TCHEBYCHEFF APPROXIMATION
IN SEVERAL VARIABLES

BY

JOHN R. RICE

1. **Introduction.** The theory of Tchebycheff approximation for functions of one real variable has been understood for some time and is quite elegant. For about fifty years attempts have been made to generalize this theory to functions of several variables. These attempts have failed because of the lack of uniqueness of best approximations to functions of more than one variable. Indeed, Mairhuber [6] has shown that best approximations cannot be unique unless our functions are defined on a space homeomorphic to a subset of the unit circle. Further negative results have been obtained by Rivlin and Shapiro [11] and the conclusion is that there is no interesting Tchebycheff approximation problem in several variables for which best approximations are unique. Schoenberg [12] gives an account of a theory in several variables, but his assumptions essentially limit the applicability to functions of one variable.

In this paper two theories of Tchebycheff approximation to functions of several variables are developed. It cannot be said that they are as satisfying as the theory for one variable, but they are workable theories. For the first theory the concept of a critical point set is given (the definition is somewhat different from that in one variable) and it is shown that the set of critical point sets of a best approximation is unique. A characterization theorem for best approximations is given in terms of critical point sets and other theorems normally found in the theory of Tchebycheff approximation are valid. Similar viewpoints have recently been investigated by Lawson [16] and Rivlin and Shapiro [17].

The second theory defines the *strict approximation* for functions defined on a *finite* point set. In addition to the results obtained from the first theory, the strict approximation is unique. It is also a best approximation in the usual sense.

There are two basic features of the one variable theory which do not generalize readily. The first of these is the idea of a Tchebycheff set. No counterpart of these sets exists in this paper. A heuristic argument is given that shows that there is no possibility of defining Tchebycheff sets in several variables with the important properties present in the case of one variable. The second feature which does not generalize readily is that of alternation or oscillation. This is an intrinsic feature of Tchebycheff approximations, but it does not appear to be possible to give a simple geometric interpretation of it for functions of several variables.

It is presented here in terms of isolability which is intimately related to irreducibly inconsistent systems of linear inequalities.

Although the theories developed here lack some of the neatness and elegance of the one variable theory, they do allow one to solve the central problem of approximation theory, namely, the computation of best approximations. In the latter part of this paper a 1 for 1 exchange algorithm is described in some detail which can be used for the actual computation of best approximations of functions of several variables.

2. **Preliminaries.** Banach spaces and sets in general are denoted by capital script letters $\mathscr{A}, \mathscr{B}, \cdots$ and elements of such sets are denoted by lower case roman letters $x, y, \cdots$. The norm of an element x in a Banach space $\mathscr{B}$ is denoted by $\|x\|_{\mathscr{B}}$. Curly brackets are used to denote sets or sequences and $\{x \mid \cdots\}$ is read as "the set of x such that $\cdots$."

Let $\mathscr{B}$ be a compact subset of a Banach space. The space of real-valued continuous functions defined on $\mathscr{B}$ is denoted by $\mathscr{C}$ and has elements $f, g, \cdots$. The norm in $\mathscr{C}$ is taken to be

$$\|f\|_{\mathscr{C}} = \max_{x \in \mathscr{B}} |f(x)|.$$

The subscript $\mathscr{C}$ may be deleted when there is no possibility of ambiguity.

Let $\mathscr{L}$ be a finite dimensional linear subspace of $\mathscr{C}$. The basis functions of $\mathscr{L}$ are denoted by $g_i(x)$, $i = 1, 2, \cdots, n$. The dimension of $\mathscr{L}$ is n. An element in $\mathscr{L}$ is denoted by

$$L(A,x) = \sum_{i=1}^{n} a_i g_i(x), \qquad |a_i| < \infty.$$

The a_i are said to be the parameters of $L(A,x)$ and A denotes the collection of the parameters.

The *Tchebycheff approximation problem* in this context is stated as follows: Given $f(x)$ in $\mathscr{C}$ determine A^* such that

$$\|f(x) - L(A^*,x)\|_{\mathscr{C}} \leq \|f(x) - L(A,x)\|_{\mathscr{C}}$$

for all A. Such a $L(A^* x)$ is said to be a *best approximation* to $f(x)$ with *deviation* $\|f(x) - L(A^*,x)\|_{\mathscr{C}}$. The elements of $\mathscr{B}$ where the norm is assumed, i.e.,

$$|f(x) - L(A^*,x)| = \|f(x) - L(A^*,x)\|$$

are said to be *extremal points*. The set of extremal points is divided into two parts as follows:

$$\mathscr{P} = \{x \mid x \in \mathscr{B}, \ f(x) - L(A^*,x) = \|f(x) - L(A^*,x)\|\},$$
$$\mathscr{N} = \{x \mid x \in \mathscr{B}, \ f(x) - L(A^*,x) = - \|f(x) - L(A^*,x)\|\}.$$

The elements of these two sets are called *positive points* and *negative points*, respectively. These designations may also be used when $L(A^*,x)$ is not a best approximation to $f(x)$.

The point sets $\mathscr{P}$ and $\mathscr{N}$ are said to be *isolable* if there is an A such that

$$L(A,x) > 0 \quad x \in \mathscr{P},$$

$$L(A,x) < 0 \quad x \in \mathscr{N}.$$

We will need to speak of the set of zeros of an element of $\mathscr{L}$. Such a set $\Gamma(A)$ is said to be an *isolating curve* and

$$\Gamma(A) = \{ x \,|\, x \in \mathscr{B}, \quad L(A,x) = 0 \}.$$

Thus the sets $\mathscr{P}$ and $\mathscr{N}$ are isolable if they lie on "opposite" sides of an isolating curve $\Gamma(A)$.

From the theory of Tchebycheff approximations in one variable it is known that the extremal points play an important role in the determination of a best approximation. In order to study Tchebycheff approximation in more than one variable our attention will be focused on particular subsets of the extremal point set. A subset $\mathscr{R}$ of extremal points is said to be a *critical point set* if the positive and negative parts of $\mathscr{R}$ are not isolable, but if any point is deleted from $\mathscr{R}$ then the positive and negative parts of $\mathscr{R}$ are isolable. Critical point sets play a central role in the theory of this paper. Note that the particular division of $\mathscr{R}$ into positive and negative parts determines whether $\mathscr{R}$ is a critical point set or not. Thus a critical point set is defined *relative* to a particular function $f(x)$ and the elements alone of a set are not sufficient to determine whether a set is critical or not.

3. **Uniqueness of critical point sets.** When one considers the Tchebycheff approximation problem for functions, the first question that arises is existence. An application of standard methods establishes the existence of a best approximation to any $f(x)$ in $\mathscr{C}$ if $\mathscr{B}$ is a compact set [1].

The second question that arises is uniqueness. A considerable amount of work has been done in this area and almost all of the results are of a negative nature. Mairhuber [6] has shown that if best approximations are to be unique for every $f(x)$ then $\mathscr{B}$ must be homeomorphic to a subset of the unit circle. Then Collatz [4] showed that best approximations by $a + bx + cy$ are unique if $\mathscr{B}$ is a strictly convex subset of the (x, y) plane and $f(x)$ is differentiable. This gave some hope that there may be a reasonably restricted approximation problem for which best approximations are unique. Recently Rivlin and Shapiro [11] have shown that no such possibility exists. They show, for example, that there are infinitely differentiable functions of two variables which do not have a unique approximation by quadratic polynomials of two variables. A general discussion of the uniqueness problems is given in [12].

In this section it is shown that there is a particular point set associated with best approximations which is unique.

The following simple lemma will be required.

LEMMA 1. *Let $L(A^*, x)$ be a best approximation to $f(x)$ and let $\mathscr{R}$ be a critical point set. Then $L(A^*, x)$ is a best approximation to $f(x)$ on $\mathscr{R}$.*

Proof. Assume that there is an $L(A_0, x)$ such that

$$\max_{x \in \mathscr{R}} \left| L(A_0, x) - f(x) \right| < \max_{x \in \mathscr{R}} \left| L(A^*, x) - f(x) \right|.$$

Then we have

$$L(A_0, x) - L(A^*, x) > 0 \qquad x \in \mathscr{P},$$

$$L(A_0, x) - L(A^*, x) < 0 \qquad x \in \mathscr{N}.$$

This contradicts the fact that $\mathscr{P}$ and $\mathscr{N}$ are not isolable.

If there are two distinct best approximations on $\mathscr{B}$ to $f(x)$, say $L(A_1, x)$ and $L(A_2, x)$ then there are critical point sets associated with each best approximation. It is possible that each best approximation has more than one critical point set associated with it. This corresponds to best approximations in one variable which have more than $n + 1$ extremal points.

THEOREM 1. *The sets of critical point sets, $\mathscr{R}_1$ and $\mathscr{R}_2$, of two distinct best approximations, $L(A_1, x)$ and $L(A_2, x)$, are identical.*

Proof. Assume that $L(A_1, x)$ and $L(A_2, x)$ are best approximations to a given function $f(x)$ from $\mathscr{C}$. Let $\mathscr{R}_1$ be a critical point set of $L(A_1, x)$ decomposed into positive and negative parts by

$$\mathscr{R}_1 = \mathscr{P}_1 \cup \mathscr{N}_1.$$

It is further assumed that $\mathscr{R}_1$ is not a critical point set associated with $L(A_2, x)$. Now consider $L(A_1, x) - L(A_2, x)$ on the sets $\mathscr{P}_1$ and $\mathscr{N}_1$. We have

$$L(A_1, x) - L(A_2, x) \geqq 0 \qquad x \in \mathscr{P}_1,$$

$$L(A_1, x) - L(A_2, x) \leqq 0 \qquad x \in \mathscr{N}_1.$$

Since $\mathscr{R}_1$ is not a critical point set of $L(A_2, x)$ there must be strict inequality at at least one point of $\mathscr{R}_1$.

It follows from Lemma 1 that $L(A_1, x)$ is a best approximation to $f(x)$ on $\mathscr{R}_1$. Consider

$$L(A_3, x) = L(A_1, x) + \lambda[L(A_1, x) - L(A_2, x)].$$

For λ suitably chosen we have

$$L(A_3, x) = L(A_1, x) \qquad x \in \mathscr{R}_1',$$

$$\left| L(A_3, x) - f(x) \right| < \left| L(A_1, x) - f(x) \right| \qquad x \notin \mathscr{R}_1',$$

where $\mathscr{R}'_1$ is a proper subset of $\mathscr{R}_1$. Since $\mathscr{R}_1$ is a critical point set the positive and negative parts of $\mathscr{R}'_1$ can be isolated and hence there is an $L(A_4, x)$ such that

$$\max_{\mathscr{R}_1} |L(A_4, x) - f(x)| < \max_{\mathscr{R}_1} |L(A_1, x) - f(x)|.$$

This is a contradiction and concludes the proof.

4. Conditions of nonuniqueness. The fact that nonuniqueness of best approximations is an integral part of the theory of Tchebycheff approximation in more than one variable is firmly established by the results of Rivlin and Shapiro [11]. However, the exact conditions under which nonuniqueness may occur are not given in [11]. The following theorems describe the nature of these conditions.

THEOREM 2. *If there is not a unique best approximation to $f(x)$, then all of the critical point sets lie on an isolating curve.*

Proof. Let $L(A_1, x)$ and $L(A_2, x)$ be distinct best approximations to $f(x)$ and let $\mathscr{R}$ be a critical point set. Then

$$|f(x) - L(A_1, x)| = |f(x) - L(A_2, x)| \qquad x \in \mathscr{R}.$$

Let $\mathscr{R}_0$ be the subset, if any, of $\mathscr{R}$, where $L(A_1, x)$ and $L(A_2, x)$ differ. Let $\mathscr{P}$ and $\mathscr{N}$ be the positive and negative parts, respectively, of $\mathscr{R}$ with reference to $L(A_1, x)$. Then we have

$$L(A_1, x) - L(A_2, x) \;>\; 0 \qquad x \in \mathscr{P}, \qquad x \in \mathscr{R}_0,$$

$$L(A_1, x) - L(A_2, x) \;<\; 0 \qquad x \in \mathscr{N}, \qquad x \in \mathscr{R}_0,$$

$$L(A_1, x) - L(A_2, x) \;=\; 0 \qquad x \in \mathscr{R}, \qquad x \notin \mathscr{R}_0.$$

This system of inequalities and equalities leads to a contradiction by the same argument as used in Theorem 1.

The converse of Theorem 2 is established by an argument similar to that used by Haar [1]. A finite set $\mathscr{R}$ of points is said to be a set of *independent* points if the system of linear equations for A

$$L(A, x) = v(x) \qquad x \in \mathscr{R}$$

has a solution for all $v(x)$.

THEOREM 3. *Let $\mathscr{R}$ be a finite set of points on an isolating curve Γ which is not independent. Then there exists an $f(x) \in \mathscr{C}$ for which best approximations are not unique and such that $\mathscr{R}$ contains a critical point set and is contained in the set of extremal points for some best approximation to $f(x)$.*

Proof. Let $\mathscr{R}$ consist of k points. Since $\mathscr{R}$ lies on an isolating curve, the mapping defined by

$$A: \; \to L(A, x) \qquad x \in \mathscr{R}$$

maps n-dimensional Euclidean space $\mathscr{E}_n$ into a proper subspace of $\mathscr{E}_k$. The image of $\mathscr{E}_n$ does not intersect the interior of some k-dimensional quadrant of $\mathscr{E}_k$. Let $v_0(x)$ be a vector in the interior of this quadrant. $\mathscr{R}$ is divided into positive and negative parts, $\mathscr{P}$ and $\mathscr{N}$, as follows:

$$\mathscr{P} = \{x \,|\, x \in \mathscr{R}, \; v_0(x) > 0\},$$

$$\mathscr{N} = \{x \,|\, x \in \mathscr{R}, \; v_0(x) < 0\}.$$

$\mathscr{P}$ and $\mathscr{N}$ are not isolable. This division may not be unique since the signs of $v_0(x)$ may not be uniquely determined.

Let the isolating curve Γ be given by

$$\Gamma = \{x \,|\, L(A_0, x) = 0\}.$$

Since $\mathscr{B}$ is compact we may assume that

$$\| L(A_0, x) \| \; < \; 1.$$

Let $\delta(x)$ be any function in $\mathscr{C}$ such that

$$\| \delta(x) \| \; \leqq \; 1,$$

$$\delta(x) \; = \; \operatorname{sgn} \, v_0(x) \qquad x \in \mathscr{R},$$

and define $f(x)$ by

$$f(x) \; = \; \delta(x) \, [1 \; - \; |\, L(A_0, x)\,|].$$

It is seen that

$$\|f(x) \; - \; L(A, x)\| \geqq \|f(x) \; - \; L(A_0, x)\|,$$

for otherwise $L(A - A_0, x)$ would isolate $\mathscr{P}$ and $\mathscr{N}$. On the other hand for $|\lambda| < 1$

$$|f(x) - \lambda L(A_0, x)| \leqq |f(x)| + |\lambda L(A_0, x)|$$

$$\leqq 1 - (1 - |\lambda|)\,|L(A_0, x)| \leqq 1.$$

Thus $f(x)$ has infinitely many best approximations, namely, $\lambda L(A_0, x)$ for $|\lambda| < 1$. This concludes the proof.

It is not necessarily true that $\mathscr{R}$ contains the set of critical point sets (though it may be possible to obtain this by further restrictions on $\delta(x)$). Nor is it true that every point of $\mathscr{R}$ must belong to a critical point set. To illustrate the latter approximation with consider

$$\mathscr{B} = \{(x, y) \,|\, 0 \leqq |x| \leqq 2, \, 0 \leqq |y| \leqq 2\}$$

and with $\mathscr{L}$ the subspace of cubic polynomials. Γ is defined by

$$\Gamma(A_0) = \{(x, y) \mid x^2(x - 1) = 0\}$$

and $\mathscr{R}$ is the set

$$\mathscr{R} = \{(0, -1), (0, 1/2), (0, 0), (0, 1/2), (0, 1), (1, 0)\}.$$

It may be shown that the point $(1, 0)$ does not belong to any critical point set. This follows from the fact that there is an element of $\mathscr{L}$, namely, x^2, which is zero on all points of $\mathscr{R}$ except $(1, 0)$.

If $\mathscr{B}$ is a closed and bounded subset of $\mathscr{E}_n$ and if $\mathscr{C}$ is replaced by the set $\mathscr{C}^1$ of functions differentiable on $\mathscr{B}$, then an even more restricted result may be established about nonuniqueness.

THEOREM 4. *Let $\mathscr{B}$ be a compact subset of $\mathscr{E}_n$ and $\mathscr{C}^1$ the set of functions differentiable on $\mathscr{B}$. If best approximations of an element $f(x) \in \mathscr{C}^1$ are not unique then the critical point sets lie on an isolating curve*

$$\Gamma = \{ x \mid L(A_0, x) = 0 \}$$

and the differential of $L(A_0, x)$ vanishes at the points of the critical point sets which are interior points of $\mathscr{B}$.

Proof. It follows from Theorem 2 that all of the critical points lie on an isolating curve Γ. If x_0 is a critical point in the interior of $\mathscr{B}$ then the differential of $L(A_1, x) - f(x)$ and $L(A_2, x) - f(x)$ vanishes at x_0. Hence the differential of $L(A_1, x) - L(A_2, x)$ vanishes at x_0 and the proof is complete.

From this theorem one can gain some insight into a theorem of Collatz [4].

THEOREM (COLLATZ). *If $f(x, y)$ has continuous first partial derivatives at all interior points of a strictly convex closed region $\mathscr{B}$ of the plane, then there exists a unique linear polynomial $a + bx + cy$ of best approximation to f in $\mathscr{B}$.*

This is a simple corollary of Theorems 2 and 4. The differential of $a + bx + cy$ can vanish only if $b = 0$ and $c = 0$. Thus the critical points must lie on the boundary. But this implies that there are only two critical points since an isolating curve can intersect the boundary of $\mathscr{B}$ at most twice. On the other hand, it is clear that two points cannot form a critical point set for the approximating function $a + bx + cy$. Thus the critical point set does not lie on an isolating curve and best approximations are unique.

A partial converse to Theorem 4 has been given by Rivlin and Shapiro [11]. The theorem is stated here in a more general form but the same proof applies.

THEOREM (RIVLIN AND SHAPIRO). *Let $\mathscr{B}$ be a compact subset of $\mathscr{E}$ and $\mathscr{L}$ a linear subspace of $\mathscr{C}$ whose elements are twice continuously differentiable. Let $\mathscr{R}$ be a finite set of points in $\mathscr{B}$ which are not independent. Assume that for some $L(A_0, x) \in \mathscr{L}$*

$$L(A_0, x) = 0 \qquad x \in \mathscr{R}$$

and the differential of $L(A_0, x)$ vanishes at every point of $\mathscr{R}$ which is not an isolated point of $\mathscr{B}$. Then there is an $f(x) \in \mathscr{C}^{\infty}$ for which best approximations are not unique and such that $\mathscr{R}$ contains a critical point set and $\mathscr{R}$ is contained in the set of extremal points for some best approximation.

The above theorems describe very closely the situations under which best approximations are not unique. It is clear that there is no restricted Tchebycheff approximation problem of interest in more than one variable for which best approximations are unique. On the other hand it is seen that the occurrence of nonuniqueness is a very unlikely event for a randomly selected $f(x)$.

5. Alternation and Tchebycheff sets. In the theory of Tchebycheff approximation for one real variable, the concepts of alternation, or oscillation, and Tchebycheff sets hold a central position. It is natural to try to discover the generalization of these concepts to functions of more than one variable. This possibility is discussed in this section and the results are mainly of a negative nature.

The first attempt to generalize from one real variable appears to have been made by Tonelli [13]. He discussed Tchebycheff approximation by a function of one complex variable. His work was simplified and expanded upon by de la Vallée Poussin [14]. Already some of the difficult nature of the problem appeared, though uniqueness is still present in this case.

If $P_n(z)$ is a best approximation to $f(z)$ by complex polynomials of degree n let

$$p = \max_{z \in \mathscr{B}} \left| P_n(z) - f(z) \right|.$$

Some of de la Vallée Poussin's results are stated (in a different form) in the following theorem.

THEOREM (DE LA VALLÉE POUSSIN). *The approximatng polynomial P_n of degree n for a finite set of points gives critical point sets containing at least $n + 2$ points and at most $2n + 3$ points.*

Thus, much of the simplicity of the theory of Tchebycheff approximation in one real variable is already lost for one complex variable.

In order to understand what Tchebycheff sets do for the theory in one real variable we may examine the concept of unisolvent functions [7]. These functions are essentially an abstraction of polynomials, with only the essential properties required to give the usual proofs of the standard theorems of Tchebycheff approximation. It has been shown [9] that under certain broad conditions these properties are also necessary for the proofs of the standard theorems. The two essential assumptions for a unisolvent function with n parameters defined on $[0,1]$ are:

SOLVENCE. *Given n distinct points in $[0,1]$ and n real numbers, then there exists a function which assumes the given values at the given points.*

PROPERTY Z. *The difference of any two unisolvent functions has at most* $n-1$ *zeros in* [0,1].

The first of these assumptions allows one to place a lower bound $(n + 1)$ on the number of points in a critical point set. The second allows one to place an upper bound (also $n + 1$) on the number of points in a critical point set. If one generalizes the definition of a Tchebycheff set to functions of several variables, then one would expect the corresponding assumptions to play the same role. However, one cannot expect the resulting upper and lower bounds on the number of points in a critical point set to be the same. This may be seen from the example of polynomials in two variables.

A polynomial of degree n in two variables has $(n + 1)(n + 2)/2$ parameters and hence one would expect a critical point set to contain $(n + 1)(n + 2)/2 + 1$ points. However, these polynomials only have $n + 1$ free parameters along any straight line, for they are essentially reduced to a polynomial in one variable. It is easy to construct functions of two variables whose critical points lie on any given line. Thus the lower bound on the number of points in a critical point set for polynomials of two variables is the same as for polynomials of one variable, namely $n + 2$. On the other hand the upper bound is $(n + 1)(n + 2)/2 + 1$.

The definition of alternation is as follows: The deviation $\max |f(x) - P_n(x)|$ is said to *alternate* n *times* in [0, 1] if there are $n + 1$ points $0 \le x_1 < x_2 < \cdots < x_{n+1} \le 1$ such that

$$f(x_i) - P_n(x_i) = -[f(x_{i+1}) - P_n(x_{i+1})] = \pm \max_{x \in [0,1]} |f(x) = P_n(x)|.$$

The concept of alternation is fundamental in the Tchebycheff theory of one real variable. Its main function is to allow one to easily recognize best approximations. A form of alternation is still present for functions of several variables. However, there is apparently no simple method of distinguishing between the alternation of a best approximation and the alternation of other approximating functions. This is due to the fact that there is no natural ordering of the critical points.

Another use of alternation for one variable is to allow one to easily extract critical point sets from the set of extremal points of a best approximation. For functions of more than one variable this selection is no longer a trivial problem.

Since neither of these fundamental items of the one variable theory appear to generalize easily, the question of what to do arises. The approach taken in this paper is to completely ignore the concept of Tchebycheff sets. That it is not a truly essential requirement for the theory of Tchebycheff approximation is seen when the theory to be developed is viewed in the special case of one real variable. Tchebycheff sets merely allow one to explicitly determine the number of points in a critical point set for a certain class of special cases. It might be remarked that the set of n-tuples of continuous functions which form Tchebycheff sets is a very small subset of the set of all n-tuples of continuous functions.

One cannot ignore alternation for it is an essential feature of Tchebycheff approximations. This concept appears in the form of "isolability" in the appoach taken here. Later it is shown how this may be interpreted in relation to the theory of systems of linear inequalities.

6. Characterization of best approximations.   Since it is not possible to explicitly construct best Tchebycheff approximations, the theorems which give characterizing properties of a best approximation hold a central position in the Tchebycheff theory. The theorem given here is extremely easy to prove, in fact it is almost a tautology, yet it is of fundamental importance in the actual determination of best approximations.

THEOREM 5. $L(A^*, x)$ *is a best approximation to* $f(x)$ *if and only if the set* $\mathscr{E}$ *of extremal points of* $L(A^*, x) - f(x)$ *contains a critical point set.*

Proof. If $\mathscr{E}$ contains no critical point set set then the positive and negative parts of $\mathscr{E}$ may be isolated, say by $L(A_0, x)$. Then λ may be chosen so that $L(A^*, x) + \lambda L(A_0, x)$ is a better approximation to $f(x)$ than $L(A^*, x)$.

Assume that $L(A_0, x)$ is a better approximation to $f(x)$ than $L(A^*, x)$ and that $\mathscr{E}$ contains a critical point set $\mathscr{R}$. Then

$$L(A^*, x) - L(A_0, x)$$

isolates the positive and negative parts of $\mathscr{E}$, and hence $\mathscr{R}$. This is a contradiction.

7. Approximation on finite subsets.   The nature of the Tchebycheff norm tends to greatly emphasize the importance of certain finite subsets of $\mathscr{B}$. These are the extremal points and, especially, the critical point sets. The intimate relation between approximation on $\mathscr{B}$ with approximation on particular subsets is brought out by the following theorem.

THEOREM 6.   *Let* $L(A^*, x)$ *be a best approximation to* $f(x)$ *with a critical point set* $\mathscr{R}$ *of* k *points. Then* $L(A^*, x)$ *is a best approximation to* $f(x)$ *on* $\mathscr{R}$ *and is characterized as a best approximation with the largest deviation* ρ *among all best approximations to* $f(x)$ *on subsets of* k *points of* $\mathscr{B}$.

Proof. It has been established in Lemma 1 that $L(A^*, x)$ is a best approximation to $f(x)$ on $\mathscr{R}$. It is clear that

$$\rho \geqq \max |L(A^*, x) - f(x)|$$

for $\mathscr{R}$ is one of the subsets of $\mathscr{B}$ of k points. On the other hand

$$\rho \leqq \max |L(A^*, x) - f(x)|$$

since the deviation of a best approximation on a subset of $\mathscr{B}$ must be less than or equal to the deviation of a best approximation on $\mathscr{B}$.

8. **De la Vallée Poussin algorithm.** If $\mathscr{B}$ contains an infinite number of points, then the exact calculation of a best approximation to a given $f(x)$ is in general impossible. However, one may use an algorithm of de la Vallée Poussin to obtain an estimate of a best approximation. The algorithm is as follows. One chooses a set

$$\mathscr{X} = \{x_\alpha \mid x_\alpha \in \mathscr{B}\}$$

and then determines a best approximation $L(A_\beta, x)$ to $f(x)$ on

$$\mathscr{X}_\beta = \{x_\alpha \mid \alpha \leqq \beta\}.$$

If $\mathscr{X}$ is dense in $\mathscr{B}$, then one expects that $L(A_\beta, x)$ tends to a best approximation of $f(x)$ on $\mathscr{B}$ as β increases. If $\mathscr{B}$ is separable and $\mathscr{X}$ can be taken as a countable set then this method can be used for the actual calculation of best approximations.

For functions of one variable the convergence of this algorithm has been established by de la Vallée Poussin [15] and the rate of convergence has been discussed by Rice [10]. For approximation on a compact metric space the convergence of the algorithm has been established by Cheney and Goldstein [3, §29]. For approximation of functions of more than one variable the rate of convergence of this algorithm is slower and hence more critical.

A function on $\mathscr{B}$ is said to satisfy a Hölder condition of exponent α if there is a λ such that

$$|f(x) - f(y)| \leqq \lambda \|x - y\|_{\mathscr{B}}^\alpha .$$

Let $\mathscr{A}$ be a compact subset of $\mathscr{B}$ and set

$$\delta = \max_{x \in \mathscr{B}} \min_{y \in \mathscr{A}} \|x - y\|.$$

Let $L(A_0, x)$ be a best approximation to $f(x)$ on $\mathscr{A}$ and $L(A^*, x)$ a best approximation to $f(x)$ on $\mathscr{B}$. Set

$$\rho^* = \max_{x \in \mathscr{B}} |L(A^*, x) - f(x)|,$$

$$\rho_0 = \max_{x \in \mathscr{A}} |L(A_0, x) - f(x)|.$$

THEOREM 7. *If $f(x)$ and the $g_i(x)$ satisfy a Hölder condition with exponent α, then*

$$|\rho^* - \rho_0| \leqq k\delta^\alpha,$$

where k is a constant depending only on $f(x)$ and $\mathscr{L}$.

Proof. Let

$$\mathscr{L}_0 = \left\{ |L \in \mathscr{L}, \max_{x \in \mathscr{A}} |L(A, x) - f(x)| \leqq \rho_0 + 1 \right\}.$$

It can be shown [1] that $\mathscr{L}_0$ is a bounded subset of $\mathscr{L}$. Hence if $L \in \mathscr{L}_0$ then $\mathscr{L}$ satisfies a Hölder condition with exponent α.

Let $x_0 \in \mathscr{B}$ be a point such that

$$\left| L(A_0, x_0) - f(x_0) \right| = \max_{x \in \mathscr{A}} \left| L(A_0, x) - f(x) \right|$$

and $y_0 \in \mathscr{A}$ be a point such that

$$\| x_0 - y_0 \| \leq \delta.$$

Then

$$\rho^* \leq \max_{x \in \mathscr{B}} \left| L(A_0, x) - f(x) \right| = \left| L(A_0, x_0) - f(x_0) \right|$$

$$\leq \left| L(A_0, x_0) - L(A_0, y_0) \right| + \left| L(A_0, y_0) - f(y_0) \right| + \left| f(y_0) - f(x_0) \right|$$

$$\leq k\delta^\alpha + \max_{x \in \mathscr{A}} \left| L(A_0, x) - f(x) \right| = k\delta^\alpha + \rho_0.$$

It is clear that $\rho_0 \leq \rho^*$ and hence the theorem is established.

COROLLARY. *Let $\mathscr{B}$ be a compact subset of $\mathscr{E}_n$ and $\mathscr{A} = \{x_i | i = 1, 2, \cdots\}$ a count-able dense subset of $\mathscr{B}$. Let ρ_m be the deviation of a best approximation to $f(x)$ on the first m points of $\mathscr{A}$. Then if $f(x)$ and $\{g_i(x)\}$ satisfy a Hölder condition of exponent α*

$$\left| \rho^* - \rho_m \right| \leq k \left(\frac{1}{m} \right)^{\alpha/n}.$$

One can easily choose a subset $\mathscr{A}$ of $\mathscr{B}$ so that

$$\delta_m = \max_{x \in \mathscr{B}} \quad \min_{1 \leq i \leq m} \| x - x_i \|$$

is of the order of $m^{-1/n}$. This establishes the corollary. One cannot in general improve on the estimate

$$\delta_m \approx m^{-1/n}$$

and hence as n becomes large the rate of convergence of the the de la Vallée Poussin algorithm becomes very slow.

In [10] a method of accelerating the convergence of this algorithm is described for approximation in one real variable. Since the acceleration is very considerable, such a process should be especially profitable for functions of several variables.

9. **Strict approximations.** In this section a new type of best Tchebycheff approximation for finite point sets is defined and some of its properties are developed. These approximations are a subset of the set of best Tchebycheff approximations. These approximations may be viewed as the natural generalization of Tchebycheff approximations for functions of several variables defined on a finite point set. The resulting theory is very similar to that for functions of one variable and the important theorems (including uniqueness) have a direct extension to these approximations.

There is some machinery which must be established in order to discuss these approximations. The n-vector $\mathbf{g}$ is defined for a point x of $\mathscr{B}$ by

$$\mathbf{g} = \mathbf{g}(x) = (g_1(x), g_2(x), \cdots, g_n(x)).$$

This defines a mapping G from $\mathscr{B}$ into Euclidean n-space $\mathscr{E}_n$.

$$G : x \to \mathbf{g}(x)$$

Square brackets, [ ], denote the closure by linear combinations of a set of vectors, or equivalently, denote the smallest linear subspace of $\mathscr{E}_n$ containing a set of vectors. Thus

$$[\mathbf{g}_1, \mathbf{g}_2] = \{\mathbf{g} \mid \mathbf{g} = a\mathbf{g}_1 + b\mathbf{g}_2, |a| \leqq \infty, \quad |b \leqq \infty\}.$$

The *dimension* of a set $\mathscr{R}$ in $\mathscr{B}$ with respect to the mapping G is defined as the dimension of the linear subspace $[G(\mathscr{R})]$. Thus the dimension of a set $\mathscr{R}$ is the dimension of the subspace of $\mathscr{E}_n$ spanned by the vectors $\mathbf{g}(x)$ for $x \in \mathscr{R}$. A point $x \in \mathscr{B}$ is said to be dependent on a set $\mathscr{R} \subset \mathscr{B}$ if $\mathbf{g}(x)$ is contained in $[G(\mathscr{R})]$.

In order to make the approximation problem interesting one must require that $\mathscr{B}$ contains a sufficient number of points. For approximation to functions of one variable by a Tchebycheff set it is sufficient to assume that $\mathscr{B}$ contains $n + 1$ points. The general situation requires a slightly more complicated condition. A set $\mathscr{R}$ of points in $\mathscr{B}$ is said to be *nondegenerate* if there is no point x_0 in $\mathscr{R}$ such that $\mathbf{g}(x_0)$ is linearly independent of the set of vectors

$$\{\mathbf{g}(x) \mid x \in \mathscr{R}, \quad x \neq x_0\}.$$

The following is assumed throughout the remainder of this section:

ASSUMPTION. $\mathscr{B}$ is nondegenerate.

In the original definition of a critical point set $\mathscr{R}$, the positive and negative parts of $\mathscr{R}$ cannot be isolated by any element of $\mathscr{L}$. We define a *critical point set with respect to* $\mathscr{L}_1 \subset \mathscr{L}$ as a set $\mathscr{R}$ whose positive and negative parts cannot be isolated by a member of $\mathscr{L}_1$.

DEFINITION. *The following construction is used to define a strict approximation to $f(x)$ on $\mathscr{B}$. Let $\mathscr{R}_1$ be the set of critical point sets of best approximations to $f(x)$ on $\mathscr{B}$. Define a set $\mathscr{B}_1$ by*

$$\mathscr{B}_1 = \{x \mid \mathbf{g}(x) \in [G(\mathscr{R}_1)]\}$$

and denote by $\mathscr{L}_1$ the set of best approximations to $f(x)$ on $\mathscr{B}$. Then $\mathscr{L}_2$ is defined as the set of $L(A^*, x)$ such that

$$\| L(A^*, x) - f(x) \|_{\mathscr{B} - \mathscr{B}_1} \leqq \| L(A, x) - f(x) \|_{\mathscr{B} - \mathscr{B}_1}$$

for all $A \in \mathscr{L}_1$. Let $\mathscr{R}_2$ be the set of critical point sets with respect to $\mathscr{L}_1$ of the approximations of $\mathscr{L}_2$. Define a set $\mathscr{B}_2$ by

$$\mathcal{B}_2 = \{x \mid \mathbf{g}(x) \in [G(\mathcal{R}_1 \cup \mathcal{R}_2)]\}$$

and denote by $\mathcal{L}_2$ the set of $L(A^*, x)$ such that

$$\| L(A^*, x) - f(x) \|_{\mathcal{B} - \mathcal{B}_2} \leqq \| L(A, x) - f(x) \|_{\mathcal{B} - \mathcal{B}_2}$$

for all $A \in \mathcal{L}_2$. This construction is carried until $\mathcal{B}_k = \mathcal{B}$. The elements of $\mathcal{L}_k$ are said to be strict approximations to $f(x)$ on $\mathcal{B}$.

On each of the sets $\mathcal{B} - \mathcal{B}_i$ the strict approximations have a certain maximum deviation. This maximum deviation is clearly assumed by every strict approximation on every set of critical point sets, $\mathcal{R}_i$. Denote this deviation by d_i and the dimension of $\mathcal{R}_i$ by m_i. The *deviation vector* $\mathbf{d}$ is defined to be the vector whose first m_1 components are d_1, whose next m_2 components are d_2, and so forth.

$$\mathbf{d} = \frac{\overbrace{m_1}}{(d_1, \cdots, d_1,}\ \frac{\overbrace{m_2}}{d_2, \cdots, d_2,}\ d_3, \cdots, d_{k-1}, \frac{\overbrace{m_k}}{d_k, \cdots, d_k)}.$$

In the analysis of approximations it is necessary to compare the "size" of the deviation of various approximations on subsets of $\mathcal{B}$. For functions of one variable and approximations using Tchebycheff sets the components of the deviation vector are all equal and thus there is a simple ordering of the magnitudes of the deviations. The natural ordering of the deviation vectors of strict approximations is the *lexicographic ordering*. Thus a strict approximation with deviation vector $(2, 2, 1, 1)$ has a smaller deviation than strict approximations with deviation vectors $(2, 2, 2, 1)$ and $(3, 0, 0, 0)$ and a larger deviation than a strict approximation with deviation vector $(2, 2, 1, 0)$.

A *strict critical point set* is the union of one critical point set with respect to $\mathcal{L}_i$ from each of the sets $\mathcal{R}_{i+1}$. One may determine independently of any strict approximation to $f(x)$ whether a given ordered collection of sets forms a critical point set or not. For if the collection $\{\mathcal{S}_i\}$ forms a strict critical point set then none of the following systems of equalities and inequalities has a solution. The positive and negative part of $\mathcal{S}_i$ are denoted by $\mathcal{P}_i$ and $\mathcal{N}_i$, respectively.

$$
\begin{aligned}
L(A_1, x) &> 0 & x &\in \mathcal{P}_1, \\
L(A_1, x) &< 0 & x &\in \mathcal{N}_1, \\[6pt]
L(A_2, x) &= 0 & x &\in \mathcal{S}_1, \\
L(A_2, x) &> 0 & x &\in \mathcal{P}_2, \\
L(A_2, x) &< 0 & x &\in \mathcal{N}_2, \\
&\qquad \cdot \quad \cdot \quad \cdot \\
L(A_i, x) &= 0 & x &\in \bigcup_{j=1}^{i-1} \mathcal{S}_j, \\
L(A_i, x) &> 0 & x &\in \mathcal{P}_i, \\
L(A_i, x) &< 0 & x &\in \mathcal{N}_i, \\
&\qquad \cdot \quad \cdot \quad \cdot
\end{aligned}
$$

(1)

Any point in $\mathscr{B} - \mathscr{B}_i$ where

$$\left| L(A^*, x) - f(x) \right| = \left\| L(A^*, x) - f(x) \right\|_{\mathscr{B} - \mathscr{B}_i}$$

for some strict approximation $L(A^*, x)$ is said to be an extremal point of the strict approximation.

The following lemma establishes several simple facts relative to the above definitions.

LEMMA 2. (i) *The sets $\mathscr{R}_i$ are not empty if $d_i > 0$.*
(ii) *The sets $\mathscr{R}_i$ and $\mathscr{B}_i$ are uniquely determined.*
(iii) *There are at most n sets $\mathscr{R}_i$.*
(iv) *The deviation vector of a strict approximation has exactly n components.*

Proof. If $d_i = 0$, then there may not be any critical point sets with respect to $\mathscr{L}_{i-1}$. However, if $d_i > 0$ then $\mathscr{B} - \mathscr{B}_{i-1}$ must contain a set of extremal points, since $\mathscr{B} - \mathscr{B}_{i-1}$ is compact. Clearly this set of extremal points contains a critical point set with respect to $\mathscr{L}_{i-1}$.

By Theorem 1 the set $\mathscr{R}_1$ is unique and hence $[G(\mathscr{R}_1)]$ and $\mathscr{B}_1$ are uniquely determined. The same argument as in Theorem 1 shows that the set $\mathscr{R}_i$ of critical point sets with respect to each $\mathscr{L}_i$ is unique. This establishes (ii).

From the definition of the sets $\mathscr{B}_i$ it follows that $\mathbf{g}(x)$, $x \in \mathscr{R}_i$ is independent of $[G(\mathscr{B}_{i-1})]$. Thus $[G(\mathscr{B}_i)]$ is a subspace of $\mathscr{E}_n$ of at least one more dimension than $[G(\mathscr{B}_{i-1})]$. Since $\mathscr{E}_n$ has dimension n there can be at most n sets $\mathscr{R}_i$. This establishes (iii).

Since $\mathscr{B}$ is nondegenerate, $\mathscr{B}_k = \mathscr{B}$ only when

$$\mathscr{E}_n = \left[G\left(\bigcup_{i=1}^{k} \mathscr{R}_i \right) \right].$$

The difference in dimension between $\mathscr{B}_{i-1}$ and $\mathscr{B}_i$ is the dimension m_i of $\mathscr{R}_i$. Thus the dimension of $\mathscr{B}_k$ is

$$\sum_{i=1}^{k} m_i = n.$$

This establishes (iv).

We can now consider for strict approximations the three basic problems of approximation theory. These are the problems of existence, uniqueness and characterization of best approximations. The question of existence is easily answered. It is well known that there exists a best approximation to $f(x)$ on $\mathscr{B}$. Hence $\mathscr{L}_1$ is not empty. The usual classical arguments show that $\mathscr{L}_i$ is not empty since the set $\mathscr{L}_{i-1}$ is closed and $\mathscr{B} - \mathscr{B}_{i-1}$ is compact.

The uniqueness of the strict approximation is now established.

THEOREM 8. *The strict approximation is unique.*

Proof. Let (A^*, x) be a strict approximation to $f(x)$ on $\mathscr{B}$. Denote by $\mathscr{R}_i$, $1 \leq i \leq k$, the ith set of critical point sets with respect to $\mathscr{L}_i$ of the set of strict critical point sets. Denote by $\mathscr{P}_i$, $\mathscr{N}_i$ and m_i the positive part, negative part and dimension of $\mathscr{R}_i$, respectively. Choose m_i linearly independent points, x_{ij} $j = 1, 2, \cdots, m_i$, from $\mathscr{R}_i$. Then $L(A^*, x)$ satisfies the following system of linear equalities

(2)
$$L(A^*, x_{ij}) = f(x_{ij}) - d_i \qquad x_{ij} \in \mathscr{P}_i,$$

$$L(A^*, x_{ij}) = f(x_{ij}) + d_i \qquad x_{ij} \in \mathscr{N}_i.$$

From the assumption that $\mathscr{B}$ is nondegenerate it follows that the set of vectors $\{\mathbf{g}(x_{ij})\}$ span the space $\mathscr{E}_n$. Hence there exists a subset containing n of them which are linearly independent. Thus the parameters A^* satisfy a subsystem of (2) of n linear equations with a nonsingular coefficient matrix. Hence $L(A^*, x)$ is the unique strict approximation which satisfies (2).

It only remains to be shown that every strict approximation satisfies (2). This follows directly by the argument of Theorem 1 which shows that every strict approximation must assume the same values at every point of a strict critical point set. This concludes the proof.

The characterization of strict approximations is similar to that of best approximations given in Theorem 5. It is also a natural generalization of the characterizing property for Tchebycheff approximation for functions of one variable.

THEOREM 9. *$L(A^*, x)$ is the strict approximation to $f(x)$ if and only if the set $\mathscr{E}$ of extremal points of $L(A^*, x) - f(x)$ contains a strict critical point set of dimension n.*

Proof. From the definition of a strict critical point set, one cannot improve upon the strict approximation associated with it. For the existence of a better approximation would imply that one of the systems (1) has a solution which is a contradiction. This establishes the "if" part of the proof.

Since $\mathscr{B}$ is nondegenerate the procedure defining the strict approximation terminates only when

$$\left[G \left(\bigcup_{i=1}^{k} \mathscr{R}_i \right) \right] = \mathscr{E}_n.$$

Thus, by construction, $\mathscr{E}$ contains a strict critical point set of dimension n.

In the theory of Tchebycheff approximation there is an intimate relation between the best approximation on $\mathscr{B}$ and best approximations on certain finite subsets of $\mathscr{B}$. This is illustrated by Theorem 6. A similar result is valid for strict approximations.

THEOREM 10. *Let $L(A^*, x)$ be the strict approximation to $f(x)$ with deviation*

vector **d*** *and whose strict critical point set $\mathscr{R}$ contains k points. Then $L(A^*, x)$ is the strict approximation to $f(x)$ on $\mathscr{R}$ and is characterized as the strict approximation with largest deviation vector* **d** *among all strict approximations to $f(x)$ on nondegenerate subsets of k points of $\mathscr{B}$.*

Proof. It is clear that $L(A^*, x)$ is the strict approximation to $f(x)$ on $\mathscr{R}$.

The deviation vector of the strict approximation on a subset of $\mathscr{B}$ is smaller than the deviation vector of the strict approximation on all of $\mathscr{B}$. Thus

$$\mathbf{d^*} \geqq \mathbf{d}.$$

Since $\mathscr{R}$ is one of the nondegenerate subsets of k points of $\mathscr{B}$, it is seen that

$$\mathbf{d^*} \leqq \mathbf{d}.$$

This concludes the proof.

In the preceding analysis the assumption that $\mathscr{B}$ is a finite point set is used at only one point. That is in part (i) of Lemma 2, where the assumption is essential.

The basic technique of strict approximation is to isolate the effect of the critical points of $\mathscr{L}_k$ from the determination of $L(A_{k+1}, x)$. If $\mathscr{B}$ is a finite point set this is simply done by deleting the critical point sets and all points of $\mathscr{B}$ linearly dependent on them. If one of the critical points of $\mathscr{L}_k$ is the limit of points of $\mathscr{B} - \mathscr{B}_k$, then by the continuity of $f(x)$ and $L(A, x)$, there is no decrease in the deviation of strict approximations on $\mathscr{B} - \mathscr{B}_k$. Furthermore, there are no critical point sets in $\mathscr{B} - \mathscr{B}_k$ and the defining process for strict approximations breaks down.

An obvious method of eliminating this difficulty is to delete a small neighborhood of each point of the critical point sets. The problem is to determine the exact neighborhood to be deleted. One can choose spheres of fixed radius as the neighborhoods and establish results analogous to those found above. However, this is not a natural choice. The neighborhoods chosen should reflect the intuitive meaning of the phrase "isolate the effect of the critical points." The choice of these neighborhoods and, more generally, the extension of the concept of strict approximation to general sets $\mathscr{B}$ is an open question.

10. **The method of ascent.** For functions of one real variable there are two basic methods for computing Tchebycheff approximations. The first is the method of descent. With this method one interprets the Tchebycheff approximation problem as a linear programming problem and then applies computation methods from linear programming. Several algorithms of this type have been described by Cheney and Goldstein [3].

The second of these methods is the method of ascent. There is some ambiguity in this distinction since some methods of ascent are the duals of a method of descent, but this relation will not be discussed here. For functions of more than one variable there are some complications which are not present for one variable. Since one apparently cannot interpret alternation geometrically for functions of

several variables, it appears that the Remez algorithm [8] cannot be easily generalized to functions of more than one variable. However, the 1 for 1 exchange algorithms can be so modified and such modifications are described in this section.

The basic idea of the 1 for 1 exchange algorithm is as follows: Given a best approximation $L(A_0, x)$ on a subset and its critical point set $\mathscr{R}_0$ one locates x_1 so that

$$\left| L(A_0, x_1) - f(x_1) \right| = \max_{x \in \mathscr{B}} \left| L(A_0, x_1) - f(x_1) \right|.$$

One then adds x_1 to the set $\mathscr{R}_0$ to form $\mathscr{S}_1$ and determines a best approximation $L(A_1, x)$ on $\mathscr{S}_1$. A critical point set $\mathscr{R}_1$ is determined in $\mathscr{S}_1$ and the process is repeated.

Normally $\mathscr{R}_0$ and $\mathscr{R}_1$ will have the same number of points and hence a 1 for 1 exchange is made. However, the number of points in a critical point set need not be fixed and thus the exchange may not always be 1 for 1. If one has an algorithm which computes the strict approximation to $f(x)$ on $\mathscr{B}$, then one also has an algorithm for a best approximation to $f(x)$ on $\mathscr{B}$. On the other hand one might hope that the computation of a best approximation is simpler than the computation of the strict approximation. Thus the computation of the strict approximation will be described and then possible simplifications will be indicated for the computation of a best approximation.

It should be noted that a difference in the methods will occur only when there is not a unique best approximation. This is not as rare in practice as one might assume from the comment at the end of §4. This is due to the fact that one most commonly approximates functions on rectangular or regularly shaped regions. This means that the boundaries are formed from isolating curves which leads to a rather frequent lack of uniqueness in practice.

The formal statement of this algorithm is the same for functions defined on a Banach space as for functions of one variable. Assume that, at the kth step, one has found a strict approximation $L(A_k, x)$ to $f(x)$ on a nondegenerate strict critical subset $\mathscr{R}_k$ of $\mathscr{B}$ with deviation vector $\mathbf{d}_k$.

Assume that $\mathscr{R}_k$ is composed of independent critical point sets $\mathscr{S}_{ik}$ as determined in the definition of the strict approximation. Let d_{ik} be the deviation of $L(A_k, x)$ on $\mathscr{S}_{ik}$. The sets $\mathscr{S}_{ik}$ are ordered according to the size of d_{ik}. The next step of the algorithm is then

1. Find $x_{k+1} \in \mathscr{B}$ such that $\mathbf{g}(x_{k+1})$ is dependent on $[G(\bigcup_j \mathscr{S}_{jk})]$ for some set of j and such that

$$\left| L(A_k, x_{k+1}) - f(x_{k+1}) \right| = d_0 > \min_j d_{jk}.$$

If no such x_{k+1} can be found then $L(A_k, x)$ is the strict approximation to $f(x)$ on $\mathscr{B}$.

2. Determine the strict approximation $L(A_{k+1}, x)$ on $\mathscr{R}_k \cup \{x_{k+1}\}$ and choose $\mathscr{R}_{k+1}$ as a strict critical point set of $L(A_{k+1}, x)$.

If $\mathscr{B}$ is a finite point set then the following lemma establishes the convergence of this algorithm.

LEMMA 3. *The deviation vector* $\mathbf{d}_{k+1}$ *determined by Steps* 1 *and* 2 *above is larger than* $\mathbf{d}_k$.

Proof. Let $\mathscr{S}_j$, $j = 1, 2, \cdots, p$ be the smallest collection of $\mathscr{S}_{ik}$ upon which $\mathbf{g}(x_{k+1})$ is dependent. If this collection contains only one set $\mathscr{S}_1 = \mathscr{S}_{jk}$ then the strict approximation on $\mathscr{R}_k \cup \{x_{k+1}\}$ will have the same deviations on $\mathscr{R}_i$, $i < j$ and clearly the deviation is increased on $\mathscr{R}_k \cup \{x_{k+1}\}$.

If this collection contains several sets then $\mathbf{g}(x_{k+1})$ is independent of each subspace $[G(\mathscr{S}_j)]$. Let $\mathscr{S}_p$ be the member of this collection with the smallest deviation. Consider the systems of equations (1). From these equations it follows that there are approximations $L(A, x)$ such that

$$L(A, x) = 0 \qquad x \in \bigcup_{j=1}^{p-1} \mathscr{S}_j.$$

Hence, from this set of approximations one cannot determine an approximation such that

$$L(A, x) \;\; = \;\; 0 \qquad x \in \mathscr{S}_p,$$

$$L(A, x_{k+1}) \;\neq\; 0,$$

for this contradicts the definition of the collection $\{\mathscr{S}_j\}$. $L(A_k, x)$ is the unique approximating function which minimizes the deviations d_{jk} and thus $L(A_{k+1}, x)$, being different from $L(A_k, x)$, has a larger deviation vector. This concludes the proof.

COROLLARY. *If* $\mathscr{B}$ *is a finite point set then the* 1 *for* 1 *exchange algorithm converges to the strict approximation in a finite number of steps.*

Proof. Since $\mathscr{B}$ is finite there are a finite number of nondegenerate subsets. The algorithm strictly increases the deviation vector at each step and by Theorem 10 the strict approximation is reached in a finite number of steps.

If one is only interested in best approximation rather than strict approximations, there is one obvious simplification that can be made in the algorithm. That is in Step 1 choose only those x_{k+1} such that

$$\left| L(A_k, x_{k+1}) - f(x_{k+1}) \right| = d_0 > d_1.$$

One would actually choose x_{k+1} so that

$$(3) \qquad\qquad \left| L(A_k, x_{k+1}) - f(x_{k+1}) \right| = \left\| L(A_k, x) - f(x) \right\|_{\mathscr{B}}.$$

Thus one may or may not arrive at the strict approximation to $f(x)$ on $\mathscr{B}$.

It is somewhat difficult to avoid computing strict approximations on the sets

$\mathscr{R}_k$. For if one does not have a strict approximation on $\mathscr{R}_k$ and if $g(x_{k+1})$ is independent of $[G(\mathscr{R}_k)]$, then one cannot discard x_{k+1}, but on the other hand one does not know where to place it. One can allow $\mathscr{R}_k$ to be degenerate and when such an x_{k+1} is obtained, put it in a set $\mathscr{S}_{jk+1}$ by itself. In this way one would either build up to a nondegenerate set $\mathscr{R}_{k+j}$ or find a point x_{k+j} such that $g(x_{k+j})$ is dependent upon $[G(\mathscr{R}_{k+j-1})]$. In any case if x_{k+1} is chosen according to (3) then the algorithm may be shown to converge to a best approximation to $f(x)$ on $\mathscr{B}$.

Since the 1 for 1 exchange algorithm operates only on finite subsets of $\mathscr{B}$, one may also apply it to a general set $\mathscr{B}$. One can then consider the convergence of the algorithm in such a case. Apparently, no one has published a proof of convergence of this algorithm even for functions of one variable, though there is no doubt that such a proof can be given.

11. **Approximation on critical point sets.** The algorithm described in the preceding section requires one to compute the strict approximation on the union of a strict critical subset $\mathscr{R}_k$ and one additional point of $\mathscr{B}$. This computation is nontrivial and some indications of a method of performing this computation are given in this section.

The computational procedures can be described with concepts already found in the theory of systems of linear inequalities. A resumé of the pertinent results which may not be generally familiar is given. See, for example, [5].

Let f_i be given linear functionals on $\mathscr{B}$ and let the α_i be given real numbers. Consider the system of linear inequalities

$$(4) \qquad\qquad f_i(x) \geqq \alpha_i \qquad i = 1, 2, \cdots, p.$$

THEOREM (PRINCIPLE OF BOUNDING SOLUTIONS). *Let* (4) *be consistent and let* $r \geqq 1$ *be the maximum number of linearly independent* f_i. *Then there exists* r *linearly independent functionals* f_{v_j}, $j = 1, 2, \cdots, r$ *such that every solution of the system*

$$f_{v_j}(x) = \alpha_{v_j} \qquad j = 1, 2, \cdots, r$$

is a solution of (4).

The concept of an irreducibly inconsistent system [2; 5] is intimately related to that of a critical point set.

$$(5) \qquad\qquad f_i(x) > \alpha_i \qquad i = 1, 2, \cdots, p.$$

DEFINITION. *The system* (5) *is irreducibly inconsistent if it is inconsistent, but every proper subsystem is consistent.*

Carver [2] has established the following theorem for a system of strict inequalities.

THEOREM. *The system* (5) *is irreducibly inconsistent if and only if*:

(i) *every $p - 1$ of the f_i are linearly independent,*

(ii) *there exist $\lambda_i > 0$ such that*

$$\sum_{i=1}^{p} \lambda_i f_i = 0,$$

$$\sum_{i=1}^{p} \lambda_i \alpha_i \geqq 0.$$

If $\mathscr{R}$ is a critical point set with positive and negative parts, $\mathscr{P}$ and $\mathscr{N}$, respectively, then the system

$$L(A, x) > 0 \qquad x \in \mathscr{P},$$

$$- L(A, x) > 0 \qquad x \in \mathscr{N},$$

is irreducibly inconsistent. Conversely, if the above system is irreducibly inconsistent then $\mathscr{R}$, with this particular division into positive and negative parts, is a critical point set.

At Step 2 of the algorithm one has a point x_{k+1} which is dependent on a collection $\{\mathscr{T}_j \mid j = 1, 2, \cdots, p\}$ of critical point sets $\mathscr{S}_{jk}$. In this section various subcollections of $\{\mathscr{T}_j\}$ are discussed, but they are also denoted by $\mathscr{T}_j$ and p is a generic number representing the number of sets in $\{\mathscr{T}_j\}$. The positive and negative parts of $\mathscr{T}_j$ and the deviation on $\mathscr{T}_j$ are denoted by $\mathscr{U}_j, \mathscr{V}_j$ and c_j, respectively. The sets $\{\mathscr{T}_j\}$ are always assumed to be ordered according to the magnitude of of the corresponding component of the deviation vector. Let $\{\mathscr{T}_j\}$ be the smallest collection from $\{\mathscr{S}_{jk}\}$ upon which x_{k+1} depends.

The actual computation is made with the systems of equations and inequalities which define a strict critical point set. The points of the sets $\mathscr{T}_j$ for which $d_j \geqq d_0$ may be eliminated from consideration for exchange with x_{k+1}. The collection $\{\mathscr{T}_j\}$ is redefined to exclude these sets. The first step in the computation of $L(A_{k+1}, x)$ is to impose the condition that

$$(6) \qquad \begin{aligned} L(A_{k+1}, x) - f(x) &= d_j & x \in \mathscr{P}_{jk} \\ L(A_{k+1}, x) - f(x) &= - d_j & x \notin \mathscr{N}_{jk} \end{aligned} \qquad x \notin \bigcup_{j=1}^{p} \mathscr{T}_j.$$

If $p = 1$ then the determination of $L(A_{k+1}, x)$ is relatively simple. $\mathscr{T}_1 = \mathscr{S}_{jk}$ contains $m_j + 1$ points. In order to determine $L(A_{k+1}, x)$ one solves the system (6) along with

$$L(A_{k+1}, x) - f(x) = d \qquad x \in \mathscr{U}_1,$$

$$L(A_{k+1}, x) - f(x) = - d \qquad x \in \mathscr{V}_1$$

for all subsets of $\mathscr{T}_1 \cup \{x_{k+1}\}$ containing $m_j + 1$ points. This gives a system of $n + 1$ equations in $n + 1$ unknowns, namely, the parameters A_{k+1} and the de-

viation component d. The subset of $\mathscr{T}_1 \cup \{x_{k+1}\}$ which maximizes d is the new critical subset of the strict approximation $L(A_{k+1}, x)$. A reordering of the set $\{\mathscr{S}_{jk+1}\}$ may be necessary.

If $p > 1$ then one approach is the following. Consider $\mathscr{T}_p \cup \{x_{k+1}\}$ and the system of inequalities (coupled with (6)).

$$
(7) \qquad
\begin{aligned}
L(A, x) - f(x) &< c_{p-1} & x \in \mathscr{U}_1, \\
L(A, x) - f(x) &> -c_{p-1} & x \in \mathscr{V}_1,
\end{aligned}
$$

where x_{k+1} has been added to both $\mathscr{U}_p$ and $\mathscr{V}_p$. If this system is consistent then one need only consider exchanging x_{k+1} with a point of $\mathscr{T}_p$. This may be done using the procedure of the previous paragraph.

If (7) is inconsistent then one considers the system

$$
(8) \qquad
\begin{aligned}
L(A, x) - f(x) &< c_{p-2} & x \in \mathscr{U}_p, \mathscr{U}_{p-1}, \\
L(A, x) - f(x) &> -c_{p-2} & x \in \mathscr{V}_p, \mathscr{V}_{p-1}.
\end{aligned}
$$

If (8) coupled with (6) is consistent then one need only consider exchanging x_{k+1} with points of $\mathscr{T}_{p-1}$ and $\mathscr{T}_p$. If (8) and (6) are inconsistent then one proceeds with enlarging the system of inequalities.

This process leads to a situation where one has a certain collection of critical point sets $\{\mathscr{T}_j\}$ and it is known that the deviation of the strict approximation on $\{x_{k+1}\} \bigcup_{j=1}^p \mathscr{T}_j$ is larger than the deviation on $\{\mathscr{T}_j\}$, but smaller than the deviation on the remaining subsets from the original collection of $\{\mathscr{T}_j\}$. Thus the following system is inconsistent:

$$
(9) \qquad
\begin{aligned}
L(A, x) &= 0 & x \in \mathscr{T}_j, \quad j = 1, 2, \cdots, p, \\
L(A, x) - f(x) &< c_1 & x \in \bigcup_{j=1}^p \mathscr{U}_j, \\
L(A, x) - f(x) &> -c_1 & x \in \bigcup_{j=1}^p \mathscr{V}_j.
\end{aligned}
$$

To determine $L(A_{k+1}, x)$ one has to find an irreducibly inconsistent subsystem of (9) which yields the largest deviation.

Note that at this point one could modify the algorithm and take any irreducibly inconsistent subsystem of (9). This would not yield the strict approximation $L(A_{k+1}, x)$, but it would yield a set $\mathscr{R}'_{k+1}$ and a corresponding $L(A'_{k+1}, x)$ with a deviation vector $\mathbf{d}'_{k+1}$ larger than $\mathbf{d}_k$. This is sufficient for the convergence of the algorithm to the strict approximation on $\mathscr{B}$.

The problem of computing the strict approximation on $\mathscr{R}_k \cup \{x_{k+1}\}$ has been reduced to that of determining the irreducibly inconsistent subsystems of (9). The details of this problem are not discussed here.

References

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

2. W. B. Carver, *Systems of linear inequalities*, Ann. of Math. **23** (1921–1922), 212–220.

3. E. W. Cheney and A. A. Goldstein, *Newton's method for convex programming and Tchebycheff approximation*, Numer. Math. **1** (1959), 253–268.

4. L. Collatz, *Approximation von Funktionen bei einer und bei mehreren unabhängigen Veränderlichen*, Z. Angew. Math. Mech. **36** (1956), 198–211.

5. Ky Fan, *On systems of linear inequalities, linear inequalities and related systems*, pp. 94–156, Annals of Mathematics Studies No. 38, Princeton Univ. Press, Princeton, N. J., 1956.

6. J. C. Mairhuber, *On Haar's theorem concerning Chebychev approximation problems having unique solutions*, Proc. Amer. Math. Soc. **7** (1956), 609–615.

7. T. S. Motzkin, *Approximation by curves of a unisolvent family*, Bull. Amer. Math. Soc. **55** (1949), 789–798.

8. Ya. L. Remez, *On a method of Tchebycheff type approximation of functions*, Ukrain. Ann. (1935).

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

10. ———, *On the convergence of an algorithm for best Tchebycheff approximations*, J. Soc. Indust. Appl. Math. **7** (1959), 133–142.

11. T. J. Rivlin and H. S. Shapiro, *Some uniqueness problems in approximation theory*, Comm. Pure Appl. Math. **13** (1960), 35–47.

12. I. J. Schoenberg, *On the question of unicity in the theory of best approximations*, Ann. New York Acad. Sci. **86** (1960), 682–691.

13. M. Tonelli, *I polinomi d'approssimozione di Tchebychev*, Ann. Mat., Sec. III, Vol. 15.

14. Ch. de la Vallée Poussin, *Sur les polynomes d'approximation à une variable complexe*, Bull. Acad. Roy. Belg. Cl. Sci. 1911, 199–211.

15. ———, *Leçons sur l'approximation des fonctions d'une variable réele*, Gauthier-Villars, Paris, 1919.

16. C. L. Lawson, *Contributions to the theory of linear least maximum approximation*, Thesis, Univ. of California, Los Angeles, Calif., 1961.

17. 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.

General Motors Research Laboratories,
   Warren, Michigan