# Combining and Updating of Local Estimates and Regional Maps Along Sets of One-Dimensional Tracks 

ALAN S. WILLSKY, MEMBER, ieee, MARTIN G. BELLO, DAVID A. CASTANON, MEMBER, IEEE,<br>BERNARD C. LEVY, MEMBER, IEEE, AND GEORGE C. VERGHESE, MEMBER, IEEE


#### 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 spatiallydistributed 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.


## I. Introduction

SUPPOSE THAT, as illustrated in Fig. 1, we are interested in the estimation of a two-dimensional field given data obtained along parallel measurement tracks. Problems of this type arise in the mapping of gravitational anomalies given data along tracks over the ocean [2], [14] and the mapping of meteorological variables from data gathered by satellites [1]. Typically, in such problems the data are collected at several times, i.e., a number of distinct surveys are performed. In some cases all data are used to produce a single map of the desired region, and as new data are collected, one is confronted with the map updating

[^0]

Fig. 1. Parallel data tracks across a two-dimensional random field.
problem: update an existing map with a new set of survey data. In other situations, different sets of data are used to produce separate maps, perhaps of the same region of the field, perhaps of overlapping regions, or perhaps of completely separate regions. The map combining problem is that of combining these maps to produce a single map of some overall region. In this paper we provide solutions to certain formulations of these two problems, and, in the process, will solve a problem in decentralized estimation. In the remainder of the introduction we provide a discussion of our formulation that should make clear the focus and the contributions of this paper.

Let $f(t, s)$ denote the two-dimensional random field, which we assume to be Gaussian and, for simplicity, zeromean. In order to formulate the two mapping problems mathematically, we assume for simplicity that there are only two surveys to be considered (the extension to $N$ surveys is immediate). As indicated in Fig. 2, $t$ is the direction in which the tracks are taken. In particular, suppose that the first survey measures $f$ along the lines $\left\{\left(t, s_{j 1}\right) \mid 0 \leqslant t \leqslant T\right\}, j=1, \cdots, M_{1}$ and that the second survey measures the field along $\left\{\left(t, s_{j 2}\right) \mid 0 \leqslant t \leqslant T\right\}, j=1, \cdots, M_{2}$. Assume further that we would like our overall map to provide estimates of $f$ along a grid of parallel lines that includes both of these survey sets. For the map combining problem, we assume further that there are two separate maps, each based on one of the surveys and covering some set of lines including those for the corresponding survey.

In our treatment we will view any collection of parallel lines of the random field (i.e., $\left.f\left(t, s_{j}\right), j=1, \cdots, M\right)$ as a vector function of $t$ and further will assume that this resulting vector process has a finite dimensional shaping filter representation. In particular, we assume that the
values of the field along the grid of lines used for the overall map can be viewed as the output of the model

$$
\begin{align*}
\dot{x}(t) & =A(t) x(t)+w(t)  \tag{1.1}\\
E\left[w(t) w(\tau)^{\prime}\right] & =Q(t) \delta(t-\tau) \tag{1.2}
\end{align*}
$$

where $w(t)$ is independent of $x(0)$ which is taken to be a zero-mean Gaussian random variable with covariance $\Sigma(0)$. The set of field lines for each survey can also be obtained as a set of linear combinations of the components of $x(t)$, and consequently we can model the set of noise-corrupted measurements from the $i$ th survey by an equation of the form

$$
\begin{equation*}
y_{i}(t)=C_{i}(t) x(t)+v_{i}(t), \quad i=1,2 \tag{1.3}
\end{equation*}
$$

where $v_{1}$ and $v_{2}$ are independent of each other and of $w$ and $x(0)$, with

$$
\begin{equation*}
E\left[v_{i}(t) v_{i}(\tau)\right]=R_{i}(t) \delta(t-\tau), \quad i=1,2 \tag{1.4}
\end{equation*}
$$

Here $C_{i}(t) x(t)$ equals the vector of values of the field along the lines of the $i$ th survey, i.e., $\left[f\left(t, s_{1 i}\right), f\left(t, s_{2 i}\right), \cdots\right]^{\prime}$.

For the map combining problem we assume that the processing algorithms for each survey are based on different models:

$$
\begin{align*}
\dot{x}_{i}(t) & =A_{i}(t) x_{i}(t)+w_{i}(t), \quad i=1,2  \tag{1.5}\\
E\left[w_{i}(t) w_{i}(\tau)^{\prime}\right] & =Q_{i}(t) \delta(t-\tau), \quad i=1,2  \tag{1.6}\\
y_{i}(t) & =H_{i}(t) x_{i}(t)+v_{i}(t), \quad i=1,2 \tag{1.7}
\end{align*}
$$

where $x_{i}(0)$ is taken to be zero-mean with covariance $\Sigma_{i}(0)$. It is important to emphasize at this point one key feature of our work. Specifically, the model for $x(t), y_{1}(t)$, and $y_{2}(t)$ in (1.1)-(1.4) should be thought of as providing an exact representation of the field along the specified grid of lines. On the other hand (1.5)-(1.7) should be thought of as providing a model-perhaps exact, perhaps only ap-proximate-for the grid of lines used in the individual mapping procedure for each survey. For example, it is possible that $x_{i}(t)=x(t)$ so that $x_{i}(t)$ exactly represents the field along the same set of lines as in the overall map. Alternatively, $x_{i}(t)$ might exactly represent the field along some subset of the lines used in the overall map. In this case $x_{i}(t)=B_{i} x(t)$ for some matrix $B_{i}$. On the other hand, it may be true that $x_{i}(t) \neq B_{i} x(t)$ for any matrix $B_{i}$. In this case either $x_{i}(t)$ exactly represents the field along a set of tracks including some not in the set used in the overall map or $x_{i}(t)$ is only an approximate model for the field along a set of lines obtained, for example, by using a reduced-order model that neglects some correlation in the field. In the following sections we will begin by assuming no relationship between the model (1.5)-(1.7) (which we will term the local model) and the model (1.1)-(1.4) (which will be called the global model), except for the assumption that the $v_{i}$ in (1.7) are the same as in (1.4) (i.e., that we have modeled the measurement noise correctly in the local model). As we
need to impose some relationship between local and global models, we will do so.

Given the models just described, we can now formulate our mapping problems precisely. In particular, let $\hat{x}_{i s}(t)$ denote the smoothed estimate of $x_{i}(t)$ based on $y_{i}(t)$ and on the corresponding local model. That is,

$$
\begin{equation*}
\hat{x}_{i s}(t)=E_{i}\left[x_{i}(t) \mid y_{i}(\tau), 0 \leqslant \tau \leqslant T\right] \tag{1.8}
\end{equation*}
$$

where " $E_{i}$ " is used to denote expectations based on the model (1.5)-(1.7). We also let $\hat{x}_{s}(t)$ denote the smoothed estimate of $x(t)$ based on both $y_{1}(t)$ and $y_{2}(t)$ :

$$
\begin{equation*}
\hat{x}_{s}(t)=E\left[x(t) \mid y_{1}(\tau), y_{2}(\tau), 0 \leqslant \tau \leqslant T\right] \tag{1.9}
\end{equation*}
$$

(" $E$ " is used for expectations with respect to the global model). The map combining problem is then the problem of determining if it is possible to compute $\hat{x}_{s}(t), 0 \leqslant t \leqslant T$ in terms of $\hat{x}_{1 s}(t)$ and $\hat{x}_{2 s}(t), 0 \leqslant t \leqslant T$ and constructing an efficient algorithm if it is. The map updating problem is one of computing $\hat{x}_{s}(t), 0 \leqslant t \leqslant T$ in terms of $\hat{x}_{1 s}(t)$ and $y_{2}(t), 0 \leqslant t \leqslant T$.
In the following sections of this paper we provide efficient procedures for updating and combining smoothed estimates. Our method of solution for these problems is based on the two-filter form for optimum smoothers presented in [4], and a key step in our procedure is the solution of the problem of combining and updating causally filtered estimates. That is, let

$$
\begin{align*}
& \hat{x}_{f}(t)=E\left[x(t) \mid y_{1}(s), y_{2}(s), 0 \leqslant s \leqslant t\right]  \tag{1.10}\\
& \hat{x}_{i j}(t)=E_{i}\left[x_{i}(t) \mid y_{i}(s), 0 \leqslant s \leqslant t\right] . \tag{1.11}
\end{align*}
$$

Then the problem of combining local estimates is that of causally computing $\hat{x}_{f}$ from the $\hat{x}_{i f}$. Note that this problem is of independent interest, as it can be interpreted as a decentralized estimation problem, where $x(t)$ is the state of a large, distributed system, $y_{1}(t)$ and $y_{2}(t)$ are measurements made at different points in the system, and $x_{1}(t)$ and $x_{2}(t)$ are exact or approximate descriptions of the portions of the dynamics of interest to the $i$ th local processor. A problem of this type was considered by Speyer [3] in the context of decentralized control and also by Chong [7], and our work in Section II represents an extension of their results. We will comment on the potential utility of these decentralized estimation results in the following section.

In the remaining sections of this paper we will take as our starting points the one-dimensional models given in (1.1)-(1.7). Before analyzing the problems we have posed in this context, it is clearly appropriate to address the questions of the restrictiveness of the assumption of the existence of a finite-dimensional model as in (1.1)-(1.4) and of the specific measurement geometry of Fig. 1. To begin this discussion, define the two-dimensional correlation function

$$
\begin{equation*}
R(t, \tau ; s, \sigma)=E\left[f(t, s) f^{\prime}(\tau, \sigma)\right] . \tag{1.12}
\end{equation*}
$$

Then one important case in which we can find a model of the desired form is when $R$ is separable:

$$
\begin{equation*}
R(t, \tau ; s, \sigma)=R_{1}(t, \tau) R_{2}(s, \sigma) \tag{1.13}
\end{equation*}
$$

with $R_{1}$ and $R_{2}$ square

$$
\begin{equation*}
R_{1}(t, \tau)=R_{1}^{\prime}(\tau, t), \quad R_{2}(s, \sigma)=R_{2}^{\prime}(\sigma, s) \tag{1.14}
\end{equation*}
$$

and $R_{1}$ itself separable

$$
\begin{equation*}
R_{\mathrm{j}}(t, \tau)=H(t) G(\tau), \quad t \geqslant \tau \tag{1.15}
\end{equation*}
$$

The conditions (1.13)-(1.15) are a slight generalization of classes of processes considered by others. For example, if we had further assumed separability for $R_{2}$ as in (1.15) and made the field stationary (so that (1.15) becomes $R_{1}(\tau)=$ $H e^{F \tau} G, \tau \geqslant 0$ ), we would have the continuous-space version of the model considered by Attasi [8]. The derivation of a model of the form of (1.1)-(1.14) for a field satisfying the above conditions is a straightforward exercise (see, for example, [16], [17]).

It is important to point out that if the random field under consideration is not separable, then the resulting model for a set of parallel tracks is infinite dimensional. While our results do not apply to such problems, they do shed significant light on the structure that map updating and combining algorithms must have in this case. In particular, any nonseparable covariance can be approximated arbitrarily closely by a separable one [8], and thus we can obtain suboptimal procedures that are arbitrarily close to the optimal one. Furthermore, initial results in [16] indicate that the discrete-space version of the results derived here are applicable to the nonseparable case. Thus, while there are interesting problems that are still outstanding for con-tinuous-space nonseparable fields, much of the conceptual nature of the problem and solution can be ascertained from the simpler problem considered here. Also, the same can be said for the case of nonparallel measurement geometries. In particular, an example in [16] indicates that discrete-space versions of problems of this type can be formulated in terms of finite-dimensional linear models and consequently can be solved using the discrete-time counterparts of the results derived in this paper. Also, one will be able to use our results to construct suboptimal finite-dimensional estimators which are arbitrarily close to the optimum for continuous-space problems.

For these reasons we feel that the results presented in this paper represent an important step in the development of efficient high-performance mapping algorithms. Since we will focus in this paper on the one-dimensional formulation in (1.1)-(1.7), we have chosen to present our results in the continuous-parameter case, where the algebra is less cumbersome, in order to highlight their nature more clearly. In a subsequent paper we will use the ideas and results developed here in order to address the two-dimensional aspects of the problem and will use both the continuous and discrete formulations.

In the next section we present and discuss the solution to the problem of combining decentralized filtered estimates, while Section III contains the solution of the problems of combining and updating smoothed estimates. In Section IV we apply these results to the problem of real-time smooth-
ing, 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

## A. The General Case

Consider the global model of (1.1)-(1.4) and local models of (1.5)-(1.7). In this section we examine the causal computation of $\hat{x}_{f}(t)$ from $\hat{x}_{1 f}(t)$ and $\hat{x}_{2 f}(t)$ [see (1.10), (1.11)]. As we have said, this is a necessary first step in the derivation of our smoothing results. In the context of decentralized estimation, the implications of this problem are several. Specifically, if we can recover $\hat{x}_{f}(t)$ from the local estimates, then much of the raw data processing can be done locally, without any loss of 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. Furthermore, allowing the local models to differ from the global models leads to several potential advantages. For example, presumably the local models are lower dimensional than (1.1). In this case the local processor can be made far less complex than the global processor. We will comment further on these issues as we proceed and also in Section V.

To begin our derivation, note that $\hat{x}_{1 f}(t)$ and $\hat{x}_{2 f}(t)$ are computed by Kalman filters based on the local models ${ }^{1}$ :

$$
\begin{equation*}
\hat{x}_{i f}(t)=\left[A_{i}-P_{i f} H_{i}^{\prime} R_{i}^{-1} H_{i}\right] \hat{x}_{i f}(t)+P_{i f} H_{i}^{\prime} R_{i}^{-1} y_{i}(t) \tag{2.1}
\end{equation*}
$$

The covariance $P_{i f}$ can be precomputed from either of the following equations:

$$
\begin{gather*}
\dot{P}_{i f}=A_{i} P_{i f}+P_{i f} A_{j}^{\prime}+Q_{i}-P_{i f} H_{i}^{\prime} R_{i}^{-1} H_{i} P_{i f}  \tag{2.2}\\
\frac{d}{d t}\left(P_{i f}^{-1}\right)=-P_{i f}^{-1} A_{i}-A_{i}^{\prime} P_{i f}^{-1}-P_{i f}^{-1} Q_{i} P_{i f}^{-1}+H_{i}^{\prime} R_{i}^{-1} H_{i} \tag{2.3}
\end{gather*}
$$

with the initial condition $P_{i f}(0)=\Sigma_{i}(0)$.
As will become clear shortly, a necessary and sufficient condition for our being able to recover $\hat{x}_{f}$ from $\hat{x}_{1 f}$ and $\hat{x}_{2 f}$ is that there exist (possibly time-varying) matrices $M_{1}$ and $M_{2}$ such that

$$
\begin{equation*}
C_{i}=H_{i} M_{i}, \quad i=1,2 \tag{2.4}
\end{equation*}
$$

This condition is equivalent to the statement 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 interrelationship. That is, if the local models (1.5)-(1.9) assume any redundancy among the sensed quantities-i.e., the components of $y_{i}$-then that redundancy must actually

[^1]exist in the global model. Note that valid choices for $M_{1}$ and $M_{2}$ are
\[

$$
\begin{equation*}
M_{i}=H_{i}^{\dagger} C_{i}, \quad i=1,2 \tag{2.5}
\end{equation*}
$$

\]

(where " $\dagger$ " denotes pseudoinverse) and the choice is unique only if $N\left(H_{i}\right)=\{0\}$.

Thus, the dynamics (1.5), (1.6) can be totally arbitrary, as long as (2.4) 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 (2.4) is sufficiently weak so that, if we desire, we can always choose a local model of this minimal dimension that satisfies the condition. Therefore, the condition does not require that there be any physical relationship between the local states, $x_{1}$ and $x_{2}$, and the global state $x$. On the other hand, (2.4) suggests an interpretation of $x_{i}$ as being a part of the global state, specifically $M_{i} x$. If this is the case, then (1.5) 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 usually true 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. If, however, it is the case that $x=M_{i} x$, 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 $\hat{x}_{f}$ :

$$
\begin{align*}
& \dot{\hat{x}}_{f}(t)=\left[A-P_{f} C_{1}^{\prime} R_{1}^{-1} C_{1}-P_{f} C_{2}^{\prime} R_{2}^{-1} C_{2}\right] \hat{x}_{f}(t) \\
&+P_{f} C_{1}^{\prime} R_{1}^{-1} y_{1}(t)+P_{f} C_{2}^{\prime} R_{2}^{-1} y_{2}(t) \tag{2.6}
\end{align*}
$$

where $P_{f}$ can be calculated from

$$
\begin{equation*}
\dot{P}_{f}=A P_{f}+P_{f} A^{\prime}+Q-P_{f} C_{1}^{\prime} R_{1}^{-1} C_{1} P_{f}-P_{f} C_{2}^{\prime} R_{2}^{-1} C_{2} P_{f} \tag{2.7}
\end{equation*}
$$

with $P_{f}(0)=\Sigma(0)$. The solution to the problem we have posed can be obtained as follows. Rearranging (2.1) we have ${ }^{2}$

$$
\begin{equation*}
H_{i}^{\prime} R_{i}^{-1} y_{i}=P_{i f}^{-1}\left\{\dot{\hat{x}}_{i f}-\left[A_{i}-P_{i f} H_{i}^{\prime} R_{i}^{-1} H_{i}\right] \hat{x}_{i f}\right\} . \tag{2.8}
\end{equation*}
$$

Examining (2.6), we see that the quantities needed in the calculation of $\hat{x}_{f}$ are $C_{1}^{\prime} R_{1}^{-1} y_{1}$ and $C_{2}^{\prime} R_{2}^{-1} y_{2}$. These can be obtained from (2.8) if and only if matrices $M_{1}$ and $M_{2}$ exist that satisfy (2.4). Assuming that this is the case, we can combine (2.4), (2.6), and (2.8) to obtain

$$
\begin{align*}
\dot{\hat{x}}_{f}= & {\left[A-P_{f} C_{1}^{\prime} R_{1}^{-1} C_{1}-P_{f} C_{2}^{\prime} R_{2}^{-1} C_{2}\right] \hat{x}_{f} } \\
& +P_{f} M_{1}^{\prime} P_{1 f}^{-1}\left\{\dot{\hat{x}}_{1 f}-\left[A_{1}-P_{1 f} H_{1}^{\prime} R_{1}^{-1} H_{1}\right] \hat{x}_{1 f}\right\} \\
& +P_{f} M_{2}^{\prime} P_{2 f}^{-1}\left\{\dot{\hat{x}}_{2 f}-\left[A_{2}-P_{2 f} H_{2}^{\prime} R_{2}^{-1} H_{2}\right] \hat{x}_{2 f}\right\} . \tag{2.9}
\end{align*}
$$

[^2]In order to simplify notation, define the following quantities:

$$
\begin{align*}
F_{f} & =A-P_{f} C_{1}^{\prime} R_{1}^{-1} C_{1}-P_{f} C_{2}^{\prime} R_{2}^{-1} C_{2}  \tag{2.10}\\
F_{i f} & =A_{i}-P_{i f} H_{i}^{\prime} R_{i}^{-1} H_{i} \quad i=1,2  \tag{2.11}\\
G_{i f} & =P_{f} M_{i}^{\prime} P_{i f}^{-1} \quad i=1,2 \tag{2.12}
\end{align*}
$$

Then it can be checked by straightforward algebra that $\hat{x}_{f}$ can be calculated from the following equations:

$$
\begin{align*}
& \dot{\xi}_{f}=F_{f} \xi_{f}+K_{1 f} \hat{x}_{1 f}+K_{2 f} \hat{x}_{2 f}  \tag{2.13}\\
& \hat{x}_{f}=\xi_{f}+G_{1 f} \hat{x}_{1 f}+G_{2 f} \hat{x}_{2 f}
\end{align*}
$$

where ${ }^{3}$

$$
\begin{align*}
K_{i f}= & F_{f} G_{i f}-\dot{G}_{i f}-G_{i f} F_{i f} \\
= & {\left[P_{f} M_{i}^{\prime} P_{i f}^{-1} Q_{i} P_{i f}^{-1}-Q M_{i}^{\prime} P_{i f}^{-1}\right] } \\
& +\left[P_{f} M_{i}^{\prime} A_{i}^{\prime} P_{i f}^{-1}-P_{f} A^{\prime} M_{i}^{\prime} P_{i f}^{-1}-P_{f} \dot{M}_{i}^{\prime} P_{i f}^{-1}\right] \\
& i=1,2 \tag{2.15}
\end{align*}
$$

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 filtered estimates. In addition, this solution can be directly adapted to the problem of computing $\hat{x}_{f}$ from $\hat{x}_{1 f}$ 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 (2.6), and instead of replacing both $C_{1}^{\prime} R_{1}^{-1} y_{1}$ and $C_{2}^{\prime} R_{2}^{-1} y_{2}$ by expressions in terms of $\hat{x}_{i f}$ and $\dot{\hat{x}}_{i f}$, we make this substitution only for $C_{1}^{\prime} R_{1}^{-1} y_{1}$. The remaining analysis is analogous to that carried out previously, and the result is

$$
\begin{equation*}
\hat{x}_{f}=\rho_{f}+G_{1 f} \hat{x}_{1 f} \tag{2.16}
\end{equation*}
$$

where

$$
\begin{equation*}
\dot{\rho}_{f}=F_{f} \rho_{f}+K_{1} \hat{x}_{1 f}+P_{f} C_{2}^{\prime} R_{2}^{-1} y_{2} . \tag{2.17}
\end{equation*}
$$

Here $F_{f}, K_{1 f}$, and $G_{1 f}$ are the same as given previously.
In the next two subsections we present two special cases which result in some simplifications and consequently allow us to interpret our result in more detail.

## B. 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,

$$
\begin{align*}
& A_{1}=A_{2}=A, \quad Q_{1}=Q_{2}=Q, \\
& C_{1}=H_{1}, \quad C_{2}=H_{2}, \quad M_{1}=M_{2}=I . \tag{2.18}
\end{align*}
$$

[^3]In this case the expressions for $K_{1 f}$ and $K_{2 f}$ simplify to

$$
\begin{equation*}
K_{i f}=P_{f} P_{i f}^{-1} Q P_{i f}^{-1}-Q P_{i f}^{-1}=\left(P_{f} P_{i f}^{-1}-I\right) Q P_{i f}^{-1} \tag{2.19}
\end{equation*}
$$

and

$$
\begin{equation*}
\hat{x}_{f}=\xi_{f}+P_{f}\left(P_{1 f}^{-1} \hat{x}_{1 f}+P_{2 f}^{-1} \hat{x}_{2 f}\right) . \tag{2.20}
\end{equation*}
$$

Note that the second term in the expression for $\hat{x}_{f}$ is the usual expression for combining independent estimates [4], [5]. However, $\hat{x}_{1 f}$ and $\hat{x}_{2 f}$ are not independent in general, and $\xi_{f}$ represents a correction for this correlation.

The reason that $\hat{x}_{1 f}$ and $\hat{x}_{2 f}$ 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 $\left(\xi_{f}\right)$ to account for the correlation in the processes $\hat{x}_{1 f}$ and $x_{2 f}$. If $Q=0$ (i.e., if $w(t)$ is not present), then $K_{i f}=0$ and hence $\xi_{f}=0$, and $\hat{x}_{f}$ is a memoryless function of $\hat{x}_{1 f}$ and $\hat{x}_{2 f}$. In this case it is straightforward to show that

$$
\begin{equation*}
\hat{x}_{f}(t)=P_{f}(t)\left[P_{1 f}^{-1}(t) \hat{x}_{1 f}(t)+P_{2 f}^{-1}(t) \hat{x}_{2 f}(t)\right] \tag{2.21}
\end{equation*}
$$

and

$$
\begin{equation*}
P_{f}^{-1}(t)=P_{1 f}^{-1}(t)+P_{2 f}^{-1}(t)-\Sigma^{-1}(t) \tag{2.22}
\end{equation*}
$$

where $\Sigma(t)$ is the unconditional covariance of $x(t)$.
Note that even in this case $\hat{x}_{1 f}$ and $\hat{x}_{2 f}$ are not independent estimates because of the correlation of the estimation errors with $x(0)$. Following the work of Wall [4], we can interpret (2.21) and (2.22) 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 $\hat{x}_{i f}$ uses $y_{i}$ and the a priori information, which, therefore is used twice. Equation (2.22) corrects for the fact that both $P_{1 j}^{-1}$ and $P_{2 f}^{-1}$ reflect the uses of this information. Also, (2.22) 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 (2.21) is replaced by

$$
\begin{equation*}
\hat{x}_{f}(t)=P_{f}(t)\left[P_{1 f}^{-1} \hat{x}_{1 f}(t)+P_{2 f}^{-1} \hat{x}_{2 f}(t)-\Sigma^{-1}(t) m(t)\right] \tag{2.23}
\end{equation*}
$$

where $m(t)$ is the unconditional mean of $x(t)$. 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 $K_{i f}=0$ also if $P=P_{i f}$. 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 ( $P_{f} P_{i f}^{-1}-I$ ) 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 $\hat{x}_{f}$ is a memoryless function of $\hat{x}_{1 f}$ and $\hat{x}_{2 f}$. Also, if one set of measurements, say set 1 , contains no information about a part of $x$, then the corresponding parts of $P_{f}$ and $P_{2 f}$ are identical.

## C. 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

$$
\begin{equation*}
x_{1}(t)=M_{1}(t) x(t) \tag{2.24}
\end{equation*}
$$

Equation (2.24) has several important implications. Since $x_{1}$ satisfies (1.5), (1.6) and $x$ satisfies (1.1), (1.2), (2.24) states that the Markov process $x(t)$ has a subprocess, namely $x_{1}(t)$, that is Markov by itself. Some simple algebra allows us to conclude that (2.24) implies the following relations:

$$
\begin{align*}
A_{1} M_{1} & =\dot{M}_{1}+M_{1} A  \tag{2.25}\\
w_{1} & =M_{1} w  \tag{2.26}\\
Q_{1} & =M_{1} Q M_{1}^{\prime}  \tag{2.27}\\
\Sigma_{1} & =M_{1} \Sigma M_{1}^{\prime} . \tag{2.28}
\end{align*}
$$

Note that from (2.28) it is clear that $\Sigma_{1}$ is invertible only if $M_{1}$ is onto (assuming that $\Sigma$ is invertible). We will assume that this is the case, since from (2.24) 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 $K_{1 f}$ simplifies as follows:

$$
\begin{align*}
K_{1 f} & =P_{f} M_{1}^{\prime} P_{1 f}^{-1} M_{1} Q M_{1}^{\prime} P_{1 f}^{-1}-Q M_{1}^{\prime} P_{1 f}^{-1} \\
& =\left[P_{f} M_{1}^{\prime} P_{1 f}^{-1} M_{1}-I\right] Q M_{1}^{\prime} P_{1 f}^{-1} \tag{2.29}
\end{align*}
$$

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, let us 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

$$
\begin{align*}
x & =\left(\frac{x_{1}}{\gamma}\right)  \tag{2.30}\\
M_{1} & =(I: 0)  \tag{2.31}\\
\binom{\dot{x}_{1}}{\gamma} & =\left(\begin{array}{cc}
A_{1} & 0 \\
A_{12} & A_{22}
\end{array}\right)\binom{x_{1}}{\gamma}+w  \tag{2.32}\\
Q & =\left(\begin{array}{ll}
Q_{1} & Q_{12} \\
Q_{12}^{\prime} & Q_{22}
\end{array}\right)  \tag{2.33}\\
y_{1} & =\left(\begin{array}{ll}
H_{1} & 0
\end{array}\right)\binom{x_{1}}{\gamma}+v_{1} \tag{2.34}
\end{align*}
$$



Fig. 2.

$$
\begin{equation*}
y_{2}=\left(C_{21} \vdots C_{22}\right)\binom{x_{1}}{\gamma}+v_{2} \tag{2.35}
\end{equation*}
$$

This form is illustrated in Fig. 2. Note from the figure that it is clear that the global system is not observable from $y_{1}$ alone. Using (2.30)-(2.35), (2.29) becomes

$$
\begin{align*}
K_{1 f} & =\left\{P_{f}\binom{P_{1 f}^{-1}}{\hdashline 0} Q_{1} P_{1 f}^{-1}-\binom{Q_{1}}{Q_{12}^{\prime}} P_{1 j}^{-1}\right\} \\
& =\binom{\left[\left(P_{f}\right)_{11} P_{1 f}^{-1}-I\right] Q_{1} P_{1 f}^{-1}}{\left[\left(P_{f}\right)_{12}^{\prime} P_{1 f}^{-1}-Q_{12}^{\prime}\right] P_{1 f}^{-1}} \tag{2.36}
\end{align*}
$$

where

$$
P_{f}=\left(\begin{array}{ll}
\left(P_{f}\right)_{11} & \left(P_{f}\right)_{12}  \tag{2.37}\\
\left(P_{f}\right)_{12}^{\prime} & \left(P_{f}\right)_{22}
\end{array}\right)
$$

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. ${ }^{4}$ Since $K_{1 f}$ is partitioned, let us consider each part individually. The first piece, $\left[\left(P_{f}\right)_{11} P_{1 f}^{-1}-I\right] Q_{1} P_{I f}^{-1}$, is exactly of the form of the gain that we saw in the preceding subsection when the local and global models were identical [see (2.19)]. 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 $\hat{x}_{1 f}$ into a global estimate of this piece of $x$, given $y_{1}$ and $y_{2}$, is the same as the problem we considered in II-B.

The second piece $\left[\left(P_{f}\right)_{12}^{\prime} P_{1 f}^{-1} Q_{1}-Q_{12}^{\prime}\right] P_{1 f}^{-1}$ 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 condi-

[^4]tions $\left(P_{f}\right)_{11}=P_{1 f}$, which merely state that local processor no. 1 produces the best filtered estimate of $x_{1}$ given $y_{1}$. From (2.36) we see that this observation is consistent with the fact that the first part of $K_{1 f}$ is zero. Also, the second piece of $K_{i f}$ becomes
$$
\left[\left(P_{f}\right)_{12}^{\prime}\left(P_{f}\right)_{11}^{-1} Q_{1}-Q_{12}^{\prime}\right]\left(P_{f}\right)_{11}^{-1}
$$
and using (2.10)-(2.14) and (2.36), the optimal estimator for $\gamma$ becomes
\[

$$
\begin{align*}
\gamma_{f} & =\eta_{f}+\left(P_{f}\right)_{12}^{\prime}\left(P_{f}\right)_{11}^{-1} \hat{x}_{1 f}  \tag{2.38}\\
\dot{\eta}_{f} & =A_{22} \eta_{f}+\left[\left(P_{f}\right)_{12}^{\prime}\left(P_{f}\right)_{11}^{-1}-Q_{12}^{\prime}\right]\left(P_{f}\right)_{11}^{-1} \hat{x}_{1 f} \tag{2.39}
\end{align*}
$$
\]

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. In this case the complexity of each of the two processors (for $\hat{x}_{1 f}$ and for $\hat{\gamma}_{f}$ ) is 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 (2.17) 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 implementation advantages in breaking the overall optimal estimator into smaller pieces.

Note also from (2.29) that $K_{1 f}=0$ if $Q=0$. In fact, from (2.31) and (2.33) (together with the fact that $Q_{12}$ must be zero if $Q_{1}$ is), we see that $K_{1 f}=0$ if $Q_{1}=0$. In this case, whether $y_{2}$ is present or not, $\hat{x}_{f}$ depends on $\hat{x}_{1 f}$ in a memoryless fashion. This is best understood by noting that with $Q_{1}=0, x_{1}$ is a time-varying bias ${ }^{5}$

$$
\begin{equation*}
x_{1}(t)=\Phi_{1}(t, 0) x_{1}(0) \tag{2.40}
\end{equation*}
$$

and it also produces a time-varying bias in $\gamma$ :

$$
\begin{align*}
\gamma(t)=\Phi_{22}(t, 0) \gamma(0) & +\int_{0}^{t} \Phi_{22}(t, \tau) A_{21}(\tau) x_{1}(\tau) d t \\
& +\int_{0}^{t} \Phi_{22}(t, \tau)[0 \vdots I] w(\tau) d \tau \tag{2.41}
\end{align*}
$$

The measurements $y_{1}$ provide information about the second term in (2.41), which can be rewritten as

[^5]\[

$$
\begin{equation*}
\left[\int_{0}^{t} \Phi_{22}(t, \tau) A_{21}(\tau) \Phi(\tau, t) d \tau\right] x_{1}(t) \tag{2.42}
\end{equation*}
$$

\]

Thus the best estimate of $\gamma$ given the measurements $y_{1}$ is simply a memoryless function of $\hat{x}_{1 f}$. For example, if we do not have a second set of measurements ( $C_{21}=C_{22}=0$ ), then from (2.14), (2.31), and (2.36) we find that

$$
\begin{equation*}
\hat{\gamma}_{f}=\left(P_{f}\right)_{12}^{\prime} P_{1 f}^{-1} \hat{x}_{1 f} . \tag{2.43}
\end{equation*}
$$

## III. The Combining and Updating of Smoothed Estimates

In this section we consider the problems of combining and updating smoothed estimates which were introduced in Section I [see, in particular, (1.8) and (1.9)]. As we will see, the updating problem always has a solution, while the combining problem can be solved only when some further mild restriction, beyond the existence of $M_{1}$ and $M_{2}$, is placed on the local models. In the next subsection we develop the basic ideas behind our approach, while in the subsequent two subsections we address the two special cases considered in Sections II-B and II-C, which are the most important for problems of random field mapping.

## A. 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] and in [15]. 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.

Let us summarize the smoother equations for each of the two local processors. The forward estimator for processor $i$ ( $i=1,2$ ) is specified by (2.1)-(2.3). The reverse time estimator involves the unconditional covariance for the local model assumed by the processor, which can be calculated from

$$
\begin{equation*}
\dot{\Sigma}_{i}=A_{i} \Sigma_{i}+\Sigma_{i} A_{i}^{\prime}+Q_{i} \tag{3.1}
\end{equation*}
$$

or

$$
\begin{equation*}
\frac{d}{d t} \Sigma_{i}^{-1}=-\Sigma_{i}^{-1} A_{i}-A_{i}^{\prime} \Sigma_{i}^{-1}-\Sigma_{i}^{-1} Q_{i} \Sigma_{i}^{-1} . \tag{3.2}
\end{equation*}
$$

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

$$
-\dot{\hat{x}}_{i r}=\left[-A_{i}-Q_{i} \Sigma_{i}^{-1}-P_{i r} H_{i}^{\prime} R_{i}^{-1} H_{i}\right] \hat{x}_{i r}+P_{i r} H_{i}^{\prime} R_{i}^{-1} y_{i}
$$

which is a backward Kalman filter, with covariance also calculated backward in time [with initial condition $P_{i r}(T)$ $\left.=\Sigma_{i}(T)\right]$ from either of the following equations:

$$
\begin{align*}
-P_{i r}= & -\left[A_{i}+Q_{i} \Sigma_{i}^{-1}\right] P_{i r}-P_{i r}\left[A_{i}+Q_{i} \Sigma_{i}^{-1}\right]^{\prime} \\
& +Q_{i}-P_{i r} H_{i}^{\prime} R_{i}^{-1} H_{i} P_{i r}  \tag{3.4}\\
-\frac{d}{d t}\left(P_{i r}^{-1}\right)= & P_{i r}^{-1}\left[A_{i}+Q_{i} \Sigma_{i}^{-1}\right]+\left[A_{i}+Q_{i} \Sigma_{i}^{-1}\right]^{\prime} P_{i r}^{-1} \\
& -P_{i r}^{-1} Q_{i} P_{i r}^{-1}+H_{i}^{\prime} R_{i}^{-1} H_{i} . \tag{3.5}
\end{align*}
$$

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

$$
\begin{equation*}
\hat{x}_{i s}=P_{i s}\left[P_{i j}^{-1} \hat{x}_{i f}+P_{i r}^{-1} \hat{x}_{i r}\right] \tag{3.6}
\end{equation*}
$$

where

$$
\begin{equation*}
P_{i s}^{-1}=P_{i f}^{-1}+P_{i r}^{-1}-\Sigma_{i}^{-1} . \tag{3.7}
\end{equation*}
$$

The overall smoothed estimate satisfies a similar set of equations

$$
\begin{align*}
\hat{x}_{s} & =P_{s}\left[P_{f}^{-1} \hat{x}_{f}+P_{r}^{-1} \hat{x}_{r}\right]  \tag{3.8}\\
P_{s}^{-1} & =P_{f}^{-1}+P_{r}^{-1}-\Sigma^{-1} \tag{3.9}
\end{align*}
$$

where $\Sigma$ is given by either of the equations

$$
\begin{align*}
\dot{\Sigma} & =A \Sigma+\Sigma A^{\prime}+Q  \tag{3.10}\\
\frac{d}{d t} \Sigma^{-1} & =-\Sigma^{-1} A-A^{\prime} \Sigma^{-1}-\Sigma^{-1} Q \Sigma^{-1} \tag{3.11}
\end{align*}
$$

Also, $P_{f}$ is given by (2.7) or equivalently by

$$
\begin{align*}
& \frac{d}{d t}\left(P_{f}^{-1}\right)=-P_{f}^{-1} A-A^{\prime} P_{f}^{-1}-P_{f}^{-1} Q P_{f}^{-1} \\
&  \tag{3.12}\\
& \quad+C_{1}^{\prime} R_{1}^{-1} C_{1}+C_{2}^{\prime} R_{2}^{-1} C_{2}
\end{align*}
$$

and $P_{r}$ satisfies both of the following:

$$
\begin{align*}
-\dot{P}_{r}= & -\left[A+Q \Sigma^{-1}\right] P_{r}-P_{r}\left[A+Q \Sigma^{-1}\right]^{\prime}+Q \\
& -P_{r} C_{1}^{\prime} R_{1}^{-1} C_{1} P_{r}-P_{r} C_{2}^{\prime} R_{2}^{-1} C_{2} P_{r}  \tag{3.13}\\
-\frac{d}{d t}\left(P_{r}^{-1}\right)= & P_{r}^{-1}\left[A+Q \Sigma^{-1}\right]+\left[A+Q \Sigma^{-1}\right]^{\prime} P_{r}^{-1} \\
& -P_{r}^{-1} Q P_{r}^{-1}+C_{1}^{\prime} R_{1}^{-1} C_{1}+C_{2}^{\prime} R_{2}^{-1} C_{2} . \tag{3.14}
\end{align*}
$$

Using the results of the previous section, we can calculate $\hat{x}_{f}$ in terms of $\hat{x}_{1 f}$ and $\hat{x}_{2 f}$ as in (2.13), (2.14), and by looking at the problem in reverse time, we can use the same results to compute $\hat{x}_{r}$ in terms of $\hat{x}_{1 r}$ and $\hat{x}_{2 r}$. The resulting equations are

$$
\begin{align*}
\hat{x}_{r} & =\xi_{r}+G_{1 r} \hat{x}_{1 r}+G_{2 r} \hat{x}_{2 r}  \tag{3.15}\\
-\dot{\xi}_{r} & =F_{r} \xi_{r}+K_{1 r} \hat{x}_{1 r}+K_{2 r} \hat{x}_{2 r}  \tag{3.16}\\
F_{r} & =-A-Q \Sigma^{-1}-P_{r} C_{1}^{\prime} R_{1}^{-1} C_{1}-P_{r} C_{2}^{\prime} R_{2}^{-1} C_{2} \tag{3.17}
\end{align*}
$$

$$
\begin{align*}
G_{i r}= & P_{r} M_{i}^{\prime} P_{i r}^{-1}  \tag{3.18}\\
K_{i r}= & {\left[P_{r} M_{i}^{\prime} P_{i r}^{-1} Q_{i} P_{i r}^{-1}-Q M_{i}^{\prime} P_{i r}^{-1}\right] } \\
& +\left\{P_{r} M_{i}^{\prime}\left[-A_{i}^{\prime}-\Sigma_{i}^{-1} Q_{i}\right] P_{i r}^{-1}\right. \\
& \left.-P_{r}\left[-A^{\prime}-\Sigma^{-1} Q\right] M_{i}^{\prime} P_{i r}^{-1}\right\} \\
& +P_{r} \dot{M}_{i}^{\prime} P_{i r}^{-1} . \tag{3.19}
\end{align*}
$$

From (2.13), (2.14), (3.8), (3.15), and (3.16), we now have an algorithm for calculating $\hat{x}_{s}$ from $\hat{x}_{1 f}, \hat{x}_{1 r}, \hat{x}_{2 f}$, and $\hat{x}_{2 r}$. What we would like is to compute $\hat{x}_{s}$ in terms of $\hat{x}_{1 s}$ and $\hat{x}_{2 s}$. To see when and how this can be done, we note first that straightforward algebra [using (2.12), (2.14), (3.6). (3.8), (3.15), and (3.18)] yields

$$
\begin{equation*}
\hat{x}_{s}=P_{s}\left[P_{f}^{-1} \xi_{f}+P_{r}^{-1} \xi_{r}+M_{1}^{\prime} P_{1 s}^{-1} \hat{x}_{1 s}+M_{2}^{\prime} P_{2 s}^{-1} \hat{x}_{2 s}\right] \tag{3.20}
\end{equation*}
$$

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 mentioned, they are correlated, and thus we have correction terms to account for this correlation.

We have now reduced the algorithm for calculating $\hat{x}_{s}$ to equations (2.13), (3.16), and (3.20). We have eliminated $\hat{x}_{i f}$ and $\hat{x}_{i r}$ and replaced them with $\hat{x}_{i s}$ in (3.20), but (2.13) and (3.16) still involve the forward and reverse estimates. Consequently, we are faced with something that resembles an inverse system problem: we want to express the term involving $\hat{x}_{i f}$ in (2.13) and the term involving $\hat{x}_{i r}$ in (3.16) by terms involving $\hat{x}_{i s}$. As we will see in the next two 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.

## B. The Special Case of Identical Local and Global Models

As in Section II-B, consider the case when (2.18) holds, i.e., when $x_{1}=x_{2}=x$. For this case we obtain considerable simplification. Also, in this context we can readily examine and solve the updating problem introduced in Section I in addition to the estimate combining problem.

Specializing (3.20) to the case when the local and global models are the same we obtain

$$
\begin{equation*}
\hat{x}_{s}=P_{s}\left[P_{f}^{-1} \xi_{f}+P_{r}^{-1} \xi_{r}+P_{1 s}^{-1} \hat{x}_{1 s}+P_{2 s}^{-1} \hat{x}_{2 s}\right] \tag{3.21}
\end{equation*}
$$

and, from Section II-B

$$
\begin{align*}
K_{i f} & =\left[P_{f} P_{i f}^{-1}-I\right] Q P_{i f}^{-1}  \tag{3.22}\\
K_{i r} & =\left[P_{r} P_{i r}^{-1}-I\right] Q P_{i r}^{-1} \tag{3.23}
\end{align*}
$$

We now see that the quantities we actually need in (2.13) and (3.16) are $Q P_{i f}^{-1} \hat{x}_{i f}$ and $Q P_{i r}^{-1} \hat{x}_{i r}$. In order to proceed, it is useful to define

$$
\begin{align*}
& z_{i f}=P_{i f}^{-\mathrm{i}} \hat{x}_{i f} \\
& z_{i r}=P_{i r}^{-1} \hat{x}_{i r} \\
& z_{i s}=P_{i s}^{-1} \hat{x}_{i s}=z_{i f}+z_{i r} . \tag{3.24}
\end{align*}
$$

Differentiating $z_{i f}$ and $z_{i r}$, using (2.1), (2.3), (3.3), and (3.10), and performing some algebra, we obtain

$$
\begin{align*}
& \dot{z}_{i f}=-\left(A^{\prime}+P_{i f}^{-1} Q\right) z_{i f}+C_{i}^{\prime} R_{i}^{-1} y_{i}  \tag{3.25}\\
& \dot{z}_{i r}=-\left(A^{\prime}+\Sigma^{-1} Q-P_{i r}^{-1} Q\right) z_{i r}-C_{i}^{\prime} R_{i}^{-1} y_{i} \tag{3.26}
\end{align*}
$$

If we add these last two equations and use (3.24) we obtain the following expressions:

$$
\begin{align*}
Q z_{i f} & =P_{i s}\left[-\dot{z}_{i s}-\left(A^{\prime}+\Sigma^{-1} Q-P_{i r}^{-1} Q\right) z_{i s}\right]  \tag{3.27}\\
Q z_{i r} & =P_{i s}\left[\dot{z}_{i s}+\left(A^{\prime}+P_{i f}^{-1} Q\right) z_{i s}\right] \tag{3.28}
\end{align*}
$$

From (2.13), (3.16), and (3.22)-(3.24), we see that we can use (3.27) and (3.28) in these equations to replace $\hat{x}_{i f}, \hat{x}_{i r}$ with $\hat{x}_{i s}$. Thus, in this case, we can obtain an algorithm of the desired form. Note that we haven't shown that we can recover $\hat{x}_{i f}, \hat{x}_{i r}$ from $\hat{x}_{i s}$, or equivalently $z_{i f}, z_{i r}$ from $z_{i s}$, but we have seen that we can recover $Q z_{i f}$ and $Q z_{i r}$, and this is all that we need for our problem. Note, however, that the expressions (3.27) and (3.28) for these quantities involve derivatives of $z_{i s}$. In order to avoid these, we must use a feedforward formulation. First of all, substituting (3.22) and (3.27) into (2.13) we obtain

$$
\begin{align*}
\dot{\xi}_{j}=F_{f} \xi_{f}-\sum_{i=1}^{2} & {\left[P_{f} P_{i f}^{-1}-I\right] } \\
& \cdot P_{i s}\left[\dot{z}_{i s}+\left(A^{\prime}+\Sigma^{-1} Q-P_{i r}^{-1} Q\right) z_{i z}\right] \tag{3.29}
\end{align*}
$$

From (3.24) we know that

$$
\begin{equation*}
P_{i s} \dot{z}_{i s}=\dot{\hat{x}}_{i s}-\dot{P}_{i s} P_{i s}^{-1} \hat{x}_{i s} \tag{3.30}
\end{equation*}
$$

Now define

$$
\begin{equation*}
q_{f}=\xi_{f}+\sum_{i=1}^{2}\left[P_{f} P_{i f}^{-1}-I\right] \hat{x}_{i s} \tag{3.31}
\end{equation*}
$$

Differentiating this, using (3.31) we obtain

$$
\begin{align*}
\dot{q}_{f}= & F_{f}\left[q_{f}-\sum_{i=1}^{2}\left(P_{f} P_{i f}^{-1}-I\right) \hat{x}_{i s}\right] \\
& +\sum_{i=1}^{2}\left[\frac{d}{d t}\left(P_{f} P_{i f}^{-1}\right)\right] \hat{x}_{i s}-\sum_{i=1}^{2}\left[P_{f} P_{i f}^{-1}-I\right] \\
& \cdot\left[-\dot{P}_{i s} P_{i s}^{-1}+P_{i s}\left(A^{\prime}+\Sigma^{-1} Q-P_{i r}^{-1} Q\right) P_{i s}^{-1}\right] \hat{x}_{i s} . \tag{3.32}
\end{align*}
$$

We have all of the equations needed to simplify this equation except for an expression for $\dot{P}_{i s}$. From (3.7)

$$
\begin{equation*}
\frac{d}{d t}\left(P_{i s}^{-1}\right)=\frac{d}{d t}\left(P_{i f}^{-1}\right)+\frac{d}{d t}\left(P_{i r}^{-1}\right)-\frac{d}{d t}\left(\Sigma^{-1}\right) \tag{3.33}
\end{equation*}
$$

and using expressions for the terms on the right-hand side [(2.3), (3.5), and (3.11), together with (2.18)], we obtain

$$
\begin{align*}
& \frac{d}{d t}\left(P_{i s}^{-1}\right)=-P_{i s}^{-1} A-A^{\prime} P_{i s}^{-1}-P_{i f}^{-1} Q P_{i s}^{-1} \\
&  \tag{3.34}\\
& \quad-P_{i s}^{-1} Q \Sigma^{-1}+P_{i s}^{-1} Q P_{i r}^{-1} .
\end{align*}
$$

A great deal of algebra then yields

$$
\begin{equation*}
\dot{q}_{f}=F_{f} q_{f}-P_{f} C_{2}^{\prime} R_{2}^{-1} C_{2} x_{1 s}-P_{f} C_{1}^{\prime} R_{1}^{-1} C_{1} x_{2 s} . \tag{3.35}
\end{equation*}
$$

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

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

$$
\begin{equation*}
q_{r}=\xi_{r}+\sum_{i=1}^{2}\left[P_{r} P_{i r}^{-1}-I\right] \hat{x}_{i s} . \tag{3.36}
\end{equation*}
$$

Then

$$
\begin{equation*}
-\dot{q}_{r}=F_{r} q_{r}-P_{r} C_{2}^{\prime} R_{2}^{-1} C_{2} x_{1 s}-P_{r} C_{1}^{\prime} R_{1}^{-1} C_{1} \hat{x}_{2 s} \tag{3.37}
\end{equation*}
$$

Substituting (3.31) and (3.36) into (3.21), we obtain

$$
\begin{align*}
\hat{x}_{s}=P_{s}\left\{P_{f}^{-1} q_{f}\right. & -\sum_{i=1}^{2}\left[P_{i f}^{-1}-P_{f}^{-1}\right] \hat{x}_{i s}+P_{r}^{-1} q_{r} \\
& \left.-\sum_{i=1}^{2}\left[P_{i r}^{-1}-P_{r}^{-1}\right] \hat{x}_{i s}+\sum_{i=1}^{2} P_{i s}^{-1} \hat{x}_{i s}\right\} . \tag{3.38}
\end{align*}
$$

Using (3.7) and (3.9), we obtain the following algorithm for combining smoothed estimates:

$$
\begin{align*}
\hat{x}_{s}= & P_{s}\left\{P_{f}^{-1} q_{f}+P_{r}^{-1} q_{r}\right\}+\hat{x}_{1 s}+\hat{x}_{2 s}  \tag{3.39}\\
\dot{q}_{f}= & F_{f} q_{f}-P_{f} C_{2}^{\prime} R_{2}^{-1} C_{2} \hat{x}_{1 s} \\
& -P_{f} C_{1}^{\prime} R_{1}^{-1} C_{1} \hat{x}_{2 s}  \tag{3.40}\\
-\dot{q}_{r}= & F_{r} q_{r}-P_{r} C_{2}^{\prime} R_{2}^{-1} C_{2} \hat{x}_{1 s} \\
& -P_{r} C_{1}^{\prime} R_{1}^{-1} C_{1} \hat{x}_{2 s} . \tag{3.41}
\end{align*}
$$

If we think of optimal estimates as orthogonal projections in spaces of random variables, then $\hat{x}_{1 s}$ is the projection of $x$ onto $\dot{y}_{1}$, the subspace spanned by the first pass measurements. Similarly, $\hat{x}_{2 s}$ is the projection onto $g_{2}$, and $x_{s}$ is the projection onto $\mathscr{y}_{1}+\mathscr{g}_{2}$. If $\mathscr{y}_{1}$ and $\mathscr{y}_{2}$ were orthogonal, i.e., independent, then $\hat{x}_{s}$ would equal $\hat{x}_{1 s}+\hat{x}_{2 s}$. However, they are not; and thus the other terms in (3.39) 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 $\hat{x}_{1 s}$ and the new data $\mathscr{g}_{2}$. The solution to this problem is readily obtained in a manner analogous to that used in deriving (2.16) and (2.17). That is, if we perform all of the analyses we have just done, leaving $y_{2}$
alone and only replacing $y_{1}$ by $\hat{x}_{1 f}, \hat{x}_{1 r}$, and eventually by $\hat{x}_{1 s}$, linearity guarantees that the input-output relation from $\hat{x}_{1 s}$ 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 as follows:

$$
\begin{align*}
\hat{x}_{s}= & P_{s}\left[P_{f}^{-1} r_{f}+P_{r}^{-1} r_{r}\right]+\hat{x}_{1 s}  \tag{3.42}\\
\dot{r}_{f}= & F_{f} r_{f}+P_{f} C_{2}^{\prime} R_{2}^{-1} y_{2} \\
& -P_{f} C_{2}^{\prime} R_{2}^{-1} C_{2} \hat{x}_{1 s}  \tag{3.43}\\
-\dot{r}_{r}= & F_{r} r_{r}+P_{r} C_{2}^{\prime} R_{2}^{-1} y_{2} \\
& -P_{r} C_{2}^{\prime} R_{2}^{-1} \hat{x}_{1 s} . \tag{3.44}
\end{align*}
$$

Here, if we let $\mathscr{F}$ denote the orthogonal complement of $\mathscr{O}_{1}$ in $\mathscr{Y}_{1}+\mathscr{Y}_{2}$-i.e., the part of $\mathscr{\mathscr { Y }}_{2}$ that is independent of $\mathscr{Y}_{1}$, then $\mathscr{y}_{1}+\mathscr{y}_{2}=\mathscr{y}_{1} \oplus \mathscr{F}$, and $\hat{x}_{1 s}$ is the projection of $x$ onto ${\sigma_{3}}_{1}$, while the remaining terms in (3.42) are the projection onto $\mathscr{F}$.

Note also that (3.43) and (3.44) can be rewritten in the following form:

$$
\begin{align*}
\dot{r}_{f} & =F_{f} r_{r}+P_{f} C_{2}^{\prime} R_{2}^{-1}\left[y_{2}-C_{2} \hat{x}_{1 s}\right]  \tag{3.45}\\
-\dot{r}_{r} & =F_{r} r_{r}+P_{r} C_{2}^{\prime} R_{2}^{-1}\left[y_{2}-C_{2} \hat{x}_{1 s}\right] . \tag{3.46}
\end{align*}
$$

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 correction terms are the projection of the estimation error $\left(x-\hat{x}_{1 s}\right)$ onto $\mathscr{\mathscr { y }}_{2}$. 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. Alternate forms for these realizations are given in [16] and [18].

## C. 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 (2.13), (3.16), and (3.20), together with (2.12) and (3.18) which define the gains needed in the algorithm. Also, we have differential equations for $z_{i f}, z_{i r}$, and $z_{i s}$ as defined in (3.24), but with slight modifications due to the fact that we have local models. Specifically, we have

$$
\begin{align*}
& \dot{z}_{i f}=-\left(A_{i}^{\prime}+P_{i f}^{-1} Q_{i}\right) z_{i f}+H_{i}^{\prime} R_{i}^{-1} y_{i}  \tag{3.47}\\
& \dot{z}_{i r}=-\left(A_{i}^{\prime}+\sum_{i}^{-1} Q_{i}-P_{i r}^{-1} Q_{i}\right) z_{i r}-H_{i}^{\prime} R_{i}^{-1} y_{i} \tag{3.48}
\end{align*}
$$

which lead to the equations

$$
\begin{equation*}
Q_{i} z_{i f}=P_{i s}\left[-\dot{z}_{i s}-\left(A_{i}^{\prime}+\Sigma_{i}^{-1} Q_{i}-P_{i r}^{-1} Q_{i}\right) z_{i s}\right] \tag{3.49}
\end{equation*}
$$

$$
\begin{equation*}
Q_{i} z_{i r}=P_{i s}\left[\dot{z}_{i s}+\left(A_{i}^{\prime}+P_{i f}^{-1} Q_{i}\right) z_{i s}\right] . \tag{3.50}
\end{equation*}
$$

In order to proceed as we did in the preceding subsection, we wish to be able to find some matrices $L_{i f}, L_{i r}$ so that

$$
\begin{align*}
& K_{i f} \hat{x}_{i f}=L_{i f} Q_{i} z_{i f}=L_{i f} Q_{i} P_{i f}^{-1} \hat{x}_{i f}  \tag{3.51}\\
& K_{i r} \hat{x}_{i r}=L_{i r} Q_{i} z_{i r}=L_{i r} Q_{i} P_{i r}^{-1} \hat{x}_{i r} . \tag{3.52}
\end{align*}
$$

This will be possible if and only if

$$
\begin{array}{r}
N\left(Q_{i} P_{i f}\right)^{-1} \subset N\left(K_{i f}\right) \\
N\left(Q_{i} P_{i r}^{-1}\right) \subset N\left(K_{i r}\right) . \tag{3.54}
\end{array}
$$

Note that if (3.53) and (3.54) hold, then $L_{i f}$ and $L_{i r}$ can be chosen to be

$$
\begin{align*}
L_{i f} & =K_{i f} P_{i f} Q_{i}^{\dagger}  \tag{3.55}\\
L_{i r} & =K_{i r} P_{i r} Q_{i}^{\dagger} . \tag{3.56}
\end{align*}
$$

In this case, we can replace $K_{i f} \hat{x}_{i f}$ and $K_{i r} \hat{x}_{i r}$ in (2.13) and (3.16) by expressions involving $Q z_{i f}$ and $Q z_{i r}$, which can, in turn, be replaced by the right-hand sides of (3.49) and (3.50). Finally, we use a set of steps similar to those used in Section III-B to remove the $z_{i s}$-term. The result of these calculations is that if we define

$$
\begin{align*}
& q_{f}=\xi_{f}+\sum_{i=1}^{2} L_{i j} \hat{x}_{i s}  \tag{3.57}\\
& q_{r}=\xi_{r}+\sum_{i=1}^{2} L_{i r} \hat{x}_{i s} \tag{3.58}
\end{align*}
$$

then some algebra yields

$$
\begin{align*}
\dot{q}_{f} & =F_{f} q_{f}+N_{1 f} \hat{x}_{1 s}+N_{2 f} \hat{x}_{2 s}  \tag{3.59}\\
-\dot{q}_{r} & =F_{r} q_{r}+N_{1 r} \hat{x}_{1 s}+N_{2 r} \hat{x}_{2 s}  \tag{3.60}\\
\hat{x}_{s} & =P_{s}\left[P_{f}^{-1} q_{f}+P_{r}^{-1} q_{r}+D_{1} \hat{x}_{1 s}+D_{2} \hat{x}_{2 s}\right] \tag{3.61}
\end{align*}
$$

where

$$
\begin{align*}
N_{i f}= & -F_{f} L_{i f}+\dot{L}_{i f}+L_{i f} \dot{P}_{i s} P_{i s}^{-1} \\
& -L_{i f} P_{i s}\left(A_{i}^{\prime}+\sum_{i}^{-1} Q_{i}-P_{i r}^{-1} Q_{i}\right) P_{i s}^{-1}  \tag{3.62}\\
N_{i r}= & -F_{r} L_{i r}-\dot{L}_{i r}-L_{i r} \dot{P}_{i s} P_{i s}^{-1} \\
& +L_{i r} P_{i s}\left(A_{i}^{\prime}+P_{i f}^{-1} Q_{i}\right) P_{i s}^{-1}  \tag{3.63}\\
D_{i}= & M_{i}^{\prime} P_{i s}^{-1}-P_{f}^{-1} L_{i f}-P_{r}^{-1} L_{i r} . \tag{3.64}
\end{align*}
$$

The analysis in this case is clearly more complex, since the
equations involve $M_{i}$ and $Q_{i}$, and if these are time-varying, we will also have to consider their time derivatives.
Note that one obvious case in which (3.53) and (3.54) hold is when $Q_{i}$ is invertible. In this case, the smoother is essentially invertible, as we can recover $z_{i f}$ and $z_{i r}$ and consequently the $H_{i}^{\prime} R_{i}^{-1} y_{i}$. 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 $x_{1}$ is an actual part of the global state. As was discussed in Section II-C, in this case (assuming that $M_{1}$ is onto) we can choose a basis for the state space so that (2.30)-(2.37) hold. Consider the form for $Q$ given in (2.33). It is relatively easy to check that since $Q$ is a covariance matrix, there must be a matrix $T_{1}$ so that

$$
\begin{equation*}
Q_{12}^{\prime}=T_{1} Q_{1} \tag{3.65}
\end{equation*}
$$

(write $w^{\prime}=\left(w_{1}^{\prime}, w_{2}^{\prime}\right)$ and let $\hat{w}_{2}=T_{1} w_{1}$ be the best linear estimate of $w_{2}$ given $w_{1}$; since $w_{2}=\hat{w}_{2}+\left(w_{2}-\hat{w}_{2}\right)$ with $w_{2}-\hat{w}_{2}$ uncorrelated with $w_{1}$, the result follows). If we substitute (3.65) into (2.36), we find that

$$
\begin{equation*}
K_{1 f}=\binom{\left(P_{f}\right)_{11} P_{i f}^{-1}-I}{\left(P_{f}\right)_{12}^{\prime} P_{1 f}^{-1}-T_{1}} Q_{1} P_{1 f}^{-1} . \tag{3.66}
\end{equation*}
$$

Therefore

$$
\begin{equation*}
L_{1 f}=\binom{\left(P_{f}\right)_{11} P_{1 f}^{-1}-I}{\left(P_{f}\right)_{12}^{\prime} P_{1 f}^{-1}-T_{1}}=P_{f}\binom{P_{1 f}^{-1}}{0}-\binom{I}{T_{1}} . \tag{3.67}
\end{equation*}
$$

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 (2.30)-(3.34), (3.19) reduces to

$$
\begin{align*}
& K_{1 r}=\left[P_{r} M_{1} P_{1 r}^{-1} M_{1}-I\right] Q M_{1}^{\prime} P_{1 r}^{-1} \\
&+P_{r}\left[\Sigma^{-1} Q M_{1}^{\prime}-M_{1}^{\prime}-M_{1}^{\prime} \Sigma_{1}^{-1} Q_{1}\right] P_{1 r}^{-1} \tag{3.68}
\end{align*}
$$

where $M_{1}=[I \vdots 0], Q$ is given by (2.33), and $\Sigma$ and $\Sigma_{1}$ are related by

$$
\Sigma=\left[\begin{array}{ll}
\Sigma_{1} & \Sigma_{12}  \tag{3.69}\\
\Sigma_{12} & \Sigma_{22}
\end{array}\right]
$$

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

$$
\Sigma^{-1}=\left[\begin{array}{c:c}
\Sigma_{1}^{-1}+\Sigma_{1}^{-1} \Sigma_{12}\left[\Sigma_{22}-\Sigma_{12}^{\prime} \Sigma_{1}^{-1} \Sigma_{12}\right]^{-1} \Sigma_{12}^{\prime} \Sigma_{1}^{-1} & -\Sigma_{1}^{-1} \Sigma_{12}\left[\Sigma_{22}-\Sigma_{12} \Sigma_{1}^{-1} \Sigma_{12}\right]^{-1}  \tag{3.70}\\
-\left[\Sigma_{22}-\Sigma_{12}^{\prime} \Sigma_{1}^{-1} \Sigma_{12}\right]^{-1} \Sigma_{12}^{\prime} \Sigma_{1} & {\left[\Sigma_{22}-\Sigma_{12}^{\prime} \Sigma_{1}^{-1} \Sigma_{12}\right]^{-1}}
\end{array}\right] .
$$

Using these relationships and (3.65), (3.68) becomes

$$
\begin{align*}
K_{1 r}= & \binom{\left[\left(P_{r}\right)_{11} P_{1 r}^{-1}-I\right] Q_{1} P_{1 r}^{-1}}{\left[\left(P_{r}\right)_{12}^{\prime} P_{1 r}^{-1}-T_{1}\right] Q_{1} P_{1 r}^{-1}} \\
& +P_{r}\binom{\Sigma_{1}^{-1} \Sigma_{12}\left[\Sigma_{22}-\Sigma_{12}^{\prime} \Sigma_{1}^{-1} \Sigma_{12}\right]^{-1}}{-\left[\Sigma_{22}-\Sigma_{12}^{\prime} \Sigma_{1}^{-1} \Sigma_{12}\right]^{-1}} \\
& \cdot\left[\Sigma_{12}^{\prime} \Sigma_{1}^{-1}-T_{1}\right] Q_{1} P_{1 r}^{-1} . \tag{3.71}
\end{align*}
$$

Therefore,

$$
\begin{align*}
L_{1 r}= & {\left[\begin{array}{l}
\left(P_{r}\right)_{11} P_{1 r}^{-1}-I \\
\left(P_{r}\right)_{12}^{\prime} P_{1 r}^{-1}-T_{1}
\end{array}\right] } \\
& +\binom{\left(P_{r}\right)_{11} \Sigma_{1}^{-1} \Sigma_{12}-\left(P_{r}\right)_{12}}{\left(P_{r}\right)_{12}^{\prime} \Sigma_{1}^{-1} \Sigma_{12}-\left(P_{r}\right)_{22}} \\
& \cdot\left[\Sigma_{22}-\Sigma_{12}^{\prime} \Sigma_{1}^{-1} \Sigma_{12}\right]^{-1}\left[\Sigma_{12}^{\prime} \Sigma_{1}^{-1}-T_{1}\right] \\
= & P_{r}\binom{P_{1 r}^{-1}}{0}-\binom{I}{T_{1}}+P_{r}\left[\Sigma^{-1}\binom{I}{T_{1}}-\binom{\Sigma_{1}^{-1}}{0}\right] . \tag{3.72}
\end{align*}
$$

Comparing (3.71) and (3.72) with (3.66) and (3.67), we see that the first terms on the right-hand sides of (3.71) and (3.72) are analogous to the right-hand sides of (3.66) and (3.67), respectively. The additional terms in (3.71) and (3.72) represent the complication that arises in constructing reverse models for $x_{1}$ by itself (as used in the calculation of $\hat{x}_{1 r}$ and $\left.\hat{x}_{I s}\right)$ and a reverse model for $x=\left(x_{1}^{\prime}, \gamma^{\prime}\right)$, 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 [6],

$$
\begin{align*}
-\dot{x}_{1}(t) & =-\left(A_{1}+Q_{1} \Sigma_{1}^{-1}\right) x_{1}(t)-\tilde{w}_{1}(t)  \tag{3.73}\\
-\binom{\dot{x}_{1}(t)}{\gamma(t)} & =-\left\{\left(\begin{array}{ll}
A_{1} & 0 \\
A_{12} & A_{22}
\end{array}\right)+Q \Sigma^{-1}\right\}\binom{x_{1}(t)}{x_{2}(t)}-\tilde{w}(t) \tag{3.74}
\end{align*}
$$

where these models can be interpreted as generating the same sample paths as the forward model. Here $\tilde{w}_{\mathrm{I}}(t)$ 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, $\tilde{\tilde{w}}(t)$ 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

$$
\begin{equation*}
\tilde{w}_{\mathrm{I}}(t) \neq \tilde{w}_{1}(t) . \tag{3.75}
\end{equation*}
$$

An equivalent way of looking at this is to view (3.73) 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, (3.74) gives the best predictor of $x_{1}$ and $\gamma$ given the future of $x_{1}$ and $\gamma$. Now although going forward in time the future of $x_{1}$ is decoupled from the past of $\gamma$, the future of $\gamma$ does
depend on the past of $x_{1}$ [see (2.32): $A_{12}=0$, but $A_{21}$ need not be zero]. Therefore, if we want to predict the past of $x_{1}$, the future of $\gamma$ 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 (2.32) is zero, the ( 1,2 ) block of the reverse-time dynamics matrix in (3.74) 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 (3.71) and (3.72) reflect this fact. In fact, it is straightforward to check that (assuming the invertibility of $P_{r}, \Sigma_{1}$, and $\Sigma$ ) these extra terms will be zero if and only if

$$
\begin{equation*}
\left(\Sigma_{12}^{\prime} \Sigma_{1}^{-1}-T_{1}\right) Q_{1}=0 \tag{3.76}
\end{equation*}
$$

or, equivalently,

$$
\begin{equation*}
\Sigma_{12}^{\prime} \Sigma_{1}^{-1} Q_{1}-Q_{12}=0 \tag{3.77}
\end{equation*}
$$

which is exactly the condition required for $x_{1}$ to be the state of a subsystem both in forward and reverse time. It is relatively easy to see that this is the case if $A_{21}=Q_{12}=0$, since in that case $x_{1}$ and $\gamma$ are independent. Other essentially equally trivial cases can be found in which (3.77) is satisfied, but the condition is quite restrictive.

These observations notwithstanding, (3.67) and (3.72) allow us to replace $\hat{x}_{1 f}$ and $\hat{x}_{1 r}$ in (2.13) and (3.16) by expressions involving $\hat{x}_{\text {Is }}$ through the use of (3.59)-(3.64). A similar analysis can be performed for local processor no. 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 (2.32) will, in general, be different).

As a final comment, we note that we can also consider the case of map updating in which we wish to compute $\hat{x}_{s}$ from $\hat{x}_{1 s}$ and $y_{2}$. These results follow in much the same manner as those derived in Section III-B:

$$
\begin{align*}
\hat{x}_{s} & =P_{s}\left[P_{f}^{-1} r_{f}+P_{r}^{-1} r_{r}+D_{1} \hat{x}_{1 s}\right]  \tag{3.78}\\
r_{f} & =F_{f} r_{f}+N_{1 g} \hat{x}_{1 s}+P_{f} C_{2} R_{2}^{-1} y_{2}  \tag{3.79}\\
-\hat{r}_{r} & =F_{r} r_{r}+N_{1 r} \hat{x}_{I s}+P_{r} C_{2}^{\prime} R_{2}^{-1} y_{2} \tag{3.80}
\end{align*}
$$

where $D_{1}, N_{1 f}$, and $N_{1 r}$ are defined in (3.62)-(3.64). These equations hold whenever $L_{1 f}$ and $L_{1 r}$ 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 (2.30), (3.62)(3.64) are computed using $M_{1}=[I \vdots 0]$ and $L_{1 f}$ and $L_{1 r}$ defined in (3.67) and (3.72). In this case, some algebra yields

$$
\begin{align*}
P_{s} D_{1}= & \binom{I}{T_{1}}  \tag{3.81}\\
N_{1 f}= & -P_{f} C_{2}^{\prime} R_{2}^{-1} C_{2}\binom{I}{T_{1}}-\binom{0}{\dot{T}_{1}} \\
& +\binom{0}{A_{21}+A_{22} T_{1}-T_{1} A_{1}} \tag{3.82}
\end{align*}
$$

$$
\begin{align*}
N_{1 r}=- & P_{r} C_{2}^{\prime} R_{2}^{-1} C_{2}^{\prime}\binom{I}{T_{1}} \\
& +\left(P_{r} \Sigma^{-1}-I\right)\left[\binom{0}{\dot{T}_{1}}+\binom{0}{A_{21}+A_{22} T_{1}-T_{1} A_{1}}\right] \tag{3.83}
\end{align*}
$$

## IV. Real-Time Smoothing

A variance of the problem addressed in the preceding section is the real-time smoothing problem. In this case, from some previous survey we have observed $y_{1}(t)$ over the interval $[0, T]$ and have produced $\hat{x}_{1 s}(t)$. 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.,

$$
\begin{equation*}
\hat{x}_{r s}(t)=E\left[x(t) \mid y_{1}(\tau), 0 \leqslant \tau \leqslant T, y_{2}(\sigma), 0 \leqslant \sigma \leqslant t\right] \tag{4.1}
\end{equation*}
$$

in terms of $\hat{x}_{1 s}$ 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 incorporate the new 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., $x^{\prime}=\left(x_{1}^{\prime}, \gamma^{\prime}\right)$, 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 the other tracks not covered in $y_{1}$. Alternatively, $\gamma$ 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 $\gamma$ could represent errors induced in an inertial navigation system aboard a ship [2], [9]. In this case we want the (real-time) estimates of $\gamma$. Clearly, we can also model in this same way the case in which $\gamma$ 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

$$
\begin{equation*}
C_{2}(s)=H_{2}(s)=0 \quad t<s \leqslant T . \tag{4.2}
\end{equation*}
$$

Also, since we are attempting to compute $\hat{x}$ directly using $y_{2}$, we have essentially a smoothing update problem. Based on these observations, (2.13), (3.16), and (3.20) can be adapted to the present situation:

$$
\begin{align*}
\hat{x}_{r s} & =P_{r s}\left[P_{f}^{-1} \xi_{f}+P_{b}^{-1} \xi_{b}+M_{1}^{\prime} P_{1 s}^{-1} \hat{x}_{1 s}\right]  \tag{4.3}\\
\dot{\xi}_{f} & =F_{f} \xi_{f}+K_{1 f} \hat{x}_{1 f}+P_{f} C_{2}^{\prime} R_{2}^{-1} y_{2}  \tag{4.4}\\
-\dot{\xi}_{b} & =F_{b} \xi_{b}+K_{1 b} \hat{x}_{1 b} \tag{4.5}
\end{align*}
$$

where $P_{f}, P_{1 s}, F_{f}$, and $K_{1 f}$ are as before, and $P_{b}$ is the reverse error covariance for $x$ based on $y_{1}$ alone:

$$
\begin{align*}
-\dot{P}_{b}=-\left[A+Q \Sigma^{-1}\right] P_{b}-P_{b} & {\left[A+Q \Sigma^{-1}\right] } \\
& +Q-P_{b} C_{1}^{\prime} R_{1}^{-1} C_{1} P_{b} . \tag{4.6}
\end{align*}
$$

Also,

$$
\begin{align*}
P_{r s}^{-1}= & P_{f}^{-1}+P_{b}^{-1}-\Sigma^{-1}  \tag{4.7}\\
F_{b}= & -A-Q \Sigma^{-1}-P_{b} C_{1}^{\prime} R_{1}^{-1} C_{1}  \tag{4.8}\\
K_{l b}= & {\left[P_{b} M_{1}^{\prime} P_{1 r}^{-1} Q_{1} P_{1 r}^{-1}-Q M_{1}^{\prime} P_{1 r}^{-1}\right] } \\
& +P_{b} M_{1}^{\prime}\left[-A_{1}^{\prime}-\Sigma_{1}^{-1} Q_{1}\right] P_{1 r}^{-1} \\
& -P_{b}\left[-A^{\prime}-\Sigma^{-1} Q\right] M_{1}^{\prime} P_{1 r}^{-1}+P_{b} \dot{M}_{1}^{\prime} P_{1 r}^{-1} . \tag{4.9}
\end{align*}
$$

## Assuming that we can write

$$
\begin{align*}
& K_{i f} \hat{x}_{1 f}=L_{1 f} Q_{1} P_{1 f}^{-1} \hat{x}_{1 f}  \tag{4.10}\\
& K_{1 b} \hat{x}_{1 r}=L_{1 b} Q_{1} P_{1 r}^{-1} \hat{x}_{1 r}, \tag{4.11}
\end{align*}
$$

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

$$
\begin{align*}
\hat{x}_{r s} & =P_{r s}\left[P_{f}^{-1} q_{f}+P_{b}^{-1} q_{b}+E_{1} \hat{x}_{1 s}\right]  \tag{4.12}\\
\dot{q}_{f} & =F_{f} q_{f}+N_{l f} \hat{x}_{1 s}+P_{f} C_{2}^{\prime} R_{2}^{-1} y_{2}  \tag{4.13}\\
-\dot{q}_{b} & =F_{b} q_{b}+N_{1 b} x_{i s} \tag{4.14}
\end{align*}
$$

where

$$
\begin{align*}
E_{1}= & M_{1}^{\prime} P_{1 s}^{-1}-P_{f}^{-1} L_{1 f}-P_{b}^{-1} L_{1 b}  \tag{4.15}\\
N_{1 f}= & -F_{f} L_{1 f}+\dot{L}_{1 f}+L_{1 f} \dot{P}_{1 s} P_{1 s}^{-1} \\
& -L_{1 f} P_{1 s}\left(A_{1}^{\prime}+\Sigma_{1}^{-1} Q_{1}-P_{1 r}^{-1} Q_{1}\right) P_{1 s}^{-1}  \tag{4.16}\\
N_{1 b}= & -F_{b} L_{1 b}-\dot{L}_{1 b}-L_{1 b} \dot{P}_{1 s} P_{1 s}^{-1} \\
& +L_{1 b} P_{1 s}\left(A_{1}^{\prime}+P_{1 f}^{-1} Q_{1}\right) P_{1 s}^{-1} . \tag{4.17}
\end{align*}
$$

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. The, using the fact that

$$
\begin{equation*}
A=A_{1}, \quad Q=Q_{1}, \quad C_{1}=H_{1}, \quad M_{1}=I, \quad \Sigma_{1}=\Sigma \tag{4.18}
\end{equation*}
$$

and comparing (3.4) and (4.6), we see that

$$
\begin{equation*}
P_{b}=P_{1 r} . \tag{4.19}
\end{equation*}
$$

This not surprising, since $P_{b}$ is the error covariance for the estimate of $x$ based on the future of $y_{1}, P_{1 r}$ 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 (4.18) and (4.19) also imply is that

$$
\begin{equation*}
K_{1 b}=0 \tag{4.20}
\end{equation*}
$$

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 $\left\{y_{1}(s), s<t\right\}$, 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 $\hat{x}_{1 s}$. Also, for this case

$$
\begin{equation*}
K_{1 f}=\left[P_{f} P_{1 f}^{-1}-I\right] Q P_{1 f}^{-1} \tag{4.21}
\end{equation*}
$$

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

$$
\begin{align*}
\hat{x}_{r s} & =P_{r s} P_{f}^{-1} q_{f}+\hat{x}_{1 s}  \tag{4.22}\\
\dot{q}_{f} & =F_{f} q_{f}+P_{f} C_{2}^{\prime} R_{2}^{-1}\left[y_{2}-C_{2} \hat{x}_{1 s}\right] . \tag{4.23}
\end{align*}
$$

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 (2.30)-(2.34) hold. In this case (3.78)-(3.83) can be adapted to yield

$$
\begin{align*}
P_{r s} E_{1}= & \binom{I}{T_{1}}  \tag{4.24}\\
N_{1 f}= & -P_{f} C_{2}^{\prime} R_{2}^{-1} C_{2}\binom{I}{T_{1}} \\
& +\binom{0}{-\dot{T}_{1}+A_{21}+A_{22} T_{1}-T_{1} A_{1}}  \tag{4.25}\\
N_{1 b}= & \left(P_{b} \Sigma^{-1}-I\right)\binom{0}{-\dot{T}_{1}+A_{21}+A_{22} T_{1}-T_{1} A_{1}} \tag{4.26}
\end{align*}
$$

Equations (4.12)-(4.14) and (4.24)-(4.26) define the algorithm for real-time smoothing. Note that the new data $\left(y_{2}\right)$ is processed only forward in time, while the reverse processing could be precomputed, since it only involves $\hat{x}_{1 s}$. 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 $\hat{x}_{1 r}$ of $x_{1}$ based on $y_{1}$. This is very much like what was discussed in Section II-C when one wishes to estimate 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 III-C, $x_{1}$ is not a substate of $x$ in reverse time. If it were, then given $\hat{x}_{1 r}$, the top block of $N_{1 b}$ would have to be zero, since the best estimate of $x_{1}$ based on the future of $y_{1}$ would have to be $\hat{x}_{1 r}$. However, the first part of $N_{1 b}$ is not zero, reflecting the fact that the reverse dynamics for $x_{1}$ alone are different from those when $x_{1}$ is viewed as a subvector 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. We are motivated in this investigation by the problems of map updating and combining described in Section I. As a first step in solving these problems, we
examined their causal versions in Section II and obtained solutions under very general conditions. Basically, the only restriction on the local processing is that the model on which it is based must have as many degrees of freedom as there are in the observations that are to be processed locally. As we indicated, the results in Section II are of independent interest, as they represent the solution to a decentralized estimation problem. We briefly 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 state. 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 and on the rate at which information is transmitted, the question of what to transmit is a complex one. Specifically, given communication capacity and local computation constraints, the problem becomes one of which 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 $y_{i}(\tau), 0 \leqslant \tau \leqslant T$ is preserved in the local smoothed estimate history $x_{i s}(\tau), 0 \leqslant \tau \leqslant T$.
Beyond this there remains the issue of interpreting the results of Sections III and IV. The very simple form of the solution in some cases, such as in (3.45) and (3.46), 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, 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. The development of substance for this idea may provide the basic insight needed to understand our results from first principles.
Finally, it is our belief that the results presented in this paper can be of great value in the development of efficient algorithms for the mapping of spatially-distributed random process, but a significant amount of work remains to be done in examining the implications of our methods for problems involving more general random field models and measurement geometries than those discussed in Section I. In a subsequent paper we will use the techniques developed here to make some initial inroads into problems of this type and consequently to provide support for our belief.

## 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. Appl. Meteorol., vol. 17, pp. 1023-1033, July 1978.
[2] R. A. Nash, Jr. and S. K. Jordan, "Statistical geodosy-An engineering perspective," Proc. IEEE, vol. 66, pp. 532-550, May 1978.
[3] J. L. Speyer, "Computation and transmission requirements for a decentralized linear-quadratic-Gaussian control problem," IEEE Trans. Automat. Contr., vol. AC-24, pp. 266-269. Apr. 1979.
[4] J. E. Wall, Jr., A. S. Willsky, and N. R. Sandell, Jr., "On the fixed-interval smoothing problem," Stochastics, vol. 5. nos. 1, 2, pp. 1-42, 1981.
[5] F. C. Schweppe, Uncertain Dynamic Systems. Englewood Cliffs, N.I.: Prentice-Hall, 1973.
[6] G. Verghese and T. Kailath, "A further note on backwards Markovian models," IEEE Trans. Inform. Theory, vol. IT-25, pp. 121-124, Jan. 1971.
[7] C. Y. Chong, "Hierarchical estimation," presented at the 2nd MIT/ONR Workshop Distributed Inform. Decision Syst. Motivated Naval Command-Contr.-Commun. ( $\mathrm{C}^{3}$ ) Problems. Naval Postgraduate School, Monterey, CA, July 1979.
[8] S. Attasi, "Modelling and recursive estimation for double indexed sequences," in System Identification: Advances and Case Studies, R. K. Mehra and D. G. Lainiotis, Eds. New York: Academic, 1976.
[9] A. S. Willsky and N. R. Sandell, Jr., "The stochastic analysis of dynamic systems moving through random fields," this issue, pp. 830-838.
[10] A. H. Haddad, "Linear filtering of singularly perturbed systems." IEEE Trans. Automat. Contr., vol. AC-21. pp. 515-519, Aug. 1976.
[11] D. Teneketzis and N. R. Sandell, Jr.. "Linear regulator design for
stochastic systems by a multiple time scales method," IEEE Trans. Automat. 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," in Proc. 1979 JACC. Denver, CO. pp. 629-633.
[13] L. Ljung, T. Kailth, and B. Friedlander, "Scattering theory and linear least squares estimation-Part I: Continuous-time problems," Proc. IEEE, vol. 64, pp. 131-139, Jan. 1976.
[14] R. F. Brammer, "Estimation of the ocean geoid near the Blake escarpment using GEOS-3 satellite altimetry data." J. Geophys. Res., submitted for publication.
[15] G. S. Sidhu and U. B. Desai, "Reversed-time lumped models, smoothing algorithms and Fredholm resolvents." in Proc. Johns Hopkins Conf. Inform. Sci. Syst., 1976, pp. 168-173.
[16] M. Bello. "Centralized and decentralized map updating and terrain masking analysis," Ph.D. dissertation. Dep. Elec. Eng. Comput. Sci., M.I.T., Cambridge, MA, Aug. 1981.
[17] S. R. Powell and L. M. Silverman, "Modeling of two-dimensional covariance functions with application to image restoration," IEEE Trans. Automat. Contr., vol. AC-19, pp. 8-13, 1974.
[18] A. S. Willsky, M. Bello, D. A. Castanon, B. C. Levy, and G. Verghese, "Combining and updating of local estimates and regional maps along sets of one-dimensional tracks," M.I.T. Lab. Inform. Decision Syst., Cambridge. MA, Rep. LIDS-P-955, Nov. 1979.


Alan S. Willsky (S'70-M'73) was born in Newark, NJ, on March 16, 1948. He received the S.B. degree in aeronautics and astronautics and the Ph.D. degree in instrumentation and control, both from the Massachusetts Institute of Technology, Cambridge, MA, in 1969 and 1973, respectively.
From 1969 through 1973 he held a Fannie and John Hertz Foundation Fellowship. He joined the faculty of the M.I.T. Department of Electrical Engineering and Computer Science in 1973 and is presently an Associate Professor of Electrical Engineering. From 1974 through 1981 he was Assistant Director of the M.I.T. Laboratory for Information and Decision Systems. From February through June of 1977 he was Science Research Council Senior Visiting Fellow at Imperial College, London, England, and from September 1980 through January 1981 he was a Professeur Associe at the Universite de Paris-Sud, Orsay, France. He is the author of one book, Digital Signal Processing and Control and Estimation Theorv: Points of Tangency, Areas of Intersection, and Parallel Directions (M.I.T. Press, 1979). The paper, "Relationships Between Digital Signal Processing and Control and Estimation Theory" (Proc. IEEE, vol. 66, Sept. 1978), for which he was awarded the 1979 Alfred Noble Prize by the American Society of Civil Engineers and the 1980 Browder J. Thompson Memorial Prize Award by the IEEE, provides an introduction to many of the topics addressed in the book. He is presently working on an undergraduate text, Signals and Systems, coauthored with Prof. A. V. Oppenheim and Dr. I. T. Young, to be published by Prentice-Hall in 1982. He has written numerous papers on algebraic system theory, nonlinear filtering, failure detection, stochastic processes, and biomedical signal processing. His present research interests are in problems involving abrupt changes in signals and systems and the related problems of detection and reliability, the modeling and processing of spatially-distributed random data, digital signal processing issues in control system design, and the analysis of stochastic processes and systems.

Dr. Willsky is a founder and member of the Board of Directors of Alphatech. Inc. In 1975 he received an award from the M.I.T. Graduate Student Council for outstanding teaching, and in August 1975 he received the Donald P. Eckman Award from the American Automatic Control Council. He is Editor of the M.I.T. Press series of books on signal processing, optimization, and control, was the Program Chairman for the 17th IEEE Conference on Decision and Control held in San Diego, CA, in January 1979, is an Associate Editor of the journals Stochastics and Control and Systems Letters, is a member of the Administrative Committee of the IEEE Control Systems Society, and was the Control Systems Society Program Chairman for the 1981 Bilateral Seminar on Control Systems sponsored by the IEEE Control Systems Society and the Chinese Association for Automation. He gave the opening plenary lecture at the 20th IEEE Conference on Decision and Control. He is a member of SIAM. AAAS, Sigma Xi, Sigma Gamma Tau, and Tau Beta Pi.


Martin G. Bello was born in Brookline, MA, on October 7, 1953. He received the B.S. degree in electrical engineering from Cornell University, Ithaca, NY, in 1975, and the M.S. and Ph.D. degrees, also in electrical engineering, from the Massachusetts Institute of Technology, Cambridge, in 1977 and 1981, respectively.

He worked as a Research Assistant at M.I.T. while completing his M.S. and Ph.D. theses, entitled "The Estimation of Delay Gradients for Purposes of Routing in Data-Communications Networks" and "Centralized and Decentralized Map-Updating and Terrain Masking Analysis," respectively. He is currently employed by Alphatech, Inc., Burlington, MA.

Dr. Bello received the Sibley Prize for academic excellence while at Cornell University. He is a member of Tau Beta Pi, Eta Kappa Nu, and Sigma Xi.


Bernard C. Levy (S'74-S'78-M'78) was born in Princeton, NJ, on July 31, 1951. He received the diploma of Ingenieur Civil des Mines from the Ecole Nationale Superieure des Mines, Paris, France, and the Ph.D. degree in electrical engineering from Stanford University, Stanford, CA.

While at Stanford University, he held an IRIA Fellowship and also worked as Teaching Assistant, Research Assistant, and Instructor. Since Fall 1978, he has been associated with the Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, first as a Research Associate and then, since June 1979, as Assistant Professor of Electrical Engineering and Computer Science. His research interests are in the area of control systems, estimation, and multidimensional signal processing.


David A. Castanon (S'68-M'79) was born in La Habana, Cuba, on November 24, 1950. He received the B.S. degree in electrical engineering. from Tulane University, New Orleans, LA, in 1971 and the Ph.D. degree in applied mathematics from the Massachusetts Institute of Technology, Cambridge, in 1976.

From 1976 to 1981 he was a Research Associate with the Laboratory for Information and Decision Systems at M.I.T. He is presently with Alphatech, Inc., Burlington, MA. His research interests include stochastic control and estimation, game theory, and qualitative properties of stochastic large scale systems.


George C. Verghese (S'74-Mㄱ8) was born in Addis Ababa, Ethiopia, on August 15, 1953. He received the B. Tech. degree in electronics from the Indian Institute of Technology, Madras, India, in 1974, the M.S. degree in electrical science from the State University of New York, Stony Brook, in 1975, and the Ph.D. degree in electrical engineering from Stanford University, Stanford, CA, in 1978.
Since January 1979 he has been Assistant Professor of Electrical Engineering at the Massachusetts Institute of Technology and the Electric Power Systems Engineering Laboratory. His research interests are in the areas of systems, control, and estimation, and their applications to power systems and power electronics.

# Square Root Covariance Ladder Algorithms 

BOAZ PORAT, sTUDENT MEMBER, IEEE, BENJAMIN FRIEDLANDER, SENIOR MEMBER, IEEE, AND MARTIN MORF, MEMBER, IEEE


#### Abstract

Square root normalized ladder algorithms provide an efficient recursive solution to the problem of multichannel autoregressive model fitting. A simplified derivation of the general update formulas for such ladder forms is presented, and is used to develop the growing memory and sliding memory covariance ladder algorithms. New ladder form realizations for the identified models are presented, leading to convenient methods for computing the model parameters from estimated reflection coefficients. A complete solution to the problem of possible singularity in the ladder update equations is also presented.


Manuscript received January 19, 1981; revised September 14, 1981. Paper recommended by H. V. Poor, Chairman of the Estimation Committee. This work was supported by the Defense Advanced Research Projects Agency Contract MDA903-80-C-0331 and NSF Grant ENG-78-1003.
B. Porat and M. Morf are with the Information Systems Laboratory, Stanford University, Stanford, CA 94305.
B. Friedlander is with Systems Control Technology, Inc., Palo Alto, CA 94304.

## I. Introduction

FITTING AN autoregressive (AR) model to an observed data sequence is a ubiquitous problem in the areas of estimation, system identification, and signal processing. Applications of AR modeling include speech processing [1], spectral estimation [2], time series analysis [3], and adaptive filtering [4]. The problem is typically formulated as follows. Let us assume that the observed sequence of $p$-vectors $\{y(t)\}$ is generated by a multichannel AR model of the type

$$
\begin{equation*}
y(t)+\sum_{i=1}^{N} A_{i} y(t-i)=u(t) \tag{1.1}
\end{equation*}
$$


[^0]:    Manuscript received December 14, 1979; revised August 6, 1981. Paper recommended by A. Ephremides, Past Chairman of the Estimation Committee.
    A. S. Willsky is with the 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. and was supported in part by the Air Force Office of Scientific Research under Grant AFOSR-77-3281B and by the National Science Foundation under Grant ECS-8012668.
    M. G. Bello was with the Laboratory for Information and Decision Systems and the Department of Electrical Engineering and Computer Science, M.I.T., Cambridge, MA 02139. He is now with Alphatech, Inc., Burlington, MA 01803. 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, Lexington, MA 02173.
    D. A. Castanon was with the Laboratory for Information and Decision Systems and the Department of Electrical Engineering and Computer Science, M.I.T., Cambridge, MA 02139. He is now with Alphatech, Inc., Burlington, MA 01803. The work of this author was performed at M.I.T. and was supported in part by the Office of Naval Research under Contract ONR/N00014-77-C-0532 and by the National Science Foundation under Grant ECS-8012668.
    B. C. Levy is with the Laboratory for Information and Decision Systems, M.I.T., Cambridge, MA 02139. The work of this author was performed at M.I.T. and was supported by the National Science Foundation under Grant ECS-8012668.
    G. C. Verghese is with the Electric Power Systems Engineering Laboratory and the Department of Electrical Engineering and Computer Science, M.I.T., Cambridge, MA 02139. The work of this author was supported in part by the National Science Foundation under Grant ECS-8012668.

[^1]:    ${ }^{1}$ From this point on, 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.

[^2]:    ${ }^{2}$ Note that we have implicitly made one other assumption about the local models, in that in (2.8) we are assuming that $P_{i f}$ is invertible. This will be guaranteed as long as $\Sigma_{i}(0)$ is invertible.

[^3]:    ${ }^{3}$ Note that in (2.15) 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 $M_{\mathrm{I}}$ and $M_{2}$ can be taken to be constants.

[^4]:    ${ }^{4}$ In the following discussion we use the notation developed previously. Thus $\hat{x}_{1 /}$ refers to the local estimate of $x_{1}$ given $y_{1}$ ( $P_{1 f}$ is its locally-computed covariance) and $\hat{x}_{f}$ refers to the global estimate of $x$ given $y_{1}$ and $y_{2}$ (global covariance $P_{f}$ ). In the particular case being examined here, $x^{\prime}=$ ( $x_{i}^{\prime} \vdots \gamma^{\prime}$ ) and therefore there is some chance of confusion. We have attempted to reduce this chance by using $\hat{x}_{1 /}$ and $\hat{x}_{\text {f }}$ only in the sense described above. Also. we have denoted the upper left-hand block of $P_{f}$ by $\left(P_{f}\right)_{11}$ [see (2.37)] to distinguish it from $P_{1 /}$. Here $\left(P_{f}\right)$ is the estimation error covariance of $x_{1}$ given $y_{1}$ and $y_{2}$, while $P_{1 j}$ is the error covariance based only on $y_{1}$.

[^5]:    ${ }^{5}$ Here $\Phi_{1}$ is the state transition matrix associated with $A_{1}$. Similarly, $\Phi_{22}$ is the state transition matrix for $A_{22}$.

