JOURNAL OF APPROXIMATION THEORY 11, 134–142 (1974)

Nonlinear Mean Approximation

CHARLES B. DUNHAM

Computer Science Department, University of Western Ontario, London 72, Canada

Communicated by E. W. Cheney

The properties of best nonlinear approximations with respect to a generalized integral norm on an interval are studied. A necessary condition for an approximation to be locally best is obtained. The interpolatory properties of best approximations are related to the dimension of a Haar subspace in the tangent space. A sufficient condition for an approximation to be best only to itself is given for a class of norms including the L_p norms, $1 < p < \infty$. A sufficient condition for the set of points at which the given approximated function and an approximant agree to be of positive measure is given. The results are applied to approximation by exponential families V_n: in the case of L_p approximation, $1 < p < \infty$, degenerate approximations are best only to themselves and the error of a best approximation is either identically zero or has $2n$ sign changes.

Let τ be a continuous nonnegative function, $\tau(0) = 0$. Let $\int$ denote the integral on $[\alpha, \beta]$. For $g \in C[\alpha, \beta]$, define

$$N(g) = \int \tau(g).$$

Let F be an approximating function with parameter $A = (a_1, ..., a_n)$ taken from a parameter space P, a subset of n-space, such that $F(A, \cdot) \in C[\alpha, \beta]$ for all $A \in P$. The approximation problem is: given $f \in C[\alpha, \beta]$, to find a parameter $A^* \in P$ minimizing $e(A) = N(f - F(A, \cdot))$ over P. Any such parameter A^* is called best and $F(A^*, \cdot)$ is called a best approximation.

The problem of linear approximation with respect to a τ-norm has been studied by Motzkin and Walsh [7], who required that $\tau(t) = \tau(-t)$ and that τ have a continuous second derivative. Motzkin later studied a more general problem in [12]. The problem of rational approximation with respect to a τ-norm has been studied by the author [2], and the results of this paper are generalizations of the results of that paper. That paper, in turn, owed much to the paper of Cheney and Goldstein [10] on mean square rational approximation.

Necessary Conditions for Local Minima

We assume henceforth that τ has a continuous first derivative when restricted to $(-\infty, 0]$ or to $[0, \infty)$. τ may not have a derivative at zero, but has a left-hand derivative $\tau_-'(0)$ and a right-hand derivative $\tau_+'(0)$ at zero. We assume further that $\operatorname{sgn}(\tau'(t)) = \operatorname{sgn}(t)$ for $t \neq 0$. For example, in the case of L_p approximation, $1 \leqslant p < \infty$, $\tau(t) = |t|^p$, $\operatorname{sgn}(\tau'(t)) = \operatorname{sgn}(t)$ for $t \neq 0$. It will be convenient to define $\tau_0'(0) = 0$.

Let us define

$$\eta(A, B) = \lim_{\lambda \to 0+} \frac{\int \tau(f - F(A + \lambda B, \cdot)) - \int \tau(f - F(A, \cdot))}{\lambda},$$
$$= \lim_{\lambda \to 0+} \int \frac{\tau(f - F(A + \lambda B, \cdot)) - \tau(f - F(A, \cdot))}{\lambda}. \quad (1)$$

Let a neighborhood of A in n-space be in P. Then a necessary condition for A to be a local minimum of e is that $\eta(A, B) \geqslant 0$ for all B. This makes it desirable to have a more convenient formula for $\eta(A, B)$. We define a parameter norm,

$$\| A \| = \max\{| a_i |: 1 \leqslant i \leqslant n\}.$$

Definition. Let there exist continuous partial derivatives F_k of F with respect to parameter component a_k of A. Define

$$D(A, B, x) = \sum_{k=1}^{n} b_k F_k(A, x),$$
$$R(A, B, x) = F(A + B, x) - F(A, x) - D(A, B, x),$$

and let $R(A, B, x) = o(\| B \|)$ as $\| B \| \to 0$. Let a neighbourhood of A in n-space be in P. We say F is *locally linear* at A.

A similar definition is used in [8] and a similar condition is assumed in [6, 306–307].

Definition. Let $Z(A) = \{x: f(x) - F(A, x) = 0\}$ and

$$\sim Z(A) = [\alpha, \beta] \sim Z(A).$$

Lemma 1. *Let F be locally linear at A and the zeros of $D(A, B, \cdot)$ be a set of measure zero, then*

$$\eta(A, B) = \int_{\sim Z(A)} - \tau'(f - F(A, \cdot))\, D(A, B, \cdot)$$
$$- \int_{Z(A)} \tau'_{-\operatorname{sgn}(D(A,B,\cdot))}(0)\, D(A, B, \cdot).$$

Proof. There is $\mu > 0$ such that $F(A + \lambda B, \cdot) \in C[\alpha, \beta]$ for $0 \leqslant \lambda \leqslant \mu$. The integrand in the formula (1) for $\eta(A, B)$ is bounded above in absolute value by JK:

$$J = \sup\{| \tau'(f(x) - F(A + \lambda B, x))| : 0 \leqslant \lambda \leqslant \mu, \alpha \leqslant x \leqslant \beta\},$$
$$K = \sup\{| F(A + \lambda B, x) - F(A, x)|/\lambda : 0 \leqslant \lambda \leqslant \mu, \alpha \leqslant x \leqslant \beta\}.$$

If we denote by $| \tau'(0)|$ the quantity $\max\{\tau_-'(0), \tau_+'(0)\}$ then $| \tau' |$ is upper semicontinuous. The quantity we take the supremum of in getting J is then upper semicontinuous and the supremum is taken on a compact set, so there is a point where the supremum is attained, hence J is finite.

$$\begin{aligned}(F(A + \lambda B, x) - F(A, x))/\lambda &= (D(A, \lambda B, x) + R(A, \lambda B, x))/\lambda \\ &= D(A, B, x) + o(\lambda)/\lambda,\end{aligned}$$

so K is the supremum of a continuous function on a compact set and hence is finite. We have JK finite, so by the Lebesque dominated convergence theorem,

$$\begin{aligned}\eta(A, B) &= \int \lim_{\lambda \to 0+} \frac{\tau(f - F(A + \lambda B, \cdot)) - \tau(f - F(A, \cdot))}{\lambda} \\ &= \int \frac{\partial}{\partial \lambda} \tau(f - F(A + \lambda B, \cdot))]_{\lambda = 0+}.\end{aligned}$$

If

$$f(x) - F(A, x) \neq 0, \qquad D(A, B, x) \neq 0,$$

the integrand is $-\tau'(f(x) - F(A, x))\, D(A, B, x)$. If $f(x) - F(A, x) = 0$, $D(A, B, x) \neq 0$ the integrand is

$$\begin{aligned}&\lim_{\lambda \to 0+} \frac{\tau(-D(A, \lambda B, x) - R(A, \lambda B, x)) - \tau(0)}{\lambda} \\ &\quad = \lim_{\lambda \to 0+} \frac{\tau(\lambda[-D(A, B, x) - R(A, \lambda B, x)/\lambda]) - \tau(0)}{\lambda} \\ &\quad = -\tau'_{-\operatorname{sgn}(D(A,B,x))}(0)\, D(A, B, x).\end{aligned}$$

As the set of zeros of $D(A, B, \cdot)$ form a set of measure zero, we do not need the integrand on this set.

Interpolating Properties of Best Approximations

The linear space $\{D(A, B, \cdot) : B \in E_n\}$ is the tangent space of F at A, which is used extensively by Meinardus and Schwedt [6, 307 ff] under different notation. In view of the theorem to be proved shortly, it is important to know how large a Haar subspace this linear space contains and in particular whether it is a Haar subspace. Unfortunately, this is known only for very few F. The author has obtained such results for several more F [4].

DEFINITION. A linear subspace of dimension m of $C[\alpha, \beta]$ is a Haar subspace on the open interval (α, β) if every non-zero element of the subspace has at most $m - 1$ zeros on (α, β).

DEFINITION. A continuous function g is said to *change sign* at x if x is a zero of g interior to $[\alpha, \beta]$ and for all sufficiently small $\epsilon > 0$, $g(x - \epsilon) \cdot g(x + \epsilon) < 0$.

The following result can be proven by elementary arguments (see, for example, the remark in [7, 1228]).

LEMMA 2. *Let L be a Haar subspace of dimension m on (α, β) then for any $p < m$ interior points there is a nonzero element of L changing sign at the p points with no other zeros in (α, β).*

THEOREM 1. *Let F be locally linear at A. Let A be a local minimum of e and $f \neq F(A, \cdot)$. If $\{D(A, B, \cdot): B \in E_n\}$ contains a Haar subspace of dimension m on (α, β), then*

(i) *$f - F(A, \cdot)$ has m sign changes, or*

(ii) $\max\{-\tau_-'(0), \tau_+'(0)\} > 0$ *and* $\mu(Z(A)) > 0$.

Proof. Suppose $f - F(A, \cdot)$ has sign changes only at p points, $x_1, \ldots, x_p$, $p < m$, and one of the following is true:

(i) $\tau'(0) = 0$,

(ii) $\mu(Z(A)) = 0$.

By Lemma 2 there exists B such that $D(A, B, \cdot)$ changes sign only at $x_1, \ldots, x_p$ and has no other zeros in (α, β). As $\tau'(f - F(A, \cdot))$ changes sign only at these points, we can assume that $D(A, B, \cdot)$ is of the same sign as $\tau'(f - F(A, \cdot))$ except possibly on $Z(A)$ or $\{\alpha, \beta\}$. We have by Lemma 1,

$$\eta(A, B) = \int_{\sim Z(A)} - \tau'(f - F(A, \cdot))\, D(A, B, \cdot).$$

The integrand is negative and continuous on $(\alpha, \beta) \sim Z(A)$, hence $\eta(A, B) < 0$ and A is not a local minimum of e.

It should be noted that interpolation results have also been obtained by Rice [11, Chap. 13]. These involve different hypotheses and an entirely different approach.

ORTHOGONAL COMPLEMENTS

In this section we show that some common nonlinear families have orthogonal complement of zero. The results are applicable to the theory to be developed in the following section.

Let $BM[\alpha, \beta]$ be the bounded measurable functions on $[\alpha, \beta]$. A family S of continuous functions is said to have *orthogonal complement* of zero (in $BM[\alpha, \beta]$) if the only elements g of $BM[\alpha, \beta]$ for which

$$\int gh = 0$$

for all $h \in S$ are elements vanishing almost everywhere.

DEFINITION. A set of powers is called *fundamental* on a set S of functions, if for any $g \in S$, there is a sequence $\{h_k\}$ of linear combinations of the powers such that $\| g - h_k \|_\infty \to 0$.

EXAMPLES.

1. The powers $\{1, x, x^2, ...\}$ are fundamental in $C[\alpha, \beta]$ by the Weierstrass theorem.
2. The even powers $\{1, x^2, x^4, ...\}$ are fundamental in $C[0, \alpha]$.
3. The odd powers $\{x, x^3, x^5, ...\}$ are fundamental in $CZ[0, \alpha]$, where $CZ[\alpha, \beta]$ is the functions continuous on $[\alpha, \beta]$ which vanish at zero.
4. The powers $\{x, x^2, x^3, ...\}$ are fundamental in $CZ[\alpha, \beta]$.

LEMMA 3. *Let $\{x^{k(0)}, x^{k(1)}, ...\}$ be fundamental in $CZ[\alpha, \beta]$. For $g \in BM[\alpha, \beta]$, the conditions*

$$\int gx^{k(i)} = 0, \qquad i = 0, 1, ...,$$

imply that $g = 0$ almost everywhere.

Proof. The convergence of a sequence with respect to the sup norm on $[\alpha, \beta]$ implies convergence with respect to the L_1 norm on $[\alpha, \beta]$. Also the continuous functions vanishing at zero are dense in $L_1[\alpha, \beta]$. We have for h a linear combination of powers

$$\int g^2 = \int gh + \int g(g - h) \leqslant \sup\{| g(x)| : \alpha \leqslant x \leqslant \beta\} \| g - h \|_1 ,$$

hence $\int g^2 = 0$ and $g = 0$ almost everywhere.

THEOREM 2. *Let ψ have a Taylor series expansion*

$$\psi(x) = \sum_{k=0}^{\infty} a_k x^k$$

about zero with radius of convergence $R > 0$. Let the coefficients of a sequence of integer powers fundamental in $CZ[\alpha, \beta]$ be nonzero. Let $\mu > 0$, then the orthogonal complement of $\{\psi(\delta x): -\mu < \delta < \mu\}$ is zero.

Proof. Let $\int g\psi(\delta x) = 0$ for $-\mu < \delta < \mu$. We have

$$0 = \int g \left(\sum_{k=0}^{\infty} a_k (\delta x)^k \right)$$

for $|\delta| < R/\max\{|\alpha|, |\beta|\}$, hence for such δ,

$$0 = \sum_{k=0}^{\infty} a_k \left[\int g x^k \right] \delta^k,$$

and since $a_{k(i)} \neq 0$

$$\int g x^{k(i)} = 0, \qquad i = 0, 1, \ldots .$$

By the previous lemma, $g = 0$ almost everywhere. A consequence of the theorem is that for any common transcendental function ψ which is analytic at zero, the orthogonal complement of $\{\psi(\delta x): -\mu < \delta < \mu\}$ is zero.

APPROXIMATIONS WHICH ARE BEST ONLY TO THEMSELVES

In standard cases of Chebyshev approximation, in particular alternating approximation [11, Chap. 7], every approximation is best to a function which is not itself. In even the simplest cases of nonlinear L_p approximation, $1 \leqslant p < \infty$, this may be no longer true. Particular cases where some approximations are best only to themselves are given in [1, 227; 2; 10, 239] and in the section on exponential approximation in this paper.

It would be desirable to have a theory telling what approximations are best only to themselves. A start at such a theory follows in this section, where we show that certain approximations are best only to themselves. To complete the theory, we would have to show that the remaining approximations were best to some other function. This has been done by Cheney and Goldstein [10, 238] for mean square approximation by ordinary rational functions.

DEFINITION. The *sum space* of an approximation $F(A, \cdot)$ is the set of functions h such that $F(A, \cdot) + \lambda h$ is an approximant for all $|\lambda|$ sufficiently small.

EXAMPLE. Let ψ be a continuous function and let

$$V_n(\psi) = \left\{ \sum_{k=1}^{n} a_k \psi(a_{n+k} x) \right\}.$$

An element of $V_n(\psi)$ with one of $a_1, \dots, a_n$ zero has $\{\psi(ax): a \text{ real}\}$ in its sum space and is an element of $V_{n-1}(\psi)$. Such an element is called *degenerate*.

THEOREM 3. *Let $\tau'(0) = 0$. If the sum space of $F(A, \cdot)$ has 0 as its orthogonal complement in $C[\alpha, \beta]$, $F(A, \cdot)$ is a best approximation only to itself.*

Proof. Let $F(A, \cdot)$ be best for $f \in C[\alpha, \beta]$. Let h be in the sum space of $F(A, \cdot)$. Let

$$I(\lambda) = \int \tau(f - F(A, \cdot) - \lambda h),$$

then

$$I'(0) = \frac{\partial}{\partial \lambda} \int \tau(f - F(A, \cdot) - \lambda h)]_{\lambda=0} = \int \frac{\partial}{\partial \lambda} \tau(f - F(A, \cdot) - \lambda h)]_{\lambda=0}$$

$$= \int \tau'(f - F(A \ \cdot))h.$$

Since $F(A, \cdot)$ is best, $I'(0)$ must be zero for all h in the sum space, hence $\tau'(f - F(A, \cdot))$ is in the orthogonal complement of the sum space and is, therefore, zero.

Remark. The proof shows that $F(A, \cdot)$ cannot even be locally best to $f \neq F(A, \cdot)$.

We have for $\psi(x) = \exp(x)$, $\psi(x) = \log(1 + x)$, $\psi(x) = \sin(x)$, $\psi(x) = \cos(x)$, that a degenerate element of $V_n(\psi)$ is best only to itself if $\tau'(0) = 0$.

COROLLARY. *Let $\tau'(0) = 0$. Let ψ be a continuous function and I be an interval such that the orthogonal complement of $\{\psi(\delta x): \delta \in I\}$ in $C[\alpha, \beta]$ is zero. Let $\rho_n(f) = \inf\{N(f - g): g \in V_n(\psi)\}$. If a best approximation exists to f in $V_n(\psi)$ and $f \notin V_n(\psi)$ then $\rho_n(f) > \rho_{n+1}(f)$.*

Proof. If $\rho_n(f) = \rho_{n+1}(f)$ then a best approximation g in V_n is a best approximation in V_{n+1}. The sum space of g in V_{n+1} is $\{\psi(ax): a \text{ real}\}$, hence the orthogonal complement is zero, and by Theorem 3, $g = f$.

A result quite close to the corollary was obtained by Hobby and Rice [5, 98–99], who used a property much more restrictive than the orthogonal complement being zero. If τ does not have a derivative at zero, $F(A, \cdot)$ can be best to f even though the orthogonal complement of the sum space of $F(A, \cdot)$ is zero. For examples, see the papers [1, 3, 9] on the L_1 case. We can, however, show that $Z(A)$ must have positive measure.

Positive Measure

Theorem 4. *Let the orthogonal complement in* $BM[\alpha, \beta]$ *of the sum space of* $F(A, \cdot)$ *be zero.* $F(A, \cdot)$ *can be best to* f *only if* $Z(A)$ *is a set of positive measure.*

Proof. Suppose $Z(A)$ is of zero measure, then by arguments used in the proof of the preceding theorem,

$$\int \tau'(f - F(A, \cdot))h = 0$$

for all h in the sum space. Hence $\tau'(f - F(A, \cdot))$ is in the orthogonal complement of the sum space. As $\tau'(f - F(A, \cdot))$ is continuous on $\{x: f(x) - F(A, x) > 0\}$ and on $\{x: f(x) - F(A, x) < 0\}$ and is bounded, $\tau'(f - F(A, \cdot)) \in BM[\alpha, \beta]$. Hence $\tau'(f - F(A, \cdot)) = 0$ almost everywhere, which contradicts $\mu(Z(A)) = 0$.

Corollary. *Let the orthogonal complement in* $BM[\alpha, \beta]$ *of the sum space of* $F(A, \cdot)$ *be zero. Let* $F(A, \cdot)$ *be analytic, then if* $F(A, \cdot)$ *is best to analytic* f, $f \equiv F(A, \cdot)$.

Exponential Approximation

In this section we consider approximation by $V_n(\exp)$, that is,

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

An approximation which can be expressed in this form with at least one of $a_1, \ldots, a_n$ equal to zero is called *degenerate*. It has $\{\exp(ax): a \text{ real}\}$ in the sum space and this has 0 as its orthogonal complement in $BM[\alpha, \beta]$ by Theorem 2. When $\tau'(0) = 0$, we have by Theorem 3 that degenerate $F(A, \cdot)$ is best only to itself, and the corollary to Theorem 3 is applicable. It is shown by

Meinardus and Schwedt [6, 313] that $\{D(A, B, \cdot): B \in E_{2n}\}$ is a Haar subspace of dimension $2n$ if $F(A, \cdot)$ is nondegenerate. By Theorem 1 and 4 we have the following theorem.

THEOREM 5. *Let $F(A, \cdot)$ be best in $V_n(\exp)$ to f. Let $\tau'(0) = 0$ or f be analytic. Then $f \equiv F(A, \cdot)$ or $f - F(A, \cdot)$ has $2n$ sign changes.*

Results on L_1 approximation are found in [3, 9].

It should be noted that Professor D. W. Kammler of Southern Illinois University, Carbondale, will have papers appearing on mean exponential approximation.

REFERENCES

1. C. B. DUNHAM, Degeneracy in mean rational approximation, *J. Approximation Theory* **4** (1971), 225–229.
2. C. B. DUNHAM, Best mean rational approximation, *Computing* **9** (1972), 87–93.
3. C. B. DUNHAM, Degeneracy in mean nonlinear approximation, *J. Approximation Theory* **9** (1973), 307–312.
4. C. B. DUNHAM, Families satisfying the Haar condition, in preparation.
5. C. R. HOBBY AND J. R. RICE, Approximation from a curve of functions, *Arch. Rat. Mech. Anal.* **24** (1967), 91–106.
6. G. MEINARDUS AND D. SCHWEDT, Nicht-lineare Approximationen, *Arch. Rat. Mech. Anal.* **17** (1964), 297–326.
7. T. S. MOTZKIN AND J. L. WALSH, Best approximations within a linear family on an interval, *Proc. Nat. Acad. Sci.* **46** (1960), 1225–1233.
8. C. B. DUNHAM, Chebyshev approximation with the local Haar condition, *J. SIAM Numer. Anal.* **8** (1971), 749–753.
9. D. BRAESS, On a paper of C. B. Dunham concerning degeneracy in mean nonlinear approximation, *J. Approximation Theory* **9** (1973), 313–315.
10. E. W. CHENEY AND A. A. GOLDSTEIN, Mean square approximation by generalized rational functions, *Math. Z.* **95** (1967), 232–241.
11. J. R. RICE, "The Approximation of Functions," Addison-Wesley, Reading, Mass., 1969, Vol. 2.
12. T. S. MOTZKIN, Approximation in the sense of a deviator integral, *J. SIAM Numer. Anal.* **3** (1966), 276–286.