第一讲 Part2：数值计算的基本特性

1. 逼近：数值计算的哲学

近似的哲学：

无限 → 有限，　复杂 → 简单

计算机只能做有限位的加减乘除，循环往复……

无限维 ——→ 有限维　(PDE → ODE)

无限步 ——→ 有限步　（积分 → 求和，导数 → 差商）

无限位 ——→ 有限位　（舍入）

非线性 ——→ 代数、线性

2. 误差：数值计算的主角

A. 误差来源：

a. 算前误差：　模型误差、数据误差

b. 计算误差：
　　截断误差　或　离散误差
　　舍入误差

例：地球表面积　　$S = 4\pi R^2$

模型误差：球形近似
数据误差：半径R的不确定性
截断误差：圆周率取3.14，R取有限小数
舍入误差：计算过程中取若干位有效数字

2. 误差：数值计算的主角

B. 误差衡量：

a. 绝对误差：　$\varepsilon_a = |x - x_h|$

b. 相对误差：　$\varepsilon_r = \varepsilon_a / |x|$

C. 前向与后向误差：

$$\begin{tikzcd}
x \arrow[rr, leftrightarrow] \arrow[d, leftrightarrow, red, "\text{(Backward Error)}"', "e_b"] \arrow[drr] & & f(x) \arrow[d, leftrightarrow, red, "e_f"', "\text{(Forward Error)}"] \\
\hat{x} = x + \delta x \arrow[rr, leftrightarrow] & & \hat{f}(x) = f(\hat{x})
\end{tikzcd}$$

前向误差：　$e_f = |\hat{f}(x) - f(x)|$

后向误差：　$e_b = |\hat{x} - x|, \quad \hat{f}(x) = f(\hat{x})$

3. 计算机算术: I

A. 科学计数法:

$$12345 = 1.2345 \times 10^4, \quad 0.00123 = 1.23 \times 10^{-3}$$

B. 有效数字（Significant digits）：

设

$x_h = \pm 10^k \times 0.a_1 a_2 \cdots a_n$ 为 x 的一个近似，$a_i \in \{0,1,\cdots,9\}$

β

$a_1 \neq 0$，如

$$|x - x_h| \le \frac{1}{2} \times 10^{k-p}$$

（β进制下则为 $\frac{1}{2} \times \beta^{k-p}$）

称用 x_h 近似 x 有 p 位有效数字.

直观上如 x（或 x_h）可由 x_h（或 x）的第 $p+1$ 位（从第一个非0的算起）由四舍五入得到相同的前 p 位，则前 p 位为有效数字.

例子： $x = 0.121, \quad x_h = 0.1205$

$$|x - x_h| = 0.0005 = \frac{1}{2} \times 10^{-3}, \quad k = 0 \Rightarrow p = 3$$

3. 计算机算术: II

C. 浮点数表示:

计算机最常用的是浮点运算. 浮点数由三部份组成

a. 进制 : 基数 (base) β

b. 字长 精度 (precision) t

c. 指数 指数 (exponent) $E \in [L, U]$ 整数

$\forall x$:

$$x = \pm \left(d_0 + \frac{d_1}{\beta} + \frac{d_2}{\beta^2} + \cdots + \frac{d_{t-1}}{\beta^{t-1}} \right) \cdot \beta^E$$

3. 计算机算术: III

C. 浮点数表示:

$\forall x$:

$$x = \pm\left(d_0 + \frac{d_1}{\beta} + \frac{d_2}{\beta^2} + \cdots + \frac{d_{t-1}}{\beta^{t-1}}\right) \cdot \beta^E$$

$0 \le d_i \le \beta-1, \quad i=1,\cdots,t-1; \quad 0 < d_0 \le \beta-1 \qquad L \le E \le U$

$d_0 d_1 \cdots d_{t-1}$ 称尾数(mantissa)或有效数(significant)

$\pm$: 符号(sign); $\qquad E$: 指数

在计算机上, 通常 $\beta=2$ (HP计算器中 $\beta=10$) [规范化仅指 $d_0 \neq 0$]

此时 $d_0=1$, 故存储时通常省略, 称为规范化(normalized) 否则称 denormalized

$d_1 d_2 \cdots d_{t-1}$ 称为有效位中的分数部分(fraction)

3. 计算机算术: IV

C. 浮点数表示:

IEEE 754 标准:

	β	t (字长)	L	U	
单精度(float)	2	24	-126	127	4字节
双精度(double)	2	53	-1022	1023	8字节
(long double)	2	64	-16382	16383	

3. 计算机算术: V

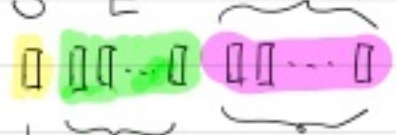
C. 浮点数表示:

t-1位 (M)

S E

float

单精度:



1bit 8bit 23bit 共32位 = 4 bytes

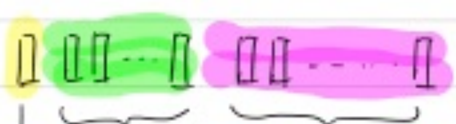
Sign Exponent Fraction

S = 0 正号; S = 1 负号 (Sign)

E为无符号型, 共8比特, 在0, 1, ⋯, 255间 (exponent)

double

双精度:



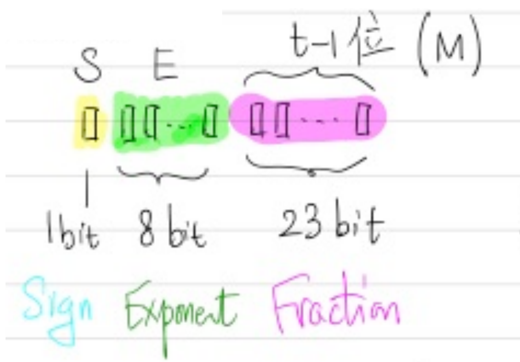
共64位 = 8 bytes

1bit 11bit 52bit

3. 计算机算术: VI

C. 浮点数表示:

单精度



定义 偏移(bias) = 127.

分三种情形:

a. E不全为0或不全为1. 实际指数 = E−127. 使得指数范围从 L=−126 到 U=127.

此时浮点数表示 $\pm \underset{\substack{\uparrow \\ d_0}}{1}.M \times 2^{E-127}$

b. E全为0. 实际指数 = 1−127 = −126

但浮点数表示 $\pm 0.M \times 2^{-126}$, 可表示更小的数.

c. E全为1.

如M全为0, 则表示 ±无穷大, (Inf)

如M不全为0, 则表示 NaN (Not a Number)

(否则比无穷大还大(或比负无穷还更小))

3. 计算机算术: VII

D. 浮点数基本性质:

a. Underflow: $UFL = \beta^{L}$ (对规范化表示,即 $d_0 \neq 0$)

或者 $UFL = \beta^{L-23}$ (SP)

$UFL = \beta^{L-52}$ (DP)

最小正数 (IEEE754表示)

此处L = −126 (SP) 或 −1022 (DP)

b. Overflow: $OFL = \beta^{U+1}(1-\beta^{-t})$

[$d_0, d_1 \cdots d_{t-1}$ 中全为1]

c. Rounding: $fl(x)$ (Rounding to nearest)

一般实数 ⟶ 浮点数

单次浮点转换带来的舍入误差大小是否可以刻划?

3. 计算机算术: VIII

D. 浮点数基本性质:

d. 机器精度 (machine precision)

$$\varepsilon_{mach} = \frac{1}{2}\beta^{1-t} \qquad (\text{"四舍五入"法})$$

(或 β^{1-t}, 如采用直接截断法)

$\varepsilon_{mach} \neq UFL$, 一般 $0 < UFL < \varepsilon_{mach} < OFL$

ε_{mach} 表征的是浮点系统的最大相对误差界, 也称 unit round-off. (也记为 u)

3. 计算机算术: IX

机器精度的直观（unit round-off）：

我们有

[即如 $x=1$, 使 $x+\varepsilon>1$ 的最小的 ε（浮点下），则 $\varepsilon=\beta^{1-t}$（直接截断）.]

$$fl(x)=x(1+\delta) \qquad |\delta|\le\varepsilon_{mach}$$

即 $|fl(x)-x|/|x| \le \varepsilon_{mach}$.

设 $x=d_0.d_1d_2\cdots d_{t-1}d_td_{t+1}\cdots\times\beta^k$, $\quad d_0\geq 1$

$fl(x)=\tilde{d}_0.\tilde{d}_1\cdots\tilde{d}_{t-1}\times\beta^k$ "四舍五入", t位有效数字

则

$$\left|\frac{fl(x)-x}{x}\right|\le\frac{1}{2}\beta^{1-t}\times\beta^k/|x|\le\frac{1}{2}\beta^{1-t}=\varepsilon_{mach}$$

$$(|x|\geq\beta^k)$$

3. 计算机算术: X

D. 浮点数的基本性质:

e. 浮点范围:

单精度: $\varepsilon_{mach} \sim 5.96\times 10^{-8}$ 范围 $10^{\pm 38}$

双精度: $\varepsilon_{mach} \sim 1.11\times 10^{-16}$ 范围 $10^{\pm 308}$

f. 运算的封闭性:

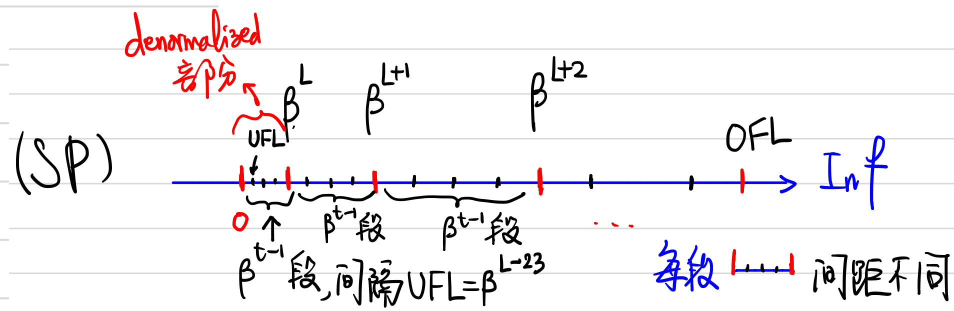
$0/0,\ 0\times\infty,\ \sqrt{-1}$: NaN

Overflow : ±Inf

除0 : ±Inf

能表示的所有非0数如下图:

（正）浮点数的分布



能表示的全体数为离散、非均匀间隔且有上、下界.

4. 舍入误差对计算的影响: I

A. 判断两个数是否相等:

由于舍入误差，两个数是否相等只需在舍入误差范围内

即 $|x-y|/|x| \le \varepsilon_{mach}$ 即可

[由于 $fl(x)=x(1+\delta)$, $\delta \le \varepsilon_{mach}$]

B. 大数吃掉小数:

$10+\underbrace{10^{-5}+\cdots+10^{-5}}_{共10^5个}$ 设 $\beta=10$, $t=6$

$10+10^{-5} \overset{c}{=} 10$ "$\overset{c}{=}$" 在浮点运算下.

但如采用 $10^{-5}+\cdots+10^{-5}+10$ 则 $\overset{c}{=} 1+10=11$ (或进一步二分递归相加)

4. 舍入误差对计算的影响: II

C.　相近数相减会丧失有效位:

a. $x = 1-(1+\varepsilon)^2 \qquad \varepsilon = 0.1$

设 $\beta=10,\ t=2.$

$x_h^1 = 1-1.1^2 \stackrel{c}{=} -0.2$

$x_h^2 = 1-(1+2\varepsilon+\varepsilon^2) = -\varepsilon^2-2\varepsilon$

$= -0.01-0.2 = -10^{-1}(0.1+2) = -2.1\times10^{-1}$

b. $x = 1-\cos 2^\circ = 1-\cos \pi/90 \qquad t=5$

$x_h^1 \stackrel{c}{=} 1-0.99939 = 6.1\times10^{-4}$

$x_h^2 = 2\sin^2\frac{\pi}{180} \stackrel{c}{=} 6.0917\times10^{-4}$

是否可以有某种理论解释?

4. 舍入误差对计算的影响: III

D. 简例:

a、一元二次方程求根:

$$ax^2+bx+c=0$$

$$x=\frac{-b\pm\sqrt{b^2-4ac}}{2a} \qquad a\neq 0$$

Case1: 如 $b^2 >> 4ac \Rightarrow \sqrt{b^2-4ac} \approx |b|$

例 $b=-(10^5+1)$, $c=1$, $a=1$; $\beta=10$, $t=4$

$$\sqrt{b^2-4ac} \overset{c}{=} |b| \overset{c}{=} 10^5$$

$$\Rightarrow x_1^h = 10^5,\quad x_2^h = 0$$

但精确解 $x_1=10^5$, $x_2=1$; x_2 误差很大.

是否可以有某种方法减小误差?

4. 舍入误差对计算的影响: IV

D. 简例:

Strategy:

先计算 $x_1^h = \dfrac{-b - \mathrm{sgn}(b)\sqrt{b^2-4ac}}{2a}$

再由韦达定理:

$$x_2^h = \frac{c}{a} / x_1^h$$

Case 2:

$a = 10^{20}$, $b = -3 \cdot 10^{20}$, $c = 2 \cdot 10^{20}$ $(x_1 = 1, x_2 = 2)$

直接用求根公式会上溢 $(\mathrm{OFL} = 10^{38}, \mathrm{SP})$

是否可以有某种方法避免困难?

此时需对方程两边同除 10^{20} 再计算.

4. 舍入误差对计算的影响: V

D. 简例:

Case 3:

$a=10^{-20},\ b=-3,\ c=2\cdot 10^{20}$

根 $x_1=10^{20},\ x_2=2\cdot 10^{20}$

直接用求根公式是否可行?

一个更好的策略:

可通过 $x=y\cdot 10^{20}$ 变为

$10^{20}y^2-3\cdot 10^{20}y+2\cdot 10^{20}=0$

再作除 10^{20} 变为一个好的方程.

无量纲化!

4. 舍入误差对计算的影响: VI

E. 计算应讲求效率:

$$P_n(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_0$$

直接计算 $\frac{n(n+1)}{2}$ 乘法.

有更好的计算策略么？

秦九韶算法: $P_n(x) = (\cdots(a_n x + a_{n-1})x + a_{n-2}) + \cdots)x + a_0$

n次乘法, n次加法. (Horner算法)

5. 其他相关

A. 软件资源：

Numerical Recipes, Matlab, Mathematica, www.netlib.org, ⋯

B. 杂志

计算数学类： Math. Comp., SINUM, SISC, Numer. Math., ⋯

计算科学类： J. Comp. Phys., J. Chem. Phys., J. Chem. Theory Comp., PLoS CB, Phys. Rev.系列, ⋯

综合科学类： Nature, Science, PNAS, PRL, PRX, NC, ⋯

课程要求

课程要求

课程网页: http://www.math.pku.edu.cn/teachers/litj
Courses: Numerical Analysis

考核要求:

手写作业: 15%

上机作业: 15%

期末考试: 70%

课程梗概

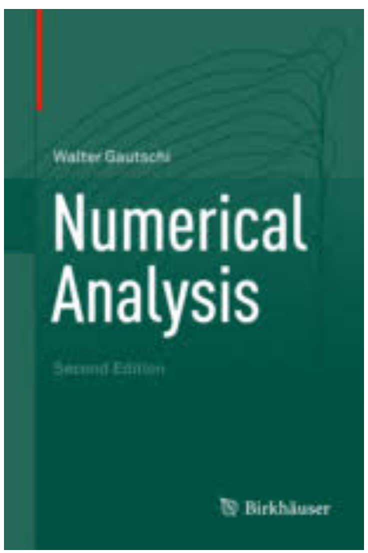
Lect01	Introduction
Lect02	Lagrange Interpolation and Newton's Formula
Lect03	Cubic and B-Splines
Lect04	Least Squares Approximation
Lect05	Uniform Approximation
Lect06	Special Topics: ENO, Sparse Grid, Neural Network
Lect07	Numerical Quadrature
Lect08	Error Analysis, Gaussian Quadrature
Lect09	Special Topics: Spectral Accuracy, Adaptivity
Lect10	Multidimensional quadrature, Numerical Differentiation
Lect11	Monte Carlo Integration: Basics
Lect12	Metropolis Algorithm
Lect13	Numerical Solution of Nonlinear Equation: Scalar
Lect14	Numerical Solution of Nonlinear Equations: System
Lect15	Fast Fourier Transform: Basics
Lect16	Fast Fourier Transform: Applications
Lect17	Fast Gauss Transform
Lect18	Numerical ODEs: Background and Basics
Lect19	Numerical Stability and Convergence Theory
Lect20	Runge-Kutta Methods
Lect21	Stiff and Multiscale ODEs
Lect22	Symplectic Methods and Boundary Value Problems
Lect23	Introduction to Numerical Software Development

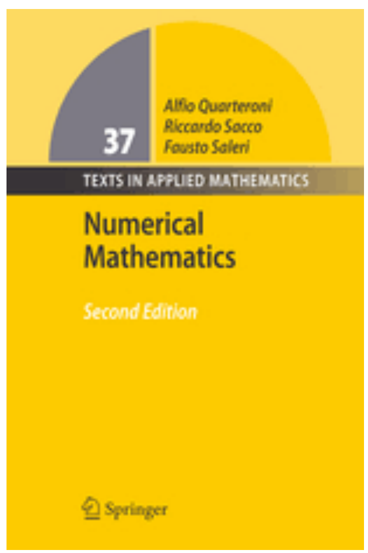
参考书

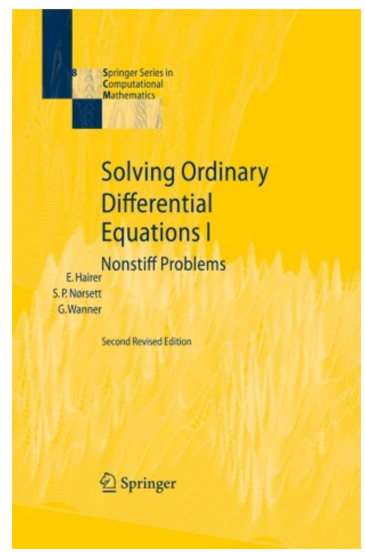
1. W. Gautschi, Numerical Analysis, Birkhauser, 2nd Ed., 2012.

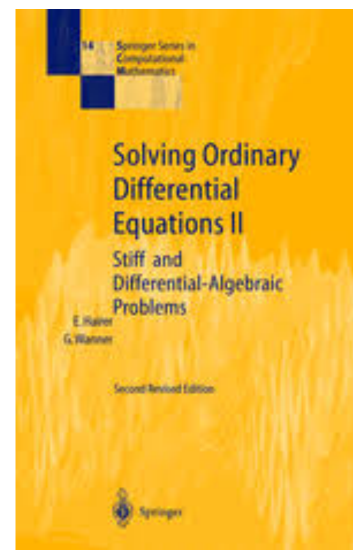
2. A. Quateroni, R. Sacco, and F. Saleri, Numerical Mathematics, Springer, 2000.

3. E. Hairer and G. Wanner, Solving ODEs: I & II, Springer.





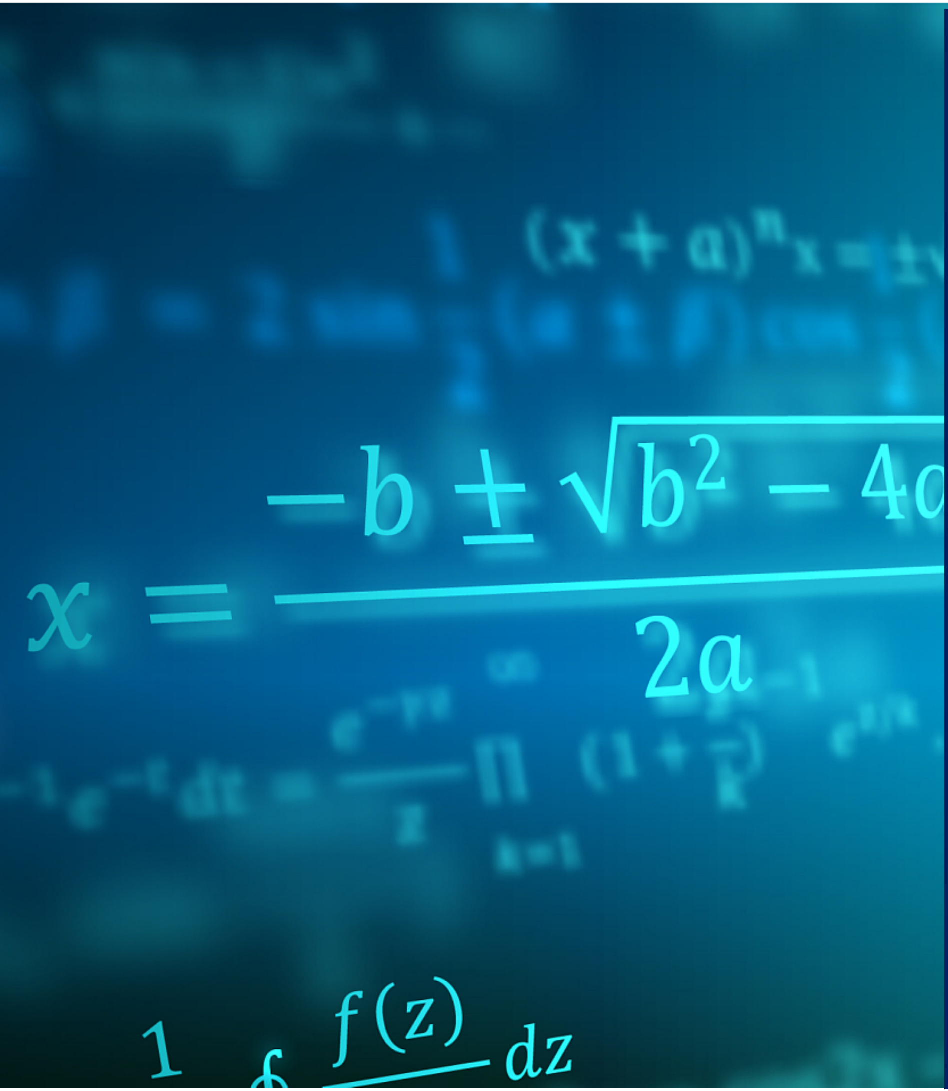




编程作业

C/C++/Python/Baltamatica: 建议大作业使用C/C++/Python;
简单编程采用Baltamatica

Baltamatica下载网址： http://www.baltamatica.com/download.html



北太天元数值计算通用软件

Baltamatica

北太天元数值计算通用软件（以下简称“北太天元”）是面向科学计算与工程计算的国产通用型科学计算软件。本软件具有自主知识产权，提供科学计算、可视化、交互式程序设计，具备丰富的底层数学函数库，支持数值计算、数据分析、数据可视化、数据优化、算法开发等工作，并通过SDK与API接口，扩展支持各类学科与行业场景，为各领域科学家与工程师提供优质、可靠的科学计算环境。

北太天元能力

研发情况

兼容 .mat 脚本文件

数学函数

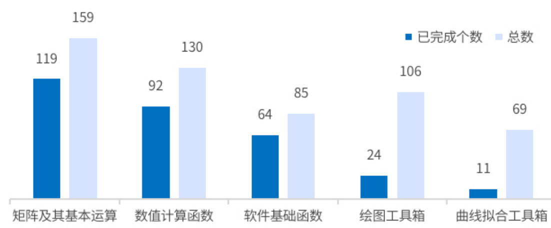
300+
常用数学函数

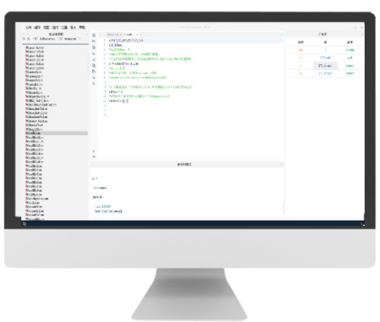
72%
国外软件常用数学函数替换率

功能函数

1300+
常用功能函数

34%
国外软件常用功能函数替换率





Baltamatica Numerical Computation Software
北太天元数值计算通用软件

数值积分
线性代数
微分方程
初等数学
优化
快速傅里叶变换
数据文件读写
插值
曲线拟合
银河麒麟
龙芯LoongArch64
飞腾
可视化呈现
符号计算
算法开发
随机数生成
数据分析
统信UOS
deepin
三维绘图函数
二维绘图函数

www.baltamatica.com

北太天元官网

www.baltamatica.com/download.html

软件下载