JOURNAL OF APPROXIMATION THEORY 4, 1–5 (1971)

Simultaneous Optimization of Absolute and Relative Errors in Functional Approximation

ALEX BACOPOULOS

Département d'Informatique, Université de Montréal, Case Postale 6128, Montréal 101

Communicated by John R. Rice

Received November 19, 1968

This paper is motivated by the problem of optimizing simultaneously the absolute and relative errors arising in functional approximation with respect to the supremum norm. Two natural formulations of the problem are studied, and answers, both positive and negative, are given to the usual questions of existence, uniqueness and characterization of best approximations.

DEFINITIONS

Let X be a compact subset of $[a, b]$ containing at least $n + 1$ points. Let $C(X)$ denote the Banach algebra of all real-valued continuous functions defined on X with the norm $\| g \| = \max\{| g(x)| : x \in X\}$, and let M be an n-dimensional Haar subspace of $C[a, b]$. That is, M is an n-dimensional linear subspace of $C[a, b]$ such that the zero function is the only function in M which vanishes at n distinct points of $[a, b]$. We shall assume throughout this paper that the functions $\varphi_1(x),..., \varphi_n(x)$ form a basis for M.

Let $f, w_1, w_2 \in C(X)$ be given functions: f is the function to be approximated, and w_1 and w_2 are positive (weight) functions. We define two new norms by

$$\| g \|_m = \max\{\| w_1 g \|, \| w_2 g \|\}$$

and

$$\| g \|_s = \| w_1 g \| + \| w_2 g \|.$$

Then $p \in M$ is said to be a best *max approximation* to f provided

$$\| f - p \|_m = \inf_{q \in M} \| f - q \|_m .$$

Similarly, p is said to be a best *sum approximation* to f provided

$$\| f - p \|_s = \inf_{q \in M} \| f - q \|_s .$$

An application of special interest which motivated this work is obtained by setting $w_1 \equiv 1$ and $w_2 = 1/f$. (The function f to be approximated is assumed here to have a constant sign on X).

Given f, w_1 and w_2 as above, we define the set X_p of "critical" points of a sum approximation $p \in M$, as follows: $X_p = X_{+1} \cup X_{+2} \cup X_{-1} \cup X_{-2}$, where

$$\begin{aligned}
X_{+1} &= \{x \in X : w_1(x)(f(x) - p(x)) = \| w_1(f - p)\|\}, \\
X_{+2} &= \{x \in X : w_2(x)(f(x) - p(x)) = \| w_2(f - p)\|\}, \\
X_{-1} &= \{x \in X : w_1(x)(f(x) - p(x)) = -\| w_1(f - p)\|\}, \\
X_{-2} &= \{x \in X : w_2(x)(f(x) - p(x)) = -\| w_2(f - p)\|\}.
\end{aligned}$$

Chebychev-Type Theories

The existence of a best max and a best sum approximation follows from the finite dimensionality of the linear space M, and the fact that $\| \cdot \|_m$ and $\| \cdot \|_s$ are bona fide norms. In what follows it is shown that max approximation is reducible to ordinary weighed Chebychev approximation while sum approximation is not. In fact, it is shown that, for the latter, uniqueness fails in general, and a generalized oscillation of $f - p$ is necessary but not sufficient for p to be a best sum approximation.

Theorem 1. *A best max approximation to f is unique and is equal to the best (ordinary) weighed Chebychev approximation to f with respect to the weight function*

$$w_3(x) = \max\{w_1(x), w_2(x)\}.$$

Proof. Denote by p_1, p_2 and p_3 the unique best (ordinary) approximation with respect to w_1, w_2 and w_3, respectively. We distinguish three cases.

Case 1.

$$\| w_2(f - p_1)\| < \| w_1(f - p_1)\|.$$

It is clear that the desired best approximation p_3 equals p_1 and, therefore, there is nothing to prove.

Case 2.

$$\| w_2(f - p_2)\| > \| w_1(f - p_2)\|.$$

Similarly, $p_3 = p_2$.

Case 3. If cases 1 and 2 do not hold, we observe that a best max approximation p has to satisfy the condition

$$\| w_1(f-p)\| = \| w_2(f-p)\|.$$

This follows from the continuity in $r \in M$ of the nonlinear functional

$$G(r) = \| w_1(f-r)\| - \| w_2(f-r)\|,$$

and from the connectivity of M. Let $x_1 < x_2 < \cdots < x_{n+1}$ be ordinary critical points of $f - p_3$ with respect to w_3, satisfying $\sigma(x_i) = (-1)^{i+1}\sigma(x_1)$, and assume that there exists a $q \in M$ such that $\|f-q\|_m < \|f-p_3\|_m$. By the above condition and the positivity of w_1 and w_2 it follows that $(-1)^i [q(x_i) - p_3(x_i)]$ is $\geqslant 0$ for all i or $\leqslant 0$ for all i. Now, a continuity argument described in [2, p. 61] and the fact that M is a Haar space imply that $p_3 = q$, a contradiction.

To complete the proof we apply the usual arguments of alternation, from which we derive that there exists no best max approximation $\neq p_3$.

THEOREM 2. *Let $f \in C(X) - M$, let $p \in M$, and consider the following statements*:

(a) *p is a best sum approximation to f.*

(b) *The origin of Euclidean n-space belongs to the convex hull of $\{\sigma(x) \cdot \hat{x} : x \in X_p\}$, where $\sigma(x) = -1$ if $x \in X_{-1} \cup X_{-2}$, $\sigma(x) = +1$ if $x \in X_{+1} \cup X_{+2}$, and $\hat{x} = (\varphi_1(x),\ldots, \varphi_n(x))$.*

(c) *There exist $n+1$ points $x_1 < x_2 < \cdots < x_{n+1}$ in X_p, satisfying $\sigma(x_i) = (-1)^{i+1}\sigma(x_1)$.*

Then, (a) $\Rightarrow$ (b) $\Rightarrow$ (c) $\Rightarrow$ (b) *but* (c) $\not\Rightarrow$ (a).

Proof. (a) $\Rightarrow$ (b). Assume that $0 \notin$ the convex hull of $\{\sigma(x) \cdot \hat{x} : x \in X_p\}$. Since X_p is compact, it follows from a theorem on linear inequalities [1, p. 19] that there exists a $q \in M$ such that $\sigma(y)\, q(y) > 0$ for all $y \in X_p$. We shall show that there exists a $\lambda > 0$ for which $r_\lambda = p + \lambda q \in M$ satisfies $\|f - r_\lambda\|_s < \|f-p\|_s$.

Let $s(x) = \operatorname{sgn}(f(x) - p(x))$ and $\delta = \min\{s(x)\, q(x) : x \in X_p\}$; then $\delta > 0$. For $i = 1, 2$, let

$$Y_i = \{x \in X : |\, w_i(x)(f(x) - p(x))| > \| w_1(f-p)\|/2 \text{ and } s(x)\, q(x) > \delta/2\}.$$

Y_i is open and contains X_p; thus $|\, w_i(x)(f(x)-p(x))| < \| w_i(f-p)\|$ on the compact set $X - Y_i$. Therefore, by continuity, there exists a $\lambda_i > 0$ such that $0 \leqslant \lambda \leqslant \lambda_i$ implies

$$\max_{x \in X - Y_i} |\, w_i(x)(f(x) - r_\lambda(x))| < \| w_i(f-p)\|.$$

Letting Z_i be the closure of Y_i, we see that $x \in Z_i$ implies $s(x)\, q(x) \geqslant \delta/2$ and $|\, w_i(x)(f(x) - p(x))| \geqslant \| w_i(f - p)\|/2$. Now choose $\mu_i > 0$ such that $0 \leqslant \lambda \leqslant \mu_i$ implies $\| w_i(p - r_\lambda)\| < \| w_i(f - p)\|/2$. Then for $x \in Z_i$ and $0 < \lambda \leqslant \mu_i$ we have that $\operatorname{sgn}(f(x) - r_\lambda(x)) = \operatorname{sgn}(f(x) - p(x))$. Setting $\lambda = \min\{\lambda_1, \lambda_2, \mu_1, \mu_2\}$, we observe that $\|f - r_\lambda\|_s < \|f - p\|_s$, a contradiction.

(b) $\Leftrightarrow$ (c). The proof is similar to that in [1, pp. 74–75]. The arguments there involving alternations and convex hull continue to be valid here if we replace ordinary extrema by the points of X_p.

(c) $\not\Rightarrow$ (a). To see that (c) does not imply that p is a best sum approximation to f, let p_1 be the polynomial of degree $\leqslant 1$ which best approximates on $[-1, 1]$, in the sup norm, the function $f(x) = x^2$. Then $2(f(x) - p_1(x))$ is the Chebychev polynomial $T_2(x) = 2x^2 - 1$. Define:

$$w_1 \equiv 1 \qquad \text{on } [-1, 1]$$

and

$$w_2 = \begin{cases} \text{const } 1/5 & \text{on } [-1, 1 - \epsilon], \\ \text{the function whose graph is the} & \\ \text{line segment joining } (1 - \epsilon, \tfrac{1}{5}) \text{ to } (1, 1) & \text{on } [1 - \epsilon, 1]. \end{cases}$$

Observe that the error function $f(x) - p_1(x) = (2x^2 - 1)/2$ satisfies (c) at -1, 0 and 1. Yet, it follows that for small enough $\epsilon > 0$,

$$\left\| f(x) - p_1(x) - \frac{x}{4} - \frac{1}{8} \right\|_s = \frac{41}{64} + \frac{1}{8} = \frac{49}{64} < \|f - p_1\|_s = \frac{1}{2} + \frac{1}{2} = 1.$$

Thus $p_1(x) + x/4 + 1/8$ is a better sum approximation to the function $f(x) = x^2$ than $p_1(x)$.

PROPOSITION. *Best sum approximations are not generally unique.*

Proof. Consider the following simple example which was communicated to the author by G. D. Taylor:

Let $f(x) = x$, $M =$ the set of constant functions on $[0, 1]$, $[a, b] = [0, 1]$, $w_1 \equiv 1$ and

$$w_2 = \begin{cases} \text{const } 1/3 & \text{on } [0, \tfrac{1}{2} - \epsilon], \\ \text{the function whose graph} & \\ \text{is the line segment joining } (\tfrac{1}{2} - \epsilon, \tfrac{1}{3}) \text{ to } (\tfrac{1}{2}, 1) & \text{on } [\tfrac{1}{2} - \epsilon, \tfrac{1}{2}], \\ \text{const } 1 & \text{on } [\tfrac{1}{2}, 1]. \end{cases}$$

It is seen that for sufficiently small $\epsilon > 0$, all A, $1/2 \leqslant A \leqslant 3/4$, are best sum approximations.

COMMENTS

For simplicity of exposition we have used a Haar space instead of more general approximation classes. Actually, in the case of sum approximation by varisolvent families, the relationship of Theorem 2 between best approximation and generalized alternation continues to hold.

Finally, we would like to remark that condition (c) of Theorem 2 is so strong that it is "almost sufficient" for p to be a best sum approximation to f. In fact, in [3], condition (c) is used to derive a nontrivial generalization of the Remes Algorithm which computes efficiently all the best polynomial sum approximations to f by scanning a closed interval of the real line.

REFERENCES

1. E. W. CHENEY, "Introduction to Approximation Theory," McGraw-Hill, New York, 1966.
2. J. R. RICE, "The Approximation of Functions," Vol. I, Addison-Wesley, Reading, Mass., 1964.
3. A. BACOPOULOS AND B. GAFF, On the reduction of a problem of minimization of $n + 1$ variables to a problem of one variable, *SIAM J. Numer. Anal.* **8** (1971).