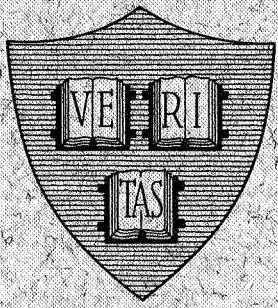


AD-669510

Office of Naval Research
Contract N00014-67-A-0298-0006 NR-372-012
NATIONAL AERONAUTICS AND SPACE ADMINISTRATION
Grant NGR 22-007-068

ON OPTIMAL AND SUBOPTIMAL LINEAR SMOOTHING



By
R. K. Mehra

GPO PRICE \$ _____

CFSTI PRICE(S) \$ _____

Hard copy (HC) _____

Microfiche (MF) _____

March 1968

ff 653 July 65

68-35746

(ACCESSION NUMBER)

34

(PAGES)

CR-97072

(NASA CR OR TMX OR AD NUMBER)

AD669510

(THRU)

(CODE)

(CATEGORY)

FACILITY FORM 602

Technical Report No. 559

Reproduction in whole or in part is permitted for any purpose of the United States Government.

Division of Engineering and Applied Physics
Harvard University • Cambridge, Massachusetts

Office of Naval Research

Contract N00014-67-A-0298-0006

NR - 372 - 012

National Aeronautics and Space Administration

Grant NGR 22-007-068

ON OPTIMAL AND SUBOPTIMAL LINEAR SMOOTHING

By

R. K. Mehra

Technical Report No. 559

Reproduction in whole or in part is permitted for
any purpose of the United States Government.

March 1968

The research reported in this document was made possible through support extended the Division of Engineering and Applied Physics, Harvard University by the U. S. Army Research Office, the U. S. Air Force Office of Scientific Research and the U. S. Office of Naval Research under the Joint Services Electronics Program by Contracts N00014-67-A-0298-0006, 0005, and 0008 and by the National Aeronautics and Space Administration under Grant NGR 22-007-068.

Division of Engineering and Applied Physics

Harvard University Cambridge, Massachusetts

ON OPTIMAL AND SUBOPTIMAL LINEAR SMOOTHING

By

R. K. Mehra

Division of Engineering and Applied Physics
Harvard University Cambridge, Massachusetts

ABSTRACT

Recursive form of results on smoothing for linear dynamic systems were first given by Bryson and Frazier (1962). Alternate formulations of the problem were given by Rauch et al (1965), Mayne (1966), Fraser (1967) and Kailath (1968). The present report shows that the results of Mayne [10] and Fraser [5] can be derived in a more general setting using the Orthogonality Principle of Linear Estimation [7]. The form of the results is particularly useful for the sensitivity analysis of the optimal smoother. Explicit equations are derived for the actual covariance of a suboptimal smoother which uses wrong information about the mean square values of noise inputs. Conditions are established under which the calculated values of the covariances provide upper bounds on the actual covariances of the smoothed estimates.

1 INTRODUCTION

Optimal smoothing of data using the Bryson-Frazier [1] or the Rauch-Tung-Striebel [2] approach requires that the system and noise parameters be known exactly. In many practical situations, these parameters are either unknown or only approximately known. It is then important to know how the performance of the optimal smoother is affected by using incorrect values for these parameters. This report presents the equations for the actual covariance of the smoother which is no longer optimal under these conditions and compares them with the calculated values using optimal smoother covariance equations. This leads to convenient error bounds for the suboptimal smoother.

To start with, a relationship between filtering and smoothing due to Fraser [5] is derived from basic principles in a more general setting. It is then applied to Gauss-Markov processes and the results of Bryson-Frazier [1] and Rauch-Tung-Striebel [2] are derived from it. The effect of errors in the system and noise parameters on forward and backward filtering is studied along the lines of Nishimura [3] and Fitzgerald [4]. These results are then applied to smoothing.

2 A GENERAL RELATIONSHIP BETWEEN LINEAR FILTERING AND SMOOTHING

In this section, a relationship between filtering and smoothing estimates due to Fraser [5] will be derived for a wider class of problems. Fraser [5] derives this relationship using Rauch-Tung-Striebel form [2] of equations. However, it can be derived easily using basic principles in a more general setting. The method of proof presented here is simple and rests on the orthogonality principle of Optimal Linear Estimation [7].

2.1. Statement of the Problem

Let $\{x(t), t_0 \leq t \leq T\}$ be a multidimensional random process and let $\{z(t), t_0 \leq t \leq T\}$ be another multidimensional random process related to $x(t)$.^{*} Denote by

$\hat{x}(t)$ = Filtering estimate of $x(t)$ based on $\{z(\tau), t_0 \leq \tau \leq t\}$

$\hat{x}_0(t)$ = Estimate of $x(t)$ based on $\{z(\tau), t < \tau \leq T\}$

$\hat{x}(t/T)$ = Smoothing estimate of $x(t)$ based on $\{z(\tau), t_0 \leq \tau \leq T\}$

All estimates are assumed to be linear minimum variance estimates. For example, $\hat{x}(t)$ which is to be obtained by a linear operation on $\{z(\tau), t_0 \leq \tau \leq t\}$ minimizes the expected value of the norm of the error, viz., $E\{(x - \hat{x})^T (x - \hat{x})\}$.^{**}

^{*}To be more precise, the random processes should be denoted as $\{x(t, \xi), t_0 \leq t \leq T\}$ and $\{z(t, \xi), t_0 \leq t \leq T\}$ where ξ is an elementary event. We shall omit ξ from the notation here.

^{**} $\hat{x}(t)$ is actually optimal for a more general loss function. See Kalman [8].

Following the notation of Fraser [5], $\hat{x}(t)$ may be called a forward filter estimate and $\hat{x}_b(t)$ may be called a backward filter estimate. Both are estimates of $x(t)$, one based on the past and the present history of z and the other based on the future history of z . Note that they have no points of z in common. Let $e(t)$ denote the error in the forward filter estimate of $x(t)$ and $e_b(t)$ denote the error in the backward filter estimate of $x(t)$. Then

$$\hat{x}(t) = x(t) - e(t) \quad (1)$$

$$\hat{x}_b(t) = x(t) - e_b(t) \quad (2)$$

Let $P(t)$ and $P_b(t)$ denote the covariance of errors $e(t)$ and $e_b(t)$ respectively.

$$P(t) = E\{e(t) e^T(t)\} \quad (3)$$

$$P_b(t) = E\{e_b(t) e_b^T(t)\} \quad (4)$$

Also let

$$P(t/T) = E\{[x(t) - \hat{x}(t/T)][x(t) - \hat{x}(t/T)]^T\} \quad (5)$$

Then $P(t/T)$ is the covariance of the error in the smoothed estimate, $\hat{x}(t/T)$.

It would be desirable to obtain the smoothing estimate $\hat{x}(t/T)$ by linearly combining the forward filtering estimate $\hat{x}(t)$ and the backward filtering estimate $\hat{x}_b(t)$. The following theorem states necessary and sufficient conditions under which this can be done.

2.2. Theorem

The necessary and sufficient condition for the smoothing estimate $\hat{x}(t/T)$ to be a linear combination of the forward filtering estimate $\hat{x}(t)$ and the backward filtering estimate $\hat{x}_b(t)$ is that the error $e(t) = x(t) - \hat{x}(t)$ in the forward filter estimate be uncorrelated to the error $e_b(t) = x(t) - \hat{x}_b(t)$ in the backward filter estimate of $x(t)$, i. e.,

$$E\{e(t) e_b^T(t)\} = 0$$

Proof

The theorem can be proved using Orthogonality Principle of Optimum Linear Estimation. The principle states that if x is a vector lying in a space M and \hat{x} is a vector lying in space M_1 where $M_1 \subseteq M$, then the minimum of norm $\|x - \hat{x}\|$ exists and is attained if and only if \hat{x} is the projection of x on M_1 . The error vector $e = \hat{x} - x$ is orthogonal to \hat{x} and lies in a subspace M_2 such that $M_1 \cup M_2 = M$ and $M_1 \perp M_2$. (M_1 is the orthogonal complement of M_2 in space M .) For a proof of the theorem, see Deutsch [7].

Notice that for the forward filter, M_1 is the subspace spanned by the vectors $\{z(\tau), t_0 \leq \tau \leq t\}$. M_2 is the orthogonal complement of M_1 . For the backward filter, let M_3 be the subspace spanned by the vectors $\{z(\tau), t < \tau \leq T\}$ and M_4 be the orthogonal complement of M_3 .

$$M_3 \cup M_4 = M, M_3 \perp M_4$$

Similarly for the smoother, let M_5 be the subspace spanned by $\{z(\tau), t_0 \leq \tau \leq T\}$ and M_6 be the orthogonal complement of M_5 .

$$M_5 \cup M_6 = M \text{ and } M_5 \perp M_6$$

By construction, $M_1 \cup M_3 = M_5$.

To prove the sufficiency part of the theorem, assume that $e(t)$ and $e_b(t)$ are uncorrelated. Therefore the vectors $e(t)$ and $e_b(t)$ are orthogonal to each other. But since $e(t)$ and $e_b(t)$ can be any vectors in the subspaces M_2 and M_4 , it follows that

$$M_2 \perp M_4$$

From the Orthogonality Principle,

$$M_2 \perp M_1 \text{ and } M_4 \perp M_3$$

Therefore

$$M_1 \perp M_3$$

We have already shown that $M_1 \cup M_3 = M_5$. Thus, M_5 is the direct sum of two orthogonal subspaces M_1 and M_3 . The projection operator from M to M_5 can, therefore, be expressed as a linear combination of projection operators from M to M_1 and from M to M_3 . Hence the smoothing estimate $\hat{x}(t/T)$ can be expressed as a linear combination of filtering estimates $\hat{x}(t)$ and $\hat{x}_b(t)$.

To prove the necessity of the condition, we reverse the argument.

Suppose that $\hat{x}(t/T)$ is expressed as a linear combination of $\hat{x}(t)$ and $\hat{x}_b(t)$. It follows that the projection operator from M to M_5 can also be expressed as a linear combination of projection operators from M to M_1 and from M to M_3 . This, in turn implies that $M_1 \perp M_3$. But since $M_1 \perp M_2$ and $M_3 \perp M_4$, it follows that $M_2 \perp M_4$. Hence the vectors $e(t)$ and $e_b(t)$ are orthogonal.

2.3 Relationship Between Filtering and Smoothing

The linear relationship between $\hat{x}(t/T)$, $\hat{x}(t)$ and $\hat{x}_b(t)$ can be derived in a number of ways. The Orthogonality Principle can be used once again or a suitable loss function J can be minimized with respect to $x(t)$.

$$J = \frac{1}{2} \{ (x - \hat{x})^T P^{-1} (x - \hat{x}) + (x - \hat{x}_b)^T P_b^{-1} (x - \hat{x}_b) \}$$

The results are well known (see Deutsch [7]) and can be expressed in a number of forms. One form of results is:

$$\hat{x}(t/T) = P(t/T) [P^{-1}(t) \hat{x}(t) + P_b^{-1}(t) \hat{x}_b(t)] \quad (6)$$

$$P^{-1}(t/T) = P^{-1}(t) + P_b^{-1}(t) \quad (7)$$

Notice that both $P(t)$ and $P_b(t)$ are positive definite matrices. Therefore, from Eq. (7),

$$P(t/T) \leq P(t)$$

$$P(t/T) \leq P_b(t)$$

In other words, the smoothing estimate is always better than the filtering estimates.

The generality of the results proved above lies in the fact that no specific form has been assumed for the forward and the backward filters. In other words, no assumptions have been made about the systems generating random processes $\{x(t), t_0 \leq t \leq T\}$ and $\{z(t), t_0 \leq t \leq T\}$. The systems can be differential or nondifferential. In Section 4, it will be shown that similar results hold for sequences of random variables. The Gaussian assumption for

random variables has not been explicitly made, but the linear estimates are strict sense optimum only for the Gaussian case. We shall apply these results to a number of cases in the next few sections.

3 FILTERING AND SMOOTHING FOR GAUSS-MARKOV RANDOM PROCESSES

A Gaussian random process whose future is dependent only on its past and its present is called a Gauss-Markov process. The output of a linear system excited by white noise is a Gauss-Markov process. Kalman and Bucy [6] derived equations of an optimal filter for a Gauss-Markov process. Bryson and Frazier [1] and Rauch, Tung and Striebel [2] derived corresponding equations for an optimal smoother. In this section, we shall rederive the results of [1] and [2] using the general relationship derived in Section 2. The approach in Section 3.1 follows closely that of Fraser [5].

3.1 White Noise in the Measurements

Consider a linear dynamic system (or the linearized equations of a nonlinear system) of the following form:

$$\dot{\mathbf{x}} = \mathbf{F}\mathbf{x} + \mathbf{G}\mathbf{u} \quad t_0 \leq t \leq T \quad (8)$$

where

\mathbf{x} = $n \times 1$ state vector of the system

\mathbf{F} = $n \times n$ matrix of time varying functions

\mathbf{u} = $q \times 1$ vector of random forcing functions (white noise process)

\mathbf{G} = $n \times q$ matrix of time varying functions

$$E\{\mathbf{u}(t)\} = 0; E\{\mathbf{u}(t) \mathbf{u}^T(\tau)\} = \mathbf{Q}(t) \delta(t - \tau)$$

$E\{ \}$ denotes expected value of the quantity within the brackets. Continuous measurements $z(t)$ are made on the system over the time interval $\{t_0 \leq t \leq T\}$ and are of the type,

$$z(t) = Hx + v \quad (9)$$

where

$z = r \times 1$ vector of measurements

$H = r \times n$ matrix of time varying functions

$v = r \times 1$ vector of random measurement errors
(white noise process).

and

$$E\{v(t)\} = 0; \quad E\{v(t) v^T(\tau)\} = R(t) \delta(t - \tau)$$

We assume that the white noise processes u and v are uncorrelated,

$$E\{uv^T\} = 0$$

The initial conditions for the system, $x(t_0)$, are also random and normally distributed with

$$E\{x(t_0)\} = 0; \quad E\{x(t_0) x^T(t_0)\} = P(t_0)$$

Kalman and Bucy [6] Filter for the above system is given by the following set of equations:

$$\frac{d}{dt} \hat{x}(t) = F\hat{x} + K(z - H\hat{x}); \quad \hat{x}(t_0) = 0 \quad (10)$$

$$K = PH^T R^{-1} \quad (11)$$

$$\frac{d}{dt} P(t) = FP + PF^T + GQG^T - PH^T R^{-1} HP; P(t_0) \text{ given} \quad (12)$$

The backward filter for this system can be derived using $\tau = (T + t_0 - t)$ as the variable of integration.

Using $\frac{d}{d\tau} = -\frac{d}{dt}$, Eqs. (8) and (9) can be written as follows to derive the backward filter:

$$\frac{d}{d\tau} x = -Fx - Gu \quad (13)$$

$$z(\tau) = Hx + v \quad (14)$$

$$\text{for } t_0 \leq \tau \leq T$$

Using the forward filter equations (10), (11) and (12), the backward filter equation can be written by changing F to $-F$, G to $-G$ and t to τ .

$$\frac{d}{d\tau} \hat{x}_b(\tau) = -F \hat{x}_b + K_b(z - H\hat{x}_b) \quad (15)$$

$$K_b = P_b H^T R^{-1} \quad (16)$$

$$\frac{d}{d\tau} P_b(\tau) = -FP_b - P_b F^T + GQG^T - P_b H^T R^{-1} HP_b \quad (17)$$

Changing the variable of integration back to t ,

$$\frac{d}{dt} \hat{x}_b(t) = F\hat{x}_b - K_b(z - H\hat{x}_b) \quad (18)$$

$$K_b = P_b H^T R^{-1} \quad (19)$$

$$\frac{d}{dt} P_b(t) = FP_b + P_b F^T - GQG^T + P_b H^T R^{-1} H P_b \quad (20)$$

The appropriate boundary conditions for the backward filter will be derived shortly.

The error $e(t)$ and $e_b(t)$ in the forward and backward filtering estimates obey the following equations.

$$\frac{d}{dt} e(t) = (F - KH)e + Gu - Kv \quad (21)$$

$$\frac{d}{d\tau} e_b(\tau) = -(F + K_b H)e_b - (Gu + K_b v) \quad (22)$$

Notice that $e(t)$ depends on $\{u(\tau), v(\tau), t_0 \leq \tau \leq t\}$, but $e_b(t)$ depends on $\{u(\tau), v(\tau), t < \tau \leq T\}$. Since both $u(t)$ and $v(t)$ are white noises, it follows that $e(t)$ and $e_b(t)$ are uncorrelated to each other. Using the theorem of Section 2.2, the smoothing estimate $\hat{x}(t/T)$ can, therefore, be expressed as a linear combination of the forward and backward filtering estimates. From Eq. (6) and (7)

$$\hat{x}(t/T) = P(t/T) \{P^{-1}(t) \hat{x}(t) + P_b^{-1}(t) \hat{x}_b(t)\} \quad (23)$$

$$P^{-1}(t/T) = P^{-1}(t) + P_b^{-1}(t) \quad (24)$$

The smoother estimate must be the same as the forward filter estimate at $t = T$, i.e.,

$$\hat{x}(T/T) = \hat{x}(T), \quad P(T/T) = P(T)$$

So, the boundary conditions for the backward filter must be $P_b^{-1}(T) = 0$. The boundary condition on $\hat{x}_b(T)$ is unknown. This makes it necessary to transform Eq. (18) by defining a new variable $y(t) = P_b^{-1}(t) \hat{x}_b(t)$. To obtain an equation

for $\frac{d}{dt} y$, we must get $\frac{d}{dt} (P_b^{-1})$. Equation (20) can be written in terms of P_b^{-1} by using the matrix differentiation lemma,

$$\frac{d}{dt} P_b^{-1}(t) = - P_b^{-1} \left(\frac{d}{dt} P_b \right) P_b^{-1} \quad (25)$$

The resulting equation is

$$\frac{d}{dt} P_b^{-1}(t) = - P_b^{-1} F - F^T P_b^{-1} + P_b^{-1} G Q G^T P_b^{-1} - H^T R^{-1} H \quad (26)$$

Using Eq. (26) and (18), we obtain

$$\frac{d}{dt} y(t) = - (F - G Q G^T P_b^{-1})^T y - H^T R^{-1} z, \quad y(T) = 0 \quad (27)$$

The Rauch-Tung-Striebel form of smoothing equations can be obtained directly by differentiating Eq. (23) and (24) and using Eq. (26) along with the following equation derived from Eq. (12).

$$\frac{d}{dt} P^{-1}(t) = - P^{-1} F - F^T P^{-1} - P^{-1} G Q G^T P^{-1} + H^T R^{-1} H$$

Differentiating Eq. (23)

$$\begin{aligned} \frac{d}{dt} P^{-1}(t/T) &= \frac{d}{dt} P^{-1}(t) + \frac{d}{dt} P_b^{-1}(t) \\ &= - (P^{-1} + P_b^{-1}) F - F^T (P^{-1} + P_b^{-1}) - P^{-1} G Q G^T P^{-1} + P_b^{-1} G Q G^T P_b^{-1} \\ &= - P^{-1}(t/T) F - F^T P^{-1}(t/T) - P^{-1}(t) G Q G^T P^{-1}(t) \\ &\quad + (P^{-1}(t/T) - P^{-1}(t)) G Q G^T (P^{-1}(t/T) - P^{-1}(t)) \\ &= - P^{-1}(t/T) (F + G Q G^T P^{-1}(t)) - (F + G Q G^T P^{-1}(t))^T P^{-1}(t/T) \\ &\quad + P^{-1}(t/T) G Q G^T P^{-1}(t/T) \end{aligned}$$

or

$$\frac{d}{dt} P(t/T) = (F + GQG^T P^{-1}(t)) P^{-1}(t/T) + P^{-1}(t/T) (F + GQG^T P^{-1}(t))^T - GQG^T \quad (28)$$

Similarly by differentiating Eq. (23) and expressing results only in terms of the forward filter estimates, it can be shown that

$$\frac{d}{dt} \hat{x}(t/T) = F \hat{x}(t/T) + GQG^T P^{-1}(t) (\hat{x}(t/T) - \hat{x}(t)) \quad (29)$$

Equation (28) and (29) are the same as derived by Rauch, Tung and Striebel [2] using a variational approach.

Bryson-Frazier [1] results can be derived by defining

$$\lambda(t) = P^{-1}(t) P(t/T) P_b^{-1}(t) (\hat{x}(t) - \hat{x}_b(t)) \quad (30)$$

where $\lambda(t)$ are the adjoint variables used by Bryson-Frazier [1].

Substituting for \hat{x}_b in Eq. (23) using Eq. (30), we get

$$\hat{x}(t/T) = \hat{x}(t) - P(t) \lambda(t) \quad (31)$$

Differentiating Eq. (30), the following equation for $\lambda(t)$ can be obtained.

$$\frac{d}{dt} \lambda(t) = - (F - P(t)H^T R^{-1}H)^T \lambda + H^T R^{-1} (z - H\hat{x}(t)) \quad (32)$$

$$\lambda(T) = 0.$$

It is also easy to show using Eq. (31) and (30) that

$$P(t/T) = P(t) - P(t) \Lambda(t) P(t) \quad (33)$$

where

$$\Lambda(t) = E\{\lambda(t) \lambda^T(t)\}$$

and

$$\frac{d}{dt} \Lambda(t) = - (F - PH^T R^{-1} H)^T \Lambda - \Lambda (F - PH^T R^{-1} H) - H^T R^{-1} H \quad (34)$$

$$\Lambda(T) = 0$$

$\Lambda(t)$ can be related to $P(t)$ and $P_b(t)$ by inverting Eq. (24)

$$\begin{aligned} P(t/T) &= (P^{-1}(t) + P_b^{-1}(t))^{-1} \\ &= P(t) - P(t) [P(t) + P_b(t)]^{-1} P(t) \end{aligned}$$

Comparing the above equation with Eq. (31)

$$\Lambda(t) = [P(t) + P_b(t)]^{-1}$$

3.2 Correlated Noise in the Measurements

This is the case treated in Ref. [12] using a variational approach. To derive it using the theorem of Section 2.2, the state of the system is augmented with correlated noise components. This leads to the case of perfect measurements for which Bryson and Johansen [9] derived the optimum filter. Their results were derived in Ref. [12] as part of the smoothing results. It

was found necessary to differentiate the perfect measurements till they contained white noise in them. The results of theorem 2.2 apply once the filtering estimates are known. The extension is straightforward once an equivalent system with white noise in the measurement is derived.

4 FILTERING AND SMOOTHING FOR SEQUENCES OF RANDOM VARIABLES

Let x_1, \dots, x_N be an unknown sequence of N random variables and let z_1, \dots, z_N be a known sequence of N random variables.

Define,

\hat{x}_i = Forward filter estimate of x_i based on z_1, \dots, z_i

\hat{x}_i^b = Backward filter estimate of x_i based on z_{i+1}, \dots, z_N

$\hat{x}_{i/N}$ = Estimate of x_i based on z_1, \dots, z_N

Notice that the backward filter estimate \hat{x}_i^b does not use z_i . The important point is that the forward and the backward filters should not use any data point in common. If the forward filter estimate uses z_i , then the backward filter estimate should not use z_i and vice versa.

The errors e_i and e_i^b can be defined as in the continuous case. It is easy to prove the following theorem which is the discrete analog of theorem 2.2.

Theorem

The necessary and sufficient condition for the smoothing estimate $\hat{x}_{i/N}$ to be a linear combination of the forward filtering estimate \hat{x}_i and the

backward filtering estimate \hat{x}_i^b is that the error $e_i = x_i - \hat{x}_i$ in the forward filter estimate be uncorrelated to the error $e_i^b = x_i - \hat{x}_i^b$ in the backward filter estimate of x_i , i. e.,

$$E\{e_i^b e_i^T\} = 0$$

5 EFFECT OF ERRORS IN $P(t_0)$, $Q(t)$ AND $R(t)$ MATRICES ON OPTIMAL FILTERING AND SMOOTHING ERROR COVARIANCE MATRICES

To illustrate the general procedure, we shall consider the specific system of Section 3.1. Nishimura [3] considered the effect of errors in $P(t_0)$, $Q(t)$ and $R(t)$ matrices on the forward filter error covariance matrix. Similar results would be derived here for the inverses of the error covariances both for the forward filter and the backward filter. These results will then be applied to an optimal smoother.

5.1 Forward Filter

Let $P_c(t_0)$, $Q_c(t)$ and $R_c(t)$ be the best known values of $P(t_0)$, $Q(t)$ and $R(t)$. In the absence of any further knowledge about these parameters, one may construct a filter assuming these to be exact values. Such a filter would be necessarily suboptimal and the orthogonality principle (of Section 2.2) would not hold for it. Consider, for example, Eq. (10), (11) and (12) with $P(t_0)$ replaced by $P_c(t_0)$, $Q(t)$ replaced by $Q_c(t)$ and $R(t)$ replaced by $R_c(t)$. The value of the gain matrix $K(t)$ would be different from its optimal value due to errors in the assumed values of $P(t_0)$, Q and R . Further, Eq. (12) would no longer give the actual covariance of the suboptimal filter. We shall derive equations for the actual covariance of the suboptimal filter. As the purpose here is to

derive equations for smoothing, it would be more convenient to work with the inverses of the filter covariance matrices. It will be assumed that the inverses of these matrices exist.

Let

$$D_c(t) = P_c^{-1}(t)$$

where $P_c(t)$ is the calculated value of the error covariance matrix obtained from Eq. (12) by replacing $P(t_0)$ by $P_c(t_0)$, $Q(t)$ by $Q_c(t)$ and $R(t)$ by $R_c(t)$. We would write down the equation in terms of $D_c(t)$. Using the matrix lemma stated in Eq. (25) .

$$\frac{d}{dt} D_c(t) = -D_c F - F^T D_c - D_c G Q_c G^T D_c + H^T R_c^{-1} H$$

or

$$\frac{d}{dt} D_c(t) = -D_c (F + N) - (F + N)^T D_c + D_c G Q_c G^T D_c + H^T R_c^{-1} H \quad (35)$$

where

$$N = G Q_c G^T D_c$$

Also

$$D_c(t_0) = P_c^{-1}(t_0)$$

The form of smoothing Eq. (23) indicates that it would be more convenient to write the filtering equation for $D_c(t) \hat{x}(t)$ instead of $\hat{x}(t)$.

Let

$$\eta(t) = D_c(t) \hat{x}(t)$$

$$\begin{aligned} \frac{d}{dt} \eta(t) &= \frac{d}{dt} D_c \hat{x} + D_c \frac{d}{dt} (\hat{x}) \\ &= (-D_c F - F^T D_c - D_c G Q_c G^T D_c + H^T R_c^{-1} H) \hat{x} + D_c (F \hat{x} + Kz - KH \hat{x}) \end{aligned}$$

where

$$K = P_c H^T R_c^{-1}$$

Simplifying the above equation,

$$\frac{d}{dt} \eta(t) = - (F + N)^T \eta + H^T R_c^{-1} z \quad (36)$$

Define e_η , the error in η as follows:

$$\begin{aligned} e_\eta(t) &= D_c(t) x(t) - \eta(t) \\ &= D_c(t) (x(t) - \hat{x}(t)) \end{aligned}$$

and

$$D(t) = E\{e_\eta(t) e_\eta^T(t)\}$$

Notice that $D(t) = D_c(t) P(t) D_c(t)$ where $P(t)$ is actual covariance of $\hat{x}(t)$. An equation for $e_\eta(t)$ can be written down using Eq. (35) and (36)

$$\frac{d}{dt} e_\eta(t) = - (F + N)^T e_\eta + D_c G u - H^T R_c^{-1} v$$

Then

$$\frac{d}{dt} D(t) = E\{\dot{e}_\eta e_\eta^T + e_\eta \dot{e}_\eta^T\}$$

Substituting for e_η and \dot{e}_η and simplifying,

$$\frac{d}{dt} D(t) = -D(F + N) - (F + N)^T D + D_c G Q G^T D_c + H^T R_c^{-1} R R_c^{-1} H \quad (37)$$

$$D(t_0) = D_c(t_0) P(t_0) D_c(t_0)$$

Subtracting Eq. (37) from Eq. (35),

$$\begin{aligned} \frac{d}{dt} (D_c(t) - D(t)) &= - (D_c - D) (F + N) - (F + N)^T (D_c - D) \\ &+ D_c G (Q_c - Q) G^T D_c + H^T R_c^{-1} (R_c - R) R_c^{-1} H \end{aligned} \quad (38)$$

Boundary conditions are

$$\begin{aligned} D_c(t_0) - D(t_0) &= D_c(t_0) [I - P(t_0) D_c(t_0)] \\ &= D_c(t_0) [P_c(t_0) - P(t_0)] D_c(t_0) \end{aligned}$$

Equation (38) is a linear differential equation. Its solution can be written down in terms of a transition matrix $\Phi(t, t_0)$ defined as follows:

$$\frac{d}{dt} \Phi(t, t_0) = - (F + N)^T \Phi(t, t_0) \quad (39)$$

$$\Phi(t_0, t_0) = I$$

$$\begin{aligned}
 D_c(t) - D(t) &= \Phi(t, t_0) [D_c(t_0) - D(t_0)] \Phi^T(t, t_0) \\
 &+ \int_{t_0}^t \Phi(t, \tau) [D_c G(Q_c - Q)G^T D_c + H^T R_c^{-1} (R_c - R)R_c^{-1} H]_{,\tau} \Phi^T(t, \tau) d\tau
 \end{aligned}
 \tag{40}$$

Positive Semidefiniteness Lemma I:

A set of sufficient conditions for $D_c(t) \geq D(t)$ and $P_c(t) \geq P(t)$ over $\{t_0 \leq t \leq T\}$ are:

- (i) $P_c(t_0) \geq P(t_0)$
- (ii) $Q_c(t) \geq Q(t)$ over $\{t_0 \leq t \leq T\}$
- (iii) $R_c(t) \geq R(t)$ over $\{t_0 \leq t \leq T\}$

Proof:

The above lemma is easily proved by noting that each term on the right hand side of Eq. (40) is positive semidefinite under conditions (i), (ii) and (iii). Therefore the sum of these terms, viz., $[D_c(t) - D(t)]$, is also positive semidefinite. To show that $P_c(t) \geq P(t)$, write $(D_c - D)$ as follows:

$$D_c - D = D_c - D_c P D_c = D_c (P_c - P) D_c$$

So if $(D_c - D)$ is positive semidefinite, so is $(P_c - P)$.

5.2 Backward Filter

Equations for the backward filter are obtained from Eq. (26) and (27) by changing Q to Q_c and R to R_c .

$$\frac{d}{dt} y(t) = - (F - GQ_c G^T P_{bc}^{-1})^T y - H^T R_c^{-1} z \quad (41)$$

$$y(T) = 0$$

$$\frac{d}{dt} P_{bc}^{-1}(t) = - P_{bc}^{-1} F - F^T P_{bc}^{-1} + P_{bc}^{-1} G Q_c G^T P_{bc}^{-1} - H^T R_c^{-1} H \quad (42)$$

$$P_{bc}^{-1}(T) = 0$$

$P_{bc}^{-1}(t)$ denotes the calculated value of the error covariance of estimate $y(t)$. For further manipulations, it would be convenient to write Eq. (41) and (42) slightly differently by using the following notation:

Let

$$B_c(t) = P_{bc}^{-1}(t)$$

$$M(t) = G Q_c G^T P_{bc}^{-1}(t)$$

Eqs. (41) and (42) can be written as

$$y(t) = - (F - M)^T y - H^T R_c^{-1} z \quad (43)$$

$$y(T) = 0$$

or

$$\frac{d}{dt} B_c(t) = - (F - M)^T B_c - B_c (F - M) - B_c G Q_c G^T B_c - H^T R_c^{-1} H \quad (44)$$

$$B_c(T) = 0$$

Let $B(t)$ denote the error covariance of $y(t)$, i. e.,

$$B(t) = E\{e_y e_y^T\}$$

where $e_y = B_c x - y$.

$$B(t) = B_c P_b B_c$$

where P_b is the actual covariance of \hat{x}_b .

Then

$$\begin{aligned} \frac{d}{dt} e_y &= (-B_c F - F^T B_c + B_c G Q_c G^T B_c - H^T R_c H)x + B_c (Fx + Gu) \\ &\quad + (F - M)^T y + H^T R_c^{-1} z \end{aligned}$$

or

$$\frac{d}{dt} e_y = - (F - M)^T e_y + H^T R_c^{-1} v + B_c Gu \quad (45)$$

$$e_y(\Gamma) = 0$$

An expression for $B(t)$ similar to Eq. (44) can be derived using Eq. (45). A change of variable to $\tau = (\Gamma + t_0 - t)$ would make the derivation easier.

$$\frac{d}{d\tau} B(\tau) = (F - M)^T B + B(F - M) + B_c G Q_c G^T B_c + H^T R_c^{-1} R_c R_c^{-1} H \quad (46)$$

$$B(\tau = t_0) = 0$$

Equation (44) may also be written in terms of τ as follows:

$$\frac{d}{d\tau} B_c(\tau) = (F - M)^T B_c + B_c (F - M) + B_c G Q_c G^T B_c + H^T R_c^{-1} H \quad (47)$$

$$B_c(\tau = t_0) = 0$$

Subtracting Eq. (46) from (47), we obtain

$$\begin{aligned} \frac{d}{d\tau} (B_c - B) &= (F - M)^T (B_c - B) + (B_c - B) (F - M) + B_c G (Q_c - Q) G^T B_c \\ &\quad + H^T R_c^{-1} (R_c - R) R_c^{-1} H \end{aligned} \quad (48)$$

with the boundary condition

$$(B_c - B)_{\tau=t_0} = 0$$

The solution to linear differential Eq. (48) can be written in terms of a transition matrix $\Phi_b(\tau, t_0)$ defined as follows:

$$\frac{d}{d\tau} \Phi_b(\tau, t_0) = (F - M)^T \Phi_b(\tau, t_0) \quad (49)$$

$$\Phi_b(t_0, t_0) = I$$

Then

$$B_c(\tau) - B(\tau) = \int_{t_0}^{\tau} \Phi_b(s, t_0) [B_c G (Q_c - Q) G^T B_c + H^T R_c^{-1} (R_c - R) R_c^{-1} H]_s \Phi_b^T(s, t_0) ds \quad (50)$$

Notice that to obtain $B_c(t)$ where t denotes time from t_0 , we put $\tau = (T + t_0 - t)$ in Eq. (50).

Positive Semidefiniteness Lemma II:

A set of sufficient conditions for $B_c(t) \geq B(t)$ and $P_{bc}(t) \geq P_b(t)$ over $\{t_0 \leq t \leq T\}$ are

(i) $Q_c(t) \geq Q(t)$ over $\{t_0 \leq t \leq T\}$

(ii) $R_c(t) \geq R(t)$ over $\{t_0 \leq t \leq T\}$

Proof:

Under conditions (i) and (ii), each term on the right hand side of Eq. (50) is positive semidefinite. Hence the sum is also positive semidefinite.

Moreover

$$\begin{aligned} B_c - B &= B_c (B_c^{-1} - P_b) B_c \\ &= B_c (P_{bc} - P_b) B_c \end{aligned}$$

Therefore $(P_{bc} - P_b)$ is positive semidefinite.

5.3 Smoother

The smoothed estimate is obtained using Eq. (23) and (24). In terms of the new notation of Sections 5.1 and 5.2, smoother equations can be written as follows:

$$P_c^{-1}(t/T) = D_c(t) + B_c(t) \tag{51}$$

and

$$P_c^{-1}(t/T) \hat{x}(t/T) = \eta(t) + y(t) \tag{52}$$

The actual covariance of the smoother denoted as $P(t/T)$ can be obtained by subtracting $P_c^{-1}(t/T) x(t)$ from both sides of Eq. (52).

$$P_c^{-1}(t/T) [x(t) - \hat{x}(t/T)] = e_\eta(t) + e_y(t)$$

Taking expectations,

$$P_c^{-1}(t/T) P(t/T) P_c^{-1}(t/T) = D(t) + B(t) \quad (53)$$

since $e_\eta(t)$ and $e_y(t)$ are uncorrelated.

Subtracting Eq. (53) from (51)

$$P_c^{-1}(t/T) [P_c(t/T) - P(t/T)] P_c^{-1}(t/T) = [D_c(t) - D(t)] + [B_c(t) - B(t)]$$

or

$$P_c(t/T) - P(t/T) = P_c(t/T) [D_c(t) - D(t)] P_c(t/T) + P_c(t/T) [B_c(t) - B(t)] P_c(t/T) \quad (54)$$

Equation (54) is an analytical expression for the error in the calculated smoother covariance. It can be used to study the sensitivity of the smoother to errors in Q and R matrices.

Positive Semidefinitenn Lemma III:

A set of sufficient conditions for $P_c(t/T) \geq P(t/T)$ over $\{t_0 \leq t \leq T\}$ are

- (i) $P_c(t_0) \geq P(t_0)$
- (ii) $Q_c(t) \geq Q(t)$ over $\{t_0 \leq t \leq T\}$
- (iii) $R_c(t) \geq R(t)$ over $\{t_0 \leq t \leq T\}$

Proof:

It has already been shown in Lemmas I and II that under conditions (i), (ii) and (iii), $D_c(t) \geq E(t)$ and $B_c(t) \geq B(t)$. Hence from Eq. (54), it follows that $P_c(t/T) \geq P(t/T)$.

5.4 Implications of Lemmas I, II and III

Let the state vector x be one dimensional (a scalar). The Lemmas proved above state that if the estimated values $P_c(t_0)$, $Q_c(t)$, $R_c(t)$ are greater than or equal to the actual values $P(t_0)$, $Q(t)$, $R(t)$ respectively, then the calculated error variances are greater than or equal to the actual variances, i. e.,

$$P_c(t) \geq P(t)$$

$$P_{bc}(t) \geq P_b(t)$$

$$P_c(t/T) \geq P(t/T)$$

for all t in the interval $\{t_0 \leq t \leq T\}$. Thus the computed values provide upper bounds on the actual values. For an n -dimensional system, the above remarks would apply to all the diagonal elements of the error covariance matrices. The diagonal elements of these matrices represent the mean square error terms for the estimates of the state variables of the system. In a number of practical situations, though the actual values of $P(t_0)$, $Q(t)$ and $R(t)$ are unknown, one can specify suitable upper bounds on these quantities. Then the above Lemmas state that the diagonal elements of the computed covariance matrices $P_c(t)$ and $P_c(t/T)$ will provide upper bounds on the diagonal elements of matrices $P(t)$ and $P(t/T)$. One advantage of Kalman filtering and recursive optimal smoothing techniques is that one can perform the error analysis without the use of actual data. In this way very useful feasibility studies can be carried out before collecting actual data. If one can estimate a state with satisfactory accuracy using

upper bounds for $P(t_0)$, Q and R then it is assured that the state would be estimated to at least that much accuracy when the actual tests are performed. If some adaptive scheme is used to improve estimates of $P(t_0)$, Q and R , then the signal would, of course, be estimated to a greater accuracy.

5.5 Error Compensation Effect In a Smoother

Equation (54) reveals another interesting fact about smoothing distinct from filtering, viz., it is possible to compensate partially for the errors in the smoother by making errors in the forward and the backward filters of opposite signs. This is done by choosing Q_c and R_c differently for the forward and the backward filter. The compensation, however, will not be uniformly good over the whole time interval unless Q_c and R_c are made time varying.

6 EFFECT OF ERRORS IN F, G AND H MATRICES ON OPTIMUM SMOOTHING

The detailed equations for this case are rather messy and will not be presented here. The method of approach is the same as in Section 5. The development is parallel to that of Fitzgerald [4] who studies the effect of errors in F , G and H on a Kalman filter. Fitzgerald shows that under certain conditions, the actual covariance matrix may diverge from the calculated covariance matrix. Same thing can happen to both the forward and the backward filters. However, the divergence for the smoother may not be so bad because the errors in the forward and the backward filters may try to compensate each other.

7 SUMMARY AND CONCLUSIONS

The results of Fraser [5] on "Smoother As A Combination of Two Kalman Filters" have been generalized and derived from basic principles. Necessary and sufficient conditions are obtained under which an optimal smoother may be expressed as a linear combination of two optimal filters, one operating forward on the data (forward filter) and the other operating backward on the data (backward filter). The smoothing results of Bryson-Frazier [1] and Rauch-Tung-Striebel [2] are derived using the general relationship between filtering and smoothing estimates. The effect of errors in system and noise parameters on optimum smoothing has been studied using the Two Kalman filter approach. An analytical expression for the error in the calculated smoother covariance matrix has been presented. This expression is then used to establish upper bounds on the actual smoother covariance matrix.

ACKNOWLEDGEMENTS

The author is very grateful to Professor Bryson of Harvard University and to S. Levine of The Analytic Sciences Corporation, Winchester, for their generous help and suggestions during the course of this research.

Page intentionally left blank

PRECEDING PAGE BLANK NOT FILMED.

- [10] Mayne, D. Q., "A Solution of the Smoothing Problem for Linear Dynamic Systems," *Automatica*, Vol. 4, pp. 73-92, Pergamon Press, 1966.
- [11] Kailath, T., "A Wold-Kolmogorov Approach to Linear Least-Squares Estimation," Part II, TR 7050-13, Stanford Electronics Laboratory, Stanford, California, January 1968.
- [12] Mehra, R. K., and A. E. Bryson, "Smoothing for Time-Varying Systems Using Measurements Containing Colored Noise," Technical Report No. 1, Harvard University, June 1967.

DOCUMENT CONTROL DATA - R & D

Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified

1. ORIGINATING ACTIVITY (Corporate author) Division of Engineering and Applied Physics Harvard University Cambridge, Massachusetts		2a. REPORT SECURITY CLASSIFICATION Unclassified	
		2b. GROUP	
3. REPORT TITLE ON OPTIMAL AND SUBOPTIMAL LINEAR SMOOTHING			
4. DESCRIPTIVE NOTES (Type of report and inclusive dates) Interim technical report			
5. AUTHOR(S) (First name, middle initial, last name) R. K. Mehra			
6. REPORT DATE March 1968	7a. TOTAL NO. OF PAGES 34	7b. NO. OF REFS 11	
8a. CONTRACT OR GRANT NO. N00014-67-A-0298-0006 and	9a. ORIGINATOR'S REPORT NUMBER(S) Technical Report 559		
b. PROJECT NO. NASA Grant NGR 22-007-068			
c.	9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)		
d.			
10. DISTRIBUTION STATEMENT Reproduction in whole or in part is permitted for any purpose of the United States Government.			
11. SUPPLEMENTARY NOTES		12. SPONSORING MILITARY ACTIVITY Office of Naval Research	
13. ABSTRACT Recursive form of results on smoothing for linear dynamic systems were first given by Bryson and Frazier (1962). Alternate formulations of the problem were given by Rauch et al (1965), Mayne (1966), Fraser (1967) and Kailath (1968). The present report shows that the results of Mayne [10] and Fraser [5] can be derived in a more general setting using the Orthogonality Principle of Linear Estimation [7]. The form of the results is particularly useful for the sensitivity analysis of the optimal smoother. Explicit equations are derived for the actual covariance of a suboptimal smoother which uses wrong information about the mean square values of noise inputs. Conditions are established under which the calculated values of the covariances provide upper bounds on the actual covariances of the smoothed estimates.			

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Optimal Suboptimal Smoothing						