JOURNAL OF APPROXIMATION THEORY 1, 167–175 (1968)

Lower Bounds for the Minimal Distance in Rational Approximation[1]

WERNER KRABS

Institut für Angewandte Mathematik der Universität, Hamburg, Germany

1. INTRODUCTION

We consider a compact Hausdorff space M and denote by $C(M)$ the vector space of real-valued continuous functions on M. As a norm in $C(M)$ we introduce the maximum norm

$$\|g\|_M = \max_{x \in M} |g(x)|, \qquad g \in C(M).$$

Let U and V be finite-dimensional subspaces of $C(M)$ which are spanned by $u_0, \ldots, u_r$ and $v_0, \ldots, v_s \in C(M)$. We assume the convex cone

$$V_M{}^+ = \{v \in V : v(x) > 0 \qquad \text{for all} \quad x \in M\}$$

to be nonempty. For each $f \in C(M)$ we define

$$\rho_M(f) = \inf_{u \in U,\ v \in V_M{}^+} \left\| f - \frac{u}{v} \right\|_M$$

and call this number the minimal distance between f and $W_M = \{u/v : u \in U, v \in V_M{}^+\}$.

The rational approximation problem consists of finding $\hat{u} \in U$ and $\hat{v} \in V_M{}^+$ such that

$$\left\| f - \frac{\hat{u}}{\hat{v}} \right\|_M = \rho_M(f).$$

$\hat{u}/\hat{v}$ is called a best approximant of f (in W_M). Each couple $u \in U$, $v \in V_M{}^+$ yields a trivial upper bound of $\rho_M(f)$. However, for an estimation of the difference between $\|f - (u/v)\|_M$ and $\rho_M(f)$ or even for an estimation of $\rho_M(f)$ itself it is important to know lower bounds of $\rho_M(f)$.

In [*5*] we have developed a principle for the computation of such lower bounds which has been originated by Collatz [*2*], [*3*], [*4*] and has been expanded to nonlinear approximation by Meinardus and Schwedt [*10*], [*11*].

In this paper we intend to develop a more general principle which can be handled in a simpler way. For that purpose we consider a nonempty closed

[1] This research was supported by the Air Force Office of Scientific Research under grant AF-AFOSR-937-67.

subset D of M. D is compact and all the functions $g \in C(M)$ can be considered as real-valued continuous functions on D provided with the norm

$$\|g\|_D = \max_{x \in D} |g(x)|.$$

If we define

$$V_D^+ = \{v \in V : v(x) > 0 \qquad \text{for all} \qquad x \in D\},$$

we get a nonempty subset of V since $V_M^+ (\subseteq V_D^+)$ is assumed to be nonempty. Furthermore, for

$$\rho_D(f) = \inf_{u \in U, v \in V_D^+} \left\| f - \frac{u}{v} \right\|_D$$

we have

$$\rho_D(f) \leqslant \rho_M(f).$$

Our aim is to compute $\rho_D(f)$ or at least to find lower bounds for $\rho_D(f)$ when D is a certain finite subset of M.

The results of this paper, without proofs, have been given in [6].

2. Lower Bounds for the Minimal Distance

Notation. By $\boldsymbol{R}^n$ we denote the real Euclidean n-space and by θ_n the zero vector of $\boldsymbol{R}^n$. For $x, y \in \boldsymbol{R}^n$ we write $x \geqslant y$ if and only if $x_i \geqslant y_i$ for $i = 1, \ldots, n$. z^T denotes the transposed vector z. By $|z|$ we mean the vector $(|z_1|, \ldots, |z_n|)^T$, where $z = (z_1, \ldots, z_n)^T$.

Lemma 2.1. *If we assume that for a subset $D = \{x_1, \ldots, x_n\}$ of M there exist two vectors $c = (c_1, \ldots, c_n)^T \neq \theta_n$ and $p = (p_1, \ldots, p_n)^T \geqslant \theta_n$ and numbers $\lambda_1, \ldots, \lambda_n \in \boldsymbol{R}$ such that*

$$\sum_{i=1}^{n} u_j(x_i)\, c_i = 0, \qquad j = 0, \ldots, r, \tag{2.1}$$

$$\sum_{i=1}^{n} f(x_i)\, v_k(x_i)\, c_i = \sum_{i=1}^{n} \lambda_i(|c_i| + p_i)\, v_k(x_i), \tag{2.2}$$

$$k = 0, \ldots, s$$

then we have

$$\min_{i=1,\ldots,n} \lambda_i \leqslant \rho_D(f). \tag{2.3}$$

(The assertion is a slight generalization of Satz 1 in [7] where all the λ_i's are assumed to be equal.)

Proof. The case $\lambda_i \leqslant 0$ for at least one i is trivial. Hence we assume

$$\min_i \lambda_i > 0.$$

For a given $u \in U$ and $v \in V_D{}^+$ the conditions (2.1), (2.2) and $p \geqslant \theta_n$ imply

$$\sum_{i=1}^n c_i\left(f(x_i) - \frac{u(x_i)}{v(x_i)}\right) v(x_i) = \sum_{i=1}^n \lambda_i(|c_i| + p_i)\, v(x_i)$$
$$\geqslant \sum_{i=1} \lambda_i |c_i|\, v(x_i) \geqslant (\min_i \lambda_i) \sum_{i=1} |c_i|\, v(x_i)$$

and because of

$$\sum_{i=1}^n |c_i|\, v(x_i) > 0$$

we have

$$\min \lambda_i \leqslant \frac{\sum_{i=1}^n c_i\left(f(x_i) - \frac{u(x_i)}{v(x_i)}\right) v(x_i)}{\sum_{i=1}^n |c_i|\, v(x_i)} \leqslant \left\| f - \frac{u}{v} \right\|_D.$$

Since $u \in U$ and $v \in V_D{}^+$ are arbitrarily chosen, we can conclude (2.3), which completes the proof.

In the case $\rho_M(f) > 0$ by Lemma 3.2 of [8] there exist for each $\lambda \in (0, \rho_M(f)]$, $n(\leqslant r+s+3)$ distinct points $x_1, \ldots, x_n$ and vectors $c \neq \theta_n$, $p \geqslant \theta_n$ such that (2.1) and (2.2) hold if we choose

$$\lambda_i = \lambda \qquad \text{for} \qquad i = 1, \ldots, n.$$

Hence, in principle, $\rho_M(f)$ can be estimated from below as best as possible by use of Lemma 2.1. However, Lemma 2.1 is not very convenient for numerical purposes.

In order to find a result which can be handled with less effort we need the following:

Assumption. We require the functions

$$u_0, \ldots, u_r, \qquad v_0 \cdot f, \ldots, v_s \cdot f \tag{2.4}$$

to be linearly independent on M. Under this condition there exist $n = r+s+2$ distinct points $x_1, \ldots, x_n \in M$ such that the functions $u_0, \ldots, u_r, v_0 \cdot f, \ldots, v_s \cdot f$ are linearly independent on

$$D = \{x_1, \ldots, x_n\}.$$

This means that the matrix

$$\tilde{A} = \begin{pmatrix} u_j(x_i) \\ v_k(x_i) f(x_i) \end{pmatrix}^{2} \tag{2.5}$$

is nonsingular. If we define the matrix B by

$$B = \begin{pmatrix} O \\ v_k(x_i) \end{pmatrix} \tag{2.6}$$

where O is a zero matrix consisting of $r+1$ rows and $r+s+2$ columns we can formulate

LEMMA 2.2. *We assume* $D = \{x_1, \ldots, x_n\} \subseteq M$, $n = r+s+2$, *to be such that the matrix* (2.5) *is nonsingular. Then for each vector* $y = (y_1, \ldots, y_n)^T \geqslant \theta_n$, $y \neq \theta_n$, *there exists exactly one vector* $c = (c_1, \ldots, c_n)^T \neq \theta_n$ *such that*

$$\tilde{A}c = By \tag{2.7}$$

and we have

$$q(y) = \min_{c_i \neq 0} \frac{y_i}{|c_i|} \leqslant \rho_D(f). \tag{2.8}$$

Remark. If we choose $y > \theta_n$ (that is $y_i > 0$ for $i = 1, \ldots, n$), we have $q(y) > 0$. Hence in this case we always get a positive lower bound of $\rho_M(f)$ by solving the linear system (2.7) and computing $q(y) > 0$. Furthermore, the assumption (2.4) yields $\rho_M(f) > 0$.

Proof. For each $y \in \boldsymbol{R}^n$, $y \geqslant \theta_n$, $y \neq \theta_n$ we have

$$By \neq \theta_n.$$

Otherwise, for each $v \in V_D^+$ we would have

$$\sum_{i=1}^{n} v(x_i) y_i = 0,$$

which is impossible. Since $\tilde{A}$ is nonsingular for each such $y \in \boldsymbol{R}^n$ there is exactly one solution $c \neq \theta_n$ of the linear system (2.7). We put $I = \{i : c_i = 0$ and $y_i = 0\}$ and define for each $i \notin I$

$$\lambda_i = \begin{cases} \dfrac{y_i}{|c_i|} & \text{if} \quad c_i \neq 0 \\[2ex] \dfrac{y_i}{p_i} & \text{if} \quad c_i = 0 \end{cases}$$

---

[2] $j = 0, \ldots, r$ and $k = 0, \ldots, s$ denote row indices and $i = 1, \ldots, r+s+2$ denotes column indices.

where $p_i > 0$ is at our disposal. If we put $p_i = 0$ for each i such that $c_i \neq 0$, the system (2.7) can be written as

$$\sum_{i \notin I} u_j(x_i)\, c_i = 0, \qquad j = 0, \ldots, r,$$
$$\sum_{i \notin I} f(x_i)\, v_k(x_i)\, c_i = \sum_{i \notin I} \lambda_i(|c_i| + p_i)\, v_k(x_i), \qquad k = 0, \ldots, s.$$

Hence by Lemma 2.1 we conclude

$$\min_{i \notin I} \lambda_i \leqslant \rho_D(f).$$

If for each $i \notin I$ such that $c_i = 0$ we choose $p_i > 0$ sufficiently small, we can achieve

$$q(y) = \min_{c_i \neq 0} \frac{y_i}{|c_i|} = \min_{i \notin I} \lambda_i,$$

which completes the proof.

THEOREM 2.1. *Under the assumption of Lemma* 2.2 *for the set* $D \subseteq M$ *we have*

$$\rho_D(f) = \mathit{max}_{y \in K}\, q(y),$$

where $q(y)$ *is defined by* (2.8) *and*

$$K = \{y \in \boldsymbol{R}^n : y \geqslant \theta_n, \qquad y \neq \theta_n\}. \tag{2.9}$$

Proof. We have to show that there exists $\hat{y} \in K$ such that

$$q(\hat{y}) = \rho_D(f).$$

Then the assertion follows by Lemma 2.2. By Satz 2 in [7] there exist $\hat{c} \in \boldsymbol{R}^n$, $\hat{c} \neq \theta_n$ and $p \in \boldsymbol{R}^n$, $p \geqslant \theta_n$ so that

$$\tilde{A}\hat{c} = \rho_D(f) \cdot B(|\hat{c}| + p)$$

where $\tilde{A}$ and B are given by (2.5) and (2.6). If we put $\hat{y} = \rho_D(f)(|\hat{c}| + p)$, then $\hat{y} \in K$, and by Lemma 2.2

$$q(\hat{y}) = \min_{\hat{c}_i \neq 0} \frac{\hat{y}_i}{|\hat{c}_i|} \leqslant \rho_D(f).$$

On the other hand, for each i such that $\hat{c}_i \neq 0$ we have

$$\frac{\hat{y}_i}{|\hat{c}_i|} = \rho_D(f) \frac{|\hat{c}_i| + p_i}{|\hat{c}_i|} \geqslant \rho_D(f),$$

whence $\rho_D(f) \leqslant q(\hat{y})$, which completes the proof.

In the setting of Theorem 2.1 the computation of $\rho_D(f)$ leads to the following nonlinear optimization problem (which is solvable): Under the conditions

$$|Ay| \leqslant \frac{1}{\lambda} y, \qquad y \geqslant \theta_n, \qquad y \neq \theta_n,$$

λ is to be maximized ($A = \tilde{A}^{-1} B$).

3. A Nonlinear Eigenvalue Problem

Let $D = \{x_1, \ldots, x_n\}$, $n = r + s + 2$, be a subset of M such that the matrix (2.5) is nonsingular. We consider the following problem: Find a number $\lambda > 0$ such that the system

$$\begin{aligned} &\sum_{i=1}^{n} u_j(x_i)\, c_i = 0, \qquad j = 0, \ldots, r, \\ &\sum_{i=1}^{n} v_k(x_i) f(x_i)\, c_i = \lambda \sum_{i=1}^{n} |c_i|\, v_k(x_i), \\ &\qquad\qquad k = 0, \ldots, s \end{aligned} \tag{3.1}$$

has a solution $c = (c_i, \ldots, c_n)^T \neq \theta_n$ For each such number λ we have, by Lemma 2.1,

$$\lambda \leqslant \rho_D(f).$$

If there is a best approximant of f in

$$W_D = \left\{\frac{u}{v}; u \in U, v \in {V_D}^{+}\right\} \subseteq C(D)$$

we know, [7], that for $\lambda = \rho_D(f)$ there is a nontrivial solution c of (3.1). Furthermore, we know by Satz 2.1 in [8] that there is a subset D of M such that for $\lambda = \rho_M(f)$ the system (3.1) admits a nontrivial solution c if the approximation problem in $C(M)$ is solvable. By the substitution

$$y = \lambda |c|, \qquad \mu = \frac{1}{\lambda} \tag{3.2}$$

the above problem turns out to be equivalent to the following nonlinear eigenvalue problem: Find a number $\mu > 0$ such that there is a solution $y \in K$ of

$$|Ay| = \mu y \tag{3.3}$$

where $A = \tilde{A}^{-1} B$, $\tilde{A}$ is given by (2.5), B by (2.6) and K by (2.9).

THEOREM 3.1. *There is a* $y \in K$ *and a* $\mu > 0$ *such that* (3.3) *holds, i.e., such that*

$$c = Ay \qquad and \qquad \lambda = \frac{1}{\mu}$$

solve problem (3.1).

Proof. (According to [9].) We define a mapping $P: \boldsymbol{R}^n \to \boldsymbol{R}^n$ by

$$P(y) = |Ay|, \qquad y \in \boldsymbol{R}^n. \tag{3.4}$$

In the proof of Lemma 2.2 we have shown

$$By \neq \theta_n \qquad \text{for each} \qquad y \in K.$$

Hence $P(y) = |Ay| = |\tilde{A}^{-1}By| \geqslant \theta_n$ and $\neq \theta_n$ for each $y \in K$; that is, $P(K) \subseteq K$. Evidently the subset

$$S = \left\{ y \in K : \|y\|_1 := \sum_{i=1}^{n} y_i = 1 \right\}$$

of K is convex and compact and we have

$$\|P(y)\|_1 := \sum_{i=1}^{n} P(y)_i > 0$$

for all $y \in S$. Hence the operator

$$\tilde{P}(y) = \frac{P(y)}{\|P(y)\|_1}$$

is defined on S, continuous and maps S into itself. By Brouwer's fixed-point theorem there is a $y^* \in S$ such that $\tilde{P}(y^*) = y^*$ or

$$P(y^*) = |Ay^*| = \mu^* y^*$$

where $\mu^* = \|P(y^*)\|_1 > 0$. This completes the proof.

For numerical purposes it would be very helpful if the operator P defined by (3.4) were monotone on $K \cup \{\theta_n\}$; that is,

$$\theta_n \leqslant y \leqslant z \qquad \text{implies} \qquad P(y) \leqslant P(z).$$

If we then start with an arbitrary $y^0 \in K$ and define a sequence $y^k \in K$ by

$$y^{k+1} = P(y^k), \qquad k = 0, 1, 2, \ldots,$$

it turns out that for the numbers

$$q_k = \min_{y_i^{k+1} \neq 0} \frac{y_i^k}{y_i^{k+1}} \qquad \text{and} \qquad \hat{q}_k = \max_{y_i^{k+1} \neq 0} \frac{y_i^k}{y_i^{k+1}}$$

we have

$$q_0 \leqslant q_1 \leqslant \ldots \leqslant q_k \leqslant \rho_D(f) \leqslant \hat{q}_k \leqslant \ldots \leqslant \hat{q}_1 \leqslant \hat{q}_0$$

and

$$\rho_D(f) = \lim_{k\to\infty} q_k = \lim_{k\to\infty} \hat{q}_k$$

if P is strictly monotone in the sense of Bohl [*1*].

If we define the matrix $|A|$ by taking the absolute values of the elements of A as elements of $|A|$, then P is obviously monotone on $K \cup \{\theta_n\}$ if we have

$$|Ay| = |A|y \qquad \text{for each} \qquad y \in K. \tag{3.5}$$

Sufficient for (3.5) is that in each row of A all the elements which are unequal to zero have the same sign. But as E. Bohl pointed out this is also necessary for P to be monotone on $K \cup \{\theta_n\}$. Bohl gave a simple (unpublished) proof for this fact, namely: Assume that for some index i and two indices j and k with $j \neq k$ we have $A_{ij} \neq 0$, $A_{ik} \neq 0$ and $\operatorname{sgn} A_{ij} = -\operatorname{sgn} A_{ik}$. Then we define two vectors y and $z \in \boldsymbol{R}^n$ by

$$y_l = \left\{\begin{array}{ll} 0 & \text{for} \quad l \neq j \text{ and } k, \\ \lambda & \text{for} \quad l = k, \\ -\dfrac{A_{ik}}{A_{ij}} & \text{for} \quad l = j, \end{array}\right\} \quad \text{and} \quad z_l = \left\{\begin{array}{ll} 0 & \text{for} \quad l \neq j \text{ and } k, \\ 1 & \text{for} \quad l = k, \\ -\dfrac{A_{ik}}{A_{ij}} & \text{for} \quad l = j, \end{array}\right\}$$

$$l = 1, \dots n.$$

If we choose $0 \leqslant \lambda < 1$, then $y, z \in K$ and

$$\theta_n \leqslant y \leqslant z.$$

However,

$$|(Ay)_i| = |-A_{ik} + \lambda A_{ik}| = |A_{ik}|(1 - \lambda),$$

$$|(Az)_i| = |-A_{ik} + A_{ik}| = 0,$$

and $|(Ay)_i| \leqslant |(Az)_i| = 0$ would imply $A_{ik} = 0$, a contradiction. The result is that the operator P defined by (3.4) is monotone if and only if the problem (3.3) is a linear eigenvalue problem of the form (3.5).

In general, P is not monotone as the following example shows: $M = [a,b]$, $r = 0$, $s = 1$, $u_0 = v_0 \equiv 1$, $v_1(x) = x$ and $f \in C(M)$ chosen such that for three different points $x_i \in [a,b]$, $i = 1, 2, 3$, we have $f(x_1) \neq 0$, $f(x_3) \neq 0$, $f(x_2) = 0$. Then the matrix $\tilde{A}$ given by (2.5) is nonsingular in this case; in fact, the determinant of $\tilde{A}$ is given by the formula

$$\det(\tilde{A}) = (x_1 - x_3) f(x_1) f(x_3)$$

and hence is not equal to zero. $A = \tilde{A}^{-1} B$, with B of (2.6), is given by

$$A = \begin{pmatrix} \dfrac{1}{f(x_1)} & \dfrac{x_2 - x_3}{(x_1 - x_3)f(x_1)} & 0 \\ -\dfrac{1}{f(x_1)} & \dfrac{(x_2 - x_1)f(x_1) + (x_3 - x_2)f(x_3)}{(x_1 - x_3)f(x_1)f(x_3)} & -\dfrac{1}{f(x_3)} \\ 0 & \dfrac{x_1 - x_2}{(x_1 - x_3)f(x_3)} & \dfrac{1}{f(x_3)} \end{pmatrix}.$$

Therefore P defined by (3.4) is monotone on $K \cup \{\theta_n\}$ if and only if

$$\operatorname{sgn} f(x_1) = \operatorname{sgn} f(x_3),$$

and

$$\operatorname{sgn}(x_1 - x_2) = \operatorname{sgn}(x_1 - x_3) = \operatorname{sgn}(x_2 - x_3).$$

Final remark: In [*6*] we have shown how the nonlinear eigenvalue problem (3.3) is related to the linear eigenvalue problem which has been investigated by Werner [*12*] in the case of the classical rational approximation problem.

References

1. E. Bohl, Eigenwertaufgaben bei monotonen Operatoren und Fehlerabschätzungen für Operatorgleichungen. *Arch. Rat. Mech. Anal.* **22** (1966), 313–332.

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

3. L. Collatz, Tschebyscheffsche Annäherung mit rationalen Funktionen. *Abhandl. Math. Sem. Univ. Hamburg* **24** (1960), 70–78.

4. L. Collatz, Inclusion theorems for the minimal distance in rational Tschebyscheff approximation with several variables. *In*: "Approximation of Functions," (H. L. Garabedian, Ed.). Elsevier, Amsterdam, 1965.

5. W. Krabs, Über ein Kriterium von Kolmogoroff bei der Approximation von Funktionen. To appear in *Internat. Ser. Numer. Math.*, Birkhäuser, Basel.

6. W. Krabs, Eine nichtlineare Eigenwertaufgabe bei rationaler Approximation. *Z. Angew. Math. Mech.* **T47** (1967) 57–60.

7. W. Krabs, Dualität bei diskreter rationaler Approximation. *Internat. Ser. Numer. Math.*, Birkhäuser, Basel. (1967), 33–41.

8. W. Krabs, Zur verallgemeinerten rationalen Approximation. *Math. Z.* **94** (1966), 84–97.

9. M. G. Krein and M. A. Rutman, Linear operators leaving invariant a cone in a Banach space. *Transl. Am. Math. Soc.* **10** (1962), 201.

10. G. Meinardus, "Approximation von Funktionen und ihre numerische Berhandlung." Springer, Berlin, 1964.

11. G. Meindarus and D. Schwedt, Nichtlineare Approximationen. *Arch. Rat. Mech. Anal.* **17** (1964), 297–326.

12. H. Werner, Rationale Tschebyscheff-Approximation, Eigenwerttheorie und Differenzenrechnung. *Arch. Rat. Mech. Anal.* **13** (1963), 330–347.