Numer. Math. 37, 297–320 (1981)

Numerische Mathematik
© Springer-Verlag 1981

Rational Chebyshev Approximation on the Unit Disk

Lloyd N. Trefethen

Computer Science Department, Stanford University, Stanford, California 94305, USA

Summary. In a recent paper we showed that error curves in polynomial Chebyshev approximation of analytic functions on the unit disk tend to approximate perfect circles about the origin [23]. Making use of a theorem of Carathéodory and Fejér, we derived in the process a method for calculating near-best approximations rapidly by finding the principal singular value and corresponding singular vector of a complex Hankel matrix. This paper extends these developments to the problem of Chebyshev approximation by rational functions, where non-principal singular values and vectors of the same matrix turn out to be required. The theory is based on certain extensions of the Carathéodory-Fejér result which are also currently finding application in the fields of digital signal processing and linear systems theory.

It is shown among other things that if $f(\varepsilon z)$ is approximated by a rational function of type (m, n) for $\varepsilon > 0$, then under weak assumptions the corresponding error curves deviate from perfect circles of winding number $m+n+1$ by a relative magnitude $O(\varepsilon^{m+n+2})$ as $\varepsilon \to 0$. The "CF approximation" that our method computes approximates the true best approximation to the same high relative order. A numerical procedure for computing such approximations is described and shown to give results that confirm the asymptotic theory. Approximation of e^z on the unit disk is taken as a central computational example.

Subject Classifications: AMS(MOS) 30D50, 30E10, 41A50.

Contents

0029-599X/81/0037/0297/$04.80

1. Introduction

Let $S \equiv \{z \in \mathbb{C}: |z| = 1\}$ be the complex unit circle, let $D \equiv \{z \in \mathbb{C}: |z| < 1\}$ be the open unit disk, and let $\bar{D} = D \cup S$. Let R_{mn} be the space of rational functions of type (m, n) that have no poles in $\bar{D}$; that is, the set of rational functions with at most m finite zeros and at most n finite poles, with all of the poles outside of the unit disk. Let $\|\cdot\|$ be the supremum norm over S. Here is the *rational Chebyshev approximation problem:* given f analytic in D and continuous on $\bar{D}$, find a rational function $r^*_{mn} \in R_{mn}$ such that $\|f - r^*_{mn}\| = \inf_{r \in R_{mn}} \|f - r\|$. Such a *best approximation* exists for any f, m, n, but it need not be unique when $n > 0$ [17]. Where clarity permits we will usually drop the subscripts m, n of r^*_{mn} and similar functions.

For given f and any r, the image of S under $f - r$ describes some curve in the plane, which we call the *error curve* corresponding to r. A best approximation r^* is a function whose error curve can be contained in a disk of minimal radius about the origin. This work began with the observation, based on numerical computations, that for smooth f, the error curve corresponding to r^* often approximates closely a perfect circle about the origin of winding number $m + n + 1$, and that this near-circularity phenomenon becomes more pronounced as $m \to \infty$ [23].

For example, consider approximation of e^z on the unit disk. Figure 1 shows error curves corresponding to Padé and Chebyshev approximations of type (1, 1). Both curves have winding number 3, but whereas the first one varies in radius considerably, the second one evidently approximates a circle to within a fraction of a percent. The plot is typical for smooth functions f. If you've seen one Chebyshev approximation error curve plot on the unit disk, you have (almost) seen them all.

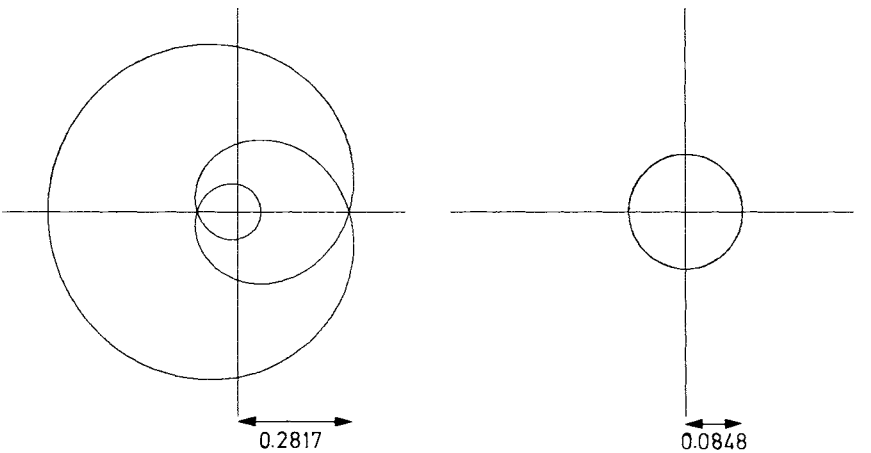

Fig. 1. Error curves for rational approximation of type (1, 1) to e^z on the unit disk. Error curves for Padé (left) and Chebyshev (right) approximation are shown plotted on the same scale. The latter varies in radius by less than 1%, and this figure decreases rapidly if the degree of the numerator is increased, as shown in Table 1

Table 1. Relative deviation α from a perfect circle (Eq. (1.1)) of error curves of best approximations r^*_{mn} to e^z on the unit disk. Various m, n. Some figures uncertain

$m=$ \ $n=$	0	1	2	3
0	2(−1)	3(−1)	3(−1)	2(−1)
1	5(−3)	7(−3)	2(−2)	5(−3)
2	4(−5)	1(−4)	?	?
3	5(−6)	5(−7)	?	?

Let us quantify such near-circularity by defining

$$\alpha \equiv \frac{\|f-r\| - \min_{z\in S} |(f-r)(z)|}{\|f-r\|} \tag{1.1}$$

as a measure of the relative deviation of an error curve from a perfect circle. Then for each pair (m,n) in the range $0 \leqq m,n \leqq 3$, the winding number of $e^z - r^*_{mn}$ on S is in fact exactly $m+n+1$, and Table 1 shows the remarkable decrease of α with m.

A previous paper [23] analyzed the near-circularity phenomenon for the case of Chebyshev approximation by polynomials $-n=0$. The purpose of this paper is to extend that analysis to rational approximation.

We begin in Sect. 2 with preliminary definitions and propositions. An extended approximation space $\tilde{R}_{mn} \supseteq R_{mn}$ is defined, and Rouché's theorem is applied to show that any approximation with a nearly circular error curve must be close to best. (Too many papers in complex Chebyshev approximation invoke the Kolmogorov criterion in places where Rouché's theorem would suffice.) The approximation problem associated with $\tilde{R}_{mn}$ is solved in Sect. 3, and it is shown that the solution is always characterized by a perfectly circular error curve (Theorem 3.2). This is the extension of the Carathéodory-Fejér Theorem, based upon the singular value decomposition of a Hankel matrix of Maclaurin series coefficients of f, that this work is founded upon. Section 4 is devoted to describing the asymptotic behavior of $\tilde{r}^*_{mn}$, the best approximation out of $\tilde{R}_{mn}$ on S to a function $f(\varepsilon z)$, as $\varepsilon \to 0$. The purpose is to show that for small ε, $\tilde{r}^*$ comes very close on S to a rational function in R_{mn}. Such a function is derived from $\tilde{r}^*$ in Sect. 5 and named the "Carathéodory-Fejér approximation" r^{cf}_{mn}. It is confirmed that r^{cf} and $\tilde{r}^*$ become almost equal on S as $\varepsilon \to 0$, hence that r^{cf} has a nearly circular error curve, hence that it is close to best.

At this point it has been established that r^{cf} is near best in the sense that $\|f-r^{cf}\|$ is not much bigger than $\|f-r^*\|$. If r^{cf} has a nearly circular error curve, however, one can show further that in fact $\|r^{cf}-r^*\|$ must be correspondingly small. The required a posteriori estimate is applied in Sect. 6 to establish the most important conclusions of this paper: that at least in the asymptotic limit $\varepsilon \to 0$, best approximations are approximated exceedingly closely by the CF approximation (Theorem 6.2), and best approximation error curves are exceedingly close to circular (Theorem 6.3). Section 6 also contains a summary and a discussion of these results.

An extensive amount of numerical experimentation has accompanied this theoretical work. Though most of our theorems are asymptotic, the CF method is astonishingly successful in many ordinary approximation problems on the unit disk. Section 7 describes an efficient method for the numerical computation of $\tilde{r}^*$ and r^{cf}. In Sect. 8 the problem of approximation of e^z is considered numerically. Section 9 concludes the paper with some final remarks.

2. Preliminaries

The sets S, D, $\bar{D}$, the space R_{mn}, and the norm $\|\cdot\|$ have already been defined, along with the best approximation $r^* \in R_{mn}$ (not necessarily unique) with respect to $\|\cdot\|$ to a function f.

Let G be the set of functions which are analytic and bounded in $1<|z|\leqq\infty$ and zero at $z=\infty$; that is, with expansions of the form $\sum_{k=-\infty}^{-1} g_k z^k$ that converge and are bounded outside the closed unit disk. A function $g\in G$ need not extend continuously to S, but it will have a non-tangential limit almost everywhere there [14]. By means of these limits we will apply the norm $\|\cdot\|$ in the obvious way to g and to sums of the form $g+f$, where f is defined on S.

For any $n\geqq 0$, define

$$\tilde{R}_{nn} \equiv R_{nn} + G, \tag{2.1}$$

where R_{nn} still includes only rational functions whose poles lie outside the disk $\bar{D}$. Further, for any $m\geqq 0$, define

$$\tilde{R}_{mn} \equiv z^{m-n} \tilde{R}_{nn}. \tag{2.2}$$

It is not hard to see that $\tilde{R}_{mn}$ is precisely the set of functions that are bounded on S and can be written in the form

$$r(z) = \sum_{k=-\infty}^{m} d_k z^k \Big/ \sum_{k=0}^{n} e_k z^k. \tag{2.3}$$

Note that it is *not* the case that $\tilde{R}_{mn} = R_{mn} + G$, unless $m \geqq n-1$.

An important tool beginning in Sect. 4 will be the *Padé approximation* of type (m, n) to f, denoted r^p_{mn}. This is the (unique) rational function of type (m, n) whose Maclaurin series matches that of f to as high an order as possible. An excellent reference on Padé approximation is the survey by Gragg [8]. We will also speak of "the" Laurent series of an analytic function $\phi(z)$. Whenever we do, ϕ will be analytic on S, and the Laurent series intended is the one that converges in a neighborhood of that circle. Its coefficients are given by Cauchy integrals on S, and are readily computed numerically by the Fast Fourier Transform (Sect. 7).

A limiting case for the rational approximation problem was already given in [15] and in [23]. Here and throughout this paper, the "winding number" is with respect to the origin.

Proposition 2.1 *(circular⇒best). Given f analytic in D and continuous on $\bar{D}$, suppose the error curve of some function $r \in R_{mn}$ is a perfect circle about the origin with winding number $\geqq m+n+1$. Then r is a best approximation to f out of R_{mn}. However, this situation can occur only if f is a rational function.*

Proof [23]. The first assertion is a consequence of Rouché's theorem and the definition of a best approximation. The second follows from the symmetry principle and the fact that any function meromorphic in the extended plane must be rational. ∎

The argument by Rouché's theorem extends immediately to give a bound on $\|f-r^*\|$ in the case where the error curve of r is not exactly circular but nearly so. This proposition is an analog of the de la Vallée Poussin theorem in real approximation [17].

Proposition 2.2 *(nearly circular⇒near best). Given f analytic in D and continuous on $\bar{D}$, suppose the error curve of some function $r \in R_{mn}$ does not pass through the origin and has winding number $\geqq m+n+1$. Then*

$$\min_{z \in S} |(f-r)(z)| \leqq \|f-r^*\| \leqq \|f-r\|. \quad \blacksquare$$

A similar argument is valid for approximation out of $\tilde{R}_{mn}$. We will need this result in the next section.

Lemma 2.3. *Given f analytic in D and continuous on $\bar{D}$, suppose the error curve of some function $\tilde{r} \in \tilde{R}_{mn}$ does not pass through the origin and has winding number $\geqq m+n+1$. Then*

$$\min_{z \in S} |(f-\tilde{r})(z)| \leqq \|f-\tilde{r}^*\| \leqq \|f-\tilde{r}\|.$$

Proof. The second inequality follows from the best approximation property of $\tilde{r}^*$. For the first, suppose to the contrary that for some $\tilde{r}' \in \tilde{R}_{mn}$

$$\|f-\tilde{r}'\| < \min_{z \in S} |(f-\tilde{r})(z)|.$$

Without loss of generality we may assume $\tilde{r}'$ is continuous on S, for if it is not, the inequality will still hold for some function $\tilde{r}'(Rz)$, where $R>1$ is sufficiently close to 1. Then clearly $\tilde{r}-\tilde{r}'$ has the same winding number as $f-\tilde{r}$, which is $\geqq m+n+1$. Now it is easy to see that $\tilde{r}-\tilde{r}'$ belongs to $\tilde{R}_{m+n,\,2n}$. However, such a function can have winding number no greater than $m+n$, for it is meromorphic in $1<|z|\leqq\infty$ with at most $2n+(m+n-2n)=m+n$ poles there. This contradiction finishes the proof. ∎

3. The Extended Best Approximation $\tilde{r}^* \in \tilde{R}_{mn}$

The Chebyshev approximation problem in R_{mn} has no closed form solution, but the same problem in $\tilde{R}_{mn}$ does. Here we present that solution. The theory has a clear beginning in the seminal paper of Carathéodory and Fejér in 1911 [5], which considered the polynomial case $m=n=0$. This original theory was

rederived and extended a short while later by Schur in 1918 [19]. The extension to rational approximation ($m=n\mp 0$) was first accomplished by Takagi in 1924 [22], who built upon the work of Schur. Later, essentially the same results were rediscovered by Akhieser in 1931 [2, 3], and rediscovered again by Clark in 1968 [6]. The most general, complete, and correct exposition can be found in the recent work of Adamian, Arov, and Krein [1].

Our own presentation will avoid functional analysis, and as a result it is closer in spirit to the development of Takagi than to that of Adamian, Arov, and Krein. Various minor modifications have been made, however, and in particular the language of singular value decompositions is used. The extension to $m\mp n$ is new, and important for the applications that follow, but mathematically trivial.

Let f be a polynomial $f(z)=c_0+\ldots+c_K z^K$, and let H_f denote the Hankel matrix

$$H_f \equiv \begin{pmatrix} c_1 & c_2 & \cdots & c_K \\ c_2 & & & \\ \vdots & & 0 & \\ c_K & & & \end{pmatrix}.$$

(A matrix is *Hankel* if its entries are constant along cross diagonals.) H_f is symmetric but if the c_k are not real, it is not Hermitian. Let

$$H_f = U\Sigma V^H$$

be *a singular value decomposition* of H_f; i.e. let the above equation hold with U, V unitary and Σ of the form $\mathrm{diag}(\sigma_1, \sigma_2, \ldots, \sigma_K)$, $\sigma_1 \geqq \sigma_2 \geqq \ldots \geqq \sigma_K \geqq 0$. (Because of the symmetry of H_f, we may require $V=\bar{U}$, but this is not necessary in the formulation that follows.) Then here is a kind of reverse generalized Carathéodory-Fejér theorem (labeled "proposition" because it will be extended further in Theorem 3.2). We will give a partial proof based on a winding number argument.

Proposition 3.1. *The polynomial $f(z)=c_0+\ldots+c_K z^K$ has a unique best approximation $\tilde{r}^*_{nn}$ out of $\tilde{R}_{nn}$. The error is*

$$\|f-\tilde{r}^*\| = \sigma_{n+1}(H_f)$$

(where $\sigma_{n+1}\equiv 0$ if $n+1>K$), and the error curve is a perfect circle about the origin whose winding number is exactly $2n+1$ if σ_{n+1} is simple. $\tilde{r}^*$ is given by

$$f(z)-\tilde{r}^*(z) = \sigma_{n+1} z^K \frac{u_1+u_2 z+\ldots+u_K z^{K-1}}{v_K+v_{K-1} z+\ldots+v_1 z^{K-1}} \tag{3.1}$$

where $u=(u_1, \ldots, u_K)^T$ is the $(n+1)$st column of U and $v=(v_1, \ldots, v_K)^T$ is the $(n+1)$st column of V in any singular value decomposition $H_f=U\Sigma V^H$.

Proof. A complete proof is given in [1]. The following argument proves everything but uniqueness under the additional assumption that all of the singular values of H_f are distinct.

Let $\tilde{r}^*$ be whatever function is defined by (3.1); we will show that it belongs to $\tilde{R}_{nn}$ and is a best approximation of that class. Without loss of generality let us assume $v=\bar{u}$, and indeed $V=\bar{U}$. We may do this because the symmetry of H_f implies $u=we^{i\phi}$ and $v=\bar{w}e^{i\phi}$ for some vector w, and the factor $e^{i\phi}$ will drop out in the quotient of (3.1). The right hand side of (3.1) is now a constant times z^K times a *finite Blaschke product*,

$$\sigma_{n+1} z^K \frac{u_1+u_2 z+\ldots+u_K z^{K-1}}{\bar{u}_K+\bar{u}_{K-1} z+\ldots+\bar{u}_1 z^{K-1}},$$

and therefore it is a rational function of type $(2K-1, K-1)$ that maps S onto a perfect circle of radius σ_{n+1}.

Multiply both sides of (3.1) by the denominator $\bar{u}_K+\ldots+\bar{u}_1 z^{K-1}$. This yields an equation

$$[(c_0+\ldots+c_K z^K)-\tilde{r}^*(z)](\bar{u}_K+\ldots+\bar{u}_1 z^{K-1})=\sigma_{n+1} z^K(u_1+\ldots+u_K z^{K-1})$$

in which each side must be a polynomial of degree at most $2K-1$. It turns out that the polynomial $\tilde{r}^*(z)(\bar{u}_K+\ldots+\bar{u}_1 z^{K-1})$ has degree only at most $K-1$, however. For if we ignore this term and compare coefficients of z^k for $k=2K-1, 2K-2, \ldots, K$, we get the system of equations

$$\begin{aligned} c_K\bar{u}_1 &= \sigma_{n+1} u_K \\ c_{K-1}\bar{u}_1+c_K\bar{u}_2 &= \sigma_{n+1} u_{K-1} \\ &\vdots \\ c_1\bar{u}_1+c_2\bar{u}_2+\ldots+c_K\bar{u}_K &= \sigma_{n+1} u_1. \end{aligned}$$

This system (in reverse order) can be written $H_f\bar{u}=u\sigma_{n+1}$, and therefore it is satisfied since $H_f=U\Sigma\bar{U}^H$.

Thus $\tilde{r}^*$ must be a rational function of type $(K-1, K-1)$ representable with denominator $\bar{u}_K+\ldots+\bar{u}_1 z^{K-1}$. Suppose that it has ν poles in $1<|z|\leqq\infty$ counted with multiplicity. Then $\tilde{r}^*$ can be written in the form $g+r_{\nu\nu}$, where $r_{\nu\nu}\in R_{\nu\nu}$ and g is analytic in $1<|z|\leqq\infty$ and zero at ∞. Now if $\bar{u}_K+\ldots+\bar{u}_1 z^{K-1}$ happens to have any zeros on the circle S, then the numerator $u_1+\ldots+u_K z^{K-1}$ has these as zeros also, so they cancel in (3.1). Therefore g is bounded in $1<|z|\leqq\infty$, which implies $g\in G$ and $\tilde{r}^*\in\tilde{R}_{\nu\nu}$.

Let μ be the number of poles of $\tilde{r}^*$ in D. Then the right hand side of (3.1) has winding number $\tau\equiv K+\nu-\mu$ on S. Since $\nu+\mu\leqq K-1$, we have $\tau\geqq 2\nu+1$. It follows now by Lemma 2.3 that $\tilde{r}^*$ is a best approximation to f in $\tilde{R}_{\nu\nu}$*.

Thus each of $\sigma_1, \ldots, \sigma_K$ is the error corresponding to a best approximation in $\tilde{R}_{\nu\nu}$ for some ν with $0\leqq\nu\leqq K-1$. By the assumption that the singular values of H_f are distinct, these values of ν must be distinct and increase monotonically with n. Hence we can only have $\nu=n$, and the theorem is proved. ∎

Proposition 3.1 gives a constructive description of best approximations to polynomials out of the sets $\tilde{R}_{nn}$. It is quite easy to extend this result to approximation out of $\tilde{R}_{mn}$. First, if $m<n$, we naturally proceed by finding the

* **Note Added in Proof:** This argument must be amended slightly to handle the possibility that $\tilde{r}^*$ has poles at ∞

best approximation $\tilde{r}_{nn}^{*\prime}$ to $f' \equiv z^{n-m} f$ out of $\tilde{R}_{nn}$. This will involve the right-shifted Hankel matrix

$$H_f^{(v)} \equiv \begin{pmatrix} 0 & \dots & c_0 & c_1 & \cdot & \cdot & c_K \\ \vdots & & c_0 & c_1 & & c_K & \\ c_0 & c_1 & & & c_K & & \\ \vdots & & \iddots & & & 0 & \\ c_K & & & & & & \end{pmatrix} (K+v \times K+v) \quad (v>0)$$

with $v=n-m$. Since z^{n-m} has constant modulus on S, it is easy to see that $\tilde{r}_{mn}^* \equiv \tilde{r}_{nn}^{*\prime}/z^{n-m}$ must be a best approximation to f out of $\tilde{R}_{mn}$, and that the corresponding error curve is a perfect circle with winding number $m+n+1$ if $\sigma_{n+1}(H_f^{(n-m)})$ is simple.

Second, if $m \geqq n$, we again proceed by finding a best approximation $\tilde{r}_{nn}^{*\prime}$ to $f' \equiv z^{n-m} f$. This time f' may possess terms of negative degree, but these have no influence on the approximation problem because they are absorbed in G, and $\tilde{R}_{nn} = R_{nn} + G$. Thus we now make use of the left-shifted Hankel matrix

$$H_f^{(v)} \equiv \begin{pmatrix} c_{1-v} & c_{2-v} & \cdots & c_K \\ c_{2-v} & & & \\ \vdots & & 0 & \\ c_K & & & \end{pmatrix} (K+v \times K+v) \quad (v \leqq 0).$$

We may sum up these results in the main theorem of this section.

Theorem 3.2 (*solution of extended approximation problem*). *The polynomial* $f(z) = c_0 + \dots + c_K z^K$ *has a unique approximation* $\tilde{r}_{mn}^*$ *out of* $\tilde{R}_{mn}$. *The error is*

$$\| f - \tilde{r}^* \| = \sigma_{n+1}(H_f^{(n-m)})$$

(*where* $\sigma_{n+1} \equiv 0$ *if* $n+1 > K+n-m$), *and the error curve is a perfect circle about the origin whose winding number is* $m+n+1$ *if* σ_{n+1} *is simple.* $\tilde{r}^*$ *is given by*

$$f(z) - \tilde{r}^*(z) = \sigma_{n+1} z^K \frac{u_1 + \dots + u_{K+n-m} z^{K+n-m-1}}{v_{K+n-m} + \dots + v_1 z^{K+n-m-1}} \tag{3.2}$$

where $u = (u_1, \dots, u_{K+n-m})^T$ *and* $v = (v_1, \dots, v_{K+n-m})^T$ *are the* $(n+1)$*st columns of* U *and* V, *respectively, in any singular value decomposition* $H_f^{(n-m)} = U \Sigma V^H$. ∎

The contrast between Theorem 3.2 and the situation for ordinary rational approximation is great; in the latter problem uniqueness is not assured, existence proofs are nonconstructive, and the error curve cannot be very satisfactorily characterized. This is why introducing the extended approximation problem is so fruitful.

Note that if $K = m+1$ and $\sigma_{n+1}(H_f^{(n-m)})$ is simple, then $\tilde{r}^*$ must have n poles in $1 < |z| < \infty$, which according to (3.2) implies that none can lie inside D. Therefore $\tilde{r}_{mn}^* \in R_{mn}$, so $\tilde{r}^* = r^*$. Thus Theorem 3.2 gives the exact solution of the Chebyshev approximation problem in this case. (This problem goes back to Chebyshev. See [3], p. 278 and [17], p. 166.)

The restriction that f must be a polynomial has been adopted for simplicity. Adamian, Arov, and Krein prove an extension of Proposition 3.1 for approximation of an arbitrary function $f \in L^\infty(S)$, and here the singular values of an infinite Hankel matrix come into play [1]. The uniqueness and winding number assertions are no longer so simple. For functions f that are analytic in a neighborhood of $\bar{D}$, this kind of result can presumably be derived from Theorem 3.2 by considering limits as K approaches ∞. In practice, we will normally be given a function of this kind, such as e^z, and will truncate it to a polynomial at some term z^K. It is also possible to base a constructive theory on the assumption that f is *rational* of some type (K, K). For this approach, and for a presentation of Proposition 3.1 in the language of linear systems theory, see for example [20].

Theorem 3.2 provides an immediate lower bound for rational best approximation errors. In most cases this is much tighter than other lower bounds that have been published:

Theorem 3.3 (σ *is a lower bound*). *Let f be analytic in a neighborhood of $\bar{D}$ and let $\sigma = \lim_{K\to\infty} \sigma_{n+1}$ from Theorem* 3.2; *or let f be analytic in D and belong to $L^\infty(S)$, with σ defined as the $(n+1)$st singular value of the infinite Hankel matrix $H_f^{(n-m)}$ with $K = \infty$. Then*

$$\sigma \leqq \| f - r^* \|. \tag{3.3}$$

Proof. (3.3) follows from Theorem 3.2 and the inclusion $R_{mn} \subseteq \tilde{R}_{mn}$. The limit in the first hypothesis must exist, and must be a lower bound for $\| f - r^* \|$, since under that hypothesis f is the uniform limit on $\bar{D}$ of its partial Maclaurin sums. For the second hypothesis, see [1]. ∎

4. Asymptotic Behavior of $\tilde{r}^*$ as $\varepsilon \to 0$

Throughout this section we shall assume the following setup. Let $\hat{f}(z) = c_0 + c_1 + \ldots + c_K z^K$ be a polynomial of degree K, and assume $c_K \neq 0$ for convenience. (For applications the assumption that $\hat{f}$ is a polynomial is unnecessary, and it will be removed in the next section.) For given $\varepsilon > 0$, define

$$f(z) \equiv \hat{f}(\varepsilon z).$$

If $c_k \equiv 0$ for $k < 0$ and $k > K$, then for any $\varepsilon > 0$, f has the Laurent series

$$f(z) = \sum_{k=-\infty}^{\infty} c_k (\varepsilon z)^k. \tag{4.1}$$

Let $m \geqq 0$ and $n \geqq 0$ be fixed nonnegative integers, and assume:

Assumption A. The Padé approximant r_{mn}^p of f has n finite poles, and its Taylor series agrees with f exactly through the term of degree $m+n$.

(If the assumption is true for any ε, of course, it is true for all ε.) Let the Taylor series of r^p be

$$r^p(z) = \sum_{k=-\infty}^{\infty} c_k^p (\varepsilon z)^k, \tag{4.2}$$

with $c_k \equiv 0$ for $k<0$; for all sufficiently small ε (so that the poles of r^p lie outside $\bar{D}$), (4.2) is also the Laurent series for r^p. (This means, with respect to S; see Section 2.) Both $\{c_k\}$ and $\{c_k^p\}$ are independent of ε. For given $\varepsilon>0$, let $\tilde{r}^*$ be the best approximation out of $\tilde{R}_{mn}$ to f on S given by Theorem 3.2. Let $\tilde{r}^*$ have the Laurent series

$$\tilde{r}^*(z) = \sum_{k=-\infty}^{\infty} \tilde{c}_k^* (\varepsilon z)^k. \tag{4.3}$$

f, r^p, $\tilde{r}^*$, and the coefficients $\tilde{c}_k^*$ depend on ε, but we shall indicate none of this in the notation.

This section is the foundation of all the asymptotic results that follow. Its main purpose is to show that when ε is small, the Laurent coefficients $(c_k - \tilde{c}_k^*)\varepsilon^k$ of $f-\tilde{r}^*$ decrease geometrically in size as k decreases from $m+n+1$ towards $-\infty$ (Lemma 4.3). To show this, we begin by showing that $\tilde{r}^*$ is close to r^p, making use of the nonvanishing of a Hankel determinant implied by Assumption A (Lemma 4.1). From this and a winding number argument based on the Blaschke representation of Theorem 3.2, it is shown that as $\varepsilon \to 0$, all the poles and zeros of $f-\tilde{r}^*$ either approach 0 like ε or approach ∞ like $1/\varepsilon$ (Lemma 4.2). Lemma 4.3 then follows by Cauchy's estimate.

The Hankel determinant argument of Lemma 4.1 is central to our results. Though we state it here not in full generality but only in the context of relating $\tilde{r}^*$ to r^p, the same reasoning will be appealed to twice more in Sect. 6 to relate r^* to r^p and r^{cf} to r^*. The idea is that in the presence of a condition like Assumption A, near equality of the first $m+n$ Maclaurin coefficients of two functions in R_{mn} or $\tilde{R}_{mn}$ implies near equality of the remainder of the coefficients. The argument is a sharpening of one used by Walsh in Theorem 1 of [24].

Lemma 4.1. *Assume the conditions of Assumption* A. *For each* ε *let* r^p *and* $\tilde{r}^*$ *be represented in the form*

$$r^p(z) = \frac{d_0^p + \ldots + d_m^p (\varepsilon z)^m}{1 + e_1^p (\varepsilon z) + \ldots + e_n^p (\varepsilon z)^n} \tag{4.4}$$

and, as in (2.3),

$$\tilde{r}^*(z) = \frac{\ldots + \tilde{d}_{-1}^* (\varepsilon z)^{-1} + \tilde{d}_0^* + \ldots + \tilde{d}_m^* (\varepsilon z)^m}{1 + \tilde{e}_1^* (\varepsilon z) + \ldots + \tilde{e}_n^* (\varepsilon z)^n}. \tag{4.5}$$

(*The numerator of* (4.5) *is understood to converge in* $1<|z|<\infty$, *while its denominator has all its zeros in* $|z|>1$. *Other than this we make no a priori assumptions about these representations, for example that they are unique or that common factors have been cancelled.*) *Then as* $\varepsilon \to 0$,

$$|\tilde{e}_k^* - e_k^p| = O(\varepsilon), \qquad 1 \leqq k \leqq n, \tag{4.6}$$

$$|\tilde{c}_{m+n+1}^* - c_{m+n+1}^p| = O(\varepsilon), \tag{4.7}$$

and

$$\sum_{k=m+n+2}^{\infty} |\tilde{c}_k^* \varepsilon^k| = O(\varepsilon^{m+n+2}). \tag{4.8}$$

Proof. Assumption A implies the bound

$$\|f-r^p\| = O(\varepsilon^{m+n+1})$$

as $\varepsilon \to 0$, and this implies in turn

$$\sigma_{n+1} = \|f-\tilde{r}^*\| = O(\varepsilon^{m+n+1}), \tag{4.9}$$

since $\tilde{r}^*$ must be at least as good an approximation to f as r^p. Combining these bounds yields

$$\|\tilde{r}^* - r^p\| = O(\varepsilon^{m+n+1}).$$

Therefore by Cauchy's estimate we must have

$$|\tilde{c}_k^* - c_k^p| = O(\varepsilon^{m+n+1-k})$$

for all k, and in particular

$$|\tilde{c}_k^* - c_k^p| = O(\varepsilon) \qquad \forall k \leqq m+n. \tag{4.10}$$

Equating (4.2) and (4.4) and multiplying through by the denominator of (4.4) leads to an identity between the numerator polynomial of (4.4) and a convergent power series. Terms in this identity can be equated in powers of z. Carrying this out for powers z^{m+1} through z^{m+n} leads to the system of equations

$$\begin{pmatrix} c_{m-n+1}^p & & c_m^p \\ & \iddots & \\ c_m^p & & c_{m+n-1}^p \end{pmatrix} \begin{pmatrix} e_n^p \\ \vdots \\ e_1^p \end{pmatrix} = -\begin{pmatrix} c_{m+1}^p \\ \vdots \\ c_{m+n}^p \end{pmatrix}. \tag{4.11}$$

The matrix here is a Hankel matrix.

By Assumption A, r^p has a full n finite poles, hence a pole at infinity of order at most $m-n$. It follows from the theory of Hankel determinants that the matrix in (4.11) is nonsingular [12, Theorem 7.5.e]. This implies that the coefficients $e_1^p, \ldots, e_n^p$, which constitute the solution of (4.11), are unique after all, as indeed could have been made clear on simpler grounds.

The same term-by-term identification can be carried out for $\tilde{r}^*$. By equating (4.3) and (4.5) we derive a second Hankel system

$$\begin{pmatrix} \tilde{c}_{m-n+1}^* & & \tilde{c}_m^* \\ & \iddots & \\ \tilde{c}_m^* & & \tilde{c}_{m+n-1}^* \end{pmatrix} \begin{pmatrix} \tilde{e}_n^* \\ \vdots \\ \tilde{e}_1^* \end{pmatrix} = -\begin{pmatrix} \tilde{c}_{m+1}^* \\ \vdots \\ \tilde{c}_{m+n}^* \end{pmatrix}. \tag{4.12}$$

Now here is the key argument. Since the matrix (4.11) is nonsingular, its condition number is finite. Moreover, from (4.10) it follows that in any norm both the right hand sides and the matrices of (4.11) and (4.12) differ by only $O(\varepsilon)$ as $\varepsilon \to 0$. Combining these facts implies that (4.12) also has a unique solution for all sufficiently small ε, and furthermore that (4.6) holds.

To prove (4.7) and (4.8), we observe that additional coefficients of $\tilde{r}^*$ satisfy the recurrence relation

$$-\tilde{c}_{k+1}^* = \tilde{c}_{k-n+1}^* \tilde{e}_n^* + \ldots + \tilde{c}_k^* \tilde{e}_1^* \qquad \forall k \geqq m+n, \tag{4.13}$$

which describes additional rows that might be added to the system (4.12). (4.7) follows directly from (4.6), (4.10), (4.13) (with $k=m+n$), and the parallel relation to (4.13) relating the Padé coefficients $\{c_j^p\}$ and $\{e_j^p\}$. (4.13) also implies

$$|\tilde{c}_{k+1}^*| \leqq \left(\sum_{j=1}^{n} |\tilde{e}_j^*|\right) \max_{k-n+1 \leqq j \leqq k} |\tilde{c}_j^*| \qquad \forall k \geqq m+n.$$

Combined with (4.6) and (4.10), this implies that the coefficients $\tilde{c}_k^*$ satisfy a bound of the form

$$|\tilde{c}_k^*| \leqq \text{const}\,(\text{const})^k \qquad \forall k > m+n$$

uniformly in ε, for all sufficiently small ε, and this implies (4.8). ∎

Lemma 4.2. *Assume the conditions of Assumption* A. *Then there exists a constant* $\beta > 0$ *such that for all sufficiently small* ε, $f - \tilde{r}^*$ *has exactly* n *poles and* $K-m-1$ *zeros in* $1 < |z| < \infty$, *and they all lie in* $\beta/\varepsilon < |z| < \infty$.

Proof. Lemma 4.1 established that the coefficients of the two polynomials $1 + e_1^p z + \ldots + e_n^p z^n$ and $1 + \tilde{e}_1^* z + \ldots + \tilde{e}_n^* z^n$ differ by only $O(\varepsilon)$ as $\varepsilon \to 0$. Since the first one has a full n zeros, so must the second, for all sufficiently small ε. Moreover, their zeros $\{\zeta_i^p\}$ and $\{\tilde{\zeta}_i^*\}$ must coalesce as $\varepsilon \to 0$, with a worst-case convergence rate of

$$|\tilde{\zeta}_i^* - \zeta_i^p| = O(\varepsilon^{1/n})$$

in the event of n-fold multiplicity. Rescaling by ε, it follows that the zeros of $1 + \tilde{e}_1^*(\varepsilon z) + \ldots + \tilde{e}_n^*(\varepsilon z)^n$ converge to those of $1 + e_1^p(\varepsilon z) + \ldots + e_n^p(\varepsilon z)^n$ at a rate $O(\varepsilon^{1/n-1})$. The latter have modulus greater than $\rho\varepsilon^{-1}$, where ρ is any number smaller than the moduli of all poles of r^p when $\varepsilon = 1$. This proves the claim about pole location, taking any $\beta \leqq \rho$.

To determine the location of the zeros of $f - \tilde{r}^*$, we use a winding number argument. Let $\|\cdot\|_r$ denote the supremum norm over the circle about the origin of radius r. Applying Cauchy's estimate to (4.9) as in the last proof, we derive

$$|c_k - \tilde{c}_k^*| < \text{const} \times \varepsilon^{m+n+1-k}.$$

This implies that for $|z| = \beta/\varepsilon$, where β is any fixed positive number,

$$|(c_k - \tilde{c}_k^*)(\varepsilon z)^k| < \text{const} \times \varepsilon^{m+n+1}(\varepsilon/\beta)^{-k},$$

and from this the bound

$$\left\| \sum_{k=-\infty}^{m+n} (c_k - \tilde{c}_k^*)(\varepsilon z)^k \right\|_{\beta/\varepsilon} < \text{const} \times \varepsilon \beta^{m+n}/(1-\varepsilon/\beta) \tag{4.14}$$

follows provided $\varepsilon < \beta$. Now since the poles of $f - \tilde{r}^*$ all lie outside $|z| = \rho/\varepsilon$, we also have

$$|c_k - \tilde{c}_k^*| < \text{const} \times \rho^{-k} \qquad \forall k \geqq m+n+2 \tag{4.15}$$

for some const > 0, and from this a complementary bound

$$\left\| \sum_{k=m+n+2}^{\infty} (c_k - \tilde{c}_k^*)(\varepsilon z)^k \right\|_{\beta/\varepsilon} < \text{const} \times \left(\frac{\beta}{\rho}\right)^{m+n+2}/(1-\beta/\rho) \tag{4.16}$$

follows, this time for any β with $0 < \beta < \rho$.

(4.7) implies that the degree $m+n+1$ term of $f-\tilde{r}^*$, on the other hand, satisfies

$$|(c_{m+n+1}-\tilde{c}^*_{m+n+1})(\varepsilon z)^{m+n+1}| = |c_{m+n+1}-c^p_{m+n+1}|\,\beta^{m+n+1}+O(\varepsilon)$$

uniformly for $|z|=\beta/\varepsilon$. If β is fixed small enough, this estimate must be greater than the sum of (4.14) and (4.16), for all sufficiently small ε. This implies that $f-\tilde{r}^*$ has the same winding number as $(\varepsilon z)^{m+n+1}$ on $|z|=\beta/\varepsilon$, namely $m+n+1$.

Now $f-\tilde{r}^*$ is the finite Blaschke product given in formula (3.2) of Theorem 3.2. Its winding number on any circle is the number of poles minus the number of zeros in the region of the extended plane outside that circle. Outside $|z|=\beta/\varepsilon$, $f-\tilde{r}^*$ has K poles at infinity (since $c_K\neq 0$), n finite poles (proved above), and hence by (3.2) at most $(K+n-m-1)-n=K-m-1$ zeros, which correspond by symmetry to poles of $f-\tilde{r}^*$ in the unit disk. Therefore on $|z|=\beta/\varepsilon$, $f-\tilde{r}^*$ must have winding number at least $(K+n)-(K-m-1)=m+n+1$, and it can only be that small, as we have just shown that in fact it is, if a full $K-m-1$ zeros lie outside $|z|=\beta/\varepsilon$. ∎

Lemma 4.3. *Assume the conditions of Assumption* A. *Then there exist constants* $M<\infty$, $\gamma<\infty$ *such that for all sufficiently small* ε,

$$|c_k-\tilde{c}^*_k| \leqq M\varepsilon^{2m+2n+2-k}(\gamma\varepsilon)^{-k}$$

for all integers $-\infty<k\leqq m+n+1$.

Proof. This result follows from Lemma 4.2 and the Blaschke product representation of Theorem 3.2. By Lemma 4.2, for some fixed β and all sufficiently small $\varepsilon>0$, $f-\tilde{r}^*$ must have n zeros $\zeta_1,\ldots,\zeta_n$ in $0<|z|<\varepsilon/\beta$ and $K-m-1$ zeros $\zeta_{n+1},\ldots,\zeta_{K+n-m-1}$ in $\beta/\varepsilon<|z|<\infty$. From Theorem 3.2, we may write $f-\tilde{r}^*$ as

$$(f-\tilde{r}^*)(z)=\sigma z^K\prod_{j=1}^{n}\left(\frac{z-\zeta_j}{\bar{\zeta}_j z-1}\right)\prod_{j=n+1}^{K+n-m-1}\left(\frac{z-\zeta_j}{\bar{\zeta}_j z-1}\right).$$

Consider the size of this expression on the circle $|z|=2\varepsilon/\beta$. Easy estimates show that on this circle the four factors

$$\sigma,\ z^K,\ \prod_{j=1}^{n}\left(\frac{z-\zeta_j}{\bar{\zeta}_j z-1}\right),\ \prod_{j=n+1}^{K+n-m-1}\left(\frac{z-\zeta_j}{\bar{\zeta}_j z-1}\right)$$

have magnitudes $O(\varepsilon^{m+n+1})$ (by (4.9)), $O(\varepsilon^K)$, $O(\varepsilon^n)$, and $O(\varepsilon^{-(K-m-1)})$, respectively. Combining these bounds gives

$$(f-\tilde{r}^*)(z)=O(\varepsilon^{2m+2n+2}) \quad \text{on } |z|=2\varepsilon/\beta.$$

By Cauchy's estimate there follows

$$|(c_k-\tilde{c}^*_k)\,\varepsilon^k| = O(\varepsilon^{2m+2n+2})\times(2\varepsilon/\beta)^{-k}$$

uniformly for all k, which proves the theorem with $\gamma=2/\beta$. ∎

A final corollary will be needed for the a posteriori argument of Sect. 6.

Lemma 4.4. *Assume the conditions of Assumption* A. *Then as* $\varepsilon \to 0$,

$$(f-\tilde{r}^*)(z)=(c_{m+n+1}-c^p_{m+n+1})(\varepsilon z)^{m+n+1}+O(\varepsilon^{m+n+2})$$

uniformly on S.

Proof. From Lemma 4.3 it follows easily that the magnitude on S of the sum of all terms of degree $\leqq m+n$ in the Laurent expansion of $f-\tilde{r}^*$ is $O(\varepsilon^{m+n+2})$ as $\varepsilon \to 0$. (4.8), on the other hand, implies the same for the sum of all terms of degree $\geqq m+n+2$. The claim follows from these observations and (4.7). ∎

5. The Near-Best Approximation $r^{cf} \in R_{mn}$

The purpose of studying the extended best approximation $\tilde{r}^* \in \tilde{R}_{mn}$ is to derive from it a nearby approximation r^{cf} that belongs to R_{mn}. The foregoing asymptotic results will show that, at least if f is smooth, such an r^{cf} can be chosen that is nearly equal to $\tilde{r}^*$ on S.

If $m=n$, then $\tilde{r}^*$ can be written in the form (2.1)

$$\tilde{r}^*_{nn}=r_{nn}+g$$

where $r_{nn} \in R_{nn}$ and $g \in G$. In this case a natural choice for r^{cf} would be $r^{cf}_{nn} \equiv r_{nn}$. A generalization of this choice can be defined in any problem in which $m \geqq n-1$. Let $\tilde{r}^*_{mn}$ be expanded as in (4.3) in a Laurent series with respect to S. Then if $m \geqq n-1$, the nonnegative degree terms of this series must define a rational function belonging to R_{mn}. For $m \geqq n$ this follows from (2.1) and (2.2), which imply in this case that the nonnegative degree portion of the series can be written in the form $z^{m-n} r_{nn} + p_{m-n-1}$, where p_{m-n-1} is a polynomial of degree at most $m-n-1$. This sum is in R_{mn}. The case $m=n-1$ is similar. For $m<n-1$, $\tilde{R}_{mn} \neq R_{mn}+G$, however, as was remarked in Section 2, and because of this, truncating negative degree terms in $\tilde{r}^*_{mn}$ does not in general yield a function in R_{mn}.

To make possible a uniform treatment for any m and n, therefore, we shall define r^{cf} by a different truncation. Let $\tilde{r}^*_{mn}$ be written as a quotient as in (2.3). r^{cf} will be constructed by simply dropping all terms of negative degree in the numerator:

$$r^{cf}_{mn}(z) \equiv \frac{\tilde{d}^*_0+\ldots+\tilde{d}^*_m z^m}{1+\tilde{e}^*_1 z+\ldots+\tilde{e}^*_n z^n}. \tag{5.1}$$

(In degenerate cases with fewer than n finite poles outside $\bar{D}$, r^{cf} is not uniquely defined.) This choice of r^{cf} will prove sufficient for deriving asymptotic results concerning near-circularity of the error curve of r^*.

The Carathéodory-Fejér theory of Sect. 3 was developed only for polynomials f, although it was remarked that comparable statements hold for arbitrary functions $f \in L^\infty(S)$. For asymptotic results with $\varepsilon \to 0$, however, any function that is analytic in a neighborhood of the origin becomes arbitrarily close to a polynomial as $\varepsilon \to 0$, so there is no need to restrict the consideration to polynomials. Specifically, given $\hat{f}$ analytic at $z=0$, let $\hat{f}^T$ be the degree-$(2m$

$+2n+2$) partial sum of its Maclaurin series, and define $f(z)=\hat{f}(\varepsilon z)$ as before and $f^T(z)=\hat{f}^T(\varepsilon z)$. Then

$$\|f-f^T\|=O(\varepsilon^{2m+2n+3}) \tag{5.2}$$

as $\varepsilon\to 0$. From now on define $\tilde{r}^*$ and r^{cf} to be the rational functions obtained by applying the CF theory and (5.1) to f^T. (K is then the degree of the largest nonzero term in f^T, at most $2m+2n+2$.) The bound (5.2) is small enough so that the strength of our subsequent asymptotic theorems will not be affected by the $f\mapsto f^T$ truncation. In a particular computational example, of course, one might choose K larger than $2m+2n+2$ and expect a slight gain in accuracy.

With these definitions it is straightforward to derive the asymptotic behavior of the error curve of r^{cf} from previous results:

Lemma 5.1. *Let f be analytic at the origin and assume the conditions of Assumption* A. *Then as $\varepsilon\to 0$,*

$$(i)\ (f-r^{cf})(z)=(c_{m+n+1}-c^p_{m+n+1})(\varepsilon z)^{m+n+1}+O(\varepsilon^{m+n+2})$$

uniformly on S, and

$$(ii)\ \|f-r^{cf}\|-\min_{z\in S}|(f-r^{cf})(z)|=O(\varepsilon^{2m+2n+3}).$$

Thus the error curve of r^{cf} is nearly circular with a relative deviation in radius that is $O(\varepsilon^{m+n+2})$, and it has winding number exactly $m+n+1$ for sufficiently small ε.

Proof. (*i*) and (*ii*) follow from Lemma 4.4 and the exact circularity of $f^T-\tilde{r}^*$ on S, respectively, together with (5.2) and the bound

$$\|\tilde{r}^*-r^{cf}\|=O(\varepsilon^{2m+2n+3}) \tag{5.3}$$

as $\varepsilon\to 0$. Let us establish this bound. By (4.5) and (5.1), $\tilde{r}^*-r^{cf}$ has an expansion of the form

$$(\tilde{r}^*-r^{cf})(z)=\frac{\ldots+\tilde{d}^*_{-2}(\varepsilon z)^{-2}+\tilde{d}^*_{-1}(\varepsilon z)^{-1}}{1+\tilde{e}^*_1(\varepsilon z)+\ldots+\tilde{e}^*_n(\varepsilon z)^n}. \tag{5.4}$$

From (4.6), each $\tilde{e}^*_k$ is bounded as $\varepsilon\to 0$, which implies that the denominator of (5.4) behaves like $1+O(\varepsilon)$ on S as $\varepsilon\to 0$. Therefore (5.3) will follow if

$$\|\ldots+\tilde{d}^*_{-2}(\varepsilon z)^{-2}+\tilde{d}^*_{-1}(\varepsilon z)^{-1}\|=O(\varepsilon^{2m+2n+3}). \tag{5.5}$$

Now by (4.3) and (4.5) we have readily

$$\tilde{d}^*_k=\tilde{c}^*_k+\tilde{c}^*_{k-1}\tilde{e}^*_1+\ldots+\tilde{c}^*_{k-n}\tilde{e}^*_n,$$

and from this, the boundedness of the $\tilde{e}^*_k$, and Lemma 4.3, (5.5) follows. ∎

By Proposition 2.2, Lemma 5.1 implies the very strong result

Proposition 5.2 (r^{cf} *is near best*). *Let f be analytic at the origin and assume the conditions of Assumption* A. *Then as $\varepsilon\to 0$,*

$$\|f-r^{cf}\|-\|f-r^*\|=O(\varepsilon^{2m+2n+3})$$

and thus $\|f-r^{cf}\|$ *exceeds the minimal error by a relative magnitude* $O(\varepsilon^{m+n+2})$ *as* $\varepsilon \to 0$. ∎

6. Main Asymptotic Results

At this point an approximation r^{cf} to f has been constructed whose error curve is nearly circular. Now we will show that it follows that $\|r^{cf}-r^*\|$ is nearly zero, hence that the error curve of r^* is also nearly circular. As in the last section, here $\hat{f}$ is any function analytic at 0, and $f(z)\equiv\hat{f}(\varepsilon z)$.

A lemma is needed on the behavior of the denominators of r^* and r^{cf}.

Lemma 6.1. *Let* f *be analytic at the origin and assume the conditions of Assumption* A. *For each* $\varepsilon>0$, *write* $r^{cf}=p^{cf}/q^{cf}$ *and* $r^*=p^*/q^*$ *with* q^{cf} *and* q^* *normalized to have constant term* 1. *Then as* $\varepsilon\to 0$,

$$q^{cf}(z)=1+O(\varepsilon) \tag{6.1}$$

and

$$q^*(z)=1+O(\varepsilon) \tag{6.2}$$

uniformly on S.

Proof. q^{cf} is also the denominator of $\tilde{r}^*$, so (6.1) follows from (4.5) and (4.6) of Lemma 4.1. (4.6) was derived by means of Hankel determinants by comparing $\tilde{r}^*$ to r^p. The only facts about $\tilde{r}^*$ used for this were $\tilde{r}^*\in\tilde{R}_{mn}$ and the best approxiamtion property $\|f-\tilde{r}^*\|\leqq\|f-r^p\|$. As both facts hold also for r^*, the same argument proves (6.2). ∎

Lemma 6.1 suggests that asymptotically as $\varepsilon\to 0$, rational approximation becomes less and less nonlinear. This fact enables us to show that $\|r^*-r^{cf}\|$ is small by the same a posteriori argument used for polynomial approximation in [23].

Theorem 6.2 ($r^{cf}\approx r^*$). *Let* f *be analytic at the origin and assume the conditions of Assumption* A. *Then as* $\varepsilon\to 0$,

$$\|r^{cf}-r^*\|=O(\varepsilon^{2m+2n+3}). \tag{6.3}$$

Proof. Let $\Delta c=c_{m+n+1}-c^p_{m+n+1}$; from Assumption A, $\Delta c\neq 0$. By Lemma 5.1*i*,

$$\frac{(f-r^{cf})(z)}{\Delta c(\varepsilon z)^{m+n+1}}=1+O(\varepsilon), \tag{6.4}$$

and by Lemma 5.1*ii* this function varies in modulus by only $O(\varepsilon^{m+n+2})$ on S. Since r^* is a best approximation to f, it follows that adding $(r^{cf}-r^*)(z)/[\Delta c(\varepsilon z)^{m+n+1}]$ to (6.4) can increase the modulus of (6.4) by no more than $O(\varepsilon^{m+n+2})$ at any point on S. In particular, we must have

$$\left|\frac{(r^{cf}-r^*)(z)}{\Delta c(\varepsilon z)^{m+n+1}}\right|=O(\varepsilon^{m+n+2})$$

uniformly over all points where the argument of this quotient has modulus less than, say, $\pi/4$. Since $r^{cf}-r^*=(p^{cf}q^*-p^*q^{cf})/q^{cf}q^*$, and since $(q^{cf}q^*)(z)=1+O(\varepsilon)$ as $\varepsilon\to 0$ by Lemma 6.1, it follows that

$$\frac{(p^{cf}q^*-p^*q^{cf})(z)}{\Delta c(\varepsilon z)^{m+n+1}}=O(\varepsilon^{m+n+2})$$

uniformly over all points where this quotient is real and positive. Now this quotient is a polynomial P in z^{-1} of degree at most $m+n+1$ with constant term 0. It follows (Lemma 5, [23]) that the image of $|z^{-1}|\leqq 1$ under P covers completely the disk about $z^{-1}=0$ of radius $2^{-(m+n+1)}\|P\|$. This necessitates the bound

$$\left\|\frac{p^{cf}q^*-p^*q^{cf}}{\Delta c(\varepsilon z)^{m+n+1}}\right\|=O(\varepsilon^{m+n+2}),$$

hence

$$\|p^{cf}q^*-p^*q^{cf}\|=O(\varepsilon^{2m+2n+3}),$$

and hence (6.3), again since $(q^{cf}q^*)(z)=1+O(\varepsilon)$. ∎

Lemma 5.1 and Theorem 6.2 imply immediately

Theorem 6.3 (the error curve of r^* is nearly circular). *Let f be analytic at the origin and assume the conditions of Assumption* A. *Then as $\varepsilon\to 0$, for all sufficiently small ε, the error curve of r^* has winding number exactly $m+n+1$, and*

$$\|f-r^*\|-\min_{|z|=1}|(f-r^*)(z)|=O(\varepsilon^{2m+2n+3}).\quad\blacksquare$$

Summary of Asymptotic Results. Let us summarize what we have established, or could readily establish with additional combinations of the foregoing arguments, about asymptotic approximation of a function f analytic at the origin that satisfies Assumption A. As $\varepsilon\to 0$

$$\|f\|=O(1), \tag{6.5}$$

whereas

$$\left.\begin{aligned}&\|f-r^*\|\\&\|f-r^p\|\\&\|f-r^{cf}\|\end{aligned}\right\}=O(\varepsilon^{m+n+1})\qquad\text{but not } O(\varepsilon^{m+n+2}).\qquad\begin{aligned}&(6.6a)\\&(6.6b)\\&(6.6c)\end{aligned}$$

r^p and r^* have relative contact $O(\varepsilon)$ on the scale of these errors,

$$\|r^p-r^*\|=O(\varepsilon^{m+n+2}), \tag{6.7}$$

while r^{cf} and r^* have relative contact $O(\varepsilon^{m+n+2})$,

$$\|r^{cf}-r^*\|=O(\varepsilon^{2m+2n+3}). \tag{6.8}$$

For sufficiently small ε, r^* and r^p and r^{cf} all have exactly n finite poles, and all of their error curves have winding number exactly $m+n+1$. The error curve of r^p deviates from a circle by a relative radius $O(\varepsilon)$,

$$\|f-r^p\|-\min_{z\in S}|(f-r^p)(z)|=O(\varepsilon^{m+n+2}), \tag{6.9}$$

and those of r^{cf} and r^* deviate by a relative radius $O(\varepsilon^{m+n+2})$,

$$\left.\begin{aligned} &\|f-r^{cf}\|-\min_{z\in S}|(f-r^{cf})(z)| \qquad (6.10a)\\ &\|f-r^*\|-\min_{z\in S}|(f-r^*)(z)| \qquad (6.10b)\end{aligned}\right\}=O(\varepsilon^{2m+2n+3}).$$

Discussion. The agreement (6.7) of r^* and r^p was essentially established by Walsh [24], through not stated in this form. It is likely that Walsh also knew of the following corollary, which follows form the winding number results for $f-r^*$ by the argument principle, but he seems not to have published it:

Corollary. *Let f be analytic at the origin and assume the conditions of Assumption* A. *Then for all sufficiently small ε, r^* interpolates f at exactly $m+n+1$ points in D, counted with multiplicity.* ∎

This extends a result of Motzkin and Walsh [18] for polynomial approximation, discussed also in [23].

It is interesting to see that although the best approximation r^* need not be unique, (6.8) implies that it is nearly so: if r_1^* and r_2^* are any two best approximations for each $\varepsilon>0$, then

$$\|r_1^*-r_2^*\|=O(\varepsilon^{2m+2n+3}).$$

Our presentation has described approximation on the fixed disk D of a function that grows smoother as $\varepsilon\to 0$, but obviously the results pertain equally to approximation of a fixed function f on the shrinking disk $|z|\leqq\varepsilon$. This was the setting considered by Walsh. He showed then that as $\varepsilon\to 0$, $r^*\to r^p$ uniformly on any compact set not containing poles of r^p [24]. Our arguments duplicate this conclusion, showing that $r^p-r^*=O(\varepsilon)$ on any such compact set. A third application of the argument of Lemma 4.1, in fact, shows the much stronger result that $r^{cf}-r^*=O(\varepsilon^{m+n+2})$ on such sets. This is another way of expressing the fact that whereas Padé approximation captures analytically the first term in an asymptotic description of r^*, CF approximation captures the first $m+n+2$ terms.

7. Numerical Computation of $\tilde{r}^*$

Here we sketch how the coefficients of $\tilde{r}^*$ as a quotient of the form (2.3) can be computed numerically. Additional details, in the context of digital filter design, are given in [10].

Step 1. First, one must decide at what degree K to truncate the Maclaurin series of f, and then find the $K+1$ required coefficients. In realistic applications the series may converge fairly slowly on $\bar{D}$, so we must assume that K may be fairly large: say, between 10 and 100. If the Maclaurin coefficients are not known analytically, they can be computed by the Fast Fourier Transform [13, § 3.1].

Step 2. The most time-consuming part of the problem is to find the $(n+1)$st singular value and vector of the $K+n-m$ by $K+n-m$ Hankel matrix $H = H_f^{(n-m)}$. The most straightforward approach to this is to compute a full singular value decomposition of H by unitary reduction to bidiagonal form followed by a QR iteration, an algorithm developed by Golub and Kahan and implemented in both EISPACK [21] and LINPACK. This will take $O(K^3)$ floating point operations. However, our problem is special in three ways: H is Hankel, it is triangular, and we only need one singular value and vector. One would like to take advantage of as much of this structure as possible. If the coefficients of f are real, then an additional fact to be exploited is that the singular value problem reduces to an eigenvalue problem.

Unfortunately, no methods are currently known that take substantial advantage of general Hankel (or Toeplitz) structure in singular value or eigenvalue problems. We have contented ourselves with reducing the $O(K^3)$ time constant by computing only some eigenvalues and one eigenvector of H via Sturm sequencing and inverse iteration [21], in the case where f has real coefficients. Even this saving is not as great as one would like, however, for while we seek the $(n+1)$st eigenvalue of H in magnitude, the Sturm sequence approach isolates eigenvalues according to magnitude and sign. Thus we are forced to search both ends of the spectrum of H in order to determine which eigenvalue it is that we want. EISPACK provides routines for this.

Step 3. Now one must extract the coefficients of $\tilde{r}^*$ from (3.2). Let us write the denominator of the Blaschke product in (3.2) in the form $q_{\text{in}}(z)\, q_{\text{out}}(z)$, where q_{in} and q_{out} are polynomials with all zeros inside and outside $\bar{D}$, respectively. Then the denominator of $\tilde{r}^*$ is precisely q_{out}, or may be taken to be such in the degenerate case in which a zero of q_{out} is cancelled by an identical zero in the numerator of the Blaschke product. Thus to find the denominator of $\tilde{r}^*$, we need to find the polynomial subfactor of $\bar{u}_k + \ldots + \bar{u}_1 z^{k-1}$ containing precisely those zeros outside $\bar{D}$. For this we have used an excellent technique proposed by Henrici [13, §3.2] based on forming the logarithmic derivative of $\bar{u}_k + \ldots + \bar{u}_1 z^{k-1}$ and computing certain of its Laurent coefficients, making use of the Fast Fourier Transform. The accuracy of this procedure depends on the zeros of $\bar{u}_k + \ldots + \bar{u}_1 z^{k-1}$ lying not too close to S, but this is a limitation one can live with, as the CF method itself will give poor approximations when some of these zeros have magnitude close to 1.

Once q_{out} is known, the numerator of $\tilde{r}^*$ can be found multiplying (3.2) by q_{out}. The resulting equation gives $\tilde{r}^* q_{\text{out}}$ in the form of a Laurent series that converges in $|z| > 1$, which is precisely the form that we seek. The fastest way to compute desired coefficients of this Laurent series is by means of another FFT.

Step 4. Finally, r^{cf} is formed by dropping the terms of negative degree in the numerator of $\tilde{r}^*$.

The total time required for these computations depends strongly on K. For approximation of e^z, $K = 20$ is ample to give r^{cf} accurate to ten places when m and n are small, and the computation requires roughly 0.1 secs on an IBM 370/168. A typical time for a function with a less quickly converging power series might be 1 second.

8. Numerical Example: Approximation of e^z on the Disk

To illustrate the foregoing results, let us see how the CF method performs in approximating e^z. Because the Maclaurin series of e^z decreases so rapidly, $K=25$ is equivalent to $K=\infty$ for our numerical purposes. We will talk as if $K=\infty$. Combining Proposition 2.2 and Theorem 3.3, therefore, we get the simple bound

$$\sigma_{n+1} \leqq \|e^z - r^*\| \leqq \|e^z - r^{cf}\| \tag{8.1}$$

provided the error curve of r^{cf} has the expected winding number $m+n+1$. (Clearly it does, although this has not been proved for all m, n.)

Sample Computations for $(m, n)=(1, 1)$

Consider approximation of type (1, 1). The simplest candidate is the Padé approximant,

$$r^p_{11}(z)=\frac{1+0.5z}{1-0.5z}. \tag{8.2a}$$

The corresponding error (to the accuracy shown) is

$$\|e^z - r^p_{11}\| = 0.282. \tag{8.2b}$$

The error curve, as shown already in Fig. 1, is not close to circular.

Next, we compute the extended best approximation $\tilde{r}^*$ by the method described in Sect. 7. It is

$$\tilde{r}^*_{11}=\frac{\ldots+0.00000983z^{-2}+0.00024668z^{-1}+0.99613054+0.58955195z}{1-0.43416584z} \tag{8.3a}$$

with corresponding error

$$\|e^z - \tilde{r}^*_{11}\| = \sigma_2(H^{(0)}_{e^z}) = 0.08455. \tag{8.3b}$$

The error curve here winds 3 times, and is a perfect circle (assuming $K=\infty$).

Now by truncating negative powers we get

$$r^{cf}_{11}(z)=\frac{0.99613054+0.58955195z}{1-0.43416584z}, \tag{8.4a}$$

with error

$$\|e^z - r^{cf}_{11}\| < 0.08493. \tag{8.4b}$$

Evidently r^{cf} approximates e^z to within 0.5% of the minimal error. Its error curve is circular to within a relative deviation of less than 1%. The Padé approximation (8.2), in contrast, is non-optimal by more than a factor of three. If we were to increase m, the comparison would become even more striking.

For a true appraisal of r^{cf}, we need to compare it to the best approximation r^*. In general, computing r^* numerically to the high accuracy required for this comparison is a difficult matter. The Remes algorithm for rational Chebyshev approximation on a real interval, for example, does not extend to

complex approximation. Stephen Ellacott (private communication) has tackled this problem with Lawson's algorithm, but as we have discussed elsewhere [23], Lawson's algorithm ceases to converge precisely as the error curve approaches a circle. Because of this problem, Ellacott's coefficients for r^* are generally not as close to correct as the more easily computed coefficients of the approximations $r^{cf} \approx r^*$ are, for $m \geqq 3$. (In defense of Lawson's algorithm, however - it does get *near* a best approximation quickly, and the last fraction of a percent is unimportant for most purposes.)

For the case $(m, n) = (1, 1)$, Ellacott's computation is sufficiently accurate, and we have checked it against a "brute force" computation by a general-purpose optimization program. We find

$$r_{11}^*(z) = \frac{0.99625 + 0.58952z}{1 - 0.43414z} \tag{8.5a}$$

with error

$$\|e^z - r_{11}^*\| = 0.08480. \tag{8.5b}$$

A comparison of (8.5) with (8.4) suggests that in practice, r^{cf} can be expected to approximate r^* very closely. Note that (8.5b) lies between (8.3b) and (8.4b), as it must.

In this example $\|e^z - r^{cf}\|$ would fall all the way to 0.08481, and the relatively poor constant term in the numerator of r^{cf} would rise to 0.99624, if the choice of r^{cf} first considered in Sect. 5 were used rather than the one finally adopted there for convenience of generalization. Undoubtedly the first choice is better, in practice, when $m \geqq n$.

Dependence on m and n

Let the measure α of relative circularity of equation (1.1) be applied to r^{cf} for various m, n. Table 2 shows the results. Note the general agreement between the numbers of Table 2 and those of Table 1. This is an indication that at least for the present problem, the CF method is an effective approach to the phenomenon of near-circularity. It may appear worrisome that in both tables α is roughly independent of n, for this seems to contradict the asymptotic results of Sect. 5. The explanation is that the constants in those results increase with n. Dependence on n is always more difficult to analyze than dependence on m in

Table 2. Relative deviation α from a perfect circle (Eq. (1.1)) of error curves of CF approximations r_{mn}^{cf} to e^z on the unit disk. Various m, n

$m=$ \ $n=$	0	1	2	3
0	1(−1)	6(−1)	6(−1)	3(−1)
1	4(−3)	1(−2)	1(−2)	9(−3)
2	3(−5)	1(−4)	6(−5)	7(−5)
3	5(−6)	5(−7)	2(−6)	3(−7)

Table 3. $\sigma_{n+1} = \|e^z - \tilde{r}^*_{mn}\|$ for various m, n (see (8.1)). Underlined digits are known to agree with corresponding digits of best approximation errors $\|e^z - r^*_{mn}\|$

$m=$ \ $n=$	0	1	2	3
0	1.25836 65707	0.39659 05141	0.11527 04209	0.02919 04410
1	0.55752 90694	0.08454 87259	0.01295 01410	0.00186 66235
2	0.17737 38152	0.01459 00251	0.00139 32413	0.00013 47402
3	0.04336 8926832	0.00218 6196115	0.00014 2307100	0.00000 9931757

Table 4. Relative deviation α from a perfect circle (Eq. (1.1)) of error curves of CF approximations r^{cf}_{11} to $e^{\varepsilon z}$ on the unit disk. Various ε

ε	$\|f - r^{cf}\|$	α	$\sqrt[4]{\alpha}/\varepsilon$
4	0.170(+2)	0.723	0.461
2	0.810	0.123	0.296
1	0.849(−1)	0.978(−2)	0.314
1/2	0.104(−1)	0.594(−3)	0.312
1/4	0.130(−2)	0.349(−4)	0.307
1/8	0.163(−3)	0.215(−5)	0.306

rational approximation, because it is there that the nonlinearity lies. As it happens, in this example α begins to decrease steadily if n is increased further.

To sum up how close to best r^{cf} may be, Table 3 shows $\|e^z - \tilde{r}^*\|$ as a function of m and n for $m, n \leqq 3$. Thus the (m, n) entry in the table is just the $(n+1)$st singular value of a Hankel matrix $H^{(n-m)}_{e^z}$. Digits in which $\|e^z - \tilde{r}^*\|$ is known to agree with $\|e^z - r^*\|$ (usually on the basis of (8.1)) have been underlined.

Asymptotic Behavior for $(m, n) = (1, 1)$

Finally, it is easy to confirm that as $\varepsilon \to 0$, $\|e^{\varepsilon z} - r^{cf}\| - \min_{z \in S} |e^{\varepsilon z} - r^{cf}| = O(\varepsilon^{2m+2n+3})$, as predicted in Lemma 5.1. Table 4 lists $\|e^{\varepsilon z} - r^{cf}\|$ and the same α as in Table 2 as a function of ε for $(m, n) = (1, 1)$. The final column shows that as expected, α decreases like $\varepsilon^{1+1+2} = \varepsilon^4$. Moreover, it shows that the constant involved is small. Evidently D is already a small disk as far as the smoothness of e^z is concerned.

9. Additional Remarks

The phenomenon of nearly circular error curves has been observed by a few people over the years, but not speculated about in print until [23]. Since the tendency to near-circularity is so strong, it is likely that interesting features of the Chebyshev approximation problem have been overlooked as a consequence. The approach described here does not yield a satisfying explanation of "why" error curves are nearly circular, but perhaps the results it leads to

can at least make this phenomenon a recognized feature of the Chebyshev approximation problem. Any such feature must have theoretical and practical consequences. One practical consequence, as mentioned in the last section, is that Lawson's algorithm is not suitable for computing Chebyshev approximations to high accuracy.

That is the geometric aspect of this work; the other theme has been the algebraic one, namely the remarkable connection with Blaschke products and the singular value analysis of a Hankel matrix of Maclaurin coefficients. The Carathéodory-Fejér theorem and related results belong traditionally to the study of function theory on the unit disk, and are only recently being borrowed for other purposes. The papers of Adamian, Arov and Krein [1] are currently inspiring active work in systems theory by Bettayeb, Bultheel, de Wilde, Genin, Kung, Silverman, and perhaps others [4, 7, 10, 16, 20]. Problems in systems theory reside naturally on the unit disk, however, whereas in approximation theory they need not. At least some of the techniques described here can be transplanted to more general regions by a conformal map, as for example in Theorem 12 of [23], but algebraic simplicity is lost in the process. It remains to be seen how fruitful such transplantation can be.

For the disk there is no doubt of the power of the CF approach. The great weakness of the theorems proved here is that with the exception of Theorem 3.3, they are entirely asymptotic. If non-asymptotic results can be found that capture the true strength of CF approximation, the method might become a powerful theoretical tool. For example, it might then be easy to prove strong theorems about best approximation in the more difficult asymptotic cases $m \to \infty$ and $n \to \infty$. Some conjectures along this line can be found in the book by Meinardus [17, e.g. (9.14)]. Many of the estimates in this paper have been far from best possible, so it has not appeared worthwhile to tackle these problems, as was done for the polynomial case in [23].

Most surprisingly, it turns out that the CF method extends with no loss of algebraic simplicity to approximation by real functions on a real interval. Now a Hankel matrix of coefficients in an expansion in Chebyshev polynomials is needed. In fact, the CF idea turns out to be even more powerful in real approximation than in complex approximation. The real CF method is presented by Gutknecht and Trefethen in [11] and discussed further by Gutknecht in [9].

Acknowledgments. Much of this work has been developed jointly with Martin Gutknecht of the Eidgenössische Technische Hochschule in Zurich, Switzerland, with whom it has been a pleasure and a valuable education to work. I am also grateful for support and advice to Gene Golub, Peter Henrici, Joseph Oliger, Edward Saff, and Paul van Dooren. Computations were performed at the Stanford Linear Accelerator Center of the U.S. Department of Energy. Work was supported in part by an NSF Graduate Followship and in part by Office of Naval Research Contract N00014-75-C-1132.

References

1. Adamian, V.M., Arov, D.Z., Krein, M.G.: Analytic properties of Schmidt pairs for a Hankel operator and the generalized Schur-Takagi problem. Math. USSR Sb. **15**, 31–73 (1971)
2. Akhieser, N.I.: On a minimum problem in the theory of functions and on the number of roots of an algebraic equation which lie inside the unit circle. Izv. Akad. Nauk. SSSR, Otd. Mat. Estestv. Nauk **9**, 1169–1189 (1931) (in Russian)

3. Akhieser, N.I.: Theory of approximation. (Appendix D) New York: Ungar, 1956
4. Bultheel, A., de Wilde, P.: On the Adamian-Arov-Krein approximation, identification and balanced realization of a system. IEEE Trans. Circuits and Systems (in press, 1981)
5. Carathéodory, C., Fejér, L.: Über den Zusammenhang der Extremen von harmonischen Funktionen mit ihren Koeffizienten und über den Picard-Landauschen Satz. Rend. Circ. Mat. Palermo **32**, 218–239 (1911)
6. Clark, D.: Hankel forms, Toeplitz forms and meromorphic functions. Trans. Amer. Math. Soc. **134**, 109–116 (1968)
7. Genin, Y.V., Kung, S.Y.: A two-variable approach to the model reduction problem with Hankel norm criterion. IEEE Trans. Circuits and Systems (in press, 1981)
8. Gragg, W.B.: The Padé table and its relation to certain algorithms of numerical analysis. SIAM Rev. **14**, 1–62 (1972)
9. Gutknecht, M.: Carathéodory-Fejér approximation. (in press, 1981)
10. Gutknecht, M.H., Trefethen, L.N.: Recursive digital filter design by the Carathéodory-Fejér method. IEEE Trans. Acoust. Speech Signal Proc. (in press, 1981)
11. Gutknecht, M.H., Trefethen, L.N.: Real polynomial Chebyshev approximation by the Carathéodory-Fejér method. SIAM J. Numer. Anal. (in press, 1981)
12. Henrici, P.: Applied and computational complex analysis, Vol. 1. New York: Wiley, 1974
13. Henrici, P.: Fast Fourier methods in computational complex analysis. SIAM Rev. **21**, 481–527 (1979)
14. Hoffman, K.: Banach spaces of analytic functions. Englewood Cliffs, N.J.: Prentice-Hall, 1962
15. Klotz, V.: Gewisse rationale Tschebyscheff-Approximationen in der komplexen Ebene. J. Approximation Theory **19**, 51–60 (1977)
16. Kung, S.Y.: New fast algorithms for optimal model reduction. Proceedings, 1980 Joint Automatic Control Conference, San Francisco. New York: IEEE, 1980
17. Meinardus, G.: Approximation of functions: theory and numerical methods. Berlin-Heidelberg-New York: Springer, 1967
18. Motzkin, T.S., Walsh, J.L.: Zeros of the error function for Tchebycheff approximation in a small region. Proc. London Math. Soc. **3**, 90–98 (1963)
19. Schur, I.: Über Potenzreihen, die im Innern des Einheitskreises beschränkt sind. J. Reine Angew. Math. **148**, 122–145 (1918)
20. Silverman, L.M., Bettayeb, M.: Optimal approximation of linear systems. IEEE Trans. Automatic Control (in press, 1981)
21. Smith, B.T., Boyle, J.M., Dongarra, J.J., Garbow, B.S., Ikebe, Y., Klema, V.C., Moler, C.B.: Matrix eigensystem routines – EISPACK guide. (Lecture Notes in Computer Sciences, Vol. 6, 2nd ed.) Berlin-Heidelberg-New York: Springer, 1976
22. Takagi, T.: On an algebraic problem related to an analytic theorem of Carathéodory and Fejér. Japan J. Math. **1**, 83–93 (1924) and **2**, 13–17 (1925)
23. Trefethen, L.: Near-circularity of the error curve in complex Chebyshev approximation. J. Approximation Theory (in press, 1981)
24. Walsh, J.L.: Padé approximants as limits of rational functions of best approximation. J. Math. Mech. **12**, 305–312 (1964)

Received October 13, 1980