Exact Smoothers for Discrete-Time Hybrid Stochastic Systems

R. J. Elliott, W. P. Malcolm and F. Dufour

Abstract—In this article we compute the exact smoothing algorithm for discrete-time Gauss-Markov models whose parameter-sets switch according to a known Markov law. The smoothing algorithm we present is general, but can be readily configured into any of the three main classes of smoothers of interest to the practitioner, that is, fixed point, fixed lag and fixed interval smoothers.

All smoothers are functions of their corresponding filter. The filter we use to develop our smoother is the exact information-state filter for hybrid Gauss-Markov models due to Elliott, Dufour and Sworder, [14]. Our approach is in contrast to some other smoothing schemes in the literature, which are often based upon ad-hoc schemes.

It is well known that the fundamental impediment in all estimation for jump Markov systems, is the management of an exponentially growing number of hypotheses. In our scheme, we propose a method to maintain a fixed number of candidate paths in a history, each identified as optimal by a probabilistic criterion. The outcome of this approach is a new and general smoothing scheme, based upon the exact filter dynamics, and whose memory requirements remain fixed in time.

I. INTRODUCTION

In this article the reference probability method is used to compute smoothed state estimates for a discrete-time hybrid dynamical system. This particular problem has received relatively little attention in the literature, that is, compared to the corresponding filtering problem for hybrid stochastic systems. This is perhaps due to the inherent complexity arising from stochastic hybrid systems and the basic complexity one generally expects in smoothing algorithms. In the articles [5] and [6], a smoothing scheme based upon the Interacting Multiple Model (IMM) algorithm is proposed. The originator of the IMM, namely Henk Blom, also studied the smoothing problem by using a time-reversal method in [4]. Further, a two-filter form of a smoother, similar in spirit to the ideas of the so-called Fraser-Potter smoother, was presented in [17].

Traditionally smoothing has been considered largely an offline processing scheme and therefore has received relatively scant attention from communities focussed on real-time estimation, such as the tracking community. However, with the advent of increasingly powerful computing and the potential benefits of smoothing schemes, it is indeed timely to revisit this particular smoothing problem. Some potential applications of smoothing in tracking are, fixed lag smoothers, fixed interval smoothing for track reconstruction and parameter estimation using the EM algorithm.

Using a general result, we propose a new suboptimal smoothing algorithm which provides an exact hypothesis management scheme, circumventing growth in algorithmic complexity.

II. STOCHASTIC DYNAMICS

Much of the detail in this section is now standard and can be found in texts such as [2], [9], [10]. Further, a companion paper to this article which considers the associate filtering problem, appears in this publication, see [20].

All processes are defined, initially, on a fixed probability space \((\Omega, \mathcal{F}, P)\).

A. Markov Chain Dynamics

We consider a time-homogeneous discrete-time \(m\)-state Markov chain \(Z\). We represent the state space of \(Z\) on a canonical basis of indicator function \(e_i\), where the vector \(e_i = (0, 0, \ldots, 1, 0, \ldots)\)' has unity in the \(i\)-th position. Our Markov chain has statistics \((\Pi, p_0)\), where \(\Pi = [\pi(j,i)]_{1\leq j\leq m}^{1\leq i\leq m}\) is the transition matrix of \(Z\), with elements

\[
\pi(j,i) = P(Z_k = e_j | Z_{k-1} = e_i), \quad \forall k \in \mathbb{N}
\]

and \(E[Z_0] = p_0\).

B. State Process Dynamics

We suppose the indirectly observed state vector \(x \in \mathbb{R}^{n \times 1}\), has dynamics

\[
x_k = \sum_{j=1}^{m} \langle Z_k, e_j \rangle A_j x_{k-1} + \sum_{j=1}^{m} \langle Z_k, e_j \rangle B_j w_k.
\]

Here \(w\) is a vector-valued Gaussian process with \(w \sim N(0, I_n)\). \(A_j\) and \(B_j\) are \(n \times n\) matrices and for each \(j \in \{1, 2, \ldots, m\}\), are nonsingular.

C. Observation Process Dynamics

Consider a vector-valued observation process with values in \(\mathbb{R}^{d \times 1}\) and dynamics

\[
y_k = \sum_{j=1}^{m} \langle Z_k, e_j \rangle C_j x_k + \sum_{j=1}^{m} \langle Z_k, e_j \rangle D_j v_k.
\]

Here \(v\) is a vector-valued Gaussian process with \(v \sim N(0, I_d)\). We suppose the matrices \(D_j \in \mathbb{R}^{d \times d}\), for each \(j \in \{1, 2, \ldots, m\}\), are nonsingular. Our filtrations are as follows:

\[
\mathcal{F}_k = \sigma\{x_\ell, 0 \leq \ell \leq k\},
\]

\[
\mathcal{Z}_k = \sigma\{Z_\ell, 0 \leq \ell \leq k\},
\]

\[
\mathcal{Y}_k = \sigma\{y_\ell, 0 \leq \ell \leq k\},
\]

\[
\mathcal{G}_k = \sigma\{Z_\ell, x_\ell, y_\ell, 0 \leq \ell \leq k\}.
\]
D. Reference Probability

The dynamics given at (2) and (3), are each defined on a measurable space \((\Omega, \mathcal{F})\), under a measure \(P\). However, consider a new measure \(P^\dagger\), under which the dynamics for the processes \(Z, x\) and \(y\), are, respectively

\[
P^\dagger \left\{ \begin{array}{l}
Z_k = \Pi Z_{k-1} + L_k, \\
x_k \text{ are iid and } N(0, I_n), \\
y_k \text{ are iid and } N(0, I_d).
\end{array} \right.
\]

\[\tag{8} \]

**Notation:**
The symbol \(\Phi(\cdot)\) will be used to denote the zero mean normal density on \(\mathbb{R}^d\):

\[
\Phi(\xi) = (2\pi)^{-d/2} \exp\left(-\frac{1}{2} \xi^\top \xi\right).
\]

Similarly we shall also use the symbol \(\Psi(\cdot)\) to denote a standardised Gaussian density. The space dimension on which these densities is defined will be clear by context. To avoid cumbersome notation with matrices, we sometimes denote the inverse of a matrix \(A\) by \(\text{inv}(A)\) and the space of \(m \times n\) matrices in our context is written as \(\mathbb{M}^{m \times n}\).

We now define the measure \(P\), by setting the restriction of a Radon-Nikodym derivative to \(\mathcal{G}_k\) to

\[
\Lambda_{0,k} \overset{\Delta}{=} \frac{dP}{dP^\dagger} |_{\mathcal{G}_k} = \prod_{\ell=0}^k \lambda_{\ell},
\]

where

\[
\lambda_{\ell} = \sum_{j=1}^m \langle Z_\ell, e_j \rangle \frac{\Phi(D_{j,-1}(y_\ell - C_j x_\ell))}{|D_j| \Phi(y_\ell)} \times \frac{\Psi(B_j^{-1}(x_\ell - A_j x_{\ell-1}))}{|B_j| \Psi(x_\ell)}.
\]

\[\tag{10} \]

**Theorem 1** Suppose \(\gamma = \{\gamma_\ell, 0 \leq \ell \leq k\}\) is an integrable \(\mathcal{G}\)-adapted process, then

\[E[\gamma_k | \mathcal{Y}_k] = \frac{E[\Lambda_{0,k}^\dagger|\mathcal{Y}_k]}{E[\Lambda_{0,k} | \mathcal{Y}_k]}\]

\[\tag{11} \]

III. Review Of Exact Hybrid Filter Dynamics

The exact state estimation filter given in [14] is written in unnormalised form, that is, dynamics satisfied by an unnormalised probability density. These dynamics are computed using reference probability techniques, see [9], [10] and [2]. Briefly, we are interesting in computing conditional probabilities for joint events of the form \(P(x \in dx, Z_k = e_j | \mathcal{Y}_k)\). Omitting the details, we assume the existence unnormalised probability densities corresponding to our events of interest, where, for example,

\[P(x \in dx, Z_k = e_j | \mathcal{Y}_k) = \frac{q^j_k(x)\, dx}{\int_{\mathbb{R}^d} q^j_k(x)\, dx},\]

\[\tag{12} \]

What we would like to compute, is recursive dynamics whose solutions are the densities \(q^j(x) : \mathbb{R}^{n \times 1} \rightarrow \mathbb{R}_+\), where \(j \in \{1, 2, \ldots, m\}\). To this end, we recall the fundamental contribution of [14] in the next Theorem.

**Theorem 2** (Elliott, Dufour, Sworder, 1996) The unnormalised probability density \(q^j_k(x)\), is computed by the equation,

\[
q^j_k(x) = \frac{\Phi(D_j^{-1}(y_k - C_j x))}{\Phi(y_k)} D_j \times \sum_{r=1}^m \pi(j, r) \int_{\mathbb{R}^d} \Psi(B_j^{-1}(x - A_j \xi)) q_{j-1}^r(\xi)\, d\xi.
\]

\[\tag{13} \]

IV. Exact Hybrid Smoother Dynamics

In this section we present our main results. We compute a general smoother by applying similar techniques to those developed in [19] and [12].

We first note that

\[E[\langle Z_k, e_j \rangle \, f(x_k) | \mathcal{Y}_{0,T}] = \frac{E[\Lambda_{0,T}^\dagger \langle Z_k, e_j \rangle \, f(x_k) | \mathcal{Y}_{0,T}]}{E[\Lambda_{0,T} | \mathcal{Y}_{0,T}]}.
\]

\[\tag{14} \]

Write

\[
\bar{F}_k \overset{\Delta}{=} \sigma\{x_\ell, Z_\ell, 0 \leq \ell \leq k\}.
\]

\[\tag{15} \]

Using repeated conditioning, the numerator in the quotient of equation (14) may be written as

\[E[\Lambda_{0,T}^\dagger \langle Z_k, e_j \rangle \, f(x_k) | \mathcal{Y}_{0,T}] = E[\Lambda_{0,k} \Lambda_{k+1,T}^\dagger \langle Z_k, e_j \rangle \, f(x_k) | \mathcal{Y}_{0,T}]\]

\[= E[\Lambda_{0,k} \Lambda_{k+1,T}^\dagger \langle Z_k, e_j \rangle \, f(x_k) | \bar{F}_k \vee \mathcal{Y}_{0,T}]| \mathcal{Y}_{0,T}\]

\[= E[\Lambda_{0,k} \langle Z_k, e_j \rangle \, f(x_k)E[\Lambda_{k+1,T}^\dagger | \bar{F}_k \vee \mathcal{Y}_{0,T}] | \mathcal{Y}_{0,T}].\]

\[\tag{16} \]

Further, the Markov property of our processes ensures the following equality

\[E[\Lambda_{k+1,T}^\dagger | \bar{F}_k \vee \mathcal{Y}_{0,T}] = E[\Lambda_{k+1,T} | Z_k, x_k, \mathcal{Y}_{0,T}].\]

\[\tag{17} \]

Write

\[v_k^j(x) \overset{\Delta}{=} E[\Lambda_{k+1,T} | Z_k = e_j, x_k = x, \mathcal{Y}_{0,T}].\]

\[\tag{18} \]

**Theorem 3** (Pardoux [24]) The exact and general, unnormalised smoother density for the stochastic hybrid system whose dynamics include equations (2), (3) and the Markov chain \(Z\), is given by the expectation

\[E[\Lambda_{0,k} \langle Z_k, e_j \rangle \, f(x_k) | \mathcal{Y}_{0,T}] = \int_{\mathbb{R}^d} \left\{ q^j_k(\xi) v_k^j(\xi) \right\} f(\xi)\, d\xi.
\]

\[\tag{19} \]

We now compute a backwards recursion for the function \(v_k^j(x)\), similar in form to the forward recursion for the function \(q_k^j(x)\), given in equation (13).
Theorem 4 The un-normalised function \( v_k^j(x) \), satisfies the backward recursion:

\[
v_k^j(x) = \sum_{r=1}^{m} \frac{\pi(r,j)}{|D_r||B_r|\Phi(y_{k+1})} \times \int_{\mathbb{R}^n} \Phi(D_r^{-1}(y_{k+1} - C_r\xi)) \Psi(B_r^{-1}(\xi - A_r x)) v_{k+1}^j(\xi) d\xi.
\]

\( (20) \)

Proof: From definition (18), we again use repeated conditioning to write

\[
v_k^j(x) = E[\Lambda_{k+1,T} | Z_k = e_j, x_k = x, \mathcal{Y}_{0,T}] = E[\Lambda_{k+1,T} | Z_k = e_j, x_k = x, \mathcal{Y}_{0,T}]
\]

\[
= E\left[ \left. \lambda_{k+1} E^\dagger \left[ \Lambda_{k+2,T} | Z_{k+1}, x_{k+1}, Z_k = e_j, x_k = x, \mathcal{Y}_{0,T} \right] \right| Z_k = e_j, x_k = x, \mathcal{Y}_{0,T} \right]
\]

\[
= \sum_{r=1}^{m} E^\dagger \left[ \lambda_{k+1} (Z_{k+1}, e_r) E^\dagger \left[ \Lambda_{k+2,T} | Z_{k+1} = e_r, x_{k+1}, Z_k = e_j, x_k = x, \mathcal{Y}_{0,T} \right] \right]
\]

\[
= \sum_{r=1}^{m} E^\dagger \left[ \lambda_{k+1} \left( Z_{k+1}, e_r \right) x, x_k = x, \mathcal{Y}_{0,T} \right] \right]
\]

\[
\times E^\dagger \left[ \Lambda_{k+2,T} | Z_{k+1} = e_r, x_{k+1}, Z_k = e_j, x_k = x, \mathcal{Y}_{0,T} \right]
\]

\[
= \sum_{r=1}^{m} E^\dagger \left[ \lambda_{k+1} \left( Z_{k+1}, e_r \right) x, x_k = x, \mathcal{Y}_{0,T} \right] \right]
\]

\[
\times E^\dagger \left[ \Lambda_{k+2,T} | Z_{k+1} = e_r, x_{k+1}, Z_k = e_j, x_k = x, \mathcal{Y}_{0,T} \right]
\]

\[
(21) \]

Since \( Z \) is a Markov chain and under \( P^\dagger \), we can continue this calculation now by noting that

\[
v_k^j(x) = \sum_{r=1}^{m} \frac{\pi(r,j)}{|D_r||B_r|\Phi(y_{k+1})} \times \int_{\mathbb{R}^n} \Phi(D_r^{-1}(y_{k+1} - C_r x_{k+1})) \Psi(B_r^{-1}(\xi - A_r x)) v_{k+1}^j(\xi) d\xi
\]

\( (22) \)

V. GAUSSIAN MIXTURE APPROXIMATIONS

The dynamics for the \( v \) process given in Theorem 4 are exact, however, to implement such dynamics, the immediate question is how one might deal with the integral over \( \mathbb{R}^n \) in equation (20). The process \( v \) is a nonnegative process. We suppose that the function \( v_{k+1}^j(\xi) \), in equation (20), can be approximated arbitrarily closely by a weighted Gaussian mixture. This assumption is supported by a fundamental approximation result in function space, which is detailed in [30] and [18].

Theorem 5 Suppose the function \( v_{k+1}^j(\xi) \), (in equation (20)), is approximated as a weighted Gaussian mixture with \( M^\dagger \in \mathbb{N} \) components. That is, suppose

\[
v_{k+1}^j(\xi) = \sum_{s=1}^{M^\dagger} \rho_{k+1,s} \frac{1}{(2\pi)^{n/2} |\Sigma_{k+1,T}|^\frac{1}{2}} \times \exp \left\{ -\frac{1}{2} (\xi - \alpha_{k+1,T}^r)^\dagger \Sigma_{k+1,T}^{-1} (\xi - \alpha_{k+1,T}^r) \right\}.
\]

\( (23) \)

Here \( \Sigma_{k+1,T} \in \mathbb{R}^{n \times n} \), and \( \alpha_{k+1,T} \in \mathbb{R}^n \), are both \( \mathcal{Y}_{k+1,T} \)-measurable functions for all pairs \( (r, s) \in \{1, 2, \ldots, m\} \times \{1, 2, \ldots, M^\dagger\} \). Using the Gaussian mixture (23), the recursion for \( v_k^j(x) \), at times \( k \in \{1, 2, \ldots, T - 1\} \), has the form

\[
v_k^j(x) = \frac{1}{(2\pi)^{n/2} |\Sigma_{k+1,T}|^\frac{1}{2}} \sum_{r=1}^{M^\dagger} \sum_{s=1}^{M^\dagger} K_{k+1,T}^r(j,r,s) \times \exp \left\{ -\frac{1}{2} (x - \text{inv}(S_{k+1,T}^r) \tau_{k+1,T}^r)^\dagger S_{k+1,T}^r (x - \text{inv}(S_{k+1,T}^r) \tau_{k+1,T}^r) \right\}.
\]

\( (24) \)

At the final time \( T \), \( \forall \ j \in \{1, 2, \ldots, m\} \),

\[
v_T^j(x) = \sum_{r=1}^{M^\dagger} \frac{\pi(r,j)}{|D_r||B_r|\Phi(y_{k+1})} \times \int_{\mathbb{R}^n} \Phi(D_r^{-1}(y_{k+1} - C_r x_{k+1})) \Psi(B_r^{-1}(\xi - A_r x)) v_{k+1}^j(\xi) d\xi
\]

\( (25) \)

Here

\[
K_{k+1,T}^r(j,r,s) \triangleq \frac{\pi(r,j)}{|D_r||B_r|\Phi(y_{k+1})} \times \left[ \sum_{r=1}^{M^\dagger} \right] |D_r||B_r|\Phi(y_{k+1}) \times
\]

\[
\left[ \frac{\pi(r,j)}{|D_r||B_r|\Phi(y_{k+1})} \times \int_{\mathbb{R}^n} \Phi(D_r^{-1}(y_{k+1} - C_r x_{k+1})) \Psi(B_r^{-1}(\xi - A_r x)) v_{k+1}^j(\xi) d\xi \right]
\]

\( (22) \)
Corollary 1 Write

\[ v^j_k \triangleq \int_{\mathbb{R}^n} v^j_k(\xi) d\xi. \] (32)

The scalar-valued quantity \( v^j_k \), is computed by the double sum

\[ v^j_k = \frac{1}{(2\pi)^{d/2}\Phi(y_{k+1})} \sum_{r=1}^{m} \sum_{s=1}^{M^v} \phi^j_{k+1,T}(j, r, s). \] (33)

Here

\[ \phi^j_{k+1,T}(j, r, s) \triangleq K^j_{k+1,T}(j, r, s)|S^r_k|^{-1/2}. \] (34)

A. Sub-Optimal Smoother Dynamics

1) Hypothesis Management: Write

\[ \Gamma^v \triangleq \{1, 2, \ldots, m\} \times \{1, 2, \ldots, M^v\}, \] (35)

\[ \tilde{S}^v_{k+1,T}(j, r, s) \triangleq \{\phi^v_{k+1,T}(j, r, s)\}_{(r, s) \in \Gamma^v}. \] (36)

We propose, at each time \( k \), to identify the \( M^v \)-best candidate functions, (components in the Gaussian mixture), for each function \( v^j_k(x) \). This maximisation procedure is as follows:

\[ \phi^v_{k+1,T}(j, r_1^*, s_1^*) \triangleq \max_{(r, s) \in \Gamma^v} \tilde{S}^v_{k+1,T}(j, r, s), \] (37)

\[ \vdots \]

\[ \phi^v_{k+1,T}(j, r^*_M, s^*_M) \triangleq \max_{(r, s) \in \Gamma^v \setminus \{r_1^*, s_1^*\} \ldots} \tilde{S}^v_{k+1,T}(j, r, s). \] (38)

The optimal index set, for function \( v^j_k(x) \), is:

\[ I_k(j) \triangleq \{(r^*_1, s^*_1), (r^*_2, s^*_2), \ldots, (r^*_M, s^*_M)\}. \] (39)

Using these indexes, the order \( M^v \) equation for \( v^j_k(x) \), whose memory requirements are fixed in time, has the form:

\[ v^j_k(x) \triangleq \frac{1}{(2\pi)^{(d+n)/2}\Phi(y_{k+1})} \times \sum_{\ell=1}^{M^v} K^v_{k+1,T}(\ell, r^*_k, s^*_k) \times \exp \left\{ -\frac{1}{2} \left( x - \text{inv}(S^{-1}_{k+1,T}(\ell, r^*_k, s^*_k)) \right)^T \right\} \times \left( x - \text{inv}(S^{-1}_{k+1,T}(\ell, r^*_k, s^*_k)) \right) \} \] (40)

Remark V.1 In the companion paper to this article, see [20], a corresponding suboptimal scheme is developed to compute the \( q \) process. Combining the suboptimal schemes for the \( q \) and the \( v \), according the Theorem of Pardoux given by equation (19), one may compute smooth state estimates with a scheme whose memory requirements remain fixed in time.

REFERENCES


