科学与工程计算基础

任课教师：黄忠亿

清华大学数学科学系

目录

简单迭代

前面已经提到, 许多大规模科学计算问题涉及到大型(且有一定带状结构的)稀疏矩阵线性方程组问题的求解. 对于此类问题如果仍用直接法求解, 计算量、存贮量均是目前计算机不可接受的. 因此, 我们一般用迭代法求解此类问题. 即构造格式:

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

最简单情形为 $m = 0$, 即只用上一步值. 当 F_k 为一线性映射, 即

$$\mathbf{x}^{(k+1)} = B_k\mathbf{x}^{(k)} + \mathbf{f}^{(k)}, \quad k = 0, 1, \cdots$$

此类方法称为单步线性迭代法. 进一步若 $B_k \equiv B$ 称为简单迭代.

迭代法的基本概念

定义 3.1 (向量序列收敛)

对于序列 $\{\mathbf{x}^{(k)}\}_{k=1}^{\infty} \subset \mathbb{C}^n$, 若存在向量 $\mathbf{x} \in \mathbb{C}^n$ *s.t.*

$$\lim_{k\to+\infty} \left\|\mathbf{x}^{(k)} - \mathbf{x}\right\| = 0,$$

则称序列 $\{\mathbf{x}^{(k)}\}_{k=1}^{\infty}$ 收敛于 $\mathbf{x}$, 简记为 $\lim\limits_{k\to\infty} \mathbf{x}^{(k)} = \mathbf{x}$.

由有限维空间上范数之等价性可知, 上述定义不依赖于范数的选取.

特别取 ∞-范数即得

$$\lim_{k\to\infty} \mathbf{x}^{(k)} = \mathbf{x} \Longleftrightarrow \lim_{k\to\infty} x_i^{(k)} = x_i, \quad i = 1, \cdots, n \text{ 即每个分量都收敛}$$

类似地，对于矩阵序列的收敛性可以如下定义。

迭代法的基本概念

定义 3.2 (矩阵序列收敛)

对于序列 $\{A^{(k)}\}_{k=1}^{\infty} \subset \mathbb{C}^{m\times n}$, 若存在 $A \in \mathbb{C}^{m\times n}$ *s.t.*

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

则称 序列 $\{A^{(k)}\}_{k=1}^{\infty}$ 收敛于 A, 简记为 $\lim\limits_{k\to\infty} A^{(k)} = A$.

同样可知, 矩阵收敛也等价于每个元素均相应收敛, 即

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

定理 3.1 (矩阵收敛与向量收敛间的关系)

设 $\{A^{(k)}\} \subset \mathbb{C}^{m\times n}$, 则 $\lim\limits_{k\to\infty} A^{(k)} = 0 \Longleftrightarrow \lim\limits_{k\to\infty} A^{(k)}\mathbf{x} = 0, \forall \mathbf{x} \in \mathbb{C}^n$.

迭代法的基本概念

◁ 由算子范数的定义: $\left\|A^{(k)}\mathbf{x}\right\| \leq \left\|A^{(k)}\right\| \cdot \|\mathbf{x}\|, \forall \mathbf{x} \in \mathbb{C}^n.$

即 $\lim\limits_{k\to\infty} A^{(k)} = 0 \Longleftrightarrow \lim\limits_{k\to\infty} \left\|A^{(k)}\right\| = 0 \Longrightarrow \lim\limits_{k\to\infty} \left\|A^{(k)}\mathbf{x}\right\| = 0,\ \forall \mathbf{x} \in \mathbb{C}^n.$

反之, 特别取 $\mathbf{x} = \mathbf{e}_j, j = 1, \cdots, n.$ 因 $\lim\limits_{k\to\infty} A^{(k)}\mathbf{e}_j = 0,$ 即 $A^{(k)}$ 的每一列都收敛于零. 自然有 $\lim\limits_{k\to\infty} A^{(k)} = 0$ ▷

对于简单迭代有

$$\begin{aligned} \mathbf{x}^{(k+1)} &= B\mathbf{x}^{(k)} + \mathbf{f}^{(k)} = B^2\mathbf{x}^{(k-1)} + B\mathbf{f}^{(k-1)} + \mathbf{f}^{(k)} = \cdots \\ &= B^{k+1}\mathbf{x}^{(0)} + B^k\mathbf{f}^{(0)} + \cdots + \mathbf{f}^{(k)} \end{aligned}$$

因此我们要研究矩阵序列 $\{B^k\}_{k=1}^{\infty}$ 的收敛性.

迭代法的基本概念

定理 3.2

设$B \in \mathbb{C}^{n\times n}$, 以下三个命题等价:

1. $\lim\limits_{k\to 0} B^k = \mathbf{0}$; 2. $\rho(B) < 1$; 3. 至少有一种范数使得 $\|B\| < 1$.

$\lhd$ 1 $\Longrightarrow$ 2: 设 λ 为 B 的特征值, $\mathbf{x} \neq 0$ 为相应特征向量: $B\mathbf{x} = \lambda\mathbf{x}$.

那么有 $B^k\mathbf{x} = \lambda^k\mathbf{x} \Longrightarrow \left\|B^k\mathbf{x}\right\| = \left\|\lambda^k\mathbf{x}\right\| = |\lambda|^k \|\mathbf{x}\|$.

由 $\lim\limits_{k\to 0} B^k = \mathbf{0} \Longrightarrow \forall \mathbf{y} \in \mathbb{C}^n, B^k\mathbf{y} \to 0 \Longrightarrow B^k\mathbf{x} \to 0$

$\Longrightarrow |\lambda|^k \|\mathbf{x}\| \to 0$ 即一定有 $|\lambda| < 1 \Longrightarrow \rho(B) < 1$.

2 $\Longrightarrow$ 3: 因 $\rho(B) < 1$, 令 $\varepsilon = \dfrac{1-\rho(B)}{2} > 0$, 由前面定理 2.9 知,

存在一种范数使得 $\|B\| \leq \rho(B) + \varepsilon < 1$.

3 $\Longrightarrow$ 1: 因 $\|B\| < 1 \Longrightarrow \left\|B^k\right\| \leq \|B\|^k \xrightarrow{k\to\infty} 0$, 即 $B^k \to 0$. $\rhd$

迭代法的基本概念

定理 3.3

设$B \in \mathbb{C}^{n\times n}$, $\|\cdot\|$ 为矩阵范数, 那么 $\lim\limits_{k\to\infty} \left\|B^k\right\|^{1/k} = \rho(B).$

◁ 因为 $\rho(B) \le \|B\| \Longrightarrow \rho(B) = [\rho(B^k)]^{1/k} \le \left\|B^k\right\|^{1/k}, \quad \forall k \in \mathbb{N}.$

下面只需证明, $\forall \varepsilon > 0$, 当 k 充分大有 $\left\|B^k\right\|^{1/k} \le \rho(B) + \varepsilon$ 即可.

令 $B_\varepsilon = [\rho(B) + \varepsilon]^{-1} B$, 显然有 $\rho(B_\varepsilon) = \dfrac{\rho(B)}{\rho(B) + \varepsilon} < 1.$

由前面定理3.2知 $\lim\limits_{k\to\infty} B_\varepsilon^k = 0.$ 即存在 $N = N(\varepsilon)$, 使得 $k > N$ 时,

$$\left\|B_\varepsilon^k\right\| = \left\|B^k\right\| \cdot [\rho(B) + \varepsilon]^{-k} < 1.$$

这样就有 $k > N$ 时, $\rho(B) \le \left\|B^k\right\|^{1/k} \le \rho(B) + \varepsilon$, 证毕. ▷

迭代法的基本概念

看一个例子（验证定理3.3的结论）：

例 3.1

设 $B = \begin{pmatrix} \frac{1}{2} & 0 \\ \frac{1}{4} & \frac{1}{2} \end{pmatrix}$，有 $\lambda_1 = \lambda_2 = \frac{1}{2}$，即 $\rho(B) = \frac{1}{2}$. 简单计算即得

$$B^k = \begin{pmatrix} \frac{1}{2^k} & 0 \\ \frac{k}{2^{k+1}} & \frac{1}{2^k} \end{pmatrix},$$ 取无穷范数有 $\left\|B^k\right\|_\infty = \frac{1}{2^k}\left(1 + \frac{k}{2}\right)$. 即

$$\left\|B^k\right\|_\infty^{1/k} = \frac{1}{2}\left(1 + \frac{k}{2}\right)^{1/k} \to \frac{1}{2} = \rho(B).$$

简单迭代公式的构造

欲求解 $A\mathbf{x}=\mathbf{b}$, 设 A 行列式不为零. 将 A 分裂成 $A=M-N$:
使得 M 为非奇异阵, 则 $A\mathbf{x}=\mathbf{b}\Longleftrightarrow(M-N)\mathbf{x}=\mathbf{b}$,

即等价于 $M\mathbf{x}=\mathbf{b}+N\mathbf{x}\Longleftrightarrow\mathbf{x}=M^{-1}N\mathbf{x}+M^{-1}\mathbf{b}$.

如果令 $B=M^{-1}N$, $\mathbf{f}=M^{-1}\mathbf{b}$, 即得 $\mathbf{x}=B\mathbf{x}+\mathbf{f}$.
这样便得到简单迭代格式

$$\mathbf{x}^{(k+1)}=B\mathbf{x}^{(k)}+\mathbf{f},\quad k=0,1,\cdots \tag{3.1}$$

给了初始猜测 $\mathbf{x}^{(0)}\in\mathbb{C}^n$, 如果上述迭代产生的序列 $\{\mathbf{x}^{(k)}\}_{k=0}^{\infty}$ 收敛, 易见有 $\mathbf{x}^{(k)}\to\mathbf{x}$.

简单迭代公式的构造

定义 3.3

若存在 $\mathbf{x}^* \in \mathbb{C}^n$, *s.t.* $\forall \mathbf{x}^{(0)} \in \mathbb{C}^n$, 上述迭代 (3.1) 产生的迭代序列 $\left\{\mathbf{x}^{(k)}\right\}_{k=0}^{\infty}$ 满足 $\lim\limits_{k\to\infty} \mathbf{x}^{(k)} = \mathbf{x}^*$, 则称该迭代法 (3.1) 收敛.

下面的问题在于, 给了矩阵 A, 如何构造 B, *s.t.* 迭代格式 (3.1):

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

是收敛的!

简单迭代公式的构造

我们先来看一个例子（简单迭代法）：

例 3.2 (Jabobi、Gauss-Seidel迭代)

设 $A=\begin{pmatrix}10 & 3 & 1\\ 3 & -10 & 3\\ 1 & 3 & 10\end{pmatrix}$, $\mathbf{b}=\begin{pmatrix}14\\ -4\\ 14\end{pmatrix}$, 有 $A\mathbf{x}=\mathbf{b}$ 的解为 $\mathbf{x}=\begin{pmatrix}1\\ 1\\ 1\end{pmatrix}$.

将A分裂为$A=M-N$: $M=\begin{pmatrix}10 & & \\ & -10 & \\ & & 10\end{pmatrix}$, $N=\begin{pmatrix}0 & -3 & -1\\ -3 & 0 & -3\\ -1 & -3 & 0\end{pmatrix}$

称之为Jacobi 迭代法,

有 $B_J=M^{-1}N=\begin{pmatrix}0 & -\frac{3}{10} & -\frac{1}{10}\\ \frac{3}{10} & 0 & \frac{3}{10}\\ -\frac{1}{10} & -\frac{3}{10} & 0\end{pmatrix}$, $\mathbf{f}_J=M^{-1}\mathbf{b}=\begin{pmatrix}\frac{14}{10}\\ \frac{4}{10}\\ \frac{14}{10}\end{pmatrix}$

简单迭代公式的构造

也可取: $M=\begin{pmatrix}10 & & \\ 3 & -10 & \\ 1 & 3 & 10\end{pmatrix}$, $N=\begin{pmatrix}0 & -3 & -1\\ 0 & 0 & -3\\ 0 & 0 & 0\end{pmatrix}$

称之为 Gauss-Seidel 迭代法,

即有 $B_{GS}=\begin{pmatrix}0 & -\frac{3}{10} & -\frac{1}{10}\\ 0 & -\frac{9}{100} & \frac{27}{100}\\ 0 & \frac{57}{1000} & -\frac{71}{1000}\end{pmatrix}$, $\mathbf{f}_{GS}=\begin{pmatrix}\frac{14}{10}\\ \frac{82}{100}\\ \frac{1014}{1000}\end{pmatrix}$

都取 $\mathbf{x}^{(0)}=(0,0,0)^T\in\mathbb{R}^3$. 用上述两种迭代格式计算有

Jacobi	$\mathbf{x}^{(1)}=(1.4,0.4,1.4)^T$	$\cdots$	$\mathbf{x}^{(4)}=(0.9834,0.9484,0.9834)^T$
G-S	$\mathbf{x}^{(1)}=(1.4,0.82,1.014)^T$	$\cdots$	$\mathbf{x}^{(4)}=(1.0013,1.0010,0.9996)^T$

简单迭代公式的构造

Gauss-Seidel 格式也可以看成(写成分量形式)

$$
\begin{aligned}
x_1^{(k+1)} &= \frac{1}{10}\left(\qquad -3x_2^{(k)} - x_3^{(k)} + 14\right) \\
x_2^{(k+1)} &= \frac{1}{10}\left(3x_1^{(k+1)} \qquad + 3x_3^{(k)} + 4\right) \\
x_3^{(k+1)} &= \frac{1}{10}\left(-x_1^{(k+1)} - 3x_2^{(k+1)} \qquad + 14\right)
\end{aligned}
$$

即把新算出来的 $x_1^{(k+1)}$、$x_2^{(k+1)}$ 用于计算后面的 $x_2^{(k+1)}$ 和 $x_3^{(k+1)}$.
这可以看成 Jacobi 迭代的某种改进.

从上例可以看出，确实Gauss-Seidel迭代比Jacobi迭代略快.

迭代法的收敛性分析

我们先来看简单迭代: $\mathbf{x}^{(k+1)} = B\mathbf{x}^{(k)} + \mathbf{f} \xrightarrow{\text{何时}} \mathbf{x}^* = B\mathbf{x}^* + \mathbf{f}$.

令 $\mathbf{e}^{(k)} = \mathbf{x}^{(k)} - \mathbf{x}^*, k = 0, 1, \cdots$

要收敛, 即要求 $\lim\limits_{k\to\infty} \mathbf{e}^{(k)} = 0, \ \forall \mathbf{e}^{(0)}$.

易见 $\mathbf{e}^{(k)} = B\mathbf{e}^{(k-1)} = B^k\mathbf{e}^{(0)}$. 由前面的定理 3.1、3.2即得

定理 3.4

下面任一个条件都是迭代法 $\mathbf{x}^{(k+1)} = B\mathbf{x}^{(k)} + \mathbf{f}$ 收敛的充要条件:

1) $\rho(B) < 1$; 2) 至少有一种范数使得 $\|B\| < 1$.

就是B^k—>0

对一般矩阵可以计算 $\|B\|_1$、$\|B\|_\infty$ 和 $\|B\|_2$ 等.

对于上例, 易见 $\|B_J\|_\infty = \frac{3}{5}, \|B_{GS}\|_\infty = \frac{2}{5}$, 均小于1.

迭代法的收敛性分析

利用矩阵范数可以给出简单迭代收敛的一些充分条件和收敛速度.

定理 3.5

设 $\mathbf{x}^*$ 是 $B\mathbf{x}+\mathbf{f}=\mathbf{x}$ 的唯一解, $\|\cdot\|$ 是任一向量范数, 如果由此定义的算子范数 $\|B\|=q<1$, 则该迭代收敛, 且有以下两个估计式

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

◁ 由定理 3.4 知该迭代法收敛. 并且

$$\mathbf{x}^{(k)}-\mathbf{x}^* = B(\mathbf{x}^{(k-1)}-\mathbf{x}^*) = B(\mathbf{x}^{(k-1)}-\mathbf{x}^{(k)}) + B(\mathbf{x}^{(k)}-\mathbf{x}^*)$$

$$\Longrightarrow \quad \left\|\mathbf{x}^{(k)}-\mathbf{x}^*\right\| \leq \|B\| \cdot \left\|\mathbf{x}^{(k-1)}-\mathbf{x}^{(k)}\right\| + \|B\| \cdot \left\|\mathbf{x}^{(k)}-\mathbf{x}^*\right\|$$

$$\overset{\|B\|=q<1}{\Longrightarrow} \left\|\mathbf{x}^{(k)}-\mathbf{x}^*\right\| \leq \frac{q}{1-q}\left\|\mathbf{x}^{(k-1)}-\mathbf{x}^{(k)}\right\|$$

再由 $\left\|\mathbf{x}^{(k)}-\mathbf{x}^{(k-1)}\right\| = \left\|B^{k-1}(\mathbf{x}^{(1)}-\mathbf{x}^{(0)})\right\| \leq q^{k-1}\left\|\mathbf{x}^{(1)}-\mathbf{x}^{(0)}\right\|$, 证毕. ▷

迭代法的收敛性分析

由误差定义 $\mathbf{e}^{(k)} = \mathbf{x}^{(k)} - \mathbf{x}^* = B^k\left(\mathbf{x}^{(0)} - \mathbf{x}^*\right) = B^k\mathbf{e}^{(0)}$,

即 $\left\|\mathbf{e}^{(k)}\right\| \leq \left\|B^k\right\| \cdot \left\|\mathbf{e}^{(0)}\right\|$.

这说明 k 次迭代后, 解的误差缩小为初始的 $\left\|B^k\right\|$ 倍.

亦即平均每一步迭代的收缩率为 $\left\|B^k\right\|^{1/k}$.

如果希望误差控制在 $\dfrac{\left\|\mathbf{e}^{(k)}\right\|}{\left\|\mathbf{e}^{(0)}\right\|} \leq \varepsilon$, 即需要 $k \geq \dfrac{-\ln\varepsilon}{-\ln\left\|B^k\right\|^{1/k}}$

另外我们知道 $\left\|B^k\right\|^{1/k} \to \rho(B)$, 因此一般用 $R(B) = -\ln\rho(B)$ 来定义渐近收敛率.

目录

Jacobi、Gauss-Seidel、SOR迭代法

我们经常遇到对角占优或对称正定矩阵, 因此我们来看对于这些矩阵如何构造简单迭代法. 上节例子中我们已经做过简单介绍.

欲求解 $A\mathbf{x}=\mathbf{b}$: 取 $D=diag(A)$ (即矩阵 A 的对角部分),

$-L$ 设为 A 的下三角部分, $-U$ 设为 A 的上三角部分,

即 $A=D-L-U$.

设D可逆 (对角严格占优和对称正定矩阵确实如此), 即$a_{ii}\neq 0$

令 $B_J=D^{-1}(L+U)=I-D^{-1}A$, $\mathbf{f}_J=D^{-1}\mathbf{b}$. 构造Jacobi迭代法如下

$$\mathbf{x}^{(k+1)}=B_J\mathbf{x}^{(k)}+\mathbf{f}_J,\quad k=0,1,\cdots \tag{3.2}$$

也可写成分量形式

$$x_i^{(k+1)}=\frac{1}{a_{ii}}\left(b_i-\sum_{j=1,j\neq i}^{n}a_{ij}x_j^{(k)}\right),\quad i=1,\cdots,n. \tag{3.3}$$

Jacobi、Gauss-Seidel、SOR迭代法

如果上面Jacobi迭代收敛, 那么计算出来 $x_i^{(k+1)}$ 之后, 如果立即用于下面计算 $x_{i+1}^{(k+1)}$ 的式子中 (即替换掉里面的 $x_i^{(k)}$), 有可能得到收敛更快的格式. 这便是Gauss-Seidel格式, 写成分量形式

$$x_i^{(k+1)} = \frac{1}{a_{ii}}\left(b_i - \sum_{j=1}^{i-1} a_{ij}x_j^{(k+1)} - \sum_{j=i+1}^{n} a_{ij}x_j^{(k)}\right), \quad i = 1, \cdots, n. \tag{3.4}$$

Gauss-Seidel格式也可写成矩阵形式, 即引入

$$B_{GS} = (D-L)^{-1}U = I - (D-L)^{-1}A, \quad \mathbf{f}_{GS} = (D-L)^{-1}\mathbf{b}.$$

Gauss-Seidel迭代法即为

$$\mathbf{x}^{(k+1)} = B_{GS}\cdot\mathbf{x}^{(k)} + \mathbf{f}_{GS}, \quad k = 0, 1, \cdots \tag{3.5}$$

把 $B_{GS}, \mathbf{f}_{GS}$ 的表达式代入上面 (3.5), 可以写成

Jacobi、Gauss-Seidel、SOR迭代法

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + (D-L)^{-1}[\mathbf{b} - A\mathbf{x}^{(k)}]$$

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

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

即 $$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + D^{-1}\left[\mathbf{b} + L\mathbf{x}^{(k+1)} - (D-U)\mathbf{x}^{(k)}\right]$$

或者说从 (3.4) 也可以得到分量形式

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

红色部分可看成是每一步的修正值, 我们也可引入一个松弛参数

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

称之为 SOR (Successive Over Relaxation, 逐次超松弛)方法.

Jacobi、Gauss-Seidel、SOR迭代法

写成矩阵形式 $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \omega D^{-1}\left[\mathbf{b} + L\mathbf{x}^{(k+1)} - (D-U)\mathbf{x}^{(k)}\right]$

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

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

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

$$\equiv B_{SOR}^{\omega}\mathbf{x}^{(k)} + \mathbf{f}_{SOR}$$

相当于 $A = M - N$, $M = \frac{1}{\omega}(D-\omega L)$, $N = \frac{1}{\omega}[(1-\omega)D + \omega U]$

例如对于上节例子, 取 $\omega = 0.94$, 发现确实比 G-S 收敛得还快, 有 $\mathbf{x}^{(4)} = (1.0008, 0.9999, 0.9999)^T$, 比G-S的 $\mathbf{x}^{(4)}$ 精度更高.

我们也可采用如下形式(SSOR, 当A是对称矩阵情形效果更好)

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

Jacobi、G-S、SOR迭代法的收敛性分析

从前面的定理 3.4、3.5 可知, 简单迭代法收敛 $\Longleftrightarrow \rho(B) < 1$. 且当 $\|B\| < 1$ 时可以给出收敛速度估计.

下面我们对一些特殊结构矩阵给出一些收敛的充分条件.

我们前面引入过以下定义

定义 3.4 (对角占优)

若 $A = (a_{ij}) \in \mathbb{C}^{n\times n}$ 满足 $|a_{ii}| > \sum\limits_{j=1, j\neq i}^{n} |a_{ij}|, i = 1, \cdots, n,$ 称 A 是严格对角占优阵. 若 A 满足 $|a_{ii}| \geq \sum\limits_{j=1, j\neq i}^{n} |a_{ij}|, \ i = 1, \cdots, n,$ 且至少有一个为严格不等号，称 A 为弱对角占优.

Jacobi、G-S、SOR迭代法的收敛性分析

定义 3.5 (可约、不可约)

设 $A \in \mathbb{C}^{n\times n}$. 如果存在排列阵 P 使得 $P^TAP = \begin{pmatrix} \bar{A}_{11} & \bar{A}_{12} \\ 0 & \bar{A}_{22} \end{pmatrix}$, 其中 $\bar{A}_{11}$, $\bar{A}_{22}$ 分别为 $r\times r$ 和 $(n-r)\times(n-r)$ 阶方阵, 则称 A 为可约矩阵。 否则就称 A 为不可约矩阵.

对于可约矩阵 A, 即在求解 $A\mathbf{x} = \mathbf{b}$ 时, 可以通过交换行列及对变量重新排序, 可以把问题分成两个低阶线性方程组求解.

如: $A = \begin{pmatrix} 2 & -1 & & \\ -1 & 2 & -1 & \\ & -1 & 2 & -1 \\ & & -1 & 2 \end{pmatrix}$ 不可约, $B = \begin{pmatrix} 2 & 0 & -1 & 1 \\ 0 & 2 & 0 & 1 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 0 & 2 \end{pmatrix}$ 可约.

Jacobi、G-S、SOR迭代法的收敛性分析

定理 3.6 (对角占优矩阵情形简单迭代的收敛性)

设 A 为严格对角占优或不可约的对角占优阵, 则求解 $A\mathbf{x} = \mathbf{b}$ 的 *Jacobi*、*Gauss-Seidel*迭代法均收敛.

$\lhd$ 1) 先看 A 为严格对角占优情形. 显然$|a_{ii}| > 0$, 且已证A可逆.

对于 $B_J = I - D^{-1}A = (b_{ij})$, 有 $b_{ij} = \begin{cases} 0, & i = j; \\ -a_{ij}/a_{ii}, & j \neq i \end{cases}$

因此有 $\|B_J\|_\infty < 1$ (每行绝对值之和都小于1), 自然Jacobi迭代收敛.

Jacobi、G-S、SOR迭代法的收敛性分析

下面欲证 $\rho(B_{GS}) < 1$. 设 λ 为 B_{GS} 的一个特征值, 即有

$$0 = \det(\lambda I - B_{GS}) = \det\left((D-L)^{-1}\right)\det\left(\lambda(D-L) - U\right).$$

若 $|\lambda| \geq 1 \Longrightarrow \lambda(D-L) - U$ 也为严格对角占优, 即行列式非零.

因而证明了 $\rho(B_{GS}) < 1$, 即 G-S 迭代收敛.

2) 再看 A 为不可约的对角占优阵情形. 仍然有 $|a_{ii}| > 0$.

对于 $B_J = I - D^{-1}A = (b_{ij})$, 欲证明 $\rho(B_J) < 1$. 设 $B_J\mathbf{y} = \mu\mathbf{y}$.

$$\text{即 } \mu y_i = -\sum_{j\neq i}\frac{a_{ij}}{a_{ii}}y_j \Longrightarrow |\mu||y_i| \leq \sum_{j\neq i}\frac{|a_{ij}|}{|a_{ii}|}|y_j|$$

Jacobi、G-S、SOR迭代法的收敛性分析

分两种情况讨论:

I) 如果 $|y_1| = |y_2| = \cdots = |y_n|$: 由上式立即有 $|\mu| \le \sum_{j\neq i} \frac{|a_{ij}|}{|a_{ii}|} \le 1$,

$i = 1, \cdots, n$. 且会至少有一个为严格不等号 $\Longrightarrow |\mu| < 1$.

II) 反之如果 $\mathbf{y}$ 的各分量绝对值不全等, 我们可以无妨设(不然可用排列阵做相似变换) $|y_1| \le \cdots \le |y_m| < |y_{m+1}| = \cdots = |y_n|$

对 $m+1 \le i \le n$, 有 $|\mu y_i| = |\mu y_n| \le \sum_{j=1}^{m} \frac{|a_{ij}|}{|a_{ii}|}|y_j| + \sum_{i\neq j=m+1}^{n} \frac{|a_{ij}|}{|a_{ii}|}|y_n|$

由于 A 的不可约性, 我们知道 $\left\{a_{ij},\ \begin{array}{l} 1 \le j \le m \\ m+1 \le i \le n \end{array}\right\}$ 不全为零.

Jacobi、G-S、SOR迭代法的收敛性分析

$\Longrightarrow$ 会有某个i, *s.t.* $|\mu||y_n| < \sum_{i\neq j=1}^{n} \frac{|a_{ij}|}{|a_{ii}|}|y_n|$

$\Longrightarrow |\mu| < 1$. 这样有 $\rho(B_J) < 1$. 即Jacobi迭代收敛.

由此我们也证明了不可约弱对角占优矩阵也非奇异: $\det A \neq 0$.

对于 B_{GS}: 因为 $A = D - L - U$ 不可约弱对角占优

$\Longrightarrow (D - L - \frac{1}{\lambda}U)$ 在 $|\lambda| \geq 1$ 时也是不可约弱对角占优阵

$$\Longrightarrow 0 \neq \det\left(D - L - \tfrac{1}{\lambda}U\right) = \det\left(\tfrac{1}{\lambda}(D-L)(\lambda I - (D-L)^{-1}U)\right)$$
$$= \det\left(\tfrac{1}{\lambda}(D-L)\right)\det\left(\lambda I - (D-L)^{-1}U\right)$$

$\Longrightarrow |\lambda| \geq 1$ 时, $\det\left(\lambda I - (D-L)^{-1}U\right) \neq 0 \Longrightarrow \rho(B_{GS}) < 1$. $\triangleright$

Jacobi、G-S、SOR迭代法的收敛性分析

例 3.3

前面求解两点边值问题得到的系数矩阵 $A=\begin{pmatrix} 2 & -1 & & \\ -1 & \ddots & \ddots & \\ & \ddots & \ddots & -1 \\ & & -1 & 2 \end{pmatrix}$ 是不可约弱对角占优阵, 因此 *Jacobi*、*Gauss-Seidel* 迭代均收敛.

定理 3.7

设 $A \in \mathbb{R}^{n\times n}$ 对称正定, 则 $Jacobi$ 迭代收敛$\Longleftrightarrow 2D-A$ 正定.

$\triangleleft$ 前面已经证过 A 对称正定 $\Longrightarrow a_{ii}>0$. 令 $D^{\frac{1}{2}}=diag(\sqrt{a_{ii}})$, 有

$$B_J = I - D^{-1}A = D^{-\frac{1}{2}}\left(I - D^{-\frac{1}{2}}AD^{-\frac{1}{2}}\right)D^{\frac{1}{2}} \quad (\text{相似})$$

$\Longrightarrow \sigma(B_J)=\sigma\left(I-D^{-\frac{1}{2}}AD^{-\frac{1}{2}}\right)$. 且易见 $D^{-\frac{1}{2}}AD^{-\frac{1}{2}}$ 为对称阵.

因此 $I-D^{-\frac{1}{2}}AD^{-\frac{1}{2}}$ 也为对称阵 $\Longrightarrow$ 其特征值均为实数.

Jacobi、G-S、SOR迭代法的收敛性分析

"$\Longrightarrow$:" 设Jacobi收敛, 即 $\rho(B_J) < 1$. 设 $D^{-\frac{1}{2}}AD^{-\frac{1}{2}}$ 的特征值为 μ.

则 $I - D^{-\frac{1}{2}}AD^{-\frac{1}{2}}$ 的特征值为 $1-\mu$. 因而有 $|1-\mu| < 1$

$\Longrightarrow \mu \in (0,2) \Longrightarrow D^{-\frac{1}{2}}AD^{-\frac{1}{2}}$ 为正定阵

又 $2D - A = D^{\frac{1}{2}}\left(2I - D^{-\frac{1}{2}}AD^{-\frac{1}{2}}\right)D^{\frac{1}{2}}$.

$\mu \in (0,2) \Longrightarrow 2I - D^{-\frac{1}{2}}AD^{-\frac{1}{2}}$ 正定.

由 $\mathbf{x}^T(2D-A)\mathbf{x} = \left(D^{\frac{1}{2}}\mathbf{x}\right)^T\left(2I - D^{-\frac{1}{2}}AD^{-\frac{1}{2}}\right)\left(D^{\frac{1}{2}}\mathbf{x}\right) > 0$

$\Longrightarrow 2D - A$也正定.

"$\Longleftarrow$:" 设 $2D - A$ 正定 $\Longrightarrow \mathbf{x}^T(2D-A)\mathbf{x} > 0$

由上式 $\Longrightarrow 2I - D^{-\frac{1}{2}}AD^{-\frac{1}{2}}$ 正定

$\Longrightarrow D^{-\frac{1}{2}}AD^{-\frac{1}{2}}$ 的特征值 $\mu \in (0,2) \Longrightarrow \rho(B_J) < 1$. $\triangleright$

Jacobi、G-S、SOR迭代法的收敛性分析

关于SOR方法我们有以下关于收敛性的必要条件:

定理 3.8 (Kahan)

设 A 的对角元非零, 则 $\rho(B_{SOR}^{\omega}) \geq |\omega - 1|$.

注: 这意味着 SOR 迭代收敛的必要条件是 $\omega \in (0, 2)$.

◁ 设 $\mu_1, \cdots, \mu_n$ 为 B_{SOR}^{ω} 的特征值, 我们有

$$\prod_{j=1}^{n} \mu_j = \det(B_{SOR}^{\omega}) = \det\left((D - \omega L)^{-1}\right) \det((1-\omega)D + \omega U)$$

(因 L-严格下三角阵, U-严格上三角阵) $= (1-\omega)^n$.

即 $\rho(B_{SOR}^{\omega}) = \max\limits_{1 \leq j \leq n} |\mu_j| \geq \sqrt[n]{\prod\limits_{j=1}^{n} |\mu_j|} = |\omega - 1|$. ▷

Jacobi、G-S、SOR迭代法的收敛性分析

对于正定矩阵, 关于SOR方法有以下收敛性的充分条件:

定理 3.9 (Ostrowski-Reich)

设 A 为正定 Hermite 阵（自然也包含了实对称正定阵情形）, 若 $0 < \omega < 2$, 则 SOR 方法收敛. 注: $\omega = 1$ 即为 G-S 此时也收敛.

(留作思考题：SSOR 此时也收敛)

◁ 欲证$\rho(B_{SOR}^{\omega}) < 1$. 设 λ 是 B_{SOR}^{ω} 的特征值, 相应特征向量为 $\mathbf{x}$.

即 $B_{SOR}^{\omega}\mathbf{x} = \lambda\mathbf{x}$, 欲证 $|\lambda| < 1$.

因 $B_{SOR}^{\omega} = (D - \omega L)^{-1}[(1-\omega)D + \omega U]$

$$\Longrightarrow [(1-\omega)D + \omega U]\mathbf{x} = \lambda(D - \omega L)\mathbf{x}.$$

Jacobi、G-S、SOR迭代法的收敛性分析

(利用 $A = D - L - U$)

整理有 $2[(1-\omega)D + \omega U] = (2-\omega)D - \omega A + \omega(U-L)$,

及 $2[D - \omega L] = (2-\omega)D + \omega A + \omega(U-L)$.

即 $[(2-\omega)D-\omega A+\omega(U-L)]\mathbf{x}=\lambda[(2-\omega)D+\omega A+\omega(U-L)]\mathbf{x}$

再左乘以 $\mathbf{x}^*$得: $\lambda = \dfrac{\mathbf{x}^*[(2-\omega)D-\omega A+\omega(U-L)]\mathbf{x}}{\mathbf{x}^*[(2-\omega)D+\omega A+\omega(U-L)]\mathbf{x}}$.

记 $d = (D\mathbf{x}, \mathbf{x})$, $a = (A\mathbf{x}, \mathbf{x})$, $s = i((U-L)\mathbf{x}, \mathbf{x}) = i\mathbf{x}^*(U-L)\mathbf{x}$.

Jacobi、G-S、SOR迭代法的收敛性分析

由于 A 为正定 Hermite 阵, 即有 $A^* = A$, $a_{ii} > 0$, 且 $d > 0$, $a > 0$, 及 $\bar{s} = -i\mathbf{x}^*(U^* - L^*)\mathbf{x} = -i\mathbf{x}^*(L - U)\mathbf{x} = s$, 即 $s \in \mathbb{R}$.

$$\text{即 } \lambda = \frac{(2-\omega)d - \omega a - i\omega s}{(2-\omega)d + \omega a - i\omega s}.$$

这样, $0 < \omega < 2$ 时, 上面分子实部的绝对值为 $|(2-\omega)d - \omega a|$, 小于分母的实部 $(2-\omega)d + \omega a$, 而分子分母虚部相同. 即 $|\lambda| < 1$.

这样证明了, $\omega \in (0, 2)$ 时, 对于正定Hermite阵, SOR 方法收敛. $\triangleright$

SOR迭代法的收敛性分析—最优松弛因子

我们看到 SOR 迭代法的收敛性与松弛因子 ω 有关. 那么如何寻找最优松弛因子呢？ 先引入一个定义.

定义 3.6 (相容次序矩阵 consistency ordered matrix)

令 $A = D - L - U$, 且 D 可逆. 若矩阵

$$C(\alpha) = \alpha D^{-1}L + \alpha^{-1}D^{-1}U, \quad \alpha \in \mathbb{C} \setminus \{0\},$$

的特征值与 α 无关, 则称 A 是相容次序矩阵.

虽然看似相容次序定义很苛刻，但事实上有很多这样的例子.

注 3.1

三对角矩阵(对角元非零)是相容次序矩阵.

SOR迭代法的收敛性分析—最优松弛因子

$\triangleleft$ 令 $S(\alpha) = diag(1, \alpha, \cdots, \alpha^{n-1})$. 对于

$$A = \begin{pmatrix} b_1 & c_1 & & \\ a_2 & \ddots & \ddots & \\ & \ddots & \ddots & c_{n-1} \\ & & a_n & b_n \end{pmatrix} = D - L - U, \text{ 有}$$

$$C(\alpha) = \alpha D^{-1}L + \alpha^{-1}D^{-1}U = -\begin{pmatrix} 0 & \frac{c_1}{\alpha b_1} & & \\ \frac{\alpha a_2}{b_2} & \ddots & \ddots & \\ & \ddots & \ddots & \frac{c_{n-1}}{\alpha b_{n-1}} \\ & & \frac{\alpha a_n}{b_n} & 0 \end{pmatrix},$$

易见 $C(\alpha) = S(\alpha)C(1)S^{-1}(\alpha)$.

自然 $C(\alpha)$ 的特征值都和 $C(1) = D^{-1}(L+U)$ 一样. $\triangleright$

SOR迭代法的收敛性分析—最优松弛因子

对于相容次序矩阵，我们有以下定理.

定理 3.10 (Young)

设 A 是相容次序矩阵, 且其*Jacobi*迭代矩阵 $B_J = D^{-1}(L+U)$ 的特征值都为实数, $\Lambda \equiv \rho(B_J) < 1$. 则 *SOR* 迭代对于 $0 < \omega < 2$ 收敛, 且 $\rho(B_{SOR}^{\omega})$ 在 ω 取 $\omega_{opt} = \frac{2}{1+\sqrt{1-\Lambda^2}}$ 时达到最小值

$$\rho(B_{SOR}^{\omega_{opt}}) = \frac{1-\sqrt{1-\Lambda^2}}{1+\sqrt{1-\Lambda^2}}.$$

◁ 记 $B(\omega) = (D-\omega L)^{-1}[(1-\omega)D+\omega U]$, 当 $0 \neq \mu$ 时, 利用以下等式

$$(I-\omega D^{-1}L)\big(\mu I - B(\omega)\big) = \mu(I-\omega D^{-1}L) - D^{-1}[(1-\omega)D+\omega U]$$

$$= (\mu+\omega-1)I - \omega\sqrt{\mu}(\sqrt{\mu}D^{-1}L + \frac{1}{\sqrt{\mu}}D^{-1}U)$$

SOR迭代法的收敛性分析—最优松弛因子

由 $I-\omega D^{-1}L$ 可逆，我们知道 μ 为 $B(\omega)$ 的特征值 $\Longleftrightarrow$

$\lambda=\dfrac{\mu+\omega-1}{\sqrt{\mu}\omega}$ 为 $C(\sqrt{\mu})=\sqrt{\mu}D^{-1}L+\dfrac{1}{\sqrt{\mu}}D^{-1}U$ 的特征值

由假设条件, A 是相容次序的，因而 λ 也是 $C(1)=D^{-1}(L+U)\equiv B_J$ 的特征值, 因而由已知条件知 $\lambda\in\mathbb{R}$.

解关于 $\sqrt{\mu}$ 的一元二次方程 : $\mu+\omega-1=\sqrt{\mu}\omega\lambda$ 得

$$\mu=\left(\frac{\omega\lambda}{2}\pm\sqrt{\frac{\omega^2\lambda^2}{4}+1-\omega}\right)^2$$

利用 $C(\alpha)$ 的表达式可知 $C(-1)=-C(1)$, 即若 λ 是 $C(\alpha)$ 的特征值, $-\lambda$ 也是其特征值.

SOR迭代法的收敛性分析—最优松弛因子

因为我们只关心 $B(\omega)$ 的谱半径, 因此我们可只考虑

$$\mu = \left(\frac{\omega|\lambda|}{2} \pm \frac{1}{2}\sqrt{\lambda^2\omega^2 - 4\omega + 4} \right)^2$$

已知 $|\lambda| < 1$, 因而 $\lambda^2\omega^2 - 4\omega + 4 = 0$ 有两个实根 $\omega_{1,2}(\lambda) = \frac{2\pm 2\sqrt{1-\lambda^2}}{\lambda^2}$

显然只有取负号的 $\omega_2(\lambda) = \frac{2-2\sqrt{1-\lambda^2}}{\lambda^2} = \frac{2}{1+\sqrt{1-\lambda^2}} \in [1, 2)$

这样当 $0 < \omega \le \omega_2(\lambda)$ 时, $\Delta = \frac{\omega^2\lambda^2}{4} + 1 - \omega \ge 0$.

即 μ 为实数, 因而 $|\mu(\omega(\lambda))| = \left(\frac{\omega|\lambda|}{2} \pm \sqrt{\frac{\omega^2\lambda^2}{4} + 1 - \omega} \right)^2$

当 $1 \le \omega_2(\lambda) < \omega < 2$ 时, $\Delta < 0$, $\mu(\omega)$ 为复数, $|\mu(\omega)| = \omega - 1$.

由上面两式可知: $|\mu(\omega(\lambda))|$ 随着 $|\lambda| \nearrow$ 而递增

SOR迭代法的收敛性分析—最优松弛因子

因此我们有

$$\rho\big(B(\omega)\big)=\begin{cases}\left(\frac{\omega\Lambda}{2}+\sqrt{\frac{\omega^2\Lambda^2}{4}+1-\omega}\right)^2, & 0<\omega\le\omega_2(\Lambda);\\ \omega-1, & \omega_2(\Lambda)<\omega<2.\end{cases}$$

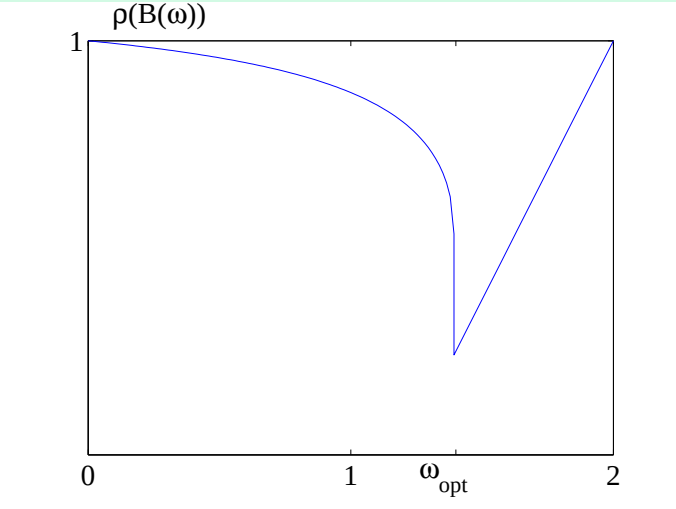
令 $f(\omega)=\frac{\omega\Lambda}{2}+\sqrt{\frac{\omega^2\Lambda^2}{4}+1-\omega}$, 显然有 $f(0)=1$, 且

$$f'(\omega)=\frac{|\Lambda|}{2}+\frac{\omega\lambda^2-2}{2\sqrt{\omega^2\Lambda^2+4-4\omega}}<0,\quad \text{当}0<\omega<\omega_2(\Lambda)\equiv\omega_{opt}$$

这样 (令 $\omega_{opt}\equiv\omega_2(\Lambda)=\frac{2}{1+\sqrt{1-\Lambda^2}}$), 有

$$0<\omega<\omega_{opt}\text{时},\rho\big(B(\omega)\big)\searrow;\quad \omega_{opt}<\omega<2\text{时},\rho(B(\omega))\nearrow$$

SOR迭代法的收敛性分析—最优松弛因子



即如左图所示, 我们证明了 $0 < \omega < 2$ 时,

$$\rho\big(B(\omega)\big) < 1.$$

且

$$\rho\big(B(\omega_{opt})\big) = \omega_{opt} - 1 = \frac{1-\sqrt{1-\Lambda^2}}{1+\sqrt{1-\Lambda^2}}. \quad \triangleright$$

推论 3.1

此时*G-S*迭代比*Jacobi*迭代快一倍. (因为 $\rho(B_{GS}) = \Lambda^2 = \rho^2(B_J)$)

而$\rho(B_{SOR}^{\omega_{opt}}) = \dfrac{\Lambda^2}{(1+\sqrt{1-\Lambda^2})^2} < \Lambda^2$, 即最优因子下SOR比G-S更快.

目录

变分问题

我们经常会遇到求解大规模稀疏矩阵方程组问题. 一般我们用有限元或者有限差分方法求解自伴算子椭圆型边值问题时, 得到的矩阵就是大规模稀疏的对称正定矩阵.

对于这类问题，我们先把求解线性方程组的问题变换成一个等价的变分问题.

对 $A\mathbf{x}=\mathbf{b}$, 设 A 对称正定, 定义二次函数

$$\mathcal{F}(\mathbf{x})=\frac{1}{2}(A\mathbf{x},\mathbf{x})-(\mathbf{b},\mathbf{x})=\frac{1}{2}\sum_{i,j=1}^{n}a_{ij}x_ix_j-\sum_{i=1}^{n}b_ix_i. \tag{3.6}$$

易见 $\nabla\mathcal{F}(\mathbf{x})=A\mathbf{x}-\mathbf{b}$.

变分问题

定理 3.11

设 A 为实对称正定矩阵，则 $A\mathbf{x}^* = \mathbf{b} \Longleftrightarrow \mathcal{F}(\mathbf{x}^*) = \min\limits_{\mathbf{x}\in\mathbb{R}^n} \mathcal{F}(\mathbf{x})$.

$\triangleleft$ “$\Longrightarrow$”若$A\mathbf{x}^* = \mathbf{b}$, 那么

$$\mathcal{F}(\mathbf{x}) - \mathcal{F}(\mathbf{x}^*) = \tfrac{1}{2}(A\mathbf{x},\mathbf{x}) - (\mathbf{b},\mathbf{x}) - \tfrac{1}{2}(A\mathbf{x}^*,\mathbf{x}^*) + (\mathbf{b},\mathbf{x}^*)$$

$$\overset{\mathbf{b}=A\mathbf{x}^*}{=} \tfrac{1}{2}[(A\mathbf{x},\mathbf{x}) - 2(A\mathbf{x}^*,\mathbf{x}) + (A\mathbf{x}^*,\mathbf{x}^*)] = \tfrac{1}{2}(A(\mathbf{x}-\mathbf{x}^*),\mathbf{x}-\mathbf{x}^*) \geq 0.$$

“$\Longleftarrow$” 设 $\mathcal{F}(\mathbf{x}^*)$ 达到最小 $\Longrightarrow \nabla\mathcal{F}(\mathbf{x}^*) = 0 \Longrightarrow A\mathbf{x}^* = b$. $\triangleright$

即求解 $A\mathbf{x} = \mathbf{b}$ 转化为等价的变分问题: 求 $\mathcal{F}(\mathbf{x})$ 的极小值点. 通常即用迭代法得到一个向量序列 $\{\mathbf{x}^{(k)}\}$ *s.t.* $\mathcal{F}(\mathbf{x}^{(k)}) \to \mathcal{F}(\mathbf{x}^*)$. 迭代法一般是通过每一步求一个局部极小值来得到序列$\{\mathbf{x}^{(k)}\}$.

一维搜索求解变分问题

高维搜索一般太复杂，无法找到精确解，通常每步做一个一维搜索：找一个向量 $\mathbf{p}^{(k)} \in \mathbb{R}^n$, 沿着 $\mathbf{p}^{(k)}$ 方向做一维搜索, 求极小值点 $\mathbf{x}^{(k+1)}$. 即求 $\alpha_k \in \mathbb{R}$, *s.t.*

$$\phi(\alpha_k) = \mathcal{F}(\mathbf{x}^{(k)} + \alpha_k \mathbf{p}^{(k)}) = \min_{\alpha \in \mathbb{R}} \mathcal{F}(\mathbf{x}^{(k)} + \alpha \mathbf{p}^{(k)}).$$

由此得到 $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \alpha_k \mathbf{p}^{(k)}$. 而

$$0 = \left.\frac{d\phi}{d\alpha}\right|_{\alpha_k} = \nabla \mathcal{F}(\mathbf{x}^{(k)} + \alpha_k \mathbf{p}^{(k)}) \cdot \mathbf{p}^{(k)} = \left(A(\mathbf{x}^{(k)} + \alpha_k \mathbf{p}^{(k)}) - \mathbf{b}, \mathbf{p}^{(k)}\right)$$

$$\Longrightarrow \alpha_k = \frac{\left(\mathbf{b} - A\mathbf{x}^{(k)}, \mathbf{p}^{(k)}\right)}{\left(A\mathbf{p}^{(k)}, \mathbf{p}^{(k)}\right)} \equiv \frac{\left(\mathbf{r}^{(k)}, \mathbf{p}^{(k)}\right)}{\left(A\mathbf{p}^{(k)}, \mathbf{p}^{(k)}\right)}$$

这里 $\mathbf{r}^{(k)} = \mathbf{b} - A\mathbf{x}^{(k)}$ 称为残差. 简单计算可得

$$\mathcal{F}(\mathbf{x}^{(k+1)}) = \mathcal{F}(\mathbf{x}^{(k)}) - \frac{1}{2}\frac{\left(\mathbf{r}^{(k)}, \mathbf{p}^{(k)}\right)^2}{\left(A\mathbf{p}^{(k)}, \mathbf{p}^{(k)}\right)} \leq \mathcal{F}(\mathbf{x}^{(k)}).$$ 确实在下降.

最速下降法

学习多元微积分时知道: 沿着负梯度方向函数值下降得最快！ 所以“最速下降法”就是每次都沿着当前点的负梯度方向搜索. 即搜索方向 $\mathbf{p}^{(k)} = -\nabla \mathcal{F}(\mathbf{x}^{(k)}) = \mathbf{b} - A\mathbf{x}^{(k)} = \mathbf{r}^{(k)}$, 取成“残差”方向.

算法 3.1 (最速下降法)

任取 $\mathbf{x}^{(0)} \in \mathbb{R}^n$, 对 $k = 0, 1, 2, \cdots$

1. 计算残差: $\mathbf{r}^{(k)} = \mathbf{b} - A\mathbf{x}^{(k)}$. 若 $\|\mathbf{r}^{(k)}\| < \varepsilon$ 则停止; 否则继续.
2. 计算 α_k: $\alpha_k = \dfrac{\left(\mathbf{r}^{(k)}, \mathbf{p}^{(k)}\right)}{\left(A\mathbf{p}^{(k)}, \mathbf{p}^{(k)}\right)}$.
3. 更新: $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \alpha_k \mathbf{r}^{(k)}$. 若 $|\alpha_k| < \delta$ 则停止, 否则继续.

由此可见, $\mathcal{F}(\mathbf{x}^{(k+1)}) = \mathcal{F}(\mathbf{x}^{(k)}) - \frac{1}{2} \dfrac{\left(\mathbf{r}^{(k)}, \mathbf{r}^{(k)}\right)^2}{\left(A\mathbf{r}^{(k)}, \mathbf{r}^{(k)}\right)}$.

最速下降法的收敛性定理

为研究最速下降法的收敛性，我们先给出以下引理

引理 3.1

设对称正定矩阵 A 的特征值为 $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n > 0$, 任给 $P_m(t)$ 为 m 次多项式, 则有

$$\|P_m(A)\mathbf{x}\|_A \leq \max_{1\leq i\leq n} |P_m(\lambda_i)| \, \|\mathbf{x}\|_A ,$$

其中 $\|\mathbf{x}\|_A = (\mathbf{x}, A\mathbf{x})^{1/2}$ 为一种向量范数, 称为 $A-$范数.

◁ 留作思考题: 可利用A的特征向量构成 $\mathbb{R}^n$ 的一组标准正交基来证明 ▷

最速下降法的收敛性定理

定理 3.12 (最速下降法的收敛性定理)

设 A 对称正定, λ_1, λ_n 为其最大、最小特征值, 则对于最速下降法有

$$\left\|\mathbf{x}^{(k)}-\mathbf{x}^*\right\|_A \leq \left(\frac{\lambda_1-\lambda_n}{\lambda_1+\lambda_n}\right)^k \left\|\mathbf{x}^{(0)}-\mathbf{x}^*\right\|_A, \text{其中} A\mathbf{x}^*=\mathbf{b}.$$

◁ 由最速下降法的迭代序列构造过程有

$$\mathcal{F}(\mathbf{x}^{(k)}) \leq \mathcal{F}(\mathbf{x}^{(k-1)}+\alpha\mathbf{r}^{(k-1)}), \quad \forall \alpha \in \mathbb{R}.$$

由 $\mathcal{F}(\mathbf{x})$ 的定义及 $\nabla\mathcal{F}(\mathbf{x}^*)=0$, 上面两边都减去 $\mathcal{F}(\mathbf{x}^*)$ 经简单计算有

$$\begin{aligned}\forall \alpha \in \mathbb{R}, \quad & (\mathbf{x}^{(k)}-\mathbf{x}^*)^T A(\mathbf{x}^{(k)}-\mathbf{x}^*) \\ \leq\ & (\mathbf{x}^{(k-1)}+\alpha\mathbf{r}^{(k-1)}-\mathbf{x}^*)^T A(\mathbf{x}^{(k-1)}+\alpha\mathbf{r}^{(k-1)}-\mathbf{x}^*) \\ =\ & [(I-\alpha A)(\mathbf{x}^{(k-1)}-\mathbf{x}^*)]^T A[(I-\alpha A)(\mathbf{x}^{(k-1)}-\mathbf{x}^*)].\end{aligned}$$

最速下降法的收敛性定理

在上面 引理3.1 中取 $P(t)=1-\alpha t$ 有

$$\begin{aligned}\left\|\mathbf{x}^{(k)}-\mathbf{x}^{*}\right\|_{A} &\leq \max_{1\leq i\leq n}|1-\alpha\lambda_i|\left\|\mathbf{x}^{(k-1)}-\mathbf{x}^{*}\right\|_{A}\\ \text{(利用连续性)} &\leq \max_{\lambda_n\leq\lambda\leq\lambda_1}|1-\alpha\lambda|\left\|\mathbf{x}^{(k-1)}-\mathbf{x}^{*}\right\|_{A}.\end{aligned}$$

要使上式右端达到最小, 有 $\alpha=\frac{2}{\lambda_1+\lambda_n}$, 此时 $\max\limits_{\lambda_n\leq\lambda\leq\lambda_1}|1-\alpha\lambda|=\dfrac{\lambda_1-\lambda_n}{\lambda_1+\lambda_n}$.

这样我们得到

$$\begin{aligned}\left\|\mathbf{x}^{(k)}-\mathbf{x}^{*}\right\|_{A} &\leq \left(\frac{\lambda_1-\lambda_n}{\lambda_1+\lambda_n}\right)\left\|\mathbf{x}^{(k-1)}-\mathbf{x}^{*}\right\|_{A}\\ &\leq \left(\frac{\lambda_1-\lambda_n}{\lambda_1+\lambda_n}\right)^{k}\left\|\mathbf{x}^{(0)}-\mathbf{x}^{*}\right\|_{A}. \quad \square\end{aligned}$$

最速下降法的收敛性定理

由此看出, 虽然最速下降法对于对称正定矩阵会保证收敛, 但当 A 条件数太大时 ($\lambda_1 \gg \lambda_n > 0$), 最速下降法收敛很慢! 另外, 其实最速下降法一开始会较快, 但是随着迭代增多, 由于舍入误差的影响, 每次搜索方向与实际残差方向会有很大差别, 会带来严重数值不稳定性.

因此我们需要构造新的搜索方向, 以使得搜索更为有效, 即希望每次搜索之后找到的是 前面所有搜索方向组成线性空间中的最小值点.

共轭梯度法

设每次搜索方向为 $\mathbf{p}^{(k)}$, 计算到第 $k+1$ 步 (无妨设 $\mathbf{x}^{(0)}=\mathbf{0}$)

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

我们自然希望

$$\mathcal{F}(\mathbf{x}^{(k+1)})=\min_{\mathbf{x}\in\mathrm{span}\{\mathbf{p}^{(0)},\cdots,\mathbf{p}^{(k)}\}}\mathcal{F}(\mathbf{x}).$$

如果令 $\mathbf{x}=\mathbf{y}+\alpha\mathbf{p}^{(k)}$, 其中 $\mathbf{y}\in\mathrm{span}\{\mathbf{p}^{(0)},\cdots,\mathbf{p}^{(k-1)}\}$. 那么

$$\mathcal{F}(\mathbf{x})=\mathcal{F}(\mathbf{y}+\alpha\mathbf{p}^{(k)})=\mathcal{F}(\mathbf{y})+\alpha(A\mathbf{y},\mathbf{p}^{(k)})-\alpha(\mathbf{b},\mathbf{p}^{(k)})+\frac{\alpha^2}{2}(A\mathbf{p}^{(k)},\mathbf{p}^{(k)})$$

如果$(A\mathbf{y},\mathbf{p}^{(k)})=0$, 则求$\min\mathcal{F}(\mathbf{x})$的问题转化为两个独立的子问题:

1) 求 $\min_{\mathbf{y}\in\mathrm{span}\{\mathbf{p}^{(0)},\cdots,\mathbf{p}^{(k-1)}\}}\mathcal{F}(\mathbf{y})$ 及

2) 求 $\alpha\in\mathbb{R}$ *s.t.* $\frac{\alpha^2}{2}(A\mathbf{p}^{(k)},\mathbf{p}^{(k)})-\alpha(\mathbf{b},\mathbf{p}^{(k)})$ 达到最小

共轭梯度法

求解上面第二个子问题即一维搜索问题

$$\Longrightarrow \alpha = \frac{(\mathbf{b}, \mathbf{p}^{(k)})}{(A\mathbf{p}^{(k)}, \mathbf{p}^{(k)})} = \frac{(\mathbf{b} - A\mathbf{x}^{(k)}, \mathbf{p}^{(k)})}{(A\mathbf{p}^{(k)}, \mathbf{p}^{(k)})}$$

因为假设了 $(A\mathbf{y}, \mathbf{p}^{(k)}) = 0 \Longrightarrow (A\mathbf{x}^{(k)}, \mathbf{p}^{(k)}) = 0.$

我们来看 $(A\mathbf{y}, \mathbf{p}^{(k)}) = 0$ 所引出的定义:

定义 3.7

设 A 对称正定, 若 $\{\mathbf{p}^{(0)}, \cdots, \mathbf{p}^{(m)}\} \subset \mathbb{R}^n$, *s.t.*

$$(A\mathbf{p}^{(i)}, \mathbf{p}^{(j)}) = 0, \quad \forall i \neq j,$$

则称 $\{\mathbf{p}^{(k)}\}_{k=0}^{m}$ 是 $\mathbb{R}^n$ 中的 A-共轭向量组 (或称为 A-正交向量组).

显然 $m < n$, 且不含零向量的共轭向量组是线性无关的.

共轭梯度法

若$A = I$, 则"A-共轭"回到普通正交情形.

类似于Gram-Schmit正交化过程, 可以构造出A-共轭的向量组.

当然, 与 $\mathbf{p}^{(0)}, \cdots, \mathbf{p}^{(k-1)}$ 都共轭的向量 $\mathbf{p}^{(k)}$ 并不唯一 (即便只从方向上看, 不算向量长度). 因此我们可以取

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

适当选取参数 β_{k-1} *s.t.* $(\mathbf{p}^{(k)}, A\mathbf{p}^{(k-1)}) = 0$, 即

$$(\mathbf{r}^{(k)} + \beta_{k-1}\mathbf{p}^{(k-1)}, A\mathbf{p}^{(k-1)}) = 0 \Longrightarrow \beta_{k-1} = -\frac{(\mathbf{r}^{(k)}, A\mathbf{p}^{(k-1)})}{(\mathbf{p}^{(k-1)}, A\mathbf{p}^{(k-1)})}.$$

共轭梯度法

这样共轭梯度法可以描述为：

算法 3.2 (共轭梯度法)

取 $\mathbf{x}^{(0)} \in \mathbb{R}^n$, 计算残差 $\mathbf{r}^{(0)} = \mathbf{b} - A\mathbf{x}^{(0)}$. 令 $k = 0$, $\mathbf{p}^{(0)} = \mathbf{r}^{(0)}$.

1. 计算 $\alpha_k = \dfrac{(\mathbf{r}^{(k)}, \mathbf{p}^{(k)})}{(A\mathbf{p}^{(k)}, \mathbf{p}^{(k)})}$, $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \alpha_k \mathbf{p}^{(k)}$.

   若 $\left\|\alpha_k \mathbf{p}^{(k)}\right\| < \varepsilon$ 则停止迭代; 否则继续.

2. 计算新残差 $\mathbf{r}^{(k+1)} = \mathbf{b} - A\mathbf{x}^{(k+1)}$, 计算 $\beta_k = -\dfrac{(\mathbf{r}^{(k+1)}, A\mathbf{p}^{(k)})}{(\mathbf{p}^{(k)}, A\mathbf{p}^{(k)})}$,

   给出新的搜索方向 $\mathbf{p}^{(k+1)} = \mathbf{r}^{(k+1)} + \beta_k \mathbf{p}^{(k)}$.

3. $k = k + 1$, 转第*1*步.

共轭梯度法

上面我们选取新搜索方向时, 仅仅保证了 $(\mathbf{p}^{(k+1)}, A\mathbf{p}^{(k)}) = 0$, 如何说明 $(\mathbf{p}^{(k+1)}, A\mathbf{p}^{(j)}) = 0, j = 0, 1, \cdots, k-1$ 都成立呢? 事实上我们有以下定理来保证:

定理 3.13

上面算法产生的向量组 $\{\mathbf{r}^{(k)}\}_{k\geq 0}$ 正交; $\{\mathbf{p}^{(k)}\}_{k\geq 0}$ 是 A-共轭的.

◁ 我们可以用归纳法来证明上述结论.

由 $\mathbf{r}^{(1)} = \mathbf{b} - A\mathbf{x}^{(1)} = \mathbf{b} - A(\mathbf{x}^{(0)} + \alpha_0\mathbf{p}^{(0)}) = \mathbf{b} - A\mathbf{x}^{(0)} - \alpha_0 A\mathbf{p}^{(0)}$

$\Longrightarrow (\mathbf{r}^{(0)}, \mathbf{r}^{(1)}) = (\mathbf{r}^{(0)}, \mathbf{r}^{(0)} - \alpha_0 A\mathbf{p}^{(0)}) = 0$ (由 $\mathbf{p}^{(0)} = \mathbf{r}^{(0)}$ 及 α_0 的定义)

$(\mathbf{p}^{(1)}, A\mathbf{p}^{(0)}) = (\mathbf{r}^{(1)}, A\mathbf{p}^{(0)}) + \beta_0(\mathbf{p}^{(0)}, A\mathbf{p}^{(0)})$ (由 β_0 的定义)

下面归纳假设 $j \leq k$ 时, $\{\mathbf{r}^{(j)}\}$ 正交; $\{\mathbf{p}^{(j)}\}$ 是 A-共轭的.

共轭梯度法

由 $\mathbf{p}^{(k+1)}$ 的定义可知 $\mathbf{p}^{(k+1)}$ 是 $\mathbf{r}^{(0)}, \mathbf{r}^{(1)}, \cdots, \mathbf{r}^{(k+1)}$ 的线性组合.

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

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

当 $j < k$ 时, 由归纳假设立即有 $(\mathbf{r}^{(k+1)}, \mathbf{r}^{(j)}) = 0$.

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

由于 $\mathbf{p}^{(k)} = \mathbf{r}^{(k)} + \sum_{0 \le j < k} c_j \mathbf{r}^{(j)} \Longrightarrow$

$$(\mathbf{r}^{(k+1)}, \mathbf{r}^{(k)}) = (\mathbf{r}^{(k)}, \mathbf{p}^{(k)}) - \alpha_k (A\mathbf{p}^{(k)}, \mathbf{p}^{(k)}) - \left(\mathbf{r}^{(k)}, \sum_{0 \le j < k} c_j \mathbf{r}^{(j)}\right) = 0$$

这用了 α_k的定义 及 $\{\mathbf{r}^{(j)}\}_{j \le k}$ 之正交性 $\overset{\text{归纳}}{\Longrightarrow}$ $\{\mathbf{r}^{(j)}\}_{j \le k+1}$ 正交.

共轭梯度法

再看 $(\mathbf{p}^{(k+1)}, A\mathbf{p}^{(j)}) = (\mathbf{r}^{(k+1)}, A\mathbf{p}^{(j)}) + \beta_k(\mathbf{p}^{(k)}, A\mathbf{p}^{(j)})$

当 $j = k$ 时, 由 β_k 的定义 可知上式为零.

当 $j < k$ 时, 由 $\mathbf{x}^{(j+1)} = \mathbf{x}^{(j)} + \alpha_j \mathbf{p}^{(j)} \Longrightarrow \mathbf{p}^{(j)} = \frac{\mathbf{x}^{(j+1)} - \mathbf{x}^{(j)}}{\alpha_j}$ 代入上式

再利用 $\{\mathbf{p}^{(j)}\}_{j \le k}$ 之 A-共轭性, 有

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

这里利用了已证明的 $\{\mathbf{r}^{(j)}\}_{j \le k+1}$ 之正交性及 $j + 1 < k + 1$. $\triangleright$

推论 3.2

如果不考虑舍入误差, 共轭梯度法至多 n 步可以得到准确解. (因为 $\mathbf{r}^{(n)}$ 必为零)

共轭梯度法

利用上面定理的结论, 我们还可以把共轭梯度法简化为

算法 3.3 (共轭梯度法)

取 $\mathbf{x}^{(0)} \in \mathbb{R}^n$, 计算残差 $\mathbf{r}^{(0)} = \mathbf{b} - A\mathbf{x}^{(0)}$. 令 $k = 0$, $\mathbf{p}^{(0)} = \mathbf{r}^{(0)}$.

1. 计算 $\alpha_k = \dfrac{\|\mathbf{r}^{(k)}\|_2^2}{(A\mathbf{p}^{(k)}, \mathbf{p}^{(k)})}$, $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \alpha_k \mathbf{p}^{(k)}$.

   若 $\left\|\alpha_k \mathbf{p}^{(k)}\right\| < \varepsilon$ 则停止迭代; 否则继续.

2. 计算新残差 $\mathbf{r}^{(k+1)} = \mathbf{r}^{(k)} - \alpha_k A\mathbf{p}^{(k)}$, 计算 $\beta_k = \dfrac{\|\mathbf{r}^{(k+1)}\|_2^2}{\|\mathbf{r}^{(k)}\|_2^2}$,

   [利用了 $A\mathbf{p}^{(j)} = \frac{1}{\alpha_j}(\mathbf{r}^{(j)} - \mathbf{r}^{(j+1)})$] $\mathbf{p}^{(k+1)} = \mathbf{r}^{(k+1)} + \beta_k \mathbf{p}^{(k)}$.

3. $\mathbf{r}^{(k+1)}$或$\mathbf{p}^{(k+1)} = 0$ 或 $k = n - 2$ 停止; 否则 $k = k + 1$, 转1

   $\alpha_{n-1} = \dfrac{\|\mathbf{r}^{(n-1)}\|_2^2}{(A\mathbf{p}^{(n-1)}, \mathbf{p}^{(n-1)})}$, $\mathbf{x}^{(n)} = \mathbf{x}^{(n-1)} + \alpha_{n-1}\mathbf{p}^{(n-1)}$.

预处理的共轭梯度法

实际计算时总会有舍入误差, 事实上随着 k 增大, 计算出的 $\{\mathbf{p}^{(j)}\}_{j=1}^{k}$ 会几乎线性相关. 因此不可能 n 步计算出精确解. 一般有以下估计式

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

这里 $\|\mathbf{x}\|_A^2=(A\mathbf{x},\mathbf{x})$, $A\mathbf{x}^*=\mathbf{b}$, $\kappa=\mathrm{cond}_2(A)$.

也就是说“共轭梯度法”比“最速下降法”收敛速度快一倍.

但是如果 $\kappa \gg 1$, 共轭梯度法收敛速度还会很慢. 我们还需做些处理以降低条件数.

因为 A 是对称正定矩阵, 我们自然希望预处理完以后的矩阵仍是对称正定的.

预处理的共轭梯度法

这样若设 $S \in \mathbb{R}^{n\times n}$ 为可逆矩阵, 令 $M = SS^T$ 自然为对称正定矩阵. 将 $A\mathbf{x} = \mathbf{b}$ 改写为等价形式

$$S^{-1}A(S^{-1})^T\mathbf{u} = S^{-1}\mathbf{b}, \qquad \text{其中 } \mathbf{u} = S^T\mathbf{x}.$$

令 $\widetilde{A} = S^{-1}AS^{-T}$, $\mathbf{f} = S^{-1}\mathbf{b}$, 那么 $\widetilde{A}$ 显然是对称正定的. 如果我们对方程组 $\widetilde{A}\mathbf{u} = \mathbf{f}$ 用共轭梯度法求解, 即有

取 $\mathbf{u}^{(0)} \in \mathbb{R}^n$, $\widetilde{\mathbf{r}}^{(0)} = \mathbf{f} - \widetilde{A}\mathbf{u}^{(0)}$, $\widetilde{\mathbf{p}}^{(0)} = \widetilde{\mathbf{r}}^{(0)}$, $k = 0, 1, \cdots$

1. $\widetilde{\alpha}_k = \dfrac{\|\widetilde{\mathbf{r}}^{(k)}\|_2^2}{(\widetilde{A}\widetilde{\mathbf{p}}^{(k)}, \widetilde{\mathbf{p}}^{(k)})}$, $\mathbf{u}^{(k+1)} = \mathbf{u}^{(k)} + \widetilde{\alpha}_k\mathbf{u}^{(k)}$.
2. $\widetilde{\mathbf{r}}^{(k+1)} = \widetilde{\mathbf{r}}^{(k)} - \widetilde{\alpha}_k\widetilde{A}\widetilde{\mathbf{p}}^{(k)}$.
3. $\widetilde{\beta}_k = \dfrac{\|\widetilde{\mathbf{r}}^{(k+1)}\|_2^2}{\|\widetilde{\mathbf{r}}^{(k)}\|_2^2}$, $\widetilde{\mathbf{p}}^{(k+1)} = \widetilde{\mathbf{r}}^{(k+1)} + \widetilde{\beta}_k\widetilde{\mathbf{p}}^{(k)}$.

预处理的共轭梯度法

再把变量 $\mathbf{u}$ 换回 $\mathbf{x}$: $\mathbf{u}^{(k)} = S^T\mathbf{x}^{(k)}$ 有

$$\tilde{\mathbf{r}}^{(k)} = \mathbf{f} - \tilde{A}\mathbf{u}^{(k)} = S^{-1}\left(\mathbf{b} - AS^{-T}S^T\mathbf{x}^{(k)}\right) = S^{-1}\mathbf{r}^{(k)}.$$

再令 $\mathbf{p}^{(k)} = S^{-T}\tilde{\mathbf{p}}^{(k)}$, $\mathbf{p}^{(0)} = M^{-1}\mathbf{r}^{(0)} = S^{-T}S^{-1}\mathbf{r}^{(0)}$, $\mathbf{z}^{(k)} = M^{-1}\mathbf{r}^{(k)}$, 便得到以下预处理的共轭梯度法 (PCG):

算法 3.4 (预处理共轭梯度法)

取 $\mathbf{x}^{(0)} \in \mathbb{R}^n$, $\mathbf{r}^{(0)} = \mathbf{b} - A\mathbf{x}^{(0)}$, $\mathbf{z}^{(0)} = M^{-1}\mathbf{r}^{(0)}$, $\mathbf{p}^{(0)} = \mathbf{z}^{(0)}$, $k = 0, 1, \cdots$

(如果修正量足够小则停止)

1. 计算 $\alpha_k = \frac{(\mathbf{z}^{(k)}, \mathbf{r}^{(k)})}{(A\mathbf{p}^{(k)}, \mathbf{p}^{(k)})}$, $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \alpha_k\mathbf{p}^{(k)}$.
2. 计算 $\mathbf{r}^{(k+1)} = \mathbf{r}^{(k)} - \alpha_k A\mathbf{p}^{(k)}$, 解 $M\mathbf{z}^{(k+1)} = \mathbf{r}^{(k+1)}$ 得 $\mathbf{z}^{(k+1)}$.
3. 计算 $\beta_k = \frac{(\mathbf{z}^{k+1}, \mathbf{r}^{(k+1)})}{(\mathbf{z}^{(k)}, \mathbf{r}^{(k)})}$, $\mathbf{p}^{(k+1)} = \mathbf{z}^{(k+1)} + \beta_k\mathbf{p}^{(k)}$.

预处理的共轭梯度法

可以证明, 上面预处理共轭梯度法得到的残差向量组 $\{\mathbf{r}^{(k)}\}$ 是 M^{-1}-共轭的, 搜索的方向 $\{\mathbf{p}^{(k)}\}$ 是 A-共轭的, 即

$$(\mathbf{p}^{(i)}, A\mathbf{p}^{(j)}) = 0, \quad (\mathbf{r}^{(i)}, M^{-1}\mathbf{r}^{(j)}) = 0, \quad \forall i \neq j.$$

因此我们有

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

其中 $\|\mathbf{x}\|_{\widetilde{A}} = (\widetilde{A}\mathbf{x}, \mathbf{x})$, $\widetilde{\kappa} = \mathrm{cond}_2(M^{-1}A)$.

因此，若选取M使得 $M^{-1}A \approx I$ 那么就可以大大降低条件数. 而且一般希望 M 为对称正定、稀疏, 这样 $M\mathbf{z} = \mathbf{r}$ 就易于求解. 对于 $M = LL^T$, 自然希望 $LL^T \approx A$ 最好. 因此可令 $A = M - N$, 其中 M 对称正定, N “尽可能小”.

预处理的共轭梯度法–Jacobi、对称Gauss-Seidel迭代预处理

最简单的即令 $A = D - L - L^T$, 其中 L 为严格下三角阵.

令 $M = D$ 即Jacobi迭代的分裂矩阵, $S = S^T = D^{\frac{1}{2}}$.

这在 A 的对角元量级差别很大时可使迭代收敛速度大大提高.

还可以取M为对称超松弛迭代的分裂矩阵, 即$M = SS^T$, 其中

$$S = [\omega(2-\omega)]^{-\frac{1}{2}}(D - \omega L)D^{-\frac{1}{2}}.$$

上述处理后, $\widetilde{A} = S^{-1}AS^{-T}$ 的条件数会大约为 A 的条件数的平方根左右. 特别, $\omega = 1$ 即 对称Gauss-Seidel迭代预处理效果就很好.

无论用Jacobi迭代还是对称Gauss-Seidel迭代预处理，一般只进行少数几次迭代就够了。

预处理的共轭梯度法–不完全Cholesky分解预处理

前面介绍的两种预处理方法，都是利用迭代法求解辅助方程组 $M\mathbf{z}=\mathbf{r}$. 下面介绍用直接法来求解上述方程组的预处理技术.

因为M对称正定，我们自然想到 M 的Cholesky分解 $M=LL^T$. 又希望M尽可能接近A, 所以可取L作为A的“近似”Cholesky因子:

预处理的共轭梯度法–不完全Cholesky分解预处理

算法 3.5

For $j = 1, n$

$$l_{jj} = (a_{jj} - \sum_{k=1}^{j-1} l_{jk}^2)^{1/2}$$

For $i = j + 1, n$

$\{$*if* $a_{ij} = 0$ *then* $l_{ij} = 0$

else$\}$ $l_{ij} = (a_{ij} - \sum_{k=1}^{j-1} l_{ik} l_{jk})/l_{jj}$

end

end

预处理的共轭梯度法–不完全Cholesky分解预处理

如果去掉上面算法中的 { } 内部分, 它就是之前介绍的Cholesky分解. 利用上述不完全 Cholesky 分解级数得到的因子 L 与 A 具有完全相同的稀疏性结构. 无论从存贮空间的需求还是从计算量的角度, 这种分解对于求解辅助方程组

$$LL^T\mathbf{z} = \mathbf{r}$$

都是十分有益的.

不过稍有遗憾的是, 并不是所有对称正定矩阵都可以有上述分解, 而且有时候也会是数值不稳定的. 另外上述分解方法使得 $M^{-1}A$ 的条件数可以大大减少的原因从理论上也还不是非常明确. 不过预处理后的效果还是相当显著的.

目录

关于线性方程组的Galerkin原理

假设要求解线性方程组

(3.7) $$A\mathbf{x} = \mathbf{b},$$

其中 $A \in \mathbb{R}^{n\times n}$ 是大型非奇异稀疏矩阵, $\mathbf{b} \in \mathbb{R}^n$ 为给定向量, 下面用到的范数 $\|\cdot\|$ 均为 2-范数. 当 A 不再是对称正定矩阵时, 近年来人们以 Galerkin 原理为基础, 研究了一系列算法求解此类问题.

记K_m和L_m是 $\mathbb{R}^n$ 中的两个m维子空间, 分别由 $\{\mathbf{v}_i\}_{i=1}^m$ 和 $\{\mathbf{w}_i\}_{i=1}^m$ 所张成. 取 $\mathbf{x}_0 \in \mathbb{R}^n$ 为任一向量, 令 $\mathbf{x} = \mathbf{x}_0 + \mathbf{z}$, 则 (3.7) 等价于

(3.8) $$A\mathbf{z} = \mathbf{r}_0,$$

其中残差 $\mathbf{r}_0 = \mathbf{b} - A\mathbf{x}_0$. 下面考虑方程组 (3.8) 的求解问题.

关于线性方程组的Galerkin原理

求解 (3.8) 的Galerkin原理可叙述为: 在子空间 K_m 中可以找到 (3.8) 的近似解 $\mathbf{z}_m$, 使得残差 $\mathbf{r}_0 - A\mathbf{z}_m$ 与 L_m 中所有向量正交, 即 $\mathbf{z}_m \in K_m$, *s.t.*

(3.9) $$(\mathbf{r}_0 - A\mathbf{z}_m, \mathbf{w}) = 0, \quad \forall \mathbf{w} \in L_m.$$

定义矩阵 $V_m = (\mathbf{v}_1, \cdots, \mathbf{v}_m)$ 及 $W_m = (\mathbf{w}_1, \cdots, \mathbf{w}_m)$, 那么 $\mathbf{z}_m$ 可表示成 $\mathbf{z}_m = V_m \mathbf{y}_m$, 其中 $\mathbf{y}_m \in \mathbb{R}^m$. 这样 (3.9) 可写成

(3.10) $$(W_m^T A V_m)\mathbf{y}_m = W_m^T \mathbf{r}_0.$$

如果 $W_m^T A V_m \in \mathbb{R}^{m\times m}$ 非奇异, 那么不难得到近似解

(3.11) $$\mathbf{z}_m = V_m (W_m^T A V_m)^{-1} W_m^T \mathbf{r}_0.$$

关于线性方程组的Galerkin原理

要使上述算法实用，那么需要解决以下几个问题:

1. 如何选择子空间 K_m 及 L_m 以及它们的基底 $\{\mathbf{v}_i\}$ 和 $\{\mathbf{w}_i\}$?
2. 如何有效地在计算机上实现上述算法?
3. 理论上，当且仅当 $m = n$ 时, $\mathbf{z}_m$ 才是精确解. 如果给定了某种 K_m, L_m 的选择方法, 当 $m \ll n$ 时, $\mathbf{z}_m$ 与 (3.8) 的精确解 $\mathbf{z}^*$ 之间的误差如何? 即怎么估计 $\|\mathbf{z}_m - \mathbf{z}^*\|$? 当 m 增加时, $\mathbf{z}_m$ 会收敛到 $\mathbf{z}^*$ 吗?

在实用和理论分析中最常用的两种选取子空间的方式为:

1. $L_m = K_m$. 此即 Arnoldi 算法.
2. $L_m = AK_m$. 此即 GMRES 算法.

Krylov子空间和Arnoldi算法

根据上面的讨论可知，用Galerkin原理求解 $A\mathbf{x}=\mathbf{b}$ 时，我们总是假设给定初值 $\mathbf{x}_0\in\mathbb{R}^n$, 并设 $\mathbf{x}=\mathbf{x}_0+\mathbf{z}$, 将原方程组化为求解 $A\mathbf{z}=\mathbf{r}_0$, 其中 $\mathbf{r}_0=\mathbf{b}-A\mathbf{x}_0$ 为残差. 然后选定两个 $m(\leq n)$ 维的子空间 K_m 和 L_m, 我们在 K_m 中寻找“近似”解 $\mathbf{z}_m$ 使得

$$\mathbf{z}_m\in K_m,\quad (\mathbf{r}_0-A\mathbf{z}_m)\perp L_m. \tag{3.12}$$

首先讨论子空间 K_m 的选取方法. 因为方程组 $A\mathbf{z}=\mathbf{r}_0$ 的解可表示为 $\mathbf{z}=A^{-1}\mathbf{r}_0$. 由矩阵论中的 Cayley-Hamilton 定理可知, 矩阵满足其本身的特征方程, 即

Krylov子空间和Arnoldi算法

(3.13) $$A^n + \alpha_1 A^{n-1} + \cdots + \alpha_{n-1} A + \alpha_n I = O,$$

这里 $\alpha_1, \cdots, \alpha_n$ 即 A 的特征多项式的系数.

因为 A 为非奇异矩阵, 将上式 (3.13) 两侧同乘以 A^{-1} 再移项得到

(3.14) $$A^{-1} = \frac{-1}{\alpha_n}\left(A^{n-1} + \alpha_1 A^{n-2} + \cdots + \alpha_{n-1} I\right),$$

从而有

(3.15) $$\mathbf{z} = A^{-1}\mathbf{r}_0 = \frac{-1}{\alpha_n}\left(A^{n-1}\mathbf{r}_0 + \alpha_1 A^{n-2}\mathbf{r}_0 + \cdots + \alpha_{n-1}\mathbf{r}_0\right),$$

也就是说, 向量 $\mathbf{z} \in \mathrm{span}\{\mathbf{r}_0, A\mathbf{r}_0, \cdots, A^{n-1}\mathbf{r}_0\}$.

受此启发，我们取 $K_m = \mathrm{span}\{\mathbf{r}_0, A\mathbf{r}_0, \cdots, A^{m-1}\mathbf{r}_0\}$, 并称此空间为由 A 和 $\mathbf{r}_0$ 生成的Krylov 子空间.

Krylov子空间和Arnoldi算法

如果我们取 $L_m = K_m$, 这样选取子空间的Galerkin方法就称为Arnoldi算法. 为简化方程 (3.10), 我们先求出 K_m 的一组标准正交基向量, 这个过程一般称为Arnoldi过程.

首先引入上Hessenberg阵的概念: 称 $H \in \mathbb{R}^{n\times n}$ 为上Hessenberg阵，是指

$$H = \begin{pmatrix} h_{11} & h_{12} & \cdots & h_{1n} \\ h_{21} & h_{22} & \cdots & h_{2n} \\ 0 & h_{32} & \ddots & \vdots \\ 0 & 0 & \cdots & h_{nn} \end{pmatrix},$$

即如果$i - j > 1$, 就有 $h_{ij} = 0$.

Krylov子空间和Arnoldi算法

下面我们来说明任何 $n \times n$ 的实方阵 A 可以用正交矩阵相似于上Hessenberg阵:

记正交矩阵 $V = (\mathbf{v}_1, \cdots, \mathbf{v}_n)$, 假设 $V^T AV = H$, 即等价于

$$AV = VH. \tag{3.16}$$

给定 $\mathbf{v}_1$ 满足 $\|\mathbf{v}_1\|_2 = 1$. 写出 (3.16) $AV = VH$ 的第一列, 有

$$A\mathbf{v}_1 = h_{11}\mathbf{v}_1 + h_{21}\mathbf{v}_2.$$

利用 $(\mathbf{v}_1, \mathbf{v}_1) = 1$ 和 $(\mathbf{v}_2, \mathbf{v}_1) = 0$ 得到

$$h_{11} = (\mathbf{v}_1, A\mathbf{v}_1), \quad h_{21}\mathbf{v}_2 = A\mathbf{v}_1 - h_{11}\mathbf{v}_1 \equiv \mathbf{r}_1.$$

Krylov子空间和Arnoldi算法

实际上就是h21=(Av1,v2)

再利用 $(\mathbf{v}_2, \mathbf{v}_2) = 1$ 得到 $h_{21} = \|\mathbf{r}_1\|_2$. 假设 $\|\mathbf{r}_1\| \neq 0$ 因而有

$$\mathbf{v}_2 = (A\mathbf{v}_1 - h_{11}\mathbf{v}_1)/\|\mathbf{r}_1\|_2 .$$

继续下去，可以写出 (3.16) 的第二列:

$$A\mathbf{v}_2 = h_{12}\mathbf{v}_1 + h_{22}\mathbf{v}_2 + h_{32}\mathbf{v}_3.$$

依照上面方法可以求出 $\mathbf{v}_3$ 和 H 的第二列. 只要 $\{\mathbf{r}_1, \mathbf{r}_2, \cdots, \mathbf{r}_{n-1}\}$ 都不为零, 就可以一直进行下去, 从而把 A 相似变换成上Hessenberg阵.

为了用Arnoldi过程求出 $\mathrm{span}\{\mathbf{r}_0, A\mathbf{r}_0, \cdots, A^{n-1}\mathbf{r}_0\}$ 的一组标准正交基，我们取 $\mathbf{v}_1 = \mathbf{r}_0/\|\mathbf{r}_0\|_2$. 可以证明以下定理

Krylov子空间和Arnoldi算法

定理 3.14

对于 $m < n$, 假设上述 **Arnoldi** 过程中 $\mathbf{r}_i \neq 0$, 产生的向量序列 $\{\mathbf{v}_i\}_{i=1}^m$ 是空间 $\mathrm{span}\{\mathbf{r}_0, A\mathbf{r}_0, \cdots, A^{m-1}\mathbf{r}_0\}$ 的一组标准正交基，而且 $\mathbf{v}_{m+1} \perp \mathrm{span}\{\mathbf{r}_0, A\mathbf{r}_0, \cdots, A^{m-1}\mathbf{r}_0\}$.

◁ 可以用数学归纳法证明（留作练习）. ▷

用矩阵符号写出上述过程有

$$AV_m = V_m H_m + h_{m+1,m}\mathbf{v}_{m+1}\mathbf{e}_m^T, \tag{3.17}$$

其中 $V_m = (\mathbf{v}_1, \cdots, \mathbf{v}_m)$, H_m 为 H 的前 $m \times m$ 阶子矩阵, $\mathbf{e}_m = (0, \cdots, 0, 1)^T \in \mathbb{R}^m$.

Krylov子空间和Arnoldi算法

在 $K_m = L_m$ 的标准正交基底 $\{\mathbf{v}_i\}_{i=1}^m$ 下, 利用 $\mathbf{v}_{m+1}$ 与 V_m 的正交性, 将 V_m^T 左乘 (3.17) 得 $V_m^T A V_m = H_m$, 及 $V_m^T \mathbf{r}_0 = \beta \mathbf{e}_1$, 这里 $\beta = \|\mathbf{r}_0\|_2$.

这样方程 (3.10) $(W_m^T A V_m)\mathbf{y}_m = W_m^T \mathbf{r}_0$ 变成 (注意此时 $W_m = V_m$)

$$H_m \mathbf{y}_m = \beta \mathbf{e}_1. \tag{3.18}$$

如果 H_m 非奇异, 那么求解上述方程组即得 $\mathbf{y}_m$, 这样就得到了一个“近似解” $\mathbf{z}_m = V_m \mathbf{y}_m$. 这就是 Arnoldi 算法的原理.

但是如果 H_m 是一个奇异阵, 那么我们什么也得不到. 我们称此时算法发生了恶性中断.

下面我们来估计一下， $A\mathbf{z} = \mathbf{r}_0$ “近似解” $\mathbf{z}_m$的残差大小.

Krylov子空间和Arnoldi算法

定理 3.15

对于 $m > 0$ (如果 $\mathbf{v}_{m+1} \neq 0$) 有

(3.19) $$\|\mathbf{r}_0 - A\mathbf{z}_m\| = |h_{m+1,m}\mathbf{e}_m^T\mathbf{y}_m|.$$

◁ 把 $\mathbf{z}_m = V_m H_m^{-1}\beta\mathbf{e}_1$ 代入 $\mathbf{r}_0 - A\mathbf{z}_m$ 有

$$\mathbf{r}_0 - A\mathbf{z}_m = \mathbf{r}_0 - AV_m H_m^{-1}\beta\mathbf{e}_1.$$

再利用 (3.17) 有

$$\begin{aligned}\mathbf{r}_0 - A\mathbf{z}_m &= \mathbf{r}_0 - (V_m H_m + h_{m+1,m}\mathbf{v}_{m+1}\mathbf{e}_m^T)H_m^{-1}\beta\mathbf{e}_1 \\ &= \mathbf{r}_0 - \beta\mathbf{v}_1 - \beta h_{m+1,m}\mathbf{v}_{m+1}\mathbf{e}_m^T H_m^{-1}\mathbf{e}_1\end{aligned}$$

Krylov子空间和Arnoldi算法

再注意到 $\beta\mathbf{v}_1 = \mathbf{r}_0$ 和 $\mathbf{y}_m = H_m^{-1}\beta\mathbf{e}_1$ 得

$$\mathbf{r}_0 - A\mathbf{z}_m = -h_{m+1,m}\mathbf{v}_{m+1}\mathbf{e}_m^T\mathbf{y}_m. \tag{3.20}$$

两边取范数, 并利用 $\|\mathbf{v}_{m+1}\| = 1$ 即得 (3.19). ▷

上面 (3.20) 式告诉我们, 如果 H_m 非奇异, 而 $\mathbf{v}_{m+1} = 0$, 那么我们得到的 $\mathbf{y}_m$ 给出的 $\mathbf{z}_m = V_m\mathbf{y}_m$ 就是方程 $A\mathbf{z} = \mathbf{r}_0$ 的准确解.

Krylov子空间和Arnoldi算法

算法 3.6 (Arnoldi算法)

1)取任意 $\mathbf{x}_0 \in \mathbb{R}^n$, 将方程化为 $A\mathbf{z} = \mathbf{r}_0$, 其中 $\mathbf{z} = \mathbf{x} - \mathbf{x}_0$, $\mathbf{r}_0 = \mathbf{b} - A\mathbf{x}_0$.

2)对 $m = 1, \cdots, n$ 完成Arnoldi过程.

如果 H_m 为奇异矩阵, 则算法产生了恶性中断. 更换 $\mathbf{x}_0$ 转 1).

否则 如果 $\mathbf{v}_{m+1} = 0$, 求解 $H_m\mathbf{y}_m = \beta\mathbf{e}_1$, 计算 $\mathbf{z}_m = V_m\mathbf{y}_m$,

得精确解 $\mathbf{x} = \mathbf{x}_0 + \mathbf{z}_m$, 停止.

否则 利用 (3.19) 估计残差

如果 $\|\mathbf{r}_0 - A\mathbf{z}_m\| < \varepsilon$ (制定误差界), 则 $\mathbf{x} = \mathbf{x}_0 + \mathbf{z}_m$, 停止.

否则 $m = m + 1$ 转 2).

GMRES算法

由于Arnoldi算法的恶性中断问题难以解决, 以及理论上很难分析其收敛性, 人们转向其他Galerkin型算法. 广义极小化残差算法(Generalized Minimal RESidual algorithm)便是其中之一, 它与各种预处理技术结合起来, 已经成为当前求解大型稀疏非对称线性方程组问题的主要手段.

仍设要求解的方程组为 (3.8): $A\mathbf{z} = \mathbf{r}_0$, 并记 $\beta = \|\mathbf{r}_0\|$. 现在取

$$K_m = \mathrm{span}\{\mathbf{r}_0, \cdot, A^{m-1}\mathbf{r}_0\}, \quad L_m = \mathrm{span}\{A\mathbf{r}_0, \cdots, A^m\mathbf{r}_0\},$$

或者简记 $L_m = AK_m$. 再利用 Galerkin 原理, 取 $\mathbf{z}_m \in K_m$, 令 $\mathbf{r}_m = \mathbf{r}_0 - A\mathbf{z}_m$ 与 L_m 中所有向量正交而得到 $\mathbf{z}_m$. 有以下引理:

GMRES算法

引理 3.2

设 $A \in \mathbb{R}^{n\times n}$ 非奇异, K_m 为 $m(\leq n)$ 维子空间, $L_m = AK_m$. 令V_m与W_m分别代表K_m与L_m中一组基向量构成的矩阵, 则 $B_m = W_m^T AV_m$ 非奇异.

◁ 设 $V_m = (\mathbf{v}_1, \cdots, \mathbf{v}_m)$ 是子空间 K_m 的一组基, $W_m = (\mathbf{w}_1, \cdots, \mathbf{w}_m)$ 是 $L_m = AK_m$ 的一组基. $\mathbf{w}_i$ 可以写成 $\mathbf{w}_i = A\mathbf{u}_i$, 其中 $\mathbf{u}_i \in K_m$. 所以

$$\mathbf{w}_i = A\mathbf{u}_i = AV_m\mathbf{g}_i, \quad \mathbf{g}_i \in \mathbb{R}^m.$$

令 $G = (\mathbf{g}_1, \cdots, \mathbf{g}_m)$, 我们有 $W_m = AV_mG$. 显然 G 是非奇异矩阵, 从而得到

$$\begin{aligned} B_m &= W_m^T AV_m = G^T V_m^T A^T AV_m \\ &= G^T (AV_m)^T (AV_m) \end{aligned}$$

因 $(AV_m)^T(AV_m)$ 是$m \times m$的对称正定阵, 且G非奇异, 因而B非奇异. ▷

GMRES算法

我们还有以下引理结论:

引理 3.3

令 A 为 $n \times n$ 阶方阵, 设 $L_m = AK_m$, 任取 $\mathbf{x}_0 \in \mathbb{R}^n$ 为初始向量. 则按照 Galerkin 原理计算近似解 $\widetilde{\mathbf{x}}$, 等价于 $\widetilde{\mathbf{x}}$ 是在 $\mathbf{x}_0 + K_m$ 中极小化泛函 $R(\mathbf{x}) = \|\mathbf{b} - A\mathbf{x}\|_2^2$, 即

$$R(\widetilde{\mathbf{x}}) = \min_{\mathbf{x} \in \mathbf{x}_0 + K_m} R(\mathbf{x}). \tag{3.21}$$

GMRES算法

◁ 先证明必要性 "$\Longrightarrow$":

对任意的 $\mathbf{x} \in \mathbf{x}_0 + K_m$, 都有

$$\|\mathbf{b} - A\mathbf{x}\|_2^2 = \|\mathbf{b} - A(\mathbf{x} - \widetilde{\mathbf{x}} + \widetilde{\mathbf{x}})\|_2^2$$

$$= \|\mathbf{b} - A\widetilde{\mathbf{x}}\|_2^2 - 2\big((\mathbf{b} - A\widetilde{\mathbf{x}}), A(\mathbf{x} - \widetilde{\mathbf{x}})\big) + \|A(\mathbf{x} - \widetilde{\mathbf{x}})\|_2^2 \tag{3.22}$$

其中 $A(\mathbf{x} - \widetilde{\mathbf{x}}) \in AK_m = L_m$. 因为 $\widetilde{\mathbf{x}}$ 是按照 (3.9) 得到的, 这样上面 (3.22) 中第二项为零, 从而得到 $\|\mathbf{b} - A\mathbf{x}\|_2^2 \geq \|\mathbf{b} - A\widetilde{\mathbf{x}}\|_2^2$.

再证充分性 "$\Longleftarrow$":

如果 $\widetilde{\mathbf{x}} \in \mathbf{x}_0 + K_m$ 使得 $R(\mathbf{x})$ 达到极小, 那么 $\forall \alpha \in \mathbb{R}$, $\forall \mathbf{v} \in K_m$, 有

$$\|\mathbf{b} - A(\widetilde{\mathbf{x}} + \alpha\mathbf{v})\|_2^2 \geq \|\mathbf{b} - A\widetilde{\mathbf{x}}\|_2^2 .$$

GMRES算法

上式左端是 α 的二次函数, 记之为 $Q(\alpha)$, 有

$$Q(\alpha) = \alpha^2 \|A\mathbf{v}\|_2^2 - 2\alpha(\mathbf{b} - A\widetilde{\mathbf{x}}, A\mathbf{v}) + \|\mathbf{b} - A\widetilde{\mathbf{x}}\|_2^2 .$$

显然 $\alpha = 0$ 时 $Q(\alpha)$ 取极小值, 即

$$\left.\frac{dQ(\alpha)}{d\alpha}\right|_{\alpha=0} = -2(\mathbf{b} - A\widetilde{\mathbf{x}}, A\mathbf{v}) = 0.$$

上式对任何 $\mathbf{v} \in K_m$ 都成立, 从而对任何 $L_m = AK_m$ 中向量成立. 这恰好是Galerkin条件的矩阵表达形式. 证毕. ▷

这个引理表明, 对于这种特定的 K_m, L_m 选取方式, 求 $\mathbf{z}_m$ 实际上等价于在 K_m 中极小化残差的 2-范数, 故称之为广义极小化残差方法.

GMRES算法

基于上面的分析, 我们仍然取 $K_m = \mathrm{span}\{\mathbf{r}_0, A\mathbf{r}_0, \cdots, A^{m-1}\mathbf{r}_0\}$.

又利用 Arnoldi 过程, 可以找到 K_m 的一组标准正交基 $\{\mathbf{v}_i\}_{i=1}^m$. 将 (3.17) 改写成 $AV_m = V_{m+1}\widetilde{H}_m$, 其中 $\widetilde{H}_m = \begin{pmatrix} H_m \\ h_{m+1,m}\mathbf{e}_m^T \end{pmatrix}$. 这样可得

$$\begin{aligned} \|\mathbf{r}_0 - A\mathbf{z}_m\| &= \|\mathbf{r}_0 - AV_m\mathbf{y}_m\| = \left\|\mathbf{r}_0 - V_{m+1}\widetilde{H}_m\mathbf{y}_m\right\| \\ &= \left\|V_{m+1}(\beta\mathbf{e}_1 - \widetilde{H}_m\mathbf{y}_m)\right\|. \end{aligned}$$

GMRES算法

由于 $V_{m+1}^T V_{m+1} = I$, 所以

$$\|\mathbf{r}_0 - A\mathbf{z}_m\| = \left\|\beta\mathbf{e}_1 - \widetilde{H}_m\mathbf{y}_m\right\|. \tag{3.23}$$

即在 $\mathbb{R}^m$ 中极小化 $\|\mathbf{r}_0 - A\mathbf{z}_m\|$ 等价于在 K_m 中极小化 $\left\|\beta\mathbf{e}_1 - \widetilde{H}_m\mathbf{y}_m\right\|$.

这样可以把广义极小化残差算法总结如下:

算法 3.7 (GMRES)

1. 选择 $\mathbf{x}_0 \in \mathbb{R}^n$ 并计算 $\mathbf{r}_0 = \mathbf{b} - A\mathbf{x}_0$, 记 $\|\mathbf{r}_0\| = \beta$, $\mathbf{v}_1 = \mathbf{r}_0/\|\mathbf{r}_0\|$.
   对于 $m = 1, \cdots$ 执行以下步骤2-4，直到收敛
2. 用 Arnoldi 过程求出 $\{\mathbf{v}_i\}_{i=1}^m$ 和 $\widetilde{H}_m$ (假设过程中 $\mathbf{r}_i \neq 0$).
3. 求解最小二乘问题 $\min\limits_{\mathbf{y}_m \in \mathbb{R}^m} \left\|\beta\mathbf{e}_1 - \widetilde{H}_m\mathbf{y}_m\right\|$ 得到 $\mathbf{y}_m$.
4. 计算 $\mathbf{x}_k = \mathbf{x}_0 + V_m\mathbf{y}_m$.

GMRES算法

从理论上讲, 如果 $\{A^i\mathbf{r}_0\}_{i=0}^{m-1}$ 线性无关, 当 $m=n$ 时 GMRES 算法应当给出准确解. 但 m 很大时, 计算中需要保存所有的 $\{\mathbf{v}_i\}_{i=1}^{m}$. 对于大型问题 ($n \gg 1$), 所需存贮量显然太大.

为了克服上面的困难, 可以采用如下循环GMRES算法:

GMRES算法

算法 3.8 (GMRES(m) 算法)

1. 选择 $\mathbf{x}_0 \in \mathbb{R}^n$ 并计算 $\mathbf{r}_0 = \mathbf{b} - A\mathbf{x}_0$, 记 $\|\mathbf{r}_0\| = \beta$, $\mathbf{v}_1 = \mathbf{r}_0/\|\mathbf{r}_0\|$.
   选择适当大小的 m.
2. 用 Arnoldi 过程求出 $\{\mathbf{v}_i\}_{i=1}^m$ 和 $\widetilde{H}_m$ (假设过程中 $\mathbf{r}_i \neq 0$).
3. 求解最小二乘问题 $\min\limits_{\mathbf{y}_m \in \mathbb{R}^m} \left\|\beta\mathbf{e}_1 - \widetilde{H}_m\mathbf{y}_m\right\|$ 得到 $\mathbf{y}_m$.
4. 计算 $\mathbf{x}_m = \mathbf{x}_0 + V_m\mathbf{y}_m$, $\mathbf{r}_m = \mathbf{b} - A\mathbf{x}_m$, $\mathbf{r}_0 = \mathbf{r}_m$.
5. 如果 $\beta = \|\mathbf{r}_m\| < \varepsilon$, 则停止迭代;
   否则 $\mathbf{x}_0 = \mathbf{x}_m$, $\mathbf{v}_1 = \mathbf{r}_m/\|\mathbf{r}_m\|$, 转第 2 步.

在该算法中, 如何选取合适的 m 及如何极小化 $\left\|\beta\mathbf{e}_1 - \widetilde{H}_m\mathbf{y}_m\right\|$ 是两个关键问题. 一般来说最有效方法是通过平面旋转把上Hessenberg矩阵 $\widetilde{H}_m$ 逐步变成上三角阵, 即可立即得到最小二乘问题的解.