BEST APPROXIMATIONS AND INTERPOLATING FUNCTIONS

BY

JOHN R. RICE[1]

1. **Introduction.** Let $f(x)$ be a continuous function on $[0, 1]$ and let $P_n(x)$ be the best least-squares nth degree polynomial approximation to $f(x)$. It is a well-known fact that $P_n(x)$ interpolates the values of $f(x)$ in at least $n+1$ points. This phenomenon is also present with other than least-squares approximations. On the other hand a classical method of obtaining polynomial approximations to $f(x)$ is to interpolate the values of $f(x)$ at a certain set of $n+1$ points. A function $F(a, x)$ depending on n parameters is said to be an interpolating function of $f(x)$ if $F(a, x)$ interpolates $f(x)$ in at least n points. Thus it appears that there is an intimate relation between best approximating polynomials and interpolating polynomials. This relation has been investigated for L_p norms by Motzkin and Walsh for finite point sets [6; 7; 8] and a closed bounded interval [9].

The purpose of this paper is to investigate the extent of this relation in more general situations. Thus an answer is sought to the following questions. What are the properties of the norm and of the approximating functions which force best approximations to be interpolation functions? How general is this connection between best approximations and interpolating functions? What properties of the norm and the interpolating are required in order that an interpolating function of $f(x)$ also be a "best" approximation to $f(x)$?

Not all of these questions are completely answered here. Much more success is achieved in showing that best approximations are interpolating functions than in the converse, i.e., in showing that interpolating functions are best approximations. More is accomplished for finite point sets than for the interval $[0, 1]$.

Three classes—Class 1, Class 2, and Class 3—of norms are considered. A norm belongs to Class 2, Class 3, or Class 1 according to whether it emphasizes values away from zero, at zero or neither. Typical members of these classes are the L_1 norm for Class 1, the Tchebycheff norm for Class 2 and the L_p norm, $0<p<1$, for Class 3. For detailed definition of these classes see §§2 and 4. Not all norms are included in these three classes.

The approximating functions $F(a, x)$ that are considered are unisolvent functions of variable degree [12] (called varisolvent functions) and linear approximating functions, i.e.

Presented to the Society, September 2, 1960; received by the editors September 29, 1960.

(1) This work was done under a National Research Council—National Bureau of Standards Postdoctoral Research Associateship.

$$P(a, x) = \sum_{i=1}^{n} a^i \phi_i(x)$$

where the $\phi_i(x)$ form a Tchebycheff set. The concept of a regular varisolvent function is introduced in §2 and regularity is an essential element in the analysis. Roughly speaking, F is regular if

$$\operatorname*{Lim}_{n\to\infty} F(a_n, x) = F(a^*, x)$$

implies that the rate of convergence is uniform in x. Linear approximating functions are regular.

The principal result for the interval $[0, 1]$ is that best approximations in the Class 1 and Class 2 norms by regular varisolvent functions are interpolating functions. The exact nature of the interpolation is different for each class however. For a finite point set the definition of an interpolating function is slightly modified to allow "interpolation" at points not in the set. It is shown for finite point sets that best approximations in any monotonic norm by any varisolvent approximating function are interpolating functions. The nature of the interpolation is studied for Class 1, Class 2, and Class 3 norms and regular varisolvent approximating functions.

The relation between best approximations and interpolating functions is very well defined for linear approximating functions, finite point sets and weighted L_p norms. The principal result states that for any $f(x)$:

1. The set of best approximations in a weighted L_p norm, $1 < p \leqq \infty$ is identical with the set of strongly interpolating functions.
2. The set of best approximations in a weighted L_1 norm is identical with the set of weakly interpolating functions.
3. The set of best approximations in a weighted L_p norm, $0 < p < 1$, is identical with the set of exactly interpolating functions.

Much more general results are obtained for the Tchebycheff norm.

The results for linear approximating functions and weighted L_p norms on a finite point set are very similar to the recent results of Motzkin and Walsh. Most of their results are contained here in the form of special cases. Some of the results are related to the much older work of D. Jackson [3; 4].

There remain two major areas of open problems. The first and most interesting is that of showing that interpolating functions on the interval $[0, 1]$ are best approximations. A related problem is to show that interpolating varisolvent, or unisolvent [5], functions on a finite point set are best approximations. The second problem area is for Class 3 norms. Only very meager results are available for the interval $[0, 1]$ and no results are known for unisolvent or varisolvent functions on a finite point set.

Recently Motzkin and Walsh [10] have considered generalizations of their work for linear approximation functions on the interval $[0, 1]$. They use the norm

$$\|f(x) - P(a, x)\| = \int_0^1 \tau(|f(x) - P(a, x)|)\omega(x)dx$$

where $\tau(t)$ is twice continuously differentiable and $\omega(x)$ is positive almost everywhere. The similarity of the results obtained here and in [10] indicates that there is a strong connection between the class of the norm (as defined here) and the behavior of $\tau'(t)$ at and near zero.

2. **Preliminaries.** Euclidean n-dimensional space is denoted by E_n; points in E_n are denoted by $a, b, \cdots$ and the coordinates of a are $a^1, a^2, \cdots, a^n$. The absolute value of a is defined by $|a|^2 = \sum_{i=1}^n |a^i|^2$. Curly brackets, $\{\ \}$, denote a set or sequence and $\{x| \cdots\}$ is read as "the set of x such that $\cdots$."

The real function $F(a, x)$ is defined for $x \in [0, 1]$ and $a \in P$ where P is an arcwise connected subset of E_n. F is continuous in the sense that given $a_0 \in P$, $x_0 \in [0, 1]$ and $\epsilon > 0$, then there is a $\delta > 0$ such that $a \in P$, $x \in [0, 1]$ and $|a_0 - a| + |x_0 - x| < \delta$ implies that $|F(a_0, x_0) - F(a, x)| < \epsilon$. It is assumed that if $a \neq a^*$ then $F(a, x) \neq F(a^*, x)$ for some $x \in [0, 1]$. $f(x)$ denotes a continuous function on $[0, 1]$.

This paper is concerned with functions F which are *varisolvent* functions. This is a generalization of the concept of *unisolvent* functions [5], and a shortened name for functions unisolvent of variable degree [12]. In order to define varisolvent functions precisely, the following definitions are required.

Definition 2.1. *F is solvent of degree m at $a^* \in P$ if given a set $\{x_j | x_j < x_{j+1},\ j = 1, 2, \cdots, m,\ x_j \in [0, 1]\}$ and $\epsilon > 0$, then there is a $\delta(a^*, \epsilon, x_1, \cdots, x_m) > 0$ such that $|y_j - F(a^*, x_j)| < \delta$, $j = 1, 2, \cdots, m$ implies that there is a solution $a \in P$ to*

$$F(a, x_j) = y_j, \qquad j = 1, 2, \cdots, m,$$

with $\max_{x \in [0,1]} |F(a, x) - F(a^*, x)| < \epsilon$.

Definition 2.2. *F has property z of degree m at $a^* \in P$ if $a \neq a^*$ implies that $F(a^*, x) - F(a, x)$ has at most $m-1$ zeros in* $[0, 1]$.

Definition 2.3. *F is unisolvent of degree m at $a^* \in P$ if F has property z of degree m at a^* and F is solvent of degree m at a^*.*

If F is unisolvent of degree $m(a)$ for every $a \in P$ then F is said to be a *varisolvent* function. The degree of unisolvence (or simply the degree) of F at a is denoted by $m(a)$.

The first part of this paper is concerned with best approximations in a general norm. Fejér [2] has defined a *monotonic norm* for a finite point set and for approximation by polynomials. This norm is suitable for the study of approximations on a finite point set. The norm to be defined here may be considered an extension of the monotonic norm to the interval $[0, 1]$. The norm of a function $f(x)$ is denoted by $\|f(x)\|$. The norm is assumed to have the following properties:

1. *It is defined for all piecewise continuous functions on* $[0, 1]$.

2. *The norm of* $f(x)$ *is the same as the norm of* $|f(x)|$.

3. *Given* $g(x)$ *let* $A=\{x \mid |g(x)| \geqq K \max_{x\in[0,1]} |g(x)|\}$, $0<K<1$. *If* $|h(x)| = |g(x)|$ *for* $x \notin A$ *and* $|g(x)| < |h(x)|$ *for* $x\in A$ *then* $\|g(x)\| > \|h(x)\|$.

If the sequence $\{f_n(x) \mid n=1, 2, \cdots\}$ converges to $f(x)$, the convergence is said to be *regular* if the rate of convergence is "uniform" with respect to x. This is stated more precisely as follows: Let Z be the closure of the set

$$\{x \mid f(x) = f_n(x) \text{ for some } n\};$$

let $\epsilon_n = \max_{x\in[0,1]} |f(x)-f_n(x)|$ and let I be a closed subinterval of $[0, 1]$ such that $I\cap Z$ is empty. The sequence $\{f_n(x)\}$ is said to *converge regularly* to $f(x)$ if

$$\min_{x\in I} |f_n(x) - f(x)| \geqq K\epsilon_n$$

for some positive K, which may depend on I.

A varisolvent function is said to be regular if

$$\operatorname{Lim}_{n\to\infty} F(a_n, x) = F(a, x)$$

for each x implies that the convergence is regular.

All approximating functions of the form

$$P(a, x) = \sum_{i=1}^{n} a^i\phi_i(x)$$

are regular. An example is given which shows that regularity is essential for the conclusions of this paper to be valid.

We introduce two special classes of norms on the interval $[0, 1]$. The properties of these norms may be stated intuitively as follows. A Class 2 norm emphasizes values of $f(x)$ away from zero and a Class 1 norm places no special emphasis on any values of $f(x)$. The typical examples of a Class 2 norm is

$$\|f(x)\|^2 = \int_0^1 |f(x)|^2 dx$$

and the typical example of a Class 1 norm is

$$\|f(x)\| = \int_0^1 |f(x)| \, dx.$$

In order to define these classes precisely, consider a function $g(x)$ and a sequence $\{g_n(x)\}$ which converges regularly to $g(x)$. Let $r(n)$ tend to zero as n tends to infinity and set

$$R_n = \{x \mid |g(x)| \leqq r(n)\}.$$

Assume that $0<g_n(x)/g(x)\leq 1$ for $x\notin R_n$. Further let Y be a fixed subset of $[0, 1]$ with following properties:

(i) $0<g_n(x)/g(x)<1$ for all n and $x\in Y$,

(ii) Y is the finite union of closed intervals,

(iii) if $g(x_0)=\max_{x\in[0,1]}|g(x)|$ then x_0 is in the interior of Y.

Set

$$s_n(x) = \begin{cases} g_n(x), & x \in Y, \\ g(x), & x \notin Y, \end{cases}$$

$$t_n(x) = \begin{cases} g_n(x), & x \in R_n, \\ g(x), & x \notin R_n. \end{cases}$$

A norm is said to be in *Class* 2 if

$$\lim_{n\to\infty} \frac{\|t_n(x) - g(x)\|}{\|s_n(x)\| - \|g(x)\|} = 0$$

for all possibilities described in the preceding paragraph. Make the further assumption that the measure of the set $\{x|g(x)=0\}$ is zero; then the norm is said to be of *Class* 1 if

$$\lim_{n\to\infty} \frac{\|t_n(x) - g(x)\|}{\|s_n(x)\| - \|g(x)\|} = 0.$$

Note that in the *Class* 1 norms the numerator will tend to zero because R_n tends to zero, as well as because $\{g_n(x)\}$ converges to $g(x)$.

A function $g(x)$ is said to have n strong sign changes on a set X if there are $n+1$ points $\{x_j|j=1, 2, \cdots, n+1; x_j<x_{j+1}; x_j\in X\}$ such that

$$g(x_1) > 0, \qquad g(x_2) < 0, \qquad g(x_3) > 0, \cdots$$

or

$$g(x_1) < 0, \qquad g(x_2) > 0, \qquad g(x_3) < 0, \cdots.$$

If the strict inequality signs are weakened to allow equality, then $g(x)$ is said to have ***n* weak sign changes** on X. A zero x_0 of $g(x)$ is said to be a simple zero if $g(x)$ changes sign at x_0 and a double zero if $g(x)$ does not change sign at x_0.

A varisolvent function $F(a, x)$ is said to be a *strongly interpolating function* of $f(x)$ if $F(a, x)-f(x)$ has $m(a)$ strong sign changes on $[0, 1]$. $F(a, x)$ is a ***weakly interpolating function*** of $f(x)$ if $F(a, x)-f(x)$ has $m(a)$ weak sign changes on $[0, 1]$.

3. Interpolating functions and best approximations on $[0, 1]$. We begin by establishing a lemma about varisolvent functions. The proof of this lemma is only indicated here; for complete details the reader is referred to [11]. We need

DEFINITION 3.1. *F has property* A *if given $a^*\in P$, $k < m(a^*)$, $\{x_j | 0 = x_0 < x_1 < \cdots < x_{k+1} = 1\}$ and ϵ with $0 < \epsilon < \min_j (x_{j+1} - x_j)$, $j=0, 1, \cdots, k$ then*

(i) *there are $a_1, a_2\in P$ such that for $x\in[0, 1]$, $F(a^*, x)-\epsilon<F(a_1, x)<F(a^*, x)<F(a_2, x)<F(a^*, x)+\epsilon$;*

(ii) *there are $a_3, a_4\in P$ such that $|F(a_3, x)-F(a^*, x)|<\epsilon$, $|F(a_4, x)-F(a^*, x)|<\epsilon$ for $x\in[0, 1]$ and $F(a_3, x)-F(a^*, x)$, $F(a_4, x)-F(a^*, x)$ change sign from $x_j-\epsilon$ to $x_j+\epsilon$, $j=1, 2, \cdots, k$ and have no zeros outside $[x_j-\epsilon, x_j+\epsilon]$. Further $F(a_3, 0)>F(a^*, 0)>F(a_4, 0)$.*

LEMMA 3.1. *If F is a varisolvent function then F has property* A.

Proof. It follows from Theorem 7 [11] that F has property NS of degree $m(a)$ at each point $a\in P$. The proof of Lemma 4 [11] may now be applied directly to establish this lemma.

It is noted that in many instances in this paper property A is sufficient for the proofs even though varisolvence has been stated as a hypothesis. Property A does not imply varisolvence.

THEOREM 3.1. *Let F be a regular varisolvent function and let the norm be of Class 1. If $F(a^*, x)$ is a best approximation of $f(x)$, then either $F(a^*, x)-f(x)$ vanishes identically on a set of positive measure or $F(a^*, x)-f(x)$ has $m(a^*)$ strong sign changes.*

Proof. Assume that $F(a^*, x)-f(x)$ does not vanish on a set of positive measure and that $F(a^*, x)-f(x)$ has $k<m(a^*)$ simple zeros, $0\leqq y_1<y_2<\cdots<y_k\leqq 1$. Since F is varisolvent, F has property A by Lemma 1. In Definition 3.1 let $\epsilon=2^{-n}$, $x_j=y_j$ (if $y_1=0$ take $x_1=2^{-n-1}$, if $y_k=1$ take $x_k=1-2^{-n-1}$) and determine $a_n\in P$ such that $|F(a_n, x)-F(a^*, x)|<2^{-n}$, $F(a_n, x)$ changes sign in $[x_j-2^{-n}, x_j+2^{-n}]$, $j=1, 2, \cdots, k$ and has no zeros outside these intervals, and such that $\max_{x\in[0,1]}|F(a^*, x)-f(x)|>\max_{x\in[0,1]}|F(a_n, x)-f(x)|$. Let

$$R_n = \{x \,|\, |F(a^*, x) - f(x)| \leqq 2^{-n}\}.$$

Since both F and $f(x)$ are continuous, it follows that the measure of R_n tends to zero as n tends to infinity. R_n will include points in the neighborhood of any double zeros of $F(a^*, x)-f(x)$. Further let

$$Y = \left\{x \,\Big|\, |F(a^*, x) - f(x)| \geqq \frac{1}{2} \max_{x\in[0,1]} |F(a^*, x) - f(x)|\right\}$$

and let

$$s_n(x) = \begin{cases} F(a_n, x) - f(x), & x \in Y, \\ F(a^*, x) - f(x), & x \notin Y, \end{cases}$$

$$t_n(x) = \begin{cases} F(a_n, x) - f(x), & x \in R_n, \\ F(a^*, x) - f(x), & x \notin R_n. \end{cases}$$

Now consider $\|F(a^*, x)-f(x)\|-\|F(a_n, x)-f(x)\|$ which equals

$$\|F(a^*, x) - f(x)\| - \|s_n(x)\| + \|s_n(x)\| - \|s_n(x) + t_n(x) + f(x) - F(a^*, x)\| \\ + \|s_n(x) + t_n(x) + f(x) - F(a^*, x)\| - \|F(a_n, x) - f(x)\|.$$

Since

$$| s_n(x) + t_n(x) + f(x) - F(a^*, x) | \geqq | F(a_n, x) - f(x) |,$$

it follows that

$$\|s_n(x) + t_n(x) + f(x) - F(a^*, x)\| - \|F(a_n, x) - f(x)\| \geqq 0.$$

It is seen that

$$\left| \frac{\|s_n(x)\| - \|s_n(x) + t_n(x) + f(x) - F(a^*, x)\|}{\|F(a^*, x) - f(x)\| - \|s_n(x)\|} \right| \leqq \left| \frac{\|t_n(x) + f(x) - F(a^*, x)\|}{\|F(a^*, x) - f(x)\| - \|s_n(x)\|} \right|.$$

Since the norm is of Class 1 it follows that the above expression tends to zero as n tends to infinity. Furthermore, by the definition of a monotonic norm it follows that

$$\|F(a^*, x) - f(x)\| - \|s_n(x)\| > 0;$$

hence for n sufficiently large

$$\|F(a^*, x) - f(x)\| - \|F(a_n, x) - f(x)\| > 0.$$

This contradicts the fact that $F(a^*, x)$ is a best approximation to $f(x)$. Thus the original assumption of this proof is false and this establishes the theorem.

Let $\{\phi_i(x) | i=1, 2, \cdots, m\}$ form a Tchebycheff set [1, Chapter 2] and set

$$P(a, x) = \sum_{i=1}^{n} a^i\phi_i(x)$$

for $a \in E_n$. It is clear that $P(a, x)$ is a regular varisolvent function.

Corollary 3.1. *Let the norm be of Class* 1. *If $P(a^*, x)$ is a best approximation to $f(x)$ then either $P(a^*, x)-f(x)$ vanishes identically on a set of positive measure or $P(a^*, x)-f(x)$ has n strong sign changes.*

This corollary should be compared with Theorem 2 of [10].

A very special case of Theorem 3.1 has been established by D. Jackson [4]. He considered approximation by polynomials in the L_1 norm:

$$\|f(x)\| = \int_c^1 |f(x)| \, dx.$$

The L_1 norm is the best known example of a Class 1 norm which is not also of Class 2. In the special case of this norm a result stronger than Theorem 3.1 may be established. We define a number p_k as follows:

$$p(a_1, a_2) = \int_0^1 | F(a_1, x) - F(a_2, x) | \, dx \Big/ \max_{x \in [0,1]} | F(a_1, x) - F(a_2, x) |$$

and

$$p_k = \inf\{ p(a_1, a_2) \mid F(a_1, x) - F(a_2, x) \text{ has exactly } k \text{ zeros} \}.$$

We can now state

THEOREM 3.2. *Let F be a varisolvent function and let $F(a^*, x)$ be a best approximation to $f(x)$ in the L_1 norm. If $F(a^*, x) - f(x)$ has exactly k strong sign changes then it must vanish on a subset of $[0, 1]$ of measure greater than $p_k/2$.*

Proof. Let $F(a_n, x)$ be such that $F(a_n, x) - F(a^*, x)$ vanishes at the simple zeros of $F(a^*, x) - f(x)$ and such that $F(a_n, x) - f(x)$ has weakly the same sign as $F(a^*, x) - f(x)$ and $\max_{x \in [0,1]} | F(a_n, x) - F(a^*, x) | \leqq 2^{-n}$. Set

$$R_n = \{ x \mid | F(a^*, x) - f(x) | \leqq 2^{-n} \}$$

and let η_n be the measure of R_n. Then $\lim_{n \to \infty} \eta_n$ is the measure of the set

$$\{ x \mid | F(a^*, x) - f(x) | = 0 \}.$$

Let I_n denote the set $[0, 1] - R_n$. Then

$$\begin{aligned}
\int_{I_n} & | F(a^*, x) - f(x) | \, dx - \int_{I_n} | F(a_n, x) - f(x) | \, dx \\
&= \int_{I_n} | F(a^*, x) - F(a_n, x) | \, dx \\
&= \int_0^1 | F(a^*, x) - F(a_n, x) | \, dx - \int_{R_n} | F(a^*, x) - F(a_n, x) | \, dx \\
&\geqq \int_0^1 | F(a^*, x) - F(a_n, x) | \, dx - \eta_n \max_{x \in [0,1]} | F(a^*, x) - F(a_n, x) | .
\end{aligned}$$

Also

$$\begin{aligned}
\int_{R_n} & | F(a_n, x) - f(x) | \, dx \\
&= \int_{R_\infty} | F(a_n, x) - F(a^*, x) | \, dx + \int_{R_n - R_\infty} | F(a_n, x) - f(x) | \, dx \\
&\leqq 2\eta_n - \eta_\infty \max_{x \in [0,1]} | F(a_n, x) - F(a^*, x) | .
\end{aligned}$$

Hence

$$\int_0^1 |F(a^*, x) - f(x)|\,dx - \int_0^1 |F(a_n, x) - f(x)|\,dx$$

$$\geqq \int_0^1 |F(a^*, x) - F(a_n, x)|\,dx - 3\eta_n - \eta_\infty \max_{x\in[0,1]} |F(a^*, x) - F(a_n, x)|\,.$$

Since $F(a^*, x)$ is a best approximation to $f(x)$, it follows that

$$\int_0^1 |F(a^*, x) - F(a_n, x)|\,dx - 3\eta_n - \eta_\infty \max_{x\in[0,1]} |F(a^*, x) - F(a_n, x)| \leqq 0$$

or

$$3\eta_n - \eta_\infty \geqq \int_0^1 |F(a^*, x) - F(a_n, x)|\,dx \Big/ \max_{x\in[0,1]} |F(a^*, x) - F(a_n, x)|.$$

This implies that

$$\operatorname*{Lim}_{n\to\infty} \eta_n \geqq \frac{1}{2}\, p_k$$

which establishes the theorem.

This theorem is not quite comparable with Theorem 3.1, since regularity is not involved. Regularity would seem to imply that the p_k are positive for $k<m(a^*)$ and Theorem 3.1 would follow for the L_1 norm. However, for non-regular functions the p_k may all be zero, in which case Theorem 3.2 would lose its significance. Motzkin and Walsh [9, Theorem 6] have established Theorem 3.2 for polynomials. In this case the p_k may be specifically calculated.

For Class 2 a theorem very similar to Theorem 3.1 may be established. Theorem 3.3 has been established for polynomial approximations in the L_p norm $1<p<\infty$ by Jackson [3].

THEOREM 3.3. *Let F be a regular varisolvent function and let the norm be of Class 2. If $F(a^*, x)$ is the best approximation to $f(x)$ then $F(a^*, x)$ is a strongly interpolating function of $f(x)$.*

Proof. Assume that $F(a^*, x)-f(x)$ has $k<m(a^*)$ strong sign changes and let $0<y_1<y_2<\cdots<y_k\leqq 1$ be the simple zeros of $F(a^*, x)-f(x)$. From property A and Lemma 1 let $\epsilon=2^{-n}$, $x_j=y_j$ (if $y_1=0$ take $x_1=2^{-n-1}$, if $y_k=1$ take $x_k=1-2^{-n-1}$) and determine $a_n\in P$ such that $|F(a_n, x)-F(a^*, x)| <2^{-n}$, $F(a_n, x)-F(a^*, x)$ changes sign in $[x_j-2^{-n}, x_j+2^{-n}]$, $j=1, 2, \cdots, k$ and has no zeros outside these intervals, and such that $\max_{x\in[0,1]} |F(a^*, x) -f(x)| >\max_{x\in[0,1]} |F(a_n, x)-f(x)|$.

Set

$$R_n = \{x \mid |F(a^*, x) - f(x)| \leqq 2^{-n}\}$$

and

$$Y = \left\{x \,\Big|\, |F(a^*, x) - f(x)| \geqq \frac{1}{2} \max_{x\in[0,1]} |F(a^*, x) - f(x)|\right\}.$$

With

$$s_n(x) = \begin{cases} F(a_n, x) - f(x), & x \in Y, \\ F(a^*, x) - f(x), & x \notin Y, \end{cases}$$

$$t_n(x) = \begin{cases} F(a_n, x) - f(x), & x \in R_n, \\ F(a^*, x) - f(x), & x \notin R_n, \end{cases}$$

consider

$$\begin{aligned} &\|F(a^*, x) - f(x)\| - \|F(a_n, x) - f(x)\| \\ &\quad = \|F(a^*, x) - f(x)\| - \|s_n(x)\| + \|s_n(x)\| - \|s_n(x) + t_n(x) + f(x) - F(a^*, x)\| \\ &\quad\quad + \|s_n(x) + t_n(x) + f(x) - F(a^*, x)\| - \|F(a_n, x) - f(x)\|. \end{aligned}$$

Since the norm is of Class 2, it follows that

$$\operatorname*{Lim}_{n\to\infty} \frac{\|t_n(x) + f(x) - F(a^*, x)\|}{\|F(a^*, x) - f(x)\| - \|s_n(x)\|} = 0$$

and hence for n sufficiently large

$$\|F(a^*, x) - f(x)\| - \|F(a_n, x) - f(x)\| > 0.$$

This is a contradiction and the original assumption that $F(a^*, x) - f(x)$ had $k < m(a^*)$ zeros is false.

COROLLARY 3.2. *Let the norm be of Class 2. If $P(a^*, x)$ is a best approximation to $f(x)$, then $P(a^*, x) - f(x)$ has n strong sign changes.*

An example is now given that illustrates the effect of nonregularity. Let the continuous function $\phi(a, x)$ be defined so that $\phi(a, x) = a^6$ in the intervals $[0, 1/2 - a^2]$, $[1/2 + a^2, 1]$, $\phi(a, 1/2) = a + a^6$ and $\phi(a, x)$ is linear in the remaining intervals. $\phi(a, x)$ is a nonregular unisolvent function of degree 1.

We compute the best approximation to the function $|x - 1/2|$ in the least squares norm. We have

$$\frac{3}{2}\left\|\left|x - \frac{1}{2}\right| - \phi(a, x)\right\|^2 = \frac{1}{8} + a^4 - a^5 - 3a^6/4 - 3a^9 + 6a^{10} + 3a^{12}/2.$$

Clearly the minimum is attained for $a = 0$, and $\phi(0, x)$ is the best approximation to $|x-1/2|$. It is seen that $|x-1/2|-\phi(0, x)$ has only one zero change. Thus it is seen that a simple nonregular approximating function need not satisfy Theorem 3.3.

The L_1 norm is a well-known example of a norm which is of Class 1 but not of Class 2 and for which Theorem 3.3 is not valid.

4. **Interpolation and approximation on finite point sets.** Let X denote a finite set of distinct points $\{x_j|j=1, 2, \cdots, M\}$ in $[0, 1]$. The remainder of this paper is primarily concerned with approximations to $f(x)$ on X. It is still assumed that all functions are defined on $[0, 1]$, even though in some cases this is not significant. In particular, the definition of a varisolvent function is unchanged. It is assumed that $f(x)$ is not identically equal on X to any approximating function.

The definitions of the classes of norms are somewhat simplified and another class is considered. Given $g(x)$ and a sequence $\{g_n(x)\}$ which converges regularly to $g(x)$, let

$$R = \{x \mid g(x) = 0,\ x \in X\}.$$

Assume that $0<g_n(x)/g(x)<1$ for $x\notin R$. Set

$$s_n(x) = \begin{cases} g_n(x), & x \notin R, \\ g(x), & x \in R, \end{cases}$$

$$t_n(x) = \begin{cases} g_n(x), & x \in R, \\ g(x), & x \notin R. \end{cases}$$

Note that the definition of $g(x)$, $\{g_n(x)\}$, $\{s_n(x)\}$ and $\{t_n(x)\}$ outside X is irrelevant since the norm is a function of values on X only.

A norm is of Class 2 if

$$\operatorname*{Lim}_{n\to\infty} \frac{\|g(x) - t_n(x)\|}{\|g(x)\| - \|s_n(x)\|} = 0$$

for all $g(x)$ and $\{g_n(x)\}$. A norm is of Class 1 if

$$\left| \frac{\|g(x) - t_n(x)\|}{\|g(x)\| - \|s_n(x)\|} \right| \leqq K$$

where K may depend on $g(x)$ and $\{g_n(x)\}$. These definitions are the direct analogs of those given in §2, and perhaps they show the intuitive nature of the classes better. If it is assumed that $g_n(x)\neq g(x)$ for all n and some $x\in R$, then a third class of norms may be defined. A norm is said to be of Class 3 if, with the above assumption,

$$\operatorname*{Lim}_{n\to\infty} \frac{\|g(x) - s_n(x)\|}{\|g(x)\| - \|t_n(x)\|} = 0.$$

The typical example of this class is the L_p norm, $0<p<1$. This class of norms can also be defined for the interval $[0, 1]$, but so far no general results have been obtained for this class. Motzkin and Walsh [9, Theorem 7] have announced a very specialized result which indicates that the situation for Class 3 norms is quite complex. Even for finite point sets the results are less complete for this class than the other two.

Strongly and weakly interpolating functions for a finite point set are defined in the same way as for the interval $[0, 1]$. In addition, $F(a, x)$ is said to be an exactly interpolating function of $f(x)$ on X if $F(a, x)-f(x)$ has $m(a)$ zeros in X. If $F(a, x)$ is a strongly interpolating function of $f(x)$ on X, then $F(a, x)$ is said to *strongly interpolate* $f(x)$ on X. Similar conventions are used for weakly and exactly interpolating functions.

The analysis begins with a lemma similar to Lemma 3.1. This lemma may also be stated for the interval $[0, 1]$ and is probably true there, but this has not been established.

LEMMA 4.1. *Let F be a varisolvent function and let $a^*\in P$, $\{y_j \mid j=1, 2, \cdots, k;\ k<m(a^*);\ y_j\in X\}$ and $\epsilon>0$ be given. Then there are $a_1, a_2\in P$ such that* (i) *$F(a^*, x)-F(a_1, x)$, $F(a^*, x)-F(a_2, x)$ change sign at the y_j and have no other sign changes in X,* (ii) *$|F(a^*, x) - F(a_1, x)| \leqq \epsilon$, $|F(a_2, x)-F(a^*, x)| \leqq \epsilon$,* (iii) *either $F(a_1, x)-F(a^*, x)=0$, $F(a_2, x)-F(a^*, x)=0$, or*

$$\operatorname{sgn}[F(a_1, x) - F(a^*, x)] = - \operatorname{sgn}[F(a_2, x) - F(a^*, x)]$$

for $x\in X$.

Proof. Let $[c, d]$ be an interval in $[0, 1]$ containing no point of X. If $m(a^*)-k$ is odd, choose $m(a^*)-1$ distinct points $\{y_j \mid j=k+1, \cdots, m(a^*)-1\}$ in the interior of $[c, d]$. Determine $a_1, a_2\in P$ by Definition 2.1 such that (1) $F(a^*, y_j)=F(a_1, y_j)=F(a_2, y_j)$, $j=1, \cdots, m(a^*)-1$; (2) for some $x_0\in X$ $F(a_1, x_0)<F(a^*, x_0)<F(a_2, x_0)$; (3) $\max_{x\in[0,1]} |F(a^*, x)-F(a_1, x)| \leqq \epsilon$, $\max_{x\in[0,1]} |F(a^*, x)-F(a_2, x)| \leqq \epsilon$.

$F(a_1, x)$ and $F(a_2, x)$ clearly satisfy (ii) and since $F(a^*, x)-F(a_1, x)$, $F(a^*, x)-F(a_2, x)$ can have at most $m(a^*)-1$ zeros, (i) is satisfied. Note that

$$\operatorname{sgn}[F(a_1, c) - F(a^*, c)] = \operatorname{sgn}[F(a_1, d) - F(a^*, d)].$$

It follows from (2) of the construction that (iii) is satisfied.

If $m(a^*) - k$ is even and $x_1 \neq 0$, $x_M \neq 1$ again choose $[c, d]$ and $\{y_i \mid i=k+1, \cdots, m(a^*)-2\}$ as above. In addition choose z_0 such that $z_0<x_1$ and choose x_0 such that $x_M<x_0\leqq 1$. Determine $a_1, a_2\in P$ by Definition 2.1 such that (1), (2), and (3) of the preceding construction are satisfied, except that y_{k+1} is taken as x_0. Then $F(a_1, x)-F(a^*, x)$, $F(a_2, x)-F(a^*, x)$ cannot have another sign change in $[x_1, x_M]$ without two sign changes, which would contradict property z.

If $x_1=0$ and $x_M=1$ and $x_1\in\{y_j\}$, $x_M\in\{y_j\}$, then x_0, z_0 and $[c, d]$ may be chosen as in the preceding paragraph.

If $x_1=0$ and $x_1\notin\{y_j\}$ consider a neighborhood $N(a^*)$ of a^* such that $a\in N(a^*)$ implies that the degree of F is at least $m(a^*)$. The existence of such a neighborhood is given in [11, Theorem 2]. Add the point $x_1=0$ to the set $\{y_j\}$ and construct a_1', $a_2'\in N(a^*)$ satisfying the requirements of the lemma by the construction for $m(a^*)-k$ odd, with ϵ replaced by $\epsilon/2$. With the original set of $\{y_j\}$ and with z_1 in the interior of $[x_1, x_2]$ we may determine a_1, $a_2\in P$ such that

(1) $|F(a_1, x)-F(a_1', x)|<\epsilon/2$, $|F(a_2, x)-F(a_2', x)|<\epsilon/2$;

(2) $F(a_1, y_j)=F(a_1', y_j)$, $F(a_2, y_j)=F(a_2', y_j)$, $j=1, 2, \cdots, m(a^*)-2$;

(3) $F(a_1, z_1)<F(a_1', z_1)<F(a^*, z_1)<F(a_2', z_1)<F(a_2, z_1)$;

(4) $F(a_1, 0)<F(a^*, 0)<F(a_2, 0)$.

It may be verified that a_1 and a_2 satisfy the conditions of the lemma.

The foregoing construction is also valid when $x_M=1$ and $x_M\notin\{y_j\}$. This concludes the proof of the lemma.

The next theorem may be considered the analog of Theorem 3.1 for Class 1 norms. However, it is seen that no assumption as to the class of the norm or as to regularity need be made. The theorem can be established for the interval $[0, 1]$, if Lemma 4.1 can be established for this interval.

THEOREM 4.1. *Let F be a varisolvent function and let the norm be monotonic. If $F(a^*, x)$ is a best approximation to $f(x)$ on X, then $F(a^*, x)$ is a weakly interpolating function of $f(x)$.*

Proof. Assume that $F(a^*, x)-f(x)$ has $k<m(a^*)$ weak sign changes on X. Let

$$d = \min_j \{ | F(a^*, x_j) - f(x_j) | \, | F(a^*, x_j) - f(x_j) \neq 0\}$$

and determine $a_1\in P$ by Lemma 4.1 such that

(i) $\max_{x\in X} |F(a_1, x)-F(a^*, x)| \leq d/2$,

(ii) $F(a^*, x_j)-f(x_j)=0$ implies $F(a_1, x_j)-f(x_j)=0$,

(iii) $F(a_1, x)-f(x)$ has weakly the same sign as $F(a^*, x)-f(x)$ for $x\in X$.

Clearly $|F(a^*, x_j)-f(x_j)|>|F(a_1, x_j)-f(x_j)|$ if $F(a^*, x_j)-f(x_j)\neq 0$, and since the norm is monotonic, $\|F(a^*, x)-f(x)\|>\|F(a_1, x)-f(x)\|$, which is a contradiction.

With the additional assumptions of regularity and of a Class 1 norm a stronger result may be established. This result is comparable to Theorem 3.2 where the number of strong sign changes is related to the measure of the set $\{x| F(a^*, x)-f(x)=0\}$. The number μ of zeros of $F(a^*, x)-f(x)$ required if $F(a^*, x)-f(x)$ has $k<m(a^*)$ is not simply defined in this general situation. For specialized cases, such as polynomials and the L_1 norm, μ might be determined explicitly as a function of k.

THEOREM 4.2. *Let F be a regular varisolvent function and let the norm be of Class* 1. *If $F(a^*, x)$ is a best approximation to $f(x)$ on X and $F(a^*, x)-f(x)$ has exactly k, $k \leqq m(a^*)$, strong sign changes on X, then $F(a^*, x)-f(x)$ has at least μ_k zeros in X, where μ_k is defined in the proof.*

Proof. Let $R=\{y_j|j=1, 2, \cdots, m\}$ be the set of zeros of $F(a^*, x)-f(x)$, and determine a sequence $\{a_n|a_n\in P\}$ by Lemma 4.1 so that the signs of $F(a_n, x)-f(x)$ and $F(a^*, x)-f(x)$ agree weakly and so that $|F(a_n, x)-F(a^*, x)| \leqq 2^{-n}$. Further, a_n is determined so that $|F(a_n, x_j)-f(x_j)| < |F(a^*, x_j)-f(x_j)|$ if $x_j \notin R$, $x_j \in X$. We may assume that $F(a_n, x_j)-F(a^*, x_j) \neq 0$ for all n and some $x_j \in R$; otherwise $F(a^*, x)$ would not be a best approximation for $f(x)$.

Define

$$s_n(x) = \begin{cases} F(a_n, x) - f(x), & x \notin R, \\ F(a^*, x) - f(x), & x \in R, \end{cases}$$

$$t_n(x) = \begin{cases} F(a_n, x) - f(x), & x \in R, \\ F(a^*, x) - f(x), & x \notin R. \end{cases}$$

Since the norm is of Class 1 we have

$$\left| \frac{\|F(a^*, x) - f(x) - t_n(x)\|}{\|F(a^*, x) - f(x)\| - \|s_n(x)\|} \right| \leqq K.$$

It is seen that

$$| F(a^*, x) - f(x) - t_n(x) | = | s_n(x) - F(a_n, x) + f(x) |$$

and hence

$$\left| \frac{\|F(a_n, x) - f(x)\| - \|s_n(x)\|}{\|F(a^*, x) - f(x)\| - \|s_n(x)\|} \right| \leqq K.$$

Set

$$K(f(x), \{a_n\}) = \inf \left\{ K \left| \left| \frac{\|F(a_n, x) - f(x)\| - \|s_n(x)\|}{\|F(a^*, x) - f(x)\| - \|s_n(x)\|} \right| \leqq K \text{ for all } n \right. \right\}.$$

$K(f(x), \{a_n\})$ is finite for each $f(x)$ and each sequence $\{a_n\}$. Let $K(f(x)) = \inf\{K(f(x), \{a_n\}) | F(a_n, x)$ tends to $F(a^*, x)$ as prescribed$\}$. Assume that $\{a_n\}$ is chosen so that $K(f(x), \{a_n\})$ is arbitrarily close to $K(f(x))$. Consider

$$\begin{aligned} 0 &\geqq \|F(a^*, x) - f(x)\| - \|F(a_n, x) - f(x)\| \\ &= \|F(a^*, x) - f(x)\| - \|s_n(x)\| + \|s_n(x)\| - \|F(a_n, x) - f(x)\|. \end{aligned}$$

It follows that

$$\begin{aligned} \|F(a^*, x) - f(x)\| - \|s_n(x)\| &\leqq \|F(a_n, x) - f(x)\| - \|s_n(x)\| \\ &\leqq K(f(x))[\|F(a^*, x) - f(x)\| - \|s_n(x)\|]. \end{aligned}$$

Thus it is seen that if $F(a^*, x)$ is a best approximation to $f(x)$ and $F(a^*, x) - f(x)$ has k strong sign changes and m zeros, $f(x)$ must be such that $K(f(x)) \geqq 1$. Let

$$K(l, m) = \inf\{K(f(x))\}$$

where the infimum is taken over all $f(x)$ such that $F(a^*, x) - f(x)$ has l strong sign changes on X and m zeros in X. Clearly, if $K(l, m) < 1$, then $F(a^*, x)$ cannot be a best approximation in such a situation. Let

$$\mu_k = \min\{m \mid K(k, m) \geqq 1\}$$

then $F(a^*, x) - f(x)$ must have at least μ_k zeros if $F(a^*, x) - f(x)$ has k strong sign changes in X.

The behavior of μ_k depends on the particular properties of F and the norm. Intuitively one feels that μ_k decreases as a function of k, but this has not been established. It is easy to see that $\mu_{m(a^*)} = 0$ and $\mu_0 > m(a^*) + 1$.

For Class 2 norms the following direct analog of Theorem 3.3 may be established.

THEOREM 4.3. *Let F be a regular varisolvent function and let the norm be of Class 2. If $F(a^*, x)$ is a best approximation to $f(x)$ then $F(a^*, x)$ is a strongly interpolating function of $f(x)$.*

Proof. Assume that $F(a^*, x) - f(x)$ has $k < m(a^*)$ strong sign changes. A set $\{y_i \mid i = 1, 2, \cdots, k\}$ is chosen in the following manner. If $F(a^*, x) - f(x)$ changes sign between x_j and x_{j+1} choose a point in the interior of $[x_j, x_{j+1}]$. If $F(a^*, x_j) - f(x_j) = 0$, $j = l, l+1, \cdots, p$ and $F(a^*, x) - f(x)$ changes sign from x_{l-1} to x_{p+1} choose a point in the interior of $[x_{l-1}, x_{p+1}]$.

The proof of this theorem may be completed along the lines of the proof of Theorem 3.3.

Theorems as strong as Theorems 4.3 and 4.2 have not been established for Class 3 norms. Of course Theorem 4.1 is applicable to Class 3 norms. $F(a^*, x)$ is said to be a *local best approximation* to $f(x)$ if there is a neighborhood $N(a^*)$ of a^* in P such that a $N(a^*)$ implies

$$\|F(a^*, x) - f(x)\| < \|F(a, x) - f(x)\|.$$

THEOREM 4.4. *Let F be a regular varisolvent function and let the norm be of Class 3. Then if $F(a^*, x)$ is an exactly interpolating function of $f(x)$ on X, $F(a^*, x)$ is a local best approximation to $f(x)$.*

Proof. Assume that $F(a^*, x)$ exactly interpolates $f(x)$ on X, and let $R = \{x_j \mid |F(a^*, x_j) - f(x_j)| = 0\}$ and set

$$d = \min_j \{ |F(a^*, x_j) - f(x_j)| \mid x_j \notin R\}.$$

Let the neighborhood $N(a^*)$ be such that $a \in N(a^*)$ implies

$$|F(a^*, x_j) - F(a, x_j)| \leqq d/2, \qquad j = 1, 2, \cdots, M.$$

Assume that there exists a sequence $\{a_n | a_n \in N(a^*)\}$ such that $\operatorname{Lim}_{n\to\infty} a_n = a^*$ and

$$\|F(a^*, x) - f(x)\| \geqq \|F(a_n, x) - f(x)\|.$$

Set

$$s_n(x) = \begin{cases} F(a_n, x) - f(x), & x \notin R, \\ F(a^*, x) - f(x), & x \in R; \end{cases}$$

then

$$\begin{aligned} \|F(a^*, x) - f(x)\| - \|F(a_n, x) - f(x)\| = \|F(a^*, x) - f(x)\| - \|s_n(x)\| \\ + \|s_n(x)\| - \|F(a_n, x) - f(x)\|. \end{aligned}$$

It is clear that $\|s_n(x)\| - \|F(a_n, x) - f(x)\| < 0$ since $|s_n(x)| \leqq |F(a_n, x) - f(x)|$ and the inequality holds for some points in R. If $F(a_n, x) = F(a^*, x)$ in R, then $a^* = a_n$, since R contains at least $m(a^*)$ points. Since the norm is of Class 3 it follows that

$$\operatorname*{Lim}_{n\to\infty} \frac{\|F(a^*, x) - f(x)\| - \|s_n(x)\|}{\|F(a_n, x) - f(x)\| - \|s_n(x)\|} = 0.$$

This implies that for n sufficiently large that

$$\|F(a^*, x) - f(x)\| - \|F(a_n, x) - f(x)\| < 0$$

which contradicts the original assumption on the sequence $\{a_n\}$.

Theorem 4.4 is of a completely different nature from the theorems proved up to this point. Instead of the assumption of best approximation implying interpolation, we have an assumption of interpolation implying a local best approximation. The natural companion of Theorems 4.2 and 4.3 would be: *Let F be a regular varisolvent function and let the norm be of Class 3. If $F(a^*, x)$ is a best approximation to $f(x)$, then $F(a^*, x) - f(x)$ has at least $m(a^*)$ zeros in X.* It is not known if this is a true statement.

On the other hand one would like to prove the converse theorems of Theorem 4.2 and Theorem 4.3. This has not been done.

5. **The L_p norms.** The remainder of this paper is concerned with approximations in weighted L_p norms by linear approximating functions on finite point sets. Motzkin and Walsh [8] have developed an extensive theory for weighted L_p norms and polynomials on finite point sets. The main points of this theory are extended to approximations by linear unisolvent functions in this section. For the Tchebycheff norm the results are extended to varisolvent functions and to the interval [0, 1].

Let X and

$$P(a, x) = \sum_{i=1}^{m} a^i \phi_i(x)$$

be as previously defined where the $\phi_i(x)$ form a Tchebycheff set.

For any finite point set X define a function $\phi_{n+1}(x)$ such that $\{\phi_i(x) \mid i=1, 2, \cdots, n+1\}$ forms a Tchebycheff set on X. $\phi_{n+1}(x)$ need not be defined except in X.

Let $Y=\{y_i \mid i=1, 2, \cdots, n+1\}$ be a subset of X. Define $n+1$ functions of the form

$$w_k(y) = P(a_k, y)/c_k + \phi_{n+1}(y), \qquad y \in Y,$$

such that

$$P(a_k, y_j) + c_k\phi_{n+1}(y_j) = \begin{cases} 0, & j \neq k, \\ (-1)^k, & k = j. \end{cases}$$

Let $\{\epsilon_j \mid j = 1, 2, \cdots, n + 1; \; |\epsilon_j| = 1\}$ be such that $\epsilon_k c_k > 0$. Since $\{\phi_i(x) \mid i=1, 2, \cdots, n\}$ is a Tchebycheff set the c_k are all nonzero.

LEMMA 5.1. *The functions $T(x)=\phi_{n+1}(x)+P(a, x)$ which weakly interpolate zero on Y are exactly the functions*

$$(1) \qquad T(x) = \sum_{j=1}^{n+1} \lambda_j \epsilon_j w_j(x), \qquad \lambda_j \geqq 0, \qquad \sum_{j=1}^{n+1} \lambda_j = 1.$$

Those which strongly interpolate zero on Y are of the same form except $\lambda_j>0$.

Proof. Clearly every function of the form (1) with $\lambda_j>0$ has n strong sign changes on Y. Those with $\lambda_j \geqq 0$ are limiting cases of $\lambda_j>0$ and hence have n weak sign changes.

Assume that $T(x)$ has n strong sign changes. Then $T(x)$ changes sign from y_j to y_{j+1} and on Y we have

$$T(x) = \sum_{i=1}^{n+1} \lambda_i \epsilon_i w_i(x)$$

where $\lambda_i = \pm |c_i T(y_i)| \neq 0$. All the λ_i have the same sign and since $T(x) =\phi_{n+1}(x)+P(a, x)$ we have $\sum_{i=1}^{n+1} \lambda_i=1$ which implies $\lambda_i>0$. The $T(x)$ with weak sign changes are limiting cases of those with strong sign changes, and hence $\lambda_i \geqq 0$.

LEMMA 5.2. *If $\phi_{n+1}(x)+P(a^*, x)$ weakly interpolates zero on Y, then there are positive weights $\{\mu_i \mid i=1, 2, \cdots, n+1\}$ such that $\phi_{n+1}(x)+P(a^*, x)$ minimizes the weighted L_1 norm with these weights on Y.*

Proof. If $\phi_{n+1}(x)+P(a^*, x)$ has n weak sign changes on Y, then by Lemma 5.1

$$P(a^*, x) + \phi_{n+1}(x) = \sum_{i=1}^{n+1} \lambda_i \epsilon_i w_i(x), \qquad \sum_{i=1}^{n+1} \lambda_i = 1, \qquad \lambda_i \geqq 0.$$

If $\lambda_j \neq 0$, set $\lambda_j = |c_j|$ and if $\lambda_j = 0$, set $\lambda_j = 2|c_j|$. Consider

$$\mu(P(a^*, x)) = \sum_{i=1}^{n+1} \mu_i | P(a^*, y_i) + \phi_{n+1}(y_i) | = \sum_{i=1}^{n+1} \mu_i | \lambda_i \epsilon_i w_i(y_i) |$$

$$= \sum_{i=1}^{n+1} \mu_i \lambda_i / | c_i | = 1.$$

Clearly a^* minimizes $\mu(P(a, x))$, and $P(a^*, x)$ is a best approximation to $-\phi_{n+1}(x)$ on Y in the weighted L_1 norm.

With these two lemmas a special case of the converse of Theorem 4.1 may be proved.

THEOREM 5.1. *If $P(a^*, x)$ weakly interpolates $f(x)$ on X, then there are positive weights $\{\mu_i\}$ such that $P(a^*, x)$ is a best approximation to $f(x)$ in the weighted L_1 norm.*

Proof. If $P(a^*, x) - f(x)$ has n weak sign changes on X, then it is simple to show that every point of X belongs to a subset Y of $n+1$ points such that $P(a^*, x) - f(x)$ has n weak sign changes on Y.

On any such Y one may determine $a_1 \in P$ and K such that on Y

$$P(a^*, x) - f(x) = K[P(a_1, x) + \phi_{n+1}(x)].$$

It follows from Lemma 5.2 that since $P(a_1, x) + \phi_{n+1}(x)$ has n weak sign changes, there are positive weights $\{\mu_i\}$ such that $P(a_1, x)$ is a best approximation on Y to $-\phi_{n+1}(x)$ in the weighted L_1 norm. Hence $P(a^*, x) - f(x)$ minimizes the same norm on Y.

If $P(a^*, x)$ is a best approximation to $f(x)$ on Y_1 with weights $\{\mu_i\}$ and a best approximation to $f(x)$ on Y_2 with weights $\{\sigma_i\}$, then $P(a^*, x)$ is a best approximation to $f(x)$ on $Y_1 \cup Y_2$ with weights $\{\rho_i\}$ determined as follows: In $Y_1 \cap Y_2$, $\rho_i = (\sigma_i + \mu_i)$, in $Y_1 - Y_1 \cap Y_2$, $\rho_i = \mu_i$ and in $Y_2 - Y_1 \cap Y_2$, $\rho_i = \sigma_i$. X has a finite number of subsets of $n+1$ points, and therefore $P(a^*, x)$ is a best approximation to $f(x)$ on X in some weighted L_1 norm.

A similar approach may be used to establish a converse of Theorem 4.3 for L_p norms, $1 < p < \infty$ and linear approximating functions. The following lemma may be easily established [7].

LEMMA 5.3. *Let $\lambda_i \geqq 0$, $\sum_{i=1}^{m} \lambda_i = 1$, and $\mu(\lambda_i) = \sum_{i=1}^{m} \mu_i \lambda_i^p$; $\mu_i > 0$, $p > 1$. The minimum of $\mu(\lambda_i)$ is given by*

$$\lambda_i = \mu_i^{1/(1-p)} \Big/ \sum_{j=1}^{m} \mu_j^{1/(1-p)}.$$

For the next lemma Y, as usual, denotes a subset of X containing $n+1$ distinct points.

LEMMA 5.4. *Let $p>1$ be given. If $P(a^*, x)+\phi_{n+1}(x)$ strongly interpolates zero on Y, then there are positive weights such that $P(a^*, x)+\phi_{n+1}(x)$ minimizes a weighted L_p norm.*

Proof. From Lemma 5.1 it is seen that $P(a^*, x)+\phi_{n+1}(x)$ is the form

$$\phi_{n+1}(x) + P(a^*, x) = \sum_{i=1}^{n+1} \lambda_i \epsilon_i w_i(x), \qquad \lambda_i > 0, \ \sum_{i=1}^{n+1} \lambda_i = 1,$$

for $x \in Y$. Set $\mu_i = \lambda_i^{1-p} |c_i|^p$. Then it follows from Lemma 5.3 that the λ_i minimize

$$\sum_{i=1}^{n} \mu_i \lambda_i^p / |c_i|^p,$$

which concludes the proof of the lemma.

THEOREM 5.2. *If $P(a^*, x)$ strongly interpolates $f(x)$ on X, then for any $p>1$ there is a set of positive weights such that $P(a^*, x)$ is a best approximation to $f(x)$ in the weighted L_p norm.*

Proof. Set $E = \{x \mid |P(a^*, x)-f(x)| >0, x \in X\}$. Every point of E belongs to a subset Y such that $P(a^*, x)-f(x)$ has n strong sign changes on Y. For any such subset Y there are $a_1 \in P$ and K such that

$$P(a^*, x) - f(x) = K[P(a_1, x) + \phi_{n+1}(x)].$$

It follows from Lemma 5.4 that there are positive weights such that $P(a^*, x)$ is a best approximation to $f(x)$ in a weighted L_p norm.

The argument of the proof of Theorem 5.1 shows that there are positive weights such that $P(a^*, x)$ is a best approximation to $f(x)$ on E in a weighted L_p norm. On the points of X not in E the weights may be assigned arbitrarily and $P(a^*, x)$ is a best approximation to $f(x)$ on X.

For L_p norm, $0<p<1$, Motzkin, Walsh, and Dvoretzky [6, Theorem 6] have established the following theorem.

THEOREM 5.3. *If $P(a^*, x)$ is a best approximation to $f(x)$ in a weighted L_p norm, $0<p<1$, then $P(a^*, x)-f(x)$ has at least n zeros on X.*

This theorem is analogous to Theorem 4.3, and one could conjecture that this theorem extends to varisolvent functions. The proof of Theorem 5.3 is of a different nature than those of this paper and apparently does not extend to varisolvent, or even unisolvent, functions without essential modification.

The analysis for L_p norms, $0<p<1$, follows the same pattern as for $p \geqq 1$. The following lemma is easily established.

LEMMA 5.5. *Let $\lambda_i>0$, $\sum_{i=1}^m \lambda_i = 1$ and set*

$$\mu(\lambda_i) = \sum_{i=1}^{m} \mu_i \lambda_i^p, \qquad \mu_i > 0, \ 0 < p < 1.$$

The minimum of $\mu(\lambda_i)$ is given by

$$\lambda_j = \begin{cases} 0, & \text{if } \mu_i > \min_j \mu_j, \\ \rho_i, & \text{if } \mu_i = \min_j \mu_j, \end{cases}$$

where $\rho_i \neq 0$ for only one value of i.

LEMMA 5.6. *The function of the form $T(x) = P(a, x) + \phi_{n+1}(x)$ which minimizes*

$$\sum_{i=1}^{n+1} \mu_i \left| T(y_i) \right|^p, \qquad \mu_i > 0, 0 < p < 1, y_i \in Y,$$

is given by $T(x) = \sum_{i=1}^{n+1} \lambda_i \epsilon_i w_i(x)$ where $\lambda_i = 0$ except at the point y_k (or a point y_k) at which $\mu_k / \left| c_k \right|^p = \min_i \mu_i / \left| c_i \right|^p$ where $\lambda_k = 1$.

Proof. It is seen that

$$\sum_{i=1}^{n+1} \mu_i \left| T(y_i) \right|^p = \sum_{i=1}^{n+1} \mu_i \lambda_i^p / \left| c_i \right|^p$$

and the proof follows directly from Lemma 5.5.

THEOREM 5.4. *If $P(a^*, x)$ exactly interpolates $f(x)$ on X, then for any p, $0 < p < 1$, there is a set of positive weights such that $P(a^*, x)$ is a best approximation to $f(x)$ in the weighted L_p norm.*

Proof. Select a subset $Y = \{ y_i \mid i = 1, 2, \cdots, n+1 \}$ such that $P(a^*, x) - f(x)$ has n zeros in Y. In Y we have

$$P(a^*, x) - f(x) = K[P(a_1, x) + \phi_{n+1}(x)],$$

and hence $P(a_1, x) + \phi_{n+1}(x)$ has n zeros in Y. Set $\mu_k = \left| c_k \right|^p$ at the point of Y where $P(a_1, x) + \phi_{n+1}(x) \neq 0$. At each of the other points of Y determine μ_i so that $\mu_i / \left| c_i \right|^p > 1$. Then by Lemma 5.6, $P(a^*, x) + \phi_{n+1}(x)$ minimizes the weighted L_p norm with these weights. Hence $P(a^*, x)$ is a best approximation to $f(x)$ on Y.

Every point of X belongs to a subset Y of the above type. By the argument previously employed it follows that $P(a^*, x)$ is a best approximation to $f(x)$ on X.

For the L_∞ or Tchebycheff norm

$$\| f(x) \| = \max_{x \in X} \left| f(x) \right|$$

the analysis is considerably simpler and extends immediately in several directions. We have

THEOREM 5.5. *If $P(a^*, x)$ strongly interpolates $f(x)$ on X, then there is a set of positive weights such that $P(a^*, x)$ is a best approximation to $f(x)$ in the weighted Tchebycheff norm.*

Proof. Choose a set $Y=\{y_i\}$ of $n+1$ points such that $P(a^*, x)-f(x)$ has n strong sign changes on Y. In Y determine $\{\mu_i | i=1, 2, \cdots, n+1\}$ by

$$\mu_i \left| P(a^*, y_i) - f(y_i) \right| = 1, \qquad i = 1, 2, \cdots, n+1.$$

For $x_j \notin Y$ determine μ_j so that

$$\mu_j \left| P(a^*, x_j) - f(x_j) \right| < 1.$$

It follows from [1, Chapter 2] that $P(a^*, x)$ is a best approximation to $f(x)$ in the weighted Tchebycheff norm with weights $\{\mu_j | j=1, 2, \cdots, M\}$.

The same proof may be used to establish the following corollary.

COROLLARY 5.1. *If $P(a^*, x)-f(x)$ has n strong sign changes in $[0, 1]$, then there is a positive continuous weight function $\mu(x)$ such that $P(a^*, x)-f(x)$ minimizes*

$$\max_{x\in[0,1]} \mu(x) \left| P(a^*, x) - f(x) \right| .$$

The next corollaries follow directly from the proof of Theorem 5.5 and [12].

COROLLARY 5.2. *Let F be a varisolvent function. If $F(a^*, x)-f(x)$ has $m(a^*)$ strong sign changes in $[0, 1]$, then there is a positive continuous weight function such that $F(a^*, x)$ is the best approximation to $f(x)$ in the weighted Tchebycheff norm.*

COROLLARY 5.3. *Let F be a varisolvent function and let $f(x)$ be given. Then the following subsets of P are identical:*

$\{a | F(a, x)-f(x)$ *has* $m(a)$ *strong sign changes in* $[0, 1]\}$.

$\{a | F(a, x)$ *is the best approximation to* $f(x)$ *in a weighted* L_∞ *norm*$\}$.

The main conclusions of this section are summarized by the following theorem.

THEOREM 5.6. *Let $P(a, x)=\sum_{i=1}^{n} a^i\phi_i(x)$, X, $0<q<1<p\leqq\infty$ and $f(x)$ be given. Then we have three pairs of identical sets:*

(1)
$\{a | P(a, x) - f(x)$ *minimizes a weighted* L_p *norm*$\}$.
$\{a | P(a, x)$ *strongly interpolates* $f(x)$ *on* $X\}$.

(2)
$\{a | P(a, x) - f(x)$ *minimizes a weighted* L_1 *norm*$\}$.
$\{a | P(a, x)$ *weakly interpolates* $f(x)$ *on* $X\}$.

(3)
$\{a | P(a, x) - f(x)$ *minimizes a weighted* L_q *norm*$\}$.
$\{a | P(a, x)$ *exactly interpolates* $f(x)$ *on* $X\}$.

Proof. The equivalence for the first pair follows from Theorem 4.3, Theorem 5.2, Theorem 5.5 and [1, Chapter 2]. The equivalence of the second pair follows from Theorem 4.1 and Theorem 5.1, the third pair from Theorem 5.3 and Theorem 5.4.

We have as a corollary

COROLLARY 5.4. *Let $P(a, x)$, X and $f(x)$ be given. Then for any p, q; $1 < p < q \leqq \infty$, the following sets are identical:*

$$\{a \mid P(a, x) - f(x) \text{ minimizes a weighted } L_p \text{ norm}\}.$$

$$\{a \mid P(a, x) - f(x) \text{ minimizes a weighted } L_q \text{ norm}\}.$$

REFERENCES

1. N. I. Achieser, *Theory of approximation*, New York, Frederick Ungar, 1956.

2. L. Fejér, *Über die Lage der Nullstellen von Polynomen, die aus Minimumforderungen gewisser Art entspringen*, Math. Ann. vol. 85 (1922) pp. 41–48.

3. D. Jackson, *On functions of closest approximation*, Trans. Amer. Math. Soc. vol. 22 (1921) pp. 117–128.

4. ———, *Note on a class of polynomials of approximation*, Trans. Amer. Math. Soc. vol. 22 (1921) pp. 320–326.

5. T. S. Motzkin, *Approximation by curves of a unisolvent family*, Bull. Amer. Math. Soc. vol. 55 (1949) pp. 789–793.

6. T. S. Motzkin and J. L. Walsh, *Least pth power polynomials on a real finite point set*, Trans. Amer. Math. Soc. vol. 78 (1955) pp. 67–81.

7. ———, *Least pth power polynomials on a finite point set*, Trans. Amer. Math. Soc. vol. 83 (1956) pp. 371–396.

8. ———, *Polynomials of best approximation on a real finite point set*. I, Trans. Amer. Math. Soc. vol. 91 (1959) pp. 231–245.

9. ———, *Polynomials of best approximation on an interval*, Proc. Nat. Acad. Sci. U.S.A. vol. 45 (1959) pp. 1523–1528.

10. ———, *Best approximations within a linear family on an interval*, Proc. Nat. Acad. Sci. U.S.A. vol. 46 (1960) pp. 1225–1233.

11. J. R. Rice, *The characterization of best nonlinear Tchebycheff approximations*, Trans. Amer. Math. Soc. vol. 96 (1960) pp. 322–340.

12. ———, *Tchebycheff approximation by functions unisolvent of variable degrees*, Trans. Amer. Math. Soc. vol. 99 (1961) pp. 298–302.

RESEARCH LABORATORIES, GENERAL MOTORS CORPORATION,
WARREN, MICHIGAN