CHEBYSHEV APPROXIMATION BY POLYNOMIALS IN THE COMPLEX PLANE

A thesis presented

by

Lloyd Nicholas Trefethen

to

The Committee on Applied Mathematics

in partial fulfillment of the honors requirements

for the degree of

Bachelor of Arts

Harvard College
Cambridge, Massachusetts

2 May 1977

Contents

I. INTRODUCTION

Here is the general problem of Chebyshev approximation: suppose that f(x) is a real- or complex-valued function to be approximated, K is a domain, and V is a space of functions in which an approximation to f is to be found. Then what are the functions $g \varepsilon V$ which minimize the quantity

$$\|f-g\| = \sup_{x \varepsilon K} |f(x)-g(x)| \ ?$$

How can one find them? How small can $\|f-g\|$ be made to become?

The word Chebyshev indicates the fact that the norm under consideration is the uniform norm, as above, in honor of the early work of P.L. Chebyshev (1821-1894) on approximation in this norm. The symbol $\| \ \|$ will refer throughout this paper to the uniform norm over the region K under consideration; notations like $\| \ \|_\infty$ or $\| \ \|_K$ which make explicit reference to norm or domain will be introduced only when necessary for clarity. The expression best approximation is equivalent to Chebyshev approximation.

Our concern will be with functions f(z) of one complex variable, which we will usually require to be continuous and often analytic. The approximating functions will be polynomials or rational functions. The notation $p_n(z)$ denotes a polynomial of degree n; we say of degree n but always strictly mean of degree *at most n*. Similarly, $r_{mn}(z)$ denotes a rational function with numerator of degree at most m and denominator of degree at most n; we say that such a function is of type (m,n). In a given problem of approximation by functions $g \varepsilon V$ there may exist a function g* V such that

$$\|f-g^*\| = \inf_{g\varepsilon V} \|f-g\| ;$$

in this case g* is a <u>best approximation function</u> to f over K, and the asterisk denotes this fact.

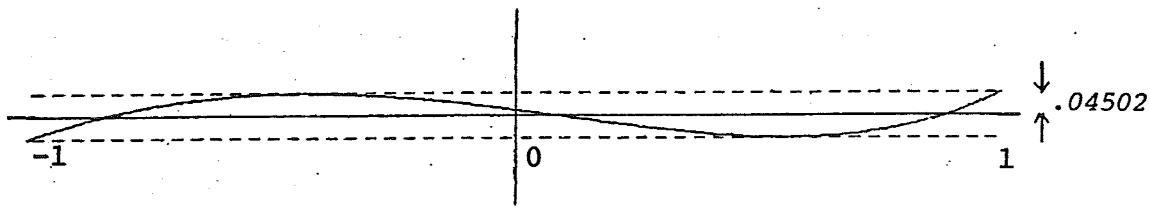
If g(z) is an approximation to f(z), of central importance is the so-called <u>error function</u>:

$$e(z) = f(z)-g(z) .$$

Our goal is always to find approximations g(z) which minimize $\|e(z)\|$ If g*(z) is such a function, we write e*(z) = f(z)-g*(z). The least attainable error $\|e^*\|$ is then denoted by E or E(f), or in particular $E_n(f)$ and $E_{mn}(f)$ for the polynomial and rational cases, respectively.

A final bit of terminology: the homeomorphic image of an interval in the plane is a <u>Jordan arc</u>, that of a circle a <u>Jordan curve</u>. If a Jordan arc or curve can be parametrized in such a way that its real and imaginary components are real-analytic functions of the parameter with non-vanishing derivatives, it is <u>analytic</u>. The closed region bounded by a Jordan curve is a <u>Jordan region</u>. We will make constant use of the fact, which follows from the maximum modulus principle, that best approximation to an analytic function over a Jordan region is equivalent to best approximation over the region's boundary.

The study of Chebyshev approximation is about a century old. Chebyshev himself proved part of the characterization theorem which remains the foundation of Chebyshev approximation in a real variable. This states that if f(x) is continuous on an interval [a,b], then for any n there exists a unique best approximation polynomial p*(z) of degree n, and it is characterized by the property that the corresponding error function $e^*(x) = f(x)-p^*(x)$ attains its maximum value $E = \|e^*\|$ at at least n+2 points $\{x_i\}$ in [a,b], with $e^*(x_{i+1}) = -e(x_i)$ if the x_i are labeled in order. To illustrate, Figure 1 shows a typical oscillating <u>error curve</u> for Chebyshev approximation in a real variable. The function considered is $f(x) = e^x$ on the interval $I = [-1,1]$, for which we have computed the best quadratic polynomial approximation, $p^*(x) \approx .98904 + 1.13018x + .55404x^2$. As predicted by Chebyshev, the error curve $e^*(I) = (f-p^*)(I)$ attains its maximum magnitude $E \approx .04502$ at $2+2 = 4$ distinct points, with alternating signs.



<u>Figure 1</u>. An error curve $e(I) = (f-p^*)(I)$ for real Chebyshev approximation by polynomials on the unit interval $I = [-1,1]$. $f(x) = e^x$, n=2.

Other mathematicians added substantially to Chebyshev's work in the early part of this century. Basic results on existence, uniqueness, possibility of approximation and degree of approximation were proved by Runge, Borel, de la Vallée Poussin, Jackson, Bernstein, and others. Most of the concern was with real approximation at this time, and still is, but progress was made on the complex case too. Beginning in the

1920's, J.L. Walsh at Harvard devoted much of a career to interpolation and approximation in the complex plane, adding much to the theory as well as compiling his encyclopedic work Interpolation and Approximation by Rational Functions in the Complex Domain (1935).

More recently, all branches of approximation theory have flourished in the era of computers, which have provided simultaneously a need for approximations and a means for computing them. Algorithms have been developed for computing best approximations in the real case, in one and in several variables. Complexities and specialties have proliferated.

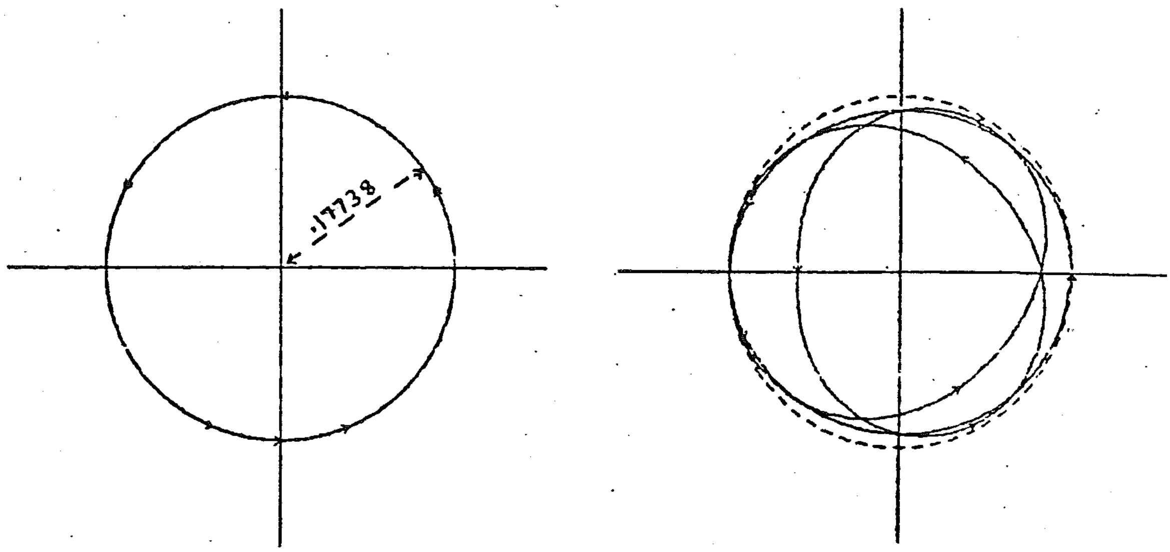
The case of complex Chebyshev approximation has received relatively little attention, partly because it is not a pressing practical problem. (Russian mathematicians have done more in this area than Western ones.) Most of the work on it has been theoretical rather than computational. In fact, I have managed to locate just one pair of algorithms for computing complex polynomial and rational best approximations (see Section III); it seems safe to say that only a handful have been produced. This work has been conducted in the belief that theory and computation can productively be mixed.

There are three main areas of discussion in this thesis. The first is the basic theory of complex Chebyshev approximation by polynomials. Section II outlines this theory in a review of the main theorems: on possibility of approximation, degree of approximation, existence and uniqueness of best approximations, and characterization of best approximations. Proofs of all except the degree of approximation theorems are given in Appendix A. These results are not new, but they are not often found collected together.

The second topic is numerical computation, discussed in Section III. Following the recent work of J. Williams and S. Ellacott in England, I have implemented a computer program for computing best polynomial approximations in the complex plane by a version of Lawson's algorithm adapted from the real case. Section III presents this algorithm and discusses its strengths and a major weakness which does not seem to have been recognized before. A program listing is given in Appendix B. It is shown in Section III that because of an essential difference between real and complex Chebyshev approximation, Lawson's algorithm is bound to converge at an extremely slow rate in many problems in the complex plane.

The essential difference just mentioned between real and complex best approximation is in the geometric behavior of the error curve, and this is the third area of discussion, presented in Section IV. By an error curve in approximation on a Jordan region K in the complex plane we mean the image of the boundary of K; because of the maximum modulus principle, uniform approximation on the boundary of K carries with it approximation in the interior. In the real case, as exemplified by Figure 1, best approximation error curves oscillate through zero between a positive and a negative extreme value. In the complex case they do not have to go through zero in order to get from one extreme value to another, and in general they do not. In fact, computed examples indicate that they tend to do just the opposite. *The error in complex Chebyshev approximation on a Jordan region tends to remain near its maximum magnitude all along the boundary of the region of approximation; geometrically, the error curve tends to approximate a multiply-winding circle about the origin.*

As an example, Figure 2a shows the error curve for degree-2 polynomial approximation to $f(z) = e^z$ again, where now we consider the complex unit disk instead of the real unit interval. (The best approximation polynomial is $p^*(z) \approx .99982 + .99783z + .54326z^2$.) To the eye, the error curve appears to be a perfect circle (winding around the origin three times). In fact it is not quite a perfect circle, but the maximum deviation from a circle is less than one part in twenty-thousand relative to the radius, as the expanded plot in Figure 2b shows. Most examples are not so dramatic, but the phenomenon of nearly-circular error curves occurs to varying degrees generally, and not just on the unit disk. That it occurs appears not to be generally known ([12], [33], [47], [54]).



(a) (b)

Figure 2. (a) An error curve $e(C) = (f-p^*)(C)$ for complex Chebyshev approximation by polynomials on the unit disk bounded by C: $|z|=1$. $f(z) = e^z$, $n=2$. (b) The same, with deviation from a circle exaggerated by a factor of 10,000.

This paper does not succeed in explaining why Chebyshev approximation error curves in the complex plane tend to approximate circles, although some partial results are given in Section IV. As far as I know this is an open problem. Because of the general occurrence of the phenomenon, because of the chance that it may be possible to characterize Chebyshev approximations fully or partially in terms of the shape of their error curves, and because the appearance of near-circles is computationally destructive at least in the implementation of Lawson's algorithm, it is a problem worth pursuing.

II. BASIC THEOREMS OF COMPLEX CHEBYSHEV APPROXIMATION

Suppose a function f defined on a domain K is given, and we are looking for best polynomial approximations to f over K. Here are four basic questions which it is natural to ask:

1) Possibility of approximation. Can f be uniformly approximated as closely as desired by polynomials of sufficiently high degree?

2) Degree of approximation. If it can, how quickly does the minimum error $E_n(f)$ approach 0 as n approaches ∞? How great is $E_n(f)$ for a fixed degree n?

3) Existence and uniqueness of best approximations. For fixed n, does a best degree-n approximation polynomial for f over K exist? If so, is it unique?

4) Characterization of best approximations. How can such a best approximation polynomial be characterized?

These questions are the natural starting point for a study of polynomial approximation over any domain, and much attention has been given to them during the past century. The most heavily studied case has been approximation of a real-valued function over a real interval; for this case all four questions had been well resolved by Weierstrass, Chebyshev, Jackson and others by 1911. For the case of approximation of a complex-valued function over a complex domain the questions have received answers more recently, but the characterizations available for best approximation polynomials are not as powerful as one would hope.

This section is devoted to an exposition of a set of answers to questions (1) - (4) in the case of complex approximation by polynomials. In order to make the section readable, proofs are deferred to Appendix A. Polynomial approximation in a complex domain is simultaneously an extension of the polynomial approximation problem on a real interval and a restriction of the rational approximation problem on a complex domain; parallel methods and results for these two neighboring problems will be pointed out whenever possible.

1. Possibility of approximation

The question of possibility of approximation by polynomials was opened by the landmark theorems of Weierstrass and Runge in 1885, was refined successively over the following sixty years, and finally was completed in a general way by Mergelyan in 1951.

Weierstrass's contribution, the Weierstrass approximation theorem, settled the question for the case of real approximation on a real interval ([52]). *Provided only that* $f(x)$ *is continuous on* $[a,b]$ *,* f *can be approximated arbitrarily closely over* $[a,b]$ *by polynomials* $p_n(x)$ *for sufficiently large degree* n. Runge's theorem in the same year took the major step for the complex case ([38]): *if* K *is a compact set in the plane whose complement is connected, and if* $f(z)$ *is analytic throughout a neighborhood of* K*, then* f *can be approximated arbitrarily closely over* K *by polynomials.* Each of these theorems can be proved with elementary arguments, and in fact the standard proof of Weierstrass's theorem via Bernstein polynomials (cf. [11], pp. 107-111) is both elementary and constructive, although the construction is very inefficient. A simple proof of Runge's theorem is given in Appendix A.

In the years following 1885 refinements in Runge's theorem were made by Walsh (1926), Lavrentiev (1934), and Keldysch (1945). Finally in 1951 S.N. Mergelyan extended Runge's theorem to a very general result, which we state here ([27]):

Theorem 1 (Mergelyan's theorem). Let K be a compact set in the plane whose complement is connected, and let f(z) be continuous on K and analytic in the interior. Then for any $\varepsilon>0$*, there exists a polynomial p(z) such that* $\|f-p\|<\varepsilon$ *over K.*

The only difference in content between Runge's theorem and Mergelyan's theorem is that the latter weakens the requirement of analyticity on a neighborhood of K to just analyticity in the interior. Why is this refinement so important? There are at least two answers to this question. First, the relaxation of the boundary restriction makes the theorem significantly more powerful. In particular, the Weierstrass approximation theorem is a special case of Mergelyan's theorem, but not of Runge's theorem. More generally, from Mergelyan's theorem it follows that if C is any set which has no interior points and does not separate the plane, then a function f(z) continuous on C can be uniformly approximated by polynomials on C; this was first proved by Lavrentiev in 1934. Second, note that to say that f can be approximated arbitrarily closely by polynomials on some set K is the same as to say that f is a uniform limit over K of a sequence of polynomials, and that this implies (since polynomials are analytic functions) that f is analytic in the interior of K. Thus Mergelyan's theorem is as general as possible in a sense in which Runge's theorem is not, for we may state it in the form of an equivalence: *if K is a compact set in the plane whose complement is connected, then f may be approximated arbitrarily closely by polynomials <u>if</u> <u>and</u> <u>only</u> <u>if</u> f is continuous on K and analytic in the interior of K.*

A proof of Mergelyan's theorem is given in Appendix A. This is Mergelyan's original proof, which proceeds from Runge's theorem as a lemma. The theorem can also be proved in a more elegant but more advanced way with the techniques of functional analysis; see [6].

Theorem 1 is the general (and entirely satisfactory) answer to question (1).

2. <u>Degree of approximation</u>

The study of degree of approximation is a large and somewhat confusing area, replete with many theorems establishing various estimates in various contexts. Here we will set forth only the two broadest results on asymptotic degree of approximation in a complex Jordan region: one for the case of f analytic within K and continuous on the boundary, and the other for the case of f analytic throughout a neighborhood of K.

The theory is rooted in the analogous study for the real variable case, which was brought most of the way to its present state in the first decade of this century by Jackson and Bernstein and others. Their work made great use of the theory of trigonometric approximation by sums of the form $\sum_{k=0}^{n}(c_k \cos kx + d_k \sin kx)$, a problem intimately related to our own question of approximation by polynomials. For example, if f(x) is continuous and suitably periodic on $[-\pi,\pi]$ then its truncated Fourier series gives an upper bound for the minimum error attainable in approximation by sums of the above form; with the change of variable g(x) = f(cosx) the same estimate can be carried over to the case of polynomial approximation to a function f(x) continuous on the interval [-1,1]. Using this kind of approach, Jackson proved ([20]) that if f(x) on [a,b] is such that $f^{(k)}(x)$ exists and satisfies a Lipschitz condition of order $\alpha\varepsilon(0,1]$, then $E_n(f) = O(\frac{1}{n^{k+\alpha}})$. The converse of this theorem is also valid (with a modification in the case $\alpha=1$), and was established by S. Bernstein ([3]). For these and other results on degree of approximation in the real variable case, see for example [8], [11], or [26].

The extension of such theorems to the complex case was achieved by J.L. Walsh and his students in the 1920's and 1930's. At this point the problem becomes considerably more complicated. Walsh's strategy was a powerful one, which is covered fully in his book ([50]). Given a Jordan region, he was able to make use of Green's functions with poles at infinity to approximate its boundary curve uniformly by lemniscates, which are the curves in the plane consisting of points which satisfy an equation of the form $|p(z)|=k$ for some polynomial p. Best approximation results for general curves now followed from results which Walsh established on approximation on lemniscates. This technique differs completely from that for the real case, but some of the results it yields are directly parallel. Here is an $O(\frac{1}{n^{k+\alpha}})$ result for the complex case, stated in not quite full generality, as proved by J.H. Curtiss and Walsh and W.E. Sewell; for an exposition see [43]:

Theorem 2. *Let K be a compact set in the plane whose boundary is an analytic Jordan curve C, and let f(z) be continuous throughout K and analytic in the interior of K. Given $\alpha \varepsilon (0,1)$ and $k \geq 0$, there exist polynomials $p_n(z)$ and a constant M such that*

$$\|f-p_n\| < \frac{M}{n^{k+\alpha}}$$

if and only if $f^{(k)}(z)$ exists throughout K and satisfies a Lipschitz condition of order α on C.

Now this result does not assume that f is analytic on the boundary curve C, but of course in many cases of interest f will be analytic there. When this is true f is infinitely differentiable on C, so that Theorem 2 implies that $n^k \|f-p_n^*\| \to 0$ as $n \to \infty$ for arbitrarily large k. In fact, in such a case $\{p_n^*\}$ converges to f geometrically ([50]):

Theorem 3. *Let K be a compact region in the plane whose boundary is an analytic Jordan curve C, and let f(z) be analytic throughout a neighborhood of K. Let the exterior of K be mapped conformally on the exterior of the unit disk in the w-plane in such a way that the points at infinity correspond to each other, and let C_ρ be the (inverse) image in the z-plane under this mapping of the circle $|w| = \rho$, $\rho > 1$. If $f^{(k)}(z)$ exists on the curve C_ρ and satisfies a Lipschitz condition of order $\alpha \varepsilon (0,1]$ there, then there exist polynomials $p_n(z)$ and a constant M such that*

$$\|f-p_n\| < \frac{M}{\rho^n n^{k+\alpha}} .$$

Conversely, if there exist polynomials such that this formula is valid *with* $\alpha\varepsilon(0,1)$, *then* $f^{(k-1)}(z)$ *exists on* C_ρ *and satisfies there a Lipschitz condition of order* α.

In the special case in which K is a circular disk this geometric convergence is easy to see. Suppose that K is a disk with radius R, and suppose f(z) converges not only in K but in a larger concentric disk with radius ρR for $\rho>1$. Then we know that the coefficients in the Taylor expansion for f around the center of K must approach zero essentially as ρ^{-n}, and so the truncated Taylor series for f should provide a series converging to f at the required geometric rate. The argument can easily be made rigorous with the use of Cauchy's inequality. It is good always to bear in mind this special case of approximation over a disk, for it helps make Theorems 2 and 3 plausible. This argument also has the advantage of showing in a loose way that best approximations on a circular region are not much better than truncated Taylor series. If K is not circular, of course, a Taylor series may not exist to be truncated.

3. Existence and uniqueness of best approximations

With these problems we move from questions of asymptotic behavior as $n\to\infty$ to questions dependent on a fixed degree n.

The existence of best polynomial approximations to continuous functions in the real case was first established by Borel in 1905 ([47]), and in the complex case by Leonida Tonelli in 1908 ([46]). This result is straightforward to show by a compactness argument — a proof is given in Appendix A.

Theorem 4 (Existence of best approximations). Let f(z) be continuous on a compact set K in the plane. Then for any $n \geq 0$, there exists a degree-n polynomial $p^*(z)$ of best approximation to f over K; that is, there exists a polynomial $p^*(z)$ of degree n such that $\|f-p^*\| \leq \|f-p\|$ for every degree-n polynomial p(z).

Best rational as well as polynomial approximations over compact subsets of the plane exist, but the proof is more difficult. It was first accomplished by Walsh in 1931 ([48]) using the techniques of normal families developed by P. Montel in the early part of the century.

Having answered the existence question in the affirmative, we proceed to do the same with uniqueness. Uniqueness in the real case was part of Chebyshev's original work of 1859 ([7]); Tonelli extended the result to the complex case in 1908 ([46]). This result is entirely elementary. The details are given in Appendix A, but we outline the argument here because it exemplifies the elementary character of the early work of Chebyshev and Tonelli and others on polynomial approximation: most of the results follow from the definition of a best approximation, the fact that a polynomial of degree n can have no more than n zeros, and the fact that such a polynomial may be interpolated through any n+1 points.

The proof rests on a lemma which will be of interest later on. Suppose that p(z) is a degree-n approximation polynomial to f(z) on some set K, and suppose that the extremal set K_o on which the error function $e(z) = f(z)-p(z)$ attains its maximum magnitude $\|f-p\|$ contains no more than n+1 points of K. Then one may construct an interpolation polynomial q(z) of degree n which agrees with e(z) at all of these points of extreme magnitude, and it can be shown that adding a sufficiently small multiple of q to p produces an approximation $p+\lambda q$ which is closer to f than p. Proceeding in this way we prove:

<u>Lemma</u>. *Let f(z) be continuous on a compact set K in the plane which contains at least n+2 points. Let p*(z) be a best degree-n polynomial approximation to f(z) over K. Then the extremal set K_o on which f(z)-p*(z) attains its maximum magnitude contains at least n+2 points.*

To show uniqueness, we now make use of the fact that a degree-n polynomial can have only n zeros. If p* and q* are both best approximations of degree n, it is easy to see that ½(p*+q*) must be a best approximation too, and from this we can show that p* and q* must attain equal maximum errors at equal points. Since there are at least n+2 such points, p* and q* must be identical.

<u>Theorem 5</u> <u>(Uniqueness of best approximations)</u>. *Let f(z) be continuous on a compact set K in the plane which contains at least n+1 points. Then the best degree-n polynomial approximation to f over K is unique.*

Uniqueness for rational functions of given type does <u>not</u> hold in general. Chebyshev ([7]) showed that it does hold in the real rational case (see [26], p. 161), but Walsh demonstrated by counterexample that it can fail to hold in the complex plane ([50], p. 356). In fact, it can even happen that a complex rational function may approximate a real-valued function on a real interval better than any real rational function of the same type, and consequently that such a best approximation is not unique. For example, the best complex rational approximation to $f(x) = x^2$ on the interval [-1,1] does <u>not</u> have real coefficients ([40]).

For polynomial approximation, however, continuity of f and compactness of K are enough to ensure existence and uniqueness.

4. Characterization of best approximations

To compute or recognize best approximations, it is important to have a more concrete characterization of them available than the definition $\|f-p^*\| \leq \|f-q\|$ $(\forall q)$. This problem is best approached in the way in which uniqueness of best approximations is established, by considering the extremal set. If p is a polynomial of degree n with extremal set K_o, and if we perturb p to consider the approximation polynomial $p+\lambda q$ for some q also of degree n, then for λ sufficiently small only points near K_o will be candidates for extremal points of $f-(p+\lambda q)$. In particular, if the addition of λq perturbs the approximation in such a way as to decrease the error magnitude at all the extremal points z_i in K_o, then by continuity this should be enough to ensure $\|f-(p+\lambda q)\| < \|f-p\|$ for sufficiently small λ. Such an approach is the standard one upon which characterizations of best approximations are based: locally, we need consider only the points of extreme error magnitude.

For real approximation on a real interval, the problem of characterization is now as good as solved. In one direction, we may easily show as in the lemma on p. 16 that $f(x)-p^*(x)$ must attain its maximum magnitude $\|f-p^*\|$ at at least n+2 points. In the other, if $f(x)-p(x)$ attains its maximum magnitude at n+2 (or more) successive points with alternating signs, then no perturbation term $\lambda q(x)$ of degree n can lessen the error at all n+2 points unless $q(x)$ crosses the x-axis n+1 times in [a,b], and hence has n+1 zeros; since this is impossible, p(x) must be the best approximation p*(x). Thus we have a simple characterization: *for real polynomial approximation on an interval [a,b], p(x) is the best degree-n polynomial approximation to a continuous function f(x)* *__if and only if__ f(x)-p(x) achieves its maximum magnitude at at least n+2 successive points in [a,b], with alternating signs of the error function f(x)-p(x).* This characterization is the __Chebyshev equioscillation theorem__. It can be extended to apply to rational approximations ([7]); in the rational case one does not know a priori how many extremal points are needed to determine a best approximation.

For complex approximation the situation is more difficult. What breaks down is the notion of "alternating signs" of the error function. Now the sign of an error is a __complex sign__ of the form $e^{i\theta}$, and since the complex sign of f(z)-p(z) may easily change values without passing through 0 as z moves from one extremal point to another, one cannot argue that a perturbation term λq(z) which reduces the error at n+2 points must have n+1 zeros. On the contrary, it is possible for an error function f(z)-p(z) to attain its maximum magnitude at n+2 points without having p(z) be the polynomial of best approximation. For example, z^4-z attains a maximum magnitude of 2 at three points on the unit disk, but the best degree-1 approximation to z^4 over the unit disk is 0, not z.

We will now give two characterizations of best approximations in the complex plane, but neither one is as powerful as Chebyshev's equioscillation theorem for the real case. For the first one, consider that we might state the argument leading to the Chebyshev result like this: p(x) is the best degree-n polynomial approximation to f(x) on

$I=[a,b]$ if and only if, for any degree-n polynomial $q(x)$,

$$\max_{x\in I_o} \{[f(x)-p(x)]q(x)\} \geq 0 ,$$

where I_o is the extremal set on which $|f(x)-p(x)| = \|f-p\|$. Kolmogorov's characterization ([22]) is an extension of this statement:

Theorem 6 (Kolmogorov's characterization). Let $f(z)$ be continuous on a compact set K in the plane. Let $p(z)$ be a polynomial of degree n, and let K_o be the extremal set of points z at which $|f(z)-p(z)| = \|f-p\|$. Then p is a best degree-n approximation to f over K if and only if, for any degree-n polynomial $q(z)$,

$$\max_{z\in K_o} \mathrm{Re}\{[f(z)-p(z)]\overline{q(z)}\} \geq 0 .$$

A proof of this theorem is given in Appendix A. It is a small modification of the earlier proof of uniqueness.

The second, more substantive characterization is due to Remes ([30]), who tightened Kolmogorov's characterization by making use of arguments of convexity:

Theorem 7 (Remes's characterization). Let $f(z)$ be continuous on a compact set K in the plane. Then $p(z)$ is a best degree-n polynomial approximation to f if and only if for some $r \leq 2n+3$ there exist r points $z_1, \ldots, z_r$ in the extremal set K_o for p, and r numbers $w_1, \ldots, w_r$ with $w_k>0$ and $\Sigma w_k = 1$, such that

$$\sum_{k=1}^{r} w_k\{f(z_k)-p(z_k)\}\overline{q(z_k)} = 0$$

for every degree-n polynomial q.

This theorem is proved in Appendix A also.

Neither of these characterizations has the simplicity or the power of Chebyshev's equioscillation theorem for the real case. This fact will be expanded upon in the next section, on computation of best approximations. However, the Remes characterization does succeed in eliminating all but extremal points $z_k \varepsilon K_o$ from consideration. As a consequence we have an immediate corollary of Theorem 7 whose importance is both theoretical and computational:

<u>Corollary</u>. *Let f(z) be continuous on a compact set K in the plane. If p*(z) is a best degree-n polynomial approximation to f over K, then p* is also a best degree-n approximation to f over some finite subset of K which consists of at most 2n+3 points.*

Unfortunately, in practice it is no easy matter to find such a set of points, or even to know how many points it must contain.

III. COMPUTATION OF COMPLEX CHEBYSHEV APPROXIMATIONS

1. Lawson's algorithm

The Chebyshev equioscillation theorem (p. 18) is a powerful tool for computing best approximations numerically in the real case. Many algorithms for the real case exist (see [9], [10], [24], [31] for surveys), and most start from this theorem.

The best known of the numerical methods is the second algorithm of Remes ([29]). The idea behind this algorithm in the polynomial case is that if any set $\{x_k\}$ of n+2 points of alternating extreme error were known, then the best approximation p* could be computed by solving the linear system of n+2 equations

$$f(x_k)-p^*(x_k) = (-1)^k \delta , \qquad (1)$$

where the unknowns are the n+1 coefficients of p* and the value of δ. The algorithm attempts to find such extremal points iteratively, as follows. First, an initial guess $\{x_k^1\}$ at the extremal points is chosen and a best approximation $p^1(x)$ over these points is computed from (1). Next a new set $\{x_k^2\}$ is constructed by choosing n+2 extremal points of $f-p^1$ at which this difference alternates in sign. Now a second approximation $p^2(x)$ is computed, and so on.

The Remes algorithm and related algorithms can be extended to rational approximation on a real interval, although the situation grows more complicated for two reasons. First, the equations (1) become nonlinear. Second, one no longer knows in advance how many extremal points $\{x_k\}$ it takes to determine a best approximation (see p. 18).

But the same approach cannot be extended directly to complex approximation. For one thing, again one does not know in advance how many extremal points $\{z_k\}$ to look for, even in the case of approximation by polynomials (see p. 20). But more fundamentally, "alternation" breaks down. The terms $(-1)^k$ in (1) now become unknown arguments $e^{i\theta}$, and so even if the points $\{x_k\}$ are given the system has more unknowns than equations. It was pointed out in Section II that no characterization of complex best approximations exists which is equal in power to the Chebyshev equioscillation theorem for the real case, and this fact shows up in the difficulty of computing Chebyshev approximations in the complex plane.

A different approach, Lawson's algorithm, has been implemented here. This algorithm was first studied for the real case by C.L. Lawson in 1961 ([23]), and an extension to complex approximation was achieved by S. Ellacott and J. Williams in 1976 ([14]). Additional discussion of it may be found in [32] and [34].

Given a Jordan region K bounded by a Jordan curve C, if f is analytic inside K it is sufficient to consider approximations over C, because of the maximum modulus principle. Going further, we shall discretize the problem and consider only a finite subset $\tilde{C}$ of C — say, 100 points evenly spaced. Such a restriction to a discrete set can be justified theoretically. If we define the density $|\tilde{C}|$ of $\tilde{C}$ in C by

$$|\tilde{C}| = \sup_{\tilde{z}\in\tilde{C}} \inf_{z\in C} |z-\tilde{z}| ,$$

then it is true that as $|\tilde{C}|\to 0$, which corresponds (confusingly) to $\tilde{C}$ becoming more dense in C, the best approximation to f over $\tilde{C}$ approaches that over C:

<u>Theorem 8</u> *(Cheney [8], p. 87). Let $f(z)$ be continuous on a compact set K in the plane which contains at least n+1 points. Let $p^*(z)$ and $\tilde{p}^*(z)$ be the best degree-n polynomial approximations to f over K and $\tilde{K}$, respectively, where $\tilde{K}$ is any subset of K. Then as the density of $\tilde{K}$ in K approaches zero, $\tilde{p}^*$ approaches p^*.*

Ellacott and Williams ([14]) discuss the characterization process in more detail. They show in particular that if C is a piecewise twice differentiable Jordan curve (say, the unit square), and if one includes all the "corner" points in each discrete subset $\tilde{C}$ used for computation, then $\|f-\tilde{p}^*\|$ converges to $\|f-p^*\|$ quadratically as $|C|\to 0$.

So let us assume that a discrete domain K has been chosen, and drop the tilde symbol. The procedure used in Lawson's algorithm is to approach p^* iteratively as a suitable limit of weighted least-squares approximations, which are relatively easy to compute. The foundation of this method is the Remes characterization theorem (Theorem 7), which was stated in Section II:

<u>Theorem</u>. *Let $f(z)$ be continuous on a compact set K in the plane. Then $p(z)$ is a best degree-n polynomial approximation to f <u>if and only if</u> for some $r \le 2n+3$ there exist r points $z_1, \ldots, z_r$ in the extremal set K_o for p, and r numbers $w_1, \ldots, w_r$ with $w_k > 0$ and $\Sigma w_k = 1$, such that*

$$\sum_{k=1}^{r} w_k \{f(z_k)-p(z_k)\} \overline{q(z_k)} = 0 \qquad (2)$$

for every degree-n polynomial q.

The linear system of equations (2) has a familiar form: taking $q(z) = 1, z, \ldots, z^n$, it becomes the system of normal equations for approximation of f by p in the weighted least-squares norm with weights $\{w_k\}$. Like the system (1), it includes more unknowns than equations, for we do not know at the outset the polynomial p, the weights $\{w_k\}$, or the points $\{z_k\}$. Lawson's algorithm is a method of finding a set $\{z_k\}$ and corresponding weights $\{w_k\}$ iteratively. First, choose an initial set $\{w_k^1\}$ of positive weights arbitrarily, where now z_k ranges over the discrete set $\tilde{C}$. Using these, solve (2) to find a weighted least-squares approximation over $\tilde{C}$. That is, find p^1 so as to minimize the quantity σ^1 defined by

$$\sigma^j = \left[\sum_{k \in \tilde{C}} w_k \{f(z_k) - p^j(z_k)\}^2 \right]^{\frac{1}{2}} \tag{3}$$

Next, adjust the weights according to this formula:

$$w_k^{j+1} = \frac{w_k{}^j |e^j(z_k)|}{\sum_i w_i{}^j |e^j(z_i)|} \tag{4}$$

where $e^j = f - p^j$, as usual. With the new weights return to compute a second weighted least-squares approximation $p^2(z)$, and so on.

It can be proven that the polynomials p^j computed in this fashion converge to the best approximation polynomial p*, with one proviso. This is that if it should happen during the iteration that w_k is set to zero by (4) for some k, the algorithm must be restrated with new nonzero weights set in a certain way specified in [14], p. 39. After a finite number of such restarts, the algorithm is guaranteed to converge. In practice, in my own experience and that of Ellacott and Williams ([14],

p. 41), the restart procedure is rarely required.

It can also be shown that as the computation proceeds the weighted least-squares error σ^j must increase monotonically to the limit $\|e^*\|$. This fact enables one to judge at each step how close the current p^j is to the best approximation p^*, by examining the bound

$$\sigma^j \le \|e^*\| \le \|e^j\| . \qquad (5)$$

2. <u>Description of program</u>

I have implemented Lawson's algorithm in FORTRAN on a PDP-10 (Aiken Computation Lab, Harvard). The algorithm converges reliably towards best approximations, and I have successfully used it to compute about a hundred best approximation polynomials and plot the corresponding error curves.

For convenience, single precision complex arithmetic has been used throughout. This means that as it stands the program is not suitable for generating approximations with the accuracy required to serve as function definitions within a computer. However, there is no reason to anticipate that any difficulty would be encountered in converting the algorithm to double or higher precision.

The linear system (2) is solved by Gaussian elimination with partial pivoting and iterative improvement, using an adaptation for complex arithmetic of the code of Forsythe and Moler ([16]). Matrices arising in least-squares approximation tend to be ill-conditioned, and specialized methods exist for coping with them; Ellacott and Williams used Golub's algorithm involving Householder transformations ([53]). This implementation could probably be made more efficient and reliable by taking advantage of these methods, but I have not experienced numerical problems with the current program.

The FORTRAN code is reproduced in Appendix B. FUNCT(Z) is the function to be approximated and CP(Z) is the approximation polynomial, whose coefficients are adjusted at each iteration. The main program CHEBY controls the sequence of operations, calling the following subroutines in order:

LAWSON. Performs the Lawson's algorithm iteration. The Forsythe and Moler subroutines DECOMP, SOLVE, IMPRUV, and SING are called to solve the linear system (2). Coefficients for $p^j(z)$ are printed at each iteration, along with the minimum error $E^j_{min} = \min_K |e^j(z)|$, the weighted least-squares error σ^j, and the maximum error $E^j = \|e^j\|$. Convergence is judged manually by the user on the basis of these data and (5).

ERRCRV. Plots the error curve e*(C) in the complex plane. Arrows showing the direction of the curve are marked at the image points of eight points evenly spaced along the domain boundary (for the unit circle, the eight roots of unity).

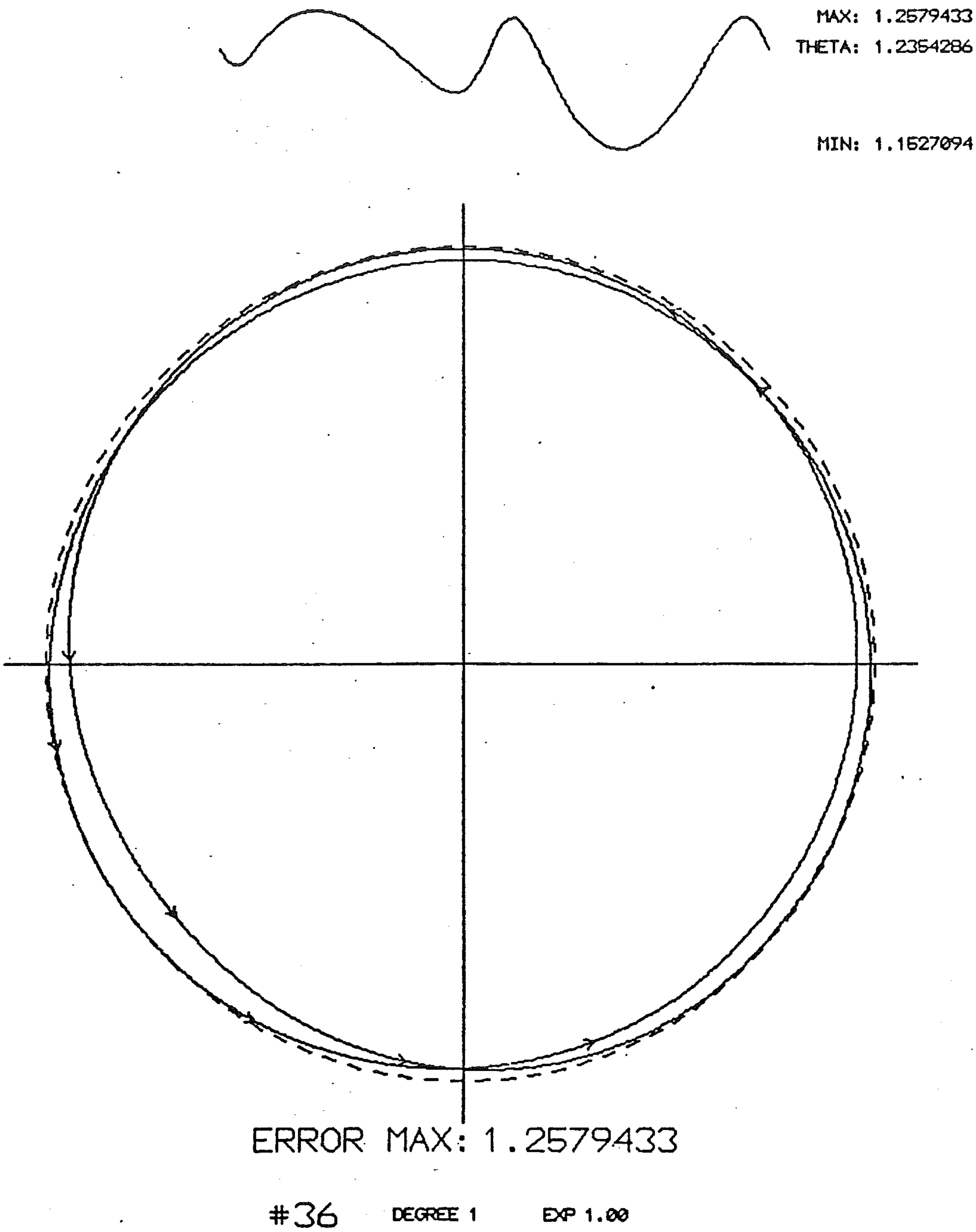
ERRMAG. Plots the error magnitude $|e^*(z)|$ on a scale from E_{min} to E at the top of the plotter page. In a case like that of Figure 2a it is only this plot which shows the true shape of the error magnitude and reveals extremal points, since the error curve itself is so nearly circular.

Figure 3 shows a typical page of output from the plotter. The example shown is the best linear approximation to the function $\frac{\ln(z+3i)}{\cos(z+.1)}$ over the unit disk. Note the characteristic presence of (at least) n+2 points of extreme error in the ERRMAG plot, and the roughly circular shape of the error curve below.

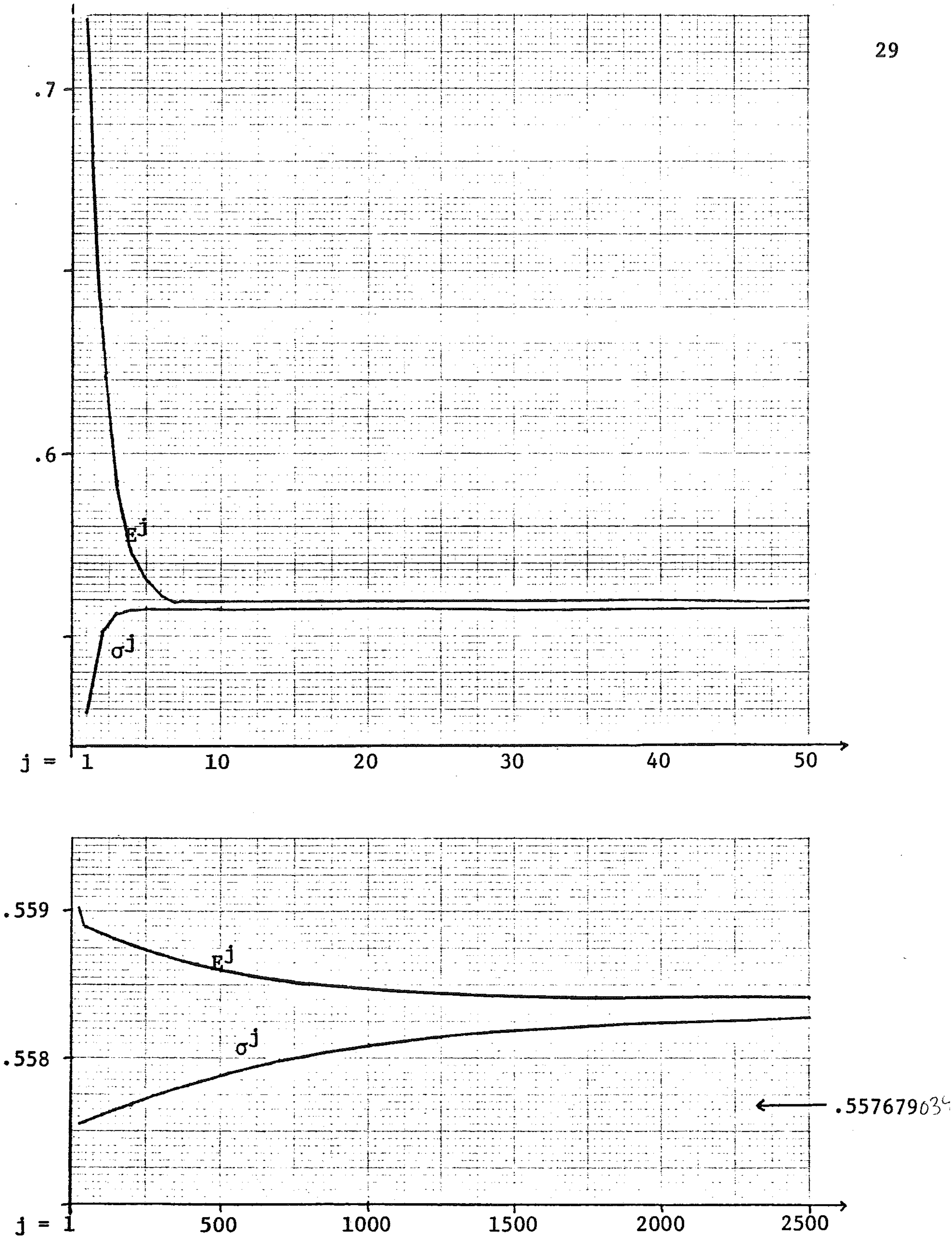
3. <u>Rate of convergence</u>

The algorithm converges towards best approximations, but it fails to do so at a satisfactory rate. In fact, as the best approximation is approached the rate of convergence often drops sharply to near zero. Figure 4 shows plots of E^j and σ^j as a function of iteration number j for the simple case of linear approximation to e^z on the set consisting of 64 evenly spaced points along the unit circle. E^j and σ^j should converge from above and below to the minimax error E, but it is clear that after an initial period of rapid convergence this process begins to take very, very long. Figure 4b extends the curves out to $j = 2500$, a computation which took half an hour of CPU time, with little improvement. Convergence has been virtually halted.

The value of .557679 marked on the plot is a conjectured limiting value for linear best approximation of e^z over the unit disk. It was computed by assuming that the minimax error E(r) in linear approximation of e^z on the disk of radius r is a real-analytic function of r, and investigating its Taylor series coefficients empirically. This approach leads eventually to an apparently clear answer of $E(r) = (2+\frac{4}{r^2})\cosh r$



<u>Figure 3</u>. Typical plotter output shows error curve $e(C) = (f-p^*)(C)$ at center of page with error magnitude plot above. The curves shown are for degree-1 approximation over the unit disk to $f(z) = \frac{\ln(z+3i)}{\cos(z+.1)}$.



<u>Figure 4</u>. E^j and σ^j as functions of iteration number j in linear approximation to e^z over the unit disk by Lawson's algorithm. The domain of approximation is 64 points evenly spaced along the boundary.

$-\frac{4\sinh r}{r} - \frac{4}{r^2}$, which agrees convincingly with computed estimates of E(r) for various r and is asymptotically correct as $r\to\infty$. It is to be expected that .557679 should lie slightly higher than the apparent limiting values in Figure 4b, since the algorithm has considered a discrete set of 64 points, not the continuous unit circle. The fact that it lies slightly lower instead is a bad mark for the conjectured formula.

The slow convergence in this example can be seen in more detail in Figure 5, which shows the distribution of weights and the error magnitude as a function of angle θ around the unit circle, for a sequence of iteration numbers. The error magnitude is plotted on a normalized scale from E_{min} to E as usual, because its absolute magnitude remains very close to .558 all around the unit circle. One sees immediately that although the curve quickly achieves an oscillatory behavior that is close to correct, it fails to converge at a reasonable rate to true solution, which would have three points of precisely equal maximum error magnitude. The same lack of convergence is evident in the weight distribution plots, which are normalized from 0 to w_{max}. Lawson's algorithm is based on the idea that the weights w_k can be made to converge to zero at all but a true set of extremal points, until one is left with a weight distribution that is zero except at a few spikes. Not only is the weight distribution in this example failing to converge at a reasonable rate to the three spikes it must attain in the limit, but for a long time the largest weights appear at points which are not ultimately in the extremal set.

This convergence problem is a result of the near-circularity of the error curve. At each step the weights are adjusted according to equation (4):

$$w_k^{j+1} = \frac{w_k^j |e^j(z_k)|}{\sum w_i^j |e^j(z_i)|} .$$

But to the extent that $|e^j(z)|$ stays nearly constant along the error curve, it is obvious that (4) will adjust the weights slowly. This problem is visible in Figure 5. An early non-optimum weight distribution is established in the first few iterations, when $|e(z)|$ varies significantly over the unit circle; it soon becomes clear that a different weight distribution is needed, but by this time $|e(z)|$ is so nearly constant around the circle that it takes hundreds of iterations for (4) to reshape the distribution. Thus in any problem where the best approximation error curve is nearly circular, which is a common occurrence in the complex case, we may expect the convergence of Lawson's algorithm ultimately to be linear with an asymptotic error constant that is only negligibly better than 1.

For example, one of the examples Ellacott and Williams mention is approximation of e^z on the unit disk by a degree-5 polynomial. Having computed a best approximation, they observe that in this instance "all the points are extremal points" along the unit circle ([14], p. 42). Mathematically this is not true (this will be proved in Theorem 12), but numerically it might as well be, for I estimate on the basis of computed results that the deviation of the true error curve from a circle in this example is less than one part in 10^{10} (see Figure 8). Even in infinite-precision arithmetic, therefore, Lawson's algorithm would require on the order of 10^{10} iterations to converge.

For the practical approximation of functions, the ultimate rate

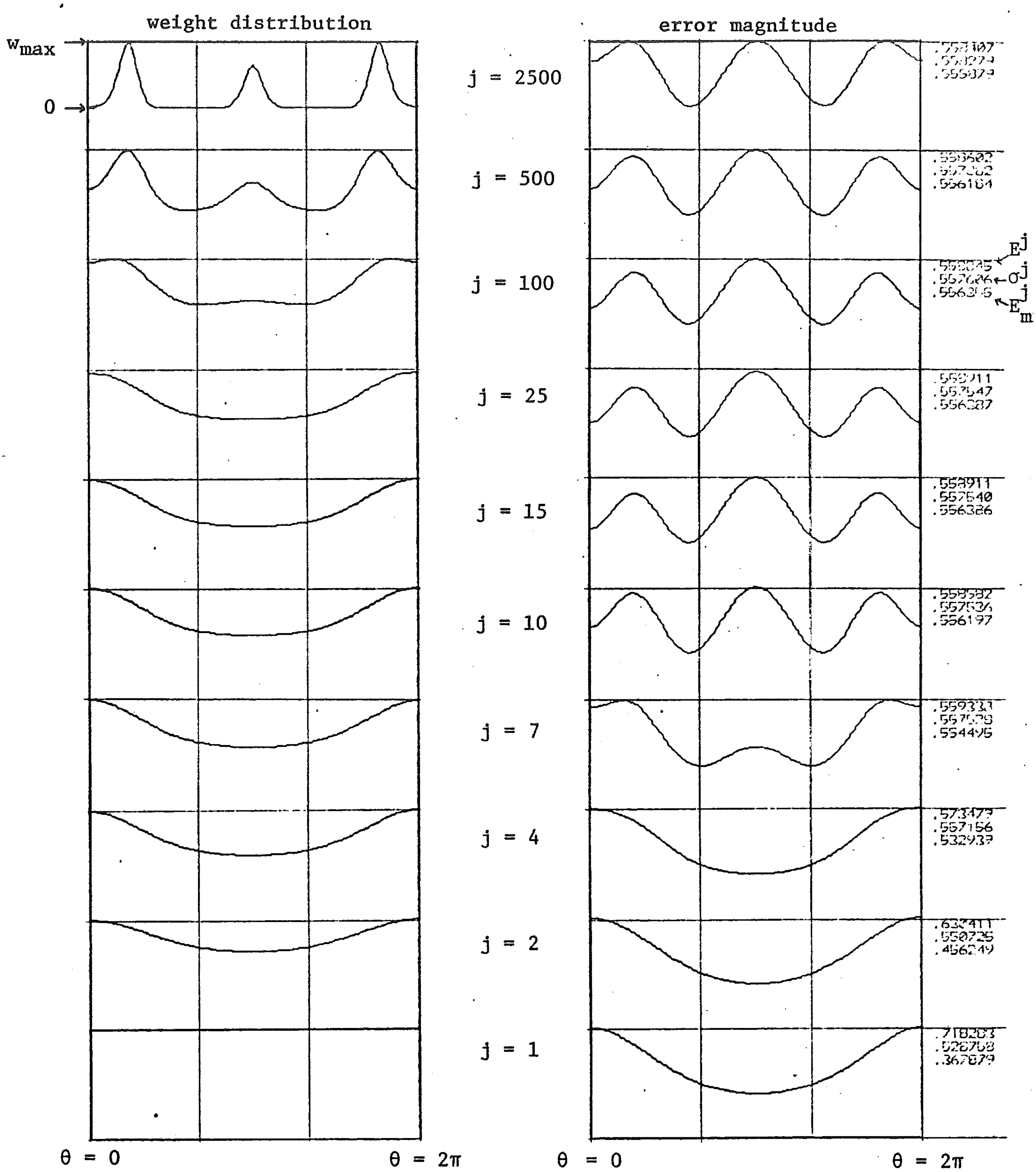


Figure 5. Weight and error magnitude distributions around the unit circle $e^{i\theta}$ for a sequence of iteration numbers j in linear approximation to e^z over the unit disk by Lawson's algorithm. Convergence towards the best approximation is very slow. The domain of approximation is 64 points evenly spaced along the boundary.

of convergence may not be critical provided that a nearly-best approximation can be computed in reasonable time. In Lawson's algorithm only when the best approximation is nearly achieved, at which point the error curve may be nearly circular, does the rate of convergence fall. As a consequence the minimax errors computed and reported here are generally accurate to at least three decimal places. Coefficients for best approximation functions such as those reported in Appendix C, on the other hand, are often an order of magnitude less accurate. Computing them to four places would usually have required the kind of excessive use of computer time required for the computation of Figures 4 and 5.

Those who have worked on Lawson's algorithm in the real case have suggested possible ways to make it converge faster (see [32]). One of these was adopted by Ellacott and Williams in their algorithm: instead of using (4) at every step, at alternate iterations they update the weights instead according to the formula

$$w_k^{j+1} = \frac{w_k^j |e^j(z_k)|^2}{\sum w_i^j |e^j(z_i)|^2} \tag{6}$$

and they show that this alternating procedure is still guaranteed to converge ([14]). Indeed, in practice the resulting convergence rate is generally higher than in the scheme using (4) exclusively. Most of the results reported here have been computed with the alternating method. But asymptotically each iteration using (6) is no better than two iterations using (4), so the modification is not enough to get around the general convergence problem caused by nearly-circular error functions.

I have experimented with various schemes to pick out the extremal points more quickly than with (4) or (6), but have not yet found one which works consistently.

IV. GEOMETRY OF THE ERROR CURVE

1. Geometric interpretation of the problem

We shall now turn to questions of the geometry of the error curve $e^*(C) = (f-g^*)(C)$, and in particular ask why it is observed to tend to approximate a circle. Unfortunately, only a partial answer to this question can be given.

What motivates our interest in the behavior of the error curve is the fact that for analytic functions, approximation over a Jordan region is equivalent to approximation over its boundary. This is an immediate consequence of the maximum modulus principle:

Theorem 9. *Let K be a Jordan region in the plane whose boundary is C, and let $f(z)$ be continuous throughout K and analytic in the interior. Then $r(z)$ is a best rational approximation to f over K of type (m,n) if and only if it is a best approximation to f over C among the rational functions of type (m,n) which have no poles in K.*

Because of this fact, the problem of approximating an analytic function f on a Jordan region by rational functions or polynomials may be interpreted geometrically as follows: what is a function g^* such that $f-g^*$ maps the boundary C into a disk of minimum radius about the origin? This interpretation is surely not original, but most people do not approach the Chebyshev approximation problem from this point of view.

To begin with, let us state a collection of simple geometric facts.

Theorem 10. *Let K be a Jordan region in the plane whose boundary is C, and let f(z) be continuous throughout K and analytic in the interior. Then we have:*

a) If g(z) is any approximation to f which is analytic within K, then the error curve e(C) = (f−g)(C) winds around the origin in the positive sense as many times as there are points of interpolation z_i *at which* $f(z_i) = g(z_i)$ *interior to K.* (Argument principle)

b) Suppose f and g are in fact analytic on a neighborhood of K. If a point z_c *on C is the vertex between two analytic Jordan arcs of C meeting at a definite angle α, then the error curve has a bend of the same angle α at the point* $(f-g)(z_c)$, *provided that* $(f'-g')(z_c) \neq 0$. (This is the interpretation of (f−g)(z) as a conformal mapping.)

c) Suppose p(z) is the best degree-n polynomial approximation to f over C, hence over K. Then the error curve e*(C) = (f−p*)(C) is contained within a circle of radius* $\|e^*\|$ *about the origin. It touches this circle at at least n+2 points of C, and attains locally minimum magnitudes at at least n+2 points interlaced with these.* (This is a restatement of the lemma preceding Theorem 5 on p. 16.)

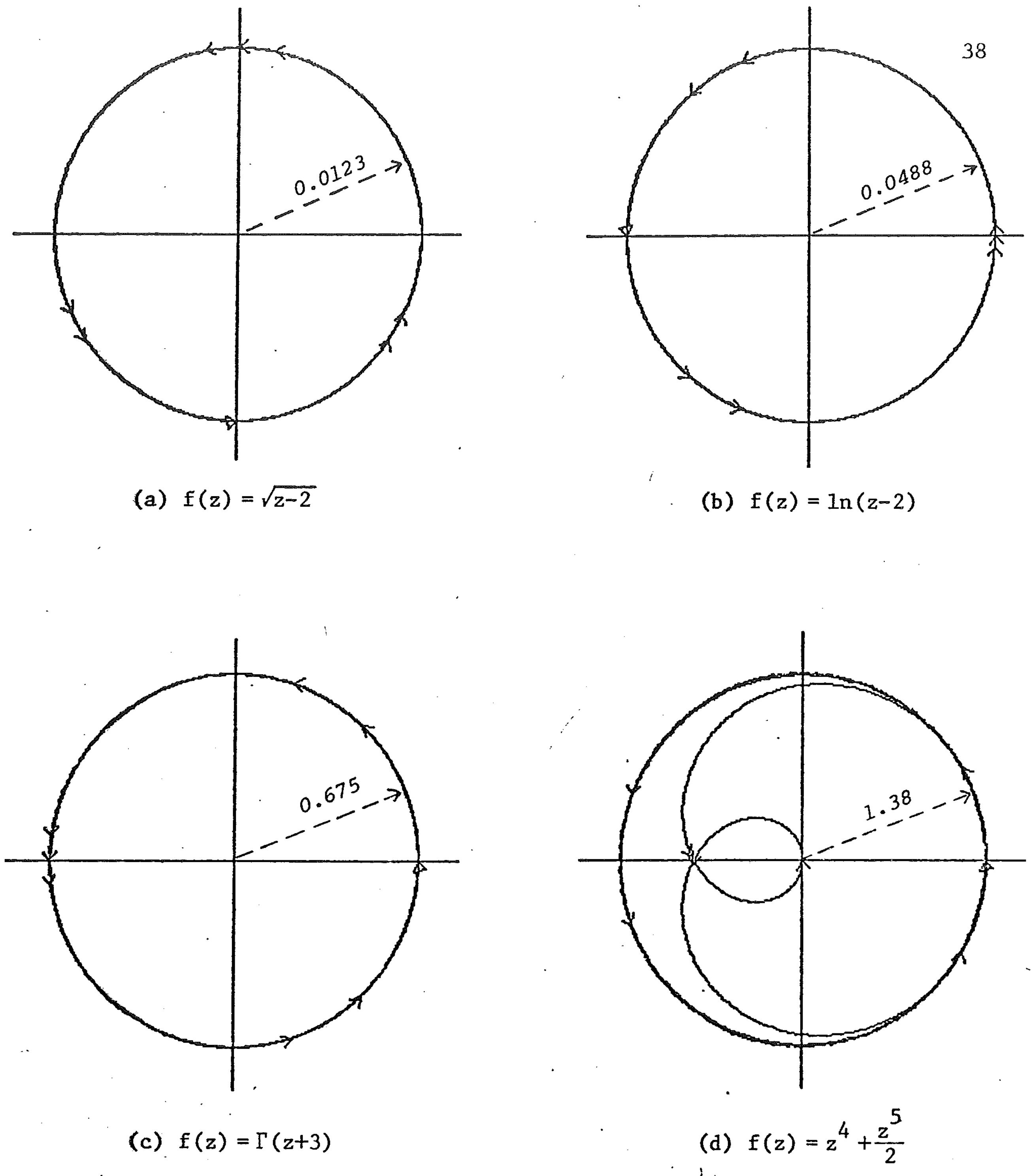
These geometric features will be illustrated in the figures which follow.

2. Computational evidence

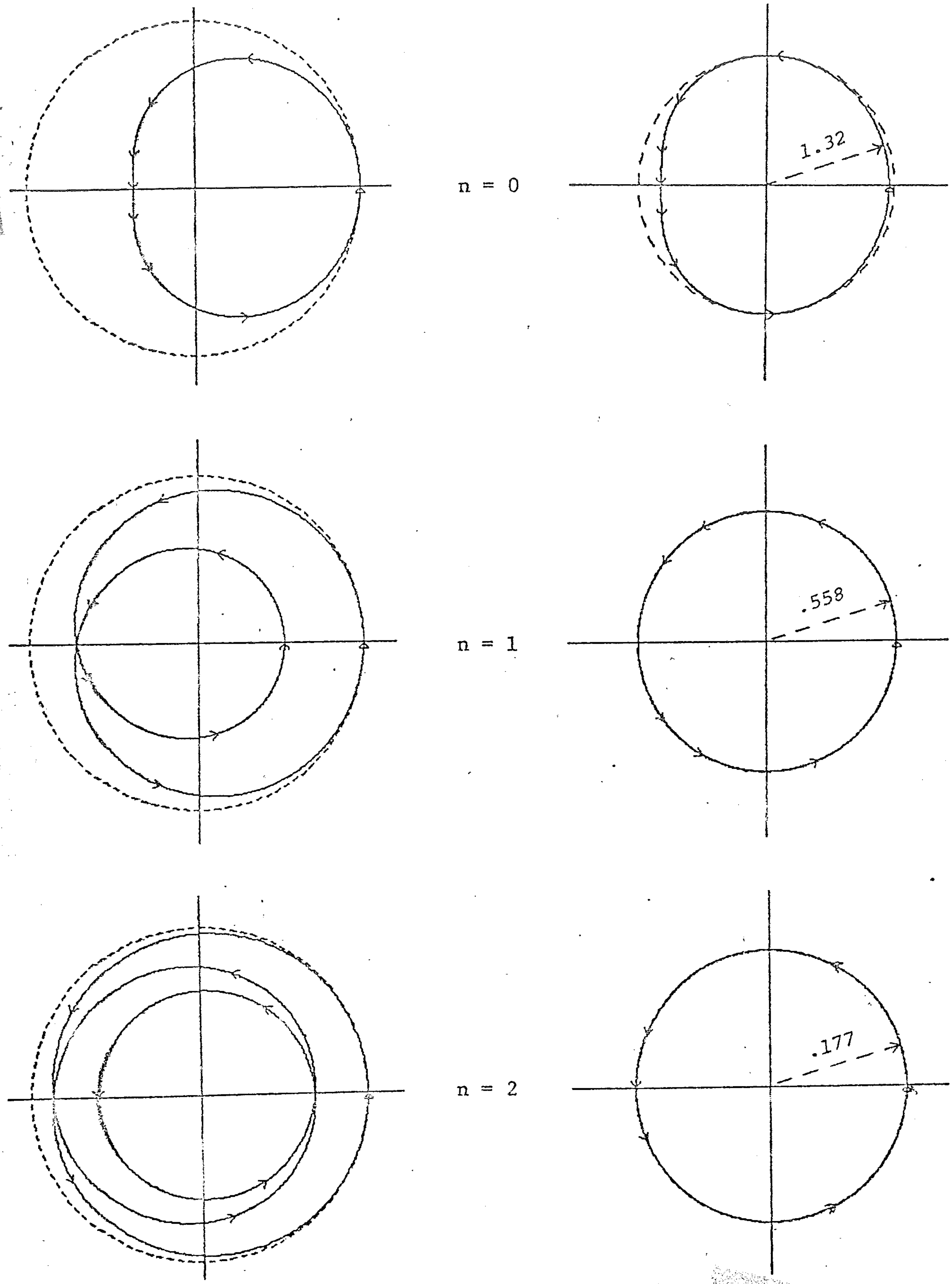
I have computed and plotted a large number of polynomial best approximations over the unit disk, and found clear evidence that in a wide range of situations the corresponding error curves closely approximate circles along most or all of their length. Judging by appearances, it would seem that Chebyshev approximation polynomials may be characterizable as those polynomials which achieve error curves that are in some sense most nearly circular. However, although this possibility has been a primary focus of my efforts, I have been unable to prove a satisfactory result along these lines.

As a first set of examples, Figure 6 shows the error curves corresponding to best quadratic approximation over the unit disk to four functions which are analytic in the disk: $\sqrt{z-2}$, $\ln(z-2)$, $\Gamma(z+3)$, and $z^4 + \frac{1}{2} z^5$. In Figures 6a-c the curve looks like a circle (with winding number 3), and in fact none of these curves deviates from a perfect circle by more than 0.5%. In Figure 6d the curve is by no means a circle, as it loops in and barely around the origin, and yet even in this case a nearly circular path is traversed during two of the three main circuits around the origin. These examples are representative of the functions I have considered.

To illustrate that the nearly-circular shape of the error curve is a feature peculiar to Chebyshev approximation, Figure 7 compares the Chebyshev approximation error curves for e^z on the unit disk, n=0,1,2, with the non-Chebyshev error curves corresponding to the first three partial sums of the Taylor series for e^z: 1, 1+z, and $1 + z + \frac{z^2}{2}$. To facilitate comparison, the two plots for each polynomial degree are drawn on the same scale. On the unit circle the least-squares approximation to an analytic function is always the same as the corresponding



<u>Figure 6</u>. Error curves $(f-p^*)(C)$ corresponding to best degree-2 polynomial approximation to $f(z)$ over the unit disk, for four functions f analytic in the disk.



<u>Figure 7</u>. Error curves for polynomial approximation to e^z over the unit disk. Left column shows partial sums of the Taylor series; right column shows Chebyshev approximations. Each pair is plotted on a single scale.

partial sum of the function's Taylor series, so from Figure 7 it is evident that least-squares approximation over the unit circle does not lead to nearly-circular error curves in the way that Chebyshev approximation does. It is clear also that the extent of circularity in Chebyshev approximation error curves is much greater than can be accounted for simply as the result of the (n+1)-degree term of the e^z Taylor series dominating the subsequent ones. Something more is going on here.

Figure 8 makes the same comparison in another form: it plots E and $E-E_{min}$ as functions of degree n for both Chebyshev approximation and approximation by truncated Taylor series. The rate at which $E-E_{min}$ approaches 0 for Chebyshev approximation is striking.

Most of my experimentation has been on the unit disk. A few examples computed over non-circular domains, however, suggest two observations. First, best approximation error curves are typically less circular than in approximation over the disk. Second, it seems true nevertheless that as the degree n increases, the error curves become increasingly circular.

To illustrate, Figure 9 considers best polynomial approximations of degrees $n = 0,2,4,8$ to the function $f(z) = \frac{1}{z-i}$ over the ellipse in the complex plane passing through the points $(2, \frac{1}{2}i)$, $(2, -\frac{1}{2}i)$, $(-2, \frac{1}{2}i)$, $(-2, -\frac{1}{2}i)$. f is analytic in this domain, but because of the pole at $z = i$ it cannot be expressed as a Taylor series about the origin. Even so, it appears that for $n \to \infty$ the error curve approaches a perfect circle.

What about regions whose boundaries are not smooth? Figure 10 shows error curves for best polynomial approximation to e^z over the unit square, again for $n = 0,2,4,8$. As predicted by Theorem 10b, four right angles appear in each error curve, and it appears that because of

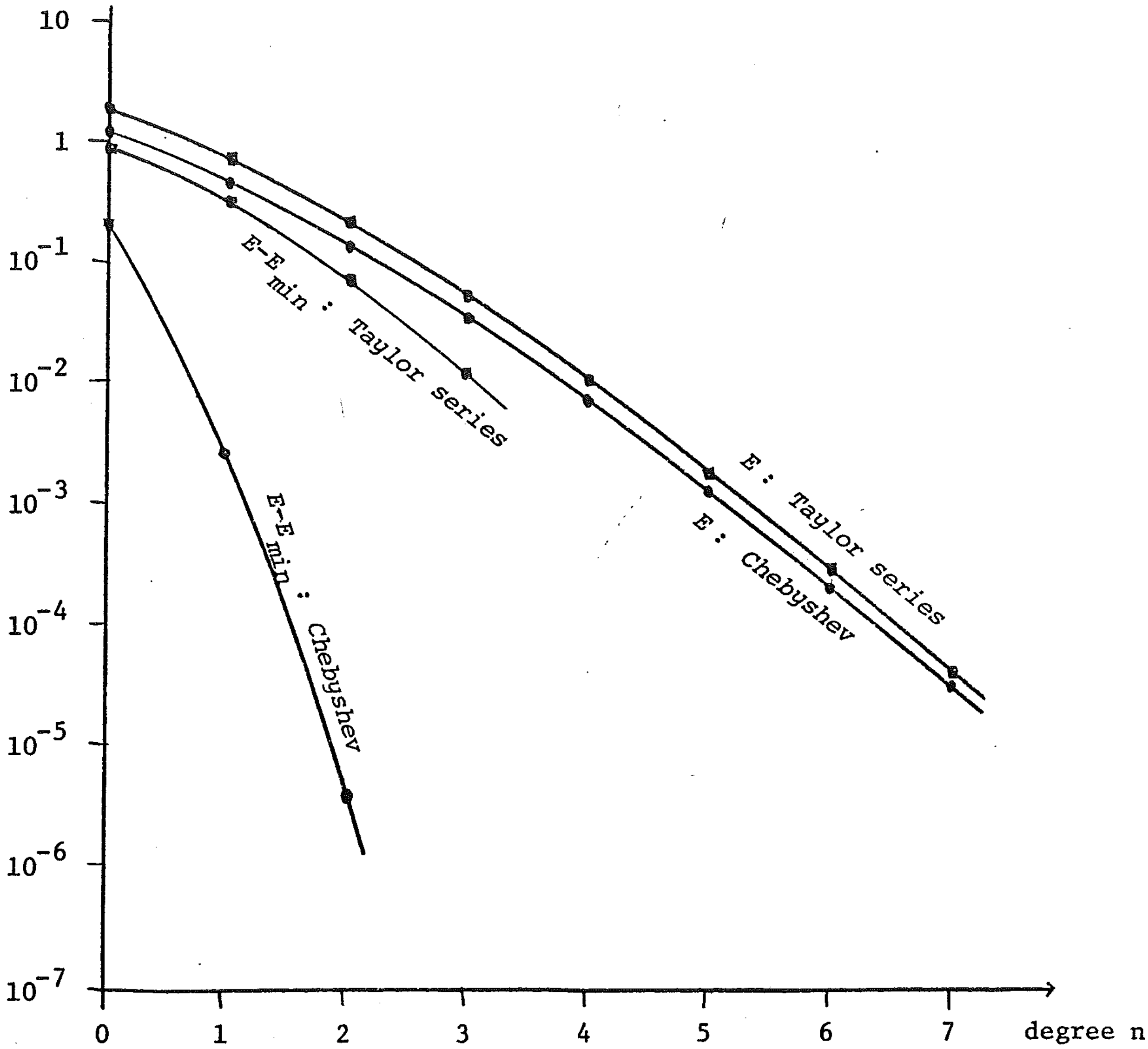


Figure 8. E and $E - E_{min}$ as functions of degree n for approximation of e^z on the unit disk by partial sums of the Taylor series and by Chebyshev approximations.

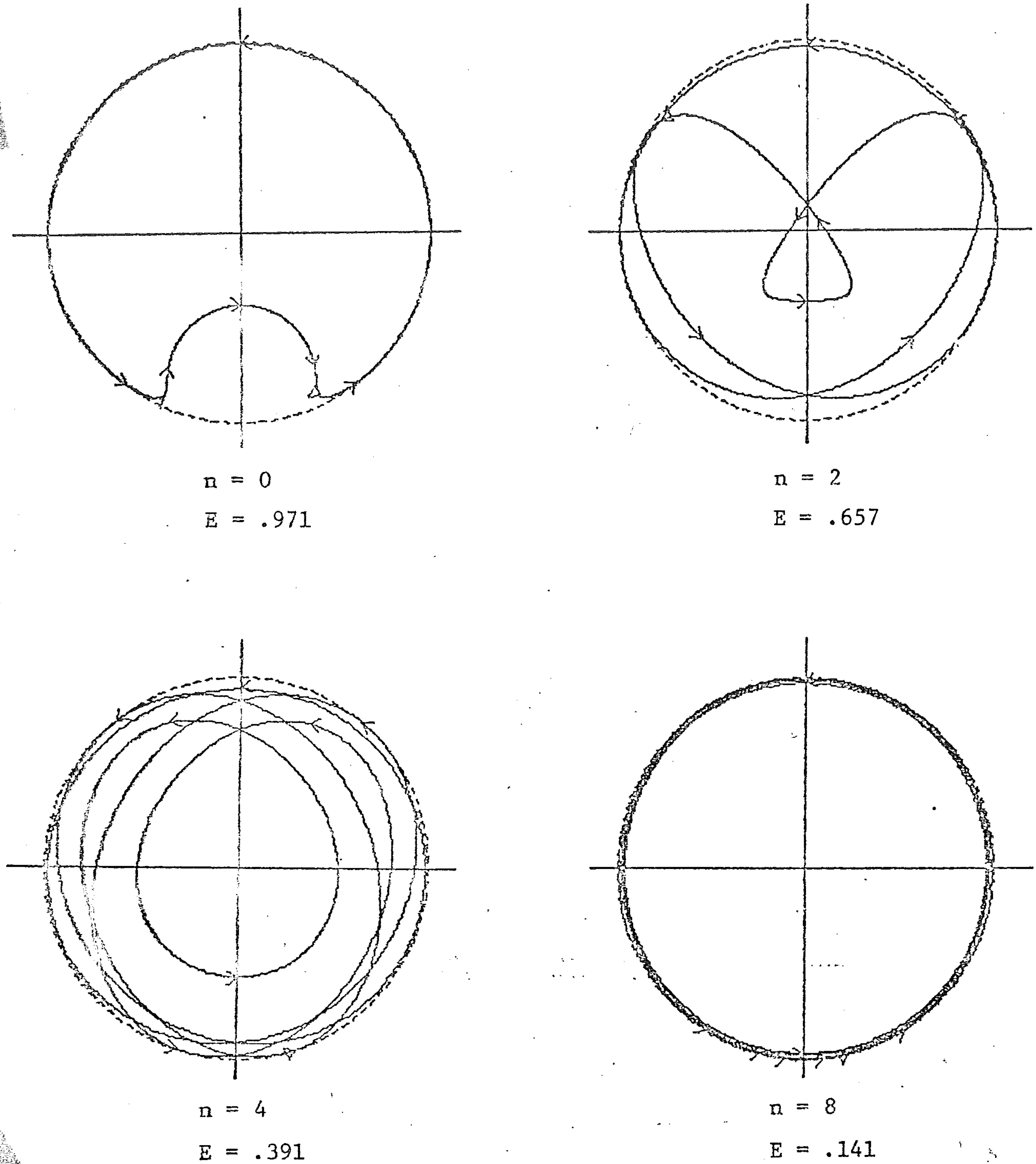
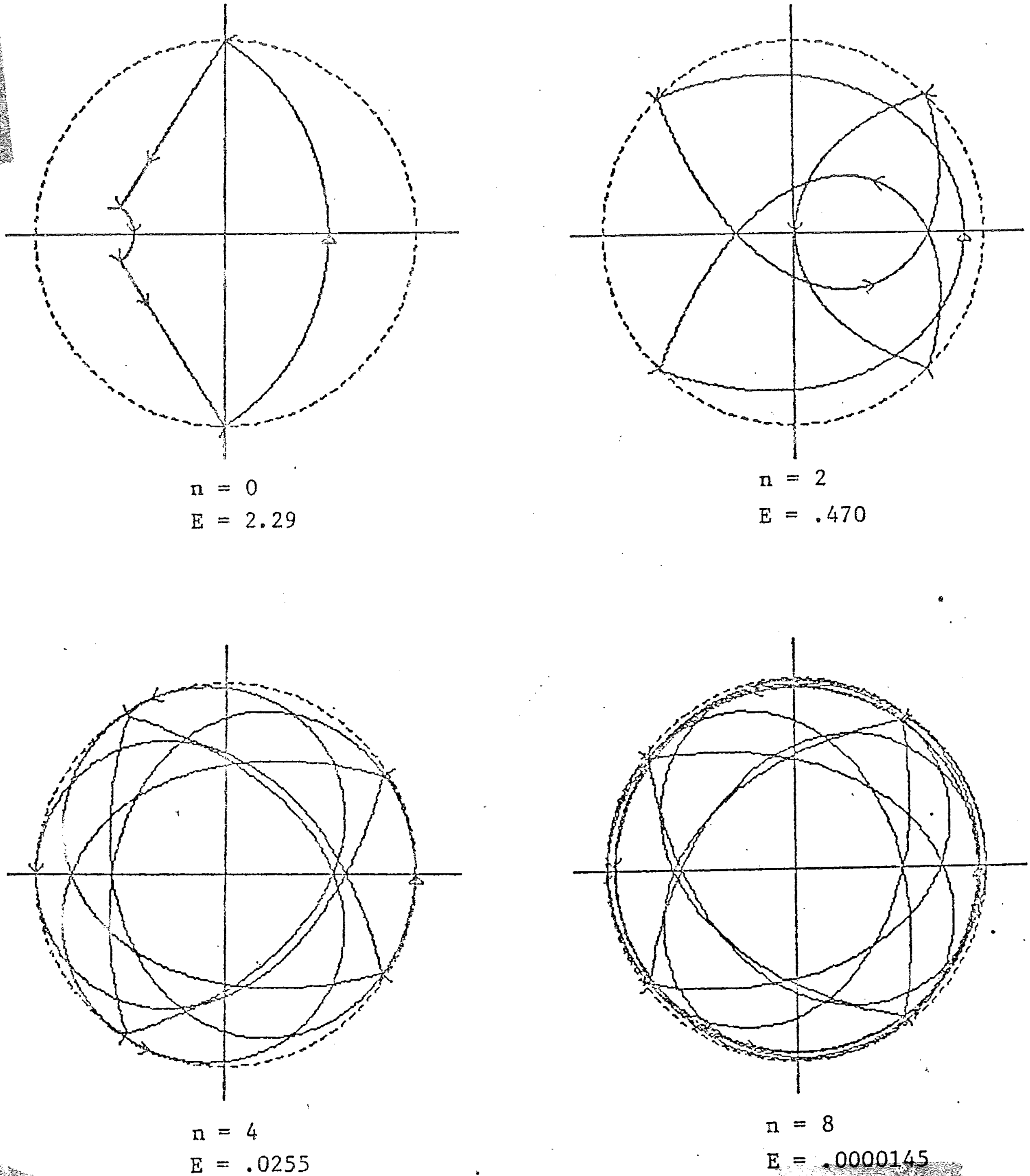


Figure 9. Error curves for best polynomial approximation to $f(z) = \frac{1}{z-i}$ over the ellipse passing through the points $(2,\frac{1}{2}i)$, $(2,-\frac{1}{2}i)$, $(-2,\frac{1}{2}i)$, $(-2,-\frac{1}{2}i)$.



<u>Figure 10</u>. Error curves for best polynomial approximation to $f(z) = e^z$ over the unit square.

them $\frac{E_{min}}{E}$ will not approach 1 as n approaches infinity. But will the average of $\frac{|e(z)|}{E}$ over the unit square approach 1? It appears that the tendency of best approximation error curves to approach a circular shape in some measure is not limited to approximation on a disk.

I have not computed any rational best approximations, but I have plotted error curves for the few numerical examples of them which I have been able to obtain from others ([13]). The curves in these examples have been very nearly circular with winding numbers m+n+1 for approximations of type (m,n). Thus the tendency for best approximation error curves to approximate circles seems not to be limited to polynomial approximation either. It would seem that this circularity is in some way a general feature of the supremum norm for complex approximation.

For further examples, with plots and numerical coefficients, see Appendix C.

3. Beginnings of an explanation

As a first step in the direction of explaining the occurrence of nearly-circular error curves, we can prove from Rouché's theorem that in the extreme case in which an error curve $e(C) = (f-r_{mn})(C)$ is perfectly circular, with sufficiently high winding number, then $r_{mn}(z)$ is necessarily a best approximation to $f(z)$. This result appeared to be new until I discovered that it has been proved independently for the case of the unit disk by V. Klotz ([21]) earlier this year. The extension from polynomial to rational approximation is his.

Theorem 11. *Let K be a Jordan region in the plane with boundary C, and let $f(z)$ be analytic in the interior of K and continuous on C. Suppose that $r(z)$ is a rational approximation to f of type (m,n) such that the error function $f(z)-r(z)$ maps C onto a perfect circle around the origin with positive winding number $\geq m+n+1$. Then r is a best rational approximation to f over K.*

Proof. Suppose on the contrary that there exists some rational function $\tilde{r}$ of type (m,n) such that $\|f-\tilde{r}\| < \|f-r\|$ over K, hence the same over C. Since $f(z)-r(z)$ is circular, it follows that $|f(z)-\tilde{r}(z)|<|f(z)-r(z)|$ for every $z \in C$. Therefore, by Rouché's theorem, $r-\tilde{r}$ must have the same number (#zeros − #poles) interior to K as $f-r$, which is at least $m+n+1$. Hence $r-\tilde{r}$ must have at least $m+n+1$ zeros within K. This is impossible since this difference is of type $(m+n,2n)$. Q.E.D.

As an immediate corollary, we may note that for any $N>n$, the best degree-n polynomial approximation to z^N over a disk centered at the origin is $p^*(z) = 0$, for $(z^N-0)(C)$ is in this case a perfect circle with positive winding number $N>n$. Thus the analogs of the Chebyshev polynomials over the unit disk rather than the interval $[-1,1]$ are just the monomials $1, z, z^2, \ldots$.

The direct applicability of Theorem 11 is limited, because only a rational function can have circular error curves on a circular domain. The following proof was suggested to me by Lars Ahlfors.

Theorem 12. *Let K be a disk in the plane bounded by a circle C, and let $f(z)$ be meromorphic in the interior of K and continuous on C. Suppose $r(z)$ is a rational function with the property that $f(z)-r(z)$ maps C onto a perfect circle around the origin. Then f must be a* *rational function.*

Proof. By the principle of symmetry ([41], p. 219), $f(z)-r(z)$ can be extended to a function meromorphic in the plane whose zeros and poles are symmetric with respect to C. Such a function is necessarily meromorphic in the extended plane, hence rational ([19], vol. 1, p. 217). Therefore f is rational also. Q.E.D.

4. Possible next steps

But the main question remains unanswered. In what sense and under what circumstances do best approximations error curves approximate circles? How do we account for Figure 2? I will close this section by mentioning some of the approaches to the problem which I have considered. Perhaps one can lead to the answers sought.

Carathéodory-Fejér. On the unit disk, the problem of approximating an analytic function by polynomials may be put this way: given the tail of a Taylor series, $f(z) = c_{n+1}z^{n+1} + c_{n+2}z^{n+2} + \ldots$, what partial sum $P^*(z) = c_0 + c_1 z + \ldots + c_n z^n$ minimizes the error $\|f-p^*\|$? Carathéodory ([5]) and Fejér in effect considered the reverse situation: if $p(z) = c_0 + c_1 z + \ldots + c_n z^n$ is given, what tail $f^*(z) = c_{n+1}z^{n+1} + c_{n+2}z^{n+2} + \ldots$ minimizes the error $\|f^*-p\|$ over the unit disk? The answer, they found, is that f^* is always a rational function, and $(f^*-p)(C)$ is a perfect circle with winding number n. (Part of this result may be proved by an application of Rouché's theorem exactly parallel to that used in the proof of Theorem 11.)

This example reinforces the idea that circular error curves are a rather general feature of best approximation in the complex plane. Can it be exploited?

Different measures of near-circularity. If best Chebyshev approximations can be characterized somehow in terms of nearly-circular error curves, it is probable that "nearly circular" should not be defined in terms of maximum relative deviation from the circle of maximum error. Examples are common like that of Figure 6d, in which the error curve is circular over most of its length but dips far from the circle of maximum error at its minimum magnitude. Other candidates for definitions of nearness are the average modulus over the circle of $|e(z)|$, and the area (computed with multiplicity) of the region enclosed by e(C). Simple and potentially useful properties of each of these quantities are reported in some of the well-known works on complex analysis (cf. [45], p. 174, [42], p. 186).

Perturbation and deformation. We know (Theorem 11) that circular error curves of sufficiently high winding numbers correspond to best rational approximations. Given a Jordan region K and an analytic function f on K, can we perturb f slightly so as to achieve a function $\tilde{f}$ for which there is an approximation function whose error curve is perfectly circular? Alternately, can we leave f fixed but deform K slightly with the same effect? In either approach one would aim to show that since the best approximation error curve in a nearby problem is exactly circular, the error curve in the given problem is nearly circular.

Expansion from a point to the given region. As $r \to 0$, the best rational approximation $R_{mn}(f;r)$ to an analytic function f over the disk of radius r approaches the Padé fraction $P_{mn}(f)$, and its error curve approaches a perfect circle ([51]). Starting instead at r=0 and moving outward, can we show that as r increases the drift of $R_{mn}(f;r)$ away from $P_{mn}(f)$ has the effect of maintaining that circular shape to the maximum extent possible?

V. CONCLUSION

Plenty remains to be done in complex Chebyshev approximation, polynomial or otherwise. Here are three questions which are not too difficult to tackle, whose answers would be very good to have.

1. No reasonably general examples of best approximations to analytic functions on the complex disk seem to be known. The sources I have seen which give explicit polynomial approximations in the complex plane are [2], [21], [28], [35], and [37]; none of these gives a reasonably nontrivial best approximation over the disk whose error curve is not perfectly circular, and examples with circular error curves are certainly extreme cases (Theorem 12). For example, can we find analytic expressions for the best polynomial approximations to e^z over the disk of radius r about the origin? Further computational work along the lines described on p. 27 might suggest probable solutions to such a question, which could then perhaps be confirmed theoretically.

2. In what sense do best polynomial approximation error curves over the unit disk approximate circles, either for fixed n or asymptotically for $n \to \infty$? How does the answer extend to other domains, or to approximation by rational functions? Can best approximations be characterized in some way in terms of the shape of their error curves?

3. We have shown that Lawson's algorithm is unsatisfactory for complex Chebyshev approximation because it converges too slowly whenever the best approximation error curve is nearly circular. Can the algorithm be altered to get around this problem? Can one design an algorithm which, rather than failing as a circular error curve is approached, takes advantage of the fact that the limit error function is nearly circular?

APPENDIX A. PROOFS OF THE BASIC THEOREMS

Runge's Theorem

(Proof from Grabiner [18] 1976.)

Lemma 1. *Let K be a compact set in the plane and let f(z) be analytic throughout a neighborhood of K. Then for any $\varepsilon>0$, there exists a rational function Q(z), all of whose poles lie in the complement of K, such that $\|f-Q\| < \varepsilon$ over K.*

Proof ([41], Lemma IV-1.1). Let H be an open region containing K in which f is analytic. We will need to make use of a rectifiable system of curves Γ contained in H-K such that Cauchy's integral formula holds on Γ:

$$\frac{1}{2\pi i}\int_{\Gamma}\frac{f(\zeta)}{\zeta-z}\,d\zeta = f(z) .$$

To find such a Γ, cover H by a grid of closed squares fine enough so that every square which has points in common with K is contained in H. It can then be shown that the boundary segments of the system of those squares which have at least one point in common with K will serve for Γ. (For details see [41], Lemma III-10.1).

Now let $\zeta=\zeta(t)$ be a parametrization of Γ, with $0\le t\le 1$. $\zeta'(t)$ exists at all but a finite number of points of Γ; let M be the maximum of $|\zeta'(t)|$ where it exists. The function $\frac{f(\zeta(t))}{\zeta(t)-z}$ is a continuous function of t and z for $z\varepsilon K$, so we can divide [0,1] into a finite number of subintervals $[t_i, t_{i+1}]$, $i=0,1,\ldots n-1$, such that

$$\left| \frac{f(\zeta(t))}{\zeta(t)-z} - \frac{f(\zeta(t_i))}{\zeta(t_i)-z} \right| < \frac{\varepsilon}{M}, \qquad t\varepsilon[t_i, t_{i+1}],\ z\varepsilon K.$$

Hence, taking

$$Q(z) = \sum_{i=0}^{n-1} \frac{f(\zeta(t_i))}{\zeta(t_i)-z} [\zeta(t_{i+1})-\zeta(t_i)],$$

we have for $z\varepsilon K$

$$\left| \int_\Gamma \frac{f(\zeta)}{\zeta-z} d\zeta - Q(z) \right|$$

$$= \left| \sum_{i=0}^{n-1} \int_{t_i}^{t_{i+1}} \left[\frac{f(\zeta(t))}{\zeta(t)-z} - \frac{f(\zeta(t_i))}{\zeta(t_i)-z}\right] \zeta'(t)\, dt \right|$$

$$< \frac{\varepsilon}{M} M = \varepsilon . \qquad \text{Q.E.D.}$$

Lemma 2. *Let B be the collection of functions on K which can be approximated arbitrarily closely by polynomials. Then for $\lambda \notin K$, $\frac{1}{z-\lambda}$ belongs to B.*

Proof. Let U be the complement of K in the plane and let $\rho = \{ \lambda : \frac{1}{z-\lambda} \varepsilon B\}$. For large enough λ the Taylor series for $\frac{1}{z-\lambda}$ converges uniformly on K, so ρ is not empty. If λ belongs to ρ and if $|\mu-\lambda| < \text{dist}(\lambda,K)$, then μ also belongs to ρ, since $\frac{1}{z-\mu} = \frac{1}{z-\lambda} (1 - \frac{\mu-\lambda}{z-\lambda})^{-1}$ is the uniformly converging sum of a series in powers of $\frac{1}{z-\lambda}$. Thus ρ is an open subset of U. Now suppose that μ is a boundary point of ρ, and choose a sequence $\{\lambda_n\}$ in ρ with limit μ. Since μ does not belong to ρ we must have $|\mu-\lambda_n| \geq \text{dist}(\lambda_n,K)$ for all n. In the limit this gives $\text{dist}(\mu,K)=0$, so that $\mu \notin U$. Thus the boundary points of $\rho \subseteq U$ are disjoint from U, and hence $\rho=U$. Q.E.D.

Similarly,

$$\left|\frac{\partial V}{\partial x}\right|, \left|\frac{\partial U}{\partial y}\right|, \left|\frac{\partial V}{\partial y}\right| < \frac{3\omega(\delta)}{\delta} .$$

Hence, if we define

$$\Psi(z) = (\frac{\partial U}{\partial x} - \frac{\partial V}{\partial y}) + i(\frac{\partial U}{\partial y} + \frac{\partial V}{\partial x}) , \tag{8}$$

then

$$|\Psi(z)| < \frac{12\omega(\delta)}{\delta} . \tag{9}$$

Now $\Phi(z)$ is not analytic throughout D_n, so Cauchy's integral formula does not apply to it. But we can use Green's identity to construct a generalized form of Cauchy's integral formula to take advantage of the fact that Φ is "nearly" analytic, as represented by the bound (9). To do this, set $f(z) = \frac{1}{\zeta - z}$ for some $z\varepsilon\Gamma$, and let Γ' be the compound arc $L_n - \Gamma_\delta - C_\varepsilon$, where C_ε is a small circle of radius ε surrounding the point z. Then f(z) is analytic in the region bounded by Γ', and $\Phi(z)$, as noted earlier, has continuous partial derivatives in that region. If $f(z) = \alpha(z) + i\beta(z)$, we may write

$$\int_{\Gamma'} f(z)\Phi(z)dz = \int_{\Gamma'} [\alpha(z)U(z) - \beta(z)V(z)]dx - [\beta(z)U(z) + \alpha(z)V(z)]dy$$

$$+ i\int_{\Gamma'} [\beta(z)U(z) + \alpha(z)V(z)]dx + [\alpha(z)U(z) - \beta(z)V(z)]dy . \tag{10}$$

To this we apply Green's identity,

$$\int P(z)dx + Q(z)dy = \iint (\frac{\partial P}{\partial y} - \frac{\partial Q}{\partial x}) \, dx\, dy .$$

Using the fact that f is analytic, (10) becomes

$$\int_{\Gamma'} f(z)\Phi(z)dz = \iint f(z)\{[\frac{\partial U}{\partial y}(z) - \frac{\partial V}{\partial x}(z)] + i[\frac{\partial U}{\partial y}(z) - \frac{\partial V}{\partial x}(z)]\} \, dx\, dy ,$$

where the area integral is over the region bounded by Γ', which is $R_{n\delta}$

minus the disk around z bounded by C_ε. Now plugging in $f(\zeta) = \frac{1}{\zeta - z}$, and taking the limit $\varepsilon \to 0$, we obtain after some manipulation

$$\Phi(z) = \frac{1}{2\pi i}\int_{L_n} \frac{\Phi(\zeta)d\zeta}{\zeta - z} - \frac{1}{2\pi i}\int_{\Gamma_\delta} \frac{\Phi(\zeta)d\zeta}{\zeta - z} + \frac{1}{2\pi}\iint_{R_{n\delta}} \frac{\Psi(\zeta)d\xi d\eta}{\zeta - z},$$

which we write in the form

$$\Phi(z) = I_1(L_n;z) + I_2(\Gamma_\delta;z) + I_3(R_{n\delta};z). \tag{11}$$

The integral $I_1(L_n;z)$ is analytic on K, and therefore by Runge's theorem it can be approximated arbitrarily closely by polynomials. The integral $I_2(\Gamma_\delta;z)$ vanishes, by virtue of (7) and the fact that ϕ is analytic on G_δ. Furthermore, $I_3(R_{n\delta};z)$ approaches $I_3(R_\delta;z)$ uniformly for z in K as $n \to \infty$. Combining these facts and taking (6) into account also, we have now shown that to prove Mergelyan's theorem it is sufficient to show that

$$\inf_{P(z)} \| P - I_3(R_\delta;\cdot)\|_\Gamma \to 0 \quad \text{as } \delta \to 0, \tag{12}$$

where P(z) ranges through all polynomials and $\| \ \|_\Gamma$ is the supremum norm over Γ.

To establish (12) we need a lemma.

<u>*Lemma*</u>. *Given δ and any point $\zeta \varepsilon R_\delta$, there exists a polynomial $P_\zeta(z)$ for which*

$$|P_\zeta(z)| < \frac{A}{\delta}, \quad z \varepsilon K \tag{13}$$

and

$$\left|P_\zeta(z) - \frac{1}{\zeta - z}\right| < \frac{B\delta^2}{|\zeta - z|^3}, \quad z\varepsilon K,\ |\zeta - z| > 16\delta, \tag{14}$$

where A and B are constants independent of δ, and for each δ the degree of $P_\zeta(z)$ is bounded by a number independent of ζ.

Proof. Without loss of generality let us assume $\zeta=0$. Since the complement of K is connected, it contains some Jordan arc ℓ with endpoints t_1 and t_2 such that $|t_1|=2\delta$, $|t_2|=3\delta$, and which lies within the disk $|z|<4\delta$. Let d be a Jordan region which contains ℓ and has no points in common with either K or the circle $|z|=4\delta$.

Let $z=\tau(w)$ map the region $|w|>4\delta$ conformally onto the complement of d in such a way that the points at infinity correspond to each other. Denote by D the image of d under the transformation $z=\frac{w}{4\delta}$. Then the function $z=\frac{\tau(4\delta w)}{4\delta}$ maps $|w|>1$ conformally onto the complement of D, and it may be written in the form

$$z = F(w) = \frac{\tau(4\delta w)}{4\delta} = aw+b+\frac{a_1}{w}+\frac{a_2}{w^2}+\dots.$$

D lies in the disk $|w|<1$, and by choice of the endpoints t_1 and t_2 its diameter is at least ¼. It can therefore be shown using arguments of capacity which we shall not go into ($|a|$ is the capacity of D; see [1], Section 2-3) that we have

$$\frac{1}{16} < |a| < 1 . \tag{15}$$

The function $F(w)-aw=b+\frac{a_1}{w}+\frac{a_2}{w^2}+\dots$ is analytic in $|w|>1$ and at infinity, and so attains its maximum on the unit circle, which must be no more than 2. By applying Cauchy's estimate to $F(\frac{1}{w})$ on the unit disk, we see that

$$|b|<2, \quad |a_k|<2, \quad k=1,2,\dots, \tag{16}$$

and therefore

$$|F(w)-aw-b| < \frac{4}{|w|} \quad \text{for } |w|>2 .$$

It follows that

$$|\tau(w)-4\delta b-aw| < 64\frac{\delta^2}{|w|} \quad \text{for } |w|>8\delta . \tag{17}$$

Now $w=\tau^{-1}(z)$ is a conformal transformation of the complement of d onto the region $|w|>4\delta$, and by applying the minimum modulus principle to $\frac{\tau^{-1}(z)}{z}$ we see that $|\tau^{-1}(z)| > |z|$ throughout the complement of d. Using this fact, for $|z|>8\delta$ we compute

$$\left|\frac{1}{z-4\delta b} - \frac{1}{a\tau^{-1}(z)}\right| = \left|\frac{a\tau^{-1}(z) - z + 4\delta b}{a\tau^{-1}(z)(z-4\delta b)}\right|$$

$$\leq \frac{64\delta^2}{|a||\tau^{-1}(z)|^2|z-4\delta b|} \qquad \text{(by (17))}$$

$$< \frac{64\delta^2}{|a|\ |z|^2\ |z-4\delta b|} \leq \frac{1024\delta^2}{|z|^2\ |z-4\delta b|} \qquad \text{(by (15))}.$$

If we take $|z|>16\delta$, then by (16) this becomes

$$\left|\frac{1}{z-4\delta b} - \frac{1}{a\tau^{-1}(z)}\right| < \frac{2048\delta^2}{|z|^3}, \qquad |z|>16\delta\ . \tag{18}$$

On the set K the function $\frac{1}{a\tau^{-1}(z)}$ is analytic, with the bound

$$\left|\frac{1}{a\tau^{-1}(z)}\right| < \frac{1}{|a|4\delta} < \frac{4}{\delta} \qquad \text{(by (15))}, \tag{19}$$

so by Runge's theorem and (18) there exists a polynomial Q(z) such that

$$\left|\frac{1}{z-4\delta b} - Q(z)\right| < \frac{2048\delta^2}{|z|^3}, \qquad z\varepsilon K,\ |z|>16\delta\ , \tag{20}$$

$$|Q(z)| < \frac{4}{\delta}, \qquad z\varepsilon K\ . \tag{21}$$

We now may compute

$$\left|\frac{4\delta b}{(z-4\delta b)^2} - 4\delta b[Q(z)]^2\right| = 4\delta|b|\ \left|\{\frac{1}{z-4\delta b} - Q(z)\}\{\frac{1}{z-4\delta b} + Q(z)\}\right|$$

$$< 8\delta\{\frac{2048\delta^2}{|z|^3}\}\left|\frac{1}{z-4\delta b} + Q(z)\right| \qquad \text{(by (16) and (20))},$$

and thus by (21), for a suitable constant A_1,

$$\left|\frac{4\delta b}{(z-4\delta b)^2} - 4\delta b[Q(z)]^2\right| < A_1 \frac{\delta^2}{|z|^3}, \qquad z\varepsilon K, \ |z|>16\delta . \tag{22}$$

From the equation

$$\frac{1}{z} = \frac{1}{z-4\delta b} - \frac{4\delta b}{(z-4\delta b)^2} + \frac{16\delta^2 b^2}{z(z-4\delta b)^2}$$

we have by (20), (22) and (15)

$$\left|\frac{1}{z} - [Q(z) - 4\delta b Q^2(z)]\right| < \frac{2048\delta^2}{|z|^3} + \frac{A_1\delta^2}{|z|^3} + \frac{64\delta^2}{|z|\ |z-4\delta b|^2}$$

$$< \frac{A_2\delta^2}{|z|^3}, \qquad z\varepsilon K, \quad |z|>16\delta$$

for some constant A_2.

By this inequality and (21), we may take $P_0(z) = Q(z) - 4\delta b Q^2(z)$ and satisfy conditions (13) and (14) of the Lemma. Now by continuity any $P_\zeta(z)$ satisfies (13) and (14) not just at ζ, but throughout some neighborhood of ζ. K is compact, so a finite number of such neighborhoods cover it. Hence a finite collection of polynomials P_ζ will satisfy (13) and (14) at every point in K, and so we may bound the degree of P_ζ by a number independent of ζ. This completes the proof of the Lemma. Q.E.D.

It remains to apply the Lemma to demonstrate (12). Given δ, the degree of P_ζ is bounded, and so the function

$$P(z) = \frac{1}{2\pi}\iint_{R_\delta} \Psi(\zeta)P_\zeta(z)\,d\xi\,d\eta$$

is a polynomial. We have

$$|P(z) - I_3(R_\delta;z)| \le \frac{1}{2\pi}\iint_{R_\delta} |\Psi(z)|\left|P_\zeta(z) - \frac{1}{\zeta-z}\right| d\xi\,d\eta .$$

To evaluate this, break R_δ into two parts: the part outside the disk

$|z| \leq 16\delta$, and the part inside. By (9) and (14), the integral over the former is of order $w(\delta)$. By (9) and (13), the integral over the latter is also of order $w(\delta)$. Thus (12) is established, and Mergelyan's theorem is proved.

Existence of best approximations

Theorem 4 (Existence of best approximations). Let $f(z)$ be continuous on a compact set K in the plane. Then for any $n \geq 0$, there exists a degree-n polynomial $p^*(z)$ of best approximation to f over K; that is, there exists a polynomial $p^*(z)$ such that $\|f-p^*\| \leq \|f-p\|$ for every degree-n polynomial $p(z)$.

Proof ([11], pp. 138-139). Given f(z) consider two functions defined on $\mathbf{C}^{n+1}$:

$$d(a_0,\ldots,a_n) = \left\| f - \sum_{k=0}^{n} a_k z^k \right\| ,$$

$$h(a_0,\ldots,a_n) = \left\| \sum_{k=0}^{n} a_k z^k \right\| .$$

Both d and h are continuous. Let S denote the sphere

$$S: \sum_{k=0}^{n} |a_k|^2 = 1$$

in $\mathbf{C}^{n+1}$. S is compact, and hence h must take on a minimum value m there. The case m=0 is possible only if $1,z,\ldots,z^n$ are linearly dependent over K, in which event K has no more than n points and f(z) can be approximated with zero error by an interpolation polynomial; so let us suppose m>0. If we now write

$$r = \left[\sum_{k=0}^{n} |a_k|^2 \right]^{\frac{1}{2}},$$

then

$$h(a_0,\ldots,a_n) = r \left\| \sum_{k=0}^{n} \frac{a_k}{r} z^k \right\| ,$$

so that

$$h(a_0,\ldots,a_n) \geq mr .$$

Moreover,

$$d = \left\| f - \sum_{k=0}^{n} a_k z^k \right\| \geq \left\| \sum_{k=0}^{n} a_k z^k \right\| - \| f \|$$

$$\geq mr - \| f \|.$$

From this inequality it is clear that as r approaches infinity, d must do the same. Accordingly, for suitably large R, d must have the same infimum over the closed ball

$$\sum_{k=0}^{n} |a_k|^2 \leq R^2$$

as over all of C^{n+1}. This ball is compact, and so d must attain a minimum at some point within it. This point is in the (n+1)-tuple of coefficients of a best approximation polynomial p*(z). Q.E.D.

Uniqueness of best approximations

Lemma. Let f(z) be continuous on a compact set K in the plane which contains at least n+2 points. Let p(z) be a best degree-n polynomial approximation to f(z) over K. Then the extremal set K_o on which f(z)−p*(z) attains its maximum magnitude contains at least n+2 points.*

Proof ([25], pp. 18-19). Suppose on the contrary that K_o contains s points z_k, with $s \le n+1$. Then we can construct an interpolation polynomial q(z) of degree n such that

$$q(z_k) = f(z_k) - p^*(z_k)\ , \qquad k=1,\dots,s\ .$$

On K_o,

$$[f(z) - p^*(z)]\,\overline{q(z)} = E^2 > 0\ ,$$

where $E = \|f-p^*\|$. f, p*, and q are all continuous, so there exists an open set K_1, $K_o \subset K_1 \subset K$, such that

$$\mathrm{Re}\{[f(z)-p^*(z)]\overline{q(z)}\} > \tfrac{1}{2}E^2\ , \qquad z\varepsilon K_1\ .$$

We will now show that for sufficiently small $\lambda > 0$, the polynomial $\tilde{p} = p^* + \lambda q$ is a closer approximation to f over K than p*, thus contradicting our assumption and proving the Lemma.

First, consider $z\varepsilon K_1$. Let M denote the maximum of $|q(z)|$ on K. Then

$$|f(z) - \tilde{p}(z)|^2 = |[f(z) - p^*(z)] - \lambda q(z)|^2$$

$$= |f(z) - p^*(z)|^2 - 2\lambda \mathrm{Re}\{[f(z)-p^*(z)]\overline{q(z)}\} + \lambda^2 |q(z)|^2$$

$$< E^2 - \frac{2\lambda E^2}{2} + \lambda^2 M^2 = E^2 - \lambda(E^2 - \lambda M^2)\ .$$

For $\lambda < \frac{E^2}{M^2}$ this quantity is less than E^2, so $\tilde{p}(z)$ is a closer approximation only if $\alpha=\tilde{\alpha}$. Thus $p^*(z_k) = \tilde{p}^*(z_k)$ for at least n+2 points of K, and this means p^* and $\tilde{p}^*$ must be identical. Q.E.D.

Kolmogorov's characterization

Theorem 6 (Kolmogorov's characterization). Let $f(z)$ be continuous on a compact set K in the plane. Let $p(z)$ be a polynomial of degree n, and let K_o be the extremal set of points z at which $|f(z)-p(z)| = \|f-p\|$. Then p is a best degree-n approximation to f over K if and only if, for any degree-n polynomial $q(z)$,

$$\max_{z\varepsilon K_o} Re\{[f(z)-p(z)]\overline{q(z)}\} \geq 0 . \tag{1}$$

Proof ([25], pp. 18-19). In the "if" direction, suppose p is given such that (1) holds for every q of degree n. For any $\tilde{p}(z)$ of degree n, take $q = p-\tilde{p}$. Then for some $z_o \varepsilon K_o$,

$$\mathrm{Re}\{[f(z_o)-p(z_o)]\overline{q(z_o)}\} \geq 0 ,$$

and therefore

$$|f(z_o)-\tilde{p}(z_o)|^2 = |f(z_o)-p(z_o)+q(z_o)|^2$$

$$= |f(z_o)-p(z_o)|^2 + 2\mathrm{Re}\{[f(z_o)-p(z_o)]\overline{q(z_o)}\} + |q(z_o)|^2$$

$$\geq |f(z_o)-p(z_o)|^2 = \|f-p\|^2 .$$

Thus p is a better approximation to f than $\tilde{p}$.

In the "only if" direction, the argument is just as in the proof of the lemma leading to Theorem 5. Suppose that $p^*(z)$ is a polynomial of best approximation but that (1) does not hold. Then there exists a polynomial q(z) of degree n such that

$$\max_{z\varepsilon K_o} \mathrm{Re}\{[f(z)-p^*(z)]\overline{q(z)}\} = -2\varepsilon$$

for some $\varepsilon>0$. f, p^*, and q are all continuous, so there exists an open set K_1, $K_o \subset K_1 \subset K$, such that

$$\mathrm{Re}\{[f(z)-p^*(z)]\overline{q(z)}\} < -\varepsilon , \qquad z\varepsilon K_1 .$$

We will now show that for sufficiently small $\lambda>0$, the polynomial $\tilde{p} = p^* + \lambda q$ is a closer approximation to f over K than p^*, thus contradicting our assumption and proving the theorem.

First, consider $z\varepsilon K_1$. Let M denote the maximum of $|q(z)|$ on K, and write $E = \|f-p^*\|$. Then

$$|f(z)-\tilde{p}(z)|^2 = |[f(z)-p^*(z)] - \lambda q(z)|^2$$

$$= |f(z)-p^*(z)|^2 = |[f(z)-p^*(z)]-\lambda q(z)|^2$$

$$< E^2 - 2\lambda\varepsilon + \lambda^2 M^2 = E^2 - \lambda(2\varepsilon - \lambda M^2) .$$

For $\lambda < \frac{2\varepsilon}{M^2}$ this quantity is less than E^2, so $\tilde{p}(z)$ is a closer approximation to $f(z)$ than $p^*(z)$ over K_1.

Second, consider $z\varepsilon K-K_1$. $K-K_1$ is a closed set, on which $|f(z)-\tilde{p}(z)|<E$, so for some $\delta>0$, $|f(z)-\tilde{p}(z)|\leq E-\delta$ on $K-K_1$. It follows that for $\lambda<\frac{\delta}{M}$,

$$|f(z)-\tilde{p}(z)| \leq |f(z)-p^*(z)| + \lambda|q(z)|$$

$$< E - \delta + \delta = E .$$

Thus $\tilde{p}(z)$ is also a closer approximation to $f(z)$ than $p^*(z)$ over $K-K_1$.

Q.E.D.

Remes's characterization

Theorem 7 (Remes's characterization). Let f(z) be continuous on a compact set K in the plane. Then p(z) is a best degree-n polynomial approximation to f if and only if for some $r \leq 2n+3$ there exist r points $z_1, \ldots, z_r$ in the extremal set K_o for p, and r numbers $w_1, \ldots, w_r$ with $w_k > 0$ and $\Sigma w_k = 1$, such that

$$\sum_{k=1}^{r} w_k \{f(z_k) - p(z_k)\} \overline{q(z_k)} = 0 \tag{1}$$

for every degree-n polynomial q.

Proof ([25], chap. 2). From Kolmogorov's characterization (Theorem 6) we know that p is best if and only if, for any polynomial q of degree n,

$$\max_{z \in K_o} \operatorname{Re}\{e(z)\overline{q(z)}\} \geq 0,$$

where $e(z) = f(z) - p(z)$ as usual. Putting $q(z) = \Sigma c_k z^k$, we see that this is equivalent to the statement that if for some coefficients $\{c_k\}$ and real a_o

$$\operatorname{Re}\{\sum_{k=0}^{n} \overline{c}_k e(z) \overline{z}^k\} \leq a_o \qquad \text{for all } z \in K_o, \tag{2}$$

then $a_o \geq 0$. Now the inequality (2) can be interpreted in terms of a closed half-space in $\mathbb{C}^{n+1}$. It states that for any $z \in K_o$, the point

$$w = (w_0, w_1, \ldots, w_n) = (e(z), e(z)\overline{z}, \ldots, e(z)\overline{z}^n), \quad z \in K_o \tag{3}$$

lies in the half-space of $\mathbb{C}^{n+1}$ defined by $\operatorname{Re} \sum_{k=0}^{n} \overline{c}_k w_k \leq a_o$. The statement that whenever (2) holds a_o must be ≥ 0 is therefore equivalent to the statement that the origin 0 in $\mathbb{C}^{n+1}$ is contained in the intersection of all half-spaces which contain all the points w given by (3).

K_o is compact, and so its continuous image $w(K_o)$ from (3) is compact also. Therefore, since the convex hull of a compact set in a finite-dimensional Euclidean space is compact, the convex hull of $w(K_o)$ is compact. Now it is the case that when a convex set is compact, it is the intersection of all the closed half-spaces containing it. Applying this fact, we see that the argument of the preceding paragraph leads to this conclusion: p is a best approximation polynomial if and only if the origin in $\mathbb{C}^{n+1}$ belongs to the convex hull of the set of points $w(K_o)$ given by (3).

The conclusion of Theorem 7 now follows from a theorem of Carathéodory ([8], p. 17) which states that in an n-dimensional Euclidean space every point of the convex hull of a set B is expressible as a convex linear combination of n+1 or fewer points of B. $\mathbb{C}^{n+1}$ is (2n+2)-dimensional, so the equation (1) is just such an expression for the point 0 in $\mathbb{C}^{n+1}$. Q.E.D.

APPENDIX B. PROGRAM LISTING

```
C PROGRAM CHEBY (CIRCULAR DISK VERSION)
C
C CONTROL PROGRAM FOR COMPUTING AND PLOTTING COMPLEX CHEBYSHEV
C APPROXIMATIONS.  CALLS THE FOLLOWING SUBROUTINES:
C   LAWSON: COMPUTES BEST POLYNOMIAL APPROXIMATION COEFFICIENTS
C   ERRCRV: PLOTS THE ERROR CURVE IN THE COMPLEX PLANE
C   ERRMAG: PLOTS THE MAGNITUDE OF THE ERROR AS A FUNCTION OF PT. NUMBER
C
C LLOYD N. TREFETHEN, FEBRUARY 1977
C
      IMPLICIT COMPLEX (C,F,Z)
      COMMON NPTS,ZSET(200),F(200),C(10),NDEG,SCAL,EMIN,SIGMA,EMAX
C
C SET CONTROL PARAMETERS:
      NPTS=128
      WRITE (6,204)
      ACCEPT 201,NDEG
C
C SET UP PLOTTER:
      CALL PLOTS(14.,11.)
C
C SET UP DISCRETE POINT NETWORK ZSET(K):
      WRITE (6,205)
      ACCEPT 206,RD
      DO 10 K=1,NPTS
      CTMP=CMPLX(0.,2.*3.14159/FLOAT(NPTS))
      ZSET(K)=CMPLX(RD,0.)*CEXP(CTMP*CMPLX(FLOAT(K),0.))
      F(K)=FUNCT(ZSET(K))
   10 CONTINUE
C
C PERFORM COMPUTATIONS AND PLOTTING:
      WRITE (6,202)
      ACCEPT 201,IGLD
      IF (IGLD.EQ.1) CALL PLOT(-.25,.5,-3)
      CALL LAWSON
      WRITE (6,207)
      ACCEPT 201,ICRV
      IF (ICRV.EQ.1) CALL ERRCRV
      IF (IGLD.EQ.1) CALL PLOT(.25,-.5,-3)
      CALL ERRMAG
      STOP
C
  201 FORMAT (I6)
  202 FORMAT (' GOULD(1) OR CALCOMP(0)?'/)
  204 FORMAT (' DEGREE?'/)
  205 FORMAT (' RADIUS OF DISK?'/)
  206 FORMAT (F10.0)
  207 FORMAT (' PLOT ERROR CURVE?'/)
      END
```

```
      COMPLEX FUNCTION FUNCT(Z)
C
C THIS IS THE FUNCTION TO BE APPROXIMATED.
C
      IMPLICIT COMPLEX (C,F,Z)
      FUNCT=CEXP(Z)
      RETURN
      END
```

```
      COMPLEX FUNCTION CP(Z)
C
C THIS IS THE APPROXIMATION POLYNOMIAL.
C
      IMPLICIT COMPLEX (C,F,W,Z)
      COMMON NPTS,ZSET(200),F(200),C(10),NDEG,SCAL,EMIN,SIGMA,EMAX
      CTMP=C(NDEG+1)
      IF (NDEG.LT.1) GO TO 6
      DO 5 I=1,NDEG
    5 CTMP=CTMP*Z+C(NDEG-I+1)
    6 CP=CTMP
      RETURN
      END
```

```
      SUBROUTINE LAWSON
C
C THIS IS AN IMPLEMENTATION OF LAWSON'S ALGORITHM FOR
C COMPLEX CHEBYSHEV APPROXIMATION IN IN A JORDAN REGION,
C BASED ON THE WORK OF S. ELLACOTT AND JACK WILLIAMS IN
C "LINEAR CHEBYSHEV APPROXIMATION IN THE COMPLEX PLANE
C USING LAWSON'S ALGORITHM," MATH. COMP. 30, NO. 133,
C PP. 35-44.
C
C LLOYD N. TREFETHEN, FEBRUARY 1977
C
      IMPLICIT COMPLEX (C,F,Z)
      COMPLEX Z(200),CMAT(10,10),CUL(10,10),CRHS(10)
      DIMENSION W(200),E(200)
      COMMON NPTS,ZSET(200),F(200),C(10),NDEG,SCAL,EMIN,SIGMA,EMAX
      NEXP=1
      N1=NDEG+1
C
C SET INITIAL WEIGHTS:
      DO 5 K=1,NPTS
    5 W(K)=1./FLOAT(NPTS)
C
C SET UP LINEAR SYSTEM OF EQUATIONS:
   10 WRITE (6,204)
      NWPR=0
      ACCEPT 205,NITT
      IF (NITT.GE.0) GO TO 12
      NITT=-NITT
      NWPR=1
   12 IF (NITT.EQ.0) GO TO 70
      DO 60 NIT=1,NITT
      DO 20 J=1,N1
      CSUM=(0.,0.)
      DO 15 K=1,NPTS
      CSUM=CSUM+CMPLX(W(K),0.)*F(K)*CONJG(ZSET(K))**(J-1)
   15 CONTINUE
   20 CRHS(J)=CSUM
      DO 30 J=1,N1
      DO 30 I=1,N1
      CSUM=(0.,0.)
      DO 25 K=1,NPTS
   25 CSUM=CSUM+CMPLX(W(K),0.)*CONJG(ZSET(K))**(I-1)*ZSET(K)**(J-1)
   30 CMAT(I,J)=CSUM
C
C SOLVE THE SYSTEM:
      CALL DECOMP(N1,CMAT,CUL)
      CALL SOLVE(N1,CUL,CRHS,C)
      CALL IMPRUV(N1,CMAT,CUL,CRHS,C,DIGITS)
      WRITE (6,202) (C(I),I=1,N1)
C
C COMPUTE ERROR FUNCTION FOR THE CURRENT SET OF COEFFICIENTS:
      EMIN=99999.
      EMAX=0.
      SIGSUM=0.
      SUM=0.
      DO 40 K=1,NPTS
      ETMP=CABS(F(K)-CP(ZSET(K)))
      EMIN=AMIN1(EMIN,ETMP)
      EMAX=AMAX1(EMAX,ETMP)
      SIGSUM=SIGSUM+W(K)*ETMP**2
      E(K)=ETMP**NEXP
   40 SUM=SUM+W(K)*E(K)
      SIGMA=SQRT(SIGSUM)
      EDIFF=EMAX-EMIN
      WRITE (6,203) EMIN,SIGMA,EMAX,EDIFF
C
C ADJUST THE WEIGHTS AND GO TO THE NEXT ITERATION:
   48 DO 50 K=1,NPTS
      W(K)=W(K)*E(K)/SUM
      IF (NWPR.EQ.1) WRITE (6,206) K,W(K),K,E(K)
   50 CONTINUE
      NEXP=3-NEXP
   60 CONTINUE
      GO TO 10
   70 CONTINUE
      RETURN
C
  202 FORMAT (/' ',5(2F11.8,' '))
  203 FORMAT (' EMIN:',F11.7,2X,'SIGMA:',F11.7,2X,'EMAX:',F11.7/
     1 ' EMAX-EMIN:',E12.5)
  204 FORMAT (/' HOW MANY MORE ITERATIONS?'/)
  205 FORMAT (I6)
  206 FORMAT (' W(',I3,') =',F9.7,3X,'E(',I3,') =',F9.7)
  207 FORMAT (' CHOP WEIGHTS TO ZERO?'/)
      END
```

```
      SUBROUTINE ERRCRV
C
C PLOTS THE ERROR CURVE (F-CP)(C) IN THE COMPLEX PLANE.
C
C PARAMETER EX CONTROLS EXAGGERATION OF THE DEVIATION FROM THE CIRCLE
C OF MAXIMUM ERROR.  WHEN EX=1 THE PLOT IS TO SCALE.
C
C LLOYD N. TREFETHEN, NOVEMBER 1976
C
      IMPLICIT COMPLEX(C,F,W,Z)
      COMMON NPTS,ZSET(200),F(200),C(10),NDEG,SCAL,EMIN,SIGMA,EMAX
    1 CONTINUE
C     WRITE (6,203)
C     ACCEPT 202,EX
      EX=1.
      WRITE (6,205)
      ACCEPT 206,NEXMP
      WRITE (6,201)
      ACCEPT 202,SCLMAX
      CALL PLOT(4.,4.0,-3)
C
C PLOT IMAGE POINTS:
      IF (SCLMAX.LT.0.) SCAL=3./(-SCLMAX*EMAX)
      IF (SCLMAX.EQ.0.) SCAL=3./EMAX
      IF (SCLMAX.GT.0.) SCAL=3./SCLMAX
      W=F(NPTS)-CP(ZSET(NPTS))
      W=SCAL*W*CMPLX((EMAX-EX*(EMAX-CABS(W)))/CABS(W),0.)
      CALL PLOT(REAL(W),AIMAG(W),3)
      DO 20 I=1,NPTS
      W=F(I)-CP(ZSET(I))
      W=SCAL*W*CMPLX((EMAX-EX*(EMAX-CABS(W)))/CABS(W),0.)
      CALL PLOT(REAL(W),AIMAG(W),2)
   20 CONTINUE
C
C MARK ARROWS:
      DO 30 I=1,8
      I1=I*NPTS/8
      I2=I1+1
      IF (I1.EQ.NPTS) I2=1
      Z=ZSET(I1)
      Z2=ZSET(I2)
      W=F(I1)-CP(Z)
      W=SCAL*W*CMPLX((EMAX-EX*(EMAX-CABS(W)))/CABS(W),0.)
      W2=F(I2)-CP(Z2)
      W2=SCAL*W2*CMPLX((EMAX-EX*(EMAX-CABS(W2)))/CABS(W2),0 )
      ZDIR=(W2-W)/CMPLX(25.*CABS(W2-W),0.)
      W3=W+(-1.5,1.)*ZDIR
      W4=W+(-1.5,-1.)*ZDIR
      CALL PLOT(REAL(W),AIMAG(W),3)
      CALL PLOT(REAL(W3),AIMAG(W3),2)
      IF (I.LE.7) CALL PLOT(REAL(W4),AIMAG(W4),3)
      IF (I.EQ.8) CALL PLOT(REAL(W4),AIMAG(W4),2)
      CALL PLOT(REAL(W),AIMAG(W),2)
   30 CONTINUE
C
C DRAW CIRCLE OF MAXIMUM ERROR:
      NPC=200
      IF (SCLMAX.GT.0.) NPC=200.*EMAX/SCLMAX
      CONST3=CMPLX(0.,2.*3.14159/FLOAT(NPC))
      IPEN=3
      DO 40 I=1,NPC
      W=EMAX*SCAL*CEXP(FLOAT(I)*CONST3)
      CALL PLOT(REAL(W),AIMAG(W),IPEN)
      IPEN=5-IPEN
   40 CONTINUE
C
C DRAW AXES:
      TMP=3.*1.1
      CALL PLOT(-TMP,0.,3)
      CALL PLOT(TMP,0.,2)
      CALL PLOT(0.,TMP,3)
      CALL PLOT(0.,-TMP,2)
C
C WRITE EMAX AND OTHER INFORMATION ON PLOT:
      CALL SYMBOL (-1.5,-3.5,.16,10HERROR MAX:,0.,10)
      CALL NUMBER (.15,-3.5,.16,EMAX,0.,7)
      CALL SYMBOL (-1.38,-4.04,.16,1H#,0.,1)
      CALL NUMBER (-1.2,-4.04,.16,FLOAT(NEXMP),0.,-1)
      CALL SYMBOL (-.5,-4.0,.08,6HDEGREE,0.,6)
      CALL NUMBER (.05,-4.0,.08,FLOAT(NDEG),0.,-1)
      CALL SYMBOL (.6,-4.0,.08,3HEXP,0.,3)
      CALL NUMBER (.9,-4.0,.08,EX,0.,2)
C
C FINISH UP:
      CALL PLOT(-4.,-4.,-3)
      RETURN
C
  201 FORMAT (' SCLMAX?'/)
  202 FORMAT (F10.0)
  203 FORMAT (' EXPANSION FACTOR:'/)
  205 FORMAT (' EXAMPLE NUMBER:'/)
  206 FORMAT (I5)
      END
```

```
      SUBROUTINE ERRMAG
C
C PLOTS THE ERROR MAGNITUDE ON A SCALE FROM EMIN TO EMAX.
C
      IMPLICIT COMPLEX(C,F,W,Z)
      COMMON NPTS,ZSET(200),F(200),C(10),NDEG,SCAL,EMIN,SIGMA,EMAX
      CALL PLOT(2.,8.2,-3)
      YCONST=4./FLOAT(NPTS)
      NDEG1=NDEG+1
C
C PLOT ERROR MAGNITUDE:
      SCAL=1.
      IF (EMAX.GT.EMIN) SCAL=1./(EMAX-EMIN)
      EMAG=SCAL*(CABS(CP(ZSET(NPTS))-FUNCT(ZSET(NPTS)))-EMIN)
      CALL PLOT(0.,EMAG,3)
      DO 20 I=1,NPTS
      Z=ZSET(I)
      EMAG=SCAL*(CABS(CP(Z)-F(I))-EMIN)
      Y=I*YCONST
      CALL PLOT(Y,EMAG,2)
   20 CONTINUE
C
C WRITE EMIN, SIGMA, AND EMAX ON PLOT:
      CALL SYMBOL (4.36,0.,.08,4HMIN:,0.,4)
      CALL NUMBER (4.76,0.,.08,EMIN,0.,7)
      CALL SYMBOL (4.2,.7,.08,6HSIGMA:,0.,6)
      CALL NUMBER (4.76,.7,.08,SIGMA,0.,7)
      CALL SYMBOL (4.36,.92,.08,4HMAX:,0.,4)
      CALL NUMBER (4.76,.92,.08,EMAX,0.,7)
C
C FINISH UP:
      CALL PLOT(6.,-8.2,999)
C
  204 FORMAT (' EMIN:',F12.7/' EMAX:',F12.7/)
      END
```

```
      SUBROUTINE DECOMP(NN,A,UL)
C
C DECOMPOSES COMPLEX MATRIX A INTO UL FORM.
C ADAPTED FROM FORSYTHE & MOLER, "COMPUTER SOLUTION OF LINEAR
C ALGEBRAIC SYSTEMS" (PRENTICE-HALL, 1967).
C
      DIMENSION SCALES(10),IPS(10)
      COMPLEX A(10,10),UL(10,10),PIVOT,EM
      COMMON /LINEAR/ IPS
      N=NN
C
C INITIALIZE IPS,UL AND SCALES:
      DO 5 I=1,N
      IPS(I)=I
      ROWNRM=0.0
      DO 2 J=1,N
      UL(I,J)=A(I,J)
      IF (ROWNRM.GE.CABS(UL(I,J))) GO TO 2
      ROWNRM=CABS(UL(I,J))
    2 CONTINUE
      IF (ROWNRM.EQ.0.) GO TO 4
      SCALES(I)=1./ROWNRM
      GO TO 5
    4 CALL SING(1)
      SCALES(I)=0.0
    5 CONTINUE
C
C GAUSSIAN ELIMINATION WITH PARTIAL PIVOTING:
      NM1=N-1
      IF (NM1.LT.1) GO TO 18
      DO 17 K=1,NM1
      BIG=0.0
      DO 11 I=K,N
      IP=IPS(I)
      SIZE=CABS(UL(IP,K))*SCALES(IP)
      IF (SIZE.LE.BIG) GO TO 11
      BIG=SIZE
      IDXPIV=I
   11 CONTINUE
      IF (BIG.NE.0.) GO TO 13
      CALL SING(2)
      GO TO 17
   13 IF (IDXPIV.EQ.K) GO TO 15
      J=IPS(K)
      IPS(K)=IPS(IDXPIV)
      IPS(IDXPIV)=J
   15 KP=IPS(K)
      PIVOT=UL(KP,K)
      KP1=K+1
      DO 16 I=KP1,N
      IP=IPS(I)
      EM=-UL(IP,K)/PIVOT
      UL(IP,K)=-EM
      DO 16 J=KP1,N
      UL(IP,J)=UL(IP,J)+EM*UL(KP,J)
   16 CONTINUE
   17 CONTINUE
   18 CONTINUE
      KP=IPS(N)
      IF (UL(KP,N).EQ.(0.,0.)) CALL SING(2)
      RETURN
      END


      SUBROUTINE SOLVE(NN,UL,B,X)
C
C SOLVES THE COMPLEX MATRIX EQUATION (UL)*X=B, WHERE UL IS
C IN U-L FORM AS PREPARED BY "DECOMP."  ADAPTED FROM FORSYTHE AND
C MOLER, "COMPUTER SOLUTION OF LINEAR ALGEBRAIC SYSTEMS."
C
      COMPLEX UL(10,10),B(10),X(10),SUM
      DIMENSION IPS(10)
      COMMON /LINEAR/ IPS
      N=NN
      NP1=N+1
C
      IP=IPS(1)
      X(1)=B(IP)
      IF (N.LT.2) GO TO 6
      DO 2 I=2,N
      IP=IPS(I)
      IM1=I-1
      SUM=(0.,0.)
      DO 1 J=1,IM1
    1 SUM=SUM+UL(IP,J)*X(J)
    2 X(I)=B(IP)-SUM
C
    6 IP=IPS(N)
      X(N)=X(N)/UL(IP,N)
      IF (N.LT.2) GO TO 5
      DO 4 IBACK=2,N
      I=NP1-IBACK
      IP=IPS(I)
      IP1=I+1
      SUM=(0.,0.)
      DO 3 J=IP1,N
    3 SUM=SUM+UL(IP,J)*X(J)
    4 X(I)=(X(I)-SUM)/UL(IP,I)
    5 RETURN
      END
```

```
      SUBROUTINE IMPRUV(NN,A,UL,B,X,DIGITS)
C
C IMPROVES THE LINEAR SOLUTION ITERATIVELY.  ADAPTED FROM FORSYTHE
C AND MOLER, "COMPUTER SOLUTION OF LINEAR ALGEBRAIC SYSTEMS."
C
      COMPLEX A(10,10),UL(10,10),B(10),X(10),R(10),DX(10),T,SUM
      DOUBLE PRECISION DAA,DAB,DXA,DXB,DSUMA,DSUMB
      N=NN
C
      EPS=1.0E-8
      ITMAX=16
      XNORM=0.
      DO 1 I=1,N
    1 XNORM=AMAX1(XNORM,CABS(X(I)))
      IF (XNORM.NE.0.) GO TO 3
      DIGITS=-ALOG10(EPS)
      GO TO 10
C
    3 DO 9 ITER=1,ITMAX
      DO 5 I=1,N
              DSUMA=0.D0
              DSUMB=0.D0
              DO 4 J=1,N
              DAA=REAL(A(I,J))
              DAB=AIMAG(A(I,J))
              DXA=REAL(X(J))
              DXB=AIMAG(X(J))
              DSUMA=DSUMA+DAA*DXA-DAB*DXB
              DSUMB=DSUMB+DAB*DXA+DAA*DXB
    4         CONTINUE
              SUMA=DSUMA
              SUMB=DSUMB
              SUM=CMPLX(SUMA,SUMB)
C             (THE ABOVE INNER COMPUTATION IS IN DOUBLE PRECISION.)
      SUM=B(I)-SUM
    5 R(I)=SUM
      CALL SOLVE(N,UL,R,DX)
      DXNORM=0.0
      DO 6 I=1,N
      T=X(I)
      X(I)=X(I)+DX(I)
      DXNORM=AMAX1(DXNORM,CABS(X(I)-T))
    6 CONTINUE
      IF (ITER.NE.1) GO TO 8
      DIGITS=-ALOG10(AMAX1(DXNORM/XNORM,EPS))
    8 IF (DXNORM.LE.EPS*XNORM) GO TO 10
    9 CONTINUE
C   ITERATION DID NOT CONVERGE:
      CALL SING(3)
   10 RETURN
      END
```

```
      SUBROUTINE SING(IWHY)
C
C PRINTS AN APPROPRIATE ERROR MESSAGE.  FROM FORSYTHE AND MOLER,
C "COMPUTER SOLUTION OF LINEAR ALGEBRAIC SYSTEMS."
C
      GO TO (1,2,3),IWHY
    1 WRITE (6,201)
      GO TO 10
    2 WRITE (6,202)
      GO TO 10
    3 WRITE (6,203)
   10 RETURN
C
  201 FORMAT (' MATRIX WITH ZERO ROW IN DECOMPOSE.')
  202 FORMAT (' SINGULAR MATRIX IN DECOMPOSE.  ZERO DIVIDE IN SOLVE.')
  203 FORMAT (' NO CONVERGENCE IN IMPRUV. MATRIX IS NEARLY SINGULAR.')
      END
```

APPENDIX C. COMPUTED EXAMPLES

The following five pages present near-best polynomial approximations over the unit disk to five functions which are analytic in the disk. The approximations have been computed with Lawson's algorithm as described in Section III. These are the functions considered:

$$f(z) = z^4 + z^5$$

$$f(z) = e^{e^z}$$

$$f(z) = \tan^{-1}(\tfrac{z}{2})$$

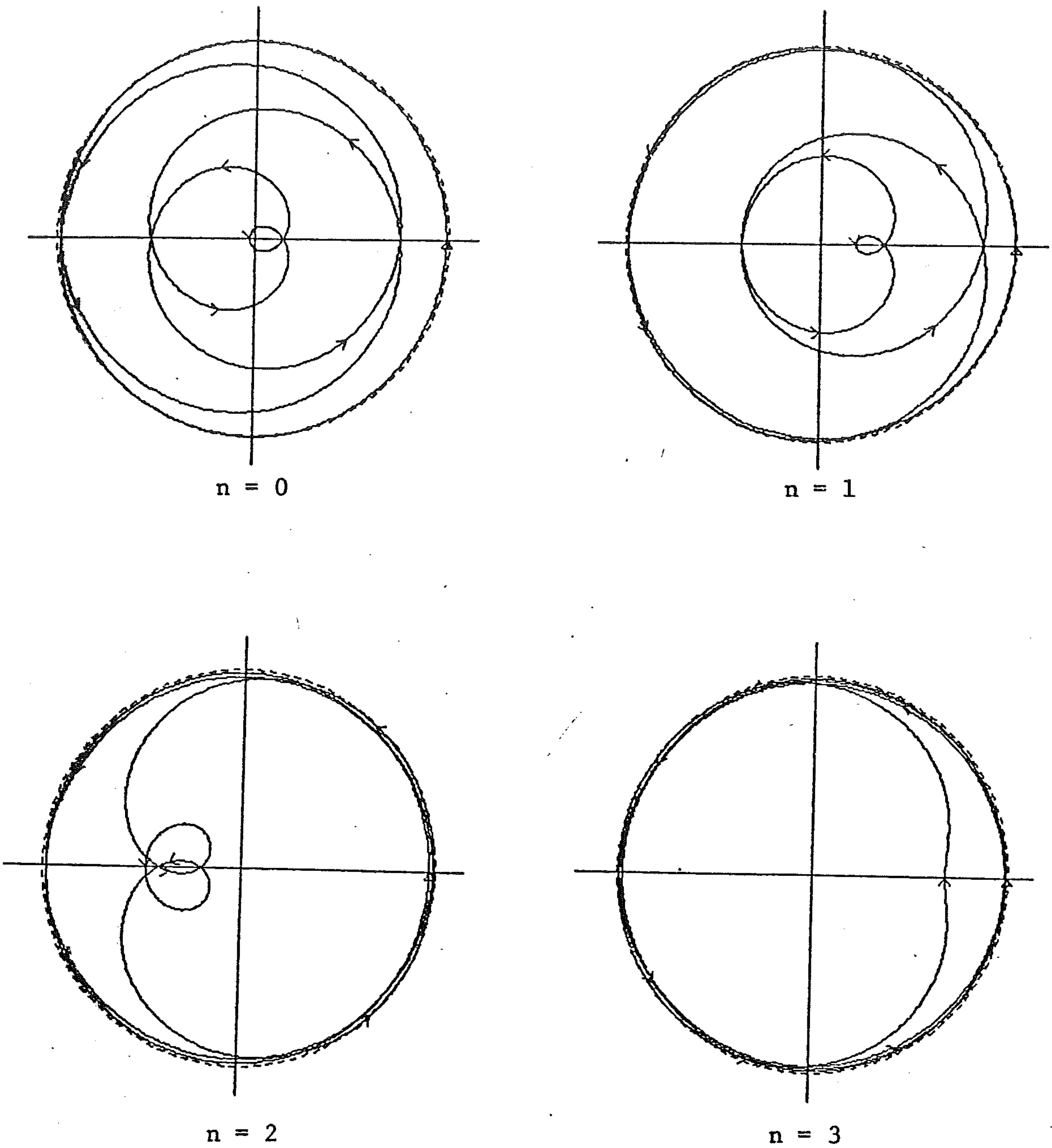
$$f(z) = \frac{1}{\Gamma(z)}$$

$$f(z) = \frac{(z-1)(z+1)(z-i)}{(z-i-1)^2}$$

All plots are drawn on normalized scales; the radius of each circle is equal to the maximum error E given below. The weighted least-squares error σ is also given, so that one may judge how nearly a given polynomial is a best approximation by considering the bound

$$\sigma \le \|e^*\| \le E$$

(see p. 25). Numerical coefficients are reported to five places, so that the polynomials might be used as given with the stated errors E. The true best approximation polynomial in each case, however, typically agrees with the near-best approximation reported to only two or three places in each coefficient.



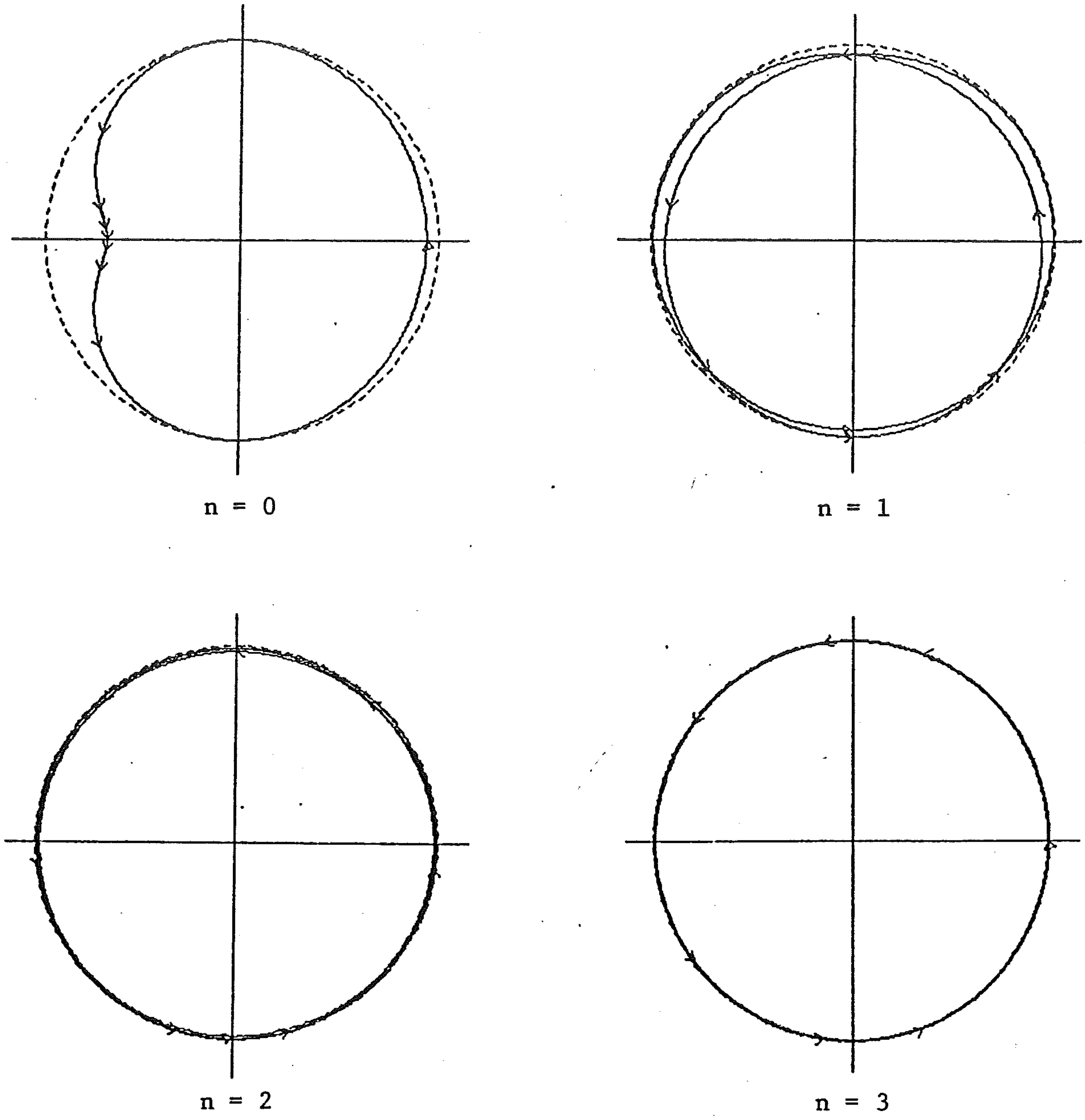
$f(z) = z^4 + z^5$

$p_0^*(z) = .05680$
$\sigma = 1.953 \qquad E = 1.975$

$p_1^*(z) = -.13690 + .21342z$
$\sigma = 1.910 \qquad E = 1.933$

$p_2^*(z) = .14216 - .35904z + .41109z^2$
$\sigma = 1.823 \qquad E = 1.863$

$p_3^*(z) = -.05973 + .15834z - .32007z^2 + .58212z^3$
$\sigma = 1.629 \qquad E = 1.662$



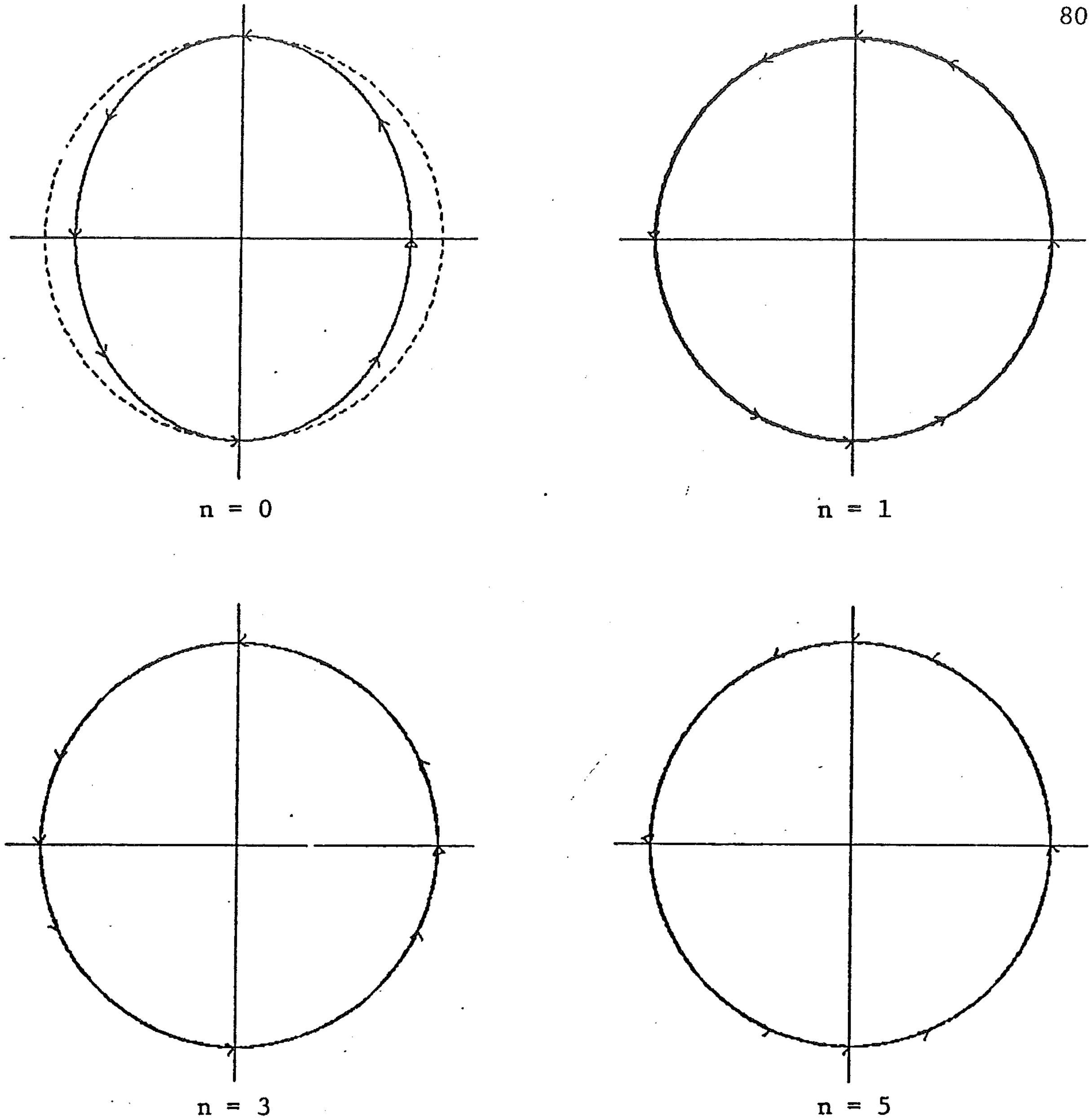
$$f(z) = e^{e^z} = 2.71828 + 2.71828z + 2.71828z^2 + 2.26523z^3 + 1.69893z^4 + \ldots$$

$p_0^*(z) = 7.23557$
$\sigma = 8.293 \qquad E = 8.438$

$p_1^*(z) = 2.15636 + 6.96589z$
$\sigma = 6.149 \qquad E = 6.254$

$p_2^*(z) = 2.59094 + 2.51918z + 5.68210z^2$
$\sigma = 4.387 \qquad E = 4.435$

$p_3^*(z) = 2.69297 + 2.67192z + 2.61649z^2 + 4.19081z^3$
$\sigma = 2.987 \qquad E = 2.998$



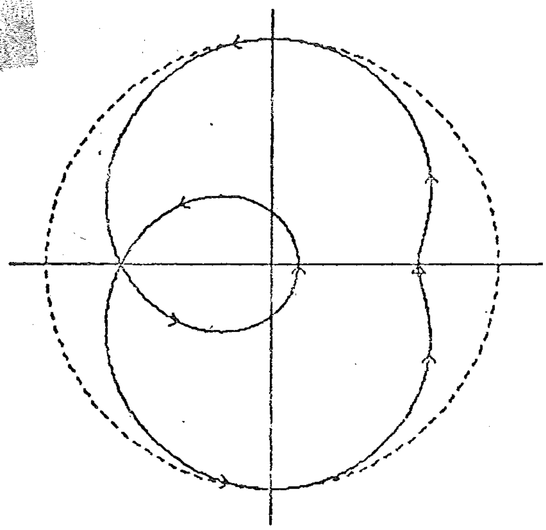
$$f(z) = \tan^{-1}\left(\frac{z}{2}\right) = .5z - .041667z^3 + .00625z^5 - .00112z^7 + \dots$$

$p_0^*(z) = 0.00000$
$\sigma = .537 \qquad E = .549$

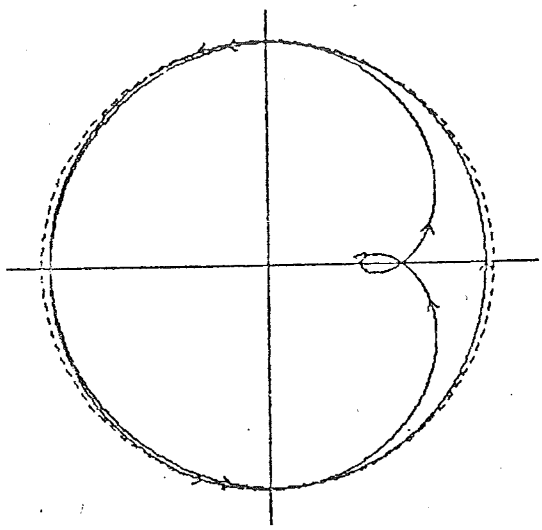
$p_1^*(z) = .50648z$
$\sigma = .04265 \qquad E = .04283$

$p_3^*(z) = .50002z - .04283z^3$
$\sigma = .006458 \qquad E = .006461$

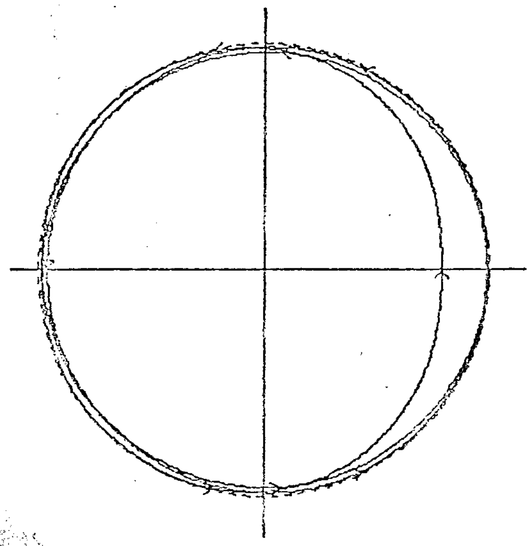
$p_5^*(z) = .50000z - .04167z^3 + .00648z^5$
$\sigma = .0011602 \qquad E = .0011603$



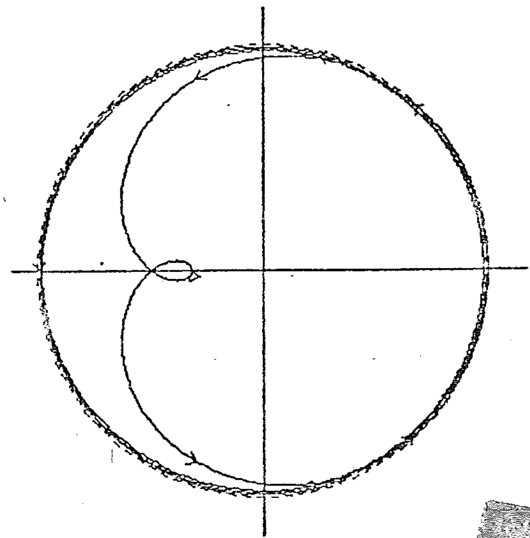
n = 0



n = 1



n = 2



n = 3

$$f(z) = \frac{1}{\Gamma(z)} = z + .57722z^2 - .65588z^3 - .04200z^4 + .16654z^5 + \ldots$$

$p_0^*(z) = -.22212$
$\sigma = 1.806 \qquad E = 1.877$

$p_1^*(z) = -.24585 + .79844z$
$\sigma = 1.064 \qquad E = 1.083$

$p_2^*(z) = -.03034 + 1.13510z + .59084z^2$
$\sigma = .713 \qquad E = .725$

$p_3^*(z) = .00939 + 1.02653z + .63195z^2 - .59962z^3$
$\sigma = .213 \qquad E = .217$

BIBLIOGRAPHY

[1] L.V. AHLFORS, Conformal Invariants: Topics in Geometric Function Theory, McGraw-Hill, 1973.

[2] S. JA. AL'PER, "Asymptotic values of best approximation of analytic functions in a complex domain," Uspehi Mat. Nauk 14, 1959, pp. 131-134. (Russian)

[3] S.N. BERNSTEIN, "Sur l'ordre de la meilleure approximation des fonctions continues par les polynômes de degré donné," Mem. Acad. Royale Belg. 4, 1912, pp. 1-104.

[4] E. BOREL, Leçons sur les Fonctions de Variables Réelles, Gauthier-Villars, Paris, 1905.

[5] C. CARATHÉODORY, "Über den Variabilitätsbereich der Fourier'schen Konstanten von positiven harmonischen Funktionen," Rend. Circ. Mat. Palermo 32, 1911, pp. 193-217.

[6] L. CARLESON, "Mergelyan's theorem on uniform polynomial approximation," Math. Scand. 15, 1964, pp. 167-175.

[7] P.L. CHEBYSHEV, "Sur les questions de minima qui se rattachent à la représentation approximative des fonctions," Oeuvres I, 1859, pp. 273-378.

[8] E.W. CHENEY, Introduction to Approximation Theory, McGraw-Hill, 1966.

[9] E.W. CHENEY & T.H. SOUTHARD, "A survey of methods for rational approximation, with particular reference to a new method based on a formula of Darboux," SIAM Review 5, 1963, pp. 219-231.

[10] W.J. CODY, "A survey of practical rational and polynomial approximation of functions," SIAM Review 12, 1970, pp. 400-423.

[11] P.J. DAVIS, Interpolation and Approximation, Dover, New York, 1975.

[12] P.J. DAVIS, private communication, 1977.

[13] S. ELLACOTT, private communication, 1977.

[14] S. ELLACOTT & J. WILLIAMS, "Linear Chebyshev approximation in the complex plane using Lawson's algorithm," Math. Comp. 30, no. 133, 1976, pp. 35-44.

[15] S. ELLACOTT & J. WILLIAMS, "Rational Chebyshev approximation in the complex plane," SIAM J. Numer. Anal. 13, 1976, pp. 310-323.

[16] G. FORSYTHE & C.B. MOLER, Computer Solution of Linear Algebraic Systems, Prentice-Hall, Englewood Cliffs, N.J., 1967.

[17] G.M. GOLUZIN, Geometric Theory of Functions of a Complex Variable, AMS Trans. v. 26, Providence, 1969.

[18] S. GRABINER, "A short proof of Runge's theorem," Am. Math. Monthly 83, 1976, pp. 807-808.

[19] E. HILLE, Analytic Function Theory, 2 vols., Chelsea, 2nd ed., 1973.

[20] D. JACKSON, The Theory of Approximation, AMS Colloq. Publ. v. XI, New York, 1930.

[21] V. KLOTZ, "Gewisse rationale Tschebyscheff-Approximationen in der komplexen Ebene," J. Approx. Th. 19, 1977, pp. 51-60.

[22] A.N. KOLMOGOROV, "A remark on the polynomials of P.L. Chebyshev deviating the least from a given function," Uspehi Mat. Nauk 3, 1948, pp. 216-221. (Russian)

[23] C.L. LAWSON, Contributions to the Theory of Linear Least Maximum Approximations, Thesis, UCLA, 1961.

[24] C.M. LEE & F.D.K ROBERTS, "A comparison of algorithms for rational ℓ_∞ approximation," Math. Comp. 27, no 121, 1973, pp. 111-121.

[25] G.G. LORENTZ, Approximation of Functions, Holt, Rinehart & Winston, 1966.

[26] G. MEINARDUS, Approximation of Functions: Theory and Numerical Methods, Springer Tracts in Natural Philosophy, v. 13, Springer-Verlag, 1967.

[27] S.N. MERGELYAN, "Uniform approximations to functions of a complex variable," AMS Trans. no. 101, Providence, R.I., 1954.

[28] S.J. POREDA, "Estimates for best approximation to rational functions," Trans. AMS 159, 1971, pp. 129-135.

[29] E.Y. REMES, "Sur le calcul effectif des polynômes d'approximation de Tschebyscheff," C.R. Acad. Sci. Paris 199, 1934, pp. 337-340.

[30] E.Y. REMES, "On approximations in the complex domain," Dokl. Akad. Nauk SSSR 77, no. 6, 1951, pp. 965-968. (Russian)

[31] J.R. RICE, The Approximation of Functions v. 1: Linear Theory, Addison-Wesley, 1964.

[32] J.R. RICE, The Approximation of Functions v. 2: Nonlinear and Multivariate Theory, Addison-Wesley, 1969.

[33] J.R. RICE, private communication, 1977.

[34] J.R. RICE & K.H. USOW, "The Lawson algorithm and extensions," Math. Comp. 22, 1968, pp. 118-127.

[35] T.J. RIVLIN, "Some explicit polynomial approximations in the complex domain," Bull. AMS 73, 1967, pp. 467-469.

[36] T.J. RIVLIN & H.S. SHAPIRO, "A unified approach to certain problems of approximation and minimization," J. SIAM 9, 1961, pp. 670-699.

[37] T.J. RIVLIN & B. WEISS, "Some best polynomial approximations in the plane," Duke. Math. J. 35, 1968, pp. 475-482.

[38] C. RUNGE, "Zur Theorie der eindeutigen analytischen Funktionen," Acta Math. 6, 1885, pp. 229-244.

[39] E.B. SAFF, "On the zeros of the error function for Tchebycheff approximation on a disk," J. Approx. Th. 9, 1973, pp. 112-117.

[40] E.B. SAFF & R.S. VARGA, "Nonuniqueness of best complex rational approximations to real functions on real intervals," to appear in J. Approx. Th.

[41] S. SAKS & A. ZYGMUND, Analytic Functions, Monografje Mat., v. 28, Warsaw, 1952.

[42] G. SANSONE & J. GERRETSEN, Lectures on the Theory of Functions of a Complex Variable v. 2: Geometric Theory, Wolters-Noordhoff, Groningen, Netherlands, 1969.

[43] W.E. SEWELL, Degree of Approximation by Polynomials in the Complex Domain, Princeton University Press, Princeton, 1942.

[44] V.I. SMIRNOV & N.A. LEBEDEV, Functions of a Complex Variable: Constructive Theory, M.I.T. Press, Cambridge, 1968.

[45] E.C. TITCHMARSH, The Theory of Functions, 2nd ed., Oxford U. Press, 1939.

[46] L. TONELLI, "I polinomi d'approssimazione di Tchebychev," Annali di Matematica (3) 15, 1908, pp. 47-119.

[47] R. VARGA, private communication, 1977.

[48] J.L. WALSH, "The existence of rational functions of best approximation," Trans. AMS 33, 1931, pp. 668-689.

[49] J.L. WALSH, Approximation by Polynomials in the Complex Domain, Mémorial des Sciences Mathématiques, Paris, 1935.

[50] J.L. WALSH, Interpolation and Approximation by Rational Functions in the Complex Domain, 2nd ed., AMS Colloq. Publ. v. 20, Providence, R.I., 1956.

[51] J.L. WALSH, "Padé approximants as limits of rational functions of best approximation," J. Math. Mech. 13, 1964, pp. 305-312.

[52] K. WEIERSTRASS, "Über die analytische Darstellbarkeit sogenannter willkürlicher Funktionen einer reelen Veränderlichen," Sitzungsberichte der Akademie zu Berlin, 1885, pp. 633-639, 789-805.

[53] J.H. WILKINSON & C. REINSCH, Handbook for Automatic Computation v. II: Linear Algebra, Springer-Verlag, 1971.

[54] J. WILLIAMS, private communication, 1977.

INDEX