THE JOHNS HOPKINS UNIVERSITY
APPLIED PHYSICS LABORATORY
8621 GEORGIA AVENUE SILVER SPRING, MD.

Operating under Contract NOrd 7386
with the Bureau of Naval Weapons, Department of the Navy

LINEAR ESTIMATION
IN
STOCHASTIC PROCESSES

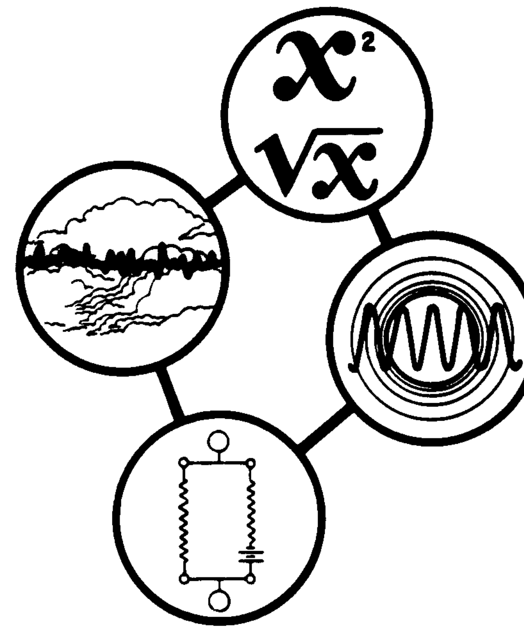
by

A. George Carlton

Bumblebee Series
Report No. 311
Copy No. 19

March 1962

BUMBLEBEE
REPORT NO. 311
MARCH 1962



LINEAR ESTIMATION IN STOCHASTIC PROCESSES

By
A. GEORGE CARLTON

THE JOHNS HOPKINS UNIVERSITY
APPLIED PHYSICS LABORATORY
8621 GEORGIA AVENUE SILVER SPRING, MARYLAND
Operating under Contract NOw 62-0604-c with the Bureau of Naval Weapons, Department of the Navy

ABSTRACT

A general theory is developed for obtaining linear estimates, of minimum mean square error, of components of random processes. The fundamental requirement for applying the theory is that the process in question be vector Markov in the wide sense, a requirement typically satisfied exactly and almost invariably satisfied approximately. This estimation theory has considerable advantages over methods of functional analysis mentioned in Section II in that: (a) it is concerned with vector observation functions which may be time-dependent linear combinations of components of the process; (b) very general nonstationary processes are treated: (c) explicit solutions are obtained without difficulty; (d) linear prediction is an immediate consequence of linear estimation; and (e) estimates are obtained sequentially. In contrast to the typical approach of considering the observation as the sum of a quantity of interest (sometimes called the signal or a linear combination of unknown parameters), together with disturbances (sometimes called noise or errors of observation), the observation is considered a linear combination of components of a vector process, and the best linear estimate of a complete state vector of the sample process is obtained.

(The essentials of the theory were presented in an internal memorandum, CLM-30, 1960, and a more complete exposition, with examples, in CLM-46 and CLM-46A, 1961.)

TABLE OF CONTENTS

I. INTRODUCTION

The basic problem considered is the sequential determination of best linear estimates of a vector $x(\tau)$ based on non-denumerable vector observations linear in $x(\tau)$, for $\{x(\tau)\}$ vector Markov processes in the wide sense; best linear estimates $\hat{x}(\tau)$ are by definition linear estimates minimizing the variance of estimate, $E\,[\hat{x}(\tau) - x(\tau)]\,[\widetilde{\hat{x}(\tau) - x(\tau)}]$.

Wide sense vector Markov processes, considered in Appendix B, are those processes $\{x(\tau)\}$ such that, for any time sequence $\tau_1 \le \tau_2 \le \cdots \le \tau_\nu$, the best linear estimate of $x(\tau_\nu)$ on the basis of $x(\tau_1), \cdots, x(\tau_{\nu-1})$ is a linear function of $x(\tau_{\nu-1})$, not depending on $x(\tau_1), \cdots, x(\tau_{\nu-2})$. This is equivalent to

$$x(\tau_\nu) = L_{\nu,\nu-1}\, x(\tau_{\nu-1}) + v_\nu,$$

with the transition matrices $L_{\nu,\nu-1}$ not depending on the sample function $x(\tau)$, and v_ν orthogonal to all preceding v and x. With slight restrictions, wide sense vector Markov processes satisfy the relations ($\tau_2 \ge \tau_1$),

$$x(\tau_2) = K(\tau_2,\tau_1)x(\tau_1) + w(\tau_2,\tau_1),$$

$$dK(\tau_2,\tau_1)/d\tau = B(\tau_2)\, K(\tau_2,\tau_1)$$

$$Ew(\tau_2,\tau_1)\, \widetilde{w}(\tau_2,\tau_1) = \int_{\tau_1}^{\tau_2} K(\tau_2,\sigma)\, Q(\sigma)\, \widetilde{K}(\tau_2,\sigma)\, d\sigma,$$

with the $w(\tau_2,\tau_1)$ corresponding to non-overlapping intervals mutually orthogonal, $B(\tau)$ and $Q(\tau)$ integrable square matrices, and $Q(\tau)$ non-negative definite. The structure of $\{x(\tau)\}$ is thus characterized by $B(\tau)$ and $Q(\tau)$.

We consider observations $y(\tau) = A(\tau)x(\tau)$ made over a finite or non-denumerable set of values of τ, and obtain, sequentially, best linear estimates $\hat{x}(\tau)$ in extremely general cases which may involve time-varying $A(\cdot)$, $B(\tau)$, and $Q(\tau)$. Best linear prediction is a simple consequence of best linear estimation. The assumption that $\{x(\tau)\}$ is a wide sense vector Markov process does not appear to be very restrictive.

The basis of the analysis is the fact that best linear estimates and observations on which they are based are orthogonal to resulting errors of estimate; this and the linear structure of $\{x(\tau)\}$ permit orthogonalization of successive observations by subtracting from each observation its best linear prediction.

Section III treats a finite set of observations; Section IV, continuous processes $\{x(\tau)\}$; Section V, continuous observations; Section VI, some aspects of Gaussian $x(\tau)$.

Standard vector notation is used, with E the symbol for expectation; A, B, C, D, F, G, .. for matrices; a, b, c, ... for column vectors; and α, β, γ, ... for scalars. The transpose of a vector or matrix is denoted by the tilde. I and i represent identity matrices and column vectors of unity components, respectively. The symbol $\triangleq$ stands for "is equal by definition to." Time derivatives are always considered as derivatives to the right. To avoid undue complexity in writing, we use the short notation:

$$F'(\tau) \triangleq \frac{d}{d\tau} F(\tau)$$

$$F'(\sigma,\tau) \triangleq \frac{d}{d\sigma} F(\sigma,\tau)$$

$$\dot{F}(\sigma,\tau) \triangleq \frac{d}{d\tau} F(\sigma,\tau).$$

That is, the prime denotes the derivative of the function with respect to the first time variable, the dot the derivative with respect to the second time variable. One further special symbol is the asterisk, representing the operation:

$$*F(\tau) \triangleq F'(\tau) + F(\tau)\,B(\tau),$$

with $B(\tau)$ a specified matrix partially characterizing $\{x(\tau)\}$. By $*^2$, we represent the repeated operation

$$*^2\,F(\tau) \triangleq *\,[*F(\tau)].$$

The concept of orthogonality of two random vectors is used extensively. Vectors x and y, whether or not of equal dimension, are said to be orthogonal if, and only if, the matrix $Ex\tilde{y}$ is zero.

II. HISTORICAL NOTE

The best linear absolutely unbiased estimate of a vector x based on a finite vector y of observations linear in x has been well known for many years. The standard demonstration, showing that the estimate minimizing a quadratic form in residuals yields the minimum variance absolutely unbiased estimate of every linear combination of x, is called the Gauss-Markov Theorem; this corresponds to Conclusions D and F of Theorem 0. Theorem 1, regarding the best linear estimate of a vector x based on a vector y of observations, is presumably also well known, but the author has not found it in the literature. To the best of the author's knowledge, the first work on sequential calculation of best absolutely unbiased linear estimates was that of J. W. Follin, Jr., about 1955, in the first case considered in Section III.5.

Kolmogorov, 1941, obtained the best linear prediction of a wide sense stationary scalar process observed over a sequence of uniformly spaced times extending indefinitely into the past. Wiener, 1949, considered scalar observations, continuously or at uniform intervals, extending indefinitely into the past, on the sum of a signal process and a noise process, each with absolutely continuous spectral distribution function, and obtained the best linear estimate of the signal; convenient instrumentation gave estimates sequentially. Estimation problems involving multiple observations were treated with considerable success, but gave rise to the difficult problem of properly factoring spectral density matrices. The reader is referred to Doob, 1953, Chapter XII and Supplement, for a discussion of estimation based on scalar observations on wide sense stationary processes observed indefinitely into the past, and remarks on contributions to such problems.

Considerable effort has been devoted to more realistic and general problems related to those solved by Wiener. The basic problem studied is the linear estimation of $\mu(\tau) + \xi_1(\tau)$, given observations made continuously over an interval $[0,\tau_1]$ of

$$\eta(\tau) = \mu(\tau) + \xi_1(\tau) + \xi_2(\tau)$$

with $\mu(\tau) = \sum_{\alpha=1}^{\nu} \lambda_\alpha \varphi_\alpha(\tau)$, where the $\varphi_\alpha(\tau)$ are known functions of time, and

$\xi_1(\tau)$ and $\xi_2(\tau)$ are zero-mean processes, not necessarily stationary. Sometimes the best linear estimate of $\mu(\tau) + \xi_1(\tau)$, subject to the restriction that the estimates of $\mu(\tau)$ be absolutely unbiased, is desired; we call this the best hybrid linear estimate. Methods employed include extensions of the Wiener-Hopf integral equation, expansion of random functions in terms of a denumerable infinity of orthogonal random variables (usually by the methods of Karhunen, 1947 and Loève, 1946), Green's function, and reproducing kernel Hilbert spaces. Zadeh and Ragazzini, 1950, considered cases with $\varphi_\alpha(\tau) = \tau^{\alpha-1}$, $\xi_1(\tau)$ and $\xi_2(\tau)$ having absolutely continuous spectral distribution functions, extended the Wiener-Hopf integral equation, and expressed best hybrid linear estimates in terms of solutions of integral equations. Grenander, 1950, obtained the explicit solution for $\bar{\lambda}$ in the case $\mu(\tau) = \lambda$, $\xi_1(\tau) = 0$, $\xi_2(\tau)$ autoregressive and stationary, using Karhunen-Loève expansions. Davis, 1952, characterized the solution of cases with $\mu(\tau)$ polynomial in time in terms of characteristic values and characteristic functions appearing in the Karhunen-Loeve expansions. Somewhat similar methods were applied by Kallianpur, 1959, in cases with $\xi_1(\tau)$ uncorrelated with $\xi_2(\tau)$. Pugachev, 1960, also uses such expansions, which he calls canonic representations, to characterize estimation problems. Dolph and Woodbury, 1952, used Green's functions in cases with $\xi_1(\tau)$ and $\xi_2(\tau)$ uncorrelated and autoregressive, and generalized to some extent the results of Grenander, 1950; in more complicated cases, they express the results in terms of solutions of integral equations. Bendat, 1955, generalized the Wiener-Hopf integral equation and solved estimation problems with $\xi_1(\tau) = 0$, $\xi_2(\tau)$ having a damped exponential cosine autocorrelation function, and $\mu(\tau)$ a finite Fourier series. Shinbrot's work was similar to that of Bendat; he set $\mu(\tau) = 0$ and solved some special cases. Parzen, 1960, characterized the problem in terms of reproducing kernel Hilbert spaces. The explicit solutions of estimation problems considered in this paragraph are limited to a few very special cases, most studies leaving difficult problems involving solution of integral equations, iterative evaluation of reproducing kernel inner products, or determining an infinite sequence of characteristic values and characteristic functions.

Problems of best linear estimation in the fictional limiting case in which observations include white noise have been completely solved, even with vector observations, although frequently in a non-rigorous manner. Important work on this problem was done by Follin (Carlton and Follin, 1955), based on sequential calculation of best absolutely unbiased linear estimates. General results for the white noise case are given in Kalman and Bucy, 1961.

The methods of the present paper are much more elementary than those used in the cited literature, and lead to explicit solutions in very general situations. The key assumption of the present paper is that the random process in question is a wide sense vector Markov process. The explicit solutions found in the literature assume this and considerably more; first, scalar observation functions, and second, particular types of wide sense vector Markov processes, e.g., autoregressive processes.

III. FINITE SET OF OBSERVATIONS

In this section we consider a finite number of observations, say y_1, y_2, .. , y_ν, which are functions of the x we wish to estimate. We consider estimates x^* which are linear combinations of the observations, and wish to choose x^* to minimize the variance of estimate

$$S^* \triangleq E(x - x^*) \widetilde{(x - x^*)}.$$

An estimate x^* is said to be unbiased if and only if $E(x - x^*) = 0$. It will be shown that the x^* minimizing S^*, denoted by $\hat{x}$ with corresponding $\hat{S}$, is unbiased. An estimate x is said to be absolutely unbiased if and only if $E(x^* | x) \equiv x$ for arbitrary x. The x^* minimizing S^* among absolutely unbiased x^* is denoted by $\bar{x}$, with corresponding $\bar{S}$. It appears reasonable to consider $\hat{x}$ the "best" estimate of x, and $\bar{x}$ the best absolutely unbiased estimate of x, and to restrict the use of $\bar{x}$ as an estimate to those cases in which it is not possible to calculate $\hat{x}$.

We shall consider first $\bar{x}$, then $\hat{x}$, and compare these estimates in the case in which either can be obtained. Sequential calculation of $\hat{x}$ and $\hat{S}$ will then be developed. We shall finally consider sequential calculation of $\bar{x}$ by use of these results. In Appendix A, the best absolutely unbiased estimate of $x(\tau)$ in a non-trivial random process is considered.

III.1 The Best Absolutely Unbiased Estimate, $\bar{x}$

The theory of $\bar{x}$ is classical least squares theory, developed by Gauss, Markov, and others. We summarize the results in Theorem 0 and its corollaries. (In Theorem 0, y and v are column vectors of μ components; x, g, $x_{F,g}$ and $\bar{x}$ are column vectors of ν components. M is $\mu \times \nu$; R, H are $\mu \times \mu$; $S_{F,g}$, $\bar{S}$ are $\nu \times \nu$; F, $\bar{F}$ are $\nu \times \mu$; A is $\sigma \times \nu$; G, B are $\sigma \times \mu$. The hypotheses imply that $\mu \geq \nu$.)

Theorem 0. Given an observation vector

$$\text{(0.0.1)} \qquad y = Mx + v,$$

v having zero mean for every x and finite variance

(0.0.2) $$R \triangleq Ev\tilde{v},$$

with R and $\tilde{M}R^{-1}M$ nonsingular; M, F, and g constants.

A. The linear estimate of x,

(0.0.3) $$x_{F,g} \triangleq Fy + g$$

is <u>absolutely unbiased</u>, i.e., for all x

$$E\, x_{F,g} | x = x \ ,$$

if and only if

(0.1.1) $$g = 0, \text{ and}$$

(0.1.2) $$FM = I.$$

B. Of the estimates $x_{F,g}$ satisfying $FM = I$, the estimate which minimizes the variance of estimate

(0.0.4) $$S_{F,g} \triangleq E(x_{F,g} - x)\widetilde{(x_{F,g} - x)}$$

is

(0.0.5) $$\bar{x} \triangleq (\tilde{M}R^{-1} M)^{-1} \tilde{M}R^{-1} y \ ,$$

i.e.,

(0.0.6) $$F \triangleq (\tilde{M}R^{-1} M)^{-1} \tilde{M}R^{-1} \ ,$$

(0.0.7) $$\bar{g} \triangleq 0 \ .$$

The minimum variance of estimate is

(0.1.3) $$\bar{S} \triangleq E(\bar{x} - x)\widetilde{(\bar{x} - x)} = FR\tilde{F} = (\tilde{M}R^{-1} M)^{-1} \ .$$

The variance of estimate associated with $x_{F,g}$ is

$$S_{F,g} = \bar{S} + (F - \bar{F})\, R(\widetilde{F - \bar{F}}) + g\tilde{g}. \tag{0.1.4}$$

C. $\bar{x} - x$ is orthogonal to x and to the residuals $y - M\bar{x}$, which have mean zero and variance

$$E(y - M\bar{x})\,(\widetilde{y - M\bar{x}}) = R - M(\tilde{M}R^{-1} M)^{-1}\, \tilde{M}\ . \tag{0.1.5}$$

D. $\bar{x}$ is the value of x which minimizes

$$(\widetilde{y - Mx})\, R^{-1}\, (y - Mx)\ .$$

E. $\bar{x}$ is invariant under nonsingular linear transformation of y.

F. The minimum variance absolutely unbiased linear estimate of Ax is $(\overline{Ax}) = A\bar{x}$, for A any constant.

Proof A. From the hypotheses and definitions,

$$Ex_{F,g}\big|x = E\left[(Fy + g)\ \big|x\right] = E\left\{\left[F(Mx + v) + g\right]\big|x\right\} = FMx + g,$$

which is identically x if and only if $g = 0$, $FM = I$.

Proof B. From (0.0.4), (0.0.3), (0.0.1), (0.1.2), and (0.0.2),

$$S_{F,g} = E(Fy + g - x)\,(\widetilde{Fy + g - x}) \tag{0.2.1}$$

$$= E\left[F(Mx + v) + g - x\right]\cdot\left[\widetilde{F(Mx + v) + g - x}\right]$$

$$= E\left[Fv + g\right]\left[\widetilde{Fv + g}\right] = FEv\tilde{v}\tilde{F} + g\tilde{g} = FR\tilde{F} + g\tilde{g}\ .$$

Substituting $\bar{F}$ and $\bar{g}$ from (0.0.6), (0.0.7) into this,

$$\bar{S} = \bar{F}R\tilde{\bar{F}} = \left[(\tilde{M}R^{-1} M)^{-1}\, \tilde{M}R^{-1}\right] R \left[R^{-1} M(\tilde{M}R^{-1} M)^{-1}\right] = (\tilde{M}R^{-1} M)^{-1} \tag{0.2.2}$$

Noting from (0.0.6) and $FM = I$ that $\bar{F}R\tilde{\bar{F}} = (\tilde{M}R^{-1} M)^{-1} \tilde{M}R^{-1} \cdot R\tilde{\bar{F}}$

$$= (\tilde{M}R^{-1} M)^{-1},$$

we have $(F - \bar{F}) R \widetilde{(F - \bar{F})} = FR\tilde{F} - \bar{F}R\tilde{\bar{F}}$.

This relation, with (0.2.1) and (0.2.2) yields

$$S_{F,g} - \bar{S} = FR\tilde{F} + g\tilde{g} - \bar{F}R\tilde{\bar{F}} = (F - \bar{F}) R \widetilde{(F - \bar{F})} + g\tilde{g} ,$$

which is (0.1.4).

$\bar{S}$ is minimum over $S_{F,g}$, since

$S_{F,g} - \bar{S}$ is clearly non-negative definite.

Proof C. By (0.0.5), (0.0.6), and (0.0.1),

(0.2.6) $$\bar{x} - x = \bar{F}y - x = \bar{F}(Mx + v) - x = \bar{F}v ,$$

and further using (0.1.2) ,

(0.2.7) $$y - M\bar{x} = Mx + v - M\bar{F} (Mx + v) = (I - M\bar{F}) v .$$

Since v has mean zero for all x, by hypothesis, (0.2.7) shows that $y - M\bar{x}$ has mean zero, and (0.2.6) shows that $\bar{x} - x$ is orthogonal to x.

To prove orthogonality of $\bar{x} - x$ and $y - M\bar{x}$, we compute from (0.2.6), (0.2.7), (0.0.2), and (0.0.6),

$$E(y - M\bar{x}) \widetilde{(\bar{x} - x)} = (I - M\bar{F}) (Ev\tilde{v})\tilde{\bar{F}}$$

$$= (I - M\bar{F}) RR^{-1} M(\tilde{M}R^{-1} M)^{-1}$$

$$= (M - M) (\tilde{M}R^{-1} M)^{-1} = 0 .$$

From (0.2.7), (0.0.2), and (0.0.6),

$$E(y - M\bar{x})\widetilde{(y - M\bar{x})} = (I - MF)(Ev\tilde{v})\widetilde{(I - MF)}$$

$$= R - R\tilde{F}\tilde{M} - MFR + MFR\tilde{F}\tilde{M}$$

$$= R - M(\tilde{M}R^{-1}M)^{-1}\tilde{M} \quad .$$

Proof D. From (0.2.7) and (0.0.6),

$$\tilde{M}R^{-1}(y - M\bar{x}) = \tilde{M}R^{-1}(I - MF)v$$

$$= \tilde{M}R^{-1}\left[I - M(\tilde{M}R^{-1}M)^{-1}\tilde{M}R^{-1}\right]v$$

$$= (\tilde{M}R^{-1} - \tilde{M}R^{-1})v = 0 \quad ,$$

from which it follows that

(0.2.8) $$\widetilde{(y - Mx)}R^{-1}(y - Mx) = \widetilde{\left[y - M\bar{x} + M(\bar{x} - x)\right]} R^{-1}\left[y - M\bar{x} + M(\bar{x} - x)\right]$$

$$= \widetilde{(y - M\bar{x})}R^{-1}(y - M\bar{x})$$

$$+ \left[\widetilde{M(\bar{x} - x)}\right]R^{-1}\left[M(\bar{x} - x)\right] \quad ,$$

proving assertion D, since the scalar term

$\left[\widetilde{M(\bar{x} - x)}\right]R^{-1}\left[M(\bar{x} - x)\right]$ is clearly non-negative definite.

Proof E. We assume, without loss of generality, that the nonsingular transformation of y is homogeneous, with matrix H. Then y, M, and v are transformed into Hy, HM, Hv, and R into $E\left[(Hv)\widetilde{(Hv)}\right] = HR\tilde{H}$. Thus $\bar{x} \triangleq (\tilde{M}R^{-1}M)^{-1}\tilde{M}R^{-1}y$ is transformed into $\left[\widetilde{(HM)}(HR\tilde{H})^{-1}(HM)\right]^{-1}\widetilde{(HM)}(HR\tilde{H})^{-1}Hy$, which is $\bar{x}$ since $(HR\tilde{H})^{-1} = \tilde{H}^{-1}R^{-1}H^{-1}$.

Proof F. By conclusion A, every absolutely unbiased linear estimate of Ax is of the form Gy, with $GM = A$. Let $G = A\bar{F} + B$; since $\bar{F}M = I$, $GM = A$ implies $BM = 0$. The variance of Gy is, by (0.2.1),

$$GR\tilde{G} = (A\bar{F} + B)\, R\, \widetilde{(A\bar{F} + B)} = A\bar{F}R\tilde{\bar{F}}\tilde{A} + BR\tilde{B}, \quad \text{(0.2.9)}$$

since from (0.0.6) and $BM = 0$, one has

$$\bar{F}R\tilde{B} = (\tilde{M}R^{-1}M)^{-1}\tilde{M}R^{-1}R\tilde{B}$$
$$= (\tilde{M}R^{-1}M)^{-1}\tilde{M}\tilde{B} = 0 \quad .$$

$BR\tilde{B}$ being non-negative definite, the variance is minimized by $(\overline{Ax}) = A\bar{F}y = A\bar{x}$.

Corollary 1 to Theorem 0. If v is transformed into γv, $0 < \gamma^2 < \infty$, $\bar{F}$ is unchanged and $\bar{S}$ is changed by the factor γ^2 .

Proof of Corollary 1.

This is evident from inspection of (0.0.2), (0.0.6), and (0.1.3).

Corollary 2 to Theorem 0. If v is normally distributed, then $\bar{x} - x$ and $y - M\bar{x}$ are normally distributed and independent; $\widetilde{(y - M\bar{x})}\, R^{-1}\, (y - M\bar{x})$ is distributed as χ^2 with degrees of freedom equal to dimensionality of y minus dimensionality of x; and $\bar{x}$ is the maximum likelihood estimate of x, if R does not depend on x.

<u>Proof of Corollary 2</u>. By (0.2.6), (0.2.7), $\bar{x} - x$ and $y - M\bar{x}$ are linear combinations of v, hence normal, and are independent since orthogonal by conclusion C.

$\bar{x}$ is the maximum likelihood estimate by conclusion D, since the likelihood of the observations is proportional to $e^{-\widetilde{(y - Mx)}\, R^{-1}\, (y - Mx)/2}$.

The quadratic form of (0.2.8),

$$\tilde{v}\, R^{-1}\, v = \widetilde{(y - M\bar{x})}\, R^{-1}\, (y - M\bar{x}) + \widetilde{(\bar{x} - x)}\, (\tilde{M} R^{-1} M)\, (\bar{x} - x),$$

is the sum of squares of orthonormalized components of v, while the second term on the right is the sum of squares of orthonormalized components of $\bar{x} - x$. Thus the first term on the right is the sum of squares of orthonormalized variables, the number of which is the dimensionality of v minus the dimensionality of x. These variables are linear combinations of $y - M\bar{x}$, hence are normal, so the sum is distributed as χ^2.

<u>Corollary 3 to Theorem 0</u>. If $R = \begin{pmatrix} R_1 & & 0 \\ & R_2 & \\ & & \ddots \\ 0 & & & R_\nu \end{pmatrix}$, with $M = \begin{pmatrix} M_1 \\ M_2 \\ \vdots \\ M_\nu \end{pmatrix}$ and

$$y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_\nu \end{pmatrix}$$ corresponding partitions of M and y, then if

$\tilde{M}_\alpha R_\alpha^{-1} M_\alpha$ is nonsingular, $\alpha = 1,2,\ldots,\nu$,

(0.3.1) $$\bar{x} = \left(\sum_{\alpha=1}^{\nu} \bar{S}_\alpha^{-1}\right)^{-1} \sum_{\beta=1}^{\nu} \bar{S}_\beta^{-1} \bar{x}_\beta ,$$

(0.3.2) $$\bar{S} = \left(\sum_{\alpha=1}^{\nu} \bar{S}_\alpha^{-1}\right)^{-1}$$

where

(0.3.3) $$\bar{x}_\beta \triangleq (\tilde{M}_\beta R_\beta^{-1} M_\beta)^{-1} \tilde{M}_\beta R_\beta^{-1} y_\beta ;$$

(0.3.4) $$\bar{S}_\alpha \triangleq (\tilde{M}_\alpha R_\alpha^{-1} M_\alpha)^{-1}$$

<u>Proof</u>. Immediate, by substitution into (0.0.5), (0.1.3), since

$$\tilde{M}R^{-1} M = (\tilde{M}_1, \tilde{M}_2, \ldots, \tilde{M}_\nu) \begin{pmatrix} R_1^{-1} & & 0 \\ & R_2^{-1} & \\ 0 & & \ddots \\ & & & R^{-1} \end{pmatrix} \begin{pmatrix} M_1 \\ M_2 \\ \vdots \\ M_\nu \end{pmatrix}$$

$$= \sum_{\alpha=1}^{\nu} \tilde{M}_\alpha R_\alpha^{-1} M_\alpha ,$$

and $\tilde{M}R^{-1} y$ decomposes similarly.

The hypothesis that the vector v of errors of observation be of zero mean for every x is essential; without it, the estimate is <u>not</u> absolutely unbiased. We warn the reader of two forms of estimate often confused with

$\bar{x}$; (1) the estimate minimizing $(y - Mx^*) \widetilde{(y - Mx^*)}$, the sum of squares of the residuals; (2) the maximum likelihood estimate minimizing the weighted sum of squares of residuals when the observations are not linear in x. The first estimate is absolutely unbiased but, in general, has greater variance of estimate than $\bar{x}$; the second is in general not even unbiased.

Case III.1.a. A frequent special case is that in which errors of observation are uncorrelated, with common variance σ^2. Then R is scalar, $\bar{F} = (\tilde{M}M)^{-1} \tilde{M}$, $\bar{S} = \sigma^2 (\tilde{M}M)^{-1}$. If the errors are further assumed to be normally distributed, $(y - M\bar{x}) \widetilde{(y - M\bar{x})}$ is a sufficient statistic for σ^2.

Case III.1.b. Another frequent special case is that in which M is nonsingular, so that x and y have equal dimensions. Then $\bar{x}$ is the only absolutely unbiased estimate (T. W. Anderson, 1962), and unless R is known, the variance of estimate cannot be calculated.

Example III.1.c. Shinbrot (1956) gives the following problem, which we shall consider exhaustively: Given successive observations of $\xi + \epsilon$, ξ constant over the sequence of observations and uniformly distributed between γ_1 and $\gamma_2 > \gamma_1$, ϵ uniformly distributed between $-\lambda$ and $+\lambda > 0$, successive ϵ mutually independent and independent of ξ.

Here, we calculate $\bar{\xi}$. We have
$$y = \begin{pmatrix} \xi \\ \xi \\ \cdot \\ \cdot \\ \xi \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \cdot \\ \cdot \\ \epsilon_\mu \end{pmatrix},$$

so that $x = \xi$, $M = i_\mu$,
$$R = \begin{pmatrix} E\,\epsilon_1^2 & & 0 \\ & \ddots & \\ 0 & & E\,\epsilon_\mu^2 \end{pmatrix} = \frac{\lambda^2}{3} I_\mu .$$

Thus
$$\bar{x} = (\tilde{M}R^{-1} M)^{-1} \tilde{M}R^{-1} y = \mu^{-1} \sum_{\alpha=1}^{\mu} \eta_\alpha .$$

$$\bar{S} = (\tilde{M}R^{-1} M)^{-1} = \frac{\lambda^2}{3\mu} .$$

The best absolutely unbiased estimate of ξ is <u>nonlinear</u> in this case, being half the sum of the largest and the smallest observation. The variance of the best nonlinear estimate decreases as μ^{-2}, rather than μ^{-1}, so that the linear estimate has an efficiency tending to zero as μ increases. (Carlton, 1946).

III.2 Best Linear Estimate, $\hat{x}$

Given the first two moments of x and y, one can calculate $\hat{x}$ without requiring linearity of y in x. The following theorem is stated for zero means of x and y, but is general since x and y can be transformed to zero mean. (In Theorem 1, y is a column vector of μ components; x, $x_{F,g}$, g, and $\hat{x}$ are column vectors of ν components. F and $\hat{F}$ are $\nu \times \mu$; $S_{F,g}$ and $\hat{S}$ are $\nu \times \nu$; A is $\sigma \times \nu$; $(Ey\tilde{y})$ and H are $\mu \times \mu$; $(Ex\tilde{y})$ is $\nu \times \mu$.)

Theorem 1. Let x and y be zero mean random variables with finite second moments,
$Ey\tilde{y}$ nonsingular, with F and g constant.

A. The linear estimate of x,

$$x_{F,g} \triangleq Fy + g \tag{1.0.1}$$

which minimizes the variance of estimate

$$S_{F,g} \triangleq E(x_{F,g} - x)\widetilde{(x_{F,g} - x)} \tag{1.0.2}$$

is

$$\hat{x} \triangleq \hat{F}y , \tag{1.0.3}$$

with

$$\hat{F} \triangleq (Ex\tilde{y})\,(Ey\tilde{y})^{-1} . \tag{1.0.4}$$

$\hat{x}$ is an unbiased estimate of x, i.e.,

$$E\hat{x} = Ex = 0. \tag{1.1.1}$$

The minimum variance is

$$\hat{S} \triangleq S_{\hat{F},\hat{g}} = Ex\tilde{x} - Ex\tilde{y}\,(Ey\tilde{y})^{-1}\,Ey\tilde{x}, \tag{1.1.2}$$

and the variance of any linear estimate is

(1.1.3) $$S_{F,g} = \hat{S} + (F - \hat{F})\, Ey\tilde{y}\, \widetilde{(F - \hat{F})} + g\tilde{g} \,.$$

B. $x - \hat{x}$ is orthogonal to y and to $\hat{x}$.

C. $\hat{x}$ is invariant under nonsingular linear transformation of y.

D. The minimum variance linear estimate of Ax is $A\hat{x}$, for A any constant.

Proof A. From (1.0.2) and (1.0.1), and hypothesized zero means of x and y,

(1.2.1) $$S_{F,g} = E(Fy + g - x)\,\widetilde{(Fy + g - x)} = E(Fy - x)\,\widetilde{(Fy - x)} + g\tilde{g}$$

$$= F(Ey\tilde{y})\,\tilde{F} - F\,(Ey\tilde{x}) - (Ex\tilde{y})\,\tilde{F} + Ex\tilde{x} + g\tilde{g}.$$

Substituting (1.0.3) and (1.0.4), i.e., $\hat{g} = 0$, $\hat{F} = (Ex\tilde{y})\,(Ey\tilde{y})^{-1}$,

(1.2.2) $$\hat{S} = Ex\tilde{x} - (Ex\tilde{y})\,(Ey\tilde{y})^{-1}\,(Ey\tilde{x}).$$

Subtracting (1.2.2) from (1.2.1), again using (1.0.4),

(1.2.3) $$S_{F,g} - \hat{S} = g\tilde{g} + F(Ey\tilde{y})\,\tilde{F} - F\,(Ey\tilde{y})\,\tilde{\hat{F}} - \hat{F}(Ey\tilde{y})\,\tilde{F} + \hat{F}(Ey\tilde{y})\,\tilde{\hat{F}}$$

$$= g\tilde{g} + (F - \hat{F})\,(Ey\tilde{y})\,\widetilde{(F - \hat{F})}\,.$$

Since (1.2.3) is clearly non-negative definite, $\hat{x}$ minimizes the variance.

Since y has zero mean, (1.0.3) shows that $\hat{x}$ has zero mean, i.e., is unbiased.

Proof B. From (1.0.3) and (1.0.4),

(1.2.4) $$E(x - \hat{x})\,\tilde{y} = Ex\tilde{y} - E\hat{F}y\tilde{y}$$

$$= Ex\tilde{y} - (Ex\tilde{y})\,(Ey\tilde{y})^{-1}\,(Ey\tilde{y}) = 0.$$

Thus $x - \hat{x}$ is orthogonal to y and, by (1.0.3), to $\hat{x}$.

Proof C. Without loss of generality, let the transformed y be Hy, H nonsingular by hypothesis.

Then, from (1.0.4), $\hat{F}$ is transformed into $\left[Ex\ (\widetilde{Hy})\right]\ \left[E(Hy)\ (\widetilde{Hy})\right]^{-1}$

$= (Ex\tilde{y})\ \tilde{H}\left[H(Ey\tilde{y})\tilde{H}\right]^{-1} = \hat{F}\ H^{-1}$, and from (1.0.3), $\hat{x}$ is transformed into $(\hat{F}H^{-1})\ (Hy) = \hat{F}y = \hat{x}$.

Proof D. From (1.0.3) and (1.0.4),

$$(\widehat{Ax}) = (EAx\tilde{y})\ (Ey\tilde{y})^{-1}\ y = A(Ex\tilde{y})\ (Ey\tilde{y})^{-1}\ y = A\hat{x}.$$

Case III.2.a. Let

(1.3.1) $$y = Mx + v,$$

v having zero mean and variance

(1.3.2) $$R \triangleq Ev\tilde{v},$$

with M constant. Define

(1.3.3) $$S \triangleq Ex\tilde{x}$$

(1.3.4) $$T \triangleq Ex\tilde{v}.$$

Then

(1.4.1) $$\hat{F} = (S\tilde{M} + T)\ (MS\tilde{M} + MT + \tilde{T}\tilde{M} + R)^{-1}\ ;$$

(1.4.2) $$\hat{S} = S - (S\tilde{M} + T)\ (MS\tilde{M} + MT + \tilde{T}\tilde{M} + R)^{-1}\ (MS + \tilde{T}).$$

These results follow from evaluation of $Ex\tilde{y}$ as $S\tilde{M} + T$, $Ey\tilde{y}$ as $MS\tilde{M} + MT + \tilde{T}\tilde{M} + R$.

III.3 Comparison of $\bar{x}$ and x

Suppose we have observations $y = Mx + v$, with hypotheses of Theorem 0 and Theorem 1 simultaneously satisfied, with S nonsingular. Thus in the special case III.2.a., $\hat{S} = S - S\tilde{M}\ (MS\tilde{M} + R)^{-1}\ MS$, since $T = 0$, while from Theorem 0, $\bar{S} = (\tilde{M}R^{-1}\ M)^{-1}$. To compare these variances of estimate, we use the following lemma.

Lemma A. For any real matrix K,

$$I - K(\tilde{K}\ K + I)^{-1}\ \tilde{K} \equiv (K\ \tilde{K} + I)^{-1}.$$

Proof. K being a real matrix, the indicated inverses surely exist. The identity can be verified by power series expansion, or developed from the evident identity

$$K\ \tilde{K}\ K + K \equiv K\ \tilde{K}\ K + K.$$

Factoring each side, we obtain

$$K(\tilde{K}\ K + I) \equiv (K\ \tilde{K} + I)K.$$

From this identity, the desired identity is obtained by post-multiplying each side by $(\tilde{K}\ K + I)^{-1}\ \tilde{K}(K\ \tilde{K} + I)$, adding $K\ \tilde{K} + I$ to each side, and pre-and-post-multiplying by $(K\ \tilde{K} + I)^{-1}$. (The identity is valid for any K such that the indicated inverses exist.)

Using Lemma A with $S^{1/2}\ \tilde{M}R^{-1/2}$ playing the role of K, we have

$$(\bar{S}^{-1} + S^{-1})^{-1} = (\tilde{M}R^{-1}\ M + S^{-1})^{-1}$$

$$= S^{1/2}\ (S^{1/2}\ \tilde{M}R^{-1}\ MS^{1/2} + I)^{-1}\ S^{1/2}$$

$$= S^{1/2} \left\{ I - S^{1/2}\ \tilde{M}R^{-1/2}(R^{-1/2}\ MS\tilde{M}\ R^{-1/2} + I)^{-1}\ R^{-1/2}\ MS^{1/2} \right\} S^{1/2}$$

$$= S - S\tilde{M}\ (MS\tilde{M} + R)^{-1}\ MS = \hat{S}.$$

In short, the inverse of $\hat{S}$ is the sum of the inverses of S and of $\bar{S}$.

Consider a sequence of values of S^{-1} tending to the limit zero. The corresponding sequence of values of $\hat{S}$ must tend to $\bar{S}$ (in fact, it can be shown that $\bar{S} - \hat{S} \leq \epsilon I$, if $S^{-1} \leq \epsilon \bar{S}^{-2}$). From Theorem 1, Conclusion A, (1.1.3), since $\bar{F}y = \bar{x}$ and $\hat{F}y = \hat{x}$, we have that

$$\bar{S} = \hat{S} + E(\bar{x} - \hat{x})\,(\widetilde{\bar{x} - \hat{x}}),$$

so that as $\hat{S}$ tends to $\bar{S}$, $\hat{x}$ must tend to $\bar{x}$ in quadratic mean. Thus $\bar{x}$ can be considered a limiting value of $\hat{x}$.

To show more directly the approach of $\hat{x}$ to $\bar{x}$ as S^{-1} tends to the limit zero, we can modify the identity of Lemma A as follows:

$$(K\tilde{K} + I)^{-1} K = K - K(\tilde{K}K + I)^{-1}\tilde{K}K$$

$$= K\{I - (\tilde{K}K + I)^{-1}\tilde{K}K\} = K(\tilde{K}K + I)^{-1}\{(\tilde{K}K + I) - \tilde{K}K)\}$$

$$= K(\tilde{K}K + I)^{-1}.$$

Using this identity, we have

$$\hat{F} = S\tilde{M}(MS\tilde{M} + R)^{-1}$$

$$= S^{1/2}(S^{1/2}\tilde{M}R^{-1/2})\left\{(R^{-1/2}MS^{1/2})(S^{1/2}\tilde{M}R^{-1/2}) + I\right\}^{-1} R^{-1/2}$$

$$= S^{1/2}(S^{1/2}\tilde{M}R^{-1}MS^{1/2} + I)^{-1} S^{1/2}\tilde{M}R^{-1/2}R^{-1/2}$$

$$= (\tilde{M}R^{-1}M + S^{-1})^{-1}\tilde{M}R^{-1},$$

which clearly tends to $\bar{F} = (\tilde{M}R^{-1}M)^{-1}\tilde{M}R^{-1}$ as S^{-1} tends to zero, so that $\hat{x} = \hat{F}y$ tends to $\bar{x} = \bar{F}y$.

III.4 Sequential Calculation of $(\hat{x}, \hat{S})$

With observations linear in x and the structure of x suitably linear, it is possible to exploit the orthogonality of $x - \hat{x}$ to $\hat{x}$ and calculate $\hat{x}$ and $\hat{S}$ sequentially. We denote by $x_1, \cdots, x_\nu$ the values of the process $x(\tau)$ at the epochs of the successive observations $y_1, y_2, \cdots, y_\nu$. The successive values of x are assumed to satisfy the relation

$$x_\alpha = L_{\alpha,\alpha-1}x_{\alpha-1} + v_\alpha,$$

with the $L_{\alpha,\alpha-1}$ specified matrices, v_α random disturbances. This relation implies that x_α is a complete state vector, that the x-process is a vector Markov process in the wide sense. The following theorem shows how one can predict $\hat{X}$ and $\hat{S}$, how transforming the observations by subtraction of predicted value of observations orthogonalizes them, and how to estimate $\hat{X}(\tau)$ and its variance sequentially. It is assumed that one has a best estimate of x_0, and known variance of estimate, at some epoch prior to the first observation y_1. (In Theorem 2, y_α and z_α are column vectors with μ_α components; x_α, v_α, $\hat{x}_{\alpha,\beta}$, and u_α are column vectors of σ_α components. (Typically, but not necessarily, $\sigma_1 = \sigma_2 = \cdots = \sigma_\nu$.) Letting μ represent $\sum_{\alpha=1}^{\nu} \mu_\alpha$, and σ represent $\sum_{\alpha=1}^{\nu} \sigma_\alpha$, one has that x and v are $\sigma \times 1$; y is $\mu \times 1$; $L_{\beta,\alpha}$ is $\sigma_\beta \times \sigma_\alpha$; $(Ey\tilde{y})$ is $\mu \times \mu$; A_α is $\mu_\alpha \times \sigma_\alpha$; $S_{\alpha,\beta}$ is $\sigma_\alpha \times \sigma_\alpha$; and $V_{\alpha,\beta}$ is $\sigma_\alpha \times \sigma_\beta$.)

<u>Theorem 2</u>. Let the random variables

$$x \triangleq x_{(\nu)} \triangleq (\widetilde{\tilde{x}_1, \tilde{x}_2, \cdots, \tilde{x}_\nu}),$$

$$y \triangleq y_{(\nu)} \triangleq (\widetilde{\tilde{y}_1, \tilde{y}_2, \cdots, \tilde{y}_\nu}),$$

$$v \triangleq v_{(\nu)} \triangleq (\widetilde{\tilde{v}_1, \tilde{v}_2, \cdots, \tilde{v}_\nu})$$

satisfy the relations, for $\alpha = 1, 2, \cdots, \nu$,

(2.0.1) $$x_\alpha = L_{\alpha,\alpha-1}\, x_{\alpha-1} + v_\alpha,$$

(2.0.2) $$y_\alpha = A_\alpha x_\alpha,$$

with $L_{\alpha,\alpha-1}$ and A_α constants; v_α mutually orthogonal zero mean random variables; second moments of v and x finite; $Ey\tilde{y}$ nonsingular. Denote by $\hat{x}_{\alpha,\beta}$ the minimum variance linear estimate of x_α based on $y_{(\beta)} \triangleq \widetilde{(\tilde{y}_1, \tilde{y}_2, \cdots, \tilde{y}_\beta)}$, and assume v orthogonal to $\hat{x}_{0,0}$ and to x_0.

(2.0.3) $$L_{\beta,\alpha} \triangleq L_{\beta,\beta-1} L_{\beta-1,\beta-2} \cdots L_{\alpha+1,\alpha}, \quad (\alpha < \beta)$$

(2.0.4) $$S_{\alpha,\beta} \triangleq E(x_\alpha - \hat{x}_{\alpha,\beta}) \widetilde{(x_\alpha - \hat{x}_{\alpha,\beta})}$$

(2.0.5) $$V_{\alpha,\beta} \triangleq E(x_\alpha - \hat{x}_{\alpha,\beta}) \widetilde{(x_\beta - \hat{x}_{\beta,\beta})}$$

(2.0.6) $$u_\alpha \triangleq x_\alpha - \hat{x}_{\alpha,\alpha-1}$$

(2.0.7) $$\hat{u}_\alpha \triangleq \hat{u}_{\alpha,\alpha}$$

(2.0.8) $$z_\alpha \triangleq A_\alpha u_\alpha \quad .$$

<u>Conclusion A</u>. For $\alpha < \beta$,

(2.1.1) $$\hat{x}_{\beta,\alpha} = L_{\beta,\alpha} \hat{x}_{\alpha,\alpha}$$

(2.1.2) $$S_{\beta,\alpha} = L_{\beta,\alpha} S_{\alpha,\alpha} \tilde{L}_{\beta,\alpha} + \sum_{\gamma=\alpha+1}^{\beta} L_{\beta,\gamma} (E v_\gamma \tilde{v}_\gamma) \tilde{L}_{\beta,\gamma} \ .$$

<u>Conclusion B</u>. The linear transformation of y into z, defined by (2.0.6) and (2.0.8) is nonsingular. The vectors $z_1, z_2 \cdots, z_\nu$ are mutually orthogonal, with mean zero and variance

(2.1.3) $$E z_\alpha \tilde{z}_\alpha = A_\alpha S_{\alpha,\alpha-1} \tilde{A}_\alpha \quad .$$

For $\alpha < \beta$, z_α is orthogonal to u_β.

Conclusion C. Estimates $\hat{x}_{1,1}, \hat{x}_{2,2}, \cdots, \hat{x}_{\nu,\nu}$ and their variances can be obtained sequentially by use of (2.1.1) and (2.1.2) with $\beta = \alpha + 1$, and

(2.1.4) $$\hat{x}_{\alpha,\alpha} = \hat{x}_{\alpha,\alpha-1} + \hat{u}_\alpha ,$$

(2.1.5) $$\hat{u}_\alpha = (Eu_\alpha \tilde{z}_\alpha)(Ez_\alpha \tilde{z}_\alpha)^{-1} z_\alpha ,$$

(2.1.6) $$S_{\alpha,\alpha} = S_{\alpha,\alpha-1} - S_{\alpha,\alpha-1} \tilde{A}_\alpha (A_\alpha S_{\alpha,\alpha-1} \tilde{A}_\alpha)^{-1} A_\alpha S_{\alpha,\alpha-1} .$$

Conclusion D. Estimates $\hat{x}_{\alpha,\alpha+1}, \hat{x}_{\alpha,\alpha+2}, \cdots, \hat{x}_{\alpha,\nu}$ and associated variances can be calculated sequentially using the additional relations, for $\alpha < \beta$,

(2.1.7) $$\hat{x}_{\alpha,\beta} = \hat{x}_{\alpha,\beta-1} + V_{\alpha,\beta-1} \tilde{L}_{\beta,\beta-1} \tilde{A}_\beta (Ez_\beta \tilde{z}_\beta)^{-1} z_\beta ;$$

(2.1.8) $$V_{\alpha,\beta} = V_{\alpha,\beta-1} \tilde{L}_{\beta,\beta-1} \left[I - \tilde{A}_\beta (Ez_\beta \tilde{z}_\beta)^{-1} A_\beta S_{\beta,\beta-1} \right] .$$

(2.1.9) $$S_{\alpha,\beta} = S_{\alpha,\beta-1} - V_{\alpha,\beta-1} \tilde{L}_{\beta,\beta-1} \tilde{A}_\beta (Ez_\beta \tilde{z}_\beta)^{-1} A_\beta L_{\beta,\beta-1} \tilde{V}_{\alpha,\beta-1} .$$

Proof A. The hypotheses of Theorem 1 are satisfied, assuming, without loss of generality, that $\hat{x}_{0,0} = Ex_0 = 0$.

Repeatedly using (2.0.1) and (2.0.3),

(2.2.1) $$x_\beta = v_\beta + L_{\beta,\beta-1} x_{\beta-1} = v_\beta + L_{\beta,\beta-1} v_{\beta-1} + L_{\beta,\beta-2} x_{\beta-2}$$
$$= v_\beta + L_{\beta,\beta-1} v_{\beta-1} + L_{\beta,\beta-2} v_{\beta-2} + \cdots + L_{\beta,\alpha+1} v_{\alpha+1} + L_{\beta,\alpha} x_\alpha .$$

By Conclusion D of Theorem 1, $\hat{x}_\beta = \hat{v}_\beta + \cdots + L_{\beta,\alpha}\hat{x}_\alpha$. By hypothesis, $v_\beta, \cdots, v_{\alpha+1}$ have mean zero and are orthogonal to x_0 and $v_1, v_2, \cdots, v_\alpha$ hence from (2.0.1) and (2.0.2) to y_α, so that $\hat{x}_\beta = L_{\beta,\alpha}\hat{x}_\alpha$.

From (2.0.4), (2.1.1), and (2.2.1), using orthogonality of the components of v,

$$S_{\beta,\alpha} \triangleq E(x_\beta - \hat{x}_{\beta,\alpha})\,\widetilde{(x_\beta - \hat{x}_{\beta,\alpha})}$$

$$= Ev_\beta\tilde{v}_\beta + L_{\beta,\beta-1}Ev_{\beta-1}\tilde{v}_{\beta-1}\tilde{L}_{\beta,\beta-1} + \cdots$$

$$+ L_{\beta,\alpha+1}Ev_{\alpha+1}\tilde{v}_{\alpha+1}\tilde{L}_{\beta,\alpha+1} + L_{\beta,\alpha}S_{\alpha,\alpha}\tilde{L}_{\beta,\alpha},$$

which is (2.1.2).

Proof B. From (2.0.8), (2.0.6) and (2.0.2).

(2.2.2)
$$z_\alpha = A_\alpha(x_\alpha - \hat{x}_{\alpha,\alpha-1})$$

$$= y_\alpha - A_\alpha\hat{x}_{\alpha,\alpha-1},$$

with $\hat{x}_{\alpha,\alpha-1}$ by definition a linear combination of $y_1, y_2, \cdots, y_{\alpha-1}$. Thus the transformation matrix is triangular with unity diagonal elements, hence nonsingular. $Ez_\alpha = 0$ since by Theorem 1, $\hat{x}_{\alpha,\alpha-1}$ is an unbiased estimate of x_α.

From (2.2.2) and (2.0.4)

$$Ez_\alpha\tilde{z}_\alpha = A_\alpha E(x_\alpha-\hat{x}_{\alpha,\alpha-1})\,\widetilde{(x_\alpha-\hat{x}_{\alpha,\alpha-1})}\,\tilde{A}_\alpha = A_\alpha S_{\alpha,\alpha-1}\tilde{A}_\alpha,$$

which is (2.1.3).

From (2.0.6), (2.0.1), and (2.1.1),

(2.2.3) $$u_\beta \triangleq x_\beta - \hat{x}_{\beta,\beta-1} = L_{\beta,\beta-1}x_{\beta-1} + v_\beta - L_{\beta,\beta-1}\hat{x}_{\beta-1,\beta-1}$$

$$= L_{\beta,\beta-1}(x_{\beta-1} - \hat{x}_{\beta-1,\beta-1}) + v_\beta \; .$$

By Theorem 1, $x_{\beta-1} - \hat{x}_{\beta-1,\beta-1}$ is orthogonal to $y_{(\beta-1)}$, and v_β was shown above to be orthogonal to $y_{(\beta-1)}$. Thus u_β is orthogonal to $y_{(\beta-1)}$, hence to $z_{(\beta-1)}$ which is a linear combination of $y_{(\beta-1)}$. Thus u_β is orthogonal to z_α, $a < \beta$, and since $z_\beta \triangleq A_\beta u_\beta$, the components of z are mutually orthogonal.

<u>Proof</u> <u>C</u>. (2.1.4) follows from Theorem 1, since

$$x_\alpha \equiv \hat{x}_{\alpha,\alpha-1} + (x_\alpha - \hat{x}_{\alpha,\alpha-1}) \triangleq \hat{x}_{\alpha,\alpha-1} + u_\alpha \; .$$

(2.1.5) is obtained from Theorem 1 and Conclusion B:

$$\hat{u}_\alpha \triangleq \hat{u}_{\alpha,\alpha} = (Eu_\alpha\tilde{z}_{(\alpha)})\,(Ez_{(\alpha)}\tilde{z}_{(\alpha)})^{-1}\,z_{(\alpha)}$$

$$= Eu_\alpha\tilde{z}_\alpha\,(Ez_\alpha\tilde{z}_\alpha)^{-1}\,z_\alpha \; .$$

(2.1.6) is obtained by using (1.1.2) with $x \triangleq x_\alpha - \hat{x}_{\alpha,\alpha-1} = u_\alpha$, $y \triangleq z_\alpha$, i.e.,

$$S_{\alpha,\alpha} = S_{\alpha,\alpha-1} - (Eu_\alpha\tilde{z}_\alpha)\,(Ez_\alpha\tilde{z}_\alpha^{-1})\,(Ez_\alpha\tilde{u}_\alpha) \; ,$$

with $Ez_\alpha\tilde{z}_\alpha = A_\alpha S_{\alpha,\alpha-1}\tilde{A}_\alpha$ by (2.1.3), and

$$Eu_\alpha\tilde{z}_\alpha = Eu_\alpha\tilde{u}_\alpha\tilde{A}_\alpha = S_{\alpha,\alpha-1}\tilde{A}_\alpha$$

by (2.0.8), (2.0.6), and (2.0.4).

Proof D. From (2.0.1), (2.0.2), and hypotheses, v_β is orthogonal to x_α, $x_{\alpha,\beta-1}$. Using this orthogonality, (2.0.8), (2.0.6), and (2.0.5),

(2.2.5)
$$E(x_\alpha - \hat{x}_{\alpha,\beta-1})\,\tilde{u}_\beta$$
$$= E(x_\alpha - \hat{x}_{\alpha,\beta-1})\left[\widetilde{L_{\beta,\beta-1}(x_{\beta-1} - \hat{x}_{\beta-1,\beta-1}) + v_\beta}\right]$$
$$= E(x_\alpha - \hat{x}_{\alpha,\beta-1})\,(\widetilde{x_{\beta-1} - \hat{x}_{\beta-1,\beta-1}})\,\tilde{L}_{\beta,\beta-1}$$
$$= V_{\alpha,\beta-1}\,\tilde{L}_{\beta,\beta-1}\ .$$

Writing $x_\alpha \equiv \hat{x}_{\alpha,\beta-1} + (x_\alpha - \hat{x}_{\alpha,\beta-1})$, using (2.2.5) and orthogonality of components of $z_{(\beta)}$, Theorem 1 gives

$$\hat{x}_{\alpha,\beta} = \hat{x}_{\alpha,\beta-1}$$
$$+ \left[E(x_\alpha - \hat{x}_{\alpha,\beta-1})\tilde{z}_{(\beta)}\right]\left[Ez_{(\beta)}\tilde{z}_{(\beta)}\right]^{-1} z_{(\beta)}$$
$$= \hat{x}_{\alpha,\beta-1} + V_{\alpha,\beta-1}\,\tilde{L}_{\beta,\beta-1}\tilde{A}_\beta\,(Ez_\beta\tilde{z}_\beta)^{-1}\,z_\beta\ ,$$

which is (2.1.7).

To verify (2.1.8), write (2.1.7) as

(2.2.6)
$$x_\alpha - \hat{x}_{\alpha,\beta} = (x_\alpha - \hat{x}_{\alpha,\beta-1})$$
$$- V_{\alpha,\beta-1}\,\tilde{L}_{\beta,\beta-1}\tilde{A}_\beta(Ez_\beta\tilde{z}_\beta)^{-1}\,z_\beta\ .$$

From (2.0.6), (2.0.7), (2.1.5),

(2.2.7)
$$x_\beta - \hat{x}_{\beta,\beta} = u_\beta - \hat{u}_\beta = \left[I - (Eu_\beta\tilde{z}_\beta)\,(Ez_\beta\tilde{z}_\beta)^{-1}\,A_\beta\right]\,u_\beta\ .$$

Taking the expectation of (2.2.6) times the transpose of (2.2.7), and using (2.2.5) and (2.0.5),

(2.2.8)
$$V_{\alpha,\beta} = \left\{V_{\alpha,\beta-1}\tilde{L}_{\beta,\beta-1} - V_{\alpha,\beta-1}\tilde{L}_{\beta,\beta-1}\tilde{A}_{\beta}\ (Ez_{\beta}\tilde{z}_{\beta})^{-1}\ (Ez_{\beta}\tilde{u}_{\beta})\right\} \cdot \left\{I - \tilde{A}_{\beta}\ (Ez_{\beta}\tilde{z}_{\beta})^{-1}(Ez_{\beta}\tilde{u}_{\beta})\right\} .$$

From the proof of C, $Ez_{\beta}\tilde{u}_{\beta} = A_{\beta}S_{\beta,\beta-1}$, while from (2.1.3), $Ez_{\beta}\tilde{z}_{\beta} = A_{\beta}S_{\beta,\beta-1}\tilde{A}_{\beta}$. Substituting these results in (2.2.8) yields (2.1.8).

Taking expectation of (2.2.6) times its tranpose, using (2.0.4), (2.2.5), (2.0.8), one obtains (2.1.9).

The theorem has been stated, preparing for later application, in terms of a decomposition of the column vector y into column vector components. In application to a finite set of observations, however, one could decompose y into scalar components, simplifying the computation by reducing to a scalar the matrix to be inverted.

<u>Example</u> <u>III.4.a</u>. (Prediction) An example from Wiener (1949) supposes two periodically observed signals with spectral densities $\Phi_{11}(\omega) = \Phi_{22}(\omega) = 1$, $\Phi_{21}(\omega) = \epsilon\ e^{i\omega}$, $\Phi_{12}(\omega) = \epsilon\ e^{-i\omega}$, with $\epsilon^2 < 1$. We wish to predict each signal at future times of observation. Let the first signal be $\xi_1 = \eta_1$, the second signal $\xi_2 = \eta_2$. Then $\begin{pmatrix}\xi_1\\ \xi_2\end{pmatrix} = x = y = \begin{pmatrix}\eta_1\\ \eta_2\end{pmatrix}$. The given spectral densities imply $x_{\alpha} = L_{\alpha,\alpha-1}\ x_{\alpha-1} + v_{\alpha}$,

with $L_{\alpha,\alpha-1} = \begin{pmatrix}0 & \epsilon\\ 0 & 0\end{pmatrix}$, $Ev_{\alpha}\tilde{v}_{\alpha} = \begin{pmatrix}1-\epsilon^2 & 0\\ 0 & 1\end{pmatrix}$.

For $\beta > \alpha$, $\hat{x}_{\beta,\alpha} = L_{\beta,\alpha}\ \hat{x}_{\alpha,\alpha} = \begin{pmatrix}0 & \epsilon\\ 0 & 0\end{pmatrix}^{\beta-\alpha} \begin{pmatrix}\eta_1\\ \eta_2\end{pmatrix}_{\alpha}$,

$$S_{\beta,\alpha} = L_{\beta,\alpha}\ S_{\alpha,\alpha}\ \tilde{L}_{\beta,\alpha} + \sum_{\gamma=\alpha+1} L_{\beta,\gamma}\ (Ev_{\gamma}\tilde{v}_{\gamma})\ L_{\beta,\gamma} .$$

Since $\begin{pmatrix} 0 & \varepsilon \\ 0 & 0 \end{pmatrix}^{\beta-\alpha} = 0, \ \beta-\alpha > 1; \ L_{\beta,\beta} = I$, and $S_{\alpha,\alpha} = 0$,

$$\hat{x}_{\alpha+1,\alpha} = \begin{pmatrix} \varepsilon(\eta_2)\ \alpha \\ 0 \end{pmatrix}, \ S_{\alpha+1,\alpha} = \begin{pmatrix} 1-\varepsilon^2 & 0 \\ 0 & 1 \end{pmatrix} ;$$

and for $\beta > \alpha+1$,

$$\hat{x}_{\beta,\alpha} = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

$$S_{\beta,\alpha} = \begin{pmatrix} 0 & \varepsilon \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 1-\varepsilon^2 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 0 \\ \varepsilon & 0 \end{pmatrix}$$

$$+ \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1-\varepsilon^2 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

$$= \begin{pmatrix} \varepsilon^2 & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} 1-\varepsilon^2 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Example III.4.b. Here we consider example III.1.c., and calculate ξ sequentially. Since ξ is constant, we denote by $\hat{\xi}_0, \hat{\xi}_1, \cdots, \hat{\xi}_\mu$ the best estimate of ξ based on no observation, one observation, $\cdots$, μ observations, and since successive ϵ_α are independent, we need not compute $\hat{\epsilon}_\alpha$ nor its variance. Assuming only the information presented, we have

$$\hat{\xi}_0 = E\xi = (\gamma_2 + \gamma_1)/2.$$

$E(\xi - E\xi)^2 = (\gamma_2 - \gamma_1)^2/12 \triangleq \sigma_0$, the variance of estimate of $\hat{\xi}_0$.

$$E\ \epsilon^2 = \lambda^2/3 \triangleq \beta.$$

$$\hat{\xi}_1 = \hat{\xi}_0 + \sigma_0\ (\sigma_0 + \beta)^{-1}\ (\eta_1 - \hat{\xi}_0).$$

Letting σ_α represent the variance of estimate of $\hat{\xi}_\alpha$,

$$\sigma_1 = \sigma_0 - \sigma_0(\sigma_0 + \beta)^{-1} \sigma_0 = \sigma_0 \beta (\sigma_0 + \beta)^{-1}.$$

$$\hat{\xi}_{\alpha+1} = \hat{\xi}_\alpha + \sigma_\alpha (\sigma_\alpha + \beta)^{-1} (\eta_{\alpha+1} - \hat{\xi}_\alpha) .$$

$$\sigma_{\alpha+1} = \sigma_\alpha - \sigma_\alpha(\sigma_\alpha + \beta)^{-1} \sigma_\alpha$$

$$= \sigma_\alpha\beta (\sigma_\alpha + \beta)^{-1} = \sigma_0\beta \left[(\alpha+1)\sigma_0 + \beta\right]^{-1} ,$$

the final expression the result of a trivial induction. We recall from example III.1.c. that the variance of $\bar{\xi}$ is $\lambda^2/3\mu = \beta/\mu$, whereas the variance of $\hat{x}$ is here shown to be $\sigma_\mu = \sigma_0\beta \left[\mu \sigma_0 + \beta\right]^{-1} = \left[1 + \beta/\mu\sigma_0\right]^{-1} \bar{S}$. It is interesting to notice that $\hat{\xi}_0$, the best linear estimate with no observation, is superior to $\bar{\xi}$ based on μ observations, when $\mu < \beta/\sigma_0 = 4\lambda^2/(\gamma_2 - \gamma_1)^2$.

It is also interesting to note that the best nonlinear estimate of ξ, which is far superior to $\hat{\xi}$ when μ is large, can be obtained sequentially. We state without proof that after the α^{th} observation, ξ is uniformly distributed over an interval $\left[\lambda_{1,\alpha}, \lambda_{2,\alpha}\right]$, with $\left[\lambda_{1,0}, \lambda_{2,0}\right] = \left[\gamma_1, \gamma_2\right]$,

$$\lambda_{1,\alpha+1} = \max (\lambda_{1,\alpha}, \eta_{\alpha+1} - \lambda)$$

$$\lambda_{2,\alpha+1} = \min (\lambda_{2,\alpha}, \eta_{\alpha+1} + \lambda).$$

The best nonlinear estimate of ξ after the α^{th} observation is $(\lambda_{1,\alpha} + \lambda_{2,\alpha})/2$, with variance of estimate $(\lambda_{2,\alpha} - \lambda_{1,\alpha})^2/12$.

III.5. <u>Sequential Calculation of $\bar{x}$</u>. Sequential calculation of $\bar{x}$ will be desirable in case one wishes to obtain $\bar{x}$ first on the basis of one set of observations, then on an enlarged set of observations, etc. It is tempting to suppose that $\bar{x}$ can be calculated sequentially by using a preliminary calculation of $\bar{x}$ for initial estimates in Theorem 2, then calculating the resulting $\hat{x}$. This supposition is not correct in general, but in view of the fact that $\bar{x}$ can be considered a limiting value of $\hat{x}$ (cf. Section III.3), the technique can be used in either of two cases: a) the errors of observation of the various sets of observations are mutually orthogonal; or, b) the initial calculation of $\bar{x}$ is based on a set of observations with nonsingular M. While these assertions are intuitively not difficult to understand, the proofs are somewhat cumbersome.

In case (a), consider two sets of observations:

$$y_1 = M_1 x + v_1$$

$$y_2 = M_2 x + v_2,$$

with $R = E \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} (\tilde{v}_1, \tilde{v}_2) = \begin{pmatrix} R_1 & 0 \\ 0 & R_2 \end{pmatrix}$.

Then $\bar{S} = \left[(\tilde{M}_1, \tilde{M}_2) \; R^{-1} \begin{pmatrix} M_1 \\ M_2 \end{pmatrix} \right]^{-1} = (\tilde{M}_1 R_1 M_1 + \tilde{M}_2 R_2 M_2)^{-1}$,

and the variance of $\bar{x}_1$, based on y_1 alone, is $\bar{S}_1 = (\tilde{M}_1 R_1 M_1)^{-1}$. Calculating $\hat{x}$ based on y_2, with initial estimates and variances $\bar{x}_1$ and $\bar{S}_1$, we have from example III.2.a,

$$\hat{S} = \bar{S}_1 - \bar{S}_1 \tilde{M}_2 \,(M_2 \bar{S}_1 \tilde{M}_2 + R_2)^{-1} \, M_2 \bar{S}_1$$

$$= \bar{S}_1^{\,1/2} \left\{ I - \bar{S}_1^{\,1/2} \, \tilde{M}_2 R_2^{-1/2} \, (R_2^{-1/2} \, M_2 \bar{S}_1 \, \tilde{M}_2 R_2^{-1/2} + I)^{-1} \, R_2^{-1/2} \, M_2 \bar{S}_1^{\,1/2} \right\} \bar{S}_1^{\,1/2},$$

by evidently justified manipulations, $\bar{S}_1$ and R_2 being positive definite. Applying Lemma A, with $K = \bar{S}_1^{\,1/2} \, \tilde{M}_2 R_2^{-1/2}$, the above relation gives

$$\hat{S} = \bar{S}_1^{\,1/2} \left\{ \bar{S}_1^{\,1/2} \, M_2 R_2^{-1} \, M_2 \bar{S}_1^{\,1/2} + I \right\}^{-1} \bar{S}_1^{\,1/2}$$

$$= (\tilde{M}_2 R_2^{-1} \, M_2 + \bar{S}_1^{\,-1})^{-1} = \bar{S}.$$

Since $\hat{S} = \bar{S}$, $\hat{x}$ must equal $\bar{x}$, the minimum variance being attained uniquely by $\hat{x}$, from Theorem 1.

In the second case we consider $y_1 = M_1x + v_1$

$$y_2 = M_2x + v_2,$$

with $R = \begin{pmatrix} R_{11} & R_{12} \\ \tilde{R}_{12} & R_{22} \end{pmatrix}$, and M_1 nonsingular. Since $\bar{v}_1 = 0$, and M_1 is nonsingular, we have $x - \bar{x}_1 = - M_1^{-1} v_1$, $z = y_2 - M_2\bar{x}_1 = v_2 - M_2M_1^{-1} v_1$, and $\bar{S}_1 = M_1^{-1} R_{11}\widetilde{M_1^{-1}}$.

We compute

$$E(x - \bar{x}_1)\,\tilde{z} = M_1^{-1} R_{11}\widetilde{M_1^{-1}}\,\tilde{M}_2 - R_{12}$$

$$Ez\tilde{z} = R_{22} - \widetilde{R_{12}}\widetilde{M_1^{-1}}\,\tilde{M}_2 - M_2M_1^{-1} R_{12} + M_2M_1^{-1} R_{11}\,\widetilde{M_1^{-1}}\,\tilde{M}_2$$

$$= (M_2M_1^{-1} R_{11}^{1/2} - \widetilde{R_{12}}\, R_{11}^{-1/2})\,(R_{11}^{1/2}\,\widetilde{M_1^{-1}}\,\tilde{M}_2 - R_{11}^{-1/2} R_{12})$$

$$+ R_{22} - \widetilde{R_{12}}\, R_{11}^{-1} R_{12}\,.$$

Thus the $\hat{S}$ calculated on the basis of y_2 and initial value of $\bar{x}_1$, $\bar{S}_1$, is

$$\hat{S} = \bar{S}_1 - \left[E(x - \bar{x}_1)\tilde{z}\right] \left[Ez\tilde{z}\right]^{-1} \left[Ez\widetilde{(x-\bar{x}_1)}\right]$$

$$= M_1^{-1} R_{11}\widetilde{M_1^{-1}} - M_1^{-1}(R_{11}\widetilde{M_1^{-1}}\,\tilde{M}_2 - R_{12}) \left\{Ez\tilde{z}\right\}^{-1} (M_2M_1^{-1} R_{11} - \widetilde{R_{12}})\,\widetilde{M_1^{-1}}.$$

Some evident factoring, and use of

$$P_{22} \triangleq R_{22} - \tilde{R}_{12}R_{11}^{-1} R_{12},$$

yields

$$\hat{S} = M_1^{-1} R_{11}^{1/2} \left\{ I - (R_{11}^{1/2} \widetilde{M_1^{-1}} \tilde{M}_2 - R_{11}^{-1/2} R_{12}) P_{22}^{1/2} \cdot \right.$$

$$\cdot \left[P_{22}^{1/2} (M_2 M_1^{-1} R_{11}^{1/2} - \tilde{R}_{12} R_{11}^{-1/2}) (R_{11}^{1/2} \widetilde{M_1^{-1}} \tilde{M}_2 - R_{11}^{-1/2} R_{12}) P_{22}^{1/2} + I \right]^{-}$$

$$\left. \cdot P_{22}^{1/2} (M_2 M_1^{-1} R_{11}^{1/2} - \tilde{R}_{12} R_{11}^{-1/2}) \right\} R_{11}^{1/2} \widetilde{M_1^{-1}} .$$

Applying Lemma A to this expression, with

$$K = (R_{11}^{1/2} \widetilde{M_1^{-1}} \tilde{M}_2 - R_{11}^{-1/2} R_{12}) P_{22}^{1/2},$$

we obtain

$$\hat{S} = M_1^{-1} R_{11}^{1/2} \left\{ (R_{11}^{1/2} \widetilde{M_1^{-1}} \tilde{M}_2 - R_{11}^{-1/2} R_{12}) P_{22} (M_2 M_1^{-1} R_{11}^{1/2} - \widetilde{R_{12}} R_{11}^{-1/2}) + I \right\}^{-1}$$

$R_{11}^{1/2} \widetilde{M_1^{-1}}$, giving

$$\hat{S}^{-1} = (\tilde{M}_2 - \tilde{M}_1 R_{11}^{-1} R_{12}) P_{22} (M_2 - \widetilde{R_{12}} R_{11}^{-1} M_1) + \tilde{M}_1 R_{11}^{-1} M_1.$$

It is easily seen that this is $\bar{S}^{-1} = \tilde{M} R^{-1} M$, since

$$R^{-1} = \begin{pmatrix} R_{11}^{-1} R_{12} P_{22} \tilde{R}_{12} R_{11}^{-1} + R_{11}^{-1} & - R_{11}^{-1} R_{12} P_{22} \\ - P_{22} \tilde{R}_{12} R_{11}^{-1} & P_{22} \end{pmatrix} .$$

Thus the equality of $\hat{S}$ and $\bar{S}$ is established, so that $\hat{x} = \bar{x}$ as in the previous case.

IV. CONTINOUS PROCESSES

Even though observations are made only at discrete time points, it is desirable to consider a continuous process $x(\tau)$ correctly, rather than as a discontinuous process. This is especially true in case it is desired to predict future values of the x process, not necessarily only at times of anticipated observation. In this section we consider continuous processes corresponding to the discrete processes specified in Theorem 2, and develop continuous extrapolation of best estimates and their variances of estimate. In IV.2 cases of simple prediction are considered; these are cases where the best predictions depend only upon the present values of scalar observations and their derivatives. In IV.3 the limiting case of observations with superposed random impulse functions is developed.

IV.1 Continuous Linear Structure

In the preceding section we considered processes with a linear structure:

$$x_\alpha = L_{\alpha,\alpha-1}\, x_{\alpha-1} + v_\alpha \,.$$

We wish now to subsume these processes by continuous processes with a similar linear structure. This is frequently done by expressing the structure as[+]

(IV.1) $\quad x'(\tau) = B(\tau)\, x(\tau) + v$

where B is a matrix function of time and v is a random impulse function (fictitious derivative of a process of orthogonal increments). Although this is a reasonable characterization, it is possible and for some purposes preferable to express the structure of the process without introducing such random impulse functions, and we shall do so in the following theorem. In practice, one may frequently view the structure as specified either way,

---

+ We recall that we denote $dF(\tau)/d\tau$ and $dF(\tau,\sigma)/d\tau$ by $F'(\tau)$ and $F'(\tau,\sigma)$ respectively.

with the symmetric non-negative definite matrix $Q(\tau)$ the rate of growth of variance of the process of orthogonal increments, of which v is the fictitious derivative.

If (IV.1) is to represent the structure of $x(\tau)$ it is clear that the vector $x(\tau)$ must be a complete state vector, including as components all existing non-zero derivatives of components, and can not in general be restricted to a particular set of linear combination of components (sometimes called the <u>signal</u>) which may be of special interest. If the α^{th} diagonal element of $Q(\tau)$ is non-zero, no derivative of $\xi_\alpha(\tau)$, the α^{th} component of $x(\tau)$, can exist. The following theorem exhibits a general process structure replacing (IV.1) and identifiable with the discrete process structure of Theorem 2, with $L_{\alpha,\alpha-1}$ non-singular, and shows the best linear predictor of $x(\tau)$, and its variance of estimate. (In Theorem 3, $x(\tau)$, $\hat{x}(\tau,\tau_\alpha)$, and $w(\tau_2,\tau_1)$ are column vectors of μ components; $K(\tau_2,\tau_1)$, $B(\sigma)$, $Q(\sigma)$, and $S(\tau,\tau_\alpha)$ are $\mu \times \mu$.)

<u>Theorem 3</u>. Over a finite interval $[\tau_*,\tau_{**}]$ let $B(\sigma)$ and $Q(\sigma)$ be bounded piecewise continuous square matrices, with $Q(\sigma)$ symmetric and non-negative definite. For any $\tau_* \le \tau_1 \le \tau_2 \le \tau_{**}$, let $K(\tau_1,\tau_1) = 1$ and

(3.0.1) $$K'(\tau_2,\tau_1) = B(\tau_2)\, K(\tau_2,\tau_1) ;$$

let $x(\tau)$ be a random process with $x(\tau_*)$ having finite variance and

(3.0.2) $$x(\tau_2) = K(\tau_2,\tau_1)\, x(\tau_1) + w(\tau_2,\tau_1),$$

where the random variables $w(\tau_2,\tau_1)$ have mean zero, variance

(3.0.3) $$Ew(\tau_2,\tau_1)\, \tilde{w}(\tau_2,\tau_1) = \int_{\tau_1}^{\tau_2} K(\tau_2,\sigma)\, Q(\sigma)\, \tilde{K}(\tau_2,\sigma)\, d\sigma,$$

are mutually orthogonal for non-overlapping intervals [i.e., for $\tau_* \le \tau_1 \le \tau_2 \le \tau_3 \le \tau_4 \le \tau_{**}$, $Ew(\tau_4,\tau_3)\, \tilde{w}(\tau_2,\tau_1) = 0$], and are orthogonal to $x(\tau)$ and $\hat{x}(\tau,\tau)$ for any $\tau \le \tau_1$.

<u>Conclusion</u> <u>A</u>. For any $\tau_0, \tau_1, \ldots, \tau_\nu$, with $\tau_* \le \tau_0 < \tau_1 < \ldots < \tau_\nu \le \tau_{**}$, the hypotheses of Theorem 2 regarding x_α, $L_{\alpha,\alpha-1}$, and v_α are satisfied by $x(\tau_\alpha)$, $K(\tau_\alpha,\tau_{\alpha-1})$, and $w(\tau_2,\tau_{\alpha-1})$, respectively.

<u>Conclusion</u> <u>B</u>. For $\tau \ge \tau_\alpha$, let $\hat{x}(\tau,\tau_\alpha)$ and $S(\tau,\tau_\alpha)$ represent the minimum variance linear estimate of $x(\tau)$ based on $y_{(\alpha)}$, and the corresponding variance of estimate.

(3.1.1) $$\hat{x}'(\tau,\tau_\alpha) \triangleq \frac{\partial\, \hat{x}(\tau,\tau_\alpha)}{\partial\, \tau} = B(\tau)\, \hat{x}(\tau,\tau_\alpha)\ .$$

(3.1.2) $$S'(\tau,\tau_\alpha) \triangleq \frac{\partial\, S(\tau,\tau_\alpha)}{\partial\, \tau} = B(\tau)\, S(\tau,\ \tau_\alpha) + S(\tau,\ \tau_\alpha)\, \breve{B}(\tau) + Q(\tau).$$

<u>Proof</u> <u>A</u>. The identification of $x(\tau_\alpha)$ with x_α, $K(\tau_\alpha,\tau_{\alpha-1})$ with $L_{\alpha,\alpha-1}$, and v_α with $w(\tau_\alpha,\tau_{\alpha-1})$ is evident on comparing (3.0.2) with (2.0.1). $K(\tau_2,\tau_1)$ is a bounded constant by (3.0.1), $B(\sigma)$ being integrable by hypothesis. The $w(\tau_\alpha,\tau_{\alpha-1})$, for $\alpha = 1, 2, \ldots, \nu$ are mutually orthogonal by hypothesis, of finite variance by (3.0.3) and hypotheses on $B(\sigma)$ and $Q(\sigma)$. The variables are orthogonal to $\hat{x}_{0,0}$ and x_0 by the final hypothesis. Finite second moments of x follow from (3.0.2) and finite variance of $x(\tau_*)$.

Proof B. From Theorem 2, Conclusion A, and (3.0.3)

(3.2.1) $$\hat{x}(\tau,\tau_1) = K(\tau,\tau_1)\,\hat{x}(\tau_1,\tau_1),$$

(3.2.2) $$S(\tau,\tau_1) = K(\tau,\tau_1)\,S(\tau_1,\tau_1)\,\tilde{K}(\tau,\tau_1) + \int_{\tau_1}^{\tau} K(\tau,\sigma)\,Q(\sigma)\,\tilde{K}(\tau,\sigma)\,d\sigma.$$

Differentiating (3.2.1) and (3.2.2) with respect to τ, and using (3.0.1), one obtains (3.1.1) and (3.1.2).

If $B(\tau)$ is constant over $[\tau_1,\tau_2]$, then $K(\tau_2,\tau_1) = e^{(\tau_2-\tau_1)B}$.

The processes $x(\tau)$ postulated in the theorem are continuous vector Markov processes in the wide sense, without discontinuities in first or second moments. Such processes are typical in practical applications. Problems in which $x(\tau)$ does not satisfy the requirements of the theorem may require the use of Theorem 2 as an alternative or supplement to Theorem 3.

Example IV.1.a. An example from Wiener [1949] assumes observation on an ergodic process with spectral density $1/(1+\omega^2)$. Here we have

$$\eta = y = x = \xi.$$

To specify problems of this nature, one can factor the spectral density into conjugate factors in $\rho = \sqrt{-1}\,\omega$, one of which has neither poles nor zeroes in the lower half plane, and regard $\eta(\tau)$ as the result of this factor operating on a random impulse function. Thus in this example, $\eta(\tau) = \frac{1}{1+\rho}v$, with v a (fictitious) random impulse function. Operational calculus gives $\eta(\tau) + \eta'(\tau) = v$, or $\xi(\tau) = -1 \cdot \xi'(\tau) + 1 \cdot v$, so that $B = -1$, $Q = 1$.

Clearly $\hat{\xi}(\tau,\tau) = \eta(\tau)$, the observation, and $S(\tau,\tau) = 0$. Assuming observations extending to time τ, we can predict $\xi(\tau+\alpha)$ and its variance by (3.1.1), (3.1.2):

$$\hat{\xi}'(\tau+\alpha,\tau) = B(\tau+\alpha)\,\hat{\xi}(\tau+\alpha,\tau) = -\,\hat{\xi}(\tau+\alpha,\tau)$$

$$S'(\tau+\alpha,\tau) = S(\tau+\alpha,\tau)\,\tilde{B}(\tau+\alpha) + B(\tau+\alpha)\,S(\tau+\alpha,\tau) + Q(\tau+\alpha) = -\,2\,S(\tau+\alpha,\tau) + 1,$$

from which $\hat{\xi}(\tau + \alpha,\tau) = \hat{\xi}(\tau,\tau)\, e^{-\alpha} = \eta(\tau)\, e^{-\alpha}$,

$$S(\tau + \alpha,\tau) = (1 - e^{-2\alpha})/2 \ .$$

IV.2. Simple Prediction.

It is well known (e.g., Doob [1953]), that prediction of an observed stable scalar process with spectral density of the form

$1/\left|\left[\sum_{\gamma=0}^{\beta} \varphi_\gamma (j\omega)^\gamma\right]\right|^2$, with $j = \sqrt{-1}$, depends only upon the values at the time of final observation, of the variable and its derivatives up to the $(\beta - 1)^{st}$, which exist. Let the observation be $\eta_1(\tau) = \xi_1(\tau)$; then the successive derivatives give $\eta_1{}'(\tau) \triangleq \eta_2(\tau) = \xi_1{}'(\tau) = \xi_2(\tau)$, etc. The complete state vector is $x(\tau) = \begin{pmatrix} \xi_1(\tau) \\ \xi_2(\tau) \\ \vdots \\ \xi_\beta(\tau) \end{pmatrix}$, and by equating $\eta(\tau)$ and its derivatives to $\xi_1(\tau), \xi_2(\tau), \ldots, \xi_\beta(\tau)$, we have $\hat{x}(\tau,\tau) = x(\tau)$, $S = 0$.

This result can be generalized to nonstationary cases. Suppose that the complete state vector is $x(\tau) = \begin{pmatrix} \xi_1(\tau) \\ \xi_1'(\tau) \\ \vdots \\ \xi^{(\beta-1)}(\tau) \end{pmatrix}$, and that over an arbitrarily short interval ending at time τ, the observation $\eta(\sigma) = \xi_1(\sigma)$. The derivatives of $\eta(\tau)$ are the derivatives of $\xi_1(\tau)$, so that $x(\tau)$ can be determined with zero error as $\begin{pmatrix} \eta(\tau) \\ \eta'(\tau) \\ \vdots \\ \eta^{(\beta-1)}(\tau) \end{pmatrix} = \hat{x}(\tau,\tau)$.

The best prediction is obtained by applying (3.1.1) with initial values $\hat{x}(\tau,\tau) = \begin{pmatrix} \eta(\tau) \\ \eta_1(\tau) \\ \vdots \\ \eta^{(\beta-1)}(\tau) \end{pmatrix}$.

The initial variance $S(\tau,\tau)$ is zero in relation (3.1.2):

$$S'(\tau + \alpha,\tau) = B(\tau + \alpha)\, S(\tau + \alpha,\tau) + S(\tau + \alpha,\tau)\, \tilde{B}(\tau + \alpha)$$

$$+ Q(\tau + \alpha) .$$

We note that the essential requirements for simple prediction, i.e., estimation with zero error based on observations over an arbitrarily short interval, are twofold: first, the vector x must consist of a component $\xi_1(\tau)$ and possibly some of its derivatives: second, the observation $\eta(\tau)$ over the interval must be $\xi_1(\tau)$. Absence of "error of observation" is <u>not</u> sufficient for simple prediction.

Example IV.2.b. An example from Wiener [1949] assumes observation of an ergodic process with spectral density $1/(1 + \omega^4)$. We have $x(\tau) = \begin{pmatrix} \xi_1(\tau) \\ \xi_2(\tau) \end{pmatrix} = \begin{pmatrix} \xi_1(\tau) \\ \xi'(\tau) \end{pmatrix}$, with $A = (1,0)$. Here the stable factor of $\frac{1}{1+\omega^4}$ is $\frac{1}{1 + \sqrt{2}\rho + \rho^2}$. Setting $\xi_1(\tau) = \eta(\tau) = \frac{1}{1 + \sqrt{2}\rho + \rho^2} v$, we have $\xi_1(\tau) + \sqrt{2}\, \xi'(\tau) + \xi_1''(\tau) = v$. Setting $\xi_1'(\tau) = \xi_2(\tau)$, this gives $\xi_2'(\tau) = \xi_1''(\tau) = -\xi_1(\tau) - \sqrt{2}\, \xi_2(\tau) - 1 \cdot v$, so that $B = \begin{pmatrix} 0 & 1 \\ -1 & -\sqrt{2} \end{pmatrix}$, $Q = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$.

This is simple prediction since $\eta(\tau) = \xi_1(\tau)$ and $x(\tau) = \begin{pmatrix} \xi_1(\tau) \\ \xi_1'(\tau) \end{pmatrix}$.

In writing the solution, to shorten notation we write $\hat{x}$ for $\hat{x}(\tau + \alpha,\tau)$, S for $S(\tau + \alpha,\tau)$, and let $S = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix}$. The prediction equations are

$$\hat{\xi}_1' = \hat{\xi}_2$$

$$\hat{\xi}_2' = -\hat{\xi}_1 - \sqrt{2}\,\hat{\xi}_2$$

$$S' = \begin{pmatrix} 2\sigma_{12} & \sigma_{22} \quad \sigma_{11} - \sqrt{2}\,\sigma_{12} \\ \sigma_{22} - \sigma_{11} - \sqrt{2}\,\sigma_{12} & 1 - 2\sigma_{12} - 2\sqrt{2}\,\sigma_{22} \end{pmatrix},$$

with initial conditions $\hat{x}(\tau,\tau) = \begin{pmatrix} \eta(\tau) \\ \eta'(\tau) \end{pmatrix}$, $S(\tau,\tau) = 0$. The solutions of the differential equations are:

$$\hat{\xi}_1(\tau+\alpha,\tau) = (\cos \alpha/\sqrt{2} + \sin \alpha/\sqrt{2})\, e^{-\alpha/\sqrt{2}}\, \eta(\tau) + \sqrt{2}(\sin \alpha/\sqrt{2})\, e^{-\alpha/\sqrt{2}}\, \eta'(\tau),$$

$$\hat{\xi}_2(\tau+\alpha,\tau) = -\sqrt{2}(\sin \alpha/\sqrt{2})\, e^{-\alpha/\sqrt{2}}\, \eta(\tau) + (\cos \alpha/\sqrt{2} - \sin \alpha/\sqrt{2})\, e^{-\alpha/\sqrt{2}}\, \eta'(\tau).$$

$$\sigma_{11}(\tau+\alpha,\tau) = \frac{\sqrt{2}}{4}\left\{1 - (2 - \cos\sqrt{2}\,\alpha + \sin\sqrt{2}\,\alpha)\, e^{-\sqrt{2}\,\alpha}\right\}$$

$$\sigma_{12}(\tau+\alpha,\tau) = \frac{1}{2}(1 - \cos\sqrt{2}\,\alpha)\, e^{-\sqrt{2}\,\alpha}$$

$$\sigma_{22}(\tau+\alpha,\tau) = \frac{\sqrt{2}}{4}\left\{1 - (2 - \cos\sqrt{2}\,\alpha - \sin\sqrt{2}\,\alpha)\, e^{-\sqrt{2}\,\alpha}\right\}.$$

__Example IV.2.c.__ From Wiener [1949] we have the above example except with spectral density $1/(1+\omega^2)^2$. $x(\tau)$, $A(\tau)$, $Q(\tau)$ are as before, but now $B = \begin{pmatrix} 0 & 1 \\ -1 & -2 \end{pmatrix}$. Routine substitution gives, using short notation as before,

$$\hat{\xi}_1' = \hat{\xi}_2$$

$$\hat{\xi}_2 = -\hat{\xi}_1 - 2\hat{\xi}_2$$

$$S' = \begin{pmatrix} 2\sigma_{12} & \sigma_{22} - \sigma_{11} - 2\sigma_{12} \\ \sigma_{22} - \sigma_{11} - 2\sigma_{12} & 1 - 2\sigma_{12} - 4\sigma_{22} \end{pmatrix}.$$

with initial conditions $\hat{x}(\tau,\tau) = \begin{pmatrix} \eta(\tau) \\ \eta'(\tau) \end{pmatrix}$, $S(\tau,\tau) = 0$. The solutions of the differential equations are:

$$\hat{\xi}_1(\tau + \alpha, \tau) = (1 + \alpha)\, e^{-\alpha}\, \eta(\tau) + \alpha e^{-\alpha}\, \eta'(\tau) \;.$$

$$\hat{\xi}_2(\tau + \alpha, \tau) = -\, \alpha e^{-\alpha}\, \eta(\tau) + (1 - \alpha)\, e^{-\alpha}\, \eta'(\tau) \;,$$

$$\sigma_{11}(\tau + \alpha, \tau) = \left\{ 1 - (1 + 2\alpha + 2\alpha^2)\, e^{-2\alpha} \right\} \Big/ 4 \;,$$

$$\sigma_{12}(\tau + \alpha, \tau) = (\alpha^2\, e^{-2\alpha})/2 \;,$$

$$\sigma_{22}(\tau + \alpha, \tau) = \left\{ 1 - (1 - 2\alpha + 2\alpha^2)\, e^{-2\alpha} \right\} \Big/ 4 \;.$$

Example IV.2.d. Wiener [1949] considers an ergodic process with spectral density $e^{-\omega^2}$, theoretically completely predictable, and approximates it by a rational spectral density for prediction purposes. This is a limiting case of simple prediction, with all derivatives existing, and the complete state vector not of finite dimensions. The B matrix is infinite, of the form

$$\begin{pmatrix} 0 & 1 & 0 & . & . & . & . & . & . & . \\ 0 & 0 & 1 & 0 & . & . & . & . & . & . \\ 0 & 0 & 0 & 1 & 0 & . & . & . & . & . \\ . & . & . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . & . & . \end{pmatrix} .$$

To predict to any desired degree of precision, one can use the observation and its derivatives to an appropriate order, and extrapolate by a finite Taylor series.

Example IV.2.e. An example from Wiener [1949], involving non-scalar observations, assumes that the two components of observed stationary processes have spectral densities $\Phi_{11}(\omega) = \Phi_{22}(\omega) = 1/(1 + \omega^2)$, $\Phi_{12}(\omega) = \epsilon/(1 - j\omega)^2$, $\Phi_{21}(\omega) = \epsilon/(1 + j\omega)^2$, with $0 < \epsilon < 1$, $j = \sqrt{-1}$. It is desired to predict each component.

To represent this problem in our terms, we note that the specified $\Phi_{11}(\omega)$ and $\Phi_{22}(\omega)$ would be obtained with $\eta_1 = (1 + j\omega)^{-1}\, v_1$ and $\eta_2 = \epsilon(1 + j\omega)^{-1}\, v_1 + \sqrt{1 - \epsilon^2}\,(1 + j\omega)^{-1}\, v_2$, with v_1 orthogonal to v_2, the forms being suggested by the ϵ in $\Phi_{12}(\omega)$. These trial values yield $\Phi_{12}(\omega)$, (expectation of η_1 times complex conjugate of η_2) as $\epsilon[1 - (j\omega)^2]^{-1}$; the desired value of $\Phi_{12}(\omega)$ is obtained if the first term of the trial η_2 is multiplied by $(1 - j\omega)\,(1 + j\omega)^{-1}$, and this does not affect Φ_2. Thus we may write

$$\eta_2 = \epsilon(1 - j\omega)\,(1 + j\omega)^{-2}\, v_1 + \sqrt{1 - \epsilon^2}\,(1 + j\omega)^{-1}\, v_2 ,$$

$$\eta_1 = (1 + j\omega)^{-1}\, v_1 = (1 + j\omega)\,(1 + j\omega)^{-2}\, v_1 .$$

From inspection of these relations, and results of earlier examples, we have

$$\eta_1 = \xi_1 + \xi_2$$

$$\eta_2 = \epsilon(\xi_1 - \xi_2) + \sqrt{1 - \epsilon^2}\,\xi_3, \text{ with}$$

$$B = \begin{pmatrix} 0 & 1 & 0 \\ -1 & -2 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad Q = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

$$A = \begin{pmatrix} 1, & 1, & 0 \\ \epsilon, & -\epsilon, & \sqrt{1 - \epsilon^2} \end{pmatrix}.$$

For predicting the observation process $y(\tau)$, this is a case of simple prediction. Using the results of examples (IV.1.a) and (IV.2.c)

$$\hat{\eta}_1(\tau + \alpha, \tau) \equiv \hat{\xi}_1(\tau + \alpha, \tau) + \hat{\xi}_2(\tau + \alpha, \tau)$$

$$= e^{-\alpha}\left[\hat{\xi}_1(\tau, \tau) + \hat{\xi}_2(\tau, \tau)\right] = e^{-\alpha}\,\eta_1(\tau);$$

$$\hat{\eta}_2(\tau + \alpha,\tau) \equiv \sqrt{1 - \epsilon^2}\ \hat{\xi}_3(\tau + \alpha,\tau) + \epsilon\left[\hat{\xi}_1(\tau + \alpha,\tau) - \hat{\xi}_2(\tau + \alpha,\tau)\right]$$

$$= e^{-\alpha}\left\{\sqrt{1 - \epsilon^2}\ \hat{\xi}_3(\tau,\tau) + \epsilon\left[(1 + 2\alpha)\ \hat{\xi}_1(\tau,\tau) - (1 - 2\alpha)\ \hat{\xi}_2(\tau,\tau)\right]\right\}$$

$$= e^{-\alpha}\left\{\eta_2(\tau) + 2\alpha\epsilon\eta_1(\tau)\right\} \quad .$$

The variance of estimate of $\hat{\eta}_1(\tau + \alpha,\tau)$ is $(1 - e^{-2\alpha})/2$, from example IV.1.a. That of $\eta_2(\tau + \alpha)$ is

$$\frac{1}{2} - e^{-2\alpha}\left\{E\eta_2^2(\tau) + 4\alpha^2\ \epsilon^2\ E\eta_1^2(\tau) + 4\alpha\epsilon\ E\eta_2(\tau)\ \eta_1(\tau)\right\}$$

$$= \frac{1}{2} - e^{-2\alpha}\left\{\frac{1}{2} + \frac{4\alpha^2\ \epsilon^2}{2}\right\} = \left\{1 - (1 + 4\alpha^2\ \epsilon^2)\ e^{-2\alpha}\right\} /2 \quad .$$

The covariance of estimate of $\hat{\eta}_1(\tau + \alpha,\tau)$ and $\hat{\eta}_2(\tau + \alpha,\tau)$ is

$$e^{-2\alpha}\ E\left\{\eta_1(\tau)\ \cdot\ \left[\eta_2(\tau) + 2\alpha\epsilon\eta_1(\tau)\right]\right\} = \epsilon\alpha e^{-2\alpha} \quad .$$

IV.3. The "White Noise" Limiting Case

A fictional limiting case of some interest is that in which, superposed on the linear function of $x(\tau)$, the observation contains so-called "white noise," that is random impulse functions, or the fictitious derivative of a process of orthogonal increments. This limiting case can easily be derived from Theorem 2, which states that for any discrete observation $y_\alpha = A_\alpha x_\alpha$,

$$\Delta\ \hat{x}_\alpha \triangleq \hat{x}_{\alpha,\alpha} - \hat{x}_{\alpha,\alpha-1} = \left[E(x_\alpha - \hat{x}_{\alpha,\alpha-1})\ \widetilde{(y_\alpha - A_\alpha\hat{x}_{\alpha,\alpha-1})}\right]$$

$$\cdot\left[E(y_\alpha - A_\alpha\hat{x}_{\alpha,\alpha-1})\ \widetilde{(y_\alpha - A_\alpha\hat{x}_{\alpha,\alpha-1})}\right]^{-1}\ (y_\alpha - A_\alpha\hat{x}_{\alpha,\alpha-1})$$

If we deal with $y_\alpha = A_\alpha x_\alpha + v_\alpha$, where v is orthogonal to x_α and $v_{\alpha-1}$, .. , etc., the relation remains valid, since $\hat{v}_{\alpha,\alpha-1}$ is zero. Now let the time interval between observations be δ, with $Ev_\alpha\tilde{v}_\alpha = N_\alpha/\delta$, nonsingular N_α, obtaining for the increment $\Delta\ \hat{x}_\alpha$:

$$\Delta \hat{x}_\alpha = S_{\alpha,\alpha-1} \tilde{A}_\alpha (A_\alpha S_{\alpha,\alpha-1}\tilde{A}_\alpha + N_\alpha/\delta)^{-1} (y_\alpha - A_\alpha \hat{x}_{\alpha,\alpha-1}).$$

Considering a sequence of decreasing δ, tending to zero, one has† with τ the time of the observation y_α,

$$\dot{\hat{x}}(\tau,\tau) = \lim_{\delta\to 0} (\Delta \hat{x}_\alpha/\delta)$$

$$= \lim_{\delta\to 0} \left\{ S_{\alpha,\alpha-1} \tilde{A}_\alpha (\delta A_\alpha S_{\alpha,\alpha-1}\tilde{A}_\alpha + N_\alpha)^{-1} (y_\alpha - A_\alpha \hat{x}_{\alpha,\alpha-1}) \right\}$$

$$= S(\tau,\tau)\, \tilde{A}(\tau)\, N^{-1}(\tau) \left[y(\tau) - A(\tau)\, \hat{x}(\tau,\tau) \right] \quad ,$$

<u>provided that $A(\tau)$ and $N(\tau)$ are continuous</u>. By a similar calculation it is easily seen that

$$\dot{S}(\tau,\tau) = - S(\tau,\tau)\, \tilde{A}(\tau)\, N^{-1}(\tau)\, A(\tau)\, S(\tau,\tau).$$

Combining these results with those of Theorem 3, we have the general solution of this limiting case in terms of differential equations. It is necessary only that the matrix functions of time, $A(\tau)$, $B(\tau)$, $Q(\tau)$, and $N(\tau)$, be piecewise continuous. There are numerous cases in practice where the simplicity of this limiting case outweighs its defects, just as there are numerous cases in which simplicity of the limiting case in which one considers his processes stationary and $A(\tau)$ constant outweighs resulting deficiencies.

<u>Example IV.3.a.</u> [Wiener, 1949.] One observes the sum of a stationary process (the "signal") with spectral density $1/(1 + \omega^2)$, and "white noise" of spectral density ϵ^2. As earlier, the signal ξ has $A = 1$, $B = -1$, $Q = 1$, while the noise term has $N = \epsilon^2$. Thus,

$$\dot{\hat{\xi}}(\tau,\tau) = S(\tau,\tau)\, \tilde{A}N^{-1} \left[\eta(\tau) - A\hat{\xi}(\tau,\tau) \right]$$

$$= S(\tau,\tau)\, \epsilon^{-2} \left[\eta(\tau) - \hat{\xi}(\tau,\tau) \right] \quad .$$

†Recall that $\dot{F}(\tau,\sigma) \triangleq dF(\tau,\sigma)/d\tau$.

$$\dot{S}(\tau,\tau) = - S(\tau,\tau)\, \tilde{A}N^{-1} AS(\tau,\tau) = - S^2(\tau,\tau)/\varepsilon^2 .$$

From Theorem 3 we have

$$\hat{\xi}'(\tau,\tau) = - \hat{\xi}(\tau,\tau)$$

$$S'(\tau,\tau) = 1 - 2S(\tau,\tau) .$$

During observation, we have

$$\frac{d\hat{\xi}(\tau,\tau)}{d\tau} = \dot{\hat{\xi}}(\tau,\tau) + \hat{\xi}'(\tau,\tau)$$

and

$$\frac{dS(\tau,\tau)}{d\tau} = \dot{S}(\tau,\tau) + S'(\tau,\tau) = 1 - 2S(\tau,\tau) - S^2(\tau,\tau)/\varepsilon^2 .$$

Suppose that one observes from time 0 to time τ, (possibly with $\hat{\xi}(0,0) = 0$, $S(0,0) = 1/2$, from the nature of the process). The differential equation for $S(\tau,\tau)$ has the solution, with

$$\gamma \triangleq \sqrt{1 + \varepsilon^2}/\varepsilon .$$

$$S(\tau,\tau) = \frac{1}{\gamma+1} + \frac{2\gamma}{\gamma^2-1}\left\{\frac{(\gamma^2-1)\, S(0,0) + \gamma + 1}{(\gamma^2-1)\, S(0,0) - \gamma + 1}\, e^{2\gamma\tau} - 1\right\}^{-1} ,$$

with $1/(\gamma + 1) \equiv \varepsilon(\sqrt{1 + \varepsilon^2} - \varepsilon)$ the asymptotic value of $S(\tau,\tau)$ as τ increases without limit.

As in example IV.1.a, prediction from time τ to time $(\tau + \alpha)$ is given by:

$$\hat{\xi}(\tau + \alpha,\tau) = \hat{\xi}(\tau,\tau)\, e^{-\alpha}$$

$$\hat{S}(\tau + \alpha,\tau) = \frac{1}{2} + \left[S(\tau,\tau) - 1/2\right] e^{-2\alpha} .$$

<u>Example</u> <u>IV.3.b</u>. [Wiener, 1949.] One observes the sum of a stationary signal with spectral density $1/(1 + \omega^4)$ and white noise of spectral density ε^4. It is desired to estimate the present value of the derivative of the signal. As in an earlier example, we have

$$x(\tau) = \begin{pmatrix} \xi_1(\tau) \\ \xi_2(\tau) \end{pmatrix}, \quad A = (1,0), \quad B = \begin{pmatrix} 0 & 1 \\ -1 & -\sqrt{2} \end{pmatrix}$$

$$Q = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}.$$

We consider observation from time 0 to time τ_1. If our only information at time 0 is the nature of the stationary process, we have $\hat{\xi}_1(0,0) = \hat{\xi}_2(0,0) = 0$, $S(0,0) = \begin{pmatrix} \sqrt{2}/4 & 0 \\ 0 & \sqrt{2}/4 \end{pmatrix}$.

We shall write $S(\tau + \alpha,\tau) = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix}_{(\tau + \alpha,\tau)}$. During the observation interval,

$$\dot{\hat{X}}(\tau,\tau) = S(\tau,\tau)\,\tilde{A}(\tau)\,N^{-1}(\tau)\,\zeta(\tau)$$

$$= \begin{pmatrix} \sigma_{11} \\ \sigma_{12} \end{pmatrix} \epsilon^{-4}\,\zeta(\tau),$$

with $\zeta(\tau) = \eta(\tau) - \hat{\xi}_1(\tau,\tau)$;

$$\dot{S}(\tau,\tau) = -\epsilon^{-4} \begin{pmatrix} \sigma_{11}^2 & \sigma_{11}\sigma_{12} \\ \sigma_{11}\sigma_{12} & \sigma_{12}^2 \end{pmatrix}_{(\tau,\tau)}$$

Both during and after the observation interval $\hat{X}'(\tau + \alpha,\tau) = B\hat{X}(\tau + \alpha,\tau)$, so that

$$\hat{\xi}_1'(\tau + \alpha,\tau) = \hat{\xi}_2(\tau + \alpha,\tau)$$

$$\hat{\xi}_2'(\tau + \alpha,\tau) = - \hat{\xi}_1(\tau + \alpha,\tau) - \sqrt{2}\, \hat{\xi}_2(\tau + \alpha,\tau) \ .$$

Also $S'(\tau + \alpha,\tau) = BS(\tau + \alpha,\tau) + S(\tau + \alpha,\tau)\tilde{B} + Q =$

$$\begin{pmatrix} 2\sigma_{12} & \sigma_{22}-\sigma_{11}-\sqrt{2}\sigma_{12} \\ \sigma_{22}-\sigma_{11}-\sqrt{2}\sigma_{12} & 1-2\sigma_{12}-2\sqrt{2}\sigma_{22} \end{pmatrix}_{(\tau + \alpha,\tau)} .$$

Equating to zero $dS(\tau,\tau)/d\tau = \dot{S}(\tau,\tau) + S'(\tau,\tau)$, one obtains the asymptotic solution as τ increases without limit:

$$\sigma_{11(\infty,\infty)} = \sqrt{2}\,\epsilon^3 \left((1 + \epsilon^4)^{1/4} - \epsilon\right) = \sqrt{2}\,\epsilon^4\,\beta$$

$$\sigma_{12(\infty,\infty)} = \epsilon^2 \left((1 + \epsilon^4)^{1/4} - \epsilon\right)^2 = \epsilon^4\,\beta^2$$

$$\sigma_{22(\infty,\infty)} = \sqrt{2}\,\epsilon(1 + \epsilon^4)^{3/4} - 2\sqrt{2}\,\epsilon^2(1 + \epsilon^4)^{1/2}$$

$$+ 2\sqrt{2}\,\epsilon^3(1 + \epsilon^4)^{1/4} - \sqrt{2}\,\epsilon^4 = \sqrt{2}\,\epsilon^4\,\beta(\beta^2 + \beta + 1),$$

with $\beta \triangleq \epsilon^{-1}(1 + \epsilon^4)^{1/4} - 1$. Substituting the asymptotic $S(\infty,\infty)$ into $d\hat{\xi}_1/d\tau$ and $d\hat{\xi}_2/d\tau$, one obtains the asymptotically optimum frequency operators on η; with $\rho = \sqrt{-1}\,\omega$,

$$\hat{\xi}_2(\tau,\tau) \simeq \frac{\beta^2\rho - \sqrt{2}\,\beta}{(\beta+1)^2 + \sqrt{2}\,(\beta+1)\rho + \rho^2}\,\eta(\tau)$$

$$\hat{\xi}_1(\tau,\tau) \simeq \frac{\sqrt{2}\,\beta\rho + \beta(\beta+2)}{(\beta+1)^2 + \sqrt{2}\,(\beta+1)\rho + \rho^2}\,\eta(\tau),$$

the asymptotically optimum operator for $\hat{\xi}_2(\tau,\tau)$ agreeing with the result in the reference.

Example IV.3.c. [Shinbrot, 1956.] A particle leaves the origin at time zero, with constant velocity of mean zero and variance β. One observes from time 0 to time τ the sum of particle position and white noise of spectral density γ.

Let the particle position be $\xi_1(\tau)$, its velocity be $\xi_2(\tau)$; we have

$$\hat{x}(0,0) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad S(0,0) = \begin{pmatrix} 0 & 0 \\ 0 & \beta \end{pmatrix},$$

$$B = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad Q = 0, \quad A = (1,0).$$

With $\alpha \geq 0$, $S(\tau + \alpha,\tau) \triangleq \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix}_{(\tau + \alpha,\tau)}$.

During observation

$$\dot{\hat{x}}(\tau,\tau) = S(\tau,\tau)\, \tilde{A}N^{-1}\, \zeta(\tau) = \begin{pmatrix} \sigma_{11} \\ \sigma_{12} \end{pmatrix} \zeta(\tau)/\gamma ,$$

with $\zeta(\tau) = \eta(\tau) - \hat{\xi}_1(\tau,\tau)$;

$$\dot{S}(\tau,\tau) = -\begin{pmatrix} \sigma_{11}^2 & \sigma_{11}\sigma_{12} \\ \sigma_{11}\sigma_{12} & \sigma_{12}^2 \end{pmatrix}_{(\tau,\tau)} \Big/ \gamma$$

Always $\hat{x}'(\tau + \alpha,\tau) = B\hat{x}(\tau + \alpha,\tau) = \begin{pmatrix} \hat{\xi}_2(\tau + \alpha,\tau) \\ 0 \end{pmatrix}$;

$$S'(\tau + \alpha,\tau) = \begin{pmatrix} 2\sigma_{12} & \sigma_{22} \\ \sigma_{22} & 0 \end{pmatrix}_{(\tau + \alpha,\tau)} .$$

Solving $dS(\tau,\tau)/d\tau = 0$ with $S(0,0) = \begin{pmatrix} 0 & 0 \\ 0 & \beta \end{pmatrix}$,

$$S(\tau,\tau) = \frac{3\gamma\beta}{3\gamma + \beta\tau^3} \begin{pmatrix} \tau^2 & \tau \\ \tau & 1 \end{pmatrix} .$$

Predicting from time τ to time $\tau + \alpha$ we have

$$\hat{\xi}_1(\tau + \alpha,\tau) = \hat{\xi}_1(\tau,\tau) + \alpha\, \hat{\xi}_2(\tau,\tau) ,$$

$$\hat{\xi}_2(\tau + \alpha,\tau) = \hat{\xi}_2(\tau,\tau) ,$$

and from $S'(\tau + \alpha,\tau)$, or directly from the above pair,

$$S(\tau + \alpha,\tau) = \frac{3\beta\gamma}{3\gamma + \beta\tau^3} \begin{pmatrix} (\tau + \alpha)^2 & \tau + \alpha \\ \tau + \alpha & 1 \end{pmatrix} .$$

Example IV.3.d. [Shinbrot, 1956.] Over a finite time interval beginning at time zero, one observes the sum of $\xi(\tau)$ and white noise of spectral density γ. $\xi(0)$ is zero; $\xi(\tau)$ is constant over the interval $(0,\tau_1)$ except for one jump of mean zero and variance $\beta\,\tau_1$, the time of the jump uniformly distributed between 0 and τ_1; for $\tau > \tau_1$, $\xi(\tau) = \xi(\tau_1)$. The best linear estimate of $\xi(\tau)$ is desired.

Here $A = 1$, $\hat{\xi}(0,0) = S(0,0) = 0$. We have $E\xi(\tau) = 0$,

$$E\xi^2(\tau) = \begin{cases} \beta\tau_1 \cdot (\tau/\tau_1) = \beta\tau & (\tau \le \tau_1) . \\ \beta\tau_1 & (\tau > \tau_1) \end{cases}$$

Thus $B = 0, Q = \begin{cases} \beta , & (\tau < \tau_1) . \\ 0 & (\tau > \tau_1) \end{cases}$

$$\hat{\xi}(\tau,\tau) = S(\tau,\tau)\, \zeta(\tau)/\gamma .$$

$$\dot{S}(\tau,\tau) = - S^2(\tau,\tau)/\gamma .$$

$$\hat{\xi}'(\tau + \alpha,\tau) = B\hat{\xi}(\tau + \alpha,\tau) = 0 .$$

$$S'(\tau + \alpha,\tau) = \begin{cases} \beta & (\tau + \alpha < \tau_1) \\ 0 & (\tau + \alpha > \tau_1) \end{cases} .$$

For $\tau \le \tau_1$, $\frac{dS(\tau,\tau)}{d\tau} = \beta - S^2(\tau,\tau)/\gamma$,

so that

$$S(\tau,\tau) = (\beta\gamma)^{1/2} \tanh \left[(\beta/\gamma)^{1/2} \tau\right] , \quad (\tau \le \tau_1) .$$

For $\tau \ge \tau_1$, $\frac{dS(\tau,\tau)}{d\tau} = - S^2(\tau,\tau)/\gamma$, giving

$$S(\tau,\tau) = \frac{\gamma\ S(\tau_1,\tau_1)}{\gamma + (\tau-\tau_1)\ S(\tau_1,\tau_1)} \qquad (\tau \ge \tau_1) .$$

For prediction from time τ to time $(\tau + \alpha)$ we have $\hat{\xi}(\tau + \alpha) = \hat{\xi}(\tau)$,

$$S\ (\tau + \alpha,\tau) = S(\tau,\tau) + \beta\ \varphi,$$

with $\varphi = \min\ (\tau_1 - \tau,\alpha)$, if $\tau < \tau_1$, and φ zero otherwise.

V. CONTINUOUS OBSERVATIONS

In this section we consider observations continuous over an interval, in fact with $A(\tau)$ possessing a derivative $A'(\tau) \triangleq dA(\tau)/d\tau$. The processing of data in such a case cannot be by numerical methods and it typically uses electronic equipment. A good example is the processing of data considered by Wiener (1949), in which processing is specified in terms of a constant coefficient frequency operator on continuous inputs. In general, the best estimate on the basis of continuous observations will require continuous data processing, but can often be approximated well enough by arithmetic methods in which the observations are lumped into approximate discrete observations or even simply sampled. Once this approximation has been made, the methods of Section III can be employed. From a mathematical standpoint, it is not advisable to ignore the case of continuous observations, as the study may shed light on theoretical points as well as practical points involving techniques and accuracy of approximation. From the standpoint of applications, the methods of processing continuous data are of great interest, since in many cases continuous processing may be cheaper, faster, and/or more precise than digital processing, despite the tremendous progress in digital computers.

In Section IV a fictional limiting case involving continuous observations with superposed white noise was considered and solved, but our primary interest is in realistic situations involving observations which are finite with probability one. We must now require that continuous $A(\tau)$ be differentiable. The simplest case is considered first, in which the matrix $A(\tau)\ Q(\tau)\ \tilde{A}(\tau)$ is nonsingular. This condition is satisfied if each scalar observation includes a simple Markov component, unless there is redundancy in the observations. In general, it is presumed that trivial singularities resulting from redundancy in the matrix $A(\tau)$ are removed by routine methods. The next case considered is that in which the observation function is scalar, $A(\tau)$ consisting of a single row vector $\tilde{a}(\tau)$, but with $\tilde{a}(\tau)\ Q(\tau)\ a(\cdot)$ possibly singular.

V.1. $A(\tau)\ Q(\tau)\ \tilde{A}(\tau)$ Nonsingular.

With continuous $A(\tau)$, it is natural to attempt to apply Theorems 2 and 3, with the time between observation $y_{\alpha-1}$ and y_α shrinking to zero. The difficulty that arises is that $S_{\alpha,\alpha}\tilde{A}_\alpha$ is zero, as can be seen from (2.1.6), so that the estimate $\hat{u}_\alpha$ of the error of estimate of x_α, given by (2.1.5) is a meaningless expression involving the product of a zero matrix times the inverse of a zero matrix. We must therefore obtain the limit of (2.1.5) as the time interval δ between the observations $y(\tau)$ and $y(\tau + \delta)$ shrinks to zero. This limit exists provided $A'(\tau)$ and $[A(\tau)\ Q(\tau)\ \tilde{A}(\tau)]^{-1}$ exist at τ, as shown in the following theorem. (In Theorem 4, $y(\tau)$ is a column vector with ν components, $x(\tau)$ and $\mu(\tau)$ column vectors with μ components. $A(\tau)$ is $\nu \times \mu$; $B(\tau)$, $Q(\tau)$, and $S(\tau,\tau)$ are $\mu \times \mu$.)

Theorem 4. Given the hypotheses of Theorems 2 and 3, with $A(\tau)$ a specified differentiable function, $A(\tau)\ Q(\tau)\ \tilde{A}(\tau)$ nonsingular. Let $\hat{u}(\tau)$ represent the increment in $\hat{x}(\tau,\tau)$ due to the observation $y(\tau) = A(\tau)\ x(\tau)$, $\dot{S}(\tau,\tau)$ the time derivative of $S(\tau,\tau)$ due to contributions $\hat{u}(\tau)$, and $z(\tau) = y(\tau) - A(\tau)\ \hat{x}(\tau,\tau)$. Then

(4.1.1) $$\hat{u}(\tau) = \left[S(\tau,\tau)\ \tilde{A}(\tau)\right]' \left\{\left[A(\tau)\ S(\tau,\tau)\ \tilde{A}(\tau)\right]'\right\}^{-1} z(\tau),$$

(4.1.2) $$\dot{S}(\tau,\tau) = -\left[S(\tau,\tau)\ \tilde{A}(\tau)\right]' \left\{\left[A(\tau)\ S(\tau,\tau)\ \tilde{A}(\tau)\right]'\right\}^{-1} \left[A(\tau)\ S(\tau,\tau)\right]' ,$$

with

(4.1.3) $$\left[S(\tau,\tau)\ \tilde{A}(\tau)\right]' = S(\tau,\tau)\ \tilde{A}'(\tau) + S(\tau,\tau)\ \tilde{B}(\tau)\ \tilde{A}(\tau) + Q(\tau)\ \tilde{A}(\tau) ,$$

(4.1.4) $$\left[A(\tau)\ S(\tau,\tau)\ \tilde{A}(\tau)\right]' = A(\tau)\ Q(\tau)\ \tilde{A}(\tau) .$$

Proof. From (2.1.6) of Theorem 2, $S(\tau,\tau)\ \tilde{A}(\tau)$ is zero. From (3.1.2) of Theorem 3, [+]

(4.2.1) $$S'(\tau,\tau) = B(\tau)\ S(\tau,\tau) + S(\tau,\tau)\ \tilde{B}(\tau) + Q(\tau) .$$

[+]Recall that $F'(\tau,\sigma) \triangleq dF(\tau,\sigma)/d\tau$.

Since $S(\tau,\tau)\,\tilde{\Lambda}(\tau) = 0 = A(\tau)\,S(\tau,\tau)$, (4.2.1) yields

$$\left[S(\tau,\tau)\,\tilde{\Lambda}(\tau)\right]' = S(\tau,\tau)\,\tilde{\Lambda}'(\tau) + B(\tau)\cdot 0$$
$$+\, S(\tau,\tau)\tilde{B}(\tau)\tilde{\Lambda}(\tau). + Q(\tau)\,\tilde{\Lambda}(\tau),$$

which is (4.1.3). Similarly, (4.1.4) is obtained from (4.2.1) and (4.1.3) by

$$\begin{aligned}\left[A(\tau)\,S(\tau,\tau)\,\tilde{\Lambda}(\tau)\right]' &= A'(\tau).0 + A(\tau)[S(\tau,\tau)\tilde{\Lambda}(\tau)]' \\ &= A(\tau)S'(\tau,\tau)\tilde{\Lambda}(\tau) \\ &= A(\tau)Q(\tau)\tilde{\Lambda}(\tau)\ .\end{aligned}$$

To determine $\hat{U}(\tau)$, we have from Theorem 2

$$\begin{aligned}\hat{x}(\tau+\delta,\,\tau+\delta) - \hat{x}(\tau+\delta,\tau) &= S(\tau+\delta,\tau)\,\tilde{\Lambda}(\tau+\delta)\cdot \\ &\cdot \left[A(\tau+\delta)\,S(\tau+\delta,\tau)\,\tilde{\Lambda}(\tau+\delta)\right]^{-1}\cdot \\ &\cdot \left[y(\tau+\delta) - A(\tau+\delta)\,\hat{x}(\tau+\delta,\tau)\right].\end{aligned} \tag{4.2.2}$$

Now $\hat{u}(\tau)$ is the limit as $\delta \to 0$ of (4.2.2). From Theorem 3 and the differentiability of $A(\tau)$ we have that $y(\tau+\delta) - A(\tau+\delta)\cdot \hat{x}(\tau+\delta,\tau) \equiv A(\tau+\delta)\,[x(\tau+\delta) - \hat{x}(\tau+\delta,\tau)]$ is continuous at $\delta = 0$ if $w(\tau+\delta,\tau)$ is continuous at $\delta = 0$. (3.0.3) indicates that the mean square of $w(\tau+\delta,\tau)$ is continuous and tends to zero as δ tends to zero, so that $w(\tau+\delta,\tau)$ is continuous with probability one. Thus the final factor on the right of (4.2.2) tends to the limit $z(\tau)$ as δ approaches zero. The limit of the other factors on the right of (4.2.2) is as given in (4.1.1), by the Mean Value Theorem, in view of (4.1.3), (4.1.4), and the hypotheses on the matrices involved. In the same manner, the limit of $[S(\tau+\delta,\tau+\delta) - S(\tau+\delta,\tau)]/\delta$ is seen to be given by (4.1.2).

<u>Example</u> <u>V.1.a</u>. [Shinbrot 1956.] A particle leaves the origin at time zero, with constant velocity of mean zero, variance β. One observes, from

time zero to some finite time, the sum of particle position and a disturbance $\xi_3(\tau)$ of mean zero and $E\,\xi_3(\tau_1)\,\xi_3(\tau_2) = \gamma e^{-\varphi|\tau_1-\tau_2|}$, with $\gamma > 0$, $\varphi > 0$. (A limiting case, with $\xi_3(\tau)$ replaced by white noise, was considered in example IV.3.c.)

The product moment function for ξ_3, with $\tau_1 = \tau_2$, shows that the constant variance of ξ_3 is γ. Comparing the product moment function with $e^{(\tau_2-\tau_1)B}\,E\,\xi_3^2(\tau_1)$, obtained from (3.0.2) on multiplication by $\xi_3(\tau_1)$ and taking expectations, shows that the element of B corresponding to ξ_3 is $-\varphi$. Finally, the element of Q corresponding to ξ_3 is $2\gamma\varphi$, by $S' = BS + S\tilde{B} + Q$, since the variance of ξ_3 is constant.

Here we have $\xi_1(\tau)$ = position, $\xi_2(\tau)$ = velocity, $\xi_3(\tau)$ = disturbance.

$$B = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -\varphi \end{pmatrix}, \quad Q = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 2\gamma\varphi \end{pmatrix},$$

$$A = (1,\ 0,\ 1).$$

Following the initial observation at time 0, when $A(\tau)$ jumped from 0 to (1, 0, 1), during the observation interval $A(\tau)$ is constant and $A(\tau)\,Q(\tau)\,\tilde{A}(\tau) = 2\gamma\varphi > 0$, so that the hypotheses of Theorem 4 are satisfied. The initial observation must be used as shown in Section III. Assuming no information other than that of the problem statement, we have $\hat{x}(0,0-) = 0$, $S(0,0-) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & \beta & 0 \\ 0 & 0 & \gamma \end{pmatrix}.$

At the initial observation, $\hat{u}(0) = S(0,0-)\,\tilde{A}(0)\left[A(0)\,S(0,0-)\,\tilde{A}(0)\right]^{-1}\eta(0) = \begin{pmatrix} 0 \\ 0 \\ \gamma \end{pmatrix}\gamma^{-1}\,\eta(0) = \begin{pmatrix} 0 \\ 0 \\ \eta(0) \end{pmatrix},$

i.e., the initial observation specifies the value of $\xi_3(0)$. From this or, if preferred, more formally, one has $S(0,0) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & \beta & 0 \\ 0 & 0 & 0 \end{pmatrix}$. Thus the continuous observation has initial values $\hat{x}(0,0) = \begin{pmatrix} 0 \\ 0 \\ \eta(0) \end{pmatrix}$, $S(0,0) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & \beta & 0 \\ 0 & 0 & 0 \end{pmatrix}$. Since $S(\tau,\tau)\,\tilde{A}(\tau) = 0$ during the observation interval, we write

$$S(\tau,\tau) = \begin{pmatrix} \sigma_{11} & \sigma_{12} & -\sigma_{11} \\ \sigma_{12} & \sigma_{22} & -\sigma_{12} \\ -\sigma_{11} & -\sigma_{12} & \sigma_{11} \end{pmatrix}_{(\tau,\tau)} .$$

With $\zeta(\tau) = \eta(\tau) - A(\tau)\,\hat{x}(\tau,\tau)$,

$$\hat{u}(\tau) = \left[S(\tau,\tau)\,\tilde{B} + Q\right] \tilde{A}(AQ\tilde{A})^{-1}\,\zeta(\tau)$$

$$= \frac{1}{2\gamma\varphi} \begin{pmatrix} \sigma_{12} + \varphi\sigma_{11} \\ \sigma_{22} + \varphi\sigma_{12} \\ 2\gamma\varphi - \varphi\sigma_{11} - \sigma_{12} \end{pmatrix}_{(\tau,\tau)} \zeta(\tau) .$$

$$\dot{S}(\tau,\tau) = -\frac{1}{2\gamma\varphi} \begin{pmatrix} (\sigma_{12}+\varphi\sigma_{11})^2 & (\sigma_{12}+\varphi\sigma_{11})(\sigma_{22}+\varphi\sigma_{12}) & (\sigma_{12}+\varphi\sigma_{11})(2\gamma\varphi-\sigma_{12}-\varphi\sigma_{11}) \\ \cdot & (\sigma_{22}+\varphi\sigma_{12})^2 & (\sigma_{22}+\varphi\sigma_{12})(2\gamma\varphi-\sigma_{12}-\varphi\sigma_{11}) \\ \cdot & \cdot & (2\gamma\varphi-\sigma_{12}-\varphi\sigma_{11})^2 \end{pmatrix}_{(\tau,\tau)}$$

Both during and after the observation interval,

$$\hat{x}'(\tau + \alpha, \tau) = B\hat{x}(\tau + \alpha, \tau) = \begin{pmatrix} \hat{\xi}_2(\tau + \alpha, \tau) \\ 0 \\ -\varphi\, \hat{\xi}_3(\tau + \alpha, \tau) \end{pmatrix}$$

$$S'(\tau + \alpha, \tau) = BS(\tau + \alpha, \tau) + S(\tau + \alpha, \tau)\, \tilde{B} + Q$$

$$= \begin{pmatrix} 2\sigma_{12} & \sigma_{22} & \varphi\sigma_{11} - \sigma_{12} \\ \sigma_{22} & 0 & \varphi\sigma_{12} \\ \varphi\sigma_{11} - \sigma_{12} & \varphi\sigma_{12} & 2\varphi(\gamma - \sigma_{11}) \end{pmatrix}_{(\tau,\tau)}$$

Solving $dS(\tau,\tau)/d\tau \equiv \dot{S}(\tau,\tau) + S'(\tau,\tau)$, with $S(0,0)$ as specified, one obtains

$$S(\tau,\tau) = \frac{6\gamma\varphi\beta}{6\gamma\varphi + 3\beta\tau + 3\varphi\beta\tau^2 + \varphi^2\,\beta\tau^3} \begin{pmatrix} \tau^2 & \tau & -\tau^2 \\ \tau & 1 & -\tau \\ -\tau^2 & -\tau & \tau^2 \end{pmatrix}$$

The problem of sequential estimation has been solved. Possible confusion may be avoided by calculating also the prediction from time τ to time $\tau + \alpha$, and then taking into account an observation at time $\tau + \alpha$, assuming the continuous observation ceased at time τ. The predicted vector is from $\hat{x}'(\tau + \alpha, \tau)$,

$$\hat{\xi}_1(\tau + \alpha, \tau) = \hat{\xi}_1(\tau,\tau) + a\, \hat{\xi}_2(\tau,\tau),$$

$$\hat{\xi}_2(\tau + \alpha, \tau) = \hat{\xi}_2(\tau,\tau)$$

$$\hat{\xi}_3(\tau + \alpha, \tau) = \hat{\xi}_3(\tau,\tau) \cdot e^{-\varphi\alpha} .$$

During the prediction interval, we do not have $S(\tau + \alpha, \tau)$ constrained by $A(\tau)$:

$$S'(\tau + \alpha,\tau) = BS(\tau + \alpha,\tau) + S(\tau + \alpha,\tau)\,\tilde{B} + Q$$

$$= \begin{pmatrix} 2\sigma_{12} & \sigma_{22} & \sigma_{23}-\varphi\sigma_{13} \\ \cdot & 0 & -\varphi\sigma_{23} \\ \cdot & \cdot & 2\varphi(\gamma - \sigma_{33}) \end{pmatrix}$$

Writing $\epsilon(\tau) = 6\gamma\varphi\beta/(6\gamma\varphi + 3\beta\tau + 3\varphi\beta\tau^2 + \varphi^2\beta\tau^3)$, the solution of the differential equation, with initial condition $S(\tau,\tau)$, is

$$S(\tau + \alpha,\tau) = \begin{pmatrix} (\tau + \alpha)^2\,\epsilon(\tau) & (\tau + \alpha)\,\epsilon(\tau) & -\tau(\tau + \alpha)\,e^{-\varphi\alpha}\,\epsilon(\tau) \\ \cdot & \epsilon(\tau) & -\tau e^{-\varphi\alpha}\,\epsilon(\tau) \\ \cdot & \cdot & \gamma-e^{-2\varphi\alpha}\,[\gamma-\tau^2\,\epsilon(\tau)] \end{pmatrix}$$

For an observation at time $(\tau + \alpha)$,

$$\hat{Q}(\tau + \alpha) = S(\tau + \alpha,\tau)\,A(\tau + \alpha)\,[A(\tau + \alpha)\,S(\tau + \alpha,\tau)\,\tilde{A}(\tau)]^{-1}\,\zeta(\tau + \alpha),$$

$$S(\tau + \alpha,\tau + \alpha) - S(\tau + \alpha,\tau)$$

$$= - S(\tau + \alpha,\tau)\,\tilde{A}(\tau + \alpha)\,[A(\tau + \alpha)\,S(\tau + \alpha,\tau)\,\tilde{A}(\tau + \alpha)]^{-1}\,A(\tau + \alpha)\,S(\tau + \alpha,\tau)\ .$$

<u>Example</u> <u>V.1.b</u>. Wiener, 1949, considers observations over $[-\infty,\tau_1]$ on a process with spectral density $\Phi(\omega) = (1 + \omega^2)/(1 + \omega^4)$, and wishes to predict the value of the process at time $(\tau_1 + \alpha)$. This is similar to Example IV.2.b, except that here $A = (1, 1)$ or perhaps $(1, -1)$, so that this is not a case of simple prediction. We shall consider separately Case I, in which $A = (1, 1)$, and Case II, in which $A = (1, -1)$; in each case we assume observation over an interval beginning at time zero, with initial estimates zero and variances of estimate the variances of the processes; namely

$$S(0, 0-) = \begin{pmatrix} \sqrt{2}/4 & 0 \\ 0 & \sqrt{2}/4 \end{pmatrix} .$$

The initial observation of $\eta(0)$ gives estimates $\hat{x}(0,0) = S(0,0-)\ \tilde{A}[AS(0,0)-\tilde{A}]^{-1}\ \eta(0)$ and $S(0,0) = S(0,0-) - S(0,0-)\tilde{A}\ [AS(0,0-)\tilde{A}]^{-1}\ AS(0,0-)$, or

Case I: $\hat{x}(0,0) = \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix} \eta(0);\qquad S(0,0) = \begin{pmatrix} \sqrt{2}/8 & -\sqrt{2}/8 \\ -\sqrt{2}/8 & \sqrt{2}/8 \end{pmatrix}$

Case II: $\hat{x}(0,0) = \begin{pmatrix} \frac{1}{2} \\ -\frac{1}{2} \end{pmatrix} \eta(0);\qquad S(0,0) = \begin{pmatrix} \sqrt{2}/8 & \sqrt{2}/8 \\ \sqrt{2}/8 & \sqrt{2}/8 \end{pmatrix}$

During the observation interval, $S(\tau,\tau)\ \tilde{A} = 0$ implies that

$S(\tau,\tau) = \begin{pmatrix} \sigma & -\sigma \\ -\sigma & \sigma \end{pmatrix}_\tau$ in Case I, $\begin{pmatrix} \sigma & \sigma \\ \sigma & \sigma \end{pmatrix}_\tau$ in Case II.

Recalling that $B = \begin{pmatrix} 0 & 1 \\ -1 & -\sqrt{2} \end{pmatrix}$, $Q = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$, we compute

$\Delta\ \hat{x}(\tau) = [S(\tau,\tau)\ \tilde{B} + Q]\ \tilde{A}(AQ\tilde{A})^{-1}\ \zeta(\tau)$, $S'(\tau,\tau) = S(\tau,\tau)\ B + BS(\tau,\tau) + Q$, $\dot{S} - [S(\tau,\tau)\ \tilde{B} + Q]\ \tilde{A}(AQ\tilde{A})^{-1}\ A[BS(\tau,\tau) + Q]$, and $\hat{x}'(\tau,\tau) = B\hat{x}(\tau,\tau)$, obtaining $\hat{x}'(\tau,\tau) = \begin{pmatrix} 0 & 1 \\ -1 & -\sqrt{2} \end{pmatrix} x(\tau,\tau)$ in both cases.

In Case I,

$$\Delta\ \hat{x}(\tau,\tau) = \begin{pmatrix} -(2-\sqrt{2})\sigma \\ 1 + (2-\sqrt{2})\sigma \end{pmatrix}_\tau \zeta(\tau)$$

$$\dot{S}(\tau,\tau) = -\begin{pmatrix} (2-\sqrt{2})^2\sigma^2 & -(2-\sqrt{2})\sigma-(2-\sqrt{2})^2\sigma^2 \\ \cdots & [1+(2-\sqrt{2})\sigma]^2 \end{pmatrix}_\tau$$

$$S'(\tau,\tau) = \begin{pmatrix} -2\sigma & \sqrt{2}\sigma \\ \sqrt{2}\sigma & 1+2(1-\sqrt{2})\sigma \end{pmatrix}_\tau$$

The solution is defined in terms of $\sigma(\tau)$, which from $dS/d\tau = S' + \dot{S}$ satisfies $d\sigma/d\tau = -2\sigma - (2-\sqrt{2})^2\sigma^2$ with initial value $\sqrt{2}/8$. The solution for $\sigma(\tau)$ is

$$\sigma(\tau) = [(3+2\sqrt{2})(e^{2\tau}-1)+4\sqrt{2}]^{-1} ,$$

tending to zero as τ increases.

In Case II,

$$\Delta\hat{x}(\tau,\tau) = \begin{pmatrix} (2+\sqrt{2})\sigma \\ (2+\sqrt{2})\sigma - 1 \end{pmatrix}_\tau \zeta(\tau),$$

$$\dot{S}(\tau,\tau) = -\begin{pmatrix} (2+\sqrt{2})^2\sigma^2 & (2+\sqrt{2})^2\sigma^2-(2+\sqrt{2})\sigma \\ \cdots & [(2+\sqrt{2})\sigma-1]^2 \end{pmatrix}_\tau$$

$$S'(\tau,\tau) = \begin{pmatrix} 2\sigma & -\sqrt{2}\sigma \\ -\sqrt{2}\sigma & 1-\sqrt{2}\sigma(1+\sqrt{2}) \end{pmatrix}_\tau .$$

In Case II $d\sigma/d\tau = 2\sigma - (2+\sqrt{2})^2\sigma^2$, which with $\sigma(0) = \sqrt{2}/8$ yields

$$\sigma(\tau) = [(3-2\sqrt{2})(1-e^{-2\tau})+4\sqrt{2}]^{-1} ,$$

tending to $3 - 2\sqrt{2}$ as τ increases.

The extrapolation to time $\tau_1 + \alpha$ is in either case given by $\hat{x}'(\tau,\tau_1) = B\hat{x}(\tau,\tau_1) = \begin{pmatrix} 0 & 1 \\ -1 & -\sqrt{2} \end{pmatrix} \hat{x}(\tau,\tau_1)$, yielding estimates of the same form as in Example IV.2.b, with $\eta(\tau)$ and $\eta'(\tau)$ replaced by $\hat{\xi}_1(\tau_1,\tau_1)$ and $\hat{\xi}_2(\tau_1,\tau_1)$ respectively. Letting σ_{ij} represent the covariance of estimate of $\hat{\xi}_i$ and $\hat{\xi}_j$, we have the same differential equation for $\sigma_{ij}(\tau + \alpha,\tau)$ as in Example IV.2.b, but different initial conditions since $S(\tau,\tau)$ is non-zero. Thus $S(\tau_1 + \alpha,\tau_1)$ is as in IV.2.b, together with the increments due to $\sigma(\tau_1)$:

Case I

$$\left\{\begin{aligned} \delta\sigma_{11}(\tau_1 + \alpha,\tau_1) &= \sigma(\tau_1)\, e^{-\sqrt{2}\alpha}\{(2 - \sqrt{2}) + (1 - \sqrt{2})(\sin\sqrt{2}\alpha - \cos\sqrt{2}\alpha)\} \\ \delta\sigma_{12}(\tau_1 + \alpha,\tau_1) &= \sigma(\tau_1)\, e^{-\sqrt{2}\alpha}\{(1 - \sqrt{2}) + \sqrt{2}(1 - \sqrt{2})\cos\sqrt{2}\alpha\} \\ \delta\sigma_{22}(\tau_1 + \alpha,\tau_1) &= \sigma(\tau_1)\, e^{-\sqrt{2}\alpha}\{(2 - \sqrt{2}) - (1 - \sqrt{2})(\sin\sqrt{2}\alpha + \cos\sqrt{2}\alpha)\} \end{aligned}\right\}$$

or Case II

$$\left\{\begin{aligned} \delta\sigma_{11}(\tau_1 + \alpha,\tau_1) &= \sigma(\tau_1)\, e^{-\sqrt{2}\alpha}\{(2 + \sqrt{2}) + (1 + \sqrt{2})(\sin\sqrt{2}\alpha - \cos\sqrt{2}\alpha)\} \\ \delta\sigma_{12}(\tau_1 + \alpha,\tau_1) &= \sigma(\tau_1)\, e^{-\sqrt{2}\alpha}\{(2 + \sqrt{2})\cos\sqrt{2}\alpha - (1 + \sqrt{2})\} \\ \delta\sigma_{22}(\tau_1 + \alpha,\tau_1) &= \sigma(\tau_1)\, e^{-\sqrt{2}\alpha}\{(2 + \sqrt{2}) - (1 + \sqrt{2})(\sin\sqrt{2}\alpha + \cos\sqrt{2}\alpha)\} \end{aligned}\right\}$$

The non-equality of $S(\tau,\mu)$ in Cases I and II is natural, although somewhat obscured in treatments merely estimating $\eta(\tau_1 + \alpha)$. The variance of estimate of $\hat{\eta}_1(\tau_1 + \alpha)$ is of course the same in Case II as in Case I, namely, $\frac{\sqrt{2}}{2}\{1 - e^{-\sqrt{2}\alpha}[(2 - \sqrt{2}) - (1 - \sqrt{2})\cos\sqrt{2}\alpha]\} + e^{-\sqrt{2}\alpha}\cdot 2(3 - 2\sqrt{2})(1 - \cos\sqrt{2}\alpha)/[(3 + 2\sqrt{2})(e^{2\tau} - 1) + 4\sqrt{2}]$.

For τ large, $\sigma(\tau)$ will be approximately equal to its asymptotic value of zero in Case I, or $3 - 2\sqrt{2}$ in Case II. The corresponding asymptotic values of $\Delta\hat{x}$ are:

$$\text{Case I,} \quad \Delta \hat{x} \simeq \begin{pmatrix} 0 \\ 1 \end{pmatrix} \zeta$$

$$\text{Case II,} \quad \Delta \hat{x} \simeq \begin{pmatrix} 2 - \sqrt{2} \\ 1 - \sqrt{2} \end{pmatrix} \zeta$$

Combining these with $\hat{x}' = B\hat{x}$, one obtains the asymptotically optimum frequency operators on $\eta(\tau)$ for $\hat{\xi}_1(\tau)$ and $\hat{\xi}_2(\tau)$:

Case I: $\quad \hat{\xi}_1 = (1 + j\omega)^{-1} \eta, \quad \hat{\xi}_2 = j\omega(1 + j\omega)^{-1} \eta \quad ;$

Case II: $\quad \hat{\xi}_1 = (1 + j\omega)^{-1} [(2 - \sqrt{2})\, j\omega + (\sqrt{2} - 1)]\, \eta \quad ,$

$$\hat{\xi}_2 = -(1 + j\omega)^{-1} [(\sqrt{2} - 1)\, j\omega + (2 - \sqrt{2})]\, \eta \ .$$

A Note on Instrumentation. One aspect of the solution for continuous observation is maintaining at zero the difference between $y(\tau)$ and $A(\tau)\, \hat{x}(\tau)$, by absorbing variations in $y(\tau)$ immediately into $\hat{x}(\tau)$. Thus we call for $\hat{u}(\tau)$, the instantaneous increment in $\hat{x}(\tau)$ to be a multiple, say $C(\tau)$, of $z(\tau) = y(\tau) - A(\tau)\, \hat{x}(\tau,\tau)$. In practice, a difficulty may arise in a system operating on $z(\tau)$, because the change in $\hat{x}(\tau)$ immediately affects $z(\tau)$, so that infinite amplification of $z(\tau)$ appears called for. The simplest reaction is to be content with an approximate solution, multiplying $z(\tau)$ by a large number, say one million, to render the inaccuracy negligible. The estimation equations can be mathematically transformed in many cases, particularly if the function $A(\tau)$ is known for all time, to a form eliminating this difficulty; it can also be avoided by assuming the existence of a convenient limiting case, such as the "white noise" case or the stationary case considered by Wiener; finally, it can be avoided by sampling the observable function. One should take care, however, that errors arising from incorrect mathematical formulation or types of instrumentation requiring great precision do not far outweigh the inaccuracies due to finite gain in a system operating on $z(\tau)$.

V.2. Scalar Observations.

We shall now consider scalar observations, with the hypotheses of Theorem 4 not necessarily satisfied. Since $y(\tau)$ has only one component, we shall now write $\eta(\tau)$, $\zeta(\tau)$, and $\tilde{a}(\tau)$ for $y(\tau)$, $z(\tau)$, and $A(\tau)$; and singularity of a matrix of the form $\tilde{a}(\tau)\ Q(\tau)\ a(\tau)$ is equivalent to a zero magnitude. We assume as in Theorem 4 that $\tilde{a}(\tau)$ is piecewise absolutely continuous, i.e., that $a(\tau)$ is differentiable except at a finite number of points, at which salti or isolated values occur (we are considering observation over a finite time).

We have previously considered the case in which $\tilde{a}(\tau)\ S(\tau,\tau-)\ a(\tau)$ is non-zero; and that in which this form is zero, but $\tilde{a}(\tau)$ is differentiable and $\tilde{a}(\tau)\ Q(\tau)\ a(\tau)$ is non-zero. This leaves the case in which $\tilde{a}(\tau)$ is differentiable but both quadratic forms are zero. It is reasonable in this case to consider differentiating $\zeta(\tau)$, possibly obtaining a non-zero form corresponding to $\tilde{a}(\tau)\ Q(\tau)\ a(\tau)$; the derivative of $\zeta(\tau)$ will in general include components of $x(\tau)$ more directly related to the disturbances represented by $Q(\tau)$. The following theorem[+] indicates that, under mild restrictions, such differentiation is possible and is a nonsingular transformation of the function $\eta(\tau)$. We use the operator $*$, such that

$$*F(\tau) \triangleq F'(\tau) + F(\tau)\ B(\tau).$$

(In Theorem 5, $a(\tau)$, $x(\tau)$, $\hat{x}(\tau,\tau)$, and $w(\sigma,\tau)$ are column vectors of μ components; $S(\tau,\tau)$, $Q(\tau)$, $J(\sigma,\tau)$, and $K(\sigma,\rho)$ are $\mu \times \mu$.)

Theorem 5. Given the hypotheses of Theorem 3, with $S(\tau,\tau)\ a(\tau)$, $Q(\tau)\ a(\tau)$, and $[Q(\tau)\ a(\tau)]'$ zero; $*\tilde{a}$ and $*^2\tilde{a}$ existing; then, with probability one,

$$\zeta'(\tau) - (*\tilde{a})\ [x(\tau) - \hat{x}(\tau,\tau)] = \eta'(\tau) - (*\tilde{a})\ \hat{x}(\tau,\tau) \tag{5.1.1}$$

exists, and is a nonsingular transformation of $\zeta(\tau)$.

Proof. Let $\sigma > \tau$, with hypotheses satisfied at time τ. From (3.0.3),

$$J(\sigma,\tau) \triangleq Ew(\sigma,\tau)\ \tilde{w}(\sigma,\tau) = \int_\tau^\sigma K(\sigma,\rho)\ Q(\rho)\ \tilde{K}(\sigma,\rho)\ d\rho. \tag{5.2.1}$$

Using (3.0.1),

+Here and later we make use of the fact that $AX\tilde{A}$ is zero if and only if AX is zero, for X symmetric and non-negative definite.

(5.2.2) $$J'(\sigma,\tau) = B(\sigma)\ J(\sigma,\tau) + J(\sigma,\tau)\ \tilde{B}(\sigma) + Q(\sigma).$$

Defining

(5.2.3) $$\gamma(\sigma,\tau) \triangleq \tilde{a}(\sigma)\ J(\sigma,\tau)\ a(\sigma),$$

and differentiating $\gamma(\sigma,\tau)$ with respect to σ, using (5.2.2) and the * operator,

(5.2.4) $$\gamma'(\sigma,\tau) = \tilde{a}'(\sigma)\ J(\sigma,\tau) + \tilde{a}(\sigma)\ [Q(\sigma) + B(\sigma)\ J(\sigma,\tau) + J(\sigma,\tau)\ \tilde{B}(\sigma)]\ a(\sigma) + \tilde{a}(\sigma)\ J(\sigma,\tau)\ a'(\sigma)$$

$$= [*\tilde{a}(\sigma)]\ J(\sigma,\tau)\ a(\sigma) + \tilde{a}(\sigma)\ J(\sigma,\tau)\ \widetilde{[*\tilde{a}(\sigma)]} + \tilde{a}(\sigma)\ Q(\sigma)\ a(\sigma).$$

Similarly,

(5.2.5) $$\gamma''(\sigma,\tau) = \left[*^2\tilde{a}(\sigma)\right] J(\sigma,\tau)\ a(\sigma) + 2\left[*\tilde{a}(\sigma)\right] J(\sigma,\tau)\ \widetilde{\left[*\tilde{a}(\sigma)\right]} + \tilde{a}(\sigma)\ J(\sigma,\tau)\ \widetilde{\left[\gamma *^2\tilde{a}(\sigma)\right]} + \left[*\tilde{a}(\sigma)\right] Q(\sigma)\ a(\sigma) + \tilde{a}(\sigma)\ Q(\sigma)\ \widetilde{\left[*\tilde{a}(\sigma)\right]} + \left[\tilde{a}(\sigma)\ Q(\sigma)\ a(\sigma)\right].$$

By (5.2.1) $J(\sigma,\tau)$ is continuous and zero at $\sigma = \tau$. Together with the hypotheses of this theorem, this makes every term of $\gamma(\sigma,\tau)$, $\gamma'(\sigma,\tau)$, and $\gamma''(\sigma,\tau)$ zero at $\sigma = \tau$.

By a familiar theorem, it follows that

(5.2.6) $$\gamma(\sigma,\tau) = \frac{1}{2} (\sigma - \tau)^2 \gamma_1(\sigma,\tau),$$

with $\gamma_1(\sigma,\tau)$ tending to zero as σ approaches τ. Consequently, from (5.2.1), (5.2.3), and (5.2.6),

(5.2.7) $$E \left[\tilde{a}(\sigma)\ w(\sigma,\tau)/(\sigma - \tau)\right] \left[\tilde{w}(\sigma,\tau)\ a(\sigma)/(\sigma - \tau)\right]$$
$$= \gamma(\sigma,\tau)/(\sigma - \tau)^2 = \gamma_1(\sigma,\tau)/2$$

has a zero limit as σ approaches τ. But this limit is the expected square of the derivative of $\tilde{a}(\sigma)\ w(\sigma,\tau)$ at $\sigma = \tau$, so that the derivative is zero with probability one. Consequently, using (3.0.2) and (3.0.1),

(5.2.8) $$\eta'(\tau) = \left[\tilde{a}(\tau)\ x(\tau)\right]' = \tilde{a}'(\tau)\ x(\tau) + \tilde{a}(\tau)\ B(\tau)\ x(\tau)$$
$$= \left[{*\tilde{a}}(\tau)\right]\ x(\tau),$$

with probability one. From (3.1.1),

(5.2.9) $$\left[\tilde{a}(\tau)\ \hat{x}(\tau,\tau)\right]' = \tilde{a}'(\tau)\ \hat{x}(\tau,\tau) + \tilde{a}(\tau)\ B(\tau)\ \hat{x}(\tau,\tau)$$
$$= \left[{*\tilde{a}}(\tau)\right]\ \hat{x}(\tau,\tau).$$

From (5.2.8) and (5.2.9), (5.1.1) is immediate. The transformation of $\zeta(\tau)$ into $\zeta'(\tau)$ is nonsingular, since $\zeta(\tau)$ is zero by the hypothesis that $S(\tau,\tau)\ \tilde{a}(\tau) = 0$.

In the hypotheses of Theorem 5, some mild restrictions have been placed on the functions $\tilde{a}(\tau)$, $B(\tau)$ and $Q(\tau)$. We recall that the hypotheses of Theorem 3 required $B(\tau)$ and $Q(\tau)$ to be piecewise continuous, while those of Theorem 4 implied $A(\tau)$ piecewise absolutely continuous. The hypotheses of Theorem 5 require that $B(\tau)$ and any existing derivative of $\tilde{a}(\tau)$ be piecewise absolutely continuous. Repeated application of Theorem 5 will require that higher existing derivatives of $\tilde{a}(\tau)$ and $B(\tau)$ be piecewise absolutely continuous.

The second restriction imposed by the hypotheses of Theorem 5 is that when $S(\tau,\tau)\ a(\tau)$ and $Q(\tau)\ a(\tau)$ are zero, we must have $[Q(\tau)\ a(\tau)]' = 0$, so that isolated zeros of $Q(\tau)\ a(\tau)$ are inadmissable. This restriction does not appear to be onerous in practice, so that Theorem 5 has been simplified by assuming such a restriction. If it is not satisfied, the proof indicates that the result is merely to change the expected square of the derivative of $\tilde{a}(\sigma)\ w(\sigma,\tau)$ at $\sigma = \tau$ from zero to $[\tilde{a}(\tau)\ Q(\tau)\ a(\tau)]'/2$, thus adding a term both to $\eta'(\tau)$ and $\zeta'(\tau)$. The practical consequence is that the observation $\eta(\tau)$ should be ignored for any τ at which $S(\tau,\tau)\ a(\tau)$ and $Q(\tau)\ a(\tau)$ are zero, but $[a(\tau)\ Q(\tau)\ a(\tau)]'$ is not zero.

When Theorem 5 is applicable, one simply replaces $\eta(\tau)$ by $n'(\tau)$, and $\tilde{a}(\tau)$ by $*\tilde{a}(\tau)$, then applies Theorem 2 or Theorem 4 if possible, otherwise one uses Theorem 5 again.

The writer has found in the literature no illustrative examples requiring the use of Theorem 5. To show the method, we apply it in Example IV.2.b., a case of simple prediction. Here $x(\tau) = \begin{pmatrix} \xi_1(\tau) \\ \xi_2(\tau) \end{pmatrix}$, $B = \begin{pmatrix} 0 & 1 \\ -1 & -\sqrt{2} \end{pmatrix}$, $Q = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$,

and $A(\tau) = (1,0)$ for $\tau \geq 0$, assuming continuous observation beginning at time zero. Assuming only the nature of the process, we have $\hat{x}(0,0-) = 0$, $S(0,0-) =$

$$\begin{pmatrix} \sqrt{2}/4 & 0 \\ 0 & \sqrt{2}/4 \end{pmatrix} = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix} (0,0-).$$

The initial observation at time zero yields $\hat{u}(0) = S(0,0-)a(0)\ [\tilde{a}(0)S(0,0-)a(0)]^{-1}\ \zeta(0) =$

$$= \begin{pmatrix} \sqrt{2}/4 \\ 0 \end{pmatrix} (\sqrt{2}/4)^{-1}\ [\eta(0) - 0] = \begin{pmatrix} \eta(0) \\ 0 \end{pmatrix}, \text{ and}$$

$$S(0,0) = S(0,0-) - \begin{pmatrix} \sqrt{2}/4 \\ 0 \end{pmatrix} (\sqrt{2}/4)^{-1}\ (\sqrt{2}/4, 0) = \begin{pmatrix} 0 & 0 \\ 0 & \sqrt{2}/4 \end{pmatrix}.$$

After the initial observation $S(\tau,\tau)a(\tau)$, $a'(\tau)$, $Q(\tau)a(\tau)$, and $[Q(\tau)a(\tau)]'$ are all zero, so Theorem 5 can be applied.

$*\tilde{a}(\tau) = \tilde{a}'(\tau) + \tilde{a}(\tau)B(\tau) = 0 + (1,0)\begin{pmatrix} 0 & 1 \\ -1, & -\sqrt{2} \end{pmatrix} = (0,1)$. The square of the expected value of $\zeta'(0)$ is

$$E\,[\zeta'(0)]^2 = *\tilde{a}(0)\,[S(0,0)]\,\widetilde{[*\tilde{a}(0)]} =$$

$$= (0,1)\begin{pmatrix} 0 & 0 \\ 0 & \sqrt{2}/4 \end{pmatrix}\begin{pmatrix} 0 \\ 1 \end{pmatrix} = \sqrt{2}/4 \neq 0.$$

The initial observation of $\dot{\eta}(\tau)$ therefore gives

$$\hat{\alpha}(0_+) = S(0,0)\,\widetilde{[*\tilde{a}(0)]}\{[*\tilde{a}(0)]\,S(0,0)\,\widetilde{[*\tilde{a}(0)]}\}^{-1}\,\zeta'(0_+)$$

$$= \begin{pmatrix} 0 \\ \sqrt{2}/4 \end{pmatrix}(\sqrt{2}/4)^{-1}\,(\eta'(0) - 0) = \begin{pmatrix} 0 \\ \eta'(0) \end{pmatrix};$$

$$S(0_+,0_+) = S(0,0) - \begin{pmatrix} 0 \\ \sqrt{2}/4 \end{pmatrix}(\sqrt{2}/4)^{-1}\,(0,\sqrt{2}/4) = 0.$$

Following the initial observation of $\eta'(\tau)$, $S(\tau,\tau)\,\widetilde{[*\tilde{a}(\tau)]}$ is zero, but $Q(\tau)\,\widetilde{[*\tilde{a}(\tau)]} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \neq 0$, so that Theorem 4 gives

$$\hat{\alpha}(\tau) = Q(\tau)\,\widetilde{[*\tilde{a}(\tau)]}\left\{[*a(\tau)]\,Q(\tau)\,\widetilde{[*a(\tau)]}\right\}^{-1}\zeta'(\tau) =$$

$$= \begin{pmatrix} 0 \\ 1 \end{pmatrix}\left\{(0,1)\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 0 \\ 1 \end{pmatrix}\right\}^{-1}\zeta'(\tau)$$

$$= \begin{pmatrix} 0 \\ \zeta'(\tau) \end{pmatrix}$$

Since $S(\tau,\tau) = 0$, $S'(\tau,\tau) = Q = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$. The above expressions for $u(\tau)$ indicate that $\hat{x}(\tau) = x(\tau)$. After the initial observation of $\eta(0)$ and $\eta'(0_+)$, one has

$$\dot{S}(\tau,\tau) = -\begin{pmatrix} 0 \\ 1 \end{pmatrix} (1)^{-1} (0,1) = -\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} = -S'(\tau,\tau),$$ showing formally that

$\hat{x}(\tau,\tau)$ has zero variance of estimate.

V.3. The General Problem. All the results needed to solve the estimation problem are now at hand. Summarizing results of V.2, when the observation is a scalar function with $a(\tau)$, $B(\tau)$ and $Q(\tau)$ possessing required piecewise continuity (absolute in some cases), at each τ we consider the sequence

$$\tilde{a}(\tau)S(\tau,\tau_-),\tilde{a}(\tau)Q(\tau), \left[*\tilde{a}(\tau)\right] S(\tau,\tau_-), \left[*\tilde{a}(\tau)\right] Q(\tau), \left[*^2\tilde{a}(\tau)\right] S(\tau,\tau_-), \ldots$$

the sequence terminating at the first non-zero term, called the <u>critical term</u>, which is of the form $*^{(\gamma)}(\tau)H(\tau)$. One replaces $\eta(\tau)$ and $\tilde{a}(\tau)$ by the γ^{th} derivative of $\eta(\tau)$ and $*^{\gamma}\tilde{a}(\tau)$, respectively, thus replacing $\zeta(\tau)$ by $\zeta^{(\gamma)}(\tau) = \eta^{(\gamma)}(\tau) - (*^{\gamma}\tilde{a})x(\tau,\tau_-)$. If the matrix $H(\tau)$ is $S(\tau,\tau_-)$ one applies Theorem 2, obtaining

$$\hat{u}(\tau) = S(\tau,\tau_-) \left[\widetilde{*^{\gamma}\tilde{a}(\tau)}\right] \left\{\left[*^{\gamma}\tilde{a}(\tau)\right] S(\tau,\tau_-) \left[\widetilde{*^{\gamma}\tilde{a}(\tau)}\right]\right\}^{-1} \zeta^{(\gamma)}(\tau),$$

$$S(\tau,\tau) = S(\tau,\tau_-) - S(\tau,\tau_-)\left[\widetilde{*^{\gamma}\tilde{a}(\tau)}\right]\left\{\left[*^{\gamma}\tilde{a}(\tau)\right] S(\tau,\tau_{-1}) \left[\widetilde{*^{\gamma}\tilde{a}(\tau)}\right]\right\}^{-1} \cdot$$

$$\cdot \left[*^{\gamma}\tilde{a}(\tau)\right] S(\tau,\tau_-) .$$

If the matrix $H(\tau)$ in the critical term is $Q(\tau)$ one applies Theorem 4, obtaining

$$\hat{u}(\tau) = \left\{S(\tau,\tau) \left[\widetilde{*^{\gamma+1}\tilde{a}(\tau)}\right] + Q(\tau) \left[\widetilde{*^{\gamma}\tilde{a}(\tau)}\right]\right\} \cdot \left\{\left[*^{\gamma}\tilde{a}(\tau)\right] Q(\tau) \left[\widetilde{*^{\gamma}\tilde{a}(\tau)}\right]\right\}^{-1} \zeta^{(\gamma)}(\tau)$$

$$\dot{S}(\tau,\tau) = -\left\{S(\tau,\tau) \left[\widetilde{*^{\gamma+1}\tilde{a}(\tau)}\right] + Q(\tau) \left[\widetilde{*^{\gamma}\tilde{a}(\tau)}\right]\right\} \cdot$$

$$\cdot \left\{ \left[*^{\gamma}\tilde{a}(\tau) \right] Q(\tau) \left[\widetilde{*^{\gamma}\tilde{a}(\tau)} \right] \right\}^{-1} \left\{ \left[*^{\gamma+1}\tilde{a}(\tau) \right] S(\tau,\tau) + \left[*^{\gamma}\tilde{a}(\tau) \right] Q(\tau) \right\} .$$

If the sequence has no critical term, the observation should be ignored. In all cases, from Theorem 3, one extrapolates continuously by

$$\hat{\mathfrak{X}}'(\tau + \alpha, \tau) = B(\tau + \alpha)\hat{\mathfrak{X}}(\tau + \alpha, \tau)$$

$$S'(\tau + \alpha, \tau) = B(\tau+\alpha)S(\tau+\alpha, \tau) + S(\tau+\alpha, \tau)\ \tilde{B}(\tau+\alpha) + Q(\tau+\alpha)$$

with $\alpha = 0$ when one is sequentially calculating $\hat{\mathfrak{X}}(\tau, \tau)$.

In considering observations which are non-scalar functions of time, we must first transform each scalar observation as discussed above, in order to avoid singularity of matrices to be inverted. To simplify notation, let us now use the symbols $\eta_{\alpha}^{+}(\tau)$, $\tilde{a}_{\alpha}^{+}(\tau)$, and $\zeta_{a}^{+}(\tau)$ for the transformed values of the α^{th} observation, coefficient row vector, and difference between observation and predicted observation respectively. Scalar observations with no critical term should be ignored.

Let $A^{S}(\tau)$ constitute the matrix formed of row vectors $\tilde{a}_{a}^{+}(\tau)$ corresponding to critical terms in $S(\tau, \tau_{-})$; and $z^{S}(\tau)$ the vector of corresponding $\zeta_{\alpha}^{+}(\tau)$. If the matrix $A^{S}(\tau)S(\tau,\tau_{-})\ \tilde{A}^{S}(\tau)$ is nonsingular, Theorem 2 can be applied, giving

$$\hat{\mathfrak{U}}^{S}(\tau) = S(\tau,\tau_{-})\ \tilde{A}^{S}(\tau) \left[A^{S}(\tau)S(\tau,\tau_{-1})\tilde{A}^{S}(\tau) \right]^{-1} z^{S}(\tau),$$

$$S(\tau,\tau) = S(\tau,\tau_{-}) - S(\tau,\tau_{-})\tilde{A}^{S}(\tau) \left[A^{S}(\tau)S(\tau,\tau_{-1})\tilde{A}^{S}(\tau) \right]^{-1} A^{S}(\tau)S(\tau,\tau_{-}).$$

Similarly, let $A^{Q}(\tau)$ and $z^{Q}(\tau)$ represent the matrix of coefficients and vector of scalar observations for which the critical terms contain $Q(\tau)$, and apply Theorem 4 if $A^{Q}(\tau)Q(\tau)\tilde{A}^{Q}(\tau)$ is nonsingular:

$$\hat{\mathfrak{U}}^{Q}(\tau) = \left\{ S(\tau,\tau) \left[\widetilde{*A^{Q}(\tau)} \right] + Q(\tau)\widetilde{A^{Q}(\tau)} \right\} \cdot \left\{ A^{Q}(\tau)Q(\tau)\widetilde{A^{Q}(\tau)} \right\}^{-1} z^{Q}(\tau),$$

with $-\dot{S}(\tau,\tau)$ of similar form, $z^{Q}(\tau)$ being replaced by the transpose of the first term.

Thus, in summary we vary $\hat{X}(\tau,\tau)$ by $\hat{X}'(\tau,\tau)$, $\hat{U}^{S}(\tau)$, and $\hat{U}^{Q}(\tau)$; and vary $S(\tau,\tau)$ by $S'(\tau,\tau)$, $\dot{S}(\tau,\tau)$ and the incremental term involving $A^{S}(\tau)S(\tau,\tau_{-1})\tilde{A}^{S}(\tau)$. There remains a question regarding possible singularity of the matrices to be inverted. The only reasonable possibility of singularity appears to be observations which are linearly dependent, in which case an elementary algebraic transformation should be used to remove the singularity.

VI. GAUSSIAN PROCESSES

In previous sections, it has not been assumed that the process $\{x(\tau)\}$ is Gaussian, as such an assumption is of no significance regarding linear estimation with minimum variance of estimate. In this section we consider an important aspect of Gaussian $x(\tau)$, namely that $\hat{x}(\tau + \alpha, \tau)$ and $S(\tau + \alpha, \tau)$ define completely the distribution of $x(\tau + \alpha)$ conditioned by the observation. This implies that a) the estimate of any function of $x(\tau + \alpha)$, which minimizes the expectation of any loss function, is specified as a function of $\hat{x}(\tau + \alpha, \tau)$ and $S(\tau + \alpha, \tau)$; and b) $\hat{x}(\tau + \alpha, \tau)$ is the minimax estimate, with respect to mean square error of estimate, over any class of possible distributions of $\{x(\tau)\}$ which includes Gaussian $\{x(\tau)\}$.

Asymptotically best estimation in ergodic Gaussian $\{x(\tau)\}$, treated by Wiener, 1949, will be considered briefly.

VI.1 Sufficiency of Best Linear Estimates

It is well known (cf. Doob, 1953, pp. 561-562) that the estimate minimizing variance of estimate is linear when $\{x(\tau)\}$ is Gaussian and observations are linear in $x(\tau)$, that is, that $\hat{x}(\tau + \alpha, \tau)$ is the expectation, conditioned by the observations, of $x(\tau + \alpha)$. It follows that $\hat{x}(\tau + \alpha, \tau)$ is a sufficient statistic for $x(\tau + \alpha)$, and that $\hat{x}(\tau + \alpha, \tau)$ and $S(\tau + \alpha, \tau)$ together specify the conditional distribution of $x(\tau + \alpha)$.

If we wish to estimate any function of $x(\tau + \alpha)$, say $f(x_{\tau + \alpha})$, by $f^*(x_{\tau + \alpha})$ so as to minimize the expectation of a specified loss function, say $\varphi\left[f(x_{\tau + \alpha}), f^*(x_{\tau + \alpha})\right]$, we need only minimize the expectation of φ over the conditional distribution of $x_{\tau + \alpha}$, which is defined by $\hat{x}(\tau + \alpha, \tau)$ and $S(\tau + \alpha, \tau)$. Consequently if $\{x(\tau)\}$ is Gaussian there is no need to restrict ourselves to minimizing the variance of estimate of linear functions of $x(\tau + \alpha)$; once $\hat{x}(\tau + \alpha, \tau)$ is determined we can minimize the expectation of any loss function of any function of $x(\tau + \alpha)$. The further problem is merely that of minimizing the integral:

$$\lambda(f^*) \triangleq (2\pi)^{-\nu/2} \, S^{-1/2} \int \varphi\left[f^*(x), f(x)\right] e^{-1/2\, \widetilde{(x - \hat{x})} S^{-1} (x - \hat{x})} dx.$$

In this expression we assume the time argument fixed, and also assume that S is nonsingular, possibly as a result of reduction of the number of components of the vector x. The minimization of the indicated $\lambda(f^*)$ is not typically difficult, although analytic methods will not always suffice. In many cases the problem is simplified by nulling the derivative of $\lambda(f^*)$ with respect to f^*. If $f(x)$ is a linear transformation of x, and φ is a symmetric non-negative function monotone non-decreasing with $|f^* - f|$ increasing, then $\hat{x}$ minimizes $\lambda(f^*)$.

<u>Example</u> <u>VI.1.a</u>. Suppose $f(x) = x$ and

$$\varphi(x^*,x) = \widetilde{(x^* - x)}\, C_1\, (x^* - x) e^{-1/2\, \widetilde{(x - c)}\, C_2\, (x-c)} dx,$$

with C_1 and C_2 symmetric and non-negative definite. Here c may be a vector chosen on the basis that for x near c, small errors of estimate are specially desirable. This is an example of minimization of a weighted mean square, a topic which has been given some attention recently. The specified loss function can represent quite a variety of situations, including the basic one considered in this paper ($C_2 = 0$). Letting γ represent any positive scalar not depending on x or x^*, we have

$$\lambda(x^*) = \gamma \int \widetilde{(x^* - x)} C_1 (x^* - x) e^{-1/2 \left[\widetilde{(x - c)} C_2 (x - c) + \widetilde{(x - \hat{x})} S^{-1} (x - \hat{x})\right]} dx.$$

The quadratic form in the exponent is

$$\widetilde{\left[x - (C_2+S^{-1})^{-1}(C_2c+S^{-1}\hat{x})\right]}\; (C_2c+S^{-1}) \left[x - (C_2+S^{-1})^{-1}(C_2c+S^{-1}\hat{x})\right] .$$

Thus $\lambda(x^*)$ is proportional to the expectation of $\widetilde{(x^* - x)}\, C_1\, (x^* - x)$ over x normally distributed with mean $(C_2 + S^{-1})^{-1}\, (C_2 c + S^{-1}\hat{x})$, so that this mean is the minimizing x^*. Note that x^* does not depend on C_1, and is a linear combination of $\hat{x}$ and c, replacing $\hat{x}$ by an x^* biased in the direction of c, proportionally to $c - \hat{x}$. Such a bias is to be expected on the basis of the prescribed loss function.

VI.2 <u>A Minimax Property of Best Linear Estimates</u>

We observed above that the best linear estimate is best among all estimates, if $\{x(\tau)\}$ is Gaussian. On the other hand, best nonlinear estimates are superior to best linear estimates for all but a few distributions of $\{x(\tau)\}$; Example III.4.b is a case in which the best linear estimate has asympotic efficiency of zero. In most cases the complete form of distribution of $\{x(\tau)\}$ is now known, although it is frequently assumed to be approximately Gaussian. For $\{x(\tau)\}$ satisfying the relevant hypotheses of our theorems, it can be seen that the best linear estimate is minimax over any class of $\{x(\tau)\}$ distributions including the Gaussian. This result follows at once from the facts that the variance of estimate of a linear estimate does not depend on the distribution of $x(\tau)$ except for the first two moments; and that the best linear estimate uniquely minimizes the variance of estimate for Gaussian $\{x(\tau)\}$.

VI.3 Ergodic Gaussian Processes.

The problem of best linear estimation in ergodic Gaussian processes was considered by Wiener, 1949, who derived for ergodic Gaussian observation processes,[+] the time-invariant operator asymptotically minimizing the variance of estimate of the observation, or of a desired component of the observation, called the signal. The examples considered by Wiener are discussed in Sections III, IV, and V of this paper. In one sense, the problem treated by Wiener is a special limiting case of the general problem considered in this paper, with $B(\tau)$ and $Q(\tau)$ constant and restricted so that $\{x(\tau)\}$ be ergodic, $A(\tau)$ constant from indefinitely in the past to the time of latest observation.

In another sense, Wiener's problems are more general than those of this paper, as he considered all stationary Gaussian processes possessing spectral density functions, whereas our stationary Gaussian processes cannot have spectral density functions which are not rational. This distinction does not appear to be extremely great, however, in view of the fact that every spectral density function can be approximated arbitrarily closely by a rational spectral density function. In Wiener's one example with a non-rational spectral density function, namely $e^{-\omega^2}$, he approximated by a rational function in order to predict approximately.

In estimating parameters of a distribution function, the statistician often uses inefficient estimates justified by asymptotic efficiency, when the asymptotically efficient estimate is easy to calculate and not too far from optimum. Similarly, one should often use Wiener's methods to estimate in random

[+] The results are valid also for best linear estimation with non-Gaussian and non-ergodic stationary processes.

processes when the simplicity of the method outweighs inaccuracies which in many cases are extremely minute. The reader can find excellent expositions of Wiener's method by Lee, 1960, and Darlington, 1958. A simple application of Theorem 2 to the initial observation will sometimes reduce the error associated with estimation using Wiener's methods.

With vector observation functions, there may be considerable difficulty in determining the asymptotically optimum linear operators. The techniques of this paper supply a method of determination, since the problem is reduced to that of finding the asymptotic value of the $S(\tau,\tau)$ matrix, with $dS(\tau,\tau)/d\tau = S'(\tau,\tau) + \dot{S}(\tau,\tau)$. This is a matrix Riccati equation, studied by Reid, 1946, and Levin, 1959. Writing the matrix Riccati equation as

$$d\,\Gamma(\tau)/d\tau = G_2(\tau) + G_1(\tau)\,\Gamma(\tau) - \Gamma(\tau)G_4(\tau) - \Gamma(\tau)G_3(\tau)\,\Gamma(\tau) ,$$

we have

$$\Gamma(\tau) = [M_1(\tau)\,\Gamma(0) + M_2(\tau)]\,[M_3(\tau)\,\Gamma(0) + M_4(\tau)]^{-1} ,$$

where $dM(\tau)/d\tau = G(\tau)M(\tau)$; more explicitly,

$$M(\tau) = \begin{pmatrix} M_1(\tau) & M_2(\tau) \\ M_3(\tau) & M_4(\tau) \end{pmatrix} ; \qquad \frac{dM(\tau)}{d\tau} \begin{pmatrix} G_1(\tau) & G_2(\tau) \\ G_3(\tau) & G_4(\tau) \end{pmatrix} M(\tau),$$

with $M(0) = I$, provided the inverse and integral indicated exist throughout the interval $(0,\tau)$. The proof given by Levin [1959] is as follows: substitute $\Gamma(\tau)$ into $d\,\Gamma(\tau)/d\tau$. With $\Gamma(\tau)$ representing $S(\tau,\tau)$, $G_4(\tau) = -\tilde{G}_1(\tau)$ and $G_2(\tau)$ and $G_3(\tau)$ are symmetric. In the Wiener case, the G matrix is constant, and we seek the asymptotic value of $\Gamma(\tau)$, with $\Gamma(0)$ the $S(0,0)$ determined from integration of spectral densities. The differential equation for the M matrix, $dM(\tau)/d\tau = GM(\tau)$ is a constant coefficient linear differential equation, with well known methods of solution. Thus the asymptotic value of $S(\tau,\tau)$ can be obtained routinely. We note that the differential equation for $\begin{pmatrix} M_1 \\ M_3 \end{pmatrix}$ is the same as that for $\begin{pmatrix} M_2 \\ M_4 \end{pmatrix}$, but that initial values are $\begin{pmatrix} I \\ 0 \end{pmatrix}$ and

$\begin{pmatrix} 0 \\ I \end{pmatrix}$, respectively. Note also that $S(\tau,\tau)$ is typically singular, and that computation will then be simplified by reduction to a smaller matrix; several examples of this paper are illustrative.

The Wiener methods can be extended, at least with scalar observations, to derive minimax estimation procedures, when the spectral density of the signal, the remainder of the observation (called noise), or both are not completely specified but subjected to linear restrictions [Carlton and Follin, 1956]. The minimax estimation procedures minimize the maximum of the variance of estimate over all spectral density functions satisfying the linear constraints. The constraints may be inequalities or equalities, of the form $\gamma_\alpha \int_{-\infty}^{\infty} \Phi(\omega)\, \theta_\alpha(\omega)\, d\omega = 1$, with θ_α a prescribed symmetric non-negative function, γ_α either unity or an arbitrary number ≥ 1. $\theta(\omega)$ proportional to 1, ω^2, or ω^4 represent bounded mean square magnitude, velocity, or acceleration of the process considered. The minimax linear operator is the Wiener operator based on the maximin spectral density. The maximin spectral density is such that the square of the absolute value of the resulting minimax frequency operator is a constant linear combination of the $\theta_\alpha(\omega)$ at all frequencies with nonzero spectral density, and equal to or less than this constant linear combination at frequencies with zero spectral density.

APPENDIX A

BEST ABSOLUTELY UNBIASED ESTIMATION IN WIDE SENSE MARKOV PROCESSES

Consider discrete linear observations on a process which is vector Markov in the wide sense, as in the hypotheses of Theorem 2, with $L^{-1}_{\alpha,\alpha-1}$ existing and denoted by $L_{\alpha-1,\alpha}$. Repeated application of

$$\left\{\begin{matrix} x_\alpha = L_{\alpha,\alpha-1} x_{\alpha-1} + v_\alpha \\ x_\alpha = L_{\alpha,\alpha+1} (x_{\alpha+1} + v_{\alpha+1}) \end{matrix}\right\}$$

yields $x_\alpha = L_{\alpha,\mu}\, x_\mu + \sum_{\beta=2}^{\nu} K_{\alpha,\beta,\mu}\, v_\beta$.

with $K_{\alpha,\beta,\mu} = L_{\alpha,\beta}(\alpha < \beta < \mu + 1)$ or $(\mu < \beta < \alpha + 1)$, zero otherwise.

Writing $x_\mu = \begin{pmatrix} x_{(1)} \\ x_{(2)} \end{pmatrix}$, where $x_{(1)}$ is the subvector of x_μ for which the best absolutely unbiased linear estimate is desired, and correspondingly writing

$L_{\alpha,\mu} = (L_{\alpha,\mu,(1)},\ L_{\alpha,\mu,(2)})$, we have

$$x_\alpha = L_{\alpha,\mu,(1)}\, x_{(1)} + L_{\alpha,\mu,(2)}\, x_{(2)} + \sum_{\beta=2}^{\nu} K_{\alpha,\beta,\mu}\, v_\beta .$$

Thus the observation η_α is

$$\eta_\alpha = \tilde{a}_\alpha x_\alpha = \tilde{a}_\alpha L_{\alpha,\mu,(1)} x_{(1)} + \tilde{a}_\alpha L_{\alpha,\mu,(2)} x_{(2)} + \sum_{\beta=2}^{\nu} \tilde{a}_\alpha K_{\alpha,\beta,\mu} v_\beta .$$

To obtain an absolutely unbiased estimate, we must have the "error of observation," $\eta_\alpha - \tilde{a}_{\alpha,\mu,(1)} x_{(1)}$, have zero mean for every $x_{(1)}$. The hypotheses on v_β are sufficient, provided $\tilde{a}_\alpha L_{\alpha,\mu,(2)} x_{(2)}$ has zero mean for every $x_{(1)}$. If $x_{(2)}$ has non-zero mean, one should replace $x_{(2)}$ by $x_{(2)} - Ex_{(2)}$, and η_α by $\eta_\alpha - \tilde{a}_\alpha L_{\alpha,\mu,(2)} Ex_{(2)}$. The requirement is then satisfied if the mean of the adjusted term, $\tilde{a}_\alpha L_{\alpha,\mu,(2)} [x_{(2)} - Ex_{(2)}]$, is zero for every $x_{(1)}$, a condition which surely is satisfied if $x_{(1)}$ and $x_{(2)}$ are mutually independent. The vector y can now be written as

$$y \triangleq \begin{pmatrix} \eta_1 \\ \cdot \\ \cdot \\ \eta_\nu \end{pmatrix} = Mx_{(1)} + G_1 x_{(2)} + \sum_{\beta=2}^{\nu} G_\beta v_\beta ,$$

with the α^{th} row of M, of G_1, and of $G_\beta (2 \leq \beta \leq \nu)$: $\tilde{a}_\alpha L_{\alpha,\mu,(1)}$; $\tilde{a}_\alpha L_{\alpha,\mu,(2)}$; and $\tilde{a}_\alpha K_{\alpha,\beta,\mu}$, respectively. Assuming that the mean of $G_1 x_{(2)}$ is zero for every $x_{(1)}$, and that $M\tilde{M}$ and $R \triangleq E(y - Mx_{(1)})\widetilde{(y - Mx_{(1)})}$ are nonsingular. Theorem 0 can be applied to give

$$\bar{x}_{(1)} = (\tilde{M} R^{-1} M) \tilde{M} R^{-1} y ,$$

with

$$R = E\left(G_1 x_{(2)} + \sum_{\beta=2}^{\nu} G_\beta v_\beta\right) \widetilde{\left(G_1 x_{(2)} + \sum_{\beta=2}^{\nu} G_\beta v_\beta\right)} = \tilde{G} V G ,$$

with $\tilde{G} = (G_1, G_2, \ldots, G_\nu)$,

and

$$V = \begin{pmatrix} E\, x_{(2)}\tilde{x}_{(2)} & & & 0 \\ & E\, v_2\tilde{v}_2 & & \\ & & \ddots & \\ 0 & & & E\, v_\nu\tilde{v}_\nu \end{pmatrix}$$

using the orthogonality of $x_{(2)}, v_2, \ldots, v_\nu$.

The computation to obtain $x_{(1)}$ is extensive, particularly if one wants to estimate such a subvector for various values of α. It has been exhibited to indicate the possibility of absolutely unbiased estimation in cases not generally considered parametric. There is a widespread tendency to seek best absolutely unbiased estimates of parameters, but best estimates of non-trivial random variables. The logical basis for such a tendency is not clear, as it would seem more difficult to obtain a priori information for complicated unknowns than for simple unknowns.

APPENDIX B

WIDE SENSE VECTOR MARKOV PROCESSES

Letting $\hat{E}$ represent the best linear estimate of random variables in terms of conditioning variables, the wide sense vector Markov processes are those with the property:

$$\hat{E}\,[x(\tau_\nu)\,|\,x(\tau_{\nu-1}),\,x(\tau_{\nu-2}),\,..\,,\,x(\tau_1)] = \hat{E}\,[x(\tau_\nu)\,|\,x(\tau_{\nu-1})], \qquad \text{(B-1)}$$

whenever $\tau_\nu \geq \tau_{\nu-1} \geq .. \geq \tau_1$. Considering first a fixed sequence $\tau_1, \tau_2, .. , \tau_\nu, \tau_{\nu+1}, ...$, denote $x(\tau_1), x(\tau_2)$ .. by $x_1, x_2, ...$ Then property (B1) is equivalent to

$$x_\nu = L_{\nu,\nu-1}\, x_{\nu-1} + v_\nu \quad , \qquad \text{(B-2)}$$

with v_ν orthogonal to $x_1, x_2, .. , x_{\nu-1}$ and to $v_1, v_2, .. , v_{\nu-1}$. It is evident that (B2) implies (B1). To show that (B1) implies (B2), we write (B1) as

$$\hat{E}\,(x_\nu | x_{\nu-1}, .. , x_1) = L_{\nu,\nu-1}\, x_{\nu-1}$$

implying that $x_\nu = L_{\nu,\nu-1}\, x_{\nu-1} + v_\nu$, with v_ν orthogonal to $x_{\nu-1}$. Similarly, $x_{\nu-1} = L_{\nu-1,\,\nu-2}\, x_{\nu-2} + v_{\nu-1}$, with $v_{\nu-1}$ orthogonal to $x_{\nu-2}$. Since v_ν is orthogonal to $x_{\nu-1}$, a linear combination of the orthogonal variables $x_{\nu-2}$ and $v_{\nu-1}$, it follows that v_ν is orthogonal to $x_{\nu-2}$ and $v_{\nu-1}$. Repeating this argument, it is seen that v_ν is orthogonal to $v_{\nu-1}, .. , \nu_1$, and to $x_{\nu-1}, .. , x_1$, establishing (B2).

To extend the property (B2) to all sequences $\tau_1, \tau_2, \ldots$, we write (B2) as

$$x(\tau_2) = K(\tau_2, \tau_1)x(\tau_1) + w(\tau_2,\tau_1), \qquad (\tau_2 \geq \tau_1) \tag{B-3}$$

with $w(\tau_2,\tau_1)$ orthogonal to $w(\tau_4,\tau_3)$ for $\tau_4 \geq \tau_3 \geq \tau_2 \geq \tau_1$, and orthogonal to $x(\tau)$ for $\tau \leq \tau_1$. Repeated application of (B3) shows that for $\tau_3 \geq \tau_2 \geq \tau_1$,

$$K(\tau_3,\tau_1) = K(\tau_3,\tau_2)K(\tau_2,\tau_1), \tag{B-4}$$

and it is clear from (B3) that $K(\tau,\tau) = I$, $w(\tau,\tau) = 0$. The <u>nonsingular</u> solution of the functional equation (B4), with $K(\tau,\tau) = I$, satisfies the relation

$$dK(\tau_2,\tau_1) = B(\tau_2)K(\tau_2,\tau_1), \tag{B-5}$$

with $B(\sigma)$ any function bounded over the interval of interest. (In Theorem 3, it is required that $B(\sigma)$ be bounded and piecewise continuous, a condition which seems necessary for practical applications.)

In order that the process $\{x(\tau)\}$ possess finite variance, $w(\tau_2,\tau_1)$ must have finite variance. The variance of $w(\tau_2,\tau_1)$, if continuous, can be written as

$$Ew(\tau_2,\tau_1)\tilde{w}(\tau_2,\tau_1) = \int_{\tau_1}^{\tau_2} K(\tau_2,\sigma)Q(\sigma)\tilde{K}(\tau_2,\sigma)d\sigma, \tag{B-6}$$

with $Q(\sigma)$ a non-negative definite function bounded over the interval of interest. (In Theorem 3, considering practical applications, it is required that $Q(\sigma)$ be bounded and piecewise continuous.)

From (B6), it is seen that $w'(\tau,\tau)$ does not exist unless $w(\sigma,\tau)$ is zero (with probability one) for σ in the neighborhood to the right of τ; for if it did exist, we would have

$$Q(\tau) = E'w(\tau,\tau)\tilde{w}(\tau,\tau) = Ew'(\tau,\tau)\tilde{w}(\tau,\tau) + Ew(\tau,\tau)\tilde{w}'(\tau,\tau),$$

with $Q(\tau)$ non-zero and $w(\tau,\tau)$ zero. Applying this result to the individual components of $w(\sigma,\tau)$, it is seen that no component of $x(\tau)$ can include the derivative of a component for which the corresponding component of $w(\sigma,\tau)$ is not zero in the neighborhood of τ, i.e. for which the corresponding rows and columns of $Q(\tau)$ are not zero.

The processes specified by (B3), (B5), (B6) clearly have continuous first and second moments, and thus are continuous in quadratic mean (Loeve, 1955, p. 470). In extending (B2) to apply simultaneously to all ordered times, we have imposed two restrictions: first, that the transition function $K(\tau_2,\tau_1)$ be nonsingular; second, that the variance $Ew(\tau_2,\tau_1)\tilde{w}(\tau_2,\tau_1)$ be continuous. In the vast majority of cases, these restrictions are of no significance. If $Ew(\tau_2,\tau_1)\tilde{w}(\tau_2,\tau_1)$ is not continuous, Theorem 2 can be applied at any point of discontinuity. Singular continuous transition functions typically imply trivial components of the x-vector, which present no problem.

As examples of wide sense vector Markov processes, we use in the body of the paper all the examples considered in the papers of Wiener, 1949 and Shinbrot, 1956. As another example, we mention the Gaussian vector processes $\{x(\tau)\}$ with first and second moments satisfying our restrictions. Letting $[\tau_*,\tau_{**}]$ be the interval of interest, the Gaussian processes satisfying (B3), (B5), and (B6) are those with $Ex(\tau) = K(\tau,\tau_*)Ex(\tau_*)$,

and for $\tau \leq \sigma$,

$$Ex(\tau)\tilde{x}(\sigma) = K(\tau,\tau_*)\,[Ex(\tau_*)\tilde{x}(\tau_*)]\,\tilde{K}(\sigma,\tau_*) + \int_{\tau_*}^{\tau} K(\tau,\rho)Q(\rho)\tilde{K}(\tau,\rho)d\rho \cdot \tilde{K}(\sigma,\tau) .$$

REFERENCES

1. T. W. Anderson, "Least Squares and Best Unbiased Estimates," Annals of Mathematical Statistics, 33, 1962, pages 266-272.

2. Julius Bendat, "A General Theory of Prediction and Filtering," Society for Industrial and Applied Mathematics Journal, 4, 1956, pages 131-151.

3. A. George Carlton, "Estimating the Parameters of a Rectangular Distribution," Annals of Mathematical Statistics, 17, 1946, pages 355-358.

4. A. G. Carlton and J. W. Follin, Jr., "Recent Developments in Fixed and Adaptive Filtering," AGARDograph 21, 1956.

5. Sidney Darlington, "Linear Least Squares Smoothing and Prediction, with Applications," Bell System Technical Journal, 37, 1958, pages 1221-1294.

6. R. C. Davis, "On the Theory of Prediction of Nonstationary Stochastic Processes," Journal of Applied Physics, 23, 1952, pages 1047-1053.

7. C. L. Dolph and M. A. Woodbury, "On the Relation Between Green's Functions and Covariances of Certain Stochastic Processes and Its Application to Unbiased Linear Prediction," Transactions of the American Society, 72, 1952, pages 519-550.

8. J. L. Doob, Stochastic Processes, Wiley, New York, 1953.

9. U. Grenander, "Stochastic Processes and Statistical Inference," Arkiv for Matematika, 1, 1950, pages 195-277.

10. G. Kallianpur, "A Problem in Optimal Filtering with Finite Data," Annals of Mathematical Statistics, 30, 1959, pages 659-669.

11. R. E. Kalman and R. S. Bucy, "New Results in Linear Filtering and Prediction Theory," Journal of Basic Engineering, 82, 1961, pages 95-108.

12. K. Karhunen, "Über Linear Methoden in der Wahrscheinlichkeitsrechnung," Annals of the Academia Scientiarum Fennicae, I Mathematica and Physica, 37, 1947.

13. A. Kolmogorov, "Interpolation und Extrapolation von Stationären Zufäligen Folgen," Bulletin of the Academy of Science, USSR, Series Mathematics, 5, 1941, pages 3-14.

14. Y. W. Lee, Statistical Theory of Communication, Wiley, New York, 1960.

15. J. J. Levin, "On the Matrix Riccati Equation," American Mathematical Society Proceedings, 10, 1959, pages 519-524.

16. Michel Loève, "Quelques Propriétés des Fonctions Aléatoires de Second Ordre," Comptes Rendus De L'Académie Des Sciences Paris, 1946, 222, pages 469-470.

17. Michel Loève, Probability Theory, Van Nostrand, New York, 1955.

18. Emanuel Parzen, "A New Approach to the Synthesis of Optimal Smoothing and Prediction Systems," Technical Report No. 34, Applied Mathematics and Statistics Laboratories, Stanford University, 1960.

19. V. S. Pugachew, "Optimum System Theory Using a General Bayes Criterion," Institute of Radio Engineers Transactions on Information Theory, 6, 1960, pages 4-7.

20. W. T. Reid, "A Matrix Differential Equation of Riccati Type," American Journal of Mathematics, 68, 1946, pages 237-246.

21. Marvin Shinbrot, "On a Method for Optimization of Time-Varying Linear Systems with Nonstationary Inputs," National Advisory Committee for Aeronautics Technical Note 3791, Ames Aeronautical Laboratory, Moffett Field, California, 1956.

22. Norbert Wiener, Extrapolation, Interpolation, and Smoothing of Stationary Time Series, Wiley, New York, 1949.

23. L. A. Zadeh and J. R. Ragazzini, "An Extension of Wiener's Theory of Prediction," Journal of Applied Physics, 21, 1950, pages 645-655.

DISTRIBUTION

The Bumblebee series consists of technical monographs that cover a reasonably complete unit of work and are a distinct phase of the Section T program. Official Laboratory review substantiates their technical validity and establishes suitability for distribution outside Section T.

Initial distribution of Bumblebee Report No. 311 has been made in accordance with a list on file in the Technical Reports Group of the Applied Physics Laboratory, The Johns Hopkins University.