N71-36946

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

Technical Memorandum 33-460

High Precision Moments for the Fourier Coefficients of Arbitrary Functions

Shalhav Zohar

JET PROPULSION LABORATORY
CALIFORNIA INSTITUTE OF TECHNOLOGY
PASADENA, CALIFORNIA

October 1, 1971

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

Technical Memorandum 33-460

High Precision Moments for the Fourier Coefficients of Arbitrary Functions

Shalhav Zohar

JET PROPULSION LABORATORY
CALIFORNIA INSTITUTE OF TECHNOLOGY
PASADENA, CALIFORNIA

October 1, 1971

Prepared Under Contract No. NAS 7-100
National Aeronautics and Space Administration

PREFACE

The work described in this report was performed by the Telecommunications Division of the Jet Propulsion Laboratory.

CONTENTS

High Precision Moments for the Fourier Coefficients of Arbitrary Functions

by Shalhav Zohar

ABSTRACT

Evaluation of the integrals

$$\int_0^{\frac{\pi}{2}} \xi^n \begin{matrix}\sin\\ \cos\end{matrix}\, k\xi \, d\xi \qquad (n,k \text{ positive integers})$$

is considered in detail. The difficulties in their direct evaluation are analyzed and overcome in a method capable of very high precision. A detailed error analysis yields a reliable error bound for the results obtained this way.

The method has been applied in computing a table of parameters allowing the evaluation of the above integrals for $n=2,3,\ldots,201$; $k=1,2,\ldots,200$, with a relative error of $2^{-132}(\approx 2\cdot 10^{-40})$. Two types of tests indicate that the actual errors are well below this upper bound.

At the other extreme, very simple coarse estimates of the integrals are also obtained and their error bounds established.

The computed parameters are also shown to be related to the (high order) derivatives of band-limited functions and to the incomplete gamma function.

1. Introduction

The construction of quadrature formulae for the evaluation of the Fourier coefficients of arbitrary functions leads to the following moments

$$(1.1) \qquad \int_0^{2\pi} \xi^n \cos m\xi \, d\xi$$

$$(1.2) \qquad \int_0^{2\pi} \xi^n \sin m\xi \, d\xi \qquad \text{(n,m positive integers)}$$

When high precision coefficients are required, precise moments have to be obtained for n ranging up to large values. But in this case (more specifically: $n > (2\pi e)m$), serious difficulties beset their computation.

An interesting solution to the problem was suggested by W. Gautschi [9] who proposed to evaluate the related integrals

$$(1.3) \qquad \int_0^{2\pi} \xi^n (1 - \cos m\xi) \, d\xi$$

$$(1.4) \qquad \int_0^{2\pi} \xi^n (1 - \sin m\xi) \, d\xi$$

via a recursive scheme. This approach, modified by some refinements, is the subject of this paper. In order to simplify the computations, we shall compute the following multiples of the above integrals:

(1.5a) $$\frac{n+1}{(2\pi)^{n+1}} \int_0^{2\pi} \xi^n \cos m\xi \, d\xi = \theta_{n+1,4m}$$

(1.5b) $$\frac{m}{(2\pi)^n} \int_0^{2\pi} \xi^n \sin m\xi \, d\xi = - \eta_{n,4m}$$

(1.5c) $$\frac{n+1}{(2\pi)^{n+1}} \int_0^{2\pi} \xi^n (1 - \cos m\xi) \, d\xi = \eta_{n+1,4m}$$

(1.5d) $$\frac{n+1}{(2\pi)^{n+1}} \int_0^{2\pi} \xi^n (1 - \sin m\xi) \, d\xi = 1 + \frac{n+1}{2\pi m} \eta_{n,4m}$$

All four integrals are computable by one simple recurrence

(1.6) $$\eta_{nk} = 1 - \frac{n(n-1)}{(k\,\frac{\pi}{2})^2} \eta_{n-2,k}$$

and the relation

(1.7) $$\theta_{nk} = 1 - \eta_{nk}$$

Furthermore, in addition to their role in (1.5), the η_{nk} parameters introduced here, allow the evaluation of integrals in which the upper bound is an integral multiple of $\frac{\pi}{2}$ rather than 2π. This means that in the formulation leading from the arbitrary function to the moments (1.1), (1.2), we do not have to work with a single approximation valid for the full period but may, instead, use four different approximations getting a better fit.

Most of this paper is devoted to the study of the related continuous function $\eta_n(x)$ ($\eta_{nk} = \eta_n(k\frac{\pi}{2})$). Our main goal in this study is to provide the foundation necessary for the computation of η_{nk} with a predictable high precision. A table of η_{nk} ($n = 2,3,\ldots,201$; $k = 1,2,\ldots,$ 200) with a relative error bound of 2^{-132} is one outcome of this undertaking. Note that most of this precision can be carried over to the moments themselves since the only seeming obstacle here is the exact computation of powers of π. As π is available to more than 10^4 decimal digits [11], the construction of a π^n table with precision compatible with that of η_{nk}, is certainly no problem.

We proceed now with a brief description of the layout of the paper. Section 2 is devoted to the derivation of the generalized version of the integrals (1.5a), (1.5b). The function $\eta_n(x)$ is introduced here.

Section 3 establishes the recurrence relation satisfied by $\eta_n(x)$ and considers its application in computing $\eta_n(x)$. We show that both forward and backward recursions are necessary. The discussions here are preliminary, pending the establishment of various important properties of

$\eta_n(x)$ in section 4. Section 5 resumes the discussion of the computation of $\eta_n(x)$. The dividing line between forward and backward recursions is considered here in detail.

Section 6 tackles problems associated with the initiation of the backward recursion. It provides detailed design information to determine the first term of the backward recursion from a prescribed error bound.

Section 7 considers the effect of truncation errors on the precision of the computed $\eta_n(x)$. An easily computable error bound is the main result here.

Section 8 describes a specific implementation of the algorithm developed in the earlier sections. The result is the η_{nk} table already mentioned. A verification of the computed values is also considered here.

Section 9 takes up various other applications of $\eta_n(x)$. It consists of three parts. In part a, the main integration results established in section 2 are extended to integrals of type (1.3), (1.4). Our main concern here is in verifying the precision in this case. Part b shows the relationship of $\eta_n(x)$ to the derivatives of band-limited functions. Part c considers the connection between $\eta_n(x)$ and the incomplete gamma function.

Section 10 establishes a relationship between η_{nk}, Riemann's zeta function, and various other related functions. The availability of high precision tables of these functions makes it possible to exploit the established relationship in a test of certain computed η_{nk} values. Results of these tests are presented.

A few topics whose analysis had been deferred in the interest of a smoother development of the subject, are discussed in section 11, the last one. In particular, the failure of direct computation of the integrals is analyzed in detail.

2. The Integration Formulae

Though the integrals we are concerned with involve trigonometric functions, it is simpler to deal initially with the exponential function. Our starting point is the following closed form integration formula [2].

$$\int \xi^n e^{i\xi} d\xi = e^{i\xi} \sum_{k=0}^{n} i^{k-1} \frac{n!}{(n-k)!} \xi^{n-k} \tag{2.1}$$

$$= i^{n-1} n! \, e^{i\xi} \sum_{r=0}^{n} \frac{(-i\xi)^r}{r!} \tag{2.2}$$

We have here a partial sum of the $e^{-i\xi}$ series. Bearing this in mind we rephrase (2.2) using the following notation

$$g_n(x) = \frac{x^n}{n!} \tag{2.3}$$

and getting

$$\int \xi^n e^{i\xi} d\xi = i^{n-1} n! \, e^{i\xi} \sum_{r=0}^{n} g_r(-i\xi) \tag{2.4}$$

$$= i^{n-1} n! \, e^{i\xi} \left[e^{-i\xi} - \sum_{r=n+1}^{\infty} g_r(-i\xi) \right] \tag{2.5}$$

Note the modification from a finite summation in (2.4) to an infinite one in (2.5). The motivation prompting this step will be dealt with in section 11.

Using (2.5) we obtain the desired definite integral

$$(2.6) \qquad I_n(x) \equiv \int_0^x \xi^n e^{i\xi}\, d\xi = i^{n+1}\, n!\, e^{ix} \sum_{r=n+1}^{\infty} g_r(-ix)$$

$$(2.7) \qquad = x^{n+1} e^{ix} \psi_n(x)$$

where

$$(2.8) \qquad \psi_n(x) = \frac{1}{(-ix)} \frac{\sum_{r=n+1}^{\infty} g_r(-ix)}{g_n(-ix)}$$

$\psi_n(x)$ is a complex valued function of the real argument x. Our main concern in the following sections will involve the closely related real function $\eta_n(x)$ which we proceed to introduce now.

Our first step is the segregation of the sum in (2.8) into two separate sums as shown below

$$(2.9) \qquad (-ix)\, \psi_n(x) = \left\{ \frac{\sum_{s=1}^{\infty} g_{n+2s}(-ix)}{g_n(-ix)} + \frac{\sum_{s=1}^{\infty} g_{(n-1)+2s}(-ix)}{g_n(-ix)} \right\}$$

Being guided by the first term inside the brackets, we introduce now

$$(2.10) \qquad \eta_n(x) = -\frac{\sum_{s=1}^{\infty} g_{n+2s}(-ix)}{g_n(-ix)}$$

$$(2.11) \qquad = \frac{\sum_{s=1}^{\infty} (-1)^{s+1} g_{n+2s}(x)}{g_n(x)}$$

Noting that (2.3) implies

$$(2.12) \qquad g_n(z) = \frac{z}{n} g_{n-1}(z)$$

we see that the second term in (2.9) is expressible in terms of $\eta_{n-1}(x)$ and thus

$$(2.13) \qquad \psi_n(x) = \frac{n}{x^2} \eta_{n-1}(x) - i\frac{\eta_n(x)}{x}$$

The formulation of $\eta_n(x)$ in (2.11) clearly shows it to be a real function of x. It also shows its relationship to the infinite series expansions of sin x and cos x. Thus, for n even, $-\eta_n(x)$ is obtained by striking out the first $[\frac{n+2}{2}]$ terms of the cos x series (sin x series for n odd) and dividing the remainder by the last eliminated term.

In some of the subsequent developments, a more explicit formulation of $\eta_n(x)$ is required. This is obtained by substituting (2.3) in (2.11)

$$(2.14) \qquad \eta_n(x) = \sum_{k=1}^{\infty} (-1)^{k+1} \frac{x^{2k}}{(n+1)_{2k}}$$

We have used here the standard notation

$$(n+1)_k = \frac{(n+k)!}{n!} \tag{2.15}$$

Note that this formulation specifically shows $\eta_n(x)$ to be an even function of x.

$$\eta_n(-x) = \eta_n(x) \tag{2.15a}$$

The integrals we have set out to evaluate can now be expressed in terms of $\eta_n(x)$. We note first that a trivial change of variable yields

$$\int_0^{\frac{x}{a}} \xi^n \, e^{ia\xi} \, d\xi = \frac{I_n(x)}{a^{n+1}} = \frac{x^{n+1}}{a^{n+1}} \, e^{ix} \, \psi_n \, (x) \tag{2.15b}$$

Hence

$$\frac{a^{n+1}}{x^{n-1}} \int_0^{\frac{x}{a}} \xi^n \cos a\xi d\xi = \mathrm{Re}\, [e^{ix} x^2 \psi_n(x)] = n\eta_{n-1}(x) \cos x + x\eta_n(x) \sin x \tag{2.16}$$

$$\frac{a^{n+1}}{x^{n-1}} \int_0^{\frac{x}{a}} \xi^n \sin a\xi d\xi = \mathrm{Im}\, [e^{ix} x^2 \psi_n(x)] = n\eta_{n-1}(x) \sin x - x\eta_n(x) \cos x \tag{2.17}$$

Eqns. (2.16), (2.17) are valid for any real x. However, we are particularly interested in those cases in which one of the terms on the right of (2.16), (2.17) vanishes. The precision of such integrals will directly depend on the precision of $\eta_n(x)$. Note that for arbitrary x this is not the case since loss of significance due to "cancellation" is certain to set in for some n, x values, no matter how precise $\eta_n(x)$ is.

In view of the above, the computed integrals will be those in which x is constrained as follows

$$x = k\frac{\pi}{2} \qquad \text{(k integer)} \tag{2.17a}$$

From (2.16), (2.17) it is obvious that under these conditions we will be dealing with

$$\eta_n(k\frac{\pi}{2}) \qquad \text{(k integer)}$$

Hence, the notation

$$\eta_{nk} = \eta_n\,(k\frac{\pi}{2}) \qquad \text{(k integer)} \tag{2.18}$$

Applying the constraint (2.17a) to (2.16), (2.17) we obtain our main results:

$$\frac{a^{n+1}}{(k\frac{\pi}{2})^{n-1}} \int_0^{\frac{k}{a}\frac{\pi}{2}} \xi^n \cos a\xi d\xi = \begin{cases} n\,\eta_{n-1,k} & (k \bmod 4 = 0) \\ (k\frac{\pi}{2})\,\eta_{nk} & (k \bmod 4 = 1) \\ -\,n\eta_{n-1,k} & (k \bmod 4 = 2) \\ -(k\frac{\pi}{2})\,\eta_{nk} & (k \bmod 4 = 3) \end{cases} \tag{2.19}$$

$$
(2.20)\quad \frac{a^{n+1}}{(k\,\frac{\pi}{2})^{n-1}} \int_0^{\frac{k}{a}\frac{\pi}{2}} \xi^n \sin a\xi d\xi = \begin{cases} -(k\,\frac{\pi}{2})\,\eta_{nk} & (k \bmod 4 = 0) \\ n\eta_{n-1,k} & (k \bmod 4 = 1) \\ (k\,\frac{\pi}{2})\,\eta_{nk} & (k \bmod 4 = 2) \\ -n\,\eta_{n-1,k} & (k \bmod 4 = 3) \end{cases}
$$

Note that (1.5b) is a special case of (2.20) with

$$
a = \text{integer} = m; \quad k = 4m
$$

The relation of (2.19) to (1.5a) will become clear later (section 5).

With the establishment of (2.19), (2.20), we have transformed our problem to that of evaluating the η_{nk} parameters. It turns out, however, to be more convenient to deal with the continuous function $\eta_n(x)$. This will occupy us in the following sections.

3. Preliminary Considerations in Computing $\eta_n(x)$

For low values of n, $\eta_n(x)$ appears to be quite elementary. To see this we recall that the infinite summation appearing in (2.11) is a direct result of the manipulation transforming (2.4) into (2.5). For the present discussion we now manipulate (2.11) in the opposite direction getting the following finite summation formulation:

$$(3.1)\qquad \eta_n(x) = \begin{cases} (-1)^{[\frac{n+2}{2}]} \cdot \dfrac{\cos x - \sum_{k=0}^{[\frac{n}{2}]} (-1)^k g_{2k}(x)}{g_n(x)} & (n \text{ even}) \\[2ex] (-1)^{[\frac{n+2}{2}]} \cdot \dfrac{\sin x - \sum_{k=0}^{[\frac{n}{2}]} (-1)^k g_{2k+1}(x)}{g_n(x)} & (n \text{ odd}) \end{cases}$$

Direct substitution yields

$$(3.2)\qquad \eta_0(x) = 1 - \cos x$$

$$(3.3)\qquad \eta_1(x) = 1 - \frac{\sin x}{x}$$

$$(3.4)\qquad \eta_2(x) = 1 - \frac{2}{x^2} + \frac{2}{x^2} \cos x = 1 - \left(\frac{\sin \frac{x}{2}}{\frac{x}{2}} \right)^2$$

$$(3.5)\qquad \eta_3(x) = 1 - \frac{6}{x^2} + \frac{6}{x^2} \frac{\sin x}{x}$$

Note that (3.1) specifies $\eta_n(x)$ as an alternating series. This raises the specter of loss of significance due to "cancellation". As a matter of fact, it will be shown in section 11 that direct application of (3.1) does, indeed, lead to very severe "cancellation" for certain n, x combinations. We conclude therefore that (3.1) has to be rejected as a general method of computing $\eta_n(x)$.

With direct computation ruled out, we consider now the possibility of evaluating $\eta_n(x)$ via a recurrence relation. Such a relation is most easily obtained from (2.14) as follows

$$(3.7) \qquad 1 - \eta_n(x) = 1 - \sum_{k=1}^{\infty} (-1)^{k+1} \frac{x^{2k}}{(n+1)_{2k}} = \sum_{k=0}^{\infty} (-1)^{k} \frac{x^{2k}}{(n+1)_{2k}}$$

Let us substitute now $k = r - 1$

$$(3.8) \quad \therefore 1-\eta_n(x) = \sum_{r=1}^{\infty} (-1)^{r+1} \frac{x^{2r-2}}{(n+1)_{2r-2}} = \frac{n(n-1)}{x^2} \sum_{r=1}^{\infty} (-1)^{r+1} \frac{x^{2r}}{(n+1)_{2r}} \qquad (n > 1)$$

$$= \frac{n(n-1)}{x^2} \eta_{n-2}(x) \qquad (n > 1)$$

Hence the simple recurrence formula

$$(3.9) \qquad \eta_n(x) = 1 - \frac{n(n-1)}{x^2} \eta_{n-2}(x) \qquad (n > 1)$$

Denoting

$$(3.10) \qquad \frac{n(n-1)}{x^2} = e_n(x), \qquad (n > 1)$$

We get the final formulation

$$(3.11) \qquad \eta_n(x) = 1 - e_n(x)\,\eta_{n-2}(x) \qquad (n > 1)$$

Eqns. (3.11), (3.10) are the basis for most of the subsequent developments.

In examining the suitability of (3.11) to the computation of a sequence of $\eta_n(x)$ values, we are assuming in this preliminary discussion infinite precision arithmetic and are concerned mainly with the effect of error propagation.

Assume that in computing $\eta_n(x)$ via (3.11), we have not used the exact $\eta_{n-2}(x)$ but rather an approximation to it, $\eta'_{n-2}(x)$ where

$$(3.11a) \qquad \eta'_{n-2}(x) - \eta_{n-2}(x) = \Delta\,\eta_{n-2}(x)$$

It is obvious then that $\Delta\,\eta_n(x)$, the resulting error in $\eta_n(x)$, is given by

$$(3.12) \qquad \Delta\,\eta_n(x) = -e_n(x)\;\Delta\,\eta_{n-2}(x)$$

We conclude that the stability of the recursion process depends on $e_n(x)$. Thus, as long as

$$e_n(x) < 1$$

for all n in the sequence generated, an error injected in, say, $\eta_n(x)$

will be progressively attenuated in $\eta_{n+2}(x)$, $\eta_{n+4}(x)$, etc. Under these conditions, the desired sequence could be generated by a simple stable process starting with ((3.2), (3.3))

$$\eta_0(x) = 1 - \cos x$$

$$\eta_1(x) = 1 - \frac{\sin x}{x}$$

We observe, however, that for $n > \frac{1}{2}$, $e_n(x)$ is a monotonically increasing function of n. This means that for every x there is a specific n value, $n_1(x)$, such that

$$(3.13) \qquad \begin{cases} e_n(x) < 1 & (n < n_1(x)) \\ e_n(x) > 1 & (n > n_1(x)) \end{cases}$$

Any error introduced past $\eta_{n_1}(x)$ will be progressively magnified in subsequent terms, eventually rendering the computed values completely meaningless.

The resolution of this difficulty is quite simple. Rewriting (3.11), (3.12) to obtain

$$(3.14) \qquad \eta_{n-2}(x) = \frac{1}{e_n(x)} [1 - \eta_n(x)] \qquad (n > 1)$$

$$(3.15) \qquad \Delta \eta_{n-2}(x) = - \frac{1}{e_n(x)} \Delta \eta_n(x)$$

we observe that for $e_n(x) > 1$, the recurrence (3.14) represents a stable process in the backward direction; that is, we can safely generate $\eta_{n-2}(x)$ from $\eta_n(x)$, etc. This, of course, immediately poses the dilemma of the unavailable initial value.

The solution is provided by a simple argument originally applied by J.C.P. Miller [3] to the computation of Bessel functions. The reasoning is as follows: If the generated function is known to be bounded, it does not really matter what estimate we adopt for the initial value starting the recursion, provided we start with an index n sufficiently removed from the first index of interest. Since the initial error is finite, it will take only a finite number of cycles to reduce it below any permissible error.

To justify the application of this idea to our case, we have to show that $\eta_n(x)$ is bounded. However, the design of an efficient guaranteed-precision algorithm calls for more than that. In short, we have to know more about the function we wish to compute.

It is with this in mind that we take up the study of $\eta_n(x)$ in the next section. Most of the properties derived there are subsequently applied in the design of the algorithm generating $\eta_n(x)$.

4. Properties of $\eta_n(x)$

The properties of $\eta_n(x)$ proved in this section may be divided into two groups. In the first group we establish the range of $\eta_n(x)$ and its asymptotic behavior near the boundaries of this range. In the second group we are concerned with the variation of $\eta_n(x)$ inside its range.

Our starting point is an integral definition of $\eta_n(x)$. Setting a = x in (2.15b) and applying (2.13), we get*

$$(4.1)\qquad \eta_n(x) = x \operatorname{Im} \int_0^1 \xi^n e^{ix(1-\xi)} d\xi = x \int_0^1 \xi^n \sin\,[x(1 - \xi)]d\xi$$

$$(4.2)\qquad = |x| \int_0^1 (1 - \xi)^n \sin\,(|x|\xi)d\xi$$

Consider now the above integral for $n, |x| > 0$. With $(1-\xi)^n \geq 0$, the integration over the first half cycle of $\sin |x|\xi$ yields a positive number. The second half cycle yields a negative number. However since $(1-\xi)^n$ is a monotonically decreasing function, this negative term is of smaller magnitude than the preceding positive term so that the net contribution is positive. Continuing this way, we conclude that the integral is positive, that is,

$$(4.3)\qquad \eta_n(x) > 0 \qquad\qquad (n > 0\colon x \neq 0)$$

Combining this with (3.9) yields

$$(4.4)\qquad 0 < \eta_n(x) < 1 \qquad\qquad (x \neq 0,\ n > 1)$$

Finally noting that (2.14) prescribes

$$(4.5)\qquad \eta_n(0) = 0,$$

we get **

$$(4.18)\qquad 0 \leq \eta_n(x) < 1 \qquad\qquad (n > 1)$$

(*) Throughout this paper, $\eta_n(x)$ is constrained to integral values of n. Eqn. (4.1) provides a convenient extension of the definition of $\eta_n(x)$ to arbitrary n.

(**) The proof of (4.18) presented here, is much shorter than the original author's proof and is due to an anonymous reviewer.

In connection with the upper bound, note that (3.9), (4.18) imply

$$\text{(4.18a)} \qquad \lim_{x\to\infty} \eta_n(x) = 1 - n(n-1) \lim_{x\to\infty} \frac{\eta_{n-2}(x)}{x^2} = 1 \qquad (n > 0)$$

We therefore conclude that the bounds (4.18) are the tightest possible. If we relax the n constraint in (4.18), it is replaced by

$$\text{(4.18b)} \qquad 0 \le \eta_n(x) \le 2 \qquad (n \ge 0)$$

as is apparent from the explicit expressions for $\eta_0(x)$, $\eta_1(x)$ ((3.2), (3.3))

The limit (4.18a) implies that the line $y(x) = 1$ is an asymptote of $\eta_n(x)$ for $x\to\infty$ (n fixed). $\eta_n(x)$ also has an asymptote for $n\to\infty$ (x fixed). To determine it, consider the explicit formulation of the recursion (3.14)

$$\text{(4.19a)} \qquad \eta_n(x) = \frac{x^2}{(n+2)(n+1)} \left[1 - \eta_{n+2}(x)\right]$$

This implies (with (4.18b))

$$\text{(4.19b)} \qquad \lim_{n\to\infty} \eta_n(x) = 0 \qquad (x \text{ fixed})$$

Returning now to (4.19a) and applying this limit to $\eta_{n+2}(x)$, we see that the function

$$\text{(4.19c)} \qquad \hat{\eta}_n(x) = \frac{x^2}{(n+2)(n+1)} = \frac{1}{e_{n+2}(x)}$$

is the asymptote of $\eta_n(x)$ for $n \to \infty$, that is,

$$(4.19d) \qquad \eta_n(x) \sim \hat{\eta}_n(x) = \frac{x^2}{(n+2)(n+1)} \qquad (n \to \infty;\ x \text{ fixed})$$

Note further that with $\eta_{n+2}(0) = 0$, (4.19a) and the continuity of $\eta_n(x)$ yield

$$(4.19e) \qquad \eta_n(x) \sim \hat{\eta}_n(x) = \frac{x^2}{(n+2)(n+1)} \qquad (x \to 0;\ n \text{ fixed})$$

Our next concern is with the variation of $\eta_n(x)$ inside its established range. We shall prove that (for $n > 2$) $\eta_n(x)$ is monotonic in both n and x. The proof employs the derivative of $\eta_n(x)$ which we evaluate now. Starting with (2.11) and noting that (2.3)

$$(4.20) \qquad \frac{dg_n(x)}{dx} = g_{n-1}(x), \qquad (n > 0)$$

we get

$$\frac{d}{dx}\left[\eta_n(x)\ g_n(x)\right] = g_n(x)\frac{d\eta_n(x)}{dx} + g_{n-1}(x)\ \eta_n(x) = \sum_{s=1}^{\infty}(-1)^{s+1} g_{(n-1)+2s}(x)$$

$$(n > 0)$$

With the use of (2.12), this yields

$$(4.21)\qquad x\,\frac{d\eta_n(x)}{dx} + n\left[\eta_n(x) - \eta_{n-1}(x)\right] = 0 \qquad (n > 0)$$

Note the symmetry of the two parts of this formula with respect to x and n. Thus, corresponding to $x\frac{d\eta_n(x)}{dx}$ associated with the continuous variable x, we have $n\,\frac{\eta_n(x)-\eta_{n-1}(x)}{n-(n-1)}$ associated with the discrete variable n.

The derivative formula tells us that if (in a given x,n region) $\eta_n(x)$ is monotonic with respect to n, it is also monotonic with respect to x.

We turn now to establish the monotonicity with respect to n or, equivalently, the sign constancy of the function

$$(4.22)\qquad F_n(x) = \eta_n(x) - \eta_{n+1}(x).$$

From (4.18a), (4.19e)

$$(4.23)\qquad F_n(0) = 0;\quad \lim_{x\to\infty} F_n(x) = 0 \qquad (n > 0)$$

Since $F_n(x)$ is continuous and not identically zero, it must have some extremum points x_n; that is,

$$(4.24)\qquad \left[\frac{dF_n(x)}{dx}\right]_{x=x_n} = 0$$

Application of the derivative formula (4.21) yields

$$(4.25)\qquad F_n(x_n) = \frac{n}{n+1}\,F_{n-1}(x_n)$$

This suggests the suitability of an inductive argument. We start with

$$(4.26)\qquad F_2(x) = \eta_2(x) - \eta_3(x) = \frac{2}{x^2}\left[(2 + \cos x) - 3\,\frac{\sin x}{x}\right]$$

Obviously

(4.27) $\quad F_2(x) > 0 \qquad\qquad (x > \pi)$

For $(0 < x \leq \pi)$, the result is not so obvious. However, since $\frac{x^2}{2} F_2(x)$ is a continuous well-behaved function, the question can be easily resolved by a computer plot covering this restricted range. The result obtained this way is

(4.28) $\quad F_2(x) > 0 \qquad\qquad (x \neq 0)$

In particular, $F_2(x_n) > 0$. Successive application of (4.25) now yields

(4.29) $\quad F_3(x_n) > 0;\ F_4(x_n) > 0;\ \ldots.\ F_n(x_n) > 0$

In other words, the curve $F_n(x)$ which touches the x axis at the points $0, \infty$, has all its other extrema above the x axis. Hence

(4.30) $\quad F_n(x) > 0 \qquad\qquad (n > 1;\ x \neq 0)$

or, equivalently (proving n-monotonicity),

(4.31) $\quad \eta_n(x) - \eta_{n+1}(x) > 0 \qquad\qquad (n > 1\ ;\ x \neq 0)$

This and (4.21) now yield the x monotonicity

(4.32) $\quad \dfrac{d\eta_n(x)}{dx} > 0 \qquad\qquad (n > 2\ ;\ x \neq 0\ )$

Both these monotonicity statements are clearly borne out in Figs. 3, 4. Note that the origin is excluded from (4.31), (4.32) only in the interest of a stronger statement. Thus, with $\eta_n(0) = 0$ we can certainly extend (4.31)

(4.33) $\quad \eta_n(x) - \eta_{n+1}(x) \geq 0 \qquad\qquad (n > 1)$

Similarly, application of (4.19e), shows that the origin derivative is zero, thus yielding

$$(4.34) \qquad \frac{d\eta_n(x)}{dx} \geq 0 \qquad (n > 2)$$

The monotonicity of $\eta_n(x)$ will now be applied to constrain $\eta_n(x)$ to x dependent tighter bounds. We start with (4.19a) expressed in terms of $\hat{\eta}_n(x)$ (4.19c)

$$(4.35) \qquad \eta_n(x) = \hat{\eta}_n(x)\left[1 - \eta_{n+2}(x)\right]$$

Applying now (4.31), we get

$$(4.36) \qquad \eta_{n+2}(x) < \eta_n(x) = \hat{\eta}_n(x)\left[1 - \eta_{n+2}(x)\right] > \hat{\eta}_n(x)\left[1-\eta_n(x)\right] \quad (n > 1;\ x \neq 0)$$

This yields the following two bounds

$$(4.37) \qquad \eta_{n+2}(x) < \hat{\eta}_n(x)\left[1 - \eta_{n+2}(x)\right];\ \therefore \eta_{n+2}(x) < \frac{\hat{\eta}_n(x)}{1+\hat{\eta}_n(x)} \qquad (n > 1;\ x \neq 0)$$

$$(4.38) \qquad \eta_n(x) > \hat{\eta}_n(x)\left[1 - \eta_n(x)\right];\ \therefore \eta_n(x) > \frac{\hat{\eta}_n(x)}{1+\hat{\eta}_n(x)} \qquad (n > 1;\ x \neq 0)$$

and hence finally,

$$(4.39) \qquad \frac{\hat{\eta}_n(x)}{1+\hat{\eta}_n(x)} < \eta_n(x) < \frac{\hat{\eta}_{n-2}(x)}{1+\hat{\eta}_{n-2}(x)} < 1 \qquad \left(\begin{array}{l} n > 1 \text{ for lower bound} \\ n > 3 \text{ for upper bound} \\ x \neq 0 \end{array}\right)$$

A simpler upper bound obtained from (4.35) and (4.4) is

$$\eta_n(x) < \hat{\eta}_n(x) \qquad (n > 1;\ x \neq 0) \tag{4.40}$$

In other words, $\eta_n(x)$, lies below its asymptote, $\hat{\eta}_n(x)$.

We conclude this section with a summary of the main results obtained:

For $n > 1$, $x \neq 0$, and with $\hat{\eta}_n(x) = \dfrac{x^2}{(n+2)(n+1)}$, $\eta_n(x)$ has the following properties:

1. $0 < \eta_n(x) < 1$
2. $\eta_n(x) < \hat{\eta}_n(x)$
3. $\dfrac{\hat{\eta}_n(x)}{1+\hat{\eta}_n(x)} < \eta_n(x) < \dfrac{\hat{\eta}_{n-2}(x)}{1+\hat{\eta}_{n-2}(x)}$ $\qquad$ ($n > 3$ for upper bound)
4. $\eta_n(x)$ is a monotonically increasing function of x $\quad (n > 2)$
5. $\eta_n(x) \sim \hat{\eta}_n(x)$ $\qquad (x \to 0;\ n \text{ fixed})$
6. $\lim\limits_{x \to \infty} \eta_n(x) = 1$
7. $\eta_n(x)$ is a monotonically decreasing function of n
8. $\eta_n(x) \sim \hat{\eta}_n(x)$ $\qquad (n \to \infty;\ x \text{ fixed})$
9. $x \dfrac{d\eta_n(x)}{dx} + n\left[\eta_n(x) - \eta_{n-1}(x)\right] = 0$

These properties are utilized in the design and performance analysis of the algorithm computing $\eta_n(x)$. Some additional properties will be derived later on.

5. The Algorithm to Compute $\eta_n(x)$

Having established the basic properties of $\eta_n(x)$ we return now to consider the application of recursion to its evaluation.

Our first subject is the dividing line between forward and backward recursion. We have already seen that this is characterized by

$$(5.1) \qquad e_n(x) = 1$$

Now, since n is an integer, (5.1) has no solution for an arbitrary x. Hence we proceed with caution replacing (5.1) by (5.2)

$$(5.2) \qquad e_\alpha(x) = \frac{\alpha(\alpha-1)}{x^2} = 1$$

where α is not constrained to be an integer. The solution is given by

$$(5.3) \qquad \alpha = \alpha(x) = \tfrac{1}{2}\left(\sqrt{1+4x^2} + 1\right)$$

It is obvious then that $n_1(x)$ introduced in (3.13) satisfies

$$(5.4) \qquad n_1(x) = \left[\alpha(x)\right]$$

This means

$$(5.5) \qquad n_1(x) \leq \alpha(x) < 1 + n_1(x)$$

(5.6) $$e_{n_1}(x) \le 1 < e_{n_1+1}(x)$$

We pause here to derive some simple subsequently needed inequalities concerning $n_1(x)$. Starting with (5.2), (5.4) we get

$$(n_1-1)^2 < n_1(n_1-1) \le \alpha(\alpha-1) = x^2$$

$$\therefore n_1(x) - 1 < x$$

Similarly,

$$(n_1+1)^2 > (n_1+1)\, n_1 > \alpha(\alpha-1) = x^2$$

$$\therefore n_1(x) + 1 > x$$

Combining, we get the required result

(5.7) $$n_1(x) - 1 < x < n_1(x) + 1$$

Let us consider now the evaluation of a specific $\eta_n(x)$ by either forward or backward recursion ((3.11), (3.14))

(5.8) $$\eta_n(x) = 1 - e_n(x)\, \eta_{n-2}(x) = \frac{1}{e_{n+2}(x)} \left[1 - \eta_{n+2}(x)\right]$$

Application of (5.6) shows that forward recursion is applicable for $n \leq n_1(x)$ while backward recursion is applicable for $n + 2 \geq n_1(x) + 1$, that is, $n \geq n_1(x) - 1$. We see that there is an overlap and the conclusion for the evaluation of $\eta_n(x)$ is therefore as follows:

$$(5.8a)\quad \begin{cases} \text{For } n \leq n_1(x) - 2 & \text{use forward recursion only} \\ \text{For } n \geq n_1(x) + 1 & \text{use backward recursion only} \\ \text{For } n = n_1(x) - 1,\ n_1(x) & \text{use either} \end{cases}$$

Note that since $\eta_n(x)$ is a monotonically decreasing function of n, this means that forward recursion should generate the high values of $\eta_n(x)$ while backward recursion should generate the low values.

Actually, with the properties of $\eta_n(x)$ so far established, it is possible to make a stronger statement regarding the value of $\eta_n(x)$ at the transition between forward and backward recursions. Thus, we prove now that

$$(5.9)\qquad \eta_{n_1+1}(x) < \tfrac{1}{2}$$

$$(5.10)\qquad \eta_{n_1-2}(x) > \tfrac{1}{2} \qquad\qquad (x \neq 0)$$

This is a direct result of combining (4.39) and (5.6) while recalling the definition (4.19c) of $\hat{\eta}_n(x)$. Specifically,

$$(5.11)\qquad \eta_{n_1+1}(x) < \frac{\hat{\eta}_{n_1-1}(x)}{1+\hat{\eta}_{n_1-1}(x)} = \frac{1}{e_{n_1+1}(x)+1} < \frac{1}{2} \qquad (x \neq 0)$$

$$(5.12)\qquad \eta_{n_1-2}(x) > \frac{\hat{\eta}_{n_1-2}(x)}{1+\hat{\eta}_{n_1-2}(x)} = \frac{1}{e_{n_1}(x)+1} \geq \frac{1}{2} \qquad (x \neq 0)$$

The property just proved makes it possible to apply a slight modification to the directive (5.8a) that would set the value $\frac{1}{2}$ as the dividing line between forward and backward recursions. To accomplish that we tentatively adopt the following algorithm:

1. Compute $\eta_n(x)$ using forward recursion up to $n \leq n_1(x)$.
2. Let $\eta_m(x)$ be the last η function computed in step 1. If $\eta_m(x) < \frac{1}{2}$ erase $\eta_m(x)$ and go to step 4.
3. Set $\eta_{m+2}(x)$ as the last η function to be computed by backward recursion. Go to step 5.
4. Set $\eta_m(x)$ as the last η function to be computed by backward recursion.
5. Initiate backward recursion.

This algorithm, while consistent with (5.8a), ensures that forward recursion generates $\eta_n(x) \geq \frac{1}{2}$ while backward recursion generates $\eta_n(x) < \frac{1}{2}$. Note that only one term has to be tested in step 2 whereas (5.9), (5.10) indicate two questionable terms ($\eta_{n_1}(x)$, $\eta_{n_1-1}(x)$). This follows from the fact that n is stepped by two in every cycle of the $\eta_n(x)$ recursion.

Note further that, in view of the limit (4.18a), some of the $\eta_n(x)$ values generated in the forward recursion will be very close to 1 and if a need for the entity $[1 - \eta_n(x)]$ will arise, loss of significant digits will result. One could easily protect against this by modifying the forward recursion to generate $[1 - \eta_n(x)]$ rather than $\eta_n(x)$. We examine this in detail now.

Denote

$$\theta_n(x) = 1 - \eta_n(x) \tag{5.15}$$

Applying the recurrence (3.11) we get

$$\theta_n(x) = e_n(x)\,[1 - \theta_{n-2}(x)] \tag{5.16}$$

The initial values for the recursion are obtained from (3.2), (3.3)

$$\theta_0(x) = \cos x \tag{5.17}$$

$$\theta_1(x) = \frac{\sin x}{x} \tag{5.18}$$

Note that while in general ((4.18), (5.15))

$$0 < \theta_n(x) \le 1, \qquad (n > 1) \tag{5.19}$$

the values of $\theta_n(x)$ generated by the forward recursion satisfy

$$(5.19a) \qquad 0 < \theta_n(x) \le \tfrac{1}{2}$$

The significance of (5.19a) is that conversion back to $\eta_n(x)$(using (5.15)) would produce an $\eta_n(x)$ having the same number of significant digits that $\theta_n(x)$ has. We conclude that the advantage of generating $\theta_n(x)$ in the forward recursion is that it guarantees essentially the same relative error in both $\eta_n(x)$ and $1 - \eta_n(x)$.

It should be pointed out that there is an alternative approach to the problem, based on the recursion (3.9)

$$(5.20) \qquad \theta_n(x) = \frac{n(n-1)}{x^2} \eta_{n-2}(x)$$

that is, if we need a high precision $[1 - \eta_n(x)]$ when $\eta_n(x)$ is close to 1, we should get it from $\eta_{n-2}(x)$ as indicated in (5.20) rather than use subtraction. We prefer the first approach based on the generation of $\theta_n(x)$ for the simple reason that when a transformation is called for, it is simpler in this case.

With the adoption of the $\theta_n(x)$ forward recursion, the algorithm to compute $\eta_n(x)$, $\theta_n(x)$ now assumes the following final form:

1. Compute $\theta_n(x)$ using forward recursion up to $n \le n_1(x)$.
2. Let $\theta_m(x)$ be the last θ function computed in step 1. If $\theta_m(x) > \frac{1}{2}$, erase $\theta_m(x)$ and go to step 4.
3. Set $\eta_{m+2}(x)$ as the last η function to be computed by backward recursion. Go to step 5.

4. Set $\eta_m(x)$ as the last η function to be computed by backward recursion.

5. Initiate backward recursion.

Before leaving this subject, we consider a somewhat different formulation of the basic integrals (2.19), (2.20), made possible by the introduction of $\theta_n(x)$. Let us denote in analogy with (2.18)

$$\theta_{nk} = \theta_n(k \tfrac{\pi}{2}) \tag{5.21a}$$

The equivalents of (5.15), (5.20) are now:

$$\theta_{nk} = 1 - \eta_{nk} \tag{5.21b}$$

$$\theta_{nk} = \frac{n(n-1)}{(k \frac{\pi}{2})^2} \eta_{n-2,k} \tag{5.21c}$$

Applying (5.21c) to all $\eta_{n-1,k}$ terms in (2.19), (2.20), we get

$$\left(\frac{2}{\pi} \cdot \frac{a}{k}\right)^{n+1} \int_0^{\frac{\pi}{2} \cdot \frac{k}{a}} \xi^n \cos a\xi \, d\xi = \begin{cases} \dfrac{\theta_{n+1,k}}{n+1} & (k \bmod 4 = 0) \\[2ex] \dfrac{\eta_{nk}}{(k \frac{\pi}{2})} & (k \bmod 4 = 1) \\[2ex] -\dfrac{\theta_{n+1,k}}{n+1} & (k \bmod 4 = 2) \\[2ex] -\dfrac{\eta_{nk}}{(k \frac{\pi}{2})} & (k \bmod 4 = 3) \end{cases} \tag{5.22}$$

$$(5.23)\quad \left(\frac{2}{\pi}\cdot\frac{a}{k}\right)^{n+1}\int_0^{\frac{\pi}{2}\cdot\frac{k}{a}} \xi^n \sin a\xi\, d\xi = \begin{cases} -\dfrac{\eta_{nk}}{(k\frac{\pi}{2})} & (k \bmod 4 = 0) \\[2ex] \dfrac{\theta_{n+1,k}}{n+1} & (k \bmod 4 = 1) \\[2ex] \dfrac{\eta_{nk}}{(k\frac{\pi}{2})} & (k \bmod 4 = 2) \\[2ex] -\dfrac{\theta_{n+1,k}}{n+1} & (k \bmod 4 = 3) \end{cases}$$

Note that (1.5a) is a special case of (5.22) with

$$a = \text{integer} = m \; ; \qquad k = 4m$$

The more general integrals (2.16), (2.17) can also be rephrased, using $\theta_n(x)$:

$$(5.24)\quad \int_0^{\frac{x}{a}} \xi^n \cos a\xi\, d\xi = \left(\frac{x}{a}\right)^{n+1} \left\{ \frac{\theta_{n+1}(x)}{n+1} \cos x + \frac{\eta_n(x)}{x} \sin x \right\}$$

$$(5.25)\quad \int_0^{\frac{x}{a}} \xi^n \sin a\xi\, d\xi = \left(\frac{x}{a}\right)^{n+1} \left\{ \frac{\theta_{n+1}(x)}{n+1} \sin x - \frac{\eta_n(x)}{x} \cos x \right\}$$

Equivalently,

$$(5.26)\qquad \int_0^{\frac{x}{a}} \xi^n e^{ia\xi} d\xi = \left(\frac{x}{a}\right)^{n+1} e^{ix} \left\{ \frac{\theta_{n+1}(x)}{n+1} - i \frac{\eta_n(x)}{x} \right\}$$

6. The Backward Recursion

The basic questions that will concern us in this section relate to the initial value for the backward recursion. As in the preliminary discussion in section 3, we are still assuming infinite precision computations. If we set out to compute $\eta_p(x)$ for

$$(6.1)\qquad n_1(x)-1 \leq p \leq n,$$

the recursion should obviously start with $\eta_{n+k}(x)$ where

$$(6.2)\qquad k > 0$$

This raises two related questions:

a. What estimate should be adopted for the unavailable $\eta_{n+k}(x)$?

b. With the adopted estimate, how large should k be so that the effect of the estimation error on $\eta_n(x)$ be negligible?

Precise answers to these questions will be obtained here.

We start with the error propagation in the backward recursion. Let the error in $\eta_{n+k}(x)$ be denoted $\Delta\eta_{n+k}(x)$. We have already seen ((3.12),(3.10)) that $\Delta\eta_{n+k-2}(x)$, the resulting error in $\eta_{n+k-2}(x)$ satisfies

$$(6.3)\qquad |\Delta\eta_{n+k-2}(x)| = \frac{x^2}{(n+k)(n+k-1)} |\Delta\eta_{n+k}(x)|$$

Successive application of this equation yields (with the notation (2.15))

$$|\Delta\eta_n(x)| = \frac{x^k}{(n+1)_k}\, |\Delta\eta_{n+k}(x)| \tag{6.4}$$

We turn our attention now to the initial error $\Delta\eta_{n+k}(x)$. Obviously, we could adopt the estimate $\eta_{n+k}(x) = 0$ and be sure that (5.9)

$$|\Delta\eta_{n+k}(x)| < \tfrac{1}{2}$$

We can, however, do better than that if we adopt $\hat{\eta}_{n+k}(x)$ (4.19c), the asymptote of $\eta_{n+k}(x)$ (for $n \to \infty$) as its estimate. We have already seen (4.40) that the asymptote $\hat{\eta}_n(x)$ lies above $\eta_n(x)$ so that the error in adopting it as an estimate is positive. To determine its bound, we rearrange (4.35), getting

$$\hat{\eta}_n(x) - \eta_n(x) = \hat{\eta}_n(x)\, \eta_{n+2}(x) \tag{6.7}$$

But (4.40), $\eta_{n+2}(x) < \hat{\eta}_{n+2}(x)$. Hence (4.19c),

$$0 < \hat{\eta}_n(x) - \eta_n(x) < \hat{\eta}_n(x)\, \hat{\eta}_{n+2}(x) = \frac{x^4}{(n+1)_4} \tag{6.8}$$

The constraint $x \neq 0$ is implied here and is to be understood throughout this section as there is nothing to compute otherwise ($\eta_n(0) = 0$).

(6.8) is directly applicable to our problem. Thus, if we adopt

$\hat{\eta}_{n+k}(x)$ as the estimate for $\eta_{n+k}(x)$, (6.8) tells us that

$$(6.9) \qquad |\Delta\eta_{n+k}(x)| < \frac{x^4}{(n+k+1)_4}$$

Substitution in (6.4) now yields

$$(6.10) \qquad |\Delta\eta_n(x)| < \frac{x^{k+4}}{(n+1)_{k+4}}$$

Our goal however is to establish bounds for the relative error $|\Delta\eta_n(x)/\eta_n(x)|$. This can be obtained from (6.10) as follows

$$(6.11) \qquad \left|\frac{\Delta\eta_n(x)}{\eta_n(x)}\right| < \frac{1}{\eta_n(x)} \frac{x^2}{(n+1)_2} \frac{x^{k+2}}{(n+3)_{k+2}} = \frac{\hat{\eta}_n(x)}{\eta_n(x)} \frac{x^{k+2}}{(n+3)_{k+2}}$$

But (4.35) prescribes

$$(6.12) \qquad \frac{\hat{\eta}_n(x)}{\eta_n(x)} = \frac{1}{1-\eta_{n+2}(x)}$$

Now, since $n \geq n_1(x) - 1$ (6.1),

$$\eta_{n+2}(x) \leq \eta_{n_1+1}(x) < \tfrac{1}{2}$$

(see (5.9)). Hence

$$(6.13) \qquad \frac{\hat{\eta}_n(x)}{\eta_n(x)} < 2 \qquad (n \geq n_1(x) - 1)$$

Substituting in (6.11), we finally get the basic result for the determination of k.

$$(6.18) \qquad \left|\frac{\Delta\eta_n(x)}{\eta_n(x)}\right| < 2 \frac{x^{k+2}}{(n+3)_{k+2}}$$

Assume now that we wish to choose k so as to guarantee

$$(6.19) \qquad \left|\frac{\Delta\eta_n(x)}{\eta_n(x)}\right| < \delta_b$$

Writing out (6.18) explicitly, suggests one way of accomplishing that

$$(6.20) \qquad \frac{1}{2}\left|\frac{\Delta\eta_n(x)}{\eta_n(x)}\right| < \frac{x}{n+3} \cdot \frac{x}{n+4} \cdot \cdots\cdots \cdot \frac{x}{n+k+4} < \frac{\delta_b}{2}$$

A computer subroutine implementing (6.20) would multiply and count the indicated factors (starting from the left) and test after each multiplication whether the bound $\frac{\delta_b}{2}$ is still exceeded.

A simpler solution can be based on the fact that in general (2.15)

$$(6.21) \qquad n_k > n^k$$

Hence

$$\left|\frac{\Delta\eta_n(x)}{\eta_n(x)}\right| < 2\,\frac{x^{k+2}}{(n+3)_{k+2}} < 2\left(\frac{x}{n+3}\right)^{k+2} = \delta_b \tag{6.22}$$

This yields

$$k = \frac{\log\left(\frac{\delta_b}{2}\right)}{\log\left(\frac{x}{n+3}\right)} - 2 \tag{6.23}$$

In a binary machine we express the realtive error bound δ_b as a power of 2

$$\delta_b = 2^{-\gamma} \tag{6.24}$$

Hence

$$k = \frac{\gamma+1}{\log_2\left(\frac{n+3}{x}\right)} - 2 \tag{6.25}$$

The disadvantage of this method is that the k prescribed by (6.25) increases without bound when $n + 3 \to x$. This is a direct result of the application of (6.21).

To overcome this difficulty we proceed now to establish an x-independent upper bound for k. This constraint means that the bound will have to be computed only once for a computed array and thus we can afford a more involved computation here.

Our first task along this line is the determination of a lower bound for the general expression $(m+1)_r$. Starting with Stirling's formula (with the correction term) [4]

(6.26) $$m! = \sqrt{2\pi}\, m^{m+\frac{1}{2}} e^{-m} f_1(m)$$

we get

(6.27) $$(m+1)_r = \frac{(m+r)!}{m!} = (1 + \frac{r}{m})^{m+\frac{1}{2}} (\frac{m+r}{e})^r \frac{f_1(m+r)}{f_1(m)}$$

We replace now the variable r by the more convenient variable s where

(6.28) $$r = sm$$

Hence

(6.29) $$(m+1)_{sm} = Q_{sm} \left[\left(\frac{m}{e}\right)^s (1+s)^{1+s} \right]^m$$

where

(6.30) $$Q_{sm} = \sqrt{1+s}\, \frac{f_1[m(1+s)]}{f_1(m)}$$

Next we establish a lower bound for Q_{sm}. Our starting point here is the observation that [4]

(6.31) $$f_1(m) \leq 1 + \frac{c}{m} \qquad (c = 0.08444;\ m \geq 1)$$

Hence

(6.32) $$\left[\frac{f_1(m)}{f_1[m(1+s)]} \right]^2 \leq f_1^2(m) = 1 + s\,\frac{c}{r}(2 + \frac{c}{m}) < 1 + s\,\frac{c(2+c)}{2} = 1 + 0.088s$$

The last inequality follows from the fact that in our case $r = k + 2 > 2$. Substituting in (6.30) we get

$$(6.33) \qquad Q_{sm} = \sqrt{\frac{1+s}{\left[\frac{f_1(m)}{f_1[m(1+s)]}\right]^2}} > \sqrt{\frac{1+s}{1+0.88s}} > 1 \qquad (s > 0)$$

Therefore

$$(6.34) \qquad (m+1)_{sm} > \left[\left(\frac{m}{e}\right)^s (1+s)^{1+s}\right]^m$$

Applying this to (6.18) and noting that in this case

$$(6.35) \qquad k + 2 = s(n+2) \ ,$$

we get

$$(6.36) \qquad \frac{1}{2}\left|\frac{\Delta\eta_n(x)}{\eta_n(x)}\right| < \left[\left(\frac{xe}{n+2}\right)^s \frac{1}{(1+s)^{1+s}}\right]^{n+2}$$

To eliminate x from (6.36), we note that (6.1), (5.7) enforce

$$(6.37) \qquad \frac{x}{n+2} \leq \frac{x}{n_1+1} < 1$$

Hence, finally

$$(6.38) \qquad \frac{1}{2}\left|\frac{\Delta\eta_n(x)}{\eta_n(x)}\right| < \left[\frac{e^s}{(1+s)^{1+s}}\right]^{n+2} = \frac{\delta_b}{2}$$

The resulting equation for s is

$$(6.39) \qquad \frac{Ln(\delta_b/2)}{n+2} = s - (1+s)\, Ln(1+s)$$

or, equivalently

$$(6.40) \qquad \frac{\gamma+1}{n+2}\, Ln2 = (1+s)\, Ln(1+s) - s$$

The implementation of (6.40) is done as follows: As part of the initialization of the program, a table of the function $[(1+s)\, Ln(1+s) - s]$ is computed and stored. The determination of an acceptable s value is accomplished by comparing successive table entries to $\frac{\gamma+1}{n+2}$ Ln2. The position of the first entry exceeding this value indicates the s to be adopted, which in turn yields the desired k (from (6.35)).

As previously indicated, this k value serves only as an upper bound against which the k value computed by the direct formula (6.23), is to be compared. Note however that (6.40) and (6.23) are truly complementary in the sense that the region where (6.23) breaks down (the vicinity of $x = n+3$) is essentially the region where (6.40) yields its best (most realistic) k value (the vicinity of $x = n+2$). (see (6.36), (6.37), (6.38)).

Having adopted a specific initial value for the backward recursion, we consider now briefly the details of the convergence of the computed value $\eta'_r(x)$ ($= \eta_r(x) + \Delta\eta_r(x)$) to the true value $\eta_r(x)$. Note first that the sign of the error alternates (3.12). As the recursion is launched with $\hat{\eta}_{n+k}(x) > \eta_{n+k}(x)$, this means that the next computed value, $\eta'_{n+k-2}(x)$, lies below the true value, $\eta'_{n+k-4}(x)$ lies above the true value and so on. In other words, the computed curve, $\eta'_r(x)$ (vs. r) oscillates about the monotonic $\eta_r(x)$ with a damped oscillation starting with the highest amplitude at $r = n+k$ and diminishing to practically zero amplitude (as guaranteed by the choice of k) at $r=n$.

Another important aspect of the $\eta'_r(x)$ curve is that it lies below $\hat{\eta}_r(x)$ throughout. To see this, recall that (in the backward recursion) $|\eta'_r(x) - \eta_r(x)| = |\Delta\eta_r(x)|$ decreases with decreasing r. On the other hand, (6.7) shows $[\hat{\eta}_r(x) - \eta_r(x)]$ increasing with decreasing r. Since the initial values of the two functions (for $r=n+k$) are identical, our contention is proved. This result is utilized in section 7.

7. The Effect of Finite Precision Arithmetic

So far we have assumed infinite precision arithmetic. We consider now the guaranteed precision attainable with a normalized floating point number representation employing a finite number of digits. The treatment here is similar to that of the previous section though the situation is more complex. Whereas section 6 dealt with the propagation of a single error injected in the initial phase of the backward recursion, now we have to deal with the cumulative effect of errors injected in each and every cycle of the recursion.

Note that we are dealing now with both backward and forward recursions.

Adopting the algorithm structure of section 5, we employ the forward recursion to generate $\theta_n(x)$ rather than $\eta_n(x)$. This means that the functional forms of the two recursions dealt with are quite similar.

$$\theta_n(x) = e_n(x)\,[1 - \theta_{n-2}(x)] \qquad (\theta_n(x) \le \tfrac{1}{2}) \tag{7.1}$$

$$\eta_n(x) = \frac{1}{e_{n+2}(x)}\,[1 - \eta_{n+2}(x)] \qquad (\eta_n(x) < \tfrac{1}{2}) \tag{7.2}$$

Both equations are constrained to n regions in which the multiplier affecting error propagation is constrained to the (0,1) interval. Also, both recursions start with an initial value close to zero and successively produce higher and higher values approaching the limit $\frac{1}{2}$. As a consequence of this, the following derivations for the two cases are quite similar. We exploit this fact, giving detailed derivations only in one case ($\eta_n(x)$), being satisfied with very little more than a mere listing of the essential equations of the development in the other case ($\theta_n(x)$).

In treating the backward recursion, we adopt the following conventions: The computed values are $\eta_p(x)$ with

$$n_1(x) - 1 \le p \le n \tag{7.3}$$

The initial value starting the recursion is $\hat{\eta}_m(x)$ as the estimate for $\eta_m(x)$.

In conformance with our goal in this section, we adopt here a slightly different definition of the error $\Delta\eta_r(x)$. Let $\bar{\eta}_r(x)$ denote the values computed with infinite precision following the prescription of section 6. The difference between $\bar{\eta}_r(x)$ and $\eta_r(x)$ is due only to the error in the initial term of the backward recursion. Note, however, that for $r \leq n$, the two may be assumed identical. When we consider now the effect of finite precision arithmetic, we find that the computed values, $\eta_r'(x)$, differ from $\bar{\eta}_r(x)$. This difference defines $\Delta\eta_r(x)$ in this section

$$\Delta\eta_r(x) = \eta_r'(x) - \bar{\eta}_r(x)$$

Let the component of $\Delta\eta_r(x)$ due to an error introduced in $\eta_s(x)$ $(s \geq r)$ be denoted $\epsilon_{rs}(x)$. Applying now the reasoning that led to (6.4), we get

$$\left|\epsilon_{r,r+k}(x)\right| = \frac{x^k}{(r+1)_k} \left| \epsilon_{r+k,r+k}(x) \right| \tag{7.4}$$

As the recursion generating $\eta_p(x)$ starts with $\hat{\eta}_m(x)$, errors of the type considered will be injected into $\bar{\eta}_m(x)$, $\bar{\eta}_{m-2}(x)$, ... $\bar{\eta}_p(x)$. Hence, the total error in $\bar{\eta}_p(x)$, being the sum of these injected errors, is bounded by

$$\left|\Delta\eta_p(x)\right| < \sum_{r=0}^{\frac{1}{2}(m-p)} \frac{x^{2r}}{(p+1)_{2r}} \left|\epsilon_{p+2r,p+2r}(x)\right| \tag{7.5}$$

We note that each cycle of the recursion involves the operation of subtraction from unity followed by multiplication. Depending on the specific computer used and its arithmetic subroutines, an upper bound δ can be determined for the relative error generated in these operations, that is, we know that the injected errors in (7.5) satisfy

$$\frac{|e_{kk}(x)|}{\overline{\eta}_k(x)} < \delta \tag{7.6}$$

Therefore, the total error in $\eta_p(x)$ satisfies

$$|\Delta\eta_p(x)| < \delta \sum_{r=0}^{\frac{1}{2}(m-p)} \frac{x^{2r}}{(p+1)_{2r}} \overline{\eta}_{p+2r}(x) \tag{7.7a}$$

and the relative error is bounded by

$$\left|\frac{\Delta\eta_p(x)}{\eta_p(x)}\right| < \delta \left\{1 + \sum_{r=1}^{\frac{1}{2}(m-p)} \frac{x^{2r}}{(p+1)_{2r}} \left(\frac{\overline{\eta}_{p+2r}(x)}{\eta_p(x)}\right)\right\} \tag{7.7b}$$

From the discussion at the end of section 6, it is obvious that the curve $\overline{\eta}_r(x)$ vs. r lies beneath the curve $\hat{\eta}_r(x)$. Therefore

$$\overline{\eta}_{p+2r}(x) \leq \hat{\eta}_{p+2r}(x) < \hat{\eta}_p(x) \qquad (r > 0) \tag{7.7c}$$

and hence (6.13)

$$\frac{\overline{\eta}_{p+2r}(x)}{\eta_p(x)} < \frac{\hat{\eta}_p(x)}{\eta_p(x)} < 2 \tag{7.7d}$$

Applying this to (7.7b) yields

$$(7.8) \qquad \left|\frac{\Delta\eta_p(x)}{\eta_p(x)}\right| < \delta\left\{1 + 2\sum_{r=1}^{\frac{1}{2}(m-p)} \frac{x^{2r}}{(p+1)_{2r}}\right\}$$

It can be shown that in most cases the multiplier of the sum in (7.8) may be reduced to 1. However, the complications inherent in such an analysis do not seem to be justified and we have adopted the higher but simpler bound of (7.8).

Expressing (7.8) in terms of factorials yields

$$(7.9) \qquad \left|\frac{\Delta\eta_p(x)}{\eta_p(x)}\right| < \delta\left\{1 + 2\,\frac{p!}{x^p}\sum_{r=1}^{\frac{1}{2}(m-p)} \frac{x^{p+2r}}{(p+2r)!}\right\}$$

We have here a partial sum of the sinh x or cosh x series and proceed now to modify it so as to get a partial sum of the e^x series instead. We start with the following identity

$$\frac{x^{k+1}}{(k+1)!} = \frac{1}{1+\frac{k+1}{x}}\left[\frac{x^k}{k!} + \frac{x^{k+1}}{(k+1)!}\right]$$

$$(7.10) \qquad < \frac{1}{2}\left[\frac{x^k}{k!} + \frac{x^{k+1}}{(k+1)!}\right] \qquad (k+1 \geq x)$$

Now, the lowest factorial appearing in the (7.9) sum is $(p+2)! \geq [n_1(x)+1]!$ (See (7.3)). But from (5.7), $n_1(x) + 1 > x$. Hence the constraint in (7.10) is satisfied by all the summation terms of (7.9) which can therefore be modified according to (7.10). The result is

$$(7.11)\qquad \left|\frac{\Delta\eta_p(x)}{\eta_p(x)}\right| < \delta\left\{1 + \frac{p!}{x^p}\sum_{r=1}^{m-p}\frac{x^{p+r}}{(p+r)!}\right\} = \delta\left\{1 + \frac{p!}{x^p}\sum_{r=p+1}^{m}\frac{x^r}{r!}\right\}$$

The partial sum of the e^x series appearing in (7.11) may now be evaluated in terms of $Q(\chi^2|\nu)$, the (tabulated) integral of the χ^2 distribution [6] as follows:

$$\sum_{r=s}^{t}\frac{x^r}{r!} = \sum_{r=0}^{t}\frac{x^r}{r!} - \sum_{r=0}^{s-1}\frac{x^r}{r!}$$

$$(7.12)\qquad = e^x\left[Q(2x|2t+2) - Q(2x|2s)\right]$$

Applying this to (7.11) yields

$$(7.13)\qquad \left|\frac{\Delta\eta_p(x)}{\eta_p(x)}\right| < \delta\left\{1 + \frac{p!e^x}{x^p}\left[Q(2x|2m+2) - Q(2x|2p+2)\right]\right\}$$

Note that the Q functions are bounded by 1 and their difference will usually assume a value close to $\frac{1}{2}$. The bound (7.13) is applicable to arbitrary p,x. When we are interested only in the maximal bound, (7.13) can be further simplified. We observe that (7.8) shows the relative error bound to be a decreasing function of p and an increasing function of x. This means that the maximal bound is associated with $p = n_1(x)-1$. Furthermore, since (5.7) $x < n_1(x) + 1$, it is safe to substitute this bound for x.

To obtain the maximal bound we therefore make the following substitutions on the right of (7.13)

$$(7.14) \qquad \begin{cases} p = n_1(x) - 1 \\ x = n_1(x) + 1 \end{cases}$$

In particular, we substitute for $\frac{p!}{x^p} e^x$

$$\frac{(n_1-1)!}{(n_1+1)^{n_1-1}} e^{n_1+1} = \left(1 + \frac{1}{n_1}\right) \frac{(n_1+1)! e^{n_1+1}}{(n_1+1)^{n_1+1}} < 1.085\sqrt{2\pi(n_1+1)}\left(1 + \frac{1}{n_1}\right)$$

(see (6.26)). The final result is

$$(7.15) \qquad \max_p \left|\frac{\Delta\eta_p(x)}{\eta_p(x)}\right| < \delta \left\{1 + 1.085\sqrt{2\pi(n_1+1)}\left(1+\frac{1}{n_1}\right)\left[Q(2n_1+2|2m+2) - Q(2n_1+2|2n_1-2)\right]\right\}$$

We turn now to the case of $\theta_n(x)$. The convention adopted here is as follows: The computed values are $\theta_p(x)$ with

$$(7.16a) \qquad m + 2 \le p \le n_1(x)$$

For even p, m=0 and the recursion starts with $\theta_0(x) = \cos(x)$. For odd p, m=1 and the recursion starts with $\theta_1(x) = \frac{\sin x}{x}$. We assume that both initial values are available with a relative error bound δ.

As previously explained, the derivation here will be sketchy. In analogy with $\epsilon_{rs}(x)$, let $\gamma_{rs}(x)$ be the (non relative) error component in $\theta_r(x)$ due to an error introduced in $\theta_s(x)$. From (7.1) we get

$$(7.16b) \qquad \left|\gamma_{p,p-k}(x)\right| = \frac{(p+1-k)_k}{x^k} \left|\gamma_{p-k,p-k}(x)\right|$$

Hence

$$(7.17) \qquad \left|\frac{\gamma_{p,p-k}(x)}{\theta_p(x)}\right| = \frac{(p+1-k)_k}{x^k}\left|\frac{\gamma_{p-k,p-k}(x)}{\theta_p(x)}\right| < \frac{(p+1-k)_k}{x^k}\left|\frac{\gamma_{p-k,p-k}(x)}{\theta_{p-k}(x)}\right|$$

The initial value is $\theta_m(x)$. Relative errors bounded by

$$(7.18) \qquad \left|\frac{\gamma_{p-k,p-k}(x)}{\theta_{p-k}(x)}\right| < \delta$$

are injected into $\theta_m(x)$, $\theta_{m+2}(x)$, ..., $\theta_p(x)$. Hence

$$(7.19) \qquad \left|\frac{\Delta\theta_p(x)}{\theta_p(x)}\right| < \delta \sum_{r=0}^{\frac{1}{2}(p-m)} \frac{(p+1-2r)_{2r}}{x^{2r}}$$

$$(7.19a) \quad \therefore \quad \left|\frac{\Delta\theta_p(x)}{\theta_p(x)}\right| < \delta \left\{1 + \frac{p!}{x^p} \cdot \sum_{r=1}^{\frac{1}{2}(p-m)} \frac{x^{p-2r}}{(p-2r)!}\right\}$$

The sum in (7.19a) is to be modified now according to

$$(7.20) \qquad \frac{x^k}{k!} = \frac{1}{1+\frac{x}{k+1}}\left[\frac{x^k}{k!} + \frac{x^{k+1}}{(k+1)!}\right] < \frac{1}{2}\left[\frac{x^k}{k!} + \frac{x^{k+1}}{(k+1)!}\right] \qquad (x \geq k+1)$$

For the application of this to (7.19a) to be valid, we must have $n_1(x)-1 < x$. This is indeed the case (5.7). Applying (7.20) to (7.19a) yields

$$(7.21) \qquad \left|\frac{\Delta\theta_p(x)}{\theta_p(x)}\right| < \delta \left\{1 + \frac{1}{2}\frac{p!}{x^p} \sum_{r=m}^{p-1} \frac{x^r}{r!}\right\}$$

$$(7.22) \qquad < \delta \left\{1 + \frac{1}{2}\frac{p!e^x}{x^p}\, Q(2x|2p)\right\}$$

(7.19) indicates the error bound is a decreasing function of x and an increasing function of p. Hence the maximal error bound is obtainable with

$$(7.23) \qquad \begin{cases} p = n_1(x) \\ x = n_1 - 1 \end{cases}$$

$$(7.24) \qquad \max_p \left|\frac{\Delta\theta_p(x)}{\theta_p(x)}\right| < \delta \left\{1 + 1.085\sqrt{\frac{\pi}{2}(n_1-1)} \quad \frac{1}{1-\frac{1}{n_1}}\ Q(2n_1-2|2n_1)\right\} \qquad (n_1 > 1)$$

The most common application of the formulae derived in this section would be in the determination of a single error bound applicable to a whole array of generated $\eta_n(x)$, $\theta_n(x)$. To illustrate this, consider the somewhat arbitrary x,n plane situation described in Fig. 1. The shaded region represents an area where $\eta_n(x), \theta_n(x)$ are to be computed. The locus $\eta_n(x) = \frac{1}{2}$ represents the separation between $\eta_n(x)$ recursion (backward) and $\theta_n(x)$ recursion (forward). We refer to this locus as the separation line. It is closely approximated for large n by the curve $n=\alpha(x)\approx x+\frac{1}{2}$(see(5.2)). Inequality (7.8) tells us that the error bound for an $\eta_n(x)$ corresponding to a point on this line is higher than the bound for any point above it (same x, larger n). (7.15) computes this maximal bound. Now, for large $n_1(x)$, the dominant term in (7.15) is $\sqrt{n_1+1}$. Therefore, the highest bound for $\left|\frac{\Delta\eta_n(x)}{\eta_n(x)}\right|$ is obtained by moving up along the separation line. This brings us to point A in Fig. 1.

The same line of reasoning can now be applied to the computation of $\theta_n(x)$ in the shaded region below the separation line. The conclusion is the same, namely, point A has the highest error bound for the forward ($\theta_n(x)$) recursion. Therefore, the overall error bound sought is the higher of expressions (7.15), (7.24) evaluated at point A. Due to the coarser approximations utilized in (7.15), it will usually prescribe the higher bound.

A specific example of the application of (7.15),(7.24) is described in the next section.

8. The η_{nk} Table

The presentation in the preceding sections has been couched in general terms so as to be applicable in a wide range of differing circumstances.

A specific implementation of the $\eta_n(x)$ algorithm will be described in this section. This involved the high precision computation of η_{nk} for $n = 2, 3, \ldots 201; k = 1, 2, \ldots 200$ on an IBM 7094 computer.

The computations were carried out in penta-precision arithmetic [7], that is, 5 computer words were allotted to each floating point number in a scheme that assigns 140 bits to its fractional part.

A choice of $\gamma = 140$ in (6.24) is certainly sufficient in this case. However, we chose to add 10 more bits as a precautionary measure and based the backward recursion starting point on $\gamma = 150$.

Our aim was to get as close as possible to the limit of 140 reliable bits in each η_{nk}. The guaranteed precision actually obtained is of course lower than that. Its determination involves a direct application of the results of section 7. We recall that these were phrased in terms of δ, the relative error produced in a single cycle of the recursion. To evaluate δ, we assume that the exact $\eta_{n+2}(x)$ is available and consider the relative error accompanying the computation of $\eta_n(x)$ according to the prescription (7.2)

$$(8.1) \qquad \eta_n(x) = \frac{x^2}{(n+2)(n+1)}\left[1 - \eta_{n+2}(x)\right]$$

Let ϵ denote the (non-relative) error in computing the difference $1 - \eta_{n+2}(x)$, and let δ_m denote the overall relative error introduced in the various multiplications and divisions. $\eta_n'(x)$, the computed $\eta_n(x)$, is given by

$$(8.2) \qquad \eta_n'(x) = (1 + \delta)\, \eta_n(x)$$

$$(8.3) \qquad = \frac{x^2}{(n+2)(n+1)}\, (1- \eta_{n+2}(x) + \epsilon)(1 + \delta_m)$$

Multiplying out (neglecting second order terms) yields

$$(8.3a) \qquad \eta_n'(x) = \eta_n(x) \left[1 + \frac{\epsilon}{1-\eta_{n+2}(x)} + \delta_m\right]$$

and hence

$$(8.3b) \qquad \delta = \frac{1}{1-\eta_{n+2}(x)}\, \epsilon + \delta_m$$

A simple δ bound can be obtained from this on recalling that in the backward recursion ((5.9), (4.4))

(8.3c) $$\frac{1}{2} < 1 - \eta_{n+2}(x) < 1$$

The bound obtained with this condition is

(8.3d) $$|\delta| < 2|\varepsilon| + |\delta_m|$$

Inequality (8.3c) is also the basis for the analysis of the subtraction error ε. Note first that (8.3c) implies that in a binary, normalized, floating point representation, $(1 - \eta_{n+2}(x))$ will always be represented with the exponent zero. Its least significant bit therefore represents in our case $2^0 \cdot 2^{-140} = 2^{-140}$. It follows that $|\varepsilon_r|$, the absolute error in the representation of $1 - \eta_{n+2}(x)$ is bounded by

(8.4) $$|\varepsilon_r| < 2^{-140}$$

Note that the $|\varepsilon_r|$ bound is independent of the specific way in which $1 - \eta_{n+2}(x)$ is computed and as such serves as a lower limit for the $|\varepsilon|$ bound. Whether this bound will be 2^{-140} or a higher number, will depend on the way we implement the subtraction. As a matter of fact, straightforward subtraction is to be avoided as it yields a higher bound. To see this, note that the number 1 is represented in the computer as $2^1 \cdot \frac{1}{2}$. This means that prior to subtraction, $\eta_{n+2}(x)$ has to be brought to an (unnormalized) representation with the exponent +1. In this representation, its least significant bit corresponds to $2^1 \cdot 2^{-140} = 2^{-139}$. Therefore,

$$0 < \varepsilon < 2^{-139}$$

Note that this bound is twice the bound of $|\varepsilon_r|$. The loss of precision involved here is directly related to the loss of bits attending the alignment for the subtraction from 1. Therefore, if we could implement the computation in a manner requiring alignment with a number smaller than 1, we would get a lower ε bound. This is, indeed, possible as indicated by the following identity

$$(8.5) \qquad 1 - \eta_{n+2}(x) = \frac{1}{2} + \left(\frac{1}{2} - \eta_{n+2}(x)\right)$$

In following the prescription on the right, note that since the representation of $\frac{1}{2}$ is $2^0 \cdot \frac{1}{2}$, the least significant bit of the aligned $\eta_{n+2}(x)$ corresponds to $2^0 \cdot 2^{-140} = 2^{-140}$. Therefore

$$(8.6) \qquad 0 < \varepsilon < 2^{-140}$$

Note, also, that since (5.9) $\frac{1}{2} - \eta_{n+2}(x) > 0$, no further error is involved in the final addition of $\frac{1}{2}$ as required by (8.5).

In the construction of the table we have adopted the "indirect subtraction" discussed here. Hence (8.6) is our ε bound.

Next we turn to δ_m. Note first that in our case

$$x^2 = k^2 \frac{\pi^2}{4} \qquad \text{(k integer)} \tag{8.6a}$$

k^2 and $(n+2)(n+1)$ were initially computed in fixed point without error and subsequently converted to floating point representation (still without error). The term $\frac{\pi^2}{4}$ is directly available with more than 140 bits. Hence the relative error in its 140 bit representation is easily determined and is found to be less than 2^{-140}. Finally, with the precomputed k^2 and $(n+2)(n+1)$, the computation according to (8.1), (8.6a), involves 2 multiplications and 1 division. Each such operation introduces an error bounded by a unit in the least significant bit position and hence a relative error bound of 2^{-139}. This leads then to

$$|\delta_m| < 7 \cdot 2^{-140} \tag{8.6b}$$

Applying (8.6) and (8.6b) to (8.3d), we get finally

$$|\delta| < 9 \cdot 2^{-140} \tag{8.7}$$

An analogous treatment of the forward recursion of $\theta_n(x)$ shows that the bound (8.7) is also valid in this case.

We turn now to determine an upper bound for the cumulative effect of these errors. The point of maximal error bound is determined in Fig. 2. The reasoning here is the same as that associated with Fig. 1 of section 7. Obviously, in this case we have n_1=201. The backward recursion leading to point A starts at point T with the (computed) value m=443. Hence we have to evaluate (7.15) with

$$n_1 = 201; \; m = 443$$

The result is

$$(8.8) \quad \max_{n,x} \left| \frac{\Delta\eta_n(x)}{\eta_n(x)} \right| < \delta \left\{ 1 + 1.085\sqrt{404\pi} \; (1 + \frac{1}{201}) \left[Q(404|888) - Q(404|400) \right] \right\}$$

With

$$Q(404|888) \approx 1 \; ; \; Q(404|400) \approx \frac{1}{2}$$

this leads to

$$(8.9) \qquad \max_{n,x} \left| \frac{\Delta\eta_n(x)}{\eta_n(x)} \right| < 20.5\delta$$

Next we apply (7.24) getting

$$(8.10) \qquad \max_{n,x} \left| \frac{\Delta\theta_n(x)}{\theta_n(x)} \right| < \delta \left\{ 1 + 1.085\sqrt{100\pi} \; \frac{1}{1-\frac{1}{201}} \; Q(400|402) \right\}$$

With

$$Q(400|402) \approx \frac{1}{2}$$

this yields

$$(8.11) \quad \max_{n,x} \left| \frac{\Delta\theta_n(x)}{\theta_n(x)} \right| < 10.65\delta$$

Increasing the bound to the closest integral power of 2 we get finally

$$(8.12) \quad \max_{n,x} \left\{ \left| \frac{\Delta\eta_n(x)}{\eta_n(x)} \right|, \left| \frac{\Delta\theta_n(x)}{\theta_n(x)} \right| \right\} < 2^{-132}$$

It should be pointed out that in view of the various simplifications that were adopted to get a manageable error analysis, this figure is definitely too pessimistic. This is borne out by two independent checks of the generated table. One is treated here, the other one in section 10.

The check considered here is quite obvious and has been incorporated in the generating program itself. Consider the curve $\eta_n(x)$ vs. n(fixed x). The program described here, essentially generates this curve in two pieces, one for low n and one for high n. An examination of the matching of these two pieces at the transition region is an obvious test. We implement this by extending each recursion to generate one (test) value past its legitimate range. This gives us two n values for which both recursions have generated specific values. Comparison of these should yield a good indication of the inherent precision.

Note that $e_n(x)$ used in generating the test term from the last legitimate term is very close to 1. Therefore the relative error in the test term is essentially the sum of the last term relative error and δ (bounded by (8.7)).

The program represents the two differences involving the test terms as $K \cdot 2^{-140}$ and prints the integer K. It also prints the maximal $|K|$. For the generated table described here

$$\max |K| = 3$$

that is, the maximal "mismatch" error is less than $|\delta|$ (8.7). This is certainly much lower than the value of 28δ upon which the claimed precision (8.12) is based.

Computer plots of $\eta_n(x)$ are presented in Figs. 3,4. As these are derived from the η_{nk} table, correct values are shown only when x is an integral multiple of $\frac{\pi}{2}$ while the plotting subroutine introduces linear interpolation between these points. In view of the logarithmic x scale, this is noticeable only for low x.

The various properties proved in section 4 are clearly visible here. Note the asymptotes. Note also the monotonicity vs. n and x. Of particular interest in this context is the expanded view provided by Fig. 3a for low n values. We recall that monotonicity vs. x was proved only for $n > 2$ (4.32). Indeed, we see here that $\eta_2(x)$ is not monotonic. The prominent rippling displayed in $\eta_2(x)$ is also visible in $\eta_3(x)$. However, its magnitude is much smaller in this case, ensuring the monotonicity prescribed by (4.32).

A more useful representation is shown in Figs. 5,6 which are more in line with the actual algorithm implemented. Note that all curves in Fig. 6 have been generated by recursions starting at the bottom and building up to the value $\frac{1}{2}$.

9. Other Applications

Our main concern in this paper has been with the integrals (5.22), (5.23). However, the function $\eta_n(x)$ which is instrumental in their evaluation, turns up in other areas too. We devote this section to a brief description of three such cases.

a. Related Integrals with a Positive Integrand

With the established properties of $\eta_n(x)$, it is now possible to extend the results (5.22), (5.23) to the following related positive-integrand integrals

(9.1) $$(n+1)\left(\frac{2}{\pi}\,\frac{a}{k}\right)^{n+1}\int_0^{\frac{\pi}{2}\frac{k}{a}} \xi^n(1-\cos a\xi)\,d\xi = \begin{cases} \eta_{n+1,k} & (k \bmod 4=0) \\ 1-\dfrac{n+1}{(k\frac{\pi}{2})}\,\eta_{nk} = 1-\dfrac{(k\frac{\pi}{2})}{n+2}\,\theta_{n+2,k} & (k \bmod 4=1) \\ 1+\theta_{n+1,k} & (k \bmod 4=2) \\ 1+\dfrac{n+1}{(k\frac{\pi}{2})}\,\eta_{nk} = 1+\dfrac{(k\frac{\pi}{2})}{n+2}\,\theta_{n+2,k} & (k \bmod 4=3) \end{cases}$$

(9.2) $$(n+1)\left(\frac{2}{\pi}\,\frac{a}{k}\right)^{n+1}\int_0^{\frac{\pi}{2}\frac{k}{a}} \xi^n(1-\sin a\xi)\,d\xi = \begin{cases} 1+\dfrac{n+1}{(k\frac{\pi}{2})}\,\eta_{nk} = 1+\dfrac{(k\frac{\pi}{2})}{n+2}\,\theta_{n+2,k} & (k \bmod 4=0) \\ \eta_{n+1,k} & (k \bmod 4=1) \\ 1-\dfrac{n+1}{(k\frac{\pi}{2})}\,\eta_{nk} = 1-\dfrac{(k\frac{\pi}{2})}{n+2}\,\theta_{n+2,k} & (k \bmod 4=2) \\ 1+\theta_{n+1,k} & (k \bmod 4=3) \end{cases}$$

Note that (1.5c), (1.5d) are special cases of (9.1), (9.2), respectively, with

$$a = \text{integer} = m \; ; \quad k = 4m$$

Given the results (5.22), (5.23) the derivation of (9.1), (9.2) is rather trivial. The non-trivial question posing itself, however, concerns the relative error in the above expressions. Given the fact that these integrals are actually obtained as a difference of two terms, would the relative error increase due to "cancellation"?

Inspecting (9.1), (9.2) and recalling that $\eta_{nk}, \theta_{nk} \geq 0$ we see that the only terms requiring further study are

$$1 \pm \frac{n+1}{(k\frac{\pi}{2})} \eta_{nk} \tag{9.3}$$

Actually, with η_{nk} being an even function of k(2.15a), we may restrict the investigation to the term

$$1 - \frac{n+1}{(k\frac{\pi}{2})} \eta_{nk} \,. \qquad (k > 0) \tag{9.4}$$

Our problem then is to determine the degree of "cancellation" in this term, or, in a slightly more general context: Determine how closely the function

$$p_n(x) = \frac{n+1}{x} \eta_n(x) \qquad (n, x > 0) \tag{9.5}$$

approaches the value 1. Since $p_n(x) > 0$, the answer will be provided by establishing a sufficiently low upper bound for $p_n(x)$. We start by considering the asymptotes of $p_n(x)$. These are derived from those of

$\eta_n(x)$ ((4.19e),(4.18a))

$$p_n(x) \sim \frac{n+1}{x} \cdot \frac{x^2}{(n+1)(n+2)} = \frac{x}{n+2} \qquad (x \to 0;\ n \text{ fixed}) \tag{9.6}$$

$$p_n(x) \sim \frac{n+1}{x} \cdot 1 = \frac{n+1}{x} \qquad (x \to \infty;\ n \text{ fixed}) \tag{9.7}$$

Thus we see that $p_n(x)$ is an increasing function for low x and a decreasing function for high x. Since, in addition to that, it is also finite and continuous, we conclude that there exists an x (denoted x_n) for which $p_n(x)$ is maximal and has a zero derivative. (See plots of $\frac{\eta_n(x)}{x} = \frac{p_n(x)}{n+1}$ in Fig. 7). Evaluating the derivative of $p_n(x)$ via (4.21) we get

$$\frac{dp_n(x)}{dx} = \frac{n+1}{x^2}\,[n\eta_{n-1}(x)-(n+1)\eta_n(x)]$$

$$= \frac{n+1}{x}\,[p_{n-1}(x)-p_n(x)] \tag{9.8}$$

We conclude therefore that

$$p_{n-1}(x_n) = p_n(x_n) \geq p_n(x) \tag{9.9}$$

that is, the peak of $p_n(x)$ is a value of $p_{n-1}(x)$. Hence, $p_{n-1}(x_{n-1})$, the peak of $p_{n-1}(x)$ must satisfy

(9.10) $$p_{n-1}(x_{n-1}) \geq p_n(x_n) \geq p_n(x)$$

Successive application of this, yields

(9.10a) $$p_n(x) \leq p_1(x_1) \qquad (n > 0)$$

x_1 is determined from (9.9)

$$p_0(x_1) = p_1(x_1)$$

Using (9.5), (3.2), (3.3) we get

$$- \cos x_1 = 1 - 2 \frac{\sin x_1}{x_1}$$

$$\therefore x_1 = \pi$$

Hence

(9.10b) $$p_1(x_1) = \frac{2}{\pi}$$

and the final result is

(9.10c) $$p_n(x) \leq \frac{2}{\pi} < \frac{3}{4} = 0.11_2 \qquad (n > 0)$$

This means that in the subtraction indicated in (9.4), the worst that could happen is cancellation of the leading bit. Upon normalization, this leads then to the introduction of a meaningless zero bit in the least significant position. In other words, we "lose" only one bit.

We conclude therefore that formulae (9.1), (9.2) yield results which at worst are correct to one bit less than the number of correct bits in η_{nk}.

b. <u>The derivatives of $\frac{\cos x}{x}$, $\frac{\sin x}{x}$, and band-limited functions</u>

Let us denote

$$h(x) = \frac{e^{ix}}{x} \tag{9.11}$$

We proceed now to show that $h^{(n)}(x)$, the n-th order derivative of h(x), can be expressed in terms of $\eta_n(x)$, $\theta_{n+1}(x)$.

Using the derivative operator D and applying Leibnitz's rule to

$$h^{(n)}(x) = D^n(x^{-1}e^{ix}), \tag{9.11a}$$

we get (see(2.3))

$$h^{(n)}(x) = i^n x^{-1} e^{ix} \frac{\sum_{r=o}^{n} g_r(-ix)}{g_n(-ix)} \tag{9.12}$$

Replacing the finite sum by an infinite sum as in section 2, we get (recalling (2.8))

$$h^{(n)}(x) = (-1)^n \frac{n!}{x^{n+1}} - e^{i[x+(n-1)\frac{\pi}{2}]} \psi_n(x) \tag{9.13}$$

Application of (2.13) and (5.20) brings out the explicit dependence on $\eta_n(x)$, $\theta_{n+1}(x)$.

$$(9.14) \qquad h^{(n)}(x) = (-1)^n \frac{n!}{x^{n+1}} - e^{i[x+(n-1)\frac{\pi}{2}]} \left\{ \frac{\theta_{n+1}(x)}{n+1} - i\,\frac{\eta_n(x)}{x} \right\}$$

Finally, separation into real and imaginary parts yields

$$(9.15) \quad \frac{d^n\left(\frac{\cos x}{x}\right)}{dx^n} = (-1)^n \frac{n!}{x^{n+1}} - \frac{\theta_{n+1}(x)}{n+1} \cos[x+(n-1)\tfrac{\pi}{2}] - \frac{\eta_n(x)}{x} \sin[x+(n-1)\tfrac{\pi}{2}]$$

$$(9.16) \quad \frac{d^n\left(\frac{\sin x}{x}\right)}{dx^n} = \frac{\eta_n(x)}{x} \cos[x+(n-1)\tfrac{\pi}{2}] - \frac{\theta_{n+1}(x)}{n+1} \sin[x+(n-1)\tfrac{\pi}{2}]$$

The precision obtained with these formulae **is severely limited** for certain n,x combinations due to "cancellation". Hence, for arbitrary n,x the method of W. Gautschi [8] which obtains the derivatives directly, is definitely better. There are, however, important applications in which one term of (9.16) always vanishes. In these cases, one could take advantage of $\eta_n(x)$. A specific example of this is presented by the derivatives of band-limited functions, our next topic.

Let $\rho(\tau)$ be a band-limited function such that its Fourier transform $p(\nu)$ vanishes for $|\nu| > \frac{1}{2}$. Using the sampling theorem, we can reconstruct $\rho(\tau)$ from its samples

$$(9.17) \qquad \rho_k = \rho(k) \qquad (k \text{ integer})$$

as follows:

$$\rho(\tau) = \sum_{k=-\infty}^{\infty} \rho_k \frac{\sin[\pi(k-\tau)]}{\pi(k-\tau)} \tag{9.18}$$

Hence

$$\frac{d^n\rho(\tau)}{d\tau^n} = (-\pi)^n \sum_{k=-\infty}^{\infty} \rho_k \left[\frac{d^n\left(\frac{\sin x}{x}\right)}{dx^n}\right]_{x=\pi(k-\tau)} \tag{9.19}$$

Usually, only the samples ρ_k are available and one is interested only in the derivatives at the sampling points (τ integer). In our case we have a specific application in mind (section 10) and exploit the fact that even when

$$\tau = \frac{m}{2} \qquad (m \text{ integer}) \tag{9.20}$$

that is,

$$x = \pi(k - \frac{m}{2})$$

eqn. (9.16) still yields a one-term derivative. Evaluating (9.19) under this constraint, we get

$$\left[\frac{d^n\rho(\tau)}{d\tau^n}\right]_{\tau=\frac{m}{2}} = \begin{cases} (-1)^{r+n} \dfrac{\pi^n}{n+1} \displaystyle\sum_{k=-\infty}^{\infty} (-1)^k \rho_k \, \theta_{n+1,2k-m} & (n-m=2r) \\ (-1)^{r+n} \, 2\pi^{n-1} \displaystyle\sum_{k=-\infty}^{\infty} (-1)^k \rho_k \dfrac{\eta_{n,2k-m}}{2k-m} & (n-m=2r+1) \end{cases} \tag{9.21}$$

Note that the continuous functions $\eta_n(x)$, $\theta_{n+1}(x)$ have been replaced by the corresponding discrete parameters.

Eqn. (9.21) gives the derivatives at the sample points as well as at points centered between them and should prove useful in dealing with band-limited functions. The functions appearing in (9.21), namely, $\frac{\eta_n(x)}{x}$, $\frac{\theta_n(x)}{n}$ are plotted (vs. x) in Figs. 7,8. Note the markedly different behavior of the two functions and their asymptotes.

Our interest in (9.21) hinges on its application to the η function testing developed in section 10. For this purpose, a more specific formulation is desirable. Furthermore, $\eta_n(x)$ (and hence $\theta_n(x)$) are even functions of x (2.15a). Taking advantage of this fact we can modify (9.21) to get the following explicit formulation:

$$\text{(9.22a)} \quad \left[\frac{d^{2k}\rho(\tau)}{d\tau^{2k}}\right]_{\tau=t} = (-1)^k \frac{\pi^{2k}}{2k+1}\left\{\rho_t + \sum_{r=1}^{\infty}(-1)^r(\rho_{t+r} + \rho_{t-r})\,\theta_{2k+1,2r}\right\}$$

$$\text{(9.22b)} \quad \left[\frac{d^{2k+1}\rho(\tau)}{d\tau^{2k+1}}\right]_{\tau=t} = (-1)^{k+1}\pi^{2k}\sum_{r=1}^{\infty}(-1)^r(\rho_{t+r} - \rho_{t-r})\,\frac{\eta_{2k+1,2r}}{r}$$

$$\text{(9.22c)} \quad \left[\frac{d^{2k}\rho(\tau)}{d\tau^{2k}}\right]_{\tau=t+\frac{1}{2}} = (-1)^{k+1}2\pi^{2k-1}\sum_{r=1}^{\infty}(-1)^r(\rho_{t+r} + \rho_{t-r+1})\,\frac{\eta_{2k,2r-1}}{2r-1}$$

$$\text{(9.22d)} \quad \left[\frac{d^{2k+1}\rho(\tau)}{d\tau^{2k+1}}\right]_{\tau=t+\frac{1}{2}} = (-1)^{k+1}\frac{\pi^{2k+1}}{2(k+1)}\sum_{r=1}^{\infty}(-1)^r(\rho_{t+r} - \rho_{t-r+1})\,\theta_{2(k+1),2r-1}$$

Note that the variable t in these equations is an integer.

c. Relationship to the incomplete gamma function

Consider eqn. (2.15b) for a = 1

$$\psi_n(x) = x^{-(n+1)} \int_0^x \xi^n e^{i(\xi - x)} d\xi \tag{9.23}$$

Transforming now from the real axis to the imaginary axis by the substitution

$$\xi = i\, t \tag{9.24}$$

we get

$$\psi_n(x) = \frac{e^{-ix}}{(-ix)^{n+1}} \int_0^{-ix} e^{-t} t^n dt \tag{9.25}$$

$$= \frac{e^{-ix}}{(-ix)^{n+1}} \gamma(n+1, -ix) \tag{9.26}$$

where $\gamma(a,z)$ is the incomplete gamma function [5]

$$\gamma(a,z) = \int_0^z e^{-t} t^{a-1} dt \tag{9.27}$$

We conclude therefore that the η functions are instrumental in evaluating the incomplete gamma function for an integral first argument and an imaginary second argument. Rearranging (9.26) and using (2.13), (5.20), we get explicitly

$$\gamma(n, ix) = (ix)^n e^{-ix} \left\{ \frac{\theta_n(x)}{n} - i\, \frac{\eta_{n-1}(x)}{x} \right\} \tag{9.28}$$

For the related function [5]

$$(9.29) \qquad \gamma^*(a,z) = \frac{z^{-a}}{\Gamma(a)} \gamma(a,z) ,$$

we get

$$(9.30) \qquad \gamma^*(n,ix) = \frac{e^{-ix}}{(n-1)!} \left\{ \frac{\theta_n(x)}{n} - i \frac{\eta_{n-1}(x)}{x} \right\}$$

The similarity between (9.28) and (9.14), suggests a direct relationship between the derivatives of section 9b and $\gamma(n,ix)$. Indeed, combining (9.26) and (9.13) yields

$$(9.31) \qquad \frac{d^n\left(\frac{\sin x}{x}\right)}{dx^n} = (-x)^{-n-1} \operatorname{Im}[\gamma(n+1,-ix)]$$

$$(9.32) \qquad \frac{d^n\left(\frac{\cos x}{x}\right)}{dx^n} = (-x)^{-n-1} \left\{ \operatorname{Re}[\gamma(n+1,-ix)] - n! \right\}$$

10. An Independent Verification

One type of test to verify the claimed precision of the generated η functions was considered in section 8. An independent test of a quite different nature will be described here. This test is based on a formula relating η_{nk} to Riemann's zeta function and various other similar functions. The method is essentially constrained to the testing of $\eta_{2n,1}$, $\eta_{2n+1,2}$, $\eta_{2n+1,4}$. Note, however, that for any $m > 4$, $\eta_{m,1}$ is the value most difficult to compute directly (see analysis associated with (11.4)).

We establish the above relationship via the notion of band-limited functions mentioned in section 9b. Consider the following complex function

$$\rho(\tau) = e^{i2\pi\nu_o\tau} \qquad (|\nu_o| \leq \tfrac{1}{2}) \tag{10.1}$$

Its n-th order derivative is obtainable by direct differentiation

$$\frac{d^n\rho(\tau)}{d\tau^n} = (i2\pi\nu_o)^n \, e^{i2\pi\nu_o\tau} \tag{10.2}$$

Now, the Fourier transform of $\rho(\tau)$ is the delta function $\delta(\nu-\nu_o)$ which vanishes for $|\nu| > \frac{1}{2}$ (see ν_o constraint in (10.1)). This means that $\rho(\tau)$ is a band-limited function as defined in section 9b and as such, its n-th order derivative is obtainable from (9.22). Applying then (9.22) to $\rho(\tau)$ and equating the result to (10.2), we obtain the following identities ((10.4a) is derived from (9.22a), (10.4b) from (9.22b), etc.)

$$\sum_{r=1}^{\infty}(-1)^r \cos\,(2\pi r\nu_o)\theta_{2k+1,2r} = (2k+1)2^{2k-1}\nu_o^{2k} - \tfrac{1}{2} \qquad (|\nu_o| \leq \tfrac{1}{2}) \tag{10.4a}$$

$$(10.4b)\quad \sum_{r=1}^{\infty}(-1)^r \sin(2\pi r\nu_o)\,\frac{\eta_{2k+1,2r}}{r} = -\pi\, 4^k \nu_o^{2k+1} \qquad (|\nu_o| \le \tfrac{1}{2})$$

$$(10.4c)\quad \sum_{r=1}^{\infty}(-1)^r \cos[2\pi(r-\tfrac{1}{2})\nu_o]\,\frac{\eta_{2k,2r-1}}{2r-1} = -\pi\, 4^{k-1}\nu_o^{2k} \qquad (|\nu_o| \le \tfrac{1}{2})$$

$$(10.4d)\quad \sum_{r=1}^{\infty}(-1)^r \sin[2\pi(r-\tfrac{1}{2})\nu_o]\quad \theta_{2k,2r-1} = -k(2\nu_o)^{2k-1} \qquad (|\nu_o| \le \tfrac{1}{2})$$

We are particularly interested in the special cases of these equations in which the trigonometric multipliers are simplified. Obviously, such desirable summation formulae are generated when ν_o is constrained to the following values

$$(10.5)\qquad \nu_o = 0,\ \tfrac{1}{4},\ \tfrac{1}{2}$$

Substituting these in (10.4) we obtain the following results

$$(10.6a)\quad T^{(1)}_{2k+1,0} \equiv \sum_{r=1}^{\infty} \theta_{2k+1,2r} = k \qquad (\text{from:}(10.4a);\ \nu_o=\tfrac{1}{2})$$

$$(10.6b)\quad T^{(2)}_{2k,0} \equiv \sum_{r=1}^{\infty} \theta_{2k,2r-1} = k \qquad (\text{from:}(10.4d);\ \nu_o=\tfrac{1}{2})$$

$$(10.6c)\quad T^{(3)}_{2k+1,0} \equiv \sum_{r=1}^{\infty} (-1)^{r+1}\theta_{2k+1,2r} = \tfrac{1}{2} \qquad (\text{from:}(10.4a);\ \nu_o=0)$$

$$(10.6d)\quad T^{(4)}_{2k+1,0} \equiv \sum_{r=1}^{\infty} (-1)^{r+1}\theta_{2k+1,4r} = \tfrac{1}{2}\left(1-\frac{2k+1}{4^k}\right) \qquad (\text{from:}(10.4a);\ \nu_o=\tfrac{1}{4})$$

$$(10.6e)\quad T^{(5)}_{2k,0} \equiv \sum_{r=1}^{\infty} (-1)^{[\frac{r}{2}]}\theta_{2k,2r-1} = 2\sqrt{2}\,\frac{k}{4^k} \qquad (\text{from:}(10.4d);\ \nu_o=\tfrac{1}{4})$$

$$(10.6f)\quad E^{(6)}_{2k,1} \equiv \sum_{r=1}^{\infty} (-1)^{r+1} \frac{\eta_{2k,2r-1}}{2r-1} = 0 \qquad \text{(from:(10.4c); } \nu_o=0)$$

$$(10.6g)\quad E^{(7)}_{2k-1,1} \equiv \sum_{r=1}^{\infty} (-1)^{r+1} \frac{\eta_{2k-1,4r-2}}{4r-2} = \frac{\pi}{2^{2k+1}} \qquad \text{(from:(10.4b);} \nu_o=\tfrac{1}{4})$$

$$(10.6h)\quad E^{(8)}_{2k,1} \equiv \sum_{r=1}^{\infty} (-1)^{\left[\frac{r-1}{2}\right]} \frac{\eta_{2k,2r-1}}{2r-1} \quad \frac{\sqrt{2}\pi}{4^{k+1}} \qquad \text{(from:(10.4c); } \nu_o=\tfrac{1}{4})$$

Though these summation formulae are quite interesting, they are not particularly suitable for the contemplated high-precision testing because of their relatively slow convergence. They do provide, however, the starting point for a recursive scheme that does yield rapidly converging summation formulae.

As already hinted by the notation in (10.6), all the sums appearing there, may be regarded as special cases of one of the following two sums

$$(10.8)\qquad E^{(i)}_{nr} = \sum_{k=1}^{\infty} v^{(i)}_{k} \frac{\eta_{nk}}{k^r}$$

$$(10.9)\qquad T^{(i)}_{nr} = \sum_{k=1}^{\infty} v^{(i)}_{k} \frac{\theta_{nk}}{k^r}$$

in which the range of all $v^{(i)}_{k}$ functions is the set $\{-1, 0, 1\}$.

Fast convergence is displayed by $E^{(i)}_{nr}$, $T^{(i)}_{nr}$ with large r. These are not available but may be computed from the available $E^{(i)}_{n1}$, $T^{(i)}_{n0}$ (10.6) by applying the recursion (3.14). Writing down this recursion for $x = k\frac{\pi}{2}$,

$$(10.10)\qquad \frac{\eta_{n-2,k}}{k^2} = \frac{\pi^2}{4n(n-1)} \left(1 - \eta_{nk}\right)$$

multiplying both sides by $v_k^{(i)}/k^r$ and summing over k, we get

$$(10.11) \qquad E_{n-2,r+2}^{(i)} = \frac{\pi^2}{4n(n-1)} \left[\sum_{k=1}^{\infty} \frac{v_k^{(i)}}{k^r} - E_{nr}^{(i)} \right]$$

Denoting

$$(10.12) \qquad H_r^{(i)} = \sum_{k=1}^{\infty} \frac{v_k^{(i)}}{k^r}$$

yields the final form

$$(10.13) \qquad E_{n-2,r+2}^{(i)} = \frac{\pi^2}{4n(n-1)} \left(H_r^{(i)} - E_{nr}^{(i)} \right)$$

Note that the introduction of the sums $H_r^{(i)}$ (10.12) facilitates a transformation between $E_{nr}^{(i)}$ and $T_{nr}^{(i)}$. Applying (5.21b) to (10.9), we obtain

$$(10.14) \qquad T_{nr}^{(i)} = \sum_{k=1}^{\infty} v_k^{(i)} \frac{1 - \eta_{nk}}{k^r} = H_r^{(i)} - E_{nr}^{(i)}$$

Applying this to (10.13) we get the recurrence for $T_{nr}^{(i)}$

$$(10.15) \qquad T_{n-2,r}^{(i)} = H_r^{(i)} - \frac{\pi^2}{4n(n-1)} T_{n,r-2}^{(i)}$$

The functions $v_r^{(i)}$ appearing in (10.8), (10.9), (10.12) are implicitly defined in (10.6). Therefore, we can obtain the specific forms of the sums $H_r^{(i)}$ by inspecting these equations. The results are listed below:

$$(10.16a) \qquad H_r^{(1)} = \sum_{k=1}^{\infty} \frac{1}{(2k)^r} = \frac{1}{2^r} + \frac{1}{4^r} + \frac{1}{6^r} + \dots \qquad = 2^{-r}\zeta(r) = 2^{-r}S_r$$

$$(10.16b) \quad H_r^{(2)} = \sum_{k=1}^{\infty} \frac{1}{(2k-1)^r} = \frac{1}{1^r} + \frac{1}{3^r} + \frac{1}{5^r} + \ldots = U_r$$

$$(10.16c) \quad H_r^{(3)} = \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{(2k)^r} = \frac{1}{2^r} - \frac{1}{4^r} + \frac{1}{6^r} - \ldots = 2^{-r} s_r$$

$$(10.16d) \quad H_r^{(4)} = \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{(4k)^r} = \frac{1}{4^r} - \frac{1}{8^r} + \frac{1}{12^r} - \ldots = 4^{-r} s_r$$

$$(10.16e) \quad H_r^{(5)} = \sum_{k=1}^{\infty} \frac{(-1)^{[\frac{k}{2}]}}{(2k-1)^r} = \frac{1}{1^r} - \frac{1}{3^r} - \frac{1}{5^r} + \frac{1}{7^r} + \frac{1}{9^r} - - + + \ldots = q_r$$

$$(10.16f) \quad H_r^{(6)} = \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{(2k-1)^r} = \frac{1}{1^r} - \frac{1}{3^r} + \frac{1}{5^r} - \ldots = u_r$$

$$(10.16g) \quad H_r^{(7)} = \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{(4k-2)^r} = \frac{1}{2^r} - \frac{1}{6^r} + \frac{1}{10^r} - \ldots = 2^{-r} u_r$$

$$(10.16h) \quad H_r^{(8)} = \sum_{k=1}^{\infty} \frac{(-1)^{[\frac{k-1}{2}]}}{(2k-1)^r} = \frac{1}{1^r} + \frac{1}{3^r} - \frac{1}{5^r} - \frac{1}{7^r} + + - - \ldots = p_r$$

The symbols on the right correspond to the notation adopted by Fletcher et al [10]. High precision tables are available for most of these functions.

The application of the above results to our testing scheme is based on equations (10.8), (10.13), (10.6). Eqn. (10.8) specifies $E_{nr}^{(i)}$ directly in terms of η_{nk}. For a sufficiently large r, this is mainly a function of the first η_{nk} in the sum, that is, $\eta_{n,1}$ or $\eta_{n,2}$. Eqn. (10.13), on the other hand, determines this same $E_{nr}^{(i)}$ through a recursion utilizing $H_r^{(i)}$ and starting with the $E_{n,1}^{(i)}$ specified by (10.6). The result is an $E_{nr}^{(i)}$ value which is a function of the tabulated $H_r^{(i)}$ functions and is totally independent of the computed η_{nk}'s. The test consists of comparing these two values of $E_{nr}^{(i)}$.

Note that when the recursion (10.13) generating $E_{nr}^{(i)}$ is used with $n \gg \frac{\pi}{2}$, the error propagation properties are very favorable. Adopting this constraint in our test, ensures the high precision needed.

The same approach could be applied to $T_{nr}^{(i)}$ leading to a test of $\theta_{n,1}$, $\theta_{n,2}$, $\theta_{n,4}$ (see (10.6)). We note, however, that for large n, these θ_{nk} values are close to 1. As indicated in section 5, our algorithm actually computes in this case a high precision $\eta_{nk} \ll 1$ and the θ_{nk} required for the test would be obtained by subtracting this η_{nk} from 1. In this subtraction, many of the bits of η_{nk} are "shifted out" so that even if we find out that all 140 bits of θ_{nk} are correct, this tells us nothing about the precision of the "shifted out" bits of η_{nk}.

We conclude that conversion of the θ_{nk} sums in (10.6) to η_{nk} sums, will give us a more stringent test. Applying (5.21c)

(10.17a) $$\frac{\eta_{n-2,k}}{k^2} = \frac{\pi^2}{4n(n-1)} \cdot \theta_{nk},$$

multiplying by $v_k^{(i)}/k^r$ and summing over k, we get

(10.17b) $$E_{nr}^{(i)} = \frac{\pi^2}{4(n+1)(n+2)} T_{n+2,r-2}^{(i)}$$

Applying this to (10.6) yields our results:

(10.18a) $$E_{2k-1,2}^{(1)} = \sum_{r=1}^{\infty} \frac{\eta_{2k-1,2r}}{(2r)^2} = \frac{\pi^2}{8(2k+1)}$$

$$(10.18b) \quad E^{(2)}_{2k,2} = \sum_{r=1}^{\infty} \frac{\eta_{2k,2r-1}}{(2r-1)^2} = \frac{\pi^2}{8(2k+1)}$$

$$(10.18c) \quad E^{(3)}_{2k-1,2} = \sum_{r=1}^{\infty} (-1)^{r+1} \frac{\eta_{2k-1,\,2r}}{(2r)^2} = \frac{\pi^2}{16k(2k+1)}$$

$$(10.18d) \quad E^{(4)}_{2k-1,2} = \sum_{r=1}^{\infty} (-1)^{r+1} \frac{\eta_{2k-1,4r}}{(4r)^2} = \frac{\pi^2}{16k} \left[\frac{1}{2k+1} - \frac{1}{4^k}\right]$$

$$(10.18e) \quad E^{(5)}_{2k,2} = \sum_{r=1}^{\infty} (-1)^{[\frac{r}{2}]} \frac{\eta_{2k,2r-1}}{(2r-1)^2} = \frac{\sqrt{2}\,\pi^2}{4^{k+2}(2k+1)}$$

We consider now a specific example to illustrate the testing procedure. We base our example on $H^{(3)}_r = 2^{-r}s_r$ for which an extensive 50 decimals table is available [1]. Starting with (10.18c)

$$(10.19a) \quad E^{(3)}_{151,2} = \frac{\pi^2}{16\cdot 76\cdot 153},$$

we apply (10.13) to obtain $E^{(3)}_{101,52}$. Having obtained this value recursively, we turn now to its direct evaluation as an infinite sum. From (10.8), (10.18c) we determine its explicit form

$$(10.19b) \quad E^{(3)}_{101,52} = \frac{\eta_{101,2}}{2^{52}} - \frac{\eta_{101,4}}{4^{52}} + \frac{\eta_{101,6}}{6^{52}} - \ldots$$

Obviously, a few terms will suffice to get a highly precise estimate of this infinite sum. To determine just how many, let the first neglected term be

$$(10.19c) \quad (-1)^{r+1} \frac{\eta_{101,2r}}{(2r)^{52}}$$

Note that (10.19b) is an alternating series whose sum is very close to its first term. Therefore the ratio of (10.19c) to the first term is a good approximation to the relative error in the truncated estimate of $E^{(3)}_{101,52}$.
As the computations are performed in penta-precision, we bound this error by 2^{-140}, getting (with (4.39), (4.19c)) the following condition for r

$$(10.20) \qquad 2^{-140} > \frac{\eta_{101,2r}}{\eta_{101,2}}\, r^{-52} > \frac{1 + \dfrac{1}{\eta_{99}(\pi)}}{1 + \dfrac{1}{\eta_{101}(\pi r)}}\, r^{-52} \approx r^{-50}$$

$$(10.21) \qquad \therefore\ r > 2^{2.8} \approx 7$$

We conclude therefore that (10.19b) may be truncated after the seventh term. In the actual computation of the sum, 9 terms were used and the partial sum was octally printed each time a new term was added. These partial sums were found to be affected by the added terms only up to the seventh term. The eighth and ninth terms had no effect in agreement with the above analysis.

Comparing the two values of $E^{(3)}_{101,52}$ we found that their fraction parts differed only by $2 \cdot 2^{-140}$. Our conclusion is therefore that the relative error in the computed value of $\eta_{101,2}$ is well within the bound prescribed by (8.12).

Note that we cannot make here any useful claim regarding $\eta_{101,4}$, the next term in (10.19b). This is due to the fact that the ratio of these two terms is about 2^{52} so that in accumulating the sum, roughly 52 of the least significant bits of the second term are "shifted out". This is equivalent to saying that the normalized 140 bit value of $E^{(3)}_{101,52}$ is independent of the 52 least significant bits of $\eta_{101,4}$.

The same argument holds even more forcefully with respect to $\eta_{101,6}$ and so on. We may say then that we have here a test of $\eta_{101,2}$ which relies on successively coarser estimates of $\eta_{101,4}$, $\eta_{101,6}$, ..., $\eta_{101,12}$ and is completely independent of the remaining terms of this infinite sequence.

What we have seen in this specific case is true in general, that is, the application of this testing procedure is limited to the testing of only the first term of $E^{(i)}_{nr}$.

Equally encouraging results were obtained with other cases. The results are summarized in Table I.

Table 1

Test Results

Tested parameter	Sum involved in test	Difference of fraction parts
$\eta_{101,2}$	$E^{(1)}_{101,52}$	$1 \cdot 2^{-140}$
$\eta_{100,1}$	$E^{(2)}_{100,68}$	0
$\eta_{101,2}$	$E^{(3)}_{101,52}$	$2 \cdot 2^{-140}$
$\eta_{101,4}$	$E^{(4)}_{101,52}$	0
$\eta_{100,1}$	$E^{(6)}_{100,40}$	$1 \cdot 2^{-140}$
$\eta_{101,2}$	$E^{(7)}_{101,40}$	0

The only test types not tried were those employing $H^{(5)}_r$, $H^{(8)}_r$ (see (10.16)) for which tables were not available.

11. Concluding Remarks

The one theme, central to all of this paper is the function $\eta_n(x)$. We have shown how it arises in connection with the problem considered, we have studied some of its properties and finally applied these to its precise computation.

Two important questions regarding this function are yet to be answered though. The first concerns its evaluation. We have seen in section 3 that $\eta_n(x)$ can be expressed in a simple form involving a finite sum (3.1) which we have actually applied to determine $\eta_n(x)$ for $n \leq 3$. Why couldn't all $\eta_n(x)$ be evaluated this way?

We have indicated earlier that severe "cancellation" is to be expected for certain n,x combinations. Having established the necessary prerequisites since then, we are now in position to substantiate this statement and show how severe this "cancellation" is.

Eqn. (3.1) prescribes $\eta_n(x)$ as a finite sum of terms. Consider now tha ratio of each of these terms to $\eta_n(x)$ and denote the maximal magnitude of these ratios, $\lambda_n(x)$. The usefulness of (3.1) as a prescription for computation, strongly depends on $\lambda_n(x)$. The larger $\lambda_n(x)$, the more severe the "cancellation", and the lower the number of meaningful digits in the computed $\eta_n(x)$. In view of this, we proceed now to establish a lower bound for $\lambda_n(x)$.

We note first that since $\eta_n(x) < \hat{\eta}_n(x)$ (4.40) we shall still obtain a lower bound for $\lambda_n(x)$ if we divide the terms of (3.1) by $\hat{\eta}_n(x)$ rather than by $\eta_n(x)$. This simplifies our task to that of determining the series expansion of $\frac{\eta_n(x)}{\hat{\eta}_n(x)}$. Applying (4.19c) and (2.3) to (3.1), we get

$$(11.2a) \quad \frac{\eta_{2m}(x)}{\hat{\eta}_{2m}(x)} = (-1)^{m+1} \left\{ \frac{(2m+2)!}{x^{2m+2}} (\cos x - 1) + \frac{1}{2} \frac{(2m+2)!}{x^{2m}} - \dots \right\}$$

$$(11.2b) \quad \frac{\eta_{2m+1}(x)}{\hat{\eta}_{2m+1}(x)} = (-1)^{m+1} \left\{ \frac{(2m+3)!}{x^{2m+2}} \left(\frac{\sin x}{x} - 1 \right) + \frac{1}{2} \frac{(2m+2)!}{x^{2m}} (1 + \frac{2}{3} m) - \dots \right\}$$

In determining a meaningful lower bound for $\lambda_n(x)$ from (11.2), it should be borne in mind that there are applications where x is an exact integral multiple of $\frac{\pi}{2}$. In some of these cases the $x^{-(2m+2)}$ term in (11.2) will simply be omitted. It is, therefore, safer and certainly legitimate to base the $\lambda_n(x)$ bound on the x^{-2m} term of (11.2). An obvious such bound is

$$(11.3) \quad \lambda_n(x) \geq \frac{1}{2} \frac{(2m+2)!}{x^{2m}} \qquad (m = [\tfrac{n}{2}])$$

On applying Stirling's formula (6.26), this yields

$$(11.4) \quad \lambda_n(x) > 2\sqrt{\pi m}\, (m+1)(2m+1) \left(\frac{2m}{ex} \right)^{2m} \qquad (m = [\tfrac{n}{2}])$$

It follows that for $n > ex$, $\lambda_n(x)$ increases very sharply with n and soon attains astronomical proportions. Consider just two examples

$$(11.5a) \quad \lambda_{30}(3) \geq 6.40 \cdot 10^{20} = 1.56 \cdot 2^{68}$$

$$(11.5b) \quad \lambda_{50}(3) \geq 5.63 \cdot 10^{43} = 1.30 \cdot 2^{144}$$

This means that a computer like the IBM 7094, for example, would not be able to compute $\eta_{30}(3)$ even when employing double precision arithmetic. Furthermore, even the use of penta-precision arithmetic would not allow computation of $\eta_{50}(3)$. In contrast, note that the specific η_{nk}'s tested in section 10 indicate a relative error of less than 2^{-138} while the associated λ has a lower bound of 2^{199} in this case.

In view of the relationship between $\eta_n(x)$ and $\gamma(n,z)$, the incomplete gamma function (9.28), the above discussion is pertinent to the computation of this function on the imaginary axis (imaginary z). It is (correctly) stated that $\gamma(n,z)$ is an elementary function for imaginary z. In view of the above discussion we may comment that it is elementary all right but just as incomputable by elementary methods as any "non-elementary" function.

The second question to be considered here, concerns the definition adopted for $\eta_n(x)$ as an infinite sum (2.11). We recall that the natural formulation of the problem (2.1) actually leads to a finite sum and the infinite summation formulation was obtained by a device (2.5) which, though legitimate, seems quite arbitrary and unwarranted. We consider now the motivation for this step. Our method consists of a brief description of the results obtained when this step is omitted.

Retracing the derivations in section 2 with the objective of a finite summation formulation, one is led to the following function replacing $\eta_n(x)$

$$E_n(x) = \frac{\sum_{k=0}^{[\frac{n}{2}]} (-1)^k g_{n-2k}(x)}{g_n(x)} = \begin{cases} (-1)^{[\frac{n}{2}]} \dfrac{\sum_{k=0}^{[\frac{n}{2}]} (-1)^k g_{2k}(x)}{g_n(x)} & \text{(n even)} \\ (-1)^{[\frac{n}{2}]} \dfrac{\sum_{k=0}^{[\frac{n}{2}]} (-1)^k g_{2k+1}(x)}{g_n(x)} & \text{(n odd)} \end{cases} \tag{11.6}$$

This function satisfies a recurrence relation identical to that of $\eta_n(x)$ so there should be no problem in its computation. Comparing now (11.6) to (3.1) we see that they differ only in the absence of the trigonometric functions in (11.6). Specifically,

$$E_n(x) - \eta_n(x) = \begin{cases} (-1)^{[\frac{n}{2}]} \dfrac{n!}{x^n} \cos x & \text{(n even)} \\ (-1)^{[\frac{n}{2}]} \dfrac{n!}{x^n} \sin x & \text{(n odd)} \end{cases} \tag{11.7}$$

This means that the curve of $E_n(x)$ vs. x will oscillate about the monotonic $\eta_n(x)$ with amplitude $\frac{n!}{x^n}$. We have already seen that for $n > ex$ such a term may attain astronomical values. Thus, we have here a function which, though sharing the same value with $\eta_n(x)$ at regular intervals, is widely different from it at others and therefore has extremely large derivatives.

To see the bearing this has on the problem considered, we apply (11.7) to the second case of (2.19) under the condition

$$n \gg \frac{k\pi e}{2} \tag{11.8}$$

getting

$$(11.9) \qquad \frac{a^{n+1}}{(k\frac{\pi}{2})^n}\int_0^{\frac{\pi}{2}\frac{k}{a}} \xi^n \cos a\xi d\xi = \eta_{nk} = E_n(k\frac{\pi}{2}) - (-1)^{[\frac{n}{2}]}\frac{n!}{(k\frac{\pi}{2})^n} \qquad (n \text{ odd}; k \bmod 4 = 1)$$

Note that the properties of $\eta_n(x)$ (4.18) bound the above expression by 1. On the other hand, the term involving n! will be extremely large under (11.8). Hence, the integral is being evaluated in this case as the small difference of two, almost equal, extremely large numbers. The same situation obtains for all other cases of (2.19), (2.20) in which n+k is even. We conclude therefore that all the integrals in (2.19), (2.20), subject to

$$(11.10) \qquad n+k \text{ even}; \; n \gg \frac{k\pi e}{2}$$

are essentially incomputable via $E_n(x)$. Note that the trouble is not with the computability of $E_n(x)$ itself which, unless we run into exponent overflow problems, can certainly be computed through its recursion. Rather, the difficulty lies in the unsuitability of $E_n(x)$ to the problem at hand.

When n+k is odd, $E_n(x)$ can be used to evaluate the integrals as this situation corresponds to the intersection of the $E_n(x)$ and $\eta_n(x)$ curves. Even in this case, however, $E_n(x)$ poses problems since, under (11.8), its derivative at these points is extremely large. This does not affect the actual computations but makes it quite difficult to justify them.

There is no question then that $\eta_n(x)$ is the better instrument to tackle the problem at hand.

We turn now to the third and last topic: coarse approximations. While our main concern throughout this paper has been with the high precision determination of $\eta_n(x)$, the results established are also applicable to the quite different situation in which only a coarse estimate of $\eta_n(x)$ is desired. We refer specifically to the bounds (4.39), (4.40) which are the basis of the normalized bounding curves shown in Figure 9. Expressing these bounds in terms of the entities

$$y(\nu) = \frac{\nu^2}{1+\nu^2} \tag{11.11}$$

$$x_n = \sqrt{n(n-1)} \quad \text{(that is, } e_n(x_n) = 1\text{)} \tag{11.12}$$

$$h_n = \sqrt{\left(1+\frac{2}{n}\right)\left(1+\frac{2}{n-1}\right)}\ , \tag{11.13}$$

we obtain

$$y\left(\frac{x}{x_{n+2}}\right) < \eta_n(x) < y\left(h_n \frac{x}{x_{n+2}}\right) \qquad \begin{pmatrix} n > 1;\ \text{lower bound} \\ n > 3;\ \text{upper bound} \\ x \neq 0 \end{pmatrix} \tag{11.14a}$$

$$\eta_n(x) < \left(\frac{x}{x_{n+2}}\right)^2 \qquad (n > 1;\ x \neq 0) \tag{11.14b}$$

Thus, the bounds on $\eta_n(x)$ can be expressed in terms of the family of curves $y(h_n\nu)$ and the curve ν^2 shown in Figure 9. Setting

$$\nu = \frac{x}{x_{n+2}}\ , \tag{11.15}$$

we see that $\eta_n(x)$ is constrained to the region defined by the following three curves of Figure 9.

left boundary	ν^2	$(n > 1)$
bottom boundary	$y(\nu)$	$(n > 1)$
top boundary	$y(h_n\nu)$	$(n > 3)$

Note that as n increases, the top boundary approaches the bottom boundary, thereby reducing the maximal relative error of an $\eta_n(x)$ estimate based on Figure 9. Note also that for a given n, this error bound is a function of x attaining its highest value at the intersection of the left and upper boundaries at the point

$$\nu^2 = 1 - \frac{1}{h_n^2} \qquad (y(h_n\nu) = \nu^2) \tag{11.16}$$

The same family of curves is also useful in dealing with $\theta_n(x)$ (directly). Application of (5.15) to (11.14a) yields

$$y\left(\frac{x_n}{x}\right) < \theta_n(x) < y\left(h_n \frac{x_n}{x}\right) \qquad \begin{pmatrix} n > 3;\ \text{lower bound} \\ n > 1;\ \text{upper bound} \\ x \neq 0 \end{pmatrix} \tag{11.17a}$$

while (5.20), (4.18), (11.12) imply

$$\theta_n(x) < \left(\frac{x_n}{x}\right)^2 \qquad (n > 3) \tag{11.17b}$$

Note the similarity of (11.17) to (11.14). Obviously Figure 9 is applicable in this case with

$$\nu = \frac{x_n}{x} \tag{11.18}$$

We illustrate the application of (11.14), (11.17) with the estimation of $\eta_{10}(x)$. Our main concern here is not with the estimate itself but rather with the accompanying error bound. From (11.16),(11.13) we find that the highest error bound will be associated with

$$\nu_1^2 = 1 - \frac{1}{h_{10}^2} = 0.32$$

that is ((11.15), (11.12)),

$$x = \nu_1 x_{12} = 6.5$$

Thus, the worst case in estimating $\eta_{10}(x)$ via (11.14) is the estimation of $\eta_{10}(6.5)$. The ratio of the two bounds of $\eta_{10}(x)$ at this point is given by

$$\frac{\nu_1^2}{y(\nu_1)} = 1 + \nu_1^2 = 1.32$$

Therefore, if we adopt the geometric mean of the two bounds as our estimate, we are assured that the true $\eta_{10}(6.5)$ lies within $\pm$ 15% of this value ($\sqrt{1.32} = 1.149 \approx 1.15$)

A better estimate will be obtained in this case by proceeding indirectly via (5.20)

$$\eta_n(x) = \left(\frac{x}{x_{n+2}}\right)^2 \theta_{n+2}(x) \tag{11.19}$$

In estimating $\theta_{12}(6.5)$ we have to enter Figure 9 this time at the value (11.18)

$$\nu_2 = \frac{x_{12}}{6.5} = 1.768$$

The ratio of the two bounds of $\theta_{12}(x)$ in this case is

$$\frac{y(h_{12}\nu_2)}{y(\nu_2)} = 1.071$$

Therefore, $\theta_{12}(6.5)$ can be estimated with an error of $\pm$ 3.5% ($\sqrt{1.071} = 1.035$). In view of (11.19), this will also be the error bound of $\eta_{10}(6.5)$ obtained from $\theta_{12}(6.5)$ via this formula.

There are various other possibilities of improving the estimate by exploiting the error reduction properties of the basic recurrence relation. Indeed, by a judicious choice of the number of recursion cycles, it is possible to obtain any type of estimate, ranging all the way from the coarse ones indicated above to the highly precise ones of the table described in section 8.

References

[1] R. Liénard, Tables Fondamentales à 50 décimales Des Sommes S_n, u_n, Σ_n, Paris, 1948.

[2] I.S. Gradshtein and I.M. Ryzhik, Tables of Integrals, Series, and Products, 4th edition, Academic Press, New York, 1965(Formula 2.321-2)

[3] "Bessel Functions, Pt. II., Functions of Positive Integer Order," Vol. X of series "Mathematical Tables", British Association for the Advancement of Science, The University Press, Cambridge, England, 1931-46.

[4] M. Abramowitz and I. Stegun, Handbook of Mathematical Functions, National Bureau of Standards, Applied Mathematics Series. 55, March 1965, Table 6.3

[5] ---Ibid. p. 260.

[6] ---Ibid. Tables 26.7, 26.8

[7] C.L. Lawson, Summary of Q-precision subroutines, as revised in October, 1968, Jet Propulsion Laboratory, Tech. Memo. 211, 1/7/69.

[8] W. Gautschi, Computation of Successive Derivatives of f(z)/z, Math. Comput. Vol. 20, 1966, pp. 209-214.

[9] ---Computational Aspects of Three-Term Recurrence Relations, SIAM Review, Vol. 9, No. 1, January 1967, pp. 74-77.

[10] A. Fletcher et al., An Index of Mathematical Tables, Second Edition, Addison-Wesley, Reading, Massachusetts, 1962, Section 4.6.

[11] ---Ibid., Section 5.1.

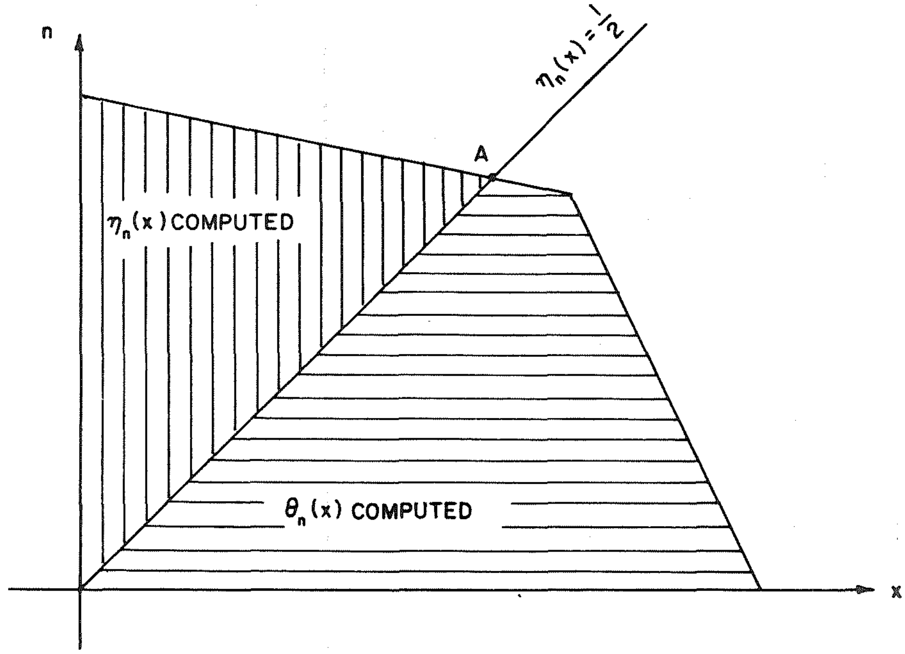

Fig. 1. An illustration of the point of maximal error bound

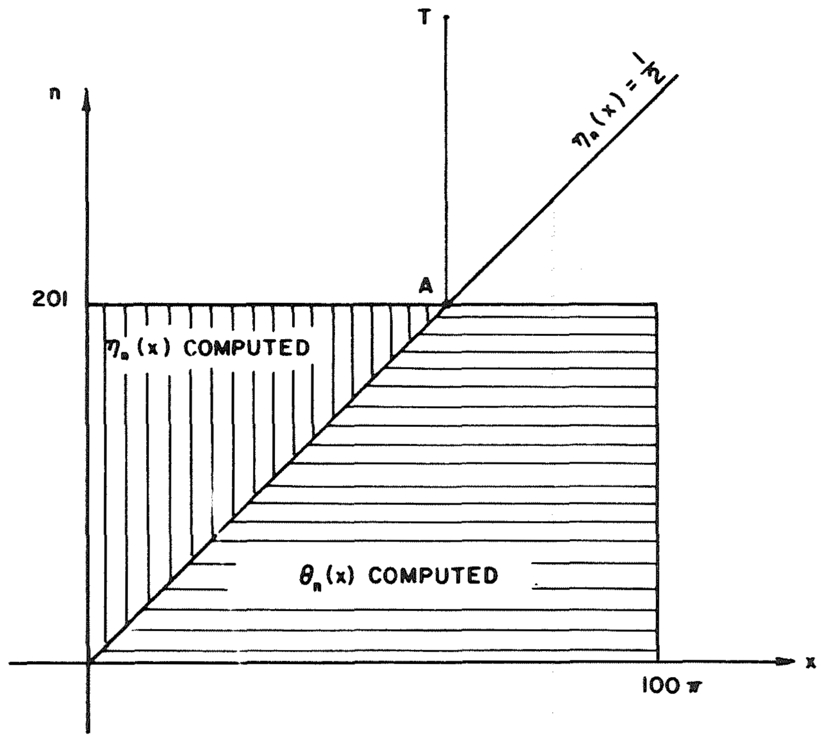

Fig. 2. Parameters of the η_{nk} table described in section 8

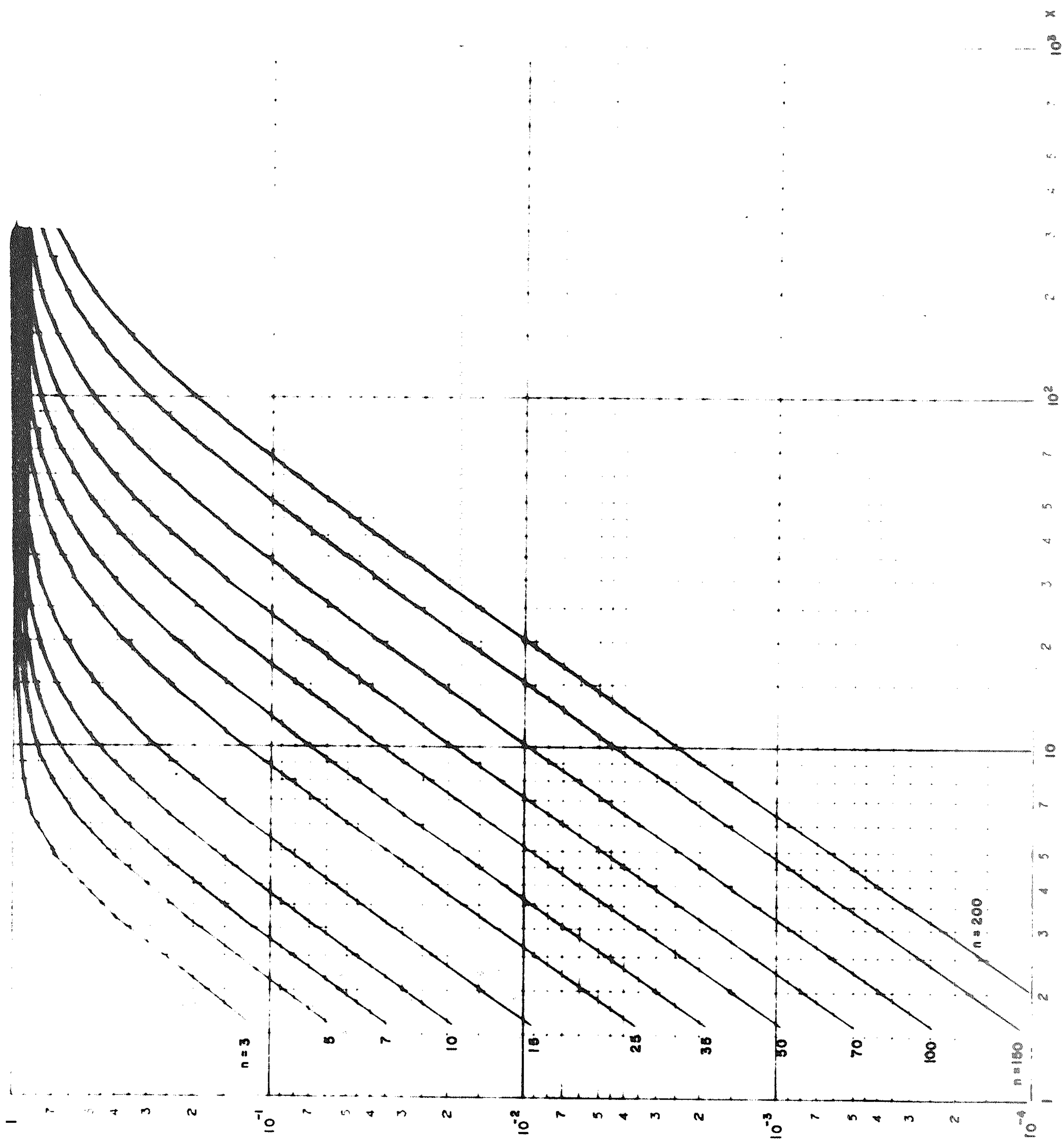

Fig. 3. $\eta_n(x)$

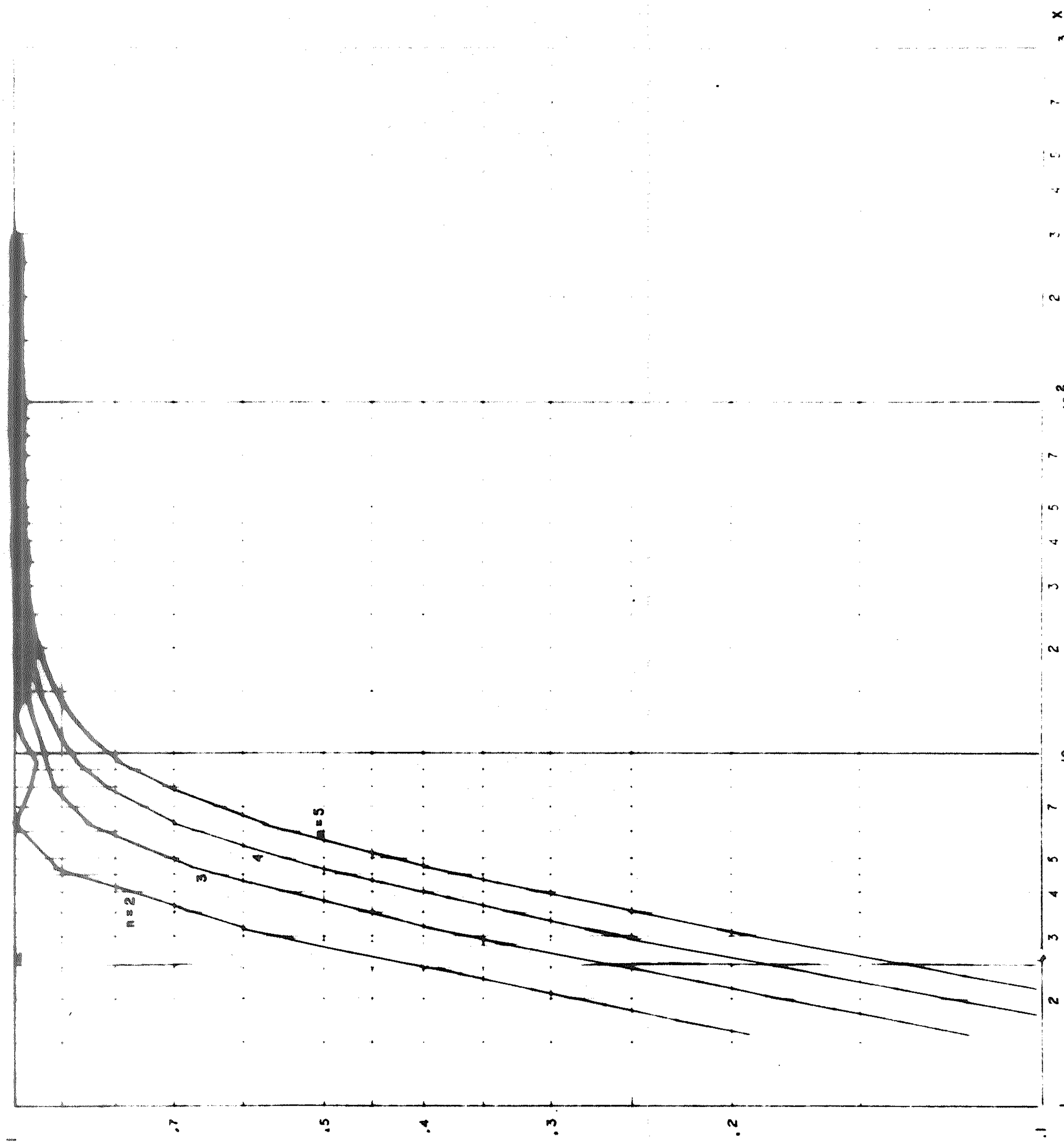

Fig. 3a. $\eta_n(x)$

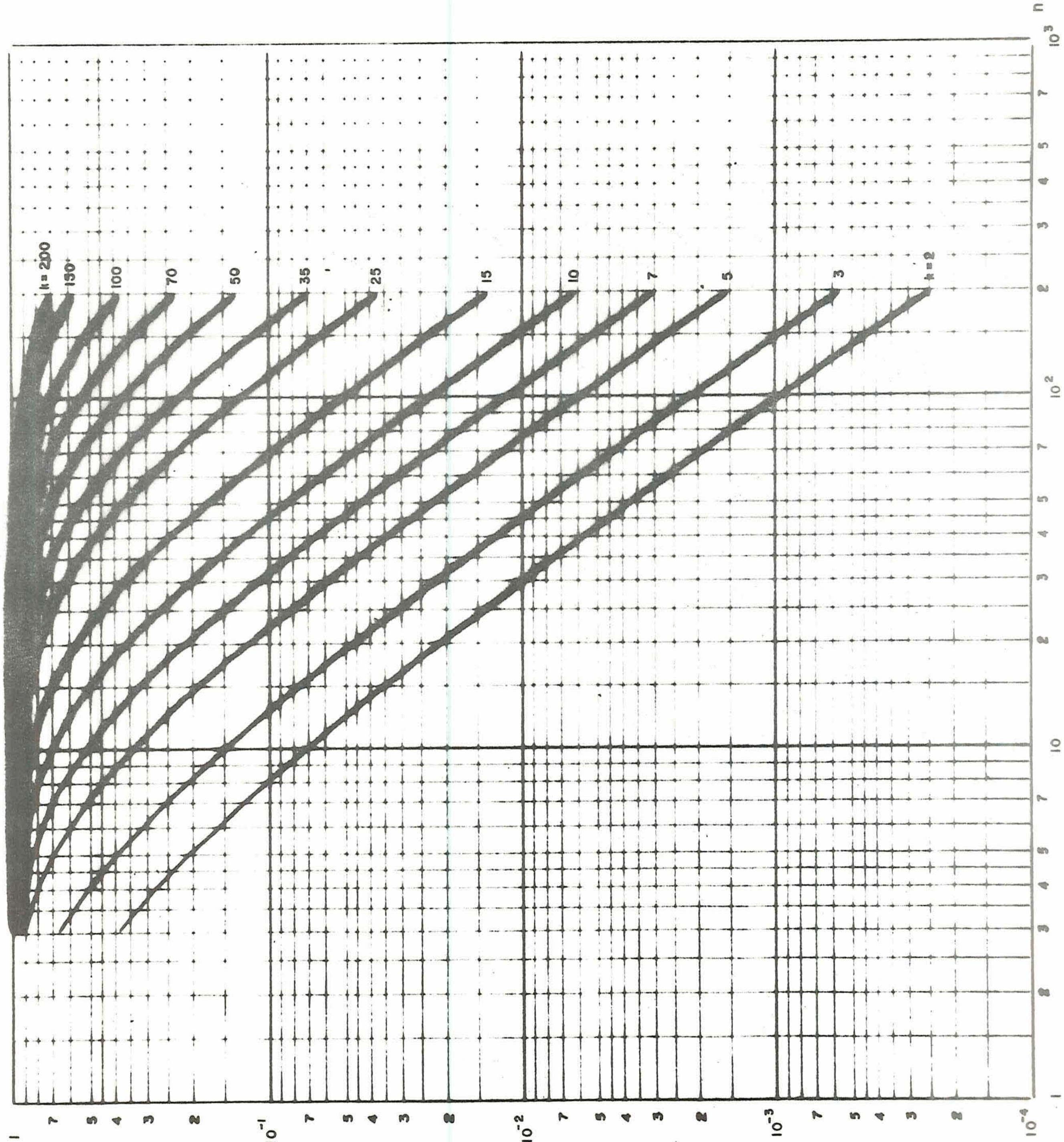

Fig. 4. η_{nk}

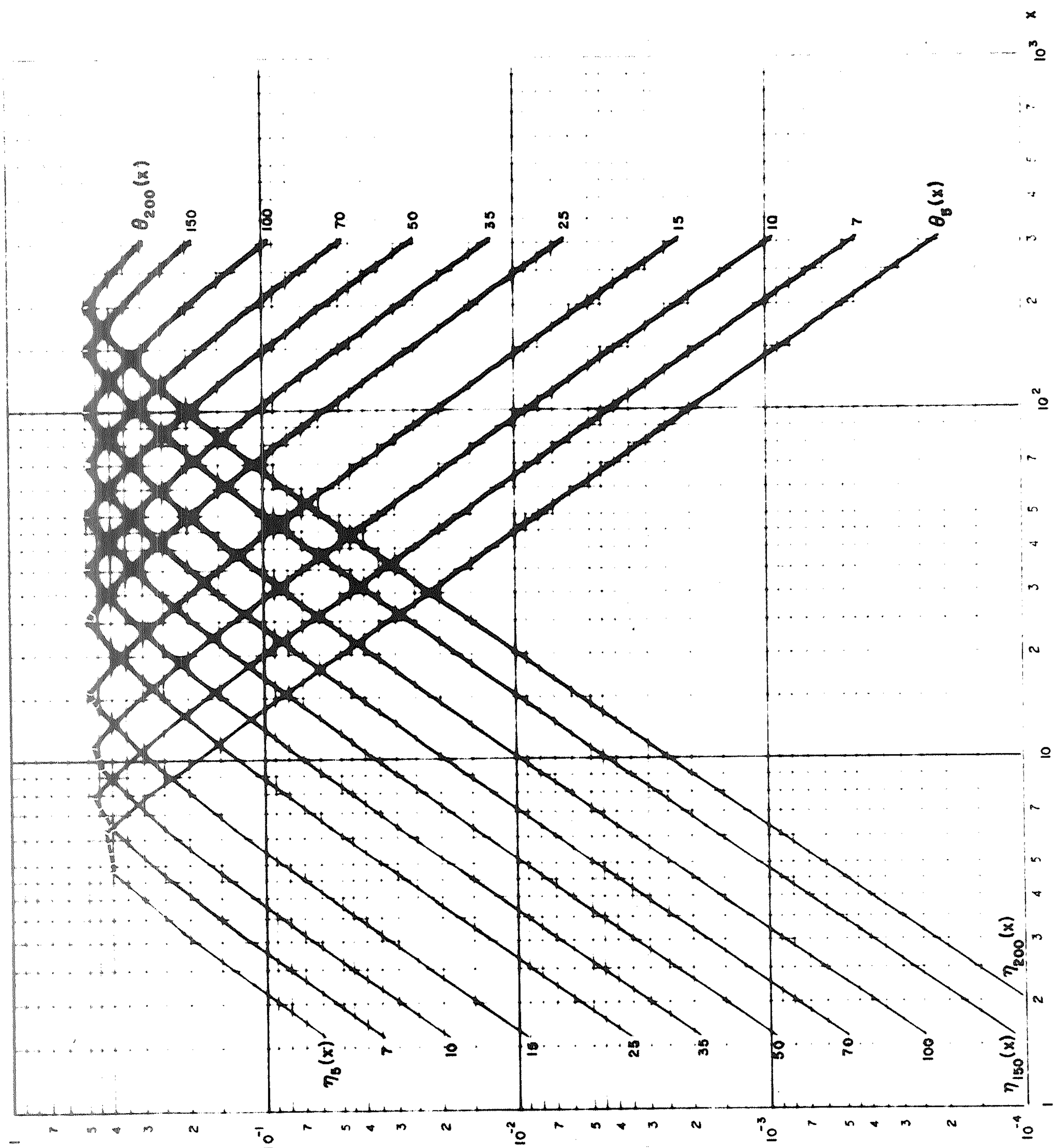

Fig. 5. $\eta_n(x)$, $\theta_n(x)$

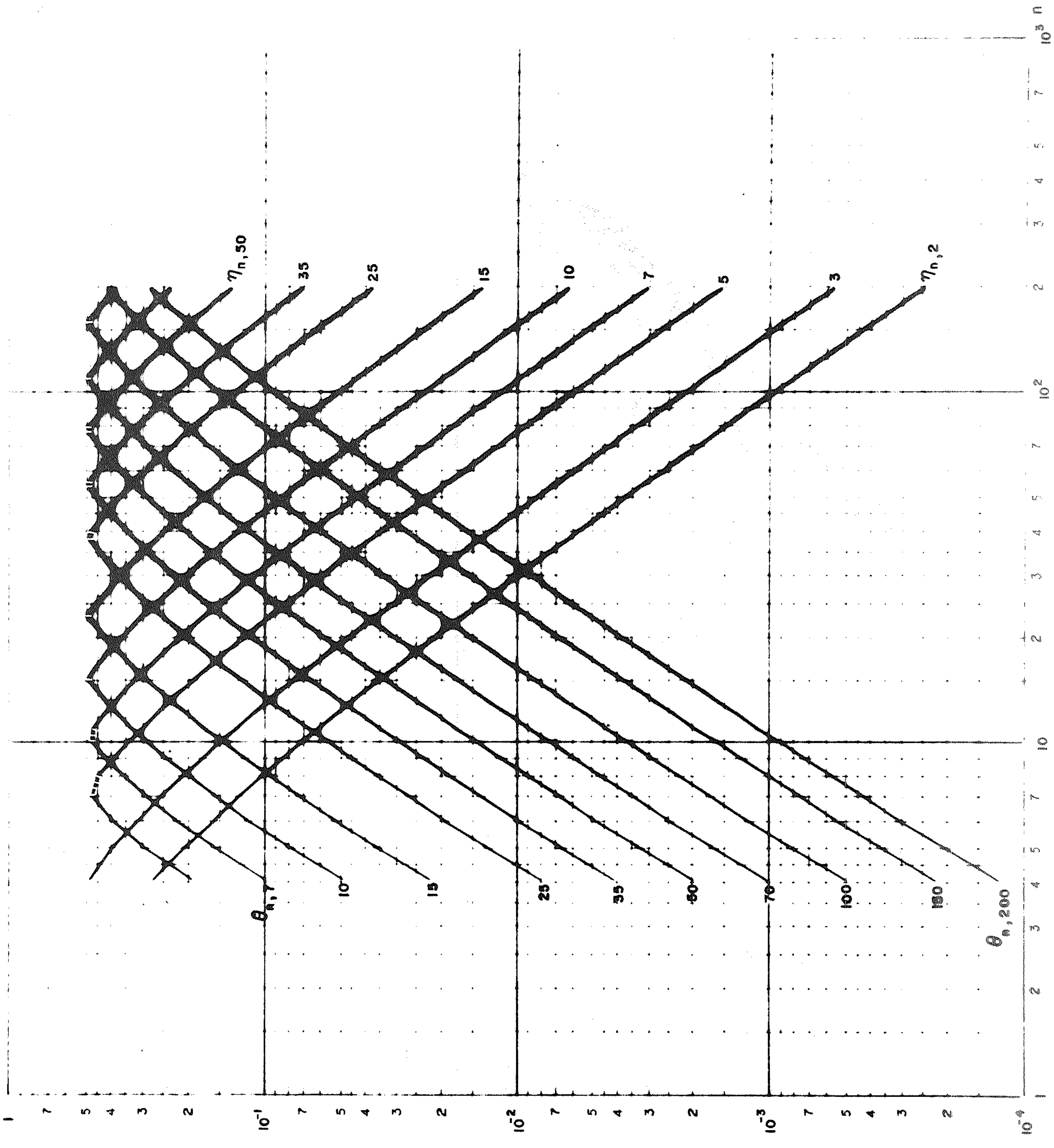

Fig. 6. θ_{nk}, η_{nk}

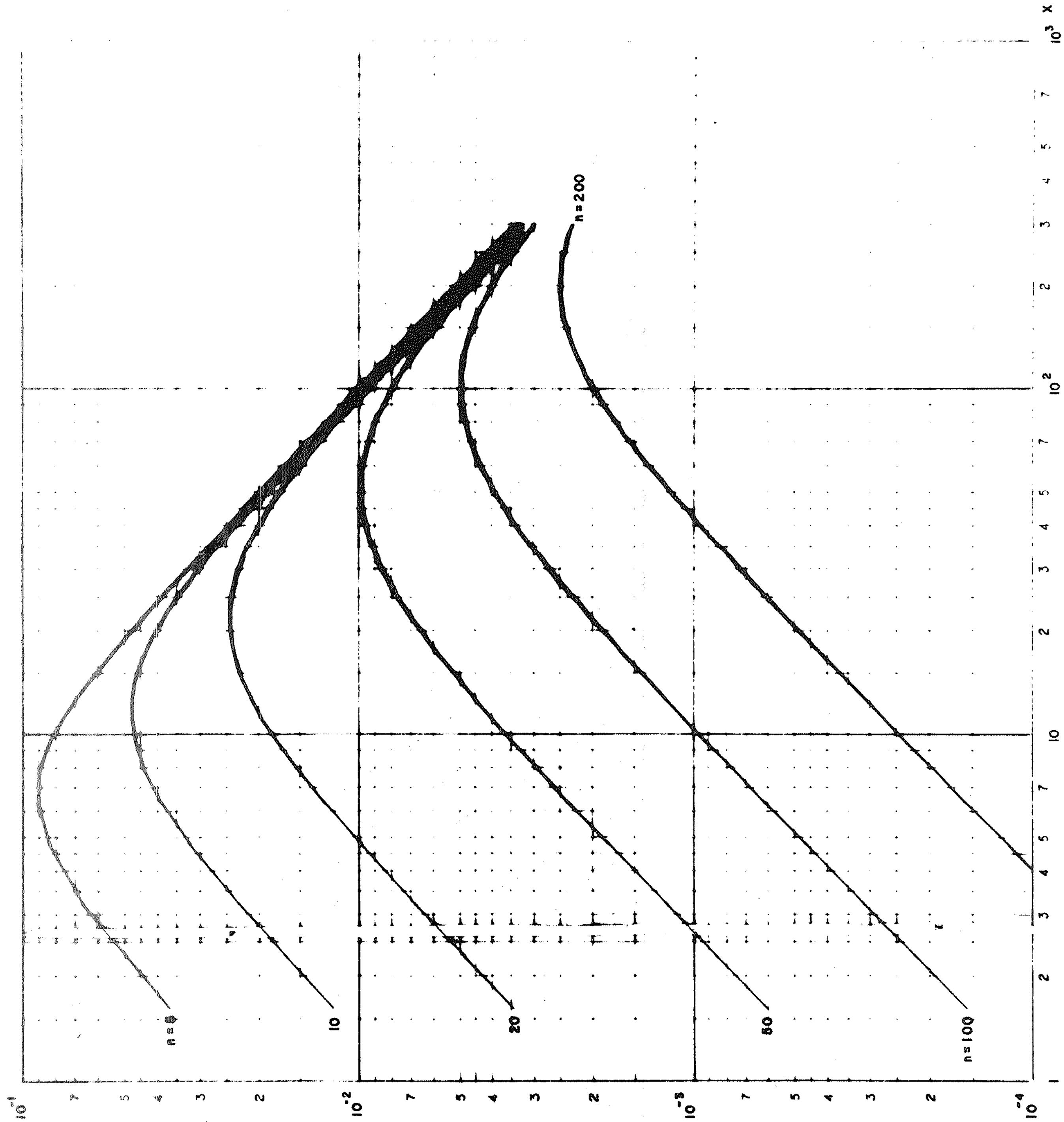

Fig. 7. $\frac{\eta_n(x)}{x}$

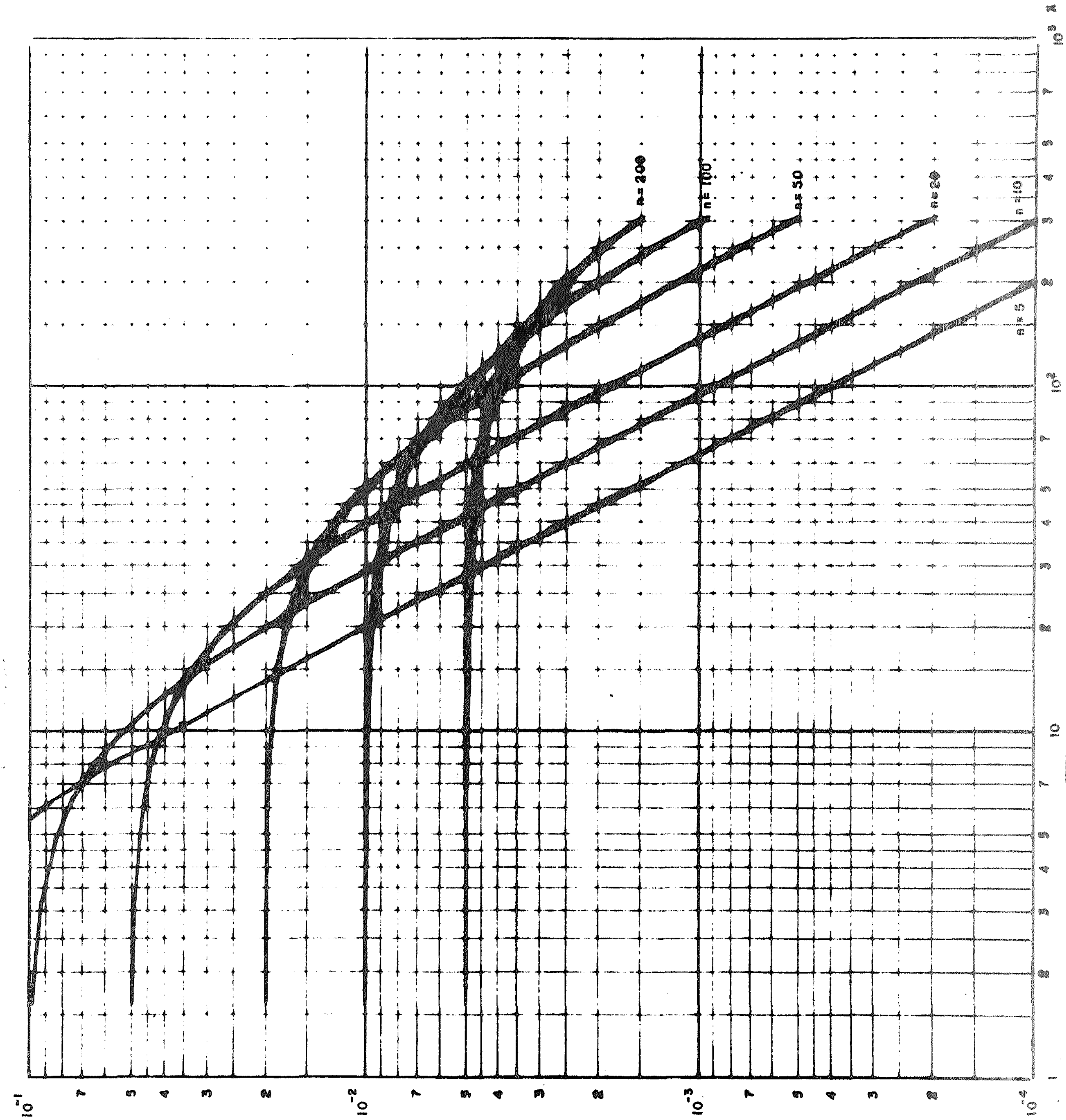

Fig. 8. $\frac{\theta_n(x)}{n}$

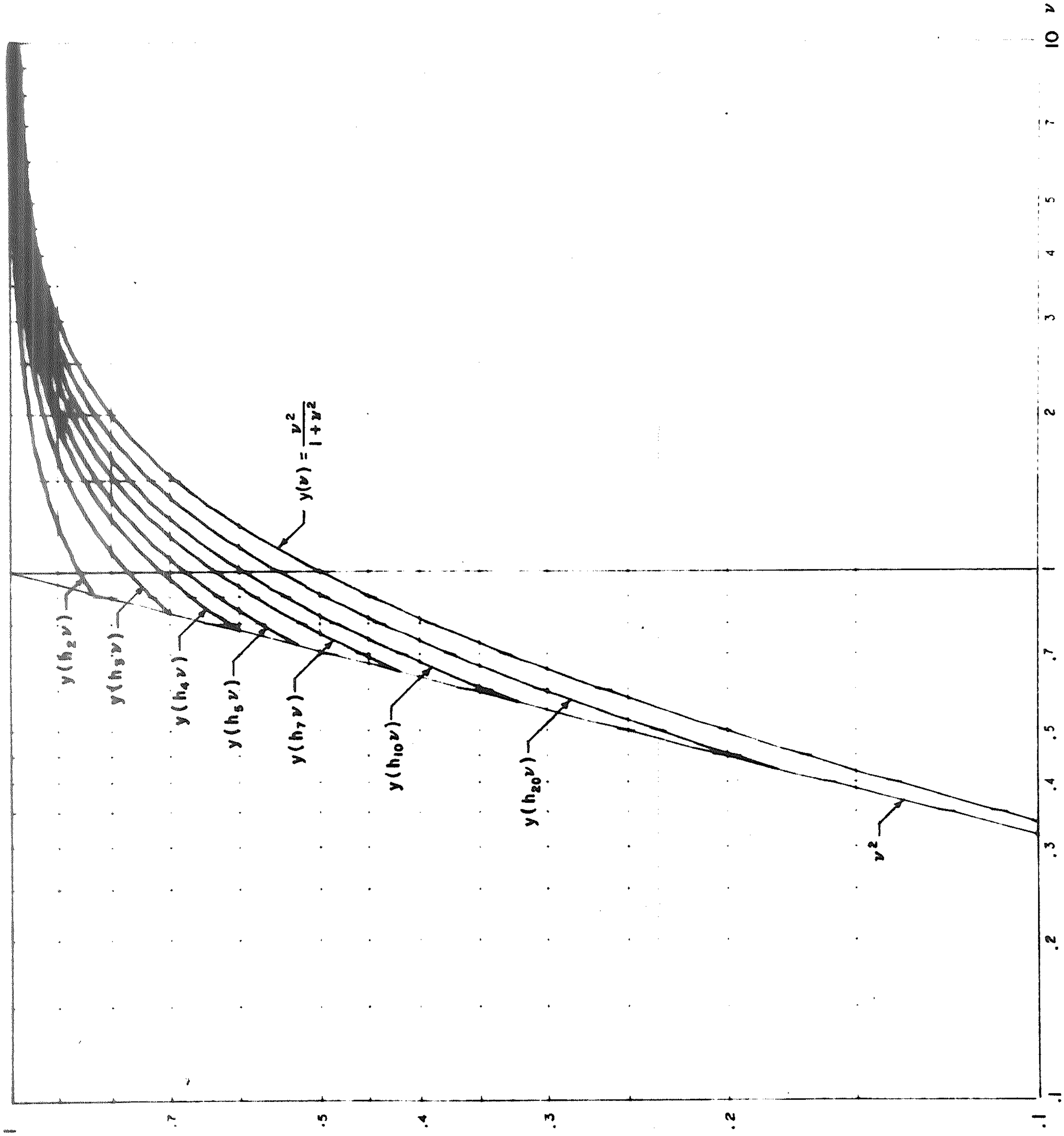

Fig. 9. Universal bounds for $\eta_n(x)$, $\theta_n(x)$

NASA — JPL — Coml., L.A., Calif.