Mathematics Notes

Note 51

7 March 1977

Fitting Noisy Data with a Complex Exponential Series

D. G. Dudley
Lawrence Livermore Laboratory

ABSTRACT

The process of fitting noisy data with a complex exponential series is considered. An analysis of the classical Prony method shows that attempts to obtain accurate answers in the presence of noise meet with considerable difficulties. Recasting the problem as one of system identification and applying the least-squares method leads to biased estimates of the parameters in the characteristic equation for the system. The difficulty centers around the fact that neither classical Prony nor least squares involves any analysis, identification, smoothing, or filtering of the noise. Some techniques for improving results are discussed, and an iterative generalized least-squares procedure, which leads to a noise filter, is recommended for further study.

CONTENTS

INTRODUCTION

The idea of fitting data with a complex exponential series is an old one, dating at least back to 1795. Many researchers in the present century have examined the idea, and much has been written on the subject. Workers involved with time-varying data are particularly interested since the method yields explicitly the poles of the system in the complex frequency plane.

Given the accuracy of modern high-speed computers, there is little doubt that the classical method (the Prony method), although well-known as an error magnifier, produces acceptable results when used in curve-fitting. Grave problems arise, however, when an attempt is made to use the method in fitting noisy data. In this application, Prony's method has many critics, whose complaints are largely justified by the erratic nature of many of the results. Nevertheless, the appeal of the method has transcended the criticism, and it continues to be used.

In this report, we consider the fitting of noisy data with exponentials both from the classical and also from the modern system identification points of view. We are able to derive some useful error criteria which definitize some of the difficulties. In addition, we recommend some techniques which we are hopeful might improve the present state of affairs. In particular, we recommend some filtering techniques which arise in doing an iterative, generalized least-squares procedure on the process.

CLASSICAL PRONY METHOD

The classical Prony method has been described by many authors. We follow essentially the development given by Weiss and McDonough.[1] Let g(t) represent a scalar process on which we are able to obtain data and let f(t) be a mathematical model which we shall attempt to make close to the process in some sense. For the model, we choose

$$f(t) = \sum_{j=1}^{N} A_j e^{s_j t} \tag{1}$$

and as criterion of closeness we force the model to be equal to the process at 2N equally spaced data points, viz.,

$$f(mT) = g(mT) \quad m = 0, 1, \ldots, 2N - 1 \tag{2}$$

where T is the sampling interval. We define a sequence of complex variables $\{z_j\}$ by

$$z_j = e^{s_j T} \quad j = 1, 2, \ldots, N, \tag{3}$$

and, without loss of generality, exhibit the $\{z_j\}$ as the roots of an N^{th} order polynomial, viz.,

$$\prod_{j=1}^{N} (z - z_j) = \sum_{m=0}^{N} a_m z^m \qquad (a_N = 1) \ . \tag{4}$$

Substitution of Eqs. (1) and (3) into Eq. (2) yields

$$g_m = \sum_{j=1}^{N} A_j \, z_j^m \qquad m = 0, 1, \ldots, 2N - 1, \tag{5}$$

where, for ease of notation, we use g_m in place of $g(mT)$. With the above framework established, we find that

$$\sum_{m=0}^{N} g_{m+k} \, a_m = 0 \qquad k = 0, 1, \ldots, N - 1, \tag{6}$$

which is easily verified by substituting Eq. (5) into Eq. (6), interchanging the order of the summations and applying Eq. (4). Since $a_N = 1$, Eq. (6) may be rearranged to produce N linear equations in N unknowns.

$$\sum_{m=0}^{N-1} g_{m+k} \, a_m = -g_{N+k} \qquad k = 0, 1, \ldots, N - 1, \tag{7}$$

which may be solved for the polynomial coefficients $\{a_m\}$. The polynomial in Eq. (4) may then be factored to give $\{z_j\}$. The exponent parameters $\{s_j\}$ are then given by

$$S_j = \frac{\ln z_j}{T} \, . \tag{8}$$

The final step is the determination of the $\{A_j\}$. Since the $\{z_j\}$ are now known, these are easily found from the first N equations in Eq. (5).

EXPONENTIAL FITTING BY A LEAST SQUARES PARAMETER ESTIMATION

In this section, we shall adopt a different point of view of the same basic problem. We again have a process g(t) and a model f(t), but in this case we define our model by the difference equation

$$\sum_{m=0}^{N} a_m \, f_{m+k} = 0 \qquad k = 0, 1, \ldots, M - 1. \tag{9}$$

The justification of this model definition is that we may prove by simple substitution that the difference equation, Eq. (9), has the solution

$$f_n = \sum_{j=1}^{N} A_j z_j^n \quad n = 0, 1, \ldots, M + N - 1, \tag{10}$$

where the $\{z_j\}$ are the roots of the characteristic equation

$$\sum_{m=0}^{N} a_m z^m = 0. \tag{11}$$

If we next let

$$z_j = e^{s_j T} \tag{12}$$

then the $\{s_j\}$ and the $\{A_j\}$ become the parameters to be determined just as before. There is, however, an important distinction. Here, the difference equation is of the model rather than of the process.

Astrom and Eykhoff,[2] in an exhaustive survey paper on system identification, have adopted a canonical form for a linear, discrete-time, single-input single-output model, and we shall follow their terminology. As a measure of closeness of the model f to the process g, we adopt a loss function V which is a functional of the model and the process, viz.,

$$V = V(f,g). \tag{13}$$

For the case herein, we choose the classical least-squares loss function defined by

$$V(f,g) = \sum_{k=0}^{M-1} e_k^2 , \tag{14}$$

where e_k is the generalized error, or "residual," which we define to be the difference between the difference equation describing the model, Eq. (9), and the same difference equation with the process substituted for the model, viz.,

$$e_k = \sum_{m=0}^{N} a_m g_{m+k} - \sum_{m=0}^{N} a_m f_{m+k} \tag{15}$$

$$= \sum_{m=0}^{N} a_m \left(g_{m+k} - f_{m+k}\right) \quad . \tag{16}$$

Substitution of Eq. (16) into Eq. (14) and application of Eq. (9) yields

$$V(f,g) = \sum_{k=0}^{M-1} \left(\sum_{m=0}^{N} a_m \, g_{m+k} \right)^2 \tag{17}$$

We next minimize the loss function V(f,g), subject to the constraint $a_N = 1$. (Some constraint is necessary in order to obtain other than the trivial solution $a_k = 0$, all k.) The minimization is carried out in Eq. (17) by requiring

$$\frac{\partial}{\partial a_m} = 0 \qquad m = 0, 1, \ldots, N - 1 \tag{18}$$

The resulting expressions are the "normal equations of least squares"

$$\sum_{k=0}^{M-1} \left(g_{k+N} + \sum_{n=0}^{N-1} a_n g_{k+n} \right) g_{k+m} = 0 \qquad m = 0,1,\ldots, N - 1 \tag{19}$$

The solution to Eq. (19) is obtained by solving the Hankel matrix equation

$$\begin{bmatrix} g_0 & g_1 & g_2 & \cdots & g_{N-1} \\ g_1 & g_2 & g_3 & \cdots & g_{N-1} \\ g_2 & g_3 & g_4 & \cdots & g_{N+1} \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ g_{M-1} & g_M & g_{M+1} & \cdots & g_{M+N-2} \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ \cdot \\ \cdot \\ a_{N-1} \end{bmatrix} = \begin{bmatrix} -g_N \\ -g_{N+1} \\ -g_{N+2} \\ \cdot \\ \cdot \\ \cdot \\ -g_{M+N-1} \end{bmatrix} \tag{20}$$

This matrix equation, for the case M = N, reduces to exactly the same result as in the classical Prony method given in Eq. (7).

The derivation given above allows us to consider Prony's method in an entirely different context than the classical one. In this sense, we shall define Prony's method as a particular system identification procedure in which the model parameters $\{a_i\}$ are estimated by minimizing a linear least-squares loss function. Such a view allows us to clearly define errors that have been observed in the Prony process. Unhappily, however, it casts doubt on results obtained by classical methods in the presence of noise.

SOLUTION BY THE MOORE-PENROSE PSEUDOINVERSE

In this section, we shall investigate the solution to the problem posed by Eq. (20). We shall attempt to determine the coefficients $\{a_j\}$, recognizing that once they are obtained, the solutions for the poles $\{s_j\}$ and the residues $\{A_j\}$ follow as outlined at the end of the section on the classical Prony method.

We write Eq. (20) in matrix form as

$$GA = \hat{G}, \tag{21}$$

where G is an M × N matrix, A is an N-term column vector of coefficients, and $\hat{G}$ is an M-term column vector. If we assume a priori that the rank of matrix G is N, the columns of G are linearly independent. Under this condition, Eq. (21) has the unique solution[3]

$$A = G^{+} \hat{G}, \tag{22}$$

where G^{+} is the Moore-Penrose pseudoinverse of G given by

$$G^{+} = \lim_{\delta \to 0} (G^{T}G + \delta^{2}I)^{-1} G^{T}, \tag{23}$$

which, in the case of column linear independence, reduces to

$$G^{+} = (G^{T}G)^{-1} G^{T}, \tag{24}$$

where G^T is the transpose of G. We state the following properties without proof[3]:

1. G^TG is square and symmetric

$$(G^TG)^T = G^TG \quad . \tag{25}$$

2. If rk(G) represents the rank of G, then

$$rk\ (G^TG) = rk(G) \quad . \tag{26}$$

3. If G is square and nonsingular,

$$G^+ = G^{-1} \quad . \tag{27}$$

Equation (24) and Eq. (22) show that the solution for the polynomial coefficients may be obtained by a matrix transposition, followed by a multiplication, followed by an inversion, followed by two more multiplications. These operations can be performed with most standard matrix computation routines. There remains the problem of determining the number of rows M in matrix G.

Matrix G, as indicated in Eq. (20), contains M rows and N columns. The number of columns is assumed to have been determined a priori and is equal to the rank of the matrix. The number of rows is determined by the number of data points we are utilizing in the process. If the process contained no errors and if we assume errorless computation, we would be justified in picking N = M and proceeding with the classic Prony method. In the presence of errors, however, the situation is not straightforward. Mann and Wald[4] have shown that if the residuals $\{e_k\}$ are independent and equally distributed with zero mean, then under certain mild restrictions the least-squares estimate of $\{a_k\}$ converges to the true parameter values as $M \to \infty$. The significance of this result is that adding more data improves the parameter estimation. Unfortunately, the estimation herein does not meet the requirements of Mann and Wald. Indeed, an examination of Eq. (16) shows immediately that the members of $\{e_k\}$ are not independent, since individual members of the sequence have noisy data terms in common. The result is that the residuals are correlated, a fact which results in biased parameter estimates.[2]

We test the biased parameter hypothesis with the following case. Let the process be given by

$$g(t) = e^{-\alpha t} \cos \omega t, \tag{28}$$

where

$$\alpha = 3.5 \times 10^{-4}$$

$$\omega = 0.25 \quad .$$

We assume that the accuracy of our computations is high enough so that the steps in the solution are noise-free. We use a 2 × 2 system and solve by classical Prony. Let T = 1 and N = 2. We obtain in Eq. (7), without showing all significant figures,

$$\begin{bmatrix} 1.00 & 0.97 \\ 0.97 & 0.88 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \begin{bmatrix} -0.88 \\ -0.73 \end{bmatrix} \tag{29}$$

A matrix inversion results in

$$\begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \begin{bmatrix} -14.34 & 15.84 \\ 15.84 & -16.35 \end{bmatrix} \begin{bmatrix} -0.88 \\ -0.73 \end{bmatrix} \tag{30}$$

which gives for the polynomial coefficients

$$\begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \begin{bmatrix} 1.001 \\ -1.938 \end{bmatrix} \tag{31}$$

We solve for the polynomial roots and obtain

$$\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \begin{bmatrix} 0.97 + i.25 \\ 0.97 - i.25 \end{bmatrix} \tag{32}$$

and with these obtain with no difficulty

$$S_1 = S_2^* = -3.5 \times 10^{-4} + i.25 \tag{33}$$

$$A_1 = A_2 = 0.5. \tag{34}$$

Substitution of Eqs. (33) and (34) into Eq. (1) reproduces the known result in Eq. (28).

We next add noise to the system by corrupting each individual data point, viz.,

$$g_k = f_k\ (1 + \delta_k), \tag{35}$$

where the δ_k are machine-generated Gaussian variates, independent, of zero mean, and of spread σ. In the presence of these errors, we attempt to reclaim

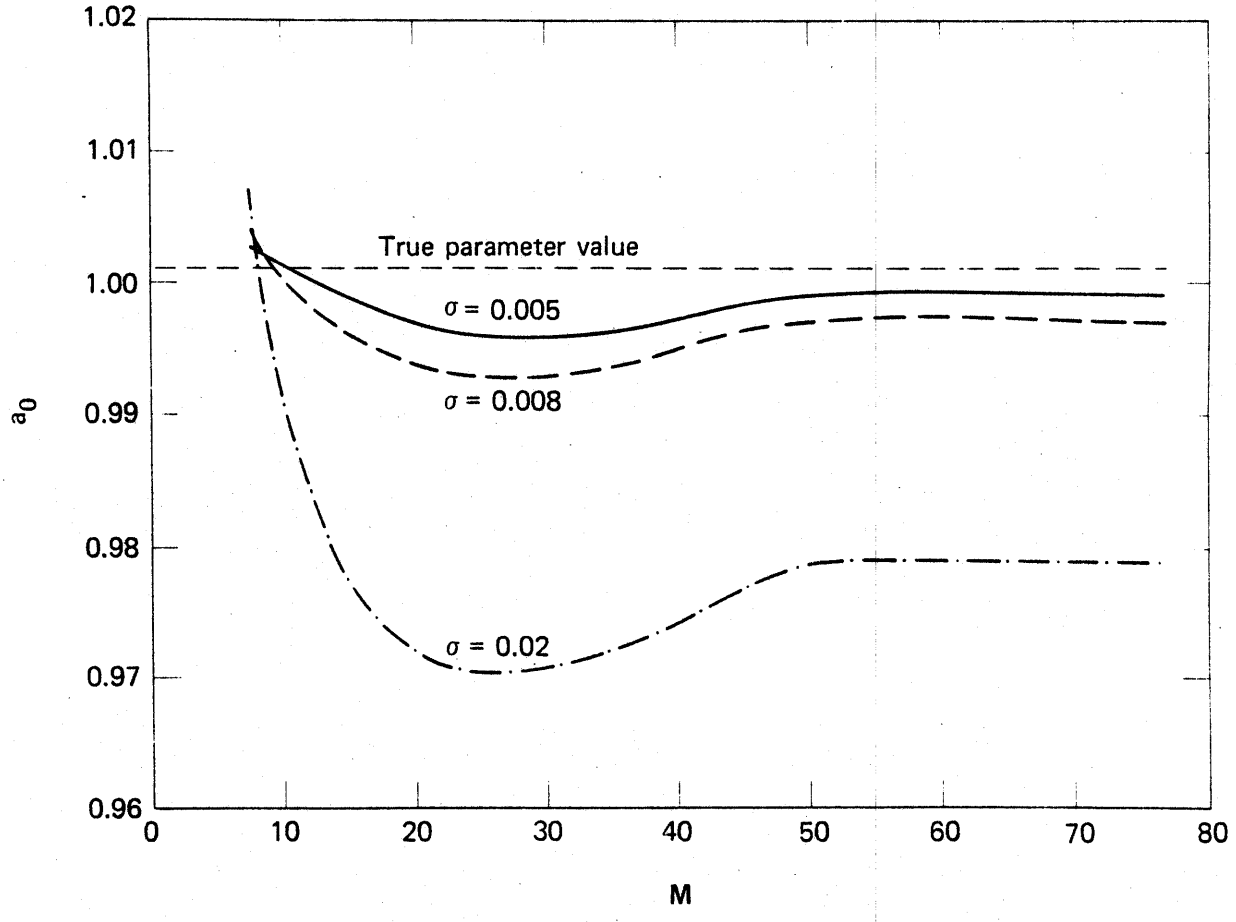

Fig. 1. Least-squares parameter estimation of polynomial coefficient a_0.

the original parameters by least-squares parameter estimation. We again let T = 1 and N = 2. For this case, however, we let M vary and solve by the Moore-Penrose pseudoinverse. An examination of the leading coefficient a_0 in the polynomial (Fig. 1) shows that indeed the least-squares process converges as M becomes large, but it converges to a biased parameter value which is a function of σ. It is unfortunate that even for small noise levels, this bias is considerably more than enough to render meaningless the ensuing calculation of the real part of s_j.

SOLUTION BY SQUARE MATRIX METHODS

Since the parameter estimation as M becomes large converges to a biased answer, we turn to a consideration of the errors if we simply let M = N. In this case, by Eq. (27), the pseudoinverse reduces to the simple inverse and we have the classic Prony method. It is common knowledge[5,6] that Prony's method is in many cases extremely sensitive to errors in the process data. In this section, we are able to definitize these errors under assumptions that the errors are small.

In Eq. (20), we assume that the errors in the data are represented by

$$g_k = f_k + \varepsilon_k, \tag{36}$$

so that Eq. (21) becomes

$$(F + E)\ (A + \alpha) = \hat{F} + \hat{E}, \tag{37}$$

where

$$F = \begin{bmatrix} f_0 & f_1 & \cdots & f_{N-1} \\ f_1 & f_2 & \cdots & f_N \\ \cdot & & & \\ \cdot & & & \\ \cdot & & & \\ f_{n-1} & f_N & \cdots & f_{2N-2} \end{bmatrix}, \quad E = \begin{bmatrix} E_0 & E_1 & \cdots & E_{N-1} \\ E_1 & E_2 & \cdots & E_N \\ \cdot & & & \\ \cdot & & & \\ \cdot & & & \\ E_{N-1} & E_N & \cdots & E_{2N-2} \end{bmatrix} \tag{38}$$

and

$$\hat{F} = \begin{bmatrix} -f_N \\ -f_{N+1} \\ . \\ . \\ . \\ f_{2N-1} \end{bmatrix}, \quad \hat{E} = \begin{bmatrix} -E_N \\ -E_{N+1} \\ . \\ . \\ . \\ -E_{2N-1} \end{bmatrix} \tag{39}$$

and

$$A = \begin{bmatrix} a_0 \\ a_1 \\ . \\ . \\ . \\ a_{N-1} \end{bmatrix}, \quad \alpha = \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ . \\ . \\ . \\ \alpha_{N-1} \end{bmatrix} \tag{40}$$

In this representation, A is the matrix of coefficients with no error, and α is the matrix containing the coefficient errors produced by the assumed errors in the data. We invert Eq. (37) by first premultiplying both sides by F^{-1} and then by $(I + F^{-1}E)^{-1}$ to give

$$A + \alpha = (I + F^{-1} E)^{-1} (A + F^{-1} \hat{E}) \quad . \tag{41}$$

We expand the inverse operator in a Taylor series, viz.,

$$(I + F^{-1} E)^{-1} = I - F^{-1} E + (F^{-1}E)^2 - \ldots \tag{42}$$

and approximate by discarding all but the first two terms to get

$$\alpha \cong (I - F^{-1} E) (A + F^{-1} \hat{E}) - A \quad . \tag{43}$$

Finally, we discard the term quadratic in the errors and obtain

$$\alpha \cong F^{-1} \hat{E} - F^{-1} E A \quad . \tag{44}$$

We may show that the error we make in our approximation is of the order $||F^{-1} E||^2$, where $||B||$ represents the norm[7] of matrix B. The proof involves simply summing the error term in the approximation of Eq. (42) followed by use of the triangle inequality and is omitted. The significance of the result in Eq. (44) is that for small errors, the error in the coefficients is a linear combination of the errors in the data.

We may carry this analysis through the process of solving for the polynomial roots by recalling from complex analysis a relationship that relates the closeness of the roots of two polynomials to the closeness of their coefficients.[7] The result is that the error in the k^{th} root z_k is given by

$$\zeta_k = \frac{\sum_{M=0}^{N-1} \alpha_m z_k^m}{P'(z_k)} , \tag{45}$$

where P(z) is the polynomial and P' its derivative. We recall also that $a_N = 1$. The significance of this result is that the error in the root becomes a linear combination of the errors in the data.

Finally, since for any complex variable w, we have, to first order in the Taylor series expansion,

$$\ln (w + \delta w) \cong \ln w + \frac{\delta w}{w} , \tag{46}$$

then

$$\tau_k = \frac{\zeta_k}{z_k} , \tag{47}$$

where τ_k is the error in the location of the k^{th} pole. We may combine Eqs. (44), (45), and (47) to obtain

$$\tau_k = \frac{\sum_{m=0}^{N-1} z_k^m \sum_{j=1}^{N+1} a_{j-1} \sum_{\ell=1}^{N} F^{-1}_{m+1,\ell} \varepsilon_{\ell+j-2}}{z_k P'(z_k)} , \tag{48}$$

where the F^{-1}_{ml} are the elements of the F^{-1} matrix. We note that Eq. (48) gives explicitly the errors in the positions of the poles related linearly to the errors in the data. This result allows a useful probabilistic interpretation of the pole position errors. It is well-known[8] that linear combinations of Gaussian variates are also Gaussian. Indeed, if it is assumed that $\{\varepsilon_k\}$ is an independent, identically distributed Gaussian sequence, then the probability density function $w_1(\varepsilon_k)$ transforms to a joint density function $w_2[\mathrm{Re}(\tau_k), \mathrm{Im}(\tau_k)]$, which is bivariate Gaussian. Therefore, for small errors, we could plot loci of constant density of the pole positions, given the statistics of the input data errors. These loci are ellipses in the complex s-plane.

Rather than pursuing this line further, let us indicate a simple computational Monto Carlo game which has far-reaching consequences. Let us assume the data to be given by Eq. (35), where the $\{f_k\}$ are obtained from known functions of the form of Eq. (1). Let $\{\delta_k\}$ be a sequence of machine-generated Gaussian variates, independent and identically distributed with zero mean and spread σ. Let the known function be characterized by $N = 2$, $A_1 = A_2 = 0.5$, $\mathrm{Im}(s_1) = -\mathrm{Im}(S_2) = 0.25$. We shall let $\mathrm{Re}(s_1) = \mathrm{Re}(s_2)$ and tabulate the results for the cases $\mathrm{Re}(s_1) = 0.00035$, 0.0035, 0.035, 0.35, and 3.5. These cases are representative of moving the poles on a locus parallel to the real axis in the s-plane. For each of the cases, we solve the Prony problem 100 times with $\sigma = 0.001$, using a new set of Gaussian data errors each time. We then combine the results using the frequency interpretation of probability. The results are given in Table 1. There are two outstanding features in Table 1.

Table 1. Monte Carlo results for two-term exponential case.[a]

Theoretical pole position	Computed pole position			
	Real part		Imaginary part	
	Mean	Spread	Mean	Spread
-0.00035 + i.25	0.0012	0.027	0.248	0.011
-0.0035 + i.25	-0.0020	.027	.248	.011
-0.035 + i.25	-0.033	.027	.248	.011
-0.35 + i.25	-0.35	.027	.248	.011
-3.5 + i.25	-3.5	0.027	0.248	0.011

[a]For all cases, $N = 2$, $A_1 = A_2 = 0.5$, $\mathrm{Im}(s_1) = \mathrm{Im}(s_2) = 0.25$, $\sigma = 0.001$.

First, as the theoretical pole position moves along a locus parallel to the real axis in the s-plane, the spread on both the real and the imaginary parts of the computed pole position does not change. Second, as a consequence of the first observation, the real part of the computed pole becomes over-whelmed by the noise as it approaches the imaginary axis. Indeed, the first entry in the table shows that the mean value of the real part of the computed pole position becomes positive, a most unfortunate occurrence.

The above result leads us to speculate on the theoretical basis for the invariance of the spread of the pole position with the pole movement. From the shifting theorem for the Laplace transform, a shift along a locus parallel to the real axis in the s-plane corresponds to multiplication by $\exp(-\alpha t)$ in the time domain. That is, if $r(t)$ has Laplace transform $R(s)$, then

$$R(s + \alpha) \Longleftrightarrow e^{-\alpha t}\, r(t) \ . \tag{49}$$

In the case of discrete data, the s-plane shift corresponds to f_k being mathematically transformed to $x^k f_k$, where $x = \exp(-\alpha)$. Concerning this transformation, we have been able to prove the following proposition:

> If the input data are represented by $f_k(1 + \delta_k)$, $k = 0, 1, \ldots 2N - 1$, and if $\delta_k << 1$ for all k, then the errors in the pole locations are invariant under the transformation
>
> $$f_k \longrightarrow x^k f_k \tag{50}$$

The proof of this proposition is straightforward but tedious and is reserved for the Appendix.

There are important consequences to the above proposition. For noisy data, the result clearly defines the sensitivity of the real part of the pole position. The concept of signal-to-noise (S/N) ratio is applicable here. We note that the noise level is established by the input noise in the data. If we analytically move the system poles toward the imaginary axis with the transformation in Eq. (49), the effect is to decrease the S/N ratio. As the movement continues, the poles become overwhelmed by the noise. At this point the Prony process fails and erratic results are to be expected. Van Blaricum and Mittra[9] mention that the real part of some of their poles seem to oscillate around their correct value. They speculate that this is probably due to noise in the data which they have made no attempt to smooth. They also comment that

some of their poles show up in the right half-plane. We feel that our proposition supports their observation. When the signal-to-noise ratio becomes low, there will be a marked sensitivity of the real part of the pole position. Further, as the S/N ratio worsens, Table 1 proves that the pole can actually appear in the right half-plane.

THE POSSIBILITIES FOR IMPROVEMENT

We have shown some definitive problems in attempts to obtain good complex exponential parameter estimates in the presence of noise. For the N × N case (classical Prony), the noise in the data produces a marked sensitivity in the real part of the pole positions, particularly for poles near the imaginary axis. For the least-squares M × N case, the correlated characteristic of the residuals leads to biased estimates of the parameters. In this section we discuss various possibilities for improving the results, still insisting on a scheme where the system identification will explicity exhibit the s-plane poles.

Golub[10] has observed that since the linear least-squares procedure, culminating in the normal equations in Eq. (19), requires $a_N = 1$ as a constraint, the procedure favors the N^{th} coefficient. Indeed, the expression

$$e_k = \sum_{m=0}^{N-1} a_m g_{m+k} + g_{n+k} \tag{51}$$

shows that in effect the residual term minus f(k + N) becomes the driving function for the difference equation with yet undetermined coefficients a_0, $a_1,\ldots,a_{N-1}$. Golub proposes that perhaps the minimization in Eq. (17) should be carried out with the constraint

$$\sum_{m=0}^{N} a_m^2 = 1 \quad . \tag{52}$$

This would have the effect of spreading the constraint among all the coefficients. The proposed method suffers from the fact that the minimization with Eq. (52) as a constraint results in a system of nonlinear equations which must be solved to produce the coefficients $\{a_k\}$. With the availability of high-speed computers, however, this disadvantage is not insurmountable.

Another possibility is based upon an examination of the error criterion adopted in Eq. (15). This definition of the error is known as "equation error"[2] since it consists of the difference between the equation governing the process and the equation governing the model. We have commented that the principal problem with this definition is the correlated nature of the residuals $\{e_k\}$ If we knew a priori that the data errors are members of an uncorrelated sequence, we could redefine the error criterion as follows:

$$e_k = g_k - f_k. \tag{53}$$

We note that in this case the $\{e_k\}$ are uncorrelated and therefore a least-squares procedure should give unbiased parameter estimates. If we minimize Eq. (53) subject to the constraint that the difference equation, Eq. (9), is satisfied, we again arrive at a system of nonlinear equations. If, however, we are able to deal computationally with the nonlinearities, we can expect elimination of the bias on the parameters. This is not a new idea; indeed, it was proposed by Householder,[11] but was abandoned because of the state of computational technology at the time. We should comment that this procedure is well-known in the general area of system identification and is classified as a generalized least squares procedure.[2]

A final possibility, and one which fits well into the framework of present systems technology is called the iterative, generalized least-squares method and is credited to Clarke,[12] who has tested it against simulated data and also against practical measurements from a system he wished to identify. We shall describe the method as we would specifically apply it to the process defined by Eq. (51). Define a shift operator q by

$$q\, f_k = f_{k+1}\ . \tag{54}$$

Then we may write Eq. (51) as

$$A(q)\ g_k = e_k, \tag{55}$$

where A(q) is a polynomial operator given by

$$A(q) = \sum_{m=0}^{N} a_m\, q^m\ . \tag{56}$$

The method is based on the following proposition. If we can find a polynomial filter B(q) such that

$$B(q)\ e_k = n_k, \tag{57}$$

where n_k is an uncorrelated noise sequence, then we can premultiply by this operator in Eq. (55) to give

$$A(q)\ B(q)\ g_k = n_k, \tag{58}$$

where we have used implicitly the fact that A commutes with B. We now define

$$\tilde{g}_k = B(q)\ g_k \tag{59}$$

and substitute into Eq. (58) to produce

$$A(q)\ \tilde{g}_k = n_k\ , \tag{60}$$

which we recognize as a tranformation of Eq. (55). The significance is that Eq. (60) has uncorrelated residuals and therefore a least-squares parameter estimation produces unbiased parameters. Since we have only estimates available, the procedure is of necessity iterative as follows:

(1) Make a least-squares estimate of A(q) in Eq. (55), which we shall call $\hat{A}(q)$

(2) Substitute $\hat{A}$ into Eq. (55) to produce an estimate of the residuals $\hat{e}_k$

(3) Use $\hat{e}_k$ in Eq. (57) and make a least-squares estimate $\hat{B}(q)$

(4) Calculate $\hat{\tilde{g}}_k$ in Eq. (59)

(5) Use $\hat{\tilde{g}}_k$ in Eq. (55) to produce a new least-squares estimate $\hat{\hat{A}}(q)$

The process continues until the loss function in Eq. (14) no longer decreases with an additional iteration.

Although there are no known proofs that the iterative, generalized least-squares process always converges, there is ample evidence that it works in many cases.[13,14] There are also extensions[14] which allow the addition of updated information in the form of additional data points without repeating the process of matrix inversion. The success of the method can be attributed to the fact that an identification procedure is performed not only on the data but also on the noise.

CONCLUSIONS AND RECOMMENDATIONS

We have examined the problem of fitting noisy data with a complex exponential series both as a classical Prony development and as a least-squares parameter estimation. We are led to the conclusion that further attempts to obtain answers in either manner in the presence of noise is an exercise in frustration. The principal problem is that the precise character of the noise in practical cases is seldom known and neither the classical nor the least-squares procedures make any attempt to do noise evaluation or filtering. In fact from the information point of view, the process is carried out in total ignorance except for the explicit data values.

There are, however, definite methods for improving the situation. The problem of fitting data with exponentials, when viewed as a problem in systems identification, becomes an autoregression in the presence of noise, viz.,

$$\sum_{m=o}^{N} a_m \, g_{m+k} = e_k \, . \tag{61}$$

Here the problem is to estimate the coefficients a_k. If this can be done with zero bias, then all will be well. The answer, however, is not in doing a simple least squares, since this leads to biased estimates. Significant improvement, however, could be obtained by considering the character of the noise, this is precisely what the iterated, generalized least-squares procedures (discussed in Sec. VI) does. The work of Hastings-James and Sage[14] is particularly definitive in this regard. It is our opinion that little can be done to improve exponential fitting without this sort of noise filtering. We strongly recommend that further efforts in this field be concentrated in analysis and numerical processing of actual and simulated data with primary emphasis on noise identification and filtering techniques.

REFERENCES

1. L. Weiss, and R. N. McDonough, "Prony's Method, Z-Transforms, and the Padé Approximation," *SIAM Review* 5, 145 (1963).
2. K. J. Astrom and P. Eykhoff, "System Identification — A Survey," in *Automatica* 7, 123 (1971).
3. A. Albert, *Regression and the Moore — Penrose Pseudoinverse* (Academic Press, New York, 1972).
4. H. B. Mann and A. Wald, "On the Statistical Treatment of Linear Stochastic Difference Equations," in *Econometrica* 11 (3 and 4) (1943).
5. F. B. Hildebrand, *Introduction to Numerical Analysis* (McGraw-Hill, New York, 1956).
6. R. W. Hamming, *Numerical Methods for Scientists and Engineers* (McGraw-Hill, New York, 1962).
7. A. Ralston, *A First Course in Numerical Analysis* (McGraw-Hill, New York, 1965).
8. A. Papoulis, *Probability, Random Variables, and Stochastic Processes,* (McGraw-Hill, New York, 1965).
9. M. L. Van Blaricum and R. Mittra, "A Technique for Extracting the Poles and Residues of a System Directly from its Transient Response," in *IEEE Trans. Ant. Prop.* AP-23, 777 (1975) and also Interaction Note 245 (1975).
10. G. Golub Stanford University, private correspondence (1976).
11. A. S. Householder, *On Prony's Method of Fitting Exponential Decay Curves and Multiple-Hit Survival Curves,* Oak Ridge National Laboratory, Oak Ridge, Tennessee, Rept. ORNL-455 (1950).
12. D. W. Clarke, "Generalized Least Squares Estimation of Parameters of a Dynamical Model," in *Proc. 1st IFAC Symposium — Identification in Automatic Control Systems, Prague,* 1967 (Academic, Prague, 1967). Available from Editor, *Automatica.*
13. J. L. Talmon, *Approximated Gauss-Markoff Estimators and Related Schemes,* Eindhoven University of Technology, Netherlands, Rept. 71-E-17 (1971).
14. R. Hastings-James and M. W. Sage, "Recursive Generalized-Least-Squares Procedure for On-Line Identification of Process Parameters," in *Proc. IEE* 116 2057 (1969).

APPENDIX
ERROR TRANSFORMATION

We shall prove the following proposition regarding errors in the input data:

If the input data are represented by $f_k(1 + \delta_k)$, $k = 0, 1, \ldots, 2N-1$, and if $\delta_k << 1$ for all k, the errors in the pole locations are invariant under the transformation

$$f_k \longrightarrow x^k f_k \ . \tag{A-1}$$

To prove the proposition, we first note that in Eq. (36)

$$\varepsilon_k = f_k \delta_k \ . \tag{A-2}$$

We define the transformation matrix T as follows:

$$T = \begin{bmatrix} x^0 & & & O \\ & x^1 & & \\ & & x^2 & \\ O & & \ddots & \\ & & & x^{N-1} \end{bmatrix} \tag{A-3}$$

In the expression for F, E, $\hat{F}$, and $\hat{E}$ in Eqs. (38) and (39), we perform the transformation indicated in Eq. (A-1) and obtain

$$F_g = TFT \tag{A-4}$$

$$E_g = TET \tag{A-5}$$

$$\hat{F}_g = x^N T \hat{F} \tag{A-6}$$

$$\hat{E}_g = x^N T \hat{E}, \tag{A-7}$$

where the subscript g indicates the transformed version of the quantity. Recalling that

$$A_g = F_g^{-1} \hat{F}_g \tag{A-8}$$

$$\alpha_g = F_g^{-1} (\hat{E}_g - E_g A_g), \tag{A-9}$$

we find

$$F_g^{-1} = T^{-1} F^{-1} T^{-1} \tag{A-10}$$

$$A_g = x^N T^{-1} A \quad . \tag{A-11}$$

$$\alpha_g = x^N T^{-1} \alpha \quad . \tag{A-12}$$

We may write the characteristic polynomial for the process in matrix form as follows

$$P_g(z) = Z_g A_g + z_g^N \tag{A-13}$$

and take its derivative to give

$$P'_g(z) = Z'_g A_g + N z_g^{N-1} , \tag{A-14}$$

where

$$Z_g = \left[z_g^0 \; z_g^1 \; z_g^2 \; \ldots \; z_g^{N-1} \right] \tag{A-15}$$

$$Z'_g = \left[0 \; z_g^0 \; 2z_g^1 \; \ldots \; (N-1) \; z_g^{N-2} \right] \tag{A-16}$$

We next substitute Eq. (A-8) to give

$$P_g(z) = x^N \left[Z_g T^{-1} A + \left(\frac{z_g}{x} \right)^N \right] \tag{A-17}$$

$$P'_g(z) = x^{N-1} \left[Z'_g \, x \, T^{-1} A + N \left(\frac{z_g}{x} \right)^{N-1} \right] , \tag{A-18}$$

which allows us to identify

$$Z = Z_g T^{-1} \tag{A-19}$$

$$Z' = Z'_g \times T^{-1} \tag{A-20}$$

and thus determine

$$Z_g = Z T \tag{A-21}$$

$$Z'_g = x^{-1} Z' T \quad . \tag{A-22}$$

Turning next to the errors in the polynomial roots, we have from Eq. (45)

$$\zeta_{gk} = \frac{Z_g \alpha_g}{Z'_g A_g + N z_g^{N-1}} \quad . \tag{A-23}$$

We substitute and obtain

$$\zeta_{gk} = \frac{xZ\alpha}{Z'A + Nz^{N-1}} \tag{A-24}$$

from which we identify

$$\zeta_{gk} = x\zeta_k \quad . \tag{A-25}$$

Since we have from Eq. (A-21)

$$z_{gk} = x\, z_k \quad , \tag{A-26}$$

we may write

$$\frac{\zeta_{gk}}{z_{gk}} = \frac{\zeta_k}{z_k} \tag{A-27}$$

and finally

$$\tau_{gk} = \tau_k \quad , \tag{A-28}$$

which proves the proposition.