JOURNAL OF APPROXIMATION THEORY 75, 175–182 (1993)

A Characterization of Best Simultaneous Approximations

G. A. WATSON

Department of Mathematics and Computer Science, University of Dundee, Dundee DD14HN, Scotland

Communicated by Will Light

Received September 11, 1991; accepted in revised form August 3, 1992

The problem of approximating a finite number of functions simultaneously is considered. For a general class of norms, a characterization of best approximations is given. The result generalizes recent work concerned specifically with the Chebyshev norm.

1. INTRODUCTION

Let X be a compact Hausdorff space and Y a normed linear space with norm $\|\cdot\|_Y$. Let $C(X, Y)$ denote the set of all continuous functions from X to Y, and let $\|\cdot\|_A$ be a norm on $C(X, Y)$. Define a norm on l-tuples of elements of $C(X, Y)$ as follows: for any $\phi_1, ..., \phi_l$ in $C(X, Y)$ let

$$\|(\phi_1, ..., \phi_l)\| = \max_{\|\mathbf{a}\|_B = 1} \left\| \sum_{i=1}^{l} a_i \phi_i \right\|_A, \tag{1.1}$$

where $\|\cdot\|_B$ is a given norm on $\mathbb{R}^l$, and $\mathbf{a} = (a_1, ..., a_l)^T$. For convenience the left-hand side of (1.1) is abbreviated as $\|\boldsymbol{\phi}\|$.

Now suppose that functions $F_1, ..., F_l$ in $C(X, Y)$ are given. Then the problem we consider here is approximating these functions simultaneously by functions in S, an n-dimensional subspace of $C(X, Y)$, in the sense of the minimization of the norm (1.1). In other words, we want to find $f \in S$ to minimize

$$\|(F_1 - f, ..., F_l - f)\|. \tag{1.2}$$

If such a function f^* exists, it is called a best simultaneous approximation to $F_1, ..., F_l$. Problems of simultaneous approximation have attracted much

640 75 2-5

0021-9045/93 $5.00

interest over many years. Of particular concern here is a characterization of best approximations for the norm (1.1): the results generalize to a wide class of norms a theorem of Tanimoto [3].

2. The Main Result

It is necessary to introduce the subdifferential or set of subgradients of $\|\cdot\|_A$ at any element of $C(X, Y)$. This is the set defined for any convex function g on $C(X, Y)$ at the point f by

$$\partial g(f)=\{w\in C^*(X, Y): g(F)\geqslant g(f)+\langle w, F-f\rangle \text{ for all } F\in C(X, Y)\}$$

(see, for example, Rockafellar [2]), where the usual inner product notation is used to link elements of $C(X, Y)$ and its dual $C^*(X, Y)$. For norms this set has a very convenient characterization: in particular if $f\in C(X, Y)$ it is readily established that $w\in\partial\|f\|_A$ if and only if

(i) $\langle w, f\rangle=\|f\|_A$,

(ii) $\|w\|_A^*\leqslant 1$, where $\|\cdot\|_A^*$ denotes the dual norm

$$\|w\|_A^*=\max_{\|f\|_A\leqslant 1}\langle w, f\rangle.$$

For a given l-tuple $\boldsymbol{\phi}=(\phi_1, ..., \phi_l)$ of functions in $C(X, Y)$, define the set

$$G(\boldsymbol{\phi})=\left\{(\mathbf{a}, w): \mathbf{a}\in\mathbb{R}^l, \|\mathbf{a}\|_B=1, \sum_{i=1}^{l} a_i\phi_i=\|\boldsymbol{\phi}\|\, u, \|u\|_A=1, w\in\partial\|u\|_A\right\}. \tag{2.1}$$

Note that if $(\mathbf{a}, w)\in G(\boldsymbol{\phi})$, then $\mathbf{a}\in\mathbb{R}^l$ is an element where the norm on the right-hand side of (1.1) is attained. Then a key result is the following characterization of the directional derivative of $\|\boldsymbol{\phi}\|$ in the direction $\boldsymbol{\psi}$. The following lemma generalizes in an obvious manner a result for matrices given in [4].

LEMMA 1. *Let $\phi_1, ..., \phi_l, \psi_1, ..., \psi_l$ be any elements in $C(X, Y)$. Then*

$$\lim_{t\to 0+}\frac{\|\boldsymbol{\phi}+t\boldsymbol{\psi}\|-\|\boldsymbol{\phi}\|}{t}=\max_{(\mathbf{a}, w)\in G(\boldsymbol{\phi})}\left\langle w, \sum_{i=1}^{l} a_i\psi_i\right\rangle.$$

Proof. For any real t,

$$\begin{aligned}
\|\boldsymbol{\phi}\| &= \max_{\|\mathbf{a}\|_B = 1} \left\| \sum_{i=1}^{l} a_i \phi_i \right\|_A \\
&\geqslant \left\| \sum_{i=1}^{l} a_i(t)\, \phi_i \right\|_A \\
&\geqslant \left\langle w(t), \sum_{i=1}^{l} a_i(t) \phi_i \right\rangle \qquad \text{for any} \quad (\mathbf{a}(t), w(t)) \in G(\boldsymbol{\phi} + t\boldsymbol{\psi}), \\
&= \left\langle w(t), \sum_{i=1}^{l} a_i(t)(\phi_i + t\psi_i) - t \sum_{i=1}^{l} a_i(t) \psi_i \right\rangle \\
&= \|\boldsymbol{\phi} + t\boldsymbol{\psi}\| - t \left\langle w(t), \sum_{i=1}^{l} a_i(t) \psi_i \right\rangle.
\end{aligned}$$

Also,

$$\begin{aligned}
\|\boldsymbol{\phi} + t\boldsymbol{\psi}\| &\geqslant \left\| \sum_{i=1}^{l} a_i (\phi_i + t\psi_i) \right\|_A \\
&\geqslant \left\langle w, \sum_{i=1}^{l} a_i(\phi_i + t\psi_i) \right\rangle \qquad \text{for any} \quad (\mathbf{a}, w) \in G(\boldsymbol{\phi}) \\
&= \|\boldsymbol{\phi}\| + t \left\langle w, \sum_{i=1}^{l} a_i \psi_i \right\rangle.
\end{aligned}$$

It follows that for all $t > 0$, and all $(\mathbf{a}, w) \in G(\boldsymbol{\phi})$, $(\mathbf{a}(t), w(t)) \in G(\boldsymbol{\phi} + t\boldsymbol{\psi})$,

$$\left\langle w, \sum_{i=1}^{l} a_i \psi_i \right\rangle \leqslant \frac{\|\boldsymbol{\phi} + t\boldsymbol{\psi}\| - \|\boldsymbol{\phi}\|}{t} \leqslant \left\langle w(t), \sum_{i=1}^{l} a_i(t) \psi_i \right\rangle.$$

If one lets t tend to zero, and uses the weak * compactness of the unit ball in the dual space (the Alaoglu–Bourbaki theorem, for example, Holmes [1]) the result follows. ∎

Now let

$$\phi_i(f) = F_i - f, \qquad i = 1, ..., l, \qquad \text{for all} \quad f \in S,$$

and let $H(f)$ denote the set of l-tuples $\{(h_1, ..., h_l)\}$ of elements in $C^*(X, Y)$ defined by

$$H(f) = \operatorname{conv}\{(a_1 w, ..., a_l w), (\mathbf{a}, w) \in G(\boldsymbol{\phi}(f))\}, \tag{2.2}$$

where as usual "conv" is used to denote the convex hull. Note that

$$\sum_{i=1}^{l} \langle h_i, \phi_i(f) \rangle = \|\boldsymbol{\phi}(f)\|, \qquad \text{for all} \quad \mathbf{h} \in H(f). \tag{2.3}$$

THEOREM 1. *$f^* \in S$ minimizes (1.2) if and only if there exists $\mathbf{h} = (h_1, ..., h_l) \in H(f^*)$ such that*

$$\sum_{i=1}^{l} \langle h_i, f \rangle = 0 \qquad \text{for all} \quad f \in S. \tag{2.4}$$

Proof. Let f^* minimize (1.2) but let the stated condition not hold. Let $\{b_1, ..., b_n\}$ be a basis for S, and define

$$D = \left\{ \mathbf{d} \in \mathbb{R}^n : d_j = \sum_{i=1}^{l} \langle h_i, b_j \rangle, j = 1, ..., n, \text{ for all } \mathbf{h} \in H(f^*) \right\}.$$

Then

$$D = \text{conv}(R),$$

where

$$R = \left\{ \mathbf{r} \in \mathbb{R}^n : r_j = \sum_{i=1}^{l} a_i \langle w, b_j \rangle, j = 1, ..., n, \text{ for all } (\mathbf{a}, w) \in G(\boldsymbol{\phi}(f^*)) \right\}.$$

Let $\mathbf{r}^{(k)} \in R$, $k = 1, 2, ...,$ so that

$$r_j^{(k)} = \sum_{i=1}^{l} a_i^{(k)} \langle w^{(k)}, b_j \rangle, \quad j = 1, ..., n, \qquad \text{where} \quad (\mathbf{a}^{(k)}, w^{(k)}) \in G(\boldsymbol{\phi}(f^*)).$$

Then there is a subsequence (not renamed) so that for some $\bar{w}$, $\|\bar{w}\|_A^* = 1$, and $\bar{\mathbf{a}}$, $\|\bar{\mathbf{a}}\|_B = 1$,

$$\langle w^{(k)}, F \rangle \to \langle \bar{w}, F \rangle, \qquad \text{as} \quad k \to \infty, \qquad \text{for all} \quad F \in C(X, Y),$$

using the Alaoglu–Bourbaki theorem, and

$$\mathbf{a}^{(k)} \to \bar{\mathbf{a}}, \qquad \text{as} \quad k \to \infty.$$

Thus

$$r_j^{(k)} \to \sum_{i=1}^{l} \bar{a}_i \langle \bar{w}, b_j \rangle, \qquad \text{as} \quad k \to \infty, \quad j = 1, ..., n.$$

Let $u^{(k)}$, $\bar{u}$ be defined by

$$\sum_{i=1}^{l} a_i^{(k)}\phi_i = \|\boldsymbol{\phi}\|\, u^{(k)}, \qquad \|u^{(k)}\|_A = 1, \qquad k = 1, 2, ...,$$

$$\sum_{i=1}^{l} \bar{a}_i\phi_i = \|\boldsymbol{\phi}\|\, \bar{u}, \qquad \|\bar{u}\|_A = 1.$$

Then

$$\langle w^{(k)}, \bar{u}\rangle = 1 + \langle w^{(k)}, \bar{u} - u^{(k)}\rangle, \qquad k = 1, 2, ...$$

If one lets $k \to \infty$, it follows that

$$\langle \bar{w}, \bar{u}\rangle = 1,$$

so that $\bar{w} \in \partial \|\bar{u}\|_A^*$. Thus $(\bar{\mathbf{a}}, \bar{w}) \in G(\boldsymbol{\phi}(f^*))$, and so R is closed. It follows that D is a closed, convex set in $\mathbb{R}^n$ which does not contain the origin, and so by a standard separation result there exists $\mathbf{c} \in \mathbb{R}^n$ such that

$$\mathbf{c}^T\mathbf{d} < 0, \qquad \text{for all} \quad \mathbf{d} \in D.$$

Thus there exists $f \in S$ such that

$$\sum_{i=1}^{l} \langle h_i, f\rangle < 0, \qquad \text{for all} \quad \mathbf{h} \in H(f^*),$$

or

$$\left\langle \sum_{i=1}^{l} a_i w, f \right\rangle < 0, \qquad \text{for all} \quad (\mathbf{a}, w) \in G(\boldsymbol{\phi}(f^*)).$$

Using Lemma 1, with $\psi_i = f$, $i = 1, ..., l$, this contradicts the fact that f^* minimizes (1.2), and establishes the necessity of the conditions (2.4).

Now let these conditions be satisfied at f^* and let f be any element of S. Then

$$\|\boldsymbol{\phi}(f)\| = \max_{\|\mathbf{a}\|_B = 1} \left\| \sum_{i=1}^{l} a_i(F_i - f) \right\|_A$$

$$\geqslant \left\langle w, \sum_{i=1}^{l} a_i(F_i - f) \right\rangle, \qquad \text{for all} \quad (\mathbf{a}, w) \in G(\boldsymbol{\phi}(f^*)).$$

Suppose that $\mathbf{h} \in H(f^*)$ satisfies (2.4). Then

$$\begin{aligned}\|\boldsymbol{\phi}(f)\| &\geqslant \sum_{i=1}^{l} \langle h_i, F_i - f\rangle \\ &= \sum_{i=1}^{l} \langle h_i, F_i - f^*\rangle, \qquad \text{using (2.4)}, \\ &= \|\boldsymbol{\phi}(f^*)\|, \qquad \text{using (2.3)}.\end{aligned}$$

The proof is complete. ∎

3. Simultaneous Chebyshev Approximation

For any l-tuple ϕ_i, $i = 1, ..., l$, in $C(X, Y)$, a Chebyshev norm may be defined by

$$\max_{1 \leqslant i \leqslant l} \max_{x \in X} \|\phi_i(x)\|_Y. \tag{3.1}$$

Now

$$\begin{aligned}\max_{1 \leqslant i \leqslant l} \max_{x \in X} \|\phi_i(x)\|_Y &\leqslant \max_{\|\mathbf{a}\|_1 = 1} \max_{x \in X} \left\| \sum_{j=1}^{l} a_j \phi_j(x) \right\|_Y \\ &\leqslant \max_{\|\mathbf{a}\|_1 = 1} \max_{x \in X} \|\mathbf{a}\|_1 \max_{1 \leqslant j \leqslant l} \|\phi_j(x)\|_Y \\ &\qquad \text{using Hölder's inequality} \\ &= \max_{1 \leqslant j \leqslant l} \max_{x \in X} \|\phi_j(x)\|_Y.\end{aligned}$$

Therefore a Chebyshev norm on l-tuples is given by (1.1) when $\|\cdot\|_A$ is the Chebyshev norm on $C(X, Y)$, and $\|\cdot\|_B$ is the l_1 norm. The minimization of (1.2) in this case is the problem considered in [3], and the same connection is made between (3.1) and (1.1) except that use is made there of the additional fact that the components of $\mathbf{a}$ may be restricted to be nonnegative. To illustrate the use of Theorem 1, we extract the result for this norm. For $f \in S$, and any $\mathbf{a} \in \mathbb{R}^l$ with $\|\mathbf{a}\|_B = 1$, define

$$E(\mathbf{a}, f) = \left\{ x \in X : \left\| \sum_{i=1}^{l} a_i \phi_i(f(x)) \right\|_Y = \|\boldsymbol{\phi}(f)\| \right\},$$

and let

$$E(f) = \left\{ x \in X : \left\| \sum_{i=1}^{l} a_i \phi_i(f(x)) \right\|_Y = \|\boldsymbol{\phi}(f)\| \text{ for some } \mathbf{a},\ \|\mathbf{a}\|_B = 1 \right\}.$$

Then for each distinct $x^k \in E(f)$, there exists $\mathbf{a}^k$, $\|\mathbf{a}^k\|_B = 1$ (not necessarily all distinct), such that

$$x^k \in E(\mathbf{a}^k, f). \tag{3.2}$$

To use the theorem, we need the form of the subdifferential of the Chebyshev norm on $C(X, Y)$ in this case. In particular, $(\mathbf{a}, w) \in G(\boldsymbol{\phi}(f))$ implies that

$$w \in \operatorname{conv}\{v(x)\,\delta(x), \text{ for all } x \in E(\mathbf{a}, f)\},$$

where δ is the delta function, and $v(x) \in \partial \|\sum_{i=1}^{l} a_i \phi_i(f(x))\|_Y$. Thus

$$\langle w, f \rangle = \sum_{k=1}^{s} \mu_k \langle v(x^k), f(x^k) \rangle_Y,$$

where $\langle \cdot, \cdot \rangle_Y$ denotes an inner product between elements of Y and its dual, where $x^k \in E(\mathbf{a}, f)$ and μ_k, $k = 1, \ldots, s$, are nonnegative numbers summing to 1. Using this expression, and Caratheodory's theorem, it is readily verified that the conditions of Theorem 1 specialize to the following.

THEOREM 2. *$f^* \in S$ minimizes (1.2) if and only if there exist $m \leqslant n+1$ distinct elements $x^1, \ldots, x^m$ of $E(f^*)$, m vectors $\mathbf{a}^1, \ldots, \mathbf{a}^m$ satisfying (3.2) with $f = f^*$, and m positive numbers $\gamma_1, \ldots, \gamma_m$ summing to 1 such that*

$$\sum_{k=1}^{m} \gamma_k \langle v_k(x^k), f(x^k) \rangle_Y = 0, \qquad \textit{for all} \quad f \in S, \tag{3.3}$$

where $v_k(x) \in \partial \|\sum_{i=1}^{l} a_i^k \phi_i(f^*(x))\|_Y$, $k = 1, \ldots, m$.

The form given by Tanimoto [3] has (3.3) replaced by

$$\sum_{k=1}^{m} \gamma_k \left\| \sum_{i=1}^{l} a_i^k \phi_i(f^*(x^k)) \right\|_Y \leqslant \sum_{k=1}^{m} \gamma_k \left\| \sum_{i=1}^{l} a_i^k \phi_i(f(x^k)) \right\|_Y, \qquad \text{for all} \quad f \in S. \tag{3.4}$$

It is easy to verify that these conditions are sufficient. We show that they are implied by (3.3). Let $f \in S$ be arbitrary, and let $v_k(x) \in \partial \|\sum_{i=1}^{l} a_i^k \phi_i(f^*(x))\|_Y$, $k = 1, \ldots, m$, satisfy (3.3). Then the left-hand side of (3.4) is just

$$\begin{aligned}
&\sum_{k=1}^{m} \gamma_k \left\langle v_k(x^k), \sum_{i=1}^{l} a_i^k \phi_i(f^*(x^k)) \right\rangle_Y \\
&\quad = \sum_{k=1}^{m} \gamma_k \left\langle v_k(x^k), \sum_{i=1}^{l} a_i^k (F_i(x^k) - f^*(x^k)) \right\rangle_Y \\
&\quad = \sum_{k=1}^{m} \gamma_k \left\langle v_k(x^k), \sum_{i=1}^{l} a_i^k (F_i(x^k) - f(x^k)) \right\rangle_Y, \qquad \text{by (3.3)}, \\
&\quad \leqslant \sum_{k=1}^{m} \gamma_k \left\| \sum_{i=1}^{l} a_i^k \phi_i(f(x^k)) \right\|_Y,
\end{aligned}$$

which gives the required result.

References

1. R. B. Holmes, "Geometric Functional Analysis and Its Applications," Springer-Verlag, New York, 1975.
2. R. T. Rockafellar, "Convex Analysis," Princeton Univ. Press, Princeton, NJ, 1970.
3. S. Tanimoto, A characterization of best simultaneous approximations, *J. Approx. Theory* **59** (1989), 359–361.
4. G. A. Watson, Characterization of the subdifferential of some matrix norms, *Linear Algebra Appl.* **170** (1992), 33–45.