A Technique for the Identification of Linear Systems

K. STEIGLITZ AND L. E. MCBRIDE

Abstract—**An iterative technique is proposed to identify a linear system from samples of its input and output in the presence of noise by minimizing the mean-square error between system and model outputs. The model chosen has a transfer function which is a ratio of polynomials in z^{-1}. Although the regression equations for the optimal set of coefficients are highly nonlinear and intractable, it is shown that the problem can be reduced to the repeated solution of a related linear problem.**

Computer simulation of a number of typical discrete systems is used to demonstrate the considerable improvement over the Kalman estimate which can be obtained in a few iterations. The procedure is found to be effective at signal-to-noise ratios less than unity, and with as few as 200 samples of the input and output records.

I. Introduction

Many investigators have considered the problem of identifying plants from input-output observation records. Indeed, the control engineer is in touch with physical reality either through the analysis of physical models or through the examination of observed system signals, and the latter is often the only information available. This is especially true in changing environments, where an adaptive approach is called for. The characterization of physical systems by a parametrized model remains today a central problem in control theory.

Of special utility when a high-speed digital computer is available is a linear sampled-data model [1], [2], which assumes that the input and output samples are related by a rational z-transform

$$N(z)/D(z)$$

where

$$N(z) = \alpha_0 + \alpha_1 z^{-1} + \cdots + \alpha_{n-1} z^{-(n-1)}$$

and

$$D(z) = 1 + \beta_1 z^{-1} + \cdots + \beta_n z^{-n}.$$

The use of a rational model permits greater freedom with fewer parameters than the polynomial obtained by linear regression [3] and requires no a priori assumption about the settling time. Since the number of parameters must be small compared to the number of data points when noise is present, the rational function permits more accurate identification when record length is limited, e.g., in slowly time-varying systems. Also, every linear, lumped-parameter physical plant has a z-transform of this form. A recently proposed identification method using the model [4] has optimal characteristics but is wasteful of data and requires an elaborate nonlinear computational technique.

Kalman [1] suggested finding the $2n$ coefficients by a linear regression on the input-output record. In Fig. 1, if x and w are the available finite records of input and output samples, respectively, the following minimization is involved:

$$\sum e_j^2 = \frac{1}{2\pi j} \oint | XN - WD |^2 \frac{dz}{z} = \min \tag{1}$$

where

$$X = X(z) = \sum x_j z^{-j}$$
$$W = W(z) = \sum w_j z^{-j}$$

Manuscript received January 25, 1965; revised July 23, 1965. This work made use of computer facilities supported in part by National Science Foundation Grant N.S.F. GP579.

The authors are with the Dept. of Electrical Engineering, Princeton University, Princeton, N. J.

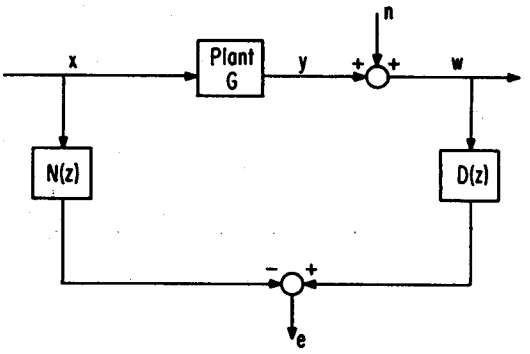

Fig. 1. The error in the linear regression solution.

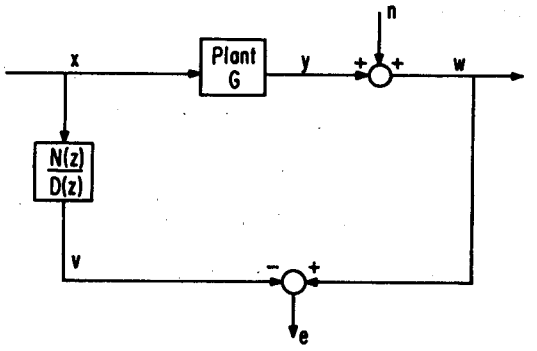

Fig. 2. True model-plant output error.

where the summations are carried out over the record length, and where the contour of integration is the unit circle. The solution to this problem is relatively simple and convenient to implement; it involves the inversion of a $2n \times 2n$ correlation matrix. In Appendix I it is shown that the solution is

$$\delta = \Phi^{-1} c \tag{2}$$

where

$$\delta = \begin{bmatrix} \alpha \\ -\beta \end{bmatrix}$$

is the coefficient vector, and Φ and c are appropriate correlation matrix and vector computed from the records of x and w.

While the minimization problem shown diagrammatically in Fig. 1 is easy to solve with a digital computer, it is not the one usually of interest. In fact, the error residual of (1) does not have any real physical interpretation. A more meaningful problem would be to minimize the error shown in Fig. 2

$$\sum e_j^2 = \frac{1}{2\pi j} \oint \left| X \frac{N}{D} - W \right|^2 \frac{dz}{z} = \min. \tag{3}$$

This is the mean-square error between the predicted output and the observed output of the plant. Unfortunately, (3) is a highly nonlinear regression problem, and to the authors' knowledge cannot be solved exactly. It is the purpose of this paper to present a technique for carrying out the minimization (3) by iteratively carrying out minimizations of the type (1). Thus the method is inherently no more difficult than the Kalman linear regression, but exchanges increased computation time for a better approximation to the solution of (3).

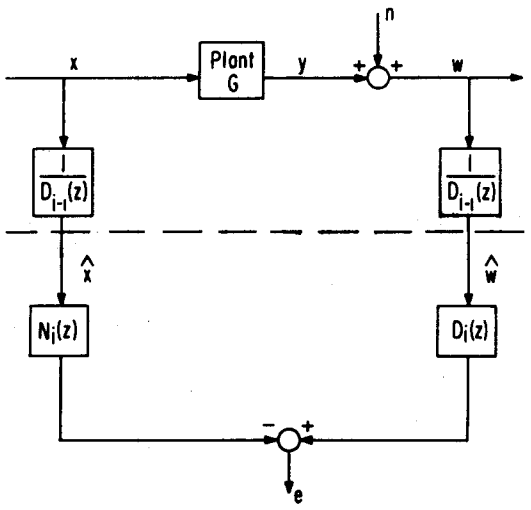

Fig. 3. The iterative prefiltering scheme.

II. The Iterative Procedure

The idea of the iterative procedure is shown diagrammatically in Fig. 3. First the minimization problem (1) is solved, using (2) and the original input-output records. The result is a "first estimate" of $N(z)$ and $D(z)$, the Kalman estimate. Call these $N_1(z)$ and $D_1(z)$. The original input and output records, x and w, are then filtered by the digital filter $1/D_1(z)$, yielding new, prefiltered input and output records $\hat{x}$ and $\hat{w}$. These prefiltered records are then used in place of the original input-output records in (2) and new estimates $N_2(z)$ and $D_2(z)$ are obtained. The digital filter $1/D_2(z)$ is then used to find $\hat{x}$ and $\hat{w}$ from x and w, and so forth. At each stage, the previous denominator is used to prefilter the input and output records, so that $N_i(z)$ and $D_i(z)$ are found such that

$$\oint \left| X \frac{N_i}{D_{i-1}} - W \frac{D_i}{D_{i-1}} \right|^2 \frac{dz}{z} = \oint \left| X \frac{N_i}{D_i} - W \right|^2 \left| \frac{D_i}{D_{i-1}} \right|^2 \frac{dz}{z} = \min,$$

$$i = 1, 2, 3, \cdots, \quad \text{and} \quad D_0 = 1.$$

If convergence is obtained, that is, if the coefficients of D_i converge as i becomes large, the error of Fig. 3 is equal to the error of Fig. 2. While no proof of convergence can be offered, every experiment with this technique conducted thus far has resulted in a decrease in mean-square error. This procedure is referred to as mode 1 iteration.

The estimate could be further improved, however, if the partial derivatives of the true error criterion were equal to zero at convergence. This condition does not hold at the end of mode 1 iteration because the two errors, though equal, are different functions of the model parameters. In Appendix II it is shown that a solution to this second problem is still given by (2), but the method of computing Q and c is altered. This procedure is called mode 2 iteration. The basic difference between the two methods is the fact that in mode 1 the measured plant output w is used as an approximation to the model output v.

Mode 2 must be started from an initial set of parameters, and is unstable (does not converge) if the initial values are far from the optimum. One or more iterations in mode 1 can be used to provide a starting point for mode 2; in many problems a more satisfactory procedure is to "slide" smoothly from 1 to 2 by using a mixture of w and v during intermediate iterations. The difference in mean-square error is usually small, but convergence in mode 2 invariably means a superior estimate, and one which has in several cases been verified by gradient methods to be the true minimum. In Section III the results of some experiments are described.

III. Experimental Results

To investigate the behavior of the technique, tests were made using the IBM 7094 digital computer. Random numbers were generated and used for the input record x; these were filtered through a known plant to produce the output signal y. To y were added independent random numbers with zero mean to produce the corrupted output signal w. The iterative identification procedures described in Section II were then carried out; with each iteration the records x and w were prefiltered and the appropriate linear regression equations were solved to yield a new estimate for N and D and a new prefilter for the next iteration. At each stage two mean-square errors were calculated and recorded; first the mean-square error between the corrupted output of the plant and the output v from the model N/D (refer to Fig. 2)

$$E_1 = \text{ave } [y + n - v]^2.$$

The second mean-square error is the error between the uncorrupted plant output and the output of the model N/D

$$E_2 = \text{ave } [y - v]^2.$$

E_1 is the error of (3) and is the one that is to be minimized. Its minimum value should be near the variance of the noise. E_2 is indicative of the accuracy of the identification, and determines whether the identification is useful. Throughout these experiments the variance of the uncorrupted output y was fixed at unity, and record lengths of 203 points were used. It was found that for a fourth-order identification, computing time was about two seconds per iteration.

The integrands which determine the errors in (1) and (3) differ by a factor of $1/|D|^2$, so that if $|D|$ is nearly constant on the unit circle the original Kalman estimate, which minimizes (1), will come close to minimizing (3) as well. This will happen if the poles of the original plant are well within the unit circle. In this case the Kalman estimate will have associated with it an E_1 which is close to the noise variance and the iterative technique cannot be expected to yield a great improvement.

Figure 4 illustrates the behavior of the mode 1 iteration when the original Kalman estimate was relatively poor (this plant had a double pole at $z=0.9$). Here both E_1 and E_2 decrease steadily, and are still decreasing at the 25th iteration.

Figures 5, 6, and 7 illustrate the behavior of the iterative technique when the original Kalman estimate is relatively good. In these cases mode 1 iteration still produces significant improvement especially in E_2. However, both E_1 and E_2 tend to reach relative minima.

As would be expected, the experimental results show that in all cases the identification is better at low noise levels.

Figure 7 shows the difference between modes 1 and 2 in a case where the actual plant is of higher order than the model. Mode 1 identification error is about 15 percent larger than the mode 2 minimum, whereas the Kalman error is 450 percent larger. These figures are typical of the relative improvement attainable by the iterative methods. Convergence of mode 2, once a rough identification is obtained, is often, as in this case, much more rapid than that of mode 1.

IV. Conclusions

The iterative identification methods described in this paper have given significant improvement over the Kalman estimate in every case considered. Successful convergence within 10–20 iterations has been obtained in more than fifty widely varying problems without failure. Each iteration is computationally identical to the Kalman linear regression except for prefiltering of the input and output records. Hence the method is well suited to high-speed digital computers and appears to have practical applications.

Further investigation in this field will include trials with input signals and noises of various spectra.

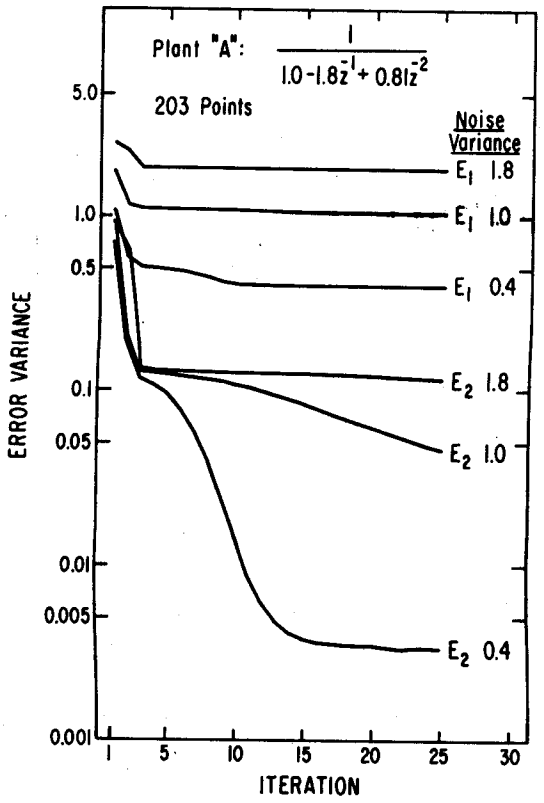

Fig. 4. Behavior of mode 1 when the original Kalman estimate is relatively poor. Second-order plant and model.

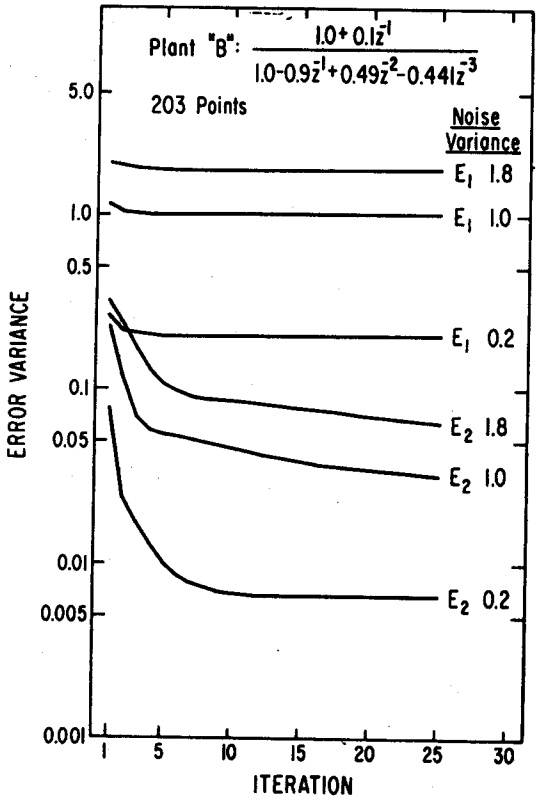

Fig. 5. Behavior of mode 1 when the original Kalman estimate is relatively good. Third-order plant and model.

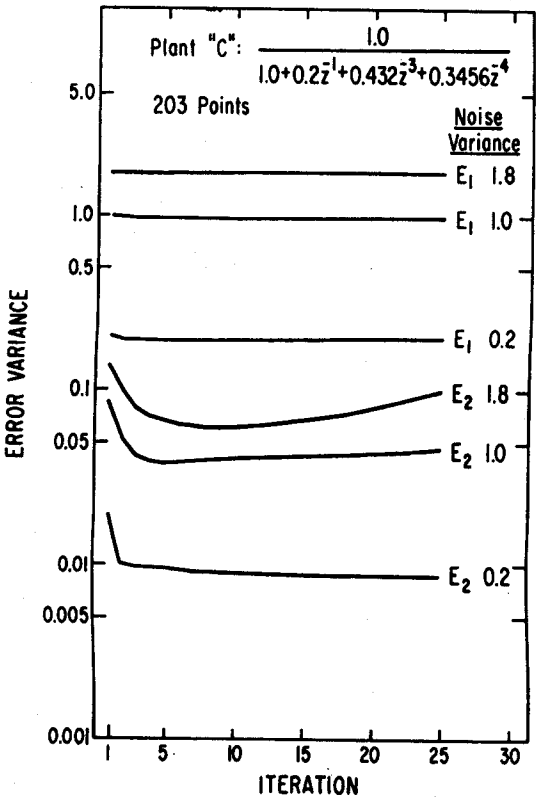

Fig. 6. Behavior of mode 1 for fourth-order plant and model.

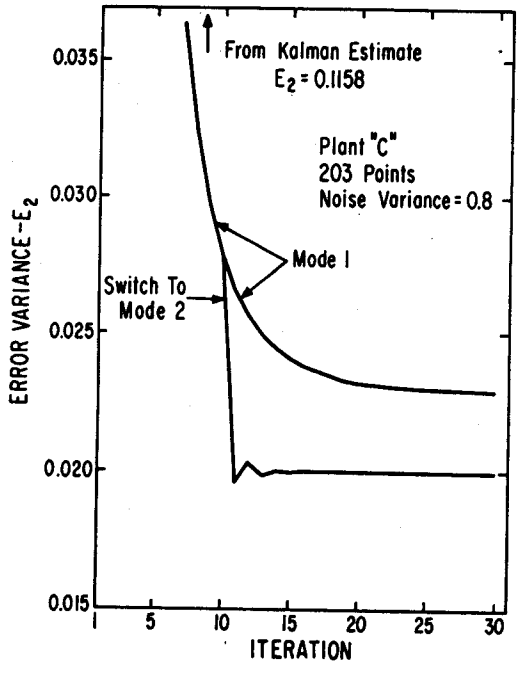

Fig. 7. The convergence of E_2 in modes 1 and 2; greatly expanded scale. At final convergence of mode 2, E_1 is a true minimum at 0.7949. The noise sample is different from that of Fig. 6, and a third-order model was used.

Appendix I

Solution of Linear Regression Equations

The error of Fig. 1 at time j is given by

$$e_j = \sum_{i=0}^{n-1} \alpha_i x_{j-i} - \sum_{i=1}^{n} \beta_i w_{j-i} - w_j.$$

If the coefficient vector δ and the input-output vector q_j are defined by

$$\delta' = [\alpha_0, \cdots, \alpha_{n-1}, -\beta_1, \cdots, -\beta_n],$$
$$q_j' = [x_j, \cdots, x_{j-n+1}, w_{j-1}, \cdots, w_{j-n}];$$

this becomes

$$e_j = q_j'\delta - w_j \tag{4}$$

where the prime denotes the transpose. Summing on j over the record length and taking the gradient with respect to δ yields

$$\text{grad}\,(\sum e_j^2) = \partial/\partial\delta(\sum e_j^2) = 2\sum (\partial e_j/\partial\delta)e_j = 2\sum q_j e_j = 0. \tag{5}$$

Substitution of (4) into (5) gives

$$(\sum q_j q_j')\delta = \sum w_j q_j.$$

If the $2n \times 2n$ correlation matrix is defined as

$$Q = \sum q_j q_j'$$

and the $2n$ correlation vector as

$$c = \sum w_j q_j$$

the solution to the original minimization problem can be written as

$$\delta = Q^{-1}c.$$

The same derivation applies in minimizing the error of Fig. 3, except that q_j is defined in terms of $\hat{x}$ and $\hat{w}$.

Appendix II

Mode 2 Iteration

Since the true error $E(z)$ is given by

$$E(z) = X(z)\frac{N(z)}{D(z)} - W(z), \tag{6}$$

its partial derivatives are given by

$$\frac{\partial E(z)}{\partial \alpha_i} = \frac{X(z)}{D(z)} z^{-i} = \hat{X}(z)z^{-i}$$
$$\frac{\partial E(z)}{\partial \beta_i} = \frac{-X(z)N(z)}{D^2(z)} z^{-i} = -\frac{V(z)}{D(z)} z^{-i} = -\hat{V}(z)z^{-i}. \tag{7}$$

Since the operations of differentiation and the inverse transform can be interchanged, if a new vector p_j is defined by

$$p_j' = [\hat{x}_j, \cdots, \hat{x}_{j-n+1}, -\hat{v}_{j-1}, \cdots, -\hat{v}_{j-n}]$$

the true error gradient becomes

$$\text{grad}\,(\sum e_j^2) = 2\sum p_j e_j = 2\sum (p_j q_j'\delta - w_j p_j) \tag{8}$$

where the second equality is true only at convergence. The procedure is then identical to that of mode 1, except that the definitions of Q and c become

$$Q = \sum p_j q_j', \qquad c = \sum w_j p_j.$$

References

[1] R. E. Kalman, "Design of a self-optimizing control system," *Trans. ASME*, vol. 80, pp. 468–478, February 1958.
[2] G. G. Lendaris, "The identification of linear systems," *Trans. AIEE (Applications and Industry)*, pt. 2, vol. 81, pp. 231–242, September 1962.
[3] M. J. Levin, "Optimum estimation of impulse response in the presence of noise," *IRE Trans. on Circuit Theory*, vol. CT-7, pp. 50–56, March 1960.
[4] M. J. Levin, "Estimation of a system pulse transfer function in the presence of noise," *IEEE Trans. on Automatic Control*, vol. AC-9, pp. 229–235, July 1964.

Reprinted from IEEE TRANSACTIONS
ON AUTOMATIC CONTROL
Volume AC-10, Number 4, October, 1965
Pp. 461-464