JOURNAL OF APPROXIMATION THEORY 58, 164–183 (1989)

# Local Lipschitz and Strong Unicity Constants for Certain Nonlinear Families

JAMES R. ANGELOS, MYRON S. HENRY, EDWIN H. KAUFMAN, JR., AND TERRY D. LENKER

*Department of Mathematics, Central Michigan University, Mount Pleasant, Michigan 48859, U.S.A.*

AND

ANDRÁS KROÒ

*Mathematical Institute of the Hungarian Academy of Sciences, Budapest, Hungary*

*Communicated by E. W. Cheney*

Received December 1, 1986; revised February 4, 1988

Let $X$ be a compact metric space, and let $V = \{F(a, x): a \in A\}$ where $A$ is an open subset of $\mathbf{R}^n$, and $F(a, x)$ and $\partial F/\partial a_i$, $1 \leqslant i \leqslant n$, are continuous on $A \times X$. Suppose $f \in C(X)$ is weakly normal; that is (i) $f$ has a best approximation $F(a^*, \cdot) = B_V(f)$ such that $N = \dim W(a^*) \equiv \dim \operatorname{span}\{(\partial F/\partial a_i)(a^*, \cdot): 1 \leqslant i \leqslant n\}$ is maximal, and (ii) certain weakened versions of the local Haar condition, a sign property equivalent to a form of asymptotic convexity, and Property Z hold. For those weakly normal functions $f$ for which $\{x \in X: |f(x) - F(a^*, x)| = \|f - F(a^*, \cdot)\|\}$ has exactly $N+1$ points, we give constructions of the local Lipschitz and strong unicity constants, as well as show that $B_V(f)$ is differentiable. © 1989 Academic Press, Inc.

## 1. INTRODUCTION

In the setting of uniform approximation by algebraic polynomials in a single real variable, characterizations for the strong unicity constant were developed in [5, 18]. In [10, 1] a characterization for the local Lipschitz constant was developed and it was shown that under certain conditions the norm of the derivative of the best approximation operator equals the local Lipschitz constant. It is the purpose of this paper to extend these results to a much more general setting, which includes, e.g., as a special case, the situation of generalized rational approximation on an arbitrary compact metric space. In this section we describe our setting, which is similar to that



0021-9045/89 $3.00

in [15, 7]; in Section 2 we prove some preliminary results; in Section 3 we review and expand some results for the linear case; and we prove our main results in Section 4.

Let $n$ be a positive integer, and let $A$ be an open subset of $\mathbf{R}^n$. For any $a=(a_1, ..., a_n) \in \mathbf{R}^n$, $\|a\|$ will denote $\max\{|a_i|: 1 \leqslant i \leqslant n\}$. Let $X$ be a compact metric space with a least $n+1$ points, and let $r(x, y)$ denote the distance between $x, y \in X$. Given $f \in C(X)$, we define $\|f\| = \sup\{|f(x)|: x \in X\}$. For any positive integer $k$ and any $S \subset X$, $\hat{S}_k$ will denote $\{(x_1, ..., x_k): x_1, ..., x_k$ distinct points in $S\}$. Let $V$ be a set of continuous functions defined on $A \times X$, where for all $F \in V$ we also assume $(\partial F/\partial a_i)(a, x) \equiv F_i(a, x)$ is continuous on $A \times X$ for $i=1, ..., n$. For any $a \in A$, set $W(a) = \{D(a, b, x) \equiv \sum_{i=1}^n b_i F_i(a, x)$: $b=(b_1, ..., b_n) \in \mathbf{R}^n\}$ and let $d(a) = \dim W(a)$. Let $N = \max\{d(a): a \in A\}$; evidently $N \leqslant n$. Given $f \in C(X)$ and $a \in A$, let $E_a(f) = \{x \in X: |f(x) - F(a, x)| = \|f - F(a, \cdot)\|\}$. We say $F(a^*, \cdot) \in V$ is a best approximation to $f \in C(X)$ on $X$ from $V$ if $\|f - F(a^*, \cdot)\| \leqslant \|f - F(a, \cdot)\|$, for all $a \in A$. If $F(a^*, \cdot)$ is unique, we will often denote it by $B_V(f)$, and we will also use the notation $e_V(f) = f - B_V(f)$ and $E_V(f) = E_{a^*}(f)$. The notation $B_V(f, S)$ will mean the unique best approximation to $f$ on $S$ from $V$, where $S \subset X$.

Definition 1. *Let* $f \in C(X)$.

(a) *The global Lipschitz constant is defined as*

$$\lambda_V(f) = \sup\left\{\frac{\|B_V(f) - B_V(g)\|}{\|f-g\|}: f \neq g,\ g \in C(X)\right\}.$$

(b) *For* $\delta > 0$, *let*

$$\lambda_V(f, \delta) = \sup\left\{\frac{\|B_V(f) - B_V(g)\|}{\|f-g\|}: 0 < \|f-g\| \leqslant \delta,\ g \in C(X)\right\}.$$

*Then*

$$\hat{\lambda}_V(f) = \lim_{\delta \to 0^+} \lambda_V(f, \delta)$$

*is the local Lipschitz constant.*

(c) *The strong unicity constant is defined as*

$$M_V(f) = \sup\left\{\frac{\|F(a, \cdot) - B_V(f)\|}{\|f - F(a, \cdot)\| - \|f - B_V(f)\|}: a \in A;\ F(a, \cdot) \neq B_V(f)\right\}.$$

(d) For $\delta > 0$, let

$$M_V(f, \delta) = \sup \left\{ \frac{\|F(a, \cdot) - B_V(f)\|}{\|f - F(a, \cdot)\| - \|f - B_V(f)\|} : a \in A; 0 < \|F(a, \cdot) - B_V(f)\| \leqslant \delta \right\}.$$

Then

$$\hat{M}_V(f) = \lim_{\delta \to 0^+} M_V(f, \delta)$$

is the local strong unicity constant.

We remark here that in case $V$ is a linear subspace it was shown in [11] that $M_V(f, \delta) = M_V(f)$ and so the local strong unicity constant only makes sense when $V$ is a nonlinear set. As stated earlier, it was demonstrated in [10, 1] that under certain conditions the local Lipschitz constant turns out to be the norm of the derivative of $B_V$. The following definition makes this concept of derivative precise.

Definition 2. The best approximation operator $B_V$ has at $f \in C(X)$ a one-sided derivative denoted by $D_f B_V : C(X) \to V$ if for each $g \in C(X)$ the limit

$$\lim_{t \to 0^+} \frac{B_V(f + tg) - B_V(f)}{t} = D_f B_V(g)$$

exists. In case $D_f B_V(g) = -D_f B_V(-g)$, we say $B_V$ is differentiable at $f$. If in addition the derivative $D_f B_V$ is a linear operator of direction $g$ then $B_V$ is called Gâteaux differentiable at $f \in C(X)$.

The study of the differentiability of $B_V$ in $C(X)$ was begun by Kroò [13] where a characterization of those $f \in C(X)$ where $B_V$ is Gâteaux differentiable is given when $X$ is an interval and $V$ is a linear space satisfying the Haar condition. These results have subsequently been extended to the setting where $X$ is an interval and $V$ is the set of ordinary rational functions [12]. In this report, studies of the local Lipschitz constant for the aforementioned general nonlinear families will lead to an extension of the differentiability of $B_V$ to these families.

We next define three properties which will appear as hypotheses in many of our results.

Definition 3. Let $a \in A$ and $S \subset X$.

(a) We say that property SIGN$(a, S)$ holds if for all $S_0 \subset S$, $S_0$ compact, for all $b \in A$ with $F(a, x) \neq F(b, x)$, for all $x \in S_0$, there exists a $\phi \in W(a)$ with $\operatorname{sgn} \phi(x) = \operatorname{sgn}(F(b, x) - F(a, x))$, for all $x \in S_0$.

(b) We say property LH($a$, $S$) holds if for every $m \leqslant d(a)$ and every set of distinct points $\{x_1, ..., x_m\} \subset S$, the set of vectors $\{[h_1(x_j), ..., h_k(x_j)]: j = 1, ..., m\}$ is linearly independent, where $k = d(a)$ and $\{h_1, ..., h_k\}$ is a basis for $W(a)$.

(c) We say property Z($a$, $S$) holds if there is a $\delta > 0$ such that for $\bar{a} \in A$ with $\|\bar{a} - a\| \leqslant \delta$ and for all $b \in A$, if $F(b, \cdot) - F(\bar{a}, \cdot)$ has more than $d(\bar{a}) - 1$ zeros in $S$, then $F(b, \cdot) \equiv F(\bar{a}, \cdot)$ on $X$.

We remark here that property SIGN($a$, $S$) is equivalent to asymptotic convexity essentially as defined in [15] restricted to subsets of $S$. Also property SIGN($a$, $S$) is similar to the definition of the "Vorzeichenbedingung" found in [7, p. 68]. Properties LH($a$, $S$) and Z($a$, $S$) are, respectively, essentially the local Haar condition and property Z as defined in [3, 7, 15] but restricted to $S$.

DEFINITION 4. Let $f \in C(X)$. We say that $f$ is weakly normal if $f$ has a best approximation $F(a^*, \cdot)$ with $d(a^*) = N$ such that SIGN($a^*$, $E_{a^*}(f)$), LH($a^*$, $E_{a^*}(f)$), and Z($a^*$, $E_{a^*}(f)$) all hold.

This weak normality condition is the same as the usual normality condition with the exception that we only require the three properties of Definition 2 to hold on the set of extreme points, $E_{a^*}(f)$. We also point out that this definition is motivated by the observation made in [15, p. 138] that the local Haar condition could in some cases be weakened by requiring the local Haar condition only on the set of extreme points.

Examples of the kinds of settings our results can be applied to are generalized rational functions, sums of exponential functions with non-coalescing frequencies on compact subsets of the real line, and of course linear subspaces. Our results also can be applied to certain generalized rational functions, which are not varisolvent as defined by Rice [16, 17].

## 2. PRELIMINARY RESULTS

We now state a lemma which collects some reslts on the set $V$ which will be needed later.

LEMMA 1. *Suppose $a^* \in A$ is fixed with $d(a^*) = k$. Then*

(a) *If $\|a - a^*\|$ is sufficiently small, then $F(a, \cdot) - F(a^*, \cdot) = D(a^*, a - a^*, \cdot) + o(\|a - a^*\|)$.*

(b) *$\|F(a, \cdot) - F(a^*, \cdot)\| = O(\|a - a^*\|)$ as $\|a - a^*\| \to 0$.*

(c) *Suppose* $k=N$. *Let* $x_1, \ldots, x_N$ *be distinct points in* $X$ *such that* $\mathrm{LH}(a^*, \{x_1, \ldots, x_N\})$ *and* $\mathrm{Z}(a^*, \{x_1, \ldots, x_N\})$ *hold. Let* $I$ *be a set of* $N$ *indices such that* $\{F_i(a^*, \cdot): i \in I\}$ *forms a basis for* $W(a^*)$. *Then for all sufficiently small* $\varepsilon > 0$, *there exists a* $\delta = \delta(\varepsilon) > 0$ *such that if* $|F(a^*, x_i) - c_i| \leqslant \delta$, $i = 1, \ldots, N$, *then there is a unique* $a \in A$ *satisfying* (i) $F(a, x_i) = c_i$, $i = 1, \ldots, N$, (ii) $a_i = a_i^*$ *for* $i \notin I$, *and* (iii) $\|a - a^*\| \leqslant \varepsilon$. *Furthermore,* $F(a, \cdot) \in V$ *is the only element of* $V$ *satisfying* (i); *that is, if also* $\bar{a} \in A$ *and* $F(\bar{a}, x_i) = c_i$, $i = 1, \ldots, N$, *then* $F(\bar{a}, x) = F(a, x)$ *for all* $x \in X$. *We may also assume that* $\|a - a^*\| = O(\|F(a, \cdot) - F(a^*, \cdot)\|)$ *as* $\|F(a, \cdot) - F(a^*, \cdot)\| \to 0$.

*Proof.* The proof of (a) follows from the fact that $F_i(a, x)$, $i = 1, \ldots, n$, are continuous on $A \times X$ and (b) follows from (a). Part (c) and its proof can be found in [3].

The first theorem we present contains results on best approximation, in particular, a "zero in the convex hull" characterization of a best approximation, an inclusion theorem, and a strong uniqueness result. However, the main purpose of the theorem is to establish a type of generalized alternation theorem in our setting.

THEOREM 1. *Let* $a^* \in A$ *be fixed with* $d(a^*) = k$.

(a) *Suppose* $f \in C(X)$ *and* $F(a^*, \cdot) \in V$ *are such that* $\mathrm{SIGN}(a^*, E_{a^*}(f))$ *holds. Then* $F(a^*, \cdot)$ *is a best approximation to* $f$ *if and only if there is no* $\phi \in W(a^*)$ *such that* $(f(x) - F(a^*, x))\, \phi(x) > 0$ *for all* $x \in E_{a^*}(f)$.

(b) *Under the hypotheses of part* (a), $F(a^*, \cdot)$ *is a best approximation to* $f$ *if and only if the zero of k-dimensional real space,* $O_k$ *lies in* $\mathrm{co}\{\mathrm{sgn}(f(x) - F(a^*, x))[h_1(x), \ldots, h_k(x)] : x \in E_{a^*}(f)\}$, *where* co *denotes the convex hull and* $\{h_1, \ldots, h_k\}$ *is any basis for* $W(a^*)$.

(c) *Suppose* $(x_0, \ldots, x_k) \in \hat{X}_k$ *is such that* $\mathrm{LH}(a^*, \{x_0, \ldots, x_k\})$ *holds. Then there is a unique set of signs* $\sigma_0, \ldots, \sigma_k$ *depending on* $a^*$ *and* $x_0, \ldots, x_k$ *such that* $\sigma_0 = 1$, $|\sigma_i| = 1$ *for* $i = 1, \ldots, k$, *and* $O_k$ *lies in* $\mathrm{co}\{\sigma_i[h_1(x_i), \ldots, h_k(x_i)]: i = 0, \ldots, k\}$, *where* $\{h_1, \ldots, h_k\}$ *is any basis for* $W(a^*)$. *Furthermore,* $\sigma_0, \ldots, \sigma_k$ *are independent of the choice of basis for* $W(a^*)$.

(d) *Suppose* $d(a^*) = N$, *and* $(x_0, \ldots, x_N) \in \hat{X}_N$ *is such that* $\mathrm{LH}(a^*, \{x_0, \ldots, x_N\})$ *holds. Then there is a* $\delta > 0$ *such that if* $a \in A$, $(y_0, \ldots, y_N) \in \hat{X}_N$ *satisfy* $\|a - a^*\| \leqslant \delta$, *and* $r(x_i, y_i) \leqslant \delta$, $0 \leqslant i \leqslant N$, *then* $d(a) = N$, $\mathrm{LH}(a, \{y_0, \ldots, y_N\})$ *holds, and the signs associated with* $a^*$ *and* $x_0, \ldots, x_N$ *are identical with those associated with* $a$ *and* $y_0, \ldots, y_N$.

(e) *Suppose* $(x_0, \ldots, x_k) \in \hat{X}_k$ *is such that* $\mathrm{LH}(a^*, \{x_0, \ldots, x_k\})$ *and* $\mathrm{SIGN}(a^*, \{x_0, \ldots, x_k\})$ *hold. Then there is no* $a \in A$ *for which* $\sigma_i(F(a, x_i) - F(a^*, x_i)) > 0$, $i = 0, \ldots, k$, *or* $\sigma_i(F(a, x_i) - F(a^*, x_i)) < 0$, $i = 1, \ldots, k$.

(f) *Suppose* $(x_0, ..., x_k) \in \hat{X}_k$ *is such that* $\mathrm{LH}(a^*, \{x_0, ..., x_k\})$ *and* $\mathrm{SIGN}(a^*, \{x_0, ..., x_k\})$ *hold. Suppose* $f \in C(X)$, *and let* $\sigma_0, ..., \sigma_k$ *be signs as in part* (c). *Suppose further that* $\mathrm{sgn}(f(x_i) - F(a^*, x_i)) = \sigma\sigma_i$, $0 \leqslant i \leqslant k$, *for some* $\sigma = -1, 0,$ *or* 1. *Then* $\inf\{\|f - F(a, \cdot)\| : a \in A\} \geqslant \min\{|f(x_i) - F(a^*, x_i)| : 0 \leqslant i \leqslant k\}$.

(g) *Suppose* $\mathrm{LH}(a^*, E_{a^*}(f))$ *and* $\mathrm{SIGN}(a^*, E_{a^*}(f))$ *hold. Then* $F(a^*, \cdot)$ *is a best approximation to* $f \in C(X)$ *if and only if* $f - F(a^*, \cdot)$ *possesses a "σ-alternant" of length* $k+1$, *that is, a* $(k+1)$*-tuple* $(x_0, ..., x_k) \in \hat{X}_k$ *with signs* $\sigma_0, ..., \sigma_k$ *as in part* (c) *such that* $x_i \in E_{a^*}(f)$ *and* $\mathrm{sgn}(f(x_i) - F(a^*, x_i)) = \sigma\sigma_i$, $i = 0, ..., k$, *for some* $\sigma = -1, 0,$ *or* 1. *We shall denote this σ-alternant by* $\sigma_V(x_0, ..., x_k; a^*)$. *For the converse part, the assumptions that* $\mathrm{LH}(a^*, E_{a^*}(f))$ *and* $\mathrm{SIGN}(a^*, E_{a^*}(f))$ *hold can be replaced by the weaker assumptions that* $\mathrm{LH}(a^*, \{x_0, ..., x_k\})$ *and* $\mathrm{SIGN}(a^*, \{x_0, ..., x_k\})$ *hold.*

(h) *Suppose* $d(a^*) = N$, *and* $f \in C(X)$ *is such that* $f - F(a^*, \cdot)$ *possesses a σ-alternant* $\sigma_V(x_0, ..., x_N; a^*)$ *of length* $N+1$ *with* $\mathrm{LH}(a^*, \{x_0, ..., x_N\})$, $\mathrm{SIGN}(a^*, \{x_0, ..., x_N\})$, *and* $Z(a^*, \{x_0, ..., x_N\})$ *all holding. Then the best approximation* $F(a^*, \cdot)$ *is strongly unique*; *that is, there is a* $\gamma > 0$ *such that*

$$\|f - F(a, \cdot)\| \geqslant \|f - F(a^*, \cdot)\| + \gamma \, \|F(a, \cdot) - F(a^*, \cdot)\|$$

*for all* $a \in A$.

*Proof.* The proof of (a) follows as a result of property $\mathrm{SIGN}(a^*, E_{a^*}(f))$ and arguments similar to those in [15, Theorem 87] and [7, Satz 5.2]. Part (b) results from part (a) and the theorem on linear inequalities [4, p. 19]. Part (f) is the usual type of inclusion result or a general de la Valleè Poussin theorem [4, p. 77] and is well known (see, e.g., [15, Theorem 85]; its proof follows immediately from (e). The proof of (h) follows by arguments similar to those in [3, 9], where part (e) above is used in place of the usual zero counting on an interval. We will give proofs of (c), (d), (e), and (g).

(c) Let $\{h_1, ..., h_k\}$ be any basis for $W(a^*)$, and consider the equation

$$\sum_{i=0}^{k} \theta_i \sigma_i [h_1(x_i), ..., h_k(x_i)] = 0_k. \tag{2.1}$$

This is an underdetermined homogeneous linear system and so it has nontrivial solutions. Now since $\mathrm{LH}(a^*, \{x_0, ..., x_k\})$ holds every set of vectors $\{[h_1(z_1), ..., h_k(z_1)], ..., [h_1(z_k), ..., h_k(z_k)]\}$ is linearly independent, where $z_1, ..., z_k$ are distinct points in $\{x_0, ..., x_k\}$. Thus for any nontrivial solution

$(\theta_0\sigma_0, ..., \theta_k\sigma_k)$ of (2.1) we must $\theta_i\sigma_i \neq 0$ for all $i = 0, 1, ..., k$. Now without loss of generality we may choose $\sigma_0 = 1$, and rewrite (2.1) as

$$\sum_{i=0}^{k} \sigma_i(\theta_i/\theta_0)[h_1(x_i), ..., h_k(x_i)] = -[h_1(x_0), ..., h_k(x_0)]. \tag{2.2}$$

The Haar condition then implies that (2.2) has a unique solution $(\sigma_1(\theta_1/\theta_0), ..., \sigma_k(\theta_k/\theta_0) = (\alpha_1, ...., \alpha_k)$, where $\alpha_i \neq 0$, $i = 1, ..., k$. Now imposing the conditions $|\sigma_i| = 1$, $\theta_i > 0$, for $i = 0, ..., k$, and $\sum_{i=0}^{k} \theta_i = 1$ determines $\sigma_1, ..., \sigma_k$ and $\theta_0, ..., \theta_k$ uniquely. To see that $\sigma_0, ..., \sigma_k$ are independent of the choice of basis for $W(a^*)$, we observe for any $h$ in any other basis for $W(a^*)$ we must have $\sum_{i=0}^{k} \theta_i\sigma_i h(x_i) = 0$ so (2.1) holds for the $\sigma_0, ..., \sigma_k, \theta_0, ..., \theta_k$ chosen above.

(d) Without loss of generality we may assume $h_i = F_i(a^*, \cdot)$, $i = 1, ..., N$, forms a basis for $W(a^*)$. Let $\bar{h}_i = F_i(a, \cdot)$, $i = 1, ..., N$. Now, $\bar{h}_i(y_j)$ depends continuously on $a$ and $y_j$, so for $\delta > 0$ sufficiently small we have that the determinant of the matrix $[\bar{h}_i(z_j)]$, $i = 1, ..., N$, $j = 1, ..., N$, is nonzero for every choice of $N$ distinct points $\{z_1, ..., z_N\}$ from $\{y_0, ..., y_N\}$. Thus $\{\bar{h}_1, ..., \bar{h}_N\}$ is linearly independent, so $d(a) \geqslant N$; but $N$ is maximal, so $d(a) = N$, and $\{\bar{h}_1, ..., \bar{h}_N\}$ is a basis for $W(a)$. Also, $W(a)$ satisfies the Haar condition on $\{y_0, ..., y_N\}$ so LH$(a, \{y_0, ..., y_N\})$ holds. Finally, by part (c) we can infer that $\alpha_1, ..., \alpha_N$ are continuous functions of $x_0, ..., x_N$, $a^*$, so small changes in $x_0, ..., x_N$, $a^*$ will leave the signs $\sigma_0, ..., \sigma_N$ associated with $x_0, ..., x_N$ unchanged. Thus if $\delta > 0$ is sufficiently small, $\sigma_0, ..., \sigma_N$ will also be the signs associated with $y_0, ..., y_N$.

(e) Suppose $\sigma_i(F(a, x_i) - F(a^*, x_i)) > 0$ for $i = 0, ..., k$. (The case where $\sigma_i(F(a, x_i) - F(a^*, x_i)) < 0$ for $i = 0, ..., k$ is similar and will be omitted.) Let $\{h_1, ..., h_k\}$ be a basis for $W(a^*)$. By SIGN$(a^*, \{x_0, ..., x_k\})$, there is a $c \in \mathbf{R}^k$ such that $\text{sgn}(\sum_{j=1}^{k} c_j h_j(x_i)) = \text{sgn}(F(a, x_i) - F(a^*, x_i))$ for $i = 0, 1, ..., k$. Thus, setting $p = \sum_{j=1}^{k} c_j h_j$, we have $\sigma_i p(x_i) > 0$ for $i = 0, ..., k$. We wish to show this is impossible; we will establish the stronger claim that if $\sigma_i p(x_i) \geqslant 0$, for $i = 0, ..., k$, then $p \equiv 0$ on $X$. From part (c) above and its proof we have that $\sum_{i=1}^{k} \theta_i\sigma_i [h_1(x_i), ..., h_k(x_i)] = [0, ..., 0]$, for some $\theta_0, ..., \theta_k$ with $\theta_i > 0$ for all $i$. So $\sum_{i=1}^{k} \theta_i\sigma_i p(x_i) = 0$. But $\theta_i\sigma_i p(x_i) \geqslant 0$, $0 \leqslant i \leqslant k$, hence $p(x_i) = 0$, $0 \leqslant i \leqslant k$, and the claim now follows from the assumption that $W(a^*)$ satisfies the Haar condition on $\{x_0, ..., x_k\}$.

(g) ($\Rightarrow$) Suppose $F(a^*, \cdot)$ is a best approximation to $f$ from $V$. If $\|f - F(a^*, \cdot)\| = 0$, then $E_{a^*}(f) = X$ and for any $(x_0, ..., x_k) \in \hat{X}_k$ with signs $\sigma_0, ..., \sigma_k$ we have $|f(x_i) - F(a^*, x_i)| = 0 = 0\sigma_i$, $i = 0, ..., k$. If $\|f - F(a^*, \cdot)\| > 0$, then by part (b) above and Caratheodory's theorem [4, p. 17], for some $m \leqslant k$, there exist $(x_0, ..., x_k) \in \hat{X}_k$ with $|f(x_i) - F(a^*, x_i)| = \|f - F(a^*, \cdot)\|$, $i = 0, ..., m$, and the zero of $\mathbf{R}^k$ is in co$\{\text{sgn}(f(x_i) -$

$F(a^*, x_i))[h_1(x_i), ..., h_k(x_i)]: i=0, ..., m\}$ where $\{h_1, ..., h_k\}$ is a basis for $W(a^*)$. But by the Haar condition, $m \geqslant k$, so $m=k$. By the uniqueness of the signs in part (c) we have $\operatorname{sgn}(f(x_0) - F(a^*, x_0)) \operatorname{sgn}(f(x_i) - F(a^*, x_i)) = \sigma_i$, so $\operatorname{sgn}(f(x_i) - F(a^*, x_i)) = \operatorname{sgn}(f(x_0) - F(a^*, x_0))\sigma_i \equiv \sigma\sigma_i$, for $i=0, ..., k$.

($\Leftarrow$) Suppose that for some $(x_0, ..., x_k) \in \hat{X}_k$, LH($a^*$, $\{x_0, ..., x_k\}$) and SIGN($a^*$, $\{x_0, ..., x_k\}$) hold, $|f(x_i) - F(a^*, x_i)| = \|f - F(a^*, \cdot)\|$, and $\operatorname{sgn}(f(x_i) - F(a^*, x_i)) = \sigma\sigma_i$, for $i=0, ..., k$, for some $\sigma = -1, 0$, or 1. Then by part (f) we have

$$\begin{aligned} \|f - F(a^*, \cdot)\| &\geqslant \inf\{\|f - F(a, \cdot)\| : a \in A\} \\ &\geqslant \min\{|f(x_i) - F(a^*, x_i)| : 0 < i < k\} \\ &= \|f - F(a, \cdot)\|, \end{aligned}$$

so $F(a^*, \cdot)$ is a best approximation to $f$ from $V$.

*Remark.* We note that if $\sigma_V(x_0, ..., x_k; a^*)$ is a $\sigma$-alternant for $f - F(a^*, \cdot)$, and if we define $M_1 = \{x_i: \sigma_i = 1\}$ and $M_2 = \{x_i: \sigma_i = -1\}$, then in the terminology of [6, 19] $M = M_1 \cup M_2$ is called an $H$-set relative to $F(a^*, \cdot)$, since by Theorem 1(e), there is no $F(a, \cdot) \in V$ with $F(a^*, \cdot) - F(a, \cdot) > 0$ on $M_1$, and $F(a^*, \cdot) - F(a, \cdot) < 0$ on $M_2$. In fact, in the terminology of [9], $\sigma_V(x_0, ..., x_k; a^*)$ is a minimal $H$-set relative to $F(a^*, \cdot)$.

As stated before Theorem 1(g) is a generalized alternation theorem, but we note there that the ordinary alternation theorem does not necessarily hold even in situations where it would appear to make sense. The following example illustrates this.

EXAMPLE 1. Let $X = \{-1, 0, \frac{1}{2}\}$, $A = \mathbf{R}^2$, and $V = \{a_1 + a_2 x^2: (a_1, a_2) \in A\}$. For every $a \in A$ we have $W(a) = V$ and $d(a) = 2$. Since $V$ is a linear space we have that property SIGN($a$, $X$) holds for all $a \in A$. Note also that LH($a$, $X$) and Z($a$, $X$) also hold for all $a \in A$. Thus every $f \in C(X)$ is normal. Define $f \in C(X)$ by $f(-1) = 0$, $f(0) = 0$, and $f(\frac{1}{2}) = -\frac{3}{4}$. Consider $a = (-\frac{1}{2}, 1)$, so $F(a, x) = x^2 - \frac{1}{2}$. We have then that $\|f - F(a, \cdot)\| = \frac{1}{2}$ with $f - F(a, \cdot)$ having the ordinary alternation property, but $F(a, \cdot)$ is not a best approximation to $f$. To see this, consider $a^* = (-\frac{3}{8}, 0)$, so $F(a^*, x) = -\frac{3}{8}$; we have $\|f - F(a^*, \cdot)\| = \frac{3}{8}$, with $f(-1) - F(a^*, -1) = \frac{3}{8}$, $f(0) - F(a^*, 0) = \frac{3}{8}$, and $f(\frac{1}{2}) - F(a^*, \frac{1}{2}) = -\frac{3}{8}$. Now we have $(\frac{1}{8})(1)[1, (-1)^2] + (\frac{3}{8})(1)[1, 0^2] + (\frac{1}{2})(-1)[1, (\frac{1}{2})^2] = [0, 0]$, so $(\sigma_0, \sigma_1, \sigma_2) = (1, 1, -1)$, and so $f - F(a, \cdot)$ alternates in the sense described in Theorem 1(g); thus $F(a^*, \cdot)$ is the best approximation to $f$. Note that if we extend this example to $[-1, 1]$ by defining $f$ to pass through $(-1, 0)$,

$(-\frac{1}{2}, -\frac{3}{8})$, $(0, 0)$, $(\frac{1}{2}, -\frac{3}{4})$, and $(1, -\frac{3}{8})$, and to be linear between these points, then $f$ is weakly normal (but not normal), and $F(a^*, x) = -\frac{3}{8}$ is still the best approximation to $f$ according to Theorem 1(h) with the same $\sigma$-alternant. Thus although the concepts of weak normality and $\sigma$-alternant are the same as ordinary normality and alternant in some common situations (e.g., ordinary rationals or exponential sum approximation on a compact subset of an interval with at least $N+1$ points), in more complicated situations they can add additional insight.

The next example demonstrates that in contrast to the set of normal functions the set of weakly normal functions need not be an open set. It also gives an example of a nonweakly normal function that has a strongly unique best approximation.

EXAMPLE 2. Let $X = [-1, 1]$, $A = \mathbf{R}^2$, and $V = \{a_1 + a_2 x^2 : (a_1, a_2) \in A\}$. Then as before $W(a) = V$ and $d(a) = 2$, for all $a \in A$, but since properties LH$(a, X)$ and Z$(a, X)$ both fail, there are no normal functions in $C(X)$ with respect to $V$. Now define $f \in C(X)$ by $f(-1) = 1$, $f(-\frac{1}{2}) = -1$, $f(0) = 1$, $f(1) = 0$, and linear in between these points. Then $a^* = (0, 0)$ gives $F(a^*, x) = 0$ as the best approximation with $\{-1, -\frac{1}{2}, 0\}$ with the signs $\sigma_0 = 1$, $\sigma_1 = -1$, $\sigma_2 = 1$ forming a $\sigma$-alternant. Note also that properties SIGN, LH, and Z all hold at $a^*$ on $\{-1, -\frac{1}{2}, 0\}$, so $f$ is weakly normal and by Theorem 1(h), $F(a^*, \cdot) = 0$ is strongly unique. Now for $0 < t < \frac{1}{2}$ define $g_t \in C(X)$ by $g_t(-1) = 1$, $g_t(-\frac{1}{2}) = -1$, $g_t(-t) = 1$, $g_t(0) = 1 - t$, $g_t(t) = 1$, $g_t(1) = 0$, and linear in between these points. Then $a^* = (0, 0)$ still gives the best approximation to $g_t$ and $\{-1, -\frac{1}{2}, -t\}$ with signs $\sigma_0 = 1$, $\sigma_1 = -1$, and $\sigma_2 = 1$ forms a $\sigma$-alternant, but properties LH and Z fail to hold at $a^*$ on the set $E_{a^*}(g_t) = \{-1, -\frac{1}{2}, -t, t\}$. So $g_t$ is not weakly normal even though $g_t \to f$ uniformly as $t \to 0$. However, we do have that for $t$ sufficiently small

$$\| g_t - F(a, \cdot)\| \leqslant \| g_t \| + ((1 - 4t^2)/(7 - 4t^2)) \| F(a, \cdot)\| \qquad \text{for all} \quad a \in A.$$

Hence $g_t$ has zero as its strongly unique best approximation.

The preceding example illustrates an important fact about the weakly normal functions. That is, if $f$ is weakly normal and $g$ is sufficiently close to $f$, then $g$ must have a strongly unique best approximation, even if $g$ is not weakly normal. This is made precise in the following lemma.

LEMMA 2. *Suppose $f \in C(X)$ is weakly normal and $\|e_V(f)\| \neq 0$. Then there exists a $\delta_0 > 0$ such that if $g \in C(X)$ and $\|f - g\| \leqslant \delta_0$, then $g$ has a strongly unique best approximation $F(a, \cdot)$, and $g - F(a, \cdot)$ possesses a*

*$\sigma$-alternant of length $N+1$. In addition, $\lambda_V(f, \delta_0) < +\infty$. Moreover, this strong uniqueness holds uniformly in that there exist a $\delta$, $0 < \delta \leqslant \delta_0$, and a $\gamma > 0$ such that $\|g - F(a, \cdot)\| \geqslant \|g - B_V(g)\| + \gamma \|F(a, \cdot) - B_V(g)\|$ for all $g$ with $\|f - g\| \leqslant \delta$ and for all $a \in A$.*

*Proof.* If $F(a^*, \cdot)$ is a best approximation to $f$, then $\mathrm{LH}(a^*, E_{a^*}(f))$ and $\mathrm{SIGN}(a^*, E_{a^*}(f))$ hold, so by Theorem 1(g) $f - F(a^*, \cdot)$ possesses a $\sigma$-alternant $\sigma_V(x_0, ..., x_N : a^*)$, and by Theorem 1(h), $F(a^*, \cdot)$ is strongly unique. Then by arguments similar to those in [3], there exist $\delta_1 > 0$ and $\beta > 0$ such that if $g \in C(X)$ and $\|f - g\| \leqslant \delta_1$, then a best approximation $F(a, \cdot)$ to $g$ exists, and for any such $F(a, \cdot)$ we have $\|F(a^*, \cdot) - F(a, \cdot)\| \leqslant \beta \|f - g\|$. We now claim that for some $\delta_0$ with $0 < \delta_0 \leqslant \delta_1$, if $\|f - g\| \leqslant \delta_0$ and $F(a, \cdot)$ is a best approximation to $g$, then $g - F(a, \cdot)$ possesses a $\sigma$-alternant of length $N+1$. Once this has been shown, strong uniqueness of $F(a, \cdot)$ will follow from Theorem 1(h), and we will also have $\lambda_V(f, \delta_0) \leqslant \beta < +\infty$.

To prove the claim, suppose that there were a sequence $\{g_m\} \subset C(X)$ with $\|g_m - f\| \to 0$, and $\{a^m\} \subset A$ with $F(a^m, \cdot)$ a best approximation to $g_m$, such that $g_m - F(a^m, \cdot)$ possesses no $\sigma$-alternant of length $N+1$. Note that $\|F(a^m, \cdot) - F(a^*, \cdot)\| \leqslant \beta \|f - g_m\|$ for $m$ sufficiently large, so $\|F(a^m, \cdot) - F(a^*, \cdot)\| \to 0$. Thus by Lemma 1(c) we can assume $\|a^m - a^*\| \to 0$. Without loss of generality, suppose $\{h_1, ..., h_N\} \equiv \{F_1(a^*, \cdot), ..., F_N(a^*, \cdot)\}$ is a basis for $W(a^*)$. Then by arguments in the proof of Theorem 1(d), for $m$ suffciently large we have $d(a^m) = N$ and $\{h_{1m}, ..., h_{Nm}\} \equiv \{F_1(a^m, \cdot), ..., F_N(a^m, \cdot)\}$ is a basis for $W(a^m)$. Now by Theorem 1(b) and Carathedory's theorem [4, p. 17], for each $m$ there is a number $k \leqslant N$ (which by going to subsequences if necessary, we can assume to be fixed), numbers $\theta_{m0}, ..., \theta_{mk}$ with $\theta_{mi} > 0$, $0 \leqslant i \leqslant k$ and $\sum_{i=0}^{k} \theta_{mi} = 1$, and points $y_{m0}, ..., y_{mk} \in E_a(g_m)$ with

$$\sum_{i=0}^{k} \theta_{mi} \operatorname{sgn}(g_m(y_{mi}) - F(a^m, y_{mi}))[h_{1m}(y_{mi}), ..., h_{Nm}(y_{mi})] = 0_k.$$

Going to subsequences if necessary, we can assume $y_{mi} \to y_i \in X$, $0 \leqslant i \leqslant k$, $\theta_{mi} \to \theta_i \geqslant 0$, $0 \leqslant i \leqslant k$, with $\sum_{i=0}^{k} \theta_i = 1$, and $\operatorname{sgn}(g_m(y_{mi}) - F(a^m, y_{mi})) \to \bar{\sigma}_i$ for $0 \leqslant i \leqslant k$; since $\|g_m - F(a^m, \cdot)\| \to \|f - F(a^*, \cdot)\| \neq 0$ we have $|\bar{\sigma}_i| = 1$ for $0 \leqslant i \leqslant k$. We also have

$$\sum_{i=0}^{k} \theta_i \bar{\sigma}_i [h_1(y_i), ..., h_N(y_i)] = 0_k \tag{2.3}$$

and $\{y_0, ..., y_k\} \subset E_{a^*}(f)$ since for $0 \leqslant i \leqslant k$ we have

$$\begin{aligned}
&|f(y_i)-F(a^*, y_i)| \\
&\quad \geqslant |g_m(y_{mi})-F(a^m, y_{mi})| - |f(y_i)-f(y_{mi})| - |f(y_{mi})-g_m(y_{mi})| \\
&\qquad - |F(a^m, y_{mi})-F(a^*, y_{mi})| - |F(a^*, y_{mi})-F(a^*, y_i)| \\
&\quad \geqslant \|g_m - F(a^m, \cdot)\| - |f(y_i)-f(y_{mi})| - \|f-g_m\| \\
&\qquad - \|F(a^m, \cdot)-F(a^*, \cdot)\| - |F(a^*, y_{mi})-F(a^*, y_i)| \\
&\quad \to \|f-F(a^*, \cdot)\|
\end{aligned}$$

as $m \to +\infty$. Now if $y_i = y_j$ for some $i \neq j$, then $y_i = y_j = x$ for some $x \in E_{a^*}(f)$. Thus $\bar{\sigma}_i = \lim_{m\to\infty} \operatorname{sgn}(g_m(y_{mi}) - F(a^m, y_{1i})) = \operatorname{sgn}(f(x) - F(a^*, x)) = \lim_{m\to\infty} \operatorname{sgn}(g_m(y_{mj}) - F(a^m, y_{mj})) = \bar{\sigma}_j$. So $\theta_i \bar{\sigma}_i [h_1(y_i), ..., h_N(y_i)] + \theta_j \bar{\sigma}_j [h_1(y_j), ..., h_N(y_i)] = (\theta_i + \theta_j) \bar{\sigma}_i [h_1(y_i), ..., h_N(y_i)]$, and $\theta_i + \theta_j > 0$ if either $\theta_i > 0$ or $\theta_j > 0$. Now coalescing terms in (2.3) as above and deleting any zero terms, we have that a nonempty set of vectors of the form $[h_1(x_i), ..., h_N(x_i)]$ is linearly dependent, where the $x_i$'s are distinct element of $E_{a^*}(f)$. But our weak normality assumption implies $W(a^*)$ satisfies the Haar condition on $E_{a^*}(f)$; thus there must be at least $N+1$ of these vectors. Thus we must have $k = N$, $\theta_i > 0$, $0 \leqslant i \leqslant N$, and $y_i \neq y_j$ for $0 \leqslant i, j \leqslant N$, $i \neq j$. By Theorem 1(c), $\sigma_V(y_0, ..., y_N; a^*)$ exists with associated signs $\sigma_0, ..., \sigma_N$ and by Theorem 1(d), for $m$ sufficiently large $\sigma_V(y_{m0}, ..., y_{mN}; a^m)$ exists and has signs $\sigma_0, ..., \sigma_N$. Also we have $0_N \in \operatorname{co}\{\operatorname{sgn}(g_m(y_{mi}) - F(a^m, y_{mi}))[h_{1m}(y_{mi}), ..., h_{Nm}(y_{mi})]\colon i = 0, ..., N\}$. Thus, for $m$ sufficiently large, the uniqueness in Theorem 1(c) gives $\operatorname{sgn}(g_m(y_{mi}) - F(a^m, y_{mi})) = \sigma^m \sigma_i$ for $0 \leqslant i \leqslant N$, where $\sigma^m = \operatorname{sgn}(g_m(y_{m0}) - F(a^m, y_{m0}))$. Thus $\sigma_V(y_{m0}, ..., y_{mN}; a^m)$ is a $\sigma$-alternant for $g_m - F(a^m, \cdot)$, and this is a contradiction. Thus, the claim is established.

To prove the strong unicity part we assume that $F(a, \cdot) \equiv B_V(g)$. Suppose the result is false. Then there is a sequence $\{g_m\} \subset C(X)$ with $\|g_m - f\| \to 0$, $B_V(g_m) = F(a^m, \cdot)$, and there is a sequence $\{b^m\} \subset A$ with

$$\gamma_m \equiv \frac{\|g_m - F(b^m, \cdot)\| - \|g_m - F(a^m, \cdot)\|}{\|F(b^m, \cdot) - F(a^m, \cdot)\|} \to 0 \qquad \text{as} \quad m \to +\infty.$$

Since $\gamma_m \to 0$ we must have that $\{\|F(b^m, \cdot)\|\}$ is bounded and $\|g_m - F(b^m, \cdot)\| - \|g_m - F(a^m, \cdot)\| \to 0$. Now $\|g_m - F(a^m, \cdot)\| \to \|f - F(a^*, \cdot)\|$, so $\|g_m - F(b^m, \cdot)\| \to \|f - F(a^*, \cdot)\|$. So, using Lemma 1(c) and the arguments in [3, Theorem 2] we can assume $\|a^m - a^*\| \to 0$ and $\|b^m - a^*\| \to 0$. We can also assume by using the arguments in Theorem 1(d) that $\{F_1(a^*, \cdot), ..., F_N(a^*, \cdot)\}$ is a basis for $W(a^*)$, $\{F_1(a^m, \cdot), ..., F_N(a^m, \cdot)\}$ is a basis for $W(a^m)$, and $\{F_1(b^m, \cdot), ..., F_N(b^m, \cdot)\}$ is a basis for $W(b^m)$; also $a_j^m = b_j^m = a_j^*$ for $j > N$. Now by the first part, we have that for all $m$ sufficiently large there is a $\sigma$-alternant $\sigma_V(x_{m0}, ..., x_{nN}; a^m)$ for $h_m - F(a^m, \cdot)$;

going to subsequences if necessary, we can assume $x_{mi} \to x_i \in X$, $0 \leqslant i \leqslant N$, and from arguments of the first part of this proof, it follows that $\{x_0, ..., x_N\}$ will be the points of a $\sigma$-alternant for $f - F(a^*, \cdot)$. By Theorem 1(d), for $m$ sufficiently large we can assume that $\sigma_V(x_{m0}, ..., x_{mN}; a^m)$ and $\sigma_V(x_0, ..., x_N; a^*)$ have the same signs $\sigma_0, ..., \sigma_N$. Further let $\sigma$ be such that $\operatorname{sgn}(f(x_i) - F(a^*, x_i)) = \sigma\sigma_i$, for $0 \leqslant i \leqslant N$. Then for $m$ sufficiently large we have $\operatorname{sgn}(g_m(x_{mi}) - F(a^m, x_{mi})) = \sigma\sigma_i$ for $0 \leqslant i \leqslant N$. We now claim that there is an $\alpha > 0$ such that for all $m$ sufficiently large,

$$\max\left\{\sigma\sigma_i \frac{F(a^m, x_{mi}) - F(b^m, x_{mi})}{\|a^m - b^m\|} : 0 \leqslant i \leqslant N\right\} \geqslant \alpha.$$

To prove the claim, suppose that, going to subsequences if necessary which we do not relabel, there were positive numbers $\alpha_m \to 0$ with

$$\max\left\{\sigma\sigma_i \frac{F(a^m, x_{mi}) - F(b^m, x_{mi})}{\|a^m - b^m\|} : 0 \leqslant i \leqslant N\right\} \leqslant \alpha_m$$

for all large $m$. Now for $m$ sufficiently large, the mean value theorem implies that $F(a^m, x_{mi}) - F(b^m, x_{mi}) = D(a^*, a^m, x_i) + o(\|a^m - b^m\|)$. Thus we have that

$$\max\left\{\sigma\sigma_i \left[\frac{D(a^*, a^m - b^m, x_{mi})}{\|a^m - b^m\|} + \frac{o(\|a^m - b^m\|)}{\|a^m - b^m\|}\right] : 0 \leqslant i \leqslant N\right\} \leqslant \alpha_m.$$

Going to subsequences if necessary, we can assume $(a_j^m - b_j^m)/(\|a^m - b^m\|) \to c_j$, $0 \leqslant j \leqslant N$, where $\max\{|c_j| : 0 \leqslant j \leqslant N\} = 1$. Thus defining $c_j = 0$ for $j > N$ we have $\max\{\sigma\sigma_i D(a^*, c, x_i) : 0 \leqslant i \leqslant N\} \leqslant 0$. But by the claim in the proof of Theorem 1(e), we have $D(a^*, c, \cdot) = 0$, which is a contradiction. Thus the claim is established. Again, going to subsequences of necessary, we now have for $m$ sufficiently large and for some $i$ with $0 \leqslant i \leqslant N$,

$$\begin{aligned}
\|g_m - F(b^m, \cdot)\| &\geqslant \sigma\sigma_i(g_m(x_{mi}) - F(b^m, x_{mi})) \\
&= \sigma\sigma_i(g_m(x_{mi}) - F(a^m, x_{mi})) \\
&\quad + \sigma\sigma_i \frac{F(a^m, x_{mi}) - F(b^m, x_{mi})}{\|a^m - b^m\|} \|a^m - b^m\| \\
&\geqslant \|g_m - F(a^m, \cdot)\| + \alpha \|a^m - b^m\|.
\end{aligned}$$

So using Lemma 1(b), there is a constant $L > 0$ such that

$$\|g_m - F(b^m, \cdot)\| \geqslant \|g_m - F(a^m, \cdot)\| + \alpha L \|F(b^m, \cdot) - F(a^m, \cdot)\|.$$

Thus $\gamma_m \geqslant \alpha L$, which contradicts the fact that $\gamma_m \to 0$.

## 3. Linear Theory

We now consider the situation where $V$ is a linear subspace. We take $A=\mathbf{R}^N, h_1, h_2, ..., h_N \in C(X)$ to be linearly independent and set $V^*=\{F(a,\cdot)=a_1h_1+\cdots+a_Nh_N\}$. We have that for all $a\in\mathbf{R}^N$ and all $S\subset X$, $d(a)=N$, $W(a)=V^*$, and SIGN$(a, S)$ holds; further LH$(a, S)$ and Z$(a, S)$, when they hold, are independent of $a$, and are equivalent if $|S|\geqslant N$. Thus $f\in C(X)$ will be weakly normal if and only if it has a best approximation $p^*\in V^*$ with $V^*$ satisfying the Haar condition on $E_{V^*}(f)$. If $V^*$ does satisfy the Haar condition on $(x_0, ..., x_N)\in\hat{X}_N$, then $\sigma_{V^*}(x_0, ..., x_N; a)$ exists and is independent of $a$, and will be denoted by $\sigma_{V^*}(x_0, ..., x_N)$.

Given $X_{V^*}=(x_0, ..., x_N)\in\hat{X}_N$ with $V^*$ satisfying the Haar condition on $\{x_0, ..., x_N\}$ and $\sigma_{V^*}(x_0, ..., x_N)$ having signs $\sigma_0, ..., \sigma_N$, we define the generalized polynomials $q_i\in V^*$ by

$$q_i(x_j)=\sigma_j, \qquad j\neq i, j=0, ..., N, i=0, ..., N. \tag{3.1}$$

The proofs of the following two lemmas are similar to the proofs given in [1, Lemma 1] and so will be omitted.

Lemma 3. *Suppose $f\in C(X)$ and $X_{V^*}=(x_0, ..., x_N)\in\hat{X}_N$ with $V^*$ satisfying the Haar condition on $\{x_0, ..., x_N\}$. Let $\sigma_{V^*}(x_0, ..., x_N)$ have signs $\sigma_0, ..., \sigma_N$ and define $q_i$ for $i=0, ..., N$ as in* (3.1). *Then*

$$\text{(i)}\quad B_{V^*}(f, X_{V^*})=\sum_{j=0}^{n}\frac{-\sigma_j f(x_j)}{1+|q_j(x_j)|}q_j,$$

$$\text{(ii)}\quad \text{if } p\in V^*, \qquad \text{then}\quad p=\sum_{j=0}^{n}\frac{-\sigma_j p(x_j)}{1+|q_j(x_j)|}q_j.$$

Lemma 4. *Suppose $V^*$ satisfies the Haar condition on $X_{V^*}=\{x_0, ..., x_N\}$, where $(x_0, ..., x_N)\in\hat{X}_N$. Then the generalized polynomials $q_i$ defined by* (3.1) *satisfy*

$$\text{(i)}\quad \sum_{j=0}^{N}\frac{1}{1+|q_j(x_j)|}=1.$$

$$\text{(ii)}\quad \sum_{j=0}^{N}\frac{q_j}{1+|q_j(x_j)|}\equiv 0.$$

The next theorem gives an explicit form for the local Lipschitz constant, $\hat{\lambda}_{V^*}(f)$, and as a consequence, shows that $B_{V^*}$ is Gâteaux differentiable at $f$, for all $f\in C(X)$ that are weakly normal, and such that $|E_{V^*}(f)|=N+1$.

THEOREM 2. *Suppose* $f \in C(X)$ *is weakly normal,* $\|e_{V^*}(f)\| \neq 0$ *and* $|E_{V^*}(f)| = N+1$. *Let* $E_{V^*}(f) = \{x_0, ..., x_N\}$. *Then*

$$\hat{\lambda}_{V^*}(f) = \lim_{\delta \to 0} \lambda_{V^*}(f, \delta) = \left\| \sum_{j=0}^{N} \frac{|q_j|}{1 + q_j(x_j)|} \right\| \equiv \Phi_{V^*}(X, E_{V^*}(f)).$$

*Proof.* The proof that $\hat{\lambda}_{V^*}(f) \leqslant \Phi_{V^*}(X, E_{V^*}(f))$ is similar to that in [1, Theorem 2] and will be omitted, except to note that Markoff's inequality in that proof is replaced by the following fact: There is a function $\psi: \mathbf{R}^+ \to \mathbf{R}^+$ with $\psi(\delta) \to 0$ as $\delta \to 0^+$ with $\omega(p, \delta) \leqslant \|p\| \, \psi(\delta)$, for all $p \in V^*$, where $\omega(p, \delta)$ is the modulus of continuity of $p$. To see this we note $S = \{q \in V^*: \|q\| = 1\}$ is compact, and is thus equicontinuous by the Arzela–Ascoli theorem; thus $\psi(\delta) = \sup\{|q(x_1) - q(x_2)|: q \in S, x_1, x_2 \in X, |x_1 - x_2| \leqslant \delta\} \to 0$ as $\delta \to 0^+$. Now for any $p \in V^*$, $p \not\equiv 0$, we have $\omega(p/\|p\|, \delta) \leqslant \psi(\delta)$, so $\omega(p, \delta) \leqslant \|p\| \, \psi(\delta)$. The proof that $\hat{\lambda}_{V^*}(f) \geqslant \Phi_{V^*}(X, E_{V^*}(f))$ can be accomplished by selecting $\bar{g} \in C(X)$ with $\|\bar{g}\| \neq 0$ and proving, using the inequalities developed in the first part of the proof with $g = f + t\bar{g}$, that $\lim_{t \to 0}(B_{V^*}(f + t\bar{g}) - B_{V^*}(f))/t$ exists and equals $\sum_{j=0}^{N} (-\sigma_j \bar{g}(x_j)/(1 + |q_j(x_j)|))q_j$; this can then be used to show that

$$\hat{\lambda}_{V^*}(f) \geqslant \frac{1}{\|\bar{g}\|} \left\| \sum_{j=0}^{N} \frac{-\sigma_j \bar{g}(x_j)}{1 + |q_j(x_j)|} q_j \right\|$$

for all $\bar{g} \in C(X)$ with $\|\bar{g}\| \neq 0$, and this implies $\hat{\lambda}_{V^*}(f) \geqslant \Phi_{V^*}(X, E_{V^*}(f))$.

The proof of Theorem 2 gives us then the following which is merely a generalization of [13].

COROLLARY. *Suppose* $f \in C(X)$ *is weakly normal,* $\|e_{V^*}(f)\| \neq 0$, *and* $|E_{V^*}(f)| = N+1$. *Let* $E_{V^*}(f) = \{x_0, ..., x_N\}$. *Then* $B_{V^*}$ *is Gâteaux differentiable at f for all* $g \in C(X)$. *Moreover,*

$$D_f B_{V^*}(g) = \sum_{j=0}^{N} \frac{-\sigma_j g(x_j)}{1 + |q_j(x_j)|} q_j.$$

*Remark.* The number $\Phi_{V^*}(X, E_{V^*}(f))$ depends not explicitly on $f$ but only on the set of extreme points of $f - B_{V^*}(f)$. Thus if we change $f$ but maintain the same set of extreme points the number $\Phi_{V^*}(X, E_{V^*}(f))$ will be the same.

The final theorem of this section gives a characterization of the strong unicity constant for $f \in C(X)$ when $f$ is weakly normal and $|E_{V^*}(f)| = N+1$. We omit the proof since it follows from the arguments of [5, Theorem 5] and the fact that $|E_{V^*}(f)| = N+1$. (See also [11].)

THEOREM 3. *Suppose $f \in C(X)$ is weakly normal, $\|e_{V^*}(f)\| \neq 0$, and $|E_{V^*}(f)| = N+1$. Let $E_{V^*}(f) = \{x_0, ..., x_N\}$. Then*

$$\lim_{\delta \to 0^+} M_{V^*}(f, \delta) = M_{V^*}(f) = \max\{\|q_i\| : 0 \leqslant i \leqslant N\}.$$

## 4. MAIN RESULTS

We are now in a position to prove our main results which extend Theorem 2, its corollary, and Theorem 3 to the nonlinear setting; the results turn out to be the same as in the linear case, with $V^* = W(a^*)$.

THEOREM 4. *Suppose $f \in C(X)$ is weakly normal with best approximation $F(a^*, \cdot) \in V$, $\|e_V(f)\| \neq 0$, and $|E_V(f)| = N+1$. Let $E_V(f) = \{x_0, ..., x_N\}$ and $V^* = W(a^*)$. Then*

$$\hat{\lambda}_V(f) = \lim_{\delta \to 0} \lambda_V(f, \delta) = \Phi_{V^*}(X, E_V(f))$$
$$= \left\| \sum_{j=0}^{N} \frac{|q_j|}{1 + |q_j(x_j)|} \right\|,$$

*where $q_j$, $j = 0, ..., N$, are the generalized polynomials in $W(a^*)$ satisfying (3.1).*

*Proof.* Let $\delta_0$ be as in Lemma 2, and let $\beta = \lambda_V(f, \delta_0) < \infty$. Now suppose $0 < \delta \leqslant \delta_0$; for $\delta$ sufficiently small and $0 < \|f - g\| \leqslant \delta$ we have from Lemma 2 that $g$ has a strongly unique best approximation $F(a, \cdot)$, with $g - F(a, \cdot)$ having a $\sigma$-alternant $\sigma_V(y_0, ..., y_N; a)$. Now using arguments like those in the proof of the first part of Lemma 2 or by a generalization of [1, Lemma 4] we can assume that $\max\{r(x_i, y_i)\colon 0 \leqslant i \leqslant N\}$ is as small as we please. We have $\|F(a, \cdot) - F(a^*, \cdot)\| \leqslant \beta \|f - g\|$ so by Lemma 1(c) we can assume that $\|a - a^*\| = O(\|F(a, \cdot) - F(a^*, \cdot)\|) = O(\|f - g\|)$. By our definitions and Theorem 1(d) we can assume that $\sigma_V(x_0, ..., x_N; a^*)$, $\sigma_V(y_0, ..., y_N; a)$, $\sigma_{V^*}(x_0, ..., x_N)$, and $\sigma_{V^*}(y_0, ..., y_N)$ all have the same signs $\sigma_0, ..., \sigma_N$. Now from Lemma 1(a) we have

$$\|F(a, \cdot) - F(a^*, \cdot) - D(a^*, a - a^*, \cdot)\| = o(\|a - a^*\|) = o(\|f - g\|). \quad (4.1)$$

Let $\bar{f} = f - F(a^*, \cdot)$ and $\bar{g} = g - F(a^*, \cdot)$; we now claim that $\|B_{V^*}(\bar{g}) - B_{V^*}(\bar{f}) - D(a^*, a - a^*, \cdot)\| = o(\|f - g\|)$. Once this has been shown, applying (4.1) we will have

$$\|F(a, \cdot) - F(a^*, \cdot)\| - \|B_{V^*}(\bar{g}) - B_{V^*}(\bar{f})\|$$
$$\leqslant \|F(a, \cdot) - F(a^*, \cdot) - (B_{V^*}(\bar{g}) - B_{V^*}(\bar{f}))\|$$
$$\leqslant \|F(a, \cdot) - F(a^*, \cdot) - D(a^*, a - a^*, \cdot)\|$$
$$+ \|B_{V^*}(\bar{g}) - B_{V^*}(\bar{f}) - D(a^*, a - a^*, \cdot)\| = o(\|f - g\|).$$

Now $\|f-g\| = \|\bar{f}-\bar{g}\|$ and we have

$$\frac{\|F(a,\cdot)-F(a^*,\cdot)\|}{\|f-g\|} \leqslant \frac{\|B_{V^*}(\bar{f})-B_{V^*}(\bar{g})\|}{\|\bar{f}-\bar{g}\|} + r_1(\|\bar{f}-\bar{g}\|),$$

where $r_1(t)\to 0$ as $t\to 0^+$. Thus we have for $\delta>0$ sufficiently small, since for any $g$ with $0<\|f-g\|\leqslant\delta$ there is a $\bar{g}$ with $0<\|\bar{f}-\bar{g}\|\leqslant\delta$,

$$\sup\left\{\frac{\|F(a,\cdot)-F(a^*,\cdot)\|}{\|f-g\|}: 0<\|f-g\|\leqslant\delta\right\}$$
$$\leqslant \sup\left\{\frac{\|B_{V^*}(\bar{g})-B_{V^*}(\bar{f})\|}{\|\bar{f}-\bar{g}\|} + r_1(\|\bar{f}-\bar{g}\|): 0<\|\bar{f}-\bar{g}\|\leqslant\delta\right\}.$$

So $\lambda_V(f,\delta) \leqslant \lambda_{V^*}(\bar{f},\delta) + \sup\{r_1(\|\bar{f}-\bar{g}\|): 0<\|\bar{f}-\bar{g}\|\leqslant\delta\}$. Therefore, $\lim_{\delta\to 0}\lambda_V(f,\delta)\leqslant\lim_{\delta\to 0}\lambda_{V^*}(\bar{f},\delta)$. Similar arguments prove the reverse inequality, and so we have $\lim_{\delta\to 0}\lambda_V(f,\delta)=\lim_{\delta\to 0}\lambda_{V^*}(\bar{f},\delta)$. But by Theorem 1(g) we have $B_{V^*}(\bar{f})=0$ and $E_{V^*}(\bar{f})=E_V(f)$, so Theorem 2 implies that $\lim_{\delta\to 0}\lambda_{V^*}(\bar{f},\delta)=\Phi_{V^*}(X,E_V(f))=\lim_{\delta\to 0}\lambda_V(f,\delta)=\hat{\lambda}_V(f)$. There remains only to prove the claim. First note that (4.1) implies

$$\|\bar{g}-D(a^*,a-a^*,\cdot)\| \leqslant \|g-F(a,\cdot)\| + o(\|f-g\|). \tag{4.2}$$

Now for $i=0,\ldots,N$,

$$|\bar{g}(y_i)-D(a^*,a-a^*,y_i)|$$
$$\geqslant |g(y_i)-F(a,y_i)| - \|F(a,\cdot)-F(a^*,\cdot)-D(a^*,a-a^*,\cdot)\|$$
$$\geqslant \|g-F(a,\cdot)\| - o(\|f-g\|). \tag{4.3}$$

Since $\operatorname{sgn}(\bar{g}(y_i)-D(a^*,a-a^*,y_i))=\operatorname{sgn}(g(y_i)-F(a,y_i))=\sigma\sigma_i$ for some $\sigma=\pm 1, 0\leqslant i\leqslant N$, and $D(a^*,a-a^*,\cdot)\in V^*$, we have form (4.3) and Theorem 1(f)

$$\|\bar{g}-B_{V^*}(\bar{g})\| \geqslant \|g-F(a,\cdot)\| - o(\|f-g\|). \tag{4.4}$$

Now applying the strong unicity part of Lemma 2 to $V^*$ there is a $\gamma>0$ such that

$$\|B_{V^*}(\bar{g})-D(a^*,a-a^*,\cdot)\| \leqslant (1/\gamma)[\|\bar{g}-D(a^*,a-a^*,\cdot)\| - \|\bar{g}-B_{V^*}(\bar{g})\|].$$

Then (4.2) and (4.4) imply

$$\|B_{V^*}(\bar{g})-D(a^*,a-a^*,\cdot)\|$$
$$\leqslant (1/\gamma)[\|g-F(a,\cdot)\| + o(\|f-g\|) - (\|g-F(a,\cdot)\| - o(\|f-g\|))]$$
$$= o(\|f-g\|),$$

which establishes the claim and the theorem.

The following corollary generalizes a result of [12] where it is shown that in the case of ordinary rational function approximation on a closed interval where $f$ is normal with $|E_V(f)| = N+1$, the $B_V$ is Gâteaux differentiable at $f$. (See also [2].)

COROLLARY. *Suppose that* $f \in C(X)$ *is weakly normal with best approximation* $F(a^*, \cdot) \in V$, $\|e_V(f)\| \neq 0$, *and* $|E_V(f)| = N+1$. *Let* $E_V(f) = \{x_0, ..., x_N\}$, $V^* = W(a^*)$, *and* $\sigma_V(x_0, ..., x_N; a^*)$ *have signs* $\sigma_0, ..., \sigma_N$. *Then* $B_V$ *is Gâteaux differentiable at* $f$ *for all* $g \in C(X)$. *Moreover,*

$$D_f B_V(g) = \sum_{j=0}^{N} \frac{-\sigma_j g(x_j)}{1 + |q_j(x_j)|} q_j, \tag{4.5}$$

*where* $q_j$, $0 \leqslant j \leqslant N$, *are the generalized polynomials in* $W(a^*)$ *satisfying* (3.1). *Thus* $\|D_f B_V\| = \Phi_{V^*}(X, E_V(f))$.

*Proof.* Let $g \in C(X)$. If $g \equiv 0$ then (4.5) holds, so assume $g \not\equiv 0$. For any nonzero $t$, let $g_t = f + tg$, $f = \bar{f} - F(a^*, \cdot)$, and $\bar{g}_t = g_t - F(a^*, \cdot)$. In the proof of Theorem 4, it was shown that (with $g_t$ replacing $g$)

$$\|B_V(g_t) - B_V(f) - (B_{V^*}(\bar{g}_t) - B_{V^*}(\bar{f}))\| = o(\|f - g_t\|).$$

This then implies

$$\left\| \frac{B_V(f+tg) - B_V(f)}{t} - \frac{B_{V^*}(\bar{f}+tg) - B_{V^*}(\bar{f})}{t} \right\| = \frac{o(t\,\|g\|)}{t}$$

which approaches zero as $t \to 0$. As in Theorem 4 we have $E_{V^*}(\bar{f}) = E_V(f)$, but by the corollary to Theorem 2, we have

$$\lim_{t \to 0} \frac{B_{V^*}(\bar{f}+tg) - B_{V^*}(\bar{f})}{t} = \sum_{j=0}^{N} \frac{-\sigma_j g(x_j)}{1 + |q_j(x_j)|} q_j,$$

so (4.5) follows.

The final theorem was proved in [11] for the situation of ordinary rational function approximation on a closed interval with normality in place of weak normality. We note also that a similar theorem has been shown to hold under somewhat different assumptions [8].

THEOREM 5. *Suppose* $f \in C(X)$ *is weakly normal with best approximation* $F(a^*, \cdot) \in V$, $\|e_V(f)\| \neq 0$, *and* $|E_V(f)| = N+1$. *Let* $E_V(f) = \{x_0, ..., x_N\}$, *and* $V^* = W(a^*)$. *Then*

$$\lim_{\delta \to 0} M_V(f, \delta) = \hat{M}(f) = \max\{\|q_i\| : 0 \leqslant i \leqslant N\},$$

*where* $q_i$, $0 \leqslant i \leqslant N$, *are the generalized polynomials in* $W(a^*)$ *satisfying* (3.1).

*Proof.* For $\delta > 0$, consider $F(a, \cdot) \in V$ with $0 < \|F(a, \cdot) - F(a^*, \cdot)\| \leqslant \delta$. Then from Lemma 1(d), for $\delta > 0$ sufficiently small we can assume $\|a - a^*\| = O(\|F(a, \cdot) - F(a^*, \cdot)\|)$. Let $\bar{f} = f - F(a^*, \cdot)$; as in the proof of Theorem 4, we have $B_{V^*}(\bar{f}) \equiv 0$ and $E_{V^*}(\bar{f}) = E_V(f)$. Now by Lemma 1(a) $F(a, \cdot) - F(a^*, \cdot) = D(a^*, a - a^*, \cdot) + o(\|a - a^*\|) = D(a^*, a - a^*, \cdot) + o(\|F(a, \cdot) - F(a^*, \cdot)\|)$, so for $\delta > 0$ sufficiently small we have $D(a^*, a - a^*, \cdot) \not\equiv 0$, so $\|\bar{f} - D(a^*, a - a^*, \cdot)\| - \|\bar{f}\| > 0$. We now claim that for $\delta > 0$ sufficiently small

$$\left| \frac{\|F(a, \cdot) - F(a^*, \cdot)\|}{\|f - F(a, \cdot)\| - \|f - F(a^*, \cdot)\|} - \frac{\|D(a^*, a - a^*, \cdot) - B_{V^*}(f)\|}{\|\bar{f} - D(a^*, a - a^*, \cdot)\| - \|\bar{f} - B_{V^*}(\bar{f})\|} \right| \leqslant u(\delta),$$

where $u(\delta) \to 0$ as $\delta \to 0^+$. To see this, let $\psi_1(a) = \|F(a, \cdot) - F(a^*, \cdot)\| - D(a^*, a - a^*, \cdot)\|$; then $|\psi_1(a)| \leqslant \|F(a, \cdot) - F(a^*, \cdot) - D(a^*, a - a^*, \cdot)\| = o(\|F(a, \cdot) - F(a^*, \cdot)\|)$. So $\sup\{|\psi_1(a)|/\|F(a, \cdot) - F(a^*, \cdot)\| : 0 < \|F(a, \cdot) - F(a^*, \cdot)\| \leqslant \delta\} \to 0$ $\delta \to 0^+$. Likewise, let $\psi_2(a) = \|f - F(a, \cdot)\| - \|f - F(a^*, \cdot)\| - (\|\bar{f} - D(a^*, a - a^*, \cdot)\| - \|\bar{f}\|)$; then $|\psi_2(a)| \leqslant \|F(a^*, \cdot) - F(a, \cdot) + D(a^*, a - a^*, \cdot)\| = o(\|F(a, \cdot) - F(a^*, \cdot)\|)$. So $\sup\{|\psi_2(a)|/\|F(a, \cdot) - F(a^*, \cdot)\| : 0 < \|F(a, \cdot) - F(a^*, \cdot)\| \leqslant \delta\} \to 0$ as $\delta \to 0^+$. We have then that

$$\begin{aligned}
&\left| \frac{\|F(a, \cdot) - F(a^*, \cdot)\|}{\|f - F(a, \cdot)\| - \|f - F(a^*, \cdot)\|} - \frac{\|D(a^*, a - a^*, \cdot) - B_{V^*}(f)\|}{\|\bar{f} - D(a^*, a - a^*, \cdot)\| - \|\bar{f} - B_{V^*}(\bar{f})\|} \right| \\
&= \left| \frac{\|D(a^*, a - a^*, \cdot)\| + \psi_1(a)}{\|\bar{f} - D(a^*, a - a^*, \cdot)\| - \|\bar{f}\| + \psi_2(a)} - \frac{\|D(a^*, a - a^*, \cdot)\|}{\|\bar{f} - D(a^*, a - a^*, \cdot)\| - \|\bar{f}\|} \right| \\
&= \left| \frac{\psi_1(a)(\|\bar{f} - D(a^*, a - a^*, \cdot)\| - \|\bar{f}\|) - \psi_2(a)\, D(a^*, a - a^*, \cdot)\|}{(\|f - F(a, \cdot)\| - \|f - F(a^*, \cdot)\|)(\|\bar{f} - D(a^*, a - a^*, \cdot)\| - \|\bar{f}\|)} \right| \\
&= \frac{\|F(a, \cdot) - F(a^*, \cdot)\|}{\|f - F(a, \cdot)\| - \|f - F(a^*, \cdot)\|} \\
&\quad \times \left| \left( \psi_1(a) - \psi_2(a) \frac{\|D(a^*, a - a^*, \cdot)\|}{\|\bar{f} - D(a^*, a - a^*, \cdot)\| - \|\bar{f}\|} \right) \middle/ (\|F(a, \cdot) - F(a^*, \cdot)\|) \right|.
\end{aligned} \tag{4.6}$$

Now by the strong uniqueness of $F(a^*, \cdot)$, there is a $\gamma_1 > 0$ such that $\|f - F(a, \cdot)\| - \|f - F(a^*, \cdot)\| \geqslant \gamma_1 \|F(a, \cdot) - F(a^*, \cdot)\|$, and by the strong uniqueness of $B_{V^*}(\bar{f})$, there is a $\gamma_2 > 0$ such that $\|\bar{f} - D(a^*, a - a^*, \cdot)\| - \|\bar{f}\| \geqslant \gamma_2 \|D(a^*, a - a^*, \cdot)\|$. So we have from (4.6),

$$\left| \frac{\|F(a, \cdot) - F(a^*, \cdot)\|}{\|f - F(a, \cdot)\| - \|f - F(a^*, \cdot)\|} - \frac{\|D(a^*, a - a^*, \cdot) - B_{V^*}(\bar{f})\|}{\|\bar{f} - D(a^*, a - a^*, \cdot)\| - \|\bar{f} - B_{V^*}(\bar{f})\|} \right|$$
$$\leqslant (1/\gamma_1)[|\psi_1(a)|/\|F(a, \cdot) - F(a^*, \cdot)\| + \psi_2(a)|/(\gamma_2 \|F(a, \cdot) - F(a^*, \cdot)\|)],$$

so setting

$$u(\delta)=\sup\{(1/\gamma_1)[|\psi_1(a)|/\|F(a,\cdot)-F(a^*,\cdot)\|+|\psi_2(a)|/(\gamma_2\,\|F(a,\cdot)-F(a^*,\cdot)\|)]:\ 0<\|F(a,\cdot)-F(a^*,\cdot)\|\leqslant\delta\},$$

the claim follows.

Now for $\delta>0$ sufficiently small, if $0<\|F(a,\cdot)-F(a^*,\cdot)\|\leqslant\delta$ then

$$0<\|D(a^*,a-a^*,\cdot)\|=\|F(a,\cdot)-F(a^*,\cdot)\|+o(\|F(a,\cdot)-F(a^*,\cdot)\|)\leqslant 2\delta$$

and it follows that

$$\sup\left\{\frac{\|F(a,\cdot)-F(a^*,\cdot)\|}{\|f-F(a,\cdot)\|-\|f-F(a^*,\cdot)\|}: a\in A,\ 0<\|F(a,\cdot)-F(a^*,\cdot)\|\leqslant\delta\right\}$$
$$\leqslant\sup\left\{\frac{\|p-B_{V^*}(\bar f)\|}{\|\bar f-p\|-\|\bar f\|}\mid p\in V^*,\ 0<\|p\|\leqslant 2\delta\right\}+u(\delta).$$

Thus $\lim_{\delta\to 0}M_V(f,\delta)\geqslant\lim_{\delta\to 0}M_{V^*}(\bar f,\delta)$.

For the reverse inequality, suppose $p\in V^*$, with $0<\|p\|\leqslant\delta$, for $\delta$ small. Without loss of generality we can assume that $\{F_1(a^*,\cdot),...,F_N(a^*,\cdot)\}$ is a basis for $W(a^*)$, so $p=D(a^*,c,\cdot)$ for some $c\in\mathbf{R}^n$, and we can take $c_i=0$ for $i>N$. Now by standard arguments from the linear independence of $\{F_1(a^*),...,F_N(a^*,\cdot)\}$ we have that for some $L>0$, $\|c\|\leqslant L\,\|p\|$. Thus, by Lemma 1(a), for $\delta>0$ sufficiently small,

$$\|F(a^*+c,\cdot)-F(a^*,\cdot)\|=\|D(a^*,c,\cdot)+o(\|c\|)\|\leqslant\|p\|+o(L\,\|p\|)\leqslant 2\delta.$$

So by arguments similar to those above, for $\delta>0$ sufficiently small, we have

$$\sup\left\{\frac{\|p-B_{V^*}(\bar f)\|}{\|\bar f-p\|-\|\bar f\|}: p\in V^*,\ 0<\|p\|\leqslant 2\delta\right\}$$
$$\leqslant\sup\left\{\frac{\|F(a,\cdot)-F(a^*,\cdot)\|}{\|f-F(a,\cdot)\|-\|f-F(a^*,\cdot)\|}: a\in A,\ 0<\|F(a,\cdot)-F(a^*,\cdot)\|\leqslant 2\delta\right\}$$
$$+u(2\delta).$$

Therefore, $\lim_{\delta\to 0}M_V(\bar f,\delta)\leqslant\lim_{\delta\to 0}M_{V^*}(f,\delta)$. Since $\|e_{V^*}(\bar f)\|\neq 0$ and $E_{V^*}(\bar f)=E_V(f)$, we have by Theorem 3 $\hat M_V(f)=\lim_{\delta\to 0}M_V(f,\delta)\leqslant\lim_{\delta\to 0}M_{V^*}(\bar f,\delta)=\max\{\|q_i\|:0\leqslant i\leqslant N\}$.

*Remark.* The strong unicity constant $M_V(f)$ and the local strong unicity constant $\hat M_V(f)=\lim_{\delta\to 0}M_V(f,\delta)$, as noted earlier, are identical if $V$ is linear. If $V$ is nonlinear then in general $\hat M_V(f)\neq M_V(f)$. This was demonstrated for ordinary rational functions on a closed interval in [11].

## REFERENCES


1. J. R. ANGELOS, M. S. HENRY, E. H. KAUFMAN, JR., AND T. D. LENKER, Local Lipschitz constants, *J. Approx. Theory* **43** (1985), 53–63.
2. J. R. ANGELOS, M. S. HENRY, E. H. KAUFMAN, JR., A. KROÒ, AND T. D. LENKER, On the continuity and differentiability of the best approximation operator, *in* "Alfred Haar Memorial Conference" (J. Szabados and K. Tandori, Eds.), pp. 127–144, Colloquia Mathematica Societatis János Bolyai, Vol. 49, North-Holland, Amsterdam, 1985.
3. R. B. BARRAR AND H. L. LOEB, On the continuity of the nonlinear Tschebyscheff operator, *Pacific J. Math.* **32** (1970), 593–601.
4. E. W. CHENEY, "Introduction to Approximation Theory," Chelsea, New York, 1982.
5. A. K. CLINE, Lipschitz conditions on uniform approximation operators, *J. Approx. Theory* **8** (1973), 160–172.
6. L. COLLATZ, Inclusion theorems for the minimal distance in rational Tchebycheff approximation with several variables, *in* "Approximation of Functions" (H. L. Garabedian, Ed.) pp. 43–56, Elsevier, New York, 1965.
7. L. COLLATZ AND W. KRABS, "Approximationstheorie. Tschebyscheffsche Approximation mit Anwendungen," Teubner, Stuttgart, 1973.
8. C. B. DUNHAM, Chebyshev approximation with the local Haar condition. II. Proceedings of the 1985 Manitoba Conference on Numerical Mathematics and Computing, *Congress. Numer.* **51** (1986), 123–136.
9. C. GEIGER, Minimale *H*-Mengen, Starke Eindeutigkeit und Stetigkeit des Tschebyscheff-Operators, *ZAMM* **54** (1974), 627–634.
10. M. S. HENRY, E. H. KAUFMAN, JR., AND T. D. LENKER, Lipschitz constants for small perturbations, in "Approximation Theory IV" (C. Chui, L. Schumaker, and J Ward, Eds.) pp. 515–520, Academic Press, New York, 1983.
11. M. S. HENRY, D. SCHMIDT, AND J. J. SWETITS, Uniform strong unicity for rational approximation, *J. Approx. Theory* **33** (1981), 131–146.
12. A. V. KOLUSOV, On differentiability of the operator of best approximation by rational functions, *Soviet Math. Dokl.* **25** (1982), 802–805.
13. A. KROÒ, Differential properties of the operator of best approximation, *Acta Math. Acad. Sci. Hungar.* **30** (1977), 319–331.
14. A. KROÒ, The Lipschitz constant of the operator of best approximation, *Acta Math. Acad. Sci. Hungar.* **35** (1980), 279–292.
15. G. MEINARDUS, "Approximation of Functions: Theory and Numerical Methods" (translated by L. L. Schumaker), Springer Tracts in Natural Philosophy, Vol. 13, Springer-Verlag, New York, 1967.
16. J. R. RICE, Tschebysheff approximations by functions unisolvent of variable degree, *Trans. Amer. Math. Soc.* **99** (1961), 298–302.
17. J. R. RICE, "The Approximation of Functions," Vol. 2, Addison–Wesley, Reading, MA, 1964.
18. D. SCHMIDT, A characterization of strong unicity constants, *in* "Approximation Theory III" (E. W. Cheney, Ed.) pp. 805–810, Academic Press, New York, 1980.
19. J. P. THIRAN, PH. DEFERT, AND E. PANIER, On *H*-sets in bivariate rational approximation, *J. Approx. Theory* **37** (1983), 356–371.