JOURNAL OF APPROXIMATION THEORY 24, 161–168 (1978)

Computation of Transformed Linear Chebyshev Approximations

CHARLES B. DUNHAM

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

Communicated by John R. Rice

Received February 14, 1977

A transformed linear approximation is a function of the form $w(x)\,\phi(L(A, x))$, where $L(A, \cdot)$ is an element of an n-dimensional linear space. Best Chebyshev approximations are characterized when ϕ is an order function. Computation of a best approximation on an $n + 1$ point set is considered. A variant of Stiefel's exchange (ascent) method is proposed for computation of best approximations on finite sets. It is shown that Stiefel's exchange increases the deviation under favorable circumstances. Best approximations on infinite sets can be obtained by discretization.

Let W be a compact Hausdorff space. Let $\{\psi_1, ..., \psi_n\}$ be a linearly independent subset of $C(W)$ and define

$$L(A, x) = \sum_{k=1}^{n} a_k \psi_k(x).$$

Let w be an element of $C(W)$. Let ϕ be a continuous mapping of the real line into the extended real line. Define

$$F(A, x) = w(x)\,\phi(L(A, x)).$$

Such an approximation is called a transformed linear approximation. Let f be given: $f(x) = w(x)\,g(x)$, $g \in C(W)$. The approximation problem is to find A^* to minimize

$$e(A) = \sup\{|f(x) - F(A, x)| : x \in W\}.$$

Such a parameter A^* is called best and $F(A^*, \cdot)$ is called a best approximation to f.

The existence, characterization, and uniqueness problems were considered in a preceding paper [3] under the assumption that $w > 0$. Approximation with respect to a continuous multiplicative weight function can be handled by

0021-9045/78/0242–0161$02.00/0

building the weight function into w. To avoid trivialities, it is assumed that an approximant with finite norm exists.

PRELIMINARIES

DEFINITION. We call ϕ, a continuous mapping from the real line into the extended real line, an *order function* if ϕ is monotonic and is strictly monotonic where it is finite.

Some order functions are:

$$\text{(i)} \quad \phi(y) = \exp(y)$$

$$\text{(ii)} \quad \phi(y) = \log(y) \qquad y > 0$$
$$= -\infty \qquad y \leqslant 0.$$

Others are given in [3]. If ϕ is not an order function, very little of the theory of this paper can be used.

Let $S(W) = \{x : w(x) \neq 0, x \in W\}$.

VANISHING w

In [3], it was assumed that $w > 0$. As there exist cases of practical interest in which this is not the case [6], it appears necessary to widen the theory of that paper.

Let $M(A) = \{x : |f(x) - F(A, x)| = e(A)\}$. $M(A)$ is a nonempty closed set.

THEOREM 1. *Let $0 < e(A) < \infty$. Let ϕ be an order function. A necessary and sufficient condition for A to be best is that no B exist such that*

$$w(x)\, L(B, x)(f(x) - F(A, x)) > 0 \qquad x \in M(A).$$

This is proven using the arguments of [3, Theorem 2], noting that multiplication by w preserves betweenness.

Let $\Phi(x) = (\psi_1(x),..., \psi_n(x))$.

COROLLARY. *Let $e(A) < \infty$ and ϕ be an order function. A necessary and sufficient condition that A be best is that 0 be in the convex hull of*

$$\{(f(x) - F(A, x))\, \Phi(x)\, w(x) : x \in M(A)\}.$$

Reference [3, Lemmas 4, 5, 6, Theorem 3 and its corollary] applies.

THEOREM 2. *Let ϕ be an order function. A sufficient condition for best A* *with finite error to be uniquely best on W is that $\{\psi_1, ..., \psi_n\}$ is a Chebyshev set on $S(W)$.*

Best Approximation on $n+1$ Points Sets

We consider how to obtain the best approximation on a $n+1$ point subset $Y = \{x_0, ..., x_n\}$ of $S(W)$ if $\{\psi_1, ..., \psi_n\}$ is a Chebyshev set on the $n+1$ point set Y and ϕ is an order function. By the corollary to Theorem 1, A is best if $\{(f(x) - F(A, x))\, \Phi(x)\, w(x), x \in Y\}$ contains 0 in its convex hull and $|f(x) - F(A, x)|$ is constant on Y. We now show how to find A.

(1) Find a nontrivial solution $(\lambda_0, ..., \lambda_n)$ to

$$\sum_{i=0}^{n} \lambda_i \Phi(x_i) \operatorname{sgn}(w(x_i)) = 0.$$

One way to do this is to set $\lambda_0 = 1$ and solve

$$\sum_{i=1}^{n} \lambda_i \Phi(x_i) = -\Phi(x_0).$$

This is a linear system of n equations in n unknowns. By the Chebyshev set assumption the matrix is nonsingular and a solution exists. None of the λ's obtained could be zero, since then the Chebyshev set assumption would be violated.

(2) Solve the (nonlinear) system

$$f(x_i) - w(x_i)\, \phi(L(A, x_i)) - \operatorname{sgn}(\lambda_i)\, d = 0 \qquad i = 0, ..., n$$

for unknowns $a_1, ..., a_n, d$. The system can be rewritten as

$$L(A, x_i) - \phi^{-1}[(f(x_i) - \operatorname{sgn}(\lambda_i)\, d)/w(x_i)] = 0 \qquad i = 0, ..., n.$$

Either system may be solved by Newton's method.

A solution to the latter is a solution to the former. Let us suppose that the former has two solutions (A, d) and (B, e), then we have

$$\phi(L(A, x_i)) - \phi(L(B, x_i)) = \operatorname{sgn}(\lambda_i)(e - d)/w(x_i), \qquad i = 0, ..., n.$$

If $e = d$, then $F(A, \cdot)$, $F(B, \cdot)$ are unbounded or $A = B$ by the Chebyshev set assumption. Assume, therefore, that $e \neq d$. Assume without loss of generality that ϕ is monotonic increasing, then for $i = 0, ..., n$

$$\operatorname{sgn}(L(A, x_i) - L(B, x_i)) = \operatorname{sgn}(L(A - B, x_i)) = \operatorname{sgn}(\lambda_i)\, w(x_i)(e - d).$$

This violates choice of $\{\lambda_i\}$ and the theorem on linear inequalities [1, p. 19]. Hence a solution to the former is unique.

If the latter nonlinear equation is solved by Newton's method, the ith row of the $(n + 1) \times (n + 1)$ matrix M of partial derivatives is given by

$$m_{ij} = \phi_j(x_i) \qquad j = 1,..., n$$

$$m_{i,n+1} = \operatorname{sgn}(\lambda_i)[(\phi^{-1})' ((f(x_i) - \operatorname{sgn}(\lambda_i)\, d)/w(x_i))]/w(x_i) = \eta_i(d).$$

THEOREM 3. *Let $N = \{t : |\phi(t)| < \infty\}$ and ϕ^{-1} have a positive continuous derivative on N. Let ϕ be an order function and $\{\psi_1 ,..., \psi_n\}$ be a Chebyshev set on $\{x_0 ,..., x_n\}$. The above matrix of partial derivatives is nonsingular in a neighborhood of the solution.*

Proof. It follows that if the matrix is singular, there exists $\{b_1 ,..., b_{n+1}\}$ not all zero such that

$$L(B, x_i) + \eta_i(d)\, b_{n+1} = 0 \qquad i = 0,..., n.$$

By the Chebyshev set assumption, b_{n+1} cannot be zero. Assume without loss of generality that $b_{n+1} < 0$, then

$$\operatorname{sgn}(L(B, x_i)) = \operatorname{sgn}(\lambda_i w(x_i)) \qquad i = 0,..., n.$$

But by choice of $\{\lambda_j\}$ and the theorem on linear inequalities [1, p. 19], this is impossible.

Remark. If the derivative is negative on N, the theorem also holds.

ERROR DETERMINING SETS

For Y a compact subset of W define

$$\| g \|_Y = \sup\{| g(x)| : x \in Y\}, \qquad \rho(Y) = \inf\{\|f - F(A, \cdot)\|_Y : A \in E_n\}.$$

It is a consequence of the corollary to Theorem 1 and the theorem of Carathéodory [1, p. 17], that

THEOREM 4. *Let there exist a best approximation on W and ϕ be an order function. There exists an $n + 1$ point subset T of W such that $\rho(W) = \rho(T)$ and any best approximation to f on W is best on T.*

A best approximation on finite W can be determined by determining the best approximation on every $n + 1$ point subset Y of W and at the same time determining $\rho(Y)$. The best approximation on the Y maximizing ρ is the best approximation on W. This is an infallible but impractical method: It is impractical because the number of $n + 1$ point subsets is usually astronomical. Reflection shows that all we need is a sequence of $n + 1$ point sets X_k such that $\rho(X_k)$ is an increasing sequence. This suggests use of a generalization of Stiefel's ascent method for linear approximation [1, pp. 46–47; 5, pp. 173–176].

The Ascent Method

(i) Choose an initial set X_0 of $n + 1$ points from $S(W)$ and set $k = 0$.

(ii) Determine a best parameter A^0 to f on X_0 and $\rho(X_0)$.

(iii) Find y_k such that $|f(y_k) - F(A^k, y_k)| = e(A^k)$.

(iv) If $\rho(X_k) = e(A^k)$, stop.

(v) Find an $n + 1$ point subset X_{k+1} of $y_k \cup X_k$ such that $\rho(X_{k+1}) > \rho(X_k)$, together with A^{k+1} best on X_{k+1}.

(vi) Add 1 to k and go to (iii).

When we stop on (iv), we have a best approximation. When applied to finite W, the algorithm (if it runs) must eventually stop on (iv) since only finitely many $n + 1$ point subsets exist. It should be noted, however, that even in the linear case, we cannot always guarantee that step (v) can be done if step (iv) is passed [5, p. 256]. We need further assumptions to ensure this.

Lemma. *If there exists a unique best approximation to f on X_k and step* (iv) *of the ascent method is passed, there exists X_{k+1} as required for step* (*v*).

Proof. In view of the preceding theorem it suffices to show that $\rho(X_k \cup y_k) > \rho(X_k)$. Since $X_k \subset X_k \cup y_k$, we have $\rho(X_k \cup y_k) \geqslant \rho(X_k)$. Thus we need only consider the possibility that $\rho(X_k \cup y_k) = \rho(X_k)$. If this is so, there exists A best on X_k with

$$|f(y_k) - F(A, y_k)| \leqslant \rho(X_k).$$

But the best approximation on X_k is unique, so this implies that we stop on step (iv).

Theorem 5. *Let there exist a best approximation to f on all $n + 1$ point subsets of finite W. Let ϕ be an order function and $\{\psi_1, \dots, \psi_n\}$ be a Chebyshev set on $S(W)$. Then the ascent method converges to the unique best approximation on X in a finite number of iterations.*

We consider how X_{k+1} is to be found. X_{k+1} is obtained by replacing a suitable element of X_k by y_k. In the linear case with a Chebyshev set, Stiefel has found an exchange rule giving a suitable point to drop. Sufficient conditions are given in the next section for Stiefel's exchange to give such a point. In the absence of such a rule, we must determine best approximations on $n + 1$ point subsets of $X_k \cup y_k$ until one with a deviation ρ larger than $\rho(X_k)$ is found.

We may wish to run the ascent method on a set W on which $\{\psi_1, ..., \psi_n\}$ is not a Chebyshev set. No trouble will occur if we do not hit a set X_k on which the Chebyshev set condition fails. Suppose that the Chebyshev set condition is satisfied on every $n + 1$ point maximum of ρ, then there is $M < \rho(W)$ such that if Y has $n + 1$ points and $\rho(Y) > M$, the Chebyshev set condition is satisfied on Y. If we then start the ascent method on a finite set W with such a Y, we get certain convergence.

Stiefel's Exchange

Let ϕ be an order function. Let $X = \{x_0, ..., x_n\}$ be a subset of $S(W)$, $F(A, \cdot)$ be best on X, $|f(x_{n+1}) - F(A, x_{n+1})| > \rho(X)$, and $\{\psi_1, ..., \psi_n\}$ be a Chebyshev set on $X \cup x_{n+1}$. The Chebyshev set hypothesis guarantees uniqueness of $F(A, \cdot)$ and by arguments used to guarantee uniqueness,

$$|f(x_i) - F(A, x_i)| = \rho(X) \qquad i = 0, ..., n.$$

To implement a one-for-one exchange algorithm, we would like an $n + 1$ point subset Y of $X \cup x_{n+1}$ such that $\rho(Y) > \rho(X)$. Such a subset Y is obtained by discarding a suitable element x_j of X and replacing it by x_{n+1}. A procedure which accomplishes this in the case of linear approximation is the exchange procedure of Stiefel (Cheney [1, p. 46], Rice [4, p. 175]). From the summary of Cheney [1, p. 46], it is clear that Stiefel's exchange can be applied to transformed linear approximation (we use the linear family to which the transformation is applied).

Theorem 6. *Let ϕ be an order function. Let $X = \{x_0, ..., x_n\}$, $F(A, \cdot)$ be best on X, $|f(x_{n+1}) - F(A, x_{n+1})| > \rho(X)$, and $\{\psi_1, ..., \psi_n\}$ be a Chebyshev set on $X \cup \{x_{n+1}\}$. Let there exist a best approximation to f on all $n + 1$ point subsets of $X \cup x_{n+1}$. The $n + 1$ point set Y produced by Stiefel's exchange has the property that $\rho(Y) > \rho(X)$.*

Proof. The argument is similar to that of Cheney [1, p. 47]. Let x_j be the point discarded by Stiefel's exchange and $Y = (X \sim x_j) \cup x_{n+1}$. Let $F(B, \cdot)$ be best on Y with deviation $\rho(Y)$. By the Chebyshev set hypothesis $F(B, \cdot)$ is unique and by arguments used to guarantee uniqueness

$$|f(x_i) - F(B, x_i)| = \rho(Y) \qquad i = 0, ..., n + 1, i \neq j.$$

Let $x_i \in X \sim x_j$, then

$$|f(x_i) - F(A, x_i)| = \rho(X)\,|f(x_{n+1}) - F(A, x_{n+1})| > \rho(X),$$
$$|f(x_i) - F(B, x_i)| = \rho(Y)\,|f(x_{n+1}) - F(B, x_{n+1})| = \rho(Y),$$

so $A \neq B$. By the characterization theorem, 0 is in the convex hull of $\sigma_i \Phi(x_i)$, $i = 0,..., n$, where $\sigma_i = \operatorname{sgn}((f(x_i) - F(A, x_i))/w(x_i))$, hence by the exchange theorem [1, p. 45], 0 is in the convex hull of

$$\{\sigma_i \Phi(x_i) : i = 0,..., n + 1, i \neq j\}. \tag{0}$$

It follows from the characterization theorem that either

$$\operatorname{sgn}(f(x_i) - F(B, x_i)) = \sigma_i \qquad i = 0,..., n + 1, i \neq j,$$

or

$$\operatorname{sgn}(f(x_i) - F(B, x_i)) = -\sigma_i \qquad i = 0,..., n + 1, i \neq j.$$

We must, therefore, have

$$f(x_i) - F(B, x_i) = \sigma_i e \qquad i = 0,..., n + 1, i \neq j. \tag{1}$$

From the previous discussion, we have

$$\begin{aligned} f(x_i) - F(A, x_i) &= \sigma_i \rho(X), \qquad i = 0,..., n, \\ f(x_{n+1}) - F(A, x_{n+1}) &> \sigma_{n+1} \rho(X). \end{aligned} \tag{2}$$

Subtracting (1) from (2), we get

$$\begin{aligned} F(B, x_i) - F(A, x_i) &= \sigma_i(\rho(X) - e) \qquad i = 0,..., n, i \neq j \\ &> \sigma_{n+1}(\rho(X) - e) \qquad i = n + 1. \end{aligned} \tag{3}$$

If $\rho(X) = e$, then (3) implies

$$F(B, x_i) - F(A, x_i) = 0,$$
$$L(A, x_i) - L(B, x_i) = L(A - B, x_i) = 0, \qquad i = 0,..., n, i \neq j,$$

and by the Chebyshev set assumption, $A = B$. But we have already proved that $A \neq B$, so this is impossible. Next let $e < \rho(X)$, then

$$\sigma_i(F(B, x_i) - F(A, x_i)) > 0 \qquad i = 0,..., n + 1, i \neq j.$$

Assume without loss of generality that ϕ is strictly increasing, then

$$\sigma_i L(B - A, x_i) > 0 \qquad i = 0,..., n + 1, i \neq j.$$

By the theorem on linear inequalities [1, p. 19], this implies that 0 is not in the convex hull of (0), contrary to what we have shown.

DISCRETIZATION

DEFINITION. Let $X_1, X_2, \ldots,$ be a sequence of closed subsets of W. We say $\{X_k\} \to W$ if for any $x \in W$, there is a sequence $\{x_k\} \to x$, $x_k \in X_k$.

THEOREM 7. *Let* $|\phi(t)| \to \infty$ *as* $|t| \to \infty$. *Let* $\{\psi_1, \ldots, \psi_n\}$ *be independent on* X_1. *Let* $X_1 \subset X_2 \subset \cdots$ *and* $\{X_k\} \to W$. *Let* A^k *be best to* f *on* X_k. *Then* $\{A^k\}$ *has an accumulation point and any accumulation point of* $\{A^k\}$ *is best on* W.

This follows by arguments similar to those of [2].

This theorem suggests that best approximations on infinite sets be determined as a limit of best approximations on finite sets. For example, if $W = [0, 1] \times [0, 1]$, we could let $Y_k = \{0, \frac{1}{2}^k, \ldots, 1 - \frac{1}{2}^k, 1\}$, $X_k = Y_k \times Y_k$.

REFERENCES

1. E. W. CHENEY, "Introduction to Approximation Theory," McGraw-Hill, New York, 1966.
2. C. B. DUNHAM, The limit of non-linear Chebyshev approximation on subsets, *Aequationes Math.* **9** (1973), 60–63.
3. C. B. DUNHAM, Transformed linear Chebyshev approximation, *Aequationes Math.* **12** (1975), 6–11.
4. J. RICE, "The Approximation of Functions," Vol. 1, Addison-Wesley, Reading, Mass., 1964.
5. J. RICE, "The Approximation of Functions," Vol. 2, Addison-Wesley, Reading, Mass., 1969.
6. J. WILLIAMS, Numerical Chebyshev approximation by interpolating rationals, *Math. Comp.* **26** (1972), 199–206.