JOURNAL OF APPROXIMATION THEORY 10, 49–63 (1974)

On Vectorial Approximation

WILLIAM B. GEARHART[1]

Department of Mathematics, University of Utah, Salt Lake City, Utah 84112

Communicated by T. J. Rivlin

Received November 18, 1971

1. INTRODUCTION

We shall consider in this paper the problem of determining an approximation to a function, when it is desired to use more than one criterion of approximation simultaneously. Such problems have been considered before. For example, the problem of approximating a function in the Chebyshev norm, where the approximation is required to satisfy certain generalized interpolation conditions was investigated by Brosowski [6]. Laurent [12] has approached the problem of simultaneously approximating a function f and its first derivative f' by minimizing the error in the norm

$$\|f\|_1 = \max_{x\in[a,b]} |f(x)| + \nu \max_{x\in[a,b]} |f'(x)|,$$

where ν is some positive constant. A generalized Remez algorithm is given in [12] for computing the best approximations in this norm. Further results concerning the approximation of a function and its derivatives were obtained by Moursund [14, 16] and Moursund and Stroud [15]. In [5], Bredendiek studied a more general situation. Here the approach is to find an approximation g from some subspace which minimizes

$$\max\{\|f - g\|, |A(f) - A(g)|\},$$

where $\| \ \|$ and $| \ |$ are norms, and A is an operator which is not necessarily linear.

Recently, another approach called vectorial approximation was introduced Bacopoulos [1]. Results on this topic can be found in [1–4]. In computational work, we have found that vectorial approximation offers an advantage over techniques such as those mentioned above (see Section 4). Thus, we consider vectorial approximation here.

The setting for our work will be the following. Let D be a compact

[1] Present address: University of Texas, San Antonio, Texas 78284.

Hausdorff space, and let $C(D)$ denote the space of continuous real-valued functions on D, with the sup norm,

$$\|f\| = \sup\{|f(x)|: x \in D\}.$$

Suppose linear operators $L_1, L_2, ..., L_k$ are given, which map a subspace X of $C(D)$ into $C(D)$. For each $f \in X$, define the vector $V(f)$ in k-space by

$$V(f) = (\| L_1(f)\|, \| L_2(f)\|, ..., \| L_k(f)\|)^T.$$

The symbol $\leqslant$ will be used to designate the usual component-wise partial ordering of k-space.

Now, let $U \subset X$ be a linear subspace of dimension n, and let $f \in X$ be given. Then $u_* \in U$ is said to be a best vectorial approximation to f out of U if there is no $u \in U$ such that $V(f - u) \leqslant V(f - u_*)$ and $V(f - u) \neq V(f - u_*)$. In general, there are many best vectorial approximations, and this fact complicates the task of computation.

We remark in passing that a more general form of vectorial approximation was studied by Bacopoulos [4].

The case $k = 2$ is of some interest, and we consider this situation in detail. Results which characterize best vectorial approximations in this case are presented, and the question of uniqueness is discussed. Moreover, the characterization given in Section 2 leads immediately to a scheme for computing the vectorial approximations. It will be seen in the last section that similar computing schemes can be used to obtain best vectorial approximations for $k \geqslant 3$. Finally, the discretization error for the method of Section 2 is analyzed.

2. Vectorial Approximations

In the main result of this section, it is shown that all best vectorial approximations can be obtained as solutions of a certain computational problem.

Let $f \in X$ be given, and define

$$\mu_i = \inf\{\| L_i(f - u)\|: u \in U\}, \qquad i = 1, 2.$$

For $t \in [\mu_1, \infty)$, set

$$\lambda(t) = \inf\{\| L_2(f - u)\|: u \in U, \| L_1(f - u)\| \leqslant t\},$$

and for $t \in [\mu_2, \infty)$, set

$$\rho(t) = \inf\{\| L_1(f - u)\|: u \in U, \| L_2(f - u)\| \leqslant t\}.$$

By a familiar compactness argument, there exist $u_i \in U$ such that $\mu_i = \| L_i(f - u_i)\|$ for $i = 1, 2$. Similarly, for each $t \in [\mu_1, \infty)$, there exists $u_t \in U$ such that $\| L_1(f - u_t)\| \leqslant t$ and $\| L_2(f - u_t)\| = \lambda(t)$. Also, the analogous statement holds for the function $\rho(t)$.

LEMMA 2.1. *If $u \in U$ is a best vectorial approximation, then*

$$\mu_1 \leqslant \| L_1(f - u)\| \leqslant \rho(\mu_2)$$

and

$$\mu_2 \leqslant \| L_2(f - u)\| \leqslant \lambda(\mu_1).$$

Proof. The inequalities on the left side follow by definition. Suppose $\| L_1(f - u)\| > \rho(\mu_2)$. Let $\tilde{u} \in U$ satisfy $\| L_1(f - \tilde{u})\| = \rho(\mu_2)$ and $\| L_2(f - \tilde{u})\| = \mu_2$. Then $\tilde{u}$ is a better vectorial approximation than u, which is a contradiction. A similar argument establishes the inequality $\| L_2(f - u)\| \leqslant \lambda(\mu_1)$.

LEMMA 2.2. *The function $\lambda(t)$ is convex on $[\mu_1, \infty)$, decreasing on $[\mu_1, \rho(\mu_2)]$, and $\lambda(t) = \mu_2$ for all $t \geqslant \rho(\mu_2)$. Similarly, $\rho(t)$ is convex on $[\mu_2, \infty)$, decreasing on $[\mu_2, \lambda(\mu_1)]$, and $\rho(t) = \mu_1$ for $t \geqslant \lambda(\mu_1)$.*

Proof. Only the proof for $\lambda(t)$ will be given, as the proof for $\rho(t)$ is similar. We begin by showing the λ is convex. Suppose $t_1, t_2 \in [\mu_1, \infty)$, and let $\alpha \in [0, 1]$ be fixed. If $u_i \in U$ for $i = 1, 2$, satisfy

$$\| L_1(f - u_i)\| \leqslant t_i,$$
$$\| L_2(f - u_i)\| = \lambda(t_i),$$

then we have

$$\begin{aligned}(1 - \alpha)\, \lambda(t_1) + \alpha\lambda(t_2) &= (1 - \alpha)\, \| L_2(f - u_1)\| + \alpha\, \| L_2(f - u_2)\| \\ &= \| L_2((1 - \alpha) f - (1 - \alpha)\, u_1)\| + \| L_2(\alpha f - \alpha u_2)\| \\ &\geqslant \| L_2(f - u_*)\|, \end{aligned} \tag{2.1}$$

where

$$u_* = (1 - \alpha)\, u_1 + \alpha u_2 .$$

But

$$\| L_1(f - u_*)\| \leqslant (1 - \alpha)\, t_1 + \alpha t_2 ,$$

so that by the definition of λ

$$\lambda((1 - \alpha)\, t_1 + \alpha t_2) \leqslant \| L_2(f - u_*)\|. \tag{2.2}$$

Hence, combining inequalities (2.1) and (2.2) proves the convexity of λ.

Next, it follows immediately from the definition that $\lambda(t)$ is nonincreasing on $[\mu_1, \infty)$. Suppose λ is not decreasing on $[\mu_1, \rho(\mu_2)]$. Then there exist $t_1, t_2 \in [\mu_1, \rho(\mu_2)]$, with $t_1 < t_2$, such that $\lambda(t_1) = \lambda(t_2)$. Now, it is clear that $\lambda(t) = \mu_2$ for all $t \geqslant \rho(\mu_2)$. Therefore, because λ is convex and nonincreasing, we have

$$\lambda(t_1) = \lambda(t_2) = \mu_2 .$$

However, by definition, $\rho(\mu_2)$ is the smallest number for which there exists $\tilde{u} \in U$ such that

$$\| L_1(f - \tilde{u})\| = \rho(\mu_2),$$
$$\| L_2(f - \tilde{u})\| = \mu_2 .$$

But there exists $u_1 \in U$ such that

$$\| L_1(f - u_1)\| \leqslant t_1 ,$$
$$\| L_2(f - u_1)\| = \lambda(t_1) = \mu_2 .$$

Hence, $\rho(\mu_2) \leqslant t_1$, so that $\rho(\mu_2) = t_1 = t_2$, which is a contradiction. The proof is thus complete.

LEMMA 2.3. *For each* $t \in [\mu_1, \rho(\mu_2)]$, $\rho(\lambda(t)) = t$. *Similarly, for each* $t \in [\mu_2, \lambda(\mu_1)]$, $\lambda(\rho(t)) = t$.

Proof. Let $t \in [\mu_1, \rho(\mu_2)]$ be fixed. By the definitions of ρ and λ, $\rho(\lambda(t)) \leqslant t$. Suppose $t_0 = \rho(\lambda(t)) < t$. Then there exists a $u_0 \in U$ such that $\| L_1(f - u_0)\| = t_0$ and $\| L_2(f - u_0)\| \leqslant \lambda(t)$. But $\lambda(t_0)$ is the smallest value of $\| L_2(f - u)\|$ over all $u \in U$ such that $\| L_1(f - u)\| \leqslant t_0$. Hence, $\lambda(t_0) \leqslant \lambda(t)$, which contradicts the property that λ is decreasing on $[\mu_1, \rho(\mu_2)]$. A similar proof shows that $\lambda(\rho(t)) = t$ for $t \in [\mu_2, \lambda(\mu_1)]$.

An immediate consequence of this lemma and the definitions of λ and ρ, is that if $t \in [\mu_1, \rho(\mu_2)]$ and $u \in U$ satisfies

$$\begin{aligned} \| L_1(f - u)\| &\leqslant t, \\ \| L_2(f - u)\| &= \lambda(t), \end{aligned} \tag{2.3}$$

then u is a best vectorial approximation. Similarly, if $u \in U$ satisfies

$$\begin{aligned} \| L_1(f - u)\| &= \rho(t), \\ \| L_2(f - u)\| &\leqslant t, \end{aligned} \tag{2.4}$$

for $t \in [\mu_2, \lambda(\mu_1)]$, then u is a best vectorial approximation. Let $\mathscr{P}_1(t)$ denote the problem of finding a $u \in U$ which satisfies condition (2.3), and let $\mathscr{P}_2(t)$ denote the problem of finding a $u \in U$ which satisfies condition (2.4).

THEOREM 2.1. *If $u \in U$ is a best vectorial approximation, then u solves $\mathscr{P}_1(t)$ for $t = \| L_1(f - u)\|$. Also, u solves $\mathscr{P}_2(t)$ for $t = \| L_2(f - u)\|$. Conversely, if u solves $\mathscr{P}_1(t)$ for some $t \in [\mu_1, \rho(\mu_2)]$, then u is a best vectorial approximation. Similarly, if u solves $\mathscr{P}_2(t)$ for some $t \in [\mu_2, \lambda(\mu_1)]$, then u is a best vectorial approximation.*

Proof. Only the proof for $\mathscr{P}_1(t)$ will be given. As noted above, it is an immediate consequence of Lemma 2.3 and the definitions of λ and ρ, that if u solves $\mathscr{P}_1(t)$ for some $t \in [\mu_1, \rho(\mu_2)]$ then u is a best vectorial approximation.

On the other hand, suppose $u \in U$ is a best vectorial approximation. By Lemma 2.1, if $t = \| L_1(f-u)\|$, then $t \in [\mu_1, \rho(\mu_2)]$. Further, $\| L_2(f-u)\| = \lambda(t)$ for otherwise, any solution of $\mathscr{P}_1(t)$ would be a better vectorial approximation.

This result shows that all best vectorial approximations can be obtained by solving problem $\mathscr{P}_1(t)$ for $t \in [\mu_1, \rho(\mu_2)]$, or equivalently, by solving $\mathscr{P}_2(t)$ for $t \in [\mu_2, \lambda(\mu_1)]$. These problems can be solved by linear programming methods, and further comments on these methods appear in Section 4.

3. CHARACTERIZATION AND UNIQUENESS

We investigate now the nature of the problems $\mathscr{P}_i(t)$, $i = 1, 2$. The results are stated explicitly only for $\mathscr{P}_1(t)$, as the corresponding statements for $\mathscr{P}_2(t)$ are then clear. We consider first the characterization of solutions, as this result is needed in the discussion of uniqueness.

For each $u \in U$ and $i = 1, 2$, define

$$E_i^+(u) = \{x \in D: \| L_i(f - u)\| = L_i(f - u)(x)\},$$
$$E_i^-(u) = \{x \in D: \| L_i(f - u)\| = -L_i(f - u)(x)\},$$

and

$$E_i(u) = E_i^+(u) \cup E_i^-(u).$$

If g is a real-valued function on D and $u \in U$, we say that g isolates $E_i(u)$ if $g(x) > 0$ for $x \in E_i^+(u)$ and $g(x) < 0$ for $x \in E_i^-(u)$. To simplify notation, we set $\hat{\rho} = \rho(\mu_2)$.

Our main interest here is with the case $\mu_1 < t \leqslant \hat{\rho}$. However, let us briefly mention the other cases.

If $\mu_1 = \hat{\rho}$, then there is a $u_* \in U$ such that $\| L_1(f - u_*)\| = \mu_1$ and $\| L_2(f - u_*)\| = \mu_2$. In this case, the characterization and uniqueness of u_* can be approached with results from the theory of Chebyshev approximation.

For the case $\mu_1 = t < \hat{\rho}$, set

$$B = \{u \in U: \| L_1(f - u)\| = \mu_1\}.$$

Then problem $\mathscr{P}_1(\mu_1)$ is equivalent to the problem of finding the best approximation to $L_2(f)$ from the convex set

$$K = \{L_2(u): u \in B\}.$$

The questions of characterization and uniqueness for this type of problem have been discussed by Deutsch and Maserick [9]. For example, from Theorem 2.5 and Corollary 2.6 of [9] and the characterization of the extreme points of the unit sphere in the dual space of $C(D)$ [10, p. 441], we have the following theorem.

THEOREM 3.1. *Problem $\mathscr{P}_1(\mu_1)$ is solved by $u_* \in U$ if and only if $u_* \in B$ and there exist points $x_1, x_2, \ldots, x_m$ in $E_2(u_*)$, and $\lambda_i > 0$, $i = 1, 2, \ldots, m$, where $m \leqslant n + 1$, such that $\sum_{i=1}^m \lambda_i = 1$ and*

$$\sum_{i=1}^{m} \lambda_i \sigma(x_i) L_2(u - u_*)(x_i) \leqslant 0$$

for all $u \in B$, where $\sigma(x_i) = +1$ if $x_i \in E_2^+(u_*)$ and $\sigma(x_i) = -1$ if $x_i \in E_2^-(u_*)$.

To obtain more concrete results on characterization and uniqueness, one would need further information concerning the nonuniqueness, if any, of the elements $u \in U$ for which $\| L_1(f - u)\| = \mu_1$.

We proceed now to the case $\mu_1 < t \leqslant \hat{\rho}$. The theorems of Deutsch–Maserick [9] apply here also; however, it is possible to obtain more detailed results for our particular problem.

THEOREM 3.2. *Suppose $\mu_1 < t \leqslant \hat{\rho}$. Then $u_* \in U$ solves $\mathscr{P}_1(t)$ if and only if $\| L_1(f - u_*)\| = t$, and there is no $u \in U$ such that $L_1(u)$ isolates $E_1(u_*)$ and $L_2(u)$ isolates $E_2(u_*)$.*

Proof. Suppose u_* is a solution of $\mathscr{P}_1(t)$ with $\mu_1 < t \leqslant \hat{\rho}$. It follows from Lemma 2.3 that $\| L_1(f - u_*)\| = t$. Assume next that $u \in U$ is such that $L_1(u)$ isolates $E_1(u_*)$ and $L_2(u)$ isolates $E_2(u_*)$. Then by a common argument in the theory of Chebyshev approximation (see, for example, [13, p. 14]), one can show that for some $\alpha > 0$, the function $\tilde{u} = u_* + \alpha u$ satisfies

$$\| L_1(f - \tilde{u})\| < t$$

and

$$\| L_2(f - \tilde{u})\| < \lambda(t),$$

which is impossible.

Conversely, suppose $u_* \in U$ is such that $\| L_1(f - u_*)\| = t$ where

$\mu_1 < t \leqslant \hat{\rho}$, and there is no $u \in U$ such that $L_1(u)$ isolates $E_1(u_*)$ and $L_2(u)$ isolates $E_2(u_*)$. Let $u_0 \in U$ be a solution of $\mathscr{P}_1(t)$. Assuming that u_* is not a solution of $\mathscr{P}_1(t)$, we have

$$\| L_2(f - u_0)\| < \| L_2(f - u_*)\|.$$

Set $u_1 = u_0 - u_*$. Then it follows that $L_2(u_1)$ isolates $E_2(u_*)$. Moreover, we have for $\lambda \in (0, 1)$,

$$\begin{aligned} \| L_1(f - u_* - \lambda u_1)\| &= \| L_1(f - (1 - \lambda)\, u_* - \lambda u_0)\| \\ &\leqslant (1 - \lambda)\, \| L_1(f - u_*)\| + \lambda\, \| L_1(f - u_0)\| \\ &\leqslant t, \end{aligned}$$

so that it must be the case that $L_1(u_1) \geqslant 0$ on $E_1^+(u_*)$ and $L_1(u_1) \leqslant 0$ on $E_1^-(u_*)$.

Next, because $\mu_1 < t$, it follows from a theorem of Kolomogorov (see [13, p. 15]) that there exists $u_2 \in U$ such that $L_1(u_2)$ isolates $E_1(u_*)$.

Set $v = u_1 + \alpha u_2$ where $\alpha > 0$. For any $\alpha > 0$, $L_1(v)$ isolates $E_1(u_*)$. Further, because $E_1(u_*)$ and $E_2(u_*)$ are compact, it follows that $L_2(v)$ isolates $E_2(u_*)$ for α sufficiently small and positive. Thus, we obtain a contradiction.

This theorem may be considered analogous to the classical theorem of Kolomogorov which characterizes Chebyshev approximations. It is not unreasonable to expect, therefore, that we can extend this result to obtain a characterization theorem based on the notion of extremal signatures, as was done by Rivlin and Shapiro [18] for the case of ordinary Chebyshev approximation.

A signature σ on D is a function with finite support whose values are either $+1$ or -1. If σ_1 and σ_2 are signatures on D, let us say that the pair $\{\sigma_1, \sigma_2\}$ is extremal with respect to problem $\mathscr{P}_1(t)$, $t \in [\mu_1, \hat{\rho}]$ if there exists a positive discrete measure ν_1 whose carrier is the support of σ_1 and a positive discrete measure ν_2 whose carrier is the support of σ_2 such that

$$\int L_1(u)\, \sigma_1\, d\nu_1 + \int L_2(u)\, \sigma_2\, d\nu_2 = 0$$

for all $u \in U$. We can now demonstrate the following theorem.

THEOREM 3.3. *Suppose $\mu_1 < t \leqslant \hat{\rho}$. Then $u_* \in U$ is a solution of $\mathscr{P}_1(t)$ if and only if $\| L_1(f - u_*)\| = t$ and there exists a pair of signatures $\{\sigma_1, \sigma_2\}$ which is extremal with respect to $\mathscr{P}_1(t)$ such that for each $i = 1, 2$, the support of σ_i is contained in $E_i(u_*)$ and $\sigma_i L_i(f - u_*) > 0$ on the support of σ_i. Moreover, the union of the supports of σ_1 and σ_2 consists of at most $n + 1$ points.*

Proof. Assume $u_* \in U$ is a solution of $\mathscr{P}_1(t)$. Then $\| L_1(f - u_*)\| = t$ by Theorem 3.2.

Let R^n designate n-dimensional Euclidean space with inner product $x \cdot y$ for $x, y \in R^n$. Assume $u_1, u_2, ..., u_n$ is a base for U and define $\Phi: D \to R^n$ and $\Psi: D \to R^n$ by

$$\Phi = (L_1(u_1), L_1(u_2), ..., L_1(u_n))^T,$$
$$\Psi = (L_2(u_1), L_2(u_2), ..., L_2(u_n))^T.$$

Let $V_1 \subset R^n$ be the union of the image of $E_1^+(u_*)$ under Φ and the image of $E_1^-(u_*)$ under $-\Phi$. Similarly, let V_2 be the union of the image of $E_2^+(u_*)$ under Ψ and the image of $E_2^-(u_*)$ under $-\Psi$. Then Theorem 3.2 states that there is no $y \in R^n$ such that $y \cdot v > 0$ for all $v \in V_1 \cup V_2$. Hence, it follows (see [7, p. 19]) that there exist positive numbers α_i and $v_i^{(1)} \in V_1$, $i = 1, 2, ..., m$, and positive numbers β_i and $v_i^{(2)} \in V_2$, $i = 1, 2, ..., l$, such that

$$\sum_{i=1}^{m} \alpha_i v_i^{(1)} + \sum_{i=1}^{l} \beta_i v_i^{(2)} = 0$$

and $0 < l + m \leqslant n + 1$. For any $c \in R^n$, therefore,

$$\sum_{i=1}^{m} \alpha_i c \cdot v_i^{(1)} + \sum_{i=1}^{l} \beta_i c \cdot v_i^{(2)} = 0.$$

In other words, there are points $x_1, ..., x_m$ in $E_i(u_*)$ and points $y_1, y_2, ..., y_l$ in $E_2(u_*)$ such that

$$\sum_{i=1}^{m} \alpha_i \sigma_1(x_i)\, L_1(u)(x_i) + \sum_{i=1}^{l} \beta_i \sigma_2(y_i)\, L_2(u)(y_i) = 0$$

for any $u \in U$, where σ_1 and σ_2 are signatures such that $\sigma_1(x_i) = +1$ if $x_i \in E_1^+(u_*)$, $\sigma_1(x_i) = -1$ if $x_i \in E_1^-(u_*)$, and similarly for σ_2. Hence, we have reached the desired conclusion.

Conversely, assume u_* satisfies $\| L_1(f - u_*)\| = t$, and there exists a pair $\{\sigma_1, \sigma_2\}$ of signatures with the stated properties. If $u_0 \in U$ is such that $L_1(u_0)$ isolates $E_1(u_*)$ and $L_2(u_0)$ isolates $E_2(u_*)$, then

$$\int L_1(u_0)\, \sigma_1 \, dv_1 + \int L_2(u_0)\, \sigma_2 \, dv_2 > 0,$$

which is impossible. Hence, by Theorem 3.2 it follows that u_* solves $\mathscr{P}_1(t)$.

We observe that in general, Theorem 3.3 is not valid for $t = \mu_1$, for there may be several $u \in U$ such that $\| L_1(f - u)\| = \mu_1$. However, if the element u_* for which $\| L_1(f - u_*)\| = \mu_1$ is unique, then we may say that Theorem 3.3 holds for $t = \mu_1$, with the understanding that the signature σ_2

does not appear, and that u_* is thus characterized by the extremal signature σ_1 alone.

In the case considered by Bacopoulos and Gaff [2], $L_1(f) = \omega_1 f$ and $L_2(f) = \omega_2 f$, where ω_1 and ω_2 are positive, continuous weight functions on the interval $[a, b]$. Also U is assumed to satisfy the Haar condition. For this case, it is not difficult to show that the only pair of signatures $\{\sigma_1, \sigma_2\}$ which can arise in problem $\mathscr{P}_1(t)$ have the following form:

(a) The supports of σ_1 and σ_2 are disjoint, and

(b) $\sigma_1 + \sigma_2$ is a primative extremal signature with respect to U (see [18]).

Since there is a unique element $u \in U$ for which $\| \omega_1(f - u)\| = \mu_1$, Theorem 2 of [2] follows from Theorem 3.3.

We are prepared now to consider the question of uniqueness. In the case of ordinary Chebyshev approximation, Newman and Shapiro [17] have shown that a best approximation is unique if the associated extremal signature is strong. The notion of a strong extremal signature can be extended to vectorial approximation. As any solution of $\mathscr{P}_1(\mu_1)$ is a solution of $\mathscr{P}_2(\lambda(\mu_1))$, it follows from Theorem 3.3 and the statement of Theorem 3.3 in the case of problem $\mathscr{P}_2(\lambda(\mu_1))$, that associated with each solution of $\mathscr{P}_1(t)$ for $t \in [\mu_1, \hat{\rho}]$, there is a pair of signatures which is extremal. Let us say that a pair of signatures $\{\sigma_1, \sigma_2\}$, which is extremal with respect to $\mathscr{P}_1(t)$, $t \in [\mu_1, \hat{\rho}]$, is strong if $u = 0$ is the only member of U for which $L_i(u) = 0$ on the support of σ_i for $i = 1, 2$.

THEOREM 3.4. *Let $u_* \in U$ be a solution of $\mathscr{P}_1(t)$, $t \in [\mu_1, \hat{\rho}]$, and suppose that an associated extremal pair of signatures $\{\sigma_1, \sigma_2\}$ is strong. Then u_* is unique.*

Proof. If $\sigma_2 = 0$, then σ_1 is itself a strong extremal signature. Thus, u_* is the unique element such that $\| L_1(f - u)\| = \mu_1$. Hence, $t = \mu_1$ and u_* is the unique solution of $\mathscr{P}_1(\mu_1)$. A similar argument holds when $\sigma_1 = 0$.

Assume now that $\sigma_1 \neq 0$ and $\sigma_2 \neq 0$. Define the functional $\mathscr{L}$ by

$$\mathscr{L}(\omega) = \int L_1(\omega)\, \sigma_1\, d\nu_1 + \int L_2(\omega)\, \sigma_2\, d\nu_2\,,$$

where ν_1 and ν_2 are the positive discrete measures corresponding to σ_1 and σ_2, respectively. We may assume that

$$\int d\nu_1 + \int d\nu_2 = 1.$$

Therefore,

$$\mathscr{L}(f - u_*) = t + \lambda(t)$$

and because $\mathscr{L}(u) = 0$ for all $u \in U$, it follows that

$$\mathscr{L}(f - u) = t + \lambda(t) \tag{3.1}$$

for any $u \in U$. However, if $u_0 \in U$ is another best vectorial approximation, then $\| L_1(f - u_0)\| = t$ and $\| L_2(f - u_0)\| = \lambda(t)$. Therefore, Eq. (3.1) with $u = u_0$ implies that

$$\sigma_1(x)\, L_1(f - u_0)(x) = t$$

for x in the support of σ_1, and

$$\sigma_2(x)\, L_2(f - u_0)(x) = \lambda(t)$$

for x in the support of σ_2. Hence, it follows that $L_1(u_* - u_0) = 0$ on the support of σ_1 and $L_2(u_* - u_0) = 0$ on the support of σ_2. Therefore, $u_* = u_0$.

For the setting considered in [2], mentioned above, all pairs of signatures which are extremal with respect to $\mathscr{P}_1(t)$ are strong. However, in general, it appears difficult to find interesting choices of operators L_1 and L_2 and subspaces U for which the solutions of $\mathscr{P}_1(t)$ are unique. In particular, for domains $D \subset R^n$ with $n \geqslant 2$, we suspect that solutions are rarely unique, as is the case for ordinary Chebyshev approximation. A setting of some interest in applications is that in which $L_1(f) = f$ and $L_1(f) = f'$. However, even for approximation by polynomials on $[a, b]$, the solutions of $\mathscr{P}_1(t)$ may not be unique.

EXAMPLE. Let $D = [-1, 1]$, $L_1(f) = f$ and $L_2(f) = f'$, and $U = \{$polynomials of degree $\leqslant 2\}$. Let f be the odd function whose graph on [0, 1] is shown in Fig. 1. Here, $\alpha > 1$ and on $[0, \tau]$, $f(x) = 8x - 16x^2$ and on $[\tau, 1]$,

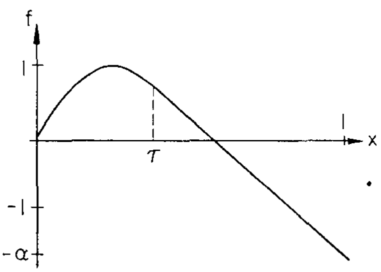

FIG. 1. Function with nonunique vectorial approximations.

the graph of f is the straight line tangent to the graph of $8x - 16x^2$ at $x = \tau$, and passing through the point $(+1, -\alpha)$.

One can show that if $\alpha - 1$ is sufficiently small and positive, then

$\mu_1 < \alpha < \hat{\rho}$. From Theorem 3.3, we see that $u_* = 0$ solves problem $\mathscr{P}_1(t)$ with $t = \alpha$. Indeed, $E_1(u_*) = \{-1, +1\}$ and $E_2(u_*) = \{0\}$, and

$$\tfrac{1}{2}u(-1) + u'(0) - \tfrac{1}{2}u(1) = 0$$

for all $u \in U$. Hence, the pair of signatures $\{\sigma_1, \sigma_2\}$ where

$$\sigma_1(x) = \begin{cases} +1 & x = -1, \\ -1 & x = +1, \\ 0 & \text{otherwise}, \end{cases}$$

$$\sigma_2(x) = \begin{cases} 1 & x = 0, \\ 0 & \text{otherwise} \end{cases}$$

is extremal, which shows that $u_* = 0$ is a solution. However, it is not difficult to check that for all β sufficiently small and positive, $\beta(x^2 - 1) \in U$ is also a solution of $\mathscr{P}_1(t)$ with $t = \alpha$.

Bacopoulos [4] has posed the problem of characterizing the best vectorial approximations for the case of the sup norm of a function and its first derivative. Theorem 3.3 would apply here; however, for a complete characterization, it would be necessary to determine all the extremal pairs of signatures. In Fig. 2 below we show a few of these signatures for the case of approximation by polynomials of degree at most two.

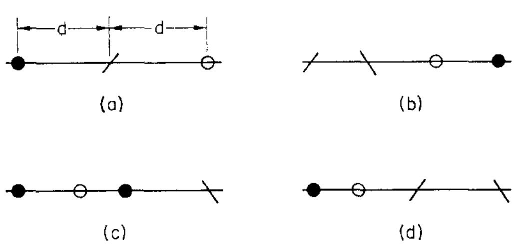

FIG. 2. Pairs of signatures which are extremal for the case $L_1(f) = f$, $L_2(f) = f'$.

The dots represent points in the support of σ_1 (where $L_1(f) = f$), and the slanted dashes represent points in the support of σ_2 (where $L_2(f) = f'$). A black dot is a point at which $\sigma_1 > 0$ and a white dot is one at which $\sigma_1 < 0$. Similarly, the slope of a dash indicates the sign of σ_2 at that point. We observe that the signatures (b), (c), and (d) are strong, while the one in (a) is not.

4. Computational Aspects

We showed in Section 2 that all vectorial approximations for $k = 2$ can be obtained as solutions of problem $\mathscr{P}_1(t)$ for $t \in [\mu_1, \hat{\rho}]$. This problem can be solved numerically by linear programming methods. The use of linear programming is quite general, and requires no special assumptions, such as the Haar condition. We present some specific techniques for implementing a linear programming approach in [11].

A basic step in using the linear programming method is to discretize the domain D, and solve problem $\mathscr{P}_1(t)$ on the discrete subset. For the case of ordinary Chebyshev approximation, Cheney and Rivlin [8] have shown that the solution of the discrete problem, for sufficiently small grid size, is a reasonable substitute for the actual solution. A similar result can be shown for problem $\mathscr{P}_1(t)$, $t \in [\mu_1, \hat{\rho}]$.

As in Cheney [7], we introduce the following notation. Assume D is a metric space with metric d. For $G \subset D$ define

$$| G | = \sup_{x \in D} \inf_{y \in G} d(x, y),$$

and for each $f \in C(D)$ define

$$\| f \|_G = \sup_{x \in G} | f(x) | .$$

Theorem 4.1. *Let $t \in [\mu_1, \hat{\rho}]$. If $G \subset D$, let $u_G \in U$ minimize $\| L_2(f - u_G)\|_G$ subject to the restriction $\| L_1(f - u_G)\|_G \leqslant t$. Then*

$$\| L_1(f - u_G)\| \to t$$

and

$$\| L_2(f - u_G)\| \to \lambda(t)$$

as $| G | \to 0$.

Proof. Using Lemmas 1 and 2 of [7, p. 85, 86], it can be shown that

$$\| L_1(f - u_G)\| \leqslant t + e_1(| G |) \tag{4.1}$$

and

$$\| L_2(f - u_G)\| \leqslant \| L_2(f - u_G)\|_G + e_2(| G |), \tag{4.2}$$

where $e_i(x) \to 0$ as $x \to 0$, for $i = 1, 2$. The proof of these results depends on showing that $\| L_1(u_G)\|$ and $\| L_2(u_G)\|$ are bounded independent of $| G |$, as $| G | \to 0$. For $\| L_1(u_G)\|$, we have

$$\| L_1(u_G)\|_G \leqslant \| L_1(f - u_G)\|_G + \| L_1(f)\| \leqslant t + \| L_1(f)\|$$

and by Lemma 1 of [7, p. 85]

$$\| L_1(u_G)\| \leqslant 2 \| L_1(u_G)\|_G$$

for $| G |$ sufficiently small. For $\| L_2(u_G)\|$, we have first

$$\| L_2(u_G)\|_G \leqslant \| L_2(f - u_G)\|_G + \| L_2(f)\|.$$

Let u_* satisfy $\| L_1(f - u_*)\| = t$ and $\| L_2(f - u_*)\| = \lambda(t)$. Because $\| L_1(f - u_*)\|_G \leqslant t$, it follows that

$$\| L_2(f - u_G)\|_G \leqslant \| L_2(f - u_*)\|_G \leqslant \lambda(t),$$

which establishes the bound on $\| L_2(u_G)\|$.

Now set

$$\Lambda_G = \{u \in U \colon \| L_1(f - u)\|_G \leqslant t\},$$

and let $\tilde{u}_G \in U$ minimize $\| L_2(f - u)\|$ over all $u \in \Lambda_G$. Since $u_G, \tilde{u}_G \in \Lambda_G$,

$$\| L_2(f - u_G)\|_G \leqslant \| L_2(f - \tilde{u}_G)\|_G \leqslant \| L_2(f - \tilde{u}_G)\| \leqslant \| L_2(f - u_G)\|.$$

Hence, using (4.2), we have

$$0 \leqslant \| L_2(f - u_G)\| - \| L_2(f - \tilde{u}_G)\| \leqslant e_2(| G |). \tag{4.3}$$

Furthermore, from the same reasoning as was used to establish (4.1), one can show that

$$\Lambda_G \subset \{u \in U \colon \| L_1(f - u)\| \leqslant t + e_1(| G |)\}.$$

Hence,

$$\| L_2(f - \tilde{u}_G)\| \geqslant \lambda(t + e_1(| G |))$$

so that

$$0 \leqslant \| L_2(f - u_*)\| - \| L_2(f - \tilde{u}_G)\| \leqslant \lambda(t) - \lambda(t + e_1(| G |)).$$

Therefore, by (4.3) and the continuity of λ on $[\mu_1, \infty)$

$$\| L_2(f - u_G)\| \to \| L_2(f - u_*)\| = \lambda(t)$$

as $| G | \to 0$.

To show that $\| L_1(f - u_G)\| \to t$, suppose first that for some $\epsilon > 0$, and some sequence $G_i \subset D$, with $| G_i | \to 0$ as $i \to \infty$,

$$\| L_1(f - u_{G_i})\| \leq t - \epsilon$$

then for each G_i,

$$\| L_2(f - u_{G_i})\| \geq \lambda(t - \epsilon) > \lambda(t),$$

which is a contradiction. Hence, from this result and (4.1), we conclude that $\| L_1(f - u_G)\| \to t$ as $| G | \to 0$.

In computing approximations based on more than one criterion of approximation, vectorial approximation has an advantage from the viewpoint of the user. For in solving problem $\mathscr{P}_1(t)$, the user may vary t between μ_1 and $\hat{\rho}$ and in this way he has direct control over the deviations $\| L_1(f - u)\|$ and $\| L_2(f - u)\|$. Such would not be the case, for example, if the criterion of approximation were to minimize

$$\| L_1(f - u)\| + \beta \| L_2(f - u)\| \tag{4.4}$$

or

$$\max\{\| L_1(f - u)\|, \beta \| L_2(f - u)\|\}, \tag{4.5}$$

where $\beta > 0$ is a fixed weight. For these criterion, the user must accept the resulting values of $\| L_1(f - u)\|$ and $\| L_2(f - u)\|$. Of course, one can vary the weight β in order to influence these deviations. However, this technique still does not provide the direct control one has in solving $\mathscr{P}_1(t)$, and, moreover, the amount of work needed to minimize (4.4) or (4.5) is, in general, at least as much as that needed to solve $\mathscr{P}_1(t)$.

For the case $k \geq 3$, we are not aware of any result such as Theorem 2.1 which provides a framework for the computation of all vectorial approximations. However, when $k \geq 3$, methods based on the use of parametric linear programming can be applied to compute vectorial approximations. We present these and related methods in [11].

References

1. A. Bacopoulos, Nonlinear Chebyshev approximation by vector-norms, *J. Approximation Theory* **2** (1969), 79–84.
2. A. Bacopoulos and B. Gaff, On the reduction of a problem of minimization of $n + 1$ variables to a problem of one variable, *SIAM J. Numer. Anal.* **8** (1971), 97–103.
3. A. Bacopoulos and G. D. Taylor, Vectorial approximation by restricted rationals, *Math. Systems Theory* **3** (1969), 232–243.

4. A. Bacopoulos, Topology of a general approximation system and applications, *J. Approximation Theory* **4** (1971), 147–158.
5. E. Bredendiek, Simultanapproximationen, *Arch. Rat. Mech. Anal.* **33** (1969), 307–330.
6. B. Brosowski, Über Tschebysheffsche approximationen mit linearen Nebenbedingungen, *Math. Z.* **88** (1965), 105–128.
7. E. W. Cheney, "Introduction to Approximation Theory," McGraw-Hill, New York, 1966.
8. E. W. Cheney and T. J. Rivlin, A comparison of uniform approximations on an interval and a finite subset thereof, *SIAM J. Numer. Anal.* **3** (1966), 311–320.
9. F. R. Deutsch and P. H. Maserick, Applications of the Hahn–Banach theorem in approximation theory, **9** (1967), 516–530.
10. N. Dunford and J. T. Schwartz, "Linear Operators, Part I: General Theory," Interscience Publishers, Inc., New York, 1964.
11. W. B. Gearhart, Use of parametric linear programming in vectorial approximation, submitted for publication.
12. P. J. Laurent, Théorèmes de caractérisation d'une meilleure approximation dans un espace normé et généralisation de l'algorithme de Remes, *Numer. Math.* **10** (1967), 190–208.
13. G. Meinardus, "Approximation of Functions: Theory and Numerical Methods," (L. L. Schumaker, transl.) Springer-Verlag, New York, 1967.
14. D. G. Moursund, Chebyshev approximation of a function and its derivatives, *Math. Comp.* **18** (1964), 382–389.
15. D. G. Moursund and A. H. Stroud, The best Cebysev approximation to a function and its derivative on $n + 2$ points, *SIAM J. Numer. Anal.* **2** (1965), 15–23.
16. D. G. Moursund, Some computational aspects of the uniform approximation of a function and its derivative, *SIAM J. Numer. Anal.* **2** (1965), 464–472.
17. D. J. Newman and H. S. Shapiro, Some theorems on Cebysev approximation, *Duke Math. J.* **30** (1963), 673–682.
18. T. J. Rivlin and H. S. Shapiro, A unified approach to certain problems of approximation and minimization, *J. Soc. Indust. Appl. Math.* **9** (1961), 670–699.