Copyright © 2002 IFAC
15th Triennial World Congress, Barcelona, Spain

ROBUST CONTINUOUS-TIME SMOOTHERS FOR BENEŠ AND PIECEWISE LINEAR STOCHASTIC SYSTEMS

Vikram Krishnamurthy [*,1] **Robert Elliott** [**]

[*] *Department of Electrical and Electronic Engineering, University of Melbourne, Victoria 3010, Australia*
[**] *Faculty of Management, University of Calgary, Alberta, Canada T2N 1N4*

Abstract:
We present a robust formulation of the smoothing equations for continuous-time partially observed Beneš and piecewise linear systems. 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. The robust smoothed state estimates are locally Lipschitz in the observations which is useful for numerical simulation.

Keywords: Nonlinear Filtering Theory, Stochastic Estimation, Maximum likelihood estimators.

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 (Bensoussan, 1992). Derivation of the fixed-interval smoothed state density is more technical as it requires the use of two sided stochastic calculus.

In this paper we derive *robust* filters and smoothers for Beneš and piecewise linear systems by using a gauge transformation, see for example (Bensoussan, 1992). 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.

Robust filtering – i.e., re-expressing the stochastic differential equation as non-stochastic differential equation with random coefficients has been used extensively in nonlinear filtering, see for example Chapter 4 of (Bensoussan, 1992). More recently, in (James *et al.*, 1996) 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

[1] This work was partially funded by an ARC large grant

Expectation Maximization (EM) algorithm.
3. As examples, robust smoothers are presented for: (i) Beneš type nonlinear dynamical systems: (The linear Gaussian case is a special case). (ii) Systems with piecewise linear dynamics.

2. MODEL AND PROBLEM FORMULATION

Consider the following continuous-time partially observed nonlinear stochastic 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 $\{x_t\} \in \mathbb{R}^m$, and observation process $\{y_t\} \in \mathbb{R}^n$, $t \geq 0$ are described by

$$dx_t = f_\theta(x_t, t)\, dt + \sigma_\theta(x_t, t) dw_t, \qquad x_0 \sim \pi_0^\theta(\cdot) \tag{1}$$

$$dy_t = h_\theta(x_t, t)\, dt + de_t, \qquad y_0 = 0 \tag{2}$$

Define the 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 $t \in [0, T]$. Here $T > 0$ denotes a fixed real number. w and e are independent standard Brownian motions independent of x_0.

We make the following standard assumptions (Bensoussan, 1992, pp.114) for all $\theta \in \Theta$:
A1. $f_\theta : \mathbb{R}^m \times [0, T] \to \mathbb{R}^m$, $h_\theta : \mathbb{R}^m \times [0, T] \to \mathbb{R}^n$ denote 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 $\alpha > 0$.
A3. f and σ are Lipschitz in x, i.e.,

$$|f(x_1, t) - f(x_2, t)| \leq k|x_1 - x_2|$$
$$\|\sigma(x_1, t) - \sigma(x_2, t)\| \leq k|x_1 - x_2|$$

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 for $0 \leq t \leq T$ to (1) with $x_. \in L^2(\Omega, \mathcal{F}, P_\theta, C(\mathbb{R}^m \times [0, T]))$. (where $C(\mathbb{R}^m \times [0, T])$ denotes the space of $\mathbb{R}^m$-valued continuous functions on $[0, T]$). Also $y_. \in C(\mathbb{R}^n \times [0, T])$ 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. 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 assume
A6. $h_\theta(x, s)$ has continuous and bounded first and second derivatives wrt x and bounded first derivative wrt t. (This is relaxed for piecewise linear $h_\theta(x)$ where we use Tanaka's formula).

Objectives 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

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

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. β is assumed once differentiable in x. Our aim is to compute the 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. The same problem is considered in (Campillo and LeGland, 1989) where two-sided stochastic calculus was used to compute $\mathbf{E}\{H_t | \mathcal{Y}_T\}$.

To motivate the robust smoothers 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 (Campillo and LeGland, 1989), 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 is interpreted in (Dembo and Zeitouni, 1989) as a generalized Stratonovich integral. In Sec.3, by expressing the smoothers in robust form, the smoothed estimate $\mathbf{E}\{H_t | \mathcal{Y}_T\}$ are computed using ordinary (non-stochastic) integration.

(iii) Using the robust smoothers in Step (ii), compute maximum likelihood parameter estimate (MLE) of θ given the observation history $\mathcal{Y}_T$.

Motivation: The EM Algorithm: The EM algorithm for ML parameter estimation serves as a primary motivation for deriving smoothers for functionals of the state of the form H_t defined in (3). Each iteration of the EM algorithm consists of two steps.
Step 1. (E–step) Set $\tilde{\theta} = \widehat{\theta}_j$ and compute $\mathcal{Q}(\theta, \tilde{\theta}) = \mathbf{E}_{\tilde{\theta}}\{\log \frac{dP_\theta}{dP_{\tilde{\theta}}} \mid \mathcal{Y}_T\}$.
Step 2. (M–step) Find $\widehat{\theta}_{j+1} \in \underset{\theta \in \Theta}{\operatorname{argmax}}\, \mathcal{Q}(\theta, \theta_j)$.

The sequence generated $\{\widehat{\theta}_j, j \geq 0\}$ gives non–decreasing values of log likelihood $\mathcal{L}(\widehat{\theta}_j)$. It is shown in (Campillo and LeGland, 1989) that $\mathcal{Q}(\theta, \tilde{\theta}) = \mathbf{E}_{\tilde{\theta}}\left\{\log \Lambda^{\theta\tilde{\theta}} \mid \mathcal{Y}_T\right\}$ where $\log \Lambda^{\theta\tilde{\theta}}$ contains terms of the form
$\int_0^T \left[h_\theta(x_s, s) - h_{\tilde{\theta}}(x_s, s)\right] [dy_s - h_\theta(x_s, s) ds]$, etc. Thus 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 (3).

Preliminaries: To simplify notation, reference to the parameter θ will be dropped until Sec.4.2. We start with $(\Omega, \mathcal{F}, \bar{P})$ 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 $\langle y \rangle_t = I$.

Consider the exponentials $\Lambda_t = \Lambda_{0,t}$ where

$$\Lambda_{t_1,t_2} = \exp\Bigg(\int_{t_1}^{t_2} (h'(x_s,s)\,dy_s - \frac{1}{2}\int_{t_1}^{t_2} h'(x_s,s)\,h(x_s,s)\,ds\Bigg), \quad t_1,t_2 \in [0,T] \tag{4}$$

If we define a measure P in terms of $\bar{P}$ by setting $\frac{dP}{d\bar{P}}|_{\mathcal{G}_t} = \Lambda_t$ then Girsanov's theorem (Bensoussan, 1992) implies that under P, $dy_t = h(x_t,t)\,dt + de_t$. and $\{x_t\}$ satisfies (1). However, $\bar{P}$ is a more convenient measure with which to work.

Let $\phi \in C^2(\mathbb{R}^m)$ be an arbitrary "test" function with compact support. Define the inner product

$$\langle \gamma_t, \delta_t \rangle \triangleq \int_{\mathbb{R}^m} \gamma_t(x)\delta_t(x)dx \tag{5}$$

The following result is standard.

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

$$\mathbf{E}\{\phi(x_t)|\mathcal{Y}_t\} = \frac{\bar{\mathbf{E}}\{\Lambda_t\,\phi(x_t)|\mathcal{Y}_t\}}{\bar{\mathbf{E}}\{\Lambda_t|\mathcal{Y}_t\}} = \frac{\langle \phi,\ q_t\rangle}{\langle 1,\ q_t\rangle} \tag{6}$$

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

$$v_t(x) = \bar{\mathbf{E}}\left\{\Lambda_{t,T}|\mathcal{Y}_{t,T} \bigvee \{x_t = x\}\right\},, \quad v_T(x) = 1. \tag{7}$$

$q_t(x)$ is the *forward* un-normalized filtered density and $v_t(x)$ is the *backward* filtered density.

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

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

3. ROBUST FIXED INTERVAL SMOOTHING

Notation: $Q = \sigma(x_t)\sigma'(x_t)$, $\bar{\epsilon}_t(x) = 1/\epsilon_t(x)$

$$\epsilon_t(x) = \exp\left[\int_0^t h'(x,s)dy_s - \int_0^t \frac{1}{2}h'(x,s)h(x,s)ds\right], \tag{8}$$

$$L(\phi) = \frac{1}{2}\mathrm{Tr}[Q\nabla^2\phi] + f'\nabla\phi \tag{9}$$

$$L^*(\phi) = \frac{1}{2}\mathrm{Tr}\left[\nabla^2(Q\phi)\right] - \mathrm{div}[f\phi]$$

3.1 *Robust Fixed Interval State Smoothers*

It is well known that $q_t(x)$ evolves according to

$$q_t(x) = q_0(x) + \int_0^t L^*\left(q_s(x)\right)ds + \int_0^t h'(x,s)q_s(x)dy_s$$

initialized by $q_0(\cdot) = \pi_0(\cdot)$. Define the robust forward filtered density

$$\bar{q}_t(x) \triangleq \bar{\epsilon}_t\, q_t(x), \qquad \bar{q}_0(x) = q_0(x) \tag{10}$$

Theorem 3.1. (Robust Forward filter). $\bar{q}_t$ satisfies the following non-stochastic parabolic partial differential equation

$$\frac{\partial \bar{q}_t(x)}{\partial t} = \bar{\epsilon}_t L^*(\epsilon_t \bar{q}_t) \tag{11}$$

Furthermore, $\hat{x}_{t|t} \triangleq \langle \bar{q}_t,\ x\rangle / \langle \bar{q}_t,\ 1\rangle$ defines a locally Lipschitz version of $\mathbf{E}\{x_t|\mathcal{Y}_t\}$.

Define the robust backward filtered process as

$$\bar{v}_t(x) = \epsilon_t v_t(x) \tag{12}$$

Theorem 3.2. $\bar{v}_t$ satisfies the non-stochastic backward parabolic pde

$$\frac{\partial \bar{v}_t}{\partial t} = -\epsilon_t L(\bar{\epsilon}_t \bar{v}_t), \qquad \bar{v}_T(x) = \epsilon_T \tag{13}$$

The fixed interval smoothed estimate is

$$\mathbf{E}\{\phi(x_t)|\mathcal{Y}_T\} = \frac{\int_{\mathbb{R}^m} \phi(x)\bar{q}_t(x)\bar{v}_t(x)dx}{\int_{\mathbb{R}^m} \bar{q}_t(x)\bar{v}_t(x)dx} = \frac{\langle \phi\bar{q}_t,\ \bar{v}_t\rangle}{\langle \bar{q}_t,\ \bar{v}_t\rangle} \tag{14}$$

PROOF. Choose $\phi(x) = 1$ in Lemma 2.2. 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 (10) and (12)

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

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. But

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

which means that $\bar{v}_t$ satisfies the backward non-stochastic parabolic pde (13).

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

We consider robust fixed interval smoothing of H_t defined in (3). Define the measure valued process $\lambda_t(x)$ and its robust version $\bar\lambda_t((x)$ as

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

From Theorem 3.2 it follows that

$$\mathbf{E}\{H_t|\mathcal{Y}_T\} = \frac{\langle \lambda_t, v_t\rangle}{\langle q_t, v_t\rangle} = \frac{\langle \bar\lambda_t, \bar v_t\rangle}{\langle \bar q_t, \bar v_t\rangle} = \frac{z_t}{\langle \bar q_t, \bar v_t\rangle} \quad (15)$$

where $z_t \triangleq \langle \bar\lambda_t, \bar v_t\rangle$ denotes the un-normalized robust fixed-interval smoothed estimate.

Theorem 3.3. (Filtered and Robust Smoothed Estimate).

$$\begin{aligned}\bar\lambda_t(x) &= H_0 \bar q_0(x) + \int_0^t \bar\epsilon_s L^*\left(\epsilon_s \bar\lambda_s\right) ds \\ &+ \int_0^t \Big[\alpha(x,y_s)\bar q_s(x) + \beta'(x,y_s) f(x,s)\bar q_s(x) \\ &- \bar\epsilon_s \mathrm{div}\left[Q\beta(x,y_s)\epsilon_s \bar q_s(x)\right]\Big] ds + \gamma'(x)\bar q_t(x) y_t \\ &- \gamma'(x)\int_0^t y_s \bar\epsilon_s L^*(\epsilon_s \bar q_s) ds \qquad (16)\end{aligned}$$

$$\begin{aligned}z_t &= \langle H_0 \bar q_0, \bar v_0\rangle + \int_0^t \langle \alpha \bar q_s + \beta' f \bar q_s + \langle \gamma \bar q_t, \bar v_t\rangle y_t \\ &- \bar\epsilon_s \mathrm{div}\left[Q\beta\epsilon_s \bar q_s\right], \bar v_s\rangle ds - \int_0^t y_s \frac{d}{ds}\langle \gamma' \bar q_s, \bar v_s\rangle ds\end{aligned}$$

Furthermore, $z_t/\langle \bar q_t, \bar v_t\rangle$ is a locally Lipschitz version of $\mathbf{E}\{H_t|\mathcal{Y}_T\}$.

4. EXAMPLE 1: ROBUST Beneš SMOOTHERS

4.1 *Robust Smoother for State*

Consider (1) and (2) with $f(x_t,t) = g(x_t,t) + F_t x_t$, $\sigma(x_t,t) = \sigma_t$, $h(x_t,t) = C$. For convenience assume y_t is a scalar valued observation process (i.e., $n = 1$). and $C \in \mathbb{R}^{1\times m}$ is time-invariant.

Assumption: (Bensoussan, 1992, pg 199): Suppose $\psi(x,t)$ in $C^{2,1}(\mathbb{R}^m, \mathbb{R}^+)$ such that

$$Q_t \nabla\psi(x,t) = g(x,t), \quad x \in \mathbb{R}^m \ . \qquad (17)$$

Assume $\psi(x,t)$ satisfies the following *Beneš non-linearity* condition $\frac{\partial\psi}{\partial t} + \frac{1}{2}\mathrm{Tr}\left(Q_t \nabla^2\psi\right)$

$$+\frac{1}{2}(\nabla\psi)' Q_t \nabla\psi + x' F_t' \nabla\psi = \frac{1}{2} x' \Lambda_t x + x' \mu_t + \delta_t$$

where $\Lambda_t \in \mathbb{R}^{m\times m}$ satisfying $\Lambda_t + C' C \geq 0$, $\mu_t \in \mathbb{R}^m$ and $\delta_t \in \mathbb{R}$.

Robust backward Beneš filter: $\bar v_T(x) = \epsilon_T(x)$ and for $t \leq T$,

$$\begin{aligned}\bar v_t(x) &= \int_{\mathbb{R}^m} \exp(\psi(\zeta,T))\, \bar s_t(\zeta)\, \bar v_t(x,\zeta)\, d\zeta \\ \bar v_t(x,\zeta) &= \exp\Big(-\psi(x,t) - \frac{1}{2}x'(\bar\Sigma_t^{-1} + C'Ct)x + \bar l_t'(\zeta)x \\ &\quad - \frac{1}{2}(\bar l_t(\zeta) - C'y_t)'\bar\Sigma_t(\bar l_t(\zeta) - C'y_t)\Big) \\ \bar s_t(\zeta) &= K_t \exp(-\frac{1}{2}\zeta' \bar S_t \zeta + \zeta' \bar\rho_t) \qquad (18)\end{aligned}$$

Here the terms $\bar l_t(\zeta)$ and $\bar S_t$ are defined as

$$\bar l_t(\zeta) = \bar\Sigma_t^{-1}\Phi_t\zeta + \tilde l_t, \qquad \bar S_t = \int_t^T \bar\Phi_s'(\Lambda + C'C)\bar\Phi_s ds$$

and the $M\times M$ matrix Φ_t satisfies the equation

$$\frac{d\bar\Phi}{dt} = -(F_t + \Sigma_t(\Lambda + C'C))\bar\Phi_t, \qquad \bar\Phi_T = I$$

The statistics $\tilde l_t$, $\bar\Sigma_t$ and $\bar\rho_t$ satisfy

$$\begin{aligned}\frac{d\tilde l_t}{dt} &= \bar\Sigma_t^{-1} Q(\tilde l_t - C'y_t) - F'(\tilde l_t - C'y_t) + \mu_t, \ \ \tilde l_T = 0 \\ \frac{d\bar\Sigma_t}{dt} &= \bar\Sigma_t(\Lambda + C'C)\bar\Sigma_t - Q + F_t\bar\Sigma_t + \bar\Sigma_t F_t', \ \ \bar\Sigma_T = 0 \\ \bar\rho_t &= -\bar\Phi_t' C' y_t + \bar\Phi_T' C' y_T + \int_t^T \bar\Phi_s' F_s' C' y_s ds \\ &+ \int_t^T \bar\Phi_s' \mu_s ds + \int_t^T \bar\Phi_s'(\Lambda_s + C'C)\Sigma_s \tilde l_s ds, \ \ \bar\rho_T = 0\end{aligned}$$

Remark: For linear dynamics with initial distribution $\pi_0(\cdot)$, simply set $\psi(x,t) = 0$, $\Lambda_t = 0$, $\mu_t = 0$ and $\delta_t = 0$. Further, if $\pi_0(\cdot) \sim N(\hat x_0, \Sigma_0)$ then the Kalman filter (in robust form) follows with conditional mean state estimate $m_t \triangleq \mathbf{E}\{x_t|\mathcal{Y}_t\} = \Sigma_t(\tilde r_t + C'y_t)$, and the Kalman state covariance $\Sigma_t \triangleq \mathbf{E}\{(x_t - m_t)(x_t - m_t)'\}$ (Riccati equation).

4.2 *Maximum Likelihood Parameter Estimation*

Consider the linear Gaussian system with $\psi = 0$, Gaussian initial conditions and (F,σ,C) in controller canonical form. Let $\theta = [c_1,\ldots,c_m]'$ denote the parameter vector. The EM algorithm outlined in Sec.2 can be used to compute the MLE of θ. The M-step yields the estimates

$$C = \left(\mathbf{E}_{\tilde\theta}\left\{\int_0^T x_s x_s' ds|\mathcal{Y}_T\right\}\right)^{-1} \mathbf{E}_{\tilde\theta}\left\{\int_0^T x_s dy_s|\mathcal{Y}_T\right\}$$

Example: Consider computing $\mathbf{E}\{\int_0^t x_s dy_s|\mathcal{Y}_T\}$ which is required above. Define $H_t^i = \int_0^t e_i' x_s dy_s$. Then from (16) with $\alpha = \beta = 0$ and $\gamma(x_s) = e_i' x_s$

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

where $\frac{d}{ds} m_{s|T}$ can be computed from the robust forward and backward Benes filters.

Remark: The above robust smoothed estimate is identical to the generalized Stratonovich integral used in (Dembo and Zeitouni, 1989).

5. EXAMPLE 2: PIECEWISE LINEAR SYSTEMS

Here we consider piecewise linear dynamics and observation equation. In general the filtered density for such models does not exist. So the Zakai equations will be considered in weak form, i.e., distributional sense. In (Pardoux and Savona, 1988) and (Savona, 1988) 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 (Pardoux and Savona, 1988) 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 (Pardoux and Savona, 1988) we show the robust formulation can be used to construct approximate smoothers.

Signal Model: Consider the scalar piecewise linear model (1), (2) where $f(x_t) = \sum_{k=1}^K I(x_t \in P_k)(a_k x_t + b_k)$, $h_\theta(x_t) = \sum_{k=1}^K I(x_t \in P_k)(c_k x_t + d_k)$, $\sigma(t)$ assumed known. Here P_k, $k = 1, 2, \dots, K$ denotes a finite partition of $\mathbb{R}$. Let $B_k \in \mathbb{R}$, $k = 1, \dots, K-1$ denote the boundary points of $P_1, \dots, P_K$. $\theta = (c_1, \dots, c_K, d_1, \dots, d_K)'$ denotes the parameter vector to be estimated. $h_\theta(x)$ is assumed continuous in x, i.e., $c_k B_k + d_k = c_{k+1} B_k + d_{k+1}$. It is well known (Pardoux and Savona, 1988) that (1) has a unique weak solution.

The EM algorithm for estimating θ requires computation of $G_k^{(i)} \triangleq \mathbf{E}_{\theta_j}\left\{\int_0^T x_s^i I(x_s \in P_k) dy_s | \mathcal{Y}_T\right\}$ and $A_k^{(i)} \triangleq \mathbf{E}_{\theta_j}\left\{\int_0^T x_s^i I(x_s \in P_k) ds | \mathcal{Y}_T\right\}$ for $k = 1, \dots, K$, $i = 0, 1, 2$. This motivates deriving smoothers for state functionals H_t with $\beta = 0$.

Define $\bar{\pi}_t(\phi H_t) = \bar{\mathbf{E}}\{\Lambda_t \phi(x_t) H_t | \mathcal{Y}_t\}$, $\bar{\pi}_{t|T}(\phi H_t) = \bar{\mathbf{E}}\{\Lambda_T \phi(x_t) H_t | \mathcal{Y}_T\}$. The weak Zakai equation is

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

Unlike the proof of Theorem 3.3, one cannot postulate the existence of the densities q_t or λ_t.

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

$$f^\delta(x,t) = \sum_{k=1}^K I\left(x_{[t/\delta]\delta} \in P_k\right)(a_k x_t + b_k)$$
$$h^\delta(x,t) = \sum_{k=1}^K I\left(x_{[t/\delta]\delta} \in P_k\right)(c_k x_t + d_k) \quad (19)$$

$f^\delta : C(\mathbb{R} \times [0,T]) \times [0,T] \to D(\mathbb{R} \times [0,T])$, and $h^\delta : C(\mathbb{R} \times [0,T]) \times [0,T] \to D(\mathbb{R}, [0,T])$.

We 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), (19) is 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 (4) define for $t_1, t_2 \in [0,T]$

$$\Lambda_{t_1,t_2}^\delta(\omega_1, \omega_2) = \exp\left(\int_{t_1}^{t_2} (h^\delta(x_s, s)\, dy_s - \frac{1}{2}\int_{t_1}^{t_2} (h^\delta(x_s, s))^2\, ds\right). \quad (20)$$

Since f^δ and h^δ have linear growth $|f^\delta(x,t) + h^\delta(x,t)| \le c(1 + |x_t|)$ for some constant $c \in \mathbb{R}$, Λ^δ is a martingale. As in Sec.2, define $\bar{P}^\delta$ by $dP^\delta/d\bar{P}^\delta|_{\mathcal{G}_t} = \Lambda_t^\delta$. Define $\epsilon_t^\delta(x)$, $\bar{\epsilon}_t^\delta(x)$ (8) for (19).

For any $\phi \in C^2(\mathbb{R}^m)$, define $\bar{\pi}_t^\delta(\phi) = \bar{\mathbf{E}}^\delta\{\Lambda_t^\delta \phi(x_t)|\mathcal{Y}_t\} = \langle \phi, q_t^\delta \rangle$, $\bar{\pi}_t^\delta(\phi H_t) = \bar{\mathbf{E}}^\delta\{\Lambda_t^\delta H_t \phi(x_t)|\mathcal{Y}_t\} = \langle \phi, \lambda_t^\delta \rangle$, $v_t^\delta(x) = \bar{\mathbf{E}}^\delta\left\{\Lambda_{t,T}^\delta | \mathcal{Y}_{t,T} \bigvee \{x_t = x\}\right\}$, $\bar{\pi}_{t|T}^\delta(\phi H_t) = \bar{\mathbf{E}}^\delta\{\Lambda_T^\delta H_t \phi(x_t)|\mathcal{Y}_T\} = \langle \phi, z_t^\delta \rangle$, $\bar{q}_t^\delta(x) = \bar{\epsilon}_t^\delta q_t^\delta(x)$, $\bar{\lambda}_t^\delta = \bar{\epsilon}_t^\delta \lambda_t^\delta$, $\bar{v}_t^\delta = \epsilon_t^\delta v_t^\delta(x)$, $z_t^\delta = \langle \bar{\lambda}_t^\delta, \bar{v}_t^\delta \rangle$

The aim of this section is to show:
(i) The robust forward filtered density $\bar{q}_t^\delta(x)$ and backward process $\bar{v}_t^\delta(x)$ of the approximate model (19) can be computed by a bank of K parallel Kalman type forward and backward filters with non-Gaussian initial conditions – see Sec.5.1.
(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) $\forall t \ge 0$. Thus the smoothed distribution for the approximate model (19) converges to smoothed distribution for the piecewise linear model. The robust formulation is used in the convergence proof – see Sec.5.2.

5.1 *Approximate Smoothing Algorithm*

For each of the piecewise linear segments $k = 1, \dots, K$ define $\bar{\epsilon}_{t,k}(x) = 1/\epsilon_{t,k}(x)$

$$\epsilon_{t,k}(x) = \exp\left[c_k(x) y_t - \frac{1}{2} c_k^2(x) t\right],$$
$$L_k(\phi) = \frac{1}{2}\mathrm{Tr}[Q\nabla^2\phi] + a_k x \nabla\phi$$

The estimate $\mathbf{E}^\delta\{\phi(x_t)|Y_T\}$ for approximate model (19) is computed by 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:

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

Step 2. Propagate: Run K robust Kalman filters for non Gaussian initial condition (see Sec.4.1) on $t \in [i\delta, (i+1)\delta)$ as

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

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 robust 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 (13); recombine at $t = i\delta$ to obtain $\bar{v}^{\delta}_{i\delta}(x)$, etc.

The smoothed state estimate is computed as $\mathbf{E}^{\delta}\{x_t|\mathcal{Y}_T\} = \langle x\bar{q}^{\delta}_t, \bar{v}^{\delta}_t\rangle / \langle \bar{q}^{\delta}_t, \bar{v}^{\delta}_t\rangle$ and $\mathbf{E}^{\delta}\{H_t|\mathcal{Y}_T\} = z^{\delta}_t / \langle \bar{q}^{\delta}_t, \bar{v}^{\delta}_t\rangle$ where z^{δ}_t is computed by (16).

5.2 *Convergence of Approximate Smoother*

Introduce a double measure change by defining

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

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

Let $\bar{\bar{P}}$ denote the standard Wiener measure. Girsanov's theorem implies that
$\langle \pi_{t|T}, H_t\phi\rangle(\omega^2) = \bar{\bar{\mathbf{E}}}\{\Lambda_T N_T H_t \phi(x_t)|\mathcal{Y}_T\}$,
$\langle \pi^{\delta}_{t|T}, H_t\phi\rangle(\omega^2) = \bar{\bar{\mathbf{E}}}\{\Lambda^{\delta}_T N^{\delta}_T H_t \phi(x_t)|\mathcal{Y}_T\}$.

Theorem 5.1. As $\delta \to 0$, the smoothed estimate $\langle \pi^{\delta}_{t|T}, H_t\phi\rangle(\omega^2) \to \langle \pi_{t|T}, H_t\phi\rangle(\omega^2)$, $\bar{\bar{P}}$ a.s. for all $t \in [0,T]$ and $\omega^2 \in \Omega^2$.

To prove Theorem 5.1 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 $\langle \pi^{\delta}_{t|T}, H_t\phi\rangle(\omega^2)$, $\omega^2 \in \Omega^2 = C(\mathbb{R} \times [0,T])$. To obtain these robust versions integrate by parts the stochastic (dY) integrals appearing in Λ_t (4) and Λ^{δ}_t (20. This requires computation of the differentials of $h(x_s)$ and $h^{\delta}(x,s)$.

Lemma 5.2. As $\delta \to 0$, $\Lambda^{\delta}_T(\cdot,\omega^2) \to \Lambda_T(\cdot,\omega^2)$, $\bar{\bar{P}}$ a.s. for all $\omega^2 \in \Omega^2$

PROOF. Because the gradient of h jumps across each boundary point B_k one needs to use Tanaka's formula for semi-martingales (instead of Ito's formula) which yields $\bar{\bar{P}}$ a.s.

$$h(x_t) = h(x_0) + \int_0^t \nabla h(x_s)dx_s + \frac{1}{2}\sum_{k=1}^{K-1}\int_0^t I(x_s = B_k)dL^{B_k}_s(x) \quad (21)$$

where $\nabla h(x) = \sum_{k=1}^K I(x_s \in P_k)c_k$ and $L^{B_k}_s(x)$ denotes the local time at B_k of the process x_t.

Thus $\Lambda_t = \exp\Big(y_t h(x_t) - \int_0^t y_s \nabla h(x_s)dx_s$
$- \frac{1}{2}\sum_{k=1}^{K}\int_0^t y_s I(x_s = B_k)dL^{B_k}_s(x) - \frac{1}{2}\int_0^t (h(x_s))^2 ds\Big)$

Consider evaluating Λ^{δ}_t: It is easily shown $\bar{\bar{P}}$ a.s.

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

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

$$\sum_{j \in J(\delta;i,t)} y_{j\delta}(\omega_2)(\Lambda_i(x_{j\delta}) - \Lambda_i(x_{(j-1)\delta}))(x_{j\delta} - x) \to \int_0^t y_s(\omega^2) I(x_s = B_k)dL^{B_k}_s(x) \quad \bar{\bar{P}} \text{ a.s. and } \forall\omega^2$$

From Lemma 5.2, $\Lambda^{\delta}_T N^{\delta}_T \to \Lambda_T N_T$, $\bar{\bar{P}}$ a.s. for all $\omega^2 \in \Omega^2$. Also $\forall\omega^2$, it can be shown that

$$\bar{\bar{\mathbf{E}}}\left\{\left(N^{\delta}_T(\cdot) Z^{\delta}_T(\cdot,\omega^2)\right)^2\right\} \le c(\cdot,\omega^2) \quad \forall\delta > 0$$

which implies that $\Lambda^{\delta}_T(\cdot,\omega^2)N^{\delta}_T$ is uniformly integrable. Hence Theorem 5.1 follows.

6. REFERENCES

Bensoussan, A. (1992). *Stochastic Control of Partially Observable Systems.* Cambridge University Press. Cambridge, UK.

Campillo, F. and F. LeGland (1989). MLE for partially observed diffusions : direct minimization vs. the EM algorithm. *Stochastic Processes and their Applications* **33**, 245–274.

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

James, M.R., V. Krishnamurthy and F. LeGland (1996). Time discretization of continuous-time filters and smoothers for HMM parameter estimation. *IEEE Transactions on Information Theory* **42**(2), 593–605.

Pardoux, E. and C. Savona (1988). Piecewise linear filtering. In: *Stochastic Differential Systems, Stochastic Control and Applications.* Vol. 10 of *IMA*. pp. 429–437. Springer. Berlin—New York.

Savona, C. (1988). Approximate nonlinear filtering for piecewise linear systems. *Systems and Control Letters* **11**, 327–332.