数值分析与算法

王泽文 邱淑芳 阮周生 编著

科 学 出 版 社

数值分析与算法

王泽文　邱淑芳　阮周生　编著

科学出版社
北京

内 容 简 介

本书共分12章，主要内容有：误差分析、非线性方程求根、线性方程组的直接解法与迭代解法、向量与矩阵范数、插值、最小二乘与函数的最佳逼近、数值积分与数值微分、常微分方程数值解法、矩阵特征值的计算方法、三角插值与快速Fourier变换、不适定问题与Tikhonov正则化方法等.

本书主要是为理工科研究生与本科生学习数值分析的理论与算法而写的，也是作者从事相关教学与科研的总结.

图书在版编目(CIP)数据

数值分析与算法/王泽文，邱淑芳，阮周生编著. —北京：科学出版社，2016.2

ISBN 978-7-03-047116-1

Ⅰ. ①数… Ⅱ. ①王… ②邱… ③阮… Ⅲ. ①数值计算 Ⅳ. ①O241

中国版本图书馆CIP数据核字(2016) 第012086号

责任编辑：胡海霞／责任校对：蒋 萍
责任印制：霍 兵／封面设计：陈 敬

科学出版社 出版
北京东黄城根北街16号
邮政编码：100717
http://www.sciencep.com

文林印务有限公司 印刷
科学出版社发行 各地新华书店经销
*
2016年2月第 一 版 开本：720×1000 1/16
2016年2月第一次印刷 印张：21 1/4
字数：428 000

定价：54.00元

(如有印装质量问题，我社负责调换)

前　言

科学计算, 是指应用计算机处理现代科学与工程研究领域中的数学计算问题, 包括构建数学模型和定量分析方法, 然后运用计算机分析并求解科学与工程问题. 其中, 数值分析与算法是科学计算中一个非常重要的内容, 主要研究计算机环境下求解科学与工程问题的数值算法与理论. 在许多科学与工程技术领域, 一个优秀的研究人员往往也是优秀的数值分析专家; 即使一个普通的研究人员或工程技术人员, 也必须掌握一些基本的数值算法, 并具备在计算机上实现算法的能力.

本书主要是为数学类本科生、非数学类理工科本科生与研究生学习数值分析的理论与算法而写的, 也是作者从事相关教学与科研工作的总结. 本书以算法为主线, 对基本理论和方法的论述尽可能做到精练简洁, 力求厘清算法构造的思维过程, 注重理论分析与计算机算法相结合, 积极融入探究性学习与教学的方法, 参考国内外文献并结合作者的教学科研, 较为系统地介绍数值分析的理论与算法. 在介绍误差分析的相关概念与基本方法之后, 本书首先安排的内容是非线性方程求根, 这是因为其内容及表述与之前学习的“高等数学”或“数学分析”课程最接近, 容易被初学者接受; 然后安排的是线性代数方程组数值解法方面的内容, 这部分内容则是“线性代数”或“高等代数”所学内容的延伸, 不易让初学者产生畏难情绪, 且其中一些数值解法是后续样条插值、最佳逼近等内容需要使用的; 接下来安排的内容是插值、最佳逼近、数值积分与数值微分、常微分方程数值解法等数值逼近方面的内容; 最后安排的内容是有一定难度的矩阵特征值的计算方法、三角插值与快速 Fourier 变换、不适定问题与 Tikhonov 正则化方法等内容. 其中, 本书最后一部分从矩阵的奇异值分解出发, 引入离散不适定问题的截断奇异值正则化方法, 然后结合正则化参数选取的模型函数方法介绍 Tikhonov 正则化方法, 最后给出三个典型的不适定问题及其数值解法.

本书由王泽文、邱淑芳、阮周生共同完成, 其中邱淑芳参与了第 2 章、第 3 章、第 12 章的部分撰写以及全书的校对工作, 阮周生参与了第 9 章的撰写和全书的校对工作, 全书其他章节由王泽文撰写. 全书由王泽文进行策划和统稿.

本书的写作得到了东华理工大学教务处领导、科技处领导、理学院领导的大力支持和帮助; 理学院刘唐伟、张文、王兵贤、孙海 (已调离)、胡彬、曾光、许小勇等教师及研究生在课程建设中付出了宝贵的时间和大量的精力, 并对本书提出了许多宝贵意见和建议, 在此一并致谢. 感谢科学出版社的编辑付出的辛勤劳动. 撰写过程中参考了众多国内外文献, 恕不一一致谢; 对于因疏漏而未能在参考文献中列出

的, 敬请谅解. 本书部分内容是王泽文访问美国 Purdue University 数学系时完成的, 为此感谢李培军教授及其领导的反问题研究团队的支持和帮助.

本书得到国家自然科学基金 (No.11161002)、江西省青年科学家培养计划 (20122BCB23024)、东华理工大学科研创新培育团队、东华理工大学重点教材等项目的资助.

限于水平和时间，书中疏漏之处在所难免，恳请广大读者批评指正.

作　者

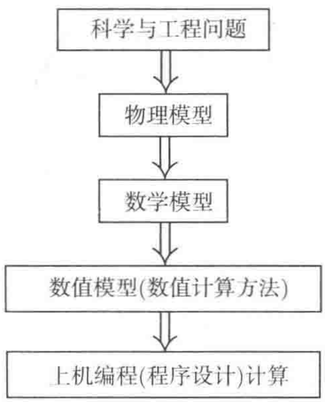
2015 年 10 月

目　录

第1章　绪　　论

1.1　数 值 分 析

数值分析主要研究数值求解各类数学问题的构造性方法. 科学与工程问题的解决大致可以分为四个步骤: ① 研究人员对具体问题建立起物理模型; ②将物理模型归结为数学模型; ③构造合适的数值计算方法, 将数学模型转化为数值模型; ④运用计算机程序设计知识和技能, 上机编程计算出结果. 在第二个步骤中, 或许能证明数学模型解的存在性和唯一性, 但能找出解的解析表达式的问题是微乎其微的. 因此, 上述第三个步骤在解决实际科学与工程问题中就显得尤为重要了. 数值分析是计算数学的重要组成部分, 计算数学是数学学科的一个主要分支, 它主要研究计算机环境下求解科学与工程问题的数值方法与理论及其软件实现.



随着科学与技术的发展, 特别是计算机技术的发展和一些边缘学科的兴起, 例如, 计算物理学、计算化学、计算生物学、计算社会学等, 在具体问题的解决中已经很难将上述三个环节给出明确的划分. 在许多科学与工程技术领域, 一个优秀的研究人员往往也是优秀的数值分析专家; 即使一个普通的研究人员或工程技术人员, 也必须掌握一些基本的数值分析方法和实现算法的计算机能力.

在一定程度上, 下述例子可以说明本课程与微积分、线性代数和常微分方程等课程的区别与联系.

众所周知, 由 Cramer 法则可以完全求出一个 n 阶的线性代数方程组的解, 但从实际计算上考虑, Cramer 法则确是毫无用处的. 读者不妨就 20 阶线性代数方程

组试试. 因此, Cramer 法则只有理论上的价值. 数值分析的一个重要任务是研究如何有效地求解线性代数方程组, 特别是大型的线性代数方程组.

由微积分中的零点定理可知, 定义在 $[a,b]$ 上的连续函数 $f(x)$, 若 $f(a)f(b)<0$, 且 $f(x)$ 是单调的, 则 (非线性) 方程 $f(x)=0$ 在 (a,b) 内存在唯一的根. 考虑如何数值地求解出这个根则是数值分析研究的内容. 再如, 对于 5 次和高于 5 次的多项式一般不存在求根的公式, 数值分析将提供多项式求根的数值计算方法.

在理论上, 对于常微分方程初值问题

$$\begin{cases} \dfrac{\mathrm{d}y}{\mathrm{d}x}=f(x,y), \quad a \leqslant x \leqslant b, \\ y(a)=y_0, \end{cases} \tag{1.1.1}$$

解的存在性和唯一性, 以及解的稳定性都有一些理论结果, 也针对右端项 $f(x,y)$ 的一些特殊类型给出了解析求解方法. 然而, 实际中提出的问题往往不属于这些类型, 即使 $f(x,y)$ 在形式上十分简单也无法得到解析解. 但是, 通过数值分析的方法, 就能得到式 (1.1.1) 的数值解, 即 $y(x)$ 在离散点上的近似值. 类似地, 还可以得到常微分方程边值问题、偏微分方程定解问题在离散点上的近似值.

数值分析是一门理论性和实践性都很强的课程, 既有数学类课程理论上的抽象性和严谨性, 又有实用性和实验性等技术特征. 对一些基本问题来说, 解的存在性和唯一性通常在纯数学范围内已得到保证或解决, 而数值分析的主要任务是提供获得解的数值方法. 但值得注意的是, 解的数值计算方法常常来源于解的存在性的构造性证明. 因此, 学习该课程需要微积分、常微分方程、线性代数和一门计算机语言 (例如, C 语言, Matlab 语言) 的知识和技能作为基础.

1.2 误 差

1.2.1 误差的概念

数值计算方法一般来说是近似的方法, 这就会产生误差. 误差就是精确值与近似值之间的偏差.

定义 1.1 设 x^* 是精确值 (或真值), x 是 x^* 的一个近似值, 称 $e=x^*-x$ 为近似值 x 的绝对误差或误差, 即

$$\textbf{绝对误差} = \text{精确值} - \text{近似值}.$$

实际上, 精确值 x^* 一般是未知的, 按定义计算不出绝对误差 e, 故通常只能给出绝对误差的绝对值 $|x^*-x|$ 的界. 例如, 当 $f(x)$ 充分光滑时, 有 Taylor(泰勒) 展开式

$$f(x)=f(0)+f'(0)x+\frac{f''(0)}{2!}x^2+\frac{f'''(0)}{3!}x^3+\cdots. \tag{1.2.1}$$

若取前三项作为 $f(x)$ 的近似, 记为

$$p(x) = f(0) + f'(0)x + \frac{f''(0)}{2!}x^2. \tag{1.2.2}$$

由于

$$f(x) = f(0) + f'(0)x + \frac{f''(0)}{2!}x^2 + \frac{f'''(\theta x)}{3!}x^3, \quad 0 < \theta < 1, \tag{1.2.3}$$

那么 $f(x)$ 与 $p(x)$ 之间的绝对误差为

$$e(x) = \frac{f'''(\theta x)}{3!}x^3. \tag{1.2.4}$$

理论上知道 θ 存在, 但实际上很难确定出 θ 的值, 也就是无法得到 $e(x)$ 的精确值. 但是, 若 $f(x)$ 满足

$$|f'''(t)| < M, \quad t\,在\,0\,和\,x\,之间,$$

则 $e(x)$ 的绝对误差有上界

$$|e(x)| \leqslant \frac{|x|^3}{3!}M, \quad x \neq 0.$$

定义 1.2 如果精确值 x^* 与近似值 x 的 (绝对) 误差的绝对值不超过某个正数 ε, 即

$$|e| = |x^* - x| < \varepsilon,$$

则称 ε 为 x 的绝对误差限或误差限.

例 1.1 若 x^* 经四舍五入得到 $x = 17.256$, 例如, $x^* = 17.25595, 17.25568, 17.25647, 17.25635$ 的四舍五入近似值都是 17.256. 它的误差限是

$$|e| = |x^* - x| < 10^{-4} \times 5 = \frac{1}{2} \times 10^{-3}.$$

若 x^* 经四舍五入得到 $x = 0.017256$, 则它的误差限是

$$|e| = |x^* - x| < 10^{-7} \times 5 = \frac{1}{2} \times 10^{-6}.$$

绝对误差具有量的概念, 但有时很难以其大小来衡量近似量的“好”与“坏”. 例如, 马拉松比赛全程为 42 公里 195 米 (1 公里 = 1 千米), 由于测量的误差, 实际路程仅为 42 公里 185 米, 绝对误差为 10 米, 按竞赛章程, 这样的误差是允许的. 但是, 若百米赛跑的实际路程为 99 米, 绝对误差仅为 1 米, 这却不能被竞赛规则所允许. 理由很简单, 相对 42 公里 195 米, 10 米的误差微不足道; 而相对 100 米, 1 米的误差却很大. 这就是相对误差的思想.

定义 1.3　设 x 是精确值 x^* 的一个近似值, 称绝对误差与精确值的比值

$$e_r = \frac{e}{x^*} = \frac{x^* - x}{x^*}$$

为近似值 x 的相对误差; 或者称绝对误差的绝对值与精确值的绝对值的比值

$$e_r = \frac{|e|}{|x^*|} = \frac{|x^* - x|}{|x^*|}$$

为近似值 x 的相对误差.

定义 1.4　如果存在正数 ε_r, 使得 $e_r = \dfrac{|e|}{|x^*|} < \varepsilon_r$, 则称 ε_r 为近似值 x 的相对误差限.

在实际计算中, 精确值 x^* 其实是不知道的, 此时往往用 x 代替 x^*, 即

$$e_r = \frac{e}{x} = \frac{x^* - x}{x};$$

而相对误差也被相对误差限代替, 即若知道 $|x^* - x| < M$, 则可用比值 $\dfrac{M}{x}$ 来确定相对误差的界限.

1.2.2　误差的来源

误差主要来源于以下三个方面.

(1) 原始误差.

原始误差包括模型误差与原始数据误差. 例如, 忽略某些次要因素, 气象模型可归结为一组偏微分方程, 这就产生了**模型误差**; 而测试到的即时气象数据因仪器精度等方面的原因, 存在一定的误差, 这就是**原始数据误差**.

(2) 截断误差.

例如, 用多项式 (1.2.2) 作为 $f(x)$ 近似公式, 它是由截断

$$T(x) = \frac{f'''(\theta x)}{3!} x^3$$

产生的, 则称 $T(x)$ 为**截断误差**, 也常称为**方法误差**.

(3) 舍入误差.

在数值计算中, 不管是人工还是计算机计算, 运算均是按有限位进行的. 例如, 当我们在采用 8 位字长的计算机上计算 2/3 时, 按照四舍五入的原则, 得到的结果就是 0.66666667, 它与真值 2/3 之间有误差, 这种误差是由于在运算中舍入产生的, 称为**舍入误差**.

1.2.3 误差的运算

本课程主要涉及截断误差和舍入误差, 而在这原始误差不是我们关心的问题. 下面考虑舍入误差的运算问题, 对截断误差的讨论将融合在后面介绍的各种数值方法中讨论.

设真值 x^* 与 y^* 的近似值分别为 x 与 y, 它们的误差分别为 e_1 与 e_2, 那么,

(1) $x^* \pm y^*$ 的近似 $x \pm y$ 的误差为

$$(x^* \pm y^*) - (x \pm y) = (x^* - x) \pm (y^* - y) = e_1 \pm e_2, \tag{1.2.5}$$

所以 $x \pm y$ 的误差限为 $|e_1| + |e_2|$, 而其相对误差为

$$\frac{|e_1 + e_2|}{|x^* \pm y^*|}. \tag{1.2.6}$$

若 $|x^* \pm y^*|$ 很小, 则其相对误差将会很大. 因此, 要尽量避免两个相近的数相减.

(2) $x^* \cdot y^*$ 的近似 $x \cdot y$ 的误差为

$$\begin{aligned} x^* \cdot y^* - x \cdot y &= (x + e_1) \cdot (y + e_2) - xy \\ &= xe_2 + ye_1 + e_1 e_2, \end{aligned} \tag{1.2.7}$$

通常将后面的高阶小量去掉, 即误差为 $xe_2 + ye_1$, 且误差限为

$$|x^* \cdot y^* - x \cdot y| \leqslant \max\{|x|, |y|\} \cdot (|e_1| + |e_2|). \tag{1.2.8}$$

(3) $\dfrac{x^*}{y^*}$ 的近似值 $\dfrac{x}{y}$ 的误差为

$$\frac{x^*}{y^*} - \frac{x}{y} = \frac{x^* y - y^* x}{yy^*} = \frac{(x + e_1)y - (y + e_2)x}{yy^*} = \frac{ye_1 - xe_2}{yy^*}, \tag{1.2.9}$$

式中分子 $ye_1 - xe_2$ 的界与 $x^* y^*$ 的近似 xy 的误差界相同. 如果 y^* 是个比较小的量, 那么误差 (1.2.9) 就可能很大. 因此, 应尽量避免小量作除数的除法运算.

1.2.4 有效数字

定义 1.5 当 x 的误差限为某一位的半个单位时, 则这一位到第一位非零位的位数称为 x 的有效位数; 若这一位到第一位非零位共有 n 位, 则称 x 有 n 位有效数字.

例如, 经四舍五入得到的 $x = 35.236, y = 0.0024563$ 均有 5 位有效数字, 而经四舍五入得到的 6.0 和 6.0000 则分别具有 2 位和 5 位有效数字.

例 1.2　求方程 $x^2 + 62.10x + 1.000 = 0$ 的根, 取四位有效数字进行运算.

解　由求根公式

$$x_{1,2} = \frac{-62.10 \pm \sqrt{62.10^2 - 4}}{2}$$

得

$$\begin{aligned} x_1 &= \frac{-62.10 - \sqrt{62.10^2 - 4}}{2} = \frac{-62.10 - \sqrt{3856 - 4}}{2} \\ &= \frac{-62.10 - 62.06}{2} = -62.10. \\ x_2 &= \frac{-62.10 + \sqrt{62.10^2 - 4}}{2} = \frac{-62.10 + \sqrt{3856 - 4}}{2} \\ &= \frac{-62.10 + 62.06}{2} = -0.0200. \end{aligned}$$

但是, 若采用

$$x_1 = \frac{-62.10 - \sqrt{62.10^2 - 4}}{2}, \quad x_2 = \frac{1}{x_1},$$

则得

$$x_1 = -62.10, \quad x_2 = -0.01610.$$

实际上, 真正的根为

$$x_1 = -62.08389\cdots, \quad x_2 = -0.016107\cdots.$$

探究　分析出现例 1.2 的计算结果的原因.

1.3　病态问题与数值稳定性

从例 1.2 可知, 舍入误差的大小决定于运算中所取位数的多少. 在计算机运算中, 它则决定于计算机的位长, 为了减少舍入误差对计算结果的影响, 有时可采用双倍位长运算来提高精度. 另一方面, 从例 1.2 又可得知, 解的误差大小还决定于计算方法 (算法). 这些是算法的数值稳定性问题, 以及在计算过程中如何避免舍入误差的扩大. 实际问题中, 一些问题的本身就对数据的输入或者说对数据的误差非常敏感, 即输入数据的微小变化将导致解的急剧变化, 这类问题常称为病态问题 (ill-posed problems), 否则称为良态问题 (well-posed). 反问题研究领域中的问题大都是病态问题, 例如, 数值微分问题、热传导方程的逆时问题等.

1.3.1　病态问题

定义 1.6　若一个问题当它的输入数据 (如参数, 初始值等) 发生微小变化时, 将引起其解的急剧变化, 则称该问题是病态的; 否则, 称为良态的.