

COMBINING AND UPDATING OF LOCAL ESTIMATES AND REGIONAL
MAPS ALONG SETS OF ONE-DIMENSIONAL TRACKS

Alan S. Willsky*
Martin Bello**
David A. Castanon[†]
Bernard C. Levy[†]
George Verghese^{††}

ABSTRACT

In this paper we consider the problem of combining and updating estimates that may have been generated in a distributed fashion or may represent estimates, generated at different times, of the same process sample path. The first of these cases has applications in decentralized estimation, while the second has applications in updating maps of spatially-distributed random quantities given measurements along several tracks. The method of solution for the second problem uses the result of the first, and the similarity in the formulation and solution of these problems emphasizes the conceptual similarity between many problems in decentralized control and in the analysis of random fields.

* Laboratory for Information and Decision Systems and the Department of Electrical Engineering and Computer Science, M.I.T., Cambridge, MA. 02139. The work of this author was performed in part at The Analytic Sciences Corporation, Reading, MA., and in part at M.I.T. with partial support provided by the Air Force Office of Scientific Research under Grant AFOSR-77-3281B.

** Laboratory for Information and Decision Systems and the Department of Electrical Engineering and Computer Science, M.I.T., Cambridge, MA. 02139. The work of this author was performed in part at The Analytic Sciences Corporation, Reading, MA., and in part at The M.I.T. Lincoln Laboratory.

[†] Laboratory for Information and Decision Systems and the Department of Electrical Engineering and Computer Science, M.I.T., Cambridge, MA. 02139. The work of these authors was performed at M.I.T. with support provided by the Office of Naval Research under Contract ONR/N00014-76-C-0346.

^{††}

Electric Power Systems Engineering Laboratory and Department of Electrical Engineering and Computer Science, Cambridge, MA. 02139

I. INTRODUCTION

The research reported in this paper was motivated by the following problem in the mapping of two-dimensional random fields, that is spatially distributed random quantities. Measurements or surveys of the field are collected at different times along sets of one-dimensional tracks across the field. The sets of tracks may differ from survey to survey. Either a local or regional map is generated for each of these surveys and the problem is either to combine these local maps optimally, or to update an overall map as each new survey becomes available. Problems of this type arise in many applications including the mapping of vertical temperature profiles of the atmosphere given data provided by satellites [1] and the mapping of anomalies in the earth's gravitational field and the effects of such anomalies on errors in inertial navigation systems [2,14].

The problem posed in the preceding paragraph is not solved completely in this paper, but a special case of it is in which the tracks are all parallel and the field along the direction of the tracks can be modeled by a finite dimensional linear shaping filter driven by white noise. In addition to solving this special case and to providing insight into the general case, the solution we obtain is of independent interest in that it provides a procedure for optimally updating smoothed estimates as more data is collected. Furthermore, one of the principle steps in our development is the construction of the optimal combined filtered (i.e. causal) estimate from several local filtered estimates. This is basically a problem in

decentralized filtering, and our results extend those of Speyer [3] and Chong [7].

In the next section we present and discuss the solution to the problem of combining decentralized filtered estimates, while Sections III contains the description and solution of the problem of updating smoothed estimates. In Section IV we apply the results of the preceding section to the problem of real-time smoothing, that is, of estimation given a previous smoothed estimate and new real-time data. The paper concludes with a discussion in Section V.

II. COMBINING DECENTRALIZED FILTERED ESTIMATES

2.1 Formulation and Solution of the General Case

Consider a linear dynamical system driven by Gaussian white noise

$$\dot{x}(t) = A(t)x(t) + w(t) \quad (2.1)$$

$$E[w(t)w(\tau)'] = Q(t)\delta(t-\tau) \quad (2.2)$$

where $w(t)$ is independent of $x(0)$ which is taken to be a zero mean Gaussian random variable with covariance $\Sigma(0)$. Suppose we have two sets of white noise-corrupted observations

$$y_1(t) = C_1(t)x(t) + v_1(t) \quad (2.3)$$

$$y_2(t) = C_2(t)x(t) + v_2(t) \quad (2.4)$$

where v_1 and v_2 are independent of each other and of w and $x(0)$, with

$$E[v_i(t)v_i(\tau)'] = R_i(t)\delta(t-\tau), \quad i=1,2 \quad (2.5)$$

The two sets of measurements (2.3), (2.4) can be thought of as representing observations taken at different locations in a distributed system or at different nodes in a network of interconnected systems. These observations are processed separately to produce local filtered estimates, and we wish to consider the problem of recovering the overall optimal filtered estimate

$$\hat{x}(t|t) = E\{x(t) | y_1(s), y_2(s), s \leq t\} \quad (2.6)$$

in terms of these local estimates. If this can be done, then much of the raw data processing can be done locally without any loss in global performance. In addition, if local filtering is performed on the data, we may reduce the required bandwidth for transmission of information to a centralized processor. A problem of this type was considered by Speyer [3] in the context of decentralized control. Our work represents an extension of the estimation portion of his results. Also, while we consider only two sets of measurements (2.3), (2.4), the preceding formulation and our analysis of it extend in an obvious manner to the case of N sets of measurements and local estimates.

In order to complete the formulation of the problem, we assume that the local processing algorithms are Kalman filters based on different models:

$$\dot{x}_i(t) = A_i(t)x_i(t) + w_i(t), \quad i=1,2 \quad (2.7)$$

$$E[w_i(t)w_i(\tau)'] = Q_i(t)\delta(t-\tau), \quad i=1,2 \quad (2.8)$$

$$y_i(t) = H_i(t)x_i(t) + v_i(t), \quad i=1,2 \quad (2.9)$$

where $x_i(0)$ is taken to be zero-mean with covariance $\Sigma_i(0)$. It is

important to emphasize here that (2.7)-(2.9) represents a model whose sole purpose is for the design of local Kalman filters. This model may not accurately reflect the actual statistics of y_i . At the moment we are assuming no relationship between the local model (2.7)-(2.9) and the correct global model (2.1)-(2.5), except for the assumption that the v_i in (2.9) are the same as in (2.5) (i.e. that they have the same statistics, so at least the measurement noise is modeled in the same fashion locally and globally). As we need to impose some relationship between local and global models, we will do so.

Given these local models, the equation for each local processor is given by the following,*

$$\dot{\hat{x}}_i(t|t) = [A_i - P_i H_i' R_i^{-1} H_i] \hat{x}_i(t|t) + P_i H_i' R_i^{-1} y_i(t) \quad (2.10)$$

The covariance P_i can be precomputed from either of the following equations:

$$\dot{P}_i = A_i P_i + P_i A_i' + Q_i - P_i H_i' R_i^{-1} H_i P_i \quad (2.11)$$

$$\frac{d}{dt}(P_i^{-1}) = -P_i^{-1} A_i - A_i' P_i^{-1} - P_i^{-1} Q_i P_i^{-1} + H_i' R_i^{-1} H_i \quad (2.12)$$

with the initial condition

$$P_i(0) = \Sigma_i(0) \quad (2.13)$$

The problem to be solved is to obtain an algorithm for computing

* From this point of the explicit time dependence of matrices will be suppressed. If a particular matrix is constant, we will explicitly state this in order to avoid confusion.

the global estimate \hat{x} in (2.6) in terms of \hat{x}_1 and \hat{x}_2 . Speyer in [3] solved this problem when the local model are the same as the global model, and we will comment on the simplifications that occur in that case shortly.

Allowing the local models to differ from the global model leads to several potential advantages. For example, presumably the local models are lower-dimensional than (2.1) and represent the important dynamics at that particular location in the distributed system or network. Therefore, the local processors can be made far less complex than the global processor. Of course, we cannot recover \hat{x} from \hat{x}_1 and \hat{x}_2 for arbitrary choices of local models, but the conditions needed are quite weak. Specifically, as we will see, the only condition that is required is that there exist (possibly-varying) matrices M_1 and M_2 such that

$$C_i = H_i M_i, \quad i=1,2 \quad (2.14)$$

Equation (2.14) and its implications deserve some comment. First note that (2.14) is equivalent to

$$R(C_i) \subset R(H_i), \quad i=1,2 \quad (2.15)$$

or equivalently that

$$R(C_i)^\perp \supset R(H_i)^\perp, \quad i=1,2 \quad (2.16)$$

What these conditions say is that if any set of components of $H_i x_i$ are linearly interrelated, then the same set of components of $C_i x$ must have exactly the same linear interrelationship. That is, if the local models

(2.7)-(2.9) assume any redundancy among the sensed quantities -- i.e. the components of y_i -- then that redundancy must actually exist in the global model. Note that if (2.15) is satisfied, valid choices for M_1 and M_2 are

$$M_i = H_i^+ C_i, \quad i=1,2 \quad (2.17)$$

(where "+" denotes pseudo-inverse) and the choice is unique only if $N(H_i) = \{0\}$.

Thus, the dynamics (2.7), (2.8) can be totally arbitrary, as long as (2.15) or (2.16) is satisfied. For example, one implication of this condition is that the dimension of x_i must be at least as large as the number of linearly independent components of the measurement vector y_i . However, the condition (2.15) is sufficiently weak that, if we desire, we can always choose a local model of this minimal dimension that satisfies the condition. Therefore, the conditions does not require that there by any physical relationship between the local states, x_1 and x_2 , and the global state x . On the other hand, (2.14) suggests an interpretation of x_i as being a part of the global state, specifically $M_i x$. If this is the case, then (2.7) implies that this part of the state is decoupled from the remaining part of x in the sense that $M_i x$ is itself a Markov process. This is, of course, not the usual case in practice, where approximations are made in assuming that the couplings between the local states can be neglected or can be replaced by additional white noise sources. What our results say is that as long as (2.14) holds, for the purposes of reconstructing \hat{x} , it doesn't

matter if (2.7) is an exact or approximate expression for the evolution of the local state. If x_i actually equals $M_i x$, we obtain some simplifications in the equations that define our algorithm, and we will discuss these at the end of this section.

As a first step in deriving our algorithm, consider the Kalman filter for the calculation of the global estimate \hat{x} :

$$\dot{\hat{x}}(t|t) = [A - PC_1'R_1^{-1}C_1 - PC_2'R_2^{-1}C_2]\hat{x}(t|t) + PC_1'R_1^{-1}y_1(t) + PC_2'R_2^{-1}y_2(t) \quad (2.18)$$

where P can be calculated from

$$\dot{P} = AP + PA' + Q - PC_1'R_1^{-1}C_1P - PC_2'R_2^{-1}C_2P \quad (2.19)$$

$$P(0) = \sum(0) \quad (2.20)$$

The solution to the problem we have posed can be obtained as follows.

Rearranging (2.10) we have*

$$H_i'R_i^{-1}y_i = P_i^{-1}\{\dot{\hat{x}}_i - [A_i - P_i H_i'R_i^{-1}H_i]\hat{x}_i\} \quad (2.21)$$

Examining (2.18), we see that the quantities needed in the calculation of $\dot{\hat{x}}$ are $C_1'R_1^{-1}y_1$ and $C_2'R_2^{-1}y_2$. These can be obtained from (2.21) only if matrices M_1 and M_2 exist that satisfy (2.14). Assuming that this is the case, we can combine (2.14), (2.18), and (2.21) to obtain

* Note that we have implicitly made one other assumption about the local models, in that in (2.21) we are assuming that P_i is invertible. This will be guaranteed as long as $\sum_i(0)$ is invertible

$$\begin{aligned}\dot{\hat{x}} = & [A - PC_1' R_1^{-1} C_1 - PC_2' R_2^{-1} C_2] \hat{x} \\ & + PM_1' P_1^{-1} \{ \hat{x}_1 - [A_1 - P_1 H_1' R_1^{-1} H_1] \hat{x}_1 \} \\ & + PM_2' P_2^{-1} \{ \hat{x}_2 - [A_2 - P_2 H_2' R_2^{-1} H_2] \hat{x}_2 \}\end{aligned}\quad (2.22)$$

In order to simplify notation, define the following quantities:

$$F = A - PC_1' R_1^{-1} C_1 - PC_2' R_2^{-1} C_2 \quad (2.23)$$

$$F_i = A_i - P_i H_i' R_i^{-1} H_i \quad i=1,2 \quad (2.24)$$

$$G_i = PM_i' P_i^{-1} \quad (2.25)$$

Then, in order to avoid differentiating \hat{x}_1 and \hat{x}_2 in (2.22), we define

$$\xi = \hat{x} - G_1 \hat{x}_1 - G_2 \hat{x}_2 \quad (2.26)$$

and differentiating, we find that

$$\dot{\xi} = F\xi + K_1 \hat{x}_1 + K_2 \hat{x}_2 \quad (2.27)$$

$$\hat{x} = \xi + G_1 \hat{x}_1 + G_2 \hat{x}_2 \quad (2.28)$$

where

$$K_i = FG_i - \dot{G}_i - G_i F_i \quad , i=1,2 \quad (2.29)$$

If we use the differential equations for P_i , P_i^{-1} and P , (2.29) becomes*

* Note that in (2.29) we have implicitly assumed that M_1 and M_2 are differentiable. Again this is not a particularly restrictive condition. For example, in the time-invariant case it is certainly true, since M_1 and M_2 can be taken to be constants.

$$K_i = [PM_i' P_i^{-1} Q_i P_i^{-1} - QM_i' P_i^{-1}] + [PM_i' A_i' P_i^{-1} - PA_i' M_i' P_i^{-1} - PM_i' P_i^{-1}], \quad i=1,2 \quad (2.30)$$

If all of the models, local and global, are time-invariant and if we consider the steady-state case, then the above solution still applies (with $\dot{M}_i=0$) and is also time-invariant.

This is the general solution to the problem of combining decentralized maps. In addition, this solution can be directly adapted to the problem of computing \hat{x} from \hat{x}_1 and y_2 . This is of interest in situations in which one local processor transmits information to a global processor that has measurements of its own. We can solve this problem by returning to (2.18), and instead of replacing both $C_1' R_1^{-1} y_1$ and $C_2' R_2^{-1} y_2$ by expressions in terms of \hat{x}_i and \hat{x}_i , we make this substitution only for $C_1' R_1^{-1} y_1$. The remaining analysis is analogous to that carried out previously, and the result is

$$\hat{x} = \rho + G_1 \hat{x}_1 \quad (2.31)$$

where

$$\rho = F\rho + K_1 \hat{x}_1 + PC_2' R_2^{-1} y_2 \quad (2.32)$$

Here F , K_1 , and G_1 are the same as given previously.

In the next two subsections we present two special cases which result in some simplifications in (2.23)-(2.32) and consequently allow us to interpret our result in more detail.

2.2 The Special Case of Identical Local and Global Models

In this section we consider the case examined by Speyer in [3]. Specifically, we assume that the models used by the local processors are identical to the global model. That is,

$$A_1=A_2=A, \quad Q_1=Q_2=Q, \quad C_1=H_1, \quad C_2=H_2, \quad M_1=M_2=I \quad (2.33)$$

In this case the expressions for K_1 and K_2 simplify to

$$K_i = PP_i^{-1}QP_i^{-1} - QP_i^{-1} = (PP_i^{-1}-I)QP_i^{-1} \quad (2.34)$$

and

$$\hat{x} = \xi + P(P_1^{-1}\hat{x}_1 + P_2^{-1}\hat{x}_2) \quad (2.35)$$

Note that the second term in the expression for \hat{x} is the usual expression for combining independent estimates [4,5]. However \hat{x}_1 and \hat{x}_2 are not independent in general, and ξ represents a correction for this correlation.

The reason that \hat{x}_1 and \hat{x}_2 are not independent estimates is that they are based not only on measurements with independent noises but also on a priori information. Specifically, both of the local estimates incorporate statistical descriptions of $x(0)$ and $w(t)$, and thus the errors in both estimates are correlated with these processes. It is the correlation with the process $w(t)$ that leads to the need for a dynamical correction (ξ) to account for the correlation in the processes \hat{x}_1 and \hat{x}_2 . If $Q=0$ (i.e. if $w(t)$ is not present), then $K_i=0$ and hence $\xi=0$, and \hat{x} is a memoryless function of \hat{x}_1 and \hat{x}_2 . In this case it is straightforward to show that

$$\hat{x}(t|t) = P(t) [P_1^{-1}(t) \hat{x}_1(t|t) + P_2^{-1}(t) \hat{x}_2(t|t)] \quad (2.36)$$

and

$$P^{-1}(t) = P_1^{-1}(t) + P_2^{-1}(t) - \Sigma^{-1}(t) \quad (2.37)$$

where $\Sigma(t)$ is the unconditional covariance of $x(t)$. In general $\Sigma(t)$ satisfies

$$\dot{\Sigma}(t) = A(t)\Sigma(t) + \Sigma(t)A'(t) + Q(t) \quad (2.38)$$

with $\Sigma(0)$ given. Equations (2.36) and (2.37) hold only in the case when Q is zero. Note that even in this case \hat{x}_1 and \hat{x}_2 are not independent estimates because of the correlation of the estimation errors with $x(0)$. Following the work of Wall [4], we can interpret (2.36) and (2.37) as follows. We have three sources of information on which to base our estimate of $x(t)$, the measurement processes y_1 and y_2 and the a priori information about $x(t)$, provided by the unconditional propagation of the mean and variance from the specified statistics of $x(0)$. The estimate \hat{x}_i uses y_i and the a priori information, which, therefore is used twice. Equation (2.37) corrects for the fact that both P_1^{-1} and P_2^{-1} reflect the uses of this information. Also, (2.36) is the correct expression under the assumption that $x(0)$ is zero mean. If this is not the case, that is if its mean $m(0) \neq 0$, then (2.36) is replaced by

$$\hat{x}(t|t) = P(t) [P_1^{-1} \hat{x}_1(t|t) + P_2^{-1} \hat{x}_2(t|t) - \Sigma^{-1}(t)m(t)] \quad (2.39)$$

where $m(t)$ is the unconditional mean of $x(t)$ which satisfies

$$\dot{m}(t) = Am(t) \quad (2.40)$$

Again we see the "subtracting out" of the effect of a priori information, so that the duplication of this information is removed.

Finally, note that $K_i=0$ also if $P=P_i$. However, this is only the case if the other set of measurements contains no information. In general, if the system is observable from each set of measurements, $(PP_i^{-1}-I)$ will be invertible. Of course, all of the previous statements have certain obvious generalizations. For example, if part of the state is uncontrollable from the noise, then the corresponding part of \hat{x} is a memoryless function of \hat{x}_1 and \hat{x}_2 . Also, if one set of measurements, say set 1, contains no information about a part of x , then the corresponding parts of P and P_2 are identical.

2.3 The Case in Which the Local Model is a Subsystem of the Global Model

In some cases the dynamics of one of the local models may, in fact, be the exact dynamics of a subsystem of the global model. Specifically, if this is true of local model 1, then

$$x_1(t) = M_1(t)x(t) \quad (2.41)$$

Equation (2.41) has several important implications. Since x_1 satisfies (2.7), (2.8), and x satisfies (2.1), (2.2), equation (2.41) states that the Markov process $x(t)$ has a subprocess, namely $x_1(t)$, that is Markov

by itself. Differentiating (2.41) and using (2.1) and (2.7) we have that

$$A_1 M_1 \dot{x} + w_1 = A_1 x_1 + w_1 = \dot{x}_1 = \dot{M}_1 x + M_1 \dot{x} = \dot{M}_1 x + M_1 A x + M_1 w \quad (2.42)$$

and from this we conclude that

$$A_1 M_1 = \dot{M}_1 + M_1 A \quad (2.43)$$

and

$$w_1 = M_1 w \quad (2.44)$$

which implies that

$$Q_1 = M_1 Q M_1' \quad (2.45)$$

Also, directly from (2.41) we have that

$$\Sigma_1 = M_1 \Sigma M_1' \quad (2.46)$$

Note that from (2.46) it is clear that Σ_1 is invertible only if M_1 is onto (assuming that Σ is invertible). We will assume that this is the case, since from (2.41) we see that any other choice for M_1 leads to an x_1 with fewer degrees of freedom than it has components. In addition, under these conditions, the expression for K_1 simplifies:

$$\begin{aligned} K_1 &= P M_1' P^{-1} M_1 Q M_1' P^{-1} - Q M_1' P^{-1} \\ &= [P M_1' P^{-1} M_1 - I] Q M_1' P^{-1} \end{aligned} \quad (2.47)$$

This equation bears some resemblance to the form of the gain when the local model is the same as the global model. In order to gain further

insight, we wish to consider a particularly convenient form for the global model. This is done by choosing a basis for the global state space so that the components of x_1 are the first components of x . Assuming without loss of generality that the global model is in this form, then

$$x = \begin{pmatrix} x_1 \\ \text{---} \\ \gamma \end{pmatrix} \quad (2.48)$$

$$M_1 = (I : 0) \quad (2.49)$$

$$\begin{pmatrix} \dot{x}_1 \\ \dot{\gamma} \end{pmatrix} = \begin{pmatrix} A_1 & 0 \\ A_{21} & A_{22} \end{pmatrix} \begin{pmatrix} x_1 \\ \gamma \end{pmatrix} + w \quad (2.50)$$

$$Q = \begin{pmatrix} Q_1 & Q_{12} \\ Q_{12}' & Q_{22} \end{pmatrix} \quad (2.51)$$

$$y_1 = \begin{pmatrix} H_1 & : & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ \gamma \end{pmatrix} + v_1 \quad (2.32)$$

$$y_2 = \begin{pmatrix} C_{21} & : & C_{22} \end{pmatrix} \begin{pmatrix} x_1 \\ \gamma \end{pmatrix} + v_2 \quad (2.53)$$

This form is illustrated in Figure 2.1. Note from the figure that it is clear that the global system is not observable from y_1 alone. This is not surprising given that x_1 is Markov by itself.

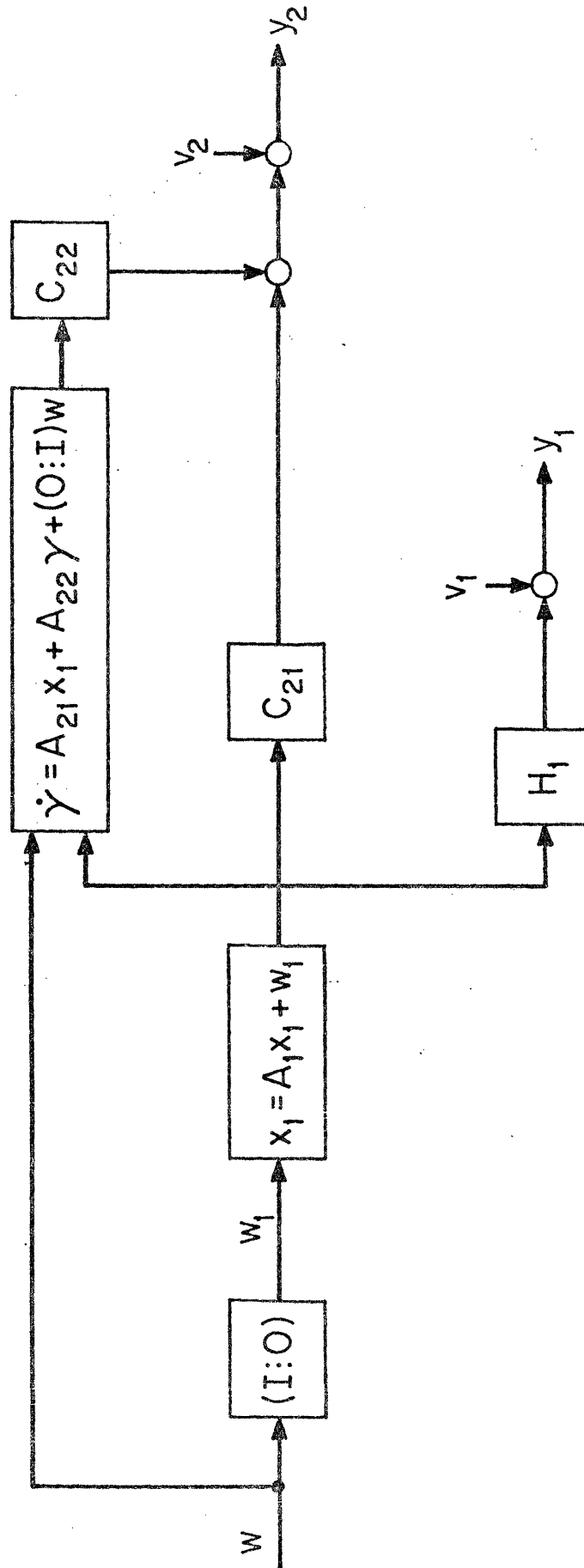


FIGURE 2.1

Using (2.48)-(2.53), equation (2.47) becomes

$$\begin{aligned}
 K_1 &= \left\{ P \begin{pmatrix} P_1^{-1} \\ \hline 0 \end{pmatrix} Q_1 P_1^{-1} - \begin{pmatrix} Q_1 \\ \hline Q_{12}' \end{pmatrix} P_1^{-1} \right\} \\
 &= \begin{pmatrix} [(P)_{11} P_1^{-1} - I] Q_1 P_1^{-1} \\ [(P)_{12}' P_1^{-1} - Q_{12}'] P_1^{-1} \end{pmatrix} \quad (2.54)
 \end{aligned}$$

where

$$P = \begin{pmatrix} (P)_{11} & (P)_{12} \\ (P)_{12}' & (P)_{22} \end{pmatrix} \quad (2.55)$$

From this and the previous equations and from the figure we can get a clearer picture of the structure of our solution in this case*. Since K_1 is partitioned, let us consider each part individually. The first piece, $[(P)_{11} P_1^{-1} - I] Q_1 P_1^{-1}$ is exactly of the form of the gain that we saw in the preceding subsection when the local and global models are identical. (see equation (2.34)). This is not surprising, as the first piece of the global state x is nothing more than x_1 , for which the local and global models agree. Therefore, the incorporation of \hat{x}_1 into a global estimate of this piece of x , given y_1 and y_2 , is the same as

* In the following discussion we use the notation developed previously. Thus \hat{x}_1 refers to the local estimate of x_1 given y_1 (P_1 is its locally-computed covariance) and \hat{x} refers to the global estimate of x given y_1 and y_2 . (global covariance P). In the particular case being examined here $x = \begin{pmatrix} x_1 \\ -\frac{x_1}{\gamma} \end{pmatrix}$, and

therefore there is some chance of confusion. We have attempted to reduce this chance by using \hat{x}_1 and \hat{x} only in the senses described above. Also, we have denoted the upper left-hand block of P by $(P)_{11}$ (see (2.55)) to distinguish it from P_1 . Here $(P)_{11}$ is the estimation error covariance of x_1 given y_1 and y_2 , while P_1 is the error covariance based only on y_1 .

the problem we considered in Subsection 2.2.

The second piece $[(P)_{12}' P_1^{-1} Q_1 - Q_{12}'] P_1^{-1}$ essentially tells us how to use the estimate of x_1 to obtain an estimate of the remaining part of the state. Consider for the moment the case in which there is no second set of measurements, that is, when $C_{21}=C_{22}=0$. In this case we have a cascade interconnection of two systems and measurements from only the first of these. It is clear that under these conditions $(P)_{11}=P_1$, which merely states that local processor #1 produces the best filtered estimate of x_1 given y_1 . From (2.54) we see that this observation is consistent with the fact that the first part of K_1 is zero. Also, the second piece of K_1 becomes

$$[P_{12}' (P)_{11}^{-1} Q_1 - Q_{12}'] (P)_{11}^{-1}$$

and using (2.23)-(2.28) and (2.54), the optimal estimator for γ becomes

$$\hat{\gamma} = \eta + (P)_{12}' (P)_{11}^{-1} \hat{x}_1 \quad (2.57)$$

$$\dot{\eta} = A_{22} \eta + [(P)_{12}' (P)_{11}^{-1} Q_1 - Q_{12}'] (P)_{11}^{-1} \hat{x}_1 \quad (2.57)$$

These equations describe how the optimal estimate of the unobservable part of a system can be constructed from the optimal estimate of the observable part. It is worth noting that this particular special case is of practical importance, for example, in navigation systems in which accelerations are sensed and in which velocities and positions are to be estimated. Our result states that the acceleration measurements can be processed first (locally) to produce optimal acceleration estimates, and

these estimates can then be used (perhaps in a centralized processor) to compute the optimal estimates of velocity and position. Again the transmission of filtered measurements may be done more efficiently than the transmission of the raw data, and the complexity of the two processors (for \hat{x}_1 and for $\hat{\gamma}$) are each less than the complexity of a global, centralized estimator for x . Such a procedure may also be of value even if y_2 is present, for example, if we do have velocity or position sensors. In this case, from eq. (2.32) we see that our results tell us how to reconstruct the optimal estimate of acceleration, velocity and position in terms of velocity and sensor measurements and the estimate of acceleration obtained by processing the accelerometers alone. Again there may be transmission savings in transmitting this estimate rather than the raw accelerometer data, and, in addition, there may be implementation advantages in breaking the overall optimal estimator into smaller pieces.

Note also from (2.47) that $K_1=0$ if $Q=0$. In fact, from (2.49) and (2.51) (together with the fact that Q_{12} must be zero if Q_1 is), we see that $K_1=0$ if $Q_1=0$. In this case, whether y_2 is present or not, \hat{x} depends on \hat{x}_1 in a memoryless fashion. This is best understood by noting that with $Q_1=0$, x_1 is a time-varying bias*

$$x_1(t) = \Phi_1(t,0)x_1(0) \quad (2.58)$$

and it also produces a time-varying bias in γ

* Here Φ_1 is the state transition matrix associated with A_1 . Similarly Φ_{22} is the state transition matrix for A_{22} .

$$\gamma(t) = \Phi_{22}(t,0)\gamma(0) + \left[\int_0^t \Phi_{22}(t,\tau)A_{21}(\tau)x_1(\tau)d\tau \right] + \int_0^t \Phi_{22}(t,\tau)[0:I]w(\tau)d\tau \quad (2.59)$$

The measurements y_1 provide information about the second term in (2.59), which can be rewritten as

$$\left[\int_0^t \Phi_{22}(t,\tau)A_{21}(\tau)\Phi(\tau,t)d\tau \right] x_1(t) \quad (2.60)$$

Thus the best estimate of γ given the measurements y_1 is simply a memoryless function of \hat{x}_1 . For example, if we do not have a second set of measurements ($C_{21}=C_{22}=0$), then (2.28) reduces to

$$\hat{x} = PM_1'P_1^{-1}\hat{x}_1 \quad (2.61)$$

where M_1 is as in (2.49) and P is given by (2.54). Therefore

$$\hat{\gamma} = P_{12}'P_1^{-1}\hat{x}_1 \quad (2.62)$$

III. THE SMOOTHING UPDATE PROBLEM

Consider the formulation described in the preceding subsection, but with the following additional features: (1) we observe the measurements over the time interval $[0,T]$; (2) the data are processed locally to produce the smoothed estimates

$$\hat{x}_{is}(t) = E[x_i(t) | y_i(\tau), 0 \leq \tau \leq T] \quad (3.1)$$

based on their local models; and (3) we wish to compute the overall smoothed estimate

$$\hat{x}_s(t) = E[x(t) | y_1(\tau), y_2(\tau), 0 \leq \tau \leq T] \quad (3.2)$$

using only the smoothed estimate histories \hat{x}_{1s} and \hat{x}_{2s} . A second, very closely related problem is that of computing \hat{x}_s in terms of \hat{x}_{1s} and y_2 . As we discussed for the filtering problem at the end of Section 2.1, the solution to the first of the smoothing problems will also provide us with a solution for the latter. Therefore we will do our analysis for the first problem and will comment on the second problem afterwards.

The motivation for these questions comes from problems in map updating. Suppose that, as illustrated in Figure 3.1, we are interested in the estimation of a two-dimensional random field given data obtained from parallel tracks. Problems of this type arise in the mapping of gravitational anomalies given data obtained along tracks over the ocean [2,14] and the mapping of meteorological variables from data gathered by satellites [1]. In this case we can think of x_i as representing the variables to be mapped along the i th set of tracks or over some portion of the field including these tracks. The global state x then represents the field along all of the tracks or regions surrounding each of the tracks. Note that our model allows repeated or overlapping surveys ($x_i = x_j$ for repeated surveys, while some components of x_i are the same as components of x_j if the surveys overlap).

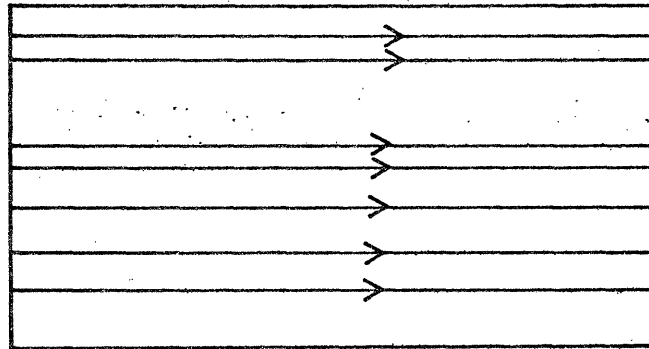


FIGURE 3.1: Parallel Data Tracks Across a Two-Dimensional Random Field

As in the preceding section we will assume dynamical models for x and x_i along the direction of the tracks. This can be done if the underlying field has a particular characterization. Specifically, let $f(t,s)$ denote the two-dimensional random field, which we assume to be Gaussian and, for simplicity, zero mean. Here t is the direction in which the tracks are taken -- i.e. we observe the field over lines of the form $\{(t,s_i) | 0 \leq t \leq T\}$ for several values of s_i . What we require is that the set of processes $f(t,s_i)$ jointly have a finite-dimensional shaping filter representation. Define the 2-D correlation function

$$R(t,\tau;s,\sigma) = E[f(t,s)f'(\tau,\sigma)] \quad (3.1)$$

As in 1-D, this function has a certain symmetry property. Using (3.1) it is readily seen that

$$R(t,\tau;s,\sigma) = R(\tau,t;\sigma,s)' \quad (3.2)$$

Thus, if we specify $R(t,\tau;s,\sigma)$ for all 4-triples (t,τ,s,σ) with $t \geq \tau$, we will have completely specified it. Other properties of f can be reflected in properties of R . For example if f is stationary, then

$$R(t,\tau;s,\sigma) = R(t-\tau,s-\sigma) \quad (3.3)$$

and if we specify $R(\tau,\sigma)$ for $\tau \geq 0$ and σ arbitrary, we will have completely specified it.

The requirement that any set of tracks $f(t,s_i)$ be realizable as the output of a finite dimensional shaping filter places additional restrictions

on R . One important case in which this is true is when R is separable

$$R(t, \tau; s, \sigma) = R_1(t, \tau) R_2(s, \sigma) \quad (3.4)$$

with R_1 and R_2 square,

$$R_1(t, \tau) = R_1^T(\tau, t) \quad , \quad R_2(s, \sigma) = R_2^T(\sigma, s) \quad (3.5)$$

and R_1 itself being separable

$$R_1(t, \tau) = H(t)G(\tau) \quad , \quad t \geq \tau \quad (3.6)$$

It is not difficult to see that (3.2), (3.4), (3.5) imply that R_1 and R_2 commute for any values of their arguments. The case (3.4)-(3.6) is a slight generalization of classes of processes considered by others. For example, if we had further assumed a separability condition for R_2 as in (3.6) and made the field stationary (so that (3.6) becomes $R_1(\tau) = H e^{F\tau} G, t \geq 0$), we would have the continuous-space version of the model considered by Attasi.

While the restriction of finite dimensionality and the scenario of Figure 3.1 are quite special, this problem is of interest in the applications cited earlier, and it opens up some interesting technical problems which we will discuss and solve. Furthermore, we feel that our results do shed some light on the issues involved in assimilating spatially-distributed data and combining regionally-processed information and as such represent a modest first step towards the goal of solving less restricted versions of these problems.

Recall that in the preceding section we solved the causal -- i.e. filtering -- version of problems of combining and updating estimates. We found that a solution existed with effectively no restrictions on the relationship between the local and global models (except the existence of M_1 in (2.14)). In this section we are interested in the noncausal versions of this problem. As we will see, the updating problem always has a solution, while the combining problem can be solved only when some further restriction, which also is not particularly severe, is placed on the local models. In the next subsection we will develop the basic ideas behind our approach and will point out where the difficulty arises. In the following two subsections we will address the two special cases considered in Subsections 2.2 and 2.3, which are the most important for random field mapping, and we will see that the difficulty can be overcome in these cases.

3.1 The General Case

The starting point for our analysis is the two-filter form for the optimal smoother. In particular, we will follow the approach described in [4]. In this approach the smoothed estimate is a weighted combination of a forward estimate, produced by the usual Kalman filter, and a reversed estimate, produced by a Kalman filter based on a reversed-time Markov model. This approach has the advantage of not involving infinite initial error covariances. For all of this we assume that $x(0)$ is zero mean and that the local

processor filters (forward and reverse) are initialized at zero. Nonzero initial conditions can easily be accommodated, because of linearity, and one should then think of all of the variables in our formulation as describing deviations from the a priori mean.

Let us summarize the smoother equations for each of the two local processors. For ease of reference, all of the relevant equations are collected here. The forward estimator for processor i ($i=1,2$) is given by

$$\dot{\hat{x}}_{if} = [A_i - P_{if} H_i' R_i^{-1} H_i] \hat{x}_{if} + P_{if} H_i' R_i^{-1} y_i \quad (3.3)$$

where P_{if} can be precomputed from either of the equations

$$\dot{P}_{if} = A_i P_{if} + P_{if} A_i' + Q_i - P_{if} H_i' R_i^{-1} H_i P_{if} \quad (3.4)$$

$$\frac{d}{dt} (P_{if}^{-1}) = -P_{if}^{-1} A_i - A_i' P_{if}^{-1} - P_{if}^{-1} Q_i P_{if}^{-1} + H_i' R_i^{-1} H_i \quad (3.5)$$

Note that these equations are essentially the same as (2.10)-(2.12).

The reverse time estimator involves the unconditional covariance for the local model assumed by the processor, which can be calculated from

$$\dot{\Sigma}_i = A_i \Sigma_i + \Sigma_i A_i' + Q_i \quad (3.6)$$

or

$$\frac{d}{dt} (\Sigma_i^{-1}) = -\Sigma_i^{-1} A_i - A_i' \Sigma_i^{-1} - \Sigma_i^{-1} Q_i \Sigma_i^{-1} \quad (3.7)$$

The reverse-time estimator operates backward in time from $t=T$ and is given by

$$\dot{\hat{x}}_{ir} = [-A_i - Q_i \sum_i^{-1} - P_{ir} H_i' R_i^{-1} H_i] \hat{x}_{ir} + P_{ir} H_i' R_i^{-1} y_i \quad (3.8)$$

which is a backward Kalman filter, with covariance also calculated backward in time (with initial condition $P_{ir}(T) = \sum_i(T)$) from either of the following equations

$$\begin{aligned} -\dot{P}_{ir} = & -[A_i + Q_i \sum_i^{-1}] P_{ir} - P_{ir} [A_i + Q_i \sum_i^{-1}]' \\ & + Q_i - P_{ir} H_i' R_i^{-1} H_i P_{ir} \end{aligned} \quad (3.9)$$

$$\begin{aligned} -\frac{d}{dt}(P_{ir}^{-1}) = & P_{ir}^{-1} [A_i + Q_i \sum_i^{-1}] + [A_i + Q_i \sum_i^{-1}]' P_{ir}^{-1} \\ & - P_{ir}^{-1} Q_i P_{ir}^{-1} + H_i' R_i^{-1} H_i \end{aligned} \quad (3.10)$$

The smoothed estimate \hat{x}_{is} is then given by

$$\hat{x}_{is} = P_{is}^{-1} [P_{if}^{-1} \hat{x}_{if} + P_{ir}^{-1} \hat{x}_{ir}] \quad (3.11)$$

where

$$P_{is}^{-1} = P_{if}^{-1} + P_{ir}^{-1} - \sum_i^{-1} \quad (3.12)$$

Note that (3.12) again reflects the fact that \hat{x}_{if} and \hat{x}_{ir} are not independent estimates, as they both utilize a priori information, in this case the information concerning $x(0)$. Here (3.12) holds for any value of Q_i .

The overall smoothed estimate satisfies a similar set of equations

$$\hat{x}_s = P_s [P_f^{-1} \hat{x}_f + P_r^{-1} \hat{x}_r] \quad (3.13)$$

$$P_s^{-1} = P_f^{-1} + P_r^{-1} - \bar{\Sigma}^{-1} \quad (3.14)$$

$$\dot{\bar{\Sigma}} = A\bar{\Sigma} + \bar{\Sigma}A' + Q \quad (3.15)$$

$$\frac{d}{dt} (\bar{\Sigma}^{-1}) = -\bar{\Sigma}^{-1}A-A'\bar{\Sigma}^{-1} - \bar{\Sigma}^{-1}Q\bar{\Sigma}^{-1} \quad (3.16)$$

$$\dot{P}_f = AP_f + P_fA' + Q - P_fC_1'R_1^{-1}C_1P_f - P_fC_2'R_2^{-1}C_2P_f \quad (3.17)$$

$$\frac{d}{dt} (P_f^{-1}) = -P_f^{-1}A-A'P_f^{-1} - P_f^{-1}QP_f^{-1} + C_1'R_1^{-1}C_1 + C_2'R_2^{-1}C_2 \quad (3.18)$$

$$\begin{aligned} \dot{P}_r = & -[A+Q]\bar{\Sigma}^{-1}P_r - P_r[A+Q]\bar{\Sigma}^{-1} + Q \\ & - P_rC_1'R_1^{-1}C_1P_r - P_rC_2'R_2^{-1}C_2P_r \end{aligned} \quad (3.19)$$

$$\begin{aligned} -\frac{d}{dt} (P_r^{-1}) = & P_r^{-1}[A+Q]\bar{\Sigma}^{-1} + [A+Q]\bar{\Sigma}^{-1}P_r^{-1} - P_r^{-1}QP_r^{-1} \\ & + C_1'R_1^{-1}C_1 + C_2'R_2^{-1}C_2 \end{aligned} \quad (3.20)$$

Using the results of the previous section we can calculate \hat{x}_f in terms of \hat{x}_{1f} and \hat{x}_{2f} , and, by looking at the problem in reverse time, we can use the same result to compute \hat{x}_r in terms of \hat{x}_{1r} and \hat{x}_{2r} . The resulting equations are

$$\hat{x}_f = \xi_f + G_{1f}\hat{x}_{1f} + G_{2f}\hat{x}_{2f} \quad (3.21)$$

$$\dot{\xi}_f = F_f \xi_f + K_{1f}\hat{x}_{1f} + K_{2f}\hat{x}_{2f} \quad (3.22)$$

where

$$F_f = A - P_f C_1' R_1^{-1} C_1 - P_f C_2' R_2^{-1} C_2 \quad (3.23)$$

$$G_{if} = P_f M_i' P_i^{-1} \quad (3.24)$$

$$\begin{aligned} K_{if} = & [P_f M_i' P_i^{-1} Q_i P_i^{-1} - Q M_i' P_i^{-1}] \\ & + [P_f M_i' A_i' P_i^{-1} - P_f A_i' M_i' P_i^{-1}] - P_f \dot{M}_i' P_i^{-1} \end{aligned} \quad (3.25)$$

and in reverse time

$$\hat{x}_r = \xi_r + G_{1r}\hat{x}_{1r} + G_{2r}\hat{x}_{2r} \quad (3.26)$$

$$-\dot{\xi}_r = F_r \xi_r + K_{1r}\hat{x}_{1r} + K_{2r}\hat{x}_{2r} \quad (3.27)$$

$$F_r = -A - Q \sum^{-1} - P_r C_1' R_1^{-1} C_1 - P_r C_2' R_2^{-1} C_2 \quad (3.28)$$

$$G_{ir} = P_r M_i' P_i^{-1} \quad (3.29)$$

$$\begin{aligned} K_{ir} = & [P_r M_i' P_i^{-1} Q_i P_i^{-1} - Q M_i' P_i^{-1}] \\ & + \{P_r M_i' [-A_i' - \sum_i^{-1} Q_i] P_i^{-1} - P_r [-A_i' - \sum_i^{-1} Q_i] M_i' P_i^{-1}\} + P_r \dot{M}_i' P_i^{-1} \end{aligned} \quad (3.30)$$

From (3.13), (3.21), (3.22), (3.26), and (3.27), we now have an algorithm for calculating \hat{x}_s from \hat{x}_{1f} , \hat{x}_{1r} , \hat{x}_{2f} , and \hat{x}_{2r} . What we would like is to compute \hat{x}_s in terms of \hat{x}_{1s} and \hat{x}_{2s} . To see when and how this

this can be done, we note first that from (3.11) and (3.24)

$$\begin{aligned} P_f^{-1} \hat{x}_f + P_r^{-1} \hat{x}_r &= P_f^{-1} \xi_f + M_1' P_{1f}^{-1} \hat{x}_{1f} + M_2' P_{2f}^{-1} \hat{x}_{2f} + P_r^{-1} \xi_r \\ &\quad + M_1' P_{1r}^{-1} \hat{x}_{1r} + M_2' P_{2r}^{-1} \hat{x}_{2r} \\ &= P_f^{-1} \xi_f + P_r^{-1} \xi_r + M_1' P_{1s}^{-1} \hat{x}_{1s} + M_2' P_{2s}^{-1} \hat{x}_{2s} \end{aligned}$$

Thus

$$\hat{x}_s = P_s [P_f^{-1} \xi_f + P_r^{-1} \xi_r + M_1' P_{1s}^{-1} \hat{x}_{1s} + M_2' P_{2s}^{-1} \hat{x}_{2s}] \quad (3.31)$$

The last two terms on the right-hand side represent the type of combination of estimates one would expect if the two sets of measurements had independent sources of error. However, as we have seen, they are correlated, and thus we have the correction terms to account for this correlation.

We have now reduced the algorithm for calculating \hat{x}_s to equations (3.31), (3.22), and (3.27). We have eliminated \hat{x}_{if} and \hat{x}_{ir} and replaced them with \hat{x}_{is} in (3.31), but (3.22) and (3.27) still involve the forward and reverse estimates. Our goal is to try to perform a combination of terms, as was done to obtain (3.31), in order to replace these estimates with \hat{x}_{is} . However, we cannot perform this in the same simple manner as was used earlier at least for the equations we have here. For example, (3.22) involves \hat{x}_{if} but not \hat{x}_{ir} , so we cannot combine terms to obtain \hat{x}_{is} . Rather, we have something that more closely resembles an inverse system problem: we want to express the term involving \hat{x}_{if} in (3.22)

by a team involving \hat{x}_{1s} . As we will see in the next few subsections this cannot always be done, but there are some very important cases in which it can be done. The most basic of these is considered in the following subsection.

3.2 The Special Case of Identical Local and Global Models

As in Subsection 2.2, consider the case when

$$\begin{aligned} A_1 = A_2 = A, \quad Q_1 = Q_2 = Q, \quad C_1 = H_1, \quad C_2 = H_2 \\ M_1 = M_2 = I \end{aligned} \quad (3.32)$$

This might correspond, for example, to two separate measurements along the same sets of one-dimensional tracks or of maps of the same region in the two dimensional field produced from measurements along two different sets of tracks.

For this case we obtain some simplification, as we did in the decentralized filtering problem. Here

$$\hat{x}_s = P_2 [P_f^{-1} \xi_f + P_r^{-1} \xi_r + P_{1s}^{-1} \hat{x}_{1s} + P_{2s}^{-1} \hat{x}_{2s}] \quad (3.33)$$

and

$$\dot{\xi}_f = F_f \xi_f + K_{1f} \hat{x}_{1f} + K_{2f} \hat{x}_{2f} \quad (3.34)$$

$$- \dot{\xi}_r = F_r \xi_r + K_{1r} \hat{x}_{1r} + K_{2r} \hat{x}_{2r} \quad (3.35)$$

where

$$K_{if} = [P_f P_{if}^{-1} - I] Q P_{if}^{-1} \quad (3.36)$$

$$K_{ir} = [P_r P_{ir}^{-1} - I] Q P_{ir}^{-1} \quad (3.37)$$

We now see that the quantities we actually need in (3.34) and (3.35) are

$QP_{if}^{-1}\hat{x}_{if}$ and $QP_{ir}^{-1}\hat{x}_{ir}$. In order to proceed, it is useful to define

$$\begin{aligned} z_{if} &= P_{if}^{-1}\hat{x}_{if} \\ z_{ir} &= P_{ir}^{-1}\hat{x}_{ir} \\ z_{is} &= P_{is}^{-1}\hat{x}_{is} \end{aligned} \tag{3.38}$$

Then, from (3.11)

$$z_{is} = z_{if} + z_{ir} \tag{3.39}$$

and, differentiating

$$\dot{z}_{if} = \left[\frac{d}{dt} (P_{if}^{-1}) \right] \hat{x}_{if} + P_{if}^{-1} \dot{\hat{x}}_{if} \tag{3.40}$$

$$\dot{z}_{ir} = \left[\frac{d}{dt} (P_{ir}^{-1}) \right] \hat{x}_{ir} + P_{ir}^{-1} \dot{\hat{x}}_{ir} \tag{3.41}$$

Substituting (3.3) and (3.5) into (3.40) and (3.8) and (3.10) into

(3.41) and performing some algebra, we obtain

$$\dot{z}_{if} = -(A' + P_{if}^{-1}Q) z_{if} + C_i' R_i^{-1} y_i \tag{3.42}$$

$$\dot{z}_{ir} = -(A' + \begin{bmatrix} -1 \\ Q - P_{ir}^{-1}Q \end{bmatrix}) z_{ir} - C_i' R_i^{-1} y_i \tag{3.43}$$

If we add these last two equations and use (3.39) we obtain two different equations for the time rate of change of z_{is} :

$$\dot{z}_{is} = -(A' + \sum^{-1} Q - P_{ir}^{-1} Q) z_{is} - P_{is}^{-1} Q z_{if} \quad (3.44)$$

$$\dot{z}_{is} = -(A' + P_{if}^{-1} Q) z_{is} + P_{is}^{-1} Q z_{ir} \quad (3.45)$$

Thus

$$Q z_{if} = P_{is} [-\dot{z}_{is} - (A' + \sum^{-1} Q - P_{ir}^{-1} Q) z_{is}] \quad (3.46)$$

$$Q z_{ir} = P_{is} [\dot{z}_{is} + (A' + P_{if}^{-1} Q) z_{is}] \quad (3.47)$$

From (3.34)-(3.38), we see that we can use (3.46), (3.47) in these equations to replace \hat{x}_{if} , \hat{x}_{ir} with \hat{x}_{is} . Thus in this case, we can obtain an algorithm of the desired form. Note that we haven't shown that we can recover \hat{x}_{if} , \hat{x}_{ir} from \hat{x}_{is} , or equivalently z_{if} , z_{ir} from z_{is} , but we have seen that we can recover $Q z_{if}$ and $Q z_{ir}$, and this is all that we need for our problem. Note, however, that the expressions (3.46), (3.47) for these quantities involve derivatives of z_{is} . In order to avoid these, we must use a feedforward formulation. First of all, substituting (3.36), (3.46) into (3.34) we obtain

$$\dot{\xi}_f = F_f \xi_f - \sum_{i=1}^2 [P_f P_{if}^{-1} - I] P_{is} [\dot{z}_{is} + (A' + \sum^{-1} Q - P_{ir}^{-1} Q) z_{is}] \quad (3.48)$$

From (3.38) we know that

$$\dot{\hat{x}}_{is} = \dot{P}_{is} P_{is}^{-1} \hat{x}_{is} + P_{is} \dot{z}_{is} \quad (3.49)$$

or

$$P_{is} \dot{z}_{is} = \dot{\hat{x}}_{is} - \dot{P}_{is} P_{is}^{-1} \hat{x}_{is} \quad (3.50)$$

Now define

$$q_f = \xi_f + \sum_{i=1}^2 [P_f P_{if}^{-1} - I] \hat{x}_{is} \quad (3.51)$$

Differentiating this, using (3.50), we obtain

$$\begin{aligned} \dot{q}_f &= F_f [q_f - \sum_{i=1}^2 (P_f P_{if}^{-1} - I) \hat{x}_{is}] + \sum_{i=1}^2 \left[\frac{d}{dt} (P_f P_{if}^{-1}) \right] \hat{x}_{is} \\ &\quad - \sum_{i=1}^2 [P_f P_{if}^{-1} - I] [-\dot{P}_{is} P_{is}^{-1} + P_{is} (A' + \sum_{r=1}^2 Q^{-1} P_{ir}^{-1} Q) P_{is}^{-1}] \hat{x}_{is} \end{aligned} \quad (3.52)$$

We have all of the equations needed to simplify this equation except for an expression for \dot{P}_{is} . From (3.12)

$$\frac{d}{dt} (P_{is}^{-1}) = \frac{d}{dt} (P_{if}^{-1}) + \frac{d}{dt} (P_{ir}^{-1}) - \frac{d}{dt} \left(\sum^{-1} \right) \quad (3.53)$$

and using expressions for the terms on the right-hand side (equations (3.5), (3.10) and (3.16) together with (3.32), we obtain

$$\frac{d}{dt} (P_{is}^{-1}) = -P_{is}^{-1} A - A' P_{is}^{-1} - P_{if}^{-1} Q P_{is}^{-1} - P_{is}^{-1} Q \sum^{-1} + P_{is}^{-1} Q P_{ir}^{-1} \quad (3.54)$$

A great deal of algebra then yields

$$\dot{q}_f = F_f q_f - P_f C_2' R_2^{-1} C_2 \hat{x}_{1s} - P_f C_1' R_1^{-1} C_1 \hat{x}_{2s} \quad (3.55)$$

Note that "2" -subscripted matrices multiply \hat{x}_{1s} and "1" -subscripted matrices multiply \hat{x}_{2s} .

We can follow exactly the same ideas for the reverse filter. Let

$$q_r = \xi_r + \sum_{i=1}^2 [P_r P_{ir}^{-1} - I] \hat{x}_{is} \quad (3.56)$$

Then

$$\dot{\hat{q}}_r = F_r \hat{q}_r - P_r C_2' R_2^{-1} C_2 \hat{x}_{1s} - P_r C_1' R_1^{-1} C_1 \hat{x}_{2s} \quad (3.57)$$

Substituting (3.51) and (3.56) into (3.33), we obtain

$$\begin{aligned} \hat{x}_s = P_2 \left\{ P_f^{-1} \hat{q}_f - \sum_{i=1}^2 [P_{if}^{-1} - P_f^{-1}] \hat{x}_{is} + P_r^{-1} \hat{q}_r \right. \\ \left. - \sum_{i=1}^2 [P_{ir}^{-1} - P_r^{-1}] \hat{x}_{is} + \sum_{i=1}^2 P_{is}^{-1} \hat{x}_{is} \right\} \end{aligned} \quad (3.58)$$

Using (3.12) and (3.14), we obtain the following algorithm for combining smoothed estimates

$$\hat{x}_s = P_s \{ P_f^{-1} \hat{q}_f + P_r^{-1} \hat{q}_r \} + \hat{x}_{1s} + \hat{x}_{2s} \quad (3.59)$$

$$\dot{\hat{q}}_f = F_f \hat{q}_f - P_f C_2' R_2^{-1} C_2 \hat{x}_{1s} - P_f C_1' R_1^{-1} C_1 \hat{x}_{2s} \quad (3.60)$$

$$\dot{\hat{q}}_r = F_r \hat{q}_r - P_r C_2' R_2^{-1} C_2 \hat{x}_{1s} - P_r C_1' R_1^{-1} C_1 \hat{x}_{2s} \quad (3.61)$$

If we think of optimal estimates as orthogonal projections in spaces of random variables, then \hat{x}_{1s} is the projection of x onto V_1 , the subspace spanned by the first pass measurements. Similarly \hat{x}_{2s} is the projection onto V_2 , and \hat{x}_s is the projection onto $V_1 + V_2$. If V_1 and V_2 were orthogonal, i.e., independent, then \hat{x}_s would equal $\hat{x}_{1s} + \hat{x}_{2s}$. However, they are not, and thus the other terms in (3.59) account for this.

We can actually see this point more clearly if we look at the smoothing update problem, that is, the problem of computing \hat{x}_s in terms of the

time history of the old smoothed estimate \hat{x}_{1s} and the new data y_2 . The solution to this problem is readily obtained in a manner analogous to that used in deriving (2.32). That is, if we perform all of the analysis we have just done, leaving y_2 alone and only replacing y_1 by \hat{x}_{1f} , \hat{x}_{1r} , and eventually by \hat{x}_{1s} , linearity guarantees that the input-output relation from \hat{x}_{1s} to \hat{x}_s is the same as that obtained already. Thus, all the work we need to do is already done, and we can simply write down the solution to the updating problem:

$$\hat{x}_s = P_s [P_f^{-1} r_f + P_r^{-1} r_r] + \hat{x}_{1s} \quad (3.62)$$

$$\dot{r}_f = F_f r_f + P_f C_2' R_2^{-1} y_2 - P_f C_2' R_2^{-1} C_2 \hat{x}_{1s} \quad (3.63)$$

$$-\dot{r}_r = F_r r_r + P_r C_2' R_2^{-1} y_2 - P_r C_2' R_2^{-1} \hat{x}_{1s} \quad (3.64)$$

Here, if we let F denote the orthogonal complement of y_1 in $y_1 + y_2$ -- i.e. the part of y_2 that is independent of y_1 , then $y_1 + y_2 = y_1 \oplus F$, and \hat{x}_{1s} is the projection of x onto y_1 , while the remaining terms in (3.62) are the projection onto F .

Note also that (3.63), (3.64) can be rewritten in the following form

$$\dot{r}_f = F_f r_f + P_f C_2' R_2^{-1} [y_2 - C_2 \hat{x}_{1s}] \quad (3.65)$$

$$-\dot{r}_r = F_r r_r + P_r C_2' R_2^{-1} [y_2 - C_2 \hat{x}_{1s}] \quad (3.66)$$

Thus these correction terms, which provide the projection onto the new information in y_2 are driven by the difference between what we observe and what we expect to observe based on our first map. Another interpretation is that these corrections terms are the projection of the estimation error $(x - \hat{x}_{1s})$ onto y_2 . What we have done with a great deal of algebra is to obtain realizations of these projections in terms of finite-dimensional forward and reverse filters.

Since they may be useful, let us note two other forms for the solution to the problem. Let

$$\eta_f = P_f^{-1} q_f, \quad \gamma_f = P_f^{-1} r_f \quad (3.67)$$

$$\eta_r = P_r^{-1} q_r, \quad \gamma_r = P_r^{-1} r_r \quad (3.68)$$

Then the equations for the combining of smoothed estimates become

$$\hat{x}_s = P_s (\eta_f + \eta_r) + \hat{x}_{1s} + \hat{x}_{2s} \quad (3.69)$$

$$\dot{\eta}_f = -(A' + P_f^{-1} Q) \eta_f - C_2' R_2^{-1} C_2 \hat{x}_{1s} - C_1' R_1^{-1} C_1 \hat{x}_{2s} \quad (3.70)$$

$$-\dot{\eta}_r = (A' + \bar{L}^{-1} Q - P_r^{-1} Q) \eta_r - C_2' R_2^{-1} C_2 \hat{x}_{1s} - C_1' R_1^{-1} C_1 \hat{x}_{2s} \quad (3.71)$$

and the corresponding updating equations are

$$\hat{x}_s = P_s (\gamma_f + \gamma_r) + \hat{x}_{1s} \quad (3.72)$$

$$\dot{\gamma}_f = -(A' + P_f^{-1} Q) \gamma_f + C_2' R_2^{-1} (y_2 - C_2 \hat{x}_{1s}) \quad (3.73)$$

$$-\dot{\gamma}_r = (A' + \bar{L}^{-1} Q - P_r^{-1} Q) \gamma_r + C_2' R_2^{-1} (y_2 - C_2 \hat{x}_{1s}) \quad (3.74)$$

If instead we use

$$\alpha_f = P_s \eta_f, \quad \beta_f = P_s \gamma_f \quad (3.75)$$

$$\alpha_r = P_s \eta_r, \quad \beta_r = P_s \gamma_r \quad (3.76)$$

we obtain

$$\hat{x}_s = \alpha_f + \alpha_r + \hat{x}_{1s} + \hat{x}_{2s} \quad (3.77)$$

$$\dot{\alpha}_f = (A+Q)^{-1}(-QP_r^{-1})\alpha_f - P_s C_2' R_2^{-1} C_2 \hat{x}_{1s} - P_s C_1' R_1^{-1} C_1 \hat{x}_{2s} \quad (3.78)$$

$$-\dot{\alpha}_r = -(A+QP_f^{-1})\alpha_r - P_s C_2' R_2^{-1} C_2 \hat{x}_{1s} - P_s C_1' R_1^{-1} C_1 \hat{x}_{2s} \quad (3.79)$$

and

$$\hat{x}_s = \beta_f + \beta_r + \hat{x}_{1s} \quad (3.80)$$

$$\dot{\beta}_f = (A+Q)^{-1}(-QP_r^{-1})\beta_f + P_s C_2' R_2^{-1} (y_2 - C_2 \hat{x}_{1s}) \quad (3.81)$$

$$-\dot{\beta}_r = -(A+QP_f^{-1})\beta_r + P_s C_2' R_2^{-1} (y_2 - C_2 \hat{x}_{1s}) \quad (3.82)$$

Note the striking symmetry between the equations using the η 's and γ 's and those using the α 's and β 's.

3.3 Conditions for Existence of a Solution to the General Case and an Important Special Case

In the preceding subsection we saw that the smoothed estimate updating problem could be solved when the local and global models are identical. In this section we look at the problem when this is not the

case. Recall that the general algorithm had been reduced to (3.22), (3.27), and (3.31), together with (3.24) and (3.29) which define the gains needed in the algorithm. Also, we have equations for z_{if} , z_{ir} , and z_{is} as before, but with slight modifications due to the fact that we have local models. Specifically, we have

$$z_{is} = z_{if} + z_{ir} \quad (3.83)$$

$$\dot{z}_{if} = -(A'_i + P_{if}^{-1} Q_i) z_{if} + H'_i R_i^{-1} y_i \quad (3.84)$$

$$\dot{z}_{ir} = -(A'_i + \sum_i^{-1} Q_i - P_{ir}^{-1} Q_i) z_{ir} - H'_i R_i^{-1} y_i \quad (3.85)$$

which lead to the equations

$$Q_i z_{if} = P_{is} [-\dot{z}_{is} - (A'_i + \sum_i^{-1} Q_i - P_{ir}^{-1} Q_i) z_{is}] \quad (3.86)$$

$$Q_i z_{ir} = P_{is} [\dot{z}_{is} + (A'_i + P_{if}^{-1} Q_i) z_{is}] \quad (3.87)$$

In order to proceed as we did in the preceding section we wish to be able to find some matrices L_{if} , L_{ir} so that

$$K_{if} \hat{x}_{if} = L_{if} Q_i z_{if} = L_{if} Q_i P_{if}^{-1} \hat{x}_{if} \quad (3.88)$$

$$K_{ir} \hat{x}_{ir} = L_{ir} Q_i z_{ir} = L_{ir} Q_i P_{ir}^{-1} \hat{x}_{ir} \quad (3.89)$$

This will be possible if and only if

$$N(Q_i P_{if}^{-1}) \subseteq N(K_{if}) \quad (3.90a)$$

$$N(Q_i P_{ir}^{-1}) \subseteq N(K_{ir}) \quad (3.90b)$$

Note that if (3.90a) and (3.90b) hold, then L_{if} and L_{ir} can be chosen to be

$$L_{if} = K_{if} P_{if} Q_i^\dagger$$

$$L_{ir} = K_{ir} P_{ir} Q_i^\dagger$$

In this case, we can substitute in for $Q_i z_{if}$ and $Q_i z_{ir}$ and $Q_i z_{is}$ using (3.86) and (3.87) and then use a set of steps similar to those used in Subsection 3.2 to remove the \dot{z}_{is} -term. Specifically, substituting (3.86)-(3.89) into (3.22) and (3.27) we obtain

$$\dot{\xi}_f = F_f \xi_f - \sum_{i=1}^2 L_{if} P_{is} [\dot{z}_{is} + (A_i' + \sum_i^{-1} Q_i - P_{ir}^{-1} Q_i) z_{is}]$$

$$-\dot{\xi}_r = F_r \xi_r + \sum_{i=1}^2 L_{ir} P_{is} [\dot{z}_{is} + (A_i' + P_{if}^{-1} Q_i) z_{is}]$$

Then, using (3.38) and (3.50),

$$\dot{\xi}_f = F_f \xi_f + \sum_{i=1}^2 L_{if} [P_{is} P_{is}^{-1} \hat{x}_{is} - \dot{\hat{x}}_{is} - P_{is} (A_i' + \sum_i^{-1} Q_i - P_{ir}^{-1} Q_i) P_{is}^{-1} \hat{x}_{is}]$$

$$-\dot{\xi}_r = F_r \xi_r + \sum_{i=1}^2 L_{ir} [\dot{\hat{x}}_{is} - P_{is} P_{is}^{-1} \hat{x}_{is} + P_{is} (A_i' + P_{if}^{-1} Q_i) P_{is}^{-1} \hat{x}_{is}]$$

Defining

$$q_f = \xi_f + \sum_{i=1}^2 L_{if} \hat{x}_{is}$$

$$q_r = \xi_r + \sum_{i=1}^2 L_{ir} \hat{x}_{is}$$

we find that

$$\dot{\mathbf{q}}_f = \mathbf{F}_f \mathbf{q}_f + \mathbf{N}_{1f} \hat{\mathbf{x}}_{1s} + \mathbf{N}_{2f} \hat{\mathbf{x}}_{2s} \quad (3.91a)$$

$$-\dot{\mathbf{q}}_r = \mathbf{F}_r \mathbf{q}_r + \mathbf{N}_{1r} \hat{\mathbf{x}}_{1s} + \mathbf{N}_{2r} \hat{\mathbf{x}}_{2s} \quad (3.91b)$$

$$\hat{\mathbf{x}}_s = \mathbf{P}_s [\mathbf{P}_f^{-1} \mathbf{q}_f + \mathbf{P}_r^{-1} \mathbf{q}_r + \mathbf{D}_1 \hat{\mathbf{x}}_{1s} + \mathbf{D}_2 \hat{\mathbf{x}}_{2s}] \quad (3.91c)$$

where

$$\mathbf{N}_{if} = -\mathbf{F}_f \mathbf{L}_{if} + \dot{\mathbf{L}}_{if} + \mathbf{L}_{if} \dot{\mathbf{P}}_{is} \mathbf{P}_{is}^{-1} - \mathbf{L}_{if} \mathbf{P}_{is} (\mathbf{A}'_i + \mathbf{L}_i^{-1} \mathbf{Q}_i - \mathbf{P}_{ir}^{-1} \mathbf{Q}_i) \mathbf{P}_{is}^{-1} \quad (3.92a)$$

$$\mathbf{N}_{ir} = -\mathbf{F}_r \mathbf{L}_{ir} - \dot{\mathbf{L}}_{ir} - \mathbf{L}_{ir} \dot{\mathbf{P}}_{is} \mathbf{P}_{is}^{-1} + \mathbf{L}_{ir} \mathbf{P}_{is} (\mathbf{A}'_i + \mathbf{P}_{if}^{-1} \mathbf{Q}_i) \mathbf{P}_{is}^{-1} \quad (3.92b)$$

$$\mathbf{D}_i = \mathbf{M}'_i \mathbf{P}_{is}^{-1} - \mathbf{P}_f^{-1} \mathbf{L}_{if} - \mathbf{P}_r^{-1} \mathbf{L}_{ir} \quad (3.93)$$

The analysis in this case is clearly more complex, since the equations involve \mathbf{M}_i and \mathbf{Q}_i , and if these are time-varying, we will also have to consider their time derivatives.

Note that one obvious case in which (3.90a) and (3.90b) hold is when \mathbf{Q}_i is invertible. In this case the smoother is, in fact, invertible, as we can recover \mathbf{z}_{if} and \mathbf{z}_{ir} and consequently the original data \mathbf{y}_i , since the forward and reverse Kalman filters are always invertible. In the remainder of this section we wish to consider one other important special case from which we can gain more insight into the nature of our solution.

Specifically we wish to consider the case in which \mathbf{x}_1 is an actual part of the global state. As was discussed in Subsection 2.3, in this

case (assuming that M_1 is onto) we can choose a (possibly time-varying) basis for the state space so that we can identify x_1 with some of the components of x

$$x = \begin{pmatrix} x_1 \\ \gamma \end{pmatrix} \quad (3.94)$$

$$\begin{pmatrix} \dot{x}_1 \\ \gamma \end{pmatrix} = \begin{pmatrix} A_1 & 0 \\ A_{21} & A_{22} \end{pmatrix} \begin{pmatrix} x_1 \\ \gamma \end{pmatrix} + w \quad (3.95)$$

$$Q = \begin{pmatrix} Q_1 & Q_{12} \\ Q'_{12} & Q_{22} \end{pmatrix} \quad (3.96)$$

$$y_1 = (H_1 : 0) \begin{pmatrix} x_1 \\ \gamma \end{pmatrix} + v_1 \quad (3.97)$$

Also, using (2.57), equation (3.25) becomes

$$K_{1f} = \begin{pmatrix} [(P_f)_{11} P_{1f}^{-1} - I] Q_1 P_{1f}^{-1} \\ [(P_f)'_{12} P_{1f}^{-1} Q_1 - Q'_{12}] P_{1f}^{-1} \end{pmatrix} \quad (3.98)$$

where

$$P_f = \begin{pmatrix} (P_f)_{11} & (P_f)_{12} \\ (P_f)'_{12} & (P_f)_{22} \end{pmatrix} \quad (3.99)$$

If we write

$$w = \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} \quad (3.100)$$

then

$$E[w_1(t)w_1'(\tau)] = Q_1 \delta(t-\tau) \quad (3.101)$$

$$E[w_2(t)w_1'(\tau)] = Q_{12}' \delta(t-\tau) \quad (3.102)$$

From this it is relatively easy to see that there must be a matrix T_1

so that

$$Q_{12}' = T_1 Q_1 \quad (3.103)$$

Specifically, if we write w_2 as $w_{21} + w_{22}$, where w_{21} is the best estimate of w_2 given w_1 , then w_{21} is a linear function of w_1 , say $w_{21} = T_1 w_1$, and $E[w_{22}(t)w_1'(\tau)] = 0$. This, together with (3.101) and (3.102) yields (3.103). If we substitute (3.103) into (3.98), we find that

$$K_{1f} = \begin{pmatrix} (P_f)_{11} P_{1f}^{-1} - I \\ (P_f)'_{12} P_{1f}^{-1} - T_1 \end{pmatrix} Q_1 P_{1f}^{-1} \quad (3.104)$$

Therefore,

$$L_{1f} = \begin{pmatrix} (P_f)_{11} P_{1f}^{-1} - I \\ (P_f)'_{12} P_{1f}^{-1} - T_1 \end{pmatrix} = P_f \begin{pmatrix} P_{1f}^{-1} \\ 0 \end{pmatrix} - \begin{pmatrix} I \\ T_1 \end{pmatrix} \quad (3.105)$$

We can perform a similar analysis in reverse time, but the situation is a bit more complex. We will comment on the reasons for this complication shortly, but first we will present the solution. For the special case described by (3.94)-(3.96), equation (3.30) reduces to

$$K_{ir} = [P_r M_1' P_{1r}^{-1} M_1 - I] Q M_1' P_{1r}^{-1} + P_r [\sum^{-1} Q M_1' - M_1' \sum_1^{-1} Q_1] P_{1r}^{-1} \quad (3.106)$$

where $M_1 = [I : 0]$, Q is given by (3.96), and \sum and \sum_i are related by

$$\sum = \begin{bmatrix} \sum_1 & \sum_{12} \\ \sum_{12}' & \sum_{22} \end{bmatrix} \quad (3.107)$$

and, using a basic formula for matrix inverses for block matrices (see, for example, [5, p.495])

$$\sum^{-1} = \begin{bmatrix} \sum_1^{-1} + \sum_1^{-1} \sum_{12} \left[\sum_{22} - \sum_{12}' \sum_1^{-1} \sum_{12} \right]^{-1} \sum_{12}' \sum_1^{-1} & - \sum_1^{-1} \sum_{12} \left[\sum_{22} - \sum_{12}' \sum_1^{-1} \sum_{12} \right]^{-1} \\ - \left[\sum_{22} - \sum_{12}' \sum_1^{-1} \sum_{12} \right]^{-1} \sum_{12}' \sum_1^{-1} & \left[\sum_{22} - \sum_{12}' \sum_1^{-1} \sum_{12} \right]^{-1} \end{bmatrix} \quad (3.108)$$

Using these relationships and also (3.103), equation (3.106) becomes

$$K_{lr} = \begin{pmatrix} [(P_r)_{11} P_{1r}^{-1} - I] Q_1 P_{1r}^{-1} \\ [(P_r)_{12} P_{1r}^{-1} - T_1] Q_1 P_{1r}^{-1} \end{pmatrix} + P_r \begin{pmatrix} \sum_1^{-1} \sum_{12} \left[\sum_{22} - \sum_{12}' \sum_1^{-1} \sum_{12} \right]^{-1} \\ - \left[\sum_{22} - \sum_{12}' \sum_1^{-1} \sum_{12} \right]^{-1} \end{pmatrix} \left[\sum_{12}' \sum_1^{-1} - T_1 \right] Q_1 P_{1r}^{-1} \quad (3.109)$$

Therefore

$$\begin{aligned}
 L_{1r} &= \begin{bmatrix} (P_r)_{11} P_{1r}^{-1} - I \\ (P_r)'_{12} P_{1r}^{-1} - T_1 \end{bmatrix} + \\
 &\quad \begin{pmatrix} (P_r)_{11} \Sigma_1^{-1} \Sigma_{12} - (P_r)_{12} \\ (P_r)'_{12} \Sigma_1^{-1} \Sigma_{12} - (P_r)_{22} \end{pmatrix} \left[\Sigma_{22} - \Sigma_{12}' \Sigma_1^{-1} \Sigma_{12} \right]^{-1} \left[\Sigma_{12}' \Sigma_1^{-1} - T_1 \right] \\
 &= P_r \begin{pmatrix} P_{1r}^{-1} \\ 0 \end{pmatrix} - \begin{pmatrix} I \\ T_1 \end{pmatrix} + P_r \left[\Sigma^{-1} \begin{pmatrix} I \\ T_1 \end{pmatrix} - \begin{pmatrix} \Sigma_1^{-1} \\ 0 \end{pmatrix} \right] \quad (3.110)
 \end{aligned}$$

Comparing (3.109) and (3.110) with (3.104) and (3.105), we see that the first terms on the right-hand sides of (3.109) and (3.110) are analogous to the right-hand sides of (3.104) and (3.105), respectively. The additional terms in (3.109) and (3.110) represent the complication that arises in constructing reverse models for x_1 by itself (as used in the calculation of \hat{x}_{1r} and \hat{x}_{1s}) and a reverse model for $x = (x_1', \gamma')'$, which is used in computing \hat{x}_r and \hat{x}_s . Since both x_1 and x are Markov, we can in fact obtain reverse-time diffusion models for each of these. Following reference [6],

$$-\dot{x}_1(t) = -(A_1 + Q_1 \Sigma_1^{-1}) x_1(t) - \tilde{w}_1(t) \quad (3.111)$$

$$-\begin{pmatrix} \dot{x}_1(t) \\ \dot{\gamma}(t) \end{pmatrix} = -\left\{ \begin{pmatrix} A_1 & 0 \\ A_{12} & A_{22} \end{pmatrix} + Q \Sigma^{-1} \right\} \begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix} - \tilde{w}(t) \quad (3.112)$$

where these models can be interpreted as generating the same sample paths as the forward model. Here $\tilde{w}_1(t)$ is a white noise process with strength Q_1 and it represents that part of $w_1(t)$ that is independent of the future of x_1 , i.e. $x_1(s)$, $s > t$. Similarly, $\tilde{w}(t)$ is a white noise process with strength Q , representing the part of $w(t)$ that is independent of $x_1(s)$, $s > t$ and $\gamma(s)$, $s > t$. Because of this difference in the two models,

$$\tilde{\tilde{w}}_1(t) \neq \tilde{w}_1(t) \quad (3.113)$$

An equivalent way of looking at this is to view (3.111) as defining (with $\tilde{w}_1=0$) an equation for the best "predictor", going in reverse time of x_1 given the future of x_1 . Similarly (3.112) gives the best predictor of x_1 and γ given the future of x_1 and γ . Now although going forward in time the future of x_1 is decoupled from the past of γ , the future of γ does depend on the past of x_1 (see equation (3.95): $A_{12}=0$, but A_{21} need not be zero). Therefore, if we want to predict the past of x_1 , the future of γ does provide us with information (for example, the future history of position does help us deduce something about the past behavior of velocity). For this reason, although the (1,2) block of the forward-time dynamics matrix in (3.95) is zero, the (1,2) block of the reverse-time dynamics matrix in (3.112) is not zero. What this implies is that in reverse time, x_1 does not represent the state of a subsystem of the global state, and the extra terms in (3.109) and (3.110) reflect this fact.

From (3.103), (3.108) and (3.112) we can evaluate the (1,2) term of the reverse dynamics matrix:

$$Q_1 \left[\sum_1^{-1} \sum_{12} + T_1' \right] \left[\sum_{22} - \sum_{12}' \sum_1^{-1} \sum_{12} \right]^{-1} \quad (3.114)$$

Comparing (3.109) and (3.114), we see that (assuming the invertibility of P_r , \sum_1 , and \sum) the last term in K_{1r} will be zero if only if (3.114) is zero, that is, when x_1 is the state of a subsystem both in forward and reverse time, and this will be true if and only if

$$\left(\sum_{12}' \sum_1^{-1} - T_1 \right) Q_1 = 0 \quad (3.115)$$

or, equivalently

$$\sum_{12}' \sum_1^{-1} Q_1 - Q_{12} = 0 \quad (3.116)$$

It is relatively easy to see that this is the case if $A_{21}=Q_{12}=0$, since in that case x_1 and γ are independent. Other, essentially equally trivial cases can be found in which (3.116) is satisfied, but the condition is quite restrictive.

These observation notwithstanding, (3.105) and (3.110) allow us to replace \hat{x}_{1f} and \hat{x}_{1r} in (3.22) and (3.27) by expressions involving \hat{x}_{1s} through the use of (3.91)-(3.93). A similar analysis can be performed for local processor #2 if x_2 is the state of a subsystem of the global

system (of course the change of basis needed on the global state space to put the system into a form analogous to (3.95) will in general be different). Also, we can consider the case of map updating in which we wish to compute \hat{x}_s from \hat{x}_{ls} and y_2 . These results follow in much the same manner as those derived in Section 3.2:

$$\hat{x}_s = P_2 [P_f^{-1} r_f + P_r^{-1} r_r + D_1 \hat{x}_{ls}] \quad (3.117)$$

$$\dot{r}_f = F_f r_f + N_{lf} \hat{x}_{ls} + P_f C_2' R_2^{-1} y_2 \quad (3.118)$$

$$-\dot{r}_r = F_r r_r + N_{lr} \hat{x}_{ls} + P_r C_2' R_2^{-1} y_2 \quad (3.119)$$

where D_1 , N_{lf} , and N_{lr} are defined in (3.92) and (3.93). These equations hold whenever L_{lf} and L_{lr} exist. For example, if x_1 is the state of a subsystem (forward in time), then (assuming that a basis has been chosen as in (3.95)) (3.92) and (3.93) are computed using $M_1 = [I; 0]$ and L_{lf} and L_{lr} defined in (3.105) and (3.110). In this case, some algebra yields

$$P_s D_1 = \begin{pmatrix} I \\ T_1 \end{pmatrix} \quad (3.120)$$

$$N_{lf} = -P_f C_2' R_2^{-1} C_2 \begin{pmatrix} I \\ T_1 \end{pmatrix} - \begin{pmatrix} 0 \\ \dot{T}_1 \end{pmatrix} + \begin{pmatrix} 0 \\ A_{21} + A_{22} T_1 - T_1 A_1 \end{pmatrix} \quad (3.121)$$

$$N_{lr} = -P_r C_2' R_2^{-1} C_2 \begin{pmatrix} I \\ T_1 \end{pmatrix} + (P_r \sum^{-1} - I) \left[\begin{pmatrix} 0 \\ \dot{T}_1 \end{pmatrix} + \begin{pmatrix} 0 \\ A_{21} + A_{22} T_1 - T_1 A_1 \end{pmatrix} \right] \quad (3.122)$$

IV. REAL-TIME SMOOTHING

A variant of the problem addressed in the preceding section is the real-time smoothing problem. In this case, from some previous pass we have observed $y_1(t)$ over the interval $[0, T]$ and have produced $\hat{x}_{1s}(t)$. We now observe y_2 up to time t , and we wish to compute the real-time smoothed estimate of the global state, i.e.

$$\hat{x}_{rs}(t) = E[x(t) | y_1(\tau), 0 \leq \tau \leq T, y_2(\sigma), 0 \leq \sigma \leq t] \quad (4.1)$$

in terms of \hat{x}_{1s} and y_2 . This formulation is motivated by problems in which a second traversing of a track across a random field is taken in real time and we wish to process the data as we get it. If $x_1 = x$, then the tracks are identical. If x_1 is the state of a subsystem of the global system, i.e. $x' = (x_1', \gamma')$, then there are two possible motivations. The first is that in which x represents several tracks across the field and y_2 may be data from one of these other tracks. Alternatively, γ may represent the state of a dynamic system which is affected by the field, modeled by x_1 , during the second pass. For example x_1 might represent anomalies in the earth's gravitational field and γ could represent errors induced in an inertial navigation system aboard a ship [2, 9]. In this case we want the (real-time) estimates of γ . Clearly we can also model in this same way the case in which γ contains two pieces, one of which models additional tracks and the other models the state of a dynamic system affected by the random field along the second track.

The solution to this problem can be obtained directly from the results in the preceding section. Specifically, at any time t we can view

(4.1) as performing two full passes over $[0, T]$, but at any time t we assume that

$$C_2(s) = H_2(s) = 0 \quad t < s \leq T \quad (4.2)$$

Also, since we are attempting to compute \hat{x} directly using y_2 , we have essentially a smoothing update problem. Based on these observations (3.31), (3.22), and (3.27) can be adapted to the present situation:

$$\hat{x}_{rs} = P_{rs}^{-1} [P_f^{-1} \xi_f + P_b^{-1} \xi_b + M_1' P_{1s}^{-1} \hat{x}_{1s}] \quad (4.3)$$

$$\dot{\xi}_f = F_f \xi_f + K_{1f} \hat{x}_{1f} + P_f C_2' R_2^{-1} y_2 \quad (4.4)$$

$$-\dot{\xi}_b = F_b \xi_b + K_{1b} \hat{x}_{1b} \quad (4.5)$$

where P_f , P_{1s} , F_f , and K_{1f} , are as before, and P_b is the reverse error covariance for x based on y_1 alone:

$$-\dot{P}_b = -[A+Q]^{-1} P_b - P_b [A+Q]^{-1}]' + Q - P_b C_1' R_1^{-1} C_1 P_b \quad (4.7)$$

Also

$$P_{rs}^{-1} = P_f^{-1} + P_b^{-1} - \sum^{-1} \quad (4.8)$$

$$F_b = -A - Q \sum^{-1} - P_b C_1' R_1^{-1} C_1' \quad (4.9)$$

$$K_{1b} = [P_b M_1' P_{1r}^{-1} Q_1 P_{1r}^{-1} - Q M_1' P_{1r}^{-1}] + \left\{ P_b M_1' [-A_1' - \sum_1^{-1} Q_1] P_{1r}^{-1} - P_b [-A_1' - \sum_1^{-1} Q_1] M_1' P_{1r}^{-1} \right\} + P_b \dot{M}_1' P_{1r}^{-1} \quad (4.10)$$

Assuming that we can write

$$K_{1f} \hat{x}_{1f} = L_{1f} Q_1 P_{1f}^{-1} \hat{x}_{1f} \quad (4.11)$$

$$K_{1b} \hat{x}_{1r} = L_{1b} Q_1 P_{1r}^{-1} \hat{x}_{1r} \quad (4.12)$$

then, as before, (4.3)-(4.5) become

$$\hat{x}_{rs} = P_{rs} [P_f^{-1} q_f + P_b^{-1} q_b + E_1 \hat{x}_{1s}] \quad (4.13)$$

$$\dot{q}_f = F_f q_f + N_{1f} \hat{x}_{1s} + P_f C_2' R_2^{-1} y_2 \quad (4.14)$$

$$\dot{q}_b = F_b q_b + N_{1b} \hat{x}_{1s} \quad (4.15)$$

where

$$E_1 = M_1' P_{1s}^{-1} - P_f^{-1} L_{1f} - P_b^{-1} L_{1b} \quad (4.16)$$

$$N_{1f} = -F_f L_{1f} + \dot{L}_{1f} + L_{1f} \dot{P}_{1s} P_{1s}^{-1} - L_{1f} P_{1s} (A_1' + \sum_1^{-1} Q_1 - P_{1r}^{-1} Q_1) P_{1s}^{-1} \quad (4.17)$$

$$N_{1b} = -F_b L_{1b} - \dot{L}_{1b} - L_{1b} \dot{P}_{1s} P_{1s}^{-1} + L_{1b} P_{1s} (A_1' + P_{1f}^{-1} Q_1) P_{1s}^{-1} \quad (4.18)$$

Again there are several special cases worth mentioning. Suppose first that $x_1 = x$, i.e. that the local and global models are the same

$$A = A_1, Q = Q_1, C_1 = H_1, M_1 = I, \sum_1 = \sum \quad (4.19)$$

Then, comparing (3.9) and (4.7) we also have that

$$P_b = P_{1r} \quad (4.20)$$

This is not surprising, since P_b is the error covariance for the estimate of x based on the future of y_1 , P_{1r} is the covariance of the estimation error for x_1 based on the future of y_1 , and in this case $x=x_1$. What (4.19) and (4.20) also imply is that

$$K_{1b}=0 \quad (4.21)$$

which in turn implies that $q_b=0$. Thus there is no backward processing in this case. Again this is not surprising, since the future data at time t is just $\{y_1(s), s < t\}$, as $y_2(s), s < t$ has not yet been collected, and since $x_1=x$, the future of y_1 has already been processed optimally in producing \hat{x}_{1s} . Also for this case

$$K_{1f} = [P_f P_{1f}^{-1} - I] Q P_{1f}^{-1} \quad (4.22)$$

and the real-time smoothing solution is recursive and is given by

$$\hat{x}_{rs} = P_{rs} P_f^{-1} q_f + \hat{x}_{1s} \quad (4.23)$$

$$\dot{q}_f = F_f q_f + P_f C_2' R_2^{-1} [y_2 - C_2 \hat{x}_{1s}] \quad (4.24)$$

The other important case of interest is that in which x_1 is the state of a subsystem of the global system. Specifically, assume that (3.94)-(3.103) hold. In this case (3.117)-(3.122) can be adopted to yield

$$P_{rs} E_1 = \begin{pmatrix} I \\ T_1 \end{pmatrix} \quad (4.25)$$

$$N_{1f} = -P_f C_2' R_2^{-1} C_2 \begin{pmatrix} I \\ T_1 \end{pmatrix} + \begin{pmatrix} 0 \\ -\dot{T}_1 + A_{21} + A_{22} T_1 - T_1 A_1 \end{pmatrix} \quad (4.26)$$

$$N_{1b} = (P_b \int^{-1} - I) \begin{pmatrix} 0 \\ -\dot{T}_1 + A_{21} + A_{22} T_1 - T_1 A_1 \end{pmatrix} \quad (4.27)$$

Equations (4.13)-(4.15), (4.25)-(4.27) define the algorithm for real-time smoothing. Note that the new data (y_2) is processed only forward in time, while the reverse processing could be precomputed, since it only involves \hat{x}_{1s} . The interpretation of this reverse processing deserves some comment. Of course it is zero if $x_1 = x$. What it does represent essentially is a reconstruction of the reverse filtered estimate of x based only on y_1 , given the reverse filtered estimate \hat{x}_{1r} of x_1 based on y_1 . This is very much like what was discussed in Section 2.3 when one wishes to reconstruct the unobservable feedforward part of x from the filtered estimate of the observable part x_1 . However there is a difference because, as mentioned in Section 3.3, x_1 is not a substate of x in reverse time. If it were, then given \hat{x}_{1r} , the top block of N_{1b} would have to be zero, since the best estimate of x_1 based on the future of y_1 would have to be \hat{x}_{1r} . However the first part of N_{1b} is not zero, reflecting the fact that the reverse dynamics for x_1 alone are different from those when x_1 is viewed as some set of the components of x . In the latter case, x_1 is not a Markov process in reverse time.

V. DISCUSSION

In this paper we have considered the problems of combining estimates obtained from several separate data sources which have been processed individually and of updating an estimate as another source of information becomes available. In Section II we examined the causal version of this problem and have obtained a solution under very general conditions. Basically, the only restriction on the local processing is that the model on which it is based have as many degrees of freedom as there are in the observations that are to be processed locally. We discussed the potential utility of these results for distributed implementation of Kalman filters and for efficient transmission of information from local processors to a central processing facility.

Several directions for further work are suggested by the results of Section II. The first is in decentralized estimation. Consider the situation in which the local models x_1 and x_2 represent different pieces of the state x . In general these pieces will be coupled, although the local processors assume that there is no coupling. Given that the global processor does take this coupling into account, is there an efficient distributed fashion in which each local estimate can be corrected using the estimate produced by the other local processor? If the coupling between x_1 and x_2 is weak, is there some asymptotic description of this correction? What if there are different time scales? For example, suppose the local processors estimate fast and slow states but all that is wanted globally is an estimate of the slow global states. The results in [10-12] on multiple time scale estimation, combined with our framework should provide the basis for a solution to such a problem.

A second problem suggested by Section II is that of efficient distributed implementation of Kalman filters. Two types of issues enter here: (1) the amount of computation that is done by each local processor; and (2) the efficient transmission of information to the central processor. If in fact the only issue were the second one, then the answer would be that each processor should whiten the observed data y_i and transmit the result. In other words, each local processor should build a global Kalman filter and transmit the resulting innovations. Remember that the local Kalman filter innovations will not be white because of discrepancies between local and global models. Given that there are constraints on the amount of computation that can be performed locally, the question of what to transmit is a complex one. Specifically, given communication capacity and local computation constraints the problem becomes one of what local processing and subsequent data transmission scheme is best in the sense of degrading the global estimate as little as possible. Our results may provide one perspective from which we can make inroads into this very difficult problem.

In Sections III and IV we considered noncausal versions of the combining and update problems. These results are of potential use in some mapping problems. In addition, they raise as many question as they answer. Specifically, the noncausal estimate combining problem does not always have a solution. The reason for this is that the noncausal local processing may lose some information that is needed for the global processing. We presented several important cases where this does not happen, but the issue

remains of characterizing precisely what information from the raw data $y_i(\tau)$, $0 \leq \tau \leq T$, is preserved in the local smoothed estimate history $\hat{x}_{is}(\tau)$, $0 \leq \tau \leq T$.

Beyond this there remains the issue of interpreting the results of Section III and IV. The very simple form of the solution in some cases, such as in (3.65) and (3.66) suggests that there must be a simpler derivation and interpretation of our results than the one we have given. For example, the framework of scattering theory [13] may provide the machinery necessary to simplify our analysis and add to our insight. Also, as suggested in the text reference to (3.65) and (3.66), one interpretation of our map updating results is that the second pass data are used to estimate the map errors from the first pass. The fact that we have been able to determine how this can be done using two recursive systems (one causal and one anticausal) suggests that this second pass processing is based on a recursive model for the map errors. This suggests the notion of conditional stochastic realizations, which at this time remains as just a notion. The development of substance for this notion map provide the basic insight needed to understand our results from first principles.

Finally, there is the extension of our map updating formulation to more general scenerios (non-parallel tracks, point measurements as well as tracks) and more general random field models. As we have discussed in Section III, the resulting problems will probably be infinite dimensional in nature. While this is a technical difficulty, it need not be a conceptual one. The results we have obtained and the notion put forth of realizations

for map error fields should be useful in general. In addition, our results should directly carry over to discrete 2-D fields, in which case the generalization to more general scenarios need not be as difficult technically. The development of a more general theory for the efficient assimilation of spatially-distributed data, either in continuous- or discrete-space, is an extremely important problem with a myriad of potential applications. It is our hope and feeling that our results have provided some concepts that can be useful in developing that theory.

REFERENCES

1. W.H. Ledsham and D.H. Staelin, "An Extended Kalman-Bucy Filter for Atmospheric Temperature Profile Retrieval with a Passive Microwave Sounder," J. Applied Meteorology, Vol. 17, No. 7, July 1978, pp. 1023-1033.
2. R.A. Nash, Jr. and S.K. Jordan, "Statistical Geodosy -- An Engineering Perspective," Proc. IEEE, Vol. 66, No. 5, May 1978, pp. 532-550
3. J.L. Speyer, "Computation and Transmission Requirements for a Decentralized Linear-Quadratic-Gaussian Control Problem," IEEE Trans. Aut. Control, Vol. AC-24, No. 2, April 1979, pp. 266-269
4. J.E. Wall, Jr., A.S. Willsky, and N.R. Sandell, Jr., "On the Fixed-Interval Smoothing Problem," submitted to Stochastics.
5. F.C. Schweppe, Uncertain Dynamic Systems, Prentice-Hall, Inc. Englewood Cliffs, N.J., 1973.
6. G. Verghese and T. Kailath, "A Further Note on Backwards Markovian Models," IEEE Trans. Inf. Theory, Vol. IT-25, No. 1, Jan. 1979, pp. 121-124.
7. C.Y. Chong, "Hierarchical Estimation," presented at the Second MIT/ONR Workshop on Distributed Information and Decision Systems Motivated by Naval Command-Control-Communication (C³) Problems, Naval Postgraduate School, Monterey, Calif., July 1979.
8. S. Attasi, "Modelling and Recursive Estimation for Double Indexed Sequences," in System Identification: Advances and Case Studies, edited by R.K. Mehra and D.G. Lainiotis, Academic Press, New York, 1976.
9. A.S. Willsky and N.R. Sandell, "The Stochastic Analysis of Dynamic Systems Moving Through Random Fields," to be submitted for publication.
10. A.H. Haddad, "Linear Filtering of Singularly Perturbed Systems," IEEE Trans. Aut. Control, Vol. AC-21, No. 4, Aug. 1976, pp. 515-519.
11. D. Teneketzis and N.R. Sandell, Jr., "Linear Regulator Design for Stochastic Systems by a Multiple Time Scales Method," IEEE Trans. Aut. Contr., Vol. AC-22, No. 4, 1977.
12. C.E. Hutchinson and N.A. Acharya, "Singularly Perturbed Estimation Using the Method of Multiple Time Scales," Proc. 1979 JACC, Denver, Colo., pp. 629-633.

13. L. Ljung, T. Kailath and B. Friedlander, "Scattering Theory and Linear Least Squares Estimation - Part I: Continuous-Time Problems," Proc. IEEE, Vol. 64, Jan. 1976, pp. 131-139.
14. R.F. Brammer, "Estimation of the Ocean Geoid near the Blake Escarpment Using GEOS-3 Satellite Altimetry Data," submitted to J. Geophys. Research.