A new and novel method for computing an upper bound on the distance of an approximate zero from an exact zero of a univariate polynomial

Siddhartha Brahma, Sudebkumar Prasant Pal
Sudhir Kumar Singh*
Department of Computer Science and Engineering
Indian Institute of Technology
Kharagpur 721302, India.
email: spp@cse.iitkgp.ernet.in
http://www.facweb.iitkgp.ernet.in/~spp

Abstract

In this paper we develop a simple and novel method for determining a sharp upper bound on the distance of a given approximate zero from an exact zero of a univariate polynomial. The computed bounds are scalable in the sense that we can compute sharper error bounds for better given approximations of zeros. We analyse the convergence of our method. We use requisite high precision computations for computing our bounds correctly and robustly.

Keywords: error bounds, Rouche's theorem, a posteriori error analysis, approximate zeros, high precision computation.

1 Introduction

Much attention has been paid to computing good approximations to zeros of polynomials. The problem of solving the polynomial equation

$$g(z) = a_0 + a_1 z + a_2 z^2 + ... + a_n z^n = 0 \qquad (1)$$

is a fundamental problem in algebra. Here, the coefficients a_i, $0 \leq i \leq n$ are complex numbers.

The *fundamental theorem of algebra* states that equation (1) always has n complex zeros, counted with multiplicities. The subject of computing or approximating these zeros has been called *algebraic aspects of the fundamental theorem of algebra* [22]. The computational challenge is in determining zeros of high degree univariate polynomials precisely . Since there is no closed form formula for finding zeros of a polynomial of degree greater than four, we need to find zeros by numerical methods. The first iterative algorithms with guaranteed

*Department of Mathematics, Indian Institute of Technology, Kharagpur, 721302, India.

convergence to all the n zeros of $g(z)$ (for any input polynomial $g(z)$ of degree n), are due to Brouwer and due to Weyl, both published in 1924 (see [16]). Pan's new $O(n^2 \log n \log bn)$ algorithm to approximate all the zeros of a n-degree polynomial within an error bound of $2^{-b} max_j |z_j|$, where z_j are the zeros of the polynomial [15], is the best known result so far. There are several algorithms for finding zeros of polynomials based on Newton's method [10], Laguerre's method [23], Jenkins and Traub's method [8], and Muller's method [16].

Since all the above methods essentially compute only approximate zeros, we need to know how close these approximations are to the actual zeros. Much work has been done on this problem, especially for the case of isolated zeros ([17, 6, 5]). For the case of multiple and clustered zeros, most of the methods are based on Gerschgorin's theorem. Smith [24], Carstensen [3], Braess and Hadeler [2], Neumaier [13] and Elsner [4] develop methods based on Gerschgorin's theorem to get bounds for multiple and clustered zeros when good approximations are available.

In the papers cited above, the polynomial and all the approximate zeros are given as input. In this paper we consider a slightly different problem. The only inputs are the polynomial and any one approximate zero of the polynomial. We develop a new and novel method for finding posteriori upper bounds on errors in the given approximate zero. More precisely, we obtain sharp upper bounds on the distance of the approximate zero from an exact zero of the polynomial. Towards this end we use *Rouché's theorem* from complex analysis [1, 6] as the basis for our approximation scheme. Based on this scheme we develop inequalities to determine a sharp upper bound on the distance of an approximate zero from an exact zero. We also analyse the convergence of our method for solving the inequalities, trying to minimize the estimated upper bound. Computations are done at requisite high precision to obtain correct results in solving the inequalities. The coefficients of the polynomial may themselves be approximate and our computations guarantee the correctness of results obtained for the precision of the coefficients of the given polynomial.

The remaining paper is organized as follows. In section 2 we state some algebraic preliminaries. In section 3 we present our scheme for computing bounds on errors in approximate zeros of univariate polynomials in one variable. In section 4 we analyze some standard methods for solving certain inequalities obtained in section 3. In section 5, we state examples where we applied our method of section 3 using **Mathematica (ver.4.0)**. We conclude the paper in section 6 with a few remarks on the results and directions for future research.

2 Preliminaries

We state the following definitions and theorems for future reference.

Definition A *complex function* $f(z)$ is said to be analytic in a region if it is defined and differentiable at every point z in the region.

Theorem 2.1 *[1, 6] Suppose the functions $f(z)$ and $g(z)$ are analytic on and inside a simple close curve C . If f and g have no zeros on C and $|f(z)-g(z)| < |f(z)|$ for all z on C, then the functions $f(z)$ and $g(z)$ have the same number of zeros inside C.*

The above stated theorem is called *Rouché's theorem*. It is a well known fact that all polynomials with complex coefficients are analytic in the complex plane. Using this fact and applying *Rouché's theorem* we derive the main theorem of the paper in the following section.

3 A novel posteriori error bound for an approximate zero

Let $g(z)$ be a univariate monic polynomial in z. In a practical situation the coefficients of $g(z)$ will not be known to us exactly, we will only have some close approximations to the actual coefficients. For example if the coefficients are derived from some data acquisition system, there will be some error in the coefficients due to error in the measurements. Also, some amount of error will always creep in because of the finite precision of representation of the coefficients. We ensure that the bounds obtained are correct for the given precision of the coefficients. We assume $g(z)$ to be monic.

Let z_0 be a given approximate zero of $g(z)$. It is our aim to find an upper bound on the distance of z_0 from an exact zero of $g(z)$. Let $f(z)$ be defined such that

$$g(z) = (z - z_0)f(z) + g(z_0) \tag{2}$$

Let $r = |z - z_0|$. Let b be the leading coefficient of $f(z)$. Therefore, we have

$$f(z) = bf_1(z)$$

where $f_1(z)$ is a monic polynomial of degree $(n-1)$ in z. Let $z_1, z_2, \cdots, z_{n-1}$ be approximate zeros of $f_1(z)$. We define the polynomial $c(z)$ as follows

$$c(z) = f_1(z) - \prod_{i=1}^{n-1}(z - z_i) \tag{3}$$

Note that since the zeros of $f_1(z)$ may be approximate , $c(z)$ is not 0 in general. Since $f_1(z)$ is a monic polynomial of degree $(n-1)$, $c(z)$ is a polynomial of degree $(n-2)$ .Therefore we have

$$c(z) = \sum_{i=0}^{n-2} c_i z^i$$

where c_i's are complex numbers. Using this fact we prove the following Lemma.

Lemma 3.1 *If $b, f_1(z), c_i$'s are defined as above and if*

$$\prod_{i=1}^{n-1} |(|z_i - z_0| - r)| \geq \sum_{i=0}^{n-2} |c_i|(|z_0| + r)^i$$

then

$$|(z - z_0)f(z)| \geq |b|r\left(\prod_{i=1}^{n-1} |\,(|z_i - z_0| - r)\,| - \sum_{i=0}^{n-2} |c_i|(|z_0| + r)^i\right) \tag{4}$$

Proof

$$\begin{aligned} |(z-z_0)f(z)| &= |z-z_0||b||f_1(z)| \\ &= r|b||c(z)+\prod_{i=1}^{n-1}(z-z_i)| \\ &\geq r|b|\left|\left(\prod_{i=1}^{n-1}|(z-z_i)|-|c(z)|\right)\right| \end{aligned}$$

We also have

$$\prod_{i=1}^{n-1}|(z-z_i)| \geq \prod_{i=1}^{n-1}|(|z_i-z_0|-r)| \tag{5}$$

$$|c(z)| \leq \sum_{i=0}^{n-2}|c_i|(|z_0|+r)^i \tag{6}$$

If

$$\prod_{i=1}^{n-1}|(|z_i-z_0|-r)| \geq \sum_{i=0}^{n-2}|c_i|(|z_0|+r)^i \tag{7}$$

then combining (5) and (6) we have

$$\prod_{i=1}^{n-1}|(z-z_i)|-|c(z)| \geq \prod_{i=1}^{n-1}|(|z_i-z_0|-r)|-\sum_{i=0}^{n-2}|c_i|(|z_0|+r)^i \geq 0 \tag{8}$$

Thus , if condition (7) is satisfied , using (8) , we have

$$|(z-z_0)f(z)| \geq |b|r\left(\prod_{i=1}^{n-1}|\,(|z_i-z_0|-r)\,|-\sum_{i=0}^{n-2}|c_i|(|z_0|+r)^i\right). \quad \blacksquare$$

Consider the two functions $(z-z_0)f(z)$ and $g(z)$. If we can force the condition of *Rouché's theorem* in a circular region of radius r centered at z_0 , since $(z-z_0)f(z)$ has z_0 as a zero in that region , by *Rouché's theorem* $g(z)$ has at least one zero in that region. Thus r gives an upper bound to the distance of z_0 from an exact zero of $g(z)$. This is the essence of our main theorem.

Theorem 3.2 *If r satisfies the following two inequalities*

$$\prod_{i=1}^{n-1}|(|z_i-z_0|-r)| \geq \sum_{i=0}^{n-2}|c_i|(|z_0|+r)^i \tag{9}$$

and

$$|b|r\left(\prod_{i=1}^{n-1}|\,(|z_i-z_0|-r)\,|-\sum_{i=0}^{n-2}|c_i|(|z_0|+r)^i\right) > |g(z_0)| \tag{10}$$

then $g(z)$ has at least one zero in the circular region of radius r centered at z_0.

Proof Consider the two functions $(z-z_0)f(z)$ and $g(z)$. From **Lemma** 3.1 if r satisfies the conditions of this theorem

$$\begin{aligned} |(z-z_0)f(z)| - |(z-z_0)f(z) - G(z)| &\geq |b|r\left(\prod_{i=1}^{n-1} |\,(|z_i - z_0| - r)\,| - \sum_{i=0}^{n-2} |c_i|(|z_0| + r)^i\right) \\ &- |g(z_0)| > 0 \end{aligned}$$

Therefore the conditions of *Rouché's theorem* are satisfied on the curve defined by $|z - z_0| = r$ with respect to the functions $(z-z_0)f(z)$ and $g(z)$. Thus $(z-z_0)f(z)$ and $g(z)$ have the same number of zeros in the circular region defined by $|z - z_0| < r$. Since z_0 is the center of this region , $g(z)$ has at least one zero in the region. Therefore r is an upper bound to the distance of z_0 from an exact zero of $g(z)$. ∎

Note that the solution (say r') of the inequality will give us an upper bound on the distance of z_0 from *at least* one exact zero of $g(z)$. If there is only one zero of $g(z)$ in the disk of radius r' centered at z_0, then r' is an upper bound on the distance of z_0 from the exact zero of $g(z)$ closest to z_0. But if there is more than one exact zero of $g(z)$ in that disk, then r' is an upper bound on the distance of z_0 from all of them, and not just the one closest to z_0.

4 Solving the inequality and convergence criteria

The main **Theorem** 3.2 suggests a method for finding an upper bound on the distance of z_0 from an exact zero of $g(z)$. All we need to do is solve inequalities (9) and (10) for a positive real value of r. It is easy to see that if inequality (10) is satisfied then inequality(9) is automatically satisfied. So, a value of r satisfying inequality (10) gives us the required upper bound. We can guarantee the existence of a solution for (10) as is shown in the following theorem.

Theorem 4.1 *Inequality (10) has a positive real solution.*

Proof We can write (10) as

$$|b|r\prod_{i=1}^{n-1} |\,(|z_i - z_0| - r)\,| - |b|r\left(\sum_{i=0}^{n-2} |c_i|(|z_0| + r)^i\right) - |g(z_0)| > 0$$

The first term in the left hand side of the inequality is of degree n, and all the other terms are of lesser degree. As r tends to $+\infty$, the term with the maximum degree dominates. Since we are interested in the modulus of its value, the dominating value is positive. Thus, inequality (10) has a positive real solution. ∎

Once we have established that (10) has a solution, we need to find it. Obtaining a good value of r reduces to the non-linear optimization problem of solving inequality (10). There are several well known methods for solving polynomial inequalities such as (10). Typically, we can adopt a polynomial zero finding algorithm to solve polynomial inequalities. In this paper we illustrate the use of a safe variant of the Newton-Raphson method [18], which combines the bisection method and the Newton-Raphson method, to solve the inequality. The hybrid algorithm takes a bisection step whenever Newton-Raphson would take the solution out of bounds or whenever Newton-Raphson is not reducing the size of the size of the brackets rapidly enough.In this algorithm, it is necessary to bracket the zero initially and it is ensured that the successive approximations to the zeros do not stray outside the brackets. Once we have the zero bracketed, we can apply the algorithm to get a very good approximation of the zero of $P(r)$.Then we can simply search in a small neighbourhood of the approximate zero to get a solution to the inequality. The above method converges always if we can bracket the root initially because the bisection method is guaranteed to converge and the Newton-Raphson method is not allowed to stray out of bounds.

In this paper, we analyse the solution of (10) for a special case when z_0 is a "good" approximation to a zero of $g(z)$ and the error in the zero is much smaller than $|z_i - z_0|, (1 \leq i \leq n-1)$. In what follows let

$$L = min\{|z_i - z_0|, 1 \leq i \leq (n-1)\}.$$

If the above assumption is valid, we can restict our attention to find a solution of (10) in the interval (0,L). Let

$$P(r) = |b|r \prod_{i=1}^{n-1} (|z_i - z_0| - r) - |b|r \left(\sum_{i=0}^{n-2} |c_i|(|z_0| + r)^i \right) - |g(z_0)|$$

Since we are working in $(0, L)$, inequality (10) can be rephrased as

$$P(r) > 0 \tag{11}$$

The zero finding procedure typically works well if we can initially bracket the zero within a small interval. Since $P(0) < 0$, for bracketing a zero of $P(r)$ which lies in $(0, L)$, we have to find a positive real $r_0 \in (0, L)$ such that $P(r_0) > 0$. Let us rewrite (11) in the following equivalent form

$$r > Q(r) \tag{12}$$

where

$$Q(r) = \frac{|g(z_0)| + |b|r \left(\sum_{i=0}^{n-2} |c_i|(|z_0| + r)^i \right)}{|b| \prod_{i=1}^{n-1} (|z_i - z_0| - r)} \tag{13}$$

Consider the equality corresponding to the inequality (12). If the above assumptions are valid,

$$r \ll |z_0|, r \ll |z_i - z_0| (1 \leq i \leq n-1)$$

then

$$Q(r) \approx \frac{|g(z_0)| + |b|r \left(\sum_{i=0}^{n-2} |c_i||z_0|^i \right)}{|b| \prod_{i=1}^{n-1} (|z_i - z_0|)}$$

Therefore, for

$$r_1 = \frac{|g(z_0)|}{|b| \left(\prod_{i=1}^{n-1} (|z_i - z_0|) - \sum_{i=0}^{n-2} |c_i||z_0|^i \right)},$$

$$r_1 \approx Q(r_1)$$

Thus, under the assumptions mentioned above, if $r_1 < L$, r_1 can serve as a good starting point for searching an r_0 such that $P(r_0) > 0$ and the method outlined above converges quickly. This is illustrated by examples in the next section. Of course, if the assumptions made above do not hold, we will have to adopt a trial and error approach or use more sophisticated methods to bracket the root initially.

The sharpness of the upper bound obtained from the solution of inequality (10) depends on many factors. The closer z_0 is to an actual zero of $g(z)$, the smaller is the value of $|g(z_0)|$. This results in better upper bounds and thus explains the scalability of our method. In this method we require the zeros of $f_1(z)$ to compute the bounds. In fact, if all the approximate zeros of $g(z)$ are given simultaneously we can avoid the computation of zeros of $f_1(z)$ under certain conditions. If $g(z)$ is a well conditioned polynomial and z_0 is a "good" approximation, it is likely that the other zeros of $g(z)$ are "good" approximations to zeros of $f_1(z)$ as well. Thus we may use them as our z_i's $(1 \leq i \leq n-1)$ as well.

Clearly, the sharpness of the upper bound also depends on how close z_i's$(1 \leq i \leq n-1)$ are to the actual zeros of $f_1(z)$. Clearly, the better the approximations are, the smaller will be the values of $|c_i|$'s. This will result in better upper bounds. The case of multiple zeros needs to be treated bit differently. If we know beforehand that z_0 is an approximation to a multiple zero of known multiplicity m, then we can simply write

$$g(z) = (z - z_0)^m f(z) + g_1(z)$$

where $f(z)$ is a polynomial of degree $(n - m)$ and $g_1(z)$ is a polynomial of degree $(m - 1)$. The rest of the derivation follows easily by using the triangle inequality on $|g(z)|$. If the multiplicity is not known , then it may happen that one of the roots (say z_t) of $f_1(z)$ is very close to z_0 and hence $|z_t - z_0|$ is very small. For this case we may require a higher precision in the internal computations to guarantee a correct solution to the inequality. If z_0 is a "good" approximation for an isolated zero of $g(z)$, the presence of other clustered zeros of $g(z)$ does not affect the computation as we are only concerned with the values of $|z_i - z_0|$ and the other zeros of $g(z)$ are likely to be close to the z_i's (for a well conditioned polynomial $g(z)$). If the zeros are clustered and the approximate zero is close to the cluster, we may again require high precision computation to deal with this case effectively.

5 Results and observations

The method oulined in the paper has been extensively tested, and the results obtained were very positive. The calculations have been carried out using **Mathematica(ver.4.0)**. We show the results obtained for the following test polynomials

1. $g(z) = (z-1)(z-2)(z-3)(z-4)(z-5)$
2. $g(z) = (z-2)^3(z-2.5)(z-3)$
3. $g(z) = (z - 2.0000000004877320)(z - 2.0000000004877460)(z - 1)$

Ser.No.	z_0	Approximate zeros of $f_1(z)$	r_1	Final Solution
1(a)	1.0000004500000000	2.0000001000000000 3.0000001000000000 4.0000001000000000 5.0000001000000000	$4.49999906 \times 10^{-7}$	$4.50002386 \times 10^{-7}$
1(b)	1.0000000000004500	2.0000001000000000 3.0000001000000000 4.0000001000000000 5.0000001000000000	$4.49999906 \times 10^{-12}$	$4.50000548 \times 10^{-12}$
2(a)	2.0000000004000000	2.5000003000000000 3.0000002000000000	$1.13548197 \times 10^{-9}$	$1.87337182 \times 10^{-9}$
2(b)	2.0000000000000040	2.5000003000000000 3.0000002000000000	$3.99999893 \times 10^{-15}$	$1.53892784 \times 10^{-14}$
3	2.0000000004877540+ 0.0000001732049710i	2.000000000487754000− 0.0000001732049710i 0.9999999999999700	$8.66024855 \times 10^{-8}$	$1.73206000 \times 10^{-7}$

The first example shows the scalability of our method. Note that the roots of $f_1(z)$ were taken to be the same for both the cases, yet the results are highly accurate for both the cases. The second example illustrates the case of a multiple root. Here it is assumed that the mutiplicity was known beforehand. Again the results are highly accurate and scalable. The third example is that of clustered roots. For both the second and third examples requisite high precision was used to guarantee the correct evaluation of the solutions.

6 Conclusions

Posteriori error bounds on approximate zeros of univariate polynomials may be used in geometric computations with high degree curves and surfaces as well as in various scientific, engineering and computational mathematics applications. Our method for computing bounds can be used for computing approximate zeros of polynomials to desired accuracies. At each stage of computing the zeros, we may check how accurate the zeros actually are and stop the computation when a desired accuracy is reached.

We have used requisite high precision in our algorithms to ensure correct evaluation of polynomials and checking of inequalities. We can deal with the case of clustered zeros by using high precision arithmetic. The case of zeros with multiplicity > 1 can also be handled with slight modification. An important advantage of the proposed scheme is its scalability. With more precise zeros, we get better upper bounds. Also, if the polynomials are well behaved, more precise the values of $z_1, z_2, \cdots, z_{n-1}$ are, the better the bounds are.

Our method works for univariate polynomials. One future research direction is that of bounding errors in approximate zeros of more general class of functions of one variable e.g. analytic functions of one variable. There is also scope of future work in developing notions of errors in approximations to zeros of multivariate polynomials and functions.

References

[1] Ahlfors L. V., "Complex Analysis", McGraw-Hill International Edition, 1979.

[2] Braess D., Hadeler K.P., Simultaneous inclusion of the zeros of a polynomial, Numer. Math. 21 (1973), 161-165.

[3] Carstensen C.,Inclusion of the roots of a polynomial based on Gerschgorin's theorem, Numer. Math. 59 (1991), 349-360.

[4] Elsner L.,A remark on the simultaneous inclusions of the zeros of a polynomial by Gerschgorin's theorem, Numer. Math. 21 (1973), 425-427

[5] Gargantini G., Henrici P., Circular arithmetic and the determination of polynomial zeros, Numer. Math. 18 (1972), 305-320.

[6] Henrici P., "Applied and Computational Complex Analysis, Vol. 1", John Wiley and Sons, New York, 1974.

[7] Henrici P., Watkins, B.O., "Finding zeros of a polynomials by the Q-D algorithm", Comm. ACM 8,9 (Sept 1965), 570-574.

[8] Jenkins M. A. and Traub J. F., Algorithm 419: Zeros of a complex polynomial, Comm. ACM 15 (1972), 97-99.

[9] Krishnan S., Foskey M., Culver T., Keyser J and Manocha D., "PRECISE: Efficient multiprecision evaluation of algebraic roots and predicates for reliable geometric computation", Proc. ACM Symp. on Comput. Geom., June 2001, 274-283.

[10] Madsen K., A root-finding algorithm based on Newton's method, BIT 13 (1973), 71-75.

[11] Mehlhorn K. and Naher S., "The LEDA platform for Combinatorial and Geometric Computing", Cambridge University Press, 1999.

[12] Mukherjee, M., Pal, S. P., Varvani M. K., Tripathi M., "Safe implementation of set operators using finite precision", CSG 96, Proceedings of the conference held in Winchester, UK, Set-theoretic Solid Modelling: Techniques and Applications, pp. 291-305, April 1996.

[13] Neumaier A., Enclosing clusters of zeros of polynomials, Jnl. of Comp. and Appl. Math 156 (2003), 389-401.

[14] POSSO-Polynomial System Solving project, http://www-sop.inria.fr/saga/POL.

[15] Pan, V.Y., "Approximating Complex Polynomial Zeros: Modified Weyl's Quad-tree Construction and Improved Newton's Iteration", J. of Complexity 16, 213-264 (2000).

[16] Pan, V.Y., "Solving a polynomial equation: Some history and recent progress", SIAM Review, vol.39, No.2, pp. 187-220, June 1997.

[17] Petkovic M.S., Herceg D.,Ilic S., Point estimation and some application to iterative methods, BIT 38 (1998) 111-128.

[18] Press,William H.,et.al.,"Numerical Recipes in C : The Art of Scientific Computing", Second Edition,Cambridge University Press,2002.

[19] Ramakrishna, P. H. D., "Bounding errors in the computation of trigonometric functions and roots of polynomials", M. Tech. (Master's Thesis), Department of Computer Science and Engineering, Indian Institute of Technology, Kharagpur, 721302, India. July, 2000.

[20] Ramakrishna, P. H. D., Bhalla, S., Basu, H., and Pal, S. P., "Computing bounds on error in solutions of algebraic equations using Rouche's theorem", Technical Report # TR/IIT/CSE/2002/SPP1, Department of Computer Science and Engineering, Indian Institute of Technology, Kharagpur, 721302, India.

[21] Ramakrishna, P. H. D., Pal, S. P., Bhalla, S., Basu, H., Singh, S. K., "Computing bounds on errors in approximate zeros of univariate polynomials", communicated for publication.

[22] Smale, S., "The Fundamental Theorem of Algebra and Complexity Theory", Bull. Amer. Math. Soc., 4 (1981), pp. 1-36.

[23] Smith, B.T., "ZERPOL, a zero finding algorithm for polynomials using Laguerre's method", Proc. 1967 Army Numerical Analysis Conference, Madison, Wis., May 1967 (Rep. 67-3, US Army Res. Office- Durham, Durham, N.C., Nov 1967). pp. 153-174.

[24] Smith, B.T., "Error Bounds for Zeros of a Polynomial Based Upon Gerschgorin's Theorem", Journal of ACM, Vol.17, No.4, Oct'1970, pp. 661-674.

[25] Sturm C., "Mémoire sur la résolution des équations numériques" , Bulletin des sciences de Férussac 11, 1929