A STUDY OF KALMAN FILTERING

763

by

JOHN Y. S. HO

B. S., National Taiwan University, 1962

A MASTER'S REPORT

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

Department of Electrical Engineering

KANSAS STATE UNIVERSITY
Manhattan, Kansas

1967

Approved by:

Lloyd W. Harris

Major Professor

TABLE OF CONTENTS

INTRODUCTION

Wiener introduced a fresh approach to the study of information transmission in the presence of perturbing noise in 1942. His pioneering work showed that the two problems, prediction of random signals, and separation of random signals from random noise, lead to the Wiener-Hopf integral equation. He also presented the solution for the special case of stationary statistics and rational spectra. Many extensions and generalizations of Wiener's work followed. Most of the work has been done in the frequency domain using transform methods to obtain the specifications of a linear dynamical system which accomplishes the prediction and filtering of the random signals. These ideas form the so-called "Classical Filtering Theory". During the past seven years, many new techniques and concepts have been introduced in this area of study because of the advances in digital computer technology and the challenge of aerospace technology. Most of this work has been done in the time domain using the concepts of state and matrix theory. These ideas form what is called "Modern Filtering Theory". There is no doubt that Kalman's work is the most notable. In 1960, Kalman presented a new approach to the standard filtering and prediction problems. His work combined two well known ideas. One is that a dynamical system is described by the "state-transition" method. The other is that linear filtering is regarded as an orthogonal projection in Hilbert space. (Hilbert space is a Banach space whose norm has the parallelogram property, Ref. 18.)

He also assumed that the required statistical data is given in such a form that determination of the optimal filter is highly simplified, with a single equation covering all cases. This single equation is called the variance equation; it is a non-linear differential (difference) equation of the Riccati type. Hence the variance equation is closely related to the Hamiltonian differential equations of the calculus of variations. An exact formula for the solution of the variance equation is available. The actual solution consists of the specification of the differential equation governing the optimal filter.

Wiener described the random process by its power spectral density or correlation function. Kalman assumes the random process to be Markovian. In other words, Kalman describes the linear dynamic system by a set of first-order differential (or difference) equations and the Wiener problem is approached from the point of view of conditional distributions. All statistical calculations and results are based on means and covariances. No other statistical data are needed.

FUNDAMENTALS OF KALMAN FILTERING

Notation

Vectors will be denoted by small underlined letters: $\underline{u}$, $\underline{v}$, . . ., $\underline{x}$, $\underline{y}$, $\underline{z}$ with coordinates u_i, v_i, . . ., x_i, y_i, z_i. Matrices will be denoted by underlined Roman and Greek capitals: $\underline{F}$, $\underline{G}$, . . ., $\underline{\Phi}$, $\underline{P}$ with elements f_{ij}, g_{ij}, . . ., p_{ij}. The unit matrix is $\underline{I}$; the prime denotes the transposed matrix. Constants will be denoted by small Greek letters. Time will be denoted by t, t_0, t_1, or τ. These may be arbitrary real numbers (continuous time case) or arbitrary integers (discrete time case). The inner product of $\underline{x}$ and $\underline{y}$ is denoted by $\underline{x'y}$. The norm $\|\underline{x}\|$ is $(\underline{x}'\underline{x})^{1/2}$ or $(\underline{x}, \underline{x})^{1/2}$. The quadratic form $\underline{x'Ax}$ is denoted by $\|x\|^2_{\underline{A}}$, if $\underline{A}$ is a symmetric nonnegative definite matrix. Numerical quantities will always be real numbers. The symbol $E\{\}$ denotes the expectation operator. Covariance matrices are denoted by cov $[\underline{x}, \underline{x}]$ and cov$[\underline{x}, \underline{y}]$, where

$$\begin{aligned}
\text{cov}\,[\underline{x}, \underline{x}] &= E\{(\underline{x} - E\{\underline{x}\})(\underline{x} - E\{\underline{x}\})'\} \\
&= E\{\underline{x}, \underline{x}'\} \qquad \text{if } E\{\underline{x}\} = \underline{0} \\
\text{cov}\,[\underline{x}, \underline{y}] &= E\{(\underline{x} - E\{\underline{x}\})(\underline{y} - E\{\underline{y}\})'\} \\
&= E\{\underline{xy}'\} \qquad \text{if } E\{\underline{x}\} = \underline{0} \text{ and } E\{\underline{y}\} = \underline{0}.
\end{aligned}$$

Preliminaries

A linear dynamical system governed by a difference equation (discrete time case) can always be described in the standard form.

$$\underline{x}(t + 1) = \underline{\Phi}(t + 1, t)\underline{x}(t) + \underline{\Delta}(t + 1, t)\underline{u}(t) \qquad (1)$$

The output equation is

$$\underline{y}(t) = \underline{H}(t)\underline{x}(t). \qquad (2)$$

The observed signal is

$$\underline{z}(t) = \underline{y}(t) + \underline{v}(t) = \underline{H}(t)\underline{x}(t) + \underline{v}(t)\,. \tag{3}$$

A linear dynamical system governed by an ordinary differential equation (continuous time case) can always be described in the standard form.

$$\frac{d\underline{x}}{dt} = \underline{F}(t)\underline{x} + \underline{G}(t)\underline{u}(t) \tag{4}$$

The output equation is

$$\underline{y}(t) = \underline{H}(t)\underline{x}(t)\,. \tag{5}$$

The observed signal is

$$\underline{z}(t) = \underline{y}(t) + \underline{v}(t) = \underline{H}(t)\underline{x}(t) + \underline{v}(t)\,. \tag{6}$$

In both cases, $\underline{x}$ is an n-vector, called the state. The coordinates x_i of $\underline{x}$ are called state variables. $\underline{u}(t)$ is an m-vector, called the control function. It is the input of the system. $\underline{v}(t)$ is a p-vector. It is an additional input. It reflects the fact that physical measurement of observables can never be made with infinite precision. It is the measurement noise. $\underline{y}(t)$ is the output of the system. It is also a p-vector whose components are linear combinations of the state variables. $\underline{z}(t)$ is the observed value of the output of the system. It is a p-vector too. $\underline{F}(t)$ is an n x n matrix. The structure of this matrix decides the nature of the state transition matrix; thus the nature of all solutions, whether forced or unforced, depend upon this matrix. $\underline{G}(t)$ is an n x m matrix. It is a coupling matrix, as the structure of this matrix determines how the input is coupled to the various state variables. $\underline{H}(t)$ is a p x n matrix. It is also a coupling matrix, coupling the state variables to the output. $\underline{\Phi}(t)$ is a transition matrix. Its properties will be discussed later. $\underline{\Delta}(t)$ is an n x m matrix. If $\underline{u}(t)$ is piecewise constant, then

$$\underline{\Delta}(t + 1, t) = \int_t^{t+1} \underline{\Phi}(t + 1, \tau)\underline{G}(\tau)d\tau \ . \tag{7}$$

If $\underline{G}$, $\underline{H}$, $\underline{F}$ are constant, then the system is said to be stationary; if $\underline{u}(t) = \underline{0}$, then the system is free.

The general solution of equation (4) has the form

$$\underline{x}(t) = \underline{\Phi}(t, t_0)\underline{x}(t_0) + \int_{t_0}^{t} \underline{\Phi}(t, \tau)\underline{G}(\tau)\underline{u}(\tau)d\tau \ , \tag{8}$$

where the transition matrix is characterized by the properties (Ref. 10)

$$\frac{d}{dt}\underline{\Phi}(t, t_0) = \underline{F}(t)\underline{\Phi}(t, t_0) \ , \tag{9}$$

$$\underline{\Phi}(t_0, t_0) = \underline{I} \qquad \text{for all } t_0 \ , \tag{10}$$

inverse rule

$$\underline{\Phi}^{-1}(t_1, t_0) = \underline{\Phi}(t_0, t_1) \quad \text{for all } t_0, t_1 \ , \tag{11}$$

product rule

$$\underline{\Phi}(t_2, t_0) = \underline{\Phi}(t_2, t_1)\underline{\Phi}(t_1, t_0) \ . \tag{12}$$

The last property of $\underline{\Phi}$ justifies its being called a transition matrix. Because of the manner in which $\underline{\Phi}$ is defined, this matrix can never be singular.

If $\underline{F}$ is a constant matrix, then

$$\underline{\Phi}(t, t_0) = \exp \underline{F}(t - t_0) = \sum_{i=0}^{\infty} \frac{[(\underline{F}(t - t_0)]^i}{i!} \qquad (13)$$

There is no simple way to compute $\underline{\Phi}$ explicitly when F is not constant.

The first term of equation (8) represents the initial condition response of the system state variables. The second term represents the forced response.

If the $\underline{u}(t)$ is piecewise constant, then any linear differential equation may be converted into a linear difference equation in such a way that at integer values of time the solutions of the differential equation agree with the solution of the difference equation.

Statement of the Problem

The statistical correlation between random signals observed over some interval of time is explained by the presence of a dynamic (linear or nonlinear) system between the primary random source and the observer. Hence random processes may be thought of as the output of a dynamic system (linear or nonlinear) excited by an independent gaussian random process. If the observed random signal $\underline{z}(t)$ is also gaussian, one may assume that the dynamic system which is between the observer and the primary source $\underline{u}(t)$ is linear.

Consider a linear dynamical system subjected to random disturbances and measurement noise. Assume that the state $\underline{x}(t)$ of the system cannot be observed directly but only through the output $\underline{y}(t)$, which can be measured only in the presence of additive gaussian white noise $\underline{v}(t)$. In addition, the system is subjected to a random input disturbance in the form of gaussian white noise $\underline{u}(t)$. Assume that physical relationships between the state $\underline{x}(t)$ and the driving noise $\underline{u}(t)$ and between the observed values $\underline{z}(t)$ and the state $\underline{x}(t)$ and the measurement noise $\underline{v}(t)$ are given. The statistical characteristics of the driving noise $\underline{u}(t)$ and measurement noise $\underline{v}(t)$ are also assumed. Then, from the actually observed values, $\underline{z}(\tau)$, over some interval of time, $t_0 \leq \tau \leq t$, one wants to find the optimal (in some sense) estimate $\hat{\underline{x}}(t_1)$ of $\underline{x}(t_1)$ at time t_1. In the case where $t_1 < t$, the problem is referred to as a data-smoothing (interpolation) problem; if $t_1 = t$, it is called the filtering problem; and if $t_1 > t$, it is called the prediction (extrapolation) problem. Collectively, these cases are referred to as the estimation problem.

Kalman made three key assumptions, which are: (A_1) The original message $\underline{x}(t)$ is assumed to be a random process generated by a mathematical model of the type described by either equation (1) or equation (4). (A_2) The observed signal $\underline{z}(t)$ is an additive combination of the output signal and a white noise described either by equations (3) or (6). (A_3) The measurement of $\underline{z}(t)$ starts at some fixed instant t_0 (which may be $-\infty$), at which time $\operatorname{cov}\left[\underline{x}(t_0), \underline{x}(t_0)\right]$ is known.

From these three assumptions, there are two cases to be discussed. These are the case where the message is a Gauss-Markov Sequence and the case where the message is a Gauss-Markov Process.

The message is a Gauss-Markov Sequence generated by the recursive relation

$$\underline{x}(t + 1) = \underline{\Phi}(t + 1, t)\underline{x}(t) + \underline{\Delta}(t + 1, t)\underline{u}(t) . \tag{1}$$

The observed signal (which is message plus noise) is

$$\underline{z}(t) = \underline{y}(t) + \underline{v}(t) = \underline{H}(t)\underline{x}(t) + \underline{v}(t) . \tag{3}$$

$\underline{u}(t)$ is a gaussian white noise sequence; $\underline{u}(t)$ evaluated at different times is independent. Hence

$$\mathrm{cov}\left[\underline{u}(t_1), \underline{u}(t_2)\right] = \underline{0} \qquad \text{if } t_1 \neq t_2 . \tag{14}$$

$\underline{u}(t)$ has zero mean, that is,

$$E\left\{\underline{u}(t)\right\} = \underline{0} \qquad \text{for all } t . \tag{15}$$

Then the covariance matrix is

$$\mathrm{cov}\left[\underline{u}(t), \underline{u}(t)\right] = \underline{Q}(t) \qquad \text{for all } t. \tag{16}$$

$\underline{x}(t)$ is a gaussian random sequence with zero mean and arbitrary covariance, independent of $\underline{u}(t)$. $\underline{x}(t)$ satisfies the Markov property, that is, the conditional probability distribution of $\underline{x}(t)$, given $\underline{x}(\tau)$, $t_0 \leq \tau \leq t$, is identical with the probability distribution of $\underline{x}(t)$ given the last observation $\underline{x}(t - 1)$ (Ref. 6).

$$P_r\left\{\underline{x}(t) \leq \Sigma_1 \,\middle|\, \underline{x}(t - 2) \;.\;.\;.\; \underline{x}(t_0)\right\}$$
$$= P_r\left\{\underline{x}(t) \leq \Sigma_1 \,\middle|\, \underline{x}(t - 1)\right\} \tag{17}$$

Hence $\underline{x}(t)$ is a Gauss-Markov sequence.

$\underline{z}(t)$ is also a gaussian random sequence, because gaussian random signals remain gaussian after passing through a linear system.

$\underline{v}(t)$ is also a gaussian white noise with zero mean and covariance $\underline{R}(t)$, independent of $\underline{u}(t)$.

$$\text{cov}\left[\underline{v}(t_1),\ \underline{v}(t_2)\right] = \underline{0} \qquad \text{if } t_1 \neq t_2 \tag{18}$$

$$E\left\{\underline{v}(t)\right\} = \underline{0} \qquad \text{for all } t \tag{19}$$

$$\text{cov}\left[\underline{v}(t),\ \underline{v}(t)\right] = \underline{R}(t) \tag{20}$$

$$\text{cov}\left[\underline{u}(t),\ \underline{v}(t)\right] = \underline{0} \qquad \text{for all } t \tag{21}$$

Thus a random sequence may be thought of as the output of a dynamical system excited by two independent gaussian random sequences.

A Gauss-Markov Process is the limiting case of Gauss-Markov sequence when the interval between successive values of time tends to zero. Then equations (1) and (3) convert to the continuous time case as follows:

$$\frac{dx}{dt} = \underline{F}(t)\underline{x}(t) + \underline{G}(t)\underline{u}(t) \tag{4}$$

$$\underline{z}(t) = \underline{y}(t) + \underline{v}(t) = \underline{H}(t)\underline{x}(t) + \underline{v}(t) \tag{6}$$

The random processes $\underline{u}$ and $\underline{v}$ are defined in such a way that at integer values of time the random processes $\underline{x}$ and $\underline{z}$ generated by equations (4) and (6) agree with those generated by equations (1) and (2). Hence the sample functions are assumed to be piecewise constant over intervals of length 1.

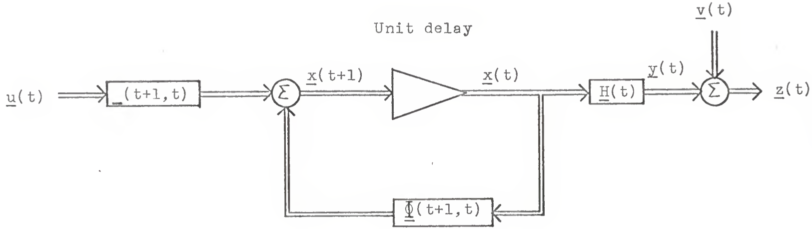

Fig. 1. Representation of Gauss-Markov sequence.

$$\underline{u}(T) = \underline{u}(T + \tau)$$
$$\underline{v}(T) = \underline{v}(T + \tau)$$

where T is an integer and

$$0 \leqslant \tau \leqslant 1.$$

It is assumed that

$$E\left\{\underline{u}(t)\right\} = \underline{0}$$
$$E\left\{\underline{v}(t)\right\} = \underline{0}$$

$$E\left\{\underline{u}(t),\ \underline{u}(t)\right\} = \underline{Q}(t)\delta(t - \tau) \tag{22}$$

$$\operatorname{cov}\left[\underline{v}(t),\ \underline{v}(\tau)\right] = \underline{R}(t)\underline{\delta}(t - \tau) \tag{23}$$

and

$$\operatorname{cov}\left[\underline{v}(t),\ \underline{u}(t)\right] = \underline{0} \qquad \text{for all } t,\ \tau\ .$$

$\delta(t)$ is the Dirac-delta function, which has the following characteristic properties (Ref. 3):

$$\delta(t - t_0) = \frac{d}{dt}\left[1(t - t_0)\right] \tag{24}$$

where $1(t)$ is a unit step function, and

$$\int_{-\infty}^{t_0} \delta(t - t_0)dt = \int_{t_0}^{\infty} \delta(t - t_0)dt = 1/2 \tag{25}$$

$$\int_{-\infty}^{\infty} f(t)\delta(t - t_0)dt = \int_{-\infty}^{\infty} f(t)\delta(t_0 - t)dt = f(t) \tag{26}$$

and

$$\int_{t_1}^{t_2} f(t)\delta(t - t_0)dt = \frac{1}{2}\, f(t_0) \tag{27}$$

if $t_1 = t_0$, or $t_2 = t_0$.

The values of the sample functions of $\underline{u}$ and $\underline{v}$ are to be regarded as Dirac-delta functions of vanishing small areas. Mathematically speaking, this definition of course is not rigorous, since $\delta(t)$ is not a well defined function.

The block diagram of this system is shown in Fig. 2.

The estimation problem is formulated as follows. Given the actually observed value of a random process $\underline{z}(\tau)$ over some interval of time, $t_0 \leq \tau \leq t$, find the optimal estimate, $\hat{\underline{x}}(t_1)$, of another related random process $\underline{x}(t_1)$ such that the estimate $\hat{\underline{x}}(t_1)$ minimizes the expected losses

$$E\left\{E\left\{L(\underline{x}(t_1) - \hat{\underline{x}}(t_1))\,\middle|\,\underline{z}(\tau),\ t_0 \leq \tau \leq t\right\}\right\} \tag{28}$$

or equivalently

$$E\left\{L(\underline{x}(t_1) - \hat{\underline{x}}(t_1))\,\middle|\,\underline{z}(\tau),\ t_0 \leq \tau \leq t\right\}. \tag{29}$$

The loss function $L(\epsilon)$ is a scalar valued, positive, non-decreasing function of the estimation error

$$\underline{\epsilon} = \underline{x}(t_1) - \hat{\underline{x}}(t_1) .$$

The loss function $L(\underline{\epsilon})$ must satisfy

$$L(\underline{0}) = 0$$

$$L(\underline{\epsilon}_1) \geq L(\underline{\epsilon}_2) \geq 0$$

where $\|\underline{\epsilon}_1\| \geq \|\underline{\epsilon}_2\| \geq 0$.

Let $\underline{x}$ be an n-dimensional random vector with mean $\underline{\xi}$ and distribution function $F(\underline{\xi})$. Suppose $F(\underline{\xi})$ has the following properties. F is symmetrical about its mean $\underline{\xi}$.

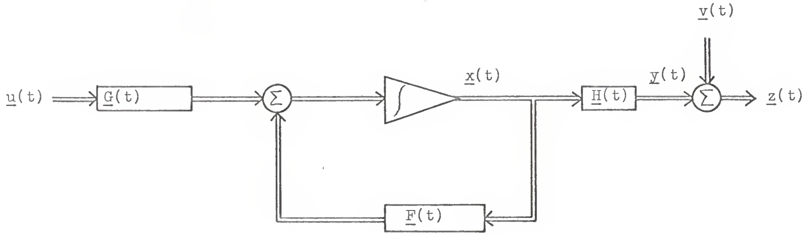

Fig. 2. Representation of Gauss-Markov process.

$$F(\underline{\xi} - \bar{\underline{\xi}}) = 1 - F(\bar{\underline{\xi}} - \underline{\xi})$$

convex for $\xi_i \leq \bar{\xi}_i$

$$F\left(\lambda \xi_{i1} + (1 - \lambda)\xi_{i2}\right) \leq \lambda F(\xi_{i1}) + (1 - \lambda)F(\xi_{i2})$$

for all $\xi_{i1}, \xi_{i2} \leq \bar{\xi}_i$ and $0 \leq \lambda \leq 1$.

Then the estimate $\hat{\underline{x}}(t_1)$ which minimizes the expected loss is the conditional expectation

$$\hat{\underline{x}}(t_1) = \hat{\underline{x}}(t_1 | t) = E\left\{ \underline{x}(t_1) \middle| \underline{z}(\tau), \quad t_0 \leq \tau \leq t \right\} \qquad (30)$$

If the loss function is defined by the mean squared error, i.e.,

$$L(\underline{\epsilon}) = (\underline{\epsilon}, \underline{\epsilon}) = \left(\left[\underline{x} - \hat{\underline{x}}\right], \left[\underline{x} - \hat{\underline{x}}\right]\right) = \sum_{i=1}^{n} (x_i - \hat{x}_i)^2$$

then the optimal estimate is also the conditional expectation,

$$\hat{\underline{x}}(t_1) = \hat{\underline{x}}(t_1 | t) = E\left\{\underline{x}(t_1) \middle| \underline{z}(\tau) \quad t_0 \leq \tau \leq t\right\}$$

without imposing the restrictions that the distribution function $F(\underline{\xi})$ be both symmetric and convex (Ref. 8).

The conditional mean, which supplies the minimum expected loss for many loss functions, plays an important role in the filtering problem. Of course, if the conditional distribution is known, the optimal estimate $\hat{\underline{x}}(t)$ can be computed for any loss function.

A gaussian distribution is both convex and symmetric about the mean. It is obvious that the optimal estimate is always the conditional expectation $\hat{\underline{x}}(t_1 | t)$. The conditional probability distribution of a gaussian process is completely described by its mean and covariance. If in addition the process is also Markovian, then it suffices to know the mean and covariance at one instant of time. The calculation of the optimal estimate involves only the means and covariance matrices of the gaussian process. Thus $\hat{\underline{x}}(t_1) = \hat{\underline{x}}(t_1|t)$ is clearly the optimal estimate for the class of all of the random processes with the same means and covariance matrices as the gaussian process.

In the gaussian process, the optimal estimate $\hat{\underline{x}}(t_1)$ of $\underline{x}(t_1)$ is

$$\hat{\underline{x}}(t_1) = \hat{\underline{x}}(t_1|t) = E\left\{\underline{x}(t_1) \middle| \underline{z}(\tau),\ t_0 \leq \tau \leq t\right\} \tag{31}$$

where $\hat{\underline{x}}(t_1|t)$ is an unbiased estimate of $\underline{x}(t_1)$. That is,

$$E\left\{\hat{\underline{x}}(t_1|t)\right\} = E\left\{\underline{x}(t_1)\right\} = \underline{0}.$$

Define

$$\tilde{\underline{x}}(t_1|t) = \underline{x}(t_1) - \hat{\underline{x}}(t_1|t) \tag{32}$$

where $\tilde{\underline{x}}(t_1|t)$ is the error between the actual value $\underline{x}(t_1)$ and its conditional expectation. The conditional expectation of gaussian random process $\underline{x}(t)$is identical with the orthogonal projection of $\underline{x}(t_1)$ upon the sample space $\underline{z}(\tau), t_0 \leq \tau \leq t$ (Ref. 8). Hence the optimal estimate $\hat{\underline{x}}(t_1|t)$ of $\underline{x}(t_1)$ is the orthogonal projection $\underline{x}(t_1)$ on a linear manifold or vector space, generated by $\underline{z}(\tau)$, $t_0 \leq \tau \leq t$. $\hat{\underline{x}}(t_1|t)$ will minimize the expected loss,

$$E\left\{L(\underline{x}(t) - \hat{\underline{x}}(t|t))\right\} = E\left\{\sum_{i=1}^{n} \left[x_i(t) - x_i(t|t)\right]^2\right\}$$

That will be discussed in more detail in the continuous case.

By gaussianness, $\hat{\underline{x}}(t_1|t)$ and $\tilde{\underline{x}}(t_1|t)$ are independent random variables. The covariance matrix of $\tilde{\underline{x}}(t_1|t)$ is defined by

$$\underline{P}(t_1|t) = \mathrm{cov}\left[\hat{\underline{x}}(t_1|t),\ \tilde{\underline{x}}(t_1|t)\right] \tag{33}$$

The quantities $\hat{\underline{v}}(t|t)$, $\tilde{\underline{v}}(t|t)$, $\hat{\underline{u}}(t|t)$ . . . $\tilde{\underline{z}}(t|t)$ are defined similarly.

Solution of the Problem

In a Gauss-Markov sequence the message is a random sequence generated by the equation

$$\underline{x}(t+1) = \underline{\Phi}(t+1,\ t)\underline{x}(t) + \underline{\Delta}(t+1,\ t)\underline{u}(t)\ .$$

Repeated use of the above equation yields

$$\underline{x}(t+2) = \underline{\Phi}(t+2,\ t+1)\underline{x}(t+1) + \underline{\Delta}(t+2,\ t+1)\underline{u}(t+1),$$

$$\begin{aligned}\underline{x}(t+3) &= \underline{\Phi}(t+3,\ t+2)\underline{x}(t+2) + \underline{\Delta}(t+3,\ t+2)\underline{u}(t+2)\\ &= \underline{\Phi}(t+3,\ t+1)\underline{x}(t+1)\\ &\quad + \underline{\Phi}(t+3,\ t+2)\underline{\Delta}(t+2,\ t+1)\underline{u}(t+1)\\ &\quad + \underline{\Phi}(t+3,\ t+3)\underline{\Delta}(t+3,\ t+2)\underline{u}(t+2)\ ,\end{aligned}$$

$$\begin{aligned}\underline{x}(t+4) &= \underline{\Phi}(t+4,\ t+3)\underline{x}(t+3) + \underline{\Delta}(t+4,\ t+3)\underline{u}(t+3)\\ &= \underline{\Phi}(t+4,\ t+1)\underline{x}(t+1)\\ &\quad + \underline{\Phi}(t+4,\ t+2)\underline{\Delta}(t+2,\ t+1)\underline{u}(t+1)\end{aligned}$$

$$+ \underline{\Phi}(t + 4,\ t + 3)\underline{\Delta}(t + 3,\ t + 2)\underline{u}(t + 2)$$

$$+ \underline{\Phi}(t + 4,\ t + 4)\underline{\Delta}(t + 4,\ t + 3)\underline{u}(t + 3)$$

$$= \underline{\Phi}(t + 4,\ t + 1)\underline{x}(t + 1) + \sum_{\tau_1=t+1}^{(t+4)-1} \underline{\Phi}(t + 4,\ \tau_1 + 1)\underline{\Delta}(\tau_1 + 1,\ \tau_1)\underline{u}(\tau_1)\ .$$

Hence by induction

$$\underline{x}(t_1) = \underline{\Phi}(t_1,\ t + 1)\underline{x}(t + 1) + \sum_{\tau_1=t+1}^{t_1-1} \underline{\Phi}(t_1,\ \tau_1 + 1)\underline{\Delta}(\tau_1 + 1,\ \tau_1)\underline{u}(\tau_1) \quad \text{for } t_1 \geq t + 2 \tag{34}$$

Taking conditional expectation of both sides with respect to $\underline{z}(\tau)$, $t_0 \leqslant \tau \leqslant t$, yields

$$E\left\{\underline{x}(t_1) \middle| \underline{z}(\tau),\ t_0 \leqslant \tau \leqslant t\right\} = E\left\{\underline{\Phi}(t_1,\ t + 1)\underline{x}(t + 1) + \sum_{\tau_1=t+1}^{t_1-1} \underline{\Phi}(t_1,\ \tau_1 + 1)\underline{\Delta}(\tau_1 + 1, \tau_1)\underline{u}(\tau_1) \middle| \underline{z}(\tau),\ t_0 \leqslant \tau \leqslant t\right\}$$

since $\underline{u}(t)$ is independent of $\underline{z}(\tau)$, $t_0 \leqslant \tau \leq t$, and $E\{\underline{u}(\tau_1)\} = \underline{0}$. Hence

$$\hat{\underline{x}}(t_1|t) = \underline{\Phi}(t_1,\ t + 1)\hat{\underline{x}}(t + 1\ t) \text{ for } t_1 \geqslant t + 1 \ . \ . \ . \ (v_d).$$

This equation applies to the optimal prediction (extrapolation) problem. So far, no similarly simple formula is known for data smoothing (interpolation) case, for $t_1 < t$.

Supposing $\hat{\underline{x}}(t|t - 1)$ is known. $\hat{\underline{x}}(t + 1|t)$ can be computed by induction.

Consider the linear manifold $M(t)$ generated by the observation $\underline{z}(t_0), \ldots, \underline{z}(t)$. This manifold can be decomposed into two parts. One part is the space $M(t - 1)$ generated by $\underline{z}(t_0), \underline{z}(t_0 + 1) \ldots \underline{z}(t - 1)$, while the second part is the space $\underline{N}(t)$ generated by $\underline{\tilde{z}}(t|t - 1)$, the component of $\underline{z}(t)$ that is orthogonal to $M(t - 1)$.

$$\underline{\tilde{z}}(t|t - 1) = \underline{z}(t) - \underline{H}(t)\underline{\hat{x}}(t|t - 1) = \underline{H}(t)\underline{\tilde{x}}(t|t - 1) + \underline{v}(t)$$

Two linear manifolds $M(t - 1)$ and $N(t)$ are said to be orthogonal, if every vector in one manifold is orthogonal to every vector in the other manifold. The two sets of gaussian random variables are independent; hence the conditional expectations may be computed separately.

$$E\{\underline{x}(t + 1)|M(t)\} = E\{\underline{x}(t + 1)|M(t - 1)\} + E\{\underline{x}(t + 1)|N(t)\}$$

$$E\{\underline{x}(t + 1)|\underline{z}(\tau),\ t_0 \leq \tau \leq t = E\{\underline{x}(t + 1)|\underline{z}(\tau),\ t_0 \leq \tau \leq t - 1\}$$
$$+ E\{\underline{x}(t + 1)|\underline{z}(t|t - 1)\}$$

Then

$$\underline{\hat{x}}(t + 1|t) = \underline{\Phi}(t + 1, t)\underline{\hat{x}}(t|t - 1)$$
$$+ \underline{\Delta}(t + 1, t)E\{\underline{u}(t)|M(t - 1)\}$$
$$+ E\{\underline{x}(t + 1)|\underline{\tilde{z}}(t|t - 1)\}\ .$$

Since $E\{\underline{u}(t)|M(t - 1)\} = \underline{0}$,

$$\underline{\hat{x}}(t + 1|t) = \underline{\Phi}(t + 1, t)\underline{\hat{x}}(t|t - 1) + E\{\underline{x}(t + 1)|\underline{\tilde{z}}(t|t - 1)\}\ . \tag{35}$$

By gaussianness and knowing that (Ref. 9)

$$E\left\{\underline{x}(t+1) \middle| \underline{\tilde{z}}(t|t-1)\right\} = E\left\{\underline{x}(t+1)\right\}$$

$$+ \operatorname{cov}\left[\underline{x}(t+1), \underline{\tilde{z}}(t|t-1)\right] \operatorname{cov}\left[\underline{\tilde{z}}(t|t-1), \underline{\tilde{z}}(t|t-1)\right]^{-1}$$

$$\cdot \left[\underline{\tilde{z}}(t|t-1) - E\left\{\underline{\tilde{z}}(t|t-1)\right\}\right],$$

one can write

$$E\left\{\underline{x}(t+1) \middle| \underline{\tilde{z}}(t|t-1)\right\} = \operatorname{cov}\left[\underline{x}(t+1), \underline{\tilde{z}}(t|t-1)\right]$$

$$\operatorname{cov}\left[\underline{\tilde{z}}(t|t-1), \underline{\tilde{z}}(t|t-1)\right]^{-1}$$

$$\cdot \left[\underline{\tilde{z}}(t|t-1)\right] \tag{36}$$

since

$$E\left\{\underline{x}(t+1)\right\} = \underline{0}$$

and

$$E\left\{\underline{\tilde{z}}(t|t-1)\right\} = E\left\{\underline{H}(t)\underline{\tilde{x}}(t|t-1) + \underline{v}(t)\right\} = \underline{0}\ .$$

The first factor on the right-hand side of equation (36) can be written

$$\operatorname{cov}\left[\underline{x}(t+1), \underline{\tilde{z}}(t|t-1)\right] = \operatorname{cov}\left[\underline{\Phi}(t+1, t)\underline{x}(t)\right.$$

$$\left. + \underline{\Delta}(t+1, t)\underline{u}(t), \underline{H}(t)\underline{\tilde{x}}(t|t-1) + \underline{v}(t)\right]$$

$$= \operatorname{cov}\left[\underline{\Phi}(t+1, t)\underline{x}(t), \underline{H}(t)\underline{\tilde{x}}(t|t-1)\right]$$

$$+ \operatorname{cov}\left[\underline{\Phi}(t+1, t)\underline{x}(t), \underline{v}(t)\right]$$

$$+ \operatorname{cov}\left[\underline{\Delta}(t+1, t)\underline{u}(t), \underline{H}(t)\underline{\tilde{x}}(t|t-1)\right]$$

$$+ \operatorname{cov}\left[\underline{\Delta}(t+1, t)\underline{u}(t), \underline{v}(t)\right]$$

which reduces to

$$\text{cov}\left[\underline{x}(t+1),\ \underline{\tilde{z}}(t|t-1)\right]$$
$$= \text{cov}\left[\underline{\Phi}(t+1,\ t)\underline{x}(t),\ \underline{H}(t)\underline{\tilde{x}}(t|t-1)\right]$$
$$= E\left\{\underline{\Phi}(t+1,\ t)\underline{x}(t)\underline{\tilde{x}}'(t|t-1)\underline{H}'(t)\right\}$$
$$= \underline{\Phi}(t+1,\ t)E\left\{\underline{x}(t)\underline{\tilde{x}}'(t|t-1)\right\}\underline{H}'(t)$$

since

$$\text{cov}\left[\underline{x}(t),\ \underline{v}(t)\right] = \text{cov}\left[\underline{u}(t),\ \underline{\tilde{x}}(t|t-1)\right] = \text{cov}\left[\underline{u}(t),\ \underline{v}(t)\right] = \underline{0}$$

and since

$$E\left\{\underline{x}(t)\right\} = E\left\{\underline{\tilde{x}}(t|t-1)\right\} = \underline{0}.$$

Furthermore, in view of the fact that

$$\underline{x}(t) = \underline{\hat{x}}(t|t-1) + \underline{\tilde{x}}(t|t-1)$$

and the fact that

$$E\left\{\underline{\hat{x}}(t|t-1)\underline{\tilde{x}}'(t|t-1)\right\} = \underline{0}\ ,$$

the first factor of the equation can be written

$$\text{cov}\left[\underline{x}(t+1),\ \underline{\tilde{z}}(t|t-1)\right]$$
$$= \underline{\Phi}(t+1,\ t)E\left\{\underline{\tilde{x}}(t|t-1)\underline{\tilde{x}}'(t|t-1)\right\}\underline{H}'(t)$$
$$= \underline{\Phi}(t+1,\ t)\underline{P}(t|t-1)\underline{H}'(t)\ . \tag{37}$$

The second factor on the right-hand side of equation (36) can be reduced in a similar manner to

$$\mathrm{cov}\left[\underline{\tilde{z}}(t|t-1),\ \underline{\tilde{z}}(t|t-1)\right]$$

$$= \mathrm{cov}\left[\underline{H}(t)\underline{\tilde{x}}(t|t-1),\ \underline{H}(t)\underline{\tilde{x}}(t|t-1)\right]$$

$$+ \mathrm{cov}\left[\underline{H}(t)\underline{\tilde{x}}(t|t-1)\ ,\ \underline{v}(t)\right]$$

$$\cdot \mathrm{cov}\left[\underline{v}(t),\ \underline{H}(t)\underline{\tilde{x}}(t|t-1)\right] + \mathrm{cov}\left[\underline{v}(t),\ \underline{v}(t)\right]$$

$$= \underline{H}(t)\mathrm{cov}\left[\underline{\tilde{x}}(t|t-1),\ \underline{\tilde{x}}(t|t-1)\right]\underline{H}'(t) + \mathrm{cov}\left[\underline{v}(t),\ \underline{v}(t)\right]$$

$$= \underline{H}(t)\underline{P}(t|t-1)\underline{H}'(t) + \underline{R}(t)\ . \tag{38}$$

Substituting equations (37) and (38) into equation (36) yields

$$E\left\{\underline{x}(t+1)\middle|\underline{\tilde{z}}(t|t-1)\right\}$$

$$= \left[\underline{\Phi}(t+1,\ t)\underline{P}(t|t-1)\underline{H}'(t)\right]$$

$$\cdot\left[\underline{H}(t)\underline{P}(t|t-1)\underline{H}'(t) + \underline{R}(t)\right]^{-1}$$

$$\cdot\left[\underline{z}(t) - \underline{H}(t)\underline{\hat{x}}(t|t-1)\right]\ . \tag{39}$$

Combining equations (35) and (39) yields

$$\underline{\hat{x}}(t+1|t) = \underline{\Phi}(t+1,\ t)\underline{\hat{x}}(t|t-1)$$

$$+ \left[\underline{\Phi}(t+1,\ t)\underline{P}(t|t-1)\underline{H}'(t)\right]$$

$$\cdot\left[\underline{H}(t)\underline{P}(t|t-1)\underline{H}'(t) + \underline{R}(t)\right|^{-1}$$

$$\cdot\left[\underline{z}(t) - \underline{H}(t)\underline{\hat{x}}(t|t-1)\right]$$

$$= \underline{\Psi}(t+1,\ t)\underline{\hat{x}}(t|t-1) + \underline{K}(t)\underline{z}(t) \tag{Id}$$

where

$$\underline{\Psi}(t+1,\ t) = \underline{\Phi}(t+1,\ t) - \underline{K}(t)\underline{H}(t)$$

and

$$\underline{K}(t) = \left[\underline{\Phi}(t+1,\ t)\underline{P}(t|t-1)\underline{H}'(t)\right] \cdot \left[\underline{H}(t)\underline{P}(t|t-1)\underline{H}'(t) + \underline{R}(t)\right]^{-1} \qquad \text{(IIId)}$$

From equation (32) one can obtain the equation

$$\underline{\tilde{x}}(t+1|t) = \underline{x}(t+1) - \underline{\hat{x}}(t+1|t) \quad . \qquad (40)$$

Substituting equations (1) and (Id) into equation (40) yields

$$\begin{aligned}
\underline{\tilde{x}}(t+1|t) &= \underline{\Phi}(t+1,\ t)\underline{x}(t) + \underline{\Delta}(t+1),\ t)\underline{u}(t) \\
&\quad - \underline{\Psi}(t+1,\ t)\underline{\hat{x}}(t|t-1) - \underline{K}(t)\underline{z}(t) \\
&= \underline{\Psi}(t+1,\ t)\underline{\tilde{x}}(t|t-1) \\
&\quad + \underline{\Delta}(t+1,\ t)\underline{u}(t) - \underline{K}(t)\underline{v}(t) \quad . \qquad \text{(IId)}
\end{aligned}$$

The solution of the filtering problem can be computed by deriving a recursion relation for the conditional covariance matrix $\underline{P}(t|t-1)$, which is the only remaining unknown in equations (Id) and (IIId).

$$\begin{aligned}
\underline{P}(t+1|t) &= \operatorname{cov}\left[\underline{\tilde{x}}(t+1|t),\ \underline{\tilde{x}}(t+1|t)\right] \\
&= \operatorname{cov}\Big[\underline{\Psi}(t+1,\ t)\underline{\tilde{x}}(t|t-1) \\
&\quad + \underline{\Delta}(t+1,\ t)\underline{u}(t) - \underline{K}(t)\underline{v}(t), \\
&\quad \underline{\Psi}(t+1,\ t)\underline{\tilde{x}}(t|t-1) + \underline{\Delta}(t+1,\ t)\underline{u}(t) \\
&\qquad - \underline{K}(t)\underline{v}(t)\Big] \\
&= \operatorname{cov}\Big[\underline{\Psi}(t+1,\ t)\underline{\tilde{x}}(t|t-1), \\
&\quad \underline{\Psi}(t+1,\ t)\underline{\tilde{x}}(t|t-1)\Big]
\end{aligned}$$

$$+ \operatorname{cov}\left[\underline{K}(t)\underline{v}(t),\ \underline{K}(t)\underline{v}(t)\right]$$
$$+ \operatorname{cov}\left[\underline{\Delta}(t+1,\ t)\underline{u}(t),\ \underline{\Delta}(t+1,\ t)\underline{u}(t)\right]$$

$$= \underline{\Psi}(t+1,\ t)\underline{P}(t\ t-1)\underline{\Psi}'(t+1,\ t)$$
$$+ \underline{K}(t)\underline{R}(t)\underline{K}'(t) + \underline{\Delta}(t+1,\ t)\underline{Q}(t)\underline{\Delta}'(t+1,\ t) \quad (41)$$

The first two terms of equation (41) are

$$\underline{\Psi}(t+1,\ t)\underline{P}(t|t-1)\underline{\Psi}'(t+1,\ t) + \underline{K}(t)\underline{R}(t)\underline{K}'(t)$$

$$= \left[\underline{\Phi}(t+1,\ t) - \underline{K}(t)\underline{H}(t)\right]\underline{P}(t|t-1)$$
$$\left[\underline{\Phi}'(t+1,\ t) - \underline{H}'(t)\underline{K}'(t)\right] + \underline{K}(t)\underline{R}(t)\underline{K}'(t)$$

$$= \underline{\Phi}(t+1,\ t)\underline{P}(t|t-1)\underline{\Phi}'(t+1,\ t)$$
$$-\underline{K}(t)\underline{H}(t)\underline{P}(t|t+1)\underline{\Phi}'(t+1,\ t)$$
$$- \underline{\Phi}(t+1,\ t)\underline{P}(t|t-1)\underline{H}'(t)\underline{K}'(t)$$
$$+ \underline{K}(t)\underline{H}(t)\underline{P}(t|t+1)\underline{H}'(t)\underline{K}'(t) + \underline{K}(t)\underline{R}(t)\underline{K}'(t)$$

$$= \underline{\Phi}(t+1,\ t)\underline{P}(t|t-1)\underline{\Phi}'(t+1,\ t)$$
$$- \underline{K}(t)\underline{H}(t)\underline{P}(t|t-1)\underline{\Phi}'(t+1,\ t)$$
$$- \underline{\Phi}(t+1,\ t)\underline{P}(t|t-1)\underline{H}'(t)\underline{K}'(t)$$
$$+ \underline{K}(t)\left[\underline{H}(t)\underline{P}(t|t-1)\underline{H}'(t) + \underline{R}(t)\right]\underline{K}'(t) \quad (42)$$

Substituting equation (IIId) into the last two terms of equation (42),

$$\underline{K}(t)\left[\underline{H}(t)\underline{P}(t|t-1)\underline{H}'(t) + \underline{R}(t)\right]\underline{K}'(t)$$
$$- \underline{\Phi}(t+1,\ t)\underline{P}(t|t-1)\underline{H}'(t)\underline{K}'(t)$$

$$= \left[\underline{\Phi}(t+1,\ t)\underline{P}(t|t-1)\underline{H}'(t)\right]\left[\underline{H}(t)\underline{P}(t|t-1)\underline{H}'(t) + \underline{R}(t)\right]^{-1}$$
$$\left[\underline{H}(t)\underline{P}(t|t-1)\underline{H}'(t) + \underline{R}(t)\right]\underline{K}'(t)$$
$$- \underline{\Phi}(t+1,\ t)\underline{P}(t|t-1)\underline{H}'(t)\underline{K}'(t)$$
$$= \underline{0}\ .$$

Then the first two terms of equation (41) are

$$\underline{\Psi}(t+1, t)\underline{P}(t|t-1)\underline{\Psi}'(t+1, t) + \underline{K}(t)\underline{R}(t)\underline{K}'(t)$$

$$= \underline{\Phi}(t+1, t)\underline{P}(t|t+1)\underline{\Phi}'(t+1, t) - \underline{K}(t)\underline{H}(t)\underline{P}(t|t-1)\underline{\Phi}'(t+1, t)$$

$$= \underline{\Phi}(t+1, t)\Big\{\underline{P}(t|t-1) - \left[\underline{P}(t|t-1)\underline{H}'(t)\right]\left[\underline{H}(t)\underline{P}(t|t-1)\underline{H}'(t) + \underline{R}(t)\right]^{-1} \cdot \underline{H}(t)\underline{P}(t|t-1)\Big\} \cdot \underline{\Phi}'(t+1, t) \qquad (43)$$

Hence equation (41) becomes

$$\underline{P}(t+1|t) = \underline{\Phi}(t+1, t)\Big\{\underline{P}(t|t-1) - \left[\underline{P}(t|t-1)\underline{H}'(t)\right] \cdot \left[\underline{H}(t)\underline{P}(t|t-1)\underline{H}'(t) + \underline{R}(t)\right]^{-1} \underline{H}(t)\underline{P}(t|t-1)\Big\} \cdot \underline{\Phi}'(t+1, t) + \underline{\Delta}(t+1, t)\underline{Q}(t)\underline{\Delta}'(t+1, t) \qquad \text{(IVd)}$$

Equations (Id) to (Vd) are the solutions for a Gauss-Markov sequence. They will be discussed later.

In a Gauss-Markov process the message is a random process generated by the equation

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

and the observed value is

$$\underline{z}(t) = \underline{y}(t) + \underline{v}(t) = \underline{H}(t)\underline{x}(t) + \underline{v}(t).$$

The solution of this problem can be directly derived from the Wiener-Hopf equation. Pugachev pointed out that the Wiener-Hopf equation is nothing more than a special case of the orthogonal-projection theorem (Ref. 20).

The orthogonal-projection lemma can be stated as follows: A necessary and sufficient condition for

$$\|\underline{x} - \underline{w}\| \geqslant \|\underline{x} - \underline{w}_0\| \qquad \text{for all } \underline{w} \text{ in } \underline{W} \tag{44}$$

is that

$$(\underline{x} - \underline{w}_0, \underline{w}) = 0 \qquad \text{for all } \underline{w} \text{ in } \underline{W}\,. \tag{45}$$

Moreover, if there is another vector $\underline{w}_1$ satisfying

$$(\underline{x} - \underline{w}_1, \underline{w}) = 0$$

then

$$\|\underline{w}_0 - \underline{w}_1\| = 0$$

where $\underline{x}$ is an element of abstract space $\underline{\mathcal{X}}$

$\underline{W}$ is a subspace of $\underline{\mathcal{X}}$

$\underline{w}$ is an element of $\underline{W}$.

Proof:

$$\|\underline{x} - \underline{w}\|^2 = \|\underline{x} - \underline{w}_0\|^2 + z(\underline{x} - \underline{w}_0, \underline{w}_0 - \underline{w}) + \|\underline{w} - \underline{w}_0\|^2 \tag{46}$$

Since

$$(\underline{x} - \underline{w}_0, \underline{w}_0 - \underline{w}) = 0$$

and

$$\|\underline{w} - \underline{w}_0\| \geqslant 0$$

it follows that

$$\|\underline{x} - \underline{w}\| \geqslant \|\underline{x} - \underline{w}_0\|\,.$$

Substituting $\underline{w} = \underline{w}_1$ into equation (46),

$$\|\underline{x} - \underline{w}_1\|^2 = \|\underline{x} - \underline{w}_0\|^2 + \|\underline{w}_1 - \underline{w}_0\|^2 \tag{47}$$

Since $\underline{w}_1$ satisfies condition (45)

$$\left\|\underline{x} - \underline{w}_0\right\|^2 = \left\|\underline{x} - \underline{w}_1\right\|^2 + 2(\underline{x} - \underline{w}_1, \underline{w}_1 - \underline{w}_0) + \left\|\underline{w}_0 - \underline{w}_1\right\|^2$$

$$= \left\|\underline{x} - \underline{w}_1\right\|^2 + \left\|\underline{w}_0 - \underline{w}_1\right\|^2 \qquad (48)$$

Combining equations (47) and (48) yields

$$\left\|\underline{x} - \underline{w}_0\right\|^2 = \left\|\underline{x} - \underline{w}_0\right\|^2 + \left\|\underline{w}_1 - \underline{w}_0\right\|^2 + \left\|\underline{w}_0 - \underline{w}_1\right\|^2$$

hence

$$\left\|\underline{w}_0 - \underline{w}_1\right\| = 0 \ .$$

Consider a vector $\underline{w}_2$ such that

$$(\underline{x} - \underline{w}_0, \underline{w}_2) = \alpha \neq 0 \ .$$

Then

$$\left\|\underline{x} - \underline{w}_0 - \beta\underline{w}_2\right\| = \left\|\underline{x} - \underline{w}_0\right\|^2 - 2\alpha\beta + \beta^2 \left\|\underline{w}_2\right\|^2$$

where α is given, but the value could vary with β. For suitable choice of β, the last two terms could be negative; then

$$\left\|\underline{x} - \underline{w}_0 - \beta\underline{w}_2\right\|^2 \leqslant \left\|\underline{x} - \underline{w}_0\right\|^2 \ .$$

This inequality contradicts the optimality of $\underline{w}_0$.

The optimal estimation problem for the multidimensional situation is one in which the values of $\underline{z}(\tau)$, $t_0 \leqslant \tau \leqslant t$ are given, and one wishes to find an estimate $\hat{\underline{x}}(t_1|t)$ of $\underline{x}(t_1)$ having the form

$$\hat{\underline{x}}(t_1|t) = \int_{t_0}^{t} \underline{A}(t_1, \tau)\underline{z}(\tau)d\tau \qquad (49)$$

which minimizes the expected squared error sum,

$$E\left\{\sum_{i=1}^{n}\left[x_i(t) - \hat{x}_i(t)\right]^2\right\}.$$

$\underline{A}(t_1, \tau)$ is an n x p matrix whose elements are continuously differentiable in both arguments. $\underline{\hat{x}}(t_1| t)$ is an unbiased minimum variance linear estimate of $\underline{x}(t_1)$. That is,

$$E\left\{\underline{x}(t_1)\right\} = E\left\{\underline{\hat{x}}(t_1|t)\right\} = \underline{0}$$

A necessary and sufficient condition that $\underline{\hat{x}}(t_1|t)$ be a minimum variance estimator for $\underline{x}(t_1)$ is that the matrix function $\underline{A}(t_1, \tau)$ satisfy the matrix form of the Wiener-Hopf equation (Ref. 17).

$$\operatorname{cov}\left[\underline{x}(t_1), \underline{z}(\sigma)\right] - \int_{t_0}^{t} \underline{A}(t_1, \tau)\operatorname{cov}\left[\underline{z}(\tau), \underline{z}(\sigma)\right]d\tau = 0 \qquad (50)$$

$$\text{for } t_0 \leqslant \sigma < t .$$

Since

$$\operatorname{cov}\left[\underline{x}(t_1), \underline{z}(\sigma)\right] = \operatorname{cov}\left[\underline{\hat{x}}(t_1|t) + \underline{\tilde{x}}(t_1 | t), \underline{z}(\sigma)\right]$$

$$= \operatorname{cov}\left[\underline{\hat{x}}(t_1 | t), \underline{z}(\sigma)\right] + \operatorname{cov}\left[\underline{\tilde{x}}(t_1 | t), \underline{z}(\sigma)\right]$$

$$= E\left\{\int_{t_0}^{t} \underline{A}(t_1, \tau)\underline{z}(\tau)\underline{z}'(\sigma)\,d\tau\right\} + \operatorname{cov}\left[\underline{\tilde{x}}(t_1 | t), \underline{z}(\sigma)\right]$$

and assuming that the interchange of expectation and integration is a valid operation, yields

$$\operatorname{cov}\left[\underline{x}(t_1), \underline{z}(\sigma)\right]$$

$$= \int_{t_0}^{t} \underline{A}(t_1, \tau)\, E\left\{\underline{z}(\tau)\underline{z}'(\sigma)\right\} d\tau$$

$$+ \operatorname{cov}\left[\underline{\tilde{x}}(t_1 | t), \underline{z}(\sigma)\right].$$

$$= \int_{t_0}^{t} \underline{A}(t_1, \tau)\,\mathrm{cov}\left[\underline{z}(\tau), \underline{z}(\sigma)\right] d\tau$$

$$+ \mathrm{cov}\left[\underline{\tilde{x}}(t_1 \mid t), \underline{z}(\sigma)\right]$$

Hence the matrix form of the Wiener-Hopf equation reduces to

$$\mathrm{cov}\left[\underline{\tilde{x}}(t_1 \mid t), \underline{z}(\sigma)\right] = 0 \qquad \text{for } t_0 \leq \sigma < t \tag{51}$$

This is equivalent to saying

$$\underline{\hat{x}}(t_1 \mid t) = \int_{t_0}^{t} \underline{A}(t_1, \tau)\underline{z}(\tau)\,d\tau$$

will be optimal estimator for $\underline{x}(t_1)$ if and only if $\underline{A}(t_1, \tau)$ satisfies (51).

Proof:

Let $\mathcal{X}$ denote the space generated by the random vectors $\underline{x}(t_1)$. The subspace $\underline{W}$ is generated by

$$\underline{w}(t_1) = \int_{t_0}^{t} \underline{B}(t_1, \tau)\underline{z}(\tau)\,d\tau \tag{52}$$

where $\underline{B}(t, \tau)$ is an m x p matrix whose elements are continuously differentiable in both arguments. Let

$$\underline{w}_0 = \underline{\hat{x}}(t_1 \mid t)$$

The orthogonal-projection lemma implies that

$$(\underline{x} - \underline{w}_0, \underline{w}) = 0$$

$$(\underline{x} - \underline{w}_0, \underline{w}) = E\left\{\left[\underline{x} - \underline{w}_0\right]\underline{w}'\right\}$$

$$= E\left\{\left[\underline{x}(t_1) - \underline{\hat{x}}(t_1 \mid t)\right]\underline{w}'(t_1)\right\}$$

$$= E\left\{ \int_{t_0}^{t} \underline{\tilde{x}}(t_1 \mid t)\underline{z}'(\sigma)\underline{B}'(t,\sigma)d\sigma \right\}$$

$$= \int_{t_0}^{t} E\left\{\underline{\tilde{x}}(t_1 \mid t)\underline{z}'(\sigma)\right\} \underline{B}'(t_1,\sigma)d\sigma$$

$$= \int_{t_0}^{t} \mathrm{cov}\left[\underline{\tilde{x}}(t_1 \mid t),\ \underline{z}(\sigma)\right]\underline{B}'(t_1,\sigma)d\sigma$$

$$= \underline{0}$$

Then $\mathrm{cov}\left[\underline{\tilde{x}}(t_1 \mid t),\ \underline{z}(\sigma)\right] = \underline{0}$ establishes the sufficiency of the condition, because $\underline{B}$ is an arbitrarily selected matrix subject only to the differentiability conditions. Let

$$\underline{B}(t_1,\sigma) = \mathrm{cov}\left[\underline{\tilde{x}}(t_1 \mid t),\ \underline{z}(\sigma)\right] ,$$

then the matrix $\underline{B}\ \underline{B}'$ is nonnegative definite. Then the integral is zero only if $\underline{B}(t,\sigma)$ vanishes identically for all $t_0 \leq \sigma < t$. Therefore the necessity condition is verified.

A canonical form for the optimal filter can be derived from the Wiener-Hopf equation. Let $t_1 = t$, then the Wiener-Hopf equation becomes

$$\mathrm{cov}\left[\underline{x}(t),\ \underline{z}(\sigma)\right] = \int_{t_0}^{t} \underline{A}(t,\ \tau)\cdot\mathrm{cov}\left[\underline{z}(\tau),\ \underline{z}(\sigma)\right]d\tau \qquad (53)$$

Differentiating both sides with respect to t and interchanging the order of operations of differentiation and expectation, then the left-hand side is

$$\frac{\partial}{\partial t}\,\mathrm{cov}\left[\underline{x}(t),\ \underline{z}(\sigma)\right] = E\left\{ \frac{\partial}{\partial t}\left[\underline{x}(t)\underline{z}'(\sigma)\right]\right.$$

$$= E\left\{ \left[\underline{F}(t)\underline{x}(t) + \underline{G}(t)\underline{u}(t)\right]\underline{z}'(\sigma)\right\}$$

$$= \underline{F}(t)E\left\{\underline{x}(t)\cdot\underline{z}'(\sigma)\right\} + \underline{G}(t)E\left\{\underline{u}(t)\underline{z}'(\sigma)\right\}$$

By assumption, $\underline{u}(t)$ is independent of $\underline{v}(\sigma)$ and $\underline{x}(\sigma)$ when $\sigma < t$. Hence

$$E\left\{\underline{u}(t)\underline{z}'(\sigma)\right\} = \underline{0}\ .$$

The matrix form of the Wiener-Hopf equation yields

$$E\left\{\underline{x}(t)\underline{z}'(\sigma)\right\} = \int_{t_0}^{t} \underline{A}(t,\ \tau)\operatorname{cov}\left[\underline{z}(\tau),\ \underline{z}(\sigma)\right]d\tau\ .$$

Hence

$$\frac{\partial}{\partial t}\operatorname{cov}\left[\underline{x}(t),\ \underline{z}(\sigma)\right]$$
$$= \int_{t_0}^{t} \underline{F}(t)\underline{A}(t,\ \tau)\operatorname{cov}\left[\underline{z}(\tau),\ \underline{z}(\sigma)\right]d\tau \qquad (54)$$

The derivative of the right side of equation (53) becomes

$$\frac{\partial}{\partial t}\int_{t_0}^{t} \underline{A}(t,\ \tau)\operatorname{cov}\left[\underline{z}(\tau),\ \underline{z}(\sigma)\right]d\tau$$
$$= \int_{t_0}^{t} \frac{\partial}{\partial t}\underline{A}(t,\ \tau)\operatorname{cov}\left[\underline{z}(\tau),\ \underline{z}(\sigma)\right]d\tau$$
$$+ \underline{A}(t,\ t)\operatorname{cov}\left[\underline{z}(t),\ \underline{z}(\sigma)\right] \qquad (55)$$

since

$$\operatorname{cov}\left[\underline{z}(t),\ \underline{z}(\sigma)\right]$$
$$= \operatorname{cov}\left[\underline{H}(t)\underline{x}(t) + \underline{v}(t),\ \underline{z}(\sigma)\right]$$
$$= \underline{H}(t)\operatorname{cov}\left[\underline{x}(t),\ \underline{z}(\sigma)\right] + \operatorname{cov}\left[\underline{v}(t),\ \underline{z}(\sigma)\right]$$
$$= \int_{t_0}^{t} \underline{H}(t)\underline{A}(t,\ \tau)\operatorname{cov}\left[\underline{z}(\tau),\ \underline{z}(\sigma)\right]d\tau\ . \qquad (56)$$

Combining equations (55) and (56), the derivative of the right side of equation (53) is

$$\frac{\partial}{\partial t}\int_{t_0}^{t} \underline{A}(t, \tau)\operatorname{cov}\left[\underline{z}(\tau), \underline{z}(\sigma)\right] d\tau$$

$$= \int_{t_0}^{t} \left[\frac{\partial}{\partial t}\underline{A}(t, \tau) + \underline{A}(t, t)\underline{H}(t)\underline{A}(t, \tau)\right] \operatorname{cov}\left[\underline{z}(\tau), \underline{z}(\sigma)\right] d\tau \tag{57}$$

Combining the results of equation (54) and equation (57) yields

$$\int_{t_0}^{t} \left[\underline{F}(t)\underline{A}(t, \tau) - \frac{\partial}{\partial t}\underline{A}(t, \tau) - \underline{A}(t, t)\underline{H}(t)\underline{A}(t, \tau)\right]\operatorname{cov}\left[\underline{z}(\tau), \underline{z}(\sigma)\right] d\tau = \underline{0} \quad \text{for } t_0 \leq \sigma < t \tag{58}$$

If the optimal matrix $\underline{A}$ is a solution of the differential equation,

$$\underline{F}(t)\underline{A}(t, \tau) - \frac{\partial}{\partial t}\underline{A}(t, \tau) - \underline{A}(t, t)\underline{H}(t)\underline{A}(t, \tau) = \underline{0} \quad \text{for } t_0 \leq \tau \leq t \tag{59}$$

then equation (58) is certainly satisfied.

If $\underline{R}(\tau)$ is positive definite in the interval $t_0 \leq \tau \leq t$, then the covariance matrix $\operatorname{cov}\left[\underline{z}(\tau), \underline{z}(\sigma)\right]$ is also positive definite in this interval $t_0 \leq \tau \leq t$. Then the condition (59) is necessary.

The optimal filter is generated by differentiating

$$\hat{\underline{x}}(t \mid t) = \int_{t_0}^{t} \underline{A}(t, \tau)\underline{z}(\tau)d\tau$$

with respect to t and combining the result with equation (59). Thus

$$\frac{d\hat{\underline{x}}(t|t)}{dt} = \frac{\partial}{\partial t} \int_{t_0}^{t} \underline{A}(t, \tau)\underline{z}(\tau)d\tau$$

$$= \int_{t_0}^{t} \frac{\partial}{\partial t} \underline{A}(t, \tau)\underline{z}(\tau)d\tau + \underline{A}(t, t)\underline{z}(t)$$

$$= \int_{t_0}^{t} \left[\underline{F}(t)\underline{A}(t, \tau) - \underline{A}(t, t)\underline{H}(t)\underline{A}(t, \tau)\right]\underline{z}(\tau)d\tau + \underline{A}(t, t)\underline{z}(t) .$$

Let $\underline{A}(t, t) = \underline{K}(t)$, then

$$\frac{d\hat{\underline{x}}(t|t)}{dt} = \int_{t_0}^{t} \left[\underline{F}(t)\underline{A}(t, \tau) - \underline{K}(t)\underline{H}(t)\underline{A}(t, \tau)\ \underline{z}(\tau)\right]d\tau + \underline{K}(t)\underline{z}(t)$$

$$= \left[\underline{F}(t) - \underline{K}(t)\underline{H}(t)\right] \int_{t_0}^{t} \underline{A}(t, \tau)\underline{z}(\tau)d\tau + \underline{K}(t)\underline{z}(t)$$

$$= \left[\underline{F}(t) - \underline{K}(t)\underline{H}(t)\right]\hat{\underline{x}}(t \mid t) + \underline{K}(t)\underline{z}(t)$$

$$= \underline{F}(t)\hat{\underline{x}}(t \mid t) + \underline{K}(t)\left[\underline{z}(t) - \underline{H}(t)\hat{\underline{x}}(t \mid t)\right] \qquad \text{(Ic)}$$

Substituting $\hat{\underline{x}}(t \mid t) = \underline{x}(t) - \tilde{\underline{x}}(t \mid t)$ into equation (Ic)

$$\frac{d}{dt}\left[\underline{x}(t) - \tilde{\underline{x}}(t \mid t)\right] = \left[\underline{F}(t) - \underline{K}(t)\underline{H}(t)\right]\left[\underline{x}(t) - \tilde{\underline{x}}(t \mid t)\right] + \underline{K}(t)\underline{z}(t)$$

$$\frac{d\underline{x}(t)}{dx} - \frac{d\tilde{\underline{x}}(t \mid t)}{dx} = \left[\underline{F}(t) - \underline{K}(t)\underline{H}(t)\right]\left[\underline{x}(t) - \tilde{\underline{x}}(t \mid t)\right] + \underline{K}(t)\left[\underline{H}(t)\underline{x}(t) + \underline{v}(t)\right]$$

Then

$$\frac{d\underline{\tilde{x}}(t|t)}{dt} = \left[\underline{F}(t) - \underline{K}(t)\underline{H}(t)\right]\underline{\tilde{x}}(t \mid t) + \underline{G}(t)\underline{u}(t) - \underline{K}(t)\underline{v}(t) \quad \text{(IIc)}$$

The next step is to derive an explicit form of the optimal gain (or optimal weighting matrix) $\underline{K}(t)$. It is obtained from the Wiener-Hopf equation

$$\text{cov}\left[\underline{x}(t),\ \underline{z}(\sigma)\right] - \int_{t_0}^{t} \underline{A}(t,\ \tau)\,\text{cov}\left[\underline{z}(\tau),\ \underline{z}(\sigma)\right] d\tau = \underline{0}$$

$$\text{cov}\left[\underline{x}(t),\ \underline{z}(\sigma)\right] - \int_{t_0}^{t} \underline{A}(t,\ \tau)\,\text{cov}\left[\underline{z}(\tau),\ \underline{z}(\sigma)\right] d\tau$$

$$= \text{cov}\left[\underline{x}(t),\ \underline{y}(\sigma) + \underline{v}(\sigma)\right] - \int_{t_0}^{t} \underline{A}(t,\ \tau)\,\text{cov}\left[\underline{y}(\tau) + \underline{v}(\tau),\ \underline{y}(\sigma) + \underline{v}(\sigma)\right] d\tau$$

$$= \text{cov}\left[\underline{x}(t),\ \underline{y}(\sigma)\right] - \int_{t_0}^{t} \underline{A}(t,\ \tau)\left\{\text{cov}\left[\underline{y}(\tau),\ \underline{y}(\sigma)\right] + \underline{R}(\tau)\delta(\tau - \sigma)\right\} d\tau$$

$$= \text{cov}\left[\underline{x}(t),\ \underline{y}(\sigma)\right] - \int_{t_1}^{t} \underline{A}(t,\ \tau)\,\text{cov}\left[\underline{y}(\tau),\ \underline{y}(\sigma)\right] d\tau - \underline{A}(t,\sigma)\underline{R}(\sigma)$$

$$= \underline{0}$$

$$t_0 \leq \sigma < t$$

Hence

$$\text{cov}\left[\underline{x}(t),\ \underline{y}(\sigma)\right] - \int_{t_0}^{t} \underline{A}(t,\ \tau)\,\text{cov}\left[\underline{y}(\tau),\ \underline{y}(\sigma)\right] d\tau = \underline{A}(t,\sigma)\underline{R}(\sigma) \quad (60)$$

Both sides of equation (60) are continuous functions of σ. Therefore that equality holds also at $\sigma = t$. Hence

$$\underline{A}(t, t)\underline{R}(t) = \text{cov}\left[\hat{\underline{x}}(t \mid t) + \tilde{\underline{x}}(t \mid t), \underline{y}(t)\right]$$
$$- \int_{t_0}^{t} \underline{A}(t, \tau)\text{cov}\left[\underline{y}(\tau), \underline{y}(t)\right] d\tau$$
$$= \text{cov}\left[\hat{\underline{x}}(t \mid t), \underline{y}(t)\right] + \text{cov}\left[\tilde{\underline{x}}(t \mid t), \underline{y}(t)\right]$$
$$- \int_{t_0}^{t} \underline{A}(t, \tau)\text{cov}\left[\underline{y}(\tau), \underline{y}(t)\right] d\tau$$

since

$$\text{cov}\left[\hat{\underline{x}}(t \mid t), \underline{y}(t)\right] = E\left\{ \int_{t_0}^{t} \underline{A}(t, \tau)\underline{z}(\tau)\underline{y}'(t) d\tau \right\}$$
$$= \int_{t_0}^{t} \underline{A}(t, \tau)E\left\{\underline{z}(\tau)\underline{y}'(t)\right\} d\tau$$
$$= \int_{t_0}^{t} \underline{A}(t, \tau)\text{cov}\left[\underline{y}(\tau), \underline{y}(t)\right] d\tau \; .$$

Therefore

$$\underline{K}(t)\underline{R}(t) = \underline{A}(t, t)\underline{R}(t) = \text{cov}\left[\tilde{\underline{x}}(t \mid t), \underline{y}(t)\right] \tag{61}$$

since

$$\text{cov}\left[\tilde{\underline{x}}(t \mid t), \underline{y}(t)\right]$$
$$= \text{cov}\left[\tilde{\underline{x}}(t \mid t), \underline{H}(t)\left(\tilde{\underline{x}}(t \mid t) + \hat{\underline{x}}(t \mid t)\right)\right]$$
$$= \text{cov}\left[\tilde{\underline{x}}(t \mid t), \tilde{\underline{x}}(t \mid t)\right]\underline{H}'(t) + \text{cov}\left[\tilde{\underline{x}}(t \mid t), \hat{\underline{x}}(t \mid t)\right]H'(t)$$
$$= \underline{P}(t)\underline{H}'(t) \; .$$

And $\underline{R}(t)$ is assumed positive definite, so that $\underline{R}^{-1}(t)$ exists. Multiplying both sides of equation (61) by $\underline{R}^{-1}(t)$, yields

$$\underline{K}(t) = \underline{P}(t)\underline{H}'(t)\underline{R}^{-1}(t) \tag{IIId}$$

The general solution of equation (IIc) is

$$\underline{\tilde{x}}(t \mid t) = \underline{\Psi}_c(t, t_0)\underline{\tilde{x}}(t_1 \mid t_0)$$

$$+ \int_{t_0}^{t} \underline{\Psi}_c(T, \tau)\left[\underline{G}(\tau)\underline{u}(\tau) - \underline{K}(\tau)\underline{v}(\tau)\right]d\tau \qquad (62)$$

where $\underline{\Psi}_c(t, \tau)$ is the common transition matrix of (Ic) and (IIc). Then one can derive the variance equation

$$\underline{P}(t) = \operatorname{cov}\left[\underline{\tilde{x}}(t \mid t), \underline{\tilde{x}}(t \mid t)\right]$$

Since $\underline{x}(t_0 \mid t_0)$ is independent of $\underline{u}(\tau)$ and $\underline{v}(\tau)$ as $t_0 \leq \tau \leq t$, this becomes

$$\underline{P}(t) = \underline{\Psi}_c(t, t_0)\operatorname{cov}\left[\underline{\tilde{x}}(t_0 \mid t_0), \underline{\tilde{x}}(t_0 \mid t_0)\right]\underline{\Psi}_c'(t, t_0)$$

$$+ E\left\{\int_{t_0}^{t} \underline{\Psi}_c(t, \tau)\left[\underline{G}(\tau)\underline{u}(\tau) - \underline{K}(\tau)\underline{v}(\tau)\right]d\tau\right.$$

$$\left.\cdot \int_{t_0}^{t} \left[\underline{u}'(\tau')\underline{G}'(\tau') - \underline{v}'(\tau')\underline{K}'(\tau')\right]\underline{\Psi}_c'(t, \tau')d\tau'\right\}$$

Hence

$$\underline{P}(t) - \underline{\Psi}_c(t, t_0)\underline{P}(t_0)\underline{\Psi}_c'(t, t_0)$$

$$= \int_{t_0}^{t} \underline{\Psi}_c(t, \tau)\left[\int_{t_0}^{t} \underline{G}(\tau)E\left\{\underline{u}(\tau)\underline{u}'(\tau')\right\}\underline{G}'(\tau')\underline{\Psi}_c'(t, \tau')d\tau'\right.$$

$$\left.+ \int_{t_0}^{t} \underline{K}(\tau)E\left\{\underline{v}(\tau)\underline{v}'(\tau')\right\}\underline{K}'(\tau)\underline{\Psi}_c'(t, \tau')d\tau'\right]d\tau$$

$$= \int_{t_0}^{t} \underline{\Psi}_c(t, \tau)\left[\int_{t_0}^{t} \underline{G}(\tau)\underline{Q}(\tau)\delta(\tau - \tau')\underline{G}'(\tau')\underline{\Psi}_c'(t, \tau')d\tau'\right.$$

$$\left.+ \int_{t_0}^{t} \underline{K}(\tau)\underline{R}(\tau)\delta(\tau - \tau')\underline{K}'(\tau') \cdot \underline{\Psi}_c'(t, \tau')d\tau'\right]d\tau$$

$$= \int_{t_0}^{t} \underline{\Psi}_c(t, \tau)\left[\underline{G}(\tau)\underline{Q}(\tau)\underline{G}'(\tau) + \underline{K}(\tau)\underline{R}(\tau)\underline{K}'(\tau)\right]\underline{\Psi}_c'(t, \tau)\,d\tau \tag{63}$$

Differentiating with respect to t, the left side of equation (63) is

$$\frac{d}{dt}\left[\underline{P}(t) - \underline{\Psi}_c(t, t_0)\underline{P}(t_0)\underline{\Psi}_c'(t, t_0)\right]$$

$$= \frac{d\underline{p}}{dt} - \frac{d}{dt}\left[\underline{\Psi}_c(t, t_0)\underline{P}(t_0)\underline{\Psi}_c'(t, t_0)\right]$$

This equation becomes

$$\frac{d}{dt}\left[\underline{P}(t) - \underline{\Psi}_c(t, t_0)\underline{P}(t_0)\underline{\Psi}_c'(t, t_0)\right]$$

$$= \frac{d\underline{p}}{dt} - \left[\underline{F}(t) - \underline{K}(t)\underline{H}(t)\right]\underline{\Psi}_c(t, t_0)\underline{P}(t_0)\underline{\Psi}_c'(t, t_0)$$
$$- \underline{\Psi}_c(t, t_0)\underline{P}(t_0)\underline{\Psi}_c'(t, t_0)\left[\underline{F}'(t) - \underline{H}'(t)\underline{K}'(t)\right] \tag{64}$$

since

$$\frac{d}{dt}\underline{\Psi}_c(t, t_0) = \left[\underline{F}(t) - \underline{K}(t)\underline{H}(t)\right]\underline{\Psi}_c(t, t_0)$$

$$\frac{d}{dt}\underline{\Psi}_c'(t, t_0) = \underline{\Psi}_c'(t, t_0)\left[\underline{F}'(t) - \underline{H}'(t)\underline{K}'(t)\right]$$

The derivative of the right side of equation (63) is

$$= \frac{d}{dt}\int_{t_0}^{t} \underline{\Psi}_c(t, \tau)\left[\underline{G}(\tau)\underline{Q}(\tau)\underline{G}'(\tau) + \underline{K}(\tau)\underline{R}(\tau)\underline{K}'(\tau)\right]\underline{\Psi}_c'(t, \tau)\,d\tau$$

$$= \int_{t_0}^{t} \frac{\partial}{\partial t}\underline{\Psi}_c(t, \tau)\left[\underline{G}(\tau)\underline{Q}(\tau)\underline{G}'(\tau) + \underline{K}(\tau)\underline{R}(\tau)\underline{K}'(\tau)\right]\underline{\Psi}_c'(t, \tau)\,d\tau.$$

$$+ \underline{\Psi}_c(t, t)\left[\underline{G}(t)\underline{Q}(t)\underline{G}'(t) + \underline{K}(t)\underline{R}(t)\underline{K}'(t)\right]\underline{\Psi}_c'(t, t)$$

$$= \left[\underline{F}(t) - \underline{K}(t)\underline{H}(t)\right] \int_{t_0}^{t} \underline{\Psi}_c(t, \tau)\left[\underline{G}(\tau)\underline{Q}(\tau)\underline{G}'(\tau) + \underline{K}(\tau)\underline{R}(\tau)\underline{K}'(\tau)\right]\underline{\Psi}_c'(t, \tau)d\tau$$

$$+ \int_{t_0}^{t} \underline{\Psi}_c(t, \tau)\left[\underline{G}(\tau)\underline{Q}(\tau)\underline{G}'(\tau) + \underline{K}(\tau)\underline{R}(\tau)\underline{K}'(\tau)\right] \underline{\Psi}_c'(t, \tau)d\tau\left[\underline{F}'(t) - \underline{H}'(t)\underline{K}'(t)\right]$$

$$+ \left[\underline{G}(t)\underline{Q}(t)\underline{G}'(t) + \underline{K}(t)\underline{R}(t)\underline{K}'(t)\right]$$

$$= \left[\underline{F}(t) - \underline{K}(t)\underline{H}(t)\right]\left[\underline{P}(t) - \underline{\Psi}_c(t, t_0)\underline{P}(t_0)\underline{\Psi}_c'(t, t_0)\right]$$

$$+ \left[\underline{P}(t) - \underline{\Psi}_c(t, t_0)\underline{P}(t_0)\underline{\Psi}_c'(t, t_0)\right] \cdot \left[\underline{F}'(t) - \underline{H}'(t)\underline{K}'(t)\right]$$

$$+ \left[\underline{G}(t)\underline{Q}(t)\underline{G}'(t) + \underline{K}(t)\underline{R}(t)\underline{K}'(t)\right] \tag{65}$$

Combining equations (64) and (65) and solving the result for $d\underline{p}/dt$ yields

$$\frac{d}{dt}\underline{P}(t) = \left[\underline{F}(t) - \underline{K}(t)\underline{H}(t)\right]\underline{P}(t) + \left[\underline{P}(t)\right]\cdot\left[\underline{F}'(t) - \underline{H}'(t)\underline{K}'(t)\right] + \left[\underline{G}(t)\underline{Q}(t)\underline{G}'(t) + \underline{K}(t)\underline{R}(t)\underline{K}'(t)\right]$$

Substituting $\underline{K}(t) = \underline{P}(t)\underline{H}'(t)\underline{R}^{-1}(t)$ into the above equation yields

$$\frac{d\underline{P}(t)}{dt} = \underline{F}(t)\underline{P}(t) + \underline{P}(t)\underline{F}'(t) - \underline{P}(t)\underline{H}'(t)\underline{R}^{-1}(t)\underline{H}(t)\underline{P}(t) + \underline{G}(t)\underline{Q}(t)\underline{G}'(t) \tag{IVc}$$

Now one can evaluate the following covariance matrix.

$$\operatorname{cov}\left[\underline{\tilde{x}}(t \mid t), \underline{u}(t)\right] = \operatorname{cov}\left[\underline{\Psi}_c(t, t_0)\underline{\tilde{x}}(t_0 \mid t_0)\right.$$

$$+ \int_{t_0}^{t} \underline{\Psi}_c(t, \tau)\left[\underline{G}(\tau)\underline{u}(\tau) - \underline{K}(\tau)\underline{v}(\tau)\right] d\tau,\ \underline{u}(\tau)\Big]$$

However, since $\underline{\tilde{x}}(t_0 | t_0)$ is independent of $\underline{u}(t)$ and $\underline{v}(t)$ is also independent of $\underline{u}(t)$, hence this equation can be reduced.

$$\mathrm{cov}\left[\underline{\tilde{x}}(t \mid t),\ \underline{u}(t)\right]$$

$$= \mathrm{cov}\left[\int_{t_0}^{t} \underline{\Psi}_c(t, \tau)\underline{G}(\tau)\underline{u}(\tau) d\tau,\ \underline{u}(t)\right]$$

$$= \int_{t_0}^{t} \underline{\Psi}_c(t, \tau)\underline{G}(\tau)E\left\{\underline{u}(\tau)\underline{u}'(t)\right\} d\tau$$

$$= \int_{t_0}^{t} \underline{\Psi}_c(t, \tau)\underline{G}(\tau)\underline{Q}(\tau)\delta(\tau - t) d\tau$$

$$= \frac{1}{2}\ \underline{G}(t)\underline{Q}(t)$$

$$\mathrm{cov}\left[\underline{\tilde{x}}(t \mid t),\ \underline{z}(t)\right]$$

$$= \mathrm{cov}\left[\underline{\tilde{x}}(t \mid t),\ \underline{H}(t)\left(\underline{\hat{x}}(t \mid t) + \underline{\tilde{x}}(t \mid t)\right) + \underline{v}(t)\right]$$

The above equation can be reduced

$$\mathrm{cov}\left[\underline{\tilde{x}}(t \mid t),\ \underline{z}(t)\right]$$

$$= \mathrm{cov}\left[\underline{\tilde{x}}(t \mid t),\ \underline{\tilde{x}}(t \mid t)\ \underline{H}'(t) + \mathrm{cov}\ \underline{\tilde{x}}(t \mid t),\ \underline{v}(t)\right]$$

$$= \underline{P}(t)\underline{H}'(t) - \int_{t_0}^{t} \underline{\Psi}_c(t, \tau)\underline{K}(\tau)E\left\{\underline{v}(\tau)\underline{v}'(t)\right\} d\tau$$

$$= \underline{P}(t)\underline{H}'(t) - \int_{t_0}^{t} \underline{\Psi}_c(t, \tau)\underline{K}(\tau)\underline{R}(\tau)\delta(\tau - t) d\tau$$

$$= \underline{P}(t)\underline{H}'(t) - 1/2\ \underline{K}(t)\underline{R}(t)$$

since

$$\mathrm{cov}\left[\underline{\tilde{x}}(t \mid t),\ \underline{\hat{x}}(t \mid t)\right] = 0\ .$$

Substituting $\underline{P}(t)\underline{H}'(t) = \underline{K}(t)\underline{R}(t)$ into the above equation yields:

$$\mathrm{cov}\left[\underline{\tilde{x}}(t \mid t),\ \underline{z}(t)\right] = \frac{1}{2}\,\underline{K}(t)\underline{R}(t)$$

To derive the formula for prediction, it is noted that if $t_1 > t$, then

$$\underline{x}(t_1) = \underline{\Phi}(t_1,\ t)\underline{x}(t) + \int_{t_0}^{t} \underline{\Phi}(t_1,\ \tau)\underline{G}(\tau)\underline{u}(\tau)d\tau$$

Since $\underline{u}(\tau)$ for $t < \tau \leq t_1$ is independent of $\underline{x}(\tau)$ in the interval $t_0 \leq \tau \leq t$, the term $\int_{t_0}^{t} \underline{\Phi}(t_1,\ \tau)\underline{G}(\tau)\underline{u}(\tau)d\tau$ vanishes in $E\left\{\underline{x}(t_1 \mid t)\right\}$; therefore

$$\underline{\hat{x}}(t_1 \mid t) = \underline{\Phi}(t_1,\ t)\underline{x}(t \mid t) \qquad \text{for } t_1 \geq t_0 \qquad \text{(Vc)}$$

If $t_1 < t$, one does not know whether $\underline{x}(\tau)$ and $\underline{u}(\tau)$ would be independent in the required interval. Hence the same conclusion cannot be drawn for the interpolation case.

These five equations are discussed in the following paragraphs.

The differential (difference) equation for optimal filter is

$$\underline{\hat{x}}(t + 1 \mid t) = \underline{\Phi}(t + 1,\ t)\underline{\hat{x}}(t \mid t - 1) + \underline{K}(t) \cdot\left[\underline{z}(t) - \underline{H}(t)\underline{\hat{x}}(t \mid t - 1)\right] \qquad \text{(Id)}$$

$$\frac{d\underline{\hat{x}}(t \mid t)}{dt} = \underline{F}(t)\underline{\hat{x}}(t \mid t) + \underline{K}(t)\left[\underline{z}(t) - \underline{H}(t)\underline{\hat{x}}(t \mid t)\right] \qquad \text{(Ic)}$$

It governs the optimal filter, which is excited by the observed signals and generates the best linear estimate of the message.

The initial state $\underline{x}(t_0|\ t_0) = \underline{0}$, since initially there are no observations and the mean of $\underline{x}(t_0)$ is zero. It is noted here that the first term of the equation is the estimate of $\underline{x}$ based on the observation before t. The quantity in the brackets is then the difference between the observation and the estimated value of the $\underline{x}(t)$. The weighting matrix $\underline{K}(t)$ weights the error signal to produce an increment to be added to the estimate; the value of $\hat{\underline{x}}(t + 1|t)$ is known immediately after time t, but it is not needed to compute the next estimate until time t + 1; this delay makes it possible to compute $\underline{K}(t)$ by digital computer. The general block diagram of the optimal filter is shown in Figs. 3 and 4. It is a feedback system. The input is the actual observation from which the latest estimate of the observable vector is subtracted. The difference is then multiplied by the weighting matrix $\underline{K}(t)$ to decrease the error of the new estimate $\hat{\underline{x}}$.

The prediction formula for Gauss-Markov sequence or process is:

$$\hat{\underline{x}}(t_1|t) = \underline{\Phi}(t_1,\ t)\hat{\underline{x}}(t|t - 1) \quad \text{for } t_1 \geqslant t + 1 \qquad \text{(Vd)}$$

$$\hat{\underline{x}}(t_1|t) = \underline{\Phi}(t_1,\ t)\underline{x}(t|t) \quad \text{for } t_1 \geqslant t \qquad \text{(Vc)}$$

These are the formulas for prediction only and are also shown in Figs. 3 and 4.

The differential (difference) equation for optimal estimation error is

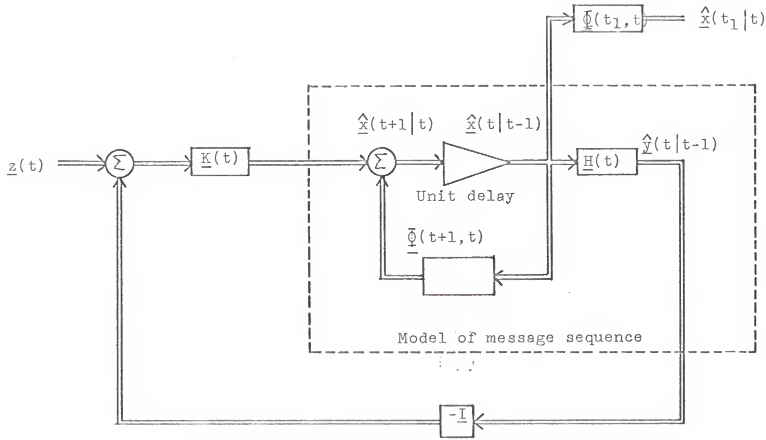

Fig. 3. General block diagram of optimal filter for discrete time case.

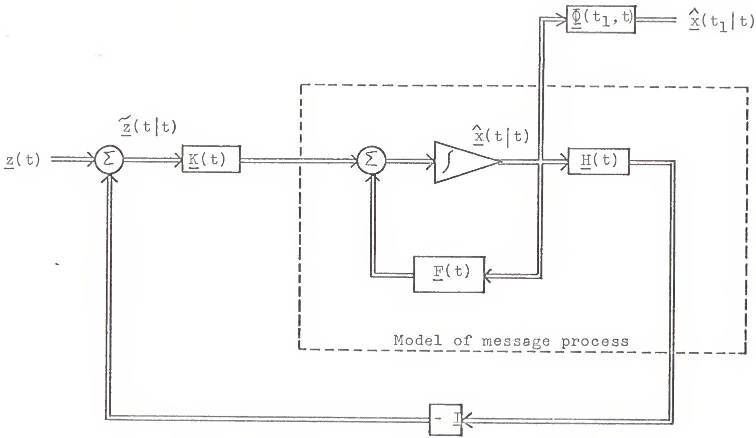

Fig. 4. General block diagram of optimal filter for continuous time case.

$$\tilde{\underline{x}}(t+1|t) = \underline{\Phi}(t+1, t)\tilde{\underline{x}}(t+1|t) + \underline{\Delta}(t+1, t)\underline{u}(t) - \underline{K}(t)\left[\underline{H}(t)\tilde{\underline{x}}(t|t-1) + \underline{v}(t)\right] \tag{IId}$$

$$\frac{d\tilde{\underline{x}}(t|t)}{dt} = \underline{F}(t)\tilde{\underline{x}}(t|t) + \underline{G}(t)\underline{u}(t) - \underline{K}(t)\left[\underline{H}(t)\tilde{\underline{x}}(t|t) + \underline{v}(t)\right] \tag{IIc}$$

It governs the optimal estimation error. General block diagram of the optimal estimation error is shown in Figs. 5 and 6.

Optimal gain formulas are:

$$\underline{K}(t) = \left[\underline{\Phi}(t+1, t)\underline{P}(t|t-1)\underline{H}'(t)\right]\left[\underline{H}(t)\underline{P}(t|t-1)\underline{H}'(t) + \underline{R}(t)\right]^{-1} \tag{IIId}$$

$$\underline{K}(t) = \underline{P}(t)\underline{H}'(t)\underline{R}^{-1}(t) \tag{IIIc}$$

They are gains of the optimal filter expressed in terms of the error covariance matrix $\underline{P}(t)$. The magnitude of the elements of $\underline{K}(t)$ are indicative of the amount of information contained in the signal $\underline{z}(t)$ at time t. In Figs. 3 and 4, it is shown that the optimal properties of the filter depend upon the proper selection of the weighting matrix $\underline{K}(t)$.

Variance equations are:

$$\underline{P}(t+1|t) = \underline{\Phi}(t+1, t)\left\{\underline{P}(t|t-1) - \left[\underline{P}(t|t-1)\underline{H}'(t)\right]\left[\underline{H}(t)\underline{P}(t|t-1)\underline{H}'(t) + \underline{R}(t)\right]^{-1}\underline{H}(t)\underline{P}(t|t-1)\right\}\underline{\Phi}'(t+1, t) + \underline{\Delta}(t+1, t)\underline{Q}(t)\underline{\Delta}'(t+1, t) \tag{IVd}$$

$$\frac{d\underline{P}(t)}{dt} = \underline{F}(t)\underline{P}(t) + \underline{P}(t)\underline{F}'(t) - \underline{P}(t)\underline{H}'(t)\underline{R}^{-1}(t)\underline{H}(t)\underline{P}(t) + \underline{G}(t)\underline{Q}(t)\underline{G}'(t) \tag{IVc}$$

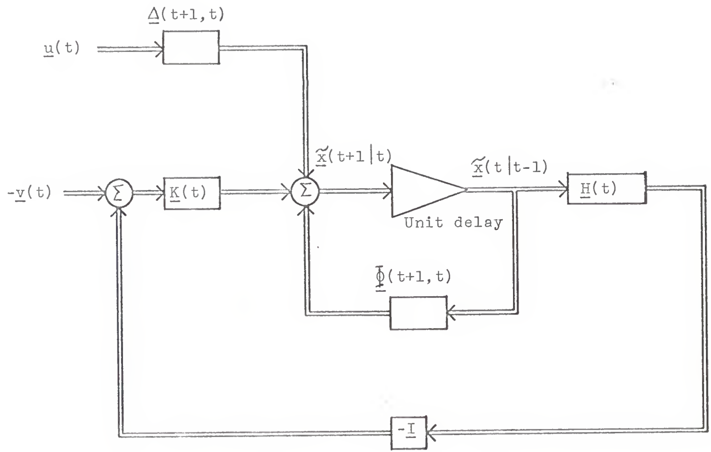

Fig. 5. General block diagram of optimal estimation error for discrete time case.

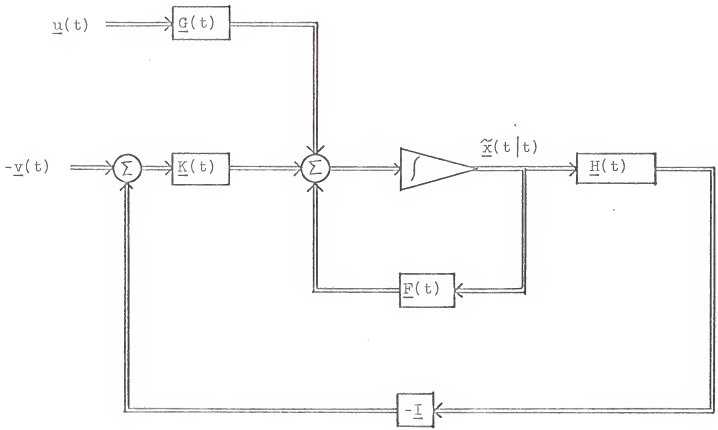

Fig. 6. General block diagram of optimal estimation error for continuous time case.

They are nonlinear differential or difference equations which govern the variance matrix of the errors of the optimal linear estimate. No $\underline{z}(t)$ terms are involved in the variance equations. This means the conditional covariance matrix does not depend on the values of conditional variables. Hence the variance equation is independent of the observations $\underline{z}(t)$. This is a special property of the Kalman filter. Since the gains of the optimal filter are governed by the variance equation, this means the structure of the optimal filter can be determined independently of the random data $\underline{z}(t)$. Given $\underline{z}(\tau)$, $t_0 \leqslant \tau \leqslant t$, one can completely determine the conditional distribution of the random variable $\hat{\underline{x}}(t)$ for all $t_1 \geqslant t$ from equations (I), (III), (IV), and (V). This is due to the gaussian and markovian assumptions. The variance equation is just another form of the Wiener-Hopf equation. Its solution yields the covariance matrix of the minimum filtering error which contains all the necessary information for the design of the optimal filter. The variance equation is also closely related to the calculus of variations as will be discussed later. The initial state of $\underline{P}$ is given as part of the problem statement. Then the solution of the variance equation is determined. If $\underline{P}_0(t)$ is nonnegative definite, then $\underline{P}(t)$ is also nonnegative definite for all $t \geqslant t_0$. In the stationary case, $\underline{\Phi}(t + 1, t)$, $\underline{\Delta}(t + 1, t)$, $\underline{H}(t)$, $\underline{R}(t)$, $\underline{Q}(t)$ are constants. Hence the covariance matrix is independent of time and can be calculated readily. The chief task in filtering theory is the study of the variance equation. This is difficult because the variance equation is nonlinear. But it is a very special one in that it is a matrix Riccati equation. A generalized matrix Riccati equation has the form

$$\frac{d\underline{\Gamma}}{dt} + \underline{\Gamma}\,\underline{G}_3\underline{P} + \underline{\Gamma}\,\underline{G}_4 - \underline{G}_1\underline{\Gamma} - \underline{G}_2 = \underline{0} ,$$

where $\underline{G}_1$, $\underline{G}_2$, $\underline{G}_3$, and $\underline{G}_4$ are $n_1 \times n_1$, $n_1 \times n_2$, $n_2 \times n_1$, and $n_2 \times n_2$ matrices, respectively. It is a classical (or scalar) Riccati equation when $n_1 = n_2 = 1$. It has the same form as the variance equation, with $n_1 = n_2 = n$; (Ref. 19), which is well known from the calculus of variations. An exact formula for the solution of the variance equation will be derived from the quadratic Hamiltonian function. Consider the Hamiltonian function $\mathcal{H}$ as defined by (Ref. 2).

$$\mathcal{H} = -\frac{1}{2}\left\|\underline{G}'(t)\underline{x}\right\|^2_{\underline{Q}(t)} - \underline{s}\,\underline{F}'(t)\underline{x} + \frac{1}{2}\left\|\underline{H}(t)\underline{s}\right\|^2 \tag{66}$$

The canonical differential equations of Hamilton are:

$$\begin{aligned} \frac{d\underline{x}}{dt} &= \operatorname{grad}_s \mathcal{H} = \frac{\partial \mathcal{H}}{\partial \underline{s}} = -\underline{F}'(t)\underline{x} + \underline{H}'(t)\underline{R}^{-1}(t)\underline{H}(t)\underline{s} \\ \frac{d\underline{s}}{dt} &= \operatorname{grad}_x \mathcal{H} = \frac{-\partial \mathcal{H}}{\partial \underline{x}} = \underline{G}(t)\underline{Q}(t)\underline{G}'(t)\underline{x} + \underline{F}(t)\underline{s} \end{aligned} \tag{67}$$

Let $\underline{S}(t)$, $\underline{X}(t)$ be the matrix solutions of the system (67) which satisfy the initial conditions.

$$\begin{aligned} \underline{X}(_0) &= \underline{I} \\ \underline{S}(t_0) &= \underline{P}(t_0) = \underline{P}_0 \end{aligned} \tag{68}$$

Assume

$$\underline{S}(t) = \underline{P}(t)\underline{X}(t) \tag{69}$$

Substituting equation (69) into equation (67), one obtains

$$\frac{d\underline{x}(t)}{dt} = \left[- \underline{F}'(t) + \underline{H}'(t)\underline{R}^{-1}(t)\underline{H}(t)\underline{P}(t)\right]\underline{X}(t) \tag{70}$$

$$\frac{d\underline{s}(t)}{dt} = \frac{d\underline{P}(t)}{dt}\,\underline{x}(t) + \underline{P}(t)\,\frac{d\underline{x}(t)}{dt}$$

$$= \left[\underline{G}(t)\underline{Q}(t)\underline{G}'(t) + \underline{F}(t)\underline{P}(t)\right]\underline{X}(t) \tag{71}$$

Combining equations (70) and (71), one gets

$$\frac{d\underline{P}(t)}{dt} = \underline{P}(t)\underline{F}'(t) + \underline{F}(t)\underline{P}(t) - \underline{P}(t)\underline{H}'(t)\underline{R}^{-1}(t)\underline{H}(t)\underline{P}(t) + \underline{G}(t)\underline{Q}(t)\underline{G}'(t) \tag{72}$$

It is identical with equation (IVc), hence the identity $\underline{S}(t) = \underline{P}(t)\underline{X}(t)$ is valid. Let $\underline{T}$ be the matrix of coefficients of the system (67).

$$\underline{T} = \begin{bmatrix} -\underline{F}'(t) & \underline{C}(t) \\ \underline{D}(t) & \underline{F}(t) \end{bmatrix}$$

where

$$\underline{C}(t) = \underline{H}'(t)\underline{R}^{-1}(t)\underline{H}(t)$$

$$\underline{D}(t) = \underline{G}(t)\underline{Q}(t)\underline{G}'(t)$$

Hence the system becomes

$$\begin{bmatrix} \frac{d\underline{x}}{dt} \\ \frac{d\underline{s}}{dt} \end{bmatrix} = \begin{bmatrix} -\underline{F}'(t) & \underline{C}(t) \\ \underline{D}(t) & \underline{F}(t) \end{bmatrix} \begin{bmatrix} \underline{x}(t) \\ \underline{s}(t) \end{bmatrix} \tag{73}$$

Let $\underline{\theta}(t, t_0)$ be 2n x 2n transition matrix of the system (73). $\underline{\theta}(t, t_0)$ be partitioned into four n x n submatrices as follows:

$$\underline{\theta}(t, t_0) = \begin{bmatrix} \underline{\theta}_{11}(t, t_0) & \underline{\theta}_{12}(t, t_0) \\ \underline{\theta}_{21}(t, t_0) & \underline{\theta}_{22}(t, t_0) \end{bmatrix}$$

where

$$\frac{d\underline{\theta}}{dt} = \underline{T}\ \underline{\theta}\ .$$

Thus

$$\begin{bmatrix} \underline{x}(t) \\ \underline{s}(t) \end{bmatrix} = \begin{bmatrix} \underline{\theta}_{11}(t, t_0) & \underline{\theta}_{12}(t, t_0) \\ \underline{\theta}_{21}(t, t_0) & \underline{\theta}_{22}(t, t_0) \end{bmatrix} \begin{bmatrix} \underline{x}(t_0) \\ \underline{s}(t_0) \end{bmatrix} \tag{74}$$

Substituting the initial condition (68) into the above equation, one has

$$\begin{aligned} \underline{x}(t) &= \underline{\theta}_{11}(t, t_0) + \underline{\theta}_{12}(t, t_0)\underline{P}_0 \\ \underline{s}(t) &= \underline{\theta}_{21}(t, t_0) + \underline{\theta}_{22}(t, t_0)\underline{P}_0 \end{aligned} \tag{75}$$

Combining equations (75) and (69), one gets an exact solution of variance equation (IVc).

$$\begin{aligned} \underline{P}(t) = &\left[\underline{\theta}_{21}(t, t_0) + \underline{\theta}_{22}(t, t_0)\underline{P}_0\right] \\ &\left[\underline{\theta}_{11}(t, t_0) + \underline{\theta}_{12}(t, t_0)\underline{P}_0\right]^{-1} \end{aligned} \tag{76}$$

EXAMPLE OF KALMAN FILTERING

Suppose a particle leaves the origin at time $t = 0$ and moves thereafter with a constant but unknown velocity of zero mean and known variance. The position of the particle is measured, the data being contaminated by the addition of white noise. What is the optimal estimate of the position and velocity of the particle at the time of the last measurement?

Let $\underline{x}_1(t)$ be the position and $\underline{x}_2(t)$ the velocity of the particle. Then the problem is represented by the model

$$\underline{x}_1(t+1) = \underline{x}_1(t) + \underline{x}_2(t)$$
$$\underline{x}_2(t+1) = \underline{x}_2(t)$$
$$\underline{z}_1(t) = \underline{x}_1(t) + \underline{v}(t) \ .$$

That is,

$$\underline{x}(t+1) = \underline{\Phi}(t+1,\ t)\underline{x}(t) + \underline{\Delta}(t+1,\ t)\underline{u}(t)$$

and

$$\underline{z}(t) = \underline{H}(t)\underline{x}(t) + \underline{v}(t)$$

where

$$\underline{\Phi}(t+1,\ t) = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \quad \underline{\Delta}(t) = 0\ , \quad \underline{H}(t) = \begin{bmatrix} 1, & 0 \end{bmatrix}$$

The additional conditions are

$$t_1 = t,\ t_0 = 0$$
$$E\left\{\underline{x}_1(0)\right\} = E\left\{\underline{x}_1{}^2(0)\right\} = E\left\{\underline{x}_2(0)\right\} = 0$$
$$E\left\{\underline{x}_2{}^2(0)\right\} = a^2$$

and

$$E\left\{\underline{v}^2(t)\right\} = \underline{R}(t) = b^2, \qquad \text{for all } t.$$

First, one wants to predict the position and velocity of the particle one step ahead.

$$\underline{P}(0) = \mathrm{cov}\left[\underline{x}(0),\ \underline{x}(1)\right]$$

$$= E\left\{\begin{bmatrix} \underline{x}_1{}^2(0) & \underline{x}_1(0)\underline{x}_2(0) \\ \underline{x}_2(0)\underline{x}_1(0) & \underline{x}_2{}^2(0) \end{bmatrix}\right\} = \begin{bmatrix} 0 & 0 \\ 0 & a^2 \end{bmatrix}$$

$$\underline{P}(1) = \underline{\Phi}\left\{\underline{P}(0) - \underline{P}(0)\underline{H}'\left[\underline{HP}(0)\underline{H}' + \underline{R}\right]^{-1} \underline{HP}(0)\right\} \underline{\Phi}$$

$$= \begin{bmatrix} a^2 & a^2 \\ a^2 & a^2 \end{bmatrix}$$

Then from equation (IVd), the covariance matrix is

$$\underline{P}(t + 1) = \underline{\Phi}\left\{\underline{P}(t) - \underline{P}(t)\underline{H}'(t)\left[\underline{H}(t)\underline{P}(t)\underline{H}'(t) + \underline{R}(t)\right]^{-1}\underline{H}(t)\underline{P}(t)\right\}\underline{\Phi}$$

$$= \frac{1}{p_{11}(t)+b^2}\begin{bmatrix} b^2\left[p_{11}(t)+2p_{12}(t)+p_{22}(t)\right]+\left[p_{11}(t)p_{22}(t)-p_{12}^2(t)\right], & b^2\left[p_{12}(t)+p_{22}(t)\right]+\left[p_{11}(t)p_{22}(t)-p_{12}^2(t)\right] \\ b^2\left[p_{22}(t)+p_{12}(t)\right]+\left[p_{11}(t)p_{22}(t)-p_{12}^2(t)\right], & b^2p_{22}(t) + \left[p_{11}(t)p_{22}(t)-p_{12}^2(t)\right] \end{bmatrix} \quad (77)$$

By simple substitutions of t = 0, t = 1, t = 2, . . ., into equation (77), it can be shown that

$$p_{11}(t)p_{22}(t) = p_{12}^2(t)$$

$$p_{11}(t) = t^2p_{22}(t) .$$

Then

$$p_{12}(t) = p_{21}(t) = tp_{22}(t) .$$

Substituting these equations into equation (77), the covariance matrix is:

$$\underline{P}(t+1) = \frac{1}{t^2p_{22}(t)+b^2}\begin{bmatrix} b^2(t^2 + 2t + 1)p_{22}(t) & b^2(t+1)p_{22}(t) \\ b^2(t + 1)p_{22}(t) & b^2p_{22}(t) \end{bmatrix}$$

$$= \frac{b^2p_{22}(t)}{t^2p_{22}(t) + b^2}\begin{bmatrix} (t + 1)^2 & (t + 1) \\ (t + 1) & 1 \end{bmatrix} \qquad \text{for } t \geqslant 0$$

where

$$p_{22}(t + 1) = \frac{b^2p_{22}(t)}{t^2p_{22}(t) + b^2} \qquad \text{for } t \geqslant 0 .$$

Assume

$$c(t) = t^2 + \frac{b^2}{p_{22}(t)}$$

and

$$c(t+1) = (t+1)^2 + \frac{b^2}{p_{22}(t+1)} = (t+1)^2 + \frac{b^2}{\dfrac{b^2p_{22}(t)}{t^2p_{22}(t)+b^2}}$$

$$= (t+1)^2 + \left[t^2 + \frac{b^2}{p_{22}(t)}\right] = (t+1)^2 + c(t)$$

where

$$c(0) = \frac{b^2}{a^2} .$$

Hence

$$\underline{P}(t+1) = \frac{b^2}{c(t)}\begin{bmatrix} (t + 1)^2 & (t + 1) \\ (t + 1) & 1 \end{bmatrix} \qquad \text{for } t \geqslant 0.$$

Hence

$$\underline{P}(t) = \frac{b^2}{c(t-1)} \begin{vmatrix} t^2 & t \\ t & 1 \end{vmatrix} \qquad \text{for } t \geq 1$$

where

$$c(0) = \frac{b^2}{a^2}$$

$$c(t) = c(t-1) + t^2 .$$

Then $\underline{K}(t)$ is found from equation (IIId).

$$\underline{K}(t) = \left[\underline{\Phi}\ \underline{P}(t)\underline{H}'\right]\left[\underline{HP}(t)\underline{H}' + \underline{R}(t)\right]^{-1}$$

$$= \frac{1}{t^2 + c(t-1)} \begin{bmatrix} t(t+1) \\ t \end{bmatrix}$$

$$\hat{\underline{x}}(t \mid t) = \underline{\Phi}\ \hat{\underline{x}}(t \mid t-1) + \underline{K}(t)\left[\underline{z}(t) - \underline{H}(t)\hat{\underline{x}}(t \mid t-1)\right]$$

where

$$\hat{\underline{x}}(t_0) = \hat{\underline{x}}(0) = \underline{0} .$$

If $t >> 1$, then

$$c(t) = c(t-1) + t^2$$

$$= c(t-2) + t^2 + (t-1)^2$$

$$= c(0) + t^2 + (t-1)^2 + \ldots + 1$$

$$= c(0) + \frac{t(t+1)(2t+1)}{6}$$

$$\cong c(0) + \frac{t^3}{3}$$

$$\cong \frac{b^2}{a^2} + \frac{t^3}{3} .$$

Hence

$$\underline{P}(t) \cong \frac{b^2}{\frac{b^2}{a^2} + \frac{t^3}{3}} \begin{bmatrix} t^2 & t \\ t & 1 \end{bmatrix} \cong \frac{a^2b^2}{b^2 + \frac{a^2t^3}{3}} \begin{bmatrix} t^2 & t \\ t & 1 \end{bmatrix}$$

and

$$\underline{K}(t) \cong \frac{1}{t^2 + (\frac{b^2}{a^2} + \frac{t^3}{3})} \begin{bmatrix} t(t+1) \\ t \end{bmatrix} = \frac{a^2}{b^2 + \frac{a^2t^3}{3}} \begin{bmatrix} t^2 \\ t \end{bmatrix}$$

for $t \gg 1$.

It is obvious that $k_1 \longrightarrow 0$ and $k_2 \longrightarrow 0$ for $t \gg 1$. This means that as the number of observations becomes large, the estimates $\hat{\underline{x}}_1(t + 1 \mid t + 1)$ and $\hat{\underline{x}}_2(t + 1 \mid t + 1)$ will depend on the previous estimate. It will be shown later that for $t \gg 1$, the discrete observation case is essentially the same as prediction based on continuous observations. Shinbrot (11), who treated this problem on the continuous data case using a completely different method, obtained the same results as here. (See Appendix A.)

If the position of the particle is continually observed in the presence of additive white noise, the problem is represented by the model,

$$\left.\begin{aligned} \frac{d\underline{x}_1}{dt} &= \underline{x}_2 \\ \frac{d\underline{x}_2}{dt} &= 0 \\ \underline{z} &= \underline{x}_1 + \underline{v} \end{aligned}\right\} \quad \frac{d\underline{x}}{dt} = \underline{F}(t)\underline{x}(t) + \underline{G}(t)\underline{u}(t)$$

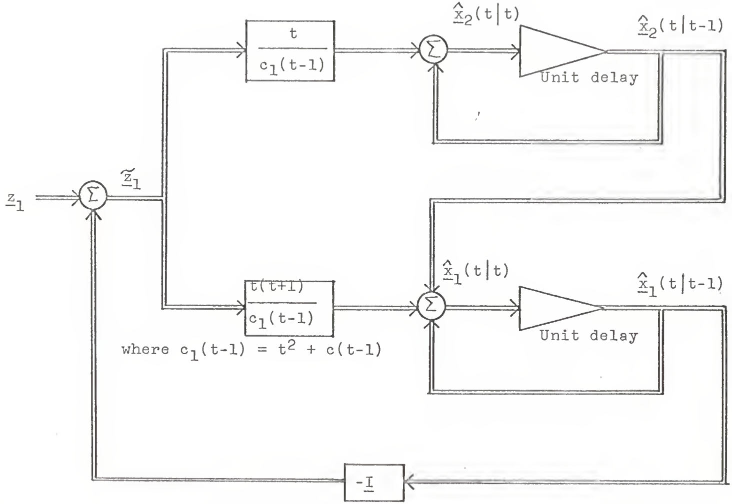

Fig. 7. Block diagram of optimal filter of the example.

where

$$\underline{F}(t) = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \quad \underline{G}(t) = \underline{0}, \quad \text{and} \quad \underline{H} = \begin{bmatrix} 1, & 0 \end{bmatrix} .$$

The initial state of covariance matrix is

$$\underline{P}(0) = \mathrm{cov}\left[\underline{x}_1(t_0),\ \underline{x}_0(t_0)\right] = \begin{bmatrix} 0 & 0 \\ 0 & a^2 \end{bmatrix}$$

$$\frac{d\underline{P}(t)}{dt} = \underline{FP} + \underline{PF}' - \underline{PH}'\underline{R}^{-1}\underline{HP} + \underline{GQG}'$$

$$= \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix} + \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}$$

$$- \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} \begin{bmatrix} \frac{1}{b^2} \end{bmatrix} \begin{bmatrix} 1, & 0 \end{bmatrix} \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix}$$

$$= \begin{bmatrix} 2p_{12} & p_{22} \\ p_{22} & 0 \end{bmatrix} - \frac{1}{b^2} \begin{bmatrix} {p_{11}}^2 & p_{11}p_{12} \\ p_{11}p_{21} & p_{12}p_{21} \end{bmatrix}$$

$$\begin{bmatrix} \frac{dp_{11}}{dt} & \frac{dp_{12}}{dt} \\ \frac{dp_{21}}{dt} & \frac{dp_{22}}{dt} \end{bmatrix} = \begin{bmatrix} 2p_{12} - \frac{1}{b^2}\,{p_{11}}^2 & p_{22} - \frac{1}{b^2}\,p_{11}p_{12} \\ p_{21} - \frac{1}{b^2}\,p_{11}p_{21} & -\frac{1}{b^2}\,p_{12}p_{21} \end{bmatrix}$$

Hence

$$\frac{dp_{11}}{dt} = 2p_{12} - \frac{{p_{11}}^2}{b^2}$$

$$\frac{dp_{12}}{dt} = \frac{dp_{21}}{dt} = p_{22} - \frac{p_{11}p_{21}}{b^2}$$

$$\frac{dp_{22}}{dt} = \frac{-p_{12}^{2}}{b^2} .$$

Using Adam's method (5), one can solve these nonlinear differential equations by digital computer. (See Appendix B.)

One can get an exact solution of the variance equation from the canonical differential equations of Hamilton.

$$\frac{d\underline{x}}{dt} = \frac{\partial \mathcal{H}}{\partial \underline{s}} = - \underline{F}'(t)\underline{x} + \underline{H}'(t)\underline{R}^{-1}(t)\underline{H}(t)\underline{s}$$

$$\frac{d\underline{s}}{dt} = \frac{-\partial \mathcal{H}}{\partial \underline{x}} = \underline{G}(t)\underline{Q}(t)\underline{G}'(t)\underline{x} + \underline{F}(t)\underline{s}$$

In this case

$$\frac{d\underline{x}}{dt} = - \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \underline{x}_1 \\ \underline{x}_2 \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} \begin{bmatrix} \frac{1}{b^2} \end{bmatrix} \begin{bmatrix} 1, & 0 \end{bmatrix} \begin{bmatrix} \underline{s}_1 \\ \underline{s}_2 \end{bmatrix}$$

$$\frac{d\underline{s}}{dt} = \frac{-\partial \mathcal{H}}{d\underline{x}} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \underline{s}_1 \\ \underline{s}_2 \end{bmatrix} .$$

Then the matrix of coefficients of the Hamiltonian equations is:

$$I = \begin{bmatrix} 0 & 0 & \frac{1}{b^2} & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

and the corresponding transition matrix is

$$\underline{\theta}(t, t_0) = \exp \underline{T}(t - t_0) = \sum_{i=0}^{\infty} \frac{\left[\underline{T}(t - t_0)\right]^i}{i!}$$

$$= \underline{T}^0 + \underline{T}t + \underline{T}^2 t^2/2 + \underline{T}^3 t^3/6 + \ldots \text{ as } t_0 = 0$$

where

$$\underline{T}^0 = \underline{I} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\underline{T} = \begin{bmatrix} 0 & 0 & \frac{1}{b^2} & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\underline{T}^2 = \begin{bmatrix} 0 & 0 & 0 & \frac{1}{b^2} \\ 0 & 0 & \frac{-1}{b^2} & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\underline{T}^3 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{-1}{b^2} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\underline{T}^4 = \underline{T}^5 = \underline{T}^6 = \ldots = \underline{0}$$

Hence

$$\underline{\theta}(t) = \begin{bmatrix} 1 & 0 & \dfrac{t}{b^2} & \dfrac{t^2}{2b^2} \\ -t & 1 & \dfrac{-t^2}{2b^2} & \dfrac{-t^3}{6b^2} \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

The four submatrices of $\underline{\theta}$ are as follows:

$$\underline{\theta}_{11}(t) = \begin{bmatrix} 1 & 0 \\ -t & 1 \end{bmatrix} \qquad \underline{\theta}_{12}(t) = \begin{bmatrix} t/b^2 & t/2b^2 \\ -t^2/2b^2 & -t^3/6b^2 \end{bmatrix}$$

$$\underline{\theta}_{21}(t) = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \qquad \underline{\theta}_{22}(t) = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$$

From equation (76), one gets

$$\underline{P}(t) = \left[\underline{\theta}_{21}(t) + \underline{\theta}_{22}(t)\underline{P}_0\right] \left[\underline{\theta}_{11}(t) + \underline{\theta}_{12}(t)\underline{P}_0\right]^{-1}$$

$$= a^2 \begin{bmatrix} 0 & t \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & \dfrac{a^2t^2}{2b^2} \\ -t & \dfrac{6b^2 - a^2t^3}{6b^2} \end{bmatrix}^{-1}$$

$$= \frac{a^2b^2}{b^2 + \dfrac{a^2t^3}{3}} \begin{bmatrix} t^2 & t \\ t & 1 \end{bmatrix}$$

$$\underline{K}(t) = \underline{P}(t)\underline{H}'(t)\underline{R}^{-1}(t)$$

$$= \frac{a^2b^2}{b^2 + \frac{a^2t^3}{3}} \begin{bmatrix} t^2 & t \\ t & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} \left(\frac{1}{b^2}\right)$$

$$= \frac{a^2}{b^2 + \frac{a^2t^3}{3}} \begin{bmatrix} t^2 \\ t \end{bmatrix}$$

This is the same result as before.

CONCLUSION

A Kalman filter is constructed using matrix theory and the state and state transition approach to linear system theory to solve the Wiener problem. Kalman's derivations are not affected by the nonstationarity of the process or the finiteness of available data. Hence Kalman filtering can be applied to both the stationary and the nonstationary cases. An important feature of the Kalman filter is that the structure of the filter can be determined independent of the random data inputs. It provides the error analysis independent of any data inputs. Further, it performs this error analysis in a very efficient way. The high-speed digital computer plays an important role in this error analysis. Theoretically speaking, Kalman filtering enlarges the realm of Wiener filtering and clarifies the fundamental assumptions and their consequences. Practically speaking, Kalman filtering is well adapted to digital computer usage, while Wiener filtering is not.

A Kalman filter can be applied to estimate the position, velocity, and altitude of a terrestrial or space navigator. For example (Ref. 13), Kalman filter estimates of the position and velocity of a space vehicle can be used for the purpose of midcourse guidance. The source of information is the sequence of measurements of three space angles. The measurements cannot be determined exactly because of the presence of additive white noise. The equation of motion is nonlinear, but it can be linearized by Taylor's expansion. The Kalman filter solves the estimation problem using a computation scheme which weights the incoming measurements in an optimal sense to produce an up-to-date optimal estimate of position and velocity. This computation scheme is readily implemented by a digital computer. The restrictions on the speed of the digital computer are that the time required to complete the computation cycle must be less than the time interval between successive observations. Hence it is clear that a Kalman filter is heavily dependent upon digital computer technology.

ACKNOWLEDGMENT

Grateful acknowledgment is made to Dr. F. W. Harris for efficient direction of the work, encouragement throughout the preparation of this report, and for aid in organizing the material.

REFERENCES

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

2. R. E. Kalman and R. S. Bucy, "New results in linear filtering and prediction theory," Basic Engineering (ASME Trans.), 83D, 1961.

3. L. A. Zadeh and C. A. Desoer, "Linear system theory," McGraw-Hill Book Co., Inc., 1963.

4. P. M. Derusso, P. J. Roy, and C. M. Close, "State Variables for Engineers," John Wiley & Sons, Inc., 1965.

5. P. Henrici, "Elements of numerical analysis," John Wiley and Sons, Inc., 1964.

6. A. Papoulis, "Probability, random variables, and stochastic processes," McGraw-Hill Book Co., Inc., 1964.

7. S. Sherman, "Non-mean-square error criteria," IRE Prof. Group on Information Theory, IT-U, 1958.

8. J. L. Doob, "Stochastic processes," John Wiley and Sons, Inc., 1953.

9. T. W. Anderson, "An introduction to multivariate statistical analysis," John Wiley and Sons, Inc., 1958.

10. M. Athans and P. L. Falb, "Optimal control," McGraw-Hill Book Co., Inc., 1965.

11. M. Shinbrot, "Optimization of time-varying linear system with nonstationary inputs," Trans. ASME, Vol. 80, 1958, pp. 457-462.

12. G. L. Smith, S. F. Schmidt, and L. A. McGee, "Application of statistical filter theory to the optimal estimation of position and velocity on board a circumlunar vehicle," Tech. Rept. R-135, Nov. 20, 1961, Ames Research Center, NASA, Moffett Field, Calif.

13. S. L. Fagin, "Recursive linear regression theory, optimal filter, and error analysis of optimal system," IEEE International Convention, 1964.

14. R. Deutsch, "Estimation theory," Prentice-Hall, Inc., 1965.

ACKNOWLEDGMENT

Grateful acknowledgment is made to Dr. F. W. Harris for efficient direction of the work, encouragement throughout the preparation of this report, and for aid in organizing the material.

REFERENCES

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

2. R. E. Kalman and R. S. Bucy, "New results in linear filtering and prediction theory," Basic Engineering (ASME Trans.), 83D, 1961.

3. L. A. Zadeh and C. A. Desoer, "Linear system theory," McGraw-Hill Book Co., Inc., 1963.

4. P. M. Derusso, P. J. Roy, and C. M. Close, "State Variables for Engineers," John Wiley & Sons, Inc., 1965.

5. P. Henrici, "Elements of numerical analysis," John Wiley and Sons, Inc., 1964.

6. A. Papoulis, "Probability, random variables, and stochastic processes," McGraw-Hill Book Co., Inc., 1964.

7. S. Sherman, "Non-mean-square error criteria," IRE Prof. Group on Information Theory, IT-U, 1958.

8. J. L. Doob, "Stochastic processes," John Wiley and Sons, Inc., 1953.

9. T. W. Anderson, "An introduction to multivariate statistical analysis," John Wiley and Sons, Inc., 1958.

10. M. Athans and P. L. Falb, "Optimal control," McGraw-Hill Book Co., Inc., 1965.

11. M. Shinbrot, "Optimization of time-varying linear system with nonstationary inputs," Trans. ASME, Vol. 80, 1958, pp. 457-462.

12. G. L. Smith, S. F. Schmidt, and L. A. McGee, "Application of statistical filter theory to the optimal estimation of position and velocity on board a circumlunar vehicle," Tech. Rept. R-135, Nov. 20, 1961, Ames Research Center, NASA, Moffett Field, Calif.

13. S. L. Fagin, "Recursive linear regression theory, optimal filter, and error analysis of optimal system," IEEE International Convention, 1964.

14. R. Deutsch, "Estimation theory," Prentice-Hall, Inc., 1965.

15. Robert C. K. Lee, "Optimal estimation identification and Control," The MIT Press, No. 28, 1964.

16. N. Wiener, "The extrapolation, interpolation, and smoothing of stationary time series," John Wiley and Sons, Inc., 1949.

17. Y. W. Lee, "Statistical theory of communication," John Wiley and Sons, Inc., 1960.

18. M. Loeve, "Probability theory," 2nd edition, D. Van Nostrand and Company, 1960.

19. J. J. Levin, "On the matrix Riccati equation," Proc. American Mathematical Society, Vol. 10, 1959.

20. V. S. Pugachev, "General condition for the minimum mean square error in a dynamic system," Avt. i Telemekh., Vol. 17, 1958.

APPENDICES

APPENDIX A

Sinbrot solved example 1 in his paper "Optimization of Time-varying Linear Systems with Nonstationary Inputs" with correlation functions as follows.

The input to the system is

$$f_i(t,\ p,\ Q) = f_m(t,\ p) + f_n(t,\ Q)$$

where f_m and f_n are message and noise, respectively.

$$f_m(t,\ p) = Pt$$

P is the unknown velocity of the particle. The mean square value of P is a^2. The hoise is white and is independent of the message. Hence

$$\phi_{dd}(t,\ \tau) = \phi_{dm}(t,\ \tau) = \phi_{di}(t,\ \tau) = \phi_{mm}(t,\ \tau) = a^2 t\tau$$

$$\phi_{nn}(t,\ \tau) = b^2\delta(t - \tau)$$

$$\phi_{ii}(t,\ \tau) = \phi_{mm}(t,\ \tau) + \phi_{nn}(t,\ \tau) = a^2 t\tau + b^2\delta(t - \tau)$$

The cross correlation of input and desired output is

$$\phi_{di}(t,\tau) = \phi_{dm}(t,\tau) = \int_0^t h(t,\sigma)\phi_{mm}(\tau,\sigma)\,d\sigma + b^2 h(t,\tau)$$
$$\text{for } 0 \leq \tau \leq t\ .$$

Hence

$$a^2 t\tau = a^2\tau \int_0^t \sigma\, h(t,\sigma)\,d\sigma + b^2 h(t,\ \tau)$$

and multiplying by τ on both sides and integrating with respect to τ from zero to t yields

$$\int_0^t a^2 t\tau^2 d\tau = \int_0^t a^2\tau^2 \left[\int_0^t \sigma h(t,\sigma)\, d\sigma\right] d\tau + \int_0^t b^2 h(t,\tau)\, d\tau \; .$$

Let

$$j(t) = \int_0^t \sigma\, h(t,\sigma)\, d\sigma \; .$$

Hence

$$a^2 t \frac{t^3}{3} = a^2 \frac{t^3}{3} j(t) + b^2 j(t)$$

$$j(t) = \frac{a^2 t^4}{a^2 t^3 + 3b^2} \quad .$$

Shinbrot pointed out that the optimal system response is

$$h(t,\tau) = a^2\tau \left[\frac{t - j(t)}{b^2}\right] = \frac{3a^2 t\tau}{a^2 t^3 + 3b^2} \qquad t \geqslant \tau$$

and the mean square error is

$$\epsilon^2 = b^2 h(t,\ t)$$

$$= \frac{3\ a^2 b^2 t^2}{a^2 t^3 + 3b^2}$$

$$= \frac{a^2 b^2 t^2}{\dfrac{a^2 t^3}{3} + b^2} \quad .$$

APPENDIX B

```
C   C
      DIMENSION X(110),Y(110),Z(110),T(110)
    2 FORMAT(8X,1HT,16X,3HP11,13X,3HP12,13X,3HP22)
    3 FORMAT(4F16.8)
      X(1)=0
      Y(1)=0
      Z(1)=1.0
      T(1)=0
      DT=0.2
      PUNCH2
      PUNCH30,T(1),X(1),Y(1),Z(1)
      DO 1 I=2,101
      K=I-1
      DX1=(2.*Y(K)-X(K)**2/16.)*DT
      DY1=(Z(K)-X(K)*Y(K)/16.)*DT
      DZ1=(-Y(K)**2/16.)*DT
      T(I)=T(K)+DT
      X(I)=X(K)+DX1
      Y(I)=Y'K)+DY1
      Z(I)=Z(K)+DZ1
      DX2=(2.*Y(I)-X(I)**2/16.)*DT
      DY2=(Z(I)-X(I)*Y(K)/16.)*DT
      DZ2=(-Y(I)**2/16.)*DT
      X(I)=X(K)+(DX1+DX2)/2.
      Y(I)=Y(K)+(DY1+DY2)/2.
      Z(I)=Z(K)+(DZ1+DZ2)/2.
    1 PUNCH30,T(I),X(I),Y(I),Z(I)
      STOP
      END
C   C
```

T	P11	P12	P22
.00000000	0.00000000	0.00000000	1.00000000
.20000000	.04000000	.20000000	.99975000
.40000000	.15967003	.39970003	.99850075
.60000000	.35673287	.59600513	.99526273
.80000000	.63418151	.79503912	.98908375
1.00000000	.98159193	.96223574	.97909446
1.20000000	1.39361610	1.16245870	.96456728
1.40000000	1.85996400	1.33023080	.94498066
1.60000000	2.36756760	1.48202350	.92007723
1.80000000	2.90116270	1.61459150	.88990312
2.00000000	3.44424330	1.72530550	.85481731
2.20000000	3.98025440	1.81242570	.81546529
2.40000000	4.49582580	1.87526970	.77271897
2.60000000	4.97183970	1.91424550	.72759249
2.80000000	5.40417120	1.93075290	.68114855
3.00000000	5.78402330	1.92698220	.63441070
3.20000000	6.10787110	1.90365480	.58829361
3.40000000	6.37509830	1.86975580	.54355782

3.60000000	6.58744450	1.82229570	.50078943
3.80000000	6.74838000	1.76612530	.46040091
4.00000000	6.86250150	1.70381110	.42264687
4.20000000	6.93500800	1.63756800	.38764847
4.40000000	6.97128450	1.56523770	.35542097
4.60000000	6.97659860	1.50030060	.32590066
4.80000000	6.95569850	1.43190750	.29896877
5.00000000	6.91369300	1.36492260	.27447133
5.20000000	6.85399580	1.29996950	.25223477
5.40000000	6.78031390	1.23747550	.23207762
5.60000000	6.69566540	1.17771220	.21381881
5.80000000	6.60261540	1.12083010	.19728331
6.00000000	6.50332050	1.06688810	.18230570
6.20000000	6.39957700	1.01587710	.16873216
6.40000000	6.29286820	.96773510	.15642138
6.60000000	6.18440900	.92238221	.14524474
6.80000000	6.07518640	.87969196	.13508592
7.00000000	5.96599570	.83954005	.12534030
7.20000000	5.85747140	.80179088	.11741411
7.40000000	5.75011460	.76630628	.10972352
7.60000000	5.64431590	.73294599	.10269374
7.80000000	5.54037470	.70158500	.09625810
8.00000000	5.43651580	.67208522	.09035723
8.20000000	5.33899280	.64432649	.08493825
8.40000000	5.24164980	.61619218	.07995407
8.60000000	5.14683050	.59357249	.07536272
8.80000000	5.05448620	.57036450	.07112678
9.00000000	4.96463220	.54847209	.06721287
9.20000000	4.87726330	.52780569	.06359114
9.40000000	4.79235770	.50828204	.06023492
9.60000000	4.70988100	.48982584	.05712034
9.80000000	4.62978910	.47235942	.05422597
10.00000000	4.55203030	.45582237	.05153260
10.20000000	4.47654760	.44015117	.04902296
10.40000000	4.40328610	.42528689	.04668149
10.60000000	4.33216420	.41118281	.04449417
10.80000000	4.26313480	.39778412	.04244837
11.00000000	4.19612610	.38504764	.04053265
11.20000000	4.13107220	.37293150	.03873665
11.40000000	4.06790770	.36139692	.03705101
11.60000000	4.00656790	.35040795	.03546720
11.80000000	3.94698950	.33993123	.03397749
12.00000000	3.88911040	.32995582	.03257465
12.20000000	3.83287610	.32039299	.03125284
12.40000000	3.77821000	.31127603	.03000562
12.60000000	3.72507320	.30256013	.02882784
12.80000000	3.67340470	.29422218	.02771460
13.00000000	3.62315120	.28624069	.02666142
13.20000000	3.57426170	.27859562	.02566419
13.40000000	3.52668660	.27126829	.02471914
13.60000000	3.48037850	.26424126	.02382278

13.80000000	3.43529170	.25749832	.02297194
14.00000000	3.39138230	.25102420	.02216366
14.20000000	3.34866820	.24480471	.02139524
14.40000000	3.30692890	.23882655	.02066416
14.60000000	3.26631550	.23307726	.01996812
14.80000000	3.22670090	.22754517	.01930495
15.00000000	3.18807920	.22221934	.01867269
15.20000000	3.15140630	.21708949	.01806948
15.40000000	3.11364920	.21214599	.01749363
15.60000000	3.07777650	.20737978	.01694353
15.80000000	3.04275790	.20278233	.01641772
16.00000000	3.00856450	.19834563	.01591481
16.20000000	2.97516850	.19406213	.01543354
16.40000000	2.94254320	.18992472	.01497270
16.60000000	2.91066300	.18592670	.01453118
16.80000000	2.87950350	.18206174	.01410795
17.00000000	2.84904110	.17832388	.01370202
17.20000000	2.81925320	.17470747	.01331250
17.40000000	2.79011810	.17120719	.01293852
17.60000000	2.76161500	.16781800	.01257929
17.80000000	2.73372400	.16453513	.01223406
18.00000000	2.70642580	.16135407	.01190213
18.20000000	2.67970200	.15827054	.01158284
18.40000000	2.65353490	.15528049	.01127557
18.60000000	2.62790750	.15238008	.01097974
18.80000000	2.60280340	.14956566	.01069480
19.00000000	2.57820680	.14683377	.01042023
19.20000000	2.55410260	.14418112	.01015554
19.40000000	2.53047620	.14160459	.00990029
19.60000000	2.50731360	.13910121	.00965402
19.80000000	2.48460140	.13666815	.00941634
20.00000000	2.46232650	.13430272	.00918687

STOP END OF PROGRAM AT STATEMENT 0010 + 01 LINES

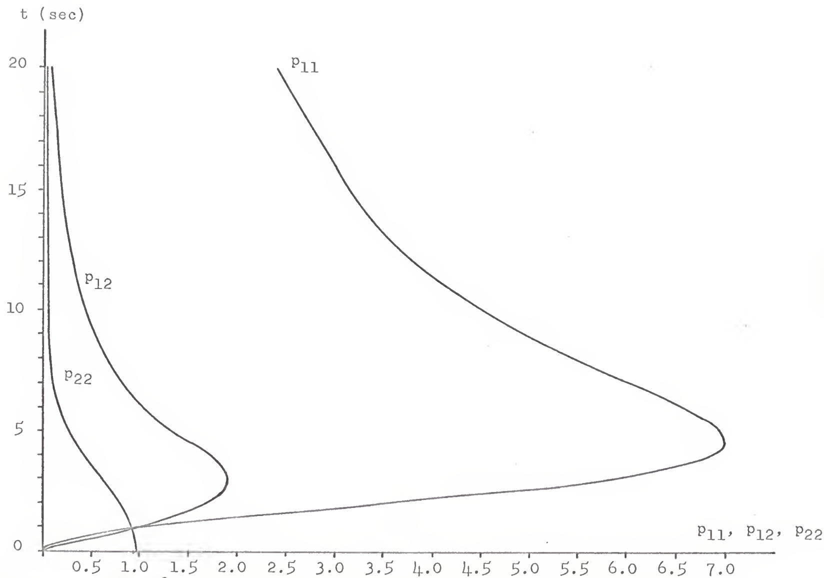

Fig. 8. Solution of the variance equation of the example, for $a^2 = 1$ and $b^2 = 16$.

A STUDY OF KALMAN FILTERING

by

JOHN Y. S. HO

B. S., National Taiwan University, 1962

AN ABSTRACT OF A MASTER'S REPORT

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

Department of Electrical Engineering

KANSAS STATE UNIVERSITY
Manhattan, Kansas

1967

Kalman approached the Wiener problem from the "state" point of view and thought of linear filtering as orthogonal projection in Hilbert space. The physical relationship among the state $\underline{x}(t)$, the white noises $\underline{u}$, $\underline{v}$, and the observation $\underline{z}(t)$ is described by a linear system which is specified by a system of first-order linear differential (difference) equations. The statistical description of the white noises $\underline{u}$, $\underline{v}$ is given as part of the problem statement. The problem is that one observes $\underline{z}(\tau)$ over some interval of time, $t_0 \leq \tau \leq t$, and one wants to find the optimal estimate $\hat{\underline{x}}(t_1)$ of $\underline{x}(t_1)$ given these observations. This optimal estimate $\hat{\underline{x}}(t_1)$ of $\underline{x}(t_1)$ will minimize the expected loss

$$E\left\{L\left(\underline{x}(t_1) - \hat{\underline{x}}(t_1)\right)\right\} = E\left\{\sum_{i=1}^{n}\left[x_i(t_1) - \hat{x}_i(t_1)\right]^2\right\} .$$

In the case where $t_1 < t$, this is called the data-smoothing (interpolation) problem; where $t_1 = t$, this is called the filtering problem; where $t_1 > t$, this is called the prediction (extrapolation) problem. Using the conditional distributions and expectations, Kalman transformed the Wiener-Hopf integral equation into a nonlinear differential equation of the Riccati type which is closely related to the Hamiltonian differential equations of the calculus of variations. The solution of this nonlinear differential equation yields the covariance matrix of the minimum filtering error which will determine the structure of the optimal filter independent of the input data. Kalman's approach is not affected by nonstationarity of the process $\underline{x}(t)$ to be estimated, or the finiteness of available data. Hence a

Kalman filter can be applied to both cases, stationary and non-stationary. Practically the main contribution of the Kalman filter is that it provides numerical procedures well adapted to digital computer usage.