2007

Lipschitz Continuity and Gateaux Differentiability of the Best Approximation Operator in Vector-Valued Chebyshev Approximation

Martin Bartelt

John Swetits
Old Dominion University

Original Publication Citation

Bartelt, M., & Swetits, J. (2007). Lipschitz continuity and Gateaux differentiability of the best approximation operator in vector-valued Chebyshev approximation. *Journal of Approximation Theory, 148*(2), 177-193. doi:10.1016/j.jat.2007.03.005

Available online at www.sciencedirect.com
ScienceDirect
Journal of Approximation Theory 148 (2007) 177–193
Journal of Approximation Theory
www.elsevier.com/locate/jat

Lipschitz continuity and Gateaux differentiability of the best approximation operator in vector-valued Chebyshev approximation

Martin Bartelt[a], John Swetits[b,*]

[a]*Department of Mathematics, Christopher Newport University, Newport News, VA 23606, USA*
[b]*Department of Mathematics and Statistics, Old Dominion University, Norfolk, VA 23529, USA*

Received 25 May 2006; received in revised form 16 October 2006; accepted 10 March 2007

Communicated by Günther Nürnberger
Available online 24 March 2007

Abstract

When G is a finite-dimensional Haar subspace of $C\left(X, R^k\right)$, the vector-valued functions (including complex-valued functions when k is 2) from a finite set X to Euclidean k-dimensional space, it is well-known that at any function f in $C\left(X, R^k\right)$ the best approximation operator satisfies the strong unicity condition of order 2 and a Lipschitz (Hőlder) condition of order $\frac{1}{2}$. This note shows that in fact the best approximation operator satisfies the usual Lipschitz condition of order 1 and has a Gateaux derivative on a dense set of functions in $C\left(X, R^k\right)$.

Keywords: Best Chebyshev approximation; Vector-valued best approximation

1. Introduction

Let X be a finite set with the discrete topology and $C\left(X, R^k\right)$ be the space of vector-valued functions from X to k-dimensional Euclidean space R^k. A natural norm for functions in $C\left(X, R^k\right)$

* Corresponding author. Fax: +1 757 6833885.
E-mail addresses: mbartelt@pcs.cnu.edu (M. Bartelt), jswetits@odu.edu (J.J. Swetits).

0021-9045/$ - see front matter
doi:10.1016/j.jat.2007.03.005

is defined as follows:

$$\|f\| := \|f\|_X := \max_{x\in X} \|f(x)\|_2 , \tag{1}$$

where $\|\cdot\|_2$ denotes the Euclidean norm on R^k.

Let G be an n-dimensional Haar subspace in $C\left(X, R^k\right)$ with $\dim G \geqslant 1$ (i.e., the trivial case $G = \{0\}$ is excluded) and basis $\{g_1, \ldots, g_n\}$. For a given function f in $C\left(X, R^k\right)$ consider the vector-valued Chebyshev approximation problem of finding a function $B(f)$ in G that is a best approximation to f, i.e., $\|f - Bf\| = dist(f, G)$, where

$$dist(f, G) := \min_{g\in G} \|f - g\| . \tag{2}$$

One special case of the above best approximation problem is the complex Chebyshev approximation problem on the set X when $k = 2$ since $C\,(X, C)$ can be identified with $C\left(X, R^2\right)$ using $f_1(x) + i f_2(x) \leftrightarrow (f_1(x), f_2(x))$. The norm in (1) is just the usual Chebyshev norm for complex functions when $k = 2$.

Say that $B(f) := B_G(f)$ is strongly unique of order α if there exists a positive constant γ (depending on f, α and G) such that

$$\|f - g\|^\alpha \geqslant dist\,(f, G)^\alpha + \gamma \cdot dist\,(g, B(f)) \quad \text{for } g \in G. \tag{3}$$

Zukhovitskii and Stechkin [14] (cf. also [2]) showed that there is a unique best approximation to every f in $C\left(X, R^k\right)$ if and only if G satisfies the (generalized) Haar condition. From now on, G is assumed to be Haar. When G is a Haar subspace there is strong unicity of order $\alpha = 2$ for $B(f)$ [2]. However, in general, there will not be strong unicity of order 1 as observed for complex approximation [7,10]. Cheney [5] showed that in a normed linear space whenever a best approximation operator B has strong unicity of order 1 at a given function f, then it satisfies at f a Lipschitz condition of order 1, i.e., there is a positive constant λ such that

$$\|Bf - Bh\| \leqslant \lambda \|f - h\| \tag{4}$$

for all h in the normed linear space. The operator B is said to satisfy a Hőlder continuity condition of order $\frac{1}{2}$ at f [2] if there exists a positive number $\lambda = \lambda(f)$ such that

$$\|B(f) - B(h)\| \leqslant \lambda \|f - h\|^{\frac{1}{2}} (1 + \|f + h\|)^{\frac{1}{2}} \tag{5}$$

for all h in $C\left(X, R^k\right)$. Equivalently

$$\|Bf - Bh\| \leqslant \lambda \|f - h\|^{\frac{1}{2}} \tag{6}$$

for all h in $C\left(X, R^k\right)$ satisfying $\|f\| \leqslant M$ for some constant M. In approximation in $C\left(X, R^k\right)$ and therefore in complex approximation, it is known [2] that B satisfies a Hőlder condition of order $\frac{1}{2}$.

Part of the original motivation for this paper comes from the well-known [12] fact that in Hilbert space even though the projection operator onto a closed subspace (the best approximation operator associated with that subspace) has strong unicity of order 2, but not of order 1 in general, it is Lipschitz continuous of order 1. This leads to the following conjecture:

Conjecture 1. *In $C\left(X, R^k\right)$ the best approximation operator from a Haar subspace has Lipschitz continuity of order* 1 *when X is finite*.

We do not prove the conjecture. However, we prove in Theorem 16 and Corollary 17 that there is a dense subset of $C\left(X, \mathbf{R}^k\right)$ where B has Lipschitz continuity of order 1 and in Theorem 19 that in the special case of $k = 2$ (complex approximation) when G is the constants that the conjecture holds.

Blatt [4] showed that the best approximation operator in complex approximation was strongly unique of order 1, and hence Lipschitz continuous, on a dense subset of $C\,(X, C)$, but this was under the assumption that X had at most $\dim G$ isolated points. Thus, Corollary 17 is an extension of Blatt's result to the case when X is finite.

In the real-valued case [8,9] ($k = 1$), the best approximation operator B has a left Gateaux derivative at any f in $C\left(X, \mathbf{R}^1\right)$, i.e., the limit

$$\lim_{t\to 0^-} \frac{B(f+t\varphi) - Bf}{t} := D_f^- B\,(\varphi)$$

exists in the sup norm for any function (direction) φ. Similarly, the right Gateaux derivative $D_f^+ B\,(\varphi)$ is defined and if $D_f^+ B\,(\varphi) = D_f^- B\,(\varphi) = D_f B\,(\varphi)$, call $D_f B\,(\varphi)$ the Gateaux derivative of B at f. The Gateaux derivative was shown to exist at f if and only if the cardinality of the set of extreme points of $f - Bf$ was exactly $1 + \dim G$. In Theorem 16 and Corollary 17 it is shown that there is a dense subset of $C\left(X, \mathbf{R}^k\right)$ on which the Gateaux derivative exists.

2. Definition and preliminaries

In this section let X be a compact Hausdorff space which is not necessarily finite.

As usual let the extreme point set be given by

$$E(f-g) := \left\{x \in X : \|(f-g)\,(x)\|_2 = \|f-g\|\right\}, \quad g \in G.$$

For completeness we give the definition of Zukhovitskii and Stechkin [14] for a Haar set in $C\left(X, R^k\right)$.

Definition 2. An n-dimensional subspace G in $C\left(X, R^k\right)$ is called a Haar set if

(i) every nonzero g in G has at most m zeroes, and
(ii) for any m distinct points $x_1, \ldots, x_m$ in X and any m vectors $v_1, \ldots, v_m$ in R^k, there is a vector-valued function g in G such that $g\,(x_i) = v_i$, $i = 1, \ldots, m$, where m is the unique maximal integer satisfying $mk < n \leqslant (m+1)k$.

We use the Kolmogorov criterion for best approximates and the strong Kolmogorov criterion for strongly unique best approximates. The notation $\langle , \rangle$ stands for the usual Euclidean inner product in R^k.

Proposition 3. *Let* $f \in C\left(X, \mathbf{R}^k\right) \backslash G$.

(i) (*Kolmogorov criterion*) *A function* g^* *in* G *is a best approximate to* f *if and only if*

$$\max_{x\in E(f-g^*)} \left\langle f(x) - g^*(x), g(x)\right\rangle \geqslant 0 \quad \textit{for all } g \textit{ in } G. \tag{7}$$

(ii) (*Strong Kolmogorov criterion*) *A function* g^* *in* G *is a strongly unique of order* $\alpha = 1$ *best approximate to* f *if and only if*

$$\max_{x \in E(f-g^*)} \langle f(x) - g^*(x), g(x) \rangle > 0 \tag{8}$$

for every nonzero g *in* G.

We also need a characterization of the best approximate which is a generalization of the notion of a reference introduced by Stiefel [13] and Blatt [4] which is closely related (see Proposition 13) to the notion of an annihilator [2,6]. Let $x_1, \ldots, x_q$ be points in X and let $S_1, \ldots, S_q$ be orthogonal linear transformations on R^k. Let $\langle\ ,\ \rangle$ denote the standard inner product on R^k and let $e_i, i = 1, \ldots, k$, denote the standard basis vectors in R^k. For $\lambda \in R^q, \lambda > 0$ means that $\lambda_i > 0, i = 1, \ldots, q$.

Definition 4. The collection $R = \{(x_i, S_i) : i = 1, \ldots, q\}$ is called a reference if the $q \times n$ matrix $B = \left(\langle S_i g_j(x_i), e_1\rangle\right)_{i=1, j=1}^{q,n}$ has rank $q-1$ and if there exists $\lambda \in R^q$, $\lambda > 0$, such that $\lambda^T B = 0$. Note that $q \leqslant n + 1$.

Definition 5. If $f \in C\left(X, \mathbf{R}^k\right)$, then a reference R is called a reference with respect to f if $S(f)(x) = \|f\| e_1$ for each $(x, S) \in R$.

Definition 6. A function $\sigma : X \to R^k$ is said to be an annihilator of G if there exist points $x_1, \ldots, x_q$ in X with $\sigma(x_i) \neq 0$ for $i = 1, \ldots, q$, such that $\sum_{i=1}^q \langle \sigma(x_i), g(x_i)\rangle = 0$ for every $g \in G$.

Remark 7. If σ is an annihilator of G, then it can be assumed that the matrix $\left(\left\langle \frac{\sigma(x_i)}{\|\sigma(x_i)\|_2}, g_j(x_i)\right\rangle\right)_{i=1, j=1}^{q,n}$, where $\{g_1, \ldots, g_n\}$ is a basis for G, has rank $q-1$.

Proof. We have

$$\sum_{i=1}^q \|\sigma(x_i)\|_2 \left\langle \frac{\sigma(x_i)}{\|\sigma(x_i)\|_2}, \ g_j(x_i)\right\rangle_{j=1}^n = 0$$

or

$$\|\sigma(x_1)\|_2 \left\langle \frac{\sigma(x_1)}{\|\sigma(x_1)\|_2}, \ g_j(x_1)\right\rangle_{j=1}^n + \sum_{i=2}^q \|\sigma(x_i)\|_2 \left\langle \frac{\sigma(x_i)}{\|\sigma(x_i)\|_2}, \ g_j(x_i)\right\rangle_{j=1}^n = 0.$$

Since a positive combination of vectors can always be replaced by a positive combination of an independent subset (cf. [11]), we can replace

$$\sum_{i=2}^q \|\sigma(x_i)\|_2 \left\langle \frac{\sigma(x_i)}{\|\sigma(x_i)\|_2}, \ g_j(x_i)\right\rangle_{j=1}^n$$

by, let us say,

$$\sum_{i=2}^r \mu_i \left\langle \frac{\sigma(x_i)}{\|\sigma(x_i)\|_2}, \ g_j(x_i)\right\rangle_{j=1}^n, \quad \mu_i > 0, \ i = 1, \ldots, r,$$

where the vectors $\left\langle \frac{\sigma(x_i)}{\|\sigma(x_i)\|_2}, g_j(x_i) \right\rangle_{j=1}^{n}$, $i = 2, \dots r$, are independent. The resulting matrix has rank $r - 1$. Just relabel r as q. The function $\widehat{\sigma}(x_1) = \sigma(x_1), \widehat{\sigma}(x_i) = \frac{\mu_1}{\|\sigma(x_i)\|_2} \sigma(x_i), i = 2, \dots, r$, is an annihilator of G. □

Recall the following characterization of best approximation.

Theorem 8 (*Deutsch [6]*). *A function $h \in G$ is a best approximation to $f \in C(X, R^k) \setminus G$ if and only if there exist points $x_1, \dots, x_q$, satisfying $\|f(x_i) - h(x_i)\|_2 = \|f - h\|$ and an annihilator σ of G satisfying $\frac{\sigma(x_i)}{\|\sigma(x_i)\|_2} = \frac{f(x_i)-h(x_i)}{\|f-h\|}$, $i = 1, \dots, q$, where $q \leqslant n + 1$.*

Call the points $x_1, \dots, x_q$ an annihilator or the support of an annihilator for $f - Bf$. We then have the following characterization of best approximation:

Theorem 9. *A function $g \in G$ is a best approximation to $f \in C(X, R^k) \setminus G$ if and only if there exists a reference R with respect to $f - g$.*

Proof. First assume that $g \in G$ is a best approximation of *f*. Then, by Theorem 8, there exist points $x_1, \dots, x_q$ and a function $\sigma : X \to R^k$ such that $\sigma(x_i) \neq 0$, $i = 1, \dots, q$,

$$\frac{\sigma(x_i)}{\|\sigma(x_i)\|_2} = \frac{f(x_i) - g(x_i)}{\|f - g\|}$$

and $\sum_{i=1}^{q} \langle \sigma(x_i), g_j(x_i) \rangle = 0$, $j = 1, \dots, n$. Thus, $\sum_{i=1}^{q} \lambda_i \langle f(x_i) - g(x_i), g_j(x_i) \rangle = 0$, $j = 1, \dots, n$, where $\lambda_i = \|\sigma(x_i)\|_2$. By Remark 7 we can assume that the matrix $B = \left(\langle f(x_i) - g(x_i), g_j(x_i) \rangle \right)_{i=1, j=1}^{q.n}$ has rank $q - 1$. Let S_i^T be an orthogonal transformation on R^k whose first column is $\frac{f(x_i)-g(x_i)}{\|f-g\|}$. Then $S_i^T e_1 = \frac{f(x_i)-g(x_i)}{\|f-g\|}$ and so $\langle f(x_i) - g(x_i), g_j(x_i) \rangle = \|f - g\| \langle S_i^T e_1, g_j(x_i) \rangle = \|f - g\| \langle e_1, S_i g_j(x_i) \rangle$. Thus, $\{(x_i, S_i) : i = 1, \dots, q\}$ is a reference with respect to $f - g$. Thus card (reference) $\leqslant n + 1$. Now assume there exists a reference $R = \{(x_i, S_i) : i = 1, \dots, q\}$ with respect to $f - g$. Then $\|f(x_i) - g(x_i)\|_2 = \|f - g\| \left\| S_i^T e_1 \right\|_2 = \|f - g\|$ and, for each $j = 1, \dots, n$,

$$\begin{aligned} \sum_{i=1}^{q} \lambda_i \langle S_i g_j(x_i), e_1 \rangle &= \sum_{i=1}^{q} \lambda_i \left\langle g_j(x_i), S_i^T e_1 \right\rangle \\ &= \sum_{i=1}^{q} \lambda_i \left\langle g_j(x_i), \frac{f(x_i) - g(x_i)}{\|f - g\|} \right\rangle \\ &= 0. \end{aligned} \tag{9}$$

Defining σ by $\sigma(x_i) = \lambda_i (f(x_i) - g(x_i))$ completes the proof. □

The following theorem shows that there is a particular set of functions in $C(X, R^k)$ at which B, by the result of Cheney, has Lipschitz continuity of order 1.

Theorem 10. *Suppose G is a generalized Haar subspace of dimension n. If there exists a reference of cardinality $n + 1$ with respect to $f - Bf$, where Bf is the unique best approximation to f, then Bf is strongly unique.*

Proof. Let $x_1, \ldots, x_{n+1}$ be points that comprise a reference. Then by (9)

$$\sum_{i=1}^{n+1} \lambda_i \langle f(x_i) - Bf(x_i), \quad g(x_i)\rangle = 0$$

for all $g \in G$, where $\lambda_i > 0$. Suppose there exists g such that

$$\max\{\langle f(x_i) - Bf(x_i), \quad g(x_i)\rangle : i = 1, \ldots, n+1\} \leqslant 0.$$

Then necessarily $\langle f(x_i) - Bf(x_i), \quad g(x_i)\rangle = 0, i = 1, \ldots, n + 1$. However, the matrix $\left(\left\langle f(x_i) - Bf(x_i), \quad g_j(x_i)\right\rangle\right)_{i=1,j=1}^{n+1,n}$ has rank n. Therefore g is identically zero. Thus, the strong Kolmogorov criterion is satisfied on $J := \{x_1, \ldots, x_{n+1}\}$. Thus there exists $c > 0$ such that

$$\|f - g\|_J \geqslant \|f - Bf\|_J + c\,\|Bf - g\|_J$$

for all $g \in G$. But $\|\,\|_J$ is a norm on G and $\|f - Bf\|_J = \|f - Bf\|$. Observing that $\|f - g\|_J \leqslant \|f - g\|$, we thus have $\gamma > 0$ such that

$$\|f - g\| \geqslant \|f - Bf\| + \gamma\,\|Bf - g\|. \qquad \square$$

It is easy to see that the following gives an equivalent condition for a reference.

Proposition 11. *A set of points $x_1, \ldots, x_q$ in $E(f - Bf)$ is a reference for $f - Bf$ if and only if there exist positive constants $\lambda_1, \ldots, \lambda_q$ such that*

(i) $$\sum_{i=1}^{q} \lambda_i \langle f(x_i) - Bf(x_i), \quad g(x_i)\rangle = 0, g \in G, \tag{10}$$

and

(ii) *the $q \times n$ matrix*

$$M := M\left(x_1, \ldots, x_q\right) := \left(\left\langle f(x_i) - Bf(x_i), g_j(x_i)\right\rangle\right)_{i=1,j=1}^{q,n} \tag{11}$$

has rank $q - 1$.

Notice that (i) implies that rank $(M) \leqslant q - 1$.

Remark 12. From the proof of Theorem 9 and Proposition 11 we see that if $x_1, \ldots, x_q$ are the points in an annihilator then some subset of them is the set of points in a reference and conversely that the points in a reference (by definition) are the points in an annihilator.

The following result clarifies the relationship between annihilator and reference.

Proposition 13. *The set of points $\left\{x_1, \ldots, x_q\right\}$ are the points in a reference for $f - Bf$ if and only if they are the support of an annihilator which has no proper subset which is the support of an annihilator.*

Proof. Suppose $\left\{x_1, \ldots, x_q\right\}$ is a reference for $f - Bf$. Let $R_1, \ldots, R_q$ be the rows of M in (11). Then there exist positive constants $\lambda_1, \ldots, \lambda_q$ such that $\sum_{i=1}^{q} \lambda_i R_i = 0$. By renumbering, if necessary, assume $x_1, \ldots, x_r$, $r < q$ is the support of an annihilator for $f - Bf$. Then there exist positive constants $\alpha_1, \ldots, \alpha_r$ such that $\sum_{i=1}^{r} \alpha_i R_i = 0$. Then R_q is dependent on $\{R_1, \ldots, R_{q-1}\}$, but also $\{R_1, \ldots, R_{q-1}\}$ is dependent. Hence rank $(M) \leqslant q - 2$ which is a contradiction. Conversely, let $\{x_1, \ldots, x_q\}$ be the support of an annihilator no proper subset of which is the support of an annihilator. If $\{x_1, \ldots, x_q\}$ does not give a reference then rank $(M) \leqslant q - 2$ so there must exist positive constants $\lambda_1, \ldots, \lambda_q$ such that $\sum_{i=1}^{q} \lambda_i R_i = 0$. We can represent (cf. [11]) $\sum_{i=1}^{q} \lambda_i R_i$ as a positive linear combination of an independent subset $\{R_1, \ldots, R_r\}$, $r \leqslant q - 2$ as $\sum_{i=1}^{r} \alpha_i R_i = 0$ for positive constants $\alpha_1, \ldots, \alpha_r$. But then $\{x_1, \ldots, x_r\}$ is the support of an annihilator which is a contradiction. □

Recall that the δ-local Lipschitz constant $\lambda_\delta(f)$ is defined (cf. [1]) for B at f in $C\left(X, R^k\right)$ by

$$\lambda_\delta(f) := \sup\left\{\frac{\|Bf - Bh\|}{\|f - h\|} : 0 < \|f - h\| < \delta\right\}. \tag{12}$$

The Lipschitz constant (of order 1) is defined by

$$\lambda(f) := \sup\left\{\frac{\|Bf - Bh\|}{\|f - h\|} : 0 < \|f - h\|\right\}. \tag{13}$$

It follows easily that $\lambda_\delta(f) < \infty$ for some $\delta > 0$ if and only if $\lambda(f) < \infty$, for if $\lambda_\delta(f) < \infty$ then since $\lim_{\|h\| \to \infty} \frac{\|Bf - Bh\|}{\|f - h\|} \leqslant 2$, it follows that for $\|f - h\| \geqslant M$ for sufficiently large M, $\frac{\|Bf - Bh\|}{\|f - h\|}$ will be bounded and $\sup\left\{\frac{\|Bf - Bh\|}{\|f - h\|} : \delta \leqslant \|f - h\| \leqslant M\right\}$ is clearly bounded.

3. Main results

Remark 14. It is known ([3], Corollary (15)) that when Bf is strongly unique R^n is the convex cone generated by $\left\{\sum_{m=1}^{n} \langle Bf(x) - f(x), g_m(x)\rangle e_m : x \in E(f - Bf)\right\}$ where $\{e_m : m = 1, \ldots, n\}$ is the standard basis for R^n and $\{g_m : m = 1, \ldots, n\}$ is a basis for the approximating subspace G. Hence if f has a strongly unique best approximate then card $(f - Bf) \geqslant n + 1$. Hence none of the functions in the next theorem have strong unicity when $q \leqslant n$.

When f satisfies the conditions of the following theorem, f has a strongly unique best approximate if and only if $q = n + 1$ by Theorem 10 and Remark 14. Also it is easy to find examples where there are functions satisfying the condition of the theorem with $q < n + 1$. One such example is G the constant in $C\left(X, R^2\right)$, $X = \{x_1, x_2, x_3\}$, $f(x_1) = \begin{pmatrix} -1 \\ 0 \end{pmatrix} = f(x_2)$, $f(x_3) = \begin{pmatrix} 1/2 \\ 0 \end{pmatrix}$ with *card* $(E(f - B(f))) = 2 < n + 1$. Furthermore, the number of points in a reference and hence in $E(f - Bf)$ is at least $\frac{n}{k} + 1$. Thus if *card* $(E(f - Bf)) = \frac{n}{k} + 1$, the minimal number, f satisfies the conditions of the following theorem. Corollary 17 will show that there actually are many functions satisfying the conditions of the following theorem.

It is well known that in the cases of real-valued and complex-valued approximations that the best approximation operator is linear if the cardinality of X is $n + 1$ and the dimension of the approximating subspace is n. The following example shows that this need not be the case in the more general vector-valued setting.

Example 15. Let $X = \{x_1, x_2\}$ and let G be the two-dimensional Haar subspace of $C(X, R^2)$ with basis $\{g_1, g_2\}$ where $g_1(x_1) = g_1(x_2) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, $g_2(x_1) = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ and $g_2(x_2) = \begin{pmatrix} 0 \\ 2 \end{pmatrix}$. Let $f(x_1) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, $f(x_2) = \begin{pmatrix} -1 \\ 0 \end{pmatrix}$, $h(x_1) = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ and $h(x_2) = \begin{pmatrix} 0 \\ -1 \end{pmatrix}$. Then it is easy to see that $Bf = 0 = Bh$. However, $B(f + h) \neq 0$ since there are no positive scalars λ_1 and λ_2 such that $\lambda_1 \langle (f + h)(x_1), g_i(x_1) \rangle + \lambda_2 \langle (f + h)(x_2), g_i(x_2) \rangle = 0$, $i = 1, 2$.

Theorem 16. *Suppose X is a finite set and G is an n-dimensional Haar subspace of $C\left(X, R^k\right)$. Let f in $C\left(X, R^k\right)$ and assume $E(f - Bf) = \left\{x_1, \ldots, x_q\right\}$ and that $\left\{x_1, \ldots, x_q\right\}$ is a reference for $f - Bf$. Then*

(i) *the best approximation operator is Gateaux differentiable at f, and*
(ii) *the best approximation operator is Lipschitz continuous of order 1 at f.*

Proof. The proof of the theorem uses the Implicit Function Theorem to show that a particular system of equations yields implicit functions which are continuously differentiable. We first give the equations and then give a long verification that the Jacobian of the system is invertible. Let $\{g_1, \ldots, g_n\}$ be a basis for G. Assume without loss of generality that $Bf = 0$ and $\|f\| = 1$. Since $\left\{x_1, \ldots, x_q\right\}$ is a reference for f, there are positive constants $\lambda_1, \ldots, \lambda_q$ such that $\sum_{i=1}^{q} \lambda_i \left\langle f(x_i), g_j(x_i)\right\rangle = 0$, $j = 1, \ldots, n$ and the $q \times n$ matrix $\left\langle f(x_i), g_j(x_i)\right\rangle_{i=1, j=1}^{q,n}$ has rank $q - 1$. We can assume without loss of generality that $\lambda_q = 1$. Let $\varphi \in C\left(X, R^k\right)$ with $\|\varphi\| = 1$. Then since X is finite there exists a $\delta > 0$ such that if $|t| \leqslant \delta$, then

$$E(f + t\varphi - B(f + t\varphi)) \subseteq E(f) \tag{14}$$

and $\left\{x_1, \ldots, x_q\right\}$ is a reference for $f + t\varphi - B(f + t\varphi)$. Therefore, there are positive constants $\lambda_i(t)$, $i = 1, \ldots, q$ such that

$$\sum_{i=1}^{q} \lambda_i(t) \left\langle f(x_i) + t\varphi(x_i) - B(f + t\varphi)(x_i), g_j(x_i)\right\rangle = 0. \tag{15}$$

By continuity, $\lim_{t\to 0} \lambda_i(t) = \lambda_i > 0$, $i = 1, \ldots, q$. Since $\lambda_q = 1$, we can normalize and assume $\lambda_q(t) = 1$ for all t, $|t| \leqslant \delta$. Then (15) becomes

$$\sum_{i=1}^{q-1} \lambda_i(t) \left\langle f(x_i) + t\varphi(x_i) - B(f + t\varphi)(x_i), g_j(x_i)\right\rangle + \left\langle f(x_q) + t\varphi(x_q) - B(f + t\varphi)(x_q), g_j(x_q)\right\rangle = 0,$$

and

$$\sum_{i=1}^{q-1} \lambda_i \left\langle f(x_i), q_j(x_i)\right\rangle + \left\langle f(x_q), g_j(x_q)\right\rangle = 0. \tag{16}$$

Let $B(f + t\varphi) = \sum_{m=1}^{n} a_m(t) g_m$. So $a_m(0) = 0$, $m = 1, \ldots, n$.

Define a function $F : R^{n+q} \times R \to R^{n+q}$ by

$$F_j(\mu_1, \ldots, \mu_{q-1}, a_1, \ldots, a_m, t, e)$$

$$= \sum_{i=1}^{q-1} \mu_i \left\langle f(x_i) + t\varphi(x_i) - \sum_{m=1}^{n} a_m g_j(x_i), g_j(x_i) \right\rangle$$
$$+ \left\langle f(x_q) + t\varphi(x_q) - \sum_{m=1}^{n} a_m g_m(x_q), g_j(x_q) \right\rangle, \ j = 1, \ldots, n$$

and

$$F_j(\mu_1, \ldots, \mu_{q-1}, a_1, \ldots, a_m, t, e) = \left\| f(x_{j-n}) + t\varphi(x_{j-n}) - \sum_{m=1}^{n} a_m g_m(x_{j-n}) \right\|_2^2 - e,$$
$$j = n+1, \ldots, n+q. \tag{17}$$

By (16), (17) satisfies $F = 0$ at $t = 0, \mu_i = \lambda_i, i = 1, \ldots, q-1, a_m = 0, m = 1, \ldots, n$, and $e = 1$. $F = 0$ is a system of $n+q$ equations in the $n+q+1$ unknowns t, $\{\mu_i\}_{i=1}^{q-1}$, $\{a_m\}_{m=1}^{n}$ and e.

Now there is a lengthy verification that the Jacobian for the system of equations (17) is invertible. For the first n equations

$$\frac{\partial F_j}{\partial \mu_i} = \langle f(x_i), g_j(x_i) \rangle, \quad i = 1, \ldots, q-1, \quad j = 1, \ldots, n,$$

$$\frac{\partial F_j}{\partial a_m} = -\sum_{i=1}^{q-1} \lambda_i \langle g_m(x_i), g_j(x_i) \rangle - \langle g_m(x_q) - g_j(x_q) \rangle, \quad m = 1, \ldots, n, \quad j = 1, \ldots, n,$$

$$\frac{\partial F_j}{\partial e} = 0, \quad j = 1, \ldots, n.$$

For the second q equations

$$\frac{\partial F_j}{\partial \mu_i} = 0, \quad i = 1, \ldots, q-1, \quad j = n+1, \ldots, n+q,$$

$$\frac{\partial F_j}{\partial a_m} = -2\langle f(x_{j-n}), g_m(x_{j-n}) \rangle, \quad j = n+1, \ldots, n+q, \quad m = 1, \ldots, n,$$

$$\frac{\partial F_j}{\partial e} = -1, \quad j = n+1, \ldots, n+q.$$

So the Jacobian of the system of equations with respect to μ_i, $i = 1, \ldots, q-1$, $a_m, m = 1, \ldots, n$ and e has the block structure

$$\begin{pmatrix} & & 0 \\ A^T & B & \vdots \\ & & 0 \\ & & -1 \\ 0 & C & \vdots \\ & & -1 \end{pmatrix} := \begin{pmatrix} A^T & \\ & D \\ 0 & \end{pmatrix},$$

where A is the $(q-1)\times n$ matrix $\left\langle f(x_i), g_j(x_i)\right\rangle_{i=1,j=1}^{q-1,n}$ which has rank $q-1$ and B is the $n\times n$ matrix $\left(-\sum_{i=1}^{q-1}\lambda_i\left\langle g_m(x_i), g_j(x_i)\right\rangle - \left\langle g_m(x_q), g_j(x_q)\right\rangle\right)_{m,j=1}^{n}$ and C is the $q\times n$ matrix $-2\left\langle f(x_i), g_m(x_i)\right\rangle_{i=1,m=1}^{q,n}$.

First we verify that B has rank n. Suppose that $B\begin{pmatrix} b_1\\ \vdots \\ b_n\end{pmatrix} = 0$. Then

$$-\sum_{i=1}^{q-1}\lambda_i\left\langle \sum_{m=1}^{n} b_m g_m(x_i), g_j(x_i)\right\rangle - \left\langle \sum_{m=1}^{n} b_m g_m(x_q), g_j(x_q)\right\rangle = 0, \quad j=1,\ldots,n.$$

Let $\hat{g} = \sum_{m=1}^{n} b_m g_m$. Because $\{g_1,\ldots,g_n\}$ is a basis for G and $\hat{g}\in G$ it follows that $-\sum_{i=1}^{q-1}\lambda_i \left\langle \hat{g}(x_i), \hat{g}(x_i)\right\rangle - \left\langle \hat{g}(x_q), \hat{g}(x_q)\right\rangle = 0$. Since $\lambda_i > 0,\ i=1,\ldots,q-1$, it follows that $\left\|\hat{g}(x_i)\right\|_2 = 0,\ i=1,\ldots,q$. Therefore $\hat{g}(x_i)=0,\ i=1,\ldots,q$. Because $\left\{x_1,\ldots,x_q\right\}$ is a reference for f and G is Haar it follows that $\hat{g}$ is identically zero. Therefore $b_i = 0,\ i=1,\ldots,n$. Thus the rank of B is n. Now the matrix

$$\begin{pmatrix} & & -1 \\ C & \vdots & \vdots \\ & & -1 \end{pmatrix} \text{ is row equivalent to } \begin{pmatrix} & & \vdots & -1 & \\ -2A & & \vdots & \vdots & \\ & & \vdots & -1 & \\ 0 & \cdots\ 0 & \vdots & -1 & -\sum_{i=1}^{q-1}\lambda_i \end{pmatrix}.$$

Therefore, the column rank of D is $n+1$ and the column rank of $\begin{pmatrix} A^T \\ 0\end{pmatrix}$ is $q-1$. Now we show that no linear combination of the columns of $\begin{pmatrix} A^T \\ 0\end{pmatrix}$ is equal to a linear combination of the columns of D except for the zero vector. Once that is done we use the fact that if S_1 and S_2 are two subspaces such that $S_1\cap S_2 = \{0\}$, then a basis for S_1 union a basis for S_2 is a linearly independent set to conclude that the Jacobian is nonsingular. Here S_1 is the column space of $\begin{pmatrix} A^T \\ 0\end{pmatrix}$ and S_2 is the column space of D. Suppose a linear combination of the columns of $\begin{pmatrix} A^T \\ 0\end{pmatrix}$ is equal to a linear combination of the columns of D. From the structure of the two matrices it is clear that the last column of D is not involved. So there are scalars $\gamma_1,\ldots,\gamma_{q-1}$ and $\alpha_1,\ldots,a_n$ such that

$$\sum_{m=1}^{q-1}\gamma_m\left\langle f(x_m), g_j(x_m)\right\rangle = \sum_{r=1}^{n}\alpha_r\left[-\sum_{i=1}^{q-1}\lambda_i\left\langle g_j(x_i), g_r(x_i)\right\rangle - \left\langle g_j(x_q), g_r(x_q)\right\rangle\right]$$

and

$$0 = \sum_{r=1}^{n}\alpha_r\left[-2\left\langle f(x_l), g_r(x_i)\right\rangle\right], \quad l=1,\ldots,q-1. \tag{18}$$

Let $\hat{g} = \sum_{r=1}^{n} \alpha_r p_r$. Then

$$0 = \left\langle f(x_l), \hat{g}(x_l) \right\rangle, \quad l = 1, \ldots, q-1, \tag{19}$$

and

$$\sum_{m=1}^{q-1} \gamma_m \left\langle f(x_m), g_j(x_m) \right\rangle = -\sum_{i=1}^{q-1} \lambda_i \left\langle g_j(x_i), \hat{g}(x_i) \right\rangle - \left\langle g_j(x_q), \hat{g}(x_q) \right\rangle, \quad j = 1, \ldots, n. \tag{20}$$

From (20) we get $\sum_{m=1}^{q-1} \gamma_m \left\langle f(x_m), \hat{g}(x_m) \right\rangle = -\sum_{i=1}^{q-1} \lambda_i \left\langle \hat{g}(x_i), \hat{g}(x_i) \right\rangle - \left\langle \hat{g}(x_q), \hat{g}(x_q) \right\rangle$.

Using (19) we get

$$0 = -\sum_{i=1}^{q-1} \lambda_i \left\langle \hat{g}(x_i), \hat{g}(x_i) \right\rangle - \left\langle \hat{g}(x_q), \hat{g}(x_q) \right\rangle. \tag{21}$$

Since $\lambda_i > 0, \ i = 1, \ldots, q-1$ we obtain from (21), $\hat{g}(x_i) = 0, \ i = 1, \ldots, q$ which implies $\hat{g} \equiv 0$ which implies $\alpha_1 = \cdots = \alpha_m = 0$ as before. Consequently, $\sum_{m=1}^{q-1} \gamma_m \left\langle f(x_m), g_j(x_m) \right\rangle = 0, \ j = 1, \ldots, n$. Since $\begin{pmatrix} A^T \\ 0 \end{pmatrix}$ has full column rank it follows that $\gamma_1 = \cdots = \gamma_m = 0$. Hence we now have verified that the Jacobian for the system of equations (17) is invertible. Now by the Implicit Function Theorem, there is a neighborhood of $t_0 = 0$ such that the system of equations (17) define $\mu_i(t)$, $a_m(t)$ and $e(t)$ as continuously differentiable functions of *t*. In particular

$$B(f + t\varphi) = \sum_{m=1}^{n} a_m(t) g_m \tag{22}$$

is differentiable at $t_0 = 0$ and thus *B* has a Gateaux derivative at *f* in any given direction φ. To prove (ii) we modify slightly the argument used to prove (i) by assuming that the $\varphi(x_i)$ values are variable. So consider the system of equations

$$\sum_{i=1}^{q-1} \mu_i \left\langle f(x_i) + t\Psi(x_i) - \sum_{m=1}^{n} a_m g_m(x_i), g_j(x_i) \right\rangle + \left\langle f(x_q) + t\Psi(x_q) - \sum_{m=1}^{n} a_m g_m(x_q), g_j(x_q) \right\rangle = 0, \quad j = 1, \ldots, n, \tag{23}$$

$$\left\| f(x_i) + t\Psi(x_i) - \sum_{m=1}^{n} a_m g_m(x_i) \right\|_2^2 - e = 0, \quad i = 1, \ldots, q, \tag{24}$$

where we now consider the values $\{\Psi(x_i)\}_{i=1}^{q}$ as variables. The system (23) is satisfied at $t_0 = 0, \mu_i = \lambda_i, \ i = 1, \ldots, q-1, a_m = 0, m = 1, \ldots, n, \Psi(x_i) = \varphi(x_i), i = 1, \ldots, q$, and $e = 1$. Thus (23) is a system of $n + q$ equations in the $n + 2q + 1$ variables t, $\left\{\mu_i\right\}_{i=1}^{q-1}$, $\{a_m\}_{m=1}^{n}$, e and $\{\Psi(x_i)\}_{i=1}^{q}$. The Jacobian of system (23) with respect to the variables e, μ_i and a_m at $e = 1$, $\mu_i = \lambda_i$, $a_m = 0$, $t_0 = 0$ and $\Psi(x_i) = \varphi(x_i)$ is the same as the Jacobian of (17) so it is invertible. By the Implicit Function Theorem then, there is a neighborhood of $t_0 = 0$ and a neighborhood $U(\varphi)$ in $R^{k \times q}$, i.e., a neighborhood of $(\varphi(x_1), \ldots, \varphi(x_q))$, such that e, μ_i and a_m are continuously differentiable functions of *t* and $\Psi(x_i), \ i = 1, \ldots, q$, for all *t* close to 0 and all $\Psi \in U(\varphi)$. Denote the neighborhood of $t_0 = 0$ by $N(0, \rho(\varphi))$. So e, μ_i and a_m are continuously differentiable for all $(t, \Psi) \in N(0, \rho(\varphi)) \times U(\varphi)$. Let $Z := \left\{\left(\varphi(x_1), \ldots, \varphi(x_q)\right) : \varphi \in C\left(X, R^k\right), \|\varphi\| = 1\right\}$. Then Z is a compact subset of $R^{k\times q}$, $\{U(\varphi)\}$ is an open covering of Z and hence there is a finite subcover $\left\{U\left(\varphi^l\right) : l = 1, \ldots, p\right\}$. Let $\rho = \min\left\{\rho\left(\varphi^l\right) : l = 1, \ldots, p\right\}$. Then e, μ_i and a_m are continuously differentiable for all $(t, \Psi) \in N(0, \rho) \times \left(\bigcup_{l=1}^{p} U\left(\varphi^l\right)\right)$, and $\overline{N(0, \rho/2)} \times Z$ is a compact subset of $N(0, \rho) \times \left(\bigcup_{l=1}^{p} U\left(\varphi^l\right)\right)$. Therefore, each $a_m(t, \varphi)$ and $\frac{\partial a_m}{\partial t}(t, \varphi)$ is uniformly bounded on $W := \overline{N(0, \rho/2)} \times Z$. Let

$$K := \max_{1 \leqslant m \leqslant n} \left\{\max\left\{\left|\frac{\partial (a_m)}{\partial t}\right| : (t, \varphi) \in W\right\}\right\}.$$

By the Mean Value Theorem, for some s between 0 and t

$$|a_m(t, \varphi)| = \left|\frac{\partial a_m(s, \varphi)}{\partial t}\right| \quad |t| \leqslant K|t|.$$

Therefore

$$\|B(f + t\varphi)\| = \left\|\sum_{m=1}^{n} a_m(t, \varphi) g_m\right\| \leqslant \left(\left(\frac{K}{2}\right) \sum_{m=1}^{n} \|g_m\|\right) |t|$$

and therefore the best approximation operator is Lipschitz continuous at *f*. □

Corollary 17. *The set of functions*

$$S = \left\{f \in C\left(X, R^k\right) : E(f - Bf) = \{x_i\}_{i=1}^{q} \text{ is a reference for } f\right\}$$

is dense in $C\left(X, R^k\right)$. *Hence* $C\left(X, R^k\right)$ *contains a dense set of functions on which B is Gateaux differentiable and has Lipschitz continuity of order* 1.

Proof. By Theorem 16 it is only necessary to show that the set S is dense in $C\left(X, R^k\right)$. Let $h \in C\left(X, R^k\right)$ and let $\left\{x_1, \ldots, x_q\right\}$ be a reference for $h - Bh$. Let $f(x_i) := h(x_i)$, $i = 1, \ldots, q$. For each $x \neq x_i$ in X, let $\varepsilon(x)$ be a vector such that

$$\|h(x) - Bh(x) + \varepsilon(x)\| < \|h - Bh\|$$

and let $f(x) = h(x) + \varepsilon(x)$. Then $Bh = Bf$ and $E(f - Bf) = \left\{x_1, \ldots, x_q\right\}$ and it is a reference for *f* so $f \in S$. □

The following is a converse to Theorem 16 (i).

Theorem 18. *Suppose X is a finite set and G is an n-dimensional Haar subspace of* $C\left(X, R^k\right)$. *If the best approximation operator has a Gateaux derivative at f, then no proper subset of* $E(f)$ *is a reference.*

Proof. Assume that $\|f\| = 1$ and $Bf = 0$ and let $E(f) = \{x_1, \ldots, x_m\}$ with $\left\{x_1, \ldots, x_q\right\}$ a reference for *f* and $q < m$. We first assume that every reference has cardinality greater than or equal to 2. Then $\{x_1, \ldots, x_q\}$ is the support of an annihilator for f so there exist positive scalars $\lambda_1, \ldots, \lambda_q$ such that

$$\sum_{i=1}^{q} \lambda_i \langle f(x_i), g(x_i)\rangle = 0 \quad \text{for all } g \text{ in } G. \tag{25}$$

Define φ in $C(X, R^k)$ by

$$\varphi(x) = \begin{cases} f(x_i), & x = x_1, \ldots, x_q, \\ -2f(x_{q+1}), & x = x_{q+1}, \\ 0, & x \notin \{x_1, \ldots, x_{q+1}\}. \end{cases}$$

Then

$$(f + t\varphi)(x) = \begin{cases} (1+t)\, f(x_i), & i = 1, \ldots, q, \\ (1-2t)\, f(x_{q+1}), & x = x_{q+1}, \\ f(x), & x \notin \{x_1, \ldots, x_{q+1}\}. \end{cases}$$

So for $t > 0$ and sufficiently small,

$$\|(f + t\varphi)(x)\|_2 = \begin{cases} 1+t, & x = x_1, \ldots, x_q, \\ |1-2t|, & x = x_{q+1}, \\ 1, & x \notin \{x_1, \ldots, x_{q+1}\}. \end{cases}$$

Hence $\|(f + t\varphi)\| = 1 + t$ and, using (24) we have

$$\sum_{i=1}^{q} \lambda_i \langle (f + t\varphi)(x_i), g(x_i)\rangle = (1+t) \sum_{i=1}^{q} \lambda_i \langle f(x_i), g(x_i)\rangle = 0$$

for all g in G and hence $B(f + t\varphi) = 0$. Therefore, $\lim_{t\to 0^+} \frac{B(f+t\varphi)}{t} = 0$. Assume now that $\lim_{t\to 0^-} \frac{B(f+t\varphi)}{t} = 0$, or equivalently that $\lim_{t\to 0^+} \frac{B(f-t\varphi)}{t} = 0$. Then if t is sufficiently small $E((f - t\varphi) - B(f - t\varphi)) \subseteq E(f)$ and one can easily verify that any reference for $f - t\varphi - B(f - t\varphi)$ contains a reference for f. Let $\{y_1, \ldots, y_l\}$ be a reference for f that is contained in a reference for $f - t\varphi - B(f - t\varphi)$. By assumption $l \geqslant 2$. Let $x \in E(f)$. Then

$$\begin{aligned} \|f - t\varphi - B(f - t\varphi)\|^2 - 1 &\geqslant \|f - t\varphi - B(f - t\varphi)(x)\|_2^2 - 1 \\ &= \|-t\varphi - B(f - t\varphi)(x)\|_2^2 + 2\langle f(x), -t\varphi(x) \\ &\quad - B(f - t\varphi)(x)\rangle, \end{aligned}$$

and if $x \in E(f - t\varphi - B(f - t\varphi))$ these are all equalities. Dividing by t and letting $t \to 0^+$ we get

$$A := \lim_{t\to 0^+} \frac{\|f - t\varphi - B(f - t\varphi)(x)\|^2 - 1}{t} = 2\langle f(y), -\varphi(y)\rangle, \quad y \in \{y_1, \ldots, y_l\},$$

and

$$A \geqslant 2\langle f(x), -\varphi(x)\rangle, \quad x \in E(f).$$

Since $x_{q+1} \in E(f)$ we then have $A \geqslant 4$. Since $l \geqslant 2$ there is a point y in $\{y_1, \ldots, y_l\}$ with $y \neq x_{q+1}$. Then $\varphi(y) = f(y)$ or $\varphi(y) = 0$ and so $A = 2$ or 0 and we have a contradiction.

Now we consider the situation where a reference can have cardinality 1 which can occur if $n < k$. There are two cases to consider.

Case I: At least one of the points in $E(f)$ is not a reference for f. Let $\{x_2\}$ be a point that is not a reference. Let $\varphi(x_1) = f(x_1)$, $\varphi(x_2) = -f(x_2)$, $\varphi(x_i) = 0$, $i = 3, \ldots, m$. Then $(f + t\varphi)(x_1) = (1+t) f(x_1)$, $(f + t\varphi)(x_2) = (1-t) f(x_2)$, $(f + t\varphi)(x_i) = f(x_i)$, $i = 3, \ldots, m$. Then if $t > 0$, $\|f + t\varphi\| = 1 + t$ and $\langle f(x_1) + t\varphi(x_1), g(x_1)\rangle = (1+t)\langle f(x_1), g(x_1)\rangle = 0$ for all $g \in G$. So $B(f + t\varphi) = 0$ if $t > 0$. Therefore $p(\varphi) = 0$ and

$$\lim_{t\to 0^-} \left(\|f + t\varphi - B(f + t\varphi)\|^2 - \|f\|^2\right)/t \leqslant 2\langle f(x_i), \varphi(x_i)\rangle$$

for every i. Therefore

$$\lim_{t\to 0^-} \left(\|f + t\varphi - B(f + t\varphi)\|^2 - \|f\|^2\right)/t = -2.$$

The consequence is that $E(f + t\varphi - B(f + t\varphi)) = \{x_2\}$ if $t < 0$ is sufficiently close to 0 which in turn implies $\{x_2\}$ is a reference for f. This contradiction shows that the best approximation operator is not Gateaux differentiable at f.

Case II: Every point in $E(f)$ is a reference for f. Note that

$$\langle f(x_i), g(x_i)\rangle = 0 \quad \text{for all } g \in G \text{ and for all } i. \tag{26}$$

Also note that for each i, the set $\left\{g_j(x_i) : j = 1, \ldots, n\right\}$ is a linearly independent set in R^k because every nontrivial element of G has no zeroes. Choose φ to satisfy

$$\langle f(x_2), \varphi(x_2)\rangle < \langle f(x_i), \varphi(x_i)\rangle < \langle f(x_1), \varphi(x_1)\rangle, \quad i = 3, \ldots, m. \tag{27}$$

Because of (25),

$$\lim_{t\to 0^+} \left(\|f + t\varphi - B(f + t\varphi)\|^2 - \|f\|^2\right)/t \geqslant 2\langle f(x_i), \varphi(x_i)\rangle \quad \text{for every } i. \tag{28}$$

Therefore because of (26), $E(f + t\varphi - B(f + t\varphi)) = \{x_1\}$ if $t > 0$ is sufficiently small. This means, using (25),

$$\langle \varphi(x_1) - p(\varphi)(x_1), g(x_1)\rangle = 0, \quad g \in G. \tag{29}$$

Also from (25),

$$\lim_{t\to 0^-} \left(\|f + t\varphi - B(f + t\varphi)\|^2 - \|f\|^2\right)/t \leqslant 2\langle f(x_i), \varphi(x_i)\rangle \quad \text{for every } i. \tag{30}$$

Therefore from (26), $E(f + t\varphi - B(f + t\varphi)) = \{x_2\}$ if $t < 0$ is sufficiently small. Therefore,

$$\langle \varphi(x_2) - p(\varphi)(x_2), g(x_2)\rangle = 0, \quad g \in G. \tag{31}$$

Let $p(\varphi) = \sum_{j=1}^{n} a_j g_j$. Let P_1 be the $n \times n$ matrix $\left(\langle g_k(x_1), g_j(x_1)\rangle\right)$ ($k=1, \ldots, n$, $j=1, \ldots, n$). P_1 is nonsingular because of the linear independence of $\left\{g_j(x_1) : j=1, \ldots, n\right\}$. Eq. (28) can be expressed as the linear system

$$P_1 \begin{pmatrix} a_1 \\ \cdot \\ \cdot \\ \cdot \\ a_n \end{pmatrix} = \begin{pmatrix} \langle \varphi(x_1), g_1(x_1)\rangle \\ \cdot \\ \cdot \\ \cdot \\ \langle \varphi(x_1), g_n(x_1)\rangle \end{pmatrix}. \tag{32}$$

In a similar way Eq. (30) can be expressed as the linear system

$$P_2 \begin{pmatrix} a_1 \\ \cdot \\ \cdot \\ \cdot \\ a_n \end{pmatrix} = \begin{pmatrix} \langle \varphi(x_2), g_1(x_2) \rangle \\ \cdot \\ \cdot \\ \cdot \\ \langle \varphi(x_2), g_n(x_2) \rangle \end{pmatrix}, \tag{33}$$

where P_2 is the $n \times n$ matrix $(\langle g_k(x_2), g_j(x_2) \rangle)$ $(k = 1, \ldots, n, \ j = 1, \ldots, n)$. Eq. (31) has a unique solution and therefore the left side of (32) is a fixed vector in R^n. All that is needed is to choose $\varphi(x_2)$ in such a way that $\langle \varphi(x_2), g_1(x_2) \rangle$ is different from the first component of the left side of (32) and still satisfies $\langle f(x_2), \varphi(x_2) \rangle < \langle f(x_1), \varphi(x_1) \rangle$. Clearly this is possible and so we again contradict the assumed Gateaux differentiability of the best approximation operator. □

We consider now the special case of approximation by complex constants.

Theorem 19. *Let G be the Haar space of constants in $C(X, R^2)$. Then the best approximation operator is Lipschitz continuous at every function in $C(X, R^2)$.*

Proof. Let $f \in C(X, R^2)$ and assume without loss of generality that $\|f\| = 1$ and $Bf = 0$. Assume $\|\varphi\| = 1$ and let $B(f + t\varphi) = (x(t), y(t))$. By Definition 6 of an annihilator and Theorem 8 it follows that if μ is an annihilator for f then $2 \leqslant \text{card}(\mu) \leqslant 3$. Let $F_t := f + t\varphi - B(f + t\varphi)$ and $\varphi(x_i) = (\varphi_{i1}, \varphi_{i2})$, for $x_i \in E(F_t)$, $i = 1, \ldots$. For t small enough it is easily shown that $E(F_t) \subseteq E(f)$ contains a reference for f. If that reference has cardinality 3 we are done since then Bf is strongly unique to f. Thus assume $\{x_1, x_2\} \subseteq E(F_t)$ is a reference for f. Since X is finite we can assume that for some sequence $\{t_j\}_{j=1}^{\infty}$ converging to zero, $\{x_1, x_2\} \subseteq E(F_{t_j})$. We know by the Kolmogorov criterion that

$$\max_{x_i \in E(F_t)} \langle F_t(x_i), (a, b) \rangle \geqslant 0, (a, b) \in G. \tag{34}$$

Since B is continuous at f [2], $\lim_{t\to 0} x(t) = \lim_{t\to 0} y(t) = 0$ and hence for some $\delta > 0$, $|x(t)| \leqslant 1$ and $|y(t)| \leqslant 1$ if $|t| < \delta$. By the definition of an annihilator there exist positive constants λ_1, λ_2 such that

$$\lambda_1 \langle f(x_1), g(x_1) \rangle + \lambda_2 \langle f(x_2), g(x_2) \rangle = 0, \quad g \in G. \tag{35}$$

Let $f(x_1) = (u, v)$ and $f(x_2) = (\overline{u}, \overline{v})$. Then letting g be $(1, 0)$ and then $(0, 1)$ in (34) gives $\overline{u} = (-\lambda_1/\lambda_2)\, u$ and $\overline{v} = (-\lambda_1/\lambda_2)\, v$. Since $\overline{u}^2 + \overline{v}^2 = 1$ we obtain $\lambda_1 = \lambda_2$ and so $\overline{u} = -u$ and $\overline{v} = -v$. Thus $f(x_1) = (u_1, v_1) = -f(x_2)$. Without loss of generality we may assume by rotation that $f(x_1) = (0, 1) = -f(x_2)$. For convenience now write $B(f + t\varphi) = (x(t), y(t)) = (x, y)$. We seek to solve

$$\min_{x,y} \max_{i>2} \Big\{ (x - t\varphi_{11})^2 + (y - (1 + t\varphi_{12}))^2, (x - t\varphi_{21})^2 + (y - (-1 + t\varphi_{22}))^2,$$
$$x - (p_i + t\varphi_{i1})^2 + (y - (g_i + t\varphi_{i2}))^2 \Big\},$$

where $f(x_i) = (p_i, q_i)$, $i > 2$ and $\mathrm{p}_i^2 + q_i^2 = 1$. Thus we seek the minimum of the maximum of

$$-2tx\varphi_{11} + t^2\varphi_{11}^2 - 2y - 2ty\varphi_{12} + 2t\varphi_{12} + t^2\varphi_{12}^2, \tag{36}$$

$$-2tx\varphi_{21} + t^2\varphi_{21}^2 + 2y - 2ty\varphi_{22} - 2t\varphi_{22} + t^2\varphi_{21}^2, \tag{37}$$

$$-2xp_i - 2xt\varphi_{i1} + 2p_it\varphi_{i1} + t^2\varphi_{i1}^2 - 2yq_i - 2yt\varphi_{i2} + 2q_it\varphi_{i2} + t^2\varphi_{i1}^2, \quad i \geqslant 3. \tag{38}$$

Now $x_1, x_2 \in E(F_t)$ implies (35) = (36) and solving that for y gives

$$y = \frac{t}{4}\Big(-2x\varphi_{11} + 2x\varphi_{21} + t\varphi_{11}^2 - 2y\varphi_{12} + 2\varphi_{12} \\ +t\varphi_{12}^2 - t\varphi_{31}^2 + 2y\varphi_{22} + 2\varphi_{22} - t\varphi_{22}\Big). \tag{39}$$

Using $|\varphi_{i1}| \leqslant \|\varphi\| \leqslant 1$ and $|x| \leqslant 1$ and $|y| \leqslant 1$ in (37) gives

$$|y| \leqslant 4|t|. \tag{40}$$

Now if $E(F_t) = \{x_1, x_2\}$, then letting $(a, b) = (1, 0)$ and then $(-1, 0)$ in (33) we find that

$$|x| \leqslant |t|. \tag{41}$$

Now assume there exists an $x_3 \in E(F_t) \setminus \{x_1, x_2\}$. From (36) = (37) with $i = 3$ in (39) we obtain

$$y(2q_3 - 2) = -2xp_3 + t\Big(2x\varphi_{11} - t\varphi_{11}^2 + 2y\varphi_{12} - 2\varphi_{12} - t\varphi_{12}^2 \\ -2x\varphi_{31} + 2p_3\varphi_{31} + t\varphi_{31}^2 - 2y\varphi_{32} + 2q_3\varphi_{32} + t\varphi_{32}^2\Big). \tag{42}$$

Thus using (38), $\|\varphi\| = 1$, $|x| \leqslant 1$, $|y| \leqslant 1$ and $|q_3| \leqslant 1$ in (40) we find that

$$|p_3 x| \leqslant 17|t|. \tag{43}$$

If $p_3 \neq 0$ then from (41),

$$|x| \leqslant (17/|p_3|)|t|. \tag{44}$$

Now suppose $p_3 = 0$ and $q_3 = 1$. Then in (41) cancel a t and let $t \downarrow 0$. This implies $\varphi_{12} = \varphi_{32}$ and hence $\varphi_{11} = \pm\varphi_{31}$. If $\varphi_{11} = \varphi_{31}$ then the point (x, y) equidistant from the three points $(f + t\varphi)(x_i)$, $i = 1$–3, i.e., satisfying (35) = (36) = (37) for $i = 3$ is the intersection of the perpendicular bisector of the sides of the triangle formed by $(f + t\varphi)(x_i)$, $i = 1$–3 if the points are distinct. When $\varphi_{11} = \varphi_{31}$, $(f + t\varphi)(x_1) = (f + t\varphi)(x_3) = (t\varphi_{11}, 1 + t\varphi_{12})$ and so $x = t\varphi_{11}$ and $|x| \leqslant |t|$. If $\varphi_{11} = -\varphi_{31}$, then $(f + t\varphi)(x_1) = (t\varphi_{11}, 1 + t\varphi_{12})$ and $(f + t\varphi)(x_3) = (-t\varphi_{11}, 1 + t\varphi_{12})$ and so $x = 0$. If $p_3 = 0$ and $q_3 = -1$ we similarly obtain the same result after setting (36) = (37) with $i = 3$. Finally then for any fixed annihilator $\{x_1, x_2\}$ for $f + t_j\varphi - B(f + t_j\varphi)$ we find

$$\|B(f + t_j\varphi) - Bf\| = \|(x(t_j), y(t_j))\|$$

satisfies $|y(t_j)| \leqslant 4|t|$ and

$$|x(t_j)| \leqslant \begin{cases} (17/|p_3|)|t| & \text{if } p_3 \neq 0, \\ |t|, & \text{if } p_3 = 0. \end{cases} \tag{45}$$

Since X is finite, when $f(x) = (p, q)$ either $p = 0$ or there exists an $\eta > 0$ such that $|p| \geqslant \eta > 0$. Thus in (44) there is an $\eta > 0$ such that $|p_3| \geqslant \eta > 0$ and thus there is a $\lambda > 0$ such that for $|t| \leqslant \delta$

$$\left\| B\left(f + t_j\varphi\right) - Bf \right\| \leqslant \lambda \left|t_j\right|.$$

Here δ depends on the sequence $\left\{t_j\right\}$. However, since there are only finitely many possible pairs $\{x_1, x_2\}$ we see that there is a $\delta > 0$ such that

$$\left\| B\left(f + t_j\varphi\right) - B(f) \right\| \leqslant \lambda |t|$$

if $|t| \leqslant \delta$. Since then B satisfies a local Lipschitz constant at f, it is Lipschitz continuous at f and the proof is complete. □

References

[1] J.R. Angelos, M.S. Henry, E.H. Kaufman, A. Kroo, T.D. Lenker, Local and global Lipschitz constants, J. Approx. Theory 46 (1986) 137–156.
[2] M. Bartelt, W. Li, Haar theory in vector-valued continuous function spaces, in: C.K. Chui, L.L. Schumaker (Eds.), Approximation Theory, VIII. vol. 1. Approximation and Interpolation, World Scientific, New York, 1995, pp. 39–46.
[3] M. Bartelt, W. Li, Abadie's constraint qualification, Hoffman's error bounds, and Hausdorff strong unicity, J. Approx. Theory 97 (1999) 140–157.
[4] H.P. Blatt, On strong uniqueness in linear complex Chebyshev approximation, J. Approx. Theory 41 (1984) 159–169.
[5] W. Cheney, Introduction to Approximation Theory, McGraw-Hill, New York, 1996.
[6] F. Deutsch, Best approximation in the space of continuous vector-valued functions, J. Approx. Theory 53 (1988) 112 –116.
[7] M. Gutknecht, Non-strong uniqueness in real and complex Chebyshev approximation, J. Approx. Theory 23 (1978) 204–213.
[8] A.V. Kolushov, Differentiability of the operator of best approximation, Math. Notes Acad. Sci. USSR 29 (1981) 295–306.
[9] A. Kroo, Differential properties of the operator of best approximation, Acta. Math. Hungar. 30 (1977) 319–331.
[10] D.J. Newman, H.S. Shapiro, Some theorems on Chebyshev approximation, Duke Math. J. 30 (1963) 673–681.
[11] R.T. Rockafellar, Convex Analysis, Princeton University Press, Princeton, NJ, 1970.
[12] W. Rudin, Real and Complex Analysis, McGraw-Hill, New York, 1966.
[13] E.L. Stiefel, Numerical methods of Tchebycheff approximation, in: R.E. Langer (Ed.), On Numerical Approximation, Madison, Wisconsin, 1959, pp. 217–232.
[14] S.I. Zukhovitskii, S.B. Stechkin, On approximation of abstract functions with values in Banach space, Dokl. Adad. Nauk. SSSR 106 (1956) 773–776.