JOURNAL OF APPROXIMATION THEORY 7, 213–225 (1973)

Uniqueness of Best Approximation of a Function and Its Derivatives

BRUCE L. CHALMERS

University of California, Riverside, California 92502

Communicated by Oved Shisha

Received October 20, 1970

1. INTRODUCTION

Let C^rI denote the space of r-times continuously differentiable functions on the interval $I = [a, b]$ of the real line $\mathbf{R}$. The question of uniqueness of best approximation of functions in C^rI by functions in a finite dimensional subspace, with respect to various norms. has been investigated in several papers. Garkavi [3] examined the problem using the ordinary supremum norm

$$\|f\|_\infty = \max_{x \in I} |f(x)|.$$

In [1] we considered the norms

$$\|f\| = \max[|f(c)|, |f^{(1)}(c)|,..., |f^{(r-1)}(c)|, \|f^{(r)}\|_p], \qquad 1 \leqslant p \leqslant \infty,$$

where $\|\cdot\|_p$ denotes the L^p norm and c is a fixed point in I. Moursund [5] and Johnson [4] studied the norm

$$\|f\| = \max[\|f\|_\infty , \|f^{(1)}\|_\infty ,..., \|f^{(r)}\|_\infty].$$

In this paper we shall further investigate this latter norm.

Moursund and Johnson show that if the $(r + 1)$st derivative of f exists everywhere on I and if p_1 and p_2 are best approximations to f in P_n, the space of polynomials of degree $\leqslant n$, then $p_1^{(r)} = p_2^{(r)}$, $r = 0, 1, 2,...$. In the case $r = 0$, Tchebycheff's classical result shows that the requirement of existence of the $(r + 1)$st derivative is unnecessary. In Section 2 we give an example to show that this requirement cannot be dropped if $r > 0$.

Garkavi showed that in order for an n-dimensional subspace V to be p-Tchebycheff (see Section 3 for definition) with respect to the usual supremum norm in C^r, $r \geqslant 1$, it is necessary and sufficient that any $k + p$

linearly independent elements of V have no more than $n - k - p$ common zeros which are also common double zeros or boundary zeros of $p + 1$ of these elements, $k = 1, 2, \ldots, n - p$. In Section 4 we shall extend the sufficiency part of Garkavi's result to the norm

$$\|f\| = \max[\|f\|_\infty, \|f^{(1)}\|_\infty, \ldots, \|f^{(r)}\|_\infty]$$

on the space of functions having an $(r + 1)$st derivative everywhere on I (Theorem 3). By use of the results of Ferguson [2] we shall see that the polynomials satisfy the conditions of this extended sufficiency result, and, thus, the result of Moursund and Johnson (Theorem 5) follows as a corollary of Theorem 3.

The results mentioned above are special cases of the more general ones discussed in Section 3 where we consider the simultaneous approximation of $r + 1$ continuous functions $f_0, f_1, \ldots, f_r$ by a function p in V and by its first r derivatives over $r + 1$ possibly different subsets of $\mathbf{R}$. In the situation where each of the $r + 1$ subsets of $\mathbf{R}$ is the same finite union of closed intervals, we shall perform a certain imbedding and then employ the methods of Rivlin and Shapiro [6] and Garkavi [3] to obtain an extension of Garkavi's result (the condition on V will be necessary and sufficient). Finally, in Section 5 we shall obtain uniqueness results for approximation by polynomials with respect to another arrangement of the $r + 1$ subsets of $\mathbf{R}$. Here we shall use again the results of Ferguson to show how the space P_n fits into the scheme.

2. Example of Nonuniqueness of Derivative of Best Approximation

Definition. If V is a subspace of a normed linear space S with norm $\|\cdot\|$, we say that $g \in V$ is a best approximation of an element f of S if $\|f - g\| = \inf_{h \in V} \|f - h\|$. It is clear that the set of such best approximations is convex.

In this section we shall demonstrate a function $f \in C^1 I$ such that the best approximations to f in P_2 with respect to the norm $\|f\| = \max[\|f\|_\infty, \|f^{(1)}\|_\infty]$ do not have identical derivatives.

Let $I = [-4.25, 4.25]$. Let $f^{(1)}(x) = |x|$ $(0 \leqslant |x| \leqslant 1.5)$ and $f^{(1)}(x) = 1.5$ $(1.5 \leqslant |x| \leqslant 4.5)$. Let $f(0) = 0$. Then $f(x) = (\operatorname{sgn} x)\, x^2/2$ for $0 \leqslant |x| \leqslant 1.5$ and $f(x) = (\operatorname{sgn} x)[1.5\,|x| - 1.125]$ for $1.5 \leqslant |x| \leqslant 4.5$; see Figs. 1 and 2. Notice that $f^{(1)}(x)$ is even and $f(x)$ is odd.

Now suppose that the derivative of a best approximation p in P_2 to f is unique. Then its graph must be horizontal. For, because of the symmetry of f and $f^{(1)}$, if $p^{(1)}(x) = ax + b$, then $p_*^{(1)}(x) = -ax + b$ is also the derivative of a best approximation in P_2 to f. We claim that $p(x) = x$.

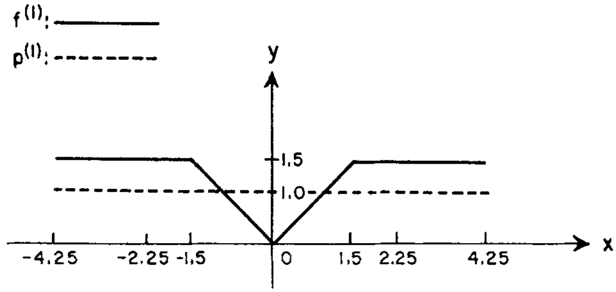

FIGURE 1.

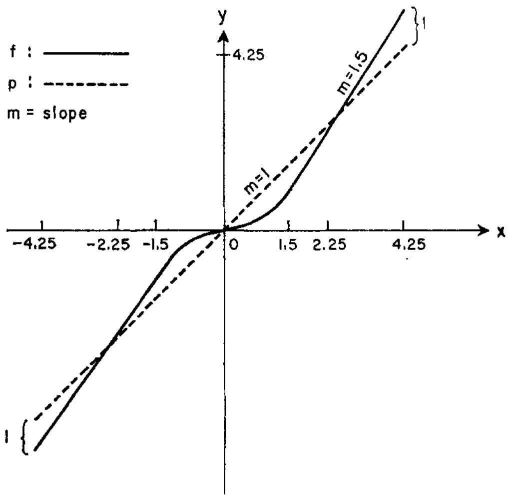

FIGURE 2.

Indeed, $f(x)$ is an increasing function with values varying on I between -5.25 and 5.25. If $p^{(1)}(x) = 1$, then $p(x) = x + c$ is an increasing function with values varying on I between $c - 4.25$ and $c + 4.25$. Thus, $p(x) = x + c$ has deviation of $1 + |c|$ from $f(x)$ at one of the endpoints. Let $c = 0$. Then it is easy to check that the maximum deviation 1 of $p(x) = x$ from $f(x)$ occurs only at the endpoints of I. Note that the maximum deviation of $p^{(1)}(x) = 1$ from $f^{(1)}(x)$ is also 1 and occurs at $x = 0$. Thus,

$$\| f - p \| = \max[\| f - p \|_\infty , \| f^{(1)} - p^{(1)} \|_\infty] = 1.$$

Further, if $p^{(1)}(x) = a > 1$, then $\| f^{(1)} - p^{(1)} \|_\infty = a > 1$. If $p^{(1)}(x) = a < 1$, then $p(x)$ has total variation $9.5a$ on I; hence, $p(x)$ must deviate from $f(x)$ by more than 1 at at least one of the endpoints 4.25 or -4.25. We conclude that if the derivative of a best approximation p in P_2 to f is unique, then p is unique and $p(x) = x$.

Now, however, consider $p_\epsilon(x) = (\epsilon/2)\, x^2 + x$, $\epsilon \geqslant 0$. Then $p_\epsilon^{(1)}(x) = \epsilon x + 1$ and it is easy to check that, for ϵ sufficiently small, $\| f^{(1)} - p_\epsilon^{(1)} \|_\infty = 1$ ($f^{(1)}(x) - p_\epsilon^{(1)}(x) = 1$, iff $x = 0$), and $\| f - p_\epsilon \|_\infty = 1$ ($f(x) - p_\epsilon(x) = 1$ iff $x = \pm 4.25$). Thus, $p(x) = x$ is not a unique best approximation in P_2 to f.

Remark 1. Note that the crux of the matter in the foregoing example is that we can slightly rotate the graph of $p^{(1)}(x)$ about the point (0, 1) without increasing $\| f^{(1)} - p^{(1)} \|_\infty$. This is because the graph of $f^{(1)}$ is wedge-shaped at $x = 0$.

Remark 2. If the length of the interval I is not greater than 2, then the requirement of existence of the $(r + 1)$st derivative in Moursund and Johnson's result can be dropped (this follows from the mean value theorem). In fact, $p^{(r)}$ is then the best Tchebycheff approximation to $f^{(r)}$ of degree $\leqslant n - r$, and $\| f - p \| = \| f^{(r)} - p^{(r)} \|_\infty$.

3. Simultaneous Approximation

Let S be a subspace of $\bigotimes_{j=0}^{r} C(E_j)$, where E_j ($j = 0, 1,..., r$) are compact subsets of $\mathbf{R}$, with norm $\| f \| = \| (f_0, f_1, ..., f_r) \| = \max[\| f_0 \|_\infty, \| f_1 \|_\infty, ..., \| f_r \|_\infty]$ where $\| f_j \|_\infty = \sup_{x \in E_j} | f_j(x) |$.

Definition. By the dimension of a convex set P (dim P) in a finite dimensional vector space we mean the largest integer k for which there exist $k + 1$ elements $g_1, g_2, ..., g_{k+1}$ in P such that

$$g_1 - g_{k+1},\ g_2 - g_{k+1}, ...,\ g_k - g_{k+1}$$

are linearly independent. (If P consists of a single point, we set $\dim(P) = 0$; if P is empty, we set $\dim(P) = -1$.) If W is a subspace of S, then, for each fixed q ($0 \leqslant q \leqslant r$), the maximum dimension of sets $P_W^{(q)}(f)$ of qth components of elements of best approximation in W of functions f in S is called the q-rank of W in S. (In the case $r = 0$ we say (following [8]) that W is s-semi-Tchebycheff or s-Tchebycheff if, for all f in S, $-1 \leqslant \dim P_W^{(0)}(f) \leqslant s$ or $0 \leqslant P_W^{(0)}(f) \leqslant s$, respectively.)

Now suppose V is an n-dimensional space of functions g defined on $E = \bigcup_{i=0}^{r} E_j$ which belong to $\bigcap_{j=0}^{r} C^j E_j$. Let $\tilde{V} = \{ \tilde{g} = (g, g^{(1)}, ..., g^{(r)});$

$g \in V\}$. We wish to investigate the r-rank of $\tilde{V}$ in S (provided, of course, $\tilde{V}$ is a subspace of S). Note that $\|f - \tilde{g}\| < \epsilon$ means $|f_j(x) - g^{(j)}(x)| < \epsilon$ for all x in E_j and all $j = 0, 1,..., r$ simultaneously.

If $f \in S$, then imbed f in $C(X)$, where $X = \bigcup_{j=0}^{r} (E_j, j)$, by $f^*(x, j) = f_j(x)$ if $x \in E_j$, $j = 0, 1,..., r$. We endow X with its natural topology. By the Hahn–Banach theorem, there exists an element L in the dual of $C(X)$, $[C(X)]^0$, such that $L(\tilde{V}) = \{0\}$, $\| L \|^0 = 1$, and $L(f) = \rho = \inf_{\tilde{g} \in \tilde{V}} \|f - \tilde{g}\|$. By the Riesz representation theorem, $L(h) = \int_x h\, d\mu$, where μ is a finite Borel measure on X. Now proceeding as in the proof of Haar's theorem (see [6]) we conclude that $\tilde{g}$ is a best approximation in $\tilde{V}$ to f if and only if $\tilde{g}^*$ is a best approximation in $\tilde{V}^*$ to f^*, and the latter implies that $f^* - \tilde{g}^* = \rho \bar{h}^*$, where $| h^* | = 1$ almost everywhere with respect to μ.

Note that $\mu |_{(E_j, j)} = \mu_j$ is a finite Borel measure on (E_j, j), $j = 0, 1,..., r$. Hence, we can write $\mu = \mu_0 + \mu_1 + \cdots + \mu_r$. We refer to an element of X as a generalized point. If $g \in V$, we call any zero of $\tilde{g}^*$ in X a generalized zero of g.

The proof of the following two theorems were obtained by combining the methods of Garkavi [3] and Rivlin and Shapiro [6] after performing the imbedding described previously.

Theorem 1 reduces to a slight generalization of Garkavi's theorem [3, p. 97], if we set $r = 0$.

THEOREM 1. *Let $S = \otimes_{j=0}^{r} \{f_j\,; f_j$ is differentiable on $E\}$ where E is a finite union of disjoint closed intervals $\{I_\alpha\}_{\alpha=1}^{m}$. Then for $\tilde{V}$ to have r-rank s in S, it is necessary and sufficient that among the common generalized zeros of k ($k = s + 1, s + 2,..., n$) linearly independent elements of V there are no more than $n - k$ generalized points which are generalized double or boundary zeros of $s + 1$ of these elements whose* rth *derivatives are linearly independent. ((x, j) is a generalized double zero of p if $p^{(j)}(x) = p^{(j+1)}(x) = 0$; (x, j) is a generalized boundary zero of p if $p^{(j)}(x) = 0$, where x is a boundary point of some I_α.*

Proof. Sufficiency. Suppose $g_{s+2}^{(r)} - g_1^{(r)}, g_{s+1}^{(r)} - g_1^{(r)},..., g_2^{(r)} - g_1^{(r)}$ are linearly independent where $\tilde{g}_1, \tilde{g}_2,..., \tilde{g}_{s+2}$ are best approximations in $\tilde{V}$ to f. Hence, among the common generalized zeros of the elements $g_{s+2} - g_1$, $g_{s+1} - g_1,..., g_2 - g_1$, there are at most $n - s - 1$ common generalized double or boundary zeros in X. But each interior generalized zero of $g_i - g_1$ in the support of μ is a generalized double zero, $i = 2, 3,..., s + 2$. This follows since, if (x, j) is interior to (E, j) and is in the support of μ, then

$$|f_j(x) - g_i^{(j)}(x)| = \rho = \max_{y \in E} |f_j(y) - g_i^{(j)}(y)|,$$

which implies

$$f_j^{(1)}(x) - g_i^{(j+1)}(x) = 0, \qquad i = 1, 2, 3,..., s+2;$$

hence,

$$g_i^{(j+1)}(x) - g_1^{(j+1)}(x) = 0, \qquad i = 2, 3,..., s+2.$$

Hence, μ has a support of $n-k+1$ $(k \geqslant s+2)$ generalized points, say $(x_1, i_1), (x_2, i_2),..., (x_{n-k+1}, i_{n-k+1})$. Thus, $L = \sum_{j=1}^{n-k+1} c_j \mathscr{L}_{(x_j, i_j)}$, where $\mathscr{L}_{(x_j, i_j)} h = h_{i_j}(x_j)$. Now $L(\tilde{V}) = \{0\}$ implies that $\sum_{j=1}^{n-k+1} c_j e_{(x_j, i_j)} = 0$ on V, where $e_{(x_j, i_j)} g = g^{(i_j)}(x_j)$ $(1 \leqslant j \leqslant n-k+1)$, and, thus, $\{e_{(x_j, i_j)}\}_{j=1}^{n-k+1}$ has rank $\leqslant n-k$ on V (i.e., $\{e_{(x_j, i_j)}\}_{j=1}^{n-k+1}$ spans a space of dimension at most $n-k$ in V^0.) Hence, there are, in V, k linearly independent elements $h_1 = g_{s+2} - g_1$, $h_2 = g_{s+1} - g_1$,..., $h_{s+1} = g_2 - g_1$, h_{s+2},..., h_k such that $h_t^{i_j}(x_j) = 0$ $(j = 1, 2,..., n-k+1)$, $t = 1, 2,..., k$. But each (x_j, i_j), $j = 1, 2,..., n-k+1$, is a common generalized double or boundary zero of $h_1, h_2,..., h_{s+1}$. Hence, among the common generalized zeros of the k $(\leqslant s+2)$ linearly independent elements $h_1, h_2,..., h_k$ of V there are $n-k+1$ generalized double or boundary zeros of $h_1, h_2,..., h_{s+1}$, and $h_1^{(r)}, h_2^{(r)},..., h_{s+1}^{(r)}$ are linearly independent—a contradiction.

Necessity. Suppose there exist linearly independent elements $g_1, g_2,..., g_k$ $(k \geqslant s+1)$ in V whose common generalized zeros include as a subset $T = \{(x_1, i_1), (x_2, i_2),..., (x_{n-k+1}, i_{n-k+1})\}$, each element of which is a generalized double or boundary zero of $g_1, g_2,..., g_{s+1}$, and $g_1^{(r)}, g_2^{(r)},..., g_{s+1}^{(r)}$ are linearly independent. Then $\{e_{(x_j, i_j)}\}_{j=1}^{n-k+1}$ is a linearly dependent system in V^0, for its rank does not exceed $n-k$, since $e_{(x_j, i_j)}(g_t) = 0$ for $t = 1, 2,..., k$, $j = 1, 2,..., n-k+1$. Hence, there exist scalars c_j $(1 \leqslant j \leqslant n-k+1)$ not all zero, such that $L = \sum_{j=1}^{n-k+1} c_j \mathscr{L}_{(x_j, i_j)} = 0$ on $\tilde{V}$. Assume, without loss of generality, that $\sum_{j=1}^{n-k+1} | c_j | = 1$. Clearly $\| L \|^0 \leqslant 1$. Now choose h_s in $C^2(E)$ such that $\| h_s \|_\infty = 1$, $h_s(x_j) = \operatorname{sgn} c_j$ for all $(x_j, s) \in T$, and $| h_s(x) | < 1$ if $(x, s) \notin T$, $s = 0, 1,..., r$. Let $h = (h_0, h_1,..., h_s) \in S$. Then clearly $| Lh | = \sum_{j=1}^{n-k+1} | c_j | = 1$, while $\| h \| = 1$. Hence, $\| L \|^0 = 1$. We may assume that $\| \tilde{g}_m \| < 1/k$, $m = 1, 2,..., k$. For $s = 0, 1,..., r$, form $f_s(x) = h_s(x)[1 - \sum_{m=1}^{k} | g_m^{(s)}(x) |]$ on $F = \bigcup_{j=1}^{n-k+1} ([\alpha_j, \beta_j], i_j)$, where $([\alpha_j, \beta_j], i_j)$ is a neighborhood of (x_j, i_j) containing no simple zeros of $g_1^{(i_j)}, g_2^{(i_j)},..., g_k^{(i_j)}$ except, possibly, boundary zeros. This is possible since either all $g_m^{(i_j)}$ $(1 \leqslant m \leqslant k)$ have a double zero at x_j, or x_j is a boundary point of E. Since $g_m^{(s)}$ has, in F, only zeros of order greater than one, except possibly at the boundary of F, $| g_m^{(s)} |$ is also differentiable in F $(1 \leqslant m \leqslant k)$. Hence, f_s is differentiable in F. Further $| f_s(x) | < 1$ if (x, s) is an (α_j, i_j) or (β_j, i_j) in the interior of (E, s). Thus, we can extend $f_s(x)$ to a function having a derivative in all of E and of absolute value $< 1 - \delta$ in $E \sim F$, $\delta > 0$ ($s = 0, 1, \ldots, r$). Let $f = (f_0, f_1, \ldots, f_r) \in S$. Then for all $\tilde{h}$ in $\tilde{V}$,

$$\|f - \tilde{h}\| \geqslant |L(f - \tilde{h})| = |Lf| = \sum_{j=1}^{n-k+1} c_j f_{i_j}(x_j)$$

$$= \sum_{j=1}^{n-k+1} c_j h_{i_j}(x_j) \left[1 - \sum_{m=1}^{k} |g_m^{(i_j)}(x_j)|\right] = \sum_{j=1}^{n-k+1} c_j h_{i_j}(x_j)$$

$$= \sum_{j=1}^{n-k+1} c_j \operatorname{sgn} c_j = 1.$$

On the other hand,

$$\left| f_s(x) - \sum_{m=1}^{k} \epsilon_m g_m^{(s)}(x) \right| \leqslant |f_s(x)| + \sum_{m=1}^{k} \epsilon_m |g_m^{(s)}(x)|$$

$$\leqslant |h_s(x)| \left[1 - \sum_{m=1}^{k} |g_m^{(s)}(x)|\right] + \sum_{m=1}^{k} \epsilon_m |g_m^{(s)}(x)|$$

$$\leqslant 1 \quad \text{if} \quad 0 \leqslant \epsilon_m \leqslant 1 \quad (1 \leqslant m \leqslant k).$$

Thus,

$$\left\{ \sum_{m=1}^{k} \epsilon_m g_m \, ; 0 \leqslant \epsilon_m \leqslant 1 \ (1 \leqslant m \leqslant k) \right\}$$

is a set of best approximations to f. But since $\{g_m^{(r)}\}_{m=1}^{s+1}$ is linearly independent we see that V has r-rank $\geqslant s + 1$ in S.

THEOREM 2. *Theorem* 1 *remains true if*

$$S = \bigotimes_{j=0}^{r} C^q E,$$

where $q \geqslant 1$.

Proof. The condition on V is, of course, still sufficient. For the necessity, observe, first, that in the case $q = 1$, the functions $f_s(x)$ in Theorem 1 are in C^1F and can, thus, be extended to be in C^1E. If $q \geqslant 2$, however, $|g_m^{(s)}(x)|$ is no longer necessarily in C^qE, $m = 1, 2, \ldots, k$. Thus, following Garkavi, we construct functions f_s ($0 \leqslant s \leqslant r$) as follows. If T is as in Theorem 1, let $T_1 = \{(x, s) \in T;\ x \in \text{boundary of } E\}$ and $T_2 = \{(x, s) \in T;\ x \in \text{interior of } E\}$. For each s ($0 \leqslant s \leqslant r$) choose an $f_s(x)$ in C^qE such that

(i) $f_s(x_j) = \operatorname{sgn} c_j$ if $(x_j, s) \in T$;

(ii) $|f_s(x)| < 1$ if $(x, s) \notin T$;

(iii) $f_s^{(1)}(x_j) \neq 0$ if $(x_j, s) \in T_1$;

(iv) $f_s^{(2)}(x_j) \neq 0$ if $(x_j, s) \in T_2$.

As before, $\|f - \tilde{g}\| \geqslant 1$ for all $\tilde{g}$ in V. For each (x_j, s) in T, let $w_j = ([\alpha_j, \beta_j], s)$ be a neighborhood of (x_j, s) such that

(i) $f_s^{(1)}(x) \neq 0$ if $(x, s) \in w_j$ and $(x_j, s) \in T_1$;

(ii) $f_s^{(2)}(x) \neq 0$ if $(x, s) \in w_j$ and $(x_j, s) \in T_2$.

Let

$$E_1{}^s = \bigcup_{(x_j, s) \in T_1} w_j \qquad \text{and} \qquad E_2{}^s = \bigcup_{(x_j, s) \in T_2} w_j .$$

Assume, without loss of generality, that

$$\sup_{(x, s) \in E_1{}^s} k \mid g_m^{(s+1)}(x) \mid < \inf_{(x, s) \in E_1{}^s} \mid f_s^{(1)}(x) \mid$$

and that

$$\sup_{(x, s) \in E_2{}^s} k \mid g_m^{(s+2)}(x) \mid < \inf_{(x, s) \in E_2{}^s} \mid f_s^{(2)}(x) \mid, \qquad m = 1, 2, \ldots, k.$$

By Taylor's formula we have, if $0 \leqslant \mid \epsilon_m \mid \leqslant 1$,

$$f_s(x) - \sum_{m=1}^{k} \epsilon_m g_m^{(s)}(x) = f_s(x_j) + \left[f_s^{(1)}(\tilde{x}) - \sum_{m=1}^{k} \epsilon_m g_m^{(s+1)}(\tilde{x}) \right] (x - x_j),$$

where (x, s) and $(\tilde{x}, s)$ belong to w_j, if $(x_j, s) \in T_1$, and

$$f_s(x) - \sum_{m=1}^{k} \epsilon_m g_m^{(s)}(x) = f_s(x_j) + \frac{1}{2} \left[f_s^{(2)}(\tilde{x}) - \sum_{m=1}^{k} \epsilon_m g_m^{(s+2)}(\tilde{x}) \right] (x - x_j)^2,$$

where (x, s) and $(\tilde{x}, s)$ belong to w_j, if $(x_j, s) \in T_2$. Since $|f_s(x_j)| = 1$, we have that $f_s^{(1)}(x_j) f_s(x_j)(x - x_j) < 0$ if $(x_j, s) \in T_1$ and $f_s^{(2)}(x_j) f_s(x_j) < 0$ if $(x_j, s) \in T_2$. Combining these facts with Taylor's formula and the fact that the first and second derivatives of f_s strongly dominate the first and second derivatives of $\sum_{m=1}^{k} \epsilon_m g_m^{(s)}$ in $E_1{}^s \cup E_2{}^s$, we obtain that

$$\left| f_s(x) - \sum_{m=1}^{k} \epsilon_m g_m^{(s)}(x) \right| \leqslant 1 \qquad \text{for all} \quad (x, s) \in E_1{}^s \cup E_2{}^s.$$

Further, in $X \sim [E_1{}^s \cup E_2{}^s]$, $|f_s(x)| \leqslant \Theta < 1$. Hence, if $\| g_m^{(s)} \|_\infty < (1 - \Theta)/k$, $1 \leqslant m \leqslant k$, we have

$$\left\| f_s - \sum_{m=1}^{k} \epsilon_m g_m^{(s)} \right\|_\infty \leqslant 1, \qquad m = 1, 2,..., k.$$

Hence

$$\left\| f - \sum_{m=1}^{k} \epsilon_m \tilde{g}_m \right\| \leqslant 1,$$

and the conclusion follows as in Theorem 1.

4. Approximation in C^rE

If $r > 0$, the subspace $V = P_{n-1}$ does not satisfy the condition in Theorem 1. In this section we examine the situation in which $f = (f_0, f_1, ..., f_r)$, $f_i = f_0^{(i)}$, $0 \leqslant i \leqslant r$. In this case the sufficient condition of Theorem 1 can be strengthened to include P_{n-1}.

Definition. If $g \in C^rE$ and $0 \leqslant i \leqslant r$, we call a generalized point (x, i), such that $g^{(i)}(x) = 0$, an r-generalized zero of g. Let $g^{(-1)} \equiv 1$. If $g^{(i)}(x) = g^{(i+1)}(x) = 0$, we may call (x, i) an r-generalized new double zero provided we agree that neither $(x, i - 1)$ nor $(x, i + 1)$ may be so labeled. If $g^{(i)}(x) = 0$ and x is a boundary point of E, then (x, i) is called an r-generalized boundary zero of g.

Theorem 3. *Let $C_{r+1}E$ denote the space of functions having an $(r + 1)$st derivative everywhere on E, a finite union of disjoint closed intervals. Suppose that the n-dimensional subspace V satisfies the condition that among the common r-generalized zeros of k ($k = s + 1, s + 2,..., n$) linearly independent elements of V, there are no more that $n - k$ generalized points which are r-generalized new double or boundary zeros of $s + 1$ of these elements whose* rth *derivatives are linearly independent. Then, with respect to the norm* $\|f\| = \max[\|f\|_\infty, \|f^{(1)}\|_\infty, ..., \|f^{(r)}\|_\infty]$, *the dimension of the set of* rth *derivatives of the best approximations in V to any f in $C_{r+1}E$ does not exceed s.*

Proof. We identify $C_{r+1}E$ with a subspace S_* of S of Theorem 2 by letting $f_* = (f, f^{(1)}, ..., f^{(r)})$. We follow a reasoning analogous to that in the sufficiency proof of Theorem 1 after we observe that if $|f^{(j)}(x) - g_i^{(j)}(x)| = \rho$ for x interior to E, then $f^{(j+1)}(x) - g_i^{(j+1)}(x) = 0 \neq \rho$, $i = 1, 2,..., s + 2$. (In the proof of Theorem 1 it is possible that $|f_j(x) - g_i^{(j)}(x)| = \rho$ and

$|f_{j+1}(x) - g_i^{(j+1)}(x)| = \rho$.) Thus, each interior r-generalized zero of $g_i - g_1$ in the support of μ is an r-generalized new double zero, $i = 2, 3,..., s+2$, according to the foregoing definition of an r-generalized new double zero. Hence, μ has a support of $n - k + 1$ ($k \geqslant s + 2$) generalized points, etc. □

Let $P_{n-1}I$ be the space of real polynomials of degree less than or equal to $n-1$ on the interval $I = [a, b]$ and let $\{x_1, x_2, ..., x_k\} \subset I$. Let $\mathscr{L}_i^j$ denote the linear functional on $P_{n-1}I$ defined by $\mathscr{L}_i^j(p) = p^{(j)}(x_i)$. Following Schoenberg [7], let $E = (e_{ij})_{i=1,2,\ldots,k}^{j=0,1,\ldots,n-1}$ be an n-incidence matrix, i.e., each e_{ij} is 0 or 1 and

$$\sum_{i,j} e_{ij} = n.$$

We say that E is poised if the set of n linear functionals $\{\mathscr{L}_i^j;\, e_{ij} = 1\}$ is linearly independent on $P_{n-1}I$. If E is an n-incidence matrix, let

$$m_j = \sum_{i=1}^{k} e_{ij}, \qquad j = 0, 1, ..., n-1,$$

and

$$M_j = \sum_{p=0}^{j} m_p, \qquad j = 0, 1, ..., n-1.$$

Then E is said to satisfy the Pólya conditions if

$$M_j \geqslant j + 1 \qquad \text{for} \quad j = 0, 1, ..., n-1.$$

In the following four theorems we assume that the n-incidence matrix E satisfies the Pólya conditions.

THEOREM A. (Pólya and Whittaker, see [2].) *If $k = 2$, then E is poised.*

THEOREM B. (Ferguson [2, p. 24].) *If $k > 2$, and if $e_{i,j-1} = e_{i,j+p} = 0$, $e_{ij} = \cdots = e_{i,j+p-1} = 1$ implies p is even, then E is poised.*

THEOREM C. (Schoenberg, see [2, p. 25].) *If $x_1 = a$ and $x_k = b$, and if $2 \leqslant i \leqslant k-1$ and $e_{ij} = 1$ imply $e_{ij'} = 1$ for each $j' \leqslant j$, then E is poised.*

By combining Ferguson's proofs of Theorems B and C we can get the following result.

THEOREM 4. *If $x_1 = a$ and $x_k = b$, and if $2 \leqslant i \leqslant k-1$ and $e_{i,j-1} = e_{i,j+p} = 0$, $e_{ij} = \cdots = e_{i,j+p-1} = 1$ imply p is even, then E is poised.*

THEOREM 5 (Moursund [5] and Johnson [4].) *Consider $C_{r+1}I$ with the norm $\|f\| = \max[\|f\|_\infty, \|f^{(1)}\|_\infty, \|f^{(2)}\|_\infty, \ldots, \|f^{(r)}\|_\infty]$, where $I = [a, b]$. If p_1 and p_2 are best approximations in $P_{n-1}I$ to f belonging to $C_{r+1}I$, then $p_1^{(r)} = p_2^{(r)}$.*

Proof. We show that $P_{n-1}I$ satisfies the condition of Theorem 3, where $E = I$ and $s = 0$. The condition in this case can be reworded as follows. If

$$p_1, p_2, \ldots, p_k \in Q_1^\perp = \{\mathscr{L}_{j_1}^{i_1}, \mathscr{L}_{j_2}^{i_2}, \ldots, \mathscr{L}_{j_{n-k+1}}^{i_{n-k+1}}\}^\perp$$

and

$$p_1 \in Q_2^\perp = \{\mathscr{L}_{j_1}^{i_1+1}, \mathscr{L}_{j_2}^{i_2+1}, \ldots, \mathscr{L}_{j_{n-k+1}}^{i_{n-k+1}+1}\}'^\perp,$$

where $Q_1 \cap Q_2$ is empty and all $i_m \leqslant r$ (here $'$ indicates that $\mathscr{L}_{j_m}^{i_m+1}$ is omitted if $x_{j_m} = a$ or $x_{j_m} = b$, and $p \in Q^\perp$ means p is in the nullspace of each of the elements of Q), then either $p_1, p_2, \ldots, p_k$ are linearly dependent or $p_1^{(r)} = 0$, $k = 1, 2, \ldots, n$. Let v be the number of elements in Q_2.

Now fix k. For $0 \leqslant q \leqslant n-1$, let $E^q = (e_{ij})_{i=1,2,\ldots,n-k+1}^{j=q,q+1,\ldots,n-1}$, where $e_{ij} = 1$ if $\mathscr{L}_j^i \in S_q \subset Q_1 \cap Q_2$, and $e_{ij} = 0$ if $\mathscr{L}_j^i \notin S_q$, where S_q will be determined. (Note: $e_{ij} = 0$ if $j > r$.) Let

$$m_j = \sum_{i=1}^{n-k+1} e_{ij}, \qquad j = 0, 1, \ldots, n-1,$$

and

$$N_p = \sum_{j=p}^{r} m_j, \qquad p = 0, 1, 2, \ldots, r.$$

Now, if for some $u \in \{0, 1, 2, \ldots, r\}$, $N_{r-t} < n - (r-t)$ for all $t = 0, 1, \ldots, u-1$, and $N_{r-u} \geqslant n - (r-u)$, then clearly $Q_1 \cup Q_2$ includes a subset S_{r-u} containing $n - (r-u)$ linear functionals such that E^{r-u} satisfies the Pólya conditions on $P_{n-(r-u)-1}$, and, hence, $p_1 \in [Q_1 \cup Q_2]^\perp$ implies $p_1^{(r-u)} = 0$, by Theorem 4.

On the other hand, if $N_{r-t} < n - (r-t)$ for all $t = 0, 1, \ldots, r$, then we can augment the set $Q_1 \cup Q_2$ by adding in $k - v - 1$ linear functionals, for example, $\mathscr{L}_{j_{n-k+2}}^0, \mathscr{L}_{j_{n-k+3}}^0, \ldots, \mathscr{L}_{j_{n-v}}^0$, so that the corresponding n-incidence matrix satisfies the Pólya conditions and the conditions of Theorem 4. Thus, Q_1 is linearly independent, and, hence, $p_1, p_2, \ldots, p_k$ are linearly dependent since $k + (n-k+1) = n+1 > n$. $\square$

5. Other Results

We return now to the general situation of Section 3 described prior to Theorem 1.

The following two theorems reduce to Rubenstein's generalization of Haar's theorem on **R** in case $r = 0$ [3, p. 94].

Theorem 6. *Let $S = \bigotimes_{j=0}^{r} C(E_j)$. Then for $\tilde{V}$ to be s-Tchebycheff in S it is necessary and sufficient that each $s + 1$ linearly independent elements of V have fewer than $n - s$ generalized zeros in common.*

Proof. This follows by applying Rubenstein's generalization of Haar's theorem [3, p. 94] to the space CX of the previous discussion. □

Theorem 7. *Let $S = \bigotimes_{j=0}^{r} C(E_j)$. Then for $\tilde{V}$ to have r-rank s in S, it is necessary and sufficient that each $s + 1$ elements of V whose* rth *derivatives are linearly independent have fewer than $n - s$ generalized zeros in common.*

Proof. Sufficiency (Sketch). Suppose $g_{s+2}^{(r)} - g_1^{(r)}, g_{s+1}^{(r)} - g_1^{(r)},..., g_2^{(r)} - g_1^{(r)}$ are linearly independent, where $\tilde{g}_1, \tilde{g}_2,..., \tilde{g}_{s+2}$ are best approximations in $\tilde{V}$ to f. Hence, there are at most $n - s - 1$ common generalized zeros of $g_{s+2} - g_1,..., g_2 - g_1$, and the proof proceeds analogously to that of Rubenstein's generalization of Haar's theorem on **R**.

Necessity (Sketch). Suppose there exist elements $g_1, g_2,..., g_{s+1}$ in V whose rth derivatives are linearly independent and which have $n - s$ common generalized zeros forming a set $T = \{(x_1, i_1), (x_2, i_2),..., (x_{n-s}, i_{n-s})\}$. Then $\{e_{(x_j, i_j)}\}_{j=1}^{n-s}$ is a linearly dependent system in V^0, and there exist scalars c_j $(1 \leqslant j \leqslant n - s)$ not all zero, such that

$$L = \sum_{j=1}^{n-s} c_j \mathscr{L}_{(x_j, i_j)} = 0,$$

on $\tilde{V}$. Assuming

$$\sum_{j=1}^{n-s} | c_j | = 1,$$

we see that $\| L \|^0 = 1$ by choosing $h \in \bigotimes_{j=0}^{r} C(E_j)$ such that $\| h \| = 1$ and $h_s(x_j) = \operatorname{sgn} c_j$ for all $(x_j, s) \in T$, $s = 0, 1,..., r$. The proof proceeds analogously to that of Rubenstein's theorem. □

Definition. $E_i \leqslant E_j$ means that $x \leqslant y$ for every x in E_i and every y in E_j.

THEOREM D (Ferguson [2, p. 27].) *Let* $E_1 < E_2 < \cdots < E_r$, *and assume that* $E_i \cap E_{i+1}$ *consists of at most one point,* $i = 1, 2,..., r - 1$. *Consider an n-incidence matrix* $E = (e_{ij})_{i=1,2,...,k}^{j=0,1,...,n-1}$, *where* $e_{ij} = 1$ *implies that* $x_i \in E_j$. *Then, if E satisfies the Pólya conditions, E is poised.*

THEOREM 8. *Let* $E_1 ,..., E_r$ *be as in the first sentence of Theorem D, and let* $S = \otimes_{j=0}^{r} C(E_j)$. *If* p_1 *and* p_2 *are in* P_{n-1} *and* $\tilde{p}_1$ *and* $\tilde{p}_2$ *are best approximations to f in S, then* $p_1^{(r)} = p_2^{(r)}$.

Proof. From Theorem D and an argument analogous to that used in the proof of Theorem 5, it follows immediately that P_{n-1} satisfies the condition of Theorem 7 with $s = 0$.

REFERENCES

1. B. L. CHALMERS AND L. O. FERGUSON, Sets of best approximation in certain classes of normed spaces, *J. Approximation Theory* **4** (1971), 194–203.
2. D. FERGUSON, The question of uniqueness for G. D. Birkhoff interpolation problems, *J. Approximation Theory* **2** (1969), 1–28.
3. A. L. GARKAVI, On dimensionality of polytopes of best approximation for differentiable functions, **23** (1959), 93–114.
4. L. JOHNSON, Unicity in approximation of a function and its derivatives, *Math. Comp.* **22** (1968), 873–875.
5. D. G. MOURSUND, Chebyshev approximation of a function and its derivatives, *Math. Comp.* **18** (1964), 382–389.
6. T. J. RIVLIN AND H. S. SHAPIRO, Some uniqueness problems in approximation theory, *Comm. Pure Appl. Math.* **13** (1960), 35–47.
7. I. J. SCHOENBERG, On Hermite–Birkhoff interpolation, *J. Math. Anal. Appl.* **16** (1966), 538–543.
8. I. SINGER, "Best Approximation in Normed Linear Spaces by Elements of Linear Subspaces," Springer-Verlag, New York, 1970.