ZASTOSOWANIA MATEMATYKI
APPLICATIONES MATHEMATICAE
XI, 2 (1970)

H. WOŹNIAKOWSKI (Warszawa)

SOME REMARKS ON BAIRSTOW'S METHOD (1)

1. To calculate the zeros of a real polynomial by certain procedures using Bairstow's method ([2], [4]), it is necessary to give numerical values to certain parameters which determine the stopping moment of the calculations. These parameters may, for instance, give the maximum number of iterations, or may determine the required accuracy of the solution. Efforts in the direction of giving numerical values to those parameters lead to numerous doubts. If the accuracy requirements are too weak (or the number of iterations too small) one obtains a bad solution, if, on the other hand, the accuracy requirements are too strong the solution procedure may last very long or even never end (because of rounding errors).

This paper contains in section 3 a description of an ,,controlled" algorithm of Bairstow's method, the purpose of which is to determine quadratic divisors of a real polynomial with maximum accuracy obtainable in the given floating-point arithmetics. In section 2 some considerations on the character of convergence of the generalized Bairstow method for the case of multiple zeros are given. The author has not found analogous considerations in the available literature.

2. Let us begin with an explanation of the character of convergence of Bairstow's method in the case when the polynomial has multiple zeros. Consider a real polynomial of degree n

(1) $$w(x) = a_n x^n + a_{n-1} x^{n-1} + \ldots + a_1 x + a_0.$$

Assume that we have to determine the divisor $m_0(x)$ of degree r of the polynomial (1)

$$m_0(x) = x^r - p_{r-1}^{(0)} x^{r-1} - \ldots - p_1^{(0)} x - p_0^{(0)} \quad (1 \leqslant r \leqslant n).$$

(1) This research has been performed under contract for the Department of Numerical Calculations, University of Warsaw, in 1967.

Editor's note: This is a modified, English version of the author's paper which has been awarded the first prize in the 1968 Polish Mathematical Society Contest for the Best Student Paper in the Theory of Probability and Applications of Mathematics.

Let the functions

$$A_i = A_i(p_{r-1}, p_{r-2}, \ldots, p_0) \qquad (i = 0, 1, \ldots, r-1)$$

denote the coefficients of the remainder while dividing the polynomial $w(x)$ by

$$m(x) = x^r - p_{r-1}\, x^{r-1} - \ldots - p_0,$$

i. e. let

(2) $$w(x) = m(x)Q(x) + A_{r-1}x^{r-1} + A_{r-2}x^{r-2} + \ldots + A_0.$$

Having an approximate divisor

$$m_1(x) = x^r - p_{r-1}^{(1)}x^{r-1} - \ldots - p_1^{(1)}x - p_0^{(1)},$$

we determine the next approximation

$$m_2(x) = x^r - p_{r-1}^{(2)}x^{r-1} - \ldots - p_1^{(2)}x - p_0^{(2)},$$

where

$$p_i^{(2)} = p_i^{(1)} + \Delta p_i^{(1)} \qquad (i = 0, 1, \ldots, r-1).$$

The quantities $\Delta p_i^{(1)}$ are defined as the solution of the following system of linear equations (see [9])

$$\frac{\partial A_0(p_{r-1}^{(1)}, \ldots, p_0^{(1)})}{\partial p_0}\Delta p_0^{(1)} + \ldots + \frac{\partial A_0(p_{r-1}^{(1)}, \ldots, p_0^{(1)})}{\partial p_{r-1}}\Delta p_{r-1}^{(1)} = -A_0(p_{r-1}^{(1)}, \ldots, p_0^{(1)}),$$

$$\vdots$$

$$\frac{\partial A_{r-1}(p_{r-1}^{(1)}, \ldots, p_0^{(1)})}{\partial p_0}\Delta p_0^{(1)} + \ldots + \frac{\partial A_{r-1}(p_{r-1}^{(1)}, \ldots, p_0^{(1)})}{\partial p_{r-1}}\Delta p_{r-1}^{(1)} = -A_{r-1}(p_{r-1}^{(1)}, \ldots, p_0^{(1)}).$$

This system gives the corrections while solving the system of equations

$$A_i(p_{r-1}, p_{r-2}, \ldots, p_0) = 0 \qquad (i = 0, 1, \ldots, r-1)$$

by Newton's method.

It is known from the theory of Newton's method that if the sequence $\{p_{r-1}^{(k)}, p_{r-2}^{(k)}, \ldots, p_0^{(k)}\}$ converges to $\{p_{r-1}^{(0)}, p_{r-2}^{(0)}, \ldots, p_0^{(0)}\}$ with $k \to \infty$ and if

$$J_0 = \left| \begin{matrix} \frac{\partial A_0}{\partial p_0}, \frac{\partial A_0}{\partial p_1}, \ldots, \frac{\partial A_0}{\partial p_{r-1}} \\ \frac{\partial A_1}{\partial p_0}, \frac{\partial A_1}{\partial p_1}, \ldots, \frac{\partial A_1}{\partial p_{r-1}} \\ \cdot \\ \frac{\partial A_{r-1}}{\partial p_0}, \frac{\partial A_{r-1}}{\partial p_1}, \ldots, \frac{\partial A_{r-1}}{\partial p_{r-1}} \end{matrix} \right|_{p_i = p_i^{(0)}} \qquad (i = 0, 1, \ldots, r-1)$$

is different from zero, then the convergence is a quadratic one, i. e. there exists such a constant L that for sufficiently large k holds

$$\sqrt{(p_0^{(0)}-p_0^{(k+1)})^2+(p_1^{(0)}-p_1^{(k+1)})^2+\ldots+(p_{r-1}^{(0)}-p_{r-1}^{(k+1)})^2}$$
$$\leqslant L[(p_0^{(0)}-p_0^{(k)})^2+(p_1^{(0)}-p_1^{(k)})^2+\ldots+(p_{r-1}^{(0)}-p_{r-1}^{(k)})^2].$$

If $J_0 = 0$ then the convergence is weaker, usually linear (i.e. comparable with the convergence of a geometric progression).

Let us differentiate (2) with respect to p_i $(i = 0, 1, \ldots, r-1)$

$$0 = -x^i Q(x)+m(x)\frac{\partial Q(x)}{\partial p_i}+\frac{\partial A_{r-1}}{\partial p_i}x^{r-1}+\ldots+\frac{\partial A_0}{\partial p_i}.$$

Hence, we have

$$x^i Q(x) = m(x)R_i(x)+\frac{\partial A_{r-1}}{\partial p_i}x^{r-1}+\ldots+\frac{\partial A_0}{\partial p_i} \quad (i = 0, 1, \ldots, r-1),$$

where $R_i(x) = \partial Q(x)/\partial p_i$ is a polynomial of the variable x of degree $n-2r+i$, as may easily be seen.

The $(i+1)$-th column of J_0 is thus composed of the coefficients of the remainder of the division of the polynomial $x^i Q_0(x)$ by $m_0(x)$, where $Q_0(x)$ is the quotient of dividing $w(x)$ by $m_0(x)$.

Now, write the equations

$$x^i Q_0(x) = m_0(x)T_i(x)+\sum_{j=0}^{r-1} A_{ji}x^j \quad (i = 0, 1, \ldots, r-1), \tag{3}$$

where $T_i(x)$ is the quotient and A_{ji} the coefficients of the remainder of the division.

It follows that

$$J_0 = |A_{ji}| \quad (j, i = 0, 1, \ldots, r-1). \tag{4}$$

Denote the zeros of the divisor $m_0(x)$ by $z_1, z_2, \ldots, z_r$. We shall prove the following

THEOREM. *The determinant J_0 is equal to zero if and only if at least one zero of the divisor $m_0(x)$ is a zero of the quotient $Q_0(x)$.*

Proof. Assume that $J_0 = 0$, i.e. that a certain linear combination of the columns is a zero vector. Thus, there exist the numbers $C_0, C_1, \ldots, C_{r-1}$ such that at least one of them is different from zero and that they satisfy (from (4)) the equality

$$\sum_{i=0}^{r-1} C_i A_{ji} = 0 \quad (j = 0, 1, \ldots, r-1). \tag{5}$$

Let us multiply the i-th equation of (3) by C_i and let us sum them for $i = 0, 1, \ldots, r-1$, having (5) in mind. We obtain

$$Q_0(x)[C_0 + C_1 x + \ldots + C_{r-1} x^{r-1}] \\ = m_0(x)[C_0 T_0(x) + C_1 T_1(x) + \ldots + C_{r-1} T_{r-1}(x)].$$

At least one of the numbers C_i is different from zero, thus $m_0(x)$ is a divisor the polynomial

$$Q_0(x)[C_0 + C_1 x + \ldots + C_{r-1} x^{r-1}].$$

But the degree of the polynomial $\sum_{i=0}^{r-1} C_i x^i$ does not exceed $r-1$, hence at tleas one of the zeros of $m_0(x)$ must be a zero of $Q_0(x)$. The necessity of the condition mentioned in the theorem is thus proved.

Let us prove now the sufficiency of this condition.

Assume that $Q_0(z_i) = 0$ for a given i, thus also that

$$z_i^j Q_0(z_i) = 0 \qquad (i = 0, 1, \ldots, r-1).$$

Multiply the k-th row of J_0 by z_i^k $(k = 1, 2, \ldots, r-1)$ and add all rows to the first one. From (4) we obtain on place j $(j = 0, 1, \ldots, r-1)$ of this row

$$\sum_{k=0}^{r-1} z_i^k A_{kj}.$$

Due to (3) this is equal to

$$z_i^j Q_0(z_i) - m_0(z_i) T_j(z_i) = 0.$$

Thus the first row consists of zeros only, hence $J_0 = 0$. This ends the proof of the theorem.

Of course, if z_i is a zero of $Q_0(x)$, it is a multiple zero of $w(x)$. It is, however, possible that z_0 is a k-fold $(k \leqslant r)$ zero of $w(x)$ and the divisor is of the form

$$m_0(x) = (x - z_0)^k m_1(x).$$

Then $Q_0(z_0) \neq 0$ and the method remains quadratic convergent. From the considerations given previously the following conclusion may be drawn:

If Bairstow's method is convergent then it is quadratic convergent if and only if every zero of both the divisor $m_0(x)$ and the polynom $w(x)$ has the same multiplicity.

3. Consider now the numerical realization of Bairstow's method for the case of quadratic divisors.

Let

$$m_1(x) = x^2 - px - r$$

be an approximation of the divisor

$$m_0(x) = x^2 - p_0 x - r_0$$

of the polynomial $w(x)$.

Divide $w(x)$ by $m_1(x)$:

$$w(x) = m_1(x) Q_1(x) + A_1 x + A_0$$

and let be

$$Q_1(x) = q_n x^{n-2} + q_{n-1} x^{n-3} + \ldots + q_3 x + q_2 .$$

The algorithm

$$\begin{aligned} q_n &= a_n, \\ q_{n-1} &= p q_n + a_{n-1}, \\ q_k &= p q_{k+1} + r q_{k+2} + a_k \quad (k = n-2, n-3, \ldots, 0) \end{aligned} \tag{6}$$

allows us to calculate the coefficients of the polynomial $Q_1(x)$, and the coefficients A_1, A_0 of the remainder may be obtained from

$$A_1 = q_1, \quad A_0 = q_0 - p q_1 .$$

The maximum accuracy obtainable while determining zeros of a function ([5], pp. 96-97) is limited mainly by the accuracy of the calculation of the function values near those zeros. It may also depend upon the accuracy of calculation of other quantities, as e.g. the derivative of the function. This phenomenon occurs often in the case of multiple zeros. A control system based on the error of the function values only may then appear to be unreliable.

In our case the accuracy of the calculation of A_1 A_0, or of the equivalent q_0, q_1, is decidive. For numerical reasons we shall consider instead of the system (q_i) given by (6) the equivalent system $(\bar{q}_i)$ given by

$$\begin{aligned} \bar{q}_n &= fl(a_n), \\ \bar{q}_{n-1} &= fl(p \bar{q}_n + a_{n-1}), \\ \bar{q}_k &= fl(p \bar{q}_{k+1} + r \bar{q}_{k+2} + a_k) \quad (k = n-2, n-3, \ldots, 0). \end{aligned} \tag{7}$$

The symbol fl denotes here the floating-point t-digit binary realization of the calculations (see [5], p. 11).

Introduce the quantities v_k and E_k defined by

$$fl(W_k) = W_k + v_k, \quad \bar{q}_k = q_k + E_k,$$

where W_k is a given algebraic expression.

The quantity v_k denotes the rounding error obtained in the floating-point calculation of W_k, and E_k denotes the rounding error which „burdens'' the calculated value $\bar{q}_k$.

Introducing this notation into (7), we obtain

$$\bar{q}_n = a_n + v_n,$$
$$\bar{q}_{n-1} = p\bar{q}_n + a_{n-1} + v_{n-1} = q_{n-1} + E_{n-1},$$
$$\bar{q}_k = p\bar{q}_{k+1} + r\bar{q}_{k+2} + a_k + v_k = q_k + E_k \quad (k = n-2, n-3, \ldots, 0).$$

Hence

$$E_n = v_n = 0,$$
$$E_{n-1} = v_{n-1},$$
$$E_k = pE_{k+1} + rE_{k+2} + v_k \quad (k = n-2, n-3, \ldots, 0).$$

It may be shown (e.g. by induction) that the quantities E_i are equal (see [3]) to

$$\text{(8)} \quad E_i = \begin{cases} \sum_{k=i}^{n-1} v_k \dfrac{z_1^{k+1-i} - z_2^{k+1-i}}{z_1 - z_2} & (z_1 \neq z_2), \\ \sum_{k=i}^{n-1} v_k (k+1-i) z_1^{k-i} & (z_1 = z_2), \end{cases} \quad (i = 0, 1, \ldots, n-1),$$

where z_1, z_2 denote now the zeros of $m_1(x)$.

We are interested in the errors E_1, E_0 which burden the quantities q_0, q_1. The maximum limiting accuracy is obtained if both $|\bar{q}_1|$ is of order $|E_1|$ and $|\bar{q}_0|$ is of order $|E_0|$. We cannot indicate any practical method of determining $|E_1|$ and $|E_0|$. We may, however, determine realistic a posteriori estimations of those quantities (see [5], pp. 37-38).

It is easy to show that v_i fulfill the inequalities

$$\text{(9)} \quad |v_i| \leqslant 2^{-t} e_i \quad (i = 0, 1, \ldots, r-1),$$

where t is the number of digits of the mantissa in the floating-point binary representation of numbers, and

$$e_{n-1} = |p\bar{q}_n| + |p\bar{q}_n + a_{n-1}|,$$
$$e_k = |p\bar{q}_{k+1}| + |r\bar{q}_{k+2}| + |p\bar{q}_{k+2}| + |p\bar{q}_{k+1} + r\bar{q}_{k+2} + a_k|$$
$$(k = n-2, n-3, \ldots, 0).$$

Let $l = \max(|z_1|, |z_2|)$. The quantities E_i may be estimated from (8) and (9) in the following way:

$$\text{(10)} \quad |E_i| \leqslant D_i = \begin{cases} \dfrac{2 \cdot 2^{-t}}{|z_1 - z_2|} \sum_{k=i}^{n-1} e_k l^{k+1-i} & (z_1 \neq z_2) \\ 2^{-t} \sum_{k=i}^{n-1} e_k (k+1-i) l^{k-i} & (z_1 = z_2) \end{cases} \quad (i = 0, 1, \ldots, n-1).$$

The quantities D_i may be calculated by Horner's algorithm.

A calculation of the system $(\bar{q}_i)$ $(i = n, n-1, \ldots, 0)$ allows a simultaneous calculation of e_i and thus also of D_i.

Instead of the condition $z_1 \neq z_2$ which is inadequate in numerical calculations the condition

$$|z_1 - z_2| > 1/5n$$

is verified in the procedures [6]. That assures that the calculated value is at most 10% greater than the true one.

Next, we verify whether $\bar{q}_0$, $\bar{q}_1$ are absolutely smaller than D_0, D_1. If so, this indicates that we are near the maximum limiting accuracy.

The next approximations of the quadratic divisor $m_0(x)$ are calculated as long as both the quantities $\bar{q}_0^2 + \bar{q}_1^2$ are diminishing and the coefficients $\bar{q}_0$, $\bar{q}_1$ do not exceed the estimations D_0, D_1.

Such a procedure allows to attain the best possible accuracy in the given arithmetics. A realization of this algorithm in the Gier Algol III system resulted in that about 36% of the realization time of the whole algorithm of Bairstow's method was spent for calculating the approximations of the errors. In the Gier Algol-double system, however, the cost of calculating errors was very small as compared with the cost of the whole algorithm (see [1], [8]).

The algorithm described in this paper is presented in the Algol language in the algorithm section of this number of the journal [7], and further details may be found in [6]. At the department of Numerical Calculations, University of Warsaw, Algol procedures of Bairstow's method including a fuller control system which consists not only in calculating the estimation of the errors of q_0, q_1 but also in calculating the errors of the coefficients of the remainder of the division of $Q_1(x)$ by $m_1(x)$ and the corrections Δp, Δr have been developed.

References

[1] H. Christiansen and others, *A manual of Gier-Algol III*, Regnecentralen, Copenhagen 1966.

[2] A. A. Grau, *Procedure bairstow, algorithm* 3, Comm. ACM (1960).

[3] F. W. J. Olver, *The evaluation of zeros of high-degree polynomials*, Phil. Trans. Roy. Soc. A 244 (1952), pp. 385-415.

[4] O. L. Rasmussen, *The procedure BAIRSTOW*, Gier System Library no. 130, 1963.

[5] J. H. Wilkinson, *Błędy zaokrągleń w procesach algebraicznych*, Polish edition of *Rounding errors in algebraic processes*, PWN, Warszawa 1967.

[6] H. Woźniakowski, *Procedury Bairstow sp i Bairstow dp*, Biblioteka ZON UW, 1967.

[7] —, *Polynomial decomposition into quadratic factors with controlled accuracy by Bairstow's method*, Algorithm 4, Zastosow. Matem. 11 (1969), pp. 221-226.

[8] J. Zachariassen, *Double precision arithmetics in Gier-Algol III*, Report no. 412, Regnecentralen, Copenhagen 1966.

[9] W. L. Zaguskin, *Przegląd metod numerycznych rozwiązywania równań*, PWN, Warszawa 1965, pp. 139-142.

Received on 15. 12. 1967

H. WOŹNIAKOWSKI (Warszawa)

UWAGI O METODZIE BAIRSTOWA

STRESZCZENIE

W pracy proponuje się algorytm metody Bairstowa dla obliczania dzielnika kwadratowego danego wielomianu rzeczywistego z maksymalną graniczną dokładnością (w arytmetyce zmienno-przecinkowej). Podane jest także twierdzenie podające warunek konieczny i dostateczny na to, by uogólniona metoda Bairstowa była zbieżna w sposób kwadratowy.

Х. ВОЗЬНЯКОВСКИ (Варшава)

ЗАМЕЧАНИЯ К МЕТОДУ БЭРСТОУ

РЕЗЮМЕ

В работе предлагается алгоритм метода Бэрстоу вычисления квадратичного делителя данного вещественного полинома с максимальной предельной точностью (в арифметике с плавающей запятой). Дается также необходимое и достаточное условие для того, чтобы сходимость обобщенного метода Бэрстоу была квадратичной.
