SIAM J. MATH. ANAL.
Vol. 16, No. 1, January 1985

015

ON CONVERGENCE AND DEGENERACY IN RATIONAL PADÉ AND CHEBYSHEV APPROXIMATION*

LLOYD N. TREFETHEN† AND MARTIN H. GUTKNECHT‡

Abstract. We study two questions associated with rational approximation of a function $f(z)$ near the origin $z=0$: continuity of the Padé approximation operator, and convergence of Chebyshev to Padé approximants as the domain of approximation shrinks to a point. Both become delicate in the case of degenerate approximations, i.e. approximations whose numerator and denominator are deficient in degree. In this situation various distinct definitions of convergence of sequences of rational functions make sense, and we give a unified treatment that explains their interrelationships. Our results show that the answers to the above questions are generally affirmative only in the nondegenerate case.

AMS-MOS subject classification (1980). Primary 41A20, Secondary 30E10, 41A21, 41A50

Introduction. This paper is concerned with two problems connected with approximation by rational functions:

(1) continuity of the Padé approximation operator;

(2) convergence of Chebyshev to Padé approximants as the domain of approximation shrinks to the origin.

The first question has been investigated previously in [4], [8], [14], [15], and the second in [2], [3], [6], [10], [11], [12]. Our purpose is to unify, correct, and extend some of the results of these papers.

Both problems turn upon questions of the convergence of sequences of functions within a fixed space R_{mn}, the set of rational functions having at most m zeros and at most n poles. Such convergence can be defined naturally in many different ways, and it is not obvious a priori which of these is most appropriate. Since each of the papers cited above considers only one or two of these definitions, the scope of the existing results, and the connections between them, have been unclear. We hope to improve this situation.

In particular we will investigate approximations involving a *degenerate* rational function $r\in R_{mn}$—that is, one with $\mu<m$ zeros and $\nu<n$ poles, hence with a *defect* $d=\min\{m-\mu, n-\nu\}$ that is positive. It is in the degenerate situation where the various definitions of convergence become distinct, and also where the answers to (1) and (2) are least obvious. The explanation for this is that in the degenerate case, r can be multiplied by one or more pole-zero pairs $(z-\zeta)/(z-\zeta')$ and the result will still belong to R_{mn}; if ζ and ζ' are nearly equal, the effect of such a perturbation will be large near these points but can be made arbitrarily small elsewhere. It is natural that this possibility should render convergence results somewhat complicated.

If $r\in R_{mn}$ has defect d, and $\tilde{r}\in R_{mn}$ is arbitrary, then $r-\tilde{r}$ belongs to $R_{m+n-d,2n-d}$ and so can have at most $m+n-d$ zeros. As a consequence the degree of agreement of r with a function $\tilde{r}$ is in some sense determined—in the absence of troublesome pole-zero pairs—by how closely they agree at any $m+n+1-d$ points. In both problems (1) and (2) above the origin is a distinguished point, and so we are led to the following notion of "H" convergence: "$r_\varepsilon \to_H r$" denotes convergence as $\varepsilon\to 0$ of the Taylor coefficients

* Received by the editors July 5, 1983.

† Courant Institute of Mathematical Sciences, New York University, New York, New York 10012. The research of this author was supported by a National Science Foundation Mathematical Sciences Postdoctoral Fellowship and by the U.S. Department of Energy under contract DE-AC02-76-ER03077-V.

‡ Seminar für Angewandte Mathematik, ETH-Zurich, Switzerland.

of degrees 0 through $m+n-d$ of $r_\varepsilon \in R_{mn}$ to the corresponding coefficients of r. This is one of a sequence of convergence definitions we consider (precise statements in §1):

cw: coefficientwise,
au: almost uniform away from poles of r,
Δ: uniform on some disk Δ about the origin,
Tay: all Taylor coefficients,
H: Taylor coefficients of degree $\le m+n-d$,
μ: measure.

In addition four other convergence definitions will be mentioned, mainly at the end of §1:

I: uniform on some interval I about the origin,
χ: chordal metric on all of $\mathbb{C}$,
χ_K: chordal metric on compact subsets of $\mathbb{C}$,
cap: capacity.

Definition I is of interest because it has been used in the papers of Werner and Wuytack [14], [15] and Chui et al. [2], [3]. Definition χ is stronger than all of the others, and becomes relevant for rational functions deficient in neither numerator nor denominator degree. Definition χ_K is equivalent to cw, and cap to μ.

Our main results can be abbreviated as follows, where cw is short for $r_\varepsilon \to_{\text{cw}} r$, and so on.

THEOREM 1. (a) *For arbitrary r one has*

$$\text{cw} \Rightarrow \text{au} \Rightarrow \Delta \Rightarrow \text{Tay} \Rightarrow H \Rightarrow \mu.$$

If r is nondegenerate, all six definitions are equivalent. (b) *If r is degenerate, they are all distinct.*

THEOREM 2. (a) *The Padé approximation operator is always H-continuous, regardless of degeneracy, hence also always μ-continuous.* (b) *It is continuous in other senses only when this follows from Theorem* 1, *i.e. only at a function f whose Padé approximant is nondegenerate.*

THEOREM 3. (a) *Chebyshev approximations on a small domain εK containing a neighborhood of the origin always converge in H as $\varepsilon \to 0$ to the Padé approximant r^p, regardless of degeneracy, hence also in μ.* (b) *If r^p is nondegenerate, K can be an arbitrary set with at least $m+n+1$ points* (*e.g.* $[-1, 1]$ *or* $[0, 1]$) *and they will still converge, in all senses.* (c) *If r^p is degenerate they do not in general converge in any sense stronger than H.*

Theorems 2-3 show that the solutions to problems (1) and (2) are closely related: desired properties typically hold in the relatively weak H sense, but hold in stronger senses only when this follows from general considerations involving sequences of rational functions.

In addition we discuss at the end a variant of the Chebyshev vs. Padé question: not whether Chebyshev approximants converge to Padé as the domain shrinks, but whether the magnitude of the error in Chebyshev approximation converges to that for Padé. One sees easily that in general it need not, even when r^p is nondegenerate.

Before beginning, it remains to make some specific remarks on how our results relate to those obtained previously.

(1) *Continuity of the Padé operator*. The basic theorem in this area is due to Werner and Wuytack [15]: in approximation of a real function f, the Padé operator is I-continuous at f if and only if $r^p(f)$ is nondegenerate. (The "if" half of this result was known earlier.) Our Theorem 2 shows that the same statement extends to continuity with respect to cw, au, Δ, and Tay, and that there is no need to restrict attention to real functions.

(2) *Convergence of Chebyshev to Padé.* In 1964 Walsh showed that $r_\varepsilon^* \to_{\text{au}} r^p$ must hold as $\varepsilon \to 0$ for complex approximation on small disks $|z| \le \varepsilon$, if r^p is nondegenerate [11]. Our Theorem 3a is a generalization of this result. In 1974 he extended the convergence statement to real approximation on $[0, \varepsilon]$ [12], but the proof he gave is erroneous. Theorem 3b here gives a correct proof of this theorem, as well as generalizing it with regard to domain and definition of convergence. On the other hand in 1974 Chui, Shisha, and Smith claimed to show $r_\varepsilon^* \to_I r^p$ for real approximation on $[0,\varepsilon]$, regardless of degeneracy [2], [3]. However, our Theorem 3c shows that this conclusion is false. The upshot of our results is that it appears there is very little difference regarding the $r_\varepsilon^* \to r^p$ problem between real and complex approximation, or between approximation on $|z| \le \varepsilon$, $[-\varepsilon, \varepsilon]$, and $[0,\varepsilon]$. In all cases convergence in cw, au, Δ, or Tay is assured only if r^p is nondegenerate.

1. Convergence of sequences of rational functions. Let $\mathbb{C}$ denote the complex plane topologized by the absolute value metric $d(w,z) = |w-z|$. Let S denote the extended plane $\mathbb{C} \cup \{\infty\}$ topologized by the *chordal* or *spherical metric* χ defined by

$$\chi(w,z) = \frac{|w-z|}{\left(1+|w|^2\right)^{1/2}\left(1+|z|^2\right)^{1/2}}$$

for w, $z \in \mathbb{C}$, and by continuity for $w = \infty$ or $z = \infty$ [1], [7]. Under this definition S is a compact 2-manifold, and $\chi(w,z)$ can be interpreted as the Euclidean distance in $\mathbb{R}^3$ between the points w and z on the Riemann sphere of diameter 1. For any two functions f, g: $S \to S$, and any set $K \subseteq S$, define the uniform-norm distance between f and g on K (possibly infinite) by

$$\|f-g\|_K = \sup_{z \in K} |f(z) - g(z)|,$$

and the chordal-metric distance $\chi_K(f,g)$ on K (at most 1) by

$$\chi_K(f,g) = \sup_{z \in K} \chi(f(z), g(z)).$$

Let χ_S be abbreviated by χ.

Let $m, n \ge 0$ be fixed integers, and let R_{mn} be the space of complex rational functions r with at most m zeros in $\mathbb{C}$ (unless $r \equiv 0$) and at most n poles in $\mathbb{C}$, counted with multiplicity, and satisfying the additional condition $r(0) \ne \infty$. A function r: $S \to S$ belongs to R_{mn} if and only if it can be written as a fraction

$$r(z) = \frac{p(z)}{q(z)} = \frac{a_0 + \cdots + a_m z^m}{b_0 + \cdots + b_n z^n}, \qquad b_0 = 1 \tag{1.1}$$

for some coefficients a_k, $b_k \in \mathbb{C}$. We assume that all common factors have been removed from p and q, which makes this representation unique, and we refer to $\{a_k\} = \{a_k(r)\}$ and $\{b_k\} = \{b_k(r)\}$ as "the coefficients of r as a rational function." Let $\mu \le m$ and $\nu \le n$ denote the exact degrees of p and q, so that if $r \not\equiv 0$, then r has exactly μ zeros and ν poles in $\mathbb{C}$. If $r \equiv 0$, then $\mu = -\infty$ and $\nu = 0$.

DEFINITION. The *defect* of r is the nonnegative integer

$$d = d(r) = \min\{m-\mu, n-\nu\}.$$

r is nondegenerate if $\mu = m$ or $\nu = n$ (i.e. $d = 0$); otherwise it is *degenerate* (i.e. $d > 0$).

Any $r\in R_{mn}$ has a Taylor series

$$r(z)=\sum_{k=0}^{\infty} c_k z^k \tag{1.2}$$

converging in a neighborhood of $z=0$. We refer to $\{c_k\}=\{c_k(r)\}$ as "the Taylor coefficients of r", and for convenience we define also $c_k=0$ for $k<0$. The coefficients $\{a_k\}$, $\{b_k\}$ and $\{c_k\}$ are related linearly. To obtain this relation, equate (1.1) and (1.2) and then multiply through by $q(z)$. The result is the following infinite system of equations:

$$\begin{bmatrix} c_{-n} & c_{-n+1} & \cdots & c_{-1} \\ c_{-n+1} & c_{-n+2} & \cdots & c_0 \\ \vdots & \vdots & \cdots & \vdots \\ c_{m-n} & c_{m-n+1} & \cdots & c_{m-1} \\ c_{m-n+1} & c_{m-n+2} & \cdots & c_m \\ c_{m-n+2} & c_{m-n+3} & \cdots & c_{m+1} \\ \vdots & \vdots & & \vdots \end{bmatrix} \begin{bmatrix} b_n \\ b_{n-1} \\ \vdots \\ b_1 \end{bmatrix} = \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_m \\ 0 \\ 0 \\ \vdots \end{bmatrix} - \begin{bmatrix} c_0 \\ c_1 \\ \vdots \\ c_m \\ c_{m+1} \\ c_{m+2} \\ \vdots \end{bmatrix}. \tag{1.3}$$

Of particular interest is the $n\times n$ subsystem

$$\begin{bmatrix} c_{m-n+1} & \cdots & c_m \\ \vdots & & \vdots \\ c_m & \cdots & c_{m+n-1} \end{bmatrix} \begin{bmatrix} b_n \\ \vdots \\ b_1 \end{bmatrix} = - \begin{bmatrix} c_{m+1} \\ \vdots \\ c_{m+n} \end{bmatrix}. \tag{1.4}$$

Let H denote the matrix in (1.4). Since H has the form $h_{ij}=h_{i+j}$, it is a *Hankel matrix*. If H is nonsingular, i.e. $\det H\neq 0$, then the coefficients $\{b_k\}$ are uniquely determined by $(c_{m-n+1},\cdots,c_{m+n})$ as the solution to (1.4). Once these are known, the coefficients $\{a_k\}$ are uniquely determined by $(c_0,\cdots,c_m)$ from the first $m+1$ rows of (1.3). All together, $\{a_k\}$ and $\{b_k\}$ depend upon $c_0,\cdots,c_{m+n}$ in this case, but not on the remaining coefficients c_k.

The following result is well known [1], [5].

PROPOSITION. *r is degenerate if and only if H is singular.*

Proof. The solutions $\{a_k\}$, $\{b_k\}$ to (1.3), (1.4) are unique if and only if H is nonsingular. Such solutions correspond to all possible representations of r as a fraction of the form (1.1), including those that are not in lowest terms. On the other hand a lowest-terms quotient p/q is the unique representation for r if and only if it can be multiplied by no fraction $(z-a)/(z-a)$ and still remain a quotient of type (m,n), which is to say, if and only if r is nondegenerate. $\square$

Thus when H is singular, the defect d is positive. It can be seen that in general $\{a_k\}$ and $\{b_k\}$ are determined by $c_0,\cdots,c_{m+n-d}$ (but not by $c_0,\cdots,c_{m+n-d-1}$).

Suppose $f(z)=\sum_{k=0}^{\infty}c_kz^k$, $c_k\in\mathbb{C}$, is a formal power series. The *Padé approximant* $r^p\in R_{mn}$ to f is defined to be that rational function in R_{mn} whose Taylor series agrees with f to as high an order as possible. It can be shown that if the matrix H formed from the coefficients $\{c_k\}$ is nonsingular, then the coefficients of r^p are the unique solution of (1.3) and (1.4), and $(f-r^p)(z)=O(z^{m+n+1})$. In general H may be singular, but r^p is always uniquely defined and satisfies $(f-r^p)(z)=O(z^{m+n+1-d})$. (Neither of these estimates need be sharp.)

Now let $\{r_\varepsilon\}_{\varepsilon>0}$ be a family of functions in R_{mn}, and let r belong to R_{mn} also. We will make use of the following six definitions of convergence of r_ε to r as $\varepsilon\to 0$.

cw ("coefficientwise") $r_\varepsilon \to_{\mathrm{cw}} r$ if $\lim_{\varepsilon\to 0} a_k(r_\varepsilon)=a_k(r)$ for $0\le k\le m$ and $\lim_{\varepsilon\to 0} b_k(r_\varepsilon)=b_k(r)$ for $1\le k\le n$.

au ("almost uniform") $r_\varepsilon\to_{\mathrm{au}} r$ if $\lim_{\varepsilon\to 0}\|r_\varepsilon - r\|_K=0$ for any compact $K\subseteq\mathbb{C}$ that contains no poles of r.

Δ ("wrt disk Δ") $r_\varepsilon\to_\Delta r$ if $\lim_{\varepsilon\to 0}\|r_\varepsilon - r\|_\Delta=0$ for some disk $\Delta=\{z\in\mathbb{C}:|z|\le\delta\}$, $\delta>0$.

Tay ("Taylor") $r_\varepsilon\to_{\mathrm{Tay}} r$ if $\lim_{\varepsilon\to 0} c_k(r_\varepsilon)=c_k(r)$ for all $k\ge 0$.

H ("Hankel") $r_\varepsilon\to_H r$ if $\lim_{\varepsilon\to 0} c_k(r_\varepsilon)=c_k(r)$ for $0\le k\le m+n-d$.

μ ("measure") $r_\varepsilon\to_\mu r$ if for any $\delta>0$ and any compact $K\subseteq\mathbb{C}$, $\lim_{\varepsilon\to 0}\mu\{z\in K: |r_\varepsilon(z)-r(z)|>\delta\}=0$, where μ is the Lebesgue measure on $\mathbb{C}$.

The following theorem describes the relationships between these definitions of convergence. In the statement "cw" is an abbreviation for $r_\varepsilon\to_{\mathrm{cw}} r$, and so on.

THEOREM 1. (a) *If r is nondegenerate, then*

$$\mathrm{cw}\Leftrightarrow\mathrm{au}\Leftrightarrow\Delta\Leftrightarrow\mathrm{Tay}\Leftrightarrow H\Leftrightarrow\mu. \tag{1.5}$$

(b) *If r is degenerate, then*

$$\mathrm{cw}\;\overset{\Rightarrow}{\not\Leftarrow}\;\mathrm{au}\;\overset{\Rightarrow}{\not\Leftarrow}\;\Delta\;\overset{\Rightarrow}{\not\Leftarrow}\;\mathrm{Tay}\;\overset{\Rightarrow}{\not\Leftarrow}\;H\;\overset{\Rightarrow}{\not\Leftarrow}\;\mu, \tag{1.6}$$

except that au $\Rightarrow$ cw *holds if r has no poles in* $\mathbb{C}$.

Proofs—arbitrary r. First we prove those implications asserted to hold regardless of whether r is nondegenerate, namely the five rightward implications in (1.5)–(1.6).

(a) cw $\Rightarrow$ au. If $r_\varepsilon\to_{\mathrm{cw}} r$, then the denominator polynomials q_ε converge coefficientwise to q, which implies that the zeros of q_ε converge to zeros of q or to ∞. If $K\subseteq\mathbb{C}$ is compact and contains no poles of r, it follows that for all sufficiently small ε, the poles of r_ε are uniformly bounded away from K. Therefore for small enough ε, the values $r_\varepsilon(z)$ $(z\in K)$ depend continuously on the coefficients of r_ε, hence on ε, in a manner uniform in z for $z\in K$. This implies $\lim_{\varepsilon\to 0}\|r_\varepsilon - r\|_K=0$.

(b) au $\Rightarrow\Delta$. Trivial.

(c) $\Delta\Rightarrow$ Tay. If $r_\varepsilon\to_\Delta r$, there is a disk Δ on which r is analytic with $\lim_{\varepsilon\to 0}\|r_\varepsilon - r\|_\Delta =0$. If $\|r_\varepsilon - r\|_\Delta<\infty$, then r_ε is analytic on Δ too. Therefore the Taylor coefficients for both r and r_ε can be computed by Cauchy integrals around $|z|=\delta$, and the uniform convergence on that circle implies that these integrals converge.

(d) Tay $\Rightarrow H$. Trivial.

(e) $H\Rightarrow\mu$. If $r_\varepsilon\to_H r$, then $c_k(\Delta r_\varepsilon)\to 0$ for $0\le k\le m+n-d$, where $\Delta r_\varepsilon=r_\varepsilon - r\in R_{m+n-d,2n-d}$. Setting $M=m+n-d$ and $N=2n-d$, we see that it is enough to show that if $r_\varepsilon\in R_{MN}$ satisfies $c_k(r_\varepsilon)\to 0$ for $0\le k\le M$, then $r_\varepsilon\to_\mu 0$.

For each ε, let $r_\varepsilon(z)$ be written as a quotient $p_\varepsilon(z)/q_\varepsilon(z)$ with the normalization $\|q_\varepsilon\|_\Delta=1$, where Δ is the unit disk. (This is a different normalization from that of (1.1). Further specification regarding common factors and a constant of modulus 1 is unnecessary.) The condition $\|q_\varepsilon\|_\Delta=1$ implies $|b_k|\le 1$ for each coefficient of q_ε, and since $p_\varepsilon(z)=q_\varepsilon(z)\Sigma_{k=0}^\infty c_k z^k$, the conclusion $\lim_{\varepsilon\to 0}\|p_\varepsilon\|_\Delta=0$ then follows by the $c_k\to 0$ hypothesis.

Now let $K\subseteq\mathbb{C}$ be compact, and let $\delta>0$ be arbitrary. Clearly $\|p_\varepsilon\|_K\to 0$ also as $\varepsilon\to 0$. On the other hand we have $\{z\in K:|r_\varepsilon(z)|>\delta\}\subseteq\{z\in K:|q_\varepsilon(z)|<\|p_\varepsilon\|_K/\delta\}$, and it is readily seen that the latter set has measure bounded by $\mathrm{const}(\|p_\varepsilon\|_K/\delta)^{2/N}$. Therefore the measure of this set goes to 0 with ε, which is just what is required to establish $r_\varepsilon\to_\mu 0$ (see [1, vol. 1, §6.6]).

Proofs—degenerate r. Next we prove those relationships asserted to hold if and only if r is degenerate, namely the leftward nonimplications in (1.6). If $r \in R_{mn}$ is degenerate, then $\mu < m$ and $\nu < n$ hold. Therefore for any $a, z_0 \in \mathbb{C}$ with $z_0 \neq 0$, the function

$$r_\varepsilon(z) = r(z)\left(1 + \frac{a}{1 - z/z_0}\right)$$

belongs to R_{mn} too. By choosing a and z_0 judiciously, we construct a sequence of counterexamples that establish the required results. Detailed verifications are left to the reader. In the case $r(z) \equiv 0$, each construction should be modified by setting simply $r_\varepsilon(z) = a/(1 - z/z_0)$.

(f) au $\not\Rightarrow$ cw. Assuming r has a finite pole at z_0, take $r_\varepsilon(z) = r(z)(1 + \varepsilon/(z - z_0))$.

(g) $\Delta \not\Rightarrow$ au. Take $r_\varepsilon(z)$ as in (f), but with z_0 equal to any nonzero complex number that is not a pole of r.

(h) Tay $\not\Rightarrow \Delta$. Take $r_\varepsilon(z) = r(z)(1 + \varepsilon^{-1/\varepsilon}/(1 - z/\varepsilon))$.

(i) $H \not\Rightarrow$ Tay. Take $r_\varepsilon(z) = r(z)(1 + \varepsilon^{m+n+1-d}/(1 - z/\varepsilon))$.

(j) $\mu \not\Rightarrow H$. Take $r_\varepsilon(z) = r(z)(1 + 1/(1 - z/\varepsilon))$.

Proof—nondegenerate r. Finally, assume that $r \in R_{mn}$ is nondegenerate. To complete the proof of Theorem 1, it is enough to show $\mu \Rightarrow$cw:

(k) $\mu \Rightarrow$cw. If r is nondegenerate, assume it has n finite poles $z_1, \cdots, z_n$; the case of m zeros is analogous. It is clear that if $r_\varepsilon \to_\mu r$, then for all sufficiently small ε, r_ε must have n poles $z_k^{(\varepsilon)}$ satisfying $z_k^{(\varepsilon)} \to z_k$ as $\varepsilon \to 0$. This implies $q_\varepsilon \to_{\text{cw}} q$, hence $q_\varepsilon \to_\mu q$. From this and $r_\varepsilon \to_\mu r$, one can conclude $p_\varepsilon \to_\mu p$, hence $p_\varepsilon \to_{\text{cw}} p$, hence $r_\varepsilon \to_{\text{cw}} r$. □

We now make some remarks on the additional notions of convergence mentioned in the Introduction. They are defined as follows:

I ("wrt interval I") $r_\varepsilon \to_I r$ if $\lim_{\varepsilon\to 0} \|r_\varepsilon - r\|_I = 0$ for some interval $I = [-\delta, \delta]$, $\delta > 0$.

χ ("chordal") $r_\varepsilon \to_\chi r$ if $\lim_{\varepsilon \to 0} \chi(r_\varepsilon, r) = 0$.

χ_K ("almost chordal") $r_\varepsilon \to_{\chi_K} r$ if $\lim_{\varepsilon\to 0} \chi_K(r_\varepsilon, r) = 0$ for any compact $K \subseteq \mathbb{C}$.

cap ("capacity") $r_\varepsilon \to_{\text{cap}} r$ if for any $\delta > 0$ and any compact $K \subseteq \mathbb{C}$, $\lim_{\varepsilon \to 0} \text{cap}\{z \in K : |r_\varepsilon(z) - r(z)| > \delta\} = 0$, where cap is the logarithmic capacity [7].

We state without proof some basic facts relating these definitions to the others.

THEOREM 1c.

(i) *If r is nondegenerate, then* $\Delta \Leftrightarrow I$. *Otherwise* $\Delta \Rightarrow I$ *but* $I \not\Rightarrow \Delta$.

(ii) *If both* $\mu = m$ *and* $\nu = n$ *hold, then* $\chi \Leftrightarrow$cw. *Otherwise* $\chi \Rightarrow$cw *but* cw $\not\Rightarrow \chi$.

(iii) $\chi_K \Leftrightarrow$cw.

(iv) cap $\Leftrightarrow \mu$.

Result (iv) is, of course, quite different from the more familiar situation cap $\Rightarrow \mu$ $\not\Rightarrow$ cap that holds for approximation by arbitrary rational functions rather than rational functions of fixed type (m, n) [1].

2. Continuity of the Padé approximation operator. Let $m, n \geq 0$ be fixed and let $f(z) = \Sigma c_k z^k$ be a formal power series. Then f has a unique Padé approximant $r^P \in R_{mn}$, and we let P denote the operator

$$P: f \mapsto r^P.$$

In fact r^P depends only on the coefficients $c_0, \cdots, c_{m+n}$, and if the defect is $d > 0$, it depends only on $c_0, \cdots, c_{m+n-d}$. (To be precise, $\tilde{f} - f = O(z^{m+n+1-d})$ implies $P(\tilde{f}) = P(f)$, but $\tilde{f} - f = O(z^{m+n-d})$ does not.) Therefore the most reasonable way to define convergence of f_ε to f in the Padé approximation context is:

$$f_\varepsilon \to f \text{ if } \lim_{\varepsilon \to 0} c_k(f_\varepsilon) = c_k(f) \quad \text{for } 0 \leq k \leq m+n.$$

Since only finitely many terms of f have any influence, we can be careless as to whether f is a full power series or just a set of numbers $c_0,\cdots,c_{m+n}$. On the other hand for defining convergence of $P(f_\varepsilon)$ to $P(f)$ all of the choices discussed in §1 are reasonable candidates. To each definition of convergence corresponds a different definition of continuity of the Padé approximation operator. We say that P is *H-continuous at f* if $f_\varepsilon \to f$ implies $P(f_\varepsilon)\to_H P(f)$, and so on.

If $r^p=P(f)$ is nondegenerate, then for most senses of continuity it is an easy matter of linear algebra to show directly that P is continuous at f. Essentially the required argument is given in [8, Thm. 3.17] (for rational interpolation), [4, Thm. 8] (for Newton–Padé approximation), in §II of [14] (under the stricter assumption that r^p is normal), and probably elsewhere too. An explicit statement that P is continuous in the nondegenerate case appears perhaps first as [15, Thm. 4.1], where I-continuity is established. The case where r^p is normal was treated earlier in [14]. Our approach here is to show that P is H-continuous regardless of degeneracy, from which continuity in other sense follows as a corollary of Theorem 1a, if r^p is nondegenerate.

THEOREM 2a. *Let f be arbitrary. The Padé approximation operator P is H-continuous at f.*

COROLLARY (*by Theorem* 1). *If r^p is nondegenerate, then P is also* cw-, au-, Δ-, *and* Tay-*continuous at f. Whether or not r^p is nondegenerate, P is μ-continuous at f.*

Werner and Wuytack have established μ-continuity previously in [15, Thm. 6].

Proof. In fact one has local Lipschitz H-continuity with a constant of exactly 1. For we have already mentioned that $r^p-f=O(z^{m+n+1-d})$, and we claim that the analogous identity holds for sufficiently nearby perturbations $\tilde{f}$ of f. To see this, observe that for either $(\mu,\nu)=(m-d,n)$ or $(\mu,\nu)=(m,n-d)$ (with the obvious modification if $r^p\equiv 0$), r^p is also the Padé approximant to f in $R_{\mu\nu}$, and is nondegenerate with respect to that class. By the Proposition of §1 the same nondegeneracy holds for nearby $\tilde{f}$, since small perturbations of a nonsingular matrix H are nonsingular. Therefore for all $\tilde{f}$ sufficiently near to f one has $P_{\mu\nu}(\tilde{f})-f=O(z^{\mu+\nu+1})=O(z^{m+n+1-d})$, hence a fortiori $P_{mn}(\tilde{f})-\tilde{f}=O(z^{m+n+1-d})$, as claimed. □

The main result of [15] is the following converse to Theorem 2a: if r^p is degenerate, then P is I-discontinuous at f. The proof involves multiplications of r^p by cleverly chosen pole-zero pairs. Our proof below generalizes this result to Tay-discontinuity, hence also discontinuity in cw, au, and Δ. Also, in [15] Werner and Wuytack present their argument only for the case in which r^p lies in a 2×2 square block in the Padé table, and they suggest that the proof for the general case will require the introduction of several pole-zero pairs rather than one. However the following proof, which has no block size restriction, shows that one is enough.

THEOREM 2b. *If r^p is degenerate, then P is* Tay-*discontinuous at f.*

COROLLARY (*by Theorem* 1b). *If r^p is degenerate, then P is also* cw-, au-, *and* Δ-*discontinuous at f.*

Proof. Let f,m,n be given, let f have the form $f(z)=c_lz^l+c_{l+1}z^{l+1}+\cdots$ with $c_l\neq 0$, and let $P(f)=r^p\in R_{mn}$ have defect $d>0$. Then

$$f(z)=r^p(z)+az^{m+n+1-d}+O(z^{m+n+2-d}) \tag{2.1}$$

for some $a\in\mathbb{C}$, possibly zero. To begin with, assume $r^p\not\equiv 0$, which implies $l\leq m+n-d$.

If $a\neq 0$, then for each $\varepsilon>0$, define

$$r_\varepsilon(z)=r^p(z)\left(1+\frac{a\varepsilon^{m+n+1-d-l}/c_l}{1-z/\varepsilon}\right). \tag{2.2}$$

Since r^p is degenerate, r_ε belongs to R_{mn} for each ε and has defect at least $d-1$. This implies that for $r_\varepsilon = P(f_\varepsilon)$ to hold for some f_ε, it is enough that f_ε satisfy

$$(2.3)\qquad f_\varepsilon(z) = r_\varepsilon(z) + O(z^{m+n+2-d}).$$

To achieve this, let f_ε be that function which has the Taylor coefficients of f for degree $\geq m+n+2-d$ and those of r_ε for degree $\leq m+n+1-d$. Now from (2.1) and (2.2) it follows that the coefficients of f and r_ε agree up to $O(\varepsilon)$ for degrees $\leq m+n+1-d$. Therefore $f_\varepsilon \to f$ as $\varepsilon \to 0$, while from (2.2), $r_\varepsilon \nrightarrow_{\mathrm{Tay}} r^p$. This establishes discontinuity as claimed.

If $a=0$, replace $a\varepsilon^{m+n+1-d-l}$ by $\varepsilon^{m+n+2-d-l}$ in (2.2).

It remains to treat the case $r^p \equiv 0$, which will occur whenever $f(z) = O(z^{m+1})$. If $m \geq n-1$, we set

$$(2.4)\qquad r_\varepsilon(z) = \frac{a\varepsilon^{m+n+1-d}}{1-z/\varepsilon},$$

or $r_\varepsilon(z) = \varepsilon^{m+n+2-d}/(1-z/\varepsilon)$ if $a=0$, and then the proof is again valid. Therefore assume $m \leq n-2$. In this event r^p has defect $d=n$, while (2.4) has defect $m \leq n-2$, and so that proof breaks down at (2.3). If $f(z) = O(z^{n+1})$, let f_ε have the Taylor coefficients of f for $k \geq n+1$ and those of $\varepsilon^{n+1}/(1-z/\varepsilon)$ for $k \leq n$. Then $r_\varepsilon = P(f_\varepsilon) = \varepsilon^{n+1}/(1-z/\varepsilon)$ $\nrightarrow_{\mathrm{Tay}} 0$ as $\varepsilon \to 0$, but $f_\varepsilon \to f$, and discontinuity is established.

On the other hand if $f(z) = az^K + O(z^{K+1})$ with $a \neq 0$ and $m+1 \leq K \leq n$, set $f_\varepsilon(z) = \varepsilon + f(z)$ and $r_\varepsilon = P(f_\varepsilon)$. Then for any $\varepsilon > 0$, r_ε will have Kth coefficient $a \neq 0$, while r^p has Kth coefficient 0. Thus again one has $r_\varepsilon \nrightarrow_{\mathrm{Tay}} r^p$. □

In summary, the Werner–Wuytack result that P is continuous at f if and only if $r^p(f)$ is nondegenerate holds not only for I-continuity, which seems after all a somewhat unnatural definition of continuity for a problem with no intrinsic restriction to the real axis, but also for continuity with respect to definitions cw, au, Δ, and Tay.

3. Best approximation on small domains. Suppose $K \subseteq \mathbf{C}$ is a compact set, f is a fixed function, and for each $\varepsilon > 0$, $r^*_{\varepsilon K}$ is a best approximation to f in R_{mn} on εK. In 1934 Walsh posed the question [10]: as $\varepsilon \to 0$, must $r^*_{\varepsilon K}$ approach r^p? We are especially interested in three choices of K:

$$\Delta = \{z : |z| \leq 1\}, \quad I = [-1,1], \quad J = [0,1].$$

In his original paper Walsh settled the question in the affirmative for polynomial approximation on these regions, showing [10, pp. 175–176]

$$(3.1)\qquad r^*_{\varepsilon\Delta}, r^*_{\varepsilon I}, r^*_{\varepsilon J} \to_{\mathrm{au}} r^p \quad \text{if } n=0$$

provided f is analytic (case Δ) or sufficiently differentiable (cases I, J) at the origin.

To obtain analogous theorems for rational approximation, it is convenient to make use of the linear system (1.3), and therefore natural to assume that r^p is nondegenerate. In 1964 Walsh extended (3.1) to

$$(3.2)\qquad r^*_{\varepsilon\Delta} \to_{\mathrm{au}} r^p \quad \text{if } r^p \text{ is nondegenerate}$$

for f analytic in a neighborhood of the origin [11]. In Theorem 3a below we generalize this result as follows: if K is any region containing a ball around the origin, then $r^*_{\varepsilon K} \to_H r^p$ as $\varepsilon \to 0$, regardless of degeneracy. Thus H- and μ-convergence always occur, by Theorem 1a, and if r^p is nondegenerate, one also has convergence with respect to cw, au, Δ, and Tay.

A decade later Walsh published the analogous result for the half-interval:

(3.3) $$r_{\varepsilon J}^{*} \to_{\mathrm{au}} r^{p} \quad \text{if } r^{p} \text{ is nondegenerate}$$

for $f \in C^{m+n+1}[0,\delta]$, some $\delta > 0$ [12]. This theorem is correct, but Walsh's proof has an error in it: his equation (13) does not follow from his equation (12), and it appears that no simple modification can get around this problem. In Theorem 3b below we give an alternate proof that avoids this error, and in the process generalize the domain: we show that if K is any bounded set with at least $m+n+1$ points, then $r_{\varepsilon K}^{*} \to_{\mathrm{cw}} r^{p}$ if r^{p} is nondegenerate. In particular K can be disconnected or discrete, and it need not contain the origin.

Walsh did not speculate as to whether the nondegeneracy condition is necessary for (3.2) and (3.3) to hold. This question was taken up by Chui, Shisha, and Smith, also in 1974. For the problem of approximation of a real function $f \in C^{m+n+1}[0,\delta]$ by rational functions with real coefficients, they claimed [2], [3]

(3.4) $$r_{\varepsilon J}^{*} \to_{I} r^{p} \quad \text{regardless of degeneracy.}$$

However, *this assertion is false*. We will demonstrate this in Theorem 3c by exhibiting a counterexample that is a modification of some related examples derived in [6]. The error in the proof of [2] comes in the last sentence of the paper, where the authors appeal to the fact cw$\Rightarrow I$ (in the notation of our Theorem 1), without having imposed the normalization $b_0 = 1$ (eq. (1.1)) in the definition of cw that is needed for this implication to hold.

In general it appears that if r^{p} is degenerate, then nothing can be said about convergence in senses stronger than H, regardless of what domain K is under consideration, and in fact Theorem 3c will give examples with $r_{\varepsilon K}^{*} \not\to_{\mathrm{Tay}} r^{p}$ for $K = \Delta$, I, and J, for both real and complex approximation.

THEOREM 3a. *Let f be analytic in a neighborhood of the origin and have the (m,n) Padé approximant r^{p} with defect d. Let $K \subset \mathbb{C}$ be a bounded set that contains a disk about the origin, and for each ε, let $r_{\varepsilon K}^{*}$ be a best approximation in R_{mn} to f on εK. Then*

(3.5) $$r_{\varepsilon K}^{*} \to_{H} r^{p} \quad \text{as } \varepsilon \to 0.$$

COROLLARY (*by Theorem* 1a). *Under the same hypotheses one has $r_{\varepsilon K}^{*} \to_{\mu} r^{p}$, and if in addition $d = 0$, one has convergence also with respect to* cw, au, Δ, *and* Tay.

Remark. The assumptions that f is analytic and that $r_{\varepsilon K}^{*}$ is the best approximation are unnecessarily strict. All that is needed for the proof is $\|r_{\varepsilon K}^{*} - r^{p}\|_{\varepsilon K} = o(\varepsilon^{m+n-d})$.

Proof. By definition r^{p} is analytic at the origin and its Taylor coefficients agree with those of f through degree $m+n-d$. Since K is bounded, this implies

$$\|f - r^{p}\|_{\varepsilon K} = o(\varepsilon^{m+n-d})$$

and therefore also

$$\|f - r_{\varepsilon K}^{*}\|_{\varepsilon K} = o(\varepsilon^{m+n-d}).$$

Subtracting these estimates yields

(3.6) $$\|r^{p} - r_{\varepsilon K}^{*}\|_{\varepsilon K} = o(\varepsilon^{m+n-d}).$$

Now without loss of generality assume K contains the disk Δ. Then (3.6) will hold in particular on the boundary of $\varepsilon\Delta$, and by a Cauchy integral this implies the estimate

$$c_k(r^{p} - r_{\varepsilon K}^{*}) = o(\varepsilon^{m+n-d-k})$$

for the kth Taylor coefficient of $r^p - r^*_{\varepsilon K}$. For $k \le m+n-d$ one therefore has $c_k(r^*_{\varepsilon K}) \to c_k(r^p)$ as $\varepsilon \to 0$, and this is the definition of $r^*_{\varepsilon K} \to_H r^p$. $\square$

THEOREM 3b. *Let f be analytic in a neighborhood of the origin and have the nondegenerate (m,n) Padé approximant r^p. Let $K \subseteq \mathbb{C}$ be an arbitrary bounded set containing at least $m+n+1$ points, which need not include the origin. Then*

$$r^*_{\varepsilon K} \to_{\mathrm{cw}} r^p \quad \text{as } \varepsilon \to 0. \tag{3.7}$$

COROLLARY (*by Theorem* 1a). *Under the same hypotheses one also has convergence with respect to* au, Δ, Tay, H, *and* μ.

Remark. The remark following Theorem 3a applies again here, and now it is more important. To guarantee $\|r^*_{\varepsilon K} - r^p\| = o(\varepsilon^{m+n})$, it will be enough for f to have $m+n+1$ derivatives at the origin with respect to the set $\cup \varepsilon K$, which may consist of a union of rays through the origin (such as J or I) rather than a complex neighborhood. Also, if $f(\bar{z}) = \overline{f(z)}$, then the conclusion holds for real best approximations $r^*_{\varepsilon K}$ as well as complex ones.

Proof. Let r^p and $r^*_{\varepsilon K}$ be represented as

$$r^p(z) = \frac{P(z)}{Q(z)}, \qquad r^*_{\varepsilon K}(z) = \frac{p_\varepsilon(z)}{q_\varepsilon(z)}$$

normalized by $Q(0) = 1$ and $\|q_\varepsilon\|_{\varepsilon K} = 1$. As in the previous proof, one obtains the estimate (3.6) with $d = 0$,

$$\left\| \frac{p_\varepsilon}{q_\varepsilon} - \frac{P}{Q} \right\|_{\varepsilon K} = o(\varepsilon^{m+n}).$$

By the normalization of Q and q_ε we can multiply through to obtain

$$\|p_\varepsilon Q - P q_\varepsilon\|_{\varepsilon K} = o(\varepsilon^{m+n}).$$

Since the function inside the norm is a polynomial of degree at most $m+n$, and since K contains at least $m+n+1$ points, this estimate can only hold if in fact

$$\|p_\varepsilon Q - P q_\varepsilon\|_{\varepsilon \Delta} = o(\varepsilon^{m+n}).$$

By a Cauchy integral over $|z| = \varepsilon$, this implies that the polynomials $p_\varepsilon Q$ and $P q_\varepsilon$ have approximately equal coefficients,

$$c_k(p_\varepsilon Q) = c_k(P q_\varepsilon) + o(1), \qquad k = 0, \cdots, m+n. \tag{3.8}$$

Now if $n = 0$, then $Q \equiv q_\varepsilon \equiv 1$, and (3.8) is the conclusion (3.7) we are looking for. Therefore assume $n > 0$. In this event the nondegeneracy assumption implies $P \not\equiv 0$, and since P is independent of ε and $\|q_\varepsilon\|_{\varepsilon K} = 1$, it follows that for each sufficiently small ε, $P q_\varepsilon$ has a coefficient bounded below in modulus by a fixed constant. Together with (3.8) this implies that the sets of zeros of $p_\varepsilon Q$ and $P q_\varepsilon$ must converge to each other as $\varepsilon \to 0$ in the following sense: if $z_1, \cdots, z_{m+n}$ and $\zeta_1, \cdots, \zeta_{m+n}$ are the zeros of these polynomials, counted with multiplicity, and padded with numbers $z_k = \infty$ or $\zeta_k = \infty$ when the degree is less than $m+n$, then for some ordering of the subscripts one has $\chi(z_k, \zeta_k) \to 0$ for each k as $\varepsilon \to 0$. Now the zeros of P and Q are independent of ε, and if r^p is nondegenerate, either there are m of the former or there are n of the latter, or both. Suppose P has m zeros; the other case is analogous. Then the convergence of the zero sets of $P q_\varepsilon$ and $p_\varepsilon Q$ implies that for all sufficiently small ε, p_ε has degree exactly m, with its zeros converging as $\varepsilon \to 0$ to those of P. It follows then that the zeros of q_ε also converge to those of Q, with possibly some additional zeros converging to ∞. From here (3.7) is a ready consequence. $\square$

Remark. Comparing Theorem 3b to Theorem 3a, one sees that in permitting a general region K, we have lost the ability to conclude $r^*_{\varepsilon K} \to_H r^P$ in the absence of nondegeneracy. An example shows that this cannot be helped: take $f(x)=x$, $(m,n)=(0,1)$, $K=\{1,-1\}$. For each ε, the best approximation is then

$$r^*_{\varepsilon K}(z)=\frac{\varepsilon^2}{z},$$

which has a pole at the origin, so $r^*_{\varepsilon K} \to_H r^P$ certainly does not hold. On the other hand it is still conceivable that some conditions on f and K weaker than the assumption $\Delta \subseteq K$ would be enough to ensure $r^*_{\varepsilon K} \to_H r^P$.

Having established $r^*_{\varepsilon K} \to r^P$ under appropriate hypotheses, we come now to the task of giving examples to show that if r^P is degenerate, then convergence in stronger senses than H will not in general take place. (Of course, degeneracy will not always cause nonconvergence; for example, the best real approximation in R_{0n} to $f(x)=x$ on εI is 0 for all ε, which converges to the degenerate Padé approximant $r^P \equiv 0$ in every sense.)

THEOREM 3c. *There exist examples of integers m,n and entire functions f with the following properties*:

(i) $r^*_{\varepsilon\Delta} \not\to_{\mathrm{Tay}} r^P$,
(ii) $r^*_{\varepsilon I} \not\to_{\mathrm{Tay}} r^P$,
(iii) $r^*_{\varepsilon I} \not\to_{\mathrm{Tay}} r^P$.

Analogous examples also exist if each problem is restricted to approximation of real functions by real rational functions, in which case one also has $r^*_{\varepsilon K} \not\to_I r^P$. (By a "real" function on Δ, we mean a function f with $f(\bar{z})=\overline{f(z)}$.)

COROLLARY (*by Theorem* 1b). *The same nonconvergence results hold with respect to* cw, au, *and* Δ.

Proof. There are six statements to prove, which we label Δ-$\mathbb{C}$, I-$\mathbb{C}$, J-$\mathbb{C}$, Δ-$\mathbb{R}$, I-$\mathbb{R}$, J-$\mathbb{R}$. Probably examples exist in each category for arbitrary $m \ge 0$, $n \ge 1$, but we will not worry about achieving this generality.

Δ-$\mathbb{C}$. Take $(m,n)=(0,1)$ and $f(z)=z^2-z^5$, hence $r^P \equiv 0$. In the proof of Theorem 4 in [6] it was shown that for all $\varepsilon, r^*_{\varepsilon\Delta}$ has a pole in the region $|z| \le \rho\varepsilon$, for some fixed constant ρ. It follows also from the arguments there that one has $\|f-0\|_{\varepsilon\Delta}-\|f-r^*_{\varepsilon\Delta}\|_{\varepsilon\Delta} \ge \mathrm{const}\,\varepsilon^5$ as $\varepsilon \to 0$. This implies $\|r^*_{\varepsilon\Delta}(\varepsilon\omega)\| \ge \mathrm{const}\,\varepsilon^5$ for each of the three roots of $\omega^3=-1$, and therefore $r^*_{\varepsilon\Delta}$ must have the form

$$r^*_{\varepsilon\Delta}(z)=\frac{a(\varepsilon)\varepsilon^6}{z-\varepsilon b(\varepsilon)} \tag{3.9}$$

with $a(\varepsilon)$ bounded below and $b(\varepsilon)$ bounded both above and below by constants. It follows that the Taylor coefficients $c_k(r^*_{\varepsilon\Delta})$ for $k \ge 6$ diverge to ∞ as $\varepsilon \to 0$, so in particular $r^*_{\varepsilon\Delta} \not\to_{\mathrm{Tay}} r^P$

Δ-$\mathbb{R}$. The example and argument above are suitable here too. Since $b(\varepsilon)$ is real, obviously $r^*_{\varepsilon\Delta} \not\to_I r^P$ also.

I-$\mathbb{C}$. Consider $f(x)=x$, $(m,n)=(0,1)$. It is shown in [6] that $r^*_{\varepsilon I}$ is not a constant, but has a pole somewhere in $\mathbb{C}$. A scale invariance argument shows that $r^*_{\varepsilon I}$ must therefore have the form

$$r_{\varepsilon I}^*(z)=\frac{a\varepsilon^2}{z-\varepsilon b} \tag{3.10}$$

for some constants a and b, independent of ε. Obviously $r_{\varepsilon I}^* \not\to_{\text{Tay}} r^p \equiv 0$.

I-$\mathbb{R}$. Consider $f(x)=x^2$, $(m,n)=(0,2)$. By the equioscillation theorem, $r_{\varepsilon I}^*$ cannot be a constant, for a constant yields at best three equioscillation points. So it must have a pair of real poles, symmetric with respect to the origin, and a scaling argument gives

$$r_{\varepsilon I}^*(x)=\frac{a\varepsilon^4}{x^2-\varepsilon^2 b^2},$$

which again implies $r_{\varepsilon I}^* \not\to_{\text{Tay}} r^p$, and also $r_{\varepsilon I}^* \not\to_I r^p$.

J-$\mathbb{C}$ *and* J-$\mathbb{R}$. For both of these situations the example and argument of case I-$\mathbb{C}$ apply. The equioscillation theorem shows that the best approximation to $f(x)=x$ on εJ is not a constant, either in real or complex approximation; therefore in each case it must have the form (3.10), which implies $r_{\varepsilon J}^* \not\to_{\text{Tay}} r^p$. For real approximation one also has $r_{\varepsilon J}^* \not\to_I r^p$, and in fact in this case the coefficients of the solution have been calculated explicitly by Maehly and Witzgall [8]: they are $a=-\frac{1}{4}$, $b=(1+\sqrt{2})/2$. □

All of the above examples have $r^p\equiv 0$, and it may seem that this might make them exceptional. However, examples with $r^p \not\equiv 0$ can also be invented. For example, consider type (2,1) approximation to $f(z)=z+z^7+z^{15}$ on $\varepsilon\Delta$. Now $r^p(z)=z$, but the arguments of [6] show that $r_{\varepsilon\Delta}^* \not\to_{\text{Tay}} r^p$ holds regardless.

* * *

Throughout this section we have investigated whether $r_{\varepsilon K}^*$ and r^p approach each other as $\varepsilon\to 0$. However, for some purposes it may be more interesting to know whether $\|f-r_{\varepsilon K}^*\|_{\varepsilon K}$ and $\|f-r^p\|_{\varepsilon K}$ approach each other. Let

$$\sigma_{\varepsilon K}=\frac{\|f-r^p\|_{\varepsilon K}}{\|f-r_{\varepsilon K}^*\|_{\varepsilon K}}\geq 1$$

be a measure of the agreement between these two. For Padé approximation to be asymptotically best, one should have $\sigma_{\varepsilon K}\to 1$ as $\varepsilon\to 0$. The following examples reveal that in general this need not occur, even if r^p is nondegenerate, but that on the other hand it may occur even if $r_{\varepsilon K}^* \not\to_{\text{Tay}} r^p$.

First, suppose K is the disk Δ and f is analytic at 0, as in the proofs Δ-$\mathbb{C}$ and Δ-$\mathbb{R}$ above. Then whether or not r^p is degenerate, for small enough ε the function $f-r^p$ maps $|z|=\varepsilon$ onto a curve of winding number at least $m+n+1-d$ whose modulus is constant up to a factor $1+O(\varepsilon)$. By Rouche's theorem one concludes $\sigma_{\varepsilon K}=1+O(\varepsilon)$ [9]. Thus $\sigma_{\varepsilon K}\to 1$ does not imply $r_{\varepsilon K}^* \to_{\text{Tay}} r^p$.

Second, let K be any bounded region that contains a disk about the origin. An extension of the above argument shows $\sigma_{\varepsilon K}\leq$const as $\varepsilon\to 0$, but it is easy to devise situations in which $\sigma_{\varepsilon K}$ is bounded away from 1, even when r^p is nondegenerate. (For example: let K be the eccentric disk $|z-\frac{1}{2}|\leq 1$, and take $f(z)=z$, $(m,n)=(0,0)$. Then $\|f-r^p\|_{\varepsilon K}=3\varepsilon/2$, but $\|f-r_{\varepsilon K}^*\|_{\varepsilon K}=\varepsilon$, so $\sigma_{\varepsilon K}=\frac{3}{2}$ for all ε.) Thus $r_{\varepsilon K}^* \to_{\text{Tay}} r^p$ does not imply $\sigma_{\varepsilon K}\to 1$.

Third, consider any of the examples in the proofs I-$\mathbb{C}$, I-$\mathbb{R}$, J-$\mathbb{C}$, J-$\mathbb{R}$ above. Here one has $\sigma_{\varepsilon K}\equiv$const$>1$ as $\varepsilon\to 0$, independent of ε, and so neither $\sigma_{\varepsilon K}\to 1$ nor $r_{\varepsilon K}^* \to_{\text{Tay}} r^p$ holds.

Finally, take $K=[\frac{1}{2},1]$ and let f be a C^∞ function on $[0,\infty)$ whose Taylor series at the origin is that of some fixed $r_\infty\in R_{mn}$, degenerate or nondegenerate, but which equals some slightly different $r_k\in R_{mn}$ on each interval $4^{-k}K$, $k\geq 0$. Then $\|f-r^*_{\varepsilon K}\|_{\varepsilon K}=0$ but $\|f-r^P\|_{\varepsilon K}\neq 0$ for each $\varepsilon=4^{-k}$, so in this case the ratios $\sigma_{\varepsilon K}$ are not even bounded as $\varepsilon\to 0$.

Acknowledgments. We are grateful to Edward Saff for several suggestions that contributed significantly to the development of this paper, and to Lothar Reichel for calling our attention to the question discussed at the end of §3.

REFERENCES

[1] G. BAKER AND P. GRAVES-MORRIS, *Padé Approximants* (2 vols.), Encyclopedia of Mathematics, vols. 13 and 14, Addison-Wesley, Reading, MA, 1981.

[2] C. K. CHUI, O. SHISHA AND P. W. SMITH, *Padé approximants as limits of best rational approximants*, J. Approx. Theory, 12 (1974), pp. 201–204.

[3] C. K. CHUI, *Recent results on Padé approximants and related problems, in* Approximation Theory III, Lorentz, Chui, and Schumaker, eds., Academic Press, New York, 1976.

[4] M. A. GALLUCCI AND W. B. JONES, *Rational approximations corresponding to Newton series (Newton-Padé approximants)*, J. Approx. Theory, 17 (1976), pp. 366–392.

[5] W. B. GRAGG, *The Padé table and its relation to certain algorithms of numerical analysis*, SIAM Rev., 14 (1972), pp. 1–62.

[6] M. H. GUTKNECHT AND L. N. TREFETHEN, *Nonuniqueness of best rational Chebyshev approximations on the unit disk*, J. Approx. Theory, 39 (1983), pp. 275–288.

[7] E. HILLE, *Analytic Function Theory*, Vol. II, Chelsea, New York, 1973.

[8] C. MAEHLY AND C. WITZGALL, *Tschebyscheff-Approximationen in kleinen Intervallen* II, Numer. Math., 2 (1960), pp. 293–307.

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

[10] J. L. WALSH, *On approximation to an analytic function by rational functions of best approximation*, Math. Z., 38 (1934), pp. 163–176.

[11] ———, *Padé approximants as limits of rational functions of best approximation*, J. Math. Mech., 13 (1964), pp. 305–312.

[12] ———, *Padé approximants as limits of rational functions of best approximation, real domain*, J. Approx. Theory, 11 (1974), pp. 225–230.

[13] H. WERNER, *On the rational Tschebyscheff operator*, Math. Z., 86 (1964), pp. 317–326.

[14] L. WUYTACK, *On the conditioning of the Padé approximation problem, in* Padé Approximation and its Applications, de Bruin and van Rossum, eds., Springer, New York, 1981.

[15] H. WERNER AND L. WUYTACK, *On the continuity of the Padé operator*, SIAM J. Numer. Anal., 20 (1983), pp. 1273–1280.