3.线性方程组的直接解法

线性方程组的解法分为**直接解法**与**迭代解法。**

3.1基本概念

- **向量的范数**：记为$||\cdot||$

  向量的范数具有**正定性、正齐次性、三角不等式。**

  对于数域$\mathbb{K}$上的线性空间S，若定义了范数，则称为**赋范线性空间。**

  $p-$**范数**：$\mathbf{x}=[x_1,x_2,\cdots,x_n]^T$，$||\mathbf{x}||_p=(\sum_{i=1}^n|x_i|^p)^{\frac{1}{p}}$，$p>1$。

  内积范数：针对**内积**$\langle x,y\rangle$，定义范数$||\mathbf{x}||=\sqrt{\langle\mathbf{x},\mathbf{x}\rangle}$。

  三种常用范数：

  1. **1-范数**：$||\mathbf{x}||_1=\sum_{i=1}^n|x_i|$，也被称为**曼哈顿范数**
  2. **2-范数**：$||\mathbf{x}||_2=\sqrt{\sum_{i=1}^n|x_i|^2}$，也被称为**欧式范数/内积范数**
  3. $\infty-$**范数**：$||\mathbf{x}||_\infty=\max_{1\leq i\leq n}|x_i|$，也被称为**"最大"范数**

  Theorem 3.6：存在$f\in C^1$，使得$||\mathbf{x}||=f(x_1,x_2,\cdots,x_n)$。

  Theorem 3.7（**不同范数的等价性**）：$c_1||\mathbf{x}||_s\leq||\mathbf{x}||_t\leq c_2||\mathbf{x}||_s$，其中$c_1,c_2>0$为与$\mathbf{x}$无关的常数。

  根据*Theorem 3.7*，**在某种范数下成立的一些结论在其他范数下也成立。**

  Theorem 3.8：$\lim_{k\to\infty}\mathbf{x}^{(k)}=\mathbf{x}^*\Leftrightarrow\lim_{k\to\infty}||\mathbf{x}^{(k)}-\mathbf{x}^*||$

  > 根据*Theorem 3.7*，为证明*Theorem 3.8*，只需说明其对于$\infty-$范数成立即可。而这是显然的。

- **矩阵的范数**

  对$\mathbb{R}^{n\times n}$空间的**矩阵范数**，需要增加一些条件：

  1. 增加对矩阵乘法的要求：$||\mathbf{AB}||\leq||\mathbf{A}||\cdot||\mathbf{B}||$
  2. **相容性条件**：$||\mathbf{Ax}||\leq||\mathbf{A}||\cdot||\mathbf{x}||$

  根据某种向量范数$||\mathbf{x}||_v$，定义矩阵的**算子范数**（也称为**向量诱导范数**）：

  $$||\mathbf{A}||_v=\max_{\mathbf{x}\neq\mathbf{0}}\frac{||\mathbf{Ax}||_v}{||\mathbf{x}||_v},\quad x\in\mathbb{R}^n,A\in\mathbb{R}^{n\times n}$$

  直观理解，也即$\mathbf{Ax}$对向量$\mathbf{x}$的最大拉伸倍数（也可能<1）。对于$||\mathbf{x}||_v=1$，$\mathbf{x}$的端点轨迹为二维空间中的单位圆，而$\mathbf{Ax}$向量端点的轨迹为**椭圆**，$||\mathbf{A}||_v$就是这个椭圆**半长轴**的长度。矩阵范数的相容性条件显然成立，对矩阵乘法要求的证明如下：

  $$\begin{aligned}||\mathbf{AB}||_v&=\max_{\mathbf{x}\neq\mathbf{0}}\frac{||\mathbf{ABx}||_v}{||\mathbf{x}||_v}=\max_{\mathbf{x}\neq\mathbf{0}}(\frac{||\mathbf{ABx}||_v}{||\mathbf{Bx}||_v}\cdot\frac{||\mathbf{Bx}||_v}{||\mathbf{x}||_v})\\&\leq\max_{\mathbf{x}\neq\mathbf{0}}\frac{||\mathbf{A(Bx)}||_v}{||\mathbf{Bx}||_v}\cdot\max_{\mathbf{x}\neq0}\frac{||\mathbf{Bx}||_v}{||\mathbf{x}||_v}\\&\leq\max_{\mathbf{x}\neq\mathbf{0}}\frac{||\mathbf{Ax}||_v}{||\mathbf{x}||_v}\cdot\max_{\mathbf{x}\neq0}\frac{||\mathbf{Bx}||_v}{||\mathbf{x}||_v}\\&=||\mathbf{A}||\cdot||\mathbf{B}||\end{aligned}$$

  常用的矩阵范数：

  1. **1-范数**：$||\mathbf{A}||_1=\max_{1\leq j\leq n}\sum_{i=1}^n|a_{ij}|$（竖列元素绝对值之和的最大值）

2. ∞**-范数**：$||\mathbf{A}||_\infty = \max_{1\le i\le n}\sum_{j=1}^n |a_{ij}|$
3. **2-范数**：$||\mathbf{A}||_2 = \sqrt{\lambda_{\max}(\mathbf{A}^T\mathbf{A})}$，其中，$\lambda_{\max}$表示取矩阵的**最大特征值。**
4. **Frobenius范数**：$||\mathbf{A}||_F = \sqrt{\sum_{i=1}^n\sum_{j=1}^n a_{ij}^2} = \sqrt{tr(\mathbf{A}^T\mathbf{A})}$

3.2问题的敏感性

$$\mathbf{Ax} = \mathbf{b} \Longrightarrow \mathbf{A}(\mathbf{x}+\Delta\mathbf{x}) = \mathbf{b}+\Delta\mathbf{b}$$

条件数$\text{cond} = \frac{||\Delta\mathbf{x}||/||\mathbf{x}||}{||\Delta\mathbf{b}||/||\mathbf{b}||} \le ||\mathbf{A}||\cdot||\mathbf{A}^{-1}||$

> 这是因为，$\Delta\mathbf{x} = \mathbf{A}^{-1}\Delta\mathbf{b}$，从而有$||\Delta\mathbf{x}|| \le ||\mathbf{A}^{-1}||\cdot||\Delta\mathbf{b}||$，同理有$||\mathbf{b}|| \le ||\mathbf{A}||\cdot||\mathbf{x}||$。

上式也被称为矩阵的条件数。设$\mathbf{A}$为**非奇异矩阵**，则定义矩阵$\mathbf{A}$的条件数$\text{cond}(\mathbf{A}) = ||\mathbf{A}||\cdot||\mathbf{A}^{-1}||$。

矩阵的条件数是**反映线性方程组求解问题的敏感性的条件数的上界**。如果矩阵条件数大，则问题很**病态**，此时称该矩阵为**病态矩阵**；反之则称问题为**良态**。

Theorem 3.11：在任一算子范数下，有：

$$\text{cond}(\mathbf{A}) = \max_{\mathbf{x}\ne\mathbf{0}}\frac{||\mathbf{Ax}||}{||\mathbf{x}||}\Big/\min_{\mathbf{x}\ne\mathbf{0}}\frac{||\mathbf{Ax}||}{||\mathbf{x}||}$$

这是因为，$||\mathbf{A}^{-1}|| = \max_{\mathbf{x}\ne\mathbf{0}}\frac{||\mathbf{A}^{-1}\mathbf{x}||}{||\mathbf{x}||} = \max_{\mathbf{x}\ne\mathbf{0}}\frac{||\mathbf{x}||}{||\mathbf{Ax}||} = \frac{1}{\min_{\mathbf{x}\ne\mathbf{0}}\frac{||\mathbf{Ax}||}{||\mathbf{x}||}}$。这体现出$\text{cond}(\mathbf{A})$的几何意义，也即**对单位圆的扭曲程度**，同时也说明$\text{cond}(\mathbf{A}) \ge 1$。此外，定义$\text{cond}(\text{奇异矩阵}) = +\infty$。

Theorem 3.12：矩阵条件数具有如下性质：

1. $\text{cond}(\mathbf{A}) \ge 1$，$\text{cond}(\mathbf{A}^{-1}) = \text{cond}(\mathbf{A})$，
   $\text{cond}(c\mathbf{A}) = \text{cond}(\mathbf{A}),\ \forall c \ne 0$
2. $\text{cond}(\mathbf{I}) = 1$
3. $\mathbf{D}$为对角阵，则在p-范数意义下，$\text{cond}(\mathbf{D}) = \frac{\max_i|d_{ii}|}{\min_i|d_{ii}|}$。
4. 采用2-范数，$\text{cond}(\mathbf{A}) = \sqrt{\frac{\lambda_{\max}(\mathbf{A^T}A)}{\lambda_{\min}(\mathbf{A^T}A)}}$
5. $\mathbf{Q}$为正交矩阵，则有
   $\text{cond}(\mathbf{Q})_2 = 1$
   $\text{cond}(\mathbf{QA})_2 = \text{cond}(\mathbf{AQ})_2 = \text{cond}(\mathbf{A})_2$

   > 对于正交阵$\mathbf{Q}$，有$||\mathbf{Qa}|| = ||\mathbf{a}||$
   > $\text{cond}(\mathbf{QA})_2 = \max_{\mathbf{x}\ne\mathbf{0}}\frac{||\mathbf{QAx}||}{||x||} = \max_{\mathbf{x}\ne\mathbf{0}}\frac{||\mathbf{Ax}||}{||\mathbf{x}||} = \text{cond}(\mathbf{A})_2$

3.3矩阵的LU分解

高斯消去法：用初等变换将$\mathbf{A}$变为单位阵，用于算**逆矩阵。**

Theorem 3.14：对方程$\mathbf{Ax} = \mathbf{b}$，其中$\mathbf{A} \in \mathbb{R}^{n\times n}$，若执行高斯消去过程中的主元 $a_{kk}^{(k)} \ne 0, (k = 1, 2, \cdots, n-1) \Longleftrightarrow$系数矩阵$\mathbf{A}$**存在唯一**的**LU分解。**

唯一性的证明：假设$\mathbf{A} = \mathbf{L_1U_1} = \mathbf{L_2U_2}$，其中$\mathbf{L_1}, \mathbf{L_2}$为单位下三角阵（非奇异）。若$\mathbf{A}$非奇异，则$\mathbf{U_1}$也非奇异，从而有$\mathbf{L_1L_2^{-1}} = \mathbf{U_2U_1^{-1}} = I \Longrightarrow \mathbf{L_1} = \mathbf{L_2}, \mathbf{U_1} = \mathbf{U_2}$，矛盾。若$\mathbf{A}$奇异，则$\mathbf{U}(n, n) = 0$，从而有

$$\begin{bmatrix}\mathbf{L_1}' & \mathbf{0} \\ \alpha_1^T & 1\end{bmatrix}\begin{bmatrix}\mathbf{U_1}' & \beta_1 \\ \mathbf{0}^T & 0\end{bmatrix} = \begin{bmatrix}\mathbf{L_2}' & \mathbf{0} \\ \alpha_2^T & 1\end{bmatrix}\begin{bmatrix}\mathbf{U_2}' & \beta_2 \\ \mathbf{0}^T & 0\end{bmatrix}$$

进而有$\mathbf{L_1}'\mathbf{U_1}' = \mathbf{L_2}'\mathbf{U_2}' \Longrightarrow \mathbf{L_1}' = \mathbf{L_2}', \mathbf{U_1}' = \mathbf{U_2}' \Longrightarrow \mathbf{L_1} = \mathbf{L_2}, \mathbf{U_1} = \mathbf{U_2}$，矛盾。

LU分解的用途：

- **单个方程求解**：

  $\mathbf{Ax} = \mathbf{b} \Rightarrow \mathbf{x} = (\mathbf{LU})^{-1}\mathbf{b} = \mathbf{U^{-1}L^{-1}}b$。

  1. 首先求解**单位下三角型方程组**$\mathbf{Ly} = \mathbf{b}$
  2. 再求解**上三角型方程组**$\mathbf{Ux} = \mathbf{y}$
- **右端项变化的问题**：

  $\mathbf{Ax_i} = \mathbf{b_i}$，只需对每个右端项执行上述两步，而不需要重复计算LU分解。

3.4选主元技术与稳定性

Theorem 3.15：对矩阵$\mathbf{A} \in \mathbb{R}^{n\times n}$，高斯消去过程中**不出现零主元**的**充要条件**是，$\mathbf{A}$的前$n-1$个顺序主子式均不为零，即$\det(A_k) \neq 0, (k = 1, 2, \cdots, n-1)$。

> 需要注意的是，如果高斯消去过程中**不出现零主元**，则**LU分解存在且唯一**，但反过来，LU分解存在不能说明**高斯消去过程一定没有零主元**。例如，矩阵$\begin{bmatrix}0 & 0 \\ 0 & 1\end{bmatrix}$就有无穷多个LU分解，但是高斯消去过程存在零主元。不过，**LU分解存在且唯一**与**高斯消去过程中不出现零主元**等价。
>
> 并且，LU分解的存在性与唯一性与$\mathbf{A}$是否奇异无关。**对奇异矩阵$\mathbf{A}$也可能完成消去、LU分解**，例如$\mathbf{A} = \begin{bmatrix}1 & 1 \\ 1 & 1\end{bmatrix}$（关键在于LU分解不要求最后一个主元非零），而**非奇异矩阵也不一定有LU分解**（可能需要先进行一些初等行变换）。

通过**选主元技术**，可以解决主元为零的问题，同时，通过将同一列中的最大元素换做主元，还可以**减小数值误差**，从而增强算法稳定性。

部分主元的LU分解：

$$\mathbf{PA} = \mathbf{LU}$$

其中，$\mathbf{P} = \mathbf{P_{n-1}P_{n-2}}\cdots\mathbf{P_1}$称为**排列阵**，其为单位阵$\mathbf{I}$重排的结果。

> 数值更稳定的**全选主元**：选取未消去子矩阵中的最大元素，通过**行**、**列**交换到主元位置。
>
> $$\mathbf{PAQ} = \mathbf{LU}$$

算法的稳定性：

高斯消去法主要受**舍入误差**的影响。通过**向后误差分析**，我们可以得到其误差的一个上界。设方程$\mathbf{Ax} = b$的数值解为$\hat{\mathbf{x}}$，且$(\mathbf{A} + \Delta\mathbf{A})\hat{\mathbf{x}} = \mathbf{b}$，有

$$\frac{||\Delta\mathbf{A}||_\infty}{||\mathbf{A}||_\infty} \lesssim n\rho\epsilon_{mach}$$

其中，ρ为增长因子，表示$\mathbf{A}^{(k)}$（第k步高斯消去法后得到的矩阵）与$\mathbf{A}$最大元素之比。对于部分选主元方法，有$\rho \leq 2^{n-1}$，但实际往往远小于这个上界；对于不选主元的方法，ρ可能任意大。

3.5对称正定矩阵的*Cholesky*分解

> 正定矩阵：**所有顺序主子式的行列式均为正**；半正定矩阵：**所有顺序主子式的行列式均非负**

对于对称矩阵$\mathbf{A}$，有

$$\mathbf{A}=\mathbf{LDU_0}=\mathbf{A}^T=\mathbf{U_0^T DL^T}$$

从而若高斯消去不中断，LU分解唯一，有$\mathbf{L}=\mathbf{U_0^T}$。

对于**对称正定矩阵**$\mathbf{A}$，其LU分解一定存在且唯一，故$\mathbf{A}=\mathbf{LDL^T}$唯一存在，且$\mathbf{D}=\mathbf{L}^{-1}\mathbf{A}\mathbf{L^T}$也为对称正定矩阵，其对角元均大于零。故有：

$$\mathbf{A}=\mathbf{LD^{\frac{1}{2}}D^{\frac{1}{2}}L^T}=\mathbf{L_1L_1^T}$$

此称为***Cholesky*分解**，由于其分解为一组对称阵的乘积，存储量可以节省一半。

***Cholesky*分解算法（平方根法）**

与直接LU分解的思想类似。

$$\begin{bmatrix} l_{11} & 0 & \cdots & 0 \\ l_{21} & l_{22} & \ddots & \vdots \\ \vdots & \vdots & \ddots & 0 \\ l_{n1} & l_{n2} & \cdots & l_{nn} \end{bmatrix}\begin{bmatrix} l_{11} & l_{21} & \cdots & l_{n1} \\ 0 & l_{22} & \cdots & l_{n2} \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & l_{nn} \end{bmatrix}=\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}$$

考虑对应位置元素相等，有（**原地存$\mathbf{L}$的结果**）

$$a_{jj}=\sqrt{a_{jj}-\sum_{k=1}^{j-1}a_{jk}^2}\quad j=1,2,\cdots,n$$

$$a_{ij}=(a_{ij}-\sum_{k=1}^{j-1}a_{ik}a_{jk})/a_{jj}\quad i=j+1,\cdots,n$$

算法的稳定性：

考虑**增长因子**

$$\rho=\frac{\max\{|U^T(i,j)|\}}{\max\{|a_{ij}|\}}=\frac{\max\{|l_{ij}l_{jj}|\}}{\max\{|a_{ij}|\}}\leq\frac{\max\{l_{ij}^2\}}{\max\{a_{ii}\}}\leq 1$$

注：由上平方根算法，可知**$\mathbf{A}$的对角元是$\mathbf{L}$一行元素的平方和。**

3.6带状矩阵解法与稀疏矩阵简介

3.6.1带状矩阵

矩阵$\mathbf{A}=(a_{ij})_{n\times n}$，若$\forall i,j$，$|i-j|>\beta$时都有$a_{ij}=0$，且$\exists k,a_{k,k-\beta}\neq 0$或$a_{k,k+\beta}\neq 0$，则称$\mathbf{A}$为**半带宽**为$\beta$的**带状矩阵**。对于带状矩阵，其LU分解中，$\mathbf{L}$、$\mathbf{U}$矩阵的**非零元**依然分布在**原始带宽范围内**。

> 注意，对于带状矩阵的LU分解，其$\mathbf{A}^{-1},\mathbf{L}^{-1},\mathbf{U}^{-1}$均稠密，因此应避免再计算逆矩阵。

按行对角占优矩阵：$|a_{ii}|\geq\sum_{j=1,j\neq i}^n|a_{ij}|$，$i=1,2,\cdots,n$，且至少有一个大于号成立。（若均为严格大于，则称为**严格对角占优矩阵**）。

Theorem 3.20：**按列严格对角占优阵**，列主元LU分解不需要交换行。

3.6.2稀疏矩阵

- **三元组**（COO）：值、行下表、列下标
- **压缩稀疏行**（CSR）：值、（按行顺序存储的）列下标的列表、每行开头在列表中的位置

  CSR是为了解决COO中连续存储的元素有相同的行（列）编号的问题
- **压缩稀疏列**（CSC）
- **若干个一维数组**：针对带状矩阵，每个数组存储一条带
- **分块压缩稀疏行**：针对分块矩阵

在进行稀疏矩阵有关的计算时，不存储零元素，**只需遍历所有非零元。**而对稀疏矩阵做高斯消去时，填入的元素可能会造成稀疏矩阵存储结构的变化。