ON THE SOLUTION OF POLYNOMIAL EQUATIONS USING CONTINUED FRACTIONS

Alkiviadis G. AKRITAS
Department of Applied Mathematics II, University of Athens, Athens 621, Greece

Received 28 August 1979

Zeroes of polynomials, continued fractions

It is well known that, in the beginning of the 19th century the mathematicians proved the impossibility of solving algebraically polynomial equations of degree greater than four and as a result their attention focused on numerical methods. During this period Fourier conceived the idea to proceed in two steps; that is, first to isolate the roots and then to approximate them to any desired degree of accuracy. Approximation is a rather trivial task and will not be discussed in this paper; moreover, we will be mainly concerned with the real roots.

Isolation of the real roots of a polynomial equation is the process of finding real, disjoint intervals such that each contains exactly one real root and every root is contained in some interval. In order to accomplish this, Sturm's method is the only one widely known and used since 1830; since it can be found in the literature [5] we describe it briefly.

For any given square-free polynomial equation $P(x) = 0$, Sturm's method works as follows: we compute an absolute upper root bound b, so that all the real roots lie in the interval (−b, b), and then we continuously subdivide (−b, b) until in each subinterval there is at most one root; that is, Sturm's method uses *bisection* in order to isolate the real roots. (As we know, with the help of Sturm's sequence – derived from the polynomials P and P′ – we can easily determine the exact number of real roots in any subinterval (p, q); more details can be found in the literature.) The basic drawback of Sturm's method is the coefficient growth: that is, if the calculations are performed in the ring of integers, the coefficients of the polynomials in the Sturm sequence become too large, and hence, the round-off errors are not at all negligible. This drawback was overcome, though, in 1970, when Sturm's method was programmed in an algebraic manipulation system [4]; moreover, it was shown [4] that its theoretical computing time bound is

$$O(n^{13}L(|P|_\infty)^3) , \qquad (1)$$

where n is the degree of the square-free, integral polynomial equation $P(x) = 0$, and $L(|P|_\infty)$ the length, in bits, of the maximum coefficient in absolute value. The reader immediately feels that Sturm's method is very slow; it has been determined that its slowness is due to the computation of the Sturm sequence.

During this decade new root isolation procedures appeared in the literature [1, pp.1–8]. However, despite the fact that their theoretical computing time bounds are better than (1), they all have one thing in common, namely, they all use *bisection* in order to isolate real roots.

Recently, in our Ph.D. dissertation [1] we developed our own method for the isolation of the real roots of a polynomial equation, a method which by far surpasses all the existing ones in beauty, simplicity and speed; moreover, as we shall see, our method is the only one with polynomial computing time bound which isolates the real roots using *continued fractions.* It is based on the following:

Theorem (Vincent–Uspensky–Akritas). Let $P(x) = 0$ be a polynomial equation of degree $n > 1$, with rational coefficients and without multiple roots, and let $\Delta > 0$ be the smallest distance between any two of its roots. Let m be the smallest index such that

$$F_{m-1}\frac{\Delta}{2} > 1 \quad \text{and} \quad F_{m-1}F_m\Delta > 1 + \frac{1}{\epsilon_n}, \tag{2}$$

where F_k is the k^{th} member of the Fibonacci sequence

1, 1, 2, 3, 5, 8, 13, 21, ...

and

$$\epsilon_n = \left(1 + \frac{1}{n}\right)^{1/(n-1)} - 1 . \tag{3}$$

Then the transformation

$$x = a_1 + \cfrac{1}{a_2 + \cfrac{}{\ddots + \cfrac{1}{a_m} + \cfrac{1}{\xi}}}, \tag{4}$$

(which is equivalent to the series of successive transformations of the form $x = a_i + 1/\xi$, $i = 1, 2, ..., m$) presented in the form of a continued fraction with arbitrary, positive, integral elements $a_1, a_2, ..., a_m$, transforms the equation $P(x) = 0$ into the equation $\tilde{P}(\xi) = 0$, which has not more than one sign variation, in the sequence of its coefficients.

The original form of this theorem (that is, without specifying the quantity m) is due to Vincent alone [6], and appeared in 1836. The proof is omitted since it can be found in the literature [1,2]. It should be mentioned that Vincent's theorem was so totally forgotten that even such a capital work as the *Enzyclopaedie der mathematischen Wissenschaften* ignores it. The author of this paper discovered it in Uspensky's *Theory of Equations* [5].

This theorem can be used in order to isolate the real roots of a polynomial equation. The fact that it holds only for equations without multiple roots does not restrict the generality, because in the opposite case all we have to do is to express P(x) in the form $P = \Pi_{i=1}^{e} S_i^i$, where each of the S_i's has only single roots [5, pp.65–69]. Each of these single roots is of multiplicity i for the polynomial P(x) and thus we see that our theorem can be applied on the S_i's. So, in the rest of this discussion it is assumed that $P(x) = 0$ is without multiple roots.

From the statement of the above theorem we know that a transformation of the form (4), with arbitrary, positive integer elements $a_1, a_2, ..., a_m$ transforms $P(x) = 0$ into an equation $\tilde{P}(\xi) = 0$, which has at most one sign variation; this transformation can be also written as

$$x = \frac{P_m\xi + P_{m-1}}{Q_m\xi + Q_{m-1}}, \tag{5}$$

where P_k/Q_k is the k^{th} convergent to the continued fraction

$$a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \ddots}}$$

(Recall that from the law of convergents we have

$P_{k+1} = a_{k+1}P_k + P_{k-1}$,

$Q_{k+1} = a_{k+1}Q_k + Q_{k-1}$.)

Since the elements $a_1, a_2, ..., a_m$ are arbitrary there is obviously an infinite number of transformations of the form (4). However, with the help of Budan's theorem we can easily determine those that are of interest to us; namely, there is a finite number of them (equal to the number of the positive roots of $P(x) = 0$) which lead to an equation with exactly one sign variation in the sequence of its coefficients. Suppose that $\tilde{P}(\xi) = 0$ is one of these equations; then from the Cardano–Descartes rule of signs we know that it has one root in the interval $(0, \infty)$. If $\hat{\xi}$ was this positive root, then the corresponding root $\hat{x}$ of $P(x) = 0$ could be easily obtained from (5). We only know though that $\hat{\xi}$ lies in the interval $(0, \infty)$; therefore, substituting ξ in (5) once by 0 and once by ∞ we obtain for the positive root $\hat{x}$ its isolating interval, whose unordered endpoints are P_{m-1}/Q_{m-1} and P_m/Q_m. In this fashion we can isolate all the positive roots of $P(x) = 0$. If we subsequently replace x by $-x$ in the original equation, the negative roots become positive and hence, they too can be isolated in the way mentioned above. Thus we see that we have a procedure for isolating all the real roots of $P(x) = 0$.

The calculation of the quantities $a_1, a_2, ..., a_m$ – for the transformations of the form (4) which lead to an equation with exactly one sign variation – constitutes the polynomial real root isolation procedure. Two methods actually result, Vincent's and Akritas', corresponding to the two different ways in which the computation of the a_i's may be performed.

Vincent's method basically consists of computing a particular a_i by a series of unit incrementations; that is, $a_i \leftarrow a_i + 1$, which corresponds to the substitution $x \leftarrow x + 1$. This 'brute force' approach results in a method with an exponential behavior, namely, for big values of the a_i's this method will take a long time (even years in a computer) in order to isolate the real roots of an equation. Therefore, Vincent's method is of little practical importance. Examples of this approach can be found in Vincent's paper [6], and in Uspensky's book [5, pp.129–137]. The reader should notice that in the preface of his book Uspensky claims that he himself invented this method. A simple comparison with Vincent's paper though makes clear that what can be considered a contribution on Uspensky's part is only the fact that he used the Ruffini–Horner method [3] in order to perform the transformations $x \leftarrow x + 1$, whereas Vincent used Taylor's expansion theorem. Moreover, Uspensky seems to ignore Budan's theorem and, while computing a particular a_i, he performs, after each translation $x \leftarrow x + 1$, the unnecessary transformation $x \leftarrow 1/(x + 1)$, something which Vincent avoids.

Akritas' method, on the contrary, is an aesthetically pleasing interpretation of the Vincent–Uspensky–Akritas theorem. Basically it consists of immediately computing a particular a_i as the lower bound b of the positive roots of a polynomial; that is, $a_i \leftarrow b$, which corresponds to the substitution $x \leftarrow x + b$ performed on the particular polynomial under consideration. It is obvious that our method is independent of how big the values of the a_i's are. (An unsuccessful treatment of the big values of the a_i's can be found in Uspensky [5, p.136]. In this discussion it is assumed that $b = \lfloor \alpha_s \rfloor$, where α_s is the smallest positive root.) Since the substitutions $x \leftarrow x + 1$ and $x \leftarrow x + b$ can be performed in about the same time [3], we easily see that our method results in enormous savings of computing time. We have implemented our method in a computer algebra system and have been able to show that its computing time bound is

$$O(n^5 L(|P|_\infty)^3) . \qquad (6)$$

Table 1

Degree	Sturm	Akritas
5	2.05	0.26
10	33.28	0.48
15	156.40	0.94
20	524.42	2.36

Comparing (1) and (6) we clearly see the superiority of our method.

In this paper we present just one table (Table 1) showing the observed times for the methods of Sturm and Akritas (more examples can be found in [1]). All times are in seconds and were obtained by using the SAC-1 computer algebra system on the IBM S/370 Model 165 computer located at the Triangle Universities Computation Center, where a subroutine CCLOCK is available, which reads the computer clock. All the coefficients of the polynomials (of degree 5, 10, 15, 20) were nonzero, each ten decimal digits long and randomly generated.

We see that the most efficient way to isolate the real roots of a polynomial equation is by using continued fractions. In our Ph.D. thesis we have given all the necessary hints as to how the above mentioned theorem may be used in order to isolate the complex roots as well. We hope to have more to say on this subject in the near future.

References

[1] A.G. Akritas, Vincent's theorem in algebraic manipulation, Ph.D. thesis, Operations Research Program, North Carolina State University, Raleigh, NC (1978).

[2] A.G. Akritas, A correction on a theorem by Uspensky, Bull. Greek Math. Soc. 19 (1978) 278–285.

[3] A.G. Akritas and S.D. Danielopoulos, On the complexity of algorithms for the translation of polynomials, Computing, to appear.

[4] L.E. Heindel, Integer arithmetic algorithms for polynomial real zero determination, J. ACM 18 (1971) 533–548.

[5] J.V. Uspensky, Theory of Equations (McGraw-Hill, New York, 1948).

[6] M. Vincent, Sur la résolution des équations numériques, J. Math. Pures Appl. 1 (1836) 341–372.