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

Martin Bartelt
John Swetits
Old Dominion University
Lipschitz continuity of the best approximation operator in vector-valued Chebyshev approximation

Martin Bartelt\textsuperscript{a}, John Swetits\textsuperscript{b,}\textsuperscript{*}

\textsuperscript{a}Department of Mathematics, Christopher Newport University, Newport News, VA 23606, USA
\textsuperscript{b}Department of Mathematics and Statistics, Old Dominion University, Norfolk, VA 23529, USA

Received 1 August 2007; accepted 30 December 2007
Communicated by Günther Nürnberger
Available online 5 January 2008

Abstract

When $G$ is a finite dimensional Haar subspace of $C(X, R^k)$, the vector-valued continuous 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 $C(X, R^k)$ the best approximation operator satisfies the strong unicity condition of order 2 and a Lipschitz (Hölder) condition of order $\frac{1}{2}$. This note shows that in fact the best approximation operator satisfies the usual Lipschitz condition of order 1.

Keywords: Best approximation; Haar subspace; Chebyshev approximation

1. Introduction

Let $X$ be a compact Hausdorff space and $C(X, R^k)$ be the space of vector-valued continuous functions from $X$ to $k$-dimensional Euclidean space $R^k$. A natural norm for functions in $C(X, R^k)$ is defined as follows:

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

(1)

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

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

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

Let $P_G(f) := \{g \in G : \|f - g\| = \text{dist}(f, G)\}$.

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 $C(X, \mathbb{R}^2)$ using $f_1(x) + if_2(x) \leftrightarrow (f_1(x), f_2(x))$. 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 $\alpha$ if it is unique and if there exists a positive constant $\gamma(f, \alpha, G)$ such that

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

Zukhovitskii and Stechkin [14] (cf. also [2]) showed that there is a unique best approximation to every $f \in C(X, \mathbb{R}^k)$ if and only if $G$ satisfies the (generalized) Haar condition. 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 [9,10]. Cheney [7] 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 $\lambda$ such that

$$\|Bf - Bh\| \leq \lambda \|f - h\| \quad (4)$$

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

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

for all $h$ in $C(X, \mathbb{R}^k)$. Equivalently

$$\|Bf - Bh\| \leq \lambda \|f - h\|^\frac{1}{2} \quad (6)$$

for all $h$ in $C(X, \mathbb{R}^k)$ satisfying $\|f\| \leq M$ for some positive constant $M$. In approximation in $C(X, \mathbb{R}^k)$ and therefore in complex approximation, it is known [2] that $B$ satisfies a Hölder 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. Bartelt and Swetits [4] showed that the best approximation operator is Lipschitz continuous of order 1 on a dense subset of $C(X, \mathbb{R}^k)$ when $X$ is finite and $G$ is a Haar subspace of $C(X, \mathbb{R}^k)$ and they conjectured that the best approximation operator is globally Lipschitz continuous of order 1 in this case. They verified the conjecture in the case $k = 2$ when $G$ is the two dimensional subspace of constant vectors. The purpose of this paper is to establish the conjecture. One consequence is that best approximation in $C(X, \mathbb{R}^k)$ for $k \geq 2$ is fundamentally different from best approximation in $C(X, \mathbb{R})$, where Lipschitz continuity of order 1 and strong uniqueness of order 1 are essentially equivalent. Specifically Bartelt and Schmidt [5] established the following. Let $G$ be an $n$-dimensional subspace of $C(X, \mathbb{R})$, where $X$ is a compact Hausdorff
space. The metric projection, $P_G$, is said to be Lipschitz continuous of order 1 at $f$ if there is a constant $\lambda$ such that $H(P_G(f), P_G(h)) \leq \lambda \|f - h\|$ for all $h \in C(X, R)$, where $H$ denotes the Hausdorff metric. They showed that $f$ has a strongly unique of order 1 best approximation from $G$ if and only if $f$ has a unique best approximation from $G$ and $P_G$ is Lipschitz continuous of order 1 at $f$.

2. Results

As usual let the extreme point set be given by

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

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

**Definition 1.** An $n$-dimensional subspace $G$ in $C(X, R^k)$ 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 $mk < n \leq (m + 1)k$.

We need a characterization of the best approximate which is a generalization of the notion of a reference introduced by Stiefel [13] and Blatt [6] which is closely related (see Proposition 13) to the notion of an annihilator [2,8]. 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 $\lambda_i > 0, \quad i = 1, \ldots, q$. Let $\{g_1, \ldots, g_n\}$ denote a basis for $G$.

**Definition 2.** The collection $R = \{(x_i, S_i) : i = 1, \ldots, q\}$ is called a reference if the $q \times n$ matrix

$$B = ([S_i g_j(x_i), e_1])_{i=1,j=1}^{q,n}$$

has rank $q - 1$ and if there exists $\lambda \in R^q$, $\lambda > 0$, such that $\lambda^T B = 0$. Note that $q \leq n + 1$.

**Definition 3.** If $f \in C(X, R^k)$, 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 4.** 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 $g \in G$.

Recall the following characterization of best approximation [8].

**Theorem 5.** 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 $x_1, \ldots, x_q$, satisfying $\|f(x_i) - h(x_i)\|_2 = \|f - h\|$ and an annihilator $\sigma$ of $G$ satisfying $\frac{\sigma(x_i)}{\|\sigma(x_i)\|_2} = \frac{f(x_i) - h(x_i)}{\|f - h\|}$, $i = 1, \ldots, q$, where $q \leq n + 1$.

Call the points $x_1, \ldots, x_q$ an annihilator or the support of an annihilator for $f - Bf$. We then have the following characterization of best approximation. The proof is in [4].
Theorem 6. A function \( g \in G \) is a best approximation to \( f \in C(\mathbb{X}, \mathbb{R}^k) \setminus G \) if and only if there exists a reference \( R \) with respect to \( f - g \).

The following theorem [4] shows that there is a particular set of functions in \( C(\mathbb{X}, \mathbb{R}^k) \) at which \( B \), by the result of Cheney, has Lipschitz continuity of order 1.

Theorem 7. 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.

The next result clarifies the relationship between a reference and an annihilator and also provides an alternative characterization of a reference. The proof is in [4].

Theorem 8. Suppose \( \{x_1, \ldots, x_q\} \subseteq E(f - Bf) \). The following are equivalent.

(i) \( \{x_i, S_i\} : i = 1, \ldots, q \) is a reference with respect to \( f - Bf \).

(ii) \( \{x_1, \ldots, x_q\} \) is the support of an annihilator and no proper subset is the support of an annihilator.

(iii) The matrix \( M := M(x_1, \ldots, x_q) := ([f(x_i) - Bf(x_i), g_j(x_i)])_{i=1,j=1}^{q,n} \) has rank \( q - 1 \) and there exists \( \lambda \in \mathbb{R}^q \), \( \lambda > 0 \), such that \( \lambda^T M = 0 \).

Remark 9. For brevity we will refer to \( \{x_1, \ldots, x_q\} \) as a reference.

We can now state and prove the main result of this paper.

Theorem 10. Let \( X \) be a finite set with the discrete topology and let \( G \) be an \( n \)-dimensional Haar subspace of \( C(X, \mathbb{R}^k) \). Then the best approximation operator, \( B : C(X, \mathbb{R}^k) \to G \), is pointwise Lipschitz continuous of order 1.

Proof. The proof is by contradiction. Assume there is a function \( f \in C(X, \mathbb{R}^k) \setminus G \) such that \( B \) is not Lipschitz continuous of order 1 at \( f \). We can assume \( \|f\| = 1 \) and \( Bf = 0 \). Then there exists a sequence \( \{\varphi_j\} \subseteq C(X, \mathbb{R}^k) \) and a sequence \( \{t_j\} \) of positive reals such that \( \|\varphi_j\| = 1 \), \( j = 1, 2, \ldots \), \( \lim_{j \to \infty} t_j = 0 \) and \( \lim_{j \to \infty} \frac{B(f + t_j \varphi_j)}{t_j} = \infty \). For simplicity let \( f_j = f + t_j \varphi_j \).

The unit sphere in \( C(X, \mathbb{R}^k) \) is compact and so we can assume there exists \( \varphi \in C(X, \mathbb{R}^k) \) such that \( \{\varphi_j\} \) converges to \( \varphi \).

Because \( X \) is finite, there exists \( \delta > 0 \) such that if \( h \in C(X, \mathbb{R}^k) \) and \( \|f - h\| < \delta \), then \( E(h - Bh) \subseteq E(f) \). We can assume \( 0 < t_j < \delta \), \( j = 1, 2, \ldots \), and we can also assume, by passing to a subsequence if necessary, that there is \( \{x_1, \ldots, x_m\} \subseteq E(f) \) which is a reference with respect to \( f_j - B(f_j) \) for every \( j = 1, 2, \ldots \). Thus for \( j = 1, 2, \ldots \), there are positive scalars \( \lambda_i(j), i = 1, \ldots, m \), such that \( \sum_{i=1}^{m} \lambda_i(j) = 1 \) and

\[
\sum_{i=1}^{m} \lambda_i(j) [f_j(x_i) - B(f_j)(x_i), g(x_i)] = 0 \quad (7)
\]

for all \( g \in G \). For each \( i = 1, \ldots, m \) the sequence \( \{\lambda_i(j)\} \) is a bounded sequence of positive reals and so we can assume \( \lim_{j \to \infty} \lambda_i(j) = \lambda_i > 0 \) exists. Because \( \{f_j\} \) converges to \( f \) and
\( \{ B(f_j) \} \) converges to \( Bf = 0 \), it follows from (7) that
\[
\sum_{i=1}^{m} \lambda_i \langle f(x_i), g(x_i) \rangle = 0
\]
(8)
for all \( g \in G \). Therefore \( \{ x_1, \ldots, x_m \} \) contains a reference with respect to \( f \). We can assume the reference is \( \{ x_1, \ldots, x_q \} \), \( \lambda_i > 0 \) for \( i = 1, \ldots, q \), and \( \sum_{i=1}^{m} \lambda_i = 1 \). Because \( B \) is Lipschitz continuous of order \( \frac{1}{2} \) on bounded sets we can assume there exists \( p \in G \) such that
\[
\lim_{j \to \infty} B(f_j) \frac{1}{\sqrt{t_j}} = p.
\]
(9)
Choose \( g = Bf_j \) in (7) and divide through by \( t_j \). It then follows that
\[
\lim_{j \to \infty} \frac{1}{t_j} \sum_{i=1}^{m} \lambda_i (j) \left( \langle f(x_i), B(f_j)(x_i) \rangle = \sum_{i=1}^{m} \lambda_i \| p(x_i) \|_2^2. \right.
\]
(10)
Now choose \( g = p \) in (7) and divide through by \( \sqrt{t_j} \) to obtain
\[
\lim_{j \to \infty} \frac{1}{\sqrt{t_j}} \sum_{i=1}^{m} \lambda_i (j) \left( \langle f(x_i), p(x_i) \rangle = \sum_{i=1}^{m} \lambda_i \| p(x_i) \|_2^2. \right.
\]
(11)
We now consider the following identity. For \( i = 1, \ldots, m \)
\[
\| f_j - B(f_j) \|^2 - \| f \|^2 = \| f_j(x_i) - B(f_j)(x_i) \|^2 - \| f(x_i) \|^2
\]
\[
= 2 \langle f(x_i), t_j \varphi_j(x_i) - B(f_j)(x_i) \rangle
\]
\[
+ \| t_j \varphi_j(x_i) - B(f_j)(x_i) \|^2.
\]
(12)
In (12) divide through by \( t_j \), multiply by \( \lambda_i (j) \) and sum from \( i = 1 \) to \( m \). From (10) it follows that
\[
\lim_{j \to \infty} \frac{\| f_j - B(f_j) \|^2 - \| f \|^2}{t_j} = \sum_{i=1}^{m} 2 \lambda_i \langle f(x_i), \varphi(x_i) \rangle - \sum_{i=1}^{m} \lambda_i \| p(x_i) \|_2^2.
\]
(13)
From (13) it follows that
\[
\lim_{j \to \infty} \frac{\| f_j - B(f_j) \|^2 - \| f \|^2}{\sqrt{t_j}} = 0
\]
and from (9) it follows that
\[
\lim_{j \to \infty} \frac{B(f_j)}{t_j^{1/4}} = 0.
\]
In (12) divide through by \( \sqrt{t_j} \). It then follows for \( i = 1, \ldots, m \) that
\[
\lim_{j \to \infty} \left( f(x_i), B(f_j) \frac{1}{\sqrt{t_j}} \right) = \langle f(x_i), p(x_i) \rangle = 0.
\]
Therefore from (11) it follows that \( p(x_i) = 0, i = 1, \ldots, q \). Since \( G \) is Haar we obtain \( p \equiv 0 \).
We can assume there exists \( \hat{g} \in G, \| \hat{g} \| = 1 \), such that \( \lim_{j \to \infty} \frac{B(f_j)}{\|B(f_j)\|} = \hat{g} \). From (12) and (13) it follows that \( \lim_{j \to \infty} \left\langle f(x_i), \frac{B(f_j)(x_i)}{t_j} \right\rangle \) exists for each \( i = 1, \ldots, m \). Because

\[
\lim_{j \to \infty} \frac{\|B(f + t_j \varphi_j)\|}{t_j} = \infty, \tag{14}
\]

it then follows that

\[
\lim_{j \to \infty} \left\langle f(x_i), \frac{B(f_j)(x_i)}{\|B(f_j)\|} \right\rangle = \left\langle f(x_i), \hat{g}(x_i) \right\rangle = 0, \quad i = 1, \ldots, m. \tag{15}
\]

In (7) choose \( g = \hat{g} \) and use (15) to obtain

\[
\lim_{j \to \infty} \sum_{i=1}^{m} \lambda_i(j) \left\langle \frac{B(f_j)(x_i)}{\|B(f_j)\|}, \hat{g}(x_i) \right\rangle = \sum_{i=1}^{m} \lambda_i \left\langle \varphi(x_i), \hat{g}(x_i) \right\rangle. \tag{16}
\]

It then follows from (14) and (16) that

\[
\lim_{j \to \infty} \sum_{i=1}^{m} \lambda_i(j) \left\langle \frac{B(f_j)(x_i)}{\|B(f_j)\|}, \hat{g}(x_i) \right\rangle = \sum_{i=1}^{m} \lambda_i \| \hat{g}(x_i) \|^2_2 = 0.
\]

Therefore \( \hat{g}(x_i) = 0, \; i = 1, \ldots, q \). Because \( G \) is Haar it follows that \( \hat{g} \equiv 0 \) which contradicts \( \| \hat{g} \| = 1 \). This contradiction establishes the result. \( \square \)

References