A Survey of Methods of Computing Minimax and Near-Minimax Polynomial Approximations for Functions of a Single Independent Variable

W. FRASER

University of Toronto, Canada*

Abstract. Methods are described for the derivation of minimax and near-minimax polynomial approximations. For minimax approximations techniques are considered for both analytically defined functions and functions defined by a table of values. For near-minimax approximations methods of determining the coefficients of the Fourier-Chebyshev expansion are first described. These consist of the rearrangement of the coefficients of a power polynomial, and also direct determination of the coefficients from the integral which defines them, or the differential equation which defines the function. Finally there is given a convenient modification of an interpolation scheme which finds coefficients of a near-minimax approximation without requiring numerical integration or the numerical solution of a system of equations.

1. *Introduction*

The determination of approximations to continuous functions for use in computer programs is affected by several factors such as the number of independent variables on which the function depends, the method of its definition (analytic function or table of values), the form of the approximating function, and the norm or criterion used to determine the suitability of a specific approximation.

The minimax polynomial approximation of degree n to a continuous function $f(x)$ over an interval (a, b) is defined to be the polynomial $P_n^*(x)$ such that

$$E^* = \max_{a \le x \le b} | P_n^*(x) - f(x)| \le \max_{a \le x \le b} | P_n(x) - f(x)|, \tag{1.1}$$

where $P_n(x)$ is any polynomial of degree n. On the other hand any polynomial $Q_n(x) \not\equiv P_n^*(x)$ for which

$$\epsilon = \max_{a \le x \le b} | P_n^*(x) - Q_n(x)| \tag{1.2}$$

is sufficiently small is called a near-minimax approximation to $f(x)$ over (a, b). Since $\max_{a \le x \le b} | Q_n(x) - f(x)| \le E^* + \epsilon$, it follows that if ϵ/E^* is small enough, for example $1/10$ or less, $Q_n(x)$ may for computing purposes be considered as virtually indistinguishable from $P_n^*(x)$.

Programs based on minimax approximations permit a guarantee of the smallest possible maximum error attainable by polynomials of the given degree or less under random sampling of the interval (a, b), and this is one reason for which their study is of computational interest. However, it requires considerable effort to determine a minimax approximation of a given degree for a given interval (a, b), and the entire calculation must be repeated if either the degree or the interval (a, b) is changed. Furthermore, if it is required to know the smallest degree which permits a stated least maximum error, this cannot be found except by going completely through the calculation. On the other hand certain near-minimax approximations are relatively

* Institute of Computer Science.

Journal of the Association for Computing Machinery, Vol. 12, No. 3 (July, 1965), pp. 295–314

easily obtained, and these require only a slight recalculation for a change of degree or range. At the same time they provide a close a priori estimate of the maximum error associated with their use, this error being as a rule close to the minimax error.

In the following paragraphs attention is restricted to minimax and near-minimax polynomial approximations to a function of a single independent variable, defined either as an analytic function or by a table of values.

The linear transformation $t = (2x - \overline{b+a})/(b - a)$ takes the interval (a, b) into the interval $(-1, 1)$, and we assume throughout that this has been done.

The difference $P_n(x) - f(x)$ is referred to as the absolute error of the approximation $P_n(x)$, while either $\{P_n(x) - f(x)\}/f(x)$ or $\ln |P_n(x)/f(x)|$ is referred to as its relative error. The minimax approximation is dependent on the definition of error used and only approximations which minimize the absolute error are considered since no large-scale modifications are required if relative rather than absolute error is used.

In the following section some properties of minimax approximations are given. Also, since the Chebyshev polynomials play such an important role in the approximation problem, a development of some of their properties is included. However, proofs of theorems are not in general presented; instead references are given to places where these proofs may be found.

2. *Minimax Approximations and Chebyshev Polynomials*

It will be taken for granted that the polynomial $P_n^*(x)$ of (1.1) actually exists; however, for a proof of this see Achieser [1956] or de la Vallée Poussin [1952]. In these references it is also shown that if $E_n^*(x) = P_n^*(x) - f(x)$, there must exist a set of at least $(n+2)$ points $-1 \leq x_0 < x_1 < \cdots < x_{n+1} \leq 1$ at which the extreme value $\pm E_n^*$ is achieved with alternating sign from point to point, i.e., $E_n^*(x_i) = \pm(-1)^i E_n^*$, and $|E_n^*(x)| \leq E_n^*, \quad -1 \leq x \leq 1$. Furthermore, if corresponding to a polynomial $Q_n(x)$ the function $E_n(x) = Q_n(x) - f(x)$ possesses $(n+2)$ or more extremes $\pm E_n$ of equal size and alternating sign, where $|E_n(x)| \leq E_n$, $-1 \leq x \leq 1$, $Q_n(x)$ must be the minimax approximation of degree n to $f(x)$. Hence the existence of at least $(n+2)$ extremes of the error $E_n(x) = P_n(x) - f(x)$ with alternating sign is both a necessary and sufficient condition for the polynomial $P_n(x)$ to be a minimax approximation to $f(x)$.

The above property which characterizes a minimax approximation can be used to establish its uniqueness. For if $P_n(x)$ and $Q_n(x)$ are two minimax approximations to $f(x)$ so is their arithmetic mean $R_n(x)$. From the relation $R_n(x) - f(x) = \frac{1}{2}[P_n(x) - f(x)] + \frac{1}{2}[Q_n(x) - f(x)]$ it follows that $P_n(x)$ and $Q_n(x)$ must have the same values in at least $(n+2)$ points, and hence must be identical.

There are various ways of attempting to find a polynomial $P_n(x)$ which approximates a given continuous $f(x)$. One approach is to choose $(n+1)$ points in $(-1, 1)$ and form the Lagrange interpolation polynomial which has these points as nodes, i.e., points at which the function and polynomial have the same value. If the nodes are the set of points $\{x_i\}$, where $-1 \leq x_0 < x_1 \cdots < x_n \leq 1$, the remainder term $R_n(x)$, defined as $R_n(x) = f(x) - P_n(x)$, is given by Kopal [1955] to be

$$R_n(x) = \frac{(x - x_0)(x - x_1) \cdots (x - x_n)}{(n + 1)!} f^{(n+1)}(\xi) \qquad (-1 < \xi < 1). \tag{2.1}$$

In making an estimate of the maximum error connected with approximating $f(x)$ by $P_n(x)$, it is first noted that the term $(n+1)!$ depends only on the degree of the approximating polynomial. Furthermore the value of ξ in (2.1) is in general not known and $|f^{(n+1)}(\xi)|$ is usually replaced by its upper bound and hence is used so as to depend only on the function being approximated and the degree of the approximating polynomial. Thus the estimate of the maximum error will be the smallest if the nodes of the interpolation are chosen so that

$$\max_{-1\leq x\leq 1} |(x - x_0)(x - x_1) \cdots (x - x_n)| = \max_{-1\leq x\leq 1} |x^{n+1} - \cdots |$$

is the smallest possible. Thus the problem of determining an interpolating polynomial of degree n with the least maximum error estimate is closely connected with the problem of finding the polynomial of degree $(n+1)$ with leading coefficient 1 which will have the smallest maximum modulus in $(-1, 1)$.

To find the polynomial $P_n(x)$ of degree n with leading coefficient 1 and smallest maximum modulus in $(-1, 1)$ is equivalent to finding the minimax approximation to x^n by polynomials of degree $(n - 1)$. The polynomial $P_n(x)$ is characterized in this case by assuming its extreme values $\pm E_n$ in exactly $(n+1)$ points, $-1 \leq x_0 < \cdots < x_n \leq 1$, and can be found by the following elementary procedure. In the first place two of the extremes must be at end points of the interval, since at an interior extreme the derivative, a polynomial of degree $(n-1)$, must be zero. Consider next the polynomial of degree at most $(2n-1)$,

$$S_{2n-1}(x) = [P_n^2(x) + ((1 - x^2)\{P_n'(x)\}^2)/n^2] - E_n^2 \tag{2.2}$$

whose derivative is

$$S_{2n-1}'(x) = 2P_n'(x)[P_n(x) - \{xP_n'(x)\}/n^2 + \{(1 - x^2)P_n''(x)\}/n^2]. \tag{2.3}$$

The polynomial $S_{2n-1}(x)$ is zero at each extreme of $P_n(x)$ and by (2.3) $S_{2n-1}'(x)$ is zero at a zero of $P_n'(x)$. Thus $S_{2n-1}(x)$ has $(n+1)$ zeros corresponding to each of the extremes of $P_n(x)$, and each of the $(n - 1)$ interior extremes is a double zero. Thus $S_{2n-1}(x)$ must be identically zero. Simplifying (2.3) and clearing of fractions leads to

$$(1 - x^2)P_n'' - xP_n' + n^2P_n = 0. \tag{2.4}$$

Writing $x = \cos\theta$, (2.4) transforms into

$$\frac{d^2P_n}{d\theta^2} + n^2P_n = 0, \tag{2.5}$$

whose solution is

$$P_n = A_n \cos n\theta + B_n \sin n\theta. \tag{2.6}$$

Extremes of P_n occur at $x = 1$ and -1, corresponding respectively to $\theta = 0$ and π. Thus, in order to have an extreme of P_n given by (2.6) at $\theta = 0$, it follows that we must have $B_n = 0$. Since the extreme has been referred to previously as E_n, the solution of (2.4) is now written in the form

$$P_n = E_n \cos n\theta. \tag{2.7}$$

It is immediately obvious that $E_1 = 1$ and $E_2 = \frac{1}{2}$. Furthermore, since

$$P_{n+1}/E_{n+1} + P_{n-1}/E_{n-1} = 2xP_n/E_n, \tag{2.8}$$

it follows by equating the leading coefficients that $E_{n+1} = \frac{1}{2}E_n$, and hence by induction that $E_n = 2^{-(n-1)}$ $(n > 0)$. Thus the polynomial of degree n with leading coefficient 1 which has the smallest maximum modulus in $(-1, 1)$ is the polynomial $2^{-(n-1)} \cos n\theta$, where $x = \cos\theta$. It is customary to define the Chebyshev polynomial of degree n, $T_n(x)$, to be the polynomial P_n/E_n and we do this henceforth. Thus we define

$$T_n(x) = \cos n\theta \tag{2.9}$$

where $x = \cos\theta$ $(0 \le \theta \le \pi$, i.e., $-1 \le x \le 1)$.

The relation (2.8) rewritten in terms of Chebyshev polynomials becomes

$$T_{n+1}(x) = 2xT_n(x) - T_{|n-1|}(x). \tag{2.10}$$

Starting with $T_0(x) = 1$, $T_1(x) = x$, we can build up the Chebyshev polynomials using the recurrence relation (2.10).

Since the range (0, 1) is sometimes more convenient than $(-1, 1)$, it is sometimes used and on this range we define

$$T_n^*(x) = T_n(2x - 1) \qquad (0 \le x \le 1). \tag{2.11}$$

From this the useful relation $T_n^*(x^2) = T_{2n}(x)$ is easily verified.

Several properties of the Chebyshev polynomials are given in Clenshaw [1962]. Since they are needed in the following sections a few of these properties are stated below, without proof.

Upon integration we have

$$\int T_n(x)\,dx = \tfrac{1}{2}\{T_{n+1}(x)/(n+1) - T_{|n-1|}(x)/(n-1)\} \qquad (n \ge 0), \tag{2.12}$$

where it is understood that the second term in the bracket is to be omitted for $n = 1$. Also

$$\int_{-1}^{1} \frac{T_m(x)T_n(x)}{\sqrt{(1-x^2)}}\,dx = \begin{cases} \pi & (m = n = 0) \\ \frac{1}{2}\pi & (m = n \ne 0) \\ 0 & (m \ne n). \end{cases} \tag{2.13}$$

In addition to orthogonality as expressed by (2.13), the Chebyshev polynomials possess a property of orthogonality with respect to summation. Letting Σ' represent a sum in which the first term is to be halved, and Σ'' a sum in which first and last terms are to be halved, we have for $x_i = \cos i\pi/n$, where $n > 0$ and $j, k \le n$,

$$\sum_{i=0}^{n}{}'' \, T_j(x_i)T_k(x_i) = \begin{cases} n & (j = k = 0 \text{ or } n) \\ \frac{1}{2}n & (j = k \ne 0 \text{ or } n) \\ 0 & (j \ne k). \end{cases} \tag{2.14}$$

3. *Minimax Polynomial Approximations*

The minimax approximation $P_n^*(x)$ of degree n to a given continuous $f(x)$ has associated with it an error $E_n^*(x) = P_n^*(x) - f(x)$ which has at least $(n+2)$ extremes with alternation of sign from one to the next. However, in case $f(x)$ is either even or odd, the minimax approximation is either even or odd (Murnaghan and Wrench, [1960]), and the error $E_n^*(x)$ has in both these cases at least $(n+3)$ extremes with alternating sign. Many routines for finding minimax approximations assume the existence of a standard situation, i.e., exactly $(n+2)$ extremes of the error curve, and when the number of extremes exceeds this, computing embarrassment is possible. For even and odd functions it is possible to compensate by computing the polynomial of degree one higher than that actually wanted, and of course the leading coefficient then turns out to be zero. For odd functions $f(x)$ it is common to approximate $f(x)/x$, thereby preserving relative accuracy near the origin.

While it simplifies the problem to assume that the end points of the interval are among the set of extremes of $E_n^*(x)$, it is not in general possible to do so. Examples of minimax approximations are known for which the absolute value of the error at each end of the interval is less than the maximum value of the error over the whole interval, and also there are examples in which this is true at just one end of the interval. However, for a function $f(x)$ whose $(n+1)$-st derivative has no zeros inside the interval of approximation, the end points must be extremes of $E_n^*(x)$. This follows from the fact that a repeated application of Rolle's Theorem shows it to be impossible for $E_n^{*\prime}(x)$ to have more than n zeros inside the interval.

If extremes $\pm E^*$ of the error of the minimax approximation $P_n^*(x) = \sum_{j=0}^n b_{nj}^* x^j$ occur with alternating signs at the points $-1 \le x_0^* < x_1^* < \cdots < x_{n+1}^* \le 1$, the coefficients $\{b_{nj}^*\}$, the points $\{x_i^*\}$ and the minimax error E^* must satisfy

$$\sum_{j=0}^{n} b_{nj}^* x_i^{*j} - f(x_i^*) = (-1)^i E^* \qquad (\mathrm{i} = 0, 1, \cdots, n+1) \quad (3.1)$$

and also

$$\sum_{j=0}^{n} b_{nj}^* x_i^{*j} - f(x_i^*) = \text{extreme}. \quad (3.2)$$

In the event that $f(x)$ is differentiable in addition to being continuous, the equations (3.2) may be replaced by

$$\sum_{j=1}^{n} b_{nj}^* j (x_i^*)^{j-1} - f'(x_i^*) = 0 \qquad (-1 < x_i^* < 1). \quad (3.3)$$

The condition at the right side of (3.3) expresses the fact that x_i^* must be an interior extreme.

The usual method for solving the system (3.1) and (3.2) or (3.3) is to choose a set of points $\{x_i\}$ as a first estimate of the set $\{x_i^*\}$, and substitute into (3.1) which then becomes a linear system for the set $\{b_{nj}\}$ and E, first estimates respectively of $\{b_{nj}^*\}$ and E^*. Using the $\{b_{nj}\}$ as coefficients of $P_n(x)$, the first estimate of $P_n^*(x)$, the system (3.2) or (3.3) is used to calculate the set of points $\{m_i\}$ at which $E(x) = P_n(x) - f(x)$ has its extremes $\{E_i\}$. This completes a cycle of the calculation. To initiate the next cycle the set $\{x_i\}$ is chosen to be the set $\{m_i\}$ determined in the previous cycle. It is shown by Novodvorskii and Pinsker [1951] that this procedure is convergent regardless of the initial choice $\{x_i\}$. Also it is shown by Veidinger [1960] that the rate of convergence of the set (3.1) and (3.3) is quadratic, i.e., if the estimate of E^* in the kth cycle is written as $E^{(k)}$, then $E^* - E^{(k+1)} = O(E^* - E^{(k)})^2$.

The accuracy of the coefficients of the approximation and the location of the extremes can always be improved by further iteration. However it is usually desirable to terminate the calculation when $E_M = \max_i | E_i |$ is less than some preassigned multiple of $E_m = \min_i | E_i |$, for example 1.05.

It follows from (3.1) and (3.3) that the partial derivatives of E^* and the coefficients $\{b_{nj}^*\}$ with respect to an interior extreme are zero. Hence the coefficients of the minimax polynomial are not too sensitive to small errors in determination of the location of the interior extremes. While convergence is independent of the initial choice of points $\{x_i\}$, it is usually found advantageous, as suggested by Section 2, to take these as the points at which $| T_{n+1}(x) | = 1$.

Example 1 which follows is a sample calculation of the mimimax approximation to a continuous function.

Example 1. Find the minimax polynomial approximation of degree 2 to $1/(1 + x)$ over $0 \leq x \leq 1$; i.e., find $P_2(x) = b_0 + b_1x + b_2x^2$ to approximate $1/(1 + x)$ for $0 \leq x \leq 1$ with the smallest maximum error.

To initiate the calculation we take (arbitrarily) the set $\{x_i\}$ to be the points (0, 0.3, 0.6, 1.0). The equations (3.1) may be written

$$\begin{aligned} b_0 \qquad\qquad\qquad\quad - E &= 1 &&= 1.00000\ 00000 \\ b_0 + 0.3b_1 + 0.09b_2 + E &= 1/1.3 &&= 0.76923\ 07692 \\ b_0 + 0.6b_1 + 0.36b_2 - E &= 1/1.6 &&= 0.62500\ 00000 \\ b_0 + \quad b_1 + \qquad b_2 + E &= 1/2 &&= 0.50000\ 00000. \end{aligned}$$

The solution of this system is $b_0 = 0.99381\ 86813$, $b_1 = -0.83104\ 39561$, $b_2 = 0.34340\ 65935$ and $E = -0.00618\ 13187$. Using these values for b_0, b_1, b_2, the first cycle of the calculation is completed by finding the extremes of $P_2(x) - 1/(1 + x)$ in $0 \leq x \leq 1$. In this case Newton's method may be used to find the zeros of $P_2'(x) + 1/(1 + x)^2 = 0$. The results of the calculation for the first and remaining cycles are given in Table I. Since $| E_M | < 1.05| E_M |$ in cycle 2, the calculation was terminated at the end of the cycle.

Next consider the case in which $f(x)$ is defined by a table of values rather than by an analytic formula; i.e., $f(x)$ is defined for the set of m points $\{x_i\}$ $(i = 1, 2, \cdots, m)$. Corresponding to any subset of $(n+2)$ points $x_{i_1} < x_{i_2} < \cdots < x_{i_{n+2}}$ a polynomial $P_n(x)$ and a number E can be found such that

$$P_n(x_{i_k}) - f(x_{i_k}) = (-1)^k E \qquad (k = 1, 2, \cdots, n+2). \quad (3.4)$$

TABLE I

Cycle	i	x_i	b_i	m_i	E_i
1	0	0.0	.99381 86813	.0	−.00618 13187
	1	0.3	−.83104 39561	.20162 93457	.00801 34415
$E = -.00618\ 13187$	2	0.6	.34340 65935	.71492 47108	−.00791 00913
	3	1.0		1.0	.00618 13187
2	0	0.0	.99264 57295	.0	−.00735 42705
	1	0.20162 93457	−.82844 47342	.20707 46091	.00736 10312
$E = -.00735\ 42705$	2	0.71492 47108	.34315 32752	.70713 12972	−.00736 29601
	3	1.0		1.0	.00735 42705

Vallée Poussin [1911] shows that the minimax approximation $P_n^*(x)$ to $f(x)$ over the set $\{x_i\}$ is that obtained by using the subset of $(n+2)$ points which provides the largest possible absolute value for the solution E of the system (3.4). The "Exchange" algorithm used to determine this polynomial is described and a proof of convergence given by Stiefel [1959].

The computational technique for determining $P_n^*(x)$ is similar to that employed in the case of a continuously defined function. A subset of $(n+2)$ points $\{x_{i_k}\}$ is chosen from the m points $\{x_i\}$, and the system (3.4) is solved. These points may be chosen initially so that the subscripts $\{i_k\}$ form as nearly as possible an arithmetic progression, including 1 and m. After solving (3.4) the residuals $r_i = P_n(x_i) - f(x_i)$ are evaluated for $i = 1, 2, \cdots, m$. If no residual is numerically greater than $|E|$, the problem is finished. Otherwise at least one more cycle of the calculation is required. To initiate the next cycle the set $\{x_{i_k}\}$ is chosen so as to correspond to the $(n+2)$ largest residuals, consistent with the requirement of alternation in sign. In general this will imply that if a local extreme of the residuals, r_i, is found at a point x_i which is not a member of the set $\{x_{i_k}\}$ used to solve (3.4), the point x_i is then made to replace the nearest x_{i_k} which provided a residual of the same sign as r_i. In the event that there is an extreme of the residuals to the right of $x_{i_{n+2}}$, of opposite sign to $r_{i_{n+2}}$, the corresponding point is included in the set $\{x_{i_k}\}$ and x_{i_1} deleted if the residual at the point is greater in magnitude than $|r_{i_1}|$, otherwise the point is not used. A comparable procedure is followed if an extreme is found to the left of x_{i_1}.

A sample calculation of a minimax approximation for a function defined over a discrete set of points is given below.

Example 2. Find the minimax polynomial approximation of degree 3 over the interval $0 \leq x \leq 3$ for the function defined by the following table of values.

x	$f(x)$	x	$f(x)$
0.0	0.0	1.6	1.26491
0.2	0.44721	1.8	1.34164
0.4	0.63245	2.0	1.41421
0.6	0.77460	2.2	1.48324
0.8	0.89443	2.4	1.54919
1.0	1.00000	2.6	1.61245
1.2	1.09545	2.8	1.67332
1.4	1.18322	3.0	1.73205

With the first set of points $\{x_{i_k}\}$ chosen to be 0.0, 0.4, 1.6, 2.6 and 3.0, the equations to be solved are

$$\begin{aligned}
b_0 \qquad\qquad\qquad\qquad\qquad\qquad\qquad - E &= 0.0\\
b_0 + 0.4b_1 + 0.16b_2 + 0.064b_3 + E &= 0.63245\\
b_0 + 1.6b_1 + 2.56b_2 + 4.096b_3 - E &= 1.26491\\
b_0 + 2.6b_1 + 6.76b_2 + 17.576b_3 + E &= 1.61245\\
b_0 + 3.0b_1 + 9.00b_2 + 27.000b_3 - E &= 1.73205.
\end{aligned}$$

The solution of this system is:

$$b_0 = 0.0424\ 2319, \quad b_1 = 1.642\ 988, \quad b_2 = -0.7356\ 752,$$

$$b_3 = 0.1268\ 209, \quad E = 0.0424\ 2328.$$

TABLE II

Cycle 1					Cycle 2					Cycle 3				
k	i_k	x_{i_k}	b_k	E	k	i_k	x_{i_k}	b_k	E	k	i_k	x_{i_k}	b_k	E
0	0	0.0	.042423	.0424	0	0	0.0	.073021	.0730	0	0	0.0	.074503	.0745
1	2	0.4	1.642988		1	1	0.2	1.659285		1	1	0.2	1.642520	
2	8	1.6	−.735675		2	6	1.2	−.796250		2	5	1.0	−.786253	
3	13	2.6	.126821		3	12	2.4	.145202		3	12	2.4	.143732	
4	15	3.0			4	15	3.0			4	15	3.0		

Cycle 1				Cycle 2				Cycle 3			
x_i	r_i	x_i	r_i	x_i	r_i	x_i	r_i	x_i	r_i	x_i	r_i
0.0	.04242	1.6	.04242	0.0	.07302	1.6	.01931	0.0	.07450	1.6	.01354
0.2	−.10460	1.8	.01419	0.2	−.07302	1.8	−.01494	0.2	−.07450	1.8	−.01981
0.4	−.04242	2.0	−.01394	0.4	−.01382	2.0	−.04600	0.4	−.01754	2.0	−.04982
0.6	.01617	2.2	−.03652	0.6	.03871	2.2	−.06754	0.6	.03341	2.2	−.07020
0.8	.05648	2.4	−.04791	0.8	.07076	2.4	−.07302	0.8	.06448	2.4	−.07450
1.0	.07656	2.6	−.04242	1.0	.08126	2.6	−.05588	1.0	.07450	2.6	−.05623
1.2	.07833	2.8	−.01425	1.2	.07302	2.8	−.00944	1.2	.06624	2.8	−.00878
1.4	.06546	3.0	.04242	1.4	.05058	3.0	.07302	1.4	.04416	3.0	.07450

The first cycle of the calculation is completed by formation of a table of residuals, and selection of the subset of the $\{x_i\}$ corresponding to their extremes, to be used as the set $\{x_{i_k}\}$ to begin cycle 2. The results of the calculations for the three cycles required are given in Table II.

In the last cycle the extreme residuals are not in fact equal in magnitude to the full word length of the machine used for the calculations. This is not surprising since they are formed by a subtraction involving two numbers of comparable size. As before the calculation could be terminated when the ratio of the largest to the smallest extreme of the residuals is less than $1+\eta$ for some acceptable positive number η.

4. *Near-Minimax Approximations; Fourier-Chebyshev Expansions*

If $f(x)$ is continuous and of bounded variation in $(-1, 1)$ it will possess an expansion in terms of Chebyshev polynomials of the form

$$f(x) = \tfrac{1}{2}a_0 + a_1T_1(x) + \cdots = \sum_{k=0}^{\infty}{}' a_kT_k(x) \tag{4.1}$$

where, using (2.13), we have

$$a_k = \frac{2}{\pi}\int_{-1}^{1}\frac{f(x)T_k(x)}{\sqrt{(1-x^2)}}\,dx \qquad (k = 0, 1, 2, \cdots). \tag{4.2}$$

If the $|a_k|$ tend monotonically to zero with sufficient rapidity, then the approximation

$$A_n(x) = \sum_{k=0}^{n}{}' a_kT_k(x)$$

obtained by truncating (4.1) after the term $a_n T_n(x)$ will have an error dominated by the term $a_{n+1}T_{n+1}(x)$, the first omitted term. Since $T_{n+1}(x)$ alternately takes the extreme values 1 and -1 at $(n+2)$ points in $(-1, 1)$ it follows that the error of the truncated Chebyshev expansion will have a character resembling that of the minimax approximation, i.e., there will be $(n+2)$ extremes of the error which alternate in sign and have nearly equal magnitude. It follows that throughout the interval $(-1, 1)$ the value of the truncated Chebyshev expansion will be close to the value of the minimax approximation.

If the function $f(x)$ has a power series expansion

$$f(x) = \sum_{n=0}^{\infty} \alpha_n x^n, \tag{4.3}$$

its Chebyshev expansion may frequently be obtained by a simple rearrangement of terms, as shown by Lanczos [1956]. If we write

$$x^n = \sum_{j=0}^{n}{}' \rho_{nj} T_j(x) = \sum_{j=0}^{n}{}' \rho_{nj}^* T_j^*(x),$$

the coefficients ρ_{nj} and ρ_{nj}^* are given by Clenshaw [1962] to be

$$\left.\begin{aligned} \rho_{nj} &= 2^{-(n-1)} \binom{n}{(n-j)/2}, \qquad && \text{if } n-j \text{ is even,} \\ &= 0, && \text{if } n-j \text{ is odd,} \\ \rho_{nj}^* &= 2^{-(2n-1)} \binom{2n}{n-j}. && \end{aligned}\right\} \tag{4.4}$$

Substituting for x^n in (4.3), $f(x)$ can be rewritten in one or other of the forms

$$f(x) = \sum_{n=0}^{\infty}{}' a_n T_n(x) = \sum_{n=0}^{\infty}{}' a_n^* T_n^*(x), \tag{4.5}$$

where

$$a_n = \sum_{k=n}^{\infty} \alpha_k \rho_{kn}, \qquad a_n^* = \sum_{k=n}^{\infty} \alpha_k \rho_{kn}^*. \tag{4.6}$$

There are several difficulties which may arise in the evaluation of the a_n or a_n^* from (4.6). For example, if the α_k are not all of one sign there is a danger of loss of significant digits by cancellation. In this case double or higher precision must be used to achieve single precision results. The convergence of (4.6) is governed by the convergence of (4.3) for $x = 1$, implying that it will be difficult to find the a_n or a_n^* if (4.3) converges slowly for $x = 1$. However Thacher [1964] has shown that it is often possible in this case to use a series transformation such as Euler's or the e_k transformation (Shanks [1955], Wynn [1956]) for the summation of (4.6). The expressions for the a_n and the a_n^* may be thought of as matrix-vector multiplications, in which the matrix (ρ_{nj}) is infinite. A sample rearrangement is carried out in the following example.

Example 3. Rearrangement of a power series into a Chebyshev series. In the interval $-1 \leq x \leq 1$ the expansion $\sum_{n=0}^{\infty} \alpha_n x^n$ may be converted to the form

$\sum_{n=0}^{\infty\prime} a_n T_n(x)$ by the following matrix-vector multiplication:

$$\begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \\ \cdot \end{pmatrix} = \begin{pmatrix} 1 & 0 & 2 & 0 & 6 & 0 & \cdots \\ 0 & 1 & 0 & 3 & 0 & 10 & \cdots \\ 0 & 0 & 1 & 0 & 4 & 0 & \cdots \\ 0 & 0 & 0 & 1 & 0 & 5 & \cdots \\ 0 & 0 & 0 & 0 & 1 & 0 & \cdots \\ 0 & 0 & 0 & 0 & 0 & 1 & \cdots \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdots \end{pmatrix} \begin{pmatrix} 2\alpha_0 \\ \alpha_1 \\ \alpha_2/2 \\ \alpha_3/4 \\ \alpha_4/8 \\ \alpha_5/16 \\ \cdot \end{pmatrix}.$$

Referring to an element of the infinite matrix as d_{ij}, further columns can be added by writing

$$\begin{aligned} d_{1j} &= 2d_{2,j-1}, \\ d_{ij} &= d_{i-1,j-1} + d_{i+1,j-1} && (2 \le i \le j), \quad (4.7) \\ d_{ij} &= 0 && (i > j). \end{aligned}$$

The column vector on the right is extended by adding elements of the form $\alpha_n/2^{n-1}$. For example, given the function

$$\cosh x = 1 + x^2/2! + x^4/4! + x^6/6! + x^8/8! + \cdots \qquad (-\infty < x < \infty),$$

we have $2\alpha_0 = 2$, $\alpha_2/2 = .25$, $\alpha_4/8 = .00520\ 83333$, $\alpha_6/32 = .00004\ 34028$, $\alpha_8/128 = .00000\ 01938$ and $\alpha_{10}/512 = .00000\ 00005$. The above matrix-vector multiplication produces the results

$$\begin{aligned} a_0 &= 2.53213\ 17480, & a_2 &= .27149\ 53387, & a_4 &= .00547\ 42399, \\ a_6 &= .00004\ 49771, & a_8 &= .00000\ 01992, & a_{10} &= .00000\ 00005. \end{aligned}$$

On the other hand, in the interval $0 \le x \le 1$, the expansion $\sum_{n=0}^{\infty} \alpha_n x^n$ may be converted to the form $\sum_{n=0}^{\infty\prime} a_n^* T_n^*(x)$ by the following matrix-vector multiplication:

$$\begin{pmatrix} a_0^* \\ a_1^* \\ a_2^* \\ a_3^* \\ a_4^* \\ a_5^* \\ \cdot \end{pmatrix} = \begin{pmatrix} 1 & 2 & 6 & 20 & 70 & 252 & \cdots \\ 0 & 1 & 4 & 15 & 56 & 210 & \cdots \\ 0 & 0 & 1 & 6 & 28 & 120 & \cdots \\ 0 & 0 & 0 & 1 & 8 & 45 & \cdots \\ 0 & 0 & 0 & 0 & 1 & 10 & \cdots \\ 0 & 0 & 0 & 0 & 0 & 1 & \cdots \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdots \end{pmatrix} \begin{pmatrix} 2\alpha_0 \\ \alpha_1/2 \\ \alpha_2/8 \\ \alpha_3/32 \\ \alpha_4/128 \\ \alpha_5/512 \\ \cdot \end{pmatrix}.$$

Referring to an element of the matrix as d_{ij}, further columns can be added by writing

$$\begin{aligned} d_{1j} &= 2(d_{1,j-1} + d_{2,j-1}), \\ d_{ij} &= d_{i-1,j-1} + 2d_{i,j-1} + d_{i+1,j-1} && (2 \le i \le j), \quad (4.8) \\ d_{ij} &= 0 && (i > j). \end{aligned}$$

The column vector on the right is extended by adding elements of the form $\alpha_n/2^{2n-1}$.

There will be some cases in which the integrals defining the coefficients (4.2) in the Chebyshev expansion of a function are known and their values tabulated, or they may otherwise be easily obtained. In such a case the expansion (4.1) can be found directly, as illustrated in the following example.

Example 4. For the functions arc tan x and $1/x$ arc tan x find expansions of the form $\sum_{n=0}^{\prime\infty} a_n T_n(x) \quad (-1 \le x \le 1)$.

To begin with $f(x) = \text{arc tan } x$ we have, since $f(x)$ is odd, $a_{2n} = 0 \quad (n = 0, 1, \cdots)$, and

$$a_{2n+1} = \frac{2}{\pi} \int_{-1}^{1} \frac{(\text{arc tan } x) T_{2n+1}(x)}{\sqrt{(1 - x^2)}}\, dx.$$

Writing $x = \cos \theta$, and performing an integration by parts, the integral for a_{2n+1} becomes

$$a_{2n+1} = \frac{1}{(2n + 1)\pi} \int_{0}^{2\pi} \frac{\sin (2n + 1)\theta \sin \theta}{1 + \cos^2 \theta}\, d\theta.$$

Making the substitution $z = e^{i\theta}$, $dz = iz\, d\theta$, $\sin \theta = (z - 1/z)/2i$, $\sin (2n + 1)\theta = (z^{2n+1} - 1/z^{2n+1})/2i$, $\cos \theta = (z + 1/z)/2$, we derive

$$a_{2n+1} = \frac{1}{(2n + 1)\pi} \int_{|z|=1} \frac{(1/2i)(z^{2n+1} - 1/z^{2n+1})(1/2i)(z - 1/z)}{1 + \frac{1}{4}(z + 1/z)^2} \left(\frac{dz}{iz}\right).$$

This can be reduced to the form

$$a_{2n+1} = \frac{1}{(2n + 1)\pi} \int_{|z|=1} \frac{iz(z^{2n+1} - 1/z^{2n+1})(z - 1/z)}{(z - ip)(z + ip)(z - i/p)(z + i/p)}\, dz,$$

where $p = \sqrt{2} - 1$. In the unit circle there are poles of the integrand at $z = ip$ $z = -ip$ and $z = 0$. The residues at $z = ip$ and $z = -ip$ are equal and their common value is

$$\frac{(-1)^n(p^{4n+2} + 1)(p^2 + 1)i}{2(p^4 - 1)p^{2n}}.$$

To determine the residue of the pole at the origin we write the integrand in the form

$$\frac{i(z^{2n+1} - 1/z^{2n+1})(z^2 - 1)}{1 - p^4} \left(\frac{1}{(1 + z^2/p^2)} - \frac{p^4}{(1 + p^2/z^2)}\right)$$

and expand in powers of z. The coefficient of $1/z$ in this expansion is found to be

$$\frac{(-1)^n i(1 + p^2)(1 - p^{4n+2})}{(1 - p^4)p^{2n}}.$$

The sum of the residues at the poles thus turns out to be

$$\frac{(-1)^{n+1} i(1 + p^2)}{p^{2n}(1 - p^4)}\, 2p^{4n+2}.$$

Recalling that $p = \sqrt{2} - 1$, we have $p^2 = 3 - 2\sqrt{2}$, $p^4 = 17 - 12\sqrt{2}$, $1 + p^2 = 2\sqrt{2}p$, and $1 - p^4 = 4\sqrt{2}p^2$. Thus the coefficient a_{2n+1} is given by

$$a_{2n+1} = \frac{2(-1)^n p^{2n+1}}{(2n + 1)}.$$

Accordingly, with $p = \sqrt{2} - 1$, we have

$$\text{arc tan } x = 2\left[pT_1(x) - \frac{p^3}{3} T_3(x) + \frac{p^5}{5} T_5(x) - \frac{p^7}{7} T_7(x) + \cdots\right] \qquad (-1 \le x \le 1).$$

To find the expansion of $1/x$ arc tan x we write the relation (2.10) in the form $(1/x)T_{n+1}(x) = 2T_n(x) - (1/x)T_{|n-1|}(x)$, and divide the above expression for arc tan x by x. This leads to

$$\frac{1}{x}\text{arc tan } x = 2\left[p.1 - \frac{p^3}{3}(2T_2(x) - 1) + \frac{p^5}{5}(2T_4(x) - 2T_2(x) + 1) - \frac{p^7}{7}(2T_6(x) - 2T_4(x) + 2T_2(x) - 1) + \cdots\right].$$

Using the relation $p + p^3/3 + p^5/5 + \cdots = \frac{1}{2}\ln\left(\frac{1+p}{1-p}\right) = -\frac{1}{2}\ln p$ (since $p = \sqrt{2} - 1$), this becomes

$$\frac{1}{x}\text{arc tan } x = -\ln p + 4\sum_{n=1}^{\infty}(-1)^n\left(-\frac{1}{2}\ln p - \sum_{k=0}^{n-1}\frac{p^{2k+1}}{2k+1}\right)T_{2n}(x).$$

Using $p = .41421\ 35623$, we find $-\frac{1}{2}\ln p = .44068\ 67933$. The first few coefficients of the Chebyshev expansion $\sum_{n=0}^{\prime\infty} a_{2n}T_{2n}(x)$ for $1/x$ arc tan x are listed below:

$$a_0 = 1.76274\ 71732, \quad \tfrac{1}{2}a_0 = .88137\ 35866, \quad a_2 = -.10589\ 29240,$$
$$a_4 = .01113\ 58416, \quad a_6 = -.00138\ 11944.$$

It is also possible to find the Chebyshev expansion of a function $f(x)$ given in the form $f(x) = \int^x \phi(t)\,dt$, where an expansion of $\phi(t)$ in the form (4.1) is known. To do this the relations (2.12) are used as illustrated in the example below.

Example 5. Find the expansion $\ln(1 + x) = \sum_{n=0}^{\prime\infty} a_nT_n^*(x)$ $(0 \le x \le 1)$. Writing $\ln(1 + x) = \sum_{n=0}^{\prime\infty} a_nT_n^*(x)$, and $1/(1 + x) = \sum_{n=0}^{\prime\infty} A_nT_n^*(x)$ $(0 \le x \le 1)$ and using the relations

$$\int_0^x 1.dt = \frac{1}{2}T_1^*(x) + \frac{1}{2},$$

$$\int_0^x T_1^*(t)\,dt = \frac{1}{8}T_2^*(x) - \frac{1}{8},$$

$$\int_0^x T_n^*(t)\,dt = \frac{1}{4}\left[\frac{T_{n+1}^*(x)}{n+1} - \frac{T_{n-1}^*(x)}{n-1}\right] + \frac{(-1)^{n+1}}{2(n^2-1)} \qquad (n \ge 2),$$

it follows that

$$\begin{aligned}\ln(1 + x) &= \int_0^x \frac{1}{1+t}\,dt \\ &= \frac{1}{4}A_0[T_1^*(x) + 1] + A_1\left[\frac{1}{8}T_2^*(x) - \frac{1}{8}\right] + \cdots \\ &\qquad + A_n\left[\frac{T_{n+1}^*(x)}{4(n+1)} - \frac{T_{n-1}^*(x)}{4(n-1)} + \frac{(-1)^{n+1}}{2(n^2-1)}\right] + \cdots \\ &= \left[\frac{1}{4}A_0 - \frac{1}{8}A_1 + \cdots + A_n\frac{(-1)^{n+1}}{2(n^2-1)} + \cdots\right] \\ &\quad + \frac{1}{4}T_1^*(x)(A_0 - A_2) + \cdots + \frac{1}{4}T_n^*(x)\left(\frac{1}{n}A_{n-1} - \frac{1}{n}A_{n+1}\right) + \cdots.\end{aligned}$$

Thus in the expansion of $\ln(1 + x)$ we have

$$a_0 = \left[\frac{1}{2}A_0 - \frac{1}{4}A_1 + \cdots + \frac{(-1)^{n+1}}{n^2 - 1}A_n + \cdots\right]$$

and

$$a_n = \frac{1}{4n}[A_{n-1} - A_{n+1}] \qquad (n \geq 1).$$

The relation

$$1/(1 + x) = \tfrac{1}{2}A_0 + A_1T_1^*(x) + A_2T_2^*(x) + \cdots$$

can be written

$$\frac{4}{T_1^* + 3} = A_0 + 2A_1 T_1^* + 2A_2 T_2^* + 2A_3 T_3^* + \cdots.$$

Multiplying through by $T_1^* + 3$, and using the relation $2T_1^*T_n^* = T_{n+1}^* + T_{|n-1|}^*$, it follows that

$$4 = (3A_0 + A_1) + (A_0 + 6A_1 + A_2)T_1^* + \cdots + (A_{n-1} + 6A_n + A_{n+1})T_n^* + \cdots.$$

The coefficient of T_k^* satisfies the relation

$$A_{k-1} + 6A_k + A_{k+1} = 0 \qquad (k \geq 1).$$

The solution of this difference equation which remains bounded as $k \to \infty$ is of the form $A_k = cq^k$, where $q = -3 + 2\sqrt{2}$. The relation $4 = 3A_0 + A_1$ implies that $c = \sqrt{2}$, from which $A_k = \sqrt{2}q^k$, where $q = -3 + 2\sqrt{2} = -.17157\ 28753$. Hence it follows that

$$a_n = (6 - 4\sqrt{2})\frac{q^{n-1}}{n} \qquad (n \geqq 1).$$

Also

$$\begin{aligned}
a_0 &= \tfrac{1}{2}A_0 - \tfrac{1}{4}A_1 - A_2/3 + \cdots + (-1)^{n+1}A_n/(n^2 - 1) + \cdots \\
&= \sqrt{2}[\tfrac{1}{2} - q/4 - q^2/3 + \cdots + (-1)^{n-1}q^n/2(n - 1) \\
&\qquad\qquad - (-1)^{n+1}q^n/2(n + 1) + \cdots] \\
&= \sqrt{2}\left[-\frac{1}{2}q\ln(1 + q) + \frac{1}{2q}\ln(1 + q)\right] \\
&= \frac{\sqrt{2}}{2}\cdot\frac{1 - q^2}{q}\ln(1 + q) = -4\ln(1 + q) = (-4)(-.18822\ 64065) \\
&= .75290\ 56260.
\end{aligned}$$

The coefficients a_n may be generated using the relation

$$a_{n+1} = \frac{n}{n + 1}qa_n \qquad (n \geq 1),$$

$$a_1 = 6 - 4\sqrt{2} = .34314\ 57505.$$

The coefficients a_0 to a_3 are found to be as follows:

$$a_0 = .75290\ 56260, \qquad a_1 = .34314\ 57505,$$

$$a_2 = -.02943\ 72515, \qquad a_3 = .00336\ 70893.$$

The Chebyshev expansion of a function which satisfies a linear differential equation whose coefficients are rational functions of the independent variable may be found by the τ-method of Lanczos [1956] or alternatively by a method described by Clenshaw [1957]. For a detailed comparison of these methods with illustrative examples see Fox [1962].

In the τ-method the solution of the equation is approximated by a polynomial of a given degree n. This is substituted in the equation leading to a set of relations for the coefficients. However, unless the function in question actually is a polynomial, this set will in general be overdetermined. This difficulty is resolved by modifying the differential equation through the addition of a term of the form $\tau_1 T_{n+1}(x) + \cdots + \tau_k T_{n+k}(x)$. Lanczos remarks that in a large number of cases of practical interest the value $k = 1$ is used, and that it is rare to require $k > 2$. The solution of the equations for the coefficients also provides a value for the parameters $\{\tau_i\}$ which have been introduced. The polynomial found in this way is of course the solution of a modified form of the differential equation rather than the original equation itself. However, if the $\{\tau_i\}$ are small, as is usually the case, the solution which is found is very close to the truncated Chebyshev expansion of the solution of the original equation. Examples of this method are discussed in detail in Lanczos [1956] and Fox [1962] and are not repeated here.

In Clenshaw's method an expansion of the form (4.1) is substituted in the differential equation. With the help of the integration formulas (2.12) and the recurrence relation (2.10) a recurrence relation for the coefficients of the Chebyshev expansion of the function is then derived. Thus if we write

$$y = \sum_{n=0}^{\infty}{}' a_n T_n(x), \qquad y' = \sum_{n=0}^{\infty}{}' a_n' T_n(x), \qquad y'' = \sum_{n=0}^{\infty}{}' a_n'' T_n(x),$$

a term-by-term integration of the series for y' (or y'') and a subsequent rearrangement of terms leads to

$$\left.\begin{aligned} 2ka_k &= a'_{|k-1|} - a'_{k+1} \\ 2ka_k' &= a''_{|k-1|} - a''_{k+1} \end{aligned}\right\} \quad (k \geq 0). \tag{4.9}$$

Also, letting $C_k(y) = a_k \quad (k = 0, 1, 2, \cdots)$, the relation (2.10) leads to

$$\left.\begin{aligned} C_k(xy^{(n)}) &= \tfrac{1}{2}(a^{(n)}_{|k-1|} + a^{(n)}_{k+1}) \\ C_k(x^2y^{(n)}) &= \tfrac{1}{4}(a^{(n)}_{|k-2|} + 2a^{(n)}_k + a^{(n)}_{k+2}) \end{aligned}\right\} \quad \begin{aligned} &(k = 0, 1, 2, \cdots), \\ &(n = 0, 1, 2, \cdots). \end{aligned} \tag{4.10}$$

If the differential equation is written $L(y) = 0$, we write $C_k[L(y)] = 0$ for $k = 0, 1, 2, \cdots$. With the help of (4.9) and (4.10) recurrences connecting the a, a' and a'' are then determined and a table of trial values is constructed. To do this it is assumed that for subscripts n greater than some arbitrarily selected N, we have $a_n = 0$. The recurrence is used backwards to provide successively trial values for $a_{N-1}, \cdots, a_1, a_0$, and finally the trial values are adjusted through the use of given boundary conditions. It should be noted that the recurrence is used backwards primarily to ensure roundoff stability. Although this procedure is easy to use there are some difficulties connected with it. For example if the polynomial coefficients of the differential equation are of high degree, the resulting difference equation will be of high order, with several solutions. Care in the selection of the proper linear combination will be required. The use of this method is illustrated in the example which follows.

Example 6. Where the Fresnel integrals are defined as

$$C(x) = \sqrt{(2/\pi)} \int_0^{\sqrt{x}} \cos t^2 \, dt$$

and $S(x) = \sqrt{(2/\pi)} \int_0^{\sqrt{x}} \sin t^2 \, dt$, find the Chebyshev series expansion for $C(x)/\sqrt{x}$ and $S(x)/\sqrt{x}$.

Consider the function $y(x) = \sqrt{(2/\pi)} \int_0^{\sqrt{x}} e^{it^2} \, dt$, and look for an expansion of $y(x)$ in the form $y(x) = \sqrt{x} \sum_{n=0}^{\prime\infty} a_n T_n(x)$. From the relation

$$u(x) = y(x)/\sqrt{x} = \sqrt{(2/\pi x)} \int_0^{\sqrt{x}} e^{it^2} \, dt$$

it follows by differentiation that

$$u'(x) = -u/2x + \sqrt{(2/\pi)} e^{ix}/2x.$$

Thus

$$2xu' = -u + \sqrt{(2/\pi)} e^{ix}$$

and

$$2xu'' + 2u' = -u' + \sqrt{(2/\pi)} i e^{ix}.$$

Multiplying the first of the above equations by $-i$ and adding to the second, it follows that $2xu'' + (3 - 2ix)u' - iu = 0$. Accordingly

$$2C_k(xu'') + 3C_k(u') - 2iC_k(xu') - iC_k(u) = 0,$$

and hence

$$a''_{|k-1|} + a''_{k+1} + 3a'_k - i(a'_{|k-1|} + a'_{k+1}) - ia_k = 0 \qquad (k = 0, 1, 2, \cdots).$$

Writing $a_k = \alpha_k + i\beta_k$, this becomes

$$\alpha''_{|k-1|} + i\beta''_{|k-1|} + \alpha''_{k+1} + i\beta''_{k+1} + 3\alpha'_k + 3i\beta'_k$$
$$- i(\alpha'_{|k-1|} + i\beta'_{|k-1|} + \alpha'_{k+1} + i\beta'_{k+1}) - i(\alpha_k + i\beta_k) = 0.$$

Separating real and imaginary parts we have

$$\left.\begin{aligned} \alpha''_{|k-1|} + \alpha''_{k+1} + 3\alpha'_k + \beta'_{|k-1|} + \beta'_{k+1} + \beta_k &= 0 \\ \beta''_{|k-1|} + \beta''_{k+1} + 3\beta'_k - \alpha'_{|k-1|} - \alpha'_{k+1} - \alpha_k &= 0 \end{aligned}\right\} \qquad (k = 0, 1, 2, \cdots).$$

If in the first of these equations the index k is increased by 2, the equation becomes

$$\alpha''_{k+1} + \alpha''_{k+3} + 3\alpha'_{k+2} + \beta'_{k+1} + \beta'_{k+3} + \beta_{k+2} = 0.$$

Subtracting and using (4.9) this gives us

$$\begin{aligned} &(2k + 1)\beta_k + (2k + 3)\beta_{k+2} + 2k\alpha'_k \\ &\quad + 6(k + 1)\alpha_{k+1} + 2(k + 2)\alpha'_{k+2} = 0 \quad (k = 0, 1, 2, \cdots). \end{aligned} \tag{4.11}$$

In the same way, from the second of the equations we derive

$$(2k + 1)\alpha_k + (2k + 3)\alpha_{k+2} - 2k\beta_k' - 6(k + 1)\beta_{k+1} - 2(k + 2)\beta_{k+2}' = 0 \quad (k = 0, 1, 2, \cdots). \tag{4.12}$$

A table of trial values of α_k, α_k', β_k, β_k' is now set up by using

$$\left.\begin{aligned} \alpha'_{|k-1|} &= 2k\alpha_k + \alpha'_{k+1} \\ \beta'_{|k-1|} &= 2k\beta_k + \beta'_{k+1} \end{aligned}\right\} \quad (k = 0, 1, 2, \cdots) \tag{4.13}$$

along with (4.11) and (4.12). For some sufficiently large N, which has been chosen in this instance to be $N = 8$, we set

$$\alpha_{N+p} = \cdots = \beta'_{N+p} = 0 \qquad (p = 1, 2, \cdots).$$

Since $C(x)/\sqrt{x}$ is even and $S(x)/\sqrt{x}$ is odd, we choose $\alpha_8 = \beta_8' = 1$ and $\alpha_8' = \beta_8 = 0$.

The table of trial values is as follows:

k	α_k	β_k	α_k'	β_k'
0	171,038,306.	0	0	53,737,684.
1	0	28,390,299.	16,759,204.	0
2	−4,289,389.6	0	0	−3,042,914.1
3	0	−513,987.74	398,353.90	0
4	50,230.123	0	0	41,012.342
5	0	4,126.5209	−3,487.0768	0
6	−291.92306	0	0	−252.86667
7	0	−18.133333	16.0	0
8	1	0	0	1

The trial values are adjusted to their final values by using the boundary condition $u(0) = \sqrt{(2/\pi)} = .7978\ 8456$. The trial value of $u(0)$ is $(1/2)\alpha_0 - \alpha_2 + \alpha_4 - \alpha_6 + \alpha_8 = 89{,}859{,}065.65$. To adjust trial values to final values all entries in the trial table are multiplied by $0.7978\ 8456/89{,}859{,}065.65 = .88792\ 88408 \times 10^{-8}$. The table of final values thus obtained is given below:

k	α_k	β_k	α_k'	β_k'
0	1.5186 9844	0	0	.4771 5238
1	0	.2520 8565	−.1488 0981	0
2	−.0380 8673	0	0	−.0270 1891
3	0	−.0045 6385	.0035 3710	0
4	.0004 4601	0	0	.0003 6416
5	0	.0000 3664	−.0000 3096	0
6	−.0000 0259	0	0	−.0000 0225
7	0	−.0000 0016	.0000 0014	0
8	.0000 0001	0	0	.0000 0001

If the expansion for $u(x)$ had been desired in a range $(-\lambda, \lambda)$, the differential equation for u, $2xu'' + (3 - 2ix)u' - iu = 0$ would have been modified by making the change of independent variable $x = \lambda t$, but otherwise the procedure would have been the same.

It should be noted that for a constant m, $C_0(m) = 2m$, while $C_k(m) = 0$ $(k = 1, 2, \cdots)$.

5. *Near-Minimax Approximations; Modified Interpolation*

In Section 2 it was mentioned that the minimum estimate of the maximum error of an interpolating polynomial occurred if the nodes of the interpolation were located at the zeros of $T_{n+1}(x)$. A similar procedure is to modify the function $f(x)$ at the points $\{x_i\}$ for which $|T_{n+1}(x)| = 1$ by the alternate addition and subtraction of an amount $\frac{1}{2}E_n$. The coefficients $\{c_{nk}\}$ of the approximating polynomial

$$C_n(x) = \sum_{k=0}^{n}{}' c_{nk} T_k(x) \tag{5.1}$$

and the number E_n are then determined by solving the equations

$$C_n(x_j) + (-1)^j E_n/2 = f(x_j) \qquad (j = 0, 1, \cdots, n+1). \tag{5.2}$$

The number $|E_n/2|$ provides an estimate of the maximum error associated with the approximation of $f(x)$ by $C_n(x)$. This estimate can never be too large (Novodvorskii and Pinsker [1951]), but for a large class of well-behaved functions it is sufficiently close to the maximum to make its use practical.

To solve the system (5.2) we first note that the points $\{x_i\}$ at which $|T_{n+1}(x)| = 1$ are given by

$$x_i = \cos \frac{i\pi}{n+1} \qquad (i = 0, 1, \cdots, n+1). \tag{5.3}$$

In addition the value of $T_k(x_j)$ is given by

$$T_k(x_j) = \cos \frac{kj\pi}{n+1}. \tag{5.4}$$

The first and the last of the equations (5.2) are multiplied by $\frac{1}{2}$. With the equations in this form it is then easily verified by the use of the orthogonality relations (2.14) that

$$c_{nk} = \frac{2}{n+1} \sum_{j=0}^{n+1}{}'' f(t_j) \cos \frac{kj\pi}{n+1} \qquad (k = 0, \cdots, n), \tag{5.5}$$

and in particular

$$\frac{1}{2} E_n = \frac{1}{n+1} \sum_{j=0}^{n+1}{}'' (-1)^j f(t_j). \tag{5.6}$$

Both the coefficients $\{c_{nk}\}$ and the maximum error estimate $(1/2)|E_n|$, associated with the approximation $C_n(x)$ can thus be determined without the necessity of a numerical matrix inversion (or of a numerical integration).

E_n and the coefficients $\{c_{nk}\}$ are related to the coefficients $\{a_k\}$ of the Fourier-Chebyshev expansion of $f(x)$ by

$$c_{ni} = a_i + a_{2(n+1)-i} + a_{2(n+1)+i} + a_{4(n+1)-i} + a_{4(n+1)+i} + \cdots \tag{5.7}$$

and

$$\tfrac{1}{2}E_n = a_{n+1} + a_{3(n+1)} + a_{5(n+1)} + \cdots. \tag{5.8}$$

Thus for a function whose coefficients $\{a_i\}$ decrease in size with sufficient rapidity as i increases, the coefficients $\{c_{ni}\}$ very nearly equal the $\{a_i\}$, while $(1/2)E_n$ will nearly equal a_{n+1}.

The following example illustrates the use of this method.

Example 7. Find a fourth degree polynomial approximation to the function $f(x) = 4/(x-1) \ln((x+3)/4)$, $(-1 \leq x < 1)$, $f(1) = 1$.

In this case the points $\{x_i\}$ are given by $x_i = \cos i\pi/(n+1) = \cos i\pi/5$ $(i = 0, 1, \cdots, 5)$. Writing c_k instead of c_{nk} and E instead of E_n, the solution (5.5) of the equations (5.2) can be written in the form:

$$\begin{pmatrix} c_0 \\ c_1 \\ c_2 \\ c_3 \\ c_4 \\ E \end{pmatrix} = \frac{2}{5} \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & \cos 36° & \cos 72° & \cos 108° & \cos 144° & \cos 180° \\ 1 & \cos 72° & \cos 144° & \cos 216° & \cos 288° & \cos 360° \\ 1 & \cos 108° & \cos 216° & \cos 324° & \cos 432° & \cos 540° \\ 1 & \cos 144° & \cos 288° & \cos 432° & \cos 576° & \cos 720° \\ 1 & \cos 180° & \cos 360° & \cos 540° & \cos 720° & \cos 900° \end{pmatrix} \begin{pmatrix} \frac{1}{2}f(\cos 0°) \\ f(\cos 36°) \\ f(\cos 72°) \\ f(\cos 108°) \\ f(\cos 144°) \\ \frac{1}{2}f(\cos 180°) \end{pmatrix}.$$

The value of $\cos 36°$ is $\frac{1}{4}(1 + \sqrt{5}) = .80901\ 69944$, and $\cos 72°$ is equal to $.30901\ 69944$. The column vector on the right becomes

$$\begin{pmatrix} .50000\ 00000 \\ 1.02466\ 10550 \\ 1.09781\ 68485 \\ 1.21125\ 33490 \\ 1.33098\ 59641 \\ .69314\ 71806 \end{pmatrix}.$$

Performing the matrix-vector multiplication we get

$$\begin{aligned} c_0 &= 2.34314\ 57589, & \tfrac{1}{2}c_0 &= 1.17157\ 28795 \\ c_1 &= -.19040\ 92176 \\ c_2 &= .02120\ 20443 \\ c_3 &= -.00268\ 62086 \\ c_4 &= .00037\ 22568 \\ E &= -.00010\ 35088, & \tfrac{1}{2}E &= -.00005\ 17544. \end{aligned}$$

With the change of variable $\frac{1}{4}(x + 3) = u$, we have

$$T_n(x) = T_n^*(2u - 1)$$

and

$$\begin{aligned} \ln u \cong (u - 1)\{&1.17157\ 28795 - .19040\ 92176T_1^*(2u - 1) \\ &+ .02120\ 20443T_2^*(2u - 1) - .00268\ 62086T_3^*(2u - 1) \\ &+ .00037\ 22568T_4^*(2u - 1)\} \end{aligned}$$

with a maximum error estimate of approximately .000 025.

6. *Conclusion*

In the preceding sections approximations to a function $f(x)$ have been obtained as follows:

(i) the truncated Chebyshev expansion

$$A_n(x) = \sum_{k=0}^{n}{}' a_k T_k(x); \tag{6.1}$$

(ii) the minimax approximation

$$B_n(x) = \sum_{k=0}^{n}{}' b_{nk} T_k(x); \tag{6.2}$$

(iii) the modified interpolation

$$C_n(x) = \sum_{k=0}^{n}{}' c_{nk} T_k(x). \tag{6.3}$$

An approximation in any of the forms (6.1), (6.2) or (6.3) is frequently better conditioned than the corresponding approximation obtained by the rearrangement into a power polynomial which can be achieved using the inverse of the matrix of example 3. The coefficients of the linear Chebyshev combination do not in general have as great a range of sizes and this is often an advantage, particularly if the function in question has zeros or relatively small minima in the range of approximation. Such values are achieved by a cancellation, and the greater the range of size of the coefficients the greater the problem of avoiding loss of significant figures.

A polynomial given in a form such as (6.1) may be evaluated by the following procedure (Clenshaw [1955]): Set $b_{n+2} = b_{n+1} = 0$, and $b_k = 2xb_{k+1} - b_{k+2} + a_k$ $(k = n, \cdots, 1, 0)$. Then $A_n(x) = \frac{1}{2}(b_0 - b_2)$.

This procedure is easily modified for the special cases (i) $A_n(x)$ is even, (ii) $A_n(x)$ is odd, (iii) $A_n(x)$ is given in terms of shifted Chebyshev polynomials.

Each of (6.1), (6.2) or (6.3) in some circumstances possesses advantages over the other two. For example the coefficients $\{a_k\}$ of (6.1) are independent of n, the degree of the approximation, and hence a change in n does not require their recalculation, as would be the case with the coefficients of (6.2) or (6.3). The use of (6.2) leads to the maximum possible precision for a given degree of approximation, or alternatively the lowest possible degree which may be used to obtain a stated precision. Also for functions defined empirically by a table of values the use of (6.2) is particularly convenient. If, on the other hand, an empirical function is defined by a graph, or values can be found for specified choices of the independent variable, the use of (6.3) is advantageous. Also for analytic functions where changes in range and precision from an established routine are required, the use of (6.3) is very convenient.

Acknowledgment. The author wishes to thank Dr. Henry C. Thacher for his detailed criticisms of this article and helpful suggestions for its improvement.

RECEIVED FEBRUARY, 1965

REFERENCES

ACHIESER, N. I. *Theory of Approximation.* Ungar, New York, 1956. English translation by C. J. Hyman.

CHENEY, E. W. Approximation theory. Dept. of Mathematics, U. of California, 1963. (Mimeo)

CLENSHAW, C. W. A note on the summation of Chebyshev series. *MTAC 9*, 1955, 118–120.

——. The numerical solution of linear differential equations in Chebyshev series. *Proc. Camb. Phil. Soc. 53* (1957), 134–149.

——. *Chebyshev Series for Mathematical Functions.* National Physical Lab., Math. Tables 5, HM Stat. Off., London, 1962.

DAVIS, P. *Interpolation and Approximation.* Blaisdell, New York, 1963.

FOX, L. Chebyshev methods for ordinary differential equations. *Comput. J. 4* (1962), 318–31.

GOLOMB, M. Lectures on theory of approximation. Appl. Math. Div., Argonne Nat. Lab., Argonne, Ill., 1962.

KOPAL, Z. *Numerical Analysis*. Chapman & Hall, London, 1955.

LANCZOS, C. *Applied Analysis*. Prentice-Hall, Englewood Cliffs, N. J., 1956.

MURNAGHAN, F. D., AND WRENCH, J. W. Report No. 1175, David Taylor Model Basin, Md., 1960.

NATANSON, I. P. *Constructive Theory of Functions*. AEC-tr-4503, Books 1 and 2. English translation by US AEC.

NATIONAL BUREAU OF STANDARDS. *Handbook of Mathematical Functions*. NBS Appl. Math. Series 55, US Govt. Printing Off., Washington, D. C., 1964.

NOVODVORSKII, E. P., AND PINSKER, I. S. The process of equating maxima. *Uspehi Mat. Nauk 6* (1951), 174–181. English translation by A. Shenitzer.

REMEZ, E. YA. General computational methods of Chebyshev approximation. In *The Problems with Linear Real Parameters*. AEC-tr-4491, Books 1 and 2, English translation by US AEC.

RICE, J. R. *The Approximation of Functions, Vol. 1, Linear Theory*. Addison-Wesley, Reading, Mass., 1964.

SHANKS, D. Non-linear transformations of divergent and slowly convergent series. *J. Math. Phys. 34* (1955), 1–42.

STIEFEL, E. L. Numerical methods of Chebyshev approximation. In *On Numerical Approximation*, R. E. Langer, Ed., U. of Wisconsin Press, Madison, 1959.

THACHER, H. C. Conversion of a power to a series of Chebyshev polynomials. *Comm. ACM 7* (1964), 181–182.

VALLÉE POUSSIN, C. J. DE LA. Sur la méthode de l'approximation minimum. *Ann. Soc. Sci. Bruxelles, Seconde Partie,* Mémoires 35, pp. 1–16, Bruxelles, 1911. English translation by H. E. Salzer.

——. *Leçons sur l'Approximation des Fonctions d'une Variable Réelle*. Gauthier-Villars, Paris, 1919.

VEIDINGER, L. On the numerical determination of the best approximations in the Chebyshev sense. *Numer. Math. 2* (1960), 99–105.

WALSH, J. L. *Interpolation and Approximation*. Amer. Math. Soc. Coll. Publ. 20, Providence, 1960.

WYNN, P. On a device for computing the $e_m(S_n)$ transformation. *MTAC 10* (Apr. 1956), 91-96.