J. F. TRAUB, Editor

A Generalized Bairstow Algorithm

G. H. Golub* and T. N. Robertson
Stanford University,† Stanford, California
and
Occidental College Los Angeles, California

The Bairstow algorithm is generalized to the case of a polynomial which is itself a linear combination of polynomials satisfying a three-term recursion. Convergence properties of the method are derived.

1. Introduction: The Bairstow Process

The basic idea for finding the roots of real polynomials by finding a quadratic factor makes use of the following identity

$$(a_0x^n + a_1x^{n-1} + \cdots + a_n) \equiv (x^2 - \alpha x - \beta)(b_0x^{n-2} + b_1x^{n-3} + \cdots + b_{n-2}) + Ax + B. \quad (1)$$

Equating coefficients gives (with $b_{-1} = b_{-2} = 0$):

$$b_k = a_k + \alpha b_{k-1} + \beta b_{k-2} \quad (2)$$

for $k = 0, 1, \cdots, n-2$ and

$$\left.\begin{aligned} A &= a_{n-1} + \alpha b_{n-2} + \beta b_{n-3} \\ B &= a_n + \beta b_{n-2} \end{aligned}\right\}. \quad (3)$$

Beginning with arbitrary α_0 and β_0, (2) and (3) can be used to define an iterative process for getting a quadratic factor. At the ith step α_i and β_i are used in (2) to provide coefficients b_k; after which, (3) is solved for α_{i+1} and β_{i+1} with A and B set to zero.

This is usually known as Lin's method [2], which was extended and studied by Friedman [3] and Luke and Ufford [4]. The convergence properties have not been fully established, but the method is often slowly convergent.

The Bairstow method [1] consists of solving the system

$$\left.\begin{aligned} A &\equiv A(\alpha, \beta) = 0 \\ B &\equiv B(\alpha, \beta) = 0 \end{aligned}\right\}$$

by Newton's process of successive approximations. The theorem of Kantorovich [5] gives conditions on A and B and on the starting values α_0, β_0 which ensure convergence. The verification of these conditions is not computationally feasible, but the method is usually quadratically convergent. More precisely, if we formulate the algorithm as in [6], viz., replacing α_n and β_n by $\alpha_n+\delta$ and $\beta_n+\epsilon$, where δ and ϵ satisfy

$$\left.\begin{aligned} A + \delta A_\alpha + \epsilon A_\beta &= 0 \\ B + \delta B_\alpha + \epsilon B_\beta &= 0 \end{aligned}\right\}$$

(subscripts denote partial differentiation), then the iteration procedure is quadratically convergent if the sequences $\{\alpha_n\}$ and $\{\beta_n\}$ have limits s and t, respectively, and further

$$D = \begin{vmatrix} A_\alpha & A_\beta \\ B_\alpha & B_\beta \end{vmatrix} \neq 0 \quad \text{at} \quad (s, t)$$

This criterion is applied in the generalization which follows.

2. Generalized Bairstow Algorithm

Consider (cf., identity (1))

$$a_0P_n + a_1P_{n-1} + \cdots + a_nP_0 \equiv (P_2 - \alpha P_1 - \beta P_0)(b_0P_{n-2} + b_1P_{n-3} + \cdots + b_{n-2}P_0) + AP_1 + BP_0 \quad (4)$$

where $P_n(x)$ are nth degree polynomials satisfying a three-term recursion

$$P_{n+1} = (c_nx + d_n)P_n + e_nP_{n-1}, \text{ with } P_{-1} = 0, \quad P_0 = 1.$$

Thus we can write $P_1P_k = l_{k+1}P_{k+1} + m_kP_k + r_{k-1}P_{k-1}$ and $P_2P_k = s_{k+2}P_{k+2} + t_{k+1}P_{k+1} + u_kP_k + v_{k-1}P_{k-1} + w_{k-2}P_{k-2}$ for appropriate l, m, r, s, etc., so that equating coefficients now gives

$$b_k = \frac{1}{s_{n-k}}\{a_k + (\alpha l_{n-k} - t_{n-k})b_{k-1} + (\beta + \alpha m_{n-k} - u_{n-k})b_{k-2} + (\alpha r_{n-k} - v_{n-k})b_{k-3} - w_{n-k}\, b_{n-4}\} \quad (4a)$$

(with $b_{-4} = b_{-3} = b_{-2} = b_{-1} = 0$) for $k = 0, 1, \cdots, n-2$, and

$$A = a_{n-1} + \alpha b_{n-2} + (\beta + \alpha m_1 - r_1)b_{n-3} + (\alpha r_1 - v_1)b_{n-4} - w_1b_{n-5} \quad (4b)$$
$$B = a_n + \beta b_{n-2} + \alpha r_0b_{n-3} - w_0b_{n-4}.$$

---

* The work of the first author was supported in part by the National Science Foundation and in part by the Office of Naval Research

† Computer Science Department

Equations (4a) and (4b) can be used to compute the factor $(P_2 - \alpha P_1 - \beta P_0)$ by a natural extension of Bairstow's process. Begin by choosing starting values α_0, β_0. Having computed α_i and β_i, (4a) will provide values for the quantities b_k and their partial derivatives with respect to α and β. These in turn are used in connection with (4b) to provide values for $A, B, A_\alpha, A_\beta, B_\alpha, B_\beta$. Newton's process will yield values for δ and ϵ, from which α_{i+1}, β_{i+1} follow.

To establish the convergence properties, write (4) as follows: $P(x) \equiv (P_2 - \alpha P_1 - \beta P_0)Q(x) + AP_1 + BP_0$.

THEOREM. *If $P_2 - sP_1 - tP_0$ is an exact factor with roots r_1 and r_2, then the convergence $\alpha \to s$ and $\beta \to t$ is quadratic if $Q(r_1)$ and $Q(r_2)$ are nonzero.*

PROOF. Differentiation of (4) w.r.t. α gives

$$0 = -P_1Q(x) + (P_2 - \alpha P_1 - \beta P_0)Q_\alpha(x) + A_\alpha P_1 + B_\alpha P_0, \tag{5}$$

so that evaluation at $\alpha = s$, $\beta = t$, and $x = r_1$ gives

$$A_\alpha P_1(r_1) + B_\alpha = P_1(r_1)Q(r_1).$$

Similarly, $A_\beta P_1(r_1) + B_\beta = Q(r_1)$. Then

$$D = \begin{vmatrix} A_\alpha & A_\beta \\ B_\alpha & B_\beta \end{vmatrix} = Q(r_1)(A_\alpha - A_\beta\, P_1(r_1)).$$

So (i) if $r_1 \neq r_2$, evaluation of (5) at $x = r_2$ gives two more equations which solve to yield

$$A_\alpha = \frac{P(r_1)Q(r_1) - P(r_2)Q(r_2)}{P_1(r_1) - P_1(r_2)},$$

$$A_\beta = \frac{Q(r_1) - Q(r_2)}{P_1(r_1) - P_1(r_2)}$$

and thus $D = Q(r_1)Q(r_2)$; (ii) if $r_1 = r_2 = r$, differentiation of (5) w.r.t. x gives

$$0 = -P_1'Q(x) - P_1Q'(x) + (P_2 - \alpha P_1 - \beta P_0)'Q_\alpha(x) + (P_2 - \alpha P_1 - \beta P_0)Q_\alpha'(x) + A_\alpha P_1',$$

and evaluation at $\alpha = s$, $\beta = t$ and $x = r$ gives

$$A_\alpha = Q(r) + P_1(r)Q'(r)/a_1.$$

Similarly, $A_\beta = Q'(r)/a_1$, so $D = Q^2(r)$.

Thus in either case $D \neq 0$ if $Q(r_1)$ and $Q(r_2)$ are nonzero.

Proof (i) above is the generalization of the result given by Henrici [6] which ensures convergence of the Bairstow algorithm to a quadratic factor of a polynomial if its roots have multiplicity one. We have shown in (ii) that a root of multiplicity two can be extracted, and the procedure remains quadratically convergent.

It is interesting to experiment with the classical Bairstow method upon polynomials having repeated roots. It has been observed, for example, that if r_1 has multiplicity two and r_2 is any other root, even if an initial approximation closer to $(x - r_1)(x - r_2)$ than to $(x - r_1)^2$ is taken, the scheme "prefers" to converge to $(x - r_1)^2$, avoiding $Q(r_1) = 0$. Similarly, in the generalization we can say that the extraction of "quadratic factors" $P_2 - sP_1 - t$ can be accompanied with quadratic convergence if the roots of the linear combination $\sum_{k=0}^{n} a_kP_k$ have multiplicity two or less.

3. Applications

(i) ORTHOGONAL POLYNOMIALS. Orthogonal polynomials are important in curve fitting, cf. [7]; they also play an important role in Gaussian quadrature. The method we have presented is applicable to finding the zeros of linear combinations of orthogonal polynomials, since such polynomials satisfy a three-term recurrence relationship.

Programming experiments using an IBM 1620 tested the method on combinations of the form $\sum_{k=0}^{n} a_kT_k$, where $T_k(x)$ is the kth degree Chebyshev polynomial, and confirmed the properties of convergence to the "quadratic factors" $T_2 - sT_1 - t$. The recursion formulas

$$T_1T_k = \tfrac{1}{2}(T_{k+1} + T_{k-1}),\ T_2T_k = \tfrac{1}{2}(T_{k+2} + T_{k-2})$$

for Chebyshev polynomials yield

$$b_k = 2a_k + \alpha b_{k-1} + 2\beta b_{k-2} + \alpha b_{k-3} - b_{k-4}$$
$$\text{with } b_{-1} = b_{-2} = b_{-3} = b_{-4} = 0$$
$$A = a_{n-1} + \alpha b_{n-2} + \left(\beta - \frac{1}{2}\right)b_{n-3} + \frac{\alpha}{2}\, b_{n-4} - \frac{1}{2}\, b_{n-5} \tag{6}$$
$$B = a_n + \beta b_{n-2} + \frac{\alpha}{2}\, b_{n-3} - \frac{1}{2}\, b_{n-4}.$$

The formulas for $\partial b_k/\partial\alpha$, $\partial b_k/\partial\beta$ follow easily from (6) and apply to provide A_α, A_β, B_α, B_β.

(ii) EIGENVALUE PROBLEMS. For the tridiagonal matrix

$$A = \begin{pmatrix} d_1 & u_1 & & & O \\ l_2 & d_2 & \cdot & & \\ & \cdot & \cdot & \cdot & \\ & & \cdot & \cdot & u_{n-1} \\ O & & & l_n & d_n \end{pmatrix}$$

the characteristic polynomial

$$\det(xI - A) = P_n(x)$$

satisfies

$$P_0 = 1, \quad P_1 = x - d_1 \quad \text{and}$$
$$P_{k+1} = (x - d_{k+1})P_k - u_kl_{k+1}P_{k-1},$$

so that the method applies to the eigenvalue problem for arbitrary tridiagonal matrices.

Programming tests on symmetric tridiagonal matrices with known eigenvalues gave good convergence and accuracy. The eigenvalues were found in pairs, each pair being deflated out before the subsequent pair was obtained.

There have been a number of algorithms proposed to reduce an arbitrary matrix to tridiagonal form; references to these algorithms are given in [8]. The method presented, used in conjunction with such a routine, can be used for solving the complete eigenvalue problem. In particular, when approximations to the eigenvalues are known this generalization of the Bairstow process is an efficient means of obtaining final values. Wilkinson [9, p. 450] has given a similar method without proof of convergence, however, and Handscomb [10] has given a generalization of the Bairstow method for Hessenberg matrices.

(iii) Symmetric Polynomials. Consider a $2n$th degree polynomial of the form

$$P_{2n}(z) = a_0 z^{2n} + a_1 z^{2n-1} + \cdots + a_n z^n + \cdots + a_0$$
$$= a_0(z^{2n} + 1) + a_1(z^{2n-1} + z) + \cdots + a_n z^n.$$

It is easy to see that if $P_{2n}(z^*) = 0$, then $P_{2n}(1/z^*) = 0$. Now $P_{2n}(z) = 0$ when

$$W(z) = \frac{P_{2n}(z)}{z^n} = a_0\left(\frac{z^n + z^{-n}}{2}\right) + a_1\left(\frac{z^{n-1} + z^{-(n-1)}}{2}\right) + \cdots + \frac{a_n}{2} = 0.$$

Let us write

$$R_k(z) = \frac{z^k + z^{-k}}{2},$$

so that

$$W(z) = a_0 R_n(z) + a_1 R_{n-1}(z) + \cdots + \frac{a_n}{2} R_0(z).$$

Note that

$$R_{k+1}(z) = 2R_1(z)R_k(z) - R_{k-1}(z).$$

It is easy to see that the method presented here is applicable even though $R_k(z)$ is not a kth degree polynomial.

Acknowledgment. The authors are pleased to acknowledge the helpful comments of Mr. Richard Bartels and wish to thank Dr. J. Traub for calling their attention to Wilkinson's work in this area.

Received January, 1967; revised February, 1967

REFERENCES

1. Bairstow, L. Investigations relating to the stability of the aeroplane. R and M 154 of Advisory Committee for Aeronautics, 1914.
2. Lin, Shih-nge. A method for finding roots of algebraic equations. *J. Math. Phys. 22* (1943), 60–77.
3. Friedman, B. Note on approximating complex zeros of a polynomial. *Comm. Pure Appl. Math. 2* (1949), 195–208.
4. Luke, Y. L., and Ufford, D. On the roots of algebraic equations. *J. Math. Phys. 30* (1951), 94–101.
5. Kantorovich, L. V. Functional analysis and applied mathematics. Nat. Bur. Standards Report 1509, Transl. by C. D. Benster and edited by G. E. Forsythe.
6. Henrici, Peter. *Elements of Numerical Analysis.* John Wiley, New York, 1964.
7. Forsythe, G. E. Generation and the use of orthogonal polynomials for data fitting with a digital computer. *J. Soc. Indust. Appl. Math. 5* (1957), 74–88.
8. Waltmann, W. L., and Lambert, R. J. T-algorithm for tridiagonalization. *J. Indust. Appl. Math. 13* (1965), 1069–1078.
9. Wilkinson, J. H. *The Algebraic Eigenvalue Problem.* Clarendon Press, Oxford, 1965.
10. Handscomb, D. C. Computation of the latent roots of a Hessenberg matrix by Bairstow's method. *Comput. J. 5* (1962), 139–141.

Algorithms

J. G. HERRIOT, Editor

ALGORITHM 303
AN ADAPTIVE QUADRATURE PROCEDURE WITH RANDOM PANEL SIZES [D1]

L. J. Gallaher (Recd. 8 Nov. 1966 and 1 Feb. 1967)
Georgia Institute of Technology, Engineering Experiment Station, Atlanta, Ga.

real procedure *Integral*(*a*, *x*, *b*, *fx*, *random number*, *error*);
value *a*, *b*, *error*;
real *a*, *x*, *b*, *fx*, *error*;
real procedure *random number*;
comment This procedure approximates the quadrature of the function *fx* on the interval $a < x < b$ to an estimated accuracy of *error*. It does this by sampling the function *fx* at appropriate points until the estimated error is less than *error*. The points to be sampled are determined by a combination of random sampling and of estimating what regions are more in need of sampling, this need being determined by the samples already taken. This process goes under the name "importance sampling" in nuclear reactor literature [for example, see J. M. Hammersley and D. C. Handscomb, *Monte Carlo Methods*, John Wiley, Inc., 1964, p. 57]. The form of importance sampling used here is based on estimates of the error contributed to the quadrature by the second derivative. That is, random samples of the average value of the second derivative of *fx* in a region are taken and used to decide if more samples are needed in that region.

Randomness here is achieved through the real procedure *random number*. This procedure is not given explicitly here but can be any random number generator available, provided only that the numbers given are distributed on the interval 0 to 1. The random numbers given need not be of particularly high quality (i.e., need not have low correlation). Further the random number generator need not be passed as a parameter but could be either global or local to the procedure *Integral*.

This procedure is meant to be used for low-accuracy estimates of quadratures, especially large dimensional multiple integrals for which the high-accuracy methods would be too time consuming and expensive. It can achieve high accuracies but not as efficiently as algorithms already in the literature. The general form of this algorithm is similar to Algorithm 145 [W. M. McKeeman, Adaptive Numerical Integration by Simpson's Rule, *Comm. ACM 5* (Dec. 1962), 604] (and others) except that in subdividing the region of integration the panel sizes are determined in part by the random-number generator.

This quadrature procedure has been found particularly effective in integrating ill-behaved functions of the following type.

A. Functions having singularities on the boundary of the region of integration. Such integrals as

$$\int_0^1 x^{-1/2}\,dx,$$

$$\int_0^2 dx \int_0^{\sqrt{1-(1-x)^2}} dy(x^2+y^2)^{-1/2},$$