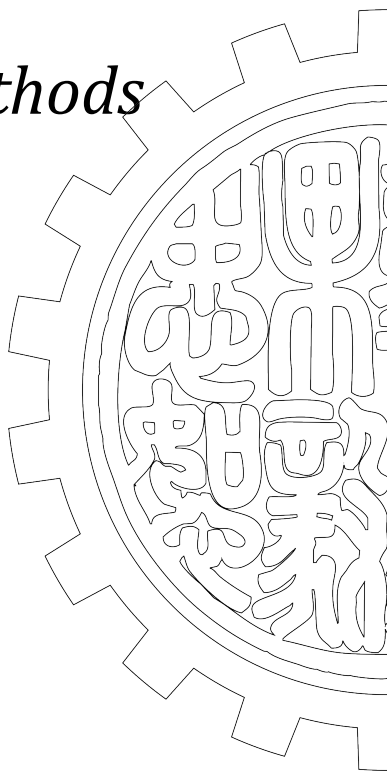
计算方法笔记

Notes on Computing Methods

作者：王天浩

2019 年 6 月 6 日



钱学森书院学业辅导中心

Qian Yuan Xue Fu

XI'AN JIAOTONG UNIVERSITY

作品信息

➤ **标题：**计算方法笔记 - *Notes on Computing Methods*
➤ **作者：**王天浩
➤ **校对排版：**xjtu-blacksmith
➤ **出品时间：**2019 年 6 月 6 日
➤ **总页数：**19

许可证说明

本作品已发布于 GitHub 之上，发布地址为：

`https://github.com/qyxf/Tutorials`

本作品的版本号为。

前言

计算方法是本校钱学森班、少年班等试验班的同学必修的一门数学基础课程，其内容包括解决一些计算问题（如解线性方程组、插值与逼近、数值微积分等）的常用算法，以及针对这些算法之误差与稳定性的相关数学理论。相较于其他的数学课程，计算方法课程具有内容丰富、实用性强等特点，但也兼有一般数学课程内容深、考点多的性质。可以说，计算方法课程是工科学生在走向专业课程之前所要翻越的最后一座数学理论“大山”。

基于以上的特点，在计算方法课程的学习与复习过程中，便非常需要一本**详略适中**、**重点突出**的课程笔记，以资参考。2014 级少年班的王天浩同学，于此方面做了相当好的工作，在学习本课程期间整理了十分详尽的课程笔记，基本上达到了以上所提到的要求。……

为方便之后的同学复习计算方法课程，经与王天浩学长协商，我自 2019 年 6 月 6 日开始将原有的纸质扫描笔记以 LaTeX 整理为电子版，补足了语言上的一些省略、缺失，并增写了许多注记、说明。……

正文格式说明

电子版笔记基本上遵循了原有纸质笔记的框架，但相较与原来的样式更为清晰、简明。这份笔记的正文中包含了这样几类内容：

零碎的知识点 用无序号列表的形式列出。

例题 以加粗的“**例**”字引导，通常是对其正上方那一个或几个知识点的具体呈现与演示。

注记 用脚注的方式给出，通常是对正文内容的进一步阐释。大部分注记标明了“——编者注”的字样，这些都是编者所加上的。

Qian Yuan Xue Fu

帮助我们改进这份笔记

一本好的教科书，来自于相关教师历经数代、数十年的逐次再版改进；一份好的笔记，同样也需要长期的维护、改进才能够最终创造出来。……

能动少 C71 尤佳睿（黑山雁）[1]

2019 年 6 月 6 日

[1] 个人博客：https://www.cnblogs.com/xjtu-blacksmith/

目录

第一章　误差

§1.1　真值与误差

➢**1.1** 有测量就会有误差。通常，将某数学量、物理量的**真值**记为 x（不加任何修饰符），而将测量或计算所得的 x 的**近似值**记作 $\tilde{x}$。

➢**1.2** 两种误差：$\begin{cases}\Delta x = x - \tilde{x} \text{ (绝对误差)} \\ \delta x = \frac{x-\tilde{x}}{x} \text{ (相对误差)}\end{cases}$

➢**1.3** 两种误差限：$\begin{cases}|\Delta x| \le \varepsilon \text{ (绝对误差限)} \\ |\delta x| \le \varepsilon_r \text{ (相对误差限)}\end{cases}$

➢**1.4** 相对误差较小时，有近似计算式[1]：$|\delta x| = \frac{|x-\tilde{x}|}{\tilde{x}} = \frac{\Delta x}{\tilde{x}} \le \frac{|\varepsilon|}{\tilde{x}}$

➢**1.5** 若 $|\Delta x| = |x - \tilde{x}| \le 0.5 \times 10^{-n}$，则称 x 的近似值 $\tilde{x}$准确到第 n 位小数。

✐**1.6** 设 $x = 0.31682$，则 $\tilde{x}_1 = 0.3$ 精确到 1 位有效数字，$\tilde{x}_2 = 0.32$ 精确到 2 位，$\tilde{x}_3 = 0.317$ 精确到 3 位，$\tilde{x}_4 = 0.3168$ 精确到 4 位。若取 $\tilde{x}_5 = 0.3169$ 为 x 的近似值，则其仅精确到 3 位小数。

§1.2　浮点运算与浮点数集

➢**1.7** 在计算机中，实数将被储存为**浮点数**，故计算机中的实数运算常被称作**浮点运算**。为此，有下面的一些概念与理论。

➢**1.8** **浮点运算量**：记一次加法和一次乘法（如 $a + b \times c$）所需的时间为一个**时间单位**，记为 flop。

✐**1.9** 设 $\mathbf{A}_1$ 为一 10×20 的矩阵，$\mathbf{A}_2$ 为一 20×50 的矩阵，欲计算 $\mathbf{A}_1 \cdot \mathbf{A}_2$，则运算量为 $10 \times 20 \times 50 = 10000$ flop，如图1.1所示。

➢**1.10** **浮点数集**：在 10 进制中，浮点数 $\tilde{x}$（或一实数 x 的近似 t 位有效数字的浮点数 $\tilde{x}$）可表示如下：

$$fl(x) = \tilde{x} = \pm\left\{\frac{x_1}{10} + \frac{x_2}{10^2} + \frac{x_3}{10^3} + \cdots + \frac{x_t}{10^t}\right\} \times 10^l \ (\tilde{x} = 0.x_1x_2x_3\cdots x_t \times 10^l)$$

[1]在估计误差时，真值 $\tilde{x}$ 往往难以确定，但绝对误差 $|\Delta x|$ 或绝对误差限 ε 往往能够确定下来。

图 1.1: 矩阵乘法运算量示意图

其中 $1 \le x_1 < 10$，$0 \le x_j < 10$，$j = 2, 3, \cdots, t$。类似的，在 β 进制中，一个数的表示方式：

$$fl(x) = \tilde{x} = \pm\left\{\frac{x_1}{\beta} + \frac{x_2}{\beta^2} + \cdots + \frac{x_t}{\beta^t}\right\} \times \beta^l$$

其中 $1 \le x_1 < \beta$，$0 \le x_j < \beta$，$j = 2, 3, \cdots, t$。β^l 称为指数部分，指数 l 满足 $L \le l \le U$，L 与 U 分别为下界与上界；$x_1, x_2 \cdots, x_t$ 称为位数。$fl(x)$ 称为一个**规格化浮点数**。

➢**1.11** 称计算机中所能表示的全体数的集合称为**浮点数集**，记为 $F(\beta, t, L, U)$。

$$F(\beta, t, L, U) = \{0\} \cup \left\{\pm\left(\frac{x_1}{\beta} + \frac{x_2}{\beta^2} + \cdots + \frac{x_t}{\beta^t}\right) \times \beta^l : L \le l \le U\right\} \tag{1.1}$$

✎**1.12** C++ 里的`float`：4 字节内存，32 个二进制 bit，如图1.2所示。可以将这一浮点数集记为 $F(2, 23, -128, 127)$。

图 1.2: C++ 中 `float` 类型变量的储存原理

➢**1.13** 浮点数集中的数的个数：$N = 2 \cdot (\beta - 1) \cdot \beta^t \cdot (U - L + 1) + 1$

➢**1.14** 浮点数 $fl(x)$ 与对应真值 x 的误差：

- 绝对误差：$|x - fl(x)| \le \frac{1}{2}\beta^{-t} \times \beta^l = \frac{1}{2}\beta^{l-t}$
- 相对误差：$\frac{|x - fl(x)|}{|x|} \le \frac{\beta^{l-t}/2}{\beta^{l-1}} = \frac{1}{2}\beta^{1-t}$ $(|x| \ge 0.1 \times \beta^l)$

此类误差称为**舍入误差**。

➢**1.15** 计算结果的错误/误差：

1. $l \notin [L, U]$：**上溢**（$l \ge L$）会出错，**下溢**（$l \le U$）变为 0。

2. 尾数多于 t 位：自动进行舍入处理，造成误差
3. 有效数字丢失（“大数吃小数”）

✍**1.16** 设计算时保留 4 位有效数字,则 $1234+0.3678 = 1234.3678 = 1234.3678 \approx 1234$，在此发生了“大数吃小数”的现象。

➢**1.17** 设计数值计算的算法时，应结合浮点数具有的特性，避免上面所提到的各类计算错误。为此，提出以下几条浮点运算原则：

1. 避免产生大结果的运算，避免小数作为除数。
2. 避免“大”、“小”数相加减，防止大数吃小数。
3. 避免相近数直接相减，防止有效数字损失。
4. 简化运算步骤，减少运算次数 [(2)]。

若原有的计算公式不符合以上的这些原则，则可以通过对原式的等价变换或近似处理，使之符合上面的原则。

✍**1.18** 设 $|x| \ll 1$，则可以简化数值计算公式 $\ln \dfrac{1-\sqrt{1-x^2}}{|x|}$ 为以下形式，以避免小数作分母：

$$\ln \frac{1-\sqrt{1-x^2}}{|x|} = \ln \frac{x^2}{|x| \cdot (1+\sqrt{1-x^2})} = \ln \frac{|x|}{1+\sqrt{1-x^2}}$$

§1.3 计算方法的研究内容

➢**1.19** 计算方法课程，并不仅仅包含各类数值计算方法。归结而言，计算方法课程的研究内容可以归纳为：

1. 某一问题的数值计算算法（即通常意义上的“计算方法”）；
2. 这些算法的误差、复杂性或收敛速度之估计。

而后者至关重要，对于算法的误差或复杂性分析使这门课程区别于一般的工具性课程。

➢**1.20** 针对一些模型，还存在着一类特定的问题，即**病态问题**。度量这类问题的性质，需要用到**条件数**。

- 根据输入数据的微小变化能引起问题之解变化的大小程度，可以将数值计算问题区别为两类：若由此能引起解的很大变化，则称问题是**病态**的；否则，称一个问题是**良态**的。病态问题不易精确求解。

[(2)] 由此避免各类误差的逐次累计。——编者注

- **条件数**：输入数据 $x,\tilde{x}$，输出 $f(x),f(\tilde{x})$。设 $x\neq 0$，$f(x)\neq 0$，若存在 $m>0$ 使：

$$\frac{|f(x)-f(\tilde{x})|}{|f(x)|}\leq m\cdot\frac{|x-\tilde{x}|}{|x|}\ (\text{输出误差}\leq m\cdot\text{输入误差})$$

  则将 m 称为该问题的**条件数**，记为 $\mathrm{Cond}(f)$。

✎**1.21** $y=\varphi(x_1,x_2,\cdots,x_n)$。输入为 $\tilde{x}_1,\tilde{x}_2,\cdots,\tilde{x}_n$，近似解 $\tilde{y}=\varphi(\tilde{x}_1,\tilde{x}_2,\cdots,\tilde{x}_n)$，则有

$$\Delta y=\varphi(x_1,x_2,\cdots,x_n)-\varphi(\tilde{x}_1,\tilde{x}_2,\cdots,\tilde{x}_n)\approx\sum_{i=1}^{n}\frac{\partial\varphi(\tilde{x}_1,\tilde{x}_2,\cdots,\tilde{x}_n)}{\partial x_i}\cdot\frac{x_i}{y}\cdot\Delta x_i \tag{1.2}$$

$$\delta y=\frac{\Delta y}{y}\approx\sum_{i=1}^{n}\frac{\partial\varphi(\tilde{x}_1,\tilde{x}_2,\cdots,\tilde{x}_n)}{\partial x_i}\cdot\frac{x_i}{y}\cdot\frac{\Delta x_i}{x_i}=\sum_{i=1}^{n}\frac{\partial\varphi(\tilde{x}_1,\tilde{x}_2,\cdots,\tilde{x}_n)}{\partial x_i}\cdot\frac{x_i}{y}\cdot\delta x_i \tag{1.3}$$

故可见 $\left|\frac{\partial\varphi(\tilde{x}_1,\tilde{x}_2,\cdots,\tilde{x}_n)}{\partial x_i}\cdot\frac{x_i}{y}\right|$ 即条件数。

➢**1.22** **稳定性**（数值稳定性）：运算中**舍入误差积累**是否影响结果的可靠性。

✎**1.23** 欲用数值计算方法求解由 $I_k=\mathrm{e}^{-1}\int_0^1 x^k\mathrm{e}^x\mathrm{d}x,\ k=0,1,\cdots,7$ 所定义的一系列定积分的值。

- 算法 1：构建递推公式

$$\begin{cases}I_0=\mathrm{e}^{-1}\displaystyle\int_0^1\mathrm{d}x=1-\frac{1}{\mathrm{e}}\\ I_k=\mathrm{e}^{-1}\displaystyle\int_0^1 x^k\mathrm{e}^x\mathrm{d}x=\mathrm{e}^{-1}x^k\cdot\mathrm{e}^{-1}\Big|_0^1-\mathrm{e}^{-1}\int_0^1 k\cdot\mathrm{e}^x x^{k-1}\mathrm{d}x=1-kI_{k-1}\end{cases} \tag{1.4}$$

  利用递推关系依次计算 $I_0\to I_1\to I_1\to\cdots\to I_7$。

- 算法 2：近似计算 I_7，利用递推关系 **(3)** 依次计算 $I_7\to I_6\to\cdots\to I_0$。

实际上就整体而言，算法 2 精度更高。对算法 1 递推公式 $I_k=1-kI_{k-1}$。若 I_{k-1} 有舍入误差 ΔI_{k-1}（或记作 Δ），则 $\tilde{I}_k=1-k(I_{k-1}+\Delta)=I_k-k\cdot\Delta$，误差被放大 **(4)**。

---

(3) 即将 (1.4) 式移项，反得 $I_{k-1}=\frac{1-I_k}{k}$。——编者注

(4) 对算法 2 的递推公式做类似分析，可见 $\tilde{I}_{k-1}=I_{k-1}-\Delta/k$，即误差被减小到原来的 $1/k$ 倍，这是大大缩小了。故算法 2 较算法 1 更为稳定。——编者注

第二章　线性方程组直接解法

§2.1　Gauss 消元法的引入

➢$_{2.1}$ 整体思路：

1. 先推得 $\mathbf{Ax}=\mathbf{b}\ \Rightarrow\ \mathbf{x}=\mathbf{A}^{-1}\mathbf{b}$，再求 $\mathbf{A}^{-1}$（初等行变换法、伴随矩阵法、Gauss-Jordan 消去法）
2. Crammer 法则：$\mathbf{Ax}=\mathbf{b}\ \Rightarrow\ x_i=\frac{|\mathbf{A}_i|}{|\mathbf{A}|}$，浮点运算数 $N=(n^2-1)\cdot n!+n$ flop（很大）

➢$_{2.2}$ Gauss 消去法：降维（$n\to n-1\to\cdots\to 1$）

图 2.1: Gauss 消去法步骤示意图

- 消去运算量：$N_1=\sum_{k=1}^{n-1}(n-k)(n-k+2)=\frac{n^3}{3}+n^2-\frac{5n}{6}$
- 回代运算量：$N_2=1+2+\cdots+n=\frac{n(n+1)}{2}$
- 总计运算量：$N=N_1+N_2=\frac{n^3}{3}+n^2-\frac{n}{3}=O(n^3)$

➢$_{2.3}$ 可能出现的问题：（主要是消去过程中）

1. $a_{kk}^{(k-1)}=0$，无法进行
2. $|a_{kk}^{(k-1)}|\ll|a_{ik}^{(k-1)}|\ (i=k+1,k+2,\cdots,n)$，误差极大（大/小 = 大，误差被放大）

对 Problem 1，只要满足：(1)**A** 是方阵；(2)$|\mathbf{A}|\neq 0$，则可通过换行达到解决问题。

☞**2.4** $a_{kk}^{(k-1)}$ 不为 0 的充要条件是 **A** 的 1 阶与 k 阶主子式均不为 0，即

$$a_{kk}^{(k-1)} \neq 0 \Leftrightarrow D_1 = a_{11} \neq 0,\ D_k = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1k} \\ a_{21} & a_{22} & \cdots & a_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{k1} & a_{k2} & \cdots & a_{kk} \end{vmatrix}$$

对 Problem 2，则不易解决[(1)]。

☞**2.5** 设矩阵 **A** 满足 $\sum_{j=1,j\neq i}^{n} |a_{ij}| < |a_{ii}|$，则称 **A** 是严格对角占优矩阵。

➢**2.6** Gauss 消去法顺利进行条件：

1. **A** 各阶顺序主子式不等于 0。
2. **A** 是对称正定阵。
3. **A** 是严格对角占优矩阵。

§2.2 Gauss 消元法的改进

➢**2.7** 列主元 Gauss 消元法：消去进行到第 k 步时如下所示：

$$\begin{pmatrix} a_{11}^{(k-1)} & a_{12}^{(k-1)} & \cdots & a_{1k}^{(k-1)} & \cdots & a_{1n}^{(k-1)} \\ 0 & a_{22}^{(k-1)} & \cdots & a_{2k}^{(k-1)} & \cdots & a_{2n}^{(k-1)} \\ \vdots & \vdots & & \vdots & & \vdots \\ 0 & 0 & \cdots & a_{kk}^{(k-1)} & \cdots & a_{kn}^{(k-1)} \\ \vdots & \vdots & & \vdots & & \vdots \\ 0 & 0 & \cdots & a_{nk}^{(k-1)} & \cdots & a_{nn}^{(k-1)} \end{pmatrix}$$

选取 $\max(|a_{ik}^{(k-1)}|)\ i = k, k+1, \cdots, n$ 的一行与第 k 行互换，继续消去（算法较稳定）

➢**2.8** Gauss 消去法矩阵形式：

$$\mathbf{A} = \mathbf{A}^{(0)} = \begin{pmatrix} * & * & * & * \\ * & * & * & * \\ * & * & * & * \\ * & * & * & * \end{pmatrix} \Rightarrow \mathbf{A}^{(1)} = \begin{pmatrix} * & * & * & * \\ & * & * & * \\ & * & * & * \\ & * & * & * \end{pmatrix}$$

[(1)]可参见下一节中的“列主元 Gauss 消元法”

则 $\mathbf{A}^{(1)}=\mathbf{L}_1\mathbf{A}^{(0)}$，其中 $L_1=\begin{pmatrix}1&0&\cdots&0\\-l_{21}&1&\cdots&0\\\vdots&\vdots&\ddots&\vdots\\-l_{n1}&0&\cdots&1\end{pmatrix}$。按此方式依次变换得

$$\mathbf{A}^{(n)}=\mathbf{L}_n\mathbf{A}^{(n-1)}=\cdots=\mathbf{L}_n\mathbf{L}_{n-1}\cdots\mathbf{L}_2\mathbf{L}_1\mathbf{A}^{(0)}$$

反推得到

$$\mathbf{A}=\mathbf{L}_1^{-1}\mathbf{L}_2^{-1}\cdots\mathbf{L}_{n-1}^{-1}\mathbf{A}^{(n-1)}$$

记 $\mathbf{U}=\mathbf{A}^{(n-1)}$，则 $\mathbf{A}=\mathbf{L}_1^{-1}\mathbf{L}_2^{-1}\cdots\mathbf{L}_{n-1}^{-1}\mathbf{U}=\mathbf{LU}$

☞**2.9** 设 $\mathbf{A}$ 为 n 阶矩阵，$D_k\neq 0$，则 $\mathbf{A}$ 可唯一分解为一单位下三角阵 $\mathbf{L}$ 与一上三角阵 $\mathbf{U}$ 之积

$$\mathbf{A}=\mathbf{LU} \tag{2.1}$$

称为 **LU 分解**（Doolittle 分解）。

➢**2.10** LU 分解的算法实现：根据式 (2.1) 可知，$\mathbf{A}$ 中的元素 α_{ij} 满足

$$\begin{aligned}\alpha_{ij}&=(l_{i1}\ l_{i2}\ \cdots\ l_{i,i-1}\ 1\ 0\ \cdots\ 0)\cdot(\mu_{1j}\ \mu_{2j}\ \cdots\ \mu_{jj}\ 0\ \cdots\ 0)^{\mathrm{T}}\\&=\begin{cases}\sum_{k=1}^{i-1}l_{ik}\mu_{kj}+\mu_{ij} & ,j\geq i,\ i=1,2,\cdots,n\\ \sum_{k=1}^{j}l_{ik}\mu_{kj} & ,j<i,\ i=1,2,\cdots,n\end{cases}\end{aligned}$$

由此可以推得迭代算式为：

$$\mu_{1j}=\alpha_{1j} \qquad j=1,2,\cdots,n \tag{2.2}$$

$$l_{i1}=\alpha_{i1}/\mu 11 \qquad i=2,3,\cdots,n \tag{2.3}$$

$$\mu_{ij}=\alpha_{ij}-\sum_{k=1}^{i-1}l_{ik}\mu_{kj} \qquad j=i,i+1,\cdots,n\ i=2,3,\cdots,n \tag{2.4}$$

$$l_{ki}=(\alpha_{ki}-\sum_{t=1}^{i-1}l_{kt}\mu_{ti})/\mu_{ii} \qquad k=i+1,\cdots,n\ i=2,3,\cdots,n \tag{2.5}$$

➢**2.11** 实用算法：

1. 照抄[(2)]系数矩阵 $\mathbf{A}$ 第 1 行；
2. 用式 (2.3) 写出第 1 列；

[(2)]即利用式 (2.2)。

3. 用式 (2.4) 写出第 2 行；
4. 用式 (2.5) 写出第 2 列；
5. 重复应用式 (2.4)、式 (2.5)，生成行与列。
6. 分成 **L**、**U** 两矩阵。

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ & & & \end{pmatrix} \Rightarrow \begin{pmatrix} \mu_{11} & \mu_{12} & \cdots & \mu_{1n} \\ l_{21} & & & \\ \vdots & & & \\ l_{n1} & & & \end{pmatrix} \Rightarrow \begin{pmatrix} \mu_{11} & \mu_{12} & \cdots & \mu_{1n} \\ l_{21} & \mu_{22} & \cdots & \mu_{2n} \\ \vdots & & & \\ l_{n1} & & & \end{pmatrix}$$

$$\overset{\cdots}{\Rightarrow} \begin{pmatrix} \mu_{11} & \mu_{12} & \cdots & \mu_{1n} \\ l_{21} & \mu_{22} & \cdots & \mu_{2n} \\ \vdots & \vdots & & \vdots \\ l_{n1} & l_{n2} & \cdots & \mu_{nn} \end{pmatrix} = \mathbf{L}+\mathbf{U} = \begin{pmatrix} & & & \\ l_{21} & & & \\ \vdots & \vdots & & \\ l_{n1} & l_{n2} & \cdots & \end{pmatrix} + \begin{pmatrix} \mu_{11} & \mu_{12} & \cdots & \mu_{1n} \\ & \mu_{22} & \cdots & \mu_{2n} \\ & & \ddots & \vdots \\ & & & \mu_{nn} \end{pmatrix}$$

➢**2.12** 需注意：重复第 5 步时，有如下的小妙招。

1. 我们注意到，按此方法生成的顺序如左下图所示，为“┌”状的生成顺序。那么我们以第 k 步时的生成为例。
2. 在第 k 步生成行时，对第 k 步的行的某个元素，首先向上找，应该能找到 $k-1$ 个元素的列向量；再向左找，能找到一个 $k-1$ 个元素的行向量，则该处应生成的值即为矩阵 **A** 对应位置的元素值减去找到的两个向量的内积。
3. 类似的，在第 k 步生成时，先向左找 $k-1$ 个元素的行向量，再向上找 $k-1$ 个元素的列向量，生成值为矩阵 **A** 对应位置元素值减两向量内积之后**除以** μ_{kk}。
4. 千万不要忘记生成列时要除以 μ_{kk}!!!

✎**2.13** 分解矩阵 **LU**，其中：$\mathbf{A} = \begin{pmatrix} 4 & -2 & 0 & 4 \\ -2 & 2 & -3 & 1 \\ 0 & -3 & 13 & -7 \\ 4 & 1 & -7 & 23 \end{pmatrix}$

（答案：$\mathbf{L} = \begin{pmatrix} 1 & & & \\ -\frac{1}{2} & 1 & & \\ 0 & -3 & 1 & \\ 1 & 3 & \frac{1}{2} & 1 \end{pmatrix}, \mathbf{U} = \begin{pmatrix} 4 & -2 & 0 & 4 \\ & 1 & -3 & 3 \\ & & 4 & 2 \\ & & & 9 \end{pmatrix}$）

➢**2.14** 利用 LU 分解可解线性方程组 $\mathbf{Ax}=\mathbf{b}$。

➢**2.15** 平方根法和改进平方根法。

➢**2.16** LDU 分解：令 $\mathbf{D}=\mathrm{diag}(\mu_{11},\mu_{22},\cdots,\mu_{nn})$，则有：

$$\mathbf{A}=\mathbf{LU}=\mathbf{L}\cdot\mathbf{I}\cdot\mathbf{U}=\mathbf{LDD}^{-1}\mathbf{U}=\mathbf{LDM}^{\mathrm{T}}$$

其中 $\mathbf{M}^{\mathrm{T}}=\mathbf{D}^{-1}\mathbf{U}$，$\mathbf{M}^{\mathrm{T}}$ 是一单位上三角阵。则：

$$\mathbf{M}=\begin{pmatrix}1&&&\\m_{21}&1&&\\\vdots&\vdots&\ddots&\\m_{n1}&m_{n2}&\cdots&1\end{pmatrix},\ m_{ji}=\frac{\mu_{ij}}{\mu_{ii}}\ \text{（即每行元素除以排头元素）}\qquad(2.6)$$

称 $\mathbf{A}=\mathbf{LDM}^{\mathrm{T}}$ 为矩阵的 LDU 分解。

➢**2.17** 对于对称阵，可分解为：$\mathbf{A}=\mathbf{LDL}^{\mathrm{T}}$（前提：各阶顺序主子式非 0）

➢**2.18** 对于对称正定阵，$\mathbf{D}$ 的元素均非负（且对角线非 0），$\mathbf{A}=\mathbf{GG}^{\mathrm{T}}$（Cholesky 分解）

➢**2.19** LDU 计算式不必死记，只需记住 LU 分解然后变换即可。

➢**2.20** 平方根法：$\mathbf{A}=\mathbf{GG}^{\mathrm{T}}\ \Rightarrow\ \mathbf{Ax}=\mathbf{b}\ \Rightarrow\ \mathbf{GG}^{\mathrm{T}}\mathbf{x}=\mathbf{b}\ \Rightarrow\ \begin{cases}\mathbf{Gy}=\mathbf{b}\ \Rightarrow\ \text{解出}\mathbf{y}\\ \mathbf{G}^{\mathrm{T}}\mathbf{x}=\mathbf{y}\ \Rightarrow\ \text{解出}\mathbf{x}\end{cases}$

➢**2.21** 改进平方根法：$\mathbf{A}=\mathbf{LDL}^{\mathrm{T}}\ \Rightarrow\ \mathbf{Ax}=\mathbf{b}\ \Rightarrow\ \mathbf{LDL}^{\mathrm{T}}\mathbf{x}=\mathbf{b}\ \Rightarrow\ \begin{cases}\mathbf{Ly}=\mathbf{b}\ \Rightarrow\ \text{解出}\mathbf{y}\\ \mathbf{Dz}=\mathbf{y}\ \Rightarrow\ \text{解出}\mathbf{z}\\ \mathbf{L}^{\mathrm{T}}\mathbf{x}=\mathbf{b}\ \Rightarrow\ \text{解出}\mathbf{x}\end{cases}$

➢**2.22** 稀疏矩阵：大量元素为 0，非零元很少。$p=q=1$ 时的带状矩阵称为三对角阵，通常为严格对角占优矩阵。

图 2.2: 稀疏矩阵做 LU 分解示意图

➢**2.23** 解三对角系数矩阵线性方程组的追赶法：设系数矩阵 $\mathbf{T}$ 为三对角阵，则有

$$\mathbf{T}=\mathbf{LU},\mathbf{Tx}=\mathbf{d}\ \Rightarrow\ \mathbf{LUx}=\mathbf{d}\ \Rightarrow\ \begin{cases}\mathbf{Ly}=\mathbf{d}\ \ \text{（追）}\\ \mathbf{Ux}=\mathbf{y}\ \ \text{（赶）}\end{cases}$$

§2.3 病态问题理论

➢**2.24** 残向量：$\mathbf{r}=\mathbf{b}-\mathbf{A}\tilde{\mathbf{x}}$

➢**2.25** 误差向量：$\mathbf{e}=\mathbf{x}^{*}-\tilde{\mathbf{x}}$

➢**2.26** 如何衡量误差（向量）的大小？可采用向量的范数衡量。

➢**2.27** 向量范数：称 $\|\mathbf{x}\|$ 为一个向量的范数，若 $\|\mathbf{x}\|\in\mathbb{R}$ 满足：

1. 非负性：$\forall\mathbf{x}\in\mathbb{R}^{n},\|\mathbf{x}\|\geq 0$且$\|\mathbf{x}\|=0\neq\mathbf{x}=\mathbf{0}$
2. 齐次性：$\forall\alpha\in\mathbb{R},x\in\mathbb{R}^{n},\|\alpha\mathbf{x}\|=|\alpha|\cdot\|\mathbf{x}\|$
3. 三角不等式：$\forall\mathbf{x},\mathbf{y}\in\mathbb{R}^{n},\|\mathbf{x}+\mathbf{y}\|\leq\|\mathbf{x}\|+\|\mathbf{y}\|$

➢**2.28** 常用范数：

- 1-范数：$\|\mathbf{x}\|_{1}=|x_{1}|+|x_{2}|+\cdots+|x_{n}|$.
- 2-范数：$\|\mathbf{x}\|_{2}=\sqrt{x_{1}^{2}+x_{2}^{2}+\cdots+x_{n}^{2}}$.
- ∞-范数：$\|\mathbf{x}\|_{\infty}=\max\limits_{1\leq i\leq n}|x_{i}|$.

➢**2.29** 矩阵范数：称 $\|\mathbf{A}\|\in\mathbb{R}$ 为一个矩阵范数，若其满足：

1. 非负性：$\forall\mathbf{A},\|\mathbf{A}\|\geq 0$ 且 $\|\mathbf{A}\|=0\Leftrightarrow\mathbf{A}=\mathbf{O}$
2. 齐次性：$\forall\alpha\in\mathbb{R},\|\alpha\mathbf{A}\|=|\alpha|\cdot\|\mathbf{A}\|$
3. 三角不等式：$\|\mathbf{A}+\mathbf{B}\|\leq\|\mathbf{A}\|+\|\mathbf{B}\|$
4. $\|\mathbf{AB}\|\leq\|\mathbf{A}\|\cdot\|\mathbf{B}\|$

➢**2.30** 定义：若 $\|\mathbf{Ax}\|\leq\|\mathbf{A}\|\cdot\|\mathbf{x}\|$，则称矩阵范数 $\|\mathbf{A}\|$ 与向量范数 $\|\mathbf{x}\|$ 为相容或协调的。

➢**2.31** 算子范数：$\|\mathbf{A}\|_{p}=\max\dfrac{\|\mathbf{Ax}\|_{p}}{\|\mathbf{x}\|_{p}}=\max\limits_{\|\mathbf{x}\|_{p}=1}\|\mathbf{Ax}\|_{p}$，容易证明 $\|\mathbf{A}\|_{p}$ 满足相容条件。

1. 1-范数：$\|\mathbf{A}\|_{1}=\max\limits_{1\leq j\leq n}\left\{\sum\limits_{i=1}^{n}|a_{ij}|\right\}$即列和最大值
2. 2-范数：$\|\mathbf{A}\|_{2}=\sqrt{\mathbf{A}^{\mathrm{T}}\mathbf{A}\text{的最大特征值}}$
3. ∞-范数：$\|\mathbf{A}\|_{\infty}=\max\limits_{1\leq i\leq n}\left\{\sum\limits_{j=1}^{n}|a_{ij}|\right\}$即行和最大值

➢**2.32** 矩阵的谱半径：$\rho(\mathbf{A})=\max\limits_{1\leq i\leq n}|\lambda_{i}|$，性质：$\rho(\mathbf{A})\leq\|\mathbf{A}\|$

➢**2.33** 定理：设 $\|\mathbf{B}\|\leq 1$，则 $\mathbf{I}-\mathbf{B}$ 可逆，且

$$\|(\mathbf{I}-\mathbf{B})^{-1}\|\leq\frac{1}{1-\|\mathbf{B}\|}\tag{2.7}$$

➢**2.34** 舍入误差对解的影响：

$$\mathbf{r}=\mathbf{b}-\mathbf{A}\tilde{\mathbf{x}}\ \Rightarrow\ \mathbf{A}\tilde{\mathbf{x}}=\mathbf{b}-\mathbf{r}\neq\mathbf{b}$$

为分析舍入误差的相对水平，首先定义 $\mathbf{e}=\mathbf{x}^*-\tilde{\mathbf{x}}$，分析其量级有

$$\tilde{\mathbf{x}}=\mathbf{A}^{-1}\mathbf{r}\ \Rightarrow\ \|\mathbf{e}\|=\|\mathbf{A}^{-1}\mathbf{r}\|\leq\|\mathbf{A}^{-1}\|\cdot\|\mathbf{r}\|$$

由于 $\mathbf{A}\mathbf{x}^*=\mathbf{b}$，故

$$\|\mathbf{b}\|=\|\mathbf{A}\mathbf{x}^*\|\leq\|\mathbf{A}\|\cdot\|\mathbf{x}^*\|\ \Rightarrow\ \frac{1}{\|\mathbf{x}^*\|}\leq\frac{\|\mathbf{A}\|}{\|\mathbf{b}\|}$$

故相对误差水平 $\dfrac{\|\mathbf{x}^*-\tilde{\mathbf{x}}\|}{\|\mathbf{x}^*\|}\leq\|\mathbf{A}\|\cdot\|\mathbf{A}^{-1}\|\cdot\dfrac{\|\mathbf{r}\|}{\|\mathbf{b}\|}$，其中的系数即可定义为矩阵的条件数 $\mathrm{Cond}(\mathbf{A})=\|\mathbf{A}\|\|\mathbf{A}^{-1}\|$

➢**2.35** 易知 $\mathrm{Cond}(\mathbf{A})>1$（$\|\mathbf{A}\|\|\mathbf{A}^{-1}\|\geq\|\mathbf{A}\cdot\mathbf{A}^{-1}\|=1$）

➢**2.36** 残向量 $\mathbf{r}$ 不能完全反映偏差水平，因 $\mathbf{r}$ 小，$\mathbf{e}$ 也不一定小。

➢**2.37** 系数矩阵扰动对解的影响：

$$(\mathbf{A}+\Delta\mathbf{A})\mathbf{x}=\mathbf{b}\ \Rightarrow\ \mathbf{r}=\mathbf{b}-\mathbf{A}\tilde{\mathbf{x}}=\Delta\mathbf{A}\tilde{\mathbf{x}}\ \Rightarrow\ \|\mathbf{r}\|\leq\|\Delta\mathbf{A}\|\|\tilde{\mathbf{x}}\|$$

可见若 $\Delta\mathbf{A}$ 小，$\mathbf{r}$ 也小。故相对误差水平

$$\frac{\|\mathbf{x}^*-\tilde{\mathbf{x}}\|}{\mathbf{x}^*}\leq\mathrm{Cond}(\mathbf{A})\frac{\|\mathbf{r}\|}{\|\mathbf{b}\|}\leq\mathrm{Cond}(\mathbf{A})\cdot\frac{\|\tilde{\mathbf{x}}\|}{\|\mathbf{x}^*\|}\cdot\frac{\|\Delta\mathbf{A}\|}{\|\mathbf{A}\|}$$

➢**2.38** 共同影响：$(\mathbf{A}-\Delta\mathbf{A})(\mathbf{x}-\Delta\mathbf{x})=\mathbf{b}-\Delta\mathbf{b}$，当 $\|\Delta\mathbf{A}\|\cdot\|\mathbf{A}^{-1}\|<1$ 时：

$$\frac{\|\mathbf{x}^*-\tilde{\mathbf{x}}\|}{\mathbf{x}^*}\leq\frac{\|\mathbf{A}^{-1}\|\cdot\|\mathbf{A}\|}{1-\|\mathbf{A}^{-1}\|\cdot\|\mathbf{A}\|\cdot\frac{\|\Delta\mathbf{A}\|}{\|\mathbf{A}\|}}\left(\frac{\|\Delta\mathbf{b}\|}{\|\mathbf{b}\|}+\frac{\|\Delta\mathbf{A}\|}{\|\mathbf{A}\|}\right)\tag{2.8}$$

$\mathrm{Cond}(\mathbf{A})$ 较大时，方程组为病态方程组；反之，$\mathrm{Cond}(\mathbf{A})$ 较小时，方程组仍为良态方程组。

➢**2.39** $\mathrm{Cond}(\mathbf{A})$ 的估计：

$$\mathbf{A}\mathbf{x}=\mathbf{b}\ \Rightarrow\ \mathbf{x}=\mathbf{A}^{-1}\mathbf{b}\ \Rightarrow\ \|\mathbf{x}\|\leq\|\mathbf{A}^{-1}\|\cdot\|\mathbf{b}\|\ \Rightarrow\ \frac{\|\mathbf{x}\|}{\|\mathbf{b}\|}\leq\|\mathbf{A}^{-1}\|$$

随机选取 p 个向量 $\mathbf{b}_1,\mathbf{b}_2,\cdots,\mathbf{b}_p$，解方程 $\mathbf{A}\mathbf{x}^{(k)}=\mathbf{b}_k$，得到 $\mathbf{x}^{(k)}$，由上面结论得到

$$\max_{1\leq k\leq p}\frac{\|\mathbf{x}^{(k)}\|}{\|\mathbf{b}_k\|}\leq\|\mathbf{A}^{-1}\|$$

故可近似认为：$\mathrm{Cond}(\mathbf{A})\approx\|\mathbf{A}\|\cdot\max\limits_{1\leq k\leq p}\dfrac{\|\mathbf{x}^{(k)}\|}{\|\mathbf{b}_k\|}$.

第三章　线性方程组迭代解法

§3.1　迭代方法概要

➢**3.1**　思想：$f(x^*)=0 \Rightarrow \cdots \Rightarrow x^*=\phi(x^*)$，给出初值 x_0 和递推公式 $x_{k+1}=\phi(x_k)$，假设 $\{x_k\}$ 收敛，求极限——设 $\lim\limits_{k\to\infty} x_k=x$，则有 $x=\phi(x)$，从而必有 $f(x)=0$。

➢**3.2**　向量序列收敛：$\mathbf{x}^{(k)}=(x_1^{(k)},x_2^{(k)},\cdots,x_n^{(k)})^{\mathrm{T}}$，若 $\lim\limits_{k\to\infty} x_i^{(k)}=x_i^*$，则 $\mathbf{x}^{(k)}\to\mathbf{x}^*\ (k\to\infty)$，记作 $\lim\limits_{k\to\infty}\mathbf{x}^{(k)}=\mathbf{x}^*$。

➢**3.3**　矩阵序列收敛：$\mathbf{A}^{(k)}=(a_{ij}^{(k)})_{m\times n}$，若 $\lim\limits_{k\to\infty} a_{ij}^{(k)}=a_{ij}$，则称 $\mathbf{A}^{(k)}\to\mathbf{A}=(a_{ij})_{m\times n}$ [(1)]，记作 $\lim\limits_{k\to\infty}\mathbf{A}^{(k)}=\mathbf{A}$。

➢**3.4**　序列收敛定理：

- 对向量，$\mathbf{x}^{(k)}\to\mathbf{x}^* \Leftrightarrow \lim\limits_{k\to\infty}\|\mathbf{x}^*-\mathbf{x}^{(k)}\|=0$
- 对矩阵，$\mathbf{A}^{(k)}\to\mathbf{A} \Leftrightarrow \lim\limits_{k\to\infty}\|\mathbf{A}-\mathbf{A}^{(k)}\|=0$

➢**3.5**　定理：设 $\mathbf{B}\in\mathbb{R}^{m\times n}$，则 $\lim\limits_{k\to\infty}\mathbf{B}^k=0 \Leftrightarrow \rho(\mathbf{B})<1$。

➢**3.6**　设 $\mathbf{Ax}=\mathbf{b}$，变形得 $\mathbf{x}=\mathbf{Bx}+\mathbf{g}$，可构造迭代格式

$$\mathbf{x}^{(k+1)}=\mathbf{B}\mathbf{x}^{(k)}+\mathbf{g} \tag{3.1}$$

则 $\mathbf{x}^*=\lim\limits_{k\to\infty}\mathbf{x}^{(k)}$ 满足 $\mathbf{x}^*=\mathbf{B}\mathbf{x}^*+\mathbf{g} \Rightarrow \mathbf{A}\mathbf{x}^*=\mathbf{b}$，故可以通过 (3.1) 式不断迭代以逼近方程的解。

§3.2　三种基本迭代法

➢**3.7**　方法一·Jacobi 迭代法：以第 i 行为例，有：

$$a_{i1}x_1+a_{i2}x_2+\cdots+a_{ii}x_i+\cdots+a_{in}x_n=b_i$$

[(1)] 不用 $\mathbf{A}^*$，是怕与伴随矩阵的符号弄混。——作者注

可解出

$$x_i = \frac{1}{a_{ii}}[b_i - (a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{i,i-1}x_{i-1} + a_{i,i+1}x_{i+1} + \cdots + a_{in}x_n)]$$
$$= \frac{1}{a_{ii}}\left[b_i - \sum_{j=1,j\neq i}^{n} a_{ij}x_j\right] \tag{3.2}$$

依上式即可构造 Jacobi 迭代格式：

$$x_i^{(k+1)} = \frac{1}{a_{ii}}\left[b_i - \sum_{j=1,j\neq i}^{n} a_{ij}x_j^{(k)}\right] \quad (i = 1, 2, \cdots, n) \tag{3.3}$$

按照 $\mathbf{x} = \mathbf{Bx} + \mathbf{g}$ 的标准格式，整理成矩阵格式：

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} 0 & -\frac{a_{12}}{a_{11}} & -\frac{a_{13}}{a_{11}} & \cdots & -\frac{a_{1n}}{a_{11}} \\ -\frac{a_{21}}{a_{22}} & 0 & -\frac{a_{23}}{a_{22}} & \cdots & -\frac{a_{2n}}{a_{22}} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -\frac{a_{n1}}{a_{nn}} & -\frac{a_{n2}}{a_{nn}} & -\frac{a_{n3}}{a_{nn}} & \cdots & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} + \begin{pmatrix} \frac{b_1}{a_{11}} \\ \frac{b_2}{a_{22}} \\ \vdots \\ \frac{b_n}{a_{nn}} \end{pmatrix} \tag{3.4}$$

➢**3.8** 方法二·Gauss-Seidel 迭代法：

$$\begin{cases} x_1^{(k+1)} = \frac{1}{a_{11}}(b_1 - a_{12}x_2^{(k)} - a_{13}x_3^{(k)} - \cdots - a_{1n}x_n^{(k)}) \\ x_2^{(k+1)} = \frac{1}{a_{22}}(b_2 - a_{21}\boxed{x_1^{(k+1)}} - a_{23}x_3^{(k)} - \cdots - a_{2n}x_n^{(k)}) \\ x_3^{(k+1)} = \frac{1}{a_{33}}(b_2 - a_{31}\boxed{x_1^{(k+1)}} - a_{32}\boxed{x_2^{(k+1)}} - \cdots - a_{3n}x_n^{(k)}) \\ \quad \cdots\cdots \\ x_n^{(k+1)} = \frac{1}{a_{nn}}(b_2 - a_{n1}\boxed{x_1^{(k+1)}} - a_{n2}\boxed{x_2^{(k+1)}} - \cdots - a_{n,n-1}\boxed{x_{n-1}^{(k+1)}}) \end{cases} \tag{3.5}$$

$\mathbf{x} = \mathbf{Bx} + \mathbf{g}$ 形式与 Jacobi 法相同，但区别在于进行下一步变量的迭代时采用了“新解”即上式中用方框框出的部分。

➢**3.9** 方法三·超松弛（SOR）迭代法：对 Gauss-Seidel 法的通用格式分析改写：若记每次迭代时的误差为 $\mathbf{r_i^{(k+1)}}$，即

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

当 $k \to \infty$ 时总有 $r_i^{(k+1)}/a_{ii} \to 0$，故可以强行乘一个系数 ω 以加快收敛：

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

整理得

$$x_i^{(k+1)} = (1-\omega)x_i^{(k)} + \frac{\omega}{a_{ii}}\left[b_i - \sum_{j=1}^{i-1} a_{ij}x_j^{(k+1)} - \sum_{j=i+1}^{n} a_{ij}x_j^{(k)}\right] \tag{3.6}$$

整体迭代格式：

$$\begin{cases} x_1^{(k+1)} = (1-\omega)x_1^{(k)} + \dfrac{\omega}{a_{11}}\left(b_1 - a_{12}x_2^{(k)} - a_{13}x_3^{(k)} - \cdots - a_{1n}x_{11}^{(k)}\right) \\ x_2^{(k+1)} = (1-\omega)x_2^{(k)} + \dfrac{\omega}{a_{11}}\left(b_2 - a_{21}x_1^{(k+1)} - a_{23}x_3^{(k)} - \cdots - a_{2n}x_{11}^{(k)}\right) \\ \quad \cdots\cdots \\ x_n^{(k+1)} = (1-\omega)x_n^{(k)} + \dfrac{\omega}{a_{n1}}\left(b_n - a_{n1}x_1^{(k+1)} - a_{n2}x_2^{(k+1)} - \cdots - a_{n,n-1}x_{11}^{(k+1)}\right) \end{cases} \tag{3.7}$$

➢**3.10** 迭代的矩阵表示法：将 $\mathbf{A}$ 分解为 $\mathbf{D},\mathbf{E},\mathbf{F}$ 三部分：$\mathbf{A} = \mathbf{D} - \mathbf{E} - \mathbf{F}$，其中

$$\mathbf{D} = \begin{pmatrix} a_{11} & & & \\ & a_{22} & & \\ & & \ddots & \\ & & & a_{nn} \end{pmatrix}, \quad \mathbf{E} = \begin{pmatrix} 0 & & & & \\ -a_{21} & 0 & & & \\ -a_{31} & -a_{32} & 0 & & \\ \vdots & \vdots & \vdots & \ddots & \\ -a_{n1} & -a_{n2} & -a_{n3} & \cdots & 0 \end{pmatrix}, \quad \mathbf{F} = \begin{pmatrix} 0 & -a_{12} & -a_{13} & \cdots & -a_{1n} \\ & 0 & -a_{23} & \cdots & -a_{2n} \\ & & 0 & \cdots & -a_{3n} \\ & & & \ddots & \vdots \\ & & & & 0 \end{pmatrix} \tag{3.8}$$

- Jacobi 法：用矩阵形式推导 Jacobi 迭代公式，有

  $$(\mathbf{D}-\mathbf{E}-\mathbf{F})\mathbf{x} = \mathbf{b} \;\Rightarrow\; \mathbf{D}\mathbf{x} = (\mathbf{E}+\mathbf{F})\mathbf{x} + \mathbf{b} \;\Rightarrow\; \mathbf{x} = \mathbf{D}^{-1}(\mathbf{E}+\mathbf{F})\mathbf{x} + \mathbf{D}^{-1}\mathbf{b}$$

  可见有

  $$\begin{cases} \mathbf{B} = \mathbf{D}^{-1}(\mathbf{E}+\mathbf{F}) = \mathbf{D}^{-1}(\mathbf{D}-\mathbf{A}) = \mathbf{I} - \mathbf{D}^{-1}\mathbf{A} \\ \mathbf{g} = \mathbf{D}^{-1}\mathbf{b} \end{cases} \tag{3.9}$$

- Gauss-Seidel 法：

  $$(\mathbf{D}-\mathbf{E}-\mathbf{F})\mathbf{x} = \mathbf{b} \;\Rightarrow\; (\mathbf{D}-\mathbf{E})\mathbf{x} = \mathbf{F}\mathbf{x} + \mathbf{b} \;\Rightarrow\; \mathbf{x} = (\mathbf{D}-\mathbf{E})^{-1}\mathbf{F}\mathbf{x} + (\mathbf{D}-\mathbf{E})^{-1}\mathbf{b}$$

  可见有

  $$\begin{cases} \mathbf{B} = (\mathbf{D}-\mathbf{E})^{-1}\mathbf{F} \\ \mathbf{g} = (\mathbf{D}-\mathbf{E})^{-1}\mathbf{b} \end{cases} \tag{3.10}$$

- SOR 法：

$$\omega(\mathbf{D}-\mathbf{E}-\mathbf{F})\mathbf{x}=\omega\mathbf{b}\ \Rightarrow\ (\mathbf{D}-\omega\mathbf{E})\mathbf{x}=[(1-\omega)\mathbf{D}+\omega\mathbf{F}]\mathbf{x}+\omega\mathbf{b}$$
$$\Rightarrow\ \mathbf{x}=(\mathbf{D}-\omega\mathbf{E})^{-1}[(1-\omega)\mathbf{D}+\omega\mathbf{F}]\mathbf{x}+(\mathbf{D}-\omega\mathbf{E})^{-1}\omega\mathbf{b}$$

可见有

$$\begin{cases}\mathbf{B}=(\mathbf{D}-\omega\mathbf{R})^{-1}[(1-\omega)\mathbf{D}+\omega\mathbf{F}]\\ \mathbf{g}=\omega(\mathbf{D}-\omega\mathbf{E})^{-1}\mathbf{b}\end{cases}\tag{3.11}$$

§3.3 迭代收敛理论

➢**3.11** 下面给出迭代格式收敛的条件（非常重要！个人猜测必考）。

➢**3.12** 定理 1：若 $\|\mathbf{B}\|\le 1$，则 $\forall\mathbf{x}^{(0)}$，迭代格式 $\mathbf{x}^{(k+1)}=\mathbf{B}\mathbf{x}^{(k)}+\mathbf{g}$ 收敛于解 $\mathbf{x}^*$，且有误差估计式：

$$\|\mathbf{x}^*-\mathbf{x}^{(k)}\|\le\frac{\|\mathbf{B}\|}{1-\|\mathbf{B}\|}\|\mathbf{x}^{(k)}-\mathbf{x}^{(k-1)}\|\quad\text{（事后估计）}\tag{3.12}$$

$$\|\mathbf{x}-\mathbf{x}^{(k)}\|\le\frac{\|\mathbf{B}\|^k}{1-\|\mathbf{B}\|}\|\mathbf{x}^{(1)}-\mathbf{x}^{(0)}\|\quad\text{（事前估计）}\tag{3.13}$$

➢**3.13** 定理 2：$\forall\mathbf{x}^{(0)}$，迭代格式 $\mathbf{x}^{(k+1)}=\mathbf{B}\mathbf{x}^{(k)}+\mathbf{g}$ 收敛于解 $\mathbf{x}^*$ 的充要条件是下列两条件至少有一个成立：

1. $\mathbf{B}^k\to\mathbf{O}$；
2. $\mathbf{B}$ 的谱半径 $\rho(\mathbf{B})<1$。

➢**3.14** 推论：SOR 法收敛的必要条件是 $0<\omega<2$。

$$|\mathbf{B}|=|(D-\omega\mathbf{E})^{-1}|\cdot|(1-\omega)\mathbf{D}+\omega\mathbf{F}|=\frac{|(1-\omega)\mathbf{D}|}{|\mathbf{D}|}=(1-\omega)^n$$
$$\rho(\mathbf{B})<1\ \Rightarrow\ |\lambda_1\cdots\lambda_n|<1\ \Rightarrow\ |1-\omega|<1\ \Rightarrow\ 0<\omega<2.$$

➢**3.15** 对于三种常用迭代法，还有更为实用的结论可供应用：

- 引理：若 $\mathbf{A}$ 是严格对角占优矩阵，$0\le\omega\le 1$ 且 $\lambda\ge 1$ 时，矩阵 $(\lambda+\omega-1)\mathbf{D}-\lambda\omega\mathbf{E}-\omega\mathbf{F}$ 也是严格对角占优矩阵。
- 推论 2：若 $\mathbf{A}$ 是严格对角占优矩阵，则 $\forall\mathbf{x}^{(0)}$，Jacobi 法、G-S 法、SOR 法（$0<\omega\le 1$）均收敛[(2)]。

[(2)]可由其上的引理推出：通过反设 $|\lambda|\ge 1$ 退出 $|\lambda\mathbf{I}-\mathbf{B}|\ne 0$，故 $|\lambda|<1$，$\rho(\mathbf{B})<1$。

- 推论 3：若 $\mathbf{A}$ 为对称正定阵，则 $\forall \mathbf{x}^{(0)}$，Jacobi 法收敛的充要条件是：$2\mathbf{D}-\mathbf{A}$ 也是对称正定阵。
- 推论 4：若 $\mathbf{A}$ 是对称正定阵，则 $\forall \mathbf{x}^{(0)}$，SOR 法收敛充要条件为 $0<\omega<2$。

➢**3.16** 对不同情况下的审敛法做总结：

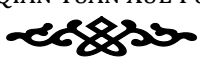
1. 对角占优矩阵 $\mathbf{A}$：Jacobi 法收敛，G-S 法收敛，$0<\omega\leq 1$ 时 SOR 法收敛。
2. 对称正定阵 $\mathbf{A}$：
   - Jacobi 法收敛 $\Leftrightarrow$ $2\mathbf{D}-\mathbf{A}$ 收敛；
   - SOR 法收敛 $\Leftrightarrow$ $0<\omega<2$。
3. 一般矩阵 $\mathbf{A}$：
   - $\|\mathbf{B}\|<1$ 时三方法均收敛；
   - $\mathbf{B}^k\to\mathbf{O}$ 和 $\rho(B)<1$ 中之一成立，则三方法均收敛。
   - SOR 法收敛的必要条件是 $0<\omega<2$。

✍**3.17** 判断系数矩阵为 $\mathbf{A}=\begin{pmatrix}2&3&4\\3&6&10\\4&10&20\end{pmatrix}$ 时各迭代法的收敛性。

第四章　插值法

§4.1　插值法思想概要

后记



索引