IOWA STATE UNIVERSITY
Digital Repository

1966

Estimation of covariance parameters for an adaptive Kalman filter

Jerome Clair Shellenbarger
Iowa State University

ESTIMATION OF COVARIANCE PARAMETERS FOR AN ADAPTIVE KALMAN FILTER

by

Jerome Clair Shellenbarger

A Dissertation Submitted to the

Graduate Faculty in Partial Fulfillment of

The Requirements for the Degree of

DOCTOR OF PHILOSOPHY

Major Subject: Electrical Engineering

Approved:

In Charge of Major Work

Head of Major Department

Dean of Graduate College

Iowa State University
Of Science and Technology
Ames, Iowa

1966

TABLE OF CONTENTS

I. INTRODUCTION

One of the basic categories of control theory is termed the estimation problem. It has to do with the problem of estimating the states of a system in a stochastic environment; that is, by means of operations performed upon the output measurements of the system, estimates of the stochastic signals and/or stochastic disturbances are formed. The operations are functions of the statistical properties of the signals and disturbances, the nature of the system, and the relationships of the measurements to the system, as determined by the solution of some criterion for what constitutes good, or optimum, estimates.

The initial significant work on this problem was by Wiener (7) who developed the condition to be satisfied for optimal estimation in a mean-squared-error sense; this condition is generally referred to as the Wiener-Hopf integral equation. He also developed the solution for the case of stationary, Gaussian statistics. This work and further extensions and modifications by others are known as "Wiener filters". The use of transformations into the frequency domain characterizes much of the work, and the results are usually implemented as linear, analog-type filters. Non-time-stationary and multiple input-output problems are difficult to solve by the Wiener approach.

Kalman (3) treated this estimation problem from a different point of view and formulated the equivalent of the Wiener-Hopf integral equation as a vector-matrix differential equation in state space. He developed the solution for the discrete, linear system with Gaussian statistics as a set of vector-matrix recursive relationships which are commonly termed the "Kalman filter". The advantages of the Kalman filter are that the

computations are performed recursively, in the time domain (thereby being well-suited for handling by computers), and are readily applicable to non-stationary and multiple input-output systems.

A host of study areas are closely related to, or are sub-categories of, the Kalman-filter theory; some of them treat problems of stability, sensitivity, smoothing, prediction, effects of different types of sampling, linear approximations, system identification, and parameter estimation. It is this last topic which is the subject of this dissertation, so further explanation is in order. It is assumed that a Kalman filter is to be developed for use on some linear system about which the following statements can be made:

1. The system is completely specified, including the "filters" needed to convert Gaussian "white-noise" sources into the actual inputs.

2. The measurement device is specified.

3. The root-mean-square amplitudes of the inputs may or may not be known.

4. The measurement errors have Gaussian, time-independent statistics with zero means and variances which may or may not be known.

In order to apply the Kalman filter, the covariances of the measurement errors and the inputs are required. If either, or both, covariance matrices are unknown, due to lack of information as mentioned in parts 3 and 4 above, then covariance parameter estimation must be employed.

Methods are developed in the later sections which enable one to make estimates of the necessary covariance parameters when the measurement-error and/or the input covariances are missing. These estimates cause the Kalman filter to be adaptive to unknown statistics, provided the statistics

do not fluctuate too rapidly. The estimation of the covariances of measure-
ables is covered in nearly any standard text on statistics, but so far as
is known to this author, no literature exists which expands the concepts
to enable estimation of the two principle covariance matrices associated
with the control-system and measurement equations.

The principle of maximum likelihood is used as a starting point for
the development of the estimation equations. Certain compromises are made
in order to obtain usable results. Several situations are studied, per-
taining to which particular set of covariances is sought, measurement
redundancy, observability of the system, etc. For each situation, two
estimation equations are developed and examined; one being rather crude
but easy to apply, the other holding promise of greater accuracy, but
being significantly more elaborate.

II. REVIEW OF KALMAN FILTERING

A review of Kalman filtering would seem to be in order, primarily so that many of the necessary symbols can be properly introduced and defined. In addition, the Kalman-filter results are required later.

The system itself is characterized by a vector-matrix differential equation (6,8),

$$\dot{x}(t) = A(t)x(t) + B(t)u(t) \qquad (1)$$

where $x(t)$ is a k-vector of the state variables

         $u(t)$ is an m-vector of the Gaussian white-noise, independent inputs

         $A(t)$ is a k-by-k coefficient matrix for the system

         $B(t)$ is a k-by-m input matrix.

The general solution of this equation is expressed as

$$x(t) = \Phi(t,t_0)x(t_0) + \int_{t_0}^{t} \Phi(t,\tau)B(\tau)u(\tau) \, d\tau \qquad (2)$$

where $\Phi(t,t_0)$ is the k-by-k transition matrix.

For convenience, the integral in the general solution is generally designated by a k-element response vector, which will be defined here as

$$g(t,t_0) = \int_{t_0}^{t} \Phi(t,\tau)B(\tau)u(\tau) \, d\tau \qquad (3)$$

Thus $x(t) = \Phi(t,t_0)x(t_0) + g(t,t_0)$ (4)

The measurements are related to the states of the system by the equation

$$y(t) = M(t)x(t) + v(t) \qquad (5)$$

where $y(t)$ is an r-vector of the measurements

v(t) is an r-vector of the measurement errors

M(t) is an r-by-k measurement matrix.

Since the measurements are made at discrete instants of time, a subscript notation is used in the equations to describe the system and the measurement relation at the times of the measurements.

$$y_n = M_n x_n + v_n$$
$$x_n = \Phi_n x_{n-1} + g_n$$
$$\left.\right\} \quad (6)$$

where $\quad y_n = y(t_n), \ x_n = x(t_n), \ v_n = v(t_n), \ M_n = M(t_n), \ \Phi_n = \Phi(t_n, t_{n-1}),$

$g_n = g(t_n, t_{n-1}).$

The circumflex, $\wedge$, and the tilde, $\sim$, are used to denote an estimated quantity and an error quantity, respectively. Thus the state vector, x_n, can be expressed as the sum of an estimate and an estimation error.

$$x_n = \hat{x}_{n/j} + \tilde{x}_{n/j} \qquad (7)$$

The subscript n/j denotes that the variable was obtained as the result of an estimation of the state at time t_n by using measurements made up to and including those at time t_j. If $j \geq n$, then $\hat{x}_{n/j}$ is commonly termed the a posteriori estimate of x_n. If $j < n$, then $\hat{x}_{n/j}$ is called the a priori estimate of x_n. Another way to classify the various situations is to say that if $j = n$, then filtering is being accomplished; if $j > n$, it is smoothing; and if $j < n$, it is predicting.

The Kalman filter equations can be derived by several different approaches. Some of these are summarized and compared by Lee (4). A method divised by Battin (1) relies on assuming the form of the estimate as

$$\hat{x}_{n/n} = \hat{x}_{n/n-1} + W_n(y_n - \hat{y}_{n/n-1}) \qquad (8)$$

where W_n is a k-by-r weighting matrix,

$$\left. \begin{array}{l} \hat{x}_{n/n-1} = \Phi_n \ \hat{x}_{n-1/n-1} \\[2mm] \hat{y}_{n/n-1} = M_n \ \hat{x}_{n/n-1} \end{array} \right\} \quad (9)$$

and W_n is determined by minimizing the estimation-error variances of the states. By definition,

$$P_{n/j} = E[\tilde{x}_{n/j} \ \tilde{x}_{n/j}^T] \qquad (10)$$

is the estimation-error covariance matrix of the states, so the criterion is

$$\frac{\partial \ tr[P_{n/n}]}{\partial \ W_n} = 0 \qquad (11)$$

When this operation is carried out, a set of recursive equations are obtained which, with equations 8 and 9, constitute the Kalman filter.

$$\left. \begin{array}{l} W_n = P_{n/n-1} \ M_n^T \ (M_n \ P_{n/n-1} \ M_n^T + V_n)^{-1} \\[3mm] P_{n/n-1} = \Phi_n \ P_{n-1/n-1} \ \Phi_n^T + H_n \\[3mm] P_{n/n} = P_{n/n-1} - W_n(M_n P_{n/n-1} \ M_n^T + V_n) \ W_n^T \end{array} \right\} \quad (12)$$

where $V_n = E[v_n v_n^T]$, the measurement-error covariance

   $H_n = E[g_n g_n^T]$, the response covariance

and $0 = E[v_i g_j^T]$, for all i, j

   $0 = E[v_i v_j^T]$, for i ≠ j

   $0 = E[g_i g_j^T]$, for i ≠ j.

The parameters which must be known in order to apply the Kalman filter are the transition matrix, Φ, the measurement matrix, M, the measure-

ment-error covariance matrix, V, the response covariance matrix, H, plus

initial conditions. If any of these quantities are inaccurately known,

but are used in these equations anyway, the resultant estimates of the

states will not be optimum and the estimation-error covariances will be

inaccurate.

Methods for estimating V and/or H are developed in the sections

which follow.

III. SELECTION OF THE ESTIMATION CRITERION

In order to make estimations of the V and/or H matrices, a criterion must be selected. For example, the criterion used in section II for the development of the Kalman filter was the minimization of the sum of the estimation-error variances with respect to the weighting matrix. A commonly-used statistical method of estimation is known as the principle of maximum likelihood (5), and it would appear that this method offers a way to approach the problem.

The likelihood function is the probability density function of the measurements, conditional upon the quantities to be estimated. For a single measurement vector, y_n, the likelihood function is simply the Gaussian density function for a variable with zero mean.

$$p(y_n) = \frac{1}{(2\pi)^{\frac{r}{2}} |Y_n|^{\frac{1}{2}}} e^{-\frac{1}{2}(y_n^T Y_n^{-1} y_n)} \tag{13}$$

where

$$Y_n = E[y_n y_n^T] = M_n X_n M_n^T + V_n$$
$$X_n = E[x_n x_n^T] = \Phi_n X_{n-1} \Phi_n^T + H_n \tag{14}$$

For two measurement vectors, y_n and y_{n-1}, it is a joint density function.

$$p(y_n, y_{n-1}) = p(y_n | y_{n-1}) \cdot p(y_{n-1})$$

$$= p(y_{n-1}) \frac{1}{(2\pi)^{\frac{r}{2}} |C_n|^{\frac{1}{2}}} e^{-\frac{1}{2}(\tilde{y}_{n/n-1}^T C_n^{-1} \tilde{y}_{n/n-1})} \tag{15}$$

where

$$\tilde{y}_{n/n-1} = y_n - \hat{y}_{n/n-1}$$

$$C_n = M_n P_{n/n-1} M_n^T + V_n \tag{16.}$$

For larger numbers of measurement vectors, the likelihood function is a joint density function of increased complexity, but still it can be expressed without too much trouble by extension of the previous work.

$$p(y_n, y_{n-1}, \ldots, y_i) = p(y_n | y_{n-1}, \ldots, y_i) \, p(y_{n-1}, \ldots, y_i)$$

$$= (2\pi)^{-\frac{r}{2}(n+1-i)} |Y_i|^{-\frac{1}{2}} e^{-\frac{1}{2}(y_i^T Y_i^{-1} y_i)}$$

$$\cdot \prod_{j=i+1}^{n} [|C_j|^{-\frac{1}{2}} e^{-\frac{1}{2}(\tilde{y}_{j/j-1}^T C_j^{-1} \tilde{y}_{j/j-1})}] \tag{17}$$

Suppose that an estimate for V_n is desired. If the single-measurement likelihood function is maximized with respect to V_n and solved for the estimate, $\hat{V}_n$, there would appear to be no problem. But if a multi-measurement likelihood function is similarly maximized, the solution for $\hat{V}_n$ depends upon $V_{n-1}, \ldots, V_i$ which are not known. The proper procedure would be to maximize the likelihood function with respect to all the V's involved, V_n, $V_{n-1}, \ldots, V_i$, and solve the resulting set of n+1-i vector-matrix equations for the $\hat{V}$'s; or, if it were known that V were constant or very nearly constant, replace V_n, $V_{n-1}, \ldots, V_i$ by just V, maximize with respect to V, and solve the resultant high-order equation for $\hat{V}$. Neither of these procedures is feasible for any nontrivial system.

The objective is to obtain relatively simple expressions for the estimates; expressions which are recursive would be especially useful for computer applications. Since it is not practical to form estimates of

$V_{n-1}, \ldots, V_i$ at the same time V_n is being estimated, then previously obtained estimates must be used. V_n is incorporated into the multi-measurement likelihood function only by way of the factor $p(y_n | y_{n-1}, \ldots, y_i)$ so maximizing the likelihood function with respect to V_n is the same as maximizing this conditional probability density function. The procedure can thus be expressed as performing the following operations and solving for $\hat{V}_n$:

$$\left. \frac{\partial \; p(y_n | y_{n-1}, \ldots, y_i)}{V_n} \right|_{\substack{V_n \to \hat{V}_n \\ P_{n/n-1} \to \overline{P}_{n/n-1}}} = 0 \qquad (18)$$

where

$$\overline{P}_{n/n-1} = \left. P_{n/n-1} \right|_{V_{n-1} \to \overline{V}_{n-1}, \; P_{n-1/n-2} \to \overline{P}_{n-1/n-2}}$$

$$\overline{V}_{n-1} = [\text{an "average" of } \hat{V}_{n-1}, \ldots, \hat{V}_i]$$

The "average" will be discussed further in the next section.

Entirely similar expressions can readily be made if H_n is to be estimated. Due to the dependence of X_n upon $H_n, H_{n-1}, \ldots, H_2, H_1$, even use of the single-measurement likelihood function must rely upon previous estimates of the H's.

Thus there are two possible ways of forming estimates of the covariance matrices. One way is to simply maximize the single-measurement likelihood function, $p(y_n)$, to form estimates for each of the measurements, and then average. The other is to maximize the conditional probability density function, $p(y_n | y_{n-1}, \ldots, y_i)$, to form estimates, and then average. The first is overly simple, but useful in some respects; the second ought

to result in better estimates, but at the cost of added complication. In order to avoid confusion with the normal method of maximum likelihood estimation, these two estimation methods will hereafter be referred to as maximum-probability (MP) estimation and maximum-conditional-probability (MCP) estimation, respectively.

IV. ESTIMATION OF THE MEASUREMENT-ERROR COVARIANCE

The first situation to be postulated is one in which all the parameters necessary to the Kalman filter are known, except for the measurement-error covariance matrices, the V's.

A. Maximum-Probability Estimation of V

The MP estimation of V involves the performance of the following operation,

$$\left. \frac{\partial \ p(y_n)}{\partial \ V_n} \right|_{V_n \to \hat{V}_n} = 0 \tag{19}$$

and solving for $\hat{V}_n$. The differentiations of various matrix expressions by matrices are performed in Appendix A and the results of part 9 have direct bearing on equation 19, leading to

$$0 = Y_n^{-1} - Y_n^{-1} \ y_n y_n^T \ Y_n^{-1} \Big|_{V_n \to \hat{V}_n} \tag{20}$$

$$0 = Y_n - y_n y_n^T \Big|_{V_n \to \hat{V}_n}$$

$$\hat{V}_n = y_n y_n^T - M_n X_n M_n^T \tag{21}$$

As can readily be seen, this is quite a rough estimate of V_n since only one measurement vector, y_n, is used. In fact, as it stands, some of the major diagonal elements of $\hat{V}_n$ may be negative-valued unless an additional restriction is imposed. However, this is of no great concern since it is highly unlikely that such a simple estimate would ever be used by itself as a reasonable estimate of V_n.

The expected value of this estimate is

$$E[\hat{V}_n] = E[y_n y_n^T - M_n X_n M_n^T]$$

$$= Y_n - M_n X_n M_n^T$$

$$= V_n \tag{22}$$

so the estimate is unbiased. The variance of the estimate is an item of interest, but just what is meant by the word "variance" when it is applied to a matrix? Here it will be used to denote a matrix whose elements are the variances of the corresponding elements of the estimate. In this instance, it is

$$(\text{variance of } \hat{V}_n) = E[(\hat{V}_n - V_n) * (\hat{V}_n - V_n)]$$

$$= Y_n * Y_n + c[Y_n] \cdot c[Y_n]^T \tag{23}$$

The non-standard matrix operations, * and c[ ], are explained in Appendix B. The evaluation of the variance is performed in Appendix C, part 1.

The variance matrix is useful for an element-by-element examination of the estimate, but what is usually sought is a scalar measure that shows just how closely the entire estimate comes to agreeing with the parameter being estimated. In section II, such a measure for the estimate of the state vector was the trace of the state-error covariance matrix. A similar measure can be specified for the estimate of the measurement-error covariance matrix. It is the trace of the expected-value of the quadradic estimation-error matrix. By quadradic estimation-error matrix is meant the product, $(\hat{V}_n - V_n)(\hat{V}_n - V_n)$, and the expected value of this is

$$E[(\hat{V}_n - V_n)(\hat{V}_n - V_n)] = Y_n Y_n + Y_n \, \text{tr}(Y_n) \tag{24}$$

as developed in Appendix C, part 2. Each major-diagonal element of this

matrix is the sum of the variance elements of the corresponding row or column of the variance matrix. Thus the trace of this matrix is equal to the sum of all the variance elements. This measure will be referred to as the $(\text{norm})^2$ of the estimation variances.

$$[(\text{norm})^2 \text{ of the variance of } \hat{V}_n] = \text{tr}(Y_n Y_n) + [\text{tr}(Y_n)]^2 \qquad (25)$$

In some respects the variance matrix is quite useful, but the quadradic matrix uses conventional notation and so is easier to manipulate. The $(\text{norm})^2$ can be expressed as either the trace of the expected quadradic matrix or as the variance matrix, pre- and post-multiplied by the unit vector (see Appendix B).

By way of comparison, note that if the measurement-error vector, v_n, were directly available for use in estimating V_n by letting $\hat{V}_n = v_n v_n^T$, then the corresponding equations would be

$$\left.\begin{aligned}
(\text{variance of } \hat{V}_n) &= V_n * V_n + c[V_n] \cdot c[V_n]^T \\
E[\text{quadradic of } (\hat{V}_n - V_n)] &= V_n \cdot V_n + V_n \, \text{tr}(V_n) \\
[(\text{norm})^2 \text{ of variance of } \hat{V}_n] &= \text{tr}(V_n \cdot V_n) + [\text{tr}(V_n)]^2
\end{aligned}\right\} \qquad (26)$$

1. Averaged estimation for time-stationary V

The usual estimation procedure, when it is known that the measurements are independent and the statistics are time-stationary, is to form an average. For example, if a number, m, of the measurement-error vectors were available, the principle of maximum likelihood would automatically lead one to the development of a "best" estimate that is

$$(\hat{V})_{\text{best}} = \overline{V} = \frac{1}{m} \sum_{i=1}^{m} (v \, v^T)_i \qquad (27)$$

$$[(norm)^2 \text{ of variance of } \bar{V}] = \frac{1}{m}[tr(V \cdot V) + [tr(V)^2] \tag{28}$$

Similarly, it would seem likely that an average of the individual estimates obtained from the measurement vectors would be a better overall estimate of a time-stationary V.

$$\bar{V}_n = \frac{1}{n} \sum_{i=1}^{n} (\hat{V}_i) = \frac{1}{n} \sum_{i=1}^{n} (yy^T - MXM^T)_i = \frac{1}{n} \hat{V}_n + (\frac{n-1}{n})\bar{V}_{n-1} \tag{29}$$

Since the measurement vectors are not independent, the expression for the norm will involve cross-product terms.

$$[(norm)^2 \text{ of } \bar{V} - variance] = tr[E\{[\frac{1}{n} \sum_{i=1}^{n} (\hat{V}_i - V)] \cdot [\frac{1}{n} \sum_{j=1}^{n} (\hat{V}_j - V)]\}]$$

$$= \frac{1}{n^2} tr\{E[\sum_{i=1}^{n} (\hat{V}_i - V)(\hat{V}_i - V)] + 2E[\sum_{i=2}^{n} \sum_{j=1}^{i-1} (\hat{V}_i - V)(\hat{V}_j - V)]\} \tag{30}$$

These cross-product terms can be evaluated as follows:

$$E[(\hat{V}_i - V)(\hat{V}_j - V)]_{i>j} = E[(yy^T - Y)_i (yy^T - Y)_j] = E[(yy^T)_i (yy^T)_u] - Y_i Y_j$$

$$= M_i \Phi_i \cdots \Phi_{j+1} [X_j (\Phi_{j+1}^T \cdots \Phi_i^T M_i^T M_j)^T X_j$$

$$+ X_j tr(\Phi_{j+1}^T \cdots \Phi_i^T M_i^T M_j X_j)]M_j^T$$

$$= (M_i \Phi_i \cdots \Phi_{j+1} X_j M_j^T)[(M_i \Phi_i \cdots \Phi_{j+1} X_j M_j^T)$$

$$+ I \, tr(M_i \Phi_i \cdots \Phi_{j+1} X_j M_j^T)]$$

Therefore,

$$[(\text{norm})^2 \text{ of } \overline{V}_n - \text{variance}] = \frac{1}{n^2}\{ \sum_{i=1}^{n} [\text{tr}(Y_i Y_i) + (\text{tr}[Y_i])^2]$$

$$+ 2 \sum_{i=2}^{n} \sum_{j=1}^{i-1} [\text{tr}\{(M_i \Phi_i \cdots \Phi_{j+1} X_j M_j^T)(M_i \Phi_i \cdots \Phi_{j+1} X_j M_j^T)\}$$

$$+ \{\text{tr}(M_i \Phi_i \cdots \Phi_{j+1} X_j M_j^T)\}^2]\} \tag{31}$$

The number of terms to be handled varies as n^2 and the coefficient of the overall experession is $\frac{1}{n^2}$, so convergence of $\overline{V}_n$ to the value V is not automatically guaranteed. The requirement for a system to be asymptotically stable (8) is that the expression $(\Phi_i \cdots \Phi_{j+1})$ approach the null matrix as the difference between i and j becomes large. Thus, if the system is asymptotically stable, the cross-product terms in the equation for the norm become insignificant if i>>j and the remaining terms do not increase in magnitude indefinitely, so convergence does occur.

2. Averaged estimation for slow-time-varying V

For those instances where V may be, or is known to be, slowly varying with time, other "averages" are more appropriate. The average used should have the property of being unbiased and should produce a result that depends most strongly on the latest measurements.

a. Truncated average   This average incorporates only the "p" most-recent estimates,

$$\left. \begin{aligned}
\overline{V}_{n;p} &= \frac{1}{p} \sum_{i=n-p+1}^{n} (\hat{V}_i) \ , \ p \le n \\
&= \frac{1}{p} \hat{V}_n + \overline{V}_{n-1;p} - \frac{1}{p} \hat{V}_{n-p}
\end{aligned} \right\} \tag{32}$$

As can be seen though, this average, when expressed in the recursive form,

requires that the most-recent p+1 estimates of V be kept in a "memory".

b. Exponentially-weighted average   By assigning exponential

weights to the measurements as follows, the more recent measurements tend

to determine the nature of the result.

$$
\left.
\begin{aligned}
(V_{n;p})_{exp.} &= \left[\frac{1-e^{-\frac{1}{p}}}{1-e^{-\frac{n+1}{p}}}\right] \sum_{i=1}^{n} (\hat{V}_i \, e^{-\frac{n-i}{p}}) \\
&= \left[\frac{1-e^{-\frac{1}{p}}}{1-e^{-\frac{n+1}{p}}}\right]\hat{V}_n + \left[\frac{1-e^{-\frac{n}{p}}}{1-e^{-\frac{n+1}{p}}}\right](\overline{V}_{n-1;p})_{exp.}
\end{aligned}
\right\} \quad (33)
$$

This average avoids the need for "memory", but the evaluation of the ex-

ponential terms may be disadvantageous.

c. Limiting case of exponentially-weighted average   If the "p"

used in the exponential average is allowed to become very large, then that

average becomes

$$
(\overline{V}_{n;p})_{exp.lim.} = \frac{1}{p}\hat{V}_n + \frac{p-1}{p}(\overline{V}_{n-1;p})_{exp.lim.} \quad (34)
$$

Examination shows that this average is valid even if p is not large.

The expression for the norm can be determined readily for the truncat-

ed average, but some complications arise for the other two averages. The

requirement for the system to be asymptotically stable in order for con-

vergence to occur also holds for these averages.

B. Maximum-Conditional-Probability Estimation of V

The MCP estimation of V involves the performance of the following

operation,

$$\left. \frac{\partial\ p(y_n | y_{n-1}, \ldots, y_1)}{\partial\ V_n} \right|_{\substack{V_n \rightarrow \hat{V}_{n/n} \\ \\ P_{n/n-1} \rightarrow \overline{P}_{n/n-1}}} = 0 \qquad (35)$$

and solving for $\hat{V}_n$. Carrying out the differentiation results in the expression

$$0 = C_n^{-1} - C_n^{-1} \tilde{y}_{n/n-1} \tilde{y}_{n/n-1}^T C_n^{-1} \Big|_{\substack{V_n \rightarrow \hat{V}_{n/n} \\ \\ P_{n/n-1} \rightarrow \overline{P}_{n/n-1}}} \qquad (36)$$

and

$$\hat{V}_{n/n} = \tilde{y}_{n/n-1} \tilde{y}_{n/n-1}^T - M_n \overline{P}_{n/n-1} M_n^T \qquad (37)$$

If the initial state estimation-error covariance, $P_{1/0}$, is known then

$$E[\hat{V}_{1/1}] = V_1 \qquad (38)$$

and by induction, using the recursive equation-set 12,

$$E[\hat{V}_{n/n}] = V_n \qquad (39)$$

Thus this estimator is also unbiased.

The variance matrix and norm cannot be easily evaluated for this estimate because $\overline{P}_{n/n-1}$ depends upon all the preceding estimates in quite a complicated fashion. However, since $P_{n/n-1}$ is optimum, the minimum possible norm would occur if $\overline{P}_{n/n-1}$ equaled $P_{n/n-1}$. Thus

$$[(\text{norm})^2 \text{ of } \hat{V}_{n/n} - \text{variance}]_{\text{min.}} = tr(C_n C_n) + [tr(C_n)]^2 \qquad (40)$$

It may be noticed that Y_n and C_n differ only to the extent that the X_n in the Y_n expression is replaced by the $P_{n/n-1}$ in the C_n expression:

$$Y_n = V_n + M_n X_n M_n^T$$

$$C_n = V_n + M_n P_{n/n-1} M_n^T$$

$$\qquad\qquad (41)$$

and, since

$$\text{tr}(P_{n/n-1} P_{n/n-1}) \leq \text{tr}(X_n X_n) \qquad\qquad (42)$$

The MCP estimator should be better than the MP estimator, provided that $\bar{P}_{n/n-1}$ does not exhibit poor "transient" behavior when only a small number of measurements are available. In a particular situation, a running evaluation and comparison can be made of X_n and $\bar{P}_{n/n-1}$ so that a decision can be made as to which estimation method should be used.

The sequence of steps to be performed for making MCP estimates of the V's of a particular system is as follows:

1. Determine the initial state vector, x_0, as accurately as possible, using whatever technique is available. Also determine the co-variance matrix for the error in this estimate of x_0 as well as the circumstances permit; i.e., estimate $P_{0/0}$.

2. Calculate these predictions:

$$\hat{x}_{1/0} = \Phi_1 \hat{x}_0 \ , \ \hat{y}_{1/0} = M_1 \hat{x}_{1/0}, \ \tilde{y}_{1/0} = y_1 - \hat{y}_{1/0}$$

$$\bar{P}_{1/0} = \Phi_1 \hat{P}_{0/0} \Phi_1^T + H_1$$

3. Estimate V_1.

$$\hat{V}_1 = \tilde{y}_1 \tilde{y}_1^T - M_1 \bar{P}_{1/0} M_1^T$$

4. Form an average estimate for V_1.

$$\bar{V}_1 = \frac{1}{q+1} \hat{V}_1 + \frac{q}{q+1} \bar{V}_0$$

where $\bar{V}_0$ is an initial rough estimate of V, and q is a scalar weight that indicates the degree of confidence in $\bar{V}_0$ with respect

to $\hat{V}_1$.

5. Calculate $\overline{W}_1$.

$$\overline{W}_1 = \overline{P}_{1/0}M_1^T(M_1\overline{P}_{1/0}M_1^T + \overline{V}_1)^{-1}$$

Note that $\hat{V}_1$ cannot be used in place of $\overline{V}_1$ in this expression because then $(M_1P_{1/0}M_1^T + \hat{V}_1) = \tilde{y}_{1/0}\tilde{y}_{1/0}^T$ which is singular.

6. $\overline{P}_{1/1} = \overline{P}_{1/0} - \overline{W}_1(M_1\overline{P}_{1/0}M_1^T + \overline{V}_1)\overline{W}_1^T$

7. $\hat{x}_{1/1} = \hat{x}_{1/0} + \overline{W}_1\tilde{y}_{1/0}$

8. $\overline{P}_{2/1} = \Phi_2\overline{P}_{1/1}\Phi_2^T + H_2$

9. $\hat{x}_{2/1} = \Phi_2\hat{x}_{1/1}$, $\hat{y}_{2/1} = M_2\hat{x}_{2/1}$, $\tilde{y}_{2/1} = y_2 - \hat{y}_{2/1}$

10. $\hat{V}_2 = \tilde{y}_{2/1}\tilde{y}_{2/1}^T - M_2\overline{P}_{2/1}M_2^T$

11. Form an average estimate for V_2 in any of several different ways, depending upon whether V is, or is not, known to be time-stationary. For example,

$$\overline{V}_2 = (\frac{1}{q+2})\hat{V}_2 + (\frac{q+1}{q+2})\overline{V}_1$$

is a suitable average if V is time-stationary.

12. Continue in this fashion by cycling back to step 5 and incrementing the subscripts by one.

V. ESTIMATION OF THE RESPONSE COVARIANCE

The situation postulated in this section is that all the parameters necessary to the Kalman filter are known, except for the response covariance matrices, the H's.

A. Maximum-Probability Estimation of H

The MP estimation of H requires the performance of the following operation,

$$\left. \frac{\partial\ p(y_n)}{\partial\ H_n} \right|_{\substack{H_n \to \hat{H}_n \\ X_{n-1} \to \bar{X}_{n-1}}} = 0 \tag{43}$$

where

$$\bar{X}_i = \Phi_i \bar{X}_{i-1} \Phi_i^T + \bar{H}_i$$

The result is

$$0 = M_n^T Y_n^{-1}(Y_n - y_n y_n^T) Y_n^{-1} M_n \Big|_{\substack{H_n \to \hat{H}_n \\ X_{n-1} \to \bar{X}_{n-1}}} \tag{44}$$

The nature of the solution for $\hat{H}_n$ depends greatly upon the dimensionality of the measurement matrix, M_n. M_n is an r-by-k matrix and, for convenience, it will be assumed that it is of maximum possible rank; that is, the rank is either r or k, whichever is smaller. The descriptive names applied to M_n are: "square" if r equals k, "horizontal" if r is less than k, and "vertical" if r is greater than k.

1. Solution when M is square

When M_n is square, the solution for $\hat{H}_n$ is quite simple.

$$0 = (Y_n - y_n y_n^T)\Big|_{\substack{H_n \to \hat{H}_n \\ X_{n-1} \to \overline{X}_{n-1}}}$$

$$M_n \hat{H}_n M_n^T = y_n y_n^T - V_n - M_n \Phi_n \overline{X}_{n-1} \Phi_n^T M_n^T$$

$$\hat{H}_n = (M_n)^{-1}(y_n y_n^T - V_n - M_n \Phi_n \overline{X}_{n-1}\Phi_n^T M_n^T)(M_n^T)^{-1} \tag{45}$$

If the initial state covariance, X_0, is known, then

$$E[\hat{H}_1] = H_1 \tag{46}$$

and by induction,

$$E[\hat{H}_n] = H_n \tag{47}$$

so the estimators are unbiased for known initial conditions.

The norm, for n equal to 1, is developed in Appendix D, part 1.

$$[(norm)^2 \text{ of } \hat{H}_1 - variance] = tr[(M^{-1}YM^{T^{-1}})_1 (M^{-1}YM^{T^{-1}})_1]$$

$$+ [tr(M^{-1}YM^{T^{-1}})_1]^2 \tag{48}$$

It becomes convenient to define a certain function as follows: Let

$$T[Z] = tr[Z \cdot Z] + [tr(Z)]^2 \tag{49}$$

Then equation 48 can be expressed more compactly as

$$[(norm)^2 \text{ of } \hat{H}_1 - variance] = T[(M^{-1}YM^{T^{-1}})_1] \tag{50}$$

The norm for other values of n depends upon just how $\overline{X}$ is formed. The

estimate of H_n involves $\bar{X}_{n-1} = \Phi_{n-1}\bar{X}_{n-2}\Phi_{n-1}^T + \bar{H}_{n-1}$, where $\bar{H}_{n-1}$ is some sort of an average of $\hat{H}_{n-1}$, $\hat{H}_{n-2}$, etc. This average greatly complicates the expression for the norm. However, if the worst-possible average is used, namely $\bar{H}_{n-1} = \hat{H}_{n-1}$, then the norm can be determined rather easily, as shown in Appendix D, parts 2, 3, and 4.

$$[(\text{norm})^2 \text{ of } \hat{H}_n - \text{variance}]_{max} = T[M_n^{-1}Y_n M_n^{T^{-1}}] - 2T[\Phi_n X_{n-1}\Phi_n^T]$$

$$+T[\Phi_n M_{n-1}^{-1}Y_{n-1}M_{n-1}^{T^{-1}}\Phi_n^T] \qquad (51)$$

The norm of the variance of $\bar{H}_n$, with $\bar{X}_{n-1} = \Phi_{n-1}\bar{X}_{n-2}\Phi_{n-1}^T + \hat{H}_{n-1}$, is expressed as

$$[(\text{norm})^2 \text{ of } \bar{H}_n - \text{variance}]_{max} = \text{tr}\{E[\frac{1}{n}\sum_{i=1}^{n}(\hat{H}_i-H_i)\cdot\frac{1}{n}\sum_{j=1}^{n}(\hat{H}_j-H_j)]\}$$

$$= \frac{1}{n^2}\text{tr}\{E[\sum_{i=1}^{n}(\hat{H}_i-H_i)(\hat{H}_i-H_i)$$

$$+ 2\sum_{i=2}^{n}\sum_{j=1}^{i-1}(\hat{H}_i-H_i)(\hat{H}_j-H_j)]\} \qquad (52)$$

But, results obtained in Appendix D, part 5, show that many of the cross-product terms are zero, so

$$[(\text{norm})^2 \text{ of } \bar{H}_n - \text{variance}]_{max} = \frac{1}{n^2}\text{tr}\{E[\sum_{i=1}^{n}(\hat{H}_i-H_i)(\hat{H}_i-H_i)$$

$$+ 2\sum_{i=2}^{n}(\hat{H}_i-H_i)(\hat{H}_{i-1}-H_{i-1})]\} \qquad (53)$$

$$
\begin{aligned}
= \frac{1}{n^2}\{ \sum_{i=1}^{n} & [T[M_i^{-1}Y_i M_i^{T^{-1}}] + T[\Phi_i M_{i-1}^{-1}Y_{i-1}M_{i-1}^{T^{-1}}\Phi_i^T] \\
& - 2T[\Phi_i X_{i-1}\Phi_i^T]] + 2 \sum_{i=2}^{n} [T[\Phi_i X_{i-1}] \\
& -T[\Phi_i M_{i-1}^{-1}Y_{i-1}M_{i-1}^{T^{-1}}]]\}
\end{aligned}
\tag{54}
$$

Since the number of terms varies as some multiple of n and the overall summation is multiplied by $\frac{1}{n^2}$, then at least one requirement for convergence of $\hat{H}_n$ is satisfied. If, also, the system is asymptotically stable, then $\overline{H}_n$ does converge to H_n.

There are a few points which should be examined if a number of estimates of H are to be averaged together in the expectation of converging towards the true response matrix. For the averaging process to be at all successful, H must be virtually constant during the effective averaging period. Examine the complete equation for H_n:

$$
\begin{aligned}
H_n = \int_{t_{n-1}}^{t_n} \int_{t_{n-1}}^{t_n} & \Phi(t_n,\tau_1)B(\tau_1)E[u(\tau_1)u^T(\tau_2)] \\
& \cdot B^T(\tau_2)\Phi^T(t_n,\tau_2)d\tau_1\, d\tau_2
\end{aligned}
\tag{55}
$$

Those factors which must remain virtually constant (assuming that a variation in one factor is not offset by variations of others) are the interval between measurements, the transition matrix for a full interval, the input matrix, and the input variances.

2. Solution when M is vertical

When M_n is vertical, i.e., redundant measurements are made, the solution of equation 44 for $\hat{H}_n$ is a bit more difficult. A matrix identity

from Appendix E is used to convert equation 44 to a form which can be simplified by pre- and post-multiplications.

$$M^T Y^{-1} = M^T (MHM^T + Q)^{-1} \tag{56}$$

where

$$Q_n = V_n + M_n \Phi_n X_{n-1} \Phi_n^T M_n^T \tag{57}$$

so

$$M^T Y^{-1} = H^{-1} [M^T Q^{-1} M + H^{-1}]^{-1} M^T Q^{-1} \tag{58}$$

Equation 44 becomes

$$0 = \{H^{-1} [M^T Q^{-1} M + H^{-1}]^{-1} M^T Q^{-1} (Y - yy^T)$$

$$\cdot Q^{-1} M [M^T Q^{-1} M + H^{-1}]^{-1} H^{-1}\}_n \Big|_{\substack{H_n \to \hat{H}_n \\ X_{n-1} \to \overline{X}_{n-1}}} \tag{59}$$

which can be reduced to

$$0 = [M^T Q^{-1} (Y - yy^T) Q^{-1} M]_n \Big|_{\substack{H_n \to \hat{H}_n \\ X_{n-1} \to \overline{X}_{n-1}}} \tag{60}$$

and solved for $\hat{H}_n$

$$\hat{H}_n = \{[(M^T Q^{-1} M)^{-1} M^T Q^{-1}](yy^T - Q)$$

$$\cdot [Q^{-1} M (M^T Q^{-1} M)^{-1}]\}_n \Big|_{X_{n-1} \to \overline{X}_{n-1}} \tag{61}$$

The only difference between this equation and equation 45, for which M is square, is that M_n^{-1} has been replaced by $[(M^T Q^{-1} M)^{-1} M^T Q^{-1}]_n$. Thus all of

the results obtained in the preceding sub-section are valid when M is vertical, provided this substitution is made.

It would seem that this estimator has the serious disadvantage of being too difficult to use, except on very simple systems.

a. Alternate estimation method   Since the direct application of the MP estimation method leads to a result of some complexity when M_n is vertical, it may be advantageous to investigate an alternative. Let the MP estimation method be applied to finding an estimate for $(MHM^T)_n$ and then solve for $\hat{H}_n$. Thus, instead of finding an estimate of the H matrix which is "most likely", an estimate of the H matrix is found that causes the estimate of (MHM^T) to be "most likely".

$$\left. \frac{\partial\ p(y_n)}{\partial\ (MHM^T)_n} \right|_{\substack{H_n \rightarrow \hat{H}_n \\ X_{n-1} \rightarrow \overline{X}_{n-1}}} = 0 \qquad (62)$$

$$0 = \left[Y^{-1}(Y-yy^T)Y^{-1} \right]_n \Big|_{H_n \rightarrow \hat{H}_n,\ X_{n-1} \rightarrow \overline{X}_{n-1}}$$

$$(MHM^T)_n = (yy^T-Q)_n \Big|_{X_{n-1} \rightarrow \overline{X}_{n-1}}$$

$$\hat{H}_n = \left\{ \left[(M^TM)^{-1}M^T \right](yy^T-Q)\left[M(M^TM)^{-1} \right] \right\}_n \Big|_{X_{n-1} = \overline{X}_{n-1}} \qquad (63)$$

This result differs from the MP estimate of H_n only to the extent that the Q^{-1} factors have been eliminated, so once again all the results obtained in the sub-section covering the square M-matrix case are valid if M_n^{-1} is replaced by $[(M^TM)^{-1}M^T]_n$. The elimination of the Q-inverses should greatly ease the computational task.

3. Solution when M is horizontal

When M_n is horizontal, the system is not observable (4,6), and there is an infinite set of right-hand inverses for it. Thus, equation 44 reduces to

$$0 = Y_n - y_n y_n^T \Big|_{H_n \to \hat{H}_n, \ X_{n-1} \to \bar{X}_{n-1}}$$

$$M_n \hat{H}_n M_n^T = y_n y_n^T - Q_n \Big|_{X_{n-1} \to \bar{X}_{n-1}} \tag{64}$$

At this point it is seen that $\hat{H}_n$ cannot be found uniquely since there is no way to pre- and post-multiply (MHM^T) such that M and M^T are eliminated. This is equivalent to the ordinary algebraic situation of having more unknown quantities (corresponding to elements of $\hat{H}_n$) to be determined than there are independent equations. This does not mean that H_n cannot be estimated. On the contrary, an infinity of estimates can be made. It just means that there is no way of knowing how good any particular estimate is, unless other information is supplied. Any of these estimates can be represented by

$$\hat{H}_n = [\text{right-hand inverse of } M_n](y_n y_n^T - \bar{Q}_n)[\text{right-hand inverse of } M_n]^T \tag{65}$$

Among the infinite set of right-hand inverses of M_n is one that causes the $(\text{norm})^2$ of the variance of $\hat{H}_n$ to be minimum, but there is no way of finding it unless H_n is already known.

One particular estimate that is unique in a certain respect is obtained by letting the right-hand inverse of M_n be $[M_n^T(M_n M_n^T)^{-1}]$, the general inverse. This results in an estimate for H_n that is "minimum norm" (2); that is, $\text{tr}(\hat{H}_n \hat{H}_n)$ is smaller when using the general inverse than when

using any other right-hand inverse. This estimate may lead to some insight about H_n, but it is doubtful that it is of much real use.

The sort of additional information needed to obtain a unique and useful estimate of H_n is prior knowledge of part of the H_n matrix. To see this, partition M_n and H_n as follows:

$$(MHM^T)_n = \left| M_a M_b \right|_n \cdot \left| \begin{matrix} H_{aa} & H_{ab} \\ H_{ab}^T & H_{bb} \end{matrix} \right|_n \cdot \left| \begin{matrix} M_a^T \\ M_b^T \end{matrix} \right|_n$$

$$= (M_a H_{aa} M_a^T + M_a H_{ab} M_b^T + M_b H_{ab}^T M_a^T + M_b H_{bb} M_b^T)_n \tag{66}$$

If the elements of the matrices have been arranged so that only the elements of $(H_{aa})_n$ are unknown, then

$$(M_a H_{aa} M_a^T)_n = y_n y_n^T - \overline{Q}_n - (M_a H_{ab} M_b^T + M_b H_{ab}^T M_a^T + M_b H_{bb} M_b^T)_n \tag{67}$$

$$= y_n y_n^T - \overline{Q}_n^* \tag{68}$$

Now, if $(M_a)_n$ is either square or vertical, $(\hat{H}_{aa})_n$ can be determined by the methods already developed; simply replace M_n by $(M_{aa})_n$ and Q_n by Q_n^*.

B. Maximum-Conditional-Probability Estimation of H

The MCP estimation of H requires the performance of the following operation,

$$\left. \frac{\partial\, p(y_n | y_{n-1}, \ldots, y_1)}{\partial\, H_n} \right|_{\substack{H_n \to \hat{H}_{n/n} \\ P_{n-1/n-1} \to \overline{P}_{n-1/n-1}}} = 0 \tag{69}$$

where

$$\bar{P}_{n-1/n-1} = P_{n-1/n-1}\Big|_{H_{n-1} \to \bar{H}_{n-1}\ ,\ P_{n-2/n-2} \to \bar{P}_{n-2/n-2}}$$

The result is

$$0 = M_n^T C_n^{-1}(C_n - \tilde{y}_{n/n-1}\tilde{y}_{n/n-1}^T)C_n^{-1}M_n\Big|_{H_n \to \hat{H}_{n/n}\ ,\ P_{n-1/n-1} \to \bar{P}_{n-1/n-1}} \tag{70}$$

Due to the similarity of this equation to equation 44, the solutions for $\hat{H}_{n/n}$ are similar in form to those obtained for $\hat{H}_n$, the MP estimate of H_n.

1. Solution when M is square

When M_n is square,

$$\hat{H}_{n/n} = (M_n)^{-1}(\tilde{y}_{n/n-1}\tilde{y}_{n/n-1}^T - V_n - M_n\Phi_n\bar{P}_{n-1/n-1}\Phi_n^T M_n^T)(M_n^T)^{-1} \tag{71}$$

2. Solution when M is vertical

When M_n is vertical, the MCP estimate for H_n is the same as equation 71 with $(M_n)^{-1}$ replaced by the factor $[(M^T\bar{R}^{-1}M)^{-1}M^T\bar{R}^{-1}]_n$

where

$$\bar{R}_n = V_n + M_n\Phi_n\bar{P}_{n-1/n-1}\Phi_n^T M_n^T \tag{72}$$

The alternate estimate is the same as equation 71 with $(M_n)^{-1}$ replaced by the factor $[(M_n^T M_n)^{-1}M_n^T]$.

3. Solution when M is horizontal

When M_n is horizontal, the "minimum norm" estimate of H_n can be obtained by replacing $(M_n)^{-1}$ in equation 71 by $[M_n^T(M_n M_n^T)^{-1}]$. If enough elements of H_n are already known, then the unknown portion can be estimated by means of matrix partition and rearrangement of terms to get the equation into the form where the methods used on the square or vertical M_n matrix can be used:

$$(M_a\hat{H}_{aa}M_a^T)_n = \tilde{y}_{n/n-1}\tilde{y}_{n/n-1}^T - \bar{R}_n^* \tag{73}$$

where

$$R_n^* = R_n + (M_a H_{ab} M_b^T + M_b H_{ab}^T M_a^T + M_b H_{bb} M_b^T)_n \tag{74}$$

4. Determination of norms

The norm of the variance of $\hat{H}_{n/n}$ cannot be easily determined due to the complexity of $\bar{P}_{n-1/n-1}$. However, as in the situation of section IV where V_n is estimated by the MCP method, the minimum possible norm would occur if $\bar{P}_{n-1/n-1}$ were exactly $P_{n-1/n-1}$.

$$[(\text{norm})^2 \text{ of } \hat{H}_{n/n} - \text{variance}]_{min} = T[(M^{-1}CM^{T^{-1}})_n] \tag{75}$$

Equation 75 is the minimum-norm case when M_n is square. The other norms can be obtained by the proper substitutions for M_n^{-1}. Since

$$tr(P_{n-1/n-1} \cdot P_{n-1/n-1}) \leq tr(X_{n-1} \cdot X_{n-1}),$$

the MCP estimation method should be better than the MP estimation method, and the actual $(\text{norm})^2$ of the $\hat{H}_{n/n}$ - variance would be expected to be between the minimum-possible MCP $(\text{norm})^2$ and the maximum MP $(\text{norm})^2$.

VI. ESTIMATION OF BOTH THE MEASUREMENT-ERROR AND RESPONSE COVARIANCES

The situation postulated in this section is that the measurement-error covariance matrix, V, and the response covariance matrix, H, are both unknown.

A. Maximum-Probability Estimation of V and H

Maximizing the unconditional probability density function with respect to both V_n and H_n results in the two equations,

$$\left.\begin{array}{l} 0 = Y_n - y_n y_n^T \\[2em] 0 = M_n^T Y_n^{-1}(Y_n - y_n y_n^T) Y_n^{-1} M_n \end{array}\right|\begin{array}{l} V_n \to \hat{V}_n \\[1em] H_n \to \hat{H}_n \\[1em] X_{n-1} \to \bar{X}_{n-1} \end{array} \Biggr\} \qquad (76)$$

There is no possibility that solutions may be obtained for $\hat{V}_n$ and $\hat{H}_n$, as things stand, since the equations are not independent. Additional conditions or information are needed.

Since V_n and Y_n are both r-by-r matrices, the number of unknown elements of V_n equals the number of known elements of Y_n or $y_n y_n^T$. The number of unknown elements of H_n may be greater than, equal to, or less than the number of known elements, depending upon whether M_n is horizontal, square, or vertical, respectively. Therefore, some way must be found to reduce the number of unknown elements in V_n and possibly some elements of H_n will have to be known, particularly if M_n is not vertical.

Until now, the assumption has been that the measurement errors might be mutually correlated (dependent), but not correlated in time. In many systems, the measurement errors will be completely uncorrelated

(independent) and V_n will be a diagonal matrix in such cases. For such a system, the number of unknown elements of V_n is reduced from $\frac{1}{2}(r+1)r$ to just r, and the equations to be solved are expressed, using notation of Appendix B, as

$$\left. \begin{array}{l} 0 = D[Y_n - y_n y_n^T] \\[2em] 0 = M_n^T Y_n^{-1}(Y_n - y_n y_n^T)Y_n^{-1}M_n \end{array} \right| \begin{array}{l} V_n \to \hat{V}_n \\[1em] H_n \to \hat{H}_n \\[1em] X_{n-1} \to \bar{X}_{n-1} \end{array} \right\} \tag{77}$$

These two equations are independent, so a solution is possible if M_n is vertical. Proceeding,

$$\hat{V}_n = D[y_n y_n^T - M_n \hat{H}_n M_n^T - M_n \Phi_n \bar{X}_{n-1}\Phi_n^T M_n^T] \tag{78}$$

and, using the methods of section V,

$$\hat{H}_n = [(M^T Q^{-1}M)^{-1}M^T Q^{-1}]_n (yy^T - Q)_n [MQ^{-1}[MQ^{-1}(M^T Q^{-1}M)]]_n \left| \begin{array}{l} V_n \to \hat{V}_n \\[1em] X_{n-1} \to \bar{X}_{n-1} \end{array} \right. \tag{79}$$

The nature of the $D[\ ]$ operation prevents one from substituting equation 78 directly into equation 79 in order to eliminate $\hat{V}_n$ and solve for $\hat{H}_n$. The presence of so many Q's, which are functions of V, also prevents one from substituting equation 79 into 78 and solving for $\hat{V}_n$.

The alternate estimation method developed in section V can be tried as a means of obtaining the solutions. Thus, instead of equation 79, use

$$\hat{H}_n = [(M^TM)^{-1}M^T]_n (yy^T-Q)_n [M(M^TM)^{-1}]_n \Big|_{\substack{V_n \to \hat{V}_n \\ X_{n-1} \to \bar{X}_{n-1}}} \tag{80}$$

If equation 80 is substituted into equation 78, considerable cancellation of terms occurs and the result is

$$\hat{V}_n - D[N_n\hat{V}_nN_n] = D[y_ny_n^T - N_ny_ny_n^TN_n] \tag{81}$$

where

$$N_n = M_n(M_n^TM_n)^{-1}M_n^T$$

(N_n is an idempotent matrix.)

A typical element of $D[N_n\hat{V}_nN_n]$ can be expressed as

$$(D[N\hat{V}N]_{ab})_n = \{\sum_i [(N)_{ai}(\hat{V})_{ii}(N)_{ib}\delta_b^a]\}_n \tag{82}$$

so a typical diagonal element is

$$(D[N\hat{V}N]_{aa})_n = \{\sum_i [(N)_{ai}^2(\hat{V})_{ii}]\}_n \tag{83}$$

This can be put back into matrix notation as

$$c[N_n\hat{V}_nN_n] = (N_n*N_n)\; c[\hat{V}_n] \tag{84}$$

Thus,

$$c[\hat{V}_n] -(N_n*N_n)\; c[\hat{V}_n] = c[y_ny_n^T - N_ny_ny_n^TN_n] \tag{85}$$

$$c[\hat{V}_n] = (I-N_n*N_n)^{-1}\; c[y_ny_n^T - N_ny_ny_n^TN_n] \tag{86}$$

With this solution for $\hat{V}_n$, it is then a relatively simple matter of substitution into equation 80 to evaluate $\hat{H}_n$.

The expected value of $c[\hat{V}_n]$ is

$$E[c[\hat{V}_n]] = (I-N_n*I*N_n)^{-1} E[c[y_n y_n^T - N_n y_n y_n^T N_n]]$$

$$= (I-N_n*I*N_n)^{-1} c[Y_n - N_n Y_n N_n]$$

$$= (I-N_n*I*N_n)^{-1} c[V_n - N_n V_n N_n]$$

$$= c[V_n] \tag{87}$$

so this estimate is unbiased.

Results obtained in Appendix F enable the norm to be expressed.

$$[(norm)^2 \text{ of } \hat{V}_n - \text{variance}] = \text{tr}\{E[(\hat{V}_n - V_n)(\hat{V}_n - V_n)]\}$$

$$= 2 \text{ tr}\{(I-N*N*N)^{-1}[Y*Y-2(YN)*(YN) + (NYN)*(NYN)](I-N*N)^{-1}\}_n \tag{88}$$

The norm of the variance of $\overline{V}_n$ involves the expected values of cross-products such as $(\hat{V}_i - V_i)(\hat{V}_j - V_j)$, as shown in equation 30, but for the particular estimation method being used in this section, all of these cross-products have expected values which are zeros.

$$\text{tr}\{E[(\hat{V}_i - V_i)(\hat{V}_j - V_j)]\} = 0 \quad, \quad \text{for } i \neq j \tag{89}$$

This is demonstrated in Appendix F. Consequently,

$$[(norm)^2 \text{ of } \overline{V}_n - \text{variance}] = \frac{1}{n^2} \sum_{i=1}^{n} [(norm)^2 \text{ of } \hat{V}_i - \text{variance}] \tag{90}$$

The development of the norm expressions for the variances of $\hat{H}_n$ and $\overline{H}_n$ is rather difficult since the estimates of the V's are involved. Fortunately, these norm expressions are not required for the purpose of investigating whether or not convergence of the averaged estimates can occur. Equations 88 and 90 indicate that averaged estimates for the V's

are expected to converge when the system is asymptotically stable. So, when n becomes large, $\overline{V}_n$ differs from V_n by negligible amounts and the situation becomes virtually the same as postulated in section V where H_n is estimated when V_n is known. Therefore, the condition for convergence of $\overline{H}_n$ is that the system be asymptotically stable.

B. Maximum-Conditional-Probability Estimation of V and H

When the conditional probability density function is maximized with respect to V_n and H_n, much the same problems arise as with the MP estimation method. The only difference is that X_n is replaced by $P_{n/n-1}$ and y_n by $\tilde{y}_{n/n-1}$. Thus,

$$c[\hat{V}_{n/n}] = (I - N_n * N_n)^{-1} \cdot c[\tilde{y}_{n/n-1} \tilde{y}^T_{n/n-1} - N_n \tilde{y}_{n/n-1} \tilde{y}^T_{n/n-1} N_n] \tag{91}$$

$$\hat{H}_{n/n} = [(M^T M)^{-1} M^T (\tilde{y}\tilde{y}^T - \hat{V}) M (M^T M)^{-1}]_n - \Phi_n \overline{P}_{n-1/n-1} \Phi^T_n \tag{92}$$

The minimum norm for the variance of $\hat{V}_n$ is

$$[(norm)^2 \text{ of } \hat{V}_n - variance]_{min} = 2 \text{ tr}\{(I - N*N)^{-1}[C*C - 2(CN)*(CN)$$

$$+ (NCN)*(NCN)](I - N*N)^{-1}\}_n \tag{93}$$

The minimum norm for the variance of $\hat{H}_n$ is

$$[(norm)^2 \text{ of } \hat{H}_n - variance]_{min} = T[(M^T M)^{-1} M^T C M (M^T M)^{-1}]_n \tag{94}$$

As has been stated in other sections, the actual norms of the variances will be greater than these minimum-norms and, if enough measurements are used to obtain good averages of estimates, the actual norms should be less than the worst-case norms obtained for the MP estimates.

VII. SUMMARY AND CONCLUSIONS

By the process of choosing those values for V_n and/or H_n which are most probable (i.e., cause a probability density function to be maximum) and then averaging, reasonable estimates can be found for the measurement-error covariance matrix, V, and/or the response covariance matrix, H. When the initial system-state covariance, X_0, is known for the MP estimation method, or $P_{0/0}$ is known for the estimate of x_0 when using the MCP estimation method, then all of the estimates are unbiased.

The individual estimates (prior to averaging) which were developed in the preceding sections are summarized as follows:

1. When V is unknown, the MP estimate is

$$\hat{V}_n = y_n y_n^T - M_n X_n M_n^T$$

and the MCP estimate is

$$\hat{V}_n = \tilde{y}_{n/n-1} \tilde{y}_{n/n-1}^T - M_n \bar{P}_{n/n-1} M_n^T$$

where $\bar{P}_{n/n-1}$ is a function of previous estimates of V.

2. When H is unknown, the MP estimate is

$$\hat{H}_n = M_n^{-1}(y_n y_n^T - \bar{Q}_n) M_n^{T-1} \ ,$$

if M_n is square;

$$\hat{H}_n = [(M^T \bar{Q}^{-1} M)^{-1} M^T \bar{Q}^{-1}]_n (y_n y_n^T - \bar{Q}_n)[\bar{Q}^{-1} M(M^T \bar{Q}^{-1} M)^{-1}]_n$$

if M_n is vertical and the inversions are not too difficult, or

$$\hat{H}_n = [(M^T M)^{-1} M^T]_n (y_n y_n^T - \bar{Q}_n)[M(M^T M)^{-1}]_n$$

if M_n is vertical and $\bar{Q}_n$ is not easily inverted; or

$$\hat{H}_n = [M^T(MM^T)^{-1}]_n(y_n y_n^T - \bar{Q}_n)[(MM^T)^{-1}M]_n$$

which is the minimum-norm estimate if M_n is horizontal. If

enough of the elements of H are known, matrix partitioning

may be used to convert from a situation in which M_n is horizontal

to one in which the techniques used for M_n being square or verti-

cal can be applied. The MCP estimates are similar to the MP

estimates: $\bar{Q}_n$ is replaced by $\bar{R}_n$ and y_n is replaced by $\tilde{y}_n$ in the

equations.

3. When both V and H are unknown, but M is vertical and V is known

to be a diagonal matrix, the MP estimates are

$$c[\hat{V}_n] = (I - N*N)^{-1} c[y_n y_n^T - N_n y_n y_n^T N_n]$$

$$\hat{H}_n = [(M^TM)^{-1}M^T]_n(y_n y_n^T - \bar{V}_n)[M(M^TM)^{-1}]_n - \Phi_n \bar{X}_{n-1}\Phi_n^T$$

and the MCP estimates are similar, with $\bar{X}_{n-1}$ replaced by $\bar{P}_{n-1/n-1}$

and y_n replaced by $\tilde{y}_n$.

The MP estimators are the easiest to use, but the accuracies to be

obtained are not expected to be as good as those obtained by use of the

MCP estimators. The MCP estimators are considerably more difficult to

apply since, at each measurement time, averages of previous estimates must

be substituted into the Kalman filter equations to find approximations to

the W and P matrices and make a priori estimates of the y's. Of course

the ultimate objective is to form these Kalman filter parameters so that

the state vectors can be estimated, so there is no particular saving in

computational labor by using the MP estimators instead of the MCP

estimators.

The examinations made of the estimation-variance norms result in the conclusion that all of the averaged estimates converge if the system is asymptotically stable and the parameters being estimated are time-stationary, or nearly so. Thus, there would appear to be no reason why these estimators could not be used to adapt a Kalman filter to a system with unknown V and/or H which may be varying in time slowly with respect to the measurement sampling rate.

VIII. LITERATURE CITED

1. Battin, R. H. A statistical optimizing navigation procedure for space flight. American Rocket Society Journal 32: 1681-1696. 1962.

2. Horst, Paul. Matrix algebra for social scientists. New York, N.Y., Holt, Rinehart and Winston, Inc. 1963.

3. Kalman, R. E. A new approach to linear filtering and prediction problems. American Society of Mechanical Engineers Transactions, Series D, Journal of Basic Engineering 82: 35-45. 1960.

4. Lee, Robert C. K. Optimal estimation, identification, and control. Massachusetts Institute of Technology Research Monograph 28. 1964.

5. Mood, A. M. Introduction to the theory of statistics. New York, N.Y., McGraw-Hill Book Company, Inc. 1950.

6. Tou, J. T. Modern control theory. New York, N.Y., McGraw-Hill Book Company, Inc. 1964.

7. Wiener, Norbert. Extrapolation, interpolation, and smoothing of stationary time series. Cambridge, Mass., The M.I.T. Press. 1949.

8. Zadeh, L. A. and C. A. Desoer. Linear system theory. New York, N.Y., McGraw-Hill Book Company, Inc. 1963.

IX. ACKNOWLEDGMENT

The author wishes to express his appreciation for the suggestions

and guidance given by his major professor, Dr. R. G. Brown, in both the

developmental and writing phases of this dissertation.

X. APPENDIX A

Differentiations of certain matrix functions with respect to a matrix are required. The procedure used is to perform the differentiation with respect to a general element of the matrix first, using summation notation, and then converting to the equivalent statement in matrix notation.

The following symbology is used: if Z is a matrix, then:

1. $|Z|$ denotes the determinant of Z,

2. z_{ij} or $(Z)_{ij}$ denotes the ij'th element of Z,

3. $|Z_{ji}|$ denotes the cofactor of z_{ij} or $(Z)_{ij}$.

1. $\dfrac{\partial |Z|}{\partial z_{mn}} = \sum\limits_{i,j}\left[\dfrac{\partial z_{ij}}{\partial z_{mn}}|Z_{ji}|\right] = \sum\limits_{i,j}(\delta^i_m\delta^j_n|Z_{ji}|) = |Z_{nm}|$

$\therefore \quad \dfrac{\partial |Z|}{\partial Z} = \text{adjoint}(Z)$

2. $\dfrac{\partial |Z+B|}{\partial z_{mn}} = \sum\limits_{i,j}\left[\dfrac{\partial(z_{ij}+b_{ij})}{\partial z_{mn}}(|Z_{ji}|+|B_{ji}|)\right] = |(Z+B)_{nm}|$

$\therefore \quad \dfrac{\partial |Z+B|}{\partial Z} = \text{adjoint}(Z+B)$

3. $\dfrac{\partial |EZF+B|}{\partial z_{mn}} = \sum\limits_{i,j}\left[\dfrac{\partial\left(\sum\limits_{k,\ell}e_{ik}z_{k\ell}f_{\ell j}+b_{ij}\right)}{\partial z_{mn}}\right]|(EZF+B)_{ji}| = \sum\limits_{i,j}\left[e_{im}f_{nj}|(EZF+B)_{ji}|\right]$

$\therefore \quad \dfrac{\partial |EZF+B|}{\partial Z} = E^T[\text{adjoint}(EZF+B)]F^T$

4. $\dfrac{\partial(a^TZb)}{\partial z_{mn}} = a^T\dfrac{\partial Z}{\partial z_{mn}}b = a^T\varepsilon_{mn}b = a_mb_n$

where ε_{mn} is a matrix that has zeros for all its elements except for the mn'th element which is one.

$$\therefore \quad \frac{\partial(a^T Z b)}{\partial Z} = ab^T$$

5. $$\frac{\partial(Z \cdot Z^{-1})}{\partial z_{mn}} = \frac{\partial Z}{\partial z_{mn}} Z^{-1} + Z \frac{\partial Z^{-1}}{\partial z_{mn}} = \frac{\partial(I)}{\partial z_{mn}} = 0$$

$$\therefore \quad \frac{\partial Z^{-1}}{\partial z_{mn}} = - Z^{-1} \frac{\partial Z}{\partial z_{mn}} Z^{-1} = - Z^{-1} \varepsilon_{mn} Z^{-1}$$

6. $$\frac{\partial(a^T Z^{-1} b)}{\partial z_{mn}} = - (a^T Z^{-1}) \varepsilon_{mn} (Z^{-1} b) = - (Z^{T^{-1}} a)_m (Z^{-1} b)_n$$

$$\therefore \quad \frac{\partial(a^T Z^{-1} b)}{\partial Z} = - Z^{T^{-1}} a \, b^T \, Z^{T^{-1}}$$

7. $$\frac{\partial(a^T (Z+B)^{-1} b)}{\partial z_{mn}} = - a^T (Z+B)^{-1} \varepsilon_{mn} (Z+B)^{-1} b$$

$$\therefore \quad \frac{\partial(a^T (Z+B)^{-1} b)}{\partial Z} = - (Z+B)^{T^{-1}} ab^T (Z+B)^{T^{-1}}$$

8. $$\frac{\partial[a^T (EZF+B)^{-1} b]}{\partial z_{mn}} = - a^T (EZF+B)^{-1} E \varepsilon_{mn} F (EZF+B)^{-1} b$$

$$\therefore \quad \frac{\partial[a^T (EZF+B)^{-1} b]}{\partial Z} = - E^T (EZF+B)^{T^{-1}} ab^T (EZF+B)^{T^{-1}} F^T$$

9. Example of matrix differentiation problem

Some of the preceding matrix derivatives are employed in finding the derivative of a somewhat more elaborate function, of which the multivariate Gaussian probability density is a special case.

$$\frac{\partial\left[|EZF+B|^{-\frac{1}{2}}e^{-\frac{1}{2}[a^T(EZF+B)^{-1}b]}\right]}{\partial Z}$$

$$= -\frac{1}{2}\frac{E^T[\text{adjoint}(EZF+B)]F^T}{|EZF+B|^{3/2}}e^{-\frac{1}{2}[a^T(EZF+B)^{-1}b]}$$

$$+ \frac{1}{2}|EZF+B|^{-\frac{1}{2}}e^{-\frac{1}{2}[a^T(EZF+B)^{-1}b]}$$

$$\cdot E^T(EZF+B)^{T^{-1}}ab^T(EZF+B)^{T^{-1}}F^T$$

$$= -\frac{1}{2}|EZF+B|^{-\frac{1}{2}}e^{-\frac{1}{2}[a^T(EZF+B)^{-1}b]}$$

$$\cdot [\ E^T(EZF+B)^{T^{-1}}F^T - E^T(EZF+B)^{T^{-1}}ab^T(EZF+B)^{T^{-1}}F^T]$$

XI. APPENDIX B

Some non-standard matrix operations are required in order to express certain relationships which the standard matrix operations cannot express.

If Z is a square matrix, then $D[Z]$ is defined to be a square matrix whose major diagonal elements are identical to the corresponding major diagonal elements of Z and whose off-diagonal elements are all zeros.

If Z is a square matrix, then $c[Z]$ is defined to be a column vector whose elements are the corresponding major diagonal elements of Z.

If T and Z are two matrices with the same numbers of rows and columns, then $(T*Z)$ is defined to be a matrix, also with the same numbers of rows and columns, whose elements are the simple products of the corresponding elements of T and Z.

Certain useful relationships between these defined operations are:

1. $D[Z] = Z*I$

2. $D[c[Z] \cdot 1^T] = D[Z]$

3. $1^T \cdot (Z*Z) \cdot 1 = tr(Z \cdot Z^T)$

4. $(Z*Z) \cdot 1 = c[Z \cdot Z^T]$

XII. APPENDIX C

The expected values of two different matrix products involving an estimate of V are required.

1. The variance of $\hat{V}$

$$E[(\hat{V}-V)*(\hat{V}-V)]_{ij} = E[(yy^T - Y)*(yy^T - Y)]_{ij}$$

$$= E[(y_iy_j)\cdot(y_iy_j) - 2(y_iy_j)\cdot(Y)_{ij} + (Y)_{ij}\cdot(Y)_{ij}]$$

$$= [2(Y)^2_{ij} + (Y)_{ii}(Y)_{jj}] - 2(Y)^2_{ij} + (Y)^2_{ij}$$

$$= (Y)^2_{ij} + (Y)_{ii}(Y)_{jj}$$

$$E[(\hat{V}-V)*(\hat{V}-V)] = (Y*Y) + c[Y]\cdot c[Y]^T$$

2. The norm of the variance of $\hat{V}$

$$[(\text{norm})^2 \text{ of } \hat{V} - \text{variance}] = \text{tr}\{E[(\hat{V}-V)(\hat{V}-V)]\}$$

$$E[(\hat{V}-V)(\hat{V}-V)]_{ij} = E[(yy^T-Y)(yy^T-Y)]_{ij}$$

$$= E[\sum_k y_iy_ky_ky_j - y_iy_k(Y)_{kj} - (Y)_{ik}y_ky_j + (Y)_{ik}(Y)_{kj}]$$

$$= \sum_k \{[2(Y)_{ik}(Y)_{kj} + (Y)_{ij}(Y)_{kk}] - (Y)_{ik}(Y)_{kj}\}$$

$$= \sum_k [(Y)_{ik}(Y)_{kj} + (Y)_{ij}(Y)_{kk}]$$

$$E[(\hat{V}-V)(\hat{V}-V)] = Y\cdot Y + Y \text{ tr}(Y)$$

$$[(\text{norm})^2 \text{ of } \hat{V} - \text{variance}] = \text{tr}[Y\cdot Y + Y \text{ tr}(Y)]$$

$$= \text{tr}(Y\cdot Y) + [\text{tr}(Y)]^2$$

XIII. APPENDIX D

The expected values of several products involving estimates of H's are needed. These are determined in the following sub-sections, for a square M.

1. Norm of $\hat{H}_1$-variance

$$[(\text{norm})^2 \text{ of } \hat{H}_1\text{-variance}] = \text{tr}\{E[(\hat{H}_1-H_1)(\hat{H}_1-H_1)]\}$$

$$\hat{H}_1-H_1 = M_1^{-1}(y_1 y_1^T - Y_1)M_1^{T^{-1}}$$

$$E[(\hat{H}_1-H_1)(\hat{H}_1-H_1)] = M_1^{-1} E[(y_1 y_1^T - Y_1)M_1^{T^{-1}} M_1^{-1}(y_1 y_1^T - Y_1)]M_1^{T^{-1}}$$

$$\{E[(yy^T-Y)M^{T^{-1}} M^{-1}(yy^T-Y)]_{ij}\}_1$$

$$= E\{ \sum_{k,\ell} [y_i y_k (M^{T^{-1}} M^{-1})_{k\ell} y_\ell y_j - y_i y_k (M^{T^{-1}} M^{-1}Y)_{kj}$$

$$-(YM^{T^{-1}} M^{-1})_{i\ell} y_\ell y_j + (YM^{T^{-1}} M^{-1}Y)_{ij}]_1 \}$$

$$= \sum_{k,\ell} \{[(YM^{T^{-1}} M^{-1}Y)_{ij} + (Y)_{i\ell}(Y)_{kj}(M^{T^{-1}} M^{-1})_{k\ell}$$

$$+(Y)_{ij}(Y)_{\ell k}(M^{T^{-1}} M^{-1})_{k\ell}] - (YM^{T^{-1}} M^{-1}Y)_{ij}\}_1$$

$$E[(\hat{H}_1-H_1)(\hat{H}_1-H_1)] = M_1^{-1}[YM^{T^{-1}} M^{-1}Y + Y \text{ tr}(YM^{T^{-1}} M^{-1})]_1 M_1^{T^{-1}}$$

$$= (M^{-1}YM^{T^{-1}})_1[(M^{-1}YM^{T^{-1}})_1 + I \text{ tr}(M^{-1}YM^{T^{-1}})_1]$$

$$[(\text{norm})^2 \text{ of } \hat{H}_1\text{-variance}] = \text{tr}[(M^{-1}YM^{T^{-1}})_1 (M^{-1}YM^{T^{-1}})_1]$$

$$+ [\text{tr}(M^{-1}YM^{T^{-1}})_1]^2$$

$$= T[(M^{-1}YM^{T^{-1}})_1],$$

as defined by equation 49.

2. Norm of $\hat{H}_2$-variance

$$\hat{H}_2 - H_2 = [M^{-1}(yy^T - V)M^{T^{-1}}]_2 - \Phi_2 X_1 \Phi_2^T - H_2$$

$$= [M^{-1}(yy^T - Y)M^{T^{-1}}]_2 - \Phi_2(\hat{H}_1 - H_1)\Phi_2^T$$

$$= [M^{-1}(yy^T - Y)M^{T^{-1}}]_2 - \Phi_2[M^{-1}(yy^T - Y)M^{T^{-1}}]_1 \Phi_2^T$$

$$\text{tr}\{E[(\hat{H}_2 - H_2)(\hat{H}_2 - H_2)]\} = T[(M^{-1}YM^{T^{-1}})_2] + T[\Phi_2(M^{-1}YM^{T^{-1}})_1 \Phi_2^T]$$

$$- 2\,\text{tr}\{M_2^{-1} E[(yy^T - Y)_2 M_2^{T^{-1}} \Phi_2 M_1^{-1}(yy^T - Y)_1]M_1^{T^{-1}} \Phi_2^T\}$$

$$E[(yy^T - Y)_2 M_2^{T^{-1}} \Phi_2 M_1^{-1}(yy^T - Y)_1]$$

$$= E\{[(M_2\Phi_2 x_1 + M_2 g_2 + v_2)(M_2\Phi_2 x_1 + M_2 g_2 + v_2)^T$$

$$- M_2\Phi_2 X_1\Phi_2^T M_2^T - M_2 H_2 M_2^T - V_2]M_2^{T^{-1}} \Phi_2 M_1^{-1}$$

$$[(M_1 x_1 + v_1)(M_1 x_1 + v_1)^T - M_1 X_1 M_1^T - V_1]\}$$

$$= M_2\Phi_2[X_1(\Phi_2^T\Phi_2)^T X_1 + X_1\,\text{tr}(\Phi_2^T\Phi_2 X_1)]M_1^T$$

$$[(\text{norm})^2 \text{ of } \hat{H}_2\text{-variance}] = T[(M^{-1}YM^{T^{-1}})_2] + T[\Phi_2(M^{-1}YM^{T^{-1}})_1 \Phi_2^T]$$

$$- 2\,T[\Phi_2 X_1\Phi_2^T]$$

3. Norm of $\hat{H}_3$-variance

$$\hat{H}_3 - H_3 = [M^{-1}(yy^T - V)M^{T^{-1}}]_3 - \Phi_3 \bar{X}_2 \Phi_3^T - H_3$$

Then, if $\quad \bar{X}_2 = \Phi_2 \hat{X}_1 \Phi_2^T + \hat{H}_2$.

$$\hat{H}_3 - H_3 = [M^{-1}(yy^T - Y)M^{T^{-1}}]_3 - \Phi_3[\Phi_2(\hat{X}_1 - X_1)\Phi_2 + (\hat{H}_2 - H_2)]\Phi_3^T$$

$$= [M^{-1}(yy^T - Y)M^{T^{-1}}]_3 - \Phi_3[\Phi_2(\hat{H}_1 - H_1)\Phi_2 + (\hat{H}_2 - H_2)]\Phi_3^T$$

$$\hat{H}_3 - H_3 = [M^{-1}(yy^T - Y)M^{T^{-1}}]_3 - \Phi_3 \Phi_2[M^{-1}(yy^T - Y)M^{T^{-1}}]_1 \Phi_2^T \Phi_3^T$$

$$-\Phi_3\{[M^{-1}(yy^T - Y)M^{T^{-1}}]_2 - \Phi_2[M^{-1}(yy^T - Y)M^{T^{-1}}]_1 \Phi_2^T\}\Phi_3^T$$

$$= [M^{-1}(yy^T - Y)M^{T^{-1}}]_3 - \Phi_3[M^{-1}(yy^T - Y)M^{T^{-1}}]_2 \Phi_3^T$$

Since the form of this equation is the same as for $(\hat{H}_2 - H_2)$, then the resulting norm of the variance of $\hat{H}_3$ should also be the same form as the norm of the variance of $\hat{H}_2$.

$$[(norm)^2 \text{ of } \hat{H}_3\text{-variance}] = T[(M^{-1}YM^{T^{-1}})_3] + T[\Phi_3(M^{-1}YM^{T^{-1}})_2 \Phi_3^T]$$

$$- 2 T[\Phi_3 X_2 \Phi_3^T]$$

4. Norm of $\hat{H}_n$-variance

By induction, the norm of the variance of $\hat{H}_n$ can be specified provided that $\bar{X}_{n-1} = \Phi_{n-1} \bar{X}_{n-2} \Phi_{n-1}^T + \hat{H}_{n-1}$. With this stipulation on how previous estimates of H are incorporated, the norm expression is

$$[(norm)^2 \text{ of } \hat{H}_n\text{-variance}] = T[(M^{-1}YM^{T^{-1}})_n] + T[\Phi_n(M^{-1}YM^{T^{-1}})_{n-1} \Phi_n^T]$$

$$- 2 T[\Phi_n X_{n-1} \Phi_n^T]$$

5. Norms of cross-products

$$E[(\hat{H}_i - H_i)(\hat{H}_{i-1} - H_{i-1})]$$

$$= E\{[[M^{-1}(yy^T - Y)M^{T^{-1}}]_i - \Phi_i[M^{-1}(yy^T - Y)M^{T^{-1}}]_{i-1}\Phi_i^T]$$

$$\cdot [[M^{-1}(yy^T - Y)M^{T^{-1}}]_{i-1} - \Phi_{i-1}[M^{-1}(yy^T - Y)M^{T^{-1}}]_{i-2}\Phi_{i-1}^T]\}$$

$$E[(\hat{H}_i - H_i)(\hat{H}_{i-1} - H_{i-1})]$$

$$= \Phi_i[X_{i-1}\Phi_i X_{i-1} + X_{i-1}\text{tr}(\Phi_i X_{i-1})]$$

$$- \Phi_i\Phi_{i-1}[X_{i-2}\Phi_{i-1}^T\Phi_i\Phi_{i-1}X_{i-2} + X_{i-2}\text{tr}(\Phi_{i-1}^T\Phi_i\Phi_{i-1}X_{i-2})]\Phi_{i-1}^T$$

$$- \Phi_i M_{i-1}^{-1}[Y_{i-1}(M_{i-1}^{T^{-1}}\Phi_i M_{i-1}^{-1})Y_{i-1} + Y_{i-1}\text{tr}(M_{i-1}^{T^{-1}}\Phi_i M_{i-1}^{-1}Y_{i-1})]M_{i-1}^{T^{-1}}$$

$$+ \Phi_i\Phi_{i-1}[X_{i-2}\Phi_{i-1}\Phi_i\Phi_{i-1}^T X_{i-2} + X_{i-2}\text{tr}(\Phi_{i-1}\Phi_i\Phi_{i-1}^T X_{i-2})]\Phi_{i-1}^T$$

$$= (\Phi_i X_{i-1})[(\Phi_i X_{i-1}) + I\,\text{tr}(\Phi_i X_{i-1})]$$

$$- (\Phi_i M_{i-1}^{-1}Y_{i-1}M_{i-1}^{T^{-1}})[(\Phi_i M_{i-1}^{-1}Y_{i-1}M_{i-1}^{T^{-1}}) + I\,\text{tr}(\Phi_i M_{i-1}^{-1}Y_{i-1}M_{i-1}^{T^{-1}})]$$

$$\{(\text{norm})^2 \text{of} E[(\hat{H}_i - H_i)(\hat{H}_{i-1} - H_{i-1})]\} = T[\Phi_i X_{i-1}] - T[\Phi_i M_{i-1}^{-1}Y_{i-1}M_{i-1}^{T^{-1}}]$$

$$E[(\hat{H}_i - H_i)(\hat{H}_j - H_j)]\Big|_{i>j+1}$$

$$= E\{[[M^{-1}(yy^T - Y)M^{T^{-1}}]_i - \Phi_i[M^{-1}(yy^T - Y)M^{T^{-1}}]_{i-1}\Phi_i^T]$$

$$\cdot [[M^{-1}(yy^T - Y)M^{T^{-1}}]_j - \Phi_j[M^{-1}(yy^T - Y)M^{T^{-1}}]_{j-1}\Phi_j^T]\}$$

$$= \Phi_i\cdots\Phi_{j+1}\{[X_j\Phi_i\cdots\Phi_{j+1}X_j + X_j\text{tr}(\Phi_i\cdots\Phi_{j+1}X_j)]$$

$$- \Phi_j[X_{j-1}\Phi_j^T\Phi_i\cdots\Phi_j X_{j-1} + X_{j-1}\text{tr}(\Phi_j^T\Phi_i\cdots\Phi_j X_{j-1})]\Phi_j^T\}$$

$$- \Phi_i \cdots \Phi_{j+1} \{ [X_j \Phi_i \cdots \Phi_{j+1} X_j + X_j \mathrm{tr}(\Phi_i \cdots \Phi_{j+1} X_j)]$$

$$+ \Phi_j [X_{j-1} \Phi_j^T \Phi_i \cdots \Phi_j X_{j-1} + X_{j-1} \mathrm{tr}(\Phi_j^T \Phi_i \cdots \Phi_j X_{j-1})] \Phi_j^T \}$$

$$= 0$$

XIV. APPENDIX E

The development of a matrix inversion lemma, essentially as given by Horst (2), is presented here because it also suffices to provide a matrix identity which is of some use.

If $Y = Q + MHM^T$

then
$$I = QY^{-1} + MHM^TY^{-1}$$

$$QY^{-1} = I - MHM^TY^{-1}$$

$$Y^{-1} = Q^{-1} - Q^{-1}MHM^TY^{-1} \tag{E1}$$

$$Y^{-1} + Q^{-1}MHM^TY^{-1} = Q^{-1}$$

$$M^TY^{-1} + M^TQ^{-1}MHM^TY^{-1} = M^TQ^{-1}$$

$$(H^{-1} + M^TQ^{-1}M)HM^TY^{-1} = M^TQ^{-1}$$

$$HM^TY^{-1} = (H^{-1} + M^TQ^{-1}M)^{-1}M^TQ^{-1} \tag{E2}$$

Thus, the matrix identity is

$$M^TY^{-1} = H^{-1}(H^{-1} + M^TQ^{-1}M)^{-1}M^TQ^{-1}$$

The matrix inversion lemma results from substituting equation E2 into equation E1.

$$Y^{-1} = Q^{-1} - Q^{-1}M(H^{-1} + M^TQ^{-1}M)^{-1}M^TQ^{-1}$$

XV. APPENDIX F

Certain norm expressions are required for the estimate of V obtained

in section VII where V is known to be diagonal and M is vertical.

1. Norm of variance of V

$$c[\hat{V}-V] = (I-N*N)^{-1} \cdot c[(yy^T-Nyy^TN) - (V-NVN)]$$

$$[(norm)^2 \text{ of } V\text{-variance}] = tr\{E[\ [\hat{V}-V] \cdot c[\hat{V}-V]^T]\}$$

$$= tr\{(I-N*N)^{-1} E[c[yy^T-Nyy^TN - V + NVN]$$

$$\cdot c[yy^T-Nyy^TN - V + NVN]^T](I-N*N)^{-1}\}$$

$$\{E[c[yy^T-Nyy^TN-V+NVN] \cdot c[yy^T-Nyy^TN-V+NVN]^T]\}_{ij}$$

$$= E\{[y_i^2 - \sum_{k,\ell} (N)_{ik} y_k y_\ell (N)_{\ell i} - (V-NVN)_{ii}]$$

$$\cdot [y_j^2 - \sum_{m,n} (N)_{jm} y_m y_n (N)_{nj} - (V-NVN)_{jj}]\}$$

$$= (Y-NYN-V+NVN)_{ii} \cdot (Y-NYN-V+NVN)_{jj}$$

$$+ 2(Y)_{ij}^2 - 2(YN)_{ij}^2 - 2(NY)_{ij}^2 + 2(NYN)_{ij}^2$$

$$= 2[(Y)_{ij}^2 - 2(YN)_{ij}^2 + (NYN)_{ij}^2]$$

$$[(norm)^2 \text{ of } V\text{-variance}] = 2\ tr\{(I-N*N)^{-1}[Y*Y-2(YN)*(YN)$$

$$+ (NYN)*(NYN)](I-N*N)^{-1}\}$$

2. Norm of cross-products

$$[(\text{norm})^2 \text{ of } (\hat{V}_n - V_n)(\hat{V}_i - V_i)] = \text{tr}\{E[c[\hat{V}_n - V_n] \cdot c[\hat{V}_i - V_i]]\}$$

$$= \text{tr}\{c[Y_n - N_n Y_n N_n - V_n + N_n V_n N_n] \cdot c[Y_i - N_i Y_i N_i - V_i - N_i V_i N_i]^T$$

$$+ 2(M_n \Phi_n \cdots \Phi_{i+1} X_i M_i^T) * (M_n \Phi_n \cdots \Phi_{i+1} X_i M_i^T)$$

$$- 4(M_n \Phi_n \cdots \Phi_{i+1} X_i M_i N_i) * (M_n \Phi_n \cdots \Phi_{i+1} X_i M_i^T N_i)$$

$$+ 2(N_n M_n \Phi_n \cdots \Phi_{i+1} X_i M_i N_i) * (N_n M_n \Phi_n \cdots \Phi_{i+1} X_i M_i N_i)\}$$

$$= 0$$

In the above development, $n > i$.