Robust Continuous-time Smoothers – without two-sided stochastic integrals

Vikram Krishnamurthy
Department of Electrical and Electronic Engineering,
University of Melbourne,
Parkville, Victoria 3052, Australia
vikram@mullian.ee.mu.oz.au

Robert Elliott
Department of Mathematical Sciences,
University of Alberta, Edmonton, T6G 2G1, Canada.
r.elliott@ualberta.ca

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 non-stochastic 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 Hidden Markov Model smoothers are derived – these finite dimensional smoothers do not involve stochastic integrals.

1 Introduction

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 [1]. Derivation of the fixed-interval smoothed state density is mathematically more formidable as it requires the use of two sided stochastic calculus [12].

In this paper we derive *robust* filters and smoothers for the state of a continuous-time stochastic dynamical system. 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 non-stochastic 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 non-stochastic differential equation with random coefficients was first developed in [3], see also [11], [5] or Chapter 4 of [1]. More recently, in [10] versions of these robust filters, probabilistic interpretations and implicit and explicit discretization schemes were developed for continuous-time Hidden Markov models.

The contributions of this paper are as follows:
1. It is shown that the smoothed state estimate can be computed via a robust forward and backward filters. Each of these filters involve non-stochastic parabolic partial differential equations.

2. Robust fixed interval smoothed estimates of functionals of the state of the system are derived. Again the equations involve non-stochastic integrals. These robust smoothers can be used in maximum likelihood parameter estimation via the Expectation Maximization (EM) algorithm.

3. As examples of the robust smoothers for the state and functionals of the state, state and maximum likelihood parameter estimation for Hidden Markov Models.

2 Model and Problem Formulation

2.1 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 $\{x_t\}$, $t \geq 0$ and the observation process $\{y_t\}$, $t \geq 0$ are described by

$$dx_t = f_\theta(x_t,t)\,dt + \sigma_\theta(x_t,t)dw_t, \qquad x_0 \in \mathbb{R}^m \qquad (1)$$
$$dy_t = h_\theta(x_t,t)\,dt + dv_t, \qquad y_0 = 0 \in \mathbb{R}^n \qquad (2)$$

Define the right-continuous filtrations $\mathcal{F}_t = \sigma(x_s,\ s \leq t)$, $\mathcal{G}_t = \sigma\{x_s, y_s : s \leq t\}$, $\mathcal{Y}_t = \sigma\{y_s : s \leq t\}$ for $t \in [0,T]$. In (1) and (2) w and v are independent standard Brownian motions. (In Sec.4, we will consider the Hidden Markov Model case where w_t is a $\mathcal{F}_t$ measurable finite state martingale increment process). Further, w and v are independent of x_0. We assume that x_0 is a random variable with normal density $\pi_0(x)$ which is $N(\hat{x}_0, P_0)$.

We make the following standard assumptions [1] for all $\theta \in \Theta$:
1. $f_\theta : \mathbb{R}^m \times [0,T] \to \mathbb{R}^m$ and $h_\theta : \mathbb{R}^m \times [0,T] \to \mathbb{R}^n$ denote bounded measurable functions and $T > 0$ denotes a fixed real number.
2. $\sigma_\theta : \mathbb{R}^m \times [0,T] \to \mathbb{R}^{m\times r}$ 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 $\alpha > 0$.

The existence and uniqueness of a weak solution (in the distributional sense) to the state equation (1) on $C^m[0,T]$ follow from the assumptions 1 and 2 above. It is assumed that f_θ, σ_θ and h_θ are continuously differentiable with respect to the parameter θ and that the derivative $\partial f_\theta/\partial\theta$ and $\partial h_\theta/\theta$ are measurable and bounded functions.

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 described in Sec.1, this is a useful property from an implementation point of view. The aim of this paper is three-fold:
(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

$$H_t = H_0 + \int_0^t \alpha(x_s,y_s)ds + \int_0^t \beta'(x_s,y_s)dx_s + \int_0^t \gamma'(x_s)dy_s \qquad (3)$$

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. 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. Such computations arise in computing the maximum likelihood parameter estimate via the EM algorithm – see Sec.2.2. The same problem is considered in [2] 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 (3). One would have liked to have interchanged the conditional expectation and the integral. However the resulting expression

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

is not an Ito integral since the integrand is not adapted to the filtration $\{\mathcal{Y}_t : 0 \leq t \leq T\}$. In [2], it is shown that the above integral can be interpreted as a Skorohod integral and requires the use of two-sided stochastic calculus. The above integral can also be interpreted as a generalized Stratonovich integral, see [6].

In Sec.3, it will demonstrated that by expressing the filters in robust form, the smoothed estimate $\mathbf{E}\{H_t|\mathcal{Y}_T\}$ can be computed using ordinary (non-stochastic) integration.
(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

$$\mathcal{L}(\theta) = \mathbf{E}_0\{\log \frac{dP_\theta}{dP_0} \mid \mathcal{Y}_T\}\ ,$$

and the MLE is defined by $\hat{\theta} \in \operatorname{argmax}_{\theta\in\Theta} \mathcal{L}(\theta)$. Application of the EM algorithm to Hidden Markov Models is covered in Sec.4.2.

2.2 Motivation: The EM Algorithm

As mentioned above 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 (3). The EM algorithm is an iterative numerical method for computing the MLE. Let $\hat{\theta}_0$ 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 $Q(\cdot,\tilde{\theta})$, where $\mathcal{Q}(\theta,\tilde{\theta}) = \mathbf{E}_{\tilde{\theta}}\{\log \frac{dP_\theta}{dP_{\tilde{\theta}}} \mid \mathcal{Y}_T\}$.
Step 2. (M–step) Find $\widehat{\theta}_{j+1} \in \operatorname*{argmax}_{\theta\in\Theta} \mathcal{Q}(\theta,\theta_j)$.

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

It is shown in [2] that $\mathcal{Q}(\theta,\tilde{\theta}) = \mathbf{E}_{\tilde{\theta}}\left\{\log \Lambda^{\theta\tilde{\theta}} \mid \mathcal{Y}_T\right\}$

where

$$\log \Lambda^{\theta\tilde{\theta}} = \int_0^T \left[f_\theta(x_s) - f_{\tilde{\theta}}(x_s)\right]' Q^{-1}(dx_s - f_{\tilde{\theta}}(x_s)ds)$$
$$- \frac{1}{2}\int_0^T \left[f_\theta(x_s) - f_{\tilde{\theta}}(x_s)\right]' Q^{-1}\left[f_\theta(x_s) - f_{\tilde{\theta}}(x_s)\right] ds$$
$$+ \int_0^T \left[h_\theta(x_s) - h_{\tilde{\theta}}(x_s)\right]\left[dy_s - h_\theta(x_s)ds\right]$$
$$- \frac{1}{2}\int_0^T \left[h_\theta(x_s) - h_{\tilde{\theta}}(x_s)\right]'\left[h_\theta(x_s) - h_{\tilde{\theta}}(x_s)\right] ds \tag{4}$$

It is clear from (4) that computing $\mathcal{Q}(\theta, \tilde{\theta})$ involves computing fixed interval smoothed estimate of functionals of the state of the form H_t in (3).

2.3 Preliminaries

To simplify notation, reference to the parameter θ will be dropped until Sec.4. 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

$$\Lambda_{t_1,t_2} = \exp\left(\int_{t_1}^{t_2} (h'(x_s,s)\,dy_s - \frac{1}{2}\int_{t_1}^{t_2} h'(x_s,s)\,h(x_s,s)\,ds\right), \quad t_1, t_2 \in [0, T] \tag{5}$$

For notational convenience define $\Lambda_t = \Lambda_{0,t}$. Then $d\Lambda_t = \Lambda_t\, h'(x_t,t)dy_t$ 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 $\frac{dP}{d\bar{P}}\big|_{\mathcal{G}_t} = \Lambda_t$ then Girsanov's theorem [7] implies that under P, v_t is a standard n-dimensional Brownian motion if we define $dv_t = dy_t - h(x_t,t)\,dt), \quad v_0 = 0$. That is, under P, $dy_t = h(x_t,t)\,dt + dv_t$. 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 $\phi : \mathbb{R}^m \to \mathbb{R}$ is an arbitrary "test" function, which is in $\mathbb{C}^2$ and has compact support. For any $\gamma_t(x)$, $\delta_t(x) \in L_2([0,T] \times \mathbb{R}^m)$ define the scalar product

$$\langle \gamma_t, \delta_t \rangle \triangleq \int_{\mathbb{R}^m} \gamma_t(x)\delta_t(x)dx \tag{6}$$

Filtering is concerned with computing $\mathbf{E}\{\phi(x_t)|\mathcal{Y}_t\}$. The following result is standard [7].

Lemma 2.1 *Suppose the measure valued process $\bar{\mathbf{E}}\{\Lambda_t\,\phi(x_t)|\mathcal{Y}_t\}$ has a $\mathcal{Y}_t$ measurable density function $q : [0, T] \times \mathbb{R}^m \times \Omega \to \mathbb{R}$. Then*

$$\mathbf{E}\{\phi(x_t)|\mathcal{Y}_t\} = \frac{\bar{\mathbf{E}}\{\Lambda_t\,\phi(x_t)|\mathcal{Y}_t\}}{\bar{\mathbf{E}}\{\Lambda_t|\mathcal{Y}_t\}} = \frac{\langle \phi,\, q_t\rangle}{\langle 1,\, q_t\rangle} \tag{7}$$

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 measured value process

$$v_t(x) = \bar{\mathbf{E}}\left\{\Lambda_{t,T}|\mathcal{Y}_t \bigvee \{x_t = x\}\right\}, \quad t \in [0,T],\ v_T(x) = 1. \tag{8}$$

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

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

$$\mathbf{E}\{\phi(x_t)|\mathcal{Y}_T\} = \frac{\langle \phi q_t,\ v_t\rangle}{\langle q_t,\ v_t\rangle} \tag{9}$$

Proof:

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

where $\mathcal{Y}_T \bigvee \mathcal{G}_t$ denotes the sigma algebra generated by $Y_T, \mathcal{G}_t$. Now $\bar{\mathbf{E}}\{\Lambda_{t,T}|\mathcal{Y}_T \bigvee \mathcal{G}_t\} = \bar{\mathbf{E}}\{\Lambda_{t,T}|\mathcal{Y}_T \bigvee \{x_t\}\} = v_t(x_t)$ by the Markovian property of the process $\Lambda_{t,T}$. Therefore,

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

■

3 Robust Fixed Interval Smoothing

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

$$\epsilon_t = \exp\left[h(x)'y_t - \frac{1}{2}h(x)'h(x)t\right], \qquad \bar{\epsilon}_t = 1/\epsilon_t \tag{10}$$

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

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

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

$$L(\phi) = \frac{1}{2}\text{Tr}[Q\nabla^2\phi] + f'\nabla\phi$$
$$L^*(\phi) = \frac{1}{2}\text{Tr}[Q\nabla^2\phi] - \text{div}[f\phi]$$

3.1 Robust Fixed Interval State Smoothers

We start with the following well known Duncan-Mortenson-Zakai (DMZ) equation, which describes the evolution of the un-normalized filtered state density, see for example [1] for a proof.

Theorem 3.1 *(DMZ equation) The un-normalized filtered density $q_t(x)$ evolves according to the following DMZ equation*

$$q_t(x) = q_0(x) + \int_0^t L^* \left(q_s(x)\right) ds + \int_0^t h(x) q_s(x) dy_s \tag{11}$$

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

$$\bar{q}_t(x) \stackrel{\triangle}{=} \bar{\epsilon}_t \, q_t(x), \qquad \bar{q}_0(x) = q_0(x) \tag{12}$$

Also let $\|y\| = \sup_{0 \le t \le T} |y(t)|$. The result below is proved in [11].

Theorem 3.2 (Robust Forward filter) *$\bar{q}_t$ satisfies the following non-stochastic parabolic partial differential equation*

$$\frac{\partial \bar{q}_t(x)}{dt} = \bar{\epsilon}_t L^*(\epsilon_t \bar{q}_t) \tag{13}$$

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

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

Define now the robust backward filtered density as

$$\bar{v}_t(x) = \epsilon_t v_t(x) \tag{14}$$

The theorem below 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) *$\bar{v}_t$ satisfies the non-stochastic backward parabolic pde*

$$\frac{\partial \bar{v}_t}{\partial t} = -\epsilon_t L(\bar{\epsilon}_t \bar{v}_t), \qquad \bar{v}_T(x) = \epsilon_T \tag{15}$$

Remark: (15) can be derived by starting with a backward Ito stochastic differential equation for v_t and then applying a robust transformation. However, the following straightforward proof derives smoothers without recourse to backward stochastic calculus.

Proof: Choose $\phi(x) = 1$ in (9). This yields $\langle q_t, v_t \rangle = \mathbf{E}\{\Lambda_T | \mathcal{Y}_T\}$ which means that $\langle q_t, v_t \rangle$ is independent of time t. Now from (12) and (14) 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 $d\langle \bar{q}_t, \bar{v}_t \rangle / dt = 0$. But

$$\frac{d}{dt} \langle \bar{q}_t, \bar{v}_t \rangle = \langle \frac{\partial \bar{q}}{\partial t}, \bar{v}_t \rangle + \langle \bar{q}_t, \frac{\partial \bar{v}_t}{dt} \rangle$$
$$= \langle \bar{\epsilon}_t L^* \left(\epsilon_t \bar{q}_t\right), \ \bar{v}_t \rangle + \langle \bar{q}_t, \frac{\partial \bar{v}_t}{dt} \rangle = \langle \bar{q}_t, \ \epsilon_t L \left(\bar{\epsilon}_t \bar{v}_t\right) \rangle + \langle \bar{q}_t, \frac{\partial \bar{v}_t}{dt} \rangle$$

which means that $\bar{v}_t$ satisfies the backward non-stochastic parabolic pde (15). ∎

Theorem 3.4 *In terms of the robust forward and backward filtered densities, the fixed interval smoothed estimate is computed as*

$$\mathbf{E}\{\phi(x_t) | \mathcal{Y}_T\} = \frac{\int_{\mathbb{R}^m} \phi(x) \bar{q}_t(x) \bar{v}_t(x) dx}{\int_{\mathbb{R}^m} \bar{q}_t(x) \bar{v}_t(x) dx} = \frac{\langle \phi \bar{q}_t, \, \bar{v}_t \rangle}{\langle \bar{q}_t, \, \bar{v}_t \rangle} \tag{16}$$

3.2 Robust Fixed-Interval Robust Smoothers for Functionals of the State

We consider robust fixed interval smoothing of H_t defined in (3). As mentioned in Sec.2.2, such computations arise in the EM algorithm for MLE.

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

$$\bar{\mathbf{E}}\{\Lambda_t H_t \phi(x_t) | \mathcal{Y}_t\} = \langle \lambda_t, \phi \rangle \tag{17}$$

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 Theorem 3.4 that $\mathbf{E}\{H_t | \mathcal{Y}_T\}$ is computed as

$$\mathbf{E}\{H_t | \mathcal{Y}_T\} = \frac{\langle \lambda_t, v_t \rangle}{\langle q_t, v_t \rangle} = \frac{\langle \bar{\lambda}_t, \bar{v}_t \rangle}{\langle \bar{q}_t, \bar{v}_t \rangle} = \frac{z_t}{\langle \bar{q}_t, \bar{v}_t \rangle} \tag{18}$$

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

Theorem 3.5

$$\lambda_t(x) = \lambda_0(x) + \int_0^t L^*(\lambda_s(x))ds + \int_0^t \Big[\alpha(x,y_s)q_s(x) + \beta'(x,y_s)f(x,s)q_s(x) - \operatorname{div}[Q\beta(x,y_s)q_s(x)] + \gamma'(x)h(x,s)q_s(x)\Big]ds + \int_0^t [\gamma'(x)q_s(x) + h'(x,s)\lambda_s(x)]\,dy_s \quad (19)$$

$$\bar\lambda_t(x) = H_0\bar q_0(x) + \int_0^t \bar\epsilon_s L^*\left(\epsilon_s\bar\lambda_s\right)ds + \int_0^t \Big[\alpha(x,y_s)\bar q_s(x) + \beta'(x,y_s)f(x,s)\bar q_s(x) - \bar\epsilon_s \operatorname{div}[Q\beta(x,y_s)\epsilon_s\bar q_s(x)]\Big]ds + \int_0^t \gamma'(x)\bar q_s(x)dy_s \quad (20)$$

$$z_t = \int_0^t \langle \alpha\bar q_s + \beta' f\bar q_s,\, \bar v_s\rangle ds - \int_0^t \langle \bar\epsilon_s \operatorname{div}[Q\beta\epsilon_s\bar q_s],\, \bar v_s\rangle ds + \int_0^t \langle \gamma\bar q_s, \bar v_s\rangle dy_s \quad (21)$$

Furthermore, the robust smoothed state estimate $\mathbf{E}\{H_t|\mathcal{Y}_T\} = z_t/\langle \bar q_t, \bar v_t\rangle$ *defines a locally Lipschitz version of* $\mathbf{E}\{H_t|\mathcal{Y}_T\}$ *in that for* $y^{(1)}, y^{(2)} \in \mathbb{R}^n$ *and constant* K *depending on* $\|y^{(1)}\|$ *and* $\|y^{(2)}\|$

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

4 Example: Robust Hidden Markov Model Smoothers

Let X_t, $t \ge 0$ be a continuous-time Markov chain defined on $(\Omega, \mathcal{F}, P)$ with finite state space $\{e_1, e_2, \dots, e_M\}$ 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 $1 \le j \le M$.

It is straightforward to show [7] that the semi-martingale representation of X_t is

$$dX_t = A^* X_t dt + W_t$$

where * denotes tranpose, W_t is a $\mathcal{F}_t$ zero mean M-vector martingale under P. Let $C = (c_1,\ c_2, \dots, c_M)' \in \mathbb{R}^M$ denote a known vector. Assume that X_t is observed via the noisy measurement process y_t as

$$dy_t = C'X_t dt + dv_t$$

4.1 Robust HMM Smoother

From (5) 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 $B \triangleq \operatorname{diag}[C]$. It is not difficult to show [7, p-p185] that the un-normalized filtered density vector $q_t = \mathbf{E}\{\Lambda_t x_t|\mathcal{Y}_t\}$ is given by the the Zakai equation

$$dq_t = A^* q_t dt + Bq_t dy_t \quad (22)$$

The above equation is the well known Wonham filter or Hidden Markov Model filter.

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

$$\mathbf{E}\{x_t|\mathcal{Y}_T\} = \frac{\sum_{i=1}^M q_t(i)v_t(i)e_i}{\langle q_t, v_t\rangle} \quad (23)$$

In analogy to (10), define the $M \times M$ diagonal exponential matrices ϵ_t as:

$$\epsilon_t(i) = \exp(c_i y_t - \frac{1}{2}c_i^2 t),$$

$$\epsilon_t = \operatorname{diag}(e_t(1), \cdots, e_t(M)) = \exp(By_t - \frac{1}{2}B^2 t)$$

Define the robust forward and backward filtered state estimates, respectively, as $\bar q_t = \bar\epsilon_t q_t$, $\bar v_t = \epsilon_t v_t$. Similar to Theorems 3.2 and 3.3, the following holds (proof omitted to save space).

Theorem 4.1 *(Robust HMM Smoother) The robust forward and backward filters evolve as*

$$\frac{d\bar q_t}{dt} = \bar\epsilon_t A^* \epsilon_t \bar q_t \quad (24)$$

$$\frac{d\bar v_t}{dt} = -\epsilon_t A \bar\epsilon_t \bar v_t, \qquad v_T = \epsilon_T \mathbf{1} \quad (25)$$

The fixed-interval smoothed estimate is computed as

$$\mathbf{E}\{x_t|\mathcal{Y}_T\} = \frac{\sum_{i=1}^M \bar q_t(i)\bar v_t(i)e_i}{\langle \bar q_t, \bar v_t\rangle} \quad (26)$$

4.2 Maximum Likelihood Parameter Estimation for HMM

Let $\theta = (a_{ij}, c_i, i \in \{1, \dots, M\}, j \in \{1, \dots, M\})$ denote the parameter vector of the HMM. By using the EM algorithm outlined in Sec.2.2 to compute the ML parameter estimates, the following re-estimation equations are obtained [10]:

$$a_{ij} = \frac{\mathbf{E}_{\tilde\theta}\left\{N_T^{ij}|\mathcal{Y}_T\right\}}{\mathbf{E}_{\tilde\theta}\left\{J_T^i|\mathcal{Y}_T\right\}}, \qquad i \ne j, \qquad c_i = \frac{\mathbf{E}_{\tilde\theta}\left\{G_T^i|\mathcal{Y}_T\right\}}{\mathbf{E}_{\tilde\theta}\left\{J_T^i|\mathcal{Y}_T\right\}} \quad (27)$$

$$N_T^{ij} = \int_0^T \langle x_{s_-}, e_i\rangle\langle dx_s, e_j\rangle, \qquad J_T^i = \int_0^T \langle x_s, e_i\rangle ds,$$

$$G_T^i = \int_0^T \langle x_s, e_i\rangle dy_s \quad (28)$$

Here N_T^i 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. Below robust smoothers are developed for evaluating these quantities which does not require two-sided Skorohod integrals.

Theorem 4.2 *Robust smoothed estimates of J_t^i, N_t^{ij} and G_t^i are given as*

$$\mathbf{E}\left\{J_t^i|\mathcal{Y}_T\right\} = \int_0^t \bar{q}_s(i)\bar{v}_s(i)ds \tag{29}$$

$$\mathbf{E}\left\{N_t^{ij}|\mathcal{Y}_T\right\} = \int_0^t \frac{\epsilon_t(i)}{\epsilon_t(j)} a_{ij}\bar{v}_s(j)\bar{q}_s(i)ds \tag{30}$$

$$\begin{aligned}\mathbf{E}\left\{G_t^i|\mathcal{Y}_T\right\} &= \bar{q}_t(i)\bar{v}_t(i)y_t \\ &- \int_0^t \left[\frac{\bar{v}_s(i)}{\epsilon_s(i)}\bar{q}_s'\epsilon_s A - \bar{q}_s(i)\epsilon_s(i)\bar{v}_s'\bar{\epsilon}_s A'\right] e_i y_s ds\end{aligned} \tag{31}$$

Remark: The EM equations (27) for HMM parameter estimation read

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

$$c_i = \frac{\bar{q}_T(i)\bar{v}_T(i)y_T - \int_0^T \left[\frac{\bar{v}_t(i)}{\epsilon_t(i)}\bar{q}_t'\epsilon_t\tilde{A} - \bar{q}_t(i)\epsilon_t(i)\bar{v}_t'\bar{\epsilon}_t\tilde{A}'\right] e_i y_t dt}{\int_0^T \bar{q}_t(i)\bar{v}_t(i)dt}$$

These expressions are apparently obtained here for the first time. In comparison, the EM equation for c_i derived in [6] 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 [10, pg.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.

References

[1] A. Bensoussan. *Stochastic Control of Partially Observable Systems.* Cambridge University Press, Cambridge, UK, 1992.

[2] F. Campillo and F. LeGland. Mle for partially observed diffusions : direct minimization vs. the em algorithm. *Stochastic Processes and their Applications*, 33:245–274, 1989.

[3] J.M.C. Clarke. The design of robust approximations to the stochastic differential equations of nonlinear filtering. In J.K. Skwirzynski, editor, *Communication Systems and Random Processes Theory, Darlington 1977.* Sijthoff and Noordhoff, Alphen aan den Rijn, 1978.

[4] F.E. Daum. Exact finite dimensional nonlinear filters. *IEEE Trans. Auto. Control*, AC-31(7):616–622, 1986.

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

[6] A. Dembo and O. Zeitouni. Parameter estimation of partially observed continuous time stochastic processes via the EM algorithm (Corrigendum). *Stochastic Processes and their Applications*, 31(1):167–169, 1989.

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

[8] R.J. Elliott and V. Krishnamurthy. Exact finite-dimensional filters for maximum likelihood parameter estimation of continuous-time linear Gaussian systems. *SIAM Journal on Control and Optimization*, 35(6):1908–1923, November 1997.

[9] 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. Auto. Control*, 44(10):1929–1933, Oct 1999.

[10] M.R. James, V. Krishnamurthy, and F. LeGland. Time discretization of continuous-time filters and smoothers for hmm parameter estimation. *IEEE Transactions on Information Theory*, 32(2):593–605, March 1996.

[11] Harold J. Kushner. A robust discrete state approximation to the optimal nonlinear filter for a diffusion. *Stochastics*, 3(2):75–83, 1979.

[12] E. Pardoux and P. Protter. A two sided stochastic integral and its calculus. *Probability theory and related fields*, 76:15–49, 1987.