数值分析学习指导

关 治 编

清华大学出版社

数值分析学习指导

关治　编

清华大学出版社
北京

内容简介

本书是与数值分析(或计算方法)课程学习配套的辅导材料. 书中总结了此课程各部分的基本内容和要点,通过典型例题阐述了对各种概念的正确理解、数值方法的合理使用以及各种性质的分析,这些典型例题既包括解题技巧,也包括方法的具体实现. 对于一些容易混淆的问题,分析了出错的原因并给出正确的解法. 各章还包括复习题和计算实习题,方便读者复习、理解及在计算机上实际计算.

本书适合学习数值分析课程的研究生和本科生使用,也可供从事科学与工程计算的技术人员参考.

图书在版编目(CIP)数据

数值分析学习指导/关治编. —北京: 清华大学出版社,2008.11

ISBN 978-7-302-18638-0

Ⅰ. 数…　Ⅱ. 关…　Ⅲ. 数值计算—高等学校—教学参考资料　Ⅳ. O241

中国版本图书馆 CIP 数据核字(2008)第 147358 号

责任编辑: 刘　颖
责任校对: 赵丽敏
责任印制: 李红英
出版发行: 清华大学出版社　　地　址: 北京清华大学学研大厦 A 座
http://www.tup.com.cn　　邮　编: 100084
社 总 机: 010-62770175　　邮　购: 010-62786544
投稿与读者服务: 010-62776969,c-service@tup.tsinghua.edu.cn
质 量 反 馈: 010-62772015,zhiliang@tup.tsinghua.edu.cn
印 刷 者: 北京市清华园胶印厂
装 订 者: 三河市金元印装有限公司
经　销: 全国新华书店
开　本: 185×230　印　张: 13.25　字　数: 286 千字
版　次: 2008 年 11 月第 1 版　印　次: 2008 年 11 月第1 次印刷
印　数: 1～4000
定　价: 20.00 元

本书如存在文字不清、漏印、缺页、倒页、脱页等印装质量问题,请与清华大学出版社出版部联系调换。联系电话:(010)62770177 转 3103　产品编号:028935-01

前　　言

本书是理工科各专业数值分析(或计算方法)课程的教学辅助读物.适合学习这类课程的本科生和研究生使用,也适合一些参加考试(例如某些专业的同等学历申请学位或博士生入学考试等)的考生复习备考用.目前,这类课程有不同的学时和要求.本书基本定位于针对硕士研究生课程或工程硕士专业学位课程.这些课程要求读者既能对数值分析的基本理论有一定的理解,又能正确应用各种方法数值求解数学问题.所以我们选材比较基本,着重各种典型例题的分析,同时也有少量的内容和例题较为深入,适合要求稍高的读者.对于学时较少或对课程要求稍低的读者,可以只参考本书各章适当的内容和例题中相对基本的部分.

本书内容包括数值分析引论、解线性代数方程组的直接解法和迭代解法、非线性方程和方程组的数值解法、矩阵特征值问题的数值解法、插值法、函数逼近、数值积分与数值微分、常微分方程初值问题的数值解法.每章都有基本内容提要,而主要部分是典型例题分析,此外还有复习题和计算实习题,便于全面掌握各章的知识、理论和方法.在第1章还有一些关于线性代数知识的内容,这是以后各章所要用到的,把这些问题集中在第1章,对理解各章内容和全面复习课程会起到一定作用.

本书内容的选取参照了众多的数值分析教科书和参考书.例如,各章基本内容提要的详细分析可在参考文献[1～3]中找到.有些例题分析结合了我们多年的教学经验,分析了一些学生经常会碰到的难点或者是常见的错误.除了一些性质的分析外,还包括了一些数值计算的题目,便于读者更好掌握各章的方法.各章还有少量计算实习题,供读者上机计算,我们推荐大家使用 MATLAB 的工具.

陆金甫教授参与了本书的讨论,进行了十分有益的交流.事实上本书是我们合作编写的教科书(参考文献[1,2])自然的延续和教学的一点总结.

希望本书对学习数值分析课程的读者会有所帮助,也十分感谢读者和同行专家的反馈信息.

关　治

2007 年 12 月

目　录

第1章　数值分析引论

1.1　基本内容提要

1.1.1　数值计算的误差

误差的分类　舍入误差、截断误差、误差在计算过程的传播.

定义 1.1　设 x_A 是实数 x 的一个近似值，$x-x_A$ 称为 x_A 的**绝对误差**(**简称误差**). 若 $x\neq 0$，$\dfrac{x-x_A}{x}$ 称为 x_A 的**相对误差**. 如果有估计

$$|x-x_A|\leqslant \varepsilon_A,$$

ε_A 称为 x_A 的**绝对误差界**. 若 $x\neq 0$，$\dfrac{\varepsilon_A}{|x|}$ 称为 x_A 的**相对误差界**.

定义 1.2　设 x_A 是 x 的一个近似值，写成

$$x_A=\pm 10^k\times 0.a_1a_2\cdots a_n\cdots,$$

其中 $a_i(i=1,2,\cdots)$ 是 $0,1,\cdots,9$ 中的一个数字，$a_1\neq 0$，k 为整数. 如果

$$|x-x_A|\leqslant 0.5\times 10^{k-n},$$

则称 x_A 为 x 的**具有 n 位有效数字的近似值**.

也就是说，如果 x_A 的误差的绝对值不超过其数字中某一位上的半个单位，而从第一位非零数字到该位数字共有 n 位，x_A 就有 n 位有效数字.

定理 1.1　设 x_A 有定义 1.2 中的形式. 如果 x_A 有 n 位有效数字，则

$$\frac{|x-x_A|}{|x_A|}\leqslant \frac{1}{2a_1}\times 10^{1-n}.$$

反之，如果

$$\frac{|x-x_A|}{|x_A|}\leqslant \frac{1}{2(a_1+1)}\times 10^{1-n},$$

则 x_A 至少有 n 位有效数字.

函数值的误差估计　若 $x_1,x_2,\cdots,x_n$ 的近似值为 $x_{1A},x_{2A},\cdots,x_{nA}$，$f$ 是关于 $x_1,x_2,\cdots,x_n$ 的具有二阶连续偏导数的 n 元函数，则

$$|f(x_1,x_2,\cdots,x_n)-f(x_{1A},x_{2A},\cdots,x_{nA})|\leqslant \sum_{k=1}^{n}\left|\left(\frac{\partial f}{\partial x_k}\right)_A\right||x_k-x_{kA}|,$$

其中$\left(\frac{\partial f}{\partial x_k}\right)_A=\frac{\partial}{\partial x_k}f(x_{1A},x_{2A},\cdots,x_{nA})$. 当$n=1$时,公式成为

$$|f(x)-f(x_A)|\leqslant|f'(x_A)||x-x_A|,$$

即函数值$f(x)$的近似值$f(x_A)$的绝对误差界为$|f'(x_A)||x-x_A|$.

1.1.2 避免误差危害

如果一种数值计算方法初始数据的微小改变将使得最后结果的改变是微小的,此方法是**数值稳定**的,否则就是数值不稳定的. 数值不稳定是一种误差危害现象,应该避免.

此外,还应注意避免使有效数字损失的计算方法. 例如,避免相近数的相减,适当调整计算公式或计算次序,尽量减少运算次数等.

1.1.3 线性代数的一些基础知识

1. 矩阵的特征值和相似变换

$\mathbb{R}^{n\times n}$中的矩阵$\boldsymbol{A}$的谱半径是

$$\rho(\boldsymbol{A})=\max_{\lambda\in\sigma(\boldsymbol{A})}|\lambda|,$$

其中$\sigma(\boldsymbol{A})$是$\boldsymbol{A}$的谱,即$\boldsymbol{A}$的全体特征值的集合.

定理 1.2 设$\boldsymbol{A}\in\mathbb{R}^{n\times n}$,则$\boldsymbol{A}$可以通过相似变换化为对角矩阵的充分必要条件是$\boldsymbol{A}$有$n$个线性无关的特征向量.

定理 1.3 设$\boldsymbol{A}\in\mathbb{R}^{n\times n}$,$\boldsymbol{A}$有$s$个不同的特征值$\lambda_1,\lambda_2,\cdots,\lambda_s(s\leqslant n)$,则$\boldsymbol{A}$可通过相似变换化为Jordan标准形$\boldsymbol{J}$,即存在可逆矩阵$\boldsymbol{P}$,使

$$\boldsymbol{P}^{-1}\boldsymbol{AP}=\boldsymbol{J}=\begin{bmatrix}\boldsymbol{J}_1 & & & \\ & \boldsymbol{J}_2 & & \\ & & \ddots & \\ & & & \boldsymbol{J}_s\end{bmatrix},$$

其中$\boldsymbol{J}_i\in\mathbb{R}^{n_i\times n_i}$,它对应于特征值$\lambda_i(i=1,2,\cdots,s)$,$n_1+n_2+\cdots+n_s=n$. $\boldsymbol{J}_i$有下面的形式:

$$\boldsymbol{J}_i=\begin{bmatrix}\boldsymbol{J}_{i1} & & & \\ & \boldsymbol{J}_{i2} & & \\ & & \ddots & \\ & & & \boldsymbol{J}_{im_i}\end{bmatrix},\quad 其中\ \boldsymbol{J}_{ik}=\begin{bmatrix}\lambda_i & 1 & & & \\ & \lambda_i & 1 & & \\ & & \ddots & \ddots & \\ & & & \lambda_i & 1 \\ & & & & \lambda_i\end{bmatrix},\quad k=1,2,\cdots,m_i.$$

如果$\boldsymbol{J}_{ik}$是1×1矩阵,则它只有对角元λ_i.

2. 线性空间、内积

数域$\mathbb{P}$(本书一般指$\mathbb{R}$或$\mathbb{C}$)上的**线性空间**指在其上定义了加法和数乘两种运算的非空

集合 V，它满足加法和数乘的规则. 在数值分析课程常见的线性空间有 n 维向量空间 $\mathbb{R}^n$（或 $\mathbb{C}^n$），$m\times n$ 矩阵的空间 $\mathbb{R}^{m\times n}$（或 $\mathbb{C}^{m\times n}$）；$[a,b]$ 上连续和有 n 阶连续导数的函数空间 $C[a,b]$，$C^n[a,b]$；以及不超过 n 次的多项式函数的空间 $\mathcal{P}_n[a,b]$ 等.

定义 1.3 对数域 $\mathbb{P}$ 上的线性空间 V 中任意的元素 u 和 v，若存在惟一的数（记为 (u,v)）与之对应，满足

(1) $(u+v,w)=(u,w)+(v,w),\quad \forall u,v,w\in V$，

(2) $(\alpha u,v)=\alpha(u,v),\quad \forall u,v\in V,\alpha\in\mathbb{P}$，

(3) $(u,v)=\overline{(v,u)},\quad \forall u,v\in V$，

(4) $(u,u)\geqslant 0,\forall u\in V$，且 $(u,u)=0\Leftrightarrow u$ 为 V 的零元素，

则称 (u,v) 为元素 u 与 v 的**内积**. 定义了内积的线性空间 V 称为**内积空间**.

$\mathbb{R}^n$ 中内积通常如下定义：设 $\boldsymbol{x}=(x_1,x_2,\cdots,x_n)^{\mathrm{T}}$，$\boldsymbol{y}=(y_1,y_2,\cdots,y_n)^{\mathrm{T}}$，

$$(\boldsymbol{x},\boldsymbol{y})=\sum_{i=1}^n x_i y_i=\boldsymbol{y}^{\mathrm{T}}\boldsymbol{x}.$$

或者是带权的内积：

$$(\boldsymbol{x},\boldsymbol{y})_w=\sum_{i=1}^n w_i x_i y_i,$$

其中权系数 $w_i>0(i=1,2,\cdots,n)$.

$C[a,b]$ 上可定义带权内积

$$(f,g)_\rho=\int_a^b \rho(x)f(x)g(x)\mathrm{d}x,\quad \forall f,g\in C[a,b].$$

式中 $\rho(x)$ 是权函数，满足

(1) $\rho(x)\geqslant 0,\quad \forall x\in(a,b)$；

(2) $\int_a^b x^k\rho(x)\mathrm{d}x$ 存在，$k=0,1,\cdots$；

(3) 若对 $[a,b]$ 上的非负连续函数 g，$\int_a^b\rho(x)g(x)\mathrm{d}x=0$ 成立，则 $g(x)\equiv 0$（这保证了在 $[a,b]$ 的任一开子区间上 $\rho(x)$ 不恒为零）.

当 $\rho(x)\equiv 1$ 时，记

$$(f,g)=\int_a^b f(x)g(x)\mathrm{d}x,\quad \forall f,g\in C[a,b].$$

3. 范数

定义 1.4 若对于数域 $\mathbb{P}$ 上的线性空间 V 上任意的元素 u，都有实数 $\|u\|$ 与之对应，满足

(1) **正定性**：$\|u\|\geqslant 0,\forall u\in V$，且

$$\|u\|=0\Leftrightarrow u \text{ 是 } V \text{ 中的零元素；}$$

(2) **齐次性**：$\|\alpha u\|=|\alpha|\|u\|$，$\forall u\in V,\alpha\in\mathbb{P}$；

(3) **三角不等式**：$\|u+v\|\leqslant\|u\|+\|v\|$，$\forall u,v\in V$，

则称 $\|\cdot\|$ 是 V 上的范数.

$\mathbb{R}^n$(或$\mathbb{C}^n$)上常用的范数有三种：

$$\|x\|_1=\sum_{i=1}^{n}|x_i|,$$

$$\|x\|_2=\left(\sum_{i=1}^{n}|x_i|^2\right)^{1/2}=\sqrt{(x,x)},$$

$$\|x\|_\infty=\max_{1\leqslant i\leqslant n}|x_i|,$$

其中 $\boldsymbol{x}=(x_1,x_2,\cdots,x_n)^{\mathrm{T}}$.

$C[a,b]$上常用的范数有

$$\|f\|_2=\left[\int_a^b\rho(x)[f(x)]^2\mathrm{d}x\right]^{1/2}=\sqrt{(f,f)_\rho},$$

$$\|f\|_\infty=\max_{x\in[a,b]}|f(x)|.$$

$\mathbb{R}^{n\times n}$(或$\mathbb{C}^{n\times n}$)上的矩阵范数，除了满足一般范数定义的三个条件外，还要求满足

$$\|\boldsymbol{AB}\|\leqslant\|\boldsymbol{A}\|\|\boldsymbol{B}\|,\quad\forall\boldsymbol{A},\boldsymbol{B}\in\mathbb{R}^{n\times n}.$$

常用的矩阵范数有

$$\|\boldsymbol{A}\|_F=\left(\sum_{i=1}^{n}\sum_{j=1}^{n}|a_{ij}|^2\right)^{1/2},$$

$$\|\boldsymbol{A}\|_\infty=\max_{1\leqslant i\leqslant n}\sum_{j=1}^{n}|a_{ij}|,$$

$$\|\boldsymbol{A}\|_1=\max_{1\leqslant j\leqslant n}\sum_{i=1}^{n}|a_{ij}|,$$

$$\|\boldsymbol{A}\|_2=[\rho(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})]^{1/2}.$$

$\|\boldsymbol{A}\|_\infty$，$\|\boldsymbol{A}\|_1$，$\|\boldsymbol{A}\|_2$ 分别是由向量范数 $\|\boldsymbol{x}\|_\infty$，$\|\boldsymbol{x}\|_1$，$\|\boldsymbol{x}\|_2$ 导出的矩阵范数，也称为从属于向量范数的矩阵范数，满足

$$\|\boldsymbol{A}\|_p=\max_{x\neq 0}\frac{\|\boldsymbol{Ax}\|_p}{\|\boldsymbol{x}\|_p}=\max_{\|\boldsymbol{x}\|_p=1}\|\boldsymbol{Ax}\|_p,\quad p=1,2,\infty,$$

$$\|\boldsymbol{Ax}\|_p\leqslant\|\boldsymbol{A}\|_p\|\boldsymbol{x}\|_p,\quad p=1,2,\infty.$$

定理 1.4　设 $\|\cdot\|$ 为$\mathbb{R}^{n\times n}$上任意一种(从属或非从属)范数，则对任意的 $\boldsymbol{A}\in\mathbb{R}^{n\times n}$ 有

$$\rho(\boldsymbol{A})\leqslant\|\boldsymbol{A}\|.$$

定理 1.5　$\mathbb{R}^n$ 上各种范数都是彼此等价的. 即对任意两种范数 $\|\cdot\|_\alpha$ 和 $\|\cdot\|_\beta$，存在实数 M,m 满足 $M\geqslant m>0$，使得

$$m\|\boldsymbol{x}\|_\alpha\leqslant\|\boldsymbol{x}\|_\beta\leqslant M\|\boldsymbol{x}\|_\alpha,\quad\forall\boldsymbol{x}\in\mathbb{R}^n.$$

同样，$\mathbb{R}^{n\times n}$上各种范数也是彼此等价的.

定理 1.6　设 $\|\cdot\|$ 为$\mathbb{R}^{n\times n}$上一种从属于向量范数的矩阵范数，矩阵 $\boldsymbol{B}\in\mathbb{R}^{n\times n}$ 满足

$\|B\|<1$,则 $I+B$ 和 $I-B$ 非奇异,且

$$\|(I\pm B)^{-1}\|\leqslant\frac{1}{1-\|B\|}.$$

4. 几种常见矩阵的性质

(1) **正交矩阵** $A\in\mathbb{R}^{n\times n}$,满足 $A^{\mathrm{T}}A=I$.

(2) **对称矩阵** $A\in\mathbb{R}^{n\times n}$,满足 $A^{\mathrm{T}}=A$. 对称矩阵的特征值均为实数,且有 n 个线性无关的特征向量. 它对应于不同特征值的特征向量必正交. 可以通过相似变换将对称矩阵化为对角矩阵. 若 A 对称,则 $\|A\|_2=\rho(A)$.

如果矩阵 A 对称,且

$$(Ax,x)>0,\quad\forall x\in\mathbb{R}^n,x\neq 0,$$

则称 A 是**对称正定矩阵**. 若 A 对称,则

$$\begin{aligned}A\text{ 正定}&\Leftrightarrow A\text{ 的特征值都大于零}\\&\Leftrightarrow A\text{ 的顺序主子式都大于零.}\end{aligned}$$

(3) **初等矩阵** $E(u,v;\sigma)=I-\sigma uv^{\mathrm{T}}$,其中 $u,v\in\mathbb{R}^n,\sigma\in\mathbb{R},\sigma\neq0$.

例如,**初等排列矩阵**

$$I_{ij}=I-(e_i-e_j)(e_i-e_j)^{\mathrm{T}}.$$

它是单位矩阵 I 对换第 i 行和第 j 行所得的矩阵. 满足 $I_{ij}^{-1}=I_{ij}$, $\det(I_{ij})=-1$(当 $i\neq j$ 时). $I_{ij}A$ 是将 A 的第 i 行和第 j 行对换所得的矩阵,AI_{ij} 是将 A 的第 i 列和第 j 列对换所得的矩阵.

若干个初等排列矩阵的乘积称**排列矩阵**.

又例如,**初等下三角矩阵**

$$L_j(l_j)=I+l_je_j^{\mathrm{T}}=\begin{bmatrix}1&&&&&\\&\ddots&&&&\\&&1&&&\\&&l_{j+1,j}&1&&\\&&\vdots&&\ddots&\\&&l_{nj}&&&1\end{bmatrix},$$

其中向量 $l_j=(0,\cdots,0,l_{j+1,j},\cdots,l_{nj})^{\mathrm{T}}$,满足

$$L_j(l_j)^{-1}=L_j(-l_j),\quad\det(L_j(l_j))=1.$$

一个单位下三角矩阵可以写成初等下三角矩阵的乘积,即

$$L=\begin{bmatrix}1&&&&\\l_{21}&1&&&\\l_{31}&l_{32}&1&&\\\vdots&\vdots&\vdots&\ddots&\\l_{n1}&l_{n2}&l_{n3}&\cdots&1\end{bmatrix}=L_1(l_1)L_2(l_2)\cdots L_{n-1}(l_{n-1}).$$

(4) **可约矩阵**. 设 $n\geqslant 2, \boldsymbol{A}\in\mathbb{R}^{n\times n}$,若存在排列矩阵 $\boldsymbol{P}\in\mathbb{R}^{n\times n}$,使

$$\boldsymbol{P}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{P}=\begin{bmatrix}\boldsymbol{A}_{11} & \boldsymbol{A}_{12}\\ \boldsymbol{0} & \boldsymbol{A}_{22}\end{bmatrix},$$

其中 $\boldsymbol{A}_{11}, \boldsymbol{A}_{22}$ 为方阵,则 $\boldsymbol{A}$ 称为**可约矩阵**,否则称 $\boldsymbol{A}$ 不可约.

(5) **对角占优矩阵**. 设 $\boldsymbol{A}=[a_{ij}]\in\mathbb{R}^{n\times n}$,若满足

$$|a_{ii}|>\sum_{\substack{j=1\\ j\neq i}}^{n}|a_{ij}|,\quad i=1,2,\cdots,n.$$

则称 $\boldsymbol{A}$ 为**严格对角占优矩阵**,若满足

$$|a_{ii}|\geqslant\sum_{\substack{j=1\\ j\neq i}}^{n}|a_{ij}|,\quad i=1,2,\cdots,n,$$

且其中至少有一个式子取严格不等号,则称 $\boldsymbol{A}$ 为**弱对角占优矩阵**.

定理 1.7　设 $\boldsymbol{A}\in\mathbb{R}^{n\times n}$,若 $\boldsymbol{A}$ 为严格对角占优矩阵,或 $\boldsymbol{A}$ 为不可约的弱对角占优矩阵,则 $\boldsymbol{A}$ 的对角元 $a_{ii}\neq 0(i=1,2,\cdots,n)$,且 $\boldsymbol{A}$ 非奇异.

1.2　典型例题分析

例 1.1　下列实数 x 的近似值为 x_A,它们各有多少位有效数字? 估计它们的相对误差界.

(1) $x=\pi, x_A=3.1$;　　(2) $x=\frac{1}{6}, x_A=0.167$;

(3) $x=\frac{\pi}{1000}, x_A=0.0031$;　　(4) $x=\frac{100}{6}, x_A=16.7$.

解　(1)根据有效数字的定义,x_A 有两位有效数字. 用定理 1.1,因 $a_1=3, n=2$,相对误差界可取为$\frac{1}{2\times 3}\times 10^{1-2}=\frac{1}{6}\times 10^{-1}$.

同理,(2)x_A 有三位有效数字,相对误差界可取为$\frac{1}{2}\times 10^{-2}$. (3)$x_A$ 有两位有效数字,相对误差界可取为$\frac{1}{6}\times 10^{-1}$. (4)$x_A$ 有三位有效数字,相对误差界可取为$\frac{1}{2}\times 10^{-2}$.

例 1.2　要使$\sqrt{13}$的近似值 x_A 的相对误差不超过 10^{-4},x_A 应取几位有效数字?

分析　相对误差和有效数字的关系,可以利用定理 1.1 来分析.

解　$\sqrt{13}=3.605\,551\cdots, a_1=3$. 根据定理 1.1,若 x_A 有 n 位有效数字,则相对误差界不超过$\frac{1}{2a_1}\times 10^{1-n}$. 令

$$\frac{1}{2a_1}\times 10^{1-n}\leqslant 10^{-4},$$

即 $10^{1-n}\leqslant 0.0006$,只要取 $n=5$,就有相对误差不超过 10^{-4}.这样得 $x_A=3.6056$.

例 1.3 下列公式要怎样变换才能使计算过程避免有效数字的损失?

(1) $I=\int_N^{N+1}\frac{\mathrm{d}x}{1+x^2}, N\gg 1$.

(2) $f(x)=\cos^2 x-\sin^2 x, x\approx\frac{\pi}{4}$.

分析 如果(1)直接用积分公式 $I=\arctan(N+1)-\arctan N$ 计算,N 大时将会出现相近数相减而损失有效数字位数的情况.类似情况也会在(2)直接计算时出现.

解 (1) 设 $\alpha=\arctan(N+1)$,$\beta=\arctan N$,则

$$\tan(\alpha-\beta)=\frac{\tan\alpha-\tan\beta}{1+\tan\alpha\tan\beta}=\frac{N+1-N}{1+N(N+1)}.$$

所以用 $I=\alpha-\beta=\arctan\frac{1}{1+N(N+1)}$来计算,将会避免误差危害.

(2) 用 $f(x)=\cos 2x=1-\frac{(2x)^2}{2!}+\frac{(2x)^4}{4!}-\cdots$计算.

举例来说,如果(1)中取 $N=100$,用计算器直接计算

$$\arctan 101-\arctan 100=1.560\,895\,66-1.560\,796\,66=9.900\times 10^{-5}.$$

而 $\arctan\frac{1}{1+100\times 101}=9.900\,009\,868\times 10^{-5}$,显然有更多的有效数字位数.

例 1.4 考虑二次方程

$$x^2+62.10x+1=0.$$

分析在四位十进制运算下怎样求其根 x_1 和 x_2 的近似值.

解 在 7 位有效数字的情形下有

$$x_1=-0.016\,107\,23,\quad x_2=-62.083\,90.$$

如果用初等代数的求根公式,并用四位十进制运算,得到近似值

$$x_{1A}=\frac{-62.10+\sqrt{(62.10)^2-4.000}}{2.000}=\frac{-62.10+62.06}{2.000}=-0.020\,00,$$

$$x_{2A}=\frac{-62.10-\sqrt{(62.10)^2-4.000}}{2.000}=\frac{-62.10-62.06}{2.000}=-62.10.$$

可见 x_{1A}误差很大.如果变换方法,同上计算 x_2 的近似值,而利用 $x_1x_2=1$ 来计算 x_1 的近似值,则有

$$\bar{x}_{2A}=-62.10,\quad \bar{x}_{1A}=\frac{1}{-62.10}=-0.016\,10.$$

得到 x_1 较好的近似值.

可以进一步思考一般的一元二次方程 $ax^2+bx+c=0(a\neq0)$ 如何数值求解.

例 1.5 如果要求计算系列的积分值 $I_n=\int_0^1 x^n e^{x-1}dx, n=0,1,\cdots,N$. 试讨论一种稳定的递推算法.

分析 由分部积分知 $I_n=1-nI_{n-1}$. 容易想到给出初值 $I_0=1-e^{-1}$ 的一个较准确的近似值 I_0^*,然后按递推公式 $I_n^*=1-nI_{n-1}^*$ 逐步计算 $I_1^*,I_2^*,\cdots,I_N^*$. 例如,按四位十进制计算,取 $I_0^*=1-0.3679=0.6321$,然后计算 $I_1^*=1-I_0^*=0.3679, I_2^*=1-2I_1^*=0.2642,\cdots$. 到 $I_8^*=-0.7280$,这是明显不对的结果(因为 $I_n>0$). 事实上,如果记

$$\varepsilon_n=I_n-I_n^*,$$

则由初值 I_0^* 的取法,有 $|\varepsilon_0|<0.5\times10^{-4}$. 但是

$$\varepsilon_1=I_1-I_1^*=1-I_0-(1-I_0^*)=-\varepsilon_0,$$
$$\varepsilon_2=I_2-I_2^*=1-2I_1-(1-2I_1^*)=-2\varepsilon_1=2\varepsilon_0.$$

递推得到

$$\varepsilon_n=I_n-I_n^*=-n\varepsilon_{n-1}=\cdots=(-1)^n n!\varepsilon_0.$$

$\{|\varepsilon_n|\}$ 是随 n 增大的数列,$|\varepsilon_8|=8!\ \varepsilon_0>4\times10^4\varepsilon_0$,这超过了 $|I_8|$ 的数值. 所以这样的递推方法是数值不稳定的,不能采用.

解 由以上分析知

$$I_{n-1}=(1-I_n)/n.$$

如果按 n 取 $N,N-1,\cdots$ 的次序来计算,若 I_n 近似值有误差 ε_n,则 I_{n-1} 近似值的误差 $|\varepsilon_{n-1}|=|\varepsilon_n|/n$,每次递推的误差都逐步减少,所以这是一种稳定的递推算法. 至于递推初值 I_N 的近似,可利用积分式

$$\frac{e^{-1}}{N+1}=\int_0^1 x^N e^{0-1}dx<I_N<\int_0^1 x^N e^{1-1}dx=\frac{1}{N+1},$$

取上式左右两端的平均值作为 I_N^*,即

$$I_N^*=\frac{1}{2}\left(\frac{e^{-1}}{N+1}+\frac{1}{N+1}\right),$$

然后递推计算 $I_{N-1}^*,I_{N-2}^*,\cdots,I_0^*$.

例 1.6 下面是两种利用 9 次 Taylor 多项式近似计算 e^{-5} 的方法,试分析哪种方法能提供较好的近似值.

(1) $e^{-5}\approx\sum_{i=0}^{9}(-1)^i\frac{5^i}{i!}$; (2) $e^{-5}\approx\left(\sum_{i=0}^{9}\frac{5^i}{i!}\right)^{-1}$.

分析 不妨记方法(1)和(2)中 e^{-5} 的近似值分别为

$$x_1=\sum_{i=0}^{9}(-1)^i\frac{5^i}{i!},\quad x_2=\left(\sum_{i=0}^{9}\frac{5^i}{i!}\right)^{-1}.$$

如果用计算器直接作数值计算,可得 $x_1\approx-1.8271, x_2\approx\frac{1}{143.69}=6.9595\times10^{-3}$(各有 5 位

有效数字)，且容易算得 $e^{-5}=6.7379\times10^{-3}$，可见 x_1 是很差的近似值，它是一个负数. 下面从截断误差来分析.

解 由 e^x 的 Taylor 展开式，有 $e^{-5}=x_1+R_1$，其中

$$R_1=\frac{5^{10}}{10!}-\frac{5^{11}}{11!}+\frac{5^{12}}{12!}-\cdots.$$

这是一个收敛的交错级数，其和的绝对值不超过首项. 为了估计得更精确，可以把上式开头三项作为首项，即 $|R_1|\leqslant\frac{5^{10}}{10!}-\frac{5^{11}}{11!}+\frac{5^{12}}{12!}\approx1.98$. 粗略地说，$x_1$ 近似 e^{-5} 的误差界取为 2，这是比较精确的估计. 可以看出截断误差比 e^{-5} 的绝对值还要大，方法(1)的近似值是不能用的.

考察方法(2)，由 Taylor 公式，$e^5=x_2^{-1}+\bar{R}_2$，其中

$$\begin{aligned}\bar{R}_2&=\frac{5^{10}}{10!}\left(1+\frac{5}{11}+\frac{5^2}{12\times11}+\frac{5^3}{13\times12\times11}+\cdots\right)\\&\leqslant\frac{5^{10}}{10!}\left[1+\frac{5}{11}+\left(\frac{5}{11}\right)^2+\left(\frac{5}{11}\right)^3+\cdots\right]\\&=\frac{5^{10}}{10!}\left(1-\frac{5}{11}\right)^{-1}\approx4.94.\end{aligned}$$

而 $e^{-5}=x_2+R_2$，误差 R_2 满足

$$R_2=e^{-5}-x_2=\frac{1}{x_2^{-1}+\bar{R}_2}-x_2=\frac{1}{x_2^{-1}(1+x_2\bar{R}_2)}-x_2.$$

因为 $|x_2\bar{R}_2|<1$，所以

$$\begin{aligned}R_2&=x_2(1-x_2\bar{R}_2+x_2^2\bar{R}_2^2-\cdots)-x_2\\&=-x_2^2\bar{R}_2+x_2^3\bar{R}_2^2-x_2^4\bar{R}_2^3+\cdots.\end{aligned}$$

这又是一个交错级数，其和的绝对值不超过首项的绝对值，即误差

$$|R_2|\leqslant|x_2^2\bar{R}_2|\leqslant2.4\times10^{-4}.$$

事实上，从开头的分析，实际的误差约为 2.2×10^{-4}，符合这个估计.

从本例可见，当 $x<0$ 时，直接用 e^x 在 $x=0$ 的 Taylor 多项式求 e^x 近似值是不适宜的.

例 1.7 数域 $\mathbb{P}$ 上的线性空间 V 的加法和数乘满足什么规则? V 中元素线性无关是什么意思? 回答上述问题并以 $\mathbb{R}^n$ 和 $C[a,b]$ 为例说明.

解 加法满足

(1) $u+v=v+u,\quad\forall u,v\in V$.

(2) $(u+v)+w=u+(v+w),\quad\forall u,v,w\in V$.

(3) 存在惟一的元素 $\theta\in V$(称为 V 中的零元素). 使得

$$u+\theta=u,\quad\forall u\in V.$$

(通常 θ 也记为 0，注意在不同情况下它与实数 0 的区别.)

(4) 对一切 $u\in V$，都在 V 中存在惟一的元素，记为 $-u$，满足