E. G. Kogbetliantz

Computation of Arcsin *N* for 0<*N*<1 Using an Electronic Computer

Abstract: All known subroutines for Arcsine are based on the relation Arcsin N = Arctan $[N/(1-N^2)^{\frac{1}{2}}]$. Therefore, Arcsine is not computed as such but as an Arctangent.

To avoid the loss of machine time caused by the computation of $N/(1-N^2)^{\frac{1}{2}}$, a direct computation of Arcsine is proposed. A subroutine yielding the first six correct significant digits in only five multiplications and divisions is described in detail to illustrate the new method's rapidity. The same number of five operations is necessary to compute, knowing N, the number $N/(1-N^2)^{\frac{1}{2}}$.

Introduction

Our results are based on a new rational approximation to Arcsin N deduced with the aid of Dr. H. Maehly's method[1] from the Chebyshev expansion of Arcsin (xSin 2θ), where $N=x\text{Sin}2\theta, 0\leq x\leq 1$:

$$\text{Arcsin}\,(x\text{Sin}2\theta) = \sum_{n=0}^{\infty} c_n(\theta)T_{2n+1}(x). \qquad (0<\theta<\pi/4) \quad (1)$$

Since $N=1$ is a singular point for Arcsin N, the rational approximation yields the required accuracy only for $N\leq N_0 = \text{Sin}\ 2\theta$, where $\theta<\pi/4$. In the sequel we take $\theta=\pi/8$, and in the remaining part of the range use the relations

$$\text{Arcsin}\ N = \frac{\pi}{4}+\frac{1}{2}\,\text{Arcsin}\,(2N^2-1) \qquad (2)$$

$$\text{Arcsin}\ N = \frac{\pi}{2} - 2\,\text{Arcsin}\,[(1-N)/2]^{\frac{1}{2}}. \qquad (3)$$

If Sin $\pi/4\leq N\leq$ Sin 3 $\pi/8$, then $2N^2-1\leq$ Sin $\pi/4$ so that our rational approximation can be used again to compute Arcsin $(2N^2-1)$. However, for $N>$ Sin $3\pi/8$, we are so near the singular point that another approximation will be used.

It was important to find an easy way for a precise computation of coefficients $c_n(\theta)$ in (1). They are functions of θ, and although in our example we take $\theta=\pi/8$, our method holds for any θ in (0, $\pi/4$) except $\theta=\pi/4$. But if $\theta=\pi/4$, then $\pi c_n(\pi/4)=(n+\frac{1}{2})^{-2}$.

The expansion (1) is deduced and the method for computing its coefficients is described in the first part of this paper. The second part contains a definition of Maehly's method and its application to Arcsine. The particular case of a subroutine giving the first six correct significant digits is studied in the last part. It is hoped that this example is sufficient to permit the construction of subroutines giving more than six correct digits. When the computation of Arcsine is to be repeated many thousands of times, as often happens, the saving of machine time due to the use of a shorter and faster subroutine becomes important.

Chebyshev expansion

Let $F_m(z)$ denote the hypergeometric function $F(m+\frac{1}{2}, m+\frac{1}{2}; 2m+2; z)$, that is

$$\Gamma^2(m+\tfrac{1}{2})F_m(z)=(2m+1)!\sum_{s=0}^{\infty}\Gamma^2(m+\tfrac{1}{2}+s)z^s/s!\ (2m+1+s)!.$$

Then, as we will prove,

$$(m+\tfrac{1}{2})m!\,c_m\Gamma(\tfrac{1}{2})=(\text{Sin}\,\theta\text{Cos}\,\theta)^{2m+1}\Gamma(m+\tfrac{1}{2})F_m(\text{Sin}^2 2\theta). \quad (4)$$

It is possible to deduce (4) from the expression of c_m as a Fourier coefficient, namely

$$\pi c_m = 2\int_{-1}^{+1}\text{Arcsin}\,(x\text{Sin}\ 2\theta)T_{2m+1}(x)\,(1-x^2)^{-\frac{1}{2}}dx, \qquad (5)$$

but the transformation of Maclaurin's series

$$\text{Arcsin}\,(x\,\text{Sin}\,2\theta)=\sum_{n=0}^{\infty}(-1)^n\binom{-\frac{1}{2}}{n}(x\ \text{Sin}\ 2\theta)^{2n+1}/(2n+1)$$

into (1) with the aid of substitutions

$$(2x)^{2n+1}/(2n+1)!=2\sum_{m=0}^{n}T_{2m+1}(x)/(n-m)!\,(n+m+1)!$$

is a much easier way to prove (4).

Denoting the general term of a double series thus obtained by $u_{n-m,\,m}$, we have

$$\text{Arcsin}\,(x\ \text{Sin}\ 2\theta)=\sum_{n=0}^{\infty}\sum_{m=0}^{\infty}u_{n-m,m}=\sum_{m=0}^{\infty}\sum_{n=0}^{\infty}u_{nm}$$

where

$$u_{nm}=\lambda_m(\text{Sin}\ \theta\ \text{Cos})\ \theta^{2m+1}T_{2m+1}(x)U_{nm}(\theta)$$

with

$$\lambda_m=\Gamma(m+\tfrac{1}{2})/m!(m+\tfrac{1}{2})\Gamma(\tfrac{1}{2})$$

and

$$\Gamma^2(m+\tfrac{1}{2})(2m+n+1)!n!U_{nm}(\theta)=(2m+1)!\ \text{Sin}^{2n}2\theta\ \Gamma^2(m+n+\tfrac{1}{2}).$$

Summing up with respect to n, we obtain

$$\sum_{n=0}^{\infty}U_{nm}(\theta)=F_m(\text{Sin}^2 2\theta)$$

and this proves (4).

Hypergeometric functions satisfy recurrence relations and therefore we should expect that our c_n's satisfy one also. They do:

$$m(2m+3)^2c_{m+1}=(2m+1)\ [2m(m+1)(\tan^2\theta+\text{cotan}^2\theta)+1]c_m-(m+1)(2m-1)^2c_{m-1}. \quad (6)$$

To prove (6) we first express $F_m(z)$ in terms of $f_m(z)=F(m+\frac{1}{2},\ m+\frac{1}{2};\ 2m+1;\ z)$ as follows

$$4(m+1)F_m(z)=4(m+1)f_m(z)-(m+\tfrac{1}{2})zf_{m+1}(z). \quad (7)$$

The identity (7) is easy to check comparing the coefficients of z^n in both members. Using now the classical relation[2]

$$(1+\sqrt{1-z})^{2m+1}f_m(z)=2^{2m+1}F(\tfrac{1}{2},\ m+\tfrac{1}{2};\ m+1;\ \zeta), \quad (8)$$

where $\zeta=(1-\sqrt{1-z})^2/(1+\sqrt{1-z})^2$, and taking $z=\text{Sin}^2 2\theta$, we transform $f_m(\text{Sin}^2 2\theta)$ as follows

$$f_m(\text{Sin}^2 2\theta)=(1+t^2)^{2m+1}H_m(t). \quad (9)$$

Here $t=\tan\theta$ and
$H_m(t)=H_m=F(\frac{1}{2},\ m+\frac{1}{2};\ m+1;\ t^4)$.

Combining (7) and (9), we obtain

$$(2m+2)F_m(\text{Sin}^2 2\theta)=(1+t^2)^{2m+1}[(2m+2)H_m-(2m+1)t^2H_{m+1}]$$

and this gives

$$(m+\tfrac{1}{2})(m+1)!\ c_m\Gamma(\tfrac{1}{2})=\Gamma(m+\tfrac{1}{2})t^{2m+1}[(m+1)H_m-(m+\tfrac{1}{2})t^2H_{m+1}]. \quad (10)$$

Now the functions H_m satisfy the recurrence relation

$$(2m+1)^2t^4H_{m+1}=4m(m+1)\ [(1+t^4)H_m-H_{m-1}] \quad (11)$$

which is easy to check by replacing the hypergeometric functions by their Maclaurin's series in t^4.

Consider now four functions H_{s-1}, H_s, H_{s+1} and H_{s+2}. They satisfy five linear equations: three of type (10) for $m=s-1$, s, $s+1$; and two of type (11) for $m=s$, $s+1$. Therefore, the determinant of this system of five homogeneous equations with five unknowns H_{s-1}, H_s, H_{s+1}, H_{s+2}, and $\rho=1$ vanishes identically. Expanding it we obtain (6).

Using (6), we need the values of the first two coefficients c_0 and c_1, the other c_n's for $n\geq 2$ being computed by (6). On the other hand, (5) proves that c_n is a linear combination of complete elliptic integrals $K=K(k)$ and $E=E(k)$ with $k=\text{Sin}\ 2\theta$

$$c_n=\alpha_nE-\beta_nK, \qquad (n\geq 0) \quad (12)$$

α_n and β_n being rational functions of k.

Integrating by parts the integral in (5) and applying the relation $T_{2m+1}(x)+T_{2m-1}(x)=2xT_{2m}(x)$, we prove that

$$(2m+1)c_m-(2m-1)c_{m-1}=2kG_m(k) \qquad (m\geq 1) \quad (13)$$

where $\pi G_m(k)=\int_0^1(1-x^2)^{\frac{1}{2}}T_{2m}(x)dx/(1-k^2x^2)^{\frac{1}{2}}$,

while

$$c_0=4kG_0(k)=4(E-k'^2K)/k\pi. \qquad (k'^2=1-k^2) \quad (14)$$

Expressing $G_1(k)$ in terms of E and K and applying (13) with $m=1$, we have also

$$c_1=4[(1+7k'^2)E-k'^2(5+3k'^2)K]/9k^3\pi. \quad (15)$$

Thus, comparing (12) for $n=0$, 1 with (14) and (15): $\alpha_0=4/k\pi$; $\beta_0=4k'^2/k\pi$; $\alpha_1=4(1+7k'^2)/9k^3\pi$ and $\beta_1=4k'^2(5+3k'^2)/9k^3\pi$. But α_m, β_m verify the same recurrence relation (6) which holds for c_m and they now can be computed recursively.

We see that the computation of c_n's is reduced to that of K and E, once $k=\text{Sin}\ 2\theta$ is known. In the last section of this paper we compute c_n's for $\theta=\pi/8$, $k=2^{-\frac{1}{2}}$.

Maehly's method

In this method the rational approximation $R(x)$ to $f(x)$ is deduced from the Chebyshev expansion of $f(x)$ as follows. Denoting the Fourier coefficients of $f(x)$ by f_n and introducing $M+N+1$ unknown parameters a_m, $0\leq m\leq M$, and b_m, $1\leq m\leq N$, with $b_0=1$, we form the function $H(x)$ and expand it into its Chebyshev series:

$$H(x)=\Big[\sum_{m=0}^{N}b_mT_m(x)\Big]\Big[\sum_{0}^{\infty}f_nT_n(x)\Big]-\sum_{m=0}^{M}a_mT_m(x)=\sum_{m=0}^{\infty}h_mT_m(x).$$

Then, the choice of a_m, b_m is fixed by letting $h_m=0$ for $0\leq m\leq M+N$. These $M+N+1$ conditions $h_m=0$ form a system of linear equations satisfied by $M+N+1$ unknowns a_m, b_m. The explicit expressions of h_m in terms of a_m, b_m and f_m are obtained by performing the multiplications and replacing the products $2T_mT_n$ by the sums $T_{m+n}+T_{m-n}$.

Thus, the expansion of $H(x)$ begins with the term $A_0T_{M+N+1}(x)$, where $A_0=h_{M+N+1}$, and the rational function

$$R(x)=\sum_{0}^{M}a_mT_m(x)\ /\ \sum_{m=0}^{N}b_mT_m(x) \qquad (b_0=1)$$

is the desired approximation to $f(x)$. Its accuracy depends on the choice of M and N. More economical (in the sense

of number of operations for the same accuracy) are the cases with $M=N-1$ and $M=N$. The range of x is $(-1, +1)$ and it is possible to choose any range for the argument of $f(z)$, letting $z=kx$ and expanding $f(kx)$ as a function of x.

Applying this idea to an odd function

$$f(kx)=\sum_{n=0}^{\infty} c_n(k)T_{2n+1}(x),$$

we form an odd function $H(x)$ also:

$$H(x)=\Big[\sum_{0}^{N} b_m T_{2m}(x)\Big]\Big[\sum_{0}^{\infty} c_n T_{2n+1}(x)\Big]-\sum_{0}^{M} a_m T_{2m+1}(x).$$

Now the system $h_m=0$ is as follows:

$$\sum_{s=1}^{N}(c_{m-s}+c_{m+s})b_s+2c_m=0 \qquad (M+1\leq m\leq M+N) \quad (16)$$

$$a_m=c_m+\tfrac{1}{2}\sum_{s=1}^{N}(c_{m-s}+c_{m+s})b_s. \qquad (0\leq m\leq M) \quad (17)$$

Here by definition $c_{-s}=c_{s-1}$, if $s\geq 1$. The absolute error $R(x)-f(kx)$ of the approximation

$$f(kx)\sim R(x)=\sum_{0}^{M} a_m T_{2m+1}(x)/\sum_{0}^{N} b_m T_{2m}(x)$$

is of the order of the first neglected term $A_0T_{2M+2N+3}(x)$ and, letting $M+N+1=\mu$,

$$A_0=\tfrac{1}{2}\sum_{s=1}^{N}(c_{\mu-s}+c_{\mu+s})b_s+c_\mu. \qquad (\mu=M+N+1) \quad (18)$$

The c_n's decrease when n increases and the rate of decrease depends on the choice of k: the smaller the k, the faster the c_n's decrease. For a given k and for a prescribed accuracy, the order of magnitude of A_0 should be studied for various choices of M and N. In general, A_0 decreases when these parameters increase, and it is desirable to choose the least values of N and $M=N-1$ (or $M=N$) compatible with the required accuracy. To know the value of A_0 for given values of M and N, it is not necessary to solve (16) and use (18). Eliminating b_m's from (16) and (18), we can express A_0 as a ratio D/Δ of two determinants, Δ being the principal minor of D. Their elements, except those in the last column of D, are the sums $c_{m-k}+c_{m+k}$. Since the c_n's decrease very rapidly, omitting the second terms c_{m+k} in these sums, a good approximation D^*/Δ^* to A_0 is easily computed. Denoting the elements of D^* by d_{ij}^*, we have $d_{ij}^*=c_{i+j+M-N-1}$. $\qquad (1\leq i,j\leq N+1)$

In the case $M=N=2$ studied in the last section, we have for instance

$$A_0\begin{vmatrix} c_1 & c_2 \\ c_2 & c_3 \end{vmatrix} \approx \begin{vmatrix} c_1 & c_2 & c_3 \\ c_2 & c_3 & c_4 \\ c_3 & c_4 & c_5 \end{vmatrix}. \qquad (19)$$

Given M and N, b_m's are computed solving (16), and then (17) gives the a_m's. Replacing T_{2m} and T_{2m+1} by their explicit expressions, the rational function $R(x)$ is then expanded into a continued fraction of the type

$$R(x)=x\Big(\alpha_0+\sum_{1}^{\nu}\frac{\alpha_m|}{|\beta_m+x^2}\Big),$$

where $\nu=N$, if $M=N-1$ or $M=N$, and $\alpha_0=0$ for $M=N-1$, but $\alpha_0\neq 0$, if $M=N$. Finally, replacing x by N/k, the approximation is as follows

$$f(N)\sim R(N/k)=N\Big[A_0+\sum_{1}^{\nu}\frac{A_m|}{|B_m+N^2}\Big].$$

Computing the upper bound B of relative error

$A_0|T_{2m+2n+3}(x)|/f(kx)$, where $f(kx)$ vanishes for $x=0$, we use the inequality $|T_n(x)|\leq n|x|$.

Thus the upper bound B is as follows

$B=(2M+2N+3)A_0|x/f(kx)|$.

Subroutine for Arcsin N, $0\leq N\leq 1$

If $\theta=\pi/8$, $k=2^{-\frac{1}{2}}=k'$, (6), (14) and (15) are as follows

$$m(2m+3)^2c_{m+1}=(2m+1)(12m^2+12m+1)c_m-(m+1)(2m-1)^2c_{m-1} \quad (6)^*$$

$$\pi c_0=2\sqrt{2}(2E-K) \quad (14)^*$$

$$\pi c_1=2\sqrt{2}(2E-13K/9). \quad (15)^*$$

The classical Legendre's relation[3]

$KE'+EK'-KK'=\pi/2$

yields in our case $2E-K=\pi/2K$, so that

$c_0=\sqrt{2}/K$; $c_1=c_0-8K\sqrt{2}/9\pi$.

To find $K=\Gamma^2(\frac{1}{4})/4\Gamma(\frac{1}{2})$, we will use the Gaussian form of Landen's Transformation[4]:

$$K_n=\int_0^{\pi/2}(\alpha_n^2\,\mathrm{Cos}^2\psi+\beta^2_n\,\mathrm{Sin}^2\psi)^{-\frac{1}{2}}d\psi=$$

$$\int_0^{\pi/2}(\alpha^2_{n+1}\,\mathrm{Cos}^2\psi+\beta^2_{n+1}\,\mathrm{Sin}^2\psi)^{-\frac{1}{2}}d\psi=K_{n+1}$$

with $\alpha_{n+1}=\frac{1}{2}(\alpha_n+\beta_n)$ and $\beta_{n+1}=(\alpha_n\beta_n)^{\frac{1}{2}}$.

In our case $K=K_0$ with $\alpha_0=1$ and $\beta_0=\sqrt{2}/2$. Tabulating the results of computation, we find that

$|\beta_n-\alpha_n|<5\times 10^{-16}$ for $n=4$:

n	$10^{15}\alpha_n$					$10^{15}\beta_n$				
1	853	553	390	593	274	840	896	415	253	714
2	847	224	902	923	494	847	201	266	746	892
3	847	213	084	835	193	847	213	084	752	766
4	847	213	084	793	980	847	213	084	793	980

Approximating $\lim_{n=\infty}\alpha_n=\lim_{n=\infty}\beta_n$ by $\alpha=\alpha_4=\beta_4$, we have

$K\approx\pi/2\alpha$, so that $c_0=2\alpha\sqrt{2}/\pi$ and $c_1=c_0-4\sqrt{2}/9\alpha$; that is,

$c_0=0.762\,759\,763\,501\,814$; $c_4=0.000\,018\,691\,074\,973$

$c_1=0.020\,869\,237\,569\,305$; $c_5=0.000\,002\,354\,064\,470$ (20)

$c_2=0.001\,586\,931\,627\,771$; $c_6=0.000\,000\,312\,577\,010$

$c_3=0.000\,160\,822\,752\,706$; $c_7=0.000\,000\,043\,092\,937$

the last six c_n's being obtained by $(6)^*$.

Another way to compute these numbers would be to use (12), starting with values of α_0, α_1, β_0, β_1, K and E. Here $\alpha_0=\alpha_1=2$, $\beta_0=1$, $\beta_1=13/9$ as well as

$K=1.85407\ 46772\ 99357$

$E=1.35064\ 38810\ 46669,$

E being computed with the aid of Legendre's relation. With the aid of (6)* we obtained $\alpha_2=5.84$; $\alpha_3=21.2$; $\alpha_4=2{,}326/27$; $\alpha_5=1{,}126/3$; $\alpha_6=869{,}148.4/507$; $\alpha_7=947{,}825.68/117$; $\beta_2=12.76/3$; $\beta_3=2{,}270.2/147$; $\beta_4=11{,}861/189$; $\beta_5=231{,}587/847$; $\beta_6=16{,}250{,}885.8/13{,}013$ and $\beta_7=17{,}721{,}952.76/3{,}003$.

The values of c_n, $0\leq n\leq 7$, recomputed by (12) agree completely with (20). This check was necessary because even small errors in the values of c_n can spoil the final results completely.

In our case $M=N=2$, and solving (16), that is

$$(c_2+c_4)b_1+(c_1+c_5)b_2+2c_3=0$$
$$(c_3+c_5)b_1+(c_2+c_6)b_2+2c_4=0,$$

we obtain $b_1=-0.31460\ 68409$ and $b_2=0.00879\ 15854$.

Now (17) gives the values of $a_0=0.69359\ 09539$, $a_1=-0.09601\ 14713$ and $a_2=0.00163\ 18472$.

Replacing in $R(x)$ the polynomials T_n by their expressions, we have

$$R(x)=x(\alpha_0+\alpha_1x^2+\alpha_2x^4)/(\beta_0+\beta_1x^2+\beta_2x^4)$$

with $\alpha_0=a_0-3a_1+5a_2=0.93578\ 46039$, $\alpha_1=4(a_1-5a_2)=-0.41668\ 28293$, $\alpha_2=16a_2=0.02610\ 95550$, $\beta_0=1-b_1+b_2=1.32339\ 84263$, $\beta_1=2b_1-8b_2=-0.69954\ 63652$ and $\beta_2=8b_2=0.07033\ 26834$.

Substituting $N\sqrt{2}$ for x, we find the following form $f(N)$ of our approximation to Arcsin N in the interval $0<N<2^{-\frac{1}{2}}$ Arcsin $N\sim f(N)=N(mN^4+nN^2+p)/(qN^4+rN^2+s)$ with

$$m=\alpha_2\sqrt{2}=0.03692\ 45063,\quad q=\beta_2=0.07033\ 26834$$

$$n=\alpha_1/\sqrt{2}=-0.29463\ 92542,\quad r=\beta_1/2=-0.34977\ 31826$$

$$p=\alpha_0/2\sqrt{2}=0.33084\ 98196,\quad s=\beta_0/4=0.33084\ 96066.$$

Since $\lim_{N=0}(N^{-1}\text{Arcsin } N)=1$, the ratio p/s is a check. Here $p/s=1.000\ 000\ 64\ldots$. The maximum of relative error is at $N=0$ and it is equal to 6.4×10^{-7}:

$$\lim_{N=0}[f(N)/\text{Arcsin } N-1]=p/s-1=6.4\times10^{-7}.$$

Transforming our rational function into a continued fraction we obtain the final formula

$$f(N)=N\left(A_0+\frac{A_1|}{|B_1-N^2}+\frac{-A_2|}{|B_2-N^2}\right). \tag{21}$$

The five constants to store are

$A_0=m/q=0.52499\ 78317$; $A_1=(rA_0-n)/q=1.57834\ 2904$; $B_2=(p-sA_0)/qA_1=1.41569\ 02913$; $B_1=-(B_2+r/q)=3.55743\ 40883$; $A_2=B_1B_2-s/q=0.33215\ 85891$.

Thus, Arcsin N is computed in only four operations, two multiplications and two divisions, the number of correct significant digits being equal to six. Example: Arcsin $(2^{-\frac{1}{2}})=\pi/4=0.785\ 398\ 16\ldots$ and $f(2^{-\frac{1}{2}})=0.785\ 398\ 04\ldots$, so that the relative error is less in absolute value than 2×10^{-7}.

Another equivalent form of $f(N)$ would be

$$f(N)=N\left(A_0+\frac{\eta_1}{\xi_1-N^2}+\frac{\eta_2}{\xi_2-N^2}\right). \tag{22}$$

where $\xi_1=1.270451499$, $\xi_2=3.702672882$, $\eta_1=0.09425018578$, $\eta_2=1.484093006$.

The detailed study of relative error reveals that it decreases from 6.4×10^{-7} to 3.8×10^{-7} when N increases from 0 to 0.1. For $0.1\leq N\leq2^{-\frac{1}{2}}$ the relative error remains in absolute value less than 3.8×10^{-7}. Thus, for some exceptional values of N in the range (0; 0.1), the sixth significant digit in $f(N)$ may exceed the corresponding digit in the exact value of Arcsin N by one unit. In our opinion, it is not worthwhile to complicate the subroutine to avoid these few exceptional cases. If it is to be done, however, approximation (21) could be replaced in the range $0<N\leq0.1$ by the sum of the first three terms of Maclaurin's series, that is by $N+N^3/6+3N^5/40$, which yields the first seven correct significant digits, if $N\leq0.1$.

The same approximation (21) is used in the interval $2^{-\frac{1}{2}}\leq N\leq(2+\sqrt{2})^{\frac{1}{2}}/2=\text{Sin }(3\pi/8)$. To compute Arcsin N in this range, we first form $N^*=2N^2-1$ and then compute $f(N^*)$ using (21). Using (2), we find that

$$\text{Arcsin } N\approx\pi/4+\tfrac{1}{2}f(N^*).\qquad(N^*=2N^2-1) \tag{23}$$

In the range $\text{Sin }(3\pi/8)<N<1$, the value of Arcsin N exceeds $3\pi/8=1.178\ldots$ so that the first six significant digits are correct if the absolute error is less than 5×10^{-6}. In this last interval we use relation (3), approximating Arcsin $[(1-N)/2]^{\frac{1}{2}}$ by the sum of the first two terms of the Chebyshev expansion (1) with $\text{Sin }2\theta\geq\text{ max. }[(1-N)/2]^{\frac{1}{2}}=\text{Sin }11°15'$, that is, with $\theta\geq5°37'.5$ and $\tan\theta\geq0.0985$.

Computing the coefficients c_n in (1), we take $t=\tan\theta=0.1$. Rounding off increases slightly the range of validity of (1), simplifying considerably the numerical computations. Using (10) for $m=0$, 1 and 2, we need the values of $H_m=H_m(0.1)$ for $m\leq3$. They are as follows:
$H_0=1.\ 000\ 025\ 001\ 406$; $H_1=1.\ 000\ 037\ 502\ 344$;
$H_2=1.\ 000\ 041\ 669\ 401$; $H_3=1.\ 000\ 043\ 752\ 953$.

Therefore $c_0=0.\ 199\ 004\ 962\ 779$, $c_1=0.\ 000\ 330\ 845\ 730$ and $c_2=0.\ 000\ 001\ 487\ 562$

and in general $0<c_{m+1}<c_m/100$. Now

$$\sum_{m=2}^{\infty}c_m|T_{2m+1}(x)|<\sum_{m=2}^{\infty}c_m<c_2/99\approx1.5\times10^{-6}.$$

Since $\text{Sin }2\theta=2t/(1+t^2)=2/10.1$, we have

$$\text{Arcsin }(2x/10.1)\approx c_0T_1(x)+c_1T_3(x) \tag{24}$$

with an absolute error less than 2×10^{-6}. The right-hand member of (24) can be written as $x(c_0-3c_1+4c_1x^2)=x(a+bx^2)$ with $a=0.198\ 012\ 4256$ and $b=0.001\ 323\ 3829$. Substituting $x=5.05[(1-N)/2]^{\frac{1}{2}}$, equation (24) takes the form

$$\phi(N)=[(1-N)/2]^{\frac{1}{2}}(A-BN) \tag{25}$$

where $A=1.085\ 180\ 421$, $B=0.085\ 217\ 6716$.

The approximation $\phi(N)$ to Arcsin $[(1-N)/2]^{\frac{1}{2}}$ is not very accurate in itself, but it is sufficiently correct to insure the required accuracy of the first six digits of Arcsin N as given by relation (3)

$$\text{Arcsin } N\simeq\pi/2-2\phi(N). \tag{3*}$$

Computation of $\phi(N)$ with the aid of (25) is now reduced to the extraction of a square root. It can be performed in two divisions and one multiplication as follows. Taking R_1 defined by

$$R_1=\alpha\left(\beta+f-\frac{\gamma}{\delta+f}\right) \tag{26}$$

as a first approximation to $\sqrt{f}$, $0.25\leq f\leq 1$, and then applying the so-called Newton's method[5] only once, we obtain an approximation

$$R_2=\tfrac{1}{2}(R_1+f/R_1) \tag{27}$$

to $\sqrt{f}$. This yields an approximation to $[(1-N)/2]^{\frac{1}{2}}$ with the first eight correct significant digits, if the numerical coefficients in (26) are:

$\alpha=0.3343\ 1261$; $\beta=2.7691\ 3454$; $\gamma=1.1903\ 1245$; $\delta=0.5316\ 4106$.

Using the foregoing values of α, β, γ and δ, the absolute error $e=e(f)=|f^{\frac{1}{2}}-R_1|$ of the first approximation does not exceed 3×10^{-4}: $e(f)\leq3\times10^{-4}$. Then, by a well-known property of Heron's method, the absolute error $E(f)=|R_2-f^{\frac{1}{2}}|$ of the last approximation R_2 is equal at most to $e^2/2R_1$ so that $E\leq e^2\leq9\times10^{-8}$, because $R_1\sim f^{\frac{1}{2}}\geq0.5$.

Here f is defined by $(1-N)/2=2^{-2h}f$ with $0.25\leq f<1$. Since $(1-N)/2\leq$ Sin^2 11°15′$=0.03806\ldots$, the positive integer h is equal at least to 2: $h\geq2$. But $[(1-N)/2]^{\frac{1}{2}}=2^{-h}f^{\frac{1}{2}}$ so that the absolute error of the approximation $2^{-h}R_2$ to $[(1-N)/2]^{\frac{1}{2}}$ is at least four times less than E, that is, less than 2.25×10^{-8} which proves our statement.

The choice of the range (0.25; 1) for f presupposes that a binary machine is used. For a decimal machine which cannot perform the extraction of square roots directly (as can the IBM 610) we let $(1-N)/2=(10)^{-2h}f$ with $0.01\leq f\leq1$; $h\geq0$. Here two sets of constants α, β, γ and δ are needed to insure the required accuracy: one for the interval $0.01\leq f\leq0.1$; another for $0.1\leq f\leq1$. Using formula (26) again where

$0.01\leq f\leq0.1$		$0.1\leq f\leq1$
1.248 114 30	α	0.394 688 40
0.194 941 49	β	1.949 414 90
0.005 523 71	γ	0.552 371 19
0.034 278 20	δ	0.342 782 00

the relative error in both subintervals is less than 3×10^{-3}. Applying (27), we reduce the relative error to no more than 4.5×10^{-6} and this is also the upper bound for the relative error in computing $[(1-N)/2]^{\frac{1}{2}}=10^{-h}f^{\frac{1}{2}}$. Multiplying 4.5×10^{-6} by the upper bound Sin 11°15′$\cong0.195\ \ldots$ of $[(1-N)/2]^{\frac{1}{2}}$, we have the desired result

$$|10^{-h}R_2-[(1-N)/2^{\frac{1}{2}}|<10^{-6}.$$

Five multiplications and divisions are performed using expression (25); six constants are needed for a binary and eleven for a decimal machine. The total number of stored constants is therefore thirteen for a binary machine and eighteen for a decimal machine, the additional constants being necessary to locate N.

References

1. Monthly Progress Report, The Institute for Advanced Study, Princeton, N. J., October 1956.
2. "Higher Transcendental Functions," Bateman Manuscript Project, Vol. 1, p. 113, form. (31), Case a=b.
3. Op. cit. (2), Vol. II, page 320, form. (15).
4. Op. cit. (2), Vol. II, page 316, last row.
5. Misnomer. Method actually originated by Greek mathematician, Heron of Alexandria.

Received March 28, 1958