Smoothing Algorithms for Nonlinear Finite-Dimensional Systems

BRIAN D. O. ANDERSON†

Department of Systems Engineering, Australian National University, Canberra, ACT 2600, Australia.

and

IAN B. RHODES‡

Department of Electrical and Computer Engineering, University of California, Santa Barbara, CA 93106.

(Accepted for publication October 26, 1982)

Systems are considered where the state evolves either as a diffusion process or as a finite-state Markov process, and the measurement process consists either of a nonlinear function of the state with additive white noise or as a counting process with intensity dependent on the state. Fixed interval smoothing is considered, and the first main result obtained expresses a smoothing probability or a probability density symmetrically in terms of forward filtered, reverse-time filtered and unfiltered quantities; an associated result replaces the unfiltered and reverse-time filtered quantities by a likelihood function. Then stochastic differential equations are obtained for the evolution of the reverse-time filtered probability or probability density and the reverse-time likelihood function. Lastly, a partial differential equation is obtained linking smoothed and forward filtered probabilities or probability densities; in all instances considered, this equation is not driven by any measurement process. The different approaches are also linked to known techniques applicable in the linear-Gaussian case.

1. INTRODUCTION

Consider a state process x_t and an observation or measurement process z_t both evolving forward in time, with the measurement process depending in part on the state process. The task of estimating x_t from z_s, $s<t$, is termed *filtering*; the task of estimating x_t from $z_s, s<T$ from some $T>t$ is

†Work supported by the Australian Research Grants Committee. Part of this work was carried out at the address of the second author.

‡Work supported by the Office of Naval Research under Grant N00014-79-C-0549.

termed *smoothing*. Since more measurements are used in smoothing than in filtering when x_t is being estimated, smoothing leads to better estimates and is therefore to be preferred if the time delay inherent in it is acceptable.

Our main goal in this paper is to relate the smoothing problem to the filtering problem for large classes of state and measurement processes. The state processes considered are of two types: processes evolving in accordance with a diffusion equation, and continuous-time, finite-state Markov processes. For the most part, the observation processes considered are also of two types: a nonlinear function of the state contaminated by additive continuous-time white Gaussian noise, and a counting process, the rate of which is dependent on the state. To help understanding of our results, we also indicate briefly some corresponding discrete-time results.

Our first result (Section 3) relates the smoothing problem to the conventional filtering problem and to a reverse-time filtering problem: this requires the estimation of x_t given z_s for $s>t$, rather than $s<t$. Associated with this result is a second one which relates the smoothing problem to the conventional filtering problem and to a reverse-time likelihood ratio—in discrete time the probability of $z_t, z_{t+1}, \ldots, z_T$ given x_t, and an analogous quantity in continuous time.

Now it is easy to note that the solution of a reverse-time filtering problem is just like that of a forward filtering problem, provided that one has a suitable reverse time signal model; accordingly, the next task is to explain how a reverse time signal model can be obtained from a forward-time signal model (Section 4). Such constructions happen to be available for stationary finite-state Markov processes, and for diffusion equation models, see [1–3]. We recall these results, and indicate various extensions to them to encompass the various state and observation process models of interest to us. In this way, we can obtain equations for the evolution of both forward- and reverse-filtered densities (or probabilities in the case of a finite-state process). With an equation also available for the corresponding unfiltered quantity, the smoothing problem is solved, at least in a formal sense.

We also present (Section 5) an equation for the evolution of the reverse-time likelihood ratio, providing a different formal solution to the smoothing problem. The solution involving the reverse-time likelihood ratio is perhaps less appealing than that involving forward- and reverse-filters, since the latter has a pleasing symmetry that reflects the symmetry of the smoothing problem itself. However, it can be that the reverse-time signal model cannot be found, and in this case one is thrown back on using the reverse-time likelihood ratio.

Yet a third style of equation (Section 6) relating forward filtered and smoothed probabilities or probability densities can be found. Actually, such an equation has been known for some time for diffusion equation state models with a measurement process containing additive white Gaussian noise, independent of the state process [4, 5]; also the equation is known for a finite-state Markov process state model with the same measurement process, see [6]. In all cases the equation is undriven by the measurement process.

We referred above to having a solution "in a formal sense" to the smoothing problem: when the state process is defined by a diffusion equation, the forward and reverse filtering densities are the solutions of stochastic partial differential equations, and the smoothed density is found using these two filtered densities. It is accordingly of interest to indicate situations when a finite-dimensional solution to the smoothing problem can be obtained. Besides the standard linear-Gaussian problem and the problem with a finite-state Markov process state model, we describe one such situation in Section 6 involving a linear diffusion equation state model and an observation model containing a mixture of linear-Gaussian observations and space-time counting process observations.

Besides the references noted above on nonlinear smoothing, there have been a number of other contributions (see, e.g. [7–12]) including some which are very significant. None, as far as we are aware, has attempted to use forward and reverse-time filters in a symmetric way, nor has developed, for as comprehensive a set of state and observation models as considered here, the equation for the evolution of the reverse-time likelihood ratio and the equation relating the smoothed and forward filtered densities which is undriven by the measurements.

The general thrust of this paper is to suggest a wide variety of algorithms, rather than to present the fine details governing an existence proof associated with one algorithm. To the extent that the different competing algorithms for linear-Gaussian smoothing are now all apparently captured in a non-linear framework, we may therefore also be setting out a representative collection of bases upon which to build finite-dimensional (near- but non-optimal) smoothers.

2. NOTATION AND BACKGROUND ASSUMPTIONS

Throughout the paper we shall consider stochastic processes defined on a fixed measurable space (Ω, F). Unless otherwise indicated, all processes are considered to be defined on the fixed interval $[0, T]$ and to be F_t-adapted, where $\{F_t, t \in [0, T]\}$ is an increasing family of sub-σ-algebras of F. Mostly, we shall be concerned with continuous-time processes in which $[0, T]$ has its usual meaning; however, we shall also consider briefly discrete-time processes in which case $t, T \in Z_+$, the set of positive integers. We denote by $X \triangleq \{x_t, F_t, P\}$ a finite-dimensional *state* process which is F_t-adapted and P-measurable on (Ω, F) under the probability measure P. We denote by $\sigma(x_t)$ the sub-σ-algebra of F_t generated by the random variable x_t, and by X_{t-} and X_{t+} the sub-σ-algebras of F_t generated by the collections of random variables $\{x_s; s[0, t)\}$ and $\{x_s; s(t, T]\}$.

We shall concentrate on two x_t-process models—a diffusion model D, and a finite state Markov model F, defined as follows: First,

$$D \qquad dx_t = f(x_t, t)\,dt + g(x_t, t)\,dv, \tag{2.1}$$

where $V = \{v_t, F_t, P\}$ is a Wiener process taking values in R^p, f and g satisfy appropriate smoothness and rate-of-growth conditions, and x_0 is independent of $\{v_t\}$.

The finite-state model F is defined as being a Markov process such that if p_t is an n-vector with i-th entry the probability that x_t is at the i-th of n levels, then

$$F \qquad dp_t = A(t)p_t\,dt \tag{2.2}$$

The matrix $A(t)$ satisfies

$$a_{ij}(t) \geqq 0 \quad \text{for} \quad i \neq j, \; \sum_{k=1}^{n} a_{ki}(t) = 0. \tag{2.3}$$

It is, of course, possible to consider combinations of D and F, and the subsequent ideas and results of the paper can be extended to such combined models.

Associated with the state process X there will also be an *observation or measurement* process $Z = \{z_t, F_t, P\}$, which is linked to x_t by some form of observation model. Like the x-process, Z is finite-dimensional, F_t-adapted and P-integrable on (Ω, F) under the probability measure P. In slight contrast to our definition of X_{t-} and X_{t+}, we define Z_{t-} and Z_{t+} to be the sub-σ-algebras of F_t generated by the increments of z_s in the intervals $[0, t)$ and $[t, T)$. We shall be interested in a model G with additive white Gaussian measurement noise and a model C with z a counting process whose intensity depends on the state. In more detail

$$G \qquad dz_t = h(x_t, t)\,dt + M(t)\,dw_t, \quad z_0(\omega) = 0, \tag{2.4}$$

where $W = \{w_t, F_t, P\}$ is a Wiener process taking values in

R^m, $M(t): R^m \to R^m$ is deterministic and invertible, and h is jointly measurable in x_t and t. The process W is independent of X and V and

$$R(t) \triangleq M(t)M'(t) \tag{2.5}$$

is uniformly bounded with a uniformly bounded inverse, and L_2-integrable on $[0, T]$. Further

$$\int_0^T \|h(x_t, t)\|^2_{R^{-1}(t)} \, dt < \infty \qquad P\text{-a.s.} \tag{2.6}$$

Next, the counting process model is

$$C \qquad dz_t = dN_t \quad z_0(\omega) = 0, \tag{2.7}$$

where $N = \{N_t, F_t, P\}$ is a counting process with intensity $\Lambda = \{\lambda_t, F_t, P\}$ satisfying

$$\lambda_t = h(x_t, t), \tag{2.8}$$

where h is jointly measurable in x_t and t, $\int_0^T h(x_s, s)\, ds < \infty$ P-a.s., and h is such that Λ is predictable and positive a.s. Several calculations involving the appropriate Ito rule also require λ_t to be uniformly bounded a.s.; for convenience we take this bound to be unity, since the extension to arbitrary upper bound is straightforward [see, e.g., 11].

A crucial and well-known consequence of the above assumptions is:

2.1. Observation

With either state model D or F and either observation model G or C, the σ-algebra $X_{t-} \vee Z_{t-}$ is conditionally independent given x_t of $X_{t+} \vee Z_{t+}$. Here $X \vee Z$ denotes the least σ-algebra containing X and Z.

In the sequel, we shall be interested in the probabilities (model F) or probability densities (model D) $p(x_t)$, $p(x_t | Z_{t-})$, $p(x_t | Z_{t+})$, and $p(x_t | Z_T)$. We shall refer to these as unfiltered, filtered or forward filtered, reverse filtered, and smoothed probabilities or probability densities, respectively. Our concern is not to pay close attention to conditions for the existence of these densities, which by and large have been dealt with exhaustively in other works; for the most part, we shall simply assume that the conditions are fulfilled which ensure their existence. At times, however, it would be satisfactory to work with distributions.

3. BASIC SMOOTHING FORMULAS

The various smoothing formulas we first obtain can be thought of as taking one of two forms:

$$p(x_t|Z_{T-})=\frac{p(x_t|Z_{t-})p(x_t|Z_{t+})}{Np(x_t)} \tag{3.1}$$

and

$$p(x_t|Z_{T-})=\frac{p(x_t|Z_{t-})}{N'}\cdot[\text{Object like } p(Z_{t+}|x_t)]. \tag{3.2}$$

Here, N and N' are normalizing quantities, and so are Z_T-dependent, but not x_t-dependent.

We can obtain a feel for these formulas by considering a discrete-time version of one of the models given earlier. Thus, suppose x_t is a continuous-state, discrete-time, Markov-process, and $z_t=h(x_t)+n_t$, where n_t is discrete-time white noise, with the $\{n_t\}$ and $\{x_t\}$ processes independent. Let Z_{t-} denote $\{z_0,z_1,\ldots,z_t\}$, Z_{t+} denote $\{z_{t+1},\ldots,z_T\}$, and Z_T denote $\{z_0,z_1,\ldots,z_T\}$. Then Observation 2.1 becomes simply

$$p(Z_T|x_t)=p(Z_{t-}|x_t)p(Z_{t+}|x_t). \tag{3.3}$$

Now, assuming all densities exist, we have from Bayes' rule and (3.3)

$$p(x_t|Z_T)=\frac{p(Z_T|x_t)p(x_t)}{p(Z_T)}=\frac{p(Z_{t-}|x_t)p(Z_{t+}|x_t)p(x_t)}{p(Z_T)}$$

$$=\frac{p(Z_{t-}|x_t)p(Z_{t+}|x_t)p(x_t)p(x_t)}{p(Z_{t-})p(Z_{t+})p(x_t)}\frac{p(Z_{t-})p(Z_{t+})}{p(Z_T)}$$

and (3.1) is immediate upon reapplication of Bayes' rule and identification of $p(Z_T)[p(Z_{t-})p(Z_{t+})]^{-1}$ as the normalizing quantity N. To derive (3.2), the last equality in the algebra above is replaced by

$$p(x_t|Z_t)=\frac{p(Z_{t-}|x_t)p(x_t)}{p(Z_{t-})}\cdot\frac{p(Z_{t-})}{p(Z_T)}\cdot p(Z_{t+}|x_t),$$

which is (3.2) with $N'=p(Z_T)p^{-1}(Z_{t-})$ and the "object like $p(Z_{t+}|x_t)$" simply $p(Z_{t+}|x_t)$ itself. If x_t is a discrete-time finite-state process, the same calculations yield (3.1) and (3.2) where the p symbols should be interpreted as probabilities rather than probability densities.

The above Bayes' rule calculations cannot be duplicated exactly for continuous-time processes. Nevertheless, tools for capturing the Bayes' idea exist in the form of representation theorems (see, e.g. [13, 14]). For the observation process *G*, define

$$L_{s_1,s_2}=\exp\left\{\int_{s_1}^{s_2} h'(x_s,s)R^{-1}(s)\,dz_s-\tfrac{1}{2}\int_{s_1}^{s_2}\|h(x_s,s)\|^2_{R^{-1}(s)}ds\right\} \tag{3.4}$$

and for observation process *C*, define

$$L_{s_1,s_2}=\exp\left\{\int_{s_1}^{s_2} \ln h(x_s,s)\,dz_s-\int_{s_1}^{s_2}[h(x_s,s)-1]\,ds\right\} \tag{3.5}$$

Then we have for any combination of state and observation models the representation theorem (see, e.g., [11], [15])

$$p(x_s=x\mid Z_{t-})=\frac{E_X[L_{0,t}\mid Z_{t-},x_s=x]}{E_X[L_{0,t}\mid Z_{t-}]}\,p(x_s=x), \tag{3.6}$$

for all $s,t\in[0,T]$. The notation $E_X[L_{0,t}\mid Z_{t-}]$ or $E_X[L_{0,t}\mid Z_{t-},x]$ means that the expectation is to be taken with respect to the process X with Z_{t-} fixed, i.e. the equality in (3.6) holds pointwise for each observation trajectory (Z_{t-}). We note that (3.6) is in the form of the Bucy representation theorem [15] involving only one probability measure, in contrast to representation theorems in which the two conditional expectations in a form similar to (3.6) are with respect to a transformed probability measure.† We note, too, that one way to derive (3.6) is via *measure transformations* [see 11].

One can think of $E_X[L_{0,t}\mid Z_{t-}]$ and $E_X[L_{0,t}\mid Z_{t-},x_s=x]$ as being like $p(Z_{t-})$ and $p(Z_{t-}\mid x_s)$, and then (3.6) becomes Bayes' theorem. This thinking, and the discrete time derivations, guide the following development.

Observe with the aid of (3.4) and (3.5) that

$$L_{0,T}=L_{0,t}L_{t,T}. \tag{3.7}$$

†These versions involving transformed measures hold under conditions that are weaker than our standing assumptions. Pardoux [7, 8, 16] has used a representation involving transformed measures to derive an expression associated with the "asymmetric form" (3.2) of the smoothed density that, for state model *D* and observation model *G*, is more general than Eq. (5.9) in Section 5 in that it includes correlated state and observation noise processes *v* and *w*.

Accordingly, using (3.4) and (3.5) again along with Observation 2.1,

$$E_X[L_{0,T}|Z_{T-},x_t=x]=E_X[L_{0,t}|Z_{t-},x_t=x]\,E_X[L_{t,T}|Z_{t+},x_t=x]. \quad (3.8)$$

With s and t in (3.6) identified with t and T, we then have

$$\begin{aligned}p(x_t=x|Z_{T-})&=\frac{E_X[L_{0,T}|Z_T,x_t=x)}{E_X[L_{0,T}|Z_T]}p(x_t=x)\\&=\frac{E_X[L_{0,t}|Z_{t-},x_t=x]p(x_t=x)}{E_X[L_{0,t}|Z_{t-}]}\\&\quad\cdot\frac{E_X[L_{t,T}|Z_{t+},x_t=x]E_X[L_{0,t}|Z_{t-}]}{E_X[L_{0,T}|Z_T}\\&=p(x_t=x|Z_{t-})E_X[L_{t,T}|Z_{t+},x_t=x]\cdot\frac{E_X[L_{0,t}|Z_{t-}]}{E_X[L_{0,T}|Z_T]}\end{aligned} \quad (3.9)$$

The second equality follows by using (3.7) and the third one using (3.6) with s and t in (3.6) both equal to t. Eq. (3.9) is the concrete form of (3.2), where of course $E_X[L_{t,T}|Z_{t+},x_t=x]$ is the "object like $p(Z_{t+}|x_t)$" and the inverse of the third factor on the right side is the normalizing quantity N'. An equation like this (but involving expectations under a transformed probability measure) is studied by Pardoux in [16] for state model D and observations G and C where a wide range of precise assumptions are set out under which the equation is valid.

To obtain (3.1), notice that if the origin is shifted to time t and we then identify s with t and t with T in (3.6) we have

$$p(x_t=x|Z_{t+})=\frac{E_X[L_{t,T}|Z_{t+},x_t=x]}{E_X[L_{t,T}|Z_{t+}]}p(x_t=x).$$

and when this is used to substitute for the second term on the right side of (3.9), we obtain

$$p(x_t=x|Z_{T-})=\frac{p(x_t=x|Z_{t-})p(x_t=x|Z_{t+})}{p(x_t=x)}\cdot\frac{E_X[L_{0,t}|Z_{t-}]E_X[L_{t,T}|Z_{t+}]}{E_X[L_{0,T}|Z_T]} \quad (3.10)$$

This is (3.1), with a more precise identification of the normalizing quantity N.

The above arguments apply for either observation model G or C, and either state model D or F, save that p must be interpreted for model F as a probability and for model D as a probability density. We note that (3.1) and a version of (3.2) have been derived via an alternative route in [21, Section 6.7] for state model D and observation model C.

4. REVERSE FILTERED DENSITIES

A key conclusion of the preceding section was the formula (3.1) expressing the smoothed probability or probability density symmetrically in terms of a forward filtered, a reverse filtered, and an unfiltered probability or probability density. In this section, we note how equations for the reverse filtered probability or density can be derived.

Let us focus for the moment on the state model D, repeated as

$$dx_t = f(x_t, t)\,dt + g(x_t, t)\,dv_t. \tag{4.1}$$

The properties of this model include independence of x_0 and v_t, and, more generally, independence of x_t and $v_s - v_t$ for any $s > t$, *but not for* $s < t$, i.e. present and past states are independent of future noise, but present and future states are not independent of past noise. The model is thought of as evolving forward in time, and (4.1) is understood as defining a forward Ito equation. What is needed for the calculation of the reverse filtered density is a signal model that evolves backwards in time with the property that, relative to this reverse time evolution, "future" noise is independent of "past" signal.

The main result of [3] is that if $p(x_t, t)$ exists for all t and x, then one may define

$$(dv_t^r)^k = (dv_t)^k + \frac{1}{p(x,t)} \cdot \left\{ \sum_j \frac{\partial}{\partial x_t^j} [p(x_t, t) g^{jk}(x_t, t)]\,dt \right\} \tag{4.2}$$

and

$$[f^r(x_t, t)]^i = [f(x_t, t)]^i - \frac{1}{p(x_t, t)} \left\{ \sum_{j,k} \frac{\partial}{\partial x_t^j} [p(x_t, t) g^{ik}(x_t, t) g^{jk}(x_t, t)] \right\} \tag{4.3}$$

with the following properties:

$$v_t^r \text{ is a Wiener process.} \tag{4.4}$$

$$dx_t \overset{b}{=} f^r(x_t, t)\,dt + g(x_t, t)\,dv_t^r, \tag{4.5}$$

where (4.5) is a backward Ito equation, i.e., the integral form of (4.5) involves a backward Ito integral, and

$$x_t \quad \text{and} \quad v_s^r - v_t^r \text{ are independent for } s < t. \tag{4.6}$$

This means that if t is the present, and $s>t$ is called the "past" for the backwards-evolving (4.5) and $s<t$ the "future", then "future" driving noise is independent of present and past state. (Henceforth, b or f above the equality sign will stress that the equation is a backward or forward equation.)

The above construction is an extension of one in [17] for the linear-Gaussian case.

Just as the state model can be conceived as evolving in reverse time, so can a measurement model. Under our assumptions, the noise W in the measurement model G of (2.4) is independent of the state process X for either state model D or F. In this case the appropriate backward measurement model is just (2.4) itself, i.e.

$$dz_t = h(x_t, t)\,dt + M(t)\,dw_t, \tag{4.7}$$

though (4.7) has two interpretations, as a forward and as a backward equation, i.e.

$$z_{t+dt} - z_t \overset{f}{=} h(x_t, t)\,dt + M(t)[w_{t+dt} - w_t] \tag{4.7f}$$

and

$$z_t - z_{t-dt} \overset{b}{=} h(x_t, t)\,dt + M(t)[w_t - w_{t-dt}]. \tag{4.7b}$$

Had the driving noise V in model D been correlated with the observation noise W in model G, the backward observation model would have differed from the forward model to reflect appropriate correlation with the process $\{v_t^*\}$ in (4.5).

The point of these observations is that they allow the setting up of equations for the reverse filtered density that are just like those for the forward filtered density. For example, for the state model defined by (4.1) and the measurement Eq. (2.4), the unnormalized forward filtered density $p_{fu}(x_t)$ is given from the Zakai [18] equation as, (see, e.g., [7, 8]),

$$dp_{fu} \overset{f}{=} \mathscr{L}[p_{fu}]\,dt + p_{fu} h' R^{-1}\,dz_t, \tag{4.8}$$

where

$$\mathscr{L}(\cdot) = -\sum_i \frac{\partial}{\partial x_t^i}[f^i(x_t, t)\cdot] + \tfrac{1}{2}\sum_{i,j}\frac{\partial^2}{\partial x_t^i \partial x_t^j}\{p[g(x_t, t)g'(x_t, t)]^{ij}\cdot\} \tag{4.9}$$

and $p_{fu}(x_0 \mid Z_{0-}) = p(x_0)$. Accordingly, we see from (4.5) and (4.7b) that the equation for the associated unnormalized reverse-time filtered density is simply

$$dp_{ru} \overset{b}{=} -\mathscr{L}_r[p_{ru}]\,dt - p_{ru}h'R^{-1}\,dz_t, \qquad (4.10)$$

with $\mathscr{L}_r$ defined like $\mathscr{L}$, save that $[-f^r]'$ replaces $[f]'$ in the definition (4.9); the boundary condition is, naturally, $p_{ru}(x_T|Z_{T+}) = p(x_T)$.

The general approach to obtain (4.10) is simple. One has backward models for the state process and measurement process, one makes a change of time variable so that forward models of related processes are obtained, one uses (4.8) to write down the forward unnormalized filtered density, and then one changes the time variable again, to get (4.10). The full details are set out in Appendix A, which also serves as a guide to similar subsequent calculations that are omitted.

As pointed out by a reviewer, this equation can almost certainly be derived from results of [16, Eq. (3.15)], since $p(x_t = x, Z_{t+})$ is given in terms of $E_X[L_{T,t}Z_{t+}, x_t = x]$ just above (3.10). However, such a derivation obscures the essential time symmetry of the smoothing problem, which underpins the development of this section.

Instead of working with the unnormalized equation, one can work if desired with the normalized equation for $p_f = p(x_t|Z_{t-})$. A version for scalar x_t and z_t may be found in [6]. Alternatively, it may be obtained in a similar manner to a calculation of [19]. The answer is standard, but we set out the brief calculation as a guide to later constructions. Let

$$n_f(t) = \int_{R^n} p_{fu}(x_t, t)\,dx_t, \qquad (4.11)$$

where x_t is an n-vector. Then (4.8) implies

$$dn_f(t) \overset{f}{=} \hat{h}_f' R^{-1}\,dz_t n_f(t), \qquad (4.12)$$

where $\hat{h}_{f_t} = E[h(x_t, t)|Z_{t-}]$, and then the Ito rules give

$$d\left(\frac{p_{fu}(x_t, t)}{n_f(t)}\right) = \frac{dp_{fu}}{n_f} - \left(\frac{p_{fu}}{n_f}\right)\left(\frac{dn_f}{n_f}\right) - \left(\frac{dp_{fu}}{n_f}\right)\left(\frac{dn_f}{n_f}\right) + \frac{p_{fu}}{n_f}\left(\frac{dn_f}{n_f}\right)\left(\frac{dn_f}{n_f}\right) \qquad (4.13)$$

which leads to

$$dp_f \overset{f}{=} \mathscr{L}(p_f)\,dt + (h - \hat{h}_f)'R^{-1}(dz_t - \hat{h}_f\,dt). \qquad (4.14)$$

The corresponding reverse equation is, with $\hat{h}_r = E[h|Z_{t+}]$,

$$dp_r \overset{b}{=} -\mathscr{L}_r(p_r)\,dt - (h - \hat{h}_r)'R^{-1}(dz_t - \hat{h}_r\,dt). \qquad (4.15)$$

Now let us turn to the finite-state Markov model F. The construction of a reverse model given a stationary forward model is undertaken in [1, 2]. Here, we note an extension. Suppose that p_t, $t\in[0,T]$, is the solution of (2.2). Suppose further that the entries of p_t are all identically nonzero for t in $[0,T]$. (It is easy to see that this is equivalent to demanding that all entries of p_0 are nonzero.) Define

$$\Pi_t=\text{diag}[p_{1t},\ldots,p_{nt}],\tag{4.16}$$

there being n different state levels. Also, define

$$A^r(t)=\Pi_t A'(t)\Pi_t^{-1}-\dot{\Pi}_t\Pi_t^{-1}.\tag{4.17}$$

Then the reverse model is a finite state Markov process $x^r(\cdot)$ and, with the i-th entry of p_t^r denoting the probability that $x^r(t)$ is in level i,

$$dp_t^r=-A^r(t)p_t^r\,dt\quad p_T^r=p_T.\tag{4.18}$$

It is quickly verified that for all t the entries of A^r satisfy $a_{ij}^r\geqq 0$ if $i\neq j$, and that $\sum_{i=1}^n a_{ij}^r=0$ for all j, as required for (4.18) to be a (backward) finite-state process. To see that (4.18) is, in fact, the reverse model associated with (2.2), it is necessary and sufficient to show that, for all i, j and t,

$$Pr[x(t)=i]=Pr[x^r(t)=i]\tag{4.19}$$

and

$$Pr[x(t)=i\,|\,x(s)=j]=Pr[x^r(t)=i\,|\,x^r(s)=j].\tag{4.20}$$

Now, it is not hard to verify that if $\Phi(t,s)$ is the transition matrix which satisfies

$$\frac{d}{dt}\Phi(t,s)=A(t)\Phi(t,s)\quad\Phi(s,s)=I\tag{4.21}$$

for all t and s, then $\Psi(t,s)=\Pi_t\Phi'(s,t)\Pi_s^{-1}$ satisfies

$$\frac{d}{dt}\Psi(t,s)=-A^r(t)\Psi(t,s)\quad\Psi(s,s)=I,\tag{4.22}$$

for all t and s. Also, observe that $[1,1,\ldots,1]\Phi(t,s)=[1,1,\ldots,1]$ for all t and s; for if p_s is the probability vector at time s, that at time t is $\Phi(t,s)p_s$ and the entries of this vector must sum to 1 for all probability vectors p_s.

Now, to establish (4.19), observe that

$$p_t^r = \Psi(t,T)p_T^r = \Pi_t \Phi'(T,t)\Pi_T^{-1} p_T = \Pi_t \Phi'(T,t)[1,1,\ldots,1]' = \Pi_t[1,1,\ldots,1]' = p_t.$$

Now consider (4.20). Let e_j denote a unit vector with 1 in the j-th position. We have, if $t > s$,

$$p[x_t^r = i \mid x_s^r = j] = \frac{p[x_s^r = j \mid x_t^r = i]\, p[x_t^r = i]}{p[x_s^r = j]} = \frac{e_j' \Psi(s,t) e_i p_{it}}{p_{js}}$$

$$= e_j' \Pi_s^{-1} \Psi(s,t) \Pi_t e_i = e_j' \phi'(t,s) e_i$$

$$= e_i' \phi(t,s) e_j = p[x_t = i \mid x_s = j].$$

A similar argument works for $t < s$ and thus establishes (4.20).

Suppose that when x_t is in the i-th state, $h(x_t, t) = h_i(t)$. Suppose also that the observation model G applies, and let p_{if} denote the i-th entry of the forward filtered probability vector. Thus the forward filter equations are (expanding [20] to the time-varying, multiple-output, case and omitting the common time argument t)

$$dp_{if} \stackrel{f}{=} \sum_j a_{ij} p_{jf}\, dt + [h_i - \sum_{j=1}^{n} h_j p_{jf}]' R^{-1} [dz_t - \sum_{j=1}^{n} h_j p_{jf}\, dt] p_{if}. \quad (4.23)$$

The observation model is the same in reverse time, and so in view of the equality of all joint densities associated with the forward and reverse models, we see also that

$$dp_{ir} \stackrel{b}{=} -\sum_j a_{ij}^r p_{jr}\, dt - [h_i - \sum_{j=1}^{n} h_j p_{jr}]' R^{-1} [dz_t - \sum_{j=1}^{n} h_j p_{jr}\, dt] p_{ir}. \quad (4.24)$$

We also note the unnormalized form of these equations:

$$dp_{ifu}(t) \stackrel{f}{=} \sum_j a_{ij}(t) p_{jfu}(t)\, dt + h_i(t) R^{-1}(t)\, dz_t p_{ifu}(t) \quad (4.25)$$

$$dp_{iru}(t) \stackrel{b}{=} -\sum_j a_{ij}^r(t) p_{jru}(t)\, dt - h_i(t) R^{-1}(t)\, dz_t p_{iru}(t). \quad (4.26)$$

We note that the connection between (4.25) and (4.23) follows easily on using the normalizing factor

$$n_f(t) = \sum_i p_{ifu}(t).$$

which satisfies

$$dn_f(t) \stackrel{f}{=} \left[\sum_{j=1}^{n} h_j(t) \frac{p_{jfu}(t)}{n_f(t)}\right]' R^{-1}(t)\, dz_t n_f(t). \tag{4.27}$$

The use of (4.13) yields (4.23) with $p_{if} = p_{ifu}/n_f$.

Now consider observation model C. When the state model is D, the forward normalized filter equation is [11]

$$dp_f \stackrel{f}{=} \mathscr{L}[p_f]\, dt + [h - \hat{h}_f][\hat{h}_f]^{-1}[dz_t - \hat{h}_f\, dt]p_f. \tag{4.28}$$

The observation model is evidently the same in reverse time as in forward time save for its interpretation as a backwards rather than a forwards equation. Consequently, using arguments analogous to those in Appendix A, the reverse equation for $p_r = p(x_t \mid Z_{t+})$ is

$$dp_r \stackrel{b}{=} -\mathscr{L}_r[p_f]\, dt - [h - \hat{h}_r][\hat{h}_r]^{-1}[dz_t - \hat{h}_r\, dt]p_r. \tag{4.29}$$

We assert that the unnormalized equation corresponding to (4.28) is

$$dp_{fu} \stackrel{f}{=} \mathscr{L}(p_{fu})\, dt + (h - 1)(dz_t - dt)p_{fu} \tag{4.30}$$

which implies that

$$dp_{ru} \stackrel{b}{=} -\mathscr{L}_r(p_{ru})\, dt - (h - 1)(dz_t - dt)p_{ru}. \tag{4.31}$$

The verification that (4.30) leads to (4.28) proceeds in a parallel fashion to the derivation of (4.14) from (4.8) via (4.11)–(4.13) using, in this case, the Ito differential rule for counting process observations that can be found, for example, in [11], [21].

Finally, for the combination of observation model C and state model F, we have for the normalized Eq. [11]

$$dp_{if} \stackrel{f}{=} \sum_j a_{ij} p_{jf}\, dt + [h_i(t) - \hat{h}_f(t)][\hat{h}_f(t)]^{-1}[dz_t - \hat{h}_f(t)\, dt]p_{if} \tag{4.35}$$

and for the unnormalized equation

$$dp_{ifu} \stackrel{f}{=} \sum_j a_{ij} p_{jfu}\, dt + (h - 1)(dz_t - dt)p_{ifu}. \tag{4.32}$$

The reverse equations are

$$dp_{ir} \stackrel{b}{=} -\sum_j a^r_{ij} p_{jr}\, dt - [h_i - \hat{h}_r][\hat{h}_r]^{-1}[dz - \hat{h}_r\, dt]p_{ir} \tag{4.33}$$

and

$$dp_{iru} \stackrel{b}{=} -\sum_j d_{ij}p_{jru}\,dt - (h_i - 1)(dz_t - dt)p_{iru}. \tag{4.34}$$

The derivation of the last three equations also proceeds in a similar fashion to the earlier derivations.

5. REVERSE LIKELIHOOD FUNCTION

Recall that one of the two basic smoothing formulas derived in Section 3 involved $p(Z_{t+}|x_t)$ for discrete time and $E_X[L_{t,T}|Z_{t+}, x_t = x]$ for continuous time. In this section, we shall present differential equations for the evolution of this latter quantity, which is in effect a likelihood function.

For state models of type D and observation models of type G, an equation for this quantity has been derived in [7, 8]. We begin with a different derivation for this case, the form of which permits easy extension to state models F and observation models C. Define

$$\phi_t = \int_t^T h'(x_s, s)R^{-1}(s)\,dz_s - \tfrac{1}{2}\int_t^T \|h(x_s, s)\|^2_{R^{-1}(s)}\,ds \tag{5.1}$$

so that

$$d\phi_t \stackrel{f}{=} -h'(x_t, t)R^{-1}(t)\,dz_t + \tfrac{1}{2}\|h(x_t, t)\|^2_{R^{-1}(t)}\,dt. \tag{5.2}$$

where this is a *forward* Ito equation. By the standard Ito rule applied to the combined process $[\phi_t, x_t']'$, it is straightforward to derive

$$dL_{t,T} = d(\exp\phi_t) \stackrel{f}{=} -L_{t,T}h'(x_t, t)R^{-1}(t)\,dz_t + L_{t,T}\|h(x_t, t)\|^2_{R^{-1}(t)}\,dt. \tag{5.3}$$

To obtain a backward equation, one uses double the adjustment required to form the Stratonovich equation; the adjustment can be obtained by applying it to the Markov process $[L_{t,T}, x_t']'$. We obtain

$$dL_{t,T} \stackrel{b}{=} -L_{t,T}h'(x_t, t)R^{-1}\,dz_t$$

as a backward equation, so that

$$L_{t,T} \stackrel{b}{=} \int_t^T L_{s,T}h'(x_s, s)R^{-1}(s)\,dz_s + 1. \tag{5.4}$$

Then

$$E[L_{t,T}|Z_{t+},x_t=x] \overset{b}{=} \int_t^T \left\{ \int_{R^n} E[L_{s,T}|Z_{s+},x_s] h'(x_s,s) \times R^{-1}(s) p(x_s|x_t=x)\, dx_s \right\} dz_s + 1. \tag{5.5}$$

Let us define

$$V_t(x) = E[L_{t,T}|Z_{t+},x_t=x]. \tag{5.6}$$

Then (5.5) yields

$$dV_t(x_t) \overset{b}{=} -V_t(x_t)h'(x_t,t)R^{-1}(t)\,dz_t + \int_t^T \left\{ \int_{R^n} V_s(x_s)h'(x_s,s)R^{-1}(s) \frac{\partial}{\partial t} p(x_s|x_t=x)\, dx_s \right\} dz_s. \tag{5.7}$$

Recall the standard backward Kolmogorov Eq., [13]:

$$\frac{\partial}{\partial t} p(x_s|x_t) = -\mathscr{L}^a[p(x_s|x_t)] \quad (s \geqq t). \tag{5.8}$$

where $\mathscr{L}^a$ is the adjoint of $\mathscr{L}$ defined in (4.10), and the x-differentiation in $\mathscr{L}^a$ is with respect to x_t, not x_s. Then, using (5.8) in (5.7) yields

$$dV_t(x_t) \overset{b}{=} -V_t(x_t)h'(x_t,t)R^{-1}(t)\,dz_t - \mathscr{L}^a \int_t^T \left\{ \int_{R^n} V_s(x_s)h'(x_s,s)R^{-1}(s)p(x_s|x_t=x)\,dx_s \right\} dz_s.$$

The last term can be replaced, via (5.5), with $-\mathscr{L}^a[V_t(x)-1]\,dt = -\mathscr{L}^a[V_t(x)]\,dt$. Thus

$$dV_t(x_t) \overset{b}{=} -\mathscr{L}^a[V_t(x_t)]\,dt - V_t(x_t)h'(x_t,t)R^{-1}(t)\,dz_t. \tag{5.9}$$

Eq. (5.9), with the boundary condition $V_T(x_T)=1$, is the equation sought. As mentioned in an earlier footnote, this equation has been derived by Pardoux [16, Eq. (3.15)].

Suppose now that the state model is the finite-state model F, and the observation model is type G. Then, a similar argument to that leading to (5.5) gives

$$E[L_{t,T}|Z_{t+},x_t=i] \stackrel{b}{=} \int_t^T \sum_j E[L_{s,T}|Z_{s+},x_s=j]$$

$$\times h_j'(s)R^{-1}(s)p(x_s=j|x_t=i)\,d\tilde{z}_s+1. \qquad (5.10)$$

Define, in similar vein to (5.6),

$$V_i(t)=E[L_{t,T}|Z_{t+},x_t=i].$$

Observe that, with $\Phi(t,s)$ the transition matrix associated with $\dot{x}=A(t)x$,

$$p(x_s=j|x_t=i)=e_j'\Phi(s,t)e_i$$

and

$$\frac{\partial}{\partial t}p(x_s=j|x_t=i)=\frac{\partial}{\partial t}e_j'\Phi^{-1}(t,s)e_i=-e_j'\Phi^{-1}(t,s)A(t)e_i$$

$$=-e_j'\Phi(s,t)\sum_{k=1}^{n}e_ke_k'A(t)e_i=-\sum_{k=1}^{n}p(x_s=j|x_t=k)a_{ki}(t).$$

Consequently, differentiation of (5.10) yields

$$dV_i(t)\stackrel{b}{=}\int_t^T\sum_j V_j(s)h_j'(s)R^{-1}(s)[-\sum_{k=1}^{n}p(x_s=j|x_t=k)a_{ki}(t)]\,d\tilde{z}_s\,dt$$

$$-V_i(t)h'(t)R^{-1}(t)\,d\tilde{z}_t$$

$$=-V_i(t)h_i'(t)R^{-1}(t)\,d\tilde{z}_t-\sum_k[V_k(t)-1]a_{ki}(t)\,dt$$

$$=-V_i(t)h_i'(t)R^{-1}(t)\,d\tilde{z}_t-\sum_k V_k(t)a_{ki}(t)\,dt. \qquad (5.11)$$

on using the fact that $\sum_k a_{ki}(t)=0$. This is the desired equation of evolution for state model F and observation model G.

Now let us consider a state model of type D, but in conjunction with the counting process observation model C. Define

$$\phi_t \stackrel{f}{=} \int_t^T \ln h(x_s,s)\,dz_s-\int_t^T(h(x_s,s)-1)\,ds. \qquad (5.12)$$

It is easily established, using the appropriate Ito differentiation rule, see e.g. [11], that the forward equation for $L_{t,T}=\exp(\phi_t)$ is

$$dL_{t,T}\stackrel{f}{=}L_{t,T}[h^{-1}-1]\,dz_t+L_{t,T}(h-1)\,dt \qquad (5.13)$$

and the corresponding backward equation is

$$dL_{t,T} \overset{b}{=} -L_{t,T}[h-1]\,dz_t + L_{t,T}(h-1)\,dt \tag{5.14}$$

so that

$$L_{t,T} \overset{b}{=} \int_t^T L_{s,T}[h-1][dz_s - ds] + 1, \tag{5.15}$$

(with a backward Ito integral applying). Then similar arguments applied to (5.15) as were applied above for observation model G lead to

$$dV_t(x_t) \overset{b}{=} -V_t(x_t)(h-1)[dz_t - dt] - \mathcal{L}^a[V_t(x_t)]\,dt;\ V_T(x_T) = 1 \tag{5.16}$$

where

$$V_t(x_t) = E[L_{t,T} \,|\, Z_{t+}, x_t = x]. \tag{5.17}$$

This equation also appears in [16, Section 5].

Finally, state model F and observation model C lead with similar arguments to

$$dV_i(t) \overset{b}{=} -V_i(t)(h_i - 1)[dz_t - dt] - \sum_k V_k(t)a_{ki}(t)\,dt. \tag{5.18}$$

6. SMOOTHING WITHOUT A REVERSE-TIME FILTER OR REVERSE TIME LIKELIHOOD FUNCTION

In [4, 5] the following equation is derived linking the smoothed density $p_s(x_t) = p(x_t \,|\, Z_T)$ and the forward filtered density $p_f(x_t) = p(x_t \,|\, Z_{t-})$ for state model D and observation model G:

$$dp_s = \left[\frac{p_s}{p_f}\mathcal{L}(p_f) - p_f\mathcal{L}^a\left(\frac{p_s}{p_f}\right)\right]dt. \tag{6.1}$$

The idea is that the forward filtering density is first found, and then (6.1) is solved backwards in time from T, using the boundary conditions $p_s(x_T) = p_f(x_T)$. Using the ideas of this paper we shall derive this equation and the modifications of it associated with different combinations of state and observation models. Of course, the various densities are all to be interpreted as functions rather than measures.

Let us recall from (3.9) and the definition of $V_t(x)$ in (5.6) that

$$p(x_t = x \,|\, Z_{T-}) = p(x_t = x \,|\, Z_{t-})\, V_t(x_t) \frac{E_x[L_{0,t} \,|\, Z_{t-}]}{E_x[L_{0,T} \,|\, Z_{T-}]}. \tag{6.2}$$

Accordingly, with p_{su} and p_{fu} denoting unnormalized densities,

$$p_{su}(x_t)=p_{fu}(x_t)V_t(x_t). \tag{6.3}$$

Now in (5.9), we found a backward equation for $V_t(x_t)$. The corresponding forward equation is

$$dV_t(x_t)\overset{f}{=}-\mathscr{L}^a[V_t(x)]\,dt-V_t(x_t)h'(x_t,t)R^{-1}(t)\,dz_t+V_t(x_t)\|h(x_t,t)\|^2_{R^{-1}(t)}\,dt. \tag{6.4}$$

Combining this with the forward Ito equation for $p_{fu}(x_t)$, see (4.8), yields

$$\begin{aligned}dp_{su}(x_t)&\overset{f}{=}d[p_{fu}(x_t)V_t(x_t)]=d[p_{fu}(x_t)]V_t(x_t)+p_{fu}(x_t)\,dV_t(x_t)+dp_{fu}(x_t)\,dV_t(x_t)\\&=V_t(x_t)\mathscr{L}[p_{fu}(x_t)]\,dt+V_t(x_t)h'R^{-1}(t)\,dz_tp_{fu}(x_t)\\&\quad-p_{fu}(x_t)\mathscr{L}^a[V_t(x)]\,dt-p_{fu}(x_t)V_t(x_t)h'R^{-1}(t)\,dz_t\\&\quad+p_{fu}(x_t)V_t(x_t)\|h\|^2_{R^{-1}(t)}\,dt-V_t(x_t)p_{fu}(x_t)\|h\|^2_{R^{-1}(t)}\,dt\\&=V_t(x_t)\mathscr{L}[p_{fu}(x_t)]\,dt-p_{fu}(x_t)\mathscr{L}^a[V_t(x_t)]\,dt\\&=\left\{\frac{p_{su}(x_t)}{p_{fu}(x_t)}\mathscr{L}[p_{fu}(x_t)]-p_{fu}(x_t)\mathscr{L}^a\left[\frac{p_{su}(x_t)}{p_{fu}(x_t)}\right]\right\}dt.\end{aligned}$$

Now it is immediate that the $\mathscr{L}$ and $\mathscr{L}^a$ operators commute with the normalizing quantity associated with p_{fu}. Accordingly,

$$dp_{su}(x_t)=\left\{\frac{p_{su}(x_t)}{p_f(x_t)}\mathscr{L}[p_f(x_t)]-p_f(x_t)\mathscr{L}^a\left[\frac{p_{su}(x_t)}{p_f(x_t)}\right]\right\}dt. \tag{6.5}$$

It is now easy to obtain the normalized equation, since the integral of the coefficient of dt over x_t space is zero, indicating that the normalizing quantity

$$n_s(t)=\int_{R^n}p_{su}(x_t)\,dx_t,$$

satisfies $dn_s(t)=0$, i.e. n_s is constant; (6.1) is then immediate.

We turn now to the finite state model F with observation model G. Using (5.11), we have the following forward equation for $V_i(t)$

$=E[L_{i,T}|Z_{t+},x_t=i]$:

$$dV_i(t)\stackrel{\mathcal{L}}{=}-V_i(t)h_i'(t)R^{-1}(t)\,dz_t-\sum_k V_k(t)a_{ki}(t)\,dt+h_i'(t)R^{-1}(t)h_i(t)V_i(t)\,dt. \tag{6.6}$$

Accordingly, with $p_{isu}(t)=p_{ifu}(t)V_i(t)$, (4.25) and (6.6) yield

$$\begin{aligned}dp_{isu}(t)&\stackrel{\mathcal{L}}{=}V_i(t)\,dp_{ifu}(t)+p_{ifu}(t)\,dV_i(t)-h_i'(t)R^{-1}(t)h_i(t)V_i(t)p_{ifu}(t)\,dt\\&=V_i(t)\sum_j a_{ij}(t)p_{jfu}(t)\,dt-\sum_k V_k(t)a_{ki}(t)p_{ifu}(t)\,dt\\&=\frac{p_{isu}(t)}{p_{ifu}(t)}\left[\sum_j a_{ij}(t)p_{jfu}(t)\right]dt-p_{ifu}(t)\left[\sum_k\frac{p_{ksu}(t)}{p_{kfu}(t)}a_{ki}(t)\right]dt.\end{aligned}$$

Since all unnormalized filtered probabilities occur as ratios, we have

$$dp_{isu}(t)=\frac{p_{isu}(t)}{p_{if}(t)}\left[\sum_j a_{ij}(t)p_{jf}(t)\right]dt-p_{if}(t)\left[\sum_j\frac{p_{jsu}(t)}{p_{jf}(t)}a_{ji}(t)\right]dt.$$

Now observe that this equation implies easily that

$$\sum_i dp_{isu}(t)=0.$$

Thus the normalization on $p_{isu}(t)$ is independent of time. Accordingly,

$$dp_{is}(t)=\frac{p_{is}(t)}{p_{if}(t)}\left[\sum_j a_{ij}(t)p_{jf}(t)\right]dt-p_{if}(t)\left[\sum_j\frac{p_{js}(t)}{p_{jf}(t)}a_{ji}(t)\right]dt. \tag{6.7}$$

Of course the initialization for the equation is provided by $p_{is}(T)=p_{if}(T)$ and the equation is solved backwards in time. This equation has been obtained by different arguments in [6].

It is interesting to note that we can rewrite (6.7), with p_s denoting the vector of p_{is},

$$dp_s=A^s p_s\,dt, \tag{6.8}$$

where $-A^s$ is a stochastic matrix and A^s is defined by

$$a_{ii}^s=\sum_{\substack{j=1\\j\neq i}}^{n}\frac{a_{ij}p_{jf}}{p_{if}} \tag{6.9a}$$

$$a_{ij}^{*}=-\frac{p_{if}a_{ji}}{p_{jf}}. \tag{6.9b}$$

Now consider a state model D with observation process C. Eq. (6.3) is still relevant. The forward equation for $V_t(x_t)$ is, from (5.16),

$$dV_t(x_t)\overset{f}{=}V_t(x_t)[h^{-1}-1]\,dz_t+V_t(x_t)[h-1]\,dt-\mathscr{L}^{a}[V_t(x_t)]\,dt. \tag{6.10}$$

[The argument is just like that for connecting (5.13) and (5.14).] Then combining this with (4.30), for $p_{fu}(x_t)$, we obtain

$$\begin{aligned} dp_{su}(x_t)&\overset{f}{=}V_t(x_t)\,dp_{fu}(x_t)+p_{fu}(x_t)+dV_t(x_t)\,dp_{fu}(x_t)\\ &=V_t(x_t)\mathscr{L}(p_{fu})\,dt+V_t(x_t)(h-1)(dz-dt)p_{fu}\\ &\quad+p_{fu}V_t(x_t)(h^{-1}-1)\,dz+p_{fu}V_t(x_t)(h-1)\,dt\\ &\quad-p_{fu}\mathscr{L}^{a}[V_t(x_t)]\,dt+p_{fu}V_t(x_t)(h-1)(h^{-1}-1)\,dz\\ &=V_t(x_t)\mathscr{L}(p_{fu}(x_t))\,dt-p_{fu}(x_t)\mathscr{L}^{a}(V_t(x_t))\,dt, \end{aligned} \tag{6.11}$$

and from this we immediately obtain the same Eq. (6.5) as was found to hold for observation model G. In other words, Eq. (6.5) holds for either observation model G or P in conjunction with state mode D.

In the same vein, similar calculations show that Eq. (6.7) holds not only for state model F and observations G as shown above, but also for state model F and observations C: one uses (4.25) for p_{ifu} and a forward time version of (5.18) to obtain (6.7).

7. FINITE DIMENSIONAL SMOOTHERS

We have already indicated in the last two sections one class of finite dimensional smoothers—those associated with a finite state model (type F). The forward filter, reverse filter, reverse likelihood function, and *a priori* probability computations are all finite-dimensional. In this section, we note some other possibilities of finite-dimensional filters.

Linear-Gaussian problems

This case is well known, and the three distinct approaches advanced here all have their specializations which have appeared in the literature. We comment, however, that the two Eqs. (3.1) and (3.2) do not seem to have been recognized as such in the linear-Gaussian case. Let m_u, m_f, m_r, m_s denote the unfiltered, forward-filtered, reverse-filtered, and smoothed means. Let Σ_u, Σ_f, Σ_r, Σ_s denote the corresponding (error) covariances. Then it is a trivial observation from (3.1), knowing the densities are Gaussian, to obtain

$$\Sigma_s^{-1}=\Sigma_f^{-1}+\Sigma_r^{-1}-\Sigma_u^{-1} \tag{7.1}$$

and

$$m_s=\Sigma_s[\Sigma_f^{-1}m_f+\Sigma_r^{-1}m_r-\Sigma_u^{-1}m_u]. \tag{7.2}$$

These equations appear to have first been derived in [22–23]. These references also include equations for the update of the mean m_r and covariance Σ_r associated with the reverse filter, and thus capture the approach of Section 4.

As described also in [22–23], the Fraser–Potter smoothing formula [23,24] can be thought of as combining a forward filtered estimate and a reverse maximum likelihood estimate. Thus (3.2), or (3.9), is the key formula applicable here. One can show in the linear Gaussian case that $E_X[L_{t,T}|Z_{t+}, x_t=x]$ is proportional to

$$\exp-\tfrac{1}{2}[(x-m_L)'\Sigma_L^{-1}(x-m_L)]$$

where

$$\Sigma_L^{-1}=\Sigma_r^{-1}-\Sigma_u^{-1} \tag{7.3}$$

and

$$\Sigma_L^{-1}m_L=\Sigma_r^{-1}m_r-\Sigma_u^{-1}m_u. \tag{7.4}$$

Again, formulas are given in the references for the update of M_L and Σ_L, and so the ideas of Section 5 are captured.

The formula (6.1) also has its counterpart, see for example [25, Section 7.5]; equations for the smoothed mean and error covariance are given in terms of a prior information and the forward filtered mean and error covariance, with no use of the measurement process, and the equations are solved backwards in time from T.

Problems with space-time point-process observations

In [26], an extension of the standard linear Gaussian problem is considered. This consists in providing an additional observation process which is a space-time point process defined on $[0, \infty)\times R^m$ as follows.

Each point occurrence is identified by a temporal coordinate t $[0, \infty)$ and a spatial coordinate r R^m. Let τ, A be Borel sets in $[0, \infty)$, R^m and denote by $N(\tau \times A)$ the number of points occurring in $\tau \times A$, with $N_t \triangleq N([0, t) \times R^m)$ being the number of points occurring before t at any location; N is taken to be a doubly stochastic Poisson counting process with intensity μ, where the stochastic processes μ and N are independent of the state process and the linear-in-the-state observation process, and μ_t is a.s. positive. Given that N has a jump at t, the spatial location r of the point is an m-dimensional Gaussian random vector with mean $H(t)x_t$ and known positive definite covariance $S(t)$, where $H(\cdot)$ is known. Given N_s and x_s for $s \geqq 0$, the spatial locations are independent random vectors that are independent of all other entities. Thus the space-time process can be thought of as having an intensity

$$\lambda_t(r, x_t, \mu_t) = \mu_t \gamma_t(r, x_t) \tag{7.5}$$

that separates into the product of a temporal component μ_t that underlies the Poisson counting process N and a spatial component

$$\gamma_t(r, x_t) = (2\pi)^{-m/2}[\det S(t)]^{-1/2} \exp\{-\tfrac{1}{2}\|r - H(t)x_t\|^2_{S^{-1}(t)}\}. \tag{7.6}$$

The main result of [26] is that the conditional density of x_t given the two observation processes up until time t is (conditionally) Gaussian; both the conditional mean and the conditional covariance satisfy nonlinear but finite-dimensional stochastic differential equations driven by the observation processes.

Now, as noted in Section 4, a reverse time model can be found in general for the linear-Gaussian state process and the linear-Gaussian observation process. The reverse models are linear and Gaussian. The point process has an unchanged reverse model. It follows that the reverse filtered density is also conditionally Gaussian (and so, by the fundamental formula (3.1), is the smoothed density). Further, the reverse conditional mean and covariance can be found in the same manner as the corresponding forward quantities.

8. CONCLUSIONS

The ideas of the paper fall into four divisions. First are the pair of equations relating the smoothed density (or probability, as the case may be) to the forward filtered, reverse filtered, and unfiltered densities or, alternatively, to the forward filtered density and a reverse-time likelihood ratio. The second division is the development of reverse-time filtered equations, using recent results on the construction of reverse time processes. Third is the development of equations for the evolution of the reverse-time likelihood ratio, and last is the development of equations that relate smoothed and filtered densities and are undriven by the measurement.

Though we have not discussed the point, it is clear that there are reverse-time equivalents of the last two ideas, which would involve a forward-evolving likelihood ratio to be used in conjunction with a reverse filtered density to get a smoothed density, and an equation relating smoothed and reverse filtered densities, undriven by the measurements.

The results have all been presented with certain independence assumptions between system state and measurement noise; for example, with state model D and measurement model G, the processes w_t and v_t have been assumed independent. Assumptions like this can doubtlessly be relaxed.

It would also be possible to develop parallels of a number of the ideas in discrete time. A quite different development could lie in the generation of bounds on the smoothed errors, such as can be done in the filtering case. Indeed, the key formula relating smoothed, forward filtered, reverse filtered, and unfiltered quantities may be of great help here.

References

[1] P. Whittle, *Reversibility and Acyclicity*, pp. 217–224, in *Perspectives in Probability and Structures*, ed. J. Gani. Academic Press, London, 1975.

[2] B. D. O. Anderson and T. Kailath, Forwards and backwards models for finite-state Markov processes, *Advances in Applied Probability* 11 (1979), 118–133.

[3] B. D. O. Anderson, Reverse-time diffusion equation models, *Stochastic Processes and their Applications* 12 (1982), 313–326.

[4] C. T. Leondes, J. B. Peller and E. B. Stear, Nonlinear smoothing theory, *IEEE Trans. Syst. Science Cyber.* SSC-6 (1970), 63–71.

[5] B. D. O. Anderson, Fixed interval smoothing for nonlinear continuous-time systems, *Information and Control* 20, No. 3, (April 1972), 294–300.

[6] R. S. Liptser and A. N. Shiryaev, *Statistics of Random Processes I*, Springer-Verlag, Berlin, 1977.

[7] E. Pardoux, Stochastic partial differential equations and filtering of diffusion processes, *Stochastics* 3 (1979), 127–167.

[8] E. Pardoux, Equations du filtrage nonlinéaire, de la prédiction et du lissage, *Stochastics* 6 (1982), 193–231.

[9] P. Frost and T. Kailath, An innovative approach to least-square estimation—Part III: Nonlinear estimation in white Gaussian noise, *IEEE Trans. Auto. Control* AC-16 (June 1971), 217–226.

[10] D. L. Snyder, Smoothing for doubly stochastic Poisson processes, *IEEE Trans. Info. Theory* IT-18 (1972), 558–562.

[11] K. P. Dunn and I. B. Rhodes, A generalized representation theorem with applications to estimation and hypothesis testing, *Proc. Eleventh Allerton Conference on Circuit and System Theory*, (1973), 304–314.

[12] D. Clements and B. D. O. Anderson, A nonlinear fixed-lag smoother for finite-state Markov processes, *IEEE Trans. Info. Theory* IT-21 (1975), 446–452.

[13] E. Wong, *Stochastic Processes in Information and Dynamical Systems*, McGraw Hill, New York, 1971.

[14] J. H. VanSchuppen, Stochastic filtering theory: A discussion of concepts, methods, and results, in *Stochastic Control Theory and Stochastic Differential Systems*, M. Kolhmann and W. Vogel, eds., New York, Springer-Verlag, 1979.

[15] R. S. Bucy and P. D. Joseph, *Filtering for Stochastic Processes with Applications to Guidance*, Wiley-Interscience, New York, 1968.

[16] E. Pardoux, Non-linear filtering, prediction and smoothing, in M. Hazewinkel and J. C. Willems (eds.), *Stochastic Systems: The Mathematics of Filtering and Identification and Applications*, Reidel, Dordrecht, 1981.

[17] A. Lindquist and G. Picci, On the stochastic realization process problem, *SIAM J. Control Opt.* 17 (May 1979), 365–389.

[18] M. Zakai, On the optimal filtering of diffusion processes, *Zeitschrift fur Wahrscheinlichkeitstheorie der Verwandten Gebiete*, 11 (1969), 230–243.

[19] M. H. A. Davis and S. I. Marcus, An introduction to nonlinear filtering, in M. Hazelwinkel and J. C. Willems, eds., *Stochastic Systems; The Mathematics of Filtering and Identification*, NATO Advanced Study Institute Series, Reidel, Dordrecht, 1980.

[20] W. M. Wonham, Some applications of stochastic differential equations to optimal nonlinear filtering, *SIAM J. Control* 2 (1965), 347–369.

[21] D. L. Snyder, *Random Point Processes*, Wiley, New York, 1975.

[22] J. E. Wall, A. S. Willsky and N. R. Sandell, On the fixed-interval smoothing problem, *IEEE Trans. Auto Control* to appear.

[23] D. C. Fraser, *A New Technique for the Optimal Smoothing of Data*, Sc.D. dissertation, Massachusetts Institute of Technology, Cambridge, MA, 1967.

[24] D. C. Fraser and J. E. Potter, The optimum linear smoother as a combination of two optimum linear filters, *IEEE Trans. Auto. Control* AC-14 (August 1969), 387–390.

[25] J. S. Meditch, *Stochastic Optimal Linear Estimation and Control*, McGraw Hill, New York, p. 261, 1969.

[26] I. B. Rhodes and D. L. Snyder, Estimation and control performance for space-time point-process observations, *IEEE Trans. Auto. Control* AC-22 (June 1977), 338–346.

APPENDIX A—EXTENDED DERIVATION OF EQ. (4.10)

We start with the backward Eq. (4.5), which can be written as

$$x_t - x_{t-dt} \overset{b}{=} f^r(x_t, t)\,dt + g(x_t, t)[v_t^r - v_{t-dt}^r].$$

To prevent confusion below, let us rewrite this as

$$x_t - x_{t-\Delta} \overset{b}{=} f^r(x_t, t)\Delta + g(x_t, t)[v_t^r - v_{t-\Delta}^r]$$

where it is understood that Δ is a vanishingly small positive quantity.

Now let

$$\bar{x}_t = x_{T-t}$$

$$\bar{v}^r_t = v^r_{T-t}$$

$$\bar{f}^r(\bar{x}_t, t) = f^r(\bar{x}_t, T-t) = f^r(x_{T-t}, T-t)$$

$$\bar{g}^r(\bar{x}_t, t) = g^r(\bar{x}_t, T-t) = g^r(x_{T-t}, T-t).$$

Then

$$\begin{aligned} \bar{x}_{t+\Delta} - \bar{x}_t &= x_{T-(t+\Delta)} - x_{T-t} \\ &= -[x_{(T-t)} - x_{(T-t)-\Delta}] \\ &\overset{b}{=} -f^r(x_{T-t}, T-t)\Delta - g(x_{T-t}, T-t)[v^r_{T-t} - v^r_{T-t-\Delta}] \\ &= -\bar{f}^r(\bar{x}_t, t)\Delta - \bar{g}(\bar{x}_t, t)[\bar{v}^r_t - \bar{v}^r_{t+\Delta}] \\ &\overset{\mathcal{L}}{=} -\bar{f}^r(\bar{x}_t, t)\Delta + \bar{g}(\bar{x}_t, t)[\bar{v}^r_{t+\Delta} - \bar{v}^r_t]. \end{aligned} \quad \text{(A.1)}$$

This of course is a forward Ito equation, since Δ is still a vanishingly small *positive* quantity.

To get an associated observation equation, first of all write (4.7b) as

$$z_t - z_{t-\Delta} \overset{b}{=} h(x_t, t)\Delta + M(t)[w_t - w_{t-\Delta}].$$

Set

$$\bar{z}_t = z_{T-t} \quad \bar{w}_t = w_{T-t}$$

$$\bar{h}(\bar{x}_t, t) = h(\bar{x}_t, T-t) = h(x_{T-t}, T-t)$$

$$\bar{M}(t) = M(T-t).$$

Then

$$\begin{aligned} \bar{z}_{t+\Delta} - \bar{z}_t &= -[z_{(T-t)} - z_{(T-t)-\Delta}] \\ &= -h(x_{T-t}, T-t)\Delta - M(T-t)[w_{T-t} - w_{T-t-\Delta}] \\ &\overset{\mathcal{L}}{=} -\bar{h}(\bar{x}_t, t)\Delta + \bar{M}(t)[\bar{w}_{t+\Delta} - \bar{w}_t]. \end{aligned} \quad \text{(A.2)}$$

Now, (A.1) and (A.2) are both forward equations to which the forward filtering equation can be applied. Thus, see (4.8)

$$p_u(\tilde{x}_{t+\Delta}=\alpha\,|\,\tilde{z}_s, 0\leq s\leq t+\Delta)-p_u(\tilde{x}_t=\alpha\,|\,\tilde{z}_s, 0\leq s\leq t)$$

$$\stackrel{\mathcal{L}}{=}\mathscr{L}[p_u(\tilde{x}_t=\alpha\,|\,Z_t)]\Delta-p_u(\tilde{x}_t=\alpha\,|\,Z_t)\hat{h}_t' R^{-1}(\tilde{z}_{t+\Delta}-\tilde{z}_t). \tag{A.3}$$

We must now convert this to a backward equation. We have at once:

$$p_u(\tilde{x}_{T-t+\Delta}=\alpha\,|\,\tilde{z}_s, 0\leq s\leq T-t+\Delta)-p_u(\tilde{x}_{T-t}=\alpha\,|\,\tilde{z}_s, 0\leq s\leq T-t)$$

$$\stackrel{\mathcal{L}}{=}\mathscr{L}[p_u(\tilde{x}_{T-t}=\alpha\,|\,Z_{T-t})]\Delta$$

$$-p_u(\tilde{x}_{T-t}=\alpha\,|\,Z_{T-t})\hat{h}_{T-t}' R^{-1}(\tilde{z}_{T-t+\Delta}-\tilde{z}_{T-t}). \tag{A.4}$$

Here

$$\mathscr{L}[p_u(\tilde{x}_{T-t}=\alpha\,|\,Z_{T-t})]=-\sum_i\frac{\partial}{\partial\alpha^i}[-f^{ri}(\alpha, T-t)p_u(\tilde{x}_{T-t}=\alpha\,|\,Z_{T-t})]$$

$$+\tfrac{1}{2}\sum_{i,j}\frac{\partial^2}{\partial\alpha^i\,\partial\alpha^j}\{[g(\alpha, T-t)g'(\alpha, T-t)]^{ij}\,p_u(\tilde{x}_{T-t}=\alpha\,|\,Z_{T-t})\}$$

$$=-\sum_i\frac{\partial}{\partial\alpha^i}[-f^{ri}(\alpha, t)p_u(x_t=\alpha\,|\,z_s, t\leq s\leq T]+\tfrac{1}{2}\sum_{i,j}\frac{\partial^2}{\partial\alpha^i\,\partial\alpha^j}$$

$$\times\{[g(\alpha, t)g'(\alpha, t)]^{ij}p_u(x_t=\alpha\,|\,z_s, t\leq s\leq T)\}=\mathscr{L}_r[p_{ru}]. \tag{A.5}$$

where $\mathscr{L}_r$ is defined like $\mathscr{L}$, save that $-f^r$ replaces f, and p_{ru} is shorthand for $p_u(x_t=\alpha\,|\,z_s, t\leq s\leq T)$.

Using (A.5) in (A.4), we have

$$p_u(x_{t-\Delta}=\alpha\,|\,z_s, t-\Delta\leq s\leq T)-p_u(x_t=\alpha\,|\,z_s, t\leq s\leq T)$$

$$\stackrel{b}{=}\mathscr{L}_r[p_{ru}]\Delta-p_{ru}\hat{h}' R^{-1}(z_{t-\Delta}-z_t)$$

or

$$dp_{ru}\stackrel{b}{=}-\mathscr{L}_r[p_{ru}]\,dt-p_{ru}\hat{h}' R^{-1}\,dz$$

as desired.