ON THE SOLUTION OF ALGEBRAIC EQUATIONS

STEPHEN KULIK

1. Introduction. A root, \( a \), of an equation \( f(x) = 0 \) can be determined as a limit of a certain expression \( Q_n(x) \), as \( n \to \infty \), where \( Q_n(x) \) depends on \( f(x) \) and its first \( n \) derivatives, and \( x \) is given an arbitrary value which is closer to \( a \) than to any other root of the equation. The purpose of this paper is to show some applications of the method to algebraic equations with real roots. A root can be evaluated with any accuracy by a simple procedure, and in many cases close upper and lower bounds of the root can be found at the same time. The main result of §2 is presented by formulas (15)–(18). Formula (22) of §3 represents a generalization of the Laguerre quadratic formula for the solution of algebraic equations with real roots. Formula (33) of §4 gives a generalization of the Laguerre quadratic formula for finding upper and lower bounds of the roots of an algebraic equation with real roots and, at the same time, of the Bernoulli method for finding the numerically largest root of such an equation.

2. A few simple cases. In this and in the following sections, let \( f(x) = 0 \) be an algebraic equation of degree \( N \) with real coefficients and real roots \( a_i, i = 1, 2, \ldots, N, a_i \neq a_j \), for \( i \neq j \). We may write the identity

\[
1/f(x) = \sum A_i/(x - a_i)
\]

and, by differentiating it \( n - 1 \) times, derive another identity

\[
P_{n-1}/[f(x)]^n = \sum A_i/(x - a_i)^n,
\]

where \( P_n \) is evaluated recursively [4; 5]:

\[
P_n = f'P_{n-1} - f"fP_{n-2}/2! + \cdots + f^{(n)}(-f)^{n-1}P_0/n!
\]

with \( P_0 = 1 \) and \( f = f(x) \); and the sum is taken over all the roots of the equation.

Now if \( x \) is a number which is closer to a root \( a \), which we shall write without the subscript, than to any other root of the equation, then

\[
x - a = \lim fP_{n-1}/P_n,
\]

and

Received by the editors October 17, 1957 and, in revised form, August 15, 1958.

1 Sponsored by the Office of Ordnance Research, U. S. Army.
(5) \[ x - a = \lim \left[ \frac{f^n}{P_n} \right]^{1/n}, \quad \text{as } n \to \infty, \]

the \( n \)th root being appropriately chosen; it can be seen by the comparison with \( fP_{n-1}/P_n \).

In the same way, if we take the identity

(6) \[ \frac{f'(x)}{f(x)} = \sum (x - a_i)^{-1}, \]

we derive

(7) \[ D_n/[f(x)]^n = \sum (x - a_i)^{-n}, \]

from which it follows [4; 5] under the same conditions as in (4) and (5)

(8) \[ x - a = \lim fD_{n-1}/D_n, \]

and

(9) \[ x - a = \lim \left[ \frac{f^n}{D_n} \right]^{1/n}, \quad \text{as } n \to \infty \]

with

\[
D_n = fD_{n-1} - f''D_{n-2}/2! + \cdots + f^{(n-1)}(-f)^{n-2}D_1/(n - 1)! \\
+ f^{(n)}(-f)^{n-1}D_0/(n - 1)!, \quad n \geq 3,
\]

\[ D_0 = 1, \quad D_1 = f', \quad D_2 = f'' - f''f. \]

By taking \( n \) large enough in (4), (5), (8) or (9) we can get an approximation to \( a \) with any required accuracy. We shall denote it by \( x_1 \) in each of the four cases and obtain

(11) \[ x_1 = x - fP_{n-1}/P_n, \]

(12) \[ x_1 = x - \left( \frac{f^n}{P_n} \right)^{1/n}, \]

(13) \[ x_1 = x - fD_{n-1}/D_n, \]

(14) \[ x_1 = x - \left( \frac{f^n}{D_n} \right)^{1/n}. \]

The following statements give an account of the character of the approximations \( x_1 \) to a root of \( f(x) = 0 \) in some particular (see also [6]) cases.

1°. If \( x \) is an upper bound of the roots of the equation \( f(x) = 0 \), and \( a_N \) is its largest root, then

(15) \[ a_N < x - fP_{n-1}/P_n, \]

(16) \[ a_N < x - f/P_n^{1/n}, \]

(17) \[ a_N > x - fD_{n-1}/D_n, \quad n \geq 2 \]

(18) \[ a_N < x - f/D_n^{1/n}. \]
2°. If \( n \) is even, then the interval \( (x - f/D_{n/n}^{1/n}, x + f/D_{n/n}^{1/n}) \) is free of roots of the equation; and the interval \( (x - fN^{1/n}/D_{n/n}^{1/n}, x + fN^{1/n}/D_{n/n}^{1/n}) \) has in it at least one root of the equation, and only one if \( n \) is large enough and \( x \) is not the mid-point of the two adjacent roots.

In case of the Equations (11), (12), and (14) we can write by using (2) and (7), \( x - a_N = (x - x_1)(1 + \alpha_n) \), where \( \alpha_n \) is a positive non-increasing function of \( n \). From this fact inequalities (15), (16), and (18) follow. In case of (13) we may write

\[
x - a_N = (x - x_1)(1 + \beta_n)/(1 + \beta_{n-1}),
\]

where \( \beta_n \) is positive, decreasing with \( n \). Therefore, \( a_N > x_1 \), and this proves (17). Statement 2° follows immediately by a similar argument.

3. A generalization of the Laguerre quadratic formula. Let \( n \) be an even number and \( u \) an arbitrary real number not equal to \( x \) or a root of the equation. We may write the identity

\[
(u - x)^n f'(x)/f(x) = \sum (u - a_i)^n/(x - a_i) + N x^{n-1} + c_1 x^{n-2} + \cdots + c_n
\]

and by differentiating it \( n - 1 \) times derive

\[
Q_n(x, u)/[f(x)]^n = \sum (u - a_i)^n/(x - a_i)^n,
\]

where the sum is taken over all the roots of the equation \( f(x) = 0 \), and \( Q_n(x, u) \) is evaluated by the formula

\[
Q_n(x, u) = \sum_{k=0}^{n-1} \binom{n}{k} (u - x)^{n-k} D_{n-k} f(x)^k + N.
\]

We can now prove that

3°. If \( x \) is an arbitrary real number, then the two real roots, \( X_1 \) and \( X_2 \), of the equation

\[
Q_n(x, u)/[f(x)]^n = (u - X)^n/(x - X)^n
\]

are separated by \( x \) and the interval \( (X_1, X_2) \) has no roots of \( f(x) = 0 \) in it. Further, at least one of these roots converges to a root of the equation \( f(x) = 0 \), as \( n \to \infty \). With a finite \( n \) the accuracy of the approximation \( X \) to a root a increases with the increase of \( n \), and it depends also on the choice of \( u \).

For a proof we shall consider the equation

\[
\sum (u - a_i)^n/(x - a_i)^n - (u - X)^n/(x - X)^n = 0,
\]

which is identical with (22). The left member of this equation is posi-
tive for \( X \) equal to any root of \( f(x) = 0 \), because the last term in it is
equal to only one term of the sum, all of whose terms are positive,
but if \( X \) is close enough to \( x \), then the last member becomes larger
than the total sum and the left side of (23) will be negative. This
proves the first half of the statement. If \( u \) satisfies the equation
\[
(u - a_1)/(x - a_1) = -(u - a_2)/(x - a_2),
\]
where \( a_1 \) and \( a_2 \) are the two
roots adjacent to \( x \), \( a_1 < x < x_2 < a_2 \), then the sum in (23) has two
dominant terms, \((u - a_1)^n/(x - a_1)^n\) and \((u - a_2)^n/(x - a_2)^n\). Therefore
\( x_1 \) and \( x_2 \) will approach \( a_1 \) and \( a_2 \) with the increase of \( n \). If \( x - a_1 > a_2 - x \), then the two dominant terms are obtained by taking
\[
u = \left[ (a_1 + a_2) x - 2 a_1 a_2 \right] / (2 x - a_1 - a_2).
\]
Any \( u \) in the interval \( x < u < \left[ (a_1 + a_2) x - 2 a_1 a_2 \right] / (2 x - a_1 - a_2) \) makes the dominant term the one
involving \( a_1 \); and the two intervals, \( -\infty < u < x \) and
\[
\left[ (a_1 + a_2) x - 2 a_1 a_2 \right] / (2 x - a_1 - a_2) < u = \infty,
\]
define \( u \) making the dominant term the one involving \( a_2 \). If \( x - a_1 = a_2 - x \), then \( u > x \) makes the dominant term the one involving \( a_1 \),
and \( u < x \) makes the dominant term the one involving \( a_2 \). All possible
situations are illustrated by diagrams 1 and 2.

Two values \( u_1 \) and \( u_2 \) necessary for approximating \( a_1 \) and \( a_2 \) can be
selected by first applying formula (14) for determining an interval
\( (x', x'') \) around \( x \) free of roots. Afterwards, \( u_1 \) can be taken to be
any number in the interval \( x < u < 2 x' - x \), and \( u_2 \) any number in
the interval \( 2 x' - x \leq u < x \). Both \( u_1 \) and \( u_2 \) would fall in the intervals
defined above; however, they would not necessarily insure the best
approximations possible by (22) with a fixed \( n \).

Denoting now
\[
S_n = Q_n(x, u) /[f(x)]^n,
\]
we get from (22)
\[
X = x - (x - u)/(1 \pm (S_n)^{1/n})
\]
the \( n \)th root being taken with plus for \( X_1 \) when \( u < x \), and with minus
when \( u > x \). A particular case, \( u \rightarrow \infty \), reduces formula (25) to (14).

In the particular case \( n = 2 \), the best approximations to the roots
\( a_1 \) and \( a_2 \) can easily be obtained as a simple problem in extrema. Two
equations determining \( X_1 \), \( X_2 \) and \( u_1 \), \( u_2 \) are as follows:
\[
(N - 1)(L^2 D_2 - f^2) - (LD_1 - f)^2 = 0,
\]
and
\[
[(N - 2)f'' - (N - 1)ff''']L^2 + 2ff'L - Nf^2 = 0,
\]
and

\[(27)\]  

\[f''m^2 + 2(N - 1)f'm + N(N - 1)f = 0,\]

where \(L = x - X\), and \(m = u - x\).

From these equations

\[(28)\]  

\[L = \frac{Nf}{f' \pm H^{1/2}},\]

and

\[(29)\]  

\[m = -\frac{N(N - 1)f}{(N - 1)f' \mp H^{1/2}},\]

where \(H = (N - 1)^2f'^2 - N(N - 1)f''\).

A few relations between \(L\) and \(m\) are helpful in calculations, namely:

\[(30)\]  

\[L = mf/[mf' + (N - 1)f],\]

\[(31)\]  

\[m = -LF(N - 1)f/[L'f' - f],\]
and

\[ 1/L + 1/m = \pm H^{1/2}/(N - 1)f. \]  

Equation (26) is the famous quadratic formula suggested by Laguerre [7, p. 97–103] and derived by Hermite [7, p. 461].

We do not know, however, a simple equation defining \( u \) which would give the best approximations in the general case (25). Before applying the general formula (22), it is useful to know the results of applying (26) and (27) in deciding what value of \( u \) should be used. The \( u \) and \( x \) values could be selected such as to make the calculations as simple as possible. This makes the use of formula (25) preferable to the repeated applications of the Laguerre quadratic formula.

4. A generalization of the Bernoulli method. While the value of \( x \) in the preceding section could also be taken as an upper or lower bound of the roots of the equation \( f(x) = 0 \), a particular case \( x \to \infty \) deserves a special treatment.

If in (22) and (23) we put \( x \to \infty \), we get

\[ Q_n(u) = (u - X)^n, \]

and

\[ \sum (u - a_i)^n - (u - X)^n = 0, \]

respectively, where

\[ Q_n(u) = Nu^n + \sum_{k=1}^{n} \binom{n}{k} D_k'. \]

Here

\[
D_k' = \begin{bmatrix}
p_1 & 1 & 0 & \cdots & \cdots & \cdots \\
p_2 & p_1 & 1 & \cdots & \cdots & \cdots \\
\vdots & \vdots & \vdots & \ddots & \ddots & \ddots \\
Np_N & p_{N-1} & p_{N-2} & \cdots & \cdots & \cdots \\
0 & p_N & p_{N-1} & \cdots & \cdots & \cdots \\
0 & 0 & 0 & \cdots & p_N & p_{N-1} \cdots p_1 \\
0 & 0 & 0 & \cdots & \cdots & \cdots
\end{bmatrix}
\]

if \( f(x) = x^N + p_1x^{N-1} + p_2x^{N-2} + \cdots + p_N \). It can be evaluated recursively

\[
D_k' = p_1D_{k-1}' - p_2D_{k-2}' + \cdots + (-1)^{k-2}p_{k-1}D_1' + (-1)^{k-1}kp_kD_0'
\]

\[ D_0' = 1, \quad D_1' = p_1, \quad D_2' = p_1^2 - 2p_2; \]
or by an explicit expression

\[ (38) \quad D'_k = k \sum (-1)^{i+j+\ldots+r-1} \frac{(i+j+\ldots+r-1)!}{i!j!\ldots r!} p_q p_s \ldots p_r, \]

the sum being taken over all the terms for which \( qi + sj + \ldots + tr = k. \)

Now, if \( a_1 \) and \( a_N \) denote the smallest and the largest roots of the equation \( f(x) = 0 \), remembering that \( n \) is even, the following conclusions can be made as to the roots of (33).

4°. If \( u \) is an arbitrary real number, then the two real roots \( X_1 \) and \( X_2 \), \( X_1 < X_2 \), of the Equation (33) are separated by all the roots of the equation \( f(x) = 0 \). If \( u = (a_1 + a_N)/2 \), then \( X_1 \) converges to \( a_1 \), and \( X_2 \) to \( a_N \), as \( n \to \infty \); if \( u > (a_1 + a_N)/2 \), then \( X_1 \) converges to \( a_1 \), and if \( u < (a_1 + a_N)/2 \) then \( X_2 \) converges to \( a_2 \) as \( n \to \infty \).

The first half of the statement follows from the fact that the last term in (34) equals or is less than one term of the sum so long as \( a_1 \leq X \leq a_N \), but it is larger than the sum for some \( X < a_1 \) and \( X > a_N \). Further, if \( u - a_1 = a_N - u \), there are two dominant terms in the sum, and therefore \( X_1 \) and \( X_2 \) approach \( a_1 \) and \( a_N \) respectively as \( n \) increases; if \( u - a_1 > a_N - u \), then only one term is dominant, \( (u - a_1)^n \), and therefore \( X_1 \) converges to \( a_1 \); for the same reason if \( u - a_1 < a_N - u \), \( X_2 \) converges to \( a_N \). The particular case \( u = 0 \) reduces (33) to

\[ (39) \quad D'_n = X^n, \]

the result which was suggested by Bernoulli [1] for approximating the root which is largest in absolute value.

The values of \( u_1 \) and \( u_2 \) giving the best approximations to \( a_1 \) and \( a_N \) can easily be found for \( n = 2 \) as a problem in extrema. The equation defining these values is

\[ (40) \quad N(N - 1)u^2 + 2(N - 1)p_1u + 2p_2 = 0, \]

and the Equation (33) becomes

\[ (41) \quad NX^2 + 2p_1X + 2(N - 1)p_2 - (N - 2)p_1^2 = 0. \]

From these equations

\[ (42) \quad u = - \frac{[(N - 1)p_1 \pm (H_1)^{1/2}]/N(N - 1)}{2p_2/[(N - 1)p_1 \mp (H_1)^{1/2}]}, \]

\[ X = \frac{(-p_1 \pm (H_1)^{1/2})/N}{[-(N - 2)p_1^2 + 2(N - 1)p_2]/(-p_1 \mp (H_1)^{1/2})}, \]
where \( II_1 = (N-1)^2p_1^2 - 2N(N-1)p_2 \). The upper sign before the radical corresponds to the largest root. The following two relations between \( u \) and \( X \) are useful in applications:

\[
(N - 1)u + X + p_1 = 0,
\]

and

\[
u - X = \pm (H_1)^{1/2}/(N - 1).
\]

The values of \( u \) determined by (40), and rounded off for convenience, can be used for the general formula (33). Formula (41) represents another Laguerre quadratic equation \([7, p. 101]\) for evaluating upper and lower bounds of the roots of algebraic equations with real roots.

References

7. E. N. Laguerre, Oeuvres de Laguerre 1, 1898.