

THE USE OF ECONOMIZED POLYNOMIALS IN MATHEMATICAL TABLES

By C. W. CLENSHAW AND F. W. J. OLVER

Communicated by E. T. GOODWIN

Received 28 September 1954

1. *Introduction and summary.* The advantages of using polynomial approximations for the purpose of constructing interpolable numerical tables on punched cards have been pointed out by Sadler (7). The object of this paper is to demonstrate the value of this method for ordinary published tables.

Powerful polynomial representations of a function $f(x)$ valid throughout a tabular interval $a \leq x \leq a + h$ are obtained by truncating the expansion of $f(x)$ in Chebyshev polynomials†, given by

$$f(a + ph) = \alpha_0 + 2\alpha_1 T_1^*(p) + 2\alpha_2 T_2^*(p) + \dots \quad (0 \leq p \leq 1), \tag{1.1}$$

at the term $2\alpha_n T_n^*(p)$, and rearranging the result in powers of p in the form

$$f(a + ph) = c_0 + c_1 p + c_2 p^2 + \dots + c_n p^n + \eta(p), \tag{1.2}$$

where

$$\eta(p) \equiv 2\alpha_{n+1} T_{n+1}^*(p) + 2\alpha_{n+2} T_{n+2}^*(p) + \dots \tag{1.3}$$

is the truncation error, normally restricted not to exceed half a unit in the last decimal place retained by making n sufficiently large. The coefficients c_0, c_1, \dots, c_n may be given in the table side by side with f , and interpolation is then carried out by application of (1.2).

From the standpoint of the *user* such a method enjoys considerable advantages over formulae involving differences or modified differences. It is both simpler and quicker; the whole of the calculation can be carried out on a desk machine without any intermediate recording, and the use of tables of interpolation coefficients is dispensed with.

A disadvantage of this method compared with those based on differences or modified differences is that it requires more space if the same interval in the argument is used. Even so, if it is desired to save space in a given table without unduly inconveniencing the user it may still be preferable to use economized polynomials, but with an increased argument interval. The addition of a term or two to the polynomial permits the tabular interval to be increased considerably, and the labour of evaluating a polynomial of slightly higher degree is often no greater than that of interpolation at the smaller interval using other methods. An example of this is given later (§9) together with a comparison of the speeds of interpolation using various types of interpolation facilities.

† The notation used here for Chebyshev polynomials is that of Lanczos (5), and is defined by

$$T_n(x) = \cos(n \cos^{-1} x), \quad T_n^*(x) = T_n(2x - 1) = T_{2n}(\sqrt{x}).$$

The coefficients $\alpha_0, \alpha_1, \dots$ in (1.1) and c_0, c_1, \dots in (1.2) may be computed in various ways. For table-making purposes they can be conveniently expressed in terms of the central differences of f . A point to be noticed is that the value of c_0 is not, in general, equal to $f(a)$, for in constructing the approximation no restrictions are made concerning the actual values to be taken at the end-points of the range. In some circumstances, for example, with tables on punched cards, it may be permissible to tabulate c_0 in place of f . Indeed, in these cases it may even be advantageous to construct *optimum-* or *maximum-interval tables*, described by Herget and Clemence (2) and Sadler (7), in which the tabular interval need not be constant and the function values are modified drastically so that the argument itself is used as the variable in the polynomial rather than the fraction of the tabular interval.

In the case of ordinary printed tables, however, it would be undesirable to tabulate c_0 in place of f , and uneconomic to tabulate them both. A preferable procedure is to modify the representation and replace the constant term by $f(a)$. This can be done in such a way that the power of the approximation is not seriously impaired, and we investigate this point in §§2 and 3.

In §4 the rounding errors associated with (1.2) are considered. In §5 the choice of interpolation polynomials for the range $0 \leq p \leq 1$ is discussed with reference to the combined effects of truncation and rounding errors. Explicit formulae for the coefficients of polynomials of degrees 2 to 6 are given in §6.

Other forms of polynomial approximations are briefly discussed in §7. In §8 the evaluation of derivatives using the economized interpolation polynomials is considered, and in the concluding section (§9) numerical examples are given together with an account of time tests that have been carried out with various standard methods of interpolation.

2. *The interpolation cubic in $0 \leq p \leq 1$.* For the purpose of illustration we construct first the interpolation polynomial of degree 3. Truncating (1.1) at the term $2\alpha_3 T_3^*(p)$, substituting the formulae given by Miller (4) for the coefficients $\alpha_0, \alpha_1, \dots$, and rearranging in powers of p , we obtain the relation (1.2) with $n = 3$ and the coefficients c_0, c_1, c_2, c_3 and η (all of which depend on n) given by

$$\left. \begin{aligned} c_0 &= f_0 - \frac{2}{2^8 \cdot 4!} \mu \delta_{\frac{1}{2}}^4 + \frac{2}{2^{10} \cdot 5!} \delta_{\frac{1}{2}}^5 + \frac{266}{2^{12} \cdot 6!} \mu \delta_{\frac{1}{2}}^6 - \dots, \\ c_1 &= \frac{4}{2^2 \cdot 1!} \delta_{\frac{1}{2}} - \frac{16}{2^4 \cdot 2!} \mu \delta_{\frac{1}{2}}^2 + \frac{32}{2^6 \cdot 3!} \delta_{\frac{1}{2}}^3 + \frac{576}{2^8 \cdot 4!} \mu \delta_{\frac{1}{2}}^4 - \frac{1124}{2^{10} \cdot 5!} \delta_{\frac{1}{2}}^5 - \frac{57584}{2^{12} \cdot 6!} \mu \delta_{\frac{1}{2}}^6 + \dots, \\ c_2 &= \frac{16}{2^4 \cdot 2!} \mu \delta_{\frac{1}{2}}^2 - \frac{96}{2^6 \cdot 3!} \delta_{\frac{1}{2}}^3 - \frac{576}{2^8 \cdot 4!} \mu \delta_{\frac{1}{2}}^4 + \frac{3360}{2^{10} \cdot 5!} \delta_{\frac{1}{2}}^5 + \frac{57584}{2^{12} \cdot 6!} \mu \delta_{\frac{1}{2}}^6 - \dots, \\ c_3 &= \frac{64}{2^6 \cdot 3!} \delta_{\frac{1}{2}}^3 - \frac{2240}{2^{10} \cdot 5!} \delta_{\frac{1}{2}}^5 + \dots, \end{aligned} \right\} \quad (2.1)$$

$$\eta(p) = 2\alpha_4 T_4^*(p) + 2\alpha_5 T_5^*(p) + \dots \quad (2.2)$$

Here, in the usual notation, the odd and mean even differences of f at the point $x = a + \frac{1}{2}h$ are denoted by $\delta_{\frac{1}{2}}^{2s-1}$ and $\mu \delta_{\frac{1}{2}}^{2s}$ ($s = 1, 2, \dots$) respectively, and $f_0 \equiv f(a)$.

The size of the interval of tabulation will invariably be such that the significant differences of f diminish steadily with increasing order. The dominant part of the error term is accordingly

$$\eta(p) \doteq 2\alpha_4 T_4^*(p) \doteq \mu \delta_{\frac{1}{2}}^4 T_4^*(p) / 3072. \tag{2.3}$$

Since $|T_4^*(p)| \leq 1$ throughout the range $0 \leq p \leq 1$, it follows that† $|\eta(p)| < \frac{1}{2}$ if

$$|\mu \delta_{\frac{1}{2}}^4| < 1536. \tag{2.4}$$

The maximum error equals $|f_0 - c_0|$ approximately. If we replace c_0 in (1.2) by f_0 , we obtain

$$f(a + ph) = f_0 + c_1 p + c_2 p^2 + c_3 p^3 + \eta_1(p), \tag{2.5}$$

where, from (2.3) and the first of (2.1),

$$\eta_1(p) \doteq \mu \delta_{\frac{1}{2}}^4 \{T_4^*(p) - 1\} / 3072. \tag{2.6}$$

To ensure that $|\eta_1(p)|$ does not exceed half a unit we must apply the restriction

$$|\mu \delta_{\frac{1}{2}}^4| < 768, \tag{2.7}$$

because the maximum value of $|T_4^*(p) - 1|$ in $0 \leq p \leq 1$ is 2. The limit (2.7) is substantially lower than (2.4). Some of the loss can, however, be recovered in the following way.

Suppose that $\phi(p)$ is a polynomial of degree 3 or less which approximates to $T_4^*(p) - 1$ throughout $0 \leq p \leq 1$, and has the property $\phi(0) = 0$. Then it is evident that if we incorporate the quantity $\mu \delta_{\frac{1}{2}}^4 \phi(p) / 3072$ in the interpolation cubic, the size of the error term will be reduced.

A suitable approximation of this kind (see §3) is given by

$$T_4^*(p) - 1 = k(4p^2 - 4p) + \psi(p), \tag{2.8}$$

where $k = 24 - 16\sqrt{2} = 1.37\dots$, and

$$|\psi(p)| \leq k \quad (0 \leq p \leq 1). \tag{2.9}$$

Substituting (2.8) in (2.6) and using (2.5), we obtain

$$f(a + ph) = f_0 + a_1 p + a_2 p^2 + a_3 p^3 + \epsilon(p), \tag{2.10}$$

where

$$a_1 = c_1 - \frac{4k\mu\delta_{\frac{1}{2}}^4}{3072}, \quad a_2 = c_2 + \frac{4k\mu\delta_{\frac{1}{2}}^4}{3072}, \quad a_3 = c_3, \tag{2.11}$$

and

$$\epsilon(p) \doteq \mu \delta_{\frac{1}{2}}^4 \psi(p) / 3072. \tag{2.12}$$

From (2.9) we see that $|\epsilon(p)| < \frac{1}{2}$ if

$$|\mu \delta_{\frac{1}{2}}^4| < 1536/k = 1120, \tag{2.13}$$

on rounding off to the nearest 10, which should be compared with (2.4) and (2.7). Explicit formulae for a_1, a_2, a_3 in terms of the differences are given in §6.

† Here and elsewhere it is supposed that the units are in terms of the last decimal place retained.

The same procedure may be used for constructing interpolation polynomials of any degree n . The expressions corresponding to (2.6) for the dominant part of the error term, when c_0 is replaced by f_0 , are

$$\frac{\delta_1^{\frac{n+1}{2}}}{2^{2n+1}(n+1)!} \{T_{n+1}^*(p) + 1\} \quad \text{or} \quad \frac{\mu \delta_1^{\frac{n+1}{2}}}{2^{2n+1}(n+1)!} \{T_{n+1}^*(p) - 1\}, \tag{2.14}$$

according as n is even or odd. In order to reduce the error term in the manner described we need polynomial approximations of degree not exceeding n to the function $T_{n+1}^*(p) - (-1)^{n+1}$. These we proceed to consider.

3. *Approximations to $T_s^*(p) - (-1)^s$.* Let us write

$$T_s^*(p) - (-1)^s = \phi(p) + \psi(p), \tag{3.1}$$

where $\phi(p)$ is a polynomial of degree not exceeding $s - 1$, and $\psi(0) = 0$. We wish to know the $\phi(p)$ for which the maximum value of $|\psi(p)|$ in $0 \leq p \leq 1$ is a minimum.

We shall solve the problem with the added condition $\psi(1) = 0$, so that the approximation $\phi(p)$ has to take the same value as $T_s^*(p) - (-1)^s$ at both $p = 0$ and $p = 1$. Although the imposition of this condition weakens the approximation, we see later (§ 5) that this is more than compensated by the favourable effect it has on the rounding error in the application of (1.2).

The problem is solved if $\psi(p)$ has $s - 1$ turning points p_1, p_2, \dots, p_{s-1} such that

$$0 < p_1 < p_2 < \dots < p_{s-1} < 1, \tag{3.2}$$

and
$$\psi(p_1) = -\psi(p_2) = \psi(p_3) = \dots = (-1)^s \psi(p_{s-1}). \tag{3.3}$$

For suppose $\phi_1(p)$ is a better approximation, and

$$\psi_1(p) \equiv T_s^*(p) - (-1)^s - \phi_1(p). \tag{3.4}$$

Then $\psi(p) - \psi_1(p) \equiv \phi_1(p) - \phi(p)$ is a polynomial of degree $s - 1$ which vanishes at $p = 0$ and 1, and changes sign successively at the points p_1, p_2, \dots, p_{s-1} . It is therefore identically zero.

The solution may now be seen to be

$$\phi(p) = T_s^*(p) - (-1)^s - kT_s \left\{ (2p - 1) \cos \frac{\pi}{2s} \right\}, \tag{3.5}$$

where k is a constant, chosen so that the term in p^s disappears. For this gives

$$\psi(p) = kT_s \left\{ (2p - 1) \cos \frac{\pi}{2s} \right\}, \tag{3.6}$$

which vanishes at $p = 0$ and 1, and has just the properties required by (3.2) and (3.3).

The value of k is readily found to be

$$k = \sec^s \frac{\pi}{2s}, \tag{3.7}$$

and the maximum value of $|\psi(p)|$ in $0 \leq p \leq 1$ is k .

Numerical formulae for the ϕ -polynomials, giving the coefficients and the corresponding values of k to two decimals, are

$$\left. \begin{aligned} s = 3 \quad k = 1.54 \quad \phi = 2p, \\ s = 4 \quad k = 1.37 \quad \phi = -5.49p + 5.49p^2, \\ s = 5 \quad k = 1.29 \quad \phi = 10.45p - 25.34p^2 + 16.89p^3, \\ s = 6 \quad k = 1.23 \quad \phi = -16.86p + 72.00p^2 - 110.28p^3 + 55.14p^4, \\ s = 7 \quad k = 1.19 \quad \phi = 24.73p - 161.55p^2 + 418.88p^3 - 466.77p^4 + 186.71p^5. \end{aligned} \right\} \quad (3.8)$$

4. *Rounding error.* We now examine the effects of rounding errors in the application of the formula

$$f(a + ph) = f_0 + a_1p + a_2p^2 + \dots + a_n p^n + \epsilon(p) \quad (0 \leq p \leq 1), \quad (4.1)$$

in which $\epsilon(p)$ is the truncation error; we may suppose that

$$\epsilon(0) = \epsilon(1) = 0 \quad \text{and} \quad |\epsilon(p)| \leq \frac{1}{2} \quad \text{in} \quad 0 \leq p \leq 1. \quad (4.2)$$

We assume that the function f and the coefficients in (4.1) are computed originally with at least one guarding figure, and rounded before being given in the final table. If no special precautions are taken the rounding error in (4.1) may amount to $\frac{1}{2}(n + 1)$ units in the last decimal place. This is, of course, undesirably large. The most obvious way of reducing it is to tabulate the coefficients a_1, a_2, \dots, a_n to one more decimal place than f , but this is uneconomic and it is preferable to proceed as follows.

Let f' and a'_m denote the rounded values of f and a_m , and let

$$f_0 = f'_0 + g_0, \quad f_1 = f'_1 + g_1, \quad a_m = a'_m + b_m \quad (m = 1, 2, \dots, n), \quad (4.3)$$

so that

$$|g_0| \leq \frac{1}{2}, \quad |g_1| \leq \frac{1}{2}, \quad |b_m| \leq \frac{1}{2}. \quad (4.4)$$

Putting $p = 1$ in (4.1) and using (4.2), we obtain

$$f_1 \equiv f(a + h) = f_0 + a_1 + a_2 + \dots + a_n. \quad (4.5)$$

The most unfavourable rounding errors will obviously all have the same sign, giving the worst possible error at $p = 1$. It is natural, therefore, to consider adjusting the roundings so that (4.5) is satisfied *exactly* by the rounded values, that is, to satisfy

$$f'_1 = f'_0 + a'_1 + a'_2 + \dots + a'_n. \quad (4.6)$$

This can be done systematically by rounding, in the usual way, f_0, f_1 and all but one, a_m say, of the coefficients a_r , and replacing a'_m by a_m^* , where

$$a_m^* = f'_1 - f'_0 - a'_1 - a'_2 - \dots - a'_{m-1} - a'_{m+1} - \dots - a'_n. \quad (4.7)$$

Then from (4.3), (4.5) and (4.7), we obtain

$$a_m - a_m^* = g_1 - g_0 - b_1 - b_2 - \dots - b_{m-1} - b_{m+1} - \dots - b_n. \quad (4.8)$$

Using this result and (4.1), we see that in the formula

$$f(a + ph) \div f'_0 + a'_1 p + \dots + a'_{m-1} p^{m-1} + a_m^* p^m + a'_{m+1} p^{m+1} + \dots + a'_n p^n, \quad (4.9)$$

the total error $E_m(p)$ —the amount we must add to produce the correct value of $f(a + ph)$ —is given by

$$\begin{aligned}
 E_m(p) &= \epsilon(p) + g_0 + b_1 p + \dots + b_{m-1} p^{m-1} + b_{m+1} p^{m+1} + \dots + b_n p^m \\
 &\quad + (g_1 - g_0 - b_1 - \dots - b_{m-1} - b_{m+1} - \dots - b_n) p^m \\
 &= \epsilon(p) + g_0(1 - p^m) + g_1 p^m + \sum_{s=m}^n b_s (p^s - p^m).
 \end{aligned}
 \tag{4.10}$$

Hence using (4.2), (4.4) and the fact that $0 \leq p \leq 1$, we see that

$$|E_m(p)| \leq \frac{1}{2} V_m(p), \tag{4.11}$$

where
$$\begin{aligned}
 V_m(p) &= 1 + (1 - p^m) + p^m + \sum_{s=1}^{m-1} (p^s - p^m) + \sum_{s=m+1}^n (p^m - p^s) \\
 &= 2 + p + p^2 + \dots + p^{m-1} + (n - 2m + 1) p^m - p^{m+1} - p^{m+2} - \dots - p^n.
 \end{aligned}
 \tag{4.12}$$

The value of m is at our disposal and we now seek the value which makes $V_m(p)$ a minimum.

From (4.12) we have

$$V_m(p) - V_{m+1}(p) = (n - 2m)(p^m - p^{m+1}), \tag{4.13}$$

which is positive or negative according as m is less or greater than $\frac{1}{2}n$, whatever the value of p in $0 < p < 1$. Thus if n is odd the least member of the sequence

$$V_m(p) \quad (m = 1, 2, \dots, n)$$

has $m = \frac{1}{2}n + \frac{1}{2}$; if n is even then $V_{\frac{1}{2}n}(p) = V_{\frac{1}{2}n+1}(p)$, and these are the least members of the sequence.

Maximum values in $0 \leq p \leq 1$ of the minimum V 's are, to two decimals, as follows:

$n = 2$	$V_1 = V_2 = 2 + p - p^2$	$\max \frac{1}{2}V_1 = 1.13,$	} (4.14)
$n = 3$	$V_2 = 2 + p - p^3$	$\max \frac{1}{2}V_2 = 1.19,$	
$n = 4$	$V_2 = V_3 = 2 + p + p^2 - p^3 - p^4$	$\max \frac{1}{2}V_2 = 1.31,$	
$n = 5$	$V_3 = 2 + p + p^2 - p^4 - p^5$	$\max \frac{1}{2}V_3 = 1.39,$	
$n = 6$	$V_3 = V_4 = 2 + p + p^2 + p^3 - p^4 - p^5 - p^6$	$\max \frac{1}{2}V_3 = 1.51.$	

These limits for the total error are a marked improvement on the limit $(\frac{1}{2}n + 1)$ units when no rounding precautions are taken.

5. *Choice of interpolation polynomials for the range $0 \leq p \leq 1$.* If, in the error term (2.14), we substitute the formula (3.1) with $s = n + 1$ and $\phi(p)$ given by (3.5), and then incorporate the term †

$$2^{-2n-1} \phi(p) \delta_{\frac{1}{2}}^{n+1} / (n + 1)!$$

in the interpolation polynomial in the manner described in §2, the new truncation error is

$$2^{-2n-1} \psi(p) \delta_{\frac{1}{2}}^{n+1} / (n + 1)!, \tag{5.1}$$

where $\psi(p)$ is given by (3.6). To ensure that this does not exceed half a unit we must apply the restriction

$$|\delta_{\frac{1}{2}}^{n+1}| < 2^{2n} (n + 1)! / k, \tag{5.2}$$

where k is defined by (3.7) with $s = n + 1$. The values of k , for $n = 2, 3, 4, 5, 6$, are given in the first row of Table 1 below (cf. (3.8)).

† If n is odd $\delta_{\frac{1}{2}}^{n+1}$ is to be replaced by $\mu \delta_{\frac{1}{2}}^{n+1}$.

To a sufficient degree of approximation the truncation error vanishes at $p = 0$ and $p = 1$. Hence if the coefficients a_1, \dots, a_n of the interpolation polynomial are rounded in accordance with equation (4.7) with $m = [\frac{1}{2}n + \frac{1}{2}]$, the maximum error in an unrounded interpolate is given by (4.14). It is shown in Table 1 in the row marked 'Error A'.

Suppose that instead of (3.5) we had used the more powerful approximation

$$\phi(p) = T_s^*(p) - (-1)^s - lT_s \left(\left(1 + \cos \frac{\pi}{2s} \right) p - \cos \frac{\pi}{2s} \right), \tag{5.3}$$

where
$$l = \sec^{2s} \frac{\pi}{4s}, \tag{5.4}$$

which, it may be verified, is the solution to the problem propounded in the opening paragraph of §3 without the condition $\psi(1) = 0$. Then in place of (5.2) we have

$$|\delta_{\frac{1}{2}}^{n+1}| < 2^{2n}(n+1)!/l, \tag{5.5}$$

Table 1

n	2	3	4	5	6
k	1.54	1.37	1.29	1.23	1.19
Error A	1.13	1.19	1.31	1.39	1.51
l	1.23	1.17	1.13	1.11	1.09
Error B	1.50	1.50	1.53	1.61	1.68
Error C	1.40	1.43	1.47	1.56	1.64

which is less restrictive than (5.2) itself because $l < k$ (see Table 1). Now, however, the truncation error does not vanish at $p = 1$, but takes its maximum value there. The analysis of §4 may be modified to take account of this, and it is found that the rounding error is increased. Assuming that the maximum truncation error is half a unit, and that the best value of m is used, in this case $[\frac{1}{2}n + 1]$, we find the numerical values of the total error to be those given in the row 'Error B' of Table 1.

A fair comparison of the two approximations can be made by assessing the maximum total error, Error C say, which results if the approximation given by (3.1) and (5.3) is used with the condition (5.2) in place of (5.5). This is given by

$$\text{Error C} = (\text{Error B} - 0.5) + 0.5 \times (l/k),$$

and appears in the last row of Table 1. In every case it exceeds Error A.

In conclusion, therefore, provided that the coefficients a_1, \dots, a_n are rounded as in §4, the approximation (3.5) is to be preferred to (5.3)†.

† *Note added in proof:* Slightly sharper estimates for Error A can be obtained by taking advantage in the analysis of §4 of the fact that the truncation error is now of the specific form (5.1). With the condition (5.2) we obtain $|\epsilon(p)| \leq \frac{1}{2} |\psi(p)/k|$, and substituting this inequality in (4.10) we find that $|E_m(p)| \leq \frac{1}{2} W_m(p)$, $W_m(p) \equiv V_m(p) - 1 + |\psi(p)/k$.

The best value of m is again $[\frac{1}{2}n + \frac{1}{2}]$ and the greatest value of $\frac{1}{2}W_{[\frac{1}{2}n + \frac{1}{2}]}(p)$ in $0 \leq p \leq 1$ is 1.09, 1.19, 1.31, 1.38, 1.47, for $n = 2, 3, 4, 5, 6$, respectively. We can treat Error B in a similar way, but it is found that the estimates are unaffected to two decimal places, except for $n = 5$ for which 1.61 is reduced to 1.59. The general conclusion at the end of the section is unaffected.

6. *Interpolation polynomials for the range* $0 \leq p \leq 1$. We now give the expansions in terms of the differences for the coefficients a_1, a_2, \dots, a_n in the formula

$$f(a + ph) = f_0 + a_1p + a_2p^2 + \dots + a_np^n + \epsilon(p) \quad (0 \leq p \leq 1), \tag{6.1}$$

together with the maximum value which the $(n + 1)$ th difference may have without the truncation error $|\epsilon(p)|$ exceeding half a unit in the last decimal place. For simplicity the suffix $\frac{1}{2}$ has been omitted from the δ 's and $\mu\delta$'s.

The quantity E given with the formulae is the maximum total error (rounding error plus truncation error) in an unrounded interpolate obtained from (6.1) with $\epsilon(p)$ neglected, on the assumption that δ^{n+1} does not exceed the stated limit (see 'Error A' of Table 1).

Quadratic: $\delta^3 < 60, E = 1.1:$

$$a_1 = f_1 - f_0 - a_2 \quad (\text{rounded values}),$$

$$a_2 = \frac{1}{2}\mu\delta^2 - 0.094\mu\delta^4 + \dots$$

Cubic: $\delta^4 < 1100, E = 1.2:$

$$a_1 = \delta - \frac{1}{2}\mu\delta^2 + \frac{1}{12}\delta^3 + 0.09196\mu\delta^4 - 0.0091\delta^5 - 0.0195\mu\delta^6 + 0.001\delta^7 + \dots,$$

$$a_2 = f_1 - f_0 - a_1 - a_3 \quad (\text{rounded values}),$$

$$a_3 = \frac{1}{6}\delta^3 - 0.0182\delta^5 + 0.003\delta^7 - \dots$$

Quartic: $\delta^5 < 24000, E = 1.3:$

$$a_1 = \delta - \frac{1}{2}\mu\delta^2 + \frac{1}{12}\delta^3 + \frac{1}{12}\mu\delta^4 - 0.008977\delta^5 - 0.01662\mu\delta^6 + 0.00135\delta^7 + 0.0036\mu\delta^8 - 0.0002\delta^9 - 0.001\mu\delta^{10} + \dots,$$

$$a_2 = f_1 - f_0 - a_1 - a_3 - a_4 \quad (\text{rounded values}),$$

$$a_3 = \frac{1}{6}\delta^3 - \frac{1}{12}\mu\delta^4 - 0.017954\delta^5 + 0.02326\mu\delta^6 + 0.00269\delta^7 - 0.0057\mu\delta^8 - 0.0005\delta^9 + 0.001\mu\delta^{10} + \dots,$$

$$a_4 = \frac{1}{24}\mu\delta^4 - 0.01163\mu\delta^6 + 0.0029\mu\delta^8 - 0.001\mu\delta^{10} + \dots$$

Quintic: $\delta^6 < 600000, E = 1.4:$

$$a_1 = \delta - \frac{1}{2}\mu\delta^2 + \frac{1}{12}\delta^3 + \frac{1}{12}\mu\delta^4 - \frac{1}{120}\delta^5 - 0.0166293\mu\delta^6 + 0.001188\delta^7 + 0.003554\mu\delta^8 - 0.00020\delta^9 - 0.00079\mu\delta^{10} + 0.0000\delta^{11} + \dots,$$

$$a_2 = \frac{1}{2}\mu\delta^2 - \frac{1}{4}\delta^3 - \frac{1}{24}\mu\delta^4 + \frac{1}{48}\delta^5 + 0.0050347\mu\delta^6 - 0.002740\delta^7 - 0.000686\mu\delta^8 + 0.00044\delta^9 + 0.00010\mu\delta^{10} - 0.0001\delta^{11} - \dots,$$

$$a_3 = f_1 - f_0 - a_1 - a_2 - a_4 - a_5 \quad (\text{rounded values}),$$

$$a_4 = \frac{1}{24}\mu\delta^4 - \frac{1}{48}\delta^5 - 0.0115946\mu\delta^6 + 0.004123\delta^7 + 0.002867\mu\delta^8 - 0.00079\delta^9 - 0.00069\mu\delta^{10} + 0.0002\delta^{11} + \dots,$$

$$a_5 = \frac{1}{120}\delta^5 - 0.001649\delta^7 + 0.00032\delta^9 - 0.0001\delta^{11} + \dots$$

Sextic: $\delta^7 < 170\,000\,000$, $E = 1.5$:

$$\begin{aligned}
 a_1 &= \delta - \frac{1}{2}\mu\delta^2 + \frac{1}{12}\delta^3 + \frac{1}{12}\mu\delta^4 - \frac{1}{120}\delta^5 - \frac{1}{60}\mu\delta^6 + 0.00118\,8702\delta^7 + 0.00357\,153\mu\delta^8 \\
 &\quad - 0.00019\,774\delta^9 - 0.00079\,37\mu\delta^{10} + 0.00003\,59\delta^{11} + 0.00018\,0\mu\delta^{12} - \dots, \\
 a_2 &= \frac{1}{2}\mu\delta^2 - \frac{1}{4}\delta^3 - \frac{1}{24}\mu\delta^4 + \frac{1}{48}\delta^5 + \frac{1}{80}\mu\delta^6 - 0.00274\,3713\delta^7 - 0.00089\,489\mu\delta^8 \\
 &\quad + 0.00043\,574\delta^9 + 0.00015\,96\mu\delta^{10} - 0.00007\,69\delta^{11} - 0.00003\,0\mu\delta^{12} + \dots, \\
 a_3 &= f_1 - f_0 - a_1 - a_2 - a_4 - a_5 - a_6 \quad (\text{rounded values}), \\
 a_4 &= \frac{1}{24}\mu\delta^4 - \frac{1}{48}\delta^5 - \frac{1}{144}\mu\delta^6 + 0.00411\,1958\delta^7 + 0.00115\,134\mu\delta^8 - 0.00078\,765\delta^9 \\
 &\quad - 0.00019\,74\mu\delta^{10} + 0.00015\,44\delta^{11} + 0.00003\,5\mu\delta^{12} - \dots, \\
 a_5 &= \frac{1}{120}\delta^5 - \frac{1}{240}\mu\delta^6 - 0.00164\,4783\delta^7 + 0.00152\,530\mu\delta^8 + 0.00031\,506\delta^9 \\
 &\quad - 0.00043\,67\mu\delta^{10} - 0.00006\,18\delta^{11} + 0.00011\,5\mu\delta^{12} + \dots, \\
 a_6 &= \frac{1}{720}\mu\delta^6 - 0.00050\,843\mu\delta^8 + 0.00014\,56\mu\delta^{10} - 0.00003\,8\mu\delta^{12} + \dots
 \end{aligned}$$

It is supposed (in accordance with §4) that a_1, a_2, \dots, a_n are computed retaining at least one guarding figure and subsequently rounded, with the exception of $a_{\lfloor \frac{n+1}{2} \rfloor}$, which is evaluated from the rounded values of the other coefficients and the next tabular value f_1 , as indicated in the formulae. For completeness and checking purposes, however, we record the expansions of these particular coefficients.

Quadratic: $a_1 = \delta - \frac{1}{2}\mu\delta^2 + 0.094\mu\delta^4 + \dots$

Cubic: $a_2 = \frac{1}{2}\mu\delta^2 - \frac{1}{4}\delta^3 - 0.09196\mu\delta^4 + 0.0273\delta^5 + 0.0195\mu\delta^6 - 0.004\delta^7 - \dots$

Quartic: $a_2 = \frac{1}{2}\mu\delta^2 - \frac{1}{4}\delta^3 - \frac{1}{24}\mu\delta^4 + 0.02693\,1\delta^5 + 0.00499\mu\delta^6 - 0.00403\delta^7 \\ - 0.0007\mu\delta^8 + 0.0007\delta^9 + 0.000\mu\delta^{10} - \dots$

Quintic: $a_3 = \frac{1}{6}\delta^3 - \frac{1}{12}\mu\delta^4 + 0.02318\,91\mu\delta^6 - 0.00092\,2\delta^7 - 0.00573\,5\mu\delta^8 \\ + 0.00023\delta^9 + 0.00138\mu\delta^{10} - 0.0001\delta^{11} - \dots$

Sextic: $a_3 = \frac{1}{6}\delta^3 - \frac{1}{12}\mu\delta^4 + \frac{1}{48}\mu\delta^6 - 0.00091\,2164\delta^7 - 0.00484\,484\mu\delta^8 \\ + 0.00023\,460\delta^9 + 0.00112\,25\mu\delta^{10} - 0.00005\,17\delta^{11} - 0.00026\,2\mu\delta^{12} + \dots$

The formulae in this section have been obtained by the method of §2, using the relations (3.1) and (3.8) to reduce the truncation error (2.14).

7. *Other interpolation polynomials.* The polynomials we have just obtained are not the only kind which can be developed and we now examine briefly two other possibilities.

(i) *Interpolation polynomials for the range $-\frac{1}{2} \leq p \leq \frac{1}{2}$.* These can be derived from the Chebyshev expansion

$$f(a + ph) = \beta_0 + 2\beta_1 T_1(2p) + 2\beta_2 T_2(2p) + \dots \quad \left(-\frac{1}{2} \leq p \leq \frac{1}{2}\right). \tag{7.1}$$

Expressions for the coefficients β_s have been given by Miller (4), in whose notation β_s is replaced by α_s and $T_s(2p)$ by $\frac{1}{2} C_s(4\theta)$.

If the series (7.1) is truncated at the term $2\beta_n T_n(2p)$ and rearranged in powers of p , polynomials of the form

$$f(a + ph) = d_0 + d_1 p + d_2 p^2 + \dots + d_n p^n + \zeta(p) \tag{7.2}$$

(cf. (1.2)) are obtained, $\zeta(p)$ being the truncation error. The coefficient d_0 differs slightly from f_0 , but by adjusting d_1, d_2, \dots, d_n we can, if desired, replace d_0 by f_0 (cf. §§ 2, 3).

These polynomials have two advantages compared with those of § 6. First, the formulae for the coefficients d_s are simpler in that mixtures of odd and even differences do not occur. Secondly, no special precautions need be taken with the rounding of the coefficients. If the coefficients are computed retaining a guarding figure and rounded in the ordinary way, then because $|p| < \frac{1}{2}$ the maximum rounding error in an interpolate is

$$\frac{1}{2} + \frac{1}{2^2} + \frac{1}{2^3} + \dots + \frac{1}{2^{n+1}} < 1. \tag{7.3}$$

Thus if the truncation error $\zeta(p)$ does not exceed half a unit, the total error in an unrounded interpolate will not exceed $1\frac{1}{2}$ units.

Both of these advantages, however, affect only the *compiler* of the tables. From the standpoint of the *user* the polynomials† for the range $-\frac{1}{2} \leq p \leq \frac{1}{2}$ are inferior because the interpolation phase p can be negative. When this happens it complicates the evaluation of the polynomial and becomes a possible source of mistakes.

The use of negative values of p can be circumvented by computing the function f and the coefficients d_s at points half-way between the tabular arguments and rearranging the polynomials obtained in powers of $p' \equiv p + \frac{1}{2}$. If adequate guarding figures are retained in the computations then the polynomials obtained in this way are identical with those given by (1.2). It may be thought at first that if the coefficients d_s are rounded *before* rearranging in powers of $p + \frac{1}{2}$, then the rounding error in the rearranged polynomial would be given by (7.3). This is not the case, however, because in the process of rearrangement the values of d_s are multiplied by fractional numbers and so further roundings are necessary, with the consequent introduction of additional error. The avoidance of these extra rounding errors by the retention of a guarding figure in the final coefficients would be uneconomic.

(ii) *Everett-type formulae.* Various formulae of the type

$$f(a + ph) = q\{f_0 + \chi(q^2)\} + p\{f_1 + \chi(p^2)\},$$

where $q = 1 - p$ and $\chi(x^2)$ is a polynomial in x^2 , can be obtained by economization of the power series in Everett's interpolation formula

$$f(a + ph) = qf_0 + \sum_{s=1}^{\infty} \binom{q+s}{2s+1} \delta_0^{2s} + pf_1 + \sum_{s=1}^{\infty} \binom{p+s}{2s+1} \delta_1^{2s}.$$

Typical among such formulae is the following quintic:

$$f(a + ph) = q\{f_0 + (1 - q^2)(d_{2,0} + d_{4,0}q^2)\} + p\{f_1 + (1 - p^2)(d_{2,1} + d_{4,1}p^2)\}, \tag{7.4}$$

where
$$\left. \begin{aligned} d_2 &= -\frac{1}{6}\delta^2 + \frac{1}{30}\delta^4 - 0.0070890\delta^6 + 0.00157\delta^8 - 0.0004\delta^{10} + \dots, \\ d_4 &= -\frac{1}{120}\delta^4 + 0.0023189\delta^6 - 0.00057\delta^8 + 0.0001\delta^{10} - \dots, \end{aligned} \right\} \tag{7.5}$$

† Expansions corresponding to those of § 6 for the coefficients of these polynomials in terms of differences have been computed by the writers. It was not until a convenient method (§ 4) had been devised for reducing the rounding error associated with the polynomials of § 6, that the use of polynomials for the range $-\frac{1}{2} \leq p \leq \frac{1}{2}$ was rejected.

and the suffixes 0, 1 indicate that these functions are to be evaluated at the points a , $a + h$ respectively. This formula is valid† if $|\mu\delta_{\frac{1}{2}}^6| < 3\,00000$ and $|\delta_{\frac{1}{2}}^7| < 27000$, under which restrictions the truncation error will not exceed half a unit in the last decimal place retained.

The attractive feature of (7.4) compared with the interpolation quintic

$$f(a + ph) = f_0 + a_1p + a_2p^2 + a_3p^3 + a_4p^4 + a_5p^5, \tag{7.6}$$

of §6, which is valid under roughly the same conditions, is that for the same argument interval it requires half as much space in a printed table; only two coefficients d_2, d_4 need be given in addition to f , compared with five in the case of (7.6).

The disadvantage of (7.4) compared with (7.6) is that because of its greater complexity it requires twice as much labour to compute. A tabular comparison of the number of operations involved is as follows:

	Recordings	Multiplications	Settings	Transfers
Formula (7.4)	3	10	7	7
Formula (7.6)	1	5	6	4
Quartic	1	4	5	3

It is assumed that in the evaluation of (7.4) the quantities $1 - q^2$ and $1 - p^2$ are calculated as $p(1 + q)$ and $q(1 + p)$ respectively, and that q and $f_0 + p(1 + q)$ ($d_{2,0} + d_{4,0}q^2$) are recorded as well as the answer.

The last row of this table refers to the quartic polynomial of the type (7.6). This comparison is not irrelevant because after the cubic formula of the type (7.4) the next stage of approximation is a quintic and advantage cannot be taken of the cases in which a quartic would suffice.

The writers believe that if the extra space in a table required by formula (7.6) cannot be spared, then rather than use (7.4) it may be preferable to double the tabular interval, using if necessary an interpolation polynomial of the type (7.6) and of degree one higher.

8. *Evaluation of derivatives.* A useful property of the interpolation polynomials is that they can, with care, be used to evaluate the derivative at any point quite quickly. Differentiating (6.1), we obtain

$$hf'(a + ph) = a_1 + 2a_2p + 3a_3p^2 + \dots + na_n p^{n-1} + \epsilon'(p). \tag{8.1}$$

The polynomial part of the right-hand side of this equation can be evaluated on a calculating machine without any intermediate recording, by arranging it in the form

$$hf'(a + ph) = [\{na_n p + (n - 1)a_{n-1}\}p + (n - 2)a_{n-2}\}p + \dots \tag{8.2}$$

The error $\epsilon'(p)$ in (8.1), however, may be considerably larger than $\epsilon(p)$ itself. From (5.1) we have

$$\epsilon'(p) \doteq \frac{\delta_{\frac{1}{2}}^{n+1} \psi'(p)}{2^{2n+1}(n + 1)!}, \tag{8.3}$$

† In producing (7.4) a device similar to that described in §§2 and 3 has been used. The formula can also be obtained by rearrangement of (9.1).

$\delta_{\frac{1}{2}}^{n+1}$ being replaced by $\mu\delta_{\frac{1}{2}}^{n+1}$ if n is odd. From (3.6), putting $s = n + 1$ and differentiating, we find

$$\psi'(p) = 2k \cos \left\{ \frac{\pi}{2(n+1)} \right\} T'_{n+1} \left\{ (2p-1) \cos \frac{\pi}{2(n+1)} \right\}. \tag{8.4}$$

The maximum value of $|\psi'(p)|$ in $0 \leq p \leq 1$ occurs at $p = 0$ and 1 , and is equal to

$$2k(n+1) \cot \frac{\pi}{2(n+1)} \doteq \frac{4}{\pi} k(n+1)^2. \tag{8.5}$$

Hence

$$\frac{|\epsilon'(p)|_{\max.}}{|\epsilon(p)|_{\max.}} \doteq \frac{4}{\pi} (n+1)^2. \tag{8.6}$$

Thus the value of $|\epsilon'(p)|$ may be quite large. However, if only a limited accuracy in the derivative is required, as, for example, in inverse interpolation (§9), formula (8.1) may be quite useful.

In applying (8.6) it need not be assumed that $|\epsilon(p)|_{\max.} = \frac{1}{2}$. For example, in the table from which Table 2 of the next section (§9) is an extract, it is known that $|\epsilon(p)|_{\max.} < 0.05$ when $0 < x < 2.2$. Thus in this range $|\epsilon'(p)|_{\max.} < 1.6$ and the values of a_1 are equal to $-0.1 \text{ Ai}'(-x)$ to within a unit or two of the eighth decimal.

9. *Examples and time tests.* Table 2 is a typical extract from an eight-decimal table of the function $\text{Ai}(-x)$ at interval 0.1 in x , which gives the coefficients a_1, a_2, a_3, a_4 of the interpolation polynomial (6.1) with $n = 4$. They were computed by means of the formulae given in §6.

Table 2

x	$\text{Ai}(-x)$	a_1	a_2	a_3	a_4
2.0	+0.22740743	-6182590	-227414	+16836	+876
2.1	+0.16348451	-6583406	-171670	+20351	+812

We calculate from this table the value of $\text{Ai}(-x)$ for $x = 2.09439510$. Taking $p = 0.9439510$ and working in units of the eighth decimal, we obtain

$$\begin{aligned} \text{Ai}(-2.09439510) &= [\{ (876 \times 0.9439510 + 16836) \times 0.9439510 - 227414 \} \times 0.9439510 \\ &\quad - 6182590] \times 0.9439510 + 22740743 \\ &= +0.16716902. \end{aligned}$$

As an example of inverse interpolation we calculate the value of x in the range $2.0 < x < 2.1$ for which $\text{Ai}(-x) = 0.2$. In effect we have to solve the polynomial equation

$$a_1 p + a_2 p^2 + a_3 p^3 + a_4 p^4 = -0.02740743.$$

As a first approximation, we have

$$p = p_1 = -0.02741/a_1 = 0.4433,$$

and as a second approximation,

$$p = p_2 = p_1 - (a_2/a_1) p_1^2 = 0.4361.$$

Taking this value, we find that

$$\begin{aligned} \text{Ai}(-2.04361) &= [\{ (876 \times 0.4361 + 16836) \times 0.4361 - 227414 \} \times 0.4361 - 6182590] \\ &\quad \times 0.4361 + 22740743 \\ &= 0.20002693, \end{aligned}$$

and, using (8.2),

$$\begin{aligned}
 -0.1 \text{Ai}'(-2.04361) &= \{(4 \times 876 \times 0.4361 + 3 \times 16\,836) \times 0.4361 - 2 \times 227\,414\} \\
 &\qquad \qquad \qquad \times 0.4361 - 6182\,590 \\
 &= -0.06371\,0.
 \end{aligned}$$

The final approximation is obtained by application of Newton's rule, which yields

$$p = p_2 + (0.20002\,693 - 0.2)/0.06371\,0 = 0.43652\,27.$$

As a check we interpolate again and find $\text{Ai}(-2.04365\,227) = 0.20000\,000_1$. Thus to eight decimals,

$$x = 2.04365\,227.$$

Time tests. In order to assess the merit of the present method of interpolation, time tests have been carried out by a number of computers. Ten (in some cases more) interpolations were performed using the table from which Table 2 above is an extract, and they were then repeated using tables of which Tables 3, 4 and 5 are extracts. The

Table 3

x	$\text{Ai}(-x)$	δ^2
2.09	+0.17005055	-3554
2.10	+0.16348451	-3433

Table 4

x	$\text{Ai}(-x)$	δ_m^2	γ^4
2.0	+0.22740743	-456984	+22
2.1	+0.16348451	-345378	+20

Table 5

x	$\text{Ai}(-x)$	d_2	d_4
2.0	+0.22740743	+76222	-180
2.1	+0.16348451	+57617	-171

speeds of the methods were compared by dividing by the time taken using Table 2 and afterwards forming the means (with respect to all the computers); these are as follows:

Table 2	Table 3	Table 4	Table 5	Lagrange
1.0	1.0	(i) 2.1 (ii) 1.8	2.3	3.3

Although the actual speeds varied from one computer to another, the time ratios did not deviate much from the means.

Details of the ways in which each table was interpolated have an important bearing on the speeds and we now describe them briefly.

Table 2. It may be thought that a transfer calculating machine would be the most suitable for evaluating the interpolation polynomial. Various machines were tried, and the quickest was usually found to be a hand machine with rapid keyboard setting but *without* a mechanical transfer.

Table 3. This is an extract from a published table (1) at interval 0.01 in x . Interpolation was performed using Bessel's formula

$$f(a + ph) = (1 - p)f_0 + pf_1 + B''(\delta_0^2 + \delta_1^2) + B'''\delta_{\frac{1}{2}}^3.$$

The value of $\delta_{\frac{3}{2}}^3$ was formed mentally, and B'' , B''' were obtained from (3). In most interpolations the term $B''\delta_{\frac{3}{2}}^3$ could be neglected.

Table 4. This is at an interval 0.1 in x and gives the modified second difference δ_m^2 and fourth-difference correction γ^4 for use in the modified Everett formula ((1), p. B7)

$$f(a + ph) = (1 - p)f_0 + pf_1 + E_0^2\delta_{m0}^2 + E_1^2\delta_{m1}^2 + M_0^4\gamma_0^4 + M_1^4\gamma_1^4. \quad (9.1)$$

(i) No satisfactory single table of all the coefficients E_0^2 , E_1^2 , M_0^4 , M_1^4 exists having an interval 0.001 in p , but by suitable pretabulation near the required values it was possible to carry out the time tests on the assumption that such a table was at hand. The actual values of p used had more than four decimals. In consequence, formula (9.1) was applied twice, using the two tabular entries nearest to the value of p in question, and the results interpolated linearly to give the correct answer. This is a little quicker than the alternative process of interpolating the interpolation coefficients before applying (9.1).

(ii) The tests were repeated on the assumption that a table of E_0^2 , E_1^2 , M_0^4 and M_1^4 having an interval 0.0001 in the argument p was available. At this interval no interpolation in M_0^4 and M_1^4 is required, and the quickest way of evaluating (9.1) is to interpolate and record E_0^2 and E_1^2 first.

Table 5. This gives the coefficients d_2 , d_4 for use with the Everett-type polynomial (7.4). The evaluation was performed in the manner suggested in §7.

Lagrange's method. This was applied to tabular values of $Ai(-x)$ at interval 0.1 in x and the eight-point formula was used. The values of the Lagrangian coefficients were extracted from (6), which is a table at an interval 0.001 in p . The two nearest tabular entries were used and the required result then obtained by linear interpolation.

Examining the results of the time tests, we notice that it takes no longer to carry out an interpolation in Table 2 than it does in Table 3, which has an argument interval one-tenth the size. Allowing for the fact that the rows in the former table are twice as long as those of the latter, we see that by using economized interpolation polynomials instead of second differences, the table of $Ai(-x)$ from which Table 3 is an extract could be reduced to one-fifth the size without detriment to the user.

Tables 4 and 5, it may be noted, take up even less space than Table 2. But the cost of the saved space is to double the labour required for an interpolation. Moreover, because of the increased number of arithmetical operations involved, the risk of making a blunder in the interpolation is magnified. Similar remarks apply to the use of Lagrange's formula.

The saving of space and the provision of convenient interpolation facilities are mutually incompatible requirements between which the compiler of mathematical tables must effect a compromise. In tables which are not linearly interpolable the usual modern practice is to provide differences or modified differences for use with Everett's formula, regarded as so much more convenient than that of Lagrange that the increased size of the table is tolerated. The writers consider economized polynomials to be even more convenient than Everett's formula and, if the same interval is used, the extra size can still be tolerated; space can be saved, however, by increasing both the degree of the interpolation polynomial and the interval of tabulation, with little extra inconvenience to the user.

The work described above has been carried out as part of the research programme of the National Physical Laboratory, and this paper is published by permission of the Director of the Laboratory.

REFERENCES

- (1) *British Association mathematical tables*, part-vol. B, *The Airy integral* (Cambridge, 1946).
- (2) HERGET, P. and CLEMENCE, G. M. Optimum-interval punched-card tables. *Math. Tab., Wash.*, 1 (1943-5), 173-6.
- (3) *Interpolation and allied tables* (H.M. Stationery Office, 1936).
- (4) MILLER, J. C. P. Two numerical applications of Chebyshev polynomials. *Proc. roy. Soc. Edinb.* 62 (1943-9), 204-10.
- (5) *National Bureau of Standards, Applied Mathematics Series*, 9. *Tables of Chebyshev polynomials $S_n(x)$ and $C_n(x)$* (Washington, 1952).
- (6) *National Bureau of Standards, Mathematical Tables Project. Tables of Lagrangian interpolation coefficients* (New York, 1944).
- (7) SADLER, D. H. Maximum-interval tables. *Math. Tab., Wash.*, 4 (1950), 129-32.

NATIONAL PHYSICAL LABORATORY
TEDDINGTON, MIDDLESEX