Nc 1

NWC-TN ~~Technical Note~~ 4070-58-75

A GUIDED TOUR OF KALMAN FILTERING

by

R. L. T. Hampton
Weapons Systems Analysis Division
Systems Development Department

December 1975

This is an informal report containing preliminary information. It is not an official report of the Naval Weapons Center and must not be used as a basis for action unless enclosed in official correspondence in which the purpose of the transmittal is made clear.

NAVAL WEAPONS CENTER
China Lake, California 93555

FOREWORD

In the past decade there has been a proliferation of signal processing algorithms proposed to reconstruct information from noisy measurements. Most of these techniques have been variations or extensions of the popular Kalman filter. Too often, the literature has obscured the basic concepts of Kalman filtering and its modifications with unnecessary mathematical abstractions. It is the purpose of this report to expose the easily understood fundamentals of Kalman filtering, and to provide the reader with the foundation required to utilize the theory in real-world engineering problems. However, the transition from theory to practice is usually a non-trivial step.

This work was performed under the Naval Air Systems Command AirTask A510-5103/008-2/6235-000-143. This report is released at the working level for information purposes only.

R. B. SEELEY, *Head*
Avionics Analysis Branch
Weapons Systems Analysis Division
12 December 1975

ACKNOWLEDGMENT

The author would like to thank Linda Miller for contributing the computer plots represented by Figures 2-2 and 2-3.

NWC Technical Note 4070-58-75

Published by . . . . Technical Information Department
Manuscript 5362/MS 75-214
Collation . . . . Cover, 50 pages
First printing . . . . 30 copies
Second printing . . . . 25 copies

CONTENTS

I. INTRODUCTION

1.1 PROLOGUE

Physical systems are designed to accomplish well-defined tasks. The importance of navigation for aircraft and orbit determination for satellites is fundamental. In order to estimate whether either is likely to accomplish its objectives, and ultimately to control its performance, the engineer must know the "state" of his system. For aircraft navigation, the state consists of position and velocity; for satellites, it consists of the orbital parameters. Regardless of the problem under consideration, the state of the system is observed, often indirectly, using a suitable set of sensors referred to as the measurement system. Inevitably, these observations are contaminated with noise that is generated by the phenomena observed and the sensors themselves. Thus, the measurements yield only crude information, which may be only indirectly related to the parameters of interest. This latter point is easily seen in orbit determination problems, where the measurements typically consist of range, range rate, and the direction cosines from the observation site. Whereas, the orbital parameters to be estimated are eccentricity, semimajor axis, argument of perigee, and the mean Kepler anomaly.

The filter theory of Kalman represents, in principle, an almost ideal solution to the above type problems of smoothing noisy measurements and reconstructing estimates for parameters which can not be directly measured. It accomplishes this by optimally* utilizing any dynamic, geometric and statistical relationships that exist between the noisy data and the desired information. In addition, it processes the data sequentially. Thus, we are led to the following definition.

DEFINITION: A Kalman filter is an optimum recursive algorithm for extracting information from noise-corrupted data.

It is the purpose of this report to provide an intuitive development of the Kalman filter and to confirm this result rigorously via maximum likelihood and minimum variance estimation theory.

At this point it is important to observe that even though Kalman filtering is an excellent solution to many data processing problems, it would be of only academic interest in aerospace applications without the rapid development of powerful airborne digital computers.

* The optimality is to be interpreted in a minimum mean-square error sense.

1.2 TYPICAL APPLICATIONS

The application of modern filter theory to multisensor avionic systems began only a few years after Kalman published his original papers in the early 1960s (Ref. 1 and 2). For example, the Kalman filter provides a systematic method for optimally employing all external measurements (position, velocity, or attitude) to improve the accuracy of inertial navigation systems (Ref. 3). The block diagram for such an augmented inertial navigation system is shown in Figure 1-1.

Another important application is in precision pointing devices (Ref. 4 and 5). These systems must be able to track in a rapidly changing dynamic environment. Conversely, the bandwidth must be narrow enough to sufficiently attenuate the servo and sensor noise to achieve the accuracy required for precision pointing. These two conflicting requirements are resolved by using a modified Kalman filter to generate an aided-tracking input to the controller as shown in Figure 1-2. Since in this case the output of the filter is a rate, the aided-tracking input, in effect, adds a derivative feature to the controller.

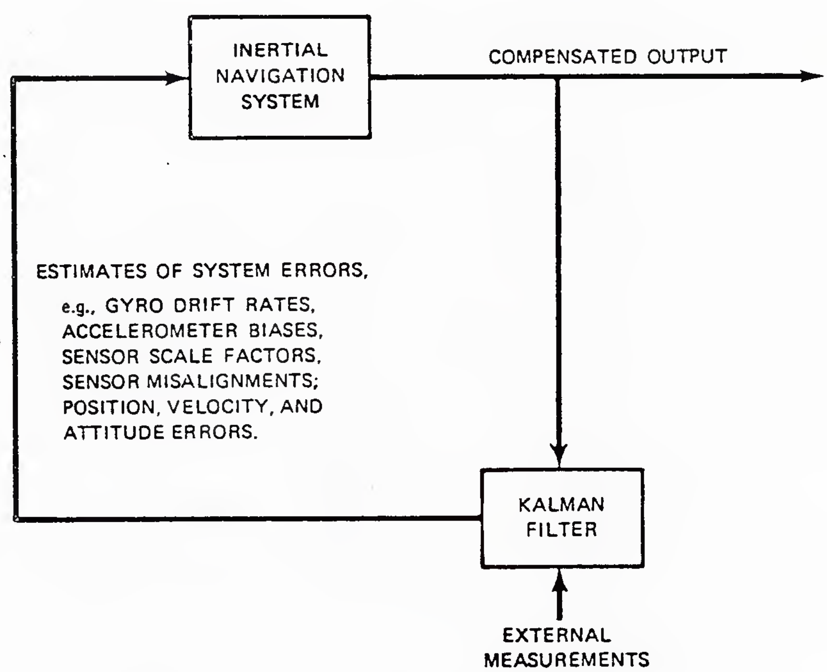

FIGURE 1-1. Feedback Compensation of Augmented Inertial Navigation Systems.

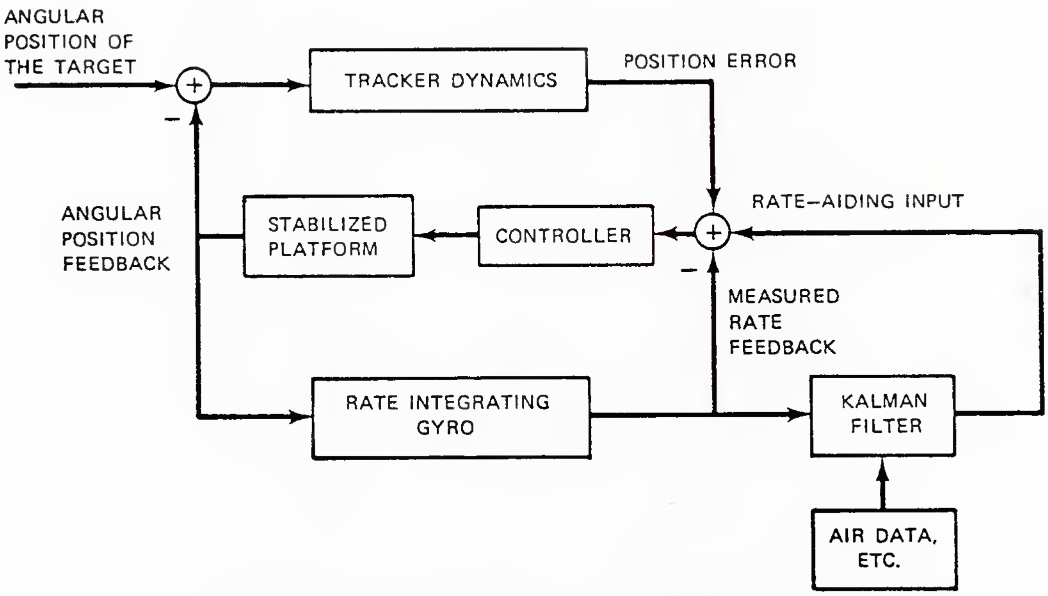

FIGURE 1-2. Rate-Aided Precision Tracking System.

1.3 SYSTEM MODELING: A CRUCIAL PRELUDE TO FILTERING

Before the filter theory of Kalman can be applied to a problem, the dynamics* of the quantities of interest and any measurements of them must be in a specific format. This format requires that all difference equations describing the dynamics be written as a set of first order difference equations. The following simple, but useful, example introduces the modeling process and will be employed later to illustrate different aspects of Kalman filtering.

Example 1.1: Radar Tracking of Nominal Constant Velocity Vehicles

The assumptions for such trajectories are:

(1) The average acceleration is zero.

(2) The acceleration between radar scans is constant, i.e.,

$$a(t) = a_{n-1} = \text{a constant} \qquad (n-1)T \leq t \leq nT$$

where T is the time between radar samples.

(3) The acceleration between scanning intervals is uncorrelated.

* Since digital computation is assumed, the dynamics will be in difference equation rather than differential equation form.

It is not possible to set the instantaneous acceleration equal to zero due to wind gusts, short-term irregularities in engine thrust, etc. Under assumptions (1) through (3), the position x and velocity $\dot{x}$ vary according to the equations

$$x_n = x_{n-1} + T\dot{x}_{n-1} + \frac{T^2}{2} a_{n-1} \tag{1-1}$$

$$\dot{x}_n = \dot{x}_{n-1} + Ta_{n-1} \qquad (n-1)T \leq t \leq nT \tag{1-2}$$

where a_{n-1} is a random sequence with zero-mean and a mean-square acceleration of

$$E\left\{a_j a_k\right\} = \begin{cases} \sigma^2 & \text{for } j = k \\ 0 & \text{otherwise} \end{cases} \tag{1-3}$$

Note that Eq. 1-1 and 1-2 are first order difference equations as required. They can be more compactly written in vector-matrix notation as

$$X_n = \Phi(n,n-1)X_{n-1} + W_{n-1} \tag{1-4}$$

where

$$X_n = \begin{bmatrix} x_n \\ \dot{x}_n \end{bmatrix} \tag{1-5}$$

is defined as the state vector at time t = nT.

$$\Phi(n,n-1) = \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix} \tag{1-6}$$

is the state transition matrix that describes the dynamic evolution of X from time (n-1)T to nT, and

$$W_{n-1} = \begin{bmatrix} T^2/2 \\ T \end{bmatrix} a_{n-1} \tag{1-7}$$

is the unknown random disturbance.

The radar measurements are represented by

$$Z_n = H(n)X_n + V_n \tag{1-8}$$

where Z_n is the observation vector at time nT, H(n) is the observation matrix, and V_n is the measurement noise. Just as with the random acceleration, we will assume that the measurement noise has zero mean and is uncorrelated from sample to sample. The H matrix is determined by the sensor set available. For example, if x(n) represents range in Eq. 1-5 and range plus range-rate sensors are available, the corresponding H matrix is

$$H = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

In the common case where only position information is present, H reduces to a row vector

$$H = [1 \quad 0] \tag{1-9}$$

Equations 1-4 and 1-8 represent the fundamental structure required in optimal filtering to estimate the state X_n in a recursive fashion from the measurements Z_n. Therefore, this form will be used as a starting point for the derivations which follow.

1.4 PHILOSOPHY AND ORGANIZATION

The discrete-time version of the Kalman filter is developed in this report; that is, we assume sampled-data observations of a dynamic system described by difference equations. The discrete time problem was chosen due to the prevalence of digital computation and because it has a number of inherent tutorial and theoretical advantages in an introductory treatment. In simple problems the discrete algorithm can be manipulated easily with a hand calculator, so that considerable insight can be gained. Furthermore, the step-by-step processing of information lends itself to a more straightforward heuristic solution than does the continuous problem. This development of Kalman filtering is presented in Section II.

Two different approaches are invoked to obtain a rigorous derivation of the optimal filter equations. The maximum likelihood method is employed for this purpose in Section III. Then minimum variance estimation theory is used to rederive the Kalman filter in Section IV. The reason for presenting two different developments, when one would be sufficient, is they illustrate the important fact that there is more than one way to view and solve the problem. More importantly, it is hoped that by approaching it from different avenues a deeper understanding of the physical and statistical features of the Kalman filtering algorithm will be achieved. In addition, Sections II, III, and IV are written so they can essentially be read independently. Also included is a section on practical aspects of Kalman filtering. Section VI summarizes the contents of this report and provides a perspective of Kalman's contribution and its relationship to problems encountered in other fields. An Appendix is included which contains several matrix definitions and theorems employed throughout the sequel.

II. A HEURISTIC DEVELOPMENT OF KALMAN FILTERING

The derivation of the filter equations presented in this section is intended to appeal primarily to intuitive reasoning. This approach is taken initially to provide insight into the characteristics of the solution that the mathematical manipulations of Sections III and IV might obscure.

2.1 PROBLEM STATEMENT

Given the discrete time dynamic system described by the state equation

$$X_n = \Phi(n,n-1)X_{n-1} + W_{n-1} \tag{2-1}$$

and measurement data related to the state by

$$Z_n = H(n)X_n + V_n \tag{2-2}$$

where

X_n is the m x 1 state vector

$\Phi(n,n-1)$ is the m x m transition matrix

W_{n-1} is the m x 1 system perturbation noise

Z_n is the p x 1 measurement vector

$H(n)$ is the p x m observation matrix

V_n is the p x 1 measurement noise

The noise sources are assumed to be zero-mean uncorrelated noise sequences with the following covariances:

$$E\left\{W_j W_k^T\right\} = Q(k)\delta_{jk}$$

$$E\left\{V_j V_k^T\right\} = R(k)\delta_{jk}$$

$$E\left\{V_j W_k^T\right\} = 0 \qquad \text{for all } j,k \tag{2-3}$$

and δ_{jk} is the Kronecker delta.* A linear unbiased recursive estimate $\hat{X}_n$

* These restrictions could be relaxed to include cross-correlated, sequentially correlated and non-zero mean noise sources, but such a generalization would significantly complicate the sequel.

of the state X_n is to be computed from the data sequence $\{Z_0, Z_1, \ldots, Z_n\}$ so that the mean-square error of the estimate is minimized.

The recursive feature restricts the solution to be an explicit function of only the present measurement Z_n and the previous estimate $\hat{X}_{n-1}$. The familiar sample-mean calculation can be used to illustrate the desired form.

Example 2.1: Sample-Mean Computation

Consider the problem of estimating a DC voltage X from a set of n-1 noisy measurements of the form

$$Z_n = X + V_n \tag{2-4}$$

Note that Eq. 2-4 is a special case of Eq. 2-2. The sample-mean estimate of the constant X results from averaging the measurements

$$\hat{X}_{n-1} = \frac{1}{n-1} \sum_{i=1}^{n-1} Z_i \tag{2-5}$$

After the next measurement, the sample mean becomes

$$\hat{X}_n = \frac{1}{n} \sum_{i=1}^{n} Z_i \tag{2-6}$$

What is needed is a recursive version of Eq. 2-6 which incorporates new data as it is received. This is accomplished by rewriting Eq. 2-6 as follows

$$\begin{aligned} \hat{X}_n &= \frac{n-1}{n}\left[\frac{1}{n-1} \sum_{i=1}^{n-1} Z_i\right] + \frac{1}{n} Z_n \\ &= \frac{n-1}{n} \hat{X}_{n-1} + \frac{1}{n} Z_n \\ &= \hat{X}_{n-1} + \frac{1}{n}\left(Z_n - \hat{X}_{n-1}\right) \end{aligned} \tag{2-7}$$

Several important facets in the structure of this equation are relevant to Kalman filtering. The estimate is an explicit function of the past estimate and the present measurement only. Thus, the necessity of storing past data and repeating the averaging operation of Eq. 2-6 is eliminated. Note also that, just before the nth measurement, the best prediction of X is simply the last estimate $\hat{X}_{n-1}$. This follows because X is a constant.

In addition, $\hat{X}_{n-1}$ is also the best predicted value for Z_n, because the noise has zero-mean. Therefore, we may interpret Eq. 2-7 as follows:

> The new estimate $\hat{X}_n$ is a linear combination of its best predicted value $\hat{X}_{n-1}$ plus an appropriately weighted difference between the new measurement Z_n and its best predicted value $\hat{X}_{n-1}$. The weighting is supplied by the time-varying gain $1/n$.

In fact, the sample-mean is the Kalman filter for the special case

$$\begin{aligned} X_{n+1} &= X_n \\ &= X \\ &= \text{a constant} \end{aligned}$$

where

$$\begin{aligned} \Phi(n+1,n) &= I \\ W_n &\equiv 0 \\ H(n) &= I \end{aligned}$$

and

$$R(n) = I$$

2.2 HEURISTIC SOLUTION

By analogy with the sample-mean example, we can proceed heuristically to a logical structure for the general Kalman filter. Suppose that at time $(n-1)T$ an optimal estimate $\hat{X}_{n-1}$ of X_{n-1} given the data $\{Z_0,\ldots,Z_{n-1}\}$ is available for the system of Eq. 2-1. Since the average value of the noise term W_n is zero, it is reasonable to extrapolate $\hat{X}_{n-1}$ forward to time nT using the dynamic model relating X_{n-1} to X_n. This gives

$$\hat{X}_{n|n-1} = \Phi(n,n-1)\hat{X}_{n-1} \tag{2-8}$$

The prediction of X_n from $\hat{X}_{n-1}$ is denoted by $\hat{X}_{n|n-1}$, and is to be interpreted as the best estimate of X_n, given measurements up through time $(n-1)T$. Based on this prediction, the anticipated value for the new observation Z_n can be expected to be

$$\hat{Z}_{n|n-1} = H(n)\hat{X}_{n|n-1} \tag{2-9}$$

This follows from the measurement equation (2-2) and the assumption that the noise V_n has zero mean. From the sample-mean example, we know that the information contained in the new measurement must also be reflected in the difference between the measurement and its predicted value. Consequently, this difference is referred to as the innovation sequence and is defined by

$$\nu_n \triangleq Z_n - \hat{Z}_{n|n-1} \tag{2-10}$$

$$= Z_n - H(n)\hat{X}_{n|n-1} \tag{2-11}$$

Obviously, the error represented by ν_n must be due to a change in the state and/or sensor noise. Thus, ν_n is proportional to the correction that we should make to the prediction $\hat{X}_{n|n-1}$ to incorporate the new information contained in Z_n. Now the filter was required to be a linear function of the present measurement Z_n and the previous filtered value $\hat{X}_{n-1}$. Motivated by this restriction, suppose we construct this estimate $\hat{X}_n$ by a linear combination of the predicted value, and the innovation, Eq. 2-8 and 2-11 respectively. Furthermore, by analogy with the sample-mean algorithm, let this linear combination have the form

$$\hat{X}_n = \hat{X}_{n|n-1} + K_n\left[Z_n - H(n)\hat{X}_{n|n-1}\right] \tag{2-12}$$

Note that Eq. 2-12 is an explicit function of $\hat{X}_{n-1}$ and Z_n only, because from Eq. 2-8 it can be rewritten as

$$\hat{X}_n = \Phi(n,n-1)\hat{X}_{n-1} + K_n\left[Z_n - H(n)\Phi(n,n-1)\hat{X}_{n-1}\right] \tag{2-13}$$

Observe that Eq. 2-12 has the intuitively satisfying structure of a predictor-corrector algorithm.

However, the m x p gain matrix K_n must still be chosen to minimize the mean-square error, where the error itself is defined by

$$e_n = X_n - \hat{X}_n \tag{2-14}$$

An expression for this error as a function of K_n can be obtained by substituting Eq. 2-12 into Eq. 2-14

$$\begin{aligned} e_n &= X_n - \hat{X}_{n|n-1} - K_n\left[Z_n - H(n)\hat{X}_{n|n-1}\right] \\ &= e_{n|n-1} - K_n\left[H(n)X_n - H(n)\hat{X}_{n|n-1} + V_n\right] \\ &= e_{n|n-1} - K_n\left[H(n)e_{n|n-1} + V_n\right] \\ &= \left[I - K_nH(n)\right]e_{n|n-1} - K_nV_n \end{aligned} \tag{2-15}$$

Equation 2-15 is a function of the error in the predicted estimate, which can be written as

$$e_{n|n-1} \triangleq X_n - \hat{X}_{n|n-1}$$

$$= \Phi(n,n-1)X_{n-1} + W_{n-1} - \Phi(n,n-1)\hat{X}_{n-1}$$

$$= \Phi(n,n-1)e_{n-1} + W_{n-1} \tag{2-16}$$

Since it was specified that the filter estimate $\hat{X}_n$ be unbiased, i.e.,

$$E\{\hat{X}_n\} = X_n$$

it follows that the average value of e_n must be zero. Thus,

$$E\{e_n\} = 0 \tag{2-17}$$

Therefore, from Eq. 2-16, we have

$$E\{e_{n|n-1}\} = \Phi(n,n-1)E\{e_{n-1}\} + E\{W_{n-1}\} = 0 \tag{2-18}$$

because W_{n-1} is also zero-mean. Equation 2-18 implies that $\hat{X}_{n|n-1}$ is unbiased and, consequently, the covariance of the error in this estimate is from Eq. 2-16

$$E\left\{e_{n|n-1}e^T_{n|n-1}\right\} = \Phi(n,n-1)E\left\{e_{n-1}e^T_{n-1}\right\}\Phi^T(n,n-1) + E\left\{W_{n-1}W^T_{n-1}\right\} \tag{2-19}$$

Using Eq. 2-3, the last expectation in Eq. 2-19 is Q(n-1). Denoting the error covariance matrix by P, Eq. 2-19 can be written

$$P_{n|n-1} = \Phi(n,n-1)P_{n-1}\Phi^T(n,n-1) + Q(n-1) \tag{2-20}$$

This equation is known as <u>the discrete Riccati equation.</u>

Employing Eq. 2-15, the error covariance of the filtered estimate error is

$$P_n \triangleq E\left\{e_n e_n^T\right\} \tag{2-21}$$

$$= E\left\{\left[I - K_nH(n)\right]e_{n|n-1}\left(e^T_{n|n-1}\left[I - K_nH(n)\right]^T + V_n^TK_n^T\right)\right.$$

$$\left. + K_nV_n\left(e^T_{n|n-1}\left[I - K_nH(n)\right]^T + V_n^TK_n^T\right)\right\} \tag{2-22}$$

As a result of the uncorrelated and zero-mean properties of the measurement noise,

$$E\left\{e_{n|n-1}V_n^T\right\} = 0 \tag{2-23}$$

Consequently, P_n reduces to

$$P_n = \left[I - K_nH(n)\right]P_{n|n-1}\left[I - K_nH(n)\right]^T + K_nR(n)K_n^T \tag{2-24}$$

Now K_n must be selected so that the mean-square error

$$\text{M.S.E.} = E\left\{e_n^Te_n\right\} \tag{2-25}$$

is minimized. By comparing Eq. 2-25 with Eq. 2-21, it is evident that the mean-square error is the trace of the error covariance matrix. Thus, one approach to choosing K_n is to minimize the trace of P_n. This is accomplished by setting the partial derivative of the trace of P_n equal to zero and solving for K_n.

$$\frac{\partial\left[\text{trace } P_n\right]}{\partial K_n} = 0 \tag{2-26}$$

First, substitute Eq. 2-24 into Eq. 2-26 and utilize the matrix relationship

$$\frac{\partial}{\partial D}\left[\text{trace}\left(DED^T\right)\right] = 2DE \tag{2-27}$$

when E is symmetric. Remembering that all covariance matrices are symmetric, the differentiation in Eq. 2-26 gives

$$-2\left[I - K_nH(n)\right]P_{n|n-1}H^T(n) + 2K_nR(n) = 0$$

Solving for K_n yields

$$K_n = P_{n|n-1}H^T(n)\left[H(n)P_{n|n-1}H^T(n) + R(n)\right]^{-1} \tag{2-28}$$

Substitution of this expression into Eq. 2-24 gives, after some manipulation,

$$\begin{aligned} P_n &= P_{n|n-1} - P_{n|n-1}H^T(n)\left[H(n)P_{n|n-1}H^T(n) + R(n)\right]^{-1}H(n)\,P_{n|n-1} \\ &= \left[I - K_nH(n)\right]P_{n|n-1} \end{aligned} \tag{2-29}$$

The development of the optimal recursive filter is almost complete. There remains only the specification of the initial conditions for the filter. These are determined from the requirement that the filter estimates be unbiased. Since V_n is unbiased, it follows from Eq. 2-15 that

$$E\left\{e_n\right\} = \left[I - K_n H(n)\right] E\left\{e_{n|n-1}\right\} \tag{2-30}$$

Clearly, $E\{e_n\}$ is zero (i.e., $\hat{X}_n$ is unbiased) if the prediction $\hat{X}_{n|n-1}$ is unbiased. And from Eq. 2-16, $\hat{X}_{n|n-1}$ is unbiased if $\hat{X}_n$ is unbiased. Carrying this argument to its logical conclusion implies that the estimates are unbiased for all n if, and only if, the initial condition for the predicted estimate is chosen to be unbiased. That is,

$$E\left\{\hat{X}_{0|-1}\right\} = E\left\{X_0\right\} \tag{2-31}$$

This initial condition and its covariance $P_{0|-1}$ must be given. In practice, an educated guess for $\hat{X}_{0|-1}$ and $P_{0|-1}$, obtained from a priori or externally computed information, is often used.

Equations 2-8, 2-12, 2-20, 2-28, 2-29, and 2-31 comprise the algorithm referred to as the Kalman filter. They are summarized for convenient reference in Table 2-1. The associated block diagram along with the system and measurement models is illustrated in Figure 2-1.

Note that the filter contains the dynamic model within it, around which is a negative feedback loop containing the measurement model. The error signal driving the filter is the innovation process, which represents the new information contained in the measurements. Thus, the Kalman filter can be viewed as a model-following feedback control system.

One significant attribute of the Kalman filter is its recursive structure. This property makes it ideally suited for sequential real-time digital processing of new data. Another feature that adds to the popularity of the Kalman filter is the generation of its own error analysis via $P_{n|n-1}$ and P_n, which are inherently calculated as part of the algorithm. However, the most important property of the filter is its ability to reconstruct the entire state vector from a noisy scalar measurement of the state vector. This capability is dependent on the observability concept. Mathematically the system of Eq. 2-1 and 2-2 is observable if, and only if, the m x m symmetric matrix M(0,N) is positive definite for some N > 0, where

$$M(0,N) = \sum_{i=1}^{N} \Phi^T(i,0)H^T(i)H(i)\Phi(i,0)$$

and

$$\Phi(i,0) = \Phi(i,i-1)\Phi(i-1,i-2)\cdots\Phi(1,0)$$

Physically, this means that the measurement must be related dynamically to all the elements in the state vector.

TABLE 2-1. Summary of the Discrete Time Kalman Filter Algorithm.

DYNAMIC MODEL	$X_n = \Phi(n,n-1)X_{n-1} + W_{n-1}$ $W_n \sim N(0,Q_n)$	(2-1)
MEASUREMENT MODEL	$Z_n = H(n)X_n + V_n$ $V_n \sim N(0,R_n)$	(2-2)
NOISE STATISTICS	$E\left\{V_j\right\} = 0,\ E\left\{W_j\right\} = 0$	
	$E\left\{V_j V_k^T\right\} = R(k)\delta_{jk},\ E\left\{V_j W_k^T\right\} = 0$	(2-3)
	$E\left\{W_j W_k^T\right\} = Q(k)\delta_{jk}$	
PREDICTED ESTIMATE	$\hat{X}_{n\|n-1} = \Phi(n,n-1)\hat{X}_{n-1}$ $\hat{x}_k(-) = \Phi_{k-1}\hat{x}_{k-1}(+)$	(2-8)
COVARIANCE FOR PREDICTION ERROR	$P_k(-)$ $P_{n\|n-1} = \Phi(n,n-1)P_{n-1}\Phi^T(n,n-1) + Q(n-1)$	(2-20)
FILTER GAIN MATRIX	$K_n = P_{n\|n-1}H^T(n)\left[H(n)P_{n\|n-1}H^T(n) + R(n)\right]^{-1}$	(2-28)
FILTERED ESTIMATE	$\hat{X}_n = \hat{X}_{n\|n-1} + K_n\left[Z_n - H(n)\hat{X}_{n\|n-1}\right]$	(2-12)
COVARIANCE FOR FILTER ERROR	$P_n = \left[I - K_n H(n)\right]P_{n\|n-1}$	(2-29)
INITIAL CONDITIONS	$E\left\{\hat{X}_{0\|-1}\right\} = E\left\{X_0\right\}$	(2-31)
	$P_{0\|-1} = E\left\{e_{0\|-1}e_{0\|-1}^T\right\}$	

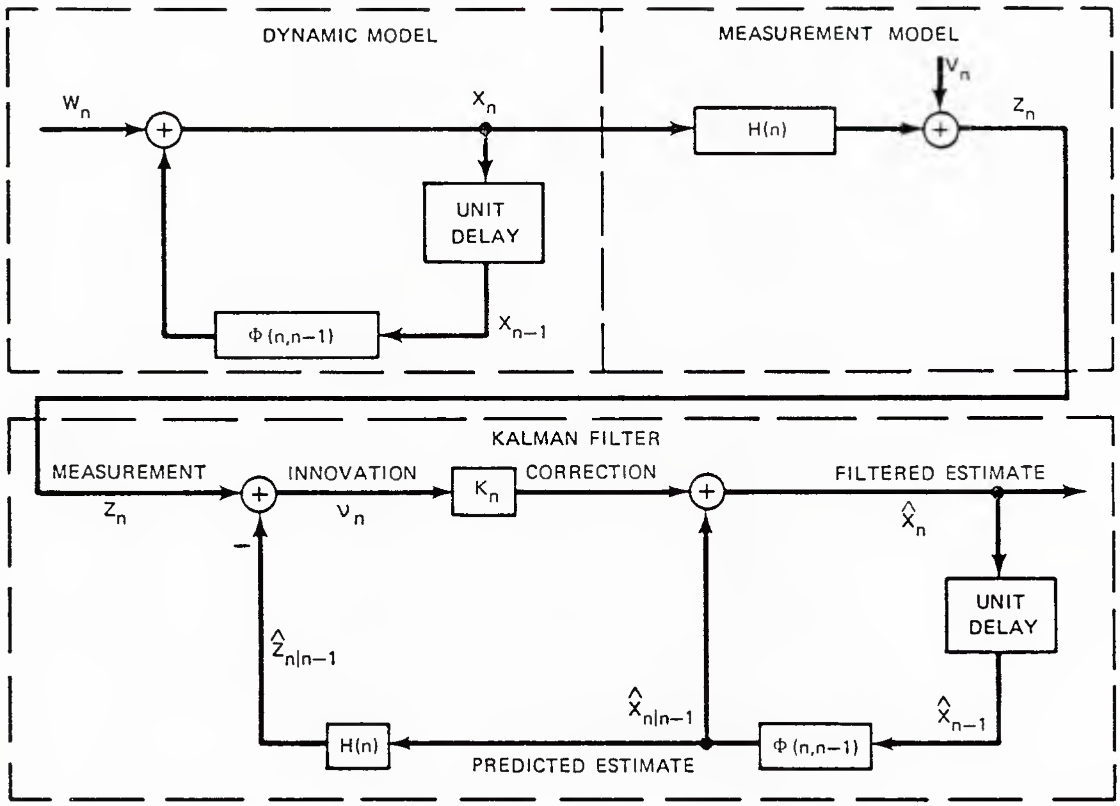

FIGURE 2-1. Discrete Time Kalman Filter and the Associated System Model.

Although our development has been heuristic, it can be made rigorous as will be shown in Sections III and IV. In fact, it will become evident that the Kalman filter is not only the best linear filter for the system of Eq. 2-1 and 2-2, but also the best filter (linear or nonlinear) if the noise sequences are gaussian or if only first and second order statistics are known.

Before proceeding to Section III, a specific problem will be solved and discussed in order to enhance the reader's intuitive feel for the characteristics of the Kalman filter. The steady-state Kalman filter solution will then be related to the familiar transfer function method of filter design.

Example 2.2: Design of an Alpha-Beta Filter Using the Kalman Algorithm

Given the model of Example 1.1, we wish to estimate position and velocity from noisy position measurements only. If the measurements are available twice per second, then the system Eq. 1-4 through 1-8 are

$$\begin{bmatrix} x_n \\ \dot{x}_n \end{bmatrix} = \begin{bmatrix} 1 & 1/2 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_{n-1} \\ \dot{x}_{n-1} \end{bmatrix} + \begin{bmatrix} 1/8 \\ 1/2 \end{bmatrix} a_{n-1} \qquad (2\text{-}32)$$

$$Z_n = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} x_n \\ \dot{x}_n \end{bmatrix} + V_n \qquad (2\text{-}33)$$

$$= x_n + V_n \qquad (2\text{-}34)$$

Now assuming the root mean-square magnitude of the perturbation acceleration is 2 meters per second and since

2

$$W_{n-1} = \begin{bmatrix} 1/8 \\ 1/2 \end{bmatrix} a_{n-1}$$

the Q matrix can be calculated using Eq. 1-7.

$$Q(n-1) = E\left\{W_{n-1} W_{n-1}^T\right\}$$

$$= \begin{bmatrix} 1/16 & 1/4 \\ 1/4 & 1 \end{bmatrix} \qquad (2\text{-}35)$$

The initial condition $P_{0|-1}$ is chosen to be

$$P_{0|-1} = \begin{bmatrix} 10 & 0 \\ 0 & 10 \end{bmatrix} \qquad (2\text{-}36)$$

To display the automatic variation of the Kalman gain matrix to compensate for a time-varying environment, let the covariance of the observation noise be

$$R(n) = \begin{cases} 4 + (-2)^n & 0 \leq n \leq 10 \\ 4 & 10 < n \end{cases} \qquad (2\text{-}37)$$

From an inspection of Eq. 2-37, it is not difficult to see that the gain matrix K_n should be increased when processing measurements at odd values of n to take advantage of the relatively less noisy observations. Thus, there will be less lag in the filtered estimates, and the error covariances should be smaller. Correspondingly, the gains should decrease for even-numbered measurements to suppress the increased noise level. This, however, will increase both the filter lag and the error covariances. Figures 2-2 and 2-3 graphically display these characteristics of the Kalman algorithm. Note from Figure 2-2 that the filter gains, computed from Eq. 2-28, reach a periodic steady-state quickly.

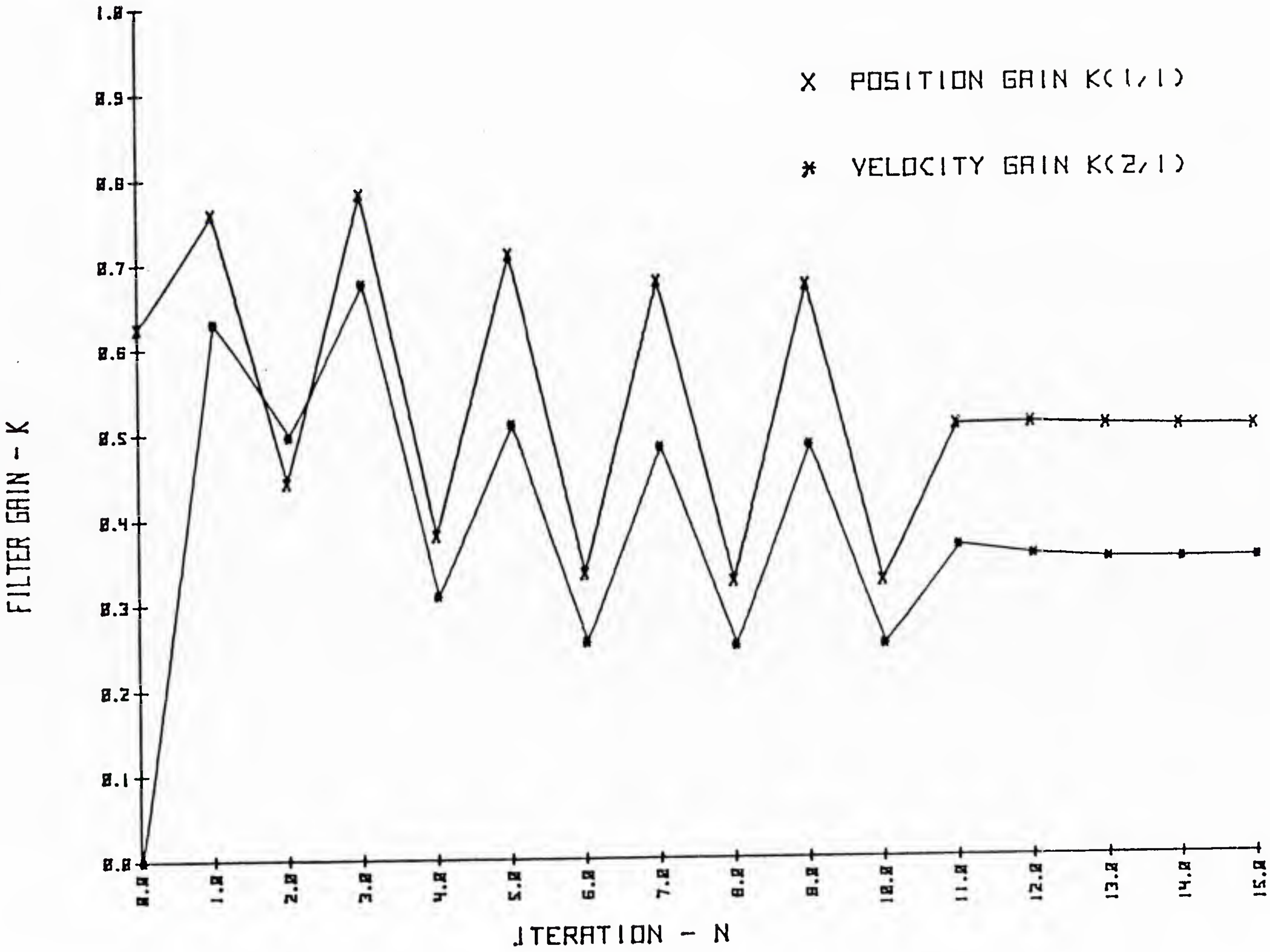

FIGURE 2-2. Filter Gain Histories.

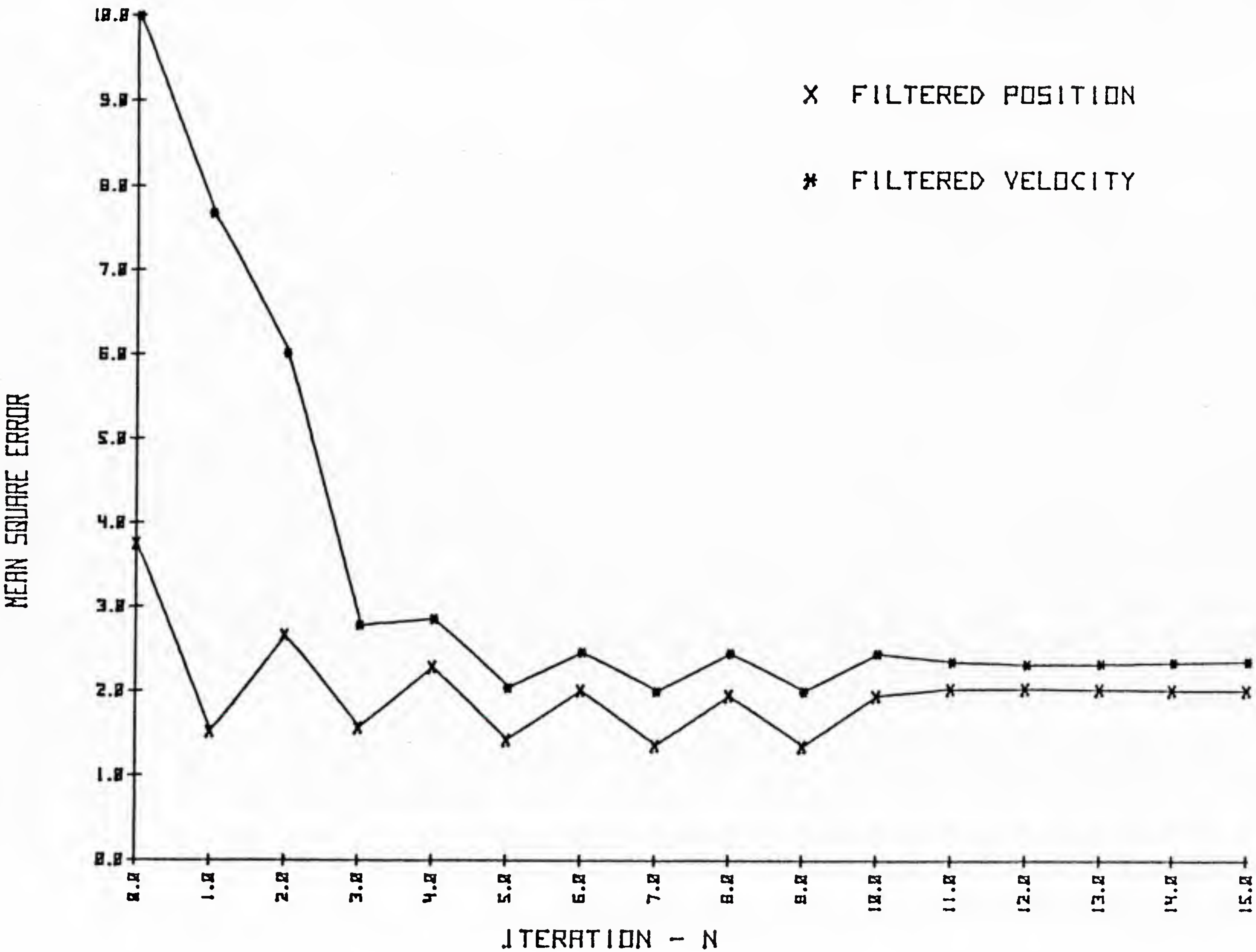

FIGURE 2-3. Filtered Mean-Square Errors.

After the 10th measurement, the covariance of the noise remains constant at four. Consequently, the filter gains become constant after the transient effects of this change die out.

This example serves to introduce another facet of the Kalman algorithm. Alpha-beta trackers were originally devised to predict position and velocity. Such a predictor can be obtained directly, without explicit filtering, by substituting Eq. 2-12 into Eq. 2-8. The result is

$$\hat{X}_{n+1|n} = \Phi(n+1,n)\hat{X}_{n|n-1} + G_{n+1}\left[Z_n - H(n)\hat{X}_{n|n-1}\right] \tag{2-38}$$

where

$$G_{n+1} = \Phi(n+1,n)K_n \tag{2-39}$$

Therefore, if the filter gain is known, the prediction gain G_{n+1} can be quickly determined from Eq. 2-39. Conversely, the filter gain may be calculated from the prediction gain through the inverse relationship

$$K_n = \Phi^{-1}(n+1,n)G_{n+1} \tag{2-40}$$

Since many graphs exist for choosing G_{n+1} given a desired damping ratio and bandwidth, Eq. 2-40 is particularly useful for obtaining the corresponding filter gains.

Under steady-state conditions, it is possible to determine the general filter transfer functions for our example. This will allow us to interpret Kalman's time domain solution in terms of familiar frequency domain concepts. Rearranging Eq. 2-8 and 2-12 and collecting the coefficients of the filtered estimates into one matrix gives

$$\begin{bmatrix} \hat{x}_n \\ \hat{\dot{x}}_n \end{bmatrix} = \begin{bmatrix} 1-K(1,1) & [1-K(1,1)]T \\ -K(2,1) & 1-TK(2,1) \end{bmatrix} \begin{bmatrix} \hat{x}_{n-1} \\ \hat{\dot{x}}_{n-1} \end{bmatrix} + \begin{bmatrix} K(1,1) \\ K(2,1) \end{bmatrix} Z_n \tag{2-41}$$

$$\begin{bmatrix} \hat{x}_{n+1|n} \\ \hat{\dot{x}}_{n+1|n} \end{bmatrix} = \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \hat{x}_n \\ \hat{\dot{x}}_n \end{bmatrix} \tag{2-42}$$

Note that the sampling time T between measurements has been retained in Eq. 2-41 and 2-42 for the sake of generality. Taking the Z transform of these two equations and applying simple matrix manipulations, the transfer functions for filtered position x_F, filtered velocity $\dot{x}_F$, predicted position x_P and predicted velocity $\dot{x}_P$ are found to be

$$\frac{x_F(Z)}{Z(Z)} = \frac{K(1,1) + \left[TK(2,1) - K(1,1)\right] Z^{-1}}{1 + \left[-2 + K(1,1) + TK(2,1)\right] Z^{-1} + \left[1 - K(1,1)\right] Z^{-2}} \qquad (2\text{-}43)$$

$$\frac{\dot{x}_F(Z)}{Z(Z)} = \frac{K(2,1)\left[1 - Z^{-1}\right]}{(\ \bullet\ )} \qquad (2\text{-}44)$$

$$\frac{x_P(Z)}{Z(Z)} = \frac{Z^{-1}\left[K(1,1) + TK(2,1) - K(1,1)Z^{-1}\right]}{(\ \bullet\ )} \qquad (2\text{-}45)$$

$$\frac{\dot{x}_P(Z)}{Z(Z)} = \frac{K(2,1)Z^{-1}\left(1 - Z^{-1}\right)}{(\ \bullet\ )} \qquad (2\text{-}46)$$

Observe that the characteristic equation is the same for all four transfer functions as indicated by the (•) notation.

From these Z-transfer functions, the corresponding Laplace transfer functions may be obtained by using a suitable inverse mapping. In low pass filters, the simple backward difference approximation to S is usually sufficient, i.e.,

$$S \doteq \frac{1 - Z^{-1}}{T}$$

or

$$Z^{-1} \doteq 1 - TS$$

The Kalman gains can then be used to find the S-plane pole-zero configuration. From this S-plane plot it is obvious that the gains determine the damping ratio and undamped natural frequency of the filter. Thus, the frequency domain effect of increasing the gains, when the measurement noise decreases, is an increase in the filter bandwidth. The converse is true when the measurement noise increases.

III. MAXIMUM LIKELIHOOD ESTIMATION FOR LINEAR DYNAMIC SYSTEMS

3.1 PROBLEM FORMULATION

Assume that the quantities to be estimated are the states of a dynamic system. The equation that describes the state propagation is

$$X_{n+1} = \Phi(n+1,n)X_n + W_n \tag{3-1}$$

where

X_n is the m x 1 state vector

W_n is an m x 1 system perturbation noise

$\Phi(n+1,n)$ is the m x m state transition matrix.

The measurements are related to the state in the following way

$$Z_n = H(n)X_n + V_n \tag{3-2}$$

where

Z_n is the p x 1 vector of measurements

H(n) is the p x m observation matrix, which relates the measurements to the state vector

V_n is a p x 1 vector which represents the noise or errors in the measurements.

The noise vectors W_n and V_n are assumed to be stochastically independent gaussian vectors with zero means and the following covariances:

$$E\left\{W_j W_k^T\right\} = Q(k)\delta_{jk}$$

$$E\left\{V_j V_k^T\right\} = R(k)\delta_{jk}$$

$$E\left\{W_j V_k^T\right\} = 0 \quad \text{for all } j,k \tag{3-3}$$

where δ_{jk} is the Kronecker delta. In addition X_0 is assumed to have a gaussian distribution.

The problem can be stated as follows: given the observations $Z_0, Z_1, \ldots, Z_n$, find the "best" estimate $\hat{X}_n$ of X_n. If $n = N$, we have a filtering problem, and if $n > N$, we have a prediction problem. Here we will be concerned primarily with filtering. However, the one-step predictor is obtained automatically in the filtering solution. The criterion for the "best" estimate will be determined from the maximum likelihood principle.

3.2 THE METHOD OF MAXIMUM LIKELIHOOD

Maximum likelihood is a classic technique for estimation developed by R. A. Fisher. The basic concept is quite simple: define a likelihood function of the state X_n and measurements $Z_0 = z_0, Z_1 = z_1, \ldots, Z_n = z_n$

$$L\left(X_n, z_0, \ldots, z_n\right)$$

This function is maximized with respect to X_n by solving

$$\left.\frac{\partial\, L\left(X_n, z_0, \ldots, z_n\right)}{\partial\, X_n}\right|_{X_n = \hat{X}_n} = 0 \tag{3-4}$$

such that

$$\left.\frac{\partial^2\, L\left(X_n, z_0, \ldots, z_n\right)}{\partial\, X_n^2}\right|_{X_n = \hat{X}_n} < 0 \tag{3-5}$$

In this development, it will be crucial to distinguish between the random sequence $Z_0, Z_1, \ldots, Z_n$ where each Z_i is a function and a particular realization of this sequence $Z_0 = z_0, Z_1 = z_1, \ldots, Z_n = z_n$ where each z_i is a number.

Generally, the likelihood function is selected to be the joint probability density function (p.d.f.),

$$L\left(X_n, z_0, \ldots, z_n\right) = p\left(X_n, Z_0 = z_0, \ldots, Z_n = z_n\right) \tag{3-6}$$

However, it is more convenient to maximize the marginal distribution of X_n conditioned on $Z_0 = z_0, \ldots, Z_n = z_n$, i.e., choose

$$L\left(X_n, z_0, \ldots, z_n\right) = p\left(X_n \middle| Z_0 = z_0, \ldots, Z_n = z_n\right) \tag{3-7}$$

The estimate $\hat{X}_n$ obtained by solving

$$\left.\frac{\partial\, p\left(X_n \middle| Z_0 = z_0, \ldots, Z_n = z_n\right)}{\partial\, X_n}\right|_{X_n = \hat{X}_n} = 0 \tag{3-8}$$

is known as the Bayesian maximum likelihood estimate (B.M.L.E). For the case of linear systems with gaussian noise with which we are dealing, both the likelihood functions of Eq. 3-6 and 3-7 yield identical results. In addition, since the logarithm of a function varies monotonically with that function, an equivalent and simpler likelihood function is

$$L\left(X_n, z_0, \ldots, z_n\right) = \ell n\ p\left(X_n \middle| Z_0 = z_0, \ldots, Z_n = z_n\right) \tag{3-9}$$

Therefore, we will work with Eq. 3-9.

3.3 SOLUTION OF THE STATE ESTIMATION PROBLEM

Using the method of maximum likelihood to find the conditional p.d.f. required in Eq. 3-9, we employ Bayes rule

$$p\left(X_n \middle| Z_0, \ldots, Z_n\right) = \frac{p\left(X_n, Z_0, \ldots, Z_n\right)}{p\left(Z_0, \ldots, Z_n\right)} \tag{3-10}$$

Applying Bayes rule, the numerator of Eq. 3-10 can be decomposed as follows

$$p\left(X_n, Z_0, \ldots, Z_n\right)$$
$$= p\left(Z_n \middle| X_n, Z_0, \ldots, Z_{n-1}\right) p\left(X_n \middle| Z_0, \ldots Z_{n-1}\right) p\left(Z_0, \ldots, Z_{n-1}\right) \tag{3-11}$$

Now

$$p\left(Z_n \middle| X_n, Z_0, \ldots, Z_{n-1}\right) = p\left(H(n)X_n + V_n \middle| X_n, Z_0, \ldots, Z_{n-1}\right) \tag{3-12}$$

$$= p\left(H(n)X_n\right) + p\left(V_n \middle| Z_0, \ldots, Z_{n-1}\right)$$

$$= p\left(H(n)X_n\right) + p\left(V_n\right) \tag{3-13}$$

The last equation follows since V_n is stochastically independent of $Z_0, \ldots, Z_{n-1}$. But Eq. 3-13 is identical to the p.d.f. of Z_n conditioned on X_n. Therefore,

$$p\left(Z_n \middle| X_n, Z_0, \ldots, Z_{n-1}\right) = p\left(Z_n \middle| X_n\right) \tag{3-14}$$

From this fact, Eq. 3-11 reduces to

$$p\left(X_n, Z_0, \ldots, Z_n\right) = p\left(Z_n | X_n\right) p\left(X_n | Z_0, \ldots, Z_{n-1}\right) p\left(Z_0, \ldots, Z_{n-1}\right) \tag{3-15}$$

Substituting Eq. 3-15 into Eq. 3-10 gives

$$p\left(X_n | Z_0, \ldots, Z_n\right) = \frac{p\left(Z_n | X_n\right) p\left(X_n | Z_0, \ldots, Z_{n-1}\right) p\left(Z_0, \ldots, Z_{n-1}\right)}{p\left(Z_0, \ldots, Z_n\right)} \tag{3-16}$$

Dividing the numerator and denominator of Eq. 3-16 by $p(Z_0, \ldots, Z_{n-1})$ simplifies it to

$$p\left(X_n | Z_0, \ldots, Z_n\right) = \frac{p\left(Z_n | X_n\right) p\left(X_n | Z_0, \ldots, Z_{n-1}\right)}{p\left(Z_n | Z_0, \ldots, Z_{n-1}\right)} \tag{3-17}$$

From this result, we can find the conditional p.d.f. of X_n given $Z_0, \ldots, Z_n$ by evaluating the p.d.f.'s on the right-hand side of Eq. 3-17. To accomplish this, it is necessary to remember that any linear transformation of a gaussian process is also a gaussian process. Consequently, both X and Z are gaussian, since they are generated by linear combinations of gaussian random sequences. This is the reason it was necessary to require the distribution for X_0 to be gaussian in Section 3.1. As a consequence of this structure for X and Z, only the mean and covariance of each p.d.f. in Eq. 3-17 need be determined.

Considering $p(Z_n | X_n)$ first, its mean value is

$$E\left\{Z_n | X_n\right\} = E\left\{H(n)X_n + V_n | X_n\right\}$$

$$= H(n)X_n \tag{3-18}$$

Equation 3-18 follows, since $\{V_n\}$ is a zero mean process that is independent of X_n. The covariance is

$$\text{Cov}\left\{Z_n | X_n\right\} = E\left\{\left[Z_n - H(n)X_n\right]\left[Z_n - H(n)X_n\right]^T | X_n\right\} \tag{3-19}$$

$$= E\left\{V_n V_n^T\right\}$$

$$= R(n) \tag{3-20}$$

Thus, the gaussian p.d.f. $p(Z_n|X_n)$ is

$$p\left(Z_n|X_n\right) = \frac{\exp\left\{-\frac{1}{2}\left[Z_n - H(n)X_n\right]^T R^{-1}(n)\left[Z_n - H(n)X_n\right]\right\}}{\sqrt{(2\pi)^p \det R(n)}} \tag{3-21}$$

Next we will define the mean and covariance of $p(X_n|Z_0,\ldots,Z_{n-1})$.

$$E\left\{X_n|Z_0,\ldots,Z_{n-1}\right\} \triangleq \hat{X}_{n|n-1} \tag{3-22}$$

Later, justification will be given for defining $E\{X_n|Z_0,\ldots,Z_{n-1}\}$ to be the same as the maximum likelihood solution to the one-step prediction problem. The error covariance for the prediction is

$$\begin{aligned}\mathrm{Cov}\left\{X_n|Z_0,\ldots,Z_{n-1}\right\} &= E\left\{\left[X_n - \hat{X}_{n|n-1}\right]\left[X_n - \hat{X}_{n|n-1}\right]^T|Z_0,\ldots,Z_{n-1}\right\} \\ &\triangleq P_{n|n-1}\end{aligned} \tag{3-23}$$

Therefore, the p.d.f. for X_n conditioned on $Z_0,\ldots,Z_{n-1}$ can be written

$$p\left(X_n|Z_0,\ldots,Z_{n-1}\right) = \frac{\exp\left\{-\frac{1}{2}\left[X_n - \hat{X}_{n|n-1}\right]^T P^{-1}_{n|n-1}\left[X_n - \hat{X}_{n|n-1}\right]\right\}}{\sqrt{(2\pi)^m \det P_{n|n-1}}} \tag{3-24}$$

Actually, it is not necessary to determine the p.d.f. $p(Z_n|Z_0,\ldots,Z_{n-1})$ in the denominator of Eq. 3-17 because in the likelihood equation the numerical values $z_0,\ldots,z_n$ for the measurement sequence $Z_0,\ldots,Z_n$ are used. Therefore, $p(Z_n = z_n|Z_0 = z_0,\ldots,Z_{n-1} = z_{n-1})$ is just a normalizing constant and does not enter into the minimization process of Eq. 3-4. However, the justification for the notational definitions in Eq. 3-22 and 3-23 is facilitated by doing so. Hence,

$$\begin{aligned}E\left\{Z_n|Z_0,\ldots,Z_{n-1}\right\} &= E\left\{H(n)X_n + V_n|Z_0,\ldots,Z_{n-1}\right\} \\ &= H(n)E\left\{X_n|Z_0,\ldots,Z_{n-1}\right\} && (3\text{-}25) \\ &= H(n)\hat{X}_{n|n-1} && (3\text{-}26)\end{aligned}$$

To obtain Eq. 3-25 the fact that $\{V_n\}$ is an independent zero-mean process has been exploited. Equation 3-26 follows from Eq. 3-22. The covariance is

$$\mathrm{Cov}\left\{Z_n \mid Z_0,\ldots,Z_{n-1}\right\} = E\left\{\left[Z_n - H(n)\hat{X}_{n|n-1}\right]\left[Z_n - H(n)\hat{X}_{n|n-1}\right]^T \mid Z_0,\ldots,Z_{n-1}\right\}$$

$$= H(n)E\left\{\left[X_n - \hat{X}_{n|n-1}\right]\left[X_n - \hat{X}_{n|n-1}\right]^T \mid Z_0,\ldots,Z_{n-1}\right\}H^T(n)$$

$$+ E\left\{V_n V_n^T\right\} \tag{3-27}$$

$$= H(n)P_{n|n-1}H^T(n) + R(n) \tag{3-28}$$

Equation 3-27 is derived by substituting for Z_n and again employing the independent zero-mean property of V_n. Using the definitions for the prediction error covariance Eq. 3-23 and the measurement covariance Eq. 3-3 yield Eq. 3-28. The gaussian p.d.f. for Z_n conditioned on $Z_0,\ldots,Z_{n-1}$ can now be written.

$$p\left(Z_n \mid Z_0,\ldots,Z_{n-1}\right) = A\exp\left\{-\frac{1}{2}\left[Z_n - H(n)\hat{X}_{n|n-1}\right]^T\left[H(n)P_{n|n-1}H^T(n) + R(n)\right]^{-1}\right.$$

$$\left.\left[Z_n - H(n)\hat{X}_{n|n-1}\right]\right\} \tag{3-29}$$

where

$$A = \left[(2\pi)^p \det\left\{H(n)P_{n|n-1}H^T(n) + R(n)\right\}\right]^{-1/2}$$

By substituting Eq. 3-21, 3-24, and 3-29 into Eq. 3-17, $p(X_n \mid Z_0,\ldots,Z_n)$ is found to be

$$p\left(X_n \mid Z_0,\ldots,Z_n\right) = B\exp\left\{-\frac{1}{2}\left(\left[Z_n - H(n)X_n\right]^T R^{-1}(n)\left[Z_n - H(n)X_n\right]\right.\right.$$

$$+ \left[X_n - \hat{X}_{n|n-1}\right]^T P_{n|n-1}^{-1}\left[X_n - \hat{X}_{n|n-1}\right]$$

$$\left.\left. - \left[Z_n - H(n)\hat{X}_{n|n-1}\right]^T\left[H(n)P_{n|n-1}H^T(n) + R(n)\right]^{-1}\left[Z_n - H(n)\hat{X}_{n|n-1}\right]\right)\right\} \tag{3-30}$$

where

$$B = \left[(2\pi)^m \det R(n) \det P_{n|n-1}\right]^{-1/2}\left[\det\left\{H(n)P_{n|n-1}H^T(n) + R(n)\right\}\right]^{1/2} \tag{3-31}$$

The likelihood function is determined by substituting Eq. 3-30 into Eq. 3-9, yielding

$$L\left(X_n, z_0, \ldots, z_n\right) = -\frac{1}{2}\left[z_n - H(n)X_n\right]^T R^{-1}(n)\left[z_n - H(n)X_n\right]$$

$$-\frac{1}{2}\left[X_n - \hat{X}_{n|n-1}\right]^T P_{n|n-1}^{-1}\left[X_n - \hat{X}_{n|n-1}\right]$$

$$+ \ell n\, B + C \tag{3-32}$$

where C is the constant

$$C = \frac{1}{2}\left[z_n - H(n)\hat{X}_{n|n-1}\right]^T\left[H(n)P_{n|n-1}H^T(n) + R(n)\right]^{-1}$$

$$\left[z_n - H(n)\hat{X}_{n|n-1}\right] \tag{3-33}$$

and $\ell n\, B$ is the natural logarithm of the constant defined in Eq. 3-31. Equation 3-33 confirms our earlier observation that $p(Z_n = z_n | Z_0, \ldots, Z_{n-1} = z_{n-1})$ is simply a normalizing constant, thereby eliminating the necessity of computing $p(Z_n | Z_0, \ldots, Z_{n-1})$. The likelihood equation is found from Eq. 3-4 to be

$$H^T(n)R^{-1}(n)\left[z_n - H(n)X_n\right] - P_{n|n-1}^{-1}\left[X_n - \hat{X}_{n|n-1}\right]\Bigg|_{X_n=\hat{X}_n} = 0 \tag{3-34}$$

The solution to Eq. 3-34 is the B.M.L.E. $\hat{X}_n$ and represents the filtered estimate of X_n given $Z_0, \ldots, Z_n$. It is

$$\hat{X}_n = \left[H^T(n)R^{-1}(n)H(n) + P_{n|n-1}^{-1}\right]^{-1}$$

$$\left[H^T(n)R^{-1}(n)z_n + P_{n|n-1}^{-1}\hat{X}_{n|n-1}\right] \tag{3-35}$$

At this point it is appropriate to justify our definitions in Eq. 3-22 and 3-23. It is not difficult to see that $p(X_n | Z_0, \ldots, Z_n)$ is gaussian.* Since a gaussian p.d.f. is symmetric and unimodal, its expected value coincides with its maximum value. Consequently, the maximum likelihood solution $\hat{X}_n$ is equal to $E\{X_n | Z_0, \ldots, Z_n\}$, i.e.,

$$\hat{X}_n = E\left\{X_n | Z_0, \ldots, Z_n\right\} \tag{3-36}$$

* In fact, Eq. 3-30 can be put in the form

$$p(X_n | Z_0, \ldots, Z_n) = \frac{\exp\{-1/2[X_n - \hat{X}_n]^T P_n^{-1} [X_n - \hat{X}_n]\}}{\sqrt{(2\pi)^m \det P_n}}$$

where

$$P_n \triangleq E\{[X_n - \hat{X}_n][X_n - \hat{X}_n]^T | Z_0, \ldots, Z_n\}$$

Similarly, the maximum likelihood estimate for the one-step prediction $\hat{X}_{n|n-1}$ is equal to the expected value of X_n conditioned on $Z_0,\ldots,Z_{n-1}$. This is true because $p(X_n|Z_0,\ldots,Z_{n-1})$ is also gaussian. Thus,

$$\hat{X}_{n|n-1} = E\left\{X_n \middle| Z_0,\ldots,Z_{n-1}\right\} \tag{3-37}$$

which corresponds to Eq. 3-22.

From Eq. 3-1, X_n is generated by

$$X_n = \Phi(n,n-1)X_{n-1} + W_{n-1} \tag{3-38}$$

Thus, Eq. 3-37 can be written

$$\hat{X}_{n|n-1} = E\left\{\Phi(n,n-1)X_{n-1} + W_{n-1} \middle| Z_0,\ldots,Z_{n-1}\right\} \tag{3-39}$$

$$= \Phi(n,n-1)E\left\{X_{n-1} \middle| Z_0,\ldots,Z_{n-1}\right\} \tag{3-40}$$

To obtain Eq. 3-40, the independent zero-mean characteristics of W_{n-1} were exploited. Using the general result of Eq. 3-36, it follows that

$$\hat{X}_{n-1} = E\left\{X_{n-1} \middle| Z_0,\ldots,Z_{n-1}\right\}$$

Therefore, Eq. 3-40 reduces to

$$\hat{X}_{n|n-1} = \Phi(n,n-1)\hat{X}_{n-1} \tag{3-41}$$

This is the desired form for the one-step Kalman predictor. It could have been derived directly from the maximum likelihood equation in a manner analogous to the method used to arrive at the filtered estimate. However, the likelihood approach requires more involved algebraic manipulations.

The definition for the prediction error covariance Eq. 3-23 is a natural consequence of Eq. 3-37. This is easily seen since the prediction error is

$$e_{n|n-1} = X_n - \hat{X}_{n|n-1} \tag{3-42}$$

and its covariance is given by $P_{n|n-1}$. Substituting Eq. 3-41 into the definition Eq. 3-23 for $P_{n|n-1}$ yields

$$P_{n|n-1} = E\left\{\left[X_n - \hat{X}_{n|n-1}\right]\left[X_n - \hat{X}_{n|n-1}\right]^T \middle| Z_0,\ldots,Z_{n-1}\right\}$$

$$= E\left\{\left[X_n - \Phi(n,n-1)\hat{X}_{n-1}\right]\left[X_n - \Phi(n,n-1)\hat{X}_{n-1}\right]^T \middle| Z_0,\ldots,Z_{n-1}\right\} \tag{3-43}$$

Substituting Eq. 3-38 into the above expression gives

$$P_{n|n-1} = \Phi(n,n-1)E\left\{\left[X_{n-1} - \hat{X}_{n-1}\right]\left[X_{n-1} - \hat{X}_{n-1}\right]^T \middle| Z_0,\ldots,Z_{n-1}\right\}\Phi^T(n,n-1)$$

$$+ E\left\{W_{n-1}W_{n-1}^T\right\} \tag{3-44}$$

Since the error in the filtered estimate (at time n-1) is

$$e_{n-1} = X_{n-1} - \hat{X}_{n-1} \tag{3-45}$$

the first expectation in Eq. 3-44 is just the covariance of the error in the filtered estimate. That is,

$$P_{n-1} \triangleq E\left\{e_{n-1}e_{n-1}^T \middle| Z_0,\ldots,Z_{n-1}\right\} \tag{3-46}$$

The second expectation in Eq. 3-44 was defined in Eq. 3-3 to be the covariance of the system perturbation noise. Therefore, Eq. 3-44 reduces to the attractive sequential relationship.

$$P_{n|n-1} = \Phi(n,n-1)P_{n-1}\Phi^T(n,n-1) + Q(n-1) \tag{3-47}$$

The inverse of $P_{n|n-1}$ is required in the estimation of $\hat{X}_n$ using Eq. 3-35. In fact, as it now stands, Eq. 3-35 involves taking the inverse of two m x m matrices and a p x p matrix at every measurement update. This represents a prohibitive computational burden. What is needed is a recursive version of Eq. 3-35 which eliminates the matrix inversions. This can be accomplished in the following manner. First define

$$P_n \triangleq \left[H^T(n)R^{-1}(n)H(n) + P_{n|n-1}^{-1}\right]^{-1} \tag{3-48)*}$$

A straightforward application of the inside-out lemma (see Theorem A.2 of the appendix) allows P_n to be rewritten

$$P_n = P_{n|n-1} - P_{n|n-1}H^T(n)\left[R(n) + H(n)P_{n|n-1}H^T(n)\right]^{-1}H(n)P_{n|n-1} \tag{3-49}$$

---

* The P_n defined here is the explicit form of the error covariance of the filtered estimate defined in the footnote on page 28. This equivalency will be established subsequently. It is introduced at this point for notational convenience only.

This form for P_n requires only one p x p matrix inversion instead of the two m x m and one p x p matrix inversions required in Eq. 3-48. Thus, the computational load is significantly reduced, especially since p is less than or equal to m. Equation 3-49 is the sought-after recursive relation for updating the error covariance matrix.

Using the definition of Eq. 3-48, the filtered estimate in Eq. 3-35 reduces to

$$\hat{X}_n = P_n\left[H^T(n)R^{-1}(n)Z_n + P^{-1}_{n|n-1}\hat{X}_{n|n-1}\right] \tag{3-50}$$

It can be simplified further by adding and subtracting $P_nH^T(n)R^{-1}(n)H(n)\hat{X}_{n|n-1}$ from the right side of Eq. 3-50

$$\begin{aligned}\hat{X}_n &= P_n\Big[H^T(n)R^{-1}(n)Z_n + P^{-1}_{n|n-1}\hat{X}_{n|n-1} + H^T(n)R^{-1}(n)H(n)\hat{X}_{n|n-1} \\ &\quad - H^T(n)R^{-1}(n)H(n)\hat{X}_{n|n-1}\Big]\end{aligned} \tag{3-51}$$

$$\begin{aligned}&= P_n\left[P_{n|n-1} + H^T(n)R^{-1}(n)H(n)\right]\hat{X}_{n|n-1} \\ &\quad + P_nH^T(n)R^{-1}(n)\left[Z_n - H(n)\hat{X}_{n|n-1}\right]\end{aligned} \tag{3-52}$$

In the last equation the coefficient of the first term is the identity matrix. This follows from Eq. 3-48. Therefore, the expression for $\hat{X}_n$ becomes

$$\hat{X}_n = \hat{X}_{n|n-1} + P_nH^T(n)R^{-1}(n)\left[Z_n - H(n)\hat{X}_{n|n-1}\right] \tag{3-53}$$

Defining the gain matrix K_n as

$$K_n \triangleq P_nH^T(n)R^{-1}(n) \tag{3-54}$$

reduces Eq. 3-53 to

$$\hat{X}_n = \hat{X}_{n|n-1} + K_n\left[Z_n - H(n)\hat{X}_{n|n-1}\right] \tag{3-55}$$

This is the desired recursive filter algorithm. Note that all matrix inversions, except the one required in computing K_n, have been eliminated. The matrix K_n is known as the Kalman gain. As will be shown below, expression 3-54 for K_n is equivalent to Eq. 2-28 of Section II. Substituting Eq. 3-49 into Eq. 3-54 gives

$$K_n = P_{n|n-1}H^T(n)R^{-1}(n) - P_{n|n-1}H^T(n)\left[R(n) + H(n)P_{n|n-1}H^T(n)\right]^{-1}$$

$$\cdot\ H(n)P_{n|n-1}H^T(n)R^{-1}(n)$$

$$= P_{n|n-1}H^T(n)\Big\{R^{-1}(n) - \left[R(n) + H(n)P_{n|n-1}H^T(n)\right]^{-1}$$

$$\cdot\ H(n)P_{n|n-1}H^T(n)R^{-1}(n)\Big\}$$

$$= P_{n|n-1}H^T(n)\left[R(n) + H(n)P_{n|n-1}H^T(n)\right]^{-1}$$

$$\left\{\left[R(n) + H(n)P_{n|n-1}H^T(n)\right]R^{-1}(n) - H(n)P_{n|n-1}H^T(n)R^{-1}(n)\right\}$$

$$= P_{n|n-1}H^T(n)\left[R(n) + H(n)P_{n|n-1}H^T(n)\right]^{-1}$$

$$\left\{I + H(n)P_{n|n-1}H^T(n) - H(n)P_{n|n-1}H^T(n)R^{-1}(n)\right\}$$

$$= P_{n|n-1}H^T(n)\left[R(n) + H(n)P_{n|n-1}H^T(n)\right]^{-1} \tag{3-56}$$

which is identical to Eq. 2-28.

To complete the Kalman filter algorithm, there remains only to show that the P_n defined in Eq. 3-48, or its equivalent Eq. 3-49, is indeed the filtering error covariance matrix. By definition the error covariance is

$$E\left\{\left[X_n - \hat{X}_n\right]\left[X_n - \hat{X}_n\right]^T \middle| Z_0,\ldots,Z_n\right\}$$

$$= E\left\{\left[X_n - \hat{X}_n\right]X_n^T \middle| Z_0,\ldots,Z_n\right\} - E\left\{\left[X_n - \hat{X}_n\right]\hat{X}_n^T \middle| Z_0,\ldots,Z_n\right\}$$

$$= E\left\{\left[X_n - \hat{X}_n\right]X_n^T \middle| Z_0,\ldots,Z_n\right\} \tag{3-57}$$

Equation 3-57 results from the fundamental property that the error in an optimal estimate is orthogonal to the estimate, i.e.,

$$E\left\{\left[X_n - \hat{X}_n\right]\hat{X}_n^T \middle| Z_0,\ldots,Z_n\right\} = 0 \tag{3-58}$$

This result can be verified by direct calculation of $E\{[X_n - \hat{X}_n]\hat{X}_n^T|Z_0,\ldots,Z_n\}$ using the expression of Eq. 3-53 for the optimal estimate. Similarly, it follows that

$$E\left\{\left[X_n - \hat{X}_{n|n-1}\right]\hat{X}_{n|n-1}^T \middle| Z_0,\ldots,Z_n\right\} = 0 \tag{3-59}$$

Adding and subtracting $\hat{X}_{n|n-1}$ in Eq. 3-57 results in*

$$\begin{aligned}
&E\left\{\left[X_n - \hat{X}_n\right]\left[X_n - \hat{X}_{n|n-1} + \hat{X}_{n|n-1}\right]^T\right\} \\
&\quad = E\left\{\left[X_n - \hat{X}_{n|n-1} - K_n\left(Z_n - H(n)\hat{X}_{n|n-1}\right)\right]\left[X_n - \hat{X}_{n|n-1} + \hat{X}_{n|n-1}\right]^T\right\} \\
&\quad = E\left\{\left[X_n - \hat{X}_{n|n-1}\right]\left[X_n - \hat{X}_{n|n-1}\right]^T\right\} + E\left\{\left[X_n - \hat{X}_{n|n-1}\right]\hat{X}_{n|n-1}^T\right\} \\
&\quad - K_n H(n) E\left\{\left[X_n - \hat{X}_{n|n-1}\right]\left[X_n - \hat{X}_{n|n-1}\right]^T\right\} - K_n E\left\{V_n\left[X_n - \hat{X}_{n|n-1}\right]^T\right\} \\
&\quad - K_n H(n) E\left\{\left[X_n - \hat{X}_{n|n-1}\right]\hat{X}_{n|n-1}^T\right\} - K_n E\left\{V_n \hat{X}_{n|n-1}^T\right\}
\end{aligned} \tag{3-60}$$

From Eq. 3-23, the first and third terms are, respectively, $P_{n|n-1}$ and $-K_nH(n)P_{n|n-1}$. The second and fifth terms are zero due to the orthogonality condition Eq. 3-59. The fourth and sixth terms are zero since V_n is a zero-mean independent random sequence. Therefore, the error covariance for the filter estimate reduces to

$$E\left\{\left[X_n - \hat{X}_n\right]\left[X_n - \hat{X}_n\right]^T \middle| Z_0,\ldots,Z_n\right\} = \left[I - K_nH(n)\right]P_{n|n-1} \tag{3-61}$$

By substituting Eq. 3-56 into Eq. 3-61 we obtain

$$\begin{aligned}
&E\left\{\left[X_n - \hat{X}_n\right]\left[X_n - \hat{X}_n\right]^T \middle| Z_0,\ldots,Z_n\right\} \\
&\quad = P_{n|n-1} - P_{n|n-1}H^T(n)\left[R(n) + H(n)P_{n|n-1}H^T(n)\right]^{-1}H(n)P_{n|n-1} \\
&\quad \triangleq P_n
\end{aligned} \tag{3-62}$$

* Henceforth, the sequence notation $Z_0,\ldots,Z_n$ on which the expectation is conditioned will be omitted when no confusion will result.

Thus, we have demonstrated that P_n as defined in Eq. 3-48 or, equivalently, Eq. 3-49 is the error covariance for the filtered estimate. We could have more easily shown this by using the method of Section II. Our purpose in employing the present approach was to expose the reader to the principle of orthogonality, Eq. 3-58. It is in fact a necessary and sufficient condition for optimal filtering and was the starting point for Kalman in his original Hilbert space derivation of the filter algorithm. Actually, Eq. 3-58 is a discrete time version of the Wiener-Hopf equation which is fundamental to many problems in mathematical physics.

We have completed the maximum likelihood derivation of the Kalman filter. The results are summarized below.

Predicted Estimate
$$\hat{X}_{n|n-1} = \Phi(n,n-1)\hat{X}_{n-1} \tag{3-41}$$

Covariance for Predicted Estimate
$$P_{n|n-1} = \Phi(n,n-1)P_{n-1}\Phi^T(n,n-1) + Q(n-1) \tag{3-47}$$

Filter Gain
$$K_n = P_{n|n-1}H^T(n)\left[H(n)P_{n|n-1}H^T(n) + R(n)\right]^{-1} \tag{3-56}$$

$$= P_n H^T(n)R^{-1}(n) \tag{3-54}$$

Filtered Estimate
$$\hat{X}_n = \hat{X}_{n|n-1} + K_n\left[Z_n - H(n)\hat{X}_{n|n-1}\right] \tag{3-55}$$

Covariance for Filter Error
$$P_n = \left[I - K_n H(n)\right]P_{n|n-1} \tag{3-61}$$

$$= P_{n|n-1} - P_{n|n-1}H^T(n)\left[R(n) + H(n)P_{n|n-1}H^T(n)\right]^{-1} \cdot H(n)P_{n|n-1} \tag{3-62}$$

Initial Conditions
$$E\left\{\hat{X}_{0|-1}\right\} = E\left\{X_0\right\}$$

$$P_{0|-1} = E\left\{\left[X_0 - \hat{X}_{0|-1}\right]\left[X_0 - \hat{X}_{0|-1}\right]^T\right\}$$

It is worth emphasizing that incorrect initial conditions have no effect on the filter after the initial transient period.

3.4 THE KALMAN FILTER AS A MAXIMUM LIKELIHOOD ESTIMATOR

Because we have derived the Kalman filter via the method of maximum likelihood, it has all the attendant attributes of maximum likelihood estimators. Some of the more important aspects are mentioned below.

The covariance of the error in any estimate can be bounded from below by the Rao-Cramer Inequality. An estimate which achieves this lower bound is called an efficient estimate. It is not difficult to show that if such an estimate exists it is a maximum likelihood estimate. For gaussian processes the maximum likelihood estimate and, therefore, the Kalman filter are efficient estimates.

We can relate maximum likelihood estimation to least squares estimation by observing that Eq. 3-32 has the form of a loss function corresponding to a least squares estimator. This implies that the Kalman filter can be derived from a least squares point of view. To accomplish this, the weighting matrices in the loss function must be selected such that the least squares estimate has minimum variance. One such development will be presented in the next section.

IV. MINIMUM VARIANCE STATE ESTIMATION FOR LINEAR DYNAMIC SYSTEMS

4.1 MINIMUM VARIANCE ESTIMATION

The principles and most important result of minimum variance estimation are embodied in the fundamental Gauss-Markov Theorem.

<u>Theorem 4.1. Gauss-Markov Theorem</u>

Let

$$Y = MX + \varepsilon \qquad t > m \tag{4-1}$$

be the measurement model, where

Y is a $t \times 1$ vector of observations,

M is a $t \times m$ mapping matrix of rank m,

X is a $m \times 1$ unknown vector to be estimated,

ε is a $t \times 1$ zero-mean random error vector with covariance

$$E\left\{\varepsilon\varepsilon^T\right\} = C \tag{4-2}$$

Then the minimum variance linear unbiased estimate (MVLUE) of X is

$$\hat{X} = \left(M^TC^{-1}M\right)^{-1}M^TC^{-1}Y \tag{4-3}$$

and its error covariance is

$$P \triangleq E\left\{(\hat{X} - X)(\hat{X} - X)^T|Y\right\} \triangleq \text{Cov } [\hat{X} - X] = \left(M^TC^{-1}M\right)^{-1} \tag{4-4}$$

From the definition of expectation,

$$\begin{aligned}
E\left\{\hat{X}\right\} &= E\left\{\left(M^TC^{-1}M\right)^{-1}M^TC^{-1}Y\right\} \\
&= \left(M^TC^{-1}M\right)^{-1}M^TC^{-1}E\left\{Y\right\} \\
&= \left(M^TC^{-1}M\right)^{-1}M^TC^{-1}E\left\{MX + E\right\} \\
&= \left(M^TC^{-1}M\right)^{-1}\left(M^TC^{-1}M\right)E\left\{X\right\} \\
&= E\left\{X\right\}
\end{aligned}$$

Hence $\hat{X}$ is unbiased.

Now let

$$X^* = AY \tag{4-5}$$

be any other unbiased estimator of X. Then

$$E\{X^*\} = E\{X\} = E\{\hat{X}\} \tag{4-6}$$

from which it follows that

$$AME\{X\} = E\{X\} = G\left(M^T C^{-1} M\right) E\{X\} \tag{4-7}$$

where

$$G \triangleq \left(M^T C^{-1} M\right)^{-1} \tag{4-8}$$

Since Eq. 4-7 is an identity in $E\{X\}$, we have

$$(AM - I) = 0 \tag{4-9}$$

$$\left(AM - GM^T C^{-1} M\right) = 0 \tag{4-10}$$

and

$$\left(GM^T C^{-1} M - I\right) = 0 \tag{4-11}$$

Now

$$\begin{aligned} \mathrm{Cov}\left[X^* - X\right] &= \mathrm{Cov}\left[AY - X\right] \\ &= \mathrm{Cov}\left[A(MX + \varepsilon) - X\right] \\ &= \mathrm{Cov}\left[(AM - I)X + A\varepsilon\right] \end{aligned} \tag{4-12}$$

Using Eq. 4-9, Eq. 4-12 becomes

$$\begin{aligned} \mathrm{Cov}\left[X^* - X\right] &= \mathrm{Cov}\left[A\varepsilon\right] \\ &= E\left\{A\varepsilon(A\varepsilon)^T\right\} \\ &= AE\left\{\varepsilon\varepsilon^T\right\}A^T \\ &= ACA^T \end{aligned} \tag{4-13}$$

Similarly,

$$\begin{aligned}
\text{Cov}\,[\hat{X} - X] &= \text{Cov}\left|GM^TC^{-1}Y - X\right| \\
&= \text{Cov}\left|GM^TC^{-1}(MX + \varepsilon) - X\right| \\
&= \text{Cov}\left|\left(GM^TC^{-1}M - I\right)X + GM^TC^{-1}\varepsilon\right|
\end{aligned} \tag{4-14}$$

From Eq. 4-11, Eq. 4-14 reduces to

$$\begin{aligned}
\text{Cov}\,[\hat{X} - X] &= \text{Cov}\left[GM^TC^{-1}\varepsilon\right] \\
&= E\left\{GM^TC^{-1}\varepsilon\left(GM^TC^{-1}\varepsilon\right)^T\right\} \\
&= GM^TC^{-1}E\left\{\varepsilon\varepsilon^T\right\}C^{-1}MG^T \\
&= GM^TC^{-1}CC^{-1}MG^T \\
&= GM^TC^{-1}MG^T \qquad (4\text{-}15) \\
&= G \qquad (4\text{-}16)
\end{aligned}$$

which establishes the claim of Eq. 4-4. The error covariance of X^* can now be related to the error covariance of $\hat{X}$. Adding and subtracting GM^TC^{-1} in Eq. 4-13 yields

$$\begin{aligned}
\text{Cov}\,[X^* - X] &= \left[A - GM^TC^{-1} + GM^TC^{-1}\right]C\left[A - GM^TC^{-1} + GM^TC^{-1}\right]^T \\
&= \left[A - GM^TC^{-1}\right]C\left[A - GM^TC^{-1}\right]^T \\
&\quad + 2\left[AM - GM^TC^{-1}M\right]G^T + GM^TC^{-1}MG^T
\end{aligned} \tag{4-17}$$

The first term in Eq. 4-17 is a non-negative quadratic form, the second term is zero from Eq. 4-10, and from Eq. 4-15 the last term is equal to $\text{Cov}[\hat{X} - X]$. Thus, $\text{Cov}[X^* - X]$ = non-negative quadratic form + $\text{Cov}[\hat{X} - X]$. Since the diagonal elements of a non-negative quadratic form are non-negative,

$$\text{Cov}\left|x_j^* - x_j\right| \geq \text{Cov}\left|\hat{x}_j - x_j\right| \qquad j = 1,\ldots,m$$

where x_j denotes the jth element of X.

Therefore, we have shown that $\hat{X}$ is the MVLUE of X whose error covariance is given by Eq. 4-4. Uniqueness follows from Eq. 4-17, for if Cov $[X^* - X]$ = Cov $[\hat{X} - X]$, it is necessary that

$$A = GM^TC^{-1}$$

which implies $X^* = \hat{X}$. Q.E.D.

4.2 A MINIMUM VARIANCE APPROACH TO KALMAN FILTERING

Consider the dynamic model

$$X_{n+1} = \Phi(n+1,n)X_n + W_n \tag{4-18}$$

and the measurements

$$Z_n = H(n)X_n + V_n \tag{4-19}$$

where

X_n is the m x 1 state vector

W_n is the m x 1 dynamic perturbation noise

$\Phi(n+1,n)$ is the m x m state transition matrix

Z_n is the p x 1 measurement vector

$H(n)$ is the p x m observation matrix

V_n is the p x 1 measurement noise vector.

Both V_n and W_n are assumed to be zero-mean random noise sequences with covariances

$$E\left\{W_jW_k^T\right\} = Q(k)\delta_{jk}$$

$$E\left\{V_jV_k^T\right\} = R(k)\delta_{jk} \tag{4-20}$$

$$E\left\{V_jW_k^T\right\} = 0 \quad \text{for all } j,k$$

Now assume that the MVLUE $\hat{X}_{n|n-1}$ of X_n given the measurements $Z_0,\ldots,Z_{n-1}$ has already been computed. Then the random error vector is

$$e_{n|n-1} = \hat{X}_{n|n-1} - X_n \tag{4-21}$$

with covariance

$$E\left\{e_{n|n-1}e^T_{k|k-1}\right\} = \begin{cases} P_{n|n-1} & n = k \\ 0 & n \neq k \end{cases} \tag{4-22}$$

Since $\hat{X}_{n|n-1}$ is unbiased, $e_{n|n-1}$ is a zero-mean sequence. Upon receiving the measurement Z_n, we want to obtain the MVLUE $\hat{X}_n$ of X_n.* But

$$Z_n = H(n)X_n + V_n$$

is a p x 1 vector, where

$$p = t < m$$

Therefore, Theorem 4.1 is not directly applicable.

However, by forming a model using all the known information, Theorem 4.1 can be invoked. This is accomplished as follows. Rewrite Eq. 4-21 and combine it with Eq. 4-19, giving

$$\begin{bmatrix} Z_n \\ \hat{X}_{n|n-1} \end{bmatrix} = \begin{bmatrix} H(n) \\ I \end{bmatrix} X_n + \begin{bmatrix} V_n \\ e_{n|n-1} \end{bmatrix} \tag{4-23}$$

By making the correspondences

$$Y = \begin{bmatrix} Z_n \\ \hat{X}_{n|n-1} \end{bmatrix} \quad M = \begin{bmatrix} H(n) \\ I \end{bmatrix} \quad \text{and } \varepsilon = \begin{bmatrix} V_n \\ e_{n|n-1} \end{bmatrix} \tag{4-24}$$

we have

$$Y = MX_n + \varepsilon \tag{4-25}$$

where

Y is a $(p + m) \times 1$ vector

M is a $(p + m) \times m$ matrix

X_n is a $m \times 1$ unknown vector to be estimated

ε is a $(p + m) \times 1$ zero-mean random error vector with covariance

* Note that $\hat{X}_{n|n-1}$ is the one-step prediction and $\hat{X}_n$ is the filtered value of X_n.

$$E\left\{\varepsilon\varepsilon^T\right\} = C = \begin{bmatrix} R(n) & 0 \\ 0 & P_{n|n-1} \end{bmatrix} \qquad (4-26)$$

All the conditions required to use Theorem 4.1 are now satisfied. Thus, the desired estimate $\hat{X}_n$ of X_n can be written down immediately,

$$\hat{X}_n = \left(M^T C^{-1} M\right)^{-1} M^T C^{-1} Y \qquad (4-27)$$

Expanding Eq. 4-27 by using the correspondences of Eq. 4-24 and 4-26 yields

$$\hat{X}_n = \left(\begin{bmatrix} H(n) \\ I \end{bmatrix}^T \begin{bmatrix} R(n) & 0 \\ 0 & P_{n|n-1} \end{bmatrix}^{-1} \begin{bmatrix} H(n) \\ I \end{bmatrix}\right)^{-1} \begin{bmatrix} H(n) \\ I \end{bmatrix}^T \begin{bmatrix} R(n) & 0 \\ 0 & P_{n|n-1} \end{bmatrix}^{-1} \begin{bmatrix} Z_n \\ \hat{X}_{n|n-1} \end{bmatrix}$$

$$= \left(\begin{bmatrix} H^T(n) & I \end{bmatrix} \begin{bmatrix} R^{-1}(n) & 0 \\ 0 & P^{-1}_{n|n-1} \end{bmatrix} \begin{bmatrix} H(n) \\ I \end{bmatrix}\right)^{-1} \begin{bmatrix} H^T(n) & I \end{bmatrix} \begin{bmatrix} R^{-1}(n) & 0 \\ 0 & P^{-1}_{n|n-1} \end{bmatrix} \begin{bmatrix} Z_n \\ \hat{X}_{n|n-1} \end{bmatrix}$$

$$= \left[H^T(n) R^{-1}(n) H(n) + P^{-1}_{n|n-1}\right]^{-1} \left[H^T(n) R^{-1}(n) Z_n + P^{-1}_{n|n-1} \hat{X}_{n|n-1}\right] \qquad (4-28)$$

The reader should observe that Eq. 4-28 is identical to Eq. 3-3?, and it can be converted into a more efficient computational form in exactly the same way. From Eq. 4-4 we know that the covariance matrix for the estimation error at time n is

$$P_n = \left(M^T C^{-1} M\right)^{-1} \qquad (4-29)$$

but the first term in Eq. 4-28 is just the expanded form of $(M^T C^{-1} M)^{-1}$. This follows from an inspection of Eq. 4-27. Therefore,

$$P_n = \left[H^T(n) R^{-1}(n) H(n) + P^{-1}_{n|n-1}\right]^{-1} \qquad (4-30)$$

Applying the inside-out lemma (see Theorem A.2 of the appendix) to Eq. 4-30 yields a less sensitive numerical expression for P_n.

$$P_n = P_{n|n-1} - P_{n|n-1} H^T(n) \left[R(n) + H(n) P_{n|n-1} H^T(n)\right]^{-1} H(n) P_{n|n-1} \qquad (4-31)$$

Substituting Eq. 4-30 into Eq. 4-28 gives

$$\hat{X}_n = P_n\left[H^T(n)R^{-1}(n)Z_n + P^{-1}_{n|n-1}\hat{X}_{n|n-1}\right] \tag{4-32}$$

Proceeding as in Section III (Eq. 3-51), add and subtract $P_nH^T(n)R^{-1}(n)H(n)\hat{X}_{n|n-1}$ from the right side of Eq. 4-32. This results in

$$\hat{X}_n = \hat{X}_{n|n-1} + P_nH^T(n)R^{-1}(n)\left[Z_n - H(n)\hat{X}_{n|n-1}\right] \tag{4-33}$$

Defining the gain term as

$$K_n = P_nH^T(n)R^{-1}(n) \tag{4-34}$$

provides the Kalman filter expression

$$\hat{X}_n = \hat{X}_{n|n-1} + K_n\left[Z_n - H(n)\hat{X}_{n|n-1}\right] \tag{4-35}$$

To complete the derivation there remains only the determination of $\hat{X}_{n|n-1}$ and its covariance $P_{n|n-1}$. From Theorem A.3 of the appendix, we know that the minimum variance unbiased estimate of X_n given the measurement set $\{Z_0,\ldots,Z_{n-1}\} \triangleq Z(n-1)$ is

$$\hat{X}_{n|n-1} = E\left\{X_n | Z(n-1)\right\} \tag{4-36}$$

$$= E\left\{\Phi(n,n-1)X_{n-1} + W_{n-1} | Z(n-1)\right\} \tag{4-37}$$

Since W_{n-1} is a zero-mean sequence that is independent of Z(n-1), Eq. 4-37 becomes

$$\hat{X}_{n|n-1} = \Phi(n,n-1)E\left\{X_{n-1} | Z(n-1)\right\} \tag{4-38}$$

Again from Theorem A.3, the MVLUE $\hat{X}_{n-1}$ of X_{n-1} is given by $E\{X_{n-1}|Z(n-1)\}$. Therefore,

$$\hat{X}_{n|n-1} = \Phi(n,n-1)\hat{X}_{n-1} \tag{4-39}$$

By definition the error covariance matrix is

$$P_{n|n-1} = E\left\{\left[X_n - \hat{X}_{n|n-1}\right]\left[X_n - \hat{X}_{n|n-1}\right]^T \Big| Z(n-1)\right\} \tag{4-40}$$

Substituting Eq. 4-18 and 4-39 into Eq. 4-40 gives

$$P_{n|n-1} = \Phi(n,n-1)E\left\{\left[X_{n-1} - \hat{X}_{n-1}\right]\left[X_{n-1} - \hat{X}_{n-1}\right]^T \middle| Z(n-1)\right\}\Phi^T(n,n-1)$$
$$+ E\left\{W_{n-1}W^T_{n-1}\right\} \tag{4-41}$$

The first expectation in Eq. 4-41 is just the covariance of the filtered estimate at time n-1, and the second expectation is defined in Eq. 4-20. Thus, the error covariance of the predicted estimate is

$$P_{n|n-1} = \Phi(n,n-1)P_{n-1}\Phi^T(n,n-1) + Q(n-1) \tag{4-42}$$

Equations 4-31, 4-34, 4-35, 4-39, and 4-42 comprise the Kalman filter. They are summarized below.

Predicted Estimate
$$\hat{X}_{n|n-1} = \Phi(n,n-1)\hat{X}_{n-1} \tag{4-39}$$

Error Covariance for Predicted Estimate
$$P_{n|n-1} = \Phi(n,n-1)P_{n-1}\Phi^T(n,n-1) + Q(n-1) \tag{4-42}$$

Filter Gain
$$K_n = P_nH^T(n)R^{-1}(n) \tag{4-34}$$
$$= P_{n|n-1}H^T(n)\left[H(n)P_{n|n-1}H^T(n) + R(n)\right]^{-1} \tag{4-43}$$

Filtered Estimate
$$\hat{X}_n = \hat{X}_{n|n-1} + K_n\left[Z_n - H(n)\hat{X}_{n|n-1}\right] \tag{4-35}$$

Error Covariance for Filter Estimate
$$P_n = \left[I - K_nH(n)\right]P_{n|n-1} \tag{4-44}$$
$$= P_{n|n-1} - P_{n|n-1}H^T(n)\left[R(n) + H(n)P_{n|n-1}H^T(n)\right]^{-1}$$
$$\cdot\, H(n)P_{n|n-1} \tag{4-31}$$

Initial Conditions
$$E\left\{\hat{X}_{0|-1}\right\} = E\left\{X_0\right\}$$
$$P_{0|-1} = E\left\{\left[X_0 - \hat{X}_{0|-1}\right]\left[X_0 - \hat{X}_{0|-1}\right]^T\right\}$$

Equations 4-43 and 4-44 were not explicitly developed in this section. They were, however, derived in Section III using straightforward matrix manipulations (see Eq. 3-56 and 3-61).

4.3 ADDITIONAL ASPECTS OF MINIMUM VARIANCE ESTIMATION

The fact that the Kalman filter can be derived as a minimum variance estimator implies that it could also be developed from the deterministic least squares approach. This follows from the least squares criterion: given the measurement model

$$Y = MX + \varepsilon \tag{4-1}$$

the least squares estimate $\hat{X}$ of X is selected such that

$$\ell = (Y - MX)^T W(Y - MX) \tag{4-45}$$

is a minimum. The result is

$$\hat{X} = \left(M^T WM\right)^{-1} M^T WY \tag{4-46}$$

By choosing the weighting matrix W to be the inverse of the error covariance matrix C, Eq. 4-46 becomes

$$\hat{X} = \left(M^T C^{-1} M\right)^{-1} M^T C^{-1} Y \tag{4-47}$$

This equation is identical to the minimum variance estimate of Eq. 4-3. Thus, the Kalman filter may be interpreted as a least squares estimate.

At this point it is worth noting that the Gauss-Markov Theorem is the discrete time analog of the Wiener-Hopf integral equation. Thus, it can also be derived from the orthogonality condition of Eq. 3-58. Furthermore, minimum variance estimation requires knowledge of only the first two moments (mean and covariance) of a probability density function (p.d.f.). The exact form of the p.d.f. is irrelevant. This is in marked contrast to the method of maximum likelihood where complete knowledge of the p.d.f. of X_n conditioned on all the measurements is necessary.

V. COMMENTS ON THE PRACTICAL APPLICATION OF KALMAN FILTERING

Kalman filter theory is defined in precise mathematical terms, but its application to practical problems is not a precise science. Considerable engineering experience is needed to properly identify the system to which the filter is to be applied, to adequately model that system, and then to generate a practical machine language software program for the on-board computer. In addition, most dynamic models are nonlinear and the Kalman filter assumes linear dynamics. Thus, a form of linearization that adequately controls the truncation error has to be used. The implementation of the filter must also include other factors which are difficult or impossible to characterize analytically, e.g., the trade-off between filter performance, computer program size, and execution time.

Another source of difficulty is specifying the statistical parameters that represent sensor noise and dynamic modeling errors. In general these parameters are complicated and not well-known. Therefore, they usually must be selected using engineering judgment coupled with test data. As a case in point, the covariance matrix of an operational Kalman filter in a navigation system contains an estimate of the root mean-square position accuracy. The performance is usually optimized if this estimate closely matches the actual root mean-square position accuracy obtained by the system. To achieve this match, the position noise parameters and/or dynamics must be somewhat arbitrarily adjusted.

Another characteristic of airborne systems is a rapidly varying environment. This condition often necessitates the incorporation of adaptive features in the filter structure. The somewhat ad hoc methods required to satisfactorily solve this problem further enhance the value of engineering intuition and judgment needed in real-world filter design.

Some of the additional features important to any practical implementation of the Kalman filter are listed below:

1. If the measurement is a scalar, no matrix inversion is required in Eq. 2-28. Therefore, when more than one measurement is available, they should generally be processed one at a time.

2. The matrix equations of Table 2-1 can be written out in scalar form. This will often permit a large reduction in computational complexity as a result of the decoupling effect of terms that are zero.

3. Theoretically, the Kalman filter is stable or "robust" in the sense that the effects of initial errors, round-off and other computational errors die out asymptotically. This is true only so long as the model is an adequate representation of reality, including the numerical implementation. Otherwise, the error covariance matrix can become ill-conditioned due to the finite nature of the computer and/or modeling errors. This seriously degrades the filter performance.

Extensive consideration of these and other problems associated with the mechanization of the Kalman filter are available in the literature, e.g., see Ref. 6-8.

VI. EPILOGUE: A SUMMARY AND PERSPECTIVE

6.1 SUMMARY

In this report the discrete time version of the Kalman filter and some typical applications have been considered. Section I introduced the underlying estimation problem to be solved and the basic format needed in its solution via Kalman filtering. In Section II a heuristic development of the Kalman filter was presented. It hopefully displayed the essential features of the Kalman algorithm in a transparent manner. To illustrate the numerical behavior of the Kalman filter, a specific problem was solved. This example was then used to relate the Kalman algorithm to familiar frequency domain and transfer function concepts.

Next the method of maximum likelihood was invoked to provide a more precise derivation. Minimum variance estimation was also used to derive the Kalman filtering equations in Section IV. The similarities and differences of the two methods were discussed and related to the Wiener-Hopf equation which plays a fundamental role in mathematical physics.

In Section V some aspects of the difficulties that arise in applying the Kalman filter to engineering problems were considered briefly. As a side effect, these comments should reinforce the fact that the transition from theory to practice is not a trivial task.

6.2 PERSPECTIVE

From the derivations and discussions of the previous sections, it is clear that the Kalman filter is simply an algorithm for the time-domain design of linear filters that are optimal in the minimum mean-square error sense. Nevertheless, the significance of this achievement should not be underestimated. It represents a complete and elegant solution to the nonstationary multidimensional Wiener-Hopf equation. For this reason it has applications or interconnections to a large number of problems encountered in other fields ranging from econometrics to astrophysics. In fact, it was the astrophysicist Chandrasekhar (Ref. 9) who first solved the stationary multidimensional Wiener-Hopf equation.

Even though its derivation assumes linearity, the Kalman filter has found its greatest application to nonlinear systems. This extension is accomplished by linearizing the nonlinear equations about the most recent estimate (or an assumed nominal solution). In spite of the success of this approximation, it must be used with caution. The exact solution involves an infinite set of simultaneous equations, and there exists no general method for satisfactorily truncating this set. In spite of the difficulties, this generalization to nonlinear dynamics has attracted wide attention, since it arises in many fields such as quantum mechanics (Ref. 10), geophysics (Ref. 11), and a majority of engineering problems.

Appendix
FUNDAMENTAL THEOREMS

Theorem A.1: Let H be a p x m matrix defined on the real field, then H^TH is

(a) Positive definite if the rank of H is m and $m \leq p$,

(b) Positive semidefinite if the rank of H is p and $p < m$.

It is worth emphasizing that positive definite matrices are nonsingular, while positive semidefinite matrices are singular.

Theorem A.2: Inside-Out Matrix Inversion Lemma. The inverse of the m x m matrix S, where

$$S = P^{-1} + H^TR^{-1}H \tag{A-1}$$

is

$$S^{-1} = P - PH^TD^{-1}HP \tag{A-2}$$

where

$D = R + HPH^T$ is a p x p matrix, $p < m$

H is a p x m observation matrix of rank p

R is a p x p positive definite symmetric matrix, and

P is a m x m positive definite symmetric matrix.

Proof: The result follows from the fact that

$$\begin{aligned}
&\left[P^{-1} + H^TR^{-1}H\right]\left[P - PH^TD^{-1}HP\right] \\
&\quad = I - H^TD^{-1}HP + H^TR^{-1}HP - H^TR^{-1}HPH^TD^{-1}HP \\
&\quad = I + \left[-H^TD^{-1} + H^TR^{-1} - H^TR^{-1}HPH^TD^{-1}\right]HP \\
&\quad = I + H^TR^{-1}\left[-RD^{-1} + I - HPH^TD^{-1}\right]HP \\
&\quad = I + H^TR^{-1}\left[I - DD^{-1}\right]HP \\
&\quad = I
\end{aligned}$$

Q.E.D.

Theorem A.3: Suppose that a random variable X is to be estimated from the data $Z(k) = \{Z_1, Z_2, \ldots, Z_k\}$ such that the estimate $\hat{X}$ is unbiased and has minimum variance. That is,

$$E\{\hat{X}\} = E\{X\}$$

and

$$E\{(X - \hat{X})^T(X - \hat{X})\}$$

is a minimum. Then

$$\hat{X} = E\{X|Z(k)\} \tag{A-3}$$

Proof: Write the error variance in terms of the conditional expectation using the identity

$$E\{(X - \hat{X})^T(X - \hat{X})\} = E_Z\{E_X[(X - \hat{X})^T(X - \hat{X})|Z(k)]\}$$

Since the expectation with respect to Z(k) does not depend on $\hat{X}$, it is sufficient to minimize the conditional expectation. Expanding it gives

$$\begin{aligned} &E_X\{(X - \hat{X})^T(X - \hat{X})|Z(k)\} \\ &= \hat{X}^T\hat{X} - 2\hat{X}^T E\{X|Z(k)\} + E\{X^TX|Z(k)\} \\ &= \left[\hat{X} - E\{X|Z(k)\}\right]^T\left[\hat{X} - E\{X|Z(k)\}\right] + E\{X^TX|Z(k)\} \\ &\quad - \left[E\{X|Z(k)\}\right]^T E\{X|Z(k)\} \\ &\geq E\{X^TX|Z(k)\} - \left[E\{X|Z(k)\}\right]^T E\{X|Z(k)\} \end{aligned}$$

with equality if, and only if,

$$\hat{X} = E\{X|Z(k)\}$$

Eq. A-3 is unbiased, since

$$E\{\hat{X}\} = E_Z\{E_X[X|Z(k)]\} = E\{X\}$$

Q.E.D.

REFERENCES

1. Kalman, R. E. "A New Approach to Linear Filtering and Prediction Problems," *J. Basic Eng.*, Vol. 82D (1960), pp. 34-45.

2. Kalman, R. E., and R. S. Bucy. "New Results in Linear Filtering and Prediction Theory," *J. Basic Eng.*, Vol. 83D (1961), pp. 95-108.

3. Air Force Avionics Laboratory. *The Evaluation of Kalman Filter Designs for Multisensor Integrated Navigation Systems*, by J. A. D'Appolito, The Analytic Sciences Corporation. Wright-Patterson Air Force Base, Ohio, AFAL, 1971. (AFAL-TR-70-271, publication UNCLASSIFIED.)

4. Naval Weapons Center. *Comments on an ARBS-Laser Spot Tracker Filter Concept*, by Roy Dale Cole. China Lake, Calif., NWC, August 1972. (NWC TN 4070-34-72, publication UNCLASSIFIED.)

5. ------. *Stabilization and Filtering of a TV Tracking System for Airborne Angle-Angular Rate Sensing*, by R. D. Berry. China Lake, Calif., NWC, June 1971. (NWC TP 5215, publication UNCLASSIFIED.)

6. Sorenson, H. W. "Kalman Filtering Techniques," Chapter 5 in *Advances in Control Systems*, Vol. 3, C. T. Leondes, ed. New York, Academic Press, 1966.

7. Bucy, R. S., and P. D. Joseph. *Filtering for Stochastic Processes With Applications to Guidance*. New York, Interscience Publishers, 1968.

8. *IEEE Transactions on Automatic Control*, "Special Issue on the Linear-Quadratic-Gaussian Problem," December 1971.

9. Chandrasekhar, S. "On the Radiative Equilibrium of a Stellar Atmosphere, Part XXI," *Astrophys. J.* (1947), pp. 152-216.

10. Klauder, J. R., and E. G. Sudarshan. *Quantum Optics*. New York, Benjamin Press, 1969.

11. Burg, J. "Maximum Entropy Spectral Analysis," in *Proceedings of the 37th Annual Meeting of the Society of Exploration Geophysics*, 1967.