数理解析研究所講究録
第310巻 1977年 83-92

ミニコンの基本外部関数の誤差曲線について

群大　工　春海佳三郎
東大　地震研　小竹美子
檜山澄子

0. はじめに

ミニコンは，今では非常に普及し，我々の身近な存在になってきている。それにもかかわらず，その基本外部関数の精度に関してはユーザは勿論，メーカ側も意外と知らないようである。そこで私達は，その精度の検定を試みた。その為に得られた誤差をもとにして誤差曲線を描いてみたところ，各組込関数の作り方がわかり，作製上のプログラムミスが容易に発見されること，又同一機種でもコンパイラによって(コンパイラ作製者によって)非常に精度に差があること等コンパイラの評価に非常に有効なことがわかったので報告する。機種は一番多く利用されている5社のミニコンの代表的コンパイラの単精度，倍精度，複素関数の基本外部関数について検定を行った。尚くわしくは文献1) 2)を参照されたい。

1. 検定の方法

(1) 単精度の場合

a) 対象機種；A, B, C, D, E

b) 検定した関数；SQRT, EXP, SIN, COS, ALOG, TANH, ATAN

c) テスト区間；1/128～200/128で1/128きざみ，また，1～100で1きざみで関数値を計算

d) 誤差の評価；

SIN, COSは上記全区間で絶対誤差

ALOG $\begin{cases} 0.5 \sim \pi/2 \text{ で絶対誤差} \\ \text{その他の区間は相対誤差} \end{cases}$

その他の関数は上記全区間で相対誤差

ただし，検定済みの計算機で計算した倍精度関数の8桁目を4捨5入したものを，真値とみなしている。

その一部をプロットしたものが，図1である．また念のため各機種のビット構成を表1にのせる．

表1 各計算機の内部構成

機種	内部構成	
	指数部(ビット)	仮数部(ビット)
A	S_1+7	S_2+23
B-a	S_1+7	S_2+23
B-b	S_1+7	S_2+23
C	S_1+7	S_2+23
D	S_2, 7	24
E	S_2, 7	24

(2) 倍精度の場合

a) 対象機種；A, B-a, B-b, D （Bにはコンパイラが2種ある．C, Eには，倍精度関数はない）

b) 検定した関数；DSQRT, DEXP, DSIN, DCOS, DLOG, DTANH, DATAN c)テスト区間；単精度の場合に同じ

d)誤差の評価；倍精度のビット構成は各コンパイラによって，まちまちであるため（表2参照），各々のミニコンが保持しているそれぞれの最終桁までを考慮して，その誤差をとった．絶対誤差，相対誤差のとり方は，単精度の場合と同じである．

表2　各計算機の内部構成（倍精度）

機種	内部構成		仮数部最終ビットの10進表示
	指数部（ビット）	仮数部（ビット）	
A-b single	S_1+7	S_2+23	$2^{-23}=1.192\times10^{-7}$
A-b	S_1+7	S_2+39	$2^{-39}=1.819\times10^{-12}$
B-a	S_1+15	S_2+31	$2^{-31}=4.657\times10^{-10}$
B-b	S_1+15	S_2+31	〃
D-b	S_2+7	56	$2^{-56}=1.388\times10^{-17}$

図2がそれぞれの誤差曲線の一部である．

(3)複素関数の場合

a)対象機種；A-a，A-b

b)検定した関数；CSQRT，CEXP，CSIN，CCOS，CLOG

c)テスト区間；$z=x+iy$として，x，yはそれぞれ1/16～25/16まで1/16きざみの計625点

d)誤差の評価；

実部，虚部ごとに相対誤差

誤差の最大値を表3に載せる．

表3　各複素関数の実部および虚部の相対誤差の最大値

［単位は $\times10^{-7}$］

	CSQRT		CEXP		CSIN		CCOS		CLOG	
	R	I	R	I	R	I	R	I	R	I
A-a	9.03	4.99	47.56 (9.3)	16.68 (1.7)	24.93 (2.1)	23.99 (1.7)	22.06 (2.2)	14.58 (1.0)	11.93 (1.2)	17.95 (2.8)
A-b	5.19	2.09	62.77	75.00	76.96 (6.0)	128.53 (22.0)	63.44 (5.8)	120.9 (24.3)	127.6 (23.2)	24.79 (3.9)

（　）内は，関数値が0.1以下の場合の絶対誤差の最大値。A-bのCEXPは，1以下

これら一連の作業を行って得た結論は，文献1，2に詳しく述べたので，ここでは省略することにする。

図 1 各計算機の単精度の誤差曲線（横軸は変数，縦軸は相対誤差（SIN，COS は絶対誤差）で，目盛りの単位は 10^{-7}）

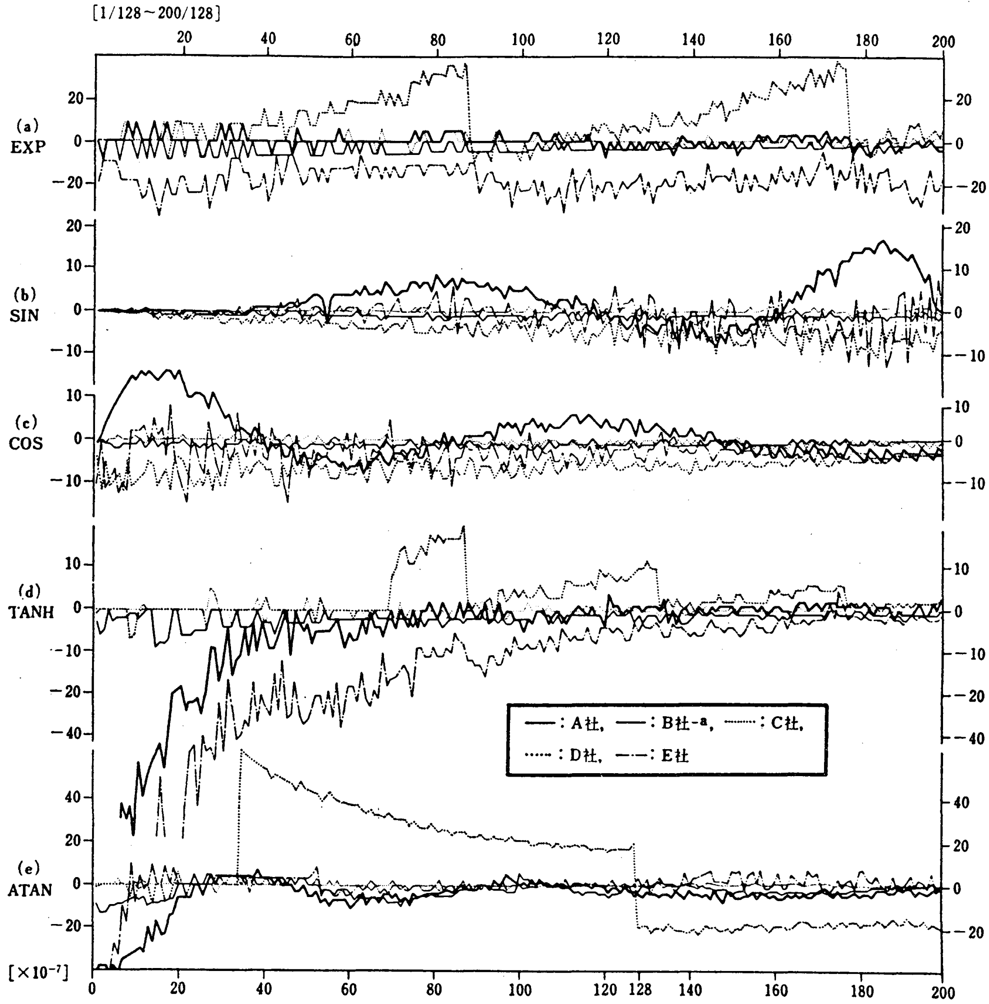
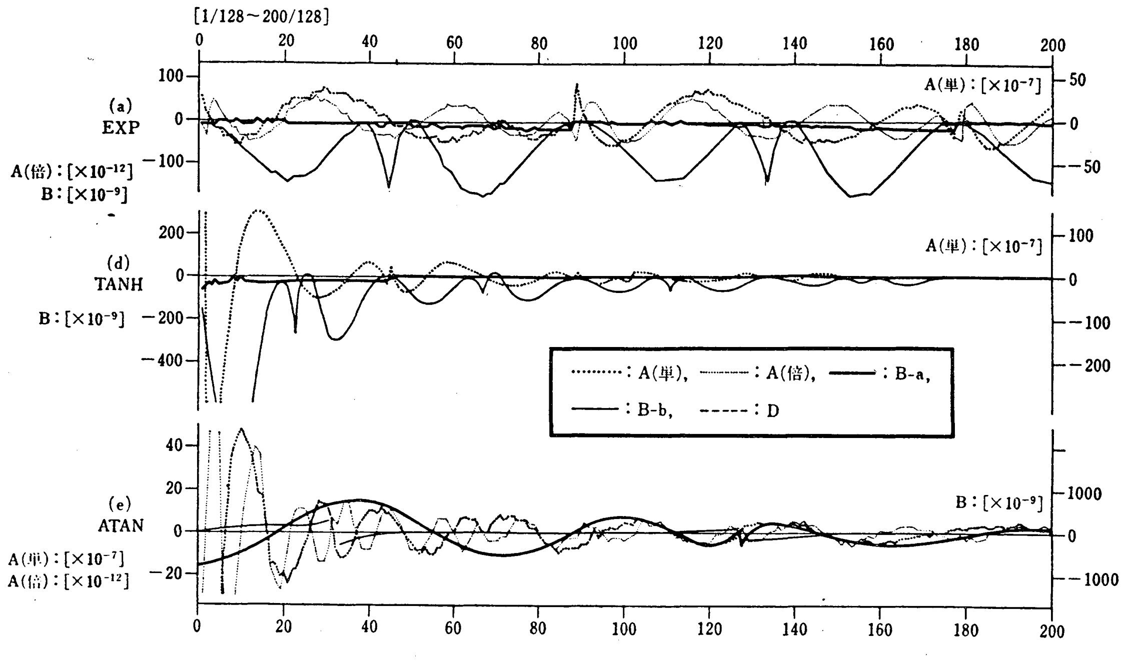


図 2 各計算機の倍精度の誤差曲線（横軸は変数，縦軸は相対誤差で目盛りの単位は $\times 10^{-7}$）



2．誤差曲線の解析

ここで，描かれた誤差曲線をながめて見ると，その中には意外と規則性のあるものがあり，この関数はどんな近似式を用いたか，またどんな手法で作られたかということを推測できるものがあった．以下主なものについて解析を試みた．

(1) 単精度－EXP－Dの場合

図3

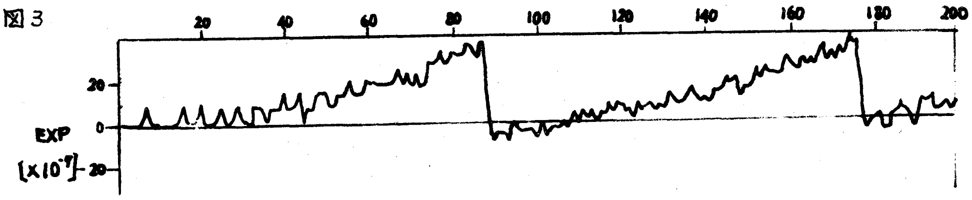


図3の誤差曲線は，88/128で最大になり89/128で0に近い負の値をとり，また177/128で最大になつて，178/128でほぼ0に近い負の値になっている。しかも $0<x\leq 88/128$ と $89/128\leq x<178/128$ では，曲線の形がよく似ている．そこで一般に $e^x=2^{x/\ln 2}$ の関係があることに注目して，$x=88/128$，$89/128$，$177/128$，$178/128$ のときの $x/\ln 2$ を計算してみると，それぞれ0.9918，1.003（≒1），1.995，2.006（≒2）となっている．つまり誤差曲線が急に0に近くなって段のつくところは $x/\ln 2$ が整数値をとるところである．このことから，この誤差曲線は $0<x<\ln 2$，$\ln 2<x<2\ln 2$ の範囲でよく似た形をくり返し，$\ln 2$ を周期にしていることがわかる。しかも2周期目の $89/128\leq x<178/128$ の範囲では

$2^{1+\alpha} = 2\cdot 2^{\alpha}$；$0<\alpha<\ln 2$ なる関係により，$0<x<\ln 2$（$\fallingdotseq 88/128$）で求めたEXP(x)を2倍していることもわかる．

(2) 単精度－SIN，COS－Aの場合

図4

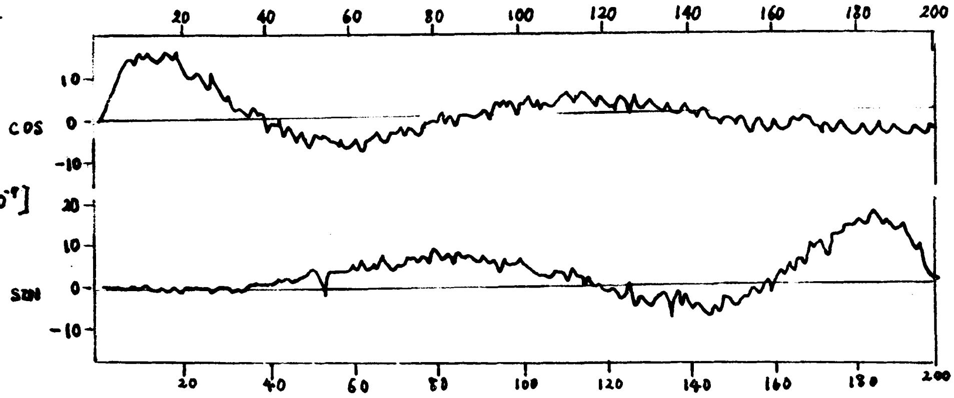
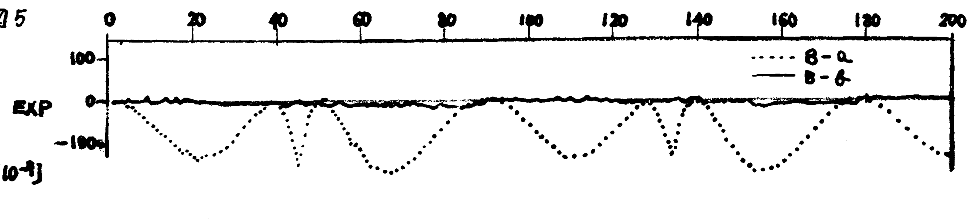


図4のSIN，COSは左右が反転している．このことからCOSの値は $\cos(x)=\sin(\pi/2-x)$ の関係をつかって，$\sin x$ から求められていることがわかる．一方SINの誤差曲線の0点は $x=0.3$，0.85，1.25，$\pi/2$ の近くの計4点である．そして山と谷の部分で $(\sin x)/x$ を計算してみるといずれでも1.2程度になっている．これらのことから，相対型の最良近似式を使っており，その次数は偶数べき乗の6次，つまり，$(\sin x)/x \simeq a_0+a_2x^2+\cdots+a_6x^6$ の型の近似式を使っている．

(3) 倍精度－EXP－ B-a，B-b の場合

図5



これは同一機種であっても，関数の作成者がちがうとその上手下手が良く表われるという例でもある．最大誤差では，B-bの方はB-aの約9倍にもなっている．しかもB-bの方はTANHの計算でも $x>45$ でオーバーフローになっていた．B-aの方はそういうことはなかった．

さてB-aの誤差曲線は，$x=\ln 2 \fallingdotseq 88/128$ の間隔で変化している．従ってこれは $e^x=2^{x/\ln 2}$ の式を使って算出されていると思われる．一方B-bは，$(\ln 2)/2 \fallingdotseq 44/128<x<3(\ln 2)/2 \fallingdotseq 132/128$ で山が3個，谷が2個あらわれていて，0点が6個であるから，6個の係数をもつ $a_0+a_1x+\cdots\cdots+a_5x^5$ の型の5次の近似式を使っていることがわかる．

(4)複素関数－CSIN－ A-bの場合

複素関数の中で，最も誤差の大きなCSIN をとりあげる．CSIN の実部，虚部の誤差を1/8 きざみで示したものが表4である．ただしこれでは，傾向がつかみにくいので，$x=3/4$ と固定して，$1/64 \leq y \leq 100/64$ の範囲で誤差曲線を描いてみたものが図6である．これは，図7のEXPの誤差曲線と非常に似た傾向をもっている．その理由は

$$c\sin(x+iy)=\sin x\cosh y+i\cos x\sinh y \quad \cdots\cdots\cdots (1)$$

の関係式があり，一方

$$\cosh y=(e^y+e^{-y})/2 \quad \cdots\cdots\cdots \quad (2)$$

表4

(a) CSIN の実部 [×10⁻⁷]

N/16												
24	-16	-18	-8	-47	-24	-35	-17	-24	-29	-25	-32	-27
22	-13	-13	-3	0	-11	-24	-14	-17	-12	-19	-19	-11
20	-10	-9	-3	-8	0	0	4	1	1	0	1	0
18	-12	-16	-10	-8	-22	-34	-26	-33	-26	-33	-33	-30
16	-5	-2	2	18	9	4	15	15	24	23	20	19
14	-4	1	9	18	9	9	20	14	18	23	25	21
12	-9	-12	-8	-6	-17	<u>-23</u>	18	-24	-17	-25	-28	-20
10	-5	-3	5	10	3	2	8	12	1	13	12	7
8	-6	-7	-1	2	-6	-9	-4	6	0	0	0	0
6	-6	-5	1	4	-3	-2	-1	-2	-2	0	0	0
4	-4	-1	6	11	5	3	10	10	10	11	11	7
2	-5	-5	1	4	-3	-5	-3	0	0	-1	0	0
	2	4	6	8	10	12	14	16	18	20	22	24

N/16

(b) CSIN の虚部 [×10⁻⁷]

N/16												
24	-31	-33	-27	-25	-21	-20	-34	-15	-3	-13	-18	-8
22	-25	-17	-19	-13	-18	0	-21	0	0	-8	-14	-.7
20	-6	3	0	0	0	0	-1.0	3	7	-3	-8	-4
18	-42	-42	-38	-36	-35	-28	-36	-20	-12	-15	-16	-7.2
16	22	30	24	20	22	21	0	<u>16</u>	17	7	-1	-1.7
14	29	30	28	26	25	23	12	18	18	8	1	-1.0
12	-34	-32	-31	-29	-27	-24	-27	-17	11	-8	-11	-4.9
10	25	25	24	23	21	20	12	15	13	7	1	-2
8	-23	-22	-21	-19	-18	-16	-18	11	-7	-8	-7	-3.2
6	-11	-11	10	-10	-9	-8	-10	-5	-3	-4	-4.5	-2.0
4	29	-28	28	19	24	23	17	16	14	8.7	4.2	1.3
2	-5	-5	-4	-5	-4	-4.1	-4.5	-2.9	-1.9	-1.9	-1.8	-7.9
	2	4	6	8	10	12	14	16	18	20	22	24

N/16

図6

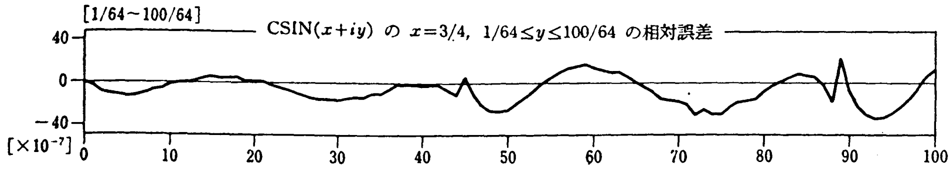
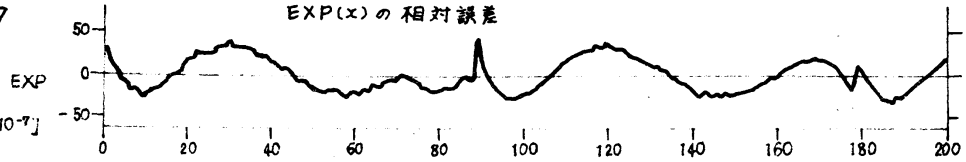


図7



従って x を固定した場合には，EXPの誤差がきれいにあらわれたものと思われる．そこで $x=y=0.75$ のとき，つまり表4のアンダーラインの部分の誤差が，EXPの誤差とどのような関係になっているかを考察してみる．(1)，(2)より

$$\mathrm{Re}(\sin(x+iy)) = (e^{y}+e^{-y})/2 \cdot \sin x \quad \cdots\cdots \quad (3)$$

また $x=y=0.75=96/128$ だから，EXP(96/128)の誤差を ε_{96} とすれば，図7から $\varepsilon_{96} = -63.5 \times 10^{-7}$

$$\therefore \cosh(0.75) = [e^{0.75}+\varepsilon_{96}+(e^{0.75}+\varepsilon_{96})^{-1}]/2$$

$$\fallingdotseq 1.2947 - 24.67 \times 10^{-7}$$

一方図2(b)より $\sin(96/128)$ の誤差は -5×10^{-7} だから

$\sin(0.75) = 0.6816 - 5\times10^{-7}$　従って(3)式は

$Re(\sin(x+iy)) = 0.8825 - 23.27\times10^{-7}$

これは表4に示した値(N=12)の −23とよくあっている。次に $x = y = 1 = 16/16$ の時の虚部を考察すると誤差は 17.35×10^{-7} となり、これも表4の値の16とよくあっている。その他の x, y についても同様に確かめてみると、CSINの実部、虚部のそれぞれの誤差はEXP, SIN, COSの誤差で説明がつけられる。従ってCSINは関係式(3), (4)を使って、単精度のEXP, SIN, COSから求められたものであることがわかる。

3. 結論

紙面の都合で我々の行った誤差曲線の解析は一部しか出来なかったのでくわしくは文献1) 2) を参照されたい。これらの解析の結果次のような結論が得られた。

1) 近似式の選び方は、必要精度ぎりぎりの近似式を選ぶよりは、少し次数の高い近似式を選んだほうがよい。次数を一次上げると1/10程度に誤差が小さくなる例が多く[3)]、演算時間が多少多くなっても、最近は計算速度は速いから誤差が少ないほうがよいからである。

2) 定数は、倍精度に入力して2進で単精度分を取り出すとかの方法をとれば、精度はよくなる。

3) 定数ミス，プログラムミスを防ぐためには，誤差曲線を書くのが最良の方法で出力を紙テープまたはカードに出し，これと倍精度または4倍精度とを比較すれば，倍精度までの誤差曲線は比較的簡単に求まる．ですからユーザも，誤差曲線を要求する習慣にするとよいと思う．

4) 使用頻度の高いプログラムに関しては，時間と手間をかけて，良いソフトウェアを作ること，そのためにはユーザがソフトウェアの良し悪しを判断し，正当な評価をしなければならない。

5) ソフトウェアの情報交換が，他社とはもとより自社内でも良くないようで，社会全体から見た場合，多大なロスをしているようで，情報交換の場がほしいものである．

参考文献

1. 春海，その他； bit，1976.7 PP 44－50
2. 春海，その他； bit，1976.8 PP 60－66
3. 一松，関数近似について（特に近似公式の選択について）第5回プログラムシンポジウム報告集，N－1，1964．