A Convergent Algorithm for Solving Polynomial Equations

J. B. MOORE

The University of Santa Clara, Santa Clara, California*

ABSTRACT. The method of steepest descent is applied in a convergent procedure to determine the zeros of polynomials having either real or complex coefficients. By expressing the polynomials in terms of the Šiljak functions, the methods are readily programmed on a digital computer. The significance of the procedures is that their application is straightforward, and not only is convergence rapid in the region of a zero but convergence is guaranteed independent of the initial values.

Introduction

In this paper the application of the steepest descent approach in a straightforward convergent method is considered for the rapid solution of polynomial equations. The method is presented as being a more efficient and thus a more reasonable alternative than presently available methods [1–3].

In many practical problems involving the solution of algebraic equations, there is sufficient a priori knowledge of the roots to permit the application of the readily applied, rapidly converging synthetic division methods. The limitation of Newton-Raphson's method, Lin's method, Bairstow's method, and other synthetic division methods is that convergence is dependent on the initial conditions.

For the problems in which the a priori information about the location of the roots is inadequate, the Lehmer-Schur method, Graeffe's root-squaring method, Bernoulli's method, or the methods of Lance may be used and convergence is guaranteed. These methods are by no means simple and straightforward, and are not rapidly convergent, and thus for efficiency are usually used to calculate only the approximate root locations so that one of the more rapid synthetic division methods may then be applied. The need for a straightforward method which always converges and which has a rapid convergence in the region of the root is thus indicated.

Kokotović and Šiljak [1] have developed the steepest descent approach used by Levine and Meissinger [2] and Lance [3] in conjunction with Mitrovic's method for the automatic analog solution of polynomials. The limitation of the analog approach is that a considerable search may be required to separate the roots and find appropriate initial conditions.

In this paper the steepest descent approach of Kokotović and Šiljak [1] is extended and adapted for the efficient digital computation of the zeros of analytic functions involving polynomials. The steepest descent approach is used to minimize a non-negative function, the minimal values of which are zero and correspond to the zeros of the analytic function under investigation. The magnitude of the increment in the

* Department of Electrical Engineering. The research reported herein was supported by the National Aeronautics and Space Administration under Contract No. NAS2-2609.

direction of steepest descent for each iteration is calculated from equations similar to those used in the rapidly converging Newton-Bairstow method.

In particular, the rapid factorization of polynomials having either real or complex coefficients is achieved. By expressing polynomials in terms of Šiljak functions, the methods are readily programmed on a digital computer.

Review of Steepest Descent Equations

Consider the equation

$$f(z) = u + iv = 0, \tag{1}$$

where $f(z)$ is an entire function of the complex variable z, where

$$z = x + iy. \tag{2}$$

The differential equations which may be used to determine the zeros of $f(z)$ by the method of steepest descent are

$$\frac{\partial x}{\partial t} = -h\frac{\partial F}{\partial x} \quad \text{and} \quad \frac{\partial y}{\partial t} = -h\frac{\partial F}{\partial y}, \tag{3}$$

where h is a positive constant and $F = F(x, y)$ is a function satisfying the following conditions:

$$\left.\begin{array}{l} \text{(1) } F \text{ is nonnegative;} \\ \text{(2) the derivatives } \partial F/\partial x \text{ and } \partial F/\partial y \text{ exist;} \\ \text{(3) the zeros of } F \text{ are located at the roots of } f(z) = 0; \\ \text{(4) these zeros are the only minima of } F. \end{array}\right\} \tag{4}$$

It may be shown that the time derivative of F is always negative, i.e.,

$$\frac{dF}{dt} = \frac{\partial F}{\partial x}\cdot\frac{dx}{dt} + \frac{\partial F}{\partial y}\cdot\frac{dy}{dt}. \tag{5}$$

Substituting eqs. (3) gives

$$\frac{dF}{dt} = -h\left[\left(\frac{\partial F}{\partial x}\right)^2 + \left(\frac{\partial F}{\partial y}\right)^2\right] < 0. \tag{6}$$

This property of dF/dt, together with the previously listed properties of F, indicates that for any process simulating eqs. (3), F is a Liapunov function and the process will always converge to a zero of $f(z)$ independent of initial conditions.

Algorithm

For the numerical evaluation of the zeros of $f(z)$, the difference equations corresponding to the differential eqs. (3) are considered, i.e.,

$$\Delta x = -h\frac{\partial F}{\partial x} \quad \text{and} \quad \Delta y = -h\frac{\partial F}{\partial y}. \tag{7}$$

The direction of the steepest descent path given by the ratio of Δx and Δy is independent of h, i.e.,

$$\frac{\Delta x}{\Delta y} = \frac{\partial F}{\partial x} \Big/ \frac{\partial F}{\partial y}. \tag{8}$$

The magnitude of the increment in the direction given by Δx and Δy is determined by h. The value of h should be chosen to minimize the function F along the direction of steepest descent. Consideration is now given to the calculation of a suitable h.

In the neighborhood of a zero of $f(z)$, the higher order terms in a series expansion for u and v are negligible, and therefore the following approximations are valid:

$$-u \simeq \frac{\partial u}{\partial x} \Delta x + \frac{\partial u}{\partial y} \Delta y, \qquad -v \simeq \frac{\partial v}{\partial x} \Delta x + \frac{\partial v}{\partial y} \Delta y, \tag{9}$$

where Δx and Δy are the distances in the x and y directions from a point in the neighborhood of the zero of $f(z)$ to the zero of $f(z)$. Using the Cauchy-Riemann equations and solving eqs. (9) for Δx and Δy yields

$$\Delta x \simeq \frac{-u \dfrac{\partial u}{\partial x} - v \dfrac{\partial v}{\partial x}}{\left(\dfrac{\partial u}{\partial x}\right)^2 + \left(\dfrac{\partial v}{\partial x}\right)^2}, \qquad \Delta y \simeq \frac{u \dfrac{\partial v}{\partial x} - v \dfrac{\partial u}{\partial x}}{\left(\dfrac{\partial u}{\partial x}\right)^2 + \left(\dfrac{\partial v}{\partial x}\right)^2}. \tag{10}$$

If the values of u, v, $\partial u/\partial x$, $\partial v/\partial x$ can be calculated, then successive application of eqs. (10) in the region of a zero of $f(z)$ gives rapid convergence to the zero location for any required accuracy. It is to be noted that eqs. (10) are similar to those used in the quadratically convergent Newton-Bairstow's method. The difference lies in the fact that eqs. (10) are of the form of the steepest descent eqs. (7), where

$$F = u^2 + v^2, \tag{11}$$

$$h = \frac{0.5}{\left(\dfrac{\partial u}{\partial x}\right)^2 + \left(\dfrac{\partial v}{\partial x}\right)^2} > 0. \tag{12}$$

This is evident, as eqs. (10) may be derived from eqs. (7) and (11), and the function u^2+v^2 satisfies the conditions (4). (It is immediately apparent that the first three conditions (4) are satisfied, and by realizing that u^2+v^2 is the square of the modulus of $f(z)$, one may apply the maximum modulus theorem to show that the fourth condition is satisfied.) It is concluded that the application of eqs. (7) and (11) or the equivalent eqs. (10) indicates the direction of the steepest descent of F. (An exception to this exists when $y = 0$; for this case, the direction indicated is along the y-axis.) For regions other than the neighborhoods of the zeros of $f(z)$, the application of eqs. (10) will give a poor estimate of the increment for which F is minimized in the direction of steepest descent of F. However, even if more than one calculation is necessary, an increment can be found in the direction of steepest descent for which the value of F is reduced, and so convergence may be guaranteed.

The following algorithm, which arises out of the above theoretical developments, is the basis for a convergent method for finding zeros of analytic functions.

1. Scale to have some roots within the unit circle.
2. Choose initial approximations other than on the real axis, i.e., x_0, y_0.

3. Compute u, v, $\partial u/\partial x$, $\partial v/\partial x$, and F.
4. Compute the new approximations from eqs. (10), i.e., $x_1 = x_0 + \Delta x$, $y_1 = y_0 + \Delta x$.
5. Take x_1 as x_0 and y_1 as y_0 and repeat steps 3, 4, and 5, including the following step 6, until convergence occurs.
6. If the value of F calculated is greater than the value for the preceding iteration, reduce the increments Δx and Δy used previously until the value of F is smaller than for the preceding iteration. The increment may be reduced by a factor of say one quarter each trial.

Solution of Algebraic Equations

Consider the algebraic equation

$$f(z) = \sum_{k=0}^{n} (a_k + ib_k)z^k = 0, \qquad a_n + ib_n \neq 0, \tag{13}$$

where a_k and b_k are real. As in eq. (1), $f(z)$ is an entire function of the complex variable $z = x + iy$. The real and imaginary parts of $f(z)$ and their partial derivatives may be conveniently calculated in terms of the Šiljak functions X_k and Y_k.

The functions $X_k(x, y)$ and $Y_k(x, y)$ are defined from

$$z^k = X_k + iY_k , \tag{14}$$

and may be calculated using the recurrence relationships[1]

$$\begin{aligned} X_{k+2} - 2xX_{k+1} + (x^2 + y^2)X_k &= 0, \\ Y_{k+2} - 2xY_{k+1} + (x^2 + y^2)Y_k &= 0, \end{aligned} \tag{15}$$

where

$$X_0 = 1, \quad X_1 = x, \quad Y_0 = 0, \quad Y_1 = y.$$

The partial derivatives of X_k and Y_k are simply calculated from the partial differentiation of eq. (14) with respect to x, i.e.,

$$\begin{aligned} \frac{\partial^n X_k}{\partial x^n} &= k(k-1) \cdots (k-n+1)X_{k-n} , \\ \frac{\partial^n Y_k}{\partial x^n} &= k(k-1) \cdots (k-n+1)Y_{k-n} . \end{aligned} \tag{16}$$

Using eqs. (14) and (16), one may calculate the values of the real part u and the imaginary part v of $f(z)$ from eq. (13) and their partial derivatives as

$$\begin{aligned} u &= \sum_{k=0}^{n} (a_k X_k - b_x Y_k), \qquad v = \sum_{k=0}^{n} (a_k Y_k + b_k X_k), \\ \frac{\partial u}{\partial x} &= \sum_{k=0}^{n} k(a_k X_{k-1} - b_k Y_{k-1}), \qquad \frac{\partial v}{\partial x} = \sum_{k=0}^{n} k(a_k Y_{k-1} + b_k X_{k-1}), \\ \frac{\partial^2 u}{\partial x^2} &= \sum_{k=0}^{n} k(k-1)(a_k X_{k-2} - b_k Y_{k-2}), \\ \frac{\partial^2 v}{\partial x^2} &= \sum_{k=0}^{n} k(k-1)(a_k Y_{k-2} + b_k X_{k-2}). \end{aligned} \tag{17}$$

[1] These are derived from a substitution for z^{k+2}, z^{k+1}, and z^k using eq. (14) into the equation $z^k(z^2 - 2xz + (x^2 + y^2)) = 0$; the real and imaginary parts are equated to zero.

For the case when the algebraic equation coefficients are real, i.e., $b_k = 0$ for $k = 0 \rightarrow n$, eqs. (17) are simplified. (*Note:* The values of u, v, $(\partial u)/(\partial x)$, and $(\partial v)/(\partial x)$ may be calculated using alternate functions [1] or from the remainder coefficients calculated from a synthetic division of $f(z)$ and $f'(z)$ with factors corresponding to the particular x and y of interest.)

Using the steepest descent methods of the previous section together with eqs. (15) and (17), one can achieve the rapid factorization of a polynomial on a digital computer. Once a zero of $f(z)$ is determined, by use of synthetic division a reduced polynomial is found, and this may be used in the evaluation of the next zero. As in other procedures, a suitable first approximation may be read into the machine or calculated from either the first three or the last three coefficients of the polynomial, or, if frequency scaling is used, at least some of the zeros will be in the vicinity of the origin and a convenient first approximation may be $x = 0$, $y = 1$.

The computer time for each iteration, when Newton's method is used, is about three quarters that for the above method. However, as mentioned previously, when one uses the methods of this paper convergence is guaranteed, and no a priori information about the root location is required. For the seventh-order equation of [1], for example, all the roots were found in seventeen iterations.

Transcendental equations involving polynomials have also been solved by use of the approach given in this paper, and thus the method has a more general application.

Conclusions

A straightforward numerical method has been presented for the solution of polynomials which is rapidly convergent in the region of a zero, and convergence is guaranteed independent of the initial values. It has been shown that by expressing the polynomials in terms of the Šiljak functions the method is convenient to implement on a digital computer and is applicable to the solution of algebraic equations having either real or complex coefficients or to transcendental equations involving polynomials. It is concluded that in terms of simplicity and convergence properties the approach is more efficient than presently available methods.

REFERENCES

1. Kokotović, P., and Šiljak, D. Automatic analog solution of algebraic equations and plotting of root loci by generalized Mitrovic method. *IEEE Trans. Appl. and Ind. 83* (1964), 324–328.
2. Levine, L., and Meissinger, H. F. An automatic analog computer method for solving polynomials and finding root loci. *National Convention Record IRE*, Pt. 4 (March 1957), 45–62.
3. Lance, G. N. Solutions of algebraic and transcendental equations on an automatic digital computer. *J. ACM 6* (1959), 97–101.

Received March, 1966; Revised May, 1966