AD NUMBER

AD441527

LIMITATION CHANGES

TO:

Approved for public release; distribution is unlimited. Document partially illegible.

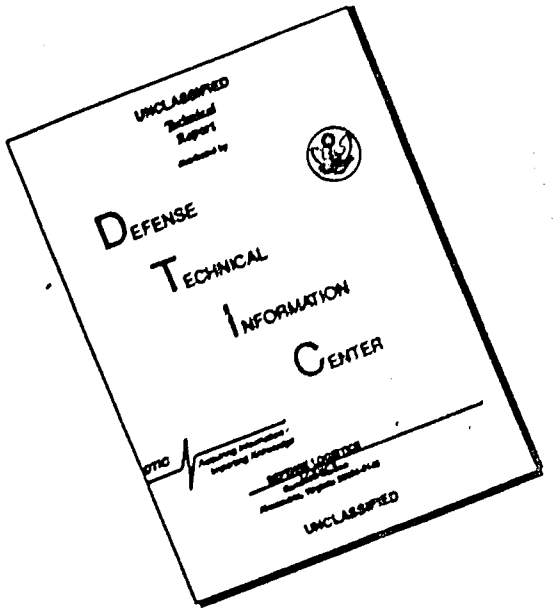
FROM:

**Distribution authorized to U.S. Gov't. agencies and their contractors;
Administrative/Operational Use; APR 1962. Other requests shall be referred to Office of Naval Research, One Liberty Center, 875 North Randolph Street, Arlington, VA 22203-1995. Document partially illegible.**

AUTHORITY

onr ltr, 19 may 1970

DISCLAIMER NOTICE



THIS DOCUMENT IS BEST QUALITY AVAILABLE. THE COPY FURNISHED TO DTIC CONTAINED A SIGNIFICANT NUMBER OF PAGES WHICH DO NOT REPRODUCE LEGIBLY.

SEL-62-058

Linear Estimation of Sampled Stochastic Processes with Random Parameters

by
H. E. Rauch

April 1962

Technical Report No. 2108-1

Prepared under Office of Naval Research Contract
Nonr 225(24), NR 373 360
Jointly supported by the U.S. Army Signal Corps, the
U.S. Air Force, and the U.S. Navy (Office of Naval Research)

STANFORD ELECTRONICS LABORATORIES
STANFORD UNIVERSITY · STANFORD, CALIFORNIA

LINEAR ESTIMATION OF SAMPLED STOCHASTIC PROCESSES WITH RANDOM PARAMETERS

By

H. E. Rauch

April 1962

Technical Report No. 2108-1
Prepared under
Office of Naval Research Contract Nonr-225(24), NR 373 360

Stanford Electronics Laboratories
Stanford University Stanford, California

ABSTRACT

In this investigation the general solution is derived for the problem of the optimum linear estimation of a sampled stochastic process, when the transition and output matrices of the model of the process are random parameters that are independent from one sample point to the next with known mean and covariance. The resulting estimate is optimum in the sense that it minimizes the trace of the covariance matrix of the error (a generalized mean-squared-error criterion).

The notation used in the following discussion is based on the state-transition approach to linear estimation developed by Kalman. In this approach the stochastic process is represented as the output of a linear (possibly time-varying) dynamic system with an independent random input.

For current estimation and prediction of the state vector, the optimum estimate is implemented by a linear dynamic filter with a matrix-valued gain the only undetermined coefficient. This matrix-valued gain, as well as the covariance matrix of the error in the optimum estimate, is determined iteratively for each sample point from a nonlinear difference equation involving the covariance of the error at the previous sample point.

The configuration of the solution for linear interpolation with delay is a linear dynamic filter similar to the one used for prediction. For each sample period the estimate is delayed, an additional weighting matrix and delay element must be added to the filter.

All of these results are derived from the sampled version of the Wiener-Hopf equation, and they apply without modification to stationary and nonstationary statistics and to growing-memory and infinite-memory filters.

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

LIST OF ILLUSTRATIONS

LIST OF ILLUSTRATIONS (Continued)

LIST OF TABLES

ACKNOWLEDGMENT

Appreciation is expressed for the guidance and assistance of Dr. Gene F. Franklin under whose direction this research was developed into a doctoral dissertation. Acknowledgment is also made to my wife, Margi, who has cheerfully lived under dissertation conditions for our entire marriage.

I. INTRODUCTION

A. STATEMENT OF THE PROBLEM

This investigation concerns the linear estimation of a sampled stochastic process when certain parameters of the process must be treated as random variables that are independent from one sample point to the next. These random parameters may be due to multiplicative noise that is corrupting the observed samples, to random variations in the sample period, or to other uncertainties in the a priori knowledge of the process. The stochastic process is represented (for average statistical properties up to second order) as the output of a linear dynamic system excited by independent gaussian processes. This model of a stochastic process is very general, and in particular it includes the important special case of stationary statistics and rational power spectra as well as a large class of nonstationary processes. Both the stochastic process and the random parameters may be stationary or nonstationary, and the linear estimation includes the case in which the number of observed samples is growing.

By way of example, consider the following problem. A satellite is telemetering data to a distant ground station. The original data are continuous, but they must be sampled before they are transmitted. Noise in the electronics or random fading in the transmission characteristics of the atmosphere can introduce multiplicative noise. The time between successive sample points may vary in a random manner because of jitter or missed samples in periodic sampling resulting from imperfections in the equipment, jamming, or natural interference. On the other hand, the data may be transmitted at random intervals intentionally because of the random character of the quantity being measured or in order to counteract jamming. In other words, the ground station may be operating on randomly sampled data with multiplicative noise. From these observed samples it is desired to obtain a continuous estimate of the original data or to predict the value of the data at some future time.

The output of a communication system such as the one described here can be represented as a stochastic process. A realistic model of the

stochastic process must include all the sources of random variation. In this investigation the model of the stochastic process is based on a combination of the Bode-Shannon representation of a random process and the "state-transition" method of analysis of dynamic systems introduced by Kalman [Ref. 1]. The output matrix and the transition matrix of the sampled model are matrix-valued random parameters because of multiplicative noise and random variations in the sample period. The random parameters have a known probability distribution (not necessarily stationary) independent from one sample period to the next.

The optimum estimate of the state variables is the linear estimate which minimizes the trace of the covariance matrix of the error. This is a natural extension of the linear least-squares estimate discussed extensively in the literature [Ref. 2]. If the stochastic process is actually generated by gaussian random excitations of a linear dynamic system, and if the probability distribution of the random parameters is unimodal and symmetric about the expected value of the parameters, then the optimum linear estimate gives the conditional expectation of the desired state variables. Sherman [Ref. 3] has shown that this conditional expectation minimizes the expected value of a large class of loss functions. Gunckel [Ref. 4] has proved that, when the state variable is not known exactly, its conditional expectation can be used in the solution to the general problem of control with a quadratic loss function. Therefore, one is led to the conclusion that the conditional expectation is the best estimate in the general control problem as well as in the estimation problem. On the other hand, the optimum linear estimate requires only the average statistics of the process up to second order, and if only the mean and the covariance of the process are known, then it is the best estimate that can be made with this information.

In addition to estimating the current value of the state variables, it may be necessary to predict the value of the state variables at some time in the future, or it may be advantageous to interpolate some past values of the state vector from more recently observed random variables. The interpolation with delay should reduce the trace of the covariance of the error because more random variables have been observed during the delay.

B. PREVIOUS WORK

The early work of Wiener [Ref. 5] showed that, for a continuous stochastic process, the problem of linear estimation leads to the Wiener-Hopf integral equation. For the practically important case of stationary statistics and rational power spectra, he demonstrated that the solution to the integral equation could be obtained by spectral factorization. Under the same conditions of stationary statistics and rational power spectra, Franklin [Ref. 6] solved the problem of linear estimation using periodically sampled data, and Amara [Ref. 7] generalized this work to include multivariable systems.

Concurrently, there has been considerable interest in the literature on the analysis and stability of systems with random parameters [Ref. 8, 9], although not much work has been concerned with the design of filters for these systems. Kalman [Ref. 10] considered the optimum control of a linear system that is randomly sampled using a quadratic error criterion and a step input. Gunckel [Ref. 4] has extended more recent work of Kalman [Refs. 11, 12] to provide a general solution to the problem of the control of a linear system with random parameters. In particular, he shows that, if it is desired to minimize the expected value of a quadratic loss function, the conditional expectation of the state variables should be used in the optimum control procedure. Gunckel's work separates effectively the problem of estimation of the state variables from the problem of control of the state variables. His results are especially important because under the conditions discussed in the previous section, the optimum linear estimate derived in the present investigation is the conditional expectation of the state variable and therefore can be used in the optimum control procedure.

For a randomly sampled, stationary stochastic process, Bergen [Ref. 13] has determined the spectral density from a convolution integral that can be evaluated in some cases by the method of residues. A synthesis procedure to determine the best linear time-invariant continuous filter for these cases is based upon the standard Wiener factorization of the sampled power spectrum [Ref. 14].

Buetler [Refs. 15, 16] has generalized the Wiener theory to include

stochastic processes with random parameters for continuous stationary systems. He obtained practical solutions for two cases--the optimum linear filter when the prediction time or lag time is a random parameter, and the optimum linear estimate when the system gain is a random parameter (multiplicative noise). Shaw [Ref. 17] has considered the problem of dual-mode filtering for a continuous stationary stochastic process when the instantaneous model of the process varies randomly between two possible models.

Kalman [Ref. 1] has formulated the whole problem of linear estimation from sampled data in matrix notation in terms of state-transition. The problem is approached from the point of view of conditional expectations rather than from the sampled version of the Wiener-Hopf equation. Kalman and Bucy [Ref. 18] have extended this formulation to the linear estimation of a continuous stochastic process when white noise is added to the measurements. In this extension the results are derived from the continuous Wiener-Hopf integral equation.

The work in this investigation represents a generalization of this approach to the problem of linear estimation in the presence of random parameters.

C. OUTLINE OF NEW RESULTS

The solution to the general problem of the optimum linear estimate of a sampled stochastic process with random parameters is derived in this investigation. Chapter II is a review of the state-transition approach to linear estimation using matrix notation, with appropriate examples of random parameters included.

In Chapter III the optimum filter for the current estimation and prediction of the process described in the previous chapter is derived from the sampled Wiener-Hopf equation; the desired weighting coefficients and the covariance of the error in estimation are determined iteratively for each sample point from the covariance of the previous estimate.

These results are extended in Chapter IV to the problem of optimum interpolation with delay (at time t_n the optimum estimate is desired of the state vector at time t where $t < t_n$). Chapter IV presents

the first thorough investigation in the literature of optimum interpolation with delay for a nonstationary stochastic process.

Chapter V applies the ideas developed in this investigation to the estimation of a stationary stochastic process with a random sample period, and for a simple example the optimum filter is compared with the best linear time-invariant filter.

For current estimation and prediction, the optimum estimate is implemented by a linear dynamic filter. The only undetermined coefficient of the filter is the matrix-valued gain, which is determined iteratively for each sample point. When the statistics of the process are stationary, the matrix-valued gain approaches a steady-state value as the number of sample points approaches infinity.

The configuration of the optimum solution for linear interpolation with delay is shown to be a linear dynamic filter similar to the one used for prediction; but, for each sample period the estimate is delayed, an additional weighting matrix must be determined.

In order to relate this investigation to the conventional approach to linear estimation, all of these results are derived from the sampled version of the Wiener-Hopf equation.

II. DESCRIPTION OF THE SYSTEM

A. INTRODUCTION

This chapter will provide an introduction to the state-transition approach to linear estimation. The state of the system is some quantitative information (such as a set of numbers) which is the minimum amount of data necessary to predict the future behavior of the system. When applied to estimation, the state is the data necessary to predict the expected value of the future behavior of the system. The state transition specifies the dynamics of the system--how the state at one instant of time is transferred to the state at a later instant of time.

This paper will consider only those stochastic processes which can be represented (for average statistical properties up to second order) as the output of a linear dynamic system excited by independent gaussian processes. Therefore, at any instant of time, the state of the system can be represented by an r-dimensional vector, and the state transition is an r x r matrix with the properties enumerated in Section II-C. This representation is a very general one, and in particular it includes the important special case of stationary statistics and rational power spectra, as well as a large class of nonstationary processes. In Sections II-B and III-C this representation is presented for two common stationary processes.

The sampled stochastic process is represented in Section II-D as a linear-difference equation with random parameters and random excitation. Examples of random parameters are discussed in Section II-E for the output matrix (caused by multiplicative noise) and the transition matrix (when the sample period is an independent random variable). It should be emphasized that the random parameters are not the only source of noise; both correlated and uncorrelated noise can be included in the vector representing the state of the system.

Finally, in Section II-F are discussed the reasons for restricting the optimum estimate to a linear combination of the observed random variables.

In the notation convention followed here, A^T is the transpose of

A, and E [A] is the expected value of A. When v is an m X 1 vector, then $v^T v$ is the scalar product resulting in a scalar, and vv^T is the vector product resulting in an m X m matrix. The elements of the A matrix are denoted by a_{ij} and the components of the vector by v_i.

B. MODEL OF THE CONTINUOUS PROCESS

The model of the stochastic process is based on the Bode-Shannon representation of a random process and the "state-transition" method of analysis of dynamic systems introduced by Kalman [Ref. 1]. A linear dynamic system can be described by the following ordinary differential equation,

$$\frac{d}{dt} x(t) = F(t)\, x(t) + G(t)\, v(t)$$

$$y(t) = M(t)\, x(t) \qquad (1)$$

where

- $x(t)$ is an r X 1 vector that is the state of the system,
- $v(t)$ is an m X 1 vector that is the input to the system,
- $y(t)$ is a p X 1 vector that is the output of the system,
- $F(t)$ is an r X r matrix representing the dynamics of the system,
- $G(t)$ is an r X m distribution matrix representing the constraints on the input,
- $M(t)$ is a p X r output matrix.

The components x_i of the state vector x are called the state variables, while the components of the output y are linear combinations of these state variables. The matrix F may be nonsingular and represents the dynamics of the system. In stationary systems the matrices F, G, and M are constants.

The matrix block diagram of the system is presented in Fig. 1. The

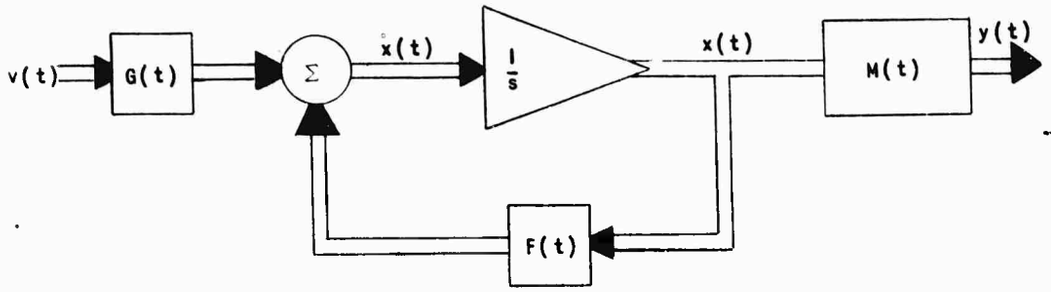


FIG. 1. MODEL OF THE CONTINUOUS STOCHASTIC PROCESS.

thick lines indicate vector-signal flow, and the transfer function $1/s$ actually stands for r integrators such that the output of each is a state variable. The dynamics matrix $F(t)$ indicates how the outputs of the integrators are fed back into the inputs of the integrators. Thus $f_{ij}(t)$ is the coefficient with which the output of the jth integrator is fed back to the input of the ith integrator.

This linear dynamic system represents a stochastic process when the input to the system, $v(t)$, is a random process. In the model used in this investigation, the input $v(t)$ is an $m \times 1$ vector-valued random process with zero mean and with the $m \times m$ covariance matrix

$$E[v(t)\, v^T(s)] = V(t)\, \delta(t - s), \tag{2}$$

where δ is the Dirac delta function. The defining property of the delta function is, for any function $V(t)$ bounded and continuous at s,

$$\delta(t - s) = 0 \qquad t \neq s$$

$$\int_{-\infty}^{\infty} V(t)\delta(t - s)dt = V(s). \tag{3}$$

This definition is satisfactory because only the integral of the covariance in Eq. (2) is ever required. Because the input $v(t)$ is an independent random variable with zero mean,

$$E[v(t)\ x^T(t_0)] = 0 \qquad t > t_0. \tag{4}$$

When the stochastic process is stationary with a rational power spectrum $\mathcal{S}(s)$, the transfer function $H(s)$ of the linear dynamic system representing the process is obtained by factoring the power spectrum

$$\mathcal{S}(s) = H(s)\ H(-s) \tag{5}$$

where $H(s)$ contains all the poles and zeros in the left half plane.

For example, consider the first order stationary Markov process with the power spectrum and autocorrelation function

$$\mathcal{S}_{xx}(s) = \frac{2\beta}{\beta^2 - s^2}$$

$$\mathcal{R}_{xx}(\tau) = e^{-\beta|\tau|} \tag{6}$$

The power spectrum is the Fourier transform of the autocorrelation function [Ref. 19]; therefore, both functions give the same information about the process. In this example the transfer function of the linear system is

$$H(s) = \frac{(2\beta)^{1/2}}{\beta + s} \tag{7}$$

The linear differential equation describing the process is

$$\frac{dx(t)}{dt} = -\beta x(t) + v(t), \tag{8}$$

with the input $v(t)$ an independent random variable with zero mean and covariance

$$E[v(t)\ v(s)] = v_{ll}\ \delta(t - s) = 2\beta\ \delta(t - s). \tag{9}$$

The model of the stationary Markov process is presented in Fig. 2.

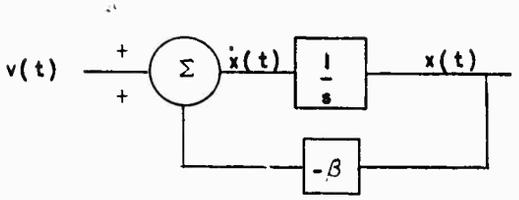


FIG. 2. MODEL OF A STATIONARY MARKOV PROCESS.

The general solution to the ordinary differential equation, Eq. (1), is

$$x(t) = \Phi(t,t_0)x(t_0) + \int_{t_0}^{t} \Phi(t,\tau)G(\tau)v(\tau)d\tau \qquad t \geq t_0, \tag{10}$$

where $\Phi(t,t_0)$ is an r X r matrix called the transition matrix of the system. From Eqs. (4) and (10), it is easily seen that

$$E[x(t)x^T(t_0)] = \Phi(t,t_0)\, E[x(t_0)x^T(t_0)] \qquad t \geq t_0, \tag{11}$$

since $x(t_0)$ is independent of $v(\tau)$ for $\tau \geq t_0$.

C. TRANSITION MATRIX

The pertinent information about the transition matrix is presented in this section. For a more complete treatment see Coddington and Levinson [Ref. 20]. The transition matrix is a nonsingular matrix satisfying the differential equation

$$\frac{d}{dt}\Phi = F(t)\Phi, \tag{12}$$

made unique by the requirement that

$$\Phi(t_0,t_0) = I, \tag{13}$$

where I is the identity matrix.

Two properties of the transition matrix are

$$\Phi(t_3,t_2)\,\Phi(t_2,t_1) = \Phi(t_3,t_1)$$

$$\Phi^{-1}(t_2,t_1) = \Phi(t_1,t_2). \tag{14}$$

If the matrices $F(t)$ and $\int_{t_0}^{t} F(\tau)d\tau$ commute, then the transition matrix can be written

$$\Phi(t,t_0) = \exp\left(\int_{t_0}^{t} F(\tau)d\tau\right) \tag{15}$$

where the definition of

$$\exp(B) = I + \sum_{k=1}^{\infty} \frac{B^k}{k!} \tag{16}$$

For a discussion of the properties of commuting matrices see Gantmacher [Ref. 21]. In particular, the two matrices commute if $F(t)$ is diagonal or if F is a constant. When the F matrix is constant, the transition matrix Φ is stationary, and it can be written

$$\Phi(t_0 + \tau,t_0) = \Phi(\tau) = \exp(F\tau) = e^{F\tau}, \tag{17}$$

where $\exp(F\tau)$ is called the exponential of the matrix $F\tau$.

When the characteristic roots of F are distinct, then the matrix F is similar to a diagonal matrix Λ with diagonal $\lambda_1,\lambda_2,\ldots,\lambda_r$, so that

$$F = D\Lambda D^{-1}, \tag{18}$$

where D is a nonsingular $r \times r$ matrix and the characteristic roots λ_i may be complex. Therefore, the transition matrix is

$$\Phi(\tau) = \exp(F\tau) = D \cdot \exp(\Lambda\tau) \cdot D^{-1}. \tag{19}$$

This is readily seen to be the case because

$$\exp(D\Lambda\tau D^{-1}) = I + \sum_{k=1}^{\infty} \frac{(D\Lambda D^{-1})^k \tau^k}{k!} = D\left\{ I + \sum_{k=1}^{\infty} \frac{(\Lambda\tau)^k}{k!} \right\} D^{-1}. \tag{20}$$

This method of taking the exponential of a matrix is very satisfactory when F is already diagonal or nearly so, but sometimes in actual practice this diagonalization may be difficult to perform, and an alternate method may be more efficient.

An alternative method for taking the exponential of a matrix, given by Friedman [Ref. 22], is based on the following theorem in his chapter on Spectral Theory of Operators:

> "If A is a matrix whose eigenvalues, arranged in order of increasing absolute value are $\lambda_1, \lambda_2, \ldots, \lambda_n$ and if $g(\lambda)$ is an analytic function of λ in a circle around the origin with radius greater than $|\lambda_n|$, then $g(A)$ equals $r(A)$, the polynomial of degree $n - 1$ for which
>
> $$g(\lambda_k) = r(\lambda_k) \qquad k = 1,2,\ldots,n." \tag{21}$$

In particular, this means that if A is an $n \times n$ matrix with distinct characteristic roots $\lambda_1, \lambda_2, \ldots, \lambda_n$, then

$$e^{At} = \alpha_0 I + \sum_{i=1}^{n} \alpha_i (At)^i, \tag{22}$$

where the α_k are evaluated by the set of equations

$$e^{\lambda kt} = \alpha_0 + \sum_{i=1}^{n} \alpha_i (\lambda_k t)^i \qquad k = 1,2,\ldots,n. \tag{23}$$

When

$$\lambda_1 = \lambda_2 = \ldots = \lambda_\nu$$

so that λ_ν is a characteristic root of multiplicity ν for A, then Eq. (21) is modified as follows:

$$g^j(\lambda_\nu) = r^j(\lambda_\nu) \qquad j = 0,1,2,\ldots,\nu-1$$

$$g(\lambda_k) = r(\lambda_k) \qquad k = \nu+1,\ \nu+2,\ldots,n. \tag{25}$$

Thus Eq. (23) is modified to

$$j!\, t^j e^{\lambda_\nu t} = \frac{d^j}{d\lambda_\nu{}^j}\left\{\alpha_0 + \sum_{i=1}^{n} \alpha_i (\lambda_\nu t)^i\right\} \qquad j = 0,1,2,\ldots,\nu-1$$

$$e^{\lambda_k t} = \alpha_0 + \sum_{i=1}^{n} \alpha_i (\lambda_k t)^i \qquad k = \nu+1,\nu+2,\ldots,n. \tag{26}$$

To illustrate this method, the exponential will be taken of the arbitrary 2 X 2 matrix At where

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} . \tag{27}$$

The characteristic roots are determined from

$$|A - \lambda I| = 0$$

$$(a_{11} - \lambda)(a_{22} - \lambda) - a_{12}a_{21} = 0$$

$$\lambda = \tfrac{1}{2}(a_{11} + a_{22}) \pm \tfrac{1}{2}\sqrt{(a_{11} - a_{22})^2 + 4a_{12}a_{21}} \qquad (28)$$

The two characteristic roots λ_1 and λ_2 are distinct if

$$(a_{11} - a_{22})^2 + 4a_{12}a_{21} \neq 0 \qquad (29)$$

For the 2 X 2 matrix Eq. (23) is

$$\begin{aligned} e^{\lambda_1 t} &= \alpha_0 + \alpha_1\lambda_1 t \\ e^{\lambda_2 t} &= \alpha_0 + \alpha_1\lambda_2 t, \end{aligned} \qquad (30)$$

with the solution

$$\begin{aligned} \alpha_0 &= \frac{\lambda_1 e^{\lambda_2 t} - \lambda_2 e^{\lambda_1 t}}{\lambda_1 - \lambda_2} \\ \alpha_1 t &= \frac{e^{\lambda_1 t} - e^{\lambda_2 t}}{\lambda_1 - \lambda_2} \end{aligned} \qquad (31)$$

Thus, when the characteristic roots of the 2 X 2 matrix A are distinct, the exponential of the matrix At is

$$e^{At} = \frac{\lambda_1 e^{\lambda_2 t} - \lambda_2 e^{\lambda_1 t}}{\lambda_1 - \lambda_2}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \frac{e^{\lambda_1 t} - e^{\lambda_2 t}}{\lambda_1 - \lambda_2}\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}. \qquad (32)$$

If the characteristic roots are not distinct, then from Eq. (26), the exponential of the matrix At is

$$e^{At} = e^{\lambda_1 t}\begin{bmatrix} a_{11} + 1 - \lambda_1 t & a_{12} \\ a_{21} & a_{22} + 1 - \lambda_1 t \end{bmatrix}. \qquad (33)$$

As an example, the exponential in Eq. (32) will be used to calculate the transition matrix of a stationary stochastic process with two state variables. Consider the process with the following power spectrum and autocorrelation

$$\mathcal{S}_{x_1x_1}(s) = \frac{(\beta + s)(\beta - s)\, L}{[(\beta + s)^2 + \gamma^2][(\beta - s)^2 + \gamma^2]}$$

$$\mathcal{R}_{x_1x_1}(\tau) = e^{-\beta|\tau|}\left(\cos \gamma\tau - \frac{\beta\gamma \sin\gamma|\tau|}{2\beta^2 + \gamma^2}\right)$$

$$L = \frac{4\beta(\beta^2 + \gamma^2)}{2\beta^2 + \gamma^2} \tag{34}$$

The transfer function of the dynamic system is

$$H(s) = \frac{(\beta + s)\, L^{1/2}}{(\beta - s)^2 + \gamma^2} \tag{35}$$

The model of this stochastic process is presented in Fig. 3, where the two state variables are $x_1(t)$, the observed random variable and the output of one integrator, and $x_2(t)$, the output of the other integrator. The two linear differential equations describing the system are

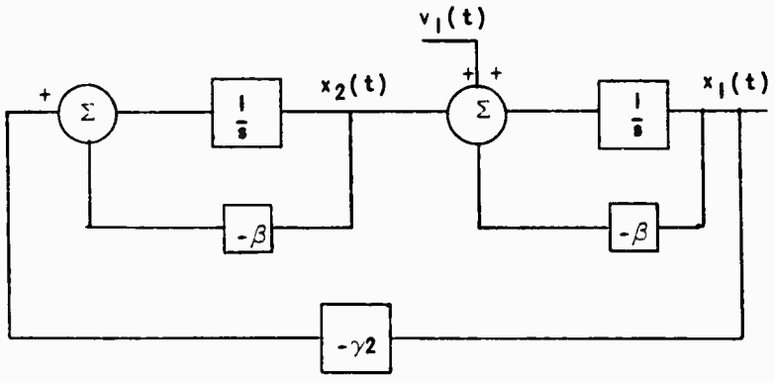


FIG. 3. MODEL OF A PROCESS WITH AN EXPONENTIAL COSINE AUTOCORRELATION.

$$\frac{dx_1(t)}{dt} = -\beta x_1(t) + x_2(t) + v_1(t)$$

and

$$\frac{dx_2(t)}{dt} = -\gamma^2 x_1(t) - \beta x_2(t). \tag{36}$$

The input $v_1(t)$ has zero mean and covariance

$$E[v_1(t)v_1(s)] = v_{11}\,\delta(t - s). \tag{37}$$

For this system the 2 X 2 dynamics matrix F is

$$F = \begin{bmatrix} -\beta & 1 \\ -\gamma^2 & -\beta \end{bmatrix} \tag{38}$$

and the characteristic roots are complex with

$$\lambda_1 = -\beta - j\gamma$$

$$\lambda_2 = -\beta + j\gamma \tag{39}$$

where

j is the imaginary unit number.

From Eq. (32) the exponential of the matrix Ft is

$$\Phi(t) = e^{Ft} = e^{-\beta t} \begin{bmatrix} \cos \gamma t & 1/\gamma \sin \gamma t \\ -\gamma \sin \gamma t & \cos \gamma t \end{bmatrix} \tag{40}$$

where the sum of the complex conjugate exponentials has been written in terms of sines and cosines

$$\sin \gamma t = \tfrac{1}{2} j(e^{-j\gamma t} - e^{+j\gamma t})$$

$$\cos \gamma t = \tfrac{1}{2} (e^{-j\gamma t} + e^{+j\gamma t}) \qquad (41)$$

The model of the first-order stationary Markov process in Fig. 2 and the model of the process with exponential cosine autocorrelation in Fig. 3 will be used in the numerical examples calculated in the following chapters.

D. MODEL OF THE SAMPLED PROCESS

In many cases, a linear dynamic system, such as that shown in Fig. 1, is sampled at points in time $t_0, t_1, \ldots t_k, \ldots t_n$ called the sample points, where the subscript k indicates the kth sample. The time between successive samples is called the sample period T_k, which is given by

$$T_k = t_{k+1} - t_k. \qquad (42)$$

It is assumed that all the switches in the sampling operation operate synchronously and that the sampling operation can be represented as the result of modulating an impulse train $\delta_T(t)$ with the output of the system $y(t)$ so that

$$y^*(t) = y(t)\, \delta_T(t), \qquad (43)$$

where the impulse train $\delta_T(t)$ is given by

$$\delta_T(t) = \sum_{k=0}^{\infty} \delta(t - t_k). \qquad (44)$$

When a linear system with an impulse response $h(t)$ follows the sampling operation, then the output of the linear system at time t is

$$z(t) = \sum_{k=0}^{n} h(t - t_k)\, y(t_k) \qquad t_n < t < t_{n+1}. \tag{45}$$

If a clamp or zero-order hold follows the sampling operation, then the output of the hold at time t is $y(t_n)$ with $t_n < t < t_{n+1}$.

When the state of the system is considered only at the sample points, the stochastic process is said to be discrete, and the model under consideration becomes

$$x(t_{n+1}) = \Phi(t_{n+1}, t_n)\, x(t_n) + u(t_n)$$

$$y(t_n) = M(t_n)\, x(t_n), \tag{46}$$

where

t_n is the time of the nth sampling instant,
$x(t_n)$ is an r X 1 vector which is the state of the system,
$u(t_n)$ is an r X 1 vector which is the effect of the random inputs to the system between the n and $n + 1$ sample points,
$y(t_n)$ is a p X 1 vector which is the observed random variable,
$\Phi(t_{n+1}, t_n)$ is the r X r transition matrix,
$M(t_n)$ is the r X p output matrix.

A matrix block diagram of the sampled system is presented in Fig. 4. The element marked DELAY permits the state of the system to change only at the sample points.

For a discrete stochastic process the properties of the sampled excitation $u(t_n)$ may be given directly, but they can also be derived from the properties of the random input to the continuous system $v(t)$. By comparing Eqs. (46) and (10) it is seen that

$$u(t_n) = \int_{t_n}^{t_{n+1}} \Phi(t_{n+1}, \tau)\, G(\tau)\, v(\tau)\, d\tau. \tag{47}$$

Thus, the sampled excitation $u(t_n)$ also has zero mean, and the

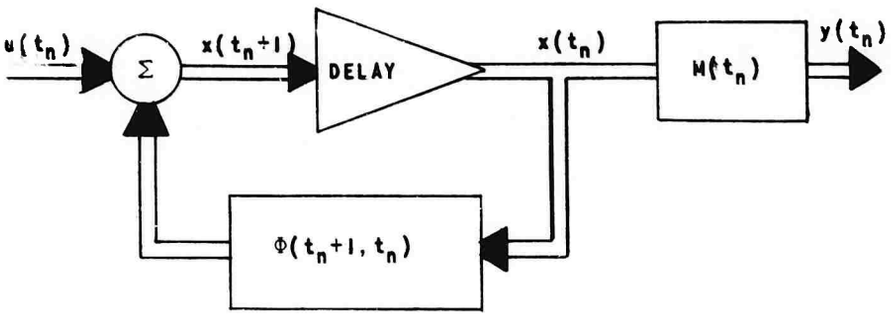


FIG. 4. MODEL OF THE SAMPLED STOCHASTIC PROCESS.

delta-function property of the covariance of the random input $v(t)$ in Eq. (2) means that the covariance matrix of the excitation is

$$E[u(t_n)u^T(t_n)] = E\left[\int_{t_n}^{t_{n+1}} \Phi(t_{n+1},\tau)G(\tau)v(\tau)d\tau \int_{t_n}^{t_{n+1}} v^T(\sigma)G^T(\sigma)\Phi^T(t_{n+1},\sigma)d\sigma\right]$$

$$= \int_{t_n}^{t_{n+1}} \Phi(t_{n+1},\tau)G(\tau)V(t)G^T(\tau)\Phi^T(t_{n+1},\tau)d\tau. \qquad (48)$$

The sampled excitation is an independent random variable, so from Eq. (4)

$$E[u(t_n)x^T(t)] = 0 \qquad t_n > t. \qquad (49)$$

In this investigation, the output matrix M and the transition matrix Φ of the model in Fig. 4 are not known exactly, but they are considered to be matrix-valued random parameters of the model with known probability distribution independent from one sample point to the next. The matrices M and Φ will be random parameters if, for instance, the observed random variable has been corrupted by multiplicative noise or if the sample period is a random variable.

More precisely, it is assumed that a known cumulative distribution function $F_M[C,n]$ is defined over the components of $M(t_n)$, and a different known cumulative distribution function $F_\Phi[D,n]$ is defined over the components of $\Phi(t_{n+1},t_n)$. When $PR[A]$ is defined as the probability that an event A occurs, and $B \leq C$, where B and C are matrices, is defined to imply that $b_{ij} \leq c_{ij}$ for all i and j, then

$$PR[M(t_n) \leq C] = F_M[C,n]$$

$$PR[\Phi(t_{n+1},t_n) \leq D] = F_\Phi[D,n], \tag{50}$$

where $F_M[C,n]$ and $F_\Phi[D,n]$ are the known cumulative distribution functions. These distribution functions are not necessarily stationary, but the random parameters must be statistically independent from one sampling instant to the next so that

$$PR[M(t_m) \leq A;\ \Phi(t_{m+1},\ t_m) \leq B;\ M(t_n) \leq C;\ \Phi(t_{n+1},t_n) \leq D]$$

$$= F_M[A,m]\ F_\Phi[B,m]\ F_M[C,n]\ F_\Phi[D,n]. \tag{51}$$

Finally, it is assumed that $E[M(t_n)\ M^T(t_n)]$, and $E[\Phi(t_{n+1},t_n)\ \Phi^T(t_{n+1},t_n)]$ exist for all n.

The expected values of the two matrices M and Φ will always exist, also, by the previous assumption, and this expected value will be denoted by a superscript bar,

$$E[M(t_n)] = \bar{M}(t_n)$$

$$E[\Phi(t_{n+1},t_n)] = \bar{\Phi}(t_{n+1},t_n). \tag{52}$$

The difference between the actual value of the matrix and the expected value will be denoted by a superscript tilde, so

$$M(t_n) = \bar{M}(t_n) + \tilde{M}(t_n)$$

$$\Phi(t_{n+1},t_n) = \bar{\Phi}(t_{n+1},t_n) + \tilde{\Phi}(t_{n+1},t_n). \tag{53}$$

In this investigation it is not necessary to know the probability distribution of the random parameters. The only information which is needed is the mean of the parameters, $\bar{M}(t_n)$ and $\bar{\Phi}(t_{n+1},t_n)$, and the covariances

$$E[\tilde{M}(t_n)x(t_n)x^T(t_n)\tilde{M}(t_n)]$$

and

$$E[\tilde{\Phi}(t_{n+1},t_n)x(t_n)x^T(t_n)\tilde{\Phi}(t_{n+1},t_n)].$$

E. EXAMPLES OF RANDOM PARAMETERS

The probability distribution of the random parameters M and Φ in the sampled model in Fig. 4 may have any form, but there are some common probability distributions which can be used to approximate situations occurring in actual practice. In this section the mean and the covariance of the random parameters will be determined for five of these situations; two cases of multiplicative noise (amplifier noise and exponential noise), and three cases of randomness in the sample periods (periodic sampling with jitter, periodic sampling with independent misses, and purely random or "Poisson" sampling). For a complete discussion of the probability laws used to approximate these situations see Parzen [Ref. 23].

In the first two cases a single state-variable is corrupted by multiplicative noise and the results can be extended to the multivariable case. Let ϵ be an independent random variable with normal distribution. The cumulative distribution function is

$$F_\epsilon(\rho) = \int_{-\infty}^{\rho} f_\epsilon(\nu)d\nu = \frac{1}{\sqrt{2\pi}\,\sigma}\int_{-\infty}^{\rho} e^{-\frac{\nu^2}{2\sigma^2}}\,d\nu \tag{54}$$

where $f_\epsilon(\nu)$ is the probability-density function. The random variable ϵ has mean zero and variance σ^2 and

$$E\,[e^{\lambda\epsilon}] = e^{\sigma^2\lambda^2/2} \tag{55}$$

One source of multiplicative noise occurs when the state variable passes through an amplifier with noisy gain. Then

$$y_1(t_n) = A(1 + \epsilon)\; x_1(t_n),$$

where A is a constant. The mean and covariance are

$$E[M(t_n)] = E\,[A(1 + \epsilon)] = A$$

$$E[\tilde{M}(t_n)x(t_n)x^T(t_n)\tilde{M}^T(t_n)] = A^2\,E[\epsilon^2]\,E[x_1^2(t_n)]$$

$$= A^2\sigma^2 E[x_1^2(t_n)] \tag{56}$$

Another source of multiplicative noise is the result of taking the exponential of a state variable with added white noise. In a communication system in which the range of the state variable is several orders of magnitude (such as the reflected pulse in a radar system where the energy is inversely proportional to the fourth power of the distance), it may be desirable to transmit the logarithm of the state variable and then convert this at the receiver to the estimate of the original state variable. In that case,

$$\log y_1(t_n) = \log x_1(t_n) + \epsilon$$

$$y_1(t_n) = e^{\epsilon}\; x_1(t_n) \tag{57}$$

and the mean and covariance are

$$E[M(t_n)] = E[e^{\epsilon}] = e^{\sigma^2/2}$$

$$E[\tilde{M}(t_n)x(t_n)x^T(t_n)\tilde{M}^T(t_n)] = \left\{ E[e^{2\epsilon}] - (E[e^{\epsilon}])^2 \right\} E[x_1^2(t_n)]$$

$$= e^{\sigma^2}(e^{\sigma^2} - 1)\, E[x_1^2(t_n)] \qquad (58)$$

In the remaining three cases, the transition matrix (t_{n+1},t_n) is the random parameter because the sample period T_n is an independent random variable. It will be assumed that the dynamics matrix F is constant or can be satisfactorily approximated by a constant over the sample period T_n, but the constant and the probability distribution of T_n can change from one sample point to the next, so the sample period is not necessarily stationary. The transition matrix (t_{n+1},t_n) is an $r \times r$ matrix-valued random variable given by

$$(t_{n+1},t_n) = e^{FT_n}. \qquad (59)$$

The exponential of the matrix $F\,T_n$ is composed of terms of the form

$$e^{\lambda_i T_n},\ T_n e^{\lambda_i T_n},\ T_n^2 e^{\lambda_i T_n},\ \ldots,\ T_n^{\nu} e^{\lambda_i T_n}$$

where λ_i is a characteristic root of F of multiplicity ν. The powers of T_n occur when λ_i is a multiple characteristic root, as in the 2×2 matrix in Eq. (33).

In order to determine the expected value and the covariance of the transition matrix, it is necessary only to know

$$E[e^{\lambda T_n}]$$

for all values of λ both real and complex. If that expected value is known, the other expected values can be calculated by the relationship

$$E[T_n^{\nu} e^{\lambda_i T_n}] = \frac{d^{\nu}}{d\lambda^{\nu}} E[e^{\lambda T_n}] \Bigg|_{\lambda = \lambda_i} \qquad (60)$$

The three situations in which the sample period can be considered as a random variable, and their associated probability distributions, are as follows: periodic sampling with jitter (normal distribution), periodic sampling with independent misses (geometric), and purely random or Poisson sampling (exponential). The pertinent information concerning these three probability laws is presented in Table 1, and their characteristics are discussed in the remainder of this section.

TABLE 1. THREE PROBABILITY LAWS FOR RANDOM SAMPLE PERIODS.

Probability Law	Normal	Geometric	Exponential
Sample Pattern	Periodic with Jitter	Periodic with Independent Misses	Purely Random or Poisson
Parameters of Probability Distribution	$0 < 3\sigma < \bar{T} < \infty$	$0 < T < \infty$ $0 \leq p = 1-q \leq 1$	$0 < \mu < \infty$
Probability Density $f(\tau)$	$\frac{\exp[-(\tau - \bar{T})^2/2\sigma^2]}{\sqrt{2\pi}\,\sigma}$	pq^{k-1} for $\tau = k\bar{T}$ $k = 0, 1, 2, \ldots$ 0 Otherwise	$\mu e^{-\mu\tau}$ for $\tau \geq 0$ 0 Otherwise
Mean $E[\tau]$	$\bar{T}$	T/p	$1/\mu$
Variance $E[\tau^2] - E^2[\tau]$	σ^2	qT^2/p^2	$1/\mu^2$
$E[e^{\lambda\tau}]$	$e^{\bar{T}\lambda} e^{\frac{1}{2}\sigma^2\lambda^2}$	$pe^{\lambda T}/(1 - qe^{\lambda T})$	$\mu/(\mu-\lambda)$

Jitter occurs when the sample period is nearly constant, but varies slightly from period to period. In certain anti-jamming applications the sample period is varied randomly in this manner, or the variation may be unintentional due to imperfections in the equipment. An approximation to the effect of jitter is given by the normal distribution of sample periods,

$$F_{T_n}[\tau] = \bar{T} + \frac{1}{\sqrt{2\pi}\,\sigma} \int_{-\infty}^{\tau} e^{-\nu^2/2\sigma^2}\, d\nu \qquad 0 < 3\sigma < \bar{T} < \infty \qquad (61)$$

or, alternatively,

$$T_n = \bar{T} + \epsilon, \tag{62}$$

where $\bar{T}$ is the mean of the sample period, and ϵ is the normal random variable with mean zero and variance defined in Eq. (54). In the normal distribution in Eq. (61), there is a finite probability that the sample period will be negative; therefore, one might argue that a better approximation would be to truncate the distribution so that it is zero outside the region

$$0 < T_n < 2\bar{T}.$$

This is a valid argument, but in Eq. (61) the standard deviation is constrained to be less than one-third of the mean of the sample period $\bar{T}$ and, under this constraint, the area of the normal distribution that must be truncated is less than 2.6×10^{-3}; therefore, the original approximation should be valid.

Periodic sampling with independent misses results when the intended sample period is constant but at each sampling instant there is a fixed probability q that the random variable will not be observed. This situation may arise when the receiver rejects the signal unless the signal-to-noise ratio is above some threshold value. The probability distribution of the actual sample period is a geometric distribution given by

$$f_{T_n}(kT) = pq^{(k-1)} \qquad k = 1,2,\ldots \qquad 0 \leq p = 1 - q \leq 1$$

$$= 0 \qquad \text{otherwise} \tag{63}$$

When the number of sample points in a given interval of time has a Poisson distribution, the samples are being received at purely random points in time. The samples may be transmitted in this way intentionally to avoid jamming, or because of the random character of the quantity being transmitted. For this case the sample period has an exponential distribution

$$F_{T_n}(\tau) = \mu e^{-\mu T} \qquad T \geq 0 \qquad 0 \leq \mu < \infty \qquad (64)$$
$$= 0 \qquad \text{otherwise}$$

F. THE OPTIMUM ESTIMATE

The purpose of the design procedure presented in this paper is to estimate the state vector of a stochastic process from information obtained from the a priori knowledge of the process and the observed random variables. The difference between the actual value of the state vector and the estimate is the error, which is written

$$\tilde{x}(t_n + \alpha/t_n) = x(t_n + \alpha) - \hat{x}(t_n + \alpha/t_n), \qquad (65)$$

where

$\tilde{x}(t_n + \alpha/t_n)$ is the $(r \times 1)$ vector that is the error in the estimate of the value of the state vector at time $t_n + \alpha$, given the observed random variables $y(t_n), y(t_{n-1}), \ldots, y(t_0)$;

$x(t_n + \alpha)$ is the actual value of the state vector,

$\hat{x}(t_n + \alpha/t_n)$ is the estimate of the value of the state vector at time $t_n + \alpha$, given the observed random variables $y(t_n), y(t_{n-1}), \ldots, y(t_0)$.

The best estimate is the one which minimizes some function of the error. For α positive the operation of estimation is called prediction; and for α negative, it is called interpolation or smoothing.

In this paper the estimate will be confined to linear combinations of the observed random variable, and the optimum estimate will be the linear estimate that minimizes the trace of the covariance matrix of the error. The covariance of the error is an $r \times r$ matrix defined as

$$P(t_n + \alpha/t_n) = E[\tilde{x}(t_n + \alpha/t_n)\tilde{x}^T(t_n + \alpha/t_n)]. \qquad (66)$$

The trace is the sum of the diagonal terms of the matrix. The trace of the covariance is the expected value of the sum of the squared error

in the estimation of the individual state variables, so the optimum estimate is a natural extension of the "linear-least-squares" estimate that is discussed extensively in the literature.

The question naturally arises as to which is better--a linear estimate or a nonlinear estimate. Doob [Ref. 24] proves rigorously that the estimate of a quantity that minimizes the expected value of the squared error is the conditional expectation of the quantity with respect to the given information. For a linear dynamic system excited by random variables with normal distribution, the conditional expectation of the state vector is indeed a linear combination of the observed random variables and the linear-least-squares estimate will be better than any nonlinear estimate in minimizing the expected value of the squared error.

Sherman [Ref. 3] has proved that the conditional expectation can minimize the expected value of a still larger class of loss functions. In particular, when e is the random variable representing the error, and the loss function $L(e)$ is a positive function, symmetric and non-decreasing about zero, so that

$$0 \leq L(e) = L(-e)$$

$$L(e_1) \leq L(e_2) \quad \text{when} \quad 0 \leq e_1 \leq e_2, \tag{67}$$

and if the conditional distribution of the quantity being estimated is symmetric and convex about its conditional expectation, then the conditional expectation minimizes the expected value of the class of loss functions included in Eq. (67). This class of loss functions includes:

$$L_1(e) = e^2$$

$$L_2(e) = e^4$$

$$L_3(e) = |e|$$

$$L_4(e) = 1 \qquad -b \leq e \leq b$$
$$= 0 \qquad \text{otherwise.} \tag{68}$$

In this investigation, Sherman's results mean that if the stochastic process is the output of a linear dynamic system with a gaussian input, and if the random parameters have a symmetric convex probability distribution, then the optimum linear estimate will be better than any other estimate, linear or nonlinear, in minimizing the expected value of the class of loss functions in Eq. (67).

The optimum linear estimate has the additional advantage that only the first- and second-order statistical averages are required. In practical situations it is often very difficult to measure more than this, and any nonlinear estimate would require more information than just the first- and second-order statistical averages to improve over the optimum linear estimate. For these reasons, in this investigation the optimum estimate will be the linear estimate that minimizes the trace of the covariance matrix of the error.

III. OPTIMUM FILTERING AND PREDICTION OF RANDOM-PARAMETER PROCESSES

A. INTRODUCTION

In Chapter II the sampled stochastic process was represented as a matrix-difference equation with random parameters. The optimum estimate was defined as the linear estimate that minimized the trace of the covariance matrix of the error. In this chapter it will be proved that the optimum estimate of the current value of the state vector at the current sample point is a linear combination of the optimum estimate at the preceding sample point and the random variable observed at the current sample point. The optimum estimate is implemented by a linear filter with the only undetermined coefficient being a matrix-valued gain. This matrix-valued gain as well as the covariance of the error in the optimum estimate is determined by a nonlinear difference equation that can be solved iteratively for each sample point. The optimum current estimate becomes the optimum prediction of a future value of the state vector when it is matrix-multiplied by the expected value of the transition matrix.

These results are obtained from the solution to the matrix-valued sampled Wiener-Hopf equation, derived in Sec. III-C, but first the Markov property of the optimum linear estimate will be discussed.

B. MARKOV PROPERTY OF THE OPTIMUM ESTIMATE

For a Markov process, the probability functions relating to the future depend on the present state, but not on the manner in which the present state has emerged from the past [Ref. 25]. The model of the stochastic process developed in Chapter III has the Markov property, and in Sec. III-D it will be proved that the optimum linear estimate has the Markov property because the optimum estimate of the current value of the state vector at the current sample point is a linear combination of the optimum estimate at the preceding sample point and the observed random variable at the current sample point. In other words,

$$\hat{x}(t_n/t_n) = \bar{\Phi}^*(t_n,t_{n-1})\hat{x}(t_{n-1}/t_n) + K(t_n)y(t_n) \tag{69}$$

where $\bar{\Phi}^*(t_n,t_{n-1})$ and $K(t_n)$ are r x r and r x p linear weighting matrices. (Note that $\bar{\Phi}^*(t_n,t_{n-1})$ is an arbitrary weighting coefficient different from the expected value of the transition matrix $\bar{\Phi}(t_n,t_{n-1})$, which does not have the asterisk.) This section will explore the reasoning which might lead one to the hypothesis of Eq. (69).

First, assume that at time t_{n-1} the exact value of the state vector is known; therefore,

$$\hat{x}(t_{n-1}/t_{n-1}) = x(t_{n-1}). \tag{70}$$

If the transition matrix in the sample period is equal to its expected value $\bar{\Phi}(t_n,t_{n-1})$ and if there are no random inputs to the system during the sample period, then the next state will be

$$x(t_n) = \bar{\Phi}(t_n,t_{n-1})x(t_{n-1}), \tag{71}$$

and the best estimate

$$\hat{x}(t_n) = \bar{\Phi}(t_n,t_{n-1})\hat{x}(t_{n-1}) = x(t_n) \tag{72}$$

will be the exact value of the state vector. If the output matrix is also equal to its expected value $\bar{M}(t_n)$, the observed random variable will be

$$y(t_n) = \bar{M}(t_n)x(t_n) = \bar{M}(t_n)\,\bar{\Phi}(t_n,t_{n-1})\hat{x}(t_{n-1}). \tag{73}$$

and no new information about the process will be gained by looking at it, because its exact value is already known. This is not a very interesting problem, and, naturally, in actuality things are not so simple. There will be errors in the estimate of the state at the previous sample instant, $\tilde{x}(t_{n-1}/t_{n-1})$; there are random excitations to the system $u(t_{n-1})$, and variations in the transition and output matrices $\tilde{\Phi}(t_n,t_{n-1})$ and $\tilde{M}(t_n)$.

The only way one discovers these random variations is by comparing the observed random variable $y(t_n)$ with its expected value which is given by Eq. (73). All the new information about the process comes from the difference between $y(t_n)$ and its expected value

$$y(t_n) - \overline{M}(t_n)\ \overline{\Phi}\ (t_n,t_{n-1})\hat{x}(t_{n-1}).$$

Because the estimate is restricted to linear operations, it is reasonable to weight the new information with some linear matrix-valued coefficient $K(t_n)$ and add it to the previous estimate in Eq. (72), so that

$$\hat{x}(t_n/t_n) = \overline{\Phi}\ (t_n/t_{n-1})\hat{x}(t_{n-1}/t_{n-1})$$

$$+ K(t_n)[y(t_n) - \overline{M}(t_n)\ \overline{\Phi}\ (t_n/t_{n-1})\hat{x}(t_{n-1}/t_{n-1})]. \qquad (74)$$

Combining terms in Eq. (74) gives

$$\hat{x}(t_n/t_n) = [I - K(t_n)\overline{M}(t_n)]\ \overline{\Phi}\ (t_n,t_{n-1})\hat{x}(t_{n-1}/t_{n-1}) + K(t_n)y(t_n)$$

where I is the identity matrix and

$$[I - K(t_n)\overline{M}(t_n)]\ \overline{\Phi}\ (t_n,t_{n-1}) = \overline{\Phi}^{*}(t_n,t_{n-1}), \qquad (75)$$

the arbitrary $r \times r$ weighting matrix in Eq. (69).

This reasoning started with the assumption, in Eq. (70), that at time t_{n-1} the exact value of the state vector was known, but all the ideas still apply if only the expected value of the state vector is known. The key idea in this discussion can be summarized as follows: When the stochastic process can be represented as an r^{th}-order linear Markov process, the optimum linear estimate of the state vector of the process can also be represented as an r^{th}-order linear Markov process.

This approach to the linear estimate of a stochastic process was first developed fully by Kalman [Ref. 1] for a process with deterministic parameters. He used conditional expectations and the projection theorem, and considered in detail the linear estimate

$$\hat{x}(t_{n+1}/t_n) = \Phi^*(t_{n+1},t_n)\hat{x}(t_n/t_{n-1}) + \Delta^*(t_n)y(t_n), \tag{76}$$

where

$$\Phi^*(t_{n+1},t_n) = \Phi(t_{n+1},t_n) - \Delta^*(t_n)M(t_n). \tag{77}$$

He proved that any other estimate $\hat{x}(t_n + \alpha/t_n)$ for α positive could be derived from the one in Eq. (76) by the relationship

$$\hat{x}(t_n+\alpha/t_n) = \Phi(t_n+\alpha,t_n)[\Phi(t_{n+1},t_n)]^{-1}\hat{x}(t_{n+1}/t_n), \tag{78}$$

where A^{-1} represents the inverse of A.

The quantity that in this investigation, is analogous to the quantity in brackets in Eq. (78) is $\Phi(t_{n+1},t_n)$, which may be singular and does not always have a unique inverse. Therefore, for the stochastic process with random parameters this approach must be modified to the linear estimate in Eq. (69).

C. THE SAMPLED WIENER-HOPF EQUATION

Any linear estimate of the state vector will be a linear combination of the observed random variables so that

$$\hat{x}(t_n+\alpha/t_n) = \sum_{\nu=0}^{n} A(t_n+\alpha,t_\nu)y(t_\nu), \tag{79}$$

where $\hat{x}(t_n+\alpha/t_n)$ is an r x 1 vector that is the estimate of the state vector $x(t_n+\alpha)$, given the observed random variables $y(t_n), y(t_{n-1}), \ldots, y(t_0)$, and

$A(t_n+\alpha,t_\nu)$ is an r x p matrix which is the νth set of weighting coefficients of the estimate.

The sampled Wiener-Hopf equation is a matrix-valued linear equation that is satisfied by the weighting coefficients $A(t_n+\alpha,t_\nu)$ when the linear estimate is optimum.

The error in the linear estimate in Eq. (79), as defined in Eq. (65), is

$$\tilde{x}(t_n+\alpha/t_n) = x(t_n+\alpha) - \hat{x}(t_n+\alpha/t_n) = x(t_n+\alpha) - \sum_{\nu=0}^{n} A(t_n+\alpha, t_\nu)y(t_\nu) \tag{80}$$

Substituting the error in Eq. (80) into the covariance of the error defined in Eq. (66) gives

$$\begin{aligned} P(t_n+\alpha/t_n) &= E[\tilde{x}(t_n+\alpha/t_n)\tilde{x}^T(t_n+\alpha/t_n)] \\ &= E[x(t_n+\alpha)x^T(t_n+\alpha)] \\ &\quad - \sum_{\nu=0}^{n} A(t_n+\alpha,t_\nu)E[y(t_\nu)x^T(t_n+\alpha)] \\ &\quad - \sum_{\nu=0}^{n} E[x(t_n+\alpha)y^T(t_\nu)]A^T(t_n+\alpha,t_\nu) \\ &\quad + \sum_{\nu=0}^{n}\sum_{\mu=0}^{n} A(t_n+\alpha, t_\nu)\, E[y(t_\nu)y^T(t_\mu)]\, A^T(t_n+\alpha,t_\mu) \end{aligned} \tag{81}$$

The trace of the covariance matrix in Eq. (81) is the sum of the diagonal terms, so that

$$TR[P(t_n+\alpha/t_n)] = \sum_{i=1}^{r} p_{ii}(t_n+\alpha/t_n) = E\left[\sum_{i=1}^{r} \tilde{x}_i^2(t_n+\alpha/t_n)\right] \tag{82}$$

which is the expected value of the sum of the squared error in the estimation of the individual state variables. From Eq. (81) it is seen that all the components of the covariance matrix are quadratic functions of the elements of the weighting coefficients $a_{jk}(t_n+\alpha,t_\nu)$, so the trace in Eq. (82) is minimized when its derivative with respect to each of the elements $a_{jk}(t_n+\alpha,t_\nu)$ is zero,

$$\frac{d\ TR[P(t_n+\alpha/t_n)]}{d\ a_{jk}(t_n+\alpha,t_\nu)} = 0 \qquad \begin{matrix} j = 1,2,\ldots,r \\ k = 1,2,\ldots,p \\ \nu = 0,1,\ldots,n \end{matrix} \tag{83}$$

Because the trace is a quadratic function of the elements, there will be at least one minimum, and the resulting equation for the optimum elements will be linear. After performing the operation indicated in Eq. (83) and combining the linear equations, the result is

$$E[x(t_n+\alpha)y^T(t_\rho)] - \sum_{\nu=0}^{n} A(t_n+\alpha,t_\nu)E[y(t_\nu)y^T(t_\rho)] = 0 \qquad \rho = 0,1,\ldots,n, \tag{84}$$

which is the sampled Wiener-Hopf equation. This result yields a minimum because

$$\frac{d^2\ TR[P(t_n+\alpha/t_n)]}{d\ a_{jk}(t_n+\alpha,t_\nu)^2} = E[y_1^2(t_\nu)] > 0 \tag{85}$$

the second derivative with respect to the elements is positive. The optimum weighting in Eq. (84) not only minimizes the trace, but it also minimizes all the components of the covariance matrix, as can be seen by carrying through the operations in Eq. (83) for the off-diagonal components of the covariance matrix. Substituting the estimate in Eq. (79) for the equivalent expression in Eq. (84) gives

$$E[x(t_n+\alpha)y^T(t_\rho)] - E[\hat{x}(t_n+\alpha/t_n)y^T(t_\rho)] = 0 \qquad \rho = 0,1,\ldots,n \tag{86}$$

as a compact way of writing the sampled Wiener-Hopf equation. When the weighting coefficients are optimum and satisfy Eq. (86), some of the terms in the expression for the covariance matrix in Eq. (81) are identically zero, and the covariance of the error can be written,

$$P(t_n+\alpha/t_n) = E[x(t_n+\alpha)x^T(t_n+\alpha)] - E[\hat{x}(t_n+\alpha/t_n)x^T(t_n+\alpha)] \tag{87}$$

where $\hat{x}(t_n+\alpha/t_n)$ is defined in Eq. (79).

D. DERIVATION OF THE OPTIMUM ESTIMATE

The sampled Wiener-Hopf Eq. (86) must be satisfied by the weighting coefficients of the optimum linear estimate for any stochastic process with any statistics. In this section it will be proved that, when the process can be represented by the model developed in Chapter II, the optimum estimate has the Markov property discussed in Sec. III-B. The assumption is first made that the optimum estimate of the current value of the state vector $\hat{x}(t_n/t_n)$ can be written in the form

$$\hat{x}(t_n/t_n) = [I - K(t_n)\overline{M}(t_n)]\,\overline{\Phi}\,(t_n,t_{n-1})\hat{x}(t_{n-1}/t_{n-1}) + K(t_n)y(t_n), \tag{88}$$

where $\hat{x}(t_{n-1}/t_{n-1})$ is the previous optimum estimate and $K(t_n)$ is an $r \times p$ weighting matrix. Next it is shown that, if the weighting coefficients of the previous estimate $\hat{x}(t_{n-1}/t_{n-1})$ satisfy the sampled Wiener-Hopf Eq. (86), the weighting of the current estimate in Eq. (88) will satisfy Eq. (86) for a particular $r \times p$ weighting matrix $K(t_n)$. Finally, the weighting matrix $K(t_n)$ is derived in terms of a priori known quantities and the covariance of the error of the previous optimum prediction $P(t_n/t_{n-1})$, which can be calculated iteratively at each sample point.

The random inputs to the system have zero mean; therefore, from Eq. (11) for α positive,

$$E[x(t_n+\alpha)y^T(t_\rho)] = E[\Phi(t_n+\alpha,t_n)x(t_n)y^T(t_\rho)] = \overline{\Phi}(t_n+\alpha,t_n)E[x(t_n)y^T(t_\rho)].$$

$$\begin{aligned} &\alpha \geq 0 \\ &\rho \leq n \end{aligned} \tag{89}$$

Substituting Eq. (89) into Eq. (86) for α positive gives

$$\overline{\Phi}(t_n+\alpha,t_n)E[x(t_n)y^T(t_\rho)] - E[\hat{x}(t_n+\alpha/t_n)y^T(t_\rho)] = 0 \quad \begin{aligned} &\alpha \geq 0 \\ &\rho = 0,1,\ldots n \end{aligned} \tag{90}$$

as the condition for optimum prediction. It is obvious from Eq. (90) that, if $\hat{x}(t_n/t_n)$ is the optimum current estimate and hence satisfies that equation for α equal to zero,

$$\hat{x}(t_n+\alpha/t_n) = \bar{\Phi}\,(t_n+\alpha, t_n)\hat{x}(t_n/t_n) \qquad \alpha \geq 0 \tag{91}$$

must satisfy Eq. (90) for α positive. Thus, the optimum prediction is obtained from the optimum current estimate by matrix multiplying the current estimate by the expected value of the transition matrix. Now, the optimum current estimate will be determined.

Assume the previous optimum estimate is

$$\hat{x}(t_{n-1}/t_{n-1}) = \sum_{0}^{n-1} A(t_{n-1}, t_\nu)y(t_\nu), \tag{92}$$

and therefore it satisfies Eq. (93),

$$E[x(t_{n-1})y^T(t_\rho)] - E[\hat{x}(t_{n-1}/t_{n-1})y^T(t_\rho)] = 0 \qquad \rho = 0,1,\ldots,n-1, \tag{93}$$

From Eqs. (91) and (92), the optimum prediction at the previous sample instant of the current value of the state vector is

$$\hat{x}(t_n/t_{n-1}) = \bar{\Phi}\,(t_{n+1}, t_n)\hat{x}(t_{n-1}/t_{n-1}), \tag{94}$$

and from Eq. (87) the covariance of the error of this prediction is

$$P(t_n/t_{n-1}) = E[x(t_n)x^T(t_n)] - \bar{\Phi}(t_n, t_{n-1})E[\hat{x}(t_{n-1}/t_{n-1})x^T(t_n)]. \tag{95}$$

The trial estimate in Eq. (88) can be written

$$\hat{x}(t_n/t_n) = \bar{\Phi}\,(t_n, t_{n-1})\hat{x}(t_{n-1}/t_{n-1}) + K(t_n)[y(t_n) - \bar{M}(t_n)\bar{\Phi}\,(t_n, t_{n-1})\hat{x}(t_{n-1}/t_{n-1})]. \tag{96}$$

For the estimate $\hat{x}(t_n/t_n)$ in Eq. (96) to be optimum, it must satisfy Eq. (97),

$$E[x(t_n)y^T(t_\rho)] - E[\hat{x}(t_n/t_n)y^T(t_\rho)] = 0 \quad . \quad \rho = 0,1,\ldots,n. \tag{97}$$

Substituting the trial estimate in Eq. (96) into Eq. (97) gives

$$E[x(t_n)y^T(t_\rho)] - \bar{\Phi}(t_n,t_{n-1})E[\hat{x}(t_{n-1}/t_{n-1})y^T(t_\rho)]$$
$$- K(t_n)\left\{E[y(t_n)y^T(t_\rho)] - \bar{M}(t_n)\,\bar{\Phi}(t_n,t_{n-1})E[\hat{x}(t_{n-1}/t_{n-1})y^T(t_\rho)]\right\} = 0$$
$$\rho = 0,1,\ldots,n \tag{98}$$

as the condition for optimum estimation.

Equation (98) will be examined first for ρ less than n. In that case,

$$E[x(t_n)y^T(t_\rho)] = \bar{\Phi}(t_n,t_{n-1})E[x(t_{n-1})y^T(t_\rho)]$$
$$E[y(t_n)y^T(t_\rho)] = \bar{M}(t_n)\,\bar{\Phi}(t_n,t_{n-1})E[x(t_{n-1})y^T(t_\rho)] \quad \rho < n \tag{99}$$

Substituting Eq. (99) into Eq. (98) for ρ less than n yields,

$$[I - K(t_n)\bar{M}(t_n)]\,\bar{\Phi}(t_n,t_{n-1})\left\{E[x(t_{n-1})y^T(t_\rho)] - E[\hat{x}(t_{n-1}/t_{n-1})y^T(t_\rho)]\right\} = 0$$
$$\rho = 0,1,\ldots,n-1 \tag{100}$$

Comparison of Eqs. (100) and (93) shows that the quantity in braces in Eq. (100) is the same as Eq. (93), which is identically zero for ρ less than n. Thus, the trial estimate in Eq. (95) satisfies the conditions for optimum estimation in Eq. (98) for ρ less than n. The remaining condition in Eq. (98) for ρ equal to n will be used to determine the r x p weighting matrix $K(t_n)$; this condition is

$$E[x(t_n)y^T(t_n)] - \bar{\Phi}(t_n,t_{n-1})E[\hat{x}(t_{n-1}/t_{n-1})y^T(t_n)]$$

$$- K(t_n)\left\{E[y(t_n)y^T(t_n)] - \bar{M}(t_n)\bar{\Phi}(t_n,t_{n-1})E[\hat{x}(t_{n-1}/t_{n-1})y^T(t_n)]\right\} = 0 \tag{101}$$

From the model of the process in Eq. (46) and the properties of the random parameters in Eq. (53), the observed random variable $y(t_n)$ can be written

$$y(t_n) = M(t_n)x(t_n) = \bar{M}(t_n)x(t_n) + \tilde{M}(t_n)x(t_n), \tag{102}$$

where $\tilde{M}(t_n)$ is an independent random variable with zero mean, so that

$$E[x(t_n)y^T(t_n)] = E[x(t_n)x^T(t_n)]\,\bar{M}^T(t_n)$$

$$E[\hat{x}(t_{n-1}/t_{n-1})y^T(t_n)] = E[\hat{x}(t_{n-1}/t_{n-1})x^T(t_n)]\,\bar{M}^T(t_n)$$

$$E[y(t_n)y^T(t_n)] = \bar{M}(t_n)E[x(t_n)x^T(t_n)]\bar{M}^T(t_n)$$

$$+ E[\tilde{M}(t_n)x(t_n)x^T(t_n)\tilde{M}^T(t_n)]$$

$$= \bar{M}(t_n)E[x(t_n)x^T(t_n)]\bar{M}^T(t_n) + \bar{R}(t_n). \tag{103}$$

In the third equation of Eq. (103), the substitution has been made that

$$\bar{R}(t_n) = E[\tilde{M}(t_n)x(t_n)x^T(t_n)\tilde{M}^T(t_n)]. \tag{104}$$

Substituting Eq. (103) into the remaining condition for optimum estimation in Eq. (101) and combining terms yields

$$[I-K(t_n)\bar{M}(t_n)]\left\{E[x(t_n)x^T(t_n)]-\bar{\Phi}(t_n,t_{n-1})E[\hat{x}(t_{n-1}/t_{n-1})x^T(t_n)]\right\}\bar{M}^T(t_n)$$

$$- K(t_n)\,\bar{R}(t_n) = 0 \tag{105}$$

The quantity in braces in Eq. (105) is equal to Eq. (95) which is the

covariance of the error in the optimum prediction at the previous sample instant. Replacing that quantity by $P(t_n/t_{n-1})$ in Eq. (95) yields

$$P(t_n/t_{n-1})\overline{M}^T(t_n) - K(t_n)\left\{\overline{M}(t_n)P(t_n/t_{n-1})\overline{M}^T(t_n) + \overline{R}(t_n)\right\} = 0, \tag{106}$$

where $\overline{R}(t_n)$ is defined in Eq. (104). When the r x p weighting matrix $K(t_n)$ satisfies Eq. (106), the trial estimate in Eq. (96) is indeed the optimum estimate because it satisfies the sampled Wiener-Hopf Eq. (97). The only quantity in Eq. (106) that is not known a priori is the covariance matrix of the error $P(t_n/t_{n-1})$. The covariance $P(t_{n+1}/t_n)$ can be determined iteratively at each sample point from the covariance $P(t_n/t_{n-1})$ at the previous sample point.

From Eq. (91) the optimum prediction $\hat{x}(t_{n+1}/t_n)$ is

$$\hat{x}(t_{n+1}/t_n) = \overline{\Phi}(t_{n+1},t_n)\hat{x}(t_n/t_n), \tag{107}$$

From Eqs. (87) and (89) the covariance of the error in this prediction is

$$\begin{aligned} P(t_{n+1}/t_n) &= E[x(t_{n+1})x^T(t_{n+1})] - E[\hat{x}(t_{n+1}/t_n)x^T(t_{n+1})] \\ &= E[x(t_{n+1})x^T(t_{n+1})] - \overline{\Phi}(t_{n+1},t_n)E[\hat{x}(t_n/t_n)x^T(t_n)]\,\overline{\Phi}^T(t_{n+1},t_n). \end{aligned} \tag{108}$$

The form of the optimum estimate $\hat{x}(t_n/t_n)$ is given by Eq. (96) and, using Eq. (94), this can be written

$$\begin{aligned} \hat{x}(t_n/t_n) &= \overline{\Phi}(t_n,t_{n-1})\hat{x}(t_{n-1}/t_{n-1}) + K(t_n)[y(t_n) - \overline{M}(t_n)\overline{\Phi}(t_n,t_{n-1})\hat{x}(t_{n-1}/t_{n-1})] \\ &= K(t_n)y(t_n) + [I - K(t_n)\overline{M}(t_n)]\hat{x}(t_n/t_{n-1}). \end{aligned} \tag{109}$$

Substituting Eq. (109) into the covariance in Eq. (108) gives

$$P(t_{n+1}/t_n) = E[x(t_{n+1})x^T(t_{n+1})] - \bar{\Phi}(t_{n+1},t_n)K(t_n)E[y(t_n)x^T(t_n)]\bar{\Phi}^T(t_{n+1},t_n)$$

$$+ \bar{\Phi}(t_{n+1},t_n)[I-K(t_n)\bar{M}(t_n)]E[\hat{x}(t_n/t_{n-1})x^T(t_n)]\ \bar{\Phi}^T(t_{n+1},t_n). \tag{110}$$

From the model of the process in Eq. (46) and the properties of the random parameters in Eq. (53), the state vector $x(t_{n+1})$ can be written

$$x(t_{n+1}) = \Phi(t_{n+1},t_n)x(t_n) + u(t_n)$$

$$= \bar{\Phi}(t_{n+1},t_n)x(t_n) + \tilde{\Phi}(t_{n+1},t_n)x(t_n) + u(t_n). \tag{111}$$

Therefore, the covariance of $x(t_{n+1})$ is

$$E[x(t_{n+1})x^T(t_{n+1})] = \bar{\Phi}(t_{n+1},t_n)E[x(t_n)x^T(t_n)]\ \bar{\Phi}^T(t_{n+1},t_n) + \bar{Q}(t_n). \tag{112}$$

where the substitution has been made that

$$\bar{Q}(t_n) = E[\tilde{\Phi}(t_{n+1},t_n)x(t_n)x^T(t_n)\ \tilde{\Phi}(t_{n+1},t_n)] + E[u(t_n)u^T(t_n)], \tag{113}$$

and the cross terms in Eq. (112) are zero because $\tilde{\Phi}(t_{n+1},t_n)$ and $u(t_n)$ have zero mean and are independent of $x(t_n)$.

Substituting Eq. (112) into the covariance in Eq. (110) gives

$$P(t_{n+1},t_n) = \bar{\Phi}(t_{n+1},t_n)[I - K(t_n)\bar{M}(t_n)]\ x$$

$$\left\{E[x(t_n)x^T(t_n)] - E[\hat{x}(t_n/t_{n-1})x^T(t_n)]\right\}\bar{\Phi}^T(t_{n+1},t_n) + \bar{Q}(t_n) \tag{114}$$

The quantity in braces in Eq. (114) is the covariance of the error $P(t_n/t_{n-1})$ in Eq. (95). Replacing that quantity in Eq. (114) by $P(t_n/t_{n-1})$ yields as the iterative equation for the covariance of the error

$$P(t_{n+1},t_n) = \bar{\Phi}(t_{n+1},t_n)\left\{ I - K(t_n)\bar{M}(t_n)\right\} P(t_n/t_{n-1})\, \bar{\Phi}^T(t_{n+1},t_n) + \bar{Q}(t_n), \tag{115}$$

where $\bar{Q}(t_n)$ which is defined in Eq. (113) and the two expected values $\bar{M}(t_n)$ and $\bar{\Phi}(t_{n+1},t_n)$ are known a priori. The covariance of the error for the current optimum prediction is given by Eq. (115) when $P(t_n/t_{n-1})$ is the covariance of the previous optimum prediction and the r x p weighting matrix $K(t_n)$ satisfies Eq. (106) [which is repeated as Eq. (116)].

$$P(t_n/t_{n-1})\bar{M}^T(t_n) - K(t_n)\left\{ \bar{M}(t_n)P(t_n/t_{n-1})\bar{M}^T(t_n) + \bar{R}(t_n)\right\} = 0. \tag{116}$$

If the quantity in braces in Eq. (116) is nonsingular, then it has a unique inverse, and the weighting $K(t_n)$ is determined uniquely by

$$K(t_n) = P(t_n/t_{n-1})\bar{M}^T(t_n)\left\{ \bar{M}(t_n)P(t_n/t_{n-1})\bar{M}^T(t_n) + \bar{R}(t_n)\right\}^{-1}. \tag{117}$$

Both $\bar{M}(t_n)P(t_n/t_{n-1})\bar{M}^T(t_n)$ and $\bar{R}(t_n)$, which is defined in Eq. (104), are positive semidefinite from their definition as the expected values of the covariance of vector-valued quantities, so, if either of these two matrices is positive definite, the whole quantity in braces in Eq. (112) will be positive definite and therefore nonsingular.

The matrix $\bar{M}(t_n)P(t_n/t_{n-1})\bar{M}^T(t_n)$ will be positive definite if the rows of $\bar{M}(t_n)$ are independent and if the covariance $P(t_n/t_{n-1})$ is positive definite. The covariance $P(t_n/t_{n-1})$ is positive definite unless there is some $x_i(t_n)$ that is known exactly; if it were known exactly; it wouldn't have to be estimated, and the estimation problem could be reduced by one dimension. When the rows of the output matrix $\bar{M}(t_n)$ are dependent and there is no multiplicative noise on the measurements, there is no information lost by reducing the number of observed random variables until the rows of $\bar{M}(t_n)$ become independent. If there is independent multiplicative noise on all the measurements, the covariance $\bar{R}(t_n)$ is positive definite and the quantity in braces in Eq. (116) will be positive definite. Therefore, even if the quantity

in braces in Eq. (116) is not nonsingular to start with, it should be possible to restate the estimation problem so that the quantity will be nonsingular in the new statement of the problem.

If for some reason the weighting $K(t_n)$ must be chosen when the quantity in braces in Eq. (116) is a singular matrix, the weighting $K(t_n)$, that satisfies Eq. (116) and has the minimum norm of all the weighting functions that satisfy Eq. (116), is given by

$$K(t_n) = P(t_n/t_{n-1})\bar{M}^T(t_n)\left\{\bar{M}(t_n)P(t_n/t_{n-1})\bar{M}^T(t_n) + \bar{R}(t_n)\right\}^\dagger . \tag{118}$$

where the matrix $A^\dagger$ is defined as the psuedoinverse of the matrix A. A particularly concise explanation of the properties of the psuedoinverse is given by Gunckel [Ref. 4] in an appendix, but more complete explanations can be found elsewhere [Ref. 26]. For this investigation, it is sufficient to say that the psuedoinverse gives the solution to a set of underspecified equations which has the minimum norm of all possible solutions to the set of equations. In the remainder of this paper it will be assumed that the matrix in braces in Eq. (116) is nonsingular and has a unique inverse. If this matrix is singular, it is necessary only to replace the expression for the inverse of this matrix by the psuedoinverse of the matrix, and all the equations will still be valid and the solution will have the minimum norm of all possible solutions.

E. SUMMARY OF RESULTS

The procedure developed in this chapter yields an iterative solution to the problem of the optimum filtering and prediction of a sampled stochastic process that can be represented by the model developed in Chapter II. The optimum estimate of the current value of the state vector $x(t_n)$ given the observed random variables $y(t_n), y(t_{n-1}), \ldots, y(t_0)$ is

$$\hat{x}(t_n/t_n) = [I - K(t_n)\bar{M}(t_n)]\,\bar{\Phi}(t_n,t_{n-1})\hat{x}(t_{n-1}/t_{n-1}) + K(t_n)y(t_n) \tag{119}$$

where $\hat{x}(t_{n-1}/t_{n-1})$ is the previous optimum estimate and the expected values $\bar{M}(t_n)$ and $\bar{\Phi}(t_n,t_{n-1})$ are known a priori. The r x p weighting matrix $K(t_n)$ is given by

$$K(t_n) = P(t_n/t_{n-1})\bar{M}^T(t_n)\left\{\bar{M}(t_n)P(t_n/t_{n-1})\bar{M}^T(t_n) + \bar{R}(t_n)\right\}^{-1}, \tag{120}$$

where the covariance

$$\bar{R}(t_n) = E[\tilde{M}(t_n)x(t_n)x^T(t_n)\tilde{M}^T(t_n)] \tag{121}$$

is known a priori.

The covariance matrix of the error in prediction $P(t_{n+1}/t_n)$ is calculated by the recursion relation

$$P(t_{n+1}/t_n) = \bar{\Phi}(t_{n+1},t_n)[I - K(t_n)\bar{M}(t_n)]P(t_n/t_{n-1})\bar{\Phi}^T(t_{n+1},t_n) + \bar{Q}(t_n), \tag{122}$$

where the covariance

$$\bar{Q}(t_n) = E[\tilde{\Phi}(t_{n+1},t_n)x(t_n)x^T(t_n)\tilde{\Phi}^T(t_{n+1},t_n)] + E[u(t_n)u^T(t_n)] \tag{123}$$

is also known a priori.

Substituting Eq. (120) into Eq. (122) gives the nonlinear difference equation that determines the covariance of the error as

$$P(t_{n+1},t_n) = \bar{\Phi}(t_{n+1},t_n)\left\{I - P(t_n/t_{n-1})\bar{M}^T(t_n)[\bar{M}(t_n)P(t_n/t_{n-1})\bar{M}^T(t_n) + \bar{R}(t_n)]^{-1}\bar{M}(t_n)\right\}P(t_n/t_{n-1})\bar{\Phi}^T(t_{n+1},t_n) + \bar{Q}(t_n). \tag{124}$$

It is interesting to compare Eq. (124) with the equivalent expression for the covariance of the error in the optimum prediction when the parameters of the process are not random. When the parameters are not random,

$$
\begin{aligned}
\bar{M}(t_n) &\to M(t_n) \\
\tilde{M}(t_n) &\to 0 \\
\bar{\Phi}(t_{n+1},t_n) &\to \Phi(t_{n+1},t_n) \\
\tilde{\Phi}(t_{n+1},t_n) &\to 0 \\
\bar{R}(t_n) &\to 0 \\
\bar{Q}(t_n) &\to Q(t_n) = E[u(t_n)u^T(t_n)],
\end{aligned} \tag{125}
$$

so that the expression equivalent to Eq. (124) is

$$
P(t_{n+1},t_n) = \Phi(t_{n+1},t_n)\Big\{ I - P(t_n/t_{n-1})M^T(t_n)[M(t_n)P(t_n/t_{n-1})M^T(t_n)]^{-1} \times M(t_n)\Big\} P(t_n/t_{n-1})\Phi^T(t_{n+1},t_n) + Q(t_n). \tag{126}
$$

Equation (126) is the same as the nonlinear difference equation derived by Kalman [Ref. 1] for the estimation problem when the parameters of the process are known a priori, instead of random parameters. The effect of the random variations in the parameters on the covariance of the error in the optimum prediction $P(t_{n+1}/t_n)$ in Eq. (124) is included in the covariance matrices $\bar{R}(t_n)$ and $\bar{Q}(t_n)$.

When the parameters of the process M and Φ are a priori known constants, Kalman [Ref. 1] says it can be shown that the optimum estimate in Eq. (119) becomes a stationary dynamic system in the limit as the number of observed random variables n approach infinity. When the matrices $\bar{M}$ and $\bar{\Phi}$ and the covariances $\bar{R}$ and $\bar{Q}$ are a priori known constants, the optimum estimate in Eq. (119) should also become a stationary dynamic system in the limit. The reasoning behind the latter statement is as follows:

> The trace of the covariance matrix of the error in Eq. (124) is bounded from above by $E[x(t_n)x^T(t_n)]$ (which would result from an estimate that the state vector is identically zero) and from below by zero. The trace for the optimum estimate should not increase from one estimate to the next, because if identical weightings are used for consecutive estimates, the trace of the covariance matrix of the error will be the same for stationary statistics.

Thus, the trace is a bounded nonincreasing sequence, so it must approach a limit as n approaches infinity. This same argument can be used for the trace of the covariance of the error in the optimum estimate of any linear combination of state variables; therefore, the individual elements of the covariance matrix must all approach a limiting value. The limiting value of the weighting matrix $K(t_n)$ is obtained from Eq. (120), therefore, the optimum estimate in Eq. (119) becomes a stationary dynamic system.

The matrix block diagram in Fig. 5 represents the linear filter that implements the optimum estimate in Eq. (119). The element marked DELAY relates the state of the filter at time t_{n+1} to the state of the filter at time t_n. The covariance of the error in the optimum current estimate in Eq. (119) is given by

$$P(t_n/t_n) = [I - K(t_n)\bar{M}(t_n)]\, P(t_n/t_{n-1}) \tag{127}$$

so that from Eq. (124),

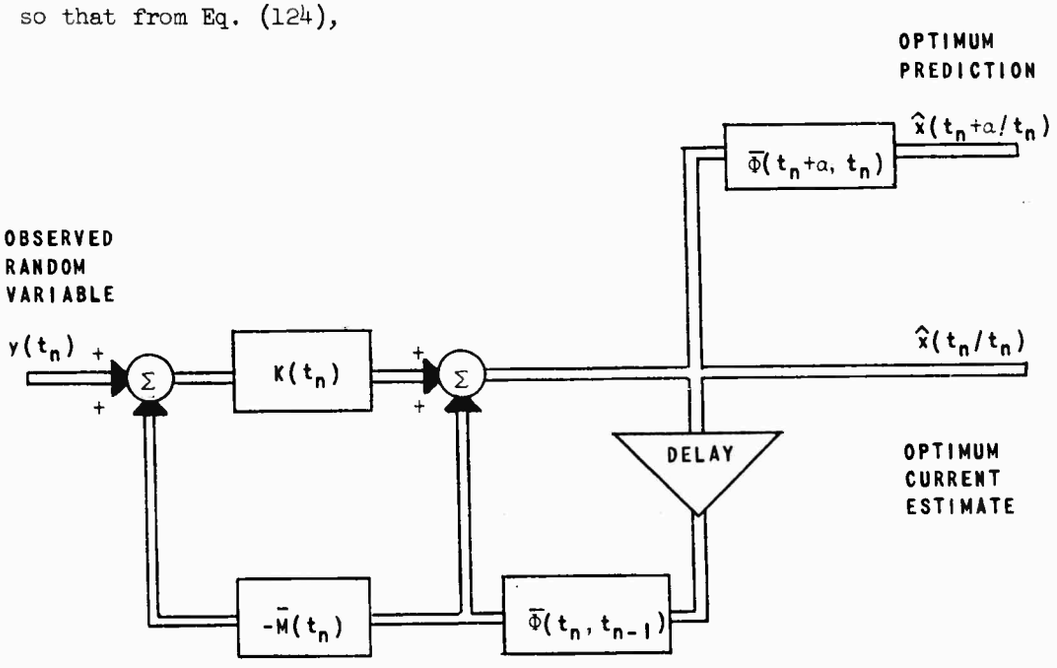


FIG. 5. MODEL OF THE OPTIMUM FILTER.

$$P(t_{n+1}/t_n) = \overline{\Phi}(t_{n+1},t_n)P(t_n/t_n)\overline{\Phi}^T(t_{n+1},t_n) + \overline{Q}(t_n). \qquad (128)$$

The optimum prediction of the state vector is

$$\hat{x}(t_n+\alpha/t_n) = \overline{\Phi}(t_n+\alpha,t_n)\hat{x}(t_n/t_n) \qquad \alpha \geq 0 \qquad (129)$$

This optimum prediction can also be used to interpolate between sample points when the α is time varying so that

$$\alpha = t - t_n \qquad t_n \leq t \leq t_{n+1}. \qquad (130)$$

The initial condition of the linear filter with no observed random variables and no other given information is the expected value of the state vector, which is zero, i.e.,

$$\hat{x}(t_o/0) = 0. \qquad (131)$$

The covariance of the error in this initial estimate is

$$P(t_o/0) = E[x(t_o)x^T(t_o)]. \qquad (132)$$

The weighting $K(t_n)$ is determined iteratively at each sample point from Eqs. (120) and (124) and the initial conditions.

The examples in the following section will illustrate the ideas of this chapter.

F. EXAMPLES

The procedure for estimation and prediction derived in this chapter uses the model of the stochastic process developed in Chapter II, where the stochastic process and the random parameters can have nonstationary statistics. The purpose of this section is to illustrate and clarify this procedure; therefore, the examples will be limited to stochastic processes and random parameters with stationary statistics. For the second-order process with exponential cosine autocorrelation, a digital

computer was used in the iteration of the equations which determine the weighting function and the covariance matrix of the error at each point.

For a stationary, stochastic process the derivatives of the covariance of the state vector is zero. This fact leads to the relation between the covariance of the state vector $E[x(t)x^T(t)]$ and the covariance of the input V where F and G are matrices defined in Chapter II.

$$F \cdot E[x(t)x^T(x)] + E[x(t)x^T(t)]F^T + GVG^T = 0. \tag{133}$$

The covariance $\bar{Q}(t_n)$ can be determined from the stationary version of Eq. (112), which is

$$E[x(t_n)x^T(t_n)] = \bar{\Phi}(T_n)E[x(t_n)x^T(t_n)]\bar{\Phi}^T(T_n) + \bar{Q}(t_n). \tag{134}$$

The solution to Eq. (134) is given by

$$\bar{q}_{il}(t_n) = E[x_i(t_n)x_l(t_n)] - \sum_{j=1}^{r}\sum_{k=1}^{r}\bar{\phi}_{ij}(T_n)E[x_j(t_n)x_k(t_n)]\bar{\phi}_{lk}(T_n). \tag{135}$$

1. Exponential Autocorrelation

The first example concerns the stationary first order Markov process which was presented in Fig. 2 and discussed in Sec. II-B. The autocorrelation function of the process is

$$R_{xx}(\tau) = e^{-\beta|\tau|}. \tag{136}$$

In this example the subscript 1 can be omitted because all the vectors and all the matrices are scalars. The linear differential equation describing the system is

$$\frac{dx(t)}{dt} = -\beta x(t) + V(t). \tag{137}$$

Therefore, the transition matrix is a scalar given by

$$\Phi(T_n) = \phi(T_n) = e^{-\beta T_n}. \tag{138}$$

The expected value of the transition matrix $\bar{\Phi}(T_n)$ will be abbreviated $\bar{\phi}$, so that

$$\bar{\phi} = E[\phi(T_n)] = E[\exp(-\beta T_n)]. \tag{139}$$

From the autocorrelation function in Eq. (136) the covariance of the state vector $E[x(t)x(t)]$ is given by

$$E[x(t)x(t)] = \mathcal{R}_{xx}(0) = 1. \tag{140}$$

Substituting Eqs. (139) and (140) into Eq. (134) gives the covariance $\bar{Q}(t_n)$ as

$$\bar{Q}(t_n) = \bar{q}(t_n) = 1 - \bar{\phi}^2. \tag{141}$$

The multiplicative noise has a mean of unity and a variance of $\bar{r}$, so that

$$\bar{M}(t_n) = E[m(t_n)] = 1$$

$$\bar{R}(t_n) = \bar{r} = E[\tilde{m}(t_n)x(t_n)x(t_n)\tilde{m}(t_n)] \tag{142}$$

The model of the optimum filter for this example is presented in Fig. 6. The weighting $k(t_n)$ is a scalar given by Eq. (120) as

$$k(t_n) = \frac{p(t_n/t_{n-1})}{p(t_n/t_{n-1}) + \bar{r}} \tag{143}$$

with the covariance of the error in prediction $p(t_{n+1}/t_n)$ calculated iteratively from the nonlinear difference Eq. (124), which is

$$p(t_{n+1}/t_n) = \bar{\phi}\left[1 - \frac{p(t_n/t_{n-1})}{p(t_n/t_{n-1}) + \bar{r}}\right]p(t_n/t_{n-1})\bar{\phi} + (1-\bar{\phi}^2). \tag{144}$$

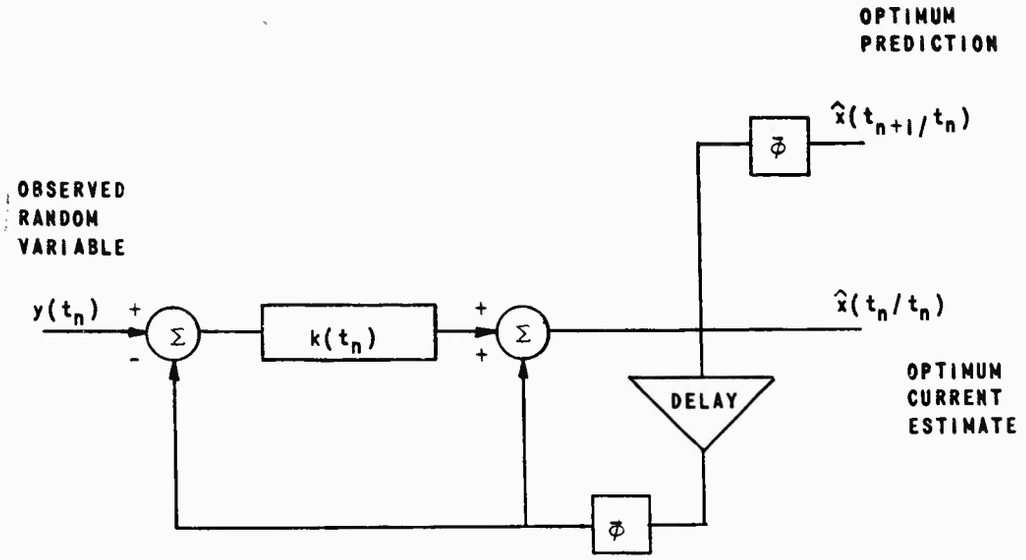


FIG. 6. MODEL OF OPTIMUM FILTER FOR A STATIONARY MARKOV PROCESS.

The covariance of the error in the current estimate is given by Eq. (127), as

$$p(t_n/t_n) = \frac{p(t_n/t_{n-1})\bar{r}}{p(t_n/t_{n-1}) + \bar{r}} . \tag{145}$$

For a nonstationary Markov process with the parameter β and the covariances $\bar{\emptyset}$ and $\bar{r}$ time-varying quantities, Eqs. (143), (144), and (145) would still be valid, but β, $\bar{\emptyset}$, and $\bar{r}$ would have different values at each sample point. For a stationary process, the covariance of the error $p(t_{n+1}/t_n)$ will approach a steady-state value as n approaches infinity.

$$\lim_{n\to\infty} p(t_{n+1}/t_n) = \lim_{n\to\infty} p(t_n/t_{n-1}) = \bar{p}(t_{n+1}/t_n). \tag{146}$$

The steady-state version of Eq. (144) is a quadratic in $\bar{p}(t_{n+1}/t_n)$,

$$\bar{p}(t_{n+1}/t_n) = \frac{\bar{\phi}^2\ \bar{r}\ \bar{p}(t_{n+1}/t_n)}{\bar{p}(t_{n+1}/t_n) + \bar{r}} + (1 - \bar{\phi}^2). \tag{147}$$

which has the solution

$$\bar{p}(t_{n+1}/t_n) = \frac{1}{2}(1 - \bar{\phi}^2)\left[(1-\bar{r}) + \sqrt{(1-\bar{r})^2 + \frac{4\bar{r}}{(1-\bar{\phi}^2)}}\right]. \tag{148}$$

(The square root in Eq. (148) has a positive sign because $\bar{p}(t_{n+1}/t_n)$ must always be positive.)

Three common proabability laws for the sample period T_n were discussed in Sec. II-E. The pertinent information for these three laws for particular numerical values of the parameters of the probability laws and for periodic sampling is presented in Table 2. The expected value of the sample period is normalized to unity, and the variances of the sample period have the values stated. The expected value of the transition matrix $\bar{\phi}$ is listed for

$$\beta = 0.1. \tag{149}$$

For all numerical calculations this value of β will be used; therefore, the system will have a time constant of ten seconds. Notice how increasing the variance of the sample period increases the expected value of the transition matrix $\bar{\phi}$.

The steady-state values of the covariance of the error $\bar{p}(t_n/t_n)$ and $\bar{p}(t_{n+1}/t_n)$ are presented in Table 3 for the same expected values of $\bar{\phi}$ listed in Table 2 for two cases--with no multiplicative noise and with the variance of the multiplicative noise equal to one.

Notice that with optimum estimation the variation in the probability distribution of the sample period has a smaller effect on the covariance of the error than does the variance of the multiplicative noise. The two covariances $\bar{p}(t_n/t_n)$ and $\bar{p}(t_{n+1}/t_n)$ are plotted in Fig. 7 as a function of the variance of the multiplicative noise $\bar{r}$ for

TABLE 2. NUMERICAL VALUES FOR FOUR PROBABILITY LAWS.

Probability Distribution of Sample Period	Periodic Sampling	Periodic Sampling with Jitter	Periodic Sampling with Independent Misses	Purely Random or Poisson Sampling
	Constant	Normal	Geometric	Exponential
Parameters of Probability Distribution	$\bar{T}=1$	$\bar{T}=1$ $\sigma=\frac{1}{3}$	$\bar{T}=\frac{1}{2}$ $p=q=\frac{1}{2}$	$\mu=1$
Mean Sample Period $E[T_n]$	$\bar{T}=1$	$\bar{T}=1$	$T/p=1$	$1/\mu=1$
Variance of Sample Period $E[T_n^2]-(E[T_n])^2$	0	$\sigma^2=\frac{1}{9}$	$\frac{q}{p^2}T^2=\frac{1}{2}$	$1/\mu^2=1$
$\phi = E[\exp(-\frac{T_n}{10})]$	0.9048	0.9053	0.9070	0.9091

TABLE 3. COVARIANCE OF THE ERROR FOR EXAMPLE 1.

Probability Distribution of Sample Period	Constant	Normal	Geometric	Exponential
$\phi=E[\exp(-\frac{T_n}{10})$	0.9048	0.9053	0.9070	0.9091
$\bar{p}(t_n/t_n)$ with $\bar{r}=0$	0.0	0.0	0.0	0.0
$\bar{p}(t_n+1/t_n)$ with $\bar{r}=0$	0.1813	0.1804	0.1773	0.1736
$\bar{p}(t_n/t_n)$ with $\bar{r}=1$	0.2986	0.2982	0.2962	0.2927
$\bar{p}(t_n+1/t_n)$ with $\bar{r}=1$	0.4258	0.4247	0.4211	0.4167

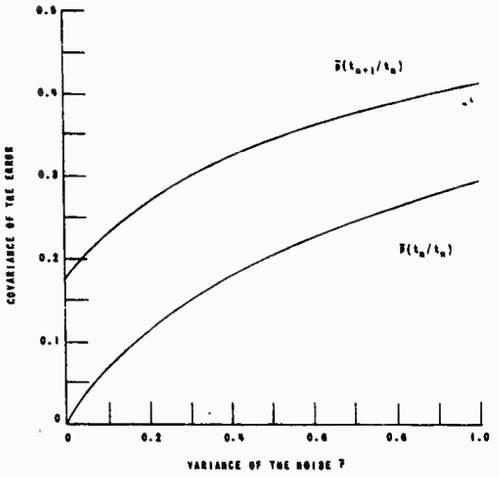


FIG. 7. COVARIANCE OF THE ERROR FOR EXAMPLE 1.

$$\bar{\phi} = 0.9091, \tag{150}$$

which is the case where the sample period has an exponential probability distribution (purely random sampling).

2. Exponential Cosine Autocorrelation

The second example concerns the stationary process with the exponential cosine autocorrelation

$$\mathcal{R}_{x_1 x_1}(\tau) = e^{-\beta|\tau|}\left(\cos \gamma\tau - \frac{\beta\gamma \sin \gamma|\tau|}{2\beta^2 + \gamma^2}\right) \tag{151}$$

that was presented in Fig. 3 and discussed in Sec. II-C. The linear differential vector equation that describes the system is

$$\frac{d}{dt}\begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} = \begin{bmatrix} -\beta & 1 \\ -\gamma^2 & -\beta \end{bmatrix}\begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} + \begin{bmatrix} V_1(t) \\ 0 \end{bmatrix} \tag{152}$$

The transition matrix is given by Eq. (40) as

$$\Phi(T_n) = e^{-\beta T_n} \begin{bmatrix} \cos \gamma T_n & \frac{1}{\gamma} \sin \gamma T_n \\ -\gamma \sin \gamma T_n & \cos \gamma T_n \end{bmatrix} . \qquad (153)$$

When the sample period has a geometric distribution, the expected value of the transition matrix can be calculated by changing the sines and cosines to complex exponentials, so that

$$E[e^{-\beta T_n} \cos \gamma T_n] = RE[\exp(-\beta T_n + j\gamma T_n)] = \frac{\mu(\mu + \beta)}{(\mu + \beta)^2 + \gamma^2}$$

$$E[e^{-\beta T_n} \sin \gamma T_n] = 1M[\exp(-\beta T_n - j\gamma T_n)] = \frac{\mu\gamma}{(\mu + \beta)^2 + \gamma^2} \qquad (154)$$

where RE and 1M represent the real and imaginary parts, respectively.

The covariance of the state vector $E[x(t)x^T(t)]$ is determined from Eq. (133). For this example the solution to Eq. (133) is

$$E[x_1(t)x_1(t)] = \Re_{x_1x_1}(0) = 1$$

$$E[x_1(t)x_2(t)] = E[x_2(t)x_1(t)] = -\beta\gamma^2/(\gamma^2 + 2\beta^2)$$

$$E[x_2(t)x_2(t)] = \gamma^4/(\gamma^2 + 2\beta^2)$$

$$E[v_1(t)v_1(t)] = v_{11} = 4\beta(\gamma^2 + \beta^2)/(\gamma^2 + 2\beta^2). \qquad (155)$$

The covariance $\bar{Q}(t_n)$ is determined directly from Eq. (135). The multiplicative noise has a mean of unity and a variance of $\bar{r}$, so that

$$E[M^T(t_n)] = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$\bar{R}(t_n) = E[\tilde{M}(t_n)x(t_n)x^T(t_n)\tilde{M}^T(t_n)] = \bar{r}\, E[x_1(t_n)x_1(t_n)] = \bar{r}. \qquad (156)$$

The numerical values of the transition matrix $\bar{\Phi}(T_n)$ and the covariances $\bar{Q}(t_n)$ and $E[x(t)x^T(t)]$ are presented in Table 4 for a geometric distribution of the sample period (purely random sampling) with $\gamma = 1$, so that

$$\mu = 1.0$$
$$\beta = 0.1$$
$$\gamma = 1.0. \tag{157}$$

TABLE 4. NUMERICAL VALUES OF MATRICES FOR EXAMPLE 2.

$\bar{\Phi}(T_n)$	$\begin{bmatrix} 0.4977 & 0.4525 \\ -0.4525 & 0.4977 \end{bmatrix}$
$\bar{Q}(t_n)$	$\begin{bmatrix} 0.5956 & -0.08964 \\ -0.08964 & 0.4886 \end{bmatrix}$
$E[x(t)x^T(t)]$	$\begin{bmatrix} 1.0 & -0.09804 \\ -0.09804 & 0.9804 \end{bmatrix}$

The expected value of the transition matrix $\bar{\Phi}(T_n)$ and the covariance $\bar{Q}(t_n)$ are needed for the iterative calculation of the covariance of the error $P(t_{n+1}/t_n)$ from the nonlinear difference Eq. (124). The initial condition for the covariance of the error $P(t_o/0)$ with no other information is

$$P(t_o/0) = E[x(t)x^T(t)]. \tag{158}$$

Because the observed random variable is a scalar, the quantity in Eq. (124), which must be inverted, is also a scalar.

$$[\bar{M}(t_n)P(t_n/t_{n-1})\bar{M}^T(t_n) + \bar{R}(t_n)]^{-1} = 1/[P_{11}(t_{n+1}/t_n) + \bar{r}] \qquad (159)$$

For this example the steady-state covariance of the error $\bar{P}(t_{n+1}/t_n)$ was obtained by iteration of Eq. (124) on a digital computer. The steady-state value of the covariance $\bar{P}(t_{n+1}/t_n)$ and the weighting matrix $\bar{K}(t_n)$ are presented for the numerical values of the matrices in Table 5 when the multiplicative noise has a variance of one-tenth,

$$\bar{r} = 0.1. \qquad (160)$$

TABLE 5. DESIGN RESULTS FOR EXAMPLE 2.

$\bar{P}(t_{n+1}/t_n)$	$\begin{bmatrix} 0.7564 & 0.0411 \\ 0.0411 & 0.6700 \end{bmatrix}$
$\bar{K}(t_n)$	$\begin{bmatrix} 0.8832 \\ 0.0480 \end{bmatrix}$
$K(t_0)$	$\begin{bmatrix} 0.9091 \\ 0.0893 \end{bmatrix}$
$K(t_1)$	$\begin{bmatrix} 0.8905 \\ 0.1190 \end{bmatrix}$
$K(t_2)$	$\begin{bmatrix} 0.8856 \\ 0.0656 \end{bmatrix}$
$K(t_3)$	$\begin{bmatrix} 0.8837 \\ 0.05121 \end{bmatrix}$

The weighting matrix $K(t_n)$ is also presented in Table 5 for the first four sample points to show how fast the iterations converge to a steady-state value.

From Eq. (127) the steady-state value of the covariance $p_{11}(t_n/t_n)$ is

$$\bar{p}_{11}(t_n/t_n) = \frac{\bar{p}_{11}(t_{n+1}/t_n)\,\bar{r}}{\bar{p}_{11}(t_{n+1}/t_n) + \bar{r}}. \qquad (161)$$

In Fig. 8 the steady-state covariances of the error $\bar{p}_{11}(t_n/t_n)$ and

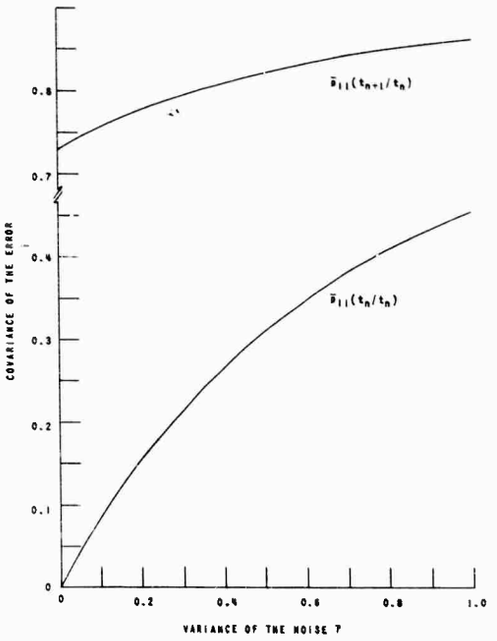


FIG. 8. COVARIANCE OF THE ERROR FOR EXAMPLE 2.

$\bar{p}_{11}(t_{n+1}/t_n)$ are plotted as a function of the variance of the multiplicative noise $\bar{r}$ for a geometric distribution of the sample period for the value of the matrices in Table 4. Notice how much the cosine term in the autocorrelation has increased the covariance of the error in Fig. 8 over the corresponding covariances in Fig. 7.

3. Nonstationary Process

The third example concerns a first-order Markov process with a particular kind of nonstationarity. The linear differential equation describing the system is

$$\frac{dx(t)}{dt} = -\beta(t)x(t) + v(t)$$

with $\beta(t)$ time varying so that

$$\beta(t) = B_1 \qquad 2k < t \leq 2k + 1$$

$$= B_2 \qquad 2k + 1 < t \leq 2k \qquad k = 0,1,2,\ldots$$

Given the value of the process at time $2k + \tau$ with

$$0 \leq \tau < 2,$$

it is desired to predict the value of the process at time $2k + \tau + 1$. From Eq. (129) the prediction is obtained by multiplying the current value of the process $x(2k + \tau)$ by the transition matrix $\phi(2k+\tau+1, 2k+\tau)$

$$\hat{x}(2k+\tau+1/2k+\tau) = \phi(2k+\tau+1, 2k+\tau)x(2k+\tau)$$

$$= e^{-B_1 T_1} e^{-B_2 T_2} x(2k+\tau)$$

where

$$T_1 = 1-\tau \qquad T_2 = \tau \qquad \text{for} \qquad 0 < \tau < 1$$

$$T_1 = \tau - 1 \qquad T_2 = 2-\tau \qquad \text{for} \qquad 1 \leq \tau < 2$$

If $v(t)$ has unity variance, the mean squared error of the estimate is

$$e^{-2B_2 T_2} \frac{1 - e^{-2B_1 T_1}}{2B_1} + \frac{1 - e^{-2B_2 T_2}}{2B_2} .$$

These examples have showed problems in prediction of random processes that can be solved by using the method developed in Chapter III.

IV. OPTIMUM INTERPOLATION WITH DELAY OF RANDOM PARAMETER PROCESSES

A. INTRODUCTION

In Chapter III the matrix-valued sampled Wiener-Hopf equation was derived, and the optimum estimate of the current value of the state vector was shown to be a linear combination of the previous optimum estimate and the newly observed random variable at the current sample point. In this chapter it will be proved that any optimum estimate that satisfies the sampled Wiener-Hopf equation can always be written in an iterative form similar to the one derived in Chapter III. The exact configuration of this iterative form is worked out for the general problem of interpolation with delay (at time t_n the optimum estimate is desired of the state vector at time t where $t < t_n$), and the iterative nonlinear difference equations are derived for the matrix-valued gain of the optimum filter with delay up to $(n - d)$ sample periods. The reason for delaying the estimate is that the trace of the covariance matrix of the error of the optimum estimate can be decreased when the estimate is delayed until additional random variables are observed.

B. ITERATIVE SOLUTION TO THE SAMPLED WIENER-HOPF EQUATION

In this section it will be proved that at any sample point t_n the optimum estimate $\hat{x}(t_n+\alpha/t_n)$ can be written in the iterative form

$$\hat{x}(t_n+\alpha/t_n) = \hat{x}(t_n+\alpha/t_{n-1}) + K(t_n+\alpha,t_n)[y(t_n) - \bar{M}(t_n)\hat{x}(t_n/t_{n-1})] \tag{162}$$

where $\hat{x}(t_n+\alpha/t_{n-1})$ is the optimum estimate at the previous sample point and $K(t_n+\alpha,t_n)$ is an $r \times p$ weighting matrix. When α is positive, the estimate $\hat{x}(t_n+\alpha/t_n)$ is called prediction, and when α is negative, it is interpolation with delay. The proof in this section is a generalization of the proof in Chapter III that, when α is zero, the optimum current estimate can be written in the form of Eq. (162).

Because the two estimates $\hat{x}(t_n+\alpha/t_{n-1})$ and $\hat{x}(t_n/t_{n-1})$ are optimum, they satisfy the appropriate versions of the sampled-Wiener-Hopf equation which are

$$E[x(t_n+\alpha)y^T(t_\rho)]-E[\hat{x}(t_n+\alpha/t_{n-1})y^T(t_\rho)] = 0 \qquad \rho = 0,1,\ldots,n-1 \tag{163}$$

and

$$E[x(t_n)y^T(t_\rho)]-E[\hat{x}(t_n/t_{n-1})y^T(t_\rho)] = 0 \qquad \rho = 0,1,\ldots,n-1. \tag{164}$$

It will be proved that there is an $r \times p$ weighting matrix $K(t_n+\alpha,t_n)$ for which the estimate in Eq. (162) satisfies Eq. (165),

$$E[x(t_n+\alpha)y^T(t_\rho)]-E[\hat{x}(t_n+\alpha/t_n)y^T(t_\rho)] = 0 \qquad \rho = 0,1,\ldots,n. \tag{165}$$

Substituting the trial estimate in Eq. (162) into Eq. (165) gives

$$\begin{aligned} &E[x(t_n+\alpha)y^T(t_\rho)]-E[\hat{x}(t_n+\alpha/t_{n-1})y^T(t_\rho)] \\ &\quad -K(t_n+\alpha,t_n)\Big\{E[M(t_n)x(t_n)y^T(t_\rho)] \\ &\quad -\bar{M}(t_n)E[\hat{x}(t_n/t_{n-1})y^T(t_\rho)]\Big\} = 0 \qquad \rho = 0,1,\ldots,n. \end{aligned} \tag{166}$$

For ρ less than n, Eq. (166) is equal to Eq. (163) minus $K(t_n+\alpha,t_n)\bar{M}(t_n)$ times Eq. (164). Both these equations are identically zero for $\rho < n$, so Eq. (166) is identically zero for $\rho < n$.

For ρ equal to n, Eq. (166) can be written

$$\begin{aligned} &\Big[E[x(t_n+\alpha)x^T(t_n)]-E[\hat{x}(t_n+\alpha/t_{n-1})x^T(t_n)]\Big]\bar{M}^T(t_n) \\ &\quad -K(t_n+\alpha,t_n)\Big\{\bar{M}(t_n)P(t_n/t_{n-1})\bar{M}^T(t_n)+\bar{R}(t_n)\Big\} = 0, \end{aligned} \tag{167}$$

where the substitution has been made that

$$E[M(t_n)x(t_n)y^T(t_n)]-E[\hat{x}(t_n/t_{n-1})y^T(t_n)$$

$$= \bar{M}(t_n)P(t_n/t_{n-1})\bar{M}^T(t_n)+\bar{R}(t_n), \tag{168}$$

with the covariance matrix of the error $P(t_n/t_{n-1})$ defined as

$$P(t_n/t_{n-1}) = E[x(t_n)x^T(t_n)]-E[\hat{x}(t_n/t_{n-1})x^T(t_n)], \tag{169}$$

and the covariance matrix $\bar{R}(t_n)$ defined in Eq. (104) as

$$\bar{R}(t_n) = E[\tilde{M}(t_n)x(t_n)x^T(t_n)\tilde{M}^T(t_n)]. \tag{170}$$

Therefore, when the $r \times p$ weighting matrix $K(t_n+\alpha,t_n)$ satisfies Eq. (167), the trial estimate in Eq. (162) is indeed optimum and satisfies the sampled-Wiener-Hopf equation (165). When the quantity in braces in Eq. (167) is nonsingular, the weighting matrix $K(t_n+\alpha,t_n)$ is determined uniquely as

$$K(t_n+\alpha,t_n) = \left\{E[x(t_n+\alpha)x^T(t_n)]-E[\hat{x}(t_n+\alpha/t_{n-1})x^T(t_n)]\right\} \bar{M}^T(t_n) \cdot$$

$$\left\{\bar{M}(t_n)P(t_n/t_{n-1})\bar{M}^T(t_n)+\bar{R}(t_n)\right\}^{-1}. \tag{171}$$

The conditions under which the second quantity in braces is nonsingular are discussed in Sec. III-C. When α is equal to zero, the weighting $K(t_n,t_n)$ for the optimum current estimate from Eq. (171) is

$$K(t_n,t_n) = \left\{E[x(t_n)x^T(t_n)]-E[\hat{x}(t_n/t_{n-1})x^T(t_n)]\right\} \bar{M}^T(t_n) \cdot$$

$$\left\{\bar{M}(t_n)P(t_n/t_{n-1})\bar{M}^T(t_n)+\bar{R}(t_n)\right\}^{-1}, \tag{172}$$

which is the same as $K(t_n)$ in Eq. (120).

It is also desirable to know the covariance matrix of the error of the optimum estimate $P(t_n+\alpha/t_n)$, which is given by Eq. (87) as

$$P(t_n+\alpha/t_n) = E[x(t_n+\alpha)x^T(t_n+\alpha)] - E[\hat{x}(t_n+\alpha/t_n)x^T(t_n+\alpha)]. \tag{173}$$

Substituting the optimum estimate in Eq. (162) into Eq. (174) gives

$$P(t_n+\alpha/t_n) = E[x(t_n+\alpha)x^T(t_n+\alpha)] - E[\hat{x}(t_n+\alpha/t_{n-1})x^T(t_n+\alpha)] - K(t_n+\alpha,t_n)\bar{M}(t_n)\left\{E[x(t_n)x^T(t_n+\alpha)]-E[\hat{x}(t_n/t_{n-1})x^T(t_n+\alpha)]\right\} \tag{175}$$

The first two terms in the expression for the covariance $P(t_n+\alpha/t_n)$ in Eq. (175) are equal to the previous covariance $P(t_n+\alpha/t_{n-1})$; thus, the covariance matrix of the error of the optimum estimate can also be written as

$$P(t_n+\alpha/t_n) = P(t_n+\alpha/t_{n-1})-K(t_n+\alpha,t_n)\bar{M}(t_n)\left\{E[x(t_n)x^T(t_n+\alpha)] - E[\hat{x}(t_n/t_{n-1})x^T(t_n+\alpha)]\right\} \tag{176}$$

When α is equal to zero, the covariance $P(t_n+\alpha/t_n)$ in Eq. (176) is equal to the covariance $P(t_n/t_n)$ in Eq. (127).

In the next section the iterative nonlinear difference equations determining the weighting matrix $K(t_n+\alpha,t_n)$ in Eq. (172) and the covariance matrix $P(t_n+\alpha/t_n)$ in Eq. (176) will be derived for interpolation with delay (when α is negative).

C. DERIVATION OF THE OPTIMUM ESTIMATE

To simplify the derivation of the optimum estimate for interpolation with a delay of $|\alpha|$ sec, the notation used in the preceding section will be changed slightly. The time $t_n+\alpha$ can be written

$$t_n + \alpha = t_d + \delta \qquad 0 \le \delta < T_d \tag{177}$$

so that, for α negative,

$$t_{\ell} \le t_n + \alpha < t_{d+1} \qquad \alpha \le 0$$
$$d \le n. \tag{178}$$

Using the new notation the optimum estimate $\hat{x}(t_n+\alpha/t_n)$ in Eq. (162) becomes

$$\hat{x}(t_d+\delta/t_n) = \hat{x}(t_d+\delta/t_{n-1}) + K(t_d+\delta,t_n)[y(t_n)-\bar{M}(t_n)\hat{x}(t_n/t_{n-1})]. \tag{179}$$

Making use of Eq. (162) once more the optimum estimate $\hat{x}(t_d+\delta/t_{n-1})$ in Eq. (179) can be written

$$\hat{x}(t_d+\delta/t_{n-1}) = \hat{x}(t_d+\delta/t_{n-2})+K(t_d+\delta,t_{n-1})[y(t_{n-1})-\bar{M}(t_{n-1})\hat{x}(t_{n-1}/t_{n-2})]. \tag{180}$$

Continuing this process for $(n - \ell)$ times the expression for the optimum estimate, $\hat{x}(t_d+\delta/t_n)$ becomes

$$\hat{x}(t_d+\delta/t_n) = \hat{x}(t_d+\delta/t_{\ell}) + \sum_{k=\ell+1}^{n} K(t_d+\delta,t_k)[y(t_k)-\bar{M}(t_k)\hat{x}(t_k/t_{k-1})]. \tag{181}$$

For a delay of $(n-d)$ sample points, it is necessary to calculate $(n-d)$ weighting matrices $K(t_d+\delta,t_k)$.

The matrix block diagram in Fig. 9 represents the linear filter which implements the optimum estimate in Eq. (181). The elements marked DELAY relate the state of the filter of time t_n to the state of the filter at time t_{n+1}. That part of the filter within the dashed rectangle duplicates the model of the optimum filter for prediction in Fig. 5, which was derived in Chapter III, and calculates the optimum prediction $\hat{x}(t_k/t_{k-1})$ in Eq. (181).

The weighting matrix $K(t_d+\delta,t_n)$, where n represents any number

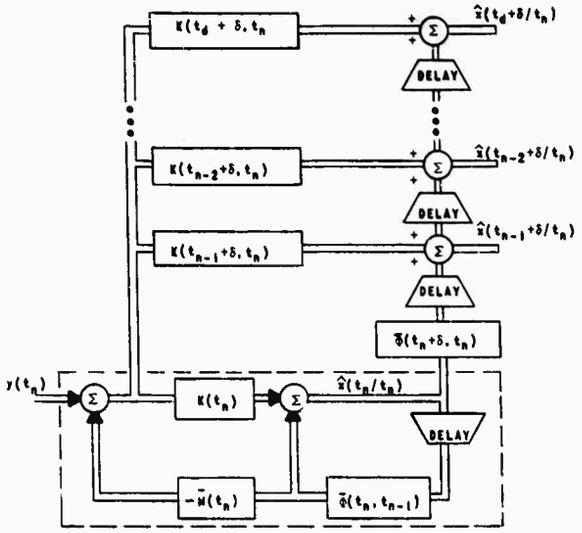


FIG. 9. MODEL OF OPTIMUM FILTER WITH DELAY.

greater than d, must satisfy Eq. (167) with $t_n+\alpha$ replaced by $t_d+\delta$; i.e.,

$$\left\{E[x(t_d+\delta)x^T(t_n)] - E[\hat{x}(t_d+\delta/t_{n-1})x^T(t_n)]\right\}\bar{M}^T(t_n) - K(t_d+\delta,t_n)[\bar{M}(t_n)P(t_n/t_{n-1})\bar{M}^T(t_n)+\bar{R}(t_n)] = 0. \tag{182}$$

The covariance of the error $P(t_k/t_{k-1})$ in Eq. (182) can be determined iteratively at each sample point by Eq. (124); thus, only the first term in braces in Eq. (182) remains unknown. This term is composed of the sum of two quantities $E[x(t_d+\delta)x^T(t_n)]$ and $E[\hat{x}(t_d+\delta/t_{n-1})x^T(t_n)]$. This sum will be determined first for the case where there is a delay of only one sample point, so that n is equal to d+1. A method will then be derived for calculating this sum in the general case where there is a delay of (n-d) sample points.

From the solution to the ordinary differential equation representing the system in Eq. (10), the state vextor $x(t_d+\delta)$ can be written

$$x(t_d+\delta) = \Phi(t_d+\delta,t_d)x(t_d)+u(t_d+\delta,t_d), \tag{183}$$

where $u(t_d+\delta,t_d)$ is an r x 1 vector defined as

$$u(t_d+\delta,t_d) = \int_{t_d}^{t_d+\delta} \Phi(t_d+\delta,\tau)G(\tau)v(\tau)d\tau \tag{184}$$

to represent the effect of the random inputs to the system between time t_d and time $t_d+\delta$. From the model of the process and the properties of the random parameters in Eq. (53) the state vectors can be written

$$\begin{aligned} x(t_d+\delta) &= \Phi(t_d+\delta,t_d)x(t_d) + u(t_d+\delta,t_d) \\ &= \bar{\Phi}(t_d+\delta,t_d)x(t_d) + \tilde{\Phi}(t_d+\delta,t_d)x(t_d) + u(t_d+\delta,t_d) \end{aligned} \tag{185}$$

and

$$\begin{aligned} x(t_{d+1}) &= \Phi(t_{d+1},t_d)x(t_d) + u(t_{d+1},t_d) \\ &= \bar{\Phi}(t_{d+1},t_d)x(t_d) + \tilde{\Phi}(t_{d+1},t_d)x(t_d)+u(t_{d+1},t_d). \end{aligned} \tag{186}$$

Substituting Eqs. (185) and (186) into the quantity $E[x(t_d+\delta)x^T(t_{d+1})]$ gives

$$E[x(t_d+\delta)x^T(t_{d+1})] = \bar{\Phi}(t_d+\delta,t_d)E[x(t_d)x^T(t_d)]\ \bar{\Phi}^T(t_{d+1},t_d)+\bar{Q}_1(t_d+\delta), \tag{187}$$

where the expected values of the cross terms are zero, and the substitution has been made of the r x r matrix $\bar{Q}_1(t_d+\delta)$ defined as

$$\bar{Q}_1(t_d+\delta) = E[\tilde{\Phi}(t_d+\delta,t_d)x(t_d)x^T(t_d)\tilde{\Phi}^T(t_{d+1},t_d)]+E[u(t_d+\delta,t_d)u^T(t_{d+1},t_d)]. \tag{188}$$

The quantity $E[\hat{x}(t_d+\delta/t_d)x^T(t_{d+1})]$ can be written

$$E[\hat{x}(t_d+\delta/t_d)x^T(t_{d+1})] = \bar{\Phi}(t_d+\delta,t_d)E[\hat{x}(t_d/t_d)x^T(t_d)]\,\bar{\Phi}^T(t_{d+1},t_d). \quad (189)$$

Thus, by substituting Eqs. (187) and (189), the sum of the two quantities $E[x(t_d+\delta)x^T(t_{d+1})]$ and $E[\hat{x}(t_d+\delta/t_d)x^T(t_{d+1})]$ can be written

$$\begin{aligned} &E[x(t_d+\delta)x^T(t_{d+1})] - E[\hat{x}(t_d+\delta/t_d)x^T(t_{d+1})] \\ &\quad = \bar{\Phi}(t_d+\delta,t_d)E[x(t_d)x^T(t_d)]\bar{\Phi}^T(t_{d+1},t_d)+\bar{Q}_1(t_d+\delta) \\ &\qquad - \bar{\Phi}(t_d+\delta,t_d)E[\hat{x}(t_d/t_d)x^T(t_d)]\bar{\Phi}^T(t_{d+1},t_d) \\ &\quad = \bar{\Phi}(t_d+\delta,t_d)P(t_d/t_d)\,\bar{\Phi}^T(t_{d+1},t_d)+\bar{Q}_1(t_d+\delta), \end{aligned} \quad (190)$$

where the substitution has been made that

$$P(t_d/t_d) = E[x(t_d)x^T(t_d)]-E[\hat{x}(t_d/t_d)x^T(t_d)]. \quad (191)$$

Returning to the general case with the first term in braces in Eq. (182) composed of the sum of $E[x(t_d+\delta)x^T(t_n)]$ and $E[\hat{x}(t_d+\delta/t_{n-1})\cdot x^T(t_n)]$, the substitution of Eq. (181), for $\hat{x}(t_d+\delta/t_{n-1})$ gives

$$\begin{aligned} &\left\{E[x(t_d+\delta)x^T(t_n)] - E[\hat{x}(t_d+\delta/t_{n-1})x^T(t_n)]\right\} \\ &\quad = E[x(t_d+\delta)x^T(t_n)] - E[\hat{x}(t_d+\delta/t_d)x^T(t_n)] \\ &\quad - \sum_{k=d+1}^{n-1} K(t_d+\delta,t_k)\left\{E[y(t_k)x^T(t_{n-1})]-\bar{M}(t_k)E[\hat{x}(t_k/t_{k-1})x^T(t_n)]\right\} \end{aligned} \quad (192)$$

For k less than n,

$$\begin{aligned} &\left\{E[y(t_k)x^T(t_{n-1})] - \bar{M}(t_k)E[\hat{x}(t_k/t_{k-1})x^T(t_n)]\right\} \\ &\quad = \bar{M}(t_k)\left\{E[x(t_k)x^T(t_k)] - E[\hat{x}(t_k/t_{k-1})x^T(t_k)]\right\}\bar{\Phi}^T(t_n,t_k) \\ &\quad = \bar{M}(t_k)P(t_k/t_{k-1})\,\bar{\Phi}^T(t_n,t_k) \end{aligned} \quad (193)$$

where the covariance of the error $P(t_k/t_{k-1})$ has been substituted into Eq. (193). For d less than n

$$E[x(t_d+\delta)x^T(t_n)] - E[\hat{x}(t_d+\delta/t_d)x^T(t_n)]$$

$$= \left\{E[x(t_d+\delta)x^T(t_{d+1})] - E[\hat{x}(t_d+\delta/t_d)x^T(t_{d+1})]\right\}\bar{\Phi}^T(t_n,t_{d+1})$$

$$= \left\{\bar{\Phi}(t_d+\delta,t_d)P(t_d/t_d)\bar{\Phi}^T(t_{d+1},t_d)+\bar{Q}_1(t_d+\delta)\right\}\bar{\Phi}^T(t_n,t_{d+1}) \quad (194)$$

where Eq. (190) has been substituted into Eq. (194).

Finally, substituting Eqs. (192 - 194) into Eq. (182), which must be satisfied by the optimum weighting matrices, yields

$$\left\{\bar{\Phi}(t_d+\delta,t_d)P(t_d/t_d)\bar{\Phi}^T(t_{d+1},t_d)+\bar{Q}_1(t_d+\delta)\right\}\bar{\Phi}^T(t_n,t_{d+1})\bar{M}^T(t_n)$$

$$-\sum_{k=d+1}^{n-1} K(t_d+\delta,t_k)\bar{M}(t_k)P(t_k/t_{k-1})\,\bar{\Phi}^T(t_n,t_k)\bar{M}^T(t_n)$$

$$- K(t_d+\delta,t_n)\left\{M(t_n)P(t_n/t_{n-1})\bar{M}^T(t_n)+\bar{R}(t_n)\right\} = 0. \quad (195)$$

where $\bar{Q}_1(t_d+\delta)$ and $\bar{R}(t_n)$ are defined in Eqs. (188) and (167). First, set n equal to d+1, and Eq. (195) becomes

$$\left\{\bar{\Phi}(t_d+\delta,t_d)P(t_d/t_d)\,\bar{\Phi}^T(t_{d+1},t_d)+\bar{Q}_1(t_d+\delta)\right\}\bar{M}^T(t_n)$$

$$-K(t_d+\delta,t_{d+1})\left\{\bar{M}(t_{d+1})P(t_{d+1}/t_d)\bar{M}^T(t_{d+1})+\bar{R}(t_{d+1})\right\} = 0. \quad (196)$$

The optimum weighting matrix $K(t_d+\delta,t_{d+1})$ is determined from Eq. (196) to be

$$K(t_d+\delta,t_{d+1}) = \left\{\bar{\Phi}(t_d+\delta,t_d)P(t_d/t_d)\,\bar{\Phi}^T(t_{d+1},t_d)+\bar{Q}_1(t_d+\delta)\right\}\bar{M}^T(t_d)$$

$$\times\left\{\bar{M}(t_{d+1})P(t_{d+1}/t_d)\bar{M}^T(t_{d+1})+\bar{R}(t_{d+1})\right\}^{-1}. \quad (197)$$

Because the delay δ is known, the expected value $\bar{\Phi}(t_d+\delta,t_d)$ and

$\bar{Q}_1(t_d+\delta)$ can be calculated directly; thus, all the quantities in Eq. (197) are known a priori except for the covariances of the error $P(t_{d+1}/t_d)$ and $P(t_d/t_d)$, which can be determined iteratively at each sample point by Eqs. (124) and (127). The conditions under which Eq. (197) uniquely determines the weighting matrix $K(t_d+\delta,t_{d+1})$ are discussed in Section III-D.

Next, set n equal to d+2 in Eq. (195) and the result is

$$\left\{\bar{\Phi}(t_d+\delta,t_d)P(t_d/t_d)\ \bar{\Phi}^T(t_{d+1},t_d)+\bar{Q}(t_d+\delta)\right\}\ \bar{\Phi}^T(t_{d+2},t_{d+1})\bar{M}^T(t_{d+2})$$
$$- K(t_d+\delta,t_{d+1})\bar{M}(t_{d+1})P(t_{d+1}/t_d)\ \bar{\Phi}^T(t_{d+2},t_{d+1})\bar{M}^T(t_{d+2})$$
$$- K(t_d+\delta,t_{d+2})\left\{\bar{M}(t_{d+2})P(t_{d+2}/t_{d+1})\bar{M}^T(t_{d+2})+\bar{R}(t_{d+2})\right\} = 0. \qquad (198)$$

The optimum weighting matrix $K(t_d+\delta,t_{d+2})$ is determined from Eq. (198), to be

$$K(t_d+\delta,t_{d+2}) = \left\{\bar{\Phi}(t_d+\delta,t_d)P(t_d/t_d)\ \bar{\Phi}^T(t_{d+2},t_d)\bar{M}^T(t_{d+2})\right.$$
$$+\bar{Q}_1(t_d+\delta)\ \bar{\Phi}^T(t_{d+2},t_{d+1})\bar{M}^T(t_{d+2})$$
$$\left.-K(t_d+\delta,t_{d+1})\bar{M}^T(t_{d+1})P(t_{d+1}/t_d)\ \bar{\Phi}^T(t_{d+2},t_{d+1})\bar{M}^T(t_{d+2})\right\}x$$
$$\left\{\bar{M}(t_{d+2})P(t_{d+2}/t_{d+1})\bar{M}^T(t_{d+2})+\bar{R}(t_{d+2})\right\}^{-1} \qquad (199)$$

with $K(t_d+\delta,t_{d+1})$ given by Eq. (198).

To calculate the optimum estimate $\hat{x}(t_d+\delta/t_n)$ in Eq. (181) with delay up to (n-d) sample periods, it is necessary to determine (n-d) weighting matrices $K(t_d+\delta,t_k)$. The weighting matrices $K(t_d+\delta,t_{d+1})$ and $K(t_d+\delta,t_{d+2})$ have already been determined from Eq. (195) as given by Eqs. (197) and (199) respectively. The matrix $K(t_d+\delta,t_{d+3})$ is obtained from Eq. (195) with n equal to 3, and this iterative procedure is continued until all the (n-d) weighting matrices have been obtained.

These calculations will be rather tedious if the optimum estimate has been delayed for many sample points and there are a large number of weighting matrices to be determined. When the optimum estimate is obtained from a stationary dynamic system (as discussed in Sec. III-E), these weighting matrices will not change from one sample point to the next; thus, the tedious calculations must be performed only once.

These ideas are illustrated in Sec. IV-D, where examples are worked out for the optimum estimate with a delay of one sample period and with a delay of two sample periods. For a delay of up to one sample period the covariance of the error $P(t_d+\delta/t_{d+1})$ is given by Eq. (175), with $t_n+\alpha$ and t_n replaced by $t_d+\delta$ and t_{d+1} respectively, so that

$$P(t_d+\delta/t_{d+1}) = E[x(t_d+\delta)x^T(t_d+\delta)]-E[\hat{x}(t_d+\delta/t_d)x^T(t_d+\delta)]$$
$$-K(t_d+\delta,t_{d+1})\bar{M}(t_{d+1})\left\{E[x(t_{d+1})x^T(t_{d+}\delta)]\right.$$
$$\left.-E[\hat{x}(t_{d+1}/t_d)x^T(t_d+\delta)]\right\}. \qquad (200)$$

By using Eq. (185) the covariance $E[x(t_d+\delta)x^T(t_d+\delta)]$ can be written

$$E[x(t_d+\delta)x^T(t_d+\delta)] = \bar{\Phi}(t_d+\delta,t_d)E[x(t_d)x^T(t_d)]\,\bar{\Phi}^T(t_d+\delta,t_d)+\bar{Q}_2(t_d+\delta), \qquad (201)$$

where the expected value of the cross terms is zero, and the substitution has been made of the $r \times r$ matrix $\bar{Q}_2(t_d+\delta)$ defined as

$$\bar{Q}_2(t_d+\delta) = E[\bar{\Phi}(t_d+\delta,t_d)x(t_d)x^T(t_d)\bar{\Phi}^T(t_d+\delta,t_d)]+E[u(t_d+\delta,t_d)u^T(t_d+\delta,t_d)] \qquad (202)$$

Both the quantity $\bar{Q}_2(t_d+\delta)$ in Eq. (202) and the quantity $\bar{Q}_1(t_d+\delta)$ defined in Eq. (188) are equal to zero when $t_d + \delta$ is equal to t_d. Substituting Eqs. (187) and (201) into Eq. (199), and using the covariance $P(t_d/t_d)$ for the equivalent quantity given by Eq. (191) yields as the covariance matrix of the error $P(t_d+\delta/t_{d+1})$

$$P(t_d+\delta/t_{d+1}) = \bar{\Phi}(t_d+\delta,t_d)P(t_d/t_d)\,\bar{\Phi}^T(t_d+\delta,t_d) + \bar{Q}_2(t_d+\delta)$$
$$- K(t_d+\delta,t_{d+1})\bar{M}(t_{d+1})\Big\{\bar{\Phi}(t_{d+1},t_d)P(t_d/t_d)\,\bar{\Phi}^T(t_d+\delta,t_d)$$
$$+ \bar{Q}_1^T(t_d+\delta)\Big\}\ , \qquad (203)$$

where $K(t_d+\delta,t_{d+1})$ is given by Eq. (197), and $Q_1(t_d+\delta)$ and $Q_2(t_d+\delta)$ are defined by Eqs. (188) and (202) respectively.

For a delay of up to two sample periods the expression for the covariance of the error $P(t_d+\delta/t_{d+2})$ can be derived from Eq. (176) by the same procedure used to derive $P(t_d+\delta/t_{d+1})$ from Eq. (176).

$$P(t_d+\delta/t_{d+2}) = P(t_d+\delta/t_{d+1}) - K(t_d+\delta,t_{d+2})\bar{M}(t_{d+2})\Big\{E[x(t_{d+2})x^T(t_d+\delta)]$$
$$- E[\hat{x}(t_{d+2}/t_{d+1})x^T(t_d+\delta)]\Big\}\ . \qquad (204)$$

Everything in Eq. (204) has already been determined except for the quantity in braces. From Eq. (162) the optimum estimate $\hat{x}(t_{d+2}/t_{d+1})$ can be written

$$\hat{x}(t_{d+2}/t_{d+1}) = \bar{\Phi}(t_{d+2},t_{d+1})\hat{x}(t_{d+1}/t_{d+1})$$
$$= \bar{\Phi}(t_{d+2},t_{d+1})\Big[\hat{x}(t_{d+1}/t_d) + K(t_{d+1})\Big\{y(t_{d+1}) - \bar{M}(t_{d+1})\hat{x}(t_{d+1}/t_d)\Big\}\Big] \qquad (205)$$

Substituting Eq. (205) into the quantity in brackets in Eq. (204) gives

$$\Big\{E[x(t_{d+2})x^T(t_d+\delta)] - E[\hat{x}(t_{d+2}/t_{d+1})x^T(t_d+\delta)]\Big\}$$
$$= \bar{\Phi}(t_{d+2},t_{d+1})[I - K(t_{d+1})\bar{M}(t_{d+1})]\ x$$
$$\Big\{E[x(t_{d+1})x^T(t_d+\delta)] - E[\hat{x}(t_{d+1}/t_d)x^T(t_d+\delta)]\Big\}\ . \qquad (206)$$

Substituting Eqs. (206) and (187) into Eq. (204) gives as the covariance of the error

$$P(t_d+\delta/t_{d+2}) = P(t_d+\delta/t_{d+1})-K(t_d+\delta,t_{d+2})\overline{M}(t_{d+2})\,\overline{\Phi}(t_{d+2},t_{d+1})\,x$$

$$[I-K(t_{d+1})\overline{M}(t_{d+1})][\,\overline{\Phi}(t_{d+1},t_d)P(t_d/t_d)\,\overline{\Phi}^T(t_d+\delta,t_d)+\overline{Q}_1^T(t_d+\delta)], \tag{207}$$

where I is the identity matrix and $P(t_d+\delta/t_{d+1})$ and $K(t_d+\delta,t_{d+2})$ are given by Eqs. (203) and (199) respectively.

For a delay of up to $(n-d)$ sample periods the general expression for the covariance matrix of the error is given by Eq. (176), and, by repeating the same procedure used to obtain Eq. (205), this covariance $P(t_d+\delta/t_n)$ can be written

$$P(t_d+\delta/t_n) = P(t_d+\delta/t_{n-1})+K(t_d+\delta,t_n)\prod_{k=d+1}^{n-1}\overline{\Phi}(t_{k+1},t_k)[I-K(t_k)\overline{M}(t_k)]\,x$$

$$[\,\overline{\Phi}(t_{d+1},t_d)P(t_d/t_d)\,\overline{\Phi}^T(t_d+\delta,t_d)+\overline{Q}_1^T(t_d+\delta)]. \tag{208}$$

Therefore, the covariance of the error $P(t_d+\delta/t_n)$ can be calculated by starting with Eq. (203) and using Eq. (208) iteratively.

In conclusion, when considering estimation with delay, it must be decided if the improvement in the estimate is worth the delay in receiving the estimate and the increased complexity of the optimum filter. The examples in the following section will serve to illustrate some of the ideas of this chapter.

D. EXAMPLES

The procedure for optimum interpolation with delay developed in this chapter is intimately related to the technique for optimum filtering and prediction derived in Chapter III. In order to clarify this relationship, the examples in this section will be concerned with the same two stationary stochastic processes that were used in Section III-F. to illustrate the ideas of Chapter III. For the first-order Markov process, the optimum interpolation will be determined for a delay of one sample period and two sample periods. For the stationary process with

exponential cosine autocorrelation, the optimum interpolation will be determined for a delay of one sample period. With a delay of a whole number of sample periods δ is equal to zero, and from Eqs. (188) and (202),

$$\bar{Q}_1(t_n+\delta)\Big|_{\delta=0} = 0$$

$$\bar{Q}_2(t_n+\delta)\Big|_{\delta=0} = 0. \tag{209}$$

1. Exponential Autocorrelation

The stationary first-order Markov process has the autocorrelation

$$\Re_{xx}(\tau) = e^{-\beta|\tau|} \tag{210}$$

The subscript 1 will not be used on the variables in this example, so the expected value of the transition matrix will be denoted $\bar{\phi}$ with

$$\bar{\phi} = E[\exp(-\beta T_n)] \tag{211}$$

and the variance of the multiplicative noise will be denoted by $\bar{r}$.

The model of the optimum filter for interpolation with a delay of both one and two sample periods is presented in Fig. 10. That part of the filter within the dashed rectangle duplicates the model of the optimum filter for prediction presented in Fig. 6 and discussed in Section III-F. The weighting $k(t_d,t_{d+1})$ is a scalar which is determined from Eq. (197) to be

$$k(t_d,t_{d+1}) = \frac{\bar{\phi}\, p(t_d/t_d)}{p(t_{d+1}/t_d)+\bar{r}} \tag{212}$$

The weighting $k(t_d,t_{d+2})$ is determined from Eq. (199) as

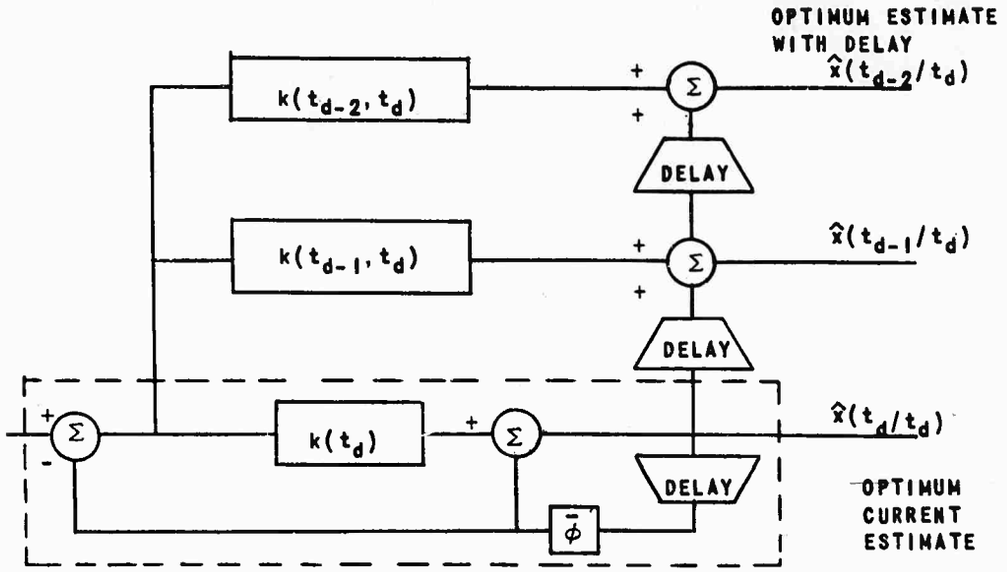


FIG. 10. MODEL OF THE OPTIMUM FILTER FOR ESTIMATION WITH DELAY FOR EXAMPLE 1.

$$k(t_d,t_{d+2}) = \frac{\bar{\phi}^2 p(t_d/t_d)}{p(t_{d+2}/t_{d+1})+\bar{r}} - \frac{k(t_d,t_{d+1})\,\bar{\phi} p(t_{d+1}/t_d)}{p(t_{d+2}/t_{d+1})+\bar{r}} \qquad (213)$$

and after substituting Eq. (212) for $k(t_d,t_{d+1})$ in Eq. (213) this can be written

$$k(t_d,t_{d+2}) = \frac{\bar{\phi}^2 p(t_d/t_d)}{p(t_{d+2}/t_{d+1})+\bar{r}} \left[1 - \frac{p(t_{d+1}/t_d)}{p(t_{d+1}/t_d)+\bar{r}} \right] . \qquad (214)$$

The covariance $p(t_{n+1}/t_n)$ and $p(t_n/t_n)$ are given by Eqs. (144) and (145) as

$$p(t_{n+1}/t_n) = \frac{\bar{\phi}^2\,\bar{r}}{p(t_n/t_{n-1})+\bar{r}} + 1 - \bar{\phi}^2$$

$$p(t_n/t_n) = \frac{-\bar{r}}{p(t_n/t_{n-1})+\bar{r}} . \qquad (215)$$

The steady-state value of the covariance of the error with a delay of one sample period $\bar{p}(t_{n-1}/t_n)$ is determined from Eq. (203) after substituting Eq. (212) for $k(t_d/t_{d+1})$ and the result is

$$\bar{p}(t_{n-1}/t_n) = \bar{p}(t_n/t_n)\left[1 - \frac{\bar{\phi}^2\bar{p}(t_n/t_n)}{\bar{p}(t_{n+1}/t_n)+\bar{r}}\right] . \tag{216}$$

The steady-state value of the covariance of the error with a delay of two sample periods $\bar{p}(t_{n-2}/t_n)$ is determined from Eqs. (207) and (216) after substituting Eq. (143) for $k(t_d)$ and Eq. (214) for $k(t_d/t_{d+2})$ and the result is

$$\bar{p}(t_{n-2}/t_n) = \bar{p}(t_n/t_n)\left[1 - \frac{\bar{\phi}^2\bar{p}(t_n/t_n)}{\bar{p}(t_{n+1}/t_n)+\bar{r}} - \frac{\bar{\phi}^4\bar{p}(t_n/t_n)^2\bar{r}^2}{(\bar{p}(t_{n+1}/t_n)+\bar{r})^3}\right] . \tag{217}$$

The steady-state value of the three covariances $\bar{p}(t_n/t_n)$, $\bar{p}(t_{n-1}/t_n)$, and $\bar{p}(t_{n-2}/t_n)$ are plotted in Fig. 11 as a function of the variance of the multiplicative noise $\bar{r}$ for

$$\bar{\phi} = 0.9091, \tag{218}$$

which is the expected value of the transition matrix when the sample period has an exponential distribution (purely random sampling) with the parameters given by Eq. (219).

$$\begin{aligned} \mu &= 1.0 \\ \beta &= 0.1. \end{aligned} \tag{219}$$

Notice how delaying the estimate decreases the covariance of the error.

2. Exponential Cosine Autocorrelation

The second example is the stationary process with exponential cosine autocorrelation

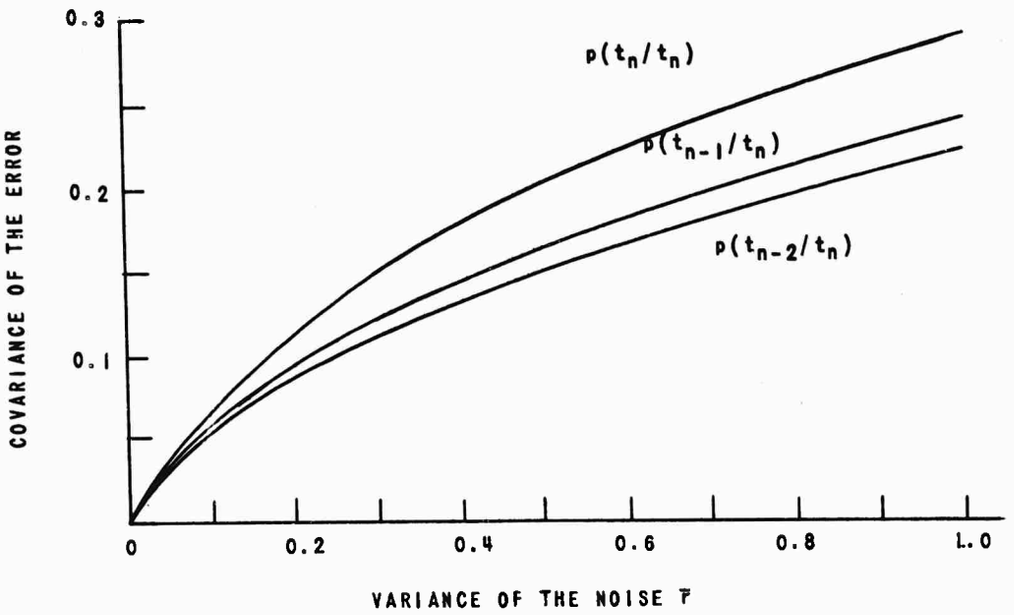


FIG. 11. COVARIANCE OF THE ERROR FOR EXAMPLE 1.

$$\mathcal{R}_{x_1x_1}(\tau) = e^{-\beta|\tau|}\left(\cos\gamma\tau - \frac{\beta\gamma\,\sin\gamma|\tau|}{(2\beta^2+\gamma^2)}\right) \tag{220}$$

which was discussed in Sect. II-C and III-F. For a delay of one sample period, the weighting matrix is determined from Eq. (197) to be

$$K(t_d/t_{d+1}) = \begin{bmatrix} \dfrac{\bar{\phi}_{11}p_{11}(t_d/t_d) + \bar{\phi}_{12}p_{12}(t_d/t_d)}{p_{11}(t_{d+1}/t_d)+\bar{r}} \\ \dfrac{\bar{\phi}_{11}p_{21}(t_d/t_d) + \bar{\phi}_{12}p_{22}(t_d/t_d)}{p_{11}(t_{d+1}/t_d)+\bar{r}} \end{bmatrix} . \tag{220}$$

The elements of the covariance matrix $P(t_d/t_d)$ can be determined from Eq. (127) as

$$p_{11}(t_d/t_d) = \frac{p_{11}(t_{d+1}/t_d)\bar{r}}{p_{11}(t_{d+1}/t_d)+\bar{r}}$$

$$p_{12}(t_d/t_d) = p_{21}(t_d/t_d) = \frac{p_{12}(t_{d+1}/t_d)\bar{r}}{p_{11}(t_{d+1}/t_d)+\bar{r}}$$

$$p_{22}(t_d/t_d) = p_{22}(t_{d+1}/t_d) - \frac{p_{12}(t_{d+1}/t_d)p_{21}(t_{d+1}/t_d)}{p_{11}(t_{d+1}/t_d)+\bar{r}} . \qquad (221)$$

The steady-state covariance of the error $\bar{p}_{11}(t_{n-1}/t_n)$ for a delay of one sample period can be obtained in terms of the elements of the covariance matrix $\bar{P}(t_{n+1}/t_n)$ by substituting Eqs. (220) and (221) into Eq. (203), and the result is

$$\bar{p}_{11}(t_{n-1}/t_n) = \frac{\bar{p}_{11}(t_{n+1}/t_n)\bar{r}}{\bar{p}_{11}(t_{n+1}/t_n)+\bar{r}} - \frac{\bar{r}^2}{[\bar{p}_{11}(t_{n+1}/t_n)+\bar{r}]^2} \times$$

$$\left\{ \sum_{i=1}^{2}\sum_{j=1}^{2} \bar{\phi}_{1i}\,\bar{\phi}_{1j}\,p_{1i}(t_{n+1}/t_n)p_{1j}(t_{n+1}/t_n) \right\} . \qquad (222)$$

The steady-state value of the two covariances $\bar{p}_{11}(t_n/t_n)$ and $\bar{p}_{11}(t_{n-1}/t_n)$ are plotted in Fig. 12 as a function of the multiplicative noise when the sample period has a geometric distribution (purely random sampling) with the parameters given below, in Eq. (223)

$$\begin{aligned} \mu &= 1.0 \\ \beta &= 0.1 \\ \gamma &= 1.0, \end{aligned} \qquad (223)$$

which means that the values of the matrices are given by Table 4 in Sect. III-F. Delaying the estimate for one sample period decreases the covariance of the error, but not quite as much as in the previous example.

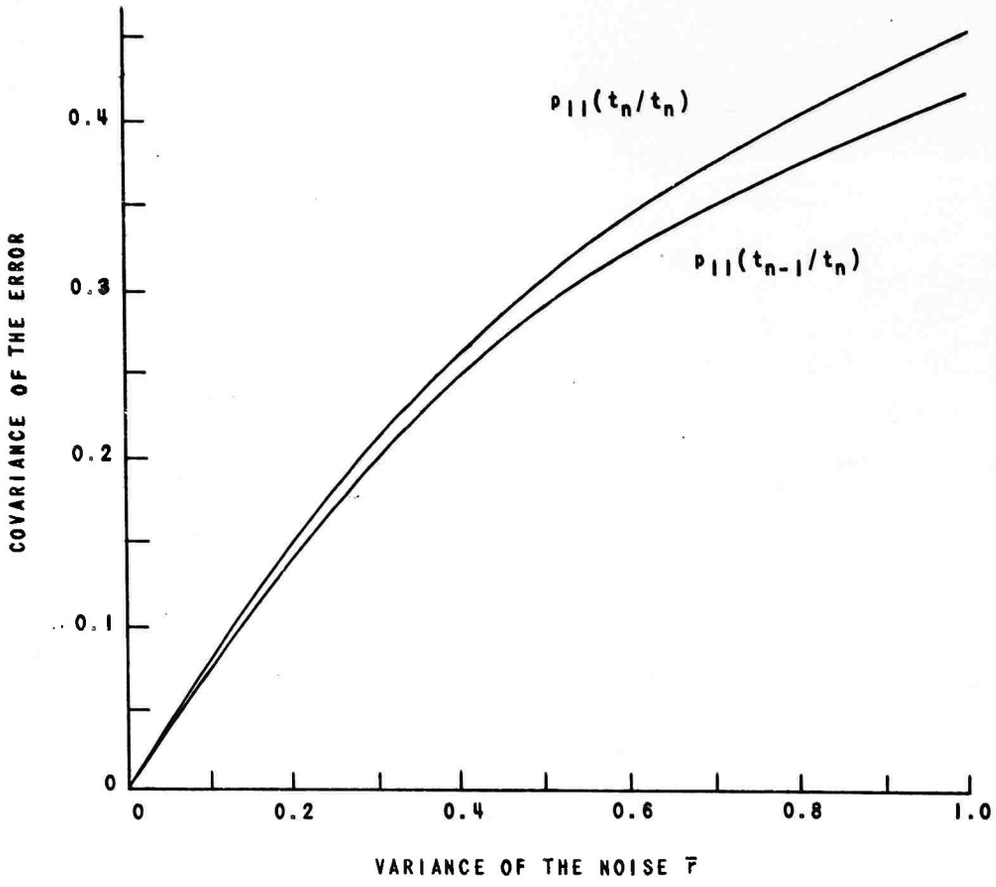


FIG. 12. COVARIANCE OF THE ERROR FOR EXAMPLE 2.

V. STATIONARY STOCHASTIC PROCESSES WITH RANDOM SAMPLE PERIODS

A. INTRODUCTION

The purpose of this chapter is to show how the techniques developed in this investigation can be applied to a problem which has been considered in the literature. A stationary stochastic process with a rational power spectrum is sampled so that the sample period is a stationary random variable with a known first-order probability distribution. There is no multiplicative noise; therefore, the only random parameter of the process is the randomness in the sample period. It is assumed that the sample points are being observed in real time; therefore, although the sample period T_n is a random parameter, the value of the random parameter T_n can be determined exactly as soon as the sample points t_n and t_{n+1} are observed.

In Sect. V-B, the continuous optimum estimate of the process is presented when the sample period can be considered as a known time-varying parameter. This optimum estimate is generated by a linear dynamic filter with a time-varying matrix-valued gain. If the optimum estimate can be delayed until an on-line computer iteratively determines the gain at each sample point, then the optimum estimate can be implemented.

When it is not practical to recompute the gain at each sample point, the linear filter that generates the optimum estimate is modified so that the matrix-valued gain is a constant. The constant value chosen is that value which minimizes the trace of the covariance matrix of the error in the limit as the number of observations n approaches infinity.

In the approach to the problem that has been considered in the literature [Refs. 13,14] transform techniques are used to design the linear time-invariant filter that minimizes the mean square error of the continuous estimate. In Sect. V-C this time-invariant filter is compared with the optimum filter for the case in which the stochastic process is a stationary first-order Markov process.

B. THE OPTIMUM ESTIMATE

When the sample period of a stationary stochastic process is a time-varying parameter that is known exactly, the optimum current estimate is given by Eq. (119) which can be written

$$\hat{x}(t_n/t_n) = \Phi(T_{n-1})\hat{x}(t_{n-1}/t_{n-1})+K(t_n)[y(t_n)-M\,\Phi(T_{n-1})\hat{x}(t_{n-1}/t_{n-1})] \quad (224)$$

where the transition matrix $\Phi(\tau)$ is given by Eq. (17) as

$$\Phi(\tau) = e^{F\tau}. \quad (225)$$

The matrix block diagram in Fig. 13 represents the linear filter that implements the optimum estimate in Eq. (224). That part of the filter within the dashed rectangle duplicates the model of the process and has the impulse response $\Phi(\tau)$. The switch on the left, representing the sampling operation, and the synchronous switch in the feedback loop of

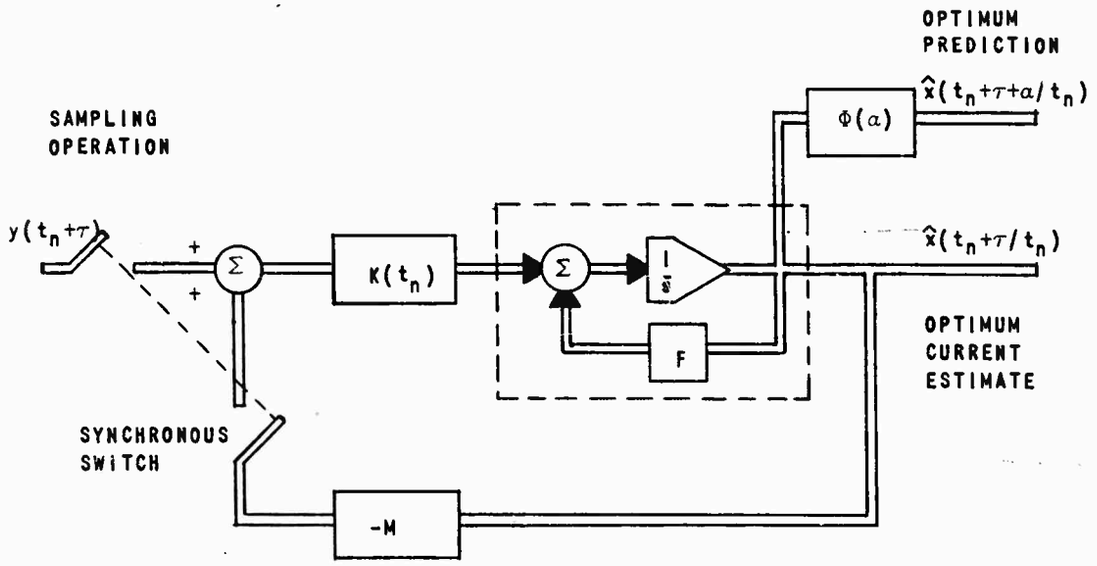


FIG. 13. MODEL OF THE OPTIMUM FILTER.

the filter close simultaneously, so that the sum of their outputs is zero except at a sample point t_n when the sum is an impulse with area

$$y(t_n) - M\Phi(T_{n-1})\hat{x}(t_{n-1}/t_{n-1}). \tag{226}$$

The weighting matrix $K(t_n)$ is determined from Eq. (120), which is

$$K(t_n) = P(t_n/t_{n-1})M^T\left\{M\,P(t_n/t_{n-1})\,M^T\right\}^{-1}. \tag{227}$$

The covariance $P(t_n/t_{n-1})$ is calculated iteratively from the nonlinear difference equation (125), which can be written

$$P(t_{n+1}/t_n) = \Phi(T_n)\left\{I - P(t_n/t_{n-1})M^T[M\,P(t_n/t_{n-1})M^T]^{-1}\,M\right\} \quad x$$
$$P(t_n/t_{n-1})\,\Phi^T(T_n) + \int_0^{T_n} \Phi(T_n-\sigma)GVG^T\,\Phi^T(T_n-\sigma)d\sigma, \tag{228}$$

where the substitution has been made from Eq. (48) that

$$Q(t_n) = \int_0^{T_n} \Phi(T_n-\sigma)GVG^T\,\Phi^T(T_n-\sigma)d\sigma. \tag{229}$$

From Eqs. (227) and (228) the covariance of the error of the optimum estimate during any time in the sample period is

$$P(t_n+\tau/t_n) = \Phi(\tau)\left\{I - K(t_n)M\right\}P(t_n/t_{n-1})\,\Phi^T(\tau)$$
$$+\int_0^{\tau} \Phi(\tau-\sigma)\,GVG^T\,\Phi^T(\tau-\sigma)\,d\sigma. \tag{230}$$

The weighting $K(t_n)$ for the optimum estimate in Eq. (227) can be determined in real time by an on-line computer if a slight delay is permitted in the estimate. The following procedure must be repeated at each sample point:

1. Determine the sample period T_n as soon as the sample point t_{n+1} is observed.

2. Calculate and store the covariance of the error $P(t_{n+1}/t_n)$ from Eq. (228) using the known sample period T_n and the stored covariance $P(t_n/t_{n-1})$.

3. Calculate the weighting matrix $K(t_{n+1})$ and set the gain of the filter accordingly.

The idea of using an on-line computer to calculate the weighting matrix at each sample point did not originate from this investigation. In particular, it follows directly from the work of Kalman [Ref. 1] when his results are carefully applied to this problem. The advantage to this procedure is that the optimum estimate is obtained, but the corresponding disadvantage is that an on-line computer must be available, and the optimum estimate must be delayed until the computer has completed the necessary calculations to determine the weighting matrix.

Sometimes it is impractical to recalculate the weighting matrix at each sample point. It would be desirable to find some constant value for the matrix that would minimize the average value of the trace of the covariance matrix of the error. The constant value of the matrix which minimizes the average value of the trace in the limit as the number of observations n approaches infinity can be determined by the techniques developed in Chapter III.

Assume that the sample period T_n is an unknown random parameter, rather than a known time-varying parameter. The series of weighting matrices $\bar{K}(t_n)$ can be determined to minimize the average value of the trace of the covariance matrix of the error. In Sect. III-E it was shown that when all the statistics of the processes were stationary, then the optimum estimate would become a stationary dynamic system in the limit. This means that in the limit the weighting matrices $\bar{K}(t_n)$ will approach the desired constant value $\bar{K}$.

The weighting matrix $\bar{K}(t_n)$ is determined from Eq. (227); thus

$$\bar{K}(t_n) = \bar{P}(t_n/t_{n-1})\, M^T \left\{ M\, \bar{P}(t_n/t_{n-1})\, M^T \right\}^{-1}. \qquad (231)$$

The average value of the covariance $\bar{P}(t_n/t_{n-1})$ is calculated iteratively from the average value of Eq. (228) which can be written

$$\bar{P}(t_{n+1}/t_n) = E[\Phi(T_n)]\left\{I - \bar{P}(t_n/t_{n-1})\, M^T [M\, \bar{P}(t_n/t_{n-1})\, M^T]^{-1} M\right\} \times$$

$$\bar{P}(t_n/t_{n-1}) E[\Phi^T(T_n)] + E[\tilde{\Phi}(T_n) \bar{P}(t_n/t_{n-1}) \tilde{\Phi}^T(T_n)] + E[Q(t_n)]; \tag{232}$$

where the covariance $Q(t_n)$ is given by Eq. (229) and all the expectations in Eq. (232) are with respect to the sample period T_n. The limiting value of the covariance

$$\lim_{n\to\infty} \bar{P}(t_{n+1}/t_n) = \lim_{n\to\infty} \bar{P}(t_n/t_{n-1}) = \bar{P} \tag{233}$$

can be obtained by iterating Eq. (232) until a steady-state solution is obtained, or it can be obtained directly from the steady-state solution of Eq. (232). The desired constant value of the weighting matrix $\bar{K}$ is obtained by substituting the limiting value of the covariance $\bar{P}$ into Eq. (231).

The filter in Fig. 13 with a constant gain $\bar{K}$ is not the optimum filter, but it is the best filter from a sub-optimum class of filters. In the following section the filter in Fig. 13 is compared with another filter from a sub-optimum class of filters--the linear time-invariant filter.

C. COMPARISON WITH THE TIME-INVARIANT FILTER

When it is desired to design the time-invariant filter with infinite memory that minimizes the mean square error of the estimate, transform techniques can be used to get a solution to the problem [Refs. 13, 14]. In the transform approach, the spectral density of the power spectrum of the sampled stochastic process is first determined from a complex convolution integral, which can be evaluated by the method of residues in some cases. Then, the synthesis procedure is based upon the standard Wiener spectral factorization of the sampled power spectrum. The resulting filter is continuous and time-invariant and gives the minimum mean square error for all time of any linear filter that is time invariant.

Even with a constant gain, the filter discussed in the previous section and shown in Fig. 13 is not time invariant. It contains a synchronous switch that closes at the sample point, so that the characteristics of the filter depend upon the sample period. A simple example will serve to show the difference between the two filters. The performance of these two filters will be compared with that of a pure stretcher or time varying hold.

Consider the stationary first-order Markov process discussed in Sect. II-B with the autocorrelation function

$$\mathcal{R}_{xx}(\tau) = e^{-\beta|\tau|}.$$

The process is sampled, the sample period is a stationary random variable, and it is desired to reconstruct the original process from the observed random variables in order to minimize the mean-square error.

The optimum filter for this process is presented in Fig. 14. The gain of the filter is unity for any probability distribution of the sample period. The filter is not time invariant because of the synchronous switch. The output of the filter is

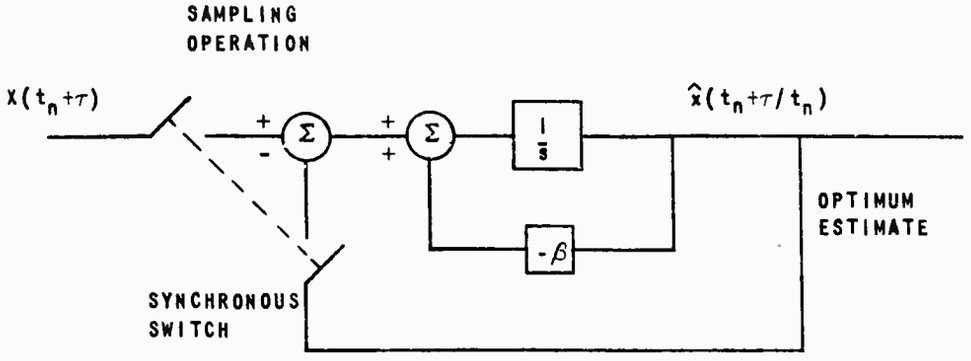


FIG. 14. OPTIMUM FILTER.

$$\hat{x}(t_n+\tau) = e^{-\beta\tau}x(t_n) \qquad t_n \le t_n + \tau < t_{n+1}, \tag{234}$$

so that at each sample point the error in the estimate is zero. During the sample period, the error is due to the random input to the process. The mean-squared error due to the covariance of the random input is obtained from the one-dimensional version of Eq. (229), which is

$$p(t_n+\tau/t_n) = \int_0^{\tau} e^{-\beta(\tau-\sigma)}\, 2\beta e^{-\beta(\tau-\sigma)} d\sigma = 1 - e^{2\beta\tau} \qquad t_n \le t_n+\tau < t_{n+1} \tag{235}$$

where the substitution has been made from Eq. (9) that

$$v_{11} = 2\beta \tag{236}$$

From Eq. (235) the average over all time of the covariance of the error (or the mean-squared error) is

$$E\left[\int_0^{T_n} p(t_n+\tau/t_n)d\tau\right] \Big/ E[T_n] = E\left[\int_0^{T_n} (1-e^{-2\beta\tau})d\tau\right] \Big/ E[T_n]$$

$$= 1 - \frac{1 - E[e^{-2\beta T_n}]}{2\beta\, E[T_n]} \tag{237}$$

The linear time-invariant filter for the stationary first-order Markov process is presented in Fig. 15. The impulse response of the filter $g(\tau)$ that minimizes the mean-squared error is dependent on the probability distribution of the sample period so that a different $g(\tau)$ must be determined for each distribution. The best time-invariant filter and the corresponding mean-squared error have been determined in closed form for two probability distributions of the sample period--the geometric and exponential distributions. The probability laws governing these two distributions have already been discussed in Sect. II-E, and the pertinent information concerning these laws is summarized in Table 1.

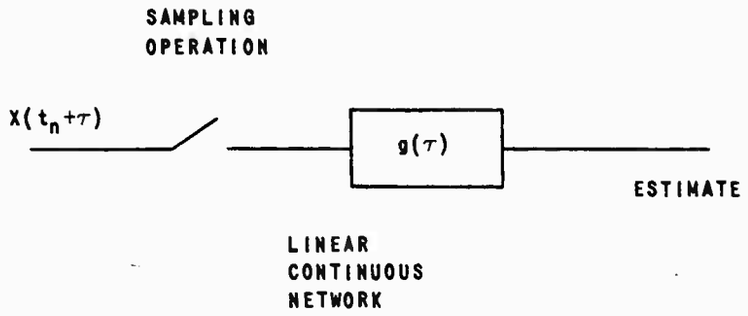


FIG. 15. TIME-INVARIANT FILTER.

For the geometric distribution of the sample period (periodic sampling with independent misses) the mean-squared error of the time-invariant filter is given by Eq. (238)

$$\text{Mean-squared error} = 1 - \frac{(1-\sigma)(e^{2\beta T}-1)}{2\beta T}, \tag{238}$$

where σ is

$$\sigma = \frac{1}{2}(1-e^{-2\beta T})(1 - \frac{qT}{p}) + \frac{1}{2}\sqrt{(1-e^{-2\beta T})^2(1 - \frac{qT}{p})^2 + 4\,\frac{qT}{p}(1-e^{-2\beta T})} \tag{239}$$

and p, q, and T are parameters of the distribution.

For the exponential distribution of the sample period (purely random sampling) the mean-squared error of the time-invariant filter is given by Eq. (240).

$$\text{Mean-squared error} = \frac{-\beta + \sqrt{\beta^2 + 2\beta\mu}}{\mu} \tag{240}$$

where μ is a parameter of the distribution.

It is interesting to compare the optimum filter and the time invariant filter with a simple time varying hold. The output of the time varying hold is

$$\hat{x}(t_n+\tau) = x(t_n) \qquad t_n \leq t_n + \tau < t_{n+1}$$

The mean squared error of an estimate at time $t_n+\tau$ would be

$$2(1 - e^{-2\beta\tau})$$

and the average over all time of the mean squared error is

$$E\left[\int_0^{T_n} 2(1-e^{-2\beta\tau})d\tau\right] \Big/ E[T_n]$$

$$= 2 - 2\left(1 - E[e^{-\beta T_n}]\right) \Big/ \beta\, E[T_n]$$

In Table 6 the mean squared error of these three filters is compared for period sampling, periodic sampling with independent misses, and purely random sampling for β equal to one-tenth. The expected value of the sample period is normalized to unity and the value of the other parameters of the probability distribution are listed in Table 6.

TABLE 6. MEAN-SQUARE ERROR FOR OPTIMUM AND TIME-INVARIANT FILTERS.

Purely Random or Poisson Sampling			
Periodic Sampling with Independent Misses			
Periodic Sampling			
Probability Distribution of Sample Period	Constant	Geometric	Exponential
Parameters of Probability Distribution	$\bar{T}=1$	$T=1/2$ $p=q=1/2$	$\mu=1$
Mean Sample Period	$\bar{T}=1$	$T/p=1$	$1/\mu=1$
Variance of Sample Period	0	1/2	1
With $\beta=0.1$ Mean-Square Error for Optimum Filter	0.0937	0.1311	0.1667
With $\beta=0.1$ Mean-Square Error for Time-Invariant Filter	0.0937	0.2458	0.3583
With $\beta=0.1$ Mean-Square Error for Time Varying Hold	0.0967	0.1400	0.1818

For periodic sampling the optimum filter and the time invariant filter are identical, but as the variance of the sample period increases, the mean squared error of the time invariant filter increases much more rapidly than that of the optimum filter. The mean squared error of the simple time-varying hold is only slightly greater than that of the optimum filter for periodic sampling and also when the sample period is a random variable. In this example, the time invariant filter does not approximate the true optimum very well for large variance of the sample period, while the simple time-varying hold does approximate the true optimum. This shows that sometimes a "non-optimum" time-varying filter is better than the best time invariant filter.

VI. CONCLUSIONS

A. SUMMARY

In this investigation the general solution has been derived for the problem of the optimum linear estimation of a sampled stochastic process with random parameters that can be adequately approximated by the model presented in Chapter II. The random parameters are independent from one sample point to the next with known mean and covariance. The solution can be implemented by a linear dynamic system with a matrix-valued gain (or gains) calculated iteratively for each sample point. For high-order complex systems these computations are most easily performed by a digital computer.

B. SUGGESTIONS FOR FUTURE WORK

The ideas presented in Chapter IV represent the first thorough investigation of optimum interpolation with delay for a nonstationary sampled stochastic process. As yet, no completely satisfactory theory exists for the optimum linear interpolation with delay for a nonstationary continuous process. It should be possible to extend the technique developed in Chapter IV to the interpolation with delay of the continuous stochastic process with white noise added to the measurements. For filtering and prediction, this problem was formulated by Kalman and Bucy [Ref. 18].

Another possibility is concerned with the area of adaptive systems. The optimum estimate in this investigation is implemented by a linear dynamic system that uses the expected value of the random parameters. The only place the covariance of the random parameters is needed is in the iterative calculation of the matrix-valued gain. If only the expected value of the random parameters were known, then the techniques of adaptive systems could be used to adjust the matrix-valued gain to react to measured variations in the covariance of the random parameters.

REFERENCES

1. Kalman, R. E., "A new approach to linear filtering and prediction problems", Journal of Basic Engineering, Trans. ASME, Series D, 82, 1960.

2. Bendat, J. S., Principles and Applications of Random Noise Theory, John Wiley, New York, 1958.

3. Sherman, S., "Non-mean-square error criteria", IRE Trans. on Information Theory, IT-4, 1958.

4. Gunckel, T. L., "Optimum design of sampled-data systems with random parameters", Stanford Electronics Laboratories, Technical Report No. 2102-2, April 24, 1961.

5. Wiener, N., The Extrapolation, Interpolation, and Smoothing of Stationary Time Series, John Wiley, New York, 1949.

6. Franklin, G. F., "The optimum synthesis of sampled-data systems", Doctoral Dissertation, Department of Electrical Engineering, Columbia University, 1955.

7. Amara, R. C., "The linear least squares synthesis of continuous and sampled data multivariable systems", Stanford Electronics Laboratories, Technical Report No. 40, July 28, 1958.

8. Rosenbloom, A., J. Heilfron, D. L. Trautman, "Analysis of linear systems with randomly varying inputs and parameters", IRE Convention Record, Part 4, 1955.

9. Bergen, A. R., "Stability of systems with randomly time-varying parameters", IRE Trans. on Automatic Control, AC-5, 4, September 1960.

10. Kalman, R. E., "Analysis and synthesis of linear systems operating on randomly sampled data", Doctoral Dissertation, Dept. of Electrical Engineering, Columbia University, August 10, 1957.

11. Kalman, R. E. and R. W. Koepcke, "Optimal synthesis of linear sampling control systems using generalized performance indexes", ASME Trans., Paper No. 58-IRD-6, 1958, p. 1821.

12. Kalman, R. E., "On the general theory of control systems", Proc. of the First IFAC Moscow Congress, 1960.

13. Bergen, A. R., "The synthesis of optimum random sampling systems", Technical Report T-T/133, Electronics Research Laboratories, Columbia University, 1957.

REFERENCES (Continued)

14. Hsieh, H. C., "On the optimum synthesis of random sampling multipole filters with stationary inputs", Report No. 61-8, University of California, Department of Engineering, Los Angeles, California.

15. Beutler, F. J., "A generalization of Wiener optimum filtering and prediction", Doctoral dissertation, California Institute of Technology, Pasadena, California.

16. Beutler, F. J., "Prediction and filtering for random parameter systems", IRE Trans. on Information Theory, IT-4, December 1958.

17. Shaw, L. G., "Dual mode filtering", Stanford Electronics Laboratories, Technical Report No. 2106-2, September 26, 1960.

18. Kalman, R. E. and R. S. Bucy, "New results in linear filtering and prediction theory", Journal of Basic Engineering, Trans. ASME, Series D, 83, 1961.

19. Davenport, W. B. and W. L. Root, An Introduction to the Theory of Random Signals and Noise, McGraw-Hill, New York, 1958.

20. Coddington, E. A. and N. Levinson, Theory of Ordinary Differential Equations, McGraw-Hill, New York, 1955.

21. Gantmacher, F. R., The Theory of Matrices, Vol. 1, Chelsea, New York, 1959.

22. Friedman, B., Principles and Techniques of Applied Mathematics, John Wiley, New York, 1956.

23. Parzen, E., Modern Probability Theory and its Application, John Wiley, New York, 1960.

24. Doob, J. L., Stochastic Processes, John Wiley, New York, 1956.

25. Feller, W., An Introduction to Probability Theory and its Applications, Vol. 1, John Wiley, New York, 1950.

26. Greville, T. N. E., "The pseudoinverse of a rectangular or singular matrix and its application to the solution of systems of linear equations", SIAM Review, 2, 1, January 1960.

BASIC DISTRIBUTION LIST
December 1961

Evans Signal Laboratory
Belmar, New Jersey
1 Attn: Head, Solid-State Dev. Br. Dr. H. Jacobs

Commanding General, USASRDL
5 Ft. Monmouth, New Jersey
1 Attn: SIGRA/SL-SC

Chief, West Coast Office
1 Pasadena 2, California

Commanding Officer, Frankford
1 Philadelphia, Pa.

Ballistics Research Laboratory
Aberdeen Proving Grd, Md.
2 Attn: L.A. Delsasso

Commanding General
Ft. Huachuca, Arizona
1 Attn: Technical Library

Chief of Naval Research
Washington 25, D.C.
2 Attn: Code 427
1 Attn: Code 463

Commanding Officer, ONR Br.Off
1 San Francisco 9, California

Chief Scientist, ONR BR. Offic
1 Pasadena, California

Comm. Officer, ONR Br. Office
1 Chicago 1, Illinois

Comm. Officer, ONR Br. Office
1 New York 13, New York

Officer-in-Charge, ONR
15 New York, New York

Director, NR Laboratory
Washington 25, D.C.
1 Attn: Code 5240
1 Attn: Code 7100
1 Attn: Code 5430
1 Attn: Code 5200
1 Attn: Code 5300
1 Attn: Code 5400
1 Attn: Code 5270
1 Attn: Code 5210 GA

Chief, Bureau of Ships
Washington 25, D.C.
3 Attn: Code 810
1 Attn: Code 816
1 Attn: Code 820
1 Attn: Code 840

Chief, Bureau of Aeronautics
Washington 25, D.C.
1 Attn: Code RAAV-33

Chief, Bureau of Naval Weapons
Washington 25, D.C.
1 Attn: RREN
1 Attn: RAAV-44

Chief of Naval Operations
Washington 25, D.C.
1 Attn: Code Op-94T

Comm. Officer and Director
1 San Diego 52, California

U.S. Naval Post Graduate Sch.
Monterey, California
1 Attn: Tech. Rpts. Librarian

Commander, U.S.N.Missile Ctr.
Pt. Mugu, California
1 Attn: 5321

U.S. Naval Weapons Lab.
Dahlgren, Virginia
1 Attn: Tech. Library

Naval Ordnance Laboratory
Corona, California
1 Attn: Library

Director, NOL
1 White Oak, Maryland

Commander, NADC
Johnsville, Pa.
1 Attn: NADC Library

Dept. of the Army
Washington 25, D.C.
1 Attn: Research Support Div.

Chief, P&O Division
Washington 25, D.C.
1 Attn: SIGOP-5

Office of the Chief ofEngr.
1 Washington 25, D.C.

Hqs., USAF (AFRDR-NU-3)
Pentagon, Washington 25, D.C.
1 Attn: Mr. Harry Mulkey

Chief of Staff, USAF
Washington 25, D.C.
2 Attn: AFDRT-ER

Commanding Officer
Aeronautical Systems Division
Wright-Patterson AFB,Ohio
1 Attn: ASRNEB
1 Attn: ASRNRS-2

Executive Director
AF Office of Scientific Res.
Washington 25, D.C.
1 Attn: Code SRYA

AF Special Wespons Ctr.
Kirtland AFB, New Mexico
2 Attn: SWOI

Director, AUL
Maxwell AFB, Alabama
1 Attn: CR 4582

Commander
Patrict AFB, Florida
1 Attn: AFMTC Tech.Lib.,MU-135

Commanding General
Rome, New York
2 Attn: RCRW

Commanding General
Bedford, Massachusetts
1 Attn: CRTOTT-2, Electronics

AF Office of Scientific Res.
Washington 25, D.C.
1 Attn: Tech. Library

AF Systems Command
Washington 25, D.C.
1 Attn: RDRLT,Capt.Zimmerman
1 Attn: RDRESE-Mr. Presnell

Hqa. ARDC
Washington 25, D.C.
1 Attn: RDRS
1 Attn: RDRC

Assistsnt Sec. of Defense
Washington 25, D.C.
1 Attn: Tech. Library

Director of Defense
1 Washington 25, D.C.

U.S. Coast Guard
Washington 25, D.C.
1 Attn: EEE

Dept. of Defense
Washington 25, D.C.
1 Attn: Code 711, Fred Miller

Advisory Grp on Electron Tubes
1 New York 13, New York

Adv. Grp on Reliability of Elec. Equipment
1 Washington 25, D.C.

Commanding Officer
Diamond Ordnance Fuze Labs.
Washington 25, D.C.
2 Attn: ORDTL 430, Dr. R. Young

U.S. Dept. of Commerce
1 Boulder, Colorado
2 Attn: Miss V. Lincoln

Director, Nat. Security Agcy
Ft. George G. Meade, Md.
1 Attn: REMP-2

California Inst. of Tech.
Pasadena, California
1 Attn: Prof. L.M. Field

California Inst. of Tech
Jet Propulsion Laboratory
Pasadena, California
1 Attn: Documents Library

University of California
Berkeley 4, California
1 Attn: Prof. R.M. Saunders,CHm

U. Of California, Radiation L.
Berkeley, California
1 Attn: Dr. R. K. Wakerling

Univ. of California
Los Angeles 24, California
1 Attn: C.T. Leondes

Georgia Institute of Tech.
Atlanta, Georgia
1 Attn: Mrs. J.H. Crosland, Lib.

Harvard University
Cambridge 38, Massachusetts
1 Attn: Mrs. E. Farkas, Lib.

Illinois Inst. of Tech.
Chicago 16, Illinois
1 Attn: Dr. Paul C. Yuen
1 Attn: Dr. George I. Cohn

University of Illinois
Urbana, Illinois
1 Attn: Prof. D. Alpert
1 Attn: Paul Coleman, EE Dept.
1 Attn: Library Serials Dept.

Johns Hopkins University
Baltimore 2, Maryland
1 Attn: Librarian

Johns Hopkins University
Silver Spring, Maryland
1 Attn: Document Library

Mass. Inst. of Technology
Cambridge 39, Mass.
1 Attn: Research Lab. of Elec.

Mass. Inst. of Tech.
Lincoln Lab
1 Lexington 73, Mass.

Director, Cooley Elecs. Lab.
1 Ann Arbor, Mich.

Univ. of Michigan
Ann Arbor, Michigan
1 Attn: Tech. Documents Serv.

University of New Mexico
Albuquerque, New Mexico
1 Attn: R.K. Moore, Chm.,

Univ. of Nevada
Reno, Nevada
1 Attn: Dr. Robt. Manhart

Northwestern University
Evanston, Illinois
1 Attn: Walter S. Toth

Ohio State University
Columbus 10, Ohio
1 Attn: Prof. E.M. Boone

Oregon State College
Corvallis, Oregon
1 Attn: H.J. Oorthuys

Purdue University
Lafayette, Indiana
1 Attn: Library

University of Rochester
Rochester 20, New York
1 Attn: Dr. Gerald H. Cohen

Stanford Research Institute
Menlo Park, California
1 Attn: Documents Center

Villanova University
Villanova, Pa.
1 Attn: Thomas C. Gabriele

Yale University
New Haven, Connecticut
1 Sloane Physics Laboratory
1 Dept. of Elec. Engr.
1 Dunham Laboratory, Engr. Lib.

University of Puerto Rico
Mayaguez, Puerto Rico
1 Attn: Dr. Braulio Dueno

University of Saskatchewan
Saskatoon, Canada
1 Attn: Prof.N.F.Moody

Uppsala University
Uppaala, Sweden
1 Attn: Dr. P.A. Tove

Admiral Corporation
Chicago 47, Illinois
1 Attn: N. Robertson, Lib.

Airborne Instruments Lab.
Deer Pk, L.I., New York
1 Attn: J. Dyer, VP

Bell Telephone Labs.
Murray Hill, New Jersey
1 Attn: Elecs. Res. Dept.

Bomac Laboratories
Beverly, Mass.
1 Attn: Wellesley J. Dodds

Columbia Radiation Lab.
1 New York 27, New York

Cook Research Labs.
1 Morton Grove, Illinois

Cornell Aeronautical Labs.,Inc
1 Buffalo 21, New York

General Electric TWT Product Section
Palo Alto, California
1 Attn: Technical Library

General Electric Company
Auburn, New York
1 Attn: Mr. Gerald C. Huth

General Electric Company
Schenectady, New York
1 Attn: G.E. Feiker
1 Attn: E.D. McArthur

Gilfillan Brothers
Los Angeles, California
1 Attn: Engineering Library

Hughes Aircraft Company
1 Culver City, California

Huggins Laboratories
1 Sunnyvale, California

ITT Federal Laboratories
Nutley 10, New Jersey
1 Attn: A.W. McEwan

LEL, Inc.
Copiague, L.I., New York
1 Attn: Mr. Robt. S. Mautner

Lockheed Aircraft Corporation
Sunnyvale, California
1 Attn: G.W. Price, Subsystem
1 Attn: Dr. Harris, Mgr.

The Maxson Elecs. Corporation
New York 1, New York
1 Attn: M. Simpson

MITRE Corporation
Bedford, Mass.
1 Attn: Constance Willis, Lib.

Monsanto Corporation
St. Louis 66, Missouri
1 Attn: E. Orban, Inorganic Development

National Company, Inc.
Maiden 48, Mass.
1 Attn: Technical Librarian

Nortronics Tech. Info. Agency
1 Hawthorne, California

Polarad Corporation
Long Island City 1, New York
1 Attn: A.H. Sonnenschein
Chief Systems Engr.

RCA Laboratories
Princeton, New Jersey
1 Attn: Dr. W.M. Webster
1 Attn: Harwick Johnson

Ramo-Wooldridge Corporation
Los Angeles 45, California
1 Attn: Dorothy K. Goss, Lib.

The Rand Corporation
Santa Monica, California
1 Attn: Margaret Anderson, Lib.

Raytheon Company
Burlington, Massachusetts
1 Attn: Librarian

Raytheon Manufacturing Co.
Waltham, Massachusetts
1 Attn: Librarian

Roger White Electron Devices
1 Haskell, New Jersey

STL Technical Library
Space Tech. Labs. Inc.
1 Los Angeles 45, California

Sperry Rand Corporation
1 Gainesville, Florida

Commanding Officer
1 Mountain View, California

Technical Research Group
1 Syosset, L.I., New York

Texas Instruments Inc.
1 Dallas, Texas
1 Attn: Dr. W. Adcock
1 Attn: M.E. Chun

Varian Associates
Palo Alto, California
1 Attn: Technical Library

Westinghouse Electric Corp.
Baltimore 3, Maryland
1 Attn: G. Ross Kilgore, Mgr.
Applied Res. Dept.
Baltimore Laboratory