JOURNAL OF APPROXIMATION THEORY **17**, 166–176 (1976)

# Nonlinear $L_1$ Approximation of Smooth Functions*

JERRY M. WOLFE

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

*Communicated by E. W. Cheney*

Received September 30, 1974

## INTRODUCTION

In this paper we shall consider certain aspects of nonlinear approximation of smooth functions in the $L_1$ norm on an interval $[a, b]$ of the real line. In particular it will be shown that under certain circumstances (which would include many practical situations) second derivative techniques are applicable. This leads to checkable sufficient conditions for a local best approximation and makes applicable certain numerical techniques such as Newton's method. We also consider the unicity problem and extend certain results in [1, 2] given for $L_p$ spaces with $p \geqslant 2$ to the setting of this paper.

The approximation problem we consider is as follows. An open set $S$ of Euclidean $N$-space $E^N$ is given and a map $A(\cdot): S \to C[a, b]$ such that the map $x \to A''(x, \cdot, \cdot)$ is continuous on $S$ where $A''(x, \cdot, \cdot)$ is the second Fréchet derivative of the map $A$ at the point $x$. Moreover we will assume that

(1) For each $x \in S$, $\{(\partial A/\partial x_1)(x),..., (\partial A/\partial x_N)(x)\}$ spans a Haar subspace of $C[a, b]$ of dimension $d(x) \leqslant N$.

(2) For each $x, y \in S$, $A(x) - A(y)$ is either identically zero or has at most $N - 1$ roots in $[a, b]$.

(3) Each $A(x)(t)$ is analytic on $[a, b]$ with respect to $t$ and moreover, if $x_v \to x_0$ then $(d/dt)\, A(x_v)(t)$ converges uniformly to $(d/dt)\, A(x_0)(t)$ on $[a, b]$.

Then given $f \in C[a, b]$ we seek $A(x^*) \in A(S) \equiv \{A(x) \mid x \in S\}$ such that $\int_a^b | A(x^*)(t) - f(t)|\, dt = \inf_{x \in S} \int_a^b | A(x)(t) - f(t)|\, dt$.

The standard nonlinear families encountered in practice such as the rational functions and the exponentials satisfy the hypotheses (1)–(3) so that the class of approximating families considered here is a useful one.

* Research for this paper was supported in part by the Graduate School, Department of Mathematics, University of Oregon, Eugene, Oregon.

Our analysis requires that we establish the Fréchet differentiability of certain maps. However, it is more convenient to calculate directional or Gateaux derivatives and so we will use the following basic fact from multivariate calculus.

LEMMA 1. *Suppose $\varphi$ is a real valued map defined on an open subset of $E^N$ such that $\varphi$ is Gateaux differentiable on a neighborhood of $x_0$ and such that for each $h$, the map $x \to \varphi'(x, h)$ is continuous at $x_0$. Then $\varphi$ is Fréchet differentiable at $x_0$.*

We shall not distinguish notationally between the two types since in each case the above lemma will be applicable. The following lemma is also standard and we shall not present a proof here. The set $Z(x)$ (defined below) is crucial to the result and we refer the reader to the work of Kripke and Rivlin [5] for a thorough analysis of its role in $L_1$ problems.

LEMMA 2. *Assume that for each $x \in S$ the set $Z(x) \equiv \{t \in [a, b] \mid A(x)(t) = f(t)\}$ has Lebesgue measure zero. Then the function $F(x) = \int_a^b |A(x)(t) - f(t)|\, dt$ is continuously Fréchet differentiable on $S$ and*

(*) $\quad F'(x, h) = \int_a^b \operatorname{sgn}(A(x)(t) - f(t))\, A'(x, h)(t)\, dt$ *for all $h \in E^N$ where $A'(x, \cdot)$ is the Fréchet derivative of $A$ at $x$.*

The following necessary condition for a best approximation is now clear.

COROLLARY 1. *Suppose $A(x_0)$ is a local best approximation to $f$ and assume $Z(x)$ has measure zero for all $x \in S$. Then $\int_a^b \operatorname{sgn}(A(x_0)(t) - f(t)) \times A'(x_0, h)(t)\, dt = 0$ for all $h \in E^N$.*

## DERIVATIVES

The following is the basic result of this paper.

THEOREM 1. *Let $f \in C[a, b]$ be differentiable on $(a, b)$ and let $A(x) \in A(S)$. Assume that $\epsilon(t) \equiv A(x)(t) - f(t)$ has precisely $K$ roots in $[a, b]$ say $t_1 < \cdots < t_K$ all in the interior of $[a, b]$ such that $(d\epsilon/dt)(t_j) \neq 0$ $j = 1,\ldots, K$. The functional $F(x) = \int_a^b |A(x)(t) - f(t)|\, dt$ is twice continuously Frechet differentiable at $x$ and in fact*

$$F''(x, h, k) = \sum_{j=1}^{K} \frac{A'(x, h)(t_j)\, A'(x, k)(t_j)}{|(d\epsilon/dt)(t_j)|} + \int_a^b \operatorname{sgn}(\epsilon(t))\, A''(x, h, k)(t)\, dt \qquad (1)$$

*for each $h, k \in E^N$ where the first sum is empty if $K = 0$.*

*Proof.* Let $h, k \in E^N$ be arbitrary and define $\epsilon(\lambda, t) \equiv A(x + \lambda k)(t) - f(t)$ for $t \in [a, b]$ and $|\lambda|$ sufficiently small that $x + \lambda k \in S$. Then

$$\begin{aligned}
&[F'(x + \lambda k, h) - F'(x, h)]/\lambda \\
&\quad = \int_a^b [\operatorname{sgn}(\epsilon(\lambda, t))\, A'(x + \lambda k, h)(t) - \operatorname{sgn}(\epsilon(0, t))\, A'(x, h)(t)\, dt]/\lambda \\
&\quad = \int_a^b [(\operatorname{sgn}(\epsilon(\lambda, t)) - \operatorname{sgn}(\epsilon(0, t)))/\lambda]\, A'(x, h)(t)\, dt \\
&\qquad + \int_a^b \operatorname{sgn} \epsilon(\lambda, t)[(A'(x + \lambda k, h)(t) - A'(x, h)(t))/\lambda] \equiv I_1(\lambda) + I_2(\lambda).
\end{aligned}$$

It is clear by the Lebesgue dominated convergence theorem that $I_2(\lambda) \to \int_a^b \operatorname{sgn}(A(x)(t) - f(t))\, A''(x, h, k)(t)\, dt$ as $\lambda \to 0$. To calculate $\lim_{\lambda \to 0} I_1(\lambda)$ we first note that $(\partial\epsilon/\partial\lambda)(\lambda, t) = A'(x + \lambda k, k)(t)$ and $(\partial\epsilon/\partial t)(\lambda, t)$ is nonzero at $(0, t_j)$ for each $j = 1,..., K$ and both partial derivatives are continuous. Letting $t_j(0) \equiv t_j$ $j = 1,..., K$ and applying the implicit function theorem in a neighborhood of each $t_j(0)$ separately we find that there is a $\lambda_0 > 0$ such that for each $\lambda$ and $j$ with $|\lambda| \leqslant \lambda_0$ there is a unique $t_j(\lambda)$ such that $\epsilon(\lambda, t_j(\lambda)) = 0$. Moreover,

$$\frac{dt}{d\lambda} j(\lambda) = -\frac{A'(x + \lambda k, k)(t_j(\lambda))}{(\partial\epsilon/\partial t)(\lambda, t_j(\lambda))}\, j = 1,..., K.$$

By perhaps restricting $\lambda$ further we may assume that $\epsilon(\lambda, t)$ has no other roots in $[0, 1]$ and that $|t_j(\lambda) - t_j(0)| < \min\{|t_{j+1}(0) - t_j(0)|, |t_{j-1}(0) - t_j(0)|\}$. Thus, letting $\alpha_j(\lambda) = \min(t_j(\lambda), t_j(0))$ and $\beta_j(\lambda) = \max(t_j(\lambda), t_j(0))$ we may write $I_1(\lambda)$ in the form

$$I_1(\lambda) = \sum_{j=1}^{K} \int_{\alpha_j(\lambda)}^{\beta_j(\lambda)} \frac{[\operatorname{sgn}(\epsilon(\lambda, t)) - \operatorname{sgn}(\epsilon(0, t))]}{\lambda} A'(x, h)(t)\, dt$$

since $\operatorname{sgn}(\epsilon(\lambda, t)) = \operatorname{sgn}(\epsilon(0, t))$ except on the intervals $(\alpha_j(\lambda), \beta_j(\lambda))$. Since $\int_{\min(\alpha,\beta)}^{\max(\alpha,\beta)} \psi(t)\, dt = \operatorname{sgn}(\beta - \alpha) \int_\alpha^\beta \psi(t)\, dt$, $I_1(\lambda)$ becomes

$$\sum_{j=1}^{K} (\operatorname{sgn} t_j(\lambda) - t_j(0)) \int_{t_j(\lambda)}^{t_j(\lambda)} \left[\frac{\operatorname{sgn}(\epsilon(\lambda, t)) - \operatorname{sgn}(\epsilon(0, t))}{\lambda}\right] A'(x, h)(t)\, dt. \qquad (*)$$

Suppose that $\epsilon(0, a) > 0$. Then to evaluate

$$\int_{t_1(0)}^{t_1(\lambda)} \left[\frac{\operatorname{sgn}(\epsilon(\lambda, t)) - \operatorname{sgn}(\epsilon(0, t))}{\lambda}\right] A'(x, h)(t)\, dt \qquad (**)$$

we consider two cases:

(a) $t_1(0) < t_1(\lambda)$. Then (**) becomes

$$2\int_{t_1(0)}^{t_1(\lambda)} \frac{A'(x, h)(t)}{\lambda}\, dt$$

$$= (-1)^{j-1}\, 2\, \mathrm{sgn}(\epsilon(0, a))\, \mathrm{sgn}(t_1(\lambda) - t_1(0)) \int_{t_1(0)}^{t_1(\lambda)} \frac{A'(x, h)(t)}{\lambda}\, dt$$

where $j = 1$.

(b) $t_1(0) \geqslant t_1(\lambda)$. Then (**) becomes

$$-2\int_{t_1(0)}^{t_1(\lambda)} \frac{A'(x, h)(t)}{\lambda}\, dt$$

$$= (-1)^{j-1} \cdot 2 \cdot (\mathrm{sgn}\, \epsilon(0, a))\, \mathrm{sgn}(t_1(\lambda) - t_1(0)) \int_{t_1(0)}^{t_1(\lambda)} \frac{A'(x, h)(t)}{\lambda}\, dt.$$

Clearly, this formula (in (a) or (b)) is still valid if $\epsilon(0, a) < 0$. Because of the alternation of $\mathrm{sgn}(\epsilon(\lambda, t))$ at $t_j(\lambda)$ it is easy to see that the general relationship is

$$\int_{t_j(0)}^{t_j(\lambda)} \left[\frac{\mathrm{sgn}(\epsilon(\lambda, t)) - \mathrm{sgn}(\epsilon(0, t))}{\lambda}\right] A'(x, h)(t)\, dt$$

$$= (-1)^{j-1}\, 2\, \mathrm{sgn}(\epsilon(0, a))\, \mathrm{sgn}(t_j(\lambda) - t_j(0)) \int_{t_j(0)}^{t_j(\lambda)} \frac{A'(x, h)(t)}{\lambda}\, dt$$

But

$$\frac{1}{\lambda}\int_{t_j(0)}^{t_j(\lambda)} A'(x, h)(t)\, dt = \frac{1}{\lambda}\left[\int_{t_j(0)}^{t_j(\lambda)} A'(x, h)(t)\, dt - \int_{t_j(0)}^{t_j(0)} A'(x, h)(t)\, dt\right]$$

and passing to the limit we obtain

$$\begin{aligned}
\lim_{\lambda\to 0} \frac{1}{\lambda}\int_{t_j(0)}^{t_1(\lambda)} & A'(x, h)(t)\, dt \\
&= \frac{d}{d\lambda}\int_{t_j(0)}^{t_j(\lambda)} A'(x, h)(t)\, dt\,\Big|_{\lambda=0} \\
&= A'(x, h)(t_j(\lambda))\, t_j'(\lambda)\,|_{\lambda=0} \\
&= \frac{-A'(x + \lambda k, k)(t_j(\lambda))}{(\partial\epsilon/\partial t)(\lambda, t_j(\lambda))} \cdot A'(x, h)(t_j(\lambda))\,\Big|_{\lambda=0} \\
&= \frac{(-1)\, A'(x, k)(t_j(0))\, A'(x, h)(t_j(0))}{(\partial\epsilon/\partial t)(0, t_j(0))}.
\end{aligned}$$

Finally it is simple to check that $\operatorname{sgn}(\partial\epsilon/\partial t)(0, t_j(0)) = (-1)^j \operatorname{sgn} \epsilon(0, a)$ $j = 1,..., K$. Thus we obtain

$$\lim_{\lambda\to 0} \left[(-1)^{j-1} 2 \operatorname{sgn}(\epsilon(0, a)) \operatorname{sgn}(t_j(\lambda) - t_j(0)) \int_{t_j(0)}^{t_j(\lambda)} \frac{A'(x, h)(t)}{\lambda}\, dt\right]$$

$$= (-1)^{j-1} 2 \operatorname{sgn}(\epsilon(0, a)) \operatorname{sgn}(t_j(\lambda) - t_j(0))$$

$$\times \frac{(-1)\, A'(x, k)(t_j(0))\, A'(x, h)(t_j(0))}{(-1)^j \operatorname{sgn}(\epsilon(0, a))\, |(\partial\epsilon/\partial t)(0, t_j(0))|}$$

$$= \frac{\operatorname{sgn}(t_j(\lambda) - t_j(0))\, A'(x, k)(t_j(0))\, A'(x, h)(t_j(0))}{|(\partial\epsilon/\partial t)(0, t_j(0))|}$$

and substituting into (*) we have the desired value for $\lim_{\lambda\to 0} I_1(\lambda)$. Thus formula (1) is valid for directional derivatives. The fact that $F(x)$ is actually twice continuously Frechet differentiable follows easily from the fact that $x \to A''(x, \cdot, \cdot)$ is continuous on $S$ and that (using assumption (3)) as

$$x_v \to x, \qquad \frac{dA}{dt}(x_v)(t) \to \frac{dA}{dt}(x)(t)$$

uniformly in $t$. ∎

*Remark* 1. It does not follow automatically from our previous assumptions that a local minimum $x_0$ of the functional $F(x) \equiv \int_a^b | A(x)(t) - f(t)|\, dt$ corresponds to a local best approximation to $f$ from $A(S)$. However, if $A^{-1}(\cdot)$ exists on a relative neighborhood of $A(x_0)$ in $A(S)$ and is continuous at $A(x_0)$ then $A(x_0)$ is a local best approximation. This leads to the following definition.

DEFINITION. A point $x_0 \in S$ is called normal if

(i) $A'(x_0, \cdot)$ is one to one on $E^N$

(ii) $A^{-1}(x)$ exists on a relative neighborhood of $A(x_0)$ in $A(S)$ and is continuous at $A(x_0)$.

If one assumes $f$ is analytic, say, on $[a, b]$ (and not in $A(S)$) then any local best approximation must be normal for any of the standard nonlinear families. (See for example [4] and [1, Theorem 8] noting that the proof can be adapted to this setting.)

As an application of Theorem 1 we have the following which provides sufficient conditions for a local minimum of the functional $F(x)$.

COROLLARY 2. *Let* $f \in C[a, b]$ *be differentiable on* $(a, b)$ *and assume* $A(x_0) \in A(S)$ *is normal and satisfies*:

(i) *$\epsilon(t) = A(x_0)(t) - f(t)$ has precisely $K$ roots in $[a, b]$ all of which are simple and lie in $(a, b)$*

(ii) $\int_a^b \operatorname{sgn}(\epsilon(t))\, A'(x_0, h)(t)\, dt = 0$ *for all* $h \in E^N$

(iii) $\left\{ \sum_{j=1}^{K} \frac{(A'(x_0, h)(t_j))^2}{|(d\epsilon/dt)(t_j)|} + \int_a^b \operatorname{sgn}(\epsilon(t))\, A''(x_0, h, h)(t)\, dt \right\} > 0$ *if* $h \neq 0$.

*Then $A(x_0)$ is an isolated local best approximation to $f$ from $A(S)$ with respect to the $L_1$ norm.*

*Proof.* Hypotheses (i)–(iii) above and Theorem 2 imply that functional $F(x) = \int_a^b | A(x)(t) - f(t)|\, dt$ satisfies

(a) $F'(x_0, h) = 0$ for all $h \in E^N$

(b) $F''(x_0, h, h) \geqslant \eta > 0$ for all $h \in E^N$ such that $\| h \| = 1$.

Thus $x_0$ is a local minimum of $F(\cdot)$. The continuity of $A^{-1}$ at $x_0$ then implies that $A(x_0)$ is also a local minimum of $\int_a^b | r(t) - f(t)|\, dt$ as $r$ ranges over $A(S)$. ∎

*Remark* 2. If the hypotheses of Corollary 1 above are satisfied then the number of roots of $\epsilon(t)$ is at least $N$ (see [4]) so that the sum

$$\sum_{j=1}^{K} \frac{(A'(x_0, h)(t_j))^2}{|(d\epsilon/dt)(t_j)|}$$

is strictly positive for nonzero $h$. (That is the space spanned by $(\partial A/\partial x_1)(x_0),\ldots, (\partial A/\partial x_N)(x_0)$ is Haar of dimension $N$ (see Remark 1) so that for nonzero $h$, $A'(x_0, h)(\cdot)$ has at most $N - 1$ roots in $[a, b]$).

*Remark* 3. The situation when $\epsilon(t)$ has multiple roots in $[a, b]$ (but only a finite number of them) is more complex. One finds in this case (provided that $f$ is analytic on $[a, b]$ say) that

$$F''_{+}(x, h, h) = \lim_{\lambda \to 0^+} \frac{F'(x + \lambda h, h) - F'(x, h)}{\lambda}$$

and

$$F''_{-}(x, h, h) = \lim_{\lambda \to 0^-} \frac{F'(x + \lambda h, h) - F'(x, h)}{\lambda}$$

both exist for every $h$ and are of the form

$$\sum_{j=1}^{K} \frac{(A'(x, h)(t_j))^2}{|(d\epsilon/dt)(t_j)|} + \text{nonnegative terms} + \int_a^b \operatorname{sgn}(\epsilon(t))\, A''(x, h, h)(t)\, dt$$

where some of the nonnegative terms may be $+\infty$ and where the finite sum runs over the simple roots of $\epsilon(t)$ in $(a, b)$.

We shall not present the full details of the above analysis but instead record the following simple example to illustrate the basic ideas.

EXAMPLE 1. Let $S = E'$ and define $A: E' \to C[-1, 1]$ by $A(x) = x$ (i.e., the constant function $x$) and let $f(t) = t^3$. Then

$$F(x) = \int_{-1}^{1} | x - t^3 | \, dt \qquad \text{and} \qquad F'(x, h) = \int_{-1}^{1} \operatorname{sgn}(x - t^3) \, h \, dt.$$

Letting $x = 0$ we note that $F'(0, h) = 0$ for all $h \in E'$ and that it is the only parameter with this property. To calculate $F''(x, h, h)$ where $h \neq 0$ we have

$$\begin{aligned} \frac{F'(0 + \lambda h, h) - F'(0, h)}{\lambda} &= \frac{F'(\lambda h, h)}{\lambda} = \frac{1}{\lambda} \int_{-1}^{1} \operatorname{sgn}(\lambda h - t^3) \, h \, dt \\ &= \frac{1}{\lambda} \int_{-1}^{(\lambda h)^{1/3}} h \, dt + \frac{1}{\lambda} \int_{(\lambda h)^{1/3}}^{1} - h \, dt \\ &= \frac{h}{\lambda} [\lambda^{1/3} h^{1/3} + 1] - \frac{h}{\lambda} [1 - \lambda^{1/3} h^{1/3}] \\ &= \frac{2h^{4/3}}{\lambda^{2/3}} \to +\infty \end{aligned}$$

as $\lambda \to 0$. Thus, $F''_{+}(0, h, h) = F''_{-}(0, h, h) = +\infty$.

We now have the following result which we only state.

COROLLARY 3. *Suppose $f$ is analytic on $[a, b]$ and suppose $A(x_0) \in A(S)$ is normal and satisfies*

(i) $\displaystyle\int_a^b \operatorname{sgn}(\epsilon(t)) \, A'(x_0, h)(t) \, dt = 0$ *for all* $h \in E^N$

(ii) $\displaystyle\inf_{\|h\|=1} \left\{ \sum_{j=1}^{K} \frac{(A'(x_0, h)(t_j))^2}{|(d\epsilon/dt)(t_j)|} + \int_a^b \operatorname{sgn}(\epsilon(t)) \, A''(x, h, h)(t) \, dt \right\} > 0$

*where the finite sum is over the simple roots of $\epsilon(t)$ in $(a, b)$. Then $A(x_0)$ is an isolated local best approximation to $f$ from $A(S)$.*

*Remark* 4. The results above should have computational significance since they will allow the application of Newton's method to the solution of many types of nonlinear $L_1$ problems. Of course the problem of locating the roots of the error curve is nontrivial but it is likely that fairly crude approximations will suffice. These computational aspects of the $L_1$ problem will be considered in a future paper.

## UNIQUENESS

We now consider the unicity problem. It is well known that, in general, not every function (even analytic) will have a unique best approximation from $A(S)$ (when its best approximation exists). In [1] the topological properties of the set of elements having unique best approximations in $A(S)$ was studied in $L_p$ norms for $p \geqslant 2$ using second derivative techniques. Using Theorem 1 some of those results will now be extended to the $L_1$ setting.

THEOREM 2. *Let $\mathcal{O}\!\ell$ denote the set of functions analytic on $[a, b]$ and let $A(S)$ be as before. Then the set of functions having unique best approximations in $A(S)$ is a dense subset of the set of functions having at least one best approximation in $A(S)$.*

*Proof.* Let $f \in \mathcal{O}\!\ell$ and assume $A(x_0) \in A(S)$ is a best approximation to $f$. Let $f_\lambda = \lambda f + (1 - \lambda) A(x_0)$ for each $\lambda \in (0, 1)$.

CLAIM. *Each $f_\lambda$ has $A(x_0)$ as its unique best approximation from $A(S)$.*

*Proof.* For some $\lambda \in (0, 1)$ suppose $A(x_1) \neq A(x_0)$ is a best approximation to $f_\lambda$. Then

$$\begin{aligned} \|f - A(x_1)\| &\leqslant \|f - f_\lambda\| + \|f_\lambda - A(x_1)\| \\ &\leqslant \|f - f_\lambda\| + \|f_\lambda - A(x_0)\| \\ &= \|f - A(x_0)\| \leqslant \|f - A(x_1)\| \end{aligned}$$

since $A(x_0)$ is best to $f$. Whence $A(x_0)$ is also a best approximation to $f_\lambda$ and $\|f - A(x_1)\| = \|f - f_\lambda\| + \|f_\lambda - A(x_1)\|$. That is,

$$\begin{aligned} &\int_a^b |f(t) - A(x_1)(t)|\, dt \\ &\quad = (1 - \lambda) \int_a^b |f(t) - A(x_0)(t)|\, dt + \int_a^b |\lambda(f(t) - A(x_0)(t)) \\ &\qquad + A(x_0)(t) - A(x_1)(t)|\, dt. \end{aligned}$$

But $\int_a^b |g + h|\, dt = \int_a^b |g|\, dt + \int_a^b |h|\, dt$ if and only if $\operatorname{sgn} g = \operatorname{sgn} h$ whenever both are nonzero if $g$ and $h$ are continuous. Thus letting $g = f - f_\lambda$ and $h = f_\lambda - A(x_1)$ we find that $\operatorname{sgn}[\lambda(f - A(x_0)) + (A(x_0) - A(x_1))] = \operatorname{sgn}[f - A(x_0)]$. But $f - A(x_0)$ changes sign at least $N$ times in the interior of

$(a, b)$ (see Remark 1). Thus we conclude that $A(x_0) - A(x_1)$ changes sign at least $N$ times as well which implies $A(x_0) = A(x_1)$ by assumption (2). This is a contradiction and the claim is proved. The conclusion of the theorem is now clear. ∎

We shall need the following standard definition.

DEFINITION 1. A subset $M$ of a normed linear space $E$ is called approximatively compact if given any $x \in E$ and any sequence $\{m_v\} \subset M$ such that $\| x - m_v \| \to \text{dist}(x, M)$ then $\{m_v\}$ has a cluster point in $M$.

To extend the techniques of [1, 2] to this setting it is required that the second derivative of $F(x)$ be jointly continuous in $x$ and $f$. The results of Theorem 1 indicate this will be true only if functions "near" $F$ have derivatives near $f'$. Thus we are led to consider the topology $\mathscr{T}$ on $\mathfrak{A}$ given by $f_v \to f$ if and only if $f_v \to f$ uniformly and $f_v' \to f'$ uniformly on $[a, b]$. Clearly, the topology $\mathscr{T}$ is generated by the norm $N(f) = \|f\|_\infty + \|f'\|_\infty$ and so we can consider $\mathfrak{A}$ to be a normed linear space.

Then we have the following local uniqueness theorem.

THEOREM 3. *Assume $\overline{A(S)}$ is approximatively compact and that $f \in \mathfrak{A}$ has a unique best approximation $A(x_0) \in A(S)$ from $\overline{A(S)}$ such that the error curve $\epsilon(t) = A(x_0)(t) - f(t)$ has only simple roots all lying in $(a, b)$ and such that*

$$(*) \qquad \inf_{\|h\|=1} \left\{ \sum_{j=1}^{K} \frac{(A'(x_0, h)(t_j))^2}{|(d\epsilon/dt)(t_j)|} + \int_a^b \text{sgn}(\epsilon(t))\, A''(x_0, h, h)(t)\, dt \right\} > 0$$

*where $t_1, \ldots, t_K$ are the roots of $\epsilon(t)$ in $[a, b]$. Then there is a $\mathscr{T}$-neighborhood $U$ of $f$ such that each $g \in U$ has a unique best approximation in $A(S)$.*

*Proof.* When we proved [2, Theorem 2] we actually showed the following: Assume that $f_0$ has a unique best approximation $A(x_0)$ and that $F''(x_0, h, h)$ is a positive definite quadratic form. If the mappings $(f, x) \to F'$ and $(f, x) \to F''$ are jointly continuous then there is a neighborhood of $f$ such that each $g$ in this neighborhood has a unique best approximation in $A(S)$. It is straightforward to show that these hypotheses are satisfied using the expressions for $F'$ and $F''$ given in Lemma 2 and Theorem 1. ∎

*Remark* 5. The condition that $\overline{A(S)}$ be approximatively compact is satisfied by the ordinary rational function, the exponential families, and most of the standard nonlinear approximating families. Thus, no great restriction has been made.

The following example illustrates that critical points may exist for $F$ which are not local minima so that examination of the second derivative is necessary to check that a given critical point is indeed a (local) minimum.

EXAMPLE 2. Let $S = \{(\alpha, \beta) \in R^2 \mid |\beta| < 1\}$, define $A: S \to C[-1, 1]$ by $A(\alpha, \beta) = \alpha/(1 + \beta t)$ and let $f(t) = t$. Then

$$F(x) = \int_{-1} \left| \frac{\alpha}{1 + \beta t} - t \right| dt \qquad \text{and}$$

$$F'(x, h) = \int_{-1}^{1} \operatorname{sgn}\left(\frac{\alpha}{1 + \beta b} - t\right)\left(\frac{\partial A}{\partial \alpha}(x)\, h_1 + \frac{\partial A}{\partial \beta}(x)\, h_2\right) dt$$

$$= \int_{-1}^{1} \operatorname{sgn}\left(\frac{\alpha}{1 + \beta t} - t\right) \frac{h_1(1 + \beta t) - \alpha h_2 t}{(1 + \beta t)^2}\, dt \quad \text{where } h = (h_1, h_2)^T.$$

Letting $\alpha = \beta = 0$ we have $F'(0, h) = \int_{-1}^{1} \operatorname{sgn}(-t) h_1\, dt = 0$ for each $h = (h_1, h_2)^T \in R^2$. To calculate $A''(0, h, h)$ we note that

$$\frac{\partial A}{\partial \alpha} = -\frac{1}{1 + \beta t}, \qquad \frac{\partial A}{\partial \beta} = -\frac{\alpha t}{(1 + \beta t)^2}$$

so that

$$\frac{\partial^2 A}{\partial \alpha^2} = 0, \qquad \frac{\partial^2 A}{\partial \alpha \partial \beta} = \frac{\partial^2 A}{\partial \beta \partial \alpha} = \frac{-t}{(1 + \beta t)^2}, \qquad \text{and} \qquad \frac{\partial^2 A}{\partial \beta^2} = \frac{2\alpha t^2}{(1 + \beta t)^3}.$$

Thus,

$$A''(0, h, h) = \left[ \begin{pmatrix} \frac{\partial^2 A}{\partial \alpha^2}(0) & \frac{\partial^2 A}{\partial \alpha \partial \beta}(0) \\ \frac{\partial^2 A}{\partial \beta \partial \alpha}(0) & \frac{\partial^2 A}{\partial \beta^2}(0) \end{pmatrix} \begin{pmatrix} h_1 \\ h_2 \end{pmatrix}, \begin{pmatrix} h_1 \\ h_2 \end{pmatrix} \right] = -2t h_1 h_2$$

where $[\cdot, \cdot]$ is the usual inner product of $R^2$.

We note also that $t = 0$ is the only root of $\epsilon(t) = -t$, that it is simple and that $|\epsilon'(0)| = 1$. Thus, using Theorem 1 we have

$$F''(0, h, h) = (A'(0, h)(0))^2 + \int_{-1}^{1} \operatorname{sgn}(-t) - 2t h_1 h_2\, dt = h_1^2 + 2h_1 h_2 \int_{-1}^{1} |t|\, dt$$

$$= h_1^2 + 2h_1 h_2 = h_1[h_1 + 2h_2].$$

Then if $h = (1, -1)^T$ we have $F''(0, h, h) < 0$ while $F''(0, h, h) > 0$ if $h = (1, 1)^T$. The map $x \to F''(x, h, h)$ is well defined for all $h \neq 0$ and continuous in $x$ on some neighborhood of $x = 0$ and so 0 is in fact a saddle point of $F(x)$ and not a local minimum. That is, $F(x)$ is strictly increasing in some directions and strictly decreasing in other directions at $x = 0$ and $F'(0, h) = 0$ for all $h \in R^2$.

As a final extension of the results in [2] to this setting we consider the best approximation operator $\mathscr{P}$. That is, for each $f \in \mathscr{O}$ having a unique best approximation in $A(S)$, $\mathscr{P}$ is the function that assigns to $f$ its best approximation. Following the methods of [2] we have

THEOREM 4. *Assume that the hypotheses of Theorem 3 are satisfied at* $f \in \mathfrak{A}$. *Then there is a* $\mathscr{T}$*-neighborhood of* $f$ *on which the operator* $\mathscr{P}$ *is continuously Frechet differentiable. In particular it is Lipschitz continuous on a* $\mathscr{T}$*-neighborhood of* $f$.

## REFERENCES

1. J. M. WOLFE, On the Unicity of Nonlinear Approximation in Smooth Spaces, *J. Approximation Theory*, **12** (1974), 165–181.
2. J. M. WOLFE, Differentiability of Nonlinear Best Approximation Operators in a Real Inner Product Space, to appear.
3. E. W. CHENEY AND A. A. GOLDSTEIN, Mean-Square Approximation by Generalized Rational Functions, *Math. Z.* **95** (1967), 232–241.
4. C. B. DUNHAM, Nonlinear Mean Approximation, *J. Approximation Theory*, **11** (1974), 134–142.
5. B. R. KRIPKE AND T. J. RIVLIN, Approximation in the Metric of $L(X, \mu)$, *Trans. Amer. Math. Soc.* **119** (1965), 101–122.