1. Iterate the function $2y^2 - 1$, starting with the number y whose arc sin is required.
2. Record the *signs* of the iterates in order.
3. Accumulate the signs; that is, record the "partial products" of the signs in order.
4. Write descending powers of 2 between the signs accumulated.
5. Multiply the series obtained by $\pi/2$.

Example: Compute arc sin $\sqrt{.75}$

1.	$\sqrt{.75}$,	.5,	$-.5$,	$-.5$,	$-.5$	$\cdots$	
2.	+	+	−	−	−	$\cdots$	
3.	+	+	−	+	−	$\cdots$	
4.	$+1/2$	$+1/4$	$-1/8$	$+1/16$	$-1/32$	$\cdots$	$= 2/3$
5.	$\pi/2 \cdot (2/3) = \pi/3$						
	arc sin $\sqrt{.75} = \pi/3$.						

5. Comparison with Other Methods. The usefulness of the method described above as compared with other methods depends, of course, on the function to be evaluated and on the features of the machine to be used.

The fact that each iteration yields exactly one binary bit may be an advantage or a disadvantage; a method where error decreases faster than 2^{-n} will converge with fewer iterations than this one. On the other hand, one iteration of this method may consist of fewer commands than an iteration of another method. The logarithm program for the CRC-102A described above has 4 commands per iteration as compared with 14 commands per iteration in another program for logarithm on the same machine. The program for arcsin by the method described above has 9 commands per iteration as compared to 22 for another arcsin program. The fact that each iteration yields exactly one binary bit also simplifies error analysis, for the number of digits of accuracy is exactly one less than the number of digits computed.

D. R. MORRISON

SANDIA Corp.
Albuquerque, New Mexico

A Method for Solving Algebraic Equations using an Automatic Computer

Introduction. Many methods have been developed for solving algebraic equations and several of these have been used with automatic computers [1, 2]. Those methods which are most suitable for use with automatic computers are ones which apply to a wide class of equations and which are relatively rapid when the degree of the equation is large. The method described here has been constructed with these considerations in mind and has been programmed for the ILLIAC computer at the University of Illinois.

A process is to be constructed to find n solutions to the general algebraic equations of nth degree

$$f(x) = a_0x^n + a_1x^{n-1} + \cdots + a_n = 0 \tag{1}$$

where the coefficients $a_0, a_1, \cdots, a_n$ are complex numbers and $a_0 \neq 0$. Each root is found by an iterative procedure. Successive iterations toward a particular root are obtained by finding the nearer root of a quadratic whose curve passes through the last three points. The quadratic will in general have complex coefficients and complex roots. This solution is accomplished by a variation of the standard quadratic formula. Although the method derived here is rather complicated, no evaluation of derivatives of $f(x)$ and only one evaluation of the polynomial $f(x)$ is required per iteration. If the degree of the equation is large a greater amount of time is spent evaluating the function than is spent in the remainder of the process. Thus, the time spent per iteration is less with this process than with iterative schemes which require the calculation of derivatives, whenever the degree of the equation is large.

The Lagrange interpolation formula will yield a quadratic

$$L_i(f(x)) = b_0x^2 + b_1x + b_2 \tag{2}$$

whose curve passes through the last three points $(x_i, f(x_i))$, $(x_{i-1}, f(x_{i-1}))$, $(x_{i-2}, f(x_{i-2}))$ where the coefficients b_0, b_1, b_2 satisfy the system of equations

$$\begin{aligned} b_0x_i^2 + b_1x_i + b_2 &= f(x_i), \\ b_0x_{i-1}^2 + b_1x_{i-1} + b_2 &= f(x_{i-1}), \\ b_0x_{i-2}^2 + b_1x_{i-2} + b_2 &= f(x_{i-2}). \end{aligned} \tag{3}$$

A somewhat more convenient representation for this quadratic is obtained by introducing the new quantities $h = x - x_i$, $h_i = x_i - x_{i-1}$, $h_{i-1} = x_{i-1} - x_{i-2}$, $\lambda = h/h_i$, $\lambda_i = h_i/h_{i-1}$, and $\delta_i = 1 + \lambda_i$. The Lagrange interpolation formula may now be written as the following quadratic in λ.

$$\begin{aligned} L_i(f(x)) = {} & \lambda^2\delta_i^{-1}[f(x_{i-2})\lambda_i^2 - f(x_{i-1})\lambda_i\delta_i + f(x_i)\lambda_i] \\ & + \lambda\delta_i^{-1}[f(x_{i-2})\lambda_i^2 - f(x_{i-1})\delta_i^2 + f(x_i)(\lambda_i + \delta_i)] + f(x_i). \end{aligned} \tag{4}$$

A single iterative step is obtained by letting x_{i+1} be a value of x which makes $L_i(f(x))$ vanish. We may solve the quadratic equation in λ obtained by setting expression (4) equal to zero. We then obtain $\lambda = \lambda_{i+1} = (x_{i+1} - x_i)/(x_i - x_{i-1})$ by using the inverse of the standard quadratic formula

$$\lambda_{i+1} = \frac{-2f(x_i)\delta_i}{g_i \pm \sqrt{g_i^2 - 4f(x_i)\delta_i\lambda_i[f(x_{i-2})\lambda_i - f(x_{i-1})\delta_i + f(x_i)]}} \tag{5}$$

where $g_i = f(x_{i-2})\lambda_i^2 - f(x_{i-1})\delta_i^2 + f(x_i)(\lambda_i + \delta_i)$. From λ_{i+1} we may obtain $h_{i+1} = \lambda_{i+1}h_i$ and $x_{i+1} = x_i + h_{i+1}$. This x_{i+1} represents a zero of the quadratic described above. The sign in the denominator of (5) is always taken so as to make the denominator have the greater magnitude. This makes λ_{i+1} and h_{i+1} each be

that one of the two possible choices having the smaller magnitude so that x_{i+1} is the root which is closer to x_i.

A convenient starting method for this process uses artificial starting values at $x_0 = -1$, $x_1 = 1$, and $x_2 = 0$:

$$\begin{aligned} &a_n - a_{n-1} + a_{n-2} && \text{is used for } f(x_0), \\ &a_n + a_{n-1} + a_{n-2} && \text{is used for } f(x_1), \\ &a_n && \text{is used for } f(x_2). \end{aligned} \tag{6}$$

thus making $\lambda_2 = -\frac{1}{2}$ and $h_2 = -1$. This choice of starting values makes

$$L_2(f(x)) = a_n + a_{n-1}x + a_{n-2}x^2$$

which approximates to $f(x)$ in the neighborhood of $x = 0$. The advantage of this starting process is that it requires no special evaluations of the polynomial $f(x)$ and is therefore rapid.

Convergence of the Process. A final value of the root x_i is taken when $|x_i - x_{i-1}|/|x_i|$ becomes less than some preassigned number. Such a convergence criterion is consistent with the use of floating point arithmetic in the calculation. As a result of this criterion we see that convergence occurs if $x_{i-1} = x_i$. This means that before convergence no two consecutive iterative results will be equal. Furthermore, if $x_i = x_{i-2}$ we have $\delta_i = \dfrac{x_i - x_{i-2}}{x_{i-1} - x_{i-2}} = 0$ so by (5) $\lambda_{i+1} = 0$ and $x_{i+1} = x_i$ also giving convergence unless $x_i = 0$. Thus in normal operation of the process we see that x_i, x_{i-1}, and x_{i-2} are distinct.

As each root is found it may be divided into the polynomial $f(x)$ thus reducing the degree of the polynomial by one. The algorithm for this reduction is the commonly used one

$$a_i' = ra_{i-1} + a_i, \quad (i = 0, 1, 2, \cdots) \tag{7}$$

where a_i' is the new coefficient to replace a_i and r is the root which has just been found. We make $a_{-1} = 0$. Errors introduced by this process will be reduced if the roots are eliminated in order of increasing magnitude. By always starting at the point $x = 0$ one will tend to find roots in roughly this order.

No general proof of convergence in the large has been obtained for this process, but convergence can be shown to occur whenever the process leads one sufficiently close to a single or double root.

In order to facilitate the study of convergence let us assume that $x_{i+1} = 0$. This loses no generality since a simple shift of origin is always possible within the system. At point x_{i+1} we also have $L_i(x_{i+1}) = 0$ so that

$$L_i(x_{i+1}) = b_0x^2{}_{i+1} + b_1x_{i+1} + b_2 = b_2 = 0. \tag{8}$$

Now each of the functions appearing on the right hand sides of equations (3) may be expanded about $x_{i+1} = 0$ as

$$f(x_{i-m}) = \sum_{k=0}^{n} x^k{}_{i-m} f^{(k)}(0)/k! \tag{9}$$

and b_2 may be written as

$$b_2 = \sum_{k=0}^{n} b_{2k} f^{(k)}(0)/k! = 0 \tag{10}$$

where b_{2k} is obtained by solving the system of equations

$$\begin{aligned} b_{0k}x_i^2 + b_{1k}x_i + b_{2k} &= x_i^k \\ b_{0k}x^2{}_{i-1} + b_{1k}x_{i-1} + b_{2k} &= x^k{}_{i-1} \\ b_{0k}x^2{}_{i-2} + b_{1k}x_{i-2} + b_{2k} &= x^k{}_{i-2}. \end{aligned} \tag{11}$$

This system of equations may be solved by elimination provided x_i, x_{i-1}, and x_{i-2} are distinct and we obtain $b_{20} = 1$, $b_{21} = b_{22} = 0$ for the first three solutions. When $k \geq 3$ the solution becomes somewhat more difficult but may be carried out by eliminating b_{1k} between the first two equations to give

$$b_{0k}x_ix_{i-1} - b_{2k} = x_ix_{i-1} \sum_{p=0}^{k-2} x_i^p x_{i-1}^{k-2-p}.$$

A similar elimination may be made between another pair and the result combined with the above equations to give

$$b_{2k} = x_ix_{i-1}x_{i-2} \sum_{p+q+s=k-3} x_i^p x_{i-1}^q x_{i-2}^s. \tag{12}$$

The sum is to be taken over all non-negative integral p, q, and s satisfying $p + q + s = k - 3$. All of these results may be inserted in (8) and (10) giving

$$f(0) = -x_ix_{i-1}x_{i-2}\left(\tfrac{1}{6}f'''(0) + \sum_{k=4}^{n} \sum_{p+q+s=k-3} x_i^p x_{i-1}^q x_{i-2}^s f^{(k)}(0)/k! \right). \tag{13}$$

Up to this point no approximations have been made and no limits have been taken. Equation (13) expresses the same relationship contained in (5). We now assume that the points x_i, x_{i-1}, x_{i-2} lie in the neighborhood of a root r. Thus if we let $\epsilon_{i+1} = x_{i+1} - r$, $\epsilon_i = x_i - r$, $\epsilon_{i-1} = x_{i-1} - r$ and $\epsilon_{i-2} = x_{i-2} - r$ the magnitudes of the last three quantities are all assumed to be less than some upper bound ϵ_m

$$|\epsilon_i| < \epsilon_m, \quad |\epsilon_{i-1}| < \epsilon_m, \quad |\epsilon_{i-2}| < \epsilon_m. \tag{14}$$

We also now make the tentative assumption

$$|\epsilon_{i+1}| < \epsilon_m, \tag{15}$$

and we shall seek to justify this assumption later. If (14) and (15) are inserted in (13) and the functions are expanded about r we obtain

$$\begin{aligned} \epsilon_{i+1}f'(r) + \epsilon_{i+1}^2 \tfrac{1}{2}f''(r) + \epsilon_{i+1}^3 \tfrac{1}{6}f'''(r) \\ = -(\epsilon_i - \epsilon_{i+1})(\epsilon_{i-1} - \epsilon_{i+1})(\epsilon_{i-2} - \epsilon_{i+1})\tfrac{1}{6}f'''(r) + 0(\epsilon_m^4). \end{aligned} \tag{16}$$

If r is a single root we see that $\epsilon_{i+1} = 0(\epsilon_m^3)$ so that we may write

$$\epsilon_{i+1} = -\epsilon_i\epsilon_{i-1}\epsilon_{i-2}\frac{f'''(r)}{6f'(r)} + 0(\epsilon_m^4). \tag{17}$$

A solution ϵ_{i+1} to this equation does exist if ϵ_m is sufficiently small and will satisfy (15). This solution will also satisfy $L_i(f(x_{i+1})) = 0$, and hence we are justified in assumption (15) and hence (17) for at least one of the two x_{i+1} for which $L_i(f(x_{i+1})) = 0$ holds. We now wish to show that (17) holds for that x_{i+1} which is actually chosen by the process described in connection with equation (5). If both x_{i+1} satisfy (17), the proof need not be given. If, however, one does and one does not, we must make some further analysis. It was pointed out that the process chooses the point x_{i+1} which is nearer to x_i. The point given by equation (17) must satisfy

$$|x_{i+1} - x_i| = |\epsilon_{i+1} - \epsilon_i| < 2\epsilon_m. \tag{18}$$

This must therefore also hold for the x_{i+1} which is chosen by the process in (5) and hence $|\epsilon_{i+1}| < 3\epsilon_m$ for this case. But $|\epsilon_{i+1}| < 3\epsilon_m$ is adequate to give (17) for sufficiently small ϵ_m and we may therefore assume (17) for the ϵ_{i+1} obtained in the process.

A general limiting formula for the ϵ_j in the neighborhood of a root may be obtained from equation (17). If logarithms are taken on both sides we obtain

$$\log \epsilon_{i+1} = \log \epsilon_i + \log \epsilon_{i-1} + \log \epsilon_{i-2} + \log\left(-\frac{f'''(r)}{6f'(r)}\right) + 0(\epsilon_m). \tag{19}$$

Neglecting the terms $0(\epsilon_m)$ we may solve (19) as a difference equation using standard techniques and obtain

$$\log \epsilon_j = c_1m_1^j + c_2m_2^j + c_3m_3^j - \tfrac{1}{2}\log\left(-\frac{f'''(r)}{6f'(r)}\right). \tag{20}$$

Where the constants c_1, c_2, and c_3 are determined by the starting values and the three orders of convergence m_1, m_2, and m_3 are roots of the characteristic equation

$$m^3 = m^2 + m + 1. \tag{21}$$

The roots are

$$m_1 = 1.84,$$
$$m_2, m_3 = -.420 \pm .606i.$$

Since the last two roots have magnitude less than 1 their effect will die out and the order of the process is given by m_1. After these approximations become valid we have from (20)

$$\epsilon_{j+1} = K\epsilon_j^{1.84} \tag{22}$$

where

$$K = \left(-\frac{f'''(r)}{6f'(r)} \right)^{-\frac{1}{2}}.$$

In the case of a double root a similar argument exists. Equation (17) is then replaced by

$$\epsilon_{i+1}^2 = -\epsilon_i \epsilon_{i-1} \epsilon_{i-2} \frac{f'''(r)}{3f''(r)} + 0(\epsilon_m^4) \tag{23}$$

and the characteristic equation becomes

$$2m^3 = m^2 + m + 1. \tag{24}$$

It has roots

$$m_1 = 1.23,$$
$$m_2, m_3 = -.367 \pm .520i,$$

and again the order of convergence is given by m_1. We therefore have in the limit

$$\epsilon_{j+1} = K\epsilon_j^{1.23} \tag{25}$$

with

$$K = \left(-\frac{f'''(r)}{3f''(r)} \right)^{-\frac{1}{2}}.$$

Convergence of the Generalized Process. One might imagine a generalized process in which an α degree Lagrange interpolation polynomial $L_i(x, \alpha)$ is used rather than the quadratic of equation (2). This presumes that some new method for obtaining the nearest root of this polynomial is to be used. Since the direct method corresponding to equation (5) would no longer be practical, one would probably use some iterative method for solution of the equation $L_i(x_{i+1}, \alpha) = 0$.

We now wish to investigate the convergence rate for such a process.

A general set of equations corresponding to equations (11) may be formed. They are

$$\sum_{j=0}^{\alpha} b_{jk} x_{i-s}^{\alpha-j} = x_{i-s}^k, \quad s = 0, 1, \cdots, \alpha. \tag{26}$$

When $k > \alpha$ the quantity $b_{\alpha k}$ may be obtained by elimination. (This direct method for obtaining $b_{\alpha k}$ was pointed out to the author by Mr. W. Scott Bartky.) Let us eliminate $b_{\alpha-1,k}$ between the first equation and each succeeding equation giving

$$\sum_{j=0}^{\alpha-2} b_{jk} x_i x_{i-s} (x_i^{\alpha-j-1} - x_{i-s}^{\alpha-j-1}) - b_{\alpha k}(x_i - x_{i-s}) = x_i x_{i-s}(x_i^{k-1} - x_{i-s}^{k-1}), \tag{27}$$
$$s = 1, 2, \cdots, \alpha.$$

Each equation may be divided by $(x_i - x_{i-s})$ giving

$$(28)\quad \sum_{j=0}^{\alpha-2} b_{jk}x_ix_{i-s}\sum_{l=0}^{\alpha-j-2} x_i^l x_{i-s}^{\alpha-2-j-l} - b_{\alpha k} = x_ix_{i-s}\sum_{l=0}^{k-2} x_i^l x_i^{k-2-l},\quad s = 1, 2, \cdots, \alpha.$$

We next eliminate $b_{\alpha-2,k}\cdots, b_{0k}$ in a similar manner until the result

$$(29)\quad b_{\alpha k} = (-1)^\alpha x_ix_{i-1}\cdots x_{i-\alpha}\sum x_i^{p_0}x_{i-1}^{p_1}\cdots x_{1-\alpha}^{p_\alpha},$$
$$p_0 + p_1 + \cdots + p_\alpha = k - (\alpha + 1)$$

is obtained. In this expression the summation is made over all terms for which the exponents $p_0, p_1, \cdots, p_\alpha$ are non-negative integers and $\sum_{j=0}^{\alpha} p_j = k - (\alpha + 1)$. We also see directly from (26) that $b_{\alpha k} = 0$ if $k \leq \alpha$, except that $b_{\alpha 0} = 1$.

We may therefore obtain a generalization of equation (17)

$$(30)\quad \epsilon_{i+1} = (-1)^{\alpha+1}\epsilon_i\epsilon_{i-1}\cdots\epsilon_{i-\alpha}\frac{f^{(\alpha+1)}(r)}{(\alpha+1)!f'(r)} + 0(\epsilon_m^{\alpha+2})$$

and obtain for the characteristic equation corresponding to (21)

$$(31)\quad m^{\alpha+1} = m^\alpha + m^{\alpha-1} + \cdots + 1$$

This equation has one root m_1 which lies between 1 and 2 on the real axis and which approaches 2 with increasing α. The remaining roots lie within the unit circle and therefore represent perturbations which die out. The order of convergence of the process to single roots is therefore given by m_1. Since this can never reach 2 we conclude that there is little to be gained in speed of convergence by letting α exceed 2.

One should not ignore the possibility of letting $\alpha = 1$. In this case the formula corresponding to (5) is greatly simplified since a linear equation rather than a quadratic now must be solved. This choice, however, suffers from a disadvantage if all the coefficients of the original equation are real. If one starts from a real point x_0 then all successive iterative results x_i will also be real and hence only real roots will be found.

Tests of the Method. The process with $\alpha = 2$ as outlined in the preceding sections was altered slightly in practice. Whenever the new value of the function $f(x_{i+1})$ is calculated the quantity $|f(x_{i+1})|/|f(x_i)|$ is formed. If this latter quantity exceeds 10 the quantity λ_{i+1} is halved and $h_{i+1}x_{i+1}$, and $f(x_{i+1})$ are recomputed accordingly. With this revision the process has produced convergence in all the cases tested.

Another alteration was made to handle the case in which the denominator of (5) is zero. This occurs whenever $f(x_i) = f(x_{i-1}) = f(x_{i-2})$ and in such cases the arbitrary value $\lambda_{i+1} = 1$ is chosen since (5) may no longer be used.

The process ($\alpha = 2$) was tested for equations of varying degree. Fourteen equations were solved starting with degree 10 and progressing in steps of ten through degree 140. Each equation was formed by choosing random points as

roots within the square having vertices $\pm 1 \pm i$. Polynomials were then formed from these roots. The solutions to these polynomials were then compared with the original random numbers which were used to generate the polynomial. Results are summarized in the table.

Degree of Equation	Accuracy of Least Accurate Root	Accuracy of Last Root to be Found	Time Taken by ILLIAC for Complete Solution in Minutes
10	10^{-7}	10^{-9}	1
20	10^{-8}	10^{-8}	2
30	10^{-5}	10^{-8}	5
40	10^{-4}	10^{-9}	6
50	10^{-4}	10^{-5}	10
60	10^{-5}	10^{-7}	12
70	10^{-4}	10^{-5}	17
80	10^{-4}	10^{-5}	20
90	10^{-1}	10^{-5}	20
100	—	10^{-6}	33
110	—	—	42
120	—	—	43
130	—	10^{-8}	48
140	—	—	60

Dashes in the table indicate that some roots were too inaccurate to be identified. In all equations some roots appeared which were correct to 10^{-8} or better. The solutions to the 100th degree equation, which had some unidentifiable roots, were used to generate a polynomial. All coefficients of this polynomial agreed with the coefficients of the original polynomial to at least 6 decimal places. This result indicates that the obtaining of accurate values of the roots of the equations of higher degree was precluded by the limited accuracy of the coefficients and independent of the method of solution.

The equation $x^{128} - 1 = 0$ was solved as an example of a special type of equation whose solution could be easily checked. The maximum error occurring in any root was of order 10^{-7} and the time for solution was 70 minutes.

No equation whose solution has been attempted has failed to yield convergence although as indicated in the table, the solutions of equations of large degree may be greatly in error. We conclude that convergence in the large does occur in most practical cases in spite of the fact that convergence has only been proved for single and double roots when the process has brought one to the neighborhood of a root.

DAVID E. MULLER

University of Illinois
Urbana, Illinois

1. R. A. BROOKER, "The solution of algebraic equations on the EDSAC," Cambridge Phil. Soc., *Proc.*, v. 48, 1952, p. 255–270.
2. HANS J. MAEHLY, "Zur Iterativen Auflösung Algebraischer Gleichungen," *Z. Angew. Math. Physik*, v. 5, 1954, p. 260–263.