JOURNAL OF APPROXIMATION THEORY 7, 48–58 (1973)

On the Best Linear One-Sided Chebyshev Approximation

G. A. WATSON

Department of Mathematics, University of Dundee, Dundee, Scotland

Communicated by C. W. Clenshaw

Received June 20, 1970

1. INTRODUCTION

The classical problems

(i) find a real vector $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, ..., \alpha_p)^T$ to minimise

$$\max | r_i |, \qquad i = 1, 2, ..., n > p,$$

where

$$\mathbf{r} = \mathbf{b} - A\boldsymbol{\alpha}, \tag{1.1}$$

and A is an $n \times p$ matrix (the discrete T-problem) and

(ii) find $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, ..., \alpha_p)^T$ to minimise

$$\max | r(x, \boldsymbol{\alpha})|, \qquad a \leqslant x \leqslant b,$$

where

$$r(x, \boldsymbol{\alpha}) = f(x) - \sum_{i=1}^{p} \alpha_i \phi_i(x), \tag{1.2}$$

with $f(x) \in C[a, b]$, $\phi_i(x) \in C[a, b]$, $i = 1, 2, ..., p$ (the continuous T-problem) are now well understood. In particular if the matrix A of equation (1.1) has rank p, then problem (i) can be solved as a linear programming problem (see, for example, Stiefel [1], Osborne and Watson [2]).

In this paper, we are concerned with problems (i) and (ii) where the solution $\boldsymbol{\alpha}$ satisfies the additional constraints

$$r_i \geqslant 0, \qquad i = 1, 2, ..., n \tag{1.3}$$

and

$$r(x, \boldsymbol{\alpha}) \geqslant 0, \qquad a \leqslant x \leqslant b, \tag{1.4}$$

respectively. In this case we have the problems of discrete and continuous one-sided Chebyshev approximation from above. If the inequalities in (1.3) and (1.4) are reversed, then we have the corresponding problems of one-sided Chebyshev approximation from below. We will be concerned in this paper entirely with approximation from above, although analogous results hold for approximations from below.

Our aim is to demonstrate the power of linear programming as a tool in the development of theory and algorithms for one-sided Chebyshev approximation problems. As is to be expected, this is particularly evident with regard to the discrete problem and in Section 2, it is shown how results analogous to those for the discrete T-problem are readily obtained. In Section 3, we consider the continuous problem and show how linear programming can be used as the basis for an algorithm which is similar to the first algorithm of Remes [3] for the continuous T-problem and which converges under a minimum of restrictions on the problem.

In the interests of clarity, we mention some details concerning the notation used in connection with partitioned vectors and matrices. For example $\begin{bmatrix} A \\ \mathbf{x}^T \end{bmatrix}$ represents the matrix A extended by a row vector $\mathbf{x}^T$ and $\begin{bmatrix} A \\ B \end{bmatrix}$ represents the matrix A extended by the rows of the matrix B. Similarly $[A\ \mathbf{x}]$ represents the matrix A extended by a column vector $\mathbf{x}$ and $[A\ B]$ represents the matrix A extended by the columns of the matrix B. A vector of the form $\begin{bmatrix} \mathbf{b} \\ \mathbf{c} \end{bmatrix}$ represents the column vector $\mathbf{b}$ extended by the elements of the column vector $\mathbf{c}$ and a vector of the form $[\mathbf{b}^T\ \mathbf{c}^T]$ represents the row vector $\mathbf{b}^T$ extended by the elements of the row vector $\mathbf{c}^T$.

The elements of the matrix denoted by $-A$ are the negatives of those of A, and $-\mathbf{b}$ and $-\mathbf{b}^T$ represent vectors whose elements are the negatives of those of $\mathbf{b}$ and $\mathbf{b}^T$ respectively. Finally, it will be necessary to make use of the null vector both as a row and column vector. For simplicity, we have just used 0 in either case. The appropriate meaning will be clear from the context.

2. Linear Programming and the Discrete Problem

In this section, it will be necessary to use standard results from linear programming theory. These will be quoted without reference but details may be obtained in, for example, Hadley [4]. It is convenient to state the problem of linear discrete one-sided Chebyshev approximation from above as follows.

Let

$$\mathbf{r} = \mathbf{b} - A\boldsymbol{\alpha}, \tag{2.1}$$

and define

$$P = \{\boldsymbol{\alpha} : r_i \geqslant 0,\ i = 1, 2, \ldots, n\}. \tag{2.2}$$

Find $\boldsymbol{\alpha} \in P$ to minimise

$$\max r_i , \qquad i = 1, 2,..., n.$$

We assume that $n > p$. Let

$$h = \max r_i , \qquad i = 1, 2,..., n. \tag{2.3}$$

Then this problem can be stated: Minimise h, subject to

$$0 \leqslant \mathbf{b} - A\boldsymbol{\alpha} \leqslant h\mathbf{e}, \tag{2.4}$$

where $\mathbf{e}$ is a vector of which each component is 1. This is clearly a linear programming problem and can be more conventionally formulated as

$$\text{minimise } \mathbf{e}_{p+1}^T \begin{bmatrix} \boldsymbol{\alpha} \\ h \end{bmatrix}$$

subject to

$$\begin{bmatrix} A & \mathbf{e} \\ -A & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha} \\ h \end{bmatrix} \geqslant \begin{bmatrix} \mathbf{b} \\ -\mathbf{b} \end{bmatrix}, \tag{2.5}$$

where $\mathbf{e}_{p+1}$ is a vector, each component of which is zero except the $(p + 1)$st which is 1.

This form is still not particularly suitable for the application of standard techniques because the matrix of contraints is such that $2n$ slack variables are required and also because the components of $\boldsymbol{\alpha}$ are not constrained to be non-negative. As in the corresponding T-problem, both these difficulties are overcome by going to the dual linear programming problem, which is

maximise $$z = [\mathbf{b}^T \; -\mathbf{b}^T]\, \mathbf{w}$$

subject to

$$\begin{bmatrix} A^T & -A^T \\ \mathbf{e}^T & 0 \end{bmatrix} \mathbf{w} = \mathbf{e}_{p+1}, \qquad \mathbf{w} \geqslant 0. \tag{2.6}$$

Remark. Since h is not unconstrained, the last equation of (2.6) does not automatically hold with equality. However, an argument similar to that given in the proof of Lemma 4.1 of Osborne and Watson [2] shows that equality must hold unless an optimum value of z exists at $\mathbf{w} = 0$. We exclude this case from consideration.

LEMMA 1. *Necessary and sufficient conditions for a solution by the simplex method of linear programming to the problem defined by* (2.1) *and* (2.2) *are that*

(i) *the set P is non-null,*

(ii) *the matrix A has rank p.*

Proof. The necessity of (i) is obvious. Suppose, therefore, that A has rank $r < p$. Then the rank of the matrix

$$\begin{bmatrix} A^T & -A^T \\ \mathbf{e}^T & 0 \end{bmatrix}$$

is less than $p + 1$ and so no basic feasible solution exists to the linear programming problem (2.6). This concludes the proof of necessity.

Now suppose that (i) and (ii) are satisfied. Then, by (ii) there exists a matrix A^* formed by $p + 1$ rows of A such that

$$\boldsymbol{\lambda}^T A^* = 0 \tag{2.7}$$

for some nontrivial vector $\boldsymbol{\lambda}$, unique up to a scalar multiple. Thus there exists a vector $\mathbf{w} \geqslant 0$ such that

$$[A^T - A^T]\,\mathbf{w} = 0,$$

where there are at most $p + 1$ non-zero components of $\mathbf{w}$, equal to values of $|\lambda_i|$, $i = 1, 2,..., p + 1$.

Further, we can introduce a scaling factor such that

$$[\mathbf{e}^T\ 0]\,\mathbf{w} = 1,$$

and it follows that $\mathbf{w}$ is a basic feasible solution to the dual constraints (2.6). Thus, since the primal problem has a feasible solution, both problems have optimal solutions and the sufficiency is proved.

LEMMA 2. *At a basic feasible solution to the Eqs.* (2.6), *at least* $(p + 1)$ *of the constraints* (2.5) *will hold with equality.*

Proof. This follows immediately from the result that if a variable is in the dual basis, then the corresponding primal constraint must hold with equality.

LEMMA 3. *If a column of* $\begin{bmatrix} A^T \\ \mathbf{e}^T \end{bmatrix}$ *and the corresponding column of* $\begin{bmatrix} -A^T \\ 0 \end{bmatrix}$ *are present together in the dual basis matrix, then the current value of* $z \leqslant 0$.

Proof. Suppose that corresponding columns are in fact present in the dual basis matrix. Then, for some i,

$$\begin{aligned} \rho_i(A)\,\boldsymbol{\alpha} + h &= b_i\,, \\ -\rho_i(A)\,\boldsymbol{\alpha} &= -b_i\,, \end{aligned}$$

by Lemma 2, where $\rho_i(A)$ denotes the ith row of A.

Thus $h = 0$, and so $z \leqslant 0$ by the relationship between the primal and dual objective functions.

COROLLARY. *By our assumption on the optimal value of z, it follows that at some stage of the calculation, we must have no corresponding columns together in the dual basis matrix.*

We are now able to place the discrete one-sided Chebyshev approximation problem on a theoretical basis completely analogous to that for the discrete T-problem. We begin by introducing a number of definitions:

1. Any set of $(p + 1)$ Eqs. of (2.1) is called a *reference*. We will write this as

$$A^\sigma \boldsymbol{\alpha} = \mathbf{b}^\sigma - \mathbf{r}^\sigma \tag{2.8}$$

2. If the rank of A^σ is p, then there exists a nontrivial vector $\boldsymbol{\lambda}$, unique up to a scalar multiple, called the λ-vector for the reference, such that

$$\boldsymbol{\lambda}^T A^\sigma = 0. \tag{2.9}$$

3. The vector $\boldsymbol{\alpha}$ is called a *reference vector* if, for all i such that $r_i^\sigma \lambda_i \neq 0$, we have

$$\operatorname{sgn}(r_i^\sigma) = \operatorname{sgn}(\lambda_i),$$

or

$$\operatorname{sgn}(r_i^\sigma) = -\operatorname{sgn}(\lambda_i).$$

4. Let

$$\mu = \operatorname{sgn}(\boldsymbol{\lambda}^T \mathbf{b}^\sigma), \tag{2.10}$$

and define a vector $\mathbf{g}$ by

$$\begin{aligned} g_i &= 1, \qquad \mu\lambda_i > 0, \\ g_i &= 0, \qquad \mu\lambda_i < 0, \end{aligned}$$

and g_i arbitrary in $[0, 1]$ if $\lambda_i = 0$.

Then the matrix

$$[A^\sigma \; \mathbf{g}]$$

is nonsingular provided that A^σ has rank p, so that the vector $\begin{bmatrix} \boldsymbol{\alpha} \\ h \end{bmatrix}$ is uniquely defined by the set of equations

$$A^\sigma \boldsymbol{\alpha} = \mathbf{b}^\sigma - h\mathbf{g}. \tag{2.11}$$

In this case, $\boldsymbol{\alpha}$ is called the *one-sided reference vector* and h is called the *reference deviation.*

LEMMA 4. *The one-sided reference vector solves the linear discrete one-sided Chebyshev approximation problem for the given reference.*

Proof. Consider the solution of the problem (2.8) by linear programming. Then since at the optimum we require equality in $(p + 1)$ primal equations,

$$A^{\sigma}\boldsymbol{\alpha} = \mathbf{b}^{\sigma} - h\mathbf{q}, \tag{2.12}$$

where $q_i = 1$ or 0, $i = 1, 2, \ldots, p + 1$. It remains to show that $\mathbf{q}$ satisfies the definition of $\mathbf{g}$ given above.

Corresponding to Eq. (2.6), as A^{σ} has rank p, we must have

$$\mathbf{w}^{*T} A^{\sigma} = 0, \tag{2.13}$$

where

$$\begin{aligned} w_i^* &= w_i, && \text{if} \quad q_i = 1, \\ w_i^* &= -w_i, && \text{if} \quad q_i = 0. \end{aligned}$$

Thus, by the uniqueness of the λ-vector for the reference, we require

$$\boldsymbol{\lambda} = \beta \mathbf{w}^*,$$

where β is a scaling factor.

Now the optimal reference deviation is given by

$$h = \mathbf{w}^{*T}\mathbf{b}^{\sigma},$$

which is greater than zero by assumption. Thus

$$\mu = \operatorname{sgn}(\boldsymbol{\lambda}^T \mathbf{b}^{\sigma}) = \operatorname{sgn}(\beta).$$

It follows immediately from this result that

(i) if $\mu\lambda_i > 0$, then $q_i = 1$,

(ii) if $\mu\lambda_i < 0$, then $q_i = 0$,

and this completes the proof.

COROLLARY. The vector $\mathbf{w}^*$ gives the λ-vector for the reference scaled so that

$$\mathbf{w}^{*T}\mathbf{g} = 1.$$

LEMMA 5. *Assume that the linear programming problem* (2.6) *is being solved by the simplex algorithm, and that a stage has been reached where the current dual basis matrix does not contain corresponding columns of* $\begin{bmatrix} A^T \\ \mathbf{e}^T \end{bmatrix}$ *and* $\begin{bmatrix} -A^T \\ 0 \end{bmatrix}$. *Let* $d = \min\{r_i, h - r_i\}$, $i = 1, 2, \ldots, n$, *where* h *is the current reference deviation. Then*

(i) $d \geqslant 0$, *the optimal solution has been obtained,*

(ii) *if* $d < 0$ *and* $d = r_j$ *for some* j, *the corresponding column of* $\begin{bmatrix} -A^T \\ 0 \end{bmatrix}$ *enters the basis.*

(iii) *if* $d < 0$ *and* $d = h - r_j$ *for some* j, *the corresponding column of* $\begin{bmatrix} A^T \\ \mathbf{e}^T \end{bmatrix}$ *enters the basis.*

Proof. Let

$$\mathbf{c}^T = [\mathbf{b}^T - \mathbf{b}^T]. \tag{2.14}$$

Let B be the current dual basis matrix and let $\mathbf{c}_B$ be the vector obtained by deleting the elements of $\mathbf{c}$ corresponding to the nonbasic variables. Further, define

$$z_i = \mathbf{c}_B{}^T B^{-1} K_i \begin{bmatrix} A^T & -A^T \\ \mathbf{e}^T & 0 \end{bmatrix}, \qquad i = 1, 2,..., 2n,$$

where $K_i[M]$ denotes the ith column of the matrix M.

Now, by Lemma 2,

$$B^T \begin{bmatrix} \boldsymbol{\alpha} \\ h \end{bmatrix} = \mathbf{c}_B, \qquad \text{and so}$$

$$z_i = \boldsymbol{\alpha}^T K_i[A^T] + h, \qquad i = 1, 2,..., n,$$
$$z_i = -\boldsymbol{\alpha}^T K_i[A^T], \qquad i = n + 1,..., 2n.$$

Thus, using equations (2.1) and (2.14)

$$z_i - c_i = h - r_i, \qquad i = 1, 2,..., n,$$
$$z_i - c_i = r_i, \qquad i = n + 1,..., 2n.$$

Now, it is a standard linear programming result that the vector to enter the basis in a maximization problem (using the simplex algorithm) is given by that corresponding to j such that

$$z_j - c_j = \min(z_i - c_i)$$

for all i such that $z_i - c_i < 0$. Further, if $z_i - c_i \geqslant 0$, $i = 1, 2,..., 2n$, the optimum has been reached.

The results (i), (ii) and (iii) follow immediately.

3. Linear Programming and the Continuous Problem

The problem of linear continuous one-sided Chebyshev approximation from above can be stated as follows:

Let

$$r(x, \boldsymbol{\alpha}) = f(x) - \sum_{i=1}^{p} \alpha_i \phi_i(x), \tag{3.1}$$

where $f(x) \in c[a, b]$, $\phi_i(x) \in c[a, b]$, $i = 1, 2, \ldots, P$, and define

$$P = \{\boldsymbol{\alpha} : r(x, \boldsymbol{\alpha}) \geqslant 0, a \leqslant x \leqslant b\}. \tag{3.2}$$

Find $\boldsymbol{\alpha} \in P$ to minimise

$$\max r(x, \boldsymbol{\alpha}) \qquad a \leqslant x \leqslant b. \tag{3.3}$$

The properties of the continuous problem do not admit an analysis through the theory of linear programming in a manner comparable with that of the discrete problem. However, the results of Lemma 5 can be used to enable an algorithm based on linear programming to be developed for the solution of the continuous problem as posed above. The algorithm, which is similar to the first algorithm of Remes for the continuous T-problem, involves the solution of a sequence of discrete problems, and the remainder of this section is devoted to a description of the algorithm and to a proof of its convergence. The proof is based on that of Cheney [5] for the convergence of the first algorithm of Remes, and the notation used is similar.

Let X denote the interval $[a, b]$ and let X^k be a finite discrete subset of X. Then, provided that X^k contains at least $p + 1$ points, we can define a discrete problem as follows.

Let

$$r(x, \boldsymbol{\alpha}) = f(x) - \sum_{i=1}^{p} \alpha_i \phi_i(x), \qquad x \in X^k \tag{3.4}$$

and let

$$P^k = \{\boldsymbol{\alpha} : r(x, \boldsymbol{\alpha}) \geqslant 0, x \in X^k\}. \tag{3.5}$$

Find $\boldsymbol{\alpha} \in P^k$ to minimise

$$\max r(x, \boldsymbol{\alpha}), \qquad x \in X^k. \tag{3.6}$$

A solution to this problem by linear programming will be obtained provided that the conditions of Lemma 1 are satisfied. In particular, if the set P defined by (3.2) is non-null, then the condition (i) will hold for all k, and we can define

$$m = \inf_{\boldsymbol{\alpha} \in P} \Delta(\boldsymbol{\alpha}),$$

where

$$\Delta(\boldsymbol{\alpha}) = \max_{x \in X} | r(x, \boldsymbol{\alpha}) | , \tag{3.7}$$

and $\boldsymbol{\alpha}^k \in P^k$ to be the vector which minimises

$$\Delta^k(\boldsymbol{\alpha}) = \max_{x \in X^k} | r(x, \boldsymbol{\alpha})| . \tag{3.8}$$

Then the steps of the algorithm are as follows.

(1) Find $\boldsymbol{\alpha}^k \in P^k$ to minimise $\Delta^k(\boldsymbol{\alpha})$.

(2) Find $x_1^k \in X$ to minimise $\Delta^k(\boldsymbol{\alpha}^k) - r(x, \boldsymbol{\alpha}^k)$, and $x_2^k \in X$ to minimise $r(x, \boldsymbol{\alpha}^k)$.

(3) Set $X^{k+1} = X^k \cup x_1^k \cup x_2^k$.

It is an immediate consequence of Lemma 5 that

$$\Delta^k(\boldsymbol{\alpha}^k) \leqslant \Delta^{k+1}(\boldsymbol{\alpha}^{k+1}), \tag{3.9}$$

for all k. Further, the sequence is bounded above by m, and so tends to a limit. We have

LEMMA 6. *Let $\boldsymbol{\gamma}$ be a limit point of the sequence $\{\boldsymbol{\alpha}^k\}$. Then $\boldsymbol{\gamma} \in P$.*

Proof. Let

$$\delta_k = \min_{x \in X} r(x, \boldsymbol{\alpha}^k) = r(\xi^k, \boldsymbol{\alpha}^k), \text{ say.}$$

Then, by the algorithm, we must have

$$r(\xi^i, \boldsymbol{\alpha}^k) \geqslant 0, \qquad k > i.$$

Now, for any $\boldsymbol{\alpha}$, $\boldsymbol{\gamma}$ and any $x \in X$,

$$| r(x, \boldsymbol{\alpha}) - r(x, \boldsymbol{\gamma})| \leqslant M \mid \boldsymbol{\alpha} - \boldsymbol{\gamma} \mid, \tag{3.10}$$

where $M = \max_i \max_{x \in X} | \phi_i(x)|$, and we define

$$| \mathbf{v} | = \sum_{i=1}^{p} | v_i | ,$$

for a vector $\mathbf{v}$ with p elements.

Thus, in particular

$$| \delta_k - r(\xi^k, \boldsymbol{\gamma})| \leqslant M \mid \boldsymbol{\alpha}^k - \boldsymbol{\gamma} \mid,$$

and so if $\boldsymbol{\gamma}$ is a limit point of $\{\boldsymbol{\alpha}^k\}$ and we define

$$\eta_k = \delta_k - r(\xi^k, \boldsymbol{\gamma}),$$

we have

$$\eta_k \to 0 \qquad \text{as} \quad k \to \infty.$$

Now suppose that

$$r(x_0, \gamma) = r_0 < 0 \tag{3.11}$$

for some $x_0 \in X$.

Then, for all k sufficiently large, say $k > k_0$, we will have

$$\eta_k > \tfrac{1}{2} r_0,$$

and consequently

$$\begin{aligned}\delta_k &= \eta_k + r(\xi^k, \gamma)\\ &\geqslant \eta_k\\ &> r(x_0, \gamma).\end{aligned}$$

This contradicts the definition of δ_k. Thus no x_0 exists satisfying (3.11) and the lemma is proved.

LEMMA 7.

$$\Delta^k(\boldsymbol{\alpha}^k) \to m.$$

Proof. By equation (3.9), $\Delta^k(\boldsymbol{\alpha}^k) \to m - \epsilon$, for some $\epsilon \geqslant 0$. It remains to show that $\epsilon = 0$.

Suppose that $\epsilon > 0$ and let γ denote a limit point of the sequence $\{\boldsymbol{\alpha}^k\}$. Thus, for any $\delta > 0$ we may find an index k such that

$$| \gamma - \boldsymbol{\alpha}^k | < \delta,$$

and an index $i > k$ such that

$$| \gamma - \alpha^i | < \delta.$$

Then $| \boldsymbol{\alpha}^i - \boldsymbol{\alpha}^k | \leqslant 2\delta$, and using Lemma 6 and equation (3.10),

$$m \leqslant \Delta(\gamma) \leqslant \Delta(\boldsymbol{\alpha}^k) + M\delta.$$

Now

$$\Delta(\boldsymbol{\alpha}^k) = | r(x_0{}^k, \boldsymbol{\alpha}^k)|,$$

where $x_0{}^k = x_1{}^k$ or $x_2{}^k$.

Thus

$$\begin{aligned} m &\leqslant | r(x_0{}^k, \boldsymbol{\alpha}^k)| + M\delta\\ &\leqslant | r(x_0{}^k, \boldsymbol{\alpha}^i)| + 3M\delta\\ &\leqslant \Delta^i(\alpha^i) + 3M\delta\\ &\leqslant m - \epsilon + 3M\delta.\end{aligned}$$

If $\delta < \epsilon/3M$ this is a contradiction, and the result follows.

It is a sufficient condition for the application of the above algorithm that the initial discrete set contains $(p + 1)$ points, and that the matrix of the corresponding discrete problem has rank p. In fact, it is a feature of the simplex algorithm that successive subsets X^k need only contain $(p + 1)$ points, as only these points corresponding to basic variables need be considered.

A sufficient condition for (3.9) to hold with strict inequality is that the successive linear programming problems have non-degenerate optimal solutions. In this case, it is necessary for the λ-vectors for the optimal references to contain no zero elements. A sufficient condition for this is that the functions $\phi_i(x)$ form a Chebyshev set in $[a, b]$, i.e. no linear combination vanishes at more than $(p - 1)$ points in $[a, b]$ (see for example [2]).

Note, finally, that since the above algorithm represents an infinite process, little can be deduced directly about the characteristics of the solution to the continuous problem. For example, it is incorrect to assume that at least $(p + 1)$ points of $[a, b]$, $r(x, \boldsymbol{\alpha})$ must either attain its maximum value or be zero, for it is possible for points represented in the successive optimal basis matrices to coalesce, as in the continuous T-problem (see Osborne and Watson [6]). Clearly there exist characterisation theorems precisely analogous to those for the continuous T-problem, but these are not available directly through the medium of linear programming.

Acknowledgment

I am indebted to the referee for his valuable advice.

References

1. E. Stiefel, Note on Jordan elimination, Linear programming and Tschebyscheff Approximation, *Numer. Math.* **2** (1960), 1–17.
2. M. R. Osborne and G. A. Watson, On the best linear Chebyshev approximation, *Comp. J.* **10** (1967), 172–177.
3. E. Ya. Remes, On the best approximation of functions in the Tchebyscheff sense, Ukrainsky Matematichesky Zhurnal, Kiev, 1935.
4. G. Hadley, Linear programming, Addison–Wesley Publishing Co. 1962.
5. E. W. Cheney, "Introduction to Approximation Theory," McGraw-Hill, New York, 1966.
6. M. R. Osborne and G. A. Watson, A note on singular minimax approximation problems, *J. Math. Anal. Appl.* **25** (1969), 692–700.