The control law as specified in (9) is open-loop in nature. It may be converted to a closed-loop control law by using the principle of optimality. Let the optimal control be given by

$$U^0 = G_j X_0 \qquad \text{for } j = 1, 2, 3, \cdots \tag{10}$$

where

$$\begin{aligned} G_j &= -(H_j^T H_j)^{-1} H_j^T \phi(N, 0) X_0 && \text{for } jm < n \\ &= F_j^T (F_j F_j^T)^{-1} X_0 && \text{for } jm \geq n. \end{aligned}$$

Let g_j^m be the $(m \times n)$ matrix which consists of the first m rows of G_j. Direct application of the principle of optimality gives the desired optimal closed-loop control law.

$$U_k^0 = g_{N-k}^m X_k \qquad \text{for } k = 0, 1, 2, \cdots, N-1. \tag{11}$$

IV. Numerical Example

The system under study is governed by the second-order differential equation

$$\frac{d^2c}{dt^2} + \frac{dc}{dt} = U(t). \tag{12}$$

If the input, $U(t)$, is constrained to be constant over one-second intervals, the system of (12) is characterized by the discrete state representation

$$X_{k+1} = AX_k + BU_k$$

where

$$X_k = \begin{bmatrix} C(t) \\ \dfrac{dC(t)}{dt} \end{bmatrix} \qquad \text{at } t = k \text{ seconds}$$

$$A = \begin{bmatrix} 1 & 1 - e^{-1} \\ 0 & e^{-1} \end{bmatrix}, \quad B = \begin{bmatrix} e^{-1} \\ 1 - e^{-1} \end{bmatrix}$$

and $U(t) = U_k$ a constant for $k \leq t < k+1$.

Problem

Find the control inputs U_0, U_1, U_2, U_3 which will drive the system of (12) from any initial state to the zero state in four sample periods.

Solution

Apply the results of Section III, Case $mN \geq n$. The vectors f_k which form the columns of F are given by

$$f_k = -A^{-k+1}B \qquad \text{for } k = 0, 1, 2, 3 \tag{13}$$

since for time-invariant systems

$$\phi(k, j) = A^{k-j}.$$

Evaluating f_0, f_1, f_2 and f_4 by means of (13), one obtains (note $mN > n$, i.e., $1 \times 4 > 2$).

$$f_0 = \begin{bmatrix} 0.7183 \\ -1.7183 \end{bmatrix}, \qquad f_1 = \begin{bmatrix} 3.6708 \\ -4.6708 \end{bmatrix}$$

$$f_2 = \begin{bmatrix} 11.6965 \\ -12.6965 \end{bmatrix}, \qquad f_3 = \begin{bmatrix} 33.5126 \\ -34.5126 \end{bmatrix}.$$

Since f_0 and f_1 are linearly independent vectors, this system is completely controllable [1].

From (5), for $N = 4$, F is given by

$$F = \begin{bmatrix} 0.7183 & 3.6708 & 11.6965 & 33.5126 \\ -1.7183 & -4.6708 & -12.6965 & -34.5126 \end{bmatrix}.$$

Thus

$$FF^T = \begin{bmatrix} 1273.8934 & -1323.4916 \\ -1323.4616 & 1377.0897 \end{bmatrix}$$

therefore

$$(FF^T)^{-1} = \begin{bmatrix} 0.5225 & 0.5022 \\ 0.5022 & 0.4833 \end{bmatrix}.$$

The open-loop optimal control law as given by (7) becomes

$$\begin{bmatrix} U_0 \\ U_1 \\ U_2 \\ U_3 \end{bmatrix} = \begin{bmatrix} -0.4876 & -0.4698 \\ -0.4275 & -0.4143 \\ -0.2643 & -0.2632 \\ 0.1794 & 0.1473 \end{bmatrix} \begin{bmatrix} C(0) \\ \dfrac{dC(0)}{dt} \end{bmatrix}. \tag{14}$$

Given any initial state $X_0^T = [C(0), dC(0)/dt]$, the optimal control is determined by performing the matrix multiplication given in (14).

V. Conclusion

A method has been presented which synthesizes the minimum energy regulator using elementary inverse matrix theory. An open-loop control law results which may be converted to a closed-loop control law by applying the principle of optimality. The optimal control given by (11) requires $mn \times n$ multiplications which in many cases may be rapidly made on existing digital computers. A real-time controller is therefore possible.

References

[1] H. Freeman, *Discrete-Time Systems*. New York: Wiley, 1965.
[2] B. Friedland, "Design of optimal controllers for linear processes with energy limitations," *J. Basic Engrg., Trans. ASME*, ser. D, vol. 85, pp. 181–196, June 1963.
[3] J. E. Bertram and P. E. Sarachik, "On optimal computer control," *Proc. 1st IFAC (Moscow) Congress*. London: Butterworths, 1960, vol. 1, pp. 419–422.
[4] J. A. Cadzow, "A study of minimum norm control for sampled-data systems," *Proc. JACC*, pp. 545–550, 1965.
[5] R. Penrose, "A generalized inverse of matrices," *Proc. Cambridge Phil. Soc.*, vol. 51, pp. 406–413, 1955.
[6] R. E. Kalman, Y. C. Ho, and K. S. Narendra, "Controllability of linear dynamical systems," in *Contributions to Differential Equations*. New York: Wiley, 1962, vol. 1, no. 2.

Power-Spectrum Identification in Terms of Rational Models

S. A. TRETTER AND K. STEIGLITZ, MEMBER, IEEE

***Abstract*—A technique is described for the identification of unknown power-spectral densities from sampled data in terms of a rational function of z. The problem is reduced to the minimization of a function of K parameters, where K is the order of the numerator of the model. This criterion, called the "minimum residual" criterion, reduces to the maximum likelihood criterion when the observed signal is Gaussian. A computational technique is described for minimizing this function which uses filtering and correlation to obtain the gradient and an iterative descent method due to M. J. D. Powell for minimization. Some computational results are given in which the method is compared with all-pole and conventional spectrum estimation techniques.**

I. Introduction

The problem of estimating the power-spectral density of a signal from a finite record is of considerable interest to the engineer and statistician [1]–[3]. The methods developed center on estimating the power of the signal in narrow frequency bands. In the continuous case, this can be accomplished by passing the signal through a bandpass filter and measuring the average power of the output. In the discrete case, a spectral window which corresponds to this bandpass filter is used to obtain spectral-density estimates from the finite cosine transform of the estimated correlation function.

The final product of these methods is a plot of the estimated spectral density as a function of frequency. This is analogous to determining the transfer function of a linear system by measuring its response

Manuscript received January 7, 1966; revised August 24, 1966, and December 22, 1966. This work was supported partly by the Army Research Office Durham under ARO Contract DA-31-124-ARO-D-292, by the National Science Foundation under Grant NSF GP-1647, and by a Princeton Predoctoral Fellowship sponsored by the Ford Foundation. Use was made of computer facilities supported in part by the National Science Foundation under Grant NSF GP-579. A more detailed version of this short paper was presented at the 1966 JACC, Seattle, Wash.

S. A. Tretter is with the Department of Electrical Engineering, University of Maryland, College Park, Md.
K. Steiglitz is with the Department of Electrical Engineering, Princeton University, Princeton, N. J.

to sinusoidal inputs at various frequencies. In many applications, this approach yields the desired information. In adaptive control, communication, or detection systems, where the law of adaptation may depend on the spectral densities of system signals, simple parametric models of the spectral density are more useful, just as parametric models for plant transfer functions are needed in adaptive control systems.

Heretofore the only sampled spectral-density model that has been investigated to any extent is the all-pole function of z [4]. This spectrum model is equivalent to an autoregressive model for the signal [1]. In this case, the inversion of an estimated covariance matrix yields the maximum likelihood estimate of the unknown parameters if the signal is Gaussian. The behavior of the all-pole model has been investigated in the adaptive matched filter problem [5] and in the adaptive reconstruction problem [6]. It has been found, however, that when the actual signal spectrum has influential zeros, quite a few poles are needed in the all-pole model to achieve a good approximation. The introduction of zeros in the model would therefore allow a more efficient representation and a more effective identification.

In this short paper, a method is described for the identification of the sampled power-spectral density in terms of a rational function of z with preassigned numerator and denominator orders. The computational procedure involved combines surface-searching for the numerator coefficients with all-pole identification for the denominator coefficients. Computational results are presented which illustrate the utility of the rational model.

II. Minimum Residual Formulation

We shall assume that we have available a sequence $x(1), \cdots, x(M)$ which is a record of samples from a discrete, wide-sense stationary, ergodic random process with rational spectral density of the form

$$\Phi_{xx}(z) = \beta^2 \frac{N(z)N(z^{-1})}{D(z)D(z^{-1})} \tag{1}$$

where

$$N(z) = \sum_{n=0}^{K} a_n z^{-n} \quad \text{and} \quad D(z) = \sum_{n=0}^{K} b_n z^{-n}$$

are polynomials in z^{-1} of order K and L, respectively, with $a_0 = b_0 = 1$ and with roots inside the unit circle. β^2 is a positive constant. This model can be justified in many physical situations—it represents the result of filtering white noise through a digital filter with a rational transfer function.

If we assume that the random process is Gaussian, the method of maximum likelihood can be used to estimate the coefficients of N, D, and β^2 in the following way. Suppose the signal x is passed through a digital filter $D(z)/N(z)$ to give the output y. The spectral density of y is

$$\Phi_{yy}(z) = \frac{D(z)D(z^{-1})}{N(z)N(z^{-1})} \Phi_{xx}(z).$$

If one could choose the coefficients of N and D so that $\Phi_{yy}(z) = \beta^2$, the spectral density of x would be given by (1). When this is the case, y is white Gaussian noise and the joint probability density of the observed output sequence would then be

$$p[y(1), \cdots, y(M)] = \frac{1}{(2\pi)^{M/2}\beta^M} \exp(-R/2\beta^2), \tag{2}$$

where

$$R = \sum_{i=1}^{M} y^2(i)$$

will be called the residual. The maximum likelihood parameter estimates are therefore given by the solution to the following set of equations [7]:

$$\frac{\partial R}{\partial a_n} = 0, \qquad n = 1, \cdots, K,$$
$$\frac{\partial R}{\partial b_n} = 0, \qquad n = 1, \cdots, L, \tag{3}$$

and

$$\beta^2 = R/M.$$

Thus the problem is equivalent to finding the minimum of a function of several variables. The conditions (3) will be called the minimum residual criterion. This condition, which, intuitively, tries to "whiten" x as much as possible, appears to be meaningful even when the signal is not Gaussian.

III. Solution in the All-Pole Case

As mentioned above, when the order of $N(z)$ is zero, the minimum residual criterion yields a set of linear equations for the b_n. We now summarize the results in this case.

Let f_i be the mean lagged products of x defined by

$$f_i = \frac{1}{M-i} \sum_{k=1}^{M-i} x(k)x(k+i), \qquad 0 \le i \le L,$$

let $\boldsymbol{F}$ be the estimated covariance matrix $\{f_{|i-j|};\ i, j = 1, \cdots, L\}$, let $\boldsymbol{f}$ be the column vector $\{f_i;\ i = 1, \cdots, L\}$, and let $\boldsymbol{b}$ be the column vector $\{b_i;\ i = 1, \cdots, L\}$. Then the minimum residual criterion leads to the estimates

$$\boldsymbol{b} = -\boldsymbol{F}^{-1}\boldsymbol{f},$$
$$\beta^2 = f_0 + \boldsymbol{b}'\boldsymbol{f}, \tag{4}$$

which requires besides the calculation of the mean lagged products only the inversion of one $L \times L$ matrix.

IV. Case when $N(z)$ is of Nonzero Order

When there are unknown parameters in the numerator, the minimum residual criterion leads to nonlinear regression equations which cannot be solved explicitly. This suggests the using of an iterative technique to optimize the free parameters of the model. Many such techniques are available, ranging from steepest descent to the Newton-Raphson algorithm. An algorithm due to M. J. D. Powell [8]–[10] was found by the authors to be quite satisfactory for this problem.

Although it is possible to perform a direct $(K+L)$-dimensional minimization for the unknown a_n and b_n, only a K-dimensional search for the a_n is necessary. This is so, because for any given $N(z)$ of order K, the optimum $D(z)$ of order L can be found directly by applying the all-pole estimation procedure specified by (4) to the signal $X(z)/N(z)$. A two-stage estimation procedure was therefore programmed with $N(z)$ and $D(z)$ being estimated alternately by Powell's method and by matrix inversion, respectively.

V. Calculation of the Gradient by Filtering

The majority of the more efficient extremal seeking algorithms, including Powell's method, employ the gradient of the function whose extreme value is being sought. While it is always possible to use finite differences to estimate the gradient, it has been found that the method of sensitivity filters [11] is more accurate and efficient.

From (3) we have

$$\frac{\partial R}{\partial a_i} = 2 \sum_{k=1}^{M} y(k) \frac{\partial y(k)}{\partial a_i}. \tag{5}$$

Now

$$Y(z) = \frac{D(z)}{N(z)} X(z), \tag{6}$$

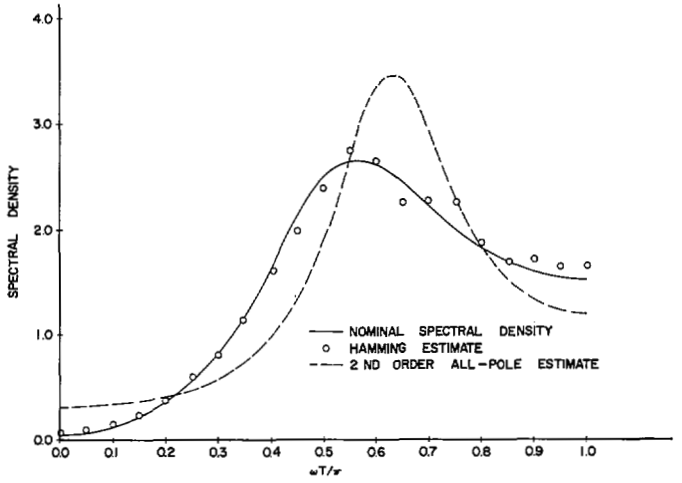

Fig. 1. Second-order all-pole estimate compared with the nominal density and the Hamming estimate.

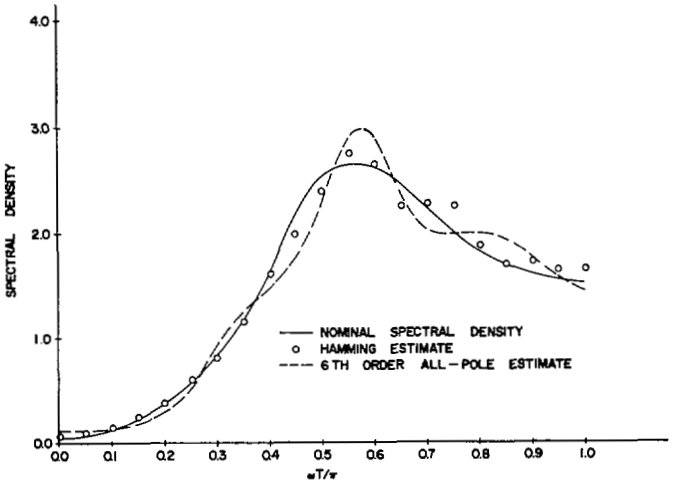

Fig. 2. Sixth-order all-pole estimate compared with the nominal density and the Hamming estimate.

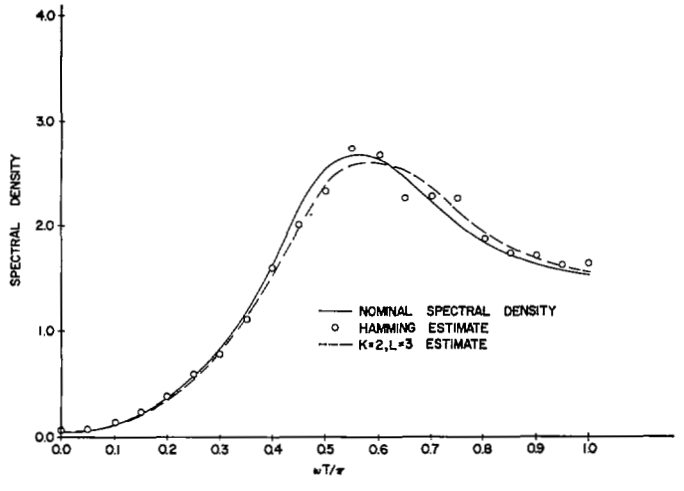

Fig. 3. Rational estimate for $K=2$ and $L=3$ compared with the nominal density and the Hamming estimate.

so that

$$\frac{\partial Y(z)}{\partial a_i} = -\frac{D(z)}{N^2(z)} z^{-i} X(z). \tag{7}$$

Hence

$$\frac{\partial R}{\partial a_i} = -2 \sum_{k=1}^{M} y(k) v(k-i), \tag{8}$$

where the signal

$$V(z) = \frac{Y(z)}{N(z)} = \frac{D(z)}{N^2(z)} X(z) \tag{9}$$

is obtained by filtering the observed signal by the indicated linear filter. Thus the gradient of R with respect to the a_i may be obtained by filtering and correlating appropriate signals. The gradient $\partial R/\partial b_i$ may be obtained in a similar manner but is not needed here.

VI. Choice of Model Order

Unless some a priori knowledge is available, the choice of K and L presents a serious problem, as in all identification problems. Normally, the effect on the minimum residual of increasing K or L by one will provide some guide as to the appropriateness of the model. Various schemes have been proposed for testing the significance of an additional parameter in the all-pole (autoregressive) case. These are discussed in Bartlett [12] and Whittle [13]. To the authors' knowledge, no results are available in the rational case.

VII. Computational Results

In this section, some typical results of a computer simulation are described. Five thousand independent normal random numbers with zero mean and unit variance were passed through the digital filter

$$H(z) = \frac{(1-0.3z^{-1})(1-0.8z^{-1})}{(1-j0.5z^{-1})(1+j0.5z^{-1})(1-0.5z^{-1})}$$

to generate a sequence with the nominal spectral density $H(z)H(z^{-1})$. This nominal density is compared with a second-order all-pole estimate and the Hamming estimate using 21 mean lagged products in Fig. 1. The estimated all-pole parameters were

$$\beta^2 = 1.115678$$
$$D(z) = 1 + 0.44699z^{-1} + 0.398878z^{-2}.$$

The nominal density is compared with a sixth-order all-pole estimate and the Hamming estimate in Fig. 2. The estimated parameters were

$$\beta^2 = 1.013372$$
$$D(z) = 1 + 0.598049z^{-1} + 0.628833z^{-2} + 0.404799z^{-3} + 0.270027z^{-4} + 0.165984z^{-5} + 0.091338z^{-6}.$$

Finally, in Fig. 3, the rational estimate for $K=2$ and $L=3$ is shown together with the nominal density and the Hamming estimate. The estimated parameters in this case were

$$\beta^2 = 1.002461$$
$$N(z) = 1 - 0.578249z^{-1} - 0.135121z^{-2}$$
$$D(z) = 1 + 0.034511z^{-1} + 0.170013z^{-2} - 0.012531z^{-3}.$$

The values for β^2 for various order models are listed below:

K	L	β^2
0	2	1.115678
0	3	1.058848
0	4	1.034949
0	5	1.021898
0	6	1.013372
0	7	1.007076
0	8	1.005578
0	9	1.002803
0	10	1.001860
0	11	1.001234
2	3	1.002461

On observing Fig. 1, it is clear that the second-order all-pole estimate is a relatively rough estimate of the nominal density. The sixth-order all-pole estimate is substantially better but still tends to exhibit ripple. The $K=2$, $L=3$ estimate, which is from a priori information the correct model, fits the nominal density quite well. Taking the residual as a measure of the quality of the fit, it appears that at least a tenth-order all-pole model is needed to surpass the $K=2$, $L=3$ model estimate. For all-pole estimates of order greater than ten, plots of the estimated densities compare well with the nominal density. This illustrates the usefulness of all-pole models of sufficiently high order even in the case of spectral densities containing zeros.

With fewer sample points in the input record, the statistical fluctuations in the spectral estimates become more pronounced and tend to swamp out the model error. In one example with the same nominal density as above but with only 250 data points, a sixth-order all-pole approximant, while more oscillatory than a $K=2$, $L=3$ approximant, seemed to be quite adequate.

VIII. Conclusions

A method has been described for identifying spectral densities in terms of rational functions of z. The special case of all-pole identification has been extended to include zeros in the model. This problem is typical of many which have become practical to do only with digital computers.

Experimental results have indicated that, for models of appropriate order, the quality of the estimation is at least equivalent to that obtained from bandpass filters and Hamming windows. Furthermore, the results in a simple parametric form are more useful in many applications, especially in the design of adaptive systems. The question of how to choose the model order remains unsettled, although the behavior of the residual for various model orders gives some indication of the appropriateness of the model. Similar problems arise in conventional spectrum analysis, where one must choose the number of mean lagged products and the bandwidth of the spectral window.

Acknowledgment

The authors would like to thank P. Woodrow for supplying his program of Powell's method, and Prof. L. McBride for many interesting discussions relating to the identification problem.

The authors would like to mention the work of K. J. Astrom [14], which recently came to their attention. Astrom treats the process identification problem with white noise excitation by constructing a likelihood function in a manner similar to ours. He describes the results of computations using the Newton–Raphson and Gauss–Newton methods.

References

[1] U. Grenander and M. Rosenblatt, *Statistical Analysis of Stationary Time Series*. New York: Wiley 1957, pp. 11–115.
[2] R. B. Blackman and J. W. Tukey, *The Measurement of Power Spectra*. New York: Dover, 1959.
[3] S. S. L. Chang, *Synthesis of Optimum Control Systems*. New York: McGraw-Hill, 1961.
[4] K. Steiglitz, "Power spectrum identification for adaptive systems," *IEEE Trans. on Application and Industry*, vol. 83, pp. 195–197, May 1964.
[5] K. Steiglitz and J. B. Thomas, "A class of adaptive matched digital filters," *Proc. NEC*, vol. 20, pp. 641–646, October 1964.
[6] S. A. Tretter and K. Steiglitz, "Adaptive signal reconstruction," *Proc. NEC*, vol. 21, pp. 487–492, October 1965.
[7] H. Cramer, *Mathematical Methods of Statistics*. Princeton, N. J.: University Press, 1946.
[8] M. J. D. Powell, "An iterative method for finding stationary values of a function of several variables," *The Computer J.*, vol. 5, pp. 147–151, July 1962.
[9] P. J. Woodrow, "Methods for locating the extrema of multivariate functions," Control Systems Lab., Dept. of Electrical Engineering, Princeton University, Princeton, N. J., Tech. Rept. 4, November 1964.
[10] P. J. Woodrow, "Description of subroutines for Powell's method," Aeronautical Research Associates of Princeton, Inc., Princeton, N. J., ARAP Working Paper 63-2, November 1963.
[11] K. S. Narendra and L. E. McBride, "Multiparameter self-optimizing systems using correlation techniques," *IEEE Trans. on Automatic Control*, vol. AC-9, pp. 31–38, January 1964.
[12] M. S. Bartlett, *An Introduction to Stochastic Processes*. Cambridge, England: University Press, 1961, ch. 9.
[13] P. Whittle, *Prediction and Regulation by Linear Least-Squares Methods*. London: English Universities Press, 1963, section 3.4.
[14] K. J. Astrom, "Control problems in papermaking," *Proc. IBM Scientific Computing Symp., Control Theory and Applications*, White Plains, N. Y., 1966.

Measurements for the Control of a Randomly Time-Varying Linear System

G. E. POLLON

***Abstract*—Measurements are presented which, performed on a stationarily-randomly time-varying linear system, yield sufficient information about the system to enable one to design linear time-invariant open-loop compensation. The measurements are demonstrated on an electronic randomly time-varying linear system and the design technique is applied to an example with interesting results.**

Introduction

Much of the literature on the measurement of time-variant linear systems is concerned with the time and frequency spread characteristics of radar targets and communication channels (scattering function, two-frequency correlation function, and tap-gain correlation function). Methods by which these characteristics can be measured have been considered by Green [2], [6], Hagfors [4], Gallagher [5], Levin [14], and others. Theories for the measurement of the statistics of a general stationarily-randomly time-varying linear system exist, but being aimed at a complete characterization, they contain great complexity [1], [7].

The measurements presented here are directed at obtaining just sufficient data on the system to enable one to proceed analytically to determine the optimum (mean-square-error criterion) linear time-invariant open-loop controller. Since it is asked only to control the system, not to uniquely characterize it, the measurements required are relatively simple. The only additional restriction of physical significance is that the system vary in a stationary manner. Significantly, there are no restrictions on the rate at which the system may vary. On the contrary, the utility of the design procedure advocated here is the greatest for systems whose parameters fluctuate rapidly, since if the system varies slowly, an adaptive approach is preferable.

A randomly time-varying linear system (RTVLS) will be defined as one whose input-output relationship can be expressed as,

$$y(t) = \int_{-\infty}^{\infty} h(t, \sigma)u(t-\sigma)d\sigma: \qquad h(t, \sigma) = 0 \quad \text{for } \sigma < 0 \tag{1}$$

where $h(t, \sigma)$ is a sample from a two-dimensional stochastic process, stationary in the t-direction, and independent of the input $u(t)$. $h(t, \sigma)$ is the system response at time t to an impulse applied at time $t-\sigma$. It is further assumed that $\int_0^\infty h^2(t, \sigma)d\sigma < \infty$.

It can easily be shown that the design problem shown in Fig. 1(a) is identical to the one depicted in Fig. 1(b), where $f(\sigma)$ is a linear time-invariant system with impulse response,

$$f(\sigma) = \overline{h(t, \sigma)} \tag{2}$$

where $\overline{(\quad)}$ denotes ensemble average, and $v(\sigma)$ is a (minimum phase) linear time-invariant system with autotranslation function, related to the "spectrum" of the variations that the system undergoes, given by,

$$I_{vv}(\tau) = \int_0^\infty \overline{h(t, \sigma)}\,\overline{h(t, \sigma-\tau)}d\sigma - \int_0^\infty \overline{h(t, \sigma)}\,\overline{h(t, \sigma-\tau)}d\sigma. \tag{3}$$

However, the problem in Fig. 1(b) is the same as Newton's saturation tendency constraint problem [3]. Thus once the hypothetical systems $f(\sigma)$ and $v(\sigma)$ have been determined, the mathematical solutions for $w_0(\sigma)$ are identical.

Manuscript received March 17, 1966; revised September 9, 1966, and December 30, 1966. This paper is based upon part of a Doctor of Science thesis in the Department of Electrical Engineering at the Massachusetts Institute of Technology and was supported by Contract NOW-61-0400.

The author is presently with General Dynamics/Pomona Division, Pomona, Calif.