INT. J. CONTROL, 1989, VOL. 50, NO. 1, 203–223

Construction and applications of discrete-time smoothing error models

MARTIN G. BELLO†, ALAN S. WILLSKY‡ and BERNARD C. LEVY§

We present a unified perspective on techniques for constructing both forward and backward markovian realizations of the error process associated with discrete-time fixed-interval smoothing. Several alternative approaches are presented, each of which provides additional insight and connections with other results in linear estimation theory. In addition, two applications of such models are described. The first is a problem in the validation of an error model for a measurement system, while the second concerns the problems of updating and combining smoothed estimates. We also present an example of the use of our updating solution for the mapping of random fields based on several sets of data tracks.

1. Introduction

This paper is concerned with the characterization, derivation, and application of markovian models for the error in fixed-interval smoothed estimates. Models of this type are required in several applications, including two that are investigated in the last few sections of this paper. In particular, we describe a model validation process in which one set of data is used to evaluate the validity of a noise model for a second set of data. We also describe a solution to estimate updating and combining problems. In the former, a smoothed estimate based on one set of data is updated with a new set of data. In the latter, estimates based on separate data sets are combined. Such problems arise in the construction of maps of random fields, and we describe an application of this type in which the data sets represent non-coincident and non-parallel tracks across a two-dimensional random field.

In each of these applications, smoothing error models are required: in the first problem in order to compute the likelihood function for the validity of the specified model; in the second, to specify a model for the error in the map based on the first set of data which is then used as the basis for incorporating the new data.

We consider the construction of smoothing error models from four perspectives:

(i) the use of Martingale difference decompositions of the process noise associated with the state dynamics:

(ii) the use of the Bierman square-root information filtering framework (Bierman 1977), and the Dyer–McReynolds smoothed estimate recursions derived in Bierman (1974);

(iii) the use of a backward markovian model for one-step prediction estimate errors;

Received 26 September 1988.

† The Analytic Sciences Corporation, Reading, Massachusetts, U.S.A.

‡ Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA 02139, U.S.A.

§ Department of Electrical and Computer Engineering, University of California at Davis, Davis, CA 95516, U.S.A.

0020-7179/89 $3.00

(iv) the use of Weinert and Desai's 'Method of Complementary Models' (Weinert and Desai 1981).

The continuous-time counterpart of approach (i) was employed in Bello *et al.* (1986) and Bello (1981) in order to obtain forward and backward Markov models for fixed-interval smoothed estimate errors. The second approach (ii), was employed by Bierman (1974) in order to derive the Dyer–McReynolds backward recursion for fixed-interval smoothed estimate error covariance matrices. In our development, we explicitly identify some of the quantities derived in Bierman (1974) with a backward smoothing error model, thereby exposing the relationship between backward smoothing error models and the Dyer–McReynolds recursions.

The continuous-time counterpart of approach (iii) was described in Bello *et al.* (1986). In addition, Badawi (1980), briefly suggests the use of approach (iii) for obtaining a backward smoothing error model associated with discrete-time problems, but does not proceed to develop the detailed results. In the development of approach (iii) that follows, it will be seen that the invertibility of the one-step prediction error dynamics matrix is a necessary assumption for the construction of a backward one-step prediction error model that has the same sample paths as a corresponding forward model. However, without this invertibility assumption, we show that approach (iii) can be modified in order to construct a backward smoothing error model that satisfies the weaker requirement of having the same second-order properties as a corresponding forward model.

Approaches (i), (ii), and (iii) described above are employed to specify backward markovian representations for the smoothing error process. Approach (iv), however, involves the use of Weinert and Desai's 'Method of Complementary Models' (Weinert and Desai 1981), in order to obtain a forward markovian representation for the smoothing error process.

In the next section we introduce some notation and perform some preliminary calculations. Sections 3–6 are then devoted to the four approaches indicated previously. In § 7 we describe the model validation application, while § 8 is concerned with the updating and combining problems; the solution to the latter exposes the connection of these results to the theory of oblique projections (Desai and Kiaei 1985). In § 9, we apply our updating results to the problem of mapping a random field given two data sets along non-parallel tracks across the field. We conclude in § 10 with some brief remarks.

2. Notation and preliminary calculations

We consider the following state model

$$\mathbf{x}(t+1) = F(t)\mathbf{x}(t) + B_1(t)\mathbf{v}(t) + B_2(t)\mathbf{w}(t) \tag{1}$$

$$\mathbf{y}(t) = H(t)\mathbf{x}(t) + D(t)\mathbf{v}(t) \tag{2}$$

where $\mathbf{v}(t)$ and $\mathbf{w}(t)$ are uncorrelated, zero-mean gaussian white noise processes with identity covariance and are also uncorrelated with the zero-mean, gaussian initial condition $\mathbf{x}(0)$. We also assume that $P(0)$, the covariance of $\mathbf{x}(0)$, is positive definite, so that the state covariance $P(t)$, which satisfies

$$P(t+1) = F(t)P(t)F'(t) + B_1(t)B_1'(t) + B_2(t)B_2'(t) \tag{3}$$

is positive definite for all t in the interval $[0, T]$ of interest. We also assume that $D(t)$ is square and invertible, so that

$$R(t) \triangleq D(t)D'(t) > 0 \tag{4}$$

We now consider several state estimates, each of which can be viewed as a projection onto an appropriate Hilbert space of random variables. Specifically if $\mathbf{z}(t)$ is a zero-mean process on $[0, T]$, let $H(\mathbf{z})$ denote the Hilbert space spanned by the components of $\mathbf{z}(\tau)$, $0 \leqslant \tau \leqslant T$, let $H_t^-(\mathbf{z})$ be the corresponding space using only $\mathbf{z}(\tau)$, $0 \leqslant \tau \leqslant t$ while $H_t^+(\mathbf{z})$ uses $\mathbf{z}(\tau)$, $t \leqslant \tau \leqslant T$. Also, let $E(\xi|H)$ denote the projection of a zero-mean random vector ξ onto the Hilbert subspace H. With this notation, the smoothed estimate of $\mathbf{x}(t)$ is

$$\hat{\mathbf{x}}_s(t) = [\mathbf{x}(t)|H(\mathbf{y})] \tag{5}$$

the one-step-ahead predicted estimate is given by

$$\hat{\mathbf{x}}_p(t) = E[\mathbf{x}(t)|H_{t-1}^-(\mathbf{y})] \tag{6}$$

and the filtered estimate by

$$\hat{\mathbf{x}}_f(t) = E[\mathbf{x}(t)|H_t^-(\mathbf{y})] \tag{7}$$

The corresponding estimation errors are denoted with tildes, e.g.

$$\tilde{\mathbf{x}}_s(t) = \mathbf{x}(t) - \hat{\mathbf{x}}_s(t) \tag{8}$$

Next, we recall the standard Kalman filtering formulae for $\hat{\mathbf{x}}_p(t)$:

$$\left.\begin{aligned} \hat{\mathbf{x}}_p(t+1) &= F(t)\hat{\mathbf{x}}_p(t) + K(t)\mathbf{v}(t) \\ \mathbf{x}_p(0) &= \mathbf{0} \end{aligned}\right\} \tag{9}$$

where $\mathbf{v}(t)$ is the innovations process

$$\mathbf{v}(t) = \mathbf{y}(t) - H(t)\hat{\mathbf{x}}_p(t) = H(t)\tilde{\mathbf{x}}_p(t) + D(t)v(t) \tag{10}$$

and the gain $K(t)$ is given by

$$K(t) = [F(t)P_p(t)H'(t) + B_1(t)D'(t)]V^{-1}(t) \tag{11}$$

where $V(t)$, the covariance of $\mathbf{v}(t)$ is given by

$$V(t) = H(t)P_p(t)H'(t) + R(t) \tag{12}$$

and $P_p(t)$ is the covariance of the error $\tilde{\mathbf{x}}_p(t)$.

Combining (1), (2), (9), and (10) we obtain a *forward* model for $\tilde{\mathbf{x}}_p(t)$

$$\tilde{\mathbf{x}}_p(t+1) = \Gamma(t)\tilde{\mathbf{x}}_p(t) + [G(t) \mid B_2(t)]\begin{bmatrix} \mathbf{v}(t) \\ \mathbf{w}(t) \end{bmatrix} \tag{13}$$

where

$$\Gamma(t) = F(t) - K(t)H(t) \tag{14}$$

$$G(t) = B_f(t) - K(t)D(t) = -\Gamma(t)P_p(t)H'(t)(D'(t))^{-1} \tag{15}$$

where the last equality in (15) follows after some additional algebra.

Note also that (1) can be rewritten as

$$\mathbf{x}(t+1) = A(t)\mathbf{x}(t) + \zeta(t) \tag{16}$$

where

$$A(t) = F(t) - B_1(t)D^{-1}(t)H(t) \tag{17}$$

$$\zeta(t) = B_1(t)D^{-1}(t)\mathbf{y}(t) + B_2(t)\mathbf{w}(t) \tag{18}$$

From this we directly deduce

$$\hat{\mathbf{x}}_p(t+1) = A(t)\hat{\mathbf{x}}_f(t) + B_1(t)D^{-1}(t)\mathbf{y}(t) \tag{19}$$

and

$$P_p(t+1) = A(t)P_f(t)A'(t) + B_2(t)B_2'(t) \tag{20}$$

Also, from the standard formula (Jazwinski 1970),

$$\hat{\mathbf{x}}_f(t) = P_f(t)[H'(t)R^{-1}(t)\mathbf{y}(t) + P_p^{-1}(t)\hat{\mathbf{x}}_p(t)] \tag{21}$$

together with (19) we find that

$$\Gamma(t) = A(t)P_f(t)P_p^{-1}(t) \tag{22}$$

Finally, we note a particular decomposition of $\tilde{\mathbf{x}}_s(t)$ that we will find of value. Specifically, the basic property of the innovations sequence $\mathbf{v}(t)$ gives us an orthogonal decomposition of $H(\mathbf{y})$

$$H(\mathbf{y}) = H_{t-1}^-(\mathbf{y}) \oplus H_t^+(\mathbf{v}) \tag{23}$$

Writing

$$\mathbf{x}(t) = \tilde{\mathbf{x}}_p(t) + \hat{\mathbf{x}}_p(t) \tag{24}$$

projecting onto $H(\mathbf{y})$, subtracting the result from (24), and using the fact that $\tilde{\mathbf{x}}_p(t)$, $H_{t-1}^-(\mathbf{y})$, and $H_t^+(\mathbf{v})$ are all orthogonal, we obtain

$$\tilde{\mathbf{x}}_s(t) = \tilde{\mathbf{x}}_p(t) - E[\tilde{x}_p(t)|H_t^+(\mathbf{v})] \tag{25}$$

3. Backwards markovian smoothing error models using a martingale difference decomposition

In this section we sketch the discrete-time counterpart of the continuous-time approach, presented in Bello *et al.* (1986), using martingale decompositions to derive markovian models for the smoothing error. Throughout this section we assume that $\Gamma(t)$, or equivalently $A(t)$, is invertible. (This will be true, for example, if the system matrix $F(t)$ is invertible and $(F(t), B_2(t))$ is a reachable pair.)

To begin, we take the projection of both sides of (16) onto $H(\mathbf{y})$, subtract the result from (16) and use the invertibility of $A(t)$ to obtain

$$\tilde{\mathbf{x}}_s(t) = A^{-1}(t)\tilde{\mathbf{x}}_s(t+1) - A^{-1}(t)\tilde{\zeta}(t) \tag{26}$$

where

$$\tilde{\zeta}(t) = \zeta(t) - E[\zeta(t)|H(\mathbf{y})] \tag{27}$$

This is not yet a backward markovian model, as $\tilde{\zeta}(t)$ is neither white nor independent of $\tilde{\mathbf{x}}_s(t)$. To obtain a markovian model, we need a backward martingale decomposition of $\tilde{\zeta}(t)$ with respect to an appropriately chosen σ-field, i.e. one consisting of all the process $\mathbf{y}(\cdot)$ and the 'future' of $\tilde{\mathbf{x}}_s(\cdot)$.

Specifically, let

$$\mathscr{F}_t = H(y) \oplus H_{t+1}^+(\tilde{\mathbf{x}}_s) \tag{28}$$

and write

$$\tilde{\zeta}(t) = E[\tilde{\zeta}(t)|\mathscr{F}_t] + \mu(t) \tag{29}$$

where, as shown in Segal (1976), $\mu(t)$ is a discrete-time white noise sequence, independent of the future as represented by $\mathscr{F}_t$. It remains then to calculate the covariance of $\mu(t)$ and the expectation in (29). From (27), the independence of $H(\mathbf{y})$ and $H_{t+1}^+(\tilde{\mathbf{x}}_s)$, and (18), we have that

$$\begin{aligned} E[\tilde{\zeta}(t)|\mathscr{F}_t] &= E[\tilde{\zeta}(t)|H_{t+1}^+(\tilde{\mathbf{x}}_s)] \\ &= B_2(t)E[\mathbf{w}(t)|H_{t+1}^+(\tilde{\mathbf{x}}_s)] \end{aligned} \tag{30}$$

$$\mu(t) = B_2(t)\{\mathrm{w}(t) - E[\mathbf{w}(t)|\mathscr{F}_t]\} \tag{31}$$

As shown in Appendix A,

$$E[\mathbf{w}(t)|H_{t+1}^+(\tilde{\mathbf{x}}_s)] = B_2'(t)P_p^{-1}(t+1)\tilde{\mathbf{x}}_s(t+1) \tag{32}$$

$$E[\mathbf{w}(t)|H(\mathbf{y})] = B_2'(t)P_p^{-1}(t+1)[\tilde{\mathbf{x}}_p(t+1) - \tilde{\mathbf{x}}_s(t+1)] \tag{33}$$

and, of course, $E[\mathbf{w}(t)|\mathscr{F}_t]$ is simply the sum of (32) and (33), so that

$$\mu(t) = B_2(t)[\mathbf{w}(t) - B_2'(t)P_p^{-1}(t+1)\tilde{\mathbf{x}}_p(t+1)] \tag{34}$$

Thus we have the backward markovian model

$$\tilde{\mathbf{x}}_s(t) = A^{-1}(t)[I - B_2(t)B_2'(t)P_p^{-1}(t+1)]\tilde{\mathbf{x}}_s(t+1) - A^{-1}(t)\mu(t) \tag{35}$$

where, from (31)–(34), we have that the covariance of the white noise $\mu(t)$ is

$$E[\mu(t)\mu'(t)] = B_2(t)[I - B_2'(t)P_p^{-1}(t+1)B_2(t)]B_2'(t) \tag{36}$$

4. Backwards markovian smoothing error models from square-root information filtering

In this section, we make explicit the connection between the backward model (34)–(36) derived in § 3 and square-root information filtering (Bierman 1974, 1977), particularly the Dyer–McReynolds backward smoothing error covariance recursion.

We begin by recalling the filtering and prediction steps of square-root information filtering. Consider the problem of estimating $\mathbf{x}(t)$, satisfying (16)–(18) and with $\mathbf{y}(t)$ given by (2). Let

$$[P_p(t)]^{-1} = S_p'(t)S_p(t) \tag{37}$$

$$[P_f(t)]^{-1} = S_f'(t)S_f(t) \tag{38}$$

$$\mathbf{z}_p(t) = S_p(t)\hat{\mathbf{x}}_p(t) \tag{39}$$

$$\mathbf{z}_f(t) = S_f(t)\hat{\mathbf{x}}_f(t) \tag{40}$$

Given $S_p(t)$, $\mathbf{z}_p(t)$, the filtering steps then consist of constructing a Householder transformation $T_1(t)$ that produces the zero block in the following equation and consequently also yields $S_f(t)$ and $\mathbf{z}_f(t)$

$$T_1(t)\left[\begin{array}{c|c} S_p(t) & \mathbf{z}_p(t) \\ \hline D^{-1}(t)H(t) & D^{-1}(t)\mathbf{y}(t) \end{array}\right] = \left[\begin{array}{c|c} S_f(t) & \mathbf{z}_f(t) \\ \hline 0 & * \end{array}\right] \tag{41}$$

where '*' indicates a non-zero quantity of no interest in our discussion.

Given $S_f(t)$, $\mathbf{z}_f(t)$ the prediction step consists of constructing a Householder transformation $T_2(t)$ to produce the indicated zero blocks (and consequently the other quantities indicated) on the right-hand side of the following

$$T_2(t)\left[\begin{array}{c|c|c} I & 0 & 0 \\ \hline -S_f(t)A^{-1}(t)B_2(t) & S_f(t)A^{-1}(t) & \mathbf{z}_f(t)+S_f(t)A^{-1}(t)B_1(t)D^{-1}(t)\mathbf{y}(t) \end{array}\right] = \begin{bmatrix} S_{\mathbf{w}}(t) & S_{\mathbf{wx}}(t+1) & \mathbf{r}(t) \\ 0 & S_p(t+1) & \mathbf{z}_p(t+1) \end{bmatrix} \quad (42)$$

where the top blocks in these matrices have m rows, where $\dim(\mathbf{w}) = m$. Also, the quantities in the top block row on the right-hand side of (42) can be given precise statistical interpretations related to the best filtered estimate of the process noise $\mathbf{w}(t)$.

For our purposes, the key point is that Bierman (1974, 1977) uses these results to derive the following algorithm for computing the smoothed estimate

$$\hat{\mathbf{x}}_s(t) = A^{-1}(t)[\hat{\mathbf{x}}_s(t+1) - B_1(t)D^{-1}(t)\mathbf{y}(t) - B_2(t)\hat{\mathbf{w}}_s(t)] \quad (43)$$

$$\hat{\mathbf{w}}_s(t) = S_{\mathbf{w}}^{-1}(t)[\mathbf{r}(t) - S_{\mathbf{wx}}(t+1)\hat{\mathbf{x}}_s(t+1)] \quad (44)$$

with

$$\hat{\mathbf{x}}_s(T) = S_f^{-1}(T)\mathbf{z}_f(T) \quad (45)$$

and $\hat{\mathbf{w}}_s(t)$ is the smoothed estimate of $\mathbf{w}(t)$.

Bierman also shows that $\mathbf{x}(t)$ satisfies the following backward model

$$\mathbf{x}(t) = A^{-1}(t)[(I + L(t)S_{\mathbf{wx}}(t+1))\mathbf{x}(t+1) - B_1(t)D^{-1}(t)\mathbf{y}(t) - L(t)\mathbf{r}(t) + L(t)\omega(t)] \quad (46)$$

where

$$L(t) = B_2(t)S_{\mathbf{w}}^{-1}(t) \quad (47)$$

and

$$\omega(t) = -S_{\mathbf{w}}(t)\mathbf{w}(t) - S_{\mathbf{wx}}(t+1)\mathbf{x}(t+1) + \mathbf{r}(t) \quad (48)$$

A by-product of the square-root algorithm is that $\omega(t)$ is a zero-mean white noise process independent of $H(\mathbf{y})$ and with identity covariance.

Combining (43), (44) to obtain a single backward equation for $\hat{\mathbf{x}}_s(t)$ and subtracting from (46) yields the desired backward model for $\tilde{\mathbf{x}}_s(t)$

$$\tilde{\mathbf{x}}_s(t) = A^{-1}(t)[I + L(t)S_{\mathbf{wx}}(t+1)]\tilde{\mathbf{x}}_s(t+1) + A^{-1}(t)L(t)\omega(t) \quad (49)$$

from which we can also derive the Dyer–McReynolds smoothing covariance recursion

$$P_s(t) = A^{-1}(t)[I + L(t)S_{\mathbf{wx}}(t+1)]P_s(t+1)[I + L(t)S_{\mathbf{wx}}(t+1)]'[A'(t)]^{-1} + A^{-1}(t)L(t)L'(t)[A'(t)]^{-1} \quad (50)$$

Finally, using an argument analogous to that of Bierman (1977) (pp. 220–221) we can show the first equality of the following

$$A^{-1}(t)[I + L(t)S_{\mathbf{wx}}(t+1)] = P_p(t)A'(t)P_p^{-1}(t+1) = A^{-1}(t)[I - B_2(t)B_2'(t)P_p^{-1}(t+1)] \quad (51)$$

where the second equality follows from (20). Thus

$$L(t)S_{\mathbf{wx}}(t+1) = -B_2(t)B_2'(t)P_p^{-1}(t+1) \tag{52}$$

Also, from (44) and (48) we see that

$$L(t)\omega(t) = -B_2(t)[[\mathbf{w}(t) - \hat{\mathbf{w}}_s(t)] + S_w^{-1}(t)S_{\mathbf{wx}}(t+1)\tilde{\mathbf{x}}_s(t+1)] \tag{53}$$

Using (52) and the expression for $\hat{\mathbf{w}}_s(t)$ in (33), and comparing with (53), we see that

$$L(t)\omega(t) = -\mu(t) \tag{54}$$

That is, the model (49) is identical with that derived in § 3, although we have now made explicit contact with quantities arising in square-root algorithms.

5. Backward smoothing error models from backwards prediction error models

In order to obtain a backwards markovian representation for the smoothing error process, a *backwards* counterpart to the forward markovian model (13) for one-step prediction errors is used together with (25). The approach presented here exposes the intimate connection between the structure of backward smoothing and one-step prediction estimate error models. In addition, it is shown how this approach can be applied to the construction of backward smoothing error models in the case when $\Gamma(t)$ is singular.

Suppose that $\Gamma(t)$ is invertible. Then, using results from Verghese and Kailath (1979), we obtain the following backward markovian counterpart of (13)

$$\begin{aligned}\tilde{\mathbf{x}}_p(t) &= \Gamma^{-1}(t)\{I - [G(t)G'(t) + B_2(t)B_2'(t)]P_p^{-1}(t+1)\}\tilde{\mathbf{x}}_p(t+1)\\ &\quad - \Gamma^{-1}(t)[G(t) \mid B_2(t)]\begin{bmatrix}\eta(t)\\ \rho(t)\end{bmatrix}\end{aligned} \tag{55}$$

where

$$\begin{bmatrix}\eta(t)\\ \rho(t)\end{bmatrix} = \begin{bmatrix}\mathbf{v}(t)\\ \mathbf{w}(t)\end{bmatrix} - \begin{bmatrix}G'(t)\\ B_2'(t)\end{bmatrix}P_p^{-1}(t+1)\tilde{\mathbf{x}}_p(t+1) \tag{56}$$

is a white noise process independent of the future of $\tilde{\mathbf{x}}_p(\cdot)$ and the covariance

$$\operatorname{Cov}\begin{pmatrix}\eta(t)\\ \rho(t)\end{pmatrix} = I - \begin{bmatrix}G'(t)\\ B_2'(t)\end{bmatrix}P_p^{-1}(t+1)[G(t) \mid B_2(t)] \triangleq M(t) \tag{57}$$

Some algebra (using (13) and its associated Lyapunov equation) yields the following equivalent forms for (55) and (56)

$$\tilde{\mathbf{x}}_p(t) = N(t)\tilde{\mathbf{x}}_p(t+1) + [P_p(t)H'(t)(D'(t))^{-1} \mid -\Gamma^{-1}(t)B_2(t)]\begin{bmatrix}\eta(t)\\ \rho(t)\end{bmatrix} \tag{58}$$

where

$$N(t) = P_p(t)\Gamma'(t)P_p^{-1}(t+1)$$

and

$$\begin{bmatrix} \boldsymbol{\eta}(t) \\ \boldsymbol{\rho}(t) \end{bmatrix} = M(t)\left\{\begin{bmatrix} \mathbf{v}(t) \\ \mathbf{w}(t) \end{bmatrix} - \begin{bmatrix} G'(t) \\ B_2'(t) \end{bmatrix}(\Gamma'(t))^{-1}P_p^{-1}(t)\tilde{\mathbf{x}}_p(t)\right\} \tag{59}$$

$$\triangleq M(t)\begin{bmatrix} \boldsymbol{\alpha}(t) \\ \boldsymbol{\beta}(t) \end{bmatrix} \tag{60}$$

Using (10) and (15) we find that

$$\boldsymbol{\alpha}(t) = D^{-1}(t)\mathbf{v}(t) \tag{61}$$

Also, from (59), (60) and standard linear estimation results we can write

$$\begin{aligned} \boldsymbol{\beta}(t) &= -B_2'(t)(\Gamma'(t))^{-1}P_p^{-1}(t)E[\tilde{x}_p(t)|\mathbf{v}(t)] + \boldsymbol{\gamma}(t) \\ &= -B_2'(t)(\Gamma'(t))^{-1}H'(t)V^{-1}(t)\mathbf{v}(t) + \boldsymbol{\gamma}(t) \end{aligned} \tag{62}$$

where $\boldsymbol{\gamma}(t)$ is independent of $\mathbf{v}(t)$, with

$$\begin{aligned} E[\boldsymbol{\gamma}(t)\boldsymbol{\gamma}'(t)] &= I + B_2'(t)(\Gamma'(t))^{-1}P_p^{-1}(t)P_f(t)P_p^{-1}(t)(\Gamma(t))^{-1}B_2(t) \\ &= I + B_2'(t)(\Gamma'(t))^{-1}[P_p^{-1}(t) - H'(t)V^{-1}(t)H(t)]\Gamma^{-1}(t)B_2(t) \end{aligned} \tag{63}$$

where we have used the standard formula

$$P_f(t) = P_p(t) - P_p(t)H'(t)V^{-1}(t)H(t)P_p(t) \tag{64}$$

Performing some algebra (again using (15), and (13) and its Lyapunov equation) and summarizing, we now have the following backward markovian model for $\tilde{\mathbf{x}}_p(t)$

$$\begin{aligned} \tilde{\mathbf{x}}_p(t) = {} & N(t)\tilde{\mathbf{x}}_p(t+1) - N(t)B_2(t)\boldsymbol{\gamma}(t) \\ & + N(t)[\Gamma(t)P_p(t)H'(t)R^{-1}(t) + B_2(t)B_2'(t)(\Gamma'(t))^{-1}H'(t)V^{-1}(t)]\mathbf{v}(t) \end{aligned} \tag{65}$$

where $\boldsymbol{\gamma}(t)$ and $\mathbf{v}(t)$ are independent white noise processes with covariances given by (63) and $V(t)$, respectively.

Note that the independence of $\boldsymbol{\gamma}(t)$ and $\mathbf{v}(t)$ implies the orthogonality of $\boldsymbol{\gamma}(t)$ and *all* of $H(\mathbf{y})$. From this fact, it is quite straightforward to obtain a backward model for the smoothing error, simply by projecting (65) onto $H_t^+(\mathbf{v})$ and subtracting from (65). This yields the desired model

$$\tilde{\mathbf{x}}_s(t) = N(t)\tilde{\mathbf{x}}_s(t+1) - N(t)B_2'(t)\boldsymbol{\gamma}(t) \tag{66}$$

Some further algebra then yields

$$N(t) = A^{-1}(t)[I - B_2(t)B_2'(t)P_p^{-1}(t+1)] \tag{67}$$

so that (66) is identical with the model (35) derived in § 3.

Finally, let us comment on the extension of these results to the case when either $\Gamma(t)$ or $P_p(t)$ are not invertible. Using the results of Verghese and Kailath (1979), one can see that the invertibility of $\Gamma(t)$ is necessary for the construction of a backward model for $\tilde{\mathbf{x}}_p$ (and ultimately for $\tilde{\mathbf{x}}_s$) that has exactly the same sample paths as the forward model; however, the invertibility of P_p is not necessary. If P_p is singular, the corresponding results are obtained by noting that

$$E[\tilde{\mathbf{x}}_p(t)|H_{t+1}^+(\tilde{\mathbf{x}}_p)] = P_p(t)\Gamma'(t)P_p^{\#}(t+1)\tilde{\mathbf{x}}_p(t+1) \tag{68}$$

where '#' denotes the Moore–Penrose pseudo-inverse. If $\Gamma(t)$ is invertible, the analysis of this section can be carried out starting from (56)–(58) with the sole change being that now

$$N(t) = P_p(t)\Gamma'(t)P_p^{\#}(t+1) \tag{69}$$

If $\Gamma(t)$ is singular, we can still obtain a backward model that has the same second-order statistics as $\tilde{\mathbf{x}}_s$. Specifically, thanks to (68), we can write

$$\tilde{\mathbf{x}}_p(t) = N(t)\tilde{\mathbf{x}}_p(t+1) + \boldsymbol{\delta}(t) \tag{70}$$

where $N(t)$ is given by (69) and $\boldsymbol{\delta}(t)$ is a white noise process independent of the future of $\tilde{\mathbf{x}}_p(t)$. Furthermore, by construction $\boldsymbol{\delta}(t)$ is independent of $H_{t+1}^{+}(\mathbf{v})$ and, thanks to (70), it is also independent of $H_{t-1}^{-}(\mathbf{v})$. Thus we can write

$$\boldsymbol{\delta}(t) = E[\boldsymbol{\delta}(t)|\mathbf{v}(t)] + \boldsymbol{\xi}(t) \tag{71}$$

where $\boldsymbol{\xi}(t)$ is a white noise process independent of the future of $\tilde{\mathbf{x}}_p$ and of $H(\mathbf{y})$. Thus, in the same manner as used previously, we obtain the backward model

$$\tilde{\mathbf{x}}_s(t) = N(t)\tilde{\mathbf{x}}_s(t+1) + \boldsymbol{\xi}(t) \tag{72}$$

All that remains is to specify the covariance of $\boldsymbol{\xi}(t)$. This is done in Appendix B.

6. Construction of a forward smoothing error model

We use the method of complementary models (Weinert and Desai 1981, Bello 1981, Adams *et al.* 1984) to derive a forward smoothing error model that requires the invertibility neither of $\Gamma(t)$ nor of $P(0)$. To begin, recall the form of the state model given by (16)–(18), and consider the following processes

$$\mathbf{x}^*(t+1) = A(t)\mathbf{x}^*(t) + B_2(t)\mathbf{w}(t) \tag{73}$$

$$\mathbf{y}^*(t) = H(t)\mathbf{x}^*(t) + D(t)\mathbf{v}(t) \tag{74}$$

with

$$\mathbf{x}^*(0) = \mathbf{x}(0) \tag{75}$$

From (2), (16), (18), and (73)–(75), we can deduce that for $t > 0$

$$\left.\begin{aligned} \mathbf{x}^*(t) &= \mathbf{x}(t) - \sum_{\tau=0}^{t-1} \Phi_A(t,\tau+1)B_1(\tau)D^{-1}(\tau)\mathbf{y}(\tau) \\ \mathbf{y}^*(0) &= \mathbf{y}(0) \end{aligned}\right\} \tag{76}$$

and for $t > 0$

$$\mathbf{y}^*(t) = \mathbf{y}(t) - \sum_{\tau=0}^{t-1} H(t)\Phi_A(t,\tau+1)B_1(\tau)D^{-1}(\tau)\mathbf{y}(\tau) \tag{77}$$

where $\Phi_A(\cdot,\cdot)$ is the transition matrix associated with A. From (77) we see that $H(\mathbf{y}) = H(\mathbf{y}^*)$ and then, from (76), that

$$\tilde{\mathbf{x}}_s^*(t) = \tilde{\mathbf{x}}_s(t) \tag{78}$$

Note that the model (73), (74) has uncorrelated process and measurement noises.

In the following development we write

$$\mathbf{x}(0) = T_\varepsilon \boldsymbol{\varepsilon} \tag{79}$$

where ε is a random vector with identity covariance. The method of complementary models consists of identifying the space Y_c, where

$$H(\varepsilon) \oplus H(\mathbf{w}) \oplus H(\mathbf{v}) = H(\mathbf{y}^*) \oplus Y_c \tag{80}$$

Then

$$\hat{\mathbf{x}}_s^*(t) = E[x^*(t)|Y_c] \tag{81}$$

Results in Weinert and Desai (1981), Bello (1981), and Adams *et al.* (1984) yield

$$Y_c = H(\boldsymbol{\theta}) \oplus H(\mathbf{z}) \tag{82}$$

where

$$\boldsymbol{\theta} = \varepsilon - T_\varepsilon' A'(0)\boldsymbol{\lambda}(0) - T_\varepsilon' H'(0)[D'(0)]^{-1}\mathbf{v}(0) \tag{83}$$

$$\mathbf{z}(t) = -B_2'(t)\boldsymbol{\lambda}(t) + \mathbf{w}(t) \tag{84}$$

$\boldsymbol{\lambda}(t)$ satisfies the backward equation

$$\boldsymbol{\lambda}(t) = A'(t+1)\boldsymbol{\lambda}(t+1) + H'(t+1)[D'(t+1)]^{-1}\mathbf{v}(t+1) \tag{85}$$

and

$$\boldsymbol{\lambda}(T) = 0 \tag{86}$$

To obtain a forward model for $\hat{\mathbf{x}}_s$ we first obtain an alternative basis for Y_c, namely

$$\boldsymbol{\psi} = \boldsymbol{\theta} - E[\theta|H(\mathbf{z})] \tag{87}$$

$$\mathbf{v}_z(t) = \mathbf{z}(t) - E[z(t)|H_{t+1}^+(\mathbf{z})] \tag{88}$$

Note that $\boldsymbol{\psi}$ and $\mathbf{v}_z(\cdot)$ are uncorrelated and $\mathbf{v}_z(t)$ is a white noise process, representing the innovation process associated with the Kalman filter for the reverse-time model (84)–(86). These quantities and their covariances can be generated using standard Kalman filtering formulae (applied backwards in time). Specifically, let

$$\hat{\boldsymbol{\lambda}}_f(t) = E[\boldsymbol{\lambda}(t)|H_t^+(\mathbf{z})] \tag{89}$$

$$\hat{\boldsymbol{\lambda}}_p(t) = E[\boldsymbol{\lambda}(t)|H_{t+1}^+(\mathbf{z})] \tag{90}$$

with $\tilde{\boldsymbol{\lambda}}_f(t)$ and $\tilde{\boldsymbol{\lambda}}_p(t)$ denoting the corresponding errors and $\Theta_f(t)$ and $\Theta_p(t)$ the corresponding error covariances. Then

$$\tilde{\boldsymbol{\lambda}}_f(t) = [I - \Theta_f(t)B_2(t)B_2'(t)]\tilde{\boldsymbol{\lambda}}_p(t) + \Theta_f(t)B_2(t)\mathbf{w}(t) \tag{91}$$

$$\tilde{\boldsymbol{\lambda}}_p(t) = A'(t+1)\tilde{\boldsymbol{\lambda}}_f(t+1) + H'(t+1)[D'(t+1)]^{-1}\mathbf{v}(t+1) \tag{92}$$

with

$$\tilde{\boldsymbol{\lambda}}_f(T) = 0 \tag{93}$$

where

$$\Theta_f(t) = \Theta_p(t) - \Theta_p(t)B_2(t)[B_2'(t)\Theta_p(t)B_2(t) + I]^{-1}B_2'(t)\Theta_p(t) \tag{94}$$

$$\Theta_p(t) = A'(t+1)\Theta_f(t+1)A(t+1) + H'(t+1)R^{-1}(t+1)H(t+1) \tag{95}$$

with

$$\Theta_f(T) = 0 \tag{96}$$

Thus, from (83), (84), (87), and (88)

$$\boldsymbol{\psi} = \boldsymbol{\varepsilon} - T_\varepsilon' A'(0)\tilde{\boldsymbol{\lambda}}_f(0) - T_\varepsilon' H'(0)[D'(0)]^{-1}\mathbf{v}(0) \tag{97}$$

$$\mathbf{v}_z(t) = \mathbf{w}(t) - B_2'(t)\tilde{\boldsymbol{\lambda}}_p(t) \tag{98}$$

and, using (94), (95)

$$E[\boldsymbol{\psi}\boldsymbol{\psi}'] = I + T_\varepsilon' A'(0)\Theta_f(0)A(0)T_\varepsilon + T_\varepsilon' H'(0)R^{-1}(0)H(0)T_\varepsilon \tag{99}$$

$$E[\mathbf{v}_z(t)\mathbf{v}_z'(t)] = I + B_2'(t)\Theta_p(t)B_2(t) \tag{100}$$

We are now in a position to compute

$$\tilde{\mathbf{x}}_s(t) = E[\mathbf{x}^*(t)\boldsymbol{\psi}']E[\boldsymbol{\psi}\boldsymbol{\psi}']^{-1}\boldsymbol{\psi} + \sum_{\tau=0}^{T-1} E[\mathbf{x}^*(\tau)\mathbf{v}_z'(\tau)]E[\mathbf{v}_z(\tau)\mathbf{v}_z'(\tau)]^{-1}\mathbf{v}_z(\tau) \tag{101}$$

From (73), (91), (92), and some algebra we find that

$$\begin{aligned} E[\mathbf{x}^*(t)\boldsymbol{\psi}'] &= \Phi_A(t,0)T_\varepsilon + \sum_{\tau=0}^{t-1} \Phi_A(t,\tau+1)B_2(\tau)B_2'(\tau)\Theta_f(\tau)\Phi_E'(0,\tau)A(0)T_\varepsilon \\ &= \Phi_G(t,0)T_\varepsilon \end{aligned} \tag{102}$$

where

$$\Phi_A(t,\tau) = \begin{cases} A(t-1)\dots A(\tau) & t > \tau \\ I & t = \tau \end{cases} \tag{103}$$

$$\Phi_E(t,\tau) = \begin{cases} E(t)E(t+1)\dots E(\tau-1) & t < \tau \\ I & t = \tau \end{cases} \tag{104}$$

with

$$E(t) = [I - \Theta_f(t)B_2(t)B_2'(t)]A'(t+1) \tag{105}$$

and

$$\Phi_G(t,\tau) = \begin{cases} G(t-1)\dots G(\tau) & t > \tau \\ I & t = \tau \end{cases} \tag{106}$$

with

$$G(t) = [I - B_2(t)B_2'(t)\Theta_f(t)]A(t) \tag{107}$$

Also

$$E[\mathbf{x}^*(t)\mathbf{v}_z'(\tau)] = \begin{cases} 0 & \tau \geqslant t \\ \Phi_G(t,\tau+1)B_2(\tau) & \tau < t \end{cases} \tag{108}$$

Combining (99)–(108), we obtain the following forward model for $\tilde{\mathbf{x}}_s(t)$.

$$\mathbf{x}_s(t+1) = G(t)\tilde{\mathbf{x}}_s(t) + B_2(t)[I + B_2'(t)\Theta_p(t)B_2(t)]^{-1}\mathbf{v}_z(t) \tag{109}$$

with

$$\tilde{\mathbf{x}}_s(0) = T_\varepsilon[I + T_\varepsilon' A'(0)\Theta_f(0)A(0)T_\varepsilon + T_\varepsilon' H'(0)R^{-1}(0)H(0)T_\varepsilon]^{-1}\boldsymbol{\psi} \tag{110}$$

7. Model validation using two measurement sets

We describe and develop one important application of smoothing error models, namely a particular problem of model validation. Specifically we consider a problem in which two measurement systems are used to provide information about the same physical process. Associated with each system, we have a dynamic model describing the error and noise sources in the system. One of these models is assumed to have been validated, and so the objective of this section is to evaluate the validity of the model for the other measurement system.

A problem of the type just described arises in the mapping of the Earth's gravitational field (Nash and Jordan 1978). A wide variety of sensors, sensitive to different ranges of spatial wavelength, is available for this application. For example, satellite orbit data provide information on very long wavelength variations, satellite altimetry data are sensitive to intermediate wavelengths, and ship-borne gravimetry data yield information at still shorter wavelengths. An overall map of the gravity field is then obtained by combining such sources of information.

Maps of this type are used in a variety of contexts (e.g. aiding navigation systems on ships). In such a context, it is necessary to have statistical models for the map errors in order to make predictions concerning navigation system accuracy. Validating such models is of great importance in such an application, and one way in which an evaluation procedure for such a model can be designed is to obtain two sets of 'measurements' for the same gravity-field-related quantity. One set is derived from map information and a second set derived from a different data set, the error model for which has been adequately determined. Using this information one would like to construct likelihood functions related to the parameters and validity of the map error model. It is this problem that we now study.

Consider the following process that is under observation:

$$\mathbf{x}_0(t+1) = F_0(t)\mathbf{x}_0(t) + B_0(t)\mathbf{w}_0(t) \tag{111}$$

$$\mathbf{z}(t) = H_0(t)\mathbf{x}_0(t) \tag{112}$$

Suppose that we have two sets of measurements

$$\mathbf{y}_i(t) = \mathbf{z}(t) + \mathbf{n}_i(t), \quad i = 1, 2 \tag{113}$$

where the noise model for $\mathbf{n}_1(t)$ is known:

$$\mathbf{x}_1(t+1) = F_1(t)\mathbf{x}_1(t) + B_1(t)\mathbf{w}_1(t) \tag{114}$$

$$\mathbf{n}_1(t) = H_1(t)\mathbf{x}_1(t) + D_1(t)\mathbf{w}_1(t) \tag{115}$$

while the corresponding model for $\mathbf{n}_2(t)$ is parametrized by a vector Λ

$$\mathbf{x}_2(t+1) = F_2(t;\Lambda)\mathbf{x}_2(t) + B_2(t;\Lambda)\mathbf{w}_2(t) \tag{116}$$

$$\mathbf{n}_2(t) = H_2(t;\Lambda)\mathbf{x}_2(t) + D_2(t;\Lambda)\mathbf{w}_2(t) \tag{117}$$

where $\mathbf{x}_i(0)$, $i = 0, 1, 2$ $\mathbf{w}_i(t)$, $i = 0, 1, 2$ are mutually independent zero-mean and gaussian, with

$$E[\mathbf{x}_i(0)\mathbf{x}_i'(0)] = P_i(0), \quad i = 0, 1 \quad E[\mathbf{x}_2(0)\mathbf{x}_2'(0)] = P_2(0;\Lambda) \tag{118}$$

$$E[\mathbf{w}_i(t)\mathbf{w}_i'(\tau)] = Q_i(t)\delta_{t,\tau} \quad i = 0, 1 \quad E[\mathbf{w}_2(t)\mathbf{w}_2'(\tau)] = Q_2(t;\Lambda)\delta_{t,\tau} \tag{119}$$

The objective is to compute the likelihood function

$$p(\mathbf{Y}_1, \mathbf{Y}_2; \Lambda) \tag{120}$$

where

$$\mathbf{Y}_i = [\mathbf{y}_i'(0) \quad \dots \quad \mathbf{y}_i'(T)]' \tag{121}$$

This can then be used for model validation and parameter estimation. What we demonstrate now is the central role played by smoothing error models in the computation of (120). Specifically, note that

$$p(\mathbf{Y}_1, \mathbf{Y}_2; \Lambda) = p(\mathbf{Y}_1)p(\mathbf{Y}_2|\mathbf{Y}_1; \Lambda) = p(\mathbf{Y}_1)p(\tilde{\mathbf{Y}}_2; \Lambda) \tag{122}$$

where

$$\tilde{\mathbf{Y}}_2 = [\tilde{\mathbf{y}}_2'(0) \quad \dots \quad \tilde{\mathbf{y}}_2'(T)]' \tag{123}$$

$$\tilde{\mathbf{y}}_2(t) = \mathbf{y}_2(t) - E[\mathbf{y}_2(t)|\mathbf{Y}_1; \Lambda] \tag{124}$$

$$= \tilde{\mathbf{z}}_{1s}(t) + \mathbf{n}_2(t) \tag{125}$$

where $\tilde{\mathbf{z}}_{1s}(t)$ is the error in estimating $\mathbf{z}(t)$ given $\mathbf{Y}_1$. A markovian model for this error can be constructed using the methods in §§ 3–6 based on a state model consisting of (111)–(115) (using only $\mathbf{y}_1(t)$ in (113)). The computation of $p(\tilde{\mathbf{Y}}_2: \Lambda)$ is then a standard problem. Specifically, if we have constructed a forward model for the smoothing error $\tilde{\mathbf{z}}_{1s}(\cdot)$, we now have an overall forward model consisting of this forward error model, (125), and (116), (117). Standard Kalman filter-based methods then provide the required recursions for computing $p(\tilde{\mathbf{Y}}_2; \Lambda)$. If we have constructed a backward model for $\tilde{\mathbf{z}}_{1s}(\cdot)$, we first obtain a backward markovian model corresponding to (116), (117) (Verghese and Kailath 1979) and then apply Kalman filter recursions in reverse time.

8. Updating and combining of smoothed estimates

We describe the discrete-time counterparts of the continuous-time results on updating and combining estimates developed in Willsky *et al.* (1982), and Bello *et al.* (1986). We also present some new insights to clarify the structure of the solution by making use of the idea of oblique projections employed in Desai and Kiaei (1985) to solve decentralized filtering problems. As discussed in Willsky *et al.* (1982) and Bello *et al.* (1986), the problem considered here is motivated by spatial data assimilation problems such as combining gravitational maps based on different data sets and updating such maps as new measurements become available. In § 9 we give an example indicating the applicability of these discrete-time results to a particular non-trivial measurement geometry, namely data sets along non-parallel tracks across a random field.

Let $\mathbf{x}(t)$, given by (1) by the process to be estimated, and suppose that two sets of measurements are available:

$$\mathbf{y}_i(t) = H_i(t)\mathbf{x}(t) + \mathbf{v}_i(t), \quad i = 1, 2 \tag{126}$$

where $\mathbf{w}(t)$, $\mathbf{v}_1(t)$, $\mathbf{v}_2(t)$ are mutually independent, and

$$E[\mathbf{v}_i(t)\mathbf{v}_i(\tau)] = R_i(t)\delta_{t,\tau} \tag{127}$$

The smoothed estimates based on $\mathbf{y}_i$ alone or on $(\mathbf{y}_1, \mathbf{y}_2)$ together are, respectively, given by

$$\hat{\mathbf{x}}_{is}(t) = E[\mathbf{x}(t)|H(\mathbf{y}_i)], \quad i = 1, 2 \tag{128}$$

$$\hat{\mathbf{x}}_s(t) = E[\mathbf{x}(t)|H(\mathbf{y}_1, \mathbf{y}_2)] \tag{129}$$

The updating problem is concerned with the computation of $\hat{\mathbf{x}}_s$ in terms of $\hat{\mathbf{x}}_{1s}$ and $\mathbf{y}_2$, i.e. with updating $\hat{\mathbf{x}}_{1s}$ given the new data $\mathbf{y}_2$. The following elementary Hilbert space argument shows that this is possible and leads to an algorithm directly involving the use of a smoothing error model.

To begin, we define the error process

$$\tilde{\mathbf{y}}_2(t) = \mathbf{y}_2(t) - E[\mathbf{y}_2(t)|H(\mathbf{y}_1)] \tag{130}$$

Then we have the following *orthogonal* sum decomposition

$$H(\mathbf{y}_1, \mathbf{y}_2) = H(\mathbf{y}_1) \oplus H(\tilde{\mathbf{y}}_2) \tag{131}$$

If we then project $\mathbf{x}(t)$ onto both sides of (131) we obtain

$$\hat{\mathbf{x}}_s(t) = \hat{\mathbf{x}}_{1s}(t) + E[\mathbf{x}(t)|H(\tilde{\mathbf{y}}_2)] \tag{132}$$

However, $\mathbf{x}(t) = \tilde{\mathbf{x}}_{1s}(t) + \hat{\mathbf{x}}_{1s}(t)$, and it is not difficult to check that

$$\tilde{\mathbf{y}}_2(t) = H_2(t)\tilde{\mathbf{x}}_{1s}(t) + \mathbf{v}_2(t) \tag{133}$$

so that $\hat{\mathbf{x}}_{1s}(t)$ is orthogonal $H(\tilde{\mathbf{y}}_2)$. Thus

$$\hat{\mathbf{x}}_s(t) = \hat{\mathbf{x}}_{1s}(t) + E[\tilde{\mathbf{x}}_{1s}(t)|H(\tilde{\mathbf{y}}_2)] \tag{134}$$

which represents the solution to the updating problem. Specifically, using results in §§ 3–6, we can construct Markov models for $\tilde{\mathbf{x}}_{1s}(t)$ which, when combined with (133), allow us to use standard smoothing algorithms to compute $E[\tilde{\mathbf{x}}_{1s}(t)|H(\mathbf{y}_2)]$.

The combining problem is concerned with the computation of $\hat{\mathbf{x}}_s$ in terms of $\hat{\mathbf{x}}_{1s}$ and $\hat{\mathbf{x}}_{2s}$. To solve this, let us first note the counterparts of (130), (133), and (134) with 1 and 2 reversed.

$$\tilde{\mathbf{y}}_1(t) = \mathbf{y}_1(t) - E[\mathbf{y}_1(t)|H(\mathbf{y}_2)] = H_1(t)\tilde{\mathbf{x}}_{2s}(t) + \mathbf{v}_1(t) \tag{135}$$

$$\hat{\mathbf{x}}_s(t) = \hat{\mathbf{x}}_{2s}(t) + E[\tilde{\mathbf{x}}_{2s}(t)|H(\tilde{\mathbf{y}}_1)] \tag{136}$$

Adding (134) and (136) and subtracting $\hat{\mathbf{x}}_s(t)$ yields

$$\hat{\mathbf{x}}_s(t) = \hat{\mathbf{x}}_{1s}(t) + \hat{\mathbf{x}}_{2s}(t) + \{E[\tilde{\mathbf{x}}_{1s}(t)|H(\tilde{\mathbf{y}}_2)] + E[\tilde{\mathbf{x}}_{2s}(t)|H(\tilde{\mathbf{y}}_1)] - \hat{\mathbf{x}}_s(t)\} \tag{137}$$

To show that this is an abstract solution to the combining problem we must verify that the bracketed term in (137) is a function of $\hat{\mathbf{x}}_{1s}$ and $\hat{\mathbf{x}}_{2s}$ alone. In the continuous-time context, this is demonstrated by purely algebraic techniques which also yield two-filter algorithms for computing the bracketed term from $\hat{\mathbf{x}}_{1s}$ and $\hat{\mathbf{x}}_{2s}$ (Wilsky *et al.* 1982, Bello *et al.* 1986). Algebraic techniques also form the basis for the discrete-time solutions presented in Bierman and Belzer (1985) and Watanabe (1986). We follow a different path, based on the theory of oblique projections, that also provides us with deeper insight into the geometric nature of the solution and of the role played by smoothing error models in computing these oblique projections.

Let us anticipate the final answers. Specifically, let L_1 and L_2 denote the linear operators such that

$$E[\tilde{\mathbf{x}}_{2s}(\cdot)|H(\tilde{\mathbf{y}}_1)] = L_1(\tilde{\mathbf{y}}_1) \tag{138 a}$$

$$E[\tilde{\mathbf{x}}_{1s}(\cdot)|H(\tilde{\mathbf{y}}_2)] = L_2(\tilde{\mathbf{y}}_2) \tag{138 b}$$

What we will show is that the smoothed estimate $\hat{\mathbf{x}}_s$ can be expressed as follows

$$\hat{\mathbf{x}}_s(\cdot) = L_1(\mathbf{y}_1) + L_2(\mathbf{y}_2) \tag{139}$$

Assuming that this is true and using (133), (135), and (137)–(139) we find that

$$\hat{\mathbf{x}}_s(t) = \hat{\mathbf{x}}_{1s}(t) + \hat{\mathbf{x}}_{2s}(t) - L_1\,(H_1\hat{x}_{2s})(t) - L_2\,(H_2\hat{\mathbf{x}}_{1s})(t) \tag{140}$$

What (140) says is the following. Recall that the computations in (138 a) and (138 b) are standard smoothing problems, because of the existence of Markov models for $\tilde{\mathbf{x}}_{1s}$ and $\tilde{\mathbf{x}}_{2s}$. Thus L_1 and L_2 can be implemented in a variety of ways such as the two-filter form for the optimal smoother. Then, according to (140), $\hat{\mathbf{x}}_s$ is computed from $\hat{\mathbf{x}}_{1s}$ and $\hat{\mathbf{x}}_{2s}$ by adding the two and subtracting the outputs of the algorithms implementing L_1 and L_2 when these have $H_1\hat{\mathbf{x}}_{2s}$ and $H_2\hat{\mathbf{x}}_{1s}$, respectively, as inputs.

What remains now is to demonstrate the validity of (139). We begin by recalling the oblique projection result stated in Bello (1981). Let G be a closed subspace of a Hilbert space H with the following direct, but not necessarily orthogonal, decomposition

$$G = M_1 + M_2 \tag{141}$$

Then for any $\alpha \in H$, the projection of α onto G, denoted by $P[\alpha|H]$ can be uniquely expressed as

$$P[\alpha|H] = \alpha_1 + \alpha_2, \quad \alpha_i \in M_i \tag{142}$$

The α_i are the oblique projections of α, *uniquely determined* by the orthogonality conditions

$$\langle P[\alpha|\tilde{M}_i] - \alpha_i, \Psi\rangle = 0 \quad \text{for all } \Psi \in \tilde{M}_i, \quad i = 1, 2 \tag{143}$$

where $\langle \cdot, \cdot \rangle$ is the inner product on H, and $\tilde{M}_i$ is defined as

$$\tilde{M}_i = \overline{\text{span}\,\{\beta - P[\beta|M_j]|\beta \in M_i\}}, \quad \text{where } j \neq i \tag{144}$$

In our context, $G = H(\mathbf{y}_1, \mathbf{y}_2)$, $M_i = H(\mathbf{y}_i)$, $\tilde{M}_i = H(\tilde{\mathbf{y}}_i)$, and $\alpha = \mathbf{x}(t)$. Then

$$\hat{\mathbf{x}}_s(t) = K_1(\mathbf{y}_1) + K_2(\mathbf{y}_2) \tag{145}$$

where K_1 and K_2 are the oblique projection operators. Also, in this context

$$P[\alpha|\tilde{M}_i] = E[\mathbf{x}(t)|H(\tilde{\mathbf{y}}_i)] = E[\tilde{\mathbf{x}}_{sj}(t)|H(\tilde{\mathbf{y}}_i)] = L_i(\tilde{\mathbf{y}}_i) \tag{146}$$

where $j \neq i$ and the second equality follows from the orthogonality of $\hat{\mathbf{x}}_{sj}$ and $\mathbf{y}_i$.

Thus in our case the orthogonality condition (143) becomes

$$[L_i(\tilde{\mathbf{y}}_i) - K_i(\mathbf{y}_i)] \perp H(\tilde{\mathbf{y}}_i) \tag{147}$$

and the problem is to find the (*unique*) operator K_i that satisfies (147). We now show that the choice is

$$K_i = L_i \tag{148}$$

Indeed if we make this choice and use (133), (135) we find that (147) reduces to

$$-L_i(H_i\hat{x}_{js}) \perp H(\tilde{\mathbf{y}}_i) \tag{149}$$

which *is* valid because of the orthogonality of $\hat{\mathbf{x}}_{js}$ and $\tilde{\mathbf{y}}_i$.

Appendix C contains an alternative independent proof of (139).

9. Map-updating example

We consider a discrete space map-updating problem corresponding to a measurement geometry with non-parallel measurement tracks through a scalar, separable, stationary, random field. As will become clear, our approach can be applied to a large class of discrete space mapping problems.

Let $\mathbf{f}(i,j)$, $i = 1, \ldots, T$, $j = 1, \ldots, M$ denote a two-dimensional zero-mean random field with correlation function

$$E[\mathbf{f}(m+i, n+j)\mathbf{f}(m,n)] \triangleq R(i,j) = p\alpha^{|i|}\beta^{|j|} \tag{150}$$

For simplicity let us consider a problem in which each data set consists of a single track of data across the field. Specifically the two data sets are defined by

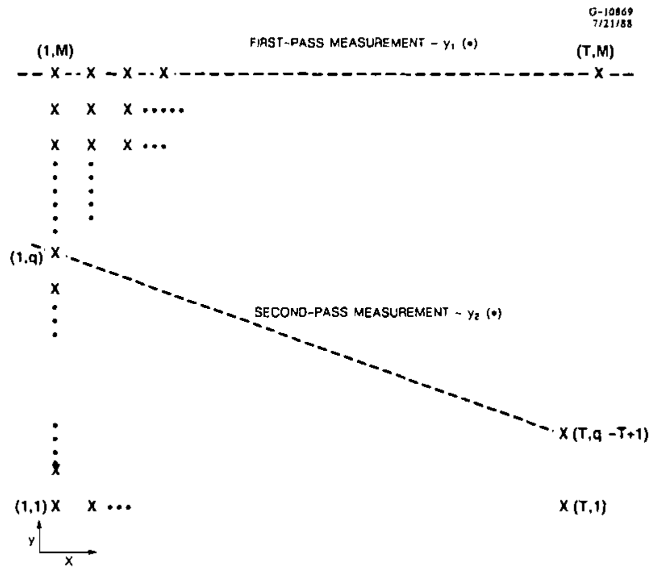
$$\mathbf{y}_1(t) = \mathbf{f}(t, M) + \mathbf{v}_1(t), \quad t = 1, \ldots, T \tag{151}$$

$$\mathbf{y}_2(t) = \mathbf{f}(t, [q - t + 1]) + \mathbf{v}_2(t), \quad t = 1, \ldots, T \tag{152}$$

where $\mathbf{v}_1(\cdot)$, $\mathbf{v}_2(\cdot)$ are independent, zero-mean white noise processes with

$$E[\mathbf{v}_i^2(t)] = r_i \quad i = 1, 2 \tag{153}$$

The measurement geometry is depicted in the Figure.



Measurement geometry for discrete-space map-updating example.

Again for simplicity we assume that

$$T \leqslant q \leqslant M \tag{154}$$

so that $\mathbf{y}_2(t)$ is defined for $t = 1, \ldots, T$, i.e. so that the second track of data is a complete one.

We now demonstrate that this problem can be cast in the framework described in the preceding sections. The state model we use, which is essentially the same as that used in Powell and Silverman (1974), describes the evolution of the set of values $\{\mathbf{f}(t,j), j = 1, \ldots, M\}$ as t increases from 1 to T. Specifically, let

$$\mathbf{z}(t) = [\mathbf{f}(t, M) \quad \ldots \quad \mathbf{f}(t, 1)]' \tag{155}$$

Then a state model for our field is

$$\mathbf{x}(t+1) = F\mathbf{x}(t) + B\mathbf{w}(t) \tag{156 a}$$

with

$$\mathbf{z}(t) = C\mathbf{x}(t) \tag{156 b}$$

where $\mathbf{x}(t)$ and $\mathbf{w}(t)$ are M–dimensional, $\mathbf{w}(t)$ is white noise with identity covariance, and

$$F = \alpha I \tag{157 a}$$

$$B = [p(1-\alpha^2)]^{1/2} I \tag{157 b}$$

$$E(\mathbf{x}(0)\mathbf{x}'(0)] = P = pI \tag{157 c}$$

and

$$C_{ij} = \begin{cases} \beta^{i-1} & \text{if } j = 1 \\ \beta^{i-j}(1-\beta^2)^{1/2} & \text{if } 1 < j \leqslant i \\ 0 & \text{otherwise} \end{cases} \tag{157 d}$$

From (151), (152), (155), and (157) we also have

$$\mathbf{y}_1(t) = H_1\mathbf{x}(t) + \mathbf{v}_1(t) \tag{158}$$

$$\mathbf{y}_2(t) = H_2(t)\mathbf{x}(t) + \mathbf{v}_2(t) \tag{159}$$

where

$$H_1 = [1 \quad 0 \quad \dots \quad 0]C = [1 \quad 0 \quad \dots \quad 0] \tag{160}$$

$$\begin{aligned} H_2(t) &= [0 \quad \dots \quad 0 \quad \overset{(m-q+t)\text{th position}}{1} \quad 0 \quad \dots \quad 0]C \\ &= [\beta^{M+t-q-1} \quad \beta^{M+t-q-2}(1-\beta^2)^{1/2} \quad \dots \quad (1-\beta^2)^{1/2} \quad 0 \quad \dots \quad 0] \end{aligned} \tag{161}$$

Note that our random field has a separable covariance. Since the first measurement set is along one of the directions of separability, we obtain a time-invariant model (158). The second measurement set, however, is not along such a direction, and the resulting model is time-varying.

Our problem is now set up exactly in the form used in § 8. For example, the map-updating solution is given by

$$\hat{\mathbf{z}}_s(t) = C\hat{\mathbf{x}}_s(t) \tag{162}$$

where

$$\hat{\mathbf{x}}_s(t) = \hat{\mathbf{x}}_{1s}(t) + E[\tilde{\mathbf{x}}_{1s}(t)|H(\tilde{\mathbf{y}}_2)] \tag{163}$$

Note that the special structure of the state model (156), (157) leads to considerable simplification. Specifically, the components of

$$\mathbf{x}(t) = [\mathbf{x}^1(t) \quad \dots \quad \mathbf{x}^M(t)]' \tag{164}$$

are independent. Since $\mathbf{y}_1(t)$ measures only the first component of $\mathbf{x}(t)$,

$$\hat{\mathbf{x}}_{1s}(t) = [\mathbf{x}^1_{1s}(t) \quad 0 \quad \dots \quad 0]' \tag{165}$$

and thus the smoothing error model for $\tilde{\mathbf{x}}_{1s}(t)$ consists of the *original* models for $\mathbf{x}^2(t), \dots, \mathbf{x}^M(t)$ and the scalar smoothing error model for $\mathbf{x}^1(t)$. Some algebra (Bello

1981) then yields the following model

$$\tilde{\mathbf{x}}_{1s}(t+1) = F(t)\tilde{\mathbf{x}}_{1s}(t) + B(t)\tilde{\mathbf{w}}(t) \tag{166}$$

$$\begin{aligned}\tilde{\mathbf{y}}_2(t) &= \mathbf{y}_2(t) - H_2(t)\hat{\mathbf{x}}_{1s}(t) \\ &= H_2(t)\tilde{\mathbf{x}}_{1s}(t) + \mathbf{v}_2(t)\end{aligned} \tag{167}$$

where

$$F(t) = \operatorname{diag}(a(t), \alpha I) \tag{168 a}$$

$$B(t) = \operatorname{diag}(b(t), [p(1-\alpha^2)]^{1/2} I) \tag{168 b}$$

where

$$a(t) = \alpha[1 - p(1-\alpha^2)P_{1b}(t)] \tag{169 a}$$

$$b(t) = \{p(1-\alpha^2)[1 - p(1-\alpha^2)P_{1b}(t)]\}^{1/2} \tag{169 b}$$

and $P_{1b}(t)$ satisfies the backward recursion

$$P_{1b}(t) = \frac{\alpha^2 P_{1b}(t+1) + (1/r_1)}{p(1-\alpha^2)[\alpha^2 P_{1b}(t+1) + (1/r_1)] + 1} \tag{170 a}$$

with

$$P_{1b}(T) = 0 \tag{170 b}$$

Finally, the initial covariance for (166) is

$$E[\tilde{\mathbf{x}}_{1s}(1)\tilde{\mathbf{x}}'_{1s}(1)] = \operatorname{diag}(P_{1s}(1), p, \ldots, p) \tag{171 a}$$

where

$$P_{1s}(1) = \left[\frac{1}{p} + \left(\alpha^2 P_{1b}(1) + \frac{1}{r_1}\right)\right]^{-1} \tag{171 b}$$

Note also that the triangular structure of C and hence of $H_2(t)$ also implies considerable simplification for the solution to the smoothing problem for (166), (167) (Bello 1981). All of this indicates that particular efficiencies may be obtained by careful choice of state representations to take advantage of the geometry of the measurements and the correlation structure of the underlying random field. The method developed here can be immediately extended to correlation models of the form

$$R(i,j) = \sum_l \phi_l(i)\psi_l(j) \tag{172}$$

where $\phi_l(\cdot)$ and $\psi_l(\cdot)$ are one-dimensional correlation functions realizable as the correlation of the output of finite-dimensional linear systems. Representations of this type can be used to approximate correlation structures of a large class of random fields (Woods and Radewan 1977).

10. Conclusions

We have considered the construction and application of markovian models for the error in fixed-interval smoothed estimates. In particular we have described three approaches to the construction of backward models, each of which provides additional insight and connection to other results in linear estimation. In addition, the third of these methods makes clear the necessity of the invertibility of the one-step prediction transition matrix for the construction of a backward model with identical sample paths. On the other hand, the forward model developed using the method of complementary models has no such requirement.

We have described two applications in which smoothing error models play a central role, namely a particular model validation problem and the problems of map updating and combining. In the case of the combining of smoothed estimates we have exposed the connection of the problem to the theory of oblique projections. We have also presented the problem of updating the map of a 2-dimensional random field, given data along non-parallel tracks. In this case, the non-parallelism manifests itself in the non-stationarity of the resulting 1-dimensional models. By judicious choice of realization, the computations required in constructing the smoothing error model and in the subsequent updating operation can be greatly simplified.

ACKNOWLEDGMENTS

M. G. Bello was supported in part by the National Science Foundation under Grant ECS-8312921. A. S. Willsky was supported in part by the National Science Foundation under Grants ECS-8312921 and ECS-8700903, in part by the Army Research Office under Grant DAAL03-86-K-0171 and part by l'Institut National de Recherche en Informatique et en Automatique during his sabbatical at l'Institut de Recherche en Informatique et Systèmes Aléatoires, Rennes, France. B. C. Levy was supported in part by the National Science Foundation under Grants ECS-8312921 and ECS-8700903 and in part by the Army Research Office under Grant DAAL03-86-K-0171.

Appendix A

In this appendix we verify (32) and (33). To demonstrate the validity of (32) we must show that

$$E\{[\mathbf{w}(t) - B_2'(t)P_p^{-1}(t+1)\tilde{\mathbf{x}}_s(t+1)]\tilde{\mathbf{x}}_s'(t+l)\} = 0, \quad l = 1, 2, \ldots \tag{A 1}$$

From (10), (13), and (25) we can write the following explicit formula for $\tilde{\mathbf{x}}_s(t)$.

$$\tilde{\mathbf{x}}_s(t) = \tilde{\mathbf{x}}_p(t) - P_p(t)\sum_{\tau=t}^{T} \Phi_\Gamma'(\tau, t)H'(\tau)V^{-1}(\tau)\mathbf{v}(\tau) \tag{A 2}$$

where $\Phi_\Gamma(\cdot,\cdot)$ is the transition matrix associated with $\Gamma(t)$. Using (A 2), (10), and (13) we can then make the following computations.

$$E[\mathbf{w}(t)\tilde{\mathbf{x}}_s'(t+l)] = B_2'(t)\Phi_\Gamma'(t+l, t+1) - \sum_{\tau=(t+l)}^{T} B_2'(t)\Phi_\Gamma'(\tau, t+1)H'(\tau)V^{-1}(\tau) \times H(\tau)\Phi_\Gamma(\tau, t+l)P_p(t+l) \tag{A 3}$$

$$E[\tilde{\mathbf{x}}_s(t+1)\tilde{\mathbf{x}}_s'(t+l)] = P_p(t+1)\Phi_\Gamma'(t+l, t+1) - P_p(t+1)\sum_{\tau=(t+l)}^{T}\Phi_\Gamma'(\tau, t+1) \times H'(\tau)V^{-1}(\tau)H(\tau)\Phi_\Gamma(\tau, t+l)\, P_p(t+l) \tag{A 4}$$

from which (A 1) follows.

To verify (33), we note that from (23) we have that

$$E[\mathbf{w}(t)|H(\mathbf{y})] = E[\mathbf{w}(t)|H_{t+1}^{+}(\mathbf{v})] \tag{A 5}$$

This, together with (10) and (13), allows us to obtain the formula

$$E[\mathbf{w}(t)|H(\mathbf{y})] = B_2'(t)\left[\sum_{\tau=(t+1)}^{T} \Phi_\Gamma'(\tau, t+1)H'(\tau)V^{-1}(\tau)\mathbf{v}(\tau)\right] \tag{A 6}$$

and the use of (A 2) then yields (33).

Appendix B

To derive the covariance of $\xi(t)$ in (72) we begin by noting the following equalities:

$$\tilde{\mathbf{x}}_s(t) = \tilde{\mathbf{x}}_f(t) - E[\tilde{\mathbf{x}}_f(t)|H_{t+1}^{+}(\mathbf{v})] \tag{B 1}$$

$$\tilde{\mathbf{x}}_s(t) = \tilde{\mathbf{x}}_p(t) - E[\tilde{\mathbf{x}}_p(t)|H_t^{+}(\mathbf{v})] \tag{B 2}$$

$$E[\tilde{\mathbf{x}}_f(t)|H_{t+1}^{+}(\mathbf{v})] = E[\tilde{\mathbf{x}}_p(t)|H_{t+1}^{+}(\mathbf{v})] \tag{B 3}$$

$$E[\tilde{\mathbf{x}}_p(t)|H_{t+1}^{+}(\mathbf{v})] = N(t)E[\tilde{\mathbf{x}}_p(t+1)|H_{t+1}^{+}(\mathbf{v})] \tag{B 4}$$

where (B 4) follows from (70). Substituting these into (72), (with (B 2) evaluated at $(t+1)$, we obtain

$$\xi(t) = \tilde{\mathbf{x}}_f(t) - N(t)\tilde{\mathbf{x}}_p(t+1) \tag{B 5}$$

Using (17), (29) we then obtain

$$E[\xi(t)\xi'(t)] = P_f(t) - L(t)S(t) - S(t)L(t) + L(t)S(t)L(t) \tag{B 6}$$

where

$$L(t) = P_p(t)P_p^{\#}(t) \tag{B 7}$$

$$S(t) = P_f(t)A'(t)P_p^{\#}(t+1)A(t)P_f(t) \tag{B 8}$$

Appendix C

We demonstrate the validity of (139) without reference to oblique projections. To begin, we define the operators

$$\hat{\mathbf{x}}_{is} = N_i(\mathbf{y}_i) \tag{C 1}$$

$$E[\mathbf{y}_i|H(\mathbf{y}_j)] = S_{ij}(\mathbf{y}_j) \tag{C 2}$$

Then using (138 a), (C 1) and (C 2), we can rewrite (136)

$$\begin{aligned}\hat{\mathbf{x}}_s &= N_2(\mathbf{y}_2) + L_1(\tilde{\mathbf{y}}_1)\\ &= N_2(\mathbf{y}_2) + L_1(\mathbf{y}_1 - S_{12}(\mathbf{y}_2))\\ &= L_1(\mathbf{y}_1) + [N_2 - L_1S_{12}](\mathbf{y}_2)\end{aligned} \tag{C 3}$$

Now the defining requirement for L_2 implies that

$$[\mathbf{x} - L_2(\tilde{\mathbf{y}}_2)] \perp H(\tilde{\mathbf{y}}_2) \tag{C 4}$$

Writing $\mathbf{x} = \hat{\mathbf{x}}_s + \tilde{\mathbf{x}}_s$ and using (C 3) and the expression

$$\tilde{\mathbf{y}}_2 = \mathbf{y}_2 - S_{21}(\mathbf{y}_1) \tag{C 5}$$

we find that (C 4) becomes

$$\{\tilde{\mathbf{x}}_s + [L_1 + L_2 S_{21}](\mathbf{y}_1) + [N_2 - L_1 S_{12} - L_2](\mathbf{y}_2)\} \perp H(\tilde{\mathbf{y}}_2) \tag{C 6}$$

Since $\tilde{\mathbf{x}}_s$ and $\mathbf{y}_1$ are both orthogonal to $H(\tilde{\mathbf{y}}_2)$ but $\mathbf{y}_2$ is not, we must have

$$L_2 = N_2 - L_1 S_{12} \tag{C 7}$$

which, together with (C 3), yields (139).

References

ADAMS, M. B., WILLSKY, A. S., and LEVY, B. C., 1984, Linear estimation of boundary value stochastic processes. Parts I and II. *I.E.E.E. Transactions on Automatic Control*, **29**, 803–821.

BADAWI, F., 1980, Structures and algorithms in stochastic realization theory and the smoothing problem. Ph.D. thesis, Department of Mathematics, University of Kentucky, Lexington, Kentucky.

BELLO, M. G., 1981, Centralized and decentralized map-updating and terrain masking analysis. Ph.D. dissertation, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, Massachusetts.

BELLO, M. G., WILLSKY, A. S., LEVY, B. C., and CASTAÑON, D. A., 1986, Smoothing error dynamics and their use in the solution of smoothing and mapping problems. *I.E.E.E. Transactions on Information Theory*, **32**, 483–495.

BIERMAN, G. J., 1974, Sequential square root filtering and smoothing of discrete linear systems. *Automatica*, **10**, 147–158; 1977, *Factorization Methods for Discrete Sequential Estimation* (New York: Academic Press).

BIERMAN, G. J., and BELZER, M. R., 1985, A decentralized square root information filter smoother. *Proceedings of the 24th I.E.E.E. Conference on Decision and Control*, Fort Lauderdale, Florida, pp. 1902–1905.

DESAI, U. B., and KIAEI, S., 1985, A stochastic realization approach to reduced order hierarchical estimation. *Proceedings of the 24th I.E.E.E. Conference on Decision and Control*, Fort Lauderdale, Florida, pp. 416–421.

JAZWINSKI, A. H., 1970, *Stochastic Processes and Filtering Theory* (New York: Academic Press).

NASH, R., and JORDAN, S., 1978, Statistical geodesy, an engineering perspective. *Proceedings of the Institute of Electrical and Electronic Engineers*, **66**, 532–550.

POWELL, S., and SILVERMAN, L., 1974, Modeling of two-dimensional covariance functions with application to image restoration. *I.E.E.E. Transactions on Automatic Control*, **19**, 8–12.

SEGAL, A., 1976, Stochastic processes in estimation theory. *I.E.E.E. Transactions on Information Theory*, **22**, 275–286.

VERGHESE, G. C., and KAILATH, T., 1979, A further note on backward markovian models. *I.E.E.E. Transactions on Information Theory*, **25**, 121–124.

WATANABE, K., 1986, Decentralized two-filter smoothing algorithms in discrete-time systems. *International Journal of Control*, **44**, 49–63.

WEINERT, H. L., and DESAI, U. B., 1981, On complementary models and fixed-interval smoothing. *I.E.E.E. Transactions on Automatic Control*, **26**, 863–867.

WILLSKY, A. S., BELLO, M. G., CASTAÑON, D. A., LEVY, B. C., and VERGHESE, G. C., 1982, Combining and updating of local estimates and regional maps along sets of one-dimensional tracks. *I.E.E.E. Transactions on Automatic Control*, **27**, 799–813.

WOODS, J., and RADEWAN, C., 1977, Kalman filtering in two dimensions. *I.E.E.E. Transactions on Information Theory*, **23**, 473–482.