The Numerical Analysis and Computational Method

数值分析与计算方法

主编 蒋 勇 李建良

科学出版社

数值分析与计算方法

主　编　蒋　勇　李建良

参　编　卢长娜　陈文兵　雷金贵

白　羽　陈允杰

科学出版社

北　京

内 容 简 介

《数值分析与计算方法》是为理工科大学的“数值分析”和“计算方法”课程而编写的教材. 本书突出“培养综合素质与能力”的核心理念, 注重理论与实际结合, 主要内容包括绪论、插值问题、曲线拟合与逼近、数值积分与数值微分、常微分方程数值解法、非线性方程求解、线性方程组的直接解法与迭代法、矩阵的特征值与特征向量计算等. 为帮助学生掌握内容, 每章附有适量习题; 为培养学生的综合素质、提高学生的实际技能, 本书专门安排了“上机实习课题”, 并配备了相应的计算实习题. 全书采用模块化结构, 章节相对独立, 脉络分明, 阐述严谨, 深入浅出, 便于教师根据学生的不同背景与教学计划灵活安排教学.

本书同时可作为理工科大学相关专业的研究生课程教材, 并可供从事科学与工程计算的科技工作者参考.

图书在版编目(CIP)数据

数值分析与计算方法/蒋勇, 李建良主编. —北京: 科学出版社, 2012

ISBN 978-7-03-032772-7

Ⅰ.①数… Ⅱ.①蒋… ②李… Ⅲ.①数值分析-高等学校-教材 ②数值计算-高等学校-教材 Ⅳ.①O241

中国版本图书馆 CIP 数据核字 (2011) 第 231664 号

责任编辑: 伍宏发 杨 锐 胡 凯 / 责任校对: 陈玉凤

责任印制: 赵 博 / 封面设计: 王 浩

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

源海印刷有限责任公司印刷

科学出版社发行 各地新华书店经销

*

2012 年 1 月第 一 版 开本: B5(720×1000)

2012 年 1 月第一次印刷 印张: 14

字数: 268 000

定价: 35.00 元

前　　言

“数值分析” 与 “计算方法” 是理工科大学各专业普遍开设的两门密切相关的核心课程. 其中, “数值分析” 作为核心的专业基础课, 主要开设对象为数学、统计、计算机科学等专业的学生; 而 “计算方法” 作为重要的专业基础课或选修课, 主要开设对象为其他理工科专业的学生. 从内容上看, 这是两门不同的课程, 各有不同的要求和侧重点. 但是, 它们又具有较大的交叉与重叠. 前者通常侧重数值方法的理论基础、算法设计与分析等, 后者通常侧重对算法的了解、实际应用等.

科学技术的不断发展与社会事业的不断进步, 对现代大学生与研究生的综合素质、创新与应用能力提出了新的要求: 一方面, 要求数理专业的学生具有更好的开拓性与实际应用能力; 另一方面, 对接受技术与应用能力训练为主的工科类学生以及非数理专业的其他理科学生, 也要求其在综合素质 (包括数理素质) 方面受到更好的训练.

作为 “十二五” 规划 “人才培养模式创新与课程体系改革”、“人才培养模式创新实验基地”、“学术与技能分类、专业培养分流 —— 人才培养的创新与实践”、“基于 ‘学术型、技能型、国际化’ 的多元化人才培养模式的研究与实践”、“数值分析精品课程建设” 等多项教学研究、教学改革与质量工程建设的成果之一, 我们在本书中尝试性地将 “数值分析” 与 “计算方法” 两门课程的内容进行了有机的结合与互补性的融合, 供这两门课程教学使用. 希望本书能够在教学中配合教师, 对学生综合素质的培养与实际技能的训练有所帮助. 应当说明, 鉴于不同学科与专业的实际情况与要求, 在教学过程中需要根据学生的背景因材施教, 对教学内容做适当的取舍与补充. 为了便于教学, 我们力求在保证系统性与层次性的基础上, 采用模块化结构, 章节相对独立; 在保持阐述严谨的基础上, 层次分明, 循序渐进, 希望有助于教师根据学生的不同背景与各专业不同的教学计划灵活安排教学. 为了帮助学生提高自学能力, 我们在编写过程中特别注意叙述的深入浅出, 希望有助于学生在教师指导下, 通过自学也能掌握相关课程的主要内容. 从这个意义上讲, 本书也特别适合于 “教师讲授与学生自学” 相结合的教学改革.

本书突出 “培养综合素质与能力” 的核心理念, 注重理论与实际结合, 主要内容包括绪论、插值问题、曲线拟合与逼近、数值积分与数值微分、常微分方程数值解法、非线性方程求解、线性方程组的直接解法与迭代法、矩阵的特征值与特征向量计算等. 为帮助学生掌握内容, 每章附有适量习题; 为培养学生的综合素质、提高学生的实际技能, 本书专门安排了 “上机实习课题”, 并配备了相应的计算实习题.

本书同时可作为理工科大学相关专业的研究生课程教材, 并可供从事科学与工程计算的科技工作者参考.

本书是在李建良教授、蒋勇教授等所编著的《计算机数值方法》(该书曾获南京理工大学优秀教材一等奖等) 的基础上, 根据上述原则由蒋勇、李建良、卢长娜、陈文兵、雷金贵、白羽、陈允杰等江苏省内外多所高校的教授、博士共同改编而成. 在改编过程中, 南京信息工程大学数学专业的部分研究生参与了部分编辑工作. 我们在此对所有参与工作的同志表示衷心的感谢. 此外, 特别感谢南京信息工程大学优秀教材出版基金对本书的立项与经济资助, 感谢计算数学、数值模拟与仿真等方面的众多专家对我们相关工作的支持与帮助!

欢迎各位读者提出宝贵意见, 以便我们在再版过程中改正.

南京信息工程大学数学与统计学院　蒋　勇
南京理工大学理学院　李建良

目　录

绪　论

计算机数值方法是一门研究数值计算的理论与方法的学问, 相应的计算实习则是将有关理论和方法在计算机上实现的系统训练.

计算机的工作依赖于以算法为核心的应用软件 (俗称程序), 而算法通常分为数值算法与非数值算法. 从计算机的出现至今, 数值算法一直受到科技与工程领域各个学科的共同关注, 人们习惯称其为数值计算方法或计算方法. 随着计算机的发展, 在内存、速度等方面已逐渐达到其物理极限. 因此, 研究、设计、应用计算方法已不仅仅要求提供可解决实际问题的算法, 而且必须考虑降低算法复杂性等问题. 例如, 如何提高算法的速度、减少占用的内存、降低算法的结构复杂性、增加算法的稳定性等, 都成为计算方法的研究内容.

虽然计算方法本身可以不依赖于计算机而作为科学与工程领域中的重要理论与方法, 但是将计算方法与计算机应用结合起来, 一方面可以使问题得到更好的解决, 另一方面可以大大提高计算机应用的水平. 因此, 进行适当的计算实习也自然成为本门课程的重要环节.

通过计算机数值方法的学习和训练, 可使学生在掌握计算方法的理论与方法的同时, 增强发现问题、分析问题、解决问题的能力, 从而达到提高综合素质的目的. 这对于 21 世纪的科技工作者显然尤为重要.

0.1　计算机数值方法的任务

为了说明计算机数值方法的任务及研究对象, 我们自然要看清它在解决实际问题中扮演的角色. 一般地, 用数学方法或计算机解决实际问题的过程可概括为如图 0-1 所示的模式.

由图 0-1 可见, 用计算机解决科学与工程领域计算问题的一般过程包括 (a) 建立数学模型、(b) 设计计算方法、(c) 设计程序、(d) 上机运算等, 以分别得到 (A) 数学模型、(B) 计算方法、(C) 程序、(D) 结果等为目标. 不难看出, 从 (A) 数学模型出发, 通过 "设计计算方法" 得到适当的 (B) 计算方法, 再经过 "设计程序" 得到可行的 (C) 程序, 这都是实际问题能否得到解决的关键过程, 也恰好就是本书所要处理的问题. 总之, 计算机数值方法的主要目的是: 根据数学模型设计出相应的计算方法, 并将计算机数值方法在计算机上实现. 除此之外, 计算方法中的有关理论与方法也在 "建立数学模型" 等环节中扮演着重要角色. 例如, 本书第 2 章中的曲线

拟合方法通常被作为建立数学模型的主要工具之一. 可见, 计算机数值方法在科学与工程设计中起着重要的作用.

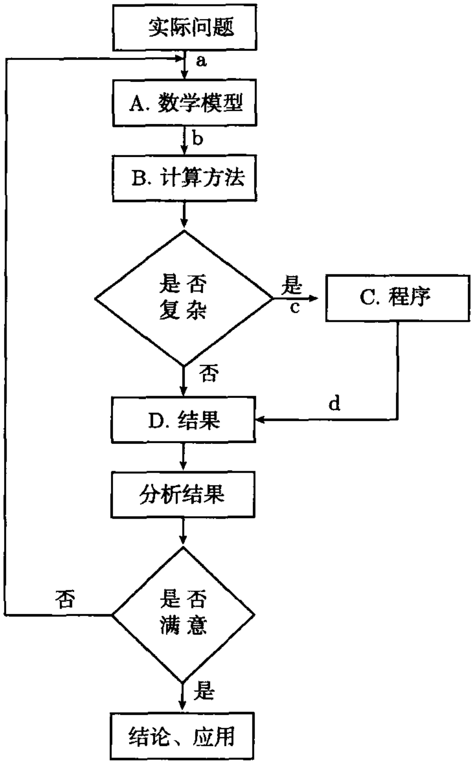

图 0-1　计算机解决实际问题的一般过程

那么, 计算机数值方法包括哪些内容呢? 可以说, 凡是有数学模型出现的地方都有一个设计计算方法的问题. 尽管实际问题复杂多变, 相应的数学模型五花八门, 但从数学理论上可对其作出比较系统的分类, 从而对应的计算方法问题可以大致概括为: 数值逼近、数值微分与数值积分、微分方程数值解、非线性方程数值解、线性方程组的数值解、特征值及特征向量计算等. 我们将通过对这些问题中比较常见的、简单的计算方法的学习, 掌握计算方法的基本理论、方法与技巧, 从而提高程序设计的能力, 为今后解决各种实际问题奠定良好的基础.

然而, 我们不能片面地将计算机数值方法理解为各种计算方法的简单罗列和堆积, 其实它本身就是一门内容丰富、研究方法深刻、有自身理论体系的学科. 它既有纯数学的高度抽象性与严密科学性, 又有应用的广泛性与实际实验的高度技术性, 是一门与计算机使用密切结合的、实用性很强的应用数学与计算机应用课程. 为了说明它与纯数学课程及计算机理论课程的不同, 我们来考虑线性方程组的数值解问

题. 在线性代数中只介绍解的存在唯一性及有关理论和精确解法, 用这些理论和方法还不能在计算机上求解上百个未知数的方程组, 甚至, 我们在第 6 章的引言中将看到, 即便是二三十个未知数的线性代数方程组, 都难以求解, 更不用说求解十几万个未知数的方程组了. 可见, 要求解这类问题, 还应根据方程的特点研究适合计算机使用的、满足精度要求的、节约计算时间的有效算法及其相关理论. 在实现这些算法时往往还要根据计算机容量、字长、速度等指标, 研究具体求解步骤的程序设计技巧. 另外, 对于被证明是行之有效的其他方法也应采用.

计算机数值方法的特点概括如下:

1) 面向计算机

要根据计算机的特点提供实际可行的有效算法, 即算法只能包括加、减、乘、除运算, 使计算机能直接处理.

2) 有可靠的理论分析

能任意逼近并达到精度要求, 对近似算法要保证收敛性和数值稳定性, 还要对误差进行分析, 这些都建立在相应的数学理论基础上.

3) 有好的计算复杂性

时间复杂性好是指节省时间, 空间复杂性好是指节省存储量, 这也是建立算法要研究的问题, 它们关系到算法能否在计算机上实现.

4) 有数值实验

任何一个算法除了从理论上要满足上述三点外, 还要通过数值实验证明是行之有效的.

根据计算机数值方法的这些特点, 学习时我们首先要注意掌握方法的基本原理和思想, 要注意方法处理的技巧及其与计算机的结合, 要重视误差分析、收敛性和稳定性等基本理论; 其次, 要通过例子, 学习使用各种计算方法来解决实际计算问题; 最后, 为了掌握本课程的内容, 还应做一定数量的理论分析与计算练习. 由于本课程内容包括了微积分、代数、常微分方程的数值方法, 读者必须掌握这几门课程的基本内容后才能学习本课程.

下面我们举两个例子, 对上述内容加以说明.

0.2 关于方程求根及其二分法

我们考虑如下问题:

问题 A. 求 $f(x) = x^4 - 2x^2 - 4x + 19\,590\,426$ 的极值.

问题 B. 求 $y = f(x) = (x^3 - x - 1)^{-1}$ 与 $y = 0, x = 0, x = 1.3$ 所包围的面积.

问题 C. 求方程 $y''' - y' - y = 0$ 的通解.

由高等数学的知识不难发现它们都归结为方程 $x^3 - x - 1 = 0$ 的求解问题. 对大多数人来说, 找出其解并不是一件容易的事. 而这只是实际问题中出现的一个非常简单的代数方程的求解问题, 我们面对的其他实际问题也不仅仅是代数方程的求解问题. 即使是代数方程的求解问题, 当方程的次数较高时人们也难以用公式、分析等方法直接求解. 事实上, 19 世纪挪威数学家阿贝尔 (Abel) 已经证明: 对次数高于 4 次的代数方程, 不可能由方程的系数直接表示出解的形式. 这就迫使人们去寻求解决各种实际问题的切实可行的数值计算方法. 这里, 我们不妨来看一下代数方程的比较经典的求解方法 —— 二分法.

设函数 $f(x)$ 在闭区间 $[a,b]$ 上连续, 且 $f(a)f(b) < 0$, 根据连续函数的性质, $f(x)$ 在开区间 (a,b) 内一定有实的零点, 即方程 $f(x) = 0$ 在开区间 (a,b) 内一定有实根. 我们这里假定它在开区间 (a,b) 内有唯一的单实根 x^*, 此即连续函数介值定理. 我们记为**定理 0.1**.

定理 0.1 设实函数 $f(x)$ 在闭区间 $[a,b]$ 上连续, 且 $f(a)f(b) < 0$, 则方程 $f(x) = 0$ 在开区间 (a,b) 内至少有 1 个实根.

考察有根区间 (a,b), 取中点 $x_0 = (a+b)/2$ 将它分为两半, 然后进行根的搜索, 即检查 $f(x_0)$[当 $f(x_0) = 0$ 时, 则 $x^* = x_0$] 与 $f(a)$ 是否同号. 如果确系同号, 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 x^* 必在 x_0 左侧, 这时令 $a_1 = a, b_1 = x_0$ (图 0-2). 不管出现哪一种情形, 新的有根区间 $[a_1, b_1]$ 的长度仅为 $[a,b]$ 的一半.

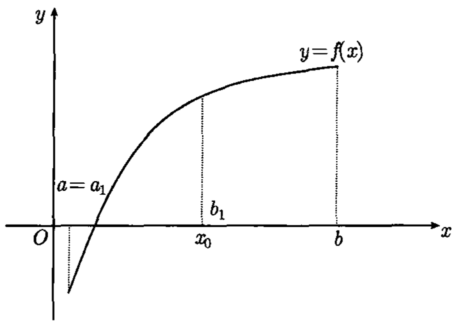

图 0-2 二分法的几何意义

对压缩后的有根区间 $[a_1, b_1]$ 又实施同样的方法, 即取中点 $x_1 = (a_1 + b_1)/2$ 将区间 $[a_1, b_1]$ 再分为两半, 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.

如此反复二分下去, 即可得出一系列有根区间:

$$[a,b] \supset [a_1, b_1] \supset [a_2, b_2] \supset \cdots \supset [a_k, b_k] \supset \cdots,$$

其中, 每个区间都是前一个区间的一半, 因此二分 k 次后的有根区间 $[a_k,b_k]$ 的长度为

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

可见, 如果二分过程无穷地进行下去, 这些有根区间最终必收缩于一点 x^*, 该点显然就是所求的根.

设第 k 次二分后, 取有根区间的中点

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

作为根的近似值, 则在二分过程中可以获得一个近似根的序列 $x_0, x_1, x_2, \cdots$, 该序列以根 x^* 为极限.

不过在实际计算时, 我们一般不可能完成这种无穷过程, 其实也没有这种必要, 因为数值分析的结果允许带有一定的误差, 由于

$$|x^* - x_k| \leqslant \frac{1}{2}(b_k - a_k) = \frac{1}{2^{k+1}}(b-a),$$

只要二分足够多次 (即 k 充分大), 便有

$$|x^* - x_k| < \varepsilon,$$

其中, ε 为预定精度.

上述求解方法称为**二分法**. 它是计算机中一种常用的算法. 我们给出其算法框图, 如图 0-3. 其中, a, b 表示有根区间的左、右端点; x 表示近似根.

各框的具体含义如下:

[框 1]　从所给区间 (a,b) 着手二分.

[框 2]　取有根区间 (a,b) 的中点 x 作为近似根.

[框 3]　确定二分后新的有根区间 (a,b).

[框 4]　检查近似根 x 是否满足精度要求.

例 0.1　用二分法求方程 $x^3 - x - 1 = 0$ 在区间 $(1, 1.5)$ 内的一个实根, 要求误差不超过 0.005.

解: ① 因为 $f(1) = -1, f(1.5) = 0.875$, 所以

$$f(1)f(1.5) < 0.$$

另外, 显然 $f(x)$ 在 $[1, 1.5]$ 连续, 故由**定理 0.1**, $f(x)$ 在区间 (1,1.5) 内有根.

② 预估所要二分的次数. 按误差估计式

$$|x^* - x_k| \leqslant \frac{1}{2^{k+1}}(b-a),$$

只需二分 6 次, 便能达到所要求的精度.

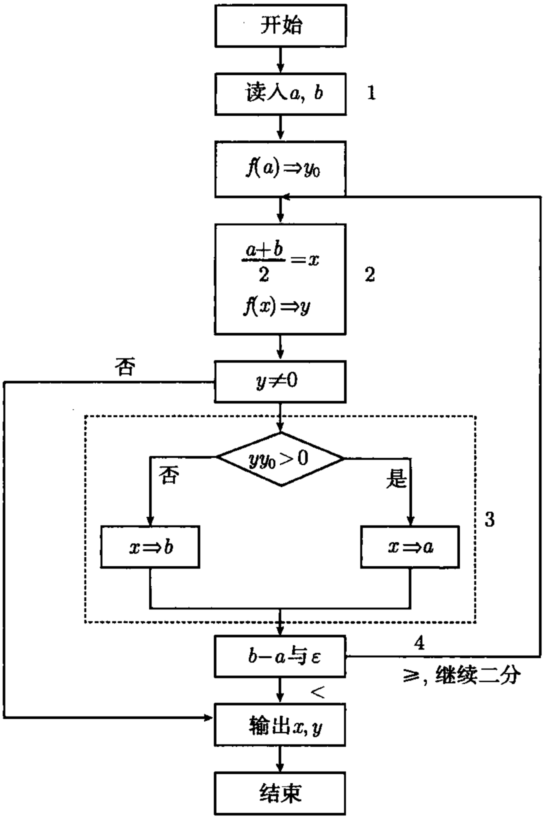

图 0-3　二分法的算法框图

③ 二分法的计算结果如表 0-1 所示.

表 0-1　二分法 (例 0.1) 的计算结果

k	a_k	b_k	x_{k+1}	$f(x_k)$ 的符号
0	1.0000	1.5000	1.2500	−
1	1.2500		1.3750	+
2		1.3750	1.3125	−
3	1.3125		1.3438	+
4		1.3438	1.3281	+
5		1.3281	1.3203	−
6	1.3203		1.3242	−

0.3　误差分析的重要性

通常人们习惯地认为, 只要从理论上给出了计算步骤, 即可大胆地计算下去, 而

不考虑计算过程中舍入误差的影响. 也有人在将复杂问题化成简单问题时误认为截去比如方程的无穷小量不会对方程的解产生很大的误差. 诸如此类, 在实际问题的求解上, 往往会产生面目全非的结果, 甚至在计算机上无法运行下去.

这里我们仅给出一个例子来说明问题的严重性, 从而引起我们在设计计算方法时对相应问题的重视.

例 0.2　建立计算 $I_n=\int_0^1 \frac{x^n}{x+25}\mathrm{d}x$ 的递推公式并计算 I_1,I_2,I_3,I_4.

解: ① 显然

$$I_n+25I_{n-1}=\int_0^1 \frac{x^n+25x^{n-1}}{x+25}\mathrm{d}x=\int_0^1 x^{n-1}\mathrm{d}x=\frac{1}{n}, \tag{0.1}$$

故有递推公式

$$I_n=\frac{1}{n}-25I_{n-1}.$$

② 保留 4 位有效数字, 计算如下:

$$\begin{aligned}
I_0&=\int_0^1 \frac{1}{x+25}\mathrm{d}x=\ln(x+25)\big|_0^1=0.039\,22,\\
I_1&=1-25I_0=1-25\times 0.039\,22=0.019\,50,\\
I_2&=\frac{1}{2}-25I_1=\frac{1}{2}-25\times 0.019\,50=0.012\,50,\\
I_3&=\frac{1}{3}-25I_2=\frac{1}{3}-25\times 0.012\,50=0.020\,83,\\
I_4&=\frac{1}{4}-25I_3=\frac{1}{4}-25\times 0.020\,83=-0.270\,75.
\end{aligned}$$

③ 分析计算结果如下.

令 $f_n(x)=\frac{x^n}{x+25}$, 由于 $f_n(x)$ 在区间 $(0,1]$ 上都是大于零的, 所以, 对任意的整数 n, 都有 $I_n=\int_0^1 f_n(x)\mathrm{d}x>0$, 这与 ② 中 $I_4=-0.270\,75<0$ 矛盾. 由此可见, 上述方法并不能保证结果的正确性, 为什么呢? 这是因为在计算 I_0 时有误差 (只保留了小数点后 5 位).

准确值为

$$I_1=1-25I_0,$$

近似值为

$$\bar{I}_1=1-25\bar{I}_0,$$

所以

$$\left|I_1-\bar{I}_1\right|=25\left|\bar{I}_0-I_0\right|.$$