The Calculation of Fourier Coefficients by the Möbius Inversion of the Poisson Summation Formula Part I. Functions whose Early Derivatives are Continuous*

By J. N. Lyness

Abstract. The Möbius inversion technique is applied to the Poisson summation formula. This results in expressions for the remainder term in the Fourier coefficient asymptotic expansion as an infinite series. Each element of this series is a remainder term in the corresponding Euler-Maclaurin summation formula, and the series has specified convergence properties.

These expressions may be used as the basis for the numerical evaluation of sets of Fourier coefficients. The organization of such a calculation is described, and discussed in the context of a broad comparison between this approach and various other standard methods.

1. Introduction. The purpose of this paper and its sequel is to derive a class of formulas suitable for the numerical evaluation of a set of Fourier coefficients

$$C^{(m)}f = \int_0^1 f(x) \cos 2\pi mx dx\,, \qquad m = 1, 2, 3, \cdots$$

$$S^{(m)}f = \int_0^1 f(x) \sin 2\pi mx dx\,, \qquad m = 1, 2, 3, \cdots .$$

In this paper we restrict ourselves to functions $f(x)$ which (preferably together with their first few derivatives) are continuous in the interval [0, 1]. In the sequel we shall provide a generalization of these results to cover functions which have algebraic or logarithmic singularities of a specified nature in the interval, and provide modifications for functions which are analytic in the interval but which have inconvenient numerical properties due to nearby poles in the complex plane.

The approximations $\tilde{C}^{(m)}f$ and $\tilde{S}^{(m)}f$ derived here differ fundamentally from other standard formulas, though they have some points in common. They seem particularly suitable in a situation in which all the Fourier coefficients are required to a uniform accuracy ϵ, a subroutine for $f(x)$ is available and (so far as Part I is concerned) $f(x)$ together with its first few derivatives are known to be continuous in [0, 1]. The points of similarity include the property that the approximations $\tilde{C}^{(m)}f$ (or $\tilde{S}^{(m)}f$), $m = 1, 2, \cdots$, are based on the same set of function values or a subset of this set. The principal difference is that there is no restriction to a particular number of points for function evaluation per period. In fact, coefficients of the type

Received June 23, 1969, revised August 11, 1969.

AMS Subject Classifications. Primary 4210, 6555; Secondary 4208, 6525, 6580.

Key Words and Phrases. Fourier coefficients, Fourier series, Poisson summation formula, Euler-Maclaurin summation formula, Möbius inversion, Filon-Luke formulas, fast Fourier transform, trapezoidal quadrature rule, Fourier coefficient asymptotic expansion.

* Work performed under the auspices of the U. S. Atomic Energy Commission.

$\cos 2\pi j/m$ do not appear in these formulas. The implementation has a degree of flexibility. If the value of the integral $\int_0^1 f(x)dx$ is known, or *approximate* values of derivatives $f^{(q)}(0)$ and $f^{(q)}(1)$ are known, this information may be incorporated in a simple manner into the formulas with a consequent reduction in the number of function values required.

The first half of this paper contains no approximation theory. In Section 2, the Poisson summation formula is introduced. In Section 3 the asymptotic expansion for the Fourier coefficient and the Euler-Maclaurin summation formula are derived. In Section 5 the Möbius inversion technique is discussed. All these results are classical, and are included here briefly to provide a proper background, and to establish an appropriate notation. A brief discussion which illustrates the danger of using the asymptotic expansion (without the remainder term) for numerical calculation is included in Section 4. This provides a proper motivation for the evidently new formulas derived in Section 6 by making use of the classical results of Sections 2, 3, and 5. These resemble the asymptotic expansion, but provide the remainder term in a completely different form. This involves an infinite series; the terms of this series may be readily calculated and the ultimate rate of convergence of the series is known.

In the second half of this paper, methods of applying this formula in actual calculations are described. This involves the appropriate assignment of various parameters occurring in the exact formula, together with the practical determination of the point at which to truncate the infinite series. In Sections 8 and 9 an implementation of an essentially practical nature is described. In Section 10 some theoretical properties of the approximation are described and a standard approximation error bound is derived. In Section 11 a discussion of what the author considers to be the essential features of the method is presented, in the form of a comparison with a finite version of the Fast Fourier Transform and with the Filon-Luke Formulas.

A suitable starting point for all the theory required in both this paper and its sequel is to assume the well-known relations between $f(x)$ and its Fourier series $\bar{f}(x)$. In order to present this theory in a relatively straightforward manner we restrict the functions $f(x)$ being considered to those to which the standard theorems of finite Fourier analysis may be applied without having to state detailed restrictions at every stage. Consequently we introduce the following overall restrictions:

(1.1) (i) $f(x)$ is absolutely integrable over the closed interval $[0, 1]$,

(1.2) (ii) $f(x)$ has at most a finite number of singularities in the interval $[0, 1]$.

The theory presented in Section 2 requires no further restriction. However, at the present time, this theory has been developed to the stage of providing a viable method for the calculation of Fourier coefficients only in the case in which these singularities are algebraic or logarithmic.

In the rest of this paper (Part I), we deal with a much smaller class of function. Here $f(x)$ has to be continuous in the closed interval $[0, 1]$ and, for the results to be more than trivial identities, some of the derivatives of $f(x)$ have to be continuous as well:

(1.3) (iii) $$f(x) \in C^p[0, 1], \qquad p \geqq 0 .$$

2. Finite Forms of the Poison Summation Formula. In this section we define the trapezoidal rule and use the fundamental theorem about Fourier series to derive a finite form of Poisson's summation formula. These results apply to functions which satisfy the first two restrictions (1.1) and (1.2) mentioned in Section 1. In Section 3 we confine our attention to continuous functions which satisfy restriction (1.3) and derive the standard Fourier coefficient asymptotic expansion and the Euler-Maclaurin summation formula.

The reader who is familiar with these formulas need only refer to these sections in order to acquaint himself with the notation.

It is convenient to emphasize the linear nature of many of the quantities occurring in this paper. This is done by using the terminology of linear operators wherever possible, though most of the expressions required are classical and more familiar in an expanded form. Consequently we denote the integral of $f(x)$ by

$$If = \int_0^1 f(x)dx \tag{2.0}$$

and we denote the Fourier coefficients of $f(x)$ by

$$C^{(r)}f = \int_0^1 f(x) \cos 2\pi rx dx , \qquad r = 1, 2, 3, \cdots \tag{2.1}$$

$$S^{(r)}f = \int_0^1 f(x) \sin 2\pi rx dx , \qquad r = 1, 2, 3, \cdots . \tag{2.2}$$

We invoke the classical theorems from the theory of Fourier analysis to define the Fourier series $\bar{f}(x)$ of $f(x)$. This is given formally by

$$\bar{f}(x) = If + 2 \sum_{r=1}^{\infty} C^{(r)}f \cos 2\pi rx + 2 \sum_{r=1}^{\infty} S^{(r)}f \sin 2\pi rx . \tag{2.3}$$

As is well known the function $\bar{f}(x)$ defined as the sum of the series in (2.3) coincides in general with the function $f(x)$. So long as $f(x)$ satisfies restrictions (1.1) and (1.2), $\bar{f}(x)$ exists at all points other than possibly those at which $f(x)$ itself is undefined. It is very well known that, if the limits in the following equations exist, then

$$\bar{f}(x) = \tfrac{1}{2} \lim_{\epsilon \to 0} (f(x + \epsilon) + f(x - \epsilon)) , \qquad 0 < x < 1 \tag{2.4}$$

and

$$\bar{f}(0) = \bar{f}(1) = \tfrac{1}{2} \lim_{\epsilon \to 0+} (f(\epsilon) + f(1 - \epsilon)) . \tag{2.5}$$

We now introduce a condensed notation for the trapezoidal quadrature rule approximations to the integral

$$If = I\bar{f} = \int_0^1 f(x)dx . \tag{2.6}$$

The conventional (end point) trapezoidal rule approximation is defined by

$$R^{[m,1]}f = \frac{1}{m}\left\{\frac{1}{2} f(0) + \sum_{j=1}^{m-1} f\left(\frac{j}{m}\right) + \frac{1}{2} f(1)\right\} . \tag{2.7}$$

We also require a general 'offset' trapezoidal rule. This is one which uses m equally spaced function values, the spacing being $1/m$ with the first abscissa at the point

$$t_\alpha/m = (1+\alpha)/2m\,, \qquad |\alpha| < 1\,. \tag{2.8}$$

General Offset Trapezoidal Rule $|\alpha| \neq 1$.

$$R^{[m,\alpha]}f = \frac{1}{m}\sum_{j=1}^{m} f\left(\frac{j+t_\alpha-1}{m}\right), \qquad t_\alpha = (1+\alpha)/2\,, \quad |\alpha| < 1\,. \tag{2.9}$$

(The special case $\alpha = 1$ is given by (2.7) above.) These rule sums exist only if $f(x)$ is defined at each of the points required for function evaluation.

The classical Poisson summation formula relates an infinite series of function values to an infinite series of Fourier transforms. It may be written in the form

$$h\sum_{j=-\infty}^{\infty} f(jh) = \sum_{r=-\infty}^{\infty}\int_{-\infty}^{\infty} f(t)\cos\,[2\pi rt/h]dt\,. \tag{2.10}$$

Clearly the function $f(x)$ has to satisfy certain properties which ensure that the various limiting processes required in this formula exist. In this paper we are concerned with various finite forms of this formula, namely (2.13) to (2.16) below. While these may be obtained directly from the classical Poisson summation formula by inserting specially chosen functions $f(x)$, it is more in keeping with our underlying approach to proceed directly from the trapezoidal rule sum (2.9) and the Fourier series (2.3). Specifically we may substitute for the quantity $\bar{f}((j+t_\alpha-1)/m)$ which occurs in (2.9), the Fourier series given by (2.3) and change the order of the summation operators. This change is permissible since one of the sums is finite. The summation over index j may be carried out analytically, making use of identities such as

$$\begin{aligned}\sum_{j=1}^{m}\cos\,[2\pi r(j+t_\alpha-1)/m] &= m\cos 2\pi rt_\alpha\,, \qquad r/m = \text{integer}\\ &= \quad 0 \quad , \qquad r/m \neq \text{integer}\,.\end{aligned} \tag{2.11}$$

The result is as follows:

General Finite Form of Poisson Summation Formula.

$$R^{[m,\alpha]}\bar{f} - I\bar{f} = 2\sum_{r=1}^{\infty}\cos 2\pi rt_\alpha C^{(rm)}f + 2\sum_{r=1}^{\infty}\sin 2\pi rt_\alpha S^{(rm)}f\,. \tag{2.12}$$

Subsequently we make use of only four simple special cases of this formula; two are obtained from (2.12) by setting $\alpha = 1$ and $\alpha = 0$; the third is a linear combination of these; the fourth is a linear combination obtained from (2.12) with $\alpha = -\frac{1}{2}$ and $\alpha = \frac{1}{2}$. These are respectively:

$$R^{[m,1]}\bar{f} - I\bar{f} = 2\sum_{r=1}^{\infty} C^{(rm)}f\,, \tag{2.13}$$

$$R^{[m,0]}\bar{f} - I\bar{f} = 2\sum_{r=1}^{\infty} (-1)^r C^{(rm)}f\,, \tag{2.14}$$

$$R^{[m,1]}\bar{f} - R^{[2m,1]}\bar{f} = 2\sum_{r=1}^{\infty} C^{((2r-1)m)}f\,, \tag{2.15}$$

$$\frac{1}{2}\left(R^{[m,-1/2]}\bar{f} - R^{[m,1/2]}\bar{f}\right) = 2\sum_{r=1}^{\infty}(-1)^{r-1}S^{((2r-1)m)}f\,. \tag{2.16}$$

These are all simple variants of a finite form of the Poisson summation formula.

The formulas of this section are valid if $f(x)$ satisfies restrictions (1.1) and (1.2) and if the rule sums are defined and so do not involve function evaluation at an abscissa for which $\bar{f}(x)$ is not defined. Thus (2.14) can be used with $f(x) = x^{-1/2}$.

3. The Fourier Coefficient and the Euler-Maclaurin Asymptotic Expansions. The results of the previous section are valid for a wide class of functions $f(x)$, which includes all those satisfying restrictions (1.1) and (1.2). We now specialize the theory to functions $f(x)$ which are continuous in the closed interval [0, 1]. The various formulas derived here require that $f(x)$ together with its first p derivatives should be continuous in the closed interval [0, 1]. We denote this condition by the statement

$$f(x) \in C^p[0,1]\,. \tag{3.1}$$

The first result we require is an asymptotic expansion for the Fourier coefficient. This may be obtained by integration by parts. Thus

$$\int_0^1 f(x)e^{2\pi imx}dx = \frac{f(1)-f(0)}{2\pi im} - \frac{1}{2\pi im}\int_0^1 f'(x)e^{2\pi imx}dx\,. \tag{3.2}$$

The integral on the right is of the same form as that on the left, but with $f'(x)$ replacing $f(x)$. Thus we may successively integrate by parts to form a finite series whose rth term includes a factor $(2\pi im)^{-r}$, together with a remainder term. Taking the real and imaginary parts we find the following formulas.

The Fourier Coefficient Asymptotic Expansion.

$$\begin{aligned} C^{(m)}f &= \int_0^1 f(x)\cos 2\pi mx dx \\ &= \sum_{q=1}^{[(p-1)/2]} \frac{(-1)^{q-1}(f^{(2q-1)}(1) - f^{(2q-1)}(0))}{(2\pi m)^{2q}} + C_p{}^{(m)}f\,, \end{aligned} \tag{3.3}$$

$$\begin{aligned} S^{(m)}f &= \int_0^1 f(x)\sin 2\pi mx dx \\ &= \sum_{q=0}^{[(p-2)/2]} \frac{(-1)^{q-1}(f^{(2q)}(1) - f^{(2q)}(0))}{(2\pi m)^{2q+1}} + S_p{}^{(m)}f\,. \end{aligned} \tag{3.4}$$

The most convenient forms of the remainder terms $C_p{}^{(m)}f$ and $S_p{}^{(m)}f$ differ according as p is even or odd. For example

$$C_{2p}^{(m)}f = \frac{(-1)^p}{(2\pi m)^{2p}}\int_0^1 f^{(2p)}(x)(\cos 2\pi mx - 1)dx\,. \tag{3.5}$$

These expansions are valid only if the process of integration by parts is also valid. A sufficient condition is that $f(x) \in C^p[0, 1]$. In this case the remainder terms satisfy

$$C_p{}^{(m)}f \sim O(m^{-p})\;; \qquad S_p{}^{(m)} \sim O(m^{-p})\,, \qquad m \to \infty\,. \tag{3.6}$$

In Section 4 we discuss several examples of the Fourier coefficient asymptotic expansion.

These asymptotic expansions may be used to derive another set of asymptotic expansions. These are variants of the classical Euler-Maclaurin summation formula. The Poisson summation formula (2.12) expresses the error functional $E^{[m,\alpha]}\bar{f} = R^{[m,\alpha]}\bar{f} - I\bar{f}$ in terms of an infinite series involving the Fourier coefficients $C^{(rm)}f$ and $S^{(rm)}f$. We may substitute for these their expressions given by the finite sums on the right-hand sides of (3.3) and (3.4). The resulting formula may be simplified by introducing the Bernoulli functions

$$(3.7)\qquad \begin{aligned} \bar{B}_{2q}(x) &= 2(-1)^{q+1}(2q)!\sum_{r=1}^{\infty}\frac{\cos 2\pi rx}{(2\pi r)^{2q}}\,, \\ \bar{B}_{2q+1}(x) &= 2(-1)^{q+1}(2q+1)!\sum_{r=1}^{\infty}\frac{\sin 2\pi rx}{(2\pi r)^{2q+1}}\,. \end{aligned}$$

This leads to the following formula.

General Euler-Maclaurin Asymptotic Expansion.

$$(3.8)\qquad \begin{aligned} E^{[m\ \alpha]}f &= R^{[m,\alpha]}f - If \\ &= \sum_{q=1}^{p-1}\frac{\bar{B}_q(t_\alpha)}{q!}\,\frac{f^{(q-1)}(1) - f^{(q-1)}(0)}{m^q} + E_p^{[m,\alpha]}f\,, \end{aligned}$$

where the remainder term is

$$(3.9)\qquad \begin{aligned} E_p^{[m,\alpha]}f &= 2\sum_{r=1}^{\infty}\cos 2\pi rt_\alpha C_p^{(mr)}f + 2\sum_{r=1}^{\infty}\sin 2\pi rt_\alpha S_p^{(mr)}f \\ &= \frac{1}{m^p}\int_0^1 f^{(p)}(x)\,\frac{\bar{B}_p(t_\alpha) - \bar{B}_p(t_\alpha - mx)}{p!}\,dx\,. \end{aligned}$$

Since the Bernoulli functions are bounded in the interval [0, 1] it follows that

$$E_p^{[m,\alpha]}f \sim O(m^{-p})\,, \qquad m \to \infty\,.$$

In the subsequent theory, we shall be interested in four special cases of this formula. These four cases are Eqs. (3.15) to (3.18) below. It is convenient to express the Bernoulli functions in terms of the Riemann zeta function and some of its variants. Following Abramowitz and Stegun [1], we define

$$(3.10)\qquad \begin{aligned} \zeta(q) &= 1 + \frac{1}{2^q} + \frac{1}{3^q} + \frac{1}{4^q} + \cdots\,, & q &> 1\,, \\ \eta(q) &= 1 - \frac{1}{2^q} + \frac{1}{3^q} - \frac{1}{4^q} + \cdots\,, & q &\geqq 1\,, \\ \lambda(q) &= 1 + \frac{1}{3^q} + \frac{1}{5^q} + \frac{1}{7^q} + \cdots\,, & q &> 1\,, \\ \beta(q) &= 1 - \frac{1}{3^q} + \frac{1}{5^q} - \frac{1}{7^q} + \cdots\,, & q &\geqq 1\,. \end{aligned}$$

Clearly

$$(3.11)\qquad \eta(q) = (1 - 2^{1-q})\zeta(q)\,; \qquad \lambda(q) = (1 - 2^{-q})\zeta(q)\,.$$

These functions are related to the Bernoulli functions in the following way:

$$\frac{B_{2q}(1)}{2q!} = \frac{2(-1)^{q-1}\zeta(2q)}{(2\pi)^{2q}}; \qquad \frac{B_{2q}(\frac{1}{2})}{2q!} = \frac{2(-1)^{q}\eta(2q)}{(2\pi)^{2q}}, \tag{3.12}$$

$$\frac{\frac{1}{2}\,(B_{2q}(1) + B_{2q}(\frac{1}{2}))}{2q!} = \frac{2(-1)^{q-1}\lambda(2q)}{(2\pi)^{2q}}, \tag{3.13}$$

$$\frac{B_{2q-1}(\frac{1}{4})}{(2q-1)!} = -\frac{B_{2q-1}(\frac{3}{4})}{(2q-1)!} = \frac{2(-1)^{q}\beta(2q-1)}{(2\pi)^{2q-1}}. \tag{3.14}$$

We now write down the particular cases of the Euler-Maclaurin asymptotic expansion which correspond to the operators introduced in Eqs. (2.13) to (2.16). In doing so we replace the Bernoulli functions in (3.8) by equivalent forms expressed in terms of the Riemann zeta function and its variants given above

$$\begin{aligned} R^{[m,1]}f - If = \sum_{q=1}^{[(p-1)/2]} & \frac{2(-1)^{q-1}\zeta(2q)}{(2\pi m)^{2q}} \\ & \cdot (f^{(2q-1)}(1) - f^{(2q-1)}(0)) + E_p^{[m,1]}f \end{aligned} \tag{3.15}$$

$$\begin{aligned} R^{[m,0]}f - If = \sum_{q=1}^{[(p-1)/2]} & \frac{2(-1)^{q}\eta(2q)}{(2\pi m)^{2q}} \\ & \cdot (f^{(2q-1)}(1) - f^{(2q-1)}(0)) + E_p^{[m,0]}f \end{aligned} \tag{3.16}$$

$$\begin{aligned} R^{[m,1]}f - R^{[2m,1]}f = \sum_{q=1}^{[(p-1)/2]} & \frac{2(-1)^{q-1}\lambda(2q)}{(2\pi m)^{2q}} \\ & \cdot (f^{(2q-1)}(1) - f^{(2q-1)}(0)) + [E_p^{[m,1]}f - E_p^{[2m,1]}f] \end{aligned} \tag{3.17}$$

$$\begin{aligned} \tfrac{1}{2}\,(R^{[m,-1/2]}f - R^{[m,1/2]}f) = \sum_{q=1}^{[p/2]} & \frac{2(-1)^{q}\beta(2q-1)}{(2\pi m)^{2q-1}} \\ & \cdot (f^{(2q-2)}(1) - f^{(2q-2)}(0)) + \tfrac{1}{2}\,[E_p^{[m,-1/2]}f - E_p^{[m,1/2]}f]. \end{aligned} \tag{3.18}$$

The first of these is the classical Euler-Maclaurin summation formula. It is interesting to note the close similarity between this formula (3.15) and the cosine Fourier coefficient asymptotic expansion (3.3). The difference is the factor $2\zeta(2q)$, which occurs in each term in (3.15) but is absent in (3.3). For large values of q, $\zeta(2q) \simeq 1$. The subsequent theory exploits this similarity and the corresponding similarity between the other expansions (3.16) to (3.18) and either (3.3) or (3.4).

4. Examples of the Fourier Coefficient Asymptotic Expansion. In Section 3 we derived two sets of asymptotic expansions. These expressed the Fourier coefficients and the error functional as a finite series, together with a remainder term. All of these expansions have a very similar structure. In this section we discuss in more detail one of these, the cosine Fourier coefficient expansion (3.3). However, this discussion applies with only minor modification to any of these expansions.

In a problem in which there is no difficulty associated with the calculation of derivatives, it would be very convenient if Eq. (3.3) could be used to evaluate the cosine Fourier coefficient. This would involve in practice truncating this series at a point at which the remainder term $C_p{}^{(m)}f$ is thought to be smaller in magnitude than ϵ, the required accuracy. The main problem in such a calculation would be that of estimating the magnitude of the remainder term $C_p{}^{(m)}f$.

If $f(x)$ happened to be a function for which $\lim_{p\to\infty} C_p{}^{(m)}f = 0$, then the series in (3.3) may converge. The numerical summation of this series could then be attempted with some sort of confidence. But it is well known that in general this is not the case. The infinite series obtained from (3.3) by allowing p to become infinite is an asymptotic expansion which is generally divergent.

The remaining sections in this paper are devoted principally to obtaining a representation for the remainder terms $C_p{}^{(m)}f$ and $S_p{}^{(m)}f$ which may be evaluated in a relatively straightforward manner. Thus it is appropriate to discuss at this stage briefly the general pattern of behavior of this expansion and its remainder term in certain simple cases. This discussion will indicate the importance of the remainder term and show how dangerous it may be to make any assumption about its size which is not rigorously justified.

As a preliminary we consider the information already available. This is that $C_p{}^{(m)}f$ is of order $O(m^{-p})$. While this is of considerable use in further analytic investigations, it is of very doubtful value in direct numerical application. Essentially we may assume the following. If we retain the first $p/2$ terms, and require some accuracy ϵ, there is some value of m, say m_0, for which the remainder term $|C_p{}^{(m)}f| < \epsilon$ for all $m > m_0$. Unfortunately, the value of m_0 as a function of ϵ is not known a priori. To determine m_0 in any particular instance requires an analytical investigation based on the particular properties of $f(x)$.

The simplest example is the polynomial. If $f(x)$ is a polynomial in x of degree d, the expansion terminates, leaving an expression for the cosine (sine) Fourier coefficient as an even (odd) polynomial in $1/m$ of degree d or less.

Another simple class of functions consists of entire functions of order 1. Thus if $f(x) = e^{\alpha x}$ it is simple to show that if $2\pi m > |\alpha|$, the series converges geometrically; on the other hand if $2\pi m < |\alpha|$, the series diverges geometrically. This behavior is typical of all entire functions of order 1.

However, the series may converge to an incorrect result. If $f(x) \in C^{\infty}[0, 1]$ and is periodic with period 1, we find that

$$f^{(q)}(1) - f^{(q)}(0) = 0 \quad \text{all } q\,. \tag{4.1}$$

Each term in the series is zero. This can happen even if the Fourier coefficient is not zero. Thus if

$$f(x) = e^{\cos 2\pi x} \tag{4.2}$$

the information given by the finite series is that

$$\begin{aligned} C^{(m)}f &= \int_0^1 f(x)\cos 2\pi m x dx = C_{2p}^{(m)}f \\ &= \frac{(-1)^p}{(2\pi m)^{2p}}\int_0^1 f^{(2p)}(x)(\cos 2\pi m x - 1)dx\,. \end{aligned} \tag{4.3}$$

Here the cosine Fourier coefficient is equal in value to the remainder term. Neither is zero. While perhaps the user might notice and suspect a series all of whose terms are zero, a dangerous situation arises in the case of a function such as

$$f(x) = e^{\cos 2\pi x} + e^{\alpha x}, \qquad |\alpha| < 2\pi m . \tag{4.4}$$

The use of the series here would produce a series which converged to the Fourier coefficient of $e^{\alpha x}$. A different function, which shares this property of periodic functions, is the 'smudge' function

$$\begin{aligned} f(x) &= e^{-1/x}e^{-1/(1-x)}g(x) , \qquad 0 < x < 1 , \\ f(0) &= f(1) = 0 , \end{aligned} \tag{4.5}$$

where $g(x) \in C^\infty[0, 1]$.

The examples mentioned above are mainly examples in which the series converges (to a correct or incorrect result) or in which the divergent nature of the expansion is at once apparent.

If $f(x)$ is an analytic function having a singularity in the complex plane at a finite distance from the origin, or is an entire function of order greater than 1, the series is almost invariably divergent for any value of m. (The exceptions to this statement arise if $f(x)$ is periodic with period 1 and $C^\infty[0, 1]$, or if some symmetry property has the consequence that the significant part of $f^{(q)}(0)$ and of $f^{(q)}(1)$, although very large in magnitude, eliminate each other when taken in the combination $f^{(q)}(1) - f^{(q)}(0)$ for all q odd (or even).) Thus with

$$f(x) = \frac{1}{x+1} \tag{4.6}$$

the nonzero terms in the series are

$$T_{2q} = \frac{K_{2q}}{m^{2q}} = \frac{(-1)^{q-1}(2q-1)!}{(2\pi m)^{2q}} \left\{1 - \frac{1}{2^{2q}}\right\} . \tag{4.7}$$

For large m, the series consists first of terms successively decreasing in magnitude. However, when terms $T_{2q'}$ where $2q' > 2\pi m$ are reached, the terms in the series successively increase in magnitude. The series diverges for all m. The function $f(x)$ given by (4.6) can be shown to have an nth derivative of constant sign in the interval [0, 1]. This information can be used to show that the series is semiconvergent, i.e., the value of $C_{2p}{}^{(m)}f$ is smaller in magnitude than the final included term T_{2p} and of the same sign as T_{2p}.

But in general one cannot expect $f(x)$ to have the property that its high-order derivatives have constant sign in the interval [0, 1]. It may be very dangerous indeed to assume that such series have 'approximate' properties of this nature. An example (which is not pathological) is given by

$$f(x) = 1/(x^2 - x + 0.26) . \tag{4.8}$$

This function has simple poles at $z = \frac{1}{2} \pm \frac{1}{10}i$. The individual terms in the expansion have a straightforward analytic expression. In Table 1, the values of T_{2q} and of the partial sums

$$\sum_{2q} = T_2 + T_4 + \cdots + T_{2q} \tag{4.9}$$

are listed for $m = 6$ and $q = 1(1)20$. Inspection of the table shows that the terms become very small (about 10^{-8}) and then increase. A plausible conclusion from this table is that

$$\int_0^1 f(x) \cos 12\pi x dx \simeq -0.02 . \tag{4.10}$$

The true value of this integral is $+0.701$.

In conclusion then, the use of this expansion for numerical computation can be very unreliable and deceptive unless some bound on the remainder term is available. Even then, the true value of the remainder term may be so large that the numerical result is not meaningful. In Section 6 an expression for the remainder term is derived, in the form of a convergent series. Thus a meaningful calculation based on the series described in this section may be carried out, if the additional work involved in calculating the value of the remainder term is included in the calculation.

TABLE 1

q	qth term: $T_{2q} = K_{2q}/6^{2q}$	qth partial sum: $\sum_{2q}$
1	−2.081713996537 − 002	−2.081713996537 − 002
2	6.240287755500 − 004	−2.019311118929 − 002
3	−4.405942671467 − 005	−2.023717061616 − 002
4	5.406335388194 − 006	−2.023176428105 − 002
5	−9.689934463007 − 007	−2.023273327446 − 002
6	2.189459081506 − 007	−2.023251432867 − 002
7	−4.886759442277 − 008	−2.023256319633 − 002
8	−1.259169531957 − 009	−2.023256445536 − 002
9	2.215382693859 − 008	−2.023254230153 − 002
10	−3.699575480714 − 008	−2.023257929715 − 002
11	5.432548918994 − 008	−2.023252497194 − 002
12	−7.975547956536 − 008	−2.023260472750 − 002
13	1.182895543519 − 007	−2.023248643789 − 002
14	−1.688494712958 − 007	−2.023265528725 − 002
15	1.911763558659 − 007	−2.023246411118 − 002
16	4.555025903741 − 008	−2.023241856077 − 002
17	−1.615341149154 − 006	−2 023403390252 − 002
18	9.184600203531 − 006	−2.022484930232 − 002
19	−4.237003493705 − 005	−2.026721933740 − 002
20	1.809352162352 − 004	−2.008628412092 − 002

The elements and partial sums in the asymptotic expansion of the sixth cosine Fourier coefficient of $f(x) = 1/(x^2 - x + 0.26)$.

5. The M¨bius Inversion Technique. One of the standard topics in the theory of numbers is the theory of Möbius inversion. This is concerned with the inversion of an infinite set of equations. We suppose that the set of numbers $G(m)$, $m = 1, 2, 3, \cdots$, is related to the set of numbers $F(m)$, $m = 1, 2, 3, \cdots$ by the set of equations

$$G(m) = a_1F(m) + a_2F(2m) + a_3F(3m) + \cdots , \qquad m = 1, 2, 3, \cdots , \tag{5.1}$$

where $a_1 \neq 0$ and the coefficients a_i are independent of m. Under certain conditions, the series (5.1) may be inverted and one may derive a set of equations

$$F(m) = b_1G(m) + b_2G(2m) + b_3G(3m) + \cdots , \qquad m = 1, 2, 3, \cdots , \tag{5.2}$$

where the set of coefficients b_i depend only on the set a_i. A theorem which gives sufficient conditions for such an inversion to be justified is the following:

Möbius Inversion Theorem. Given a set of numbers $a_1, a_2, \cdots$ $(a_1 \neq 0)$ a second set $b_1, b_2, \cdots$ may be determined recursively using

$$(5.3) \qquad a_1 b_1 = 1; \qquad \sum_{r \mid d} a_r b_{d/r} = 0\,, \qquad d = 2, 3, 4, \cdots .$$

If the set of Eqs. (5.1) is valid, the set (5.2) is also valid under the sufficient condition

$$(5.4) \qquad \sum_{l=1}^{\infty} \sum_{k=1}^{\infty} |a_k b_l F(klm)| < \infty\,, \qquad m = 1, 2, 3, \cdots .$$

Several alternate sets of sufficient conditions are known. The necessary conditions do not appear to be known.

One of the first applications of the technique defined by Eq. (5.3) is to find the set b_i which corresponds to the set $a_1 = a_2 = a_3 = \cdots = 1$. This leads to the Möbius numbers μ_j (Möbius function $\mu(j)$), defined by

$$(5.5) \qquad \begin{array}{ll} \mu_1 = 1\,, & \\ \mu_j = 0 & \text{if } j \text{ has a square factor other than } 1\,, \\ \mu_j = (-1)^r & \text{if } j \text{ is the product of } r \text{ distinct prime numbers} \\ & \text{(not including 1)}\,. \end{array}$$

(The first ten Möbius numbers are $+1, -1, -1, 0, -1, +1, -1, 0, 0, +1$.) If we refer back to the special cases of the Poisson summation formula given in Eqs. (2.13) to (2.16) we see that each consists of a set of equations of precisely the form of (5.1). Each may be inverted using the Möbius inversion technique. A minor variant of this inversion is carried out in Section 6. For the moment we simply determine the appropriate coefficients. We list here the values of b_j corresponding to four different sets of a_j. These turn out to be Möbius numbers or simple functions of them.

THEOREM 5.6. *The solutions of Eq.* (5.3) *for the four following specified sets of a_j are the corresponding sets b_j defined as follows:*

$$(5.6) \quad (1) \quad a_j = 1\,; \qquad b_j = \mu_j\,,$$

$$(5.7) \quad (2) \quad \begin{array}{llll} a_j = 1 & (j \textit{ odd})\,; & b_j = \nu_j = \mu_j & (j \textit{ odd})\,, \\ = -1 & (j \textit{ even})\,; & b_j = \nu_j = 2^{n-1}\mu_k & (j = 2^n k; k \textit{ odd})\,, \end{array}$$

$$(5.8) \quad (3) \quad \begin{array}{llll} a_j = 1 & (j \textit{ odd})\,; & b_j = \mu_j & (j \textit{ odd})\,, \\ = 0 & (j \textit{ even})\,; & b_j = 0 & (j \textit{ even})\,, \end{array}$$

$$(5.9) \quad (4) \quad \begin{array}{llll} a_j = 1 & (j = 4k+1)\,; & b_j = \mu_j & (j = 4k+1)\,, \\ a_j = -1 & (j = 4k+3)\,; & b_j = -\mu_j & (j = 4k+3)\,, \\ a_j = 0 & (j \textit{ even})\,; & b_j = 0 & (j \textit{ even})\,. \end{array}$$

The proof of (1) above may be found in any standard textbook on Number Theory. (See for example Hardy and Wright [9].) The results (3), (4), and (2) are successively more complicated consequences of (1).

In Section 6 we shall use some results which require the inversion of the formula for the Riemann zeta function

$$\zeta(q) = \sum_{s=1}^{\infty} \frac{1}{s^q}, \qquad q > 1 \tag{5.10}$$

and the variants given in Section 3. These results constitute a standard application of the above theory and are derived in the following manner. Definition (5.10) may be written in the form

$$\frac{\zeta(q)}{m^q} = \sum_{s=1}^{\infty} \frac{1}{(ms)^q}, \qquad m = 1, 2, 3, \cdots . \tag{5.11}$$

This may be identified with (5.1) by setting

$$\begin{gathered} G(m) = \zeta(q)/m^q ; \qquad F(m) = 1/m^q , \\ a_1 = a_2 = a_3 = \cdots = 1 . \end{gathered} \tag{5.12}$$

Equation (5.2) then follows formally, the values of b_j being given by (5.6) above.

$$\frac{1}{m^q} = \sum_{s=1}^{\infty} \frac{\mu_s \zeta(q)}{(ms)^q}, \qquad m = 1, 2, \cdots . \tag{5.13}$$

This is in fact the form in which we require this identity. The standard form is

$$\frac{1}{\zeta(q)} = \sum_{s=1}^{\infty} \frac{\mu_s}{s^q} . \tag{5.14}$$

The validity of (5.13) and so of (5.14) is established by showing that condition (5.4) is satisfied. Here we have

$$\begin{aligned} \sum_{l=1}^{\infty} \sum_{k=1}^{\infty} |a_k b_l F(klm)| &= \sum_{l=1}^{\infty} \sum_{k=1}^{\infty} |\mu_l/(klm)^q| \\ &< \frac{1}{m^q} (\zeta(q))^2 < \infty , \qquad q > 1 , \end{aligned} \tag{5.15}$$

the first inequality being obtained by replacing $|\mu_l|$ by its upper bound 1.

We require in Section 6 the result corresponding to (5.13) for the variants of the Riemann zeta function. These are

$$\eta(q) = \sum_{s=1}^{\infty} \frac{(-1)^{s-1}}{s^q} ; \qquad \lambda(q) = \sum_{s=1}^{\infty} \frac{1}{(2s-1)^q} ; \qquad \beta(q) = \sum_{s=1}^{\infty} \frac{(-1)^{s-1}}{(2s-1)^q} \tag{5.16}$$

and the inverted formulas are

$$\frac{1}{m^q} = \sum_{s=1}^{\infty} \frac{\nu_s \eta(q)}{(ms)^q} = \sum_{s=1}^{\infty} \frac{\mu_{2s-1}\lambda(q)}{((2s-1)m)^q} = \sum_{s=1}^{\infty} \frac{(-1)^{s-1}\mu_{2s-1}\beta(q)}{((2s-1)m)^q}, \qquad q > 1 . \tag{5.17}$$

These may all be established formally following the procedure by which (5.13) was established. However, condition (5.4) is not satisfied for the first expression (involving the coefficients ν_s) for $q \leqq 2$ and a separate proof is required for $1 < q \leqq 2$.

6. Formulas for Fourier Coefficients. The variants of the Poisson summation formula (2.13) to (2.16) express the error function $R^{[m,1]}f - If$ or sums of function values such as $R^{[m,1]}f - R^{[2m,1]}f$ as series whose elements are Fourier coefficients. We are interested in obtaining formulas of the opposite nature, formulas which express Fourier coefficients in terms of sums of function values.

The fundamental idea on which the following theory is based is that the Poisson summation formula is a formula to which the Möbius inversion technique may be applied. This appears to have been previously unnoticed, except by Goldberg and Varga [7]. The inverted formula obtained in this way is (6.10) below which by itself is only useful if $f(x)$ happens to be periodic and $C^\infty[0, 1]$. But the principle, which uses the Möbius inversion formula to obtain formulas for Fourier coefficients in terms of function values, is very useful and credit for this idea belongs to Goldberg and Varga.

Instead of proceeding directly to invert the Poisson summation formula, it is more convenient to invert the corresponding formulas (3.9) which express the remainder term in the Euler-Maclaurin expansion in terms of remainder terms for the Fourier coefficient asymptotic expansion. The four special cases of (3.9) we use are:

$$E_p^{[m,1]}f = 2\sum_{r=1}^{\infty} C_p^{(rm)}f\,; \qquad E_p^{[m,0]}f = 2\sum_{r=1}^{\infty}(-1)^r C_p^{(rm)}f\,, \tag{6.1}$$

$$E_p^{[m,1]}f - E_p^{[2m,1]}f = 2\sum_{r=1}^{\infty} C_p^{((2r-1)m)}f\,, \tag{6.2}$$

$$\frac{1}{2}\,(E_p^{[m,-1/2]}f - E_p^{[m,1/2]}f) = 2\sum_{r=1}^{\infty}(-1)^{r-1}S_p^{((2r-1)m)}f\,. \tag{6.3}$$

Each of these is of precisely the form (5.1), namely

$$G(m) = a_1F(m) + a_2F(2m) + a_3F(3m) + \cdots\,, \tag{6.4}$$

and each may be inverted to give a formula of the form (5.2),

$$F(m) = b_1G(m) + b_2G(2m) + b_3G(3m) + \cdots\,. \tag{6.5}$$

The set of numbers a_i are different in each of the four cases and coincide with the four sets listed in relations (5.6) to (5.9) of Theorem 5.6 in Section 5. This theorem provides the appropriate values of the set of numbers b_i. Substituting into (6.5) we obtain four formulas, namely

$$\begin{aligned} 2C_p^{(m)}f &= \sum_{s=1}^{\infty}\mu_s E_p^{[ms,1]}f = -\sum_{s=1}^{\infty}\nu_s E_p^{[ms,0]}f \\ &= \sum_{s=1}^{\infty}\mu_{2s-1}[E_p^{[(2s-1)m,1]}f - E_p^{[2m(2s-1),1]}f]\,; \qquad p \geqq 2\,, \end{aligned} \tag{6.6}$$

$$2S_p^{(m)}f = \sum_{s=1}^{\infty}(-1)^{s-1}\mu_{2s-1}[\tfrac{1}{2}(E_p^{[(2s-1)m,-1/2]}f - E_p^{[(2s-1)m,1/2]}f)]\,; \qquad p \geqq 1\,. \tag{6.7}$$

The validity of these inversions follows for most of the stated values of p from the Möbius inversion theorem (5.3). Since

$$E_p^{[m,\alpha]} f \sim O(m^{-p}) \quad \text{and} \quad |\mu_s| \leqq 1 \,, \quad |\nu_s| < s$$

the sufficient condition (5.4) is satisfied when $p > 2$ since then

$$\sum_{l=1}^{\infty} \sum_{k=1}^{\infty} |a_k b_l F(klm)| < \frac{K}{m^p} \zeta(p)\zeta(p-1)$$

in all four cases. The first and third equation of (6.6) may be validated for $p > 1$ using this same condition (5.4). However, the second equation, that involving the coefficient ν_s, requires a separate proof for the case $1 < p \leqq 2$. This proof is straightforward, but tedious. Equation (6.7) is in fact valid for $p \geqq 1$, but for $p \leqq 2$, the proof is extremely sophisticated. This rather special case depends on Eq. (6.17) below; there is a brief discussion following that equation.

These equations, expressed in a different form, are suitable for calculating Fourier coefficients. Each remainder term occurring in these equations was originally defined (in (3.3), (3.4), and (3.15) to (3.18)) as the difference between some functional such as $C^{(m)}f$ and the first p terms of its asymptotic expansion. Thus the next step is to substitute for these remainder terms into (6.6) and (6.7). It is convenient to describe this in detail for only one of these equations. The corresponding results for the others are given towards the end of this section. The first equation in (6.6) is

$$(6.8) \qquad 2C_p^{(m)} f = \sum_{s=1}^{\infty} \mu_s E_p^{[ms,1]} f \,.$$

We may substitute for these remainder terms using (3.3) and (3.15). This gives

$$\begin{aligned} 2\int_0^1 f(x) \cos 2\pi m x dx &= 2 \sum_{q=1}^{n} \frac{(-1)^{q-1}(f^{(2q-1)}(1) - f^{(2q-1)}(0))}{(2\pi m)^{2q}} \\ (6.9) \qquad &+ \sum_{s=1}^{\infty} \mu_s \Bigg[R^{[ms,1]} f - If - \sum_{q=1}^{n} \\ &\times \frac{(-1)^{q-1} 2\zeta(2q)(f^{(2q-1)}(1) - f^{(2q-1)}(0))}{(2\pi m s)^{2q}} \Bigg] . \end{aligned}$$

This formula is of interest because it provides a remainder term, in the form of an infinite series, for the truncated asymptotic series for the Fourier coefficient. It appears that (6.9) may be generalized. The special case with $n = 0$ may be written

$$(6.10) \qquad 2\int_0^1 f(x) \cos 2\pi m x dx = \sum_{s=1}^{\infty} \mu_s [R^{[ms,1]} f - If] \,.$$

We may add to each side of this equation different multiples of (5.13) with different values of q, and in view of the absolute convergence of all the series involved, we may combine these to give

$$(6.11) \quad 2\int_0^1 f(x) \cos 2\pi m x dx = \sum_{q=1}^{n} \frac{\tilde{K}_{2q}}{m^{2q}} + \sum_{s=1}^{\infty} \mu_s \Bigg[R^{[ms,1]} f - If - \sum_{q=1}^{n} \frac{\tilde{K}_{2q}\zeta(2q)}{(ms)^{2q}} \Bigg] ,$$

where $\tilde{K}_2, \tilde{K}_4, \cdots, \tilde{K}_{2n}$ are arbitrary. Equation (6.9) is merely a special case of this more general relation, obtained by setting $\tilde{K}_{2q} = K_{2q}$, where

$$K_{2q} = 2(-1)^{q-1}(f^{(2q-1)}(1) - f^{(2q-1)}(0))/(2\pi)^{2q}. \tag{6.12}$$

In one sense this is an optimum choice as it makes the term in square brackets a term of order $O(s^{-2(n+1)})$, ensuring the maximum ultimate rate of convergence for the series.

This equation and the others like it given below ((6.13), (6.14), and (6.15)) have several outstanding features which make them eminently suitable as a basis for numerical computation of the Fourier coefficients. Their possible use in this manner is described in considerable detail in the remaining sections of this paper. However, it is pertinent at this point to draw the reader's attention to certain aspects of these formulas.

The most interesting feature is that Eq. (6.9) is quite independent of round-off errors in the calculation of the elements $f^{(2q-1)}(1) - f^{(2q-1)}(0)$. Since it is simply a special case of (6.11) in which the $\tilde{K}_{2q}$ are arbitrary, Eq. (6.9) is true quite independently of the values which are assigned to these elements. The penalty for using incorrect values may be that the series converges more slowly, but it converges in such a way as to give a correct result for the Fourier coefficient.

A second feature is that the same set of numerical quantities $E_{2p}^{[s,1]}f$, $s = 1, 2, 3, \cdots, \bar{s}$, is required for all the different Fourier coefficients. Here $\bar{s}$ is the value of s for which $E_{2p}^{[s,1]}f$ is so small and is evidently steadily decreasing in such a manner that the computer is prepared to disregard $E_{2p}^{[s,1]}f$ for $s > \bar{s}$. The calculation of the first Fourier coefficient ($m = 1$) requires all these (except those for which $\mu_s = 0$). The second coefficient requires alternate members of the set, and so on. The Fourier coefficients with $m > \bar{s}$ do not need any of these values. They are computed from the first p terms of the standard asymptotic expansion only, and with confidence in spite of the fact that the approximations to the derivatives being used need not be accurate.

The third feature of Eq. (6.9) or of (6.11) is that, although they are exact equations, they are of interest only if a numerical calculation is envisioned. The immediate reaction of a competent mathematician faced with Eq. (6.11) is to cancel out all the terms involving $\tilde{K}_{2q}$ and to reduce it, correctly, to the simpler form (6.10). If the sum over index s were to be evaluated analytically, this would be an obvious first stage. The importance of (6.9) is that, with these additional terms present on the right-hand side, the series converges at a rapid rate. And, aside from one or two special circumstances, the only reason for making a series more complicated in order to ensure more rapid convergence is that one intends to use it in a numerical calculation.

In the case that $f(x)$ happens to be a $C^\infty[0, 1]$ function and periodic with period 1, it follows that

$$f^{(q)}(1) = f^{(q)}(0)$$

and $K_{2q} = 0$. In this case (6.9) reduces to the much simpler form (6.10) and in this case the convergence of the right-hand side of (6.10) is relatively rapid, each term being $O(s^{-p})$ for any value of p. In the general case though, the convergence of the right-hand side of (6.10) is too slow for comfort, being $O(s^{-2})$, and in practice it is necessary to use a formula such as (6.9) or (6.11) to obtain a formula suitable for computation.

The above remarks apply equally to the variants of (6.9), (6.10), and (6.11) given below. We conclude this section by deriving these variants from the remaining equations in (6.6) and (6.7). Following a directly analogous procedure, we find

$$(6.13)\qquad 2\int_0^1 f(x)\cos 2\pi mx dx = \sum_{q=1}^{n} \frac{\tilde{K}_{2q}}{m^{2q}} - \sum_{s=1}^{\infty} \nu_s \left[R^{[ms,0]}f - If - \sum_{q=1}^{n} \frac{\eta(2q)\tilde{K}_{2q}}{(ms)^{2q}} \right], \qquad n \geqq 0$$

$$(6.14)\qquad 2\int_0^1 f(x)\cos 2\pi mx dx = \sum_{q=1}^{n} \frac{\tilde{K}_{2q}}{m^{2q}} + \sum_{s=1;\, s \text{ odd}}^{\infty} \mu_s \left[R^{[ms,1]}f - R^{[2ms,1]}f - \sum_{q=1}^{n} \frac{\lambda(2q)\tilde{K}_{2q}}{(ms)^{2q}} \right], \qquad n \geqq 0$$

and

$$(6.15)\qquad 2\int_0^1 f(x)\sin 2\pi mx dx = \sum_{q=1}^{n} \frac{\tilde{K}_{2q-1}}{m^{2q-1}} + \sum_{s=1;\, s \text{ odd}}^{\infty} (-1)^{(s-1)/2}\mu_s \times \left[\frac{1}{2}\left(R^{[ms,-1/2]}f - R^{[ms,1/2]}f\right) - \sum_{q=1}^{n} \frac{\beta(2q-1)\tilde{K}_{2q-1}}{(ms)^{2q-1}} \right], \qquad n \geqq 0,$$

where the numbers $\tilde{K}_q$ ($q = 1, 3, \cdots, 2n - 1$) are arbitrary. The choice of $\tilde{K}_{2q}$ which gives forms for the remainder term in the cosine Fourier coefficient is given by (6.12). The analogous choice for the sine Fourier coefficient is $\tilde{K}_{2q+1} = K_{2q+1}$ where

$$(6.16)\qquad K_{2q+1} = 2(-1)^{q-1}(f^{(2q)}(1) - f^{(2q)}(0))/(2\pi)^{2q+1}.$$

The proof of Eq. (6.15) in the case in which $\tilde{K}_1$ is arbitrary and not given by (6.16) requires the validity of the inversion

$$(6.17)\qquad \beta(q) = \sum_{s=1}^{\infty} \frac{(-1)^{s-1}}{(2s-1)^q}; \qquad \frac{1}{\beta(q)} = \sum_{s=1}^{\infty} \frac{\mu_{2s-1}(-1)^{s-1}}{(2s-1)^q}, \qquad q = 1.$$

This follows in an elementary manner for $q > 1$ from the Möbius inversion formula. To establish this for $q = 1$ the author has found it necessary to follow the method given in Landau [11, pp. 157–159] in which the corresponding result

$$(6.18)\qquad \sum_{s=1}^{\infty} \frac{\mu_s}{s} = \lim_{q\to 1} \frac{1}{\zeta(q)} = 0$$

is derived.

This case ($\tilde{K}_1$ arbitrary) is of little interest from a numerical point of view since in applications the value of $f(1) - f(0)$ would normally be available.

7. Implementation (General Remarks). In the preceding sections, no use of approximation theory has been made. The formulas derived in the previous section are all exact. Their immediate use is precluded since each includes an infinite sum over index s. They differ from the simpler asymptotic series of Section 3 in one respect only. The 'infinite tail' of the asymptotic expansion (which normally diverges) has been replaced by a convergent infinite sum, the sth term in the sum having order $O(s^{-2n-2})$.

In this section and in subsequent sections we discuss the implementation of a calculation based on one of these formulas. In many important respects the same treatment may be applied to each of these formulas. It is convenient to describe in detail only one, namely (6.11). Any significant differences between different members of the set of formulas will be mentioned in passing.

While there may be many different ways of implementing these formulas, we shall confine our attention to a specific type of problem. We shall assume that we have available, in the form of a subroutine or an analytic expression, the function $f(x)$. We wish to calculate approximations to a set of cosine Fourier coefficients

$$(7.1) \qquad C^{(m)}f = \int_0^1 f(x)\cos 2\pi mx dx\,, \qquad m = 1, 2, 3, \cdots, \overline{m}\,.$$

We require each approximation to have an error smaller in magnitude than a given tolerance ϵ. We wish to calculate all the cosine Fourier coefficients which are greater in magnitude than ϵ. Thus the value of $\overline{m}$ in (7.1) depends on ϵ and may be determined in the course of the calculation.

The exact formula (6.11), on which we shall base an approximate formula (7.4) below, may be written in the following form:

$$(7.2) \quad 2C^{(m)}f = 2\int_0^1 f(x)\cos 2\pi mx dx = \frac{\tilde{K}_2}{m^2} + \frac{\tilde{K}_4}{m^4} + \cdots + \frac{\tilde{K}_{2n}}{m^{2n}} + \sum_{s=1}^{\infty} \mu_s \tilde{E}_{2n+2}^{[ms,1]} f\,,$$

where

$$(7.3) \qquad \tilde{E}_{2n+2}^{[s,1]}f = R^{[s,1]}f - If - \frac{\zeta(2)\tilde{K}_2}{s^2} - \frac{\zeta(4)\tilde{K}_4}{s^4} - \cdots - \frac{\zeta(2n)\tilde{K}_{2n}}{s^{2n}}\,.$$

Here the numbers $\tilde{K}_{2q}$ are arbitrary. The particular choice $\tilde{K}_{2q} = K_{2q}$ where

$$(7.4) \qquad K_{2q} = 2(-1)^{q-1}(f^{(2q-1)}(1) - f^{(2q-1)}(0))/(2\pi)^{2q}$$

leads to the identification of $\tilde{E}_{2n+2}^{[s,1]}f$ with $E_{2n+2}^{[s,1]}f$ and in this case the sth term of the sum is $O(s^{-2n-2})$. In general, when $\tilde{K}_2 \neq K_2$, this sth term is $O(s^{-2})$ and the ultimate rate of convergence of this series is slower. (This is discussed in some detail in Section 9.)

The approximation to (7.2) which we shall consider has the following form:

$$(7.5) \qquad 2\tilde{C}^{(m)}f = \frac{\tilde{K}_2}{m^2} + \frac{\tilde{K}_4}{m^4} + \cdots + \frac{\tilde{K}_{2n}}{m^{2n}} + \sum_{s \leqq s/m} \mu_s \tilde{E}_{2n+2}^{[ms,1]} f\,.$$

This differs from the exact result (7.2) in that all terms $\tilde{E}_{2n+2}^{[s,1]}f$ with s greater than $\bar{s}$ have been removed from the right-hand side of (7.2) to form the right-hand side of (7.5).

This set of approximations is specified once the following information is available:

(7.6) (i) The value of n.

(7.7) (ii) The values of If, $\tilde{K}_2$, $\tilde{K}_4$, $\cdots$, $\tilde{K}_{2n}$.

(7.8) (iii) The value of $\bar{s}$.

The general discussion falls into three parts. In Section 8 we take the view that we want to construct a method of procedure using which the values of $\bar{s}$ and n are determined in the course of the calculation. In Section 9 we discuss standard practical procedures for determining the values of If and the parameter $\tilde{K}_{2q}$ and their relevance in this particular problem. In Section 10 we discuss theoretical properties of the approximation (7.5).

8. Determination of n and $\bar{s}$. Before dealing with the practical aspect of this calculation we derive first a simple theorem which relates $\bar{s}$ to ϵ, the required tolerance.

THEOREM 8.1. *In terms of definitions* (7.2), (7.3), *and* (7.5), *if $\bar{s}$ is an integer for which*

$$\sum_{s=\bar{s}+1}^{\infty} |\tilde{E}_{2n+2}^{[s,1]}f| < 2\epsilon, \tag{8.1}$$

then the set of approximation errors satisfy

$$|\tilde{C}^{(m)}f - C^{(m)}f| < \epsilon. \tag{8.2}$$

The proof is direct: we take the difference between (7.2) and (7.5). Using standard manipulation of inequalities, we find

$$\begin{aligned} |\tilde{C}^{(m)}f - C^{(m)}f| &= \left| \frac{1}{2} \sum_{s>\bar{s}/m} \mu_s \tilde{E}_{2n+2}^{[ms,1]} f \right| \\ &\leqq \frac{1}{2} \sum_{s>\bar{s}} |\tilde{E}_{2n+2}^{[s,1]}f| < \epsilon. \end{aligned} \tag{8.3}$$

Here we have used the inequality $|\mu_s| \leqq 1$, redefined the summation index, introduced nonnegative terms into the sum and applied (8.1).

Condition (8.1) is a sufficient, but not a necessary condition. Since $\tilde{E}_{2n+2}^{[s,1]}f \sim O(s^{-p})$ where $p \geqq 2$, there always exists a value of $\bar{s}$ satisfying (8.1). If any particular value of $\bar{s}$ satisfies (8.1), so does any greater integer.

In a practical implementation, the determination of n and $\bar{s}$ and the methods used to determine If and $\tilde{K}_{2q}$ are related to each other. For descriptive purposes it is convenient to suppose for the moment that a value of n has been assigned and the required numbers If, $\tilde{K}_{2q}$ $(q = 1, 2, \cdots, n)$ are already available. In this case one may proceed as follows. We calculate successively the values $\tilde{E}_{2n+2}^{[s,1]}f$, $s = 1, 2, 3, \cdots$. These values are given by (7.3) and each calculation requires the rule sum evaluation $R^{[s,1]}f$. This calculation is to be terminated at a point when we have just calculated $E_{2n+2}^{[s,1]}f$, $s = \bar{s}$ and we have reason to believe that criterion (8.1) is satisfied. It is necessary in practice to replace (8.1) by a practical convergence criterion. There are many ways of constructing such a criterion, but none are foolproof. A simple form might have four parts

(8.4) P.C. 1. Round-off error check,

(8.5) P.C. 2. Physical limit check,

(8.6) P.C. 3. $|E_{2n+2}^{[\bar{s},1]}f| < 2\epsilon$,

(8.7) P.C. 4. $|E_{2n+2}^{[s,1]}f| s = \bar{s} - 2, \bar{s} - 1, \bar{s}$ seem to form a suitable sequence.

P.C. 1 and P.C. 2 are normal guards which will terminate the calculation altogether if the round-off level is clearly higher than ϵ or if some physical limit set in the code is about to be exceeded. P.C. 3 is a simple criterion, which clearly must be satisfied before P.C. 4 can be considered. P.C. 4 may be as complicated as the user wishes. It should guard against a condition in which the sequence is converging very slowly, or the possibility of a single value of $E_{2n+2}^{[s,1]}f$ being very small. We do not go into any detail about these practical convergence criteria here.

Thus the calculation of these elements is terminated with the $\bar{s}$th term. $\bar{s}$ satisfies some practical convergence criterion P.C. 4 and hopefully it satisfies condition (8.1).

At this stage the set of numbers

$$\tilde{E}_{2n+2}^{[s,1]}f, \qquad s = 1, 2, \cdots, \bar{s}, \tag{8.8}$$

are available. For the calculation of $\tilde{C}^{(m)}f$ using (7.5) only a subset of this set is required, namely the set

$$\tilde{E}_{2n+2}^{[ms,1]}f, \qquad ms \leqq \bar{s}, \quad \mu_s \neq 0. \tag{8.9}$$

In fact if $m > \bar{s}$, the set (8.9) is empty and the sum in (7.5) may be replaced by zero. The theorem assures us that, so long as $\bar{s}$ in fact satisfies (8.1), then the calculated approximation $\tilde{C}^{(m)}f$ differs from the true value $C^{(m)}f$ by less than ϵ for all m.

The 'cost' of this calculation includes the following principal items

(8.10) (i) Evaluation of If,

(8.11) (ii) Assignment of $\tilde{K}_2, \tilde{K}_4, \cdots, \tilde{K}_{2n}$,

(8.12) (iii) Evaluation of $R^{[s,1]}f, \qquad s = 1, 2, 3, \cdots, \bar{s}$.

It is important to note that the same set of function values is used for all the Fourier coefficients, though all are not used explicitly in the evaluation of each coefficient. For example, in the cases where $m > \bar{s}$, no function values appear in the formula. But they were required previously in order to show that $m > \bar{s}$ by establishing the value of $\bar{s}$. Also, function values may have been used to determine the values of $\tilde{K}_{2q}$.

The description given above is restricted to a simplified situation in which the value of n is assigned and the values of $\tilde{K}_{2q}$ $(q = 1, 2, \cdots, n)$ are immediately available. As described above the values of $\tilde{E}_{2n+2}^{[s,1]}f$ actually encountered are used to determine the value of $\bar{s}$.

In a realistic situation, the use of an appropriate value of n is very important. The value of $\bar{s}$ depends on n as well as on ϵ and may be quite different for different values of n. Thus one may 'cut costs' under item (ii) by using a small value of n to find that this involves a large $\bar{s}$ and an increase in cost under item (iii).

To illustrate this dependence we have treated the example

$$f(x) = 1/(x^2 - x + (5/8)^2). \tag{8.13}$$

The values of If and K_{2q} have simple analytic expressions and these have been used to calculate $E_{2n+2}^{[s,1]}f$ for $2n = 0, 2, 4, 6, 8, 10$, $s = 1, 2, 3, \cdots, 10$. These numbers are set out in Table 2. The round-off level in the table is about 10^{-10}.

TABLE 2

s	$E_2{}^{[s,1]}f$	$E_4{}^{[s,1]}f$	$E_6{}^{[s,1]}f$	$E_8{}^{[s,1]}f$	$E_{10}^{[s,1]}f$	$E_{12}^{[s,1]}f$
1	−2.38557 + 000	−1.29331 + 000	−1.44990 + 000	−1.68387 + 000	−8.18568 − 001	−1.46674 − 001
2	−1.10019 − 001	1.63048 − 001	1.53261 − 001	1.49605 − 001	1.52985 − 001	1.53641 − 001
3	−1.33478 − 001	−1.21153 − 002	−1.40485 − 002	−1.43694 − 002	−1.42376 − 002	−1.42262 − 002
4	−6.62583 − 002	2.00841 − 003	1.39674 − 003	1.33962 − 003	1.35282 − 003	1.35346 − 003
5	−4.35555 − 002	1.35175 − 004	−1.15365 − 004	−1.30339 − 004	−1.28124 − 004	−1.28055 − 004
6	−3.02033 − 002	1.37471 − 004	1.66475 − 005	1.16326 − 005	1.21478 − 005	1.21589 − 005
7	−2.22253 − 002	6.59047 − 005	6.87138 − 007	−1.30160 − 006	−1.15150 − 006	−1.14912 − 006
8	−1.70275 − 002	3.91793 − 005	9.49938 − 007	5.73997 − 008	1.08976 − 007	1.09601 − 007
9	−1.34605 − 002	2.42762 − 005	4.09790 − 007	−3.04726 − 008	−1.03711 − 008	−1.01784 − 008
10	−1.09068 − 002	1.58851 − 005	2.26396 − 007	−7.57739 − 009	1.07562 − 009	1.14281 − 009

Let us suppose that we wanted a uniform accuracy in the result of 5.10^{-7} and that the practical convergence criterion included as part P.C. 3 (8.6) the requirement

$$|E^{[\bar{s},1]}_{2n+2}f| < 5.10^{-7}\,. \tag{8.14}$$

If we had assigned $n = 0$, we would have proceeded to calculate the elements in the first column of Table 2 until one element satisfied (8.14) with $n = 0$. This sequence converges as s^{-2} and reaches the desired level at about $\bar{s} = 1473$. If we had assigned $2n = 2$, we would require the value of K_2 to calculate the elements in the second column of Table 2, but we satisfy (8.14) with $\bar{s} = 24$. With $2n = 4, 6, 8, 10$ we find (8.14) is satisfied with $\bar{s} = 9, 8, 8, 8$, respectively. In retrospect therefore, in this calculation an appropriate value of $2n$ is 4 or 6. To use a smaller value involves an excessive number of function evaluations, while to use a larger value involves the calculation of further values of K_{2q} with no saving in the number of function evaluations.

In an automatic code, this information is not available at the start. Thus the code has to be arranged in such a way that it determines both n and $\bar{s}$ on the basis of the values of $\tilde{E}^{[s,1]}_{2q}f$ actually encountered. The initial aim of such a routine is to find a pair of values n, s which satisfy P.C. 3 (8.6). After this it may retain this value of n and proceed to attempt to satisfy the entire convergence criterion, increasing the value of s if necessary.

This first stage has a superficial resemblance to a minimization routine in two variables $\hat{q}$, $\hat{s}$, the function treated being $|\tilde{E}^{[s,1]}_{2q}f|$. Only unit steps in positive directions $\hat{q}$ and $\hat{s}$ are allowed in the search and at any moment, the list $\tilde{E}^{[s,1]}_{2\hat{q}}f$ $(s = 1, 2, \cdots, \hat{s})$ is available. If the next step involves increasing $\hat{s}$, one additional entry in the list should be made. If the next step involves increasing $\hat{q}$ the entire list is updated by the addition of the terms $\zeta(2q)\tilde{K}_{2\hat{q}}/s^{2\hat{q}}$. The routine should expect relatively smooth behavior of this discrete function in the direction of increasing $\hat{s}$, but not in the direction of increasing $\hat{q}$. Also there might be an adjustment built in by which the search routine assessed the cost of a step in the $\hat{s}$ direction against the cost of a step in the $\hat{q}$ direction. In straightforward cases all that is really needed is an upper bound on $\hat{q}$.

While a poorly constructed code can lead to unnecessary work, a high level in sophistication for this part of the code is not necessary. Any terminal value $\hat{q} = n$, $\hat{s} = \bar{s}$ gives results of suitable accuracy so long as the fourth part of the practical convergence criterion P.C. 4 (8.7) is adequate. All that this first stage should be capable of doing is to choose a value of n which is not totally unreasonable. In the example illustrated in Table 2, it should be able to increase n beyond $2n = 2$ and should not increase n beyond $2n = 8$.

Finally, using these particular values of n and $\bar{s}$, the calculation of any cosine Fourier coefficient is effected by direct substitution into (7.5). The values of $\tilde{E}^{[ms,1]}_{2n+2}f$ required have just been calculated while the values of $\tilde{K}_{2q}$ $(q = 2, 4, \cdots, n)$ were calculated or assigned as a by-product in that calculation.

If it is known beforehand that $f(x)$ is periodic having period 1, the calculation is much simpler since $K_{2q} = 0$ and the appropriate value of n is zero. However, the procedure described above may be used. If the function is periodic, it should appear that the values of $\tilde{K}_{2q}$ are all small; good progress in the minimization is occurring by increasing $\hat{s}$, while no progress is being made by increasing $\hat{q}$. The routine should therefore adjust to a periodic function automatically.

The approximation $\tilde{C}^{(3m)}f$ given by (7.5) is based on the exact expression (6.11) for $C^{(m)}f$. All that is necessary to obtain (7.5) from (6.11) is to replace the infinite sum over index s by a finite sum, the restriction being $s \leqq \bar{s}/m$. Two further approximations for $C^{(m)}f$ and an approximation $\tilde{S}^{(m)}f$ for $S^{(m)}f$ may be based on Eqs. (6.13), (6.14), and (6.15), respectively, by restricting the sum in an identical manner. We list here the resulting formulas:

$$(8.15)\quad 2\tilde{C}^{(m)}f = \frac{\tilde{K}_2}{m^2} + \frac{\tilde{K}_4}{m^4} + \cdots + \frac{\tilde{K}_{2n}}{m^{2n}} - \sum_{s \leqq s/m} \nu_s \tilde{E}_{2n+2}^{[ms,0]} f ,$$

$$(8.16)\quad 2\tilde{C}^{(m)}f = \frac{\tilde{K}_2}{m^2} + \frac{\tilde{K}_4}{m^4} + \cdots + \frac{\tilde{K}_{2n}}{m^{2n}} + \sum_{s \leqq \bar{s}/m \ (s \text{ odd})} \mu_s (\tilde{E}_{2n+2}^{[ms,1]} f - \tilde{E}_{2n+2}^{[2ms,1]} f) ,$$

$$(8.17)\quad \begin{aligned} 2\tilde{S}^{(m)}f &= \frac{\tilde{K}_1}{m} + \frac{\tilde{K}_3}{m^3} + \cdots + \frac{\tilde{K}_{2n-1}}{m^{2n-1}} \\ &\quad + \sum_{s \leqq \bar{s}/m \ (s \text{ odd})} (-1)^{(s-1)/2} \mu_s \frac{1}{2} (\tilde{E}_{2n+1}^{[ms,-1/2]} f - E_{2n+1}^{[ms,1/2]} f) , \end{aligned}$$

where the terms in the sum are

$$(8.18)\quad \tilde{E}_{2n+2}^{[s,0]} f = R^{[s,0]} f - If - \frac{\eta(2)\tilde{K}_2}{s^2} - \frac{\eta(4)\tilde{K}_4}{s^4} - \cdots - \frac{\eta(2n)\tilde{K}_{2n}}{s^{2n}} ,$$

$$(8.19)\quad \tilde{E}_{2n+2}^{[s,1]} f - \tilde{E}_{2n+2}^{[2s,1]} f = R^{[s,1]} f - R^{[2s,1]} - \frac{\lambda(2)\tilde{K}_2}{s^2} - \cdots - \frac{\lambda(2n)\tilde{K}_{2n}}{s^{2n}} ,$$

$$(8.20)\quad \begin{aligned} \frac{1}{2} (\tilde{E}_{2n+1}^{[s,-1/2]} f - \tilde{E}_{2n+1}^{[s,1/2]} f) &= \frac{1}{2} (R^{[s,-1/2]} f - R^{[s,1/2]} f) - \frac{\beta(1)\tilde{K}_1}{s} \\ &\quad - \frac{\beta(3)\tilde{K}_3}{s^3} - \cdots - \frac{\beta(2n-1)\tilde{K}_{2n-1}}{s^{2n-1}} . \end{aligned}$$

With minor modifications, the discussion of Sections 7 and 8 applies to any of these formulas. The quantity $E_{2n+2}^{[s,1[}f$ is simply replaced by one of the quantities on the right of Eqs. (8.18), (8.19) or (8.20). Theorem 8.1 is valid, except that in the case of (8.15) the numbers ν_s satisfy

$$(8.21)\qquad |\nu_s| \leqq s , \qquad s > 1$$

in place of $|\mu_s| \leqq 1$. Thus (8.1) is replaced by

$$(8.22)\qquad \sum_{s=\bar{s}+1}^{\infty} s |\tilde{E}_{2n+2}^{[s,0]} f| < 2\epsilon .$$

This in turn leads to a slightly more stringent practical convergence criterion P.C. 3, P.C. 4. The author has not come across any case in practice in which (8.15) seems to be preferable to (7.5).

The final two formulas (8.16) and (8.17) do not involve If, but involve slightly more sophisticated summation operators. These are alternating sums and can be expressed in various forms. For example

$$\begin{aligned} R^{[s,1]}f - R^{[2s,1]}f &= \tfrac{1}{2}\,(R^{[s,1]}f - R^{[s,0]}f) \\ &= (1/2s)[\tfrac{1}{2}\,f(0) - f(1/2s) + f(1/s) - f(3/2s) \\ &\qquad + \cdots + \tfrac{1}{2}\,f(1)]\,. \end{aligned} \tag{8.23}$$

This is sometimes called an alternating trapezoidal sum and uses the same function values as a trapezoidal rule, but with alternating signs. Similarly

$$\begin{aligned} \tfrac{1}{2}\,(R^{[s,-1/2]}f - R^{[s,1/2]}f) &= (1/2s)[f(1/4s) - f(3/4s) \\ &\qquad + f(5/4s) - \cdots - f((4s-1)/4s)] \end{aligned} \tag{8.24}$$

is known also as the alternating midpoint sum.

These formulas have a slightly different 'cost structure' from that listed in (8.10), (8.11), (8.12). Item (i) does not appear. Item (iii) is about twice as expensive; to obtain the same accuracy roughly the same value of $\bar{s}$ is involved, but $R^{[s,1]}f$ has to be replaced by one of the operators (8.23) or (8.24), which involve about twice the number of function values. In the calculation of $\tilde{S}^{(m)}f$ this additional expense is unavoidable. In the calculation of $C^{(m)}f$ if the value of If is known, the use of (7.5) in place of (8.16) leads to a much shorter calculation.

In formula (8.17), (8.20) the parameter $\tilde{K}_1$ is arbitrary. However, it should invariably be replaced by K_1 since this involves only function values:

$$K_1 = -2(f(1) - f(0))/(2\pi)\,. \tag{8.25}$$

9. Calculation of Parameter $\tilde{K}_{2q}$. In the implementation described in the previous section, the numbers $\tilde{K}_{2q}$ ($q = 1, 2, \cdots, n$) have been treated as parameters. In fact, the exact formulas such as (7.2), (7.3) are identities in the set of numbers $\tilde{K}_{2q}$ and are valid whatever choice is made. The choice $\tilde{K}_{2q} = K_{2q}$ where K_{2q} is given by

$$K_{2q} = 2(-1)^{q-1}(f^{(2q-1)}(1) - f^{(2q-1)}(0))/(2\pi)^{2q} \tag{9.1}$$

is suggested because this choice leads to a faster ultimate rate of convergence of the sequence $\tilde{E}^{[s,1]}_{2n+2}f$. Specifically, if we define ΔK_{2q} by

$$\tilde{K}_{2q} = K_{2q} + \Delta K_{2q} \tag{9.2}$$

then we have

$$E^{[s,1]}_{2n+2}f \sim O(s^{-(2n+2)}) \quad \text{as } s \to \infty\ , \tag{9.3}$$

but using (7.3)

$$\tilde{E}^{[s,1]}_{2n+2}f = -\frac{\zeta(2)\Delta K_2}{s^2} - \frac{\zeta(4)\Delta K_4}{s^4} - \cdots - \frac{\zeta(2n)\Delta K_{2n}}{s^{2n}} + E^{[s,1]}_{2n+2}f\,, \tag{9.4}$$

so that, unless $\tilde{K}_2 = K_2$ giving $\Delta K_2 = 0$, we have

$$\tilde{E}^{[s,1]}_{2n+2}f \sim O(s^{-2}) \quad \text{as } s \to \infty\ . \tag{9.5}$$

In the example of the previous section, we have seen the effect of choosing $\tilde{K}_2 = \tilde{K}_4 = \cdots = \tilde{K}_{2n} = 0$. This is the same as choosing $n = 0$ and is done at the cost of introducing a high value of $\bar{s}$ and consequently a large number of function evaluations.

If the function $f(x)$ is known in analytic form, the derivatives may in principle be calculated analytically. Depending on the structure of the function and the time available, it may be too tedious to do this after perhaps some low-order derivative has been expressed analytically. Hopefully, automatic algebraic manipulators may become more readily available and remove the necessity for the rest of this section.

The subsequent discussion is restricted to the cases in which analytic differentiation is not a viable alternative, and some numerical expedient based on function values $f(x_i)$ has to be used.

Before commencing such a calculation, or including the facility for such a calculation in an automatic code, one must give some attention to the accuracy required for these derivatives. The general situation here is one of balancing the cost of calculating K_{2q} accurately against the cost of calculating a possibly large number of the rule sum approximations $R^{[s,1]}f$. In fact, the discussion in the previous section about the choice of n represents an extreme case of precisely this sort of balance. There the choice presented was between extremes. Either $\tilde{K}_{2q} = 0$ or $\tilde{K}_{2q} = K_{2q}$. Here it is more delicate. With increasing effort we may attempt to make $\tilde{K}_{2q}$ successively closer to K_{2q}. At what point should we be content with the accuracy attained? The reason we calculate $\tilde{K}_{2q}$ at all is to try to arrange that $\bar{s}$, the value of s for which

$$(9.6) \qquad |\tilde{E}^{[s,1]}_{2n+2}f| < 2\epsilon ,$$

is as small as possible. A glance at (9.4) indicates that we would like the effect of the terms $\zeta(2q)\,\Delta K_{2q}/s^{2q}$ to have died out by the time the value $s = \bar{s}$ is reached. But in general the value of $\bar{s}$ is not known at this stage. However, if some estimate is available, the accuracy required might be chosen so as to satisfy

$$(9.7) \qquad \frac{\zeta(2q)\Delta K_{2q}}{\bar{s}^{2q}} < \frac{\epsilon}{n} , \qquad q = 1, 2, \cdots, n$$

or some similar criterion. If we define

$$(9.8) \qquad F(x) = f(x + 1) - f(x)$$

this requirement becomes

$$(9.9) \qquad |\Delta F^{(2q-1)}(0)| < \frac{(2\pi\bar{s})^{2q}}{2\zeta(2q)n}\,\epsilon , \qquad q = 1, 2, \cdots, n .$$

While this should not be treated as a precise relationship, it is qualitatively illuminating. For example if we are willing to go as far as $\bar{s} = 6$ (a total of 13 function evaluations for the rule sums) we find that the accuracy requirement for $F'(0)$ may be relaxed by a factor of about 900 for the calculation of $F'''(0)$. (However, one should bear in mind that it is the absolute accuracy which is under consideration here. The actual values of $|F^{(2q-1)}(0)|$ may increase with increasing q, leaving a much smaller factor in any calculation based on relative accuracy criteria.) If subsequently the estimate $\bar{s} = 6$ turns out to be too high, the use of these inaccurate derivatives may force the actual value of $\bar{s}$ up to 6. On the other hand, if subsequently the estimate $\bar{s} = 6$ turns out to be too low, we have used over-accurate approximations for the derivatives. In neither case need the calculation be abandoned or modified.

We now mention briefly three numerical methods which might be employed to calculate the derivatives. The first two could be applied directly to the function

$$F(x) = f(x+1) - f(x) \tag{9.10}$$

to evaluate the set $F^{(q)}(0)$, $q = 1, 2, \cdots, 2n$. (The even-ordered derivatives are required if the sine Fourier coefficients are also being calculated.) Whether or not the calculation is arranged to calculate $F^{(q)}(0)$, or to calculate $f^{(q)}(1)$ and $f^{(q)}(0)$ separately, any accuracy check at intermediate stages should be based on the value of $F^{(q)}(0)$. For example, if $f(x)$ is nearly periodic, $F^{(q)}(0)$ may be small while $f^{(q)}(1)$ and $f^{(q)}(0)$ are nearly equal larger numbers.

Method 1. *Finite-difference approximations.* Standard formulas and codes exist for the evaluation of derivatives in terms of tabular points. These are rarely used because of the undue amplification of round-off error in the final result. In this calculation, the use of inaccurate approximations for the derivatives is corrected at a later stage in the calculation. Essentially a formula of the type

$$F^{(q)}(0) \simeq \sum_{j=-N}^{N} a_{q,j} F(jh) \tag{9.11}$$

may be used, the approximation being exact apart from round-off error if $F(x)$ is a polynomial of degree $2N$ or less. These techniques are described in Milne-Thompson [15], Bickley [2], Kopal [10], and Ballester and Pereyra [16].

Method 2. *Interpolation for derivatives in the complex plane.* A different approach, which is convenient for obtaining approximations to a set of 'normalized' Taylor coefficients $r^s f^{(s)}(x_0)/s!$ with a uniform accuracy $\epsilon_{\text{T.C.}}$, is described in Lyness [13]. This requires that $f(z)$ is analytic within a region in the complex plane which includes the circle $|z - x_0| \leqq r$ and is based on complex function evaluations $f(z_i)$ at points on the circle $|z - x_0| = r$. The formula used for these approximations is

$$\frac{r^s f^{(s)}(x_0)}{s!} \simeq \frac{1}{N} \sum_{j=1}^{N} e^{-2\pi ijs/N} f(x_0 + re^{2\pi ij/N}), \qquad s = 0, 1, 2, \cdots, N-1, \tag{9.12}$$

and on the basis of the same set of N complex function values this formula provides approximations of polynomial degree $N - 1$ to $f(x_0)$ and its first $N - 1$ derivatives at x_0. So long as $f(z)$ is a real function of z when z is real and x_0 is real, only about $N/2$ separate complex function evaluations are required since advantage may be taken of conjugate pairs, i.e., $f(x_0 + re^{i\theta}) = [f(x_0 + re^{-i\theta})]^-$.

While a particular formula is specified once r and N are provided, an automatic code may be constructed (see [13]) in which r and $\epsilon_{\text{T.C.}}$ are provided and the routine attempts to determine N in such a way that the error $\Delta f^{(s)}(x_0)$ in the result satisfies

$$\frac{r^s |\Delta f^{(s)}(x_0)|}{s!} < \epsilon_{\text{T.C.}}, \qquad s = 0, 1, 2, 3, \cdots. \tag{9.13}$$

The routine then returns a set of normalized Taylor coefficients together with an error estimate which may be larger than $\epsilon_{\text{T.C.}}$ if round-off error has necessitated this but which is generally smaller than $\epsilon_{\text{T.C.}}$.

An automatic routine of this type requires input parameters r and $\epsilon_{\text{T.C.}}$. In view of (9.9) and (9.13) we should choose these to satisfy the set of inequalities

$$(9.14)\qquad \frac{\epsilon_{\text{T.C.}}(2q-1)!}{r^{2q-1}} < \frac{(2\pi\bar{s})^{2q}}{2\zeta(2q)n}\,\epsilon\,, \qquad q = 1, 2, \cdots, n\,.$$

It appears that if this inequality is satisfied for $q = 1$ and $q = n$ it is automatically satisfied for $q = 2, 3, \cdots, n - 1$. If $n = 1$ any choice satisfying

$$(9.15)\qquad \frac{\epsilon_{\text{T.C.}}}{r} < \frac{(2\pi\bar{s})^{2}}{2\zeta(2)}\,\epsilon$$

is satisfactory. For other values of n, an approximate solution of the equations obtained from (9.14) for $q = 1$ and $q = n$ by replacing the inequality by an equality is

$$(9.16)\qquad r \simeq \frac{2n+3}{2\pi\bar{s}e}\,; \qquad \epsilon_{\text{T.C.}} \simeq \bar{s}\epsilon \qquad (n \leqq 6)\,.$$

Method 3. *Global polynomial approximation based on trapezoidal rule approximations.* There is a technique, described in Lyness and Moler [14], which is designed to calculate precisely the quantities required. This is based on treating the Euler-Maclaurin formula in the same way as we treated the Poisson summation formula in Section 6. This leads to what is essentially a modification of Romberg integration. This technique makes use of precisely the rule sums which are being calculated in any case, and at first sight it seems that the derivatives may be calculated at an insignificant additional cost.

To illustrate the theory we consider Eq. (3.17), which may be written in the form

$$(9.17)\qquad R^{[m,1]}f - R^{[2m,1]}f = \sum_{q=1}^{[(p-1)/2]} \frac{\lambda(2q)K_{2q}}{m^{2q}} + E_p{}^{[m,1]}f - E_p{}^{[2m,1]}f\,.$$

If we set $p = 2N + 2$ and write this equation down for N distinct values of m, say $m_1, m_2, \cdots, m_N$, and disregard the remainder terms, the resulting N equations may be considered to be a set of linear equations in the unknowns K_{2q}, $q = 1, 2, \cdots, N$. In fact, should $f(x)$ be a polynomial of degree $2N + 1$ or less, these equations would be exact since in that case the remainder terms are precisely zero.

This set of equations has an associated matrix of the Vandermonde type which may be inverted analytically. However, there exists (see [14]) a generalization of the Neville-Romberg algorithm. Using this, the calculation may be undertaken in a manner which is a slight generalization of Romberg integration. That is, a solution for $K_2, K_4, \cdots, K_{2N}$ based on mesh ratios $m_1, m_2, \cdots, m_N$ may be up-dated after the calculation of $R^{[m,1]}f - R^{[2m,1]}f$, $m = m_{N+1}$, by extending a generalized T-table.

If Eq. (3.15) is used instead of (3.17), the procedure includes a standard Romberg integration as a subset of the calculation.

This method fits very neatly into the general theory. In earlier versions of an automatic code this method was used and its defects were discovered experimentally. Like the other methods described above, it provides an approximation for $f^{(q)}(1) - f^{(q)}(0)$ of polynomial degree $2N + 1$. However, this method relies on global polynomial approximation (over the whole interval [0, 1]) rather than local polynomial approximation in the neighbourhood of an end point. Thus to provide accurate approximations the function $f(x)$ should approximate a polynomial over the entire interval. Otherwise, grossly inaccurate approximations are obtained.

A second extremely annoying feature of this method is that it may interact with the rest of the calculation. This is illustrated below by a simple example, in which Eq. (7.5) is being used.

We suppose that $2n = 8$ and that the derivatives and the integral are calculated, using mesh ratios m_1, m_2, m_3, m_4, m_5. In this case we have calculated approximations $\tilde{I}f$ and $\tilde{K}_{2q}$, $q = 1, 2, 3, 4$, which satisfy the equations

$$R^{[m_i,1]}f - \tilde{I}f = \sum_{q=1}^{4} \frac{\zeta(2q)\tilde{K}_{2q}}{m_i^{2q}}, \qquad i = 1, 2, 3, 4, 5 .$$

These may be written

$$\tilde{E}_{10}^{[m_i,1]}f = 0 , \qquad i = 1, 2, 3, 4, 5 .$$

Thus when we come to calculate the set of values

$$\tilde{E}_{2n+2}^{[s,1]}f , \qquad s = 1, 2, 3, \cdots , 2n + 2 = 10$$

we should find the five members of this set for which $s = m_1, m_2, m_3, m_4$, and m_5 to be identically zero (rounding errors apart). That is to say we have managed to choose approximations to the derivatives in a manner specifically designed to upset the convergence criterion.

Once this situation is noted, it is quite easy to take care to see that it is not taken as an indication of convergence. However, the interaction may not be as specific as this extreme example indicates; the apparent gain (obtaining derivative approximations at no additional cost) may be completely offset by having to use a much more carefully constructed practical convergence criterion and consequently additional function evaluations.

10. The Approximation Error $\tilde{C}^{(m)}f - C^{(m)}f$. In the previous two sections, the emphasis of the discussion is on how to apply the approximation formula to obtain results of specified numerical accuracy. In this section we look at the resulting approximation and derive some simple theoretical properties of the error functional $\tilde{C}^{(m)}f - C^{(m)}f$. The approximation $\tilde{C}^{(m)}f$ is specified once the following information is available:

(i) The values of parameters n and $\bar{s}$.

(ii) The values of parameters $\tilde{K}_{2q}$ $(q = 1, 2, \cdots, n)$.

The principal results in this section involve $\tilde{K}_{2q}$ only through $\Delta K_{2q} = \tilde{K}_{2q} - K_{2q}$, where K_{2q} is given by (7.4).

We now discuss the polynomial degree of the approximation $\tilde{C}^{(m)}f$. If $f(x)$ happens to be a polynomial of degree $2n + 1$ both the Fourier coefficient asymptotic expansion (3.3) and the Euler-Maclaurin asymptotic expansion (3.15) are finite series having n terms. The remainder terms satisfy

$$C_{2n+2}^{(m)}f = 0 , \qquad E_{2n+2}^{[s,1]}f = 0 \tag{10.1}$$

since both integral representations (3.5) and (3.9) involve an integrand with factor $f^{(2n+2)}(x)$ and this is zero. This introduces considerable simplifications into many of the formulas we have derived. Thus we may write in place of the exact result (7.2), (7.3) the simpler formula

$$(10.2) \qquad 2\,C^{(m)}f = \sum_{q=1}^{n} \frac{K_{2q}}{m^{2q}}.$$

Also, since

$$(10.3) \qquad \tilde{E}^{[s,1]}_{2n+2}f = -\sum_{q=1}^{n} \frac{\zeta(2q)\Delta K_{2q}}{s^{2q}} + E^{[s,1]}_{2n+2}f$$

we may express (7.5) in the form

$$(10.4) \qquad 2\tilde{C}^{(m)}f = \sum_{q=1}^{n} \frac{\tilde{K}_{2q}}{m^{2q}} - \sum_{s\leqq \bar{s}/m} \mu_s \sum_{q=1}^{n} \frac{\zeta(2q)\Delta K_{2q}}{(ms)^{2q}}.$$

It follows from (10.2) and (10.4) that

$$(10.5) \qquad 2(\tilde{C}^{(m)}f - C^{(m)}f) = \sum_{q=1}^{n} \frac{\Delta K_{2q}}{m^{2q}} - \sum_{s\leqq \bar{s}/m} \mu_s \sum_{q=1}^{n} \frac{\zeta(2q)\Delta K_{2q}}{(ms)^{2q}}.$$

If the quantities ΔK_{2q} $(q = 1, 2, \cdots, n)$ are also zero then the right-hand side of (10.5) is zero and the approximation $\tilde{C}^{(m)}f$ is exact. Naturally, ΔK_{2q} is zero if exact values $\tilde{K}_{2q} = K_{2q}$ have been used. Also ΔK_{2q} is zero for the functions under consideration (polynomials of degree $2n + 1$) if the derivatives $f^{(2q-1)}(1) - f^{(2q-1)}(0)$ which occur in K_{2q} have been approximated using a method which is exact if $f(x)$ is a polynomial of degree $2n + 1$. Specifically, all three methods given in the previous section have this property so long as the parameters N, $2N - 1$, N in Methods 1, 2, and 3, respectively, exceed $2n$. We state this result as a theorem.

THEOREM 10.6. *The approximation $\tilde{C}^{(m)}f$ given by* (7.5) *is exact for polynomial functions $f(x)$ of degree $2n + 1$ so long as $\Delta K_{2q} = K_{2q} - \tilde{K}_{2q}$ $(q = 1, 2, \cdots, n)$ is zero for such functions.*

We now consider the trigonometric degree of the approximation. If $f(x)$ is a trigonometric polynomial of degree $\bar{s}$, it has the form

$$(10.7) \qquad f(x) = A_0 + \sum_{r=1}^{\bar{s}} A_r \cos 2\pi rx + \sum_{r=1}^{\bar{s}} B_r \sin 2\pi rx\,.$$

Two results follow readily. These are:

$$(10.8) \qquad K_q = 0\,, \qquad q \geqq 1\,,$$

$$(10.9) \qquad R^{[s,1]}f = If = A_0\,, \qquad s > \bar{s}\,.$$

Consequently, if we set $\tilde{K}_{2q} = K_{2q}$ in the exact result (7.2), (7.3) we find

$$(10.10) \qquad 2C^{(m)}f = \sum_{s\leqq \bar{s}/m} \mu_s(R^{[ms,1]}f - If)$$

while the approximation (7.5) has the form

$$(10.11) \quad 2\tilde{C}^{(m)}f = \sum_{q=1}^{n} \frac{\Delta K_{2q}}{m^{2q}} + \sum_{s\leqq \bar{s}/m} \mu_s\left(R^{[ms,1]}f - If - \sum_{q=1}^{n} \frac{\zeta(2q)\Delta K_{2q}}{(ms)^{2q}}\right).$$

These are clearly identical so long as $\Delta K_{2q} = 0$ $(q = 1, 2, \cdots, n)$. This happens automatically if Methods 1 and 2 for the derivatives have been applied to

$$(10.12) \qquad F(x) = f(x + 1) - f(x)$$

since all function evaluations of $F(x)$ are zero; but it generally does not happen if Method 3 is used, nor if separate applications of the same method have been used to approximate $f^{(q)}(1)$ and $f^{(q)}(0)$.

THEOREM 10.13. *The approximation* $\tilde{C}^{(m)}f$ *given by* (7.5) *is exact for trigonometric polynomials* $f(x)$ *of degree* $\bar{s}$ ((10.7) *above*) *so long as* ΔK_{2q} $(q = 1, 2, \cdots, n)$ *is zero for such functions* (*or if* $n = 0$).

We now derive an approximation error bound of a conventional nature. This bound is similar to standard error bounds for quadrature rules of specified degree in that it contains a term with a factor

$$(10.14) \qquad M_{2n+2} = \max_{0 \leqq x \leqq 1} |f^{(2n+2)}(x)| .$$

It also contains terms having coefficients ΔK_{2q}, since unless these are zero, the result is not of polynomial degree $2n + 1$.

We deal with the simpler case in which $m > \bar{s}$ first. In this case, the calculated value $\tilde{C}^{(m)}f$ is simply

$$(10.15) \qquad 2\tilde{C}^{(m)}f = \sum_{q=1}^{n} \frac{\tilde{K}_{2q}}{m^{2q}}, \qquad m > \bar{s}$$

while the exact value may be expressed in terms of its asymptotic expansion (3.3) or

$$(10.16) \qquad 2C^{(m)}f = \sum_{q=1}^{n} \frac{K_{2q}}{m^{2q}} + 2C^{(m)}_{2n+2}f .$$

The approximation error is therefore

$$(10.17) \qquad 2(\tilde{C}^{(m)}f - C^{(m)}f) = \sum_{q=1}^{n} \frac{\Delta K_{2q}}{m^{2q}} - 2C^{(m)}_{2n+2}f, \qquad m > \bar{s} .$$

Applying the intermediate value theorem to expression (3.5) for the remainder term, we find

$$(10.18) \qquad |\tilde{C}^{(m)}f - C^{(m)}f| \leqq \frac{1}{2} \sum_{q=1}^{n} \frac{|\Delta K_{2q}|}{m^{2q}} + \frac{M_{2n+2}}{(2\pi m)^{2n+2}}, \qquad m > \bar{s} .$$

If we express ΔK_{2q} in terms of $\Delta F^{(2q-1)}(0)$ as in (9.8) this gives

$$(10.19) \qquad |\tilde{C}^{(m)}f - C^{(m)}f| \leqq \sum_{q=1}^{n} \frac{|\Delta F^{(2q-1)}(0)|}{(2\pi m)^{2q}} + \frac{M_{2n+2}}{(2\pi m)^{2n+2}}, \qquad m > \bar{s} .$$

This depends on $\bar{s}$ in the sense that it is valid only if $m > \bar{s}$.

We now proceed to the more complicated case, that in which $m \leqq \bar{s}$. Here we shall obtain a bound of the same general structure. The difference is that factor $(2\pi m)^{2q}$ occurring in the denominator will be replaced by $(2\pi(\bar{s} + 1))^{2q-1}$ and certain different multiplying constants occur as coefficients in each term. The bound is rather pessimistic since certain sums and integrals are bounded in magnitude by sums and integrals of the corresponding absolute quantities in a conventional manner.

We proceed, as in the derivation of Theorem 8.1, to take the difference between $C^{(m)}f$ given by (7.2), (7.3) and $\tilde{C}^{(m)}f$ given by (7.5). Taking into account relation (9.4), we find

$$(10.20)\qquad 2(\tilde{C}^{(m)}f - C^{(m)}f) = -\sum_{s>\bar{s}/m} \mu_s\left(E_{2n+2}^{[ms,1]}f - \sum_{q=1}^{n}\frac{\zeta(2q)\Delta K_{2q}}{(ms)^{2q}}\right).$$

This is of course an exact relation. In fact, in the case $m > \bar{s}$, Eq. (10.17) above may be derived from it by Möbius inversion. We proceed to calculate separate bounds for each of the $n + 1$ terms on the right-hand side of (10.20). Some of the details of this calculation are given in the Appendix. In particular, we use the inequality

$$(10.21)\qquad \left|\sum_{s>\bar{s}/m} \mu_s/(ms)^{2q}\right| < \frac{7}{4}\frac{1}{(\bar{s}+1)^{2q-1}}, \qquad m \geqq 1, \quad \bar{s} \geqq 1, \quad q \geqq 1$$

whose proof involves placing a bound on the generalized zeta function $\zeta(s, a)$. Using this we see that the magnitude of the qth term in the sum over q in (10.20) is bounded by

$$(10.22)\qquad \zeta(2q)|\Delta K_{2q}|\left|\sum_{s>\bar{s}/m}\frac{\mu_s}{(ms)^{2q}}\right| < \frac{7}{4}\frac{\zeta(2q)|\Delta K_{2q}|}{(\bar{s}+1)^{2q-1}} = \frac{7}{4\pi}\frac{\zeta(2q)|\Delta F^{(2q-1)}(0)|}{(2\pi(\bar{s}+1))^{2q-1}}.$$

The first term on the right of (10.20) (and the only term which occurs if $\Delta K_{2q} = 0$) may be bounded if we use the integral representation (3.9) for the error functional. Thus

$$(10.23)\qquad \begin{aligned} |E_{2n+2}^{[s,1]}f| &= \frac{1}{s^{2n+2}}\left|\int_0^1 f^{(2n+2)}(x)\,\frac{B_{2n+2} - \bar{B}_{2n+2}(1 - sx)}{(2n+2)!}\,dx\right| \\ &< \frac{1}{s^{2n+2}} M_{2n+2}\frac{|B_{2n+2}|}{(2n+2)!} = \frac{2M_{2n+2}\zeta(2n+2)}{(2\pi s)^{2n+2}}. \end{aligned}$$

Here M_{2n+2} is an upper bound on $|f^{(2n+2)}(x)|$ given by (10.14) and we have used the fact that the kernel function in the integrand is of definite sign and also the identity (3.12). A calculation similar to that which led to (10.22) yields

$$(10.24)\qquad \begin{aligned} \left|\sum_{s>\bar{s}/m}\mu_s E_{2n+2}^{[ms,1]}f\right| &< \sum_{s=\bar{s}+1}^{\infty}|E_{2n+2}^{[s,1]}f| < \frac{2M_{2n+2}\zeta(2n+2)}{(2\pi)^{2n+2}}\sum_{k=\bar{s}+1}^{\infty}\frac{1}{k^{2n+2}} \\ &< \frac{7}{4\pi}\cdot\frac{M_{2n+2}\zeta(2n+2)}{(2\pi(\bar{s}+1))^{2n+1}}. \end{aligned}$$

Introducing inequalities (10.22) and (10.24) into (10.20) we find the approximation error bound to be

$$(10.25)\qquad |\tilde{C}^{(m)}f - C^{(m)}f| < \sum_{q=1}^{n}\frac{7\zeta(2q)}{8\pi}\,\frac{|\Delta F^{(2q-1)}(0)|}{(2\pi(\bar{s}+1))^{2q-1}} + \frac{7}{8\pi}\,\frac{\zeta(2n+2)M_{2n+2}}{(2\pi(\bar{s}+1))^{2n+1}}.$$

The factor $7\zeta(2q)/8\pi$ lies between 1/2 and 1/4. This bound is valid for all m. It is natural to compare it with the bound (10.19), which is valid for $m > \bar{s}$ only. That bound is clearly less extravagant. The inequality (10.25) is independent of m, while that in (10.19) depends on $\bar{s}$ implicitly through the condition $m > \bar{s}$. A rough overall bound which retains the essential features of both is obtained by replacing $\bar{s} + 1$ in (10.25) by the quantity max $(m, \bar{s} + 1)$ wherever it occurs. In this way the overall accuracy is reflected in the bound. There is a roughly constant accuracy for $m \leqq \bar{s} + 1$. For higher values of m the accuracy increases with m.

11. Discussion. The methods described in this paper for the calculation of Fourier coefficients all stem from one of Eqs. (6.11), (6.13), (6.14), and (6.15). These equations are fundamentally variant forms of the result of Möbius inversion of the Poisson summation formula. It is convenient to refer to these methods collectively as "The Calculation of Fourier Coefficients by Möbius Inversion of the Poisson Summation Formula" which will be abbreviated by the initial letters MIPS.

In recent years there have been many different methods suggested in the literature. In the interests of brevity we consider only the two which are possibly most familiar. These are

1. *Finite Version of the Fast Fourier Transform* (FFT).

$$\tilde{C}^{(m)}f = R^{[2s,1]}\phi_m \tag{11.1}$$

where

$$\phi_m(x) = f(x) \cos 2\pi m x\,. \tag{11.2}$$

2. *Filon-Luke Formulas* (FLF). These are of the form

$$\tilde{C}^{(m)}f = \alpha(\theta_m)R^{[2s,1]}\phi_m \tag{11.3}$$

$$\tilde{C}^{(m)}f = \beta(\theta_m)R^{[s,1]}\phi_m + \gamma(\theta_m)R^{[s,0]}\phi_m \tag{11.4}$$

where $\phi_m(x)$ is defined above, and

$$\theta_m = 2\pi m/2s \tag{11.5}$$

$$\alpha(\theta) = (\sin 2\theta/2\theta)^2 \tag{11.6}$$

$$\beta(\theta) = 2[\theta(1 + \cos\theta) - 2\sin\theta\cos\theta]/\theta^3 \tag{11.7}$$

$$\gamma(\theta) = 4[\sin\theta - \theta\cos\theta]/\theta^3\,. \tag{11.8}$$

Formula (11.4) is known as Filon's Rule (Filon [5]). A set of formulas of which (11.3) and (11.4) are the first two members has been derived by Luke [12].

The Fast Fourier Transform (FFT) method is designed for a particular set of circumstances. In general, an infinite integral is being approximated by a finite integral. Thus any polynomial approximation is not really appropriate since the functions involved do not approximate polynomials globally. Then in the calculation of a set $\tilde{C}^{(m)}f$ ($m = 1, 2, \cdots, 2s$), the user is not interested in individual accuracy, but rather in the properties of this set of numbers as a whole (Gentleman and Sande [6]). Usually the general situation is one in which function values $f(x_i)$ at regularly spaced intervals may be obtained at virtually no cost. The principal cost is the organization of the calculation of the set of quantities $\tilde{C}^{(m)}f$, $m = 1, 2, \cdots, 2s$ from the two sets $f(j/2s)$, $j = 1, 2, \cdots, 2s$, and $\cos(2\pi j/2s)$, $j = 1, 2, \cdots, s/2$. A great amount of ingenuity has been expended on this particular data handling problem (Cooley and Tukey [3]).

The Filon-Luke formulas (FLF), and the methods based on Möbius inversion (MIPS) described here are more appropriate in a rather different set of circumstances. Here a general function, rather than one derived from approximating an infinite interval by a finite interval, is being considered. The intention is to obtain accurate individual approximations. And the cost of function evaluations is the significant cost.

The possible user should obviously decide to what extent his particular problem conforms to either of these two significantly different situations.

We give under several headings below what we consider to be the significant properties of the three methods. Together with the remarks made above, a possible user should check this list to see which method seems to be most appropriate in his case. The list also brings out some major theoretical differences.

1. Each of the rules may, if necessary, be expressed in the form

$$\tilde{C}^{(m)} f = \sum_j W_j^{(m)} f(x_j) \,, \qquad m = 1, 2, 3, \cdots .$$

That is, in terms of a single set of function values, a set of Fourier coefficients may be calculated by assigning different weights to each function value for each different Fourier coefficient.

2. Unless $f(x)$ and some of its early derivatives are continuous, none of these methods is particularly efficient. However, each is 'robust' in the sense that each will ultimately give a sufficiently close approximation if enough function values are used so long as $f(x)$ is continuous. Thus,

$$\lim_{s \to \infty} \tilde{C}^{(m)} f = C^{(m)} f \,; \qquad \lim_{\bar{s} \to \infty} \tilde{C}^{(m)} f = C^{(m)} f \,; \qquad f \in C[0, 1] \,.$$

3. *Simplicity of Calculation.* Both the FFT and the FLF require the set of coefficients cos $(2\pi j/2s)$. In each of these a calculation such as that carried out by the Cooley-Tukey algorithm is necessary. The FLF require in addition the evaluation of coefficients such as $\beta(\theta_m)$, $\gamma(\theta_m)$ and a subsequent calculation. On the other hand, the MIPS calculation requires as data the values of μ_s $(s = 1 \cdots \bar{s})$ and the Bernoulli numbers B_{2q} $(q = 1, 2, \cdots, n)$ (or in the case of the sine Fourier coefficient, Euler numbers E_{2q-1} $(q = 1, 2, \cdots, n)$). The coefficients cos $(2\pi j/2s)$ do not occur explicitly.

4. *Flexibility and Error Criterion.* In an actual calculation it is sometimes necessary to subsequently improve the accuracy of the approximation. In fact, if the intention of the user is to obtain an approximation of specified accuracy ϵ, it is difficult to use either the FFT or FLF methods unless approximations corresponding to different values of s are obtained and the accuracy estimated by comparing these different numerical results. In either case, the only reasonable option is to use values $s = s_1, s_2, \cdots$ where $s_i = 2s_{i-1}$. In this way all previously calculated function values are used, but the cost of each step is approximately the same as the cost of all the previous steps put together, and provides a considerable increase in accuracy.

On the other hand, the accuracy of the MIPS approximation may be increased by increasing the value of $\bar{s}$ by 1, as described in Section 8. This obtains a marginal improvement at a marginal additional cost. Proceeding in this way one ultimately uses an appropriate value of $\bar{s}$ automatically.

5. *Additional Information.* A property of the MIPS method not shared by other methods is that information such as the value of If or the values of the derivatives $f^{(q)}(0), f^{(q)}(1)$ may be incorporated in a simple manner into the formula. This has the effect of reducing the number of function values required.

6. *Polynomial and Trigonometric Degrees.* It was shown in Section 10 that under certain conditions the MIPS approximation has degrees $2n + 1$, $\bar{s}$ respectively. The

corresponding degrees for the FFT are 0, $2s - m$. The FLF are constructed to have specific polynomial degrees. The first two members of this sequence (11.3) and (11.4) have polynomial degrees 1 and 3 respectively. All have zero trigonometric degree.

The comparison of (11.1) and (11.3) is interesting in this context. They differ only through the factor $\alpha(\theta_m)$. Thus (11.1) is exact for trigonometric polynomials of degree $2s - m$, but not for the function $f(x) = x$, while (11.3) is exact for the function $f(x) = x$, but not for the trigonometrical polynomials. On the other hand (11.1) gives an absurd result for $m > 2s$, i.e., $\tilde{C}^{(m+2s)}f = \tilde{C}^{(m)}f$ while (11.3) may give an inaccurate result, but one of the correct order $O(m^{-2})$.

7. *Location of Abscissas.* Both the FFT and the FLF require function values $f(j/2s)$ located at equal intervals. This is particularly convenient if $f(x)$ is tabulated at equal intervals and there is an integer number of such intervals in the interval [0, 1]. It may also be convenient if the function values have to be derived using a recurrence relation, as perhaps in the solution of a differential equation. The MIPS method is not so convenient. The function values required are $f(j/k)$ for $j = 0, 1, \cdots, k$ and $k = 1, 2, \cdots, \bar{s}$. While in general many fewer function values may be required, the particular location of the abscissas may introduce some complication at an earlier stage of a large scale calculation.

In the cases in which a function subroutine is available, the actual location of the abscissas is not important, and all methods are equally convenient in this respect.

8. *Number of Function Values per Period.* It is stated in many books such as Davis and Rabinowitz [4], Hamming [8] that a requirement for a meaningful calculation is that the function values occur sufficiently densely so that each period of the function $f(x) \cos 2\pi mx$ includes more than one function value. This certainly seems to be valid if either the FFT or the FLF are used. It is an interesting feature of the MIPS method that there is no restriction of this type. In the example in Section 8 the first 1000 Fourier coefficients are greater than $\epsilon = 10^{-6}$. However, these were calculated using only 33 function values together with the exact derivatives and the value of the exact integral. If a formula based on (8.23) is used, 241 function values are required explicitly and a further 24 to obtain adequate derivatives numerically. The integral If is not required. Based on 265 function values, any of the integrals $C^{(m)}f$, $1 < m < 1000$ may be calculated in a meaningful manner. The period $1/m$ of any particular integrand does not enter into the calculation at the stage when function evaluations are being made and so does not affect their location.

The author does not wish to give any value judgement on the respective merits of the three methods discussed in this section. Several sets of numerical calculations have confirmed that there are examples in which any of these might be considered superior. In Part II various extensions of the MIPS method will be presented to handle problems for which no standard method exists. These extensions introduce more sophisticated coefficients, but otherwise have a close resemblance to the methods described here.

Appendix 1. *Incidental Constants Occurring in Formulas.* The MIPS routines require surprisingly few constants. The Riemann zeta function and its variants occur only in the following combinations:

$$\frac{2(-1)^{q-1}\zeta(2q)}{(2\pi)^{2q}} = \frac{B_{2q}}{(2q)!},$$

$$\frac{2(-1)^{q}\eta(2q)}{(2\pi)^{2q}} = \frac{2-2^{2q}}{2^{2q}}\,\frac{B_{2q}}{(2q)!},$$

$$\frac{2(-1)^{q-1}\lambda(2q)}{(2\pi)^{2q}} = \frac{2^{2q}-1}{2^{2q}}\,\frac{B_{2q}}{(2q)!},$$

$$\frac{2(-1)^{q-1}\beta(2q-1)}{(2\pi)^{2q-1}} = \frac{E_{2q-2}}{4^{2q-1}(2q-2)!},$$

B_{2q} and E_{2q} are Bernoulli and Euler numbers (see Abramowitz and Stegun [1, p. 810]). The early values (those needed in all but the most extensive calculation) are:

$$B_{2q}\ (q = 1 \cdots 6): \frac{1}{6}, -\frac{1}{30}, \frac{1}{42}, -\frac{1}{30}, \frac{5}{66}, -\frac{691}{2730};$$

$$E_{2q}\ (q = 0 \cdots 6): 1, -1, 5, -61, 1385, -50521, 2702765\,.$$

The other constants required are Möbius numbers (see (5.5)), and the value of 2π.

Appendix 2. *Bound on* $\zeta(q, a)$, $a > \frac{1}{2}$, $q > 1$. The bound given here is useful for large values of a. The function $f(x) = x^{-q}$ is convex downwards ($f''(x) > 0$) for $x > 0$. Consequently any midpoint trapezoidal rule approximation to the integral

$$\int_{a-1/2}^{\infty} x^{-q}dx = \frac{(a-\frac{1}{2})^{1-q}}{1-q}$$

gives a lower bound to this integral. It follows that

$$\zeta(q, a) = \frac{1}{a^q} + \frac{1}{(a+1)^q} + \cdots < \frac{(a-\frac{1}{2})^{1-q}}{1-q}$$

and by elementary manipulation

$$\zeta(q, a) < \frac{1}{a^q}\left\{1 + \frac{a+\frac{1}{2}}{q-1}\right\} \leqq \frac{1}{a^{q-1}}\left\{\frac{3/2}{a(q-1)} + \frac{1}{q-1}\right\}; \qquad q \geqq 2\,.$$

Appendix 3. *Proof of Inequality* (10.21). Here $\bar{s}$, q and m are all positive integers. Let a be the smallest integer greater than $\bar{s}/m$. Then $a \geqq (\bar{s}+1)/m$ and the following set of inequalities are valid:

$$\left|\sum_{s>\bar{s}/m} \mu_s/(ms)^{2q}\right| < \frac{1}{m^{2q}}\sum_{s>\bar{s}/m} 1/s^{2q} = \frac{\zeta(2q, a)}{m^{2q}}$$

$$< \frac{1}{m^{2q}}\zeta(2q, (\bar{s}+1)/m)\,.$$

Applying the inequality in Appendix 2, we find

$$\frac{1}{m^{2q}}\zeta(2q, (\bar{s}+1)/m) < \frac{1}{(\bar{s}+1)^{2q-1}}\left[\frac{3/2}{(\bar{s}+1)(2q-1)} + \frac{1}{(2q-1)m}\right].$$

The quantity in square brackets is less than 7/4 when $\bar{s}$, q, and m are all positive integers. This establishes inequality (10.21).

Argonne National Laboratory
Argonne, Illinois 60439

1. M. Abramowitz & I. A. Stegun (Editors), *Handbook of Mathematical Functions, with Formulas, Graphs, and Mathematical Tables,* Dover, New York, 1966. MR **34** #8606.
2. W. G. Bickley, "Formulae for numerical differentiation," *Math. Gaz.,* v. 25, 1941, pp. 19–27. MR **2**, 240.
3. J. W. Cooley & J. W. Tukey, "An algorithm for the machine calculation of complex Fourier series," *Math. Comp.,* v. 19, 1965, pp. 297–301. MR **31** #2843.
4. P. Davis & P. Rabinowitz, *Numerical Integration,* Blaisdell, Waltham, Mass., 1967. MR **35** #2482.
5. L. N. G. Filon, "On a quadrature formula for trigonometric integrals," *Proc. Roy. Soc. Edinburgh,* v. 49, 1929, pp. 38–47.
6. W. M. Gentleman & G. Sande, *Fast Fourier Transforms for Fun and Profit,* Proc. AFIPS 1966 Fall Joint Computer Conf., v. 29, 1966, pp. 563–578.
7. R. R. Goldberg & R. S. Varga, "Moebius inversion of Fourier transforms," *Duke Math. J.,* v. 23, 1956, pp. 553–559. MR **18**, 304.
8. R. W. Hamming, *Numerical Methods for Scientists and Engineers,* McGraw-Hill, New York, 1962. MR **25** #735.
9. G. H. Hardy & E. M. Wright, *An Introduction to the Theory of Numbers,* Clarendon Press, Oxford, 1954. MR **16**, 673.
10. Z. Kopal, *Numerical Analysis,* Wiley, New York, 1955. MR **17**, 1007.
11. E. Landau, *Vorlesungen über Zahlentheorie.* Band II, Chelsea, New York, 1947.
12. Y. L. Luke, "On the computation of oscillatory integrals," *Proc. Cambridge Philos. Soc.,* v. 50, 1954, pp. 269–277. MR **15**, 992.
13. J. N. Lyness, "Quadrature methods based on complex function values," *Math. Comp.,* v. 23, 1969, pp. 601–619.
14. J. N. Lyness & C. B. Moler, "Generalised Romberg methods for integrals of derivatives," *Numer. Math.* (To appear.)
15. L. M. Milne-Thompson, *The Calculus of Finite Differences,* Macmillan, London, 1933.
16. C. Ballester & V. Pereyra, "On the constructions of discrete approximations to linear differential expressions," *Math. Comp.,* v. 21, 1967, pp. 297–302.