JOURNAL OF APPROXIMATION THEORY 23, 299–317 (1978)

An Algorithm for the Construction of Best Approximations Based on Kolmogorov's Criterion

GERHARD OPFER

*Institute for Applied Mathematics, University of Hamburg, Hamburg, West Germany**

Communicated by Lothar Collatz

Received December 9, 1976

A descent algorithm for approximating continuous functions having values in a unitary space by functions in a convex class is given. Applications include complex, multidimensional, monotone approximations and the approximation of kernels of integral equations. Numerous complex Chebyshev polynomials are computed numerically.

Für stetige Funktionen mit Werten in einem unitären Raum wird ein Abstiegsalgorithmus zur Berechnung einer besten Approximation aus einer konvexen Menge angegeben. Die Anwendungen umfassen komplexe, mehrdimensionale, monotone Approximationen, sowie Approximationen von Kernen in Integralgleichungen. Zahlreiche komplexe Tschebyscheff-Polynome wurden mit dem angegebenen Algorithmus berechnet.

1. INTRODUCTION

Recently there has been some interest in computing complex approximations; both the linear and rational cases have been considered. Ellacott and Williams [2, 3] treat both cases and derive an algorithm which is a modified version of Lawson's algorithm [11], an algorithm for computing weighted L_2-approximations. Krabs and Opfer [10] gave a description of a descent algorithm which was applied to conformal mapping problems. Approximations on a disk were considered by Klotz [8], and Gutknecht [7] derived a fairly general descent algorithm which he applied to the construction of digital filters.

We follow Krabs' [9] idea of obtaining a descent algorithm by making the step size optimal in a certain sense. However, we allow the functions which are to be approximated to have values in unitary spaces, so that we can include quite general types of approximation problems. Also, we do not require the domain of definition to be discrete.

* This research was partially carried out while the author was a visitor at Oregon State University, Corvallis, Oreg.

0021-9045/78/0234–0299$02.00/0

Let B be a compact metric space, $(H, \langle , \rangle)$ a unitary space, and $C(B, H)$ the set of functions continuous on B and having values in H. The norm in H is designated by $\| \|_H$. It is well known that $C(B, H)$ becomes a normed space by defining

$$\| x \|_\infty = \max_{t \in B} \| x(t) \|_H \qquad \text{for all} \quad x \in C(B, H). \tag{1.1}$$

Let us assume throughout this paper that $f \in C(B, H)$ is a given function and $V \in C(B, H)$ is a given nonempty set. One is interested in statements about functions $\hat{v} \in V$ with the property

$$\| f - \hat{v} \|_\infty = \inf_{v \in V} \| f - v \|_\infty = \rho_V(f). \tag{1.2}$$

Any such function $\hat{v}$ is called a best approximation of f with respect to V. We assume in general that f is not contained in the closure of V so that $\rho_V(f) > 0$. If $\hat{v} \in V$ is any function then we define

$$E_{\hat{v}} = \{ t \in B : \| f - \hat{v} \|_\infty = \| f(t) - \hat{v}(t) \|_H \} \tag{1.3}$$

and call $E_{\hat{v}}$ set of extreme points of $f - \hat{v}$.

If

$$\min_{t \in E_{\hat{v}}} \operatorname{Re} \langle f(t) - \hat{v}(t), v(t) - \hat{v}(t) \rangle \leqslant 0 \qquad \text{for all} \quad v \in V \tag{1.4}$$

holds for a certain $\hat{v}$, then one knows that $\hat{v}$ is a best approximation of f with respect to V. Conversely, if $\hat{v}$ is a best approximation of f with respect to V then it is known that (1.4) is not true in general. The criterion (1.4) interpreted as a function with the values "true" or "false" is called the Kolmogorov criterion. Although in general the Kolmogorov criterion is not true for best approximations, there are classes of sets V for which a best approximation $\hat{v}$ of f with respect to V is characterized by Kolmogorov's criterion. In particular, this is true if V is a convex subset of $C(B, H)$ which includes the cases where V is a linear subspace of $C(B, H)$ or if V is an affine space in $C(B, H)$. The material listed so for can be found, e.g., in a little book by Singer [19].

2. Constructive Proofs of the Necessity of Kolmogorov's Criterion

We keep the notation of the foregoing section here, but assume that $V \in C(B, H)$ is a convex set. To be able to write the Kolmogorov criterion in a shorter form we use the abbreviation

$$\mu(t, v, \hat{v}) = \operatorname{Re} \langle f(t) - \hat{v}(t), v(t) - \hat{v}(t) \rangle, \qquad t \in B \tag{2.1}$$

and also simply write $\mu(t)$ instead of $\mu(t, v, \hat{v})$ when the meaning of v, $\hat{v}$ is clear from the context.

In all cases (V convex or not) where the Kolmogorov criterion characterizes any best approximation, the existence of a $v \in V$ such that for a particular $\hat{v} \in V$

$$\mu(t, v, \hat{v}) > 0 \qquad \text{for all } t \in E_{\hat{v}} \tag{2.2}$$

is equivalent to $\hat{v}$ not being a best approximation of f with respect to V. Because of the compactness of $E_{\hat{v}}$ condition (2.2) is equivalent to

$$\min_{t \in E_{\hat{v}}} \mu(t) > 0. \tag{2.3}$$

For practical purposes it turned out to be better to work also with a set E slightly larger than set $E_{\hat{v}}$. If for a given $\hat{v}$ there is an element $v \in V$, a set $E \supset E_{\hat{v}}$, and a positive number μ such that

$$\mu(t, v, \hat{v}) \geqslant \mu > 0 \qquad \text{for all } t \in E, \tag{2.4}$$

then $\hat{v}$ is not a best approximation of f with respect to V, since

$$0 < \mu \leqslant \inf_{t \in E} \mu(t) \leqslant \min_{t \in E_{\hat{v}}} \mu(t).$$

THEOREM 1. *Let V be convex. If for a given $\hat{v} \in V$ and a given set $E \supset E_{\hat{v}}$ there is an element $v \in V$ such that* (2.4) *is valid, then there is a largest number $\lambda_1 \in]0, 1]$ with the property that*

$$\| f - (\hat{v} + \lambda(v - \hat{v}))\|_\infty^2 \leqslant \| f - \hat{v} \|_\infty^2 - \lambda\mu \qquad \textit{for all} \quad \lambda \in [0, \lambda_1]. \tag{2.5}$$

Proof. Because of the norm definition (1.1) the inequality in (2.5) is equivalent to the set of inequalities

$$\| f(t) - \hat{v}(t) - \lambda(v(t) - (\hat{v}(t))\|_H^2 - \| f - \hat{v} \|_\infty^2 + \lambda\mu \leqslant 0 \qquad \text{for all} \quad t \in B. \tag{2.6}$$

For every $t \in B$ the left-hand side of (2.6) is a quadratic polynomial in λ which reads

$$p(\lambda) = a(t)\,\lambda^2 - 2b(t)\,\lambda - c(t), \tag{2.7}$$

where

$$a(t) = \| v(t) - \hat{v}(t)\|_H^2, \tag{2.8}$$

$$b(t) = \mu(t) - \tfrac{1}{2}\mu, \tag{2.9}$$

$$c(t) = \| f - \hat{v} \|_\infty^2 - \| f(t) - \hat{v}(t)\|_H^2. \tag{2.10}$$

Because $a(t) \geqslant 0$ and $c(t) \geqslant 0$ for all $t \in B$ the polynomial $p(\lambda)$ has only real roots, the greatest of which is designated by $\lambda(t)$. Explicitly from (2.6) one obtains

$$\lambda(t) = \frac{c(t)}{\mu} \qquad \text{in case} \quad a(t) = 0, \tag{2.11a}$$

$$= \frac{1}{a(t)} \{b(t) + (b^2(t) + a(t)\, c(t))^{1/2}\} \qquad \text{else.} \tag{2.11b}$$

In case $a(t) = 0$, it follows that $\mu(t) = 0$, so that $b(t)$ reduces to $-\frac{1}{2}\mu$, which gives (2.11a). Because of the form of $p(\lambda)$ it is clear that for $\lambda \geqslant 0$ the conditions $p(\lambda) \leqslant 0$ and $\lambda \leqslant \lambda(t)$ are equivalent.

The remainder of the proof consists of two parts: (I) We show that $\lambda(t) > 0$ for all $t \in B$. (II) We show that $\lambda(t)$ is continuous on the compact set B which implies $\min_{t \in B} \lambda(t) > 0$.

(I) In case $a(t) = 0$ we must have $t \notin E$, because we have already seen that in this case $\mu(t) = 0$, which would otherwise contradict (2.3). In addition, $t \notin E_{\hat{v}}$ follows, which implies that $c(t) > 0$ by (1.3) and (2.9). Now assume that $a(t) \neq 0$. If in this case $b(t) > 0$ then $\lambda(t) > 0$ by inspection from (2.11). Assume therefore that $b(t) \leqslant 0$. In this case $t \notin E$ because otherwise we obtain $0 < \frac{1}{2}\mu = \mu - \frac{1}{2}\mu \leqslant \mu(t) - \frac{1}{2}\mu = b(t) \leqslant 0$ from (2.4). If $t \notin E$ then $t \notin E_{\hat{v}}$ also, which implies that $c(t) > 0$, as we have already seen above. Thus we have $a(t) > 0$ and $c(t) > 0$, which implies that $\lambda(t) > 0$.

(II) To show the continuity of $\lambda(t)$ we first remark that this is clear from (2.11) when $a(t) = 0$ for all $t \in B$ or if $a(t) \neq 0$ for all t. In the remaining cases some straightforward analysis gives the required result. Combining these two parts we obtain

$$\hat{\lambda}_1 = \min_{t \in B} \lambda(t) > 0 \tag{2.12a}$$

and we have $p(\lambda) \leqslant 0$ for all $t \in B$ when $\lambda \leqslant \hat{\lambda}_1$.

Because we do not know whether $\hat{\lambda}_1 \leqslant 1$ we define

$$\lambda_1 = \min(1, \hat{\lambda}_1) \tag{2.12b}$$

and hence (2.5) is true. If $\lambda_1 < 1$ then from (2.12a) it is clear that λ_1 is the largest number for which (2.5) is true. ∎

If in Theorem 1 in inequality (2.5) the squares were omitted, then that theorem would not necessarily be true. But we have

THEOREM 2. *Let V be convex. If for a given $\hat{v} \in V$ and a set $E \supset E_{\hat{v}}$ there is an element $v \in V$ such that* (2.4) *is valid, then for any real M with* $0 < M <$

$\mu/\|f - \hat{v}\|_\infty$ *there is a largest number* $\lambda_2 \in\,]0, 1]$ *with the property that*

$$\|f - (\hat{v} + \lambda(v - \hat{v}))\|_\infty \leqslant \|f - \hat{v}\|_\infty - \lambda M \text{ for all } \lambda \in [0, \lambda_2]. \tag{2.13}$$

Proof. From Theorem 1, inequality (2.5), we conclude that

$$\|f - (\hat{v} + \lambda(v - \hat{v}))\|_\infty \leqslant \|f - v\|_\infty - \frac{\lambda\mu}{2\,\|f - \hat{v}\|_\infty} \qquad \text{for all} \quad \lambda \leqslant \lambda_1 . \tag{2.14}$$

The existence of λ_2 with the required properties is thus shown. To compute an explicit expression for λ_2 we have to go through an argument similar to that used in the proof of Theorem 1. We omit all details and simply state the results: Let

$$A(t) = \| v(t) - \hat{v}(t)\|_H^2 - M^2, \tag{2.15}$$

$$B(t) = \mu(t) - M\,\|f - \hat{v}\|_\infty , \tag{2.16}$$

$$C(t) = \|f - \hat{v}\|_\infty^2 - \|f(t) - \hat{v}(t)\|_H^2 , \tag{2.17}$$

$$D(t) = B^2(t) + A(t)\,C(t), \tag{2.18}$$

$$\Lambda(t) = -\frac{C(t)}{2B(t)} \qquad \text{for} \quad A(t) = 0, \tag{2.19a}$$

$$= \frac{B(t) + (D(t))^{1/2}}{A(t)} \qquad \text{for} \quad A(t) \neq 0, \quad D(t) \geqslant 0. \tag{2.19b}$$

Then

$$\hat{\lambda}_2 = \min_{D(t)\geqslant 0} \Lambda(t)$$

and

$$\lambda_2 = \min(1, \hat{\lambda}_2 , \|f - \hat{v}\|_\infty/M). \quad \blacksquare \tag{2.20}$$

Remark. If $\hat{v}$ is a best approximation of f, then any condition following from that fact is a necessary condition for $\hat{v}$ to be a best approximation. The two preceding theorems contain such conditions in contrapositive form. Furthermore these theorems, including the proofs, contain explicit information on how to improve an approximation $\hat{v}$ which is not best. Thus, the title of this section is justified.

3. Derivation of Descent Algorithms

By using either of the two preceding theorems we can define a sequence of elements in V which are decreasing in norm and which is expected to have best approximations as accumulation points. We keep the requirement that V be convex.

For any $\delta \geqslant 0$ we define

$$E_{\hat{v}}(\delta) = \{t \in B : C(t) \leqslant \delta\}, \tag{3.1}$$

where $C(t)$ is defined in (2.18) (and in (2.10)). This $E_{\hat{v}}(\delta)$ is a compact set in B and will play the role of E of the preceding section.

We assume that $\hat{v}^{(j)} \in V$ and $\delta^{(j)} > 0$ are given. In the following we define $\hat{v}^{(j+1)}$ and $\delta^{(j+1)}$, $j = 1, 2, \ldots$. We set

$$E^{(j)} = E_{\hat{v}^{(j)}}(\delta^{(j)}). \tag{3.2}$$

Case I

There is a $v^{(j)} \in V$ and a $\mu^{(j)}$ such that (2.4) is valid; i.e., we have

$$\mu(t, v^{(j)}, \hat{v}^{(j)}) \geqslant \mu^{(j)} > 0 \qquad \text{for all } t \in E^{(j)}.$$

Then we define

$$\hat{v}^{(j+1)} = \hat{v}^{(j)} + \lambda^{(j)}(v^{(j)} - \hat{v}^{(j)}), \tag{3.3}$$

where $\lambda^{(j)}$ is computed either by (2.12b) or by (2.20). The decision whether to use (2.12b) or (2.20) has to be made once and for all in the beginning of the computation. Furthermore

$$\delta^{(j+1)} = \delta^{(j)}/2 \qquad \text{if } \mu^{(j)} \leqslant \delta^{(j)}, \tag{3.4a}$$
$$= \delta^{(j)} \qquad \text{if } \mu^{(j)} > \delta^{(j)}. \tag{3.4b}$$

Case II

There is no $v \in V$ such that (2.4) is valid.

Subcase IIa. $E^{(j)} = E_{\hat{v}^{(j)}}$. In this case $\hat{v}^{(j)}$ is a best approximation of f with respect to V and the sequence $\{\hat{v}^{(j)}\}$ terminates. (Formally we could define $\hat{v}^{(j+1)} = \hat{v}^{(j)}$, $\delta^{(j+1)} = \delta^{(j)}$.)

Subcase IIb. $E_{\hat{v}^{(j)}} \subsetneqq E^{(j)}$. Here we define

$$\hat{v}^{(j+1)} = \hat{v}^{(j)}; \qquad \delta^{(j+1)} = \delta^{(j)}/2. \tag{3.5}$$

In (3.5) other choices of $\delta^{(j+1)}$ are possible. For example, one could try to choose $\delta^{(j+1)} \in \,]0, \delta^{(j)}/2]$ as large as possible such that $E_{\hat{v}^{(j+1)}} = E^{(j+1)}$ becomes true. In the discrete case (i.e., B consisting of finitely many points only) such a choice of $\delta^{(j+1)}$ is always possible.

The algorithms are herewith described, but we have not yet said how to solve problem (2.4). We remark in this connection that the Kolmogorov criterion (1.4) is equivalent to

$$\min_{t \in E_{\hat{v}}} \operatorname{Re}\langle f(t) - \hat{v}(t), d_v(v(t) - \hat{v}(t))\rangle \leqslant 0 \qquad \text{for all} \quad v \in V, \tag{1.4'}$$

where d_v is any positive number which may depend on v. From inequalities (2.5) and (2.13) it is evident that μ should be chosen as large as possible under the condition that the directions $v - \hat{v}$ are normalized in some sense, say

$$\| v - \hat{v} \| \leqslant 1, \qquad v \in V, \tag{3.6}$$

where $\| \ \|$ is any suitable norm defined in a linear space F containing V. Then (2.4) can be regarded as an optimization problem for the unknowns $(\mu, v) \in \mathbb{R} \times V$ consisting of (3.6) and

$$\mu - \mu(t, v, \hat{v}) \leqslant 0 \qquad \text{for } t \in E, \tag{3.7}$$

$$\mu = \max! \tag{3.8}$$

The optimization problem (3.6) through (3.8) always has a feasible point $(\mu, v) = (0, \hat{v})$ which implies that the maximal μ is never negative. Furthermore, if E is any nonempty subset of B and (3.6) defines a compact subset of V, then the optimization problem has a solution.

That the normalization (3.6) is always possible follows from (1.4′) by letting

$$\begin{aligned} d_v &= 1 && \text{in case} \quad \| v - \hat{v} \| \leqslant 1, \\ &= \frac{1}{\| v - \hat{v} \|} && \text{in case} \quad \| v - \hat{v} \| > 1. \end{aligned} \tag{3.9}$$

In the case $\| v - \hat{v} \| > 1$, the effect of the multiplication of $(v(t) - \hat{v}(t))$ by d_v can be regarded as a replacement of v by $(1 - d_v)\,\hat{v} + d_v v \in V$.

THEOREM 3. *If V is a convex subset of a finite-dimensional subspace F of $C(B, H)$ and if E is any nonempty subset of B, then the optimization problem* (3.5) *through* (3.8) *has a solution.*

Proof. Under the stated assumptions (3.6) defines a compact subset of V. The rest follows from the preceding remarks. ∎

Now it is clear how the given algorithms must be completed. If $\hat{v}^{(j)}$ and $\delta^{(j)} > 0$ are given, we solve the optimization problem for $v^{(j)}$ and $\mu^{(j)}$ and continue as described.

We will postpone remarks on the practical execution of the given algorithms until after a discussion of their convergence behavior.

4. CONVERGENCE BEHAVIOR

We assume here that V is a convex and closed subset of a finite-dimensional space F in $C(B, H)$. This guarantees the existence of at least one best approximation to any $f \in C(B, H)$. But because such a best approximation may not be unique, we cannot expect the given algorithms to converge. We can expect, however, that any accumulation point of the generated sequence $\{\hat{v}^{(j)}\}$, $j = 1, 2, \ldots$, is a best approximation of f.

The norm $\| \ \|$ used in (3.6) is from now on required to have the property that there is a constant $w > 0$ such that

$$\| x \| \leqslant 1 \Rightarrow \| x \|_\infty \leqslant w \qquad \text{for all } x \in F,$$

which means that $\| \ \|$ is a stronger (or equivalent) norm than $\| \ \|_\infty$. From Theorems 1 and 2 it follows that

$$\| f - \hat{v}^{(j+1)} \|_\infty \leqslant \| f - \hat{v}^{(j)} \|_\infty, \; j = 1, 2, \ldots, \tag{4.1}$$

which in turn implies the existence of

$$\lim_{j \to \infty} \| f - \hat{v}^{(j)} \|_\infty = \rho \tag{4.2}$$

and

$$\lim_{j \to \infty} \lambda^{(j)} \mu^{(j)} = 0. \tag{4.3}$$

Throughout the remainder of this section, we assume that the sequence $\{\hat{v}^{(j)}\}$ does not terminate with Subcase IIa of the preceding section.

THEOREM 4. $\lim_{j \to \infty} \delta^{(j)} = 0.$ (4.4)

Proof. The construction of the sequence $\{\delta^{(j)}\}$ implies the existence of

$$\lim_{j \to \infty} \delta^{(j)} = \delta \geqslant 0.$$

Let us assume that $\delta > 0$. This can happen if and only if there is an integer j_0 such that formula (3.4b) has to be used for all $j \geqslant j_0$. We have therefore

$$\mu^{(j)} > \delta^{(j)} \geqslant \delta > 0 \qquad \text{for } j \geqslant j_0. \tag{4.5}$$

We shall show now that $\delta > 0$ implies the existence of a positive lower bound for $\{\lambda^{(j)}\}$ which together with (4.5) would contradict (4.3).

Let us first look at the case which is described in Theorem 1. If $t \in E^{(j)}$, $j \geqslant j_0$, then from (2.11b) it follows that

$$\lambda(t) \geqslant \frac{\mu^{(j)}}{\| v^{(j)} - \hat{v}^{(j)} \|_\infty^2} > \frac{\delta}{\| v^{(j)} - \hat{v}^{(j)} \|_\infty^2} = \lambda_1^{(j)},$$

where $\lambda^{(j)}$ is an abbreviation for the expression on the left-hand side of the "=" sign.

If $t \notin E^{(j)}$ and $j \geq j_0$ then by definition $C(t) > \delta^{(j)}$. From (2.7) we deduce in this case that

$$p(\lambda) < \| v^{(j)} - \hat{v}^{(j)} \|_\infty^2 \lambda^2 + 3 \| f - \hat{v}^{(j)} \|_\infty \| v^{(j)} - \hat{v}^{(j)} \|_\infty - \delta = \tilde{p}(\lambda),$$

where $\tilde{p}(\lambda)$ is defined by this equation.

Hence $p(\lambda) < 0$ if $\tilde{p}(\lambda) \leq 0$, which occurs if we choose

$$\lambda = \lambda_2^{(j)} = \frac{1}{2 \| v^{(j)} - \hat{v}^{(j)} \|_\infty} \{-3 \| f - \hat{v}^{(1)} \|_\infty + (9 \| f - \hat{v}^{(1)} \|_\infty^2 + 4\delta)^{1/2}\}.$$

Because $\| v^{(j)} - \hat{v}^{(j)} \|_\infty \leq w$ for all j, we see that $\lambda_1^{(j)}$ and $\lambda_2^{(j)}$ have positive lower bounds which are independent of j, which then also follows for $\lambda^{(j)}$ because $\lambda^{(j)} \geq \min(\lambda_1^{(j)}, \lambda_2^{(j)})$. We will not treat the case covered in Theorem 2 in detail. ∎

THEOREM 5. *The sequence $\{\hat{v}^{(j)}\}$ has an accumulation point $\hat{v} \in V$. Any accumulation point of $\{\hat{v}^{(j)}\}$ is a best approximation of f with respect to V. Furthermore we have $\rho = \rho_V(f)$, where these quantities are defined in* (4.2) *and in* (1.2), *respectively.*

Proof. Because of (4.1) we have

$$\hat{v}^{(j)} \in \hat{V} = \{v \in V : \| f - v \|_\infty \leq \| f - \hat{v}^{(1)} \|_\infty\}, \qquad j = 1, 2, \ldots,$$

and $\hat{V}$ is a nonempty compact subset of V. This implies that all accumulation points of $\{v^{(j)}\}$ are contained in V and at least one accumulation point exists. Without loss of generality we may assume the existence of $\lim_{j\to\infty} \hat{v}^{(j)} = \hat{v}$. We observe that $\lim_{j\to\infty} \delta^{(j)} = 0$ (Theorem 4) implies the existence of a subsequence of $\{\mu^{(j)}\}$ which converges to zero. Again, without loss of generality we may assume that $\lim_{j\to\infty} \mu^{(j)} = 0$.

Using the definition (2.1) we see that for any $v \in V$ we have

$$\mu(t, \hat{v}, v) \leq \mu(t, \hat{v}^{(j)}, v) + \eta^{(j)},$$

where $\lim_{j\to\infty} \eta^{(j)} = 0$. Therefore

$$\begin{aligned} \min_{t\in E_{\hat{v}}} \mu(t, \hat{v}, v) &\leq \min_{t\in E_{\hat{v}}} \mu(t, \hat{v}^{(j)}, v) + \eta^{(j)} \\ &= \min_{t\in E^{(j)}} \mu(t, \hat{v}^{(j)}, v) \\ &\quad + (\min_{t\in E_{\hat{v}}} \mu(t, \hat{v}^{(j)}, v) - \min_{t\in E^{(j)}} \mu(t, \hat{v}^{(j)}, v)) + \eta^{(j)} \\ &\leq \mu^{(j)} + \eta^{(j)} + (\min_{t\in E_{\hat{v}}} \mu(t, \hat{v}^{(j)}, v) - \min_{t\in E^{(j)}} \mu(t, \hat{v}^{(j)}, v)). \end{aligned}$$

Because the last three terms all converge to zero for $j \to \infty$ we have $\min_{t \in E_{\hat{v}}} \mu(t, \hat{v}, v) \leqslant 0$, which implies that $\hat{v}$ is a best approximation of f with respect to V. Finally, $\rho = \lim_{j \to \infty} \| f - \hat{v}^{(j)} \|_\infty = \| f - \hat{v} \|_\infty = \rho_V(f)$. ∎

The rate of convergence is still an open question.

5. Applications

We shall describe some possible applications.

5.1. *Integral Equations*

Assume that $(X, \| \ \|)$ is a normed linear space and $K, \tilde{K} : X \to X$ are continuous linear mappings with $\| K \| < 1$ and $\| \tilde{K} \| < 1$. Assume further that we are interested in solving

$$x - Kx = g \in X.$$

Because this problem may be too difficult, suppose we solve instead

$$\tilde{x} - \tilde{K}\tilde{x} = g \in X.$$

Then

$$\| x - \tilde{x} \| \leqslant \frac{\| K - \tilde{K} \| \, \| g \|}{(1 - \| K \|)(1 - \| \tilde{K} \|)},$$

and from this inequality, it is clear that $\tilde{K}$ should be chosen such that $\| K - \tilde{K} \|$ is as small as possible.

We assume now that $(X, \| \ \|) = (C(I), \| \ \|_\infty)$, with $I = [-1, 1]$ and

$$Kx(t) = \int_{-1}^{1} k(t, \tau) \, x(\tau) \, d\tau,$$

where for simplicity k is a continuous kernel. Furthermore

$$\tilde{K}\tilde{x}(t) = \int_{-1}^{1} \tilde{k}(t, \tau) \, \tilde{x}(\tau) \, d\tau,$$

where $\tilde{k}$ is a continuous and degenerate kernel, i.e.,

$$\tilde{k}(t, \tau) = \sum_{j=1}^{m} r_j(t) \, s_j(\tau).$$

The original approximation problem requires minimization of

$$\| K - \tilde{K} \| = \max_{t \in I} \int_{-1}^{1} | k(t, \tau) - \tilde{k}(t, \tau) | \, d\tau. \tag{5.1}$$

This may be regarded as a mixed L_1-approximation problem, which is not contained in our scheme [1, p. 33].

Instead we could try to minimize

$$\| k(t, \tau) - \tilde{k}(t, \tau)\|_\infty = \max_{(t,\tau)\in I\times I} | k(t, \tau) - \tilde{k}(t, \tau)|, \tag{5.2}$$

which would imply

$$\| K - \tilde{K} \| \leqslant 2 \| k - \tilde{k} \|_\infty$$

and could be regarded as a two-dimensional T-approximation problem [1, p. 28] which now can be formulated in terms of (1.1) and (1.2). From Hölder's inequality we can deduce from (5.1) that

$$\| K - \tilde{K} \| \leqslant (2)^{1/2} \max_{t\in I} \left\{\int_{-1}^{1} (k(t, \tau) - \tilde{k}(t, \tau))^2 \, d\tau\right\}^{1/2}, \tag{5.3}$$

which results in a mixed L_2–T-approximation problem. This problem also fits into our scheme. A numerical example is given by Schultz [18, p. 50–54].

5.2. *Linear, Complex Approximations*

Let R be a bounded region of the complex plane. Then $B = \bar{R}$ where $\bar{R}$ is the closure of R and $H = \mathbb{C}$ with $\langle x, y\rangle = x \cdot \bar{y}$. Besides the approximation of certain functions on B by elements of certain linear subspaces V of $C(B, \mathbb{C})$ there are some special problems of interest: the conformal mapping problem which results in solving

$$\| t - (a_2 t^2 + a_3 t^3 + \cdots + a_n t^n)\|_\infty = \min \tag{5.4}$$

[16, 10] and the computation of complex Chebyshev polynomials (for short T-polynomials) which are defined by

$$\| t^n - (a_{n-1} t^{n-1} + a_{n-2} t^{n-2} + \cdots + a_0)\|_\infty = \min. \tag{5.5}$$

For this case we shall give some detailed numerical results in the next section. The T-polynomials play an important rôle, e.g., in solving differential equations by series methods [14]. T-polynomials on an interval of the real line are well known (e.g., [17]) in contrast to the complex case, where there are explicit results for special regions only such as ellipses and lemniscates [4]. There are also known results in the two-dimensional case [5].

5.3 *Nonlinear, Convex Approximations*

We mention here the case of the so-called monotone approximations which have been investigated by various authors (e.g., Lorentz and Zeller [12]), although numerical results seem to be scarce. Here $B = [a, b]$ with $a < b$, $H = \mathbb{R}$, $\langle x, y\rangle = x \cdot y$, and V consists of all polynomials with degree at most n such that certain prescribed derivatives have only one sign in B. Here V is even a convex cone but not a linear space.

5.4. *Nonconvex Problems*

Although these problems are not included in our derivations we are able to solve some of these by linearization. If $A \subset \mathbb{R}^n$ or $A \subset \mathbb{C}^n$ is a parameter set and $V = F(A)$, where $F : A \to C(B, H)$ has certain differentiability properties, we could work with the so-called local Kolmogorov criterion [1]. The computation of the step length as given in Section 2 has to be changed to some sort of heuristic method which may also be applied to the convex case. Usually one takes a step length λ such that $\| f - (\hat{v} + \lambda(v - \hat{v}))\|_\infty$ attains a local minimum under the condition $\hat{v} + \lambda(v - \hat{v}) \in V$, where one starts with a certain λ_0 and tries $\lambda_0/2$, $\lambda_0/4$, $\lambda_0/8$,... or $2\lambda_0$, $4\lambda_0$, $8\lambda_0$,..., etc.

6. Numerical Results

We computed complex T-polynomials of various degrees and for the following sets $B \subset \mathbb{C}$:

(a) Confocal ellipses E_r with foci ± 1 and semiaxes

$a = \frac{1}{2}(1/r + r)$, $b = \frac{1}{2}(1/r - r)$ and certain r, $0 < r \leqslant 1$, including $E_1 = [-1, 1]$,

(b) a square $Q = \{(x, y) : |x| \leqslant 1, |y| \leqslant 1\}$,

(c) a rectangle $R = \{(x, y) : |x| \leqslant 2, |y| \leqslant 1\}$,

(d) circular sectors $S_\alpha = \{t : |t| \leqslant 1, |\arg t| \leqslant \alpha\}$

for $\alpha = 0°(5°)\,95°$ which include the interval $S_0 = [0, 1]$.

Here we have $f(t) = t^n$ and $V = P_{n-1}$ the set of all polynomials with degree $n - 1$ or less and complex coefficients. Because of $V = \{v - \hat{v} : v \in V\}$ for all $\hat{v} \in V$ we can replace $v(t) - \hat{v}(t)$ in (1.4) by $v(t)$ such that condition (3.6) reads $\| v \| \leqslant 1$. If we choose $\| v \| = \max_{i=0,1,\ldots,n-1} (|\operatorname{Re} a_i|, |\operatorname{Im} a_i|)$ then (3.6) is equivalent to the linear conditions $|\operatorname{Re} a_i| \leqslant 1$, $|\operatorname{Im} a_i| \leqslant 1$, $i = 0, 1,\ldots, n - 1$ where $v(t) = \sum_{i=0}^{n-1} a_i t^i$. In all the cases treated B is symmetric to the x-axis, which implies that the best approximation of t^n by elements of P_{n-1} has only real coefficients [14, Theorem 27]. The computation of the step length was carried out according to (2.12), (2.11).

In order to facilitate solving the optimization problem (3.6) by (3.8) we always discretized the set B where we also used the fact that B could be replaced by its boundary ∂B. Thus, actually we replaced B by finitely many points of ∂B. Doing so means that the optimization problem mentioned is now an ordinary linear programming problem which can be solved by standard techniques.

That these results are also obtainable without discretizing B was shown by Vollstedt [21]. Here again (compare the corresponding remarks in Krabs and Opfer [10]) it turned out to be advisable to start with a very coarse discretization of ∂B to compute a starting vector for the same case but a finer discretization. For most of the cases we used 10, 100, and 500 points on ∂B.

Clearly everywhere in the algorithms where a comparison with zero is made, some tolerance has to be admitted. As a consequence, the quantity δ occurring in (3.1) should not be allowed to become too small. That means, practically, that the algorithms should be restarted with a new δ from time to time where the size of the new (and also beginning) δ can be computed in a reasonable manner from $|f(t) - \hat{v}(t)|$ such that $E_{\hat{v}}(\delta)$ contains all relative maxima of $|f(t) - \hat{v}(t)|$ which are close to $\|f - \hat{v}\|_\infty$. Usually it turned out here that the relative maxima of $|f(t) - \hat{v}(t)|$ were very pronounced and all very close to $\|f - \hat{v}\|$. In other cases the situation may be different, however. (Compare the remarks for the ellipse case (a).)

(a) *Confocal ellipses* E_r. In this case we know (e.g., [20, p. 360]) that for fixed n all T-polynomials are alike (independent of r) and are usually called T-polynomials of the first kind. We computed these T-polynomials up to degree $n = 15$ for various r including the degenerate case $r = 1$ where $B = [-1, 1]$. We compared the computed coefficients with the exact coefficients as given in [13, p. 458] (adjusted to the normalization used here). In most of the cases there were no observable errors at all. For $r = 2^{1/2} - 1$, for example, the error never exceeded 10^{-8}. Therefore a more detailed listing of the results seems unnecessary.

However, one remark seems necessary. If n becomes large, one observes that $\|f - \hat{v}\|_\infty^2 - |f(t) - \hat{v}(t)|^2$ becomes simultaneously almost constant and small. That means that for large n the set $E_{\hat{v}}(\delta)$ will contain almost all points of B which is not very desirable. In the case $r = 2^{1/2} - 1$ and $n = 15$ we observed $\|f - \hat{v}\|_\infty^2 - |f(t) - \hat{v}(t)|^2 < 10^{-7}$ for all $t \in B$ and have $\|f - \hat{v}\|_\infty = 16.8$. Nevertheless the computed coefficients of T_{15} have errors which are less than 10^{-9}.

(b) *Square* Q. Because of the symmetry of Q we have

$$T_n(t) = t^n + a_{j-1}^{(n)} t^{n-4} + a_{j-2}^{(n)} t^{n-8} + a_0^{(n)} t^{n-4 \cdot j}, \qquad j = [n/4]. \tag{6.1}$$

For $n = 4, 5, \ldots, 21$ the extreme points of $|T_n(t)|$ on Q are given in Fig. 1 and Table 1, while the coefficients $a_i^{(n)}$ and the norm $\|T_n\|_\infty$ are given in Table 2 for $4 \leqslant n \leqslant 16$.

TABLE I

Numerical Values of the Location of the Extreme Points on the Square Q

n	$y_n^{(1)}$	$y_n^{(2)}$	$y_n^{(3)}$
7	0.40		
8	0.42		
9	0.55		
10	0.63		
11	0.68		
12	0.26	0.72	
13	0.29	0.75	
14	0.31	0.78	
15	0.34	0.80	
16	0.46	0.82	
17	0.48	0.84	
18	0.51	0.85	
19	0.53	0.86	
20	0.20	0.59	0.87
21	0.23	0.61	0.88

TABLE II

Norm and Coefficients of T-Polynomials on the Square Q

n	$\|T_n\|_\infty$	$a_0^{(n)}$	$a_1^{(n)}$	$a_2^{(n)}$	$a_3^{(n)}$
4	2.50	1.50			
5	2.93	1.93			
6	3.33	2.33			
7	3.80	2.66			
8	4.38	0.38	3.00		
9	5.08	0.79	3.30		
10	5.93	1.34	3.59		
11	6.93	2.03	3.90		
12	8.09	0.00	2.85	4.21	
13	9.48	0.17	3.76	4.51	
14	11.12	0.49	4.78	4.82	
15	13.05	1.00	5.91	5.13	
16	15.31	0.04	1.71	7.13	5.44

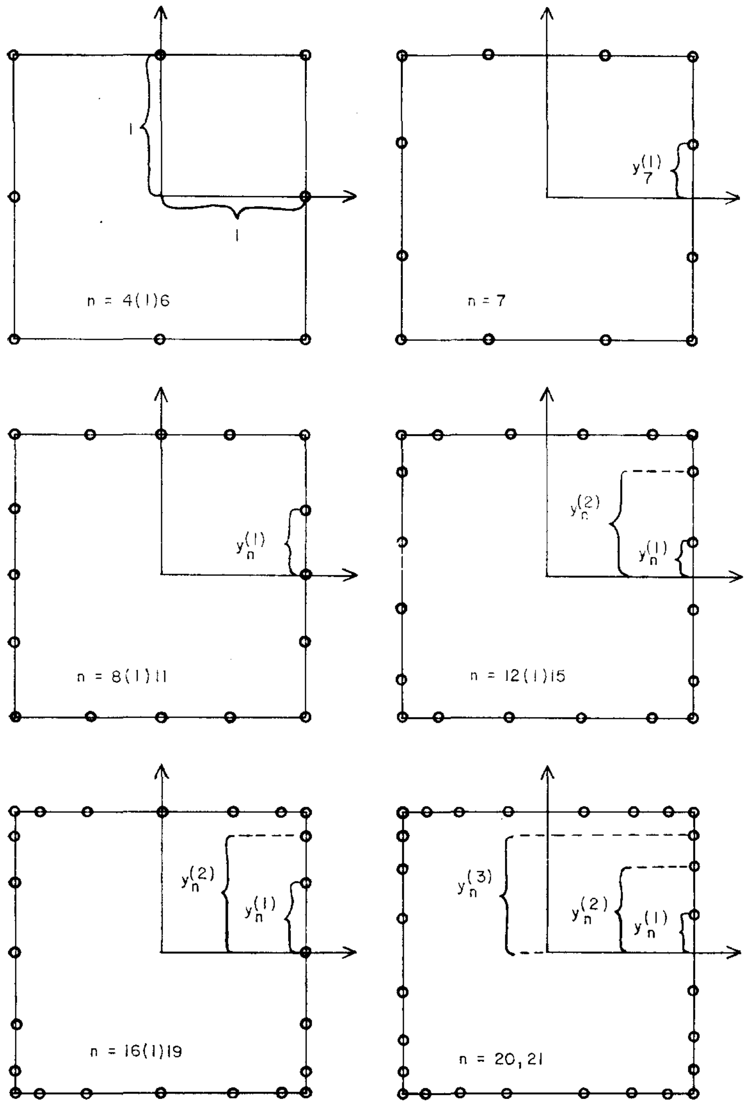

FIG. 1. Extreme point distributions on the square Q for $n = 4(1)21$.

(c) *Rectangle R.* Because of the symmetry of R we have

$$T_n(t) = t^n - a_{j-1}^{(n)} t^{n-2} + a_{j-2}^{(n)} t^{n-4} - \cdots + (-1)^j a_0^{(n)} t^{n-2\cdot j}, \qquad j = [n/2]. \tag{6.2}$$

For $n = 2, 3,\ldots, 10$ the extreme points of $T_n(t)$ are marked in Fig. 2; the coefficients $a_i^{(n)}$ and the norms $\| T_n \|_\infty$ are given in Table 3.

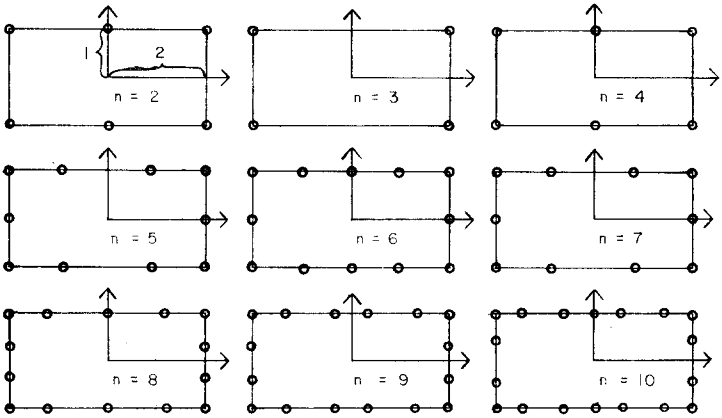

FIG. 2. Extreme points on the rectangle R for $n = 2(1)10$.

TABLE III

Norms and Coefficients of T-Polynomials on the Rectangle R

n	$\|T_n\|_\infty$	$a_0^{(n)}$	$a_1^{(n)}$	$a_2^{(n)}$	$a_3^{(n)}$	$a_4^{(n)}$
2	4.00	3.00				
3	8.94	3.00				
4	13.25	8.60	3.66			
5	20.37	13.71	4.88			
6	35.06	8.38	19.52	6.17		
7	60.30	24.78	25.21	6.87		
8	102.1	15.16	46.34	31.88	7.74	
9	175.8	51.0	75.2	39.7	8.7	
10	303.5	19.1	113.9	111.7	48.3	9.7

(d) *Circular sectors* S_α. Here we have

$$T_n(t) = t^n - a_{n+1}^{(n)} t^{n-1} + a_{n-2}^{(n)} t^{n-2} - \cdots + (-1)^n a_0^{(n)}. \tag{6.3}$$

For $n = 1, 2, 3, 4$ the extreme points of $|T_n(t)|$ for certain α are shown qualitatively in Fig. 3; the coefficients $a_i^{(n)}$, $i = 0, 1, \ldots, n - 1$ and the norms $\|T_n\|_\infty$ are given in Table 4. For $\alpha = 0$ we obtain the usual T-polynomials on $[0, 1]$ the exact coefficients of which can be found for $1 \leqslant n \leqslant 20$ in [13, p. 462]. In many of the cases mentioned the computed results give hints for exact T-polynomials and more general statements on T-polynomials which in a separate step can be proved to be true [6].

All computations were carried out on the AEG-TELEFUNKEN TR 440 of the University of Hamburg and the CONTROL DATA CYBER 70 MODEL 73 of Oregon State University.

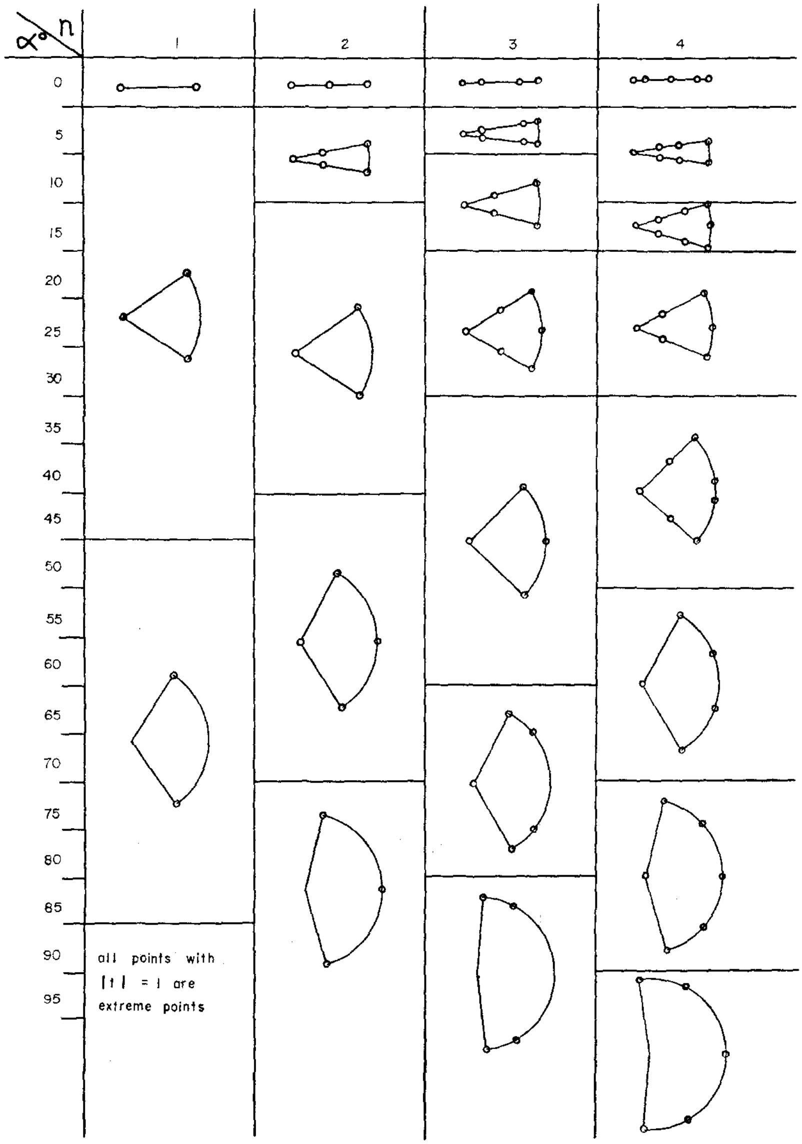

FIG. 3. Extreme point distributions of T-polynomials on sectors S_α.

TABLE IV

Norms and Coefficients of T-Polynomials on Sectors S_α

	$n = 1$	$n = 2$		$n = 3$			$n = 4$			
	$\|T_1\|_\infty$	$\|T_2\|_\infty$		$\|T_3\|_\infty$			$\|T_4\|_\infty$			
α	$a_0^{(1)}$	$a_0^{(2)}$	$a_1^{(2)}$	$a_0^{(3)}$	$a_1^{(3)}$	$a_2^{(3)}$	$a_0^{(4)}$	$a_1^{(4)}$	$a_2^{(4)}$	$a_3^{(4)}$
0°	0.5000	0.1250		0.0313			0.0078			
	0.5000	0.1250	1.0000	0.0313	0.5625	1.5000	0.0078	0.2500	1.2500	2.0000
10°	0.5077	0.1537		0.0561			0.0160			
	0.5077	0.1537	1.0913	0.0561	0.7886	1.7131	0.0160	0.4069	1.6663	2.2616
20°	0.5321	0.2549		0.0813			0.0299			
	0.5321	0.2549	1.1792	0.0813	0.9194	1.7569	0.0299	0.5880	2.0219	2.4339
30°	0.5774	0.3333		0.1187			0.0512			
	0.5774	0.3333	1.1547	0.1187	1.0792	1.8418	0.0512	0.7526	2.1878	2.4353
40°	0.6527	0.3913		0.1848			0.0804			
	0.6527	0.3913	1.0658	0.1848	1.1848	1.8152	0.0804	0.8931	2.2812	2.3975
50°	0.7660	0.4375		0.2535			0.1263			
	0.6428	0.4375	1.0000	0.2535	1.1861	1.6790	0.1263	1.0648	2.3712	2.3508
60°	0.8660	0.5000		0.3170			0.1891			
	0.5000	0.5000	1.0000	0.3170	1.1340	1.5000	0.1891	1.1303	2.2395	2.1596
70°	0.9397	0.5938		0.3830			0.2554			
	0.3420	0.5938	1.0000	0.3830	1.1387	1.3900	0.2554	1.1063	1.9856	1.8874
80°	0.9848	0.7169		0.4738			0.3283			
	0.1736	0.5321	0.8152	0.4738	1.1417	1.2904	0.3283	1.1413	1.8351	1.6938
90°	1.0000	0.8284		0.6006			0.4286			
	0.0000	0.4142	0.5858	0.4864	1.0000	1.0870	0.4286	1.1714	1.6857	1.5143

REFERENCES

1. L. COLLATZ AND W. KRABS, "Approximationstheorie," Teubner, Stuttgart, 1973.
2. S. ELLACOTT AND J. WILLIAMS, Linear Chebychev approximation in the complex plane using Lawson's algorithm, *Math. Comp.* **30** (1976), 35–44.
3. S. ELLACOTT AND J. WILLIAMS, Rational Chebychev approximation in the complex plane, *SIAM J. Numer. Anal.* **13** (1976), 310–323.
4. G. FABER, Über Tschebyscheffsche Polynome, *J. Reine Angew. Math.* **150** (1920), 79–106.
5. W. B. GEARHART, Some Chebychev approximations by polynomials in two variables, *J. Approximation Theory* **8** (1973), 195–209.
6. C. GEIGER AND G. OPFER, Complex Chebyshev polynomials on circular sectors, Universität Hamburg, Institut für Angewandte Mathematik, Preprint 77/1.
7. M. GUTKNECHT, "Ein Abstiegsverfahren für gleichmäßige Approximation, mit Anwendungen", Dissertation, ETH Zürich, 1973.
8. V. KLOTZ, "Polynomiale und rationale Tschebyscheff-Approximation in der komplexen Ebene", Dissertation, Ruhr-Universität Bochum, 1974.
9. W. KRABS, Ein Pseudo-Gradientenverfahren zur Lösung des diskreten linearen Tschebyscheff-Problems, *Computing* **4** (1969), 216–224.
10. W. KRABS AND G. OPFER, Eine Methode zur Lösung des komplexen Approximationsproblems mit einer Anwendung auf konforme Abbildungen, *ZAMM* **55** (1975), T208–T211.
11. C. L. LAWSON, "Contributions to the Theory of Linear Least Maximum Approximation," Ph. D. Dissertation, University of California, Los Angeles, 1961.
12. G. G. LORENTZ AND K. ZELLER, Monotone approximation by algebraic polynomials, *Trans. Amer. Math. Soc.* **149** (1970), 1–18.
13. Y. L. LUKE, "Mathematical Functions and Their Approximations," Academic Press, New York/San Francisco/London, 1975.
14. G. MEINARDUS, Über die Approximation asymptotischer Entwicklungen, I, *Computing* **1** (1966), 39–49.
15. G. MEINARDUS, "Approximation of Functions: Theory and Numerical Methods", Springer, Berlin/Heidelberg/New York, 1967.
16. G. OPFER, Über die Approximation der Identität im Komplexen, *ISNM* **26** (1975), 93–99.
17. T. J. RIVLIN, "The Chebyshev Polynomials," Wiley, New York/London/Sydney/Toronto, 1974.
18. R. SCHULTZ, "Charakterisierung von Minimallösungen und Verbesserung nichtoptimaler Näherungen bei Approximationsaufgaben in normierten Räumen", Diplomarbeit, Universität Hamburg, 1975.
19. I. SINGER, "The Theory of Best Approximation and Functional Analysis," Regional Conference Series in Applied Mathematics, Vol. 13, SIAM, Philadelphia, 1974.
20. V. I. SMIRNOV AND N. A. LEBEDEV, "Functions of a Complex Variable," M.I.T. Press, Cambridge, Mass., 1968.
21. H. VOLLSTEDT, "Ein Abstiegsverfahren für die nichtlineare Tschebyscheff-Approximation, mit Anwendungen", Diplomarbeit, Universität Hamburg, 1975.