JOURNAL OF APPROXIMATION THEORY **20**, 257–263 (1977)

Efficiency of Alternating Chebyshev Approximation on Finite Subsets

CHARLES B. DUNHAM

Computer Science Department, University of Western Ontario, London, Ontario, Canada

Communicated by E. W. Cheney

Received March 24, 1975

Chebyshev approximation on an interval and on its closed subsets by a nonlinear family with Haar tangent space is considered. The closeness of best approximations on subsets to the best approximation on the interval is examined. It is shown that under favorable conditions the difference is $O(d^2)$, where d is the density of the closed subset, making it unnecessary to use very large finite subsets to get good approximations on the interval.

Let X be a closed finite interval $[\alpha, \beta]$ and Y a closed subset of X. Let $C(X)$ be the space of continuous functions on X. For $h \in C(X)$ define

$$\|h\|_Y = \sup\{|h(x)| : x \in Y\}, \qquad \|h\| = \|h\|_X .$$

Let F be an approximating function with parameter space P such that $F(A, \cdot) \in C(X)$ for $A \in P$. The approximation problem on Y is, given $f \in C(X)$, to find $A^* \in P$ to minimize $\|f - F(A, \cdot)\|_Y$ over $A \in P$. Such a parameter A^* is called *best* and $F(A^*, \cdot)$ is called a *best approximation* to f on Y.

Denote the best approximation on Y (if it exists and is unique for all Y) by P_Y. Define

$$\sigma(Y) = \|f - P_Y\| - \|f - P_X\| .$$

$\sigma(Y)$ is a measure of the goodness of the best approximation on Y as an approximation on X. We consider the dependence of P_Y and $\sigma(Y)$ on f, Y, and, in particular, on the *density* $|Y|$ of Y:

$$|Y| = \sup[\inf\{|y - x| : y \in Y\} : x \in X].$$

The case of linear approximation has already been studied by the author in [14].

ISSN 0021-9045

Preliminaries

We will assume henceforth that P is a subset of real n-space with the norm

$$\|A\| = \max\{|a_k| : k = 1,\dots, n\}, \qquad \text{where } A = (a_1, a_2, \dots, a_n).$$

Definition. Let F have a continuous partial derivative F_j with respect to a_j, $j = 1,\dots, n$. Define

$$D(A, B, x) = \sum_{k=1}^{n} b_k F_k(A, x), \qquad B = (b_1, b_2, \dots, b_n) \in E_n .$$

For some A, let

$$R(A, B, x) = F(A + B, x) - F(A, x) - D(A, B, x)$$

be $O(\|B\|^2)$ as $\|B\| \to 0$. Assume P is open. Then we say F is *locally linear* at A.

Definition. F has *property* Z of degree m at A if $F(A, \cdot) - F(B, \cdot)$ having m zeros implies $F(A, \cdot) = F(B, \cdot)$.

We assume henceforth that for all $A \in P$, F is locally linear at A and there is a positive number $\rho(A)$ (the *degree* of F at A) such that F has property Z of degree $\rho(A)$ at A and, $\{D(A, B, \cdot) : B \in E_n\}$ is a Haar subspace of dimension $\rho(A)$. It follows from the theory of Meinardus and Schwedt [6, p. 310] that $F(A, \cdot)$ is best to f on Y if and only if $f - F(A, \cdot)$ alternates $\rho(A)$ times on Y, and best approximations on Y are unique. It has been shown by Barrar and Loeb [1] that our hypotheses imply that F is unisolvent of degree $\rho(A)$ at A: that is, F is *varisolvent*.

We now give some approximating families (F, P) satisfying these hypotheses. Families of ordinary rational functions (with a constant term in the denominator fixed equal to one) satisfy them. It is shown by Meinardus and Schwedt that sums of n exponentials satisfy them [6, p. 312]. Some families of the form $F(A, x) = a_1\phi(a_2x)$ [10] and many families of the form $F(A, x) = a_1 + a_2\phi(a_3x)$ [11] satisfy them.

Let us consider transformations of the approximating family. Let ϕ be a continuous mapping of the real line into the extended real line and define

$$G(A, x) = \phi(F(A, x)) \qquad P' = \{A : A \in P, \|G(A, \cdot)\| < \infty\}.$$

If ϕ is an order function [8] whose first derivative does not vanish and whose second derivative is continuous (where ϕ is finite), then (F, P) satisfying the hypotheses implies that (G, P') satisfies the hypotheses with the same degree [8; 9, Theorem 3]. Of particular utility are transformed polynomials and transformed rationals [8, 12, 13]. Use of an ordinary weight function $w > 0$ does not affect matters, since if (F, P) satisfies the hypotheses, so does (wF, P).

In [5] we proved

THEOREM 1. *Let $F(A, \cdot)$ be best to f on X and $\rho(A) = n$. On all sufficiently dense subsets a best approximation of degree n exists to f. Let $| Y_k | \to 0$ and A^k be best on Y_k, then $\| F(A, \cdot) - F(A^k, \cdot) \| \to 0$.*

If $\rho(A) < n$ such a result may not be possible. In particular it is known for a class of cases including ordinary rational approximation and exponential approximation that there exists Y_k, $| Y_k | \to 0$, with uniform convergence not occurring [3]. More generally, failure of existence or uniform convergence can often be shown to occur [4, Theorem 7].

Define $E(A, x) = f(x) - F(A, x)$. In case the best approximation on $\{x_0, ..., x_n\}$ is of degree n, it can be obtained by solving the system

$$E(A, x_i) = (-1)^i \lambda \qquad i = 0, ..., n \tag{0}$$

for unknowns $a_1, ..., a_n, \lambda$. A solution is unique [2, middle of page 228]. The system is fundamental to Remez' algorithm [2, p. 228]. We can attempt to solve the system by Newton's method.

LEMMA. *Let there exist a solution A^*, λ^* to (0) with $F(A^*, \cdot)$ of degree n. The matrix of partial derivatives for Newton's method is nonsingular at the solution (A^*, λ^*) of system (0).*

Proof. If it is singular, there is $B = (b_1, ..., b_{n+1})$ not identically zero such that

$$D(A^*, B, x_i) + (-1)^i b_{n+1} = 0 \qquad i = 0, ..., n.$$

If $b_{n+1} = 0$, $D(A^*, B, x_i)$ has $n + 1$ zeros, contradicting the Haar subspace hypothesis. If $b_{n+1} \neq 0$, $D(A^*, B, \cdot)$ alternates in sign on $\{x_0, ..., x_n\}$, contradicting the Haar subspace hypothesis.

DEPENDENCE UNDER FAVORABLE CONDITIONS

THEOREM 2. *Let f have $F(A^*, \cdot)$ as its best approximation on X and let $\rho(A^*) = n$. Let endpoints be in Y if they are extrema of $f - F(A^*, \cdot)$. Let E have continuous second partial derivatives about (A^*, x) for all $x \in [\alpha, \beta]$. Then $\| P_X - P_Y \| = O(| Y^2 |)$, and $\sigma(Y) = O(| Y |^2)$.*

Proof. Let $\{x_0^*, ..., x_n^*\}$ be an alternant of $f - F(A^*, \cdot)$ on X. The reader is asked to review the notation of W. Kahan and the author [2, middle of page 229] and the proof [2, p. 230]. We note that matrix (4) of [2, 229] is nonsingular by the previous lemma. Using the techniques of [2, pp. 229–230], we associate with given $\{x_0, ..., x_n\}$ near an alternant of $f - F(A^*, \cdot)$ on X, a change of deviation $\delta\lambda = \lambda - \lambda^*$, where λ is the deviation of the best approximation on $\{x_0, ..., x_n\}$. We infer from the cited proof, with $\{x_0, ..., x_n\}$ chosen closest in Y to $\{x_0^*, ..., x_n^*\}$, that $\delta\lambda = O(|Y|^2)$.

By Theorem 1, there is $\epsilon > 0$ such that if $|Y| < \epsilon$, then there is a best approximation P_Y to f on Y and P_Y is of degree n. Assume henceforth that $|Y| < \epsilon$. An alternant $\{x_0, ..., x_n\}$ of $f - P_Y$ on Y is characterized by the deviation of the best approximation on it being maximal over deviations of best approximations on $n + 1$ point subsets of Y. Hence, if $\{x_0, ..., x_n\}$ is an alternant of $f - P_Y$ on Y, we have $\delta\lambda = O(|Y|^2)$ also.

Let $|Y_k| \to 0$; then for all k sufficiently large there is by Theorem 1 A^k best to f on Y_k, $\rho(A^k) = n$. Further $\| F(A^k, \cdot) - F(A^*, \cdot)\| \to 0$ and by results of Barrar and Loeb [1, Theorem 2], $\{A^k\} \to A^*$. Let $\{x_0^k, ..., x_n^k\}$ be an alternant of $f - F(A^k, \cdot)$ on Y_k. By uniform convergence of $F(A^k, \cdot)$ to $F(A, \cdot)$, the sequence of $(n+1)$-tuples $(x_0^k, ..., x_n^k)$ has a subsequence converging to an alternant $x_0^*, ..., x_n^*$ of $f - F(A^*, \cdot)$. Assume that the sequence of $(n+1)$-tuples is convergent. Assume without loss of generality that $f(x_0^*) - F(A^*, x_0^*) > 0$, then we have

$$(-1)^i (f(x_i^*) - F(A^*, x_i^*)) = \lambda^* \qquad i = 0, ..., n,$$

$$(-1)^i (f(x_i^k) - F(A^*, x_i^k)) \leq \lambda^* \qquad i = 0, ..., n, \qquad (1)$$

$$(-1)^i (f(x_i^k) - F(A^k, x_i^k)) = \lambda^k = \lambda^* + O(|Y_k|^2) \qquad i = 0, ..., n. \qquad (2)$$

Subtracting (1) from (2), we obtain

$$(-1)^i [F(A^k, x_i^k) - F(A^*, x_i^k)] \leq O(|Y_k|^2) \qquad i = 0, ..., n,$$

which can be rewritten as

$$(-1)^i [D(A^*, A^k - A^*, x_i^k) + R(A^*, A^k - A^*, x_i^k)] \leq O(|Y_k|^2), \quad i = 0, ..., n.$$

We can assume without loss of generality that $\{A^k - A^*\}$ is a nonzero sequence. Define

$$C^k = (A^k - A^*)/\| A^k - A^* \|,$$

then $\| C^k \| = 1$. Assume without loss of generality that $\{C^k\} \to C$, $\| C \| = 1$. Suppose

$$(-1)^i D(A^*, C, x_i) \leq 0 \qquad i = 0, ..., n.$$

then the Haar condition is violated. Hence there is i such that $(-1)^i D(A^*, C, x_i) > 0$. By arguments similar to those of Rice [7, p. 24] and continuity, there is $\gamma > 0$ such that

$$(-1)^i D(A^*, A^k - A^*, x_i^k) > \gamma \| A^k - A^* \|$$

for all k sufficiently large. Now by hypothesis on R,

$$\| R(A^*, A^k - A^*, \cdot)\| = O(\| A^k - A^* \|^2),$$

hence

$$(-1)^i D(A^*, A^k - A^*, x_i^k) = O(\| Y_k \|^2) \qquad i = 0,..., n. \tag{3}$$

We now state and prove a generalization of a result of Rice [7, top of page 64]. This generalization was used but not proved in [14].

ASSERTION. Let L be a linear approximating function generated by a Chebyshev set. Let $\{x_i^k\} \to x_i$ for $i = 0,..., n$. Let A^k satisfy

$$(-1)^i L(A^k, x_i^k) \geq -2\delta, \qquad i = 0,..., n. \tag{4}$$

There exists a constant K (independent of δ) such that for all k sufficiently large

$$\| L(A^k, \cdot) \| < K\delta.$$

Proof. By Lemma 2 of [4], (4) implies that $\{L(A^k, \cdot)\}$ is uniformly bounded. Independence of K follows from linearity of L.

From the assertion and (3) we get

$$\| D(A^*, A^k - A^*, \cdot) \| = O(\| Y_k \|^2),$$

hence by Lemma 1-1 of Rice [7, p. 24],

$$\| A^k - A^* \| = O(\| Y_k \|^2),$$

hence

$$\| F(A^k, \cdot) - F(A^*, \cdot)\| = O(\| Y_k \|^2).$$

Let us now consider how restrictive are the hypotheses of Theorem 2. The first major assumption is that the best approximation is of maximum degree. There are two main reasons why this is likely. First, there is usually a low probability that approximations of less than maximum degree are best. For example, in rational approximation the set of functions whose best approximation is of less than maximum degree is nowhere dense [4, p. 109]. Second, approximations of lower degree are usually of simpler form: If they are best, it is usually an indication that the approximating family is inappropriate for approximating the function, since a simpler form approximates just as well. The next assumption is that Y contains all endpoints which are extrema of the error.

Simple cases show that having the endpoints in Y is a good practice; hence this is not restrictive. The final assumption is that E have continuous second partial derivatives about (A^*, x) for $x \in [\alpha, \beta]$. For most f and F of practical interest, continuous partial derivatives of all orders exist.

It appears that the rate of convergence of the theorem cannot be improved even under more restrictive hypotheses, since exactly quadratic convergence has been observed in an example [14, p. 312] of approximation of an analytic function by constants.

Choice of Subset

A more careful look at the proof of Theorem 2 shows that the quality of the best approximation on a subset Y depends only on the closeness of Y to an alternant of $f - F(A^*, \cdot)$. Density merely guarantees closeness. Let $d(Y)$ be the distance of the closest alternant of $f - F(A^*, \cdot)$ from Y. Then in the start of proof of Theorem 2, we have $\delta\lambda = O(d^2(Y))$ and we end with $\| P_X - P_Y \| = O(d^2(Y))$. It follows that it only pays to make Y dense near the extrema of $f - F(A^*, \cdot)$ and points of Y elsewhere are of no benefit (except in assuring us that a larger error does not occur there). It would, therefore, be useful to know where error extrema are most probable, so that our points can be concentrated there. Frequently the extrema of the Chebyshev polynomial of degree n are good estimates of error extrema.

References

1. R. Barrar and H. L. Loeb, On the continuity of the non-linear Tschebyscheff operator, *Pacific J. Math.* **32** (1970), 593–601.
2. C. B. Dunham, Chebyshev approximation with respect to a weight function, *J. Approximation Theory* **2** (1969), 223–232.
3. C. B. Dunham, Approximation by alternating families on subsets, *Computing (Arch. Elektron. Rechnen)* **9** (1972), 261–265.
4. C. B. Dunham, Alternating Chebyshev approximation, *Trans. Amer. Math. Soc.* **178** (1973), 95–109.
5. C. B. Dunham, Varisolvent Chebyshev approximation on subsets, *in* "Approximation Theory" (G. G. Lorentz, ed.), pp. 337–340, Academic Press, New York, 1973.
6. G. Meinardus and D. Schwedt, Nicht-lineare Approximationen, *Arch. Rational Mech. Anal.* **17** (1964), 297–326.
7. J. Rice, "The Approximation of Functions," Vol. 1, Reading, Mass., 1964.
8. C. B. Dunham, Transformed rational Chebyshev approximation, *Numer. Math.* **10** (1967), 147–152.

9. C. B. DUNHAM, Chebyshev approximation with the local Haar condition, *SIAM J. Numer. Anal.* **8** (1971), 749–753.
10. C. B. DUNHAM, Chebyshev approximation by $A\phi(BX)$, *Z. Angew. Math. Mech.* **53** (1973), 353–354.
11. C. B. DUNHAM, Chebyshev approximation by $A + B\phi(CX)$, *Computing, (Arch. Elektron. Rechnen)* **13** (1974), 205–213.
12. G. TEMES AND J. BINGHAM, Iterative Chebyshev approximation technique for network synthesis, *IEEE Trans. Circuit Theory*, **14** (1967), 31–37.
13. J. WILLIAMS, Numerical Chebyshev approximation by interpolating rationals, *Math. Comp.* **26** (1972), 199–206.
14. C. B. DUNHAM, Efficiency of Chebyshev approximation on subsets, *J. Assoc. Comput. Mach.* **21** (1974), 311–313.
15. M. E. BURKE, Nonlinear best approximation on discrete sets, *J. Approximation Theory* **16** (1976), 133–141.