On Bounding the Range of Some Elementary Functions in FORTRAN–77

Chenyi Hu, R. Baker Kearfott, and Abdulhamid Awad*

We present the techniques we have used to bound the range of the arcsine, arccosine, arctangent, arccotangent, and hyperbolic sine functions in our portable FORTRAN–77 library INTLIB. The design of this library is based on a balance of simplicity and efficiency, subject to rigor and portability.

О вычислении границ множества значений некоторых элементарных функций в Фортране 77

Ч. Ху, Р. Б. Кирфотт, А. Авед

Представлена техника, используемая нами для ограничения множества значений функций арксинуса, арккосинуса, арктангенса и гиперболического синуса в нашей переносимой библиотеке INTLIB на Фортране 77. При проектировании библиотеки преследовалась цель достижения баланса между простотой и эффективностью, точностью и переносимостью.

*This research was partially supported by National Science Foundation Grants No. CCR-9203730 and No. DMS-9205680.
© C. Hu, R. B. Kearfott, A. Awad, 1994
1 Introduction

INTLIB ([7] and [8]) is a portable FORTRAN–77 library to support interval arithmetic evaluation of the elementary functions. Here, we describe the subroutines in INTLIB for bounding the arcsine, arccosine, arctangent, arccotangent, and hyperbolic sine functions. These subroutines are called IASIN, IACOS, IATAN, IACOT, and ISINH respectively.

Our basic algorithms are based on Taylor’s theorem, with special argument reduction and choice of expansions. Our choice of the Taylor approximation allows us to achieve

1) ease of obtaining mathematically rigorous bounds on the truncation error;
2) portability;
3) fast convergence rates, and
4) ease with which the FORTRAN source code can be read and maintained.

There are numerous alternatives, such as polynomial interpolation, rational interpolation or minimax approximation, etc. Polynomial interpolants have error bounds similar to those of Taylor polynomials, and share other similarities, but are slightly more complicated. Rational functions have been used in interval computations to approximate elementary functions; for example, see [2, §2.4], following [12]. However, to our knowledge, in general all coefficients change if we increase the degree of approximation. Furthermore, we know of no simple error formulas amenable to interval computation. For example, in [11, problem 19, p. 312], the error term for rational interpolation of a function \( f \) by \( P(x)/Q(x) \) at \( x_i, 1 \leq i \leq n \) is given by

\[
e(x) = \frac{p_n(x)}{n!Q^2(x)} \frac{d^n}{dx^n} \left[ Q^2(x)f(x) \right] \bigg|_{x=\xi}
\]

where \( p_n(x) = \prod_{i=1}^{n} (x - x_i) \). Contrast this with the simple error term for the arcsine in equation (4) below.

Related rationales are given in [3] and [9] for employing Taylor series in portable interval libraries for the standard functions. There are, however,
some techniques beyond the Taylor approximations that will be useful in our context; see §5 below.

Specifically, the following well-known series expansions underlie the approximations in this paper.

\[
\begin{align*}
\arcsin x &= x + \frac{x^3}{2\cdot3} + \frac{1\cdot3x^5}{2\cdot4\cdot5} + \cdots + \frac{(2n)!x^{2n+1}}{(2n\cdot n!)^2(2n + 1)} + \cdots \tag{1} \\
\arctan x &= x - \frac{x^3}{3} + \frac{x^5}{5} - \cdots + (-1)^{n-1}\frac{x^{2n-1}}{2n - 1} + \cdots \tag{2} \\
\sinh x &= x + \frac{x^3}{3!} + \frac{x^5}{5!} + \cdots + \frac{x^{2n+1}}{(2n+1)!} + \cdots \tag{3}
\end{align*}
\]

We bound these elementary functions with interval arithmetic.\footnote{Throughout this paper, we use boldface letters to denote intervals; we use \(x\) and \(\overline{x}\) to denote the lower bound and upper bound, respectively, for an interval \(x\).} In interval computations, both the number of computations and the widths of the input intervals affect the width of the result interval. Hence, we want to

(a) use narrow intervals \(x\) in the above series and

(b) accelerate the rate of convergence to reduce the number of computations.

Let us address (a) first. The range of a continuous monotone function \(f(x)\) is either \([f(x), f(\overline{x})]\) or \([f(\underline{x}), f(x)]\). Therefore, we may bound \(f(x)\) by replacing both \(x\) and \(\overline{x}\) by the smallest machine representable intervals\footnote{This can be done by using the subroutine RNDOUT in [6].} \(l\) and \(u\) containing \(x\) and \(\overline{x}\), respectively, then bounding \(f(l)\) and \(f(u)\). If \(f(x)\) is a monotonically increasing function, then \(f(x) \subseteq [f(l), f(u)]\), where \(f(x)\) denotes the exact range of \(f\) over \(x\). If \(f(x)\) is a monotonically decreasing function, then \(f(x) \subseteq [f(u), f(l)]\). Since the widths of \(l\) and \(u\) are small, the overestimation in the bounds for \(f(x)\) due to the width of the operand will be limited. This idea is used for all subroutines in this paper, since all functions considered in this paper are monotone.

To guarantee that the computed range contains \(f(x)\) for all \(x \in x\), we need to over- or underestimate the series as appropriate. For example, we need to overestimate the series (1) if we replace \(x\) by \(u\) and to underestimate (1) if we replace \(x\) by \(l\).
We address (b) and error estimations for each of these subroutines in the following sections.

2 IASIN and IACOS

IASIN and IACOS bound the arcsine function and arccosine function respectively. The core subroutine of IASIN evaluates $s_i$, the partial sum of series (1), by interval arithmetic, where

$$s_i = \sum_{n=0}^{i} \frac{(2n)!x^{2n+1}}{(2n)!^2(2n+1)}$$

and $0 \leq x \leq 0.5$. Since arcsin $x$ is an odd function, we assume $x \geq 0$. If $x < 0$, we compute arcsin $x$ by arcsin $x = -\text{arcsin} (-x)$. We assume $x \leq 0.5$ for fast convergence. If $x > 0.5$, we define $y = \sqrt{1-x^2} < 0.5$ and use $y$ in equation (1). Then, we apply the identity arcsin $x = \pi/2 - 2 \text{arcsin} \sqrt{1-x^2}$ to find arcsin $x$.

Suppose $i$ terms are used in evaluating (1), so that the interval truncated from (1) is $e_i = \sum_{n=i+1}^{\infty} \frac{(2n)!x^{2n+1}}{(2n)!^2(2n+1)}$. Since $\frac{(2n)!}{(2n)!^2(2n+1)}$ is a decreasing sequence, we may apply the geometric series with common ratio $r = 0.5 \geq x$ to overestimate the maximum absolute truncation error $e_i = \max_{e \in e_i} |e|$. We thus obtain

$$e_i = \sum_{n=i+1}^{\infty} \frac{(2n)!x^{2n+1}}{(2n)!^2(2n+1)} \leq \frac{(2i+2)!x^{2i+3}}{(2(i+1)(i+1)!^2(2i+3)} \sum_{n=1}^{\infty} x^n \leq \frac{2(2i+2)!x^{2i+3}}{(2(i+1)(i+1)!^2(2i+3)}.$$ (4)

If $s = 0$, we use sin 0 = 0 to avoid computing equation (1). If $s \neq 0$, the maximum relative error is less than $\frac{\max e_i}{\min s_i} \leq \frac{(2i)!2\pi^{2i+1}}{(2i)!^2(2i+1)\pi}$. We terminate accumulation of the sum $s$ when this relative error bound is less than a tolerance $\epsilon$ (denoted TOL0 in the code), related to the relative error in a single floating-point operation. If we need to overestimate $s$, we add
the maximum absolute truncation error to the partial sum. If we need to underestimate the series, we simply truncate all terms from the \( i + 1 \) term.

Here is the algorithm to bound \( \text{arcsin} \, x \). In the algorithm, we use a logical variable \texttt{Negative} to indicate if \( x < 0 \) or not. We use another logical variable \texttt{Over} to indicate if we need to over estimate \( s \) or not. The variable \texttt{Code} records argument reductions on \( x \), so that the result may be transformed correctly. Some irrational numbers such as \( \pi/2, \sqrt{2}, \ldots \) are used in our algorithms. They are defined as interval constants in INTLIB.

In Algorithm 1, Algorithm 2, and Algorithm 3, Step 3 is analogous to Step 2. However, the rounding direction differs, and depends on what argument reduction was previously done. We do not include details for Step 3 of these algorithms here. However, we have a complete writeup of Step 3 (and indeed, the FORTRAN source) for those interested in implementing these algorithms.

**Algorithm 1.**

1. Input the interval \( x = [\underline{x}, \overline{x}] \).

2. Compute the lower bound of \( \text{arcsin} \, x \).
   
   
   (a) If \( \underline{x} < 0 \), then \texttt{Negative} \( \leftarrow \text{True} \), \( \underline{x} \leftarrow -\underline{x} \), \texttt{Over} \( \leftarrow \text{True} \),
   else \texttt{Negative} \( \leftarrow \text{False} \), \texttt{Over} \( \leftarrow \text{False} \).
   
   (b) Represent \( x \) by an interval \( I \).
   
   (c) If \( I > 0.5 \), then \texttt{Code} \( \leftarrow 12 \), \( I \leftarrow \sqrt{(1 - I)/2} \),
   else \texttt{Code} \( \leftarrow 11 \).
   
   (d) Compute \( s = 1 + \frac{1^3}{1!} + \cdots + \frac{(2i)!l^{2i+1}}{(2i)!^2(2i+1)} \) until \( \frac{(2i + 2)!l^{2i+3}}{(2(i+1)(i + 1)!)^2(2i + 3)\bar{s}} < \epsilon \).
   
   (e) If \texttt{Over} = \text{True} then \( s \leftarrow s + \frac{(2i + 2)!l^{2i+3}}{(2(i+1)(i + 1)!)^2(2i + 3)\bar{s}} \).
   
   (f) If \texttt{Code} = 12 then \( s \leftarrow \pi/2 - s \).
   
   (g) If \texttt{Negative} = \text{True} then \( \text{arcsin} \, x \leftarrow -\bar{s} \) else \( \text{arcsin} \, x \leftarrow s \).

3. Compute the upper bound of \( \text{arcsin} \, x \), analogously to Step 2.

Since \( \text{arccos} \, x = \frac{\pi}{2} - \text{arcsin} \, x \), we may first find \( \text{arcsin} \, x \) then subtract it from \( \pi/2 \) to obtain \( \text{arccos} \, x \). Therefore, we do not present IACOS in detail.
3 IATAN and IACOT

IATAN and IACOT bound ranges of the inverse tangent function and inverse cotangent function, respectively. The series (2) with coefficient $O(1/n)$ converges too slowly. For faster convergence, we multiply both sides of (2) by $x^2$. We obtain

$$x^2 \arctan x = x^3 - \frac{x^5}{3} + \frac{x^7}{5} - \cdots + (-1)^{n-1} \frac{x^{2n+1}}{2n-1} + \cdots. \quad (5)$$

By adding (2) and (5), we have

$$(1 + x^2) \arctan x = x + \left(1 - \frac{1}{3}\right)x^3 - \cdots$$

$$+ (-1)^{n+1} \left(\frac{1}{2n-1} - \frac{1}{2n+1}\right)x^{2n+1} + \cdots. \quad (6)$$

Dividing both sides of (6) by $1 + x^2$, we obtain

$$\arctan x = \frac{x + 2 \left(\frac{1}{13}x^3 - \frac{1}{35}x^5 + \cdots + (-1)^{n+1} \frac{x^{2n+1}}{(2n-1)(2n+1)} + \cdots\right)}{1 + x^2}. \quad (7)$$

Letting $s = \frac{1}{13}x^3 - \frac{1}{35}x^5 + \cdots + (-1)^{n+1} \frac{x^{2n+1}}{(2n-1)(2n+1)} + \cdots$, (7) becomes

$$\arctan x = \frac{x + 2s}{1 + x^2}. \quad (8)$$

The series $s$ converges much faster than (2) does, since its coefficient $a_n = O(1/n^2)$. In our program, we use equation (8) with interval arithmetic to bound $\arctan x$. We assume $x \geq 0$, because the inverse tangent function is an odd function; if $x < 0$, we use $\arctan x = -\arctan (-x)$. We also assume $|x| \leq 1$. If $|x| > 1$, we apply $\arctan x = \pi/2 - \arctan \frac{1}{x}$ to find $\arctan x$.

The series $s$ also converges at $|x| = 1$. However, to further increase the rate of convergence, we only evaluate series (8) when $0 \leq x \leq \frac{\sqrt{3}}{3}$. If $x > \frac{\sqrt{3}}{3}$, we let $y = \frac{\sqrt{3}x - 1}{\sqrt{3}+x}$. It is easy to prove that $0 \leq y \leq \frac{\sqrt{3}}{3}$ when $\frac{\sqrt{3}}{3} \leq x \leq 1$. We then apply the identity $\arctan x = \pi/6 + \arctan \frac{\sqrt{3}x-1}{\sqrt{3}+x}$ to find $\arctan x$. 
The maximum absolute truncation error of \( s \) from the \( i \)-th term is less than \( \frac{2^{i+1}}{(2i-1)(2i+1)} \). This is because \( s \) is an alternating series. We use \( \frac{2^{i+3}}{(2i+1)(2i+3)} / \frac{2^i}{(2i-1)(2i+1)} \) to estimate the maximum relative error. If the relative error less than \( \epsilon \), we either terminate the computation or add one more term to \( s \), depending on the sign of the last term and whether we need to overestimate \( s \) or underestimate it. When we need to overestimate the series \( s \), we should chop the series after the \( i \)-th term if \( i \) is an odd number. When we need to underestimate the series \( s \), we should chop the series after the \( i \)-th term if \( i \) is an even number.

Here is the algorithm to bound \( \arctan x \). In the algorithm, a logical variable \( \text{Negative} \) is used to indicate if \( x < 0 \) or not. Another logical variable \( \text{Even} \) indicates whether or not we need to overestimate \( s \). The variable \( \text{Code} \) records argument reduction transformations that have been applied to \( x \).

**Algorithm 2.**

1. Input the interval \( x = [\underline{x}, \overline{x}] \).
2. Compute the lower bound for \( \arctan x \).
   
   (a) Transform \( \underline{x} \) to make \( 0 \leq \underline{x} \leq \frac{\sqrt{3}}{3} \).
      
      i. If \( \underline{x} < 0 \), then \( \text{Negative} \leftarrow True \), \( \underline{x} \leftarrow -\underline{x} \),
        else \( \text{Negative} \leftarrow False \).
      
      ii. If \( \underline{x} > 1 \) then \( \underline{x} \leftarrow \frac{1}{\underline{x}} \).
      
      iii. If \( \underline{x} > \frac{\sqrt{3}}{3} \) then \( \underline{x} \leftarrow \frac{\sqrt{3}\underline{x}-1}{\sqrt{3}+\underline{x}} \).
      
      iv. We record the transformations by the following:
          If neither (b) nor (c) has been performed, \( \text{Code} \leftarrow 11 \).
          If only (c) has been performed, \( \text{Code} \leftarrow 12 \).
          If only (b) has been performed, \( \text{Code} \leftarrow 21 \).
          If both (b) and (c) have been performed, \( \text{Code} \leftarrow 22 \).
   
   (b) Determine the value of \( \text{Even} \) according to the above transformation.
   
   If we need to overestimate \( s \) then \( \text{Even} \leftarrow False \),
   else \( \text{Even} \leftarrow True \).
   
   \[
   \begin{array}{c|ccccc}
   \text{Negative} \setminus \text{Code} & 11 & 12 & 21 & 22 \\
   \hline
   \text{False} & True & True & False & False \\
   \text{True} & False & False & True & True \\
   \end{array}
   \]
   
   (c) Represent \( \underline{x} \) by an interval \( l \).
   
   (d) Compute \( s = \frac{1}{1^3} - \frac{1}{3^3} + \cdots + (-1)^{i-1} \frac{1}{(2i-1)(2i+1)} \) until \( \frac{2^{i+3}}{(2i+1)(2i+3)} / \underline{x} < \epsilon \).
   
   (e) If \( \text{Even} = True \) and \( i \) is an odd number, then add the next term on \( s \).
(f) If $Even = False$ and $i$ is an even number, then add the next term on $s$.

(g) $s \leftarrow \frac{1+2s}{1+1}$.

(h) Transform $s$ according to the argument reduction transformations to make $0 \leq x \leq \frac{\sqrt{3}}{3}$.
   - If Code = 12, $s \leftarrow \pi/6 + s$.
   - If Code = 21, $s \leftarrow \pi/2 - s$.
   - If Code = 22, $s \leftarrow \pi/3 - s$.

(i) If $Negative = True$ then $\arctan x \leftarrow -s$,
   else $\arctan x \leftarrow s$.

3. Compute the upper bound for $\arctan x$, analogously to Step 2.

Since $\arccot x = \frac{\pi}{2} - \arctan x$, to find $\arccot x$ we first find $\arctan x$ then subtract it from $\pi/2$.

4 ISINH

We discuss ISINH in this section. The series (3) converges for all real numbers. To further increase the rate of convergence in our program, we assume $0 \leq x \leq 1$. We use $\sinh x = -\sinh(-x)$ when $x < 0$. In our program, we repeatedly apply the identity $\sinh x = 3\sinh(x/3) + 4\sinh^3(x/3)$ to evaluate $\sinh x$ if $1 < |x| \leq 3^3 = 27$. For arguments with $|x| > 27$, we use INTLIB’s exponential function IEXP. IEXP, in turn, has error checking for arguments that would overflow.

Here is our algorithm.

**Algorithm 3.**

1. Input the interval $x = [x, \pi]$.

2. Compute the lower bound for $\sinh x$.
   (a) If $x < 0$, then $Negative \leftarrow True$, $x \leftarrow -x$, $Over \leftarrow True$,
   else $Negative \leftarrow False$, $Over \leftarrow False$.
   (b) If $0 \leq x \leq 1$, then Code $\leftarrow 11$.
   - If $1 < x \leq 3$, then Code $\leftarrow 12$, $x \leftarrow x/3$.
   - If $3 < x \leq 9$, then Code $\leftarrow 13$, $x \leftarrow x/9$.
   - If $9 < x \leq 27$, then Code $\leftarrow 14$, $x \leftarrow x/27$.
   - If $27 < x$, then branch to IEXP.
(c) Represent $x$ by an interval $I$.

(d) Compute $s = 1 + \frac{l_3^3}{3!} + \frac{l_5^5}{5!} + \cdots + \frac{l_{2i+1}^{2i+1}}{(2i+1)!}$ until $\frac{2^i l_{2i+3}^{2i+3}}{(2i+3)!} \leq \epsilon$.

(e) If $Over = True$, $s \leftarrow s + \frac{2^i l_{2i+3}^{2i+3}}{(2i+3)!}$.

(f) For $k = 12$ to Code do: $s \leftarrow 3s + 4s^3$.

(g) If $Negative = True$ then $\sinh x \leftarrow -s$,
else $\sinh x \leftarrow s$.

3. Compute the upper bound for $\sinh x$, analogously to Step 2.

We have also have an algorithm, ICOSH, for the hyperbolic cosine function. However, we have not yet perfected it for inclusion in INTLIB at the time we submit this paper.

5 Additional improvements

The Ph. D. dissertations [3] and [9] have recently become available to us. These works provide a careful consideration of implementation of the elementary functions for arbitrary floating-point systems. As here, much (though not all) of these works are based on Taylor approximations. These works emphasize high or maximal precision (approximation by the nearest machine number) in the function values. That requires higher than working precision to attain. Nonetheless, much of their development can be beneficially applied to INTLIB. For instance, formulas appear for the total roundoff error in the evaluation of a polynomial using Horner’s scheme in terms of roundoff error in a single operation, as well as for the roundoff error occurring from certain types of argument reductions. With such formulas, we may do the computations in steps 2d and 3d of algorithms 1, 2, and 3 in the machine’s floating-point arithmetic, instead of in our portable interval arithmetic. Though some rearrangement of the computations will be necessary, this change will clearly greatly increase the speed of these algorithms.

Another possible improvement is the use of Chebyshev economization; see e.g. [4, pp. 456–459]. The additional error terms introduced by dropping the trailing terms in the Chebyshev basis can be easily bounded using interval arithmetic. However, the resulting polynomial coefficients are no longer as simply representable, causing possible minor difficulties. For example,
some of the roundoff error bounds in [3] and [9] depend on the zero-th or-
der coefficient of the polynomial being exactly representable in the machine.
Also, computing these coefficients complicates maintenance of the routines,
should it be desirable to increase or decrease the degree of approximation.

6 Summary

Based on the above algorithms, we have developed subroutines: IASIN,
IACOS, IATAN, IACOT, and ISINH in FORTRAN–77 ([1]). These subrou-
tines have been tested and exhaustively documented. They are included in
the readily available and portable software library INTLIB ([7]). INTLIB
has been submitted to ACM Transactions on Mathematical Software ([8]).
We are continuing to test and improve it. A test version is available through
the Internet via FTP at ftp.ucs.usl.edu in the directory
/pub/interval_math/intlib.

Acknowledgements

We wish to acknowledge the referees for their careful reading and useful
suggestions. We also wish to thank George Corliss for editorial changes that
made the paper, and in particular for making Step 2b of Algorithm 2 easier
to read.

References

[1] American national standard programming language FORTRAN ANSI

Intervallmathematik – Theorie und Anwendungen. Teubner, Leipzig,
1987.

Punkte und Intervalle in beliebigen Gleitpunktrastern. Ph. D. disserta-


C. Hu, A. Awad  
Department of Computer and  
Mathematical Sciences  
University of Houston-Downtown  
Houston, TX 77002  
USA

R. B. Kearfott  
Department of Mathematics  
University of Southwestern Louisiana  
Lafayette, LA 70504  
USA