21世纪高等学校计算机规划教材

21st Century University Planned Textbooks of Computer Science

# 计算方法

## Computational Methods

徐士良 编著

- 具备微积分与线性代数基础知识即可
- 侧重介绍工程常用数值计算方法
- C语言描述算法程序

名家系列

人民邮电出版社
POSTS & TELECOM PRESS

21世纪高等学校计算机规划教材

21st Century University Planned Textbooks of Computer Science

# 计算方法

# Computational Methods

徐士良 编著



名家系列

人民邮电出版社
北京

**图书在版编目（CIP）数据**

计算方法 / 徐士良编著. —北京：人民邮电出版社，2009.4
（21世纪高等学校计算机规划教材）
ISBN 978-7-115-19533-3

Ⅰ. 计… Ⅱ. 徐… Ⅲ. 计算方法－高等学校－教材
Ⅳ. 0241

中国版本图书馆CIP数据核字（2009）第016404号

## 内 容 提 要

本书着重介绍工程实际中常用的一些数值计算方法。主要内容包括：数值计算的误差，线性代数方程组与矩阵，矩阵的特征值与特征向量，非线性方程，插值法，函数逼近，曲线拟合，数值积分，数值微分，常微分方程的初值问题，常微分方程的边值问题。

对于主要的数值计算方法，还给出了用C语言编写的计算机程序，读者可直接使用这些程序。

本书可作为高等理工科院校非数学专业的“数值分析”或“计算方法”等课程的教材，也可供广大工程技术人员学习参考。

21世纪高等学校计算机规划教材

**计算方法**

◆ 编　　著　徐士良
责任编辑　滑　玉
执行编辑　刘　博

◆ 人民邮电出版社出版发行　　北京市崇文区夕照寺街14号
邮编　100061　　电子函件　315@ptpress.com.cn
网址　http://www.ptpress.com.cn
三河市海波印务有限公司印刷

◆ 开本：787×1092　1/16
印张：16.25
字数：425千字　　2009年4月第1版
印数：1－3 000册　　2009年4月河北第1次印刷

ISBN 978-7-115-19533-3/TP

定价：29.00元

**读者服务热线：(010)67170985　印装质量热线：(010)67129223**
**反盗版热线：(010)67171154**

# 前　言

计算方法属于数值计算的范畴，在数学理论的基础上讨论计算的方法。现在又有计算机的帮助，可以将数值计算的方法用计算机来实现。事实上，对于一般的工程技术人员来说，并不一定需要知道数学理论的全部，而只需掌握解决工程技术问题的具体方法，以及如何用计算机来实现这些方法。

本书基于上述认识，并不着重讲解数学理论，主要介绍工程中一些常用的数值计算方法，而对于主要的方法还直接给出了用C语言描述的算法程序。也就是说，本书是从应用的角度来描述数值方法，又直接用计算机来实现这些方法，这不仅对于学生，而且对于广大工程技术人员来说，都是很有帮助的。

本书主要内容包括：数值计算的误差，线性代数方程组与矩阵，矩阵的特征值与特征向量，非线性方程，插值法，函数逼近，曲线拟合，数值积分，数值微分，常微分方程的初值问题，常微分方程的边值问题。

书中所有算法均用C语言描述，并通过了实际调试。

阅读本书只需要具备微积分与线性代数方面的基础知识。当然，还需要熟悉C语言方面的知识。

本书可作为高等理工科院校非数学专业的“数值分析”或“计算方法”等课程的教材，也可作为广大工程技术人员的自学参考书。

限于编者水平，书中难免会有错误和不当之处，恳请读者批评指正。

编　者

2008年9月

# 出版者的话

推介名师好书，共享教育资源。为促进专业教材的建设，同时满足各学科建设和教学的需要，我社经过前期充分调研并征求多方意见，规划了本套教材。该套教材汇集精华，凝练智慧，旨在传承一线教学名师的教学精髓，提高年轻教师的教学水平，从本质上培养学生的分析问题、解决问题的能力。

本套教材主要体现了如下一些基本原则和特点。

**作者权威** 本套教材的作者均为国内计算机学科中的学术泰斗或高校教学一线的教学名师，他们有着深厚的科研功底和丰富的教学经验，其中不乏国内最具盛名的经典著作的撰写人。

**定位准确** 本套教材是为普通高等院校的学生量身定做的精品教材。具体体现在：一是本套教材的作者长期从事一线科研和教学工作，对高校教学有着深刻而独到的见解；二是本套教材在选题策划阶段便多次召开调研会，对普通高校的教学需求和教材建设情况进行充分摸底，从而保证教材在内容组织和结构安排更加贴近实际教学；三是组织有关作者到较为典型的普通高等院校讲授课程教学方法，深入了解教师的教学需求，充分把握学生的理解能力，以组织教材内容，引导教师严格按照科学方法实施教学。

**教材内容与时俱进** 本套教材在充分吸收国内外最新计算机教学理念和教育体系的同时，更加注重基础理论、基本知识和基本技能的培养，集思想性、科学性、启发性、先进性和适应性于一身。本套教材内容取舍科学合理，而匠心独具的装帧设计也都恰如其分地体现了教材的内在水准，二者相得益彰，同时也显示了出版者倾心打造精品教材的良苦用心。

**一纲多本，合理配套** 根据不同的教学法，同一门课程可以有多本不同的教材，教材内容各具特色。专业基础课和专业主干课教材要配套。处理好教材统一性与多样化的关系，主教材与辅助教材以及教学参考书的关系；文字教材与软件教材的关系，实现教材系列资源配套。

本套教材反映了各位教学名师的教学水平，折射着名师的教学思想，表达着名师的教学风格，能够启发年轻教师们真正领悟教学精髓。学生通过本套教材可以掌握计算机专业的基本理论和知识，通过实践深化对理论的理解，学会理论方法的实际运用。同时，在本套教材的出版过程中，聚众人之精华，充分显示了本套教材的格调和品位。无论您是刚入杏坛的年轻教师，还是象牙塔内的莘莘学子，细细品读本套丛书，相信您总会有收获。

我们希望本套教材的出版能够为普通高等教育的教材建设工作做出贡献，欢迎各位老师和读者给我们的工作提出宝贵意见。

# 目　录

# 第 1 章 数值计算的误差

误差在数值计算中是不可避免的。也就是说，在实际的数值计算过程中，绝大多数情况下不存在绝对的严格和精确。对于一个好的计算工作者，要能够正确认识计算过程中所产生的误差。本章主要讨论误差的来源，误差的一些基本概念，运算误差以及数值计算的稳定性问题。

## 1.1 误 差 类 型

在数值计算过程中，误差的产生是不可避免的，其误差的类型也是各种各样的，它们会直接影响到计算结果的准确性。

下面简单介绍几种主要的误差。

**1. 模型误差与观测误差**

在解决工程实际问题时，为了便于进行数值计算，一般首先需要将实际问题归纳为数学问题，这就是工程上常说的需要建立一个合适的数学模型。

一般来说，在将实际问题归纳为数学问题时，总要附加某些条件限制，并且还要忽略一些次要因素，以便建立起一个“理想化”的数学模型。因此，这样得到的数学模型实际上只是客观现象的一种近似描述。而这种经过归纳后的数学描述上的近似，必然也就引进了误差。这种数学描述上的近似所引进的误差称为模型误差。

在构造数学模型时，为了对问题本身做抽象近似，除了忽略一些次要因素外，还需要对某些主要因素通过实验观测取得各种有效数据，根据实验观测到的数据进行分析总结，从而确定数学模型中的各种参数。由于条件的限制，通过实验观测到的数据与真值之间往往是有一定差异的，这也就给计算引进了一定的误差，这种误差称为观测误差。

**2. 截断误差与方法误差**

数学模型建立后，计算机还不能直接处理。这是因为，对于计算机来说，只能做一些它所规定的，并且是有限次的运算或判断，以及在一些规定的设备上进行输入与输出。因此，还必须为数学模型建立一个便于用计算机进行计算的近似公式。

大家知道，许多数学运算（如微分、积分与无穷级数求和等）是通过极限过程来定义的，而实际上计算机只能完成有限次的算术运算与逻辑运算。因此，在实际应用时，还需要将数学模型变成实际可行的解题方案，即将数学模型加工成算术运算与逻辑运算的有限序列，而这种加工又往往表现为对某种无穷过程的“截断”或计算方法的近似。例如，对于收敛的无穷级数，通常用它前面的有限项之和来近似代替无穷级数的和，实际上抛弃了无穷级数后面的无穷多项，由此便

产生了误差，这类误差称为截断误差。又例如，用梯形公式计算积分的近似值，这方法本身就有一定的误差，这类误差称为方法误差。

**3. 舍入误差**

解题方案确定之后，就可以通过某种工具来具体描述解题步骤，然后编制计算机程序，调试通过后就可以在计算机上正式运行，最后得到所需要的结果。

计算机与其他任何计算工具一样，总是受有效数字位数的限制，在进行数值计算时，其处理的数据总是近似的。在计算机中，任何数据都要转换成二进制形式才能进行处理，而绝大部分的数值型数据是无法精确地用二进制形式表示的，也就是说，即使是一个准确的数，为了用计算机进行处理，在转换成二进制数时就变成近似的，如实数 0.1 在计算机中就不能被精确表示。某些实数在计算机中即使能精确表示，但对它们做运算后，其运算结果有可能不能被精确表示。例如，实数 1.0 与 3.0 虽然在计算机中都能精确表示，但 1.0/3.0 却不能被精确表示。因此，在计算机中，由于参加运算的数据只能具有有限位的有效数字，其超过部分都将被无情地舍掉，这就产生了误差，这种误差称为舍入误差。

虽然数值计算中的误差是不可避免的，但是，在解决实际问题时，应该尽量减少产生误差的机会，尽量减小某些误差或将它们限制在许可的范围之内。这是因为，误差在计算过程中会产生负面的效应。例如，某个参数由于观测引进的误差可能是微不足道的，或者少量的舍入误差对中间的计算结果影响不大，但是，这些误差经过计算机的千百万（甚至更多）次的运算以后，误差的积累就可能大得惊人。初始数据的微小误差也可能会引起严重错误，甚至会导致完全错误的结果。

## 1.2 误 差 定 义

**1. 绝对误差**

**【定义 1-1】** 设 $x$ 为准确值，$x^*$为其近似值。则

$$E(x)=x-x^*$$

称为近似值 $x^*$关于准确值 $x$ 的绝对误差。

一般来说，由于准确值 $x$ 是未知的，因此，无法根据定义 1-1 准确地计算出某个近似值的绝对误差，而只能根据测量或计算的具体情况估计出绝对误差值的一个范围，也就是估计出$|E(x)|$的一个上界。

设

$$|E(x)|=|x-x^*|\leqslant \eta$$

则称 $\eta$为近似值 $x^*$关于准确值 $x$ 的绝对误差限。

前面说过，一般无法计算出由定义 1-1 所定义的绝对误差，因此，工程上就将绝对误差限称为绝对误差。在本书中，如果没有特殊说明，绝对误差即指绝对误差限，有时就简称为误差。

当估计出近似值 $x^*$关于准确值 $x$ 的绝对误差限 $\eta$后，工程上可以用以下两种方法表示准确值 $x$ 所在的范围。

$$x-x^*\leqslant x\leqslant x^*+\eta$$

或

$$x=x^*\pm\eta$$

在计算函数值 $f(x)$时，当自变量 $x$ 有一个误差时，其计算得到的函数值也有一个误差。如果

给出了自变量 $x$ 的绝对误差为 $E(x)$，则函数值的绝对误差可以用下式来估计。

$$E[f(x)]=f'(x)E(x)$$

**2. 相对误差**

绝对误差的大小反映了近似值偏离准确值的程度，还不能完全反映近似值的准确程度。例如，设有两个量 $x$ 和 $y$，其中 $x$=10±1，$y$=1000±5。显然，近似值 $y^*$=1000 的绝对误差比近似值 $x^*$=10 的绝对误差大了 4 倍，但并不能说 $y^*$的准确程度要比 $x^*$差，实际上正好相反，$y^*$的准确程度要优于 $x^*$。

为了能够确切地表示一个近似值的准确程度，我们引进一个相对误差的概念。

**【定义 1-2】** 设 $x$ 为准确值，$x^*$为其近似值。则

$$E_r(x)=\frac{E(x)}{x}=\frac{x-x^*}{x}$$

称为近似值 $x^*$关于准确值 $x$ 的相对误差。

实际上，由于准确值 $x$ 一般是不知道的，因此，相对误差通常又定义为

$$E_r(x)=\frac{E(x)}{x^*}=\frac{x-x^*}{x^*}$$

由上述定义可以看出，相对误差说明了近似值 $x^*$关于准确值 $x$ 的绝对误差 $E(x)$与近似值本身比较起来所占的比例，因而更客观地反映了该近似值的准确程度。

和绝对误差一样，由于准确值 $x$ 一般不知道，其绝对误差 $E(x)=x-x^*$无法准确地算出，因此也就无法确定出相对误差 $E_r(x)$的准确值，而只能估计出它的一个范围。

如果

$$|E_r(x)|=\left|\frac{x-x^*}{x^*}\right|\leqslant\delta$$

则称 $\delta$ 为近似值 $x^*$关于准确值 $x$ 的相对误差限。

在实际应用中，就将相对误差限称为相对误差。同样，在本书中，如果没有特殊说明，相对误差即指相对误差限。

根据相对误差的定义，相对误差限 $\delta$ 与绝对误差限 $\eta$ 之间有如下关系：

$$\delta=\left|\frac{\eta}{x^*}\right|$$

在工程实际中，一般用百分比来表示相对误差。即

$$E_r(x)=\frac{E(x)}{x^*}\times100\%=\frac{x-x^*}{x^*}\times100\%$$

**【例 1-1】** 设 $x$ 的绝对误差为 $\eta$，试估计 $f(x)=e^x$ 的相对误差。

**解：** $|E_r[f(x)]|=\left|\frac{E[f(x)]}{f(x)}\right|=\left|\frac{f'(x)E(x)}{f(x)}\right|=\left|\frac{e^xE(x)}{e^x}\right|=|E(x)|=\eta$

**【例 1-2】** 计算球体积要使相对误差限为 1%，问度量半径 $R$ 时允许的相对误差限是多少？

**解：** 球体积的计算公式为

$$V(R)=\frac{4}{3}\pi R^3$$

当半径 $R$ 有一误差 $E(R)$时，球体积的相对误差为

$$|E_r[V(R)]|=\left|\frac{E[V(R)]}{V(R)}\right|=\left|\frac{V'(R)E(R)}{V(R)}\right|=\left|\frac{4\pi R^2E(R)}{\frac{4}{3}\pi R^3}\right|=3\left|\frac{E(R)}{R}\right|=3|E_r(R)|$$

现要求$|E_r[V(R)]|=3|E_r(R)|\leqslant 1\%$，即$|E_r(R)|\leqslant 0.3333\%$。

**【例 1-3】** 在一个化学反应中，已知一个反应物的转换分数的真值为 0.69，但操作员在他的分析中却认为是 0.63。估计在他的分析中的误差。

**解**：误差为

$$|E|=0.69-0.63=0.06$$

相对误差为

$$E_r=\frac{0.06}{0.69}\times 100\%=8.7\%$$

# 1.3 有 效 数 字

在实际应用中，除了用相对误差来反映一个近似值的准确程度外，还经常用有效数字的位数来反映近似值的准确程度。

**【定义 1-3】** 设 $x$ 为准确值，$x^*$为其近似值。若

$$|x-x^*|\leqslant\frac{1}{2}\times 10^{-k}$$

则称用 $x^*$近似表示 $x$ 时准确到小数点后第 $k$ 位；并称从小数点之后的第 $k$ 位数字起直到最左边的非零数字之间的所有数字为有效数字；称有效数字的位数为有效数位。

**【定义 1-4】** 设 $x$ 为准确值，$x^*$为其近似值，且表示成如下形式：

$$x^*=\pm 10^m(x_1\times 10^{-1}+x_2\times 10^{-2}+\cdots+x_n\times 10^{-n})$$

其中 $x_1$，$x_2$，…，$x_n$都是 0～9 这 10 个数字之一，且 $x_1\neq 0$，$n$ 是正整数，$m$ 是整数。若

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

则称近似值 $x^*$具有 $n$ 位有效数字。

**【例 1-4】** 设$\sqrt{20}$=4.472136 具有 7 位有效数字，试确定下列各近似值的有效数字位数。

（1）$\sqrt{20}\approx 4.42$；（2）$\sqrt{20}\approx 4.471\,64$；（3）$\sqrt{20}\approx 4.469\,576$。

**解**：设 $x=\sqrt{20}=4.472136$，$x^*$为各近似数。

（1）设$\sqrt{20}\approx x^*=4.42=10^1\times 0.442$，其中 $m=1$。

$$|x-x^*|=|4.472136-4.42|=0.052136\leqslant 0.5\times 10^0=0.5\times 10^{m-1}$$

因此，根据定义 1-3 有 $k=0$，即该近似数准确到个位数，共有 1 位有效数字。根据定义 1-4 有 $n=1$，同样是具有 1 位有效数字。

（2）设$\sqrt{20}\approx x^*=4.47164=10^1\times 0.447164$，其中 $m=1$。

$$|x-x^*|=|4.472136-4.47164|=0.000396\leqslant 0.5\times 10^{-3}=0.5\times 10^{m-4}$$

因此，根据定义 1-3 有 $k=3$，即该近似数准确到小数点后第 3 位，共有 4 位有效数字。根据定义 1.4 有 $n=4$，同样是具有 4 位有效数字。

（3）设$\sqrt{20}\approx x^*=4.469576=10^1\times 0.4469576$，其中 $m=1$。

$$|x-x^*|=|4.472136-4.469576|=0.00256\leqslant 0.5\times 10^{-2}=0.5\times 10^{m-3}$$

因此，根据定义 1-3 有 $k=2$，即该近似数准确到小数点后第 2 位，共有 3 位有效数字。根据定义 1-4 有 $n=3$，同样是具有 3 位有效数字。

由这个例子可以看出，关于有效数字的定义 1-3 与定义 1-4 是等价的。

【例 1-5】　设近似值 $x^*$=0.937 具有 3 位有效数字，估计用 $x^*$代替 $x$ 时的相对误差。并估计计算函数 $f(x)=\sqrt{1-x}$ 值时的绝对误差与相对误差。

**解**：因为 $x^*$=0.937 具有 3 位有效数字，所以

$$|E(x)|=|x-x^*|\leqslant 0.0005$$

由此可得用 $x^*$代替 $x$ 时的相对误差为

$$|E_r(x)|=\left|\frac{x-x^*}{x^*}\right|\leqslant\frac{0.0005}{0.937}\approx 0.000534=0.0534\%$$

函数 $f(x)=\sqrt{1-x}$ 值时的绝对误差为

$$|E[f(x)]|=|f(x)-f(x^*)|=|f'(x^*)E(x)|=\left|\frac{-1}{2\sqrt{1-x^*}}E(x)\right|=\frac{0.0005}{2\sqrt{1-0.937}}\approx 0.001$$

函数 $f(x)=\sqrt{1-x}$ 值时的相对误差为

$$|E_r[f(x)]|=\frac{|E[f(x)]|}{|f(x^*)|}=\frac{0.001}{\sqrt{1-0.937}}\approx 0.004=0.4\%$$

最后需要说明的是，在书写或表示一个近似值时，通常有以下两种方式。

① 注明该近似值 $x^*$及其绝对误差 $\eta$，即将近似值写成 $x^*\pm\eta$。

② 在没有注明近似值的绝对误差时，则默认该近似值准确到末位数字。在这种情况下，要求从其最左边的非零数字起，直到最右边的一位数字止，都是有效数字。例如，0.00203 具有 3 位有效数字，分别为 2，0，3；3.14 也具有 3 位有效数字，分别为 3，1，4。特别需要指出的是，在这种表示方式中，0.23 与 0.2300 的有效数字的位数是不一样的，前者具有 2 位有效数字，其绝对误差不超过 0.005，而后者具有 4 位有效数字，其绝对误差不超过 0.00005。

在工程实际中，有效数字的位数反映了已知值的准确程度，通常情况下，已知值的最后一位是不准确的。例如，假设一个金环的重量被报为 10.5g（即 3 位有效数字），这预示重量大概在 10.45g 和 10.55g 之间。但是，如果重量被报为 10.50g（4 位有效数字），则我们可以认为重量在 10.495g 和 10.505g 之间。

在某些实际问题中，初始数据的微小误差对计算结果的影响会很大，即在这种情况下，计算结果强烈依赖于初始数据的有效数字位数。下面的例子说明了这个问题。

【例 1-6】　设有下列矩阵

$$\tilde{H}=\begin{bmatrix}1 & 1/2 & 1/3\\ 1/2 & 1/3 & 1/4\\ 1/3 & 1/4 & 1/5\end{bmatrix}$$

求这个矩阵的逆。

**解**：矩阵求逆的方法将在第 2 章中介绍，现在只是给出结果。

如果对元素 1/3 只保留 2 位有效数字，则矩阵和它的逆分别为

$$\tilde{H}=\begin{bmatrix}1 & 0.5 & 0.33\\ 0.5 & 0.33 & 0.25\\ 0.33 & 0.25 & 0.2\end{bmatrix},\quad \tilde{H}^{-1}=\begin{bmatrix}55.56 & -277.78 & 255.56\\ -277.78 & 1446.03 & -1349.21\\ 255.56 & -1349.21 & 1269.84\end{bmatrix}$$

如果对元素 1/3 保留 4 位有效数字，则矩阵和它的逆分别为

$$\tilde{H}=\begin{bmatrix}1 & 0.5 & 0.3333\\ 0.5 & 0.3333 & 0.25\\ 0.3333 & 0.25 & 0.2\end{bmatrix},\quad \tilde{H}^{-1}=\begin{bmatrix}9.06 & -36.32 & 30.3\\ -36.32 & 193.68 & -181.56\\ 30.3 & -181.56 & 181.45\end{bmatrix}$$

如果我们在矩阵求逆的过程中保持用分数进行运算，则矩阵和它的逆分别为

$$\tilde{H}=\begin{bmatrix}1 & 1/2 & 1/3\\ 1/2 & 1/3 & 1/4\\ 1/3 & 1/4 & 1/5\end{bmatrix},\quad \tilde{H}^{-1}=\begin{bmatrix}9 & -36 & 30\\ -36 & 192 & -180\\ 30 & -180 & 180\end{bmatrix}$$

由此可以看出，对元素 1/3 取的有效数字位数彻底改变了矩阵求逆的结果。这个矩阵称为 Hilbert 矩阵，它是一个坏条件矩阵。像 Hilbert 矩阵那样，对微小误差很敏感的系统称为病态系统。在第 2 章中将进一步介绍病态系统的概念。

# 1.4 数值计算的运算误差

前面已经提到，在数值计算过程中，误差的存在是不可避免的。特别需要指出的是，由于受计算机（其他计算工具也是如此）有效数字位数的限制，在实际运算过程中的每一步都不可避免地要产生误差。因此，尽量减小运算过程中每一步的误差，或者将误差限制在最小的范围之内，是数值计算中要重点考虑的问题之一。

有运算就会产生误差，但不同的运算，或不同的运算步骤所产生的误差是不一样的。下面举例说明不同的运算产生不同误差的情况。

**【例 1-7】** 设函数

$$g(x)=10^3(1-\cos x)$$

用四位数学用表计算 $g(2^\circ)$的近似值。

**解**：求解这个问题可以用以下两种方法。

解法 1

查四位数学用表得 $\cos 2^\circ=0.9994$。因此

$$g(2^\circ)=10^3(1-\cos 2^\circ)=10^3(1-0.9994)=0.6$$

解法 2

利用三角恒等式

$$1-\cos x=2\sin^2(x/2)$$

得到

$$g(x)=10^3(1-\cos x)=2\times 10^3\sin^2(x/2)$$

查四位数学用表得 $\sin 1^\circ=0.0175$。因此

$$g(2^\circ)=2\times 10^3\sin^2 1^\circ=2\times 10^3\times(0.0175)^2=0.6125$$

在以上两种解法中，初始数据（查四位数学用表得到）都准确到小数点后的第 4 位，但最后得到的结果并不相同。究竟哪个结果更准确一些呢？下面用相对误差的概念进行分析。

在解法 1 中，假设准确值 $A=\cos 2^\circ$，则函数的准确值为

$$g(2^\circ)=10^3(1-A)$$

$A$ 的近似值为 $A^*=0.9994$，则函数的近似值为

$$g^*(2^\circ)=10^3(1-A^*)$$

其中 $|A-A^*|\leqslant\frac{1}{2}\times 10^{-4}$。由此可以计算函数的相对误差为

$$|E_r[g(2^\circ)]|=\left|\frac{E[g(2^\circ)]}{g^*(2^\circ)}\right|=\left|\frac{g(2^\circ)-g^*(2^\circ)}{g^*(2^\circ)}\right|=\left|\frac{10^3(1-A)-10^3(1-A^*)}{10^3(1-A^*)}\right|$$

$$=\left|\frac{A-A^*}{1-A^*}\right|\leqslant\frac{\frac{1}{2}\times10^{-4}}{1-0.9994}=\frac{1}{12}=8.3\%$$

在解法 2 中，假设准确值 $B=\sin1°$ ，则函数的准确值为

$$g(2°)=2\times10^3\times B^2$$

B 的近似值为 $B^*=0.0175$，则函数的近似值为

$$g^*(2°)=2\times10^3\times(B^*)^2$$

其中 $|B-B^*|\leqslant\frac{1}{2}\times10^{-4}$ 。由此可以计算函数的相对误差为

$$|E_r[g(2°)]|=\left|\frac{E[g(2°)]}{g^*(2°)}\right|=\left|\frac{g(2°)-g^*(2°)}{g^*(2°)}\right|=\left|\frac{2\times10^3\times B^2-2\times10^3\times(B^*)^2}{2\times10^3\times(B^*)^2}\right|$$

$$=\left|\frac{B^2-(B^*)^2}{(B^*)^2}\right|=\frac{|B-B^*|\,|B+B^*|}{(B^*)^2}$$

$$=\frac{2|B-B^*|}{B^*}\leqslant\frac{2\times\frac{1}{2}\times10^{-4}}{0.0175}=\frac{1}{175}=0.57\%$$

其中 $B+B^*\approx2B^*$。

由以上分析可以看出，解法 2 所得结果的相对误差比解法 1 所得结果的相对误差要小，即解法 2 所得的结果更准确一些。实际上，通过更精确的计算，具有 6 位有效数字的答案为 0.609173。

上述例子说明，同一个计算问题，可以有多个运算过程，而不同的运算过程所得到的结果其准确程度是不同的。因此，在进行数值计算时，应该考虑各种运算的误差。

在分析运算误差时，通常要考虑以下一些原则。

（1）两个相近的近似数相减，会严重丢失有效数字。

当两个近似数相近时，这两个近似数的最左边若干位数字相同，因此，它们做减法运算后，最左边的有效数字变成了 0，因而最终结果中的有效数字位数会减少。对于这种现象，可以通过相对误差的概念来说明。

假设要做如下减法运算：

$$y=x-A$$

其中，$A$ 与 $x$ 均为准确值。为了简单起见，假设 $A$ 在运算时不发生误差，而 $x$ 在运算时存在误差，其近似值为 $x^*$。即实际运算为

$$y^*=x^*-A$$

因此，当用 $x^*$近似代替 $x$ 时，$y$ 的相对误差为

$$|E_r(y)|=\left|\frac{E(y)}{y^*}\right|=\left|\frac{(x-A)-(x^*-A)}{x^*-A}\right|=\left|\frac{x-x^*}{x^*-A}\right|=\left|\frac{E(x)}{x^*-A}\right|$$

由上式可以看出，在 $x$ 的绝对误差 $E(x)$不变时，如果 $x^*$越接近 $A$，则 $y$ 的相对误差的绝对值$|E_r(y)|$会变得越大（因为分母上 $x^*-A$ 的绝对值越小），而相对误差的增大必然会导致有效数字位数的减少。在例 1-7 的解法 1 中，$\cos2°=0.9994$ 具有 4 位有效数字，但在做 $1-\cos2°$ 的运算后，其结果为 0.0006，丢失了最左边的 3 位有效数字，结果至多保留了 1 位有效数字。这就说明了在两个相近的近似数作减法运算时，其结果的准确程度会大大降低。

由以上分析可知，在做数值计算时，为了避免运算过程中丢失有效数字，应避免做减法运算，特别要避免两个相近的近似数做减法运算。例 1-7 中的解法 2 就是利用三角恒等式避免了两个相

近的数做减法运算，减小了运算结果的相对误差，从而提高了运算结果的准确程度。在实际应用时，一般都可以利用数学中的恒等式或等价关系，尽量消去计算过程中的减法运算。下面是一些常见的公式变换的例子。

当 $x_1$ 很接近 $x_2$ 时，变换公式为

$$\lg x_1 - \lg x_2 = \lg \frac{x_1}{x_2} \quad 与 \quad \ln x_1 - \ln x_2 = \ln \frac{x_1}{x_2}$$

当 $x$ 接近于 0 时，变换公式为

$$\frac{1-\cos x}{\sin x} = \frac{\sin x}{1+\cos x}$$

当 $x$ 充分大时，变换公式为

$$\operatorname{arctg}(x+1) - \operatorname{arctg} x = \operatorname{arctg} \frac{1}{1+x(x+1)}$$

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

当 $f(x)$ 与 $f(x^*)$ 很接近，但又要做 $f(x) - f(x^*)$ 运算时，为避免有效数字的丢失，可以用台劳(Taylor)展开式，即

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

**【例 1-8】** 求一元二次方程

$$x^2 - (10^{12}+1)x + 10^{12} = 0$$

的两个实根。如果用通常的求根公式，并且假设计算工具具有 7 位有效数字，则计算过程如下：

$$\begin{aligned} x_{1,2} &= \frac{(10^{12}+1) \pm \sqrt{(10^{12}+1)^2 - 4\times 1\times 10^{12}}}{2\times 1} \\ &= \frac{10^{12} \pm \sqrt{10^{24} - 4\times 10^{12}}}{2} \\ &= \frac{10^{12} \pm 10^{12}}{2} = \begin{cases} 10^{12} \\ 0 \end{cases} \end{aligned}$$

最后的计算结果为

$$x_1 = 10^{12}, \quad x_2 = 0$$

经检验发现，$x_1 = 10^{12}$ 满足原方程，是方程的根，而 $x_2 = 0$ 不满足原方程。

这个例子表明，理论上的解题方案，在有些情况下不一定能用。实际上，一元二次方程

$$Ax^2 + Bx + C = 0$$

的求根公式

$$x_{1,2} = \frac{-B \pm \sqrt{B^2 - 4AC}}{2A}$$

对于一般的系数 $A$，$B$，$C$ 来说可以使用，但如果对于某些系数 $A$，$B$，$C$ 使求根公式分子中的两项 $(-B)$ 与 $\sqrt{B^2 - 4AC}$ 很接近时，计算结果就会严重丢失有效数字，在这种情况下就不能简单地使用求根公式了。

一般来说，对于一元二次方程

$$Ax^2 + Bx + C = 0$$

可以用以下方法求出两个实根：利用求根公式先计算一个实根（使求根公式分子中的两项为同号

相加，从而避免两个同号数相减的情况），然后利用韦达定理计算另一个实根。即

$$\begin{cases} x_1 = \dfrac{-B - \mathrm{sgn}(B)\sqrt{B^2 - 4AC}}{2A} \\ x_2 = \dfrac{C}{Ax_1} \end{cases}$$

其中 sgn(*B*)为 *B* 的符号函数，即

$$\mathrm{sgn}(B) = \begin{cases} +1, & B \geqslant 0 \\ -1, & B < 0 \end{cases}$$

由这个例子可以看出，有些数学上的计算公式有时不能真正用于实际计算，这些公式理论上是正确的，但由于参与公式中各种运算的数是有误差的近似数，而误差在运算过程中会发生某些不良效应（两个相近的近似数相减，会严重丢失有效数字），从而导致计算结果的误差变得很大，使结果不可靠。这就导致理论上的解题方案与实际能用性之间会存在很大的差异。

（2）除数绝对值较小时，商的绝对误差会增大。

假设

$$z = \frac{x}{y}$$

其中 $x$ 和 $y$ 的近似值分别为 $x^*$和 $y^*$，则

$$z^* = \frac{x^*}{y^*}$$

$z$ 的绝对误差为

$$E(z) = z - z^* = \frac{x}{y} - \frac{x^*}{y^*} = \frac{xy^* - yx^*}{yy^*} = \frac{y^*(x - x^*) - x^*(y - y^*)}{yy^*}$$

$$\doteq \frac{y^*E(x) - x^*E(y)}{yy^*} \approx \frac{y^*E(x) - x^*E(y)}{(y^*)^2}$$

由上式可以看出，在做除法运算时，如果分母（即除数 $y^*$）的绝对值越小，则商的绝对误差就越大。

（3）在运算过程中，必须注意合理安排运算顺序，以便提高运算的精度或保护重要的参数。

在数值计算中，运算顺序对于结果的可靠性有时是至关重要的。由于受计算机有效数字位数的限制，公理系统中的许多规律在实际计算过程中就不再适用了。例如，在公理系统中，加法运算满足交换律，即

$$a+b+c \equiv a+c+b$$

但在实际计算时就不一定满足了。假设 $a=10^{12}$，$b=1$，$c=-10^{12}$，如果计算机系统具有 6 位有效数字（各程序设计语言中的单精度运算就是如此），则按照该顺序运算结果为

$$a+b+c = 10^{12}+1+(-10^{12}) = 0$$

这是因为，在作 $a+b$ 运算时，$a=10^{12}$ “吃掉”了 $b=1$，其结果为 $10^{12}$，再与 $c=-10^{12}$ 相加时，正好互相抵消，最后结果为 0。但如果上述计算按 $a+c+b$ 顺序运算，其运算结果为

$$a+c+b = 10^{12}+(-10^{12})+1 = 1$$

称这种运算顺序保护了对结果起重要作用的参数（即 1）。

由此可以看出，在做数值运算时，应事先分析一下参与运算的各数值的数量级，然后合理地安排它们的运算顺序，这样，一些重要的参数就不致于在运算过程中被其他参数“吃掉”。特别是在做连加运算时，合理安排各运算对象的运算顺序，可以获得较高精度的结果。

下面举例说明在运算过程中大数“吃掉”小数的现象。

【例 1-9】 在具有 4 位有效数字的计算工具上做下列运算：

① $10^3(0.8961)+10^{-5}(0.4688)$

$\rightarrow 10^3(0.8961)+10^3(0.0000)$ （对阶）

$\rightarrow 10^3(0.8961)$ （规格化）

其结果是大数“吃掉”了小数。

② $10^0(0.6108)+10^3(0.6871)$

$\rightarrow 10^3(0.0006)+10^3(0.6871)$ （对阶）

$\rightarrow 10^3(0.6877)$ （规格化）

其结果是大数“吃掉”了部分的小数。

③ $10^{-1}(0.3311)-10^{-1}(0.3307)$

$\rightarrow 10^{-1}(0.0004)$

$\rightarrow 10^{-4}(0.4000)$ （规格化）

其结果的有效数字位数大大减少，此时尽管印出的结果为 $10^{-4}(0.4000)$（这是系统进行了规格化的结果），但包括 4 在内都不一定是有效数字。

（4）注意计算步骤的简化，减少算术运算的次数。

如前所述，数值运算过程中的每一步都有可能产生误差。而且，运算过程中每一步的误差都还有可能传递到下一个运算步骤中，误差的这种传递有时是增大的，有时是减小的。同时，运算过程中每一步产生的误差，也还会积累到最终的结果中去，只不过这种误差的积累有时是增加的，有时则因互相抵消而减少。总之，在数值计算过程中都有可能引起使结果误差增大的误差传播或误差积累问题。如果在数值计算中简化了计算步骤，一方面可以减小计算工作量；另一方面还由于减少了算术运算的次数，从而可以减少产生误差的机会，同时也可以减少误差的积累。

（5）“坏条件”函数值的判别。

当函数 $f(x)$ 在区间 $[a, b]$ 上满足

$$|f'(x)|\leqslant 1$$

时，自变量 $x$ 的微小变化所引起的函数值 $f(x)$ 的变化更微小。如果对于某一点 $\hat{x}$，$|f'(\hat{x})|$ 的值很大，则称函数 $f(x)$ 在 $\hat{x}$ 这一点的计算在绝对误差意义下是坏条件的。$|f'(\hat{x})|$ 的值究竟多大才算坏条件，可以随解题的精度要求而定。

【例 1-10】 对于函数

$$f(x)=\frac{1}{n}\sin(n^2x)$$

当 $n$ 很大时，在 $x=0$ 附近，其一阶导数

$$f'(x)=n\cos(n^2x)$$

的绝对值 $|f'(x)|$ 很大。因此，在 $x=0$ 附近计算函数值 $f(x)$ 时，在绝对误差意义下是坏条件的。

如上所述，对于实际问题的计算效果可以用绝对误差来衡量。但在更多的问题中，其计算效果要用相对误差来衡量，因为，只有相对误差才从本质上反映近似值的准确程度。

如果 $x\neq 0$，设

$$E(x)=x-x^*,\quad E[f(x)]=f(x)-f(x^*)$$

则

$$E_r(x)=\frac{E(x)}{x}=\frac{x-x^*}{x}$$