UNCLASSIFIED

AD 411832

DEFENSE DOCUMENTATION CENTER

FOR

SCIENTIFIC AND TECHNICAL INFORMATION

CAMERON STATION, ALEXANDRIA, VIRGINIA

UNCLASSIFIED

NOTICE: When government or other drawings, specifications or other data are used for any purpose other than in connection with a definitely related government procurement operation, the U. S. Government thereby incurs no responsibility, nor any obligation whatsoever; and the fact that the Government may have formulated, furnished, or in any way supplied the said drawings, specifications, or other data is not to be regarded by implication or otherwise as in any manner licensing the holder or any other person or corporation, or conveying any rights or permission to manufacture, use or sell any patented invention that may in any way be related thereto.

AD No. 411832
DDC FILE COPY

Reprint of Technical paper from the 1963 Joint Automatic Control Conference.

ESTIMATION OF A SYSTEM PULSE TRANSFER FUNCTION IN THE PRESENCE OF NOISE

Morris J. Levin
Lincoln Laboratory,* Massachusetts Institute of Technology
Lexington 73, Massachusetts

1. Introduction

Accurate measurements of system characteristics from limited amounts of data are important in many applications. Some adaptive control systems carry out measurements of plant parameters during normal operation. In communication systems an analogous situation arises in the utilization of time-varying channels (1). Often the application of a special test signal is undesirable so the information must be obtained from ordinary input-output data. The presence of random noise and instrumentation errors can render many measurement schemes ineffective. Statistical estimation theory provides powerful methods for dealing with this type of problem. Some of these methods are applied ~~below~~ to the estimation of the pulse transfer function of a linear system.

For control system applications Kalman (2) has shown that the characterization of a linear system in terms of the coefficients of its pulse transfer function offers many advantages. Although the complete impulse response or frequency response also conveys the same information a good estimate of one of these functions is not readily translated into a good estimate of another. In addition, the assumptions required to express these different functions in terms of a finite number of parameters suitable for estimation are generally not equivalent. Therefore the estimation procedure should be formulated directly in terms of the desired parameters.

Estimates of the coefficients of both conventional (Laplace transform) and pulse transfer functions have been considered by previous authors. Ellington and McCallion (3) and Shinbrot (4) applied non-linear curve fitting techniques to this problem. Corbin (5), Lendaris (6), and Zaborszky and Berger (7) obtained estimates by solving sets of simultaneous linear equations in derivatives and integrals of the input and output. Kalman (2) described a least squares fitting method which was investigated experimentally by Bigelow and Ruge (8). A similar technique was applied by Kaya and Yamamura (9). Joseph, Lewis, and Tou (10) used a closely related method which avoids bias errors due to correlated disturbances at input and output. Kushner (11) examined in detail the properties of a computationally simple recursive scheme.

Thus many different types of estimates have been examined. However, none of this previous work has attacked the fundamental question of what basic form the estimates should take to make optimum use of the data available. The present paper considers this question by assuming additive Gaussian noise and utilizing the method of maximum likelihood to derive estimates having certain optimal properties. An evaluation of the Cramer-Rao lower bound provides an approximation to the sampling variances. Due to mathematical difficulties a complete solution is obtained only for a suitably restricted formulation of the problem which does not exploit all the available information. However, the solution is easily modified to incorporate the remaining information. The results provide considerable insight into the properties of other previously suggested methods. For a similar analysis of impulse response estimation, see (12). Details of some of the results below are contained in (13).

2. Results from Mathematical Statistics

The expectation (mean value) is denoted by E. Consider a sequence of S independent vector random variables whose probability density is known except for a parameter α. Let α_e denote any estimate of α. The bias of α_e is $(E\alpha_e - \alpha)$ and the variance is

$$\mathrm{Var}\ \alpha_e = E(\alpha_e)^2 - (E\alpha_e)^2$$

Under general regularity conditions, the minimum possible value of Var α_e is given by the Cramer-Rao lower bound (14). α_e is said to be a consistent estimate if α_e converges in probability (p lim) to α as $S \to \infty$.

Cramer states, "From a theoretical point of view, the most important general method of estimation so far known is the method of maximum likelihood." Under general conditions maximum likelihood estimates, denoted as $\hat{\alpha}$, are consistent, asymptotically Gaussian and asymptotically efficient. This implies that as S becomes large $\hat{\alpha}$ converges in a certain sense to a Gaussian distribution with mean α and variance given by the Cramer-Rao lower bound. A useful property of maximum likelihood estimates (15) is that if $\beta = f(\alpha)$ and the transformation is one-to-one then $\hat{\beta} = f(\hat{\alpha})$. Analogous properties apply in the case of multiple parameters.

The following theorem of Slutsky (14) is used later: "If $\xi_n, \eta_n, \ldots, \rho_n$ are random variables converging in probability to the constants, x, y, ..., r, respectively, any rational function $R(\xi_n, \eta_n, \ldots, \rho_n)$ converges in probability to the constant $R(x, y, \ldots, r)$ provided that the latter is finite."

---

* Operated with support from the U. S. Army, Navy and Air Force.

NO OTS

3. Assumptions

The situation analyzed is shown in Fig. 1. The following assumptions are initially made:

a) The input r(n) and output c(n) are sampled quantities with sampling interval unity.

b) r(n) and c(n) are related by a stable linear constant-coefficient difference equation,

$$c(n) + \beta_1 c(n-1) + \ldots + \beta_K c(n-K) - \alpha_0 r(n) - \alpha_1 r(n-1) - \ldots - \alpha_K r(n-K) = 0 \qquad (1)$$

where $\beta_K \neq 0$, K is known and the α_k and β_k are to be estimated. The pulse transfer function is then

$$H(z) = \frac{\alpha_0 + \alpha_1 z^{-1} + \ldots + \alpha_K z^{-K}}{1 + \beta_1 z^{-1} + \ldots + \beta_K z^{-K}}$$

with $z = e^s$

c) The quantities

$$x(n) = r(n) + u(n) \qquad y(n) = c(n) + v(n)$$

are observed for $0 \leq n \leq N$.

d) The obscuring noise sequences u(n) and v(n) are each sequences of independent Gaussian random variables with mean zero and known variances σ_u^2 and σ_v^2 respectively. The covariance

$$\rho = E\, u(n)\, v(n)$$

is not necessarily zero and is known.

4. Maximum Likelihood Estimates

The problem is now cast into a form for which maximum likelihood estimates can be obtained. Consider a $P = 2K+2$ dimensional Euclidean space with axes r(n), r(n-1), ..., r(n-K), c(n), c(n-1), ..., c(n-K). (If some of the coefficients are known to be zero the dimensionality of the space is correspondingly reduced.) For each value of n the corresponding set of values from the r(n) and c(n) sequences determines a point in this space. By virtue of (1) these points all lie in a hyperplane passing through the origin. If no noise is present any P-1 linearly independent points determine a hyperplane whose equation provides the P-1 values of the α_k and β_k. If noise is present and the x(n) and y(n) sequences are considered then the observed points are scattered about the hyperplane and a method of fitting a hyperplane to these points is required to estimate the coefficients.

Since the observed points have added random disturbances in all coordinate directions the standard least squares method, which assumes random errors along one coordinate axis only, is not entirely appropriate. The problem of fitting a hyperplane when the random errors occur along more than one coordinate axis has been examined by many authors (16, 17). The most pertinent analysis has been made by Koopmans (18), who derived the maximum likelihood solution for the case of Gaussian errors, independent from point to point. It turns out to be a generalization of the standard least squares fit.

Koopmans' results will now be applied to the estimation of the pulse transfer function. To obtain independent errors in each coordinate of each observed point the x(n) and y(n) sequences are initially split into non-overlapping sets. Each observed point consists of K+1 consecutive values of x(n) and the corresponding values of y(n). Take

$$x_0^{(1)} = x(K), \qquad x_1^{(1)} = x(K-1), \ldots \qquad x_K^{(1)} = x(0)$$

$$x_0^{(2)} = x(2K+1), \qquad x_1^{(2)} = x(2K), \ldots \qquad x_K^{(2)} = x(K+1)$$

$$\vdots$$

$$x_k^{(s)} = x(s[K+1] - 1 - k)$$

etc. with a similar notation for y(n) and for the other sequences. Here s indexes the observed points and k indexes their coordinates. Let $s = 1, 2, \ldots, S$ so that S is the total number of observed points where we must have

$$2K+1 \leq S \leq N/K+1$$

This notation is illustrated in Fig. 2 for K = 2.

The following vectors are now defined where T indicates the transpose:

$$\underline{\gamma}^T = [1 \beta_1 \ldots \beta_K - \alpha_0 \ldots - \alpha_K]$$

$$\underline{\mu}^{(s)T} = [c_0^{(s)}\ c_1^{(s)} \ldots c_K^{(s)}\ r_0^{(s)} \ldots r_K^{(s)}]$$

$$\underline{\lambda}^{(s)T} = [y_0^{(s)}\ y_1^{(s)} \ldots y_K^{(s)}\ x_0^{(s)} \ldots x_K^{(s)}]$$

$$\underline{\zeta}^{(s)T} = [v_0^{(s)}\ v_1^{(s)} \ldots v_K^{(s)}\ u_0^{(s)} \ldots u_K^{(s)}]$$

Thus $\underline{\lambda}^{(s)} = \underline{\mu}^{(s)} + \underline{\zeta}^{(s)}$ and from (1)

$$\underline{\mu}^{(s)T} \underline{\gamma} = 0 \qquad \text{for } s = 1, 2, \ldots, S \qquad (2)$$

Therefore the points $\underline{\mu}^{(s)}$ lie in a hyperplane. The coefficients of the equation of this hyperplane are the elements of $\underline{\gamma}$. Considered as a vector, $\underline{\gamma}$ passes through the origin and is perpendicular to this hyperplane.

For each observed point $\underline{\lambda}^{(s)}$ the random components have covariance matrix

$$\underline{Z} = E\, \underline{\zeta}^{(s)} \underline{\zeta}^{(s)T} = \left[\begin{array}{c|c} \sigma_v^2 \underline{I} & \rho \underline{I} \\ \hline \rho \underline{I} & \sigma_u^2 \underline{I} \end{array}\right]$$

where $\underline{I}$ is a (K+1) dimensional identity matrix. For all S points the probability density of the $\underline{\lambda}^{(s)}$ is the multivariate Gaussian distribution

$$\mathrm{Prob}\Big\{\underline{\ell}^{(1)} \le \underline{\lambda}^{(1)} < \underline{\ell}^{(1)} + d\underline{\ell}^{(1)}, \dots,$$

$$\underline{\ell}^{(S)} \le \lambda^{(S)} < \underline{\ell}^{(S)} + d\underline{\ell}^{(S)}\Big\}$$

$$= (2\pi)^{-SP/2}\, |\underline{Z}|^{-S/2}$$

$$\exp\Big\{-\frac{1}{2}\sum_{s=1}^{S}\Big[\underline{\ell}^{(s)} - \underline{\mu}^{(s)}\Big]^{T} \underline{Z}^{-1}\Big[\underline{\ell}^{(s)} - \underline{\mu}^{(s)}\Big]\Big\} \quad (3)$$

Maximum likelihood estimates are those values of the unknown parameters maximizing the likelihood function. This function is obtained by substituting the observed values of the random variables into the probability density (3). In the present case this maximization is equivalent to the minimization of

$$D = \sum_{s=1}^{S}\Big[\underline{\lambda}^{(s)} - \underline{\mu}^{(s)}\Big]^{T} \underline{Z}^{-1}\Big[\underline{\lambda}^{(s)} - \underline{\mu}^{(s)}\Big]$$

The solution is complicated by the fact that the points $\underline{\mu}^{(s)}$ are not explicit functions of $\underline{\gamma}$ but are merely restricted by (2) to lie in a hyperplane with coefficients $\underline{\gamma}$.

We now briefly sketch Koopmans' solution (18) for the maximum likelihood estimate $\hat{\underline{\gamma}}$. The minimization of D is carried out in two steps. First, for any trial hyperplane with coefficients $\underline{\gamma}_t$, points $\underline{\omega}^{(s)}(\underline{\gamma}_t)$ which lie in this hyperplane are substituted for the $\underline{\mu}^{(s)}$ and those which minimize D are determined. It is found that the resulting value of D becomes

$$\min_{\underline{\omega}^{(s)}(\underline{\gamma}_t)} \quad D(\underline{\gamma}_t) = \frac{\underline{\gamma}_t^{T}\,\underline{\Lambda}\,\underline{\gamma}_t}{\underline{\gamma}_t^{T}\,\underline{Z}\,\underline{\gamma}_t} \quad (4)$$

where

$$\underline{\Lambda} = \frac{1}{S}\sum_{s=1}^{S}\underline{\lambda}^{(s)}\,\underline{\lambda}^{(s)T}$$

$$= \frac{1}{S}\left[\begin{array}{ccc|ccc}
\sum (y_0^{(s)})^2 & \sum y_0^{(s)} y_1^{(s)} \dots & & \sum y_0^{(s)} x_0^{(s)} & \dots & \sum y_0^{(s)} x_K^{(s)} \\
\sum y_1^{(s)} y_0^{(s)} & \sum (y_1^{(s)})^2 \dots & & \sum y_1^{(s)} x_0^{(s)} & \dots & \sum y_1^{(s)} x_K^{(s)} \\
\vdots & \vdots & & \vdots & & \vdots \\
\hline
\sum x_0^{(s)} y_0^{(s)} & \sum x_0^{(s)} y_1^{(s)} \dots & & \sum (x_0^{(s)})^2 & \dots & \sum x_0^{(s)} x_K^{(s)} \\
\vdots & \vdots & & \vdots & & \vdots \\
\sum x_K^{(s)} y_0^{(s)} & \sum x_K^{(s)} y_1^{(s)} \dots & & \sum x_K^{(s)} x_0^{(s)} & \dots & \sum (x_K^{(s)})^2
\end{array}\right]$$

and all sums run over $s = 1, 2, \dots, S$. The elements of $\underline{\Lambda}$ are seen to be sums of cross-products of the x(n) and y(n) sequences. Also used later is the related matrix

$$\underline{M} = \frac{1}{S}\sum_{s=1}^{S}\underline{\mu}^{(s)}\,\underline{\mu}^{(s)T}$$

Note that

$$E\,\underline{\Lambda} = \underline{M} + \underline{Z} \quad (5)$$

Second, the expression (4) is minimized with respect to $\underline{\gamma}_t$ to provide $\hat{\underline{\gamma}}$. This is done by employing the extremal properties of generalized eigenvectors (19). It is found that $\hat{\underline{\gamma}}$ is given by the solution of P simultaneous linear equations

$$[\underline{\Lambda} - \Theta_1 \underline{Z}]\,\hat{\underline{\gamma}} = 0 \quad (6)$$

where Θ_1 is the smallest value of Θ satisfying the determinantal equation

$$|\underline{\Lambda} - \Theta\,\underline{Z}| = 0 \quad (7)$$

It can be shown that Θ_1 is non-negative. If $\underline{Z} = \underline{I}$, Θ_1 is the smallest eigenvalue of $\underline{\Lambda}$ and $\hat{\underline{\gamma}}$ is the corresponding eigenvector. Otherwise this is a generalized eigenvalue problem. Note that $\underline{Z}$ need be known only to within a constant multiplier. If $\underline{Z}$ is singular the derivation must be modified but the solution is still valid.

5. Geometric Interpretation

It is now demonstrated that $\hat{\underline{\gamma}}$ satisfies a generalized least squares fitting criterion. Define a generalized squared distance between any two points $\underline{\lambda}^{(s)}$ and $\underline{\omega}^{(s)}$ as

$$d_s^2 = \sum_{i,j=1}^{P} (\lambda_i^{(s)} - \omega_i^{(s)}) Z_{ij}^{-1} (\lambda_j^{(s)} - \omega_j^{(s)})$$

$$= [\underline{\lambda}^{(s)} - \underline{\omega}^{(s)}]^T \underline{Z}^{-1} [\underline{\lambda}^{(s)} - \underline{\omega}^{(s)}] \qquad (8)$$

where the Z_{ij}^{-1} are the elements of $\underline{Z}^{-1}$. Consider an observed point $\underline{\lambda}^{(s)}$, some trial hyperplane, and the "adjusted point" $\underline{\omega}^{(s)}$ which lies in this hyperplane in such a position that d_s^2 is a minimum. For example, if $\underline{Z}^{-1}$ is the unit matrix, d_s is the length of the perpendicular from $\underline{\lambda}^{(s)}$ to the hyperplane and $\underline{\omega}^{(s)}$ is the point lying at the foot of this perpendicular. For a set of observed points the sum D of the d_s^2 depends upon the hyperplane. The generalized least squares criterion selects the hyperplane minimizing D.

The standard least squares fit along the y_0 axis corresponds to the matrix

$$Z_{ij}^{-1} = \begin{cases} 1 & i = 1,\ j = 1 \\ \infty & i = j;\ i,j \neq 1 \\ 0 & \text{otherwise} \end{cases}$$

The sum of squared deviations is measured along the $(i = 1,\ j = 1)$ axis only. Deviations along any other axis are weighted by $Z_{ii}^{-1} = \infty$ and are therefore forced to be zero.

If the maximum likelihood estimates are to be reliable the observed points must not satisfy, even approximately, more than one relation of the type expressed by (2). In other words the observed points must not be concentrated in any linear subspace of dimension less than P - 1 or the hyperplane of best fit will not be well defined. This requires linear independence among the $r_k^{(s)}$ for each value of s and therefore it is necessary that the r(n) sequence not be the solution of any linear constant-coefficient difference equation of order K+1 or less. Therefore exponential or low-order polynomial inputs are undesirable for estimation puposes.

6. Properties of the Maximum Likelihood Estimates

(1) Consistency. Maximum likelihood estimates are, in general, consistent so that

$$\underset{S\to\infty}{\text{p lim}}\ \hat{\underline{\gamma}} = \underline{\gamma}$$

(2) Bias. For finite S, $\hat{\underline{\gamma}}$ is generally biased. However, Koopmans has shown that if

$$Z_{ii} \ll M_{ii} \qquad \text{for all } i \qquad (9)$$

so that the noise variance is small compared with the mean-square values of r(n) and c(n) then the bias is negligible compared with the standard deviation of $\hat{\underline{\gamma}}$.

(3) Variance. Under the condition (9) but without using the assumption of Gaussian noise Koopmans has obtained, by an involved matrix series representation, an approximation to the covariance matrix of the $\hat{\gamma}_i$,

$$\left[\text{Cov}\ \hat{\gamma}_i, \hat{\gamma}_j\right] \approx \frac{1}{S} (\underline{\gamma}^T \underline{Z}\, \underline{\gamma})\ \underline{M}_{11}^{-1} \qquad i,j \neq 1 \qquad (10)$$

Here M_{11} is the matrix formed by deleting the first row and first column of $\underline{M}$. The values for $i,j = 1$ do not appear in this covariance matrix since $\gamma_1 = 1$ by assumption. The matrix $\underline{M}_{11}^{-1}$ is proportional to the covariance matrix of the estimates that would be obtained if the errors occurred along just one coordinate axis so that the standard least squares estimates were appropriate. The scale factor $\frac{1}{S}\ (\underline{\gamma}^T\ \underline{Z}\,\underline{\gamma})$ depends upon the true parameter values and the noise covariance matrix and is inversely proportional to the number of observations. We have established by a rather intricate computation which will not be repeated here the basic result that for Gaussian noise (10) is the same as the covariance matrix given by the Cramer-Rao lower bound for joint unbiased estimates.

The quantity θ_1 in (7) is the sum of the squared deviations from the hyperplane of best fit. If (9) holds then it can be shown that $E\ \theta_1 \approx (S-P)/S$ and the order of magnitude of the standard deviation of θ_1 is $\sqrt{2/S}$. Thus θ_1 indicates how well the data fits the estimated coefficients. An excessively large value may suggest that the order of the system which has been assumed is not large enough. Alternatively, if the scale factor of $\underline{Z}$ is unknown it can be estimated by θ_1.

7. Estimates with Overlapping Sets of Values

The estimates $\hat{\underline{\gamma}}$ are maximum likelihood only with respect to the observed points constructed from the non-overlapping sets of values of the x(n) and y(n) defined in Section 4. Since these points do not contain all the information in the data it appears that improved results would be obtained by taking as observed points every successive set of (K+1) values of the x(n) and the corresponding y(n) which would increase the number of points S by a factor of (K+1). The noise components are then no longer independent from point to point and although the maximum likelihood equations are easy to derive it has not been found possible to solve them in a useful form. If the matrix $\underline{A}$ is calculated from overlapping sets of values and employed with (6) and (7) it can be shown that no additional bias errors are introduced and it appears that the variance is reduced by a factor of almost 1/(K+1). It is conjectured that when the noise components are large compared with r(n) and c(n) this procedure is efficient but that when they are small a better method may exist. For this procedure, which seems most useful for practical purposes, $\underline{A}$ becomes

$$\underline{\Lambda} = \frac{1}{S}\left[\begin{array}{ccc|ccc} \sum y^2(n) & \sum y(n)\,y(n-1) & \cdots & \sum y(n)\,x(n) & \cdots & \sum y(n)\,x(n-K) \\ \sum y(n-1)\,y(n) & \sum y^2(n-1) & \cdots & \sum y(n-1)\,x(n) & \cdots & \sum y(n-1)\,x(n-K) \\ \vdots & \vdots & & \vdots & & \vdots \\ \hline \vdots & \vdots & & \vdots & & \vdots \\ \sum x(n-K)\,y(n) & \sum x(n-K)\,y(n-1) & \cdots & \sum x(n-K)\,x(n) & \cdots & \sum x^2(n-K) \end{array}\right]$$

where all summations run over $n=K, K+1, \ldots, N$. The elements of this matrix are measured auto- and cross-correlation functions of the x(n) and y(n) except that different summations include slightly different sets of values of the products.

8. Properties of the Estimates for Other Types of Noise

If the noise obeys the assumptions of Section 3 except that it is non-Gaussian then the estimates (6) and (7) are no longer maximum likelihood. However, the geometrical interpretation and the fact that the variance is primarily influenced by only the covariance matrix of the noise suggests that these estimates are still reasonably good. It can be shown that under general conditions these estimates remain consistent.

If u(n) and v(n) are sequences of correlated random variables then the noise components of the $\underline{\lambda}^{(s)}$ are not independent and again the maximum likelihood estimates are not known. If they are stationary time series and the covariance matrix $\underline{Z}$ is used with (6) and (7) then consistent estimates are still obtained.

9. Discussion of Other Estimates

It is of interest to compare the properties of the simpler standard least squares estimates $\underline{\gamma}^*$ described by Kalman (2). These estimates minimize the sum of squared distances measured along a single coordinate axis and are given by the solution of a set of simultaneous linear equations. With the distance measured along the y_0 axis and $\gamma_1 = 1$, they satisfy

$$\underline{\Lambda}\,\underline{\gamma}^* = 0 \tag{11}$$

so the i'th component of $\underline{\gamma}^*$ is

$$\gamma_i^* = \frac{\mathcal{L}_{1i}}{\mathcal{L}_{11}} \tag{12}$$

where the $\mathcal{L}_{1i}$ are the cofactors of $\underline{\Lambda}$.

It can be shown that the variances for these estimates are approximately the same as for $\hat{\underline{\gamma}}$. Unfortunately they are not consistent when noise is present. To demonstrate this suppose S is large and that r(n) has reasonable characteristics so that $\underline{M}$ converges to some constant matrix. Then under general conditions

$$\underset{S\to\infty}{\mathrm{p\ lim}}\ \underline{\Lambda} = \underline{M} + \underline{Z} \tag{13}$$

By Slutsky's theorem (Section 2)

$$\mathrm{p\ lim}\ \gamma_i^* = \frac{\mathrm{p\ lim}\ \mathcal{L}_{1i}}{\mathrm{p\ lim}\ \mathcal{L}_{11}} \tag{14}$$

so that knowing $\underline{M}$ and $\underline{Z}$ these values can be calculated. The asymptotic bias introduced by the non-zero elements of $\underline{Z}$ can be evaluated by noting that from (2)

$$\underline{M}\,\underline{\gamma} = 0 \tag{15}$$

Whether this bias is significant depends upon the magnitude of the noise and the desired accuracy. An example is given in the next Section.

It is apparent that the solution (6) subtracts out from the matrix $\underline{\Lambda}$ the best estimate $\theta_1\,\underline{Z}$ of the components due to noise. A simpler estimate which is not asymptotically biased is given by the solution of

$$[\,\underline{\Lambda} - \underline{Z}\,]\,\underline{\gamma} = 0 \tag{16}$$

but this is presumably not so efficient as the maximum likelihood estimates.

If no noise is present in the x(n) sequence then it has been shown (13) that a set of simultaneous linear equations can be formed which provides consistent estimates of $\underline{\gamma}$ without further knowledge of $\underline{Z}$. With noise present in both the x(n) and y(n) sequences the method of Joseph, Lewis and Tou (10) provides estimates without requiring a knowledge of $\underline{Z}$. They form a set of simultaneous linear equations in terms of the cross-correlation functions of x(n) and y(n) with a signal elsewhere in the system related to x(n) and having uncorrelated noise components. If such a signal is available the method appears quite useful although any optimum properties remain to be established. Since an unfavorable input signal could cause the equations to become singular or poorly conditioned and therefore produce estimates with large variances the necessary restrictions on the input should be investigated.

10. Example

The calculations of the properties of $\hat{\underline{\gamma}}$ and $\underline{\gamma}^*$ are now demonstrated by a simple example. Consider the pulse transfer function

$$H(z) = \frac{\alpha_0}{1+\beta_1 z^{-1}} \tag{17}$$

where $|\beta_1| < 1$, α_0 and β_1 are to be estimated, and u(n) and v(n) obey assumption d) of Section 3 with $\rho = 0$. Denote auto- and cross-correlation functions of the actual r(n) and c(n) sequences by

$$\phi_{rr}(m) = \frac{1}{N+1} \sum_{n=0}^{N} r(n)\, r(n+m) \tag{18}$$

$$\phi_{cc}(m) = \frac{1}{N+1} \sum_{n=0}^{N} c(n)\, c(n+m) \tag{19}$$

$$\phi_{rc}(m) = \frac{1}{N+1} \sum_{n=0}^{N} r(n)\, c(n+m) \tag{20}$$

$[\mathrm{Cov}\, \hat{\gamma}_i, \hat{\gamma}_j]$ can be obtained by using the approximation, valid for large S,

$$\underline{M} \approx \begin{bmatrix} \phi_{cc}(0) & \phi_{cc}(1) & \phi_{rc}(0) \\ \phi_{cc}(1) & \phi_{cc}(0) & \phi_{rc}(-1) \\ \phi_{rc}(0) & \phi_{rc}(-1) & \phi_{rr}(0) \end{bmatrix} \tag{21}$$

The elements of $\underline{M}$ can be calculated from

$$\phi_{rc}(m) = \sum_{p=0}^{\infty} h(p)\, \phi_{rr}(m-p) \tag{22}$$

and

$$\phi_{cc}(m) = \sum_{p=0}^{\infty} \sum_{q=0}^{\infty} h(p)\, h(q)\, \phi_{rr}(m-p+q) \tag{23}$$

where h(p) is the impulse response given by the inverse z-transform of H(z).

The simplest case is when r(n) is a white-noise-like sequence such that for large N

$$\begin{aligned} &\phi_{rr}(0) \neq 0 \\ &\phi_{rr}(m) \approx 0 \qquad m \neq 0 \end{aligned} \tag{24}$$

Then it is found that

$$\phi_{cc}(0) = \phi_{rr}(0)\, \alpha_0^2/(1 - \beta_1^2) \tag{25}$$

$$\phi_{cc}(1) = -\beta_1\, \phi_{cc}(0) \tag{26}$$

There follows

$$\underline{M} \approx \begin{bmatrix} \phi_{cc}(0) & -\beta_1 \phi_{cc}(0) & \alpha_0 \phi_{rr}(0) \\ -\beta_1 \phi_{cc}(0) & \phi_{cc}(0) & 0 \\ \alpha_0 \phi_{rr}(0) & 0 & \phi_{rr}(0) \end{bmatrix} \tag{27}$$

From (10),

$$\mathrm{Var}\, \hat{\beta}_1 \approx \frac{\sigma_v^2 (1+\beta_1^2) + \sigma_u^2\, \alpha_0^2}{S\, \phi_{cc}(0)} \tag{28}$$

$$\mathrm{Var}\, \hat{\alpha}_0 \approx \frac{\sigma_v^2 (1+\beta_1^2) + \sigma_u^2\, \alpha_0^2}{S\, \phi_{rr}(0)} \tag{29}$$

The asymptotic values of the standard least squares estimates along the y_0 axis are found from (14) to be

$$\mathrm{p\,lim}\ \beta_1^* = \beta_1 \frac{1}{1 + \sigma_v^2/\phi_{cc}(0)} \tag{30}$$

$$\mathrm{p\,lim}\ \alpha_0^* = \alpha_0 \frac{1}{1 + \sigma_u^2/\phi_{rr}(0)} \tag{31}$$

The biases are seen to depend on the ratio of the noise variance to the mean-square input or output.

11. Conclusions

The contribution of the present paper lies in applying the method of maximum likelihood to the problem at hand by utilization of Koopmans' general solution to the hyperplane-fitting problem. Some of the properties of the estimates which have been discussed are based on Koopmans' work and others are original results.

These estimates are valid for arbitrary inputs and automatically take into account the initial conditions (stored energy) of the system. The method can easily be extended to include an unknown additive constant (d.c. level) in x(n) and y(n). A continuous system can be handled by approximating it as a sampled-data system. However, the optimum choice of the sampling interval remains to be investigated.

Maximum likelihood estimates of the poles and zeros of the system can be obtained from the maximum likelihood estimates of the coefficients by virtue of the transformation property mentioned in Section 2. The same applies to parameters of a controller which are functions of the coefficients.

Some sampling experiments have been carried out on a desk calculator and have generally supported the theoretical analysis. In applications a digital computer could solve the equations (6) and (7) routinely. Experience indicates that estimates of this nature which are not sensitive to errors in the observed data nevertheless require accurate solutions of the resulting equations. The introduction of approximations such as (16) will often deteriorate the estimates considerably, especially for small S.

REFERENCES

1. R. Price and P. E. Green, Jr., "A Communication Technique for Multipath Channels," Proc. I.R.E. 46, pp.555-570, March 1958.

2. R. E. Kalman, "Design of a Self-Optimizing Control System," Trans. A.S.M.E., 80, pp. 468-478, February 1958.

3. J. P. Ellington and H. McCallion, "The Determination of Control System Characteristics from a Transient Response," Proc. I.R.E., 105, Part C, pp. 370-373, September 1958.

4. M. Shinbrot, "A Description and Comparison of Certain Non-Linear Curve Fitting Techniques with Application to Analysis of Transient Response Data," National Advisory Comm. for Aeronautics Technical Note 2622.

5. Adaptive Control Systems, edited by E. Mishkin and L. Braun, Jr., McGraw-Hill Book Co., New York, pp. 350-354, 1961.

6. G. C. Lendaris, "The Identification of Linear Systems," Applications and Industry, A.I.E.E., 62, pp. 231-242, September 1962.

7. J. Zaborszky and R. L. Berger, "An Integral Square Self-Optimalizing Adaptive Control," Applications and Industry, A.I.E.E., 63, pp. 256-262, November 1962.

8. S. C. Bigelow and H. Ruge, "An Adaptive System Using Periodic Estimation of the Pulse Transfer Function," I.R.E. National Convention Record, Part 4, pp. 25-38, 1961.

9. Y. Kaya and S. Yamamura, "A Self-Adaptive System with a Variable Parameter PID Controller," Applications and Industry, A.I.E.E., 58, pp. 378-386, January 1962.

10. P. Joseph, J. Lewis, and J. Tou, "Plant Identification in the Presence of Disturbances and Applications to Digital Adaptive Systems," Applications and Industry A.I.E.E., No. 53, pp. 18-24, March 1961.

11. H. Kushner, "A Simple Iterative Procedure for the Identification of the Unknown Parameters of a Linear Time-Varying Discrete System," 1962 Joint Automatic Control Conference, New York.

12. M. J. Levin, "Optimum Estimation of Impulse Response in the Presence of Noise," I.R.E. Trans. on Circuit Theory, CT-7, No. 1, pp. 50-56, March 1960.

13. M. J. Levin, "Estimation of the Characteristics of Linear Systems in the Presence of Noise," Eng. Sc. D. dissertation, Dept. of E. E., Columbia University New York, July 1959.

14. H. Cramer, Mathematical Methods of Statistics, Princeton University Press, 1951.

15. T. W. Anderson, An Introduction to Multivariate Statistical Analysis, John Wiley and Sons, Inc., New York, p. 48, 1958.

16. D. V. Lindley, "Regression Lines and the Linear Functional Relationship," Journal of the Royal Statistical Society, (Supplement) 9, pp. 218-244, 1947.

17. C. Villegas, "Maximum Likelihood Estimation of a Linear Functional Relationship," Annals of Math. Stat., 32, pp. 1048-1062, December 1961.

18. T. Koopmans, Linear Regression Analysis of Economic Time Series, DeErven F. Bohn, N. V. Haarlem; 1937.

19. F. B. Hildebrand, Methods of Applied Mathematics, Prentice-Hall, Example 78, p. 115, 1952.

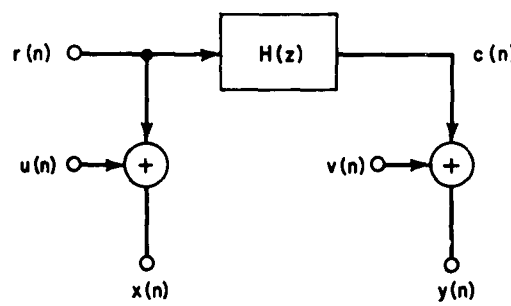

Fig. 1. The Model Assumed for Estimation of the Pulse Transfer Function

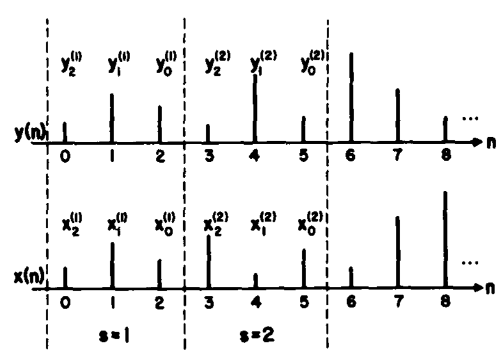

Fig. 2. Arrangement of the Observed Points for K = 2.