

## 前言

由于计算机的普及，科学计算已成为各学科领域的一项重要工作。学习和掌握数值计算方法的基本原理及应用已成为现代科学工作者不可缺少的一个环节。

用计算机解决科学计算问题需经历几个过程：由实际问题建立数学模型，根据数学模型提出求解的数值计算方法，编出程序、上机求出结果。通过以上过程，可以看出，数值计算方法 是计算机、数学和应用科学之间的桥梁，是程序设计和对数值结果进行分析的依据，是用计算机进行科学计算全过程的一个重要环节。目前，数值计算方法已经成为理工院校（非数学专业）硕士学位研究生的学位课。在农业科学研究中，数值计算方法已经成为不可缺少的有力工具。

学生通过本课程的学习，能掌握科学计算中常用的算法，能独立地用学过的算法编程，测试。并能解决工作中遇到的实际问题。

本教材同时也适当增加一些只供阅读，而不在课堂教授的内容，这样在规定的课时内可完成基本内容的讲授，又可以作为今后科研的参考书。

各章节要点及授课时数

**本教材的特点：**对现有的已知的数学模型建立起相应的一系列数值求解方法，力求设计计算量小、存储量少、精度高，特别是计算机所能接受且执行计算方法，兼顾分析方法的收敛性、稳定性及误差估计。在不失科学性的前提下，尽量做到深入浅出地介绍计算机上常用的数值方法，使学生能用数值计算方法对建立的数学模型实际求解。

本教材既坚持介绍数值计算方法基本原理，又兼顾应用学科的特点。

**使用范围：**理工科非数学学科公共专业课教材。

本教材主要介绍计算机上常用的各种数值计算方法以及相关的基本概念及理论。内容包括误差分析初步，方程求根，线性方程组的直接解法与迭代法，插值法，最小二乘曲线拟合，数值积分计算，常微分方程数值解法和偏微分方程数值解法。本课程中对主要基本算法的推导、构造原理、收敛性、误差估计进行了讨论。

本教材的另一个特色是侧重于计算机应用，各章均有例题及数值算例，并指出应掌握的基本问题，对每一个算法都给出伪代码，以便于学生编制程序的需要，且有适当的书面练习及一些上机计算题。

本教材可作为理工科（非数学）各专业研究生及大学本科高年级的数值计算方法教材，也可供工程技术人员参考。

作者在中国农业大学从事硕士研究生学位课《数值计算方法》教学已经 30 余年，在多年的教学中发现给农业院校的学生寻找一本合适的教材很难。一方面，和其它理工科院校的学生相比，农业院校的学生数学基础课学得不够，因此需要浅显易懂的教材；另一方面，农业科学的发展，对数学和计算科学提出了很高的要求，因为农业科学的复杂性和不确定性，需要有尽可能多的内容和深度，这样互相矛盾的需求给讲课教师提出了很高的要求，教师要深入的理解数值计算的理念，要有深厚的计算数学理论知识和计算经验才能讲好这门课。

在学习了许多国内外教材的基础上，根据农业院校学生的特殊要求，作者于 1984 年编写了《数值计算方法》讲稿，每年都有小的改动，在 1993 年、1999 年、2005 年和 2009 年对讲稿进行了几次较大的修改，以适应科学发展需要和农业院校学生的基础。许多研究生通过学习，在毕业论文中引用了数值计算方法解决应用问题，提高了论文水平，也有许多在职教师和科研人员学习这门课程后，将数值方法引用到科研课题中，取得了较好的成果。

超星公司在 2009 年为作者做了《数值计算方法》的教学视频，几年来在网上听众甚多，作者时常收到理工科选修数值分析学生的邮件，表达了大家对作者讲课的认可，本讲义也可以作为理工科非数学专业学生学习《数值计算方法》的参考书。

# 目录

## 第一章 预篇

- 1.1 数值计算方法的研究对象和特点
- 1.2 误差分析
- 1.3 算法概述

## 第二章 非线性方程求根

- 2.1 二分法
- 2.2 迭代法的一般原则
- 2.3 牛顿法(切线法)
- 2.4 弦截法

## 第三章 解线性方程组的直接法

- 3.1 Gauss 消去法
- 3.2 矩阵的三角分解及其在解方程组中的应用
- 3.3 解对称正定矩阵方程组的平方根解法
- 3.4 解三对角方程组的追赶法
- 3.5 向量和矩阵的范数
- 3.6 方程组的性态、条件数

## 第四章 解线性方程组的迭代法

- 4.1 Jacobi 迭代法
- 4.2 Gauss-Seidel
- 4.3 SQR 方法
- 4.4 迭代法的收敛性
- 4.5 共轭梯度法
- 4.6 最小二乘法

## 第五章 矩阵特征值问题的计算方法

- 5.1 矩阵特征值问题
- 5.2 乘幂法和反幂法
- 5.3 Household 方法

#### 5.4 QR 方法

### 第六章 函数插值

#### 6.1 Lagrange 插值

#### 6.2 Newton 插值

#### 6.3 等距节点的插值

#### 6.4 Hermite 插值

#### 6.5 分段低次多项式插值

#### 6.6 三次样条插值

### 第七章 最佳平方逼近

#### 7.1 正交多项式

#### 7.2 切比雪夫多项式

#### 7.3 曲线拟合的最小二乘法

### 第八章 数值积分

#### 8.1 Newton-Cotes 求积公式

#### 8.2 Romberg 求积公式

#### 8.3 Gauss 型求积公式

### 第九章 常微分方程数值解

#### 9.1 Euler 法和改进 Euler 法

#### 9.2 Runge-Kutta 法

#### 9.3 线性多步法

#### 9.4 解二阶常微分方程边值问题的差分法

### 第十章 偏微分方程数值解

#### 10.1 椭圆型方程差分法

#### 10.2 抛物型方程差分法

#### 10.3 双曲型方程差分法

#### 10.4 有限元方法初步

# 第一章 预篇

数学是研究数与形的科学。其中研究怎样利用手指、算盘、计算尺、计算器、计算机等工具，来求出数学问题数值解的学问，就是数值计算方法。它是数学中最古老的部分，但只是在计算机出现以后，人们获得了高速度、自动化的计算工具，才为众多浩繁的数值计算问题的解决展现了光明的前景。从此，科学研究与工程设计的手段，发生了由模型试验向数值计算的巨大转变。

近年来，由于计算机的发展，计算性的学科新分支，如计算力学、计算物理、计算化学、计算生物学、计算地质学、计算经济学以及众多工程科学的计算分支纷纷兴起。因为任何具体学科中的计算过程，不论其目的、背景和含义如何，终归是数学的计算过程，数值计算方法是各种计算性学科的联系纽带和共性基础。

学习数值计算方法可以知道如何用计算机解决数学问题，特别是我们在微积分和线性代数中没有学过的一些解决问题的方法，还介绍一些用不同的方法解决以前用传统的数学方法解决的问题。我们选择一些现实世界存在的例子，用解析的方法能够解出来，以便将数值方法和解析方法做一个对比。

## 1.1 数值计算方法的起源和意义

数值计算方法是数学的一个古老的分支，虽然数学不仅仅是计算，但推动数学产生和发展的最直接原因还是计算问题。人类社会发展的初期，就常常遇到各种各样的计算问题。一开始没有任何计算工具，最得心应手的就是人类的一双手了。于是人们采取了扳手指头和结绳计数的方法进行计算。随着社会的进一步向前发展，问题越来越复杂，原始的工具不敷使用。人们越来越迫切地希望有更先进的技术和理论来进行计算，于是数学应运而生。可以说，记数术是最原始的数学，数学的源头就是计算，计算自古以来就是数学的一个重要组成部分。

中国古代数学曾经有过辉煌的成就，它不像古希腊数学那样注重逻辑和推理，而是具有显著的计算性和实用性的特点，从《九章算术》等中国古典数学名著中就可以看出这一点。早在商代中国就形成了十进制这一方便的计数和运算机制。从最早发明的算筹到后来的算盘以及相应发展起来的珠算方法，是古代中国对世界计算技术的最重要贡献，至今还在中国和其他一些国家都发挥着作用。

到了二十世纪四十年代，生产高性能计算工具的技术条件已经成熟。当时适逢第二次世界大战，军事上对高速计算机有迫切的需求。这就迎来了世界上第一台电子计算机的诞生。

计算机的问世，从根本上改变了计算工具落后的局面。古老的计算数学借助计算机这一强有力的工具，一下子焕发出了青春。随着计算技术的发展，社会需求的急剧增加，计算数学的应用领域越来越广泛，这就使得越来越多的、以前不能设想的、难度和规模日益增大的计算问题得以解决。在这样新的条件下，计算在整个科学技术以至经济生活中的重要性得到前所未有的提高；同时，以原来分散在数学各分支的计算方法为基础的一门新的数学科学一开始形成并迅速发展。计算数学和计算机一起已经成为众多领域研究工作中不可或缺的工具和手段。

当代计算能力的大幅度提高既来自计算机的进步，也来自计算方法的进步。计算机和计算方法的发展是相辅相成、相互制约和相互促进的。计算方法的发展启发了新的计算机体系结构，而计算机的更新换代也对计算方法提出了新的标准和要求。自计算机诞生以来，经典的计算方法业已经历了一个重新评价、筛选、改造和创新的过程；与此同时，涌现了许多新概念、新课题和许多能够充分发挥计算机潜力、有更大解题能力的新方法；这就构成了现代意义下的计算数学（数值计算方法）。

## 1.2 数值计算方法的研究对象

科学计算问题的出发点，往往是研究现实世界中的问题或现象。例如：物理、自然或者社会问题。在一定的假设下，可以将这些实际问题列成方程，也就是数学模型来描述或进行分析。从简单的代数方程到复杂的非线性偏微分方程应有尽有。这些方程一旦建立起来，下一步就是要解方程，以预测这些现象将来朝什么方向发展。为了获取数据，我们也需要做一些实验。如果模型预测的结果和实测数据一致或接近，就认为所建立的模型是合适的，否则就要对所建立的模型加以调整和改进。

在求解阶段，最理想的状况是求出方程的解析解，遗憾的是，仅对最简单的模型才能求出解析解。我们通过几个例子来说明在微积分和线性代数中学过的数学模型而在这两门课程中所没有学到可行的解法。

**例 1.1** 求解线性方程组  $AX=b$ ，其中系数矩阵  $A$  是  $n \times n$  的方阵，设  $n=20$ ；

**解：** 本例的行列式  $D \equiv \det(A) \neq 0$ 。按照 Cramer 法则，此方程的解为：

$x_i = \frac{D_i}{D}, i=1,2,3 \cdots 20$ 。如解  $n$  阶方程组，要计算  $n+1$  个  $n$  阶行列式的值，每个行列式按 Laplace 展开计算，总共需要做  $(n+1)n!$  次乘法运算。本题  $n=20$ ，需要进行  $21! \approx 5.11 \times 10^{19}$  以上的乘法运算。设用每秒可做一亿次乘法的计算机，一年可以做

$365 \times 24 \times 60 \times 60 \times 10^8 \approx 3.15 \times 10^{15}$  次乘法。所以在此计算机上用 Gramer 法则解 20 阶的线性方程组，需要的时间在  $(5.11 \times 10^{19}) \div (3.15 \times 10^{15}) \approx 1.62 \times 10^4 = 1.62$  万年以上。这当然是没有实际意义的。

**例 1.2** 求超越方程  $\operatorname{tg} x + x = 0$  和  $0.25 + \operatorname{tg} x - 4.8889 \sin x = 0$  的根

**例 1.3** 不用计算器，求  $\sqrt{7}$  的近似值。

x	1	4	9
$y = \sqrt{x}$	1	2	3

**例 1.4** 计算定积分

$$I = \int_0^1 e^{-x^2} dx$$

**例 1.5** 在 7 块并排、形状大小相同的试验田上施化肥对水稻产量影响的试验，得到如下表所示的一组数据（单位：kg）

施化肥量 $x$	15	20	25	30	35	40	45
水稻产量 $y$	330	345	365	405	445	450	455

求施肥和产量的关系。

**例 1.6** 研究对象是连续的，我们只能了解到其有限个数据，比如，温度的变化，时间是连续的，温度是时间的函数， $T = f(t)$  如何画出温度的曲线图？

大量的根据实际问题所建立的模型只能求出近似解，也就是用“逼近”技术求出问题的解。本书的目的就是设计一些算法，对可以用数学模型描述的现实世界的现象求出近似解（数值解）。

用计算机解决科学计算问题需经历几个过程：由实际问题建立数学模型，根据数学模型提出求解的数值计算方法，编出程序，上机求出结果。通过以上过程，可以看出：数值计算方法是计算机、数学和应用科学之间的桥梁，是程序设计对数值结果进行分析的依据，是用计算机进行科学计算全过程的一个重要环节。

计算机实质上只会做加减乘除等基本运算。研究怎样通过计算机所执行的基本运算，求得各类问题的数值解或近似数值解，就是计算数学的根本课题。

**数值计算方法的研究对象：**数值计算方法主要研究适合于在计算机上使用的计算方法以及与此相关的理论，包括方法的收敛性、稳定性及误差分析。还要根据计算机的特点研究时间最短、需要计算机内存最少的计算方法。

### 1.3 算法

**算法的定义：**解决问题的一系列有序的步骤。

描述算法可以有不同的方式。例如，可以用日常语言和数学语言加以叙述，也可以借助形式语言（算法语言）给出精确的说明，也可以用框图直观地显示算法的全貌。下面举个简单的例子。

**例 1.7：**求解二元一次联立方程组

$$\begin{cases} a_{11}x_1 + a_{12}x_2 = b_1 \\ a_{21}x_1 + a_{22}x_2 = b_2 \end{cases}$$

这个方程组的行列式解法可表述如下：

首先判别

$$D = a_{11}a_{22} - a_{21}a_{12}$$

是否为零，存在两种可能：

(1) 如果  $D \neq 0$ ，则令计算机计算

$$x_1 = (b_1a_{22} - b_2a_{12})/D, \quad x_2 = (b_2a_{11} - b_1a_{21})/D$$

输出计算的结果  $x_1, x_2$ 。

(2) 如果  $D=0$ ，则或是无解，或有无穷多组解，这是奇异的情形。

下面用框图（也称流程图）来形象地描述上面的算法。



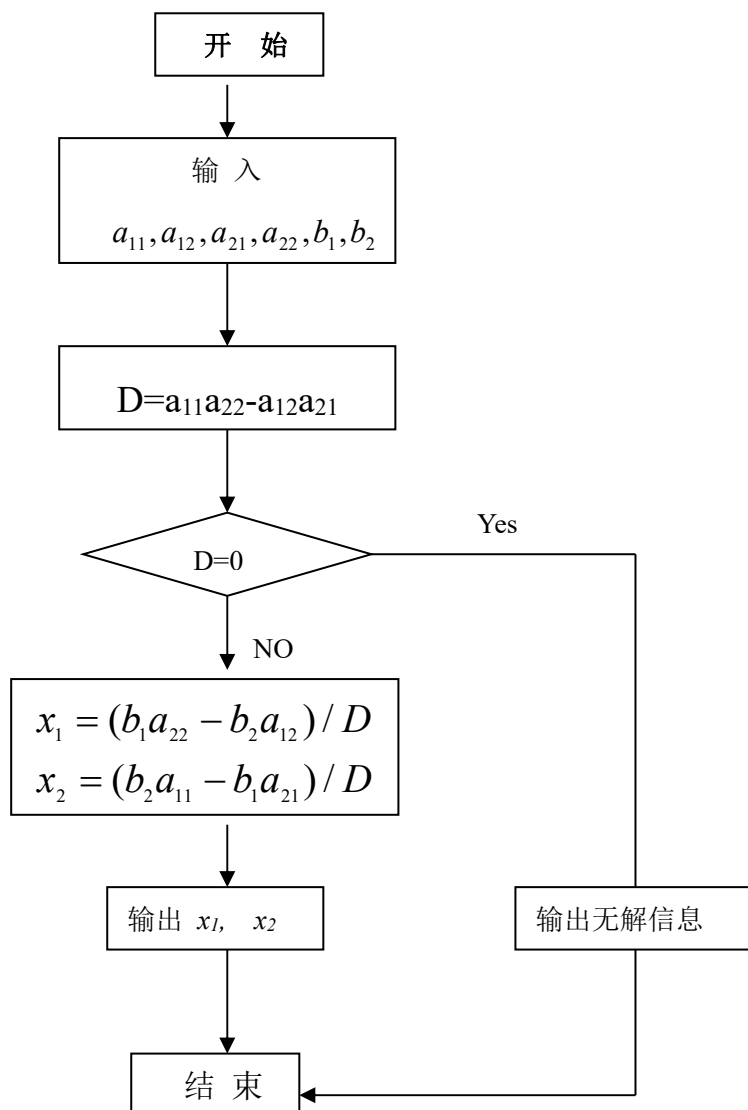

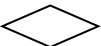


图 1.1 框图

这里我们使用了两种形式的框，一种是矩形框 ，称为叙述框，计算公式就填在这种框内；另一种是菱形框 ，称为判断框，表示算法的判断检查部分。

#### 1.4 数值计算中的误差

用数值方法解决科学研究或工程技术中的实际问题，一般来说，产生误差是不可避免的，根本不存在绝对的严格和精确。但是我们可以认识误差，从而控制误差，使之局限于最小（或尽量小）的范围内。

### 1.4.1 误差的来源

运用数学工具解决实际问题，可能产生的误差主要有以下几类。

#### 1. 描述误差

在将实际问题转化为数学模型的过程中，为了使数学模型尽量简单，以便于分析或计算，往往要忽略一些次要的因素，进行合理的简化。这样，实际问题与数学模型之间就产生了误差，这种误差称为描述误差。由于这类误差难于作定量分析，所以在计算方法中，总是假定所研究的数学模型是合理的，对描述误差不作深入的讨论。

#### 2. 观测误差

在数学模型中，一般都含有从观测（或实验）得到的数据，如温度、时间、速度、距离、电流、电压等等。但由于仪器本身的精度有限或某些偶然的客观因素会引入一定的误差，这类误差叫做观测误差。通常根据测量工具或仪器本身的精度，可以知道这类误差的上限值，所以无须在计算方法课程中作过多的研究。

#### 3. 截断误差

在数值计算中，常用收敛的无穷级数的前几项来代替无穷级数进行计算。如

$$e^x = 1 + x + \frac{x^2}{2!} + \cdots + \frac{x^n}{n!} + \cdots \quad (1.5)$$

当 $|x|$ 很小时，可以取（1.5）的前两项来近似代替 $e^x$ 的计算，即

$$e^x \approx 1 + x \quad (|x| \text{ 很小时}) \quad (1.6)$$

由泰勒定理可知，这时 $E(x) = 1 + x$ 与 $e^x$ 的误差是

$$e^x - E(x) = \frac{x^2}{2} e^{\theta x} \quad (0 < \theta < 1) \quad (1.7)$$

在计算中被抛弃了，这类误差叫做截断误差。截断误差的大小，直接影响数值计算的精度，所以它是数值计算中必须十分重视的一类误差。

#### 4. 舍入误差

无论用计算机、计算器计算还是笔算，都只能用有限位小数来代替无穷小数或用位数较少的小数来代替位数较多的有限小数，如：

$$\pi = 3.1415926\cdots \quad \frac{1}{3} = 0.3333\cdots$$

$$x = 8.123456$$

经四舍五入后取含小数点后四位数字的数来表示它们时，便有误差

$$\varepsilon_1 = \pi - 3.1416 = -0.0000074\cdots$$

$$\varepsilon_2 = \frac{1}{3} - 0.333 = 0.000033\cdots$$

$$\varepsilon_3 = x - 8.1235 = 0.000044$$

这类误差叫做舍入误差。当然，计算过程中，这类误差往往是有舍有入的，而且单从一次的舍入误差来看也许不算大，但数值计算中，往往要进行成千上万次四则运算，因而就会有成千上万个舍入误差产生，这些误差一经迭加或传递，对精度可能有较大的影响。所以，作数值计算时，对舍入误差应予以足够的重视。

显然，上述四类误差都会影响计算结果的准确性，但描述误差和观测误差往往是计算工作者不能独立解决的，是需要与各有关学科的科学工作者共同研究的问题，因此，在计算方法课程中，主要研究截断误差和舍入误差（包括初始数据的误差）对计算结果的影响。

### 1.4.2 绝对误差、相对误差和有效数字

在不同的场合下，表示近似数精确度的方法各有不同，下面将介绍数值运算误差的基本概念。

#### 1. 绝对误差与绝对误差限

**定义 1:** 设  $x$  是准确值， $x^*$  是  $x$  的一个近似值，称

$$e(x) = x - x^* \quad (1.8)$$

是近似值  $x^*$  的**绝对误差**，简称为**误差**。

由于准确值  $x$  往往是未知的，因此也无法求出绝对误差  $e(x)$  是多少，但实际问题往往可以估计出绝对误差的一个上界。

**定义 2:** 若已知  $\varepsilon^* > 0$ ，使

$$|e(x)| = |x - x^*| \leq \varepsilon^* \quad (1.9)$$

则称  $\varepsilon^*$  近似值  $x^*$  的**绝对误差限**，简称为**误差限**。

对一般情况  $|x - x^*| \leq \varepsilon^*$ ，即

$$x^* - \varepsilon^* \leq x \leq x^* + \varepsilon^*$$

有时也表示为

$$x = x^* \pm \varepsilon^*$$

绝对误差的大小，在许多情况下还不能完全刻画一个近似值的准确程度，例如测量 1000 米和 1 米两个长度，若它们的绝对误差都是 1 厘米，显然前者的测量比较准确。由此可见，决定一个量的近似值的精确度，除了考虑绝对误差的大小外，还需考虑该量本身的大小，为此引入相对误差的概念。

## 2. 相对误差和相对误差限

**定义 3:** 称近似值  $x^*$  的误差  $e(x)$  与准确值  $x$  的比值

$$\frac{e(x)}{x} = \frac{x - x^*}{x} \quad (1.10)$$

为近似值  $x^*$  的相对误差，记作  $e_r^*$ ，（在实际计算时，(1.10) 式中的  $x$  通常用近似值  $x^*$  代替）。

类似定义 2，有

**定义 4** 相对误差的上界，称为**相对误差限**，即

$$|e_r^*| = \left| \frac{x - x^*}{x} \right| \leq \frac{\varepsilon^*}{|x^*|} = \varepsilon_r^* \quad (1.11)$$

## 3. 有效数字

众所周知，当准确数  $x$  有很多位数时，常常按“四舍五入”的原则去得到  $x$  的近似数。例如  $\pi = 3.1415926 \dots$  按舍入原则，若取 4 位小数得  $\pi = 3.1416$ ，取 5 位则有  $\pi = 3.14159$ 。用有效数字这个概念来表示一个近似数的准确程度，其定义为

**定义 5:** 如果

$$|e^*| = |x - x^*| \leq \frac{1}{2} \times 10^{-n} \quad (1.12)$$

则说  $x^*$  近似表示  $x$  准确到小数后第  $n$  位，并从这第  $n$  位起直到最左边的非零数字之间的一切数字都称为有效数字，并把有效数字的位数称为有效位数。

由上述定义

$$|\pi - 3.1416| \leq \frac{1}{2} \times 10^{-4} \quad |\pi - 3.14159| \leq \frac{1}{2} \times 10^{-5}$$

$$\pi - 3.14 = 0.0015926 \quad \text{有效数位为 3 位}$$

$$\pi - 3.1416 = -0.0000074 \quad \text{有效数位为 5 位}$$

$$\pi - 3.1415 = 0.0000926 \quad \text{有效数位为 4 位}$$

近似值的有效数字位数越多，近似程度就越好。

绝对误差、相对误差及有效数字都是反映误差或近似程度的量，它们相互之间存在某些关系。

定义 4 描述了绝对误差限与相对误差限的关系，下面的定义 6 表明有效数字与绝对误差限的关系。

**定义 6:** 若将准确值  $x$  的近似值  $x^*$  表示成标准形式

$$x^* = \pm 0.a_1 a_2 \cdots a_n \times 10^m \quad (1.13)$$

而其误差限

$$|x - x^*| \leq \frac{1}{2} \times 10^{m-n} \quad (1.14)$$

则说近似值  $x^*$  具有  $n$  位有效数字。这里  $n$  为正整数， $m$  为整数，每个  $a_i (i=1,2,\dots,n)$

均为 0,1, ..., 9 中的一个数字， $a_1 \neq 0$ 。

**例 1.8** 若  $x^*=3587.64$  是  $x$  的具有 6 位有效数字的近似值，求误差限。

**解** 将  $x^*$  按 (1.13) 式写成标准形式

$$x = 3587.64 = 0.358764 \times 10^4$$

于是，在定义 6 中  $m=4$ ,  $n=6$

$$|x - x^*| \leq \frac{1}{2} \times 10^{m-n} = \frac{1}{2} \times 10^{4-6} = \frac{1}{2} \times 10^{-2}$$

下面讨论有效数字和相对误差限的关系。

**定理 1.1:** 若  $x^*$  具有  $n$  位有效数字，则其相对误差限满足

$$e_r^* \leq \frac{1}{2a_1} \times 10^{-(n-1)} \quad (1.15)$$

反之，若  $x^*$  的相对误差限  $e_r^*$  满足

$$e_r^* \leq \frac{1}{2(a_1 + 1)} \times 10^{-(n-1)} \quad (1.16)$$

则近似值  $x^*$  至少具有  $n$  位有效数字。

**证明** 若  $x^* = 0.a_1a_2 \cdots a_n \times 10^m$  具有  $n$  位有效数字，由定义 6，

$$|x - x^*| \leq \frac{1}{2} \times 10^{m-n}$$

相对误差

$$|e_r^*| = \frac{|x - x^*|}{|x^*|} \leq \frac{1}{2a_1} \times 10^{m-n} \bullet \frac{1}{|x^*|}$$

由于

$$|x^*| = |0.a_1a_2 \cdots a_n \times 10^m| \geq a_1 \times 10^{m-1}$$

所以

$$|e_r^*| \leq \frac{1}{2a_1} \times 10^{-(n-1)}$$

反之，若 (1.16) 成立，即

$$e_r^* \leq \frac{1}{2(a_1 + 1)} \times 10^{-(n-1)}$$

利用

$$|x^*| \leq (a_1 + 1) \times 10^{m-1}$$

得

$$|x - x^*| = |e_r^*| \bullet |x^*| \leq \frac{1}{2(a_1 + 1)} \times 10^{-(n-1)} \times (a_1 + 1) \times 10^{m-1} = \frac{1}{2} \times 10^{m-n}$$

根据定义 6， $x^*$  至少具有  $n$  位有效数字。

**例 1.9:**  $x^* = 2.71828$ ， $x = e$ ，求相对误差限。

**解:** 因为:  $x^* = 2.71828 = 0.271828 \times 10^1$ ，由  $|e - 2.71828| \leq \frac{1}{2} \times 10^{-5}$ ， $a_1 = 2$ ， $n = 6$ ，由

定理 1 前一部分，有

$$e_r^* \leq \frac{1}{2a_1} \times 10^{-(n-1)} = \frac{1}{2 \times 2} \times 10^{-(6-1)} = \frac{1}{4} \times 10^{-5}$$

**例 1.10** 要使  $\sqrt{20}$  的近似值的相对误差小于 0.1%，应取几位有效数字？

**解:** 由于  $4 < \sqrt{20} < 5$ ，可确定  $a_1 = 4$ ，若相对误差限

$$\varepsilon_r^* \leq 0.1\%$$

利用

$$e_r^* \leq \frac{1}{2(a_1 + 1)} \times 10^{-(n-1)} = \varepsilon_r^*$$

则有效数字位数  $n$  应满足不等式

$$\frac{1}{2(4+1)} \times 10^{-(n-1)} < 0.1\%$$

解出  $n=4$ , 即应有 4 位有效数字。

### 1.5 数值计算中应该注意的一些原则

由上述讨论可知,误差分析在数值计算中是一个很重要又很复杂的问题。因为在数值计算中每一步运算都可能产生误差,而解决一个科学计算问题往往要经过成千上万次运算,如果每一步运算都分析误差,显然是不可能的,也是不必要的。人们经常通过对误差的某种传播规律的分析,指出在数值计算中应该注意的一些原则,有助于鉴别计算结果的可靠性并防止误差危害现象的产生,下面我们给出在数值计算中应该注意的一些原则。

#### 1.5.1 要使用数值稳定的算法

所谓算法,就是给定一些数据,按着某种规定的次序进行计算的一个运算序列。它是一个近似的计算过程,我们选择一个算法,主要要求它的计算结果能达到给定的精度。一般而言,在计算过程中初始数据的误差和计算中产生的舍入误差总是存在的,而数值解是逐步求出的,前一步数值解的误差必然要影响到后一步数值解的精度。我们把运算过程中舍入误差增长可以控制的计算公式称为数值稳定的,否则是数值不稳定的。只有稳定的数值方法才可能给出可靠的计算结果,不稳定的数值方法毫无实用价值。下面用一个例子来简单介绍一下稳定性的概念。

**例 1.11:** 求  $I_n = \int_0^1 \frac{x^n}{x+5} dx$  ( $n = 0, 1, 2, \dots, 8$ ) 的值。

解: 由于

$$I_n + 5I_{n-1} = \int_0^1 \frac{x^n + 5x^{n-1}}{x+5} dx = \int_0^1 x^{n-1} dx = \frac{1}{n}$$

初值

$$I_0 = \int_0^1 \frac{1}{x+5} dx = \ln 6 - \ln 5 = \ln(1.2)$$

于是可建立递推公式

$$\begin{cases} I_0 = \ln(1.2) \\ I_n = \frac{1}{n} - 5I_{n-1}, \end{cases} \quad (n = 1, 2, \dots, 8) \quad (1.17)$$

若取  $I_0 = \ln(1.2) \approx 0.182$

按 (1.17) 式就可以逐步算得

$$I_1 = 1 - 5I_0 \approx 0.09$$

$$I_2 = \frac{1}{2} - 5I_1 \approx 0.05$$

$$I_3 = \frac{1}{3} - 5I_2 \approx 0.083$$

$$I_4 = \frac{1}{4} - 5I_3 \approx -0.165$$

因为在  $[0, 1]$  上被积函数  $\frac{x^n}{x+5} \geq 0$  (仅当  $x=0$  时为零), 且当  $m > n$  时,  $\frac{x^n}{x+5} \geq \frac{x^m}{x+5}$  (仅当

$x=0$  时, 等号成立), 所以  $I_n (n=0, 1, 2, \dots, 8)$  是恒正的, 并有  $I_0 > I_1 > I_2 > \dots > I_8 > 0$ 。

在上述计算结果中,  $I_4$  的近似值是负的, 这个结果显然是错的。为什么会这样呢? 这就是误差传播所引起的危害。由递推公式 (1.13) 可看出,  $I_{n-1}$  的误差扩大了 5 倍后传给  $I_n$ , 因而初值  $I_0$  的误差对以后各步计算结果的影响, 随着  $n$  的增大愈来愈严重。这就造成  $I_4$  的计算结果严重失真。

如果改变计算公式, 先取一个  $I_n$  的近似值, 用下面的公式倒过来计算  $I_{n-1}, I_{n-2}, \dots$

即:

$$I_{k-1} = \frac{1}{5K} - \frac{1}{5} I_K \quad (K = n, n-1, \dots, 1) \quad (1.18)$$

情况就不同了。我们发现  $I_k$  的误差减小到  $\frac{1}{5}$  后传给  $I_{k-1}$ , 因而初值的误差对以后各步的计算结果的影响是随着  $n$  的增大而愈来愈小。

由于误差是逐步衰减的, 初值  $I_n$  可以这样确定, 不妨设  $I_9 \approx I_{10}$ , 于是由

$$I_9 = \frac{1}{50} - \frac{1}{5} I_{10}$$

可求得  $I_9 \approx 0.017$ , 按公式 (1.14) 可逐次求得

$$I_8 \approx 0.019 \quad I_7 \approx 0.021$$

$$I_6 \approx 0.024 \quad I_5 \approx 0.028$$



$$I_4 \approx 0.034 \quad I_3 \approx 0.043$$

$$I_2 \approx 0.058 \quad I_1 \approx 0.088$$

$$I_0 \approx 0.182$$

显然，这样算出的  $I_0$  与  $\ln(1.2)$  的值比较符合。虽然初值  $I_9$  很粗糙，但因为用公式 (1.18) 计算时，误差是逐步衰减的，所以计算结果相当可靠。

比较以上两个计算方案，显然，前者是一个不稳定的算法，后者是一个稳定算法。对于一个稳定的计算过程，由于舍入误差不增大，因而不具体估计舍入误差也是可用的。而对于一个不稳定的计算过程，如计算步骤太多，就可能出现错误结果。因此，在实际应用中应选用数值稳定的算法，尽量避免使用数值不稳定的算法。

### 1.5.2 要避免两个相似数相减

在数值计算中，两个相近的数作减法时有效数字会损失。例如，求：

$$y = \sqrt{x+1} - \sqrt{x} \quad (1.19)$$

之值，当  $x = 1000$ ， $y$  的准确值为 0.01580。若两者直接相减

$$y = \sqrt{1001} - \sqrt{1000} = 31.64 - 31.62 = 0.02$$

这个结果只有一位有效数字，损失了三位有效数字，从而绝对误差和相对误差都变得很大，严重影响计算结果的精度。若将公式 (1.19) 处理成

$$y = \sqrt{x+1} - \sqrt{x} = \frac{1}{\sqrt{x+1} + \sqrt{x}}$$

按此公式可求得  $y = 0.01581$ ，则  $y$  有 3 位有效数字，可见改变计算公式，可以避免两相近数相减引起有效数字损失，而得到较精确的结果。

类似地，

$$\ln x - \ln y = \ln \frac{x}{y}$$

$$\sin(x + \varepsilon) - \sin x = 2 \cos\left(x + \frac{\varepsilon}{2}\right) \sin \frac{\varepsilon}{2}$$

当  $\varepsilon$  很小时，当  $x$  和  $y$  很接近时，采用等号右边的算法，有效数字就不损失。

一般地，当  $f(x) = f(x^*)$  时，可用 *Taylor* 展开

$$f(x) - f(x^*) = f'(x^*)(x - x^*) + \frac{f''(x^*)}{2!}(x - x^*)^2 + \dots$$

取右端的有限项近似左端。若无法改变算法，直接计算时就要多保留几位有效数字。

### 1.5.3 绝对值太小的数不宜作除数

算法语言中已讲述，在机器上若用很小的数作除数会溢出，而且当很小的数稍有一点误差时，对计算结果影响很大。

**例 1.12:**

$$\frac{2.7182}{0.001} = 2718.2$$

如分母变为 0.0011，也即分母只有 0.0001 的变化时

$$\frac{2.7182}{0.0011} = 2471.1$$

商却引起了巨大变化，因此，在计算过程中既要避免两个相近数相减，更要避免再用这个差作除数。

### 1.5.4 采用递推的算法

计算机上使用的算法常采用递推化的形式，递推化的基本思想是把一个复杂的计算过程归结为简单过程的多次重复。这种重复在程序上表现为循环。递推化的优点是简化结构和节省计算量。

下面用多项式求值问题说明递推化方法。

**例 1.13:** 设对于给定的  $x$  求下列  $n$  次多项式的值。

$$P(x) = a_0 + a_1x + a_2x^2 + \dots + a_nx^n \quad (1.16)$$

解：1. 用一般算法，即直接求和法，可知乘法的次数为：  $1 + 2 + 3 + \dots + n = \frac{n(n+1)}{2}$ ；

加法次数为：  $n$

#### 2. 逐项求和法

令  $t_k = x^k$ ，则  $t_{k-1} = x^{k-1}$

所以  $t_k = xt_{k-1}$

记：  $u_k = a_0 + a_1x + \dots + a_{k-1}x^{k-1} + a_kx^k$

可以看出前  $K+1$  项部分和  $u_k$  等于前  $K$  项部分和  $u_{k-1}$  再加上第  $K+1$  项  $a_kx^k$ ，因此有

$$\begin{cases} t_k = xt_{k-1} \\ u_k = u_{k-1} + a_k t_k \end{cases} \quad k = 1, 2, \dots, n \quad (1.17)$$

初值应取为

$$\begin{cases} t_0 = 1 \\ u_0 = a_0 \end{cases} \quad (1.18)$$

容易看出逐步求和法所用乘法的次数为  $2n$ ，加法次数为  $n$ 。当  $n \geq 4$  后， $2n < \frac{n(n+1)}{2}$ 。

### 3. 秦九韶方法

首先将多项式改写为

$$\begin{aligned} P(x) &= a_n x^n + a_{n-1} x^{n-1} + \dots + a_1 x + a_0 \\ &= (a_n x^{n-1} + a_{n-1} x^{n-2} + \dots + a_1) x + a_0 \\ &= ((a_n x^{n-2} + a_{n-1} x^{n-3} + \dots + a_2) x + a_1) x + a_0 \\ &= \dots \\ &= (\dots (a_n x + a_{n-1}) x + a_{n-2}) x + \dots + a_1) x + a_0 \end{aligned}$$

$$\text{令 } v_k = (\dots (a_n x + a_{n-1}) x + \dots + a_{n-(k-2)}) x + a_{n-(k-1)} x + a_{n-k}$$

则递推公式为：

$$\begin{cases} v_k = v_{k-1} x + a_{n-k} \\ v_0 = a_n \end{cases} \quad k = 1, 2, \dots, n \quad (1.19)$$

此方法称为秦九韶方法,也是多项式求值中常用的方法。

此方法的计算量为：乘法  $n$  次，加法  $n$  次。

同 1、2 相比，秦九韶方法的计算量小，且逻辑结构简单。

#### 例 1.14：用秦九韶方法求多项式

$$P(x) = 1 + x + 0.5x^2 + 0.16667x^3 + 0.04167x^4 + 0.00833x^5$$

在  $x = -0.2$  的值。

解：

$K$	$a_{5-K}$	$v_K$	
0	0.00833	0.00833	$v_0 = a_5$
1	0.04167	0.04	$v_1 = v_0 x + a_4$
2	0.16667	0.15867	$v_2 = v_1 x + a_3$
3	0.5	0.46827	$v_3 = v_2 x + a_2$

4	1	0.90635	$v_4 = v_3x + a_1$
5	1	0.81873	$v_5 = v_4x + a_0$

$$P(-0.2) = 0.81873$$

## 1.6 算法的优劣

我们知道，计算机的特点是运算速度快，存贮的信息量大，并能自动完成极其复杂的计算过程。计算机功能虽然很强，但是否可降低对算法的要求呢？许多事实说明，如果算法选择不当，计算机的利用率就得不到充分发挥，有时甚至不能得到满意的解答。一个好的算法，要求有以下几个优点：

### 1. 计算量小

因此，计算量大小是衡量算法优劣的一项重要标准。

而后面将要介绍的高斯消去法求解同样的方程组，乘除法的运算次数不超过 3074 次，与 Gramer 法则比较计算量有天壤之别。方程组的阶数增大，计算量的差别还会更大。应当指出，数值方法的计算时间，由计算机速度和数值方法的效率而定。从某种意义上来说，对于减少计算时间，提高数值方法的效率甚至比提高计算机速度更为重要，因为算法研究所需要的代价要小得多。

在估计计算量时，我们将区分主次抓住计算过程中费时较多的环节。比如，由于加减操作的机器时间比乘除少得多，对和式  $S = \sum a_k b_k$  可以忽略加法而只统计乘法的次数。又如算式  $S = \sum_K a_k f(x_k)$  中需要多次计算函数值  $f(x_k)$ ，每求一次  $f(x_k)$  时，通常需要进行多次加减乘除运算，因此对算式  $S = \sum_K a_k f(x_k)$  只要统计调用  $f$  的次数就可以了。

### 2. 存贮量小

计算程序所占用工作单元的数量称为算法的内存量。尽管计算机能贮存大量信息，但计算大型算题时有些微机仍不能使用。因此，尽量节约存贮量有经济价值，是衡量算法质量的又一标准。

### 3. 逻辑结构简单

设计算法时应考虑的另一个因素是逻辑结构问题。虽然计算机能够执行极其复杂的计算方案，但是计算方案的每个细节都需要编程人员制定。因此算法的逻辑结构应尽量简单，才

能使编程序度、维修程序和使用程序比较方便。

由此可见，虽然计算机是一种强有力的计算工具，但不能因此忽略对算法的研究。应该以计算量大小、存贮量多少、逻辑结构是否简单作为评定算法优劣的标准。

#### 4. 数值稳定

综上所述，我们总结出一个高质量的算法的特点：数值稳定，计算量小，存储量小，逻辑结构简单。

### 1.7 复习几个常用公式

微分中值定理和泰勒公式是学习数值计算方法的重要工具，许多数值解法可以直接从泰勒公式中得到，在使用这些方法时，所包含的误差估计也可以直接得到。同学们在微积分中已熟悉这些内容，但我们仍将作一个简略的介绍。

#### 1.7.1 微分中值定理

微分中值定理是反映函数与导数之间联系的重要定理。导数是由极限定义的，而计算机只能进行有限次运算，在计算机上不得不用函数值来进行微分或积分运算，它们是通向微分学的许多重要应用的桥梁。为了使大家能了解中值定理在几何上的直观背景，我们考察下述几何事实：曲线  $\widehat{AB}$  中间，有一点  $P$ ，在那里，曲线的切线平行于割线  $\overline{AB}$ 。

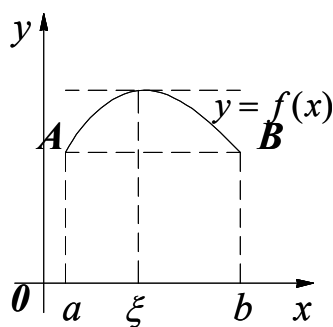


图 1.2

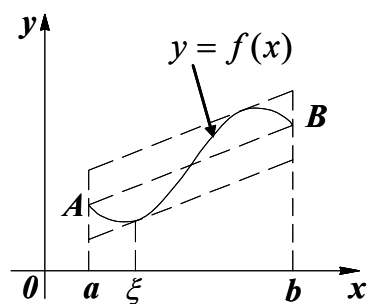


图 1.3

现在就来揭示这一几何事实所包含的数量关系。

设曲线  $\widehat{AB}$  的方程为：

$$y = f(x) \quad a \leq x \leq b$$

换句话说, 曲线段  $\widehat{AB}$  是函数  $f(x)$  的图像, 则曲线在点  $P(\xi, f(\xi))$  处的切线斜率就是  $f'(\xi)$ , 而割线  $\overline{AB}$  的斜率等于

$$\frac{f(b) - f(a)}{b - a}$$

点  $P$  处的切线平行于割线  $\overline{AB}$ , 也就是二者斜率相等, 于是得到一个公式

$$f(b) - f(a) = f'(\xi)(b - a)$$

可归纳出以下定理。

### 定理 1 (Roll 中值定理)

设函数  $f(x)$  在  $[a, b]$  上连续, 在  $(a, b)$  内可微, 且  $f(a) = f(b)$ , 则存在  $\xi \in (a, b)$ , 使得  $f'(\xi) = 0$ 。

### 定理 2 (Lagrange 中值定理)

设函数  $f(x)$  在  $[a, b]$  上连续, 在  $(a, b)$  内可微, 则存在  $\xi \in (a, b)$ , 使得

$$f'(\xi) = \frac{f(b) - f(a)}{b - a} \quad (1.20)$$

注1.1 若  $f(a) = f(b)$ , 则 Lagrange 中值定理化为 Roll 中值定理。

注1.2 公式 (1.20) 称为 Lagrange 中值公式, 它还有其他表示形式, 若令  $\theta = \frac{\xi - a}{b - a}$

则  $0 \leq \theta \leq 1$ , 这时  $\xi = a + \theta(b - a)$ , 公式 (1.20) 可以写成

$$f(b) - f(a) = f'(a + \theta(b - a))(b - a) \quad (1.21)$$

### 1.7.2. 泰勒 (Taylor) 公式

无论是手算或用其它计算工具去计算函数值, 都归结为做有限次的四则运算, 由此可见, 研究如何用多项式近似地表达一个给定函数的问题是很必要的。一个给定的函数, 如果能用多项式近似表达, 那么, 通过计算多项式的值, 就可以得到给定函数的近似值。

用多项式近似地表达一个函数, 方式很多, 我们这里要谈的只是其中一种, 我们的问题是: 给了一个函数  $f(x)$ , 要找一个在指定点  $x = x_0$  (为简单计, 以下先设  $x_0 = 0$ ) 附近与  $f(x)$  很近似的多项式

$$P(x) = a_0 + a_1x + \cdots + a_nx^n \quad (1.22)$$

那么, 要找的多项式应满足什么样的条件呢?

从几何上看,  $y=p(x)$ 与 $y=f(x)$ 代表两条曲线, 要想多项式 $p(x)$ 在点 $x=0$ 附近与函数 $f(x)$ 很近似, 就是要曲线 $y=p(x)$ 与曲线 $y=f(x)$ 在点 $(0, f(0))$ 附近很靠近。

很明显, 首先应该要求曲线 $y=p(x)$ 与曲线 $y=f(x)$ 交于点 $(0, f(0))$  (图 1.4) 也就是要求多项式 $p(x)$ 满足条件

$$p(0) = f(0) \quad (1.23)$$

要想曲线 $y=p(x)$ 与曲线 $y=f(x)$ 在 $(0, f(0))$ 附近靠得更近, 应该进而要求它们在点 $(0, f(0))$ 处有公共切线 (图 1.4), 也就是要求多项式 $p(x)$ 同时满足下列两个条件:

$$p(0) = f(0) \quad p'(0) = f'(0) \quad (1.24)$$

满足这两个条件的多项式 $p(x)$ 可以有很多。试看图(1.4)曲线 $y=p_i(x)$ 都与曲线 $y=f(x)$ 在点 $(0, f(0))$ 处有公共切线, 但与曲线 $y=f$ 的靠近程度却可以不一样。

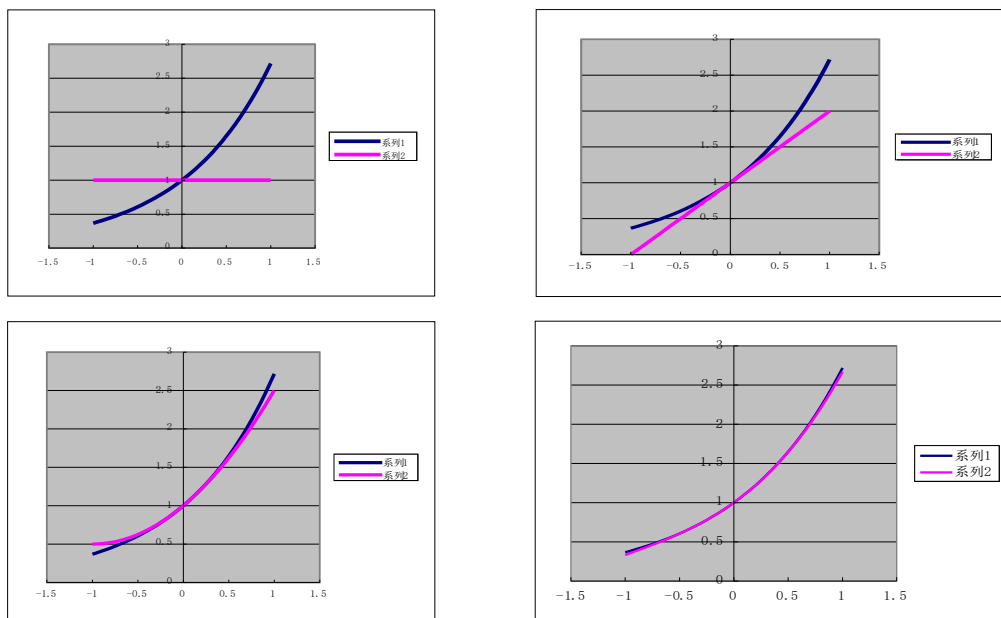


图 1.4

要想从中找出与曲线 $y=f(x)$ 靠得更近的, 就要进一步来考察它们的弯曲情况。我们要找的显然是与曲线 $y=f(x)$ 向同一方向弯曲, 这条曲线 $y=p(x)$ 应该满足条件 $p''(0) = f''(0)$ , 即 $p(x)$ 应同时满足下列三个条件:

$$p(0) = f(0), \quad p'(0) = f'(0), \quad p''(0) = f''(0) \quad (1.25)$$

由此可以推想, 假如在点 $x=0$ 处,  $p(x)$ 与 $f(x)$ 的三阶导数, 以至更高阶导数都相等, 那么在点 $(0, f(0))$ 附近, 曲线 $y=p(x)$ 与曲线 $y=f(x)$ 的靠近程度就会更高。

综上所述，所要找的多项式  $p(x)$  应满足下列条件：

$$p(0) = f(0), \quad p'(0) = f'(0), \quad p''(0) = f''(0), \dots, p^{(n)}(0) = f^{(n)}(0) \quad (1.26)$$

现在，根据条件 (1.26)，我们就可以确定所要找的近似多项式的具体形状了。

首先，由 (1.22) 得到

$$p(0) = a_0, \quad p'(0) = a_1, \quad p''(0) = 2!a_2, \dots, p^{(n)}(0) = n!a_n$$

于是，条件 (1.26) 就变成

$$a_0 = f(0), \quad a_1 = f'(0), \quad 2!a_2 = f''(0), \dots, n!a_n = f^{(n)}(0)$$

即

$$a_0 = f(0), \quad a_1 = f'(0), \quad a_2 = \frac{f''(0)}{2!}, \dots, \\ a_n = \frac{1}{n!} f^{(n)}(0)$$

代入(1.22)式，就得到：

$$p(x) = f(0) + f'(0)x + \frac{f''(0)}{2!}x^2 + \dots + \frac{f^{(n)}(0)}{n!}x^n \quad (1.27)$$

这就是我们要找的多项式，这个多项式称为  $f(x)$  在  $x_0 = 0$  这点的泰勒多项式，记为  $p_n(x)$ 。为

了进一步研究这个多项式和  $f(x)$  究竟差多少，我们必须分析误差项

$$r_n(x) = f(x) - p_n(x)$$

为推导  $r_n(x)$  的表达式，我们假定  $f(x)$  在点  $x_0 = 0$  附近有直到  $n+1$  阶的连续导数，注意从  $r_n(x)$  的定义可见

$$r_n(0) = r'_n(0) = \dots = r^{(n)}(0) = 0$$

$$r^{(n+1)}(x) = f^{(n+1)}(x)$$

于是由微积分学基本定理及分部积分法，得

$$\begin{aligned} r_n(x) &= \int_0^x r'_n(t) dt \\ &= (t-x) r'_n(t) \Big|_0^x - \int_0^x r''_n(t)(t-x) dt \\ &= -\int_0^x r''_n(t)(t-x) dt = \int_0^x r''_n(t)(x-t) dt \end{aligned}$$

继续实行分部积分，最后利用  $r^{(n+1)}(x) = f^{(n+1)}(x)$ 。便得：



$$r_n(x) = \frac{1}{n!} \int_0^x (x-t)^n f^{(n+1)}(t) dt \quad (1.28)$$

(1.28)式称为 $f(x)$ 在点 $x_0 = 0$ 的积分型 Taylor 余项公式。由(1.23)式, 我们还可以导出另一种形式的余项公式:

设 $f^{(n+1)}(x)$ 在 $0 \leq t \leq x$ 上的最大值为 $M$ , 最小值为 $m$ , 不妨设 $x > 0$ , 此时 $(x-t)^n \geq 0$  ( $0 \leq t \leq x$ )。

因此

$$m \cdot \frac{1}{n!} \int_0^x (x-t)^n dt \leq r_n(x) \leq M \cdot \frac{1}{n!} \int_0^x (x-t)^n dt$$

而

$$\int_0^x (x-t)^n dt = \frac{1}{n+1} x^{n+1}$$

所以上式即

$$m \leq \frac{(n+1)!}{x^{n+1}} r_n \leq M$$

由连续函数的介值定理, 应有 $\xi, 0 \leq \xi \leq x$ , 使

$$\frac{(n+1)!}{x^{n+1}} r_n(x) = f^{(n+1)}(\xi)$$

即

$$r_n(x) = \frac{f^{(n+1)}(\xi)}{(n+1)!} x^{n+1} \quad (1.29)$$

此式为 $f(x)$ 在 $x_0 = 0$ 点的拉格朗日型余项公式, 也是最常用的一种 Taylor 余项公式。

$r_n(x)$ 通常称为 $n$ 阶余项, 利用 $r_n(x)$ , 我们可以写成:

$$\begin{aligned} f(x) &= p_n(x) + r_n(x) \\ &= f(0) + f'(0)x + \cdots + \frac{f^{(n)}(0)}{n!} x^n + r_n(x) \end{aligned} \quad (1.30)$$

(1.30)式称为 $f(x)$ 在 $x_0 = 0$ 点的 Taylor 公式。对任意点 $x_0$ ,  $f(x)$ 在 $x_0$ 点的 Taylor 公式是:

$$f(x) = f(x_0) + f'(x_0)(x-x_0) + \cdots + \frac{f^{(n)}(x_0)}{n!} (x-x_0)^n + r_n(x)$$

### 例 1.15 Taylor 展开

#### 1.8 函数计算的误差估计

设一元函数  $f(x)$  具有二阶导数, 自变量  $x$  的一个近似值  $x^*$ ,  $f(x)$  的近似值为  $f(x^*)$ , 用  $f(x)$  在  $x^*$  点的 Taylor 展开估计误差, 可得

$$|f(x) - f(x^*)| \leq |f'(x^*)(x - x^*)| + \frac{1}{2} |f''(\xi)(x - x^*)^2|$$

其中  $\xi$  在  $x$  与  $x^*$  这间, 如果  $f'(x^*) \neq 0$ ,  $|f''(\xi)|$  与  $|f'(x^*)|$  有相同数量级, 而  $e_r^* \geq |x - x^*|$  很小, 则可得

$$\begin{aligned} ef(x^*) &\approx |f'(x^*)| e_r(x^*), \\ e_r f(x^*) &\approx \left| \frac{f'(x^*)}{f(x^*)} e_r(x^*) \right| \end{aligned} \quad (1.31)$$

分别为  $f(x^*)$  的一个近似误差限与相对误差限。

如果  $f$  为多元函数, 自变量为  $x_1, \dots, x_n$ , 其近似值为  $x_1^*, \dots, x_n^*$ , 则类似于一元函数可用多元函数  $f(x_1, x_2, \dots, x_n)$  的 Taylor 展开, 取一阶近似的误差限

$$ef(x_1^*, \dots, x_n^*) \approx \sum_{i=1}^n \left| \frac{\partial f(x_1^*, \dots, x_n^*)}{\partial x_i} \right| e(x_i^*) \quad (1.32)$$

及相对误差限

$$e_r f(x_1^*, \dots, x_n^*) \approx \sum_{i=1}^n \left| \frac{\partial f(x_1^*, \dots, x_n^*)}{\partial x_i} \right| \frac{e(x_i^*)}{|f(x_1^*, \dots, x_n^*)|} \quad (1.33)$$

若把 (1.33) 用到两个或多个数的算术运算中, 则可得到近似数  $x_1^*$  及  $x_2^*$  的四则运算误差估计:

$$e(x_1^* \pm x_2^*) = e(x_1^*) + \delta(x_2^*) \quad (1.34)$$

$$e(x_1^* x_2^*) = |x_1^*| e(x_2^*) + |x_2^*| e(x_1^*) \quad (1.35)$$

$$e\left(\frac{x_1^*}{x_2^*}\right) = \frac{|x_1^*| e(x_2^*) + |x_2^*| e(x_1^*)}{|x_2^*|^2}, \quad x_2^* \neq 0 \quad (1.36)$$

**例 1.16** 已测得某场地长  $l$  的值为  $l^* = 110\text{m}$ , 宽  $d$  的值为  $d^* = 80\text{m}$ , 已知  $|l - l^*| \leq 0.2\text{m}$ ,  $|d - d^*| \leq 0.1\text{m}$ , 试求面积  $S = ld$  的绝对误差限与相对误差限。

解 因  $S = ld$ ,  $\frac{\partial S}{\partial l} = d$ ,  $\frac{\partial S}{\partial d} = l$ , 由 (1.2.5) 知

$$e(S^*) = \left| \left( \frac{\partial S}{\partial l} \right)^* \right| e(l^*) + \left| \left( \frac{\partial S}{\partial d} \right)^* \right| e(d^*)$$

其中  $\left( \frac{\partial S}{\partial l} \right)^* = d^* = 80\text{m}$ ,  $\left( \frac{\partial S}{\partial d} \right)^* = l^* = 110\text{m}$ ,  $e(l^*) = 0.2\text{m}$ ,  $e(d^*) = 0.1\text{m}$ , 从而有

$$e(S^*) = 80 \times 0.2 + 110 \times 0.1 = 27\text{m}^2$$

相对误差限为

$$e_r(S^*) = \frac{e(S^*)}{|S^*|} = \frac{27}{80 \times 110} = 0.31\%$$

### 病态问题与条件数

对一个数值问题, 往往由于问题本身而使计算结果相对误差很大, 这种问题就是病态问题。例如计算函数值  $f(x)$ , 若  $x$  的近似值为  $x^*$ , 其相对误差为  $\frac{x - x^*}{x}$ , 函数值  $f(x^*)$  的相对误差为  $\frac{f(x) - f(x^*)}{f(x)}$ , 它们相对误差之比的绝对值为

$$\left| \frac{[f(x) - f(x^*)]/f(x)}{(x - x^*)/x} \right| \approx \left| \frac{xf'(x)}{f(x)} \right| = C_p \quad (1.37)$$

$C_p$  称为计算函数值  $f(x)$  的条件数, 如果  $C_p$  很大, 将引起函数值  $f(x^*)$  的相对误差很大, 出现这种情况时, 就认为问题是病态的。例如  $f(x) = x^n$ ,  $f'(x) = nx^{n-1}$ , 则  $C_p = n$ , 它表示相对误差可能放大  $n$  倍, 如  $n = 10$ , 有  $f(1) = 1$ ,  $f(1.02) \approx 1.24$ , 若  $x = 1$ ,  $x^* = 1.02$ , 则自变量相对误差为 2%, 而函数值  $f(1.02)$  的相对误差为 24%, 这时就认为问题是病态的。一般情况下若条件数  $C_p \geq 10$ , 则认为是问题病态,  $C_p$  越大病态越严重。

## 第一章 习题

1. 现代科学研究的基本方法是什么?
2. 什么是数值计算方法? 它与数学科学和计算机的关系如何?

3. 数值计算方法的主要研究对象是什么？
4. 什么是算法？如何判断数值算法的优劣？
5. 误差为什么是不可避免的？用什么标准来衡量近似值是准确的？
6. 科学计算中的误差来源有几种？各举出一个与讲义上不同的例子。计算方法中主要研究哪些误差对计算结果的影响？为什么？
7. 什么是绝对误差与相对误差？什么是近似数的有效数字？它与绝对误差和相对误差有何关系？
8. 什么是数值稳定的算法？如何判断算法稳定？为什么不稳定算法不能使用？
9. 下列各数都是经过四舍五入得到的近似数，试分别指出其绝对误差限、相对误差限和有效数字位数

$$x_1^* = 1.1021 \quad x_2^* = 0.031$$

$$x_3^* = 56.430 \quad x_4^* = 7 \times 10^5$$

10. 求方程  $x^2 + 62x + 1 = 0$  的两个根，使它们具有 4 位有效数字。

11. 计算下列式子，要求具有 4 位有效数字。

$$(1) \sqrt{101.1} - \sqrt{101} \quad (2) 1 - \cos 1^\circ$$

12. 序列  $\left\{\left(\frac{1}{3}\right)^n\right\}$  中由下列两种递推公式生成

$$(1) x_0 = 1, \quad x_n = \frac{1}{3}x_{n-1}, \quad (n=1, 2, \dots)$$

$$(2) y_0 = 1, \quad y_1 = \frac{1}{3}, \quad y_n = \frac{3}{5}y_{n-1} - \frac{4}{9}y_{n-2} \quad (n=2, 3, \dots)$$

采用 5 位有效数字舍入计算，试分别考察递推计算  $\{x_n\}$  与  $\{y_n\}$  是否稳定。

13. 对于  $n = 0, 1, 2, \dots, 20$  计算定积分：

$$y_n = \int_0^1 x^n e^{x-1} dx$$

14. 判断下列命题的正确性：

- (1) 一个问题的病态性如何，与求解它的算法有关系。
- (2) 解对数据的微小变化高度敏感是病态的。
- (3) 高精度运算可以改善问题的病态性。

(4) 用一个稳定的算法计算良态问题一定会得到好的近似值。

(5) 用一个收敛的迭代法计算良态问题一定会得到好的近似值。

(6) 两个相近数相减必然会使有效数字损失。

(7) 计算机上将 1000 个数量级不同的数相加，不管次序如何结果都是一样的。

15. 考虑二次代数方程的求解问题

$$ax^2 + bx + c = 0.$$

求解公式是

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}.$$

与之等价地有

$$x = \frac{2c}{-b \mp \sqrt{b^2 - 4ac}}.$$

对于

$$a = 1, \quad b = -100\,000\,000, \quad c = 1$$

应当如何选择算法？

16. 考虑数列  $x_i, i = 1, \dots, n$ ，它的统计平均值定义为

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

它的标准差

$$\sigma = \left[ \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \right]^{\frac{1}{2}}$$

数学上它等价于

$$\sigma = \left[ \frac{1}{n-1} \left( \sum_{i=1}^n x_i^2 - n\bar{x}^2 \right) \right]^{\frac{1}{2}}$$

作为标准差的两种算法，你如何评价它们的得与失？

17. 已知  $y_n = \int_0^1 \frac{x^n}{4x+1} dx$ ，试建立一个具有数值稳定性的求  $y_n$  ( $n = 1, 2, \dots$ ) 的递推算法。

18. 一个班级第一小组 10 名同学的期末数学考试成绩为：

85    78    91    73    65    55    80    70    82    73

请编程序计算

- (1) 该组的平均成绩和标准差；
- (2) 将成绩从高到低排序。
19. 设  $P(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n$ ，试对给定  $x$  设计导数  $P'(x)$  的求值算法。
20. 设计求平方值小于 1000 的最大整数的算法并编程序计算。
21. 一球从 100 m 高度落下，每次落地后反弹回原高度的一半，再落下。在球第十次落地时，共经过多少路程？第十次下落多高？请设计算法并编程序计算。
22. 找出 100 到 999 之间的整数中所有等于它每位数字立方和的数。
- 例如  $153 = 1^3 + 5^3 + 3^3$
23. 求所以满足条件的四位数 abcd:
- ① 这四位数是 11 的倍数；
  - ② a,b, c,d 是小于 10 的互不相同的自然数；
  - ③  $b+c=a$ ;
  - ④ bc 是完全平方数（例如  $b=2, c=5$ ，则  $bc=25$ ，是完全平方数）。
24. 已知四位数 3025 有一个特殊的性质：它的前两位数字 30 和后两位数字 25 的和是 55，而 55 的平方刚好等于该数（ $55^2=3025$ ）。试编程输出具有这种性质的所有四位数。
25. 编程找出四个互不相同的四位数，它们中任意两数之和为偶数，任意三数之和可以被 3 整除，而且这四个数的和越小越好。（已知它们的和不大于 50）。
26. 求完全数：如果一个数（除去本身）等于它的所有约数的和，这个数就称为完全数。
- 试输出 M 和 N 之间完全数。
27. 孪生数：如果 A 的约数之和等于 B，B 的约数之和等于 A，A 和 B 称为孪生数。
- 试找出 M 和 N 之间的孪生数。
28. 埃及分数：古埃及人有一个非常奇特的习惯，他们喜欢把一个分数表示为若干个分子为 1，分母不同的分数的和的形式。例如： $7/8 = 1/2 + 1/3 + 1/24$ 。
- 试设计用埃及分数算法求分数，并验证。
29. 用 1—9 这 9 个数组成三个三位的平方数，要求每个数字只准使用一次。
30. 某班 30 个人，在开新年联欢会时，全班按学号从 1 排到 30，围成一个圆圈。大家商定从 1 号开始报数，报到 5，出来演个节目，演过节目的同学，不再参加报数。然后接着报数。试编程输出最后演节目的同学的学号。
31. 求守形数： 某数的平方其底位与该数本身相同，则称该数为守形数。

求 2—1000 中的守形数。 例如：25\*25=625，625 的底位 25 相同，称 25 为守形数。

32. 设矩阵

$$A = \begin{pmatrix} 2 & 3 & 1 & 5 & 4 \\ 8 & 4 & 6 & 7 & 1 \\ 3 & 9 & 0 & 2 & 5 \\ 10 & 3 & 5 & 7 & 6 \end{pmatrix} \quad B = \begin{pmatrix} 12 & 3 & 5 \\ 9 & 15 & 5 \\ 3 & 2 & 8 \\ 4 & 9 & 7 \\ 1 & 3 & 8 \end{pmatrix}$$

请编程计算  $C = A \times B$

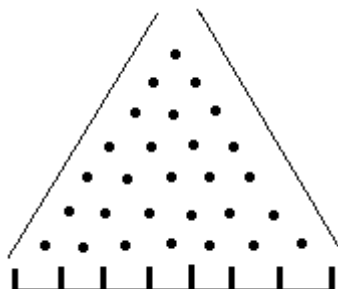
33.  $\pi$  的值可以由下列无穷级数得到：

$$\pi = 4 \sum_{n=0}^{\infty} \frac{(-1)^n}{2n+1} = 4 \left( 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \cdots \right)$$

构造一个算法，求  $\pi$  的近似值， $\varepsilon = \frac{1}{2} \times 10^{-3}$ 。

34. 豆子下落问题：

如图所示一个容器，有 1000 粒豆子从入口一个一个地落下，通过六层隔板落入下面 7 个格子，豆子经过隔板向左右两个方向的机会相等，编一个程序计算每个格子中豆子的数目。



## 第二章 非线性方程求根

### 2.1 引言

我们很熟悉一次、二次代数方程以及某些特殊的高次方程或超越方程的解法。这些方法都是代数解法，也是精确法。但是在实际的应用问题中，常遇到一些高次代数方程或超越方程，例如：

$$\operatorname{tg} x + x = 0$$

$$0.25 + \operatorname{tg} x - 4.8889 \sin x = 0$$

$$x^5 - 4x - 2 = 0$$

以上这些方程叫非线性方程。这些方程看起来很简单，却不容易求得精确解。另一方面，对于一些应用问题，只要能获得具有预先给定的误差限内的近似值就可以了。因此，需要引进能够达到一定精度要求的求方程的根近似值的方法。

设非线性方程

$$f(x)=0 \quad (2.1)$$

其中  $f(x)$  是变量  $x$  的非线性函数。若有  $x^*$  使  $f(x^*)=0$ ，则称  $x^*$  为方程 (2.1) 的根或函数  $f(x)$  的零点。

**定义 2.1** 设  $x^*$  是方程 (2.1) 的根，则  $f(x^*)=0$ 。若存在正整数  $m$ ，使

$$f(x) = (x - x^*)^m g(x) \quad (2.2)$$

且  $0 < |g(x^*)| < +\infty$ ，则称  $x^*$  为 (2.1) 式的  $m$  重根；当  $m=1$  时， $x^*$  为单根；

**定理 2.1** 若  $f(x)$  是连续函数，且  $f(x)$  的  $m$  阶导数连续，则  $x^*$  为 (2.1) 式  $m$  重根的充要条件是： $x^*$  满足

$$f(x^*) = f'(x^*) = f''(x^*) = \cdots = f^{(m-1)}(x^*) = 0 \quad \text{且} \quad f^{(m)}(x^*) \neq 0$$

求方程 (2.1) 根的近似值包括以下三方面内容：

1. 根的存在性。方程有没有根？如果有根，有几个根？
2. 这些根大致在哪里？如何把根隔离开来？
3. 根的精确化

具体求根的工作通常分为两步走：

- (1) 用大范围搜索确定出根的存在区间  $[a, b]$ ，称为根的隔离；



(2) 在有根区间 $[a, b]$ 上用确定的数值方法进行根的精确化计算, 称为近似值的逼近。

**定理 2.2 (代数基本定理)**

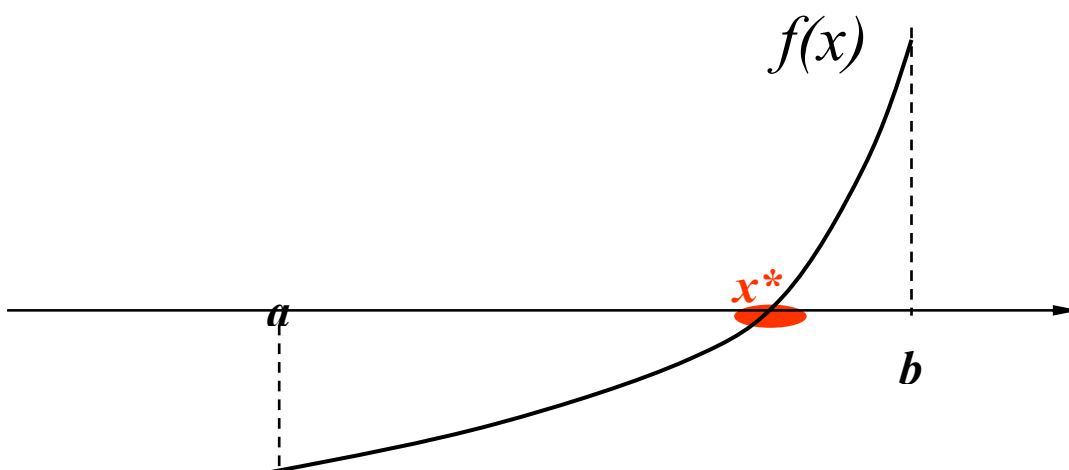
设  $f(x)=0$  为具有复系数  $c_0, c_1, c_2, \dots, c_n$  的代数方程

$$f(x) = c_0 + c_1x + c_2x^2 + \dots + c_nx^n = 0$$

则  $f(x)$  与复平面上恰有  $n$  个根 ( $m$  重根按  $m$  个根算)。若  $f(x)=0$  为具实系数  $a_0, a_1, a_2, \dots, a_n$  的代数方程, 则复根成对出现, 即若  $\alpha + i\beta (\beta \neq 0)$  是  $f(x)=0$  的复根, 则  $\alpha - i\beta (\beta \neq 0)$  也是  $f(x)=0$  的根。

**定理 2.3 (根的存在定理)**

设函数  $f(x)$  在区间  $[a, b]$  上连续, 如果  $f(a) \cdot f(b) < 0$ , 则方程  $f(x) = 0$  在  $[a, b]$  内至少有一实根  $x^*$ 。



**定理 2.3** 是微积分中连续函数介值定理的推论。

由于不知道方程  $f(x)=0$  根的分布情况, 但根据根的存在定理, 根的隔离的实际步骤为:

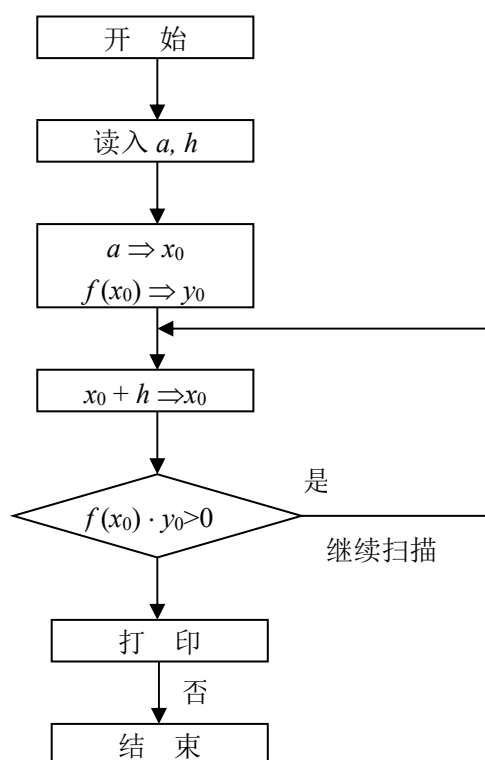
1. 画出  $y = f(x)$  的略图, 从而看出曲线  $y = f(x)$  与  $x$  轴交点的位置。
2. 设左端点  $x_0 = a$ , 计算  $f(x_0)$ ;
3. 确定搜索步长  $h$ ;
4. 计算  $f(x_0 + h)$ ;

5. 判别是否成立 
$$f(x_0) \cdot f(x_0 + h) < 0 \quad (2.3)$$

若不等式 (2.3) 成立, 那么所求的根  $x^*$  一定在  $[x_0, x_0+h]$  内, 这里可取  $x_0$  或  $x_0+h$  作为根的初始

近似。否则  $x_0 \leftarrow x_0 + h$ ，转 4.

以下框图描述了这种逐步扫描法：



**例 2.1：**考察方程

$$f(x) = x^3 - 11x^2 + 38.8x - 41.77 = 0$$

设从  $x = 0$  出发，取  $h = 1$  为步长向右进行根的扫描，列表记录各个结点上函数值的符号，

$x$	0	1	2	3	4	5	6
$f(x)$ 的符号	—	—	+	+	-	-	+

我们发现，在区间  $(1,2)$ ， $(3,4)$ ， $(5,6)$  内有实根。

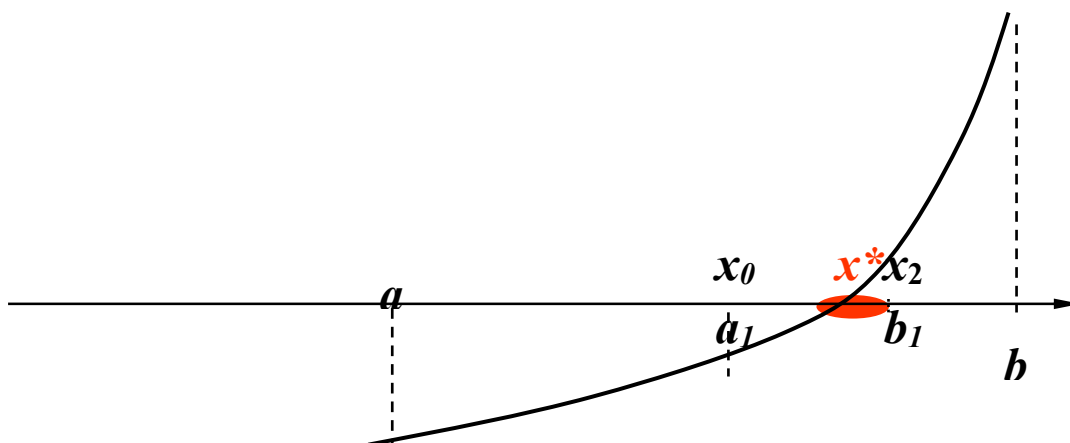
在具体运用上述方法时，步长的选择是个关键。如果步长  $h$  过小，在区间长度大时，会使计算量增大， $h$  过大，又可能出现漏根的现象。因此，这种根的隔离法，只适用于求根的初始近似。

根的逐步精确化的方法，包括二分法、迭代法、牛顿法和弦截法。我们将在以下几节介绍上述方法，将要着重学习迭代法的思想。

## 2.1 二分法

首先, 假定方程  $f(x)=0$  在区间  $[a, b]$  内有唯一的实根  $x^*$ 。

考察有根区间  $[a, b]$ , 取中点  $x_0 = (a+b)/2$  将它分为两半, 如果分点  $f(x_0)=0$ , 则  $x_0$  是根; 如果  $x_0$  不是  $f(x)=0$  的根, 检查  $f(x_0)$  与  $f(a)$  是否同号, 如确系同号, 说明所求的根  $x^*$  在  $x_0$  的右侧, 这时令  $a_1 = x_0, b_1 = b$ ; 否则  $a_1 = a, b_1 = x_0$ , 新的有根区间的长度是原区间长度的一半。



对压缩了的有根区间  $[a_1, b_1]$ , 又可以施行同样的手续, 用中点  $x_1 = (a_1 + b_1)/2$  将  $[a_1, b_1]$  再分两半, 然后通过根的搜索判定所求的根在  $x_1$  的哪一侧, 从而又确定一个新的有根区间  $[a_2, b_2]$ , 其长度是  $[a_1, b_1]$  的一半。

反复执行以上步骤, 便可得到一系列有根区间:

$$(a, b), (a_1, b_1), \dots, (a_k, b_k), \dots$$

其中每个区间都是前一个区间的一半, 因此区间长度为

$$b_k - a_k = \frac{1}{2^k} (b - a)$$

显然, 二分过程如果能够无限地继续下去, 这些区间最终必收敛于一点  $x^*$ , 该点就是所求的根。

不过, 无限过程实际上是不可能实现的, 也没有这种必要, 因为数值分析的结果允许有一定的误差。由于

$$|x_k - x^*| \leq \frac{1}{2} (b_k - a_k) = b_{k+1} - a_{k+1}$$

只要有根区间  $(a_{k+1}, b_{k+1})$  的长度小于预先给定的误差  $\varepsilon$ , 那么就可以取

$$x_{k+1} = \frac{1}{2} (a_k + b_k)$$

作为所求根  $x^*$  的第  $k+1$  次近似值。其误差估计为:

$$|x^* - x_{k+1}| \leq \frac{1}{2^{k+1}}(b-a) \quad (2.4)$$

综上所述，设  $f(x)$  在  $[a, b]$  上存在一阶导数且不变号，如果  $f(a)f(b) < 0$ ，则由 (2.4) 所知，当  $k \rightarrow \infty$  时， $|x^* - x_k| \rightarrow 0$ ，即  $x_k \rightarrow x^*$ ， $|x^* - x_k|$  收敛于零的速度，相当于以  $1/2$  为公比的等比数列。

**例 2.2：** 求方程

$$f(x) = x^3 - x - 1 = 0$$

在区间  $[1, 2]$  内的实根。要求准确到小数点后第 2 位。

解：已知  $f(1) = -1 < 0$ ， $f(2) = 5 > 0$ ，由定理 2.3，方程  $f(x) = x^3 - x - 1 = 0$  在区间  $(1, 2)$  中存在实根。

令  $x_1 = \frac{1+2}{2} = 1.5$ ，计算  $f(1.5) = 0.875 > 0$ ，区间  $(1, 1.5)$  中有根；

令  $x_2 = \frac{1+1.5}{2} = 1.25$ ，计算  $f(1.25) = -0.29688 < 0$ ，区间  $(1.25, 1.5)$  中有根；

令  $x_3 = \frac{1.25+1.5}{2} = 1.375$ ，计算  $f(1.375) = 0.224609 > 0$ ，区间  $(1.25, 1.375)$  中有根；

... ..

如此反复二分下去，我们预先估计一下二分的次数：按误差估计式

$$|x^* - x_k| \leq b_{k+1} - a_{k+1} = \frac{1}{2^{k+1}}(b-a) < 10^{-2}$$

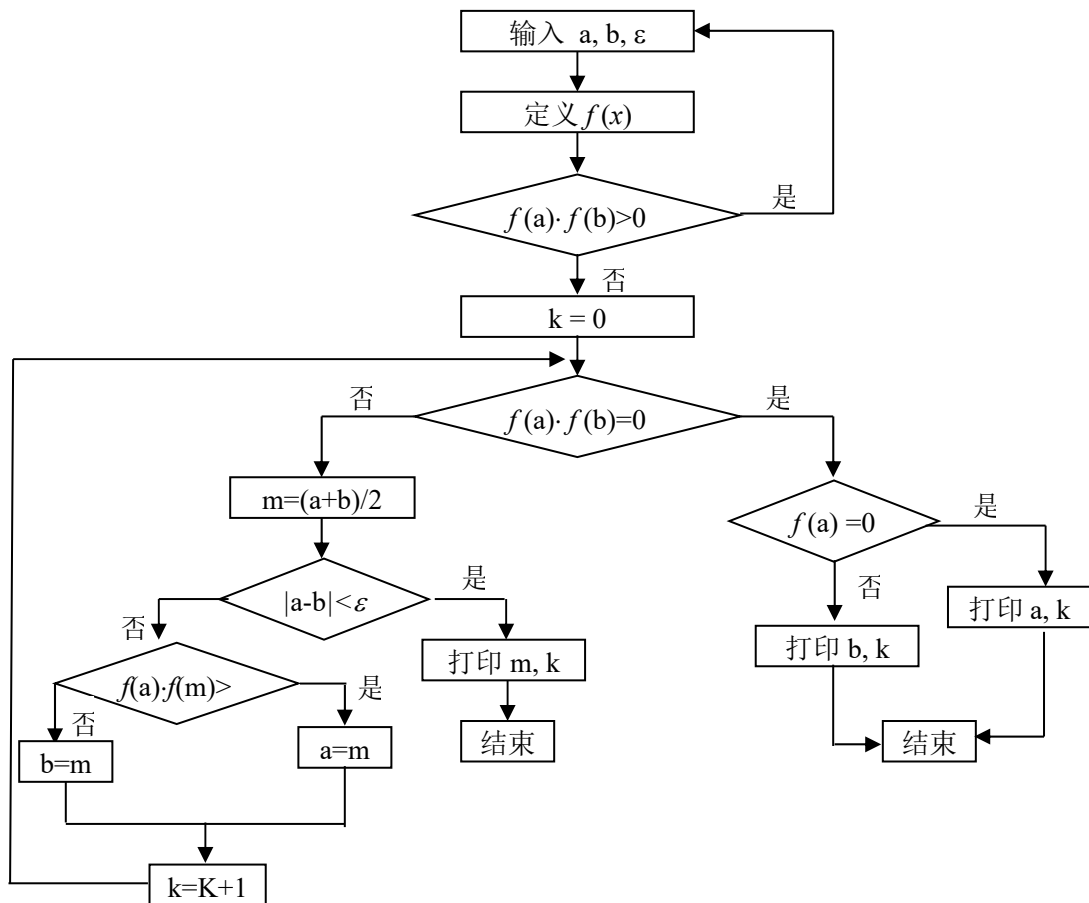
解得  $k = 7$ ，即只要二分 7 次，即达所求精度。计算结果如下表：

$k$	$a_k$	$b_k$	$x_k$	$f(x_k)$ 的符号
0	1	2	1.5	+
1	1	1.5	1.25	-
2	1.25	1.5	1.375	+
3	1.25	1.375	1.3125	-
4	1.3125	1.375	1.3438	+
5	1.3125	1.3438	1.3281	+
6	1.3125	1.3281	1.3203	-
7	1.3203	1.3281	1.3242	-

**算法：**

1. 输入欲求根的函数  $f(x)$ ，区间左右端点  $a, b$ ，误差限  $\varepsilon$ ，最大迭代次数  $N_{\max}$ ；
2. 计算  $f(x)$  在有解区间  $[a, b]$  端点处的值， $f(a)$ ， $f(b)$ 。
3. 计算  $f(x)$  在区间中点  $m = \frac{a+b}{2}$  处的值  $f(m)$ 。
4. 判断若  $f(m) = 0$ ，则  $m = \frac{a+b}{2}$  即是根，否则检验：
  - (1) 若  $f(m)$  与  $f(a)$  异号，则知解位于区间  $[a, m]$ ，以  $m$  代替  $b$ ；
  - (2) 若  $f(m)$  与  $f(a)$  同号，则知解位于区间  $[m, b]$ ， $m$  代替  $a$ 。

下面给出用二分法求  $f(x) = 0$  在  $[a, b]$  上的实根的框图。



其中： $a, b$  为区间端点

$\varepsilon$  为预先给定的误差限

$k$  为对分次数

## 2.2 不动点迭代法及其收敛性

不动点迭代法是求方程近似根的一个重要方法，它的算法简单，应用范围广，也是计算方法中的一种基本方法。

### 2.2.1. 不动点迭代法

将方程  $f(x)=0$  改写为一个等价的方程

$$x = \varphi(x) \quad (2.5)$$

若有  $x^*$  满足  $x^* = \varphi(x^*)$ ，则称  $x^*$  是  $\varphi$  的一个不动点，也是方程 (2.1) 的一个根，如果  $\varphi$  连续，可构造迭代序列

$$x_{k+1} = \varphi(x_k) \quad k = 0, 1, 2, \dots, \quad (2.6)$$

若给定一个初值  $x_0$ ，由 (2.6) 可算得  $x_1 = \varphi(x_0)$ ，再将  $x_1$  代入 (2.6) 的右端又可得  $x_2 = \varphi(x_1)$ ，…。我们称  $\{x_k\}$  为迭代序列，而称 (2.5) 中的  $\varphi(x)$  为迭代函数，(2.6) 为迭代格式。如果  $\varphi(x)$  连续，迭代序列  $\{x_k\}$  收敛于  $x^*$ ，对 (2.6) 式两端取极限，得

$$x^* = \lim_{k \rightarrow \infty} x_{k+1} = \lim_{k \rightarrow \infty} \varphi(x_k) = \varphi\left(\lim_{k \rightarrow \infty} x_k\right) = \varphi(x^*)$$

即  $x^*$  为  $\varphi$  的不动点，或者，也就是说  $f(x^*) = 0$ 。 $x^*$  是  $f(x)$  的根。

如果迭代序列  $\{x_k\}$  收敛，总能收敛于原方程的解。这种求根方法称为不动点迭代法。

(加不动点的图 P-11 图 2.2)

**例 2.3:** 求方程  $f(x) = x - 10^x + 2 = 0$  的一个根

解：因为  $f(0) = 1 > 0$   $f(1) = -7 < 0$ ，由定理 2.3 知方程在  $[0, 1]$  中必有一实根，现将原方程改为同解方程

$$10^x = x + 2 \quad x = \lg(x + 2)$$

由此得迭代格式

$$x_{k+1} = \lg(x_k + 2)$$

取初始值  $x_0 = 1$ ，可逐次算得

$$\begin{aligned} x_1 &= 0.4771, & x_2 &= 0.3939 \\ &\dots \end{aligned}$$

$$x_6 = 0.3758, \quad x_7 = 0.3758$$

因为  $x_6$  和  $x_7$  已趋于一致，所以取  $x_7 = 0.3758$  为原方程在  $[0, 1]$  内的一个根的近似值。

一个方程的迭代格式并不是唯一的，且迭代也不总是收敛的。如例 2.3 的方程也可改写成

$$x = 10^x - 2$$

得迭代格式

$$x_{k+1} = 10^{x_k} - 2$$

仍取  $x_0 = 1$  算得：

$$x_1 = 10 - 2 = 8$$

$$x_2 = 10^8 - 2 \approx 10^8$$

$$x_3 = 10^{10^8} - 2, \dots$$

显然，该迭代序列不趋向于某个定值，这种不收敛的迭代过程称为发散。那么迭代格式要满足哪些条件才能保证迭代收敛呢？下面我们来讨论这个问题。

### 2.2.2. 迭代过程的收敛性

请看下面几个图形

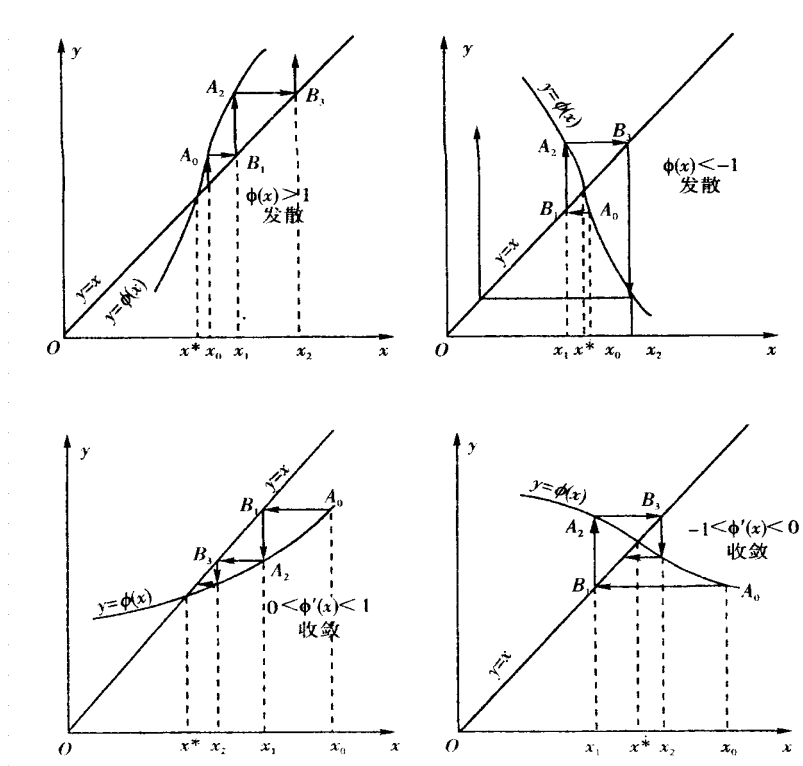


图 2.1

由以上图形可以看出, 序列  $\{x_k\}$  的收敛速度, 取决于曲线  $y = \varphi(x)$  在根附近的斜率  $\varphi'(x)$ , 由拉格朗日定理可知

$$\frac{x_{k+1} - x_k}{x_k - x_{k-1}} = \frac{\varphi(x_k) - \varphi(x_{k-1})}{x_k - x_{k-1}} = \varphi'(\xi_k)$$

其中  $\xi_k$  在  $x_k$  和  $x_{k-1}$  之间。所以在根  $x^*$  附近,  $|\varphi'(x)|$  恒小于 1, 则此迭代序列收敛, 若  $|\varphi'(x)| \geq 1$ , 则此序列发散。由此得**定理 2.4**。

**定理 2.4:** 如果  $\varphi(x)$  满足下列条件

- (1) 当  $x \in [a, b]$  时,  $\varphi(x) \in [a, b]$
- (2) 当任意  $x \in [a, b]$ , 存在  $0 < L < 1$ , 使

$$|\varphi'(x)| \leq L < 1 \quad (2.7)$$

则方程  $x = \varphi(x)$  在  $[a, b]$  上有唯一的根  $x^*$ , 且对任意初值  $x_0 \in [a, b]$  时, 迭代序列  $x_{k+1} = \varphi(x_k)$  ( $k = 0, 1, \dots$ ) 收敛于  $x^*$ 。

证明: 设  $x^*$  是方程  $f(x)=0$  的根, 即  $x^* = \varphi(x^*)$ , 由拉格朗日定理

$$x^* - x_{k+1} = \varphi(x^*) - \varphi(x_k) = \varphi'(\xi)(x^* - x_k)$$

其中  $\xi$  在  $x^*$  与  $x_k$  之间, 由 (2.7)

$$\begin{aligned} |x^* - x_{k+1}| &= |\varphi(x^*) - \varphi(x_k)| = |\varphi'(\xi)| |x^* - x_k| \\ &\leq L |x^* - x_k| \leq L^2 |x^* - x_{k-1}| \leq \dots \\ &\leq L^{k+1} |x^* - x_0| \end{aligned}$$

因为  $0 < L < 1$ , 由  $\lim_{k \rightarrow \infty} L^{k+1} = 0$  知

$$|x^* - x_{k+1}| \rightarrow 0 \quad (k \rightarrow \infty)$$

所以  $\lim_{k \rightarrow \infty} x_{k+1} = x^*$

即  $x_{k+1} = \varphi(x_k)$  收敛

证完

### 2.2.3. 迭代法的结束条件

在求根过程中,  $x^*$  是不知道的, 因此不可能用  $|x^* - x_k| < \varepsilon$  来做为迭代结束的条件。那么, 用什么方法来控制迭代次数呢?

由收敛性定理中的条件 (2.7)



$$\begin{aligned} |x_{k+1} - x_k| &= |\varphi(x_k) - \varphi(x_{k-1})| = |\varphi'(\xi)| |x_k - x_{k-1}| \\ &\leq L |x_k - x_{k-1}| \end{aligned}$$

一般地

$$|x_{k+r} - x_{k+r-1}| \leq L^r |x_k - x_{k-1}|$$

于是, 对于任意正整数  $p$ , 有

$$\begin{aligned} |x_{k+p} - x_k| &= |x_{k+p} - x_{k+p-1} + x_{k+p-1} - x_{k+p-2} + x_{k+p-2} - \cdots + x_k| \\ &\leq |x_{k+p} - x_{k+p-1}| + |x_{k+p-1} - x_{k+p-2}| + \cdots + |x_{k+1} - x_k| \\ &\leq L^p |x_k - x_{k-1}| + L^{p-1} |x_k - x_{k-1}| + \cdots + L |x_k - x_{k-1}| \\ &= (L^p + \cdots + L) |x_k - x_{k-1}| \end{aligned}$$

故定  $n$ , 令  $p \rightarrow \infty$ , 得

$$|x^* - x_k| = \frac{L}{1-L} |x_k - x_{k-1}|$$

只要  $|x_k - x_{k-1}|$  充分小, 就可以保证  $|x^* - x_k|$  足够小。因此可用条件

$$|x_k - x_{k-1}| < \varepsilon$$

来控制迭代过程的结束。

**例 2.4:** 求方程  $x^3 - 3x + 1 = 0$  在  $[0, 0.5]$  内的根, 精确到  $10^{-5}$ 。

解: 将方程变形

$$x = \frac{1}{3}(x^3 + 1) = \varphi(x)$$

因为  $\varphi'(x) = x^2 > 0$ , 在  $[0, 0.5]$  内为增函数, 所以

$$L = \max |\varphi'(x)| = 0.5^2 = 0.25 < 1$$

满足收敛条件, 取  $x_0 = 0.25$ , 用公式 (2.6) 算得

$$x_1 = \varphi(0.25) = 0.3385416$$

$$x_2 = \varphi(x_1) = 0.3462668$$

$$x_3 = \varphi(x_2) = 0.3471725$$

$$x_4 = \varphi(x_3) = 0.3472814$$

$$x_5 = \varphi(x_4) = 0.3472945$$

$$x_6 = \varphi(x_5) = 0.3472961$$

$$x_7 = \varphi(x_6) = 0.3472963$$

取近似根为  $x^* = 0.347296$

#### 2.2.4. 迭代过程的收敛速度

一种迭代格式要具有实用意义，不但需要肯定是收敛的，还要求它收敛得比较快，就是说要有一定的收敛速度。那么，何谓收敛速度呢？下面就来介绍这个概念：

**定义 2.2：** 设由某方法确定的序列  $\{x_k\}$  收敛于方程的根  $x^*$ ，如果存在正实数  $p$ ，使得

$$\lim_{k \rightarrow \infty} \frac{|x^* - x_{k+1}|}{|x^* - x_k|^p} = C \quad (C \text{ 为非零常数})$$

则称序列  $\{x_k\}$  收敛于  $x^*$  的收敛速度是  $p$  阶的，或称该方法具有  $p$  阶敛速。当  $p=1$  时，称该方法为线性（一次）收敛；当  $p=2$  时，称方法为平方（二次）收敛；当  $1 < p < 2$  时，称方法为超线性收敛。

由该定义易见，一个方法的收敛速度实际就是绝对误差的收缩率，敛速的阶  $p$  越大，绝对误差缩减得越快，也就是该方法收敛得越快。

在定理 2.3 中，若  $\varphi'(x)$  连续，且  $\varphi'(x^*) \neq 0$ ，则迭代格式  $x_{k+1} = \varphi(x_k)$  必为线性收敛。因为由

$$|x^* - x_{k+1}| = |\varphi(x^*) - \varphi(x_k)| = |\varphi'(\xi)| |x^* - x_k|$$

$$\lim_{k \rightarrow \infty} \frac{|x^* - x_{k+1}|}{|x^* - x_k|} = \lim_{k \rightarrow \infty} |\varphi'(\xi)| = |\varphi'(x^*)| \neq 0$$

如果  $\varphi'(x^*)=0$ ，则收敛速度就不止是线性的了。

### 2.3 牛顿法

牛顿法是求解方程  $f(x)=0$  的一种重要的迭代法。它是一种将非线性方程  $f(x)=0$  逐步线性化的方法，是解代数方程和超越方程的有效方法之一。

#### 2.3.1 牛顿法的迭代公式

设：已知方程  $f(x)=0$  的一个近似根  $x_0$ ，把  $f(x)$  在  $x_0$  处作泰勒展开，

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{f''(x_0)}{2!}(x - x_0)^2 + \dots$$

若取前两项来近似代替  $f(x)$ （称为  $f(x)$  的线性化），则得近似的线性方程

$$f(x_0) + f'(x_0)(x - x_0) = 0$$

设  $f'(x_0) \neq 0$ ，解之得

$$x = x_0 - \frac{f(x_0)}{f'(x_0)}$$

我们取  $x$  作为原方程  $f(x) = 0$  的近似根  $x_1$ ，即

$$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)}, \quad \text{一般地 } f(x) \neq 0。$$

再重复用上述方法得

$$x_2 = x_1 - \frac{f(x_1)}{f'(x_1)} \quad \dots$$

一般地，有迭代公式

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)} \quad (k = 0, 1, 2, \dots) \quad (2.8)$$

公式 (2.8) 称为求解  $f(x) = 0$  的牛顿迭代公式。

牛顿法有明显的几何意义。方程  $f(x) = 0$  的根  $x^*$  在几何上表示曲线  $y = f(x)$  与  $x$  轴的交点。

当我们求得  $x^*$  的近似值  $x_k$  以后，过曲线  $y = f(x)$  上对应点  $(x_k, f(x_k))$  作  $f(x)$  的切线，其切线方

程为  $y = f(x_k) + f'(x_k)(x - x_k)$

求此切线方程和  $x$  轴的交点，即得  $x^*$  的新的近似值  $x_{k+1}$  必须满足方程

$$f(x_k) + f'(x_k)(x - x_k) = 0$$

这就是牛顿法的迭代公式  $x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}$  的计算结果。

由图 2.2 可知，只要初值取得合适，点列  $\{x_k\}$  就会很快收敛于  $x^*$ 。

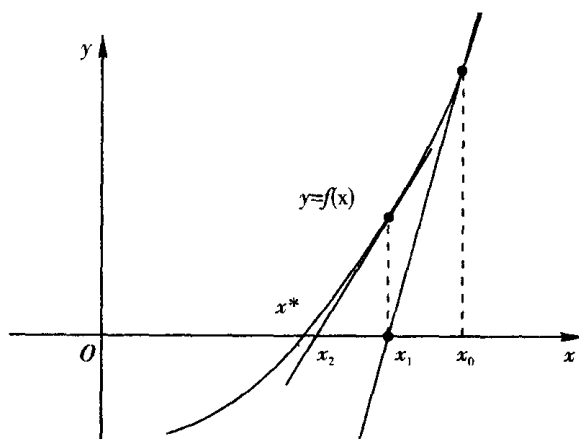


图 2.2

正因为牛顿法有这一明显的几何意义，所以牛顿法也称为切线法。

### 2.3.2 牛顿法的收敛性

事实上，牛顿迭代法的迭代函数是

$$\varphi(x) = x - \frac{f(x)}{f'(x)}$$

由于

$$|\varphi'(x)| = \frac{|f''(x)|}{[f'(x)]^2} |f(x)|$$

如果在方程  $f(x) = 0$  的根  $x^*$  的某个邻域  $R: |x^* - x| \leq \delta$  内， $f'(x) \neq 0$ ， $f''(x)$  连续从而有界，这时

只要  $|f(x)|$  充分小，亦即初值  $x_0$  足够靠近  $x^*$ ，就能使  $|\varphi'(x)| \leq L < 1$ ，牛顿迭代法收敛于  $x^*$ 。

可见牛顿迭代法对初值的要求较高：要求初值  $x_0$  充分接近  $x^*$  才能保证局部收敛性。若要保证较大范围  $[a, b]$  内的收敛性，还要增加一些条件。下面我们给出这方面的一个充分条件。

**定理 2.5：** 设  $f(x)$  在  $[a, b]$  上满足下列条件：

- (1)  $f(a) \cdot f(b) < 0$ ;
- (2)  $f'(x) \neq 0$ ;
- (3)  $f''(x)$  存在且不变号;
- (4) 取  $x_0 \in [a, b]$ ，使得  $f''(x)f(x_0) > 0$

则由 (2.8) 确定的牛顿迭代序列  $\{x_k\}$  收敛于  $f(x)$  在  $[a, b]$  上的唯一根  $x^*$ 。

事实上，条件 (1) 保证根的存在；(2) 表示  $f(x)$  单调，所以根唯一；(3) 保证曲线的凹向不变；(4) 保证  $x \in [a, b]$ ， $\varphi(x) \in [a, b]$ 。图 2.3 的四种情况都满足定理 2.5 的条件。

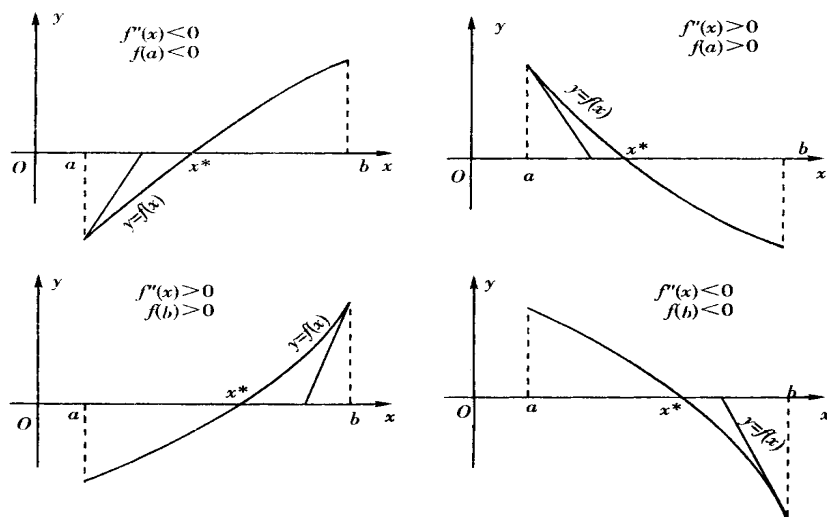


图 2.3

牛顿迭代法的主要优点是收敛速度较快，它是平方收敛的。事实上，若  $f(x)$  连续，将  $f(x)$  在  $x_k$  处按泰勒展开，并将  $x^*$  代替  $x$  可得

$$f(x^*) = f(x_k) + f'(x_k)(x^* - x_k) + \frac{f''(\xi)}{2!}(x^* - x_k)^2 = 0$$

即

$$f(x_k) + f'(x_k)(x^* - x_k) = -\frac{1}{2}f''(\xi)(x^* - x_k)^2$$

用  $f'(x_k)$  除等式两端得

$$x^* - x_k + \frac{f(x_k)}{f'(x_k)} = -\frac{1}{2} \frac{f''(\xi)}{f'(x_k)} (x^* - x_k)^2$$

即

$$x^* - x_{k+1} = -\frac{1}{2} \frac{f''(\xi)}{f'(x_k)} (x^* - x_k)^2$$

$$\frac{|x^* - x_{k+1}|}{|x^* - x_k|^2} = \left| \frac{f''(\xi)}{2f'(x_k)} \right| \rightarrow \left| \frac{f''(x^*)}{f'(x^*)} \right| \quad (k \rightarrow \infty)$$

至于牛顿法的误差估计，仍可以按 § 2 中所述的方法处理。

**例 2.5** 用牛顿迭代法建立求平方根  $\sqrt{c}$  ( $c > 0$ ) 的迭代公式，用该公式求  $\sqrt{0.78265}$

解：(1) 设  $f(x) = x^2 - c$ , ( $x > 0$ ) 则  $c$  就是  $f(x) = 0$  的正根。因为  $f'(x) = 2x$ ，所以由 (2.5) 得迭代公式

$$x_{k+1} = x_k - \frac{x_k^2 - c}{2x_k} \quad \text{或} \quad x_{k+1} = \frac{1}{2} \left( x_k + \frac{c}{x_k} \right) \quad (2.9)$$

由于  $x > 0$  时， $f'(x) > 0$ ，且  $f''(x) > 0$ ，根据定理 2.5 知：取任意初值  $x_0 > \sqrt{c}$ ，(2.5) 所确定的

迭代序列  $\{x_k\}$  必收敛于  $\sqrt{c}$ 。

(2) 求  $\sqrt{0.78265}$

取初值  $x = 0.88$ ，利用公式 (2.6) 的计算结果见表

$k$	$x_k$
0	0.88
1	0.88469
2	0.88468
3	0.88468

故可取  $\sqrt{0.78265} \approx 0.88468$

### 2.3.4. 牛顿下山法

由牛顿法的收敛性定理知，牛顿法对初始值  $x_0$  的选取要求是很高的，为保证收敛，常要利用条件  $f(x_0) \cdot f''(x_0) > 0$  来选取初值  $x_0$ ，但对有些问题，往往很难检验满足条件的初  $x_0$ ，这时我们可利用所谓下山法来扩大初值的选取范围，将牛顿的迭代公式 (2.5) 修改为

$$x_{k+1} = x_k - \lambda \frac{f(x_k)}{f'(x_k)} \quad (k = 0, 1, 2, \dots) \quad (2.10)$$

其中  $\lambda$  是一个参数， $\lambda$  的选取应使

$$|f(x_{k+1})| < |f(x_k)| \quad \text{成立} \quad (2.11)$$

当  $|f(x_{k+1})| < \varepsilon_1$  或  $|x_{k+1} - x_k| < \varepsilon_2$  时 (其中  $\varepsilon_1, \varepsilon_2$  为事先给定的精度，称  $\varepsilon_1$  为残量精确度， $\varepsilon_2$  为根的误差限)，就停止迭代，且取  $x^* \approx x_{k+1}$ ，否则再减小  $\lambda$ ，继续迭代。

按上述迭代过程计算，实际上得到一个以零为下界的严格单调下降的函数值序列  $\{f(x_k)\}$ 。这个方法就称为牛顿下山法。 $\lambda$  称为下山因子，要求满足  $0 < \varepsilon_\lambda \leq \lambda \leq 1$ ， $\varepsilon_\lambda$  为下山因子下界，为了方便，一般开始时可简单地取  $\lambda = 1$ ，然后逐步分半减少，即可选取  $\lambda = 1, \frac{1}{2}, \frac{1}{2^2}, \dots, \lambda \geq \varepsilon_\lambda$ ，且使  $|f(x_{k+1})| < |f(x_k)|$

牛顿下山法计算步骤可归纳如下：

- (1) 选取初始近似值  $x_0$ ；
- (2) 取下山因子  $\lambda = 1$ ；

$$(3) \text{ 计算 } x_{k+1} = x_k - \lambda \frac{f(x_k)}{f'(x_k)}$$

(4) 计算  $f(x_{k+1})$ , 并比较  $|f(x_{k+1})|$  与  $|f(x_k)|$  的大小, 分以下二种情况:

1) 若  $|f(x_{k+1})| < |f(x_k)|$ , 则当  $|x_{k+1} - x_k| < \varepsilon_2$  时, 取  $x^* \approx x_{k+1}$ , 计算过程结束; 当  $|x_{k+1} - x_k| \geq \varepsilon_2$  时, 则把  $x_{k+1}$  作为新的  $x_k$  值, 并重复回到 (3)。

2) 若  $|f(x_{k+1})| \geq |f(x_k)|$ , 则当  $\lambda \leq \varepsilon_\lambda$  且  $|f(x_{k+1})| < \varepsilon_1$ , 取  $x^* \approx x_k$ , 计算过程结束; 否则若  $\lambda \leq \varepsilon_\lambda$ , 而  $|f(x_{k+1})| \geq \varepsilon_1$  时, 则把  $x_{k+1}$  加上一个适当选定的小正数, 即取  $x_{k+1} + \delta$  作为新的  $x_k$  值, 并转向 (3) 重复计算; 当  $\lambda > \varepsilon_\lambda$ ; 且  $|f(x_{k+1})| \geq \varepsilon_1$ , 则将下山因子缩小一半, 取  $\lambda/2$  代入, 并转向 (3) 重复计算。

牛顿下山法不但放宽了初值  $x_0$  的选取, 且有时对某一初值, 虽然用牛顿法不收敛, 但用牛顿下山法却可能收敛。

**例 2.6** 已知方程  $f(x) = x^3 - x - 1 = 0$  的一个根为  $x^* = 1.32472$ , 若取初值  $x_0 = 0.6$ , 用牛顿法  $x_1 = x_0 - \frac{f(x_0)}{f'(x_0)} = 17.9$ , 反而比  $x_0 = 0.6$  更偏离根  $x^*$ 。若改用牛顿下山法

$$x_{k+1} = x_k - \lambda \frac{f(x_k)}{f'(x_k)} \quad (k = 0, 1, 2, \dots)$$

计算, 仍取  $x_0 = 0.6$  计算结果如下

$k$	$\lambda$	$x_k$
0	1	0.6
1	$1/2^5$	1.14063
2	1	1.36681
3	1	1.32628
4	1	1.32472

由此可见, 牛顿下山法使迭代过程收敛加速。

## 2.4 弦截法

牛顿迭代法虽然有较高的收敛速度, 但要计算导数值  $f'(x_k)$ , 这对复杂的函数  $f(x)$  是不方便的, 因此构造即有较高的收敛速度, 又不含  $f(x)$  的导数的迭代公式是十分必要的。

### 2.4.1 单点弦截法

为避免导数的计算，用平均变化率

$$\frac{f(x_k) - f(x_0)}{x_k - x_0}$$

来替代迭代公式 (2.5) 中的导数  $f'(x_k)$ ，于是得

$$x_{k+1} = x_k - \frac{f(x_k)}{f(x_k) - f(x_0)}(x_k - x_0) \quad (2.12)$$

按公式 (2.12) 进行迭代计算就称单点弦截法。单点弦截法的几何意义如图 2.4 所示。

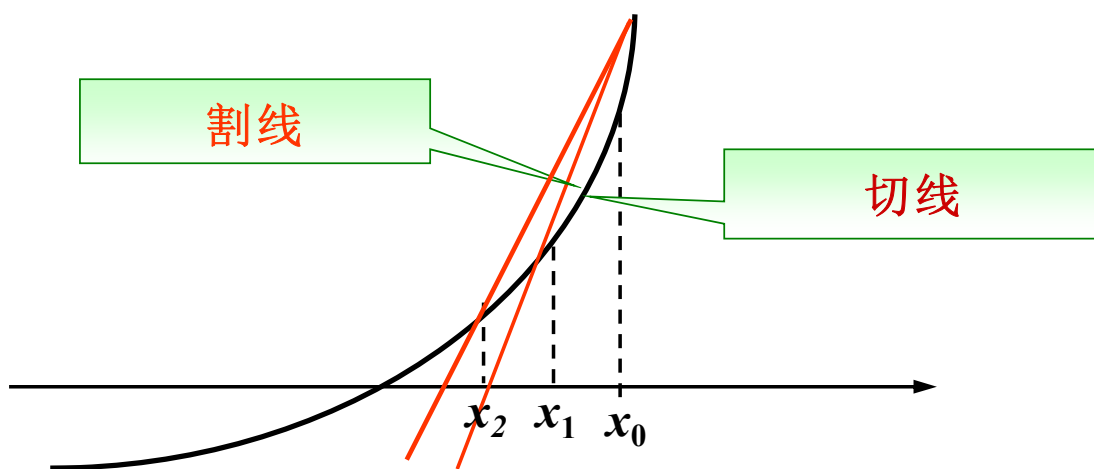


图 2.4

按公式 (2.12) 求得的  $x_{k+1}$  实际上是弦 AB 与  $x$  轴交点的横坐标，下一步再以点  $(x_{k+1}, f(x_{k+1}))$  和  $(x_0, f(x_0))$  作弦交  $x$  轴得  $x_{k+2}$  等等。每次作新的弦都以  $(x_0, f(x_0))$  作为一个端点，只有一个端点不断更换，故名为单点弦截法。

由 (2.12) 式易知其迭代函数是：

$$\varphi(x) = x - \frac{f(x)}{f(x) - f(x_0)}(x - x_0)$$

因为  $f(x^*) = 0$ ，故求导数得

$$\varphi'(x) = 1 + \frac{f'(x^*)}{f(x_0)}(x^* - x_0) = 1 - \frac{f'(x^*)}{\frac{f(x^*) - f(x_0)}{x^* - x_0}}$$

当初值  $x_0$  充分接近  $x^*$  时， $\frac{f(x^*) - f(x_0)}{x^* - x_0}$  很接近  $f'(x^*)$ ，且符号也相同，所以  $0 < |\varphi'(x^*)| < 1$ ，

所以 (2.12) 仅为线性收敛的。



### 2.4.2 双点弦截法

如果 (2.8) 中的导数  $f'(x_k)$  改用  $\frac{f(x_k) - f(x_{k-1})}{x_k - x_{k-1}}$  来代替, 就可以得到迭代公式:

$$x_{k+1} = x_k - \frac{f(x_k)}{f(x_k) - f(x_{k-1})}(x_k - x_{k-1}) \quad (2.13)$$

由公式 (2.13) 确定的迭代法称双点弦截法。

双点弦截法的几何意义如图 2.5 所示。它是用弦 AB 与  $x$  轴交点的横坐标  $x_{k+1}$  代替曲线  $y = f(x)$  与  $x$  轴交点的横坐标  $x^*$  的近似值。

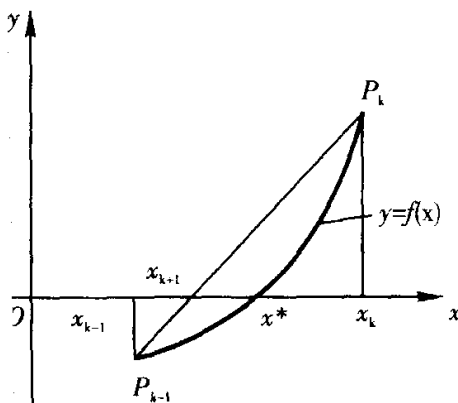


图 2.5

双点弦截法的收敛性与牛顿迭代法一样, 即在根的某个邻域内,  $f(x)$  有直至二阶的连续导数, 且  $f'(x) \neq 0$ , 具有局部收敛性, 同时在邻域任取初值  $x_0, x_1$  迭代均收敛。可以证明, 双点弦截法具有超线性敛速, 收敛的阶  $\frac{1}{2}(1 + \sqrt{5}) \approx 1.618$ 。

**例 2.7** 用双点弦截法求方程  $x^3 - x - 1 = 0$  在  $x = 1.5$  附近的根, 使绝对误差精确到  $10^{-4}$ 。

解: 取初值  $x_0 = 1.5, x_1 = 1.4$ , 按公式 (2.13) 得迭代格式

$$x_{k+1} = x_k - \frac{(x_k^3 - x_k - 1)(x_k - x_{k-1})}{(x_k^3 - x_{k-1}^3 - 1)(x_k - x_{k-1})}$$

按上式计算得:

$$x_2 = 1.33522 \quad x_3 = 1.32541$$

$$x_4 = 1.32472 \quad x_5 = 1.32472$$

$$\text{取 } x^* \approx 1.3247$$

## 习题

1. 判断如下命题是否正确：

- (a) Newton 法的收敛阶高于简化 Newton 法；
- (b) 任何方法的收敛阶都不可能高于 Newton 法；
- (c) Newton 法总是比简化 Newton 法更节省计算时间；
- (d) 如果函数的导数难于计算，则应当考虑选择简化 Newton 法；
- (e) Newton 法有可能不收敛；
- (f) 考虑迭代法  $x_{k+1} = g(x_k)$ ，其中  $x^* = g(x^*)$ 。如果  $|g'(x^*)| < 1$ ，则对任意的初始值，上述迭代都收敛。

2. 设方程  $f(x) = x^4 - 3x + 1 = 0$ ，用二分法求方程在  $[0.3, 0.4]$  内的一个实根，使精确到小数点后 5 位。

3. 分析方程  $f(x) = \sin x - \frac{x}{2} = 0$  正根的分布情况，并用二分法求正根的近似值，使误差不超过  $10^{-2}$ 。

4. 估计用二分法求方程  $f(x) = x^3 + 4x^2 - 10 = 0$  在区间  $[1, 2]$  内根的近似值，为使误差不超过  $10^{-5}$  时所需要的对分区间的次数。

5. 证明方程  $1 - x - \sin x = 0$  在  $[0, 1]$  中有一个根，使用二分法求误差不大于  $\frac{1}{2} \times 10^{-4}$  的根要迭代多少次？

6.  $x^2 - 3x + 2 = 0$ ，它的根显然为  $x = 1, x = 2$ 。考虑迭代方法

$$x_{k+1} = \frac{1}{3}(x_k^2 + 2), \quad k = 1, 2, \dots$$

- (1) 确定该方法满足迭代法收敛性充分条件的区间；
- (2) 确定该迭代法的收敛阶；
- (3) 设法找出最大的区间，使其中任何值作为  $x_0$ ，迭代都收敛到  $x^* = 1$ ；
- (4) 试将实际的数值计算结果与如下的理论估计比较。

$$|x_k - x^*| \leq \frac{1}{1-L} |x_{k+1} - x_k|$$

- (5) 能否找到一个区间，使从其中任何值  $x_0$  出发，迭代都收敛到  $x^* = 2$ ，为什么？

7. 仍考虑方程  $x^2 - 3x + 2 = 0$ ，用迭代方法

$$x_{k+1} = x_k - a \frac{x_k^2 - 3x_k + 2}{2x_k - 3}, \quad k = 1, 2, \dots$$

分别求出使该迭代法在  $x^* = 2$  和  $x^* = 1$  有局部收敛性的  $a$  范围。

8. 设有方程  $f(x) = 0$ , 其中  $f'(x)$  存在, 且对一切满足  $0 < m \leq f'(x) \leq M$ , 构造迭代过程

$$x_{k+1} = x_k - \lambda f(x_k), \quad (k = 0, 1, \dots)$$

试证明当  $\lambda$  满足  $0 < \lambda < \frac{2}{M}$  时, 对任取初值  $x_0$ , 上述迭代过程收敛。

9. 为求方程  $f(x) = x^3 - x^2 - 1 = 0$  在  $x_0 = 1.5$  附近的一个根, 可将方程改写成下列等价形式, 并建立相应的迭代公式

$$(1) \quad x = 1 + \frac{1}{x^2}, \quad \text{迭代公式为 } x_{k+1} = 1 + \frac{1}{x_k^2};$$

$$(2) \quad x = (1 + x^2)^{1/3}, \quad \text{迭代公式为 } x_{k+1} = (1 + x_k^2)^{1/3};$$

$$(3) \quad x = \left( \frac{1}{x-1} \right)^{1/2}, \quad \text{迭代公式为 } x_{k+1} = \frac{1}{\sqrt{x_k - 1}}$$

试分析每一种迭代公式的敛散性。

10. 求  $x^3 + 4x^2 - 10 = 0$  在  $[1, 2]$  内的根, 精确到  $10^{-8}$ . 试用下列三种迭代格式:

$$(a) \quad x_{k+1} = \frac{1}{2}(10 - x_k^3)^{1/2};$$

$$(b) \quad x_{k+1} = \left( \frac{10}{4 + x_k} \right)^{1/2};$$

$$(c) \quad x_{k+1} = x_k - \frac{x_k^3 + 4x_k^2 - 10}{3x_k^2 + 8x_k}$$

都取  $x_0 = 1.5$ , 比较迭代次数, 分析为什么?

11. 比较求  $e^x + 10x - 2 = 0$  在区间  $[0, 1]$  的根到三位小数所需的计算量:

- (1) 在区间  $[0, 1]$  内用二分法;
- (2) 构造一个收敛的迭代格式, 用不动点方法;
- (3) 选取合适的初值, 用牛顿迭代法。

12. 设  $f(x) = 0$  的三种等价形式的  $x = g(x)$  如下

$$(1) \quad g_1(x) = (4 + 2x^3)/(x^2) - 2x$$

$$(2) \quad g_2(x) = \sqrt{4/x}$$

$$(3) \quad g_3(x) = (16 + x^3)/(5x^2)$$

(a) 求  $f(x)$ ;

(b) 那种迭代格式收敛? 求出收敛区间。

13. 已知方程  $e^x - 4x = 0$ , 试求:

(1) 方程的有根区间;

(2) 在有根区间上构造不动点迭代公式。

14. 解方程  $12 - 3x + 2\cos x = 0$  的迭代公式为  $x_{n+1} = 4 + \frac{2}{3}\cos x_n$

(1) 证明: 对任意实数  $x_0$ , 均有  $\lim_{n \rightarrow \infty} x_n = x^*$  ( $x^*$  为方程的根)

(2) 次迭代法的收敛阶是多少?

15. 能否用简单迭代法求解下列方程:

$$(1) \quad x = \varphi_1(x) = \frac{1}{4}(\cos x + \sin x);$$

$$(2) \quad x = \varphi_2(x) = 4 - 2^x,$$

若不能, 试将原方程改成能用迭代法求解的形式。

16. 欲求方程  $x - \ln x - 3 = 0$  在区间  $[3, 5]$  上的根, 有下列迭代法

$$(1) \quad x_{k+1} = 3 + \ln x_k; \quad (2) \quad x_{k+1} = e^{x_k - 3}$$

初始近似值  $x_0 = 3$ 。试分析每一种迭代法的收敛性。对于收敛的迭代法, 迭代多少步才能

使  $|x_k - x^*| \leq \varepsilon$  ?

17. 求  $f(x) = x - \cos x = 0$  在  $x_0 = 1$  附近的实根, 要求满足精度

$$|x_{k+1} - x_k| < 0.001$$

18. 证明方程  $x = 1 + \arctg x$  有根  $x^*$ , 找出一个包含  $x^*$  的区间  $[a, b]$ , 使迭代公式

$$\begin{cases} x_{n+1} = 1 + \arctg x_n \\ \text{任取 } x_0 \in [a, b] \end{cases} \quad n = 0, 1, \dots \quad \text{收敛}$$

19. 选择函数  $g(x)$ , 使迭代公式

$$\begin{cases} x_{n+1} = g(x_n) \\ x_0 \in [2, 4] \end{cases}$$

产生的序列  $\{x_n\}$  收敛于方程  $x - e^{x-2} = 0$  在区间  $[2, 4]$  内的根, 并说明收敛理由

20. 不用开方和除法运算, 求  $\frac{1}{\sqrt[3]{a}}$  ( $a > 1$ ) 的计算公式。(提示: 用牛顿迭代法)

21. 已知  $x = \varphi(x)$  在区间  $[a, b]$  内只有一根, 而当  $a < x < b$  时

$$|\varphi'(x)| \geq k > 1$$

试问如何将  $x = \varphi(x)$  化为迭代能收敛的等价形式。

22. 分别用二分法和牛顿法求  $x = \operatorname{tg} x$  的最小正根。

23. 研究求  $\sqrt{a}$  的牛顿公式

$$x_{k+1} = \frac{1}{2} \left( x_k + \frac{a}{x_k} \right), \quad x_0 > 0$$

证明对一切  $k = 1, 2, \dots$ ,  $x_k \geq \sqrt{a}$  且序列  $x_1, x_2, \dots$  是单调递减的。

24. 什么叫做一个迭代法是二阶收敛的? Newton 法收敛时, 它的收敛阶是否总是二阶的?

25. 求解单变量非线性方程的单根, 下面的 3 种方法, 它们的收敛阶由高到低次序如何?

(a) 二分法

(b) Newton 方法

(c) 割线方法

26. 考虑函数  $f(x) = \frac{x^2}{1+x^2}$ , 对怎样的初始值它的 Newton 迭代收敛。

27. 用割线法求方程  $f(x) = x^3 - 3x^2 - x + 9 = 0$  在区间  $[-2, -1]$  内的一个实根近似值  $x_k$ ,

使  $|f(x_k)| \leq 10^{-5}$ 。

28. 用割线法求方程  $x^3 - 3x - 1 = 0$  在  $x_0 = 2$  附近的实根, 设取  $x_0 = 1.9$ , 计算到四位有效数字为止。

29. 利用适当的迭代法证明

$$\lim_{k \rightarrow \infty} \sqrt{2 + \sqrt{2 + \sqrt{2 + \cdots \sqrt{2}}}} = 2$$

30. 对于给定正数  $C$ , 应用牛顿法求二次方程  $x^2 - C = 0$  的算术根. 证明这种迭代对于任意初值  $x_0 > C$  都是收敛的。

31. 证明迭代公式

$$x_{k+1} = \frac{x_k(x_k^2 + 3a)}{3x_k^2 + a}$$

是计算  $\sqrt{a}$  的三阶方法.

#### 上机计算实习:

1. 编写方程求根的二分法标准程序. 并求方程

$$e^x + 2^x + 2 \cos x - 6 = 0$$

在区间[1, 2]内的根, 精确到  $10^{-5}$ .

2. 编写方程求根的牛顿迭代法标准程序, 并求题 12 三种迭代格式满足同样条件的根。比较收敛的速度.

## 第三章 解线性方程组的直接法

### 3.1 引言

许多科学技术问题要归结为解含有多个未知量  $x_1, x_2, \dots, x_n$  的线性方程组。例如，用最小二乘法求实验数据的曲线拟合问题，三次样条函数问题，解非线性方程组的问题，用差分法或有限元法解常微分方程、偏微分方程的边值等，最后都归结为求解线性代数方程组。关于线性方程组的数值解法一般有两类：直接法和迭代法。

#### 1. 直接法

直接法就是经过有限步算术运算，可求得线性方程组精确解的方法（假设计算过程中没有舍入误差）。但实际计算中由于舍入误差的存在和影响，这种方法也只能求得线性方程组的近似解。本章将阐述这类算法中最基本的高斯消去法及其某些变形。

#### 2. 迭代法

迭代法就是用某种极限过程去逐步逼近线性方程组精确解的方法，迭代法需要的计算机存储单元少、程序设计简单、原始系数矩阵在计算过程中不变，这些都是迭代法的优点；但是存在收敛性和收敛速度的问题。迭代法适用于解大型的稀疏矩阵方程组。

为了讨论线性方程组的数值解法，需要复习一些基本的矩阵代数知识。

#### 3.1.1 向量和矩阵

用  $\mathbf{R}^{m \times n}$  表示全部  $m \times n$  实矩阵的向量空间， $\mathbf{C}^{m \times n}$  表示全部  $m \times n$  复矩阵的向量空间。

$$\mathbf{A} \in \mathbf{R}^{m \times n} \Leftrightarrow \mathbf{A} = (a_{ij}) = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}$$

此实数排成的矩形表，称为  $m$  行  $n$  列矩阵。

$$\mathbf{x} \in \mathbf{R}^n \Leftrightarrow \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \quad \mathbf{x} \text{ 称为 } n \text{ 维列向量}$$

矩阵  $\mathbf{A}$  也可以写成

$$\mathbf{A} = (\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n)$$

其中  $\mathbf{a}_i$  为  $\mathbf{A}$  的第  $i$  列。同理

$$\mathbf{A} = \begin{pmatrix} \mathbf{b}_1^T \\ \mathbf{b}_2^T \\ \vdots \\ \mathbf{b}_n^T \end{pmatrix}$$

其中  $\mathbf{b}_i^T$  为  $\mathbf{A}$  的第  $i$  行。

**矩阵的基本运算：**

(1) 矩阵加法  $\mathbf{C} = \mathbf{A} + \mathbf{B}$ ,  $c_{ij} = a_{ij} + b_{ij}$  ( $\mathbf{A} \in \mathbf{R}^{m \times n}, \mathbf{B} \in \mathbf{R}^{m \times n}, \mathbf{C} \in \mathbf{R}^{m \times n}$ ).

(2) 矩阵与标量的乘法  $\mathbf{C} = \alpha \mathbf{A}$ ,  $c_{ij} = \alpha a_{ij}$

(3) 矩阵与矩阵乘法  $\mathbf{C} = \mathbf{AB}$ ,  $c_{ij} = \sum_{k=1}^n a_{ik} b_{kj}$  ( $\mathbf{A} \in \mathbf{R}^{m \times n}, \mathbf{B} \in \mathbf{R}^{n \times p}, \mathbf{C} \in \mathbf{R}^{m \times p}$ )

(4) 转置矩阵  $\mathbf{A} \in \mathbf{R}^{m \times n}, \mathbf{C} = \mathbf{A}^T$ ,  $c_{ij} = a_{ji}$

(5) 单位矩阵  $\mathbf{I} = (\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n) \in \mathbf{R}^{n \times n}$ , 其中

$$\mathbf{e}_k = (0, \dots, 0, 1, 0, \dots, 0)^T \quad k=1, 2, \dots, n$$

(6) 非奇异矩阵 设  $\mathbf{A} \in \mathbf{R}^{n \times n}$ ,  $\mathbf{B} \in \mathbf{R}^{n \times n}$ 。如果  $\mathbf{AB} = \mathbf{BA} = \mathbf{I}$ , 则称  $\mathbf{B}$  是  $\mathbf{A}$  的逆矩阵, 记为

$\mathbf{A}^{-1}$ , 且  $(\mathbf{A}^{-1})^T = (\mathbf{A}^T)^{-1}$ 。如果  $\mathbf{A}^{-1}$  存在, 则称  $\mathbf{A}$  为非奇异矩阵。如果  $\mathbf{A}, \mathbf{B} \in \mathbf{R}^{n \times n}$  均

为非奇异矩阵, 则  $(\mathbf{AB})^{-1} = \mathbf{B}^{-1} \mathbf{A}^{-1}$ 。

(7) 矩阵的行列式  $\mathbf{A} \in \mathbf{R}^{n \times n}$ , 则  $\mathbf{A}$  的行列式可按任一行 (列) 展开。即

$$\det(\mathbf{A}) = \sum_{j=1}^n a_{ij} A_{ij}$$

其中  $A_{ij}$  为  $a_{ij}$  的代数余子式,  $A_{ij} = (-1)^{i+j} M_{ij}$ ,  $M_{ij}$  为元素  $a_{ij}$  的余子式。

行列式的性质:

①  $\det(\mathbf{AB}) = \det(\mathbf{A}) \det(\mathbf{B})$ ,  $\mathbf{A}, \mathbf{B} \in \mathbf{R}^{n \times n}$ 。

②  $\det(\mathbf{A}^T) = \det(\mathbf{A})$ ,  $\mathbf{A} \in \mathbf{R}^{n \times n}$

③  $\det(c\mathbf{A}) = c^n \det(\mathbf{A})$ ,  $\mathbf{A} \in \mathbf{R}^{n \times n}$

④  $\det(\mathbf{A}) \neq 0 \Leftrightarrow \mathbf{A}$  是非奇异矩阵。



### 3.1.2 矩阵的特征值与谱半径

**定义 3.1**  $\mathbf{A} = (a_{ij}) \in \mathbf{R}^{n \times n}$  , 若存在数  $\lambda$  (实数或复数) 和非零向量  $x = (x_1, x_2, \dots, x_n)^T \in \mathbf{R}^n$  , 使

$$\mathbf{A}x = \lambda x$$

则称  $\lambda$  为  $\mathbf{A}$  的**特征值**,  $x$  为  $\mathbf{A}$  对应  $\lambda$  的**特征向量**,  $\mathbf{A}$  的全体特征值称为  $\mathbf{A}$  的**谱**, 记作  $\sigma(\mathbf{A})$  , 即

$$\sigma(\mathbf{A}) = \{\lambda_1, \lambda_2, \dots, \lambda_n\}, \text{ 记}$$

$$\rho(\mathbf{A}) = \max_{1 \leq i \leq n} |\lambda_i|$$

称为矩阵  $\mathbf{A}$  的**谱半径**。

由特征值的定义可知  $\lambda$  可是齐次线性方程组

$$(\lambda \mathbf{I} - \mathbf{A})x = \mathbf{0}$$

有非零解, 故系数行列式  $\det(\lambda \mathbf{I} - \mathbf{A})x = \mathbf{0}$  , 记

$$p(\lambda) = \det(\lambda \mathbf{I} - \mathbf{A}) = \begin{vmatrix} \lambda - a_{11} & -a_{12} & \cdots & -a_{1n} \\ -a_{21} & \lambda - a_{22} & \cdots & -a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ -a_{n1} & -a_{n2} & \cdots & \lambda - a_{nn} \end{vmatrix} \quad (3.0)$$

$$= \lambda^n + c_1 \lambda^{n-1} + \cdots + c_{n-1} \lambda + c_n = 0$$

$p(\lambda)$  称为矩阵  $\mathbf{A}$  的**特征多项式**, 方程  $p(\lambda) = \lambda^n + c_1 \lambda^{n-1} + \cdots + c_{n-1} \lambda + c_n = 0$  称为矩阵  $\mathbf{A}$  的

**特征方程**。因为  $n$  次代数方程  $p(\lambda)$  在复数域中有  $n$  个根  $\lambda_1, \lambda_2, \dots, \lambda_n$  , 故

$$p(\lambda) = (\lambda - \lambda_1)(\lambda - \lambda_2) \cdots (\lambda - \lambda_n)$$

由 (3.0) 式中的行列式展开可得

$$-c_1 = \lambda_1 + \lambda_2 + \cdots + \lambda_n = \sum_{i=1}^n a_{ii}$$

$$c_n = (-1)^n \lambda_1 \lambda_2 \cdots \lambda_n = (-1)^n \det \mathbf{A}$$

故矩阵  $\mathbf{A} = (a_{ij}) \in \mathbf{R}^{n \times n}$  的  $n$  个特征值  $\lambda_1, \lambda_2, \dots, \lambda_n$  是它们特征方程 (3.0) 的  $n$  个根。并有

$$\det \mathbf{A} = \lambda_1 \lambda_2 \cdots \lambda_n$$

及

$$\text{tr}\mathbf{A} = \sum_{i=1}^n a_{ii} = \sum_{i=1}^n \lambda_i$$

称  $\text{tr}\mathbf{A}$  为  $\mathbf{A}$  的迹。

$\mathbf{A}$  的特征值  $\lambda$  和特征向量  $\mathbf{x}$  还有以下性质:

- (1)  $\mathbf{A}^T$  与  $\mathbf{A}$  有相同的特征值  $\lambda$  及特征向量  $\mathbf{x}$ .
- (2) 若  $\mathbf{A}$  非奇, 则  $\mathbf{A}^{-1}$  的特征值为  $\lambda^{-1}$ , 特征向量为  $\mathbf{x}$ .
- (3) 相似矩阵  $\mathbf{B}=\mathbf{P}^{-1}\mathbf{A}\mathbf{P}$  有相同的特征多项式。

### 3.1.3 特殊矩阵

设  $\mathbf{A} = (a_{ij}) \in \mathbf{R}^{n \times n}$ .

- (1) 对角矩阵 如果当  $i \neq j$  时,  $a_{ij} = 0$ 。
- (2) 三对角矩阵 如果当  $|i - j| > 1$  时,  $a_{ij} = 0$
- (3) 上三角矩阵 如果当  $i > j$  时,  $a_{ij} = 0$
- (4) 上 海森伯格矩阵 (Hessenberg) 矩阵  $i > j + 1$  时,  $a_{ij} = 0$
- (5) 对称矩阵 如果  $\mathbf{A} = \mathbf{A}^T$ 。
- (6) 对称正定矩阵 如果①  $\mathbf{A} = \mathbf{A}^T$ , ② 对任意非零向量  $\mathbf{x} \in \mathbf{R}^n$ ,  $(\mathbf{A}\mathbf{x}, \mathbf{x}) = \mathbf{x}^T \mathbf{A} \mathbf{x} > 0$ 。
- (7) 正交矩阵如果  $\mathbf{A}^{-1} = \mathbf{A}^T$ 。
- (8) 初等置换阵 由单位矩阵  $\mathbf{I}$  交换第  $i$  行与第  $j$  行 (交换第  $i$  列与第  $j$  列), 得到的矩阵记为  $\mathbf{I}_{ij}$ , 且

$$\mathbf{I}_{ij} \mathbf{A} = \tilde{\mathbf{A}} \quad (\text{为交换 } \mathbf{A} \text{ 第 } i \text{ 行与第 } j \text{ 行得到的矩阵})$$

$$\mathbf{A} \mathbf{I}_{ij} = \mathbf{B} \quad (\text{为交换 } \mathbf{A} \text{ 第 } i \text{ 列与第 } j \text{ 列得到的矩阵})$$

- (9) 置换阵 由初等置换阵的乘积得到的矩阵。

**定理 3.1** 设  $\mathbf{A} \in \mathbf{R}^{n \times n}$ , 则下述命题等价:

- (1) 对任何  $\mathbf{b} \in \mathbf{R}^n$ , 方程  $\mathbf{A}\mathbf{x}=\mathbf{b}$  有唯一解;
- (2) 齐次方程组  $\mathbf{A}\mathbf{x}=0$  只有唯一解  $\mathbf{x}=0$ ;
- (3)  $\det(\mathbf{A}) \neq 0$

(4)  $\mathbf{A}^{-1}$  存在;

(5)  $\mathbf{A}$  的秩  $\text{rank}(\mathbf{A}) = n$ .

**定理 3.2** 设  $\mathbf{A} \in \mathbf{R}^{n \times n}$  为对称正定矩阵, 则

(1)  $\mathbf{A}$  为非奇异矩阵,  $\mathbf{A}^{-1}$  亦是对称正定矩阵;

(2) 记  $\mathbf{A}_k$  为  $\mathbf{A}$  的顺序主子阵, 则  $\mathbf{A}_k$  ( $k=1,2,\dots,n$ ) 亦是对称正定矩阵, 其中

$$A_k = \begin{pmatrix} a_{11} & \cdots & a_{1k} \\ \vdots & \ddots & \vdots \\ a_{k1} & \cdots & a_{kk} \end{pmatrix}, \quad k=1,2,\dots,n.$$

(3)  $\mathbf{A}$  的特征根  $\lambda_i > 0$  ( $i=1,2,\dots,n$ ).

(4)  $\mathbf{A}$  的顺序主子式都大于零, 即  $\det(\mathbf{A}_k) > 0$  ( $k=1,2,\dots,n$ ).

**定理 3.3** 设  $\mathbf{A} \in \mathbf{R}^{n \times n}$  为对称矩阵, 如果  $\det(\mathbf{A}_k) > 0$  ( $k=1,2,\dots,n$ ), 或  $\mathbf{A}$  的特征值  $\lambda_i > 0$  ( $i=1,2,\dots,n$ ), 则  $\mathbf{A}$  为对称正定矩阵。

### 3.2 高斯消去法

设线性方程组

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \\ \cdots \quad \cdots \quad \cdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n = b_n \end{cases} \quad (3.1)$$

这里  $a_{ij}$  ( $i, j=1, 2, \dots, n$ ) 为方程组的系数,  $b_i$  ( $i=1, 2, \dots, n$ ) 为方程组自由项。方程组 (3.1) 的矩阵形式为:

$$\mathbf{A}\mathbf{X} = \mathbf{b}$$

其中

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix} \quad X = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \quad b = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}$$

本章讨论计算机上常用而有效的直接解法——高斯消去法和矩阵的三角分解等问题。为方便计, 设所讨论的线性方程组的系数行列式不等于零 ( $\det(\mathbf{A}) \neq 0$ )。

高斯（Gauss）消去法是解线性方程组最常用的方法之一，它的基本思想是通过逐步消元，把方程组化为系数矩阵为三角形矩阵的同解方程组，然后用回代法解此三角形方程组得原方程组的解。下面先讨论三角形方程组的解法。

### 3.2.1. 三角形方程组的解法

三角形方程组是指下面两种形式的方程组

$$\begin{cases} a_{11}x_1 & = b_1 \\ a_{21}x_1 + a_{22}x_2 & = b_2 \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n & = b_n \end{cases} \quad (3.2)$$

和

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\ a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \\ \cdots \\ a_{nn}x_n = b_n \end{cases} \quad (3.3)$$

方程组（3.2）叫做下三角形方程组，方程组（3.3）叫做上三角形方程组，三角形方程组的求解是很简单的。

如果  $a_{ii} \neq 0, i = 1, 2, \cdots, n$ ，则（3.2）的解为

$$\begin{cases} x_1 = b_1 / a_{11} \\ x_k = (b_k - a_{k1}x_1 - a_{k2}x_2 - \cdots - a_{k,k-1}x_{k-1}) / a_{kk} \end{cases} \quad k = 2, 3, \cdots, n \quad (3.4)$$

此过程称为前推过程。

同样地，若  $a_{ii} \neq 0, i = 1, 2, \cdots, n$ ，则（3.3）的解为

$$\begin{cases} x_n = b_n / a_{nn} \\ x_k = (b_k - a_{k,k+1}x_{k+1} - \cdots - a_{kn}x_n) / a_{kk} \end{cases} \quad k = n-1, n-2, \cdots, 1 \quad (3.5)$$

此过程称为回代过程。

从上面的公式来看，求出  $x_k$ ，需要作  $k - 1$  次乘法和加减法及一次除法，总共完成  $1 + 2 + \cdots + n = \frac{n^2}{2}$  次乘法、加法及  $n$  次除法。

从（3.4）、（3.5）可以看出，求解三角形方程组是很简单的，只要把方程组化成了等价的三角形方程组，求解过程就很容易完成。

### 3.2.2 高斯消去法

为便于叙述，先以一个三阶线性方程组为例来说明高斯消去法的基本思想。

$$\begin{cases} 2x_1 + 3x_2 + 4x_3 = 6 & \text{(I)} \\ 3x_1 + 5x_2 + 2x_3 = 5 & \text{(II)} \\ 4x_1 + 3x_2 + 30x_3 = 32 & \text{(III)} \end{cases}$$

把方程 (I) 乘  $(-\frac{3}{2})$  后加到方程 (II) 上去, 把方程 (I) 乘  $(-\frac{4}{2})$  后加到方程 (III) 上去,

即可消去方程 (II)、(III) 中的  $x_1$ , 得同解方程组

$$\begin{cases} 2x_1 + 3x_2 + 4x_3 = 6 & \text{(I)} \\ 0.5x_2 - 4x_3 = -4 & \text{(II)} \\ -3x_2 + 22x_3 = 20 & \text{(III)} \end{cases}$$

将方程 (II) 乘  $(\frac{3}{0.5})$  后加于方程 (III), 得同解方程组:

$$\begin{cases} 2x_1 + 3x_2 + 4x_3 = 6 & \text{(I)} \\ 0.5x_2 - 4x_3 = -4 & \text{(II)} \\ -2x_3 = -4 & \text{(III)} \end{cases}$$

由回代公式 (3.5) 得  $x_3 = 2$   $x_2 = 8$   $x_1 = -13$

下面考察一般形式的线性方程组的解法, 为叙述问题方便, 将  $b_i$  写成  $a_{i,n+1}$ ,  $i = 1, 2, \dots, n$ 。

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + \dots + a_{1n}x_n = a_{1,n+1} \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + \dots + a_{2n}x_n = a_{2,n+1} \\ \dots \quad \dots \quad \dots \\ a_{n1}x_1 + a_{n2}x_2 + a_{n3}x_3 + \dots + a_{nn}x_n = a_{n,n+1} \end{cases} \quad (3.6)$$

如果  $a_{11} \neq 0$ , 将第一个方程中  $x_1$  的系数化为 1, 得

$$x_1 + a_{12}^{(1)}x_2 + \dots + a_{1n}^{(1)}x_n = a_{1,n+1}^{(1)}$$

其中  $a_{1j}^{(1)} = a_{1j}^{(0)} / a_{11}^{(0)} \quad j = 1, \dots, n+1$

(记  $a_{ij}^{(0)} = a_{ij} \quad i = 1, 2, \dots, n; \quad j = 1, 2, \dots, n+1$ )

从其它  $n-1$  个方程中消  $x_1$ , 使它变成如下形式

$$\left\{ \begin{array}{l} x_1 + a_{12}^{(1)}x_2 + \cdots + a_{1n}^{(1)}x_n = a_{1,n+1}^{(1)} \\ a_{22}^{(1)}x_2 + \cdots + a_{2n}^{(1)}x_n = a_{2,n+1}^{(1)} \\ \cdots \\ a_{n2}^{(1)}x_2 + \cdots + a_{nn}^{(1)}x_n = a_{n,n+1}^{(1)} \end{array} \right. \quad (3.7)$$

其中  $a_{ij}^{(1)} = a_{ij} - m_{i1} \cdot a_{ij}^{(1)} \quad i = 2, \cdots, n$

$$m_{i1} = \frac{a_{i1}^{(1)}}{a_{11}} \quad j = 2, 3, \cdots, n+1$$

由方程 (3.6) 到 (3.7) 的过程中, 元素  $a_{11}$  起着重要的作用, 特别地, 把  $a_{11}$  称为主元素。

如果 (3.7) 中  $a_{22}^{(1)} \neq 0$ , 则以  $a_{22}^{(1)}$  为主元素, 又可以把方程组 (3.7) 化为:

$$\left\{ \begin{array}{l} x_1 + a_{12}^{(1)}x_2 + \cdots + a_{1n}^{(1)}x_n = a_{1,n+1}^{(1)} \\ x_2 + a_{23}^{(2)}x_3 + \cdots + a_{2n}^{(2)}x_n = a_{2,n+1}^{(2)} \\ a_{33}^{(2)}x_3 + \cdots + a_{3n}^{(2)}x_n = a_{3,n+1}^{(3)} \\ \vdots \\ a_{n3}^{(2)}x_3 + \cdots + a_{nn}^{(2)}x_n = a_{n,n+1}^{(2)} \end{array} \right. \quad (3.8)$$

针对 (3.8) 继续消元, 重复同样的手段, 第  $k$  步所要加工的方程组是:

$$\left\{ \begin{array}{l} x_1 + a_{12}^{(1)}x_2 + a_{13}^{(1)}x_3 + \cdots + a_{1n}^{(1)}x_n = a_{1,n+1}^{(1)} \\ x_2 + a_{23}^{(2)}x_3 + \cdots + a_{2n}^{(2)}x_n = a_{2,n+1}^{(2)} \\ \cdots \\ x_{k-1} + a_{k-1,k}^{(k-1)}x_k + \cdots + a_{k-1,n}^{(k-1)}x_n = a_{k-1,n+1}^{(k-1)} \\ a_{kk}^{(k-1)}x_k + \cdots + a_{kn}^{(k-1)}x_n = a_{k,n+1}^{(k-1)} \\ \cdots \\ a_{nk}^{(k-1)}x_k + \cdots + a_{nn}^{(k-1)}x_n = a_{n,n+1}^{(k-1)} \end{array} \right.$$

设  $a_{kk}^{(k-1)} \neq 0$ , 第  $k$  步先使上述方程组中第  $k$  个方程中  $x_k$  的系数化为 1:

$$x_k + a_{k,k+1}^{(k)}x_{k+1} + \cdots + a_{kn}^{(k)}x_n = a_{k,n+1}^{(k)}$$

然后再从其它  $(n-k)$  个方程中消  $x_k$ , 消元公式为:

$$\begin{cases} a_{kj}^{(k)} = \frac{a_{kj}^{(k-1)}}{a_{kk}^{(k-1)}} & j = k, k+1, \dots, n+1 \\ a_{ij}^{(k)} = a_{ij}^{(k-1)} - a_{ik}^{(k-1)} \cdot a_{kj}^{(k)} & j = k+1, \dots, n+1 \\ j = k+1, \dots, n+1 \\ i = k+1, \dots, n \end{cases} \quad (3.9)$$

按照上述步骤进行  $n$  次后，将原方程组加工成下列形式：

$$\begin{cases} x_1 + a_{12}^{(1)}x_2 + a_{13}^{(1)}x_3 + \dots + a_{1n}^{(1)}x_n = a_{1,n+1}^{(1)} \\ x_2 + a_{23}^{(2)}x_3 + \dots + a_{2n}^{(2)}x_n = a_{2,n+1}^{(2)} \\ \dots \\ x_{n-1} + a_{nn}^{(n-1)}x_n = a_{n-1,n+1}^{(n-1)} \\ x_n = a_{n,n+1}^{(n)} \end{cases}$$

回代公式为：

$$\begin{cases} x_n = a_{n,n+1}^{(n)} \\ x_k = a_{k,n+1}^{(k)} - \sum_{j=k+1}^n a_{kj}^{(k)} x_j & k = n-1, \dots, 1 \end{cases} \quad (3.10)$$

综上所述，高斯消去法分为消元过程与回代过程，消元过程将所给方程组加工成上三角形方程组，再经回代过程求解。

由于计算时不涉及  $x_i, i = 1, 2, \dots, n$ ，所以在存贮时可将方程组  $AX = b$ ，写成增广矩阵  $(A, b)$  存贮。

下面，我们统计一下高斯消去法的工作量；在(3.9)第一个式子中，每执行一次需要  $n - (n - k)$  次除法，在(3.9)第二个式子中，每执行一次需要  $[n - (k - 1)] \times (n - k)$  次除法。因此在消元过程中，共需要

$$\begin{aligned} & \sum_{k=1}^n [(n - k + 1) \times (n - k) + (n - k + 1)] \\ &= \sum_{k=1}^n (n - k + 1)^2 = \frac{1}{6} n(n+1)(2n+1) \end{aligned}$$

次乘法。此外，回代过程共有

$$\sum_{k=1}^n (n - k) = \frac{n}{2} (n - 1)$$

次乘法。汇总在一起，高斯消去法的计算量为：

$$\frac{n}{3}(n^2 + 3n - 1) = \frac{n^3}{3} + n^2 - \frac{n}{3}$$

次乘除法。

### 3.2.3 主元素消去法

前述的消去过程中，未知量是按其出现于方程组中的自然顺序消去的，所以又叫顺序消去法。

实际上已经发现顺序消去法有很大的缺点。设用作除数的  $a_{kk}^{(k-1)}$  为主元素，首先，消元过程中可能出现  $a_{kk}^{(k-1)}$  为零的情况，此时消元过程亦无法进行下去；其次如果主元素  $a_{kk}^{(k-1)}$  很小，由于舍入误差和有效位数消失等因素，其本身常常有较大的相对误差，用其作除数，会导致其它元素数量级的严重增长和舍入误差的扩散，使得所求的解误差过大，以致失真。

我们来看一个例子：

例

$$\begin{cases} 0.0001x_1 + 1.00x_2 = 1.00 \\ 1.00x_1 + 1.00x_2 = 2.00 \end{cases}$$

它的精确解为：

$$\begin{aligned} x_1 &= \frac{10000}{9999} \approx 1.00010 \\ x_2 &= \frac{9998}{9999} \approx 0.99990 \end{aligned}$$

用顺序消去法，第一步以 0.0001 为主元，从第二个方程中消  $x_1$  后可得：

$$-10000x_2 = -10000 \quad x_2 = 1.00$$

回代可得  $x_1 = 0.00$

显然，这不是解。

造成这个现象的原因是：第一步主元素太小，使得消元后所得的三角形方程组很不准确所致。

如果我们选第二个方程中  $x_1$  的系数 1.00 为主元素来消去第一个方程中的  $x_1$ ，则得出如下方程式：

$$1.00 x_1 = 1.00 \quad x_1 = 1.00$$

这是真解的三位正确舍入值。

从上述例子中可以看出，在消元过程中适当选取主元素是十分必要的。误差分析的理论 and 计



算实践均表明：顺序消元法在系数矩阵  $A$  为对称正定时，可以保证此过程对舍入误差的数值稳定性，对一般的矩阵则必须引入选取主元素的技巧，方能得到满意的结果。

### 列主元消去法

在列主元消去法中，未知数仍然是顺序地消去的，但是把各方程中要消去的那个未知数的系数按绝对值最大值作为主元素，然后用顺序消去法的公式求解。

例：用列主元高斯消去法求解方程

$$\begin{cases} 2x_1 - x_2 + 3x_3 = 1 \\ 4x_1 + 2x_2 + 5x_3 = 4 \\ x_1 + 2x_2 = 7 \end{cases}$$

由于解方程组取决于它的系数，因此可用这些系数（包括右端项）所构成的“增广矩阵”作为方程组的一种简化形式。对这种增广矩阵施行消元手续：

$$\begin{pmatrix} 2 & -1 & 3 & 1 \\ 4^* & 2 & 5 & 4 \\ 1 & 2 & 0 & 7 \end{pmatrix}$$

第一步将 4 选为主元素，并把主元素所在的行定为主元行，然后将主元行换到第一行得到

$$\begin{pmatrix} 4 & 2 & 5 & 4 \\ 2 & -1 & 3 & 1 \\ 1 & 2 & 0 & 7 \end{pmatrix} \xrightarrow{\text{第一步消元}} \begin{pmatrix} 1 & 0.5 & 1.25 & 1 \\ 0 & -2^* & 0.5 & -1 \\ 0 & 1.5 & -1.25 & 6 \end{pmatrix}$$

$$\xrightarrow{\text{第二步消元}} \begin{pmatrix} 1 & 0.5 & 1.25 & 1 \\ 0 & 1 & -0.25 & 0.5 \\ 0 & 0 & -0.875 & 5.25 \end{pmatrix} \xrightarrow{\text{第三步消元}} \begin{pmatrix} 1 & 0.5 & 1.25 & 1 \\ 0 & 1 & -0.25 & 0.5 \\ 0 & 0 & 1 & -6 \end{pmatrix}$$

消元过程的结果归结到下列三角形方程组：

$$\begin{cases} x_1 + 0.5x_2 + 1.25x_3 = 1 \\ x_2 - 0.25x_3 = 0.5 \\ x_3 = -6 \end{cases}$$

回代，得

$$\begin{cases} x_1 = 9 \\ x_2 = -1 \\ x_3 = -6 \end{cases}$$

列主元消去法计算步骤：

- 1、输入矩阵阶数  $n$ ，增广矩阵  $A(n, n+1)$

2、对  $k=1,2,\cdots,n$

(1) 按列选主元：选取  $l$  使

$$|a_{lk}| = \max_{k \leq i \leq n} |a_{ik}| \neq 0$$

(2) 如果  $l \neq k$ ，交换  $A(n, n+1)$  的第  $k$  行与底  $l$  行元素

(3) 消元计算：

$$m_{ik} \leftarrow \frac{a_{ik}}{a_{kk}} \quad i = k+1, \cdots, n$$

$$a_{ij} \leftarrow a_{ij} - m_{ik} a_{kj} \quad i, = k+1, \cdots, n \quad j = k+1, \cdots, n+1$$

3、回代计算：

$$x_i \leftarrow a_{i,n+1} - \sum_{j=i+1}^n a_{ij} x_j \quad i = n, n-1, \cdots, 1$$

4、输出解向量  $x_1, x_2, \cdots, x_n$ 。

### 3.2.4 无回代过程的主元消去法

设有线性代数方程组

$$\mathbf{AX} = \mathbf{b}$$

其中  $\mathbf{A}$  为  $n \times n$  阶非奇异矩阵， $\mathbf{b}$  为  $n \times 1$  阶矩阵（列向量），由  $\mathbf{A}^{-1}\mathbf{AX} = \mathbf{A}^{-1}\mathbf{b}$  得  $\mathbf{X} = \mathbf{A}^{-1}\mathbf{b}$ 。

因此，只要求出  $\mathbf{A}^{-1}$  就可以得到解  $\mathbf{X}$ 。另外，有许多问题需要求矩阵的逆，例如常用的回归分析方法中，要求出相关矩阵的逆，因此，有必要讨论在计算机上常用的求逆方法。

**步骤：**

第一步选主元，在第一列中选绝对值最大的元素  $a_{k1} = \max_{1 \leq j \leq n} \{a_{j1}\}$ ，设第  $k$  行为主元行，将主元行换至第一行，将第一个方程中  $x_1$  的系数变为 1，并从其余  $n-1$  个方程中消去  $x_1$ 。

第二步：在第二列后  $n-1$  个元素中选主元，将第二个方程中  $x_2$  的系数变为 1，并从其它  $n-1$  个方程中消去  $x_2$ 。

.....

依此类推，直到每个方程中仅有一个变量为止。此过程进行到第  $k-1$  步后，方程组将变成如下形式：

$$\left\{ \begin{array}{l} x_1 + a_{1,k}^{(k-1)} x_k + \cdots + a_{1,n}^{(k-1)} x_n = a_{1,n+1}^{(k-1)} \\ x_2 + a_{2,k}^{(k-1)} x_k + \cdots + a_{2,n}^{(k-1)} x_n = a_{2,n+1}^{(k-1)} \\ \vdots \qquad \qquad \qquad \dots \qquad \qquad \dots \\ x_{k-1} + a_{k-1,k}^{(k-1)} x_k + \cdots + a_{k-1,n}^{(k-1)} x_n = a_{k-1,n+1}^{(k-1)} \\ a_{kk}^{(k-1)} x_k + \cdots + a_{kn}^{(k-1)} x_n = a_{k,n+1}^{(k-1)} \\ \vdots \\ a_{nk}^{(k-1)} x_k + \cdots + a_{nn}^{(k-1)} x_n = a_{n,n+1}^{(k-1)} \end{array} \right.$$

第  $k$  步: 在第  $k$  列后  $n-k$  个元素中选主元, 换行, 将第  $k$  个方程  $x_k$  的系数变为 1, 从其它方程中消去变量  $x_k$ , 消元公式为:

$$\begin{cases} a_{kj}^{(k)} = \frac{a_{kj}^{(k-1)}}{a_{kk}^{(k-1)}} & j = k, k+1, \dots, n+1 \\ a_{ij}^{(k)} = a_{ij}^{(k-1)} - a_{ik}^{(k-1)} a_{kj}^{(k)} & \\ j = k+1, \dots, n+1 \\ i = 1, 2, \dots, k-1, k+1, \dots, n \end{cases} \quad (3.11)$$

对  $k=1, 2, \cdots$ , 按上述步骤进行到第  $n$  步后, 方程组变为:

$$\begin{cases} x_1 = a_{1,n+1}^{(n)} \\ x_2 = a_{2,n+1}^{(n)} \\ \vdots \\ x_n = a_{n,n+1}^{(n)} \end{cases}$$

即为所求的解

### 3.2.5 无回代消去法的应用

### (1) 解线性方程组系

设要解的线性方程组系为:

$$AX = b_1, \quad AX = b_2, \quad AX = b_m$$

其中

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix} \quad X = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \quad b_i = \begin{pmatrix} a_{1,n+i}^{(i)} \\ a_{2,n+i}^{(i)} \\ \vdots \\ a_{n,n+i}^{(i)} \end{pmatrix}$$

上述方程组系可以写为

$$AX = B = (b_1, \dots, b_m)$$

因此  $X = A^{-1}B$

即为线性方程组系的解。在计算机上只需要增加几组右端常数项的存储单元，其结构解一个方程组时一样。

行	系数				
1	$a_{11}$	$a_{12}$	$\cdots$	$a_{1n}$	$a_{1,n+1}^{(1)} \quad \cdots \quad a_{1,n+1}^{(m)}$
2	$a_{21}$	$a_{22}$	$\cdots$	$a_{2n}$	$a_{2,n+1}^{(1)} \quad \cdots \quad a_{2,n+1}^{(m)}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$n$	$a_{n1}$	$a_{n2}$	$\cdots$	$a_{nn}$	$a_{n,n+1}^{(1)} \quad \cdots \quad a_{n,n+1}^{(m)}$

## (2) 求逆矩阵

设  $A = (a_{ij})_{n \times n}$  是非奇矩阵， $|A| \neq 0$ ，且令

$$X = A^{-1} = (X_{ij})_{n \times n}$$

由于  $AA^{-1} = AX = I$

因此，求  $A^{-1}$  的问题相当于解下列线性方程组

$$A \begin{pmatrix} x_{11} \\ x_{21} \\ \vdots \\ x_{n1} \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad \dots, \quad A \begin{pmatrix} x_{1n} \\ x_{2n} \\ \vdots \\ x_{nn} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}$$

也就是相当于 (I) 中  $m = n, B = I$  的情形。

## (3) 求行列式的值

由于行列式任意一行(列)的元素乘以同一个数后，加到另一行(列)的对应元素上，其行列式的值不变；任意对换两行(列)的位置其值反号；三角矩阵的行列式之值等于其主对角元素的乘积。因此，可用高斯消去法将  $|A|$  化成

$$|A| = \begin{vmatrix} a_{11}^{(1)} & & & \\ & a_{22}^{(2)} & & \\ & & \ddots & \\ & & & a_{nn}^{(n)} \end{vmatrix} = a_{11}^{(1)} \cdots a_{nn}^{(n)}$$

## 3.3 解三对角方程组的追赶法

在实际问题中，经常遇到以下形式的方程组

$$\left\{ \begin{array}{lcl} b_1 x_1 + c_1 x_2 & & = d_1 \\ a_2 x_1 + b_2 x_2 + c_2 x_3 & & = d_2 \\ & \dots & \\ & a_k x_{k-1} + b_k x_k + c_k x_{k+1} & = d_k \\ & \dots & \\ & a_{n-1} x_{n-2} + b_{n-1} x_{n-1} + c_{n-1} x_n & = d_{n-1} \\ & a_n x_{n-1} + b_n x_n & = d_n \end{array} \right. \quad (3.12)$$

这种方程组的系数矩阵称为三对角矩阵

$$A = \begin{pmatrix} b_1 & c_1 & & & & \\ a_2 & b_2 & c_2 & & & \\ \ddots & & & & & \\ & a_k & b_k & c_k & & \\ & & \ddots & & b_{n-1} & c_{n-1} \\ & & & a_n & b_n & \end{pmatrix}$$

以下针对这种方程组的特点提供一种简便有效的算法—追赶法。

追赶法实际上是高斯消去法的一种简化形式，它同样分消元与回代两个过程。

先将 (3.12) 第一个方程中  $x_1$  的系数化为 1

$$x_1 + \frac{c_1}{b_1} x_2 = \frac{d_1}{b_1}$$

$$\text{记} \quad r_1 = \frac{c_1}{b_1} \quad y_1 = \frac{d_1}{b_1} \quad (3.13)$$

$$\text{有} \quad x_1 + r_1 x_2 = y_1$$

注意到剩下的方程中，实际上只有第二个方程中含有变量  $x_1$ ，因此消元手续可以简化。利用 (3.13)

可将第二个方程化为

$$x_2 + r_2 x_3 = y_2$$

这样一步一步地顺序加工 (3.12) 的每个方程，设第  $k-1$  个方程已经变成

$$x_{k-1} + r_{k-1} x_k = y_{k-1} \quad (3.14)$$

再利用 (3.14) 从第  $k$  个方程中消去  $x_{k-1}$ ，得：

$$(b_k - r_{k-1} a_k) x_k + c_k x_{k+1} = d_k - y_{k-1} a_k$$

同除  $(b_k - r_{k-1} a_k)$ ，得

$$x_k + \frac{c_k}{b_k - r_{k-1}a_k} x_{k+1} = \frac{d_k - y_{k-1}a_k}{b_k - r_{k-1}a_k} \quad k = 2, 3, \dots, n$$

记

$$r_k = \frac{c_k}{b_k - r_{k-1}a_k} \quad y_k = \frac{d_k - y_{k-1}a_k}{b_k - r_{k-1}a_k}$$

则有

$$x_k + r_k x_{k+1} = y_k$$

这样做  $n-1$  步以后, 便得到:

$$x_{n-1} + r_{n-1}x_n = y_{n-1}$$

将上式与 (3.12) 中第  $n$  个方程联立, 即可解出

$$x_n = y_n$$

这里

$$y_n = \frac{d_n - y_{n-1}a_n}{b_n - r_{n-1}a_n}$$

于是, 通过消元过程, 所给方程组 (3.12) 可归结为以下更为简单的形式:

$$\begin{cases} x_1 + r_1 x_2 = y_1 \\ \vdots \\ x_k + r_k x_{k+1} = y_k \\ \vdots \\ x_n = y_n \end{cases} \quad (3.15)$$

这种方程组称作二对角型方程组, 其系数矩阵中的非零元素集中分步在主对角线和一条次主对角线上

$$\begin{pmatrix} 1 & r_1 & & & \\ & 1 & r_2 & & \\ & & \ddots & 1 & r_k \\ & & & \ddots & 1 & r_{n-1} \\ & & & & & 1 \end{pmatrix}$$

对加工得到的方程组 (3.15) 自下而上逐步回代, 即可依次求出  $x_n, x_{n-1}, \dots, x_1$ , 计算公式为:

$$\begin{cases} x_n = y_n \\ x_k = y_k - r_k x_{k+1} \end{cases} \quad k = n-1, n-2, \dots, 1 \quad (3.16)$$

上述算法就是追赶法，它的消元过程与回代过程分别称作“追”过程与“赶”过程。综合追与赶的过程，得如下计算公式：

$$\begin{cases} r_1 = \frac{c_1}{b_1} & y_1 = \frac{d_1}{b_1} \\ r_k = \frac{c_k}{b_k - r_{k-1}a_k} \\ y_k = \frac{d_k - y_{k-1}a_k}{b_k - r_{k-1}a_k} \end{cases} \quad k = 2, 3, \dots, n \quad (3.17)$$

$$\begin{cases} x_n = y_n \\ x_k = y_k - r_k x_{k+1} \end{cases} \quad k = n-1, n-2, \dots, 1 \quad (3.18)$$

### 3.4 矩阵的三角分解及其在解方程组中的应用

下面我们进一步用矩阵理论来分析高斯消去法，从而建立矩阵的三角分解定理，而这个定理在解方程组的直接解法中起着重要作用，我们将利用它来推导某些具有特殊的系数矩阵方程组的数值解法。

考虑线性方程组

$$AX = b \quad A \in \mathbb{R}^{n \times n}$$

设解此方程组用高斯消去法能够完成（不进行变换两行的初等变换），由于对  $A$  施行初等变换相当于用初等矩阵左乘  $A$ ，于是，高斯消去法的求解过程用矩阵理论来叙述如下：

记：

$$L_k = \begin{pmatrix} 1 & & & \\ & 1 & & \\ & -l_{k+1,k} & 1 & \\ & \vdots & & \\ & -l_{n,k} & & 1 \end{pmatrix} \quad L_k^{-1} = \begin{pmatrix} 1 & & & \\ & 1 & & \\ & +l_{k+1,k} & 1 & \\ & \vdots & & \\ & +l_{n,k} & & 1 \end{pmatrix}$$

$$\text{其中 } l_{ji} = \frac{a_{ji}^{(i)}}{a_{ii}^{(i)}} \quad (j = i+1, \dots, n) \quad \text{记 } A^{(1)} \equiv A$$

于是

$$\begin{aligned}
LA^{(1)} &= \begin{pmatrix} 1 & & & \\ -\frac{a_{21}^{(1)}}{a_{11}^{(1)}} & 1 & & \\ \vdots & & \ddots & \\ -\frac{a_{n1}^{(1)}}{a_{11}^{(1)}} & & & \ddots 1 \end{pmatrix} \begin{pmatrix} a_{11}^{(1)} & a_{12}^{(1)} & \cdots & a_{1n}^{(1)} \\ a_{21}^{(1)} & A_{22}^{(1)} & & \\ \vdots & & & \\ a_{n1}^{(1)} & & & \end{pmatrix} \\
&= \begin{pmatrix} 1 & 0 \\ -\frac{c_1}{a_{11}^{(1)}} & I_{n-1} \end{pmatrix} \begin{pmatrix} a_{11}^{(1)} & r_1^T \\ c_1 & A_{22}^{(1)} \end{pmatrix} = \begin{pmatrix} a_{11}^{(1)} & r_1^T \\ 0 & A_{22}^{(1)} - \frac{c_1 r_1^T}{a_{11}^{(1)}} \end{pmatrix} \\
&= \begin{pmatrix} a_{11}^{(1)} & a_{12}^{(1)} & \cdots & a_{1n}^{(1)} \\ a_{22}^{(2)} & \cdots & a_{2n}^{(2)} \\ \vdots & & \vdots \\ a_{2n}^{(2)} & \cdots & a_{nn}^{(2)} \end{pmatrix} = \begin{pmatrix} A_{11}^{(2)} & A_{12}^{(2)} \\ 0 & A_{22}^{(2)} \end{pmatrix} = A^{(2)}
\end{aligned}$$

$$L_2 A^{(2)} = \begin{pmatrix} 1 & & 0^T \\ & 0 & 0^T \\ 0 & -\frac{c_2}{a_{22}^{(2)}} & I_{n-2} \end{pmatrix} \begin{pmatrix} a_{11} & r_1^T \\ a_{22}^{(2)} & r_2^T \\ 0 & c_2 & A_{33}^{(2)} \end{pmatrix} = \begin{pmatrix} A_{11}^{(3)} & A_{12}^{(3)} \\ 0 & A_{22}^{(3)} \end{pmatrix} = A^{(3)}$$

如此继续下去，到  $n-1$  步时有：

$$L_{n-1} A^{(n-1)} = \begin{pmatrix} a_{11}^{(1)} & \cdots & a_{1n}^{(1)} \\ & a_{22}^{(2)} & \cdots & a_{2n}^{(2)} \\ & & \ddots & \\ & & & a_{nn}^{(2)} \end{pmatrix} \equiv U$$

也就是说

$$L_{n-1} L_{n-2} \cdots L_2 L_1 A = U$$

记

$$L = L_1^{-1} L_2^{-1} \cdots L_{n-1}^{-1} = \begin{pmatrix} 1 & & \\ l_{21} & 1 & \\ l_{31} & l_{32} & \\ \vdots & \vdots & \ddots \\ l_{n1} & l_{n2} & & 1 \end{pmatrix}$$

则有  $A = LU$



其中  $L$  为单位下三角矩阵,  $U$  为上三角矩阵。

总结上述讨论得到重要性定理

**定理 3.4:** (矩阵的三角分解) 设  $A$  为  $n \times n$  实矩阵, 如果解  $AX = b$  用高斯消去法能够完成 (限制不进行行的交换, 即  $a_{kk}^{(k)} \neq 0, k = 1, 2, \dots, n$ ), 则矩阵  $A$  可分解为单位下三角矩阵  $L$  与上三角矩阵  $U$  的乘积。

$$A = LU$$

且这种分解是唯一的。

**证明:** 由上述的讨论, 存在性已得证, 现在证唯一性。

$$\text{设 } A = L_1 U_1 = LU$$

其中  $L_1, L$  为单位下三角阵,  $U_1, U$  为上三角阵, 设  $U_1^{-1}$  存在, 于是有

$$L^{-1}L_1 = UU_1^{-1}$$

上式右端为上三角矩阵, 左边为单位下三角阵, 故应为单位矩阵。即

$$L_1 = L \quad U_1 = U \quad \text{证完}$$

那么矩阵  $A$  在什么条件下才能保证约化主元素  $a_{kk}^{(k)} \neq 0 \quad (k = 1, 2, \dots, n)$ ? 为此给出以下

**定理 3.5:** 约化主元素  $a_{ii}^{(i)} \neq 0 \quad (i = 1, 2, \dots, k)$

充要条件是矩阵  $A$  的顺序主子式

$$D_1 = a_{11} \neq 0, \quad D_2 = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} \neq 0, \dots$$

$$D_k = \begin{vmatrix} a_{11} & \cdots & a_{1k} \\ \vdots & \ddots & \vdots \\ a_{k1} & \cdots & a_{kk} \end{vmatrix} \neq 0$$

由上述讨论知, 解  $AX = b$  的高斯消去法相当于实现了  $A$  的三角分解, 如果我们能直接从矩阵  $A$  的元素得到计算  $L, U$  的元素的公式, 实现  $A$  的三角分解, 而不需要任何中间步骤, 那么求解  $AX = b$  的问题就等价于求解两个三角形矩阵方程组

$$(1) Ly = b \quad \text{求 } y$$

$$(2) UX = y \quad \text{求 } x$$

下面来说明  $L, U$  的元素可以由  $A$  的元素直接计算确定。显然, 由矩阵乘法。

$$a_{li} = u_{li} \quad (i = 1, 2, \dots, n)$$

得到  $U$  的第一行元素；由  $a_{i1} = l_{i1}u_{11}$  得

$$l_{i1} = \frac{a_{i1}}{u_{11}} \quad (i = 1, 2, \dots, n)$$

即  $L$  的第一列元素。设已经求出  $U$  的第 1 行~第  $r-1$  行元素， $L$  的第 1 列~第  $r-1$  列元素，由矩阵乘法可得：

$$a_{ri} = \sum_{k=1}^n l_{rk}u_{ki} = \sum_{k=1}^{r-1} l_{rk}u_{ki} + u_{ri} \quad (l_{rk} = 0, r < k)$$

$$a_{ir} = \sum_{k=1}^n l_{ik}u_{kr} = \sum_{k=1}^{r-1} l_{ik}u_{kr} + l_{ir}u_{rr}$$

即可计算出  $U$  的第  $r$  行元素， $L$  的第  $r$  列元素。

综上所述，可得到用直接三角分解法解  $AX = b$  的计算公式。

$$(1) \quad u_{li} = a_{li} \quad i = 1, 2, \dots, n \quad (3.19)$$

$$l_{i1} = \frac{a_{i1}}{u_{11}} \quad i = 1, 2, \dots, n \quad (3.20)$$

对于  $r = 2, 3, \dots, n$  计算

(2) 计算  $U$  的第  $r$  行元素

$$u_{ri} = a_{ri} - \sum_{k=1}^{r-1} l_{rk}u_{ki} \quad (i = r, r+1, \dots, n) \quad (3.21)$$

(3) 计算  $L$  的第  $r$  列元素 ( $r \neq n$ )

$$l_{ir} = \frac{(a_{ir} - \sum_{k=1}^{r-1} l_{ik}u_{kr})}{u_{rr}} \quad (i = r+1, \dots, n) \quad (3.22)$$

$$(4) \quad \begin{cases} y_1 = b \\ y_i = b_i - \sum_{k=1}^{i-1} l_{ik}y_k \end{cases} \quad (i = 2, 3, \dots, n) \quad (3.23)$$

$$(5) \quad \begin{cases} x_n = \frac{y_n}{u_{nn}} \\ x_i = \left( y_i - \sum_{k=i+1}^n u_{ik}x_k \right) / u_{ii} \end{cases} \quad (i = n-1, \dots, 2, 1) \quad (3.24)$$

(1)、(2)、(3) 是矩阵  $A$  的 LU 分解公式,称为杜利特尔 (Doolittle) 分解。同理,可推出矩

阵  $A = LU$  分解的另一种计算公式，其中  $L$  为下三角， $U$  为单位上三角，这种矩阵的分解公式称为矩阵的克劳特（Crout）分解。

例：用直接三角分解法解

$$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 5 & 2 \\ 3 & 1 & 5 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 14 \\ 18 \\ 20 \end{pmatrix}$$

解：（1）对于  $r=1$ ，利用公式（3.19）、（3.20）计算

$$u_{11} = 1 \quad u_{12} = 2 \quad u_{13} = 3$$

$$l_{21} = 2 \quad l_{31} = 3$$

（2）对于  $r=2$ ，利用（3.21）计算

$$u_{22} = a_{22} - l_{21}u_{12} = 5 - 2 \times 2 = 1$$

$$u_{23} = a_{23} - l_{21}u_{13} = 2 - 2 \times 3 = -4$$

$$l_{32} = \frac{(a_{32} - l_{31}u_{12})}{u_{22}} = \frac{(1 - 3 \times 2)}{1} = -5$$

（3） $r=3$

$$u_{33} = a_{33} - (l_{31}u_{13} + l_{32}u_{23}) = 5 - (3 \times 3 + (-5) \cdot (-4)) = -24$$

于是

$$A = \begin{pmatrix} 1 & & \\ 2 & 1 & \\ 3 & -5 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 \\ & 1 & -4 \\ & & -24 \end{pmatrix} = LU$$

（4）求解：

$$Ly = b \quad \text{得到}$$

$$y_1 = 14$$

$$y_2 = b_2 - l_{21}y_1 = 18 - 2 \times 14 = -10$$

$$y_3 = b_3 - (l_{31}y_1 + l_{32}y_2) = 20 - (3 \times 14 + (-5)(-10)) = -72$$

从而  $y = (14, -10, -72)^T$

由  $Ux = y$  得到

$$x_3 = \frac{y_3}{u_{33}} = \frac{-72}{-24} = 3$$

$$x_2 = \frac{(y_2 - u_{23}x_3)}{u_{22}} = \frac{-10 - (-4 \times 3)}{1} = 2$$

$$x_1 = \frac{y_1 - (u_{12}x_2 + u_{13}x_3)}{u_{11}} = \frac{14 - (2 \times 2 + 3 \times 3)}{1} = 1$$

$$x = (1, 2, 3)^T$$

### 3.5 平方根法

#### 3.5.1 矩阵的 LDR 分解

以上介绍了矩阵的三角分解及唯一性定理，为使分解式规范化，我们给出矩阵 LDR 分解。

**定理 3.6** 如果  $n$  阶方程  $A$  的所有顺序主子式均不等于零，则矩阵  $A$  存在唯一的分解式  $A = LDR$ ；其中  $L$  和  $R$  分别是  $n$  阶单位下三角阵和单位上三角阵， $D$  是  $n$  阶对角元素的不为零的对角阵，上述分解也称为  $A$  的 LDR 分解。

证明：充分性：因为  $A$  的顺序主子式均不为零，Gauss 消去法得以完成，即实现了一个 Doolittle 分解  $A = LU$ （或 Crout 分解  $A = \tilde{L}\tilde{U}$ ） $U$  的对角元素  $u_{ii} = a_{ii}^{(i)} \neq 0$  ( $i = 1, 2, \dots, n$ )，这时，令

$$D = \text{diag}(a_{11}^{(1)}, a_{22}^{(2)}, \dots, a_{nn}^{(n)}) \quad \text{则}$$

$$A = LU = LD(D^{-1}U) = LDR$$

其中  $R = D^{-1}U \quad D^{-1} = \text{diag}\left(\frac{1}{a_{11}^{(1)}}, \frac{1}{a_{22}^{(2)}}, \dots, \frac{1}{a_{nn}^{(n)}}\right)$

存在性得证。

唯一性：当  $A$  非奇时， $L$ 、 $D$ 、 $R$  皆非奇，若还存在另一个 LDR 分解  $A = L_1 D_1 R_1$ ，这里  $L_1$ 、 $D_1$ 、 $R_1$  也非奇，则有

$$LDR = L_1 D_1 R_1$$

即  $L_1^{-1}L = D_1 R_1 R^{-1} D^{-1}$

上式左端是单位下三角矩阵，右端是上三角矩阵，所以应该是单位矩阵，即

$$L_1^{-1}L = I \quad D_1 R_1 R^{-1} D^{-1} = I$$

$$L_1 = L \quad R_1 R^{-1} = D_1^{-1} D$$

从而必有  $R_1 R^{-1} = I \quad D_1^{-1} D = I$

也即  $R_1 = R \quad D_1 = D$  唯一性得证。

必要性：假定  $A$  有唯一的  $LDR$  分解，且  $L, D, R$  均非奇，从而  $L_i D_i R_i$  均非奇  $i = 1, 2, \dots, n$ ，而  $A_i = L_i D_i R_i$ ，所以  $A_i (i = 1, 2, \dots, n)$  也非奇。证完

### 3.5.2 平方根法

在科学研究和工程技术的实际计算中遇到的线性代数方程组，其系数矩阵往往具有对称正定性。对于系数矩阵具有这种特殊性质的方程组，上面介绍的直接三角分解还可以简化。得到“平方根法”。这是计算机上常用的有效方法之一。下面讨论对称正定矩阵的三角分解。

设  $A$  是  $n$  阶实矩阵，由线性代数知识知  $A$  是对称矩阵，即  $A = A^T$ ， $A$  是正定矩阵，即对于任意  $n$  维非零列向量  $X \neq 0$ ， $X \in R^n$ ，恒有  $X^T A X > 0$ ，对称正定矩阵有以下性质：

若  $A$  为对称正定矩阵，则  $A$  的各阶顺序主子式  $D_k > 0 (k = 1, 2, \dots, n)$ 。根据这条性质，我们就可以来讨论对称正定矩阵的三角分解，从而给出求解方程组的平方根法。

**定理 3.7：**（对称正定矩阵的三角分解）

如果  $A$  为对称正定矩阵，则存在一个实的非奇异下三角矩阵  $\tilde{L}$ ，使  $A = \tilde{L} \tilde{L}^T$ ，且当限定  $\tilde{L}$  的对角元素为正时，这种分解是唯一的。

证明：由  $A$  的对称正定性，则  $A$  的顺序主子式  $D_k \neq 0 (k = 1, 2, \dots, n)$ ，于是由定理 3 可知， $A$  总存在唯一的  $LDR$  分解。即  $A = LDR$ 。其中  $L$  是单位下三角阵， $D$  是非奇异的对角阵， $R$  是单位上三角阵。

又由  $A$  的对称性， $A^T = A$ ，则  $(LDR)^T = R^T D L^T = LDR$ ，由分解唯一性，于是有  $L = R^T$ ，从而得  $A = LDL^T$ ，这表明对称正定矩阵  $A$  的  $LDR$  分解具有特殊形式。

$$A = LDL^T$$

设  $D = \text{diag} (d_1, d_2, \dots, d_n)$ ， $d_j \neq 0 (j = 1, 2, \dots, n)$

下面我们进一步证明  $D$  的对角元素均为正数，即  $d_j > 0$ 。

由于  $L$  是单位下三角阵，所以对于单位坐标向量  $e_j = (0, \dots, 0, 1, 0, \dots, 0)^T$  存在非零向量  $X_j$ ，使  $L^T X_j = e_j (j = 1, 2, \dots, n)$ 。

$$\text{因此， } X_j^T A X_j = X_j^T (LDL^T) X_j = (L^T X_j)^T D (L^T X_j) = e_j^T D e_j = d_j$$

根据  $A$  是对称正定阵的定义，有  $X_j^T A X_j > 0$ ，从而  $d_j > 0 (j = 1, 2, \dots, n)$

这就证明了  $D$  的对角元皆为正数。

$$\text{现设 } D^{1/2} = \text{diag}(\sqrt{d_1}, \sqrt{d_2}, \dots, \sqrt{d_n})$$

注意，在这里我们将  $D^{1/2}$  的对角元素全取为正数，即

$$D = \begin{pmatrix} d_1 & & & \\ & d_2 & & \\ & & \ddots & \\ & & & d_n \end{pmatrix} = \begin{pmatrix} \sqrt{d_1} & & & \\ & \sqrt{d_2} & & \\ & & \ddots & \\ & & & \sqrt{d_n} \end{pmatrix} \begin{pmatrix} \sqrt{d_1} & & & \\ & \sqrt{d_2} & & \\ & & \ddots & \\ & & & \sqrt{d_n} \end{pmatrix}$$

则

$$A = LDL^T = LD^{\frac{1}{2}}D^{\frac{1}{2}}L^T = \left(LD^{\frac{1}{2}}\right)\left(LD^{\frac{1}{2}}\right)^T = \tilde{L} \cdot \tilde{L}^T$$

其中  $\tilde{L} = LD^{\frac{1}{2}}$ ，显然是对角元全为正的 nonsingular 的下三角阵。

由于分解式  $A = LDL^T$  是唯一的，又限定  $D^{\frac{1}{2}}$  的对角元为正数，从而分解  $D = D^{\frac{1}{2}} \cdot D^{\frac{1}{2}}$  也是唯一的，所以说在限定  $L$  的对角线元素皆为正时，三角分解是唯一的。

对称正定矩阵  $A$  的三角分解  $A = \tilde{L}\tilde{L}^T$  称为正定矩阵  $A$  的乔列斯基 (Cholesky) 分解，又称  $LL^T$  分解。

将  $A = \tilde{L}\tilde{L}^T$  记为  $A = LL^T$

那么解线性代数方程组

$$AX = b \Leftrightarrow \text{解 } Ly = b, \quad L^T x = y$$

下面给出用平方根法解线性代数方程组的公式

(1) 对矩阵  $A$  进行 Cholesky 分解，即  $A = LL^T$ ，由矩阵乘法：

对于  $i = 1, 2, \dots, n$  计算

$$l_{ii} = \left( a_{ii} - \sum_{k=1}^{i-1} l_{ik}^2 \right)^{\frac{1}{2}} \quad (3.25)$$

$$l_{ij} = \left( a_{ij} - \sum_{k=1}^{j-1} l_{ik} l_{kj} \right) / l_{jj} \quad j = 1, 2, \dots, i-1 \quad (3.26)$$

(2) 求解下三角形方程组  $LY = b$

$$y_i = \left( b_i - \sum_{k=1}^{i-1} l_{ik} y_k \right) / l_{ii} \quad i = 1, 2, \dots, n \quad (3.27)$$

(3) 求解  $L^T X = y$

$$x_i = \left( y_i - \sum_{k=i+1}^n l_{ki} x_k \right) / l_{ii} \quad (i = n, n-1, \dots, 1) \quad (3.28)$$

由于此法要将矩阵  $A$  作  $LL^T$  三角分解，且在分解过程中含有开方运算，故称该称为

$LL^T$ 分解法或平方根法。

由于  $L^T$  是  $L$  的转置，所以计算量只是一般直接三角分解的一半多一点。另外，由于  $A$  的对称性，计算过程只用到矩阵  $A$  的下三角部分的元素，而且一旦求出  $l_{ij}$  后， $a_{ij}$  就不需要了，所以  $L$  的元素可以存贮在  $A$  的下三角部分相应元素的位置，这样存贮量就大大节省了，在计算机上进行计算时，只需用一维数组  $A[n(n+1)/2]$  对应存放  $A$  的对角线以下部分相应元素。且由

$$a_{ii} = \sum_{k=1}^{i-1} l_{ik}^2$$

$$\text{可知} \quad |l_{ik}| \leq \sqrt{a_{ii}} \quad (k=1, 2, \dots, n \quad i=1, 2, \dots, n)$$

这表明  $L$  的元素的绝对值一般不会很大，所以计算是稳定的，这是 Cholesky 分解的又一个优点。其缺点是需要做一些开方运算。

### 3.5.3 改进平方根法

由于用平方根法解对称正定方程组时，计算  $L$  的对角元素  $l_{ii}$  时需要用到开方运算，为了避免开方运算，我们也可以直接采用对称正定矩阵的  $A = LDL^T$  分解式，即

$$A = \begin{pmatrix} 1 & & & \\ l_{21} & 1 & & \\ l_{31} & l_{32} & 1 & \\ \vdots & & & \\ l_{n1} & \dots & \dots & 1 \end{pmatrix} \begin{pmatrix} d_{11} & & & \\ & d_{22} & & \\ & & \ddots & \\ & & & d_{nn} \end{pmatrix} \begin{pmatrix} 1 & l_{12} & l_{13} & \dots & l_{1n} \\ & 1 & l_{23} & \dots & l_{2n} \\ & & \ddots & & \\ & & & \ddots & \\ & & & & 1 \end{pmatrix}$$

$$= \begin{pmatrix} d_{11} & & & \\ s_{21} & d_{22} & & \\ \vdots & & \ddots & \\ s_{n1} & s_{n2} & \dots & d_{nn} \end{pmatrix} \begin{pmatrix} 1 & l_{21} & \dots & l_{n1} \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{pmatrix}$$

$$\text{其中 } s_{ik} = l_{ik} d_{kk} \quad k < i$$

由矩阵乘法和比较对应元素得

$$\begin{cases} d_{11} = a_{11} \\ \text{对于 } i = 2, 3, \dots, n \\ s_{ij} = a_{ij} - \sum_{k=1}^{j-1} s_{ik} l_{kj} \\ l_{ij} = s_{ij} / d_{jj} \\ d_{ii} = a_{ii} - \sum_{k=1}^{i-1} s_{ik} l_{ik} \end{cases}$$

$d_{ii}, l_{ij}$  的计算应按上列顺序进行

$$\begin{vmatrix} d_{11} \\ l_{21} \\ l_{31} \\ \vdots \\ l_{n1} \end{vmatrix} \begin{vmatrix} d_{22} \\ l_{32} \\ \vdots \\ l_{n2} \end{vmatrix} \vdots \begin{vmatrix} d_{nn} \end{vmatrix}$$

由  $LDL^T$  分解法先求得单位下三角阵  $L$  和对角阵  $D$ , 因为  $A = LDL^T$ , 所以对称方程组

$$AX = b$$

成为

$$LD(L^T X) = b$$

令  $L^T X = y$ , 先解下三角形方程组  $LDY = b$  得

$$y_i = \left( b_i - \sum_{k=1}^{i-1} d_{kk} l_{ik} y_k \right) / d_{ii} \quad (i = 1, 2, \dots, n) \quad (3.29)$$

最后解上三角形方程组  $L^T X = Y$  得

$$x_i = \left( y_i - \sum_{k=i+1}^n l_{ik} x_k \right) \quad (i = n, n-1, \dots, 2, 1) \quad (3.30)$$

$LDL^T$  分解法解对称方程组所含的乘除法运算约为  $n^3/4$  次,  $LL^T$  分解法解对称正定方程组约需乘除法  $n^3/6$  次。 $LDL^T$  法虽然增加了计算量, 但避免了开方运算且扩大了使用范围, 优点是明显的。

例: 用改进平方根法解

$$\begin{pmatrix} 1 & 2 & 1 & -3 \\ 2 & 5 & 0 & -5 \\ 1 & 0 & 14 & 1 \\ -3 & -5 & 1 & 15 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 16 \\ 8 \end{pmatrix}$$

解:  $i = 1 \quad d_{11} = 1$

$$i = 2 \quad S_{21} = a_{21} = 2$$

$$l_{21} = S_{21}/d_{11} = 2/1 = 2$$

$$d_{22} = a_{22} - S_{21}/l_{21} = 5 - 2 \times 2 = 1$$

$$i = 3$$

$$s_{31} = a_{31} - \sum_{k=1}^0 a_{3k} l_{kj} = 1$$



$$s_{32} = a_{32} - \sum_{k=1}^1 a_{3k} l_{k2} = 0 - s_{31} l_{12} = -2$$

$$l_{31} = s_{31} / d_{11} = 1/1 = 1$$

$$l_{32} = -2$$

$$d_{33} = a_{33} - \sum_{k=1}^2 s_{ik} l_{ik} = 9$$

$$i = 2$$

$$s_{41} = -3 \quad s_{42} = 1 \quad s_{43} = 6$$

$$l_{41} = -3 \quad l_{42} = 1 \quad l_{43} = 2/3$$

$$d_{44} = 1$$

求解  $LDY = b$

$$\text{由公式 } y_i = b_i - \sum_{k=1}^{i-1} l_{ki} y_k \quad i = 1, 2, 3, 4$$

$$\text{得} \quad y_1 = b_1 = 1, \quad y_2 = b_2 - l_{12} y_1 = 0 \quad y_3 = 1.6667 \quad y_4 = 1$$

$$\text{再由公式 } x_i = y_i / d_{ii} - \sum_{k=i+1}^n l_{ik} x_k \quad i = 4, 3, 2, 1$$

$$\text{得:} \quad x_4 = 1 \quad x_3 = 1 \quad x_2 = 1 \quad x_1 = 1$$

### 3.6 向量和矩阵的范数

在线性代数方程组的数值解法中，经常需要分析解向量的误差，需要比较误差向量的“大小”或“长度”。那么怎样定义向量的长度呢？我们在初等教学里知道，定义向量的长度，实际上就是对每一个向量按一定的法则规定一个非负实数与之对应，这一思想推广到  $n$  维线性空间里，就是向量的范数或模。

#### 3.6.1 向量的范数

范数的最简单的例子，是绝对值函数。

$$|x| = \sqrt{x^2}$$

并且有三个熟知的性质：

$$(1) \quad x \neq 0 \Rightarrow |x| > 0 \quad |x| = 0 \text{ 当且仅当 } x = 0$$

$$(2) \quad |ax| = |a| \cdot |x| \quad a \text{ 为常数}$$

$$(3) |x+y| \leq |x| + |y|$$

范数的另一个简单例子是二维欧氏空间的长度

$$|OM| = \sqrt{x^2 + y^2} \quad (\text{勾股定理})$$

欧氏范数也满足三个条件:

设  $x = (x_1, x_2)$

$$(1) x \neq 0 \Rightarrow \|x\| > 0$$

$$(2) \|ax\| = |a| \cdot \|x\|_2 \quad a \text{ 为常数}$$

$$(3) \|x+y\|_2 \leq \|x\|_2 + \|y\|_2$$

前两个条件显然, 第三个条件在几何上解释为三角形一边的长度不大于其它两边长度之和。

因此, 称之三角不等式。

下面我们给出  $n$  维空间中向量范数的概念:

设  $X = (x_1, x_2, \dots, x_n)^T$ , 记为  $X \in R^n$

**定义 3.2:** 设  $X \in R^n$ ,  $\|X\|$  表示定义在  $R^n$  上的一个实值函数, 称之为  $X$  的范数, 它具有下列性质:

1) 非负性: 即对一切  $X \in R^n$ ,  $X \neq 0$ ,  $\|X\| > 0$

2) 齐次性: 即为任何实数  $a \in R$ ,  $X \in R^n$ ,

$$\|aX\| = |a| \cdot \|X\|$$

3) 三角不等式: 即对任意两个向量  $X, Y \in R^n$ , 恒有  $\|X+Y\| \leq \|X\| + \|Y\|$

从以上规定范数的三种基本性质、立即可以推出  $R^n$  中向量的范数必具有下列性质:

4)  $\|0\| = 0$

5)  $\| -X \| = |-1| \|X\| = \|X\|$

6) 对任意的  $X, Y \in R^n$ , 恒有:

$$| \|X\| - \|Y\| | \leq \|X - Y\|$$

证明: 根据范数的三角不等式

$$\|X\| = \|(X - Y) + Y\| \leq \|X - Y\| + \|Y\|$$

所以  $\|X\| - \|Y\| \leq \|X - Y\|$

同理可证  $\|Y\| - \|X\| \leq \|X - Y\| = \|Y - X\|$

因此必有： $\|X\| - \|Y\| \leq \|X - Y\|$  证完

三个常用的范数：

设  $X = (x_1, x_2, \dots, x_n)^T$ ，则有

$$(I) \quad \|X\|_1 = |x_1| + |x_2| + \dots + |x_n|;$$

$$(II) \quad \|X\|_2 = \sqrt{X^T X} = \sqrt{x_1^2 + x_2^2 + \dots + x_n^2}$$

$$(III) \quad \|X\|_\infty = \max_{1 \leq i \leq n} |x_i|$$

不难验证，上述三种范数都满足定义的条件。

**定理 3.8：** 定义在  $R^n$  上的向量范数  $\|X\|$  是变量  $X$  分量的一致连续函数。（ $\|X\| = f(x)$ ）

证明：设  $H \in R^n$  为任意向量， $e_1, e_2, \dots, e_n$  为  $R^n$  中的一个基底，且

$$H = h_1 e_1 + h_2 e_2 + \dots + h_n e_n$$

再假设

$$N = \sum_{i=1}^n \|e_i\|$$

显然  $N$  为定常数，则当  $\max_i |h_i| < \delta$  时，由三角不等式得

$$\|H\| = \sum_{i=1}^n \|h_i e_i\| = \sum_{i=1}^n |h_i| \|e_i\| \leq N \max |h_i| < N\delta$$

任给正数  $\varepsilon > 0$ ，取  $\delta = \varepsilon / N$ ，则有：

$$\|X + h\| - \|X\| \leq \|h\| < \varepsilon / \delta = \varepsilon \quad \text{证毕}$$

**定理 3.9：** 在  $R^n$  上定义的任一向量范数  $\|X\|$  都与范数  $\|X\|_1$  等价，即存在正数  $M$  与  $m (M > m)$

对一切  $X \in R^n$ ，不等式

$$m \|X\|_1 \leq \|X\| \leq M \|X\|_1$$

成立。

证明：设  $\xi \in R^n$ ，则  $\xi$  的连续函数  $\|\xi\|$  在有界闭区域  $G\{\xi \mid \|\xi\|_1 = 1\}$ （单位球面）上有界，且一定能达到最大值及最小值。设其最大值为  $M$ ，最小值为  $m$ ，则有

$$m \leq \|\xi\| \leq M \quad \xi \in R^n \quad (3.31)$$

考虑到  $\|\xi\|$  在  $G$  上大于零, 故  $m > 0$

设  $X \in R^n$  为任意非零向量, 则

$$\frac{X}{\|X\|_1} \in G$$

代入 (3.31) 得

$$m \leq \left\| \frac{X}{\|X\|_1} \right\| \leq M$$

所以  $m\|X\|_1 \leq \|X\| \leq M\|X\|_1$  证完

由此定理可得

**推论:**  $R^n$  上定义的任何两个范数都是等价的。对常用范数, 容易验证下列不等式:

$$\frac{1}{n}\|X\|_1 \leq \|X\|_\infty \leq \|X\|_1$$

$$\|X\|_\infty \leq \|X\|_1 \leq n\|X\|_\infty$$

$$\|X\|_\infty \leq \|X\|_2 \leq \sqrt{n}\|X\|_\infty$$

有了范数的概念, 我们就可以讨论向量序列的收敛性问题。

**定义 3.3:** 设给定  $R^n$  中的向量序列  $\{X_k\}$ , 即

$$X_0, X_1, \dots, X_k, \dots,$$

其中

$$X_k = (x_1^{(k)}, x_2^{(k)}, \dots, x_n^{(k)})^T$$

若对任何  $i$  ( $i = 1, 2, \dots, n$ ) 都有

$$\lim_{k \rightarrow \infty} x_i^{(k)} = x_i^*$$

则向量

$$X^* = (x_1^*, \dots, x_n^*)^T$$

称为向量序列  $\{X_k\}$  的极限, 或者说向量序列  $\{X_k\}$  依坐标收敛于向量  $X^*$ , 记为

$$\lim_{k \rightarrow \infty} X_k = X^*$$

**定理 3.10:** 向量序列  $\{X_k\}$  依坐标收敛于  $X^*$  的充分条件是

$$\lim_{k \rightarrow \infty} \|X_k - X^*\| = 0$$

如果一个向量序列  $\{X_k\}$  与向量  $X^*$ ，满足上式，就说向量序列  $\{X_k\}$  依范数收敛于  $X^*$ ，于是便得：

向量序列依范数收敛与依坐标收敛是等价的。

### 3.6.2 矩阵的范数

**定义 3.4:** 设  $A$  为  $n$  阶方阵， $R^n$  中已定义了向量范数  $\|\cdot\|$ ，则称

$$\|A\| = \max_{\substack{x \in R^n \\ x \neq 0}} \frac{\|Ax\|}{\|x\|}$$

为矩阵  $A$  的范数或模，记为  $\|A\|$ 。

$$\|A\| = \max_{\substack{\|x\|=1 \\ x \in R^n}} \|Ax\|$$

矩阵范数有下列基本性质：

(1) 当  $A = 0$  时， $\|A\| = 0$ ，当  $A \neq 0$  时， $\|A\| > 0$

(2) 对任意实数  $k$  和任意  $A$ ，有

$$\|kA\| = |k| \|A\|$$

(3) 对任意两个  $n$  阶矩阵  $A$ 、 $B$  有

$$\|A + B\| \leq \|A\| + \|B\|$$

(4) 对任意向量  $X \in R^n$ ，和任意矩阵  $A$ ，有

$$\|AX\| \leq \|A\| \|X\|$$

(5) 对任意两个  $n$  阶矩阵  $A$ 、 $B$ ，有

$$\|AB\| \leq \|A\| \cdot \|B\|$$

前三个性质可对照向量范数，下面来证明 (4)：设  $G = \{X; \|X\| = 1\}$ ，当  $X \in R^n$  时，根据定义

3.4，(4) 显然成立，当  $X \in R^n$  时，若  $X = 0$ ，则 (4) 成为式，若  $X \neq 0$  时，则  $\frac{X}{\|X\|} \in G$ ，故

$$\left\| A \frac{X}{\|X\|} \right\| \leq \|A\|$$

所以 
$$\frac{1}{\|X\|} \|AX\| \leq \|A\|$$

从而使得 (4) 中的不等式。

对于 (5)，由定义 3 及 (4) 知：

$$\|ABx\| \leq \|A\| \cdot \|Bx\| \leq \|A\| \cdot \|B\| \cdot \|x\|$$

从而得

$$\|AB\| = \max_{\substack{x \in R \\ x \neq 0}} \frac{\|ABx\|}{\|x\|} \leq \|A\| \cdot \|B\|$$

特别地，满足 (4) 的矩阵范数与向量范数，称为相容的，或协调的，(4) 称为相容性条件。

使用矩阵范数与向量范数时，必须满足相容性条件。

与常用向量范数相容的矩阵范数如下：

**定理 3.11：** 设  $n$  阶方阵  $A = (a_{ij})_{n \times n}$ ，则

(I) 与  $\|x\|_1$  相容的矩阵范数是

$$\|A\|_1 = \max_j \sum_{i=1}^n |a_{ij}|$$

(II) 与  $\|x\|_2$  相容的矩阵范数是

$$\|A\|_2 = \sqrt{\lambda_1}$$

其中  $\lambda_1$  为矩阵  $A^T A$  的最大特征值。

(III) 与  $\|x\|_\infty$  相容的矩阵范数是

$$\|A\|_\infty = \max_i \sum_{j=1}^n |a_{ij}|$$

**$A$  的范数  $\|A\|$  与  $A$  的特征值间的关系**

设  $\lambda$  为矩阵  $A$  的任一特征值，向量  $e$  为相应的特征向量，则

$$\lambda e = Ae$$

因为

$$|\lambda| \|e\| = \|Ae\| \leq \|A\| \|e\|$$

所以

$$|\lambda| \leq \|A\|$$

从而得

**定理 3.12:** 矩阵  $A$  的任一特征值的绝对值不超过  $A$  的范数  $\|A\|$ 。

**定义 3.5:** 矩阵  $A$  的诸特征值的最大绝对值称为  $A$  的谱半径，记为：

$$\rho(A) = \max_{1 \leq i \leq n} |\lambda_i|$$

由定理 9 可知：  $\rho(A) \leq \|A\|$

### 3.7 线性方程组的性态和解的误差分析

线性代数方程组  $AX=b$  的系数矩阵  $A$  和右端向量  $b$ ，往往是观测来的，因此它们不可避免地带有误差。这种原始数据的误差对方程组求解的影响如何，是必须探讨的，此即所谓方程组的条件问题。

1. 假设系数矩阵  $A$  精确，且非奇，今讨论右端  $b$  的误差对方程组解的影响。

设  $\delta b$  为  $b$  的误差，而相应的解的误差是  $\delta X$ ，则有

$$A(X + \delta X) = b + \delta b$$

所以

$$\delta X = A^{-1} \delta b$$

$$\|\delta X\| \leq \|A^{-1}\| \cdot \|\delta b\|$$

但  $\|b\| = \|AX\| \leq \|A\| \cdot \|X\|$

所以  $\|\delta X\| \cdot \|b\| \leq \|A^{-1}\| \|\delta b\| \|A\| \cdot \|X\| = \|A\| \|A^{-1}\| \|X\| \|\delta b\|$

当  $b \neq 0, X \neq 0$  时，有

$$\frac{\|\delta X\|}{\|X\|} \leq \|A\| \|A^{-1}\| \frac{\|\delta b\|}{\|b\|}$$

即解  $X$  的相对误差是初始数据  $b$  的相对误差的  $\|A\| \|A^{-1}\|$  倍。

2. 假设右端  $b$  精确，系数矩阵  $A$  有误差，今讨论  $A$  的误差对解的影响。

设矩阵  $A$  的误差为  $\delta A$ ，而相应的解的误差为  $\delta X$ ，则有

$$(A + \delta A)(X + \delta X) = b$$

设  $A$  及  $A + \delta A$  非奇（当  $\|A^{-1} \delta A\| < 1$  时即可），则

$$AX + (\delta A)X + A\delta X + \delta A\delta X = b$$

$$A\delta X = -(\delta A)X - \delta A\delta X$$

$$\delta X = -A^{-1}(\delta A)X - A^{-1}\delta A\delta X$$

根据范数性质

$$\|\delta X\| \leq \|A^{-1}\| \|\delta A\| \|X\| + \|A^{-1}\| \|\delta A\| \|\delta X\|$$

$$(1 - \|A^{-1}\| \|\delta A\|) \|\delta X\| \leq \|A^{-1}\| \|\delta A\| \|X\|$$

于是有

$$\frac{\|\delta X\|}{\|X\|} \leq \frac{\|A^{-1}\| \|\delta A\|}{1 - \|A^{-1}\| \|\delta A\|} = \frac{\|A^{-1}\| \|A\| \frac{\|\delta A\|}{\|A\|}}{1 - \|A^{-1}\| \|A\| \frac{\|\delta A\|}{\|A\|}}$$

若  $\|A^{-1}\| \|A\| \frac{\|\delta A\|}{\|A\|}$  很小, 则  $\|A^{-1}\| \|A\|$  表示相对误差的近似放大率。

由 1, 2 可知,  $b$  及  $A$  有微小改动时, 数  $\|A^{-1}\| \|A\|$  可标志着方程组解  $X$  的敏感程度。解  $X$  的相对误差可能随  $\|A^{-1}\| \|A\|$  的增大而增大。所以系数矩阵  $A$  刻画了线性代数方程组的性态。

**定义 3.6:** 设  $A$  为  $n$  阶非奇矩阵, 称数  $\|A^{-1}\| \|A\|$  为矩阵  $A$  的条件数, 记为  $\text{cond}(A)$ 。

条件数有下列性质是很容易证明的:

- i)  $\text{cond}(A) \geq 1$
- ii)  $\text{cond}(kA) = \text{cond}(A)$   $k$  为非零常数
- iii) 若  $\|A\| = 1$ , 则  $\text{cond}(A) = \|A^{-1}\|$

当  $\text{cond}(A)$  相对地大时, 称方程组  $AX = b$  为病态的, 否则称为良态的。

若方程组为病态的, 则求解过程中的舍入误差对解会有严重的影响。

例如: 对方程组

$$\begin{pmatrix} 1.001 & 0.25 \\ 0.25 & 0.0625 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1.501 \\ 0.375 \end{pmatrix}$$

其解  $X^* = (1.2)^T$ , 但是如果把系数及右端取成近似数, 比如:



$$\begin{pmatrix} 1 & 0.25 \\ 0.25 & 0.063 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1.5 \\ 0.37 \end{pmatrix}$$

则其解为  $\bar{X} = (4, -10)^T$ 。系数及右端绝对误差最大变化为  $\frac{1}{2} \times 10^{-2}$ ，而解的变化却较大。以下看条件数

$$A = \begin{pmatrix} 1.001 & 0.25 \\ 0.25 & 0.0625 \end{pmatrix} \quad A^{-1} = \begin{pmatrix} 1000 & -4000 \\ -4000 & 16016 \end{pmatrix}$$

因为

$$\|A\|_{\infty} = 1.251 \quad \|A^{-1}\|_{\infty} = 20016$$

所以  $\text{cond}(A) = 25040$

表明所给的方程组是病态的。

## 第三章 习题

### 一. 问答题

1. 一个方程组如果是病态的，其系数矩阵有何特点？
2. 求解线性方程组  $A\mathbf{x}=\mathbf{b}$ ，什么时候使用直接法？什么时候使用迭代法？
3. 什么是矩阵的条件数？它的意义是什么？

### 二. 综合题

1. 分别用不选主元消去法和列主元消去法解方程组

$$\begin{cases} 0.001000x_1 + 2.000x_2 + 3.000x_3 = 1.000 \\ -1.000x_1 + 3.712x_2 + 4.623x_3 = 2.000 \\ -2.000x_1 + 1.072x_2 + 5.643x_3 = 3.000 \end{cases}$$

（要求保证四位有效数字），并将结果与含四位有效数字的准确解  $x = (-0.4904, -0.05104, 0.3675)^T$  相比较。

2. 用高斯列主元消去法解方程组

$$\begin{bmatrix} -3 & 2 & 6 \\ 10 & -7 & 0 \\ 5 & -1 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 4 \\ 7 \\ 6 \end{bmatrix}$$

3. 用平方根法解方程组

$$\begin{bmatrix} 4 & 2 & -2 \\ 2 & 2 & -3 \\ -2 & -3 & 14 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 10 \\ 5 \\ 4 \end{bmatrix}$$

3. 证明两个下三角阵之积仍为下三角阵。

4. 设

$$A = \begin{bmatrix} 5 & 7 & 9 & 10 \\ 6 & 8 & 10 & 9 \\ 7 & 10 & 8 & 7 \\ 5 & 7 & 6 & 5 \end{bmatrix}, \quad b = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

用杜利特尔分解  $A = LU$  解方程组  $Ax = b$ 。

5. 用列主元消去法求下列矩阵的逆:

$$(1) \begin{bmatrix} 1 & 1 & -1 \\ 2 & 1 & 0 \\ 1 & 1 & 0 \end{bmatrix}, \quad (2) \begin{bmatrix} 2 & 2 & 3 \\ 1 & -1 & 0 \\ -1 & 2 & 1 \end{bmatrix}$$

6. 用追赶法解方程组

$$\begin{bmatrix} 4 & -1 & & & \\ -1 & 4 & -1 & & \\ & -1 & 4 & -1 & \\ & & -1 & 4 & -1 \\ & & & -1 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} 100 \\ 200 \\ 200 \\ 200 \\ 100 \end{bmatrix}$$

7. 用改进平方根法解下列方程组

$$(1) \begin{bmatrix} 4 & & & & \\ 2.4 & 5.44 & & & \\ 2 & 4 & 5.21 & & \\ 3 & 5.8 & 7.45 & 19.66 & \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 12.280 \\ 16.928 \\ 22.957 \\ 50.945 \end{bmatrix}$$

$$(2) \begin{bmatrix} 10 & 2 & 3 & 1 & 1 \\ 2 & 10 & 1 & 2 & 1 \\ 3 & 1 & 10 & 2 & 3 \\ 1 & 2 & 2 & 10 & 2 \\ 1 & 1 & 3 & 2 & 10 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} 15 \\ 17 \\ 18 \\ 19 \\ 25 \end{bmatrix}$$

8. 求解矩阵方程  $AX=B$ , 其中

$$A = \begin{bmatrix} 1 & 2 & -1 \\ 2 & -1 & 2 \\ -1 & -1 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 4 & 0 \\ 3 & -1 & 1 \\ -2 & -2 & 0 \end{bmatrix}$$

9. 设  $x \in R^n$ , 证明

$$(1) \|x\|_{\infty} \leq \|x\|_2 \leq \sqrt{n} \|x\|_{\infty}$$

$$(2) \|x\|_2 \leq \|x\|_1 \leq \sqrt{n} \|x\|_2$$

$$(3) \|x\|_{\infty} \leq \|x\|_1 \leq n \|x\|_{\infty}$$

10. 设

$$A = \begin{bmatrix} 1 & -1 \\ 2 & 1 \end{bmatrix}, \quad x = \begin{bmatrix} -2 \\ 1 \end{bmatrix}$$

求向量范数  $\|x\|_\infty$ 、 $\|x\|_1$ 、 $\|x\|_2$  及矩阵范数  $\|A\|_\infty$ 、 $\|A\|_1$ 、 $\|A\|_F$  及  $\|A\|_2$ 。

11. 证明

$$\frac{1}{n}\|A\|_\infty \leq \|A\|_1 \leq n\|A\|_\infty$$

12. 设  $A$  可逆, 试证

$$\frac{1}{\|A^{-1}\|_\infty} = \min_{\substack{x \in R^n \\ x \neq 0}} \frac{\|Ax\|_\infty}{\|x\|_\infty}$$

13. 试证明, 如果  $A$  为正交矩阵, 则

$$\text{Cond}(A)_2 = 1$$

14. 设方程组  $Ax = b$  为

$$\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

若  $\delta b = (1, 1)^T$ , 试计算  $\|A\|_2$ ,  $\|A^{-1}\|_2$ ,  $\text{Cond}(A)_\infty$  及解的相对误差  $\|\delta x\|_2 / \|x\|_2$  的界。

15. 设  $A$  是  $n \times n$  阶矩阵, 求证

$$\text{Cond}(A) \geq 1 \quad \text{Cond}(kA) = \text{Cond}(A)$$

16. 线性方程组

$$\begin{pmatrix} 1.001 & 0.25 \\ 0.25 & 0.0625 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1.501 \\ 0.375 \end{pmatrix}$$

有解  $x^* = (1, 2)^T$ , 若将此方程组系数及右端取成近似数

$$\begin{pmatrix} 1 & 0.25 \\ 0.25 & 0.063 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1.5 \\ 0.37 \end{pmatrix}$$

求其解  $\bar{x} = (\bar{x}_1, \bar{x}_2)^T$ , 这两组不同的解说明了什么问题?

17. 设

$$A = \begin{bmatrix} 2.0001 & -1 \\ -2 & 1 \end{bmatrix}, \quad b = \begin{bmatrix} 7.0003 \\ -7 \end{bmatrix}$$

已知方程组  $AX=B$  的精确解为  $X^* = (3, -1)^T$ 。

(1) 计算条件数  $\text{cond}_\infty(A)$ ;

(2) 若近似解  $\bar{X} = (2.97, -1.01)^T$ , 计算剩余  $r = b - A\bar{X}$ ;

(3) 计算解的相对误差估计式

$$\frac{\|X^* - \bar{X}\|_{\infty}}{\|X^*\|_{\infty}} \leq \text{Cond}_{\infty}(A) \frac{\|r\|_{\infty}}{\|b\|_{\infty}}$$

的右端，并与不等式左端比较，此结果说明什么？

18. 线性方程组

$$\begin{cases} 7x_1 + 10x_2 = 1 \\ 5x_1 + 7x_2 = 0.7 \end{cases}$$

(1) 试求出系数矩阵  $A$  的条件数  $\text{Cond}_{\infty}(A)$

(2) 若右端向量有扰动  $\delta(b) = (0, 0.1, -0.01)^T$ ，试估计解的相对误差

19. 若  $n$  阶矩阵  $A$  可逆，证明

$$\frac{1}{n} \leq \frac{\text{Cond}(A)_{\infty}}{\text{Cond}(A)_2} \leq n$$

20. 设  $A$  是  $n \times n$  阶矩阵，求证

$$\|A^{-1}\| \geq \frac{1}{\|A\|}$$

21. 设  $A$  是  $n \times n$  阶矩阵，求证

$$\text{Cond}(kA) = \text{Cond}(A) \quad k \text{ 是常数}$$

22. 设  $A$  非奇矩阵，方程组  $AX = B$  系数矩阵有扰动  $\delta A = \alpha A (\alpha \neq -1, \text{实数})$ ，试证解的相对误差

$$\frac{\|\delta X\|}{\|X\|} = \frac{|\alpha|}{|1 + \alpha|}$$

23. 设  $A$ 、 $B$  为  $n$  阶非奇异实矩阵，求证

$$\text{Cond}(AB) \leq \text{Cond}(A) \cdot \text{Cond}(B)$$

24. 设  $A$  为正交矩阵，求  $\text{cond}(A)_2$ ，并由此判断  $A$  的性态。

## 上机计算题

一、编制解  $Ax = b$  的通用子程序

(1) 列主元消去法；

(2) 改进平方根法；

(3) 实现矩阵三角分解的杜利特尔及乔利斯基方法及用此方法解  $Ax = b$  的过程。

二、用列主元消去法程序第一题中 (1) 解方程组

$$\begin{bmatrix} 1.1348 & 3.8326 & 1.1651 & 3.4017 \\ 0.5301 & 1.7875 & 2.5330 & 1.5435 \\ 3.4129 & 4.9317 & 8.7643 & 1.3142 \\ 1.2371 & 4.9998 & 10.6721 & 0.0147 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 9.5342 \\ 6.3941 \\ 18.4231 \\ 16.9237 \end{bmatrix}$$

并比较计算结果精度（准确解为  $x_1 = x_2 = x_3 = x_4 = 1$ ）

三、用平方根法解线法方程组习题 9 中（1）。

四、用改进平方根法程序（第一题中（2））求解第三题中方程组。

五、利用矩阵的杜利特尔三角分解  $A = LU$  及  $Ly = b$ ,  $Ux = y$  的解方程组  $Ax = b$  的方法解方程组

$$Ax = b_k \quad (k = 1, 2, \dots, 10)$$

其中

$$A = \begin{bmatrix} 1 & 2 & 4 & 7 & 11 & 16 \\ 2 & 3 & 5 & 8 & 12 & 17 \\ 4 & 5 & 6 & 9 & 13 & 18 \\ 7 & 8 & 9 & 10 & 14 & 19 \\ 11 & 12 & 13 & 14 & 15 & 20 \\ 16 & 17 & 18 & 19 & 20 & 21 \end{bmatrix}_{4 \times 4}$$

$b_1$  为任一非零的六元向量；若记  $x_k$  为  $Ax_k = b_k$  的解向量，则取  $b_{k+1} = \frac{x_k}{\|x_k\|_\infty}$ ，请输出矩阵  $L$ 、 $U$  及

$b_k$ ,  $x_k$  ( $k = 1, 2, \dots, 10$ ) 并观察之结果。

## 第四章 解线性方程组的迭代法

对于阶数不高的方程组,直接法非常有效,对于阶数高,而系数矩阵稀疏的线性方程组却存在着困难,在这类矩阵中,非零元素较少,若用直接法求解,就要存贮大量零元素。为减少运算量、节约内存,使用迭代法更有利。本章介绍迭代法的初步内容。

### §1 雅克比法、赛得尔法、超松弛法

#### 1. 雅克比 (Jacobi) 迭代法

设有  $n$  阶方程组

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \\ \cdots \cdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n = b_n \end{cases} \quad (4.1)$$

若系数矩阵非奇异,且  $a_{ii} \neq 0$  ( $i = 1, 2, \cdots, n$ ), 将方程组 (4.1) 改写成

$$\begin{cases} x_1 = \frac{1}{a_{11}}(b_1 - a_{12}x_2 - a_{13}x_3 - \cdots - a_{1n}x_n) \\ x_2 = \frac{1}{a_{22}}(b_2 - a_{21}x_1 - a_{23}x_3 - \cdots - a_{2n}x_n) \\ \cdots \cdots \\ x_n = \frac{1}{a_{nn}}(b_n - a_{n1}x_1 - a_{n2}x_2 - \cdots - a_{n,n-1}x_{n-1}) \end{cases}$$

然后写成迭代格式

$$\begin{cases} x_1^{(k+1)} = \frac{1}{a_{11}}(b_1 - a_{12}x_2^{(k)} - a_{13}x_3^{(k)} - \cdots - a_{1n}x_n^{(k)}) \\ x_2^{(k+1)} = \frac{1}{a_{22}}(b_2 - a_{21}x_1^{(k)} - a_{23}x_3^{(k)} - \cdots - a_{2n}x_n^{(k)}) \\ \cdots \cdots \\ x_n^{(k+1)} = \frac{1}{a_{nn}}(b_n - a_{n1}x_1^{(k)} - a_{n2}x_2^{(k)} - \cdots - a_{n,n-1}x_{n-1}^{(k)}) \end{cases} \quad (4.2)$$

(4.2) 式也可以简单地写为

$$x_i^{(k+1)} = \frac{1}{a_{ii}} \left( b_i - \sum_{\substack{j=1 \\ j \neq i}}^n a_{ij}x_j^{(k)} \right) \quad (i = 1, 2, \cdots, n) \quad (4.3)$$

对 (4.2) 或 (4.3) 给定一组初值  $X^{(0)} = (x_1^{(0)}, x_2^{(0)}, \cdots, x_n^{(0)})^T$  后, 经反复迭代可得到一向量序列  $X^{(k)} = (x_1^{(k)}, \cdots, x_n^{(k)})^T$ , 如果  $X^{(k)}$  收敛于  $X^* = (x_1^*, x_2^*, \cdots, x_n^*)^T$ , 则  $x_i^* (i = 1, 2, \cdots, n)$  就是方程组 (4.1) 的解。这一方法称为雅克比 (Jacobi) 迭代法或简单迭代法, (4.2) 或 (4.3)

称为 Jacobi 迭代格式。

下面介绍迭代格式的矩阵表示：

设  $D = \text{diag}(a_{11}, a_{22}, \dots, a_{nn})$ ，将方程组  $AX = b$  中的系数矩阵表示成三个特殊矩阵的代数和矩阵：

$$A = D - L - U$$

其中  $L = \begin{pmatrix} 0 & & & \\ -a_{21} & 0 & & \\ -a_{31} & -a_{32} & & \\ \vdots & \vdots & \ddots & \\ -a_{n1} & -a_{n2} & & 0 \end{pmatrix} \quad U = \begin{pmatrix} 0 & -a_{12} & \cdots & -a_{1n} \\ & 0 & \cdots & -a_{2n} \\ & & \ddots & \\ & & & 0 \end{pmatrix}$

由于  $a_{ii} \neq 0$  ( $i = 1, 2, \dots, n$ )，D 为可逆对角阵，L、U 分别为严格上、下三角阵，于是

$$Ax = b \Leftrightarrow (D - L - U)x = b \Leftrightarrow Dx = (L + U)x + b$$

利用 D 可逆，得到等价方程组

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

则迭代格式的向量表示为

$$X^{(k+1)} = B_J X^{(k)} + f_J$$

$$B_J = D^{-1}(L + U), \quad f_J = D^{-1}b$$

$B_J$  称为雅克比迭代矩阵。

## 2. 高斯——赛得尔 (Gauss-Seidel) 迭代法

显然，如果迭代收敛， $x_i^{(k+1)}$  应该比  $x_i^{(k)}$  更接近于原方程的解  $x_i^*$  ( $i = 1, 2, \dots, n$ )，因此

在迭代过程中及时地以  $x_i^{(k+1)}$  代替  $x_i^{(k)}$  ( $i = 1, 2, \dots, n-1$ )，可望收到更好的效果。这样 (4.2)

式可写成：

$$\begin{cases} x_1^{(k+1)} = \frac{1}{a_{11}}(b_1 - a_{12}x_2^{(k)} - a_{13}x_3^{(k)} - \cdots - a_{1n}x_n^{(k)}) \\ x_2^{(k+1)} = \frac{1}{a_{22}}(b_2 - a_{21}x_1^{(k+1)} - a_{23}x_3^{(k)} - \cdots - a_{2n}x_n^{(k)}) \\ \dots\dots\dots \\ x_n^{(k+1)} = \frac{1}{a_{nn}}(b_n - a_{n1}x_1^{(k+1)} - a_{n2}x_2^{(k+1)} - \cdots - a_{n,n-1}x_{n-1}^{(k+1)}) \end{cases} \quad (4.5)$$

(4.5) 式可简写成

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

此为 G-S 迭代格式。

$G-S$  迭代格式的矩阵表示:

$$Ax = b \Leftrightarrow (D - L - U)x = b \Leftrightarrow (D - L)x = Ux + b$$

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

$$x^{(k+1)} = B_S x^{(k)} + f_S \quad (4.6)$$

$$B_S = (D - L)^{-1}U, \quad f_S = (D - L)^{-1}b$$

$B_S$  称为高斯-赛德尔迭代矩阵。

关于上述迭代法的误差控制, 可按类似于第二章非线性方程求根的迭代法处理, 设  $\varepsilon$  为允许的绝对误差限, 可以检验

$$\max_{1 \leq i \leq n} |x_i^{(k+1)} - x_i^{(k)}| < \varepsilon$$

是否成立, 以决定计算是否终止, 进一步的讨论稍后进行。

实际计算时, 如果线性方程组的阶数不高, 建立迭代格式也可以不从矩阵形式出发, 以避免求逆矩阵的计算。

### 3. 超松弛法

使用迭代法的困难是计算量难以估计, 有些方程组的迭代格式虽然收敛, 但收敛速度慢而使计算量变得很大。

松弛法是一种线性加速方法。这种方法将前一步的结果  $x_i^{(k)}$  与高斯——赛得尔方法的迭代值  $\tilde{x}_i^{(k+1)}$  适当进行线性组合, 以构成一个收敛速度较快的近似解序列。改进后的迭代方案是:

迭代

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

加速

$$x_i^{(k+1)} = (1 - \omega)x_i^{(k)} + \omega \tilde{x}_i^{(k+1)} \quad (i = 1, 2, \dots, n)$$

所以

$$x_i^{(k+1)} = (1 - \omega)x_i^{(k)} + \frac{\omega}{a_{ii}} \left( b_i - \sum_{j=1}^{i-1} a_{ij} x_j^{(k+1)} - \sum_{j=i+1}^n a_{ij} x_j^{(k)} \right) \quad (4.7)$$

这种加速法就是松弛法。其中系数  $\omega$  称松弛因子。可以证明, 要保证迭代格式 (4.7) 收敛必须要求  $0 < \omega < 2$

当  $\omega = 1$  时, 即为高斯——赛得尔迭代法, 为使收敛速度加快, 通常取  $\omega > 1$ , 即为超松弛法。

松弛因子的选取对迭代格式 (4.7) 的收敛速度影响极大。实际计算时, 可以根据系数矩阵的性质, 结合经验通过反复计算来确定松弛因子  $\omega$ 。



## §2 迭代法的收敛条件

由 §1 中迭代格式的矩阵形式知, 方程组  $AX = b$  的雅克比迭代法、高斯——赛得尔迭代法和松弛法的矩阵形式都可以写成下式:

$$X^{(k+1)} = BX^k + F \quad (4.8)$$

当然, 不同的迭代法其迭代矩阵  $B$  和  $F$  的元素是不同的。所以我们讨论迭代格式 (4.8) 的收敛性, 就具有普遍意义。

下面, 我们不加证明地给出迭代格式 (4.8) 收敛的充分必要条件。

**定理 1:** 对任意初始向量  $X^{(0)}$  及常向量  $F$ , 迭代格式 (4.8) 收敛的充分必要条件是迭代矩阵  $B$  的谱半径  $\rho(B) < 1$ 。

这一结论在理论上是颇为重要的, 但实际用起来不甚方便, 为此我们着重研究更为实用的判别迭代格式收敛的充分条件。

考虑迭代向量序列  $\{X^{(k)}\}$  的收敛问题:

若 
$$\lim_{k \rightarrow \infty} X^{(k)} = X^*$$

$$X^* = BX^* + F$$

于是

$$\begin{aligned} X^{(k)} - X^* &= B(X^{(k-1)} - X^*) = B^2(X^{(k-2)} - X^*) \\ &= \cdots = B^k(X^{(0)} - X^*) \end{aligned}$$

收敛的意思是:

$$\lim_{k \rightarrow \infty} (X^{(k)} - X^*) = \lim_{k \rightarrow \infty} B^k (X^{(0)} - X^*) = 0$$

依范数收敛是

$$\|X^{(k)} - X^*\| \leq \|B\|^k \|X^{(0)} - X^*\| \rightarrow 0 \quad \text{当 } k \rightarrow \infty, \text{ 从而得以下定理:}$$

**定理 2:** 若迭代矩阵  $B$  的某种范数  $\|B\| < 1$  则 (4.8) 确定的迭代法对任意初值  $X^{(0)}$  均收敛于方程组  $X = BX + F$  的唯一解  $x^*$ 。

下面给出直接计算  $x_i^{(k+1)}$  时的收敛性定理。为给出这个定理, 先介绍对角占优的概念。

**定义 1:** 如果矩阵的每一行中, 不在主对角线上的所有元素绝对值之和小于主对角线上元素的绝对值, 即

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

则称矩阵  $A$  按行严格对角占优, 类似地, 也有按列严格对角占优。

**定理 3:** 若线性方程组  $AX = b$  的系数矩阵  $A$  按行严格对角占优, 则雅克比迭代法和高斯——赛得尔迭代法对任意给定初值均收敛。

证明: 记 
$$e_k = \max_{1 \leq j \leq n} |x_j^{(k)} - x_j^*|$$

为第  $k$  次近似值  $(x_1^{(k)}, x_2^{(k)}, \dots, x_n^{(k)})^T$  的误差,

(1) 由雅克比迭代法

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

记

$$L = \sum_{\substack{j=1 \\ j \neq i}}^n \frac{|a_{ij}|}{|a_{ii}|}$$

则有

$$\left| x_i^{(k+1)} - x_i^* \right| \leq L \cdot \max_{1 \leq j \leq n} |x_j^{(k)} - x_j^*|$$

上式对  $i = 1, 2, \dots, n$  成立, 故有

$$e_{k+1} \leq L e_k \leq \dots \leq L^{k+1} e_0$$

因为  $A$  严格对角占优, 故  $L < 1$ , 从而有

$$\lim_{k \rightarrow \infty} |x_i^{(k+1)} - x_i^*| \leq \lim_{k \rightarrow \infty} L^{k+1} |x_i^{(0)} - x_i^*| = 0$$

即雅克比方法收敛。

(2) 高斯——赛得尔迭代法

考虑高斯——赛得尔方法的误差

$$\begin{aligned} \left| x_i^{(k+1)} - x_i^* \right| &= \frac{\sum_{j=1}^{i-1} a_{ij} (x_j^{(k+1)} - x_j^*) + \sum_{j=i+1}^n |a_{ij}| |x_j^{(k)} - x_j^*|}{|a_{ii}|} \\ &\leq \frac{\sum_{j=1}^{i-1} |a_{ij}| |x_j^{(k+1)} - x_j^*|}{|a_{ii}|} + \frac{\sum_{j=i+1}^n |a_{ij}| |x_j^{(k)} - x_j^*|}{|a_{ii}|} \end{aligned}$$

记

$$L = \sum_{j=1}^{i-1} \frac{|a_{ij}|}{|a_{ii}|} \quad S = \sum_{j=i+1}^n \frac{|a_{ij}|}{|a_{ii}|}$$

则

$$\left| x_i^{(k+1)} - x_i^* \right| \leq L \cdot \max_{1 \leq j \leq i-1} |x_j^{(k+1)} - x_j^*| + S \cdot \max_{i+1 \leq j \leq n} |x_j^{(k)} - x_j^*|$$

从而

$$e_{k+1} \leq L e_{k+1} + S e_k$$

$$e_{k+1} \leq \frac{S}{1-L} e_k \leq \dots \leq \left( \frac{S}{1-L} \right)^{k+1} \cdot e_0$$

而

$$\left( \frac{S}{1-L} \right) = \frac{S-L+L}{1-L} = \frac{S+L}{1-L} - \frac{L}{1-L} < \frac{1}{1-L} - \frac{L}{1-L} = 1$$

所以

$$\lim_{k \rightarrow \infty} |x_i^{(k+1)} - x_i^*| = 0$$

即：高斯——赛得尔迭代法收敛。 证完

例：用雅克比迭代法和高斯——赛得尔迭代法解线性方程组

$$\begin{pmatrix} 9 & -1 & -1 \\ -1 & 8 & 0 \\ -1 & 0 & 9 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 7 \\ 7 \\ 8 \end{pmatrix}$$

解：所给线性方程组的系数矩阵按行严格对角占优，故雅克比迭代法和高斯——赛得尔迭代法都收敛。

$$D = \text{diag}(9, 8, 9) \quad D^{-1} = \text{diag}(1/9, 1/8, 1/9)$$

$$I - D^{-1}A = \begin{pmatrix} 0 & 1/9 & 1/9 \\ 1/8 & 0 & 0 \\ 1/9 & 0 & 0 \end{pmatrix} \quad D^{-1}b = \begin{pmatrix} 7/9 \\ 7/8 \\ 7/9 \end{pmatrix}$$

雅克比迭代法的迭代公式为：

$$X^{(k+1)} = \begin{pmatrix} 0 & 1/9 & 1/9 \\ 1/8 & 0 & 0 \\ 1/9 & 0 & 0 \end{pmatrix} X^{(k)} + \begin{pmatrix} 7/9 \\ 7/8 \\ 7/9 \end{pmatrix}$$

取  $X^{(0)} = (0, 0, 0)^T$ ，由上述公式得逐次近似值如下：

$k$	0	1	2	3	4
$X^{(i)}$	$\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$	$\begin{pmatrix} 0.7778 \\ 0.8750 \\ 0.8889 \end{pmatrix}$	$\begin{pmatrix} 0.9738 \\ 0.9723 \\ 0.9753 \end{pmatrix}$	$\begin{pmatrix} 0.9942 \\ 0.9993 \\ 0.9993 \end{pmatrix}$	$\begin{pmatrix} 0.9993 \\ 0.9993 \\ 0.9993 \end{pmatrix}$

高斯——赛得尔迭代法：

$$\begin{cases} x_1^{(k+1)} = \frac{1}{9}(x_2^{(k)} + x_3^{(k)} + 7) \\ x_2^{(k+1)} = \frac{1}{8}(x_1^{(k+1)} + x_3^{(k)} + 7) \\ x_3^{(k+1)} = \frac{1}{9}(x_1^{(k+1)} + 0 \cdot x_2^{(k+1)} + 8) \end{cases}$$

迭代结果为：

$k$	0	1	2	3	4
$x^{(i)}$	$\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$	$\begin{pmatrix} 0.7778 \\ 0.9722 \\ 0.9753 \end{pmatrix}$	$\begin{pmatrix} 0.9942 \\ 0.9993 \\ 0.9993 \end{pmatrix}$	$\begin{pmatrix} 0.9998 \\ 1.0000 \\ 1.0000 \end{pmatrix}$	$\begin{pmatrix} 1.000 \\ 1.000 \\ 1.000 \end{pmatrix}$

如果矩阵  $A$  严格对角占优，那么高斯——赛得尔迭代法的收敛速度快于雅克比迭代法的收敛速度。

以上定理 2、3 只是雅克比迭代法和高斯——赛得尔迭代法收敛的充分条件，对于一个给定的系数矩阵  $A$ ，两种方法可能都收敛，也可能都不收敛；还可能是雅克比方法收敛而高斯——赛得尔方法不收敛；亦或相反。在计算机上，高斯——赛得尔方法只需要一套存放迭

代向量的单元，而雅克比方法都需两套。

### § 3 迭代法的误差估计

在 § 1 中曾以检验

$$\max_{1 \leq i \leq n} |x_i^{(k+1)} - x_i^{(k)}| < \varepsilon$$

是否成立的办法来估计误差并确定迭代是否终止。它的理论依据是：

**定理 4：** 设  $X^*$  是方程组  $AX = b$  的同解方程  $X = BX + F$  的准确解，若迭代公式  $X^{(k+1)} = BX^{(k)} + F$  中迭代矩阵  $B$  的某种范数  $\|B\| = q < 1$ ，则有

$$1) \quad \|X^{(k)} - X^*\| \leq \frac{q}{1-q} \|X^{(k)} - X^{(k-1)}\|$$

$$2) \quad \|X^{(k)} - X^*\| \leq \frac{q^K}{1-q} \|X^{(1)} - X^{(0)}\|$$

证明：先证 1) 因为

$$X^{(k+1)} = BX^{(k)} + F \quad (4.9)$$

$$X^{(k)} = BX^{(k-1)} + F \quad (4.10)$$

由 (4.9)、(4.10) 相减得

$$\|X^{(k+1)} - X^{(k)}\| \leq q \|X^{(k)} - X^{(k-1)}\| \quad (4.11)$$

因为  $X^* = BX^* + F$ ，故

$$\|X^{(k+1)} - X^*\| \leq q \|X^{(k)} - X^*\|$$

另一方面，

$$\begin{aligned} \|X^{(k+1)} - X^{(k)}\| &= \|(X^{(k)} - X^*) - (X^{(k+1)} - X^*)\| \\ &\geq \|X^{(k)} - X^*\| - \|X^{(k+1)} - X^*\| \geq \|X^{(k)} - X^*\| - q \|X^{(k)} - X^*\| \\ &= (1-q) \|X^{(k)} - X^*\| \end{aligned}$$

再由 (4.11) 便得：

$$\|X^{(k)} - X^*\| \leq \frac{q}{1-q} \|X^{(k)} - X^{(k-1)}\| \quad (4.12)$$

反复运用 (4.11) 可得

$$\|X^{(k)} - X^{(k-1)}\| \leq q^{k-1} \|X^{(1)} - X^{(0)}\| \quad (4.13)$$

将 (4.13) 代入 (4.12) 即有

$$\|X^{(k)} - X^*\| \leq \frac{q^k}{1-q} \|X^{(1)} - X^{(0)}\|$$

## 第四章 习题

1. 什么是求解线性方程组的直接法和迭代法？这两种方法的特点是什么？如果解一个实际问题，你根据什么决定用直接法还是迭代法？

2. 设方程组  $Ax = b$  对应的迭代格式为

$$\begin{cases} x_1^{(k+1)} = 0.5x_1^{(k)} + \omega(4 - x_2^{(k)}) \\ x_2^{(k+1)} = 0.5x_2^{(k)} + \omega(3 - x_1^{(k)} - x_3^{(k)}) \\ x_3^{(k+1)} = 0.5x_3^{(k)} + \omega(2 - x_2^{(k)}) \end{cases} \quad (k = 0, 1, \dots)$$

讨论此迭代格式当参数  $\omega$  取何值时收敛，取何值时发散。

3. 已知线性方程组的系数阵  $A$  为

$$(1) \quad A = -\begin{bmatrix} 1 & 0 & 1 \\ -1 & 1 & 0 \\ 1 & 2 & -3 \end{bmatrix}, \quad (2) \quad A = \begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 1 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & 1 \end{bmatrix}$$

证明关于矩阵(1)雅可比迭代法收敛，高斯-塞德尔迭代法不收敛；关于矩阵(2)高斯-塞德尔迭代法收敛，雅可比迭代法不收敛。

4. 对方程组

$$\begin{cases} kx_1 + x_2 = 1 \\ x_1 + kx_2 + x_3 = 2 \\ x_2 + kx_3 = 3 \end{cases} \quad k \neq 0$$

1° 写出相应的雅可比和高斯-塞德尔迭代格式；

2° 关于参数  $k$ ，讨论 1° 中的两种格式的敛散性。

5. 设有方程组

$$\begin{cases} 3x_1 - 10x_2 = -7 \\ 9x_1 - 4x_2 = 5 \end{cases}$$

(1). 问用 Jacobi 迭代法和 Gauss-Seidel 迭代法求解此方程组是否收敛？

(2). 若把上述方程组交换次序得到新的方程组，再用 Jacobi 迭代法和 Gauss-Seidel 迭代法求解此方程组是否收敛？

(3). 用一个收敛的格式求解此方程组。

6. 对方程组  $Ax = b$

$$\begin{bmatrix} 4 & -1 & 0 & -1 & 0 & 0 \\ -1 & 4 & -1 & 0 & -1 & 0 \\ 0 & -1 & 4 & 0 & 0 & -1 \\ -1 & 0 & 0 & 4 & -1 & 0 \\ 0 & -1 & 0 & -1 & 4 & -1 \\ 0 & 0 & -1 & 0 & -1 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{bmatrix} = \begin{bmatrix} 0 \\ 5 \\ 0 \\ 6 \\ -2 \\ 6 \end{bmatrix}$$

其精确解为  $x^* = (1, 2, 1, 2, 1, 2)^T$ 。写出解  $Ax = b$  的雅可比、高斯-塞贝尔迭代格式，并讨论其敛散性。

7. 设给定方程组  $Ax=b$  的雅可比迭代阵  $B_J = D^{-1}(L + U)$  为

$$B_J = \begin{bmatrix} 0 & -2 & 2 \\ -1 & 0 & -1 \\ -2 & -2 & 0 \end{bmatrix}$$

试证明解  $Ax = b$  的雅可方法收敛，但相应的高斯-塞德尔方法发散。

8. 设方程组  $Ax = b$  中系数矩阵  $A$  为

$$A = \begin{bmatrix} 1 & a & a \\ a & 1 & a \\ a & a & 1 \end{bmatrix}$$

证当  $-\frac{1}{2} < a < 1$  成立时， $A$  为对称正定的，但解  $Ax = b$  的雅可比方法只对  $-\frac{1}{2} < a < \frac{1}{2}$  是收敛的。

9. 为求解方程组

$$\begin{cases} -x_1 + 8x_2 = 7 \\ 9x_1 - x_2 - x_3 = 7 \\ -x_1 + 9x_3 = 8 \end{cases}$$

试写出收敛的迭代公式，并说明收敛的原因。解此方程组。  $x^{(0)} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$

$$\varepsilon = 10^{-4}$$

10. 设方程组

$$\begin{cases} 5x_1 + 2x_2 + x_3 = -12 \\ 2x_1 - 3x_2 + 10x_3 = 3 \\ -x_1 + 4x_2 + 2x_3 = 20 \end{cases}$$

试构造一个收敛的迭代格式，并求解此方程。  $\|\varepsilon\|_\infty \leq 10^{-3}$ ,  $x^{(0)} = (0, 0, 0)^T$

11. 用 SOR 方法解方程组

$$\begin{cases} 4x_1 - x_2 = 1 \\ -x_1 + 4x_2 - x_3 = 4 \\ -x_2 + 4x_3 = -3 \end{cases}$$

松弛因子分别取为  $\omega = 1.03$  和  $\omega = 1.1$ , 精确解为  $x^* = \left(\frac{1}{2}, 1, -\frac{1}{2}\right)^T$ 。要求当

$\|x^* - x^{(k)}\|_\infty < 0.5 \times 10^{-6}$  时迭代终止, 并且对每一个  $\omega$  值确定出迭代次数。

12. 设  $X^*$  是方程组  $AX = b$  的同解方程  $X = BX + F$  的准确解, 若迭代公式  $X^{(k+1)} = BX^{(k)} + F$  中迭代矩阵  $B$  的某种范数  $\|B\| < 1$ , 则有

$$\|X^{(k)} - X^*\| \leq \frac{\|B\|^k}{1 - \|B\|} \|X^{(1)} - X^{(0)}\|$$

#### 第四章 上机实习题

一、编制通用子程序:

- (1) 雅可比迭代格式;
- (2) 高斯-塞德尔迭代格式;
- (3) SOR 方法格式;
- (4) 变带宽平方根法解大型稀疏方程组格式。

二、对习题 15 中的方程组  $Ax = b$ , 取初值  $x^{(0)} = (1, 1, 1, 1, 1)^T$ , 要求  $\|x^{(k+1)} - x^{(k)}\|_2 \leq 10^{-5}$

- (1) 用雅可比方法计算
- (2) 用高斯-塞德尔方法计算;
- (3) 用 SOR 迭代法计算 ( $\omega = 1.334, 1.95, 0.95$ )

最后输出近似解及迭代次数  $k$ 。

## 第五章 矩阵特征问题的求解

### 5.1 引言

在科学技术的应用领域中,许多问题都归为求解一个特征系统。如动力学系统和结构系统中的振动问题,求系统的频率与振型;物理学中的某些临界值的确定等等。

设  $A$  为  $n$  阶方阵,  $A = (a_{ij}) \in R^{n \times n}$ , 若  $x \in R^n (x \neq 0)$ , 有数  $\lambda$  使

$$Ax = \lambda x \quad (5.1)$$

则称  $\lambda$  为  $A$  的特征值,  $x$  为相应于  $\lambda$  的特征向量。因此,特征问题的求解包括两方面:

1. 求特征值  $\lambda$ , 满足

$$\varphi(\lambda) = \det(A - \lambda I) = 0 \quad (5.2)$$

2. 求特征向量  $x \in R^n (x \neq 0)$ , 满足齐方程组

$$(A - \lambda I)x = 0 \quad (5.3)$$

称  $\varphi(\lambda)$  为  $A$  的特征多项式,它是关于  $\lambda$  的  $n$  次代数方程。

关于矩阵的特征值,有下列代数理论,

**定义 1** 设矩阵  $A, B \in R^{n \times n}$ , 若有可逆阵  $P$ , 使

$$B = P^{-1}AP$$

则称  $A$  与  $B$  相似。

**定理 1** 若矩阵  $A, B \in R^{n \times n}$  且相似, 则

(1)  $A$  与  $B$  的特征值完全相同;

(2) 若  $x$  是  $B$  的特征向量, 则  $Px$  便为  $A$  的特征向量。

**定理 2** 设  $A \in R^{n \times n}$  具有完全的特征向量系, 即存在  $n$  个线性无关的特征向量构成  $R^n$  的一组基底, 则经相似变换可化  $A$  为对角阵, 即有可逆阵  $P$ , 使

$$P^{-1}AP = D = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix}$$

其中  $\lambda_i$  为  $A$  的特征值,  $P$  的各列为相应于  $\lambda_i$  的特征向量。

**定理 3**  $A \in R^{n \times n}$ ,  $\lambda_1, \dots, \lambda_n$  为  $A$  的特征值, 则

(1)  $A$  的迹数等于特征值之积, 即

$$\text{tr}(A) \equiv \sum_{i=1}^n a_{ii} = \sum_{i=1}^n \lambda_i$$

(2)  $A$  的行列式值等于全体特征值之积, 即

$$\det(A) = \lambda_1 \lambda_2 \cdots \lambda_n$$

**定理 4** 设  $A \in R^{n \times n}$  为对称矩阵, 其特征值  $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ , 则

(1) 对任  $A \in R^n$ ,  $x \neq 0$ ,



$$\lambda_n \leq \frac{(Ax, x)}{(x, x)} \leq \lambda_1$$

$$(2) \lambda_n = \min_{x \neq 0} \frac{(Ax, x)}{(x, x)}$$

$$(3) \lambda_1 = \max_{x \neq 0} \frac{(Ax, x)}{(x, x)}$$

**定理 5** (Gerschgorin 圆盘定理) 设  $A \in R^{n \times n}$ , 则

(1)  $A$  的每一个特征值必属于下述某个圆盘之中,

$$|z - a_{ii}| \leq \sum_{\substack{j=1 \\ j \neq i}}^n |a_{ij}|, \quad i = 1, 2, \dots, n \quad (5.4)$$

(5.4) 式表示以  $a_{ii}$  为中心, 以半径为  $\sum_{\substack{j=1 \\ j \neq i}}^n |a_{ij}|$  的复平面上的  $n$  个圆盘。

(2) 如果矩阵  $A$  的  $m$  个圆盘组成的并集  $S$  (连通的) 与其余  $n - m$  个圆盘不连接, 则  $S$  内恰包含  $m$  个  $A$  的特征值。

定理 4 及定理 5 给出了矩阵特征值的估计方法及界。

例 1 设有

$$A = \begin{bmatrix} 4 & 1 & 0 \\ 1 & 0 & -1 \\ 1 & 1 & -4 \end{bmatrix}$$

估计  $A$  的特征值的范围。

解 由圆盘定理,  $A$  的 3 个圆盘为

图 5.1

$$D_1: |z - 4| \leq 1$$

$$D_2: |z - 0| \leq 2$$

$$D_3: |z + 4| \leq 2$$

见图 5.1。

$D_2$  为孤立圆盘且包含  $A$  的一个实特征值  $\lambda_1$  (因为虚根成对出现的原理), 则  $3 \leq \lambda_1 \leq 5$ 。

而  $\lambda_2, \lambda_3 \in D_1 \cup D_2$ , 则  $\rho(A) = \max |\lambda_i| \leq 6$ , 即

$$3 \leq \rho(A) \leq 6$$

## 5.2 乘幂法与反幂法

在实际工程应用中,如大型结构的振动系统中,往往要计算振动系统的最低频率(或前几个最低频率)及相应的振型,相应的数学问题便为求解矩阵的按模最大或前几个按模最大特征值及相应的特征向量问题,或称为求主特征值问题。

### 5.2.1 乘幂法

乘幂法是用于求大型稀疏矩阵的主特征值的迭代方法,其特点是公式简单,易于上机实现。

乘幂法的计算公式为:

设  $A \in R^{n \times n}$ , 取初始向量  $x^{(0)} \in R^n$ , 令  $x^{(1)} = Ax^{(0)}$ ,  $x^{(2)} = Ax^{(1)}$ ,  $\dots$ , 一般有

$$x^{(k)} = Ax^{(k-1)} \quad (5.5)$$

形成迭代向量序列  $\{x^{(k)}\}$ 。由递推公式 (5.5), 有

$$\begin{aligned} x^{(k)} &= A(Ax^{(k-2)}) \\ &= A^2 x^{(k-1)} \\ &= \dots \\ &= A^k x^{(0)} \end{aligned} \quad (5.6)$$

这表明  $x^{(k)}$  是用  $A$  的  $k$  次幂左乘  $x^{(0)}$  得到的, 因此称此方法为乘幂法, (5.5) 或 (5.6) 式称为**乘幂公式**,  $\{x^{(k)}\}$  称为**迭代序列**。

下面分析乘幂过程, 即讨论当  $k \rightarrow \infty$  时,  $\{x^{(k)}\}$  与矩阵  $A$  的主特征值及相应特征向量的关系。

设  $A = (a_{ij})_{n \times n}$  有完全的特征向量系, 且  $\lambda_1, \lambda_2, \dots, \lambda_n$  为  $A$  的  $n$  个特征值, 满足

$$|\lambda_1| \geq |\lambda_2| \geq \dots \geq |\lambda_n|$$

$v_1, v_2, \dots, v_n$  为相应的特征向量且线性无关, 从而构成  $R^n$  上的一组基底。

对任取初始向量  $x^{(0)} \in R^n$ , 可由这组基底展开表示为

$$\begin{aligned} x^{(0)} &= \alpha_1 v_1 + \alpha_2 v_2 + \dots + \alpha_n v_n \\ &= \sum_{j=1}^n \alpha_j v_j \end{aligned} \quad (5.7)$$

其中  $\alpha_1, \alpha_2, \dots, \alpha_n$  为展开系数。将  $x^{(0)}$  的展开式 (5.7) 代入乘幂公式 (5.6) 中, 得

$$x^{(k)} = A^k \sum_{j=1}^n \alpha_j v_j = \sum_{j=1}^n \alpha_j (A^k v_j) \quad (5.8)$$

利用  $A^k v_j = \lambda_j^k v_j$

(5.8) 式为

$$x^{(k)} = \sum_{j=1}^n \alpha_j \lambda_j^k v_j \quad (5.9)$$

(1) 如果  $A$  有唯一的主特征值, 即  $|\lambda_1| > |\lambda_2| \geq \dots$ , 设  $\lambda_1 \neq 0$ , 且由 (5.9) 式, 有

$$\begin{aligned} x^{(k)} &= \lambda_1^k \left[ \alpha_1 v_1 + \sum_{j=2}^n \alpha_j \left( \frac{\lambda_j}{\lambda_1} \right)^k v_j \right] \\ &= \lambda_1^k (\alpha_1 v_1 + \varepsilon_k) \end{aligned}$$

其中  $\varepsilon_k = \sum_{j=2}^n \alpha_j \left( \frac{\lambda_j}{\lambda_1} \right)^k v_j$ , 由于  $\left| \frac{\lambda_j}{\lambda_1} \right| < 1, j=2, 3, \dots, n$ , 故当  $k$  充分大时,  $\varepsilon_k \approx 0$ , 此时

$$x^{(k)} \approx \lambda_1^k \alpha_1 v_1 \quad (5.10)$$

对  $i=1, 2, \dots, n$ , 若  $(\alpha_1 v_1)_i \neq 0$ , 考虑相邻迭代向量的对应分量比值,

$$\frac{x_i^{(k+1)}}{x_i^{(k)}} \approx \frac{\lambda_1^{(k+1)} (\alpha_1 v_1)_i}{\lambda_1^k (\alpha_1 v_1)_i} = \lambda_1 \quad (5.11)$$

即对  $i=1, \dots, n$

$$\lim_{k \rightarrow \infty} \frac{x_i^{(k+1)}}{x_i^{(k)}} = \lambda_1 \quad (5.12)$$

这表明主特征值  $\lambda_1$  可由 (5.11) 或 (5.12) 式得到。

由于迭代序列  $x^{(k)}$ , 当  $k$  充分大时, (5.10) 式成立,  $x^{(k)}$  与  $v_1$  只相差一个常数因子, 故可取  $x^{(k)}$  作为相应于主特征值  $\lambda_1$  的特征向量的近似值。

迭代序列  $x^{(k)}$  的收敛速度取决于  $\left| \frac{\lambda_2}{\lambda_1} \right|$  的大小。

(2) 如果  $A$  的主特征值不唯一, 且  $|\lambda_1| = |\lambda_2| > |\lambda_3| \geq \dots$  可分三种情况讨论:

a)  $\lambda_1 = \lambda_2$ ; b)  $\lambda_1 = -\lambda_2$ ; c)  $\lambda_1 = \bar{\lambda}_2$

情况 a) 当  $\lambda_1 = \lambda_2$  时,  $A$  的主特征值为二重根, 根据 (5.9) 式

$$\begin{aligned} x^{(k)} &= \lambda_1^k \left[ \alpha_1 v_1 + \alpha_2 v_2 + \sum_{j=3}^n \alpha_j \left( \frac{\lambda_j}{\lambda_1} \right)^k v_j \right] \\ &= \lambda_1^k (\alpha_1 v_1 + \alpha_2 v_2 + \varepsilon_k) \end{aligned}$$

当  $k$  充分小时, 由于  $\left| \frac{\lambda_j}{\lambda_1} \right| < 1, j=3, \dots, n$ ,  $\varepsilon_k \approx 0$ , 则

$$x^{(k)} \approx \lambda_1^k (\alpha_1 v_1 + \alpha_2 v_2)$$

对  $i=1, 2, \dots, n$ , 如果  $(\alpha_1 v_1 + \alpha_2 v_2)_i \neq 0$ , 则

$$\lim_{k \rightarrow \infty} \frac{x_i^{(k+1)}}{x_i^{(k)}} = \lambda_1 \quad (\text{主特征值})$$

且  $x^{(k)}$  收敛到相应于  $\lambda_1 (= \lambda_2)$  的特征向量的近似值。

这种重主特征值的情况, 可推广到  $A$  的  $r$  重主特征值的情况, 即当

$$\lambda_1 = \lambda_2 = \cdots = \lambda_r \quad \text{且} \quad |\lambda_1| > |\lambda_{r+1}|$$

时, 上述讨论的结论仍然成立。

情况 b) 当  $\lambda_1 = -\lambda_2$  时,  $A$  的主特征值为相反数, (5.9) 式为

$$\begin{aligned} x^{(k)} &= \alpha_1 \lambda_1^k v_1 + \alpha_2 \lambda_2^k v_2 + \sum_{j=3}^n \alpha_j \lambda_j^k v_j \\ &= \alpha_1 \lambda_1^k v_1 + (-1)^k \alpha_2 \lambda_1^k v_2 + \sum_{j=3}^n \alpha_j \lambda_j^k v_j \\ &= \lambda_1^k \left[ \alpha_1 v_1 + (-1)^k \alpha_2 v_2 + \sum_{j=3}^n \alpha_j \left( \frac{\lambda_j}{\lambda_1} \right)^k v_j \right] \\ &= \lambda_1^k (\alpha_1 v_1 + (-1)^k \alpha_2 v_2 + \varepsilon_k) \end{aligned}$$

当  $k$  充分大时,  $\left| \frac{\lambda_j}{\lambda_1} \right| < 1, j = 3, 4, \cdots, n, \varepsilon_k \approx 0$ , 则

$$x^{(k)} \approx \lambda_1^k (\alpha_1 v_1 + (-1)^k \alpha_2 v_2) \quad (5.13)$$

由于 (5.13) 式中出现因子  $(-1)^k$ , 则当  $k$  变化时,  $x^{(k)}$  出现振荡、摆动现象, 不收敛, 利用  $(-1)^k$  的特点, 连续迭代两步, 得

$$\begin{aligned} x^{(k+1)} &\approx \lambda_1^{(k+2)} (\alpha_1 v_1 + (-1)^{k+2} \alpha_2 v_2) \\ &= \lambda_1^{(k+2)} (\alpha_1 v_1 + (-1)^k \alpha_2 v_2) \end{aligned}$$

从而, 对  $i = 1, 2, \cdots, n$ , 若  $(\alpha_1 v_1 + (-1)^k \alpha_2 v_2)_i \neq 0$ , 则

$$\lim_{k \rightarrow \infty} \frac{x_i^{(k+2)}}{x_i^{(k)}} = \lambda_1^2 \quad (5.14)$$

开方之后, 便得到  $A$  的以上主特征值  $\lambda_1, \lambda_2 = -\lambda_1$ 。

为计算相应于  $\lambda_1, \lambda_2$  的特征向量, 采取组合方式,

$$x^{(k+1)} + \lambda_1 x^{(k)} \approx 2\lambda_1^{(k+1)} \alpha_1 v_1 = C_k^1 v_1 \quad (5.15)$$

$$x^{(k+1)} - \lambda_1 x^{(k)} \approx (-1)^{k+1} 2\lambda_1^{(k+1)} \alpha_2 v_2 = C_k^2 v_2 \quad (5.16)$$

可见  $C_k^1 v_1, C_k^2 v_2$  分别为相应于  $\lambda_1$  与  $\lambda_2$  的特征向量。

情况 c) 当  $\lambda_1 = \bar{\lambda}_2$  时,  $A$  的主特征值为共轭复根。

因  $A$  为实矩阵,  $\bar{A} = A$ , 于是由

$$Av_1 = \lambda_1 v_1$$

有

$$\overline{Av_1} = A\bar{v}_1 = \lambda_2 \bar{v}_1$$

即  $\bar{v}_1 = v_2$  ( $v_1$  与  $v_2$  为互为共轭向量)。

设  $\lambda_1 = \rho e^{i\theta}$ ,  $\lambda_2 = \rho e^{-i\theta}$ , 对任取  $x^{(0)} \in R^n$ , 展开式 (5.7) 可为

$$x^{(0)} = \alpha_1 v_1 + \bar{\alpha}_1 \bar{v}_1 + \sum_{j=3}^n \alpha_j v_j \quad (5.17)$$

将 (5.17) 式代入 (5.9) 式,

$$\begin{aligned} x^{(k)} &= \alpha_1 \lambda_1^k v_1 + \bar{\alpha}_1 \lambda_2^k \bar{v}_1 + \sum_{j=3}^n \alpha_j \lambda_j^k v_j \\ &= \alpha_1 \rho^k e^{ik\theta} v_1 + \bar{\alpha}_1 \rho^k e^{-ik\theta} v_2 + \rho^k \sum_{j=3}^n \alpha_j \left( \frac{\lambda_j}{\rho} \right)^k v_j \end{aligned}$$

同理, 当  $k$  充分大时

$$x^{(k)} \approx \rho^k (\alpha_1 v_1 e^{ik\theta} + \bar{\alpha}_1 \bar{v}_1 e^{-ik\theta}) \quad (5.18)$$

对  $j = 1, 2, \dots, n$ , 设复数表示

$$(\alpha_1 v_1)_j = r_j e^{i\varphi}, \quad (\bar{\alpha}_1 \bar{v}_1)_j = r_j e^{-i\varphi}$$

则 (5.18) 式的复数表示可为

$$x_j^{(k)} \approx \rho^k (r_j e^{i(\varphi+k\theta)} + r_j e^{-i(\varphi+k\theta)})$$

连续迭代, 得

$$\begin{cases} x_j^{(k)} \approx 2\rho^k r_j \cos(\varphi + k\theta) \\ x_j^{(k+1)} \approx 2\rho^{k+1} r_j \cos(\varphi + (k+1)\theta) \\ x_j^{(k+2)} \approx 2\rho^{k+2} r_j \cos(\varphi + (k+2)\theta) \end{cases} \quad (5.19)$$

利用三角函数运算性质及  $\lambda_1$ 、 $\lambda_2$  的复数表示, 不难验证。

$$x_j^{(k+2)} - (\lambda_1 + \lambda_2)x_j^{(k+1)} + \lambda_1 \lambda_2 x_j^k \approx 0$$

令

$$p = -(\lambda_1 + \lambda_2), \quad q = \lambda_1 \lambda_2 \quad (5.20)$$

解方程 ( $j = 1, 2, \dots, n$ )

$$x_j^{(k+2)} + px_j^{(k+1)} + qx_j^{(k)} = 0 \quad (5.21)$$

求出  $p, q$  后, 再解出主特征值  $\lambda_1$ 、 $\lambda_2$ , 得

$$\begin{cases} \lambda_1 = -\frac{p}{2} + i\sqrt{q - \left(\frac{p}{2}\right)^2} \\ \lambda_2 = -\frac{p}{2} - i\sqrt{q - \left(\frac{p}{2}\right)^2} \end{cases} \quad (5.22)$$

同样, 采取组合方式求相应于  $\lambda_1$ 、 $\lambda_2$  的特征向量。由于

$$x^{(k+1)} - \lambda_2 x^{(k)} \approx \lambda_1^k (\lambda_1 - \lambda_2) \alpha_1 v_1 = C_k^1 v_1 \quad (5.23)$$

$$x^{(k+1)} - \lambda_1 x^{(k)} \approx \lambda_2^k (\lambda_2 - \lambda_1) \alpha_2 v_2 = C_k^2 v_2 \quad (5.24)$$

则可分别取 (5.23)、(5.24) 左端的组合表达式作为相应于  $\lambda_1$ 、 $\lambda_2$  的特征向量的近似值。

通过上述分析, 有

**定理 6** 设  $A \in R^{n \times n}$  有完全特征向量系, 若  $\lambda_1, \lambda_2, \dots, \lambda_n$  为  $A$  的  $n$  个特征值且满足

$$|\lambda_1| \geq |\lambda_2| \geq \dots \geq |\lambda_n|$$

对任取初始向量  $x^{(0)} \in R^n$ , 对乘幂公式

$$x^{(k+1)} = Ax^{(k)}$$

确定的迭代序列  $\{x^{(k)}\}$ , 有下述结论:

(1) 当  $|\lambda_1| > |\lambda_2|$  时, 对  $i = 1, 2, \dots, n$

$$\lim_{k \rightarrow \infty} \frac{x_i^{(k+1)}}{x_i^{(k)}} = \lambda_1$$

收敛速度取决于  $r = \left| \frac{\lambda_2}{\lambda_1} \right| < 1$  的程度,  $r \ll 1$  收敛快,  $r \approx 1$  收敛慢, 且  $x^{(k)}$  (当  $k$  充分大时)

为相应于  $\lambda_1$  的特征向量的近似值。

(2) 当  $|\lambda_1| = |\lambda_2| > |\lambda_3|$  时

a) 若  $\lambda_1 = \lambda_2$ , 则主特征值  $\lambda_1$  及相应特征向量的求法同 (1);

b) 若  $\lambda_1 = -\lambda_2$ , 对  $i = 1, 2, \dots, n$

$$\lim_{k \rightarrow \infty} \frac{x_i^{(k+1)}}{x_i^{(k)}} = \lambda_1^2$$

收敛速度取决于  $r = \left| \frac{\lambda_3}{\lambda_1} \right| < 1$  的程度。向量  $x^{(k+1)} + \lambda_2 x^{(k)}$ 、 $x^{(k+1)} - \lambda_1 x^{(k)}$  分别为主特征值  $\lambda_1$ 、

$\lambda_2$  相应的特征向量的近似值。

c) 若  $\lambda_1 = \bar{\lambda}_2$ , 则连续迭代两次, 计算出  $x^{(k+1)}$ ,  $x^{(k+2)}$ , 然后对  $j = 1, 2, \dots, n$  解方程

$$x_j^{(k+2)} + px_j^{(k+1)} + qx_j^{(k)} = 0$$

求出  $p$ 、 $q$  后, 由公式

$$\lambda_1 = -\frac{p}{2} + i\sqrt{q - \left(\frac{p}{2}\right)^2}$$

$$\lambda_2 = -\frac{p}{2} - i\sqrt{q - \left(\frac{p}{2}\right)^2}$$

解出主特征值  $\lambda_1$ 、 $\lambda_2$ 。此时收敛速度取决于  $r = \left| \frac{\lambda_3}{\lambda_1} \right| < 1$  的程度。向量  $x^{(k+1)} - \lambda_2 x^{(k)}$ 、

$x^{(k+1)} - \lambda_1 x^{(k)}$  分别为相应于  $\lambda_1, \lambda_2$  的特征向量的近似值。

从分析乘幂过程可见，乘幂法可用于求矩阵按模最大的一个（或几个）特征值及相应的特征向量，当比值  $r = \left| \frac{\lambda_2}{\lambda_1} \right| \ll 1$  时，收敛速度快， $r \approx 1$  时，收敛速度慢，且计算公式简便，

便于上机实现。分析中的假设  $(\alpha_1 v_1)_i \neq 0$ 、 $(\alpha_1 v_1 + \alpha_2 v_2)_i \neq 0$ 、 $\dots$ ，在计算时可不用考虑，如果此条件不满足，则可通过迭代误差自行调整。

在用乘幂法求矩阵的主特征值  $\lambda_1$  及对应的特征向量时，迭代向量的分量  $x_i^{(k)}$  可能会出现绝对值非常大的现象，从而造成计算中溢出的可能。为此，需对迭代向量  $x^{(k)}$  进行规范化。

令  $\max(x)$  表示向量  $x$  分量中绝对值最大者。即如果有某  $i_0$ ，使

$$|x_{i_0}| = \max_{1 \leq i \leq n} |x_i|$$

则

$$\max(x) = x_{i_0}$$

对任取初始向量  $x^{(0)}$ ，记

$$y^{(0)} = x^{(0)} / \max(x^{(0)})$$

则

$$x^{(1)} = Ay^{(0)}$$

一般地，若已知  $x^{(k)}$ ，称公式

$$\begin{cases} y^{(k)} = x^{(k)} / \max(x^{(k)}) \\ x^{(k+1)} = Ay^{(k)} \end{cases} \quad (k = 0, 1, \dots) \quad (5.25)$$

为规范化的乘幂法公式或改进乘幂法公式，这里，乘幂迭代序列  $y^{(k)}$  的分量绝对值最大者 1。

类似前面的分析乘幂过程，有

**定理 7** 设  $A \in R^{n \times n}$  具有完全特征向量系， $\lambda_1, \lambda_2, \dots, \lambda_n$  为  $A$  的  $n$  个特征值，且满足

$$|\lambda_1| > |\lambda_2| \geq \dots \geq |\lambda_n|$$

则对任初始向量  $x^{(0)}$ ，由规范化的乘幂法公式 (5.25) 确定的向量序列  $y^{(k)}$ ， $x^{(k)}$  满足

$$(1) \lim_{k \rightarrow \infty} \max(x^{(k)}) = \lambda_1 \quad (5.26)$$

(2)  $y^{(k)}$  为相应于主特征值  $\lambda_1$  的特征向量近似值

$$y^{(k)} \approx v_1, \quad (Av_1 = \lambda_1 v_1) \quad (5.27)$$

例 2 用规范化乘幂法计算矩阵  $A$  的主特征值及相应特征向量

$$A = \begin{bmatrix} -4 & 14 & 0 \\ -5 & 13 & 0 \\ -1 & 0 & 2 \end{bmatrix}$$

解  $A$  的特征值  $\lambda_1 = 6, \lambda_2 = 3, \lambda_3 = 2$

取初始值  $x^{(0)} = (1, 1, 1)^T$ ，用规范化乘幂法公式（5.25）计算

$$\max(x^{(0)}) = 1$$

$$y^{(0)} = x^{(0)} / \max(x^{(0)}) = (1, 1, 1)^T$$

$$x^{(1)} = Ay^{(0)} = (10, 8, 1)^T$$

其它结果见表 5.1（表中的向量均为转置向量）。

表 5.1

$k$	$\max(y^{(k)})$	$x^{(k)} = y^{(k)} / \max(x^{(k)})$	$x^{(k+1)} = Ay^{(k)}$
0	1	(1, 1, 1)	(10, 8, 1)
1	10	(1, 0.8, 0.1)	(7.2, 5.4, -0.8)
2	7.2	(1, 0.75, -0.111111)	(6.5, 4.75, -1.222222)
3	6.57	(1, 0.730769, -0.203704)	(6.230766, 4.499997, -1.407408)
4	6.230766	(1, 0.722222, -0.225880)	(6.111108, 4.388886, -1.1451767)
5	6.111108	(1, 0.718182, -0.237561)	(6.054548, 4.336336, -1.475122)
6	6.054548	(1, 0.716216, -0.243639)	(6.027024, 4.310808, -1.487278)
7	6.027024	(1, 0.715247, -0.246768)	(6.013458, 4.298211, -1.483536)
8	6.013458	(1, 0.714765, -0.248366)	(6.00671, 4.291945, -1.496732)
9	6.00671	(1, 0.714525, -0.249177)	(6.00335, 4.28825, -1.496354)
10	6.00335	(1, 0.714405, -0.249586)	(6.00167, 4.287265, -1.499172)
11	6.00167	(1, 0.714345, -0.239792)	(6.00083, 4.286485, -1.499584)
12	6.00083	(1, 0.714315, -0.249896)	

取  $\max(x^{(12)}) = 6.00083$  作为主特征值  $\lambda_1$  的近似值，与真值  $\lambda_1 = 6$  相比，有较好的近似程度，相应于  $\lambda_1$  的特征向量的近似值取为  $y^{(2)} = (1, 0.714315, -0.249896)^T$ 。

### 5.2.2 乘幂法的加速及降价

当  $\lambda_i$  ( $i = 1, 2, \dots, n$ ) 为矩阵  $A \in R^{n \times n}$  的  $n$  个特征值，且

$$|\lambda_1| > |\lambda_2| \geq \dots \geq |\lambda_n|$$

时，乘幂法的收敛速度由  $r = \frac{|\lambda_2|}{|\lambda_1|}$  决定， $r \ll 1$  收敛得快。因此为提高收敛速度或改善  $r \approx 1$

的状况，可以采取原点移位的方法，改变原矩阵  $A$  的状态。

取  $\lambda_0$ （常数），用矩阵  $B = A - \lambda_0 I$  来代替  $A$  进行乘幂迭代。设  $\mu_i$  ( $i = 1, 2, \dots, n$ ) 为矩阵  $B$  的特征值，则  $B$  与  $A$  特征值之间应有关系式：

$$\mu_i = \lambda_i - \lambda_0 \quad (i = 1, 2, \dots, n)$$

且若  $v_i$  是  $A$  相应于  $\lambda_i$  的特征向量，则  $v_i$  亦是  $\mu_i$  的特征值，即对  $i = 1, 2, \dots, n$

$$Bv_i = (A - \lambda_0 I)v_i = Av_i - \lambda_0 v_i = (\lambda_i - \lambda_0)v_i$$

因此，对任取  $x^{(0)} \in R^n$ ，关于矩阵  $B$  的乘幂公式（5.6）可为



$$\begin{aligned}
x^{(k)} &= B^k x^{(0)} = (A - \lambda_0 I)^k x^{(0)} \\
&= \mu_1^k \left[ \alpha_1 v_1 + \sum_{j=2}^n \alpha_j \left( \frac{\mu_j}{\mu_1} \right)^k v_j \right] \\
&= (\lambda_1 - \lambda_0)^k \left[ \alpha_1 v_1 + \sum_{j=2}^n \alpha_j \left( \frac{\lambda_j - \lambda_0}{\lambda_1 - \lambda_0} \right)^k v_j \right]
\end{aligned}$$

为加快收敛速度，适当选择参数  $\lambda_0$ ，使

$$\omega(\lambda_0) = \max_{2 \leq j \leq n} \left| \frac{\lambda_j - \lambda_0}{\lambda_1 - \lambda_0} \right|^k \quad (5.28)$$

达到最小值。如当  $\lambda_i (i = 1, 2, \dots, n)$  为实数，且  $\lambda_1 > \lambda_2 \geq \dots \geq \lambda_n$  时，取

$$\lambda_0^* = \frac{1}{2}(\lambda_2 + \lambda_n)$$

则  $\lambda_0^*$  为  $\omega(\lambda_0)$  的极小值点。这时

$$\left| \frac{\lambda_2 - \lambda_0^*}{\lambda_1 - \lambda_0^*} \right| = \left| \frac{\lambda_2 - \frac{1}{2}\lambda_2 - \frac{1}{2}\lambda_n}{\lambda_1 - \frac{1}{2}\lambda_2 - \frac{1}{2}\lambda_n} \right| = \left| \frac{\lambda_2 - \lambda_n}{2\lambda_1 - \lambda_2 - \lambda_n} \right| < \left| \frac{\lambda_2}{\lambda_1} \right|$$

原点移位法是一个矩阵变换过程，变换简单且不破坏原矩阵的稀疏性。但由于预先不知道特征值的分布，所以应用起来有一定困难，通常对特征值的分布有一个大略估计，设定一个参数  $\lambda_0$  进行试算，当所取  $\lambda_0$  对迭代有明显加速效应以后再进行确定计算。

例 3 计算  $A$  的主特征值

$$A = \begin{bmatrix} 1.0 & 1.0 & 0.5 \\ 1.0 & 1.0 & 0.25 \\ 0.5 & 0.25 & 2.0 \end{bmatrix}$$

解 先用规范化乘幂法计算，得表 5.2

表 5.2

$k$	$y^{(k)} = x^{(k)} / \max(x^{(k)})$	$\lambda_1 \approx \max(x^{(k)})$
0	(1, 1, 1)	
1	(0.9091, 0.8182, 1) <sup>T</sup>	2.75
19	(0.7482, 0.6497, 1) <sup>T</sup>	2.5365374
20	(0.7482, 0.6497, 1) <sup>T</sup>	2.5365323

主特征值  $\lambda_1$  及特征向量  $v_1$  为（8 位有效数字）

$$\lambda_1 = 2.5362258$$

$$v_1 = (0.74822116, 0.64966116, 1)^T$$

而用规范化乘幂法计算的相应近似值为：

$$\lambda_1 \approx \max(x^{(20)}) = 2.5365323$$

$$v_1 \approx y^{(20)} = (0.7482, 0.6497, 1)^T$$

如果采用原点位移的加速法求解，取  $\lambda_0 = 0.75$ ，矩阵  $b = A - \lambda_0 I$

$$B = \begin{bmatrix} 0.25 & 1 & 0.5 \\ 1 & 0.25 & 0.25 \\ 0.5 & 0.25 & 1.25 \end{bmatrix}$$

对矩阵  $B$  应用规范化乘幂法公式见表 5.3。

表 5.3

$k$	$y^{(k)} = x^{(k)} / \max(x^{(k)})$	$\lambda_1 \approx \max(x^{(k)})$
0	$(1, 1, 1)^T$	
9	$(0.7483, 0.6497, 1)^T$	1.7866587
10	$(0.7483, 0.6497, 1)^T$	1.7865914

可见

$$\lambda_1 = \mu_1 + \lambda_0 \approx 2.5365914$$

此结果与未加速的规范化乘幂法公式计算结果相比，收敛速度要快得多。

在已经求出主特征值  $\lambda_1$  及特征向量  $v_1$  以后，可将原矩阵进行修改，使修改后的矩阵按模最大特征值是原矩阵的按模次大特征值，再用乘幂法去求按模次大特征值及特征向量，此方法称为降阶过程。

为使问题简单，设  $A \in R^{n \times n}$  为对称矩阵。假定已求出主特征值  $\lambda_1$  及特征向量  $v_1$ ，记  $A^{(1)} = A$ ，构造矩阵

$$A^{(2)} = A^{(1)} - \lambda_1 v_1 v_1^T / v_1^T v_1 \quad (5.29)$$

因  $A$  对称，则具完全特征向量系，且特征向量  $v_i$  可两两相互正交，即满足  $v_1^T v_i = 0$  ( $i = 2, \dots, n$ )，于是

$$A^{(2)} v_1 = A^{(1)} v_1 - \lambda_1 v_1 (v_1^T v_1) / v_1^T v_1 = 0$$

$$A^{(2)} v_i = A^{(1)} v_i - \lambda_1 v_1 (v_1^T v_i) / v_1^T v_1 = A^{(1)} v_i = \lambda_i v_i \quad (i = 2, 3, \dots, n)$$

这表明矩阵  $A^{(2)}$  除一个特征值  $\lambda_1 = 0$  与矩阵  $A^{(1)}$  不一样之外，其余与  $A^{(1)}$  具相同特征值，且  $A^{(2)}$  的按模最大特征值是  $\lambda_2$ ，用  $A^{(2)}$  代替  $A^{(1)}$  进行乘幂迭代即可求得  $\lambda_2$  及相应的特征向量  $v_2$ ，而  $\lambda_2$  又为  $A^{(1)}$  的按模次大特征值。这种做法称为降阶法。

注意，降阶法实际上只可用少数几次，求矩阵的前几个按模最大特征值及特征向量。因为，每降阶一次，计算精度就会损失或降低一些。

### 5.2.3 反幂法

设  $A \in R^{n \times n}$  可逆，则无零特征值，由

$$Ax = \lambda x \quad (x \neq 0)$$

有

$$A^{-1}x = \frac{1}{\lambda}x$$

即若 $\lambda$ 为矩阵 $A$ 的特征值, 则 $\frac{1}{\lambda}$ 必为矩阵 $A^{-1}$ 的特征值, 且特征向量相同。

如果 $A$ 的 $n$ 个特征值 $\lambda_i (i = 1, 2, \dots, n)$ 为

$$|\lambda_1| \geq |\lambda_2| \geq \dots \geq |\lambda_n|$$

则 $A^{-1}$ 的 $n$ 个特征值更为

$$\left| \frac{1}{\lambda_n} \right| \geq \left| \frac{1}{\lambda_{n-1}} \right| \geq \dots \geq \left| \frac{1}{\lambda_1} \right|$$

因此, 若乘幂法可求 $A$ 的主特征值 $\lambda_1$ , 则用 $A^{-1}$ 做乘幂矩阵, 由乘幂迭代格式

$$x^{(k+1)} = A^{-1}x^{(k)} \quad (5.30)$$

便可求出 $A^{-1}$ 的按模最大特征值 $\frac{1}{\lambda_n}$ , 取倒数, 即为矩阵 $A$ 的按模最小特征值。因此, 对任

取初始向量 $x^{(0)} \in R^n$ , 称公式(5.30)为求矩阵 $A$ 按模最小特征值的反幂法。

在应用公式(5.30)计算时, 由于要计算 $A$ 的逆矩阵 $A^{-1}$ , 一方面计算复杂、麻烦, 另一方面, 有时会破坏 $A$ 的稀疏性, 故改写(5.30)式为:

$$Ax^{(k+1)} = x^{(k)} \quad (k = 0, 1, \dots) \quad (5.31)$$

类似于公式(5.25)的规范化乘幂法公式为

$$\begin{cases} y^{(k+1)} = x^{(k)} / \max(x^{(k)}) \\ Ax^{(k+1)} = y^{(k)} \end{cases} \quad (k = 0, 1, \dots) \quad (5.32)$$

如果考虑到利用原点移位加速的反幂法, 则记 $B = A - \lambda_0 I$ , 对任取初始向量 $x^{(0)} \in R^n$ ,

$$\begin{cases} y^{(k)} = x^{(k)} / \max(x^{(k)}) \\ Bx^{(k+1)} = y^{(k)} \end{cases} \quad (k = 0, 1, \dots) \quad (5.33)$$

由于反幂法的主要工作量是每迭代一步都要解一个线性方程组(5.31), 且系数矩阵 $A$  (或 $B$ )是不变的, 故可利用矩阵的三角分解 $A = LU$  (或 $B = LU$ ), 则每次迭代只需解二个三角形方程组

$$\begin{cases} L\tilde{x} = y^{(k)} \\ Ux^{(k+1)} = \tilde{x} \end{cases} \quad (5.34)$$

且当 $|\lambda_1| \geq |\lambda_2| \geq \dots \geq |\lambda_{n-1}| > |\lambda_n| > 0$  时

$$\lim_{k \rightarrow \infty} y^{(k)} = \frac{1}{\lambda_n} \quad (5.35)$$

同时 $x^{(k+1)}$ 便为所求的特征向量, 收敛速度为 $\left| \frac{\lambda_n}{\lambda_{n-1}} \right|$ 。

反幂法的主要应用是已知矩阵的近似特征值后, 求矩阵的特征向量, 且收敛快, 精度高, 是目前求特征向量最有效的方法之一。

### 5.3 子空间迭代法

子空间迭代法也称为平行迭代法，是乘幂法的推广。乘幂法每次只能求出矩阵的一个主特值及特征向量，而子空间迭代法一次可求出矩阵的前几个按模最大特值及特征向量。

首先介绍斯密特（Schmidt）正交化过程。

以三维为例，设  $\alpha_1, \alpha_2, \alpha_3$  为  $R^3$  上的三个线性无关的向量，

令  $\beta_1 = \alpha_1 / \|\alpha_1\|_2$ ，则  $\beta_1$  为单位长度的向量，再令

$$\beta'_2 = \alpha_2 - (\alpha_2, \beta_1)\beta_1, \quad \beta_2 = \beta'_2 / \|\beta'_2\|_2$$

可以验证  $(\beta_1, \beta_2) = 0$ ，即  $\beta_1$  与  $\beta_2$  正交。若令

$$\beta'_3 = \alpha_3 - (\alpha_3, \beta_1)\beta_1 - (\alpha_3, \beta_2)\beta_2$$

则

$$(\beta'_3, \beta_1) = (\beta'_3, \beta_2) = 0$$

即  $\beta'_3$  与  $\beta_1, \beta_2$  正交，将其单位化为

$$\beta_3 = \beta'_3 / \|\beta'_3\|_2$$

于是向量组  $\beta_1, \beta_2, \beta_3$  构成  $R^3$  上一组标准正交基，且

$$\begin{aligned} [\alpha_1, \alpha_2, \alpha_3] &= [\beta_1, \beta_2, \beta_3] \begin{bmatrix} \|\alpha_1\|_2 & (\alpha_2, \beta_1) & (\alpha_3, \beta_1) \\ & \|\beta'_2\|_2 & (\alpha_3, \beta_2) \\ & & \|\beta'_3\|_2 \end{bmatrix} \\ &= QR \end{aligned}$$

其中  $Q = [\beta_1, \beta_2, \beta_3]$  为正交矩阵， $R$  是上三角阵。

对  $n$  维向量空间，设  $\alpha_1, \dots, \alpha_n$  为  $R^n$  上  $n$  个线性无关的向量，类似有

$$\begin{aligned} \alpha_1 &= \alpha_1, & \beta_1 &= \alpha_1 / \|\alpha_1\|_2 \\ \beta'_2 &= \alpha_2 - (\alpha_2, \beta_1)\beta_1, & \beta_2 &= \beta'_2 / \|\beta'_2\|_2 \\ \beta'_3 &= \alpha_3 - (\alpha_3, \beta_1)\beta_1 - (\alpha_3, \beta_2)\beta_2, & \beta_3 &= \beta'_3 / \|\beta'_3\|_2 \\ \dots & \dots & & \\ \beta'_n &= \alpha_n - \sum_{j=1}^{n-1} (\alpha_n, \beta_j)\beta_j, & \beta_n &= \beta'_n / \|\beta'_n\|_2 \end{aligned}$$

即

$$[\alpha_1, \dots, \alpha_n] = [\beta_1, \dots, \beta_n] = \begin{bmatrix} \|\alpha_1\|_2 & (\alpha_2, \beta_1) & (\alpha_3, \beta_1) & \cdots & (\alpha_n, \beta_1) \\ & \|\beta_2\|_2 & (\alpha_3, \beta_2) & \cdots & (\alpha_n, \beta_2) \\ & & \|\beta'_3\|_2 & \cdots & (\alpha_n, \beta_3) \\ & & & \ddots & \vdots \\ & & & & \|\beta'_n\|_2 \end{bmatrix}$$

$$= QR$$

$Q$  为正交阵,  $R$  为上三角阵, 将  $n$  个线性无关向量变换为  $n$  个两两正交向量的方法称为斯密特正交化方法。即斯密特正交化过程可将可逆阵  $A$  分解为正交阵与上三角阵的乘积。

当向量组  $[\alpha_1, \dots, \alpha_p]$  的个数  $p < n$  时, 上述过程仍然成立。

子空间迭代法的做法是, 若要求  $A$  的前  $p$  个特征值及相应的特征向量 ( $p < n$ ), 先取  $p$  个线性无关的向量  $x_1, x_2, \dots, x_p$  构成一个  $n \times p$  的初始矩阵

$$Y_0 = [x_1, x_2, \dots, x_p]$$

用矩阵  $A$  左乘上式,

$$Z_1 = AY_0$$

$Z_1$  的各列线性无关, 但不一定正交, 采用斯密特正交化过程, 相当于正交分解  $Z_1$  为

$$Z_1 = Y_1 R_1 \quad \text{或} \quad Y_1 = Z_1 R_1^{-1}$$

$Y_1$  的各列正交, 且  $R_1$  为可逆上三角阵, 或

$$AY_0 = Y_1 R_1$$

若已知第  $k$  步迭代近似阵为  $Y_k$ , 则子空间迭代法公式,

$$\begin{cases} Z_{k+1} = AY_k \\ Y_{k+1} = Z_{k+1} R_{k+1}^{-1} \end{cases} \quad \text{或} \quad AY_k = Y_{k+1} R_{k+1} \quad k = 1, 2, \dots \quad (5.36)$$

子空间迭代法适用于大型对称稀疏矩阵特征问题的求解。

当  $A \in R^{n \times n}$  对称时, 可按下述正交化步骤进行计算,

1. 计算

$$Z_{k+1} = AY_k$$

2. 计算  $p$  阶对称阵  $G_{k+1} = Z_{k+1}^T Z_{k+1}$ ;

3. 求矩阵  $U_{k+1} = [u_1, u_2, \dots, u_p]_{n \times p}$  使

$$U_{k+1}^T G_{k+1} U_{k+1} = D_{k+1}^2 = \begin{bmatrix} (d_1^{(k)})^2 & & \\ & \ddots & \\ & & (d_p^{(k)})^2 \end{bmatrix}$$

其中对角阵  $D_{k+1}^2$  的元素  $(d_1^{(k)})^2, \dots, (d_p^{(k)})^2$  为  $G_{k+1}$  的全部特征值,  $U_{k+1}$  的各列  $u_1, \dots, u_p$  为相应特征向量。

4. 计算

$$Y_{k+1} = Z_{k+1} U_{k+1} D_{k+1}^{-1}$$

可以验证  $Y_{k+1}$  为各列规范正交，即

$$Y_{k+1}^T Y_{k+1} = I$$

5. 若

$$|d_j^{(k+1)} - d_j^{(k)}| \leq \varepsilon, \quad (j = 1, 2, \dots, p)$$

则  $d_1^{(k+1)}, \dots, d_p^{(k+1)}$  便为  $A$  的前  $p$  个按模最小特征值， $Y_{k+1}$  便为相应的特征向量阵，否则

$Y_{k+1} \rightarrow Y_k$ ，转 (1)。

注：当  $p < n$  时，可用其它方法（见下节雅可比方法）求低阶对称矩阵的特征值，特征向量。

## 5.4 对称矩阵的雅可比 (Jacobi) 旋转法

雅克比 (Jacobi) 方法是求实对称矩阵全部特征值及对应的特征向量的方法。它也是一种迭代法，其基本思想是把对称矩阵  $A$  经一系列正交相似变换约化为一个近似对角阵，从而该对角阵的对角元就是  $A$  的近似特征值，由各个正交变换阵的乘积可得对应的特征向量。

### 1. 预备知识

雅克比方法涉及较多的代数知识，要承认如下一些主要结论：

- 1) 若  $B$  是上（或下）三角阵或对角阵，则  $B$  的主对角元素即是  $B$  的特征值。
- 2) 若矩阵  $P$  满足  $P^T P = I$ ，则称  $P$  为正交矩阵。显然  $P^T = P^{-1}$ ，且  $P_1, P_2, \dots$ ，是正交阵时，其乘积  $P = P_1 P_2 \dots P_k$  仍为正交矩阵。

3) 称矩阵

$$P_{ij} = \begin{pmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 1 & & & \\ & & \cos \theta & \sin \theta & & \\ & & 1 & & & \\ & & & \ddots & & \\ & & & & 1 & \\ & & -\sin \theta & \cos \theta & & \\ & & & & & 1 & \ddots & \\ & & & & & & & 1 \end{pmatrix} \quad (5.37)$$

为旋转矩阵，它是在单位阵  $I$  的  $i$  行，且  $j$  行和  $i$  列、 $j$  列的四个交叉位置上分别置上  $\cos \theta$ ， $\sin \theta$ ， $-\sin \theta$  和  $\cos \theta$  而成的。容易验证旋转矩阵是正交矩阵，即  $R^T(i, j) = R^{-1}(i, j)$ ，所以用它作相似变换阵时十分方便。雅克比方法就是用这种旋转矩阵对实对称阵  $A$  作一系列的旋转相似变换，从而将  $A$  约化为对角阵的。

用  $P_{ij}(i, j)$  作旋转变换的几何意义是：在维空间中，以  $i, j$  轴形成的平面上，把  $i, j$  轴旋转一个角度  $\theta$ 。

### 2. 雅克比方法

先以二阶矩阵为例：

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$$

旋转矩阵为

$$R = \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix}$$

$$\begin{aligned} B &= RAR^T = \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} \\ &= \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} \end{aligned}$$

其中

$$b_{11} = a_{11} \cos^2 \theta + a_{22} \sin^2 \theta + a_{12} \sin 2\theta$$

$$b_{12} = b_{21} = \frac{1}{2}(a_{22} - a_{11}) \sin 2\theta + a_{12} \cos 2\theta$$

$$b_{22} = a_{11} \sin^2 \theta + a_{22} \cos^2 \theta - a_{12} \cos 2\theta$$

为使  $A$  的相似矩阵  $B$  成为对角阵, 只须适当选取  $\theta$ , 使

$$b_{12} = b_{21} = \frac{1}{2}(a_{22} - a_{11}) \sin 2\theta + a_{12} \cos 2\theta = 0$$

即  $\operatorname{tg} 2\theta = \frac{2a_{12}}{a_{11} - a_{22}}$ , 其中  $|\theta| \leq \frac{\pi}{4}$ ,  $a_{11} = a_{22}$  时, 取  $\theta = \frac{\pi}{4}$ 。由此  $\theta$  可以确定, 从而旋

转矩阵  $P$  确定。 $A$  的特征值为:

$$\lambda_1 = b_{11}, \quad \lambda_2 = b_{22}$$

关于特征向量的计算

$$\text{设 } P = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \text{ 其中 } x_1, x_2 \text{ 为 } P \text{ 的行向量, 因为 } PAR^T = B \quad AP^T = P^TB$$

或  $(Ax_1, Ax_2) = (\lambda_1 x_1, \lambda_2 x_2)$ , 即  $Ax_i = \lambda_i x_i \quad (i = 1, 2, 3)$ , 所以对应于  $\lambda_1, \lambda_2$  的特征

向量是  $x_1 = (\cos \theta, \sin \theta)^T, \quad x_2 = (-\sin \theta, \cos \theta)^T$

考虑  $n$  阶矩阵的情况:

设矩阵  $A \in R^{n \times n}$  是对称矩阵, 记  $A_0 = A$ , 对  $A$  作一系列旋转相似变换, 即

$$A_k = P_k A_{k-1} P_k^T \quad (k = 1, 2, \dots) \quad (5.38)$$

其中  $A_k (k = 1, 2, \dots)$  仍是对称矩阵,  $P_k$  的形式

$$P_k = \begin{pmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 1 & & & \\ & & & \cos \theta & \sin \theta & \\ & & & 1 & & \\ & & & & \ddots & \\ & & & & & 1 \\ & & & -\sin \theta & \cos \theta & \\ & & & & & \ddots & \\ & & & & & & 1 \\ & & & & & & & i & & j & & \\ & & & & & & & & & & 1 \end{pmatrix}$$

$$P_{ii}^{(k)} = P_{jj}^{(k)} \quad P_{ij}^{(k)} = -P_{ji}^{(k)}$$

$$\text{也就是} \quad p_{pp}^{(k)} = p_{qq}^{(k)} = \cos \theta \quad p_{pq}^{(k)} = -p_{qp}^{(k)} = -\sin \theta$$

$$p_{ii}^{(k)} = 1 \quad p_{ij}^{(k)} = 0 \quad i, j \neq p, q$$

对任何角  $\theta$ , 可以验证:  $P_k$  是一个正交阵, 我们称它是  $(i, j)$  平面上的旋转矩阵, 相应地把变换 (5.38) 称为旋转变换;  $P_k$  和  $I$  仅在  $(ii)$ 、 $(jj)$ 、 $(ij)$  和  $(ji)$  上不同,  $P_k A_{k-1}$  只改变  $A_{k-1}$  的第  $p$  行, 第  $q$  行的元素,  $P_k A_{k-1} P_k$  只改变  $A$  的第  $p$  行、 $q$  行、 $p$  列、 $q$  列的元素;  $A_k$  和  $A_{k-1}$  的元素仅在第  $p$  行 (列) 和第  $q$  行 (列) 不同, 它们之间有如下的关系:

$$\begin{cases} a_{ip}^{(k)} = a_{ip}^{(k-1)} \cos \theta + a_{iq}^{(k-1)} \sin \theta = a_{pi}^{(k)} \\ a_{iq}^{(k)} = -a_{ip}^{(k-1)} \sin \theta + a_{iq}^{(k-1)} \cos \theta = a_{qi}^{(k)} \end{cases} \quad i \neq p, q \quad (5.39)$$

$$\begin{cases} a_{pp}^{(k)} = a_{pp}^{(k-1)} \cos^2 \theta + 2a_{pq}^{(k-1)} \sin \theta \cos \theta + a_{qq}^{(k-1)} \sin^2 \theta \\ a_{qq}^{(k)} = a_{pp}^{(k-1)} \sin^2 \theta - 2a_{pq}^{(k-1)} \sin \theta \cos \theta + a_{qq}^{(k-1)} \cos^2 \theta \\ a_{pq}^{(k)} = (a_{pp}^{(k-1)} - a_{qq}^{(k-1)}) \sin \theta \cos \theta + a_{pq}^{(k-1)} (\cos^2 \theta - \sin^2 \theta) \end{cases} \quad (5.40)$$

我们选取  $P_k$ , 使得  $a_{pq}^{(k)} = 0$ , 因此需使  $\theta$  满足

$$\operatorname{tg} 2\theta = \frac{2a_{pq}^{(k-1)}}{a_{pp}^{(k-1)} - a_{qq}^{(k-1)}} \quad (5.41)$$

常将  $\theta$  限制在下列范围内

$$-\frac{\pi}{4} \leq \theta \leq \frac{\pi}{4}$$

如果  $a_{pp}^{(k-1)} - a_{qq}^{(k-1)} = 0$ , 当  $a_{pq}^{(k-1)} > 0$  时, 取  $\theta = \frac{\pi}{4}$ ; 当  $a_{pq}^{(k-1)} < 0$  时, 取  $\theta = -\frac{\pi}{4}$ 。实际上

不需要计算  $\theta$ , 而直接从三角函数关系式计算  $\sin \theta$  和  $\cos \theta$ , 记

$$\begin{cases} y = |a_{ii}^{(k-1)} - a_{jj}^{(k-1)}| \\ x = \operatorname{sgn}(a_{ii}^{(k-1)} - a_{jj}^{(k-1)}) \cdot 2a_{ij}^{(k-1)} \end{cases} \quad (5.42)$$

则



$$\operatorname{tg} 2\theta = \frac{x}{y}$$

当  $|\theta| \leq \frac{\pi}{4}$  时, 有下面三角恒等式:

$$2 \cos^2 \theta - 1 = \cos 2\theta = \frac{1}{\sqrt{1 + \operatorname{tg}^2 2\theta}} = \frac{y}{\sqrt{x^2 + y^2}}$$

于是 
$$2 \cos^2 \theta = 1 + \frac{y}{\sqrt{x^2 + y^2}}$$

$\cos \theta$  始终取正值。关于  $\sin^2 \theta$  的计算有几种方法, 最简单的一种是利用公式  $\sin^2 \theta = 1 - \cos^2 \theta$ , 这个方程有一个缺点, 当  $\cos^2 \theta$  接近于 1 时,  $1 - \cos^2 \theta$  的有效位数就不多了, 为避免这个缺点, 采用下面公式计算  $\sin \theta$ 。

$$\sin 2\theta = 2 \sin \theta \cos \theta = \operatorname{tg} 2\theta \cdot \cos 2\theta = \frac{x}{\sqrt{x^2 + y^2}}$$

由于  $A_k$  的对称性, 实际上只要计算  $A_k$  的上三角元素, 而下三角元素由对称性获得, 这样即节省了计算量, 又能保证  $A_k$  是严格对称的。

雅克比方法的优点是可以容易地计算特征向量, 如果经过  $k$  次旋转变换后, 迭代就停止了, 即:

$$P_k \cdots P_2 P_1 A P_1^T P_2^T \cdots P_K^T = A_k$$

记 
$$P_k = P_1^T P_2^T \cdots P_K^T$$

则 
$$AP = PA_k$$

因为  $A_k$  可以被看作对角阵 (非对角元相当小), 所以矩阵  $P_k$  的第  $j$  列就是特征值  $a_{jj}^{(k)}$  所对应的近似特征向量, 并且所有特征向量都是正交规范化的。

在旋转变换中可以逐步形成  $P_k$ , 记

$$P_0 = I$$

则 
$$P_k = P_{k-1} P_k^T$$

即

$$\begin{cases} P_{ip}^{(k)} = P_{ip}^{(k-1)} \cos \theta + P_{iq}^{(k-1)} \sin \theta \\ P_{iq}^{(k)} = -P_{ip}^{(k-1)} \sin \theta + P_{iq}^{(k-1)} \cos \theta \\ P_{ij}^{(k)} = P_{ij}^{(k-1)} & j \neq p, q \end{cases}$$

这就不需要保存每一次的变换矩阵  $P_k$ , 若不需要计算特征向量, 则 ( ) 式所示的那一步可以省略。

算法:

1. 从  $A^{(k-1)}$  中找出绝对值最大元  $a_{p,q}^{(k-1)}$ ,  $p \neq q$

2. 若  $|a_{pq}^{(k-1)}| \leq \varepsilon$ , 则为对角阵, 停

若  $|a_{pq}^{(k-1)}| > \varepsilon$

(1) 令  $y = |a_{pp}^{(k-1)} - a_{qq}^{(k-1)}|$

$$x = 2a_{pq}^{(k-1)} \cdot \text{sign}(a_{pp}^{(k-1)} - a_{qq}^{(k-1)})$$

(2)  $\cos 2\theta = \frac{y}{\sqrt{x^2 + y^2}}$  当  $y = 0$  时,  $\theta = \frac{\pi}{4}$

$$\sin 2\theta = \frac{x}{\sqrt{x^2 + y^2}} \quad \sin \theta = \cos \theta = \frac{1}{\sqrt{2}}$$

(3)  $C = \cos \theta = \sqrt{\frac{1}{2}(1 + \cos 2\theta)}$

$$S = \sin \theta = \frac{\sin 2\theta}{2C}$$

(4) 
$$\begin{cases} a_{ip} = a_{ip}C + a_{iq}S = a_{pi} \\ a_{iq} = -a_{ip}S + a_{iq}C = a_{qi} \end{cases}$$

$$\begin{cases} a_{pp} = a_{pp}C^2 + 2a_{pq}C \cdot S + a_{qq}S^2 \\ a_{qq} = a_{pp}S^2 - 2a_{pq}C \cdot S + a_{qq}C^2 \\ a_{pq} = (a_{pp} - a_{qq})C \cdot S + a_{pq}(C^2 - S^2) = a_{qp} \end{cases}$$

(5) 计算特征向量

$$\begin{cases} P_{ip} = P_{ip}C + P_{iq}S \\ P_{iq} = P_{ip}S + P_{iq}C \\ P_{ij} = P_{ij} \end{cases} \quad \begin{matrix} (P_0 = I) \\ j \neq p, q \end{matrix}$$

转 1

**定理 8** 设  $A \in R^{n \times n}$  为对称矩阵,  $P = P_{ij} (i \neq j)$  为  $R^n$  上平面旋转阵, 用  $P$  关于  $A$  做正交相似变换, 得

$$C = PAP^T$$

则只要选  $\theta$  满足  $\lg 2\theta = \frac{2a_{ij}}{a_{ii} - a_{jj}} \quad \left( |\theta| \leq \frac{\pi}{4} \right)$  式, 便有  $c_{ij} = c_{ji} = 0$ , 且

$$(1) \quad \|C\|_F^2 = \|A\|_F^2$$

即

$$\sum_{l,s=1}^n c_{ij}^2 = \sum_{l,s=1}^n a_{ij}^2 \quad (5.43)$$

$$(2) \quad c_{ii}^2 + c_{jj}^2 = a_{ii}^2 + a_{jj}^2 + 2a_{ij}^2 \quad (5.44)$$

$$(3) \quad c_{il}^2 + c_{jl}^2 = a_{il}^2 + a_{jl}^2 \quad (l \neq i, j) \quad (5.45)$$

(4) 若用  $D(A) = \sum_{l=1}^n a_{ll}^2$  表示  $A$  对角元素平方和, 则

$$D(C) = D(A) + 2a_{ij}^2 \quad (5.46)$$

(5) 若用  $S(A) = \sum_{l \neq s}^n a_{ls}^2$  表示  $A$  的非对角元素平方和, 则

$$S(C) = S(A) - 2a_{ij}^2 \quad (5.47)$$

证明 (1) 因  $A$  对称,  $A^T A = A^2$ , 利用矩阵迹数性质 (定理 3 中 (1) 式), 有

$$\|A\|_F^2 = \sum_{i,j=1}^n a_{ij}^2 = \text{迹数}(A^T A) = \text{迹数}(A^2) = \sum_{i=1}^n \lambda_i^2(A)$$

$$\|C\|_F^2 = \text{奇迹}(C^T C) = \text{奇迹}(C^2) = \sum_{i=1}^n \lambda_i^2(A)$$

这里  $\lambda_i(A)$ ,  $\lambda_i(C)$  分别为  $A$ 、 $C$  的特征值。

由相似变换不改变矩阵特征值性质, 从  $A$  与  $C$  相似知,

$$\|A\|_F^2 = \|C\|_F^2$$

关于结论 (2) ~ (5), 直接验证计算公式 (5.43) ~ (5.47) 便可得证。

定理 8 除给定  $\theta$  的取法之外, 还说明经过一次正交相似变换:  $A \rightarrow C = PAP^T$ , 矩阵  $A$  与  $C$  的全体元素平方和总数未变, 但  $C$  矩阵对有元素平方和比  $A$  矩阵对角元素平方和增加了  $2a_{ij}^2$ 。这一性质, 将为研究雅可比旋转法的收敛性提供了保证。

从理论上讲, 每一次变换都会将  $A$  的两个非对角元素化为零, 则经过  $\frac{n^2-n}{2}$  次变换便可将  $A$  化为对角阵。但事实上, 如果已将  $a_{ij} = a_{ji}$  化为零元素, 在下次变换时,  $(i, j)$ 、 $(j, i)$  位置的元素可能又会变成非零, 这就要考虑雅可比方法是否收敛的问题。

如果  $S(A_k) = \sum_{l \neq s}^n (a_{ls}^{(k)})^2$  为  $A_k$  的非对角元素的平方和, 当  $k \rightarrow \infty$  时,  $S(A_k) \rightarrow 0$ , 则矩阵序列

$\{A_k\}$  便会收敛到对角矩阵。

**定理 9** 设  $A = (a_{ij}) \in R^{n \times n}$  为对称矩阵, 则由雅可比方法产生的矩阵序列  $\{A_k\}$ ,

$$A_{k+1} = P_k \cdots P_2 P_1 A P_1^T P_2^T \cdots P_k^T$$

收敛于对角阵  $D = \text{diag}(\lambda_1, \cdots, \lambda_n)$ , 其中  $\lambda_1, \cdots, \lambda_n$  为  $A$  的特征值的近似。

证明 只需证极限式

$$\lim_{k \rightarrow \infty} S(A_k) = 0 \quad (5.48)$$

成立即可。由定理 8 结论 (5) 知

$$S(A_{k+1}) = S(A_k) - 2(a_{ij}^{(k)})^2 \quad (5.49)$$

其中

$$|a_{ij}^{(k)}| = \max_{l \neq s} |a_{ls}^{(k)}|$$

而

$$S(A_k) = \sum_{l \neq s}^n (a_{ls}^{(k)})^2 \leq n(n-1) |a_{ij}^{(k)}|^2$$

即

$$\frac{S(A_k)}{n(n-1)} \leq |a_{ij}^{(k)}|^2 \quad (5.50)$$

将 (5.50) 式代入 (5.49) 式, 得

$$\begin{aligned} S(A_{k+1}) &\leq S(A_k) - \frac{2}{n(n-1)} S(A_k) \\ &= \left(1 - \frac{2}{n(n-1)}\right) S(A_k) \leq \cdots \\ &\leq \left(1 - \frac{2}{n(n-1)}\right)^k S(A) \rightarrow 0 \quad (k \rightarrow \infty) \end{aligned}$$

即 (5.48) 式成立。

例 用雅可比旋转法求矩阵  $A$  的特征值及特征向量,

$$A = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}$$

解 先将  $a_{12} = a_{21}$  化为零, 取  $i = 1, j = 2$ ,

$$2\theta = \frac{\pi}{2}$$

于是

$$\cos \theta = \sin \theta = \frac{1}{\sqrt{2}}$$

确定平面旋转阵  $P_{12}$ ,

$$P_{12} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

则

$$A_2 = P_{12} A P_{12}^T = \begin{bmatrix} 1 & 0 & -\frac{1}{\sqrt{2}} \\ 0 & 3 & -\frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 2 \end{bmatrix}$$

再取  $i=1, j=3$ , 则  $\operatorname{tg} 2\theta = \sqrt{2}$ , 从而

$$\sin \theta \approx 0.45969, \quad \cos \theta \approx 0.88808$$

$$P_{13} = \begin{bmatrix} 0.88808 & 0 & 0.45969 \\ 0 & 1 & 0 \\ -0.45969 & 0 & 0.88808 \end{bmatrix}$$

$$A_3 = P_{13} A_2 P_{13}^T = \begin{bmatrix} 0.63398 & -0.32505 & 0 \\ -0.32505 & 3 & 0.62797 \\ 0 & -0.62797 & 2.36603 \end{bmatrix}$$

如此这样, 经过九次旋转变换, 得

$$A_{10} = \begin{bmatrix} 0.58578 & 0.00000 & 0.00000 \\ 0.00000 & 2.00000 & 0.00000 \\ 0.00000 & 0.00000 & 3.41421 \end{bmatrix}$$

$A$  的特征值的近似值为

$$\lambda_1 = 0.58578, \quad \lambda_2 = 2.00000, \quad \lambda_3 = 3.41421$$

相应的特征向量近似值为

$$\begin{aligned} P^T &= P_1^T P_2^T P_3^T = [p_1, p_2, p_3] \\ &= \begin{bmatrix} 0.50000 & 0.70710 & 0.50000 \\ 0.70710 & 0.00000 & -0.70710 \\ 0.50000 & -0.70710 & 0.50000 \end{bmatrix} \end{aligned}$$

其中  $p_i$  相当于  $\lambda_i$  的特征向量。

雅可比方法是求实对称矩阵全部特征值和特征向量的一个较适用的方法, 结果精度高, 且求到的特征向量正交性好, 但计算量较大, 并破坏原矩阵的稀疏性。因此适用于低阶满矩阵的情形。

### 3 雅可比过关法

前面介绍的雅可比法每变换一次，都要在非对角元素中扫描选取绝对值最大者，这样难免要增加很大工作量，为提高计算运行速度，可将此方法进行改进。

如先计算矩阵  $A$  的非对角元素平方和  $S(A)$  记为

$$\gamma_0 = \left( \sum_{l \neq s}^n a_{ls}^2 \right)^{1/2} = [S(A)]^{1/2}$$

取值  $\gamma_1 = \frac{\gamma_0}{n}$ ，然后对  $A$  的非对角元素（因  $A$  对称，故可只考虑上三角或下三角部分）进行扫描，

$$\begin{array}{cccc} a_{12} & a_{13} & \cdots & a_{1n} \\ & a_{23} & \cdots & a_{2n} \\ & & \ddots & \vdots \\ & & & a_{(n-1)n} \end{array}$$

对  $|a_{ij}| \geq \gamma_1$  的元素进行旋转变换，将  $a_{ij}$  化为零，否则，过关（不变换）。当所有非对角元素绝对值都小于  $\gamma_1$  后，缩小关口，取  $\gamma_2 = \frac{\gamma_1}{n}$ ，重复前面的步骤，按此方法，关口  $\gamma_1 \geq \gamma_2 \geq \cdots$  不断缩小，直到满足

$$\gamma_r = \frac{1}{n^r} S(A) \leq \varepsilon$$

停止计算。这种方法称为雅可比过关法。

### 5.5 QR 算法

乘幂法（或子空间迭代法）可用于求矩阵按模最大的一个（或前几个）特征值和特征向量，雅可比旋转法可用于求实对称矩阵的全部特征值及特征向量。而对一般实矩阵  $A \in R^{n \times n}$ ，若  $A$  可逆，则可用 QR 方法求  $A$  的全部特征值。

QR 方法是一种变换方法，按 9.3 节中的斯密特正交化过程，可将  $A = A_1$  进行正交分解为正交矩阵  $Q_1$  和上三角矩阵  $R_1$  的乘积

$$A_1 = Q_1 R_1$$

交换因式矩阵  $Q_1, R_1$  得

$$A_2 = R_1 Q_1$$

由于  $Q_1$  为正交阵， $Q_1^T = Q_1^{-1}$ ，于是

$$A_2 = Q_1^T A_1 Q_1$$

这表明矩阵  $A_2$  与  $A_1$  正交相似。

用  $A_2$  代替  $A_1$ ，重复上述步骤可得  $A_3$ ，一般地，若已知矩阵  $A_k$ ，且已分解为

$$A_k = Q_k R_k \quad (5.51)$$

其中  $Q_k$  为正交阵， $R_k$  为上三角阵，将  $Q_k, R_k$  换序相乘，得

$$A_{k+1} = R_k Q_k = Q_k^T A_k Q_k \quad (5.52)$$

称公式 (5.51)、(5.52) 为求矩阵  $A$  全部特征值的  $QR$  算法。

**定理 10**  $A \in R^{n \times n}$ , 由公式 (5.51)、(5.52) 确定的  $QR$  算法产生的矩阵序列  $\{A_k\}$  具有两个基本特征,

- (1) 正交相似传递性, 即对每一个  $k$ ,  $A_k$  与  $A$  正交相似;
- (2) 若记

$$\tilde{Q} = Q_1 Q_2 \cdots Q_k, \quad \tilde{R} = R_k \cdots R_2 R_1 \quad (5.53)$$

则  $A^k$  有  $QR$  分解式

$$A^k = \tilde{Q} \tilde{R}$$

**证明** (1) 因  $A_2 = Q_1^T A_1 Q_1$ , 则  $A_2$  与  $A$  正交相似,

$$\begin{aligned} A_3 &= Q_{k-1}^T A_{k-1} Q_{k-1} = Q_{k-1}^T Q_{k-2}^T A_{k-2} Q_{k-2} Q_{k-1} \\ &= \cdots = Q_{k-1}^T \cdots Q_2^T Q_1^T A_1 Q_1 Q_2 \cdots Q_{k-1} \\ &= (Q_1 Q_2 \cdots Q_{k-1})^T A (Q_1 Q_2 \cdots Q_{k-1}) = \tilde{Q}_{k-1}^T A \tilde{Q}_{k-1} \end{aligned} \quad (5.54)$$

其中  $\tilde{Q} = Q_1 Q_2 \cdots Q_{k-1}$ , 则  $\tilde{Q}_{k-1}$  为正交矩阵, 于是  $A_k$  与  $A$  正交相似 (此时  $A_k$  与  $A$  有相同的特征值, 见定理 1)。

(2) 用归纳法证:

当  $k=1$  时,  $A$  的正交分解为

$$A = \tilde{Q}_1 \tilde{R}_1 = Q_1 R_1$$

设  $A^{k-1}$  的正交分解式为

$$A^{k-1} = \tilde{Q}_{k-1} \tilde{R}_{k-1}$$

则由 (5.54) 式知

$$A \tilde{Q}_{k-1} = \tilde{Q}_{k-1} A_k$$

从而

$$A^k = A(A^{k-1}) = A(\tilde{Q}_{k-1} \tilde{R}_{k-1}) = \tilde{Q}_{k-1} A \tilde{R}_{k-1}$$

再利用分解式 (5.51)

$$A_k = Q_k R_k$$

得

$$A^k = \tilde{Q}_{k-1} Q_k R_k \tilde{R}_{k-1} = \tilde{Q}_k \tilde{R}_k$$

**定义 2** 设  $\{A_k\}$  为  $R^{n \times n}$  上的矩阵序列, 如果当  $k \rightarrow \infty$  时,  $A_k$  在某种意义下收敛到上 (或下) 三角型矩阵 (或分块上 (或下) 三角型矩阵), 其中, 上 (或下) 三角 (分块三角) 型矩阵的对角元素 (或对角子块) 为确定极限值, 无论对角线 (或对角子块) 上方 (或下方) 的元素 (或子块) 是否有极限存在, 都称  $\{A_k\}$  是本质收敛的。

如

$$A_k \xrightarrow{\text{本质收敛}} \begin{bmatrix} \lambda_1 & * & \cdots & * \\ & \lambda_2 & \ddots & \vdots \\ & & \ddots & * \\ & & & \lambda_n \end{bmatrix}, \quad k \rightarrow \infty$$

或分块形式

$$A_k \xrightarrow{\text{本质收敛}} \begin{bmatrix} \Lambda_1 & * & \cdots & * \\ & \Lambda_2 & \ddots & \vdots \\ & & \ddots & * \\ & & & \Lambda_r \end{bmatrix}, \quad k \rightarrow \infty$$

其中  $\Lambda_i (i = 1, 2, \dots, r)$  为分块矩阵, \*号位置元素或子块可以没有极限。

这里不加证明, 给出一定条件下  $QR$  算法的收敛定理。

**定理 11** (QR 方法的收敛性) 设  $A = (a_{ij}) \in R^{n \times n}$ ,

1°  $A$  的  $n$  个特征值  $\lambda_1, \dots, \lambda_n$  满足

$$|\lambda_1| > |\lambda_2| > \cdots > |\lambda_n| > 0$$

2° 有可逆矩阵  $P$ , 将  $A$  相似变换为对角阵  $J$ , 即

$$A = PJP^{-1}$$

其中

$$J = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix}$$

3° 矩阵  $P^{-1}$  有三角分解

$$P^{-1} = LU$$

其中  $L$  为单位下三角阵,  $U$  为上三角阵, 则由  $A_1 = A$  进行的  $QR$  算法序列  $\{A_k\}$  本质收敛于上三角矩阵, 即

$$A_k \xrightarrow{\text{本质收敛}} \begin{bmatrix} \lambda_1 & * & \cdots & * \\ & \lambda_2 & \ddots & \vdots \\ & & \ddots & * \\ & & & \lambda_n \end{bmatrix}, \quad k \rightarrow \infty$$

**定理 12** 若  $A \in R^{n \times n}$  为对称矩阵且满足定理 11 中条件, 则由  $QR$  算法确定的  $\{A_k\}$ , 收敛于对角阵  $\Lambda$ , 且  $\Lambda$  对角线元素为  $A$  的全部特征值。

例 求矩阵  $A$  的特征值

$$A = \begin{bmatrix} 5 & -2 & -5 & -1 \\ 1 & 0 & -3 & 2 \\ 0 & 2 & 2 & -3 \\ 0 & 0 & 1 & -2 \end{bmatrix}$$

解 矩阵  $A$  的特征多项式为

$$\lambda^4 - 5\lambda^3 + 7\lambda^2 - 7\lambda - 20 = 0$$



其特征值为

$$\lambda_1 = 4, \quad \lambda_{2,3} = 1 \pm 2i, \quad \lambda_4 = -1$$

用  $QR$  算法求  $A$  的特征值。

令  $A = A_1$ ，进行正交  $QR$  分解。由施密特正交化过程，得

$$A_1 = Q_1 R_1$$

其中

$$Q_1 = \begin{bmatrix} 0.9806 & -0.0377 & 0.1923 & -0.1038 \\ 0.1961 & 0.1887 & -0.8804 & -0.4192 \\ 0 & 0.9813 & 0.1761 & 0.0740 \\ 0 & 0 & 0.3962 & -0.8989 \end{bmatrix}$$

$$R = \begin{bmatrix} 5.0992 & -1.9612 & -5.4912 & -0.3922 \\ 0 & 2.0381 & 1.5852 & -2.5288 \\ 0 & 0 & 2.5242 & -3.2736 \\ 0 & 0 & 0 & 0.7822 \end{bmatrix}$$

将  $Q_1$ 、 $R_1$  逆序相乘，得  $A_2$

$$A_2 = \begin{bmatrix} 4.6157 & 5.9508 & 1.5922 & 0.2390 \\ 0.3997 & 1.9401 & -2.5171 & 1.5361 \\ 0 & 2.4770 & -0.8525 & 3.1294 \\ 0 & 0 & 0.3099 & -0.7031 \end{bmatrix}$$

接下来，分解  $A_2 = Q_2 R_2$ ，重复上述步骤，迭代 11 次得  $A_{12}$ ，

$$A_{12} = \begin{bmatrix} 4.000 & * & * & * \\ & 1.8789 & -3.5910 & * \\ & 1.3290 & 0.1211 & * \\ & & & -1.000 \end{bmatrix}$$

$$= \begin{bmatrix} \Lambda_1 & * & * \\ & \Lambda_2 & * \\ & & * \end{bmatrix} \quad (\text{分块上三角阵})$$

其中  $\Lambda_1$ 、 $\Lambda_2$  为一阶子块， $\Lambda_2$  为二阶子块，从而  $A$  的特征值的近似值

$$\tilde{\lambda}_1 = 4.000, \quad \tilde{\lambda}_4 = -1.000$$

而  $\tilde{\lambda}_2, \tilde{\lambda}_3$  为特征方程

$$|\Lambda_2 - \lambda I| = \begin{vmatrix} 1.8789 - \lambda & -3.5910 \\ 1.3290 & 0.1211 - \lambda \end{vmatrix} = 0$$

的两个根，解出为  $\tilde{\lambda}_{2,3} = 1 \pm 2i$ 。

通常  $QR$  算法的收敛速度是线性的，且计算工作量较大，实际应用不很方便。

**定义 3** 设  $B = (b_{ij}) \in R^{n \times n}$ ，如果当  $i > j + 1$  时， $b_{ij} = 0$ ，则称  $B$  为上赫森伯格 (Hessenberg)

阵，形如，

$$B = \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ & \ddots & \ddots & \vdots \\ & & b_{nn-1} & b_{nn} \end{bmatrix}$$

对一般矩阵  $A \in R^{n \times n}$ ，首先约化成上赫森伯格阵  $B$ （可采用豪斯荷尔德（Horseholder）变换），然后再用  $QR$  方法计算矩阵  $B$  的全部征值。这样可提高收敛速度，达到平方收敛。当矩阵  $A \in R^{n \times n}$  对称时，上赫森伯格阵  $B$  便为对称的三对角矩阵，这时用  $QR$  算法求解，可达三阶收敛速度且算法稳定。

本章介绍了求矩阵按模最大、最小特征值及相应特征向量的乘幂法与反幂法公式，分析了乘幂法乘幂过程及特点，并介绍了工程上常用的子空间迭代法。乘幂法（子空间迭代法）适用于求解大型稀疏矩阵的最大（最小）特征值问题。

雅可比旋转法利用矩阵正交相似变换的原理，给出了求实对称矩阵的全部特征值及特征向量的方法，且算法稳定，精度较高，但计算量较大，破坏稀疏性，通常用来求解阶数低且稠密的矩阵特征问题。

$QR$  算法是计算矩阵（中小型矩阵）全部特征值的最有效方法之一，主要用于计算上赫森伯格阵和对称三对角阵的全部特征值问题。如果配合原点位移，采用双步  $QR$  方法，效果会更好。

本章应掌握的基本问题：

- （1）乘幂法、反幂法计算公式，特点及适用范围；
- （2）分析乘幂过程， $|\lambda_1| > |\lambda_2|$  及  $\lambda_2 = -\lambda_1$  且  $|\lambda_2| > |\lambda_3|$  的情况；
- （3）雅可比旋转方法的基本思想及计算步骤。

## 第五章 习题

1. 用乘幂法计算下列矩阵的主特征值及相应的特征向量，

$$A_1 = \begin{bmatrix} 4 & 2 & 2 \\ 2 & 5 & 1 \\ 2 & 1 & 6 \end{bmatrix}, \quad A_2 = \begin{bmatrix} 7 & 3 & -2 \\ 3 & 4 & -1 \\ -2 & -1 & 3 \end{bmatrix}$$

当特征值有 3 位小数相同时，迭代终止。

2. 已知

$$A = \begin{bmatrix} -11 & 11 & 1 \\ 11 & 9 & -2 \\ 1 & -2 & 13 \end{bmatrix}$$

先用乘幂法作适当迭代，然后用带有原点移位的乘幂法求按模最大特征值（可参考取 $\lambda_0=15$ ）。

3. 用反幂法求下列矩阵的指定特征值及特征向量，

$$(1) \quad A = \begin{bmatrix} 4 & 1 & 4 \\ 1 & 10 & 1 \\ 4 & 1 & 10 \end{bmatrix}, \quad \text{接近 } \lambda = 12 \text{ 的特征值与特征向量,}$$

$$(2) \quad A = \begin{bmatrix} 6 & 2 & 1 \\ 2 & 3 & 1 \\ 1 & 1 & 1 \end{bmatrix} \quad \text{接近 } 6 \text{ 的特征值与特征向量。}$$

4. 用雅可比方法求下列矩阵的全部特征值及特征向量，

$$A_1 = \begin{bmatrix} 1 & 1 & 0.5 \\ 1 & 1 & 0.25 \\ 0.5 & 0.25 & 2 \end{bmatrix}, \quad A_2 = \begin{bmatrix} 4 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 1 \end{bmatrix}$$

5. 用 QR 算法求下列矩阵的特征值（用 Matlab 内部函数）

$$A_1 = \begin{bmatrix} 2 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 2 \end{bmatrix}, \quad A_2 = \begin{bmatrix} 0 & 5 & 0 & 0 & 0 & 0 \\ 1 & 0 & 4 & 0 & 0 & 0 \\ 0 & 1 & 0 & 3 & 0 & 0 \\ 0 & 0 & 1 & 0 & 2 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

### 上机计算题

#### 一、编制通用子程序

- (1) 规范化乘幂法计算公式；
- (2) 雅可比过关法计算公式；
- (3) 斯密特正交化过程计算公式；
- (4) QR 方法公式。

二、用乘幂法（子程序（1））求下列矩阵的按模最大特征值及特征向量。

$$A_1 = \begin{bmatrix} \frac{49}{8} & -\frac{131}{8} & -\frac{43}{4} \\ \frac{11}{8} & -\frac{17}{8} & -\frac{9}{4} \\ -\frac{1}{2} & \frac{7}{2} & 3 \end{bmatrix}, \quad A_2 = \begin{bmatrix} 7 & 3 & -2 \\ 3 & 4 & -1 \\ -2 & -1 & 3 \end{bmatrix}$$

三、用雅可比过关法求习题九第五题中矩阵  $A_1$  的全部特征值与特征向量。