On the Constrained Chebyshev Approximation Problem on Ellipses

Roland Freund
Bernd Fischer

August 1988

Research Institute for Advanced Computer Science
NASA Ames Research Center

RIACS Technical Report 88.20

Research Institute for Advanced Computer Science

(NASA-CR-183209) ON THE CONSTRAINED CHEBYSHEV APPROXIMATION PROBLEM ON ELLIPSES (NASA) 20 p
CSCL 12A
N89-13200
Unclas
G3/64 0161500

On the Constrained Chebyshev Approximation Problem on Ellipses

*Roland Freund**
*Bernd Fischer***

Research Institute for Advanced Computer Science
NASA Ames Research Center

RIACS Technical Report 88.20
August 1988

This paper was prepared for submittal to *Journal of Approximation Theory*.

*Roland Freund is on leave from the Institut Angewandte Mathematik und Statistik, Universitat Wurzburg, D-8700 Wurzburg, F.R.G.

**Bernd Fischer is located at the Institut fur Angewandte Mathematik, Universitat Hamburg, D-2000 Hamburg 13, F.R.G.

Work reported herein was supported in part by Cooperative Agreement NCC 2-387 between the National Aeronautics and Space Administration (NASA) and the Universities Space Research Association (USRA).

Abstract

We are concerned with constrained Chebyshev approximation problems of the type

$$\min_{p\in\Pi_n:p(c)=1} \max_{z\in\mathcal{E}} |p(z)| \quad . \tag{P}$$

Here Π_n denotes the set of all complex polynomials of degree at most n, $\mathcal{E}$ is any ellipse in the complex plane, and $c \in \mathbb{C} \setminus \mathcal{E}$. Such approximation problems arise in the context of optimizing semi-iterative methods for the solution of large, sparse systems of linear equations $Ax = b$ with complex non-Hermitian coefficient matrices A. The problem of obtaining optimal ploynomial preconditioners for conjugate gradient type methods for $Ax = b$ also leads to problems of type (P). In this paper, we introduce a new family of polynomials $q_n(z; c)$, $n \in \mathbb{N}$, $c \in \mathbb{C}\setminus \mathcal{E}$, as the polynomials which are optimal for a modified version of (P) with Π_n replaced by a certain subfamily. Some simple properties of q_n are also listed. We then derive a necessary and sufficient condition for q_n to be the extremal polynomial of the approximation problem (P). Finally, it is shown that q_n is indeed optimal for (P) for all fixed n whenever the distance between c and $\mathcal{E}$ is sufficiently large. Results of some numerical tests are presented.

1. Introduction

In this paper, we are concerned with constrained Chebyshev approximation problems of the type

$$(D_n(r,c) :=) \quad \min_{p\in\Pi_n:p(c)=1} \max_{z\in\mathcal{E}_r} |p(z)| \quad . \tag{1}$$

Here Π_n denotes the set of all complex polynomials of degree at most n,

$$\mathcal{E}_r := \{z \in \mathbb{C} \quad | \quad |z-1| + |z+1| \le r + \frac{1}{r}\} \quad , \quad r \ge 1 \quad , \tag{2}$$

is any ellipse (including its interior) in the complex plane with foci at ± 1, and it is always assumed that is any ellipse (including its interior) in the complex plane $c \in \mathbb{C} \setminus \mathcal{E}_r$. Since all polynomials $p \in \Pi_n$ with $p(c) = 1$ can be parametrized in the form

$$p(z) = 1 - (\gamma_1(z-c) + \gamma_2(z-c)^2 + \cdots + \gamma_n(z-c)^n) \quad , \quad \gamma_1, \gamma_2, \ldots, \gamma_n \in \mathbb{C} \quad , \tag{3}$$

the condition $c \in \mathbb{C} \setminus \mathcal{E}_r$ guarantees that Haar's condition is satisfied. Thus, there always exists a unique optimal polynomial for (1) which will be denoted by $p_n(z; r, c)$ in the sequel. However, these extremal polynomials are explicitly known only for special cases. The solution of (1) is classical for real c:

$$p_n(z;\, r, c) \quad = \quad \frac{T_n(z)}{T_n(c)} \quad , \quad c \in \mathbb{R} \setminus \mathcal{E}_r \quad , \tag{4}$$

where T_n is the nth Chebyshev polynomial (of the first kind).

Constrained approximation problems (1) with complex c arise in the context of optimizing semi-iterative methods for the solution of non-Hermitian systems of linear equations (e.g. Manteuffel [4] and Eiermann, Niethammer, and Varga [1]). Mainly motivated by this application, in some recent papers, problem (1) was studied for complex c and the optimal polynomials were found for certain special cases. For $n = 1$, Opfer and Schober [6] obtained a complete solution of a more general version of (1) with $\mathcal{E} \subset \mathbb{C}$ any compact set not containing c. For ellipses, their result can be rewritten in the following form:

$$p_1(z;\, r, c) \quad = \quad \frac{Bz + i\ \sin\gamma}{A\,(B\cos\gamma + i\ A\sin\gamma)} \quad , \tag{5}$$

where

$$c \quad = \quad A\cos\gamma + i\ B\sin\gamma \quad (\in \partial\mathcal{E}_R) \tag{6}$$

with $0 \leq \gamma < 2\pi$ and

$$A \;=\; \frac{1}{2}\left(R + \frac{1}{R}\right) \quad , \quad B \;=\; \frac{1}{2}\left(R - \frac{1}{R}\right) \quad , \quad R > r \geq 1 \quad ,$$

(by $\partial\mathcal{E}_R$ we denote the boundary of $\mathcal{E}_R$). Freund and Ruscheweyh [3] investigated (1) for the case $r = 1$ of the line segment $\mathcal{E}_1 = [-1, 1]$. They determined $p_2(z;\, 1, c)$ for arbitrary c and $p_n(z;\, 1, c)$ for $n \in \mathbb{N}$ and purely imaginary c. In both cases, the optimal polynomials are suitable linear combinations of T_n, T_{n-1}, and T_{n-2}. Finally, Fischer [2] showed that for nondegenerate ellipses $\mathcal{E}_r$, $r > 1$, and purely imaginary c the normalized Chebyshev polynomial (4) is optimal for (1), if n is even and $|c|$ is sufficiently large compared to r.

Note that, except for the cases solved in [3], all the other explicitly known optimal polynomials are of the form

$$q(z) \quad = \quad \frac{T_n(z) + \alpha}{T_n(c) + \alpha} \quad , \quad \alpha \in \mathbb{C} \quad . \tag{7}$$

It is thus natural to ask, whether polynomials of type (7) lead to explicit solutions of (1) also for the case of general complex c and $n \in \mathbb{N}$. The purpose of this note is to answer this question.

The paper is organized as follows. In Sect. 2, we introduce a new family of polynomials $q_n(z;\, c)$, $n \in \mathbb{N}$, $c \in \mathbb{C} \setminus \mathcal{E}_r$, as the polynomials of the form (7) with minimal uniform norm on $\mathcal{E}_r$. Some simple properties of q_n are also listed. In Sect. 3, we derive a necessary and sufficient condition for q_n to be the extremal polynomial of the approximation problem (1). Finally, Sect. 4 contains the main result of this paper. We show that indeed $p_n(z;\, r, c) = q_n(z; c)$ for all fixed $n \in \mathbb{N}$, $r > 1$ and all $c \in \mathbb{C}$ whose parameter R in the representation (6) is sufficiently large, i.e. $R \geq R_0(n,r)$. An explicit formula for $R_0(n,r)$ is given.

2. A Class of Extremal Polynomials

Throughout this paper, let $n \in \mathbb{N}$, $r \geq 1$, $\mathcal{E}_r$ be the ellipse defined in (2), and it is assumed that $c \in \mathbb{C} \setminus \mathcal{E}_r$ with representation (6). We will make use of the parametrization

$$z_r(\phi) \;=\; a\, \cos\phi \;+\; i\, b\, \sin\phi \quad , \quad \phi \in \mathbb{R} \quad ,$$

of the boundary $\partial\mathcal{E}_r$ of $\mathcal{E}_r$. Here $a := a_1$, $b := b_1$, where

$$a_k := \frac{1}{2}(r^k + \frac{1}{r^k}) \quad \text{and} \quad b_k := \frac{1}{2}(r^k - \frac{1}{r^k}) \quad , \quad k = 1, 2, \ldots \quad . \tag{8}$$

$T_k(z)$ denotes the kth Chebyshev polynomial which by means of the Joukowsky map is given by

$$T_k(z) = \frac{1}{2}(v^k + \frac{1}{v^k}) \quad , \quad z = \frac{1}{2}(v + \frac{1}{v}) \quad . \tag{9}$$

By (6) and (9), one has

$$c_k := T_k(c) \;=\; A_k \cos(k\gamma) + i\; B_k\; \sin(k\gamma) \quad , \quad k = 1, 2, \ldots \quad , \tag{10}$$

where

$$A_k := \frac{1}{2}(R^k + \frac{1}{R^k}) \quad \text{and} \quad B_k := \frac{1}{2}(R^k - \frac{1}{R^k}) \quad . \tag{11}$$

The relations

$$A_k^2 - B_k^2 = 1 \quad , \quad a_k^2 - b_k^2 = 1 \quad , \quad k = 1, 2, \ldots \; ,$$

will be used repeatedly in the sequel. Moreover, note that, since $R > r$,

$$A_k > a_k \quad , \quad B_k > b_k \quad , \quad k = 1, 2, \ldots \quad .$$

We consider the extremal problem

$$(M_n(r,c) :=) \quad \min_{\alpha \in \mathbb{C}} \quad \max_{z \in \mathcal{E}_r} \quad \left| \frac{T_n(z) + \alpha}{T_n(c) + \alpha} \right| \quad . \tag{12}$$

Since $w = T_n(z)$ maps $\mathcal{E}_r$ onto $\mathcal{E}_{r^n}$, (12) is equivalent to

$$\min_{p\in\Pi_1 : p(c_n)=1} \quad \max_{w\in\mathcal{E}_{r^n}} \quad |p(w)| \quad .$$

Thus, by (5) and (10) (for $k = n$),

$$q_n(z;c) := p_1(T_n(z); r^n, c_n) \; = \; \frac{B_n T_n(z) + i \;\sin(n\gamma)}{A_n(B_n \;\cos(n\gamma) + i \; A_n \;\sin(n\gamma))} \tag{13}$$

is the unique extremal polynomial of (12). Next, we determine $M_n(r,c)$ and the corresponding extremal points, i.e. $z \in \mathcal{E}_r$ with

$$|q_n(z;c)| \; = \; M_n(r,c) \quad .$$

From the maximum modulus principle it follows that all such points lie on $\partial\mathcal{E}_r$. By (9) and (8) (both for $k = n$), one has

$$T_n(z_r(\phi)) = a_n \cos(n\phi) + i \; b_n \sin(n\phi) \quad .$$

Using this identity, we deduce from (13) the relation

$$|q_n(z_r(\phi);c)|^2 = \frac{a_n^2}{A_n^2}\left(1 - \frac{(B_n \;\sin(n\phi) - b_n \;\sin(n\gamma))^2}{a_n^2(B_n^2 + \sin^2(n\gamma))}\right) \quad , \quad \phi \in I\!R \, . \tag{14}$$

Therefore $M_n(r,c) = a_n \,/\, A_n$, and the extremal points are just the $z_r(\phi)$ with ϕ satisfying

$$B_n \;\sin(n\phi) \quad = \quad b_n \;\sin(n\gamma) \quad . \tag{15}$$

We set

$$d_n \quad := \quad \frac{b_n}{B_n}\sin(n\gamma) \tag{16}$$

and define ψ_n by

$$\sin\psi_n \quad = \quad d_n \quad , \quad -\frac{\pi}{2} \; < \; \psi_n \; < \; \frac{\pi}{2} \quad . \tag{17}$$

Note that

$$|d_n| \; \le \; \frac{b_n}{B_n} \; < \; 1 \quad . \tag{18}$$

All solutions (mod 2π) of (15) are then given by

$$\phi_l = \frac{l}{n}\pi \; + \; (-1)^l \frac{\psi_n}{n} \quad , \quad l = 1, 2, \ldots, 2n \quad .$$

Remark that for $r > 1$ (resp. $r = 1$) this leads to precisely $2n$ (resp. $n + 1$) distinct extremal points of q_n on $\partial\mathcal{E}_r$. We summarize these results in the following

Theorem 1. *$q_n(z;c)$ is the unique extremal polynomial of (12), and the corresponding minimal norm is*

$$M_n(r,c) \quad = \quad \frac{r^n + 1\,/\,r^n}{R^n + 1\,/\,R^n} \quad .$$

On $\mathcal{E}_r$, $r > 1$, $q_n(z;c)$ has precisely $2n$ extremal points:

$$z_l = \frac{1}{2}(r + \frac{1}{r})\ \cos\phi_l + \frac{i}{2}(r - \frac{1}{r})\sin\phi_l \quad , \quad \phi_l = \frac{l}{n}\pi + (-1)^l\frac{\psi_n}{n} \quad , \quad l = 1, 2, \ldots, 2n \ .$$

The extremal points of $q_n(z;c)$ on $\mathcal{E}_1 = [-1,1]$ are

$$z_l \ = \ \cos\frac{l\pi}{n} \quad , \quad l = 0, 1, \ldots, n \quad .$$

Remark 1. The optimal polynomial of (12) is identical for all $\mathcal{E}_r$, $1 \le r < R$. $M_n(r,c)$ depends only on the parameter R of $\mathcal{E}_R$, but not on the position of c on $\partial\mathcal{E}_R$.

The family of polynomials $q_n(z;c)$ also leads to upper and lower bounds for the minimal deviation $D_n(r,c)$ of (1).

Theorem 2. *Let $r \ge 1$, $c \in \partial\mathcal{E}_R$, $R > r$. Then,*

$$D_n(r,c) \ \le \ \frac{a_n}{A_n} \ = \ \frac{r^n + 1\,/\,r^n}{R^n + 1\,/\,R^n} \quad , \quad n = 1, 2, \ldots \quad , \tag{19}$$

and

$$D_n(r,c) \ \ge \ \frac{a_n}{A_n}\sqrt{1 - \frac{(B_n + b_n\ |\sin(n\gamma)|)^2}{a_n^2(B_n^2 + \sin^2(n\gamma))}} \tag{20}$$

for all n satisfying

$$|\sin(n\gamma)| \quad \le \quad b_n B_n \quad . \tag{21}$$

Remark 2. Clearly, (21) is true if n is sufficiently large.

Proof. (19) is an immediate consequence of Theorem 1. A standard technique (Trefethen [8], Manteuffel [4]) to obtain lower bounds for complex approximation problems is based on Rouché's theorem. Applied to (1) and q_n, this yields

$$D_n(r,c) \quad \ge \quad \min_{z\in\partial\mathcal{E}_r} \ |q_n(z;c)| \quad , \tag{22}$$

if it is guaranteed that all zeros of q_n are contained in $\mathcal{E}_r$. In view of (14), the right-hand side of (22) is just the bound stated in (20). By (13), the zeros of q_n are the solutions of the equation

$$T_n(z) \ = \ -i\ \frac{\sin(n\gamma)}{B_n} \quad .$$

Using (9) (for $k = n$), one easily verifies that all these solutions lie on the boundary $\partial\mathcal{E}_\rho$ of an ellipse of type (2) whose parameter $\rho \geq 1$ is defined by

$$\beta_n := \frac{1}{2}(\rho^n - \frac{1}{\rho^n}) = \frac{|\sin(n\gamma)|}{B_n}$$

Therefore, $\partial\mathcal{E}_\rho$ (and hence the zeros of q_n) is contained in $\mathcal{E}_r$ iff $\beta_n \leq b_n$. This concludes the proof of Theorem 1. ∎

3. A Criterion for Optimality

As mentioned in the introduction, it is known that

$$p_n(z; r, c) = q_n(z; c) \quad , \quad z \in \mathbb{C} \quad , \tag{23}$$

for some special cases as $n = 1$ or $c \in \mathbb{R} \setminus \mathcal{E}_r$. In this section, we present a necessary and sufficient condition for (23) for the general case $n \in \mathbb{N}$, $c \in \mathbb{C} \setminus \mathcal{E}_r$. This criterion allows to check (23) by computing $2n$ real numbers for which explicit formulas are derived.

First, consider the case $r = 1$ of the degenerate ellipse $\mathcal{E}_1 = [-1, 1]$. It was shown in [3] that $p_n(z; 1, c)$ has precisely $n + 1$ extremal points

$$1 = z_0 > z_1 > \cdots > z_n = -1$$

and there is a $s_n \in \mathbb{C}$ such that

$$p_n(z_l; 1, c) = s_n(-1)^l \, \frac{z_l - c}{|z_l - c|} \quad , \quad l = 0, 1, \ldots, n \quad . \tag{24}$$

By Theorem 1 and (13) (with $z = z_l$), $q_n(z; c)$ has the extremal points

$$z_l = \cos\frac{l\pi}{n} \quad \text{and} \quad q_n(z_l; c) = t_n((-1)^l + i\,\frac{\sin(n\gamma)}{B_n}) \quad , \quad l = 0, 1, \ldots, n \quad , \tag{25}$$

for some $t_n \in \mathbb{C}$. By comparing (24) and (25), it is straightforward to verify that, for $r = 1$, (23) holds iff $n = 1$ or $c \in \mathbb{R} \setminus \mathcal{E}_r$. So, except for the already known cases, $q_n(z; c)$ is not optimal for (1) with $r = 1$.

Therefore, for the rest of this paper, we assume that $r > 1$. By Theorem 1, the extremal points of $q_n(z; c)$ on $\mathcal{E}_r$ are

$$z_l := a\,\cos\phi_l + i\,b\,\sin\phi_l \quad , \quad \phi_l := \frac{l}{n}\pi + (-1)^l\frac{\psi_n}{n} \quad , \quad l = 1, 2, \ldots, 2n \quad , \tag{26}$$

with ψ_n defined by (17) and (16). We list some properties of the points (26), which will be needed for the derivation of the main result of this section, in the following

Lemma 1. *a) For $l = 1, 2, \ldots, 2n$:*

$$\sin(n\phi_l) \;=\; d_n \quad , \quad \cos(n\phi_l) \;=\; (-1)^l\sqrt{1-d_n^2} \quad , \tag{27}$$

and

$$q_n(z_l;c) \;=\; t_n((-1)^l + i\,\frac{a_n\,d_n}{b_n\sqrt{1-d_n^2}}) \tag{28}$$

where

$$t_n \;=\; \frac{a_n\,B_n\sqrt{1-d_n^2}}{A_n(B_n\cos(n\gamma)+i\,A_n\,\sin(n\gamma))} \quad .$$

b) For $j = 0, 1, \ldots, 2n$:

$$\sum_{l=1}^{2n} e^{ij\phi_l} = 2n \times \begin{cases} 1 & \text{if} \quad j=0 \\ i\,d_n & \text{if} \quad j=n \\ 1-2d_n^2 & \text{if} \quad j=2n \\ 0 & \text{otherwise} \end{cases} \tag{29a}$$

and

$$\sum_{l=1}^{2n}(-1)^l e^{ij\phi_l} = 2n\sqrt{1-d_n^2} \times \begin{cases} 1 & \text{if} \quad j=n \\ 2id_n & \text{if} \quad j=2n \\ 0 & \text{otherwise} \end{cases} \quad . \tag{29b}$$

Proof. (27) follows immediately from (17) and the definition of ϕ_l in (26). (28) is obtained from (13) (with $z = z_l$) by using (9) (for $k = n$), (16), and (27).

We now turn to the proof of part b). Recall that

$$\sum_{k=1}^{n}(e^{\frac{2\pi ij}{n}})^k \;=\; \begin{cases} n & \text{if} \quad j \in n\mathbb{Z} \\ 0 & \text{if} \quad j \notin n\mathbb{Z} \end{cases} \quad . \tag{30}$$

Let $0 \le j \le 2n$ and $\delta = \pm 1$. Since, by (26),

$$\phi_l \;=\; \begin{cases} \dfrac{2k\pi}{n} + \dfrac{\psi_n}{n} & \text{if} \quad l = 2k \\ \dfrac{(2k-1)\pi}{n} - \dfrac{\psi_n}{n} & \text{if} \quad l = 2k-1 \end{cases}$$

and with (30), we get

$$\sum_{l=1}^{2n} \delta^l e^{ij\phi_l} = e^{ij\frac{\psi_n}{n}}\sum_{k=1}^{n} e^{\frac{2\pi ij}{n}k} + \delta\; e^{-ij\frac{\psi_n}{n}}\sum_{k=1}^{n} e^{\frac{\pi ij}{n}(2k-1)}$$

$$= (e^{ij\frac{\psi_n}{n}} + \delta e^{\frac{-\pi ij}{n}} e^{-ij\frac{\psi_n}{n}}) \sum_{k=1}^{n} e^{\frac{2\pi ij}{n}k} = n \times \begin{cases} 1+\delta & \text{if} \quad j=0 \\ e^{i\psi_n} - \delta e^{-i\psi_n} & \text{if} \quad j=n \\ e^{2i\psi_n} + \delta e^{-2i\psi_n} & \text{if} \quad j=2n \\ 0 & \text{otherwise} \end{cases} .$$

Using (17), one easily verifies that these are just the formulas (29a) ($\delta = 1$) and (29b) ($\delta = -1$). ■

In view of (3), (1) is a linear Chebyshev approximation problem: We seek the best uniform approximation to $f(z) \equiv 1$ on $\mathcal{E}_r$ out of all functions of the linear space

$$\Pi_n(c) \; := \; \{p \in \Pi_n \mid p(c) = 0\} \quad .$$

Therefore, the characterization of best approximations due to Rivlin and Shapiro [7] can be applied. The following criterion results:

Criterion 1: *$q_n(z;c)$ is the optimal polynomial for (1) iff there exist nonnegative real numbers $\sigma_1, \sigma_2, \ldots, \sigma_{2n}$ (not all zero) such that*

$$\sum_{l=1}^{2n} \sigma_l \, \overline{q_n(z_l;c)} \, p(z_l) = 0 \quad \textit{for all} \quad p \in \Pi_n(c) \quad . \tag{31}$$

We now determine all real $\sigma_1, \ldots, \sigma_{2n}$ which fulfill (31). Note that $q_n(z_l;c)$ is given explicitly in (28). Furthermore, $\Pi_n(c)$ is spanned by the polynomials

$$T_k(z) - c_k \quad , \quad k = 1, 2, \ldots, n \quad ,$$

and, by (26), (9),

$$T_k(z_l) = \frac{1}{2}(r^k e^{ik\phi_l} + \frac{1}{r^k} e^{-ik\phi_l}) \quad .$$

Thus, (31) can be rewritten in the form

$$\sum_{l=1}^{2n} \sigma_l((-1)^l - ie_n)(r^k e^{ik\phi_l} + r^{-k} e^{-ik\phi_l} - 2c_k) = 0 \quad , \quad k = 1, 2, \ldots, n \; , \tag{31'}$$

where

$$e_n \; := \; \frac{a_n \, d_n}{b_n \, \sqrt{1 - d_n^2}} \quad . \tag{32}$$

Next, we remark that any numbers $\sigma_1, \ldots, \sigma_{2n} \in \mathbb{R}$ admit a representation of the type

$$\begin{aligned} \sigma_l &= \sum_{j=0}^{n} (\lambda_j \cos(j\phi_l) + \mu_j \sin(j\phi_l)) \\ &= \sum_{j=0}^{n} (\nu_j e^{ij\phi_l} + \overline{\nu_j} e^{-ij\phi_l}) \quad , \quad l = 1, 2, \ldots, 2n \quad , \end{aligned} \tag{33}$$

with real numbers λ_j, μ_j, $j = 0, \ldots, n$, $\mu_0 := 0$, and

$$\nu_j \ := \ \frac{\lambda_j - i\mu_j}{2} \quad . \tag{34}$$

This follows from the fact that the linear space spanned by

$$1 \ , \ \cos\phi \ , \ \cos(2\phi) \ , \ \ldots \ , \ \cos(n\phi) \ , \ \sin\phi \ , \ \sin(2\phi) \ , \ \ldots \ , \ \sin(n\phi)$$

satisfies Haar's condition on any interval of the form $[\alpha, \alpha+2\pi), \alpha \in \mathbb{R}$, and since, by (26) and (17), the numbers ϕ_l, $l = 1, \ldots, 2n$, are distinct and all contained in such an interval. By (33), (31′) leads to a system of equations for $\nu_0, \nu_1, \ldots, \nu_n$:

$$\sum_{j=0}^{n}\sum_{l=1}^{2n}((-1)^l - ie_n)(\nu_j e^{ij\phi_l} + \overline{\nu_j} e^{-ij\phi_l})(r^k e^{ik\phi_l} + r^{-k} e^{-ik\phi_l} - 2c_k) = 0 \ ,$$
$$k = 1, 2, \ldots, n \ . \tag{31''}$$

A routine calculation, making use of (29a,b), (32), and (34), shows that (31″) reduces to

$$\nu_{n-k} r^k (b_n + \frac{d_n^2}{r^n}) + \overline{\nu_{n-k}} \frac{1}{r^k}(b_n - r^n d_n^2) - i a_n d_n (\frac{\nu_k}{r^k} + \overline{\nu_k} r^k)$$
$$= 2c_k(\lambda_n b_n (1 - d_n^2) - i a_n d_n (\lambda_0 + d_n \mu_n)) \quad , \quad k = 1, 2, \ldots, n-1 \quad , \tag{35a}$$

and, for $k = n$, to

$$a_n(b_n + ic_n d_n)(\lambda_0 + d_n\mu_n) - (1 - d_n^2)(b_n c_n + id_n)\lambda_n = 0 \quad . \tag{35b}$$

Note that λ_0 and μ_n only occur in the combination

$$\tau \ := \ a_n(\lambda_0 + d_n\mu_n) \quad ; \tag{36}$$

moreover, we set

$$\lambda \ := \ (1 - d_n^2)\lambda_n \quad . \tag{37}$$

By taking its real and imaginary part, respectively, each of the complex equations (35) yields two real equations. Using (34), (8), (36), and (37), we thus arrive at

$$\begin{aligned} &(a_k b_n - b_{n-k} d_n^2)\lambda_{n-k} + a_n b_k d_n \mu_k = 2(Re\, c_k) b_n \lambda + 2(Im\, c_k) d_n \tau \quad , \\ &a_k a_n d_n \lambda_k + (b_k b_n + a_{n-k} d_n^2)\mu_{n-k} = 2(Re\, c_k) d_n \tau - 2(Im\, c_k) b_n \lambda \quad , \end{aligned} \tag{35'a}$$

for $k = 1, \ldots, n-1$, and

$$\begin{aligned} (b_n - (Im\, c_n)d_n)\tau \quad &- \quad (Re\, c_n)b_n\lambda \; = \; 0 \quad , \\ (Re\, c_n)d_n\tau \quad &- \quad (b_n(Im\, c_n) + d_n)\lambda \; = \; 0 \quad . \end{aligned} \tag{35'b}$$

With (16) and (10) (for $k = n$), the two equations of (35′b) can be written as

$$\cos(n\gamma)(\tau\cos(n\gamma) - \lambda A_n) = 0 \quad , \quad \sin(n\gamma)(\tau\cos(n\gamma) - \lambda A_n) = 0 \quad .$$

Therefore, the 2×2 system (35′b) is of rank 1 and its solutions are described by

$$\lambda \; = \; \frac{\cos(n\gamma)}{A_n}\tau \quad , \quad \tau \in I\!R \quad . \tag{38}$$

Now assume that $\tau \in I\!R$ is arbitrary, but fixed, and let λ be defined by (38). It remains to solve the system (35′a) of $2(n-1)$ linear equations for the $2(n-1)$ unknowns λ_k and μ_k, $k = 1, \ldots, n-1$. First, we note that, by combining the first equation of (35′a) with the second one of (35′a) (with k replaced by $n-k$), the system (35′a) is equivalent to the $n-1$ decoupled 2×2 systems

$$C_k \begin{pmatrix} \lambda_{n-k} \\ \mu_k \end{pmatrix} \; = \; 2b_n\tau \begin{pmatrix} f_k \\ g_k \end{pmatrix} \quad , \quad k = 1, \ldots, n-1 \; , \tag{39}$$

where

$$C_k = \begin{pmatrix} a_k b_n - b_{n-k}d_n^2 & a_n b_k d_n \\ a_{n-k}a_n d_n & b_{n-k}b_n + a_k d_n^2 \end{pmatrix}$$

and

$$f_k = (Re\, c_k)\frac{\cos(n\gamma)}{A_n} + (Im\, c_k)\frac{\sin(n\gamma)}{B_n} \; , \; g_k = (Re\, c_{n-k})\frac{\sin(n\gamma)}{B_n} - (Im\, c_{n-k})\frac{\cos(n\gamma)}{A_n} \quad . \tag{40}$$

Here, the formulas (40) were obtained by using (38) and (16). With (8), it is easily verified that

$$det\; C_k \; = \; a_k b_{n-k}(b_n^2 + d_n^2)(1 - d_n^2) \quad . \tag{41}$$

Thus, in view of (18), all matrices C_k, $k = 1, \ldots, n-1$, are nonsingular, and by Cramer's rule we deduce from (39) and (41) that

$$\lambda_k \; = \; \tau\lambda_k(1) \quad , \quad \mu_k \; = \; \tau\mu_k(1) \quad , \tag{42}$$

where

$$\lambda_k(1) = \frac{2b_n}{(b_n^2 + d_n^2)(1 - d_n^2)} \left((\frac{b_n}{a_{n-k}} + \frac{d_n^2}{b_k})f_{n-k} - \frac{a_n d_n \; b_{n-k}}{b_k \; a_{n-k}} g_{n-k} \right) \tag{43a}$$

and

$$\mu_k(1) = \frac{2b_n}{(b_n^2+d_n^2)(1-d_n^2)}\left((\frac{b_n}{b_{n-k}} - \frac{d_n^2}{a_k})g_k \;-\; \frac{a_n d_n}{a_k}\frac{a_{n-k}}{b_{n-k}} f_k \right) \quad , \tag{43b}$$

$k = 1, \ldots, n-1$. Finally, note that, by (27), (36), (37), and (38), summing up of the first ($j = 0$) and the last ($j = n$) term in (33) yields

$$\lambda_0 + \lambda_n \cos(n\phi_l) + \mu_n \sin(n\phi_l) = \tau(\frac{1}{a_n} + \frac{(-1)^l}{A_n}\;\frac{\cos(n\gamma)}{\sqrt{1-d_n^2}}) \quad , \; l = 1, \ldots, 2n \; . \tag{44}$$

Summarizing, we have proved that the set of all solutions $\sigma = (\sigma_1, \sigma_2, \ldots, \sigma_{2n})^T \in I\!R^{2n}$ of (31) is given by the one dimensional linear space

$$\sigma \;=\; \frac{\tau}{a_n}\sigma^\star \quad , \quad \tau \in I\!R \quad ,$$

where, by (33), (42), and (44),

$$\sigma_l^\star := 1 + (-1)^l \frac{\cos(n\gamma)}{\sqrt{1-d_n^2}}\;\frac{a_n}{A_n} + a_n \sum_{k=1}^{n-1} (\lambda_k(1)\cos(k\phi_l) + \mu_k(1)\sin(k\phi_l)) \; ,$$
$$l = 1, 2, \ldots, 2n \; . \tag{45}$$

Hence, Criterion 1 can be restated as follows.

Theorem 3. *Let $n \in I\!N$, $r > 1$, $c \in \mathbb{C} \setminus \mathcal{E}_r$. Then, the polynomial (13) $q_n(z;c)$ is optimal for (1) iff the numbers (45) $\sigma_l^\star$, $l = 1, 2, \ldots, 2n$, are either all nonnegative or all nonpositive.*

Remarks 3. For given n, r, $c \in \partial\mathcal{E}_R$, $R > r$, the numbers $\sigma_l^\star$, $l = 1, \ldots, 2n$, can easily be computed numerically by means of the formulas (6), (8), (11), (16), (26), and (43). We have done that in a number of cases. These numerical tests indicated that the polynomials $q_n(z;c)$ are indeed optimal for (1) whenever R (for fixed r, n) resp. n (for fixed r, R) is sufficiently large. We were not able to characterize explicitly all n, r, R for which q_n is optimal. However, in the next section, a necessary condition for the optimality of q_n is derived.

4. For the simplest case $n = 1$, the sum in (45) does not occur. It is easily verified that $R > r$ guarantees $\sigma_l^\star \geq 0$, $l = 1, 2$, and thus we have reobtained the result of Opfer and Schober [6] for the case $n = 1$.

5. It follows from Meinardus's invariance theorem [5, Theorem 27] that the extremal polynomials of (1) corresponding to c and its reflections $\bar{c}$ resp. $-\bar{c}$ on the real resp. imaginary axis are connected through

$$p_n(z;r,\bar{c}) \;=\; \overline{p_n(\bar{z};r,c)} \quad \text{resp.} \quad p_n(z;r,-\bar{c}) \;=\; \overline{p_n(-\bar{z};r,c)} \quad , \quad z \in \mathbb{C} \; .$$

This symmetry is also reflected in the following relations for the numbers (45). For fixed n and r, we consider $\sigma_l^\star = \sigma_l^\star(c)$ as a function of c. Then,

$$\sigma_l^\star(\bar{c}) = \sigma_{2n-l}^\star(c) \quad , \quad l = 0, 1, \ldots, 2n \quad ,$$

and

$$\sigma_l^\star(-\bar{c}) = \begin{cases} \sigma_{n-l}^\star(c) \quad , & l = 0, 1, \ldots, n \\ \sigma_{3n-l}^\star(c) \quad , & l = n+1, \ldots, 2n \end{cases} \quad ,$$

where $\sigma_0^\star := \sigma_{2n}^\star$. These identities can be verified by a routine calculation using the definition of $\sigma_l^\star$.

4. Optimal Polynomials for the Constrained Chebyshev Problem

In this section, we present a simple inequality involving n, r, R which guarantees the optimality of q_n for (1). For that purpose, a lower bound for the numbers (45) is derived which finally leads to a necessary condition for the nonnegativity of $\sigma_l^\star$, $l = 1, 2, \ldots, 2n$.

Throughout this section, it is assumed that $n \geq 2$, $R > r > 1$, and that $c \in \partial\mathcal{E}_R$ is represented in the form (6). Moreover, we recall the definitions of a_k, b_k (in (8)), A_k, B_k (in (11)), d_n (in (16)), and f_k, g_k (in (40)). In the following lemma, some estimates, which will be used in the sequel, for these numbers are listed.

Lemma 2. *a) For $k = 1, 2, \ldots, n-1$:*

$$|f_k| \leq A_k \frac{B_n(1-d_n^2)}{A_n^2 - a_n^2} \quad , \quad |g_k| \leq A_{n-k} \frac{B_n(1-d_n^2)}{A_n^2 - a_n^2} \quad .$$

b)

$$\begin{aligned} \sum_{k=1}^{n-1} \left(\frac{1}{a_k} + \frac{1}{b_k}\right) A_k &< \frac{4r^5}{(r^4-1)(R-r)} \left(\frac{R}{r}\right)^n \quad , \\ \sum_{k=1}^{n-1} \left(\frac{1}{a_k} + \frac{1}{b_k}\right) A_{n-k} &< \frac{4r^4}{(r^4-1)(Rr-1)} R^n \quad , \\ \sum_{k=1}^{n-1} \left(\frac{a_{n-k}}{a_k b_{n-k}} + \frac{b_{n-k}}{a_{n-k} b_k}\right) A_k &< \frac{2r(2r^2+1)}{(r^2-1)(R-r)} \left(\frac{R}{r}\right)^n \quad . \end{aligned} \qquad (46)$$

Proof. a) By Cauchy's inequality, it follows from (40) that

$$|f_k| \leq |c_k| \sqrt{g(x)} \quad , \quad |g_k| \leq |c_{n-k}| \sqrt{g(x)} \quad ,$$

where

$$g(x) := \frac{B_n^2 + x}{A_n^2 B_n^2} = \frac{\cos^2(n\gamma)}{A_n^2} + \frac{\sin^2(n\gamma)}{B_n^2} \quad , \quad x := \sin^2(n\gamma) \quad .$$

From (10), we obtain $|c_k| \le A_k$, $k = 1, 2, \ldots$, and hence it remains to show that

$$\sqrt{g(x)} \quad \le \quad \frac{B_n(1 - d_n^2)}{A_n^2 - a_n^2} \quad . \tag{47}$$

By (16),

$$1 - d_n^2 \quad = \quad 1 - \frac{b_n^2}{B_n^2} x \quad =: \quad f(x) \quad .$$

Using standard calculus, one verifies

$$\frac{\sqrt{g(x)}}{f(x)} \le \frac{\sqrt{g(1)}}{f(1)} = \frac{B_n}{A_n^2 - a_n^2} \quad , \quad 0 \le x \le 1 \quad ,$$

and thus (47) holds true.

b) First, we recall that

$$\sum_{k=1}^{n-1} x^k \quad = \quad \frac{x^n - x}{x - 1} \quad , \quad x \ne 1 \quad . \tag{48}$$

Moreover, for $k = 1, 2, \ldots$, one has $A_k < R^k$, $a_k + b_k = r^k$, and

$$\frac{r^{4k}}{r^{4k} - 1} \quad \le \quad \frac{r^4}{r^4 - 1} \quad . \tag{49}$$

Together with (48) (for $x = R/r$), we obtain

$$\sum_{k=1}^{n-1} (\frac{1}{a_k} + \frac{1}{b_k}) A_k \; < \; 4 \sum_{k=1}^{n-1} (\frac{R}{r})^k \frac{r^{4k}}{r^{4k} - 1} \; < \; \frac{4r^5}{(r^4 - 1)(R - r)} (\frac{R}{r})^n \quad .$$

Similarly, (49) and (50) (with $x = 1/(Rr)$) lead to

$$\sum_{k=1}^{n-1} (\frac{1}{a_k} + \frac{1}{b_k}) A_{n-k} \; < \; 4R^n \sum_{k=1}^{n-1} \frac{1}{(rR)^k} \frac{r^{4k}}{r^{4k} - 1} \; < \; \frac{4r^4}{(r^4 - 1)(Rr - 1)} R^n \quad .$$

We prove (46) by verifying that

$$\sum_{k=1}^{n-1} \frac{a_{n-k}}{a_k b_{n-k}} A_k \; < \; \frac{2r(r^2 + 1)}{(r^2 - 1)(R - r)} (\frac{R}{r})^n$$

and

$$\sum_{k=1}^{n-1} \frac{b_{n-k}}{a_{n-k}b_k} A_k \ < \ \frac{2r^3}{(r^2-1)(R-r)} (\frac{R}{r})^n \quad .$$

The first of these inequalities follows from

$$A_k < R^k \quad , \quad a_k > \frac{r^k}{2} \quad , \quad \frac{a_{n-k}}{b_{n-k}} = \frac{r^{2(n-k)}+1}{r^{2(n-k)}-1} \leq \frac{r^2+1}{r^2-1} \quad , \quad k = 1, \ldots, n-1 \quad ,$$

and (48) (with $x = R/r$). The second one is obtained by making use of

$$A_k < R^k \quad , \quad \frac{b_{n-k}}{a_{n-k}} < 1 \quad , \quad \frac{1}{b_k} \leq \frac{2r^2}{(r^2-1)r^k} \quad , \quad k = 1, \ldots, n-1 \quad ,$$

and again (48) (with $x = R/r$). This concludes the proof of the lemma. ∎

Next, we turn to the derivation of a lower bound for the numbers $\sigma_l^\star$, $l = 1, \ldots, 2n$. Using the fact that, by (16),

$$\frac{|\cos(n\gamma)|}{\sqrt{1-d_n^2}} = \sqrt{\frac{1-\sin^2(n\gamma)}{1-\sin^2(n\gamma)b_n^2/B_n^2}} \leq 1 \quad ,$$

and part a) of Lemma 2, one obtains from (45) and (43) the following inequalities:

$$\sigma_l^\star \geq 1 - \frac{a_n}{A_n} - a_n \sum_{k=1}^{n-1} (|\lambda_k(1)| + |\mu_k(1)|)$$

$$\geq 1 - \frac{a_n}{A_n} - \frac{2b_n^2}{b_n^2+d_n^2} (\frac{A_n}{a_n} - \frac{a_n}{A_n})^{-1} \left(\frac{B_n}{A_n} \sum_{k=1}^{n-1} (\frac{1}{a_{n-k}} + \frac{1}{b_{n-k}}) A_{n-k} \right. \tag{50}$$

$$\left. + \frac{B_n d_n^2}{A_n b_n} \sum_{k=1}^{n-1} (\frac{1}{a_k} + \frac{1}{b_k}) A_{n-k} + \frac{a_n B_n |d_n|}{A_n b_n} \sum_{k=1}^{n-1} (\frac{a_{n-k}}{a_k b_{n-k}} + \frac{b_{n-k}}{a_{n-k} b_k}) A_k \right) \quad , \quad l = 1, \ldots, 2n \quad .$$

We set $\gamma := (r/R)^n$. With (8) and (11), one easily verifies that

$$\frac{b_n}{B_n} < \gamma \quad , \quad \frac{a_n}{A_n} < 2\gamma \quad ,$$

and, together with (18), the estimates

$$\frac{B_n}{A_n} < 1 \quad , \quad \frac{B_n d_n^2}{A_n b_n} < \frac{2\gamma}{R^n} \quad , \quad \frac{a_n B_n |d_n|}{A_n b_n} < 2\gamma \tag{51}$$

follow. Furthermore, from now on it is assumed that $\gamma < 1/2$, and then

$$(\frac{A_n}{a_n} - \frac{a_n}{A_n})^{-1} < \frac{2\gamma}{1-4\gamma^2} \tag{52}$$

is guaranteed. By using (51), (52), and the inequalities stated in Lemma 2b), we finally deduce from (50) the following lower bound:

$$\sigma_l^\star > 1 - 2\gamma - \frac{1}{1-4\gamma^2} \cdot \frac{8r^5}{(r^4-1)(R-r)} \left(1 + \frac{(2r^2+1)(r^2+1)}{r^4}\gamma + \frac{2(R-r)}{r(Rr-1)}\gamma^2 \right) \quad ,$$

$$l = 1, 2, \ldots, 2n \quad .$$

In view of Theorem 3 and Theorem 1, this estimate leads to part a) of the following

Theorem 4. *Let $n \geq 2$, $c \in \partial\mathcal{E}_R$, and $R > r > 1$. Then:*
a) $q_n(z;c)$ is the optimal polynomial for (1) with corresponding minimal norm

$$D_n(r,c) \quad = \quad \frac{r^n + 1/r^n}{R^n + 1/R^n} \quad ,$$

if

$$\gamma \quad = \quad \gamma(r,R,n) \quad := \quad (\frac{r}{R})^n$$

is such that $\gamma < 1/2$ and

$$(1-2\gamma)^2(1+2\gamma) \geq \frac{8r^5}{(r^4-1)(R-r)}\left(1 + \frac{(2r^2+1)(r^2+1)}{r^4}\gamma + \frac{2(R-r)}{r(Rr-1)}\gamma^2\right) \qquad (53)$$

holds.
b) There exists a number $R_0(n,r)$ such that $q_n(z;c)$ is the extremal polynomial of (1) for all

$$c \in \partial\mathcal{E}_R \quad \textit{with} \quad R \geq R_0(n,r) \quad .$$

c) Let $c \in \partial\mathcal{E}_R$ be such that

$$R \; > \; r\,\frac{9r^4-1}{r^4-1} \quad . \qquad (54)$$

Then, there exists an integer $n_0(r,R)$ such that $q_n(z;c)$ is the extremal polynomial of (1) for all $n \geq n_0(r,R)$.

Proof. Only parts b) and c) remain to be proved. For fixed r and n, $\gamma(r,R,n) \to 0$ if $R \to \infty$, and (53) is clearly satisfied if R is sufficiently large. Similarly, if r and R are fixed, the condition (54) guarantees that (53) is true if n is large enough. This concludes the proof of Theorem 4. ∎

Remarks 6. Supported by numerical tests, we conjecture that part c) of Theorem 4 is true for arbitrary $R > r > 1$.

7. It follows from $R > r > 1$ that

$$1 + \frac{(2r^2+1)(r^2+1)}{r^4}\gamma + \frac{2(R-r)}{r(Rr-1)}\gamma^2 \quad \leq \quad 1 + 6\gamma + 2\gamma^2 \quad < \quad \frac{9}{4}\,(1+2\gamma)$$

for all $0 \leq \gamma < 1/2$. Thus (53) is true if $\gamma < 1/2$ satisfies the stronger condition

$$(1-2\gamma)^2 \quad \geq \quad \frac{18r^5}{(r^4-1)(R-r)} \quad . \qquad (55)$$

Using (55), one easily obtains explicit formulas for numbers $R_0(n,r)$ with the property stated in Theorem 4b). E.g. set

$$R_0(n,r) \ := \ r \ \max\left\{4^{1/n}, \frac{73r^4-1}{r^4-1}\right\} \quad . \tag{56}$$

Then, for all $R \geq R_0(n,r)$

$$(1-2(\frac{r}{R})^n)^2 \ \geq \ \frac{1}{4} \ \geq \ \frac{18r^5}{(r^4-1)(R-r)}$$

and, in particular $(r/R)^n < 1/2$. Hence, $R_0(n,r)$ is suitable for Theorem 4b).

8. Let $\mathcal{G}_n(r)$ denote the set of all points $c \in \mathbb{C} \setminus \mathcal{E}_r$ for which $q_n(z;c)$ is the optimal polynomial for (1). By Theorem 4b, $\mathcal{G}_n(r)$ is an unbounded set. More precisely, we proved that

$$c \in \mathcal{G}_n(r) \quad \text{for all } c \in \mathbb{C} \quad \text{with} \quad |c| \geq \frac{1}{2}(R_0 + \frac{1}{R_0}) \quad ,$$

where $R_0 = R_0(n,r)$ is given by (56). The boundary of $\mathcal{G}_n(r)$ is a closed Jordan curve which, in view of Theorem 3, is composed of pieces of $\partial\mathcal{E}_r$ and of pieces of the curves

$$\sigma_l^*(c) = 0 \quad , \quad l = 1, 2, \ldots, 2n \quad .$$

We have computed these curves numerically for a number of cases. Some typical pictures (for $r = 2$, $n = 2, \ldots, 5$) are shown in Fig. a–d. Because of the symmetry with respect to the real and imaginary axis, we have only plotted the first quadrant. $\mathcal{G}_n(r)$ is the region exterior to these curves including the parts of its boundary which are described by the curves $\sigma_l^*(c) = 0$. Note that near the real axis the boundary of $\mathcal{G}_n(r)$ is given by $\partial\mathcal{E}_r$. This led us to the following conjecture: For all $n \in \mathbb{N}$ and $r > 1$ there exists a number $\rho_0(n,r) > 0$ such that $q_n(z;c)$ is the optimal polynomial of (1) for all $c \in \mathbb{C} \setminus \mathcal{E}_r$ with $|\arg c| \leq \rho_0(n,r)$.

References

1. M. Eiermann, W. Niethammer, and R.S. Varga, A study of semiiterative methods for nonsymmetric systems of linear equations, *Numer. Math.* **47** (1985), 505-533

2. B. Fischer, Über komplexe Tschebyscheff-Approximationsprobleme, die bei Iterationsverfahren zur Lösung linearer Gleichungssysteme entstehen, Dissertation, Universität Hamburg, 1987.

3. R. FREUND AND S. RUSCHEWEYH, On a class of Chebyshev approximation problems which arise in connection with a conjugate gradient type method, *Numer. Math.* **48** (1986), 525-542.

4. T.A. MANTEUFFEL, The Tchebychev iteration for nonsymmetric linear systems, *Numer. Math.* **28** (1977), 307-327.

5. G. MEINARDUS, "Approximation of Functions: Theory and Numerical Methods ", Springer-Verlag, Berlin, Heidelberg, New York, 1967.

6. G. OPFER AND G. SCHOBER, Richardson's iteration for nonsymmetric matrices, *Linear Algebra Appl.* **58** (1984), 343-361.

7. T.J. RIVLIN AND H.S. SHAPIRO, A unified approach to certain problems of approximation and minimization, *J. Soc. Indust. Appl. Math.* **9** (1961), 670-699.

8. L.N. TREFETHEN, Near-circularity of the error curve in complex Chebyshev approximation, *J. Approx. Theory* **31** (1981), 344-367.

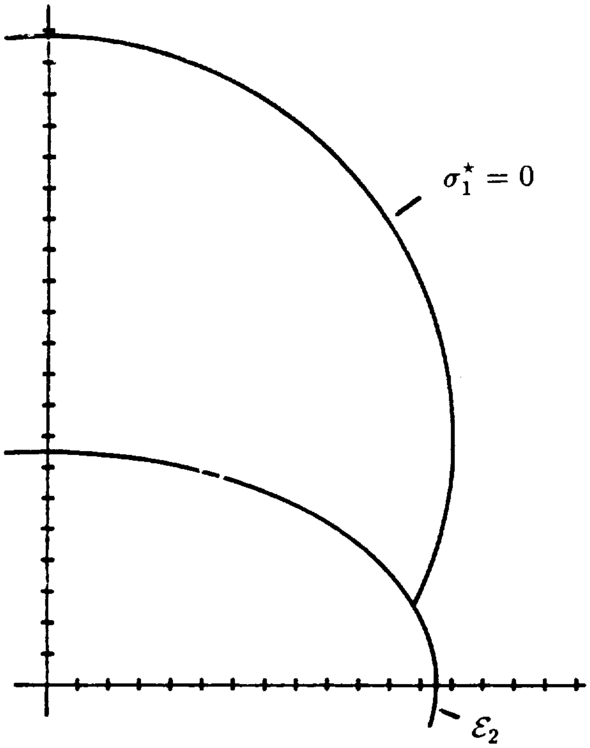

Fig. a $(r = 2,\ n = 2)$

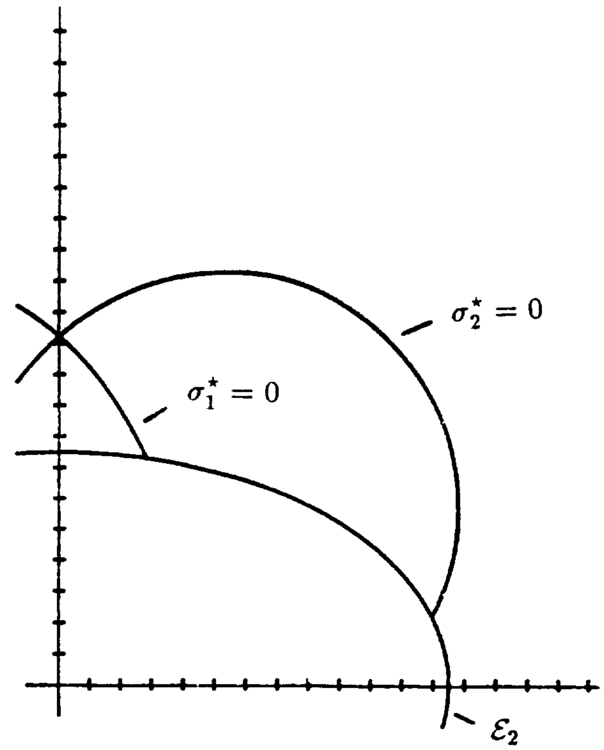

Fig. b $(r = 2,\ n = 3)$

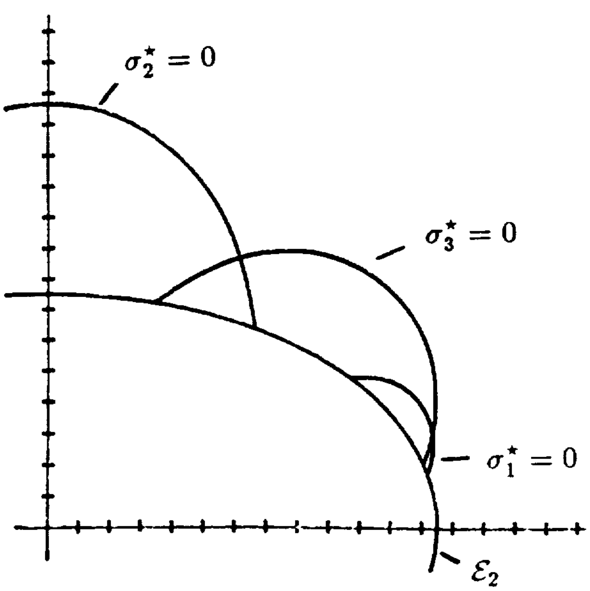

Fig. c $(r = 2,\ n = 4)$

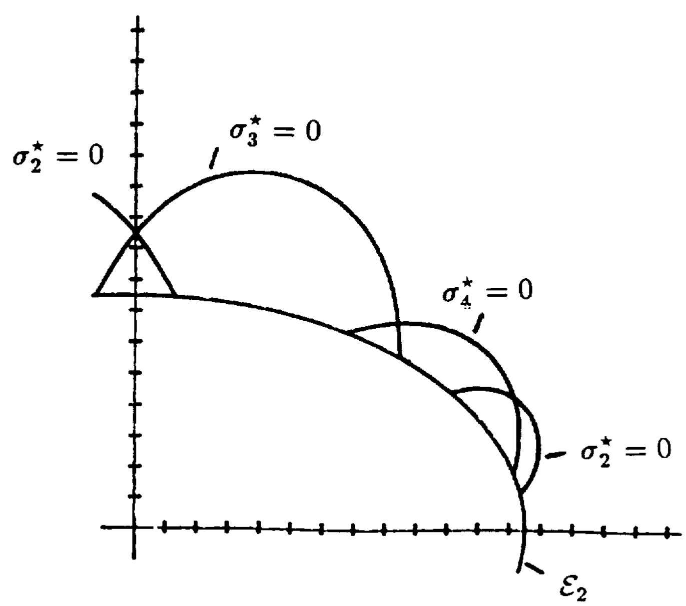

Fig. d $(r = 2,\ n = 5)$