View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Elsevier - Publisher Connector

JOURNAL OF APPROXIMATION THEORY 2, 223–232 (1969)

Chebyshev Approximation with Respect to a Weight Function

CHARLES B. DUNHAM

Department of Computer Science, University of Western Ontario, London, Ontario, Canada

In this note we consider a generalization of the Chebyshev approximation problem on a closed interval $[\alpha,\beta]$. Let R be the real line and let w be a function mapping $[\alpha,\beta] \times R \times R$ into the extended real line. Let F be an approximating function, with parameter A taken from a parameter space P. Given $f \in C[\alpha,\beta]$, the generalized Chebyshev problem is to find a parameter $A^* \in P$ minimizing

$$e(A) = \|E(A, .)\|,$$

where the error curve E is given by

$$E(A,x) = w(x, f(x), F(A,x))$$

and

$$\|g\| = \sup\{|g(x)| : \alpha \leqslant x \leqslant \beta\}.$$

Such an approximant $F(A^*, .)$ is called a best approximation to $f(x)$ with respect to w on $[\alpha,\beta]$. In this note we consider the characterization and computation of best approximations.

1. DEFINITIONS AND PRELIMINARIES

To be able to treat the approximation problem, we must put some restrictions on w and F. We assume in this note that w is a continuous mapping into the extended real line, i.e., for every $y_0 \in [\alpha,\beta] \times R \times R$, $\lim_{y \to y_0} w(y) = w(y_0)$, whether $w(y_0)$ is finite or not. It should be noted that there exist cases (such as the case of onesided approximation) of both practical and theoretical importance which do not satisfy this requirement. To obtain an alternating theory and to ensure that the error gets larger as the approximant moves away from the function, it is necessary to have monotonicity conditions on w.

DEFINITION. A continuous mapping w of $[\alpha,\beta] \times R \times R$ into the extended real line is called an *ordering function* if for x and a fixed, $w(x,a,b)$ is a monotonic function of b (strictly monotonic when it is finite) and

$$\operatorname{sgn}(w(x,a,b)) = \operatorname{sgn}(a-b).$$

© 1969 by Academic Press, Inc.

15

Examples. In the case of ordinary Chebyshev approximation we have

$$w(x,a,b) = a - b. \tag{1}$$

A slight generalization of this involves a positive continuous weighting function s,

$$w(x,a,b) = s(x) \qquad (a-b).$$

Moursund [*3*, *4*] has considered "generalized weight functions"

$$w(x,a,b) = v(x, b-a)$$

where for fixed first argument x, $v(x,y)$ is strictly monotonic in y and $\operatorname{sgn}(v(x,y)) = \operatorname{sgn}(y)$.

An ordering function of particular interest is the r-biased weight function

$$\begin{aligned} w_r(x,a,b) &= b-a & b &\geqslant a \\ &= r(b-a) & b &< a. \end{aligned}$$

The limit of r-biased weight functions is the onesided weight function

$$\begin{aligned} w(x,a,b) &= b-a & b &\geqslant a \\ &= -\infty & b &< a. \end{aligned}$$

The onesided weight function is not an ordering function, but from the fact that it is a limit of ordering functions we can deduce much.

Let ϕ be an order function ([*9*], 149). A different generalization of the Chebyshev criterion consists in choosing

$$w(x,a,b) = \phi(a) - \phi(b).$$

By use of such an ordering function, we can convert the problem of Chebyshev approximation by $\phi(F)$ of a function g taking values in $(\inf\phi, \sup\phi)$ into the problem of approximation with respect to w by F of a function $f = \phi^{-1}(g)$, thus avoiding the necessity of developing a separate theory for approximation by the approximating function $\phi(F)$.

2. Characterization of Best Approximations

In this section we show that if an alternating theory exists for F in ordinary Chebyshev approximation, then an alternating theory exists also for F in generalized Chebyshev approximation with an ordering function w.

Definition. A sequence $\{x_0, \ldots, x_n\}$, $\alpha \leqslant x_0 < \ldots < x_n \leqslant \beta$, is said to be an *alternant* of a function g, if for $i = 0, 1, \ldots, n$,

$$\begin{aligned} |g(x_i)| &= \|g\| \\ g(x_i) &= (-1)^i g(x_0). \end{aligned}$$

g is then said to *alternate* n times on $[\alpha, \beta]$.

An approximating function F is said to have *degree* $\rho(A)$ at A, if a necessary and sufficient condition for $F(A, .)$ to be a best Chebyshev approximation to f in the ordinary sense (1), is that $f - F(A, .)$ alternates $\rho(A)$ times. We say for short that $F(A, .)$ is of degree $\rho(A)$. It is well known that every element of the family

$$F(A,x) = \sum_{k=0}^{n} a_k x^k$$

of nth degree polynomials is of degree $n+1$ in the above sense, and there exists a degree for rational families ([7], 301–302; [8], 78) as well. Rice ([6], 324–327) has obtained necessary and sufficient conditions for F to have a degree at every parameter-point A in the parameter space P, that is, necessary and sufficient conditions for an alternating theory.

DEFINITION. F has property Z of degree m at A, if the fact that $F(A, .) - F(B, .)$ has m zeros implies $F(A, .) \equiv F(B, .)$.

DEFINITION. F has property $\mathscr{A}$ of degree n at A, if for any integer $m < n$, any sequence $\{x_1, \ldots, x_m\}$ with

$$\alpha = x_0 < x_1 < \ldots < x_{m+1} = \beta,$$

any sign σ, and any real ϵ with

$$0 < \epsilon < \min\{x_{j+1} - x_j : j = 0, \ldots, m\},$$

there exists a $B \in P$, such that

$$\|F(A, .) - F(B, .)\| < \epsilon,$$

$$\begin{aligned} \operatorname{sgn}(F(A,x) - F(B,x)) &= \sigma & \alpha \leqslant x \leqslant x_1 - \epsilon \\ &= \sigma(-1)^j & x_j + \epsilon \leqslant x \leqslant x_{j+1} - \epsilon \\ &= \sigma(-1)^m & x_m + \epsilon \leqslant x \leqslant \beta. \end{aligned}$$

In case $m = 0$, the above sign condition reduces to

$$\operatorname{sgn}(F(A, .) - F(B, .)) = \sigma.$$

We now require that for all $A \in P$, the degree of property $\mathscr{A}$ at A equals the degree of property Z and A; by Rice's result ([6], 324–327) we can then say that F has that degree at A.

DEFINITION. A zero x of a continuous function g on $[\alpha, \beta]$ is called a double zero, if x is an interior point of $[\alpha, \beta]$ and g does not change sign at x.

LEMMA 1. *Let F have degree n at A. Then if $F(A, .) - F(B, .)$ has n zeros, counting double zeros twice, $F(A, .) \equiv F(B, .)$.*

The proof is similar to the corresponding proof for varisolvent F, given in ([7], 299). We show that if $F(A,.)-F(B,.)$ has n zeros, counting double zeros twice, then there exists a C such that $F(A,.)-F(C,.)$ has n simple zeros.

LEMMA 2. *Let F have degree n at A, $e(A)<\infty$, and let $E(A,.)$ alternate in sign on the sequence $\{x_0,\ldots,x_n\}$; then for $F(B,.)\not\equiv F(A,.)$,* $\min\{E(A,x_i): i=0,\ldots,n\}<\max\{|E(B,x_i)|: i=0,\ldots,n\}$.

Proof. Suppose the inequality fails for some B and assume, without loss of generality, that $E(A,x_0)>0$. We have

$$F(A,x_0)\leqslant F(B,x_0)$$
$$F(A,x_1)\geqslant F(B,x_1)$$
$$F(A,x_2)\leqslant F(B,x_2)$$
$$\cdots\cdots$$

and it is seen that $F(A,.)-F(B,.)$ must have n zeros, counting double zeros twice, on (x_0,x_n). But by Lemma 1, this can only happen if $F(A,.)\equiv F(B,.)$, which proves the lemma.

Two consequences of the lemma are that n alternations of $E(A,.)$ are sufficient for $F(A,.)$ to be best to f, and that an approximation $F(A,.)$ whose error $E(A,.)$ has n alternations is a unique best approximation. One can prove that n alternations are also necessary, as we show in the following theorem. First we need a continuity result.

LEMMA 3. *Let f and g be continuous functions such that w is continuous at all points $(x,f(x),g(x))$, $\alpha\leqslant x\leqslant\beta$. For a given $\epsilon>0$ there exists a $\delta>0$ such that if h is a continuous function for which $\|g-h\|<\delta$, then*

$$\|w(.,f,h)-w(.,f,g)\|<\epsilon.$$

Proof. If the lemma is false, there exists an $\epsilon<0$, a sequence $\{h_k\}$ with $\|g-h_k\|<1/k$, and a sequence $\{x_k\}$ with

$$\gamma_k=|w(x_k,f(x_k),h_k(x_k))-w(x_k,f(x_k),g(x_k))|>\epsilon.$$

As $\{x_k\}$ is bounded, it has a limit point x_0; assume, without loss of generality, that $x_k\to x_0$. In this case, the sequences

$$\{(x_k,f(x_k),h_k(x_k))\}\qquad\text{and}\qquad\{(x_k,f(x_k),g(x_k))\}$$

both converge to $(x_0,f(x_0),g(x_0))$. By the continuity of w at this point it follows that γ_k must tend to zero, which contradicts choice of $\{\gamma_k\}$.

THEOREM 1. *Let F have degree n at A, and let w be an ordering function. If $e(A) < \infty$, a necessary and sufficient condition for $F(A, .)$ to be best is that $E(A, .)$ have n alternations.*

Proof. Sufficiency has already been established by Lemma 2; we now prove necessity. Suppose that $E(A, .)$ alternates exactly $m < \rho(A)$ times. There exists points $x_1, \ldots, x_m$ such that

$$E(A, x_i) = 0 \qquad i = 1, \ldots, m$$

and $E(A, .)$ does not alternate once on any one of the intervals $I_i = [x_i, x_{i+1}]$, $i = 0, \ldots, m$, $x_0 = \alpha$, $x_{m+1} = \beta$. Assume, without loss of generality, that $E(A, .)$ attains $e(A)$ on I_0; then $E(A, .)$ attains $(-1)^i e(A)$ on I_i, and hence does not attain $(-1)^{i+1} e(A)$ on I_i. Let us set

$$\eta = e(A) - \max\{(-1)^{i+1} E(A, x) : x \in I_i, \ i = 0, \ldots, m\};$$

we have then $\eta > 0$. Define

$$J_i = \{x : |E(A, x)| \geqslant e(A) - \eta/2, \ x \in I_i\}.$$

By the choice of J_i, we have $\operatorname{sgn} E(A,x) = (-1)^i$ for $x \in J_i$. Let

$$\delta = \inf\left\{|f(x) - F(A, x)| : x \in \bigcup_{i=1}^{m} J_i\right\}.$$

Using Lemma 3, choose $\epsilon_1 > 0$ such that $\|F(A, .) - F(B, .)\| < \epsilon_1$ implies $\|w(., f, F(A, .)) - w(., f, F(B, .)\| < \eta/2$. Choose ϵ such that $\epsilon < |x_j - y|$ for $j = 1, \ldots, m$ and any $y \in J_i$, $i = 1, \ldots, m$, and $\epsilon < \max\{\epsilon_1, \delta\}$. Now choose $F(B, .)$ as in the definition of property $\mathscr{A}$, with $\sigma = -1$. By the choice of σ, δ and ϵ, we have $F(B,x)$ strictly between $F(A,x)$ and $f(x)$ for $x \in J_i$, hence for such x,

$$|w(x, f(x), F(B, x))| < |w(x, f(x), F(A, x))| \leqslant e(A).$$

For $x \notin \bigcup_{j=1}^{m} J_i$, we have

$$|w(x, f(x), F(B, x))| < |w(x, f(x), F(A, x))| + \eta/2 < e(A) - \eta/2 + \eta/2 = e(A).$$

Thus we have $|w(., f, F(B, .))| < e(A)$, proving the theorem.

Since a best approximation $F(A^*, .)$ must alternate $\rho(A)$ times, it follows from Lemma 2 that a best approximation is unique.

3. THE REMEZ ALGORITHM

For approximating functions F of practical interest there exists a maximum degree. We denote this degree by n for the remainder of this note. A function

f for which the best approximation is of degree n, and whose error curve of the best approximation has exactly $n+1$ relative maxima and minima (which serve as an alternant of length n) is called normal. Sufficient conditions for a function to be normal in rational or exponential approximation have been obtained by Meinardus and Schwedt ([*2*], 315–320). Related results for rational approximation were obtained by Loeb ([*10*], 44). Experience suggests that most functions are normal. If the function f is normal, a simple variant of the Remez algorithm may be used to determine the best approximation. We assume henceforth that f is normal.

Many variants of the Remez algorithm have been described in the literature. For example, Remez-type algorithms for linear and unisolvent approximating functions, respectively, have been described by Rice ([*8*], 176–177) and Novodvorskii and Pinsker [*5*]. Both Kahan [*1*] and Meinardus and Schwedt [*2*] have described algorithms for general approximating functions. We briefly describe a variant of the second algorithm of Remez for the approximation problem of this note.

The Remez algorithm is an iterative procedure for determining an alternant $\{x_0^*,\ldots,x_n^*\}$ of the error curve $E(A^*,.)$ of the best approximation $F(A^*,.)$, and hence determining the best approximation. We start with a sequence $\{x_0^0,\ldots,x_n^0\}$ as an estimate of an alternant of the best approximation. The kth iteration consists of two stages. In the first stage we solve the system

$$E(A^k, x_i^{k-1}) = (-1)^i \lambda^k \qquad i=0,\ldots,n \tag{2}$$

for the unknowns A^k and λ^k. The solution, if it exists, is unique; for suppose $\{A,\lambda\}$, $\{B,\mu\}$ to be two distinct solutions. If $\lambda=\mu$, we must have

$$F(A, x_i^{k-1}) = F(B, x_i^{k-1}), \qquad i=0,\ldots,n,$$

and hence, $F(A,.) \equiv F(B,.)$. If $\lambda \neq \mu$, $F(A,.) - F(B,.)$ alternates in sign on $x_0,\ldots,x_n$, and so has n zeros, contrary to hypothesis. We discuss the solution stage later in more detail.

Set $x_{-1}^{k-1} = \alpha, x_{n+1}^{k-1} = \beta$. Stage (ii) consists of finding a point x_i^k on the interval $[x_{i-1}^{k-1}, x_{i+1}^{k-1}]$, at which $(-1)^i E(A^k,.)\lambda^k$ attains its maximum, $i=0,\ldots,n$. The usual case is where x_i^k is unique, $E(A^k,.)$ is differentiable, and $E'(A^k,.)$ vanishes only at interior points x_i^k. For this case, we can obtain the maxima $\{x_0^k,\ldots,x_n^k\}$ by solving the equations

$$(x_i^k - \alpha)(x_i^k - \beta)E'(A^k, x_i^k) = 0 \qquad i=0,\ldots,n, \tag{3}$$

with the constraint

$$\alpha \leqslant x_0^k < \ldots < x_n^k \leqslant \beta.$$

We then start the $(k+1)$th iteration, with the points $\{x_0^k,\ldots,x_n^k\}$.

We now consider the convergence of the algorithm. In the case F is unisolvent of degree n and w is an ordering function such that for any $x \in [\alpha, \beta]$, $w(x, f(x), .)$ is a $1-1$ mapping of the real line into itself, it can be shown by the arguments of Novodvorskii and Pinsker [5] that the approximants $\{F(A^k, .)\}$ of the algorithm converge uniformly to the best approximation, no matter what starting sequence $\{x_0{}^0, \ldots, x_n{}^0\}$ is chosen. Moursund [4] has given a proof of this for a generalized weight function and a linear unisolvent family. In more general cases, the only results known are local convergence results, which guarantee convergence only if the initial sequence is sufficiently close to the alternant. We extend an unpublished theorem of Kahan [1] from the ordinary case of Chebyshev approximation to the case of this paper. The following notation is useful in developing the result.

$$X = (x_0, \ldots, x_n), \qquad A = (a_1, \ldots, a_n)$$

$$\delta x_i = x_i - x_i{}^*, \qquad \delta a_i = a_i - a_i{}^* \qquad \delta\lambda = \lambda - \lambda^*$$

$$E_x(A, x) = \frac{\partial}{\partial x} E(A, x), \qquad E_j(A, x) = \frac{\partial}{\partial a_j} E(A, x), \qquad F_j(A, x) = \frac{\partial}{\partial a_j} F(A, x)$$

$$E_{xj}(A, x) = \frac{\partial E(A, x)}{\partial x\, \partial a_j}, \qquad E_{jk}(A, x) = \frac{\partial E(A, x)}{\partial a_j\, \partial a_k}, \ldots$$

$$E_x{}^i = E_x(A^*, x_i{}^*), \qquad E_j{}^i = E_j(A^*, x_i{}^*), \qquad E^i_{xj}(A^*, x_i{}^*), \ldots$$

THEOREM 2. *Under the following hypotheses, the infinite sequence* $\{x_0{}^k, \ldots, x_n{}^k\}$, $k = 0, 1, 2, \ldots$, *of the Remez algorithm, converges quadratically to the alternant of the best approximation, if the starting points are sufficiently close.*

(i) *The best approximation to f is of degree n.*

(ii) *The best approximation has a unique alternant* $\{x_0{}^*, \ldots, x_n{}^*\}$ *of length* $n+1$.

(iii) *The parameter space P of F is an open subset of real n-space.*

(iv) *The left-hand sides of equations* (2) *and* (3) *can be expanded in Taylor series in* $\delta x_i{}^k$, $\delta a_j{}^k$, *and* $\delta\lambda^*$ *about the solution* X^*, A^*, λ^* *to the approximation problem.* (X^* *is the alternant of the best approximation, and* $|\lambda^*|$ *is the error of the best approximation.*)

(v) *The matrix*

$$\begin{bmatrix} E_1(A, x_1) \ldots E_n(A, x_1) & -1 \\ \ldots \qquad \ldots & \ldots \\ E_1(A, x_n) \ldots E_n(A, x_n) & (-1)^{n+1} \end{bmatrix} \tag{4}$$

is nonsingular at the solution X^*, A^*.

(vi) *If* x_i *is an interior point,* $E^i_{xx} \neq 0$, *and if* x_i *is an endpoint,* $E_x{}^i \neq 0$.

Proof. We outline Kahan's proof which needs no changes. Expanding the left-hand-side of (2), we obtain,

$$(-1)^i\,\delta\lambda^k = E_x{}^i\,\delta x_i^{k-1} + \sum_j E_j{}^i\,\delta a_j{}^k + \tfrac{1}{2}E_{xx}^i(\delta x^{k-1})^2$$
$$+\,\delta x_i^{k-1}\sum_j E_{xj}^i\,\delta a_j{}^k + \tfrac{1}{2}\sum_j\sum_m E_{jm}\,\delta a_j{}^k\,\delta a_m{}^k + \ldots$$

Now consider the term $E_x{}^i\,\delta x_i^{k-1}$. If $E_x{}^i \neq 0$, $x_i{}^*$ must be an endpoint. If $(\delta A^{k-1}, \delta X^{k-1})$ is sufficiently small, we have $E_x(A^{k-1}, x_i^{k-1}) \neq 0$ also, in which case (3) ensures that x_i^{k-1} is the same endpoint and $\delta x_i^{k-1} = 0$. Hence, all the terms $E_x{}^i\,\delta x_i^{k-1}$ vanish. We rearrange the expansion to get

$$\sum_j (E_j{}^i + \ldots)\,\delta a_j{}^k - (-1)^i\,\delta\lambda^k = -\tfrac{1}{2}(\delta x_i^{k-1})^2\,(E_{xx}^i + \ldots). \tag{5}$$

The first-order terms of the left-hand-side may be considered as the left-hand-side of a set of linear equations in unknowns δA^k and $\delta\lambda^k$, with matrix (4). As the matrix is nonsingular, $(\delta A^k, \delta\lambda^k)$ must be $O(\delta X^{k-1})^2$ for δA^k, $\delta\lambda^k$ small.

We now expand the left-hand-side of (3), obtaining

$$(x_i{}^k - \alpha)\,(x_i{}^k - \beta)\left[E_x{}^i + E_{xx}^i\,\delta x_i{}^k + \sum_{j=1}^n E_{xj}^i\,\delta a_j{}^k + \ldots\right] = 0.$$

If $\delta x_i{}^k \neq 0$ then, as before, $E_x{}^i = 0$ and

$$\delta x_i{}^k = -\left(\sum_{j=1}^n E_{xj}^i\,\delta a_j{}^k + \ldots\right)\Big/ E_{xx}^i = O(\delta X^{k-1})^2,$$

giving quadratic convergence and proving the theorem.

We now consider the hypotheses of the theorem in more detail. With hypotheses (i) and (ii) we are simply assuming that the function f is normal. Hypotheses (iv) and (vi) are essential to most quadratic convergence proofs. We consider hypothesis (v) in more detail. Suppose that hypothesis (iii) holds, and that for any distinct points $x_0, \ldots, x_n$, the matrix

$$\begin{bmatrix} F_1(A, x_1) \ldots F_n(A, x_1) \\ \cdots \qquad \cdots \\ F_1(A, x_n) \ldots F_n(A, x_n) \end{bmatrix} \tag{6}$$

is nonsingular at a parameter-point where the approximant has degree n. It follows that $\{F_1(A, .), \ldots, F_n(A, .)\}$ is a Chebyshev system on $[\alpha, \beta]$, and by a standard argument ([*8*], 65), the matrix

$$\begin{bmatrix} F_1(A, x_0) \ldots F_n(A, x_0) & -s_1 \\ \cdots \qquad \cdots & \cdots \\ F_1(A, x_n) \ldots F_n(A, x_n) & (-1)^{n+1} s_n \end{bmatrix} \tag{7}$$

is nonsingular for distinct $x_0 < \ldots < x_n$ and for $s_i > 0$, $i = 1, \ldots, n$. Now

$$E_j(A, x_i) = \frac{\partial}{\partial a_j} w(x_i, f(x_i), F(A, x_i))$$
$$= w'(x_i, f(x_i), F(A, x_i))\, F_j(A, x_i),$$

where the prime denotes differentiation with respect to b of $w(x,a,b)$. In the case $w_i = w'(x_i, f(x_i), F(A,x_i)) > 0$, we can divide the ith row of (4) by w_i, $i = 0, \ldots, n$, obtaining a matrix of the form (7) with $s_i = 1/w_i$. This matrix is nonsingular, and hence, if the hypotheses on w' and the matrix (6) are satisfied, (4) is nonsingular. In the case where $F(A, .)$ is a linear unisolvent function,

$$F(A, x) = \sum_{k=1}^{n} a_k\, \phi_k(x),$$

the matrix (6) is a generalized Vandermonde matrix, and is nonsingular for all distinct $x_1, \ldots, x_n$.

Let us consider Newton's method applied to (2). Define

$$r_i^k(A, \lambda) = E(A, x_i^{k-1}) - (-1)^i \lambda \qquad i = 0, \ldots, n.$$

With parameters A, λ as estimates of the solution (we choose as initial estimates A^{k-1}, λ^{k-1}), new parameters $A + \Delta A$, $\lambda + \Delta\lambda$ are obtained by solving the linear system

$$\sum_{j=1}^{n} E_j(A, x_i^{k-1})\, \Delta a_j - (-1)^i \Delta\lambda = -r_i^k(A, \lambda) \qquad i = 0, \ldots, n. \tag{8}$$

We can solve the system $r_i^k(A, \lambda) = 0$ by iteration. However, we could try only one iteration of solving. If we expand the left hand side of (8), we obtain

$$\sum_{j=1}^{n} \left(E_j^i + E_{xj}^i\, \delta x_i + \sum_{m=1}^{n} E_{jm}^i\, \delta a_m + \ldots \right) \Delta a_j - (-1)^i \Delta\lambda,$$

and from (5) we see that the right-hand side is

$$\sum_{j=1}^{n} (E_j^i + E_{xj}^i\, \delta x_i + \ldots)\, \delta a_j - (-1)^i\, \delta\lambda - \tfrac{1}{2}(\delta x_i^{k-1})^2\, (E_{xx}^i + \ldots).$$

This gives

$$\Delta A^k - \delta A^k = O(\delta X^{k-1})^2, \qquad \Delta\lambda - \delta\lambda = O(\delta X^{k-1})^2.$$

The argument of the last paragraph of the proof of Theorem 2 may be used to obtain $\delta x^k = O(\delta X_i^{k-1})^2$. This shows that if we do only one iteration of Newton's method in stage (ii) of our algorithm, we still have quadratic convergence. Such a conclusion was obtained by Meinardus and Schwedt [2] in the ordinary case.

References

1. W. Kahan, Unpublished communication, 1964.
2. G. Meinardus and D. Schwedt, Nicht-lineare Approximationen. *Arch. Rational Mech. Anal.* **17** (1964), 297–326.
3. D. Moursund, Chebyshev approximation using a generalized weight function. *J. SIAM Numer. Anal.* **3** (1966), 435–450.
4. D. Moursund, Computational aspects of Chebyshev approximation using a generalized weight function. *SIAM J. Numer. Anal.* **5** (1968), 126–137.
5. E. Novodvorskii and I. Pinsker, On a process of equalization of maxima (in Russian). *Usp. Mat. Nauk* **6** (1951), 174–181.
6. J. R. Rice, The characterization of best nonlinear Tchebycheff approximations. *Trans. Am. Math. Soc.* **96** (1960), 322–340.
7. J. R. Rice, Tchebycheff approximation by functions unisolvent of variable degree. *Trans. Am. Math. Soc.* **99** (1961), 298–302.
8. J. R. Rice, "The Approximation of Functions," Vol. 1. Addison-Wesley, Reading, Mass., 1964.
9. C. B. Dunham, Transformed Rational Chebyshev Approximation. *Numer. Math.* **10** (1967), 147–152.
10. H. L. Loeb, Approximation by Generalized Rationals. *SIAM J. Numer. Anal.* **3** (1966), 34–55.