METHODS OF SEARCH FOR SOLVING POLYNOMIAL EQUATIONS

BY

PETER HENRICI

TECHNICAL REPORT NO. CS 145
DECEMBER 1969

COMPUTER SCIENCE DEPARTMENT
School of Humanities and Sciences
STANFORD UNIVERSITY

METHODS OF SEARCH FOR SOLVING POLYNOMIAL EQUATIONS

By

Peter Henrici

. December 1969

METHODS OF SEARCH FOR SOLVING POLYNOMIAL EQUATIONS

By Peter Henrici*

Eidgenössische Technische Hochschule

Zürich, Switzerland

Dedicated to D. H. Lehmer on his 65th birthday

Abstract

The problem of determining a zero of a given polynomial with guaranteed error bounds, using an amount of work that can be estimated a priori, is attacked hereby means of a class of algorithms based on the idea of systematic search. Lehmer's "machine method" for solving polynomial equations is a special case. The use of the Schur-Cohn algorithm in Lehmer's method is replaced by a more general proximity test which reacts positively if applied at a point close to a zero of a polynomial. Various such tests are described, and the work involved in their use is estimated. The optimality and non-optimality of certain methods, both on a deterministic and on a probabilistic basis, are established.

Key words

polynomials, zeros, proximity test, covering, search algorithm, work function, optimal search, optimal covering, Schur-Cohn algorithm, convergence function, linear convergence.

---

*This research was partially supported by the National Science Foundation, under Grant No. GP 7657 at the Computer Science Department, Stanford University, and by the Office of Naval Research under project NR 044-211.

1. Introduction

In 1961 D. H. Lehmer [6] proposed a "machine method" for solving polynomial equations. His algorithm was guaranteed to approximate a zero of any given complex polynomial with an arbitrarily small error. The amount of work necessary to compute a zero to a given precision could be estimated a priori.

In the present paper we shall describe a class of algorithms for polynomial zerofinding which contains Lehmer's method as a special case. Our algorithms borrow from Lehmer's method the basic idea of enclosing zeros in disks of decreasing radius, and of covering disks containing a zero by smaller disks,. However, instead of using a special procedure to determine whether or not a given disk contains a zero of a polynomial, the algorithms discussed here merely require a "proximity test" (§2) which reacts positively if applied at a point close to a zero of the given polynomial. Very simple such proximity tests exist, and as a consequence some of our algorithms are arithmetically simpler than Lehmer's method (§3).

The convergence of the general search algorithm is established (§4), and the maximum amount of work necessary to determine a zero to a preassigned accuracy is estimated (§5).

Among the class of all proximity tests, we then identify a subclass for which the convergence of the resulting algorithms is linear. Among these tests, the classical Schur-Cohn test (which forms the basis for Lehmer's method) is shown to enjoy a certain property of optimality (§6). We finally discuss the best covering strategy if coverings by disks of constant radius, are used. From a deterministic point of view, the best strategy consists in covering a disk of radius r by eight disks of radius $q_0 r$, where

$q_0 = (1 + 2 \cos 2\pi/7)^{-1} \doteq 0.44504$. From a probabilistic point of view, if coverings by disks of variable radius are permitted, Lehmer's original covering is slightly better, although not optimal.

Besides Lehmer's paper, the present study was inspired by the methods of search used in the constructive proofs of the fundamental theorem of algebra due to Brouwer [3, 4] and Rosenbloom [10].

2. Proximity tests

For positive integers N , let P_N denote the class of all monic polynomials of degree N with complex coefficients,

$$p(z) = z^N + a_{N-1}z^{N-1} + \cdots + a_0 ,$$

whose zeros ζ_1 , ζ_2 , ... , ζ_N satisfy $|\zeta_i| \leq 1$, $i = 1, 2, \ldots, N$. It is our objective to study a class of algorithms for solving the following problem: Given any $p \in P_N$ and any $\varepsilon > 0$, to construct a disk D of radius ε which contains a zero of p . The algorithms to be discussed are uniformly convergent on P_N , in the following sense: The amount of work necessary to construct D is bounded by a quantity which depends on ε and N , but not on the individual polynomial p .

The basic tool of the algorithms to be described is a proximity test $T = T(r)$, which can be applied to any polynomial $p \in P_N$ at any point z such that $|z| \leq 1$, and which the polynomial either passes or fails. The test must be such that it is passed at all points z sufficiently close to a zero, and failed at all points sufficiently far away. (There may be an in-between region where the test may be passed or failed.) The parameter r regulates the difficulty of the test. The smaller r is, the more difficult it becomes to pass the test.

Speaking formally, a test T(r) is called a proximity test if there exist two positive functions ϕ and ψ, defined on some interval $0 < r \leq r_0$ and having the following properties:. If p is any polynomial in P_N, and if ζ is any zero of p, then for all $r \in (0, r_0]$

(i) p passes T(r) at all points z such that $|z| \leq 1$ and

$$|z - \zeta| \leq \phi(r) \ ;$$

(ii) p fails T(r) at all points z such that $|z| \leq 1$ and

$$|z - \zeta| > \psi(r) \ .$$

The above evidently implies that $\phi(r) \leq \psi(r)$; we do not require that $\phi = \psi$, We postulate that T(r) becomes arbitrarily difficult to pass for $r \to 0$, i.e.,

(iii) $$\lim_{r \to 0} \psi(r) = 0 \ .$$

We furthermore require

(iv) ψ is continuous and strictly monotonically increasing.

The functions ϕ and ψ are called, respectively, the inner and outer convergence function of the test T(r)'.

The following test, to be denoted by T_1, may serve as a first example of a proximity test:

" p passes $T_1(r)$ at z " $\Longleftrightarrow$ $|p(z)| \leq r$.

To show that this test has the required properties for $0 < r < 1$, let

$$p(z) = \prod_{i=1}^{N} (z - \zeta_i).$$

If p fails the test at z, then

$$|p(z)| = \prod_{i=1}^{N} |z - \zeta_i| > r .$$

Hence for every i,

$$|z - \zeta_i| > r \prod_{\substack{j=1 \\ j \neq i}}^{N} |z - \zeta_j|^{-1}.$$

Since $|\zeta_j| \leq 1$, $|z| \leq 1$, every factor of the product on the right is at least $1/2$, and we find that

$$|z - \zeta_i| < 2^{-N+1} r , \quad i = 1, \ldots, N .$$

Hence $T_1(r)$ cannot be failed if $|z - \zeta_i| \leq 2^{-N+1} r$ for some i, and (i) is true for

$$\phi(r) = 2^{-N+1} r .$$

If, on the other hand, p passes $T_1(r)$ at z, then

$$\prod_{i=1}^{N} |z - \zeta_i| \leq r ,$$

and it follows that

$$|z - \zeta_i| \leq r^{1/N}$$

for at least one index i . . Thus the test cannot be passed if $|z - \zeta_i| > r^{1/N}$ for all i , and we find that (ii) is true for

$$\psi(r) = r^{1/N} .$$

(By considering a polynomial with a single zero of multiplicity N , we see that (ii) is not true for any smaller function ψ .) It is clear that ψ has the properties (iii) and (iv).

Two tests are called equivalent if they are defined on the same domain of r and if they produce identical results for all polynomials p at all points z and for all values r .

Example: The test T_1 is equivalent to a test which is declared passed if and only if $|p(z)|^2 \leq r^2$.

Two proximity tests T and T* are called similar if there exists an increasing function r* mapping $[0,r_0]$ onto an interval $[0,r_0^*]$ such that the test T(r) is equivalent to $T^*(r) = T(r^*(r))$. Similar tests thus differ only in the choice of the parameter. It is clear that the similarity of tests, too, is an equivalence relation.

Example: The test T_1 is similar to the test $T_1^*(r)$ which is passed if and only if $|p(z)| \leq r^N$. Convergence functions for T_1^* are $\phi(r) = 2^{-N+1}r^N$ and $\psi(r) = r$.

By (iv), every proximity test is similar to a test with outer convergence function $\psi(r) = r$.

3. The search algorithm

We require the notion of an s-covering. If ε is any positive number, and if S is any set in the complex plane, an e-covering of S is any system of closed disks of radius $\leq \varepsilon$ whose union contains S . The covering is said to be centered in S if the midpoints of the covering disks belong to s . The construction of a minimal s-covering of a given bounded set (i.e., a covering containing the least number of disks) can raise intricate questions of elementary geometry. Of course, one can always use coverings whose centers form a square or hexagonal grid.

Let $p \varepsilon P_N$, let T be a proximity test, and let $\{q_k\}$ be a monotonic sequence of positive numbers converging to zero such that $q_0 = 1$. We shall describe an algorithm for constructing a sequence of points $\{z_k\}$ such that each of the disks

$$D_k = \{z: \quad |z - z_k| \leq q_k\} ,$$

$k = 0 , \quad 1, \quad 2, \quad \ldots , \quad$ contains at least one zero of p .

Let $z_0 = 0$. Then D_0 certainly contains a zero, for it contains all zeros. The algorithm now proceeds by induction. Suppose we have found a point z_{k-1} such that D_{k-1} contains a zero. To construct z_k , we cover the set $D_{k-1} \cap D_0$ with an ε_k-covering centered in it and apply a test $T(r_k)$ at the center of each covering disk. The parameters ε_k and r_k are chosen such that the following two conditions are met:

(A) The test is passed at the center of each disk of the covering which contains a zero.

(B) Any point at which the test is passed is at a distance $\leq q_k$ from a zero.

Condition (A) is satisfied if $\varepsilon_k \leq \phi(r_k)$. Condition (B) is satisfied if $\psi(r_k) \leq q_k$. Thus both conditions are fulfilled if

$$r_k = \psi^{-1}(q_k) ,$$

(1)

$$\varepsilon_k = \phi(r_k) = \phi(\psi^{-1}(r_k)) ,$$

where ψ^{-1} denotes the inverse function of ψ.

At least one of the covering disks contains a zero, since D_{k-1} contains one, and since all disks are contained in D. Thus by (A), the test $T(r_k)$ is passed at least once. We let z_k be the first center at which the test is passed. There is no assurance that the disk of radius ε_k surrounding z_k actually contains a zero, but by (B), the disk D_k does.

The whole algorithm thus may be summarized as follows: Let $z_0 = 0$. Having constructed z_{k-1}, cover the set $D_{k-1} \cap D_0$ by an ε_k-covering centered in it, and apply $T(r_k)$ at the center of each covering disk, where, ε_k and r_k are given by (1). Let z_k be the first center which passes the test.

Provided that identical systems of converings are used, the above algorithm remains unchanged if the test T is replaced by a "similar" test T* .

4. Convergence

By construction, the centers z_k of successive disks D_k satisfy $|z_{k+1} - z_k| \leq q_k$, where $q_k \to 0$. This in itself does not imply the convergence of the sequence $\{z_k\}$. Nevertheless, there holds

THEOREM 1. The sequence $\{z_k\}$ converges, and its limit is a zero of p .

Proof. Let

$$\delta = \min_{\zeta_i \neq \zeta_j} |\zeta_i - \zeta_j|$$

be the minimum distance between distinct zeros of p . Let m be an integer such that $2q_m < \delta$. Let $n \geq m$. The disk Dk contains a zero, say ζ_i . The disk D_{k+1} likewise contains a zero, say ζ_j . From

$$|z_n - \zeta_i| \leq q_n , \quad |z_{n+1} - \zeta_j| \leq q_{n+1} ,$$

it follows by the monotonicity of the sequence $\{q_n\}$ that

$$|\zeta_i - \zeta_j| \leq q_n + q_{n+1} \leq 2q_n < \delta$$

and hence that $\zeta_i = \zeta_j$. Thus for all $n = m$, $|z_n - \zeta_i| \leq q_n$, proving that

$$\lim_{n \to \infty} z_n = \zeta_i .$$

5. Amount of work

We measure the amount of work required to approximate a zero with an error $\leq \varepsilon$ by estimating the number of applications of the test T required to construct the first disk D_k such that its radius q_k is less than ε . For reasons of simplicity we assume until further notice that the centers of the covering disks always form a square grid.

The area of D_{m-1} is πq_{m-1}^2. In a square k-covering, the centers of the covering disks must be not more than $\sqrt{2}\,\varepsilon_m$ apart. Neglecting boundary effects, approximately

$$\frac{\pi}{2}\,\frac{q_{m-1}^2}{\varepsilon_m^2}$$

disks of radius ε_m are thus required to cover D_{m-1}. (Working with a hexagonal grid, the constant $\frac{\pi}{2}$ could be replaced by $\frac{2\pi}{3\sqrt{3}}$.) Within the same degree of approximation, this also is the maximum number of applications of the test to proceed from z_{m-1} to z_m.

For the given sequence $\{q_k\}$ and for $\varepsilon > 0$, let k(e) denote the smallest k such that $q_k \leqq \varepsilon$. By the above, the total number of applications of the test necessary to approximate a zero with an error $\leqq \varepsilon$ does not exceed a quantity of the order of

$$w(T, \{q_k\}, \varepsilon) = \frac{\pi}{2} \sum_{m=1}^{k(\varepsilon)} \frac{q_{m-1}^2}{\varepsilon_m^2} \quad . \tag{2}$$

We axiomatically define the above function w as the work function of the search algorithm based on the proximity test T and the sequence $\{q_k\}$. The work function does not change if the test T is replaced by a similar test T* .

From the fact that w does not depend on p it already follows that the search algorithms described earlier are uniformly convergent in the sense described earlier.

Example. For the test T_1, choosing a geometric mode of subdivision ($q_k = q^k$, $0 < q < 1$, $k = 0, 1, 2, \ldots$) we have in view of $\phi(r) = 2^{-N+1}r$, $\psi(r) = r^{1/N}$

$$\varepsilon_m = \phi(\psi^{-1}(q_m)) = 2^{-N+1}q^{mN},$$

hence

$$w(T_1, \{q^k\}, \varepsilon) = \frac{\pi}{2} \; 2^{2N-2\sum_{m=1}^{k(\varepsilon)}} q^{2m-2-2mN} \sim C_N \, q^{-(2N-2)k(\varepsilon)}$$

$(\varepsilon \to 0)$, where

$$C_N = \frac{\pi}{2} \frac{2^{2N-2}}{q^2 - q^{2N}} \quad (N \geq 2).$$

For the determination of a zero of a polynomial of degree 10 with an error $\leq 10^{-6}$ working with $q = \frac{1}{2}$ (which requires $k = 20$) the function w yields an upper bound of approximately $2^{397}\pi \doteq 10^{120}$ applications of the test. Since on the average we can't expect to do much better than use one half of the maximum number of tests, a search algorithm based on T_1 certainly is not practical.

6: Proximity tests with linear convergence functions

Suppose the convergence functions of a proximity test T are linear,

(3) $$\phi(r) = ar, \quad \psi(r) = br$$

$(0 < a \leq b)$. Then by (1),

$$\varepsilon_m = \phi(\psi^{-1}(q_m)) = \frac{a}{b} q_m ,$$

and the work function (2) becomes

$$w(T, \{q_k\}, \varepsilon) = \frac{\pi}{2} \frac{b^2}{a^2} \sum_{m=1}^{k(\varepsilon)} \frac{q_{m-1}^2}{q_m^2} . \tag{4}$$

In particular, if $q_k = q^k$,

$$w(T, \{q^k\}, \varepsilon) = \frac{\pi b^2}{2a^2 q^2} k(\varepsilon) , \tag{5}$$

and the work necessary to compute a zero to a given accuracy is proportional to the number of decimals required. This convergence behavior is known as linear convergence.

We now shall give some examples of proximity tests with linear convergence functions. For arbitrary z and h , let

$$p(z + h) = b_0 + b_1 h + b_2 h^2 + \ldots + b_N h^N$$

($b_N = 1$) . It will be convenient to suppress the argument z in the Taylor coefficients b_i .

6.1. The test T_2 . Let

$$B = B(z) = \min_{1 \leq k \leq N} \left| \frac{b_0}{b_k} \right|^{1/k} .$$

The polynomial p is said to pass the test $T_2(r)$ at z if and only if $B(z) \geq r$. To determine the convergence functions of this test, let

(6)

The relations of Vieta imply, as is well known,

$$\rho \leq \left| \binom{N}{k} \frac{b_0}{b_k} \right|^{1/k} , \quad k = 1 , \ldots , N .$$

Since $\binom{N}{k}^{1/k} \leq N$, this implies $\rho \leq NB(z)$. Hence if $\rho > Nr$, then $B(z) > r$, and p fails $T_2(r)$ at z . It follows that

$$\psi(r) = Nr$$

is outer convergence function for T_2 . On the other hand, let p fail the test at z . Then $B > r$ and hence

$$\left| \frac{b_k}{b_0} \right| < r^{-k} , \quad k = 1 , 2 , \ldots , N .$$

If $p(z + h) = 0$ and $|h| = \rho$, the Taylor expansion shows that

$$\frac{\rho}{r} + \frac{\rho^2}{r^2} + \ldots + \frac{\rho^N}{r^N} \geq 1$$

and hence that $\frac{\rho}{r} > \frac{1}{2}$. It follows that the test cannot be failed if $\rho \leq \frac{1}{2} r$, i.e.,

$$\phi(r) = \tfrac{1}{2} r$$

is inner. convergence function for T_2 .

Thus T_2 has convergence functions of the form (3); we note that $\frac{b}{a} = 2N$. In the numerical example considered earlier ($N = 10$, $\varepsilon = 10^{-6}$, $q_k = 2^{-k}$) , (4) now furnishes an upper bound of some 50,000 applications of the test,

6.2. <u>The test</u> T_3 . The polynomial is said to pass $T_3(r)$ at z if and only if

$$|b_0| \leq |b_1|r + |b_2|r^2 + \ldots + |b_N|r^N .$$

Let ρ be defined by (6). Then for some h such that $|h| = \rho$ we have $p(z + h) = 0$, hence

$$|b_0| \leq |b_1|\rho + |b_2|\rho^2 + \ldots + |b_N|\rho^N ,$$

and p passes $T_3(\rho)$. Thus $(b(r) = r$ is inner convergence function for this test. On the other hand, a theorem of G. D. Birkhoff [2] implies that the test cannot be passed if $\rho > (2^{1/N} - 1)^{-1}r$. Thus

$$\psi(r) = \frac{1}{2^{1/N}-1}\, r$$

is outer convergence function. For this pair of convergence functions,

$$\frac{b}{a} = \frac{1}{2^{1/N}-1} \sim \frac{N}{\log 2} \quad (N \to \infty) .$$

For a given sequence $\{q_k\}$, and for linear convergence functions (3), the value of the work function for a given ε 'is proportional to b^2/a^2 .

For both tests T_2 and T_3 this ratio is $O(N^2)$ as $N \to \infty$. This situation is typical for any test that depends only on the absolute values $|b_i|$, for it is known [9, 1] that the maximum of the ratio of the largest and smallest absolute value which the smallest zero of a polynomial of degree N can have if the absolute values of the coefficients are fixed is precisely $(2^{1/N} - 1)^{-1}$. It follows that smaller values of b/a can be achieved only with tests that do not merely use the absolute values of the Taylor coefficients.

6.3. The test T_4. This test makes use of the sums

$$s_k = \sum_{i=1}^{N} (\zeta_i - z)^{-k}, \quad k = 1, 2, \ldots . \tag{7}$$

It is easily shown by means of a generating function argument that these quantities can be computed from the Taylor coefficients at z by means of the following recurrence relation:

$$s_k = -b_0^{-1}(kb_k + s_1 b_{k-1} + s_2 b_{k-2} + \cdots + s_{k-1} b_1),$$

$$k = 1, 2, \ldots .$$

Let ρ bedefinedby(6). Then $|s_k| \leqq N\rho^{-k}$, $k = 1, 2, \ldots$, and it follows that

$$\rho \leqq \left| \frac{N}{s_k} \right|^{1/k}, \quad k = 1, 2, \ldots . \tag{8}$$

Let

$$s = \min_{1 \leq k \leq N} \left| \frac{N}{s_k} \right|^{1/k} .$$

We say that p passes the test $T_4(r)$ at z if and only if $S \leq r$. It follows from (8) that

$$\psi(r) = r$$

is outer convergence function for this test. Moreover, a rather deep result of Buckholtz [5] states that $S < (2 + 2\sqrt{2})\rho$, where the numerical constant is best possible. It follows that

$$\phi(r) = (2 + 2\sqrt{2})^{-1} r$$

is inner convergence function. For this pair of convergence functions, the ratio $b/a = 2 + 2\sqrt{2} \doteq 4.8284$ is independent of N .

6.4. Sharp tests. For a given sequence $\{q_k\}$, and for linear convergence functions ϕ and ψ , the value of the work function (4) for given ε is a minimum for a test such that $b = a$. Without loss of generality it may be assumed that $b = a = 1$. A test with convergence functions $\phi(r) = \psi(r) = r$ will be called sharp. A sharp test reacts positively if and only if the closed disk of radius r about the testing point z contains a zero. Thus all sharp tests belong to the same class of equivalent tests.

There exist several realizations of sharp tests. They are based either on a conformal mapping of the disk onto the left half-plane, followed by the Routh-Hurwitz algorithm, or (more directly and efficiently) on the well-known Schur-Cohn algorithm ([8], p. 195) for counting the number of zeros in a given disk. Lehmer's method [6, 7], the first search algorithm of the type considered here, was based on the Schur-Cohn algorithm.

In our numerical example ($N = 10$, $q_k = 2^{-k}$, $\varepsilon = 10^{-6}$), (5) now yields a maximum of a mere 129 tests in an algorithm based on a sharp test. Due to neglect of boundary effects, the true maximum is somewhat higher; see below.

The mere fact that the work function is smallest for the Schur-Cohn test does not in itself imply that this test defines the computationally most efficient algorithm, since the work function does not take into account the work required to carry out the test. In the absence of rigorous results concerning the minimum number of arithmetic operations required to administer the various tests, precise results are difficult. Suffice it to say that all tests described in this section require, among other things, all Taylor coefficients at z. If performed by the Horner algorithm, their computation requires $\frac{1}{2}N^2 + O(N)$ multiplications. The Schur-Cohn algorithm, if programmed in the superior fashion recommended by Stewart [11], requires another $\frac{1}{2}N^2 + O(N)$ multiplications and divisions, roughly the same as the computation of the sums s_k required for T_4. Thus the Schur-Cohn test requires only about twice as much work as T_2 or T_3, and about the same as T_4.

7. Optimum choice of $\{q_k\}$

Suppose the search algorithm is based on a test with linear convergence functions (3). If ε is given, for what choice of the sequence $\{q_k\}$ is the work function $w(T, \{q_k\}, \varepsilon)$ a minimum?

We first answer this question when k(e) is prescribed. Let $\varepsilon > 0$, Let k be a given positive integer, and let $\{q_m\}$ be any decreasing sequence such that $q_0 = 1$, $q_k = \varepsilon$. Then, by the inequality of the arithmetic and geometric mean,

$$w(T, \{q_m\}, \varepsilon) = C \sum_{m=1}^{k} \frac{q_{m-1}^2}{q_m^2} \qquad \left(C = \frac{mb^2}{2a^2}\right)$$

$$\geq Ck \left| \prod_{m=1}^{k} \frac{q_{m-1}^2}{q_m^2} \right|^{1/k}$$

$$= Ck\varepsilon^{-2/k}$$

$$= w(T, \{\varepsilon^{m/k}\}, \varepsilon) ,$$

and we have proved:

THEOREM 2. Let $\varepsilon > 0$ and k 3 0 be given. On the space of all monotonic sequences $\{q_m\}$ such that $q_0 = 1$ and $q_k = \varepsilon$, the work function (4) assumes its smallest value for the geometric sequence, $q_m = \varepsilon^{m/k}$, m = O , 1 , 2 , ...

On the basis of this result, we now restrict our attention to geometric sequences, $q_m = q^m$ $(0 < q < 1)$, and ask for the optimal value of q to achieve a given accuracy ε. As a function of q and ε, $k(\varepsilon)$ is now the smallest integer such that $q^k \leq \varepsilon$ or

$$k(\varepsilon) = - \left[- \frac{\log \varepsilon}{\log q} \right] ,$$

where $[x]$ denotes the largest integer $\leq x$. Neglecting a fractional part, we thus have approximately

$$w(T,\{q^k\},\varepsilon) \doteq C\frac{\log \varepsilon}{q^2 \log q}$$

(C defined as above). By differentiation we easily find that the minimum of the above expression is attained for $q = e^{-1/2} \doteq 0.60653$, and that the value of the minimum is $2\, e\, C \log \frac{1}{\varepsilon}$.

Unfortunately, the above result does not indicate accurately the maximum number of tests to be applied, because the method of counting the covering disks underlying (2) becomes increasingly inaccurate (due to the neglect of boundary effects) if the ratio of the radii of the covering disks and of the disk to covered approaches 1. To determine the exact maximum, let, for $0 < x \leq 1$, $f(x)$ denote the minimum number of disks of radius x that are required to cover the unit disk. The function f is non-increasing, piecewise constant, and continuous from the right; no simple analytical expression for it exists. To proceed from z_m to z_{m+1} in a search algorithm based on a test with linear convergence functions and on a geometric sequence $\{q^m\}$ requires covering a disk of radius q^m by disks of radius $\frac{a}{b} q^{m+1}$. Hence, if an optimal covering is used, at most $f(\frac{a}{b} q)$ -applications of the test are necessary. The actual maximum number of tests to attain an error $\leq \varepsilon$ thus equals

$$W(a,b,q,\varepsilon) = f(\tfrac{a}{b}q)\left\lceil \frac{\log \varepsilon}{\log q} \right\rceil .$$

We shall determine the minimum of W as a function of q for the Schur-Cohn test (a=b=1) .

THEOREM 3. For sufficiently small fixed values of ε , the function $F(q,\varepsilon) = W(1,1,q,\varepsilon)$ assumes its minimum at $q = q_0 = (1 + 2\cos\frac{2\pi}{7})^{-1}$. The value of the minimum is

$$F(q_0,\varepsilon) = -8\left[-\frac{\log\varepsilon}{\log q_0}\right] \doteq -8\left[-\frac{\log\varepsilon}{0.8096}\right] .$$

Proof. We first determine the minimum of the function

$$G(q) = f(q)\frac{\log\varepsilon}{\log q} .$$

Let the points of discontinuity of f be, in decreasing order, $1 = x_0 > x_1 > x_2 > \ldots$, and let the constant value of f in the interval $x_m \leq x < x_{m-1}$ be denoted by f_m $(m = 1, 2, \ldots)$. Then G(q) is increasing in each of the intervals $x_m \leq q < x_{m-1}$, and has a downward jump at the points x_m $(m = 1, 2, \ldots)$. It thus is smallest where

$$G(x_m) = f_{m-1}\log x_m$$

is smallest. It can be shown that

$$x_m = (2\cos\tfrac{\pi}{m+2})^{-1} , \quad f_m = m + 2 \quad \text{for} \quad m = 1, 2, 3 ;$$

$$x_m = (1 + 2\cos\tfrac{2\pi}{m+2})^{-1} , \quad f_m = m + 3 \quad \text{for} \quad m = 4, 5, 6 .$$

From these values and from the trivial estimate $f(x) \geq x^{-2}$ it follows by computation that the minimum is assumed only at $q_0 = x_5 = (1 + 2 \cos \frac{2\pi}{7})^{-1} \doteq 0.44504$, and that it has the value

$$G(q_0) = 8 \frac{\log \varepsilon}{\log q_0} \doteq 9.882 \log \varepsilon^{-1} .$$

The function F has the form $F(q) = f(q)h(q)$, where

$$h(g) = - \left[- \frac{\log \varepsilon}{\log q} \right] .$$

The function h is piecewise constant, nondecreasing, and continuous from the left. We denote its points of discontinuity by $0 < h_0 < h_1 < h_2 < \ldots$. Evidently, $F(q) \geq G(q)$, with equality holding if and only if $q=h_n$ for some n. Let n* be the smallest index n such that $h_n \geq q_0$. For sufficiently small values of ε, the points h_n are arbitrarily dense, hence $h_{n*} < x_4$, and furthermore

$$F(h_{n*}) < G(x_m), \; m \neq 5 .$$

It follows that $F(h_{n*})$ is the smallest value of F. If $h_{n*} = q_0$, the Theorem is established. If $h_{n*} > q_0$, the Theorem follows from the fact that F(q) is constant for $q_0 \leq q \leq h_{n*}$.

The optimal covering of the unit disk by 8 disks of radius q_0 consists of a disk centered at the origin, surrounded by 7 disks centered at the points

$$z_k = R e^{\frac{2\pi i k}{7}}, \quad k = 0, 1, \ldots, 6,$$

where

$$R = \frac{2 \cos \frac{\pi}{7}}{1 + 2 \cos \frac{2\pi}{7}} \doteq 0.80194 \ .$$

8. Non-uniform converings

So far in this study, it was assumed that the covering of each disk D_k consists of disks of constant radius. It is a trivial matter to modify the definition of the basic search algorithm to permit coverings of variable radius and to extend the convergence theorem to this case. Also the upper bounds for the amount of work are easily adapted to extend to such non-uniform coverings.

However, the optimality considerations of section 7 strongly depend on the constancy of the radii of the covering disks, and it is far from obvious how they should be modified for non-uniform coverings. It appears certain, however, that the methods using uniform coverings are not optimal in the class of methods using arbitrary coverings.

The efficiency of an algorithm can also be judged from a probabilistic point of view, for instance by computing the average number Z of applications of the test required to improve the accuracy of a zero by one decimal digit. Here again the methods using uniform coverings are not optimal. For the optimal method using uniform coverings determined in Theorem 3, it can be shown that

$$Z \doteq 11.168 \ .$$

Lehmer's method covers the unit disk by a disk of radius $\frac{1}{2}$ centered at 0, and by 8 disks of radius $\frac{5}{12}$ centered on a circle of radius $\frac{5}{6}$. For this covering, if the sequence of surrounding disks is chosen optimally as suggested in [6],

$$z \doteq 11.143 .$$

It can be shown that Lehmer's coverings is again not optimal, if only by -- the trivial reason that it has some built-in slack to counteract rounding. The detailed investigation of optimal non-uniform coverings must, however, wait for another-paper.

REFERENCES

[1] Batschelet, E.: Untersuchugen über die absoluten Beträge der Wurzeln algebraischer, insbesondere kubischer Gleichungen. Verhandlungen der Naturforschenden Gesellschaft in Basel 55, pp. 158-179 (1944).

[2] Birkhoff, G. D.: An elementary double inequality for the roots of an algebraic equation having greatest absolute value. Bull. Amer. Math. Soc. 21, pp. 494-495 (1914).

[3] Brouwer, L. E. J., and B. de Loor: Intuitionistischer Beweis des Fundamentalsatzes der Algebra. Amsterdam Konigl. Akad. van Wetenschappen, Proc. 27, pp. 186-188 (1924).

[4] Brouwer, L. E. J.: Intuitionistische Ergänzung des Fundamentalsatzes der Algebra. Amsterdam Konigl. Akad. van Wetenschappen, Proc. 27, pp. 631-634 (1924).

[5] Buckholtz, J. D.: Sums of powers of complex numbers. J. Math. Anal. Appl. 17, pp. 269-279 (1967).

[6] Lehmer, D. H.: A machine method for solving polynomial equations. J. Assoc. Comp. Mach. 8, pp. 151-162 (1961).

[7] Lehmer, D. H.: Search procedures for polynomial equation solving. Constructive aspects of the fundamental theorem of algebra (B. Dejon and P. Henrici, ed.), pp. 193-208. Wiley, London, 1969.

[8] Marden, M.: Geometry of polynomials. Math. Surveys No. 3, Second edition. Amer. Math. Soc., Providence, 1966.

[9] Ostrowski, A.: Recherches sur la méthode de Graeffe et les zéros des polynômes et les séries de Laurent. Acta Math. 72, pp. 99-257 (1940).

[10] Rosenbloom, P. C.: An elementary constructive proof of the fundamental theorem of algebra. Amer. Math. Monthly 52, pp. 562-570 (1945).

[11] Stewart, G. W. III: Some Topics in Numerical Analysis. Oak Ridge National Laboratory Report ORNL-4303. Oak Ridge, Tennessee, September, 1968.