Stable Fixed-Lag Smoothing of Continuous-Time Processes

SURAPONG CHIRARATTANANON, MEMBER, IEEE, AND BRIAN D. O. ANDERSON, MEMBER, IEEE

Abstract—Attempts to realize optimal fixed-lag smoothers for continuous-time processes have encountered two major difficulties: the presence of delay lines in the smoother, and the instability of most smoothers. This paper considers suboptimal smoothers which still outperform filters and have performance very close to the optimum. Design procedures are exhibited for smoothers that are both finite-dimensional and asymptotically stable, with the procedures depending on the delay line being approximated by a finite-dimensional dynamical system.

I. Introduction

As a LOGICAL consequence of the successful introduction of the Kalman–Bucy filter [1], [2], Rauch [3] and Meditch [4], respectively, presented what are now termed discrete-time and continuous-time optimum fixed-lag smoothers. The key idea distinguishing fixed-lag smoothing from filtering is that in fixed-lag smoothing, the estimate of the state or signal at time t is not available until some later time $t + \Delta$, where t is the running variable and Δ a fixed lag. The advantage of the smoothing approach is that measurements up to time $t + \Delta$ may be used to obtain the estimate, whereas in filtering measurements only up to time t may be used. The extra measurements lead to lower error variance in the estimate [5], [6].

Unfortunately, unlike the Kalman–Bucy filter where asymptotic stability is guaranteed by a few physically reasonable conditions [7], [8], the fixed-lag smoothers in the form presented by Rauch and Meditch are unstable exactly under the conditions guaranteeing the asymptotic stability of the filter [9]. This stability problem, however, is resolved in the discrete-time case in [10] and [11], where it is shown that by writing the filter and the fixed-lag smoother as a finite-dimensional linear system, the resulting fixed-lag smoother is stable. Recently, a stable continuous-time fixed-lag smoother has been obtained [12]. However, despite the fact that the impulse response of the optimum fixed-lag smoother is easily shown to be bounded-input and bounded-output (at least when the associated filter is asymptotically stable), the smoother of [12] operates only by imposing some limitations of the effects of internal instability. As with the smoother of Meditch, instability is present, although controlled. Further, the smoother of [12] is complex, involving switching and delay lines, in addition to the usual analog components required to simulate a finite-dimensional system.

Manuscript received October 5, 1972; revised May 14, 1973. This work was supported by the Australian Research Grants Committee.
S. Chirarattananon was with the Department of Electrical Engineering, University of Newcastle, Newcastle, N.S.W., Australia. He is now with the Department of Electrical Engineering, Konkaen University, Konkaen, Thailand.
B. D. O. Anderson is with the Department of Electrical Engineering, University of Newcastle, Newcastle, N.S.W., Australia.

If one turns to the earlier literature, one can find, as is well known, descriptions of fixed-lag smoothers (for stationary problems only) in the work of Wiener [13]. The optimal fixed-lag smoothers are defined from an input–output point of view by a transfer function, which, though perhaps rational for zero lag (i.e., normal filtering), is irrational for nonzero lag. Note that no instability problem arises directly with this transfer function; as remarked previously, the optimum fixed-lag smoother is bounded-input bounded-output stable. The instability which arises is Lyapunov instability, being an internal instability of the associated state-variable realizations.

To illustrate the point, let us anticipate the later material to remark that the various theories when applied to a stationary one-pole message process generate a need to realize a transfer function $(e^{-a\Delta} - e^{-s\Delta})(s - a)^{-1}$ for positive a. Now the corresponding impulse response is easily found, and is of compact support. The transfer function itself, though at first glance having a pole in Re $[s] > 0$, does not if one makes a pole-zero cancellation. All this suggests input–output stability. However, any attempt at a state-variable realization of the transfer function, permitting delay elements, apparently leads to instability (with ingenuity, the instability can actually be controlled [12]).

The purpose of this paper is to present two basic methods of continuous-time suboptimal fixed-lag smoothing, with a number of variations, together with constructive procedures and performance analyses. The main properties of these suboptimal smoothers are as follows.

i) Simplicity: The basic building blocks are finite-dimensional linear systems. In the majority of the suboptimum smoothers, pure delay elements are not needed (although in some they are permitted), whereas implicitly in the optimum smoother a delay element is always required [12], [14].

ii) Good Performance: The complexity of the suboptimal smoother (in terms of the dimensions of the basic building blocks) determines its performance. The higher the complexity, the closer the performance of the suboptimal smoother is to the optimum smoother. Although the design procedures do not of necessity lead to suboptimal smoothers which always perform better than the optimum filter, performance superior to that of the optimum filter can always be achieved by a sufficiently complex smoother. In the majority of situations only simple arrangements are necessary to obtain quite good performance.

iii) Stability: The suboptimal smoothers are all asymptotically stable under the conditions which guarantee asymptotic stability of the filters.

The outline of this paper is as follows. In Section II we review the optimum continuous-time smoother in the integral form of Kailath [14], together with a smoothing formula given earlier by the authors [18]. In Section III, we derive a universal relationship between the error covariance of suboptimal smoothers, the error covariance of the optimum smoother, and an integral measuring the difference between operators defining the suboptimal and optimal smoothers. In this way, we obtain a direct measure of approximation of the optimal smoother by a suboptimal smoother.

The first method of suboptimal smoothing together with variations is presented in Section IV, the approach having been suggested to us by Moore. The second method is introduced in Section V. Performance analyses and examples are included in each section. It is hoped that among the number of variations presented one can find a practical smoothing procedure suitable for every situation. Section VI unifies the treatments of suboptimal fixed-lag smoothing explored in Sections IV and V. It is shown that the problem of suboptimal fixed-lag smoothing can be formulated precisely as a constrained minimization problem, of which the methods of Sections IV and V are specialized solutions. This might suggest that Sections IV and V are superfluous. However, this is not so, because the method of Section VI is computationally quite infeasible in any but stationary problems (and even then may be very awkward to apply), while that of Section IV at least extends to time-varying problems (and can even be applied to nonlinear problems); that of Section V is also much simpler computationally than that of Section VI for stationary problems. We conclude our discussion in Section VII. A summary of important aspects of suboptimal smoothing is also presented in this section.

One might well ask what applications there are of smoothing. In the communications context they would appear to coincide roughly with the applications of filtering, and in practical terms could include optimum receiver design. In a stochastic control context, [15] suggests application to the control of aircraft utilizing terrain-following radar, and a separation theorem involving smoothing as opposed to filtering can be established.

As will be seen by examples subsequently, the improvement which smoothing can offer over filtering can be substantial (smoothing error covariances equal to one third of the corresponding filter error covariance, for example). It is a little unfortunate that such improvement is greatest in the high SNR situation [6] when it is perhaps least needed. Nonetheless, improvements in receiver design should be welcomed, provided they can be easily implemented. In the laboratory, we have found it straightforward to build smoothers with performance agreeing well with theoretical prediction.

II. Review of Optimum Filtering and Fixed-Lag Smoothing

We begin by considering the following system model:

$$\frac{dx(t)}{dt} = F(t)x(t) + G(t)u(t) \tag{1a}$$

where

$$z(t) = y(t) + w(t) \tag{1b}$$

$$y(t) = H'(t)x(t) \tag{1c}$$

and $u(\cdot)$ and $w(\cdot)$ are zero-mean white Gaussian processes with intensities $Q(\cdot)$ and $R(\cdot)$.

Unless explicitly stated to the contrary we shall assume that $u(\cdot)$ and $w(\cdot)$ are uncorrelated

$$E[u(t)w'(\tau)] = 0. \tag{2a}$$

However, when correlation between $u(\cdot)$ and $w(\cdot)$ is allowed, we shall assume it is of the form

$$E[u(t)w'(\tau)] = C(t)\,\delta(t-\tau). \tag{2b}$$

Furthermore, we also assume that $R(\cdot)$ is nonsingular, that (1) applies for all $t \geq t_0$, and that $x(t_0)$ is a Gaussian random variable of mean x_0 and covariance P_0 that is independent of $u(\cdot)$ and $w(\cdot)$. If F, G, H', Q, and R are independent of t and $t_0 = -\infty$ and if (1a) is asymptotically stable then (1) is a time-invariant stationary system and $x(\cdot)$, $y(\cdot)$, and $z(\cdot)$ are stationary processes.

We define the optimum filtered state estimate $\hat{x}_f(t \mid t)$ of $x(t)$ and the optimum fixed-lag smoothed state estimate $\hat{x}_s(t \mid t+\Delta)$ of $x(t)$ as

$$\hat{x}_f(t \mid t) = E[x(t) \mid z(\tau), \quad t_0 \leq \tau < t]$$

$$\hat{x}_s(t \mid t+\Delta) = E[x(t) \mid z(\tau), \quad t_0 \leq \tau < t+\Delta].$$

The error covariance of the filtered state estimate is denoted by $P(t \mid t)$ and of the fixed-lag smoothed estimate by $P_s(t \mid t+\Delta)$; these are, respectively, defined as

$$P(t \mid t) = E\{[x(t) - \hat{x}_f(t \mid t)][x(t) - \hat{x}_f(t \mid t)]'\}$$

$$P_s(t \mid t+\Delta) = E\{[x(t) - \hat{x}_s(t \mid t+\Delta)][x(t) - \hat{x}_s(t \mid t+\Delta)]'\}.$$

As is known, $\hat{x}_f(\cdot \mid \cdot)$ is given by

$$\frac{d\hat{x}_f(t \mid t)}{dt} = F(t)\hat{x}_f(t \mid t) + K(t)v(t), \qquad \hat{x}_f(t_0 \mid t_0) = x_0 \tag{3}$$

where $v(\cdot)$ is the innovations process [16] and is a zero-mean white Gaussian process with $E[v(t)v'(\tau)] = R(t)\,\delta(t-\tau)$, given as

$$v(t) = z(t) - H'(t)\hat{x}_f(t \mid t). \tag{4}$$

The quantity $K(t)$ is given by $K(t) = P(t \mid t)H(t)R^{-1}(t)$, where $P(t \mid t)$ is obtained from the matrix Riccati differential equation

$$\frac{dP(t \mid t)}{dt} = F(t)P(t \mid t) + P(t \mid t)F'(t) - K(t)R(t)K'(t) + G(t)Q(t)G'(t)$$

$$P(t_0 \mid t_0) = P_0. \tag{5}$$

In the case where (2b) applies, $K(t)$ is given as

$$K(t) = [P(t \mid t)H(t) + G(t)C(t)]R^{-1}(t).$$

Reference [17] gives for $\hat{x}_s(t \mid t+\Delta)$ and $P_s(t \mid t+\Delta)$

$$\hat{x}_s(t \mid t+\Delta) = \hat{x}_f(t \mid t) + P(t \mid t)\int_t^{t+\Delta} \phi_f'(\tau,t)H(\tau)R^{-1}(\tau)\nu(\tau)\,d\tau \tag{6a}$$

and

$$P_s(t \mid t+\Delta) = P(t \mid t) - P(t \mid t) \cdot \int_t^{t+\Delta} \phi_f'(\tau,t)H(\tau)R^{-1}(\tau)H'(\tau)\phi_f(\tau,t)\,d\tau P(t \mid t) \tag{7}$$

where $\phi_f(t,\tau)$ is the transition matrix associated with $F - KH'$, the system matrix of the filter.

Standard assumptions on the system (1) (see, e.g., [8]) guarantee exponential asymptotic stability of (3); the same assumptions then imply that $\phi_s(t,\tau) = \phi_f'(\tau,t)$ is the transition matrix of an unstable system matrix.

Equation (6a) may be written in the form

$$\hat{x}_s(t \mid t+\Delta) = \hat{x}_f(t \mid t) + \int_t^{t+\Delta} A(t,\tau)\nu(\tau)\,d\tau \tag{6b}$$

so that $\hat{x}_s(\cdot \mid \cdot + \Delta)$ can be regarded as the sum of a delayed filtered estimate and the output of a linear system driven by $\nu(\cdot)$.[1] If one examines the impulse-response $A(\cdot,\cdot)$ of the linear system, neglecting the integration limits in (6b), one can see that $A(\cdot,\cdot)$ appears to be unstable (in view of the instability of $\phi_s(t,\tau) = \phi_f'(\tau,t)$). The fact that the integration takes place over an interval of finite length Δ, however, means that the instability is in a sense truncated, and overall, the operator mapping $\nu(\cdot)$ into $\int_t^{t+\Delta} A(t,\tau)\nu(\tau)\,d\tau$ is bounded-input and bounded-output, at least if $P(\cdot \mid \cdot)$, $H(\cdot)$, and $R^{-1}(\cdot)$ are bounded and $\phi_f(\cdot,\cdot)$ is exponentially asymptotically stable. Nevertheless, obtaining a physical *internally stable* realization of (6a) is the core problem of smoother construction.

An alternative expression for $\hat{x}_s(t \mid t+\Delta)$ is given in [18] as

$$\hat{x}_s(t \mid t+\Delta) = P(t \mid t)\phi_f'(t+\Delta, t)P^{-1}(t+\Delta \mid t+\Delta)\hat{x}_f \cdot (t+\Delta \mid t+\Delta) + P(t \mid t) \cdot \int_t^{t+\Delta} \phi_f'(\tau,t)P^{-1}(\tau \mid \tau)G(\tau)Q(\tau)G'(\tau) \cdot P^{-1}(\tau \mid \tau)\hat{x}_f(\tau \mid \tau)\,d\tau. \tag{8}$$

The same comments made on (6) apply *mutatis mutandis* to (8). Note, however, that (8) does not contain a white noise process and intuitively, approximate realization (through replacement of the integral by a sum) of (8) would be preferable to (6). Note, however, that, in contrast to (6), (8) is not valid when $u(\cdot)$ and $w(\cdot)$ are correlated.

[1] Of course, the filtered estimate itself can be considered as the output of a linear system driven by $\nu(\cdot)$.

In passing, we note that the filtered signal estimate and the fixed-lag smoothed signal estimate of $y(t)$ are given, when $\hat{x}_f(t \mid t)$ and $\hat{x}_s(t \mid t+\Delta)$ are known, as

$$\hat{y}_f(t \mid t) = H'(t)\hat{x}_f(t \mid t)$$

and

$$\hat{y}_s(t \mid t+\Delta) = H'(t)\hat{x}_s(t \mid t+\Delta) \tag{9}$$

by the linearity of the estimates.

III. Properties of Linear Suboptimal Smoothers

In this section, our aim is to point out a useful relation that will give an insight into the qualitative behavior of linear suboptimal smoothers. In view of the equivalence between the measurements process $z(\cdot)$ and the innovations process $\nu(\cdot)$ [16], we write, in obvious notation

$$\hat{x}_s(t \mid t+\Delta) = \int_{t_0}^{t+\Delta} A_0(t,\Delta,\tau)\nu(\tau)\,d\tau \tag{10a}$$

for the linear optimum smoothed estimate, and suppose a linear suboptimal smoothed estimate is generated by some system or other driven by the innovations

$$\hat{x}_{sa}(t \mid t+\Delta) = \int_{t_0}^{t+\Delta} A_a(t,\Delta,\tau)\nu(\tau)\,d\tau. \tag{10b}$$

[In (10b), $\nu(\cdot)$ is still the true innovations process, because optimality of the filter is being retained.] Define $P_{sa}(t \mid t+\Delta)$ as the error covariance of the suboptimal estimate

$$P_{sa}(t \mid t+\Delta) = E\{[x(t) - \hat{x}_{sa}(t \mid t+\Delta)] \cdot [x(t) - \hat{x}_{sa}(t \mid t+\Delta)]'\}.$$

Then, writing $x(t) - x_{sa}(t \mid t+\Delta)$ as $\hat{x}(t) - \hat{x}_s(t \mid t+\Delta) + \hat{x}_s(t \mid t+\Delta) - \hat{x}_{sa}(t \mid t+\Delta)$ and noting by the projection theorem [16] that $x(t) - \hat{x}_s(t \mid t+\Delta)$ is independent of any linear functional of the measurements process (or equivalently, the innovations process), we obtain

$$P_{sa}(t \mid t+\Delta) = P_s(t \mid t+\Delta) + E\{[\hat{x}_s(t \mid t+\Delta) - \hat{x}_{sa}(t \mid t+\Delta)] \cdot [\hat{x}_s(t \mid t+\Delta) - \hat{x}_{sa}(t \mid t+\Delta)]'\}. \tag{11a}$$

Substituting (10a) and (10b), and using the fact that $\nu(\cdot)$ is a white process, we have

$$P_{sa}(t \mid t+\Delta) = P_s(t \mid t+\Delta) + \int_{t_0}^{t+\Delta} [A_0(t,\Delta,\tau) - A_a(t,\Delta,\tau)]R(t)[A_0(t,\Delta,\tau) - A_a(t,\Delta,\tau)']\,d\tau. \tag{11b}$$

Remarks

i) Equation (11b) shows that the performance of the suboptimal smoother is directly related to the closeness to which the suboptimal operator approximates the optimum operator.

ii) When A_0 and A_a are time invariant (both are assumed bounded in an integral-square sense) and $t_0 = -\infty$, $\nu(\cdot)$

is a stationary process, and $P_{sa}(\cdot \mid \cdot + \Delta)$ and $P_s(\cdot \mid \cdot + \Delta)$ are time invariant. A relation for P_{sa} and P_s equivalent to (11b) exists in the frequency domain and is of the form

$$P_{sa} = P_s + \frac{1}{2\pi j} \cdot \int_{-j\infty}^{j\infty} [H_0(s) - H_a(s)]R[H_0^*(s) - H_a^*(s)]\, ds \tag{11c}$$

where $H_0(s)$ and $H_a(s)$ are the transfer functions of the optimum and suboptimal operators; $H_0^*(s)$ and $H_a^*(s)$ are their corresponding Hermitian transposes. Equations (11b) and (11c) may serve as a qualitative guide to approximation of the optimum operator by suboptimal operators, giving rough information as to the sensitivities of the error at different values of the time index or frequency.

IV. Exact Filters for Approximate Models

We shall now derive several state and signal smoothing methods that belong to the same conceptual framework. We aim to develop a particular method in detail, presenting a performance analysis and two examples. We shall describe variations merely in outline form. The derivation works for nonstationary and stationary problems, and also provides a general method of attack on nonlinear problems, since it really shows how any smoothing problem can be replaced by a filtering problem.

A. Signal Smoothing: System Description and Smoother Definition

In addition to (1) we consider the following linear system

$$\dot{x}_a = F_a x_a + G_a y \tag{12a}$$

$$y_a = H_a' x_a + J_a y. \tag{12b}$$

Equations (12) represent a finite-dimensional linear system with F_a, G_a, H_a, and J_a chosen so that the system approximates[2] an ideal delay operator with a delay time Δ. For obvious reasons we restrict (12) to be asymptotically stable and for practical convenience F_a, G_a, H_a', and J_a are time invariant. Approximation of the delay operator by (12) means that

$$y_a(t + \Delta) \approx y(t).$$

Hence the optimum *filtered* estimate of $y_a(t + \Delta)$ is in fact an approximation of the *fixed-lag smoothed* estimate of $y(t)$, i.e.,

$$\begin{aligned}\hat{y}_a(t + \Delta) &= E[y_a(t + \Delta) \mid z(\tau), t_0 \leq \tau < t + \Delta] \\ &\approx \hat{y}(t \mid t + \Delta) \\ &= E[y(t) \mid z(\tau), t_0 \leq \tau < t + \Delta].\end{aligned}$$

[2] Scalar rational transfer functions approximating $e^{-s\Delta}$ can be found in many network synthesis books, see, e.g., [19]. The approximation normally works well for signals of restricted bandwidth, and the larger the delay-bandwidth product, the higher must be the degree of the rational transfer function if a good approximate delay is to be achieved. Of course, once a transfer function is known, the matrices $\{F_a, G_a, H_a, J_a\}$ of a state-variable realization of that transfer function can be obtained by standard means. See the examples for typical approximating transfer functions.

Let us see how the estimate on the left can be computed. To begin, let us combine (1) and (12) in a compact form

$$\begin{bmatrix}\dot{x} \\ \dot{x}_a\end{bmatrix} = \begin{bmatrix}F & 0 \\ G_a H' & F_a\end{bmatrix}\begin{bmatrix}x \\ x_a\end{bmatrix} + \begin{bmatrix}G \\ 0\end{bmatrix}u \tag{13a}$$

$$y_a = [J_a H' \quad H_a'] \begin{bmatrix}x \\ x_a\end{bmatrix} \tag{13b}$$

$$z = [H' \quad 0]\begin{bmatrix}x \\ x_a\end{bmatrix} + w. \tag{13c}$$

We note that $z(\cdot)$ remains the actual measurements process. The time variable is dropped from (12) and (13) for convenience.

We denote by $\hat{x}_a(t + \Delta)$ the optimum filtered estimate of $x_a(t + \Delta)$ and by $\hat{y}_a(t + \Delta)$ the optimum filtered estimate of $y_a(t + \Delta)$. Applying the filter algorithm (3) to (13a) and (13c) we have

$$\begin{bmatrix}\dot{\hat{x}}_f \\ \dot{\hat{x}}_a\end{bmatrix} = \begin{bmatrix}F & 0 \\ G_a H' & F_a\end{bmatrix}\begin{bmatrix}\hat{x}_f \\ \hat{x}_a\end{bmatrix} + \begin{bmatrix}K \\ K_a\end{bmatrix}v \tag{14a}$$

$$\hat{y}_a = [J_a H' \quad H_a'] \begin{bmatrix}\hat{x}_f \\ \hat{x}_a\end{bmatrix} \tag{14b}$$

where

$$\begin{bmatrix}K \\ K_a\end{bmatrix} = \begin{bmatrix}P & P_b' \\ P_b & P_a\end{bmatrix}\begin{bmatrix}H \\ 0\end{bmatrix}R^{-1} = \begin{bmatrix}PHR^{-1} \\ P_b H R^{-1}\end{bmatrix} \tag{14c}$$

and

$$\begin{bmatrix}P & P_b' \\ P_b & P_a\end{bmatrix}$$

is the solution of the matrix Riccati differential equation

$$\begin{aligned}\begin{bmatrix}\dot{P} & \dot{P}_b' \\ \dot{P}_b & \dot{P}_a\end{bmatrix} &= \begin{bmatrix}F & 0 \\ G_a H & F_a\end{bmatrix}\begin{bmatrix}P & P_b' \\ P_b & P_a\end{bmatrix} + \begin{bmatrix}P & P_b' \\ P_b & P_a\end{bmatrix}\begin{bmatrix}F' & HG_a' \\ 0 & F_a'\end{bmatrix} \\ &\quad - \begin{bmatrix}K \\ K_a\end{bmatrix} R [K \quad K_a'] + \begin{bmatrix}GQG' & 0 \\ 0 & 0\end{bmatrix}.\end{aligned} \tag{15}$$

Before proceeding we make the following remarks.

i) The model described by (12) has no intrinsic physical significance, except for the basic system (1). The augmented system (13) is merely a technical artifice used to solve the smoothing problem for the originally given system. Consequently, it is natural to assume zero values for the initial conditions of x_a and $\hat{x}_a$. In fact, since the system (12) simulates an ideal delay operator with a delay of Δ, what the system (14) and (15) produces before $t_0 + \Delta$ is irrelevant and understandably the initial condition for $\hat{x}_a$ should be zero to minimize the initial transient term in $\hat{x}_a$. Similar arguments can be used to give $P_b(t_0) = P_a(t_0) = 0$. (Of course, if $t_0 = -\infty$, these concerns disappear.) In contrast, the quantity P in (15) is the error covariance of the filtered estimate $\hat{x}_f$ of x, and the comments made in Section II apply.

ii) Only the two quantities P and P_b are needed in the construction of the system described by (14). Equations for P and P_b alone follow from (15) as

$$\dot{P} = FP + PF' - KRK' + GQG', \qquad P(t_0) = P_0 \tag{5}$$

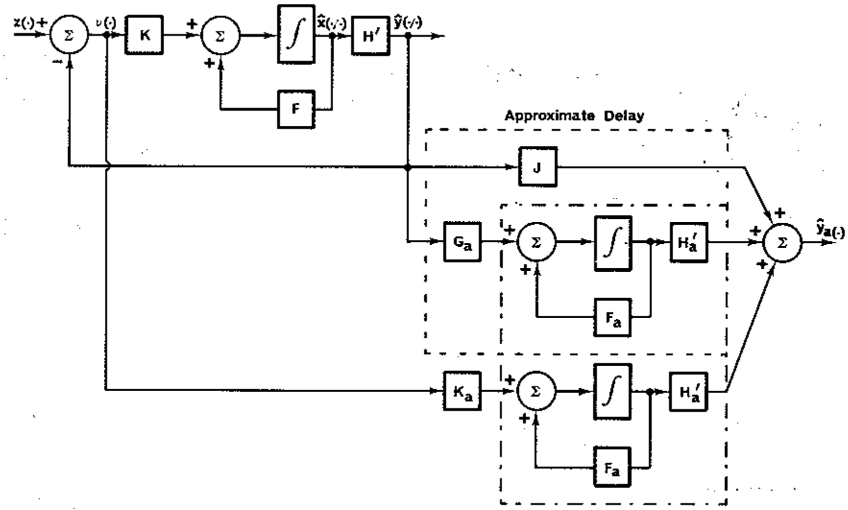

Fig. 1. Signal smoothing system of 4(a); note two sections enclosed in light dotted line are physically same system shown separated here for illustration purposes.

(with $K = PHR^{-1}$) and

$$\dot{P}_b = P_bF' + F_aP_b' - K_aRK' + G_aH'P, \qquad P_b(t_0) = 0 \quad (16)$$

(with $K_a = P_bHR^{-1}$). If the system (1) is nonstationary, the quantity P obtained as a solution of (5) is time varying and so is P_b, from (16), irrespective of the time invariance of (12). (Both matrices can most easily be obtained simultaneously.) This leads to a time-varying nonstationary smoother (14). Of course, when (1) is stationary and (12) is time invariant, we obtain a stationary smoother. Then it is easier probably to find P first, and then solve (16), with $\dot{P}_b$ replaced by zero.

iii) In the case where there is correlation between $u(\cdot)$ and $w(\cdot)$ in (1), i.e., case (2b) applies, (5), (14), (15), and (16) are still valid provided that we replace $K = PHR^{-1}$ by $K = (PH + GC)R^{-1}$ with K_a unaltered.

iv) The closed-loop system matrix of (14), when z is considered the input to the system, is

$$\begin{bmatrix} F - KH' & 0 \\ (G_a - K_a)H' & F_a \end{bmatrix}.$$

It is immediately clear that this matrix is asymptotically stable, as standard assumptions made on (1) guarantee asymptotic stability of $F - KH'$ [8], and our restriction imposes the stability requirement on F_a.

v) Of significance is the question of convergence of $\hat{y}_a(t + \Delta)$ to $\hat{y}_s(t \mid t + \Delta)$ as the delay approximation is improved. When one builds the smoothing system (14) based on the delay approximation system (12) one wants to know whether by increasing the complexity of the system (12), thus improving its approximation of the delay operator, the resulting estimate $\hat{y}_a$ will correspondingly produce a better approximation to the optimum fixed-lag smoothed estimate $\hat{y}_s$. The answer to this is "yes" as can be shown. Obviously, the better the approximation, the more complex the smoother.

vi) Fig. 1 shows the structure of the suboptimal smoother. As implied by (14) and as is obvious from the figure, $\hat{y}_a$ consists of two components. One component is $\hat{y}_f$, approximately delayed by Δ. Reference to (6) and (9) implies that the other component is an approximation of the second term of $\hat{y}_s(t \mid t + \Delta)$ [see (6a)].

This suggests one variation of the suboptimal smoother, achieved by replacing the approximate delay system in Fig. 1 by a pure delay element. However, this ad hoc change need not necessarily result in an improved smoother.

B. Signal Smoothing: Error Covariance

We define the error covariance of the suboptimal smoothed estimate as

$$P_{sa}(t \mid t + \Delta) = E\{[y(t) - \hat{y}_a(t + \Delta)][\hat{y}(t) - \hat{y}_a(t + \Delta)]'\}.$$

Since y is $H'x$ and $\hat{y}_a$ is $J_aH'\hat{x}_f + H_a'\hat{x}_a$, $P_{sa}(t \mid t + \Delta)$ can be obtained from the state covariance matrices of the following augmented system constructed from (1a) and (14a):

$$\begin{bmatrix} \dot{x} \\ \dot{\hat{x}}_f \\ \dot{\hat{x}}_a \end{bmatrix} = \begin{bmatrix} F & 0 & 0 \\ KH' & F - KH' & 0 \\ K_aH' & (G_a - K_a)H' & F_a \end{bmatrix} \begin{bmatrix} x \\ \hat{x}_f \\ \hat{x}_a \end{bmatrix} + \begin{bmatrix} G & 0 \\ 0 & K \\ 0 & K_a \end{bmatrix} \begin{bmatrix} u \\ w \end{bmatrix} \quad (17)$$

or

$$\dot{\boldsymbol{X}} = \boldsymbol{FX} + \boldsymbol{GU} \quad (18)$$

in obvious notation. Define $\boldsymbol{P}(t,\tau) = E[x(t)x'(\tau)]$, and then $\boldsymbol{P}(t,t)$ can be calculated from

$$\dot{\boldsymbol{P}}(t,t) = \boldsymbol{FP}(t,t) + \boldsymbol{P}(t,t)\boldsymbol{F}' + \boldsymbol{GQG}', \qquad \boldsymbol{P}(t_0,t_0) = \boldsymbol{P}_0 \quad (19)$$

where

$$\boldsymbol{P}_0 = \begin{bmatrix} P_0 & P_0 & 0 \\ P_0 & P_0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \qquad \boldsymbol{GQG}' = \begin{bmatrix} GQG' & 0 & 0 \\ 0 & KRK' & KRK_a' \\ 0 & K_aRK' & K_aRK_a' \end{bmatrix}$$

[or a modified $\boldsymbol{GQG}'$ in case condition (2b) applies]. As is known, $\boldsymbol{P}(t,\tau)$ can be obtained from

$$\boldsymbol{P}(t,\tau) = \psi(t,\tau)\boldsymbol{P}(t,t)1(t - \tau) + \boldsymbol{P}(t,t)\psi'(\tau,t)1(\tau - t) \quad (20$$

TABLE I
SIGNAL SMOOTHING: EXAMPLE I

Quantity	Value	Comments
F	$\begin{bmatrix} 0 & 1 \\ -2 & -2 \end{bmatrix}$	
G	$\begin{bmatrix} 0 \\ 1 \end{bmatrix}$	Single-input single-output system. Transfer function is $\frac{1}{s^2+2s+2}$.
H'	$[1 \quad 0]$	
Q	1000	
R	1	
P	$\begin{bmatrix} 5.961 & 17.76 \\ 17.76 & 153.3 \end{bmatrix}$	Filter error covariance.
$F - KH'$	$\begin{bmatrix} -5.961 & 1 \\ -19.76 & -2 \end{bmatrix}$	Closed-loop filter matrix.
$\alpha \pm j\beta$	$-3.98 \pm j3.98$	Roots of $F - KH'$.
Δ	0.5	Approximately $2/\alpha$.
P_s	$\begin{bmatrix} 2.014 & 0.0007 \\ 0.0007 & 63.18 \end{bmatrix}$	Optimum fixed-lag smoothed error covariance ($\Delta = 0.5$).
F_a	$\begin{bmatrix} 0 & 1 \\ -48 & -12 \end{bmatrix}$	
G_a	$\begin{bmatrix} 0 \\ 1 \end{bmatrix}$	Second-order bandpass approximation to delay. Design delay $\Delta = 0.5$. Transfer function is $\frac{s^2-12s+48}{s^2+12s+48}$.
H_a'	$[0 \quad -24]$	
J_a	1	
P_y	5.961	Filtered signal error covariance.
P_{sy}	2.014	Optimum fixed-lag smoothed signal error covariance.
P_{say}	2.082	Suboptimal fixed-lag smoothed signal error covariance.

where $\psi(t,\tau)$ is the transition matrix associated with F, and $1(\cdot)$ is the Heaviside unit step function. The quantities required to compute $P_{sa}(t \mid t + \Delta)$ are given by the submatrices of $P(t,t)$, $P(t + \Delta, t + \Delta)$, and $P(t + \Delta, t)$.

We remark that in case (17) is a stationary system (when (1) and (14) are stationary), (18) and (19) are also stationary. Furthermore we have $P(t,t) = P(t + \Delta, t + \Delta) = P$ and $P(t + \Delta, t) = \psi(\Delta)P$. In this case only P and $\psi(\Delta)$ need be calculated.

C. State Smoothing by Delay of States

Instead of approximately delaying the signal in (12) we consider an approximate delay of the states, as follows:

$$\dot{x}_a = F_a x_a + G_a x \tag{21a}$$

$$y_a = H_a' x_a + J_a x. \tag{21b}$$

Here $y_a(t + \Delta)$ is an approximate version of $x(t)$; straightforward modification of (14) provides the suboptimal fixed-lag smoothed state estimate. The procedures and remarks made in a) and b) apply *mutatis mutandis* to this case.

D. State Smoothing by Delay of Inputs

We consider the arrangement of Fig. 2. Here, we postulate that the input $u(\cdot)$ to the basic system also enters a system approximating a delay Δ; the delayed version of $u(\cdot)$ called $u_d(\cdot)$ then drives a subsystem identical to the basic system, except for time translation. Thus referring to the figure one has $F_d(t + \Delta) = F(t)$, $G_d(t + \Delta) = G(t)$, and $H_d'(t + \Delta) = H'(t)$. The output $y_d(\cdot)$ of the subsystem is approximately a delayed version of $y(\cdot)$, at least if F is exponentially asymptotically stable. Thus we obtain

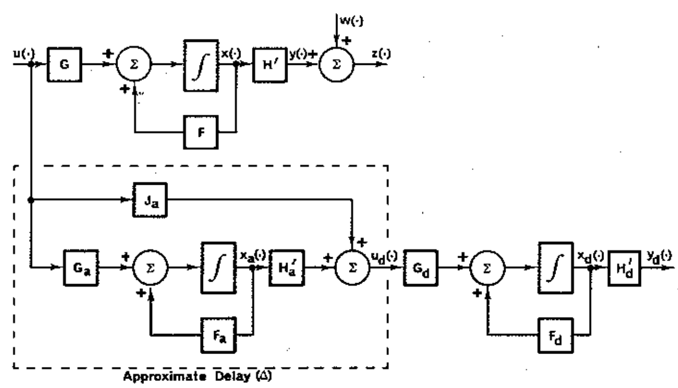

Fig. 2. System model for 4(d).

$$\begin{bmatrix} \dot{x} \\ \dot{x}_a \\ \dot{x}_d \end{bmatrix} = \begin{bmatrix} F & 0 & 0 \\ 0 & F_a & 0 \\ 0 & G_d H_a' & F_d \end{bmatrix} \begin{bmatrix} x \\ x_a \\ x_d \end{bmatrix} + \begin{bmatrix} G \\ G_a \\ J_a \end{bmatrix} u \tag{22a}$$

$$y_d = [0 \quad 0 \quad H_d'] \begin{bmatrix} x \\ x_a \\ x_d \end{bmatrix} \tag{22b}$$

$$z = [H' \quad 0 \quad 0] \begin{bmatrix} x \\ x_a \\ x_d \end{bmatrix} + w. \tag{22c}$$

TABLE II
SIGNAL SMOOTHING: EXAMPLE II

Quantity	Value	Comments
F	$\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -1 & -3 & -3 \end{bmatrix}$	
G	$\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$	Single-input single-output system. Transfer function $\frac{1}{(s+1)^3}$.
H'	$[1 \ \ 0 \ \ 0]$	
Q	1000	
R	1	
P	$\begin{bmatrix} 3.645 & 6.643 & -0.227 \\ 6.643 & 24.44 & 22.07 \\ -0.227 & 22.07 & 144.7 \end{bmatrix}$	Filter error covariance.
$F - KH'$	$\begin{bmatrix} -3.645 & 1 & 0 \\ -6.643 & 0 & 1 \\ -0.773 & -3 & -3 \end{bmatrix}$	Closed-loop filter matrix.
$\lambda, \alpha \pm j\beta$	$-3.2, -1.72 \pm j2.15$	Approximate roots of $F - KH'$
Δ	1	Approximately twice the dominant time constant ($2/1.72 = 1.16$).
P_s	$\begin{bmatrix} 1.038 & 0.0372 & -5.032 \\ 0.0372 & 4.903 & 0.1121 \\ -5.032 & 0.1121 & 101.1 \end{bmatrix}$	Optimum fixed-lag smoothed error covariance ($\Delta = 1$).
F_a	$\begin{bmatrix} 0 & 1 \\ -12 & -6 \end{bmatrix}$	
G_a	$\begin{bmatrix} 0 \\ 1 \end{bmatrix}$	Second-order bandpass approximation to delay. Design delay $\Delta = 1$. Transfer function is $\frac{s^2 - 6s + 12}{s^2 + 6s + 12}$.
H_a'	$[0 \ \ -12]$	
J_a	1	
P_y	3.645	Filtered signal error covariance.
P_{sy}	1.038	Optimum fixed-lag smoothed signal error covariance.
P_{say}	1.048	Suboptimal fixed-lag smoothed signal error covariance.

We then build the filter for (22) in the following way. Define $\hat{x}_a$ as the filtered estimate of x_a, $\hat{x}_d$ as the filtered estimate of x_d, and as usual, $\hat{x}_f$ as the filtered estimate of x; then write (22) in the more convenient form

$$\dot{X} = FX + GU$$
$$y_d = H_d'X$$
$$z = H'X + w.$$

Applying the filter equation (3) to this we have

$$\dot{\hat{X}} = F\hat{X} + K\eta \tag{23a}$$
$$\hat{y}_d = H_d'\hat{X} \tag{23b}$$
$$\hat{y}_f = H'\hat{X} \tag{23c}$$

where $\hat{X}$ is the filtered estimate of X, $\hat{y}_d$ is the filtered estimate of y_d, and $\hat{y}_f$ is the filtered estimate of y. Since $y_d(t + \Delta) \approx y(t)$, it follows that $\hat{y}_d(t + \Delta)$ is an approximation to $\hat{y}_s(t \mid t + \Delta)$.

The quantity K is obtained from the following relations:

$$K = \begin{bmatrix} K \\ K_a \\ K_d \end{bmatrix} = PH'R^{-1}$$
$$= \begin{bmatrix} P & P_b' & P_c' \\ P_b & P_a & P_e' \\ P_c & P_e & P_d \end{bmatrix} \begin{bmatrix} H \\ 0 \\ 0 \end{bmatrix} R^{-1} = \begin{bmatrix} PHR^{-1} \\ P_bHR^{-1} \\ P_cHR^{-1} \end{bmatrix}. \tag{23d}$$

The matrix P is obtained from the matrix Riccati differential equation

$$\dot{P} = FP + PF' - KRK' + GQG'. \tag{23e}$$

The following points should be noted.

i) Initial conditions are treated much as in Section IV-A. Note that in any case values of the initial conditions will not greatly affect the performance of the smoother when t is large enough, say several times the dominant time constants of F_a and F_d, as their effect will gradually be forgotten. In case where (23) is a stationary system, i.e., when it is time invariant and asymptotically stable, $t_0 = -\infty$ and v is stationary, the initial conditions will be entirely forgotten.

ii) Only the matrices P, P_b, and P_c need be computed for the construction of the smoother. Note that P is obtained from the original matrix Riccati equation, while P_b is obtained from

$$\dot{P}_b = F_aP_b + P_bF' - K_bRK' + G_aQG', \qquad P_b(t_0) = 0 \tag{24a}$$

and P_c is obtained from

$$\dot{P}_c = F_dP_c + P_cF' + G_dH_a'P_b - K_cRK' + J_aQG',$$
$$P_c(t_0 + \Delta) = P_0, \qquad t \geq t_0 + \Delta. \tag{24b}$$

In the nonstationary case, simultaneous solution for P, P_b, and P_c is to be preferred, and in the stationary case, sequential solution.

iii) When there is correlation between $u(\cdot)$ and $w(\cdot)$, i.e., when case (2b) applies, we modify (23d) to

$$K = \begin{bmatrix} K \\ K_a \\ K_d \end{bmatrix} = \begin{bmatrix} (PH + GC)R^{-1} \\ (P_bH + G_aC)R^{-1} \\ (P_cH + J_aC)R^{-1} \end{bmatrix}.$$

iv) We now ask whether improving the approximation of the system $[F_a,G_a,H_a',J_a]$ to the ideal delay operator will give a better suboptimal smoothed estimate. It is intuitively reasonable that the answer is "yes" with one reservation. This can be shown straightforwardly, with the reservation concerning the value $\hat{x}_d(t_0 + \Delta)$. It can be shown that if $\hat{x}_d(t_0 + \Delta) = x_0$, then the claim is true. Of course in the case where (23) is stationary, we need not worry about the initial conditions.

v) The closed-loop system matrix of (23) when z is considered the input is

$$\begin{bmatrix} F - KH' & 0 & 0 \\ -K_aH' & F_a & 0 \\ -K_dH' & G_dH_a' & F_d \end{bmatrix}.$$

Straightforwardly, one sees that if $F - KH'$, F_a, and F_d are individually asymptotically stable, we have also an asymptotically stable closed-loop system matrix, and hence the smoother is asymptotically stable.

vi) To obtain the error covariance of $\hat{x}_d(t + \Delta)$, we form an augmented system as in Section IV-B. This is

$$\begin{bmatrix} \dot{x} \\ \dot{\hat{x}}_f \\ \dot{\hat{x}}_a \\ \dot{\hat{x}}_d \end{bmatrix} = \begin{bmatrix} F & 0 & 0 & 0 \\ KH' & F - KH' & 0 & 0 \\ K_aH' & -K_aH' & F_a & 0 \\ K_dH' & -K_dH' & G_dH_a' & F_d \end{bmatrix} \begin{bmatrix} x \\ \hat{x}_f \\ \hat{x}_a \\ \hat{x}_d \end{bmatrix} + \begin{bmatrix} G & 0 \\ 0 & K \\ 0 & K_a \\ 0 & K_d \end{bmatrix} \begin{bmatrix} u \\ w \end{bmatrix}.$$

By a procedure similar to that used in Section IV-B, we can then obtain the individual quantities for the error covariance $E\{[x(t) - \hat{x}_d(t \mid t + \Delta)][x(t) - \hat{x}_d(t \mid t + \Delta)]'\}$.

For examples of signal smoothing, see Tables I and II.

E. Summary

We have in effect presented three different practical suboptimal smoothers in this section, all of which are linked within the same conceptual framework of building an exact filter for the original system and an associated approximate delay model. Yet the resulting suboptimal smoothers may differ a great deal depending on the structure of the system model. Out of convenience we have represented the approximate delay of the appropriate (vector) quantity by the system $[F_a,G_a,H_a',J_a]$; in practice, this approximate delay would probably be obtained using a number of decoupled single-input single-output delay systems each of which approximately delayed one scalar quantity, because most is known about finite-dimensional systems which approximately delay scalar quantities. Thus an n-dimensional state in the system model (1) would imply use of n individual delay scalar systems. In this context the method of Section IV-A presents an attractive solution in signal smoothing in the case where dimension of the signal vector is much less than the dimension of the state vector. In case state smoothing is required, the method of Section IV-D may be more attractive than the method of Section IV-C as the dimension of the input vector is usually less than the dimension of the state vector. However, since the input variables in practice have a spectrum of much wider bandwidth than the state variables, more accurate approximation for the delay operator is needed. The common features of the smoothers discussed are as follows.

1) They are all finite dimensional without any pure delay element. The dimension of the suboptimal smoothers depends on the particular method used for design, and on the dimension of the finite-dimensional system realizing the approximate delay. In turn, the latter dimension depends on the degree to which it is desired that the suboptimal performance approach the optimal performance. Note that, although it is not necessary that these smoothers will perform better than the Kalman–Bucy filter, they can be made to do so by increasing the accuracy of approximation of the ideal delay operator. A possible paradox arises here; by increasing the dimension of the subsystem approximately realizing the delay, one tends in the limit to the optimal smoother. The procedure here suggests stability and the earlier remarks suggest instability for this smoother. However, the instability was of certain realizations of this smoother, which are *not* obtained in the limiting operation.

2) The design delay value of Δ affects the approximate delay system dimension in the sense that the larger the delay the higher the dimension of the approximate delay system needed to produce a given accuracy in approximating a pure delay element. Greater Δ lead to lower optimum smoothing error covariances, but essentially all the improvement that smoothing offers over filtering can be obtained by taking Δ equal to several times the dominant filter time constant [5].

3) The smoothers are all stable under the usual conditions guaranteeing filter stability.

4) The smoothers can be defined irrespective of whether there is a correlation between the input and output noise processes, i.e., whether condition (2a) or (2b) applies.

V. Approximate Smoothers for Exact Models

In this section we present several suboptimal smoothing methods linked by a conceptual framework distinct from that of Section IV. In practice, *these methods are applicable only to time-invariant stationary systems* (1). The general theme is to approximate the operators occurring in (6) or (8). In this context it is noted that the methods we are about to present are not exhaustive.

A. General Approximation of (6) and (8)

We first write down the time-invariant stationary version of (6) and (8) for easy reference

$$\hat{x}_s(t \mid t + \Delta) = \hat{x}_f(t \mid t) + P \cdot \int_0^\infty [\phi_f'(\Delta - \tau)1(\Delta - \tau)]HR^{-1}\nu(t + \Delta - \tau)\, d\tau \tag{6c}$$

and

$$\hat{x}_s(t \mid t+\Delta) = P\phi_f'(\Delta)P^{-1}\hat{x}_f(t+\Delta \mid t+\Delta) + P \cdot \int_0^\infty [\phi_f'(\Delta-\tau)1(\Delta-\tau)]P^{-1}GQG'P^{-1}\hat{x}_f \cdot (t+\Delta-\tau \mid t+\Delta-\tau)\, d\tau \tag{8b}$$

where P is the stationary filter error covariance and $1(\cdot)$ is the Heaviside unit step function.

Our objective here is to approximate the impulse response

$$A(\Delta - t) = \phi_f'(\Delta - t)1(\Delta - t), \qquad 0 \leq t < \infty \tag{29}$$

which has a Laplace transform

$$L(s) = [sI + (F - KH')']^{-1}[e^{(F-KH')'\Delta} - e^{-s\Delta}]. \tag{30}$$

From (29) and (30) those familiar with approximation theory may already visualize several methods of approximation. Notice, however, that although the operator is well defined, it is generally a matrix and this complicates the task of approximation.

In this context it is obvious that when any method of approximation is applicable to (6c), it is also applicable to (8b), at least when the input and output noise processes are not correlated.

B. Specific Approximation of (6) and (8)

The straightforward way to find a realizable operator approximating the operator A in (29) or a transform matrix approximating L in (30) is to find a single-input single-output asymptotically stable dynamical system approximating each element of A (or equivalently of L). Hence, for an n-dimensional state, the number of these single-input single-output systems is n^2. To reduce the number of these individual dynamical systems we are thus forced to reduce the number of nonzero elements of A (or equivalently of L). This can be done by finding a similarity transformation matrix which, e.g., partially diagonalizes A.

To illustrate this point, we describe a particular procedure in detail. We assume that $(F - KH')$ has distinct characteristic roots; the roots are either real or occur in complex conjugate pairs. This assumption is not particularly restrictive, as in physical systems this more often than not is the case. We partially diagonalize $(F - KH')$ in the following manner:

$$(F - KH') = T\Lambda T^{-1}$$

and

$$\Lambda = \begin{bmatrix} \lambda_1 & & & & & \\ & \ddots & & & 0 & \\ & & \lambda_k & & & \\ & & & \begin{bmatrix} \alpha_1 & -\beta_1 \\ \beta_1 & \alpha_1 \end{bmatrix} & & \\ & & & & \ddots & \\ & 0 & & & & \begin{bmatrix} \alpha_e & -\beta_e \\ \beta_e & \alpha_e \end{bmatrix} \end{bmatrix} \tag{31}$$

where T is a nonsingular matrix, and where λ_i, $i = 1,\cdots,k$ are the real roots, $\alpha_i + j\beta_i$, $\alpha_i - j\beta_i$, $i = 1,\cdots,e$ are the complex conjugate roots. Since $F - KH'$ is the matrix associated with the filter (3) we see that λ_i and α_j are all negative real. In view of (31), the operator A takes the form

$$A(\Delta - t) = (T')^{-1} \exp[\Lambda'(\Delta - t)]T'1(\Delta - t), \qquad 0 \leq t < \infty. \tag{32}$$

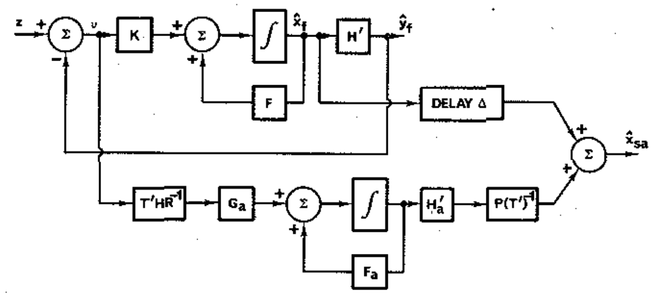

Fig. 3. Suboptimal smoother derived by approximating exact smoother.

From here we proceed to approximate each element of $e^{\Lambda'(\Delta-t)}1(\Delta - t)$, $0 \leq t < \infty$, by an asymptotically stable finite-dimensional linear system. Equation (11) provides a broad guide as to the sensitivity of the error of approximation.

Once the impulse response matrix approximating $e^{\Lambda'(\Delta-t)}1(\Delta - t)$, $0 \leq t < \infty$, is obtained we then obtain an augmented linear system description in terms of the matrices F_a, G_a, H_a' (and J_a). Fig. 3 shows the form to be expected when starting with (6c). (Note the presence of the delay element, which could be replaced by an approximating finite-dimensional block.)

To obtain the error covariance when either (6c) or (8b) is used, one follows the method of Section IV-B, *mutatis mutandis*.

Note that this diagonalization procedure reduces the problem of approximating the impulse response of an *arbitrary* smoother to the problem of approximating one of two *standard* forms, obtained with real and complex λ_i, respectively. The forms in effect contain one and two free parameters, respectively, so that a one-parameter family and a two-parameter family of approximations will yield approximations for all smoothers.

For further examples of signal smoothing, see Tables III and IV.

VI. Unified Approach to Suboptimal Fixed-Lag Smoothing

In this section we reformulate the problem of suboptimal fixed-lag smoothing in a broad framework and show that the solutions obtained in Sections V and VI are only specialized solutions of this more general problem. Define $\hat{x}_{sa}(t \mid t+\Delta)$ as the suboptimal fixed-lag smoothed estimate of $x(t)$ given measurements up to $t + \Delta$ and let

$$\hat{x}_{sa}(t \mid t+\Delta) = J_{a1}\hat{x}_f(t \mid t) + J_{a2}\hat{x}_f(t+\Delta \mid t+\Delta) + H_a'\hat{x}_a(t+\Delta) \tag{33}$$

where J_{a1}, J_{a2}, and H_a' are matrices of appropriate dimensions, $\hat{x}_a(t+\Delta)$ is the output of a finite-dimensional linear dynamical system for the moment arbitrary with input

TABLE III
SIGNAL SMOOTHING: EXAMPLE III—SUBOPTIMAL FIXED-LAG SMOOTHER CONSTRUCTED ON THE BASIS OF EQUATION (6c)

Quantity	Value	Comments
F	$\begin{bmatrix} 0 & 1 \\ -2 & -2 \end{bmatrix}$	
G	$\begin{bmatrix} 1 \\ 0 \end{bmatrix}$	Same system as that of Example I, Section IV.
H'	$[1 \quad 0]$	
Q	1000	
R	1	
P	$\begin{bmatrix} 5.961 & 17.76 \\ 17.76 & 153.3 \end{bmatrix}$	Filter error covariance.
$F - KH'$	$\begin{bmatrix} -5.961 & 2 \\ -19.76 & -2 \end{bmatrix}$	Closed-loop filter matrix.
Δ	0.5	Design delay (or lag).
P_s	$\begin{bmatrix} 2.014 & 0.0007 \\ 0.0007 & 63.18 \end{bmatrix}$	Optimum fixed-lag smoothed error covariance ($\Delta = 0.5$).
F_a	$\begin{bmatrix} -2.332 & -4.377 & 0 & 0 \\ 4.377 & -2.332 & 0 & 0 \\ 0 & 0 & -2.332 & -4.377 \\ 0 & 0 & 4.377 & -2.332 \end{bmatrix}$	
G_a	$\begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix}$	See Fig. 3 for illustration of F_a, G_a, and H_a'.
H_a'	$\begin{bmatrix} -0.6063 & 0.2515 & -0.2243 & 0.3004 \\ 0.2243 & -0.3004 & -0.6063 & 0.2515 \end{bmatrix}$	
P_{sa}	$\begin{bmatrix} 3.537 & 5.111 \\ 5.111 & 81.92 \end{bmatrix}$	Suboptimal fixed-lag smoothed error covariance.

comprising the innovations process v (passing through a gain block K_a), and the filtered estimate $\hat{x}_f$ (passing through a gain block G_a). The arrangement of (33) effectively allows only elements of a finite-dimensional linear dynamical system and one pure delay element. Together with (1) we write the dynamical equation for $\hat{x}_f$ and $\hat{x}_a$ as

$$\begin{bmatrix} \dot{x} \\ \dot{\hat{x}}_f \\ \dot{\hat{x}}_a \end{bmatrix} = \begin{bmatrix} F & 0 & 0 \\ KH' & F - KH' & 0 \\ K_aH' & (G_a - K_a)H' & F_a \end{bmatrix} \begin{bmatrix} x \\ \hat{x}_f \\ \hat{x}_a \end{bmatrix} + \begin{bmatrix} G & 0 \\ 0 & K \\ 0 & K_a \end{bmatrix} \begin{bmatrix} u \\ w \end{bmatrix}. \quad (34)$$

We also write, in obvious notation

$$\dot{\hat{X}} = F\hat{X} + GV \quad (35a)$$

$$x(t) = [I \quad 0 \quad 0]\hat{X}(t) \quad (35b)$$

$$\hat{x}_{sa}(t \mid t + \Delta) = [0 \quad J_{a1} \quad 0]\hat{X}(t) + [0 \quad J_{a2} \quad H_a']\hat{X}(t + \Delta). \quad (35c)$$

We define the error covariance of the suboptimal fixed-lag smoothed estimate $P_{sa}(t \mid t + \Delta)$ as

$$P_{sa}(t \mid t + \Delta) = E\{[x(t) - \hat{x}_{sa}(t \mid t + \Delta)] \cdot [x(t) - \hat{x}_{sa}(t \mid t + \Delta)]'\}. \quad (36)$$

This quantity can be evaluated using the quantities defined in (34) and (35) by standard procedures.

Our objective is to minimize the mean-square error of $\hat{x}_{sa}(t \mid t + \Delta)$. We therefore formulate the problem as follows.

Given F, G, H', Q, R, and K, and the dimension of F_a, find F_a, G_a, K_a, H_a', J_{a1}, and J_{a2} such that 1) F_a is an asymptotically stable system matrix; and 2) the trace of the matrix P_{sa} in (36) is minimized.

Remarks

i) The initial conditions are treated much the same as in earlier situations.

ii) If F_a, G_a, and K_a are specified *a priori*, the minimization problem is a quadratic one without constraints, and the solution is simply found. However, the general problem is not simple to solve. In view of this difficulty, it is fair to assert that general minimization is not practical in the situation where $\hat{x}$ is nonstationary and the related quantities, e.g., P_{sa}, are not time invariant.

iii) If we let $J_{a1} = 0$ or $J_{a2} = 0$ or both, a form for $\hat{x}_{sa}$ identical to those of Sections IV and V is obtained, except for the exact value of the unknowns. One can therefore take the corresponding solution in those sections as a partial solution to our problem here. (Note, however, that the solutions of Sections IV and V do not necessarily yield minimum tr (P_{sa}).)

iv) It is intuitively reasonable that as the dimension of F_a increases the estimate $\hat{x}_{sa}$ approaches the optimum estimate $\hat{x}_s$ (in the mean-square sense). The argument relies on the convergence of the suboptimal estimate obtained in

TABLE IV
SIGNAL SMOOTHING: EXAMPLE IV

Quantity	Value	Comments
F	$\begin{bmatrix} 0 & 1 \\ -1 & -9 \end{bmatrix}$	
G	$\begin{bmatrix} 0 \\ 1.42 \end{bmatrix}$	Single-input single-output system. Transfer function $\dfrac{1.42}{s^2 + as + 1}$.
H'	$[1 \quad 0]$	
Q	1000	
R	1	
P	$\begin{bmatrix} 3.993 & 7.974 \\ 7.974 & 107.6 \end{bmatrix}$	Filter error covariance.
$F - KH'$	$\begin{bmatrix} -3.993 & 1 \\ -8974 & -9 \end{bmatrix}$	Closed-loop filter matrix.
$\alpha \pm j\beta$	$-6.49 \pm j1.65$	Roots of $F - KH'$.
Δ	0.308	$\Delta \approx 2/\alpha$.
P_s	$\begin{bmatrix} 1.866 & 0.7479 \\ 0.7479 & 81.65 \end{bmatrix}$	Optimum fixed-lag smoothed error covariance ($\Delta = 0.308$).
F_a	$\begin{bmatrix} -3.956 & -6.606 & 0 & 0 \\ 6.606 & -3.956 & 0 & 0 \\ 0 & 0 & -3.956 & -6.606 \\ 0 & 0 & 6.606 & -3.956 \end{bmatrix}$	
G_a	$\begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix}$	See Fig. 3 for illustration of F_a, G_a, and H_a'
H_a'	$\begin{bmatrix} -0.6243 & 0.4837 & -0.044 & 0.1015 \\ 0.044 & -0.1015 & -0.6243 & 0.4837 \end{bmatrix}$	
P_{sa}	$\begin{bmatrix} 2.289 & 1.734 \\ 1.734 & 84.19 \end{bmatrix}$	Suboptimal fixed-lag smoothed error covariance.

Section IV to $\hat{x}_s$ in conjunction with (11) of Section III and is as follows. The estimate $\hat{x}_{sa}$ obtained in this section is a better approximation to $\hat{x}_s$ than the corresponding estimates of Section IV when the corresponding dynamics have equal dimension, since $\operatorname{tr}(P_{sa})$ obtained from the minimization problem is smaller than that of Section IV. As the dimension of the dynamics increases, the estimates in Section IV converge to the optimum estimate. Thus the estimate obtained in this section also converges to the optimum estimate. All this of course assumes that the minimization problem of this section is practically solvable. In theory this section offers a superior approach to suboptimal smoother design, but from the practical (computational) point of view, Sections IV and V offer a superior approach.

v) The modifications necessary when the signal estimate is required or in the case where there is correlation between the input and the output noise processes are obvious.

VII. Conclusion

The underlying feature of the smoothers described in Sections IV–VI is the approximation of the operator associated with optimum fixed-lag smoothing. This approach enables us to obtain simple and practical solutions to the fixed-lag smoothing problem while achieving asymptotic stability for all the smoothers. The attractiveness of these smoothers includes good performance for quite simple arrangements, at least for the examples given.

One can conceivably utilize another approach in continuous-time fixed-lag smoothing, by sampling the continuous-time measurement process and performing a recursive fixed-lag estimation for the state (or signal) vector. This idea is used in [20], in connection with filtering for a nonlinear model.

The key idea of Section IV was to replace the smoothing problem by the problem of finding the exact filter for an approximately correct signal model. This idea lends itself to a number of other applications, including nonlinear smoothing. The key idea of Section V was to replace the exact smoother by an approximately correct smoother for the exact signal model. This procedure demands knowledge of the exact smoother, and, as such, would be more difficult to extend to, say, a nonlinear smoothing problem.

References

[1] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Trans. ASME, Series D: J. Basic Eng.*, vol. 82, pp. 35–45, 1960.

[2] R. E. Kalman and R. S. Bucy, "New results in linear filtering and prediction theory," *Trans. ASME, Series D: J. Basic Eng.*, vol. 83, pp. 95–107, 1961.

[3] H. E. Rauch, "Solutions to the linear smoothing problem," *IEEE Trans. Automat. Contr.* (Short Paper), vol. AC-8, pp. 371–372, Oct. 1963.

[4] J. S. Meditch, "On optimal linear smoothing theory," *J. Inform. Contr.*, vol. 10, pp. 598–615, 1967.

[5] B. D. O. Anderson, "Properties of optimal linear smoothing," *IEEE Trans. Automat. Contr.*, vol. AC-14, pp. 114–115, Feb. 1969.

[6] B. D. O. Anderson and S. Chirarattananon, "Smoothing as an improvement on filtering: A universal bound," *Electron. Lett.*,

vol. 7, 1971.

[7] R. E. Kalman, "New methods and results in linear filtering and prediction theory," in *Proc. Symp. Eng. Appl. Probability and Random Functions*. New York: Wiley, 1961.

[8] B. D. O. Anderson, "Stability properties of Kalman–Bucy filters," *J. Franklin Inst.*, vol. 291, pp. 137–144, 1971.

[9] C. N. Kelly and B. D. O. Anderson, "On the stability of fixed-lag smoothing algorithms," *J. Franklin Inst.*, vol. 291, pp. 271–281, 1971.

[10] S. Chirarattananon and B. D. O. Anderson, "The fixed-lag smoother as a finite-dimensional linear system," *Automatica*, vol. 7, pp. 657–665, 1971.

[11] J. B. Moore, "Discrete-time fixed-lag smoothing algorithms," *Automatica*, vol. 9, pp. 163–174, 1973.

[12] P. K. S. Tam, J. B. Moore, and B. D. O. Anderson, "Stable realizations of the optimal continuous-time fixed-lag smoothing equations," in *Proc. 3rd Symp. Nonlinear Estimation Theory and Its Applications*, to be published.

[13] N. Wiener, *Extrapolation, Interpolation, and Smoothing of Stationary Time Series with Engineering Applications*. Cambridge, Mass.: M.I.T. Press, 1949.

[14] J. S. Meditch, *Stochastic Optimal Linear Estimation and Control*. New York: McGraw-Hill, 1969.

[15] A. Lindquist, "On optimal stochastic control with smoothed information," *Inform. Sci.*, vol. 1, pp. 55–85, 1968.

[16] T. Kailath and P. Frost, "An innovations approach to least-squares estimation—Part I: linear filtering in additive white noise," *IEEE Trans. Automat. Contr.*, vol. AC-13, pp. 646–655, Dec. 1968.

[17] ——, "An innovations approach to least-squares estimation—Part II: linear smoothing in additive white noise," *IEEE Trans. Automat. Contr.*, vol. AC-13, pp. 655–661, Dec. 1968.

[18] B. D. O. Anderson and S. Chirarattananon, "New linear smoothing formulas," *IEEE Trans. Automat. Contr.*, vol. AC-17, pp. 160–161, Feb. 1972.

[19] L. Weinberg, *Network Analysis and Synthesis*. New York: McGraw-Hill, 1962.

[20] C. N. Kelly and S. C. Gupta, "Discrete-time demodulation of continuous-time signals," *IEEE Trans. Inform. Theory*, vol. IT-18, pp. 488–493, July 1972.