Robust EM Algorithms For Markov Modulated Poisson Processes

R. J. Elliott [1] W. P. Malcolm [2]

Abstract

In this article we consider robust filtering and smoothing for Markov Modulated Poisson Processes (MMPPs). Using the EM algorithm, these filters and smoothers can be applied to estimate the parameters of our model. Our dynamics do not involve stochastic integrals and our new formulae, in terms of time integrals, are easily discretized.

Key Words: Expectation Maximization Algorithm, Counting Processes, Change of Measure, Martingales

1 Introduction

The well known EM algorithm [5] provides a scheme for solving a problem common in signal processing: estimating the parameters of a probability distribution for a known, partially observed dynamical system. This problem has received considerable attention for common signal models, such as the discrete time Gauss-Markov model, or the observation of a Markov process through a Brownian motion, [14][10]. In this article we propose EM algorithms for the so called Markov modulated Poisson process (MMPP).

An MMPP is a Poisson counting process whose rate of arrivals depends upon the state of a Markov chain. These models have enjoyed many successful applications in queueing theory and recently have been studied in the context of packet traffic estimation and biomedical, and optical signal processing.

The parameter estimation problem we consider, concerns computing estimates for the rate matrix of the Markov chain and the vector of Poisson intensities for the observation process. Traditionally the EM algorithm is implemented by maximizing a log likelihood function over a parameter space. In some applications this approach can lead to technical difficulties. For example the form of the log likelihood function could be complicated, or the maximization of this function might be difficult. Also the storage of estimated quantities can be a problem with the EM algorithm.

The implementations of the EM algorithms we present are the so called filter-based and smoother-based EM algorithms, [7][10]. In the filter-based scheme, the parameter estimates are computed by running a bank of four recursive filters whose only storage requirements are previous estimates. A fundamental feature of our EM algorithms is that no stochastic integrations are required. By using gauge transformations we compute Duncan-Mortenson-Zakai equations, where the observation processes appear as parameters, rather than as stochastic integrators.

The paper is organised as follows: in §2 the signal models for the state process and the observation process are defined. In §3 we briefly recall the EM algorithm. In §4 we compute a robust filter-based EM algorithm for an MMPP. Finally, in §5, we compute a smoother-based EM algorithm for an MMPP.

2 Signal Models

Initially we suppose that all processes are defined on the measurable space $(\Omega, \mathcal{F})$ with probability measure P.

2.1 The State Process

Suppose the state process $X = \{X_t, 0 \leq t\}$ is a finite state Markov chain. We use the canonical representation introduced in [7], so without loss of generality the state space of X is $\mathcal{L} = \{e_1, e_2, \ldots, e_m\}$, where e_i denotes a column vector in $\mathbb{R}^m$ with unity in the i^{th} position and zero elsewhere. The dynamics for this process are

$$X_t = X_0 + \int_0^t A\, X_s\, ds + M_t, \tag{1}$$

where M_t is a martingale and $A \in \mathbb{R}^{m \times m}$ is a rate matrix.

2.2 The Observation Process

Suppose that the state process X is observed through a counting process whose Doob–Meyer decomposition

[1] Department of Mathematical Sciences, University of Alberta, Edmonton, Canada, T6G21, R.Elliott@ualberta.ca, www.math.ualberta.ca/People/Facultypages/Elliott.R.html, and Department of Applied Mathematics, University of Adelaide, Adelaide Australia 5055, relliott@maths.adelaide.edu.au

[2] (i) Defence Science and Technology Organisation, PO Box 1500 Salisbury South Australia 5108, paul.malcolm@dsto.defence.gov.au

is

$$N_t = \int_0^t \langle X_s, \lambda \rangle ds + V_t. \tag{2}$$

Write $\mathcal{Y} = \{\mathcal{Y}_t\}$ for the filtration generated by N, where $\mathcal{Y}_t = \sigma\{N_s, 0 \le s \le t\}$. Here V is a $(P, \mathcal{Y})$ martingale and λ is a vector in $\mathbb{R}^m$ with strictly positive components $\lambda^{(1)}, \lambda^{(2)}, \dots, \lambda^{(m)}$.
Write

$$\mathcal{F}_t = \sigma\{X_s; 0 \le s \le t\}, \tag{3}$$
$$\mathcal{G}_t = \sigma\{N_s, X_s; 0 \le s \le t\}. \tag{4}$$

2.3 Reference Probability

We have under the 'real world' probability P dynamics of the form:

$$P \quad \begin{cases} dX_t = A\,X_t dt + dM_t, \\ dNt = \langle X_t, \lambda \rangle\, dt + dV_t. \end{cases} \tag{5}$$

Suppose $P^\dagger$ is a reference probability, under which X is still a Markov chain with dynamics (1) and the observation process N is a standard Poisson process. That is, under $P^\dagger$ the observation process is a Poisson process with unit intensity, independent of the process X. Denote by Λ the Radon-Nikodym derivative

$$\left.\frac{dP}{dP^\dagger}\right|_{\mathcal{G}_t} = \Lambda_{0,t}, \tag{6}$$

where

$$\begin{aligned} \Lambda_{0,t} &= \prod_{0<s\le t} \langle X_s, \lambda \rangle^{\Delta N_s} \exp\Big(\int_0^t (1 - \langle X_s, \lambda \rangle) ds\Big) \\ &= 1 + \int_0^t \Lambda_{s-}(\langle X_s, \lambda \rangle - 1)\,(dN_s - ds). \end{aligned} \tag{7}$$

3 The EM Algorithm

The EM algorithm is a two step iterative process for computing maximum likelihood estimates (MLEs). This process is usually terminated when some measure of convergence for the sequence of MLEs is observed.

Let θ index a given family of probability measures P_θ where $\theta \in \Theta$. All such measures P_θ, defined on a measurable space $(\Omega, \mathcal{F})$, are assumed absolutely continuous with respect to a fixed probability measure P_0. Suppose $\mathcal{Y} \subset \mathcal{F}$.
The two iterative steps in the EM algorithm are:

Expectation Step: Set $\theta^* = \hat{\theta}_k$ and compute $Q(\cdot, \theta^*)$, where

$$Q(\theta, \theta^*) = E_{\theta^*}\Big[\log \frac{dP}{dP_{\theta^*}} \mid \mathcal{Y}\Big], \tag{8}$$

Maximization Step: Maximize $Q(\theta, \theta^*)$ over the space Θ,

$$\hat{\theta}_{k+1} \in \underset{\theta \in \Theta}{\operatorname{argmax}} Q(\theta, \theta^*). \tag{9}$$

The so called filter-based form of the EM algorithm for a continuous time Markov chain observed in Brownian motion was presented in [7] and a robust version given in [10]. This method is based essentially on four quantities, each concerning the indirectly observed Markov process X and each computed using the information up to and including time t.

1. X_t, the state of the Markov chain. This quantity can be estimated from the observed counting process N by solving the stochastic equation

$$\begin{aligned} q_t = q_0 &+ \int_0^t A\, q_s\, ds \\ &+ \int_0^t \operatorname{diag}\{\langle \lambda, e_l \rangle - 1\}\,(dN_s - ds). \end{aligned} \tag{10}$$

   Here $\operatorname{diag}\{\langle \lambda, e_l \rangle - 1\}$ denotes a diagonal matrix with entries $\langle \lambda, e_1 \rangle - 1, \dots, \langle \lambda, e_m \rangle - 1$, A is the rate matrix for the process X and $P(X_t = e_i \mid \mathcal{Y}_t) = \dfrac{\langle q_t, e_i \rangle}{\sum_{l=1}^m \langle q_t, e_l \rangle}$.
   A proof of equation (10) is given in the Appendix.

2. J_t^i, the amount of time spent by the process X in state e_i up to time t:

$$J_t^i = \int_0^t \langle X_s, e_i \rangle\, ds. \tag{11}$$

3. $N_t^{i,j}$, the number of transitions $e_i \to e_j$ of X where $i \ne j$, up to time t:

$$N_t^{i,j} = \int_0^t \langle X_{s-}, e_i \rangle\, \langle dX_t, e_j \rangle. \tag{12}$$

4. G_t^i, the level integrals for the state e_i,

$$G_t^i = \int_0^t \langle X_s, e_i \rangle\, dN_s. \tag{13}$$

Using Bayes' Theorem, (see [9]), if $H = \{H_t, 0 \le t\}$ is any $\mathcal{G}$ adapted integrable process

$$E[H_t \mid \mathcal{Y}_t] = \frac{E^\dagger[\Lambda_{0,t} H_t \mid \mathcal{Y}_t]}{E^\dagger[\Lambda_{0,t} \mid \mathcal{Y}_t]}. \tag{14}$$

Here $E^\dagger[\cdot]$ denotes an expectation under the measure $P^\dagger$.
Write

$$\sigma(H_t) = E^\dagger[\Lambda_{0,t} H_t \mid \mathcal{Y}_t]. \tag{15}$$

Indexing the sequence of passes of the EM algorithm by $k = 1, 2, 3 \ldots$, the update formulae for the parameter estimates are:

$$[\hat{A}_{k+1}]_{i,j} = \frac{E[N_T^{i,j} \mid \mathcal{Y}_T]}{E[J_T^i \mid \mathcal{Y}_T]}, \tag{16}$$

and by (14), this is $\dfrac{\sigma(N_T^{i,j})}{\sigma(J_T^i)}$ and

$$\langle \hat{\lambda}_{k+1}, e_i \rangle = \frac{E[G_T^i \mid \mathcal{Y}_T]}{E[J_T^i \mid \mathcal{Y}_T]} = \frac{\sigma(G_T^i)}{\sigma(J_T^i)}. \tag{17}$$

The conditional expectations in equations (16) and (17) are computed using the *previous* (at k) parameter estimates for A and λ. This process is repeated, thereby generating a sequence of local maxima on the log likelihood surface.

4 A Robust Filter-Based EM Algorithm

In the previous section, the updates $[\hat{A}_k]_{(i,j)}$ and $\langle \hat{\lambda}_k, e_i \rangle$ were computed by evaluating the expectations in equations (16) and (17) respectively. However, it is in general not possible to compute individual dynamics for the processes $\sigma(J^i)$, $\sigma(N^{i,j})$ and $\sigma(G^i)$. It is however, possible to compute dynamics for the associated quantities $\sigma(J_t^i\, X_t)$, $\sigma(N_t^{i,j}\, X_t)$ and $\sigma(G_t^i\, X_t)$, where, for example,

$$\sigma(G_t^i\, X_t) = E^{\dagger}[\Lambda_t\, G_t^i\, X_t \mid \mathcal{Y}_t]. \tag{18}$$

This is the central idea behind the filter-based EM algorithm.

4.1 Vector-Valued Process Dynamics

Theorem 1 *The vector process $\sigma(J^i\, X) \in \mathbb{R}^m$ satisfies the stochastic integral equation*

$$\sigma(J_t^i\, X_t) = \int_0^t A\, \sigma(J_s^i\, X_s)\, ds + \int_0^t \langle q_s, e_i \rangle\, ds\, e_i + \int_0^t \mathit{diag}\{\langle \lambda, e_l \rangle - 1\} \sigma(J_{s-}^i\, X_{s-})\, (dN_s - ds), \tag{19}$$

where $\sigma(J_0^i\, X_0) = 0$ and q is the solution of equation (10).

Theorem 2 *The vector process $\sigma(N^{i,j}\, X) \in \mathbb{R}^m$ satisfies the stochastic integral equation*

$$\sigma(N_t^{i,j}\, X_t) = \int_0^t A \sigma(N_s^{i,j}\, X_s)\, ds + \int_0^t \langle q_s, e_i \rangle \langle A\, e_i, e_j \rangle\, ds\, e_j + \int_0^t \mathit{diag}\{\langle \lambda, e_l \rangle - 1 \rangle\} \sigma(N_{s-}^{i,j}\, X_{s-})\, (dN_s - ds) \tag{20}$$

where $\sigma(N_0^{i,i}\, X_0) = 0$ and q is the solution of equation (10).

Theorem 3 *The vector process $\sigma(G^i\, X) \in \mathbb{R}^m$ satisfies the stochastic integral equation*

$$\sigma(G_t^i\, X_t) = \int_0^t A\, \sigma(G_s^i\, X_s)\, ds + \int_0^t \langle q_{s-}, e_i \rangle \langle \lambda, e_i \rangle\, dN_s e_i + \int_0^t \mathit{diag}\{\langle \lambda, e_l \rangle - 1\} \sigma(G_{s-}^i\, X_{s-})\, (dN_s - ds) \tag{21}$$

where $\sigma(G_0^i\, X_0) = 0$ and q is the solution of equation (10).

A proof of Theorem 3 is given in the Appendix. Theorems 1 and 2 can be readily proven by similar means. By using the solutions of equation (19) (20) and (21), the updates for the parameter estimates are:

$$[\hat{A}_{k+1}]_{i,j} = \frac{\langle \sigma(N_T^{i,j}\, X_T), \mathbf{1} \rangle}{\langle \sigma(J_T^i\, X_T), \mathbf{1} \rangle}, \tag{22}$$

and

$$\langle \hat{\lambda}_{k+1}, e_i \rangle = \frac{\langle \sigma(G_T^i\, X_T), \mathbf{1} \rangle}{\langle \sigma(J_T^i\, X_T), \mathbf{1} \rangle}. \tag{23}$$

Here $\mathbf{1} = (1, 1, \ldots, 1)' \in \mathbb{R}^m$.

4.2 Observation-Parameterized Process Dynamics

The dynamics given by equations (19), (20) and (21), each contain stochastic Lebesgue-Stieltjes integral terms. These stochastic integrals with respect to N can be eliminated by using a version of a gauge transformation due to Clark [3].

Consider the diagonal matrix

$$\Gamma_t = \mathrm{diag}\{\gamma_t^i\} \in \mathbb{R}^{m \times m} \tag{24}$$

where $\gamma_t^i = \langle \lambda, e_i \rangle^{N_t} \exp\{(\langle \lambda, e_i \rangle - 1)\, t\}$, with $\gamma_0^i = 0$. Note that the matrix Γ_t^{-1} is nonsingular. Using the Itô rule one can show

$$\Gamma_t^{-1} = \int_0^t \mathrm{diag}\{\langle \lambda, e_l \rangle - 1\} \Gamma_s^{-1}\, ds + \int_0^t \Gamma_{s-}^{-1} \mathrm{diag}\Big\{ \frac{1}{\langle \lambda, e_l \rangle} - 1 \Big\}\, dN_s. \tag{25}$$

With $\overline{q}_t \triangleq \Gamma_t^{-1} q_t$ we have

$$\overline{q}_t = \overline{q}_0 + \int_0^t \Gamma_s^{-1}\, A\, \Gamma_s\, \overline{q}_s\, ds. \tag{26}$$

Equation (26) was established in [11].
For any $\mathcal{F}$-adapted integrable process H, write

$$\overline{\sigma}(H) = \Gamma_t^{-1} \sigma(H_t). \tag{27}$$

Let us first consider the process $\sigma(G^i X)$. Dynamics for the gauge transformed process $\overline{\sigma}(G^i_t X_t) = \Gamma_t^{-1}\sigma(G^i_t X_t)$ can be computed by applying the product rule,

$$d(\Gamma_t^{-1}\sigma(G^i_t X_t)) = \Gamma_{t-}^{-1} d(\sigma(G^i_t X_t)) + d\Gamma_t^{-1}\sigma(G^i_{t-} X_{t-}) + \Delta\Gamma_t^{-1}\Delta\sigma(G^i_t X_t). \tag{28}$$

$$\Delta H_t = H_t - H_{t-}. \tag{29}$$

The result of this calculation is,

$$\begin{aligned}\Gamma_t^{-1}\sigma(G^i_t X_t) = &\int_0^t \Gamma_s^{-1} A\,\Gamma_s \overline{\sigma}(G^i_s X_s)\,ds \\ &+ \int_0^t \langle \overline{q}_s, e_i\rangle\langle\lambda, e_i\rangle\,dN_s e_i \\ &+ \int_0^t \mathrm{diag}\Big\{\langle\lambda, e_l\rangle - 1\Big\}\overline{\sigma}(G^i_{s-} X_{s-})\,(dN_s - ds) \\ &+ \int_0^t \mathrm{diag}\Big\{\langle\lambda, e_l\rangle - 1\Big\}\overline{\sigma}(G^i_s X_s)\,ds \\ &+ \int_0^t \mathrm{diag}\Big\{\frac{1}{\langle\lambda, e_l\rangle} - 1\Big\}\overline{\sigma}(G^i_{s-} X_{s-})\,dN_s \\ &+ \int_0^t \mathrm{diag}\Big\{\frac{1}{\langle\lambda, e_l\rangle} - 1\Big\}\langle\overline{q}_s, e_i\rangle\langle\lambda, e_i\rangle\,dN_s\,e_i \\ &+ \int_0^t \mathrm{diag}\Big\{\frac{1}{\langle\lambda, e_l\rangle} - 1\Big\}\mathrm{diag}\{\langle\lambda, e_l\rangle - 1\} \times \\ &\qquad \overline{\sigma}(G^i_{s-} X_{s-})\,dN_s.\end{aligned} \tag{30}$$

Several stochastic integrals in equation (30) cancel, noting

$$\mathrm{diag}\Big\{\langle\lambda, e_l\rangle - 1\Big\} + \mathrm{diag}\Big\{\frac{1}{\langle\lambda, e_l\rangle} - 1\Big\} + \mathrm{diag}\Big\{\frac{1}{\langle\lambda, e_l\rangle} - 1\Big\}\mathrm{diag}\Big\{\langle\lambda, e_l\rangle - 1\Big\} = 0 \in \mathbb{R}^{m\times m}, \tag{31}$$

giving

$$\overline{\sigma}(G^i_t X_t) = \int_0^t \Gamma_s^{-1} A\,\Gamma_s\overline{\sigma}(G^i_s X_s)\,ds + \int_0^t \langle\overline{q}_s, e_i\rangle\,dN_s\,e_i. \tag{32}$$

The stochastic integral in equation (32) can be simplified by stochastic integration by parts,

$$\int_0^t \langle\overline{q}_s, e_i\rangle\,dN_s\,e_i = \langle\overline{q}_t, e_i\rangle\,N_t - \int_0^t N_s\langle d\overline{q}_s, e_i\rangle. \tag{33}$$

Finally our dynamics for $\overline{\sigma}(G^i_t X_t)$ read

$$\begin{aligned}\overline{\sigma}(G^i_t X_t) = &\int_0^t \Gamma_s^{-1}\,A\,\Gamma_s\overline{\sigma}(G^i_s X_s)\,ds \\ &+ \langle\overline{q}_t, e_i\rangle\,N_t e_i \\ &- \int_0^t N_s\langle d\overline{q}_s, e_i\rangle\,e_i.\end{aligned} \tag{34}$$

Similarly, one can apply the product rule to compute process dynamics for the quantities $\overline{\sigma}(J^i_t X_t)$ and $\overline{\sigma}(N^{i,j}_t X_t)$. The results of these calculations are, respectively,

$$\overline{\sigma}(J^i_t X_t) = \int_0^t \Gamma_s^{-1} A\,\Gamma_s\overline{\sigma}(J^i_s X_s)\,ds + \int_0^t \langle\overline{q}_s, e_i\rangle\,ds\,e_i \tag{35}$$

and

$$\begin{aligned}\overline{\sigma}(N^{i,j}_t X_t) = &\int_0^t \Gamma_s^{-1} A\,\Gamma_s\overline{\sigma}(N^{i,j}_s X_s)\,ds \\ &+ \int_0^t \langle\overline{q}_s, e_i\rangle\langle A\,e_i, e_j\rangle\,ds\,e_j.\end{aligned} \tag{36}$$

4.3 Discrete-Time Filters

To compute the update formulae given by equations (22) and (23), we must first compute the filtered quantities $\sigma(J^i_t X_t)$, $\sigma(N^{i,j}_t X_t)$ and $\sigma(G^i_t X_t)$, at $t = T$. Note also that the filters for some of these quantities depend upon q_t. To implement these filters on a digital computer or device, we must first choose an approximation in discrete time. Consider a regular partition on the time interval $[0, T]$. For $K \in \mathbb{N}_+$, write $\Delta = T/K$ and $t_n = n\Delta$, where $n = 1, 2, \ldots, K$. From equation (26) we have

$$\overline{q}_{tn} \approx \overline{q}_{t_{n-1}} + \Gamma_{t_{n-1}} A \Gamma_{t_{n-1}}^{-1}\overline{q}_{t_{n-1}}\,\Delta, \tag{37}$$

so $q_{t_n} = \Gamma_{t_n}\overline{q}_{t_n} \approx \Gamma_{t_n}\Gamma_{t_{n-1}}^{-1}[\mathbf{I} + \Delta\,A]\,q_{t_{n-1}}$. Writing the dynamics given by equation (34) recursively at sampling instants t_n and t_{n-1}, we get

$$\begin{aligned}\overline{\sigma}(G^i_{t_n} X_{t_n}) = &\,\overline{\sigma}(G^i_{t_{n-1}} X_{t_{n-1}}) \\ &+ \int_{t_{n-1}}^{t_n} \Gamma_s^{-1}\,A\,\Gamma_s\overline{\sigma}(G^i_s X_s)\,ds \\ &+ \langle\overline{q}_{t_n}, e_i\rangle\,N_{t_n} e_i - \langle\overline{q}_{t_{n-1}}, e_i\rangle\,N_{t_{n-1}}\,e_i \\ &- \int_{t_{n-1}}^{t_n} N_s\langle d\overline{q}_s, e_i\rangle\,e_i.\end{aligned} \tag{38}$$

Making the approximation

$$\begin{aligned}\int_{t_{n-1}}^{t_n} N_s\,\langle d\overline{q}_s, e_i\rangle\,e_i &= \int_{t_{n-1}}^{t_n} N_s\langle\Gamma_s^{-1}\,A\,\Gamma_s\overline{q}_s\,ds, e_i\rangle\,e_i \\ &\approx N_{t_{n-1}}\Gamma_{t_{n-1}}^{-1}\,\langle A\,q_{t_{n-1}}, e_i\rangle\,\Delta\,e_i\end{aligned} \tag{39}$$

and with some algebraic manipulation,

$$\begin{aligned}&\langle\overline{q}_{t_n}, e_i\rangle\,N_{t_n}\,e_i - \langle\overline{q}_{t_{n-1}}, e_i\rangle\,N_{t_{n-1}}\,e_i = \\ &\Gamma_{t_{n-1}}^{-1}\langle q_{t_{n-1}}, e_i\rangle\,(N_{t_n} - N_{t_{n-1}})\,e_i + \Delta\,\Gamma_{t_{n-1}}^{-1}\langle A\,q_{t_{n-1}}, e_i\rangle\,N_{t_n}\,e_i\end{aligned} \tag{40}$$

we see that

$$\begin{aligned}\overline{\sigma}(G^i_{t_n} X_{t_n}) \approx\ & \overline{\sigma}(G^i_{t_{n-1}} X_{t_{n-1}}) \\ & + \Gamma^{-1}_{t_{n-1}} A\, \Gamma_{t_{n-1}} \overline{\sigma}(G^i_{t_{n-1}} X_{t_{n-1}})\, \Delta \\ & + \Gamma^{-1}_{t_{n-1}} \langle q_{t_{n-1}}, e_i \rangle\, (N_{t_n} - N_{t_{n-1}})\, e_i \quad (41) \\ & + \Delta\, \Gamma^{-1}_{t_{n-1}} \langle A\, q_{t_{n-1}}, e_i \rangle\, N_{t_n}\, e_i \\ & - N_{t_{n-1}} \Gamma^{-1}_{t_{n-1}} \langle A\, q_{t_{n-1}}, e_i \rangle\, \Delta\, e_i .\end{aligned}$$

Now, by multiplying both sides of equation (41) on the left by the matrix Γ_{t_n}, we get,

$$\begin{aligned}\sigma(G^i_{t_n} X_{t_n}) =\ & \Gamma_{t_n} \Gamma^{-1}_{t_{n-1}} \sigma(G^i_{t_{n-1}} X_{t_{n-1}}) \\ & + \Gamma_{t_n} \Gamma^{-1}_{t_{n-1}} A\, \sigma(G^i_{t_{n-1}} X_{t_{n-1}}) \Delta \\ & + \Gamma_{t_n} \Gamma^{-1}_{t_{n-1}} \langle q_{t_{n-1}}, e_i \rangle\, (N_{t_n} - N_{t_{n-1}})\, e_i \\ & + \Gamma_{t_n} \Gamma^{-1}_{t_{n-1}} \langle A\, q_{t_{n-1}}, e_i \rangle\, (N_{t_n} - N_{t_{n-1}})\, e_i \\ =\ & \Psi_{t_n, t_{n-1}} [\mathbf{I} + \Delta\, A]\, \sigma(G^i_{t_{n-1}} X_{t_{n-1}}) \\ & + \Psi_{t_n, t_{n-1}} [\langle q_{t_{n-1}}, e_i \rangle + \Delta \langle A\, q_{t_{n-1}}, e_i \rangle] \times \\ & \quad (N_{t_n} - N_{t_{n-1}})\, e_i ,\end{aligned} \quad (42)$$

where $\Psi_{t_n, t_{n-1}} = \Gamma_{t_n} \Gamma^{-1}_{t_{n-1}}$.
After similar calculations the remaining discretized filters read

$$\begin{aligned}\sigma(N^{i,j}_{t_n} X_{t_n}) =\ & \Psi_{t_n, t_{n-1}} [\mathbf{I} + \Delta\, A]\, \sigma(N^{i,j}_{t_{n-1}} X_{t_{n-1}}) \\ & + \Psi_{t_n, t_{n-1}} \langle q_{t_{n-1}}, e_i \rangle \langle A\, e_i, e_j \rangle\, \Delta\, e_i ,\end{aligned} \quad (43)$$

$$\begin{aligned}\sigma(J^i_{t_n} X_{t_n}) =\ & \Psi_{t_n, t_{n-1}} [\mathbf{I} + \Delta\, A]\, \sigma(J^i_{t_{n-1}} X_{t_{n-1}}) \\ & + \Psi_{t_n, t_{n-1}} \langle q_{t_{n-1}}, e_i \rangle\, e_i\end{aligned} \quad (44)$$

and

$$q_{t_n} = \Psi_{t_n, t_{n-1}} [\mathbf{I} + \Delta A]\, q_{t_{n-1}} . \quad (45)$$

Summarizing the results from the previous sections, our filter-based EM Algorithm reads

Step 1 Choose $[\hat{A}_0]_{i,j}$ and $\hat{\lambda}_0$.
Step 2 Using (22) and (23), compute the MLEs, $[\hat{A}_{k+1}]_{i,j}$ and $\hat{\lambda}_{k+1}$.
Step 3 Decide to stop, or, continue from step 2.

5 A Robust Smoother-Based EM Algorithm

In many implementations of the EM algorithm, for example [14], the expectation step is completed with smoothed, rather than filtered estimates. Typically the smoothing scheme used is the so called 'fixed interval smoother'. Computing smoothing schemes for MMPPs can be particularly difficult, see for example [16]. One source of this difficulty is the task of developing backwards dynamics. This task usually leads to constructing stochastic integrals evolving backwards in time. However, the approach we use to develop smoothing algorithms completely avoids these difficulties. To compute our smoothers we exploit a duality between forward and backwards robust dynamics and, as a consequence, do not need to consider backward stochastic integration at all.

Recall the state estimation MMPP smoother presented in [12]. For a smoothed estimate for the process $X \in \mathbb{R}^m$, we wish to evaluate the expectation $E[X_t \mid \mathcal{Y}_T]$, where $0 \le t \le T$. By Bayes' rule, [9],

$$E[X_t \mid \mathcal{Y}_T] = \frac{E^\dagger[\Lambda_{0,T} X_t \mid \mathcal{Y}_T]}{E^\dagger[\Lambda_{0,T} \mid \mathcal{Y}_T]} . \quad (46)$$

Consider the numerator of equation (46),

$$\begin{aligned} r_t & \triangleq E^\dagger[\Lambda_{0,T} X_t \mid \mathcal{Y}_T] \\ & = E^\dagger[\Lambda_{0,t} \Lambda_{t,T} X_t \mid \mathcal{Y}_T] \\ & = E^\dagger[E^\dagger[\Lambda_{0,t} \Lambda_{t,T} X_t \mid \mathcal{Y}_T \vee \mathcal{F}_t] \mid \mathcal{Y}_T] \\ & = E^\dagger[\Lambda_{0,t} X_t E^\dagger[\Lambda_{t,T} \mid \mathcal{Y}_T \vee \mathcal{F}_t \mid \mathcal{Y}_T] .\end{aligned} \quad (47)$$

Under the measure $P^\dagger$, X is a Markov process, so the inner expectation in the previous line of (47) is

$$E^\dagger[\Lambda_{t,T} \mid \mathcal{Y}_T \vee \mathcal{F}_t] = E^\dagger[\Lambda_{t,T} \mid \mathcal{Y}_T \vee \sigma\{X_t\}] . \quad (48)$$

Write

$$v^i_t = E^\dagger[\Lambda_{t,T} \mid \mathcal{Y}_T \text{ and } X_t = e_i] . \quad (49)$$

Omitting further calculations, it can be shown, see [12], that

$$\begin{aligned} r_t = \langle q_t, e_1 \rangle \langle v_t, e_1 \rangle\, e_1 + \langle q_t, e_2 \rangle \langle v_t, e_2 \rangle\, e_2, \dots \\ + \langle q_t, e_m \rangle \langle v_t, e_m \rangle\, e_m \in \mathbb{R}^m .\end{aligned} \quad (50)$$

The normalized smoothed state estimate of X is then

$$E[X_t \mid \mathcal{Y}_t] = \frac{r_t}{\langle r_t, \mathbf{1} \rangle} . \quad (51)$$

Note that

$$\begin{aligned} \langle r_t, \mathbf{1} \rangle & = \langle q_t, v_t \rangle \\ & = E^\dagger[\Lambda_{0,T} \langle X_t, \mathbf{1} \rangle \mid \mathcal{Y}_T] \\ & = E^\dagger[\Lambda_{0,T} \mid \mathcal{Y}_T] \end{aligned} \quad (52)$$

is independent of t. Therefore,

$$\frac{d}{dt} \langle r_t, \mathbf{1} \rangle = \frac{d}{dt} \langle q_t, v_t \rangle = 0 . \quad (53)$$

The vector $v_t = (\langle v_t, e_1 \rangle, \langle v_t, e_2 \rangle, \dots, \langle v_t, e_m \rangle)$ incorporates the *extra* information obtained from the observations between t and T. Computing dynamics for v can be difficult. However, by exploiting a duality we can compute dynamics for its dual process $\overline{v}$. What we must do, is find a process $\overline{v}$ such that the following duality holds

$$\langle \overline{q}_t, \overline{v}_t \rangle = \langle \Gamma_t^{-1} q_t, \Gamma_t v_t \rangle = \langle q_t, v_t \rangle, \qquad \text{for all } t \in [0,T] \tag{54}$$

That is, $\overline{v}_t \triangleq \Gamma_t v_t$. Using (53) one can show that,

$$\frac{d\overline{v}_t}{dt} = \Gamma_t A^* \Gamma_t^{-1} \overline{v}_t \tag{55}$$

where $\overline{v}_T = \Gamma_T v_T = \Gamma_T \mathbf{1}$.
Following the same strategy above, we consider the duality

$$\begin{aligned} \langle \sigma(G_t^i X_t), v_t \rangle &= \langle \Gamma_t^{-1} \sigma(G_t^i X_t), \Gamma_t v_t \rangle \\ &= \langle \overline{\sigma}(G_t^i X_t), \overline{v}_t \rangle. \end{aligned} \tag{56}$$

Now define

$$\widetilde{\sigma}(G_t^i X_t) \triangleq \overline{\sigma}(G_t^i X_t) - \langle \overline{q}_t, e_i \rangle N_t e_i. \tag{57}$$

Then

$$\begin{aligned} d\widetilde{\sigma}(G_t^i X_t) &= \Gamma_t^{-1} A \Gamma_t \overline{\sigma}(G_t^i X_t)\, dt \\ &\quad - N_t \langle \Gamma_t^{-1} A \Gamma_t \overline{q}_t, e_i \rangle e_i \, dt. \end{aligned} \tag{58}$$

Now

$$\langle \overline{\sigma}(G_t^i X_t), \overline{v}_t \rangle = \langle \widetilde{\sigma}(G_t^i X_t), \overline{v}_t \rangle + N_t \langle \overline{q}_t, e_i \rangle \langle \overline{v}_t, e_i \rangle \tag{59}$$

and

$$N_t \langle \overline{q}_t, e_i \rangle \langle \overline{v}_t, e_i \rangle = N_t \langle q_t, e_i \rangle \langle v_t, e_i \rangle. \tag{60}$$

From the dynamics of $\widetilde{\sigma}(G_t^i X_t)$ we have

$$\begin{aligned} d\langle \widetilde{\sigma}(G_t^i X_t), \overline{v}_t \rangle &= \langle \Gamma_t^{-1} A \Gamma_t \overline{\sigma}(G_t^i X_t), \overline{v}_t \rangle \, dt \\ &\quad - N_t \langle \Gamma_t^{-1} A \Gamma_t \overline{q}_t, e_i \rangle \langle e_i, \overline{v}_t \rangle \, dt \\ &\quad - \langle \widetilde{\sigma}(G_t^i X_t), \Gamma_t A \Gamma_t^{-1} \overline{v}_t \rangle \, dt \\ &= \langle \Gamma_t^{-1} A \Gamma_t \overline{\sigma}(G_t^i X_t), \overline{v}_t \rangle \, dt \\ &\quad - N_t \langle \Gamma_t^{-1} A \Gamma_t \overline{q}_t, e_i \rangle \langle e_i, \overline{v}_t \rangle \, dt \\ &\quad - \langle \overline{\sigma}(G_t^i X_t) - \\ &\qquad \langle \overline{q}_t, e_i \rangle N_t e_i, \Gamma_t A^* \Gamma_t^{-1} \overline{v}_t \rangle \, dt \\ &= -N_t \langle \Gamma_t^{-1} A \Gamma_t \overline{q}_t, e_i \rangle \langle e_i, \overline{v}_t \rangle \, dt \\ &\quad + N_t \langle \overline{q}_t, e_i \rangle \langle \Gamma_t^{-1} A \Gamma_t e_i, \overline{v}_t \rangle \, dt \end{aligned} \tag{61}$$

i.e.

$$\begin{aligned} \langle \widetilde{\sigma}(G_T^i X_T), \overline{v}_T \rangle &= -\int_0^T N_s \langle \Gamma_s^{-1} A \Gamma_s \overline{q}_s, e_i \rangle \langle e_i, \overline{v}_s \rangle \, ds \\ &\quad + \int_0^T N_s \langle \overline{q}_s, e_i \rangle \langle \Gamma_s^{-1} A \Gamma_s e_i, \overline{v}_s \rangle \, ds. \end{aligned} \tag{62}$$

Therefore

$$\begin{aligned} \langle \sigma(G_T^i X_T), v_T \rangle &= \langle \overline{\sigma}(G_T^i \,, X_T), \overline{v}_T \rangle \\ &= \langle \widetilde{\sigma}(G_T^i X_T), \overline{v}_T \rangle + N_T \langle \overline{q}_T, e_i \rangle \langle \overline{v}_T, e_i \rangle \\ &= -\int_0^T N_s \langle \Gamma_s^{-1} A \Gamma_s \overline{q}_s, e_i \rangle \langle e_i, \overline{v}_s \rangle \, ds \\ &\quad \int_0^T N_s \langle \overline{q}_s, e_i \rangle \langle \Gamma_s^{-1} A \Gamma_s e_i, \overline{v}_s \rangle \, ds + \\ &\quad N_T \langle q_T, e_i \rangle \langle v_T, e_i \rangle \\ &= -\int_0^T N_s \langle A q_s, e_i \rangle \langle e_i, v_s \rangle \, ds \\ &\quad + \int_0^T N_s \langle q_s, e_i \rangle \langle e_i, A^* v_s \rangle \, ds \\ &\quad + N_T \langle q_T, e_i \rangle \langle v_T, e_i \rangle. \end{aligned} \tag{63}$$

By using similar calculations one can also show that,

$$\langle \overline{\sigma}(J_T^i X_T), \overline{v}_T \rangle = \int_0^T \langle q_s, e_i \rangle \langle v_s, e_i \rangle \, ds, \tag{64}$$

and

$$\langle \overline{\sigma}(N_T^{i,j} X_T), \overline{v}_T \rangle = \int_0^T \langle A\, e_i, e_j \rangle \langle q_s, e_i \rangle \langle v_s, e_j \rangle \, ds. \tag{65}$$

Recalling (16) and (17), our smoother-based update equations read

$$[\hat{A}_{k+1}]_{i,j} = [\hat{A}_k]_{i,j} \frac{\int_0^T \langle q_s, e_i \rangle \langle v_s, e_j \rangle \, ds}{\int_0^T \langle q_s, e_i \rangle \langle v_s, e_i \rangle \, ds} \tag{66}$$

and

$$\hat{\lambda}_{k+1} = \frac{\int_0^T N_s \langle q_s, e_i \rangle \langle e_i, A^* v_s \rangle \, ds}{\int_0^T \langle q_s, e_i \rangle \langle v_s, e_i \rangle \, ds} - \frac{\int_0^T N_s \langle A\, q_s, e_i \rangle \langle e_i, v_s \rangle \, ds + N_T \langle q_T, e_i \rangle \langle v_T, e_i \rangle}{\int_0^T \langle q_s, e_i \rangle \langle v_s, e_i \rangle \, ds}. \tag{67}$$

Summarizing, our smoother-based EM Algorithm reads

Step 1 Choose $\hat{A}_0$ and $\hat{\lambda}_0$.
Step 2 Using (66) and (67), compute the MLEs, $\hat{A}_{k+1}$ and $\hat{\lambda}_{k+1}$.
Step 3 Decide to stop, or, continue from step 2.

6 Appendix

Derivation of equation (10):

We wish to estimate X given the observations $\mathcal{Y}$ of N. By Bayes' rule

$$E[X_t \mid \mathcal{Y}_t] = \frac{E^\dagger[\Lambda_{0,t} X_t \mid \mathcal{Y}_t]}{E^\dagger[\Lambda_{0,t} \mid \mathcal{Y}_t]}. \tag{68}$$

Note $\langle X_t, \mathbf{1}\rangle = 1$.
So,

$$\begin{aligned} \langle E^\dagger[\Lambda_{0,t} X_t \mid \mathcal{Y}_t], \mathbf{1}\rangle &= E^\dagger[\Lambda_{0,t}\langle X_t, \mathbf{1}\rangle \mid \mathcal{Y}_t] \\ &= E^\dagger[\Lambda_{0,t} \mid \mathcal{Y}_t]. \end{aligned} \tag{69}$$

That is, if we write

$$q_t = E^\dagger[\Lambda_{0,t} X_t \mid \mathcal{Y}_t], \tag{70}$$

then

$$p_t \triangleq E[X_t \mid \mathcal{Y}_t] = \frac{q_t}{\langle q_t, \mathbf{1}\rangle}. \tag{71}$$

To compute the expectation at (70), we first apply the product rule to determine the decomposition for the process ΛX:

$$\begin{aligned} \Lambda_{0,t} X_t &= X_0 + \int_0^t \Lambda_{0,s} A X_s \, ds + \int_0^t \Lambda_{s-} \, dM_s + \\ &\quad \int_0^t X_{s-}(\langle X_{s-}, \lambda\rangle - 1)\Lambda_{0,s-}\,(dN_s - ds) \\ &= X_0 + \int_0^t \Lambda_{0,s}\, A\, X_s\, ds + \int_0^t \Lambda_{s-}\, dM_s + \\ &\quad \sum_{i=1}^m \int_0^t \langle X_{s-}, e_i\rangle(\langle \lambda, e_i\rangle - 1)\Lambda_{0,s-}(dN_s - ds)\, e_i \end{aligned} \tag{72}$$

By conditioning both sides of equation (72) on $\mathcal{Y}_t$ under the reference probability $P^\dagger$, it then follows that the process q has the dynamics

$$\begin{aligned} q_t &= q_0 + \int_0^t A\, q_s\, ds \\ &\quad + \int_0^t \mathrm{diag}\{\langle \lambda, e_l\rangle - 1\} q_{s-}\,(dN_s - ds). \end{aligned} \tag{73}$$

$\square$.

Proof of Theorem 3

To compute the dynamics of the process $\sigma(G^i X)$, we must evaluate the expectation $E^\dagger[\Lambda_t\, G_t^i\, X_t \mid \mathcal{Y}_t]$. Using the product rule, we compute the decomposition of the process $G X \Lambda$,

$$\begin{aligned} \Lambda_t G_t^i X_t &= \int_0^t \Lambda_{s-} X_s \langle X_s, e_i\rangle\, dN_s \\ &\quad + \int_0^t \Lambda_s G_s^i\, A\, X_s\, ds \\ &\quad + \int_0^t \Lambda_s G_s^i\, dM_s \\ &\quad + \int_0^t G_s^i\, X_s\, \Lambda_{s-}(\langle X_s, \lambda\rangle - 1)\,(dN_s - ds) \\ &\quad + \int_0^t X_s \langle X_s, e_i\rangle \Lambda_{s-}(\langle X_s, \Lambda\rangle - 1)\,(dN_s - ds). \end{aligned} \tag{74}$$

The result follows by conditioning both sides of equation (74) on $\mathcal{Y}_t$, using a version of Fubini's theorem [17] and noting that under the measure $P^\dagger$ the process N is a standard Poisson process. Consequently we see that

$$\begin{aligned} E^\dagger[\Lambda_t G_t^i X_t \mid \mathcal{Y}_t] &= \int_0^t E^\dagger[\Lambda_{s-} X_s \langle X_s, e_i\rangle \mid \mathcal{Y}_s]\, dN_s \\ &\quad + \int_0^t E^\dagger[\Lambda_s G_s^i\, A\, X_s \mid \mathcal{Y}_s]\, ds \\ &\quad + \int_0^t E^\dagger[\Lambda_s G_s^i \mid \mathcal{Y}_s]\, dM_s \\ &\quad + \int_0^t E^\dagger[G_s^i\, X_s\, \Lambda_{s-}(\langle X_s, \lambda\rangle - 1) \mid \mathcal{Y}_s]\,(dN_s - ds) \\ &\quad + \int_0^t E^\dagger[X_s \langle X_s, e_i\rangle \Lambda_{s-}(\langle X_s, \Lambda\rangle - 1) \mid \mathcal{Y}_s]\,(dN_s - ds) \\ &= \int_0^t A\, \sigma(G_s^i X_s)\, ds \\ &\quad + \int_0^t \mathrm{diag}\{\langle \lambda, e_l\rangle - 1\}\, \sigma(G_s^i\, X_s)\,(dN_s - ds) \\ &\quad + \int_0^t \langle q_s, e_i\rangle \langle \lambda, e_i\rangle\, dN_s e_i. \end{aligned} \tag{75}$$

$\square$.

References

[1] Brémaud, P., *Point Processes and Queues, Martingale Dynamics*, Springer Verlag, 1981.

[2] Campillo F., Le Gland F., MLE for partially observed diffusions: Direct maximization vs. the EM algorithm, Stochastic Processes and their Applications, volume 33, number 2, pp. 245-274, 1989.

[3] Clark, J. M., The Design of Robust Approximations to the Stochastic Differential Equations for Nonlinear Filtering, in J. K. Skwirzynski Ed, Communications Systems and Random Process Theory, Darlington 1977, Alphen aan den Rijn, the Netherlands, Sijthoff and Noorhoff 1978, pp. 721-734.

[4] Dembo A., Zeitouni O., Parameter Estimation of partially observed continuous time stochastic processes via the EM algorithm, Stochastic Processes and their Applications, volume 23, pp. 91-113, 1986. Parameter Estimation of partially observed continuous time stochastic processes via the EM algorithm, Stochastic Processes and their Applications, volume 23, pp. 91-113, 1986.

[5] Demster, A. D., Laird N. M., Rubin D. B., Maximum Likelihood from Incomplete Data via the EM Algorithm, J. Roy. Stat. Soc., volume B-39, pp. 1-37, 1997.

[6] Elliott, R. J., *Stochastic Calculus and its Applications*, Springer Verlag 1982.

[7] Elliott, R. J., New Finite Dimensional Filters and Smoothers for Noisily Observed Markov Chains, IEEE Transactions on Information Theory, Vol 39, Number 1, pp. 265-271, January 1993.

[8] Elliott, R. J., Exact adaptive filters for Markov chains observed in Gaussian noise, Automatica, volume 30, number 9, pp. 1399-1408, September 1994.

[9] Elliott, R. J., Aggoun, L., Moore, J. B., *Hidden Markov Models Estimation and Control*, Springer Verlag Applications of Mathematics Series 29, 1995.

[10] James, M. R., Krishnamurthy, V., Le Gland, F., Time Discretization for Continuous Time Filters and Smoothers for HMM Parameter Estimation, Transactions on Information Theory, volume 42, number 2, pp. 593-605, March 1998.

[11] Malcolm, W. P., Robust Filtering And Estimation With Poisson Observations, PhD Thesis, The Australian National University, 1999.

[12] Malcolm, W. P., Elliott R. J., A General Smoothing Equation For Poisson Observations, 38th I.E.E.E Conference on Decision and Control, December 1999, Phoenix USA.

[13] Royden, H., *Real Analysis*, Third Edition, Macmillan Publishing Company 1988.

[14] Shumway, R. H., Stoffer D. S., An Approach to Time Series Smoothing and Forecasting using the EM Algorithm, Journal of Time Series Analysis, volume 3, pp. 682-687, May 1989.

[15] D. L. Snyder., Filtering and Detection for Doubly Stochastic Poisson Processes, I.E.E.E. Transactions on Information Theory, volume, IT-18, number 1, pp. 91-102, 1972.

[16] D. L. Snyder., Smoothing for Doubly Stochastic Poisson Processes, I.E.E.E. Transactions on Information Theory, volume, IT-18, number 5, pp. 558-562, 1972.

[17] Wong E., Hajek, B., Stochastic Processes in Engineering Systems, Springer Verlag, New York, 1985.

[18] Zeitouni O., Dembo A, Exact filters for the estimation of the number of transitions of finite-state continuous time Markov processes, IEEE Transactions on Information Theory, volume 34, pp. 890-893, July 1988.