JOURNAL OF COMPUTATIONAL PHYSICS 54, 57–73 (1984)

The Calculation of Trigonometric Fourier Coefficients*

J. N. LYNESS

Mathematics and Computer Science Division,
Argonne National Laboratory, Argonne, Illinois 60439

Received June 15, 1981

A method for the numerical evaluation of trigonometric Fourier coefficients is described. This method is suitable for sufficiently smooth functions and is essentially a modification of the standard method based on trapezoidal rule sums using, optionally, the fast Fourier transform.

INTRODUCTION

In this paper we provide a description and justification of a method for the numerical evaluation of sets of Fourier coefficients

$$\frac{2}{B-A}\int_A^B f(x) \begin{matrix}\cos\\ \sin\end{matrix} \left(\frac{2\pi r(x-A)}{B-A}\right) dx, \qquad r = 0, 1, 2, \ldots,$$

in terms of equally spaced sets of function values

$$f(A + j(B-A)/m), \qquad j = 0, 1, 2, \ldots, m.$$

The method is restricted to functions $f(x)$ which are sufficiently smooth in an interval containing the fundamental interval $[A, B]$, that is,

$$f(x) \in C^{(p)}[A, B].$$

The method is a modification of the standard approach and also employs trapezoidal rule sums and optionally the fast Fourier transform to calculate these. In many cases, the use of this modification results in a significant reduction in the number $(m+1)$ of function values required to attain comparable accuracy.

Most of the material in this paper has been presented before. However, here the author concentrates on the underlying ideas and, particularly, on relating these to

* This work was supported by the Applied Mathematical Sciences Research Program (KC–04–02) of the Office of Energy Research, U.S. Department of Energy, Contract W–31–109–Eng–38.

0021-9991/84 $3.00

practical techniques for gauging the accuracy of the numerical approximations and the cost in terms of function values of attaining a specified accuracy. To simplify the presentation, we have treated the fundamental interval $[0, 1]$, that is, we set

$$A = 0, \qquad B = 1.$$

Results for general A and B can be obtained by linear scaling.

In Section 1, various simple classical formulas associated with the calculation of Fourier coefficients using the trapezoidal rule are presented, and the use of the fast Fourier transform together with accuracy estimation is set into this context. The material in Section 1 does not require that $f(x)$ be continuous. In Section 2, the scope is reduced to sufficiently smooth functions, the Fourier coefficient asymptotic expansion is introduced, and this is used to estimate the cost of using the FFT in terms of the required accuracy. The theory is developed in Section 3, which introduces Bernouilli functions and the Euler expansion. This is sufficient to define and justify the proposed modification, which is described in detail in Section 4.

1. Definitions and Classical Background

Given a function $f(x)$, it is almost always possible to construct a convergent *Fourier series* of the form

$$\bar{f}(x) = If + 2\sum_{r=1}^{\infty} C^{(r)}f \cos 2\pi rx + 2\sum_{r=1}^{\infty} S^{(r)}f \sin 2\pi rx. \tag{1.1}$$

The coefficients in these series are termed *Fourier coefficients*, and may be defined by

$$\begin{aligned} C^{(0)}f &= If = \int_0^1 f(x)\,dx, \\ C^{(r)}f &= \int_0^1 f(x)\cos 2\pi rx\,dx, \\ S^{(r)}f &= \int_0^1 f(x)\sin 2\pi rx\,dx. \end{aligned} \tag{1.2}$$

The standard definition of Fourier coefficients coincides with $2C^{(r)}f$ and $2S^{(r)}f$ when $r \geqslant 1$ and with $C^{(0)}f$ when $r = 0$, but we use the term Fourier coefficient loosely here. The purpose of this paper is to describe a family of methods for calculating sets of Fourier coefficients numerically. That is, to obtain approximations to $C^{(r)}f$ and $S^{(r)}f$ based only on function values $f(x_i)$.

The function $\bar{f}(x)$ is clearly periodic. It can be shown that when $f \in C[0, 1]$, and satisfies a limited fluctuation condition, then

$$\begin{aligned} \bar{f}(x) &= f((x) \quad \text{for all} \quad x \in (0, 1), \\ \bar{f}(x+1) &= \bar{f}(x) \quad \text{for all} \quad x, \\ \bar{f}(1) &= \bar{f}(0) = \tfrac{1}{2}(f(0) + f(1)). \end{aligned} \tag{1.3}$$

One method for approximating integrals (1.2) is to employ the m-panel *end point trapezoidal rule*, defined by

$$R_x^{[m,1]} f(x) = \frac{1}{m}\left(\frac{1}{2} f(0) + \sum_{j=1}^{m-1} f(j/m) + \frac{1}{2} f(1)\right). \tag{1.4}$$

When this quadrature rule is used to approximate the integrals in (1.2) we obtain approximations, sometimes called *finite Fourier transforms*; we denote these by

$$\begin{aligned} a_r^{[m,1]} f &= R_x^{[m,1]}(f(x) \cos 2\pi r x), \\ b_r^{[m,1]} f &= R_x^{[m,1]}(f(x) \sin 2\pi r x). \end{aligned} \tag{1.5}$$

It is convenient to treat in detail only the case in which m is even. The theory is similar when m is odd.

The evaluation of the set of trapezoidal rule sums $a_r^{[m,1]}f$, $b_r^{[m,1]}f$, $r = 0, 1, \dots, m/2$, from the function values $f(j/m)$, $j = 0, 1, \dots, m$, can be accomplished efficiently using a *fast Fourier transform* (FFT) routine. For this reason, the method is often referred to as the FFT method. However, the reader should bear in mind that use of the FFT is not necessary for the implementation.

Having decided on this very straightforward set of formulas, one of our first theoretical tasks must be to attempt to glean information about how accurate these approximations to the true Fourier coefficients $C^{(r)}f$ and $S^{(r)}f$ are. At the very least, such information may be helpful in choosing an appropriate value of m. We may substitute into (1.5) the Fourier series $\bar{f}(x)$ for $f(x)$, and so obtain expressions for $a_r^{[m,1]}f$ and $b_r^{[m,1]}f$ which contain only Fourier coefficients. The ensuing calculation is straightforward. As a preliminary it is helpful to establish

$$\begin{aligned} R_x^{[m,1]}(\cos 2\pi r x) &= 1 \quad \text{when } r/m \text{ is an integer}, \\ &= 0 \quad \text{otherwise}, \\ R_x^{[m,1]}(\sin 2\pi r x) &= 0 \quad \text{for all } r. \end{aligned} \tag{1.6}$$

Then we find, without difficulty, that

$$\begin{aligned} a_r^{[m,1]} f &= \sum_{l=-\infty}^{\infty} C^{(lm+r)} f, \\ b_r^{[m,1]} f &= \sum_{l=-\infty}^{\infty} S^{(lm+r)} f. \end{aligned} \tag{1.7}$$

When $|r| < m$, these may be written in the form

$$\begin{aligned} a_r^{[m,1]}f &= C^{(r)}f + \sum_{l=1}^{\infty} (C^{(lm+r)}f + C^{(lm-r)}f), \\ b_r^{[m,1]}f &= S^{(r)}f + \sum_{l=1}^{\infty} (S^{(lm+r)}f - S^{(lm-r)}f). \end{aligned} \tag{1.8}$$

These formulas are sometimes known as *aliasing* formulas. Setting $r = 0$ in the first of these gives a version of the classical *Poisson summation formula*, namely,

$$R^{[m,1]}f = If + 2\sum_{l=1}^{\infty} C^{(lm)}f. \tag{1.9}$$

(The classical formula relates an infinitely extended trapezoidal sum to Fourier transforms and could be derived from this formula. This formula can be obtained from the classical formula by defining $f(x)$ to be zero outside the integration interval. An alternative derivation of the aliasing formulas can be obtained by replacing $f(x)$ by $f(x)\cos 2\pi rx$ in (1.9).)

The aliasing formulas (1.7) provide a theoretical expression for the difference between the calculated approximations, such as, $a_r^{[m,1]}f$ and the corresponding exact quantities $C^{(r)}f$. Like many theoretical expressions for the discretation error, these involve quantities (higher order Fourier coefficients) whose numerical values are unknown. Nevertheless, we can proceed to make some heuristic deductions. These are based on the circumstance that

the Fourier series is known to converge. (1.10)

This means that the magnitude of $C^{(r)}f$ approaches zero with increasing r. As a basis for our heuristic discussion, let us think in terms of the sequence of Fourier coefficients being a steadily decreasing sequence and the approximation error $|a_r^{[m,1]}f - C^{(r)}f|$ being approximated by the term in (1.8) which is of lowest order. When $r < m/2$, this term is $C^{(m-r)}f$, which we might expect to be smaller than $C^{(r)}f$ the expression whose value we are seeking. When $r > m/2$, the error term $C^{(m-r)}f$ is probably greater than $C^{(r)}f$ and so the approximation $a_r^{[m,1]}f$ is useless. So one is led quite naturally to the heuristic result that $a_r^{[m,1]}f$ is only likely to be a meaningful approximation to $C^{(r)}f$ when $r < m/2$. In addition, the same assumption implies that the approximation error deteriorates as r is increased from $r = 0$ to $r = m/2$.

When $r = m/2$, a somewhat anomalous situation occurs. Equation (1.8) reduces to

$$a_{m/2}^{[m,1]}f = 2C^{(m/2)}f + 2C^{(3m/2)}f + 2C^{(5m/2)}f + \cdots, \tag{1.11}$$

so while $a_{m/2}^{[m,1]}f$ is a ridiculous approximation to $C^{(m/2)}f$ it is, in fact, an exceptionally good approximation to $2C^{(m/2)}f$.

Heuristic arguments along the lines given lead one to the conclusion that the efficiency of the trapezoidal rule in approximating Fourier coefficients depends on how rapidly the sequence of Fourier coefficients approaches zero. Our straightforward approach, based on the m-panel trapezoidal rule, using $(m+1)$ function values $f(j/m)$, $j=0, 1,..., m$, will yield meaningful approximations to $C^{(r)}f$, $r=0, 1,..., m/2$, and $S^{(r)}f$, $r=1, 2,..., (m/2)-1$. The accuracy of each approximation will be of magnitude about $C^{(m-r)}f$ or $S^{(m-r)}f$, respectively, except for the approximation to $C^{(m/2)}f$ which is much more accurate.

It is important to differentiate in one's mind between rigorously established results (such as all the numbered equations) and heuristic but plausible consequences, such as the foregoing remarks about which approximations should be treated as meaningful and which should be disregarded. It is possible by means of contrived examples to make the discussion following statement (1.10) above seem ridiculous. It should not be possible to provide an example which contradicts any of the equations. The rest of this paper will describe an approach to the problem involving modifications which exploit these circumstances when $f(x)$ is sufficiently continuous. Before going on to this, we make some comments.

First, it follows immediately either from the definition (1.5) or from (1.7) that both $a_r^{[m,1]}f$ and $b_r^{[m,1]}f$ are periodic in r. Specifically,

$$\begin{aligned} a_r^{[m,r]}f &= a_{r+m}^{[m,1]}f = a_{-r}^{[m,1]}f, \\ b_r^{[m,1]}f &= b_{r+m}^{[m,1]}f = -b_{-r}^{[m,1]}f. \end{aligned} \tag{1.12}$$

As r becomes large then $a_r^{[m,1]}f$ does not approach zero but oscillates while $C^{(r)}f$ does approach zero. The heuristic discussion merely indicates how large to allow r to become before abandoning the approximation.

Second, we may look at the set of finite Fourier transforms, disregard those which give meaningless approximations, and form an approximation to $\bar{f}(x)$ by replacing $C^{(r)}f$ and $S^{(r)}f$ by their finite Fourier transforms. This gives a trigonometric polynomial

$$\begin{aligned} T(x) = a_0^{[m,1]}f + 2 \sum_{j=1}^{m/2-1} a_r^{[m,1]}f \cos 2\pi rx + a_{m/2}^{[m,1]}f \cos 2\pi mx/2 \\ + 2 \sum_{j=1}^{m/2-1} b_r^{[m,1]}f \sin 2\pi rx. \end{aligned} \tag{1.13}$$

A trigonometric polynomial $t(x)$ of joint degree (d_c, d_s) is one for which

$$C^{(r)}t = 0, \quad r > d_c; \qquad S^{(r)}t = 0, \quad r > d_s. \tag{1.14}$$

The following results about $T(x)$ are readily established:

(i) if $\bar{f}(x)$ is any trigonometric polynomial of joint trigonometric degree $(m/2, (m/2)-1)$ then

$$T(x) = \bar{f}(x) \qquad \text{for all } x; \tag{1.15}$$

(ii) for any function $f(x)$,

$$T(j/m) = \bar{f}(j/m) \qquad \text{all integers } j. \tag{1.16}$$

We refer to $T(x)$ defined in (1.13) as the *interpolating trigonometric polynomial* of $f(x)$ of joint degree $(m/2, (m/2)-1)$.

Consequently the finite Fourier transforms $a_r^{[m,r]}f$, $b_r^{[m,1]}f$, $r < m/2$, have two distinct definitions. First, they are the trapezoidal rule approximations to the exact Fourier coefficients $C^{(r)}f$ and $S^{(r)}f$. Second, they are the Fourier coefficients $C^{(r)}T$, $S^{(r)}T$ of a precisely specified interpolating trigonometric polynomial $T(x)$ which interpolates $\bar{f}(x)$.

This interpretation of these coefficients allows us to provide a numerical but heuristic *a posteriori* bound to the error in cases where a subroutine for $f(x)$ is available. Since $a_r^{[m,1]}f$ is the exact Fourier coefficient of $T(x)$ we have

$$a_r^{[m,1]}f = \int_0^1 T(x) \cos 2\pi rx \, dx, \qquad r < m/2, \tag{1.17}$$

and so

$$\begin{aligned} |a_r^{[m,1]}f - C^{(r)}f| &= \left| \int_0^1 (T(x) - f(x)) \cos 2\pi rx \, dx \right| \\ &\leqslant \max_{0 \leqslant x \leqslant 1} |T(x) - f(x)| = \varepsilon_{\max}. \end{aligned} \tag{1.18}$$

Now $T(x) - f(x)$ is zero at the evaluation points $x = j/m$ and is, in general, oscillatory. Using the subroutine for $f(x)$ and (1.13) for $T(x)$, one can obtain estimates of $\varepsilon_{\max}$ by evaluating $T(x) - f(x)$ at a few points between evaluation points. One might choose amongst others $x = 1/2m$ and $x = 1 - 1/2m$. In practice, $\varepsilon_{\max}$ is usually an unduly pessimistic bound on the error. However, this procedure provides an underestimate of $\varepsilon_{\max}$ and so is heuristic. In general, this method is very reliable (but not completely reliable).

As a final word of caution, the reader should note that the joint trigonometric degree of the trapezoidal quadrature rule $R^{[m,1]}$ is $(m-1, \infty)$. That is, this rule integrates exactly the functions $\cos 2\pi jx$ for all $0 \leqslant j \leqslant m-1$, and $\sin 2\pi jx$ for all j. This should be distinguished from the joint trigonometric degree of the interpolation polynomial $(m/2, (m/2)-1)$ from which the rule can be derived.

2. The Fourier Coefficient Asymptotic Expansion

The formulas given in Section 1 are valid for a very wide class of functions $f(x)$. A list of sufficient conditions is given in Whittaker and Watson [3, p. 164]. Roughly speaking, these formulas are valid for any function for which the quantities involved can be defined. At discontinuities in $f(x)$, $\bar{f}(x)$ is defined by either

$$\bar{f}(x) = \tfrac{1}{2} \lim_{\varepsilon \to 0} (f(x+\varepsilon) + f(x-\varepsilon)) \tag{2.1}$$

or

$$\bar{f}(0) = \tfrac{1}{2} \lim_{\varepsilon \to 0} (f(\varepsilon) + f(1-\varepsilon)),$$

if these exist, or $\bar{f}(x)$ diverges at points where $f(x)$ has an infinite singularity. To proceed further, we impose on $f(x)$ the restrictions mentioned in the introduction. We restrict ourselves to sufficiently smooth functions. For some nonnegative values of p,

$$f(x) \in C^{(p)}[0, 1], \tag{2.2}$$

that is, $f(x)$ and its early derivatives are continuous in the integration interval $[0, 1]$. In general, $f(x)$ is continuous in a much larger interval which contains the interval $[0, 1]$ and is usually continuous for all real x. It is important to note that this does not imply that $\bar{f}(x)$ is continuous in the larger interval. Since $\bar{f}(x+1) = \bar{f}(x)$, $\bar{f}(x)$ consists of an infinite sequence of similar segments of unit length and coincides with $f(x)$ only in the open interval $(0, 1)$. In general, this means that $\bar{f}(x)$ has a discontinuity at every integer value of x.

However, there is a class of functions for which $f(x)$ coincides with $\bar{f}(x)$ everywhere and both are analytic for all real x. These include the trigonometric functions such as $\cos 2\pi rx$ and many others. An example we shall use latter is

$$f_3(x) = (1-\rho^2)/(1 - 2\rho \cos 2\pi x + \rho^2), \qquad |\rho| < 1. \tag{2.3}$$

We shall refer to such functions $f(x)$ simply as *periodic functions*. (In fact, we should refer to them as periodic C^∞ functions of period $B - A$.)

The circumstance that $f(x)$ is sufficiently smooth allows us to make use of the Fourier coefficient asymptotic expansion (FCAE) given in (2.4). In this section we shall first use it to provide estimates of the asymptotic behavior of Fourier coefficients for large r. Then we shall draw attention to some of its drawbacks as a calculation device in its own right. Then in Section 3 we shall employ it as a basis to derive further formulas which are useful computationally.

Straightforward integration by parts of definition (1.2) gives the *Fourier coefficient asymptotic expansion*,

$$\begin{aligned} C^{(r)}f = {} & \frac{f'(1) - f'(0)}{(2\pi r)^2} - \frac{f^{(3)}(1) - f^{(3)}(0)}{(2\pi r)^4} + \cdots \\ & + (-1)^{n-2} \frac{f^{(2n-3)}(1) - f^{(2n-3)}(0)}{(2\pi r)^{2n-2}} \\ & + \frac{(-1)^{n-1}}{(2\pi r)^{2n}} \int_0^1 f^{(2n)}(t)(1 - \cos 2\pi rt)\, dt, \end{aligned} \tag{2.4a}$$

$$S^{(r)}f = -\frac{f(1)-f(0)}{2\pi r} + \frac{f^{(2)}(1)-f^{(2)}(0)}{(2\pi r)^3} + \cdots \tag{2.4b}$$
$$+ (-1)^n \frac{f^{(2n-2)}(1)-f^{(2n-2)}(0)}{(2\pi r)^{2n-1}} + \frac{(-1)^n}{(2\pi r)^{2n}} \int_0^1 f^{(2n)}(t) \sin 2\pi rt\, dt.$$

We use this first to estimate the asymptotic behavior of Fourier coefficients. We see immediately

LEMMA 2.5.

$$\textit{When} \quad f'(0) \neq f'(0), \qquad C^{(r)}f = O(r^{-2}) \quad \text{as} \quad r \to \infty,$$
$$\textit{when} \quad f(1) \neq f(0), \qquad S^{(r)}f = O(r^{-1}) \quad \text{as} \quad r \to \infty.$$

An example of a function with this behavior is

$$f_1(x) = e^{\alpha x}/(e^\alpha - 1); \qquad C^{(r)}f_1 = \frac{\alpha}{\alpha^2 + 4\pi^2 r^2}, S^{(r)}f_1 = \frac{2\pi r}{\alpha^2 + 4\pi^2 r^2}. \tag{2.5}$$

However, there are functions for which $f(1) = f(0)$ and $f'(1) = f'(0)$.

LEMMA 2.6. *When $f^{(q)}(1) = f^{(q)}(0)$, $q = 0, 1, 2,..., p-2$, then*

$$C^{(r)}f = O(r^{-p_E}), \qquad S^{(r)}f = O(r^{-p_O}) \quad \text{as} \quad r \to \infty,$$

where p_E and p_O are the smallest even and odd integers, respectively, which exceed $p-1$.

Functions of this type include any of the form $f(x) = x^{p-1}(1-x)^{p-1} g(x)$, where $g(x)$ is analytic, the Bernoulli polynomial $B_p(x)$ (see (3.1)), and a function $g_p(x)$ which we construct at the end of Section 3.

For functions for which $f^{(s)}(1) = f^{(s)}(0)$ for all s, this lemma tells us merely that the ultimate rate of convergence of the Fourier coefficients is faster than any negative integer power of r. A stronger result is

LEMMA 2.7. *When $f^{(s)}(1) - f^{(s)}(0) = 0$ for all s and $f(z)$ is analytic, there exist finite K and L so that*

$$|C^{(r)}f| < Ke^{-2\pi Lr}, \qquad |S^{(r)}f| < Ke^{-2\pi Lr}, \tag{2.7}$$

where L is any number for which $f(z)$ is analytic for all z in the infinite strip $|\mathrm{Im}\, z| \leqslant L$, and K may be taken to be

$$K = \int_0^1 |f(x + iL)|\, dx.$$

This result can be readily proved by contour integration using

$$\int_0^1 f(x)\, e^{2\pi irx}\, dx = \int_0^{iL} + \int_{iL}^{iL+1} + \int_{iL+1}^{1} f(x)\, e^{2\pi irz}\, dz, \tag{2.8}$$

and noting that since $f(x)$ is periodic, the first and third integrals on the right eliminate each other.

The function $f_3(x)$ in (2.3) illustrates this theorem. For this function

$$C^{(r)}f_3 = \rho^r, \qquad S^{(r)}f_3 = 0, \tag{2.9}$$

and $f_3(z)$ has a sequence of poles on the lines $|\mathrm{Im}\, z| = (1/2\pi) \ln \rho$ and no singularity nearer to the real axis than these.

Parenthetically, we remark that the FCAE (2.4) can be derived from (2.8) and that though this is more complicated, such a derivation provides information about the remainder term which is useful in understanding the misleading behavior described below. For details see Lyness [1, pp. 90–93].

In the discussion in Section 1 we emphasized that the efficiency of the trapezoidal rule for calculating Fourier coefficients is directly related to the rate of decay of the Fourier coefficients. In particular, when an m panel trapezoidal rule is used, the accuracy can be assessed by the magnitudes of the coefficients $C^{(r)}f$ and $S^{(r)}f$, where r is close to $m/2$. In the lemmas above we have collected together some information about these magnitudes. The brief discussion which follows indicates the computational significance of this information.

If we require six-figure accuracy for the sine FCs of function $f_1(x)$ given by (2.5) we require $m \approx 10^5$. If we only required the cosine FCs of this function to this accuracy we require $m \approx 400$. However, when $f(x)$ is periodic, for example, $f_3(x)$ with $\rho = 0.4$, we can obtain six-figure accuracy with $m \approx 40$ and twelve-figure accuracy with $m \approx 80$. Between these extremes lie functions which satisfy Lemma 2.6 for moderate values of p. Not unrealistic estimates for examples with $p = 8$ yield six-figure accuracy with $m \approx 40$ but twelve-figure accuracy with $m \approx 140$.

Now it is often the case that a user does not know to which class his function $f(x)$ belongs. However, the only information required to determine this is the value of derivative differences $f^{(s)}(1) - f^{(s)}(0)$ for $s = 0, 1, 2, \ldots$. These could be estimated numerically before the calculation using numerical differentiation based on a handful of function values. In the case of the function $f_1(x)$, two function evaluations $f(1) = e^\alpha/(e^\alpha - 1)$ and $f(0) = 1/(e^\alpha - 1)$ yield the information that it will need 10^5 function values to obtain six-figure accuracy. In other cases, a marginally more sophisticated calculation is necessary.

The technique described in Section 4 employs this approach. There, by a process resembling subtracting out a singularity, the onus of the calculation is shifted from a function like $f_1(x)$ to one which satisfies the hypotheses of Lemma 2.6 for a moderate value of p.

At first sight one may feel tempted to use the Fourier coefficient asymptotic expansion to calculate $C^{(r)}f$ by evaluating some of the derivatives of $f(x)$ at $x=0$ and at $x=1$ and truncating the expansion. For moderate or large values of r the error resulting from inaccurate numerical differentiation is not critical since successively higher derivatives $f^{(2s-1)}(1)-f^{(2s-1)}(0)$ appear with successively smaller cofactors $(2\pi r)^{-2s}$. However, there is a major objection to the uncritical use of this expansion. This is that it usually diverges, and when it converges it may converge to the wrong result. To see this we note that when $f_3(x)$ is an analytic periodic function with period 1, the individual terms in the expansion for $C^{(r)}(f)$ and for $C^{(r)}(f+f_3)$ coincide but the remainder terms are different. But even more misleading is the circumstance that it may have the property that for moderate r the magnitudes of the individual terms decrease at first and then increase. However, truncating the series at its smallest term may produce a completely erroneous result

TABLE I[a]

Approximations to $C^{(r)}f$ with $r=6$ and $f(x)=(x^2-x+0.26)^{-1}$ Obtained by Truncating the FCAE (2.4a).

q	T_{2q} [b]	Σ_{2q} [c]
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

[a] This table is taken from J. N. Lyness, *Math. Comp.* **24** (1970), p. 110.

[b] The second column contains the qth term, namely, $T_{2q}=(f^{(2q-1)}(1)-f^{(2q-1)}(0))/(2\pi\cdot 6)^{2q}$.

[c] The third contains the truncated sum $\Sigma_{2q}=\sum_{r=1}^{q} T_{2q}$.

(wrong both in magnitude and sign). This effect is illustrated in Table I. One might well conclude from this table that, when $f(x) = 1/(x^2 - x + 0.26)$, the value of $C^{(6)}f$ is -2.033×10^{-2}. In fact, it is $+7.01 \times 10^{-1}$. Nevertheless, when r is sufficiently large, this series can be used. However, one has to have some prescription for how large this value of r must be and the number of terms to retain based on some other calculation. It will turn out that the technique described in Section 4 provides such a prescription.

3. Bernoulli Polynomials and the Euler Expansion

To proceed, it will be convenient to introduce the Bernoulli polynomials and functions. The standard definition by means of the generating function is somewhat cumbersome for the present purposes. A more straightforward approach is to define them by means of their Fourier series.

The *Bernoulli functions* $\bar{B}_q(x)$ $q = 0, 1, 2, \ldots$, are defined by:

$$\begin{aligned} \bar{B}_0(x) &= 1, \\ \frac{\bar{B}_{2q+1}(x)}{(2q+1)!} &= 2(-1)^{q+1} \sum_{r=1}^{\infty} \frac{\sin 2\pi r x}{(2\pi r)^{2q+1}}, \qquad q \geqslant 0, \\ \frac{\bar{B}_{2q}(x)}{(2q)!} &= 2(-1)^{q-1} \sum_{r=1}^{\infty} \frac{\cos 2\pi r x}{(2\pi r)^{2q}}, \qquad q \geqslant 1. \end{aligned} \tag{3.1}$$

As a preliminary we identify $\bar{B}_1(x)$,

$$\bar{B}_1(x) = x - \frac{1}{2} = -2 \sum_{r=1} \frac{\sin 2\pi r x}{2\pi r}, \qquad 0 < x < 1. \tag{3.2}$$

This is established simply by calculating the Fourier coefficients of $x - \frac{1}{2}$ and putting these together as a Fourier series. From these definitions it follows that

$$\frac{d}{dx} \frac{\bar{B}_q(x)}{q!} = \frac{\bar{B}_{q-1}(x)}{(q-1)!}, \qquad 0 < x < 1,\ q \geqslant 1. \tag{3.3}$$

(For $q > 1$, this follows from (3.1); for $q = 1$ it follows from (3.2).) Moreover, by integrating term by term in (3.1) we find

$$\begin{aligned} \int_0^1 \frac{\bar{B}_q(x)}{q!}\, dx &= 1, \qquad q = 0, \\ &= 0, \qquad q \geqslant 1. \end{aligned} \tag{3.4}$$

From (3.3) we can obtain a recursive definition, i.e.,

$$\frac{\bar{B}_q(x)}{q!} = \int_0^x \frac{\bar{B}_{q-1}(t)}{(q-1)!}\,dt + K_q, \tag{3.5}$$

where K_q, the constant of integration, is determined by (3.4).

Now $\bar{B}_1(x)$ coincides with a monic polynomial

$$B_1(x) = x - \tfrac{1}{2},$$

in the interval $0 < x < 1$. Since $\bar{B}_q(x)/q!$ may be obtained by integrating $\bar{B}_1(x)$ $(q-1)$ times, it follows that $\bar{B}_q(x)$ coincides with a monic polynomial of degree q in this interval. These polynomials are the *Bernoulli polynomials* denoted by $B_q(x)$. Finally, the ***Bernoulli numbers*** B_q are simply $B_q(0)$. Extensive tables of Bernoulli polynomials exist.

The only properties of the Bernouilli polynomials required subsequently in this paper are (3.3) and the result

$$\begin{aligned} \frac{B_q(1)}{q!} - \frac{B_q(0)}{q!} &= 0, \qquad q = 0, 2, 3, 4, \ldots, \\ &= 1, \qquad q = 1, \end{aligned} \tag{3.6}$$

which follows from (3.4).

We now construct the Euler expansion. We take the Fourier series (1.1) and substitute for $C^{(r)}f$ and $S^{(r)}f$ the Fourier coefficient asymptotic expansion (with remainder term) (2.4). When we do this, there appear sums of the form $\sum (\cos 2\pi r x/(2\pi r)^{2q})$ which may be expressed in terms of Bernoulli functions (3.1). After some elementary manipulation we find

$$f(x) = If + \sum_{q=1}^{p-1} (f^{(q-1)}(1) - f^{(q-1)}(0)) \frac{B_q(x)}{q!} + \int_0^1 f^{(p)}(t) \frac{\bar{B}_p(x) - \bar{B}_p(x-t)}{p!}\,dt. \tag{3.7}$$

This is known as the *Euler expansion.* (Incidentally, if we apply the trapezoidal rule operator $R^{[m,1]}$ (given in (1.4)) to this, we uncover the better known *Euler–Maclaurin expansion.*) These expansions share the undesirable convergence properties of the Fourier coefficient asymptotic expansion. We may write the Euler expansion in the form

$$f(x) = h_{p-1}(x) + g_p(x), \tag{3.8}$$

where

$$h_{p-1}(x) = \sum_{q=1}^{p-1} (f^{(q-1)}(1) - f^{(q-1)}(0)) \frac{B_q(x)}{q!} \tag{3.9}$$

is a polynomial of degree $(p-1)$ in x.

We now come to the purpose of constructing the Euler expansion. Using (3.3) and (3.6) above we readily show that the derivative differences of $h_{p-1}(x)$ satisfy

$$h_{p-1}^{(q)}(1) - h_{p-1}^{(q)}(0) = f^{(q)}(1) - f^{(q)}(0), \qquad q = 0, 1, \ldots, p-2,$$

and it then follows immediately from (3.8) that

$$g_p^{(q)}(1) - g_r^{(q)}(0) = 0, \qquad q = 0, 1, \ldots, p-2. \tag{3.10}$$

This result recalls the discussion in Section 2. The function $g_p(x)$ satisfies the hypotheses of Lemma 2.6 and so has the property that its Fourier coefficients decay reasonably rapidly. In accordance with this lemma we have

$$C^{(r)}g_p = O(r^{-p_E}), \qquad S^{(r)}g_p = O(r^{-p_O}),$$

where p_E and p_O are the smallest even and odd integers exceeding $p-1$. The function $g_p(x)$ may be obtained from $f(x)$ simply by subtracting a known polynomial $h_{p-1}(x)$. And it can be obtained from any function $f(x)$, however slowly the Fourier coefficients of $f(x)$ may converge.

Leaving aside for the moment some minor but nontrivial difficulties, the thrust of our approach is this. Take some moderate value of p, say $p = 8$ or 9. Calculate the parameters of the function $h_{p-1}(x)$ in (3.9). Express $C^{(r)}f$ in the form

$$C^{(r)}f = C^{(r)}h_{p-1} + C^{(r)}g_p. \tag{3.11}$$

Use the exact result for $C^{(r)}h_{p-1}$. And use the trapezoidal rule (the FFT) to evaluate the Fourier coefficients of $g_p(x)$. And because of (3.10), we are assured this is a moderate calculation and is not like the first example $f_1(x)$ of Section 2, an excessive one.

4. Approximations for Fourier Coefficients

The proposed method then is based on choosing values of p and of m, setting $f(x) = h_{p-1}(x) + g_p(x)$ and using the m-panel trapezoidal rule to approximate $C^{(r)}g_p$ by $a_r^{[m,1]}g_p$. As mentioned at the end of Section 1, these approximations to $C^{(r)}g_p$ are, in fact, exact Fourier coefficients $C^{(r)}G_p$, where $G_p(x)$ is the trigonometric interpolant of joint degree $(m/2, (m/2)-1)$. So, in effect, we are approximating $C^{(r)}f$ by $C^{(r)}F$, where

$$F(x) = h_{p-1}(x) + G_p(x) \tag{4.1}$$

and $G_p(x)$ is the trigonometric interpolant of $g_p(x)$; $F(x)$ is an approximation to $f(x)$ of the form

$$F(x) = \sum_{q=1}^{p-1} \lambda_q \frac{B_q(x)}{q!} + 2 \sum_{r=0}^{m/2}{}'' \mu_r \cos 2\pi rx + \nu_r \sin 2\pi rx, \tag{4.2}$$

with

$$\lambda_q = f^{(q-1)}(1) - f^{(q-1)}(0) \tag{4.3}$$

and

$$\begin{aligned} \mu_r &= R^{[m,1]}(G_p(x) \cos 2\pi rx) \equiv a_r^{[m,1]} g_p, & r &\leqslant m/2, \\ \nu_r &= R^{[m,1]}(G_p(x) \sin 2\pi rx) \equiv b_r^{[m,1]} g_p, & r &< m/2, \\ \mu_r &= \nu_r = 0, & r &> m/2. \end{aligned} \tag{4.4}$$

Note that this definition implies

$$\nu_0 = \nu_{m/2} = 0.$$

We note also that, while λ_q depends only on the function $f(x)$, the values of μ_r and ν_r depend on p, m, and the values of λ_q as well.

The Fourier coefficients of $F(x)$ are readily evaluted using the known Fourier coefficients of the Bernoulli polynomials given in (3.1). In terms of quantities defined in (4.3) and (4.4) they are given by

$$\begin{aligned} C^{(0)}F &= \mu_0, \\ C^{(r)}F &= \sum_{q=2}^{p-1^{\mathrm{E}}} (-1)^{q/2+1} \lambda_q/(2\pi r)^q + \mu_r, & r &= 1, 2, \ldots, \frac{m}{2} - 1, \\ C^{(m/2)}F &= \sum_{q=2}^{p-1^{\mathrm{E}}} (-1)^{q/2+1} \lambda_q/(2\pi r)^q + \tfrac{1}{2}\mu_{m/2}, \\ C^{(r)}F &= \sum_{q=2}^{p-1^{\mathrm{E}}} (-1)^{q/2+1} \lambda_q/(2\pi r)^q, & r &> \frac{m}{2} - 1, \\ S^{(r)}F &= \sum_{q=1}^{p-1^{\mathrm{O}}} (-1)^{(q+1)/2} \lambda_q/(2\pi r)^q + \nu_r, & r &= 1, 2, \ldots, \frac{m}{2} - 1, \\ S^{(r)}F &= \sum_{q=1}^{p-1^{\mathrm{O}}} (-1)^{(q+1)/2} \lambda_q/(2\pi r)^q, & r &\geqslant \frac{m}{2}, \end{aligned} \tag{4.5}$$

where the superscripts E (or O) attached to the summation symbol indicate that the sum is restricted to even (or odd) values of the summation index.

The reader will note that for $r > m/2$, the expressions for $C^{(r)}F$ and $S^{(r)}F$ coincide with the first $p/2$ terms of the Fourier coefficient asymptotic expansions of $f(x)$ given in (2.4). Thus the method effectively provides a prescription for determining when it is "safe" to use this expansion.

For all r, the approximation error is given by

$$\begin{aligned} C^{(r)}F - C^{(r)}f &= (C^{(r)}h_{p-1} + C^{(r)}G_p) - (C^{(r)}h_{p-1} + C^{(r)}g_p) \\ &= C^{(r)}G_p - C^{(r)}g_p, \end{aligned} \tag{4.6}$$

and so depends only on the accuracy of the approximation of g_p by its trigonometric interpolation polynomial. In Section 1 this accuracy was discussed in some detail. The discussion following (1.10) and the formulas (1.17) and (1.18) are equally valid in this context, so long as one replaces $f(x)$ and $T(x)$ by $g_p(x)$ and $G_p(x)$, respectively. Consequently, the numerical accuracy of (4.5), when used as approximations for $C^{(r)}f$ and $S^{(r)}f$ can be estimated either by spot checks of $g_p(x) - G_p(x)$ or by noting the magnitude of the final few calculated approximations to the Fourier coefficients of $g_p(x)$, that is, $\mu_{m/2}$, $\mu_{(m/2)-1}$, $\nu_{(m/2)-1}$. Denoting this accuracy by ε for the moment, it is important to note that $C^{(r)}F$ and $S^{(r)}F$ are approximations of this accuracy to $C^{(r)}f$ and $S^{(r)}f$, respectively, for all values of r, and not just for $r < m/2$. There is no abrupt break in the quality of the approximation at $r = m/2$. For example, when $f(x)$ is a function whose Fourier coefficients decay slowly, one finds for small r, that terms like μ_r may be large and even predominate the term involving the summation over q. As r is increased, the magnitude of μ_r rapidly decays, reaching the tolerance ε at $r = m/2$ and is not required for $\tau > m/2$. On the other hand, the term involving the summation decays slowly with increasing r, is predominant at $r = m/2$ and continues to provide a meaningful approximation to $C^{(r)}f$ for higher values of r until r is so large that $C^{(r)}F$ itself is of magnitude only ε.

While the results can be interpreted as adjustments to the FCAE, the interpretation along the lines of subtracting out the polynomial $h_{p-1}(x)$ is helpful. This is akin to subtracting out the singularity in numerical integration. In fact it is precisely equivalent to subtracting out singularities in $\bar{f}(x)$. If we denote by $\bar{h}_{p-1}(x)$ and $\bar{g}_p(x)$ the Fourier series of $h_{p-1}(x)$ and $g_p(x)$, respectively, it follows that

$$\bar{f}(x) = \bar{h}_{p-1}(x) + \bar{g}_p(x). \tag{4.7}$$

In view of (3.10), $\bar{g}_p(x)$ is a function whose derivatives of order $(p-2)$ or less are continuous for all x, and so the singularities in $\bar{f}(x)$ have been mitigated by subtracting out the known function.

In practice, numerical differentiation is required to provide approximations $\tilde{\lambda}_q$ to λ_q in (4.3) and it would be only prudent for a user to seek reassurance that this will not introduce gross inaccuracies into the results. A detailed analysis of the effect of inaccurate derivatives is given in Lyness [2], where numerical estimates are derived. In fact, what happens is that the values of μ_r and ν_r depend on the values of λ_q. One finds in a perturbation analysis of formulas like (4.5), that successively higher derivatives occur with successively smaller coefficients. Under the assumption that five binary digits are lost in each numerical differentiation, the error ε which would have been attained using exact derivatives is compromised to the extent of attaining an error 3ε.

A moment's reflection will convince the reader that one could use any subtraction function $\tilde{h}(x)$, set

$$f(x) = \tilde{h}(x) - \tilde{g}(x), \tag{4.8}$$

and calculate the Fourier coefficients of $f(x)$ using

$$C^{(r)}f = C^{(r)}\tilde{h} + C^{(r)}\tilde{g}, \tag{4.9}$$

with $C^{(r)}\tilde{h}$ being calculated exactly and $C^{(r)}g$ being calculated using the trapezoidal rule. The particular choice of $\tilde{h}(x)$ as $h_{p-1}(x)$ defined in (3.9) is merely the choice which leads to the most rapid calculation for $C^{(r)}\tilde{g}$. If one cannot evaluate $h_{p-1}(x)$ exactly but uses a function as close to $h_{p-1}(x)$ as is feasible, $\tilde{g}(x)$ will be close to $g_p(x)$ and the worst that is likely to happen is that the numerical evaluation of $C^{(r)}\tilde{g}$ will be marginally more expensive than the corresponding evaluation of $C^{(r)}g_p$ would have been.

5. Concluding Remarks

The method described in this paper was first described in Lyness [2]. In that long paper we dealt at length with mathematical details. For example, the convergence properties of the Euler–Maclaurin asymptotic expansion was discussed in detail; the theory was presented more generally using offset trapezoidal rules; and a thorough analysis of the effect of using numerical derivatives was presented. Subsequently, in a technical memorandum (Lyness [4]), we describe generalizations of the method; this time we address a reader who may want to program a calculation; the technique is generalized to include the case in which function values are available only at irregularly spaced abscissas which may or may not lie within the fundamental interval; corresponding formulas for approximating sets of trigonometrical integrals of the form

$$\int_A^B f(x)\, e^{\gamma_j x}\, dx, \qquad j = 0, 1,...,$$

where γ_j may be real or complex are also given.

In the present paper the author has attempted to present the same method in a way in which the underlying ideas stand out and a scientist who is used to calculating Fourier coefficients can readily relate these modifications to the familiar theory.

REFERENCES

1. J. N. LYNESS, *Math. Comp.* **25** (1971), 87–104.
2. J. N. LYNESS, *Math. Comp.* **28** (1974), 81–123.
3. E. T. WHITTAKER AND G. N. WATSON, "A Course of Modern Analysis," 1961.
4. J. N. LYNESS, Technical Memorandum 370, Applied Mathematics Division, Argonne National Laboratory, Argonne, Ill., 1981.