Thus

$$H(z) = \frac{D(z)}{N(z)}\left[\frac{zN(z)}{N(z)} - z\right] = z\left[1 - \frac{D(z)}{N(z)}\right].$$

The prediction error at any instant k is

$$\epsilon(k) = \hat{x}(k+1) - x(k+1)$$

or

$$E(z) = H(z)X(z) - zX(z) = -z\frac{D(z)}{N(z)}X(z) = -zY(z)$$

where

$$Y(z) = \frac{D(z)}{N(z)}X(z)$$

as in the maximum likelihood formulation. Therefore,

$$E\{\epsilon^2(k)\} = E\{y^2(k)\} = \frac{1}{2\pi j}\oint \frac{D(z)D(z^{-1})}{N(z)N(z^{-1})}S_{xx}(z)\frac{dz}{z} = \beta^2.$$

Choosing $N(z)$ and $D(z)$ to minimize R causes

$$H(z) = z\left[1 - \frac{D(z)}{N(z)}\right]$$

to be the optimum filter for predicting one step ahead. Hence even in the non-Gaussian case, the minimum residual criterion is valid since if $H(z)$ is the optimum prediction filter, $N(z)$ and $D(z)$ must give a good representation of the actual spectral density. In practice the model numerator and denominator orders K and L are not known a priori. Sufficient values for these parameters can be determined by observing the behavior of the minimum residual for various values of K and L. As K and L are increased the minimum residual will decrease and eventually become essentially constant. At this point it can be assumed that K and L are sufficiently large. This conclusion was verified experimentally by computer simulation. Several nominal and corresponding estimated spectral densities were compared graphically for various values of K and L. Typical results can be found in Tretter and Steiglitz.[2],[4]

STEVEN A. TRETTER
Dept. of Elec. Engrg.
University of Maryland
College Park, Md.

REFERENCES

[1] K. Steiglitz, "Power spectrum identification for adaptive systems," *IEEE Trans. Applications and Industry*, vol. 83, pp. 195–197, May, 1964.
[2] S. A. Tretter and K. Steiglitz, "Power spectrum identification in terms of rational models," *IEEE Trans. Automatic Control* (*Short Papers*), vol. AC-12, pp. 185–188, April 1967.
[3] P. Whittle, *Prediction and Regulation by Linear Least-Squares Methods*. London: English Universities Press, 1963.
[4] S. A. Tretter, "Some problems in the reconstruction and processing of sampled-data," Ph.D. dissertation, Dept. of Elec. Engrg., Princeton University, Princeton, N. J., December 1965.

On the Maximum Likelihood Estimation of Rational Pulse Transfer-Function Parameters

In this correspondence, it is pointed out that the implicit transition from independent to dependent observations contained in the derivation of the likelihood function in a recent paper by Rogers and Steiglitz[1] and the use of the maximum likelihood criterion for dependent observations are not fully justified.

In this case the consistency of the estimates has to be proved as has been done in the previous work of Åström *et al.*[5]

Rogers and Steiglitz[1] considered a method for estimation of pulse transfer-function parameters from input–output data.[2],[3] They noted correctly that if the parameters of the polynomials $A(z)$, $B(z)$, $C(z)$, and $D(z)$ were set to their correct values, i.e., if the parameters to be estimated were known, then the errors $e_i(\theta)$ are independent and

$$L(\theta) = (2\pi\lambda^2)^{-T/2}\exp\left(-\frac{1}{2\lambda^2}\sum e_i^2(\theta)\right) \quad (1)$$

is the likelihood function, the true parameters θ and λ satisfying

$$\sum_i e_i^2(\theta) = \min$$
$$\lambda^2 = \min\left(\frac{1}{T}\sum_i e_i^2(\theta)\right) \quad (2)$$

where T is the record length.

During the estimation procedure, however, it generally holds that the nth estimate $\theta_n \neq \theta$, and the errors $e_i(\theta_n)$ cease to be independent. Hence the preceding assumptions are not true and the estimates generally do not have the desirable maximum likelihood properties[4] claimed by the authors.

Only under certain (nontrivial) assumptions[3] it can be shown that for large record lengths the estimates are consistent and the method asymptotically efficient (properties of the maximum likelihood estimator).

The proof is rather difficult and can be found in Åström *et al.*[5] A similar problem is treated in Mann and Wald.[6]

V. PANUSKA
Dept. of Engrg. Science
Oxford University
Oxford, England

REFERENCES

[1] A. E. Rogers and K. Steiglitz, "Maximum likelihood estimation of rational transfer function parameters," *IEEE Trans. Automatic Control* (*Short Papers*), vol. AC-12, pp. 594–597, October 1967.
[2] K. J. Åström, "Control problems in paper making," *Proc. IBM Sci. Computing Symp.—Control Theory and Applications* (October 1964), pp. 135–167.
[3] K. J. Åström and T. Bohlin, "Numerical identification of linear dynamic systems from normal operating records," *Proc. Internat'l Fed. Automatic Control Symp. on Adaptive Control* (Teddington, Middlesex, England, 1965).
[4] H. Cramér, *Mathematical Methods of Statistics*. Princeton, N. J.: Princeton University Press, 1946.
[5] K. J. Åström, T. Bohlin, and S. Wensmark, "Automatic construction of linear stochastic dynamic models for stationary industrial processes with random disturbances using operating records," IBM Nordic Lab., Tech. Paper TP 18. 150, 1965.
[6] H. B. Mann and A. Wald, "On the statistical treatment of linear stochastic difference equations," *Econometrica*, vol. 11, nos. 3 and 4, 1943.

Manuscript received January 19, 1968.

Authors' Reply[1]

Two separate questions are raised by Panuska. The first concerns the construction of the correct likelihood function, and the second deals with the properties of the estimates obtained from its maximization.

With regard to the first question, the likelihood function derived in our paper[1] is indeed correct; it is the probability density of the observations z conditioned on the parameter vector θ which is, of course, unknown. The premise that it is known is used only to write the density function $p(z|\theta)$. The purpose of our paper was to present a method for calculating and maximizing this likelihood function and to report the results of numerical examples which demonstrate that the method is superior to least-square fitting of the output sequence when the noise is correlated. No properties of this maximum likelihood estimation procedure were claimed in our paper.[1]

Panuska is quite right, however, in bringing up the second question, and in citing the theoretical results in Åström *et al.*[5] The restrictions which ensure consistency and asymptotic efficiency are, as mentioned in Åström and Bohlin,[3] mild. Presumably, the results are valid for our (slightly different) model as well. These conditions concern regularity assumptions about the input sequence, the excitation characteristics of the input, and the stability and controllability of the system. All were satisfied in our experiments.

We might take this opportunity to mention that an experimental comparison between our method[1] and the eigenvalue method of Koopmans[7] and Levin[8] has been made with both input and output noise,[9] which is currently being prepared for publication. Of particular interest for further investigation is the relative quality of the methods using stochastic approximation[10],[11] and instrumental variables[12],[13] reported recently.

A. E. ROGERS
Dept. of Statistics and Computer Science
University of Delaware
Newark, Del. 19711
K. STEIGLITZ
Dept. of Elec. Engrg.
Princeton University
Princeton, N. J. 08540

REFERENCES

[7] T. Koopmans, *Linear Regression Analysis of Economic Time Series*. N. V. Haarlem, Netherlands: De Erven F. Bohn, 1937.
[8] M. J. Levin, "Estimation of a system pulse transfer function in the presence of noise," *IEEE Trans. Automatic Control*, vol. AC-9, pp. 229–235, 1964.
[9] A. E. Rogers, "Maximum likelihood estimation of dynamic system parameters," Ph.D. dissertation, Dept. of Elec. Engrg., Princeton University, Princeton, N. J., 1968.
[10] D. J. Sakrison, "The use of stochastic approximation to solve the system identification problem," *IEEE Trans. Automatic Control*, vol. AC-12, pp. 563–567, October, 1967.
[11] G. N. Saridis and G. Stein, "Stochastic approximation algorithms for linear discrete-time system identification," *1967 Proc. Nat'l Electronics Conf.*, vol. 23, pp. 45–50.

[1] Manuscript received February 12, 1968. This work was supported in part by the Army Research Office, Durham, N. C., under Contract DA-31-124-ARO-D-292.

[12] P. Joseph, J. Lewis, and J. Tou, "Plant identification in the presence of disturbances and applications to digital adaptive systems," *AIEE Trans. (Applications and Industry)*, vol. 80, pp. 18–24, March 1961.

[13] K. Y. Wong and E. Polak, "Identification of linear discrete time systems using the instrumental variable method," *IEEE Trans. Automatic Control*, vol. AC-12, pp. 707–718, December 1967.

On the Evaluation and Sensitivity of Integral Squared Functions

The evaluation of the integral square

$$I = \int_0^\infty y^2(t)dt \tag{1}$$

has long occupied the interest of engineers, dating from the text presentation of James, Nichols, and Phillips,[1] and Newton, Gould, and Kaiser[2] in which tables for the evaluation of the integral in terms of polynomial coefficients are given. Recently Effertz,[3] with comments by MacMillan and Pazdera,[4] has indicated that these integrals may be evaluated in terms of a generalized form of a Routh's array involving the polynomial coefficients. Kalman and Bertram,[5] and Jain and Koenig[6] have also considered the evaluation of these integrals using state methods. It is interesting to note that the integral can be expressed directly in terms of the numerator and denomination polynomials of $y(s)$ and derivatives evaluated at pole location. In addition to being in closed form, this formulation lends itself to sensitivity considerations.

If

$$y(t) = \sum_{i=1}^{m} \alpha_i \epsilon^{-\lambda_i t} \tag{2}$$

and

$$y(s) = \sum_{i=1}^{m} \frac{\alpha_i}{s + \lambda_i} \tag{3}$$

then it can be shown, by direct substitution or by matrix methods, that (1) becomes

$$I = \sum_{i,j=1}^{m} \frac{\alpha_i \alpha_j}{\lambda_i + \lambda_j} = \sum_{i=1}^{m} \alpha_i y(\lambda_i). \tag{4}$$

If

$$Y(s) = \frac{P(s)}{Q(s)} = \frac{K \prod_{j=1}^{n} (s + z_j)}{\prod_{k=1}^{m} (s + \lambda_k)}, \quad n < m \tag{5}$$

then α_i is given by

$$\alpha_i = \frac{P(-\lambda_i)}{Q'(-\lambda_i)} = \frac{K \prod_{j=1}^{n} (z_j - \lambda_i)}{\prod_{\substack{k=1 \\ k \neq i}}^{m} (\lambda_k - \lambda_i)} \tag{6}$$

and (4) can be expressed as

$$I = \sum_{i=1}^{m} \frac{P(-\lambda_i)P(\lambda_i)}{Q'(-\lambda_i)Q(\lambda_i)} = \sum_{i=1}^{m} \frac{K^2}{2\lambda_i} \frac{\prod_{j=1}^{n} (z_j^2 - \lambda_i^2)}{\prod_{\substack{k=1 \\ k \neq i}}^{m} (\lambda_k^2 - \lambda_i^2)}. \tag{7}$$

If $Y(s)$ has a kth order pole at $s = -\lambda_1$,

$$Y(s) = \frac{\alpha_1}{s + \lambda_1} + \cdots + \frac{\alpha_k}{(s + \lambda_1)^k} + \sum_{i=k+1}^{m} \frac{\alpha_i}{(s + \lambda_i)} \tag{8}$$

then (1) can be expressed as

$$I = -\alpha_2 Y'(\lambda_1) + \alpha_3 Y''(\lambda_1) + \cdots + \alpha_k (-1)^{k-1} Y^k(\lambda_1) + \sum_{\substack{i=1 \\ i \neq 2,3,\cdots,k}}^{m} \alpha_i Y(\lambda_i). \tag{9}$$

Equation (7) expresses the integral square in terms of the pole locations of $Y(s)$. Should the pole locations of λ_i vary, the incremental affect upon I may be calculated directly. If

$$I_i = \frac{P(-\lambda_i)}{Q'(-\lambda_i)} \frac{P(\lambda_i)}{Q(\lambda_i)} \tag{10}$$

then

$$I = \sum_{i=1}^{m} I_i \tag{11}$$

and

$$dI = \sum_{i=1}^{m} \frac{\partial I_i}{\partial \lambda_i} d\lambda_i. \tag{12}$$

From (7) it can be shown that

$$\frac{1}{I_i} \frac{\partial I_i}{\partial \lambda_i} = -\frac{P'(-\lambda_i)}{P(-\lambda_i)} + \frac{P'(\lambda_i)}{P(\lambda_i)} - \frac{Q'(\lambda_i)}{Q(\lambda_i)} + \frac{Q''(-\lambda_i)}{Q'(-\lambda_i)}. \tag{13}$$

Alternately if

$$I_i = \frac{K^2}{2\lambda_i} \frac{\prod_{j=1}^{n} (z_j^2 - \lambda_i^2)}{\prod_{\substack{k=1 \\ k \neq 1}}^{m} (\lambda_k^2 - \lambda_i^2)} \tag{14}$$

then

$$\ln I_i = \ln \frac{K^2}{2\lambda_i} + \sum_{j=1}^{n} \ln(z_j^2 - \lambda_i^2) - \sum_{\substack{k=1 \\ k \neq i}}^{m} \ln(\lambda_k^2 - \lambda_i^2) \tag{15}$$

and

$$\frac{1}{I_i} \frac{\partial I_i}{\partial \lambda_i} = -\frac{1}{\lambda_i} - \sum_{j=1}^{n} \frac{2\lambda_i}{z_j^2 - \lambda_i^2} + \sum_{\substack{k=1 \\ k \neq i}}^{m} \frac{2\lambda_i}{\lambda_k^2 - \lambda_i^2}. \tag{16}$$

Equations (12) and (13) or (16) express the sensitivity of the integral square function for variations of a pole location. If a parameter x in $Y(s)$ varies, then the effect of this variation can also be calculated. Thus from (11)

$$\frac{dI}{dx} = \sum_{i=1}^{m} \frac{dI_i}{dx} = \sum_{i=1}^{m} \frac{\partial I_i}{\partial \lambda_i} \frac{\partial \lambda_i}{\partial x}. \tag{17}$$

If $Q(s) = C(s) + xD(s)$, where $C(s)$ and $D(s)$ are polynomials in s, then, using a result of Kuo[7] it can be shown that

$$\frac{\partial \lambda_i}{\partial x} = -\frac{D(-\lambda_i)}{Q'(-\lambda_i)} \tag{18}$$

where $D(-\lambda_i)/Q'(-\lambda_i)$ represents the residue of $D(s)/Q(s)$ evaluated at $s = -\lambda_i$. Combining (12), (13), and (18) yields the result

$$\frac{dI_i}{dx} = \left[\frac{Q'(\lambda_i)}{Q(\lambda_i)} - \frac{Q''(-\lambda_i)}{Q'(\lambda_i)} + \frac{P'(-\lambda_i)}{P(-\lambda_i)} - \frac{P'(\lambda_i)}{P(\lambda_i)}\right] \frac{D(-\lambda_i)}{Q'(-\lambda_i)} I_i. \tag{19}$$

Sydney R. Parker
United States Naval Postgraduate School
Monterey, Calif.

Manuscript received February 7, 1968.

References

[1] M. H. James, N. B. Nichols, and R. S. Phillips, *Theory of Servomechanisms.* New York: McGraw Hill, 1947, pp. 323–325, 333–339, 369–370.

[2] G. C. Newton, Jr., L. A. Gould, and J. F. Kaiser, *Analytical Design of Linear Feedback Controls.* New York: Wiley, 1957, pp. 366–381.

[3] F. H. Effertz, "On two coupled matrix algorithms for the evaluation of the rms error criterion in linear systems," *Proc. IEEE (Letters)*, vol. 54, pp. 879–880, June 1966.

[4] A. J. MacMillan, F. H. Effertz, and J. S. Pazdera, "Comments on two coupled matrix algorithms for the evaluation of rms error criterion of linear systems," *Proc. IEEE (Letters)*, vol. 54, pp. 1626–1629, November 1966.

[5] R. E. Kalman and J. E. Bertram, "Control system analysis and design via the second method of Lyapounov," pt. 1, *Trans. ASME, J. Basic Engrg.*, ser. D, vol. 82, p. 388, June 1966.

[6] V. K. Jain and H. E. Koenig, "On parameter optimization of linear systems," *IEEE Trans. Automatic Control (Correspondence)*, vol. AC-11, pp. 753–754, October 1966.

[7] F. F. Kuo, "Pole zero sensitivity network functions," *IRE Trans. Circuit Theory (Correspondence)*, vol. CT-5, pp. 372–373, December 1958.

On a Bounded Input–Bounded Output Sufficient Condition for Nonlinear Systems

***Abstract*—Using some theorems of the theory of nonlinear Volterra integral equations, a sufficient condition is derived for the boundedness of response of a class of nonlinear control systems. As a consequence, an estimate is given for the upper bound of the response of systems subjected to amplitude limited signals.**

Introduction

The problem of finding upper bounds for the responses of particular classes of nonlinear systems has been considered by many authors.[1]–[15] In this correspondence, feedback systems are considered which contain a nonlinear element characterized by a zero-memory function.

The aim here is to derive a sufficient condition for the boundedness of the responses to amplitude limited input signals.

Manuscript received December 28, 1967.