普通高等教育"十一五"规划教材

大学数学教学丛书

计算方法简明教程

主 编 王新民 董小刚
主 审 冯果忱

科学出版社
www.sciencep.com

普通高等教育“十一五”规划教材

大学数学教学丛书

计算方法简明教程

主编　　王新民　董小刚

主审　　冯果忱

科学出版社

北京

内 容 简 介

本书着重介绍了能够在计算机上得以实现的一些数值解法. 主要包括一元与二元函数代数插值, 样条函数插值; 正交多项式及其应用, 函数的最佳一致逼近与最佳平方逼近; 数值积分及应用; 线性代数方程组的直接解法与迭代解法; 非线性方程和方程组的迭代方法; 矩阵特征值与特征向量的计算; 常微分方程初值问题的数值解法; 偏微分方程初、边值问题的有限差分法和有限元法. 并且针对各种算法讨论了误差估计以及方法的收敛性和稳定性等问题.

本书内容丰富, 取材精练; 阐述严谨, 脉络分明; 推导翔实, 重点突出. 具有广泛的应用性和极强的可读性. 本书可作为非数学专业研究生和高年级本科生的教材使用, 也可供从事数值计算的科技工作者参考.

图书在版编目(CIP)数据

计算方法简明教程/王新民, 董小刚主编. —北京: 科学出版社, 2010
普通高等教育"十一五"规划教材·大学数学教学丛书

ISBN 978-7-03-028729-8

Ⅰ. ①计… Ⅱ. ①王… ②董… Ⅲ. ①计算方法-高等学校-教材 Ⅳ. O241
中国版本图书馆 CIP 数据核字 (2010) 第 162212 号

责任编辑: 张中兴 于俊杰 / 责任校对: 陈玉凤
责任印制: 张克忠 / 封面设计: 耕者设计工作室

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
双 青 印 刷 厂 印刷
科学出版社发行 各地新华书店经销
*
2010年9月第 一 版 开本: B5(720 × 1000)
2010年9月第一次印刷 印张: 21 1/4
印数: 1—3 000 字数: 420 000

定价: 38.00 元

(如有印装质量问题, 我社负责调换)

《大学数学教学丛书》编委会

(以姓氏笔画为序)

丛　书　序

本丛书是为普通高等院校本科学生所编写的数学系列教材，是由长春地区五所普通高校具有丰富教学和科研经验的教师联合编写的，是集体智慧的结晶. 本丛书从酝酿到出版经历了近十年的时间，几经修改终于成稿. 在教材内容的编排上，我们一方面借鉴了国内一些品牌教材的先进模式，另一方面结合新形势下的新要求，并根据五所普通高校本科学生的特点，先后编写了逾百万字的教材与讲义，在多年使用过程中不断提炼修订，逐步趋于完善. 应该说，本套教材凝聚着五所高校几代数学教师的心血和汗水，希望能培养出更多的创新性人才.

本套教材包括《微积分(经管类)》、《概率论与数理统计》(两本)、《线性代数》、《计算方法简明教程》、《数学建模》、《复变函数与积分变换》. 编者在取材上着眼于本科生未来的发展和当今世界科学技术的发展，充分反映国内外教学前沿信息和最新学术动态，本着“夯实基础、适当延伸，注重应用、强化实践”的原则，大胆摆脱了普通高等院校教材编写的传统套路，使这套教材具有很强的实用性、一定的可读性、较高的艺术性和丰富的实践性；同时还保持了数学知识的系统性、严密性、连贯性等特点，内容翔实，清晰易读，便于教学与自学. 另外，本套教材充分考虑了有志报考研究生同学的需要，每一本教材都配备了丰富的、梯度配置的例题与习题，紧扣学生学习和报考研究生复习的需要，既具有明显的启发性，又具有典型的应用意义，可供普通高校理工科各专业使用.

本套教材从选题、大纲、组织编写到编辑出版，自始至终得到了科学出版社数理出版分社的支持，同时也得到了长春工业大学、吉林建筑工程学院、长春大学、长春工程技术师范学院、长春建筑学院教务处及数学系各位领导的支持和帮助，在此，我们一并表示衷心的感谢.

编　者

2010 年 3 月

前　言

计算方法又称数值分析, 是借助计算机进行科学研究和工程设计的一门交叉性学科, 属于计算数学的范畴. 它伴随着计算机的发展和普及而日益活跃在自然科学、军事科学、社会科学以及其他科学部门. 当把一个实际问题转化为数学模型或数值问题时, 计算方法就有了广阔的应用空间. 它在科学技术蓬勃发展的过程中起着不可忽视的作用. 例如, 样条插值方法在航空、造船等工程设计的许多领域都被认为是一种有效的数学工具; 求极值的共轭梯度法在建立经济发展的最优计划模型中起着重要作用; 微分方程数值解法在预测地下的矿藏储量等问题中发挥着巨大作用; 地震预报、天气预报以及地下水、地表水水质预测等问题往往也离不开有限差分法、有限元法等数值技术. 由此可见, 以数值计算为主的各种算法与技术已成为科学研究与工程设计的一个重要手段.

在工程技术等领域中, 常常要把一个实际问题归结为一个数学模型 (如微分方程定解问题), 而由于实际问题的复杂性, 常常得不到模型的准确解, 只能将它离散化后通过解一个大型线性 (或非线性) 代数方程组求其近似解, 这个过程没有计算机是不可想象的. 所以本书提供了能在计算机上方便实现的算法. 对实际问题来说, 要对数学模型提供一种算法并不是微积分和线性代数就能解决的, 远的不说, 就说对一个函数的性态作研究, 如果函数表达式很复杂, 我们就无从计算函数值, 更不能对这个函数作积分运算, 也就不能对它有任何认识, 因此这又赋予了本书一个任务, 就是讨论函数的逼近以及积分的数值计算问题. 本书还介绍了如何在计算机上计算矩阵的特征值和特征向量, 如何求解微分方程数值解等内容. 所有这些理论和方法都是解决工程问题时必不可少的工具.

本书是作者在多年的教学与科研工作的基础上完成的. 在编写过程中, 充分利用了王新民在长春地质学院和吉林大学工作时所出版的计算数学方面的相关教材, 并且本着与时俱进的精神, 精心选择材料, 尽可能征求任课教师的意见, 力求完善. 然而由于水平限制, 一定有疏漏不妥之处, 欢迎来自各方面的意见和建议.

本书由王新民、董小刚主编, 冯果忱先生担任了本书主审. 参加本书编写工作的还有王朝勇、高海龙和康素梅几位老师.

在完成本书的过程中, 科学出版社的同志们给予了很多帮助, 在此表示衷心的感谢. 还要感谢尹慧、陶占盛等同学, 他们为本书的出版做了许多有益的工作.

编　者

2010 年 7 月

目 录

绪　论

0.1　数值计算方法的研究对象

数值计算方法 (或称数值分析) 属于计算数学的范畴, 它伴随着电子计算机科学与技术的进步及推广应用而成长壮大. 近几十年来随着科学技术的发展与计算机计算能力的飞速发展, 数值计算方法的应用已经深入到各个学科领域, 很多复杂的和大规模的计算问题都可以在计算机上予以实现, 计算机高效的科学计算能力是其他计算工具无法比拟的. 可以说, 没有计算机上的这些数值方法的神通广大, 就没有今天如此之恢弘的航天事业以及如此之雄伟壮观的小浪底水利枢纽工程. 现在, 科学与工程中的数值计算已经成为自然科学、社会科学、军事科学以及工程技术科学发展的一种重要手段, 成为与实践和理论并列的一个不可缺少的环节.

可以认为, 数值计算方法实际上是一门数学与计算机科学的交叉学科, 它既有数学的抽象性与严密性, 又有计算机科学的实践性与技术性. 也就是说, 数值计算方法以纯数学为基础, 但不只研究数学本身的理论, 而着重研究解决问题的具体方法及产生的效果. 一种优秀的数值计算方法应该具有较快的计算速度和占有较小的存储量, 同时还应具有收敛性、稳定性以及较高的计算精度. 虽然有些方法在理论上还不够完善与严密, 但如果通过对比分析、实际计算和实践检验等手段被证明是行之有效的, 也常常被采用. 由此可见, 数值计算方法又是一门更注重实用性、可靠性、准确性和应用性的数学学科.

本书主要讨论在工程技术等领域中常用的数值方法. 这些方法是在计算机技术的基础上发展起来的, 因为在许多工程问题中, 我们常常要把实际问题归结为数学模型, 而由于问题的复杂性, 常常得不到模型的准确解, 只能将它离散化后求其数值解, 这个过程不使用计算机进行运算是不可想象的. 所以本门课程研究的内容是利用计算机求解数学问题近似解的方法, 主要包括函数的插值与逼近、积分的近似计算、线性和非线性代数方程与方程组的数值解法、矩阵特征值与特征向量问题的计算、常微分方程与偏微分方程初边值问题的数值解法等. 本书所介绍的数值计算方法的基本思想和基本技术有着广泛的实际应用, 它的诞生曾一度使科学发展产生了巨大飞跃, 从宏观天体运动学到微观分子细胞学说, 从工程系统到非工程系统, 都与这些数值算法密不可分. 它使各科学领域从定性分析阶段走向定量分析阶段, 从粗糙走向精密.

0.2 数值计算方法的研究思路

数学中常用的思想方法很多, 诸如公理化方法、模型法、关系映射反演法、构造法、类比法、归纳法、统计法、对偶法、逐次逼近法、反证法等. 在数值计算方法的研究方面, 最基本的且应用普遍的思想方法可概括为以下几种.

0.2.1 逐次逼近法

逐次逼近法又称迭代法, 在数值分析中应用得十分广泛, 这种方法是指按同一公式重复计算的一个数值过程.

例 0.2.1 *求三次方程 $x^3-3x+1=0$ 的根.*

解 将方程改写为

$$x=\frac{1}{3-x^2}, \tag{0.2.1}$$

取 $x_0=0$ 作为方程的根, 代入式 (0.2.1) 的右端并求出 x 的一个值记为 x_1, 即有

$$x_1=\frac{1}{3-0^2}=0.333\cdots,$$

将 $x_1=0.333\cdots$ 代入式 (0.2.1) 的右端, 得

$$x_2=\frac{1}{3-(1/3)^2}=\frac{9}{26}=0.3461\cdots,$$

将 $x_2=0.3461\cdots$ 代入式 (0.2.1) 的右端, 得

$$x_3=\frac{1}{3-(9/26)^2}=\frac{676}{1947}=0.3472\cdots,$$

等等, 如此继续下去, 得数列

$$x_0,x_1,x_2,\cdots,x_n,\cdots.$$

这个数列 $\{x_n\}$ 收敛于方程 (0.2.1) 的解 $x^*=0.3473\cdots$. □

对于给定的线性方程组

$$\boldsymbol{x}=\boldsymbol{B}\boldsymbol{x}+\boldsymbol{f}, \tag{0.2.2}$$

任取 $\boldsymbol{x}^{(0)}\in\mathbf{R}^n$, 代入式 (0.2.2) 的右端, 将其计算结果记为 $\boldsymbol{x}^{(1)}$, 再以 $\boldsymbol{x}^{(1)}$ 代入式 (0.2.2) 的右端, 算得的结果记为 $\boldsymbol{x}^{(2)}$, 如此进行下去, 便得到迭代格式

$$\boldsymbol{x}^{(k+1)}=\boldsymbol{B}x^{(k)}+\boldsymbol{f},\quad k=0,1,\cdots, \tag{0.2.3}$$

当 $||\boldsymbol{B}|| < 1$ 时, 对任意右端向量 $\boldsymbol{f}$ 和初始向量 $\boldsymbol{x}^{(0)}$, 迭代格式 (0.2.3) 收敛于方程组 (0.2.2) 的解 $\boldsymbol{x}^*$.

从以上的分析可以看出, 为了解决一个数学问题, 可以从一个与该问题的实质内容有着本质联系的较大范围入手, 再逐步缩小范围, 逐步逼近, 以致最后达到问题所要求的解. 这种对问题的处理方法就是逐次逼近法. 在数值计算中, 当所遇到的函数不便于计算或处理, 如有时函数关系没有明显的解析表达式, 需要根据实验观测或其他方法来确定与自变量的某些值相应的函数值; 有时虽然有表达式, 但使用起来很不方便, 而从实际需要出发, 又允许计算结果有一定的误差, 在这种情况下, 采用逐次逼近法往往可以较方便地解决问题.

0.2.2 以直代曲

"以直代曲" 的思想开拓了数学的 "无穷无尽的道路". 高等数学的诞生和发展就得力于直与曲的转化. 实际上, 直与曲相互转化的方法是微积分学乃至全部高等数学中的重要的、必不可少的一种方法.

例如, 由 Taylor 公式有

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

当 $x - x_k$ 很小时, 就有近似式

$$f(x) \approx f(x_k) + f'(x_k)(x - x_k).$$

这表明在 x_k 附近 $f(x)$ 可以用线性函数近似地代替. 于是对非线性方程 $f(x) = 0$ 的求解问题就可近似地转化为对线性方程

$$f(x_k) + f'(x_k)(x - x_k) = 0 \tag{0.2.4}$$

的求解. 如果将线性方程 (0.2.4) 的解记为 x_{k+1}, 则结合逐次逼近法就可由迭代格式

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

产生一个序列 $\{x_k\}$, 在一定条件下, 序列 $\{x_k\}$ 收敛于非线性方程 $f(x) = 0$ 的解.

又如, 对于给定的定积分 $\int_a^b f(x)\mathrm{d}x$, 如果用直线代替曲线 $y = f(x)$, 则在误差允许的范围内, 可用梯形面积近似, 即

$$\int_a^b f(x)\mathrm{d}x \approx \frac{b-a}{2}\left[f(a) + f(b)\right].$$

如果将以直代曲的思想加以推广, 则在上述问题中可选用一个简单函数 $\varphi(x)$(如代数多项式) 逼近函数 $f(x)$, 这就产生了函数逼近方法. 在此基础上, 对积分的计算、方程的求根、微分方程的求解等问题都可以给出各种不同的数值算法.

0.2.3 类比法

类比法是数学问题的又一个重要研究思路, 它是提出新问题和作出新发现的一个重要源泉. 一般地, 把需要求解的问题与以前已经解决的问题进行比较, 看看条件和目标有哪些相同、相似、接近的地方, 通过联想, 使解决老问题的方法重现, 在解决老问题方法的启发下, 打开新问题的研究思路, 这就是类比.

例如, 已知当 $\boldsymbol{A}$ 为对称正定矩阵时, 求线性代数方程组 $\boldsymbol{Ax}=\boldsymbol{b}$ 的解与求二次函数

$$J(\boldsymbol{x})=\frac{1}{2}(\boldsymbol{Ax},\boldsymbol{x})-(\boldsymbol{b},\boldsymbol{x}),\quad \boldsymbol{x}\in\mathbf{R}^n$$

的极小值问题等价, 由此可以猜想, 当微分算子 L 对称正定时, 对微分方程 $Lu=f$ 的求解可否转化为求泛函数

$$J(u)=\frac{1}{2}(Lu,u)-(f,u),\quad u\in H^1$$

的极值问题. 实践证明, 这一猜想是正确的. 正是这一猜想的准确性, 才有了变分原理, 进而产生了有限元方法.

对于解一般的问题, 如果一时难以下手, 则可先解决它的特殊情况, 然后通过类比把解决特殊情况的思路、方法或者结果推广到一般问题上, 从而使待求的问题得到解决. 例如, 给定 $y=f(x)$ 在定义区间上的两个点 $(x_0,y_0),(x_1,y_1)$, 可唯一确定一条直线

$$P_1(x)=\sum_{j=0}^{1}l_j(x)y_j=\sum_{j=0}^{1}\left(\prod_{\substack{i=0\\ i\neq j}}^{1}\frac{x-x_i}{x_j-x_i}\right)y_j,$$

其中, $l_0(x)$ 和 $l_1(x)$ 是两个线性无关的一次式, 且具有如下性质:

$$l_j(x_i)=\delta_{ij}=\begin{cases}1, i=j,\\ 0, i\neq j,\end{cases}\quad i,j=0,1.$$

若给定 $y=f(x)$ 在定义区间上的三个点 $(x_0,y_0),(x_1,y_1),(x_2,y_2)$, 则可唯一确定一条抛物线

$$P_2(x)=\sum_{j=0}^{2}l_j(x)y_j=\sum_{j=0}^{2}\left(\prod_{\substack{i=0\\ i\neq j}}^{2}\frac{x-x_i}{x_j-x_i}\right)y_j,$$

其中, $l_j(x)(j=0,1,2)$ 是三个线性无关的二次式, 也具有如下性质:

$$l_j(x_i)=\delta_{ij}=\begin{cases}1, i=j,\\ 0, i\neq j,\end{cases}\quad i,j=0,1,2.$$

根据上述特殊情况, 可以猜测并进一步得到证明, 如果给定 $y=f(x)$ 在定义区间上的 $n+1$ 个点 $(x_0,y_0),(x_1,y_1),\cdots,(x_n,y_n)$, 则可唯一确定一个 n 次式

$$P_n(x)=\sum_{j=0}^{n} y_j l_j(x)=\sum_{j=0}^{n}\left(\prod_{\substack{i=0\\ i\neq j}}^{n}\frac{x-x_i}{x_j-x_i}\right)y_j,$$

其中, $l_j(x)$ 是 n 次式, 它满足条件

$$l_j\left(x_i\right)=\delta_{ij}=\begin{cases}1, i=j,\\ 0, i\neq j,\end{cases}\qquad i,j=0,1,\cdots,n.$$

这个问题就是 n 次代数插值问题.

综合以上可以看出, 解决一个数学问题常常是联合几种思路和方法. 这里还需强调的是, 在解决一个具体问题时, 我们不要只追求或满足于问题的答案, 而要搞清楚解题的原理和过程, 要把所应用到的知识点加以总结、归纳, 充分注意推理的严密性、计算的准确性、条件的可靠性以及解法的简洁性. 要在不断的学习中熟悉知识、积累经验; 在应用中悟出道理、拓宽思路, 只有这样, 才能更好地提高解决问题的能力.

0.3 数值计算中的误差分析

一般来讲, 在对实际问题进行求解时, 所经历的每一个过程都存在有误差. 因此, 误差分析是数值计算研究领域中的核心内容之一.

0.3.1 误差的来源与分类

用数值方法进行科学计算时, 往往得到的是实际问题的近似解, 我们把近似解与准确解之差称为误差. 产生误差的原因很多, 但通常来源于以下四个方面.

1. 模型误差

我们在解决一个实际工程问题时, 首先要把实际问题归结为一个数学模型, 在建立数学模型时, 往往总要忽略一些次要因素, 附加上许多限制条件, 因此, 这样建立起来的理想化的数学模型, 其实只是客观现象的一种近似而粗糙的描述. 所以即使能求出这个数学模型的准确解, 也与实际问题的真正解不同. 这种数学模型的解与实际问题的解之间的误差称为模型误差. 这种误差难以用数量表示, 通常都假定数学模型是合理的.

2. 观测误差

在数学模型中往往还有一些根据观测得到的物理量, 如温度、长度、水位, 等等, 这些通过观测得到的参量难免有误差, 这种由观测产生的误差称为观测误差 (数据误差或参量误差). 在数值计算过程中我们不能控制这种误差, 但是我们要分析它对计算结果的影响. 分析的方法和分析舍入误差影响的方法大体上是一样的.

3. 截断误差

当数学模型得不到解析解时, 通常要用数值方法来求其近似解, 其近似解与精确解之间的误差称为**截断误差** (**或方法误差**). 这类误差往往是由有限过程代替无限过程时所产生的, 或者是用容易计算的问题代替不易计算的问题所产生的. 譬如, 函数 $f(x)=\ln(1+x)\ (x>0)$ 可展开为幂级数形式

$$\ln(1+x)=x-\frac{1}{2}x^2+\frac{1}{3}x^3-\cdots+(-1)^{n-1}\frac{x^n}{n}+\cdots, \tag{0.3.1}$$

如果用式 (0.3.1) 右边前 n 项

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

来近似 $\ln(1+x)$ 的无穷多项的和, 所产生的误差就是这一问题的截断误差. 据 Taylor 余项定理可知, 其截断误差为

$$\ln(1+x)-S_n(x)=\frac{(-1)^n}{(n+1)!}\frac{n!}{(1+\theta x)^{n+1}}x^{n+1},\quad \theta\in(0,1).$$

又如, 用差商 $\dfrac{f(x_0+h)-f(x_0)}{h}$ 代替导数 $f'(x_0)$ 时所产生的误差也是截断误差.

4. 舍入误差

有了求解数学模型的计算公式以后, 用计算机做数值计算时, 由于计算机所能表示的数字的位数有限, 而实际所用的数据可能位数很多, 甚至是无穷小数, 所以必须进行舍入. 此时产生的误差称为舍入误差 (或计算误差). 譬如, 用 3.14159 近似代替 π 所产生的误差

$$R=\pi-3.14159\approx 0.0000026$$

就是舍入误差.

计算结果的误差是上述四种误差累积影响的结果. 但在计算方法中我们只研究从数学问题转化为数值问题的算法时所产生的误差, 即观测误差、截断误差和舍入误差对计算结果的影响, 而一般不考虑模型误差.

0.3.2 绝对误差与相对误差

定义 0.3.1 设 x^* 为精确值, $\tilde{x}$ 为 x^* 的一个近似值, 称 $e=x^*-\tilde{x}$ 为近似值 $\tilde{x}$ 的绝对误差. 简称误差.

一般来说, 误差 $x^*-\tilde{x}$ 的具体数值是很难确定的, 只能根据具体测量工具或计算的情况估计出它的取值范围, 即给出误差绝对值 $|x^*-\tilde{x}|$ 的一个上界. 若 $\exists\varepsilon>0$, 使

$$|x^*-\tilde{x}|\leqslant\varepsilon,$$

则称 ε 为近似值 $\tilde{x}$ 的**绝对误差限**或简称**误差限**. 有了误差限 ε, 就可知道精确值 x^* 的范围

$$\tilde{x}-\varepsilon \leqslant x^* \leqslant \tilde{x}+\varepsilon. \tag{0.3.2}$$

在工程技术中, 常将不等式 (0.3.2) 表示成

$$x^* = \tilde{x} \pm \varepsilon.$$

误差限的大小在许多情况下还不能很好地表示近似值的精确程度. 例如, 测得光速的近似值为 $\tilde{x} = 299796\text{km/s}$, 误差限为 4km/s, 约为光速本身的十万分之一, 显然测量是非常精确的; 如果测量运动员的跑速, 误差限为 0.01km/s, 即 10m/s, 接近运动员的真正跑速, 显然这是十分粗糙的测量. 可见要刻画近似值的精度, 除了要看误差限的大小之外, 还应当考察这个值本身的大小, 即误差与精确值的比值.

定义 0.3.2 设 $\tilde{x}$ 为精确值 x^* 的一个近似值, 称比值

$$e_{\mathrm{r}} = \frac{x^*-\tilde{x}}{x^*} \tag{0.3.3}$$

为近似值 $\tilde{x}$ 的相对误差. 若 $\exists\delta>0$, 使 $\left|\dfrac{x^*-\tilde{x}}{x^*}\right| \leqslant \delta$, 则称 δ 为近似值 $\tilde{x}$ 的**相对误差限**.

在实际计算中, 精确值 x^* 是待求的, 因而也常将相对误差定义为

$$e_{\mathrm{r}} = \frac{x^*-\tilde{x}}{\tilde{x}}. \tag{0.3.4}$$

0.3.3 有效数字

通常当 x^* 有很多位数字时, 常常采用四舍五入的原则取前几位数字作为 x^* 的近似值 $\tilde{x}$. 例如, $x^* = \sqrt{2} = 1.41421356237\cdots$, 取前五位数字得 $\sqrt{2}$ 的近似值为

$$\tilde{x} = 1.4142,$$

其误差为 $0.00001356\cdots$, 误差限为 $0.00005 = \dfrac{1}{2}\times 10^{-4}$. 此时称 $\tilde{x}$ 精确到小数后第四位, 并称由此算起的前五位数字 14142 为 $\tilde{x}$ 的有效数字.

定义 0.3.3 如果 $\tilde{x}$ 的误差限为 $\dfrac{1}{2}\times 10^{-n}$, 即

$$|x^*-\tilde{x}| \leqslant \frac{1}{2}\times 10^{-n}, \tag{0.3.5}$$

则称 $\tilde{x}$ 精确到小数后第 n 位, 并称 $\tilde{x}$ 的第一个非零数字到这一位的全部数字为 $\tilde{x}$ 的有效数字.

显然, 若 $\tilde{x}$ 的形式为

$$\tilde{x} = \pm x_1 x_2 \cdots x_m . \alpha_1 \alpha_2 \cdots \alpha_n \cdots, \quad x_1 \neq 0,$$

则 $\tilde{x}$ 具有 $n+m$ 位有效数字; 若 $\tilde{x}$ 的形式为

$$\tilde{x} = \pm 0.00\cdots 0\alpha_{m+1}\alpha_{m+2}\cdots\alpha_n\cdots, \quad \alpha_{m+1}\neq 0,$$

则 $\tilde{x}$ 具有 $n-m$ 位有效数字.

例 0.3.1 **按四舍五入的原则分别写出数** 0.03783551, $e = 2.718281828\cdots$, 0.002030002 **具有 5 位有效数字的近似数.**

解 按有效数字的定义, 上述各数具有 5 位有效数字的近似数分别为

$$0.037836,\ 2.7183,\ 0.0020300.$$

它们的误差限分别为

$$|x^* - \tilde{x}| \leqslant \frac{1}{2}\times 10^{-1-5} = \frac{1}{2}\times 10^{-6};$$

$$|x^* - \tilde{x}| \leqslant \frac{1}{2}\times 10^{1-5} = \frac{1}{2}\times 10^{-4};$$

$$|x^* - \tilde{x}| \leqslant \frac{1}{2}\times 10^{-2-5} = \frac{1}{2}\times 10^{-7}.$$

值得注意的是, 有效数 0.00203 与 0.0020300 是不同的, 前者具有三位有效数字, 其误差限为 $\frac{1}{2}\times 10^{-2-3} = \frac{1}{2}\times 10^{-5}$, 而后者具有 5 位有效数字, 其误差限为 $\frac{1}{2}\times 10^{-7}$. 由此可见, 有效数字越多, 绝对误差越小; 容易验证, 有效数字越多, 相对误差也越小. 因此, 在计算过程中, 我们要尽量保留多的有效数字. □

0.4 数值计算中应注意的若干问题

0.4.1 病态问题与条件数

对数学问题而言, 如果输入数据有微小扰动 (即误差), 引起输出数据 (即数学问题的解) 相对误差很大, 这就是病态问题. 例如, 计算函数值 $f(x)$ 时, 若自变量 x 用 $\tilde{x}$ 代替, 其相对误差为 $\frac{x-\tilde{x}}{x}$, 函数值 $f(\tilde{x})$ 的相对误差为 $\frac{f(x)-f(\tilde{x})}{f(x)}$. 将相对误差之比的绝对值

$$\left|\frac{f(x)-f(\tilde{x})}{f(x)}\right| \Big/ \left|\frac{x-\tilde{x}}{x}\right| \approx \left|\frac{xf'(x)}{f(x)}\right| = C$$

称为计算函数值问题的条件数. 一般情况下, 自变量的相对误差不会太大, 如果条件数 C 很大, 就会引起函数值的相对误差很大. 出现这种情况的问题就是病态问题.