JOURNAL OF APPROXIMATION THEORY 16, 341-346 (1976)

Differentiability of Nonlinear Best Approximation Operators in a Real Inner Product Space

JERRY M. WOLFE

Department of Mathematics, University of Oregon, Eugene, Oregon 97403

Communicated by E. W. Cheney

Received June 8, 1974

One of the basic problems of approximation theory is the behavior of the best approximation operator. Indeed, the stability theory of computational processes [1] leads one to the problem of determining when the best approximation operator is Lipschitz continuous. In this paper, the implicit function theorem is used to analyze the continuity properties of the best approximation operator in a real inner product space for a certain class of nonlinear approximating families that include ordinary rational functions and the exponential family as examples. Under appropriate hypotheses it will be determined that the best approximation operator is in fact Frechét differentiable, and hence, Lipschitz continuous. The results can be extended to any space with a twice Frechét differentiable norm.

Let H denote a real inner product space, E^N real N-space, S an open subset of E^N, and A a continuous map from S to H. Then given $f \in H$, one seeks to approximate it by elements of $A(S) = \{A(x) \mid x \in S\}$. The best approximation operator $\mathscr{F}$ is the set valued map that assigns to each $f \in H$ the set of closest points to f in $A(S)$. To consider $\mathscr{F}$ as a function we restrict it to the subset of H on which its value is a singleton.

The problem of approximating f is equivalent to finding a minimum for the functional $F(f, x) = [A(x) - f, A(x) - f]$ as x ranges over S where $[\cdot, \cdot]$ is the inner product on H. If one assumes that the map A has two Frechét derivatives at each point of S, then necessary conditions for an element $x \in S$ to minimize F are given by:

(1) $F'(f, x, h) = 0$, for all $h \in E^N$,

(2) $F''(f, x, h, h) \geqslant 0$, for all $h \in E^N$.

Here $F'(f, x, \cdot)$ and $F''(f, x, \cdot, \cdot)$ are, respectively, the first and second Frechét derivatives of the map F with respect to x. Using the chain rule and partial differentiation [2, p. 685] we have

$$\tfrac{1}{2}F'(f, x, h) = [A(x) - f, A'(x, h)],$$
$$\tfrac{1}{2}F''(f, x, h, h) = [A'(x, h), A'(x, h)] + [A(x) - f, A''(x, h, h)],$$

where

$$A'(x, h) = \sum_{j=1}^{N} (\partial A/\partial x_j)(x)\, h_j\,,$$

$$A''(x, h, h) = \sum_{i=1}^{N} \sum_{j=1}^{N} (\partial^2 A/\partial x_i\, \partial x_j)(x)\, h_i h_j\,,$$

$$h = (h_1, \ldots, h_N)^T.$$

Using the above formulas, conditions (1) and (2) can be recast in the form

(1)′ $\psi(f, x) = 0$

(2)′ $\langle h, D\psi_x(f, x)(h)\rangle \geqslant 0$, for all $h \in E^N$,

where

$$\psi(f, x) \equiv ([A(x) - f, (\partial A/\partial x_1)(x)], \ldots, [A(x) - f, (\partial A/\partial x_N)(x)])^T,$$

$D\psi_x(f, x)(\cdot)$ is the derivative of $\psi(f, x)$ with respect to x, and $\langle\cdot, \cdot\rangle$ is the usual inner product on E^N. The properties of $\mathscr{F}$ will be determined by examining the solutions of (1)′.

EXAMPLE 1. Let $H = L_2[0, 1]$ and $R_m^n[0, 1] \equiv \{a_0 + \cdots + a_n x^n / b_0 + \cdots + b_m x^m \mid b_0 + \cdots + b_m x^m > 0$, for all $x \in [0, 1]\}$ and $S = \{(a_0, \ldots, a_n, b_1, \ldots, b_m) \mid 1 + b_1 x + \cdots + b_m x^m > 0$ for all $x \in [0, 1]\}$. Define $A: S \to H$ by $A(a_0, \ldots, a_n, b_1, \ldots, b_m) = (a_0 + \cdots + a_n x^n)/(1 + b_1 x + \cdots + b_m x^m)$. Since every $r \in R_m^n[0, 1]$ has a representation with $b_0 = 1$ we have that $A(S) = R_m^n[0, 1]$. For the calculation of the necessary derivatives, see [4].

EXAMPLE 2. Let $H = L_2[0, 1]$ and $S = E^{2N}$ and define A by $A(a_1, \ldots, a_n, t_1, \ldots, t_N) = \sum_{j=1}^{N} a_j e^{t_j x}$.

THEOREM 1. *Let $x_0 \in S$ and $f_0 \in H$ be such that*

(i) $\psi(f_0, x_0) = 0$

(ii) $\inf_{\|h\|=1} \langle h, D\psi_x(f_0, x_0)(h)\rangle > 0$, *and assume that the map $x \to A''(x, \cdot, \cdot)$ is continuous on S. Then there exists a neighborhood U of f_0 and a neighborhood V of x_0 with $V \subset S$ and a function $x(\cdot)\colon U \to E^N$ such that*

(a) $\psi(f, x(f)) = 0$, *for all $f \in U$*

(b) $x(f) \in V$, *for all $f \in U$*

(c) $\psi(f, x) = 0$, *with $f \in U$ and $x \in V$ implies that $x = x(f)$*

(d) *$x(\cdot)$ is differentiable on U with $x'(f, g) = -D\psi_x^{-1}(f, x(f)) \times (D\psi_f(f, x(f))(g))$, for all $g \in H$, where $D\psi_f(f, x)(\cdot)$ is the partial derivative of $\psi(f, x)$ with respect to f.*

Proof. By direct calculation, the matrix representing $D\psi_x(f_0, x_0)(\cdot)$ is given by $(a_{ij}) \equiv ([A(x_0) - f, (\partial^2 A/\partial x_i \partial x_j)(x_0)] + [(\partial A/\partial x_i(x_0), (\partial A/\partial x_j)(x_0)])$, while $D\psi_f(f_0, x_0)(g) = ([-g, (\partial A/\partial x_1)(x_0)],..., [-g, (\partial A/\partial x_N)(x_0)])^T$. By (ii), $D\psi_x(f_0, x_0)$ is positive definite so that $D\psi_x^{-1}(f_0, x_0)$ exists. Moreover, from the above formulas, the maps $(f, x) \to D\psi_x(f, x)(\cdot)$ and $(f, x) \to D\psi_f(f, x)(\cdot)$ are easily seen to be continuous on $H \times S \to B(E^N, E^N)$ and $H \times S \to B(H, E^N)$, respectively, where $B(F, G)$ denotes the set of bounded linear operators from the normed linear space F to the normed linear space G. Thus, using the (general) implicit function theorem [3, p. 230] we have that (a)–(d) hold. ▮

Remark 1. It is not difficult to show that if the map A has k continuous derivatives, then the map $x(\cdot)$ has at least $k - 1$ continuous derivatives on U. In particular, it is continuously differentiable on U.

COROLLARY 1. *Let f_0, x_0, U, and V be as in Theorem* 1. *Then the map $x(\cdot)$ is Lipschitz continuous at f_0.*

Proof. $D\psi_x(f_0, x_0)$ is symmetric and positive definite so that $\| D\psi_x^{-1}(f_0, x_0)\| = 1/\lambda^*(f_0, x_0)$, where $\lambda^*(f_0, x_0)$ is the smallest eigenvalue of $D\psi_x(f_0, x_0)$. (Here, of course, we are using the spectral norm.) As noted in the proof of Theorem 1, the elements (and hence, the eigenvalues) of $D\psi_x(f, x)$ are jointly continuous functions of f and x. Hence, there is a neighborhood $U_0 \times V_0 \subset U \times V$ of (f_0, x_0) such that for all $(f, x) \in U_0 \times V_0$, we have that $\| D\psi_x^{-1}(f, x)\| \leqslant \delta < \infty$. Also, it is clear that we can assume that $\| D\psi_f(f, x)(\cdot)\|$ is bounded on $U_0 \times V_0$. Thus, for some constant $K > 0$ depending on f_0 we have that

$$\| x'(f, \cdot)\| \leqslant \| D\psi_x^{-1}(f, x(f))(\cdot)\| \cdot \| D\psi_f(f, x(f))(\cdot)\| \leqslant K < \infty \quad \text{for all} \quad f \in U_0.$$

Hence, by the generalized mean value theorem [3, p. 149] we have that for any $f \in U_0$, $\| x(f) - x(f_0)\| \leqslant K \| f - f_0 \|$. ▮

COROLLARY 2. *Let f_0, x_0, U, and V be as in Theorem* 1 *and let ρ: $U \to A(S)$ be defined by $\rho(f) = A(x(f))$. Then ρ is differentiable on U.*

Proof. Chain rule. ▮

We shall need the following definitions.

DEFINITION 1. An element $A(x) \in A(S)$ is called a normal point if

(i) $A'(x, \cdot)$ is one to one.

(ii) $A^{-1}(\cdot)$ exists and is continuous on a relatively open neighborhood of $A(x)$.

DEFINITION 2. A subset M of a normed linear space X is called approximatively compact if for each $x \in X$ and each sequence $\{m_v\} \subset M$ with the property that $\| x - m_v \| \to \inf_{x \in M} \| x - m \|$ there exists of subsequence of $\{m_v\}$ converging in the norm topology to some element of M.

Remark 2. It is simple to show that if M is approximatively compact in S, then each $x \in X$ has at least one closest point in M.

It is evident that to establish smoothness of the operator T at a point $f \in H$ at which it is single valued, it is necessary to know that T is single valued on some neighborhood of f. We will now determine conditions under which this is true. The proof of the following lemma may be found in [6, p. 388].

LEMMA 1. *Let M be an approximatively compact subset of a normed linear space X and suppose $x \in X$ is such that x has a unique closest point m^* in M. Then, if $\{x_v\}$ is any sequence converging to x and if $\{m_v\}$ is any set of corresponding closest points in M, $m_v \to m^*$.*

THEOREM 2. *Assume that the closure of $A(S)$ in H, denoted by* $\mathrm{cl}(A(S))$ *is approximatively compact and that $A''(x, \cdot, \cdot)$ is continuous on S. Suppose $f_0 \in H$ is such that f_0 has a unique closest point in* $\mathrm{cl}(A(S))$, *which lies in $A(S)$, say $A(x_0)$, which is a normal point. Assume also that*

$$\inf_{\|h\|=1} \langle h, D\psi_x(f_0, x_0)(h) \rangle = \eta > 0.$$

Then there is a neighborhood U of f_0 such that each $g \in U$ has a unique best approximation in $A(S)$.

Proof. The proof is a modification of [4, Theorem 2]. Suppose the result is false. Then there is a sequence $\{g_v\} \subset H$, $g_v \to f_0$ such that g_v does not have a unique best approximation in $A(S)$. Assume first that some subsequence of $\{g_v\}$, which we do not relabel, fails to have a best approximation in $A(S)$. Let $\{m_v\}$ be any corresponding sequence of best approximations from $\mathrm{cl}(A(S))$. By Lemma 1, $m_v \to A(x_0)$, and by Theorem 1, there is a neighborhood U_0 of f_0 and a neighborhood V_0 of x_0 with $V_0 \subset S$ and a map $x(\cdot)$: $U_0 \to E^N$ such that (a)–(d) of Theorem 1 hold. Moreover, from the joint continuity of the map $(f, x) \to \langle \cdot, D\psi_x(f, x)(\cdot) \rangle$, there exists a neighborhood $U_1 \times V_1$ of (f_0, x_0) such that for all $(f, x) \in U_1 \times V_1$, $\inf_{\|h\|=1} \langle h, D\psi_x(f, x)(h) \rangle \geqslant \eta/2 > 0$. According to the hypothesis $\| g_v - m_v \| < \| g_v - A(x_v) \|$ (where x_v denotes $x(g_v)$ for all v, and since $m_v \in \mathrm{cl}(A(S))$, there is a $y_v \in S$ such that $\| A(y_v) - m_v \| < 1/v$ and $\| g_v - A(y_v) \| < \| g_v - A(x_v) \|$. Clearly, $A(y_v) \to A(x_0)$, and by normality, $y_v \to x_0$. Hence, for sufficiently large v, $(g_v, y_v) \in (U_0 \cap U_1) \times (V_0 \cap V_1) \equiv U \times V$. But then using Taylor's theorem,

$$\begin{aligned} F(g_v, y_v) &= F(g_v, x_v) + F'(g_v, x_v, y_v - x_v) \\ &\quad + \tfrac{1}{2} F''(g_v, \mathscr{E}_v, y_v - x_v, y_v - x_v), \end{aligned}$$

where $\mathscr{E}_v = \theta_v x_v + (1 - \theta_v) y_v$ for some $\theta_v \in [0, 1]$. Since

$$F'(g_v, x_v, y_v - x_v) = 0$$

by (a) of Theorem 1 and $\mathscr{E}_v \in V$ for sufficiently large v we conclude that $F(g_v, y_v) > F(g_v, x_v)$, a contradiction. Thus, a neighborhood U exists for which each $g \in U$ has a best approximation in $A(S)$.

To establish uniqueness, again assume the conclusion of the theorem is false and find a sequence $\{g_v\} \subset U$ converging to f such that each g_v has at least two distinct best approximations say $A(y_v)$ and $A(y_v')$ in $A(S)$. By Lemma 1 and normality of $A(x_0)$, $y_v \to x_0$ and $y_v' \to x_0$. Then $0 = \psi(g_v, y_v) = \psi(g_v, y_v')$ and so by Theorem 1 for large enough v, $y_v = y_v' = x(g_v)$, a contradiction. ∎

THEOREM 3. *Let A and $f_0 \in H$ be as in Theorem 2, where again $A(x_0)$ denotes the unique best approximation to f_0 from* cl($A(S)$). *Then there is a neighborhood U of f_0 on which the best approximation operator $\mathscr{F}$ is continuously differentiable.*

Proof. By Theorem 2 there is a neighborhood U of f_0 on which $\mathscr{F}$ is uniquely defined and it is clear from the proof that U can be chosen so that for each $g \in U$, $\mathscr{F}g = A(x(g))$. The theorem then follows immediately from Theorem 1, Corollary 2, and Remark 1. ∎

For the special case when $A(S) = R_m^n[0, 1]$, we have the following stronger result.

COROLLARY 3. *Let $A(S) = R_m^n[0, 1]$ and $H = L_2[0, 1]$, where A and S are as in Example* 1. *Then there is an open and dense subset of $L_2[0, 1]$ on which the best approximation operator T is continuously differentiable.*

Proof. The set $R_m^n[0, 1]$ is approximatively compact [5] and by [4, Theorem 4] the set of elements f satisfying the hypotheses of Theorem 2 contains an open and dense subset of H. Applying Theorem 3 to each such f we obtain the required open and dense subset of H. ∎

Corollary 3 indicates a fairly strong uniqueness result for rational approximation. However, we have the following result regarding local best approximations.

COROLLARY 4. *Suppose $f \in H$ and $x_1, ..., x_k$ are such that $D\psi_x(f, x_i)$ is positive definite and $\psi(f, x_i) = 0$ for $i = 1, ..., k$, Then there is an open ball B around f in H such that for each $g \in B$ the function $[A(x) - g, A(x) - g]$ has at least k isolated local minima in S.*

Proof. Applying Theorem 1 k times, we find neighborhoods U_j of f_0 and corresponding neighborhoods V_j of x_j $j = 1,..., k$ and maps $x_j(\cdot)$ $j = 1,..., k$ sending U_j into V_j. By Lipschitz continuity of $x_j(\cdot)$, we can restrict each U_j so that $V_j \cap V_i = \varnothing$ unless $i = j$. Then, letting $U = \bigcap_{i=1}^{k} U_i$, we have that for each $g \in U$ the functional $[A(x) - g, A(x) - g]$ has an isolated local minimum in V_j $j = 1,..., k$. ∎

Remark 3. The existence of functions f satisfying the hypotheses of Corollary 4 for arbitrary k is shown in [4] for the case when $A(S)$ is the set $R_m{}^n[0, 1]$. Moreover, the minima may be arbitrarily close together so that the practical problem of calculating a global minimum of $\| A(\cdot) - f \|^2$ can be quite difficult. That is, a unique global minimum may have several local minima located nearby.

References

1. E. Isaacson and H. B. Keller, "Analysis of Numerical Methods," Wiley, New York, 1966.
2. L. W. Kantorovich and G. P. Arilov, "Functional Analysis in Normed Linear Spaces," MacMillan, New York, 1964.
3. L. H. Loomis and S. Sternberg, "Advanced Calculus," Addison–Wesley, Reading, Massachusetts, 1968.
4. J. M. Wolfe, On the unicity of nonlinear approximation in smooth spaces, *J. Approximation Theory*, **12** (1974), 165–181.
5. J. M. Wolfe, L_p rational approximation, *J. Approximation Theory*, **12** (1974), 1–5.
6. I. Singer, "Best Approximation in Normed Linear Spaces by Elements of Linear Subspaces," Springer–Verlag, New York, 1970.