関数の精度保証付き計算法

2006MI039 稲垣康之
指導教員：杉浦洋

1 はじめに

この研究の目的は１変数関数の精度保証付き計算法を開発することである．精度保証付き計算とは，関数値を含む十分に狭い区間を計算することである．関数値の計算において大きく分けて二つの誤差が発生する。すなわち理論誤差と丸め誤差である．

理論誤差は，関数や方程式がその近似に置き換えられるときに発生する．一方、丸め誤差は近似式を計算する際の近似四則演算から発生する．

本研究では，多項式による関数近似を想定し，多項式をホーナー法で計算する際の丸め誤差を厳格に評価するアルゴリズムについて研究する．

2 IEEE754

IEEE754 は，機械実数と機械演算に関する規格であり，ほとんどの計算機がこれに従って設計されている．機械実数には使用ビット数により，単精度，拡張単精度，倍制度，拡張倍制度の規格がある．ここでは数値実験で用いる倍精度の規格について述べる．

2.1 機械実数

倍精度機械実数は，正規数，副正規数，０，$\pm\infty$，NaN からなる．通常の計算では，正規数と０が使われるので，

$$\mathbb{F} = \{\ \text{正規数全体}\ \} \cup \{0\} \tag{1}$$

を改めて機械実数と呼ぶことにする．

正規数は次の形の２進浮動小数点数である．

$$a = \pm 2^e (1.d_1 d_2 \cdots\cdots d_{52})_2. \tag{2}$$

ここで，$-1022 \le e \le 1023, d_i \in \{0,1\}$ であり，$(\quad)_2$ は２進少数を表す．

2.2 機械演算

機械演算は $\mathbb{F}$ 上の加減乗除と平方根の５種である．機械演算は正確に計算した後，$\mathbb{F}$ 上に丸めたものと同じ結果になるように定められている．$\bigcirc : \mathbb{R} \longrightarrow \mathbb{F}$ を丸め演算子，$\cdot \in \{+,-,\times,/\}$ を四則演算子，$\odot$ を対応する機械演算子とすると

$$x \odot y = \bigcirc(x \cdot y) \qquad (x, y \in \mathbb{F}) \tag{3}$$

である．

2.3 丸めモード

丸めには，次の４種類のモードがある．$c \in \mathbb{R}$ とする．

1. 丸め上げ (round upward)：c 以上の最小の浮動小数点数に丸める．これを $\triangle : \mathbb{R} \longrightarrow \mathbb{F}$ と表す．
2. 丸め下げ (round downward)：c 以下の最大の浮動小数点数に丸める．これを $\triangledown : \mathbb{R} \longrightarrow \mathbb{F}$ と表す．
3. 丸め込み (round to nearest)：c に最も近い浮動小数点数に丸める．これを $\square : \mathbb{R} \longrightarrow \mathbb{F}$ と表す．もし、このような浮動小数点が２つある場合には仮数部の最後のビットが０である浮動小数点に丸める．
4. 切捨て (round toward 0)：$|c|$ 以下の浮動小数点数の中で c に最も近いものに丸める．

3 丸め誤差

[定理 1] $x \in \mathbb{R}$ に対して，

$$|\square x - x| \le x(1+\epsilon_1), |\epsilon_1| \le u \tag{4}$$

$$|\square x - x| \le \frac{x}{1+\epsilon_2}, |\epsilon_2| \le u \tag{5}$$

が成立するような ϵ_1, ϵ_2 が存在する．ここで $u = 2^{-53}$ は丸め誤差単位である．//

[定理 2]　$|\delta_i| \le u$ で $nu < 1$ のとき

$$\prod_{i=1}^{n}(1+\delta_i) = 1 + \theta_n \tag{6}$$

とおくと，

$$|\theta_n| \le \gamma_n \equiv \frac{nu}{1-nu} \tag{7}$$

が成り立つ．//

4 ホーナー法

ホーナー法とは，与えられた x に対し，多項式の値

$$p(x) = q_n = \sum_{i=0}^{n} a_i x_i = a_0 + a_1 x + \cdots\cdots\cdots + a_n x^n$$

を計算するアルゴリズムである．

4.1 事前誤差解析

事前誤差解析とは，実際に計算をする前に，計算をしたらどれくらいの誤差が発生するかを推測し，その上限を押さえるもので，荒く誤差を評価する時に有効である．

多項式 $p(x)$ を計算するホーナー法のアルゴリズムは，

$$\begin{cases} q_n = a_n; \\ for(i = n-1; i \ge 0; i--) \\ \quad q_i = x q_{i+1} + a_i; \end{cases} \tag{8}$$

である．求める関数値は $q_0 = p_n(x)$ である．このアルゴリズムを定理 2 を用いて解析すると，補助多項式を

$$\bar{p}(x) = \sum_{i=0}^{n-1} \gamma_{2i+1} |a_i| x^i + \gamma_{2n} |a_n| x^n \tag{9}$$

として，

$$|p(x) - \hat{q}_0| \le \bar{p}(|x|) \tag{10}$$

によって事前誤差評価が得られる．

4.2 事後誤差解析

事後誤差解析とは，計算の途中結果の情報を計算結果を誤差評価に取り入れる方法である．

4.2.1 x と a_k に丸め誤差が無い場合

$$\begin{aligned}\mu_n &= \frac{1}{2}|\hat{q}_n| \\ \mu_i &= |x|\mu_{i+1} + |\hat{q}_i|(i = n-1, n-2, \cdots, 0)\end{aligned} \tag{11}$$

ここで $\pi_0 = 2\mu_0 - |\hat{q}_0|$ として，$|e_0| \leq \pi_0 u$ である．

4.2.2 x と a_k に丸め誤差がある場合

$\tilde{x} \cong x, \tilde{a}_k \cong a_k$ は，真値 x, a_k を丸め込んだものとすると，

$$\tilde{x} = (1+\delta)x, \tilde{a}_k = a_k/(1+\delta_k) \quad (|\delta|, |\delta_k| \leq u) \tag{12}$$

となる．ホーナー法の機械演算は，

$$\begin{aligned}\tilde{q}_k &= \tilde{x} \boxtimes \tilde{q}_{k+1} \boxplus \tilde{a}_k \\ &= \frac{(\tilde{x}\tilde{q}_{k+1}(1+\epsilon_k) + \tilde{a}_k)}{1+\epsilon'_k}.\end{aligned} \tag{13}$$

ここで $1+\epsilon''_k = (1+\delta)(1+\epsilon_k)$ とおくと，定理 2 より $|\epsilon''_k| \leq \gamma_2 = 2u/(1-2u)$．これより，

$$\tilde{q}_k = x\tilde{q}_{k+1} + \tilde{a}_k + \epsilon''_k x\tilde{q}_{k+1} - \epsilon'_k \tilde{q}_k. \tag{14}$$

今，$e_k = \tilde{q}_k - q_k$ と置くと，式 (12) より，

$$e_k = xe_{k+1} + \epsilon''_k x\tilde{q}_{k+1} - \delta_k \tilde{a}_k - \epsilon'_k \tilde{q}_k.$$

ここで，三角不等式により

$$|e_k| \leq |x||e_{k+1}| + \gamma_2|x||\tilde{q}_{k+1}| + u(|\tilde{a}_k| + |\tilde{q}_k|).$$

これより，$\pi_n = |\tilde{a}_n|, \xi = (1+u)|\tilde{x}| \geq |x|$ とし，

$$\pi_i = \xi\pi_{i+1} + \frac{\gamma_2}{u}\xi|\tilde{q}_{k+1}| + |\tilde{a}_k| + |q_k|$$

で数列 $\{\pi_i\}_{i=0}^n$ を定義すると，

$$|e_i| \leq u\pi_i \quad (0 \leq i \leq n). \tag{15}$$

によって誤差評価できる．

5 数値実験

Mathematica 上で，以下の 2 つの実験を行った．

実験 1　n 次の Chebyshev 多項式 $y = T_n(x)$　$(n = 10, 20, 30, 40)$ の値を精度保証した．引数 x にも，丸め誤差の生じない 2 進有限小数 $x = x_i = i/128$　$(0 \leq i \leq 128)$ を用いた．真値 $y_i = T_n(x_i)$ はホーナー法により 10 進 40 桁計算で求めたものを用いた．計算値 $\tilde{y}_i = fl(T_n(x_i))$ はホーナー法により倍精度で求めた．その結果を図 1 に示す．絶対誤差を $e_i = |\tilde{y}_i - y_i|$ とし，点で表す．事前誤差評価，事後誤差評価で求めた e_i の上界をそれぞれ E_i^b, E_i^a と書き，実線，破線で表す．実験 2　指数関数 $y = e^x$ を Taylor 展開により

$$z = \sum_{k=0}^{n} \frac{1}{k!}x^k \cong e^x \tag{16}$$

で近似する．この実験では，式 (16) 左辺をホーナー法により倍精度で計算し，それを精度保証する．引数 x として，$x = x_i = 1/200(0 \leq i \leq 200)$ を用いた．$x = x_i$ における z の値 $z_i = e^{x_i}$ は 10 進 40 桁で求めた．その倍精度による計算値 $\tilde{z}_i$ とし，絶対誤差を $e_i = |\tilde{z}_i - z_i|$ と書く．図 2 に e_i を点，事後誤差解析による上界 E_i を実線で示す．

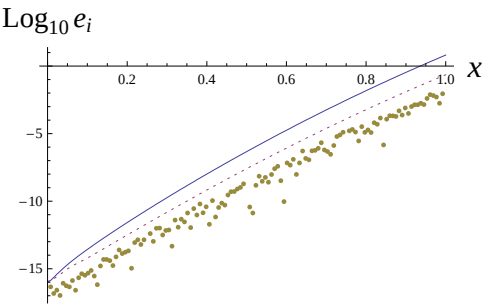

図 1　実験 1　誤差と推定誤差　n=40

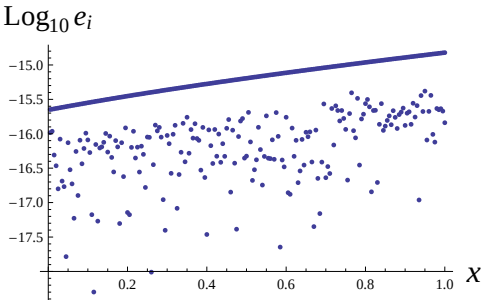

図 2　実験 2　誤差と推定誤差　n=18

実験 1，2 共，求めた上界は e_i より常に大きく精度保証に成功している．また実験 1 では常に $E_i^b \geq E_i^a$ であり，事後誤差評価式の方が事前誤差評価式より精密であると分かる．

6 おわりに

事前誤差解析，事後誤差解析，係数と x に誤差がある場合の 3 つのプログラムを作成したが，その全てにおいて精度保証に成功した．

係数と x に丸め誤差より大きい誤差が混入したときの精度保証は，今後の課題である．

参考文献

[1] 大石進一：『精度保証付き数値計算』．コロナ社，東京，2000

[2] Nicholas J. Higham：”Accuracy and Stability of Numerical Algorithms ”，siam(2002).