Robust Continuous-Time Smoothers Without Two-Sided Stochastic Integrals

Vikram Krishnamurthy, *Senior Member, IEEE,* and Robert Elliott

***Abstract*—We consider the problem of fixed-interval smoothing of a continuous-time partially observed nonlinear stochastic dynamical system. Existing results for such smoothers require the use of two-sided stochastic calculus. The main contribution of this paper is to present a robust formulation of the smoothing equations. Under this robust formulation, the smoothing equations are nonstochastic parabolic partial differential equations (with random coefficients) and, hence, the technical machinery associated with two sided stochastic calculus is not required. Furthermore, the robust smoothed state estimates are locally Lipschitz in the observations, which is useful for numerical simulation. As examples, finite dimensional robust versions of the Benes and hidden Markov model smoothers and smoothers for piecewise linear dynamics are derived; these finite-dimensional smoothers do not involve stochastic integrals.**

***Index Terms*—Continuous time, hidden Markov models (HMMs), maximum likelihood estimation, nonlinear smoothing, piecewise linear models, stochastic differential equations.**

I. Introduction

FILTERING is another word for conditional mean estimation of the state at time t of a given dynamical stochastic system, based on the available incomplete information (observations) until the same time t. Fixed-interval smoothing refers to the problem when given a trajectory of observations up to some fixed time $T > 0$, one wishes to compute the conditional mean estimate of the underlying state at times t in the interval $0 \leq t \leq T$.

For continuous-time dynamical stochastic systems, the filtered state density can be expressed as a stochastic partial differential equation called the Duncan–Mortenson–Zakai (DMZ) equation [2]. Derivation of the fixed-interval smoothed state density is mathematically more formidable as it requires the use of two sided stochastic calculus [19].

In this paper we derive *robust* filters and smoothers for the state of a continuous-time stochastic dynamical system by using a gauge transformation, see for example [6], [8]. By robust we mean that the resulting filtering and smoothing equations are locally Lipschitz continuous in the observations, i.e., the equations depend continuously on the observation path. Indeed, the equations turn out to be nonstochastic parabolic partial differential equations whose coefficients depend on the observations. Apart from not requiring the intricacies of two-sided stochastic calculus, these robust equations are useful from a practical point of view; their numerical solution via time discretization can be performed without worrying about the Ito terms.

The idea of robust filtering, i.e., re-expressing the stochastic differential equation as nonstochastic differential equation with random coefficients has been used extensively in the context of nonlinear filtering; see, for example, [6], [16], [8], [18], or [2, Ch. 4]. More recently, in [14], versions of these robust filters, probabilistic interpretations and implicit and explicit discretization schemes were developed for continuous-time hidden Markov models (HMMs).

The contributions of this paper are as follows.

1) It is shown in Section III that the smoothed state estimate can be computed via robust forward and backward filters. Each of these filters involve nonstochastic parabolic partial differential equations.
2) Robust fixed interval smoothed estimates of functionals of the state of the system are derived in Section III. Again, the equations involve nonstochastic integrals. These robust smoothers can be used in maximum likelihood parameter estimation via the expectation maximization (EM) algorithm. The EM algorithm (see Section II-B) is a widely used numerical method for computing the maximum likelihood parameter estimate for partially observed stochastic dynamical systems; see, for example, [23], [4], and [14]. Unlike this paper, in [14] and [9], two-sided stochastic calculus involving Skorohod and generalized Stratonovich integrals are used to derive smoothers for computing estimates of the functionals required in the EM algorithm for HMMs and linear Gaussian state space models, respectively.
3) As examples of the robust smoothers for the state and functionals of the state, we present state and maximum likelihood parameter estimation for three classes of stochastic dynamical systems: 1) Benes type nonlinear dynamical systems with non Gaussian initial conditions (see Section IV), 2) HMM (see Section V), and 3) systems with piecewise linear dynamics (see Section VI).

Instead of using fixed-interval smoothing for cases 1) and 2), finite-dimensional *filters* have been derived in [12], [13], and [14] to compute estimates of the functionals required in the EM algorithm. However, the computational complexity of these filters are $O(m^4)$ for some of the functionals (e.g., for the number

Manuscript received July 18, 2001; revised March 9, 2002, April 18, 2002, and July 21, 2002. Recommended by Associate Editor T. Parisini. This work was supported by the Australian Research Council and the NSERC, Canada.

V. Krishnamurthy is with the Department of Electrical and Computer Engineering, University of British Columbia, Vancouver, BC V6T 1Z4, Canada (e-mail: vikramk@ece.ubc.ca), and also with the University of Melbourne, Melboune, Australia. Portions of this work were conducted while the author was on sabbatical at the Department of Signals Sensors and Systems, Royal Insitute of Technology, 100 44 Stockholm, Sweden.

R. Elliott is with the Haskayne School of Business, University of Calgary, Alberta, AB T2N 1N4, Canada (e-mail: relliott@ucalgary.ca). Portions of this work were completed while the author was visiting the Department of Applied Mathematics, University of Adelaide, Australia.

Digital Object Identifier 10.1109/TAC.2002.804481

0018-9286/02$17.00 © 2002 IEEE

of jumps in an HMM) at each time instant where m denotes the state dimension. In comparison, computing estimates of these functionals via fixed-interval smoothers involves a complexity of $O(m^2)$ but requires storage memory of $O(T)$ where T is the length of the observation data sequence. Approximate filtering for piecewise linear systems via a bank on Kalman filters is presented in [20] and [21]. We extend these results to derive robust smoothers for the state and functionals of the state required in the EM algorithm, see Section VI for details.

II. Model and Problem Formulation

A. Signal Model and Objectives

Consider the following continuous-time partially observed nonlinear stochastic dynamical system defined on the measurable space $(\Omega, \mathcal{F})$. Let $\{P_\theta : \theta \in \Theta\}$, where Θ denotes a compact subset of $\mathbb{R}^p$, denote a family of parametrized probability measures. Under P_θ, the state $\{x_t\}$ taking values in $\mathbb{R}^m$, and the observation process $\{y_t\}$ taking values in $\mathbb{R}^n, t \geq 0$ are described by

$$dx_t = f_\theta(x_t,t)\,dt + \sigma_\theta(x_t,t)dw_t, \qquad x_0 \sim \pi_0^\theta(\cdot) \quad (1)$$
$$dy_t = h_\theta(x_t,t)\,dt + d\nu_t, \qquad y_0 = 0 \in \mathbb{R}^n. \quad (2)$$

Let $T > 0$ denote a fixed real number. For $t \in [0,T]$, define the right-continuous filtrations $\{\mathcal{F}_t\}, \{\mathcal{G}_t\}$, and $\{\mathcal{Y}_t\}$ with

$$\begin{aligned}\mathcal{F}_t &= \sigma(x_s,\ 0 \leq s \leq t)\\ \mathcal{G}_t &= \sigma\{x_s, y_s : 0 \leq s \leq t\}\\ \mathcal{Y}_t &= \sigma\{y_s : 0 \leq s \leq t\}.\end{aligned} \quad (3)$$

In (1) and (2), $\{w_t\}$ and $\{\nu_t\}$ are independent standard Brownian motions. Further, $\{w_t\}$ and $\{\nu_t\}$ are independent of x_0. (In Section V, we will consider the HMM case where w_t is a $\mathcal{F}_t$ measurable finite state zero mean martingale process.

We make the following standard assumptions [2, pp. 114] for all $\theta \in \Theta$.

A1) $f_\theta : \mathbb{R}^m \times [0,T] \to \mathbb{R}^m$ and $h_\theta : \mathbb{R}^m \times [0,T] \to \mathbb{R}^n$ are bounded Borel measurable functions

A2) $\sigma_\theta : \mathbb{R}^m \times [0,T] \to \mathbb{R}^{m\times r}$ is continuous and bounded such that $Q = \sigma_\theta\sigma_\theta'$ is a uniformly positive definite $m \times m$ matrix, i.e., $Q > \alpha I$ for some real $\alpha > 0$. This ensures that the backward operator L (defined in (15)) is uniformly elliptic. This condition can be somewhat relaxed with Q^{-1} replaced by its pseudoinverse $Q^{\#}$, see Section IV-B.

A3) f_θ and σ_θ are Lipschitz in x, i.e.,

$$\begin{aligned}|f_\theta(x_1,t) - f_\theta(x_2,t)| &\leq k_\theta|x_1 - x_2|\\ \|\sigma_\theta(x_1,t) - \sigma_\theta(x_2,t)\| &\leq k_\theta|x_1 - x_2|.\end{aligned}$$

A4) The probability measures on $\mathbb{R}^m$ with densities $(\pi_0^\theta(\cdot) : \theta \in \Theta)$ with respect to the Lebesgue measure are mutually absolutely continuous. We assume $\int_{\mathbb{R}^m} |x|^2\pi_0^\theta(x)dx < \infty$ and $\pi_0^\theta \in L^2(\mathbb{R}^m)$.

Then there exists a unique strong solution $\{x_t, 0 \leq t \leq T\} \in C(\mathbb{R}^m \times [0,T])$ to the state (1) (where $C(\mathbb{R}^m \times [0,T])$ denotes the space of $\mathbb{R}^m$-valued continuous functions on $[0,T]$). Also $\{y_t, 0 \leq t \leq T\} \in C(\mathbb{R}^n \times [0,T])$ endowed with the sup-norm, i.e., $\|y\| = \sup_{0\leq t\leq T} |y_t|$.

We also assume throughout that for all $\theta \in \Theta$, A5) holds.

A5) f_θ, σ_θ and h_θ are continuously differentiable with respect to the parameter θ. The derivatives $\partial f_\theta/\partial\theta$ and $\partial h_\theta/\partial\theta$ are measurable and bounded functions.

To introduce the gauge transformation we shall assume A6).

A6) $h_\theta(x,s)$ has continuous and bounded first and second derivatives with respect to x and bounded first derivative with respect to t. The differentiability w.r.t. x is not required in the finite-state Markov case considered in Section V.

In Section VI, the assumption of continuous first and second derivatives is relaxed. In particular Section VI assumes that $h_\theta(x)$ is piecewise linear and continuous in x. Tanaka's formula, which is roughly speaking an extension of Ito's formula to the nondifferentiable case, will be used.

Objectives : In this paper, we will derive robust filtering and smoothing equations. By robust, we mean that the solution to the resulting equations are locally Lipschitz continuous in the observation y. As mentioned in Section I, this is a useful property from an implementation point of view. The aim of this paper is threefold.

i) Derive robust fixed-interval smoothers for $\mathbf{E}\{x_t|\mathcal{Y}_T\}$ that do not involve stochastic integrals.

ii) Derive robust fixed interval smoothers for functionals of the form

$$H_t = H_0 + \int_0^t \alpha(x_s,y_s)ds + \int_0^t \beta'(x_s,y_s)dx_s + \int_0^t \gamma'(x_s)dy_s \quad (4)$$

where $\alpha : \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}, \beta : \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}^m, \gamma : \mathbb{R}^m \to \mathbb{R}^n$ are Borel measurable and bounded functions. β is assumed once differentiable in x. Our aim is to compute the fixed-interval smoothed estimate $\mathbf{E}\{H_t|\mathcal{Y}_T\}$, $t \in [0,T]$ using robust forward and backward filters. These smoothed estimates are required in computing the maximum likelihood parameter estimate via the EM algorithm; see Section II-B. The same problem is considered in [4] where two-sided stochastic calculus was used to compute $\mathbf{E}\{H_t|\mathcal{Y}_T\}$.

To motivate the robust smoothers presented below, consider computing the smoothed estimate of the last term in (4). One would have liked to have interchanged the conditional expectation and the integral. However, the resulting expression

$$\text{"} \int_0^t \mathbf{E}\{\gamma'(x_s)|\mathcal{Y}_T\}dy_s \text{"}$$

is not an Ito integral since the integrand is not adapted to the filtration $\{\mathcal{Y}_t : 0 \leq t \leq T\}$. In [4], it is shown that the above integral can be interpreted as a Skorohod integral and requires the use of two-sided stochastic calculus. The aforementioned integral is interpreted in [9] as a generalized Stratonovich integral.

In Section III, it will be demonstrated that by expressing the filters in robust form, the smoothed estimate $\mathbf{E}\{H_t|\mathcal{Y}_T\}$ can be computed using ordinary (nonstochastic) integration. Thus, two-sided stochastic calculus is not required. For example, Theorem 3.4 of this paper shows that

$$\mathbf{E}\left\{\int_0^t \gamma'(x_s)dy_s|\mathcal{Y}_T\right\} = \frac{1}{K}\left[y_t\int_{\mathbb{R}^m}\gamma'(x)\bar{q}_t(x)\bar{v}_t(x)dx - \int_0^t y_s\int_{\mathbb{R}^m}\gamma'(x)\frac{d}{ds}\left[\bar{q}_s(x)\bar{v}_s(x)\right]dx\right].$$

Here, K is a normalization factor and $\bar{q}_t, \bar{v}_t$ [defined in (17) and (19)] are robust forward and backward filtered densities that evolve according to nonstochastic partial differential equations.

iii) Using the robust smoothers in Step ii), we will address the problem of computing the maximum likelihood parameter estimate (MLE) of θ given the observation history $\mathcal{Y}_T$. The MLE is defined as follows: Suppose the family of measures P_θ were absolutely continuous with respect to a fixed probability measure P_0. The log likelihood function for computing an estimate of the parameter θ based on the information available in $\mathcal{Y}_T$ is $\mathcal{L}(\theta) = \log \mathbf{E}_0\{(dP_\theta/dP_0) \mid \mathcal{Y}_T\}$, and the MLE is defined by $\hat{\theta} \in \arg\max_{\theta\in\Theta}\mathcal{L}(\theta)$. Application of the EM algorithm to the Benes type nonlinear dynamical systems HMMs and piecewise linear systems are covered in Section IV-B,Section V-B, and Section VI, respectively.

B. Motivation:The EM Algorithm

As previously mentioned, the EM algorithm serves as a primary motivation for deriving fixed-interval smoothers for the state x_t and functionals of the state of the form H_t defined in (4). The EM algorithm is an iterative numerical method for computing the MLE. Let $\widehat{\theta}_0$ be the initial parameter estimate. Each iteration of the EM algorithm consists of two steps.

Step 1) (E-step) Set $\tilde{\theta} = \widehat{\theta}_j$ and compute $\mathcal{Q}(\cdot, \tilde{\theta})$, where $\mathcal{Q}(\theta,\tilde{\theta}) \stackrel{\triangle}{=} \mathbf{E}_{\tilde{\theta}}\{\log(dP_\theta/dP_{\tilde{\theta}}) \mid \mathcal{Y}_T\}$.

Step 2) (M-step) Find $\widehat{\theta}_{j+1} \in \arg\max_{\theta\in\Theta}\mathcal{Q}(\theta,\widehat{\theta}_j)$.

The sequence generated $\{\widehat{\theta}_j, j \geq 0\}$ gives nondecreasing values of $\mathcal{L}(\widehat{\theta}_j)$ with equality if and only if $\widehat{\theta}_{j+1} = \widehat{\theta}_j$.

Under the assumptions A1)–A5), for all $T \geq 0$, the measures $(P_\theta : \theta \in \Theta)$ when restricted to [0, T] are mutually absolutely continuous on $(\Omega, \mathcal{F})$ with Radon–Nikodym derivative $\Lambda^{\theta\tilde{\theta}} \stackrel{\triangle}{=} dP_\theta/dP_{\tilde{\theta}}$. It is shown in [4] that $\mathcal{Q}(\theta,\tilde{\theta}) = \mathbf{E}_{\tilde{\theta}}\left\{\log\Lambda^{\theta\tilde{\theta}} \mid \mathcal{Y}_T\right\}$ where (5), as shown at the bottom of the page, holds. It is clear from (5) that computing $\mathcal{Q}(\theta,\tilde{\theta})$ in the E-step involves computing fixed interval smoothed estimate of functionals of the state of the form H_t in (4).

C. Preliminaries

To simplify notation, reference to the parameter θ will be dropped until Section IV-B. We start with a reference probability space $(\Omega, \mathcal{F}, \bar{P})$ such that under $\bar{P}$

i) w is r-dimensional Brownian motion and $\{x_t\}$ is defined by (1);

ii) $\{y_t\}$ is n-dimensional Brownian motion, independent of w and x_0, and having quadratic variation $\langle y\rangle_t = I$.

Consider the exponentials

$$\Lambda_{t_1,t_2} = \exp\left(\int_{t_1}^{t_2} h'(x_s,s)dy_s - \frac{1}{2}\int_{t_1}^{t_2} h'(x_s,s)h(x_s,s)\,ds\right), t_1, t_2 \in [0,T]. \quad (6)$$

For notational convenience, define $\Lambda_t = \Lambda_{0,t}$. Then, from Ito's formula

$$\Lambda_t = 1 + \int_0^t \Lambda_s h'(x_s,s)dy_s \quad (7)$$

and $\bar{\mathbf{E}}\{\Lambda_t\} = 1$, where $\bar{\mathbf{E}}$ denotes expectation under $\bar{P}$. If we define a measure P in terms of $\bar{P}$ by setting $(dP/d\bar{P})\big|_{\mathcal{G}_t} = \Lambda_t$ then Girsanov's theorem [11] implies that under P, ν_t is a standard n-dimensional Brownian motion if we define $d\nu_t = dy_t - h(x_t,t)dt$, $\nu_0 = 0$. That is, under P, $dy_t = h(x_t,t)\,dt + d\nu_t$. Under P, the process $\{x_t\}$ still satisfies (1). Consequently, under P the processes $\{x_t\}$ and $\{y_t\}$ satisfy the real world dynamics (1) and (2). However, $\bar{P}$ is a more convenient measure with which to work.

In the sequel, we assume that $\phi \in C^2(\mathbb{R}^m)$ is an arbitrary "test" function with compact support. For any $\gamma_t(x), \delta_t(x) \in L_2([0,T]\times\mathbb{R}^m)$ define the inner product

$$\langle\gamma_t,\delta_t\rangle \stackrel{\triangle}{=} \int_{\mathbb{R}^m}\gamma_t(x)\delta_t(x)dx. \quad (8)$$

Filtering is concerned with computing $\mathbf{E}\{\phi(x_t)|\mathcal{Y}_t\}$. Define the density function $q_t(x)$ as $\int_{\mathbb{R}^m}\phi(x)q_t(x)dx = \bar{\mathbf{E}}\{\Lambda_t\,\phi(x_t)|\mathcal{Y}_t\}$. The following result is standard [11].

$$\begin{aligned}\log\Lambda^{\theta\tilde{\theta}} = &\int_0^T \left[f_\theta(x_s,s) - f_{\tilde{\theta}}(x_s,s)\right]' Q^{-1}(dx_s - f_{\tilde{\theta}}(x_s,s)ds)\\ &-\frac{1}{2}\int_0^T \left[f_\theta(x_s,s) - f_{\tilde{\theta}}(x_s,s)\right]' Q^{-1}\left[f_\theta(x_s,s) - f_{\tilde{\theta}}(x_s,s)\right]ds\\ &+\int_0^T \left[h_\theta(x_s,s) - h_{\tilde{\theta}}(x_s,s)\right]\left[dy_s - h_{\tilde{\theta}}(x_s,s)ds\right]\\ &-\frac{1}{2}\int_0^T \left[h_\theta(x_s,s) - h_{\tilde{\theta}}(x_s,s)\right]'\left[h_\theta(x_s,s) - h_{\tilde{\theta}}(x_s,s)\right]ds.\end{aligned} \quad (5)$$

Lemma 2.1: The filtered estimate $\mathbf{E}\{\phi(x_t)|\mathcal{Y}_t\}$ is given by

$$\mathbf{E}\{\phi(x_t)|\mathcal{Y}_t\} = \frac{\bar{\mathbf{E}}\{\Lambda_t\,\phi(x_t)|\mathcal{Y}_t\}}{\bar{\mathbf{E}}\{\Lambda_t|\mathcal{Y}_t\}} = \frac{\int_{\mathbb{R}^m}\phi(x)\,q_t(x)dx}{\int_{\mathbb{R}^m} q_t(x)dx} = \frac{\langle\phi,\ q_t\rangle}{\langle 1,\ q_t\rangle}. \tag{9}$$

We will subsequently refer to $q_t(x)$ as the *forward* unnormalized filtered density.

Fixed interval smoothing is concerned with computing conditional mean estimates of the form $\mathbf{E}\{\phi(x_t)|\mathcal{Y}_T\}, t\in[0,\ T]$. Consider the measure valued process

$$v_t(x) = \bar{\mathbf{E}}\left\{\Lambda_{t,T}|\mathcal{Y}_{t,T}\bigvee\{x_t=x\}\right\}, \qquad t\in[0,T], \quad \text{initialized by } v_T(x)=1. \tag{10}$$

We will subsequently refer to $v_t(x)$ as the *backward* filtered process.

Lemma 2.2: The fixed-interval smoothed estimate $\mathbf{E}\{\phi(x_t)|\mathcal{Y}_T\}$ is given by

$$\mathbf{E}\{\phi(x_t)|\mathcal{Y}_T\} = \frac{\int_{\mathbb{R}^m}\phi(x)q_t(x)v_t(x)dx}{\int_{\mathbb{R}^m} q_t(x)v_t(x)dx} = \frac{\langle\phi q_t,\ v_t\rangle}{\langle q_t,\ v_t\rangle}. \tag{11}$$

Proof: By the smoothing property of conditional expectations

$$\begin{aligned}\bar{\mathbf{E}}\{\phi(x_t)\Lambda_T|\mathcal{Y}_T\} &= \bar{\mathbf{E}}\left\{\bar{\mathbf{E}}\{\phi(x_t)\Lambda_t\Lambda_{t,T}|\mathcal{Y}_T\bigvee\mathcal{G}_t\}\mid\mathcal{Y}_T\right\}\\ &= \bar{\mathbf{E}}\left\{\phi(x_t)\Lambda_t\bar{\mathbf{E}}\{\Lambda_{t,T}|\mathcal{Y}_T\bigvee\mathcal{G}_t\}\mid\mathcal{Y}_T\right\}\end{aligned} \tag{12}$$

where $\mathcal{Y}_T\bigvee\mathcal{G}_t$ denotes the sigma algebra generated by $\mathcal{Y}_T,\mathcal{G}_t$. Now

$$\bar{\mathbf{E}}\{\Lambda_{t,T}|\mathcal{Y}_T\bigvee\mathcal{G}_t\} = \bar{\mathbf{E}}\{\Lambda_{t,T}|\mathcal{Y}_T\bigvee\{x_t\}\} = \bar{\mathbf{E}}\{\Lambda_{t,T}|\mathcal{Y}_{t,T}\bigvee\{x_t\}\} = v_t(x_t)$$

by the Markovian property of the process $\Lambda_{t,T}$ and the fact that under $\bar{P}$, y_t is standard Brownian motion. Therefore

$$\bar{\mathbf{E}}\{\phi(x_t)\Lambda_T|\mathcal{Y}_T\} = \bar{\mathbf{E}}\left\{\phi(x_t)\Lambda_t v_t(x_t)|\mathcal{Y}_t\right\}.$$

From [2, pp. 134] or [18, Lemma 3.10], it follows that

$$\bar{\mathbf{E}}\left\{\phi(x_t)\Lambda_t v_t(x_t)|\mathcal{Y}_t\right\} = \int_{\mathbb{R}^m}\phi(x)q_t(x)v_t(x)dx.$$

□

III. Robust Fixed Interval Smoothing

Notation: $Q=\sigma(x_t,t)\sigma'(x_t,t)$

$$\begin{aligned}\epsilon_t(x) &\stackrel{\triangle}{=} \exp\left[\int_0^t h'(x,s)dy_s - \frac{1}{2}\int_0^t h'(x,s)h(x,s)ds\right]\\ \bar{\epsilon}_t(x) &\stackrel{\triangle}{=} \frac{1}{\epsilon_t(x)} \qquad \forall x\in\mathbb{R}^m\end{aligned} \tag{13}$$

For convenience, we will use ϵ_t instead of $\epsilon_t(x)$, etc.

$$\nabla = \left(\frac{\partial}{\partial x_1},\frac{\partial}{\partial x_2},\ldots,\frac{\partial}{\partial x_n}\right)'$$
$$\nabla^2 = \begin{bmatrix}\frac{\partial^2}{\partial x_1^2} & \cdots & \frac{\partial^2}{\partial x_1\partial x_m}\\ \vdots & \vdots & \vdots\\ \frac{\partial^2}{\partial x_m\partial x_1} & \cdots & \frac{\partial^2}{\partial x_m^2}\end{bmatrix}. \tag{14}$$

For a vector field $g(x)=[g_1(x)\ g_2(x)\ \ldots\ g_m(x)]'$ defined on $\mathbb{R}^m$, define

$$\operatorname{div}(g) = \frac{\partial g_1}{\partial x_1}+\frac{\partial g_2}{\partial x_2}+\cdots+\frac{\partial g_m}{\partial x_m}.$$

Define the backward elliptic operator (infinitesimal generator) L and its adjoint L^* for any test function ϕ as

$$\begin{aligned}L(\phi) &= \frac{1}{2}\mathrm{Tr}[Q\nabla^2\phi] + f'\nabla\phi\\ L^*(\phi) &= \frac{1}{2}\mathrm{Tr}\left[\nabla^2(Q\phi)\right] - \mathrm{div}[f\phi].\end{aligned} \tag{15}$$

A. Robust Fixed Interval State Smoothers

We start with the following well-known DMZ equation, which describes the evolution of the unnormalized filtered state density; see, for example, [2] for a proof.

Theorem 3.1 (DMZ Equation): The unnormalized filtered density $q_t(x)$satisfies the stochastic integral equation

$$\begin{aligned}q_t(x) &= q_0(x) + \int_0^t L^*\left(q_s(x)\right)ds + \int_0^t h'(x,s)q_s(x)dy_s\\ q_0(\cdot) &= \pi_0(\cdot).\end{aligned} \tag{16}$$

The existence of a unique strong solution $q_t(x)$ is guaranteed under assumptions A1), A2), A3), A4), and A6), see [2, Sec.4.6]. In Section VI, where A6) is violated because $h(x,s)$ piecewise linear in x, a strong solution does not necessarily exist.

Our aim is to derive a *robust* version of the above DMZ filtering equation by introducting the following gauge transformation. Define the robust forward filtered density

$$\bar{q}_t(x) \stackrel{\triangle}{=} \bar{\epsilon}_t\,q_t(x), \qquad \bar{q}_0(x)=q_0(x). \tag{17}$$

The followingresult is proved in [16].

Theorem 3.2 (Robust Forward Filter): $\bar{q}_t$ satisfies the following nonstochastic parabolic partial differential equation:

$$\frac{\partial\bar{q}_t(x)}{\partial t} = \bar{\epsilon}_t L^*(\epsilon_t\bar{q}_t), \qquad \bar{q}_0(\cdot)=\pi_0(\cdot). \tag{18}$$

Furthermore, the robust filtered state estimate $\hat{x}_{t|t}\stackrel{\triangle}{=}\langle\epsilon_t\bar{q}_t,\,x\rangle/\langle\epsilon_t\bar{q}_t,\,1\rangle$ defines a locally Lipschitz version of $\mathbf{E}\{x_t|\mathcal{Y}_t\}$ in that for any two observation trajectories $y^{(1)}$, $y^{(2)}\in C(\mathbb{R}^n\times[0,T])$ and for some constant K depending on $\|y^{(1)}\|$ and $\|y^{(2)}\|$

$$|\hat{x}_{t|t}(y^{(1)}) - \hat{x}_{t|t}(y^{(2)})| \le K\|y^{(1)}-y^{(2)}\|.$$

Remark: Equation (18) follows straightforwardly from applying Ito's formula to $\bar{q}_t=\bar{\epsilon}_t q_t$. In [16], (18) is established by integrating Λ_t (defined in (6)) by parts. While both methods yield the same formula (18), it is worthwhile noting that in ϵ_t defined in (13), $x \in \mathbb{R}^m$ is merely a parameter, whereas in Λ_t defined in (6), x_t is a stochastic process. Finally, [16] also shows uniform continuity (robustness) in terms of an approximation parameter.

Define now the robust backward filtered process as

$$\bar{v}_t(x) \stackrel{\triangle}{=} \epsilon_t v_t(x). \tag{19}$$

The following theorem shows that one can derive the evolution of $\bar{v}_t(x)$ directly from the forward robust density $\bar{q}_t(x)$. In particular, one does not need to worry about the evolution of $v_t(x)$, which is governed by a backward stochastic partial differential equation.

Theorem 3.3 (Robust Backward Filter and Fixed-Interval Smoother): $\bar{v}_t$ satisfies the nonstochastic backward parabolic pde

$$\frac{\partial \bar{v}_t(x)}{\partial t} = -\epsilon_t L(\bar{\epsilon}_t \bar{v}_t), \qquad \bar{v}_T(x) = \epsilon_T. \tag{20}$$

The fixed interval smoothed estimate is computed as

$$\mathbf{E}\{\phi(x_t)|\mathcal{Y}_T\} = \frac{\int_{\mathbb{R}^m} \phi(x)\bar{q}_t(x)\bar{v}_t(x)dx}{\int_{\mathbb{R}^m} \bar{q}_t(x)\bar{v}_t(x)dx} = \frac{\langle \phi\bar{q}_t,\, \bar{v}_t\rangle}{\langle \bar{q}_t,\, \bar{v}_t\rangle}. \tag{21}$$

Remarks:

1) Reference [18] also presents a similar result (in French). However, the results in [18] are not exploited in computing functionals of the state which is one of the main aims of this paper. Existence and uniqueness of v_t holds under A1), A2), A3), A4), and A6); see [2, Ch. 4.6.4].
2) Equation (20) can be derived by starting with the following backward Ito stochastic differential equation for v_t:

$$v_t = v_T - \int_t^T L(v_s)ds - \int_t^T h v_s d\tilde{y}_s, \qquad v_T = 1$$

where $\tilde{y}_t = y_t - y_T$ and the last integral is a backward stochastic integral. Then apply the backward Ito formula of [2, pg.124] to (19). However, the following straightforward proof derives smoothers without recourse to backward stochastic calculus.

Proof: Choose $\phi(x) = 1$ in (11). This yields $\langle q_t, v_t\rangle = \bar{\mathbf{E}}\{\Lambda_T|\mathcal{Y}_T\}$ which means that $\langle q_t, v_t\rangle$ is independent of time t. Now from (17) and (19), we have

$$\langle q_t, v_t\rangle = \langle \epsilon_t\bar{q}_t, v_t\rangle = \langle \bar{q}_t, \epsilon_t v_t\rangle = \langle \bar{q}_t, \bar{v}_t\rangle$$

meaning that $\langle \bar{q}_t, \bar{v}_t\rangle$ is independent of time t. Thus, $d\langle \bar{q}_t, \bar{v}_t\rangle/dt = 0$, $\bar{P}$ a.s. However

$$\begin{aligned}\frac{d}{dt}\langle \bar{q}_t, \bar{v}_t\rangle =& \langle \frac{\partial \bar{q}}{\partial t}, \bar{v}_t\rangle + \langle \bar{q}_t, \frac{\partial \bar{v}_t}{dt}\rangle \\ =& \langle \bar{\epsilon}_t L^*\left(\epsilon_t \bar{q}_t\right),\ \bar{v}_t\rangle + \langle \bar{q}_t, \frac{\partial \bar{v}_t}{dt}\rangle \\ =& \langle \bar{q}_t,\ \epsilon_t L\left(\bar{\epsilon}_t\bar{v}_t\right)\rangle + \langle \bar{q}_t, \frac{\partial \bar{v}_t}{dt}\rangle \qquad (\bar{P} \text{ a.s.})\end{aligned}$$

which means that $\bar{v}_t$ satisfies the backward nonstochastic parabolic pde (20). Finally, Lemma 2.2 and (20) immediately yield (21). □

B. Robust Fixed-Interval Smoothers for Functionals of the State

We consider robust fixed interval smoothing of H_t defined in (4). As mentioned in Section II-B, such computations arise in the EM algorithm for MLE.

Define the measure valued process $\lambda_t(x)$ associated with H_t as

$$\bar{\mathbf{E}}\{\Lambda_t H_t \phi(x_t)|\mathcal{Y}_t\} = \langle \lambda_t, \phi\rangle. \tag{22}$$

Define the robust measure valued processes

$$\bar{\lambda}_t(x) = \bar{\epsilon}_t \lambda_t(x).$$

In terms of λ_t or its robust version $\bar{\lambda}_t$, it follows from a virtually identical proof to Lemma 2.2 (instead of (12) we now have $\bar{\mathbf{E}}\{\phi(x_t)\Lambda_T H_t|\mathcal{Y}_T\} = \bar{\mathbf{E}}\left\{\phi(x_t)\Lambda_t H_t \bar{\mathbf{E}}\{\Lambda_{t,T}|\mathcal{Y}_T \bigvee \mathcal{G}_t\} \mid \mathcal{Y}_T\right\}$) that $\mathbf{E}\{H_t|\mathcal{Y}_T\}$ is computed as

$$\mathbf{E}\{H_t|\mathcal{Y}_T\} = \frac{\langle \lambda_t, v_t\rangle}{\langle q_t, v_t\rangle} = \frac{\langle \bar{\lambda}_t, \bar{v}_t\rangle}{\langle \bar{q}_t, \bar{v}_t\rangle} = \frac{z_t}{\langle \bar{q}_t, \bar{v}_t\rangle} \tag{23}$$

where $z_t \stackrel{\triangle}{=} \langle \bar{\lambda}_t, \bar{v}_t\rangle$ denotes the unnormalized robust fixed-interval smoothed estimate.

Theorem 3.4 (Filtered and Robust Smoothed Estimate): See (24)–(26), as shown at the bottom of the next page. Furthermore, the robust smoothed state estimate $\hat{H}_{t|T} \stackrel{\triangle}{=} z_t/\langle \bar{q}_t, \bar{v}_t\rangle$ defines a locally Lipschitz version of $\mathbf{E}\{H_t|\mathcal{Y}_T\}$ in that for $y^{(1)}, y^{(2)} \in C(\mathbb{R}^n \times [0,T])$ and constant K depending on $\|y^{(1)}\|$ and $\|y^{(2)}\|$

$$|\hat{H}_{t|T}(y^{(1)}) - \hat{H}_{t|T}(y^{(2)})| \le K\|y^{(1)}) - y^{(2)}\|.$$

Proof: Starting with (4) and (7), it follows that for any test function $\phi \in C^2(\mathbb{R}^m)$

$$\begin{aligned}\Lambda_t\phi(x_t)H_t =& \phi(x_0)H_0 + \int_0^t \Lambda_s H_s L(\phi(x_s))ds \\ &+ \int_0^t \Lambda_s H_s \frac{\partial \phi(x)}{\partial x}\sigma(x_s,s)dw_s \\ &+ \int_0^t \Lambda_s \phi(x_s)\alpha(x_s,y_s)ds \\ &+ \int_0^t \Lambda_s \phi(x_s)\beta(x_s,y_s)dx_s \\ &+ \int_0^t \Lambda_s \phi(x_s)\gamma(x_s,y_s)dy_s \\ &+ \int_0^t \Lambda_s \beta'(x_s,y_s)Q\frac{\partial \phi(x_s)}{\partial x}ds \\ &+ \int_0^t \Lambda_s \phi(x_s)H_s h(x_s,s)dy_s \\ &+ \int_0^t \phi(x_s)\gamma(x_s,y_s)\Lambda_s h(x_s,s)ds.\end{aligned}$$

Conditioning on $\mathcal{Y}_t$ under the measure $\bar{P}$ (see [22, Lemma 3.2, p. 261]), it follows that

$$\begin{aligned}\bar{\mathbf{E}}\{\Lambda_t\phi(x_t)H_t|\mathcal{Y}_t\} =&\bar{\mathbf{E}}\{\phi(x_0)H_0|\mathcal{Y}_0\} \\ &+\int_0^t \bar{\mathbf{E}}\{\Lambda_s H_s L(\phi(x_s))|\mathcal{Y}_s\}ds \\ &+\int_0^t \bar{\mathbf{E}}\{\Lambda_s\phi(x_s)\alpha(x_s,y_s)|\mathcal{Y}_s\}ds \\ &+\int_0^t \bar{\mathbf{E}}\{\Lambda_s\phi(x_s)\beta(x_s,y_s)|\mathcal{Y}_s\}dx_s \\ &+\int_0^t \bar{\mathbf{E}}\{\Lambda_s\phi(x_s)\gamma(x_s,y_s)|\mathcal{Y}_s\}dy_s \\ &+\int_0^t \bar{\mathbf{E}}\{\Lambda_s\beta'(x_s,y_s)Q\frac{\partial\phi(x_s)}{\partial x}|\mathcal{Y}_s\}ds \\ &+\int_0^t \bar{\mathbf{E}}\{\Lambda_s\phi(x_s)H_s h(x_s,s)|\mathcal{Y}_s\}dy_s \\ &+\int_0^t \bar{\mathbf{E}}\{\phi(x_s)\gamma(x_s,y_s)\Lambda_s h(x_s,s)|\mathcal{Y}_s\}ds. \end{aligned} \tag{27}$$

Now, $\bar{\mathbf{E}}\{\Lambda_t\phi(x_t)H_t|\mathcal{Y}_t\}$ is a linear continuous functional on C^2 and therefore is a measure. With λ_t denoting the associated density, i.e., $\langle\phi,\ \lambda_t\rangle = \bar{\mathbf{E}}\{\Lambda_t\phi(x_t)H_t|\mathcal{Y}_t\}$, it follows that

$$\begin{aligned}\langle\phi,\ \lambda_t\rangle =&\langle\phi,\ \lambda_0\rangle + \int_0^t\langle L(\phi),\ \lambda_s\rangle ds + \int_0^t\langle\phi,\alpha q_s\rangle ds \\ &+\int_0^t\langle\phi,\ \beta' f q_s\rangle ds + \int_0^t\langle\phi,\ \gamma q_s\rangle dy_s \\ &+\int_0^t\langle\nabla'\phi,\ Q\beta q_s\rangle ds + \int_0^t\langle\phi,\ h\lambda_s\rangle dy_s \\ &+\int_0^t\langle\phi,\ \gamma h q_s\rangle ds \end{aligned} \tag{28}$$

which implies that λ_t satisfies (24).

Applying Ito's rule to $\bar{\lambda}_t = \bar{\epsilon}_t\lambda_t$ with λ_t satisfying (24), it follows that the second equation shown at the bottom of the page holds. Since the integrand $\gamma'(x)\bar{q}_s(x)$ of the last term is a finite variation process, the integral can be expressed as an ordinary (nonstochastic) integral using integration by parts as follows: $\int_0^t\gamma'(x)\bar{q}_s(x)dy_s = \gamma'(x)\bar{q}_t(x)y_t - \gamma'(x)\int_0^t y_s d\bar{q}_s$ which together with (18) implies (25).

$$\begin{aligned}\lambda_t(x) =&\lambda_0(x) + \int_0^t L^*(\lambda_s(x))ds \\ &+\int_0^t\left[\alpha(x,y_s)q_s(x) + \beta'(x,y_s)f(x,s)q_s(x) - \operatorname{div}\left[Q\beta(x,y_s)q_s(x)\right] + \gamma'(x)h(x,s)q_s(x)\right]ds \\ &+\int_0^t\left[\gamma'(x)q_s(x) + h'(x,s)\lambda_s(x)\right]dy_s \end{aligned} \tag{24}$$

$$\begin{aligned}\bar{\lambda}_t(x) =&H_0\bar{q}_0(x) + \int_0^t\bar{\epsilon}_s L^*\left(\epsilon_s\bar{\lambda}_s\right)ds \\ &+\int_0^t\left[\alpha(x,y_s)\bar{q}_s(x) + \beta'(x,y_s)f(x,s)\bar{q}_s(x) - \bar{\epsilon}_s\operatorname{div}\left[Q\beta(x,y_s)\epsilon_s\bar{q}_s(x)\right]\right]ds \\ &+\gamma'(x)\bar{q}_t(x)Y_t - \gamma'(x)\int_0^t Y_s\bar{\epsilon}_s L^*(\epsilon_s\bar{q}_s)ds \end{aligned} \tag{25}$$

$$\begin{aligned}Z_t =&\langle H_0\bar{q}_0,\bar{v}_0\rangle + \int_0^t\langle\alpha\bar{q}_s + \beta' f\bar{q}_s - \bar{\epsilon}_s\operatorname{div}\left[Q\beta\epsilon_s\bar{q}_s\right],\ \bar{v}_s\rangle ds \\ &-\int_0^t y_s\frac{d}{ds}\langle\gamma'\bar{q}_s,\bar{v}_s\rangle ds + \langle\gamma\bar{q}_t,\bar{v}_t\rangle y_t \end{aligned} \tag{26}$$

$$\begin{aligned}\bar{\lambda}_t(x) =&H_0\bar{q}_0(x) + \int_0^t\bar{\epsilon}_s L^*\left(\epsilon_s\bar{\lambda}_s\right)ds \\ &+\int_0^t\left[\alpha(x,y_s)\bar{q}_s(x) + \beta'(x,y_s)f(x,s)\bar{q}_s(x) - \bar{\epsilon}_s\operatorname{div}\left[Q\beta(x,y_s)\epsilon_s\bar{q}_s(x)\right]\right]ds \\ &+\int_0^t\gamma'(x)\bar{q}_s(x)dy_s. \end{aligned}$$

To prove (26), define $\bar{\xi}_t(x) \stackrel{\triangle}{=} \bar{\lambda}_t(x) - \gamma'(x)\bar{q}_t(x)y_t$. From (25), $\bar{\xi}_t$ satisfies the nonstochastic pde

$$\frac{\partial \bar{\xi}_t(x)}{\partial t} = \bar{\epsilon}_t L^*(\epsilon_t \bar{\lambda}_t) + \Big[\alpha(x,y_t)\bar{q}_t(x) + \beta'(x,y_t)f(x,t)\bar{q}_t(x) - \bar{\epsilon}_t \mathrm{div}\left[Q\beta(x_t,y_t)\epsilon_t\bar{q}_t(x)\right]\Big] - \gamma'(x)y_t\bar{\epsilon}_t L^*(\epsilon_t\bar{q}_t).$$

Also

$$\begin{aligned}\frac{\partial\langle\bar{\xi}_t,\bar{v}_t\rangle}{\partial t} =& \langle\frac{\partial\bar{\xi}_t}{\partial t},\bar{v}_t\rangle + \langle\bar{\xi}_t,\frac{\partial\bar{v}_t}{\partial t}\rangle\\ =&\langle\bar{\epsilon}_t L^*(\epsilon_t\bar{\lambda}_t),\bar{v}_t\rangle\\ &+\langle\alpha\bar{q}_t + \beta' f\bar{q}_t - \bar{\epsilon}_t\mathrm{div}\left[Q\beta_t\epsilon_t\bar{q}_t\right],\bar{v}_t\rangle\\ &-\langle\gamma'\bar{\epsilon}_t L^*(\epsilon_t\bar{q}_t),\bar{v}_t\rangle y_t + \langle\bar{\lambda}_t,\frac{\partial\bar{v}_t}{\partial t}\rangle\\ &-\langle\gamma'\bar{q}_t,\frac{\partial\bar{v}_t}{\partial t}\rangle y_t.\end{aligned} \tag{29}$$

Using (20), the following expressions hold:

$$\begin{aligned}\langle\bar{\lambda}_t,\frac{\partial\bar{v}_t}{\partial t}\rangle =& \langle\bar{\lambda}_t,-\epsilon_t L(\bar{\epsilon}_t\bar{v}_t)\rangle = -\langle\bar{\epsilon}_t L^*(\epsilon_t\bar{\lambda}_t),\bar{v}_t\rangle\\ \langle\gamma'\bar{q}_t,\frac{\partial\bar{v}_t}{\partial t}\rangle =& \langle\gamma'\bar{q}_t,-\epsilon_t L(\bar{\epsilon}_t,\bar{v}_t)\rangle = -\langle\bar{\epsilon}_t L^*(\gamma'\epsilon_t\bar{q}_t),\bar{v}_t\rangle\end{aligned}$$

because L^* is the adjoint of L. Substituting these expressions into (29) yields

$$\begin{aligned}\langle\bar{\xi}_t,\bar{v}_t\rangle =& \langle H\bar{q}_0,\bar{v}_0\rangle\\ &+\int_0^t\langle\alpha\bar{q}_s + \beta' f\bar{q}_s - \epsilon_s\mathrm{div}\left[Q\beta_s\epsilon_s\bar{q}_s\right],\ \bar{v}_s\rangle ds\\ &+\int_0^t\langle\bar{\epsilon}_s L^*(\gamma'\bar{q}_s\epsilon_s) - \bar{\epsilon}_s\gamma' L^*(\epsilon_s\bar{q}_s),\bar{v}_s\rangle y_s ds.\end{aligned}$$

However, it can be shown that

$$\langle\bar{\epsilon}_s L^*(\gamma'\bar{q}_s\epsilon_s) - \bar{\epsilon}_s\gamma' L^*(\epsilon_s\bar{q}_s),\ \bar{v}_s\rangle = -\frac{d}{ds}\langle\gamma'\bar{q}_s,\bar{v}_s\rangle$$

by evaluating, for example, the right-hand side of the previous equation. Therefore, $z_t = \langle\bar{\lambda}_t,\bar{v}_t\rangle = \langle\bar{\xi}_t,\bar{v}_t\rangle + \langle\gamma'\bar{q}_t,\bar{v}_t\rangle y_t$ which yields (26).

Since $\bar{q}_t$ and $\bar{v}_t$ are locally Lipschitz, so is z_t. □

IV. Example 1: Robust Benes Smoothers

A. *Robust Smoother for State*

The signal model we consider is the following special case of (1) and (2):

$$dx_t = (g(x_t,t) + F_t x_t)\,dt + \sigma_t dw_t, \qquad x_0 \sim \pi_0(\cdot) \tag{30}$$

$$dy_t = Cx_t\,dt + d\nu_t, \qquad y_0 = 0. \tag{31}$$

Here, σ_t in $\mathbb{R}^{m\times r}$ is no longer a function of x. For convenience assume y_t is a scalar valued observation process (i.e., $n = 1$). Also, $C \in \mathbb{R}^{1\times m}$ is assumed time-invariant for simplicity.

Assumption: We assume that $g(x,t)$ in (30) satisfies the following condition [2, p. 199]. Suppose that there exists $\psi(x,t)$ in $C^{2,1}(\mathbb{R}^m,\mathbb{R}^+)$ such that

$$Q_t\,\nabla\psi(x,t) = g(x,t), \quad x\in\mathbb{R}^m\,. \tag{32}$$

Assume that $\psi(x,t)$ satisfies the following Benes nonlinearity condition [2, p. 198]:

$$\frac{\partial\psi}{\partial t} + \frac{1}{2}\mathrm{Tr}\left(Q_t\,\nabla^2\psi\right) + \frac{1}{2}(\nabla\psi)'Q_t\nabla\psi + x'F_t'\nabla\psi = \frac{1}{2}x'\Gamma_t\,x + x'\mu_t + \kappa_t \tag{33}$$

where $\mathbf{\Gamma}_t \in \mathbb{R}^{m\times m}$ is an arbitrary symmetric matrix satisfying $\Gamma_t + C'C \geq 0, \mu_t\in\mathbb{R}^m$ is an arbitrary vector and $\kappa_t\in\mathbb{R}$ is an arbitrary scalar.

Remark: Several examples of nonlinearities $g(x,t)$ satisfying the aforementioned assumption are given in [2, p. 199]. For scalar valued processes ($m = n = 1$) examples include $\tan h(x)$ and $(x + K_1)/(x^2/2 + K_1x - (t/2 + K_2))$ where K_1, K_2 are arbitrary constants; see also [7].

Robust Forward Benes Filter: When the nonlinearity $g(x,t)$ satisfies (33), then for initial density $\pi_0(\cdot)$ the explicit solution of (18) is

$$\begin{aligned}\bar{q}_t(x) =& (2\pi)^{-m/2}\int_{\mathbb{R}^m}\exp(-\psi(\zeta,0))\,s_t(\zeta)\,\bar{q}_t(x,\zeta)\,\pi_0(\zeta)\,d\zeta\\ \bar{q}_0(\cdot) =& \pi_0(\cdot)\end{aligned} \tag{34}$$

where (35)–(36), as shown at the bottom of the page, hold. Here, the terms $\bar{r}_t(\zeta)$, S_t and K_t are defined as (37), as shown at the bottom of the page, and the $M\times M$ matrix Φ_t satisfies the equation

$$\frac{d\Phi}{dt} = (F_t - \Sigma_t(\Gamma_t + C'C))\Phi_t, \qquad \Phi_0 = I. \tag{38}$$

$$\bar{q}_t(x,\zeta) = \exp\left(\psi(x,t) - \frac{1}{2}x'(\Sigma_t^{-1} - C'Ct)x + \bar{r}_t'(\zeta)x - \frac{1}{2}\bar{r}_t'(\zeta)\Sigma_t\bar{r}_t(\zeta)\right) \tag{35}$$

$$s_t(\zeta) = K_t\exp\left(-\frac{1}{2}\zeta' S_t\zeta + \zeta'\rho_t\right). \tag{36}$$

$$\begin{aligned}\bar{r}_t(\zeta) =& \Sigma_t^{-1}\Phi_t\zeta + \tilde{r}_t, \qquad S_t = \int_0^t\Phi_s'(\Gamma_s + C'C)\Phi_s ds\\ K_t =& \exp\Bigg(-\int_0^t\frac{d}{ds}[\tilde{r}_s'\Sigma_s]C'y_s ds - \frac{C}{2}\left[\int_0^t y_s^2\frac{d\Sigma_s}{ds}ds - \int_0^t\Sigma_s ds\right]C'\\ &-\frac{1}{2}\int_0^t\left[(\tilde{r}_s + C'y_s)'\Sigma_s(\Gamma_s + C'C)\Sigma_s(\tilde{r}_s + C'y_s) + \mathrm{Tr}[\Sigma_s\Gamma_s] + 2\kappa_s + (\tilde{r}_s + C'y_s)'\Sigma_s\mu_s\right]ds\Bigg)\end{aligned} \tag{37}$$

The statistics $\tilde{r}_t, \Sigma_t$ and ρ_t satisfy

$$\begin{aligned}\frac{d\tilde{r}_t}{dt} =& - \Sigma_t^{-1}Q(\tilde{r}_t + C'y_t) - F_t'(\tilde{r}_t + C'y_t) + \mu_t \\ \tilde{r}_0 =& 0 \end{aligned} \tag{39}$$

$$\begin{aligned}\frac{d\Sigma_t}{dt} =& - \Sigma_t(\Gamma_t + C'C)\Sigma_t + Q + F_t\Sigma_t + \Sigma_t F_t' \\ \Sigma_0 =& 0 \end{aligned} \tag{40}$$

$$\begin{aligned}\rho_t =& - \int_0^t \Phi_s' F_s' C' y_s ds - \int_0^t \Phi_s'(\Gamma_s + C'C)\Sigma_s \tilde{r}_s ds \\ & - \int_0^t \Phi_s' \mu_s ds \\ \rho_0 =& 0. \end{aligned} \tag{41}$$

Robust Backward Benes Filter: The explicit solution for (20) is

$$\begin{aligned}\bar{v}_t(x) =& \int_{\mathbb{R}^m} \exp(\psi(\zeta, T))\, \bar{s}_t(\zeta)\, \bar{v}_t(x,\zeta)\, d\zeta \\ \bar{v}_T(x) =& \epsilon_T(x) \end{aligned} \tag{42}$$

where (43)–(44), as shown at the bottom of the page, hold. The terms $\bar{l}_t(\zeta), \bar{S}_t$ and $\bar{K}_t$ are defined as shown in (45) at the bottom of the page, and the $M \times M$ matrix $\bar{\Phi}_t$ satisfies the equation

$$\frac{d\bar{\Phi}}{dt} = (F_t + \Sigma_t(\Gamma_t + C'C))\bar{\Phi}_t, \qquad \bar{\Phi}_T = I. \tag{46}$$

The statistics $\tilde{l}_t, \bar{\Sigma}_t$ and $\bar{\rho}_t$ satisfy

$$\begin{aligned}\frac{d\tilde{l}_t}{dt} =& \bar{\Sigma}_t^{-1}Q(\tilde{l}_t - C'y_t) - F_t'(\tilde{l}_t - C'y_t) + \mu_t \\ \tilde{l}_T =& 0 \end{aligned} \tag{47}$$

$$\begin{aligned}\frac{d\bar{\Sigma}_t}{dt} =& \bar{\Sigma}_t(\Gamma_t + C'C)\bar{\Sigma}_t - Q + F_t\bar{\Sigma}_t + \bar{\Sigma}_t F_t' \\ \bar{\Sigma}_T =& 0 \end{aligned} \tag{48}$$

$$\begin{aligned}\bar{\rho}_t =& C'y_T - \int_t^T \bar{\Phi}_s' F_s' C' y_s ds \\ & - \int_t^T \bar{\Phi}_s' \mu_s ds - \int_t^T \bar{\Phi}_s'(\Gamma_s + C'C)\Sigma_s \tilde{l}_s ds \\ \bar{\rho}_T =& 0. \end{aligned} \tag{49}$$

Remarks:

1) Verifying that the previous robust filter equations satisfy (18) and (20) is straightforward but tedious and provides little insight. In the Appendix, the forward and backward robust Benes filter equations are derived starting from their nonrobust versions.
2) The aforementioned expressions for $\bar{q}_t$ and $\bar{v}_t$ do not involve stochastic integrals or any terms involving y_t outside of time integrals. As a result, computation of $d\bar{q}_t/dt$ and $d\bar{v}_t/dt$ for non-Gaussian initial conditions is straightforward. In Section VI, we will require $d\bar{q}_t/dt$ and $d\bar{v}_t/dt$.
3) For linear dynamics with initial distribution $\pi_0(\cdot)$, simply set $\psi(x,t) = 0, \Gamma_t = 0, \mu_t = 0$ and $\kappa_t = 0$. Further, if $\pi_0(\cdot) \sim N(\hat{x}_0, \Sigma_0)$, then the Kalman filter follows with conditional mean state estimate $m_t \stackrel{\triangle}{=} \mathbf{E}\{x_t|\mathcal{Y}_t\} = \Sigma_t(\tilde{r}_t + C'y_t)$, and the Kalman state covariance $\Sigma_t \stackrel{\triangle}{=} \mathbf{E}\{(x_t - m_t)(x_t - m_t)'\}$ given by the Riccati equation. For linear dynamics and Gaussian initial conditions, the conditional mean fixed-interval smoothed state estimate and associated covariance satisfy

$$m_{t|T} \stackrel{\triangle}{=} \mathbf{E}\{x_t|\mathcal{Y}_T\} = \left(\Sigma_t^{-1} + \bar{\Sigma}_t^{-1}\right)^{-1}\left(\tilde{r}_t + \tilde{l}_t\right) \tag{50}$$

$$\begin{aligned}\Sigma_{t|T} \stackrel{\triangle}{=}& \mathbf{E}\{(x_t - m_{t|T})(x_t - m_{t|T})'\} \\ =& (\Sigma_t^{-1} + \bar{\Sigma}_t^{-1})^{-1} \end{aligned} \tag{51}$$

B. Maximum Likelihood Parameter Estimation for Linear System

Consider the linear Gaussian system (30), (31) with $\psi = 0$, Gaussian initial conditions and time invariant parameters (F, σ, C) in controller canonical form, i.e.,

$$F = \begin{bmatrix} a_1 & a_2 & \cdots & a_m \\ I_{(m-1)\times(m-1)} & & & 0 \end{bmatrix}, \quad \sigma = \begin{bmatrix} 1 & 0_{1\times(m-1)} \\ 0_{(m-1)\times 1} & 0 \end{bmatrix}$$
$$C = [c_1, \ldots, c_m]'.$$

Let $\theta = [a_1, \ldots, a_m, c_1, \ldots, c_m]'$ denote the parameter vector.

$$\bar{v}_t(x,\zeta) = \exp\left(-\psi(x,t) - \frac{1}{2}x'(\bar{\Sigma}_t^{-1} + C'Ct)x + \bar{l}_t'(\zeta)x - \frac{1}{2}\bar{l}_t'(\zeta)\bar{\Sigma}_t\bar{l}_t(\zeta)\right) \tag{43}$$

$$\bar{s}_t(\zeta) = \bar{K}_t \exp(-\frac{1}{2}\zeta'\bar{S}_t\zeta + \zeta'\bar{\rho}_t). \tag{44}$$

$$\begin{aligned}\bar{l}_t(\zeta) =& \bar{\Sigma}_t^{-1}\Phi_t\zeta + \tilde{l}_t, \qquad \bar{S}_t = \int_t^T \bar{\Phi}_s'(\Gamma_s + C'C)\bar{\Phi}_s ds \\ \bar{K}_t =& \exp\Bigg(- \int_t^T \frac{d}{ds}[\tilde{l}_s'\bar{\Sigma}_s]C'y_s ds - \frac{C}{2}\left[-\int_t^T y_s^2 \frac{d\bar{\Sigma}_s}{ds} ds + \int_t^T \bar{\Sigma}_s ds\right]C' \\ & - \frac{1}{2}\int_t^T \left[(\tilde{l}_s + C'y_s)'\bar{\Sigma}_s(\Gamma_s + C'C)\bar{\Sigma}_s(\tilde{l}_s + C'y_s) + \mathrm{Tr}[\bar{\Sigma}_s\Gamma_s] + 2\kappa_s + (\tilde{l}_s + C'y_s)'\bar{\Sigma}_s\mu_s\right] ds\Bigg) \end{aligned} \tag{45}$$

The EM algorithm outlined in Section II will be used to compute the MLE of θ. It follows from (5), with Q^{-1} replaced by its pseudoinverse $Q^{\#}$ (see [4] for a justification of this), that

$$\begin{aligned}\mathcal{Q}(\theta,\tilde{\theta}) =& \mathbf{E}_{\tilde{\theta}}\{\int_0^T x_s' F' Q^{\#}\, dx_s - \frac{1}{2}\int_0^T x_s' F' Q^{\#} F x_s\, ds|\mathcal{Y}_T\} \\ &+ \mathbf{E}_{\tilde{\theta}}\left\{\int_0^T x_s' C'\, dy_s - \frac{1}{2}\int_0^T x_s' C' C x_s\, ds|\mathcal{Y}_T\right\} \\ &+ \mathbf{E}_{\tilde{\theta}}\{R(\tilde{\theta})|\mathcal{Y}_T\}. \end{aligned} \tag{52}$$

where $R(\tilde{\theta})$ does not involve θ.

To implement the M-step set, $\partial\mathcal{Q}/\partial\theta = 0$. This yields

$$\begin{aligned}[a_1,\ldots,a_m]' =& \left(\mathbf{E}_{\tilde{\theta}}\left\{\int_0^T x_s x_s' ds|\mathcal{Y}_T\right\}\right)^{-1} \\ &\times \mathbf{E}_{\tilde{\theta}}\left\{\int_0^T x_s dx_s(1)|\mathcal{Y}_T\right\} \end{aligned} \tag{53}$$

$$\begin{aligned}C =& \left(\mathbf{E}_{\tilde{\theta}}\left\{\int_0^T x_s x_s' ds|\mathcal{Y}_T\right\}\right)^{-1} \\ &\times \mathbf{E}_{\tilde{\theta}}\left\{\int_0^T x_s dy_s|\mathcal{Y}_T\right\}. \end{aligned} \tag{54}$$

In the following, we use (26) to compute the previous expressions. For convenience, the subscript $\tilde{\theta}$ is omitted.

Example 1: Consider computing $\mathbf{E}\{\int_0^t x_s x_s' ds|\mathcal{Y}_T\}$ which is required in (53) and (54). For $i \in \{1,2,\ldots,M\}$, $j \in \{1,2,\ldots,M\}$ define $H_t^{ij} = \int_0^t e_i' x_s x_s' e_j\, ds$ where e_i denotes the unit vector with 1 in the ith position. Then from (26) with $\alpha(x_s,y_s) = e_i' x_s x_s' e_j$, $\beta = \gamma = 0$ we have $z_t = \int_0^t \langle e_i' x_s x_s' e_j \bar{q}_s,\ \bar{v}_s\rangle ds$. Therefore

$$\mathbf{E}\{\int_0^T x_s x_s' ds|\mathcal{Y}_T\} = \int_0^T \left[\Sigma_{s|T} + m_{s|T} m_{s|T}'\right] ds$$

where $m_{s|T}$ and $\Sigma_{s|T}$ are the smoothed state estimate and covariance defined in (50) and (51).

Example 2: Consider computing $\mathbf{E}\{\int_0^t x_s dy_s|\mathcal{Y}_T\}$ which is required in (54). Define $H_t^i = \int_0^t e_i' x_s dy_s$. Then from (26) with $\alpha = \beta = 0$ and $\gamma(x_s) = e_i' x_s$

$$z_t = e_i'\left[\langle x\bar{q}_t,\ \bar{v}_t\rangle y_t - \int_0^t y_s \frac{d}{ds}\langle x\bar{q}_s,\ \bar{v}_s\rangle ds\right]$$

and, hence

$$\mathbf{E}\{H_t^i|\mathcal{Y}_T\} = e_i'\left[m_{t|T} y_t - \int_0^t y_s \left(\frac{d}{ds} m_{s|T}\right) ds\right] \tag{55}$$

where

$$\frac{d}{ds} m_{s|T} = \frac{d}{ds}\left([\bar{M}_s + \bar{N}_s]^{-1}[\bar{l}_s + \bar{r}_s]\right)$$

can be computed from the robust forward and backward Kalman filters.

Remark: It is interesting to note that the robust smoothed estimate (55) is identical to the generalized Stratonovich integral used in [9].

Example 3: Consider computing $\mathbf{E}\{\int_0^t x_s dx_s'|\mathcal{Y}_T\}$ which is required in (53). Let $H_t^{ij} = \int_0^t x_s' e_i e_j' dx_s$. Then from (26) with $\alpha = \gamma = 0$ and $\beta(x_s) = e_j e_i' x_s$ it follows that

$$\begin{aligned} z_t =& \int_0^t \langle x_s' e_i e_j' f(x_s,s)\bar{q}_s,\ \bar{v}_s\rangle ds \\ &- \int_0^t \langle \bar{\epsilon}_s \mathrm{div}\left[Q e_j e_i' x_s \epsilon_s \bar{q}_s\right],\bar{v}_s\rangle ds. \end{aligned}$$

Substituting the expression (which follows after some tedious algebra)

$$\begin{aligned}\mathrm{div}\left[Q e_j e_i' x_s \epsilon_s \bar{q}_s\right] =& Q_{ij}\epsilon_s \bar{q}_s + x_s' e_i e_j' Q \bar{q}_s \epsilon_s (C y_s - C'C x_s s) \\ &+ \epsilon_s \bar{q}_s x_s' e_i e_j' \left(-Q\bar{M}_s x + Q\bar{r}_s\right)\end{aligned}$$

it follows (after a few more steps) that

$$\begin{aligned}&\mathbf{E}\left\{\int_0^t x_s' e_i e_j' dx_s|\mathcal{Y}_T\right\} = \\ &- Q_{ij} t - \int_0^t e_j' Q e_i' m_{s|T}(C y_s + \bar{r}_s) ds \\ &+ \int_0^t e_j' Q\left(C'Ct + \bar{M}_t\right)\left(P_{s|T} + m_{s|T} m_{s|T}'\right) e_i ds.\end{aligned}$$

V. Example 2: Robust HMM Smoothers

Let x_t, $t \geq 0$ be a continuous-time Markov chain defined on $(\Omega, \mathcal{F}, P)$ with finite state–space $\{e_1, e_2, \ldots, e_m\}$ where e_i denotes the unit m-vector with 1 in the ith position. Let A denote the $m \times m$ transition rate matrix (infinitesimal generator), so that $\sum_{i=1}^m a_{ij} = 0$ for $1 \leq j \leq m$.

It is straightforward to show [11] that the semimartingale representation of x_t is

$$dx_t = A' x_t dt + dw_t, \qquad x_0 \sim \pi_0 \tag{56}$$

where w_t is a $\mathcal{F}_t$ zero mean m-vector martingale under P. Let $C = (c_1,\ c_2, \ldots, c_m) \in \mathbb{R}^{1\times m}$. Assume that x_t is observed via the scalar measurement process y_t as

$$dy_t = C x_t dt + d\nu_t, \qquad y_0 = 0 \tag{57}$$

where ν_t is Brownian motion independent of x_0. Equation (57) denotes the observation trajectory of a continuous time HMM, see [11] for applications of such models. Let $\theta = (a_{ij}, c_i, i \in \{1,\ldots,m\}, j \in \{1,\ldots,m\})$ denote the parameter vector of the HMM.

A. Robust HMM Smoother

Assume θ is known. From (6), it follows that

$$\Lambda_t = \exp\left(\int_0^t C x_s dy_s - \frac{1}{2}\int_0^t (C x_s)^2 ds\right).$$

Let $B \stackrel{\triangle}{=} \mathrm{diag}[C]$ denote a diagonal $m \times m$ matrix. Analogous to (16) it follows that the unnormalized filtered density $q_t = \mathbf{E}\{\Lambda_t x_t|\mathcal{Y}_t\}$ (note q_t is a m dimensional vector) is given by the Zakai equation [11, p. 185]

$$dq_t = A' q_t dt + B q_t dy_t, \qquad q_0 = \pi_0. \tag{58}$$

This equation is the well-known Wonham filter or HMM filter [11].

For any two vectors $\gamma, \delta \in \mathbb{R}^m$, let $\langle \gamma, \delta \rangle$ denote their scalar product. With $v_t \stackrel{\triangle}{=} \mathbf{E}\{\Lambda_{t,T} | \mathcal{Y}_{t,T} \bigvee x_t\}$, in complete analogy to Lemma 2.2 , the HMM smoothed-state estimate is computed as

$$\mathbf{E}\{x_t | \mathcal{Y}_T\} = \frac{\sum_{i=1}^{m} q_t(i) v_t(i) e_i}{\langle q_t, v_t \rangle}. \tag{59}$$

In analogy to (13), define the $m \times m$ diagonal exponential matrices ϵ_t as

$$\begin{aligned} \epsilon_t(i) &= \exp(c_i y_t - \frac{1}{2} c_i^2 t), \quad i = 1, 2, \ldots, m \\ \epsilon_t &= \operatorname{diag}(\epsilon_t(1), \ldots, \epsilon_t(m)) = \exp(B y_t - \frac{1}{2} B^2 t) \\ \bar{\epsilon}_t &= \epsilon_t^{-1}. \end{aligned}$$

Define the robust forward and backward filtered state estimates, respectively, as

$$\bar{q}_t = \bar{\epsilon}_t q_t, \qquad \bar{v}_t = \epsilon_t v_t.$$

Similar to Theorems 3.2 and 3.3, the following holds (proof omitted to save space).

Theorem 5.1:(Robust HMM Smoother): The robust forward and backward filters evolve as

$$\frac{d\bar{q}_t}{dt} = \bar{\epsilon}_t A' \epsilon_t \bar{q}_t, \qquad \bar{q}_0 = q_0 \tag{60}$$

$$\frac{d\bar{v}_t}{dt} = -\epsilon_t A \bar{\epsilon}_t \bar{v}_t, \qquad \bar{v}_T = \epsilon_T \mathbf{1}. \tag{61}$$

The fixed-interval smoothed estimate is computed as

$$\mathbf{E}\{x_t | \mathcal{Y}_T\} = \frac{\sum_{i=1}^{m} \bar{q}_t(i) \bar{v}_t(i) e_i}{\langle \bar{q}_t, \bar{v}_t \rangle}. \tag{62}$$

Remark: Equation (60) was derived in [6], where it was shown that $\bar{q}_t$ is a locally Lipschitz continuous function of $(y(s), 0 \leq s \leq t)$, and (60) can be used to define a version of the conditional probability distribution which enjoys this continuity property. An identical proof holds for the continuity of (61)

B. Maximum Likelihood Parameter Estimation for HMM

By using the EM algorithm outlined in Section II-B to compute the ML parameter estimate of θ, the following re-estimation equations are obtained [14]:

$$\begin{aligned} a_{ij} &= \frac{\mathbf{E}_{\tilde{\theta}} \left\{ N_T^{ij} | \mathcal{Y}_T \right\}}{\mathbf{E}_{\tilde{\theta}} \left\{ J_T^i | \mathcal{Y}_T \right\}}, \qquad i \neq j \\ c_i &= \frac{\mathbf{E}_{\tilde{\theta}} \left\{ G_T^i | \mathcal{Y}_T \right\}}{\mathbf{E}_{\tilde{\theta}} \left\{ J_T^i | \mathcal{Y}_T \right\}} \end{aligned} \tag{63}$$

$$\begin{aligned} N_T^{ij} &= \int_0^T \langle x_{s_-}, e_i \rangle \langle dx_s, e_j \rangle, \qquad J_T^i = \int_0^T \langle x_s, e_i \rangle ds \\ G_T^i &= \int_0^T \langle x_s, e_i \rangle dy_s. \end{aligned} \tag{64}$$

Here, N_T^{ij}, $i \neq j$, denotes the number of jumps from state i to state j, J_T^i denotes the duration time in state i and G_T^i denotes the "level integral" from time 0 to T. Note that by interchanging conditional expectation and integral in the computation of the level integral $\mathbf{E}_{\tilde{\theta}}\{G_T^i | \mathcal{Y}_T\}$, the resulting expression $\int_0^T \langle \mathbf{E}_{\tilde{\theta}} \{x_s \mid \mathcal{Y}_T\}, \ e_i \rangle dy_s$ is not an Ito integral; it needs to be interpreted as a Skorohod integral. In the following, robust smoothers are developed for evaluating these quantities which does not require two-sided Skorohod integrals.

Theorem 5.2: Robust smoothed estimates of J_t^i, N_t^{ij}, and G_t^i are given as (65)–(67), shown at the bottom of the page. These robust estimates are locally Lipschitz continuous in y_t

Remarks :

1) The EM (63) for HMM parameter estimation read (68)–(69), as shown at the bottom of the page. These equations are the continuous-time counterpart of the

$$\mathbf{E}_{\tilde{\theta}} \left\{ J_t^i | \mathcal{Y}_T \right\} = \int_0^t \bar{q}_s(i) \bar{v}_s(i) ds \tag{65}$$

$$\mathbf{E}_{\tilde{\theta}} \left\{ N_t^{ij} | \mathcal{Y}_T \right\} = \int_0^t \frac{\epsilon_s(i)}{\epsilon_s(j)} a_{ij} \bar{v}_s(j) \bar{q}_s(i) ds, \qquad i \neq j \tag{66}$$

$$\begin{aligned} \mathbf{E}_{\tilde{\theta}} \left\{ G_t^i | \mathcal{Y}_T \right\} =& \bar{q}_t(i) \bar{v}_t(i) y_t \\ &- \int_0^t y_s \sum_{j=1}^{M} \left[a_{ji} \frac{\epsilon_s(j)}{\epsilon_s(i)} \bar{q}_s(j) \bar{v}_s(i) - a_{ij} \frac{\epsilon_s(i)}{\epsilon_s(j)} \bar{v}_s(j) \bar{q}_s(i) \right] ds. \end{aligned} \tag{67}$$

$$a_{ij} = \tilde{a}_{ij} \frac{\int_0^T \frac{\epsilon_t(i)}{\epsilon_t(j)} \bar{q}_t(i) \bar{v}_t(j) dt}{\int_0^T \bar{q}_t(i) \bar{v}_t(i) dt}, \quad i \neq j \tag{68}$$

$$c_i = \frac{\bar{q}_T(i) \bar{v}_T(i) y_T - \int_0^T y_t \sum_{j=1}^{M} \left[\tilde{a}_{ji} \frac{\epsilon_t(j)}{\epsilon_t(i)} \bar{q}_t(j) \bar{v}_t(i) - \tilde{a}_{ij} \frac{\epsilon_t(i)}{\epsilon_t(j)} \bar{v}_t(j) \bar{q}_t(i) \right] dt}{\int_0^T \bar{q}_t(i) \bar{v}_t(i) dt}. \tag{69}$$

discrete-time Baum–Welch equations, which are widely used for discrete-time MLE of HMMs; see, for example, [5]. The expressions (68) and (69) are apparently obtained here for the first time. In comparison, the EM equation for c_i derived in [9] is

$$c_i = \frac{\int_0^T q_t(i)v_t(i) \circ dy_t}{\int_0^T q_t(i)v_t(i)dt}$$

where the integral in the numerator is a generalized Stratonovich integral. The EM equation derived in [14, p.600] is

$$c_i = \frac{\int_0^T q_t(i)v_t(i) \cdot dy_t + \tilde{c}_i \int_0^T q_t(i)v_t(i)dt}{\int_0^T q_t(i)v_t(i)dt}$$

where the integral in the numerator is a two-sided Skorohod integral. The derivation of $\mathbf{E}_{\tilde{\theta}}\{N_t^{ij}|\mathcal{Y}_T\}$ and $\mathbf{E}_{\tilde{\theta}}\{G_t^{i}|\mathcal{Y}_T\}$ in [14] uses two-sided stochastic calculus of [19].

2) *Euler discretization*: We consider here numerical discretization of the forward and backward HMM filtering (60), (61). Consider a regular partition $0 = t_0 < t_1 < \cdots < t_{n-1} < t_n < \cdots$ with constant time step $\Delta = t_n - t_{n-1}$. Let $z_n^{\Delta} = (y_{t_n} - y_{t_{n-1}})/\Delta$ denote the discrete-time sampled observations. Define the discrete-time observation probability matrix

$$B(z_n^{\Delta}) = \text{diag}(b_1(z_n^{\Delta}), \ldots, b_m(z_n^{\Delta}))$$
$$b_i(z_n^{\Delta}) = \frac{\sqrt{\Delta}}{\sqrt{2\pi}} \exp(-\frac{\Delta}{2}(z_n^{\Delta} - c_i)^2).$$

A first-order (Euler) explicit discretization of $\bar{q}_t$ in (58) yields: $\bar{q}_{t_{n+1}} = (I + \Delta \bar{\epsilon}_{t_n} A' \epsilon_{t_n})\bar{q}_{t_n}$. Multiplying both sides by ϵ_{t_n} yields

$$q_{t_{n+1}} = B(z_{n+1}^{\Delta})(I + \Delta A')q_{t_n} \tag{70}$$

which is identical to the standard discrete-time HMM filter. Similarly, a first-order discretization of $\bar{v}_t$ in (61) yields the backward recursion

$$v_{t_n} = (I + \Delta A)B(z_{n+1}^{\Delta})v_{t_{n+1}} \tag{71}$$

which is identical to the standard discrete-time HMM backward filter.

Note that providing Δ is sufficiently small so that $(I + \Delta A)$ is a stochastic matrix, the robustified estimates q_{t_n} in (70) and v_{t_n} in (71) are guaranteed to be nonnegative. In contrast, a first order discretization of the nonrobust equations can yield negative values for q_{t_n} (58) and v_{t_n} for a fixed Δ. Similarly, the summation approximation to $\mathbf{E}_{\tilde{\theta}}\{N_t^{ij}|\mathcal{Y}_T\}$ and $\mathbf{E}_{\tilde{\theta}}\{J_t^{i}|\mathcal{Y}_T\}$ in (66) and (65), using (70) and (71), are guaranteed to be nonnegative.

Proof: Let H_t denote either J_t^i, G_t^i or N_t^{ij}. In analogy to (22), define the m-dimensional vectors $\lambda_t = \bar{\mathbf{E}}\{\Lambda_t H_t X_t|\mathcal{Y}_t\}$ and its robust version $\bar{\lambda}_t = \bar{\epsilon}_t \lambda_t$. Then (23) follows with $z_t = \langle \bar{\lambda}_t, \bar{v}_t \rangle$.

Consider first the case $H_t = G_t^i$. A similar proof to (25))see also [14, eqs. (2.14) and (2.17)]) shows that

$$\bar{\lambda}_t = \int_0^t \bar{\epsilon}_s A' \epsilon_s \bar{\lambda}_s ds + \langle \bar{q}_t, e_i \rangle e_i y_t - \int_0^t y_s \langle \bar{\epsilon}_s A' \epsilon_s \bar{q}_s, e_i \rangle e_i ds. \tag{72}$$

In analogy to the proof of Theorem 3.4, define the m vector $\bar{\xi}_t \triangleq \bar{\lambda}_t - \langle \bar{q}_t, e_i \rangle e_i y_t$ so that $\bar{\xi}_0 = \bar{\lambda}_0 = 0$. Then, using (68), it follows that $\bar{\xi}_t$ satisfies the nonstochastic ordinary differential equation (ode)

$$\frac{d\bar{\xi}_t}{dt} = \bar{\epsilon}_t A' \epsilon_t \bar{\lambda}_t - y_t \langle \bar{\epsilon}_t A' \epsilon_t \bar{q}_t, e_i \rangle e_i, \qquad \bar{\xi}_0 = 0.$$

Also similar to the proof of Theorem 3.4, it can be shown that

$$\langle \bar{\xi}_t, \bar{v}_t \rangle = -\int_0^t y_s \frac{d}{ds} [\bar{q}_s(i)\bar{v}_s(i)] \, ds.$$

Therefore, $\langle \bar{\lambda}_t, \bar{v}_t \rangle = \langle \bar{\xi}_t, \bar{v}_t \rangle + \langle \bar{q}_t, e_i \rangle \langle e_i, \bar{v}_t \rangle y_t$ yields (67).

Then, consider the case $H_t = N_t^{ij}$. Along the lines of (25), it follows that

$$\bar{\lambda}_t = \int_0^t \bar{\epsilon}_s A' \epsilon_s \bar{\lambda}_s ds + \int_0^t \bar{q}_s(i) e_j a_{ij} \frac{\bar{\epsilon}_s(i)}{\bar{\epsilon}_s(j)} ds.$$

Then, using $z_t = \int_0^t \bar{\lambda}_t' d\bar{v}_s + \int_0^t \bar{v}_s' d\bar{\lambda}_s$ yields (66). The proof for J_t^i follows similarly and is omitted.

The Lipschitz continuity follows trivially from the Lipschitz continuity of $\bar{q}_t$ and $\bar{v}_t$. □

VI. Example 3: Piecewise Linear Systems

In this section, we consider a partially observed system with piecewise linear dynamics and observation equation. For such systems, there is no finite-dimensional filter for computing the optimal state estimate (see [1] for a nonstandard type filtering formula in terms of Green's functions). Unlike previous sections of this paper, in general, the filtered density for such models does not exist. Therefore, the Zakai equations will be considered in weak form, i.e., distributional sense.

In [20] and [21], it is shown that the robust formulation of the weak Zakai equation allows for the construction of a suboptimal filter for computing state estimates of the piecewise linear system. The approximate filter in [20] consists of a bank of linear Kalman type filters with non-Gaussian initial conditions, each filter operating on one of the piecewise linear segments. In the same spirit as [20], we show how the robust formulation can be used to construct approximate smoothers for the state and functionals of the state for such piecewise linear systems. These smoothers are used in the EM algorithm to compute the MLE of the piecewise linear segments.

Signal Model and Parameter Estimation Problem: Consider the following scalar piecewise linear dynamical model (1), (2), where $\sigma(t)$ known

$$f(x_t) = \sum_{k=1}^K I(x_t \in P_k)(a_k x_t + b_k)$$
$$h_\theta(x_t) = \sum_{k=1}^K I(x_t \in P_k)(c_k x_t + d_k)$$
$$\sigma(t) \quad \text{known}. \tag{73}$$

Here, $P_k, k = 1, 2, \ldots, K$ denotes a finite partition of $\mathbb{R}$, and a_k, b_k are assumed to be known constants. Let $B_k \in \mathbb{R}, k = 1, \ldots, K-1$ denote the boundary points (change points) of $P_1, \ldots, P_K$. Let $\theta = (c_1, \ldots, c_K, d_1, \ldots, d_K)'$ denote the parameter vector to be estimated. We assume that $h_\theta(x)$ is continuous in x, so that at the boundary points

$$c_i B_i + d_i = c_{i+1} B_i + d_{i+1}, \quad i = 1, 2, \ldots, K-1. \tag{74}$$

Under these conditions, it is well known [20] that (1) has a unique weak solution.

Now, consider computing the MLE of θ via the EM algorithm. From (5), we have

$$\mathcal{Q}(\theta, \widehat{\theta}_j) = \mathbf{E}_{\widehat{\theta}_j} \left\{ \int_0^T \sum_{k=1}^K (c_k x_s + d_k) I(x_s \in P_k) dy_s \right. \\ \left. -\frac{1}{2} \int_0^T \left[\sum_{k=1}^K (c_k x + d_k) I(x \in P_k) \right]^2 ds | \mathcal{Y}_T \right\}. \tag{75}$$

Then, the M-step which requires computation of $\max_\theta \mathcal{Q}(\theta, \widehat{\theta}_j)$ with continuity constraint (74) can be expressed as the following equality constrained quadratic optimization problem:

$$\text{Minimize } \frac{1}{2}\theta' A \theta - G\theta, \qquad \text{subject to } R\theta = 0 \tag{76}$$

where A, G, R are, respectively, $2K \times 2K, 2K \times 1$, and $(K-1) \times 2K$-dimensional matrices with the equation shown at the bottom of the next page holding true. The elements $A_k^{(i)}$, $G_k^{(i)}$ for $i = 0, 1, 2$ and $k = 1, 2, \ldots, K$, are

$$A_k^{(i)} = \mathbf{E}_{\widehat{\theta}_j} \left\{ \int_0^T x_s^i I(x_s \in P_k) ds | \mathcal{Y}_T \right\} \\ G_k^{(i)} = \mathbf{E}_{\widehat{\theta}_j} \left\{ \int_0^T x_s^i I(x_s \in P_k) dy_s | \mathcal{Y}_T \right\}. \tag{77}$$

Assuming A is positive definite, the solution to the quadratic program (76)[3, pp. 201—204] yields the EM parameter estimate at iteration $(j+1)$ as

$$\widehat{\theta}_{j+1} = [I - A^{-1}R'(RA^{-1}R')^{-1}R]A^{-1}G. \tag{78}$$

(Because the $K-1$ rows of R are linearly independent for all boundary points B_i, hence, $RA^{-1}R'$ is invertible providing A is positive definite).

As described in Section II, computing (77) in the EM update (78) motivates the need to derive smoothers for the state x_t and functionals of the state H_t defined in (4) with $\beta = 0$. To simplify notation, in the sequel the subscript θ will be omitted in $\mathbf{E}_\theta$.

For any test function $\phi \in C^2(\mathbb{R}^m)$, define the filtered and smoothed distributions

$$\bar{\pi}_t(\phi H_t) = \bar{\mathbf{E}}\{\Lambda_t \phi(x_t) H_t | \mathcal{Y}_t\} \\ \bar{\pi}_{t|T}(\phi H_t) = \bar{\mathbf{E}}\{\Lambda_T \phi(x_t) H_t | \mathcal{Y}_T\}.$$

From (27), the Zakai equation in weak form (with $\beta = 0$) is

$$\bar{\pi}_t(\phi H_t) = \bar{\pi}_0(\phi H_0) \\ + \int_0^t [\bar{\pi}_s(L\phi) + \bar{\pi}_s(\phi\alpha) + \bar{\pi}_s(\phi\gamma h)] ds \\ + \int_0^t [\bar{\pi}_s(\phi\gamma) + \bar{\pi}(\phi H_s h)] dy_s. \tag{79}$$

Unlike the proof of Theorem 3.4, one cannot postulate the existence of the densities q_t or λ_t.

Approximate Model: For any $a \in \mathbb{R}$ let $[a]$ denote the integer part of a. Let $\Delta > 0$ denote a fixed real number. Consider the approximated version of the piecewise linear model (73) on (Ω, $\mathcal{F}$, P^Δ) with

$$f^\Delta(x, t) = \sum_{k=1}^K I\left(x_{[t/\Delta]\Delta} \in P_k\right)(a_k x_t + b_k) \\ h^\Delta(x, t) = \sum_{k=1}^K I\left(x_{[t/\Delta]\Delta} \in P_k\right)(c_k x_t + d_k) \tag{80}$$

$f^\Delta : C(\mathbb{R} \times [0, T]) \times [0, T] \to D(\mathbb{R}, [0, T])$, and $h^\Delta : C(\mathbb{R} \times [0, T]) \times [0, T] \to D(\mathbb{R}, [0, T])$. Here, $D(\mathbb{R}, [0, T])$ denotes the space of functions from $[0, T]$ into $\mathbb{R}$ which are right continuous and possess left limits at each point $t \in [0, T]$.

In the sequel, we will need to explicitly refer to the trajectories of x_t and y_t. Let $\Omega^1 = C(\mathbb{R} \times [0, T])$ and $\Omega^2 = C(\mathbb{R} \times [0, T])$ with elements $\omega_t^1 = x_t(\omega) \in \Omega^1$ and $\omega_t^2 = y_t(\omega) \in \Omega^2$ where $\omega = (\omega_t^1, \omega_t^2) \in \Omega = \Omega^1 \times \Omega^2$.

Since (1) with f replaced with f^Δ in (80) is a linear stochastic differential equation on each interval $[i\Delta, (i+1)\Delta)$ with coefficients depending on $x_{i\Delta}$, it has a unique strong solution. Similar to (6) define

$$\Lambda_{t_1,t_2}^\Delta(\omega_1, \omega_2) \\ = \exp\left(\int_{t_1}^{t_2} (h^\Delta(x_s, s) dy_s - \frac{1}{2}\int_{t_1}^{t_2} (h^\Delta(x_s, s))^2 ds \right) \\ t_1, t_2 \in [0, T]. \tag{81}$$

Since f^Δ and h^Δ satisfy the linear growth condition $|f^\Delta(x, t) + h^\Delta(x, t)| \leq c(1 + |x_t|)$ for some constant $c \in \mathbb{R}$, Λ^Δ is a martingale. Thus, as in Section II-C, define the measure $\bar{P}^\Delta$ by $dP^\Delta / d\bar{P}^\Delta|_{\mathcal{G}_t} = \Lambda_t^\Delta$. Define $\epsilon_t^\Delta(x)$ and $\bar{\epsilon}_t^\Delta(x)$ as in (13) for the model (80).

For any test function $\phi \in C^2(\mathbb{R}^m)$, define the following densities and distributions:

$$\bar{\pi}_t^\Delta(\phi) = \bar{\mathbf{E}}^\Delta\{\Lambda_t^\Delta \phi(x_t) | \mathcal{Y}_t\} = \langle \phi, q_t^\Delta \rangle \\ \bar{\pi}_t^\Delta(\phi H_t) = \bar{\mathbf{E}}^\Delta\{\Lambda_t^\Delta H_t \phi(x_t) | \mathcal{Y}_t\} = \langle \phi, \lambda_t^\Delta \rangle \\ v_t^\Delta(x) = \bar{\mathbf{E}}^\Delta \left\{ \Lambda_{t,T}^\Delta | \mathcal{Y}_{t,T} \bigvee \{x_t = x\} \right\} \\ \bar{\pi}_{t|T}^\Delta(\phi H_t) = \bar{\mathbf{E}}^\Delta\{\Lambda_T^\Delta H_t \phi(x_t) | \mathcal{Y}_T\} = \langle \phi, z_t^\Delta \rangle \\ \bar{q}_t^\Delta(x) = \bar{\epsilon}_t^\Delta q_t^\Delta(x), \quad \bar{\lambda}_t^\Delta = \bar{\epsilon}_t^\Delta \lambda_t^\Delta \\ \bar{v}_t^\Delta = \epsilon_t^\Delta v_t^\Delta(x), \qquad z_t^\Delta = \langle \bar{\lambda}_t^\Delta, \bar{v}_t^\Delta \rangle. \tag{82}$$

The aim of this section is to show the following.

i) The robust forward filtered density $\bar{q}_t^{\Delta}(x)$ and backward density $\bar{v}_t^{\Delta}(x)$ of the approximate model (80) can be computed by a bank of K parallel Kalman type forward and backward filters with non-Gaussian initial conditions. Details are given in Section VI-A.

ii) As $\Delta \to 0$, $\bar{\pi}_{t|T}^{\Delta}(\phi H_t)(\omega^2) \to \bar{\pi}_{t|T}(\phi H_t)(\omega^2)$, $\forall \omega^2 \in \Omega^2$ (i.e., pathwise) $\forall t \geq 0$. That is, the smoothed dis-

$$A = \left[\begin{array}{cccc|cccc}
A_1^{(2)} & & & & A_1^{(1)} & & & \\
& A_2^{(2)} & & & & A_2^{(1)} & & \\
& & \ddots & & & & \ddots & \\
& & & A_k^{(2)} & & & & A_k^{(1)} \\
\hline
A_1^{(1)} & & & & A_1^{(0)} & & & \\
& A_2^{(1)} & & & & A_2^{(0)} & & \\
& & \ddots & & & & \ddots & \\
& & & A_k^{(1)} & & & & A_k^{(0)}
\end{array}\right]$$

$$G = \left[\begin{array}{c}
G_1^{(1)} \\
G_2^{(1)} \\
\vdots \\
G_k^{(1)} \\
\hline
G_1^{(2)} \\
G_2^{(2)} \\
\vdots \\
G_k^{(2)}
\end{array}\right]$$

$$R = \left[\begin{array}{ccccc|ccccc}
B_1 & -B_1 & & & & 1 & -1 & & & \\
& B_2 & -B_2 & & & & 1 & -1 & & \\
& & \ddots & \ddots & & & & \ddots & \ddots & \\
& & & B_{k-1} & B_{k-1} & & & & 1 & -1
\end{array}\right]$$

tribution for the approximate model (80) converges to the smoothed distribution for the piecewise linear model (73). The robust formulation is used in the convergence proof. Details are given in Section VI-B.

A. Approximate Smoothing Algorithm

For each of the piecewise linear segments $k = 1, \dots, K$ define

$$\begin{aligned}\epsilon_{t,k}(x) &= \exp\left[c_k(x)y_t - \frac{1}{2}c_k^2(x)t\right]\\ \bar{\epsilon}_{t,k}(x) &= \frac{1}{\epsilon_{t,k}(x)} \qquad \text{for} x \in \mathbb{R}^m\\ L_k(\phi) &= \frac{1}{2}\mathrm{Tr}[Q\nabla^2\phi] + a_k x\nabla\phi\\ L_k^*(\phi) &= \frac{1}{2}\mathrm{Tr}\left[\nabla^2(Q\phi)\right] - \mathrm{div}[a_k x\phi]. \end{aligned} \tag{83}$$

The smoothed estimate $\mathbf{E}^\Delta\{\phi(x_t)|\mathcal{Y}_T\}$ for the approximate model (80) can be computed via the following algorithm.

Robust Forward Filter over $t \in [i\Delta, (i+1)\Delta)$, where $i = 0, 1, \dots, [T/\Delta]$:

Step 1. **Re-initialize:** At $t = i\Delta$ initalize with non-Gaussian initial condition:

$$\bar{q}^{i\Delta}_{i\Delta,k} = \bar{q}^{\Delta}_{i\Delta}(x)I(x \in P_k), \quad k = 1, 2, \dots, K.$$

Step 2. **Propagate:** Run K Kalman filters for non Gaussian initial condition (see **Section IV-A**) on $t \in [i\Delta, (i+1)\Delta)$ as

$$\frac{\partial \bar{q}^{i\Delta}_{t,k}(x)}{\partial t} = \bar{\epsilon}_{t,k} L_k^*(\epsilon_t \bar{q}^{i\Delta}_{t,k}).$$

Step 3. **Recombine:** At time $t = (i+1)\Delta$,

$$\bar{q}^{\Delta}_{(i+1)\Delta}(x) = \sum_{k=1}^{K} \bar{q}^{i\Delta}_{i+1\Delta,k}(x).$$

Step 4. Set $i := i+1$, go to Step 1.

The backward filter over $t \in (i\Delta, (i+1)\Delta]$ is similar. Reinitialize at $t = (i+1)\Delta$ as $\bar{v}^{(i+1)\Delta}_{(i+1)\Delta,k} = \bar{v}^{\Delta}_{(i+1)\Delta}(x)I(x \in P_k)$; propagate according to (20); recombine at $t = i\Delta$ to obtain $\bar{v}^{\Delta}_{i\Delta}(x)$, etc.

The smoothed-state estimate is computed as $\mathbf{E}^\Delta\{x_t|\mathcal{Y}_T\} = \langle x\bar{q}^\Delta_t, \bar{v}^\Delta_t\rangle / \langle \bar{q}^\Delta_t, \bar{v}^\Delta_t\rangle$. The estimates $A_k^{\Delta(i)}$, $G_k^{\Delta(i)}$ of (77) are computed using the generic formula $\mathbf{E}^\Delta\{H_t|\mathcal{Y}_T\} = z^\Delta_t / \langle \bar{q}^\Delta_t, \bar{v}^\Delta_t\rangle$ where z^Δ_t is computed by (26). Finally, the approximate EM update for the parameter estimate $\widehat{\theta}^\Delta_{j+1}$ at iteration $(j+1)$ is given by (78) with $A_k{}^{(i)}$, $G_k{}^{(i)}$ replaced by the estimates $A_k^{\Delta(i)}$, $G_k^{\Delta(i)}$. These are computed using (26) with $\alpha(x,y) = x^i I(x \in P_k)$ and $\gamma(x) = x^i I(x \in P_k)$, respectively

$$\begin{aligned} A_k^{\Delta(i)} &= \frac{1}{\langle \bar{q}^\Delta_T, \bar{v}^\Delta_T\rangle}\int_0^T \langle x^i I(x \in P_k)\bar{q}^\Delta_s, \bar{v}^\Delta_s\rangle ds\\ G_k^{\Delta(i)} &= \frac{1}{\langle \bar{q}^\Delta_T, \bar{v}^\Delta_T\rangle}\Big[\langle x^i I(x \in P_k)\bar{q}^\Delta_T, \bar{v}^\Delta_T\rangle y_T\\ &\qquad - \int_0^T y_s \frac{d}{ds}\langle x^i I(x \in P_k)\bar{q}^\Delta_s, \bar{v}^\Delta_s\rangle ds\Big]. \end{aligned}$$

Expressions for the forward and backward Kalman filters $\bar{q}^\Delta_t$ and $\bar{v}^\Delta_t$ are given in (34) and (42). Since there are no stochastic integrals in (34), (42), and y_s only appears within integrals over time, the derivatives $d\bar{q}^\Delta_s/ds$ and $d\bar{v}^\Delta_s/ds$ are easily computed from (34) and (42).

Proposition 6.1: The aforementioned algorithm computes $\bar{q}^\Delta_t$, $\bar{v}^\Delta_t$ and, hence, z^Δ_t (defined in (82)) for the approximate signal model (80).

Proof: Similar arguments to the proof of Lemma 2.2 for $t \in [i\Delta, (i+1)\Delta]$ yield

$$\bar{\mathbf{E}}^\Delta\left\{\Lambda^\Delta_{0,T}\phi(x_t)|\mathcal{Y}_T\right\} = \bar{\mathbf{E}}^\Delta\Big\{\Lambda^\Delta_{0,t}\phi(x_t)\bar{\mathbf{E}}^\Delta\Big\{\Lambda^\Delta_{t,T}|\mathcal{Y}_T \bigvee\{x_t\}\Big\}|\mathcal{Y}_T\Big\}.$$

However, the first equation shown at the bottom of the page holds. Therefore, $v^\Delta_t = \sum_{j=1}^K v^{(i+1)\Delta}_{i\Delta,k}$ is obtained by running a bank of K backward Kalman filters, each with non Gaussian initial condition $\bar{v}^{(i+1)\Delta}_{(i+1)\Delta,k} = \bar{v}^\Delta_{(i+1)\Delta}(x)I(x \in P_k)$. Converting to robust form as in (20) yields the backward equation in the previous algorithm. Hence, at time t, the second equation shown at the bottom of the page holds, where q^Δ_t is computed as in the previous algorithm. $\square$

$$\begin{aligned}\bar{\mathbf{E}}^\Delta\{\Lambda^\Delta_{t,T}|\mathcal{Y}_T, \bigvee\{x_t = x\}\} &= \bar{\mathbf{E}}^\Delta\Big\{\Lambda^\Delta_{t,(i+1)\Delta}\bar{\mathbf{E}}^\Delta\Big\{\Lambda^\Delta_{(i+1)\Delta,T}|Y_T\bigvee\{x_{(i+1)\Delta}|\mathcal{Y}_T\bigvee\{x_t\}\Big\}\Big\}\\ &= \bar{\mathbf{E}}^\Delta\Big\{\Lambda^\Delta_{t,(i+1)\Delta}v^\Delta_{(i+1)\Delta}(x_{(i+1)\Delta})|\mathcal{Y}_T\bigvee\{x_t\}\Big\}\\ &= \sum_{j=1}^K \bar{\mathbf{E}}^\Delta\Big\{\Lambda^\Delta_{t,(i+1)\Delta}I(x_{(i+1)\Delta} \in P_k)v^\Delta_{(i+1)\Delta}(x_{(i+1)\Delta})|\mathcal{Y}_T\bigvee\{x_t\}\Big\}.\end{aligned}$$

$$\begin{aligned}\bar{\mathbf{E}}^\Delta\{\Lambda^\Delta_{0,T}\phi(x_t)|\mathcal{Y}_T\} &= \bar{\mathbf{E}}^\Delta\left\{\Lambda^\Delta_{0,t}\, v^\Delta_t(x_t)\phi(x_t)|\mathcal{Y}_t\right\}\\ &= \bar{\mathbf{E}}^\Delta\left\{\Lambda^\Delta_{0,i\Delta}\, v^\Delta_t(x_t)\bar{\mathbf{E}}^\Delta\{\Lambda^\Delta_{i\Delta,t}\phi(x_t)|x_{i\Delta}\}|\mathcal{Y}_{0,t}\right\}\\ &= \sum_{k=1}^K \bar{\mathbf{E}}^\Delta\left\{\Lambda^\Delta_{0,i\Delta}v^\Delta(x_t)I(x_{i\Delta} \in P_k)\bar{\mathbf{E}}^\Delta\{\Lambda^\Delta_{i\Delta,t}\phi(x_t)|x_{i\Delta} \in P_k\}|Y_{0,t}\right\}\\ &= \langle\phi q^\Delta_t, v^\Delta_t\rangle\end{aligned}$$

B. Convergence of Approximate Smoother

It is convenient to introduce a double measure change. Define

$$N_T(\omega^1) = \exp\left[\int_0^T f(x,s)dx_s - \frac{1}{2}\int_0^T (f(x,s))^2 ds\right]$$
$$N_T^\Delta(\omega^1) = \exp\left[\int_0^T f^\Delta(x,s)dx_s - \frac{1}{2}\int_0^T (f^\Delta(x,s))^2 ds\right].$$

Let $\bar{\bar{P}}$ denote the standard Wiener measure. Girsanov's theorem implies that

$$\langle \pi_{t|T}, H_t\phi\rangle(\omega^2) = \bar{\bar{\mathbf{E}}}\{\Lambda_T N_T H_t \phi(x_t)|\mathcal{Y}_T\}$$
$$\langle \pi_{t|T}^\Delta, H_t\phi\rangle(\omega^2) = \bar{\bar{\mathbf{E}}}\{\Lambda_T^\Delta N_T^\Delta H_t \phi(x_t)|\mathcal{Y}_T\}.$$

Theorem 6.2: As $\Delta \to 0$, the smoothed estimate $\langle \pi_{t|T}^\Delta, H_t\phi\rangle(\omega^2) \to \langle \pi_{t|T}, H_t\phi\rangle(\omega^2)$, $\bar{\bar{P}}$ a.s. for all $t \in [0,T]$ and $\omega^2 \in \Omega^2$.

To prove the aforementioned theorem, we work with robust versions (i.e., Lipschitz continuous versions in ω^2) of $\langle \pi_{t|T}, H_t\phi\rangle(\omega^2)$ and $\langle \pi_{t|T}^\Delta, H_t\phi\rangle(\omega^2)$, $\omega^2 \in \Omega^2 = C(\mathbb{R}\times[0,T])$. In order to obtain these robust versions, we need to integrate by parts the stochastic (dy) integrals appearing in Λ_t(6) and Λ_t^Δ(81); see the Remark following Theorem 3.2. This, in turn, requires computation of the differentials of $h(x_s)$ and $h^\Delta(x,s)$.

Lemma 6.3: As $\Delta \to 0, \Lambda_T^\Delta(\cdot,\omega^2) \to \Lambda_T(\cdot,\omega^2)$, $\bar{\bar{P}}$ a.s. for all $\omega^2 \in \Omega^2$

Proof: Note that the gradient of h jumps across each boundary point B_k (see (73)). Thus, Ito's formula does not apply. Instead one needs to use Tanaka's formula for semi-martingales [15, Sec. 3.6] which yields $\bar{\bar{P}}$ a.s.

$$h(x_t) = h(x_0) + \int_0^t \nabla h(x_s)dx_s + \frac{1}{2}\sum_{k=1}^{K-1}\int_0^t I(x_s = B_k)dL_s^{B_k}(x) \quad (84)$$

where $\nabla h(x) = \sum_{k=1}^K I(x_s \in P_k)c_k$ and $L_s^{B_k}(x)$ denotes the local time at B_k of the process x_t. Thus, $\bar{\bar{P}}$ a.s.; see the first equation shown at the bottom of the page.

Consider the evaluation of Λ_t^Δ. It is easily shown that $\bar{\bar{P}}$ a.s.; see the second equation shown at the bottom of the page, where $\nabla h^\Delta(x,s) = \sum_{k=1}^K I(x_{i\Delta} \in P_k)c_k$ if $i\Delta \le s \le (i+1)\Delta$; $J(\Delta; i,t)$ is the set of integers j such that $j\Delta \le t$ and the line segment joining $x_{(i-1)\Delta}$ and $x_{i\Delta}$ intersects B_k. Finally, from Tanaka's formula it follows that (see [20, p. 435]) $\bar{\bar{P}}$ a.s. and $\forall \omega^2 \in \Omega^2$

$$\sum_{j\in J(\Delta;i,t)} y_{j\Delta}(\omega_2)(\Lambda_i(x_{j\Delta}) - \Lambda_i(x_{(j-1)\Delta}))(x_{j\Delta} - x) \to \int_0^t y_s(\omega^2) I(x_s = B_k)dL_s^{B_k}(x)$$

□

From the previous Lemma, it follows that $\Lambda_T^\Delta N_T^\Delta \to \Lambda_T N_T$, $\bar{\bar{P}}$ a.s. for all $\omega^2 \in \Omega^2$. Also $\forall \omega^2$, it can be shown [20] that

$$\bar{\bar{\mathbf{E}}}\left\{\left(N_T^\Delta(\cdot)Z_T^\Delta(\cdot,\omega^2)\right)^2\right\} \le c(\cdot,\omega^2) \quad \forall \Delta > 0$$

which implies that $\Lambda_T^\Delta(\cdot,\omega^2)N_T^\Delta$ is uniformly integrable. Hence, Theorem 6.2 follows.

Remarks: The numerical integration required for the non-Gaussian initial condition at each $t = i\Delta$ can be performed by sequential Monte Carlo methods (particle filters). Using the aforementioned result and the methods in [17], one can then show spatial convergence (as the number of particles go to infinity the integration becomes exact) and temporal convergence (as $\Delta \to 0$, the forward and backward estimates converge weakly to the true estimates; see Theorem 6.2). Finally, in the numerical implementation of the aforementioned algorithms we found that working with the logarithms of the normalization factors yielded better numerical behavior.

VII. Conclusion

In this paper, we have presented fixed-interval smoothers for continuous-time stochastic dynamical systems. The main contribution was to present these smoothers in robust form, that is, in terms of nonstochastic partial differential equations with random coefficients. The advantage of this robust formulation is that one not need to work with the technically complicated machinery of two-sided stochastic calculus. We are currently examining extension of the methods in this paper to particle filters [10] and smoothers where state estimates are computed by sequential Monte Carlo methods. Also, we are examining use of the piecewise linear filtering in bearings-only target tracking where a $\tan^{-1}(\cdot)$ nonlinearity is approximated by piecewise linear segments and then the bank of Kalman filters together with sequential Monte Carlo methods is used.

$$\Lambda_t = \exp\left\{y_t h(x_t) - \int_0^t y_s \nabla h(x_s)dx_s - \frac{1}{2}\sum_{k=1}^K \int_0^t y_s I(x_s = B_k)dL_s^{B_k}(x) - \frac{1}{2}\int_0^t (h(x_s))^2 ds\right\}.$$

$$\Lambda_t^\Delta = \exp\left\{y_t h^\Delta(x,t) - \int_0^t y_s \nabla h^\Delta(x,s)dx_s - \sum_{k=1}^{K-1}\sum_{j\in J(\Delta;i,t)} y_{j\Delta}\left(\Lambda_i(x_{j\Delta}) - \Lambda_i(x_{(j-1)\Delta})\right)(x_{j\Delta} - x) - \frac{1}{2}\int_0^t |h^\Delta(x,s)|^2 ds\right\}$$

APPENDIX
VERIFICATION OF ROBUST BENES SMOOTHER EQUATIONS

Forward Filter: We start with the result from [2, Eq. 6.2.22] which in strong form states that

$$q_t(x) = (2\pi)^{-m/2} \int_{\mathbb{R}^m} \exp(-\psi(\zeta,0))\, \tilde{s}_t(\zeta)\, \tilde{q}_t(x,\zeta)\, \pi_0(\zeta)\, d\zeta$$

where with $\delta(\cdot)$ denoting the Dirac delta function, $\tilde{q}_0(x,\zeta) = \exp(\psi(x,0))\delta(x-\zeta)$, the first equation shown at the bottom of the page holds true, and $m_t(\zeta)$ satisfies [2, Eq. 6.2.20]. Then, substituting $\bar{q}_t = \bar{\epsilon}_t q_t$, we have

$$\bar{q}_t(x) = (2\pi)^{-m/2} \int_{\mathbb{R}^m} \exp(-\psi(\zeta,0))\, \tilde{s}_t(\zeta)\, \tilde{q}_t(x,\zeta)\, \pi_0(\zeta)\, d\zeta$$

$$\bar{q}_0(\cdot) = \pi_0(\cdot) \quad (85)$$

where $\bar{r}_t(\zeta) = \Sigma_t^{-1} m_t(\zeta) - C' y_t$ and $\tilde{q}_0(x,\zeta) = q_0(x,\zeta)$, the second equation shown at the bottom of the page holds. Then, using the result [2, Th. 6.2.20, p. 200] that $m_t(\zeta) = \Phi_t \zeta + \tilde{m}_t$ where Φ_t satisfies (38) and

$$d\tilde{m}_t = [(F_t - \Sigma_t\Gamma_t)\tilde{m}_t - \Sigma_t\mu_t]dt + \Sigma_t C_t'(dy_t - C\tilde{m}_t dt) \quad (86)$$

we obtain (37) for $\bar{r}_t$ where $\tilde{r}_t$ satisfies (39). We also obtain (87)–(89), as shown at the bottom of the page. Integrating the first term on the right-hand side of (89) by parts yields

$$\int_0^t \Phi_s' C' dy_s = \Phi_t' C' y_t - \int_0^t \frac{d\Phi_s'}{ds} C' y_s ds.$$

Substituting (38) in the aforementioned equation, (89) yields

$$\begin{aligned}\tilde{\rho}_t =& \Phi_t' C' y_t - \int_0^t \Phi_s' F_s' C' y_s ds \\ &- \int_0^t \Phi_s'(\Gamma_s + C'C)\Sigma_s \tilde{r}_s ds - \int_0^t \Phi_s' \mu_s ds \\ \tilde{\rho}_0 =& 0.\end{aligned}$$

The first subterm in the normalization constant $\tilde{K}_t$ (88) can be expressed as follows:

$$\int_0^t \tilde{r}_s' \Sigma_s C' dy_s = \tilde{r}_t' \Sigma_t C' y_t - \int_0^t \frac{d}{ds}[\tilde{r}_s'\Sigma_s] C' y_s ds$$

where, from (39) and (40)

$$\begin{aligned}\frac{d}{ds}[\tilde{r}_s'\Sigma_s] =& -\Sigma_t(\Gamma + C'C)\Sigma_t\tilde{r}_t \\ &+ F\Sigma_t\tilde{r}_t - QC'y_t - \Sigma_t F' C' y_t + \Sigma_t\mu_t.\end{aligned}$$

The other subterm in the first term of $\tilde{K}_t$ is $C\int_0^t \Sigma_s y_s dy_s C'$. Ito calculus yields

$$\begin{aligned}\int_0^t \Sigma_s y_s dy_s =& \Sigma_t y_t^2 - \int_0^t y_s d(\Sigma_s y_s) - [\Sigma y, y]_t \\ =& \Sigma_t y_t^2 - \int_0^t \frac{d\Sigma_s}{ds} y_s^2 ds - \int_0^t \Sigma_s y_s dy_s - \int_0^t \Sigma_s ds\end{aligned}$$

where $[\cdot,\cdot]_t$ denotes quadratic variation. Thus

$$\int_0^t \Sigma_s y_s ds = \frac{1}{2}\left(\Sigma_t y_t^2 - \int_0^t \frac{d\Sigma_s}{ds} y_s^2 ds - \int_0^t \Sigma_s ds\right).$$

Substituting these expressions into (85), all the terms involving y_t (outside of time integrals) cancel out, and we obtain the robust forward Benes filter (34).

Backward Filter: In complete analogy to the forward filter, the backward Benes filter is of the form

$$\bar{v}_t(x) = \int_{\mathbb{R}^m} \exp(\psi(\zeta,T))\tilde{\bar{s}}_t(\zeta)\tilde{v}_t(x,\zeta)d\zeta \quad \bar{v}_T(x) = \epsilon_T(x).$$

We will derive expressions for $\tilde{v}_t(x,\zeta)$ (step 1) and $\tilde{\bar{s}}_t$ (step 2) as follows.

$$\begin{aligned}\tilde{q}_t(x,\zeta) =& \exp\left(\psi(x,t) - \frac{1}{2}(x - m_t(\zeta))'\Sigma_t^{-1}(x - m_t(\zeta))\right) \\ \tilde{s}_t(\zeta) =& \exp\left\{\int_0^t m_s'(\zeta) C' dy_s - \frac{1}{2}\int_0^t [m_s'(\zeta)(\Gamma_s + C'C)m_s(\zeta) + 2m_s'(\zeta)\mu_s + \mathrm{Tr}[\Sigma_s\Gamma_s] + 2\kappa_s]\, ds\right\}\end{aligned}$$

$$\tilde{q}_t(x,\zeta) = \exp\left(\psi(x,t) - \frac{1}{2}x'(\Sigma_t^{-1} - C'Ct)x + \bar{r}_t'(\zeta)x - \frac{1}{2}(\bar{r}_t(\zeta) + C'y_t)'\Sigma_t(\bar{r}_t(\zeta) + C'y_t)\right).$$

$$\tilde{s}_t(\zeta) = \tilde{K}_t \exp(-\frac{1}{2}\zeta' S_t \zeta + \zeta'\tilde{\rho}_t) \quad (87)$$

$$\begin{aligned}\tilde{K}_t =& \exp\Bigg(\int_0^t (\tilde{r}_s + C'y_s)'\Sigma_s C' dy_s \\ &- \frac{1}{2}\int_0^t \Big[(\tilde{r}_s + C'y_s)'\Sigma_s(\Gamma + C'C)\Sigma_s(\tilde{r}_s + C'y_s) + \mathrm{Tr}[\Sigma_s\Gamma] + 2\kappa_s + (\tilde{r}_s + C'y_s)'\Sigma_s\mu_s\Big] ds\Bigg)\end{aligned} \quad (88)$$

$$\tilde{\rho}_t = \int_0^t \Phi_s' C' dy_s - \int_0^t \Phi_s'(\Gamma + C'C)\tilde{m}_s ds - \int_0^t \Phi_s'\mu_s ds. \quad (89)$$

Step 1) $\tilde{v}_t(x,\zeta)$ by definition is the explicit solution to (20) with terminal condition $\tilde{v}_T(x,\zeta) = \exp(-\psi(x,T))\delta(x,\zeta)\epsilon_T$. It is shown as follows that $\tilde{v}_t(x,\zeta)$ satisfies the first equation shown at the bottom of the page. In particular, we will show that

$$\tilde{v}_t(x,\zeta) = \exp\left(-\psi(x,t) - \frac{1}{2}x'\bar{N}_t x + \bar{l}_t' x + \bar{\zeta}_t\right) \tag{90}$$

where $\bar{N}_t = \bar{\Sigma}_t^{-1} + C'Ct, \bar{N}_t, \bar{l}_t$ and $\bar{\zeta}_t$ are given by the ordinary differential equations

$$\begin{aligned} \frac{d\bar{N}_t}{dt} =& (\bar{N}_t - C'Ct)Q(\bar{N}_t - C'Ct) - \Gamma_t \\ &- F_t'(\bar{N}_t - C'Ct) - (\bar{N}_t - C'Ct)F_t \qquad (91)\\ \frac{d\bar{l}_t}{dt} =& \mu_t + (\bar{N}_t - C'Ct)Q(\bar{l}_t - C'y_t) \\ &- F_t'(\bar{l}_t - C'y_t) \qquad (92)\\ \frac{d\bar{\zeta}_t}{dt} =& \kappa_t - \frac{1}{2}(\bar{l}_t - C'y_t)'Q(\bar{l}_t - C'y_t) \\ &+ \frac{1}{2}\mathrm{Tr}\left[Q(\bar{N}_t - C'Ct)\right]. \qquad (93) \end{aligned}$$

Evaluating the right-hand side of (20), we have (writing $\tilde{v}(x,\zeta)$ as $\tilde{v}$ for convenience)

$$\begin{aligned} \nabla\bar{\epsilon}_t =& \bar{\epsilon}_t\left[C'Cxt - Cy_t\right] \\ \epsilon_t\nabla\left[\bar{\epsilon}_t\tilde{v}_t\right] =& \left[C'Cxt - Cy_t\right]\tilde{v}_t + \nabla\tilde{v}_t \\ \epsilon_t\nabla^2\left[\bar{\epsilon}_t\tilde{v}_t\right] =& \tilde{v}_t\left[C'Cxt - Cy_t\right]\left[C'Cxt - Cy_t\right]' \\ &+ \left[C'Cxt - Cy_t\right]\nabla'\tilde{v}_t + \tilde{v}_t tC'C \\ &+ \nabla\tilde{v}_t\left[C'Cxt - Cy_t\right]' + \nabla^2\tilde{v}_t. \end{aligned}$$

Consequently, (20) reads

$$\begin{aligned} \frac{\partial\tilde{v}}{\partial t} =& -f'\left[C'Cxt - Cy_t\right]\tilde{v}_t - f'\nabla\tilde{v} \\ &- \frac{1}{2}\tilde{v}\left[C'Cxt - Cy_t\right]'Q\left[C'Cxt - Cy_t\right] \\ &- \left[C'Cxt - Cy_t\right]'Q\nabla\tilde{v} - \frac{1}{2}\tilde{v}tCQC' \\ &- \frac{1}{2}\mathrm{Tr}\left[\nabla^2\tilde{v}Q\right]. \qquad (94) \end{aligned}$$

We need to find an explicit solution to the above pde. In analogy to the solution of the forward robust Benes filter, we can show that (90) is an explicit solution to the aforementioned pde. To see this, we first note that

$$\begin{aligned} \nabla\tilde{v} =& \tilde{v}\left[-\nabla\psi - \bar{N}x + \bar{l}\right] \\ \nabla^2\tilde{v} =& \tilde{v}\left[-\nabla\psi - \bar{N}x + \bar{l}\right]\left[-\nabla\psi - \bar{N}x + \bar{l}\right]' \\ &+ \tilde{v}\left[-\nabla^2\psi - \bar{N}\right] \\ \frac{\partial\tilde{v}}{\partial t} =& \tilde{v}\left[-\frac{d\psi}{dt} - \frac{1}{2}x'\frac{d\bar{N}}{dt}x + \frac{d\bar{l}'}{dt}x + \frac{d\bar{\zeta}}{dt}\right]. \end{aligned}$$

Substituting these expressions into (94) yields

$$\begin{aligned} &-\left[\frac{\partial\psi}{\partial t} + \frac{1}{2}\mathrm{Tr}\left(Q_t\nabla^2\psi\right) + \frac{1}{2}(\nabla\psi)'Q_t\nabla\psi + x'F_t'\nabla\psi\right] \\ =& \frac{1}{2}x'\frac{d\bar{N}}{dt}x - \frac{d\bar{l}'}{dt}x - \frac{d\bar{\zeta}}{dt} - x'F_t'\left[C'Cxt - Cy_t\right] \\ &+ x'F_t'\bar{N}_t x - x'F_t'\bar{l}_t \\ &- \frac{1}{2}\left[C'Cxt - Cy_t\right]'Q\left[C'Cxt - Cy_t\right] \\ &+ \left[C'Cxt - Cy_t\right]'QNx - \left[C'Cxt - Cy_t\right]'Qr - \frac{1}{2}tCQC' \\ &- \frac{1}{2}x'\bar{N}_tQ\bar{N}_t x + x'\bar{N}_tQ\bar{l}_t - \frac{1}{2}\bar{l}_t'Q\bar{l}_t + \frac{1}{2}\mathrm{Tr}\left[\bar{N}_tQ\right] \\ &. \end{aligned} \tag{95}$$

However, according to the assumption on the Benes nonlinearity (33), the left-hand side of the previous equation is $-[1/2x'\Gamma_t x + x'\mu_t + \kappa_t]$. Equating coefficients of the terms in x^2, x and constants, yields the above robust Benes backward filter (91), (92).

In analogy to the forward filter set $\bar{l}_t(\zeta) = \bar{\Sigma}_t^{-1}\Phi_t\zeta + \bar{l}_t$. Subsituting this into (91) and using (92) yields (46).

Step 2) From [2, pp. 130—134], in complete analogy to the forward Benes filter it follows that the second equation shown at the bottom of the page holds, where the first integral is a backward Ito integral, $\bar{m}_s(\zeta) \triangleq \bar{\Sigma}_s(l_s(\zeta) - C'y_s)$. Substituting $\bar{l}_s(\zeta) = \bar{\Sigma}_s^{-1}\Phi_s\zeta + \bar{l}_s$ yields

$$\tilde{\bar{s}}_t(\zeta) = \tilde{\bar{K}}_t\exp(-\frac{1}{2}\zeta'\bar{S}_t\zeta + \zeta'\tilde{\bar{\rho}}_t)$$

where

$$\begin{aligned} \tilde{\bar{K}}_t =& \exp\bigg(\int_t^T(\tilde{l}_s - C'y_s)'\bar{\Sigma}_sC'dy_s \\ &- \frac{1}{2}\int_t^T\bigg[(\tilde{l}_s - C'y_s)'\bar{\Sigma}_s(\Gamma + C'C)\bar{\Sigma}_s(\tilde{l}_s - C'y_s) \\ &+ \mathrm{Tr}[\bar{\Sigma}_s\Gamma] + 2\kappa_s + (\tilde{l}_s - C'y_s)'\bar{\Sigma}_s\mu_s\bigg]ds\bigg) \end{aligned}$$

$$\tilde{v}_t(x,\zeta) = \exp\left(-\psi(x,t) - \frac{1}{2}x'(\bar{\Sigma}_t^{-1} + C'Ct)x + \bar{l}_t'(\zeta)x - \frac{1}{2}(\bar{l}_t(\zeta) - C'y_t)'\bar{\Sigma}_t(\bar{l}_t(\zeta) - C'y_t)\right).$$

$$\tilde{\bar{s}}_t(\zeta) = \exp\bigg\{\int_t^T\bar{m}_s'(\zeta)C'dy_s - \frac{1}{2}\int_t^T\bigg[\bar{m}_s'(\zeta)(\Gamma_s + C'C)\bar{m}_s(\zeta) + 2\bar{m}_s'(\zeta)\mu_s + \mathrm{Tr}[\Sigma_s\Gamma_s] + 2\kappa_s\bigg]ds\bigg\}$$

$$\begin{aligned}\tilde{\bar{\rho}}_t =& \bar{\Phi}'_T C' y_T - \bar{\Phi}'_t C' y_t \\ &- \int_t^T \bar{\Phi}'_s F'_s C' y_s ds - \int_t^T \bar{\Phi}'_s \mu_s ds \\ &- \int_t^T \bar{\Phi}'_s (\Gamma_s + C'C) \bar{\Sigma}_s \tilde{l}_s ds \\ \tilde{\bar{\rho}}_T =& 0.\end{aligned}$$

For deterministic integrands, the forward and backward Ito rules are identical. Hence, the first subterm in $\tilde{\bar{K}}_t$ is

$$\int_t^T \tilde{l}'_s \bar{\Sigma}_s C' dy_s = \tilde{l}_T \bar{\Sigma}_T C' y_T - \tilde{l}_t \bar{\Sigma}_t C' y_t - \int_t^T \frac{d(\bar{\Sigma}_s \tilde{l}_s)'}{ds} C' y_s ds.$$

Applying the backward Ito formula [2, p. 124] to the second subterm of the first term yields

$$C \int_t^T \bar{\Sigma}_s y_s dy_s C' = \frac{C}{2} \left[\bar{\Sigma}_T y_T^2 - \bar{\Sigma}_t y_t^2 + \int_t^T \left(\bar{\Sigma}_s - \frac{d\bar{\Sigma}_s}{ds} y_s^2 \right) ds \right] C'.$$

Substituting these expressions into (42), all the terms involving y_t (outside of time integrals) cancel out and we obtain the robust backward Benes filter (42).

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers and P. Malcolm of the Defense Science and Technology Organization (DSTO) Adelaide, Australia, for proofreading the manuscript. The second author would like to thank the Department of Applied Mathematics, University of Adelaide, Australia, for its hospitality.

REFERENCES

[1] V. Benes and I. Karatzas, "Filtering for piecewise linear drift and observation," *Proc. 20th IEEE Conf. Decision Control*, pp. 583–589, 1981.

[2] A. Bensoussan, *Stochastic Control of Partially Observable Systems*, Cambridge, U.K.: Cambridge Univ. Press, 1992.

[3] D. Bertsekas, *Nonlinear Programming*. Belmont, MA: Athena Scientific, 2000.

[4] F. Campillo and F. LeGland, "MLE for partially observed diffusions : Direct minimization vs. the EM algorithm," *Stoch. Processes Appl.*, vol. 33, pp. 245–274, 1989.

[5] S. H. Chung, V. Krishnamurthy, and J. B. Moore, "Adaptive processing techniques based on HMMs for characterising very small channel currents buried in noise and deterministic interferences," *Phil. Trans. Royal Soc., London B*, vol. 334, pp. 357–384, 1991.

[6] J. M. C. Clark, "The design of robust approximations to the stochastic differential equations of nonlinear filtering," in *Communication Systems and Random Processes Theory*, J. K. Skwirzynski, Ed, The Netherlands: Sijhtjoff and Noordhoff, 1977.

[7] F. E. Daum, "Exact finite dimensional nonlinear filters," *IEEE Trans. Automat. Contr.*, vol. AC-31, pp. 616–622, July 1986.

[8] M. H. A. Davis, "On a multiplicative functional transformation arising in nonlinear filtering theory," *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, vol. 54, pp. 125–139, 1980.

[9] A. Dembo and O. Zeitouni, "Parameter estimation of partially observed continuous time stochastic processes via the EM algorithm (Corrigendum)," *Stoch. Processes Appl.*, vol. 31, no. 1, pp. 167–169, 1989.

[10] A. Doucet, N. Gordon, and V. Krishnamurthy, "Particle filters for state estimation of jump Markov linear systems," *IEEE Trans. Signal Processing*, vol. 49, pp. 613–624, Mar. 2001.

[11] R. J. Elliott, L. Aggoun, and J. B. Moore, *Hidden Markov Models-Estimation and Control*. New York: Springer-Verlag, 1995.

[12] R. J. Elliott and V. Krishnamurthy, "Exact finite-dimensional filters for maximum likelihood parameter estimation of continuous-time linear gaussian systems," *SIAM J. Control Optim.*, vol. 35, no. 6, pp. 1908–1923, Nov. 1997.

[13] R. J. Elliott, V. Krishnamurthy, and H. V. Poor, "Exact filters for certain moments and stochastic integrals of the state of systems with benes nonlinearity," *IEEE Trans. Automat. Contr.*, vol. 44, pp. 1929–1933, Oct 1999.

[14] M. R. James, V. Krishnamurthy, and F. LeGland, "Time discretization of continuous-time filters and smoothers for HMM parameter estimation," *IEEE Trans. Inform. Theory*, vol. 42, no. (2, pp. 593–605, Mar. 1996.

[15] I. Karatzas and S. Shreve, *Brownian Motion and Stochastic Calculus*, 2nd ed. New York: Springer-Verlag, 1991.

[16] H. J. Kushner, "A robust discrete state approximation to the optimal nonlinear filter for a diffusion," *Stochastics*, vol. 3, no. 2, pp. 75–83, 1979.

[17] P. Del Moral, J. Jacod, and P. Protter, "The Monte-Carlo method for filtering with discrete-time observations," *Probab. Theory Related Fields*, vol. 120, pp. 346–368, 2001.

[18] E. Pardoux, "Equations du filtrage nonlineaire de la prediction et du lissage," *Stochastics*, vol. 6, pp. 193–231, 1982.

[19] E. Pardoux and P. Protter, "A two sided stochastic integral and its calculus," *Probab. Theory Related Fields*, vol. 76, pp. 15–49, 1987.

[20] E. Pardoux and C. Savona, "Piecewise linear filtering," in *Stochastic Differential Systems, Stochastic Control and Applications*. New York: Springer-Verlag, 1988, vol. 10 , IMA, pp. 429–437.

[21] C. Savona, "Approximate nonlinear filtering for piecewise linear systems," *Syst. Control Lett.*, vol. 11, pp. 327–332, 1988.

[22] E. Wong and B. Hajek, *Stochastic Processes in Engineering Systems*, second ed. Berlin, Germany: Springer-Verlag, 1985.

[23] C. F. J. Wu, "On the convergence properties of the EM algorithm," *Ann. Statist.*, vol. 11, no. 1, pp. 95–103, 1983.

Vikram Krishnamurthy (S'90–M'91–SM'99) was born in 1966. He received the B.S. degree in electrical engineering from the University of Auckland, Auckland, New Zealand, and the Ph.D. degree from the Australian National University, Canberra, Australia, in 1988 and 1992, respectively.

Since July 2002, he has been a Professor and Canada Research Chair in the Department of Electrical Engineering, University of British Columbia, Vancouver, Canada. Prior to this, he was a Professor in the Department of Electrical and Electronic Engineering, University of Melbourne, Australia. His research interests span several areas, including stochastic scheduling and network optimization, time-series analysis, and statistical signal processing. He is an Associate Editor of *Systems and Control Letters*.

Dr. Krishnamurthy has served on the Technical Program Committee of several conferences, including the 37th IEEE Conference on Decision and Control, Tampa, FL (1998), and the IFAC Symposium on System Identification, Santa Barbara, CA (2000). He is currently an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING.

Robert Elliott received the B.S. and M.S. degrees from Oxford University, Oxford, U.K., and the Ph.D. and D.Sc. degrees from Cambridge University, Cambridge, U.K.

He has held positions and visiting positions at various institutions around the world. Currently, he is the Royal Bank Professor of Finance at the University of Calgary, Calgary, AB, Canada, and is an Adjunct Professor in the Department of Mathematics. He has authored six books and over 300 papers, including the books *Hidden Markov Models: Estimation and Control* (New York: Springer-Verlag, 1995; reprinted 1997), coauthored with L. Aggoun and J. Moore, and *Mathematics of Financial Markets* (New York: Springer-Verlag, 1999), coauthored with P. E. Kopp and reprinted three times. His work in recent years has investigated stochastic processes in engineering and finance.