普通高等教育“十一五”国家级规划教材

数值逼近

（第二版）

蒋尔雄　赵风光　苏仰锋　编著

博学·数学系列

复旦大学出版社
www.fudanpress.com.cn

普通高等教育“十一五”国家级规划教材

数值逼近

（第二版）

蒋尔雄　赵风光　苏仰锋　编著

博学·数学系列

复旦大学出版社
www.fudanpress.com.cn

图书在版编目(CIP)数据

数值逼近(第二版)/蒋尔雄,赵风光,苏仰锋编著.
—上海:复旦大学出版社,2008.7
(复旦博学·数学系列)
ISBN 978-7-309-06133-8

Ⅰ.数… Ⅱ.①蒋…②赵…③苏… Ⅲ.数值逼近 Ⅳ.0174.41

中国版本图书馆 CIP 数据核字(2008)第 095772 号

数值逼近(第二版)
蒋尔雄 赵风光 苏仰锋 编著

出版发行 复旦大学出版社 上海市国权路 579 号 邮编 200433
86-21-65642857(门市零售)
86-21-65100562(团体订购) 86-21-65109143(外埠邮购)
fupnet@fudanpress.com http://www.fudanpress.com

责任编辑 范仁梅
出 品 人 贺圣遂

印 刷 上海华文印刷厂
开 本 787×960 1/16
印 张 16.5
字 数 287 千
版 次 2008 年 7 月第二版第一次印刷
印 数 1—4 100

书 号 ISBN 978-7-309-06133-8/O·415
定 价 32.00 元

内容提要

本书是大学计算数学专业的基础课程——数值逼近的教材，主要讲述了数值逼近的理论和各种数值逼近方法．全书内容包括：函数的插值、样条插值和曲线拟合、最佳逼近、数值积分、快速Fourier变换、函数方程求根等．学生仅需具备数学分析或高等数学、高等代数的预备知识即可阅读．

本书作者根据自己连续多年的教学经验，结合信息与科学计算专业对学生编程能力的要求，在本书的修订过程中重视学生的动手能力．一方面学生通过本教材的学习能够提高Matlab编程的水平；另一方面学生可以通过本教材所附的程序，观察、理解教材中的理论、算法在实际计算时的表现及效果，使学生在学习中获得成就感，提高学生的学习兴趣．

再 版 前 言

本书的第一版已经在很多高校中使用过. 本人也使用该教材在复旦大学数学科学学院连续使用过多年. 本次修订的主要内容是修改了第一版中的一些错误, 并添加了一些 Matlab[①]程序供参考.

数值逼近是信息与计算科学专业的第一门基础专业课. 学生除了掌握基本的计算理论和方法外, 必须通过用计算机语言将课本上的算法进行实现来进一步理解这些算法及相关的理论. 选择 Matlab 作为工具来实现算法是基于如下的想法: 第一, Matlab 本身是一种科学计算环境, 很多科学工程计算中都使用 Matlab 进行数值计算. 第二, Matlab 具有丰富的数学函数, 学生在学习、理解当前的教学内容时可以直接使用其他的函数. 例如, 在学习样条插值、最佳均方逼近时, 可以直接使用 `b=A\x` ; 来求解线性方程组或最小二乘问题 $Ax = b$. 第三, Matlab 体现了当前科学计算的基本要求, 如尽量使用向量运算而不是用循环. 第四, Matlab 具有丰富的图形工具, 容易对于计算结果进行可视化.

本书所附的光盘中包含有各章所用的 Matlab 文件. 核心程序使用 Matlab 指令 `pcode` 编译成 `.p` 文件. 该类型文件与 Matlab 标准的 `.m` 文件具有相同的功能, 但是代码不是可读的. 这样做的目的是希望该书的使用者能够自己编写这些 `.m` 文件, 并与 `.p` 文件运行结果相比较, 提高编程能力.

本次的书稿先经第一版的重新输入再修改所得. 图像也经过重新绘制. 虽然经过多人反复校阅, 仍然可能出现各种排版错误. 希望读者加以指正. 特别要感谢何力、朱一超在校对、绘图、排版等方面付出的辛勤劳动. 还要特别感谢复旦大学出版社的范仁梅老师, 没有她的多方面的努力, 本次再版是不可能的.

苏仰锋
yfsu@fudan.edu.cn
复旦大学数学科学学院
2007 年 12 月 27 日于复旦

① Matlab 是 Mathworks 公司的注册商标.

前　言

1977 年 10 月在上海召开的教材会议上, 决定把计算数学专业的基础课**计算方法**分成三部分, 其中第一部分就是**数值逼近**, 从那以后, 复旦大学每年都开设数值逼近课, 我自己也教过这门课, 有两点体会. 第一点: 在听课的几十名学生中, 不会有很多, 也不必要有很多学生以后会去从事数值逼近的研究, 因此没有必要讲得太专门, 太专门的东西待他自己需要时再学也不迟. 第二点: 数值逼近的内容很多, 很多理论的产生都有它的客观需要: 或是实际问题的需要或是理论本身的需要. 在向学生介绍这些内容时, 应尽量使学生知道问题是怎样提出来的, 有问题才能引起学生的思考, 然后再引导学生探讨这些问题是如何解决的. 这样不但能促进学生学好这些内容, 而且还能促进学生学习如何提出问题和解决问题, 有利于培养学生的能力. 这两点体会反映在我们这次编写的数值逼近教材中.

计算数学专业的数值逼近课, 常安排在二年级下学期或三年级上学期, 学生尚未学过泛函分析, 因此本书在学生仅有**数学分析**和**高等代数**的基础上编写的, 有**高等数学**基础的学生也能学好.

本书介绍了很多算法, 但没有将它们的算法语言写在书中, 这是因为市场上已有很多算法手册, 一般算法的算法语言都能找到. 对每种算法的算法语言, 在明白了其算法思想后大都是容易看懂的. 但是书中有些习题还是希望读者用计算机算一下, 以获得有关算法的感性认识.

国内外数值逼近教材或介绍数值逼近内容的教材, 已看到不少, 各有各的特点和创新之处. 本书的特点, 除上述介绍外, 还有些自己的东西. 譬如: 第一章中浮点数四则预算的基本定理 (本书第一章定理 5), 看似简单, 很多书都有介绍, 但没有分析清楚, 本书给出了严格证明. 第二章中的重节点差商, 很多书也都提到, 但常常不够系统化, 没有指出可以用重节点差商来直接构造复杂的埃尔米特插值多项式. 再有第五章中对于振荡函数的积分, 本书给出一个有效的三角梯形公式等等.

本书共 7 章, 每周 4 学时, 一个学期可以教完, 在正式出版前我们已试教过. 其中第一、二、三章和第七章由蒋尔雄编写, 第四、五、六章由赵风光编写, 蒋尔雄作了适当修改和补充.

蒋尔雄

1995 年 2 月 20 日于复旦

目　录

第一章 绪 论

§1.1 什么是数值分析

数值分析 (numerical analysis) 是对各种数学问题通过数值运算, 得到数值解答的方法和理论. 因为研究的是数学问题, 所用方法是数学方法, 因此也称之为数值数学 (numerical mathematics), 数值分析是总称, 对一个数学问题通过数值运算得到数值解答的方法, 称为数值方法 (numerical method), 如果这数值方法可以在计算机上实现, 就称为数值算法 (numerical algorithm).

数值逼近, 即为各种逼近问题的数值分析, 包括插值样条、最佳一致逼近、最佳平方逼近、数值积分、线性或非线性方程组求解等内容.

要对一个数学问题求数值解答, 除了在数学上对问题进行预处理以外, 最后都要化成各种数据的一系列 $+$、$-$、$\times$、$\div$ 的四则运算, 由此获得数值结果. 一个数在十进制表示下, 可以是有限位的, 也可以是无限位的. 譬如, 边长为 1 的正方形, 它的对角线长是

$$\sqrt{2} = 1.4142\cdots,$$

这是一个无限位的数, $\pi = 3.1415926\cdots$ 也是一个无限位数. 数值运算的一个特点就是: 参与运算的数必须是有限位的, 而且位数往往是预先规定的, 譬如规定为 8 位、16 位等等, 如果运算的数是无限位的或超过规定, 那么要用“四舍五入”规则或“截断”规则, 将它们处理成规定的位数. 对于运算结果也要处理成规定的位数.

所谓“四舍五入”规则就是: 将超过规定位数的部分按下述原则去掉.

(1) 如果舍弃的部分小于保留数的最后一单位的 1/2, 那么保留的数不变. 例如

$$\pi = 3.141592\ 6\cdots,$$

如果取两位小数, 那么保留数的最后一位单位是 10^{-2}, 舍弃部分是 $0.15926\cdots \times 10^{-2}$, 小于 $1/2 \times 10^{-2}$, 因此取为 3.14.

(2) 如果舍弃的部分大于所保留数的最后一位单位的 1/2, 那么将保留数最后一位数字加 1. 例如限制 π 取 4 位小数, 最后一位单位为 10^{-4}, 但去掉的部分是 $0.926\cdots \times 10^{-4}$, 大于 $1/2 \times 10^{-4}$, 因此取成 3.1416.

(3) 如果舍弃的部分恰等于所保留数的最后一位单位的 1/2, 此时如果保留数的最后一位是奇数, 那么加 1 成偶数; 如果保留的数最后一位是偶数, 则就不动了,

例如取两位小数, 0.675 为 0.68, 而 0.605 为 0.60.

大多数计算机是采用“四舍五入”规则处理舍弃的位数的, 但也有的计算机是采用“截断”规则的.

所谓“截断”规则就是: 将超过规定位数的部分无条件地去掉, 这样 π 取 4 位小数就为 3.1415.

考虑 $\sqrt{2}$ 与 π 相乘作数值运算, 取 5 位数字进行运算, 按“四舍五入”规则 $\sqrt{2}$ 成为 1.4142, 而 π 成为 3.1416, $1.4142\times 3.1416 = 4.44285072$ 是 9 位数字. 如果结果也限制为 5 位数字, 也按“四舍五入”规则, 则要处理成 4.4429, 它就是 $\sqrt{2}\pi$ 数值计算的结果. 这样做当然会失真, 但是却也非常接近. 实际上 $\sqrt{2}\pi = 4.442882938\cdots$ 与 4.4429 是非常接近的. 数值分析的特点就是又失真, 又接近. 不失真是不可能的, 我们追求的目的是少失真、多接近, 当然是在一定代价意义下.

因为有限位运算会带来失真, 因此原来数学上的一些性质、结论, 在有限位运算下, 就会有修正, 下面看些例子.

例1 数学上知道, $f(x)$ 可导时, 有 $\lim_{h\to 0}\big(f(x+h)-f(x)\big)/h = f'(x)$, 也即当 h 为充分小的数时, $\big(f(x+h)-f(x)\big)/h$ 很接近 $f'(x)$. 现在考察 $f(x)=\mathrm{e}^x$, 取 $x=1$ 的情况, 则自然成立

$$f'(1) = \lim_{h\to 0}\frac{f(1+h)-f(1)}{h} = \mathrm{e} = 2.71828\cdots.$$

可以证明: $g(h) = \frac{f(1+h)-f(1)}{h}$ 是 h 的单调上升函数, 也即 h 越小, $g(h)$ 越接近 e. 可是取 16 位数字在计算机上计算的结果是: 当 $h = 10^{-8}$ 时较好, 当 h 更小时, 越来越差, 见表 1.1.[①] 表 1.1 中 `-1.95e+000` 表示 $-1.95\times 10^{+000}$, 下同. 细节请参阅第 §1.3.2 节浮点数.

例2 考察下面 3 个式子

$$A = \frac{1-\cos x}{x^2}, \qquad B = \frac{[\sin x/x]^2}{1+\cos x}, \qquad C = 2\left[\frac{\sin(x/2)}{x}\right]^2.$$

数学上容易验证, 它们是恒等的, 也即对不同的 x 都相同. 但在数值运算下就不完全相同. 在 Matlab 中的运算结果见表 1.2. 表中数字下方画线的表示不正确.

从这个例子可以看出, 对于一个数学公式, 不同的表达形式 (在计算中代表不同的算法), 导致的计算效果是不一样的. 由此可以想到, 对于一个数学问题可能有很多种数值方法, 但是采用各种数值方法的效果可能不完全一样. 对于一个数学问题, 追求好的数值方法, 也是数值分析的基本任务.

① 这是一个标准的结论: 用差分方法逼近导数时, 最精确的值在 $h\approx\sqrt{\varepsilon}$ 时取得, 其中 ε 是机器精度, 也就是在计算机上满足 $1+\varepsilon = 1$ 的最大的 ε. 在 Matlab 中 $\varepsilon \approx$`2.22e-16`.

表 1.1

h	$g(h)$	e	误差
1e+000	4.67077427047160	2.71828182845905	-1.95e+000
1e-001	2.85884195487388	2.71828182845905	-1.41e-001
1e-002	2.73191865578708	2.71828182845905	-1.36e-002
1e-003	2.71964142253278	2.71828182845905	-1.36e-003
1e-004	2.71841774707848	2.71828182845905	-1.36e-004
1e-005	2.71829541991231	2.71828182845905	-1.36e-005
1e-006	2.71828318698653	2.71828182845905	-1.36e-006
1e-007	2.71828196396484	2.71828182845905	-1.36e-007
1e-008	2.71828177744737	2.71828182845905	5.10e-008
1e-009	2.71828159981169	2.71828182845905	2.29e-007
1e-010	2.71827893527643	2.71828182845905	2.89e-006
1e-011	2.71827005349223	2.71828182845905	1.18e-005
1e-012	2.71827005349223	2.71828182845905	1.18e-005
1e-013	2.71338507218388	2.71828182845905	4.90e-003
1e-014	2.66453525910038	2.71828182845905	5.37e-002
1e-015	2.66453525910038	2.71828182845905	5.37e-002

§1.2 误差和有效数字

§1.2.1 绝对误差与相对误差

在有限位运算下, 理想的正确数值是比较少的, 参与运算的往往是正确值的近似值. 例如

$$A = \sqrt{3} = 1.732050807568877293527\cdots$$

是一个无限位的数. 取成 5 位小数的近似值 $A^* = 1.73205$, 取成 4 位小数的近似值是 $A^* = 1.7321$. 这里采用了“四舍五入”规则.

每个正确值 A 跟近似值之 A^* 之间总有关系

$$A = A^* + \eta. \tag{1.1}$$

例如上面的 $\sqrt{3}$, 对 5 位小数的近似值 A^*, 它的 $\eta = 0.0000008075\cdots$, 而对 4 位小数的近似值 A^*, 它的 $\eta = -0.0000491924\cdots$.

定义1 (1.1) 式中的 η 称为 A 取近似值 A^* 时的绝对误差, 简称误差.

绝对误差的正确值, 常常是无限位的, 我们没有必要、通常也没有可能得到它的正确值. 对实际有用的是知道它的绝对值的上界, 这个上界称为绝对误差限, 或称为误差限. 实用上为了方便, 只要不混淆, 常把它说成“绝对误差”.

表 1.2

x	A	B	C
0.20000	0.49833555396896	0.49833555396896	0.49833555396896
0.02000	0.49998333355561	0.49998333355555	0.49998333355555
0.00200	0.49999983334215	0.49999983333336	0.49999983333336
0.00020	0.49999999696126	0.49999999833333	0.49999999833333
0.00002	0.50000004137019	0.49999999998333	0.49999999998333
3.14159	0.20264270961412	0.20264444437844	0.20264270961412
3.16159	0.20006700855095	0.20006700855100	0.20006700855095
3.14359	0.20238474101246	0.20238474100960	0.20238474101246
3.14179	0.20261690881243	0.20261690926896	0.20261690881243

例如 $\sqrt{3}$ 的 5 位小数近似值, 有 $|\eta| \leqslant 8.1 \times 10^{-7}$, 绝对误差限即为 8.1×10^{-7}, 而对 4 位小数的近似值, 有 $|\eta| \leqslant 5 \times 10^{-5}$, 它的绝对误差限即为 5×10^{-5}.

容易知道, 如果 A 的近似值 A^* 是由 A 按"四舍五入"规则得来的, 那么必有

$$|A - A^*| \leqslant \frac{1}{2}\alpha,$$

这里 α 是 A^* 的最后一位的单位.

定义2 *一个数通过"四舍五入"或"截断"产生的近似数与它本身的误差称为舍入误差.*

上面提到的 $\eta = 0.0000008075\cdots$ 和 $\eta = -0.0000491924\cdots$ 就是舍入误差.

衡量一个近似值的精确程度, 光有绝对误差是不够的. 譬如测量一段路程, 其长度为 1000km, 知道有误差 20m; 另外测量一条 400m 的跑道, 也有 20m 的误差. 显然后者的精度差多了. 这时用相对误差的概念, 就清楚了.

定义3 *在 (1.1) 式中, 记*

$$r = \eta/A,$$

称为相对误差. 如果 R 满足 $|r| \leqslant R$ 则称 R 为相对误差限.

由此知道测量 1000km 的路程有 20m 误差, 它的相对误差为 $20/10^6 = 2\times10^{-5}$, 而测量 400m 的跑道有 20m 误差的, 它的相对误差为 $20/400 = 5\%$, 可见比前者大多了.

在数值计算中, 正确值常常不知道, 而常将相对误差改成

$$r = \eta/A^*.$$

这是因为

$$\begin{aligned}\frac{\eta}{A} &= \frac{\eta}{A^*+\eta} = \frac{\eta}{A^*}\left(1-\frac{\eta}{A^*}+\left(\frac{\eta}{A^*}\right)^2+\cdots\right)\\ &= \frac{\eta}{A^*}-\left(\frac{\eta}{A^*}\right)^2+\left(\frac{\eta}{A^*}\right)^3-\cdots\end{aligned}$$

因此当 η/A^* 较小时, 两者差别很小.

§1.2.2 有效数字与可靠数字

$$A=\sqrt{3}=1.732050807568877\cdots$$

取近似值为 1.73, 我们说它有 3 位有效数字, 取近似值为 1.732, 它有 4 位有效数字, 1.7320 也只有 4 位有效数字, 而 1.7321 则有 5 位有效数字.

定义4 一个近似值, 在十进制数字表示中, 其误差小于某位单位的一半, 这位数字就称为有效数字. 用数学语言来说, 即若 A 的近似值

$$A^*=\pm(x_1 10^{-1}+x_2 10^{-2}+\cdots+x_k 10^{-k}+\cdots+x_n 10^{-n})\times 10^m, \qquad (1.2)$$

其中 m 是整数, k 是不超过 n 的正整数, $x_i,(i=1,2,\cdots,n)$ 是 0 到 9 的某一整数, 并且 $x_1\neq 0$, 这样 x_k 这一位的单位即为 10^{m-k}, 于是如果

$$|A-A^*|\leqslant\frac{1}{2}\times 10^{m-k},$$

则称 x_k 为有效数字. 显然, 如果 x_k 是有效数字, 那么 $x_1,x_2,\cdots,x_{k-1}$ 都是有效数字. 如果 A^* 的每一位都是有效数字, 那么 A^* 称为有效数.

从定义可知, 若 A^* 是从 A 经“四舍五入”而来的, 那它必定是有效数.

有效数字与相对误差之间有密切关系, 有下述定理.

定理1 A 的近似值 A^* 表示成 (1.2) 式, 如果已知 x_k 是有效数字, 那么相对误差限不超过 $1/2\times 10^{-(k-1)}$; 反之如果已知相对误差 r, 且有 $|r|\leqslant 1/2\times 10^{-k}$, 那么 x_k 必为有效数字.

证明 因为 x_k 是有效数字, 于是

$$|A-A^*|\leqslant\frac{1}{2}\times 10^{m-k},$$

而对相对误差 r, 有

$$|r|=\frac{|A-A^*|}{|A^*|}\leqslant\frac{1}{2}\times 10^{m-k}\Big/|A^*|.$$

但 $|A^*| \geqslant x_1 \times 10^{-1} \times 10^m \geqslant 10^{m-1}$, 于是

$$|r| \leqslant \frac{1}{2} \times 10^{m-k} \Big/ 10^{m-1} = \frac{1}{2} \times 10^{-(k-1)}.$$

又若已知 $|r| \leqslant 1/2 \times 10^{-k}$, 那么

$$|A - A^*| \leqslant |r| \cdot |A^*| \leqslant \frac{1}{2} \times 10^{-k}|A^*|.$$

但 $|A^*| \leqslant 10 \times 10^{-1} \times 10^m = 10^m$, 故

$$|A - A^*| \leqslant \frac{1}{2} \times 10^{m-k}.$$

于是 x_k 是有效数字. □

定理2 如果 A^* 最多只有 k 位有效数字, 即 x_k 是有效数字, 而 x_{k+1} 不是有效数字的必要条件是 A^* 的相对误差的绝对值必大于 $1/2 \times 10^{-k}$, 充分条件是 A^* 的相对误差的绝对值大于 $1/2 \times 10^{-k}$.

证明 先证必要条件. 因 x_{k+1} 不是有效数字, 所以

$$|A - A^*| \geqslant \frac{1}{2} \times 10^{m-k-1},$$

$$|r| = \frac{|A - A^*|}{|A^*|} \geqslant \frac{1}{2} \times \frac{10^{m-k-1}}{|A^*|} > \frac{1}{2} \times \frac{10^{m-k-1}}{10^m} = \frac{1}{2} \times 10^{-k-1}.$$

再证充分条件. 当 A^* 的相对误差 $|r| > 1/2 \times 10^{-k}$ 时, 如果 x_{k+1} 是有效数字, 则根据定理 1, 相对误差绝的对值不超过 $1/2 \times 10^{-k}$, 于是有矛盾, 从而可知 x_{k+1} 不是有效数字. □

推论1 若 A^* 的相对误差的绝对值大于 1/2, 则就没有有效数字; 反过来, 若 A^* 没有有效数字, 则它的相对误差绝对值必大于 0.05.

有效数字的概念, 是由“四舍五入”规则得出的, 而对应于“截断”规则, 有可靠数字的概念.

定义5 在 (1.2) 式中, 如果

$$|A - A^*| \leqslant 10^{m-k},$$

则称 x_k 为可靠数字.

如果 A^* 是由 A 按“截断”规则而得到, 那么有

$$|A - A^*| \leqslant \alpha,$$

这里 α 是 A^* 的最后一位的单位.

可靠数字与相对误差之间也有密切的关系, 可以建立类似定理 1、定理 2 的定理. 我们将此作为练习题.

§1.2.3 误差的来源

要计算的数学问题, 常常是从自然现象或社会现象中归纳出来的, 是反映自然界和社会中某些规律的数量关系. 譬如: 已造好的一座高楼大厦, 如何使它在风吹雨打及可能发生的地震等条件下都不会倒下, 且造价又比较低? 这就要在许多条件和要求下将它归结成数学问题进行计算; 又如经济学中的“投入与产出”的问题, 也可归结成数学问题进行计算; 再如, 保险公司推出一种保险, 怎样定它的保险率, 既使得顾客愿意, 又使得保险公司有利可图? 这也可归结为数学问题. 由自然现象、社会现象归结出来的数学问题成为数学模型, 但数学模型所描写的东西, 跟实际的自然现象和社会现象是有差别的, 这种差别称为模型误差, 它指导我们在进行计算时控制的精度. 也即如果模型误差是 1% 的数量级, 则我们在计算时, 能保证 1% 的精度就很好了, 没有必要计算得更精确.

模型误差有多大, 是各门科学研究的课题, 不属本书范围. 我们的任务是研究: 在对一个数学问题进行计算时, 最后获得的数值结果与原来数学问题的解答之间差别有多大? 这是数值分析的重要和困难任务, 需要具体问题具体分析. 我们这里先了解一下数学问题计算时的误差来源. 它有如下 3 种来源.

1. 数值方法的截断误差.

在对数学问题进行计算时, 就要选用数值方法, 就会有截断误差. 譬如我们已知 $f(x)$ 的一些值要计算它的导数 $f'(x)$, 选用的计算公式是

$$\frac{f(x+h) - f(x)}{h},$$

但

$$\frac{f(x+h) - f(x)}{h} = f'(x) + \frac{f''(\tilde{x})}{2!}h, \qquad x \leqslant \tilde{x} \leqslant x + h,$$

$hf''(\tilde{x})/2$ 就是截断误差. 因为所用的数值方法把这一项去掉了, 所以叫截断误差.

2. 运算传播误差.

初始数据的误差对计算结果的传播; 每步产生的舍入误差对计算结果的传播, 这部分误差统称为运算传播误差.

例如求: $A\cdot B/C$, 而 A,B,C 各由它们的近似值 A^*,B^*,C^* 参加运算, 它们有初始数据误差 η_A,η_B,η_C, 即

$$A=A^*+\eta_A,\qquad B=B^*+\eta_B,\qquad C=C^*+\eta_C.$$

$A^*\cdot B^*$ 得到的结果有舍入误差 δ_1, 即 $A^*\cdot B^*$ 的计算结果为 $A^*B^*-\delta_1$, 再用 C^* 除之, 又有舍入误差 δ_2, 即得到的数值结果为

$$\frac{A^*\cdot B^*-\delta_1}{C^*}-\delta_2. \tag{1.3}$$

这里告诉我们计算结果既不是 $A\cdot B/C$, 也不是 $A^*\cdot B^*/C^*$, 而是由 (1.3) 式所表示的, 也即

$$\frac{A^*\cdot B^*}{C^*}-\frac{\delta_1}{C^*}-\delta_2.$$

再来看 η_A,η_B,η_C 是如何传播的. 实际上

$$\begin{aligned}\frac{A^*\cdot B^*}{C^*} &= \frac{(A-\eta_A)(B-\eta_B)}{C-\eta_C}\\ &= \frac{AB-A\eta_B-B\eta_A+\eta_A\eta_B}{C(1-\eta_C/C)}\\ &= \frac{AB}{C}\left(1+\frac{\eta_C}{C}+\left(\frac{\eta_C}{C}\right)^2+\cdots\right)-\frac{A\eta_B+B\eta_A-\eta_A\eta_B}{C-\eta_C}\\ &= \frac{AB}{C}-\frac{A\eta_B+B\eta_A-\eta_A\eta_B}{C-\eta_C}+\frac{AB}{C}\frac{\eta_C}{C}+\frac{AB}{C}\left(\frac{\eta_C}{C}\right)^2+\cdots,\end{aligned}$$

最后的传播运算误差是

$$\begin{aligned}E=&-\frac{\delta_1}{C^*}-\delta_2-\frac{A\eta_B+B\eta_A-\eta_A\eta_B}{C-\eta_C}\\ &+\frac{AB}{C}\frac{\eta_C}{C}+\frac{AB}{C}\left(\frac{\eta_C}{C}\right)^2+\cdots,\end{aligned}$$

计算结果是:

$$\frac{AB}{C}+E.$$

这种把要计算的对象 $A\cdot B/C$ 与计算结果 $A\cdot B/C+E$ 之差的 E 或 E 的上界求出来的过程, 称为向前误差分析, 这是一种运算传播误差的表示方式.

另一种表示方式是将计算结果表示成原始数据的扰动. 对于 $A\cdot B/C$ 来说, 即将计算结果表示成

$$(A+\delta_A)(B+\delta_B)/(C+\delta_C),$$

而把 $\delta_A,\delta_B,\delta_C$ 的绝对值上界求出来, 这种过程称为向后误差分析. 对于 $A\cdot B/C$ 这个例子来说, $A^*\cdot B^*$ 的结果可以表示成 $(A^*+\varepsilon_1)(B^*+\varepsilon_2)$, 容易估计出 $\varepsilon_1,\varepsilon_2$ 的