A Nonlinear Fixed-Lag Smoother for Finite-State Markov Processes

DAVID CLEMENTS, STUDENT MEMBER, IEEE, AND BRIAN D. O. ANDERSON, FELLOW, IEEE

Abstract—**The fixed-lag smoothing of random telegraph type signals is studied. The smoothers are derived by first obtaining fixed-point smoothing equations and then using a time discretization. Simulation results are described that verify the qualitative carry-over of known results for the linear-Gaussian problem: the greater the lag, the greater the improvement; beyond a certain lag, no further improvement is obtained by the increase of lag; and the higher the signal-to-noise ratio (SNR), the greater is the improvement over filtering obtained through the use of smoothing. Smoothing errors of one-half the corresponding filtering error are demonstrated.**

I. Introduction

THE KNOWN results on the filtering of finite-state continuous-time Markov processes to the fixed-lag smoothing of such processes is extended in this paper. In broad terms, the justification for this extension is that the resulting estimation errors can be greatly reduced.

The main results on the filtering of finite-state continuous-time Markov processes are those due to Wonham [1]. He presented a set of simultaneous stochastic differential equations describing the evolution of the filtered probabilities, i.e., the probabilities of the process being in each of the possible states at each time conditioned on the measurements up to that time. Also he was able to show for a specific example that the use of these filtering equations, as opposed to suboptimal filtering equations derived using Wiener filter ideas, gave significant improvement over the Wiener filter approach. This paper builds on Wonham's filtering results to tackle the fixed-lag smoothing problem.

Most past efforts in fixed-lag smoothing problems have been associated with the linear-Gaussian problem and commenced with those of Wiener [2]. Wiener was concerned with stationary processes, derived the optimal transfer function of a fixed-lag smoother, and observed its normally nonrational character. He said little about the problem of actually building a device that would possess as its transfer function the optimal transfer function.

As a logical consequence of the introduction of the Kalman–Bucy filter, other workers, particularly Rauch [3] and Meditch [4], presented discrete-time and continuous-time optimum fixed-lag smoothers. Though applicable to both stationary and nonstationary smoothing problems, these smoothers suffered from the difficulty that they were internally unstable and, therefore, in practical terms, not usable. More recently, the stability problem for the linear-Gaussian processes has been resolved in the discrete-time case in [5] and [6], where the fixed-lag smoothers have been found to be stable when the associated filter has been stable. For the continuous-time case, practical and experimentally proven suboptimal finite-dimensional fixed-lag smoothers have been obtained in [7], while [8] contains an optimal fixed-lag smoother that, in essence, operates by limiting the effects of internal instability.

In the linear-Gaussian case, several properties of fixed-lag smoothers are observed. First, the longer the lag interval, the greater the improvement that smoothing offers over filtering. Second, it is found that beyond a certain lag, the improvement obtained by further increase of the lag is marginal, and in fact tends to zero as the lag goes to infinity. The particular time-constant involved here is the same as that associated with the filter, as opposed to the smoother. The third point is that the higher the signal-to-noise ratio (SNR), the greater the improvement that smoothing offers relative to filtering.

In this paper, besides deriving fixed-lag smoothers for finite-state continuous-time Markov processes, we also attempt to explain to what extent these properties carry over to the nonlinear case. Indeed the carry-over is substantial, and one might reasonably conjecture that these properties are indeed very general.

A detailed outline of the paper is as follows. In Section II, we review the results of Wonham on the filtering of finite-state continuous-time Markov processes and present a small generalization that is needed in a later development. In Section III, fixed-point equations are derived, and these are used in Section IV to derive fixed-lag equations. In strict terms, the recursive equations we present for the fixed-lag smoother are discrete-time equations rather than continuous-time equations. The applicability of these equations to the continuous-time problem is discussed in Section IV. Section V discusses some simulation results obtained where the process under consideration is a random telegraph wave, and some concluding remarks are contained in Section VI.

II. Review of Filtering Results

Consider the continuous-time stochastic process $\{x(t); t \geq 0\}$ where, for each $t \geq 0$, $x(t)$ is a random variable taking on only a finite number of possible values $a_1,\cdots,a_n$. Each a_i is called a state of the process. Let the transition

Manuscript received November 7, 1973; revised February 11, 1975. This work was supported in part by the Australian Radio Research Board and in part by the Australian Research Grants Committee.
The authors are with the Department of Electrical Engineering, University of Newcastle, Newcastle, N.S.W., Australia.

properties be denoted by

$$p_{ij}(h) \triangleq \Pr\{x(t+h) = a_j \mid x(t) = a_i\}, \qquad i,j = 1,\cdots,n, \quad h \geq 0.$$

Now assume that the process is Markov and also satisfies the following transition property:

$$p_{ij}(h) = \begin{cases} 1 - v_i h + o(h), & i = j \\ v_{ij} h + o(h), & i \neq j \end{cases} \tag{1}$$

where v_i, v_{ij} are nonnegative constants satisfying

$$\sum_{\substack{j=1 \\ j \neq i}}^{n} v_{ij} = v_i$$

for all i. This transition property guarantees that for almost all realizations of the process $\{x(t); t \geq 0\}$, the number of jumps in any finite interval of time will be finite.

Let the observation process $\{y(t); t \geq 0\}$ be described by the Ito equation

$$dy(t) = x(t)\,dt + \beta(t)\,dw(t), \qquad y(0) = 0 \tag{2}$$

where $\beta(t)$ is a continuously differentiable function of t, bounded away from zero for $t \geq 0$, and $\{w(t); t \geq 0\}$ is a scalar Wiener process independent of the x-process. Upon formally dividing by dt and interpreting $\dot{w}(t)$ as Gaussian white noise, this equation corresponds to the more familiar version of the observation process consisting of the signal with additive Gaussian white noise. Furthermore, suppose the probability distribution of $x(0)$ is $\{p_1(0), \cdots, p_n(0)\}$ and introduce the conditional probabilities $p_i(t) = \Pr\{x(t) = a_i \mid y(\sigma), 0 \leq \sigma < t\}$, $i = 1,\cdots,n$. Then these quantities evolve according to the set of stochastic differential equations

$$dp_i(t) = \left\{-v_i p_i(t) + \sum_{\substack{j=1 \\ j \neq i}}^{n} v_{ji} p_j(t)\right\} dt + \beta(t)^{-2} p_i(t) \cdot [a_i - \bar{x}(t)][dy(t) - \bar{x}(t)\,dt,] \qquad i = 1,\cdots,n \tag{3}$$

with the initial conditions given by the probability distribution of $x(0)$ and $\bar{x}(t) = \sum_{l=1}^{n} a_l p_l(t)$. These equations were initially derived by Kushner [9] but only formally. Subsequently, Wonham [1] presented a rigorous derivation. For each i, the right side of (3) is the sum of two terms. The first, $\{-v_i p_i(t) + \sum_{j \neq i} v_{ji} p_j(t)\}\,dt$, is easily recognizable as the change in $p_i(t)$ computable *a priori* from the dynamics of the x-process and is independent of the measurement process, while the second term is the change in $p_i(t)$ due to the measurement $dy(t)$. Also, note that by its definition $\bar{x}(t) = E\{x(t) \mid y(\sigma), 0 \leq \sigma < t\}$ and thus is the optimal[1] nonlinear filtered estimate of $x(t)$. Hence $dy(t) - \bar{x}(t)\,dt$ is the differential of the innovations process described by Frost and Kailath [10].

[1] The criterion of optimality is the standard minimum variance estimate.

Now suppose that the observation equation (2) is generalized slightly so that it now becomes

$$dy(t) = h(x(t))\,dt + \beta(t)\,dw(t), \qquad y(0) = 0$$

where $h(\cdot)$ is any real-valued function defined on an appropriate domain. Since $x(t)$ can take only one of a finite set of values $a_1,\cdots,a_n$, let the image of the function h be the finite set of values denoted by $h_1,\cdots,h_n$, not necessarily distinct. Then the conditional probabilities $p_1(t),\cdots,p_n(t)$ can be shown to satisfy

$$dp_i(t) = \left\{-v_i p_i(t) + \sum_{\substack{j=1 \\ j \neq i}}^{n} v_{ji} p_j(t)\right\} dt + \beta(t)^{-2} p_i(t) \cdot [h_i - \bar{h}(t)][dy(t) - \bar{h}(t)\,dt], \qquad i = 1,\cdots,n \tag{4}$$

where the initial conditions are as for (3) and $\bar{h}(t) = \sum_{l=1}^{n} h_l p_l(t)$. The proof of (4) parallels that given by Wonham for (3).

It can now be seen that there is no need to demand that $x(t)$ be scalar but only that h is a real-valued function defined on the states of a finite-state process with transition properties defined by (1).

III. Fixed-Point Smoothing Equations

We now develop the *fixed-point probabilities* for the x-process conditioned on the measurements. We make use of an idea employed by Zachrisson [11] for the continuous-time linear-Gaussian fixed-point smoothing problem and subsequently used by Willman [12] for the corresponding discrete-time problem.

Begin by defining a vector process for each $t \geq 0$, $X(t \mid \tau) = [x(\tau)\; x(t)]'$, with t fixed and τ variable, $\tau \geq t$. Some of the properties of this process follow as a simple consequence of the properties of the x-process.

Clearly, $\{X(t \mid \tau); \tau \geq t\}$ is a finite-state process with $N = n^2$ states $A_1,\cdots,A_N$, with each A_I corresponding to the pair of states (a_i,a_j) and I corresponding to the pair of indices (i,j). (In the remainder of this paper, these two descriptions of a state of the X-process will be used interchangeably, depending on which is the more convenient.) Moreover, it is not hard to show that the X-process is Markov. Now denote the transition probabilities of the X-process by

$$P_{IJ}(h) \triangleq \Pr\{X(t \mid \tau + h) = A_J \mid X(t \mid \tau) = A_I\}, \qquad h \geq 0.$$

Suppose that $I = (i_1,j_1)$ and $J = (i_2,j_2)$, then

$$\begin{aligned} P_{IJ}(h) &= \Pr\{x(\tau + h) = a_{i_2}, x(t) = a_{j_2} \mid x(\tau) = a_{i_1}, x(t) = a_{j_1}\} \\ &= \delta_{j_1 j_2} \Pr\{x(\tau + h) = a_{i_2} \mid x(\tau) = a_{i_1}\} \\ &= \delta_{j_1 j_2} p_{i_1 i_2}(h). \end{aligned}$$

Define the constants V_I, $V_{IJ} \geq 0$ by

$$V_I = v_{i_1}$$

and

$$V_{IJ} = \delta_{j_1 j_2} v_{i_1 i_2}, \quad \text{for } I \neq J. \tag{5}$$

We then have

$$P_{IJ}(h) = \begin{cases} 1 - V_I h + o(h), & \text{for } I = J \\ V_{IJ} h + o(h), & \text{for } I \neq J \end{cases}$$

and

$$V_I = \sum_{\substack{J=1 \\ J \neq I}}^{N} V_{IJ}.$$

Although, we have now shown that *the X-process satisfies properties of the same form as the x-process from which it was derived.*

To obtain smoothing equations, we begin by defining the following quantities for $\tau \geq t$:

$$p_j(t \mid \tau) \triangleq \Pr\{x(t) = a_j \mid y(\sigma), 0 \leq \sigma < \tau\}, \qquad j = 1,\cdots,n$$

$$P_J(t \mid \tau) \triangleq \Pr\{x(t) = a_j, x(\tau) = a_i \mid y(\sigma), 0 \leq \sigma < \tau\}, \qquad J = 1,\cdots,N, \quad J = (i,j).$$

The quantities $p_j(t \mid \tau)$, $j = 1,\cdots,n$, $\tau \geq t$, are the fixed-point probabilities of the states of the x-process that we are trying to find, while the quantities $P_J(t \mid \tau)$, $J = 1,\cdots,N$, $\tau \geq t$, are the *filtered* probabilities of the states of the X-process conditioned on the measurements $\{y(\sigma), 0 \leq \sigma < \tau\}$. Observe that

$$p_j(t \mid \tau) = \sum_{i=1}^{n} \Pr\{x(t) = a_j, x(\tau) = a_i \mid y(\sigma), 0 \leq \sigma < \tau\}$$
$$= \sum_{i=1}^{n} P_I(t \mid \tau), \qquad I = (i,j), \tau \geq t. \tag{6}$$

We now define a suitable measurement process so that we can apply the generalization of Wonham's equations to the X-process and this associated measurement process to obtain the quantities $P_J(t \mid \tau)$, $J = 1,\cdots,N$, $\tau \geq t$.

For $\tau \geq t$, define the y-process by the Ito equation

$$dy(\tau) = h(X(t \mid \tau))\, d\tau + \beta(\tau)\, dw(\tau)$$

where $h(X(t \mid \tau)) = [1\ 0][x(\tau)\ x(t)]' = x(\tau)$. Furthermore, set the initial condition at time $\tau = t$ as

$$P_I(t \mid \tau)|_{\tau=t} = \Pr\{x(t) = a_j, x(t) = a_i \mid y(\sigma), 0 \leq \sigma < t\}$$
$$= \delta_{ij} p_i(t).$$

Thus the initial condition for the evolution of the fixed-point probabilities of the states at time t depends on the filtered probabilities at time t.

These observations now allow the writing of the following set of stochastic differential equations for the quantities $P_I(t \mid \tau)$, $I = 1,\cdots,N$, $\tau \geq t$:

$$dP_I(t \mid \tau) = \left[-V_I P_I(t \mid \tau) + \sum_{\substack{J=1 \\ J \neq I}}^{N} V_{JI} P_J(t \mid \tau) \right] d\tau$$
$$+ \beta(\tau)^{-2} P_I(t \mid \tau)[h(A_I) - \bar{h}(\tau)]$$
$$\cdot [dy(\tau) - \bar{h}(\tau)\, d\tau] \tag{7}$$

with the initial condition $P_I(t \mid t) = \delta_{ij} p_i(t)$ and where

$$\bar{h}(\tau) = \sum_{I=1}^{N} h(A_I) P_I(t \mid \tau) = \sum_{i=1}^{n} \sum_{j=1}^{n} a_i P_I(t \mid \tau)$$
$$= \sum_{i=1}^{n} a_i p_i(\tau) = \bar{x}(\tau).$$

Equation (7) together with (6) and the fact that $\bar{h}(\tau) = \bar{x}(\tau)$, show that the fixed-point probabilities $p_1(t \mid \tau), \cdots, p_n(t \mid \tau)$ of the x-process can be, in theory, produced as the outputs of a nonlinear system driven by the innovations process. Since the fixed-point probabilities $P_I(t \mid \tau)$ sum to one, at most $N - 1$ of the equations in (7) are independent.

IV. Fixed-Lag Equations

At this point, we have available, for each fixed time t, the fixed-point smoothing equations describing the quantities $P_I(t \mid \tau)$, $I = 1,\cdots,N$, with τ variable and $\tau \geq t$. For the fixed-lag smoothing problem we require knowledge of the quantities $P_I(t \mid t + L)$ with L fixed and t varying. One approach, which suggests itself for determining stochastic differential equations for these quantities, is to take the stochastic differential equation (7) for $P_I(t \mid \tau)$, to express $P_I(t \mid t + L)$ as an integral using this equation, and then to compute the differential now letting t rather than τ vary. Although this approach will work for a Gauss–Markov process [13], unfortunately it fails here. Those familiar with the details of [13] might perceive part of the difficulty: in the linear Gauss–Markov case, the fixed-lag estimate can be written as the sum of the filtered estimate and an integral involving an unconditional expectation, whereas in the general nonlinear problem, the integral involves a conditional expectation [10]. This is also illustrated in a later example.

We will thus treat only the discrete-time case and derive the fixed-lag estimate as a linear combination of the states of a discrete-time nonlinear system driven by the measurements. Of course, this still has significance as far as the continuous-time problem goes, since it is frequently both natural and acceptable to sample the continuous-time measurements and use these samples to obtain filtered and smoothed estimates of the state.

Let the time be discretized into intervals of length T such that the instants of time are kT, $k = 0,1,2,\cdots$. Denote $P_I(kT \mid lT)$ by $P_I(k \mid l)$, for $l \geq k$. Then, with $t = kT$ and $\tau = lT$, (7) is discretized, giving

$$P_I(k \mid l + 1)$$
$$= P_I(k \mid l) + \left[-V_I P_I(k \mid l) + \sum_{\substack{J=1 \\ J \neq I}}^{N} V_{JI} P_J(k \mid l) \right] T$$
$$+ \beta(l)^{-2}[h(A_I) - \bar{h}(l)] P_I(k \mid l) W(l),$$
$$\text{with } P_I(k \mid k) = \delta_{ij} p_j(k). \tag{8}$$

Here k is fixed, l takes the values $k, k + 1, k + 2, \cdots$, and $W(l) = \Delta y(l) - \bar{h}(l)T$, with $\Delta y(l)$ the increment in the observation process.

For $l = k, k + 1, \cdots, k + L$, (8) generates $P_I(k \mid k)$, $P_I(k \mid k + 1), \cdots, P_I(k \mid k + L)$, for each fixed k and for all

I. Suppose, however, that we fix l in these equations and write down (8) for each k in the range l to $l - L + 1$. The probabilities appearing in these equations are the $N(L+1)$ quantities $P_I(l - j + 1 \mid l)$, where $I = 1,\cdots,N$ and $j = 1,\cdots, L+1$. Let us now rearrange these equations and, in doing so, define the new variables $F_{Ij}(l)$ by $F_{Ij}(l) = P_I(l - j + 1 \mid l), I = 1,\cdots,N$ and $j = 1,\cdots, L+1$. Clearly, from (8) the following set of equations generates the $F_{Ij}(l+1)$ from the $F_{Ij}(l)$,

$$\begin{aligned} F_{I,j+1}(l+1) &= F_{Ij}(l) + \left[-V_I F_{Ij}(l) + \sum_{\substack{J=1\\ J\neq I}}^{N} V_{JI} F_{Jj}(l) \right] T \\ &\quad + \beta(l)^{-2}[h(A_I) - \bar{h}(l)] F_{Ij}(l) W(l), \\ &\qquad j = 1,\cdots,L, \quad I = 1,\cdots,N \quad (9) \end{aligned}$$

where

$$\bar{h}(l) = \sum_{I=1}^{N} h(A_I) F_{I1}(l), \tag{10}$$

while the quantities F_{I_1} are none other than the filtered probabilities associated with the original x-process. Accordingly, these quantities are updated via the discrete-time version of the usual filtering equation (3).

The fixed-lag estimate of $x(l - L)$ given $\{y(k), k = 0,\cdots,l\}$ is then given by the straightforward formula

$$\sum_{I=1}^{N} h(A_I) F_{I,L+1}(l). \tag{11}$$

This follows immediately from the fact that $F_{I,L+1}(l) = P_I(l - L \mid l)$. The initial value for $F_{Ij}(l)$ is $F_{Ij}(L)$, for each I and j.

There is essentially nothing in (9)–(11) not contained in (8). What we have done though is to write down the discrete-time fixed-lag smoothing equations explicitly in state-space form with the state vector defined by the $N(L+1)$ quantities[2] $F_{Ij}(l)$, $I = 1,\cdots,N$, $j = 1,\cdots, L+1$; and the output of the system, the fixed-lag estimate, is given as a linear combination of the components of the state $F(l)$ by (11).

There is an alternative approach to the generation of the fixed-lag smoothing equations which proceeds as follows. Begin by replacing the continuous-time signal model by an approximating discrete-time model. This involves setting up a Markov chain (corresponding to the x-process) and an observation process. The filtering problem is easily solved. Following an approach developed in [6] for discrete-time linear smoothing, one can solve the smoothing problem by augmenting the signal model and constructing a filter for the augmented model; this filter contains within it a fixed-lag smoother for the original (unaugmented) signal model.

[2] Since the fixed-point probabilities must always sum to unity, at most $(N-1)$ fixed-point smoothing equations are independent. In general, the fixed-lag smoothing equations can then be written in terms of $(N-1)(L+1)$ quantities only.

Random Telegraph Wave Example

The simplest example for the preceding x-process is the random telegraph wave, i.e., a process with two states $a_1 = 1$, $a_2 = -1$, and transition constants $v_i = v_{ij} = v$, for all i,j. The filtering equation for this process is then given by

$$dq(t) = -2vq(t) + \beta^{-2}[1 - q(t)^2][dy(t) - q(t)\,dt], \qquad q(0) = 0$$

with $q(t) = p_1(t) - p_2(t)$ corresponding to $\bar{x}(t)$ in (3). For fixed-point smoothing, define the states $A_1,\cdots,A_4$ by

$$A_1 = \begin{bmatrix}1\\1\end{bmatrix} \qquad A_2 = \begin{bmatrix}1\\-1\end{bmatrix}$$
$$A_3 = \begin{bmatrix}-1\\1\end{bmatrix} \qquad A_4 = \begin{bmatrix}-1\\-1\end{bmatrix}.$$

Using (5), we obtain

$$\begin{aligned} V_1 &= V_2 = V_3 = V_4 = v \\ V_{IJ} &= v, \qquad \text{for } I \neq J,\ |I - J| \text{ even} \\ &= 0, \qquad \text{for } I \neq J,\ |I - J| \text{ odd}, \quad I, J = 1,\cdots,4. \end{aligned}$$

We have at most three independent variables for the fixed-point smoothing equations so, for convenience, we define q, Z, and D by

$$\begin{aligned} q &= P_1 + P_2 - P_3 - P_4 \\ Z &= P_1 - P_2 + P_3 - P_4 \\ D &= P_1 + P_4 \end{aligned}$$

where q and Z are the filtered and fixed-point estimates, respectively. In terms of these variables, (7) becomes, for the random telegraph wave,

$$dq(\tau) = -2vq(\tau) + \beta^{-2}[1 - q(\tau)^2][dy(\tau) - q(\tau)\,d\tau], \qquad q(0) = 0 \quad \text{(12a)}$$

$$dZ(t \mid \tau) = \beta^{-2}[2D(t \mid \tau) - 1 - Z(t \mid \tau)q(\tau)] \cdot [dy(\tau) - q(\tau)\,d\tau], \qquad Z(t \mid t) = q(t) \quad \text{(12b)}$$

$$dD(t \mid \tau) = v[1 - 2D(t \mid \tau)]\,d\tau + \beta^{-2} \cdot [\tfrac{1}{2}q(\tau) + \tfrac{1}{2}Z(t \mid \tau) - D(t \mid \tau)q(\tau)] \cdot [dy(\tau) - q(\tau)\,d\tau], \qquad D(t \mid t) = 1. \quad \text{(12c)}$$

Time discretization of these fixed-point smoothing equations (12) then forms the basis for the construction of a set of $3(L+1)$ fixed-lag smoothing equations, following the method used in the previous section.

If we integrate (12b) from t to $t + L$, we have

$$Z(t \mid t + L) = q(t) + \int_t^{t+L} \beta^{-2}[2D(t \mid \tau) - 1 - Z(t \mid \tau)q(\tau)][dy(\tau) - q(\tau)\,d\tau]. \tag{13}$$

This explicitly shows that the fixed-lag estimate is the filtered estimate plus a correction term involving an integral with respect to the innovations process. Thus (13) is a specific example of the general form of the solution of the nonlinear estimation problem obtained by Frost and Kailath [10]. Also, it is particularly easy to see here how any attempt to find an Ito equation for $Z(t \mid t + L)$ would involve terms of the form $(\partial/\partial t)D(t \mid \tau)$ so that a backward conditional operator of some form would be required.

V. Simulation, Discussion, and Results

The fixed-lag smoother was simulated on a digital computer for the special case of the random telegraph wave. The input generated was a Markov chain $s(kT)$, $k = 0,1,2,\cdots$, taking values a_i, $i = 1,\cdots,n$, with transition probabilities given by (1) with h replaced by T; this process is a discrete-time approximation t the $x(\cdot)$ process of Section II. The measurement process was taken to be $z(k) = s(kT)T + \beta(k)w(k)\sqrt{T}$, with $w(k)$ a white Gaussian sequence of zero mean and unit variance; in this way, (2) is approximated. Before discussing the results we consider two points of interest that arose during the simulation and that are important in obtaining realistic results.

For illustration we will consider (14), the discrete-time filtering equation for the random telegraph wave (see (12a)) though the following comments apply equally to the fixed-point (and hence the fixed-lag) smoothing equations:

$$q(k+1) - q(k) = -2\nu q(k)T + \beta^{-2}[1 - q(k)^2] \cdot [x(k) - q(k)]T + \beta^{-1} \cdot [1 - q(k)^2]w(k)\sqrt{T},$$
$$q(0) = 0. \quad (14)$$

The definition of q suggests that we demand $|q(k)| \leq 1$, for all k. However, from (14) we have that, for each k, there is a positive probability that $|q(k+1)| > 1$; all that is needed is a sufficiently large value of $|w(k)|$. Also, as the length of the sample functions being considered increases, the probability of the set of sample functions that violate this bound approaches one. Moreover, once this bound has been broken, the quantity $[1 - q(k)^2][x(k) - q(k)]$ in (14) acts as a "destabilizing" term, and if the noise w then assumes a series of positive or negative values (depending on the sign of $q(k)$), it is clear that q may quickly become arbitrarily large. Thus, in this sense, the discrete-time simulation is unstable. More important, it is an unsatisfactory simulation of the associated continuous-time filtering equation. For further discussion (including analytical results) of this type of behavior exhibited by another example, see [14].

For the simulation of the fixed-lag smoothing equations, this unstable behavior[3] is apparent within 200 sample

[3] For values of νT comparable with one (or even larger), the tendency towards instability is heightened.

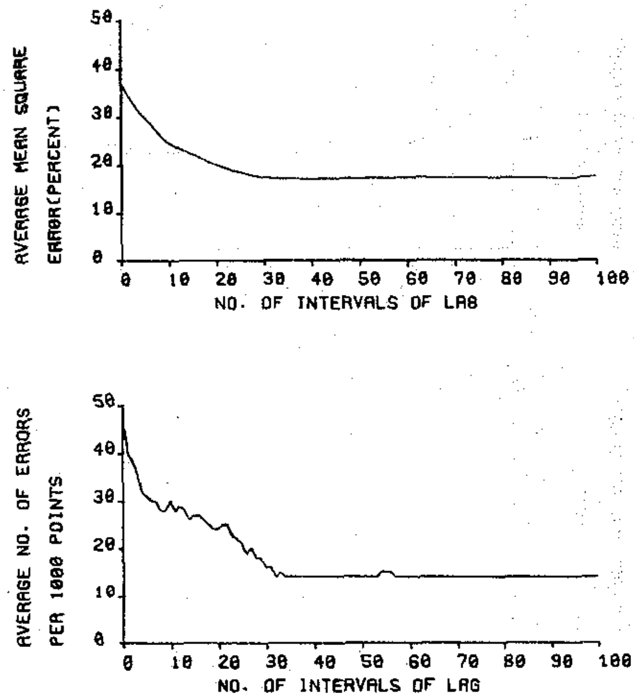

Fig. 1. Variation of error with lag ($\nu = 40$, $\beta = 0.05$, $T = 0.0003$).

points for values of β and T such that the factor $\beta^{-1}\sqrt{T}$ is of the order of magnitude one.

A heuristic argument suggests that instability will always eventually occur. The problem for the filtering equation is easily remedied by the following ad hoc procedure; after each iteration redefine $q(k)$ as $+1$, if $q(k) > 1$, and as -1, if $q(k) < -1$, and leave it unaltered, otherwise. More generally, for the fixed-point (and fixed-lag) equations, the quantities D and Z are similarly forced to satisfy $0 \leq D \leq 1$ and $|Z| \leq 1$.

Since this instability problem occurs because of large values of the noise, an alternative solution is to bound the magnitude of the noise samples by let us say V, and then try to find a sampling interval T^* such that, for each $T < T^*$, $|q(k)| \leq 1$, for all k. For (14) such a T^* can easily be found in terms of V, β, and ν and in principle can also be found for the fixed-lag equations. This method has not been used for the elimination of the stability problem in simulation studies.

The other point of interest is associated with the requirement that the discrete-time equation (14) must be an approximation to its counterpart continuous-time equation (12a). For this we require each term on the right side of (14) to the "small." An ad hoc approach to this bounding which was found to be a successful guideline for the simulation is as follows. For β small, $q(k)$ should be close to $x(k)$ away from a jump region, so assume $|q(k)| = 1$. Then by (14) $|q(k+1)| = 1 - 2\nu T$. Now choose T so that νT is small, from which it follows that the term $1 - q(k+1)^2$ is well approximated by $4\nu T$. Furthermore, if a jump occurs at $k + 1$, the worst possible value of $|x(k+1) - q(k+1)|$ is two. Thus, altogether, the middle term on the right side of (5.1) is probably overbounded by $8\beta^{-2}\nu T^2$. Arbitrarily

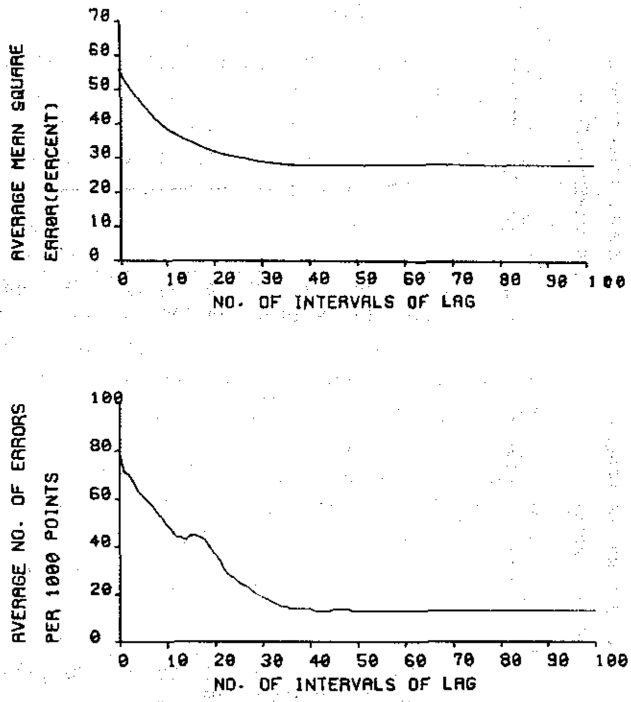

Fig. 2. Variation of error with lag ($\nu = 10$, $\beta = 0.1$, $T = 0.001$).

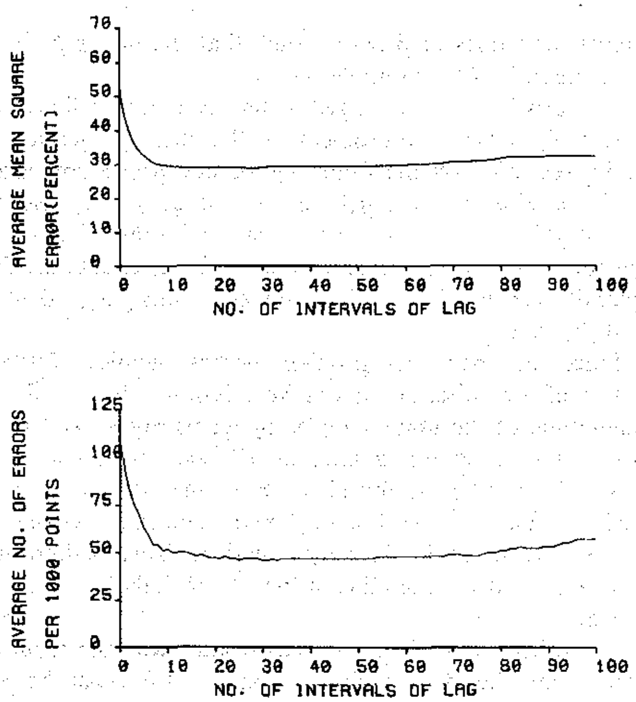

Fig. 3. Variation of error with lag ($\nu = 4$, $\beta = 0.2$, $T = 0.01$).

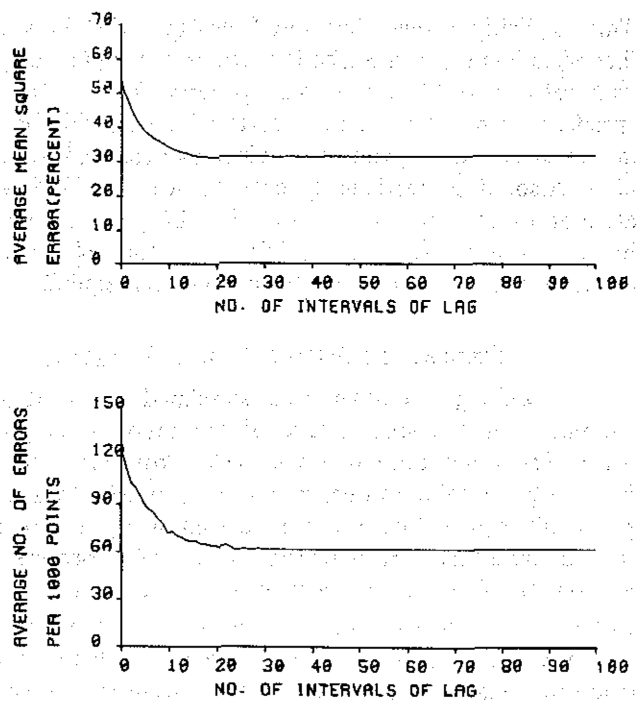

Fig. 4. Variation of error with lag ($\nu = 50$, $\beta = 0.07$, $T = 0.0005$).

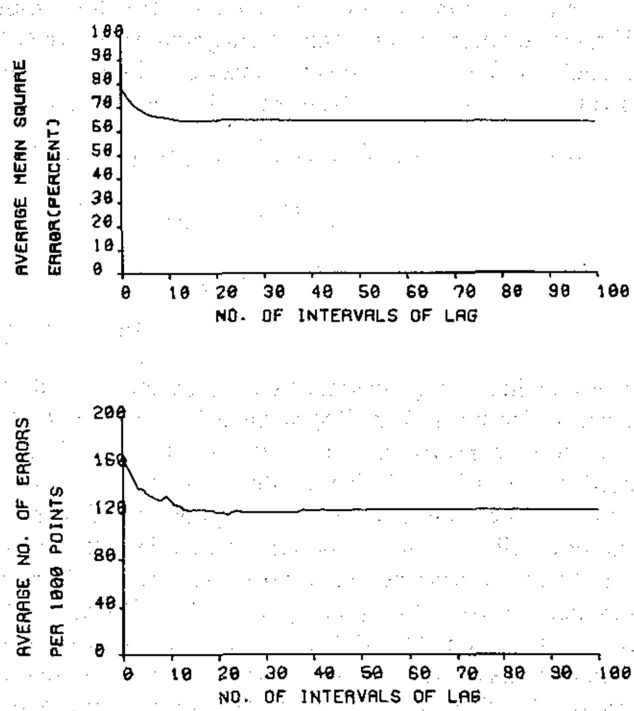

Fig. 5. Variation of error with lag ($\nu = 50$, $\beta = 0.1$, $T = 0.0005$).

choosing the bound 1/50, this implies that T should be chosen so that

$$T < \beta/20\sqrt{\nu}. \tag{15}$$

This bound on T, together with the requirement that νT be small, also guarantees that the final term of (14) is small for the more probable values of w. From (15) we conclude that the smaller is the noise intensity, the greater is the amount of computation required per unit of time; or, equivalently, the smaller is the required sampling interval.

The results of the simulation are shown in Figs. 1–5 and are in order of increasing μ, where $\mu = \nu\beta^2$. The bound (15) is satisfied for each of the figures except Fig. 3 though (15) is not violated excessively. In each of Figs. 1–5, the upper diagram results from a mean-square error criterion (for which the theory was developed) while the lower diagram results from a threshold criterion. The most striking feature of the graphs is the rapid initial improvement in estimation error with lag that then evens out to a

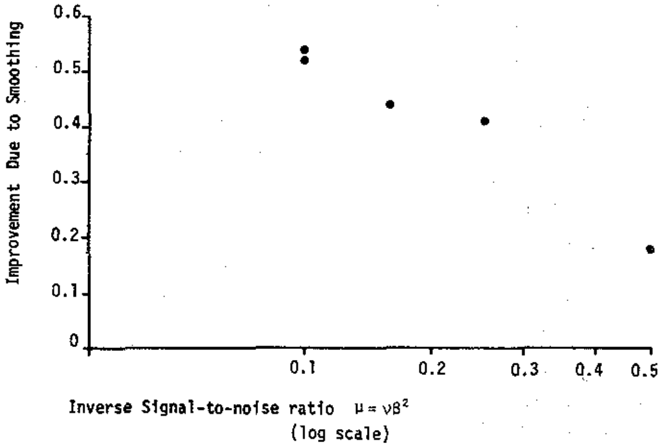

Fig. 6. Variation of improvement due to smoothing with inverse SNR. [Improvement = 1 − (smoothing mean-square error/filtering mean-square error).] Lag is taken to be that giving maximum possible improvement.

steady-state value. This property is well known for Gauss–Markov processes, with the minimum lag required for practically all the improvement being of the order of the dominant time-constant of the Kalman–Bucy filter [15]. For our case, we can associate with the filter the time-constant $1/2\nu$ or, equivalently, $1/2\nu T$ sample intervals for the discrete-time equation (14). For each of Figs. 1–5 it can be seen that the number of intervals of lag for which all improvement is obtained is considerably less than the corresponding value of $1/2\nu T$.

Furthermore, it is shown in [16] that, for stationary processes, the upper bound on the improvement possible with linear fixed-lag smoothing as compared with filtering is a monotonically increasing function of the SNR. Fig. 6 shows a plot of the improvement obtained versus the parameter μ for each of Figs. 1–5. It thus seems plausible that the property for linear smoothing carries over to the case here since, as we now argue, μ^{-1} can be thought of as an SNR. The quantity ν measures the mean number of switchings per second of the x-process and thus defines an effective bandwidth of the system as ν. The signal is confined to this bandwidth, and being always one in magnitude, has an associated signal power of one, whereas the noise power, being β^2 per *unit* bandwidth, is $\nu\beta^2 = \mu$ across the system effective bandwidth. Hence the SNR computed across the effective bandwidth is μ^{-1}.

Finally, it should be noted that physical implementation of the proposed fixed-lag smoother essentially requires the construction of a number of parallel fixed-point smoothers equal to the number of steps of lag. Thus computer speed will place a lower bound on the attainable T, and computer speed (in the absence of a parallel processor) and/or memory an upper bound on the number of steps of lag L.

Despite the plethora of contraints on the size of T and the ad hoc nature of the technique used to eliminate stability, it is gratifying to observe, in higher SNR's, mean-square smoothing errors of less than half the corresponding filtering error and maximum *a posteriori* error rates of less than one-third the corresponding filter error rates.

VI. Conclusions

In the previous sections, we have shown how an experimentally satisfactory fixed-lag smoother can be developed for a discrete-state continuous-time Markov process. The smoother exhibits characteristics similar to those of smoothers that arise in the linear-Gaussian problem, including the fact that there is a certain lag beyond which it is unnecessary to operate in order to achieve all the advantages that smoothing offers over filtering, and the fact that in higher SNR situations, the improvement obtainable using smoothing over filtering is very substantial.

A number of questions still remain however. One might well wonder whether or not there are true continuous-time structures that would serve as fixed-lag smoothers, either through being suboptimal or, though internally unstable in a sense, controlling the instability so that the net effect is still to have a stable arrangement as for the linear case in [8]. A second problem area stems from the consideration of a standard technological situation; suppose a discrete-state process is used as the input to a linear system, even a finite-dimensional linear system. Suppose noisy measurements at the output of this system are available. Then the problem is to estimate, with a fixed-lag smoother, the input process to the linear system. This problem in some sense is a statement of the intersymbol interference problem encountered with telephone channels.

References

[1] W. M. Wonham, "Some applications of stochastic differential equations to optimal nonlinear filtering," *SIAM J. Contr.*, vol. 2, pp. 347–369, 1965.
[2] N. Wiener, *Extrapolation, Interpolation and Smoothing of Stationary Time Series*. Cambridge, Mass.: M.I.T. Press, 1949.
[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] S. Chirarattananon and B. D. O. Anderson, "The fixed-lag smoother as a finite-dimensional linear system," *Automatica*, vol. 7, pp. 657–665, 1971.
[6] J. B. Moore, "Discrete-time fixed-lag smoothing algorithms," *Automatica*, vol. 9, no. 2, pp. 163–174, 1973.
[7] S. Chirarattananon and B. D. O. Anderson, "Stable fixed-lag smoothing of continuous-time processes," *IEEE Trans. Inform. Theory*, vol. IT-19, pp. 25–36, Nov. 1973.
[8] P. K. S. Tam and J. B. Moore, "Stable realization of fixed-lag smoothing equations for continuous-time signals," *IEEE Trans. Automat. Contr.*, vol. AC-19, pp. 84–87, Feb. 1974.
[9] H. J. Kushner, "On the differential equations satisfied by conditional probability densities of Markov processes, with applications," *SIAM J. Contr.*, vol. 2, pp. 106–119, 1964.
[10] P. Frost and T. Kailath, "An innovations approach to least-squares estimation—Part III: Nonlinear estimations in white Gaussian noise," *IEEE Trans. Automat. Contr.*, vol. AC-16, pp. 217–226, June 1971.
[11] L. E. Zachrisson, "On optimal smoothing of continuous-time Kalman processes," *Inform. Sci.*, vol. 1, pp. 143–172, 1969.
[12] W. W. Willman, "On the linear smoothing problem," *IEEE Trans. Automat. Contr.* (Corresp.), vol. AC-14, pp. 116–117, Feb. 1969.
[13] T. Kailath and P. Frost, "An innovations approach to least-squares estimation—Part II: Linear smoothing in additive white noise," *IEEE Trans. Automat. Contr.*, vol. AC-13, pp. 655–660, Dec. 1968.
[14] D. J. Clements and B. D. O. Anderson, "Well-behaved Itô equations with simulations that always misbehave," *IEEE Trans. Automat. Contr.* (Corresp.), vol. AC-18, pp. 676–677, Dec. 1973.
[15] B. D. O. Anderson, "Properties of optimal linear smoothing," *IEEE Trans. Automat. Contr.* (Corresp.), vol. AC-14, pp. 114–115, Feb. 1969.
[16] B. D. O. Anderson and S. Chirarattananon, "Smoothing as an improvement on filtering: A universal bound," *Electron. Lett.*, vol. 7, no. 18, 1971.