

如果 A 是不可约 M-矩阵, 则我们有下面的结论.

推论 设 $A \in \mathbb{R}^{n\times n}$ 是不可约 M-矩阵, M_1 和 M_2 分别是将 A 的某些非对角元素设置为 0 后得到的. 若矩阵分裂 $A = M_1 - N_1 = M_2 - N_2$ 满足 $N_2 \gneq N_1 \gneq 0$, 则

$$0 < \rho(M_1^{-1}N_1) < \rho(M_2^{-1}N_2) < 1.$$

更多比较定理

定理 设 $A = M_1 - N_1 = M_2 - N_2$ 是 $A \in \mathbb{R}^{n\times n}$ 的两个正则分裂.

(1) 若 $A^{-1} \geq 0$, 且存在一个正整数 k 使得 $(A^{-1}N_2)^k A^{-1} \geq (A^{-1}N_1)^k A^{-1}$, 则

$$0 \leq \rho(M_1^{-1}N_1) \leq \rho(M_2^{-1}N_2) < 1.$$

(2) 若 $A^{-1} > 0$, 且存在一个正整数 k 使得 $(A^{-1}N_2)^k A^{-1} > (A^{-1}N_1)^k A^{-1}$, 则

$$0 < \rho(M_1^{-1}N_1) < \rho(M_2^{-1}N_2) < 1.$$

(留作课外自习)

更多比较定理

根据前面的引理和定理, 我们立即可以得到下面的结论.

定理 设 $A = M_1 - N_1 = M_2 - N_2$ 是 $A \in \mathbb{R}^{n\times n}$ 的两个正则分裂.

(1) 若 $A^{-1} \geq 0$ 且 $M_1^{-1} \geq M_2^{-1}$, 则

$$0 \leq \rho(M_1^{-1}N_1) \leq \rho(M_2^{-1}N_2) < 1.$$

(2) 若 $A^{-1} > 0$ 且 $M_1^{-1} > M_2^{-1}$, 则

$$0 < \rho(M_1^{-1}N_1) < \rho(M_2^{-1}N_2) < 1.$$

✍ 需要指出的是, 条件 $M_1^{-1} \geq M_2^{-1}$ 比 $N_2 \geq N_1 \geq 0$ 要更弱一些.

弱正则分裂

定理 设 $A = M - N$ 是 $A \in \mathbb{R}^{n\times n}$ 的一个弱正则分裂, 则 A 非奇异且 $A^{-1} \geq 0$ 当且仅当 $\rho(M^{-1}N) < 1$. (板书)

对称正定 M-矩阵的弱正则分裂

设 A 是 M-矩阵, $A = M - N$ 是弱正则分裂. 若 A 对称正定且 N 对称, 则

$$\rho(M^{-1}N) \leq \frac{\rho(A^{-1})\rho(N)}{1+\rho(A^{-1})\rho(N)} < 1.$$

非负分裂

定理 设 $A = M - N$ 是 $A \in \mathbb{R}^{n\times n}$ 的一个非负分裂, 则下面的结论是等价的:

(1) $\rho(M^{-1}N) < 1$;

(2) $I - M^{-1}N$ 是单调的;

(3) A 非奇异且 $A^{-1}N \geq 0$;

(4) A 非奇异且

$$\rho(M^{-1}N) = \frac{\rho(A^{-1}N)}{1 + \rho(A^{-1}N)}.$$

3-3-2 正则分裂与迭代收敛

如果 $A \in \mathbb{R}^n$ 是一个 M-矩阵, 则有 $A^{-1} \geq 0$, 则由前面的结论立即可得:

定理 设 $A \in \mathbb{R}^{n\times n}$ 是 M-矩阵. 如果 $A = M - N$ 是正则分裂, 则对应的矩阵分裂迭代格式 $x^{(k+1)} = M^{-1}Nx^{(k)} + M^{-1}b$ 收敛.

Jacobi 迭代法的收敛性

设 $A \in \mathbb{R}^{n\times n}$ 是 M-矩阵, 则其所有对角线元素都是正的, 而且 $L + U \geq 0$. 因此 $A = M - N$ 是一个正则分裂, 其中 $M = D$, $N = L + U$. 因此我们有下面的结论.

推论 设 $A \in \mathbb{R}^{n\times n}$ 是 M-矩阵, 则 Jacobi 迭代收敛.

设 A 是 H-矩阵, 则其比较矩阵 $\langle A\rangle = |D| - |L| - |U|$ 是 M-矩阵, 而且

$$\rho(G_{\mathrm{J}}) = \rho(D^{-1}(L+U)) \leq \rho(|D^{-1}(L+U)|) = \rho(|D|^{-1}(|L|+|U|)) < 1.$$

推论 设 $A \in \mathbb{R}^{n\times n}$ 是 H-矩阵, 则 Jacobi 迭代收敛.

事实上, 我们有下面更强的结论.

定理 设 $A \in \mathbb{R}^{n\times n}$ 的对角线元素均非零, 则 A 是 H-矩阵的充要条件是 $\rho(|G_{\mathrm{J}}|) < 1$.

(留作课外自习)

G-S 迭代法的收敛性

设 A 是 M-矩阵, 则 $(D-L)^{-1} \geq 0$. 因此, G-S 所对应的矩阵分裂也是正则分裂.

推论 设 $A \in \mathbb{R}^{n\times n}$ 是 M-矩阵, 则 Gauss-Seidel 迭代法收敛.

记 $\tilde{L} \triangleq D^{-1}L$, 则 $\tilde{L}$ 严格下三角, 故 $\rho(\tilde{L}) = 0 < 1$ 且当 $k \geq n$ 时有 $\tilde{L}^k = 0$. 我们有

$$(D-L)^{-1} = (I-\tilde{L})D^{-1} = (I+\tilde{L}+\tilde{L}^2+\cdots+\tilde{L}^{n-1})D^{-1}.$$

所以

$$|(D-L)^{-1}U| \leq |(D-L)^{-1}| \cdot |U| \leq \sum_{k=0}^{n-1} |\tilde{L}|^k |D|^{-1} \cdot |U| = (|D|-|L|)^{-1}|U|.$$

定理 设 $A \in \mathbb{R}^{n\times n}$ 是 H-矩阵, 则 Gauss-Seidel 迭代收敛.

H-矩阵与 AOR 的收敛性

设 $A=[a_{ij}]\in\mathbb{R}^{n\times n}$, 定义 A 的**等模矩阵**集合

$$\Omega(A)\triangleq\left\{B=[b_{ij}]\in\mathbb{C}^{n\times n}\,:\,|b_{ij}|=|a_{ij}|,\ 1\leq i,j\leq n\right\}.$$

下面的定理给出了 AOR 迭代法和 SAOR 迭代法的收敛定理.

定理 设 $A\in\mathbb{R}^{n\times n}$ 是 H 矩阵, 且对角线元素均非零, 参数 ω 和 γ 满足 $0\leq\gamma\leq\omega$. 则对任意 $B\in\Omega(A)$ 和任意 $0<\omega<\dfrac{2}{\rho(|G_{\mathrm{J}}|)+1}$, 都有

$$\rho(G_{\mathrm{AOR}}(B))<1\quad\text{和}\quad\rho(G_{\mathrm{SAOR}}(B))<1,$$

即求解线性方程组 $Bx=f$ 的 AOR 迭代法和 SAOR 迭代法都收敛. 这里 $G_{\mathrm{AOR}}(B)$ 和 $G_{\mathrm{SAOR}}(B)$ 分别表示 AOR 和 SAOR 所对应的迭代矩阵.

3-3-3 P-正则分裂

定义 设 $A \in \mathbb{C}^{n\times n}$, 如果 $M+N$ 是正定的, 则称 $A=M-N$ 是 A 的一个 P-正则分裂.

✍ 易知, $M+N$ 正定当且仅当 M^*+N 正定. 特别地, 如果 A 是 Hermitian 的, 则 $A=M-N$ 是 P-正则分裂的充要条件是 $M+M^*-A$ 是 Hermitian 正定.

Stein Theorem

定理 设 $G \in \mathbb{C}^{n\times n}$, 则 $\rho(G) < 1$ 的充要条件是存在一个 Hermitian 正定矩阵 $A \in \mathbb{C}^{n\times n}$ 使得 $A - G^*AG$ 也是 Hermitian 正定的. (留作练习)

Householder-John heorem

定理 设 $A \in \mathbb{C}^{n\times n}$ 是非奇异的 Hermitian 矩阵. 如果 $A = M - N$ 是一个 P-正则分裂, 则 $\rho(M^{-1}N) < 1$ 当且仅当 A 正定. (板书)

下面我们给出一个更一般的结论.

定理 (Ostrowski) 设 $A = D - E - E^* \in \mathbb{C}^{n\times n}$, 其中 D 是 Hermitian 正定的, 且对于所有 $\omega \in (0, 2)$, 矩阵 $D - \omega E$ 都非奇异, 则 $\rho(G_\omega) < 1$ 的充要条件是 A 正定且 $0 < \omega < 2$, 其中

$$G_\omega \triangleq (D - \omega E)^{-1}[(1-\omega)D + \omega E^*].$$

✍ 需要指出的是, 该定理中的矩阵 E 不需要是严格下三角或严格上三角矩阵.

Hermitian 正定矩阵与 P-正则分裂

下面的结论给出了一个 Hermitian 正定矩阵的分裂是 P-正则分裂的充要条件.

定理 设 $A \in \mathbb{C}^{n\times n}$ 是 Hermitian 正定的, 则 $A = M - N$ 是 P-正则分裂的充要条件是 $\|M^{-1}N\|_{A^{\frac{1}{2}}} < 1$. (板书)

第三讲 定常迭代法

目录

3.1 定常迭代法

3.2 收敛性分析

3.3 正则分裂

3.4 交替方向迭代法

3.5 加速技巧

3-4 交替方向迭代法

3.4 交替方向迭代法

3.4.1 多步迭代法

3.4.2 交替方向迭代法

3.4.3 HSS 算法

https://math.ecnu.edu.cn/~jypan/Teaching/MatrixIter

3-4-1 多步迭代法

设 $A = M_1 - N_1 = M_2 - N_2$ 是 A 的两个矩阵分裂, 则可以构造迭代格式

$$\begin{cases} M_1 x^{(k+\frac{1}{2})} = N_1 x^{(k)} + b, \\ M_2 x^{(k+1)} = N_2 x^{(k+\frac{1}{2})} + b, \end{cases} \qquad k = 0, 1, 2, \ldots. \tag{3.3}$$

这就是**两步迭代法**, 对应的分裂称为二重分裂/两步分裂.

- 易知, 两步迭代格式 (3.3) 的迭代矩阵为

$$G = M_2^{-1} N_2 M_1^{-1} N_1.$$

因此, 其收敛的充要条件是 $\rho(M_2^{-1} N_2 M_1^{-1} N_1) < 1$.

多步迭代法

类似地, 我们可以推广到多步迭代算法. 设 l 是一个正整数, 则 A 的 l 重分裂为

$$A = M_1 - N_1 = M_2 - N_2 = \cdots = M_l - N_l,$$

相应的**多步迭代**算法为

$$\begin{cases} M_1 x^{(k+\frac{1}{l})} = N_1 x^{(k)} + b, \\ M_2 x^{(k+\frac{2}{l})} = N_2 x^{(k+\frac{1}{l})} + b, \\ \quad \cdots \quad \cdots \\ M_l x^{(k+1)} = N_l x^{(k+\frac{l-1}{l})} + b, \end{cases} \qquad k = 0, 1, 2, \ldots.$$

3-4-2 交替方向法

设 $A = A_1 + A_2$, 则交替方向法 (ADI, Alternating-direction implicit) 迭代格式为

$$\begin{cases} (\alpha I + A_1)x^{(k+\frac{1}{2})} = (\alpha I - A_2)x^{(k)} + b, \\ (\alpha I + A_2)x^{(k+1)} = (\alpha I - A_1)x^{(k+\frac{1}{2})} + b, \end{cases} \qquad k = 0, 1, 2, \ldots,$$

其中 $\alpha \in \mathbb{R}$ 是迭代参数.

- 易知 ADI 算法的迭代矩阵为

$$G_{\mathrm{ADI}} = (\alpha I + A_2)^{-1}(\alpha I - A_1)(\alpha I + A_1)^{-1}(\alpha I - A_2).$$

定理 设 $A \in \mathbb{R}^{n\times n}$ 对称正定, $A = A_1 + A_2$, 其中 A_1 和 A_2 中有一个是对称正定, 另一个是对称半正定, 则对任意正数 $\alpha > 0$, 有 $\rho(G_{\mathrm{ADI}}) < 1$, 即 ADI 迭代法收敛. (板书)

3-4-3 HSS 方法

设 $A = H + S$, 其中 H 和 S 分别是 A 的对称与斜对称部分, 即

$$H = \frac{A + A^{\intercal}}{2}, \quad S = \frac{A - A^{\intercal}}{2}.$$

该分裂就称为**对称与斜对称分裂** (HSS, Hermitian and Skew-Hermitian Splitting).

类似于 ADI 方法, 我们可得下面的 HSS 方法

$$\begin{cases} (\alpha I + H)x^{(k+\frac{1}{2})} = (\alpha I - S)x^{(k)} + b, \\ (\alpha I + S)x^{(k+1)} = (\alpha I - H)x^{(k+\frac{1}{2})} + b, \end{cases} \qquad k = 0, 1, 2, \ldots.$$

迭代矩阵: $G_{\mathrm{HSS}}(\alpha) = (\alpha I + S)^{-1}(\alpha I - H)(\alpha I + H)^{-1}(\alpha I - S)$.

定理 设 $A \in \mathbb{R}^{n\times n}$ 正定, 则对任意 $\alpha > 0$, HSS 迭代法都收敛.

参数 α 的选取

通过分析可知

$$\rho(G_{\text{HSS}}) \leq \max_{\lambda\in\sigma(H)} \left|\frac{\alpha-\lambda}{\alpha+\lambda}\right| \triangleq \sigma(\alpha).$$

定理 设 $A \in \mathbb{R}^{n\times n}$ 正定, 则极小极大问题

$$\min_{\alpha>0} \max_{\lambda_{\min}(H)\leq\lambda\leq\lambda_{\max}(H)} \sigma(\alpha)$$

的解为

$$\alpha_* = \sqrt{\lambda_{\max}(H)\lambda_{\min}(H)}.$$

此时

$$\sigma(\alpha_*) = \frac{\sqrt{\lambda_{\max}(H)} - \sqrt{\lambda_{\min}(H)}}{\sqrt{\lambda_{\max}(H)} + \sqrt{\lambda_{\min}(H)}} = \frac{\sqrt{\kappa(H)}-1}{\sqrt{\kappa(H)}+1}.$$

HSS 的推广

正规与斜对称分裂法 (NSS)

$$\begin{cases} (\alpha I + N)x^{(k+\frac{1}{2})} = (\alpha I - S)x^{(k)} + b, \\ (\alpha I + S)x^{(k+1)} = (\alpha I - N)x^{(k+\frac{1}{2})} + b, \end{cases} \qquad k = 0, 1, 2, \ldots,$$

其中 $A = N + S$, 这里 N 是正规矩阵, S 是斜对称矩阵.

正定与斜对称分裂 (PSS)

$$\begin{cases} (\alpha I + P)x^{(k+\frac{1}{2})} = (\alpha I - S)x^{(k)} + b, \\ (\alpha I + S)x^{(k+1)} = (\alpha I - P)x^{(k+\frac{1}{2})} + b, \end{cases} \qquad k = 0, 1, 2, \ldots,$$

其中 $A = P + S$, 这里 P 是正定矩阵, S 是斜对称矩阵.

第三讲 定常迭代法

目录

3.1 定常迭代法

3.2 收敛性分析

3.3 正则分裂

3.4 交替方向迭代法

3.5 加速技巧

3-5 迭代法的加速

3.5 迭代法的加速

3.5.1 外推算法

3.5.2 Chebyshev 加速方法

https://math.ecnu.edu.cn/~jypan/Teaching/MatrixIter

出发点

当迭代解 $x^{(0)}, x^{(1)}, x^{(2)}, \ldots, x^{(k)}$ 已经计算出来后, 我们可以对其进行组合, 得到一个新的近似解, 这样就可能对原算法进行加速.

3-5-1 外推算法

设原迭代格式为

$$x^{(k+1)} = Gx^{(k)} + b$$

由 $x^{(k)}$ 和 $x^{(k+1)}$ 加权组合后可得新的近似解

$$\boxed{x^{(k+1)} = (1-\omega)x^{(k)} + \omega(Gx^{(k)} + b)},$$

其中 ω 是参数. 这种加速方法就称为**外推算法**.

如何加速

选择 ω 使得迭代矩阵 $\boxed{G_\omega \triangleq (1-\omega)I + \omega G}$ 的谱半径尽可能地小.

假设 G 的特征值都是实数, 且最大特征值和最小特征值分别为 λ_1 和 λ_n. 于是

$$\rho(G_\omega) = \max_{\lambda\in\sigma(G)} |(1-\omega)+\omega\lambda| = \max\{|1-\omega+\omega\lambda_1|, |1-\omega+\omega\lambda_n|\}.$$

定理 设 G 的特征值都是实数, 其最大和最小特征值分别为 λ_1 和 λ_n, 且 $1 \notin [\lambda_n, \lambda_1]$, 则

$$\omega_* = \arg\min_{\omega} \rho(G_\omega) = \frac{2}{2-(\lambda_1+\lambda_n)},$$

此时

$$\rho(G_{\omega_*}) = 1 - |\omega_*| d,$$

其中 d 是 1 到 $[\lambda_n, \lambda_1]$ 的距离, 即当 $\lambda_1 < 1$ 时, $d = 1-\lambda_1$, 当 $\lambda_n > 1$ 时, $d = \lambda_n - 1$.

(留作课外自习, 证明见讲义)

注记

- ✍ 由定理可知, $\rho(G_{\omega_*}) = 1 - |\omega_*|d$, 且当 $\omega_* \neq 1$ 时, 外推迭代比原迭代法收敛要更快一些.
- ✍ 最优参数依赖于原迭代矩阵 G 的特征值, 因此实用性不强. 在实际应用时可以估计特征值所在的区间 $[a, b]$, 然后用 a, b 代替 λ_n 和 λ_1.

JOR 算法

对 Jacobi 迭代进行外推加速, 则可得 JOR (Jacobi over-relaxation) 算法:

$$\begin{aligned} x^{(k+1)} &= (1-\omega)x^{(k)} + \omega(D^{-1}(L+U)x^{(k)} + D^{-1}b) \\ &= x^{(k)} + \omega D^{-1}(b - Ax^{(k)}), \quad k = 0, 1, 2, \ldots. \end{aligned}$$

定理 (JOR 的收敛性) 设 A 对称正定. 若

$$0 < \omega < \frac{2}{\rho(D^{-1}A)},$$

则 JOR 算法收敛.

3-5-2 Chebyshev 加速方法

对外推技巧进行推广, 假定原迭代格式已经计算出

$$x^{(0)}, x^{(1)}, \ldots, x^{(k)},$$

我们可以考虑如何将这些近似解进行组合, 以便得到更精确的近似解.

- 记 $\varepsilon_k = x^{(k)} - x_*$ 为第 k 步迭代解的误差, 则有

$$\varepsilon_k = G\varepsilon_{k-1} = G^2\varepsilon_{k-2} = \cdots = G^k\varepsilon_0.$$

设 $\tilde{x}^{(k)}$ 为 $x^{(0)}, x^{(1)}, \ldots, x^{(k)}$ 的一个线性组合, 即

$$\boxed{\tilde{x}^{(k)} = \alpha_0 x^{(0)} + \alpha_1 x^{(1)} + \cdots + \alpha_k x^{(k)}},$$

其中 α_i 为待定系数, 且满足 $\sum\limits_{i=0}^{k} \alpha_i = 1$.

于是

$$\tilde{x}^{(k)} - x_* = \alpha_0\varepsilon_0 + \alpha_1 G\varepsilon_0 + \cdots + \alpha_k G^k \varepsilon_0 \triangleq p_k(G)\varepsilon_0, \tag{3.4}$$

其中 $p_k(t) = \sum\limits_{i=0}^{k} \alpha_i t^i$ 为 k 次多项式, 且满足 $p_k(1) = 1$.

- 我们希望通过适当选取参数 α_i, 使得 $\tilde{x}^{(k)} - x_*$ 尽可能地小, 即使得 $\tilde{x}^{(k)}$ 收敛到 x_* 速度远远快于 $x^{(k)}$ 收敛到 x_* 速度.

- 这种加速方法就称为**多项式加速** 或 **半迭代方法**

例 设 $p_n(t)$ 为 G 的特征多项式, 则 $p_n(G) = 0$, 所以选取 α_i 为 p_n 的系数, 则 $\tilde{x}^{(n)} - x_* = 0$.
但这种选取方法不实用, 原因是:
(1) $p_n(t)$ 的系数并不知道;

(2) 我们通常希望收敛所需的迭代步数 $\ll n$.

下面讨论参数 α_i 的较实用的选取方法. 通过分析可知

$$\|\tilde{x}^{(k)} - x_*\|_2 = \|p_k(G)\varepsilon_0\|_2 \leq \|p_k(G)\|_2 \cdot \|\varepsilon_0\|_2.$$

因此我们需要求解下面的极小化问题

$$\boxed{\min_{p\in\mathbb{P}_k, p(1)=1} \|p(G)\|_2},$$

其中 $\mathbb{P}_k$ 表示所有次数不超过 k 的多项式组成的集合.

极小化问题求解

一般来说, 这个问题是非常困难的. 但在一些特殊情况下 (比如 G 对称), 我们可以给出其 (近似) 最优解.

假设迭代矩阵 G 对称, 即 G 存在特征值分解:

$$G = Q\Lambda Q^{\mathsf{T}},$$

其中 Λ 是实对角矩阵, Q 是正交矩阵. 于是有

$$\begin{aligned}
\min_{p\in\mathbb{P}_k, p(1)=1} \|p(G)\|_2 &= \min_{p\in\mathbb{P}_k, p(1)=1} \|p(\Lambda)\|_2 \\
&= \min_{p\in\mathbb{P}_k, p(1)=1} \max_{1\le i\le n} \{|p(\lambda_i)|\} \\
&\le \min_{p\in\mathbb{P}_k, p(1)=1} \max_{\lambda\in[\lambda_n,\lambda_1]} \{|p(\lambda)|\},
\end{aligned} \tag{3.5}$$

其中 λ_1, λ_n 分别表示 G 的最大和最小特征值.

这是带归一化条件的多项式最佳一致逼近问题 (与零的偏差最小).
该问题的解与著名的 **Chebyshev 多项式**有关.

Chebyshev 多项式的性质

定理 设 $\eta \in \mathbb{R}$ 满足 $|\eta| > 1$, 则下面的最小最大问题

$$\min_{p(t)\in\mathbb{P}_k,p(\eta)=1} \max_{-1\leq t\leq 1} |p(t)|$$

的唯一解为

$$\tilde{T}_k(t) \triangleq \frac{T_k(t)}{T_k(\eta)}.$$

定理 设 $\alpha, \beta, \eta \in \mathbb{R}$ 满足 $\alpha < \beta$ 且 $|\eta| \notin [\alpha, \beta]$. 则下面的最小最大问题

$$\min_{p(t)\in\mathbb{P}_k, p(\eta)=1} \max_{\alpha\leq x\leq\beta} |p(t)|$$

的唯一解为

$$\hat{T}_k(t) \triangleq \frac{T_k\left(\dfrac{2t-(\beta+\alpha)}{\beta-\alpha}\right)}{T_k\left(\dfrac{2\eta-(\beta+\alpha)}{\beta-\alpha}\right)}.$$

Chebyshev 加速方法

考虑迭代格式 $x^{(k+1)} = Gx^{(k)} + b$, 我们假定:

(1) 迭代矩阵 G 的特征值都是实数;

(2) 迭代矩阵谱半径 $\rho = \rho(G) < 1$, 故 $\lambda(G) \in [-\rho, \rho] \subset (-1, 1)$.

于是最小最大问题 $\boxed{\min\limits_{p\in\mathbb{P}_k, p(1)=1} \max\limits_{\lambda\in[\lambda_n,\lambda_1]} \{|p(\lambda)|\}}$ 的解为

$$p_k(t) = \frac{T_k(t/\rho)}{T_k(1/\rho)}.$$

具体实施

实际计算时, 是否需要先计算 $x^{(0)}, x^{(1)}, \ldots, x^{(k)}$, 然后通过线性组合

$$\tilde{x}^{(k)} = \alpha_0 x^{(0)} + \alpha_1 x^{(1)} + \cdots + \alpha_k x^{(k)}$$

来计算 $\tilde{x}^{(k)}$?

事实上, 我们可以通过 Chebyshev 多项式的三项递推公式

$$T_k(t) = 2tT_{k-1}(t) - T_{k-2}(t), \qquad p_k(t) = \frac{T_k(t/\rho)}{T_k(1/\rho)}$$

由 $\tilde{x}^{(k-1)}$ 和 $\tilde{x}^{(k-2)}$ 直接计算出 $\tilde{x}^{(k)}$.

推导过程

令 $\mu_k = \dfrac{1}{T_k(1/\rho)}$, 即 $T_k(1/\rho) = \dfrac{1}{\mu_k}$. 由三项递推公式可得

$$\frac{1}{\mu_k} = \frac{2}{\rho} \cdot \frac{1}{\mu_{k-1}} - \frac{1}{\mu_{k-2}}.$$

所以

$$\begin{aligned}
\tilde{x}^{(k)} - x_* &= p_k(G)\,\varepsilon_0 = \mu_k T_k(G/\rho)\,\varepsilon_0 \\
&= \mu_k \left[\frac{2G}{\rho} \cdot \frac{1}{\mu_{k-1}} (\tilde{x}^{(k-1)} - x_*) - \frac{1}{\mu_{k-2}} (\tilde{x}^{(k-2)} - x_*) \right].
\end{aligned}$$

整理后可得

$$\boxed{\tilde{x}^{(k)} = \frac{2\mu_k}{\mu_{k-1}} \cdot \frac{G}{\rho}\,\tilde{x}^{(k-1)} - \frac{\mu_k}{\mu_{k-2}}\tilde{x}^{(k-2)} + d_k,}$$

其中

$$\begin{aligned} d_k &= x_* - \frac{2\mu_k}{\mu_{k-1}} \cdot \frac{G}{\rho}\,x_* + \frac{\mu_k}{\mu_{k-2}} x_* = x^* - \frac{2\mu_k}{\mu_{k-1}} \cdot \frac{x_* - g}{\rho} + \frac{\mu_k}{\mu_{k-2}} x_* \\ &= \mu_k \left(\frac{1}{\mu_k} - \frac{2}{\rho\mu_{k-1}} + \frac{1}{\mu_{k-2}} \right) x_* + \frac{2\mu_k g}{\mu_{k-1}\rho} = \frac{2\mu_k g}{\mu_{k-1}\rho}. \end{aligned}$$

迭代格式 $x^{(k+1)} = Gx^{(k)} + b$ 的 Chebyshev 加速算法.

算法 4 Chebyshev 加速算法

Set $\mu_0 = 1$, $\mu_1 = \rho = \rho(G)$, $\tilde{x}^{(0)} = x^{(0)}$, $k = 1$

compute $\tilde{x}^{(1)} = Gx^{(0)} + g$

while not converge **do**

$\quad k = k + 1$

$$\mu_k = \left(\frac{2}{\rho} \cdot \frac{1}{\mu_{k-1}} - \frac{1}{\mu_{k-2}} \right)^{-1}$$

$$\tilde{x}^{(k)} = \frac{2\mu_k}{\mu_{k-1}} \cdot \frac{G}{\rho} \tilde{x}^{(k-1)} - \frac{\mu_k}{\mu_{k-2}} \tilde{x}^{(k-2)} + \frac{2\mu_k}{\mu_{k-1}\rho} \cdot g$$

end while

该算法的每步迭代的整体运算量与原迭代格式基本相当.


谢谢
THANK YOU