Chebyshev approximation by exponential-polynomial sums (*)

Charles B. Dunham (**)

ABSTRACT

It is shown that best Chebyshev approximations by exponential-polynomial sums are characterized by (a variable number of) alternations of their error curve and are unique. Computation of best approximations via the Remez algorithm and Barrodale approach is considered.

Let $[\alpha, \beta]$ be a closed finite interval and $C[\alpha, \beta]$ be the space of continuous functions on $[\alpha, \beta]$. For $g \in C[\alpha, \beta]$ define

$$\|g\| = \max \{|g(x)| : \alpha \leqslant x \leqslant \beta\} .$$

Let w be a positive element of $C[\alpha, \beta]$, an ordinary multiplicative weight function. For given fixed $n > 0$, $m \geqslant 0$ define

$$F(A, x) = \sum_{k=1}^{n} a_k \exp(a_{n+k}x) + \sum_{k=1}^{m} a_{2n+k}x^{k-1} .$$

The Chebyshev problem is given $f \in C[\alpha, \beta]$ to find a parameter A^* to minimize $\|w(f - F(A,.))\|$. Such a parameter A^* is called best and $F(A^*,.)$ is called a best approximation to f.

It might perhaps be thought that the case $m > 0$ is not of practical interest. However, Spath has developed algorithms for discrete L_2 approximation for the case $n = 1$, $m = 1$ and 2 in Spath [26] and Spath [27] respectively. Further, the model $n = 1$, $m = 2$ has been found to apply to a biochemical system (Heitkamp et al. [19] equation (17)) and Spath [27] states that it is often needed in chemistry and radiation chemistry. The author has been informed by Dr. Alan Miller of CSIRO that models with $m = 1$, $n = 2$ were of interest. Such a model was used by Cantraine and Jortay [6]. Models with $m = 1$ and n larger are given by Cook and Taylor [10] and by Lemaitre and Malenge [22]. Models which can be converted by change of variable into the case $n = 1$, $m = 1$ or 2, are discussed by Shah and Khatri [29].

The same model with $m = 1$ is used by Gregory [30]. The case where $m = 0$, $w = 1$, has been studied by Meinardus and Schwedt [25, p. 312-313] and also appears in Meinardus [24, p. 177-178]. We analyze our problem with a theory also due to Meinardus and Schwedt.

A parameter A will be called *standard* if $a_{n+1},\dots, a_{2n}$ are distinct and nonzero (in the case $m = 0$ we will permit one of $a_{n+1},\dots, a_{2n}$ to be zero). It is easily seen that any parameter has an equivalent standard parameter. The *degeneracy* $d(A)$ of a standard parameter A is the number of zero coefficients in $a_1,\dots, a_n$. The *degree* of F at standard parameter A is defined to be $2n + m - d(A)$.

Haar subspaces

Definition

A linear subspace of $C[\alpha, \beta]$ of dimension ℓ is a Haar subspace on $[\alpha, \beta]$ if only the zero element has ℓ zeros on $[\alpha, \beta]$.

Lemma 1

The linear space generated by

$\{\exp(a_i x), x\exp(a_i x),\dots, x^{m(i)} \exp(a_i x)\}$,

$i = 1,\dots, n$, $a_1 < \dots < a_n$, is a Haar subspace of dimension $\sum_{i=1}^{n} (m(i) + 1)$.

The lemma is given in Meinardus and Schwedt [25, p. 313] and also proven in Meinardus [24, p. 177).

Property Z

Lemma 2

Let F be of degree ℓ at standard parameter A, then $F(A,.) - F(B,.)$ has at most $\ell - 1$ zeros or vanishes identically.

Proof

$F(A,.) - F(B,.)$ is a linear combination of $\{1,\dots, x^{m-1}\}$ and at most $2n - d(A)$ non-constant exponentials. By Lemma 1 it has at most $2n + m - 1 - d(A)$ zeros or vanishes identically.

(*) This work has been assisted by a grant from the National Research Council of Canada.

(**) Ch. B. Dunham, The University of Western Ontario, Dept. of Computer Science, London Ontario N6A 5B9, Canada.

The partial derivatives

Define the partial derivatives

$$F_k(A, x) = \frac{\partial}{\partial a_k} F(A, x),$$

then they are seen to be continuous in A, x.

The tangent space

Define

$$D(A, B, x) = \sum_{k=1}^{2n+m} b_k \frac{\partial}{\partial a_k} F(A, x)$$

$$= \sum_{k=1}^{n} b_k \exp(a_{n+k}x) + \sum_{k=1}^{n} b_{n+k} a_k \, x \exp(a_{n+k}x)$$

$$+ \sum_{k=1}^{m} b_{2n+k} x^{k-1}$$

Lemma 3

Let F be of degree ℓ at standard parameter A, then $\{D(A, ., .)\}$ is a Haar subspace of dimension ℓ. This follows immediately from Lemma 1.

Characterization of best approximations

Fundamental to much of the characterization theory of Chebyshev approximation, both linear and non-linear, is the alternation (equioscillation, equal ripple) property.

Definition

$g \in C[\alpha, \beta]$ alternates ℓ times on $[\alpha, \beta]$ if there exists $\{x_0, \ldots, x_\ell\}$, where $\alpha \leqslant x_0 < \ldots < x_\ell \leqslant \beta$, such that

$$|g(x_i)| = \|g\| \qquad g(x_i) = (-1)^i g(x_0) \qquad i = 0, \ldots, \ell.$$

Such a set $\{x_0, \ldots, x_\ell\}$ is called an alternant of g.

Theorem 1

Let F have degree ℓ at standard parameter A. A necessary and sufficient condition that F (A, .) be best to f is that w (f - F (A, .)) alternate ℓ times on $[\alpha, \beta]$. Best approximations are unique.

The theorem follows from Lemma 2, continuity of the partial derivatives, Lemma 3, and theory of Meinardus and Schwedt [25, theorems 13 and 14], also given in Meinardus [24, p. 144-147] in the case w = 1. The extension to a general w follows from theorem 1 of Dunham [12].

Computation of best approximations

In case the best approximation is of maximum degree (2n + m), its error curve alternates at least 2n + m times. This suggests that best approximations can be determined by a variant of the Remez algorithm, general versions of which are given by Barrar and Loeb ([1], p. 389) and Dunham ([12], p. 228).

The Remez algorithm for our problem proceeds roughly as follows :

(i) set j = 0 and choose $\{x_0, \ldots, x_{2n+m}\}$, where $\alpha \leqslant x_0 < \ldots < x_{2n+m} \leqslant \beta$,

(ii) solve the (non-linear) system of 2n + m + 1 levelling equations

$$(*) \quad f(x_i) - F(A, x_i) - (-1)^i \lambda / w(x_i) = 0, \qquad i = 0, \ldots, 2n + m,$$

for 2n + m + 1 unknowns A^j and λ^j,

(iii) find the alternating extrema of the error curve $w(f - F(A^j, .))$ and call them $\{x_0, \ldots, x_{2n+m}\}$.

This process of simultaneous exchanges is described in Meinardus [24, p. 107]

(iv) add 1 to j and go to (ii).

This process is an infinite process. In an actual program it is terminated when the extrema of the error curve found in (iii) are sufficiently equal in absolute value. Finding the alternating extrema can be done exactly as in variants of the Remez algorithm for other approximation problems, for example Cody, Fraser, and Hart [9] for rational approximation. The major difficulty is in solving the levelling equations (*). The most straightforward way to solve them is to use Newton's method directly. In the special case n = 1, it is also possible to re-arrange the system (*) to make the unknowns enter less nonlinearly and then apply Newton's method. The case n = m = 1, w = 1, is considered in Dunham [11] and it is not difficult to see from this analysis how to cover the general case of n = 1. Both approaches appear to work reasonably well in the case n = 1 in the author's experiments. A problem with both approaches is that Newton's method may converge to a "pseudo-solution". Pseudo-solutions for the case n = m = 1 are given in Dunham [11, p. 211] for the method with re-arrangement. Whether the levelling equations (*) can easily be solved in the case where $n \geqslant 2$ and optimal $\{a_{n+k}\}$ are not well separated is still an open question. It should be noted that the levelling equations may not even have a solution.

Example

Let n = 1, m = 0, w = 1 and $f(x_0) = -1$, $f(x_1) = f(x_2) = 1$. Suppose a solution A, λ to (*) exists. The first possibility is that $a_1 \leqslant 0$, in which case $F(A, .) \leqslant 0$, and we cannot have $f(x_1)\ f(x_2) > 0$. The second possibility is that $a_1 > 0$, in which case $F(A, .) > 0$ and it can be seen by drawing a diagram that $f(x_0)$, $f(x_1)$ cannot take their values.

The variant of Meinardus and Schwedt [25, p. 320 ff], also given in Meinardus [24, p. 150-151], uses *one* iteration of Newton's method to approximately solve the levelling equations. Barrar and Loeb [1] have proven that if the best approximation is of maximum degree, and the starting point (estimate of the parameters and deviation of the best approximation and an alternant) is sufficiently close, the Remez algorithm will converge (neglecting analytical and numerical difficulties in solving the levelling equations). Dunham [12] proves that under favorable conditions, the convergence of the Remez algorithm and the variant of Meinardus and Schwedt is quadratic. The author has programmed a version of the Remez algorithm for the case $n=1$ which appears to work reasonably well, provided the parameters of the best approximation can be closely estimated.

Another approach is to reduce part of the problem to a *linear* problem. Associate with each set $a_{n+1}, \ldots, a_{2n}$ of non-linear parameters the minimal error with these parameters. Determining the minimal error is a *linear* Chebyshev problem, involving approximation by a linear combination of

$$\{\exp(a_{n+1}x), \ldots, \exp(a_{2n}x), 1, \ldots, x^{m-1}\} :$$

this is a Chebyshev set of $n + m$ elements if $a_{n+1}, \ldots, a_{2n}$ are distinct and, if $m > 0$, nonzero. The minimal error can be determined by use of the *linear* Remez algorithm, which always converges [24, p. 108] and has quadratic convergence under favorable conditions [24, p. 111-113]. Some of the implementations of the linear Remez algorithm assume that the number of extrema equals 1+ the number of linear parameters. Such programs will have difficulty in our case, as the characterization theorem guarantees at least $2n + m + 1$ alternating extrema at the optimum parameters, if the best approximation is of maximum degree. A program for the linear Remez algorithm which does not make assumptions about the number of extrema is Golub and Smith [18]. We minimize the minimal error associated with non-linear parameters by varying the non-linear parameters. This is particularly easy when $n=1$, in which case there is only one non-linear parameter and the error can be minimized by a one-dimensional search, as in Barrodale et al. [2]. It might be thought that there is some possibility of a purely local minimum being found by this approach. However, it follows from the analysis of Meinardus and Schwedt [25, p. 318], also in Meinardus [24, p. 144 ff], that any local minimum of the error has an alternating error curve and thus is a global minimum. A global minimum can be recognized by use of the characterization theorem.

In computing, the Remez approach should probably be tried first, as it usually has quadratic convergence and proved quite fast in practice (typically less than one second execution time for the case $n=1$, $m=2$ on a CDC Cyber 73 ($\approx$ CDC 6400)). It may fail, possibly due to a poor estimate of the best parameters or their error extrema, in which case the reduction to linear approach should be tried. It should always work, providing the minimization with respect to the non-linear parameters can be accomplished (this is simple in the case $n=1$). Once a rough minimum is located, it may be best to try the Remez algorithm, due to its usual quadratic convergence.

We can attempt to compute best approximations on $[\alpha, \beta]$ by computing best approximations on a sequence $\{X_k\}$ of finite sets filling out $[\alpha, \beta]$. This should work if the best approximation is of maximum degree [13].

Discrete approximation

Consider instead approximation on a finite set $X = \{x_1, \ldots, x_j\}$, $j > 2n + m$, $x_1 < \ldots < x_j$. Define

$$\|g\|_X = \max\{|g(x)| : x \in X\}.$$

We wish to minimize $\|w(f - F(A, .))\|_X$.

Theorem 2

Let F have degree ℓ at standard parameter A. A necessary and sufficient condition that $F(A, .)$ be best to f on X is that $w(f - F(A, .))$ alternate ℓ times on X. Best approximations are unique.

The theorem follows from F being an alternating approximation function on $[\alpha, \beta]$ containing X and the result of Dunham [14] with the weight function being zero off X.

Best approximations can be computed by the discrete analogue of the Remez algorithm (the author has programmed this for (i) the case $n=1$ with the Remez one-for-one (single) exchange [24, p. 107] and it appears to work reasonably well, provided the parameters of the best approximation can be closely estimated and (ii) the case $n = 2$, $m = 0$ [15, 16].

Best approximations can also be computed by reduction to a linear problem. The linear sub-problem can be solved by use of the discrete analogue of the linear Remez algorithm, variants of Stiefel's exchange (ascent) method [7, p. 46-47], or linear programming [3]. It should be noted that Barrodale et al. [2] have used the last approach for the case $n = 1$, $m = 1$. The algorithm of Osborne and Watson could also be used (McBride and Rigler [23] report an experiment with finding the best approximation in a case with $n = 2$, $m = 0$).

A difficulty with exponential approximation

A potential source of difficulty with approximation by sums of exponentials is that sums of exponentials with quite different parameters may be represented by almost the same curve (a trivial example is the closeness of exponentials with large negative arguments to zero on positive points). This is discussed by Lanczos [21, p. 276-279], whose results are also cited in Clenshaw [8, p. 11]. The fact that exponentials may be closely approximated by polynomials on a short range means that the situation is worse (for the same n) with exponential-polynomial sums. One implication is that it may be impossible to find the "true" parameters from experimental data. It is also possible that some numerical methods, in particular Newton's method for solving the levelling equations (*) of the Remez algorithm, will work poorly if $n > 1$ and nonlinear parameters (frequencies) are not well separated.

An open question is how large n and m need be for this difficulty to be serious. The example of Lanczos is for $n = 3$. Marsaglia has presented the author with a plot (unpublished) that shows conclusively that the sums of three exponentials cannot be distinguished by visual inspection. It would be of interest to see how close the sums of two exponentials could be made.

Mean approximation

An alternative to Chebyshev approximation is mean (L_p) approximation, in particular mean-square (L_2) approximation. Possible difficulties inherent in this alternative include lack of a characterization of best approximations (so we are seldom sure if an approximation is best), local minima which are not global minima, and non-uniqueness of best approximations [17]. By theorem 4 of that paper, which makes use of lemma 3 of this paper, best discrete approximations have an error identically zero or with 2 n + m sign changes. Degenerate approximations are best only to themselves.

Even and odd approximation

Let $\alpha = -\beta$ and f be even. If approximation is on finite X, let X be symmetric about zero. Let F (A, .) be best. By symmetry the approximation with the arguments of the exponentials negated and the coefficients of odd powers negated is also best. By uniqueness of the best approximation, these must be the same approximation. Hence the best approximation must be of the form of a sum of hyperbolic cosines and even powers. It may be desirable to approximate f directly by that form on $[0, \beta]$, see the next section. Benefits include reduced computing and increased numerical stability. It should be noted that McBride and Rigler [23] approximate (even) $f(x) = x^2 + 4$ on $\{-1, -24/25, \ldots, -1/25, 0, 1/25, \ldots, 24/25, 1\}$ with n = 2, m = 0. By our observation, the best approximation is actually of the form $a_1 \cosh (a_2 x)$ and approximation could be done on $\{0, 1/25, \ldots, 24/25, 1\}$. A best approximation of this form was computed by the author, as reported in the next section.

Let $\alpha = -\beta$ and f be odd. If approximation is on finite X, let X be symmetric about zero. Let F(A, .) be best. By symmetry, the approximation with multipliers and arguments of the exponentials negated and coefficient of even powers negated is also best. By uniqueness of the best approximation, these must be the same approximation. Hence the best approximations must be of the form of a sum of hyperbolic sines and odd powers. It may be desirable to approximate f directly by that form on $[0, \beta]$, see the next section. Benefits include reduced computing and increased numerical stability.

Approximation by hyperbolic functions

Consider instead the case when $\alpha \geqslant 0$ and

$$F(A, x) = \sum_{k=1}^{n} a_k \cosh (a_{n+k} x) + \sum_{k=1}^{m} a_{2n+k} x^{2(k-1)}$$

If we use example 9 of Dunham [32] in place of lemma 1, we get lemma 2, the analogue of lemma 3, and the theorems holding. The discussion of computation and discrete sets applies. A program to compute best approximations on [0, 1] in the case n = m = 1 by the Remez algorithm, using the approach of Dunham [11], was written and run successfully. A program to compute best approximations on a finite point set in the case n = 1, m = 0 by the Remez algorithm, using an approach of the type of Dunham [11] to solve the levelling equations, was written. It was tested on the example of McBride and Rigler [23] as modified in our discussion of even approximation, namely approximation on the 26-point set $\{0, 1/25, \ldots, 24/24, 1\}$ of $f(x) = x^2 + 4$ by $a_1 \cosh(a_2 x)$ with w = 1. The optimum coefficients obtained were $a_1 = 4.005$, $a_2 = .6927$ and the maximum error was $.49 (10^{-2})$. The alternant of the error curve of the best approximation consisted of the first, 19th, and 26th (last) points.

Consider next the case where $\alpha \geqslant 0$ and

$$F(A, x) = \sum_{k=1}^{n} a_k \sinh (a_{n+k} x) + \sum_{k=1}^{m} a_{2n+k} x^{2k-1}$$

The standard alternating theory does not apply since $F(A, 0) \equiv 0$. The remedy is to only consider the approximation of functions vanishing at zero in the case $\alpha \geqslant 0$, to not count this zero at zero, and to use the theory of Dunham [31] in place of the theory of Meinardus and Schwedt. We use example 8 of Dunham [32] in place of lemma 1 and get the analogues of lemmas 2 and 3 (with the zero at zero not counted) and get the theorems holding for f with f (0) = 0. The discussion of computation and discrete sets applies, providing we take care with the zero at zero. A program to compute best approximations on [0, 1] in the case n = 1, m = 0 by the Remez algorithm was written and run successfully.

REFERENCES

1. BARRAR, R. and LOEB, H. : "On the Remez algorithm for non-linear families", Numer. Math. 15 (1970), 382-391.
2. BARRODALE, I.; ROBERTS, F. and HUNT, C. : "Computing best ℓ_p approximations by functions nonlinear in one parameter", Computer J. 13 (1970), 382-386.
3. BARRODALE, I. and PHILLIPS, C. : "Algorithm 495, Solu tion of an overdetermined system of linear equations in the Chebyshev norm", ACM Trans. Math. Software 1 (1975), 264-270.
4. BRAESS, D. : "Die Konstruktion der Tschebyscheff-Approximierden bei der Anpassung mit exponentiale Summen", J. Approx. Theory 3 (1970), 261-273.
5. BRAESS, D. : *Analysis of decay processes and approximation by exponentials, in C. MICCHELLI and T. J. RIVLIN (eds.), Optimal estimation in approximation Theory,* Plenum Publishing (1977), 259-266.
6. CANTRAINE, F. R. L. and JORTAY, A. M. : "Computer study of iodine transport in thyroid slices in vitro", Computers and Biomedical Res. 8 (1975), 405-422.
7. CHENEY, E. W. : *Introduction to approximation theory,* McGraw-Hill (1966).
8. CLENSHAW, C. : *General considerations, in J. HAYES (ed.), Numerical approximation to functions and data,*

Athlone Press London (1970), 4-15.

9. CODY, W.; FRASER, W. and HART, J. : "Rational Chebyshev approximation using linear equations", Numer. Math. 12 (1968), 242-251.

10. COOK, D. A. and TAYLOR, G. S. : "The use of the APL/360 system in pharmacology. A computer assisted analysis of efflux data", Computers and Biomedical Res. 4 (1971), 157-166.

11. DUNHAM, C. : "Chebyshev approximation by $A + B^*\phi(CX)$", Computing 13 (1974), 205-213.

12. DUNHAM, C. : "Chebyshev approximation with respect to a weight function", J. Approx. Theory 2 (1969), 223-232.

13. DUNHAM, C. : *Varisolvent Chebyshev approximation on subsets, in G. G. LORENTZ (ed.), Approximation theory,* Academic Press (1973), New York.

14. DUNHAM, C. : "Alternating Chebyshev approximation with a non-continuous weight function", Canad. Math. Bull. 19 (1976), 155-157.

15. DUNHAM, C. : "Minimax approximation by two exponentials", Computer J. 20 (1977), 191.

16. DUNHAM, C. : "Discrete minimax approximation by 2 exponentials", Research Report 32 (1977), Computer Science Department, University of Western Ontario.

17. DUNHAM, C. : "Discrete nonlinear mean approximation", Z. Angw. Math. Mech. 57 (1977), 407-412.

18. GOLUB, G. and SMITH, L. : "Algorithm 414, Chebyshev approximation of continuous functions by a Chebyshev system of functions", Comm. ACM 14 (1971), 737-746.

19. HEITKAMP, D.; MERWITZ, O. and SPATH, H. : "Zur Kinetik der strahleninduzierten Wasserstoffabspaltung aus Thymin", Z. Naturforschung 23 (1968), 403-412.

20. JULIUS, R. S. : "The sensitivity of exponentials and other curves to their parameters", Computers and Biomedical Res. 5 (1972), 473-478.

21. LANCZOS, C. : *Applied analysis,* Prentice-Hall (1956), Englewood Cliffs, N. J.

22. LEMAITRE, A. and MALENGE, J. P. : "An efficient method for exponential fitting with a computer", Computers and Biomedical Res. 4 (1971), 555-560.

23. McBRIDE, W. and RIGLER, A. : "A modification of the Osborne and Watson algorithm for nonlinear minimax approximation", Computer J. 19 (1976), 79-81.

24. MEINARDUS, G. : *Approximation of functions : Theory and numerical methods,* Springer-Verlag (1967), New York.

25. MEINARDUS, G. and SCHWEDT, D. : "Nicht-lineare Approximationen", Arch. Rat. Mech. Anal. 17 (1964), 297-326.

26. SPATH, H. : "Algorithm 295, Exponential curve fit", Comm. ACM 10 (1967), 87.

27. SPATH, H. : "Algorithm 33, Fitting data to an exponential with a straight line as background", Computer J. 11 (1968), 144-115.

28. WERNER, H. : *Tschebyscheff-approximation with sums of exponentials, in A. TALBOT (ed.), Approximation theory,* Academic Press (1970), London, 109-136.

29. SHAH, B. K. and KHATRI, C. G. : "A method of fitting the regression curve $E(y) = \alpha + \delta x + \beta\rho^x$", Technometrics 7 (1965), 59-65.

30. GREGORY, W. C. : "Design procedures and use of prior information in the estimation of parameters of the nonlinear model $\eta = \alpha - \beta\gamma^x$", Dissertation, North Carolina State University, Raleigh, 1970.

31. DUNHAM, C. : "Chebyshev approximation with a null point", Z. Angw. Math. Mech. 52 (1972), 239.

32. DUNHAM, C. : "Families satisfying the Haar condition", J. Approx. Theory 12 (1974), 291-298.