Approximating sums of infinite series

Kara Garrison* and Thomas E. Price
Department of Mathematics
The University of Akron
Akron, OH 44325

May 25, 2009

ABSTRACT

The Euler-Maclaurin summation formula is frequently used to efficiently estimate sums of infinite series of the form $\sum_{j=1}^{\infty} f(j)$. The purpose of this article is to describe a modification of this numerical technique designed to simplify and reduce the computational effort required to obtain an acceptable estimate of the sum. The modified formula is obtained by replacing $f(x)$ with an easily constructed polynomial like interpolating function $a(x)$ designed to simplify the calculation of the integral and derivatives associated with Euler-Maclaurin. This approach provides a more tractable algorithm which can be written as a matrix equation. Examples are provided to demonstrate that the accuracy of the new algorithm compares favorably with that of the traditional formula. The paper concludes with a brief discussion of a method for approximating the error incurred when replacing the exact value of the sum of the original series with the estimate.

1 INTRODUCTION

Let $f : [1, \infty) \longrightarrow \mathbb{R}$ and $k \geq 1$. Under suitable conditions, the Euler-Maclaurin summation formula

$$\sum_{j=1}^{\infty} f(j) \approx \sum_{j=1}^{k-1} f(j) + \int_{k}^{\infty} f(x)dx + \frac{1}{2}f(k) - \frac{1}{12}f'(k) + \frac{1}{720}f'''(k) - \cdots \tag{1}$$

is used to approximate sums of infinite series ([Stoer and Bulirsch, 1995]). The coefficient of $f^{(j-1)}(k)$ in the above expression is given by $(-1)^{j-1} B_j/j!, j = 1, 2, \ldots$, where B_j denotes

*Part of the research presented in this paper was completed while this author was an undergraduate student at the University of Akron.

the j-th Bernoulli number. Recall that $B_{2j+1} = 0$ for $j = 1, 2, \ldots$. Formula (1) was evidently independently discovered by Leonhard Euler and Colin Maclaurin ([Goldstein, 1977]). In practice, we let

$$E_{k,d}(f) = \int_k^\infty f(x)dx + \frac{1}{2}f(k) - \frac{1}{12}f'(k) + \cdots + \frac{B_{d+1}}{(d+1)!}f^{(d)}(k) \tag{2}$$

so that

$$\sum_{j=1}^{\infty} f(j) \approx \sum_{j=1}^{k-1} f(j) + E_{k,d}(f) \tag{EM}$$

serves as an estimate for the sum of the series.

Example 1 *Using $k = 10$ and no derivative terms in the EM formula gives the estimate*

$$\sum_{j=1}^{\infty} j^{-2} \approx \sum_{j=1}^{10} j^{-2} + E_{11,0}(x^{-2}) = 1.6448$$

which compares favorably with the true answer of $\pi^2/6 \approx 1.6449$.

It would require summing almost 10,000 terms of $\sum_{j=1}^{\infty} j^{-2}$ to obtain the same accuracy. The estimate using $E_{11,1}(x^{-2})$, which includes the first derivative term, is correct to six decimal places. Such improvements are common thus encouraging the inclusion of derivative information when computationally feasible.

The formula may produce incorrect results if used without care. For example, using $E_{1,3}(x^{-2})$ in EM yields the unacceptable estimate 1.633333. Including higher derivatives terms eventually worsens the accuracy. In fact, the EM formula *always* produces the **divergent** series

$$\frac{1}{2}f(k) - \frac{1}{12}f'(k) + \frac{1}{720}f'''(k) - \cdots$$

when $f(x) = x^{-2}$. This difficulty can not be avoided but its effect can frequently be numerically eliminated by first summing a few terms of the series $\sum_{j=1}^{\infty} f(j)$ before employing EM. You are encouraged to examine the material in [Ralston and Rabinowitz, 1978] and [Stoer and Bulirsch, 1995] to learn more about these issues related to the *art* of numerical mathematics.

There are alternatives to using formula (1). For example, Longman ([Longman, 1987]) in an effort to accelerate the rate of convergence of $\sum_{j=1}^{\infty} f(j)$, derived a new series which normally converges much faster than the original. His algorithm enjoys the property that no integrals or derivatives of $f(x)$ are needed but it does require an approximation to $f(x)$ of the form

$$f(x) \approx \sum_{\nu=1}^{N} \frac{b_j}{x^{\nu+\alpha}}, \qquad \alpha > 1. \tag{3}$$

Longman correctly observed that his technique is not a reformulation of formula (1) but it can be shown that it is equivalent to applying this formula to the sum in (3).

The calculation of the integral and derivatives can be prominent obstacles when using the EM formula. Our goal is to develop an alternate algorithm that mimics the impressive numerical advantages of this estimate while lowering these hurdles. Essentially, we approximate $f(x)$ with a *polynomial like* function $a(x)$ and then replace $E_{k,d}(f)$ in (EM) with $E_{k,d}(a)$ to produce the *Modified Euler-Maclaurin* formula

$$\sum_{j=1}^{\infty} f(j) \approx \sum_{j=1}^{k-1} f(j) + E_{k,d}(a). \tag{MEM}$$

The advantage of this strategy is that the function a is easily derived and requires only elementary calculus to determine its integral and derivatives.

2 THE ALGORITHM

The function $a(x)$ must approximate $f(x)$ on an interval of the form $[k, \infty)$ where $k \in \mathbb{N}$. Our first step in defining $a(x)$ is to set

$$g(x) = c^{-1} x^{-\beta} f(1/x), \quad x \in (0, 1/k]. \tag{4}$$

The constant $c \in \mathbb{R}$ and exponent $\beta \in \mathbb{R}, \beta > 1$, are chosen so that $g(0) = \lim_{x \to 0^+} g(x) = 1$. We assume that g is analytic at $z = 0$ in the complex variable sense. Fix $k \geq 2$. The approximating function $a(x)$ is constructed by replacing $g(x)$ in formula (4) with its polynomial interpolant $p_\mu(x)$ of degree $\mu = k - n$ in the nodes $0, 1/n, 1/(n+1), \ldots, 1/(k-1)$ where $1 \leq n \leq k - 1$. Since formula (4) is equivalent to

$$f(x) = c x^{-\beta} g(1/x), \quad x \in [k, \infty) \tag{5}$$

and $p_\mu(1/x) \approx g(1/x)$ for $x \in [k, \infty)$, the function

$$a(x) = c x^{-\beta} p_\mu(1/x) \tag{6}$$

serves as an approximation to $f(x)$. Determining $p_\mu(x)$ requires no additional functional evaluations since $f(n), \ldots, f(k-1)$ are needed in the sum $\sum_{j=1}^{k-1} f(j)$. Since $p_\mu(x)$ interpolates to $g(x)$ at $x = 0$ this strategy exploits the behavior of f at $z = \infty$. The function $a(x)$ is *polynomial like* (β may not be an integer) in $1/x$ so its integral and derivatives are reasonably simple to calculate. The following example illustrates the MEM formula.

Example 2 *Let* $f(x) = \dfrac{1}{2x^3 + x^2 + 1}$. *Then*

$$f(1/x) = \frac{1}{\frac{2}{x^3} + \frac{1}{x^2} + 1} = \frac{1}{2} x^3 \left(\frac{1}{1 + \frac{x}{2} + \frac{x^3}{2}} \right)$$

so $c = \frac{1}{2}$, $\beta = 3$, *and* $g(x) = (1 + x/2 + x^3/2)^{-1}$. *Using the nodes* $0, 1/8, 1/9,$ *and* $1/10$, *(*$n = 8$ *and* $k = 11$*) produces the degree three polynomial interpolant*

$$p_3(x) = 1 - \frac{7270\,699}{14\,559\,930} x + \frac{1692\,191}{7279\,965} x^2 - \frac{74\,488}{161\,777} x^3$$

to $g(x)$. The coefficients of p_3 are solutions to the system

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1/8 & 1/64 & 1/512 \\ 0 & 1/9 & 1/81 & 1/729 \\ 0 & 1/10 & 1/100 & 1/1000 \end{bmatrix} \begin{bmatrix} 1 \\ a_1 \\ a_2 \\ a_3 \end{bmatrix} = \begin{bmatrix} 1 \\ g(1/8) - 1 \\ g(1/9) - 1 \\ g(1/10) - 1 \end{bmatrix}. \tag{7}$$

Setting $a(x) = \frac{1}{2}x^{-3}p_3(1/x)$ we obtain the MEM estimate

$$\sum_{j=1}^{\infty} f(j) \approx \sum_{j=1}^{10} f(j) + E_{11,3}(a) = 0.331\,491\,171.$$

The estimate in Example 2 compares favorably with the value 0.331 491 163 produced by (EM) using $E_{11,3}(f)$. The fourth degree interpolant in the nodes $0, 1/7, 1/8, 1/9$, and $1/10$ yields about the same accuracy obtained by applying (EM). The true answer to nine decimal places is $\sum_{j=1}^{\infty} f(j) = 0.331\,491\,164$.

It is easy to reduce the 4×4 system in (7) to an equivalent 3×3 system. The motivation for using the larger dimension will become apparent in the sequel. The value $k = 11$ was chosen in Example 2 to be large enough to reduce the error but small enough to limit the number of function evaluations of f. Two derivative terms were used in the calculations above and including a third has virtually no effect on either of the EM or MEM formulas.

3 COMPUTATIONAL MATTERS

At first glance it may seem reasonable to compute p_μ using, say, the Newton form of the interpolating polynomial since this formulation uses the results of previous calculations when additional nodes are added. (See, for example, [Mathews and Fink, 1999].) However, as Example 2 illustrates, representing the interpolant as a Taylor section $p_\mu(x) = 1 + \sum_{j=1}^{\mu} a_j x^j$ has the desirable advantage of further simplifying the calculation of integrals and derivatives. In addition, as we shall see later in this section, we will explicitly solve the matrix equation needed to compute the coefficients of p_μ.

3.1 A linear system

In general, the coefficients for p_μ are given by the linear equation

$$\begin{bmatrix} 1 \\ a_1 \\ a_2 \\ \vdots \\ a_\mu \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & \frac{1}{n} & \frac{1}{n^2} & \cdots & \frac{1}{n^{k-n}} \\ 0 & \frac{1}{(n+1)} & \frac{1}{(n+1)^2} & \cdots & \frac{1}{(n+1)^{k-n}} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \frac{1}{k-1} & \frac{1}{(k-1)^2} & \cdots & \frac{1}{(k-1)^{k-n}} \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ g(\frac{1}{n}) - 1 \\ g(\frac{1}{n+1}) - 1 \\ \vdots \\ g(\frac{1}{k-1}) - 1 \end{bmatrix}. \tag{8}$$

Let $\mathbf{0}_{m\times n}$ represent the $m \times n$ zero matrix, set $M_{n,k} = \left[(n+i-1)^{-j}\right]_{i,j}, i,j = 1,\ldots,\mu$, and write (8) more succinctly as the block matrix equation

$$\mathbf{a}_{n,k} = \begin{bmatrix} 1 & \mathbf{0}_{1\times\mu} \\ \mathbf{0}_{\mu\times 1} & M_{n,k} \end{bmatrix}^{-1} \mathbf{g}_{n,k} = \begin{bmatrix} 1 & \mathbf{0}_{1\times\mu} \\ \mathbf{0}_{\mu\times 1} & M_{n,k}^{-1} \end{bmatrix} \mathbf{g}_{n,k}. \tag{9}$$

Setting $\mathbf{x}_j^T(x) = \begin{bmatrix} 1 & x^{-1} & \cdots & x^{-j} \end{bmatrix}, j = 0,1,\ldots$, we have

$$p_\mu(x) = \mathbf{x}_\mu^T(x) \begin{bmatrix} 1 & \mathbf{0}_{1\times\mu} \\ \mathbf{0}_{\mu\times 1} & M_{n,k}^{-1} \end{bmatrix} \mathbf{g}_{n,k}. \tag{10}$$

3.2 Computing $M_{n,k}^{-1}$

To determine $M_{n,k}^{-1}$ we will need the *fundamental Lagrange polynomials* ([?]) in the nodes $\{1-\ell\}_{\ell=1}^{\mu}$ given by

$$q_j(x) = \frac{(-1)^{-j-1}}{(j-1)!\,(\mu-j)!} \prod_{\substack{\ell=1 \\ \ell\neq j}}^{\mu} (x+\ell-1), j = 1,\ldots,\mu.$$

For a fixed $j, 1 \le j \le \mu$, the polynomial $q_j(x)$ is of degree μ, has roots $\{\ell : 1 \le \ell \le \mu, \ell \neq j\}$, and satisfies $q_j(1-j) = 1$.

Proposition 3 *Let $B_{n,k}$ be the matrix with i,j-th entry*

$$(B_{n,k})_{i,j} = \frac{(-1)^i (n+j-1)^\mu}{(\mu-i)!} q_j^{(\mu-i)}(n), \qquad i,j = 1,\ldots,\mu.$$

Then $B_{n,k} = M_{n,k}^{-1}$.

Proof. We prove this proposition by showing that $M_{n,k}B_{n,k} = I_{\mu\times\mu}$ where $I_{\mu\times\mu}$ denotes the $\mu \times \mu$ identity matrix. The i,j-th entry of this product is

$$(M_{n,k}B_{n,k})_{i,j} = \sum_{s=1}^{\mu} \frac{(-1)^s (n+j-1)^\mu}{(\mu-s)!\,(n+i-1)^s} q_j^{(\mu-s)}(n).$$

Reindexing the above sum, factoring, and using the properties of powers of (-1) yield

$$(M_{n,k}B_{n,k})_{i,j} = \frac{(n+j-1)^\mu}{(n+i-1)^\mu} \sum_{s=0}^{\mu-1} \frac{((1-i)-n)^s}{s!} q_j^{(s)}(n).$$

The sum in the above expression is the Taylor representation for $q_j(x)$ centered at n and evaluated at $(1-i)$. Consequently,

$$(M_{n,k}B_{n,k})_{i,j} = \frac{(n+j-1)^\mu}{(n+i-1)^\mu} q_j(1-i).$$

We have already observed that $q_j(1-i)=0$ if $i \neq j$ and $q_i(1-i)=1$ so $M_{n,k}B_{n,k}=I_{\mu\times\mu}$. Hence, $B_{n,k}=M_{n,k}^{-1}$. ■

As an example we have

$$B_{8,11}=\begin{bmatrix} 256 & -729 & 500 \\ -4864 & 13\,122 & -8500 \\ 23\,040 & -58\,320 & 36\,000 \end{bmatrix}. \tag{11}$$

3.3 The MEM as a matrix equation

It follows from (6), (10), and Proposition 3 that $a(x)$ can be computed using the matrix formula

$$a(x)=cx^{-\beta}\mathbf{x}_\mu^T(x)\begin{bmatrix} 1 & \mathbf{0}_{1\times\mu} \\ \mathbf{0}_{\mu\times 1} & B_{n,k} \end{bmatrix}\mathbf{g}_{n,k}.$$

Substituting this into (MEM) gives

$$\sum_{j=1}^{\infty} f(j)\approx\sum_{j=1}^{k-1} f(j)+c\mathbf{E}_{k,d}^T\left(x^{-\beta}\mathbf{x}_\mu(x)\right)\begin{bmatrix} 1 & \mathbf{0}^T \\ \mathbf{0} & B_{n,k} \end{bmatrix}\mathbf{g}_{n,k} \tag{12}$$

where $\mathbf{E}_{k,d}^T\left(x^{-\beta}\mathbf{x}_\mu(x)\right)$ denotes the vector obtained by applying (MEM) to each component of $x^{-\beta}\mathbf{x}_\mu(x)$. For example,

$$\mathbf{E}_{11,3}^T\left(x^{-2}\mathbf{x}_3(x)\right)=\begin{bmatrix} \frac{459\,799}{4831\,530} & \frac{48\,097}{10\,629\,366} & \frac{5586}{19\,487\,171} & \frac{35\,007}{1714\,871\,048} \end{bmatrix}. \tag{13}$$

Notice that $\mathbf{E}_{k,d}^T\left(x^{-\beta}\mathbf{x}_\mu(x)\right)\begin{bmatrix} 1 & \mathbf{0}^T \\ \mathbf{0} & B_{n,k} \end{bmatrix}$ does not depend on f (although it does depend on β) so this vector can be saved for future use.

Example 4 *Let $f(x)=\left(\sin x^{-1}\right)/x$. Then*

$$f(1/x)=x\sin x=x^2\left(\frac{\sin x}{x}\right)$$

so $c=1$, $\beta=2$, and

$$g(x)=\begin{cases} \frac{\sin x}{x} & \text{if } x\neq 0 \\ 1 & \text{if } x=0 \end{cases}.$$

Using the nodes 0,1/8,1/9, and 1/10 with $d=3$ and formula (13) produces

$$\begin{aligned}\sum_{j=1}^{\infty} f(j) &\approx \sum_{j=1}^{10} f(j)+\mathbf{E}_{11,3}^T\left(x^{-2}\mathbf{x}_3(x)\right)\begin{bmatrix} 1 & \mathbf{0}_{1\times 3} \\ \mathbf{0}_{3\times 1} & B_{8,11} \end{bmatrix}\mathbf{g}_{9,11} \\ &= 1.472\,828\,238.\end{aligned}$$

This result compares favorably with the EM estimate using $E_{11,3}(f)$ as well as with the true value $\sum_{j=1}^{\infty} f(j)=1.472\,828\,232$.

4 Using $g'(0)$

Formula (12) can be improved if the first derivative of the function g at 0 is known. Here, we are interested in determining the polynomial interpolant

$$\widehat{p}_{\mu+1}(x) = 1 + g'(0)\,x + \sum_{j=1}^{\mu} a_{j+1}x^{j+1}$$

that agrees with g at the nodes $0, \frac{1}{n}, \cdots, \frac{1}{k-1}$ and with g' at 0. Define the new approximating function

$$\widehat{a}(x) = cx^{-\beta}\widehat{p}_{\mu+1}(1/x) = cx^{-\beta}\mathbf{x}_{\mu+1}^{T}(x)\begin{bmatrix} I_{2\times 2} & \mathbf{0}_{2\times\mu} \\ \mathbf{0}_{\mu\times 2} & B_{n,k} \end{bmatrix} D_{n,k}\widehat{\mathbf{g}}_{n,k}$$

where $D_{n,k} = \operatorname{diag}(1, 1, n, \cdots, k-1)$ and

$$\widehat{\mathbf{g}}_{n,k} = \begin{bmatrix} 1 \\ g'(0) \\ g\left(\frac{1}{n}\right) - 1 - g'(0)\frac{1}{n} \\ \vdots \\ g\left(\frac{1}{k-1}\right) - 1 - g'(0)\frac{1}{k-1} \end{bmatrix}.$$

Hence,

$$\sum_{j=1}^{\infty} f(j) \approx \sum_{j=1}^{k-1} f(j) + c\mathbf{E}_{k,d}^{T}\left(x^{-\beta}\mathbf{x}_{\mu+1}(x)\right)\begin{bmatrix} I_{2\times 2} & \mathbf{0}_{2\times\mu} \\ \mathbf{0}_{\mu\times 2} & B_{n,k} \end{bmatrix} D_{n,k}\widehat{\mathbf{g}}_{n,k}. \tag{14}$$

The next example illustrates how formula (14) can be used to estimate Euler's constant

$$\gamma = \lim_{n\to\infty}\left(1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n} - \ln n\right) = 0.577\,215\,665.$$

Example 5 *Longman [Longman, 1987] notes that γ can be written as*

$$\gamma = \lim_{n\to\infty}(1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n} - \ln(n+1))$$

so that Euler's constant can be expressed as the sum of the series

$$\gamma = \sum_{n=1}^{\infty}\left(\frac{1}{n} - \ln\frac{n+1}{n}\right).$$

Set

$$f(x) = \frac{1}{x} - \ln\left(\frac{x+1}{x}\right)$$

so that

$$f(1/x) = x - \ln\left(\frac{\frac{1}{x}+1}{\frac{1}{x}}\right) = x - \ln(1+x).$$

Since

$$\lim_{x\to 0}\frac{x-\ln(1+x)}{x^2}=\frac{1}{2}$$

it follows that $c=\frac{1}{2}$, $\beta=2$, *and*

$$g(x)=\begin{cases} 2^{\frac{x-\ln(1+x)}{x^2}} & \text{if } x>-1, x\neq 0 \\ 1 & \text{if } x=0 \end{cases}.$$

Moreover, $g'(0)=\lim_{x\to 0}\frac{g(x)-1}{x}=-\frac{2}{3}$ *so*

$$\widehat{\mathbf{g}}_{8,11}^T=\left[\begin{array}{cccccc} 1 & -\frac{2}{3} & \frac{181}{12}-128\ln\frac{9}{8} & \frac{461}{27}-162\ln\frac{10}{9} & \frac{286}{15}-200\ln\frac{11}{10}\end{array}\right].$$

Since

$$\mathbf{E}_{11,3}^T\left(x^{-2}\mathbf{x}_4(x)\right)=\left[\begin{array}{ccccc}\frac{459\,799}{4831\,530} & \frac{48\,097}{10\,629\,366} & \frac{5586}{19\,487\,171} & \frac{35\,007}{1714\,871\,048} & \frac{54\,806}{35\,369\,215\,365}\end{array}\right]$$

we have

$$\gamma\approx\sum_{j=1}^{10}f(j)+\frac{1}{2}\mathbf{E}_{11,3}^T\left(x^{-2}\boldsymbol{x}_4(x)\right)\left[\begin{array}{cc} I_{2\times 2} & \mathbf{0}_{2\times 3} \\ \mathbf{0}_{3\times 2} & B_{8,11}\end{array}\right]D_{8,11}\widehat{\mathbf{g}}_{8,11}=0.577\,215\,662$$

which compares favorably with the approximation $\gamma\approx 0.577\,215\,664$ *obtained by using* $E_{11,3}(f)$ *in (EM). The true answer to nine decimal places is* $\gamma=0.577215665$.

This strategy can be extended to include additional derivative values of g at 0. For example, if $g''(0)$ is available we have the estimate

$$\sum_{j=1}^{\infty}f(j)\approx\sum_{j=1}^{k-1}f(j)+c\mathbf{E}_{k,d}^T\left(x^{-\beta}\mathbf{x}_{\mu+2}(x)\right)\left[\begin{array}{cc} I_{3\times 3} & \mathbf{0}_{3\times\mu} \\ \mathbf{0}_{\mu\times 3} & B_{n,k}\end{array}\right]D_{n,k}^2\widehat{\widehat{\mathbf{g}}}_{n,k}$$

where

$$\widehat{\widehat{\mathbf{g}}}_{n,k}=\left[\begin{array}{c} 1 \\ g'(0) \\ g''(0)/2 \\ g\left(\frac{1}{n}\right)-1-g'(0)\frac{1}{n}-\frac{g''(0)}{2n^2} \\ g\left(\frac{1}{n+1}\right)-1-g'(0)\frac{1}{n+1}-\frac{g''(0)}{2(n+1)^2} \\ \vdots \\ g\left(\frac{1}{k-1}\right)-1-g'(0)\frac{1}{k-1}-\frac{g''(0)}{2(k-1)^2}\end{array}\right].$$

Numerical evidence suggests that using $g''(0)$ does not significantly improve the accuracy although it did produce a somewhat better approximation to γ than did the EM estimate with $k=11$ and $d=3$.

5 ERROR ESTIMATES

The error introduced when using (EM) to estimate the sum of a series is no more than twice the absolute value of the first omitted term ([Stoer and Bulirsch, 1995]). For example, the error in the approximation given in Example 1 does not exceed $\frac{2}{120}\left|\frac{d}{dx}x^{-2}\right|_{x=10} = \frac{1}{30000}$. Error bounds, however, are often too restrictive and can be difficult to compute. Perhaps a more useful tool for determining the goodness of an approximation when a true value t is replaced by an estimate e is to approximate the relative error

$$r_t = \frac{t-e}{t}.$$

Notice that this error measures how well e/t estimates 1 and is used to determine the number of *significant digits* in an approximation. An estimate is correct to s significant digits if $|r_t| < 5 \times 10^{-s}$. (See [5] for more details.) Of course, the true value t is not available so it is often replaced by a second estimate τ that is (*hopefully!*) more accurate than e. A convenient option in our setting is to replace the true value $t = \sum_{j=1}^{\infty} f(j)$ with formula (14). This gives the following estimate for the relative error when using (MEM),

$$r_\tau = \frac{\tau - e}{\tau}$$

where e and τ are given by formulas (MEM) and (14) respectively.

To demonstrate the effectiveness of this estimate for $f(x) = (2x^3 + x^2 + 1)^{-1}$ recall that the value $e = 0.331\,491\,171$ was given in Example 2. A quick calculation using formula (14) with $k = 11$ and $d = 3$ yields $\tau = 0.331\,491\,164$ so the estimated relative error is

$$r_\tau = -2.11 \times 10^{-8}.$$

This suggests that e is accurate to eight significant digits which is correct since $r_t = -2.11 \times 10^{-8}$. Although this strategy measures the relative error of e the approximation τ is used as the final estimate to t thus providing a little extra insurance on the accuracy. It is worth noting that for this function the value of τ is a bit more accurate than the equivalent EM estimate.

In Example 4 we computed $e = 1.472\,828\,238$ for the function $f(x) = \frac{\sin x^{-1}}{x}$. Using formula (14) with $k = 11$ and $d = 3$ we obtain $\tau = 1.472\,828\,231$. These values give $r_\tau = -4.75 \times 10^{-9}$ suggesting that e is accurate to nine significant digits. This conclusion is correct since the true relative error is $r_t = -4.07 \times 10^{-9}$.

The approximation represented by τ for Euler's constant was given in Example 5. A simple calculation with $k = 11$ and $d = 3$ yields $e = 0.577\,215\,769$ which should be correct to seven significant digits since $r_\tau = -1.85 \times 10^{-7}$. The true relative error is $r_t = -1.80 \times 10^{-7}$.

References

[1] Goldstein, H. (1977). *A History of Numerical Analysis.* New York: Springer-Verlag.

[2] Longman, I.M. (1987). "Increasing the Convergence Rate of Series." New York: Elsevier Science.

[3] Mathews, J. and K. Fink (1999). *Numerical Methods Using MATLAB.* Upper Saddle River: Prentice-Hall, Inc.

[4] Ralston, A. and P. Rabinowitz. (1978). *A First Course in Numerical Analysis.* New York: McGraw-Hill.

[5] Stoer, J. and R. Bulirsch. (1995). *Introduction to Numerical Analysis.* New York: Springer-Verlag.