Accelerated Convergence, Divergence, Iteration, Extrapolation, and Curve Fitting

J. Ross Macdonald
Texas Instruments Incorporated, Dallas, Texas
(Received 3 February 1964; in final form 11 May 1964)

This paper discusses some applications of the epsilon algorithm (EA), a sequential procedure for calculating Padé approximants. The EA may be used to: (1) accelerate the convergence of slowly converging series and iterations; (2) obtain useful results from divergent series and iterations; (3) obtain the limits of iterated vector and matrix sequences; (4) aid in the solution of differential and integral equations; (5) carry out numerical integration in a new way; (6) extrapolate; (7) fit a curve to a polynomial or to a constant plus sum of exponentials.

As an illustration of curve fitting and extrapolation, we present results obtained with exact polynomial data plus random noise combined additively or proportionately. For such nonstationary data, the results are comparable, and in some cases superior, to least squares in yielding good estimates of the exact polynomial coefficients. One important advantage of the EA is that it builds up polynomials whose lower-order coefficients are independent of higher-order ones. This property is valuable when the degree of the polynomial is unknown. Finally, a simple empirical equation is given relating the precision of least-squares-calculated polynomial coefficients to the degree of the fitted polynomial and the number of effective decimal digits carried in the calculation.

INTRODUCTION

THE majority of chemists and physicists would probably agree with Abel, who wrote in 1828: "Divergent series are the invention of the devil, and it is shameful to base on them any demonstration whatsoever." This view is long outmoded among the mathematicians[1,2]; thus, it appears useful to bring to the attention of those who may need to use them certain simple but powerful techniques applying to the subjects of the title. Mathematical justification of these techniques and more precise statements concerning their range of applicability will be found in the later references.

The solutions of many problems of physical interest are obtained or may be written (after change of variable, multiplication by some power of z, etc.) as an infinite power series of the form

$$P(z)=\sum_{i=0}^{\infty} a_i z^i, \tag{1}$$

which may or may not be rapidly convergent and is frequently divergent for some finite values of z. Let us define the mth partial sum of (1) as

$$S_m(z)=\sum_{i=0}^{m} a_i z^i (m=0,1,2\cdots), \tag{2}$$

and restrict attention to reasonably regular, nonpathological types of (convergent or divergent) series, such as those arising from physical problems. We shall initially be interested in the sum of (1) for some value of z. When (1) is divergent, the "sum" (termed the antilimit by Shanks[3]) will be understood to be the value of the function most naturally associated with the series. Thus, the antilimit of

$$\sum_{i=0}^{\infty} z^i,$$

for $z=2$, is -1 since this is the value of $(1-z)^{-1}$, the "natural" function associated with this series, at $z=2$.

Under quite widely applicable conditions, an excellent analytical or numerical approximation to $P(z)$ may be obtained from its first (or any sequential) $2n+1$ partial sums by applying the epsilon algorithm[4] (EA) to them.

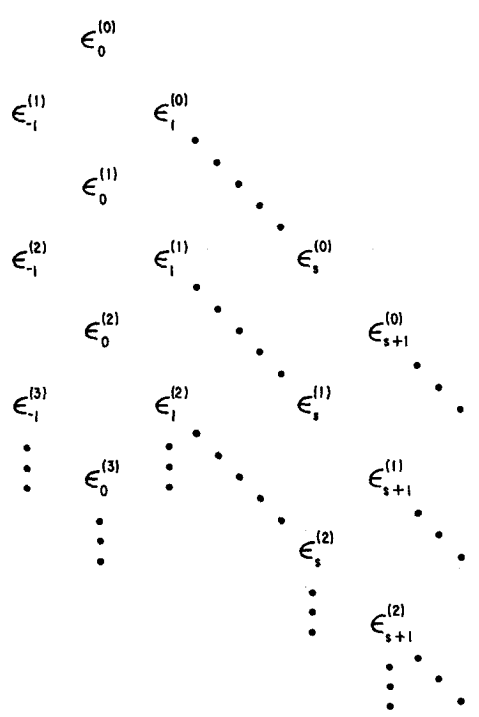

FIG. 1. Epsilon algorithm lozenge diagram.

[1] G. H. Hardy, *Divergent Series* (Clarendon Press, Oxford, England, 1956).
[2] R. P. Agnew, Mich. Math. J. 4, 105 (1957).
[3] D. Shanks, J. Math. Phys. 34, 1 (1955).
[4] P. Wynn, Math. Tables and Other Aids to Computation 10, 91 (1956). For EA instability conditions see P. Wynn, Nieuw Archief voor Wiskunde 9, 117 (1961).

This nonlinear sequence-to-sequence transformation is given by

$$\epsilon_{s+1}^{(m)} = \epsilon_{s-1}^{(m+1)} + [\epsilon_s^{(m+1)} - \epsilon_s^{(m)}]^{-1}, \qquad (3)$$

with $\epsilon_{-1}^{(m)} = 0$ and $\epsilon_0^{(m)} = S_m$. The quantities $\epsilon_{s+1}^{(m)}(s = 1,2,\cdots)$ generated from the $2n+1$ S_m's may conveniently be arranged in a lozenge diagram, as in Fig. 1, with s constant for the columns and m constant for the upper diagonals. Then, the quantities $\epsilon_0^{(m)}\{m = 0, 1, 2, \cdots, 2n\}$, $\epsilon_2^{(m)}\{m = 0, 1, 2, \cdots, (2n-2)\}, \cdots, \epsilon_{2n}^{(0)}$ are successively better and better approximations to $P(z)$. In general, the larger n, the better the final approximation, although in some cases of interest the exact answer may be obtained with n finite and relatively small (see later discussion).

The reason for the above results lies in the circumstance[3,5] that for the above initial conditions $\epsilon_{2r}^{(m)}$ is the $[r, r+m]$ Padé approximant to $P(z)$. Such an approximant is a quotient of polynomials with numerator of degree M and denominator of degree N and has the property that its power series expansion agrees exactly with the power series from which it was derived up to a higher power of z than does the series obtained from any other rational function with the same or lower degrees of numerator and denominator. The Padé approximant generally has, however, a much larger region of convergence than that of the original series (assuming the original series was not that of an entire function), and it usually yields a much better approximation to $P(z)$, whether the original series is convergent or divergent, than do the equivalent S_m's.[3,5-8] Here we shall be principally concerned with approximants of the form $[n,n] = \epsilon_{2n}^{(0)}$, the final result of the application of the EA to an odd number of inputs. Baker, Gammel, and Wills[7] have given transformations of $P(z)$ which take it into a form for which $[n,n]$ goes to the exact value of the transformed function as n approaches infinity. Where appropriate, we assume such a transformation has been made. Figure 2 shows the important $s =$ even integer columns of the EA array when $2n+1 = 7$. The $[N,M]$ approximants also shown in Fig. 2 form part of the Padé table. Its lower half may be obtained through the use of transformations discussed by Shanks[3] and Wynn.[8]

It is of interest to note that determinantal quotients for Padé approximants were first derived by Frobenius[9] and were much later rediscovered independently by Schmidt[10] and Shanks.[3] The EA, devised by Wynn,[4] furnishes a generally more convenient method of calculating these approximants. The connection between the Padé table and the approximants of continued

$\epsilon_0^{(0)} = [0,0]$

$\epsilon_0^{(1)} = [0,1]$ $\quad \epsilon_2^{(0)} = [1,1]$

$\epsilon_0^{(2)} = [0,2]$ $\quad \epsilon_2^{(1)} = [1,2]$ $\quad \epsilon_4^{(0)} = [2,2]$

$\epsilon_0^{(3)} = [0,3]$ $\quad \epsilon_2^{(2)} = [1,3]$ $\quad \epsilon_4^{(1)} = [2,3]$ $\quad \epsilon_6^{(0)} = [3,3]$

$\epsilon_0^{(4)} = [0,4]$ $\quad \epsilon_2^{(3)} = [1,4]$ $\quad \epsilon_4^{(2)} = [2,4]$

$\epsilon_0^{(5)} = [0,5]$ $\quad \epsilon_2^{(4)} = [1,5]$

$\epsilon_0^{(6)} = [0,6]$

FIG. 2. Epsilon algorithm array for seven $\epsilon_0^{(m)}$ inputs. Only even-integer columns are shown, and the quantities in brackets are Padé approximants.

fractions is well known[11]; further, Wynn[8] has given sequences of Padé convergents which are successive convergents of certain types of continued fractions. Thus, the EA is a recursive device for transforming the partial sums of a power series into the convergents of its corresponding continued fraction.[12] Therefore, previous treatments of such physical problems as summation of infinite imaging series[13] or the calculation of Regge poles[14] by continued fraction methods could, alternatively, be treated by calculating Padé approximants directly, for example, by using the EA.

Although the EA is particularly appropriate for calculating numerical values of the $[r, r+m]$ Padé approximants to $P(z)$, it may also be used directly to obtain the algebraic form of $[r, r+m]$. A much more practical procedure (for appreciable r, m) in this case has, however, been outlined in detail by Wynn,[8] and other rational approximation approaches are discussed in the papers listed in Ref. 15. Analytical Padé approximants have recently been applied to several physical

[5] P. Wynn, Math. Comp. **15**, 151 (1961).
[6] G. A. Baker, Jr., and J. L. Gammel, J. Math. Anal. Appl. **2**, 21 (1961).
[7] G. A. Baker, Jr., J. L. Gammel, and J. G. Wills, J. Math. Anal. Appl. **2**, 405 (1961).
[8] P. Wynn, Math. Comp. **14**, 147 (1960).
[9] G. Frobenius, J. für Math. **90**, 6 (1881).
[10] R. J. Schmidt, Phil. Mag. **32**, 369 (1941).
[11] H. S. Wall, *Analytic Theory of Continued Fractions* (D. Van Nostrand Company, Inc., Princeton, New Jersey, 1948), pp. 377–413.
[12] The transformation of divergent power series into convergent continued fractions has been studied by (among others) L. Euler, Nova Acta Acad. Sci. Imper. Petropol. *pro annum* 1784; A. Markoff, Ann. Math. **19**, (1895); T. J. Stieltjes, Ann. Fac. Sci. Toulouse **3**, 1 (1889); and T. Carleman, *Les Functions Quasi-analytiques* (Gauthiers Villars, Paris, 1926).
[13] C. V. Jones, Proc. Inst. Elec. Engrs. (London) **106C**, 140 (1959).
[14] C. Lovelace and D. Masson, Nuovo Cimento **26**, 472 (1962).
[15] H. J. Maehly, J. Assoc. Comput. Mach. **7**, 150 (1960); **10**, 257 (1963); E. W. Cheney and T. H. Southard, SIAM Rev. **5**, 219 (1963); R. Sankar and V. Malini, Math. Comp. **17**, 414 (1963); J. L. Walsh, J. Math. Mech. **13**, 305 (1964); D. A. Pierre, J. Soc. Ind. Appl. Math. **12**, 93 (1964).

problems,[6,16–18] especially certain aspects of the Ising model. For present purposes, we shall only consider numerical rather than algebraic transformations effected by means of the EA.

The EA is not strongly nonlinear, as the following useful properties show. If all S_m's are multiplied by a constant C, then all $\epsilon_s^{(m)}(s=1, 2\cdots)$ are correspondingly multiplied by C. If C is added to all S_m's, all resulting $\epsilon_s^{(m)}$'s are made up of C plus the corresponding $\epsilon_s^{(m)}$'s obtained by operating on the original S_m's. The nonlinearity shows up, however, in the result that if $(\epsilon_s^{(m)})_1$ is obtained from p sequential values of $(S_m)_1$ and $(\epsilon_s^{(m)})_2$ from p sequential values of $(S_m)_2$, $(\epsilon_s^{(m)})_3$ obtained by operating on the p sequential $(S_m)_1+(S_m)_2$ values is not equal to $(\epsilon_s^{(m)})_1+(\epsilon_s^{(m)})_2$, although both are approximations to the desired result. Of course, in the limit $s=2n\rightarrow\infty$, these quantities are equal when the EA is applicable.

INITIAL ILLUSTRATIONS AND APPLICATIONS

Let us first consider a rational function of z whose denominator is a polynomial of degree $M(M>0)$. When this polynomial is divided out, an infinite series is obtained which may be summed exactly by the EA if $2n+1\geq 2M+1$. If $2M+1$ (or more) sequential partial sums are used as inputs, the exact sum will be given by $\epsilon_{2M}^{(0)}$ since the corresponding $[M,M]$ Padé convergent will involve the original polynomial in its denominator. Thus, for example, the EA will sum a (convergent or divergent) geometric series exactly with only $n=1$ since such a series is generated by expansion of $(1\pm z)^{-1}$. The value of the series for $(1\pm z)^{-2}$ for some specific value of z correspondingly requires $n=2$ for an exact summation since the degree of the denominator polynomial is $M=2$. The above property may be employed to yield a useful test of whether a given series, whose closed-form sum is unknown, is associated with a rational function or not.

When $P(z)$ is irrational, $\epsilon_{2n}^{(0)}$ still generally gives a better and better approximation as n is increased, but only as $n\rightarrow\infty$ does $\epsilon_{2n}^{(0)}$ yield the exact result. If a series is considered which arises from an unknown irrational function whose degree of algebraic irrationality p/q is known, however, the series may be transformed to that of a rational function by raising it to the q/p power. The EA may then be used to obtain the exact sum with finite n. In cases of practical interest, it is easier to apply the EA directly and increase n until the desired accuracy of the result is obtained. When the series under consideration is associated with a transcendental function $P(z)$, the exact sum is only approached as $n\rightarrow\infty$ since a finite number of algebraic operations cannot yield a transcendental number exactly.

TABLE I. Accuracy of epsilon algorithm results for $\ln(1+z)$. Numbers in table are: digits of error/significant figure of error.

	$z=1$			
$2n+1$	(a)	(b)	$z=2$	$z=10$
1	3/1	3/1	1/1	8/1
3	1/1	7/3	4/3	1/1
5	9/2	2/4	3/4	2/2
7	7/2	5/6	2/5	6/3
9	5/2	1/7	1/6	2/3
11	4/2	2/8[a]	8/8	5/4
Iteration	···	2/8[a]	0/8	1/4

[a] Round-off error only.

The EA is of most value for series which approach the geometric type as m increases so that the ratio of successive terms approaches a constant. It will accelerate the convergence (or yield the antilimit of a divergent series) of series for which this condition is not fulfilled, however, such as

$$\sum_{n=1}^{\infty}(n!)^{-1} \quad \text{or} \quad \sum_{n=0}^{\infty}(-1)^n n!.$$

An illustration of the power of the EA is given in Table I where it is applied to the series calculation of the transcendental $\ln(1+z)$ for $z=1$, 2, and 10. The series only converges for $z\leq 1$, and column (a) for $z=1$ shows the slow increase in accuracy obtained directly from the partial sums. Column (b) and the other columns show the result of computer calculation of $\epsilon_{2n}^{(0)}$ using eight decimal places in the calculation. Thus, for example, for $z=1$ the EA with $2n+1=7$ yields an error of 5 digits in the sixth significant figure of the sum while S_7 shows an error of 7 digits in the second significant figure. As z increases, the convergence of the EA results toward the exact value is retarded, but it is clear that it has still been of great help in yielding a close value of ln11 from the divergent series whose 11th partial sum is approximately 10^{10}.

The final values marked "iteration" in Table I have been obtained by treating the results $\epsilon_{2n}^{(0)}\equiv{}^1\epsilon_{2n}^{(0)}$ $(n=0, 1, 2, \cdots, 5)$ as new partial sum inputs, ${}^2\epsilon_0^{(n)}={}^1\epsilon_{2n}^{(0)}$, for the EA. Application of the EA to this upper diagonal in many, if not all, cases[19] leads to an even more accurate value for the sum. If the original value of $2n+1$ is large enough and sufficient significant figures are carried in the calculation, the process can often be further repeated with additional gain in accuracy.

Wynn[20] has shown that the EA is a "stronger" method for accelerating convergence than the Euler transformation. It should be mentioned that for functions of unbounded variation such as e^z, where for large $|z|$ the individual terms in the series expansion increase in magnitude for some time before they begin to de-

[16] G. A. Baker, Jr., Phys. Rev. **124**, 768 (1961).
[17] G. A. Baker, Jr., Phys. Rev. **129**, 99 (1963).
[18] J. W. Essam and M. E. Fisher, J. Chem. Phys. **38**, 802 (1963).
[19] P. Wynn, Chiffres 4, 19 (1961).
[20] P. Wynn, Chiffres 4, 23 (1961).

crease, the $2n+1$ number of terms used in the application of the EA must be large enough that an appreciable number of the last terms used have begun to decrease in magnitude if the EA is to yield useful results.

There are a number of methods available for obtaining an exact explicit closed form for the sum of a series.[21] They are not generaly applicable, but even when they are it may sometimes be more convenient to approximate the (unknown) function associated with the series by a Padé convergent or to use the EA to obtain approximate numerical values for the sum. Another application is the evaluation of convergent integrals which cannot be conveniently integrated in closed form. Rather than calculate them by ordinary numerical integration, sufficient functions in the integrand may be expanded in power series until it is possible to integrate term by term. The resulting infinite series may be divergent for values of the dependent variable of interest, but the value of the integral, particularly if the integral is of bounded variation, may still be evaluated quite accurately and rapidly using the EA. Alternatively, if the function represented by the integral must be further manipulated analytically, it may be convenient to approximate it from the power series by an analytic Padé convergent.

The EA is a local operator as shown by Eq. (3). This equation may be transformed to show that $\epsilon_{2s+2}{}^{(m-1)}$ may be expressed in terms of $\epsilon_{2s}{}^{(m-1)}$, $\epsilon_{2s}{}^{(m)}$, $\epsilon_{2s}{}^{(m+1)}$, and the two values $\epsilon_{2s-1}{}^{(m+1)}$ and $\epsilon_{2s-1}{}^{(m)}$ in the preceding column. The new transformation is

$$\epsilon_{2s+2}{}^{(m-1)}=\frac{D_{2s+2}{}^{(m-1)}}{1+E_{2s+2}{}^{(m-1)}}+\left[\frac{E_{2s+2}{}^{(m-1)}}{1+E_{2s+2}{}^{(m-1)}}\right]\epsilon_{2s}{}^{(m)}, \quad (4)$$

where

$$\begin{aligned} D_{2s+2}{}^{(m-1)} &\equiv \frac{\epsilon_{2s}{}^{(m+1)}\epsilon_{2s}{}^{(m-1)}-(\epsilon_{2s}{}^{(m)})^2}{\epsilon_{2s}{}^{(m+1)}-2\epsilon_{2s}{}^{(m)}+\epsilon_{2s}{}^{(m-1)}} \\ &= \epsilon_{2s}{}^{(m+1)}-\frac{(\epsilon_{2s}{}^{(m+1)}-\epsilon_{2s}{}^{(m)})^2}{\epsilon_{2s}{}^{(m+1)}-2\epsilon_{2s}{}^{(m)}+\epsilon_{2s}{}^{(m-1)}}, \end{aligned} \quad (5)$$

and

$$E_{2s+2}{}^{(m-1)}\equiv[\epsilon_{2s-1}{}^{(m+1)}-\epsilon_{2s-1}{}^{(m)}]\times[D_{2s+2}{}^{(m-1)}-\epsilon_{2s}{}^{(m)}]. \quad (6)$$

When $E_{2s+2}{}^{(m-1)}=0$, (4) shows that $\epsilon_{2s+2}{}^{(m-1)}=D_{2s+2}{}^{(m-1)}$, where $D_{2s+2}{}^{(m-1)}$ is just the Aitken δ^2 process.[22] Thus, when $s=0$ and $\epsilon_{-1}{}^{(m+1)}$ and $\epsilon_{-1}{}^{(m)}$ are zero, the condition $\epsilon_{2s-1}{}^{(m+1)}=\epsilon_{2s-1}{}^{(m)}$ is satisfied and the $2s+2=2$ column of the EA is the same as that obtained by the Aitken δ^2 process. Equation (4) shows, however, that, in general, even EA columns will differ from those obtained by repeated application of the δ^2 process. Note that (4) may be readily modified to yield a similar expression for quantities such as $\epsilon_{2s+1}{}^{(m-1)}$. Although the process of calculating individual values in a $2s+2$ or $2s+1$ EA column requires only local values of quantities in earlier columns, it is important to emphasize that $\epsilon_{2n}{}^{(0)}$, frequently the desired end result of operating on $2n+1$ S_m's, is a global, nonlinear function of all the inputs.

The δ^2 process has been applied to speed the convergence of convergent iterative sequences[22,23] and has even been used with divergent iterative sequences by Samuelson[23] and Wegstein.[24] Such sequences may, for example, arise from the iteration $z_{m+1}=f(z_m)$ $(m=0, 1, 2\cdots)$. The EA is applicable to all such situations, is usually superior to a single or iterated δ^2 process, and has been applied to complicated iterative schemes for operational formulas,[5,25] ordinary differential equations,[25,26] integral equations,[25–27] and iterated vector and matrix sequences.[26] In such iterations, which are a type of extrapolation, any consecutive sequence of iterates may be used as inputs to the EA.

APPLICATION TO EXTRAPOLATION AND CURVE FITTING

We shall now consider the application of the EA to extrapolation and curve fitting. It has been shown[4] that if $\epsilon_{-1}{}^{(m)}=0$, $\xi_i\neq 0$, and

$$\epsilon_0{}^{(m)}=a_0+\sum_{r=1}^{n} a_r\xi_r{}^m (m=0,1,2,\cdots), \quad (7)$$

then $\epsilon_{2n}{}^{(m)}=a_0$, $\epsilon_{2n+1}{}^{(m)}=\infty$ $(m=0, 1, 2\cdots)$. Note that $(2n+1)$ input values are necessary to yield the exact value of a_0. Further, a single δ^2 process yields a_0 only in the simpler case $n=1$ and is usually stated[28] to apply for the condition $|\xi_1|<1$.

Equation (7) includes the situation where $\epsilon_0{}^{(m)}$ is made up of a constant and sum of exponentials, as shown by setting $\xi_i\equiv\exp(\lambda_i)$ and replacing m on the right by $z\equiv z_0+m\Delta z$. It turns out that the EA may be used to obtain all the unknown coefficients and λ_i exponents in such a representation if sufficient accurate $\epsilon_0{}^{(m)}$ values are available. A number of techniques have been developed for finding these parameters,[29] and here

[21] A. D. Wheelon, J. Appl. Phys. **25**, 113 (1954) and references cited therein.

[22] A. C. Aitken, Proc. Roy. Soc. Edinburgh **46**, 289 (1926); *ibid*, **63A**, 174 (1951). See I. Marx, J. Math. Phys. **42**, 334 (1963) for sufficient conditions for convergence of the δ^2 process.

[23] P. A. Samuelson, J. Math. Phys. **24**, 131 (1945).

[24] J. H. Wegstein, Comm. ACM **1**, 9 (1958).

[25] D. Shanks, Naval Ordnance Laboratory Memorandum No. 9994, 26 July 1949, White Oak, Maryland.

[26] P. Wynn, Math. Comp. **16**, 301 (1962).

[27] J. S. R. Chisholm, J. Math. Phys. **4**, 1506 (1963).

[28] A. S. Householder, *Principles of Numerical Analysis* (McGraw-Hill Book Company, Inc., New York, 1953), p. 117; J. Todd, *Survey of Numerical Analysis* (McGraw-Hill Book Company, Inc., New York, 1962), p. 5.

[29] E. T. Whittaker and G. Robinson, *The Calculus of Observations* (Blackie & Son, London, 1944); C. Lanczos, *Applied Analysis* (Prentice Hall, Inc., Englewood Cliffs, New Jersey, 1956); G. R. Keepin, T. F. Wimett, and R. K. Zeigler, J. Nucl. Energy **6**, 1 (1957); D. G. Gardner, J. C. Gardner, G. Laush, and W. W. Meinke, J. Chem. Phys. **31**, 978 (1959); J. Cornfield, J. Steinfield, and S. W. Greenhouse, Biometrics **16**, 212 (1960); R. G. Cornell, Biometrics **18**, 104 (1962); B. H. Worsley, Comm. ACM **7**, 39 (1964).

we shall only mention that the EA method involves the limit as $m \to \infty$ of ratios of such terms as $\epsilon_{2r}^{(m+1)}$ and $\epsilon_{2r}^{(m)}$. Since these ratios themselves will approach a form of (7) as m increases, (7) may be applied to sequences of such ratios to obtain the required limits even with data for a z region where a single exponential has not yet become dominant.

The situation is considerably simpler when polynomials are considered. Let us first consider noise or error-free data. If we substitute $\xi_r \equiv z^r$ in (7), we obtain

$$\epsilon_0^{(m)} = a_0 + \sum_{r=1}^{n} a_r z^{rm} (m = k, k+1, k+2, \cdots), \quad (8)$$

where k is an integer, or

$$\left.\begin{aligned} &\cdots\cdots\cdots\cdots\cdots\cdots \\ \epsilon_0^{(0)} &= a_0 + a_1 + a_2 + a_3 + \ldots + a_n \\ \epsilon_0^{(1)} &= a_0 + a_1 z + a_2 z^2 + a_3 z^3 + \ldots + a_n z^n \\ \epsilon_0^{(2)} &= a_0 + a_1 z^2 + a_2 z^4 + a_3 z^6 + \ldots + a_n z^{2n} \\ \epsilon_0^{(3)} &= a_0 + a_1 z^3 + a_2 z^6 + a_3 z^9 + \ldots + a_n z^{3n} \\ &\cdots\cdots\cdots\cdots\cdots\cdots \end{aligned}\right\} \quad (9)$$

Thus $\epsilon_0^{(m)}$ is a polynomial in z, and $|z|$ need not be less than unity. The quantity a_0 may now be obtained exactly using the $2n+1$ input values of $\epsilon_0^{(m)}$ associated with z^m values which change in geometrical steps. Unlike most other methods, the z values themselves are not used in the determination of a_0.

In many physical or digital computing situations, it is expected that the measured or computed quantity will well approximate to a truncated Maclaurin series, such as that of (8), as $(|z|^r)^m$ becomes smaller and smaller (m increasing if $|z|<1$; m decreasing if $|z|>1$). Frequently, it is impossible to measure a physical quantity in the limit of some parameter, such as zero absolute temperature or infinite magnetic field strength. Alternatively, when a very complicated function is being computed which begins to approximate to the form of (8) as $(|z|^r)^m$ becomes smaller, it frequently happens that the limit when $(|z|^r)^m \to 0$ cannot be calculated analytically and further that computing time (and cost) increases strongly as $(|z|^r)^m$ is decreased. In both of these cases, (8) may be employed to get a good estimate of the final extrapolated value a_0.

When the function to be extrapolated is only represented by (8) with $n=\infty$, it is obvious that increasing the number of accurate $\epsilon_0^{(m)}$ input values used will progressively increase the accuracy with which a_0 is obtained provided the infinite series converges for all z^m values. If, on the other hand, (8) with $n<\infty$ (yielding a simple polynomial) is an exact representation of the function measured or computed and the function values are exact, then the a_0 obtained with $2n+1$ input values is exact (except for possible round-off errors).

With accurate data, the EA may also be used as a substitute for least squares to obtain all a_i values, not just a_0. After obtaining a_0 as above, one first forms

$$\epsilon_0^{(m)}/z^m = (a_0/z^m) + \sum_{r=1}^{n} a_r z^{(r-1)m}$$

and applies the EA to this new sequence, yielding a_1 providing $2n+1 \geq 3$ so that the inverse terms contribute nothing. Continuation of this process yields all remaining a_i's. In actual applications of this process to exact cubic data, the EA yielded values for the coefficients correct to from one to two more significant figures than did least squares. Reasons for this result will be discussed later.

In all practical cases of extrapolation and polynomial curve fitting, the data will be contaminated with experimental, or round-off and truncation errors. Now it is known that least squares gives the best linear unbiased estimates (BLUE) of the coefficients of a linear regression equation when the Gauss–Markoff assumptions[30] concerning the data are well satisfied. For all polynomial fitting with degrees greater than unity, however, it has not been proved that a BLUE is a maximum likelihood estimator. In fact, one might guess that for such problems an adaptive, nonlinear process might exist which would yield better estimates of the polynomial coefficients than would even least squares with fixed weighting. Of course, it would not then yield a minimum sum of squared deviations, but it might yield better coefficient estimates just by discriminating nonlinearly against data points with appreciable error (operationally defined as deviation from regularity).

The EA is a nonlinear process and involves automatic adaptive weighting. Therefore, it is of interest to compare its performance on contaminated data with that obtained from least squares (LS). When $2n+1$ values of a dependent variable with error are available as well as the associated values of the independent variable, the degree of the underlying error-free polynomial will frequently be unknown. We shall here assume that it is finite or may be so approximated or, alternatively, that the infinite series in question converges for all values of the independent variable used.

In the application of LS with increasing polynomial degree p, the resulting standard deviation s (involving the sum of squared deviations, the number of input values, and the polynomial degree) should, for exact polynomial input data, reach a minimum (zero except for round-off error) at the degree of the input polynomial, and should increase slightly with succeeding higher degrees because of the presence of the degree in the definition of s. For exact polynomial data contaminated with errors, however, s should generally continually decrease until, with the maximum polynomial degree possible for a given set of input data (one less

[30] O. Kempthorne, *The Design and Analysis of Experiments* (John Wiley & Sons, Inc., New York, 1952), p. 32.

than the number of inputs), calculated values for the dependent variable agree exactly (except for round-off errors) with input values, and s is then indeterminate.

Unfortunately, the standard computer LS programs, which use a modified Gauss elimination technique for solving the LS normal equations, do not always yield results which conform with the above expectations. For exact data, s does reach a minimum at the correct p but thereafter it increases strongly with degree. Even when computation is carried on with floating-point numbers approximately 25 decimal digits long, s is found to increase by orders of magnitude as p increases beyond the correct value. In this case, the desired coefficients are found to remain correct to far more than eight significant figures but, even so, small changes in them and the appearance of very small higher coefficients change s appreciably.

For noisy polynomial data, more than one relative minimum in s may be observed as p increases, even when the numbers used have a great many decimal digits so that the coefficients obtained are highly accurate solutions of the normal equations. Even more peculiar behavior may be expected if the estimated a_i's, denoted as $\hat{a}_i$, are inaccurate solutions of the normal equations. For exact or noisy polynomial data with $a_0 \neq 0$, we have found, by comparing results obtained with differing computer word lengths, a simple approximate relation between the number of accurate decimal digits which may be expected in the LS $\hat{a}_i$'s, A, the number of decimal digits carried in the calculation, C, and p. The relation is

$$A \simeq C - p \pm 1 \qquad \text{(Gauss elimination)}. \qquad (10)$$

This equation holds for both geometric and equal spacing of the independent variable and applies to all $\hat{a}_i$'s of significance; few if any will be better than $A \simeq C-p$; few if any will be worse than $A \simeq C-p-1$. The use of an LS program based on orthogonal polynomials, which avoids matrix manipulations, has been found, however, to yield the considerably superior results,

$$A \simeq C - p + 1 \pm 1 \qquad \text{(orthogonal polynomials, geometric spacing)}, \qquad (11)$$

$$A \simeq C - p + 2 \pm 1 \qquad \text{(orthogonal polynomials, equal spacing)}. \qquad (12)$$

Although most of the LS results discussed later were calculated using the Gauss elimination method, care was taken to ensure that all $\hat{a}_i$'s presented for noisy data are accurate solutions of the normal equations to at least the number of decimal digits quoted.

Since noisy data LS $\hat{a}_i$'s vary with p even when adequate precision (i.e., solution accuracy) is ensured by proper selection of C and p, it is important to pick that p which will yield sufficiently accurate $\hat{a}_i$'s closest to the a_i's of the underlying exact polynomial. Such a choice is, of course, easy with exact data, but no general method of making it is known to the author for noisy data. Here, we shall generally use that p which leads to the first relative minimum in s. This choice will by no means always be the p of the underlying exact polynomial, generally unknown in practical cases, or that p which gives the best $\hat{a}_0$.

The above problems do not arise with the EA because it generates polynomials which have the useful property that although all coefficients depend globally on all input values, higher-order coefficients are independent of all lower-order ones. Thus, the calculation of a new coefficient does not affect the values or accuracy and precision of the earlier ones and $\hat{a}_0$, for example, does not vary with p. One may continue to extract new coefficients indefinitely although they will eventually become meaningless.

If the $2n+1$ number of input values is appreciable and/or may be increased at will, as in usual experimental or calculational situations, a test of the adequacy of either the EA or least-squares determination of $\hat{a}_0$ is to make several calculations with different values of $2n+1$ (and/or partitioning of the total available input set) and observe the stability of the resulting $\hat{a}_0$'s. If the noise in the data is not too great, considerable stability should be found after $2n+1$ exceeds a certain minimum number, M. This same comparison may be applied to higher-order $\hat{a}_i$'s to decide when to stop extracting further coefficients. When $\hat{a}_i$ begins to vary strongly and randomly with $2n+1 \gg M$, the $\hat{a}_i$ in question should be discarded as meaningless. When several such $\hat{a}_i$'s are found in succession, the process may be terminated. Should the noise or error in the data seem considerable, a smoothing process should be used before applying either the EA or least squares. Such smoothing, if properly done, will usually improve the value of $\hat{a}_0$, but one runs the risk of smoothing out true variations in the data and thus making the estimates of some of the higher-order coefficients worse. Smoothing of this type is equivalent to a filtering process in the time domain, whether abscissa values are distributed at equal or geometric intervals, and the techniques of optimum linear filter theory may be used whether the data is a discrete time series or not. In the same sense, the EA process can be employed to yield a type of nonlinear filtering.

To test the applicability of the EA and LS for extrapolation (estimation of a_0) and polynomial higher coefficient estimation on data with errors, we first generated with a computer normally distributed quasi-random errors from a population of variance σ^2. For $\sigma^2 = 10^{-3}$, for example, the largest error magnitude appearing was about 0.06 and the smallest 0.002. These errors were then either added, or added proportionally, to exact (except for round-off) values of the quadratic $y = 1+z+0.1z^2$, evaluated at geometric intervals between $z = \frac{1}{8}$ and 16 using $\alpha \equiv \sqrt{2}$ as the constant multiplier. We can represent the exact polynomial for this

case by (8) with $z=\alpha$, $n=2$, $k=-6$, and $m_{\max}=8$. Typical equations for y_m for m near zero [compare Eq. (9)] are then

$$\left.\begin{aligned} &\cdots\cdots\cdots\cdots\cdots \\ y_{-1}&=a_0+(1/\sqrt{2})a_1+(1/2)a_2 \\ y_0&=a_0+a_1+a_2 \\ y_1&=a_0+\sqrt{2}a_1+2a_2 \\ y_2&=a_0+2a_1+4a_2 \\ &\cdots\cdots\cdots\cdots\cdots, \end{aligned}\right\} \tag{13}$$

where $a_0=a_1=1$ and $a_2=0.1$, and these three quantities are the unknowns in an EA or LS run employing exact data associated with the quadratic. To make the test using noisy data as stringent as possible, no initial smoothing was carried out. Although only $2n+1=5$ $\epsilon_0^{(m)}\equiv y_m$ values would be needed for exact data, we used a maximum of $15(=m_{\max}-k+1)$ to investigate any EA smoothing of errors.

Clearly, when of 15 EA inputs even one contains an error from, say, exact quadratic data, the resulting data points can, in general, only be fitted exactly by a 14th or higher degree polynomial. To obtain the corresponding exact a_i's (different from the exact quadratic a_i's) for this 14th degree polynomial would require 29 EA inputs associated with it. Since these are unavailable in practical cases, the most important remaining question is how well do the $\hat{a}_i$'s obtained from the 15 not all error-free data points approximate the original exact quadratic a_i's.

Extensive numerical results for different error σ^2 values have been calculated with a digital computer. A small part of the results is summarized in Table II. Here, all noise or error data were drawn from a population with $\sigma^2=10^{-3}$. This results in errors in y generally larger than those encountered in a good experiment. The values of $\hat{a}_i$ shown in Table II were estimated by the methods listed in column 2. All EA results shown were calculated with $C=8$, and the first set of LS results in each numbered row was also obtained with $C=8$ and is for that p which led to the first minimum in s. The second set of LS results was calculated (and checked by the orthogonal polynomial method, frequently using double precision: $C\simeq 25$) with $C\simeq 11$ and $p=6$ in order to show the change of coefficients with degree and to yield higher LS $\hat{a}_i$'s for comparison with higher EA $\hat{a}_i$'s. Note that in the large majority of cases, the EA $\hat{a}_i$'s ($i\geq 3$) are considerably smaller in magnitude than the corresponding LS $\hat{a}_i$'s and hence are better estimates of the higher $\hat{a}_i$'s even though the EA results have been obtained using a smaller C than the $p=6$ LS results.

It is evident from rows 2 and 3 that for exact (except for round-off) data, LS generally yielded slightly better $\hat{a}_i$'s ($0\leq i\leq 2$) than did the EA. The appearance of nonzero EA $\hat{a}_i$'s ($i\geq 3$) stems from the necessity of computing with a finite C and also because the input data in this case had only eight decimal digits and was not always quite exact.

All the results in Table II were calculated with 15 input values of y except those of rows 8–12, which represent partitioned input data of seven values each. The results shown in rows 4–15 all were obtained from input values using the same noise or error set. This was added directly for 4–13 and added in proportionately to obtain the input data used in obtaining the results in rows 14 and 15. If we had an exact estimation procedure, all $\hat{a}_i$'s would equal the exact a_i's shown in row 1.

If 29 proper input values were available so that a 14th degree polynomial could be fitted exactly by the EA, then subtraction of any constant from all the input values and then division of each resulting value by the appropriate z^m would make no difference in the final $\hat{a}_0\equiv\epsilon_{28}^{(0)}$ and higher $\hat{a}_i$'s. In the present noisy case with only 15 input values, the final EA value of a run, $\epsilon_{14}^{(0)}$, is not generally equal to $\epsilon_{28}^{(0)}$, and the above subtractions can and will make a difference in $\epsilon_{14}^{(0)}$ and thus in the values of $\hat{a}_i$ obtained. The results in rows 5–7 illustrate the effect. In row 5, which yields the best EA $\hat{a}_i$ results, the exact a_i values were subtracted before division by z^m in order to obtain higher $\hat{a}_i$ values. Thus, to obtain $\hat{a}_i$, $a_0=1$ was first subtracted from the polynomial values before dividing by z^m and applying the EA to obtain the $\hat{a}_1$ shown, etc. In row 6, the EA results, e.g. $\hat{a}_0=1.0712$, etc., were subtracted at each stage. This procedure generally turned out to be worst, as shown. Finally, in row 7 and in all other EA runs, no subtraction at all was carried out, in accordance with the method described earlier to obtain all higher a_i's. Since the a_i's will be unknown in practical cases of interest, the method of row 5 will be inapplicable and that of row 7 appears to be the best available.

Comparison of rows 4 and 7 again shows that the LS results for $\hat{a}_i$ ($0\leq i\leq 2$) are generally superior to those of the EA. This was usually found to be the case for both exact and additive noisy data when the LS value of A (the number of accurate decimal digits obtained from an LS run) was high enough. For fixed C, the EA can clearly yield more accurate and precise values of the $\hat{a}_i$'s than can LS when the desired p is such that A is inadequate. Note that the EA $\hat{a}_i$'s ($i\geq 4$) of row 7 are considerably superior to the corresponding LS $\hat{a}_i$'s of row 4 even when, as here, the LS results are accurate solutions of the normal equations (sufficiently large A).

The input data for rows 8–11 were selected from the full noisy set of 15 elements in four different ways. P_1 used every other one, skipping the first two (smallest) ones. P_2 again used every other one but omitted the first value. P_3 used only the last seven values and P_2 used only the first (smallest) seven. Thus, P_1 and P_2, and P_3 and P_4 were nonoverlapping. It will be noted that in spite of the relatively small number of input terms used, the $\hat{a}_i$'s ($0\leq i\leq 2$) are relatively stable and near the exact values, while there is very appreciable instability in the $\hat{a}_i$'s ($i\geq 3$). Comparison of the latter with those of row 7, where the full 15 data points were used, also shows up large differences. These results are

TABLE II. Numerical results of the application of the EA and LS to originally quadratic data.

Data/Method	$\hat{a}_0$	$\hat{a}_1$	$\hat{a}_2$	$\hat{a}_3$	$\hat{a}_4$	$\hat{a}_5$	$\hat{a}_6$
1 Exact	1	1	0.1	0	0	0	0
2 Exact/LS	0.99999990 1.0000000	1.0000002 0.99999973	0.099999982 0.10000027	... -9.972×10^{-8}	... 1.634×10^{-8}	... 1.188×10^{-9}	... 3.099×10^{-11}
3 Exact/EA	0.99999979	1.0000049	0.099999917	-3.429×10^{-8}	-9.217×10^{-10}	-5.862×10^{-11}	-4.055×10^{-11}
4 Noisy/LS	0.9947 1.0094	1.0088 0.9611	0.0994 0.1281	... -6.334×10^{-3}	... 6.307×10^{-4}	... -2.887×10^{-5}	... 4.939×10^{-7}
5 Noisy/EA[a]	1.0712	0.9946	0.1008	-1.009×10^{-4}	-2.208×10^{-5}	-9.315×10^{-8}	-4.661×10^{-9}
6 Noisy/EA[b]	1.0712	-4.370×10^{-3}	1.6738×10^{-3}	2.469×10^{-3}	-3.029×10^{-5}	2.923×10^{-9}	-9.704×10^{-1}
7 Noisy/EA	1.0712	1.1040	0.1022	-1.210×10^{-2}	-3.572×10^{-5}	-8.797×10^{-7}	-3.282×10^{-8}
8 P_1-noisy/EA	1.0825	1.2046	0.1090	3.627×10^{-3}	2.306×10^{-4}	1.808×10^{-5}	1.443×10^{-6}
9 P_2-noisy/EA	1.1092	1.1809	0.1035	3.135×10^{-3}	1.764×10^{-4}	1.278×10^{-5}	1.029×10^{-6}
10 P_3-noisy/EA	1.5756	1.3209	0.1078	1.773×10^{-3}	5.108×10^{-5}	2.105×10^{-6}	1.041×10^{-7}
11 P_4-noisy/EA	0.9506	0.9276	0.4628	0.3519	0.3090	0.2906	0.2834
12 P_4-noisy/LS	0.9152 1.3647	1.8175 −4.4775	−1.7730 25.7243	1.124 −36.6670	... −21.424	... 82.565	... −45.00
13 Noisy[c]/EA	1.0616	1.1163	0.1045	2.981×10^{-3}	1.253×10^{-4}	6.526×10^{-6}	-3.382×10^{-7}
14 Pro-noisy/LS	1.003 1.0073	0.87 0.8640	0.21 0.2165	-2.7×10^{-2} -2.902×10^{-2}	2.5×10^{-3} 2.660×10^{-3}	-7.5×10^{-5} -7.997×10^{-5}	... 5.017×10^{-8}
15 Pro-noisy/EA	1.0544	1.0956	0.1086	2.413×10^{-3}	1.104×10^{-4}	7.120×10^{-6}	4.300×10^{-7}

[a] Subtracted $a_0=1$, $a_1=1$, $a_2=0.1$, $a_i=0$ for $i\geq 3$ to obtain successive $\hat{a}_i$'s shown.
[b] Subtracted $\hat{a}_0=1.0712$, $\hat{a}_1=-4.370\times10^{-3}$, $\hat{a}_2=1.6738\times10^{-3}$, etc., to obtain successive $\hat{a}_i$'s shown.
[c] Input data rounded off to three significant figures only.

sufficient to allow the conclusion that all a_i $(i\geq 3)$ should be zero. It will further be noted that in the vast majority of instances the results of row 7 are superior to those of rows 8 and 9 or 10 and 11, thus indicating the added accuracy arising from a larger input set. Data set P_4 was obviously rather peculiar. Comparison of rows 11 and 12 shows, moreover, that in this case the EA yielded results far superior throughout to those of LS.

Rows 7 and 13 should be compared. The only difference in the input data was that the data for 13 had been reduced to a word length of three from the original eight decimal digits per input. It will be noted that there is little change in $\hat{a}_i$ $(0\leq i\leq 2)$ but considerably more, as expected, in $\hat{a}_i$ $(i\geq 3)$. This method too could be used to help identify higher $\hat{a}_i$'s for which $a_i\equiv 0$.

Rows 14 and 15 show results for errors added proportionately. The resulting data will more nearly approximate experimental results in many cases. Note that here $p=5$ led to the first minimum in s. As shown, the first line of LS results is only accurate to approximately two significant figures, in agreement with $A\simeq C-p-1$. Out of five runs with differing noise components (but all drawn from a population with $\sigma^2=10^{-3}$), one minimum s was obtained with $p=3$ and all the rest required fifth-degree polynomials. It was usually found, as illustrated by rows 14 and 15, that LS gave better estimates of a_0 nevertheless, and the EA gave better estimates of all higher a_i's. More numerical analysis with different values of noise σ^2, different types of noise, and more replication would be desirable, but the present results are still sufficient to illustrate some of the valuable features of the EA. In addition to its obvious applications to the analysis of ordinary data, it may be used for the nonlinear smoothing and extrapolation of nonstationary time series.

ACKNOWLEDGMENTS

The author much appreciates numerous discussions with C. A. Barlow, Jr., and C. R. Ratliff, and the computer programming carried out by the latter. Comments by Dr. P. Wynn on an earlier version of the paper and his suggestion of several early references were very valuable.