NSG-334
1 M.I.T.

WESCON/66
TECHNICAL PAPERS

Presented at the Western Electronic Show and Convention, August 23-26, 1966

SESSION 7

3 The Application of State-Variable Techniques to Communication and Radar Problems

N67-29901 (ACCESSION NUMBER)
61 (PAGES)
CR-76277 (NASA CR OR TMX OR AD NUMBER)

N67-29906 (THRU)
1 (CODE)
07 (CATEGORY)

FACILITY FORM

THE APPLICATION OF STATE-VARIABLE TECHNIQUES TO COMMUNICATION AND RADAR PROBLEMS

Session Organizer & Session Chairman
Harry L. Van Trees
Dept. of Electrical Engineering & Research Laboratory of Electronics
Massachusetts Institute of Technology
Cambridge, Massachusetts

DETECTION AND CONTINUOUS ESTIMATION: THE FUNDAMENTAL ROLE OF THE OPTIMUM REALIZABLE LINEAR FILTER*

H. L. Van Trees
Department of Electrical Engineering and Research Laboratory of Electronics
Massachusetts Institute of Technology, Cambridge, Mass. 02139

I. INTRODUCTION

The importance of the differential equation (or state variable) technique of representing dynamic systems for optimal control problems is well-known. In the communications area the technique appears useful in two classes of problems.

(i) In the first class, the optimum receiver may be either a linear or nonlinear system. It contains as a component a linear filter which is a solution to the familiar Wiener-Hopf equation.

(ii) In the second class, the fundamental problem may be one of signal design or message shaping. The problem is formulated in terms of some equivalent optimal control problem which may then be solved.

In this paper, we discuss briefly the first class of problem. The discussion is tutorial in nature. Typical problems in the second class are discussed in refs. 1 and 2.

In section 2, we discuss the estimator equations. In section 3, we study the application to detection problems. In section 4, we study nonlinear modulation problems.

II. ESTIMATOR EQUATIONS

The basic linear estimation problem is

$$r(u) = c(u)\,a(u) + n(u) \qquad T_i \le u \le T_f \quad (1)$$

where $a(u)$ and $n(u)$ are sample functions from independent zero-mean, Gaussian random processes with known statistical properties and $c(t)$ is a deterministic carrier. The desired signal is denoted by $d(t)$. The optimum estimate is obtained by passing $r(u)$, $T_i \le u \le T_f$ through a continuous linear filter. We denote the output of this linear device as $\hat{d}(t)$.

We choose the linear processor so that the mean-square error,

$$\xi(t) \triangleq E\left[(\hat{d}(t) - d(t))^2\right], \qquad (2)$$

is minimized.

If

$$d(t) = a(t) \qquad (3)$$

then we are trying to estimate the message.

Three special cases arise:

(i) $t > T_f$ prediction,

(ii) $t = T_f$ realizable filtering,

(iii) $t < T_f$ filtering with delay.

There are two alternative ways of solving the problem. The first method characterizes the processes in terms of their covariance functions and the linear processor in terms of a time-varying impulse response. We will label this method the impulse response method.

2.1 Impulse Response Method

We write,

$$\hat{d}(t) = \int_{T_i}^{T_f} h(t,u)\, r(u)\, du. \qquad (4)$$

The process statistics are described by,

$$\begin{aligned} K_a(t,u) &\triangleq E\left[a(t)\, a(u)\right] \\ K_n(t,u) &\triangleq E\left[n(t)\, n(u)\right] \end{aligned} \qquad (5)$$

and
$$\begin{aligned} K_r(t,u) &\triangleq E\left[r(t)\, r(u)\right] \\ &= c(t)\, K_a(t,u)\, c(u) + K_n(t,u). \end{aligned}$$

Then, the optimum linear filter must satisfy the integral equation (e.g., ref. 3)

$$\int_{T_i}^{T_f} h_o(t,u)\, K_r(u,z)\, du = K_{dr}(t,z) \qquad (6)$$

$$T_i \le z < T_f$$
$$T_i \le t \le T_f$$

The optimum filter is shown in Fig. 1.

The error using the optimum filter is

$$\xi_o(t) = K_d(t,t) - \int_{T_i}^{T_f} h_o(t,\tau)\, K_{dr}(t,\tau)\, d\tau. \qquad (7)$$

* This work was supported principally by the National Aeronautics and Space Administration (Grant NsG-334).

There are a number of special cases that arise frequently in practice. The pertinent equations are summarized in Table 1.

A second approach might be termed the differential equation (or state-variable) method. We will confine our discussion to the realizable filtering problem. The prediction problem is a trivial modification (e.g., ref. 4). The filtering with delay is quite involved (e.g., refs. 5 or 6).

2.2 State-Variable Method

Here we characterize the message and noise processes in terms of the vector-differential equation describing the linear system which would generate them if it were excited by "white" noise.

To characterize the message we write

$$\dot{\underline{x}}(t) = \underline{F}(t)\,\underline{x}(t) + \underline{G}(t)\,u(t) \tag{8}$$

where $\underline{x}(t)$ is the state vector $(n \times 1)$.

A canonic message generator is shown in Fig. 2. The double lines denote a vector path. The matrices $\underline{F}(t)$ and $\underline{G}(t)$ describe the system dynamics. The forcing function $u(t)$ is a white noise input used to generate $a(t)$. The actual message $a(t)$ is some linear combination of the state variables.

$$E\left[u(t)\,u(\tau)\right] = q\,\delta(t-\tau). \tag{9}$$

The observation (or modulation) equation describes how $\underline{x}(t)$ is transmitted.

Table 1. Summary of Equations.

Case	Assumptions	Integral Equation	Description
1	(a) $d(t) = a(t)$ (b) $T_i \le t \le T_f$	$\int_{T_i}^{T_f} h_o(t,u)\,K_r(u,z)\,du = K_a(t,z)$ $T_i \le z \le T_f$ $T_i \le t \le T_f$	Interior point estimator Fixed endpoint
2	$d(t) = a(t)$ $t = T_f$	$\int_{T_i}^{t} h_o(t,u)\,K_r(u,z)\,du = K_a(t,z)$ $T_i \le z < T_f$	Realizable point estimator
3	$d(t) = a(t)$ $t = T_f$ $K_n(t,u) = \frac{N_o}{2}\,\delta(t-u)$	$\frac{N_o}{2} h_o(t,z) + \int_{T_i}^{t} h_o(t,u)\,K_a(u,z)\,du = K_a(t,z)$ $T_i \le z < t$	Realizable point estimator in white noise
4	$d(\tau) = a(\tau)$	$\int_{T_i}^{t} h_o(\tau,u)\,K_r(u,z)\,du = K_a(\tau,z)$ $T_i \le z < t$ $T_i \le \tau \le t$	Interior point estimator Variable endpoint
5	$K_n(t,u) = \frac{N_o}{2}\,\delta(t-u)$ $d(\tau) = a(\tau)$	$\frac{N_o}{2} h_o(\tau,z:t) + \int_{T_i}^{t} h_o(\tau,u:t)\,K_a(u,z)\,du = K_a(t,z)$ $T_i \le z < t$ $T_i \le \tau \le t$	Interior point estimator White noise Variable endpoint

$$r(u) = \underline{c}(u)\,\underline{x}(u) + w(u) \qquad T_i \le u \le t \quad (10)$$

To make eqs. (10) and (1) agree, we require

$$c(u)a(u) = \underline{c}(u)\,\underline{x}(u). \qquad (11)$$

Frequently, $a(u)$ is the first component of the state vector.

Then

$$\underline{c}(u) = \left[c(u) \,\vdots\, 0 \,\vdots\, 0 \cdots \vdots\, 0\right]. \qquad (12)$$

The additive noise $w(t)$ is assumed to be a sample function from a "white" zero-mean Gaussian process of spectral height $N_0/2$ (double-sided). If there is a colored noise component, it is included in an augmented state vector.

The optimum estimate $\hat{d}(t)$ is described in terms of a differential equation whose forcing term is $r(t)$. In the cases of interest, $\hat{d}(t)$ can be expressed as a linear combination of the state variables.

$$d(t) = \sum_{i=1}^{n} d_i(t)\,x_i(t) \triangleq \underline{D}(t)\,\underline{x}(t). \qquad (13)$$

Since minimum mean-square estimation commutes over linear transformations, we estimate the state variables and use

$$\hat{d}(t) = \sum_{i=1}^{n} d_i(t)\,\hat{x}_i(t) \triangleq \underline{D}(t)\,\hat{\underline{x}}(t). \qquad (14)$$

The differential equation describing the optimum estimate of the state vector is: (This result is due to Kalman and Bucy, ref. 4)

$$\dot{\hat{\underline{x}}}(t) = \underline{F}(t)\,\underline{x}(t) + \underline{z}(t)\left[r(t) - \underline{c}(t)\,\underline{x}(t)\right] \quad (15)$$

We see that the equation has the same structure as the message generation equation with the following associations

$$\underline{z}(t) \leftrightsquigarrow \underline{G}(t) \qquad (16)$$

$$r(t) - \underline{c}(t)\,\underline{x}(t) \leftrightsquigarrow u(t) \qquad (17)$$

The optimum estimator is shown in Fig. 3.

The matrix $\underline{z}(t)$ is specified by the <u>gain equation</u>,

$$\underline{z}(t) = \frac{2}{N_0}\underline{\xi}_p(t)\,\underline{c}^T(t) \qquad (18)$$

where $\underline{\xi}_p(t)$ is the error covariance matrix in estimating the state variables. It satisfies the equation,

$$\dot{\underline{\xi}}_p(t) = \underline{F}(t)\,\underline{\xi}_p(t) + \underline{\xi}_p(t)\,\underline{F}^T(t)$$
$$- \frac{2}{N_0}\underline{\xi}_p(t)\,\underline{c}^T(t)\,\underline{c}(t)\,\underline{\xi}_p(t) + q\,\underline{G}(t)\,\underline{G}^T(t). \quad (19)$$

The last equation is a nonlinear equation which is commonly referred to as the variance equation.

For the case in which $d(t) = a(t)$, the system in Fig. 3 is identical to the system in Fig. 1 under the assumptions of case 3.

The obvious advantage of the state-variable approach is that the required functions are easily computable.

We now look at two classes of communications problems in which the linear filters shown in Figs. 1 and 3 play a fundamental role.

III. DETECTION PROBLEM

We consider two common detection problems.

3.1 Known Signal in Colored Noise and White Noise

In the binary hypothesis case,

$$\begin{aligned} H_1 &: r(t) = s(t) + n_c(t) + w(t) \qquad 0 \le t \le T \\ H_0 &: r(t) = \phantom{s(t) +{}} n_c(t) + w(t) \qquad 0 \le t \le T \end{aligned} \qquad (20)$$

The signal $s(t)$ is a known deterministic function. The noise $n_c(t)$ is a zero-mean colored Gaussian noise with a square-integrable covariance function $K_c(t,u)$. The noise $w(t)$ is a zero-mean white Gaussian noise with a covariance function $\frac{N_0}{2}\,\delta(t-u)$.

As is well known, the optimum receiver performs a likelihood ratio test.

Using the impulse response approach,

$$L(T) = \frac{2}{N_0}\int_0^T s(\tau)\,r(\tau)\,d\tau$$
$$- \frac{2}{N_0}\int_0^T\!\!\int_0^T s(u)\,h_0(u,z)\,r(z)\,dz\,du \;\overset{H_1}{\underset{H_0}{\gtrless}}\; \gamma \quad (21)$$

where $h_0(u,z)$ satisfies the equation for case 3 in Table 1 with the associations

$$n_c(t) \sim a(t)$$
$$K_c(t,u) \sim K_a(t,u). \tag{22}$$

This can be simplified by defining,

$$g(\tau) = \frac{2}{N_0}\int_0^T s(u)\left[\delta(\tau-u) - h_0(u,\tau)\right]du. \tag{23}$$

Then

$$L(T) = \int_0^T g(\tau)\,r(\tau)\,d\tau \tag{24}$$

and the receiver is the correlator shown in Fig. 4.

We now demonstrate how one translates this result into a structure containing the realizable linear filters of Figs. 1 and 3. The technique used carries over to a large number of interesting cases.

We start by writing,

$$L(T) = \int_0^T \frac{dL(t)}{dt}\,dt \tag{25}$$

where

$$L(t) = \frac{2}{N_0}\int_0^t s(\tau)\,r(\tau)\,d\tau - \frac{2}{N_0}\int_0^t\!\!\int_0^t s(u)\,h_0(u,z:t)\,r(z)\,dz\,du \tag{26}$$

and $h_0(u,z:t)$ is the optimum filter (case 5 in Table 1.

Differentiating eq. (26) with respect to t, we have,

$$\frac{\partial L(t)}{\partial t} = \frac{2}{N_0}s(t)\,r(t) - \frac{2}{N_0}s(t)\int_0^t h_0(t,z:t)\,r(z)\,dz - \frac{2}{N_0}\int_0^t s(u)\,du\left\{h_0(u,t:t)\,r(t) + \int_0^t \frac{\partial h_0(u,z:t)}{\partial t}\,r(z)\,dz\right\}. \tag{27}$$

One can show easily that,

$$\frac{\partial h_0(u,z:t)}{\partial t} = -h_0(u,t:t)\,h_0(t,z:t) \tag{28}$$

(where $h_0(\bullet,\bullet:t)$ is given by case 5 in Table 1 and is symmetric in its first two arguments).

Re-arranging eq. (27), we have,

$$\frac{\partial L(t)}{\partial t} = \frac{2}{N_0}\left\{s(t) - \int_0^t h_0(t,z:t)\,s(z)\,dz\right\}\left\{r(t) - \int_0^t h_0(t,z:t)\,r(z)\,dz\right\}. \tag{29}$$

The resulting receiver has the simple form shown in Fig. 5. (Observe that this particular structure could also have been obtained using a "whitening" argument, e.g., ref. 7.)

The performance follows easily,

$$d^2 = \frac{2}{N_0}\int_0^T dt\left\{s(t) - \int_0^t h_0(t,z)\,s(z)\,dz\right\}^2 \tag{30}$$

A second case of interest arises when the signal component is random.

3.2 Gaussian Signals in Gaussian Noise

In the simplest binary hypothesis case,

$$\begin{aligned} H_1: r(t) &= s_{\Omega_1}(t) + w(t) \qquad T_i \le t \le T_f \\ H_0: r(t) &= w(t) \qquad\qquad\quad\; T_i \le t \le T_f \end{aligned} \tag{31}$$

Here $s_{\Omega_1}(t)$ is a sample function from a zero-mean Gaussian random process.

The likelihood ratio test is:

$$L = \int_0^T\!\!\int_0^T d\tau\,du\,r(\tau)\,h_0(\tau,u)\,r(u) \underset{H_0}{\overset{H_1}{\gtrless}} \gamma \tag{32}$$

Proceeding as in section 3.1, we obtain

$$L = \int_0^T\left[2r(t)\hat{s}_0(t)\,dt - \hat{s}_0^2(t)\right]dt \underset{H_0}{\overset{H_1}{\gtrless}} \gamma \tag{33}$$

where

$$\hat{s}_0(t) = \int_0^t h_0(t,\tau:t)\,r(z)\,dz. \tag{34}$$

(This result is due to Schweppe, ref. 8.)

The receiver structure is shown in Fig. 6. Once again we can find all of the point estimators using the state variable approach.

We now turn to another class of communications problems.

IV. ESTIMATION OF CONTINUOUS WAVEFORMS: NONLINEAR MODULATIONS

The continuous estimation problem is:

$$r(u) = s[u:\theta(u)] + w(u) \qquad -\infty < \mu < t \quad (35)$$

where $\theta(u)$ is a sample function from a Gaussian random process. The transmitted signal $s[u:\theta(u)]$ depends on $\theta(u)$ in a deterministic nonlinear no-memory manner.

In order to be explicit we will confine our comments to the angle modulation case.

Then,

$$s[u:\theta(u)] = \sqrt{2}\,\sin[w_c u + \theta(u)]. \quad (36)$$

The function $\theta(u)$ is related to the message $a(\tau)$ through some linear operation.

For example, in FM,

$$s[u:a(\tau)] = \sqrt{2}\,\sin\left[w_c u + \int_{-\infty}^{u} a(\tau)\,d\tau\right]. \quad (37)$$

Then

$$s[u:\theta(u)] \triangleq \sqrt{2}\,\sin[w_c u + \theta(u)]. \quad (38)$$

If $\underline{\theta}(u)$ is the state vector of $\theta(u)$ process, then, in general,

$$\dot{\underline{\theta}}(u) = \underline{F}_a(u)\,\underline{\theta}(u) + \underline{G}_a(u)\,a(u). \quad (39)$$

The actual phase $\theta(u)$ is the first component of this vector process.

Now let $\underline{x}_m(t)$ be the state vector of the message process.

$$\dot{\underline{x}}_m(t) = \underline{F}_m(t)\,\underline{x}_m(t) + \underline{G}_m(t)\,u(t). \quad (40)$$

Then, the total state vector of concern is

$$\underline{x}(t) = \begin{bmatrix} \underline{\theta}(t) \\ \text{------} \\ \underline{x}_m(t) \end{bmatrix} \quad (41)$$

Then one can show that an approximation to the optimum demodulator is the phase-lock loop shown in Fig. 7. The loop filter is simply precisely the same as the loop filter in Fig. 3.

There are several obvious advantages to the state-variable formulation.

(i) The actual nonlinear demodulator is constrained to be in the form of a closed loop system. We are trying to design the optimum loop filter. This is what state-variable approach gives automatically.

(ii) There are two quantities that we are trying to estimate simultaneously. The first is the phase angle. This estimate is needed to keep the loop operating in its linear region as long as possible. The second desired quantity is the message. Since it is a linear combination of the state variables and we automatically estimate all of the state variables, the message estimate comes out as a by-product of the loop filter design.

Similar arguments can be applied to other nonlinear modulation schemes.

V. SUMMARY

In this paper we have outlined the fundamental role of the optimum linear filter in several interesting communication problems. Since state-variable techniques frequently provide the most efficient method for finding the optimum linear filter, their importance in the above problems is obvious.

Our discussion has been very brief. A detailed development of the ideas is contained in ref. 9.

REFERENCES

1. J. K. Omura, "Signal Optimization for Additive Noise Channels with Feedback," WESCON Paper 7.4; Aug. 1966.
2. F. C. Schweppe, "A Modern Systems Approach to Signal Design," WESCON Paper 7.5; Aug. 1966.
3. L. A. Zadeh and J. R. Ragazzini, "Optimum Filters for the Detection of Signals in Noise," Proc. IRE, vol. 40, no. 10, pp. 1123-1131; Oct. 1952.
4. R. E. Kalman and R. Bucy, "New Results in Linear Filtering and Prediction Theory," ASME Jour. of Basic Engr., Vol. 83, pp. 95-108; March 1961.
5. A. E. Bryson and M. Frazier, "Smoothing for Linear and Nonlinear Dynamic Systems," USAF Tech. Rpt. ASD-TDR-63-119; Feb. 1963.
6. A. Baggeroer, "Maximum a posteriori Interval Estimation," WESCON Paper 7.3; Aug. 1966.
7. L. D. Collins, "A Realizable Whitening Filter," Internal Memo. IM-LC-3, M.I.T.; April 1966.
8. F. C. Schweppe, "Evaluation of Likelihood Functions for Gaussian Signals," IEEE Trans. on Information Theory, vol. IT-11, no. 1, pp. 61-70; Jan. 1965.
9. H. L. Van Trees, Detection, Estimation and Modulation Theory, John Wiley and Sons, Inc., New York (to appear).

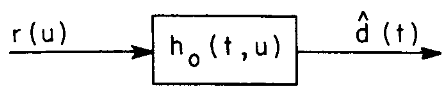


Fig. 1. Optimum filter, impulse response realization.

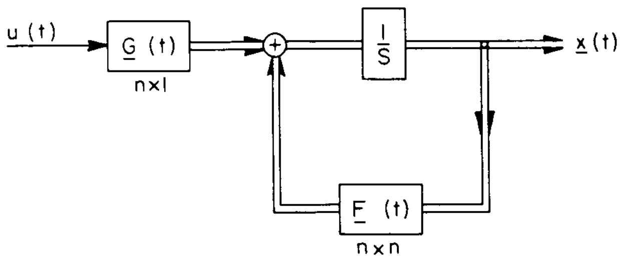


Fig. 2. Canonic message generator.

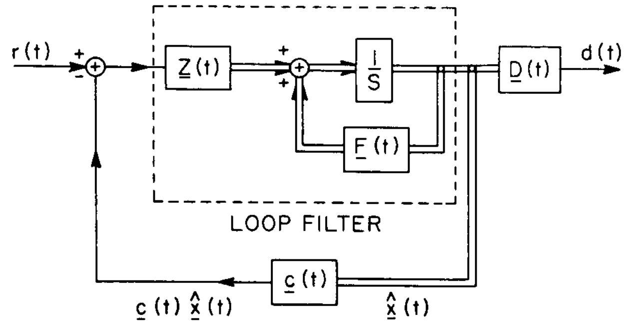


Fig. 3. Optimum estimator, state-variable realization.

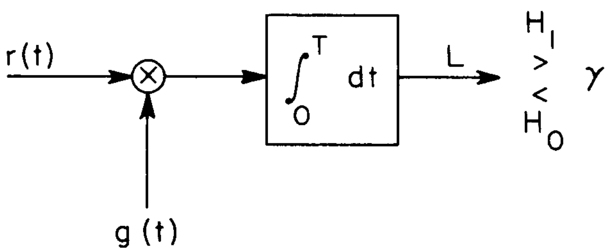


Fig. 4. Correlation receiver.

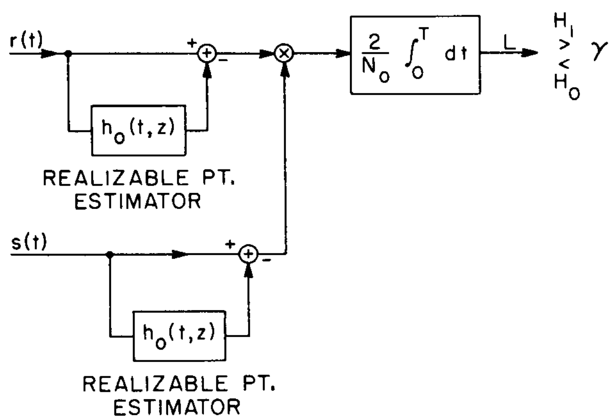


Fig. 5. Receiver for known signal in colored noise.

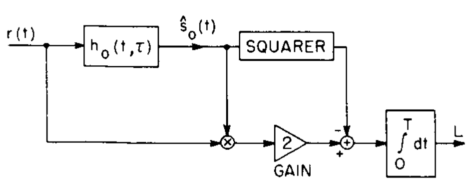


Fig. 6. Optimum realizable processor: Gaussian signal in Gaussian noise.

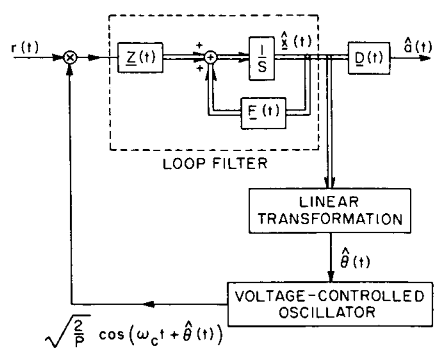


Fig. 7. Demodulator for an angle-modulation system.

A THEORY OF CONTINUOUS NONLINEAR RECURSIVE-FILTERING

WITH APPLICATION TO OPTIMUM ANALOG DEMODULATION*

Donald L. Snyder

Research Laboratory of Electronics

Massachusetts Institute of Technology

Cambridge, Massachusetts

ABSTRACT

A new approach is presented for the continuous nonlinear filtering or estimation problem. The approach is based on the use of Markov processes and state-variable concepts. Equations are derived for approximate minimum-mean-square-error estimates of a Markovian state vector observed in a signal in which it is imbedded nonlinearly. A general model for analog communication via randomly time-varying channels is defined and related to the state vector estimation problem. The model includes as special cases such linear and nonlinear modulation schemes as AM, PM, FM, and PM_n/PM; and such continuous channels as Rayleigh and Rician channels, fixed channels with memory and diversity channels. The approach leads automatically to physically realizable demodulators whose outputs are approximate MMSE estimates of the message and, if desired, the channel disturbances. Special consideration is given to PM and FM.

NOTATION

$\underline{v}(t)$	Lower-case, underscored letters denote column vectors
$v_i(t)$	The i^{th} component of $\underline{v}(t)$
$\frac{d}{dt}\underline{v}(t)$	A vector whose i^{th} component is $\frac{d}{dt} v_i(t)$
$\underline{M}(t)$	Capital, underscored letters denote matrices
$\underline{M}'(t)$	Transpose of $\underline{M}(t)$
$\underline{M}^{-1}(t)$	Inverse of $\underline{M}(t)$
$\underline{f}[t:\underline{v}(t)]$	A column vector whose components are nonlinear, no-memory, time-varying transformations of the vector $\underline{v}(t)$
$\underline{D}[\underline{f}(t:\underline{v})]$	The Jacobian matrix associated with $\underline{f}[t:\underline{v}(t)]$, the (i-row, j-column) element of the matrix is $\frac{\partial}{\partial v_i} f_j[t:\underline{v}(t)]$
$\hat{\underline{v}}(t)$	Circumflex denotes the exact minimum-mean-square-error estimate
$\underline{v}^*(t)$	Asterisk denotes the approximate minimum-mean-square estimate.
$\underline{v}_{t_0,t}$	Denotes the collection of waveforms $\{\underline{v}(\tau):t_0 \leq \tau \leq t\}$

*This work was supported principally by the National Aeronautics and Space Administration (Grant NSG-334).

INTRODUCTION

An approach is presented in this paper for continuously estimating a Markovian state vector based on a noisy observed signal in which it is imbedded nonlinearly. The approach can be applied in many diverse disciplines where the nonlinear filtering problem arises, so to present it in a general context, we shall first define an "Estimation Model" and associate with it the formal mathematical development of the theory. Applications are then made to analog communication theory. For this purpose, a broad "Communication Model" is defined for representing analog communication via randomly time-varying channels. It is a special case of the estimation model and can represent such linear-and nonlinear-modulation schemes as: AM, PM, FM, pre-emphasized FM, and PM_n/PM; and such continuous channels as: additive noise channels, Rayleigh and Rician channels, fixed channels with memory, and multi-link channels. Special consideration is given to angle-modulation schemes for which quasi-optimum demodulators are presented.

Discrete counterparts to the estimation model, or to special cases of it, have been studied by Wonham[1,2], Weaver[3], Cox[4,5], and Mowery[6]. Special cases of the continuous model have also been studied. Kalman and Bucy[6] examined the estimation of linearly transformed vector Gaussian processes. Bucy[8] examined the estimation of nonlinearly transformed one-dimensional Markov processes. Several related, not widely known studies have been made in the U.S.S.R.[9-17]. These, again, are generally for the case of nonlinearly transformed one-dimensional Markov processes. Applications to communication theory are given by Weaver and in the studies of the U.S.S.R. We shall study the estimation of nonlinearly transformed multidimensional Markov processes by a technique employing linearization and a conditional-mean argument.

Lehan and Parks[18], Youla[19], Van Trees[20-22], and Thomas and Wong[23], among others, have used an alternative approach to study communication models that are equivalent to special cases or our model. Their approach, called the MAP approach, is based on maximizing the suitably defined a posteriori probability density of a desired waveform. We shall indicate the relationship between the demodulators so obtained and ours. Recall that the MAP approach leads to an integral equation for the estimate and that the equation corresponds to a physically unrealizable demodulator. Van Trees[22] suggests making an approximation to the unrealizable demodulator for the purpose of implementation. It consists of a cascade of a nonlinear physically realizable demodulator and a linear physically unrealizable filter. On the other hand, the recursive-filtering approach leads automatically to a physically realizable demodulator. It is equivalent to the nonlinear physically realizable portion of the cascade approximation to the MAP demodulator.

THE ESTIMATION MODEL

The Estimation Model is shown in Fig. 1.

Let $\underline{x}(t)$ be a continuous m-dimensional vector Markov process described by the Ito stochastic differential equation:†

$$d\underline{x}(t) = \underline{f}[t:\underline{x}(t)]\,dt + d\underline{\chi}(t) \qquad (1)$$

where $\underline{f}[t:\underline{x}(t)]$ is an m-dimentional vector whose components are memoryless, nonlinear transformations of $\underline{x}(t)$ and $\underline{\chi}(t)$ is an m-dimensional vector whose components are Wiener processes. Let the covariance matrix associated with $\underline{\chi}(t)$ be:

$$E[\underline{\chi}(t)\underline{\chi}'(u)] = \underline{X}(t)\min(t,u) \qquad (2)$$

where $\underline{X}(t)$ is a symmetric, non-negative definite, m X m matrix. The elements of $\underline{X}(t)$, denoted by $X_{ij}(t)$, may be time-varying.

Observe that more than one vector process can be represented by (1) simply by adjoining the individual vectors to form $\underline{x}(t)$. Observe also that $\underline{x}(t)$ can have deterministic components (e.g., constant and time-varying parameters and signals), in which case the corresponding elements of $\underline{X}(t)$ are zero.

It is known (e.g., see Bharucha-Reid[25]) that the a priori probability density, $p(\underline{x};t)$, associated with the Markov process, $\underline{x}(t)$, defined by (1) satisfies the Fokker-Planck equation

---

† These equations were first given a rigorous interpretation by Ito[24]. A more recent discussion of his formulation is given in the engineering literature by Wonham[1,2]. Alternative interpretations based on new definitions for stochastic integrals have been suggested by Stratonvich[29] and Wong and Zakai[30]. We shall sometimes formally divide (1) by dt obtaining a white Gaussian process from the Wiener process. The interpretation will always be as (1).

$$\frac{\partial}{\partial t} p(\underline{x};t) = -\sum_{i=1}^{m} \frac{\partial}{\partial x_i}\left[f_i(t:\underline{x})p(\underline{x};t)\right] + \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m} X_{ij}(t)\frac{\partial^2}{\partial x_i \partial x_j} p(\underline{x};t) \qquad (3)$$

with the initial condition $p(\underline{x};t_o) = p(\underline{x}_o;t_o)$.

In the sequel we shall be interested in estimating scalar Gaussian processes with rational spectra. These processes can be represented in the form of (1) by letting $\underline{f}[t:\underline{x}(t)] = \underline{F}\ \underline{x}(t)$, where $\underline{F}$ is a time-invariant, m X m matrix. $\underline{X}(t) = \underline{X}$ must also be time invariant. By carefully choosing such a state representation, one component of $\underline{x}(t)$ can be made to correspond directly to the scalar process. Moreover, when several scalar processes are represented by adjoining their individual state vectors, each will correspond directly to one of the components of $\underline{x}(t)$. A particularly convenient state representation for scalar Gaussian processes is presented in the Appendix. We shall use this representation exclusively in the applications to follow.

We shall now define the noisy observed process. Let $\underline{y}(t)$ be a continuous, p-dimensional vector random process described by the Ito equation:

$$d\underline{y}(t) = \underline{g}[t:\underline{x}(t)]dt + d\underline{\eta}(t), \qquad (4)$$

where $\underline{g}[t:\underline{x}(t)]$ is a p-dimentional vector whose components are memoryless, nonlinear transformations of $\underline{x}(t)$ and $\underline{\eta}(t)$ is a p-dimensional vector whose components are Wiener processes. Let the covariance matrix associated with $\underline{\eta}(t)$ be given by

$$E\left[\underline{\eta}(t)\underline{\eta}'(u)\right] = \underline{N}(t)\min(t,u), \qquad (5)$$

where $\underline{N}(t)$ is a symmetric, positive-definite, p X p matrix. It is assumed that $\underline{N}^{-1}(t)$ exists; this implies that noise-free observations cannot be made. The elements of $\underline{N}(t)$, denoted by $N_{ij}(t)$, may be time-varying.

Simply for the convenience of notation, we assume that $\underline{\chi}(t)$ and $\underline{\eta}(t)$ are uncorrelated:

$$E\left[\underline{\chi}(t)\underline{\eta}'(u)\right] = 0.$$

Some of the statistics of $d\underline{y} = \underline{y}(t+dt) - \underline{y}(t)$ will be required later. We shall cite them here for convenience. Observe that to terms of order dt

$$E\left[d\underline{y}\,d\underline{y}'\right] = \underline{N}(t)dt, \qquad (6)$$

as can be demonstrated by using (4) and (5). Furthermore, all higher order moments of $d\underline{y}\ d\underline{y}'$ are of order greater than dt. This implies that $d\underline{y}\ d\underline{y}'/dt$ is essentially deterministic and equal to $\underline{N}(t)$ for dt vanishingly small. Thus, to terms of order dt

$$d\underline{y}\,d\underline{y}' = E\left[d\underline{y}\,d\underline{y}'\right] = \underline{N}(t)dt \text{ (dt infinitesimal)} \quad (7)$$

A more rigorous discussion justifying (7) is given by Kushner[31]. Eqs. (1) and (4) jointly define a continuous, (m+p)-dimensional, vector Markov process whose components are the combined components of $\underline{x}(t)$ and $\underline{y}(t)$. Formally dividing the equations by dt results in the more familiar looking expressions:

$$\frac{d}{dt}\underline{x}(t) = \underline{f}\left[t:\underline{x}(t)\right] + \underline{\xi}(t) \qquad (8)$$

and

$$\frac{d}{dt}\underline{y}(t) \equiv \underline{r}(t) = \underline{g}\left[t:\underline{x}(t)\right] + \underline{n}(t), \qquad (9)$$

where $\underline{\xi}(t) = d\underline{\chi}(t)/dt$ and $\underline{n}(t) = d\underline{\eta}(t)/dt$ are m- and p-dimensional vectors whose components are white Gaussian processes. The associated convariance matrices are $\underline{X}(t)\ \delta(t-u)$ and $\underline{N}(t)\ \delta(t-u)$, respectively.

We shall assume that the actually observed process, $\underline{r}(t) = d\underline{y}(t)/dt$, is available from an initial observation time, t_o, until the present time, t. The entire observed waveform, $\{\underline{r}(\tau):t_o \le \tau \le t\}$, will be denoted by $\underline{r}_{t_o,t}$. Similarly, the entire waveform $\{\underline{y}(\tau):t_o \le \tau \le t\}$, will be denoted by $\underline{y}_{t_o,t}$. With the observation of $\underline{r}(t)$, the a priori probability density, $p(\underline{x};t)$, evolves to the a posteriori density, $p(\underline{x};t|\underline{r}_{t_o,t}) = p(\underline{x};t|\underline{y}_{t_o,t})$, for which the following equation has been correctly derived by Kushner[26] .

$$p(\underline{x};t+dt|\underline{y}_{t_0,t+dt}) - p(\underline{x};t|\underline{y}_{t_0,t})$$

$$= -\sum_{i=1}^{m} \frac{\partial}{\partial x_i}\left[f_i(t:\underline{x})\,p(\underline{x};t|\underline{y}_{t_0,t})\right]dt$$

$$+ \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m} X_{ij}(t)\frac{\partial^2}{\partial x_i \partial x_j} p(\underline{x};t|\underline{y}_{t_0,t})\,dt \quad (10)$$

$$+ p(\underline{x};t|\underline{y}_{t_0,t})\left[\underline{g}(t:\underline{x}) - E\underline{g}(t:\underline{x})\right]'$$

$$\cdot\, \underline{N}^{-1}(t)\left[d\underline{y}(t) - E\underline{g}(t:\underline{x})\,dt\right],$$

where E indicates expectation with respect to $p(\underline{x};t|\underline{y}_{t_o,t})$. The left side along with the first two terms of the right side of (10) are recognized as the Fokker-Planck equation associated with $\underline{x}(t)$, as given by (3). The last term on the right represents the modification to the Fokker-Planck equation resulting from the observation of $\underline{r}(t)$. When $\underline{g}[t:\underline{x}(t)]$, and hence $\underline{r}(t)$, does not depend on $\underline{x}(t)$, then the last term is zero and the equation reduces the original Fokker-Planck equation as expected.

DERIVATION OF THE ESTIMATION EQUATIONS

An equation for the exact minimum-mean-square-error estimate of $\underline{x}(t)$, given the accumulated observations, $\underline{r}_{t_o,t}$, can be derived by using (10) and the fact that the estimate, denoted by $\hat{\underline{x}}(t)$, is the conditional mean‡

$$\hat{\underline{x}}(t) = \underbrace{\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}}_{\text{m-fold}} \underline{x}\,p\left(\underline{x};t|\underline{y}_{t_0,t}\right)dx_1\ldots dx_m. \quad (11)$$

‡ The minimum-mean-square-error estimate of $\underline{x}(t)$ is a vector whose i^{th} component is the minimum-mean-square error estimate of $x_i(t)$.

Multiplying both sides of (10) by $\underline{x}$ and integrating results in

$$\hat{\underline{x}}(t+dt) - \hat{\underline{x}}(t) \equiv d\hat{\underline{x}}(t) = E\underline{f}(t:\underline{x})\,dt + E\left[\left\{\underline{x} - \hat{\underline{x}}(t)\right\}\right.$$
$$\left.\cdot\, \underline{g}'(t:\underline{x})\right]\underline{N}^{-1}(t)\left[d\underline{y} - E\underline{g}(t:\underline{x})\,dt\right], \quad (12)$$

where E denotes expectation with respect to $p\left(\underline{x};t|\underline{y}_{t_o,t}\right)$ and integration by parts has been used.

We now assume that the following Taylor expansions for $\underline{f}[t:\underline{x}(t)]$ and $\underline{g}[t:\underline{x}(t)]$ exist:

$$\underline{f}(t:\underline{x}) = \underline{f}(t:\hat{\underline{x}}) + \sum_{i=1}^{m}(x_i - \hat{x}_i)\frac{\partial}{\partial x_i}\underline{f}(t:\underline{x})\Big|_{\hat{\underline{x}}} \quad (13)$$
$$+ \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}(x_i - \hat{x}_i)(x_j - \hat{x}_j)\frac{\partial^2}{\partial x_i \partial x_j}\underline{f}(t:\underline{x})\Big|_{\hat{\underline{x}}} + \ldots$$

$$\underline{g}(t:\underline{x}) = \underline{g}(t:\hat{\underline{x}}) + \sum_{i=1}^{m}(x_i - \hat{x}_i)\frac{\partial}{\partial x_i}\underline{g}(t:\underline{x})\Big|_{\hat{\underline{x}}} \quad (14)$$
$$+ \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}(x_i - \hat{x}_i)(x_j - \hat{x}_j)\frac{\partial^2}{\partial x_i \partial x_j}\underline{g}(t:\underline{x})\Big|_{\hat{\underline{x}}} + \ldots$$

The second term is each expansion may be written as $\underline{D}'[\underline{f}(t:\hat{\underline{x}})](\underline{x}-\hat{\underline{x}})$ and $\underline{D}'[\underline{g}(t:\hat{\underline{x}})](\underline{x}-\hat{\underline{x}})$, respectively. $\underline{D}[\underline{f}(t:\underline{x})]$ is the Jacobian matrix associated with the vector, $\underline{f}[t:\underline{x}(t)]$; its (i-row, j-column)-element is $\frac{\partial}{\partial x_i} f_j[t:\underline{x}(t)]$.

The equation for the exact estimate can be obtained by substituting these expansions in (12). The resulting expression can neither be solved nor readily implemented because of the general existence of an infinite number of terms. It is natural, therefore, to consider the truncation of the expansions on the assumption that the components of the error vector, $\hat{\underline{x}}(t)$-$\hat{\underline{x}}(t)$, are small. This assumption can be expected to be valid when the disturbance processes introduce only small perturbations in the observed processes.

Let $\underline{x}^*(t)$ be the approximate minimum-mean-square-error estimate of $\underline{x}(t)$ which is specified by the substitution of the expansions for $\underline{f}[t:\underline{x}(t)]$ and $\underline{g}[t:\underline{x}(t)]$ in (12) and the retention of the most significant terms. Whenever $\underline{f}[t:\underline{x}(t)]$ and $\underline{g}[t:\underline{x}(t)]$ are linear functions of $\underline{x}(t)$, no approximation is involved and the exact and approximate estimates are identical. The equation that we obtain for $\underline{x}^*(t)$ is

$$d\underline{x}^*(t) = \underline{f}\left[t:\underline{x}^*(t)\right]dt + \underline{V}^*(t)\underline{D}\left[\underline{g}(t:\underline{x}^*)\right]\underline{N}^{-1}(t) \cdot \left\{d\underline{y} - \underline{g}\left[t:\underline{x}^*(t)\right]dt\right\} \tag{15}$$

or, formally dividing by dt

$$\frac{d}{dt}\underline{x}^*(t) = \underline{f}\left[t:\underline{x}^*(t)\right] + \underline{V}^*(t)\underline{D}\left[\underline{g}(t:\underline{x}^*)\right]\underline{N}^{-1}(t) \cdot \left\{\underline{r}(t) - \underline{g}\left[t:\underline{x}^*(t)\right]\right\}, \tag{16}$$

where $\underline{V}^*(t)$ is a symmetric, non-negative definite, m X m error-covariance matrix defined by $\underline{V}^*(t)=[\underline{x}-\underline{x}^*(t)][\underline{x}-\underline{x}^*(t)]'$. We shall refer to (16) as the "Processor Equation." All quantities in it are known except for the error-covariance matrix, $\underline{V}^*(t)$, for which an equation will be derived. The associated initial condition, $\underline{x}^*(t_o)$, is determined from

$$\underline{x}^*(t_o) = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\underline{x}\,p(\underline{x};t_o|\underline{y}_{t_o,t_o})\,dx_1\ldots dx_m = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\underline{x}\,p(\underline{x};t_o)\,dx_1\ldots dx_m, \tag{17}$$

where $p(\underline{x};t_o)$ is the a priori probability density of $\underline{x}$ at time t_o. That is $\underline{x}^*(t_o)$ is the best estimate of $\underline{x}(t_o)$ without any observations.

We note that the terms of (14) having the most significant effect on the processor equation are the first two of each expansion. Consequently, the approximation is, in effect, a linearization about the current estimate. This implies that within the approximation, $p\left(\underline{x};t|\underline{y}_{t_o},t\right)$ is normal with mean $\underline{x}^*(t)$.

We now turn to the derivation of an equation for $\underline{V}^*(t)$. An equation for $v^*_{k\ell}(t)$, the (k,ℓ)-element of $\underline{V}^*(t)$, is first obtained by the following procedure: (i) multiply the equation for $p\left(\underline{x};t|\underline{y}_{t_o},t\right)$ by $[x_k-\hat{x}_k(t)][x_\ell-\hat{x}_\ell(t)]$; (ii) integrate to obtain an equation for the (k,ℓ)-element of the exact error-covariance matrix; (iii) use the expansions for $\underline{f}[t:\underline{x}(t)]$ and $\underline{g}[t:\underline{x}(t)]$ and keep only the most significant terms.

Proceeding with steps i and ii, we use

$$\left[x_k-\hat{x}_k(t)\right]\left[x_\ell-\hat{x}_\ell(t)\right] = \left[x_k-\hat{x}_k(t+dt)\right] \cdot \left[x_\ell-\hat{x}_\ell(t+dt)\right] + d\hat{x}_k(t)\,d\hat{x}_\ell(t) - \left[x_k-\hat{x}_k(t+dt)\right] \cdot d\hat{x}_\ell(t) - \left[x_\ell-\hat{x}_\ell(t+dt)\right]d\hat{x}_k(t) \tag{18}$$

to obtain

$$dv_{k\ell}(t) + d\hat{x}_k(t)\,d\hat{x}_\ell(t) = E\left\{\underline{f}(t:\underline{x})\left[\underline{x}-\hat{\underline{x}}(t)\right]' + \left[\underline{x}-\hat{\underline{x}}(t)\right]\underline{f}'(t:\underline{x})\right\}_{k\ell}dt + X_{k\ell}(t)\,dt + E\left\{x_k-\hat{x}_k(t)\right\} \cdot \left\{x_\ell-\hat{x}_\ell(t)\right\}\left[\underline{g}(t:\underline{x}) - E\underline{g}(t:\underline{x})\right]' \cdot \underline{N}^{-1}(t)\left[d\underline{y} - E\underline{g}(t:\underline{x})\,dt\right], \tag{19}$$

where integration by parts has been used to obtain the first three terms on the right. We now substitute the expansions for $\underline{f}[t:\underline{x}(t)]$ and $\underline{g}[t:\underline{x}(t)]$ in (19) and keep only the most significant terms. We also use the fact that within the approximation, $p\left(\underline{x};t|\underline{y}_{t_o},t\right)$ is normal with mean $\underline{x}^*(t)$; consequently, odd moments of the components of the error vector, $\underline{x}-\underline{x}^*(t)$, are zero and even moments factor into products of second moments. The equation we obtain for $v^*_{k\ell}(t)$ is

$$dv^*_{k\ell}(t) + dx^*_k(t)\,dx^*_\ell(t) = \left\{\underline{D}'\left[\underline{f}(t:\underline{x}^*)\right]\underline{V}^*(t) + \underline{V}^*(t)\underline{D}\left[\underline{f}(t:\underline{x}^*)\right] + \underline{X}(t)\right\}_{k\ell}dt + \left[\sum_{i=1}^{m}\sum_{j=1}^{m} v^*_{ki}(t)\,v^*_{\ell j}(t)\,\frac{\partial^2}{\partial x^*_i\,\partial x^*_j}\,\underline{g}'\left[t:\underline{x}^*(t)\right]\right] \cdot \underline{N}^{-1}(t)\left\{d\underline{y} - \underline{g}\left[t:\underline{x}^*(t)\right]dt\right\}. \tag{20}$$

The second term on the left, $dx_k^* dx_\ell^* = [(d\underline{x}^*)(d\underline{x}^*)']_{k\ell}$, remains to be examined. Using (15) and keeping terms to the order of dt, we have

$$(d\underline{x}^*)(d\underline{x}^*)' = \underline{V}^*(t)\,\underline{D}\left[\underline{g}(t:\underline{x}^*)\right]\underline{N}^{-1}(t)\,d\underline{y}\,d\underline{y}' \cdot \underline{N}^{-1}(t)\,\underline{D}'\left[\underline{g}(t:\underline{x}^*)\right]\underline{V}^*(t). \quad (21)$$

Since we are retaining only those terms of order and dt and dt is infinitesimal, $d\underline{y}d\underline{y}'$ may be replaced by $\underline{N}(t)dt$ as indicated by (7). Hence, to terms of order dt

$$(d\underline{x}^*)(d\underline{x}^*)' = \underline{V}^*(t)\,\underline{D}\left[\underline{g}(t:\underline{x}^*)\right]\underline{N}^{-1}(t) \cdot \underline{D}'\left[\underline{g}(t:\underline{x}^*)\right]\underline{V}^*(t)\,dt. \quad (22)$$

Substituting this result in (20), we have

$$dv_{k\ell}^*(t) = \Big\{\underline{D}'\left[\underline{f}(t:\underline{x}^*)\right]\underline{V}^*(t) + \underline{V}^*(t)\underline{D}\left[\underline{f}(t:\underline{x}^*)\right] + \underline{X}(t) + \underline{V}^*(t)\underline{D}\left[\underline{g}(t:\underline{x}^*)\right]\underline{N}^{-1}(t)\underline{D}'\left[\underline{g}(t:\underline{x}^*)\right]\underline{V}^*(t)\Big\}_{k\ell}dt + \left[\sum_{i=1}^{m}\sum_{j=1}^{m} v_{ki}^*(t)v_{\ell j}^*(t)\frac{\partial^2}{\partial x_i^*\partial x_j^*}\underline{g}'\left[t:\underline{x}^*(t)\right]\right]\underline{N}^{-1}(t) \cdot \left\{d\underline{y} - \underline{g}\left[t:\underline{x}^*(t)\right]dt\right\} \quad (23)$$

$$= \Big\{\underline{D}'\left[\underline{f}(t:\underline{x}^*)\right]\underline{V}^*(t) + \underline{V}^*(t)\underline{D}\left[\underline{f}(t:\underline{x}^*)\right] + \underline{V}^*(t) \cdot \underline{D}\left[\underline{D}[\underline{g}(t:\underline{x}^*)]\,\underline{N}^{-1}\left\{\underline{r}(t) - \underline{g}(t:\underline{x}^*)\right\}\right]\underline{V}^*(t)\Big\}_{k\ell}dt. \quad (24)$$

That (23) and (24) are equal may be demonstrated by expanding the matrix expressions. Formally dividing (24) by dt results in the following equation for $\underline{V}^*(t)$:

$$\frac{d}{dt}\underline{V}^*(t) = \underline{D}'\left[\underline{f}(t:\underline{x}^*)\right]\underline{V}^*(t) + \underline{V}^*(t)\underline{D}\left[\underline{f}(t:\underline{x}^*)\right] + \underline{X}(t) + \underline{V}^*(t)\underline{D}\left[\underline{D}[\underline{g}(t:\underline{x}^*)]\,\underline{N}^{-1}(t) \cdot \left\{\underline{r}(t) - \underline{g}(t:\underline{x}^*)\right\}\right]\underline{V}^*(t). \quad (25)$$

We shall refer to (25) as the "Variance Equation." The associated initial condition is determined from

$$\underline{V}^*(t_0) = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\left[\underline{x} - \underline{x}^*(t_0)\right]\left[\underline{x} - \underline{x}^*(t_0)\right]' \cdot p(\underline{x};t_0)\,dx_1\ldots dx_m. \quad (26)$$

The processor equation (16) and the variance equation (25) jointly define the quasi-optimum estimate, $\underline{x}^*(t)$. In general, the equations are coupled and both depend on the observation vector, $\underline{r}(t)$. When $\underline{f}[t:\underline{x}(t)]$ and $\underline{g}[t:\underline{x}(t)]$ are linear transformations of $\underline{x}(t)$, the equations reduce to those of Kalman and Bucy[7] in which case the equations are uncoupled and $\underline{V}^*(t)$ does not depend on $\underline{r}(t)$. We shall see that this also occurs in angle-modulation schemes in which the transmitted-signal spectrum is essentially disjoint from the message spectrum. This is of practical significance, since it implies that $\underline{V}^*(t)$, and hence the structure of the quasi-optimum demodulator, can be determined before any observations.

THE COMMUNICATION MODEL

The Communication Model is shown in Fig. 2.

Let $\underline{a}(t)$ be an n-dimensional state vector representing the output of an analog message source. $\underline{a}(t)$ is a continuous vector Markov process defined by the Ito equation:

$$d\underline{a}(t) = \underline{f}_a\left[t:\underline{a}(t)\right]dt + d\underline{\alpha}(t), \quad (27)$$

where $\underline{\alpha}(t)$ is an n-dimensional vector whose components are Wiener processes. Let the covariance matrix associated with $\underline{\alpha}(t)$ be given by

$$E\left[\underline{\alpha}(t)\underline{\alpha}'(u)\right] = \underline{A}(t)\min(t,u), \quad (28)$$

where $\underline{A}(t)$ is non-negative definite, n X n matrix which may be time-varying. More than one message can be represented simply by adjoining their individual state vectors in the formation of $\underline{a}(t)$. Of course, Gaussian messages with rational spectra are a special case of (27) with $\underline{f}_a[t:\underline{a}(t)] = \underline{F}_a\underline{a}(t)$.

The message vector, $\underline{a}(t)$, is transformed by a modulator into c signals appropriate for transmission over the channel. The modulator consists of linear filtering followed by a memoryless, nonlinear modulator. The linear filtering may be time-varying and is described by the state equation

$$d\underline{u}(t) = \underline{F}_u(t)\underline{u}(t)dt + \underline{L}_a(t)\underline{a}(t)dt, \tag{29}$$

where $\underline{u}(t)$ is an ℓ-dimensional vector, and $\underline{F}_u(t)$ and $\underline{L}_a(t)$ are matrices of dimensional $\ell \times \ell$ and $\ell \times n$, respectively. The c signals at the modulator output are represented by $\underline{s}[t:\underline{u}(t)]$.

A second linear-filtering operation follows the modulator. It is described by the state equation

$$d\underline{z}(t) = \underline{F}_z(t)\underline{z}(t)dt + \underline{L}_s(t)\underline{s}\left[t:\underline{u}(t)\right]dt, \tag{30}$$

where $\underline{z}(t)$ is a q-dimensional vector, and $\underline{F}_z(t)$ and $\underline{L}_s(t)$ are matrices of dimensional $q \times q$ and $q \times c$, respectively. We shall allow this filtering to be associated with either the modulator or the channel, the choice depending upon the particular application.

The modulator, including possible linear filtering at its output, contains as special cases: linear-modulation schemes, such as AM, AM-DSB/SC, and AM-SSB; nonlinear-modulation schemes, such as PM, FM, and preemphasized FM; diversity-modulation schemes, such as frequency-diversity PM and FM; and multi-level-modulation schemes, such as PM_n/PM.

The channel inputs are transformed into p signals that are represented by the vector $\underline{g}[t:\underline{x}(t)]$. Each component of $\underline{g}[t:\underline{x}(t)]$ is observed in additive white Gaussian noise. The observed process can be described by the Ito equation

$$d\underline{y}(t) = \underline{g}\left[t:\underline{x}(t)\right]dt + d\underline{\eta}(t), \tag{31}$$

where $\underline{\eta}(t)$ is a p-dimensional vector whose components are Wiener processes. Let the covariance matrix associated with $\underline{\eta}(t)$ be given by

$$E\left[\underline{\eta}(t)\,\underline{\eta}'(u)\right] = \underline{N}(t)\min(t,u), \tag{32}$$

where $\underline{N}(t)$ is a symmetric, positive-definite, $p \times p$ matrix that may be time-varying. We assume that $\underline{N}^{-1}(t)$ exists. The actually observed process is $\underline{r}(t) = d\underline{y}(t)/dt$. Note that we have defined $\underline{y}(t)$ for the communication model in exactly the same way as $\underline{y}(t)$ for the estimation model.

Disturbance processes, such as additive and multiplicative processes, are introduced in the randomly time-varying portion of the channel. These processes can be Markovian in general and are described by the Ito equation:

$$d\underline{b}(t) = \underline{f}_b\left[t:\underline{b}(t)\right]dt + d\underline{\beta}(t), \tag{33}$$

where $\underline{b}(t)$ and $\underline{\beta}(t)$ are k-dimensional vectors. The components of $\beta(t)$ are Wiener processes and the associated covariance matrix is given by

$$E\left[\underline{\beta}(t)\underline{\beta}'(u)\right] = \underline{B}(t)\min(t,u), \tag{34}$$

where $\underline{B}(t)$ is a symmetric, non-negative definite, $k \times k$ matrix that may be time-varying. Of course, as a special case, the disturbance processes can be Gaussian processes with rational spectra.

The channel, including possible linear filtering at its input, contains as special cases: simple additive channels; Gaussian multiplicative channels, such as Rayleigh and Rician channels; fixed channels with memory; multilink channels; and other commonly occurring channels. The Markovian disturbance processes that we include in the model cannot be treated with any alternative approach.

The vector, $\underline{x}(t)$, of the communication model is obtained by adjoining the individual state vectors $\underline{a}(t)$, $\underline{u}(t)$, $\underline{z}(t)$, and $b(t)$. Let

$$\underline{x}(t) = \begin{bmatrix} \underline{a}(t) \\ \hline \underline{u}(t) \\ \hline \underline{z}(t) \\ \hline \underline{b}(t) \end{bmatrix}; \quad \underline{f}\left[t:\underline{x}(t)\right] = \begin{bmatrix} \underline{f}_a\left[t:\underline{a}(t)\right] \\ \hline \underline{F}_u(t)\underline{u}(t) + \underline{L}_a(t)\underline{a}(t) \\ \hline \underline{F}_z(t)\underline{z}(t) + \underline{L}_s(t)\underline{s}\left[t:\underline{u}(t)\right] \\ \hline \underline{f}_b\left[t:\underline{b}(t)\right] \end{bmatrix} \tag{35}$$

and

$$\underline{\chi}(t) = \begin{bmatrix} \underline{\alpha}(t) \\ \hline 0 \\ \hline 0 \\ \hline \underline{\beta}(t) \end{bmatrix}.$$

Then

$$d\underline{x}(t) = \underline{f}\left[t:\underline{x}(t)\right]dt + d\underline{\chi}(t). \qquad (36)$$

describes $\underline{x}(t)$ for the communication model and is identical to (1) describing $\underline{x}(t)$ for the estimation model. The order in which $\underline{a}(t)$, $\underline{u}(t)$, $\underline{z}(t)$ and $\underline{b}(t)$ are placed in forming $\underline{x}(t)$ is arbitrary.

With the definition of the communication model now completed, we turn our attention to the consideration of applications. The procedure is (i) specify the particular communication model for the application; (ii) identify $\underline{x}(t)$, $\underline{f}[t:\underline{x}(t)]$, $\underline{X}(t)$, $\underline{r}(t) = d\underline{y}(t)/dt$, $\underline{g}[t:\underline{x}(t)]$, and $\underline{N}(t)$; (iii) use the Processor and Variance equations (16 and 25) to determine the structure of the demodulator.

APPLICATIONS

1. Gaussian Message - No Modulation

Consider the communication model shown in Fig. 3a. This is a simple situation to which the Wiener approach to the filtering problem is often applied. It provides some insight into the results that we shall obtain for angle modulation and, at the same time, into the relationship between the structure of Wiener and Kalman-Bucy filters.

a(t) is a stationary Gaussian message and n(t) is a white Gaussian process of spectral height N_o watts/cps. a(t) and n(t) are uncorrelated.

The equations describing the model are (with the representation for a(t) given in the Appendix):

$$\frac{d}{dt}\underline{x}(t) = \underline{F}\underline{x}(t) + \underline{\xi}(t) \qquad (37)$$

and

$$\frac{d}{dt}y(t) = r(t) = x_1(t) + n(t) \qquad (38)$$

where $\underline{x}(t)$ is an m-dimensional vector with $x_1(t) = a(t)$. $\underline{F}$ and $\underline{\xi}(t)$ are defined in the Appendix.

We assume that $E[\underline{\xi}(t)\underline{\xi}'(u)] = X\delta(t-u)$ is known. From (38) we have $\underline{g}[t:\underline{x}(t)] = x_1(t)$.

The processor and variance equations (16 and 25) become

$$\frac{d}{dt}\underline{\hat{x}}(t) = \underline{F}\underline{\hat{x}}(t) + \frac{1}{N_0}\begin{bmatrix} v_{11}(t) \\ v_{12}(t) \\ \vdots \\ v_{1m}(t) \end{bmatrix}\left\{r(t) - \hat{x}_1(t)\right\} \qquad (39)$$

and

$$\frac{d}{dt}\underline{V}(t) = \underline{F}(t)\underline{V}(t) + \underline{V}(t)\underline{F}'(t) + \underline{X}(t) - \frac{1}{N_0}\underline{M}(t) \qquad (40)$$

where $\underline{M}(t)$ is a symmetric m X m matrix whose (i,j)-element is $v_{1i}(t)\, v_{1j}(t)$. By comparing (37) and (39), we obtain the optimum processor shown in Fig. 3b. We observe that it depends only on the first column of $\underline{V}(t)$.

$\underline{V}(t)$ can be determined numerically or can be generated as the output of the system specified by (40). If desired, $\underline{V}(t)$ can be determined before any actual observations. The components of $\underline{V}(t)$ are of interest for two reasons: first, they complete the structure of the processor; second they describe the performance of the processor. We shall not give solutions to the variance equation here. Rather, we shall be interested only in obtaining the general structure of the optimum processor.

A special case arises when $t_o = -\infty$ so that steady-state conditions exist.** In this instance, the optimum filter has the alternative form shown in Fig. 4. The structure of the optimum filter which would arise most naturally through application of the Wiener approach is, of course, the closed-loop version of the filter of Fig. 4.

2. <u>Gaussian Message - Phase Modulation</u>

Consider the communication model shown in Fig. 5 in which a stationary Gaussian message, a(t), phase modulates a sinusoidal carrier whose nominal frequency is large compared with significant frequencies of a(t). We shall assume that the variance of a(t) is unity so that β can be interpreted as the modulation index. The phase-modulated signal is observed in additive white Gaussian noise of spectral height N_0 watts/cps.

The equations describing the communication model are (with the representation for a(t) given in the Appendix)

$$\frac{d}{dt}\underline{x}(t) = \underline{F}\,\underline{x}(t) + \underline{\xi}(t) \tag{41}$$

and

$$\frac{d}{dt}y(t) = r(t) = C\sin\left[\omega_0 t + \beta a(t)\right] + n(t), \tag{42}$$

where $\underline{x}(t)$ is an m-dimensional vector, and $\underline{F}$ and $\underline{\xi}(t)$ are as defined in the Appendix. Observe that $x_1(t) = a(t)$.

We assume that $E[\underline{\xi}(t)\underline{\xi}'(u)] = \underline{X}\delta(t-u)$ is known. In this instance, $\underline{g}[t:x(t)] = C\sin[\ _0 t + \beta x_1(t)]$, a $\bar{s}$calar. Hence

$$\underline{D}\left[\underline{g}(t:\underline{x})\right] = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \beta C\cos\left[\omega_0 t + \beta x_1(t)\right]. \tag{43}$$

After some manipulation, the processor and variance equations (16 and 25) become

$$\frac{d}{dt}\underline{x}^*(t) = \underline{F}\underline{x}^*(t) + \frac{1}{N_0}\begin{bmatrix} v_{11}^*(t) \\ v_{12}^*(t) \\ \vdots \\ v_{1m}^*(t) \end{bmatrix} \beta C\cos\left[\omega_0 t + \beta x_1^*(t)\right] \cdot\left\{r(t) - C\sin\left[\omega_0 t + \beta x_1^*(t)\right]\right\} \tag{44}$$

and

$$\frac{d}{dt}\underline{V}^*(t) = \underline{F}\,\underline{V}^*(t) + \underline{V}^*(t)\underline{F}' + \underline{X} - \frac{1}{N_0}\beta^2 C\left\{r(t)\sin\left[\omega_0 t + \beta x_1^*(t)\right] + C\cos\left[2\omega_0 t + 2\beta x_1^*(t)\right]\right\}\underline{M}(t) \tag{45}$$

where $\underline{M}(t)$ is a symmetric m X m matrix whose (i,j)-element is $v_{1i}^*(t)\ v_{1j}^*(t)$.

We shall examine the variance equation first. From (45), the (i,j)-element of $\underline{V}^*(t)$ satisfies

$$\frac{d}{dt}v_{ij}^*(t) = -\psi_i v_{1j}^*(t) - \psi_j v_{1i}^*(t) + v_{i+1,j}^*(t) + v_{j+1,i}^*(t) + X_{ij} - \frac{1}{N_0}\beta^2 C v_{1i}^*(t)\,v_{1j}^*(t) \cdot\left\{r(t)\sin\left[\omega_0 t + \beta x_1^*(t)\right] + C\cos\left[2\omega_0 t + 2\beta x_1^*(t)\right]\right\}. \tag{46}$$

$v_{ij}^*(t)$ can be realized as the output of the system diagrammed in Fig. 6. Let us now conjecture that the components of V*(t) are alowly varying. We shall find that to a close approximation this is, in fact, true. Then the double-frequency terms associated with $\cos[2\omega_0 t+2\beta x_1^*(t)]$ will not propagate through the lowpass filtering. Consequently, $\cos[2\omega_0 t+2\beta x_1^*(t)]$ has negligible effect and can be dropped.

**Sufficient conditions for the existence of a unique steady-state solution are given by Kalman and Bucy[7].

The input to the multiplier is then $r(t)\ \sin[\omega_0 t+\beta x_1^*(t)]$. It is through this term that the variance equation is coupled to both $r(t)$ and $\underline{x}^*(t)$. This coupling is a great disadvantage practically because $\underline{V}^*(t)$ and, therefore, the structure of the demodulator, cannot be determined before making observations. For this reason, it is worthwhile to examine $r(t)\ \sin[\omega_0 t+\beta x_1^*(t)]$ critically so as to obtain any possible simplification. We shall find that a significant simplification is possible.

Observe that that coupling term may be rewritten

$$r(t)\sin\left[\omega_0 t+\beta x_1^*(t)\right] = n(t)\sin\left[\omega_0 t+\beta x_1^*(t)\right]$$

$$+ C\sin\left[\omega_0 t+\beta x_1(t)\right]\sin\left[\omega_0 t+\beta x_1^*(t)\right]$$

$$= n(t)\sin\left[\omega_0 t+\beta x_1^*(t)\right] + \tfrac{1}{2}C\cos\beta\left[x_1(t)-x_1^*(t)\right]$$

$$-\tfrac{1}{2}C\cos\left[2\omega_0 t+\beta x_1(t)+\beta x_1^*(t)\right]. \tag{47}$$

Again, the double-frequency term can be disregarded . The second term on the right can be expanded:

$$\tfrac{1}{2}C\cos\beta\left[x_1^*(t)-x_1(t)\right]$$
$$= \tfrac{1}{2}C - \tfrac{1}{4}C\beta^2\left[x_1(t)-x_1^*(t)\right]^2 + \ldots \tag{48}$$

Within the approximation for which the demodulator is optimum, all terms of the expansion except the first can be neglected; the others lead to terms of the order of the sixth moment of the error at the output of the multiplier. Thus, to a good approximation for small error, we have

$$r(t)\sin\left[\omega_0 t+\beta x_1^*(t)\right]$$
$$\approx \tfrac{1}{2}C\left\{1+\tfrac{2}{C}n(t)\sin\left[\omega_0 t+\beta x_1^*(t)\right]\right\},$$

where $n(t)$ is a white process, by which we mean that it has a flat spectrum at least over the frequency range where it has effect. In reality, $n(t)$ has a finite variance given by N_0W_c, where W_c is the channel or receiver input bandwidth. By increasing the channel signal-to-noise ratio, $C^2/2N_0W_c$, it is possible to make the probability of excursions of $2n(t)/C$ outside a range around its mean, zero, as small as desired. Since the magnitude of sin(.) is bounded by unity, this implies

$$\tfrac{1}{2}C\left\{1+\tfrac{2}{C}n(t)\sin\left[\omega_0 t+\beta x_1^*(t)\right]\right\} \approx \tfrac{1}{2}C$$

almost always when the signal-to-noise ratio is sufficiently large. We conclude that for large channel signal-to-noise ratio

$$r(t)\sin\left[\omega_0 t+\beta x_1^*(t)\right] \approx \tfrac{1}{2}C. \tag{49}$$

The approximations have effected an uncoupling of the variance equation from $r(t)$ and $x^*(t)$, thereby making a practical simplification of importance. The variance equation becomes:

$$\frac{d}{dt}\underline{V}^*(t) = \underline{F}\,\underline{V}^*(t) + \underline{V}^*(t)\underline{F}' + X - \frac{\beta^2C^2}{2N_0}\underline{M}(t). \tag{50}$$

This equation is nearly identical to the variance equation associated with the no-modulation case of Example 1(see Eq. 40). Only the noise level must be modified. $\underline{V}^*(t)$ can be determined before making any observations, just as in the no-modulation case.

In the steady-state, the Processor equation (44) leads to the quasioptimum PM demodulator of Fig. 7. It is seen that the subtractive sinusoidal signal results only in double-frequency terms at the output of the multiplier. Since these will not propagate through the filter, the subtractive branch can be discarded. The simplified demodulator is a phase-locked loop.

3. <u>Gaussian Message - Frequency Modulation</u>

Consider the communication model shown in Fig. 8 in which $a(t)$ now frequency modulates a sinusoidal carrier. We assume that $a(t)$ has unit variance;

d_f is then the standard deviation of the modulation frequency. The spectral height of n(t) is N_o watts/cps.

If we let $x_o(t) = u(t)$, the integrated message, then the equations describing the communication model are (with the representation for a(t) given in the Appendix):

$$\frac{d}{dt}\underline{x}(t) = \underline{F}\,\underline{x}(t) + \underline{\xi}(t) \qquad (51)$$

and

$$r(t) = C\sin\left[\omega_0 t + d_f\int_{t_0}^{t} a(\tau)d\tau\right] + n(t)$$
$$= C\sin\left[\omega_0 t + d_f x_0(t)\right] + n(t), \qquad (52)$$

where

$$\underline{x}(t) = \begin{bmatrix} u(t) \\ \hline \underline{a}(t) \end{bmatrix} = \begin{bmatrix} x_0(t) \\ x_1(t) \\ \vdots \\ x_m(t) \end{bmatrix}$$

and

$$\underline{F} = \begin{bmatrix} 0 & 1 & 0 & 0 & & \\ 0 & -\psi_1 & 1 & 0 & & \\ 0 & -\psi_2 & 0 & 1 & & \\ & \vdots & & & & 1 \\ 0 & -\psi_m & 0 & 0 & \dots & 0 \end{bmatrix}; \quad \underline{\xi}(t) = \begin{bmatrix} 0 \\ \lambda_1\,\xi(t) \\ \lambda_2\,\xi(t) \\ \vdots \\ \lambda_m\,\xi(t) \end{bmatrix}$$

Note that $u(t) = x_o(t)$ and $a(t) = x_1(t)$.

We assume that $E[\underline{\xi}(t)\underline{\xi}'(u)] = \underline{X}\delta(t-u)$ is known. In this instance, $\underline{g}[t:\underline{x}(t)] = C\sin[\omega_o t + d_f x_o(t)]$. Hence

$$\underline{D}\left[\underline{g}(t:\underline{x})\right] = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} d_f C\cos\left[\omega_0 t + d_f x_0(t)\right].$$

After some manipulation, the Variance and Processor equations (16 and 25) become

$$\frac{d}{dt}\underline{x}^*(t) = \underline{F}\,\underline{x}^*(t) + \frac{1}{N_0}\begin{bmatrix} v^*_{00}(t) \\ v^*_{01}(t) \\ \vdots \\ v^*_{0m}(t) \end{bmatrix} d_f C$$
$$\cdot\cos\left[\omega_0 t + d_f x^*_0(t)\right]\left\{r(t) - C\sin\left[\omega_0 t + d_f x^*_0(t)\right]\right\} \qquad (54)$$

and

$$\frac{d}{dt}\underline{V}^*(t) = F\underline{V}^*(t) + \underline{V}^*(t)F + X$$
$$-\frac{1}{N_0}d_f^2 C\left\{r(t)\sin\left[\omega_0 t + d_f x^*_0(t)\right]\right.$$
$$\left. + C\cos\left[2\omega_0 t + 2d_f x^*_0(t)\right]\right\}\underline{M}(t) \qquad (55)$$

where $\underline{M}(t)$ is a symmetric (m+1) x (m+1) matrix whose (i,j)-element is $v_{oi}^*(t)\, v_{oj}^*(t)$. We observe that (55) is equivalent to (45), the variance equation for the PM case. Therefore, the arguments leading to the simplified variance equation, (50), carry over and (55) becomes

$$\frac{d}{dt}\underline{V}^*(t) = \underline{F}\,\underline{V}^*(t) + \underline{V}^*(t)\underline{F}' + \underline{X} - \frac{d_f^2C^2}{2N_0}\underline{M}(t) \qquad (56)$$

Eq. 56 also arises in connection with a linear filtering problem in which a(t) is integrated before being observed in additive white Gaussian noise.

In the steady state, the Processor equation (54) leads to the quasi-optimum FM demodulator of Fig. 9(the subtractive sinusoidal term of (54) has been omitted, since it has no effect). This demodulator can be placed in the form of a phase-locked loop. which is optimum for estimating u(t), and a realizable post-loop filter, whose output is a*(t). It is this last structure that arises most naturally with the MAP approach and is probably more familiar. The demodulator of Fig. 9 has the advantage of requiring one less filter.

CONCLUSION

The usefulness of the state-variable approach in treating problems of analog communication theory has been illustrated by considering angle-modulation schemes. Such schemes have also been treated by the MAP approach, so we note here the relative advantages and disadvantages associated with the two approaches. Some advantages of the state-variable approach are that (i) considerable insight into the structure of the demodulator is provided; (ii) the differential equations associated with the approach are more amenable to numerical evaluation than the integral equations of the MAP approach; (iii) realizable demodulators result directly; and (iv) a class of non-Gaussian message and channel disturbances can be treated. In the communication theory context, it is not yet clear what usefulness (iv) has; however, applications in control theory can be given. These arise when we wish to estimate the state variables of a nonlinear, dynamic system based on noisy observations of the state variables.

Some disadvantages are that (i) It is necessary that random processes and linear filtering be representable by equations of state. Thus, Gaussian processes with nonrational spectra cannot be treated. A particular linear operation which arises in array problems, for example, and cannot be treated directly is that of pure delay. (ii) The unrealizable filtering problem cannot be treated easily.

In addition to the applications to analog communication theory presented here, we have also considered the following[37] problems:

1. FM signals transmitted over several diversity channels;
2. FM signals transmitted via Rayleigh fading channels;
3. FM signals transmitted via fixed channels with memory; and
4. PM signals transmitted via a random-phase channel (i.e., unstable local oscillator).

The structure of the realizable quasi-optimum demodulator for each of these cases can be determined by a straightforward application of the processor and variance equations given above (16 and 25).

APPENDIX: STATE REPRESENTATION FOR GAUSSIAN PROCESSES

Any stationary, scalar Gaussian process, $x(t)$, with a rational spectrum approaching zero for high frequencies can be represented by the differential equation

$$\frac{d^m}{dt^m}x(t) + \psi_1\frac{d^{m-1}}{dt^{m-1}}x(t) + \dots + \psi_m x(t) = \lambda_1\frac{d^{m-1}}{dt^{m-1}}\xi(t) + \lambda_2\frac{d^{m-2}}{dt^{m-2}} + \dots + \lambda_m\xi(t) \tag{57}$$

where $\psi_1, \dots, \psi_m$ and $\lambda_1, \dots, \lambda_m$ are constants, and $\xi(t)$ is a white Gaussian process. As is well-known, $x(t)$ can be realized by the passage of $\xi(t)$ through the filter shown in Fig. 10a. Alternative realizations can be obtained by representing $x(t)$ by one of several possible equations of state. A particular state representation that we shall use, of which a detailed account is given by Zadeh and Desoer[28].

$$\begin{aligned}
\frac{d}{dt}x_1(t) &= -\psi_1 x_1(t) + x_2(t) + \lambda_1\xi(t)\\
\frac{d}{dt}x_2(t) &= -\psi_2 x_1(t) + x_3(t) + \lambda_2\xi(t)\\
&\vdots\\
\frac{d}{dt}x_{m-1}(t) &= -\psi_{m-1} x_1(t) + x_m(t) + \lambda_{m-1}\xi(t)\\
\frac{d}{dt}x_m(t) &= -\psi_m x_1(t) + \lambda_m\xi(t),
\end{aligned} \tag{58}$$

where

$$x(t) = x_1(t).$$

Eq. (58) leads to the alternative realization shown in Fig. 10b. We shall represent the equation in matrix notation as

$$\frac{d}{dt}\underline{x}(t) = F\underline{x}(t) + \underline{\xi}(t), \qquad (59)$$

where

$$F = \begin{bmatrix} -\psi_1 & 1 & 0 & 0 & \cdots & \\ -\psi_2 & 0 & 1 & 0 & \cdots & \\ -\psi_3 & 0 & 0 & 1 & \ddots & \\ \vdots & & & & & 1 \\ -\psi_m & 0 & \cdots & & & 0 \end{bmatrix}; \quad \underline{\xi}(t) = \begin{bmatrix} \lambda_1 \xi(t) \\ \lambda_2 \xi(t) \\ \lambda_3 \xi(t) \\ \vdots \\ \lambda_m \xi(t) \end{bmatrix}.$$

Observe that F contains all of the denominator coefficients associated with the rational polynomial realization and, correspondingly, $\underline{\xi}(t)$ contains all of the numerator coefficients. Because of this feature, the rational polynomial representation can be obtained by inspection from the state representation, and vice versa. Also observe that the scalar process, x(t), corresponds directly to one of the components of $\underline{x}(t)$.

A nonstationary scalar Gaussian process can be represented by (58 or 59) with time-varying coefficients, $\psi_1(t), \ldots, \psi_m(t)$ and $\lambda_1(t), \ldots, \lambda_m(t)$. The filter of Fig. 10b with varying gains can be used to realize the process.

ACKNOWLEDGEMENT

It is with pleasure that I acknowledge the invaluable guidance given me in this study by Professor Harry L. Van Trees of the Massachusetts Institute of Technology. His frequent discussions with me are sincerely appreciated.

REFERENCES

1. W. M. Wonham, "Some applications of stochastic differential equations to optimal nonlinear filtering," RIAS TR 64-3; Feb., 1964.

2. W. M. Wonham, "Some applications of stochastic differential equations to optimal nonlinear filtering," J.SIAM Control, Ser. A, Vol. 2, No. 3, pp. 347-369; 1965.

3. C. S. Weaver, "Estimating and detecting the outputs of linear dynamical systems," TR No. 6302-7, Systems Theory Lab., Stanford University, Stanford, California; 1964. (Also available as AD 464023 from DDC.)

4. H. Cox, 'Estimation of state-variables and parameters for noisy dynamic systems," Sc.D. Theses, Massachusetts Institute of Technology, Cambridge, Mass.; 1963.

5. H. Cox, "On the estimation of state variables and parameters for noisy dynamic systems," IEEE Trans. on Automatic Control, Vol. AC-9, No. 1; Jan., 1964.

6. V. O. Mowrey, "Least squares recursive differential-correction estimation in nonlinear problems," IEEE Trans. on Automatic Control, Vol. AC-10, No. 4, pp. 399-407; Oct., 1965.

7. R. E. Kalman and R. Bucy, "New results in linear filtering and prediction theory," ASME J. Basic Eng., Vol. 83, pp. 95-108; March, 1961.

8. R. S. Bucy, "Nonlinear filtering theory," IEEE Trans. on Automatic Control (Correspondence), Vol.AC-10, No. 2, p.198; April, 1965.

9. R. L. Stratonvich, "On the theory of optimal nonlinear filtration of random functions," Th. of Prob. and Its App., Vol. 4, pp. 223-225;1959.

10. R. L. Stratonvich, "Optimum nonlinear systems which bring about a separation of a signal with constant parameters from noise," Radiofizika, Vol. II, No. 6, pp. 892-901; 1959.

11. R. L. Stratonvich, "Application of the theory of Markov processes to optimal signal discrimination," Radio Eng. and Electronics, Vol. 5, No. 11, pp. 1751-1763; 1960.

12. N. K. Kul'man and R. L. Stratonovich, "Phase automatic frequency control and optimal measurement of narrowband signal parameters with nonconstant frequency in the presence of noise," Radio Eng. and Elec. Physics, No. 1, pp. 52-60; Jan.,1964.

13. N. K. Kul'man, "Optimum reception of a signal of variable frequency and amplitude in the presence of noise," Radio Eng. and Elec. Physics, Vol.9, No. 5, pp. 1285-1292; May, 1964.

14. I.A. Bol'shakov and V.G. Repin, "Problems of nonlinear filtration: I. The case of one parameter," Automation and Remote Control, Vol. 22, No. 4, pp. 397-408; Apr., 1961.

15. I. A. Bol'shakov and V.G. Repin, "Questions of nonlinear filtering: II. Multidimensional case," Automation and Remote Control, Vol. 25, No. 12, pp. 1499-1511; May, 1965.

16. Y. N. Bakayev and A. A. Guzh, "Optimal reception of doppler-shifted frequency-modulated signals," Radio Eng. and Elec. Physics, Vol.10, No. 1, pp. 135-139; Jan., 1965.

17. V. I. Tikhonov, "Nonlinear Filtration and Quasioptimal Nature of Frequency Phase Autotuning," News of the Acad. of Sci., USSR, 121-138; July, 1965. (Also available as NASA STAR Report N65-27987.)

18. F. W. Lehan and R. J. Parks, "Optimum demodulation," IRE National Convention Record, Pt. 8, pp. 101-103; 1953.

19. D. C. Youla, "The use of maximum likelihood in estimating continuously modulated intelligence which has been corrupted by noise," IRE Trans. on Information Theory, Vol. IT-3, pp. 90-105; March, 1954.

20. H. L. Van Trees, Detection, Estimation, and Modulation Theory, John Wiley and Sons, Inc., New York; (to appear).

21. H. L. Van Trees, "Analog communication over randomly time-varying channels," IEEE Trans. on Information Theory, Vol. IT-12, No. 1, pp. 51-63; Jan., 1966.

22. H. L. Van Trees, "The structure of efficient demodulators for multidimensional phase modulated signals," IEEE Trans. on Communication Systems, pp. 261-271; Sept., 1963.

23. J. B. Thomas and E. Wong, "On the statistical theory of optimum demodulation," IRE Trans. on Information Theory, Vol. IT-6, pp. 420-425; Sept., 1960.

24. K. Ito, "On stochastic differential equations," Mem. Am. Math. Soc., Vol. 4, pp. 1-51; 1951.

25. A. T. Bharucha-Reid, Elements of the Theory of Markov Processes and Their Applications, McGraw-Hill Book Co., New York; 1960.

26. H. J. Kushner, "On the dynamical equations of conditional probability density functions, with applications to optimal stochastic control theory," J. Math. Anal. Appl., pp. 332-344; April, 1964.

27. D. L. Snyder, "The state-variable approach to continuous estimation," Ph.D. Thesis, Department of Electrical Engineering, M.I.T., Cambridge, Mass.; February, 1966.

28. L. Zadeh and C. Desoer, Linear System Theory, McGraw-Hill Book Co., New York; 1963.

29. R. L. Stratonvich, "A new form of representing stochastic integrals and equations," Vestnik Moskov Univ., Ser. I. Mat. Meh., pp. 3-12; 1964.

30. E. Wong and M. Zakai, "On convergence of the solutions of differential equations involving Brownian motion," Report No. 65-5, E. R. L., University of California, Berkeley, Calif., Jan., 1965.

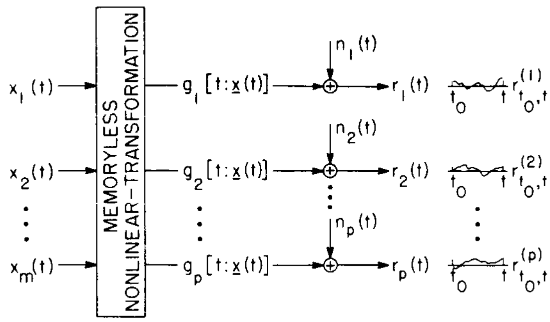


Fig. 1. The estimation model.

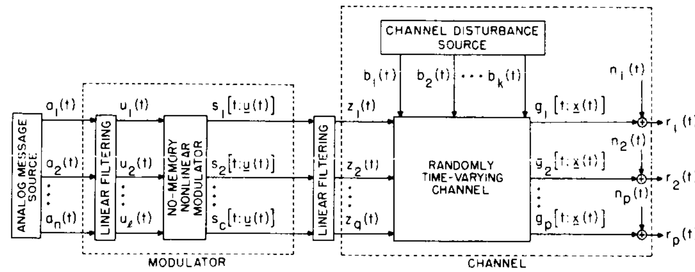


Fig. 2. The communication model.

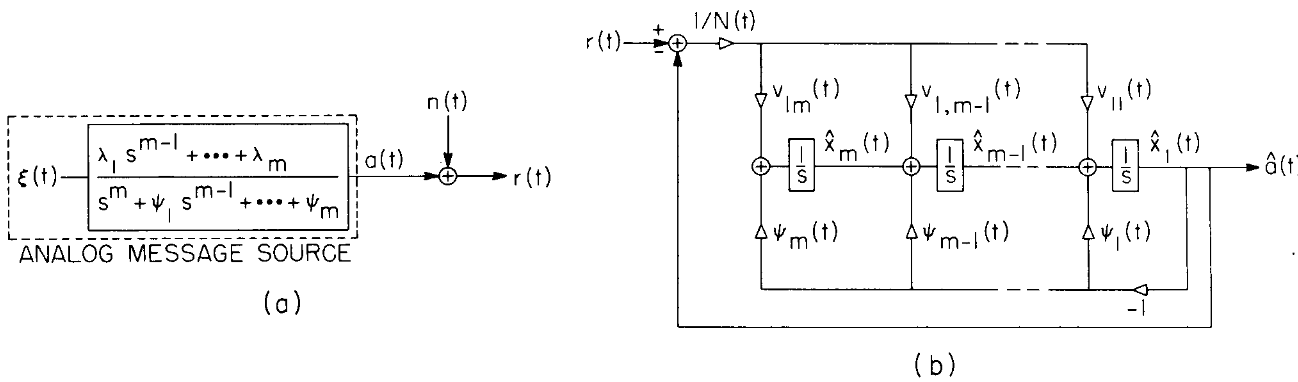


Fig. 3. (a) Stationary Gaussian message observed in an additive white noise channel.
(b) Optimum filter for estimating a stationary Gaussian message observed in additive white noise: transient conditions.

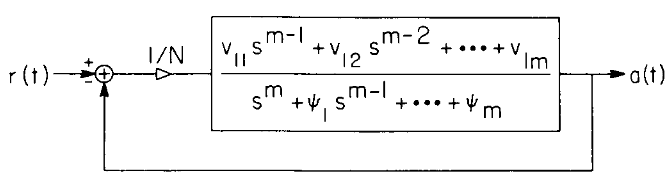


Fig. 4. Optimum filter for estimating a stationary Gaussian message observed in additive white noise: steady-state conditions.

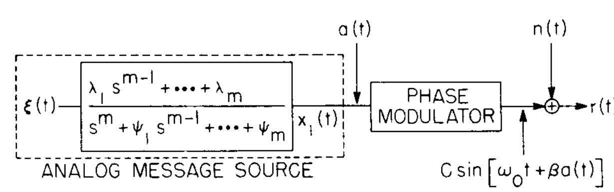


Fig. 5. Stationary Gaussian message transmitted in an additive white noise channel by phase modulation.

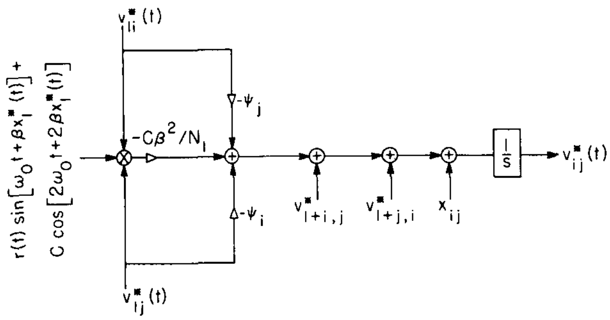


Fig. 6. A realization for the (i, j)-element of $\underline{V}^*(t)$ for the phase modulation case.

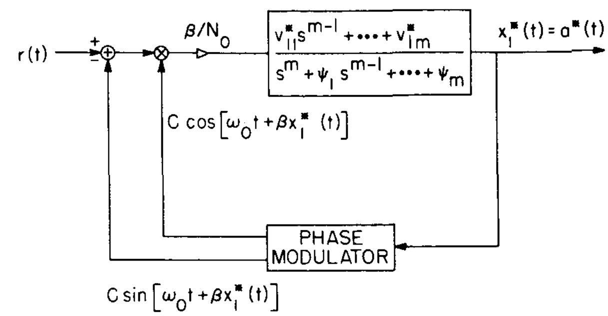


Fig. 7. Quasi-optimum demodulator for a stationary Gaussian message transmitted in an additive white noise channel by phase modulation.

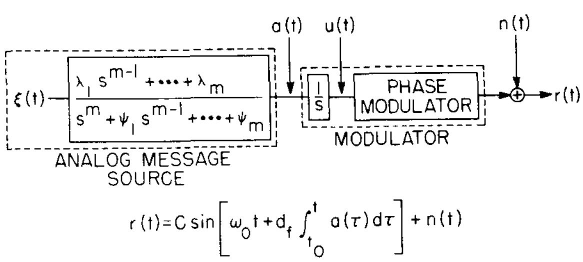


Fig. 8. Stationary Gaussian message transmitted in an additive white noise channel by frequency modulation.

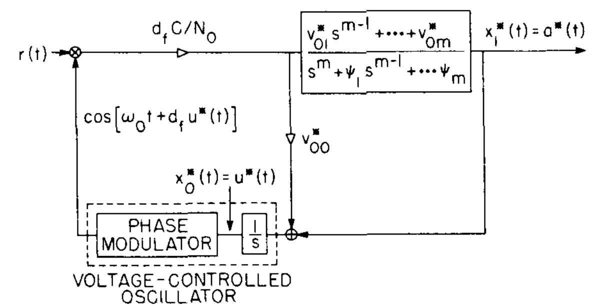


Fig. 9. Quasi-optimum demodulator for a stationary Gaussian message transmitted in an additive white noise channel by frequency modulation.

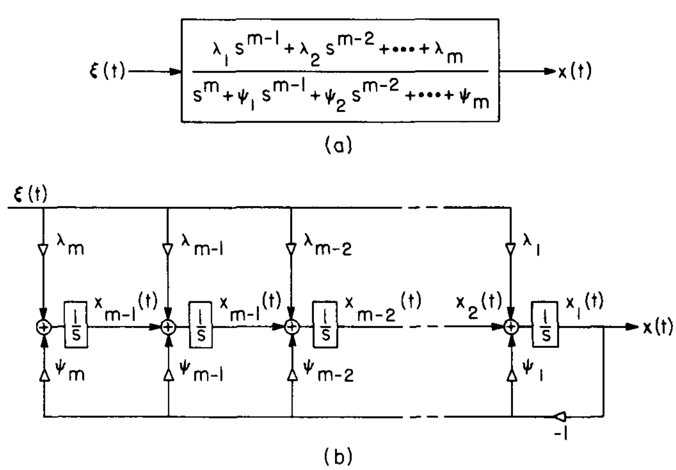


Fig. 10. Two realizations for any Gaussian process with a rational spectrum approaching zero for high frequencies.

MAXIMUM A POSTERIORI INTERVAL ESTIMATION

Arthur B. Baggeroer

Research Laboratory of Electronics
Massachusetts Institute of Technology
Cambridge, Massachusetts

N67-29904

ABSTRACT

The problem of determining the maximum a posteriori estimate of a state vector of a random process within an interval is considered. The state vector is characterized as the response of a vector differential equation to a white Gaussian forcing function. A modulator produces a signal from this state vector which is then observed over an additive white Gaussian channel.

A set of differential equations which the optimal estimate must satisfy is derived by using Lagrangian multipliers and the calculus of variations. The derivation is analogous to methods in optimal control theory.

In the case of a linear state-vector equation and a linear modulator these equations can be solved explicitly and uniquely. The estimate at the interval end point is shown to be identical to the realizable estimate, and then a convenient means of implementing the MAP receiver by using this estimate is shown. The solution to the problem of filtering with a fixed delay is also derived from the MAP estimation equations.

For the linear case, a differential equation satisfying the error is derived. From this, a differential equation for the covariance of the error matrix is derived. A solution to this equation is given by finding the appropriate integrating factor and then using the covariance of error matrix for the end-point estimate.

INTRODUCTION

The method of characterizing an optimal receiver by a set of differential, or difference, equations has been very useful. Most previous applications have been limited, however, to estimation at the end point of the interval by using just the past of the received signal. This corresponds to the realizable filtering problem. If we desire to estimate the signal over the entire observation interval by using all of the received data, we require an interval estimation procedure. This is analogous to the unrealizable filtering problem in which both past and future data are used.

One approach is to estimate coefficents of a Karhunen-Loève expansion of the message. This approach leads to a set of integral equations which the optimum estimate must satisfy. Another approach is to apply the calculus of variations to the a posteriori probability density in order to maximize it. This leads to a set of differential equations which the optimal estimate must satisfy. This approach has the advantage that the set of differential equations may be easier to implement in order to actually obtain the estimate.

In this paper we shall be concerned with maximum a posteriori (hereafter abbreviated MAP) estimation over the entire observation interval.

DERIVATION OF THE MAP ESTIMATION EQUATIONS

We shall assume that the message source may be characterized in a finite dimensional state variable form. Consequently, we represent this message source as the solution to the vector differential equation

$$\frac{d\underline{x}(t)}{dt} = \underline{f}(\underline{x}(t), \underline{u}(t), t) \quad T_o<t<T_f)$$

$$\underline{x}(T_o) = \underline{x}_o \qquad (1)$$

in which

$\underline{x}(t)$ is an n×l state-variable vector characterizing the message

$\underline{x}(T_o)$ is the n×l state-variable vector at T_o

$\underline{u}(t)$ is an m×l vector forcing function

t is the independent time variable within the interval $T_o<t<T_f$

* This work was supported by the National Aeronautics and Space Administration (Grant NsG-334), and by a National Science Foundation fellowship.

$\underline{f}$ is a (n×1) vector function of the vectors $\underline{x}(t)$ and $\underline{u}(t)$ and the scalar t.

The message source vector $\underline{x}(t)$ is used to produce a signal vector $\underline{s}(t;\underline{x}(t))$, which is transmitted over an additive noise channel. Therefore, the received signal is given by

$$\underline{r}(t) = \underline{s}(t;\underline{x}(t) + \underline{w}(t). \qquad (2)$$

A diagram of the system is illustrated in Fig. 1.

For the maximum a posteriori (MAP) estimate we wish to maximize $p(\underline{x}(t)|\underline{r}(t))$ as a function of $\underline{x}(t)$ over the interval $T_o<t<T_f$, when $\underline{r}(t)$ has been observed at the receiver. We now want to show that maximizing this quantity is equivalent to maximizing the quantity $p(\underline{r}(t)|\underline{s}(t))$ $p(\underline{u}(t))$ $p(\underline{x}_o)$. To show this we first apply Bayes' rule to the MAP density. This gives

$$p(\underline{x}(t)|\underline{r}(t)) = \frac{p(\underline{r}(t)|\underline{x}(t))\ p(\underline{x}(t))}{p(\underline{r}(t))} =$$

$$c\ p(\underline{r}(t)\ \underline{x}(t))\ p(\underline{x}(t)) \qquad (3)$$

where c is independent of $\underline{x}(t)$. Now since $\underline{x}(t)$ is a state-variable vector, it is uniquely determined by the initial state $\underline{x}_o$ and the forcing function $\underline{u}(t)$. Therefore, by assumming independence of $\underline{x}_o$ and $\underline{u}(t)$, we obtain

$$p(\underline{x}(t)) = p(\underline{x}_o,\underline{u}(t)) = p(\underline{x}_o)\ p(\underline{u}(t)) \qquad (4)$$

We also note that

$$p(\underline{r}(t)|\underline{x}(t)) = p(\underline{r}(t)|\underline{s}(t,\underline{x}(t)), \qquad (5)$$

Since the observed signal density is conditioned upon the transmitted signal $\underline{s}(t;\underline{x}(t))$, which is completely determined by $\underline{x}(t)$. Finally, we have

$$p(\underline{x}(t)|\underline{r}(t) = cp(\underline{r}(t)|\underline{s}(t;\underline{x}(t))$$
$$p(\underline{u}(t))\ p(\underline{x}_o) \qquad (6)$$

with the constraints

$$\frac{d\underline{x}(t)}{dt} = \underline{f}(\underline{x}(t),\ \underline{u}(t),\ t)$$

$$\underline{x}(T_o) = \underline{x}_o\ .$$

In general the maximization of Eq. (6) is difficult; however, if the various factors are quadratic forms, the problem is considerably more tractable. With this in mind, we are led to the assumption of Gaussian distributions for $\underline{x}_o$, $\underline{u}(t)$ and $\underline{w}(t)$. We shall assume that the estimate of the state at $t = T_o$ is $\overline{x}_o$,

$$E[\underline{x}(T_o)] = \overline{\underline{x}}_o\ . \qquad (7)$$

The error in this estimate is assumed to be Gaussianly distributed with a covariance matrix of P_o

$$E[(\underline{x}(T_o) - \overline{\underline{x}}_o)(\underline{x}(T_o) - \overline{\underline{x}}_o)^T] = P_o\ . \qquad (8)$$

We now assume that the source function $\underline{u}(t)$ is a white Gaussian random process with mean $\underline{m}_u(t)$ and covariance $Q(t)\delta(t-\tau)$,

$$E[\underline{u}(t)] = \underline{m}_u(t) \qquad (9)$$

and

$$E[(\underline{u}(t)-\underline{m}_u(t))(\underline{u}(\tau)-\underline{m}_u(\tau)^T] =$$
$$Q(t)\ \delta(t-\tau) \qquad (10)$$

The final assumption is that the observation noise is also white Gaussian random process with mean $\underline{m}_w(t)$ and covariance $R(t)\delta(t-\tau)$

$$E[\underline{w}(t)] = \underline{m}_w(t) \qquad (11)$$

and

$$E[(\underline{w}(t)-\underline{m}_w(t))(\underline{w}(\tau)-\underline{m}_w(\tau))^T] =$$
$$R(t)\ \delta(t-\tau).$$

With these assumptions, we have

$$p(\underline{x}_o)\ \alpha\ e^{-\frac{1}{2}[\underline{x}_o-\ \overline{\underline{x}}_o]^T P_o^{-1}[\underline{x}_o-\overline{\underline{x}}_o]} \qquad (12)$$

$$p(\underline{u}(t))\ \alpha\ e^{-\frac{1}{2}\int_{T_o}^{T_f}[\underline{u}(t)-\underline{m}_u(t)]^T Q^{-1}(t)[\underline{u}(t)-\underline{m}_u(t)]dt} \qquad (13)$$

$$p(\underline{r}(t)|\underline{s}(t:\underline{x}(t))\ \alpha\ e^{-\frac{1}{2}\int_{T_o}^{T_f}[\underline{r}(t)-\underline{m}_w(t)-\underline{s}(t:\underline{x}(t))]^T R^{-1}(t)[\underline{r}(t)-\underline{m}_w(t)-\underline{s}(t:\underline{x}(t))]dt} \quad (14)$$

Instead of maximizing the quantity $p(\underline{r}(t)|\underline{s}(t:\underline{x}(t))\ p(\underline{u}(t))\ p(\underline{x}_o)$, we can minimize the negative of its logarithm. Consequently, we wish to minimize the functional

$$J(\underline{u}(t),\underline{x}_o) = \frac{1}{2}[\underline{x}_o-\overline{\underline{x}}_o]^T P_o^{-1}[\underline{x}_\sigma\overline{\underline{x}}_o] + \frac{1}{2}\int_{T_o}^{T}\{(\underline{r}(t)-\underline{m}_w(t)-s(t;\underline{x}(t))^T R^{-1}(t)$$
$$(\underline{r}(t)-\underline{m}_w(t)-\underline{s}(t;\underline{x}(t)) +$$
$$(\underline{u}(t)-\underline{m}_u(t))^T Q^{-1}(t)(\underline{u}(t)-\underline{m}_u(t))\}\ dt, \quad (15)$$

with constraint

$$\frac{d\underline{x}(t)}{dt} = f(\underline{x}(t),\underline{u}(t),t) \quad T_o<t<T_f$$
$$\underline{x}(T_o) = \underline{x}_o\ .$$

We can incorporate the constraints by using the Lagrangian multiplier technique. Therefore, $J(\underline{u}(t), \underline{x}_o)$ becomes

$$J(\underline{u}(t),\underline{x}_o) = \frac{1}{2}[\underline{x}_o- \overline{\underline{x}}_o]^T P_o^{-1}[\underline{x}_\sigma\overline{\underline{x}}_o] + \frac{1}{2}\int_{T_o}^{T}\{(\underline{r}(t)-\underline{m}_w(t)-\underline{s}(t;\underline{x}(t))^T$$
$$R^{-1}(t)(\underline{r}(t) - \underline{m}_w(t) - \underline{s}(t;\underline{x}(t)) +$$
$$(\underline{u}(t)-\underline{m}_u(t))^T Q^{-1}(t)(u(t)-m_u(t)) +$$
$$\underline{p}^T(t)\ (\frac{d\underline{x}(t)}{dt} - f(\underline{x}(t),\underline{u}(t), t))\}\ dt \quad (16)$$

Before considering the minimization procedure, we want to examine the term

$$\int_{T_o}^{T_f} \underline{p}^T(t)\ \frac{d\underline{x}(t)}{dt}\ dt.$$

Integrating this by parts yields

$$\int_{T_o}^{T_f} \underline{p}^T(t)\ \frac{d\underline{x}(t)}{dt}\ dt = \underline{p}^T(T_f)\underline{x}(T_f) - \underline{p}^T(T_o)\underline{x}(T_o)- \int_{T_o}^{T_f}\frac{d\underline{p}^T(t)}{dt}\ \underline{x}(t)\ dt \quad (17)$$

Now, let us denote the optimal estimates of $\underline{x}_o$, $\underline{x}(t)$, $\underline{u}(t)$, and $\underline{s}(t)$ by $\hat{\underline{x}}_o$, $\hat{\underline{x}}(t)$, $\hat{\underline{u}}(t)$, and $\hat{\underline{s}}(t)$, respectively. Now let us **extend** $J(\underline{u}(t),\ x_o)$ around the optimal **estimates.** We get

$$\underline{u}(t) = \hat{\underline{u}}(t) + \varepsilon\delta\underline{u}(t). \quad (18)$$

The response of the message source to this

$$\underline{x}(t) = \hat{\underline{x}}(t) + \varepsilon\delta\underline{x}(t) + O(\varepsilon), \quad (19)$$

where

$$\frac{d}{dt}(\delta\underline{x}) = \frac{\partial \underline{f}}{\partial \underline{x}}\Big|_\wedge \delta\underline{x} + \frac{\partial \underline{f}}{\partial \underline{u}}\Big|_\wedge \delta\underline{u}(t)$$
$$\delta\underline{x}(T_o) = \delta\underline{x}_o$$
$$O(\varepsilon) \rightarrow 0 \quad \text{at least as } \varepsilon^2 \text{ as } \varepsilon\rightarrow 0.$$

The resulting signal $\underline{s}(t;\underline{x}(t))$

$$\underline{s}(t) = \hat{\underline{s}}(t) + \varepsilon\frac{\partial \underline{s}}{\partial \underline{x}}\Big|_\wedge \delta\underline{x} + O(\varepsilon) \quad (20)$$

[The notation $\frac{\partial \underline{f}}{\partial \underline{x}}\Big|_\wedge$ is interpreted as $(\nabla_{\underline{x}}\underline{f})$ evaluated along the optimum trajectory, i.e., the derivative with respect to each component of $\underline{x}$ of each component of $\underline{f}$. The result is an n×n matrix.]

Consequently, up to terms of order ε we have

$$J(\underline{u}(t), \underline{x}_o) = J(\underline{\hat{u}}(t), \underline{\hat{x}}_o) +$$

$$\varepsilon\{[\underline{\hat{x}}_o - \underline{\bar{x}}_o]^T P_o^{-1} \delta \underline{x}_o +$$

$$\int_{T_o}^{T_f} [-(\underline{r}(t)-\underline{m}_w(t)-\underline{\hat{s}}(t,\underline{x}(t))^T R^{-1}(t) \times$$

$$\frac{\partial \underline{s}}{\partial \underline{x}}\Big|_\Lambda \delta\underline{x}(t) + (\underline{\hat{u}}(t)-\underline{m}_u(t))\ Q^{-1}(t)$$

$$\delta\underline{\hat{u}}(t) - \frac{d\underline{p}(t)}{dt}\ \delta\underline{x}(t)-\underline{p}^T(t)\ \frac{\partial \underline{f}}{\partial \underline{x}}\Big|_\Lambda$$

$$\delta\underline{x}(t)-\underline{p}^T(t)\ \frac{\partial \underline{f}}{\partial \underline{u}}\Big|_\Lambda\ \delta\underline{u}(t)]\ dt +$$

$$\underline{p}^T(T_f)\ \delta\underline{x}(T_f)-\underline{p}^T(T_o)\ \delta\underline{x}(T_o)\} \tag{21}$$

Since $\underline{\hat{u}}(t)$ and $\underline{\hat{x}}_o$ are the optimum estimates, we must have

$$J(\underline{u}(t), \underline{x}_o)-J(\underline{\hat{u}}(t), \underline{\hat{x}}_o) \geq 0. \tag{22}$$

Therefore, by combining the various variations, we have

$$J(\underline{u}(t), \underline{x}_o) - J(\underline{\hat{u}}(t), \underline{\hat{x}}_o) =$$

$$\varepsilon\{([\underline{\hat{x}}_o - \underline{\bar{x}}_o]^T P_o^{-1} - p^T(T_o))\ \delta\underline{x}_o +$$

$$\int_{T_o}^{T_f} \{(-[\underline{r}(t)-\underline{m}_w(t)-\underline{\hat{s}}(t:\underline{x}(t))]^T R^{-1}(t)$$

$$\frac{\partial \underline{s}}{\partial \underline{x}}\Big|_\Lambda - \frac{d\underline{p}^T(t)}{dt} - \underline{p}^T(t)\ \frac{\partial \underline{f}}{\partial \underline{x}}\Big|_\Lambda)\ \delta\underline{x}(t) +$$

$$(\underline{\hat{u}}(t)-\underline{m}_u(t))\ \phi^{-1}(t)-\underline{p}^T(t)\ \frac{\partial \underline{f}}{\partial \underline{u}}\Big|_\Lambda)$$

$$\delta\underline{u}(t)\}\ dt + \underline{p}^T(T_f)\ \delta\underline{x}(T_f)\}. \tag{23}$$

Since ε is arbitrary and we may neglect the terms of higher order, this last term, the factor multiplied by ε, must be identically equal to zero in order for $\underline{\hat{u}}(t)$ and $\underline{\hat{x}}_o$ to be optimum.

We now require that $\underline{p}(t)$ satisfy the differential equation

$$\frac{d\underline{p}^T(t)}{dt} = -\underline{p}^T(t)\ \frac{\partial \underline{f}}{\partial \underline{x}}\Big|_\Lambda +$$

$$(\underline{r}(t)-\underline{m}_w(t)-\underline{\hat{s}}(t:\underline{x}(t)))^T R^{-1}(t)\ \frac{\partial \underline{s}}{\partial \underline{x}}\Big|_\Lambda \tag{24}$$

or equivalently

$$\frac{d\underline{p}(t)}{dt} = -\frac{\partial \underline{f}^T}{\partial x}\Big|_\Lambda \underline{p}(t) + \frac{\partial \underline{s}}{\partial \underline{x}}\Big|_\Lambda R^{-1}(t) \times$$

$$(\underline{r}(t)-\underline{m}_w(t) - \underline{s}(t;\ \underline{\hat{x}}(t))). \tag{25}$$

As a boundary condition on $\underline{p}(t)$, we also require

$$\underline{p}(T_f) = 0. \tag{26}$$

With this restriction on $\underline{p}(t)$, we have

$$([\underline{\hat{x}}_o-\underline{\bar{x}}_o]^T P_o^{-1}-\underline{p}^T(T_o))\ \delta\underline{x}_o +$$

$$\int_{T_o}^{T_f} (\underline{\hat{u}}(t)-\underline{m}_u(t))^T Q^{-1}(t)-\underline{p}^T(t)\ \frac{\partial \underline{f}}{\partial \underline{u}}\Big|_\Lambda)$$

$$\delta\underline{u}(t)\ dt = 0 \tag{27}$$

Now $\delta\underline{x}_o$ and $\delta\underline{u}(t)$ are arbitrary. Therefore,

$$[\hat{x}_o-\bar{x}_o]^T P_o^{-1} - \underline{p}^T(T_o) = 0 \tag{28}$$

and

$$(\underline{\hat{u}}(t)-\underline{m}_u(t))^T Q^{-1}(t)-\underline{p}^T(t)\ \frac{\partial \underline{f}}{\partial \underline{u}}\Big|_\Lambda = 0 \tag{29}$$

Equivalently,

$$(\underline{\hat{x}}_o - \underline{\bar{x}}_o) = P_o\ \underline{p}(T_o) \tag{28a}$$

and

$$\underline{\hat{u}}(t)-\underline{m}_u(t) = Q(t)\ \frac{\partial \underline{f}^T}{\partial \underline{u}}\Big|_\Lambda \underline{p}(t) \tag{29a}$$

By using this last equation, we can solve for $\underline{u}(t)$ and eliminate it in the equations which the optimal estimate must satisfy. Summarizing the results, we have

$$\frac{d\hat{\underline{x}}(t)}{dt} = \underline{f}(\hat{\underline{x}}(t),\underline{m}_u(t) + Q(t) \left.\frac{\partial \underline{f}}{\partial \underline{u}}\right|_{\wedge} \underline{p}(t)), \tag{30}$$

$$\frac{d\underline{p}(t)}{dt} = -\left.\frac{\partial \underline{f}}{\partial \underline{x}}\right|_{\wedge} \underline{p}(t) + \left.\frac{\partial \underline{s}}{\partial \underline{x}}\right|_{\wedge} R^{-1}(t) \times$$

$$(\underline{r}(t)-\underline{m}_w(t)-\underline{s}(t, \hat{\underline{x}}(t)), \tag{31}$$

and

$$\underline{u}(t) = \underline{m}_u(t) + Q(t) \left.\frac{\partial \underline{f}}{\partial \underline{u}}\right|_{\wedge} \underline{p}(t),$$

with

$$\underline{p}(T_f) = 0$$

$$(\hat{\underline{x}} - \bar{\underline{x}}_o) = P_o\, \underline{p}(T_o).$$

These are essentially the equations derived by Bryson and Frazier. The derivation, however, is complete in that we have derived the equations that must be satisfied at an extremum of the functional J. We have not shown that this extremum is in fact a minimum. There is an analogy here with the minimum principle for an optimal control problem. The major difference is that $\hat{\underline{x}}(t)$ is unconstrained at both ends of the interval. If we wished, however, to convert this problem into an exact dual of the control problem, we could require that $\hat{\underline{x}}(T_f)$ equal $\hat{\underline{x}}_{filt}(T_f)$, where $\hat{\underline{x}}_{filt}(T_f)$ the optimal filtered, or end point estimate, at T_f. We can impose this condition because both estimates $\hat{x}(T_f)$ and $\hat{x}_{filt}(T_f)$ have the same amount of data to operate upon. For further discussion of the relevence to optimal control problems, we refer to ref. 3.

As a consequence of the derivation, we find that in order to obtain the MAP estimate, we must solve a z_n-dimensional matrix differential equation. In general these equations are nonlinear and there is no general technique available to solve them. The difficult feature of implementing them on a computer is the boundary condition

$$\underline{p}(T_f) = 0.$$

We conclude by illustrating the estimator equations for a phase modulation system. Consider the following system:

$$\frac{dx(t)}{dt} = -k\, x(t) + u(t)$$

$$E[u(t)] = 0$$

$$E[x(T_o)] = 0$$

$$E[x^2(T_o)] = P$$

$$E[u(t)\, u(\tau)] = 2kP\, \delta(t-\tau)$$

or $\quad E[x^2(t)] = P$

$$r(t) = A \sin(\omega_o t + \beta\, (t))$$

$$E[\omega(t)] = 0$$

$$E[\omega(t)\, \omega(\tau)] = \frac{N_o}{2}\, \delta(t-\tau)$$

The estimation equations are

$$\frac{d\hat{x}(t)}{dt} = -k\hat{x}(t) + 2kP\, p(t)$$

$$\frac{dp(t)}{dt} = kp(t) - \frac{2A\beta}{N_o} \times \cos(\omega_o t+\beta\hat{\underline{x}}(t))$$

$$(r(t)- A \sin(\omega_o t + \beta\hat{x}(t))$$

with the boundary conditions

$$p(T_f) = 0$$

$$\hat{x}(T_o) = P\, p(T_o).$$

MAP INTERVAL ESTIMATION FOR LINEAR SYSTEMS

In general, analytic calculations or direct computer implementation of the estimation equations are not possible. In the case of linear systems, however, we may proceed considerably further. The assumption of linearity requires

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

(linear message source) (32)

and

$$\underline{s}(t;\underline{x}(t)) = C(t)\,\underline{x}(t)$$

{linear modulation}

(33)

Consequently, the estimation equations are

$$\frac{d\hat{\underline{x}}(t)}{dt} = F(t)\hat{\underline{x}}(t) + G(t)\,Q(t)\,G^T(t)\underline{p}(t) + G(t)\,\underline{m}_u(t) \qquad (34)$$

$$\frac{d\underline{p}(t)}{dt} = C^T(t)\,R^{-1}(t)C(t)\,\hat{\underline{x}}(t) - F^T(t)\underline{p}(t) - C^T(t)\,R^{-1}(t)(\underline{r}(t) - \underline{m}_w(t)) \qquad (35)$$

In matrix form, these equations are

$$\frac{d}{dt}\begin{bmatrix}\hat{\underline{x}}(t)\\ \underline{p}(t)\end{bmatrix} = \begin{bmatrix}-F^T(t) & G(t)Q(t)G^T(t)\\ G^T(t)R^{-1}(t)G(t) & F^T(t)\end{bmatrix}\begin{bmatrix}\hat{\underline{x}}(t)\\ \underline{p}(t)\end{bmatrix} + \begin{bmatrix}G(t)\underline{m}_u(t)\\ -C^T(t)R^{-1}(t)(\underline{r}(t)-\underline{m}_w(t))\end{bmatrix} \qquad (36)$$

Because of the linearity, we can satisfy the initial and final boundary conditions by the superposition of homogeneous and particular solutions to the estimation equations.

Before proceeding, we want to discuss briefly the homogeneous system. For convenience, let us denote the matrix

$$W(t) = \begin{bmatrix}F(t) & G(t)Q(t)G^T(t)\\ C^T(t)R^{-1}(t)C(t) & -F^T(t)\end{bmatrix} \qquad (37)$$

by $W(t)$. Let $\Theta(t,T_o)$ be the transition matrix associated with the homogeneous version of the estimation equations,

$$\frac{d\,\Theta(t,T_o)}{dt} = W(t)\;\Theta(t,T_o) \qquad (38)$$

with

$$\Theta(T_o, T_o) = I_{2n}$$

At this point we with to emphasize the importance of this transition matrix. Virtually everything associated with linear estimation of state variable can be related to or determined from it. We also will have cause to consider a partition of this matrix of the form

$$\Theta(t, T_o) = \left[\begin{array}{c|c}\Theta_{xx}(t,T_o) & \Theta_{xp}(t,T_o)\\ \hline \Theta_{px}(t,T_o) & \Theta_{pp}(t,T_o)\end{array}\right], \qquad (39)$$

where the matrices of the partition are n×n. We now consider implementing the estimation equations in order to obtain the MAP estimate of $\hat{\underline{x}}(t)$. Let $\underline{x}_p(t)$ and $\underline{p}_p(t)$ be solution to the estimation equations with the initial conditions

$$\underline{x}_p(T_o) = \bar{\underline{x}}_o \qquad (40)$$

$$\underline{p}_p(T_o) = \underline{0}\,. \qquad (41)$$

In order to satisfy the conditions

$$\hat{\underline{x}}(T_o) - \bar{\underline{x}}_o = P_o\underline{p}(t_o)$$

$$\underline{p}_p(T_f) = 0,$$

we add to $\underline{x}_p(t)$ and $\underline{p}_p(t)$, a linear combination of the columns of the transition matrix

$$\begin{bmatrix}\hat{\underline{x}}(t)\\ \hline \underline{p}(t)\end{bmatrix} = \begin{bmatrix}\hat{\underline{x}}_p(t)\\ \hline \underline{p}_p(t)\end{bmatrix} + \Theta(t,T_o)\begin{bmatrix}\underline{a}\\ \hline \underline{b}\end{bmatrix} \qquad (42)$$

or equivalently

$$\hat{\underline{x}}(t) = \underline{x}_p(t) + \Theta_{xx}(t, T_o)\,\underline{a} + \Theta_{xp}(t,T_o)\,\underline{b} \qquad (43)$$

$$\underline{p}(t) = \underline{P}_p(t) + \Theta_{px}(t,T_o)\,\underline{a} + \Theta_{pp}(t,T_o)\,b. \tag{44}$$

Applying the initial boundary condition yields

$$P_o b = \underline{a}\ . \tag{45}$$

Applying the final boundary condition yields

$$\underline{b} = -[\Theta_{px}(T_f,T_o)P_o + \Theta_{pp}(T_f,T_o)]^{-1}\,\underline{p}_p(T_f) \tag{46}$$

Therefore, if we let

$$\Phi_x(t,T_o) = \Theta_{xx}(t,T_o)\,P_o + \Theta_{xp}(t,T_o). \tag{47}$$

$$\Phi_p(t,T_o) = \Theta_{px}(t,T_o)\,P_o + \Theta_{pp}(t,T_o), \tag{48}$$

the solutions to the estimation equations are

$$\underline{\hat{x}}(t) = \underline{x}_p(t) - \Phi_x(t,T_o)\,\Phi_p^{-1}(T_f,T_o)\underline{p}_p(T_f) \tag{49}$$

$$\underline{p}(t) = \underline{p}_p(t) - \Phi_p(t,T_o)\,\Phi_p^{-1}(T_f,T_o)\,\underline{p}_p(T_f) \tag{50}$$

Notice that we could have specified $\Phi_x(t), T_o)$ and $\Phi_p(t,T_o)$ as solutions to the homogeneous equation with initial conditions of P_o and I_n, respectively. This would not indicate the relation to the transition matrix of the equations, and, therefore, to the other aspects of the linear estimation problem. Since the transition matrix may be precomputed, the only terms that must be computed by using the received signal $\underline{r}(t)$ are $\underline{x}_p(t)$ and $\underline{p}_p(t)$.

In spite of the apparent simplicity of the solution, there are practical problems in its implementation. The system of equations represented by the estimator equations is an unstable system. Consequently, for large time intervals, i.e., (T_F-T_o) is "large," $\underline{x}_p(t)$ and $\underline{p}_p(t)$ become rather large. Therefore, to find the MAP estimate $\underline{\hat{x}}(t)$, one must take the difference of two large numbers, which implies computational difficulties. Another difficulty is that $\Phi_x(t,T_o)$ and $\Phi_p(t,T_o)$ tend to approach singular matrices for large $(t-T_o)$. This leads to inaccuracies in the matrix inversion required for the MAP estimate. As a result of these difficulties, discussed qualitatively here, we are led to another means of implementing the solution to the equations.

If we consider the estimate at T_f, the end point of the interval, we see that this estimate is based upon only past data. Therefore, this estimate should correspond exactly to the realizable filter estimate as formulated by Kalman and Bucy.

To prove this, we consider the estimate at T_f. This is given by

$$\underline{\hat{x}}(T_f) = \underline{x}_p(T_f) - \varepsilon(T_f,T_o)\,\underline{p}_p(T_f) \tag{51}$$

where

$$\varepsilon(T_f,T_o) = \Phi_x(T_f,T_o)\,\Phi_p^{-1}(T_f,T_o) \tag{52}$$

We now will derive differential equations which $\underline{\hat{x}}(T_f)$ and $\varepsilon(T_f,T_o)$ must satisfy. If we differentiate the equation for $\varepsilon(T_f,T_o)$ with respect to T_f, and make use of the differential equations which $\Phi_x(T_f,T_o)$ and $\Phi_p(T_f,T_o)$ satisfy, we obtain

$$\frac{d}{dT_f}\left(\varepsilon(T_f,T_o)\,\Phi_p(T_f,\ T_o)\right) = \frac{d\ \ \Phi_x(T_f,T_o)}{dT_f} \tag{53}$$

$$\frac{d\varepsilon(T_f,T_o)}{dT_f}\,\Phi_p(T_f,\ T_o) + \varepsilon(T_f,\ T_o) \times [C'^T(T_f)R^{-1}(T_f)C(T_f)\,\Phi_x(T_f,T_o) - F^T(T_f)\,\Phi_p(T_f,\ T_o)] = F(T_f)\,\Phi_x(T_f,T_o) + G(T_f)\,\phi(T_f)\,G^T(T_f)\,\Phi_p(T_f,T_o)] \tag{54}$$

Multiplying both sides of above by $\Phi_p(T_f,\ T_o)$ gives

$$\frac{d\varepsilon(T_f,T_o)}{dT_f} = F(T_f)\varepsilon(T_f,T_o) + \varepsilon(T_f,T_o) \times$$

$$F^T(T_f) - \varepsilon(T_f,T_o)\, C(T_f)R^{-1}(T_f) \times$$

$$C(T_f)\ \varepsilon(T_f,T_o) + G(T_f)\ Q(T_f)\ F^T(T_f). \tag{55}$$

with the initial condition

$$\varepsilon(T_o,\ T_o) = P_o\ . \tag{56}$$

We now differentiate the estimation equation of $x(T_f)$. This gives

$$\frac{d\underline{\hat{x}}(T_f)}{dT_f} = \frac{d\underline{\hat{x}}_p(T_f)}{dT_f} - \frac{d\varepsilon(T_f,T_o)}{dT_f} \times$$

$$\underline{p}_p(T_f) - \varepsilon(T_f,T_o)\ \frac{d\underline{p}_p(T_f)}{dT_f} \tag{57}$$

Substituting the various expressions for the derivatives, we obtain

$$\frac{d\underline{\hat{x}}(T_f)}{dT_f} = F(T_f)\ \underline{\hat{x}}(T_f) + \varepsilon(T_f,T_o) \times$$

$$C^T(T_f)R^{-1}(T_f)\ (\underline{r}(T_f)-C(T_f)\underline{\hat{x}}(T_f)) \tag{58}$$

with

$$\underline{\hat{x}}(T_f)\Big|_{T_f=T_o} = \underline{\bar{x}}_o$$

Equation 56 and 58 are exactly those derived by Kalman and Bucy for the optimal realizable filter. The end point estimate $\hat{x}(T_f)$ is completely specified by the differential equation 56. In turn, this differential equation is completely specified by $\varepsilon(T_f,T_o)$, the solution of Eq. 56. In the filtering context, $\varepsilon(T_f,T_o)$ has been shown to be the error covariance matrix of the estimate of $\underline{\hat{x}}(T_f)$. We note the important point here that the transition matrix of the MAP estimator equations, completely specifies $\varepsilon(T_f,T_o)$ and consequently the structure of the optimal filter, i.e., once we have determined this transition matrix we can completely specify both the optimal filter and optimal MAP interval estimator for arbitrary p_o and $\underline{\bar{x}}_o$.

Equation 56 is a matrix Riccate equation. Levin has described a method of solving this type of equation by associating a set of linear equations with it.[4] We have found that this set of linear equations is identical with the homogeneous version of those specifying the MAP estimate over the interval.

The optimal filter allows us to solve for the MAP estimate in a very convenient fashion. We first perform optimal filtering upon the data to obtain $\underline{\hat{x}}_{filt}(T_f)$. We then use this estimate to solve the estimator equations backward in time from T_f by using the complete set of boundary conditions at T_f.

$$\underline{\hat{x}}(T_f) = \underline{\hat{x}}_{filt}(T_f)$$

$$\underline{p}(T_f) = 0\ . \tag{59}$$

In many practical cases, one wishes to do filtering with a fixed delay, not smoothing over the entire interval $T_o<t<T_f$. By using the MAP estimation equations in conjunction with the realizable, or end-point filter, we can determine a structure for such a filter with fixed delay.

For a given set of boundary conditions at some t', the solution to the estimation equations (not necessarily the optimal solution) at t can be written

$$\begin{bmatrix} \underline{x}(t) \\ \underline{p}(t) \end{bmatrix} = \Theta(t,t_o) \begin{bmatrix} \underline{x}(t_o) \\ \underline{p}(t_o) \end{bmatrix} +$$

$$\int_{t_o}^{t} \Theta(t,\tau) \begin{bmatrix} G(\tau)\ m_u(\tau) \\ -C^T(\tau)R^{-1}(\tau)(\underline{r}(\tau)-\underline{m}_w(\tau)) \end{bmatrix} d\tau \tag{60}$$

Now consider the optimal estimate at $t = t_f-\Delta$, where Δ is a fixed delay $(T_f-T_o)>\Delta$. By using the boundary conditions specified by the optimal filter, this estimate is given by

$$\begin{bmatrix}\hat{\underline{x}}(T_f-\Delta)\\ \hat{\underline{p}}(T_f-\Delta)\end{bmatrix} = \Theta(T_f-\Delta,\ T_f)\begin{bmatrix}\hat{\underline{x}}_{filt}(T_f)\\ 0\end{bmatrix} - \int_{T_f-\Delta}^{T_f}\Theta(T_f-\Delta,\ \tau)\ \times \begin{bmatrix}G(\tau)\ \underline{m}_u(\tau)\\ -C^T(\tau)R^{-1}(\tau)(\underline{r}(\tau)-\underline{m}_u(\tau))\end{bmatrix}d\tau \qquad (61)$$

If we differentiate this equation with respect to T_f, we obtain

$$\frac{d}{dT_f}\begin{bmatrix}\hat{\underline{x}}(T_f-\Delta)\\ \underline{p}(T_f-\Delta)\end{bmatrix} = \frac{d}{dT_f}(\Theta(T_f-\Delta,\ T_f))\begin{bmatrix}\hat{\underline{x}}_{filt}(T_f)\\ 0\end{bmatrix} + \Theta(T_f-\Delta,T_f)\begin{bmatrix}\dfrac{dx_{filt}(T_f)}{dT_f}\\ 0\end{bmatrix} - \Theta(T_f-\Delta,\ T_f)\begin{bmatrix}G(T_f)\underline{m}_u(T_f)\\ -C^T(T_f)R^{-1}(T_f)(\underline{r}(T_f)-\underline{m}_w(T_f))\end{bmatrix}$$

$$+ \begin{bmatrix}G(T_f-\Delta)\ \underline{m}_u(T_f-\Delta)\\ -C^T(T_f-\Delta)R^{-1}(T_f-\Delta)(\underline{r}(T_f-\Delta)-m_w(T_f-\Delta)\end{bmatrix} - \int_{T_f-\Delta}^{T_f}\frac{d}{dT_f}\Theta(T_f-\Delta,\tau)\begin{bmatrix}G(\tau)\ \underline{m}_u(\tau)\\ -C(\tau)R^{-1}(\tau)(\underline{r}(\tau)-\underline{m}_w(\tau)\end{bmatrix}d\tau \qquad (62)$$

This last term can be evaluated by using the relation

$$\frac{d\Theta}{dT_f}(T_f-\Delta,\tau) = W(T_f-\Delta)\ \Theta(T_f-\Delta,\tau). \qquad (63)$$

Since $W(T_f-\Delta)$ is independent of the integration variable τ, it can be taken outside the integral. The value of the resulting integral is given, however, by our original expression for the optimal estimate. Consequently, this last term is given by

$$-\int_{T_f-\Delta}^{T_f}\frac{d\Theta(T_f-\Delta,\tau)}{dT_f}\begin{bmatrix}G(\tau)\ \underline{m}_u(\tau)\\ -C(\tau)R^{-1}(\tau)(r(\tau)-m_w(\tau))\end{bmatrix}d\tau$$

$$W(T_f-\Delta)\left\{\begin{bmatrix}\hat{\underline{x}}(T_f-\Delta)\\ \underline{p}(T_f-\Delta)\end{bmatrix} - \Theta(T_f-\Delta,\ T_f)\ \times \begin{bmatrix}\hat{\underline{x}}_{filt}(T_f)\\ 0\end{bmatrix}\right\} \qquad (64)$$

By substituting this in Eq. 63, we obtain the differential equation

$$\frac{d}{dT_f}\begin{bmatrix}\hat{\underline{x}}(T_f-\Delta)\\ \underline{p}(T_f-\Delta)\end{bmatrix} = \frac{d}{dT_f}(\Theta(T_f-\Delta,T_f)) + \begin{bmatrix}\hat{\underline{x}}_{filt}(T_f)\\ 0\end{bmatrix} + \Theta(T_f-\Delta,T_f)\begin{bmatrix}\dfrac{d\underline{x}_{filt}(T_f)}{dT_f} - G(T_f)\ \underline{m}_u(T_f)\\ C^T(T_f)R^{-1}(T_f)(\underline{r}(T_f)-\underline{m}_w(T_f))\end{bmatrix}$$

$$\begin{bmatrix}G(T_f-\Delta)\ \underline{m}_u(T_f-\Delta)\\ -C^T(T_f-\Delta)R^{-1}(T_f-\Delta)(\underline{r}(T_f-\Delta)-\underline{m}_w(T_f-\Delta)\end{bmatrix} + W(T_f-\Delta)\begin{bmatrix}\hat{x}(T_f-\Delta)\\ \underline{p}(T_f-\Delta)\end{bmatrix} - \Theta(T_f-\Delta,\ T_f)\begin{bmatrix}\hat{\underline{x}}_{filt}(T_f)\\ 0\end{bmatrix} \qquad (65)$$

In order to specify the initial conditions at $T_f = \Delta$, one must actually solve for the MAP estimate over the interval $(T_o, T_o+\Delta)$. The form of the filter structure is rather complex; however, even in the infinite interval, stationary case the filter structure is usually complex. In the case of a time-invariant system the equation simplifies considerably. In this case

$$\Theta\ (t,t_o) = e^{W(t-t_o)} . \tag{66}$$

By assuming $\underline{m}_u(t) = \underline{m}_w(t) = 0$, we obtain

$$\frac{d}{dT_f}\left[\begin{array}{c}\hat{\underline{x}}(T_f-\Delta)\\ \hline \underline{p}(T_f-\Delta)\end{array}\right] = e^{-W\Delta}$$

$$\left[\begin{array}{c}\dfrac{d\hat{\underline{x}}_{filt}(T_f)}{dT_f} - F\ \hat{\underline{x}}_{filt}(T_f)\\ \hline C\ R^{-1}(\underline{r}(T_f)-C\ \hat{\underline{x}}_{filt}(T_f))\end{array}\right] +$$

$$W\left[\begin{array}{c}\hat{\underline{x}}(T_f-\Delta)\\ \hline \underline{p}(T_f-\Delta)\end{array}\right] + \left[\begin{array}{c}0\\ \hline -C^TR^{-1}(\underline{r}(T_f-\Delta))\end{array}\right] \tag{67}$$

for the optimal estimate with a fixed delay from the end point of the interval. By using the result of eq. 58

$$\frac{d\hat{\underline{x}}_{filt}(T_f)}{dT_f} = F\ \hat{\underline{x}}_{filt}(T_f) =$$

$$\varepsilon(T_f,T_o)\ C^TR^{-1}(\underline{r}(T_f)-C\ \hat{\underline{x}}_{filt}(T_f)),$$

we have

$$\frac{d}{dT_f}\left[\begin{array}{c}\hat{x}(T_f-\Delta)\\ \hline \underline{p}(T_f-\Delta)\end{array}\right] = e^{-W\Delta}\left[\begin{array}{c}\varepsilon(T_f,T_o)\\ \hline I\end{array}\right] \times$$

$$C^T\ R^{-1}(\underline{r}(T_f)-C\ \underline{x}_{filt}(T_f)) +$$

$$W\left[\begin{array}{c}\hat{\underline{x}}(T_f-\Delta)\\ p(T_f-\Delta)\end{array}\right] + \left[\begin{array}{c}0\\ -C^TR^{-1}(\underline{r}(T_f-\Delta))\end{array}\right] \tag{67a}$$

where the initial conditions $\hat{\underline{x}}'(0)$ and $\underline{p}'(0)$ are determined by solving the MAP equations over the interval $(T_o, T_o+\Delta)$.

COVARIANCE OF THE ERROR MATRIX FOR LINEAR SYSTEMS

An important aspect of MAP interval estimation for linear systems is its performance. This is commonly expressed in terms of the covariance of the error matrix. We now will determine a differential equation which this matrix must satisfy. Then we will solve this differential equation in terms of the transition matrix of the estimation equations and in terms of the covariance of error matrix for the linear filtering problem.

Let us indicate the estimation error at t by

$$\underline{\varepsilon}(t) = \hat{\underline{x}}(t) - \underline{x}(t)$$

The covariance matrix is then defined as

$$\varepsilon(t,T_o) = E[\underline{\varepsilon}(t)\underline{\varepsilon}^T(t)]$$

By differentiating the expression for $\underline{\varepsilon}(t)$ and by substituting eq. (✓), we obtain

$$\frac{d\underline{\varepsilon}(t)}{dt} = \frac{d\hat{\underline{x}}(t)}{dt} - \frac{d\underline{x}(t)}{dt} = F(t)\underline{\varepsilon}(t) + G(t)\phi(t)G^T(t)\underline{p}(t) - G(t)[\underline{u}(t) - \underline{m}_u(t)] \tag{68}$$

When the expression for $\underline{r}(t)$ is substituted in the costate equation, we have

$$\frac{d\underline{p}}{dt} = C^T(t)R^{-1}(t)G(t)\underline{\varepsilon}(t) - F^T(t)\underline{p}(t) - C^T(t)R^{-1}(t)[\underline{w}(t) - \underline{m}_w(t)]$$

In matrix notation these equations, which will be called the error equations, are

$$\frac{d}{dt}\begin{bmatrix}\underline{\varepsilon}(t)\\ \hline \underline{p}(t)\end{bmatrix} = W(t)\begin{bmatrix}\underline{\varepsilon}(t)\\ \underline{p}(t)\end{bmatrix} - \begin{bmatrix}G(t)[\underline{u}(t) - \underline{m}_u(t)]\\ C^T(t)R^{-1}(t)[\underline{w}(t)-\underline{m}_w(t)]\end{bmatrix} \tag{69}$$

The boundary conditions which must be satisfied are

$$\underline{p}(T_f) = 0$$

$$\hat{\underline{x}}(T_o) - \bar{\underline{x}}_o = P_o\underline{p}(T_o)$$

Rewriting the initial boundary condition at T_o yields

$$(\hat{\underline{x}}(T_o) - \underline{x}(T_o) - (\bar{\underline{x}}_o - \underline{x}(T_o)) = P_o\underline{p}(t_o) , \tag{70a}$$

$$\text{or}\quad \underline{\varepsilon}(T_o) - \underline{\varepsilon}_I(T_o) = P_o\underline{p}(T_o) , \tag{70b}$$

where $\underline{\varepsilon}(T_o)$ is the actual error at T_o;

$\underline{\varepsilon}_I(T_o)$ is the a priori initial error.

The original hypothesis assumes $\underline{\varepsilon}_I(T_o)$ is an independent random variable with

$$E[\underline{\varepsilon}_I(T_o)\underline{\varepsilon}_I^T(T_o)] = P_o \tag{71}$$

We now want to consider briefly the solution of the error equations. The development is exactly parallel to the solution of the estimation equations. We specify the particular solutions

$$\begin{bmatrix}\underline{\varepsilon}_p(t)\\ \hline \underline{p}_p(t)\end{bmatrix}$$

to be solutions to the non-homogeneous error equations with initial conditions

$$\underline{\varepsilon}_p(T_o) = \underline{\varepsilon}_I(T_o)$$

$$\underline{p}_p(T_o) = 0$$

(It can be verified that $p_p(t)$ is the same for both the estimation and error equations.) We add to this particular solution a solution to the homogeneous version of the equations. In order that the boundary conditions be satisfied, we find that this added term is the same as the corresponding term in the estimation equations, so that the total solution is

$$\underline{\varepsilon}(t) = \underline{\varepsilon}_p(t) - \Phi_x(t,T_o)\Phi_p^{-1}(T_f,T_o)\underline{p}_p(T_f) \tag{74}$$

$$\underline{p}(t) = \underline{p}_p(t) - \Phi_p(t,T_o)\Phi_p^{-1}(T_f,T_o)\underline{p}_p(T_f) \tag{75}$$

A quick check will show that the boundary conditions are satisfied.

Now consider the expectation

$$E\left\{\left[\frac{\underline{\varepsilon}(t)}{\underline{p}(t)}\right]\left[\frac{\underline{\varepsilon}(t)}{\underline{p}(t)}\right]^T\right\} = P(t,T_o) =$$

$$\left[\begin{array}{c|c} \varepsilon(t,T_o) & P_{\varepsilon p}(t,T_o) \\ \hline P_{p\varepsilon}(t,T_o) & P_{pp}(t,T_o) \end{array}\right] , \qquad (76)$$

that is, the covariance matrix that we desire is one of the partitions of the matrix

$$\varepsilon(t,T_o) = P_{\varepsilon\varepsilon}(t,T_o) \qquad (77)$$

By differentiating the expression for $P(t,T_o)$, we have

$$\frac{dP(t,T_o)}{dt} = E\left\{\left[\frac{d}{dt}\left[\frac{\underline{\varepsilon}(t)}{\underline{p}(t)}\right]\right]\left[\frac{\underline{\varepsilon}(t)}{\underline{p}(t)}\right]^T + \right.$$

$$\left.\left[\frac{\underline{\varepsilon}(t)}{\underline{p}(t)}\right]\left[\frac{d}{dt}\left[\frac{\underline{\varepsilon}(t)}{\underline{p}(t)}\right]\right]^T\right\}. \qquad (78)$$

When (78) is substituted, we obtain

$$\frac{dP(t,T_o)}{dt} = W(t)P(t,T_o) + P(t,T_o)W^T(t) +$$

$$K(t) + K^T(t) \qquad (79)$$

where

$$K(t) = -E\left\{\left[\frac{\underline{\varepsilon}(t)}{\underline{p}(t)}\right]\begin{bmatrix} G(t)[\underline{u}(t) - \underline{m}u(t)] \\ C(t)R^{-1}(t)[\underline{w}(t) - \underline{m}_w(t)] \end{bmatrix}^T\right\}. \qquad (80)$$

We shall now determine the term, K(t). Since the transition matrix is the same for both the estimation and error equations, we may write K(t) as

$$K(t) = E\left\{\left[\Theta(t,T_o)\left[\frac{\underline{\varepsilon}(T_o)}{\underline{p}(T_o)}\right] + \right.\right.$$

$$\int_{T_o}^{t} \Theta(t,\tau)\begin{bmatrix} G(\tau)[\underline{u}(\tau) - \underline{m}u(\tau)] \\ C(\tau)R(\tau)[\underline{w}(\tau) - \underline{m}_w(\tau) \end{bmatrix} d\tau \ x$$

$$\left.\begin{bmatrix} G(t)[\underline{u}(t) - \underline{m}_u(t)] \\ C(t)R^{-1}(t)[\underline{w}(t) - \underline{m}_w(t)] \end{bmatrix}^T\right\}. \qquad (81)$$

Since $\underline{u}(t)$ and $\underline{w}(t)$ are independent white Gaussian processes, we may evaluate the second term of this expression quite easily. This is given by

$$K_2(t) = \int_{T_o}^{t} \Theta(t,\tau)$$

$$\begin{bmatrix} G(t)Q(t)G^T(t) & 0 \\ 0 & C^T(t)R^{-1}(t)C(t) \end{bmatrix}\delta(t-\tau)d\tau. \qquad (82)$$

The only nonzero part of the integrand is at the upper limit. Therefore, by making use of the symmetrical properties of the delta function, we have for this second term

$$K_2(t) = \frac{1}{2}\begin{bmatrix} G(t)Q(t)G^T(t) & 0 \\ 0 & C^T(t)R^{-1}(t)C(t) \end{bmatrix}. \qquad (83)$$

Now we must consider the first term

$$K_1(t) = -E\left\{\left[\frac{\underline{\varepsilon}(T_o)}{\underline{p}(T_o)}\right]\begin{bmatrix} G(t)[\underline{u}(t) - \underline{m}_u(t)] \\ C^T(t)R^{-1}(t)[\underline{w}(t) - \underline{m}_w)t)] \end{bmatrix}^T\right\} \qquad (84)$$

By evaluating the solution to the error equations at T_o, we have

$$\underline{\varepsilon}(T_o) = \varepsilon_I(T_o) - \Phi_x(T_o,T_o)\Phi_p^{-1}(T_f,t_o)\underline{p}_p(T_f) \qquad (85)$$

$$\underline{p}(T_o) = \quad - \Phi_p(T_o,T_o)\Phi_p^{-1}(T_f,T_o)\underline{p}_p(T_f) \qquad (86)$$

Since

$$\Phi_x(T_o,T_o) = P_o \qquad (87)$$

$$\Phi_p(T_o,T_o) = I \qquad (88)$$

these equations may be written

$$\begin{bmatrix} \underline{\varepsilon}(T_o) \\ \underline{p}(T_o) \end{bmatrix} = \begin{bmatrix} \underline{\varepsilon}_I(T_o) \\ 0 \end{bmatrix} - \begin{bmatrix} P_o \\ I \end{bmatrix} \Phi_p^{-1}(T_f,T_o)\underline{p}_p(T_f) \tag{89}$$

$$= \begin{bmatrix} \underline{\varepsilon}_I(T_o) \\ 0 \end{bmatrix} - \begin{bmatrix} P_o \\ I \end{bmatrix} \Phi_p^{-1}(T_f,T_o)[0 \vdots I] \begin{bmatrix} \underline{\varepsilon}_p(T_f) \\ \underline{p}_p(T_f) \end{bmatrix} \tag{90}$$

We are now in a position to evaluate $K_2(t)$. First, we note that when we perform the expectation, the term involving $\varepsilon_I(T_o)$ will vanish because it is independent of $\underline{u}(t)$ and $\underline{w}(t)$. Consequently, from Eq. (95) we are led to the term

$$(K'_1(t) = E\left\{ \begin{bmatrix} \underline{\varepsilon}_p(T_f) \\ \underline{p}_p(T_f) \end{bmatrix} \begin{bmatrix} G(t)[\underline{u}(t) - \underline{m}_u(t)] \\ C^T(t)R^{-1}(t)[\underline{w}(t) - \underline{m}_w(t)] \end{bmatrix}^T \right\} \tag{91}$$

We are able to evaluate this term in a manner similar to that for $K_2(t)$. First we write

$$\begin{bmatrix} \underline{\varepsilon}_p(T_f) \\ \underline{p}_p(T_f) \end{bmatrix} = \Theta(T_f,T_o) \begin{bmatrix} \underline{\varepsilon}_I(T_o) \\ 0 \end{bmatrix} - \int_{T_o}^{T_f} \Theta(T_f,\tau) \times$$

$$\begin{bmatrix} G(\tau)[\underline{u}(\tau) - \underline{m}_u(\tau)] \\ C^T(\tau)R^{-1}(\tau)[\underline{w}(\tau) - \underline{m}_w(\tau)] \end{bmatrix} d\tau \tag{92}$$

Now we perform the indicated expectation of Eq. (97). Again we note that the term involving $\underline{\varepsilon}_I(T_o)$ vanishes because of its independence. Therefore we are left with the term

$$K'_1(t) = E \; - \int_{T_o}^{T_f} \Theta(T_f,\tau)$$

$$\begin{matrix} G(\tau)[\underline{u}(\tau) - \underline{m}_u(\tau)] \\ C^T(\tau)R^{-1}(\tau)[\underline{w}(\tau) - \underline{m}_w(\tau)] \end{matrix} \; \begin{matrix} G(t)[\underline{u}(t) - \underline{m}_u(t) \\ C^T(t)R^{-1}(t)[w(t) - \underline{m}_w(t)] \end{matrix}^T d\tau \tag{93}$$

which because of the white Gaussian assumption for $\underline{u}(t)$ and $w(t)$ becomes

$$K'_1(t) = -\int_{T_o}^{T_f} \Theta(T_f,\tau)$$

$$\begin{bmatrix} G(\tau)Q(\tau)G^T(\tau) & \\ 0 & C^T(\tau)R^{-1}(\tau)C(\tau) \end{bmatrix} \delta(t-\tau)d\tau \; . \tag{94}$$

Noting that t is always within the integration region, we integrate over the δ function to obtain

$$K'_1(t) = -2\Theta(T_f,t)K_2(t) \quad . \tag{95}$$

Finally by using Eq. (95), we have

$$K_1(t) = 2\Theta(t,T_o)\begin{bmatrix} P_o \\ I \end{bmatrix} \Phi_p^{-1}(T_f,T_o)$$

$$\begin{bmatrix} 0 \\ I \end{bmatrix}^T \Theta(T_f,t)K_2(t) \tag{96}$$

The complete expression for K(t) is given by

$$K(t) = -2\Theta(t,T_o)\begin{bmatrix} P_o \\ I \end{bmatrix} \Phi_p^{-1}(T_f,T_o)$$

$$\begin{bmatrix} 0 \\ I \end{bmatrix}^T \Theta(T_f,t)K_2(t) + K_2(t) \tag{97}$$

with

$$\frac{dP(t,T_o)}{dt} = W(t)P(t,T_o) + P(t,T_o)W^T(t) +$$

$$K(t) + K^T(t) \tag{98}$$

The boundary conditions for the equation are determined from the boundary conditions for the error equations. The condition $\underline{p}_p(T_f) = 0$ implies

$$P_{pp}(T_f,T_o) = P_{\varepsilon p}(T_f,T_o) = P_{p\varepsilon}(T_f,T_o) = 0 \tag{99}$$

If we multiply the initial condition by its transpose, and then take the expected value, we obtain

$$\epsilon(T_o,T_o) - P_oP_{p\varepsilon}(T_o,T_o) - P_{\varepsilon p}(T_o,T_o)P_o +$$

$$P_oP_{pp}(T_o,T_o)P_o = P_o \tag{100}$$

The most convenient way of solving this equation is to determine $\varepsilon(T_f,T_o)$ from the variance Eq. (56), of the end-point estimate, and then solve MAP covariance equation backwards from T_f.

We shall now consider the solution to this matrix differential equation for the covariance of the error. First, we want to note some properties of the adjoint system. The adjoint system is defined to be the solution to the system

$$\frac{d\psi(t,T_o)}{dt} = W^T(t,T_o)\psi(t,T_o) \tag{101}$$

with the initial condition $\psi(T_o,T_o) = I$.
The relation to the transition matrix is given by

$$\psi^T(t,T_o) = \Theta^{-1}(t,T_o). \tag{102}$$

If we premultiply the covariance of error equation by $\psi^T(t,T_o)$ and then post-multiply by $\psi(t,T_o)$, we obtain

$$\psi^T(t,T_o)\frac{dP(t,T_o)}{dt}\psi(t,T_o) - \psi^T(t,T_o)$$

$$[W(t)P(t,T_o) + P(t,T_o)W^T(t)]\psi(t,T_o) =$$

$$\psi^T(t,T_o)[K(t) + K^T(t)]\psi(t,T_o) \tag{103}$$

The left side of Eq. 103 is a derivative, that is

$$\frac{d}{dt}[\psi^T(t,T_o)P(t,T_o)\psi(t,T_o)] =$$

$$\psi^T(t,T_o)[K(t) + K^T(t)]\psi(t,T_o) \tag{104}$$

Integrating this yields

$$\psi^T(t,T_o)P(t,T_o)\psi(t,T_o) =$$

$$P^1 + \int_{T_o}^{t} \psi^T(\tau,T_o)[K(\tau) + K^T(\tau)]\,\psi(\tau,T_o)d\tau \tag{105}$$

where P^1 is a constant to be determined. We determine this term by specifying at T_f, that

$$P(T_f,T_o) = \begin{bmatrix} \epsilon(T_f,T_o) & 0 \\ 0 & 0 \end{bmatrix}, \tag{106}$$

where $\epsilon(T_f,T_o)$ is the variance of the realizable estimate. Consequently, we have

$$\psi^T(t,T_o)P(t,T_o)\psi(t,T_o) =$$

$$\psi^T(T_f,T_o)P(T_f,T_o)\psi(T_f,T_o)$$

$$- \int_t^{T_f} \psi^T(\tau,T_o)[K(\tau) + K^T(\tau)]\psi(\tau,T_o)d\tau \tag{107}$$

We now want to show that evaluating the integral can be reduced to a single integral. Consider the term

$$L(t) = \int_t^{T_f} \psi^T(\tau,T_o)K(\tau)\psi(\tau,T_o)d\tau \tag{108}$$

(the other term is the transpose, $L^T(t)$).

This is given by

$$L(t) = \int_t^{T_f} \psi^T(\tau,T_o)K_1(\tau)\psi(\tau,T_o)d\tau =$$

$$2\int_t^{T_f} \psi^T(\tau,T_o)\Theta(\tau,T_o)\begin{bmatrix} P_o \\ I \end{bmatrix}\Phi_p^{-1}(T_f,T_o)\begin{bmatrix} 0 \\ I \end{bmatrix}^T \times$$

$$\Theta(T_f,\tau)K_2(\tau)\psi(\tau,T_o)d\tau \tag{109}$$

We note that

$$\psi^T(\tau,T_o)\Theta(\tau,T_o) = I \tag{110}$$

and

$$\Theta(T_f,\tau) = \Theta(T_f,T_o)\Theta(T_o,\tau) =$$

$$\Theta(T_f,T_o)\Theta^{-1}(\tau,T_o) = \Theta(T_f,T_o)\psi^T(\tau,T_o) \tag{111}$$

Consequently, we now have

$$L(t) = \left[I - 2\begin{bmatrix} P_o \\ I \end{bmatrix}\Phi_p^{-1}(T_f,T_o)\begin{bmatrix} 0 \\ I \end{bmatrix}^T \Theta(T_f,T_o) \right] \times$$

$$\int_t^{T_f} \psi^T(\tau,T_o)K_2(\tau)\psi(\tau,T_o)d\tau \tag{112}$$

Noting that

$$\psi^{-1^T}(t,T_o)\psi^T(T_f,T_o) = \Theta(t,T_f) \tag{113}$$

and

$$\psi(T_f,T_o)\ \psi^{-1}(t,T_o) = \theta^T(t,T_f)$$

we finally obtain

$$\begin{aligned} P(t,T_o) &= \Theta(t,T_f)P(T_f,T_o)\theta^T(t,T_f) \\ &- [\Theta(t,T_o)L(t)\theta^T(t,T_o)] - \\ &- [\theta(t,T_o)L(t)\theta^T(t,T_o)]^T \quad , \end{aligned} \tag{114}$$

where

$$L(t) = [I-2\begin{bmatrix} P_o \\ \hline I \end{bmatrix} \Phi_p^{-1}(T_f,T_o)\begin{bmatrix} 0 \\ \hline I \end{bmatrix}^T \Theta(T_f,T_o)] \quad x$$

$$\int_t^{T_f} \theta\ (T_o,\tau)K_2(\tau)\ \Theta^T(T_o,\tau)d\tau$$

$$K_2(t) = \frac{1}{2}\left[\begin{matrix} G(t)Q(t)G^T(t) & 0 \\ 0 & C^T(t)R^{-1}(t)C(t) \end{matrix}\right]. \tag{115}$$

In the case of a constant parameter system, the results simplify because of the exponential nature of the transition matrix.

CONCLUSION

We have found a set of differential equations which the optimal estimate must satisfy. These differential equations had a set of mixed boundary conditions associated with them. It is this feature that made the solution difficult to implement in the general nonlinear case.

In the linear case we could solve the equations by superposition methods. We found that a convenient method of solution was to perform a filtering operation and then solve the estimation equations backward from the end point.

By differentiating with respect to the end point, we could determine a filter with a fixed delay from the end point.

We found a set of equations which the error in the optimal estimate satisfies when the system is linear. The forcing functions for these equations were the white processes driving the message source and corrupting the observation. By performing an expectation upon these equations, we determined a differential equation involving the covariance of error matrix. We then integrated this equation to obtain the solution in terms of the end point covariance matrix and the transition matrix of the estimation equations.

REFERENCES

1. A. E. Bryson and M. Frazier, "Smoothing for Linear and Nonlinear Dynamic Systems", Proceedings Optimum Systems Conference, Wright-Patterson AFB; Sept. 1962.

2. R. E. Kalman and R. S. Bucy, "New Results in Linear Filtering and Prediction Theory", ASME Paper 60-JAC; May 31, 1960.

3. M. Athans and P. Falb, Optimal Control, New York, McGraw-Hill; 1966.

4. J. J. Levin, "On the Matrix Riccati Equation", Proc. Am. Math. Soc., Vol. 10, pp. 519-524; 1959.

5. H. L. Van Trees, Course 6.576 Notes, M.I.T., Cambridge, Mass., April 1965 (printed).

ACKNOWLEDGEMENT

This work is the outgrowth of an attempt to apply a paper by Bryson and Frazier to the processing of seismic data. Consequently, many of the results derived here may be found in ref. 1. It is hoped that this paper develops the results presented there in a clear fashion.

I also wish to acknowledge the patient encouragement of Prof. H. L. Van Trees, and the many helpful discussions with Lewis D. Collins and Donald L. Snyder.

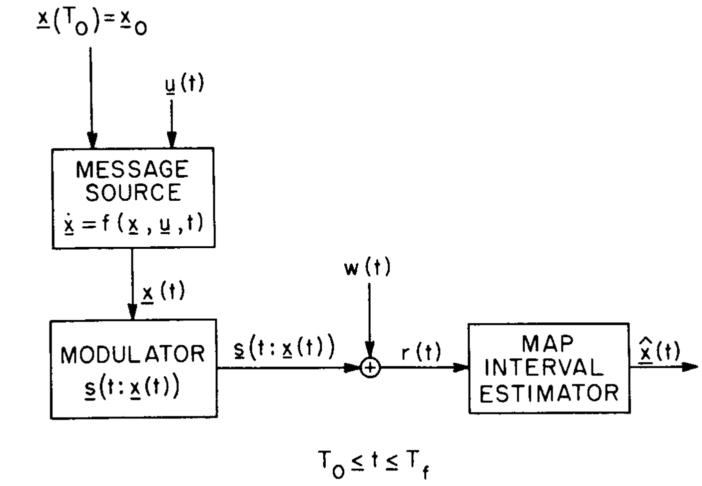


Fig. 1. Illustration of the communication model.

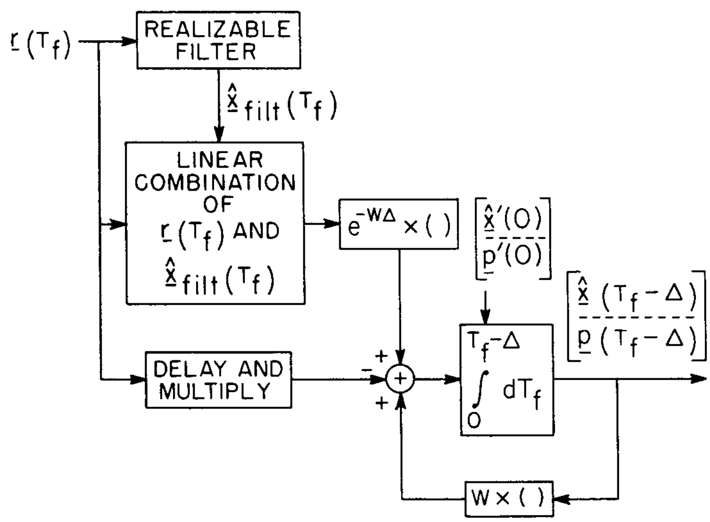


Fig. 2. Block diagram of Eq. (67a).

7/3

SIGNAL OPTIMIZATION FOR ADDITIVE NOISE CHANNELS WITH FEEDBACK*

N67-29905

Jim K. Omura,† Stanford University

SUMMARY

In this work, a communication receiver is regarded as a dynamical system described by a difference equation where the last state is the test statistics upon which a decision is based. When noiseless feedback to the transmitter is allowed it is observed that the signal selection problem is essentially a stochastic control problem. With an appropriate criterion signals are found that exploit the feedback to achieve considerable reduction in coding and decoding complexity over what would be needed for comparable performance with the best known signals for the one way channel. The schemes developed could be very important for satellite communications since it allows for a substantial decrease in the coding effort while permitting the satellite to transmit its information at a rate arbitrarily close to channel capacity.

This control theory approach depends only on the first and second order statistics of the noise, handles multiplicative noise in addition to additive noise in the forward channel, and naturally extends to consideration of noise in the feedback link.

INTRODUCTION

Recently there has been a considerable amount of interest in feedback communication systems; in particular, cases where the feedback link is noiseless. One of the main reasons for this is the advent of space communication where the power in the ground-to-satellite direction can be so much larger than in the reverse direction that the first link can be taken to be an (essentially) noiseless link. Similar situations may also arise elsewhere.

The usual approach in designing one-way communication systems is to first select signals to be used by the transmitter and then find an optimum receiver based on these signals. For example, if the transmitter is to send one of M messages at any given time over an additive Gaussian noise channel one first selects M signals to represent the messages. The minimum probability of error receiver is then designed around these signals which in this case is a linear operation on the received signal followed by a decision process. With the availability of feedback, however, one can view the problem from the opposite point of view. Namely, fix the receiver and design signals around the receiver so as to minimize probability of error. In particular, if one regards the receiver as a dynamical system whose state is the test statistics upon which a decision is based, the signal problem is essentially a stochastic control problem.

In this work, this control point of view is taken in designing signals for various channels. Motivated by the work of Schalkwijk and Kailath,[1,2] this work generalizes much of their results and extends them to consideration of multiplicative channels and noisy feedback channels.

The word "channel" stands for physical perturbation in the transmission medium and in the receiver front end, as well as for transmitter constraints. Examples of transmitter constraints are an average power constraint, a peak power constraint, a constraint on the signal bandwidth, etc.

Before discussion of the main results, a brief discussion of related work is given next.

Background

Most previous work on feedback communication systems consider only noiseless feedback. It seems reasonable that the availability of a noiseless feedback link should substantially improve communication over the noisy forward link. Therefore, Shannon's result[3] that the channel capacity of a memoryless noisy channel is not increased by noiseless feedback is rather surprising. Still, some advantages should accrue from the presence of a noiseless feedback link and, in fact, the advantage is that noiseless feedback enables a substantial reduction in the complexity of coding and decoding required to achieve a given performance over the noisy link.

A general discussion of feedback communication systems, with reference to earlier work by Chang and others, is given by Green[4] who distinguishes between post- and predecision feedback systems. In postdecision feedback systems the transmitter is informed only about the receiver's decision; in predecision feedback systems, the state of uncertainty of the receiver as to which message was sent is fed back. Postdecision feedback systems require less capacity in the backward direction; however, the improvement over one-way transmission will also be less than that obtainable with predecision feedback.

Viterbi[5] discusses a postdecision feedback system for the white Gaussian noise channel. A decision is made when the a posteriori probability computed by the receiver exceeds a certain threshold determined by the probability of error. The transmitter is informed by means of postdecision feedback that the receiver has made its decision, and it then starts sending the next message. For rates higher than half the channel capacity, the reliability is increased roughly by a factor of four as compared to one-way communication.

*This work was supported by the Air Force under Contract AF 49(683)-1517 and by the Navy under Contract Nonr 225(83).

†The author is now with the Stanford Research Institute in Menlo Park, California.

Turin[6,7] has a predecision feedback scheme applying to the white Gaussian noise channel, and giving an even greater improvement over one-way communication than Viterbi's scheme does. The receiver computes a likelihood ratio and makes a decision when this likelihood ratio exceeds a threshold set by the probability of error. The value of the likelihood ratio is fed back to the transmitter continually during the decision-making process. The transmitted signal is a function of the binary digit (that is, 0 or 1) being sent and of the value of the likelihood ratio, and is adjusted so as to make this ratio increase as fast as possible. Average and peak power constraints are imposed. The average time $\bar{T}$ for deciding on a binary digit turns out to be $\bar{T} = (P_{av}/N_o)^{-1}\ell n 2$ where P_{av} is the average power and N_o is the (one-sided) noise power spectral density. The probability of error P_e vanishes if infinite peak power and infinite bandwidth are allowed. Hence, a rate is achieved that is equal to the channel capacity

$$C = \frac{P_{av}}{N_o} \text{ nats/sec} \quad .^{*} \tag{1}$$

Schalkwijk and Kailath[1,2] developed a predecision feedback scheme motivated by the Robbins-Munro[8] stochastic approximation procedure. With a noiseless feedback link available, they considered the problem where the transmitter has to send one of M possible messages to a receiver where each message takes T seconds to send. Defining signaling rate as $R = (\ell n\ M)/T$ nats/sec and having only the transmitter constraint of average power, P_{av}, this scheme achieves rates up to channel capacity, $C = P_{av}/N_o$, with error probability given by

$$P_e = 2 \text{ erfc} \left\{ \frac{\sqrt{3}\ e^{[C-R]T}}{e^{1.577}} \right\} \quad .^{\dagger} \tag{2}$$

Schalkwijk[2] modified this scheme by requiring the transmitted average power to be constant at each iteration. Imposing both an average power constraint, P_{av}, and a signal bandwidth restriction, W, this scheme achieved rates up to channel capacity, $C = W\ \ell n\ (1 + P_{av}/N_o W)$, with the error probability

$$P_e = 2 \text{ erfc} \left\{ \sqrt{3} \left[\frac{\left[\frac{2WT - 1}{2WT} + \frac{P_{av}}{N_o W}\right]^{WT}}{e^{\left(\frac{R}{W}\right)}} \right] \right\} \quad . \tag{3}$$

This coding scheme developed by Schalkwijk gave the first deterministic procedure to achieve rates up to capacity for the band limited white Gaussian noise channel.

These schemes of Schalkwijk and Kailath motivated the work presented here.

The Problem

Here discrete-time channels that are derived from the continuous-time channels are considered. Following the control theory point of view a communication received is regarded as a dynamical system described by a difference equation. The variable (state) of this equation is the test statistics upon which a decision is based at some fixed terminal iteration, N. If noiseless feedback to the transmitter is allowed, it is observed that the signal selection problem is a stochastic control problem where the state of the system is completely observable. With noise in the feedback it is a stochastic control problem with noisy observations of the states.

Considering the signal selection problem from this control point of view the following assumptions are made:

1. The receiver is linear and discrete in time with its states given by the difference equation,

$$\begin{aligned} \chi_{k+1} &= \Phi_k \chi_k + G_k r_k \qquad k = 0,1,2,\ldots,N-1 \\ \chi_o &= 0 \end{aligned} \tag{4}$$

where χ_k is the state at the k^{th} iteration,

$$\left\{\Phi_k\right\}_{k=0}^{N-1} \quad , \quad \left\{G_k\right\}_{k=0}^{N-1}$$

are free parameters of the receiver,

r_k is the received signal from the channel at the k^{th} iteration,

and there are N iterations taking a total time of T seconds.

2. One of M possible equally likely messages is sent at any given time.

3. The receiver bases its decision on χ_N, where the decision regions consist of the unit

---

* "Nats" is defined as natural units of information in accordance with IEEE standards.

† This is a corrected version of their result. In Ref. 1, Eq. 11 becomes

$$P_{av}\ T = \frac{\alpha^2}{12} + \sigma^2\ (\ell n\ N + .577)$$

giving the relation

$$N = e^{-\left(\frac{\alpha^2}{12\sigma^2} + .577\right) + \frac{P_{av}T}{\sigma^2}}$$

so that optimizing with respect to α^2 gives $\alpha_o^2 = 12\ \sigma^2 = 6\ N_o$. (Compare with Eq. 15 in Ref. 1.)

interval, [- 1/2, 1/2], divided into M equal length disjoint subintervals. The jth message is chosen only if χ_N lies in the jth subinterval of [- 1/2, 1/2].

4. The criterion for choosing signals is the minimization of $E(\chi_N - \theta)^2$ under the constraint that the average power of the transmitter, P_{av}, is fixed. Here θ is the center of the subinterval corresponding to the sent message. The signal sequence is denoted

$$\{m_k\}_{k=0}^{N-1}$$

and the constraint equation is

$$P_{av} = \frac{1}{T} E \sum_{k=0}^{N-1} m_k^2 \quad . \tag{5}$$

5. The message points, $\theta \epsilon \Theta$, are essentially uniformly distributed over [- 1/2, 1/2] (for large M) with variance,

$$\mathrm{Var}(\theta) = \sigma_o^2 = \frac{1}{12} \quad . \tag{6}$$

This situation is illustrated in Fig. 1, where as yet no assumptions have been made concerning the feedback link. As an example, if the feedback link is noiseless, then $y_k = S_k = \chi_k$ for $k = 1, 2, \ldots, N - 1$.

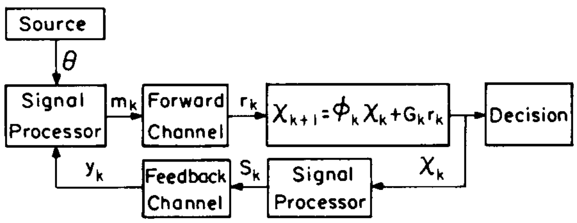


Fig. 1

Figure 2 is a sketch of a typical sequence showing how

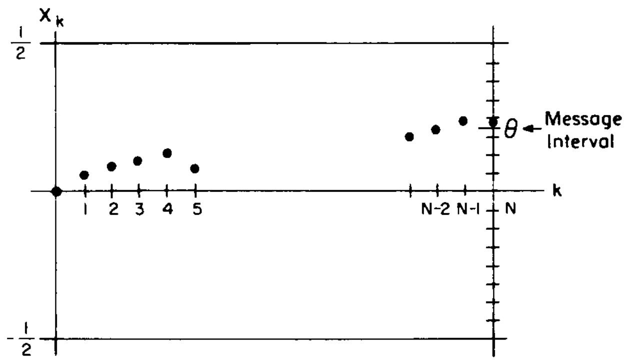
$$\{\chi_k\}_{k=0}^{N-1}$$

might behave and how the decision is made.

With these assumptions, the goal is then to find optimum signals,

$$\{m_k\}_{k=0}^{N-1} \quad ,$$

as defined by assumption 4, for a given forward and feedback channel. In general at the kth iteration the signal component is based on θ, and the observations of the receiver up to the kth



Decision based on X_N.

Fig. 2

iteration,

$$\{y_j\}_{j=1}^{k} \quad .$$

Instead of minimizing probability of error, the criterion chosen is to minimize the distance $E(\chi_N - \theta)^2$. This is closely related to minimizing probability of error for the Gaussian additive noise channels. One can see from Fig. 2 that for the noiseless feedback case the transmitter tries to "control" the states so as to get χ_N as close to θ as possible. With noisy feedback this controlling is hampered by poor observations of the states.

Stated in this manner, this is essentially a nonsequential estimation problem upon which is imposed a multiple hypothesis structure. It is nonsequential since the time of decision is fixed rather than a random variable. The criterion $E(\chi_N - \theta)^2$ is really an estimation criterion so that if a transmitter (satellite) is to send some measurement data normalized to [- 1/2, 1/2], this data would not be quantized into one of M levels but sent directly. However, by imposing a quantization one can then interpret the result as a multiple hypothesis problem where probability of errors and rates of information are evaluated.

Finally it should be noted that in designing a communication scheme one should attach a cost to complexity of equipment. It will turn out that solutions to this problem result in very simple schemes without considering such costs. Another important byproduct is the insensitivity of the schemes to the particular noise statistics. Also once the optimum signal sequence is found a second order optimization with respect to receiver parameters

$$\{\Phi_k\}_{k=0}^{N-1} \quad \text{and} \quad \{G_k\}_{k=0}^{N-1}$$

is possible. This is done for the additive noise channel and the multiplicative noise channel when noiseless feedback is assumed.

Results

This problem is solved for the additive noise channel with noiseless feedback. Assuming further that the noise is a white Gaussian noise process with spectral density $N_o/2$, the probability of error for the wideband channel (no bandwidth restriction on the signals in the channel) is found to be

$$P_e = 2 \text{ erfc} \left\{ \sqrt{3}\, e^{[C_\infty - R]T} \right\} \qquad (7)$$

where

$$C_\infty = \frac{P_{av}}{N_o} \quad .$$

When signals are limited in bandwidth to [- W, W], the error probability is

$$P_e = 2 \text{ erfc} \left\{ \sqrt{3}\, e^{[C_W - R]T} \right\} \qquad (8)$$

where

$$C_W = W \ln \left(1 + \frac{P_{av}}{N_o W} \right) \quad .$$

The optimum schemes developed in this work are dependent only on the 1st and 2nd order statistics of all random variables, although all error probabilities are evaluated under a Gaussian assumption. Throughout this work a control theory approach is taken using dynamic programming as the main tool. This approach is new and versatile as made evident by its ability to handle noisy feedback and multiplicative noise problems as well as the usual additive noise forward channel with noiseless feedback. These cases will appear soon in a Stanford Electronics Laboratories report.

ADDITIVE NOISE CHANNELS WITH NOISELESS FEEDBACK

This paper is devoted to developing an optimal feedback communication scheme for the additive noise forward channel with a noiseless feedback link. The additive noise is assumed to be white with double-sided spectral density $N_o/2$. The optimization is carried out in two steps: first, signal optimization based on the control theory point of view, and then receiver parameter optimization using ordinary calculus. This optimum scheme is then evaluated in terms of probability of error and information rates for the white Gaussian noise channel.

The Discrete-Time Channel

It is convenient to work with discrete-time channels that are equivalent to the continuous-time channels under consideration. This makes it possible to work with finite sequences of numbers rather than with continuous-time functions. In particular, the transmitted signal representing a message will be a sequence of numbers,

$$\left\{ m_k \right\}_{k=0}^{N-1} \quad ,$$

so that signal optimization consists of finding N optimum numbers rather than finding a function of time.

Consider the zero mean additive noise channel in Fig. 3. To obtain a discrete-time channel from

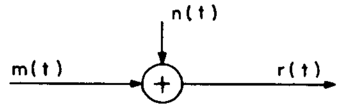


Fig. 3

this continuous-time channel, assume that the message information is transmitted by suitably amplitude modulating the amplitude of a known basic waveform $\Phi(t)$. The signal in the channel will thus be of the form

$$m(t) = \sum_{k=0}^{N-1} m_k \Phi(t - k \tfrac{T}{N}) \quad , \qquad (9)$$

where T/N will be specified later. The basic waveform, $\Phi(t)$, is required to have unit energy and to be orthogonal for shifts of T/N; that is, $\Phi(t)$ should satisfy

$$\int \Phi(t - i \tfrac{T}{N}) \Phi(t - j \tfrac{T}{N}) dt = \delta_{ij} \quad . \qquad (10)$$

Reception will be achieved by using a filter matched to $\Phi(t)$, that is, a filter with impulse response $h(t) = \Phi(- t)$. The output to this matched filter at $t = k(T/N)$, $k = 0, 1, \ldots, N - 1$ will be the sequence

$$\left\{ r_k \right\}_{k=0}^{N-1}$$

where $r_k = m_k + n_k$, and

$$n_k = \int n(t) \Phi(t - k \tfrac{T}{N}) dt \quad . \qquad (11)$$

With this modulation and reception, the discrete-time channel shown in Fig. 4 is obtained.

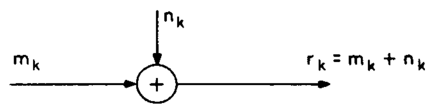


Fig. 4

If the additive noise of the continuous-time channel is white with double-sided spectral density $N_0/2$, it is easy to show that the noise

$$\{n_k\}_{k=0}^{N-1}$$

will be uncorrelated zero mean random variables with

$$E\, n_i n_j = \sigma^2 \delta_{ij} \quad , \tag{12}$$

where $\sigma^2 = N_0/2$. When the additive noise is Gaussian, these random variables will be Gaussian and therefore independent. In the Gaussian case, it is easy to see that the discrete-time channel thus obtained is completely equivalent to the original continuous-time channel. This follows from the fact that the matched filter is the ideal receiver for the white Gaussian noise channels and therefore preserves all the information in the received waveform that is relevant to the decision-making process.

Finally, note that by virtue of the orthonormality of

$$\{\phi[t - k(T/N)]\}_{k=0}^{N-1}$$

the transmitted energy of

$$m(t) = \sum_{k=0}^{N-1} m_k \phi(t - k\frac{T}{N}) \quad \text{is} \quad \sum_{k=0}^{N-1} m_k^2 \quad .$$

Signal Optimization

The signal optimization is done from the control theory point of view where the receiver is regarded as a dynamical system which can be partially controlled by the transmitted signal sequence

$$\{m_k\}_{k=0}^{N-1} \quad .$$

The problem is to choose the N numbers

$$\{m_k\}_{k=0}^{N-1}$$

in some optimum manner when the transmitter has complete knowledge through noiseless feedback of how the receiver is behaving.

Recall from the Introduction that the receiver first does a linear operation on the received signal sequence

$$\{r_k\}_{k=0}^{N-1}$$

described by the difference equation

$$x_{k+1} = \phi_k x_k + G_k r_k \quad , \quad k = 0,1,2,\ldots,N-1 \tag{13}$$

where

$$x_0 = 0 \quad ,$$

$$r_k = m_k + n_k \quad .$$

A decision is based on x_N where the j^{th} message is chosen only if x_N lies in the j^{th} subinterval on $[-1/2, 1/2]$. The only way the transmitter can control the value of x_N is through the signal sequence

$$\{m_k\}_{k=0}^{N-1} \quad .$$

What the transmitter would ideally like to do is to choose the sequence that forces x_N into the correct subinterval (corresponding to the message the transmitter wants to send) with minimum probability of error. Instead of minimum probability of error, which is difficult to work with, however, a minimum mean square distance criterion is used. Choosing θ to be the center point of the correct subinterval, the criterion is to choose the signal sequence that minimizes $E(x_N - \theta)^2$. Here the expectation is taken over all the noise random variables

$$\{n_k\}_{k=0}^{N-1} \quad .$$

Since the transmitter power is limited, it is necessary to impose some sort of power constraint on the signal sequence. A time and statistical average power constraint is imposed so that

$$P_{av}(\theta) = \frac{1}{T} E \sum_{k=0}^{N-1} m_k^2 \tag{14}$$

is the constraint equation. Here again expectation is taken over the random variables

$$\{n_k\}_{k=0}^{N-1} \quad .$$

Letting λ be a Lagrange multiplier, the total criterion is

$$J = E(x_N - \theta)^2 + \frac{\lambda}{T} E \sum_{k=0}^{N-1} m_k^2 \quad . \tag{15}$$

Thus, signal optimization consists of finding the sequence

$$\{m_k\}_{k=0}^{N-1}$$

that minimizes Eq. (15) where λ is found through Eq. (14). This signal sequence will be referred to as the optimum signal and labeled

$$\left\{\overset{o}{m}_k\right\}_{k=0}^{N-1} .$$

The solution to this discrete-time stochastic control problem where the dynamical system is linear [Eq. (13)] and the criterion is quadratic [Eq. (15)] is well known in the control theory literature.[9] This solution, however, requires that

$$\left\{n_k\right\}_{k=0}^{N-1}$$

be uncorrelated random variables, which is the same as assuming the additive noise is white in the continuous-time channel. Making this white noise assumption, the optimum signals are now derived using dynamic programming.[10]

Define

$$f_{N-k}(x_k) = \underset{\{m_j\}_{j=k}^{N-1}}{\text{Min}} \; \underset{\{n_j\}_{j=k}^{N-1}}{E} \left\{(x_N-\theta)^2 + \frac{\lambda}{T}\sum_{j=k+1}^{N} m_{j-1}^2\right\} \tag{16}$$

for

$$j = 0,1,2,\ldots,N-1$$

and

$$f_0(x_N) = (x_N - \theta)^2 \quad . \tag{17}$$

Note that $f_{N-k}(x_k)$ is the minimum expected cost from the k^{th} iteration to the N^{th} iteration assuming x_k is the state of the receiver at the k^{th} iteration. A systematic solution procedure may be obtained by making use of the fundamental principle of dynamic programming: The Principle of Optimality. This states:[9]

> An optimal policy has the property that whatever the initial state and the initial decision are, the remaining decisions must constitute an optimal policy with regard to the state resulting from the initial decision.

Here initial decision refers to earlier choices of m_0, m_1, ..., m_{k-1} when considering the choice of m_k, m_{k+1}, ..., m_{N-1}.

The Principle of Optimality, which describes the basic properties of optimal solutions, is based upon the fundamental approach of invariant imbedding. This implies that to solve a specific optimization problem, the original problem is imbedded within a family of similar problems. Thus, the multistage optimization problem is reduced to a sequence of single optimization problems. To be able to apply this principle to the specific problem stated requires the assumption that the state process

$$\left\{x_k\right\}_{k=0}^{N}$$

be a weak-sense Markov process.* This is equivalent to requiring

$$\left\{n_k\right\}_{k=0}^{N-1}$$

to be uncorrelated.

Invoking the Principle of Optimality, the following recursive relation is found:

$$f_{N-k}(x_k) = \underset{m_k}{\text{Min}} \; \underset{n_k}{E} \left[\frac{\lambda}{T} m_k^2 + f_{\overline{N-k+1}}(x_{k+1})\right] \tag{18}$$

for

$$k = 0,1,2,\ldots,N-1 \quad .$$

Starting at the end where $f_0(x_N) = (x_N - \theta)^2$ and working backward, some algebra will show that

$$f_{\overline{N-k+1}}(x_{k+1}) = P(k+1)\left(x_{k+1} - \frac{\theta}{\prod_{j=k+1}^{N-1}\Phi_j}\right)^2 + \sum_{j=k+2}^{N} P(j)G_{j-1}^2\sigma^2 \tag{19}$$

where $\sigma^2 = En_k^2$. Using this in Eq. (18) gives

$$f_{N-k}(x_k) = \underset{m_k}{\text{Min}} \; \underset{n_k}{E} \left\{\frac{\lambda}{T} m_k^2 + P(k+1)\left(x_{k+1} - \frac{\theta}{\prod_{j=k+1}^{N-1}\Phi_j}\right)^2\right\} + \sum_{j=k+2}^{N} P(j)G_{j-1}^2\sigma^2 \quad . \tag{20}$$

By Eq. (13) x_{k+1} can be written in terms of x_k and n_k. After taking the expectation with respect to n_k, f_{N-k} is

$$f_{N-k}(x_k) = \underset{m_k}{\text{Min}}\left\{\frac{\lambda}{T} m_k^2 + P(k+1)\left[\Phi_k\left(x_k - \frac{\theta}{\prod_{j=k}^{N-1}\Phi_j}\right) + G_k m_k\right]^2\right\} + \sum_{j=k+1}^{N} P(j)G_{j-1}^2\sigma^2 \quad . \tag{21}$$

---

*A process is a weak-sense Markov process if the expected value of the processes at some time given the values at some previous times depends only on the last given value.

Now differentiating the term on the right with respect to m_k gives

$$m_k^o = B(k)\left(x_k - \frac{\theta}{\prod_{j=k}^{N-1} \Phi_j}\right) \quad , \tag{22}$$

where

$$B(k) = -\frac{P(k+1)\Phi_k G_k}{\frac{\lambda}{T} + P(k+1)G_k^2} \quad , \tag{23}$$

as the optimum k^{th} term of the optimum signal sequence. Putting this optimum value into Eq. (21) results in

$$f_{N-k}(x_k) = P(k)\left(x_k - \frac{\theta}{\prod_{j=k}^{N-1} \Phi_j}\right)^2 + \sum_{j=k+1}^{N} P(j)G_{j-1}^2\sigma^2 \quad , \tag{24}$$

where

$$P(k) = P(k+1)[\Phi_k^2 + \Phi_k G_k B(k)] \tag{25}$$

$$P(N) = 1 \quad .$$

Repeating this single stage optimization procedure N times gives the optimal signal

$$\left\{m_k^o\right\}_{k=0}^{N-1}$$

in a form where the constants

$$\left\{B(k)\right\}_{k=0}^{N-1}$$

and

$$\left\{P(k)\right\}_{k=1}^{N}$$

are in recursive form. One can get these constants in closed form by iterating Eqs. (23) and (25), which will yield

$$P(k) = -\frac{\lambda}{T}\frac{\Phi_k}{G_k} B(k) \tag{26}$$

and

$$B(k) = -\frac{\frac{G_k}{\Phi_k}\prod_{j=k}^{N-1}\Phi_j^2}{\sum_{\ell=k+1}^{N-1} G_{\ell-1}^2 \prod_{j=\ell}^{N-1}\Phi_j^2 + G_{N-1}^2 + \frac{\lambda}{T}} \quad . \tag{27}$$

Thus, the optimal signal sequence is given in terms of the receiver parameters

$$\left\{\Phi_k\right\}_{k=0}^{N-1}$$

and

$$\left\{G_k\right\}_{k=0}^{N-1}$$

and λ, where λ is found from the constant equation

$$P_{av}(\theta) = \frac{1}{T}\sum_{k=0}^{N-1} B^2(k)\, E\left(x_k - \frac{\theta}{\prod_{j=k}^{N-1}\Phi_j}\right)^2 \quad . \tag{28}$$

The solution given by Eq. (22) has several interesting properties. First of all, the solution is nonparametric in the sense that all that is required is the uncorrelatedness of the additive noise. Except for a finite variance, σ^2, no other noise statistics are required. Even if the variance of the noise changes at each iteration, this analysis may be carried through. A second property is that the signal at the k^{th} iteration is a simple linear function of the current state, x_k, of the receiver. Knowledge of this exact state is made available to the transmitter by the noiseless feedback link. Finally, and perhaps most important, is the property that if the feedback link is noisy such that the transmitter has only noisy observations of the receiver's current state, this solution is still optimum with x_k replaced by $\hat{x}_{k|k}$, $\hat{x}_{k|k}$ being the least mean square error estimate of x_k based on noisy observations up to the k^{th} iteration. In other words, for the problem with noisy feedback, estimation and optimization separate.[11,12]

So far the optimization is carried out in terms of the receiver parameters

$$\left\{\Phi_k\right\}_{k=0}^{N-1}$$

and

$$\left\{G_k\right\}_{k=0}^{N-1} \quad .$$

In the next section optimum receiver parameters are found using ordinary calculus.

Receiver Parameter Optimization

In the last section the optimal signal sequence

$$\left\{m_k^o\right\}_{k=0}^{N-1}$$

was found in terms of the receiver parameters

$$\left\{\Phi_k\right\}_{k=0}^{N-1}$$

and

$$\left\{G_k\right\}_{k=0}^{N-1} \quad .$$

Using this optimum signal sequence, after some algebra the minimum distance $E(x_N - \theta)^2$ and the constraint equation can now be found in terms of these parameters.

$$E(x_N-\theta)^2 = \left(\frac{\frac{\lambda}{T}}{\sum_{\ell=1}^{N-1} G_{\ell-1}^2 \prod_{k=\ell}^{N-1} \phi_j^2 + G_{N-1}^2 + \frac{\lambda}{T}} \right)^2 \theta^2$$

$$+ \sum_{k=1}^{N-1} \left(\frac{\frac{\lambda}{T}}{\sum_{\ell=k+1}^{N-1} G_{\ell-1}^2 \prod_{j=\ell}^{N-1} \phi_j^2 + G_{N-1}^2 + \frac{\lambda}{T}} \right)^2 G_{k-1}^2 \prod_{j=k}^{N-1} \phi_j^2 \sigma^2 + G_{N-1}^2 \sigma^2 , \tag{29}$$

$$P_{av}(\theta) = \frac{1}{T}\left(G_{N-1}^2 + \sum_{j=1}^{N-1} G_{j-1}^2 \prod_{\ell=j}^{N-1} \phi_\ell^2 \right) \left(\frac{1}{\sum_{j=1}^{N-1} G_{j-1}^2 \prod_{\ell=j}^{N-1} \phi_\ell^2 + G_{N-1}^2 + \frac{\lambda}{T}} \right)^2 \theta^2$$

$$+ \frac{1}{T} \sum_{k=1}^{N-2} \left(G_{N-1}^2 + \sum_{j=k+1}^{N-1} G_{j-1}^2 \prod_{l=j}^{N-1} \phi_l^2 \right) \left(\frac{1}{\sum_{j=k+1}^{N-1} G_{j-1}^2 \prod_{\ell=j}^{N-1} \phi_\ell^2 + G_{N-1}^2 + \frac{\lambda}{T}} \right)^2 G_{k-1}^2 \prod_{\ell=k}^{N-1} \phi_\ell^2 \sigma^2$$

$$+ \frac{1}{T} \frac{\phi_{N-1}^2 G_{N-1}^2 G_{N-2}^2 \sigma^2}{\left(G_{N-1}^2 + \frac{\lambda}{T} \right)^2} . \tag{30}$$

Thus, the minimum distance $E(x_N - \theta)^2$ and the constraint equation are now written in terms of the parameters of the receiver and the Lagrange multiplier λ. Note, however, that both these equations depend on the particular value of θ. Any sort of optimum parameter set must be independent of the message point θ so that parameter optimization will be carried out on the averaged equations. The averaged equations,

$$\overline{E(x_N - \theta)^2} = \underset{\theta}{E}\, E(x_N - \theta)^2 \tag{31}$$

and

$$P_{av} = \underset{\theta}{E}\, P_{av}(\theta) \tag{32}$$

are found by simply replacing θ^2 by $\sigma_0^2 = E\theta^2$. From this point on, all expectations will be taken with respect to both the channel noise and all possible message points θ. Rather than a constraint equation for each θ, only one averaged (over θ too) power constraint is imposed. The distance $E(x_N - \theta)^2$ is also now averaged over θ as well as the channel noise. This additional averaging does not change the results in any important way, but it does reduce the computation required by a considerable amount.

In searching for the optimum parameters

$$\left\{\phi_k\right\}_{k=0}^{N-1}$$

and

$$\left\{G_k\right\}_{k=0}^{N-1}$$

that minimize $E(x_N - \theta)^2$ and meet the average power constraint, first note that in both Eqs. (29) and (30) these parameters appear in the form

$$G_{k-1}^2 \prod_{j=k}^{N-1} \phi_j^2 \tag{33}$$

$$k = 1,2,\ldots,N \quad .$$

It is clear from this that any change in the parameters

$$\left\{\phi_k\right\}_{k=0}^{N-1}$$

may be absorbed by corresponding changes in

$$\left\{G_k\right\}_{k=0}^{N-1} \quad .$$

Hence, without loss of generality, choose

$$\phi_k = 1 \quad , \quad k = 0,1,2,\ldots,N-1 \quad . \tag{34}$$

This choice reduces the main equations to simpler forms.

$$E(x_N-\theta)^2 = \left(\frac{\frac{\lambda}{T}}{\frac{\lambda}{T} + \sum_{k=0}^{N-1} G_k^2} \right)^2 \sigma_o^2 + \sum_{k=1}^{N-1} \left(\frac{\frac{\lambda}{T}}{\frac{\lambda}{T} + \sum_{j=k}^{N-1} G_j^2} \right)^2 G_{k-1}^2 \sigma^2 + G_{N-1}^2 \sigma^2 \tag{35}$$

and

$$P_{av} = \frac{1}{T} \left(\sum_{k=0}^{N-1} G_k^2 \right) \left(\frac{1}{\frac{\lambda}{T} + \sum_{k=0}^{N-1} G_k^2} \right)^2 \sigma_o^2 + \sum_{k=1}^{N-1} \left(\sum_{j=k}^{N-1} G_j^2 \right) \left(\frac{1}{\frac{\lambda}{T} + \sum_{j=k}^{N-1} G_j^2} \right)^2 G_{k-1}^2 \sigma^2 . \tag{36}$$

Regarding λ as just another parameter, the problem of finding optimum parameters reduces to a straightforward calculus problem where the conditions for the optimum parameters are

$$\frac{\partial E(x_N - \theta)^2}{\partial (\frac{\lambda}{T})} = - \frac{\lambda}{T} \frac{\partial (P_{av} T)}{\partial (\frac{\lambda}{T})} \tag{37}$$

and

$$\nabla_{\underline{G}} E(x_N - \theta)^2 = - \frac{\lambda}{T} \nabla_{\underline{G}} (P_{av} T) \tag{38}$$

where $\nabla_{\underline{G}}$ is the N-dimensional gradient with respect to the parameters $\left\{G_k\right\}_{k=0}^{N-1}$.

Solving these N + 1 equations the optimum parameters are found to be

$$\frac{\lambda}{T} = \frac{b^2\rho^{2N}}{1 - \rho^2} \tag{39}$$

$$G_k = b\rho^k \quad , \quad k = 0,1,\ldots,N - 1 \tag{40}$$

where

$$b = \frac{\sqrt{P_{av}TN\sigma_0^2}}{P_{av}T + N\sigma^2}$$
$$\rho = \left(\frac{N\sigma^2}{P_{av}T + N\sigma^2}\right)^{\frac{1}{2}} \quad . \tag{41}$$

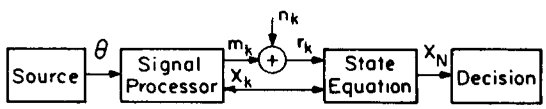
The overall scheme using the optimum signals and the parameters given above will now be referred to as the optimum scheme. This choice of parameters gives the optimum signal sequence

$$m_k^o = -\frac{(1 - \rho^2)}{b\rho^k}(x_k - \theta) \quad , \quad k = 0,\ldots,N - 1 \tag{42}$$

and the minimum distance

$$E(x_N - \theta)^2 = \sigma_0^2\rho^{2N} \quad . \tag{43}$$

This optimum scheme is summarized in Fig. 5.



State Equation: $x_{k+1} = x_k + b\rho^k r_k, \; k = 0,1,\ldots,N-1$

$$x_0 = 0$$

Signal Sequence: $m_k^o = -\dfrac{(1 - \rho^2)}{b\rho^k}(x_k - \theta)$

$$k = 0,1,\ldots,N-1$$

Fig. 5

One important property of this optimal scheme is the fact that the total power, $P_{av}T$, is uniformly distributed over the N iterations. From Eq. (13) with the optimum parameters given in Eqs. (40), (41), and (42), it follows that for the k^{th} state

$$E(x_k - \theta)^2 = \sigma_0^2\rho^{4k} + \frac{b^2\sigma^2}{\rho^2(1 - \rho^2)}(\rho^{2k} - \rho^{4k}) \quad . \tag{44}$$

Noting that $(b^2\sigma^2)/\rho^2(1 - \rho^2) = \sigma_0^2$, this becomes

$$E(x_k - \theta)^2 = \sigma_0^2\rho^{2k} \quad . \tag{45}$$

Thus, the k^{th} iteration of the signal sequence has averaged power

$$Em_k^2 = \frac{(1 - \rho^2)^2\sigma_o^2}{b^2} = \frac{P_{av}T}{N} \quad . \tag{46}$$

P_e for the White Gaussian Noise Channel

Although an optimal scheme has been developed for the additive white noise channel with noiseless feedback, little has been said about how well it performs. In particular, what sort of information rate can it have, and how does the probability of error behave? In this section these questions will be answered for the white Gaussian noise channel. Also a comparison with the best known schemes without feedback will be made.

For the white Gaussian noise channel the noise components

$$\left\{n_k\right\}_{k=0}^{N-1}$$

are independent zero mean Gaussian random variables with variance

$$En_k^2 = \sigma^2 = \frac{N_o}{2} \tag{47}$$

where $N_o/2$ is the double-sided spectral density. Because of the linearity of the state equation and of the optimum signal sequence, the states

$$\left\{x_k\right\}_{k=0}^{N}$$

are also Gaussian random variables. Since the decision as to what message is sent is based on x_N and the statistics of x_N are known, the probability of error, P_e, is easy to compute.

Consider the state x_N for the optimal scheme shown in Fig. 5.

$$\begin{aligned} x_N &= x_{N-1} - (1 - \rho^2)(x_{N-1} - \theta) + b\rho^{N-1}n_{N-1} \\ &= \rho^2 x_{N-1} + (1 - \rho^2)\theta + b\rho^{N-1}n_{N-1} \quad . \end{aligned} \tag{48}$$

Using the state equation for each state successively gives

$$x_N = (1 - \rho^{2N})\theta + b\rho^{2(N-1)} \sum_{j=0}^{N-1} \rho^{-j} n_j \quad . \quad (49)$$

The mean value of x_N given θ is clearly

$$E(x_N\ \theta) = (1 - \rho^{2N})\theta \quad (50)$$

with conditional variance

$$\mathrm{Var}\ (x_N|\theta) = \frac{b^2\sigma^2}{\rho^2(1 - \rho^2)} \rho^{2N}(1 - \rho^{2N}) \quad . \quad (51)$$

Noting that

$$\frac{b^2\sigma^2}{\rho^2(1 - \rho^2)} = \sigma_0^2 \quad , \quad (52)$$

this conditional variance is rewritten

$$\mathrm{Var}\ (x_N|\theta) = \sigma_0^2\rho^{2N}(1 - \rho^{2N}) \quad . \quad (53)$$

Note that from Eq. (50) it is clear that x_N is a biased estimator of θ. This bias resulted from the fact that an average power constraint was imposed when trying to minimize $E(x_N - \theta)^2$. For large N and T, however, $\rho^{2N} \ll 1$ so that a good approximation for the conditional mean and variance is

$$E(x_N|\theta) = \theta \quad (54)$$

and

$$\mathrm{Var}\ (x_N|\theta) = \sigma_0^2\rho^{2N} \quad . \quad (55)$$

The conditional mean and variance given in Eqs. (54) and (55) will be used instead of the actual mean and variance of Eqs. (50) and (51). This will result in a slight upper bound of the true probability of error, but the difference is negligible for large values of T and N.

Recall that the receiver bases its decision as to which one of M messages is sent on x_N in the following manner. The j^{th} message is chosen only if x_N lies in the j^{th} subinterval of [- 1/2, 1/2]. Since there are M equal length subintervals, each one has length 1/M. If x_N lies in the subinterval containing θ, then a correct decision is made. Noting that x_N is a Gaussian random variable with mean θ and variance $\sigma_0^2\rho^{2N}$, the probability density of x_N is presented in Fig. 6. The probability of

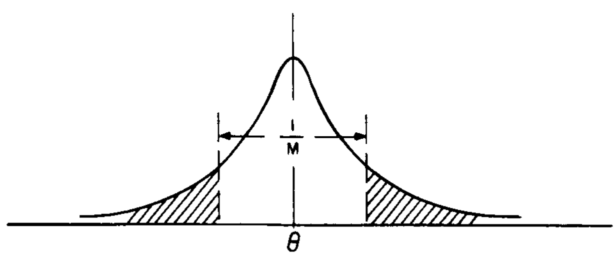


Fig. 6. The Error Probability is the Shaded Area

error, P_e, is the probability that x_N is outside the interval containing θ (shaded area). Thus,

$$\begin{aligned} P_e &= 2\ \mathrm{erfc} \left\{ \frac{\frac{1}{2M}}{\sqrt{\mathrm{Var}\ (x_N|\theta)}} \right\} \\ &= 2\ \mathrm{erfc} \left\{ \left(\frac{1}{4M^2\sigma_0^2\rho^{2N}} \right)^{\frac{1}{2}} \right\} \quad . \end{aligned} \quad (56)$$

This equation shows that P_e can be driven to zero by increasing T (and N). Since increasing T allows more total expected power, $P_{av}T$, per message this result is expected. Suppose now, however, that M is increasing while T (and N) are fixed; then in this case P_e approaches one. This is also expected since the total energy per message is fixed while the number of possible messages increases. If both T (and N) and M are increasing, what is the trade-off point where these two opposite effects cancel? In particular, is it possible for M to increase with T as

$$M = e^{RT} \quad (57)$$

for R a positive constant and still have P_e go to zero with increasing T? Shannon pointed out that it is possible to have P_e go to zero as long as R is less than a critical constant C which he called channel capacity. In the following discussion it will be shown that the optimal scheme can achieve this critical rate and that it has a probability of error that decreases much more quickly with increasing T than the best known nonfeedback communication schemes.

Defining as the message rate

$$R = \frac{\ln M}{T} \quad \text{Nats per second} \quad (58)$$

and recalling Eqs. (56), (57), and the fact that $\sigma_0^2 = 1/12$, the probability of error is rewritten,

$$P_e = 2\ \mathrm{erfc}\ \{\sqrt{3}\ e^{[C(N,T)-R]T}\} \quad (59)$$

where

$$C(N,T) = \frac{N}{2T} \ln \left(1 + \frac{2P_{av}T}{NN_o}\right) \quad . \quad (60)$$

As yet N and T have not been specified. In the continuous-time channel these schemes require N orthogonal carrier signals of duration not more than T seconds. Hence, for a fixed time T, the value of N is determined by the number of orthogonal carriers allowed. Two cases will now be investigated.

1. The Wideband Scheme. Typically in space communication, the channel places no restrictions on the carrier signal bandwidth. Under this condition of no bandwidth restriction, the additive white Gaussian noise channel has the channel capacity given by[13]

$$C_\infty = \frac{P_{av}}{N_o} \quad \text{Nats per second} \quad . \quad (61)$$

Without feedback, the best known code for this channel is a "regular-simplex" set of code words (that is, a set of M equal-energy signals with mutual cross-correlation of $-1/(M-1)$. For large M, an orthogonal signal set (for which the cross-correlations are zero rather than $-1/(M-1)$ performs almost as well. The ideal receiver for such signals is a bank of M correlation detectors, whose outputs are scanned to determine the correlator yielding the largest output. The error probability for an orthogonal (or simplex) signal set has been evaluated numerically for values of M from 2 to 10^6. For larger values of M, the following asymptotic expression can be used. If T is the duration of each of the M signals, assumed equally likely a priori, then[14]

$$P_{e,orth} = \frac{\text{constant}}{T^{\beta}} - TE(R) \qquad (62)$$

where

$$E(R) = \begin{cases} \frac{C_\infty}{2} - R & , \quad 0 \le R \le \frac{C_\infty}{4} \\ (\sqrt{C_\infty} - \sqrt{R})^2 & , \quad \frac{C_\infty}{4} \le R \le C_\infty \end{cases} \qquad (63)$$

$$1 \le \beta \le 2 \quad .$$

This equation shows that the error probability for orthogonal codes decreases essentially exponentially with T. As a result, for large T, the choice of a suitable pair of values R and R to achieve a given P_e is essentially determined by the quantity E(R).

Consider now the optimal scheme with the error probability given by Eqs. (59) and (60). When the channel places no restrictions on the carrier bandwidth, the number of possible orthogonal carrier signals of duration less than T seconds is unlimited. Hence, for a fixed T, N can be made arbitrarily large so that Eq. (60) becomes

$$\lim_{N\to\infty} C(N,T) = C_\infty = \frac{P_{av}}{N_o} \qquad (64)$$

resulting in the error probability

$$P_e = 2 \operatorname{erfc} \left\{ \sqrt{3}\, e^{(C_\infty - R)T} \right\} \quad . \qquad (65)$$

From the well known bounds on erfc (x)[15]

$$\frac{1}{\sqrt{2\pi}} e^{-1/2x^2} \left\{ \frac{1}{x} - \frac{1}{x^3} \right\} < \operatorname{erfc}(x) < \frac{1}{\sqrt{2\pi}} e^{-1/2x^2} \frac{1}{x} \quad . \qquad (66)$$

it is clear that P_e decreases essentially in a double exponential manner with increasing T. Hence, although the message rate is bounded by C_∞ in both cases, the rate at which the error probability decreases with increasing T is dramatically more rapid with the optimal feedback scheme than the best nonfeedback scheme. As a simple comparison between the feedback and nonfeedback cases, consider the value of T required to achieve

$$P_e = P_{e,orth} = 10^{-7} \qquad (67)$$

for

$$C = 1 \quad \text{bit/sec}$$

$$R = .8C \quad .$$

The nonfeedback orthogonal code scheme gives

$$T_{orth} = 2030 \text{ seconds} \qquad (68)$$

while the optimal feedback scheme requires

$$T_{fb} = 8.1 \text{ seconds} \quad . \qquad (69)$$

Although the optimal scheme has been evaluated here for the white Gaussian noise channel, it does not depend on the statistics of the additive white noise. If the additive noises

$$\{n_k\}_{k=0}^{N-1}$$

are independent random variables but otherwise unspecified, then for this wideband channel the P_e given by Eq. (65) is still correct. This can easily be shown by applying the central limit theorem. In this case, however, $C_\infty = P_{av}/N_o$ is only a lower bound for the channel capacity of the additive white noise channel of spectral density $N_o/2$ and transmitted power P_{av}. The actual capacity of such channels may be much larger than this, but the capacity is usually too complicated to evaluate analytically. At any rate, regardless of the statistics of the additive white noise, the optimal scheme should give considerable improvement over the best nonfeedback scheme as is demonstrated in the Gaussian case.

2. The Band-Limited Scheme. Suppose now the channel is band-limited to bandwidth W; that is, all carrier signals are restricted in bandwidth to [- W, W]. With this additional transmitter constraint imposed, the channel capacity is no longer P_{av}/N_o as in Eq. (61), but is now given by

$$C_w = W \ln \left(1 + \frac{P_{av}}{N_o W} \right) \qquad (70)$$

nats per second. For small values of P_{av}/N_oW this capacity approaches that of Eq. (61) as it should, for when $W \to \infty$ both channels are identical.

Shannon derived this capacity formula, C_w, by a random coding argument, and until the work of Schalkwijk[2] last year, no deterministic way was known for constructing a code achieving the critical rate for a band-limited white Gaussian noise channel with or without feedback. The optimum scheme for this band-limited channel is found to be essentially the same as Schalkwijk's scheme.

Now suppose the transmitter is required to use carrier signals that are restricted in bandwidth to [- W, W], where W is in cycles per second. In this case the number, N, of orthogonal carrier signals of duration less than T seconds is limited. The highest number of orthogonal carrier signals is approximately equal to 2WT.[15] Thus, the bandwidth restriction of the channel imposes the restriction between N and T to be

$$N = 2WT \quad . \tag{71}$$

From Eq. (60) with this restriction on N,

$$C(N,T)\Big|_{N=2WT} = W \ln \left(1 + \frac{P_{av}}{N_o W}\right) = C_w \tag{72}$$

and the probability of error becomes

$$P_e = 2 \text{ erfc} \left\{\sqrt{3}\, e^{(C_w - R)T}\right\} \quad . \tag{73}$$

This equation for P_e is essentially the same as the one derived by Schalkwijk.* Schalkwijk gives curves of this error probability and compares it with Slepian's work[17] which gives theoretical lower bounds on the error probabilities for this channel without feedback. In this case, too, the optimum scheme gives a considerable improvement over the theoretically optimum nonfeedback schemes. The best nonfeedback schemes have error probabilities that decrease essentially exponentially with T much like the wideband orthogonal code scheme. The optimal scheme, however, exhibits the same double exponential behavior here as in the wideband channel.

REFERENCES

1. J. P. M. Schalkwijk and T. Kailath, "A Coding Scheme for Additive Noise Channels with Feedback," IEEE International Symposium on Information Theory, 1966.

2. J. P. M. Schalkwijk, "Coding Schemes for Additive Noise Channels with Feedback," Ph.D. Thesis, Rept. SEL-65-073, Stanford Electronics Laboratories, Stanford, California, August 1965.

*In Eq. (3), if 2WT-1/2WT is replaced by one, it becomes Eq. (73).

3. C. E. Shannon, "The Zero-Error Capacity of a Noisy Channel," IRE Trans. on Information Theory, IT-2, 3, September 1951, pp. 8-19.

4. P. E. Green, "Feedback Communication Systems," Lectures on Communication System Theory, Baghdady (ed.) McGraw-Hill Book Company, 1961.

5. A. J. Viterbi, "The Effect of Sequential Decision Feedback on Communication over the Gaussian Channel," Info. and Contr., 8, 1, February 1965.

6. G. L. Turin, "Signal Design for Sequential Detection Systems with Feedback," IRE Trans. on Information Theory, IT-11, 3, July 1965, pp. 401-408.

7. ______________, "Comparison of Sequential and Nonsequential Detection Systems with Uncertainty Feedback," IRE Trans. on Information Theory, IT-12, 1, January 1966, pp. 5-8.

8. H. Robbins and S. Munro, "A Stochastic Approximation Method," Ann. Math. Stat., 22, 1951, pp. 400-407.

9. J. T. Tou, "Optimum Design of Digital Control Systems," Academic Press, New York, 1963.

10. R. Bellman, "Dynamic Programming," Princeton University Press, Princeton, New Jersey, 1957.

11. P. D. Joseph and J. T. Tou, "On Linear Control Theory," Trans. AIEE, Pt. II 80, 18 (1961).

12. T. L. Gunkel, "Optimum Design of Sampled-Data Systems with Random Parameters," Ph.D. Dissertation, Tech. Rept. No. 2102-2, Stanford Electronics Laboratories, Stanford, California, April 1961.

13. C. E. Shannon, "Probability of Error for Optimal Codes in a Gaussian Channel," Bell Sys. Tech. J., 38, 3, 1959, pp. 611-656.

14. R. M. Fano, "Transmission of Information," John Wiley and Sons, Inc., New York, 1961.

15. W. Feller, "An Introduction to Probability Theory and Its Application," John Wiley and Sons, Inc., New York, 1961.

16. J. M. Wozencraft and I. M. Jacobs, "Principles of Communication Engineering," John Wiley and Sons, Inc., New York, 1965.

17. D. Slepian, "Bounds on Communication," Bell Sys. Tech. J., 42, 1963, pp. 681-707.

A MODERN SYSTEMS APPROACH TO SIGNAL DESIGN

by

Fred C. Schweppe, Staff Member

, Lincoln Laboratory, * Massachusetts Institute of Technology

N67-29906

Introduction:

Modern systems theory can be characterized by the use of state variable concepts and optimization techniques such as Pontryagin's maximum principle (the minimum principle). This paper is intended to provide a feeling for why modern systems theory is a viable approach to many signal design problems of communication and radar systems. The basic ideas and available results are summarized; the details are left to the cited references. The discussion is restricted to the use of modern systems theory for a particular class of signal design problems, no attempt is made to survey the whole signal design field.

Author

Problem Definition:

The general signal design concept can be applied to a wide range of physical problems. However, the present discussion is concentrated on communication and radar systems. The term "channel" is applied to both radar and communication systems as a radar reflector is considered to be a channel. The term "signal" is also used in a general sense. In a particular application, the signal may be an amplitude modulation, a frequency modulation, an observation program, or simply a time function.

A signal is to be transmitted over some finite time interval $0 \leq t \leq T$. The signal must satisfy possible constraints on peak amplitude, total energy, and "bandwidth". The received output of the channel is put through a data processor to obtain the desired output at time T (or T plus system delays). There are two cases of particular interest:

1. Decision Making: The transmitted signal is one of M possible signals. The data processor's output is a decision as to which signal was transmitted. The performance of the system is measured by the probability of making an error.

2. State Estimation: The transmitted signal is used to "observe" the channel. The data processor's output is an estimate of the state of the channel at time T. The performance of the system is measured by the covariance matrix of the errors in the estimated state.

The decision making case arises both in communication systems and in the detection aspects of radar systems. The state estimation case is usually associated with radar systems wherein the channel's state corresponds to the radar target's position, velocity, size, spin rate, etc.

It is desired to find the signal (or set of M signals) that optimizes the performance subject to the imposed signal constraints. This design problem requires an assumption on the relationship between the signal (signal set) and the data processor. One possible approach assumes a fixed data processor. The approach discussed here assumes the data processor is always "matched" to the signal (signal set) in the sense that the data processor is always the optimum (physically realizable) system corresponding to the signal.

The signal design problem can be summarized as follows:

> Given a channel structure and the desired constraints on signal amplitude, energy, and bandwidth. Find the signal or set of M signals that optimizes the performance measure assuming the data processor is always optimally "matched " to the signal (signal set).

A modern systems theory approach to this signal design problem can be partitioned into three steps.

1. The channel is modeled as a dynamical system represented by state variable differential equations (possibly with stochastic inputs). The constraints are also modeled by state variable equations.

2. The data processor is modeled as the optimum dynamical system corresponding to the

* Operated with support from the U. S. Air Force.

channel model and the signal (signal set). The performance of this data processor is determined as a function of the signal (signal set).

3. Optimal control theory, in particular Pontryagin's maximum principle, is used to find the optimum signal (signal set).

More detailed discussions will now be provided on these three steps.

<u>Channel and Constraint Modeling:</u>

References 1 and 2 contain some general discussions on the representation of channels and signal constraints by state variable models. This is often a straightforward procedure as the channel and constraints can often be modeled using lumped parameter linear systems. These linear systems are then just represented by state variable equations. Good discussions on the state space modeling of linear systems can be found, for example, in Refs. 3 and 4.

Linear systems can enter a channel model in various ways. Correlated Gaussian noise can be assumed to be generated by passing white noise through a linear system. A bandwidth limited channel can be modeled as linear system with a bandpass frequency response. Stochastic (incoherent) channels such as multipath communication links and extended radar targets like clutter and planets can be modeled by a tapped delay line with correlated noise multiplying tap outputs and a final summing bus. The delays can be approximated by lumped parameter linear systems and the correlated noise obtained from white noise and linear systems.

Constraints on the allowable peak amplitude and total energy of the signal usually fit naturally into the overall analysis. There are many possible definitions of bandwidth but most of the interesting bandwidth constraints can be incorporated into a state variable framework. Two such possibilities are based on the energy contained in the signal's time derivative and on the energy transfer of a low pass linear system.

Of course, not all channels are easily modeled using linear systems and special techniques may be required. For example, channel modeling for a moving radar point reflector requires linearization of nonlinear equations. The development of "equivalent" linear models for an accelerating point reflector is discussed in Ref. 5 for both amplitude and frequency modulations.

<u>Data Processor Performance:</u>

For state estimation, the data processor performance is measured by the covariance matrix of the errors in the estimate of the state. The optimum dynamical system for data processing and its associated performance are available in the Kalman-Bucy formulation of the Weiner-Hopf filtering problem for Gaussian processes, see Refs. 6 and 7.

For decision making the data processor performance is measured by the probability of error. The optimum dynamical system for data processing (likelihood function generation) is discussed in Ref. 8 for Gaussian processes. Performance measures for the optimum processor are discussed in Ref. 9 in terms of two "distance" measures; the divergence and the Bhattacharyya distance. These measures are not always equivalent to the probability of error but they can provide bounds and are felt to be adequate for signal design.

For both state estimation and decision making, the performance is evaluated in terms of the solution of a matrix Riccati equation which is a first order, non-linear, matrix system of ordinary differential equations with time as the independent variable. This matrix Riccati equation is the one associated with the optimum time varying linear filter. The signal (signal set) appears in this Riccati equation as a time varying "coefficient".

<u>Signal Optimization:</u>

The Riccati equation which governs performance is a matrix system of first order differential equations. This Riccati equation can be considered to be a state space representation of some hypothetical dynamical system. The signal (signal set) can be considered to be the input to this hypothetical dynamical system and the performance measure can be considered to be its output. Viewed in this light, the signal design problem is the same as the "classical" optimum control problem of designing the input which gives the best output of a dynamical system.

A major tool of optimum control theory is Pontryagin's maximum principle (see for example, Ref. 3). Pontryagin's maximum principle yields necessary conditions which the optimum signal (signal set) must satisfy. These necessary conditions provide general information on the overall structure of the optimum. For numerical results, a two point boundary value problem must be solved and computer techniques are often required. Some of the results obtained using the maximum principle will now be summarized.

Optimum signals have been calculated for the state estimation case of a radar observing an accelerating point reflector in the presence of additive white noise. In Ref. 10, it is shown that the optimum frequency modulation is frequency switching between the allowable bandwidth limits. In Ref. 11, it is shown that the optimum amplitude modulation under peak power and total energy constraints is a pulse train of at most three pulses; all pulses having the maximum allowable peak power. In both references, the performances of the optimum signals are compared with more conventional signals.

References 10 and 11 contain results for specific problems. Reference 12 contains a more general result which is called the on-off principle. A loose statement of the on-off principle is now given. Assume the channel is as shown in Figure 1. The signal, u(t) $0 \le t \le T$, is considered to be an "instantaneous power". The signal constraints are:

$$\int_0^T u(t)\, dt = E \qquad \text{(total energy)}$$

$$0 \le u(t) \le U \qquad \text{(peak power)} \quad .$$

There are two problems of interest:

1. State Estimation: The switch is closed and the state of the dynamical system is to be estimated.

2. Decision Making: The position of the switch is to be estimated.

The necessary conditions which the optimum signal must satisfy prove that for either problem the optimum u(t) at any t, $0 \le t \le T$, is either 0 or U ; that is the optimum power level switches back and forth between full power and zero with no intermediate values. Thus the on-off principle states that the general structure of the optimum is independent of the details of the dynamical system and the correlated noise (the pulse train of Ref. 11 is a special case of the on-off principle). Of course, the actual "switch times" depend on the details of the dynamical system and the correlated noise and switch time calculation requires the solution of a two point boundary value problem.

Discussion:

The use of modern system concepts for signal design is really a philosophy of approach rather than a single technique. The crux of this approach is the use of state variable models so that the system performance can be considered to be the output of a dynamical system (the Riccati equation) whose input is the signal (signal set). Given such a model, the signal design proceeds directly using the techniques of optimal control theory.

In one sense this modern system approach is just a reformulation of a problem in new terms. However, this reformulation is extremely valuable as it makes available the powerful engineering and mathematical tools of modern systems theory. These tools enable the actual solution of difficult signal design problems involving finite time intervals, time varying systems and realistic signal constraints.

The many possibilities for further work include:

1. Development of other general structural properties like the on-off principle.

2. Development of computation algorithms tailored to solving the necessary two point boundary value problems.

3. Calculation of explicit solutions for specific problems.

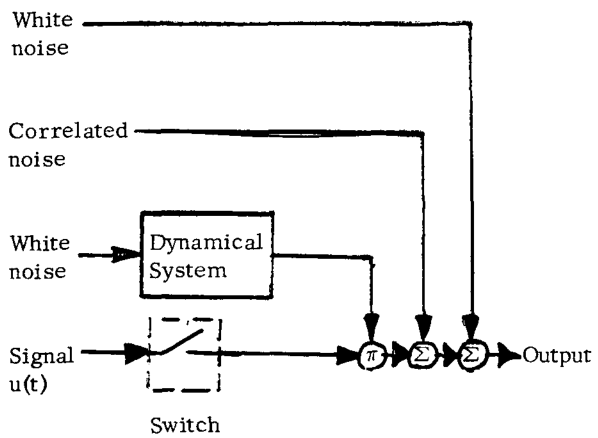


Figure 1

Channel Model for On-Off Principle

References

[1] Schweppe, F. C., "Optimization of Signals," 1964-4 Lincoln Laboratory, MIT; also presented at International Conference on Microwaves, Circuit Theory, and Information Theory, Tokyo, Japan, June 1964.

[2] Athans, M., and Schweppe, F. C., "On Optimal Waveform Design via Control Theoretic Concepts I: Formulation," submitted to Information and Control; also presented at 1966 Joint Automatic Control Conference.

[3] Athans, M., and Falb, P. L., Optimal Control; An Introduction to the Theory and Its Application, McGraw-Hill Book Co., N. Y., 1966.

[4] Zadeh, L. A., and Desoer, C. A., Linear System Theory; The State Space Approach, McGraw-Hill Book Co., N. Y., 1963.

[5] Schweppe, F. C., "On the Accuracy and Resolution of Radar Signals," IEEE Trans. on Aerospace and Electronic Systems, Vol. AES-1, No. 3, December 1965, pp. 235-245.

[6] Kalman, R. E., "New Methods and Results in Linear Prediction and Filtering Theory," RIAS Technical Report 61-1, 1961.

[7] Kalman, R. E., and Bucy, R. S., "New Results in Linear Filtering and Prediction Theory," Trans. ASME J. of Basic Engineering, 830, pp. 95-108, 1961.

[8] Schweppe, F. C., "Evaluation of Likelihood Functions for Gaussian Signals," IEEE Trans. on Information Theory, Vol. IT-11, No. 1, January 1965, pp. 61-70.

[9] Schweppe, F. C., "On the Distance Between Gaussian Processes: The State Space Approach," submitted to Information and Control; also presented at 1966 Joint Automatic Control Conference.

[10] Schweppe, F. C., "Radar Frequency Modulations for Accelerating Targets Under a Bandwidth Constraint," IEEE Trans. on Military Electronics, Vol. MIL-9, No. 1, January 1965, pp. 25-32.

[11] Schweppe, F. C., and Gray, D., "Radar Signal Design Subject to Simultaneous Peak and Average Power Constraints," IEEE Trans. on Information Theory, Vol. IT-12, No. 1, January 1966, pp. 13-25.

[12] Athans, M., and Schweppe, F. C., "On Optimal Waveform Design via Control Theoretic Concepts II: The On-Off Principle," in preparation.