EM algorithm for stochastic hybrid systems

Masaaki Fukasawa

Graduate School of Engineering Science
Osaka University

560-8531, Osaka, Japan
fukasawa@sigmath.es.osaka-u.ac.jp

Abstract

A stochastic hybrid system, also known as a switching diffusion, is a continuous-time Markov process with state space consisting of discrete and continuous parts. We consider parametric estimations of the Q matrix for the discrete state transitions and of the drift coefficient for the diffusion part based on a partial observation where the continuous state is monitored continuously in time, while the discrete state is unobserved. Extending results for hidden Markov models developed by Elliot et al. [1], we derive a finite-dimensional filter and the EM algorithm for stochastic hybrid systems.

Keywords: EM algorithm; filtering; stochastic hybrid system.

1 Introduction

A stochastic hybrid system (SHS, hereafter), also known as a switching diffusion [2], is a continuous-time Markov process $Z$ with state space $S = \{e_1, \ldots, e_k\} \times \mathbb{R}^d$ consisting of both discrete and continuous parts, namely, $\{e_1, \ldots, e_k\}$ and $\mathbb{R}^d$ respectively. The elements $\{e_i\}$ are, without loss of generality, specified as the standard basis of $\mathbb{R}^k$ in this article. Denoting by $\langle \cdot, \cdot \rangle$ the inner product of $\mathbb{R}^k$ or $\mathbb{R}^d$, $\langle e_i, e_j \rangle = \delta_{ij}$, where $\delta_{ij}$ is Kronecker’s delta. The discrete part of $Z$, denoted by $X$, can be seen as a continuous-time semi-Markov chain with state space $\{e_1, \ldots, e_k\}$ and “Q matrix” of the form $Q(Y_t) = [q_{ij}(Y_t)]$, where $Y$ is the continuous part of $Z$. In other words,

$$P(X_{t+h} = e_j|X_t = e_i, Y_t = y) = (\delta_{ji} + q_{ji}(y))h + o(h)$$

as $h \to 0$. Here, $Q(y) = [q_{ij}(y)]$ is a Q matrix for each $y$, that is, $q_{ij}(y) \geq 0$ for $j \neq i$ and $\sum_{j \neq i} q_{ji}(y) = -q_{ii}(y)$ for all $y \in \mathbb{R}^d$. The continuous part $Y$ is a semi-Markov
process on \( \mathbb{R}^d \) and defined as the solution of a stochastic differential equation

\[
\dot{Y} = \mu(X, Y) + \epsilon W
\]

for some \( \mathbb{R}^d \)-valued function \( \mu \), where \( \epsilon > 0 \) and \( W \) is a \( d \) dimensional white noise. The generator \( \mathcal{L} \) of this Markov process is given by

\[
\mathcal{L} f(e_i, y) = \langle \mu, \nabla_y f(e_i, y) \rangle + \frac{1}{2} \epsilon^2 \Delta_y f(e_i, y) + \sum_{j=1}^k (f(e_j, y) - f(e_i, y)) \langle e_j, Q(y)e_i \rangle.
\] (2)

There is a huge amount of literature on the analysis and applications of SHS. See e.g., \([3, 4, 7, 5]\) and the references therein. The author’s motivation to study SHS is its potential application to the analysis of single-molecule dynamics which has several unobservable switching states \([2]\). In this article, we consider parametric estimations of the Q matrix \( Q(y) = Q^\theta(y) \) and of the drift coefficient \( \mu(z) = \mu^\theta(z) \) based on a partial observation where \( Y \) is monitored continuously in time, while \( X \) is unobserved. When both \( Q \) and \( \mu \) do not depend on \( y \), the system is a hidden Markov model studied in Elliot et al. \([1]\).

Extending results developed in \([1]\), we derive a finite-dimensional filter and the EM algorithm for the SHS. In Section 2, we describe the basic properties of SHS as the solution of a martingale problem. In Section 3, we derive the likelihood function under complete observations of both \( X \) and \( Y \) on a time interval \([0, T]\). In Section 4, we consider the case where the discrete part \( X \) is unobservable, and construct a finite dimensional filter extending Elliot et al. \([1]\). In Section 5, again by extending Elliot et al. \([1]\), we construct the EM algorithm for parametric estimations under the partial observation.

## 2 A construction as a weak solution

Here we construct a SHS as a weak solution, that is, we construct a distribution on the path space \( D([0, T]; S) \) which is a solution of the martingale problem with the generator (2).

A direct application of Theorem (5.2) of Stroock \([6]\) provides the following.

**Theorem 2.1** Let \( \mu \) be a bounded Borel function and \( q_{ij} \), \( 1 \leq i, j \leq k \) be bounded continuous functions. Then, for any \( z \in S \), there exists a unique probability measure \( P_z \) on \( D([0, T]; S) \) such that \( Z_0 = z \) and

\[
f(Z_t) - \int_0^t \mathcal{L} f(Z_s) ds
\]

is a martingale under \( P_z \) for any \( f \in C^1_{0, \infty}((e_1, \ldots, e_k) \times \mathbb{R}^d) \), where \( Z : t \mapsto Z_t \) is the canonical map on \( D([0, T]; S) \). Moreover, \( Z \) is a strong Markov process with \( \{P_z\}_{z \in S} \).
The uniqueness part of Theorem 2.1 is important in this article. For the existence, we give below an explicit construction, which plays a key role to solve a filtering problem later.

First, we construct a SHS with $\mu = 0$ in a pathwise manner. Without loss of generality, assume $\epsilon = 1$. Note that $Y$ is then a $d$ dimensional Brownian motion. Let $(\Omega, \mathcal{F}, P^0)$ be a probability space on which a $d$ dimensional Brownian motion $Y$ and an i.i.d. sequence of exponential random variables $\{E_n\}$ that is independent of $Y$ are defined. Conditionally on $Y$, a time-inhomogeneous continuous-time Markov chain $X$ with (1) is defined using the exponential variables. More specifically, given $X_0 = e$, let

$$
\tau_1 = \min_{1 \leq j \leq k} \tau^j_1, \quad \tau^j_1 = \inf \left\{ t > 0; \int_0^t q_{ji}(Y_s)ds > E_j \right\}
$$

and $X_t = X_0$ for $0 \leq t < \tau_1$, $X_{\tau_1} = e_j$ with $j = \arg\min_1^k \tau^j_1$. The construction goes in a recursive manner; given $X_{\tau_n} = e_i$, let

$$
\tau_{n+1} = \min_{1 \leq j \leq k} \tau^j_{n+1}, \quad \tau^j_{n+1} = \inf \left\{ t > \tau_n; \int_{\tau_n}^t q_{ji}(Y_s)ds > E_{nk+j} \right\}
$$

and $X_t = X_{\tau_n}$ for $\tau_n \leq t < \tau_{n+1}$, $X_{\tau_{n+1}} = e_j$ with $j = \arg\min_1^k \tau^j_{n+1}$. Properties of the exponential distribution verifies the following lemma.

**Lemma 2.1** Assume $q_{ji}(y)$ is bounded and continuous in $y \in \mathbb{R}^d$ for each $(i, j)$. Then,

$$
P^0(X_{t+h} = e_j|X_t = e_i, Y) = (\delta_{ji} + q_{ji}(Y_t))h + o(h)
$$

and (1) with $P = P^0$.

By Itô’s formula, for any $f \in \mathcal{C}_{\text{b}}^{0,2}([e_1, \ldots, e_k] \times \mathbb{R}^d)$, we have

$$
f(X_{t+h}, Y_{t+h}) = f(X_t, Y_t) + \int_t^{t+h} \langle \nabla_y f(X_s, Y_s), dY_s \rangle + \frac{1}{2} \int_t^{t+h} \Delta_y f(X_s, Y_s)ds
$$

$$
+ \sum_{i<s \leq t+h} (f(X_s, Y_s) - f(X_{s-}, Y_s)),
$$

from which together with (1) it follows

$$
\lim_{h \to 0} \frac{E^0[f(X_{t+h}, Y_{t+h})|X_t = e_i, Y_t = y] - f(e_i, y)}{h} = \mathcal{L}^0 f(e_i, y),
$$

where $E^0$ is the expectation under $P^0$ and $\mathcal{L}^0 f = \mathcal{L} f$ with $\mu = 0$ in (2). Note only this, we have also that

$$
\mathcal{L}^0 f_i := f(X_t, Y_t) - \int_0^t \mathcal{L}^0 f(X_s, Y_s)ds
$$

is a martingale with respect to the filtration $\{\mathcal{F}_t\}$ generated by $Z = (X, Y)$. Even more importantly, Lemma 2.1 implies the following.
Lemma 2.2 Under the same conditions of Lemma 2.1, for any \( g \in C_0(\{e_1, \ldots, e_k\}) \),
\[
V_{i}^{0,g} := g(X_i) - \int_{0}^{t} \mathcal{L}_0^g(X_s)ds
\]
is a martingale with respect to the natural filtration of \( X \) under the conditional probability measure given \( Y \), where \( \mathcal{L}_0^g = \mathcal{L}_0^f \) with \( f(x, y) = g(x) \). In particular,
\[
X_i - \int_{0}^{t} Q(Y_s)X_s ds
\]
is a martingale under the conditional probability measure \( P^0(\cdot | Y) \).

Now we construct a SHS for a general bounded Borel function \( \mu \). Let
\[
\Lambda_T = \exp \left\{ \frac{1}{\varepsilon^2} \int_{0}^{T} \langle \mu(X_t, Y_t), dY_t \rangle - \frac{1}{2\varepsilon^2} \int_{0}^{T} |\mu(X_t, Y_t)|^2 dt \right\}.
\]
By the boundedness of \( \mu \), Novikov’s conditions is satisfied and so, \( \Lambda \) is an \( \{\mathcal{F}_t\} \)-martingale under \( P^0 \). Therefore,
\[
\frac{dP}{dP^0} = \Lambda_T
\]
defines a probability space \((\Omega, \mathcal{F}_T, P)\).

Theorem 2.2 Let \( Q = [q_{ij}] \) be a \( Q \) matrix-valued bounded continuous function and \( \mu \) be an \( \mathbb{R}^d \)-valued bounded Borel function. Under \( P \), \( Z = (X, Y) \) is a Markov process with generator (2). Further for any \( f \in C_0^0(\{e_1, \ldots, e_k\} \times \mathbb{R}^d) \),
\[
U_{i}^{f} := f(Z_i) - \int_{0}^{t} \mathcal{L}_f(Z_s)ds
\]
is an \( \{\mathcal{F}_t\} \) martingale.

Proof: By the Bayes formula,
\[
E[f(Z_{t+h})|\mathcal{F}_t] = \frac{E^0[\Lambda_T f(Z_{t+h})|\mathcal{F}_t]}{E^0[\Lambda_T|\mathcal{F}_t]} = E^0\left[ \frac{\Lambda_{t+h}}{\Lambda_t} f(Z_{t+h})|\mathcal{F}_t \right].
\]
Since
\[
\frac{\Lambda_{t+h}}{\Lambda_t} = \exp \left\{ \frac{1}{\varepsilon^2} \int_{t}^{t+h} \langle \mu(Z_s), dY_s \rangle - \frac{1}{2\varepsilon^2} \int_{t}^{t+h} |\mu(Z_s)|^2 ds \right\}
\]
and \( Z \) is Markov under \( P^0 \), \( E[f(Z_{t+h})|\mathcal{F}_t] = E[f(Z_{t+h})|Z_t] \), meaning that it is Markov under \( P \) as well. By Itô’s formula,
\[
d\Lambda_t = \frac{1}{\varepsilon^2} \Lambda_t \mu(Z_t) dY_t
and

$$
\Lambda_{t+h} f(Z_{t+h}) = \Lambda_t f(Z_t) + \int_t^{t+h} f(Z_s) d\Lambda_s + \int_t^{t+h} \Lambda_s dU_s^0 / f
+ \int_t^{t+h} \Lambda_s \mathcal{L}^0 f(Z_s) ds + \int_t^{t+h} \Lambda_s \langle \mu, \nabla_y f \rangle (Z_s) ds.
$$

Therefore,

$$
\Lambda_{t+h} U_{t+h}^f = \Lambda_t U_t^f + \int_t^{t+h} U_s^f d\Lambda_s + \int_t^{t+h} \Lambda_s dU_s^f,
$$

meaning that $\Lambda U^f$ is a martingale under $P^0$. The Bayes formula then implies that $U^f$ is a martingale under $P$. In particular, the generator is given by $\mathcal{L}$. ////

**Corollary 2.1** Under the same condition of Theorem 2.2,

$$
V_t := X_t - \int_0^t Q(Y_s) X_s ds
$$

is a martingale.

By the uniqueness result of Theorem 2.1, the law of $Z$ under $P$ coincides with $P_x$ with $x = Z_0$.

### 3 The likelihood under complete observations

Here we consider a statistical model $\{P^\theta\}_{\theta \in \Theta}$ and derive the likelihood under complete observation of a sample path $Z = (X, Y)$ on a time interval $[0, T]$. For each $\theta \in \Theta$, $P^\theta$ denotes the distribution on $D([0, T]; S)$ induced by a Markov process $Z$ with generator

$$
\mathcal{L}^\theta f(e_i, y) = \langle \mu^\theta, \nabla_y f \rangle (e_i, y) + \frac{1}{2} \epsilon^2 \Delta_y f(e_i, y) + \sum_{j=1}^k (f(e_j, y) - f(e_i, y)) (q^\theta_{ij}(y) e_i),
$$

where $\mu^\theta$ is a family of $\mathbb{R}^d$-valued bounded Borel functions and $Q^\theta = [q^\theta_{ij}]$ is a family of $Q$ matrix-valued bounded continuous functions. Note that $\epsilon > 0$ is almost surely identified from a path of $Y$ by computing its quadratic variation. It is therefore assumed to be known hereafter. The initial distribution $P^\theta \circ Z_0^{-1}$ is also assumed to be known and not to depend on $\theta$.

**Theorem 3.1** Let $\theta, \theta_0 \in \Theta$ and assume that

$$
y \mapsto \frac{q^\theta_{ij}(y)}{q^{\theta_0}_{ij}(y)}, \quad y \mapsto \frac{q^{\theta_0}_{ij}(y)}{q^{\theta}_{ij}(y)}
$$

5
are bounded for each \((i, j)\), where \(0/0 = 1\). Then, \(P^\theta\) is equivalent to \(P^{\theta_0}\), and the log likelihood

\[
L_T(\theta, \theta_0) := \log \frac{dP^\theta}{dP^{\theta_0}}(\{Z_t\}_{t \in [0, T]})
\]

is given by

\[
L_T(\theta, \theta_0) = \sum_{i, j=1}^k \left\{ \int_0^T \log \frac{q_{ji}^\theta(Y_t)}{q_{ji}^{\theta_0}(Y_t)} dN_i^\theta - \int_0^T (q_{ji}^\theta(Y_t) - q_{ji}^{\theta_0}(Y_t))(X_t, e_i) dt \right\}
+ \frac{1}{e^2} \int_0^T \langle \mu^\theta(Z_t) - \mu^{\theta_0}(Z_t), dY_t \rangle - \frac{1}{2e^2} \int_0^T \langle |\mu^\theta(Z_t)|^2 - |\mu^{\theta_0}(Z_t)|^2 \rangle dt,
\]

where \(N_i^\theta\) is the counting process of the transition from \(e_i\) to \(e_j\):

\[
N_i^\theta = \int_0^t \langle X_s, e_i \rangle (e_j, dX_s).
\]

**Proof:** The proof is standard but given for the readers’ convenience. Let

\[
L_i^\theta = \int_0^T \log \frac{q_{ji}^\theta(Y_t)}{q_{ji}^{\theta_0}(Y_t)} dN_i^\theta - \int_0^T (q_{ji}^\theta(Y_t) - q_{ji}^{\theta_0}(Y_t))(X_t, e_i) dt
\]

and

\[
L_i^0 = \frac{1}{e^2} \int_0^T \langle \mu^\theta(Z_t) - \mu^{\theta_0}(Z_t), dY_t \rangle - \frac{1}{2e^2} \int_0^T \langle |\mu^\theta(Z_t)|^2 - |\mu^{\theta_0}(Z_t)|^2 \rangle dt
= \frac{1}{e^2} \int_0^T \langle \mu^\theta(Z_t) - \mu^{\theta_0}(Z_t), dY_t - \mu^{\theta_0}(Z_t) dt \rangle - \frac{1}{2e^2} \int_0^T |\mu^\theta(Z_t) - \mu^{\theta_0}(Z_t)|^2 dt.
\]

By Itô’s formula,

\[
\exp[L_i^\theta] = 1 - \int_0^T \exp[L_i^\theta] (q_{ji}^\theta(Y_t) - q_{ji}^{\theta_0}(Y_t))(X_t, e_i) dt + \sum_{0 < j < t} \exp[L_i^\theta] - \exp[L_{i-}^\theta]
= 1 + \int_0^T \exp[L_i^\theta] \left( \frac{q_{ji}^\theta(Y_t)}{q_{ji}^{\theta_0}(Y_t)} - 1 \right) dN_i^\theta - q_{ji}^{\theta_0}(Y_t)(X_t, e_i) dt
\]

and by (5),

\[
dN_i^\theta - q_{ji}^{\theta_0}(Y_t)(X_t, e_i) dt = \langle X_t, e_i \rangle (e_j, dX_t - Q^{\theta_0}(Y_t) X_t dt).
\]

Therefore, by Corollary 2.1 \(\exp[L_i^\theta]\) and \(\exp[L_i^0]\) are orthogonal local martingales under \(P^{\theta_0}\). The assumed boundedness further implies that they are martingales. This implies that \(\mathcal{E}_t := \exp[L_t(\theta, \theta_0)]\) is a martingale under \(P^{\theta_0}\).
It only remains to show that $\mathcal{E}U^{0,f}$ is a martingale under $P^0$ for any $f \in C_b^{0,2}$, where

$$U_t^{0,f} = f(Z_t) - f(Z_0) - \int_0^t \mathcal{L}^0 f(Z_s) \, ds.$$ 

By Itô’s formula,

$$\mathcal{E}_t U_t^{0,f} = \int_0^t \mathcal{E}_s U_s^{0,f} \, ds + \int_0^t U_s^{0,f} \, d\mathcal{E}_s + \int_0^t (\mu^{0} - \mu^{0_b}, \nabla_y f)(Z_t) \, dt + \sum_{0 \leq j \leq t} \Delta \mathcal{E}_j \Delta U_j^{0,f}$$

and

$$\Delta \mathcal{E}_t = \mathcal{E}_t - \mathcal{E}_{t^-} = \sum_{i,j=1}^k \left( \frac{q_{ij}^0(Y_i)}{q_{jj}^0(Y_i)} - 1 \right) (N_i^{ij} - N_i^{ij^-}) \Delta U_t^{0,f} = f(X_t, Y_t) - f(X_{t^-}, Y_t).$$

Since

$$\Delta \mathcal{E}_t \Delta U_t^{0,f} = \mathcal{E}_t - \mathcal{E}_{t^-} \sum_{i,j=1}^k \left( \frac{q_{ij}^0(Y_i)}{q_{jj}^0(Y_i)} - 1 \right) (N_i^{ij} - N_i^{ij^-})(f(e_j, Y_i) - f(e_i, Y_i)).$$

we have

$$\sum_{0 \leq j \leq t} \Delta \mathcal{E}_j \Delta U_j^{0,f}$$

$$= \int_0^t \mathcal{E}_s - \mathcal{E}_{s^-} \sum_{i,j=1}^k \left( \frac{q_{ij}^0(Y_i)}{q_{jj}^0(Y_i)} - 1 \right) (f(e_j, Y_i) - f(e_i, Y_i)) \, dN_i^{ij}$$

$$= \int_0^t \mathcal{E}_s - \mathcal{E}_{s^-} \sum_{i,j=1}^k \left( \frac{q_{ij}^0(Y_i)}{q_{jj}^0(Y_i)} - 1 \right) (f(e_j, Y_i) - f(e_i, Y_i))(dN_i^{ij} - q_{jj}^0(Y_i)(X_{t^-}, e_i) \, dt]$$

$$+ \int_0^t \mathcal{E}_s \sum_{i,j=1}^k (q_{jj}^0(Y_i) - q_{jj}^0(Y_i)) (f(e_j, Y_i) - f(e_i, Y_i))(X_{t}, e_i) \, dt.$$ 

Consequently, we have

$$\mathcal{E}_t U_t^{0,f} = \int_0^t \mathcal{E}_s \, ds + \sum_{0 \leq j \leq t} \mathcal{E}_j \mathcal{E}_j U_j^{0,f}$$

$$\int_0^t \mathcal{E}_s \sum_{i,j=1}^k \left( \frac{q_{ij}^0(Y_i)}{q_{jj}^0(Y_i)} - 1 \right) (f(e_j, Y_i) - f(e_i, Y_i))(dN_i^{ij} - q_{jj}^0(Y_i)(X_{t^-}, e_i) \, dt],$$

which is a martingale under $P^0$ by (5).
4 A finite-dimensional filter

Here we extend the filtering theory of hidden Markov models developed by Elliot et al. [1] to the SHS

\[ dX_t = Q(Y_t)X_t dt + dV_t, \]
\[ dY_t = \mu(X_t, Y_t) dt + c dW_t, \]

where \( V \) is a martingale (recall Corollary 2.1). In this section we assume we observe only a continuous sample path \( Y \) on a time interval \([0, T]\) while \( X \) is hidden. The system is a hidden Markov model in [1] when both \( Q \) and \( \mu \) do not depend on \( Y \). By this dependence, \( V \) is not independent of \( W \) and so, the argument in [1] cannot apply here any more. We however show in this and the next sections that the results in [1] remain valid. Namely, a finite-dimensional filter and the EM algorithm can be constructed for the SHS. A key for this is Lemma 2.2.

Denote by \( \mathcal{F}_Y \) the natural filtration of \( Y \). The filtering problem is to infer \( X \) from the observation of \( Y \), that is, to compute \( E[X_t | \mathcal{F}_Y_t] \). The smoothing problem is to compute \( E[X_t | \mathcal{F}_Y_T] \) for \( t \leq T \). Denote \( E^0[H] = E^0[H | \mathcal{F}_Y_t] \) for a given integrable random variable \( H \), where \( E^0 \) is the expectation under \( P^0 \) in Section 2.

For a given process \( H \), the Bayes formula gives

\[ E^0[H_t | \mathcal{F}_Y_t] = E^0[H_t | \Lambda_t] E^0[\Lambda_t | \mathcal{F}_Y_t], \]

where \( \Lambda \) is defined by (4). Denoting \( \langle e_i, \mu(e_j, y) \rangle = c_{ij}(y) \), \( C(y) = [c_{ij}(y)] \), we can write \( \mu(Z_s) = C(Y_s)X_s \).

**Theorem 4.1** Under the same conditions of Theorem 2.2, if \( H \) is of the form

\[ dH_t = \alpha_t dt + \langle \beta_t, dX_t \rangle + \langle \delta_t, dY_t \rangle, \]

where \( \alpha, \beta, \delta \) are bounded predictable processes, then

\[ E^0[\Lambda_t H_t] = H_0 + \frac{1}{c^2} \int_0^t \langle C(Y_s)E^0_s[\Lambda_s H_s X_s] + c^2 E^0_s[\Lambda_s \delta_s], dY_s \rangle + \int_0^t E^0_s[\Lambda_s \alpha_s] + E^0_s[\Lambda_s \langle \beta_s, Q(Y_s)X_s \rangle] + E^0_s[\Lambda_s \langle \delta_s, C(Y_s)X_s \rangle] ds. \]  

**Proof:** Itô’s formula gives

\[ \Lambda_t H_t = H_0 + \int_0^t H_s d\Lambda_s + \int_0^t \Lambda_s dH_s + \int_0^t \Lambda_s \langle \delta_s, \mu(Z_s) \rangle ds. \]

Take the conditional expectation under \( P^0 \) given \( \mathcal{F}_Y_t \) to get (9). Here, we have used the fact that \( Y/e \) is a \( d \) dimensional Brownian motion under \( P^0 \) as well as Lemma 2.2.
Theorem 4.2 Under the same conditions of Theorem 2.2, for each \( i = 1, \ldots, k \),

\[
\langle e_i, E_0^t[A_tX_t] \rangle = \langle e_i, X_0 \rangle + \frac{1}{\epsilon^2} \int_0^t \langle e_i, E_0^s[A_sX_s] \rangle (C(Y_s)e_t, dY_s) + \int_0^t \langle e_i, Q(Y_s)E_0^s[A_sX_s] \rangle ds, \tag{10}
\]

and

\[
E_0^t[A_t] = 1 + \int_0^t \langle C(Y_s)E_0^s[A_sX_s], dY_s \rangle. \tag{11}
\]

Proof: Let \( H_t = \langle e_i, X_t \rangle \) and \( H_t = 1 \) in (9) to get (10) and (11) respectively. Here we have used that \( \langle e_i, X_s \rangle C(Y_s)X_s = \langle e_i, X_s \rangle C(Y_s)e_t \).

Note that

\[
X_s = \sum_{i=1}^k \langle e_i, X_s \rangle e_i
\]

and so, (10) is a linear equation on the vector valued process \( E_0^t[A_tX_t] \) that is easy to solve. Then (11) is also solved, and \( E[X_t|F^T] \) is obtained from (7).

Theorem 4.3 Under the same conditions of Theorem 2.2, for each \( i = 1, \ldots, k \), for any \( \tau \leq t \),

\[
\langle e_i, E_0^\tau[A_\tau X_\tau] \rangle = \langle e_i, E_0^\tau[A_\tau X_\tau] \rangle + \frac{1}{\epsilon^2} \int_\tau^t \langle e_i, E_0^s[A_sX_s] \rangle (C(Y_s)e_t, dY_s).
\]

Proof: Let \( H_t = \langle e_i, X_{t\wedge \tau} \rangle \) in (9).

This is also a linear equation and so, the smoothing problem \( E[X_t|F^T] \) is easily solved via (7).

5 The EM algorithm

Here we consider again the parametric family \( \{P^\theta\} \) introduced in Section 3. We assume that a continuous sample path \( Y \) is observed on a time interval \([0, T]\) while \( X \) is hidden. We construct the EM algorithm to estimate \( \theta \). Under the same assumptions as in Theorem 3.1, the law of \( Y \) under \( P^\theta \) is equivalent to that under \( P^{\hat{\theta}} \) and the log likelihood function is given by

\[
L^Y(\theta, \theta_0) = \log E^{\hat{\theta}} \left[ \frac{dP^\theta}{dP^{\theta_0}} | F^T \right].
\]

The maximum likelihood estimator is therefore given by

\[
\hat{\theta} = \arg\max_{\theta \in \Theta} L^Y(\theta, \theta_0).
\]
Note that \( \hat{\theta} \) does not depend on the choice of \( \theta_0 \) because by the Bayes formula,

\[
L^Y(\theta, \theta_0) = \log E^{\theta_0} \left[ \frac{dP^\theta_0}{dP^\theta_0} | \mathcal{F}^Y_t \right] - \log E^{\theta_0} \left[ \frac{dP^\theta_0}{dP^\theta_0} | \mathcal{F}^Y_{t-1} \right] = L^Y(\theta, \theta_1) - L^Y(\theta_0, \theta_1)
\]

for any \( \theta_1 \in \Theta \). Now, we recall the idea of the EM algorithm. Let

\[
Q(\theta^*, \theta) = E^\theta \left[ \log \frac{dP^\theta}{dP^\theta_0} | \mathcal{F}^Y_t \right].
\]

By Jensen’s inequality and (12),

\[
Q(\theta^*, \theta) \leq \log E^\theta \left[ \frac{dP^\theta}{dP^\theta_0} | \mathcal{F}^Y_t \right] = L^Y(\theta^*, \theta) = L^Y(\theta^*, \theta_0) - L^Y(\theta, \theta_0),
\]

which means that the sequence defined by

\[
\theta_{n+1} = \text{argmax}_{\theta \in \Theta} Q(\theta, \theta_n)
\]

makes \( L^Y(\theta_n, \theta_0) \) increasing. Under an appropriate condition the sequence \( \{\theta_n\} \) converges to the maximum likelihood estimator \( \hat{\theta} \), for which we refer to Wu [8].

The computation of \( Q(\theta, \theta_0) \) is a filtering problem for which can apply the results in Section 3. Now we state the main result of this article.

**Theorem 5.1** Let \( \Lambda \) be defined by (4) with \( \mu = \mu^0 \). Under the condition of Theorem 3.1 we have

\[
Q(\theta, \theta_0) = \frac{E^0_t[\Lambda_t L_t(\theta, \theta_0)]}{E^0_t[\Lambda_t]},
\]

\[
E^0_t[\Lambda_t L_t(\theta, \theta_0)] = \frac{1}{\epsilon^2} \int_0^t \langle c_i, E^0_s[\Lambda_s L_s(\theta, \theta_0)X_s]\rangle (C^0_s(Y_s)c_i, dY_s) + \int_0^t (A_s(\theta, \theta_0) + B_s(\theta, \theta_0) + D_s(\theta, \theta_0)) E^0_s[\Lambda_s X_s] ds,
\]

and for \( i = 1, \ldots, k, \)

\[
\langle e_i, E^0_t[\Lambda_t L_t(\theta, \theta_0)X_t] \rangle = \frac{1}{\epsilon^2} \int_0^t \langle e_i, E^0_s[\Lambda_s L_s(\theta, \theta_0)X_s]\rangle (C^0_s(Y_s)e_i, dY_s) + \int_0^t \langle e_i, E^0_s[\Lambda_s X_s]\rangle (C_s(\theta, \theta_0)e_i, dY_s) + \int_0^t \langle e_i, Q^0_s(Y_s)E^0_s[\Lambda_s L_s(\theta, \theta_0)X_s] + F_s(\theta, \theta_0) E^0_s[\Lambda_s X_s] \rangle ds,
\]

10
where

\[ A_s(\theta, \theta_0) = (A_{s,1}(\theta, \theta_0), \ldots, A_{s,k}(\theta, \theta_0)), \]

\[
A_{s,j}(\theta, \theta_0) = - \left\{ \sum_{j=1}^{k} (q^0_{ij}(Y_j) - q^0_{ij}(Y_s)) + \frac{1}{2e^2} \sum_{j=1}^{d} (c^0_{ij}(Y_s)^2 - c^0_{ij}(Y_s)^2) \right\} \langle e_i, X_t \rangle
\]

\[ B_s(\theta, \theta_0) = (B_{s,1}(\theta, \theta_0), \ldots, B_{s,k}(\theta, \theta_0)), \]

\[
B_{s,j}(\theta, \theta_0) = \sum_{j=1}^{k} q^0_{ij}(Y_s) \log \frac{q^0_{ij}(Y_s)}{q^0_{ij}(Y_s)} \]

\[ C_s(\theta, \theta_0) = C^0(Y_s) - C^0(\theta_s), \]

\[ D_s(\theta, \theta_0) = (D_{s,1}(\theta, \theta_0), \ldots, D_{s,d}(\theta, \theta_0)), \]

\[
D_{s,i}(\theta, \theta_0) = \frac{1}{e^2} \sum_{j=1}^{k} (c^0_{ij}(Y_s) - c^0_{ij}(Y_s)) c^0_{ij}(Y_s),
\]

\[ F_s(\theta, \theta_0) = [f_i(Y_s)]. \]

\[
f_{ij}(Y_s) = \begin{cases} q^0_{ij}(Y_s) \log \frac{q^0_{ij}(Y_s)}{q^0_{ij}(Y_s)} & \text{if } i \neq j, \\ -q^0_{ij}(Y_s) \log \frac{q^0_{ij}(Y_s)}{q^0_{ij}(Y_s)} & \text{if } i = j. \end{cases}
\]

and \( c^0(y) = \langle e_i, \mu^0(e_i, y) \rangle \), \( C^0(y) = [c^0(y)] \). Further, \( E_1[\Lambda_i X_t] \) and \( E_1[\Lambda_i] \) are respectively given by (10) and (11) with \( C = C^0 \) and \( Q = Q^0 \).

**Proof:** By Theorem 3.1, \( H_t := L_t(\theta, \theta_0) \) is of the form (8) with

\[
\alpha_t = -\sum_{i=1}^{k} \left\{ \sum_{j=1}^{k} (q^0_{ij}(Y_s) - q^0_{ij}(Y_s)) + \frac{1}{2e^2} \sum_{j=1}^{d} (c^0_{ij}(Y_s)^2 - c^0_{ij}(Y_s)^2) \right\} \langle e_i, X_t \rangle
\]

\[
= A_t(\theta, \theta_0) X_t,
\]

\[
\beta_t = \sum_{i,j=1}^{k} \log \frac{q^0_{ij}(Y_i)}{q^0_{ij}(Y_i)} (X_{i-1}, e_i) e_j,
\]

\[
\delta_t = \frac{1}{e^2} (C^0(Y_s) - C^0(\theta_s)) X_t = \frac{1}{e^2} C_t(\theta, \theta_0) X_t.
\]

Here we have used that \( \mu^0(Z_t) = C^0(Y_t)X_t \) and so,

\[
|\mu^0(Z_t)|^2 = \sum_{a,b,c} c^0_{ab}(Y_t) c^0_{ac}(Y_t) (e_b, X_t)(e_c, X_t) = \sum_{a,b} c^0_{ab}(Y_t)^2 (e_b, X_t).
\]
Since

\[ E^0_s[\Lambda_s \alpha_s] = A_s(\theta, \theta_0)E^0_s[\Lambda_s X_s], \]
\[ E^0_s[\Lambda_s(\beta_s, Q^{0b}(Y_s)X_{s-})] = B_s(\theta, \theta_0)E^0_s[\Lambda_s X_s], \]
\[ e^2 E^0_s[\Lambda_s \delta_s] = C_s(\theta, \theta_0)E^0_s[\Lambda_s X_s], \]
\[ E^0_s[\Lambda_s(\delta_s, C^{0b}(Y_s)X_s)] = D_s(\theta, \theta_0)E^0_s[\Lambda_s X_s], \]

(13) follows from (9).

By Itô’s formula, \( H_t = L_t(\theta, \theta_0)(e_i, X_i) \) is of the form (9) with

\[ \alpha_i = \langle e_i, X_i \rangle A_i(\theta, \theta_0) e_i, \]
\[ \beta_i = \langle e_i, X_i \rangle \sum_{j=1}^{k} \log \frac{q^0_{ij}(Y_i)}{q^{0i}_{ij}(Y_i)} e_j \]
\[ + \left( L_t(\theta, \theta_0) + (1 - 2\langle e_i, X_{i-} \rangle) \sum_{j=1}^{k} \log \frac{q^0_{ij}(Y_i)}{q^{0i}_{ij}(Y_i)} \langle X_{i-}, e_j \rangle \right) e_i \]
\[ = \langle e_i, X_{i-} \rangle \sum_{j=1}^{k} \log \frac{q^0_{ij}(Y_i)}{q^{0i}_{ij}(Y_i)} e_j + \sum_{j=1}^{k} \langle e_i, X_{i-} \rangle \log \frac{q^0_{ij}(Y_i)}{q^{0i}_{ij}(Y_i)} e_i \]
\[ + \left( L_t(\theta, \theta_0) - 2\langle e_i, X_{i-} \rangle \log \frac{q^0_{ij}(Y_i)}{q^{0i}_{ij}(Y_i)} \right) e_i \]
\[ \delta_i = \frac{1}{e^2} (e_i, X_i) C_i(\theta, \theta_0) e_i. \]

Note that

\[ E^0_s[\Lambda_s \alpha_s] = A_s(\theta, \theta_0) e_i, \]
\[ e^2 E^0_s[\Lambda_s \delta_s] = C_s(\theta, \theta_0) e_i, \]
\[ E^0_s[\Lambda_s(\delta_s, C^{0b}(Y_s)X_s)] = D_s(\theta, \theta_0) e_i, \]

and that

\[ E^0_s[\Lambda_s(\beta_s, Q^{0b}(Y_s)X_{s-})] \]
\[ = B_s(\theta, \theta_0) e_i + \langle e_i, Q^{0b}(Y_s) \rangle E^0_s[\Lambda_s L_s(\theta, \theta_0) X_s] \]
\[ + \sum_{j=1}^{k} \langle e_i, E^0_s[\Lambda_s X_{s-}] \rangle q^{0b}_{ij}(Y_s) \log \frac{q^0_{ij}(Y_s)}{q^{0i}_{ij}(Y_s)} - 2\langle e_i, E^0_s[\Lambda_s X_{s-}] \rangle q^{0b}_{ij}(Y_s) \log \frac{q^0_{ij}(Y_s)}{q^{0i}_{ij}(Y_s)} \]
\[ = B_s(\theta, \theta_0) e_i + \langle e_i, Q^{0b}(Y_s) \rangle E^0_s[\Lambda_s L_s(\theta, \theta_0) X_s] + F_s(\theta, \theta_0) E^0_s[\Lambda_s X_{s-}]. \]

Since

\[ E^0_s[\Lambda_s L_s(\theta, \theta_0)(e_i, X_s) X_s] = E^0_s[\Lambda_s L_s(\theta, \theta_0)(e_i, X_s)] e_i, \]

(9) implies (14).
Conflict of Interest: The author states that there is no conflict of interest.

References

[1] R. J. Elliot, L. Aggoun and J.B. Moore, *Hidden Markov Models*, Springer, 1995.

[2] K. Fujita, M. Ohmachi, K. Ikezaki, T. Yanagida and M. Iwaki, Direct visualization of human myosin II force generation using DNA origami-based thick filaments, *Communications Biology*, doi: 10.1038/s42003-019-0683-0.

[3] M. Ghosh, A. Arapostathis and S. Marcus, Ergodic control of switching diffusions, *SIAM J. Control Optim.* 35, 1952-1988 (1997).

[4] J.P. Hespanha, A model for stochastic hybrid systems with application to communication networks, *Nonlinear Analysis* 62, 1353-1383 (2005).

[5] X. Ki, O. Omotere, L. Qian and E.R. Dougherty, Review of stochastic hybrid systems with applications in biological systems modeling and analysis, *EURASIM J. Bioinform. Syst. Biol.* doi: 10.1186/s13637-017-0061-5.

[6] D. Stroock, Diffusion Processes Associated with Lévy Generators, *Z. Wahrscheinlichkeitstheorie verw. Gebiete* 32, 209-244 (1975).

[7] A.R. Teel, A. Subbaraman and A. Sferlazza, Stability analysis for stochastic hybrid systems: A survey, *Automatica* 50, 2435-2456 (2014).

[8] C.F.J. Wu, On the convergence properties of the EM algorithm, *Ann. Statist.* 11, 95-103 (1983).