A Globally Convergent Method for Simultaneously Finding Polynomial Roots*

By L. Pasquini and D. Trigiante

Abstract. A new method for the simultaneous approximation of all the roots of a polynomial is given. The method converges for almost every initial approximation, the set of the exceptional starting points being a closed set of measure zero, at least if all the polynomial roots are real and simple. The method exhibits quadratic convergence not only to simple, but also to multiple roots.

1. Introduction. One of the more effective methods for the simultaneous approximation of the roots of a polynomial, the origin of which is attributed to Weierstrass [1], is the one published by Kerner [2], [3] and, independently, by Durand [4]. In a modified form it was also studied by Ehrlich [5] and Aberth [6]. This method can be defined as follows. Let

$$(1) \qquad f(s) = \prod_{k=1}^{N} (s - \alpha_k) = \sum_{k=0}^{N} a_k s^{N-k}, \qquad a_0 = 1,$$

be the polynomial with real roots α_k ($k = 1, 2, \ldots, N$). The method, which, in short, will always be referred to as the W-method, can be defined by the sequences

$$(2) \qquad x_k^{(n)} = x_k^{(n-1)} - \frac{f\left(x_k^{(n-1)}\right)}{\prod_{i \neq k}\left(x_k^{(n-1)} - x_i^{(n-1)}\right)},$$

where $k = 1, 2, \ldots, N$, $n = 1, 2, \ldots$ and $x_1^{(0)}, x_2^{(0)}, \ldots, x_N^{(0)} \in \mathbf{R}$.

It can be proved [2], [7] that the W-method is equivalent to a Newton method in an N-dimensional space by considering the function

$$(3) \qquad F: \mathbf{R}^N \to \mathbf{R}^N, \quad F(x) = \left(F_1(x), F_2(x), \ldots, F_N(x)\right),$$

where the functions F_k are defined as follows:

$$(4) \qquad F_k(x) = S_k(x) + (-1)^{k-1} a_k, \qquad k = 1, 2, \ldots, N,$$

and S_k is the elementary symmetric function of degree k. By putting $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_N)$, one has $F(\alpha) = 0$ as a consequence of the fact that, if $x = \alpha$, the $F_k(x) = 0$ ($k = 1, 2, \ldots, N$) turn out to be the well-known relations between the roots and the coefficients of the polynomial (1), and the W-method described by (2)

135

is, in effect, equivalent to Newton's method

$$(5) \qquad\qquad x^{(n)} = x^{(n-1)} - J^{-1}(x^{(n-1)}) F(x^{(n-1)})$$

applied to the equation $F(x) = 0$. In (5) J^{-1} denotes the inverse of the Jacobian matrix J of the function F. The matrix J is a generalized Vandermonde-type matrix and an explicit expression of J^{-1} can be given [7]. One property of the method is the following

P1: $$\sum_{k=1}^{N} x_k^{(n)} = -a_1, \qquad n = 1, 2, \ldots,$$

essentially due to the definition of F_1, $F_1(x) = \sum_{k=1}^{N} x_k + a_1$, and its linearity.

As the W-method is the same as a Newton method, the convergence will be quadratic whenever $\det J(\alpha) \neq 0$, i.e., whenever the roots α_k are all simple.

Moreover, as (2) always converges in practice, the conjecture has been advanced that the W-method converges with almost every starting point. Except for the case of $N = 2$, however, [1], [7], a proof of this conjecture still has not been given.

In Section 2 a new method will be presented which not only preserves all the above-mentioned favorable properties of the W-method but also presents other characteristics that make it more interesting. In fact, it will be shown that the new method:

(A) is globally convergent (see Section 5), at least when applied to polynomials with real and simple roots, the exceptional set of starting points being a closed set in $\mathbf{R}^N$ of measure zero;

(B) can be used also if the function f is not a polynomial but any sufficiently smooth function;

(C) is at least locally convergent to the real roots if the polynomial (1) has also some complex roots;

(D) can yield quadratic convergence not only in the case of simple roots, but also when multiple roots occur.

In the last section numerical examples will be given which illustrate some of the above properties.

2. The New Method. Let $x_1, x_2, \ldots, x_N \in \mathbf{R}$, $x = (x_1, x_2, \ldots, x_N)$ and let us modify the definition of the function (3) by replacing (4) with

$$(6) \qquad\qquad F_k(x) = f[x_1, x_2, \ldots, x_k], \qquad k = 1, 2, \ldots, N,$$

where $f[x_1, x_2, \ldots, x_k]$ is the divided difference of f with respect to the arguments $x_1, x_2, \ldots, x_k$.

PROPOSITION 1. *Except for permutations of the coordinates, $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_N)$ is the unique solution of the equation*

$$(7) \qquad\qquad\qquad F(x) = 0.$$

Proof. Proposition 1 can be easily proved by considering that, by using the Newton interpolation formula, one obtains, $\forall s \in \mathbf{R}$,

$$f(s) = \sum_{k=1}^{N} F_k(x) \prod_{i=1}^{k-1} (s - x_i) + \prod_{i=1}^{N} (s - x_i). \quad \square$$

Applying the Newton method to the equation (7):

$$(8) \qquad x^{(n)} = \Phi(x^{(n-1)}), \qquad \Phi(x) = x - J^{-1}(x)F(x),$$

the Jacobian matrix $J(x)$ comes out to be the following:

$$J(x) = \begin{pmatrix} f[x_1, x_1] & & & \\ f[x_1, x_1, x_2] & f[x_1, x_2, x_2] & & 0 \\ \cdots\cdots\cdots & \cdots\cdots & \cdots & \\ f[x_1, x_1, x_2,\ldots,x_N] & f[x_1, x_2, x_2,\ldots,x_N] & \cdots & f[x_1, x_2,\ldots,x_N, x_N] \end{pmatrix}$$

and it is important to observe that J is now a lower triangular matrix.

The new method just described (see (3), (6), (8)), as it is defined by the divided differences of the function f, remains meaningful also when f is not a polynomial but any sufficiently smooth function. For example, the following Proposition 2 can be proved with arguments very similar to those used in the proof of Proposition 1.

PROPOSITION 2. *Let f be a smooth enough function. If $F_k(x) = 0$, $k = 1, 2, \ldots, M$, then $x_1, x_2, \ldots, x_M$ are roots of the equation $f(s) = 0$ and vice versa.*

Therefore, whenever it is not explicitly specified that f is the polynomial (1), it will be understood in the following that f is, more generally, any sufficiently smooth function. In that case, N will simply continue to mean the chosen number of scalar equations in (7) and, consequently, the dimension of the space in which the problem will be considered.

PROPOSITION 3. *If f is the polynomial (1), all the approximations $x^{(n)}$ generated by (8) verify property P1, i.e.,*

$$(9) \qquad \sum_{k=1}^{N} x_k^{(n)} = -a_1, \qquad n = 1, 2, \ldots.$$

Proof. Equation (9) follows from the definition of F_N: $F_N(x) = \sum_{k=1}^{N} x_k + a_1$ and from its linearity. □

THEOREM 1. *Let $\Phi(x) = (\Phi_1(x), \Phi_2(x), \ldots, \Phi_N(x))$, where Φ is the function defined by (8). The functions Φ_k can be written as*

$$\Phi_k(x) = x_k - \Delta_k(x), \qquad k = 1, 2, \ldots, N,$$

where the Δ_k satisfy the recurrence relation

$$(10) \quad \Delta_k(x) = \frac{f[x_1, x_2, \ldots, x_k] - \sum_{l=1}^{k-1} f[x_1, x_2, \ldots, x_l, x_l, \ldots, x_k] \Delta_l(x)}{f[x_1, x_2, \ldots, x_k, x_k]},$$

with $\sum_{1}^{0} = 0$.

Proof. The elements J_{kl}^{-1} of J^{-1} can be expressed as follows:

$$(11) \qquad J_{kl}^{-1} = \begin{cases} -\dfrac{1}{J_{kk}} \displaystyle\sum_{h=1}^{k-1} J_{kh} J_{hl}^{-1} & \text{if } k > l, \\ 0 & \text{if } k < l, \\ \dfrac{1}{J_{kk}} & \text{if } k = l. \end{cases}$$

Indeed, it is very easy to check that $(J \cdot J^{-1})_{kl} = \delta_{kl}$ holds in the case of $k \leqslant l$ if J^{-1} is defined by (11). In the remaining case $k > l$, one has:

$$(J \cdot J^{-1})_{kl} = \sum_{h=1}^{k-1} J_{kh} J_{hl}^{-1} + J_{kk} J_{kl}^{-1} = \sum_{h=1}^{k-1} J_{kh} J_{hl}^{-1} - J_{kk} J_{kk}^{-1} \sum_{h=1}^{k-1} J_{kh} J_{hl}^{-1} = 0.$$

Since

$$\Delta_k = \sum_{l=1}^{k} J_{kl}^{-1} F_l,$$

there results, in view of (11),

$$\Delta_k = \frac{1}{J_{kk}} \left[F_k - \sum_{l=1}^{k-1} \sum_{h=1}^{k-1} J_{kh} J_{hl}^{-1} F_l \right] = \frac{1}{J_{kk}} \left[F_k - \sum_{h=1}^{k-1} J_{kh} \sum_{l=1}^{h} J_{hl}^{-1} F_l \right]$$

$$= \frac{1}{J_{kk}} \left[F_k - \sum_{h=1}^{k-1} J_{kh} \Delta_h \right],$$

which concludes the proof. $\square$

COROLLARY 1.1. *Since $\Delta_1(x) = f(x_1)/f'(x_1)$, the first component Φ_1 of Φ is exactly the iteration function of Newton's method applied to the equation $f(x_1) = 0$.*

COROLLARY 1.2. *The functions Δ_k and Φ_k depend only on $x_1, x_2, \ldots, x_k$.*

An important consequence of the above result is that the behavior, for increasing n, of the first coordinates $x_1^{(n)}, x_2^{(n)}, \ldots, x_{m-1}^{(n)}$, is in no way influenced by that of the following coordinates. The main advantage resulting from this property consists of the possibility to reorder the coordinates during the procedure by putting in the first positions those with better convergence properties, in such a way that the possible nonconvergence or slow convergence of some coordinates has no influence on the others.

It will be shown that this property can be useful not only if the equation $f(s) = 0$ is solved by the method defined by (3), (6) and (8) with a value of N greater than the total number of the real roots, but also in the general case, especially if not all the roots α_k are simple or, more generally, if real roots of different multiplicities occur. In these cases the above-described technique of reordering coordinates results in an appreciable convergence improvement (see also, for a relevant numerical example, Tables 2.1 and 2.2 in Section 6).

3. Local Convergence. A detailed study of the other properties of the method requires the following four lemmas.

LEMMA 1. *Let $\psi \colon \mathbf{R} \to \mathbf{R}$, $q_m \colon \mathbf{R} \to \mathbf{R}$, be the functions defined by*

$$\psi(s) = \prod_{i=1}^{m-1} (s - \alpha_i), \qquad f(s) = \psi(s) q_m(s).$$

Then

$$f[\alpha_1, \alpha_2, \ldots, \alpha_{m-1}, x_m, x_{m+1}, \ldots, x_k] = q_m[x_m, x_{m+1}, \ldots, x_k].$$

Proof. Using the Leibniz formula for the divided differences [9, p. 5], since ψ is a monic polynomial of degree $m - 1$, one obtains

$$f[x_1, x_2, \ldots, x_k] = \sum_{l=1}^{m-1} \psi[x_1, x_2, \ldots, x_l] q_m[x_l, x_{l+1}, \ldots, x_k]$$

$$+ q_m[x_m, x_{m+1}, \ldots, x_k],$$

and the thesis follows by observing that $\psi[\alpha_1, \alpha_2, \ldots, \alpha_l] = 0$, $l = 1, 2, \ldots, m - 1$, by virtue of Proposition 1. $\square$

Interesting information on the local convergence toward the root $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_N)$ of (7) will also be obtained by considering the multiplicity of α and by studying some differential properties of the function F defined by (3) and (6). In this paper we will refer to the multiplicity definition given in [10].

Let $F^{(i)}(x)$ be the ith derivative of F at the point x and $\mathcal{N}(F^{(i)}(x))$ its null space. Furthermore, let $\mathcal{N}_i$ $(i = 1, 2, \ldots)$ be the spaces

$$(12) \quad \mathcal{N}_1 = \mathcal{N}(F'(\alpha)), \quad \mathcal{N}_i = \mathcal{N}_{i-1} \cap \mathcal{N}(F^{(i)}(\alpha)), \quad i = 2, 3, \ldots,$$

and d_i the dimension of $\mathcal{N}_i$:

$$(13) \quad d_i = \dim(\mathcal{N}_i), \quad i = 1, 2, \ldots.$$

One can then say [10] that the root α of the equation (7) is of multiplicity μ if

$$(14) \quad d_i \geqslant 1, \quad i = 1, 2, \ldots, \mu - 1, d_\mu = 0.$$

LEMMA 2. *Let M, $1 \leqslant M \leqslant N$, be the number of different roots of the equation $f(s) = 0$, and I_h $(h = 1, 2, \ldots, M)$ the set of the indices of those components of α which are equal to the hth of the above-considered different roots of $f(s) = 0$. Finally, let*

$$m_h = \max I_h, \quad h = 1, 2, \ldots, M.$$

Then the elements $J_{kl}(\alpha)$ of $J(\alpha)$ are such that

$$J_{kl}(\alpha) = f[\alpha_1, \ldots, \alpha_l, \alpha_l, \ldots, \alpha_k]$$

$$= \begin{cases} 0 & if \ l \in I_h, k < m_h, \\ f[\alpha_1, \ldots, \alpha_{m_h}, \alpha_{m_h}, \ldots, \alpha_k] \neq 0 & if \ l \in I_h, k \geqslant m_h. \end{cases}$$

Proof. Note, first of all, that for any fixed k the value of $J_{kl}(\alpha)$ is the same for all the $l \in I_h$ by virtue of the symmetry of the divided differences with respect to their arguments. From Proposition 1 it can then be deduced that $f[\alpha_1, \ldots, \alpha_l, \alpha_l, \ldots, \alpha_k]$ may be different from zero only if the argument α_l appears repeated a number of times greater than the multiplicity of α_l. This is possible only in the case $l \in I_h$, $k \geqslant m_h$, and it is easy to see, by using for example Lemma 1, that in this case $f[\alpha_1, \ldots, \alpha_l, \alpha_l, \ldots, \alpha_k]$ is different from zero. $\square$

LEMMA 3. *The Fréchet derivative $F'(\alpha)$ at α is nonsingular, and therefore α is a root of multiplicity $\mu = 1$, if, and only if, all the roots α_k are simple roots of the equation $f(s) = 0$. In general, the null space of $F'(\alpha)$ is given by*

$$(15) \quad \mathcal{N}_1 = \left\{ x \in \mathbf{R}^N \mid \sum_{k \in I_h} x_k = 0, h = 1, 2, \ldots, M \right\},$$

and, hence (see (12) *and* (13)),

$$d_1 = N - M.$$

Proof. This lemma is a consequence of the preceding one. In fact, $F'(\alpha)$ can be represented by $J(\alpha)$ and therefore the condition

$$(16) \qquad\qquad F'(\alpha)x = 0$$

is equivalent to $J(\alpha)x = 0$ and hence to

$$(17) \qquad\qquad \sum_{l=1}^{k} J_{kl}(\alpha)x_l = 0, \qquad k = 1, 2, \ldots, N.$$

Lemma 2 shows that, only if the α_k are all simple roots of the equation $f(s) = 0$, is $J(\alpha)$ nonsingular and one has $d_1 = 0$, $\mu = 1$ (see (14)). By imposing in succession the N conditions (17) and using Lemma 2 again, it is then easy to check that x satisfies the condition (16) if and only if it belongs to the set described on the right-hand side of (15). $\square$

LEMMA 4. *Let f be the polynomial* (1). *The null space of the second Fréchet derivative $F''(\alpha)$ at the point α is given by*

$$(18) \qquad \mathcal{N}(F''(\alpha)) = \left\{ x \in \mathbf{R}^N \mid x_1 = x_2 = \cdots = x_{N-1} = 0 \right\},$$

and, hence (see (12) *and* (13)), $d_2 = 0$.

Proof. The second derivative $F''(\xi)$ at $\xi \equiv (\xi_1, \xi_2, \ldots, \xi_N)$ can be represented by the array

$$(19) \qquad \left(\frac{\partial J(\xi)}{\partial \xi_1} \;\middle|\; \frac{\partial J(\xi)}{\partial \xi_2} \;\middle|\; \cdots \;\middle|\; \frac{\partial J(\xi)}{\partial \xi_N} \right).$$

Analogously, $F''(\xi)x$ may be represented by the $N \times N$ matrix, the jth column of which is $(\partial J(\xi)/\partial \xi_j)x$. If f is the polynomial (1) the two last rows of (19) are, for any ξ,

$$(20) \qquad \begin{pmatrix} 211 \cdots 10 & 121 \cdots 10 & \cdots\cdots & 111 \cdots 20 & 000 \cdots 00 \\ 000 \cdots 00 & 000 \cdots 00 & & 000 \cdots 00 & 000 \cdots 00 \end{pmatrix}$$

and, hence, the condition $F''(\xi)x = 0$ is satisfied if and only if x belongs to the set described on the right-hand side of (18). Indeed, since all the matrices $\partial J(\xi)/\partial \xi_j$ which form (19) are, as the matrix J, lower triangular, the condition $x_1 = x_2 = \cdots = x_{N-1} = 0$ implies (see also (20)) that $F''(\xi)x = 0$. But the converse is also true as $F''(\xi)x = 0$ implies (see (20) again)

$$\begin{pmatrix} 211 \cdots 1 \\ 121 \cdots 1 \\ \cdots\cdots\cdots \\ 111 \cdots 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_{N-1} \end{pmatrix} = 0$$

and the determinant of the above matrix is equal to N. $\square$

THEOREM 2. *The root α of the equation* (7) *is of multiplicity $\mu = 1$ if and only if the α_k are all simple roots of $f(s) = 0$. If f is the polynomial* (1), *one always has $\mu \leqslant 2$.*

Proof. The theorem is a direct consequence of the preceding definitions and Lemmas 3 and 4. $\square$

The next theorem holds not only for the particular Newton method defined by (3), (6) and (8) but, in general, for any Newton method applied to an equation (7) if F is smooth enough. It is contained in the results shown in [10] by L. B. Rall who also analyzes the behavior of the component $\zeta^{(n)}$ of the error decomposition considered in the theorem (see (21)), proving, essentially, that it also converges to zero, though only linearly.

THEOREM 3. *Let X be the orthogonal complement of $\mathcal{N}_1$ and*

$$(21) \qquad x^{(n)} = \alpha + \varepsilon^{(n)}, \qquad \varepsilon^{(n)} = \eta^{(n)} + \zeta^{(n)}, \qquad \eta^{(n)} \in X, \zeta^{(n)} \in \mathcal{N}_1.$$

Then,

$$(22) \qquad\qquad\qquad \eta^{(n+1)} = O\big(\|\varepsilon^{(n)}\|^2\big).$$

Proof. The assertion (22) follows essentially from the definition of Newton's method

$$F'(x^{(n)})\varepsilon^{(n+1)} = F'(x^{(n)})\varepsilon^{(n)} - F(x^{(n)})$$

by expanding F' and F:

$$\left[F'(\alpha) + F''(\alpha)\varepsilon^{(n)} + \frac{F'''(\alpha)}{2}(\varepsilon^{(n)})^2 + \cdots \right]\varepsilon^{(n+1)}$$

$$= \frac{F''(\alpha)}{2}(\varepsilon^{(n)})^2 + \frac{F'''(\alpha)}{3}(\varepsilon^{(n)})^3 + \cdots$$

and by observing that $F'(\alpha)$ as an operator on X is nonsingular. $\square$

The result stated in Theorem 3, in the case of the method defined by (3), (6) and (8), acquires an especially significant meaning, due to the particular form of the subspaces $\mathcal{N}_1$ (see Lemma 3) and X (see the following Lemma 5). This gives rise to one of the main properties of the method, stated below in Theorem 4.

LEMMA 5. *The orthogonal complement X of the subspace $\mathcal{N}_1$ stated in Lemma 3 is*

$$X = \big\{ x \in \mathbf{R}^N \mid x_i = x_k \, \forall i, k \in I_h; \, h = 1, 2, \ldots, M \big\}.$$

Proof. It is easy to check that the dimension of X is equal to M and that the scalar product between any two elements of the subspaces $\mathcal{N}_1$ and X is equal to zero. $\square$

THEOREM 4. *For each h, $1 \leqslant h \leqslant M$, let ν_h be the number of elements belonging to I_h (i.e., ν_h is the multiplicity of the hth of the M different roots of the equation $f(s) = 0$), and let $\hat{x}^{(n)}$ be the "nth average-vector"*

$$\hat{x}^{(n)} = \big(\hat{x}_1^{(n)}, \hat{x}_2^{(n)}, \ldots, \hat{x}_N^{(n)} \big)$$

having the components $\hat{x}_k^{(n)}$, $k \in I_h$, equal to each other and equal to the average $\sum_{l \in I_h} x_l^{(n)}/\nu_h$:

$$\hat{x}_k^{(n)} = \frac{1}{\nu_h} \sum_{l \in I_h} x_l^{(n)} \quad \forall k \in I_h \, (h = 1, 2, \ldots, M).$$

Then one has

(23) $$\hat{x}^{(n+1)} - \alpha = O\left(\|\varepsilon^{(n)}\|^2\right).$$

Proof. Consider $\hat{x}^{(n)}$ and α. They both belong to the subspace X defined in Lemma 5 and, by using this definition of X again and that of $\mathcal{N}_1$ stated in Lemma 3, it is easy to check that $\hat{x}^{(n)} - \alpha$ comes out to be exactly equal to the $\eta^{(n)}$ defined in (21):

$$\hat{x}^{(n)} - \alpha = \eta^{(n)}.$$

The proof is now concluded by using Theorem 3.   □

Remark 1. Theorem 4 shows that the method defined by (3), (6) and (8) is essentially a procedure for obtaining, by averaging the coordinates of $x^{(n)}$ which converge to the same root of $f(s) = 0$, a convergence of quadratic type to all the roots α_k of $f(s) = 0$, independently of their multiplicity.

Remark 2. By virtue of the triangular structure of F (see (6)) and J, Lemmas 3 and 5, as well as Theorem 4, can also be referred, more generally, to the space $\mathbf{R}^{\bar{N}}$ spanned by the coordinates $x_1, x_2, \ldots, x_{\bar{N}}$, where $1 \leqslant \bar{N} \leqslant N$. The arguments involved need only a simple generalization which can be briefly summed up by describing the generalized forms assumed by the subspaces $\mathcal{N}_1$ and X. These can be easily found, bearing in mind the arguments used in the proof of Lemma 2,

$$\mathcal{N}_1 = \left\{ x \in \mathbf{R}^{\bar{N}} \mid \sum_{k \in I_h} x_k = 0 \text{ if } m_h \leqslant \bar{N} \right\},$$

$$X = \left\{ x \in \mathbf{R}^{\bar{N}} \mid x_i = x_k \, \forall i, k \in I_h, \text{ if } m_h \leqslant \bar{N}; \, x_k = 0 \, \forall k \in I_h, \text{ if } m_h > \bar{N} \right\}.$$

This generalization allows one to improve the information given by Theorem 4 on the local convergence properties of $x^{(n)}$ near α from the description of the local convergence properties near $\bar{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_{\bar{N}})$ of the projection of $x^{(n)}$ into $\mathbf{R}^{\bar{N}}$. Such an analysis suggests the opportunity of reordering the coordinates of $x^{(n)}$ when multiple roots of $f(s) = 0$ occur by putting in the first positions those coordinates which converge to the simple roots. This results in a better use of the result of Theorem 4 and, consequently, in an appreciable improvement of the convergence properties.

Remark 3. When f is the polynomial (1), more detailed information can be obtained by using Proposition 3 in some particular cases. It is worth noting the following:

1) If the polynomial (1) has a unique root α_* of multiplicity N: $\alpha = (\alpha_*, \alpha_*, \ldots, \alpha_*)$, it follows from (9) that

$$\sum_{k=1}^{N} x_k^{(n)}/N = -a_1/N = \alpha_*, \qquad n = 1, 2, \ldots.$$

By using the notations introduced in Theorem 4, this means $\hat{x}^{(n)} - \alpha = 0$, $n = 1, 2, \ldots$, which is obviously stronger than (23). (See, for example, Table 3 in Section 6.)

2) If the polynomial (1) has a simple root α_1 and a root α_* of multiplicity $N - 1$: $\alpha = (\alpha_1, \alpha_*, \alpha_*, \ldots, \alpha_*)$, the coordinates of $\hat{x}^{(n)} - \alpha$, considered in Theorem 4, come

out to be the following:

$$\hat{x}_1^{(n)} - \alpha_1 = x_1^{(n)} - \alpha_1 = \varepsilon_1^{(n)}; \qquad \hat{x}_k^{(n)} - \alpha_k = \sum_{l=2}^{N} x_l^{(n)}/(N-1) - \alpha_*,$$

$$k = 2, 3, \ldots, N.$$

Then Proposition 3 allows one to improve the information given by (23). In fact, by considering that (9) can be written in the equivalent form

$$\sum_{l=1}^{N} x_l^{(n)} - \sum_{l=1}^{N} \alpha_l = 0,$$

one gets

$$\hat{x}_k^{(n)} - \alpha_* = \sum_{l=2}^{N} x_l^{(n)}/(N-1) - \alpha_* = -\varepsilon_1^{(n)}/(N-1), \qquad k = 2, 3, \ldots, N.$$

This means that the error relative to the average of the coordinates of $x^{(n)}$ which converge to the root α_* of multiplicity $N-1$ is $N-1$ times smaller than the error $\varepsilon_1^{(n)}$ of the coordinate $x_1^{(n)}$ that converges to the simple root α_1. Note (see Corollary 1.1) that $\varepsilon_1^{(n)}$ converges to zero as the error of the Newton method applied to the equation $f(s_1) = 0$ in a neighborhood of a simple root. (See, for example, Table 2.2 in Section 6.)

4. Global Convergence. In this section the method defined by (3), (6) and (8) is examined in the case of a polynomial (1) with real and simple roots.

Interesting information which completes that already stated in Corollary 1.1 and which concerns the behavior of the method near the subset $S_m \subset \mathbf{R}^N$: $S_m = \{ x \in \mathbf{R}^N | x_k = \alpha_k, k = 1, 2, \ldots, m-1 \}$, with $1 < m \leqslant N$, can be obtained from Theorems 5 and 6.

A study of the method near S_m means that one hypothesizes intermediate situations in which only a certain number $m-1$ of the roots α_k have as yet been accurately approximated by the first $m-1$ coordinates of $x^{(n)}$. This is a theoretical hypothesis which will allow us to show that the function Φ_m in (8) tends to the iteration function of the Newton method applied to the equation $q_m(s) = 0$ (see Lemma 1 for the definition of q_m), and, consequently, also $x_m^{(n)}$ must converge to another root α_m of $f(s) = 0$. Finally, a result due to Barna [11] (see Lemma 6 below) is applied to prove global convergence (Theorem 7).

Note, however, that this theoretical hypothesis actually represents a realistic model for the convergence of $x^{(n)}$ only in the case of particular starting points $x^{(0)}$ in which the accuracy of the first $m-1$ coordinates is much better than that of the remaining ones. In other words, $x_m^{(n)}$ does not normally have to wait for the preceding coordinates $x_1^{(n)}, x_2^{(n)}, \ldots, x_{m-1}^{(n)}$ to practically coincide with $\alpha_1, \alpha_2, \ldots, \alpha_{m-1}$ before starting to converge toward α_m.

To make the statement of Theorem 5 simpler it is convenient to slightly modify the notations, making evident the dependence of Φ on the polynomial and the dimension of the space in which the method defined by (3), (6) and (8) is applied. The function Φ in (8) will thus be denoted by $\Phi(x; f, N)$. Moreover, ξ denotes a point belonging to S_m and to the open set $\Omega \subset \mathbf{R}^N$ in which $\Phi(x; f, N)$ is defined.

Note that $\Omega = \mathbf{R}^N \setminus J_0$, where J_0 is the closed set of measure zero: $J_0 = \{x \in \mathbf{R}^N | \det J(x) = 0\}$, namely, the set in which at least one of the divided differences $f[x_1, x_2, \ldots, x_k, x_k], k = 1, 2, \ldots, N$, vanishes.

THEOREM 5. *Let ξ be any point belonging to $\Omega \cap S_m$ and q_m the function considered in Lemma 1. Then*

1) $\Phi_k(\xi; f, N) = \alpha_k, k = 1, 2, \ldots, m - 1$;

2) $\Phi_m(\xi; f, N) = x_m - q_m(x_m)/q_m'(x_m)$;

3) $\Phi_{m+k-1}(\xi; f, N) = \Phi_k((x_m, x_{m+1}, \ldots, x_N); q_m, N - m + 1), \quad k = 1, 2, \ldots, N - m + 1$.

Proof. The assertions 1), 2) and 3) easily follow from Theorem 1 and Lemma 1. □

Remark 4. Since the case of the polynomial (1) is being considered in this section, it is worth noting that Theorem 5 can be proved exactly with the same arguments even if f is only supposed to be, as in Theorem 1 and Lemma 1, a sufficiently smooth function, assuming that $\alpha_1, \alpha_2, \ldots, \alpha_{m-1}$ are simple roots of the equation $f(s) = 0$.

Now, let β_k $(k = 1, 2, \ldots, N - m)$ be the roots of $q_m'(s) = 0$, Ω_m the set $\Omega_m = \{s \in \mathbf{R} | s \neq \beta_k, k = 1, 2, \ldots, N - m\}$, and C any closed subset of Ω_m. Moreover, let $\Delta_m^{(n)}$ be the functions defined by

$$\Delta_m^{(n)}(s) = \Delta_m\left(x_1^{(n)}, x_2^{(n)}, \ldots, x_{m-1}^{(n)}, s\right).$$

The following theorem holds.

THEOREM 6. *Suppose that $x_k^{(n)}$ converges to α_k, $k = 1, 2, \ldots, m - 1$. Then for any fixed C, a constant $L = L(m, C)$ and an integer $\gamma = \gamma(m, C)$ exist such that*

(24) $$\left|\Delta_m^{(n)}(s) - q_m(s)/q_m'(s)\right| \leq L\rho_{m-1}^{(n)} \quad \forall s \in C, n > \gamma,$$

where

$$\rho_{m-1}^{(n)} = \left(\sum_{i=1}^{m-1} \left(\varepsilon_i^{(n)}\right)^2\right)^{1/2}.$$

Proof. Consider first the polynomial of degree $N - m, f[x_1^{(n)}, x_2^{(n)}, \ldots, x_{m-1}^{(n)}, s, s]$, which appears in the denominator of $\Delta_m^{(n)}(s)$ (see (10)), and note that, as a consequence of Lemma 1, one has $f[\alpha_1, \alpha_2, \ldots, \alpha_{m-1}, s, s] = q_m'(s)$. Therefore, the assumption $x_k^{(n)} \to \alpha_k$ $(k = 1, 2, \ldots, m - 1)$ can be used for asserting that a $\gamma = \gamma(m, C)$ exists such that the above denominator does not vanish in C for $n > \gamma$. Then, for any $n > \gamma$, $s \in C$, we can write, considering that, as a consequence of Lemma 1, one has $f[\alpha_1, \alpha_2, \ldots, \alpha_{m-1}, s] = q_m(s)$,

$$\Delta_m^{(n)}(s) - q_m(s)/q_m'(s) = \sum_{i=1}^{m-1} \frac{\partial \Delta_m}{\partial x_i}(\tau_1, \tau_2, \ldots, \tau_{m-1}, s)\varepsilon_i^{(n)},$$

where the τ_k depend on s and n and $0 \leq |\tau_k - \alpha_k| \leq |x_k^{(n)} - \alpha_k|$. Thus, the τ_k can be considered to belong to compact sets. On the other hand, one easily checks that the $\partial \Delta_m/\partial x_i$ (which are continuous functions), for fixed values of the first $m - 1$ variables, are the ratio of polynomials of the same degree $2(N - m)$ in the mth variable s. Thus the proof can be concluded by standard arguments. □

Theorem 6 essentially says that if the first $m - 1$ coordinates of $x^{(n)}$ converge, i.e., if $\rho_{m-1}^{(n)}$ converges to zero (note that it will then converge quadratically, see Theorem 4 and Remark 2), the sequence of the functions $\Delta_m^{(n)}$ converges uniformly to q_m/q_m' in any closed set C contained in Ω_m. The consideration of sets $C(\sigma)$ of the type $C(\sigma) = \{ s \in \mathbf{R} \mid |s - \beta_i| \geq \sigma, i = 1, 2, \ldots, N - m \}$, with $\sigma > 0$, is of particular interest. One can derive from Theorem 6 the following corollary which brings into evidence that $\Delta_m^{(n)}$ can be considered arbitrarily close to q_m/q_m' in an unbounded set arbitrarily close to Ω_m if n is large enough.

COROLLARY 6.1. *Let* $\{\sigma_i\}$, $\{\tau_i\}$ *be any prefixed sequences which decrease and converge to zero. Then, for each i, an integer $\bar{n}_i$ exists such that*

$$\left| \Delta_m^{(n)}(s) - q_m(s)/q_m'(s) \right| < \tau_i \quad \forall s \in C(\sigma_i), n > \bar{n}_i.$$

Finally, a result due to Barna [11] and reported in [12] by S. Smale, is stated in the following Lemma 6. An outline of the proof can be found in [12]; it is based on the fact that, on the hypothesis of the lemma, $T(s) = s - f(s)/f'(s)$ defines an Axiom A dynamical system on the one-dimensional projective space.

LEMMA 6. *If f is a polynomial with all roots real, Newton's method applied to the equation $f(s) = 0$ converges to a root with almost every starting point, the exceptional set E_1 of starting points being homeomorphic to the Cantor set.*

We are now in a position to prove

THEOREM 7. *Except for starting points $x^{(0)}$ belonging to a closed subset of $\mathbf{R}^N$ of measure zero, the equations (3), (6) and (8) define a sequence $\{x^{(n)}\}_{n=1,2,\ldots}$ which converges to a root of the equation (7).*

Proof. By virtue of Lemma 6 and Theorem 5 one does not have to consider the set of the $x^{(0)}$ which have the 1st coordinate $x_1^{(0)}$ belonging to E_1 or, analogously, which belong to a certain S_m and have the mth coordinate $x_m^{(0)}$ belonging to the corresponding exceptional set E_m of the Newton method applied to the equation $q_m(s) = 0$. Furthermore, the set of the $x^{(0)}$ such that $x^{(n)}$ will belong to J_0 for some value of n must obviously be left out. The union of these sets is closed and has measure zero and with respect to any one of the $x^{(0)}$ of the complementary set the following is true: the equations (3), (6) and (8) actually define a sequence $\{x^{(n)}\}_{n=1,2,\ldots}$ and the 1st coordinate $x_1^{(n)}$ converges to a root α_1 of $f(s) = 0$.

The assertion can then be derived by the results already stated in this section. They allow one to exclude the possibility that a certain number $m - 1$ of the first coordinates $x_1^{(n)}, x_2^{(n)}, \ldots, x_{m-1}^{(n)}$ of $x^{(n)}$ converge to $m - 1$ roots $\alpha_1, \alpha_2, \ldots, \alpha_{m-1}$ without the next coordinate $x_m^{(n)}$ converging to another root α_m.

In fact, since Lemma 6 can also be applied to the equation $q_m(s) = 0$, then $T_m(s) = s - q_m(s)/q_m'(s)$ also defines an Axiom A dynamical system in the projective line and, furthermore, $\Phi_m(x_1^{(n)}, x_2^{(n)}, \ldots, x_{m-1}^{(n)}, s)$ can ultimately be assumed arbitrarily close to $T_m(s)$ in a set $C(\sigma)$ arbitrarily close to the whole Ω_m. [It can also be observed, for completeness, that (see also (10)) σ can be assumed so small and n so large that if the sequence of the $x_m^{(n)}$ goes once out of $C(\sigma)$ then it will necessarily start to converge to the largest, or to the smallest, root of $q_m(s) = 0$.] Thus $x_m^{(n)}$ too

will converge with the same convergence properties of the Newton method applied to the equation $q_m(s) = 0$. $\square$

5. An Effective Formulation of the Method. The formulation of the method can be further simplified if the function f is the polynomial (1). In fact, by using (1) and Proposition 3, one easily obtains from (10) of Theorem 1 the following representation of the method:

$$x_k^{(n)} = x_k^{(n-1)} - \Delta_k, \qquad k = 1, 2, \ldots, N - 1,$$

$$x_N^{(n)} = -a_1 - \sum_{k=1}^{N-1} x_k^{(n)},$$

$$\Delta_k = \left(P_k - \sum_{l=1}^{k-1} \pi_{lk} \Delta_l \right) / \pi_{kk} \qquad \left(\sum_{1}^{0} = 0 \right),$$

where

$$P_k = \sum_{i=0}^{N-k+1} a_{N-k+1-i} p_{ik}, \qquad \pi_{lk} = \sum_{i=0}^{N-k} a_{N-k-i} q_{ik}^{[l]}$$

and

$$p_{hr} = p_{h,r-1} + x_r^{(n-1)} p_{h-1,r}, \qquad p_{0r} \equiv 1 \; \forall r \geqslant 1, p_{h0} \equiv 0 \; \forall h \geqslant 0,$$

$$q_{hr}^{[m]} = p_{hr} + x_m^{(n-1)} q_{h-1,r}^{[m]}, \qquad q_{0r}^{[m]} \equiv 1 \; \forall r \geqslant 1, m \geqslant 1.$$

The above formulae describe the method in its elementary or basic form. This basic form can then be implemented by some procedures which are suggested by the results stated in the preceding sections.

For example, it has been seen that it is advantageous to get into the first positions the coordinates of $x^{(n)}$ which converge to simple roots when multiple roots of the equation $f(s) = 0$ are present, and such an arrangement of the $x_k^{(n)}$ can be obtained by proper rearrangements during the execution. More generally, it seems reasonable to implement the basic form of the method by a procedure that forces the sequence $\{ x^{(n)} \}$ to converge to a root α of the equation (7) in which the roots α_k of $f(s) = 0$ are arranged in increasing order of their multiplicity.

Suitable tests can be inserted to recognize the convergence to multiple roots of $f(s) = 0$ and to determine the multiplicities during the execution. Those multiple roots will finally be computed by the averages described in Section 3.

These procedures, which have been employed in [8], can also be used when f is a sufficiently smooth function.

6. Numerical Results. Some numerical results are reported in this section. They illustrate properties of the method described in the preceding sections.

In Table 1 the results obtained in the case of the polynomial with simple roots $f(s) = (s^2 - 1)(s^2 - 4)(s^2 - 9)$ are listed.

In Tables 2.1 and 2.2 the results which one obtains with the polynomial $f(s) = (s + 2)(s - 1)^2$ are shown. In the first case the method is applied without rearrangements and the first two coordinates of $x^{(n)}$ converge to the double root, the third to the simple root. In the second case, the same starting point used in Table 2.1 has been rearranged to obtain the convergence of the first coordinate $x_1^{(n)}$ to the simple root.

TABLE 1

n	$x_1^{(n)}$	$x_2^{(n)}$	$x_3^{(n)}$	$x_4^{(n)}$	$x_5^{(n)}$	$x_6^{(n)}$
0	-35.0000000	36.0000000	-37.0000000	38.0000000	-39.0000000	40.0000000
1	-29.1889314	29.8646775	-25.0593088	25.4017966	- 2.2956706	1.2774367
2	-24.3508293	24.8063405	-15.6391913	15.6742371	200.1157521	-200.6063091
3	-20.3244251	20.6308473	- 7.5310374	7.4239548	99.1387122	-99.3380518
4	-16.9755082	17.1812846	2.1628441	-2.3510974	48.7213685	-48.7388915
5	-14.1924529	14.3304426	-17.8209756	17.5668480	23.4329040	-23.3167661
6	-11.8824995	11.9749196	-12.6342887	12.4734984	6.2875665	- 6.2191962
7	- 9.9686656	10.0304974	- 8.6986625	8.5907444	-8.1819410	8.2280274
8	- 8.3871868	8.4285091	- 5.6546319	5.5767515	0.5788853	- 0.5423272
9	- 7.0854039	7.1129838	- 3.1675530	3.1052242	-33.9307322	33.9654811
10	- 6.0200352	6.0384070	- 0.6984303	0.6380236	-16.7238216	16.7658567
11	- 5.1557902	5.1679849	7.3498595	-7.4390178	- 8.2871488	8.3641125
12	- 4.4643138	4.4723499	5.4791405	- 5.5272171	- 2.2094094	2.2494499
13	- 3.9234913	3.9287031	4.1280941	- 4.1546787	2.5778363	-2.5564636
14	- 3.5171780	3.5204332	3.1830384	- 3.1981566	- 0.1697105	0.1815735
15	- 3.2351724	3.2370203	2.5625662	- 2.5712744	7.2957225	-7.2888621
16	- 3.0711495	3.0719580	2.2045591	- 2.2092279	3.6379196	-3.6340593
17	- 3.0089560	3.0091273	2.0452085	- 2.0469271	1.9272740	-1.9257266
18	- 3.0001657	3.0001715	2.0032629	- 2.0034650	1.2196225	-1.2194262
19	- 3.0000001	3.0000001	2.0000177	- 2.0000195	1.0197329	-1.0197311
20	- 3.0000000	3.0000000	2.0000000	- 2.0000000	1.0001909	-1.0001909

TABLE 2.1

n	$x_1^{(n)}$	$x_2^{(n)}$	$x_3^{(n)}$	$(x_1^{(n)} + x_2^{(n)})/2$
0	1.1000000000000	0.9000000000000	-2.1000000000000	
1	1.0507936507937	0.9491516146689	-1.9999452654625	0.9999726327313
2	1.0256064999099	0.9743867459325	-1.9999932458424	0.9999966229212
3	1.0128572002852	0.9871418997619	-1.9999991000471	0.9999995500236
4	1.0064422877835	0.9935575963212	-1.9999998841047	0.9999999420524
5	1.0032245913762	0.9967753939100	-1.9999999852863	0.9999999926431
6	1.0016131607924	0.9983868373537	-1.9999999981462	0.9999999990731
7	1.0008067970788	0.9991932026886	-1.9999999997673	0.9999999998837
8	1.0004034527610	0.9995965472099	-1.9999999999709	0.9999999999855
9	1.0002017399423	0.9997982600541	-1.9999999999964	0.9999999999982
10	1.0001008733624	0.9998991266372	-1.9999999999995	0.9999999999996

TABLE 2.2

n	$x_1^{(n)}$	$x_2^{(n)}$	$x_3^{(n)}$	$(x_2^{(n)}+x_3^{(n)})/2$
0	-2.1000000000000	1.1000000000000	0.9000000000000	
1	-2.0060606060606	0.9121212121212	1.0939393939394	1.0030303030303
2	-2.0000243397834	0.9548089327005	1.0452154070829	1.0000121698917
3	-2.0000000003949	0.9773983897929	1.0226016106021	1.0000000001975
4	-2.0000000000000	0.9886991947977	1.0113008052023	1.0000000000000

TABLE 3

n	$x_1^{(n)}$	$x_2^{(n)}$	$x_3^{(n)}$	$(x_1^{(n)}+x_2^{(n)}+x_3^{(n)})/3$
0	3.0000000000000	4.0000000000000	-5.0000000000000	
1	2.0000000000000	1.5454545454545	-3.5454545454545	0.0000000000000
2	1.3333333333333	0.4096320346320	-1.7429653679654	0.0000000000000
3	0.8888888888889	-0.1127640061464	-0.7761248827425	0.0000000000000
4	0.5925925925926	-0.4282289352801	-0.1643636573125	0.0000000000000
5	0.3950617283951	0.0692161459124	-0.4642778743075	0.0000000000000
6	0.2633744855967	-0.0714512816179	-0.1919232039788	0.0000000000000
7	0.1755829903978	-0.2016201668740	0.0260371764762	0.0000000000000
8	0.1170553269319	-0.0815869673689	-0.0354683595629	0.0000000000000
9	0.0780368846212	0.0237275084445	-0.1017643930658	0.0000000000000
10	0.0520245897475	-0.0067710891325	-0.0452535006150	0.0000000000000

TABLE 4

n	$x_1^{(n)}$	$x_2^{(n)}$	$x_3^{(n)}$	$(x_2^{(n)}+x_3^{(n)})/2$
0	-6.000000000	7.000000000	4.000000000	5.500000000
1	-4.825607064	5.137746057	2.551330164	3.844538110
2	-3.898731033	3.716519296	1.846675309	2.781597303
3	-3.178412455	2.616048485	1.608298128	2.112173306
4	-2.638603528	1.724521777	1.619086890	1.671804333
5	-2.268836533	0.702397312	1.965783509	1.334090410
6	-2.067484332	0.894417323	1.402923552	1.148670437
7	-2.005485235	0.952999575	1.114329101	1.033664338
8	-2.000039772	0.974832332	1.029171758	1.002002045
9	-2.000000002	0.987501048	1.012547400	1.000024224
10	-2.000000000	0.993776559	1.006224079	1.000000319

TABLE 5

n	$x_1^{(n)}$	$x_2^{(n)}$	$x_3^{(n)}$	$(x_2^{(n)}+x_3^{(n)})/2$
0	.500000+01	.100000+01	.400000+01	.250000+01
1	.433931+01	.342099+00	.310249+01	.172229+01
2	.377532+01	-.197105+00	.219090+01	.996898+00
3	.334737+01	-.724932+00	.142242+01	.348744+00
4	.309431+01	.213345-01	.251823+00	.136579+00
5	.300883+01	-.514314-01	.528705-01	.719550-03
6	.300009+01	-.186657-01	.185708-01	-.474500-04
7	.300000+01	-.928988-02	.928986-02	-.100000-07
8	.300000+01	-.464497-02	.464497-02	.000000

Table 3 shows the results obtained in the case of the polynomial $f(s) = s^3$. The average of the three coordinates $x_1^{(n)}$, $x_2^{(n)}$ and $x_3^{(n)}$ is equal to the triple root for any value of n.

Table 4 concerns the case of a polynomial which also has complex roots. The polynomial is $f(s) = (s + 2)(s - 1)^2(s^2 + 2)$.

Finally, an example for the case in which f is not a polynomial is given in Table 5. The function is $f(s) = (\cosh s - 1)(s - 3)$.

Dipartimento di Metodi e Modelli Matematici
     per le Scienze Applicate
Facoltà di Ingegneria della 1ª Università
Via A. Scarpa 10
00161 Rome, Italy

Dipartimento di Matematica
Università di Bari
Via Nicolai 2
70121 Bari, Italy

1. M. W. GREEN, A. J. KORSAK & M. C. PEASE, "Simultaneous iteration towards all roots of a complex polynomial," *SIAM Rev.*, v. 18, 1976, pp. 501–502.

2. I. O. KERNER, "Ein Gesamtschrittverfahren zur Berechnung der Nullstellen von Polynomen," *Numer. Math.*, v. 8, 1966, pp. 290–294.

3. I. O. KERNER, "Algorithm 283," *Comm. ACM*, v. 9, 1966, p. 273.

4. E. DURAND, *Solution Numérique des Equations Algébriques*, Tome I, Masson, Paris, 1968.

5. L. W. EHRLICH, "A modified Newton method for polynomials," *Comm. ACM*, v. 10, 1967, pp. 107–109.

6. O. ABERTH, "Iteration methods for finding all zeros of a polynomial simultaneously," *Math. Comp.*, v. 27, 1973, pp. 339–344.

7. L. PASQUINI & D. TRIGIANTE, "Il metodo di continuazione e l'approssimazione simultanea degli zeri di un polinomio," Seminario dell'Ist. di Mat. Appl., Fac.di Ing., Univ. di Roma, 1981, pp. 128–146.

8. M. L. LOCASCIO, L. PASQUINI & D. TRIGIANTE, "Un polialgoritmo a convergenza rapida per la determinazione simultanea degli zeri reali di un polinomio e delle loro molteplicità," Monografie di Soft. Matem., N. 30, Pubbl. dell'IAC, 1984.

9. C. DE BOOR, *A Practical Guide to Splines*, Springer-Verlag, Berlin and New York, 1978.

10. L. B. RALL, "Convergence of the Newton process to multiple solutions," *Numer. Math.*, v. 9, 1966, pp. 23–37.

11. B. BARNA, "Über die Divergenzpunkte des Newtonschen Verfahrens zur Bestimmung von Wurzeln algebraischer Gleichungen. II," *Publ. Math. Debrecen*, v. 4, 1956, pp. 384–397.

12. S. SMALE, "The fundamental theorem of algebra and complexity theory," *Bull. Amer. Math. Soc* (*N.S.*), v. 4, 1981, pp. 1–36.