JOURNAL OF APPROXIMATION THEORY 37, 147–154 (1983)

Critical Points in Nonlinear L_1 Approximation

JERRY M. WOLFE

Department of Mathematics, University of Oregon, Eugene, Oregon 97403, U.S.A.

Communicated by E. W. Cheney

Received November 5, 1981

Let H be a finite-dimensional Haar subspace of $C[a, b]$ and let $f \in C[a, b]$ be arbitrary. Then the well-known result of Jackson [1] asserts that f has a unique closest point in H in the L_1 sense (using Lebesgue measure, say). In moving from linear to nonlinear L_1 approximation it is natural to consider the uniqueness question for nonlinear families whose local approximating tangent spaces are finite-dimensional Haar spaces. In studying the uniqueness question the critical points of the error functional are of central importance since the best approximations are among them.

In this paper we shall present examples which show that even with a single nonlinear parameter it is possible (at least for a certain class of such families including the exponential and rational families) to produce a continuous function such that the corresponding error functional has a continuum of critical points. A simple modification of the construction yields a continuous function with (at least) a countably infinite set of local minima. Thus the situation is quite different from that of the uniform case [see [2], for example] or the case of L_2 [3, 4].

GENERAL SETTING

A map $A: S \subset E^N \to C[a, b]$ (S open) is given with properties:

(a) $A(x)$ is continuously differentiable on $[a, b]$ for each $x \in S$ and the function $x \to dA(x)(\cdot)/dt$ is continuous from S to $C[a, b]$ using the uniform norm on $C[a, b]$.

(b) $\partial^2 A(x)/\partial x_i \, \partial x_j$ exists and is continuous on S using the uniform norm on $C[a, b]$ for each $1 \leqslant i, j \leqslant N$.

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

(d) $A(x) - A(y)$ has at most $N - 1$ roots in $[a, b]$ if $A(x) \neq A(y)$.

0021-9045/83/020147–08$03.00/0

Two important examples of nonlinear families satisfying (a)–(d) are the (ordinary) rational and exponential families (see [7]). See also the examples later in this paper.

The approximation problem is then:

Problem. Given $f \in C[a, b]$ find $x^* \in S$ such that

$$F(x) \equiv \int_a^b |A(x)(t) - f(t)|\, dt \text{ is minimized when } x = x^*. \tag{1}$$

The following result is well known and we shall not provide a proof.

LEMMA 1. *For each $x \in S$ and $h \in E^N$ let $A'(x, h)$ denote the directional derivative $\sum_{j=1}^N h_j \partial A(x)/\partial x_j$ where $h = (h_1, ..., h_N)$. Then a necessary condition that $x \in S$ be a local minimum of $F(x)$ given by (1) above is that*

$$\left| \int_a^b \operatorname{sgn}(A(x)(t) - f(t))\, A'(x, h)(t)\, dt \right| \leqslant \int_Z |A'(x, h)(t)|\, dt \tag{2}$$

for every $h \in E^N$ where $Z \equiv \{t \mid A(x)(t) - f(t) = 0\}$ and $\operatorname{sgn}(r) = r/|r|$ if $r \neq 0$ and 0 otherwise.

In particular, if $\mu(Z) = 0$ where μ denotes Lebesgue measure, then the Frechet derivative $F'(x, \cdot)$ exists and

$$F'(x, h) = \int_a^b \operatorname{sgn}(A(x)(t) - f(t))\, A'(x, h)(t)\, dt = 0 \tag{3}$$

for all $h \in E^N$ if $x \in S$ is a local minimum of F.

The following result may be found in [5].

LEMMA 2. *Let $x \in S$ and $f \in C[a, b]$ be such that f is differentiable on some open set containing the zeros of $A(x) - f$ and that there are exactly M such zeros all simple. Then the second Frechet derivative $F''(x, \cdot, \cdot)$ exists and in particular*

$$F''(x, h, k) = 2 \sum_{j=1}^M \frac{A'(x, h)(t_j)\, A'(x, k)(t_j)}{|dE(x)(t_j)/dt|} + \int_a^b \operatorname{sgn}(E(x)(t)) \cdot A''(x, h, k)(t)\, dt \tag{4}$$

for all $h, k \in E^N$ where $\{t_1, ..., t_M\}$ are the roots of $A(x) - f$ and $E(x) \equiv A(x) - f$.

Remark 1. In [5] the factor 2 appearing in (4) above inexplicably disappeared in the middle of a calculation and was left out of the statement of the crucial results. Also we now have that if $x \in S$ satisfies (3) and

$$F''(x, h, h) > 0 \qquad \text{for all } h \in E^N,\ h \neq 0 \tag{5}$$

then x is a local minimum of F.

Canonical Points

Let H be a Haar subspace of $C[a,b]$ of dimension N, say $H = \operatorname{span}\{h_1,\ldots,h_N\}$. Then it is known [6] that there exist unique points $\{t_1,\ldots,t_N\}$ with $a = t_0 < t_1 < \cdots < t_N < t_{N+1} = b$ such that

$$\sum_{j=0}^{N} (-1)^j \int_{t_j}^{t_{j+1}} h_i(t)\,dt = 0, \qquad i = 1,\ldots,N. \tag{6}$$

These are called the canonical points for H.

Lemma 3. *Let H be as above and suppose $f \in C[a,b]$ and $h \in H$ are such that $f - h$ changes sign exactly at the canonical points of H. Then $\int_a^b \operatorname{sgn}(f(t) - h(t))\,h_i(t)\,dt = 0$, $i = 1,\ldots,N$, and hence h is the unique closest point to f from H.*

Proof. Since $f - h$ changes sign exactly at the canonical points of H we have that

$$\int_a^b \operatorname{sgn}(f(t) - h(t))\,h_i(t)\,dt = (-1)^\varepsilon \sum_{j=0}^{N} (-1)^j \int_{t_j}^{t_{j+1}} h_i(t)\,dt = 0,$$

$i = 1,\ldots,N$, where $\varepsilon = 0$ or $\varepsilon = 1$. ∎

In order to construct the examples of this paper we need to first investigate the behavior of the canonical points for the local approximating tangent spaces of the nonlinear family as x varies. The following definition is crucial for this.

Definition. For each $x \in S$ let $T(x) = \operatorname{span}\{\partial A(x)/\partial x_1,\ldots,\partial A(x)/\partial x_N\}$ and $d(x) = \dim(T(x))$. Then the point x is called normal if $d(x) = N$.

The importance of the idea of normality lies partly in the fact that for the standard nonlinear families such as the rationals and exponentials any minimum point x satisfying (3) must be normal if $f \neq A(x)$ [see Theorem 8 of [7] noting that it holds for $p = 1$ if (3) holds].

To study the behavior of the canonical points we define a function

$\Phi(x, \mathbf{t}) = (\Phi_1(x, \mathbf{t}),..., \Phi_N(x, \mathbf{t}))$ on $S \times \mathscr{U} \to R^N$ where $\mathscr{U} = \{(t_1,..., t_N) \mid a = t_0 < t_1 < \cdots < t_N < t_{N+1} = b\}$ and where

$$\Phi_i(x, \mathbf{t}) = \sum_{j=0}^{N} (-1)^j \int_{t_j}^{t_{j+1}} \frac{\partial A}{\partial x_i}(x)(t)\, dt, \qquad i = 1,..., N. \tag{7}$$

Then assuming x is normal

$$\Phi_i(x, \mathbf{t}(x)) = 0, \qquad i = 1,..., N \tag{8}$$

where $\mathbf{t}(x) = (t_1(x),..., t_N(x))$ and $\{t_1(x),..., t_N(x)\}$ are the canonical points for $T(x)$. A simple calculation shows that

$$\frac{\partial \Phi_i}{\partial t_k}(x, \mathbf{t}) = (-1)^{k-1} 2 \frac{\partial A}{\partial x_i}(x)(t_k), \qquad i = 1,..., N; k = 1,..., N. \tag{9}$$

$$\frac{\partial \Phi_i}{\partial x_k}(x, \mathbf{t}) = \sum_{j=0}^{N} (-1)^j \int_{t_j}^{t_{j+1}} \frac{\partial^2 A}{\partial x_i\, \partial x_k}(x)(t)\, dt, \qquad i = 1,..., N; k = 1,..., N. \tag{10}$$

Let $\partial\Phi/\partial\mathbf{t}$ and $\partial\Phi/\partial x$ denote the Jacobian matrices $(\partial\Phi_i/\partial t_k)_{1 \leqslant i,k \leqslant N}$ and $(\partial\Phi_i/\partial x_k)_{1 \leqslant i,k \leqslant N}$, respectively. Then using elementary properties of determinants we have

$$\det\left(\frac{\partial \Phi}{\partial \mathbf{t}}\right) = 2^N (-1)^{N(N-1)/2} \det\left(\frac{\partial A}{\partial x_i}(x)(t_k)\right)_{1 \leqslant i,k \leqslant N} \neq 0 \tag{11}$$

if x is normal since $T(x)$ is then a Haar space of dimension N and so $(\partial A(x)(t_k)/\partial x_i)_{1 \leqslant i,k \leqslant N}$ is nonsingular by the unique interpolation property of a Haar space. Thus, we have the following result.

LEMMA 4. *Let $x \in S$ be normal and let Φ be as above. Then the canonical point map $x \to \mathbf{t}(x)$ is differentiable on some open neighborhood of x and in fact*

$$\frac{\partial \mathbf{t}}{\partial x}(x) = -\left[\frac{\partial \Phi}{\partial \mathbf{t}}(x, \mathbf{t}(x))\right]^{-1} \frac{\partial \Phi}{\partial x}(x, \mathbf{t}(x)). \tag{12}$$

Proof. Equations (7)–(11) imply that the implicit function theorem applies and by the uniqueness of canonical points the resulting map must in fact be the canonical point map $x \to \mathbf{t}(x)$. The implicit function theorem then yields (12). ∎

Equation (12) shows that the canonical points change smoothly as x varies over a sufficiently small neighborhood of any normal point. We will now specialize to the case where the approximating family contains one linear and one nonlinear parameter although the construction would generalize easily to a larger number of linear parameters.

Assume now that $A(x)(t)$ is of the form

$$A(x)(t) = x_1 \gamma(x_2)(t) \tag{13}$$

where for each $x = (x_1, x_2) \in S = E \times U$ (U open in E) $\gamma(x_2)$ is either a nonzero constant function or a strictly monotone and nonzero function on $[a, b]$. In addition to properties (a)–(d) we shall assume

(e) If $x_\nu \to x_0$, $\{x_\nu\} \subset S$, $x_0 \in S$, then the first derivative of $A(x_\nu)$ (with respect to t) converges uniformly to the first derivative of $A(x_0)$ as $\nu \to \infty$.

(f) For each x_2, $\{\gamma(x_2), \partial\gamma(x_2)/\partial x_2, \partial^2\gamma(x_2)/\partial x_2^2\}$ spans a Haar space of dimension 3.

EXAMPLES. Let $\gamma(x_2)(t) = e^{x_2 t}$ (where $U = E$) on an arbitrary interval $[a, b]$ or let $\gamma(x_2)(t) = 1/(1 + x_2 t)$, $t \in [-1, 1]$, where $U = (-1, 1)$. Then assumptions (a)–(f) are easily seen to be satisfied.

The following formulas are clear

$$\frac{\partial A}{\partial x_1}(x) = \gamma(x_2),$$

$$\frac{\partial A}{\partial x_2}(x) = x_1 \frac{\partial \gamma}{\partial x_2}(x_2), \tag{14}$$

$$\frac{\partial^2 A}{\partial x_1^2}(x) = 0; \qquad \frac{\partial^2 A}{\partial x_1\, \partial x_2}(x) = \frac{\partial^2 A}{\partial x_2\, \partial x_1}(x) = \frac{\partial \gamma}{\partial x_2}(x_2); \qquad \frac{\partial^2 A}{\partial x_2^2} = x_1 \frac{\partial^2 \gamma}{\partial x_2^2}(x_2).$$

For each normal $x = (x_1, x_2)$ the corresponding canonical points $t_1(x)$ and $t_2(x)$ depend only on x_2. That is, if $x' = (x_1', x_2)$ and $x = (x_1, x_2)$ with x_1 and x_1' both nonzero, then x and x' have the same canonical points since $\text{span}\{\gamma(x_2), x_1'\, \partial\gamma(x_2)/\partial x_2\} = \text{span}\{\gamma(x_2), x_1\, \partial\gamma(x_2)/\partial x_2\}$. Thus we shall denote the canonical points by $t_1(x_2)$ and $t_2(x_2)$ where $a < t_1(x_2) < t_2(x_2) < b$. Applying formulas (9), (10), and (14) we obtain

$$\frac{\partial \Phi}{\partial \mathbf{t}}(x, \mathbf{t}(x)) = 2 \begin{pmatrix} \gamma(x_2)(t_1(x_2)) & -\gamma(x_2)(t_2(x_2)) \\ x_1 \dfrac{\partial \gamma}{\partial x_2}(x_2)(t_1(x_2)) & -x_1 \dfrac{\partial \gamma}{\partial x_2}(x_2)(t_2(x_2)) \end{pmatrix},$$

$$\frac{\partial \Phi}{\partial x}(x, \mathbf{t}(x)) = \begin{pmatrix} 0 & 0 \\ 0 & x_1 \displaystyle\int_a^b \text{sgn}(\varepsilon(t)) \frac{\partial^2 \gamma}{\partial x_2^2}(x_2)(t)\, dt \end{pmatrix}, \tag{15}$$

where $\varepsilon(t) = 1$ on $[a, t_1(x_2))$, -1 on $(t_1(x_2), t_2(x_2))$, and 1 on $[t_2(x_2), b]$. Then applying (12) we arrive at

$$\frac{\partial \mathbf{t}}{\partial x} = \begin{pmatrix} 0 & \gamma(x_2)(t_2(x_2)) \cdot \dfrac{\sigma}{\Delta} \\ 0 & \gamma(x_2)(t_1(x_2)) \cdot \dfrac{\sigma}{\Delta} \end{pmatrix}, \tag{16}$$

where

$$\sigma = \int_a^b \operatorname{sgn}(\varepsilon(t)) \frac{\partial^2 \gamma}{\partial x_2^2}(x_2)(t)\, dt$$

and

$$\Delta = \det \begin{pmatrix} \gamma(x_2)(t_1(x_2)) & \dfrac{\partial \gamma}{\partial x_2}(x_2)(t_1(x_2)) \\ \gamma(x_2)(t_2(x_2)) & \dfrac{\partial \gamma}{\partial x_2}(x_2)(t_2(x_2)) \end{pmatrix} \neq 0.$$

Since $\{\gamma(x_2), \partial\gamma(x_2)/\partial x_2, \partial^2\gamma(x_2)/\gamma x_2^2\}$ spans a Haar subspace H of $C[a, b]$ of dimension 3 and since $\varepsilon(t)$ has exactly two sign changes in $[a, b]$ and is nonzero almost everywhere, σ is *not* zero. (That is there is an element $h \in H$ such that $\operatorname{sgn} h = \operatorname{sgn} \varepsilon$ almost everywhere). Thus $\partial t_1(x_2)/\partial x_2$ and $\partial t_2(x_2)/\partial x_2$ are both nonzero and have the same sign. We are now ready for the main result of this paper.

THEOREM. *There exists an $f \in C[a, b]$ whose error functional $F(x) = \int_a^b |A(x)(t) - f(t)|\, dt$ has a continuum of critical points for the nonlinear family $\{A(x) \mid x \in S\}$ given by* (13).

Proof. Let $\bar{x}_2 \in U$ be such that $\gamma(\bar{x}_2)$ is strictly monotone on $[a, b]$ and let $t_1(\bar{x}_2)$ and $t_2(\bar{x}_2)$ be the corresponding canonical points and define t_* by $t_* = (t_1(\bar{x}_2) + t_2(\bar{x}_2))/2$. Without loss of generality assume that $\partial t_1(\bar{x}_2)/\partial x_2$ and $\partial t_2(\bar{x}_2)/\partial x_2$ are positive. Then by continuity there is an interval $[\bar{x}_2, \bar{x}_2 + \delta] \equiv I$ so that the images $t_1(I)$ and $t_2(I)$ are of form

$$\begin{aligned} t_1(I) &= [t_1(\bar{x}_2), t_1(\bar{x}_2) + \varepsilon_1] \equiv I_1, \\ t_2(I) &= [t_2(\bar{x}_2), t_2(\bar{x}_2) + \varepsilon_2] \equiv I_2, \end{aligned} \tag{17}$$

where $t_* \notin I_1 \cup I_2$, $\gamma(\bar{x}_2)$ is strictly monotone on $[a, b]$ and $\partial t_1/\partial x_2$ and $\partial t_2/\partial x_2$ are both positive for each $x_2 \in I$. More precisely, the continuity of the maps $x_2 \to t_1(x_2)$ and $x_2 \to t_2(x_2)$ on $I \to (a, b)$, the continuity of the maps

$x_2 \to \gamma(x_2)$, $x_2 \to \partial t_1(x_2)/\partial x_2$, and $x_2 \to \partial t_2(x_2)/\partial x_2$ from $U \to C[a, b]$ (uniform norm), and property (e) are being invoked to obtain the above conclusion.

From the above we conclude that there are inverse maps x_2^1: $t \to x_2^1(t)$ and $x_2^2(t)$: $t \to x_2^2(t)$ defined on I_1 and I_2 with range I for the maps t_1 and t_2. Now for each $x_2 \in I$, define x_1 by $x_1 = 1/\gamma(x_2)(t_*)$ and let $A(x)(t) = x_1 \gamma(x_2)(t)$ where $x = (x_1, x_2)$. Then $A(x)(t_*) = 1$ for each $x_2 \in I$ and by assumption (d), if $x_2 \neq x_2'$, then $A(x) - A(x')$ has t_* as its only zero when $x = (x_1, x_2)$ and $x' = (x_1', x_2')$. Since $A(x)$ only depends on x_2 we shall let $\tau(x_2)$ denote $A(x)$ for $x = (1/\gamma(x_2)(t_*), x_2)$.

Then by strict monotonicity of $\tau(x_2)$ on $[a, b]$ for each $x_2 \in I$ and the fact that $\tau(x_2)(t_*) = 1$ for all $x_2 \in I$ we have that one of the following two situations must occur.

(i) $a \leqslant t < t_*$ and $\bar{x}_2 \leqslant x_2 < \tilde{x}_2 \leqslant \bar{x}_2 + \delta \Rightarrow \tau(x_2)(t) > \tau(\tilde{x}_2)(t)$, $t_* \leqslant t \leqslant b$ and $\bar{x}_2 \leqslant x_2 < \tilde{x}_2 \leqslant \bar{x}_2 + \delta \Rightarrow \tau(x_2)(t) < \tau(\tilde{x}_2)(t)$;

(ii) $a \leqslant t < t_*$ and $\bar{x}_2 \leqslant x_2 < \tilde{x}_2 \leqslant \bar{x}_2 + \delta \Rightarrow \tau(x_2)(t) < \tau(\tilde{x}_2)(t)$, $t_* < t \leqslant b$ and $\bar{x}_2 \leqslant x_2 < \tilde{x}_2 \leqslant \bar{x}_2 + \delta \Rightarrow \tau(x_2)(t) > \tau(\tilde{x}_2)(t)$.

Without loss of generality assume (i) holds, the construction being completely analogous for case (ii). We are now ready to define a continuous function f that has each $\tau(x_2)$ as a critical point, $x_2 \in I$. To do this consider the following five intervals. $J_1 = [a, t_1(\bar{x}_2))$; $J_2 = [t_1(\bar{x}_2), t_1(\bar{x}_2) + \varepsilon_1]$; $J_3 = (t_1(\bar{x}_2) + \varepsilon_1, t_2(\bar{x}_2))$; $J_4 = [t_2(\bar{x}_2), t_2(\bar{x}_2) + \varepsilon_2]$; $J_5 = (t_2(\bar{x}_2) + \varepsilon_2, b]$. Now define f to be continuous on $[a, b]$ with the following properties. On J_2 let $f(t) = \tau(x_2^1(t))(t)$ and on J_4 let $f(t) = \tau(x_2^2(t))(t)$ while on the remaining three intervals the inequalities $f(t) > \tau(\bar{x}_2)(t)$ $t \in J_1$; $f(t) < \min\{\tau(\bar{x}_2 + \delta)(t), \tau(\bar{x}_2)(t)\}$ $t \in J_3$; and $f(t) > \tau(\bar{x}_2 + \delta)(t)$, $t \in J_5$ should hold.

The functions $\tau(x_2^1(t))(\cdot)$ and $\tau(x_2^2(t))(\cdot)$ are continuous on J_2 and J_4, respectively, being the composition of continuous functions. That a continuous f exists with the required properties on $J_1 \cup J_3 \cup J_5$ is then clear by considering (i).

Claim. For each $x_2 \in I$, $f - \tau(x_2)$ changes sign exactly at $t_1(x_2)$ and $t_2(x_2)$.

Proof. We first check that $f - \tau(x_2)$ changes sign at $t_1(x_2)$ and $t_2(x_2)$. First $f(t_1(x_2)) - \tau(x_2)(t_1(x_2)) = \tau(x_2^1(t_1(x_2)) - \tau(x_2)(t_1(x_2)) = \tau(x_2)(t_1(x_2)) - \tau(x_2)(t_1(x_2)) = 0$ since by definition x_2^1 and t_1 are inverse functions of each other. Similarly $f - \tau(x_2)$ vanishes at $t_2(x_2)$. Assume that $t_1(x_2) > t_1(\bar{x}_2)$. Then if $t_1(\bar{x}_2) < t < t_1(x_2)$, $f(t) = \tau(x_2^1(t))(t) > \tau(x_2)(t) = \tau(x_2^1(t_1(x_2))(t)$ since $x_2^1(t) < x_2 = x_2^1(t_1(x_2))$ if $t < t_1(x_2)$. The same reasoning shows that $f(t) < \tau(x_2)(t)$ if $t > t_1(x_2)$. Thus $f - \tau(x_2)$ changes sign at x_2 (the case $t_1(x_2) = t_1(\bar{x}_2)$ is similar) and *only* at $t_1(x_2)$ on J_2. A completely analogous argument shows that $f - \tau(x_2)$ changes sign on J_4 exactly at $t_2(x_2)$.

Finally it remains to show that these are the only sign changes that occur for $f - \tau(x_2)$ on $[a, b]$. But by definition, $f - \tau(x_2)$ is positive on J_1, negative on J_3, and positive on J_5 and hence can have no other sign changes. Since $f - \tau(x_2)$ changes sign exactly at the canonical points for x_2, $x = (x_1, x_2)$ is a critical point of $F(x)$. Since x_2 was arbitrary in I, $F(x)$ has a continuum of critical points. ∎

Remark 2. In the construction above, the function f is continuously differentiable on $J_2 \cup J_4$ and can be constructed to be in $C^1[a, b]$. Also, if in the construction we use a sequence of points $\mathbf{t}_v = (t_{1v}, t_{2v})$ where $t_{iv} \downarrow t_i \equiv t_i(\bar{x}_2)$ with $t_{iv+1} < t_{iv}$, $v = 1, 2, \ldots$, $i = 1, 2$ and where $\mathbf{t}_v$ represents the canonical points for, say, x_{2v}, then referring to Remark 1 we can define f so that $f \in C^1[a, b]$ and f has each $x_v = (x_{1v}, x_{2v})$ as an isolated local minimum. It is an open question whether or not the x_v's could be made to be global minima.

References

1. E. W. Cheney, "Introduction to Approximation Theory," McGraw–Hill, New York, 1966.
2. J. R. Rice, "The Approximation of Functions," Vol. II, Addison–Wesley, Reading, Mass., 1969.
3. J. Spiess, "Best Approximation in Strictly Convex Spaces," Thesis, University of Hamburg, 1969.
4. D. Braess, On rational L_2-approximation, *J. Approx. Theory* **18** (1976), 136–152.
5. J. Wolfe, Nonlinear L_1 approximation of smooth functions, *J. Approx. Theory* **17** (1976), 166–176.
6. R. B. Barrar and H. L. Loeb, The fundamental theorem of algebra and the interpolating envelope for totally positive perfect splines, *J. Approx. Theory* **34** (1982), 167–186.
7. J. Wolfe, On the unicity of nonlinear approximation in smooth spaces, *J. Approx. Theory* **12** (1974), 165–181.