Stochastics, 1981, Vol. 5, pp. 1–41
0090-2234/81/0502-0001 $06.50/0

On the Fixed-Interval Smoothing Problem†

JOSEPH E. WALL, Jr.

Honeywell Systems and Research Center, 2600 Ridgway Parkway, Minneapolis, Minnesota 55413, U.S.A.

ALAN S. WILLSKY

M.I.T. Laboratory for Information and Decision Sciences, Cambridge, Massachusetts 02139, U.S.A.

and

NILS R. SANDELL, Jr.

Alpha Tech, Arlington, Massachusetts, U.S.A.

(*Accepted for publication April 3, 1980*)

After a review of the development of the Mayne–Fraser two-filter smoother, a first principle argument is used to rederive this smoother. Reversed-time Markov models play a key role in forming a state estimate from future observations. The built-in asymmetry of the Mayne–Fraser smoother is pointed out, and it is shown how this asymmetry may be removed. Additionally, a covariance analysis of the two-filter smoother is provided, and reduced-order smoothers are analyzed.

I. INTRODUCTION

The fixed-interval (FI) smoothing problem is of particular interest in post-experimental data analysis and has been the subject of much attention [1], [2]. Smoothing refers to estimating a state vector at a time point intermediate to a span of measurements. Consequently, there is an essential element of noncausality in smoothing since some of the

†This work was partially supported by the Department of Energy under Contract ERDA-E-(49-18)-2087.

measurements occur in the future. Fixed-interval smoothing involves measurements over a given, fixed time interval. Estimates of the state are desired throughout this time interval.

Problem formulation

Consider the continuous-time linear dynamic system

$$\frac{d}{dt}x(t)=A(t)x(t)+w(t) \tag{1.1}$$

with observations

$$y(t)=C(t)x(t)+v(t) \tag{1.2}$$

where $x(0)=x_0$ and

$$Ex_0=Ew(t)=Ev(t)=0, \quad Ex_0x_0'=\Sigma(0),$$
$$Ew(t)w'(\tau)=Q(t)\delta(t-\tau), \quad Ev(t)v'(\tau)=R(t)\delta(t-\tau),$$
$$Ex_0w'(t)=Ex_0v'(t)=Ew(t)v'(\tau)=0.$$

Also it is assumed that all the random variables x_0, $w(t)$, $v(t)$ are Gaussian. The FI smoothing problem is to compute, for all $t\in[0,T]$, the conditional expectation of $x(t)$ given the observations over $[0,T]$, i.e., the smoothed estimate is

$$\hat{x}_s(t)=E\{x(t)|y(\tau),0\leqq\tau\leqq T\} \tag{1.3}$$

and the corresponding smoothed error covariance is

$$P_s(t)=E\{[x(t)-\hat{x}_s(t)][x(t)-\hat{x}_s(t)]'\} \tag{1.4}$$

As is well-known, the estimate $\hat{x}_s(t)$ is a linear functional of the observations and is also the maximum *a posteriori* estimate and the linear least-squares estimate.

Motivation

In order to gain some insight into the smoothing problem and to motivate the development of this paper, consider the time-invariant version of the smoothing problem with observations over the interval $(-\infty,+\infty)$. In the sequel, this will be referred to as the time-invariant infinite-lag smoothing problem. Also, it is assumed here that x and y are scalar random

processes and

$$y(t)=x(t)+v(t). \tag{1.5}$$

The Wiener filter will be used to examine the relationship between past and future observations in estimating $x(t)$.

The Wiener filter provides an estimate of $x(t)$ given $y(\tau)$, $\tau \leqq t$, in terms of a convolution integral as

$$\hat{x}(t|t)=\int_{-\infty}^{t} h(t-\tau)y(\tau)\,d\tau \tag{1.6}$$

where the filter inpulse response obeys the Wiener-Hopf Equation,

$$R_{xy}(t-\sigma)=\int_{-\infty}^{t} h(t-\tau)R_y(\tau-\sigma)\,d\sigma, \quad \sigma \leqq t \tag{1.7}$$

with $R_{xy}(t)=E[x(t)y(0)]$ and $R_y(t)=E[y(t)y(0)]$. A similar anti-causal expression can be obtained to provide an estimate $\hat{x}_r(t|t)$ of $x(t)$ from future observations,

$$\hat{x}_r(t|t)=\int_{t}^{\infty} h_r(t-\tau)y(\tau)\,d\tau \tag{1.8}$$

where the reserved-time filter obeys the Wiener-Hopf equation

$$R_{xy}(t-\sigma)=\int_{t}^{\infty} h_r(t-\tau)R_Y(\tau-\sigma)\,d\tau, \quad \sigma \geqq t. \tag{1.9}$$

Since the cross-correlation function between $x(t)$ and $y(t)$ is an even function of time, the filter impulse responses $h(t)$ and $h_r(t)$ are easily related,

$$\begin{aligned}
\int_{-\infty}^{t} h(t-\tau)R_y(\tau-\sigma)\,d\tau &= R_{xy}(t-\sigma), \quad \sigma \leqq t \qquad (1.10)\\
&= R_{xy}(t-s), \quad s-t=t-\sigma\\
&= \int_{t}^{\infty} h_r(t-\tau)R_y(\tau-s)\,d\tau, \quad \text{from (1.9)}\\
&= \int_{-\infty}^{-t} h_r(t+\mu)R_y(\mu+s)\,d\mu, \quad \mu=-\tau\\
&= \int_{-\infty}^{t} h_r(-t+\tau)R_y(s-2t+\tau)\,d\tau, \quad \tau=2t+\mu\\
&= \int_{-\infty}^{t} h_r(-t+\tau)R_y(\tau-\sigma)\,d\tau\\
&\Rightarrow h(t-\tau)=h_r(\tau-t). \qquad (1.11)
\end{aligned}$$

Equation (1.11) says that the weight $h(-\tau)$ given to $y(\tau)$ in forming $\hat{x}(0|0)$ is the same as the weight $h_r(\tau)$ given to $y(-\tau)$ in forming $\hat{x}_r(0|0)$. In this sense, the *same* linear filter is used to estimate $x(t)$ from either the past or the future observations. The only place that the assumption $x(t)$ and $y(t)$ are scalar processes is used is for the relation $R_{xy}(t)=R_{xy}(-t)$. Whenever $R_x(t)=R_x'(t)$, it follows that $R_x(t)$ and $R_{xy}(t)$ are even functions of time, and so the same proof will work for vector processes in this case. The vector case is addressed more completely in Section IV. As far as their relative performance, it is straightforward to show that both estimates $\hat{x}(t|t)$ and $\hat{x}_r(t|t)$ have the same mean-square error. Therefore the past and future contain equal amounts of information about $x(t)$, and one would expect equal weightings on both when forming the smoothed estimate.

That this is exactly the case can be seen from the Wiener smoother. The smoothed estimate is

$$\hat{x}_s(t)=\int_{-\infty}^{+\infty} h_s(t-\tau)y(\tau)\,d\tau \tag{1.12}$$

where

$$F\{h_s(t)\}=\frac{F\{R_{xy}(t)\}}{F\{R_y(t)\}}$$

and the operator $F\{\cdot\}$ is the Fourier transform. Since $R_{xy}(t)$ and $R_y(t)$ are even functions of t, their transforms will be purely real. Thus the transform of $h_s(t)$ will also be real, and so $h_s(t)$ must be even. This proves that the past and future contribute equally to the smoothed estimate of $x(t)$.

One popular solution to the FI smoothing problem is the Mayne–Fraser two-filter smoother [3], [4]. Section II provides an historical review of the two-filter smoother discussing the work of Mayne, Fraser and Mehra. The two-filter smoother gives the smoothed estimate as a combination of a forward and a backward estimate. Both estimates come from Kalman filters. A surprising fact, however, is that in the infinite-lag case when the state dimension equals one, the steady-state covariance of the backward is *always* larger than the covariance of the forward filter. (See (2.22) for the backwards covariance.) This is surprising in view of the previous development where the past and future observations were seen to be equally valuable in estimating $x(t)$.

The reason for this apparent contradiction is that the Mayne–Fraser two-filter smoother has a *built-in asymmetry* that is absent from the original problem. In Section III, it is shown that this asymmetry is due to the way in which the *a priori* information enters into the estimate of $x(t)$.

New forms of the two-filter smoother will be presented which are symmetric with respect to forward- and reversed-time. These smoothers are obtained from simple, first principles arguments using reversed-time realizations of the state process. Badawi *et al.* [5] have very recently derived similar smoothing formulas based on stochastic realizations. It is to be hoped that the analysis and discussion in Sections III and IV will enable the reader to obtain a clear understanding of how future observations are incorporated in the FI smoothing problem.

Section IV examines the symmetry between forward- and reversed-time in detail. It will be shown that this symmetry breaks down for the problem of change of initial conditions. In Section V, one of the forms of the two-filter smoother will be used to analyze reduced-order smoothers and to perform a sensitivity analysis.

Preliminaries

This introductory section closes with two well-known results from probability theory. The first deals with combining estimates that have independent errors.

PROPOSITION 1. *Let x, y_1 and y_2 be zero-mean Gaussian random variables, and let $\hat{x}_1$ and $\hat{x}_2$ be the Bayesian (maximum likelihood) estimates of x given y_1 and y_2, respectively, with associated error covariances P_1 and P_2. If the errors $x-\hat{x}_1$ and $x-\hat{x}_2$ are independent, then the Bayesian (maximum likelihood) estimate of x given both y_1 and y_2 is*

$$\hat{x}=P[P_1^{-1}\hat{x}_1+P_2^{-1}\hat{x}_2] \tag{1.13}$$

where the error covariance P is given by

$$P=[P_1^{-1}+P_2^{-1}]^{-1} \tag{1.14}$$

Proof. See Schweppe [6].

By abuse of terminology, the estimates $\hat{x}_1$ and $\hat{x}_2$ are often referred to as independent estimates. The second result is simply the formula for the conditional expectation of a Gaussian random variable.

PROPOSITION 2. *Let x be a Gaussian random variable with mean m and covariance Σ, and let y be an observation of x,*

$$y=Hx+v \tag{1.15}$$

where v is a zero-mean Gaussian random variable (with covariance R) that is independent of x and the rank of H equals the dimension of x. The

Bayesian estimate of x given y is

$$\begin{aligned}\hat{x} &= E[x|y] \\ &= P[(H'R^{-1})y + \Sigma^{-1}m]\end{aligned} \tag{1.16}$$

where the error covariance P satisfies

$$P = [(H'R^{-1}H) + \Sigma^{-1}]^{-1} \tag{1.17}$$

Proof. See Schweppe [6].

It is noted that Proposition 2 can be interpreted as meaning that the Bayesian estimate of x given y equals the maximum likelihood estimate of x given both y and the *a priori* mean m and covariance Σ. This interpretation is obtained by forming two independent maximum likelihood estimates of x, one based on y and one based on m and Σ. Combining these two maximum likelihood estimates by Proposition 1 yields (1.16) and (1.17). This idea of forming a Bayesian estimate as the combination of two maximum likelihood estimates, one based on observations and one based on *a priori* data, is one the keys to the solution of the smoothing problem in Section III.

II. HISTORICAL REVIEW OF THE TWO-FILTER SMOOTHER

The first solution of the FI smoothing problem as a combination of two estimates was presented by Mayne [3] in 1966. Fraser [4] pursued this idea and in 1967 showed how both of these estimates could be obtained from separate Kalman filters. Mehra [7] then attempted to derive the two-filter smoother from basic principles. The work of Mayne, Fraser, and Mehra is reviewed in this section and a critical analysis of this work will provide motivation for the development in the sequel.

The system considered by Mayne is a discrete-time analog of (1.1) and (1.2),

$$x(k+1) = \Phi(k+1, k)x(k) + w(k) \tag{2.1}$$

$$y(k) = C(k)x(k) + v(k) \tag{2.2}$$

where $x(0) = x_0$ and

$$Ex_0 = Ew(k) = Ev(k) = 0, \quad Ex_0x_0' = \Sigma_0$$

$$Ew(i)w'(k) = Q(k)\delta_{i,k}, \quad Ev(i)v'(k) = R(k)\delta_{i,k}$$

$$Ex_0w'(k) = Ex_0v'(k) = Ew(i)v'(k) = 0.$$

His starting point was the conditional probability density of the states given the observations,

$$p(x(0),\ldots,x(T)|y(0),\ldots,y(T)) = \frac{p(y(0),\ldots,y(T)|x(0),\ldots,x(T))p(x(0),\ldots,x(T))}{p(y(0),\ldots,y(T))}. \tag{2.3}$$

Because of the independence of the observation noise process $\{v(k)\}$, the likelihood function $p(y(0),\ldots,y(T)|x(0),\ldots,x(T))$ may be written as

$$\begin{aligned} p(y(0),\ldots,y(T)|x(0),\ldots,x(T)) &= \prod_{k=0}^{T} p(y(k)|x(k)) \\ &= K_1 \prod_{k=0}^{T} \exp\{-\tfrac{1}{2}\|y(k)-C(k)x(k)\|^2_{R^{-1}(k)}\}. \end{aligned} \tag{2.4}$$

Also, since the sequence $\{x(k)\}$ is a Markov process,

$$\begin{aligned} p(x(0),\ldots,x(T)) &= p(x(0)) \prod_{k=0}^{T-1} p(x(k+1)|x(k)) \\ &= K_2 \exp\{-\tfrac{1}{2}\|x(0)\|^2_{\Sigma_0^{-1}}\} \\ &\quad \times \prod_{k=0}^{T-1} \exp\{-\tfrac{1}{2}\|x(k+1)-\Phi(k+1,k)x(k)\|^2_{Q^{-1}(k)}\}. \end{aligned}$$

Therefore, substituting into (2.3) and realizing that $p(y(0),\ldots,y(T))$ is just a normalization constant yields

$$\begin{aligned} p(x(0),\ldots,x(T)|y(0),\ldots,y(T)) &= K_3 \exp\Bigg\{-\tfrac{1}{2}\|x(0)\|^2_{\Sigma_0^{-1}} - \tfrac{1}{2}\sum_{k=0}^{T} \|y(k)-C(k)x(k)\|^2_{R^{-1}(k)} \\ &\quad -\tfrac{1}{2}\sum_{k=0}^{T-1} \|x(k+1)-\Phi(k+1,k)x(k)\|^2_{Q^{-1}(k)}\Bigg\} \end{aligned} \tag{2.6}$$

where K_1, K_2, and K_3 are constants. The optimal smoothed estimates $\{\hat{x}(k|T)\}$ maximize the conditional density given in (2.6).

Consider now the negative of the exponent in the right-hand-side of (2.6),

$$\begin{aligned} J(x(0),\ldots,x(T)) &= \tfrac{1}{2}\|x(0)\|^2_{\Sigma_0^{-1}} + \tfrac{1}{2}\sum_{k=0}^{T} \|y(k)-C(k)x(k)\|^2_{R^{-1}(k)} \\ &\quad + \tfrac{1}{2}\sum_{k=0}^{T-1} \|x(k+1)-\Phi(k+1,k)x(k)\|^2_{Q^{-1}(k)} \end{aligned} \tag{2.7}$$

The smoothed estimates can be obtained from the *minimization* of the functional J. Mayne's approach to this minimization was to consider some fixed integer r between 0 and T and to define

$$J_r(x(r)) = \min_{\{x(k)|k \neq r\}} J(x(0), \ldots, x(T)) \tag{2.8}$$

$J_r(x(r))$ will be a quadratic form in $x(r)$, and therefore given $J_r(x(r))$ it is an easy matter to compute $\hat{x}(r|T)$ and $P(r|T)$. Hence the problem of interest is the determination of an expression for $J_r(x(r))$.

Mayne decomposes the minimization over $\{x(k)|k \neq r\}$ into two separate minimizations—one over $\{x(0), \ldots, x(r-1)\}$ and the other over $\{x(r+1), \ldots, x(T)\}$. Thus

$$J_r(x(r)) = J_{0,r}(x(r)) + J_{r,T}(x(r)) \tag{2.9}$$

where

$$\begin{aligned} J_{0,r}(x(r)) = \min_{\{x(0),\ldots,x(r-1)\}} \Bigg\{ & \tfrac{1}{2}\|x(0)\|^2_{\Sigma_0^{-1}} \\ & + \tfrac{1}{2}\sum_{k=0}^{r-1} \|y(k) - C(k)x(k)\|^2_{R^{-1}(k)} \\ & + \tfrac{1}{2}\sum_{k=0}^{r-1} \|x(k+1) - \Phi(k+1,k)x(k)\|^2_{Q^{-1}(k)} \Bigg\} \end{aligned} \tag{2.10}$$

and

$$\begin{aligned} J_{r,T}(x(r)) = \min_{\{x(r+1),\ldots,x(T)\}} \Bigg\{ & \tfrac{1}{2}\sum_{k=r}^{T} \|y(k) - C(k)x(k)\|^2_{R^{-1}(k)} \\ & + \tfrac{1}{2}\sum_{k=r}^{T-1} \|x(k+1) - \Phi(k+1,k)x(k)\|^2_{Q^{-1}(k)} \Bigg\} \end{aligned} \tag{2.11}$$

Both $J_{0,r}(x(r))$ and $J_{r,T}(x(r))$ are quadratic forms in $x(r)$, (say)

$$J_{0,r}(x(r)) = \tfrac{1}{2}x(r)'F_{0,r}x(r) + g'_{0,r}x(r) + h_{0,r} \tag{2.12}$$

$$J_{r,T}(x(r)) = \tfrac{1}{2}x(r)'F_{r,T}x(r) + g'_{r,T}x(r) + h_{r,t}. \tag{2.13}$$

Therefore,

$$\begin{aligned} J_r(x(r)) = \tfrac{1}{2}x(r)'(F_{0,r} + F_{r,T})x(r) \\ + (g_{0,r} + g_{r,T})'x(r) + (h_{0,r} + h_{r,T}) \end{aligned} \tag{2.14}$$

and so the smoothed estimate and covariance are simply

$$\hat{x}_s(r) = -(F_{0,r} + F_{r,T})^{-1}(g_{0,r} + g_{r,T}) \tag{2.15}$$

$$P_s(r) = (F_{0,r} + F_{r,T})^{-1} \tag{2.16}$$

What remains is to determine recursive expressions for $J_{0,r}(x(r))$ and $J_{r,T}(x(r))$.

First consider $J_{0,r}(x(r))$ defined by (2.10). Note that this is just the cost functional one would minimize to obtain the maximum *a posteriori* estimate $\hat{x}(r|r-1)$, i.e.

$$\tfrac{1}{2}x(r)'F_{0,r}x(r) + g'_{0,r}x(r) = \tfrac{1}{2}[x(r) - \hat{x}(r|r-1)]'P^{-1}(r|r-1) \cdot [x(r) - \hat{x}(r|r-1)] + \text{constant}. \tag{2.17}$$

Therefore,

$$F_{0,r} = P^{-1}(r|r-1) \tag{2.18}$$

$$g_{0,r} = P^{-1}(r|r-1)(\hat{x}(r|r-1)) \tag{2.19}$$

and $h_{r,0}$ is of no real interest. Moreover, the Kalman filter provides a recursive computation of $\hat{x}(r|r-1)$ and $P(r|r-1)$. Thus recursive expressions for $F_{0,r}$ and $g_{0,r}$ are available.

Second, consider $J_{r,T}(x(r))$. Clearly,

$$\begin{aligned} J_{r,T}(x(r)) &= \min_{\{x(r+1),\ldots,x(T)\}} \left\{ \tfrac{1}{2}\sum_{k=r}^{T} \|y(k) - C(k)x(k)\|^2_{R^{-1}(k)} \right. \\ &\quad \left. + \tfrac{1}{2}\sum_{k=r}^{T-1} \|x(k+1) - \Phi(k+1,k)x(k)\|^2_{Q^{-1}(k)} \right\} \\ &= \min_{\{w(r),\ldots,w(T-1)\}} \left\{ \tfrac{1}{2}\sum_{k=r}^{T} \|y(k) - C(k)x(k)\|^2_{R^{-1}(k)} \right. \\ &\quad \left. + \tfrac{1}{2}\sum_{k=r}^{T-1} \|w(k)\|^2_{Q^{-1}(k)} \right\} \end{aligned} \tag{2.20}$$

subject to the constraint $x(k+1) = \Phi(k+1,k)x(k) + w(k)$, $k = r, \ldots, T-1$. But (2.20) is just a linear-quadratic optimal control problem and can be solved using dynamic programming from T backwards to r. The well-known solution to this problem yields a recursion for $F_{r,T}$ and $g_{r,T}$ ($h_{r,T}$ is of no interest) in terms of $F_{r+1,T}$, $g_{r+1,T}$, and $y(r)$. Thus there exist recursive relations for $J_{0,r}(x(r))$ and $J_{r,T}(x(r))$, and so $\hat{x}(r|T)$ and $P(r|T)$ can be found from (2.15) and (2.16).

This approach to the fixed-interval smoothing problem is easily extended to the continuous-time case. The details may be found in Reference 3.

Mayne interprets this solution of the smoothing problem as a combination of two estimates. One estimate is based on past observations and

is obtained from Kalman filtering; optimal control theory is used to obtain a second estimate from future observations. From (2.10), the estimate based on past observations is a Bayesian estimate, but Mayne does not say what kind of estimate is the one based on future observations. In Section III, this second estimate will be shown to be a maximum likelihood estimate. The idea of expressing the smoothed estimate as a linear combination of the two estimates was pursued in 1967 by Donald Fraser [4] for both continuous-time and discrete-time.

One of Fraser's two estimates is based on past observations. This estimate and the corresponding covariance are just the outputs of a standard Kalman filter working forward over the data. Fraser's second estimate is obtained from a backwards Kalman filter, i.e., a filter operating on future observations from T to the present time t. The idea is to then combine these two estimates using the formulas (1.13) and (1.14) for the optimal combination of independent estimates.

The appropriate continuous-time backward filter is [4]

$$-\frac{d}{dt}\hat{x}_b(t) = -A(t)\hat{x}_b(t) + P_b(t)C'(t)R^{-1}(t)[y(t) - C(t)\hat{x}_b(t)] \tag{2.21}$$

$$-\frac{d}{dt}P_b(t) = -A(t)P_b(t) - P_b(t)A'(t) + Q(t) - P_b(t)C'(t)R^{-1}(t)C(t)P_b(t) \tag{2.22}$$

where $P_b^{-1}(T) = 0$ and $\lim_{t\to T}[P_b^{-1}(t)\hat{x}_b(t)] = 0$. The interpretation given by Fraser and Potter [8] is that $\hat{x}_b(t)$ is, "...the best estimate of the state at time t based upon all the measurements from time t to the end of the data interval." The terminal condition of an infinite covariance matrix is intended to reflect complete uncertainty about the state estimate at time T because of the complete lack of information about $x(T)$. Thus no terminal estimate can be made; only the limit of $P_b^{-1}(t)\hat{x}_b(t)$ can be specified. Because of these initial conditions, the filter must be implemented in the so-called "information filter" form [8].

The smoothed estimate is formed by combining the "independent" past and future estimates according to (1.13) and (1.14), viz.

$$\hat{x}_s(t) = P_s(t)[P^{-1}(t|t)\hat{x}(t|t) + P_b^{-1}(t)\hat{x}_b(t)] \tag{2.23}$$

$$P_s(t) = [P^{-1}(t|t) + P_b^{-1}(t)]^{-1}. \tag{2.24}$$

This is the same formula as Mayne's (2.15) and (2.16) in continuous-time if $P_b^{-1}(t) = F_{t,T}$ and $P_b^{-1}(t)\hat{x}_b(t) = -g_{t,T}$. Rather than showing these two equalities, Fraser's method of proof consists of re-deriving the smoothing

formulas of Rauch, Tung, Striebel from (2.21)–(2.24) and the usual Kalman filter equations. This is certainly a valid method of proof, and it does show that the smoothed estimate is given by (2.23) and (2.24). What is not clear, however, is *why* the estimate $\hat{x}(t|t)$ and $\hat{x}_b(t)$ can be combined by (2.23) and (2.24) or *why* the estimate $\hat{x}_b(t)$ should be given by the backward Kalman filter (2.21) and (2.22). It would be desirable to answer these questions starting from basic principles.

Mehra [7] attempts to clarify these points in his doctoral thesis. First consider the backward filtering equations (2.21) and (2.22). By multiplying the state equation (1.1) times -1, Mehra obtains

$$-\frac{d}{dt}x(t)=[-A(t)]x(t)-w(t). \tag{2.25}$$

He then applies the usual Kalman filter equations to this backwards system by letting $\tau=T-t$ and thereby obtains (2.21) and (2.22). This same argument was later adopted by Kailath and Frost [10]. It is incorrect, however, because "future" (with respect to τ) values of the driving noise w are correlated with the present state (see Ljung and Kailath [11] where this observation was first made). That is, (2.25) is not a usual Markovian realization. Therefore, it is not possible to blindly apply the Kalman filter to (2.25) and obtain the backward filter (2.21) and (2.22).

Mehra also addresses the question of independence of the estimates $\hat{x}(t|t)$ and $\hat{x}_b(t)$. His approach is to write the differential equations for the forward error $\tilde{x}(t|t)$ and the backward error $\tilde{x}_b(t)$,

$$\frac{d}{dt}\tilde{x}(t|t)=[A(t)-K(t)C(t)]\tilde{x}(t|t)+w(t)-K(t)v(t) \tag{2.26}$$

$$-\frac{d}{dt}\tilde{x}_b(t)=[-A(t)-K_b(t)C(t)]\tilde{x}_b(t)-w(t)-K_b(t)v(t). \tag{2.27}$$

Equation (2.26) is integrated from 0 to t while (2.27) is integrated from T to t. Thus Mehra points out that $\tilde{x}(t|t)$ depends on $\{w(\tau),v(\tau)|0<\tau\leqq t\}$ and $\tilde{x}_b(t)$ depends on $\{w(\tau),v(\tau)|t<\tau\leqq T\}$—two independent sets of noises. Is this sufficient for the conclusion that $\tilde{x}(t|t)$ and $\tilde{x}_b(t)$ are independent? Obviously not, $\tilde{x}(t|t)$ and $\tilde{x}_b(t)$ may be dependent because of their *initial values*. For example, the random variable $x(0)$ is correlated with (in fact, equal to) $\tilde{x}(0|0)$ and therefore with $\tilde{x}(t|t)$. Is $\tilde{x}_b(T)$ also correlated with $x(0)$? Mehra can't say because at this point in his development he has not specified any initial values for $\tilde{x}_b(T)$ or $P_b(T)$. It should be clear that without such a specification, the independence of $\tilde{x}(t|t)$ and $\tilde{x}_b(t)$ is indeterminate. Nevertheless, Mehra prematurely de-

clares that they are independent because they are functions of independent sets of noises. The independence of these two estimates will be examined further in Section III.

The behavior of the two-filter smoother when there are errors in the various model parameters (such as the system matrix or initial covariance) was also considered by Mehra [7]. Following the work of Nishimura [12] and Fitzgerald [13], Mehra performs a covariance analysis to obtain an expression for the covariance of the forward and backward Kalman filters when the true system and the model on which the filters are based differ.

In order to obtain the smoothed error covariance, he combines these two covariances assuming the forward and backward errors of the mismatched filters are uncorrelated. This is not the case, and in Section V an expression for the smoothed error covariance is found which includes the correlation between the forward and backward errors.

In summary, this section has presented the two-filter smoother as developed by Mayne [3] and Fraser [4]. This solution of the FI smoothing problem is unique compared to the Rauch, Tung, Striebel [9] smoother, and others [2], in that it is not given as a correction to the Kalman filter estimate at the same point. Rather, it takes the form of a combination of two optimal linear filter estimates. The work of Mehra [7] was primarily directed toward deriving this smoother from basic principles. The problem with Mehra's derivation is the use of an incorrect reversed-time model. In Section III, reversed-time Markov models are employed in an attempt at obtaining the two-filter smoother from basic principles by carefully considering the use of future observations for estimating the present state.

III. A NEW SOLUTION TO THE FIXED-INTERVAL SMOOTHING PROBLEM

Motivation

When forming the smoothed Bayesian estimate of $x(t)$, there are three separate sets of information,

i) *a priori* data, $Ex(t)=0$ and $Ex(t)x'(t)=\Sigma(t)$,

ii) past observations, $\{y(\tau)|0\leqq\tau\leqq t\}$,

iii) future observations, $\{y(\tau)|t<\tau\leqq T\}$.

Intuitively, the smoothed estimate should incorporate each of these sets exactly once. In this section, it will be shown how this incorporation takes place for the two-filter smoother. The main contributions here are:

a) use of a backwards filter that has finite covariance at all times;
b) derivation of the two-filter smoother from first principles.

Section III contains the derivation; discussion is deferred to Section IV. Also, it is noted that the assumption of a zero-mean process is not at all essential and will be relaxed later.

The filtered estimate $\hat{x}(t|t)$ is based on the *a priori* data and the past observations. This is easily obtained via the Kalman filter. Also, it is a simple matter to form an estimate of $x(t)$ from just the *a priori* data, i.e., the estimate is zero and the covariance is $\Sigma(t)$. What is less well-known, however, is how to use *future observations* in forming an estimate of $x(t)$. To this end, reversed-time Markov models will be introduced in the next subsection. When combined with the Kalman filter, these reversed-time models yield the expression for a Bayesian estimate of $x(t)$ *based on* a priori *data and future observations.*

Only the continuous-time problem is considered in Section III. The analogous results for the discrete-time version are presented in Appendix A. Note that with respect to the continuous-time problem, the present observation $y(t)$ is a linear measurement of $x(t)$ corrupted by additive noise having an *infinite* covariance. Thus the isolated observation $y(t)$ contains no information about the process $x(t)$. This remark is purely formal, of course, as is the entire development of this section. These arguments can be made rigorous, but for ease of presentation and understanding, a formal development is deemed preferable. The future observations, therefore, can be defined as $\{y(\tau)|t \leqq \tau \leqq T\}$, where now $y(t)$ is included in the future observations, without altering the analysis. This definition effectively symmetrizes the smoothing problem with respect to forward- and reversed-time. This situation is in contrast to the discrete-time case where the present observation contains non-zero information and which is basically *asymmetric.* It is the present observation which is the major cause of any differences between the equations of Section III and those of Appendix A.

Reversed-time Markov models

Essentially simultaneously in the summer of 1976, several authors introduced reversed-time Markov models [11], [14], [15]. These reversed-time models generated processes that had the same second-order statistics as the corresponding forward-time processes. More recently, these results have been strengthened by Verghese and Kailath [16] and Lindquist and Picci [17] to provide backwards Markov models with sample path equivalence to the forward-time processes.

Corresponding to the forward system of (1.1), consider the reversed-time model

$$-\frac{d}{dt}x_r(t)=[-A(t)-Q(t)\Sigma^{-1}(t)]x_r(t)-\xi(t) \tag{3.1}$$

where

$$Ex_r(T)=0, \quad E\xi(t)=0, \quad Ex_r(T)x_r'(T)=\Sigma(T)$$
$$E\xi(t)\xi'(\tau)=Q(t)\delta(t-\tau), \quad E\xi(t)x_r'(T)=0.$$

Equation (3.1) is meant to denote that the reversed-time process $x_r(t)$ propagates backwards from T to 0.

THEOREM 1. *The stochastic process $x(t)$ of* (1.1) *and the stochastic process $x_r(t)$ of* (3.1) *are sample path equivalent if*

$$\begin{aligned}\xi(t)&=w(t)-E[w(t)|x(\tau), t\leqq\tau\leqq T]\\&=w(t)-Q(t)\Sigma^{-1}(t)x(t).\end{aligned}$$

Proof See Reference 16.

As pointed out by Verghese and Kailath, the process $\xi(t)$ is a white process with the same variance as $w(t)$ and is uncorrelated with $x(t)$. It differs from $w(t)$, however, by the conditional expectation of the noise given the state. It is exactly this conditioned expectation which is missing from Mehra's backwards system (2.25).

Since $x(t)$ and $x_r(t)$ are completely indistinguishable, (1.1) and (3.1) are simply two representations of the same stochastic process. One implication of this equivalence is that the observations $y_r(t)$,

$$y_r(t)=C(t)x_r(t)+v(t) \tag{3.2}$$

are sample path equivalent to the observations $y(t)$ on $x(t)$. Another implication is that any least-squares linear estimator of $x(t)$ is also a least-squares linear estimator for $x_r(t)$, and vice versa. That is, given any set of observations, the estimate of $x(t)$ is the *same* functional on these observations as is the estimate of $x_r(t)$. This is a key point in the development of the sequel.

Recall the time-varying Lyapunov equation describing the state covariance $\Sigma(t)$,

$$\frac{d}{dt}\Sigma(t)=A(t)\Sigma(t)+\Sigma(t)A'(t)+Q(t). \tag{3.3}$$

Using this equation, the reversed-time system matrix $-A(t)-Q(t)\Sigma^{-1}(t)$ may be written as

$$-A(t)-Q(t)\Sigma^{-1}(t)=\Sigma(t)A'(t)\Sigma^{-1}(t)+\left(\frac{d}{dt}\Sigma^{-1}(t)\right)\Sigma(t). \tag{3.4}$$

The stability of the reversed-time system can now be examined from (3.4). Consider first the case of a time-invariant system in the steady-state. Then the reversed-time system matrix is simply $\Sigma A'\Sigma^{-1}$. But A must be a stability matrix and has the same eigenvalues as A'. Moreover, $\Sigma A'\Sigma^{-1}$ is just a similarity transformation of A' and therefore has the same eigenvalues. The conclusion then is that the reversed-time system matrix $\Sigma A'\Sigma^{-1}$ is also a stability matrix and has the same eigenvalues as the forward system matrix A. Therefore the forward-time process $x(t)$ and the reversed-time process $x_r(t)$ both have stable realizations.

In the time-varying case, it is necessary to consider the adjoint system of (1.1)

$$-\frac{d}{dt}p(t)=A'(t)p(t). \tag{3.5}$$

This system propagating backwards in time has the same stability properties as the original forward system. Let $z(t)=\Sigma(t)p(t)$. Then

$$-\frac{d}{dt}z(t)=\left[\Sigma(t)A'(t)\Sigma^{-1}(t)+\left(-\frac{d}{dt}\Sigma(t)\right)\Sigma^{-1}(t)\right]z(t). \tag{3.6}$$

The resulting reversed-time system matrix is, according to (3.4), the same as the system matrix for the reversed-time process x_r. The system (3.1) will have the same stability properties as (3.5) and hence as (1.1), if $z(t)=\Sigma(t)p(t)$ is a Lyapunov transformation [18]. For this transformation to be a Lyapunov transformation, the following conditions must hold [18]:

i) Σ has a continuous derivative;

ii) Σ and $(d/dt)\Sigma$ are bounded;

iii) there exists a constant m such that $0<m\leqq|\det\Sigma(t)|$, $\forall t$.

Assuming these conditions are met, the forward- and reversed-time realizations (1.1) and (3.1) possess identical stability properties.

An estimate based on future observations plus *a priori* information

The conditional expectation of $x_r(t)$ (or $x(t)$) given the future observations is denoted $\hat{x}_r(t|t)$,

$$\hat{x}_r(t|t)=E\{x_r(t)|y(\tau),t\leqq\tau\leqq T\}. \tag{3.7}$$

The process $x_r(t)$ is a Gauss–Markov process in reversed-time as given by (3.1). Therefore, this Bayesian estimate can be computed from the Kalman filter for the reversed-time system model. Explicitly,

$$-\frac{d}{dt}\hat{x}_r(t|t)=[-A(t)-Q(t)\Sigma^{-1}(t)]\hat{x}_r(t|t)+K_r(t)[y(t)-C(t)\hat{x}_r(t|t)] \tag{3.8}$$

$$\begin{aligned}-\frac{d}{dt}P_r(t|t)=&[-A(t)-Q(t)\Sigma^{-1}(t)]P_r(t|t)\\&+P_r(t|t)[-A(t)-Q(t)\Sigma^{-1}(t)]'+Q(t)\\&-P_r(t|t)C'(t)R^{-1}(t)C(t)P_r(t|t)\end{aligned} \tag{3.9}$$

$$K_r(t)=P_r(t|t)C'(t)R^{-1}(t) \tag{3.10}$$

where $\hat{x}_r(T|T)=0$ and $P_r(T)=\Sigma(T)$. Note that the conditions at time T for this filter are finite, in contrast with the initial conditions of Fraser's backward filter (2.21) and (2.22).

The Bayesian estimate $\hat{x}_r(t|t)$ of $x(t)$ is a combination of *a priori* information and the future observations. The possibility of forming two estimates of $x(t)$ from these two separate data will now be addressed. The estimate of $x(t)$ from only the *a priori* data is just

$$\hat{x}_{\text{a.p.}}(t)=0 \tag{3.11}$$

$$P_{\text{a.p.}}(t)=\Sigma(t) \tag{3.12}$$

because it is being assumed that $x(t)$ is a zero-mean process. An estimate of $x(t)$ based only on the future observations is considered next.

View the future observations, $\{y(\tau)|t\leqq\tau\leqq T\}$, as one aggregate linear observation of $x(t)$ corrupted by the noises $\{w(\tau),v(\tau)|t\leqq\tau\leqq T\}$. If $x(t)$ is treated as an unknown parameter instead of as a random variable with a known probability distribution, one can speak of the maximum likelihood estimate of $x(t)$. Let $\hat{x}_{\text{future}}(t)$ denote the maximum likelihood of $x(t)$ based only on the future observations, i.e., *a priori* data about $x(t)$ is not used. Denoted by $P_{\text{future}}(t)$ the corresponding error covariance. If $x(t)$ is observable over the interval $[t,T]$, then $\hat{x}_{\text{future}}(t)$ and $P_{\text{future}}(t)$ are well defined, and the estimate may be written as

$$\hat{x}_{\text{future}}(t)=x(t)+\mu(t) \tag{3.13}$$

where the error $\mu(t)$ is solely due to the driving noise and observation noise, i.e., $Ex(t)\mu'(t)=0$, and $E\mu(t)\mu'(t)=P_{\text{future}}(t)$. (See Appendix A for an

elaboration of this discussion in the discrete-time case.) Because the estimates $\hat{x}_{\text{future}}(t)$ and $\hat{x}_{\text{a.p.}}(t)$ have independent errors, Propositions 1 and 2 may be used formally to write the Bayesian estimate of $x(t)$ given the future observations as the combination of $\hat{x}_{\text{future}}(t)$ and $\hat{x}_{\text{a.p.}}(t)$, viz

$$\hat{x}_r(t|t)=P_r(t|t)[P^{-1}_{\text{future}}(t)\hat{x}_{\text{future}}(t)+\Sigma^{-1}(t)\hat{x}_{\text{a.p.}}(t)]$$
$$=P_r(t|t)P^{-1}_{\text{future}}(t)\hat{x}_{\text{future}}(t) \tag{3.14}$$
$$P_r(t|t)=[P^{-1}_{\text{future}}(t)+\Sigma^{-1}(t)]^{-1}. \tag{3.15}$$

An interesting feature of (3.14) and (3.15) is that they are invertible in that it is possible to solve for $\hat{x}_{\text{future}}(t)$ and $P_{\text{future}}(t)$ in terms of the other quantities. This yields

$$\hat{x}_{\text{future}}(t)=P_{\text{future}}(t)P_r^{-1}(t|t)\hat{x}_r(t|t) \tag{3.16}$$
$$P_{\text{future}}(t)=[P_r^{-1}(t|t)-\Sigma^{-1}(t)]^{-1}. \tag{3.17}$$

Differentiating (3.16) and (3.17) with respect to $-t$ yields differential equations for the maximum likelihood estimate $\hat{x}_{\text{future}}(t)$ and $P_{\text{future}}(t)$ propagating backwards from T. The result is Fraser's backwards filter.

THEOREM 2. *The maximum likelihood estimate $\hat{x}_{\text{future}}(t)$ and covariance $P_{\text{future}}(t)$ are identically equal to Fraser's $\hat{x}_b(t)$ and $P_b(t)$. That is*

$$\hat{x}_b(t)=P_b(t)P_r^{-1}(t|t)\hat{x}_r(t|t) \tag{3.18}$$
$$P_b(t)=[P_r^{-1}(t|t)-\Sigma^{-1}(t)]^{-1}. \tag{3.19}$$

Proof (Appendix C).

This result says that the *a priori* information can be "subtracted out" from the conditional expectation of $x(t)$ to form the backward estimate. Moreover, this backward estimate is the maximum likelihood estimate of $x(t)$. The conditional expectation comes from a reversed-time Kalman filter. This reversed-time Kalman filter has a finite initial covariance. Using this Kalman filter together with the Lyapunov equation for the state covariance has yielded a differential equation for the maximum likelihood estimate of $x(t)$ based on future observations.

The solution

THEOREM 3 *The smoothed Bayesian estimate and covariance satisfy*

$$\hat{x}_s(t)=P_s(t)[P^{-1}(t|t)\hat{x}(t|t)+P_r^{-1}(t|t)\hat{x}_r(t|t)] \tag{3.20}$$
$$P_s(t)=[P^{-1}(t|t)+P_r^{-1}(t|t)-\Sigma^{-1}(t)]. \tag{3.21}$$

Proof Two representations of the process $x(t)$ have been given, a forward-time realization (1.1) and a reversed-time realization (3.1). In the proof of this theorem, a third representation—a "combined" realization—is introduced which combines (1.1) and (3.1) to propagate the state process both forward and backward from time t. Consider a fixed time t and let the process $x_c(t)$ be generated by

$$\frac{d}{d\tau}x_c(\tau)=A(\tau)x_c(\tau)+w(\tau), \quad \tau>t$$

$$-\frac{d}{d\tau}x_c(\tau)=[-A(\tau)-Q(\tau)\Sigma^{-1}(\tau)]x_c(\tau)+\xi(\tau), \quad \tau<t$$

where

$$Ex_c(t)=0, \quad Ew(\tau)=0, \quad E\xi(\tau)=0, \quad Ex_c(t)x_c'(t)=\Sigma(t),$$
$$Ew(\tau)w'(\sigma)=E\xi(\tau)\xi'(\sigma)=Q(\tau)\delta(\tau-\sigma),$$
$$Ew(\tau)x_c'(t)=E\xi(\tau)x_c'(t)=Ew(\tau)\xi(\sigma)=0.$$

These differential equations are meant to denote that $x_c(\tau)$ may be written as

$$x_c(\tau)=\begin{cases}\int_t^\tau A(\sigma)x_c(\sigma)\,d\sigma+\int_t^\tau dw(\sigma)+x_c(t), & \tau>t\\ \int_t^\tau[-A(\sigma)-Q(\sigma)\Sigma^{-1}(\sigma)]x_c(\sigma)\,d\sigma+\int_t^\tau d\xi(\sigma)+x_c(t), & \tau<t\end{cases}.$$

The process $x_c(\tau)$ is easily shown to be sample path equivalent to $x(\tau)$ and $x_r(\tau)$. Let $\hat{x}_{\text{past}}(t)$ and $P_{\text{past}}(t)$ be the maximum likelihood estimate of $x(t)$ and the error covariance given the past observations only, i.e., without the *a priori* information. By applying the same argument that was used in Section III for $\hat{x}_{\text{future}}(t)$ and $P_{\text{future}}(t)$ to $\hat{x}_{\text{past}}(t)$ and $P_{\text{past}}(t)$ one obtains

$$\hat{x}_{\text{past}}(t)=P_{\text{past}}(t)P^{-1}(t|t)\hat{x}(t|t)$$

$$P_{\text{past}}(t)=[P^{-1}(t|t)-\Sigma^{-1}(t)]^{-1}.$$

Consider now the third realization, i.e., the process $x_c(\tau)$. The errors in the maximum likelihood estimates $\hat{x}_{\text{past}}(t)$ and $\hat{x}_{\text{future}}(t)$ are caused by $\{\xi(\tau), v(\tau)|0\leqq\tau<t\}$ and $\{w(\tau), v(\tau)|t<\tau\leqq T\}$ respectively. Therefore these estimates are independent estimates, and Proposition 1 can be used to obtain the maximum likelihood estimate of $x_c(t)$ (or $x(t)$) given all the

observations,

$$\hat{x}_{\mathrm{ML}}(t)=P_{\mathrm{ML}}(t)[P_{\mathrm{past}}^{-1}(t)\hat{x}_{\mathrm{past}}(t)+P_{\mathrm{future}}^{-1}(t)\hat{x}_{\mathrm{future}}(t)]$$
$$P_{\mathrm{ML}}(t)=[P_{\mathrm{past}}^{-1}(t)+P_{\mathrm{future}}^{-1}(t)]^{-1}.$$

Since $x(t)$ is zero-mean, the smoothed Bayesian estimate is found from Proposition 2,

$$\begin{aligned}\hat{x}_s(t)&=P_s(t)P_{\mathrm{ML}}^{-1}(t)\hat{x}_{\mathrm{ML}}(t)\\&=P_s(t)[P_{\mathrm{past}}^{-1}(t)\hat{x}_{\mathrm{past}}(t)+P_{\mathrm{future}}^{-1}(t)\hat{x}_r(t|t)]\\&=P_s(t)[P^{-1}(t|t)\hat{x}(t|t)+P_r^{-1}(t|t)\hat{x}_t(t|t)]\\P_s(t)&=[P_{\mathrm{ML}}^{-1}(t)+\Sigma^{-1}(t)]^{-1}\\&=[P_{\mathrm{past}}^{-1}(t)+P_{\mathrm{future}}^{-1}(t)+\Sigma^{-1}(t)]^{-1}\\&=[P^{-1}(t|t)+P_r^{-1}(t|t)-\Sigma^{-1}(t)]^{-1}. \qquad \text{Q.E.D.}\end{aligned}$$

Aside An alternative proof of Theorem 3 is to note that substitution of (3.18) into (2.23) yields (3.20) and substitution of (3.19) into (2.24) yields (3.21).

IV. DISCUSSION AND EXTENSIONS

Theorem 3 expresses the smoothed estimate as a combination of two filtered estimates:

a) one estimate from a forward Kalman filter for the forward system model;

b) one estimate from a reversed-time Kalman filter for the reversed-time model.

These estimates are not independent, however, because they both include the *a priori* information. In essence, the effects of the *a priori* must be subtracted out once when obtaining the smoothed estimate. The two sets (past and future) of observations may be said to be independent observations of $x(t)$ by considering the "combined" representation and noting that the two sets of noises, $\{\xi(\tau), v(\tau)|0\leqq\tau<t\}$ and $\{w(\tau), v(\tau)|t<\tau\leqq T\}$, are independent.

In order for the maximum likelihood estimates $\hat{x}_{\mathrm{past}}(t)$ and $\hat{x}_{\mathrm{future}}(t)$ to be well-defined, it is necessary that $x(t)$ be observable over $[0,t]$ and $[t,T]$, respectively. In case $x(t)$ is not observable, it is still possible to form a maximum likelihood estimate of the observable part of $x(t)$. Then $P_{\mathrm{past}}^{-1}(t)$ and/or $P_{\mathrm{future}}^{-1}(t)$ must be replaced with pseudo-inverses in the proof of Theorem 3. Thus, the symmetric form of the smoothing Eqs. (3.20) and (3.21) holds without any observability conditions on $x(t)$.

We note that even though this development assumed the random process $x(t)$ was zero-mean, the case of a nonzero-mean process is easily handled. Letting $z(t)$ equal $x(t)$ minus the mean value of $x(t)$ yields a zero-mean process obeying the same state Eq. (1.1). Then taking as observations of $z(t)$ the observations $y(t)$ minus the mean value of $y(t)$ produces a smoothing problem of the form studied here. The smoothed estimate of $x(t)$ is simply the smoothed estimate of $z(t)$ plus the mean of $x(t)$.

An important characteristic of this smoother is that both the forward and backward models used in forming the two Kalman filters are stable (assuming the original forward realization is stable). This feature will allow a covariance analysis in Section V that requires the integration of only *stable* differential equations. This is different than earlier studies.

One striking characteristic of the smoother in Theorem 3 is the complete symmetry with respect to forward-time vs. reversed-time. Equations (3.20) and (3.21) are called symmetric because the estimates $\hat{x}(t|t)$ and $\hat{x}_r(t|t)$ are conditional expectations of $x(t)$ given the past and the future observations, respectively. This is certainly in contrast with the usual two-filter smoother of Mayne and Fraser. This symmetry between forward- and reversed-time will be developed and discussed in the remainder of this section.

For the special case of smoothing over the interval $(-\infty, +\infty)$ with a time-invariant system, it is possible to investigate the symmetry between forward- and reversed-time in detail. If the system matrix A is a scalar, then the reversed-time realization is identical to the forward-time realization. Thus the two Kalman filters are identical and so the two steady-state error covariances P and P_r are equal. From (3.20), it is clear that this implies the two estimates $\hat{x}(t|t)$ and $\hat{x}(t|t)$ are weighted equally in forming the smoothed estimate. This confirms the intuitive expectation of Section I that the future and past should be equivalent.

The reversed-time system (3.1) equals the forward-time system (1.1) in the time-invariant infinite-lag case if and only if the autocorrelation function of x is symmetric, i.e., $\Sigma A'\Sigma^{-1} = A$ if and only if $R_x(t) = R'_x(t)$. Of course $R_x(t) = R'_x(t)$ is equivalent to $R_x(t) = R_x(-t)$, the condition needed in Section I to show that the same Wiener filter can be used to estimate $x(t)$ from either the future or the past. Therefore, whenever the autocorrelation function is even, the future and past observations are equally weighted in forming an estimate of the present.

Having considered the LTI infinite-lag smoothing problem, we now return to the general case. Two key ingredients of the two-filter smoother are the maximum likelihood estimates $\hat{x}_{\text{past}}(t)$ and $\hat{x}_{\text{future}}(t)$. Theorem 2 showed that $\hat{x}_{\text{future}}(t)$ equals Fraser's backward estimate $\hat{x}_b(t)$. There is

also a backwards estimate $\hat{x}_{r_b}(t)$ of the reversed-time process $x_r(t)$. The estimate $\hat{x}_{r_b}(t)$ is based on the observations $\{y(\tau)|0 \leqq \tau \leqq t\}$ and may be shown (by repeating the proof of Theorem 2) to equal $\hat{x}_{\text{past}}(t)$. The differential equations for $\hat{x}_{r_b}(t)$ and $P_{r_b}(t)$ are just Fraser's (2.21) and (2.22) with the reversed-time system matrix in place of the forward system matrix,

$$\frac{d}{dt}\hat{x}_{r_b}(t) = [-A(t) - Q(t)\Sigma^{-1}(t)]\hat{x}_{r_b}(t) + K_{r_b}(t)[y(t) - C(t)\hat{x}_{r_b}(t)] \quad (4.1)$$

$$\begin{aligned}\frac{d}{dt}P_{r_b}(t) = {} & [-A(t) - Q(t)\Sigma^{-1}(t)]P_{r_b}(t) \\ & + P_{r_b}(t)[-A(t) - Q(t)\Sigma^{-1}(t)]' \\ & + Q(t) - P_{r_b}(t)C'(t)R^{-1}(t)C(t)P_{r_b}(t) \end{aligned} \quad (4.2)$$

where

$$K_{r_b}(t) = P_{r_b}(t)C'(t)R^{-1}(t),$$
$$P_{r_b}^{-1}(0) = 0,$$

and

$$\lim_{t \to 0} [P_{r_b}^{-1}(t)\hat{x}_{r_b}(t)] = 0.$$

This filter must be implemented as an information filter because of the initial conditions.

Using the estimate $\hat{x}_{r_b}$, it is possible to obtain another version of the two-filter smoother

$$\hat{x}_s(t) = P_s(t)[P_{r_b}^{-1}(t)\hat{x}_{r_b}(t) + P_r^{-1}(t|t)\hat{x}_r(t|t)] \quad (4.3)$$

$$P_s(t) = [P_{r_b}^{-1}(t) + P_r^{-1}(t|t)]^{-1}. \quad (4.4)$$

This is essentially Fraser's smoother (2.23) and (2.24) applied to the reversed-time realization instead of the usual forward realization. The *a priori* information is combined with the future observations to form one estimate; the second estimate is formed from the past observations alone. All of the smoothing algorithms presented by Sidhu and Desai [14] are in this same spirit—they are obtained by applying a standard smoothing algorithm to the reversed-time model.

Ljung and Kailath [19] have addressed the problem of converting linear least-squares filtered and smoothed estimates derived for one set of initial conditions to estimates valid for some other set. They showed how these estimates are altered because of changes in the assumed values for the mean and variance of the initial value $x(0)$. The motivation for their

work was, "...the possibility of deliberately using an incorrect initial value in order to use the fast Chandrasekhar-type equations..."

One could consider using some of the formulas derived here to solve the change of initial conditions problem. For example, propagate $\hat{x}_{r_b}(t)$ and then later incorporate the covariance Σ using (1.16) and (1.17) to construct the filtered estimate $\hat{x}(t|t)$. Indeed, applying the smoothing formula (4.3) and (4.4) at time T, the filtered estimate equals the smoothed estimate and can be written as

$$\hat{x}(T|T)=P(T|T)P_{r_b}^{-1}(t)\hat{x}_{r_b}(T) \tag{4.5}$$

$$P(T|T)=[P_{r_b}^{-1}(T)+\Sigma^{-1}(T)]^{-1}. \tag{4.6}$$

Certainly (4.5) and (4.6) seem to accomplish what was just proposed.

This scheme, however, does not quite work because of the equation for $\hat{x}_{r_b}(t)$. From (4.1), the computation of this estimate involves the reversed-time system matrix $-A(t)-Q(t)\Sigma^{-1}(t)$. That is to say, $\hat{x}_{r_b}(t)$ depends on the covariance Σ, and if the covariance were unknown, then it would be impossible to compute $\hat{x}_{r_b}$. The point here is that $\hat{x}_{r_b}(t)$ is a maximum likelihood estimate of $x_r(t)$ based on its "future" observations (future with respect to $-t$, i.e., $\{y(\tau)|t\geqq\tau\geqq 0\}$. When the covariance of $x(t)$ is $\Sigma(t)$ as used in the reversed-time realization, the two processes $x(t)$ and $x_r(t)$ are sample path equivalent. But if $\Sigma(0)$ is unknown, then one does not know which reversed-time model to use in the computation of $\hat{x}_{r_b}(t)$.

One conclusion from the above analysis is that the smoothing formulas developed here do not yield a simple change of initial conditions result for the forward filtered estimate. It will now be shown that such a result can be obtained for the reversed-time estimate $\hat{x}_r(t|t)$.

The problem of interest is to compute $\hat{x}_r(t|t)$, the filtered estimate of the reversed-time realization given the observations from T to t, by first processing the observations assuming the state covariance is π and then correcting this result for the actual value, Σ, of the covariance. The backward Kalman filter is designed for the reversed-time system

$$-\frac{d}{dt}x_r^{\pi}(t)=[-A(t)-Q(t)\pi^{-1}(t)]x_r^{\pi}(t)+\xi(t) \tag{4.7}$$

where $Ex_r^{\pi}(T)x_t^{\pi}(T)'=\pi(T)$ and

$$\frac{d}{dt}\pi(t)=A(t)\pi(t)+\pi(t)A'(t)+Q(t) \tag{4.8}$$

instead of the reversed-time system (3.1), i.e., Σ, the true state covariance, is replaced with π, a quantity which also obeys the Lyapunov equation.

The output of this reversed-time Kalman filter, initialized at time T with covariance $\pi(t)$, is denoted $\hat{x}_r^\pi(t|t)$. From (3.14), (3.15) and Theorem 2, the estimate $\hat{x}_r^\pi(t|t)$ and covariance $P_r^\pi(t|t)$ are related to Fraser's $\hat{x}_b(t)$ and $P_b(t)$ by

$$\hat{x}_r^\pi(t|t) = P_r^\pi(t|t)P_b^{-1}(t)\hat{x}_b(t) \tag{4.9}$$

$$P_r^\pi(t|t) = [P_b^{-1}(t) + \pi^{-1}(t)]^{-1}. \tag{4.10}$$

Solving (4.9) and (4.10) for $\hat{x}_b(t)$ and $P_b^{-1}(t)$ and using (3.14) and (3.15) again yields

$$\hat{x}_r(t|t) = P_r(t|t)P_r^\pi(t|t)^{-1}\hat{x}_r^\pi(t|t) \tag{4.11}$$

$$P_r(t|t) = [P_r^\pi(t|t)^{-1} - \pi^{-1}(t) + \Sigma^{-1}(t)]^{-1}. \tag{4.12}$$

The very natural interpretation of this result is that the incorrect covariance π is removed from the estimate and then the correct covariance Σ is added.

Thus in contrast to the forward-time case, very simple change of initial conditions formulas are obtained for the reversed-time estimate and covariance. The explanation for this is that $\hat{x}_b(t)$, the maximum likelihood estimate of $x(t)$ given the future observations, does not depend on the state covariance—the system matrix used in the filter for $\hat{x}_b(t)$ is just $-A(t)$. Using (4.11) and (4.12), it is possible to generalize (3.20) and (3.21) of Theorem 3 in that the covariance Σ can be replaced by an arbitrary function π satisfying (4.8).

$$\hat{x}_s(t) = P_s(t)[P^{-1}(t|t)\hat{x}(t|t) + P_r^\pi(t|t)^{-1}\hat{x}_r^\pi(t|t)] \tag{4.13}$$

$$P_s(t) = [P^{-1}(t|t) + P_s^\pi(t|t)^{-1} - \pi^{-1}(t)]^{-1}. \tag{4.14}$$

These expressions are similar to ones obtained by Ljung and Kailath [1] using the relationship between linear least-squares estimation and scattering theory.

Consider the implementation implications of this last observation. Basically any legitimate covariance function can be used in the reversed-time system matrix and Kalman filter. This added flexibility may be quite useful, especially when the forward system is time-invariant. In this case, one could use the steady-state covariance and thereby attain a time-invariant reversed-time model. This eliminates some of the problems involved with directly implementing the reversed-time filter.

V. COVARIANCE ANALYSIS FOR MISMATCH BETWEEN SYSTEM AND SMOOTHER

Covariance analysis is concerned with the increase in the smoothed error caused by using incorrect model parameters. For example, if a smoother is implemented with the system matrix $A^*(t)$ in place of the correct matrix $A(t)$, what is the resulting error covariance? Another example would be a reduced order smoother that uses a model of lower dimension than the actual system. The issue addressed by covariance analysis is of considerable practical importance. Griffin and Sage [20] have treated the sensitivity aspects of covariance analysis for discrete-time processes by considering the Rauch, Tung, Striebel [9] smoother. The analysis given here is thought to be the first correct treatment of the two-filter smoother and is performed for both continuous- and discrete-time (see Appendix A for the discrete-time results). Throughout this section, explicit time dependence will often be suppressed for ease of presentation.

The model used by the smoother for the dynamics and observations is

$$\frac{d}{dt}x^* = A^*x^* + w^* \tag{5.1}$$

$$y^* = C^*x^* + v^* \tag{5.2}$$

where w^* and v^* are independent white noise processes with covariances Q^* and R^*, respectively and $Ex^*(0)x^{*\prime}(0) = \Sigma^*(0)$. The superscript asterisk will be used consistently to denote model parameters as distinguished from the true system parameters. It is assumed that there is an output z of the actual system defined by

$$z = Hx \tag{5.3}$$

which is approximated by the output z^* of the model,

$$z^* = H^*x^*. \tag{5.4}$$

The only restrictions imposed on the model are that y^* and z^* have the same dimensions as y and z, respectively.

A smoothed estimate of the output is obtained as $\hat{z}_s^* = H^*\hat{x}_s^*$, and the question is, "What is the covariance of $z - \hat{z}_s^*$?" In order to determine an expression for this covariance, first the forward-time system and filter are jointly analyzed, and then a similar analysis is performed for the reversed-time system and filter. Next, the correlation between the forward- and reversed-time estimates are obtained. All these results are finally combined to yield the output error covariance.

Forward-time system and filter The model (5.1) and (5.2) can be used to design a reduced order Kalman filter,

$$\frac{d}{dt}\hat{x}^* = A^*\hat{x}^* + K^*[y - C^*\hat{x}^*] \tag{5.5}$$

$$K^* = P^*C^{*\prime}R^{*-1} \tag{5.6}$$

$$\frac{d}{dt}P^* = A^*P^* + P^*A^{*\prime} + Q^* - P^*C^{*\prime}R^{*-1}C^*P^* \tag{5.7}$$

where $\hat{x}^*(0)=0$ and $P^*(0)=\Sigma^*(0)$. Notice that the input to this filter is the actual observations y, of course. By combining the estimate $\hat{x}^*$ with the actual state x, one obtains an augmented state vector having dynamics

$$\frac{d}{dt}\begin{bmatrix} x \\ \hat{x}^* \end{bmatrix} = \begin{bmatrix} A & 0 \\ K^*C & A^* - K^*C^* \end{bmatrix}\begin{bmatrix} x \\ \hat{x}^* \end{bmatrix} + \begin{bmatrix} u \\ K^*v \end{bmatrix} \tag{5.8}$$

The covariance of this augmented state is defined as

$$\begin{bmatrix} \Sigma & M \\ M' & N \end{bmatrix} = E\left\{\begin{bmatrix} x \\ \hat{x}^* \end{bmatrix}[x' \quad \hat{x}^{*\prime}]\right\} \tag{5.9}$$

and must obey the Lyapunov equation

$$\frac{d}{dt}\begin{bmatrix} \Sigma & M \\ M' & N \end{bmatrix} = \begin{bmatrix} A & 0 \\ K^*C & A^* - K^*C^* \end{bmatrix}\begin{bmatrix} \Sigma & M \\ M' & N \end{bmatrix} + \begin{bmatrix} \Sigma & M \\ M' & N \end{bmatrix}\begin{bmatrix} A & 0 \\ K^*C & A^* - K^*C^* \end{bmatrix}' + \begin{bmatrix} Q & 0 \\ 0 & K^*RK^{*\prime} \end{bmatrix} \tag{5.10}$$

Reversed-time system and filter The reduced order reversed-time system corresponding to the model (5.1) is

$$-\frac{d}{dt}x_r^* = [-A^* - Q^*\Sigma^{*-1}]x_r^* + \xi^* \tag{5.11}$$

where ξ^* is a white noise process with covariance Q^*, $Ex_r^*(T)x_r^{*\prime}(T) = \Sigma^*(T)$, and Σ^* is given by

$$\frac{d}{dt}\Sigma^* = A^*\Sigma^* + \Sigma^*A^{*\prime} + Q^*. \tag{5.12}$$

By analogy with the forward-time case, there exists a reduced-order reversed-time Kalman filter having gain K_r^* and producing the estimate

$\hat{x}_r^*$. The reversed-time augmented system is

$$-\frac{d}{dt}\begin{bmatrix} x_r \\ \hat{x}_r^* \end{bmatrix} = \begin{bmatrix} -A-Q\Sigma^{-1} & 0 \\ K_r^*C & -A^*-Q^*\Sigma^{*-1}-K_r^*C^* \end{bmatrix}\begin{bmatrix} x_r \\ \hat{x}_r^* \end{bmatrix} + \begin{bmatrix} \xi \\ K_r^*v \end{bmatrix}. \tag{5.13}$$

Let the covariance be

$$\begin{aligned} \begin{bmatrix} \Sigma & M_r \\ M_r' & N_r \end{bmatrix} &= E\left\{\begin{bmatrix} x_r \\ \hat{x}_r^* \end{bmatrix}[x_r' \quad \hat{x}_r^{*\prime}]\right\} \\ &= E\left\{\begin{bmatrix} x \\ \hat{x}_r^* \end{bmatrix}[x_r' \quad \hat{x}_r^{*\prime}]\right\} \end{aligned} \tag{5.14}$$

since x and x_r are stochastically indistinguishable. Then

$$\begin{aligned} -\frac{d}{dt}\begin{bmatrix} \Sigma & M_r \\ M_r' & N_r \end{bmatrix} &= \begin{bmatrix} -A-Q\Sigma^{-1} & 0 \\ K_r^*C & -A^*-Q^*\Sigma^{*-1}-K_r^*C^* \end{bmatrix}\begin{bmatrix} \Sigma & M_r \\ M_r' & N_r \end{bmatrix} \\ &\quad + \begin{bmatrix} \Sigma & M_r \\ M_r' & N_r \end{bmatrix}\begin{bmatrix} -A-Q\Sigma^{-1} & 0 \\ K_r^*C & -A^*-Q^*\Sigma^{*-1}-K_r^*C^* \end{bmatrix}' \\ &\quad + \begin{bmatrix} Q & \\ 0 & K_r^*RK_r^{*\prime} \end{bmatrix}. \end{aligned} \tag{5.15}$$

Cross-correlation of $\hat{x}^*$ and $\hat{x}_r^*$ The preceding analysis has shown how solving the two time-varying Lyapunov Eqs. (5.10) and (5.15) yields the covariances of $\hat{x}^*$ and $\hat{x}_r^*$ and their cross-correlations with x. Before one can obtain an expression for the smoothed error covariance, it is also necessary to know the cross-correlation of $\hat{x}^*$ and $\hat{x}_r^*$.

LEMMA 1

$$E\hat{x}^*(t)\hat{x}_r^{*\prime}(t) = \alpha^*(t)\Sigma(t)\beta^*(t) \tag{5.16}$$

where

$$\frac{d}{dt}\alpha^*(t) = [A^*-K^*C^*]\alpha^*(t) + \alpha(t)[-A-Q\Sigma^{-1}] + K^*C \tag{5.17}$$

$$-\frac{d}{dt}\beta^*(t) = [A]'\beta^*(t) + \beta^*(t)[-A^*-Q^*\Sigma^{*-1}-K_r^*C^*]' + C'K_r^{*\prime} \tag{5.18}$$

with initial conditions $\alpha^*(0)=0$ and $\beta^*(T)=0$.

Proof This lemma is proved in Appendix B.

The cross-correlation between $\hat{x}^*(t)$ and $\hat{x}_r^*(t)$ is given by the relatively simple expression (5.16) of Lemma 1 where α^* and β^* obey differential equations of the Lyapunov type. The coefficients of α^* and β^* in these equations are:

a) $A^* - K^*C^*$: forward-time *filter* matrix;
b) $-A - Q\Sigma^{-1}$: reversed-time *system* matrix;
c) A': forward-time *system* matrix transposed;
d) $[-A^* - Q^*\Sigma^{*-1} - K_r^*C^*]'$: reversed-time *filter* matrix transposed.

Notice the striking symmetry. In the linear time-invariant infinite-lag case, all four of these matrices are stability matrices. The steady-state algebraic version of (5.17) and (5.18) will, therefore, always have unique solutions.

Everything necessary for the evaluation of the smoothed output error covariance is now available. The next theorem puts it all together.

THEOREM 4 *The error covariance of the smoothed output estimate is given by*

$$\begin{aligned}\operatorname{cov}[z(t)-\hat{z}_s^*(t)] = {} & H\Sigma H' - H^*\Sigma_s^*[P^{*-1}M' \\ & + P_r^{*-1}M_r']H' - H[MP^{*-1} + M_rP_r^{*-1}]\Sigma_s^*H^{*\prime} \\ & + H^*\Sigma_s^*[P^{*-1}NP^{*-1} + P_r^{*-1}\beta^{*\prime}\Sigma\alpha^{*\prime}P^{*-1} \\ & + P^{*-1}\alpha^*\Sigma\beta^*P_r^{*-1} + P_r^{*-1}N_rP_r^{*-1}]\Sigma_s^*H^{*\prime} \qquad (5.19)\end{aligned}$$

where Σ, M, N *come from* (5.9); M_r *and* N_r *come from* (5.14); *and* α^* *and* β^* *come from* (5.17) *and* (5.18).

Proof (Appendix C).

This theorem is the main result of this section. While tedious, the proof is basically straightforward; it consists of finding the various cross-covariances. In fact, the only difficult part of the derivation was finding an expression for $E\hat{x}^*\hat{x}_r^{*\prime}$.

As an aside, it is noted that the cross-correlations between the estimates, $\hat{x}^*$ and $\hat{x}_r^*$, and the state x could have been evaluated in an analogous fashion to the way $E\hat{x}^*(t)\hat{x}_r^*(t)$ was obtained. This is done in Appendix B where it is shown that

$$E\hat{x}^*(t)x'(t) = \alpha^*\Sigma. \qquad (5.20)$$

But $E\hat{x}^*(t)x'(t) = M'(t)$ from (5.9). Therefore

$$\alpha^* = M'\Sigma^{-1}. \qquad (5.21)$$

Similarly,

$$\beta^* = \Sigma^{-1} M_r. \tag{5.22}$$

Since the Lyapunov equations (5.10) and (5.15) have to be solved and yield M and M_r, (5.21) and (5.22) allow the evaluation of the output covariance without the solution of the additional Eqs. (5.17) and (5.18) for α^* and β^*.

In the case where the dimension of the model (5.1) equals the dimension of the actual system (1.1), Theorem 4 can be used to obtain an expression for the covariance of $x(t) - \hat{x}_s^*(t)$. This results from simply setting H and H^* equal to the identity. Alternatively, the expression for the smoothed error covariance could be obtained by expressing the smoothed error as a linear combination of three errors—forward error, reversed-time error, and *a priori* error,

$$x - \hat{x}_s^* = \Sigma_s^* [P^{*-1}(x - \hat{x}^*) + P_r^{*-1}(x - \hat{x}_r^*) - \Sigma^{*-1}(x)]. \tag{5.23}$$

The covariances of these errors and their cross-correlations can be found from (5.10), (5.15), and (5.16). By making these substitutions and performing some tedious algebraic manipulations, one is able to arrive at an expression for the smoothed error covariance from (5.23). The details are omitted.

In summary, this section has addressed the problem of fixed-interval smoothing using an incorrect model. The actual smoothed output error covariance is given in Theorem 4. A special case of this result is a sensitivity analysis expression for the smoother error covariance. In either case, it is necessary to solve the forward- and reversed-time Lyapunov Eqs. (5.10) and (5.15). The quantities α^* and β^* also obey Lyapunov equations, but can be computed (perhaps more conveniently) from (5.21) and (5.22).

VI. CONCLUSIONS

The two-filter smoother expresses the smoothed state estimate as a linear combination of two optimal estimates. One of the main contributions of this paper has been to obtain this smoother from first principles. Other derivations of the two-filter smoother have proceeded by showing equivalence with some other smoothing algorithms. Because of these derivations, it has never been clear exactly what type of estimate is the backwards estimate $\hat{x}_b(t)$. Therefore, perhaps more important than the derivation of a new two-filter smoother in this paper is the insight gained from this approach. The backwards estimate is simply the maximum likelihood

estimate. The backwards filter comes from removing the *a priori* information from a reversed-time Kalman filter. This reversed-time Kalman filter, a key element throughout the paper, is designed from a reversed-time realization of the state process. Other authors [14] have used the reversed-time model to obtain smoothing formulas, but these results essentially just applied standard smoothing formulas to the reversed-time model. Section III used the reversed-time filter in conjunction with the forward filter to obtain the resulting expression for the smoothed estimate. It should be noted that some of the equations in Section III are quite similar to ones obtained by Ljung and Kailath [1] by using the relationship between linear least-squares estimation and scattering theory. The approach taken here seems to be a natural one for addressing the smoothing problem and yields the very simple change of initial conditions formula (4.13) and (4.14) for $\hat{x}_s(0)$.

The smoothing formulas presented here are symmetric with respect to forward-time vs. reversed-time. This is not meant to imply that the two filter error covariances $P(t)$ and $P_r(t)$ are equal, but rather the form of the smoother is the same for both the past and the future. For example, the two estimates $\hat{x}(t|t)$ and $\hat{x}_r(t|t)$ that are combined to produce the smoothed estimate are both conditional expectations of $x(t)$. As discussed in Section I, intuitively there is an equivalence between past and future observations. Where the difference between forward- and reversed-time became apparent was in Section IV when the question of uncertain initial covariance was considered. The reversed-time system matrix $-A(t)-Q(t)\Sigma^{-1}(t)$ obviously depends on the state covariance; the forward system matrix $A(t)$ is independent of Σ. Therefore, when considering change of initial covariance problems, the fact that the original system model is given in forward-time introduces a distinction between forward- and reversed-time. One can form a maximum likelihood estimate of $x_r(t)$ (given the observations from t to T) which can be combined with the *a priori* data to provide a simple change of initial conditions formula. There does not exist an analogous formula for the forward estimate $\hat{x}(t|t)$. This formula for $\hat{x}_r(t|t)$ can be used to provide a change of initial conditions formula for $\hat{x}_s(0)$ since at time 0, all the observations are used in estimating $x_r(0)$.

In order to implement the two-filter smoother given in Theorem 3, it is very convenient to use the information filter form of the forward- and reversed-time Kalman filters. This means one should compute $P^{-1}(t|t)$, $P_r^{-1}(t|t)$, $P^{-1}(t|t)\hat{x}(t|t)$, and $P_r^{-1}(t|t)\hat{x}_r(t|t)$ instead of the usual Kalman filter estimate and covariance. If these quantities are available, then only one matrix inversion is needed in the computation of the smoothed estimate—the inverse of $P^{-1}(t|t)+P_r^{-1}(t|t)-\Sigma^{-1}(t)$ is all that is required.

The final contribution of this paper is the analysis of reduced-order smoothers and the sensitivity of two-filter smoothers. The approach taken here is similar to that of Mehra [7] except that Mehra erroneously assumed the forward and backward filtered errors were uncorrelated. Hence the main contribution of Section V is Lemma 1 which gives the cross-correlation between $\hat{x}^*(t)$ and $\hat{x}_r^*(t)$, the forward- and reversed-time estimates. Another important aspect of the analysis in Section V is that both Lyapunov equations (5.10) and (5.15) can correspond to table systems. In particular, for the time-invariant infinite-lag problem, the forward- and reversed-time augmented systems are both stable and, therefore, algebraic equations can be solved to yield the covariances.

References

[1] L. Ljung and T. Kailath, A unified approach to smoothing formulas, *Automatica* **12** No. 2 (1976), 147–157.

[2] J. S. Meditch, A survey of data smoothing for linear and nonlinear dynamic systems, *Automatica* **9** No. 2 (1973), 151–162.

[3] D. Q. Mayne, A solution of the smoothing problem for linear dynamic systems, *Automatica* **4** No. 2 (1966), 73–92.

[4] D. C. Fraser, *A New Technique for the Optimal Smoothing of Data*, Sc.D. Dissertation, Mass. Inst. Tech., Cambridge, MA, January 1967.

[5] F. Badawi, A. Lindquist and M. Pavon, A stochastic realization approach to the smoothing problem, to appear.

[6] F. C. Schweppe, *Uncertain Dynamic Systems*, Prentice-Hall, Inc., Englewood Cliffs, N.J. (1973).

[7] R. K. Mehra, *Studies in Smoothing and in Conjugate Gradient Methods Applied to Optimal Control Problems*, Ph.D. dissertation, Harvard University, Cambridge, MA, May 1967.

[8] D. C. Fraser and J. E. Potter, The optimum linear smoother as a combination of two optimum linear filters, *IEEE Trans. Automat. Contr.* **AC-14,** No. 4 (1969), 387–390.

[9] H. E. Rauch, F. Tung and C. T. Striebel, On the maximum likelihood estimates for linear dynamic systems, *AIAA Journal* **3** No. 8 (1965), 1445–1450.

[10] T. Kailath and P. Frost, An innovations approach to least-squares estimation—Part II: Linear smoothing in additive white noise, *IEEE Trans. Automat. Contr.* **AC-13,** No. 6 (1968), 646–660.

[11] L. Ljung and T. Kailath, Backward Markovian models for second-order stochastic processes, *IEEE Trans. Info. Th.* **IT-22** No. 4 (1976), 488–491.

[12] R. Nishimura, Error bounds of continuous Kalman filters and the application to orbit determination problems, *IEEE Trans. Automat. Contr.* **AC-12** No. 3 (1967), 268–275.

[13] R. J. Fitzgerald, Error divergence in optimal filtering problems, *Proc. 2nd IFAC Symp. on Automat. Contr. in Space* Vienna, Austria, September 1967.

[14] G. S. Sidhu and U. B. Desai, New smoothing algorithms based on reversed-time lumped models, *IEEE Trans. Automat. Contr.* **AC-21** No. 4 (1976), 538–541.

[15] G. D. Lainiotis, General backward Markov models, *IEEE Trans. Automat. Contr.* **AC-21** No. 4 (1976), 595–598.

[16] G. Verghese and T. Kailath, A further note on backwards Markovian models, *IEEE Trans. Info. Th.* **IT-25** No. 1 (1979), 121–124.

[17] A Lindquist and G. Picci, On the stochastic realization problem, *SIAM J. Cont. and Optim.* **17** No. 3 (1979).
[18] R. W. Brockett, *Finite Dimensional Linear Systems*, John Wiley & Sons, New York, N.Y. (1970).
[19] L. Ljung and T. Kailath, Efficient change of initial conditions, dual Chandrasekhar equations, and some applications, *IEEE Trans. Automat. Contr.* **AC-22** No. 3 (1977), 443–447.
[20] R. E. Griffin and A. P. Sage, Sensitivity analysis of discrete filtering and smoothing algorithms, *AIAA Journal* **7** No. 10 (1969), 1890–1897.
[21] B. Friedlander, T. Kailath and L. Ljung, Scattering theory and linear least-squares estimation—II. Discrete-time problems, *J. Franklin Inst.* **301** No. 1 (1976), 71–82.

Appendix A.

DISCRETE-TIME SMOOTHING FORMULAS

Consider the discrete-time linear system

$$x(k+1)=\phi(k+1,k)x(k)+w(k) \tag{A.1}$$

$$y(k)=C(k)x(k)+v(k) \tag{A.2}$$

where $x(0)$, $w(k)$, and $v(k)$ are all independent, zero-mean, Gaussian random variables and

$$Ex(0)x'(0)=\Sigma(0), \quad Ew(k)w'(i)=Q(k)\delta_{i,k}, \quad Ev(k)v'(i)=R(k)\delta_{i,k}.$$

The discrete-time FI smoothing problem is to compute the conditional expectation of $x(k)$ given the observations $\{y(i)|0\leqq i\leqq T\}$. The estimate is denoted $\hat{x}_s(k)$ and the error covariance is $P_s(k)$. The system covariance $Ex(k)x'(k)$ at time k is denoted by $\Sigma(k)$.

The reversed-time system corresponding to (A.1) is

$$\begin{aligned} x_r(k)&=[\Sigma(k)\phi'(k+1,k)\Sigma^{-1}(k+1)+\xi(k+1)\\ &=\phi_r(k,k+1)x_r(k+1)+\xi(k+1) \end{aligned} \tag{A.3}$$

where the covariance of the reversed-time white noise driving process is

$$\begin{aligned} Q_r(k+1)&=E\xi_r(k+1)\xi_r'(k+1)\\ &=\Sigma(k)-\Sigma(k)\phi'(k+1,k)\Sigma^{-1}(k+1)\phi(k+1,k)\Sigma(k) \end{aligned} \tag{A.4}$$

If the state transition matrix $\phi(k+1, k)$ is invertible, this covariance may be written as

$$\begin{aligned} Q_r(k+1) &= \Sigma(k)\phi'(k+1, k)\Sigma^{-1}(k+1)Q(k)\phi^{-1}(k+1, k)' \\ &= \phi_r(k, k+1)Q(k)\phi^{-1}(k+1, k)' \end{aligned} \tag{A.5}$$

The random variables $x_r(T)$ and $\xi(k)$ are independent, zero-mean, Gaussian and $Ex_r(T)x_r'(T) = \Sigma(T)$. Under these conditions, the processes $x(k)$ of (A.1) and $x_r(k)$ of (A.3) have the same covariance function and, therefore, the same joint probability density functions. It should be noted that Friedlander, Kailath, Ljung [21] and Sidhu, Desai [14] have previously given incorrect reversed-time realizations of the discrete-time process $x(k)$.

For the smoothing problem, there are four disjoint sets of information about $x(k)$,

1) past observations: $\{y(i) | 0 \leqq i < k\}$
2) future observations: $\{y(i) | k < i \leqq T\}$
3) present observation: $\{y(k)\}$
4) *a priori* data: mean 0 and covariance $\Sigma(k)$.

By grouping these sets of information in various ways, one arrives at the variety of estimated quantities below:

Kalman filter estimate	$\hat{x}(k \mid k) = 1) + 3) + 4)$
Kalman filter one-step predictor	$\hat{x}(k \mid k-1) = 1) + 4)$
reversed-time Kalman filter estimate	$\hat{x}_r(k \mid k) = 2) + 3) + 4)$
reversed-time Kalman filter one-step predictor	$\hat{x}_r(k \mid k+1) = 2) + 4)$
backwards estimate	$\hat{x}_b(k \mid k) = 2) + 3)$
present conditional expectation	$\hat{x}_{\text{p.a.p.}}(k \mid k) = 3) + 4)$
a priori estimate	$\hat{x}_{\text{a.p.}}(k) = 4)$

and others. The reversed-time Kalman filter estimate and one-step predicted estimate are obtained from applying the Kalman filter equations to the reversed-time realization (A.3). The estimate $\hat{x}_{\text{p.a.p.}}(k)$ is just $E[x(k) | y(k)]$. The backwards estimate $\hat{x}_b(k \mid k)$ is used in the discrete-time Mayne–Fraser smoother.

In Section III, the idea of forming an estimate from the future observations without including *a priori* information was discussed for the continuous-time case. We now digress in this paragraph in order to elaborate and expand on that discussion for the discrete-time case. The

future observations, $\{y(i)|i<k<T\}$, can be expressed as one aggregate observation, y, of $x(k)$,

$$y=\begin{bmatrix} y(k+1) \\ y(k+2) \\ \vdots \\ y(T) \end{bmatrix}$$

$$=\begin{bmatrix} C(k+1)\{\phi(k+1,k)x(k)+w(k)\}+v(k+1) \\ C(k+2)\{\phi(k+2,k)x(k)+\phi(k+2,k+1)w(k)+w(k+1)\}+v(k+2) \\ \vdots \\ C(T)\{\phi(T,k)x(k)+\sum_{i=k}^{T-1}\phi(T,i+1)w(i)\}+v(T) \end{bmatrix}$$

$$=Hx(k)+v \tag{A.6}$$

where $\phi(l,m)$ denotes $\prod_{j=m}^{l-1}\phi(j+1,j)$. This aggregate observation is simply a linear observation of $x(k)$ corrupted by additive noise—the additive noise being a function of $\{w(i-1),v(i)|k<i\leqq T\}$. To estimate $x(k)$ from y without using *a priori* data, view $x(k)$ *not* as a random variable, but rather as an unknown parameter. As an unknown parameter, one can speak of the maximum likelihood estimate of $x(k)$ given the future observations. This estimate is well-defined only if $x(k)$ is observable over $(k+1,T)$, i.e., if H in (A.6) is full rank. In this case, the well-known result is

$$\hat{x}_{\text{future}}(k)=(H'R^{-1}H)^{-1}H'R^{-1}y \tag{A.7}$$

with error covariance

$$P_{\text{future}}(k)=(H'R^{-1}H)^{-1} \tag{A.8}$$

where $R=Evv'$. This estimate corresponds to Fraser's $\hat{x}_b(k|k+1)$. When H is not of full rank, the pseudo-inverse of $R^{-1/2}H$ can be used to estimate the observable part of $x(k)$. It is easily seen from (A.6) and (A.7) that the error in $\hat{x}_{\text{future}}(k)$ is due solely to the additive noise in (A.6), i.e., the error is independent of the random variable $x(k)$. Hence, in either the full rank or non-full rank case, the maximum likelihood estimate $\hat{x}_{\text{future}}(k)$ can be combined with the *a priori* data using (1.13) and (1.14) to yield the Bayesian conditional expectation of $x(k)$ given $\{y(i)|k<i\leqq T\}$.

The smoothed estimate, of course, must incorporate all four sets of information exactly once. There obviously exists a plethora of ways to

combine these various estimates to obtain $\hat{x}_s(k)$,

$$\hat{x}_s(k)=\Sigma_s(k)[P^{-1}(k|k)\hat{x}(k|k)+P_b^{-1}(k|k+1)\hat{x}_b(k|k+1)] \tag{A.9a}$$

$$=\Sigma_s(k)[P^{-1}(k|k)\hat{x}(k|k)+P_r^{-1}(k|k+1)\hat{x}_r(k|k+1)] \tag{A.10a}$$

$$=\Sigma_s(k)[P^{-1}(k|k)\hat{x}(k|k)+P_r^{-1}(k|k)\hat{x}_r(k|k) -P_{\text{p.a.p.}}^{-1}(k|k)\hat{x}_{\text{p.a.p.}}(k|k)] \tag{A.11a}$$

$$=\Sigma_s(k)[P^{-1}(k|k-1)\hat{x}(k|k-1)+P_r^{-1}(k|k+1)\hat{x}_r(k|k+1) +P_{\text{p.a.p.}}^{-1}(k|k)\hat{x}_{\text{p.a.p.}}(k|k)] \tag{A.12a}$$

$$\Sigma_s(k)=[P^{-1}(k|k)+P_b^{-1}(k|k+1)]^{-1} \tag{A.9b}$$

$$=[P^{-1}(k|k)+P_r^{-1}(k|k+1)-\Sigma^{-1}(k)]^{-1} \tag{A.10b}$$

$$=[P^{-1}(k|k)+P_r^{-1}(k|k)-P_{\text{p.a.p.}}^{-1}(k|k)]^{-1} \tag{A.11b}$$

$$=[P^{-1}(k|k-1)+P_r^{-1}(k|k+1)+P_{\text{p.a.p.}}^{-1}(k|k)-2\Sigma^{-1}(k)]^{-1} \tag{A.12b}$$

The proof of these results is analogous to the proof of Theorem 3 and is omitted. Equation (A.9) is just the Mayne–Fraser two-filter smoother. Equation (A.10) expresses the smoothed estimate as a combination of two Kalman filter estimates. The other two formulas, (A.11) and (A.12), are included to show that the smoothed estimate can be written in terms of an expression that is symmetric with respect to forward- and reversed-time. There are, of course, many other possibilities besides (A.9)–(A.12) for giving the smoothed estimate.

For the reduced-order smoother analysis, the formula (A.10) will be used. Notice that this expression is asymmetric with respect to forward- and reversed-time, and so it is to be anticipated that the resulting reduced-order covariance expressions will also have some asymmetry. It is assumed that the model used in reduced-order smoothing is

$$x^*(k+1)=\phi^*(k+1,k)x^*(k)+w^*(k) \tag{A.13}$$

$$y^*(k)=C^*(k)x^*(k)+v^*(k) \tag{A.14}$$

$$z^*(k)=H^*(k)x^*(k). \tag{A.15}$$

The actual process $x(k)$ and observations $y(k)$ are generated by (A.1) and (A.2), and the actual output $z(k)$ is given by

$$z(k)=H(k)x(k). \tag{A.16}$$

The approach and assumptions are the same as in Section 4.4. Also, explicit dependence on time will be suppressed.

Forward time system and filter A reduced-order Kalman filter is designed on the basis of the model (A.13) and (A.14). Let K^* be the gain of this filter. Then consider the augmented state vector consisting of x and $\hat{x}^*$,

$$\begin{bmatrix} x(k+1) \\ \hat{x}^*(k+1|k+1) \end{bmatrix} = \begin{bmatrix} \phi & 0 \\ K^*C\phi & (I-K^*C^*)\phi^* \end{bmatrix} \begin{bmatrix} x(k) \\ \hat{x}^*(k|k) \end{bmatrix} + \begin{bmatrix} I & 0 \\ K^*C & K^* \end{bmatrix} \begin{bmatrix} u(k) \\ v(k+1) \end{bmatrix}. \tag{A.17}$$

Let

$$\begin{bmatrix} \Sigma & M \\ M' & N \end{bmatrix} = E\left\{ \begin{bmatrix} x \\ \hat{x}^* \end{bmatrix} [x' \quad \hat{x}^{*\prime}] \right\}. \tag{A.18}$$

This augmented state covariance is given by the discrete-time Lyapunov equation.

Reversed-time system and filter Let K_r^* be the filter gain of the reduced-order, reversed-time Kalman filter. Then

$$\begin{bmatrix} x_r(k) \\ \hat{x}_r^*(k|k+1) \end{bmatrix} = \begin{bmatrix} \phi_r & 0 \\ \phi_r^* K_r^* C & \phi_r^*(I-K_r^*C^*) \end{bmatrix} \begin{bmatrix} x_r(k+1) \\ \hat{x}_r^*(k+1|k+2) \end{bmatrix} + \begin{bmatrix} I & 0 \\ 0 & \phi_r^* K_r^* \end{bmatrix} \begin{bmatrix} \xi_r(k+1) \\ v(k+1) \end{bmatrix}. \tag{A.19}$$

Let the corresponding system covariance be

$$\begin{bmatrix} \Sigma & M_r \\ M_r' & N_r \end{bmatrix} = E\left\{ \begin{bmatrix} x_r \\ x_r^* \end{bmatrix} [x_r' \quad x_r^*] \right\}. \tag{A.20}$$

Cross-correlation of $x^*(k|k)$ and $x_r^*(k|k+1)$ Using the discrete-time versions of the arguments given in Section IV for continuous time yields

$$E[\hat{x}^*(k|k)x_r^*(k|k+1)] = \alpha^*(k)\Sigma(k)\beta^*(k) \tag{A.21}$$

where α^* and β^* are recursively computed from

$$\alpha^*(k) = K^*(k)C(k) + \{[I-K^*(k)C^*(k)]\phi^*(k,k-1)\} \times \alpha^*(k-1)\{\Sigma(k-1)\phi'(k,k-1)\Sigma^{-1}(k)\} \tag{A.22}$$

and

$$\beta^*(k)=\phi'(k+1,k)C(k)K_r^*(k+1)\Sigma^{*-1}(k+1)\phi^*(k+1,k)\Sigma^*(k)$$
$$+\{\phi'(k+1,k)\}\beta^*(k+1)$$
$$\times\{\Sigma^*(k)\phi^{*\prime}(k+1,k)\Sigma^{*-1}(k+1)[I-K_r^*(k+1)C^*(k+1)]\}' \tag{A.23}$$

with initial conditions $\alpha^*(-1)=\beta^*(T)=0$.

The smoothed output error covariance is therefore

$$\operatorname{cov}[z(k)-z^*(k)]=[H-H^*]E\left\{\begin{bmatrix} x(k) \\ \hat{x}_s^*(k)\end{bmatrix}[x'(k) \quad x_s^{*\prime}(k)]\right\}\begin{bmatrix} H' \\ H^{*\prime}\end{bmatrix}$$
$$=H\Sigma H'-H^*E\{\hat{x}_s^*x'\}H'-HE\{x\hat{x}_s^{*\prime}\}H^{*\prime}$$
$$+H^*E\{\hat{x}_s^*\hat{x}_s^{*\prime}\}H^{*\prime}. \tag{A.24}$$

Using (A.10) for the smoothed estimate,

$$E\{x\hat{x}_s^{*\prime}\}=E\{x[\hat{x}^{*\prime}P^{*-1}+\hat{x}_r^{*\prime}P_r^{*-1}]\Sigma_s^*$$
$$=[MP^{*-1}+M_rP_r^{*-1}]\Sigma_s^* \tag{A.25}$$

and

$$E\{\hat{x}_s^*\hat{x}_s^{*\prime}\}=E\{\Sigma_s^*[P^{*-1}\hat{x}^*+P_r^{*-1}\hat{x}_r^*][\hat{x}^{*\prime}P^{*-1}+\hat{x}_r^{*\prime}P_r^{*-1}]\Sigma_s^*\}$$
$$=\Sigma_s^*[P^{*-1}NP^{*-1}+P_r^{*-1}\beta^{*\prime}\Sigma\alpha^{*\prime}P^{*-1}$$
$$+P^{*-1}\alpha^*\Sigma\beta^*P_r^{*-1}+P_r^{*-1}N_rP_r^{*-1}]\Sigma_s^*. \tag{A.26}$$

Substituting (A.25) and (A.26) into (A.24) yields

$$\operatorname{cov}[z(k)-z^*(k)]=H\Sigma H'-H^*\Sigma_s^*[P^{*-1}M'+P_r^{*-1}M_r']H'$$
$$-H[MP^{*-1}+M_rP_r^{*-1}]\Sigma^*H^{*\prime}$$
$$+H^*\Sigma_s^*[P^{*-1}NP^{*-1}+P_r^{*-1}\beta^{*\prime}\Sigma\alpha^{*\prime}P^{*-1}$$
$$+P^{*-1}\alpha^*\Sigma\beta^*P_r^{*-1}+P_r^{*-1}N_rP_r^{*-1}]\Sigma_s^*H_s^{*\prime}. \tag{A.27}$$

The sensitivity analysis problem is solved by taking H and H^* equal to the identity matrix.

Appendix B.

CROSS-CORRELATION BETWEEN $\hat{x}^*(t)$ AND $\hat{x}_r^*(t)$

The estimate $\hat{x}^*$ can be written in integral form from the variation of constants formula as

$$\hat{x}^*(t)=\int_0^t \Phi_{A^*-K^*C^*}(t,\sigma)K^*(\sigma)y(\sigma)\,d\sigma \tag{B.1}$$

where $\Phi_{A^*-K^*C^*}$ is the state transition matrix of the forward Kalman filter,

$$\frac{d}{dt}\Phi_{A^*-K^*C^*}(t,0)=[A^*(t)-K^*(t)C^*(t)]\Phi_{A^*-K^*C^*}(t,0) \tag{B.2}$$

with the identity initial condition $\Phi_{A^*-K^*C^*}(0,0)=I$. Similarly, the reversed-time estimate can be written

$$\hat{x}_r^*(t)=\int_T^t \psi_{-A^*-Q^*\Sigma^{*-1}-K_r^*C^*}(t,\tau)K_r^*(\tau)y(\tau)\,d\tau \tag{B.3}$$

where

$$-\frac{d}{dt}\psi_{-A^*-Q^*\Sigma^{*-1}-K_r^*C^*}(t,T)$$
$$=[-A^*(t)-Q^*(t)\Sigma^{*-1}(t)]\psi_{-A^*-Q^*\Sigma^{*-1}-K_r^*C^*}(t,T) \tag{B.4}$$

and $\psi_{-A^*-Q^*\Sigma^{*-1}-K_r^*C^*}(T,T)=I$. Combining the integral expression (B.1) and (B.3) and taking the expectation yields

$$E\hat{x}^*(t)\hat{x}_r^{*\prime}(t)=\int_0^t\int_T^t \Phi_{A^*-K^*C^*}(t,\sigma)K^*(\sigma)$$
$$\times E[y(\sigma)y'(\tau)]K_t^{*\prime}(\tau)\psi'_{-A^*-Q^*\Sigma^{*-1}-K_r^*C^*}(t,\tau)\,d\tau\,d\sigma. \tag{B.5}$$

The autocorrelation function of y is evaluated in the following lemma:

LEMMA B.1 *For* $\sigma\leqq\tau$,

$$Ey(\sigma)y'(\tau)=C(\sigma)\psi_{-A-Q\Sigma^{-1}}(\sigma,t)\Sigma(t)\phi'_A(\tau,t)C'(\tau)+R(\sigma)\delta(\sigma-\tau) \tag{B.6}$$

where $t\in[\sigma,\tau]$ *and*

$$\frac{d}{dt}\Phi_A(\tau,t)=A(\tau)\Phi_A(\tau,t),\quad \Phi_A(t,t)=I \tag{B.7}$$

$$-\frac{d}{dt}\psi_{-A-Q\Sigma^{-1}}(\sigma,t)=[-A(\sigma)-Q(\sigma)\Sigma^{-1}(\sigma)]\,\psi_{-A-Q\Sigma^{-1}}(\sigma,t),$$
$$\psi_{-A-Q\Sigma^{-1}}(t,t)=I \tag{B.8}$$

i.e., Φ_A *and* $\psi_{-A-Q\Sigma^{-1}}$ *are the state transition matrices of the forward- and reversed-time systems, respectively.*

Proof

$$\begin{aligned} Ey(\sigma)y'(\tau) &= E\{[C(\sigma)x(\sigma)+v(\sigma)][C(\tau)x(\tau)+v(\tau)]'\} \\ &= C(\sigma)E\{x(\sigma)x'(\tau)\}C'(\tau)+E\{v(\sigma)v'(\tau)\} \\ &= C(\sigma)E\{[\psi_{-A-Q\Sigma^{-1}}(\sigma,t)x(t)][\Phi_A(\tau,t)x(t)]'\}C'(\tau)+R(\sigma)\delta(\sigma-\tau) \\ &= C(\sigma)\psi_{-A-Q\Sigma^{-1}}(\sigma,\tau)\Sigma(t)\Phi'_A(\tau,t)C'(\tau)+R(\sigma)\delta(\sigma-\tau). \qquad \text{Q.E.D.} \end{aligned}$$

Equation (B.6) can now be substituted into (B.5). Note that the term involving the delta function drops out because of the limits on the double integral. Thus

$$\begin{aligned} E\hat{x}^*(t)\hat{x}_r^{*\prime}(t) = \int_0^t \int_t^T & \Phi_{A^*-K^*C^*}(t,\sigma)K^*(\sigma)C(\sigma)\psi_{-A-Q\Sigma^{-1}}(\sigma,t)\Sigma(t) \\ & \times \Phi'_A(\tau,t)C'(\tau)K_r^{*\prime}(\tau)\psi'_{-A^*-Q^*\Sigma^{*-1}}(t,\tau)\,d\tau\,d\sigma. \end{aligned} \tag{B.9}$$

This result is now expressed as

LEMMA B.2

$$E\hat{x}^*(t)\hat{x}_r^{*\prime}(t) = \alpha^*(t)\Sigma(t)\beta^*(t) \tag{B.10}$$

where

$$\alpha^*(t) = \int_0^t \Phi_{A^*-K^*C^*}(t,\sigma)K^*(\sigma)C(\sigma)\psi_{-A-Q\Sigma^{-1}}(\sigma,t)\,d\sigma \tag{B.11}$$

$$\beta^*(t) = \int_t^T \Phi'_A(\tau,t)C'(\tau)K_r^{*\prime}(\tau)\psi'_{-A^*-Q^*\Sigma^{*-1}-K_r^*C^*}(t,\tau)\,d\tau \tag{B.12}$$

Proof Lemma B.2 is an immediate consequence of (B.9).

The integral expressions for α^* and β^* may be replaced by differential equations.

LEMMA B.3

$$\frac{d}{dt}\alpha^*(t) = [A^*-K^*C^*]\alpha^*(t)+\alpha^*(t)[-A-Q\Sigma^{-1}]+K^*C \tag{B.13}$$

$$-\frac{d}{dt}\beta^*(t) = [A]'\beta^*(t)+\beta^*(t)[-A^*-Q^*\Sigma^{*-1}-K_r^*C^*]'+C'K_r^{*\prime} \tag{B.14}$$

with initial conditions $\alpha^*(0)=0$ *and* $\beta^*(T)=0$.

Proof Differentiating (B.11) with respect to t yields

$$\begin{aligned}\frac{d}{dt}\alpha^*(t) &= \int_0^t \frac{d}{dt}\{\Phi_{A^*-K^*C^*}(t,\sigma)K^*(\sigma)C(\sigma)\psi_{-A-Q\Sigma^{-1}}(\sigma,t)\}\,d\sigma + K^*(t)C(t)\\ &= \int_0^t \{[A^*(t)-K^*(t)C^*(t)]\Phi_{A^*-K^*C^*}(t,\sigma)\\ &\quad \times K^*(\sigma)C(\sigma)\psi_{-A-Q\Sigma^{-1}}(\sigma,t)\\ &\quad + \Phi_{A^*-K^*C^*}(t,\sigma)K^*(\sigma)C(\sigma)\psi_{-A-Q\Sigma^{-1}}(\sigma,t)\\ &\quad \times[-A(t)-Q(t)\Sigma^{-1}(t)]\}\,d\sigma + K^*(t)C(t)\\ &= [A^*(t)-K^*(t)C^*(t)]\alpha^*(t) + \alpha^*(t)[-A(t)-Q(t)\Sigma^{-1}(t)]\\ &\quad + K^*(t)C(t).\end{aligned}$$

Equation (B.14) is obtained in a completely analogous fashion. Q.E.D.

Combining Lemmas B.2 and B.3 yields Lemma 1 of Section V.
Finally, notice that from (B.1)

$$\begin{aligned}E\hat{x}(t)x'(t) &= E\left\{\int_0^t \Phi_{A^*-K^*C^*}(t,\sigma)K^*(\sigma)y(\sigma)\,d\sigma x'(t)\right\}\\ &= \int_0^t \Phi_{A^*-K^*C^*}(t,\sigma)K^*(\sigma)C(\sigma)\psi_{-A-Q\Sigma^{-1}}(\sigma,t)\,d\sigma\Sigma(t)\\ &= \alpha^*(t)\Sigma(t). \qquad \text{(B.15)}\end{aligned}$$

Similarly,

$$Ex(t)\hat{x}_r^{*\prime}(t) = \Sigma(t)\beta^*(t). \qquad \text{(B.16)}$$

Appendix C.

PROOFS OF THEOREMS 2 AND 4

Proof of Theorem 2

Explicit dependence on t is suppressed throughout this proof (covariance). It is shown that P_b^{-1} equals $P_r^{-1}-\Sigma^{-1}$. At time T, $P_r^{-1}(T\,|\,T)-\Sigma^{-1}(T)=0=P_b^{-1}(T)$, so it suffices to show that the derivatives are

equal.

$$
\begin{aligned}
-\frac{d}{dt}P_b^{-1} &= -P_b^{-1}\left[-\frac{d}{dt}P_b\right]P_b^{-1} \\
&= P_b^{-1}A + A'P_b^{-1} - P_b^{-1}QP_b^{-1} + C'R^{-1}C \quad \text{from (2.22)} \\
&= (P_r^{-1} - \Sigma^{-1})A + A'(P_r^{-1} - \Sigma^{-1}) - (P_r^{-1} - \Sigma^{-1}) \\
&\quad \times Q(P_r^{-1} - \Sigma^{-1}) + C'R^{-1}C \quad \text{by hypothesis} \\
&= P_r^{-1}(A + Q\Sigma^{-1}) + (A + Q\Sigma^{-1})'P_r^{-1} - P_r^{-1}QP_r^{-1} \\
&\quad + C'R^{-1}C - \Sigma^{-1}A - A'\Sigma^{-1} - \Sigma^{-1}Q\Sigma^{-1} \quad \text{rearranging} \\
&= -\frac{d}{dt}P_r^{-1} + \frac{d}{dt}\Sigma^{-1} \quad \text{from (3.3) and (3.9)} \\
&= -\frac{d}{dt}[P_r^{-1} - \Sigma^{-1}].
\end{aligned}
$$

(estimate). The proof is completed by showing $P_b^{-1}\hat{x}_b$ equals $P_r^{-1}\hat{x}_r$. Once again, it suffices to demonstrate the equality of the derivatives since the quantities are equal at time T.

$$
\begin{aligned}
-\frac{d}{dt}[P_b^{-1}\hat{x}_b] &= \left(-\frac{d}{dt}P_b^{-1}\right)\hat{x}_b + P_b^{-1}\left(-\frac{d}{dt}\hat{x}_b\right) \\
&= (P_b^{-1}A + A'P_b^{-1} - P_b^{-1}QP_b^{-1} + C'R^{-1})\hat{x}_b \\
&\quad + P_b^{-1}(-A\hat{x}_b + P_bC'R^{-1}[y - C\hat{x}_b])
\end{aligned}
$$

from (2.21) and (2.22)

$$= (A' - P_b^{-1}Q)P_b^{-1}\hat{x}_b + C'R^{-1}y$$

combining terms

$$= (A' - [P_r^{-1} - \Sigma^{-1}]Q)P_r^{-1}\hat{x}_r + C'R^{-1}y$$

by hypothesis

$$= ([A + Q\Sigma^{-1}]' - P_r^{-1}Q)P_r^{-1}\hat{x}_r + C'R^{-1}y$$

rearranging terms

$$
\begin{aligned}
&= (P_r^{-1}[A + Q\Sigma^{-1}] + [A + Q\Sigma^{-1}]'P_r^{-1} \\
&\quad - P_r^{-1}QP_r^{-1} + C'R^{-1}C)\hat{x}_r \\
&\quad + P_r^{-1}([-A - Q\Sigma^{-1}]\hat{x}_r + P_rC'R^{-1}[y - C\hat{x}_r])
\end{aligned}
$$

adding and substracting $P_r^{-1}[A+Q\Sigma^{-1}]\hat{x}_r+C'R^{-1}C\hat{x}_r$

$$=\left(-\frac{d}{dt}P_r^{-1}\right)\hat{x}_r+P_r^{-1}\left(-\frac{d}{dt}\hat{x}_r\right)$$

$$=-\frac{d}{dt}[P_r^{-1}\hat{x}_r]. \qquad \text{Q.E.D.}$$

Proof of Theorem 4

$$\operatorname{cov}[z(t)-\hat{z}_s^*(t)]=\operatorname{cov}[Hx-H^*\hat{x}_s^*]$$

$$=[H \quad -H^*]E\left\{\begin{bmatrix} x \\ \hat{x}_s^* \end{bmatrix}[x' \quad \hat{x}_s^{*\prime}]\right\}\begin{bmatrix} H' \\ -H^{*\prime} \end{bmatrix}$$

$$=HE\{xx'\}H'-H^*E\{\hat{x}_s^*x'\}H'$$

$$-HE\{x\hat{x}_s^{*\prime}\}H^{*\prime}+H^*E\{\hat{x}_s^*\hat{x}_s^{*\prime}\}H^{*\prime}$$

(I.) $E\{xx'\}=\Sigma$

(II.) $E\{\hat{x}_s^*x'\}=\Sigma_s^*E\{[P^{*-1}\hat{x}^*+P_r^{*-1}\hat{x}_r^*]x\}$

$$=\Sigma_s^*[P^{*-1}M'+P_r^{*-1}M_r']$$

from (5.9) and (5.14).

(III.) $E\{\hat{x}_s^*\hat{x}_s^*\}=\Sigma_s^*E\{[P^{*-1}\hat{x}^*+P_r^{*-1}\hat{x}_r^*][P^{*-1}\hat{x}^*+P_r^{*-1}\hat{x}_r^*]'\}\Sigma_s^*$

$$=\Sigma_s^*[P^{*-1}NP^{*-1}+P_r^{*-1}\beta^{*\prime}\Sigma\alpha^{*\prime}P^{*-1}$$

$$+P^{*-1}\alpha^*\Sigma\beta^*P_r^{*-1}+P_r^{*-1}N_rP_r^{*-1}]\Sigma_s^*$$

from (5.9), (5.14) and (5.16).

$$\therefore \operatorname{cov}[z-\hat{z}_s^*]=H\Sigma H'-H^*\Sigma_s^*[P^{*-1}M'+P_r^{*-1}M_r']H'$$

$$-H[MP^{*-1}+M_rP_r^{*-1}]\Sigma_s^*H^{*\prime}$$

$$+H^*\Sigma_s^*[P^{*-1}NP^{*-1}+P_r^{*-1}\beta^{*\prime}\Sigma\alpha^{*\prime}P^{*-1}$$

$$+P^{*-1}\alpha^*\Sigma\beta^*P_r^{*-1}+P_r^{*-1}N_rP_r^{*-1}]\Sigma_s^*H^{*\prime}. \qquad \text{Q.E.D.}$$