数理解析研究所講究録
第149巻1972年34-45

標準関数の精度の検定

東大(理)　高橋秀俊
東大(理)　国井利泰
東大(大型計算機センター)　名取亮
東大(地震研)　桧山澄子

§1　序

計算機に組み込まれている基本外部関数、および関数表の精度を検定するために、原器となるプログラムを作成したので報告する。まず原器を作る際の目標は

(1) 多重精度をもつこと

(2) 汎用性があること(大抵の計算機にかけられること)

(3) 関連のある関数は、できるだけその関係を利用すること

(4) 時間はこの際問題にしないこと

(5) HARP(5020E)の4倍精度基本外部関数の検定を行なうこと

の5つであった。そのために次のようなことを行なった。

(1)について；一応任意の桁まで求まるようになっているが、ここでは目標(5)のために（HARPの4倍精度は35.8桁）40桁までを正確に求めて検定を行なうことにした。そのために計算はすべて42桁で行なった。また優れた精度を得るために連分数展開式、またはニュートン近似式を用いて計算した。

(2)について；多重精度の四則演算・入出力、4倍精度を多重精度になおすプログラムが必要になるが、これらはアセンブラーでなくすべてFORTRANで用意した。

(3)について；関数間に関連があっても、良い精度が得られるものでなければ、使わない。そのため図1に示すような計算手順になった。

(4)について；時間はこの際問題でなかったが、$\sqrt{2}$，$\tan(\pi/4)$の40桁までの計算は、それぞれ1.620秒，12.709秒であった。

(5)について；表1の区間に対し関数の相対誤差をとり検定を行なった。結果は図2に示す。

§2　各関数の計算方法とアルゴリズム

(1) $\tan(x)$

$0 \leq x \leq \frac{\pi}{4}$ の x に対し

$$\tan(x) = \frac{x}{1-}\ \frac{x^2}{3-}\ \frac{x^2}{5-\cdots}\ \frac{x^2}{(2k+1)-\cdots} \qquad (1)-(\mathrm{i})$$

なる連分数展開を行なう。このとき第何項まで計算するかが問題になるが、§3で述べる方法で決めた項まで計算する。また連分数展開式の計算方法についても§3で述べる。次に区間の拡張を次式により行なう。

$$\tan(x) = 1/\tan(\tfrac{\pi}{2}-x) \qquad (\tfrac{\pi}{4}+n\pi < x < \tfrac{\pi}{2}+n\pi)$$

$$\tan(x) = -\tan(-x) \qquad (x<0)$$

サブルーチンMTAN1は(i)までの計算であり、サブルーチンMTANは区間の拡張を行なった全区間での計算である。以下の関数の計算についてもサブルーチン名に1がついたものは、限られた閉区間のxについての計算を行なうサブルーチンである。

(2) $\sin(x)$

$0 \leq x \leq \frac{\pi}{2}$のとき

$$\sin(x) = 2\tan(\tfrac{x}{2}) / (1+\tan^2(\tfrac{x}{2}))$$

以下これを用いて

$\frac{\pi}{2} < x \leq \pi$のとき　$\sin(x) = \sin(\pi - x)$

$\pi < x \leq \frac{3}{2}\pi$のとき　$\sin(x) = -\sin(x-\pi)$

$\frac{3}{2}\pi < x < 2\pi$のとき　$\sin(x) = -\sin(2\pi - x)$

(3) $\cos(x)$

$0 \leq x \leq \frac{\pi}{4}$のとき

$\cos(x) = (1-\tan^2(\frac{x}{2}))/(1+\tan^2(\frac{x}{2}))$

以下これを用いて

$\frac{\pi}{4} < x \leq \frac{3}{4}\pi$のとき　　$\cos(x) = \sin(\frac{\pi}{2} - x)$

$\frac{3}{4}\pi < x \leq \frac{5}{4}\pi$のとき　　$\cos(x) = -\cos(\pi - x)$

$\frac{5}{4}\pi < x \leq \frac{7}{4}\pi$のとき　　$\cos(x) = \sin(x - \frac{3}{2}\pi)$

$\frac{7}{4}\pi < x < 2\pi$のとき　　$\cos(x) = \cos(2\pi - x)$

(4) $\arctan(x)$

$0 \leq x \leq \sqrt{2}-1$のとき

$$\tan^{-1}(x) = \frac{x}{1+}\,\frac{x^2}{3+}\,\frac{4x^2}{5+\cdots}\,\frac{k^2x^2}{(2k+1)+\cdots} \qquad (4)-(i)$$

をある有限項（§3参）まで計算する。以下

$\sqrt{2}-1 < x < \sqrt{2}+1$のとき　　$y = \frac{\pi}{4} + \tan^{-1}(x-1)/(x+1)$

$\sqrt{2}+1 < x$ のとき　　$y = \frac{\pi}{2} - \tan^{-1}(\frac{1}{x})$

より(4)−(i)に帰着させる。

(5) $\arcsin(x)$

$|x| < \frac{\sqrt{2}}{2}$ のとき

$$\sin^{-1}(x) = \tan^{-1}(x/\sqrt{1-x^2})$$

$\frac{\sqrt{2}}{2} \leq |x| \leq 1$のとき

$$\sin^{-1}(x) = \mathrm{sgn}(x)\left[\frac{\pi}{2} - \tan^{-1}\left(\frac{\sqrt{1-x^2}}{|x|}\right)\right]$$

(6) $\arccos(x)$

$$\cos^{-1}(x) = \frac{\pi}{2} - \sin^{-1}(x)$$

(7) $\tanh(x)$

$0 \leq x \leq 1$ のとき

$$\tanh(x) = \frac{x}{1+} \frac{x^2}{3+} \frac{x^2}{5+\cdots} \frac{x^2}{(2k+1)+\cdots} \qquad (7)-(i)$$

をある有限項まで計算する。

$x > 1$ のとき二倍角

$$\tanh(2x) = 2\tanh(x) / (1+\tanh^2(x))$$

を用い(7)−(i)に帰着して求める。

(8) $\sinh(x)$

$|x| \leq 1$ のとき

$$\sinh(x) = 2\tanh(\tfrac{x}{2}) / (1-\tanh^2(\tfrac{x}{2}))$$

$|x| > 1$ のとき 二倍角

$$\sinh(2x) = 2\sinh(x)\sqrt{1+\sinh^2(x)}$$

を用いて計算する。

(9) $\cosh(x)$

$|x| \leq 1$ のとき

$$\cosh(x) = (1+\tanh^2(\tfrac{x}{2})) / (1-\tanh^2(\tfrac{x}{2}))$$

$|x| > 1$ のとき二倍角

$$\cosh(2x) = 2\cosh^2(x) - 1$$

を用いて計算する。

(10) $\exp(x)$

$0 \leq x \leq 1$ のとき

$$e^x=(1+\tanh(\tfrac{x}{2}))\,/\,(1-\tanh(\tfrac{x}{2})) \qquad (10)-(\mathrm{i})$$

$x>1$のとき　$e^x=(e^{\frac{x}{2}})^2$

$x<0$のとき　$e^x=(e^{-\frac{x}{2}})^{-2}$

を用いて(10)-(i)に帰着させる。

(11)　$ln\,(x)$

$0.5\leq x\leq 2$のとき、即ち　$-\frac{1}{3}\leq z\leq\frac{1}{3}$のとき

$$ln\,(x)=ln\left(\frac{1+z}{1-z}\right)=\frac{2z}{1-}\ \frac{z^2}{3-}\ \frac{4z^2}{5-\cdots}\ \frac{k^2z^2}{(2k+1)-\cdots} \qquad (11)-(\mathrm{i})$$

なる連分数展開を用いて計算する。

$0<x<0.5$のとき

$x\cdot 2^n>0.5$なるnに対して

$y=\log(x\cdot 2^n)-n\log 2$

$x>2$のとき

$\frac{x}{2^n}<2$なるnに対して

$y=\log(\frac{x}{2^n})+n\log 2$

を計算する。

(12)　$\sqrt{x}$

$0.01\leq x<1$に対してニュートンの近似式

$$y_{n+1}=\frac{1}{2}\left(y_n+\frac{x}{y_n}\right)$$

を用いて計算する。第何項まで計算するかは、求める桁数をNとしたとき

$|y_{n+1}-y_n|\,/\,y_n\leq 10^{-N}$

になるまで計算する。また区間の拡張は

$x \geq 1$ のとき

$$y = 100^{[\frac{k}{2}]} \cdot 10^{k-2[\frac{k}{2}]} \sqrt{x \cdot 100^{-k}}$$

$x < 0.01$ のとき

$$y = 100^{[\frac{k+1}{2}]}\, 10^{k-2[\frac{k+1}{2}]} \sqrt{x \cdot 100^{-k}}$$

より行なう。

§3　連分数展開式の計算

$\tan(x)$、$\arctan(x)$、$\tanh(x)$、$\ln(x)$ の計算は連分数(1)−(i)、(4)−(i)、(7)−(i)、(11)−(i)、で計算された。

さて、それらの連分数を実際に計算するには、今連分数

$$V_m = \frac{x_0}{y_0+}\,\frac{x_1}{y_1+}\,\frac{x_2}{y_2+\cdots}\,\frac{x_m}{y_m}$$

において

$$p_{-1} = 0, \quad p_0 = x_0, \quad q_{-1} = 1, \quad q_0 = y_0$$

$$p_k = y_k\, p_{k-1} + x_k\, p_{k-2}$$

$$q_k = y_k\, q_{k-1} + x_k\, q_{k-2} \qquad (k = 1, 2, \cdots, m)$$

とすれば

$$V_m = p_m / q_m$$

となり、$2m-1$ 回の加法、$2m-1$ 回の乗算、1回の除法で計算される。

次に無限な連分数をどこで打ち切るかが問題になるが、40

桁まで正しく求めたとき、表2のような結果が得られた。

これより打ち切り項数の近似式MAXは

$\tan(x)$、$\tanh(x)$については

$$MAX = IFIX(10x)+10$$

$\arctan(x)$については

$$MAX = IFIX(40x)+12$$

$\ln(x)$については

$$MAX = IFIX(-50z)+13$$

が出るから、実際の計算では無条件にMAXまで連分数を計算し、以後

$$|V_m - V_{m+1}| / |V_m| < 10^{-N}$$

になるまでiterationを行なうことにした。

なお、プログラムのチェックは、$\sqrt{x}$以外は互に独立なプログラムを用いることで行なった。例えば $\arctan(x)$と$\tan(x)$のチェックでは

$$x = \arctan(\tan(x))$$

かどうかで行なった。

$\sqrt{x}$は

$$x = (\sqrt{x})^2$$

になるかどうかでたしかめた。

§4　HARP（5020E）との比較検定

HARP（5020E）の4倍精度の基本外部関数の精度の検定は、多重精度の関数値との相対誤差を取ることで行なった。

その結果QSQRT、QLOGは10^{-35}まで正確に求まるが、QSIN，QCOS，QATAN，QEXPは10^{-34}までの精度しかないことがわかった。

特に、QATANでは$x \geq \frac{\pi}{2}$で精度が落ちる。

結果は図2に示す。

4倍精度の検定区間

QSQRT	$0 \leq X \leq 1$	QATAN	$0 \leq X \leq 2$
QSIN	$0 \leq X \leq 6$	QLOG	$0.5 \leq X \leq 2$
QCOS	$0 \leq X \leq 6$	QEXP	$-1 \leq X \leq 1$

表1

x	tanh	tan	artan	ln	z
0.125	10	10	16	>50	-0.7
0.25	12	12	21	43	-0.6
0.375	13	13	26	32	$-0.\dot{4}\dot{5}$
0.5	14	14	31	26	$-0.\dot{3}$
0.625	15	15	37	21	$-0.\dot{2}3076\dot{2}$
0.75	15	16	41	20	$-0.\dot{1}4285\dot{7}$
0.875	16	16	47	13	$-0.0\dot{6}$
1.0	17	17	52	0	0.0

表2

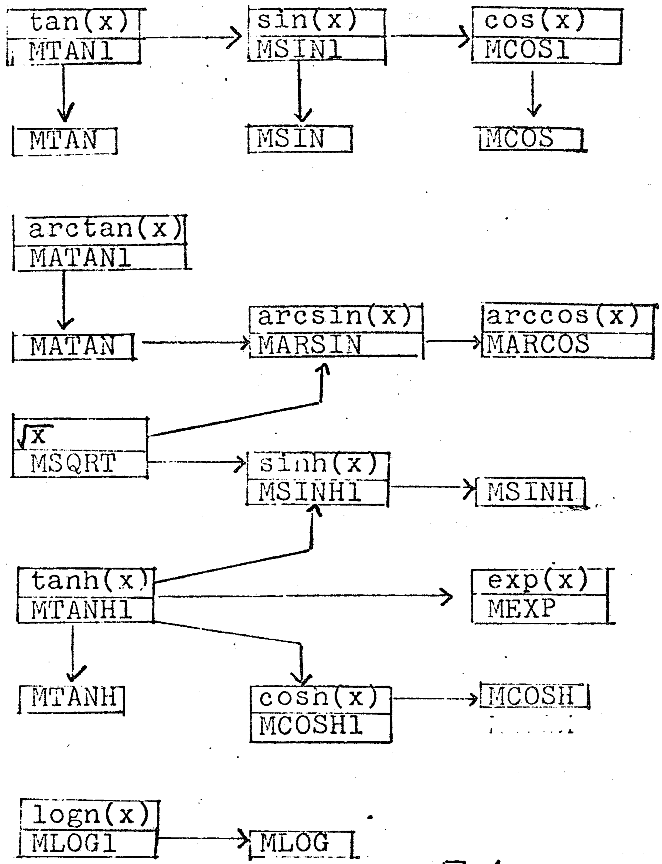
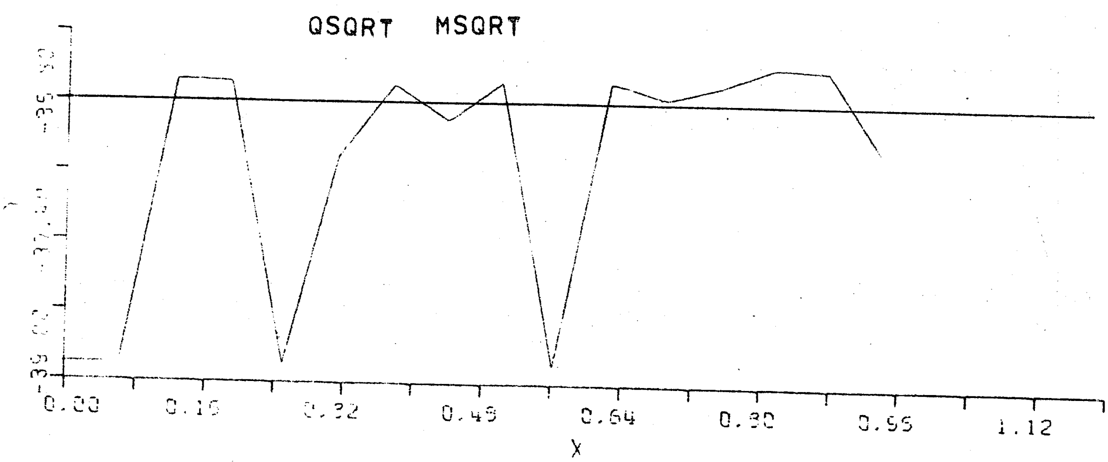
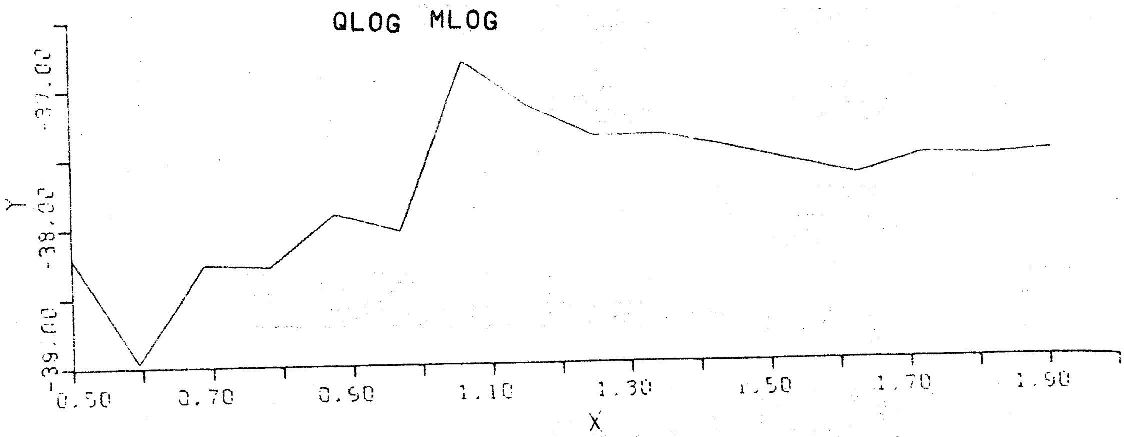


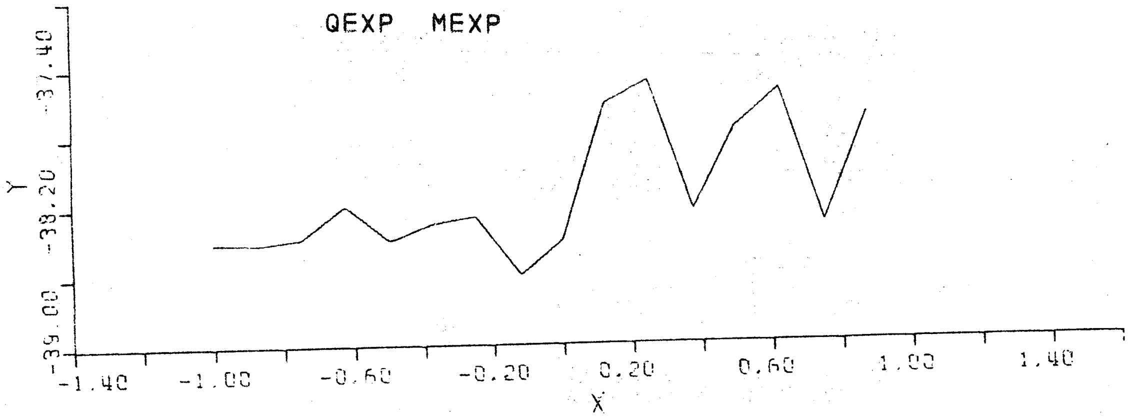
図 1

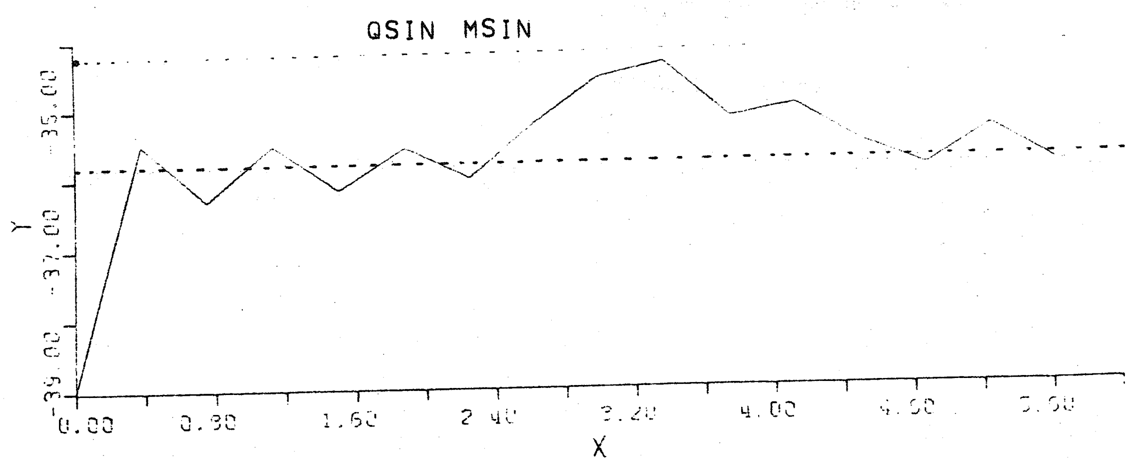
計算手順の流れ図

4倍精度の相対誤差




QLOG MLOG
-37.00
-38.00
-39.00
Y
0.50
0.70
0.90
1.10
1.30
1.50
1.70
1.90
X


QEXP MEXP
-37.40
-38.20
-39.00
Y
-1.40
-1.00
-0.60
-0.20
0.20
0.60
1.00
1.40
X


QSIN MSIN
-35.00
-37.00
-39.00
Y
0.00
0.80
1.60
2.40
3.20
4.00
X

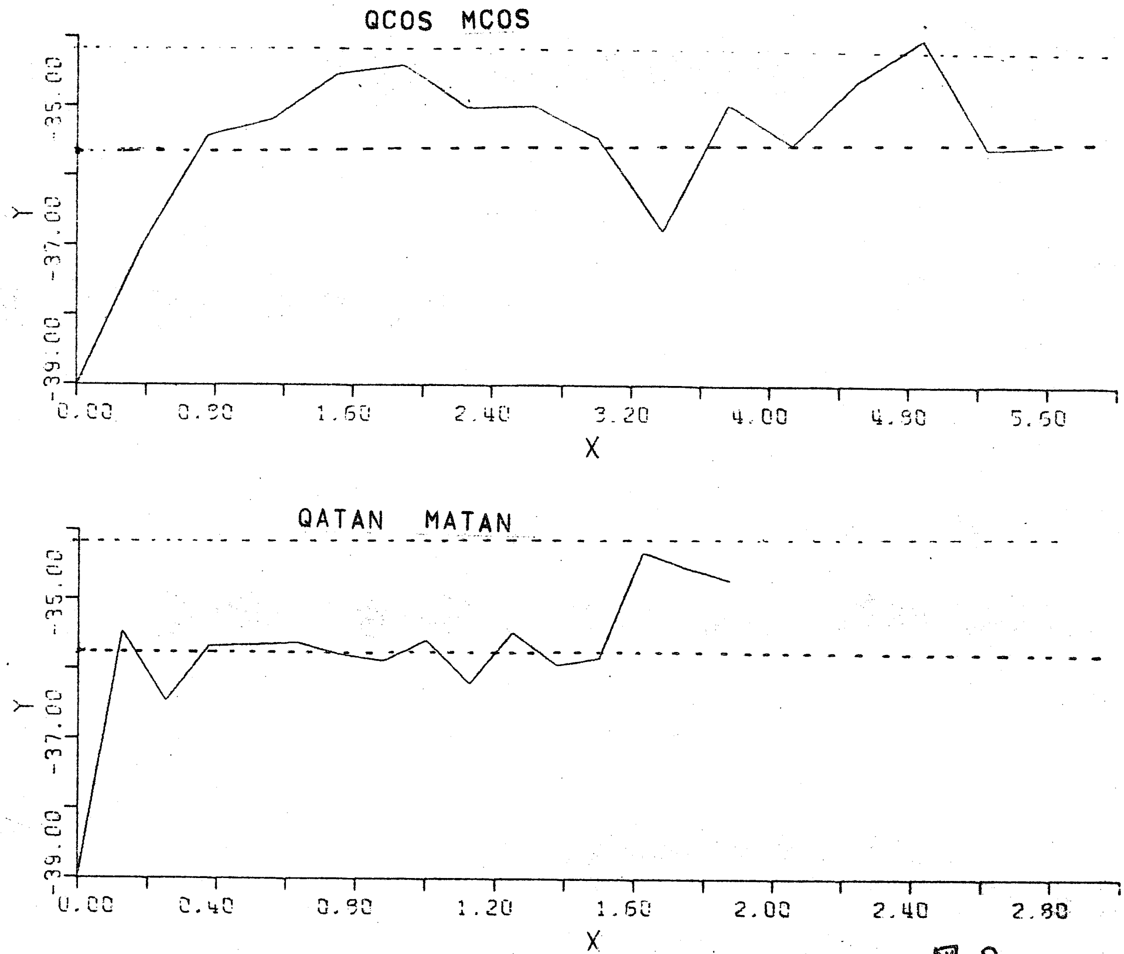


図2

References

Hart;Computer approximations,1968,John Wiley
I.D.Hill;Procedures for the basic arithmetical operations in multiple-length warking,Comp.J.,1968 PP232-235
C.T.Fike;Computer evaluation of mathematical functions; 1968,Prentice Hall