Discrete-Time Expectation Maximization Algorithms for Markov-Modulated Poisson Processes

Robert J. Elliott and W. P. Malcolm

***Abstract*—In this paper, we consider parameter estimation Markov-modulated Poisson processes via robust filtering and smoothing techniques. Using the expectation maximization algorithm framework, our filters and smoothers can be applied to estimate the parameters of our model in either an online configuration or an offline configuration. Further, our estimator dynamics do not involve stochastic integrals and our new formulas, in terms of time integrals, are easily discretized, and are written in numerically stable forms in W. P. Malcolm, R. J. Elliott, and J. van der Hoek, "On the numerical stability of time-discretized state estimation via clark transformations," presented at the IEEE Conf. Decision Control, Mauii, HI, Dec. 2003.**

***Index Terms*—Change of measure, counting processes, expectation maximization (EM) algorithm, martingales.**

I. Introduction

The well-known expectation maximization (EM) algorithm [8], [17] 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 Gaus-Markov model or the observation of a Markov process through a Brownian motion, [10], [24]. In this paper, we propose EM algorithms for the so-called Markov-modulated Poisson process (MMPP).

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

Since our hidden-state process models are continuous-time Markov chains, 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 [11], [21], [22]. In some applications, this approach can lead to technical difficulties. For example, the form of the log-likelihood function could be complicated or the operation of maximization of this function might be difficult.

The implementations of the EM algorithms we present are the so-called filter-based and smoother-based EM algorithm [4], [10]. In the filter-based scheme, the parameter estimates are computed online by running a set of four recursive filters whose only storage requirements are previous estimates. Adapting the transformation techniques introduced by Clark [1], we compute the so-called robust versions of these filter, where the observation processes appear as parameters rather than as stochastic integrators. These formulations have been shown to have some numerical advantages [16]. Our smoother-based EM algorithm exploits a type of identity between the forward robust filter and its reverse-time counterpart. Smoothed estimates are obtained without recourse to stochastic integration.

The paper is organized as follows. In Section II, the signal models for the state and observation processes are defined; our reference probability measure is also defined in this section. In Section III, we briefly recall the EM algorithm and compute a filter-based EM algorithm for MMPPs. In this section, we also compute robust filter dynamics that do not include stochastic integrals. In Section IV, we compute a robust smoother-based EM algorithm for an MMPP. Finally, in Section V, we compute a discrete-time data-recursive smoother-based EM algorithm for an MMPP.

II. Dynamics and Reference Probability

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

A. State Process Dynamics

Suppose that the state process $X = \{X_t, 0 \le t\}$ is a continuous-time finite-state Markov chain with rate matrix A and an initial probability distribution p_0. We now use the well-known canonical representation for a Markov chain, that is, without loss of generality, the state space of X is $\mathcal{L} = \{\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n\}$, where $\boldsymbol{e}_i$ denotes a column vector in $\mathbb{R}^n$ with unity in the ith position and zero elsewhere. The dynamics for this process are

$$X_t = X_0 + \int_0^t A\, X_u\, du + M_t. \tag{2.1}$$

Here, M is a $(\sigma\{X_t, 0 \le t\}, P)$-martingale and the matrix $A \in \mathbb{R}^{n\times n}$ is a rate matrix for the process X.

Manuscript received Jun 7, 2004; revised December 9, 2005 and December 1, 2006. Recommended by Associate Editor xxxx. The work of W. P. Malcolm was supported by the National Information and Communication Technology (ICT) Australia under the Grant given by the Australian Government's Department of Communications, Information Technology and the Arts and the Australian Research Council through Backing Australia's Ability and the ICT Centre of Excellence program.

R. J. Elliott was with the Haskayne School of Business, University of Calgary, Calgary, AB T2N 1N4, Canada. He is now with the University of Alberta, Edmonton, AB T6G 2E1, Canada (e-mail: relliott@ucalgary.ca).

W. P. Malcolm is with the Statistical Machine Learning Program of the National Information and Communication Technology (ICT) Australia (NICTA), Canberra, A.C.T. 2601, Australia (e-mail: paul.malcolm@nicta.com.au).

Digital Object Identifier 10.1109/TAC.2007.914305

B. Observation Process Dynamics

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

$$N_t = \int_0^t \langle X_u, \lambda\rangle du + V_t. \tag{2.2}$$

Here, V is a $(\sigma\{N_u, 0 \le u \le t\}, P)$-martingale, $\langle\cdot,\cdot\rangle$ denotes an inner product, and $\lambda \in \mathbb{R}^n_+$ is a vector of n nonnegative Poisson intensities. Our filtrations are given by

$$\mathbb{F}_t = \{\mathcal{F}_t\}, \quad \text{where } \mathcal{F}_t \triangleq \sigma\{X_u; 0 \le u \le t\} \tag{2.3}$$

$$\mathbb{Y}_{0,t} = \{\mathcal{Y}_{0,t}\}, \quad \text{where } \mathcal{Y}_{0,t} \triangleq \sigma\{N_u; 0 \le u \le t\} \tag{2.4}$$

$$\mathbb{G}_{0,t} = \big\{\mathcal{G}_{0,t}\big\}, \quad \text{where } \mathcal{G}_{0,t} \triangleq \sigma\{N_u, X_u; 0 \le u \le t\}. \tag{2.5}$$

C. Reference Probability

We define a probability measure $P^\dagger$ on the space $(\Omega, \mathcal{F})$ such that, under $P^\dagger$, the following two conditions hold.

1) The state process X is a Markov process with intensity matrix A and initial probability distribution p_0.
2) The observation process N is a standard Poisson process, that is, N has a fixed intensity of unity.

The real-world probability measure P is defined by setting

$$\left.\frac{dP}{dP^\dagger}\right|_{\mathcal{G}_{0,t}} = \Lambda_{0,t} \tag{2.6}$$

where

$$\begin{aligned}\Lambda_{0,t} &= \prod_{0<u\le t} \langle X_u, \lambda\rangle^{\Delta N_u} \exp\left\{\int_0^t (1 - \langle X_u, \lambda\rangle)du\right\} \\ &= 1 + \int_0^t \Lambda_{u-}\big(\langle X_u, \lambda\rangle - 1\big)\big(dN_u - du\big).\end{aligned} \tag{2.7}$$

Here,

$$\begin{aligned}\Delta N_\tau &\triangleq N_\tau - \lim_{\epsilon\downarrow 0} N_{\tau-\epsilon} \\ &= N_\tau - N_{\tau-}.\end{aligned} \tag{2.8}$$

Lemma 1: Under the measure P, the dynamics for the Markov process X are unchanged and given by (2.1).

A proof of Lemma 1 is given in the Appendix. Further detail on the theory of Girsanov's theorem and its application to estimation problems for stochastic dynamical systems can be found in the texts [2] and [3].

III. Filter-Based EM Algorithm

A. EM Algorithm

The EM algorithm is a two-step iterative process for computing maximum likelihood (ML) estimates. This process is usually terminated when some imposed measure of convergence for the sequence of maximum likelihood estimators (MLEs) is satisfied. 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. Suppose $\mathcal{Y} \subset \mathcal{F}$.

The two iterative steps in the EM algorithm are as follows.

1) *Expectation step:* Fix $\theta^* = \widehat{\theta}_\tau$, then compute $Q(\cdot, \theta^*)$, where

$$Q(\theta, \theta^*) = E_{\theta^*}\left[\log \frac{dP_\theta}{dP_{\theta^*}} | \mathcal{Y}\right]. \tag{3.9}$$

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

$$\widehat{\theta}_{\tau+1} \in \operatorname*{argmax}_{\theta\in\Theta} Q(\theta, \theta^*). \tag{3.10}$$

B. State Estimation Filters

The so-called filter-based form of the EM algorithm for a continuous-time Markov chain observed in Brownian motion was presented in [4] and a robust version is given in [10]. In this paper, we develop a version of the techniques used in [10] for parameter estimation with MMPPs. This method is based essentially on four quantities, each concerning the indirectly observed Markov process X and each computed by using the information up to and including time t. We now list the four quantities of interest for the filter-based EM algorithm.

1) X_t, the state of the Markov chain. We are interested in $E[X_t|\mathcal{Y}_t]$. By Bayes' Theorem this is $E^\dagger[\Lambda_t X_t|\mathcal{Y}_{0,t}]/E^\dagger[\Lambda_t|\mathcal{Y}_{0,t}]$. Write

$$q_t \triangleq E^\dagger\Big[\Lambda_t X_t|\mathcal{Y}_{0,t}\Big]. \tag{3.11}$$

Then,

$$\begin{aligned}q_t = q_0 &+ \int_0^t A q_u du \\ &+ \int_0^t \operatorname{diag}\Big\{\langle\lambda, \boldsymbol{e}_\ell\rangle - 1\Big\}\Big(dN_u - du\Big) \in \mathbb{R}^n.\end{aligned} \tag{3.12}$$

Here,

$$\operatorname{diag}\Big\{\langle\lambda, \boldsymbol{e}_\ell\rangle - 1\Big\} = \begin{bmatrix}\langle\lambda, \boldsymbol{e}_1\rangle - 1 & & & \\ & \langle\lambda, \boldsymbol{e}_2\rangle - 1 & & \\ & & \ddots & \\ & & & \langle\lambda, \boldsymbol{e}_n\rangle - 1\end{bmatrix} \tag{3.13}$$

and A is the rate matrix for the process X. The unnormalized probability q is converted to its corresponding normalized probability by noting that $\sum_{i=1}^n \langle X_t, \boldsymbol{e}_i\rangle = 1$, so $E^\dagger[\Lambda_t|\mathcal{Y}_t] = \langle q_t, \mathbf{1}\rangle$. Here, $\mathbf{1} = (1, 1, \ldots, 1)' \in \mathbb{R}^n$. Therefore,

$$P(X_t = \boldsymbol{e}_i|\mathcal{Y}_{0,t}) = \frac{\langle q_t, \boldsymbol{e}_i\rangle}{\sum_{\ell=1}^n \langle q_t, \boldsymbol{e}_\ell\rangle}. \tag{3.14}$$

A proof of (3.12) is given in the Appendix.

2) J_t^i, the cumulative sojourn time spent by the process X in state $\boldsymbol{e}_i$ is

$$J_t^i = \int_0^t \langle X_u, \boldsymbol{e}_i\rangle du. \tag{3.15}$$

3) $N_t^{(j,i)}$, the number of transitions $\boldsymbol{e}_i \to \boldsymbol{e}_j$ of X, where $i \neq j$, up to time t is

$$N_t^{(j,i)} = \int_0^t \langle X_{u-}, \boldsymbol{e}_i \rangle \, \langle dX_u, \boldsymbol{e}_j \rangle. \tag{3.16}$$

4) G_t^i, the level integrals for the state $\boldsymbol{e}_i$, is

$$G_t^i = \int_0^t \langle X_u, \boldsymbol{e}_i \rangle dN_u. \tag{3.17}$$

Using Bayes' Theorem, if $H = \{H_t, 0 \leq t\}$ is any $\mathcal{G}$ adapted process

$$E[H_t | \mathcal{Y}_{0,t}] = \frac{E^\dagger[\Lambda_{0,t} H_t | \mathcal{Y}_{0,t}]}{E^\dagger[\Lambda_{0,t} | \mathcal{Y}_{0,t}]}. \tag{3.18}$$

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

$$\sigma(H_t) = E^\dagger[\Lambda_{0,t} H_t | \mathcal{Y}_{0,t}]. \tag{3.19}$$

Indexing the sequence of passes of the EM algorithm by $\tau = 1, 2, 3 \ldots$, the update formulas for the parameter estimates are as follows:

$$[\widehat{A}_{\tau+1}]_{(i,j)} = \frac{E[N_T^{(j,i)} | \mathcal{Y}_{0,T}]}{E[J_T^i | \mathcal{Y}_{0,T}]} = \frac{\sigma(N_T^{i,j})}{\sigma(J_T^i)} \tag{3.20}$$

and

$$\langle \widehat{\lambda}_{\tau+1}, \boldsymbol{e}_i \rangle = \frac{E[G_T^i | \mathcal{Y}_{0,T}]}{E[J_T^i | \mathcal{Y}_{0,T}]} = \frac{\sigma(G_T^i)}{\sigma(J_T^i)}. \tag{3.21}$$

The conditional expectations in equations (3.20) and (3.21) are computed using the *previous* (at pass τ) parameter estimates for A and λ.

The updates for $[\hat{A}_k]_{(i,j)}$ and $\langle \hat{\lambda}_k, \boldsymbol{e}_i \rangle$ are computed by evaluating the expectations in (3.20) and (3.21), respectively. However, it is, in general, not possible to compute recursive dynamics for the processes J^i, $N^{(j,i)}$, and G^i. It is, however, possible to compute dynamics for the associated product quantities $\sigma(J_t^i X_t)$, $\sigma(N_t^{(j,i)} X_t)$, and $\sigma(G_t^i X_t)$, where, for example,

$$\sigma(G_t^i X_t) = E^\dagger[\Lambda_t \, G_t^i \, X_t | \mathcal{Y}_t] \in \mathbb{R}^n. \tag{3.22}$$

The fundamental idea behind the filter-based EM algorithm is to compute recursive filters for quantities such as (3.22), then marginalize the state variable X to evaluate the estimators given by (3.20) and (3.21). We now give recursive filters to estimate, respectively, the product quantities $J^i X$, $N^{(j,i)} X$, and $G^i X$.

Theorem 1: The vector-valued process $\sigma(J^i X) \in \mathbb{R}^n$ satisfies the stochastic integral equation

$$\begin{aligned} \sigma(J_t^i X_t) &= \int_0^t A\sigma(J_u^i X_u) du + \int_0^t \langle q_u, \boldsymbol{e}_i \rangle du \, \boldsymbol{e}_i \\ &\quad + \int_0^t \mathrm{diag}\big\{\langle \lambda, \boldsymbol{e}_\ell \rangle - 1\big\} \sigma(J_{u-}^i X_{u-}) \big(dN_u - du\big). \end{aligned} \tag{3.23}$$

Here, $\sigma(J_0^i X_0) = 0$ and q is the solution of (3.12).

Theoremm 2: The vector-valued process $\sigma(N^{(j,i)} X) \in \mathbb{R}^n$ satisfies the stochastic integral equation

$$\begin{aligned} \sigma(N_t^{(j,i)} X_t) &= \int_0^t A\sigma(N_u^{(j,i)} X_u) du \\ &\quad + \int_0^t \langle q_u, \boldsymbol{e}_i \rangle \langle A\boldsymbol{e}_i, \boldsymbol{e}_j \rangle du \, \boldsymbol{e}_j \\ &\quad + \int_0^t \mathrm{diag}\big\{\langle \lambda, \boldsymbol{e}_\ell \rangle - 1\big\} \\ &\quad \times \sigma(N_{u-}^{(j,i)} X_{u-}) \big(dN_u - du\big). \end{aligned} \tag{3.24}$$

Here, $\sigma(N_0^{(j,i)} X_0) = 0$ and q is the solution of (3.12).

Theorem 3: The vector-valued process $\sigma(G^i X) \in \mathbb{R}^n$ satisfies the stochastic integral equation

$$\begin{aligned} \sigma(G_t^i X_t) &= \int_0^t A\sigma(G_u^i X_u) du \\ &\quad + \int_0^t \langle q_{u-}, \boldsymbol{e}_i \rangle \langle \lambda, \boldsymbol{e}_i \rangle dN_u \boldsymbol{e}_i \\ &\quad + \int_0^t \mathrm{diag}\big\{\langle \lambda, \boldsymbol{e}_\ell \rangle - 1\big\} \sigma(G_{u-}^i X_{u-}) \big(dN_u - du\big). \end{aligned} \tag{3.25}$$

Here, $\sigma(G_0^i X_0) = 0$ and q is the solution of (3.12).

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 (3.23), (3.24), and (3.25), the updates for the parameter estimates are given by

$$[\widehat{A}_{\tau+1}]_{(i,j)} = \frac{\langle \sigma(N_T^{(j,i)} \, X_T), \mathbf{1} \rangle}{\langle \sigma(J_T^i X_T), \mathbf{1} \rangle} \tag{3.26}$$

and

$$\langle \widehat{\lambda}_{\tau+1}, \boldsymbol{e}_i \rangle = \frac{\langle \sigma(G_t^i X_T), \mathbf{1} \rangle}{\langle \sigma(J_T^i \, X_T), \mathbf{1} \rangle}. \tag{3.27}$$

C. Robust State Estimation Filters

Each of the dynamics given by (3.23)–(3.25) contain stochastic Lebesgue–Stieltjes integral terms. These stochastic integrals, with respect to the observation process N, can be eliminated by using a version of a gauge transformation due to Clark [1]. Consider the diagonal matrix

$$\Gamma_t \triangleq \mathrm{diag}\{\gamma_t^i\} \in \mathbb{R}^{n \times n}. \tag{3.28}$$

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

$$\begin{aligned} \Gamma_t^{-1} &= \int_0^t \mathrm{diag}\big\{\langle \lambda, \boldsymbol{e}_\ell \rangle - 1\big\} \Gamma_u^{-1} du \\ &\quad + \int_0^t \Gamma_{u-}^{-1} \mathrm{diag}\big\{\langle \lambda, \boldsymbol{e}_\ell \rangle^{-1} - 1\big\} dN_u. \end{aligned} \tag{3.29}$$

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

$$\overline{q}_t = \overline{q}_0 + \int_0^t \Gamma_u^{-1} A \Gamma_u \overline{q}_u \, du. \tag{3.30}$$

Equation (3.30) was established in [13]. For any $\mathcal{F}$-adapted integrable process H, we write

$$\overline{\sigma}(H) = \Gamma_t^{-1} \sigma(H_t). \tag{3.31}$$

Now, our objective is to compute filters to estimate the product processes $G^i X$, $N^{(j,i)} X$, and $J^i X$. Let us first consider the process $\sigma(G^i X)$. Dynamics for the gauge transformed process $\overline{\sigma}(G_t^i X_t) = \Gamma_t^{-1} \sigma(G_t^i X_t)$ can be computed by applying the product rule

$$\begin{aligned} d(\Gamma_t^{-1} \sigma(G_t^i X_t)) = \Gamma_{t-}^{-1} d(\sigma(G_t^i X_t)) + d\Gamma_t^{-1} \sigma(G_{t-}^i X_{t-}) \\ + \Delta \Gamma_t^{-1} \Delta \sigma(G_t^i X_t). \end{aligned} \tag{3.32}$$

The result of this calculation is

$$\begin{aligned} \Gamma_t^{-1} \sigma(G_t^i X_t) = & \int_0^t \Gamma_u^{-1} A \Gamma_u \overline{\sigma}(G_u^i X_u) du \\ & + \int_0^t \langle \overline{q}_u, \boldsymbol{e}_i \rangle \langle \lambda, \boldsymbol{e}_i \rangle dN_u \boldsymbol{e}_i \\ & + \int_0^t \operatorname{diag}\{\langle \lambda, \boldsymbol{e}_\ell \rangle - 1\} \overline{\sigma}(G_{u-}^i X_{u-})(dN_u - du) \\ & + \int_0^t \operatorname{diag}\{\langle \lambda, \boldsymbol{e}_\ell \rangle - 1\} \overline{\sigma}(G_u^i X_u) du \\ & + \int_0^t \operatorname{diag}\{\langle \lambda, \boldsymbol{e}_\ell \rangle^{-1} - 1\} \overline{\sigma}(G_{u-}^i X_{u-}) dN_u \\ & + \int_0^t \operatorname{diag}\{\langle \lambda, \boldsymbol{e}_\ell \rangle^{-1} - 1\} \langle \overline{q}_u, \boldsymbol{e}_i \rangle \langle \lambda, \boldsymbol{e}_i \rangle dN_u \, \boldsymbol{e}_i \\ & + \int_0^t \operatorname{diag}\{\langle \lambda, \boldsymbol{e}_\ell \rangle^{-1} - 1\} \\ & \times \operatorname{diag}\{\langle \lambda, \boldsymbol{e}_\ell \rangle^{-1} - 1\} \overline{\sigma}(G_{u-}^i X_{u-}) \, dN_u. \end{aligned} \tag{3.33}$$

Several stochastic integrals in (3.33) cancel, noting

$$\begin{aligned} & \operatorname{diag}\{\langle \lambda, \boldsymbol{e}_\ell \rangle - 1\} + \operatorname{diag}\{\langle \lambda, \boldsymbol{e}_\ell \rangle^{-1} - 1\} \\ & + \operatorname{diag}\{\langle \lambda, \boldsymbol{e}_\ell \rangle^{-1} - 1\} \operatorname{diag}\{\langle \lambda, \boldsymbol{e}_\ell \rangle - 1\} = 0 \in \mathbb{R}^{n \times n} \end{aligned} \tag{3.34}$$

giving

$$\overline{\sigma}(G_t^i X_t) = \int_0^t \Gamma_u^{-1} A \Gamma_u \overline{\sigma}(G_u^i X_u) du + \int_0^t \langle \overline{q}_u, \boldsymbol{e}_i \rangle dN_u \boldsymbol{e}_i. \tag{3.35}$$

The stochastic integral in (3.35) can be simplified by stochastic integration by parts

$$\int_0^t \langle \overline{q}_u, \boldsymbol{e}_i \rangle dN_u \boldsymbol{e}_i = \langle \overline{q}_t, \boldsymbol{e}_i \rangle N_t - \int_0^t N_u \langle d\overline{q}_u, \boldsymbol{e}_i \rangle. \tag{3.36}$$

Finally, our dynamics for $\overline{\sigma}(G_t^i X_t)$ read

$$\begin{aligned} \overline{\sigma}(G_t^i X_t) = & \int_0^t \Gamma_u^{-1} A \Gamma_u \overline{\sigma}(G_u^i X_u) du \\ & + \langle \overline{q}_t, \boldsymbol{e}_i \rangle N_t \boldsymbol{e}_i - \int_0^t N_u \langle d\overline{q}_u, \boldsymbol{e}_i \rangle \boldsymbol{e}_i. \end{aligned} \tag{3.37}$$

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

$$\overline{\sigma}(J_t^i X_t) = \int_0^t \Gamma_u^{-1} A \Gamma_u \overline{\sigma}(J_u^i X_u) du + \int_0^t \langle \overline{q}_u, \boldsymbol{e}_i \rangle du \, \boldsymbol{e}_i \tag{3.38}$$

and

$$\begin{aligned} \overline{\sigma}(N_t^{(j,i)} X_t) = & \int_0^t \Gamma_u^{-1} A \Gamma_u \overline{\sigma}(N_u^{(j,i)} X_u) du \\ & + \int_0^t \langle \overline{q}_u, \boldsymbol{e}_i \rangle \langle A \boldsymbol{e}_i, \boldsymbol{e}_j \rangle du \, \boldsymbol{e}_j. \end{aligned} \tag{3.39}$$

D. Discrete-Time Filters

For all time discretizations, we will consider a partition on an interval $[0, T]$ and write

$$\Pi_{[0,T]}^{(K)} \triangleq \{0 = t_0, t_1, \ldots, t_K = T\}. \tag{3.40}$$

Here, the partition is strict, that is, $t_0 < t_1 < \cdots < t_K = T$. To denote the mesh of the partition, we write

$$\|\Pi_{[0,T]}^{(K)}\| = \max_{1 \le k \le K} \{t_k - t_{k-1}\}. \tag{3.41}$$

For brevity, we shall use the notation $\xi_k \triangleq \xi_{t_k}$, where ξ_k denotes a process ξ at a time point t_k. Further, we write $\Delta_{(k-1,k)} = t_k - t_{k-1}$. Approximating the integral in (3.30), we get

$$\overline{q}_{t_k} \approx \overline{q}_{t_{k-1}} + \Gamma_{t_{k-1}}^{-1} A \Gamma_{t_{k-1}} \overline{q}_{t_{k-1}} \, \Delta_{(k-1,k)} \tag{3.42}$$

so

$$q_{t_k} = \Gamma_{t_k} \overline{q}_{t_k} \approx \Gamma_{t_k} \Gamma_{t_{k-1}}^{-1} \left[\mathbf{I} + \Delta_{(k-1,k)} A\right] q_{t_{k-1}}. \tag{3.43}$$

This suggests the recursion

$$\widehat{q}_k \triangleq \Gamma_k \Gamma_{k-1}^{-1} \left[\mathbf{I} + \Delta_{(k-1,k)} A\right] \widehat{q}_{k-1}. \tag{3.44}$$

Here, $\widehat{q}$ denotes an estimate of the unnormalized probability generated by the suboptimal discrete-time recursion at (3.43).

Remark 1: An important feature of the filter formulation at (3.44) is that the sampling interval or $\Delta_{(k-1,k)}$ can be chosen to ensure a certain type of numerical stability. Here, numerical stability is taken to mean $\langle q, \boldsymbol{e}_i \rangle \ge 0$ for all $i \in \{1, 2, \ldots, n\}$. The details of this property are given in [16].

Writing the dynamics given by (3.37) recursively at sampling instants t_k and t_{k-1}, we get

$$\begin{aligned} \overline{\sigma}(G_{t_k}^i X_{t_k}) = & \, \overline{\sigma}(G_{t_{k-1}}^i X_{t_{k-1}}) \\ & + \int_{t_{k-1}}^{t_k} \Gamma_u^{-1} A \Gamma_u \overline{\sigma}(G_u^i X_u) du \end{aligned}$$

$$+ \langle \overline{q}_{t_k}, \boldsymbol{e}_i \rangle N_{t_k} \boldsymbol{e}_i - \langle \overline{q}_{t_{k-1}}, \boldsymbol{e}_i \rangle N_{t_{k-1}} \boldsymbol{e}_i - \int_{t_{k-1}}^{t_k} N_u \langle d\overline{q}_u, \boldsymbol{e}_i \rangle \boldsymbol{e}_i. \tag{3.45}$$

Making an Euler–Maruyama[1] approximation, we have

$$\int_{t_{k-1}}^{t_k} N_u \langle d\overline{q}_u, \boldsymbol{e}_i \rangle \boldsymbol{e}_i = \int_{t_{k-1}}^{t_k} N_u \langle \Gamma_u^{-1} A \Gamma_u \overline{q}_u du, \boldsymbol{e}_i \rangle \boldsymbol{e}_i \approx N_{t_{k-1}} \Gamma_{t_{k-1}}^{-1} \langle A q_{t_{k-1}} \boldsymbol{e}_i \rangle \Delta_{(k-1,k)} \boldsymbol{e}_i \tag{3.46}$$

and with some algebraic manipulation

$$\begin{aligned} \langle \overline{q}_{t_k}, \boldsymbol{e}_i \rangle N_{t_k} \boldsymbol{e}_i - \langle \overline{q}_{t_{k-1}}, \boldsymbol{e}_i \rangle N_{t_{k-1}} \boldsymbol{e}_i &= \Gamma_{t_{k-1}}^{-1} \langle q_{t_{k-1}}, \boldsymbol{e}_i \rangle (N_{t_k} - N_{t_{k-1}}) \boldsymbol{e}_i \\ &\quad + \Delta_{(k-1,k)} \Gamma_{t_{k-1}}^{-1} \langle A q_{t_{k-1}}, \boldsymbol{e}_i \rangle N_{t_k} \boldsymbol{e}_i \end{aligned} \tag{3.47}$$

we see that

$$\begin{aligned} \overline{\sigma}(G_k^i X_k) \approx \overline{\sigma}(G_{k-1}^i X_{k-1}) &+ \Gamma_{k-1}^{-1} A \Gamma_{k-1} \overline{\sigma}(G_{k-1}^i X_{k-1}) \Delta_{(k-1,k)} \\ &+ \Gamma_{k-1}^{-1} \langle \widehat{q}_{k-1}, \boldsymbol{e}_i \rangle (N_k - N_{k-1}) \boldsymbol{e}_i \\ &+ \Delta_{(k-1,k)} \Gamma_{k-1}^{-1} \langle A \widehat{q}_{k-1}, \boldsymbol{e}_i \rangle N_k \boldsymbol{e}_i \\ &- N_{k-1} \Gamma_{k-1}^{-1} \langle A \widehat{q}_{k-1}, \boldsymbol{e}_i \rangle \Delta_{(k-1,k)} \boldsymbol{e}_i. \end{aligned} \tag{3.48}$$

Now, by multiplying both sides of (3.48) on the left-hand side by the matrix Γ_k, we get

$$\begin{aligned} \sigma(G_k^i X_k) \approx \Gamma_k \Gamma_{k-1}^{-1} \sigma(G_{k-1}^i X_{k-1}) &+ \Gamma_k \Gamma_{k-1}^{-1} A \sigma(G_{k-1}^i X_{k-1}) \Delta_{(k-1,k)} \\ &+ \Gamma_k \Gamma_{k-1}^{-1} \langle \widehat{q}_{k-1}, \boldsymbol{e}_i \rangle (N_k - N_{k-1}) \boldsymbol{e}_i \\ &+ \Gamma_k \Gamma_{k-1}^{-1} \langle A \widehat{q}_{k-1}, \boldsymbol{e}_i \rangle (N_k - N_{k-1}) \boldsymbol{e}_i. \end{aligned} \tag{3.49}$$

Our estimator of the quantity $\sigma(G_k^i X_k)$ has dynamics

$$\begin{aligned} \widehat{\sigma}(G_k^i X_k) \triangleq\, & \Gamma_k \Gamma_{k-1}^{-1} \left[\mathbf{I} + \Delta_{(k-1,k)} A \right] \widehat{\sigma}(G_{k-1}^i X_{k-1}) \\ &+ \Gamma_k \Gamma_{k-1}^{-1} \left[\langle \widehat{q}_{k-1}, \boldsymbol{e}_i \rangle + \Delta_{(k-1,k)} \langle A \widehat{q}_{k-1}, \boldsymbol{e}_i \rangle \right] \\ &\times (N_k - N_{k-1}) \boldsymbol{e}_i. \end{aligned} \tag{3.50}$$

After similar calculations, the remaining discretized filters read

$$\begin{aligned} \widehat{\sigma}(N_k^{(j,i)} X_k) = \, & \Gamma_k \Gamma_{k-1}^{-1} \left[\mathbf{I} + \Delta_{(k-1,k)} A \right] \widehat{\sigma}(N_{k-1}^{(j,i)} X_{k-1}) \\ &+ \Gamma_k \Gamma_{k-1}^{-1} \langle \widehat{q}_{k-1}, \boldsymbol{e}_i \rangle \langle A \boldsymbol{e}_i, \boldsymbol{e}_j \rangle \Delta_{(k-1,k)} \boldsymbol{e}_i \end{aligned} \tag{3.15}$$

$$\begin{aligned} \widehat{\sigma}(J_k^i X_k) = \, & \Gamma_k \Gamma_{k-1}^{-1} \left[\mathbf{I} + \Delta_{(k-1,k)} A \right] \widehat{\sigma}(J_{k-1}^i X_{k-1}) \\ &+ \Gamma_k \Gamma_{k-1}^{-1} \langle \widehat{q}_{k-1}, \boldsymbol{e}_i \rangle \boldsymbol{e}_i. \end{aligned} \tag{3.52}$$

[1]We will use this particular approximation throughout this paper, however, other approximations could also be used.

E. Discrete-Time Filter-Based EM Algorithm

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

Initialization	$\forall (i,j) \in \{(1,1),(1,2),\ldots,(n,n)\}$, Choose $[\widehat{A}_0]_{(i,j)}$, for each $i \in \{1,2,\ldots,n\}$ choose $\langle \widehat{\lambda}, \boldsymbol{e}_i \rangle$.
Step 1	Using (3.26) and (3.27), compute the MLEs, $[\widehat{A}_{\tau+1}]_{i,j}$ and $\widehat{\lambda}_{\tau+1}$.
Step 2	Decide to stop or continue from step 2.

IV. Smoother-Based EM Algorithm for MMPPs

In many implementations of the EM algorithm, for example, [24] and [29], the expectation step is completed with smoothed rather than (online) filtered estimates. Typically, the smoothing scheme used is the so-called "fixed interval smoother." Computing smoothing schemes for MMPPs can be particularly difficult [23], [26]. One source of this difficulty is the task of developing backwards dynamics. This task usually leads to constructing stochastic integrals evolving backward in time. However, the approach we use to develop smoothing algorithms completely avoids these difficulties. To compute our smoothers we exploit a special identity between forward and backward robust dynamics, and as a consequence, do not need to consider the backward stochastic integration at all.

A. Smoothed State Estimation for the Process X

We first briefly recall the state estimation MMPP smoother presented in [14]. For a smoothed estimate for the process $X \in \mathbb{R}^n$, we wish to evaluate the expectation $E[X_t | \mathcal{Y}_{0,T}]$, where $0 \le t \le T$. By Bayes' rule [3], we have

$$E[X_t | \mathcal{Y}_{0,T}] = \frac{E^\dagger[\Lambda_{0,T} X_t | \mathcal{Y}_{0,T}]}{E^\dagger[\Lambda_{0,T} | \mathcal{Y}_{0,T}]}. \tag{4.53}$$

Consider the numerator of (4.53)

$$\begin{aligned} r_t &\triangleq E^\dagger[\Lambda_{0,T} X_t | \mathcal{Y}_{0,T}] \\ &= E^\dagger[\Lambda_{0,t} \Lambda_{t,T} X_t | \mathcal{Y}_{0,T}] \\ &= E^\dagger[E^\dagger[\Lambda_{0,t} \Lambda_{t,T} X_t | \mathcal{Y}_{0,T} \vee \mathcal{F}_t] | \mathcal{Y}_{0,T}] \\ &= E^\dagger[\Lambda_{0,t} X_t E^\dagger[\Lambda_{t,T} | \mathcal{Y}_{0,T} \vee \mathcal{F}_t | \mathcal{Y}_{0,T}]. \end{aligned} \tag{4.54}$$

Under the measure $P^\dagger$, X is a Markov process, so the inner expectation in the previous line of (4.54) is

$$E^\dagger[\Lambda_{t,T} | \mathcal{Y}_{0,T} \vee \mathcal{F}_t] = E^\dagger[\Lambda_{t,T} | \mathcal{Y}_{0,T} \vee \sigma\{X_t\}]. \tag{4.55}$$

Write

$$v_{t,T}^i \triangleq E^\dagger[\Lambda_{t,T} | \mathcal{Y}_{0,T} \text{ and } X_t = \boldsymbol{e}_i]. \tag{4.56}$$

Omitting further calculations, it can be shown [14] that

$$\begin{aligned} r_t = \langle q_t, \boldsymbol{e}_1 \rangle \langle v_{t,T}, \boldsymbol{e}_1 \rangle \, \boldsymbol{e}_1 + \langle q_t, \boldsymbol{e}_2 \rangle \langle v_{t,T}, \boldsymbol{e}_2 \rangle \, \boldsymbol{e}_2 + \cdots \\ + \langle q_t, \boldsymbol{e}_m \rangle \langle v_{t,T}, \boldsymbol{e}_n \rangle \, \boldsymbol{e}_n \in \mathbb{R}^n. \end{aligned} \tag{4.57}$$

The normalized smoothed-state estimate of X is then

$$E[X_t | \mathcal{Y}_{0,t}] = \frac{r_t}{\langle r_t, \mathbf{1} \rangle}. \tag{4.58}$$

Note that

$$\begin{aligned}\langle r_t, \mathbf{1}\rangle &= \langle q_t, v_{t,T}\rangle \\ &= E^{\dagger}[\Lambda_{0,T}\langle X_t, \mathbf{1}\rangle | \mathcal{Y}_{0,T}] \\ &= E^{\dagger}[\Lambda_{0,T}|\mathcal{Y}_{0,T}] \end{aligned} \tag{4.59}$$

is independent of t. Therefore

$$\begin{aligned}\frac{d}{dt}\langle r_t, \mathbf{1}\rangle &= \frac{d}{dt}\langle q_t, v_{t,T}\rangle \\ &= \langle d\overline{q}_t, \overline{v}_{t,T}\rangle + \langle \overline{q}_t, d\overline{v}_{t,T}\rangle \\ &= \langle \Gamma_t^{-1} A \Gamma_t \overline{q}_t, \overline{v}_{t,T}\rangle + \langle \overline{q}_t, d\overline{v}_{t,T}\rangle \\ &= 0. \end{aligned} \tag{4.60}$$

The vector $v_{t,T} = \big(\langle v_{t,T}, \boldsymbol{e}_1\rangle, \langle v_{t,T}, \boldsymbol{e}_2\rangle, \ldots, \langle v_{t,T}, \boldsymbol{e}_n\rangle\big)$ incorporates the *extra* information obtained from the observations between t and T. Computing dynamics for v can be difficult [18], [19]. However, by exploiting a special identity between the forward dynamics and the corresponding backward, process $\overline{v}$, one can directly compute robust dynamics for the process v. What we must do is consider the process $\overline{v}$, such that the following identity holds

$$\langle \overline{q}_t, \overline{v}_{t,T}\rangle = \langle \Gamma_t^{-1} q_t, \Gamma_t v_{t,T}\rangle = \langle q_t, v_{t,T}\rangle, \quad \text{for all } t \in [0, T]. \tag{4.61}$$

That is, $\overline{v}_{t,T} \triangleq \Gamma_t v_{t,T}$. Using (4.60), one can show that

$$\frac{d\overline{v}_{t,T}}{dt} = -\Gamma_t A' \Gamma_t^{-1} \overline{v}_{t,T}, \quad \overline{v}_{T,T} = v_{T,T} = \mathbf{1} \tag{4.62}$$

so that

$$\overline{v}_{t,T} = \mathbf{1} + \int_t^T \Gamma_u A' \Gamma_u^{-1} \overline{v}_{t,T}\, du. \tag{4.63}$$

Further, using the time discretization of (3.40)

$$\begin{aligned}\overline{v}_{k-1,T} &= \overline{v}_{k,T} + \int_{t_{k-1}}^{t_k} \Gamma_u A' \Gamma_u^{-1} \overline{v}_{u,T}\, du \\ &\approx \overline{v}_{k,T} + \Gamma_k A' \Gamma_k^{-1} \overline{v}_{k,T} \Delta_{(k-1,k)}, \end{aligned} \tag{4.64}$$

so, our suboptimal estimator $\widehat{v} \approx v$ has dynamics

$$\widehat{v}_{k-1,T} \triangleq \Gamma_{k-1}^{-1} \Gamma_k \left[\boldsymbol{I} + \Delta_{(k-1,k)} A'\right] \widehat{v}_{k,T}. \tag{4.65}$$

B. Smoothers for the Quantities N_t^i, J_t^i, and G_t^i

Following the same strategy as before, we consider the identity

$$\begin{aligned}\langle \sigma(G_t^i X_t), v_{t,T}\rangle &= \langle \Gamma_t^{-1} \sigma(G_t^i X_t), \Gamma_t v_{t,T}\rangle \\ &= \langle \overline{\sigma}(G_t^i X_t), \overline{v}_{t,T}\rangle. \end{aligned} \tag{4.66}$$

Now, define

$$\widetilde{\sigma}(G_t^i X_t) \triangleq \overline{\sigma}(G_t^i X_t) - \langle \overline{q}_t, \boldsymbol{e}_i\rangle N_t \boldsymbol{e}_i. \tag{4.67}$$

Then

$$\begin{aligned}d\,\widetilde{\sigma}(G_t^i X_t) &= \Gamma_t^{-1} A \Gamma_t \overline{\sigma}(G_t^i X_t)\, dt \\ &\quad - N_t \langle \Gamma_t^{-1} A \Gamma_t \overline{q}_t, \boldsymbol{e}_i\rangle \boldsymbol{e}_i\, dt. \end{aligned} \tag{4.68}$$

Now,

$$\langle \overline{\sigma}(G_t^i X_t), \overline{v}_{t,T}\rangle = \langle \widetilde{\sigma}(G_t^i X_t), \overline{v}_{t,T}\rangle + N_t \langle \overline{q}_t, \boldsymbol{e}_i\rangle \langle \overline{v}_{t,T}, \boldsymbol{e}_i\rangle \tag{4.69}$$

and

$$N_t \langle \overline{q}_t, \boldsymbol{e}_i\rangle \langle \overline{v}_{t,T}, \boldsymbol{e}_i\rangle = N_t \langle q_t, \boldsymbol{e}_i\rangle \langle v_{t,T}, \boldsymbol{e}_i\rangle. \tag{4.70}$$

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

$$\begin{aligned}&d\langle \widetilde{\sigma}(G_t^i X_t), \overline{v}_{t,T}\rangle \\ &\quad = \langle \Gamma_t^{-1} A \Gamma_t \overline{\sigma}(G_t^i X_t), \overline{v}_{t,T}\rangle\, dt \\ &\qquad - N_t \langle \Gamma_t^{-1} A \Gamma_t \overline{q}_t, \boldsymbol{e}_i\rangle \langle \boldsymbol{e}_i, \overline{v}_{t,T}\rangle\, dt \\ &\qquad - \langle \widetilde{\sigma}(G_t^i X_t), \Gamma_t A \Gamma_t^{-1} \overline{v}_{t,T}\rangle\, dt \\ &\quad = \langle \Gamma_t^{-1} A \Gamma_t \overline{\sigma}(G_t^i X_t), \overline{v}_{t,T}\rangle\, dt \\ &\qquad - N_t \langle \Gamma_t^{-1} A \Gamma_t \overline{q}_t, \boldsymbol{e}_i\rangle \langle \boldsymbol{e}_i, \overline{v}_{t,T}\rangle\, dt \\ &\qquad - \langle \overline{\sigma}(G_t^i X_t) - \langle \overline{q}_t, \boldsymbol{e}_i\rangle N_t \boldsymbol{e}_i, \Gamma_t A' \Gamma_t^{-1} \overline{v}_{t,T}\rangle\, dt \\ &\quad = -N_t \langle \Gamma_t^{-1} A \Gamma_t \overline{q}_t, \boldsymbol{e}_i\rangle \langle \boldsymbol{e}_i, \overline{v}_{t,T}\rangle\, dt \\ &\qquad + N_t \langle \overline{q}_t, \boldsymbol{e}_i\rangle \langle \Gamma_t^{-1} A \Gamma_t \boldsymbol{e}_i, \overline{v}_{t,T}\rangle\, dt \end{aligned} \tag{4.71}$$

i.e.,

$$\begin{aligned}&\langle \widetilde{\sigma}(G_T^i X_T), \overline{v}_{T,T}\rangle \\ &\quad = -\int_0^T N_u \langle \Gamma_u^{-1} A \Gamma_u \overline{q}_u, \boldsymbol{e}_i\rangle \langle \boldsymbol{e}_i, \overline{v}_{u,T}\rangle du \\ &\qquad + \int_0^T N_u \langle \overline{q}_u, \boldsymbol{e}_i\rangle \langle \Gamma_u^{-1} A \Gamma_u \boldsymbol{e}_i, \overline{v}_{u,T}\rangle du. \end{aligned} \tag{4.72}$$

Therefore,

$$\begin{aligned}\langle \sigma(G_T^i X_T), v_{T,T}\rangle &= \langle \overline{\sigma}(G_T^i X_T), \overline{v}_T\rangle \\ &= \langle \widetilde{\sigma}(G_T^i X_T), \overline{v}_T\rangle + N_T \langle \overline{q}_T, \boldsymbol{e}_i\rangle \langle \overline{v}_T, \boldsymbol{e}_i\rangle \\ &= -\int_0^T N_u \langle \Gamma_u^{-1} A \Gamma_u \overline{q}_u, \boldsymbol{e}_i\rangle \langle \boldsymbol{e}_i, \overline{v}_{u,T}\rangle du \\ &\quad + \int_0^T N_u \langle \overline{q}_u, \boldsymbol{e}_i\rangle \langle \Gamma_u^{-1} A \Gamma_u \boldsymbol{e}_i, \overline{v}_{u,T}\rangle du \\ &\quad + N_T \langle q_T, \boldsymbol{e}_i\rangle \langle v_{T,T}, \boldsymbol{e}_i\rangle \\ &= -\int_0^T N_u \langle A q_u, \boldsymbol{e}_i\rangle \langle \boldsymbol{e}_i, v_{u,T}\rangle du \\ &\quad + \int_0^T N_u \langle q_u, \boldsymbol{e}_i\rangle \langle \boldsymbol{e}_i, A' v_{u,T}\rangle du \\ &\quad + N_T \langle q_T, \boldsymbol{e}_i\rangle \langle v_{T,T}, \boldsymbol{e}_i\rangle. \end{aligned} \tag{4.73}$$

By using similar calculations, one can also show that

$$\langle \overline{\sigma}(J_T^i X_T), \overline{v}_{T,T}\rangle = \int_0^T \langle q_u, \boldsymbol{e}_i\rangle \langle v_{u,T}, \boldsymbol{e}_i\rangle dt \tag{4.74}$$

and

$$\langle \overline{\sigma}(N_T^{(j,i)} X_T), \overline{v}_{T,T}\rangle = \int_0^T \langle A\,\boldsymbol{e}_i, \boldsymbol{e}_j\rangle \langle q_u, \boldsymbol{e}_i\rangle \langle v_{u,T}, \boldsymbol{e}_j\rangle du. \tag{4.75}$$

C. Smoother-Based EM Algorithm

Recalling (3.20) and (3.21), our smoother-based update equations are

$$[\widehat{A}_{\tau+1}]_{(i,j)} = [\widehat{A}_{\tau}]_{(i,j)} \frac{\int_0^T \langle q_u, \boldsymbol{e}_i \rangle \langle v_{u,T}, \boldsymbol{e}_j \rangle du}{\int_0^T \langle q_u, \boldsymbol{e}_i \rangle \langle v_{u,T}, \boldsymbol{e}_i \rangle du} \tag{4.76}$$

and

$$\begin{aligned} \langle \widehat{\lambda}_{\tau+1}, \boldsymbol{e}_i \rangle = & \frac{\int_0^T N_u \langle q_u, \boldsymbol{e}_i \rangle \langle \boldsymbol{e}_i, A' v_{u,T} \rangle du}{\int_0^T \langle q_u, \boldsymbol{e}_i \rangle \langle v_{u,T}, \boldsymbol{e}_i \rangle du} \\ & - \frac{\int_0^T N_u \langle A q_u, \boldsymbol{e}_i \rangle \langle \boldsymbol{e}_i, v_{u,T} \rangle du}{\int_0^T \langle q_u, \boldsymbol{e}_i \rangle \langle v_{u,T}, \boldsymbol{e}_i \rangle du} \\ & + \frac{N_T \langle q_T, \boldsymbol{e}_i \rangle \langle v_{T,T}, \boldsymbol{e}_i \rangle}{\int_0^T \langle q_u, \boldsymbol{e}_i \rangle \langle v_{u,T}, \boldsymbol{e}_i \rangle du}. \end{aligned} \tag{4.77}$$

V. Discrete-Time Smoothers

A. Discrete-Time Smoother Formulas

Suppose that one observes data on the set $[0, T]$ and parameter estimates are computed by using these data. Further, suppose one receives a subsequent observation data on the set $[T, T']$, where $T' > T$. What we would like to do is incorporate the new data on $[T, T']$ so as to reestimate the model parameters, but without complete recalculation from the origin. To utilize the information on $[T, T']$, we consider a time discretization on the total interval $[0, T] \cup [T, T']$, that is,

$$0 = t_0 < t_1 \cdots < t_K = T < t'_0 < t'_1 \cdots < t_{\widetilde{K}} = T'.$$

Here, we denote this augmented partition by $\Pi^{(K)}_{[0,T]} \cup \Pi^{(\widetilde{K})}_{[T,T']}$, where $\widetilde{K} \in \mathbb{N}$ and

$$\Pi^{(\widetilde{K})}_{[T,T']} \triangleq \{T = t'_0, t'_1, \ldots, t'_{\widetilde{K}} = T'\}. \tag{5.78}$$

Recalling the discrete-time, (backward) recursion for the estimator $\widehat{v}$, we see

$$\begin{aligned} \widehat{v}_{k-1,T} &= \Gamma^{-1}_{k-1} \Gamma_k \left[\boldsymbol{I} + \Delta_{(k-1,k)} A'\right] \Gamma^{-1}_{k-1} \Gamma_k \widehat{v}_{k,T} \\ &= \Gamma^{-1}_{k-1} \Gamma_k \left[\boldsymbol{I} + \Delta_{(k-1,k)} A'\right] \Gamma^{-1}_{k} \Gamma_{k+1} \left[\boldsymbol{I} + \Delta_{(k,k+1)} A'\right] \\ &\qquad , \ldots, \Gamma^{-1}_{K-1} \Gamma_K \left[\boldsymbol{I} + \Delta_{(k-1,k)} A'\right] \widehat{v}_{T,T}. \end{aligned} \tag{5.79}$$

Recall here that $K = t_K = T$.

Write

$$\begin{aligned} \Psi_{k-1,T} \triangleq & \Gamma^{-1}_{k-1} \Gamma_k \left[\boldsymbol{I} + \Delta_{(k-1,k)} A'\right] \Gamma^{-1}_{k} \Gamma_{k+1} \\ & \times \left[\boldsymbol{I} + \Delta_{(k,k+1)} A'\right], \ldots \Gamma^{-1}_{K-1} \Gamma_K \left[\boldsymbol{I} + \Delta_{(k-1,k)} A'\right] \\ & \in \mathbb{R}^{n \times n}. \end{aligned} \tag{5.80}$$

Further, for two epochs T and T', where $T < T'$, it follows that

$$\Psi_{k-1,T'} = \Psi_{k-1,T} \Psi_{T,T'}, \quad k \in \{0, 1, 2, \ldots, T\}. \tag{5.81}$$

At a boundary T', we set $\Psi_{T',T'} \triangleq \mathrm{diag}\{1, 1, \ldots, 1\} \in \mathbb{R}^{n \times n}$.

Remarks 2: The transitivity property for Ψ shown by equation (5.82) is critical in our development of data-recursive smoother update formulas.

Using the matrix Ψ, the backward recursion for $v_{k-1,T}$ may be written in the following compact form:

$$v_{k-1,T} = \Psi_{k-1,T} \mathbf{1}, \quad \mathbf{1} = (1, 1, \ldots, 1)' \in \mathbb{R}^n. \tag{5.82}$$

Equation (5.82) and the transitivity property of Ψ can be exploited to compute a data-recursive smoother, that is, a smoother that does not require complete recalculation from the origin upon the arrival of new observation data. Since Ψ is an $n \times n$ matrix, it can be easily stored in memory. It is immediate from the dynamics at (5.82) that the boundary T, upon which $v_{k-1,T}$ depends, is only "fixed" by the action of multiplication on the right-hand side by the vector $\mathbf{1}$. To extend this boundary upon the arrival of subsequent data, the $n \times n$ matrix $\Psi_{k-1,T}$ can be recalled from memory and the updated quantity $v_{k-1,T'}$ is calculated by the recursion

$$v_{k-1,T'} = \Psi_{k-1,T} \Psi_{T,T'} \mathbf{1}, \quad \mathbf{1} = (1, 1, \ldots, 1)' \in \mathbb{R}^n. \tag{5.83}$$

Consider, for example, the following smoothing problem. Suppose one first observes data on $[0, T]$ and computes the smoothed estimate $P(X_t = \boldsymbol{e}_i | \mathcal{Y}_{0,T})$ for some $t \in [0, T]$. Using Ψ, this estimation can be written as

$$\begin{aligned} P(X_t = \boldsymbol{e}_i | \mathcal{Y}_{0,T}) &= \frac{\langle q_t, \boldsymbol{e}_i \rangle \langle v_{t,T}, \boldsymbol{e}_i \rangle}{\sum_{\ell=1}^n \langle q_t, \boldsymbol{e}_\ell \rangle \langle v_{t,T}, \boldsymbol{e}_\ell \rangle} \\ &= \frac{\langle q_t, \boldsymbol{e}_i \rangle \left(\boldsymbol{e}'_\ell \Psi_{t,T}\right) \mathbf{1}}{\left\{\sum_{\ell=1}^n \langle q_t, \boldsymbol{e}_\ell \rangle \left(\boldsymbol{e}'_\ell \Psi_{t,T}\right)\right\} \mathbf{1}}. \end{aligned} \tag{5.84}$$

Now, suppose subsequent data are received on $[T, T']$ and we wish to compute $P(X_t = \boldsymbol{e}_i | \mathcal{Y}_{0,T'})$. Using $\Psi_{T,T'}$, this estimate may be computed by

$$\begin{aligned} P(X_t = \boldsymbol{e}_i | \mathcal{Y}_{0,T'}) &= \frac{\langle q_t, \boldsymbol{e}_i \rangle \langle v_{t,T'}, \boldsymbol{e}_i \rangle}{\sum_{\ell=1}^n \langle q_t, \boldsymbol{e}_\ell \rangle \langle v_{t,T'}, \boldsymbol{e}_\ell \rangle} \\ &= \frac{\langle q_t, \boldsymbol{e}_i \rangle \left(\boldsymbol{e}'_\ell \Psi_{t,T}\right) \Psi_T, T' \mathbf{1}}{\left\{\sum_{\ell=1}^n \langle q_t, \boldsymbol{e}_\ell \rangle \left(\boldsymbol{e}'_\ell \Psi_{t,T}\right)\right\} \Psi_{T,T'} \mathbf{1}}. \end{aligned} \tag{5.85}$$

Equation (5.85) shows that the smoother probability can be computed without the recalculation of v from the origin, provided the $n \times n$ matrix $\Psi_{t,T}$ has been stored in memory.

B. Discrete-Time Smoother-Based EM Algorithm

To compute discrete-time approximations of update formulas (4.76) and (4.77), we approximate the integrals in these estimators by the Trapezoidal rule. These approximations can also be written in a data-recursive form. To approximate $[\widehat{A}_{\tau+1}]_{(i,j)}$ on the interval $[0, T]$, we write (5.86), as shown at the bottom the next page.

Similarly, (5.87) as shown at the bottom of the next page.

Consider again the scenario of new observation data and the two time intervals $[0, T]$ and $[T, T']$. For brevity, we write the normalization constant as

$$\begin{aligned}
M_{T'} &\triangleq \int_0^{T'} \langle q_u, \boldsymbol{e}_i\rangle\langle v_{u,T}, \boldsymbol{e}_j\rangle du \\
&\approx \Bigg\{ \sum_{\ell=1}^{K} \frac{1}{2}\Delta_{(\ell-1,\ell)}\big(\langle \widehat{q}_{\ell-1}, \boldsymbol{e}_i\rangle\big(\boldsymbol{e}_i'\Psi_{\ell-1,T}\big) \\
&\quad + \langle \widehat{q}_{\ell}, \boldsymbol{e}_i\rangle\big(\boldsymbol{e}_i'\Psi_{\ell,T}\big)\Bigg\}\Psi_{T,T'}\mathbf{1} \\
&\quad + \Bigg\{ \sum_{\ell=K+1}^{\widetilde{K}} \frac{1}{2}\Delta_{(\ell-1,\ell)}\big(\langle \widehat{q}_{\ell-1}, \boldsymbol{e}_i\rangle\big(\boldsymbol{e}_i'\Psi_{\ell-1,T'}\big) \\
&\quad + \langle \widehat{q}_{\ell}, \boldsymbol{e}_i\rangle\big(\boldsymbol{e}_i'\Psi_{\ell,T'}\big)\big)\Bigg\}\mathbf{1}. \qquad (5.88)
\end{aligned}$$

The update formulas incorporating the information on $[T, T']$ in the estimates (4.76) and (4.77) are, respectively,

$$\begin{aligned}
[\widehat{A}_{\tau+1}]_{(i,j)} &= [\widehat{A}_{\tau}]_{(i,j)} \\
&\quad\times \Bigg\{ \sum_{\ell=1}^{K} \frac{1}{2}\Delta_{(\ell-1,\ell)}\big(\langle \widehat{q}_{\ell-1}, \boldsymbol{e}_i\rangle\big(\boldsymbol{e}_j'\Psi_{\ell-1,T}\big) \\
&\quad + \langle \widehat{q}_{\ell}, \boldsymbol{e}_i\rangle\big(\boldsymbol{e}_j'\Psi_{\ell,T}\big)\big)\Bigg\}\Psi_{T,T'}\mathbf{1} \Big/ M_{T'} \\
&\quad + [\widehat{A}_{\tau}]_{(i,j)}\Bigg\{ \sum_{\ell=K+1}^{\widetilde{K}} \frac{1}{2}\Delta_{(\ell-1,\ell)}\big(\langle \widehat{q}_{\ell-1}, \boldsymbol{e}_i\rangle\big(\boldsymbol{e}_j'\Psi_{\ell-1,T'}\big) \\
&\quad + \langle \widehat{q}_{\ell}, \boldsymbol{e}_i\rangle\big(\boldsymbol{e}_j'\Psi_{\ell,T'}\big)\big)\Bigg\}\mathbf{1} \Big/ M_{T'} \qquad (5.89)
\end{aligned}$$

and

$$\begin{aligned}
&\langle \widehat{\lambda}_{\tau+1}, \boldsymbol{e}_i\rangle = \\
&\Bigg\{ \sum_{\ell=1}^{K} \frac{1}{2}\Delta_{(\ell-1,\ell)}\big(N_{\ell-1}\langle \widehat{q}_{\ell-1}, \boldsymbol{e}_i\rangle\big(\boldsymbol{e}_i'A'\Psi_{\ell-1,T}\big) \\
&\quad + N_{\ell}\langle \widehat{q}_{\ell}, \boldsymbol{e}_i\rangle\big(\boldsymbol{e}_i'A'\Psi_{\ell,T}\big)\big)\Bigg\}\Psi_{T,T'}\mathbf{1} \Big/ M_{T'} \\
&\quad + \Bigg\{ \sum_{\ell=K+1}^{K+\widetilde{K}} \frac{1}{2}\Delta_{(\ell-1,\ell)}\big(N_{\ell-1}\langle \widehat{q}_{\ell-1}, \boldsymbol{e}_i\rangle\big(\boldsymbol{e}_i'A'\Psi_{\ell-1,T'}\big) \\
&\quad + N_{\ell}\langle \widehat{q}_{\ell}, \boldsymbol{e}_i\rangle\big(\boldsymbol{e}_i'A'\Psi_{\ell,T'}\big)\big)\Bigg\}\mathbf{1} \Big/ M_{T'} \\
&\quad - \Bigg\{ \sum_{\ell=1}^{K} \frac{1}{2}\Delta_{(\ell-1,\ell)}\big(N_{\ell-1}\langle A\widehat{q}_{\ell-1}, \boldsymbol{e}_i\rangle\big(\boldsymbol{e}_i'\Psi_{\ell-1,T}\big) \\
&\quad + N_{\ell}\langle A\widehat{q}_{\ell}, \boldsymbol{e}_i\rangle\big(\boldsymbol{e}_i'\Psi_{\ell,T}\big)\big)\Bigg\}\Psi_{T,T'}\mathbf{1} \Big/ M_{T'} \\
&\quad - \Bigg\{ \sum_{\ell=K+1}^{K+\widetilde{K}} \frac{1}{2}\Delta_{(\ell-1,\ell)}\big(N_{\ell-1}\langle A\widehat{q}_{\ell-1}, \boldsymbol{e}_i\rangle\big(\boldsymbol{e}_i'\Psi_{\ell-1,T'}\big) \\
&\quad + N_{\ell}\langle A\widehat{q}_{\ell}, \boldsymbol{e}_i\rangle\big(\boldsymbol{e}_i'\Psi_{\ell,T'}\big)\big)\Bigg\}\mathbf{1} \Big/ M_{T'} \\
&\quad + N_{T'}\langle \widehat{q}_{T'}, \boldsymbol{e}_i\rangle\langle \widehat{v}_{T',T'}, \boldsymbol{e}_i\rangle \Big/ M_{T'} \qquad (5.90)
\end{aligned}$$

$$\begin{aligned}
[\widehat{A}_{\tau+1}]_{(i,j)} &= [\widehat{A}_{\tau}]_{(i,j)} \frac{\int_0^T \langle q_u, \boldsymbol{e}_i\rangle\langle v_{u,T}, \boldsymbol{e}_j\rangle du}{\int_0^T \langle q_u, \boldsymbol{e}_i\rangle\langle v_{u,T}, \boldsymbol{e}_i\rangle du} \\
&\approx [\widehat{A}_{\tau}]_{(i,j)} \frac{\Big[\sum_{\ell=1}^{K} \frac{1}{2}\Delta_{(\ell-1,\ell)}\big(\langle \widehat{q}_{\ell-1}, \boldsymbol{e}_i\rangle\langle \widehat{v}_{\ell-1,T}, \boldsymbol{e}_j\rangle + \langle \widehat{q}_{\ell}, \boldsymbol{e}_i\rangle\langle \widehat{v}_{\ell,T}, \boldsymbol{e}_j\rangle\big)\Big]}{\Big[\sum_{\ell=1}^{K} \frac{1}{2}\Delta_{(\ell-1,\ell)}\big(\langle \widehat{q}_{\ell-1}, \boldsymbol{e}_i\rangle\langle \widehat{v}_{\ell-1,T}, \boldsymbol{e}_i\rangle + \langle \widehat{q}_{\ell}, \boldsymbol{e}_i\rangle\langle \widehat{v}_{\ell,T}, \boldsymbol{e}_i\rangle\big)\Big]} \\
&= [\widehat{A}_{\tau}]_{(i,j)} \frac{\Big[\sum_{\ell=1}^{K} \frac{1}{2}\Delta_{(\ell-1,\ell)}\big(\langle \widehat{q}_{\ell-1}, \boldsymbol{e}_i\rangle\big(\boldsymbol{e}_j'\Psi_{\ell-1,T}\big) + \langle \widehat{q}_{\ell}, \boldsymbol{e}_i\rangle\big(\boldsymbol{e}_j'\Psi_{\ell,T}\big)\big)\Big\}\mathbf{1}\Big]}{\Big[\sum_{\ell=1}^{K} \frac{1}{2}\Delta_{(\ell-1,\ell)}\big(\langle \widehat{q}_{\ell-1}, \boldsymbol{e}_i\rangle\big(\boldsymbol{e}_i'\Psi_{\ell-1,T}\big) + \langle \widehat{q}_{\ell}, \boldsymbol{e}_i\rangle\big(\boldsymbol{e}_i'\Psi_{\ell,T}\big)\big)\Big]\mathbf{1}}. \qquad (5.86)
\end{aligned}$$

$$\begin{aligned}
&\langle \widehat{\lambda}_{\tau+1}, \boldsymbol{e}_i\rangle \\
&= \frac{\Big[\sum_{\ell=1}^{K} \frac{1}{2}\Delta_{(\ell-1,\ell)}\big(N_{\ell-1}\langle \widehat{q}_{\ell-1}, \boldsymbol{e}_i\rangle\big(\boldsymbol{e}_i'A'\Psi_{\ell-1,T}\big) + N_{\ell}\langle \widehat{q}_{\ell}, \boldsymbol{e}_i\rangle\big(\boldsymbol{e}_i'A'\Psi_{\ell,T}\big)\big)\Big]\mathbf{1}}{\Big[\sum_{\ell=1}^{K} \frac{1}{2}\Delta_{(\ell-1,\ell)}\big(\langle \widehat{q}_{\ell-1}, \boldsymbol{e}_i\rangle\big(\boldsymbol{e}_i'\Psi_{\ell-1,T}\big) + \langle \widehat{q}_{\ell}, \boldsymbol{e}_i\rangle\big(\boldsymbol{e}_i'\Psi_{\ell,T}\big)\big)\Big]\mathbf{1}} \\
&\quad - \frac{\Big[\sum_{\ell=1}^{K} \frac{1}{2}\Delta_{(\ell-1,\ell)}\big(N_{\ell-1}\langle A\widehat{q}_{\ell-1}, \boldsymbol{e}_i\rangle\big(\boldsymbol{e}_i'\Psi_{\ell-1,T}\big) + N_{\ell}\langle A\widehat{q}_{\ell}, \boldsymbol{e}_i\rangle(\boldsymbol{e}_i'\Psi_{\ell,T})\big)\Big]\mathbf{1}}{\Big[\sum_{\ell=1}^{K} \frac{1}{2}\Delta_{(\ell-1,\ell)}\big(\langle \widehat{q}_{\ell-1}, \boldsymbol{e}_i\rangle \times \big(\boldsymbol{e}_i'\Psi_{\ell-1,T}\big) + \langle \widehat{q}_{\ell}, \boldsymbol{e}_i\rangle\big(\boldsymbol{e}_i'\Psi_{\ell,T}\big)\big)\Big]\mathbf{1}} \\
&\quad + \frac{N_T\langle \widehat{q}_T, \boldsymbol{e}_i\rangle\langle \widehat{v}_{T,T}, \boldsymbol{e}_i\rangle}{\Big[\sum_{\ell=1}^{K} \frac{1}{2}\Delta_{(\ell-1,\ell)}\big(\langle \widehat{q}_{\ell-1}, \boldsymbol{e}_i\rangle\big(\boldsymbol{e}_i'\Psi_{\ell-1,T}\big) + \langle \widehat{q}_{\ell}, \boldsymbol{e}_i\rangle\big(\boldsymbol{e}_i'\Psi_{\ell,T}\big)\big)\Big]\mathbf{1}}. \qquad (5.87)
\end{aligned}$$

Q2 Note that the sums in these two formulas are approximating integrals and need not be completely recalculated. Write, for example,

$$B \triangleq \Bigg\{ \sum_{\ell=1}^{K} \frac{1}{2} \Delta_{(\ell-1,\ell)} \big(\langle \widehat{q}_{\ell-1}, \boldsymbol{e}_i \rangle \big(\boldsymbol{e}_j' \Psi_{\ell-1,T} \big) + \langle \widehat{q}_{\ell}, \boldsymbol{e}_i \rangle \big(\boldsymbol{e}_j' \Psi_{\ell,T} \big) \big) \Bigg\}. \tag{5.91}$$

This $n \times n$ matrix appears in (5.89) and can be stored in memory. Upon the arrival of new information on $[T, T']$, the matrix B can be recalled from memory and multiplied on the right-hand side by $\Psi_{T,T'}\mathbf{1}$. Similarly, the corresponding matrix in (5.90) can be stored in memory, avoiding a complete pass through the data as in previous algorithms cited in the bibliography.

Appendix I

Proof of Lemma 1

Proof: To establish Lemma 1, we first show that M is $(P, \mathcal{F})$-martingale. Since, under $P^{\dagger}$, the process ΛM has dynamics

$$\Lambda_t M_t = M_0 + \int_0^t \Lambda_u \, dM_u + \int_0^t M_u \, d\Lambda_u \tag{A1}$$

it follows that ΛM is a $P^{\dagger}$-martingale. Using the abstract form of Bayes' rule, we see that, for $t \geq s$,

$$E\Big[M_t | \mathcal{F}_s\Big] = \frac{E^{\dagger}[\Lambda_t M_t | \mathcal{F}_s]}{E^{\dagger}\Big[\Lambda_t | \mathcal{F}_s\Big]} = \frac{\Lambda_s M_s}{\Lambda_s} = M_s. \tag{A2}$$

Therefore,

$$E\Big[M_t - M_s | \mathcal{F}_s\Big] = \mathbf{0} \in \mathbb{R}^n. \tag{A3}$$

So,

$$E\Big[X_t - X_s - \int_s^t AX_u \, du | \mathcal{F}_s\Big] = \mathbf{0} \in \mathbb{R}^n. \tag{A4}$$

Then,

$$\begin{aligned} Z_t \triangleq E\Big[X_t | \mathcal{F}_s\Big] &= X_s + \int_s^t AE\Big[X_u | \mathcal{F}_u\Big] du \\ &= X_s + \int_s^t AZ_u \, du. \end{aligned} \tag{A5}$$

Therefore,

$$Z_t = \exp\big(A(t-s)\big) X_s = E\Big[X_t | \mathcal{F}_s\Big] = E\Big[X_t | X_s\Big]. \tag{A6}$$

Equation (A6) shows that, under the measure $P^{\dagger}$, the process X satisfies the Markov property and that its rate matrix is again A. To complete the proof, we note that

$$E\Big[X_0 | \Big\{\Omega, \emptyset\Big\}\Big] = \frac{E^{\dagger}\Big[\Lambda_0^{-1} X_0 | \Big\{\Omega, \emptyset\Big\}\Big]}{E^{\dagger}\Big[\Lambda_0^{-1} | \Big\{\Omega, \emptyset\Big\}\Big]} = E\Big[X_0\Big] = p_0. \tag{A7}$$

Appendix II

Derivation of the Stochastic Integral Equation (3.12)

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

$$E[X_t | \mathcal{Y}_{0,t}] = \frac{E^{\dagger}[\Lambda_{0,t} X_t | \mathcal{Y}_{0,t}]}{E^{\dagger}[\Lambda_{0,t} | \mathcal{Y}_{0,t}]}. \tag{B1}$$

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

$$\begin{aligned} \langle E^{\dagger}[\Lambda_{0,t} X_t | \mathcal{Y}_{0,t}], \mathbf{1} \rangle &= E^{\dagger}[\Lambda_{0,t} \langle X_t, \mathbf{1} \rangle | \mathcal{Y}_{0,t}] \\ &= E^{\dagger}[\Lambda_{0,t} | \mathcal{Y}_{0,t}]. \end{aligned} \tag{B2}$$

That is, if we write

$$q_t = E^{\dagger}[\Lambda_{0,t} X_t | \mathcal{Y}_{0,t}] \tag{B3}$$

then

$$P(X_t = \boldsymbol{e}_i) \triangleq E[X_t = \boldsymbol{e}_i | \mathcal{Y}_{0,t}] = \frac{1}{\langle q_t, \mathbf{1} \rangle} \langle q_t, \boldsymbol{e}_i \rangle. \tag{B4}$$

To compute the expectation at (B3), we first apply the product rule to determine the decomposition for the process ΛX

$$\begin{aligned} \Lambda_{0,t} X_t &= X_0 + \int_0^t \Lambda_{0,u} AX_u \, du + \int_0^t \Lambda_{u-} dM_u \\ &\quad + \int_0^t X_{u-} \big(\langle X_{u-}, \lambda \rangle - 1 \big) \Lambda_{0,u-} \big(dN_u - du \big) \\ &= X_0 + \int_0^t \Lambda_{0,u} AX_u \, du + \int_0^t \Lambda_{u-} dM_u \\ &\quad + \sum_{i=1}^{n} \int_0^t \langle X_{u-}, \boldsymbol{e}_i \rangle (\langle \lambda, \boldsymbol{e}_i \rangle - 1) \Lambda_{0,u-} (dN_u - du) \boldsymbol{e}_i. \end{aligned} \tag{B5}$$

By conditioning both sides of (B5) on $\mathcal{Y}_{0,t}$ under the reference probability $P^{\dagger}$, it then follows that the process q has the dynamics

$$\begin{aligned} q_t &= q_0 + \int_0^t Aq_u \, du \\ &\quad + \int_0^t \operatorname{diag}\Big\{ \langle \lambda, \boldsymbol{e}_\ell \rangle - 1 \Big\} q_{u-} \Big(dN_u - du \Big). \end{aligned}$$

Appendix III

Proof of Theorem 3

To compute the dynamics of the process $\sigma(G^i X)$, we must evaluate the expectation $E^{\dagger}[\Lambda_t G_t^i X_t | \mathcal{Y}_{0,t}]$. Using the product rule, we compute the decomposition of the process $GX\Lambda$

$$\begin{aligned} &\Lambda_t G_t^i X_t \\ &= \int_0^t \Lambda_{u-} X_u \langle X_u, \boldsymbol{e}_i \rangle \, dN_u + \int_0^t \Lambda_u G_s^i AX_u \, du \\ &\quad + \int_0^t \Lambda_u G_u^i dM_u + \int_0^t G_u^i X_u \Lambda_{u-} \big(\langle X_u, \lambda \rangle - 1 \big) \, (dN_u - du) \\ &\quad + \int_0^t X_u \langle X_u, \boldsymbol{e}_i \rangle \Lambda_{u-} \big(\langle X_u, \Lambda \rangle - 1 \big) \big(dN_u - du \big). \end{aligned} \tag{C1}$$

The result follows by conditioning both sides of (C1) on $\mathcal{Y}_{0,t}$, using a version of Fubini's theorem [27] and noting that, under the measure $P^\dagger$, the process N is a standard Poisson process. Consequently we see that

$$\begin{aligned}
&E^\dagger[\Lambda_t G_t^i X_t|\mathcal{Y}_{0,t}] \\
&\quad= \int_0^t E^\dagger[\Lambda_{u-} X_u \langle X_u, \boldsymbol{e}_i\rangle|\mathcal{Y}_{0,s}]dN_u \\
&\qquad+ \int_0^t E^\dagger[\Lambda_u G_u^i A X_u|\mathcal{Y}_{0,s}]du + \int_0^t E^\dagger[\Lambda_u G_u^i|\mathcal{Y}_{0,s}]dM_u \\
&\qquad+ \int_0^t E^\dagger[G_u^i X_u \Lambda_{u-}\big(\langle X_u, \lambda\rangle - 1\big)|\mathcal{Y}_{0,s}]\big(dN_u - du\big) \\
&\qquad+ \int_0^t E^\dagger[X_u\langle X_u, \boldsymbol{e}_i\rangle \Lambda_{u-}\big(\langle X_u, \Lambda\rangle - 1\big)|\mathcal{Y}_{0,s}]\big(dN_u - du\big) \\
&\quad= \int_0^t A\,\sigma(G_u^i X_u)du \\
&\qquad+ \int_0^t \operatorname{diag}\Big\{\langle\lambda, \boldsymbol{e}_\ell\rangle - 1\Big\}\sigma(G_u^i X_u)\big(dN_u - du\big) \\
&\qquad+ \int_0^t \langle q_u, \boldsymbol{e}_i\rangle\langle\lambda, \boldsymbol{e}_i\rangle dN_u \boldsymbol{e}_i. \qquad\qquad (C2)
\end{aligned}$$

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers, whose suggestions have greatly improved the manuscript.

W. P. Malcolm would also like to acknowledge: Prof. D. J. Daley, for his helpful technical discussions concerning the doubly stochastic Poisson processes.

REFERENCES

[1] J. M. C. Clark, "The design of robust approximations to the stochastic differential equations for nonlinear filtering," in *Communications Systems and Random Process Theory*, J. K. Skwirzynski, Ed. Alphen aan den Rijn, The Netherlands: Sijthoff and Noorhoff, 1978, pp. 721–734.

[2] R. J. Elliott, *Stochastic Calculus and its Applications*. New York: Springer-Verlag, 1982.

[3] L. Aggoun and R. J. Elliott, *Measure Theory and Filtering*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[4] R. J. Elliott, "New finite dimensional filters and smoothers for noisily observed markov chains," *IEEE Trans. Inf. Theory*, vol. 39, no. 1, pp. 265–271, Jan. 1993.

[5] P. Bremaud, *Point Processes and Queues, Martingale Dynamics*. New York: Springer-Verlag, 1981.

[6] F. Campillo and F. L. Gland, "Maximum likelihood estimation for partially observed diffusions: Direct maximization versus the em algorithm," *Stochastic Processes Appl.*, vol. 33, no. 2, pp. 245–274, 1989.

[7] A. Dembo and O. Zeitouni, "Parameter estimation of partially observed continuous time stochastic processes via the em algorithm," *Stochastic Processes Appl.*, vol. 23, pp. 91–113, 1986.

[8] A. D. Demster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the em algorithm," *J. Royal Statist. Soc.*, vol. B-39, pp. 1–37, 1997.

[9] R. J. Elliott, "Exact adaptive filters for Markov chains observed in Gaussian noise," *Automatica*, vol. 30, no. 9, pp. 1399–1408, Sep. 1994.

[10] M. R. James, V. Krishnamurthy, and F. L. Gland, "Time discretization for continuous time filters and smoothers for hmm parameter estimation," *IEEE Trans. Inf. Theory*, vol. 42, no. 2, pp. 593–605, Mar. 1998.

[11] F. Konecny, "Parameter estimation for point processes with partial observations: Afiltering approach," *Syst. Control Lett.*, vol. xx, no. 4, pp. 281–286, 1984.

[12] G. Lindgren and U. Holst, "Recursive estimation of parameters in Markov-modulated Poisson processes," *IEEE Trans. Commun.*, vol. 43, no. 11, pp. 2812–2820, Nov. 1995.

[13] W. P. Malcolm, "Robust filtering and estimation with Poisson observations" Ph.D. dissertation, The Australian Nat. Univ., Canberra, Australia, 1999.

[14] W. P. Malcolm and R. J. Elliott, "A general smoothing equation for Poisson observations," presented at the IEEE Conf. Decision Control, Phoenix, AZ, Dec. 1999.

[15] R. J. Elliott and W. P. Malcolm, "Robust smoother dynamics for Poisson processes driven by an itô diffusion," presented at the IEEE Conf. Decision Control, Orlando, FL, Dec. 2001.

[16] W. P. Malcolm, R. J. Elliott, and J. van der Hoek, "On the numerical stability of time-discretized state estimation via clark transformations," presented at the IEEE Conf. Decision Control, Mauii, HI, Dec. 2003.

[17] J. McLachlan and T. Krishnan, *The EM Algorithm and Extensions*. New York: Wiley, 1997.

[18] E. Pardoux, "Equations du filtrage nonlineare de la predictions et du lissage," *Stochastics*, vol. xx, no. 6, pp. 193–231, 1982.

[19] E. Pardoux and P. Protter, "A two-sided stochastic integral and its calculus," *Probab. Theory Relat. Fields*, vol. 76, pp. 15–49, 1987.

[20] H. Royden, *Real Analysis*, 3rd ed. New York: MacMillan, 1988.

[21] T. Ryden, "Parameter estimation for markov modulated Poisson processes," *Commun. Statist. Stochastic Models*, vol. 10, no. 4, pp. 795–829, 1994.

[22] T. Ryden, "An em algorithm for estimation in Markov-modulated poisson processes," *Comput. Statist. Data Anal.*, vol. 21, pp. 431–447, 1996.

[23] A. Segall, M. H. A. Davis, and T. Kailath, "Nonlinear filtering with counting process observations," *IEEE Trans. Inf. Theory*, vol. 21, no. 2, pp. 143–149, Mar. 1975.

[24] R. H. Shumway and D. S. Stoffer, "An approach to time series smoothing and forecasting using the em algorithm," *J. Time Ser. Anal.*, vol. 3, no. 4, pp. 682–687, 1989.

[25] D. L. Snyder, "Filtering and detection for doubly stochastic poisson processes," *IEEE Trans. Inf. Theory*, vol. 18, no. 1, pp. 91–102, Jan. 1972.

[26] D. L. Snyder, "Smoothing for doubly stochastic poisson processes," *IEEE Trans. Inf. Theory*, vol. 18, no. 5, pp. 558–562, Sep. 1972.

[27] E. Wong and B. Hajek, *Stochastic Processes in Engineering Systems*. New York: Springer-Verlag, 1985.

[28] O. Zeitouni and A. Dembo, "Exact filters for the estimation of the number of transitions of finite-state continuous time Markov processes," *IEEE Trans. Inf. Theory*, vol. 34, no. 4, pp. 890–893, Jul. 1988.

[29] D. H. I. Ziskind, "Maximum likelihood localization of narrow band sources via the em algorithm," presented at the IEEE Int. Conf. Acoust. Speech Signal Process., Minneapolis, MN, Apr. 1993.

[30] L. Deng and J. W. Mark, "Parameter estimation for Markov modulated Poisson processes via the em algorithm with time discretization," *Telecommun. Syst.*, vol. 1, pp. 321–328, 1993.

Robert J. Elliott received the Bachelor's and Master's degrees from Oxford University, Oxford, U.K., and the Ph.D. and D.Sc. degrees from Cambridge University, Cambridge, U.K.

He has held positions at Newcastle, Yale, Oxford, Warwick, Hull, and Alberta, and visiting positions in Toronto, Northwestern, Kentucky, Brown, Paris, Denmark, Hong Kong, and Australia. He is currently the AF Collins Professor of Finance at the University of Alberta, Edmonton, Canada, and also jointly holds a position in the Mathematical Sciences. He is a frequent visitor to Adelaide. He is the author or coauthor of more than 250 papers published in various international journals and six books, including *Mathematics of Financial Markets* (Springer). His current research interests include stochastic processes in engineering and finance.

W. P. Malcolm received the Bachelor's degree in applied physics from the Royal Melbourne Institute of Technology, Malbourne, Australia, the Master's degree in applied mathematics from the Flinders University, Adelaide, South Australia, and the Ph.D. degree in electrical engineering from the Australian National University, Canberra, Australia.

He is currently a Senior Researcher in the Statistical Machine Learning Program of the National ICT Australia (NICTA), Canberra, Australia. His current research interests include filtering, estimation, stochastic control, and measures of risk in decision making with uncertainty. In particular, filtering and stochastic control with point process observations and marked point process observations.

QUERIES

Q1. Author: Please provide the IEEE membership details, if any, for both the authors.

Q2. Author: Fraction given in solidus form in Eq. (5.90) have been set as display fraction. Please check if equation is ok as typeset.

Q3. Author: Please provide the subjects and the years in which R. J. Elliott received the Bachelor's, Master's, Ph.D. and D.Sc. degrees. Also, please provide the names of the institutes in Newcastle, Yale, Oxford, Warwick, Hull, Alberta, Toronto, Northwestern, Kentucky, Brown, Paris, Denmark, Hong Kong, and Australia, where he held various positions.

Q4. Author: Please provide the years in which W. P. Malcolm received the Bachelors, Masters, and Ph.D. degrees.