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 queuing 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 \(X = \{X_t, 0 \leq t\}\) 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 \(\mathcal{L} = \{e_1, e_2, \ldots, e_n\}\), where \(e_i\) denotes a column vector in \(\mathbb{R}^n\) with unity in the \(i\)th position and zero elsewhere. The dynamics for this process are

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

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

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

\[
N_t = \int_0^t \langle X_u, \lambda \rangle du + V_t.
\]

(2.2)

Here, \( V \) is a \((\sigma\{N_u, 0 \leq u \leq t\}, P)\)-martingale, \( \langle \cdot, \cdot \rangle \) denotes an inner product, and \( \lambda \in \mathbb{R}^+_0 \) is a vector of \( n \) nonnegative Poisson intensities. Our filtrations are given by

\[
\mathcal{F}_t = \{ F_1 \}, \quad \text{where } F_t = \sigma \{ X_u; 0 \leq u \leq t \}
\]

(2.3)

\[
\mathcal{Y}_{0,t} = \{ Y_{0,t} \}, \quad \text{where } Y_{0,t} = \sigma \{ N_u; 0 \leq u \leq t \}
\]

(2.4)

\[
\mathcal{G}_{0,t} = \{ G_{0,t} \}, \quad \text{where } G_{0,t} = \sigma \{ N_u, X_u; 0 \leq u \leq t \}.
\]

(2.5)

C. Reference Probability

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

1) The state process \( X \) is a Markov process with intensity matrix \( \Lambda \) 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

\[
\frac{dP}{dP^\theta} \bigg|_{\mathcal{G}_{0,t}} = \Lambda_{0,t}
\]

(2.6)

where

\[
\Lambda_{0,t} = \prod_{0 < u \leq t} \langle X_u, \lambda \rangle^{N_u} \exp \left\{ \int_0^t (1 - \langle X_u, \lambda \rangle) du \right\} = 1 + \int_0^t \Lambda_u - (\langle X_u, \lambda \rangle - 1) \left( dN_u - du \right).
\]

(2.7)

Here,

\[
\Delta N_r \triangleq N_r - \lim_{\epsilon \to 0} N_{r-\epsilon} = N_r - N_{r-}.
\]

(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 \( \theta \) index a given family of probability measures \( P_\theta \), where \( \theta \in \Theta \). All such measures \( P_\theta \) 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 \( \theta^* = \hat{\theta}_r \), then compute \( Q(\cdot, \theta^*) \), where

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

(3.9)

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

\[
\hat{\theta}_{r+1} \in \arg \max_{\theta \in \Theta} Q(\theta, \theta^*)
\]

(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 \( E[X_t | Y_t] \). By Bayes’ Theorem this is \( E'[A_t X_t | Y_{0,t}] / E'[A_t | Y_{0,t}] \). Write

\[
q_t \triangleq E'[A_t X_t | Y_{0,t}].
\]

(3.11)

Then,

\[
q_t = q_0 + \int_0^t A_q d\lambda_u du
\]

(3.12)

where

\[
\text{diag} \{ (\lambda, e_i) - 1 \} = \begin{bmatrix}
(\lambda, e_1) - 1 \\
(\lambda, e_2) - 1 \\
\vdots \\
(\lambda, e_n) - 1
\end{bmatrix}
\]

(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 \( \sum_{i=1}^n \langle X_t, e_i \rangle = 1 \), so \( E'[A_t | Y_t] = \langle q_t, 1 \rangle \). Here, \( 1 = (1, 1, \ldots, 1) \in \mathbb{R}^n \). Therefore,

\[
P(X_t = e_i | Y_{0,t}) = \frac{\langle q_t, e_i \rangle}{\sum_{i=1}^n \langle q_t, e_i \rangle}.
\]

(3.14)

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

2) \( J_i \), the cumulative sojourn time spent by the process \( X \) in state \( e_i \), is

\[
J_i = \int_0^t \langle X_u, e_i \rangle du.
\]

(3.15)
3) \(N_{t}^{(j,i)}\), the number of transitions \(e_{i} \rightarrow e_{j}\) of \(X\), where \(i \neq j\), up to time \(t\) is
\[
N_{t}^{(j,i)} = \int_{0}^{t} (X_{u-}, e_{i}) \langle dX_{u}, e_{j} \rangle.
\] (3.16)

4) \(G_{t}^{i}\), the level integrals for the state \(e_{i}\), is
\[
G_{t}^{i} = \int_{0}^{t} (X_{u}, e_{i}) dN_{u}.
\] (3.17)

Using Bayes’ Theorem, if \(H = \{H_{t}, 0 \leq t\}\) is any \(G\) adapted process
\[
E[H_{t} | Y_{0,t}] = \frac{E[I_{t} | H_{t}, Y_{0,t}]}{E[I_{t} | Y_{0,t}]}.
\] (3.18)

Here, \(E[I_{\cdot}]\) denotes an expectation taken under the measure \(P_{\cdot}\).

Write
\[
\sigma(H_{t}) = E[I_{t} | H_{t}, Y_{0,t}].
\] (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:
\[
[A_{\tau+1}]_{(i,j)} = \frac{E[N_{t}^{(j,i)} | Y_{0,t}]}{E[J_{t}^{i} | Y_{0,t}]} = \sigma(N_{t}^{(j,i)}) \sigma(J_{t}^{i})
\] (3.20)

and
\[
(\lambda_{\tau+1}, e_{i}) = \frac{E[G_{t}^{i} | Y_{0,t}]}{E[J_{t}^{i} | Y_{0,t}]} = \frac{\sigma(G_{t}^{i})}{\sigma(J_{t}^{i})}.
\] (3.21)

The conditional expectations in equations (3.20) and (3.21) are computed using the previous (at pass \(\tau\)) parameter estimates for \(A\) and \(\lambda\).

The updates for \([A_{\tau}]_{(i,j)}\) and \((\lambda_{\tau}, e_{i})\) 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 \(\sigma(J^{i} X_{t}), \sigma(N_{t}^{(j,i)} X_{t}),\) and \(\sigma(G_{t}^{i} X_{t})\), where, for example,
\[
\sigma(G_{t}^{i} X_{t}) = E[I_{t} | G_{t}^{i} X_{t}, Y_{t}] \in \mathbb{R}^{n}.
\] (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 \(J^{i} X_{t}, N^{(j,i)} X_{t},\) and \(G^{i} X_{t}\).

**Theorem 1:** The vector-valued process \(\sigma(J^{i} X_{t}) \in \mathbb{R}^{n}\) satisfies the stochastic integral equation
\[
\sigma(J^{i} X_{t}) = \int_{0}^{t} A \sigma(J^{i} X_{u}) du + \int_{0}^{t} \langle q_{u}, e_{i} \rangle du e_{i}
\]
\[+ \int_{0}^{t} \text{diag} \{\langle \lambda, e_{\ell} \rangle - 1\} \sigma(J^{i} X_{u} -) (dN_{u} - du).
\] (3.23)

Here, \(\sigma(J^{0}_{0} X_{0}) = 0\) and \(q\) is the solution of (3.12).

**Theorem 2:** The vector-valued process \(\sigma(N^{(j,i)} X_{t}) \in \mathbb{R}^{n}\) satisfies the stochastic integral equation
\[
\sigma(N^{(j,i)} X_{t}) = \int_{0}^{t} A \sigma(N^{(j,i)} X_{u}) du
\]
\[+ \int_{0}^{t} \langle q_{u}, e_{i} \rangle \langle A e_{i}, e_{j} \rangle du e_{j}
\]
\[+ \int_{0}^{t} \text{diag} \{\langle \lambda, e_{\ell} \rangle - 1\} \sigma(N^{(j,i)} X_{u} -) (dN_{u} - du).
\] (3.24)

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

**Theorem 3:** The vector-valued process \(\sigma(G^{i} X_{t}) \in \mathbb{R}^{n}\) satisfies the stochastic integral equation
\[
\sigma(G^{i} X_{t}) = \int_{0}^{t} A \sigma(G^{i} X_{u}) du
\]
\[+ \int_{0}^{t} \langle q_{u}, e_{i} \rangle \langle \lambda, e_{i} \rangle du e_{i}
\]
\[+ \int_{0}^{t} \text{diag} \{\langle \lambda, e_{\ell} \rangle - 1\} \sigma(G^{i} X_{u} -) (dN_{u} - du).
\] (3.25)

Here, \(\sigma(G^{0}_{0} X_{0}) = 0\) 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
\[
[A_{\tau+1}]_{(i,j)} = \frac{\sigma(N^{(j,i)} X_{T})}{\sigma(J^{i} X_{T})}, (3.26)
\]

and
\[
(\lambda_{\tau+1}, e_{i}) = \frac{\sigma(G^{i} X_{T})}{\sigma(J^{i} X_{T})}, (3.27)
\]

**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
\[
\Gamma_{i} = \text{diag} \{\gamma_{i}^{T}\} \in \mathbb{R}^{n \times n}.
\] (3.28)

Here, \(\gamma_{i}^{T} = \exp \{t (\lambda_{i}, e_{i})\} \{\lambda_{i}, e_{i}\}^{N_{i}}\) with \(\gamma_{0}^{T} = 0\). Note that the matrix \(\Gamma_{i}^{-1}\) is nonsingular. Using the Itô rule, one can show that
\[
\Gamma_{i}^{-1} = \int_{0}^{t} \text{diag} \{\langle \lambda, e_{\ell} \rangle - 1\} \Gamma_{u}^{-1} du
\]
\[+ \int_{0}^{t} \Gamma_{u}^{-1} \text{diag} \{\langle \lambda, e_{\ell} \rangle - 1\} dN_{u}.
\] (3.29)
Finally, our dynamics for $\mathbf{\sigma}(G_t^i X_t)$ read:

$$\mathbf{\sigma}(G_t^i X_t) = \int_0^t \Gamma_{u}^{-1} A \Gamma_{u} \mathbf{\sigma}(G_u^i X_u) du$$

$$+ \langle \theta_{t}, e_i \rangle N_{t} e_i - \int_0^t N_u \langle d\theta_u, e_i \rangle e_i.$$  

(3.37)

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

$$\mathbf{\sigma}(J_t^i X_t) = \int_0^t \Gamma_{u}^{-1} A \Gamma_{u} \mathbf{\sigma}(J_u^i X_u) du + \int_0^t \langle \theta_u, e_i \rangle du e_i$$

(3.38)

and

$$\mathbf{\sigma}(N_t^{j,i} X_t) = \int_0^t \Gamma_{u}^{-1} A \Gamma_{u} \mathbf{\sigma}(N_u^{j,i} X_u) du + \int_0^t \langle \theta_u, e_i \rangle \langle A e_i, e_j \rangle du e_j.$$  

(3.39)

D. Discrete-Time Filters

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

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

(3.40)

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

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

(3.41)

For brevity, we shall use the notation $\xi_k \Delta \hat{\xi}_{t_k}$, where $\xi_k$ denotes a process $\xi$ 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

$$\hat{\theta}_{t_k} \approx \theta_{t_{k-1}} + \Gamma_{t_{k-1}}^{-1} A \Gamma_{t_{k-1}} \hat{\theta}_{t_{k-1}} \Delta_{(k-1,k)}.$$  

(3.42)

so

$$q_k = G_{t_k} \hat{\theta}_{t_k} \approx G_{t_k} \Gamma_{t_{k-1}}^{-1} [ I + \Delta_{(k-1,k)} A ] \hat{q}_{t_{k-1}}.$$  

(3.43)

This suggests the recursion

$$\hat{q}_k \Delta \Gamma_{k-1} [ I + \Delta_{(k-1,k)} A ] \hat{q}_{k-1}.$$  

(3.44)

Here, $\hat{q}$ 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 $\langle q, e_i \rangle \geq 0$ 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

$$\mathbf{\sigma}(G_t^i X_t) = \mathbf{\sigma}(G_{t_{k-1}}^i X_{t_{k-1}})$$

$$+ \int_{t_{k-1}}^{t_k} \Gamma_{u}^{-1} A \Gamma_{u} \mathbf{\sigma}(G_u^i X_u) du$$

(3.36)
Making an Euler–Maruyama\(^1\) approximation, we have
\[\int_{t_{k-1}}^{t_k} N_u(d\tilde{\eta}_u, e_i) e_i = \int_{t_{k-1}}^{t_k} N_u(\Gamma_u^{-1} A \tilde{\eta}_u, d\tilde{\eta}_u) e_i e_i \approx N_{t_{k-1}}^{-1} \langle A q_{t_{k-1}} e_i \rangle \Delta_{(k-1,k)} e_i\]
(3.46)

and with some algebraic manipulation
\[\langle q_{t_{k-1}}, e_i \rangle \langle A \tilde{q}_{t_{k-1}} e_i \rangle N_{t_{k-1}} e_i - \langle q_{t_{k-1}}, e_i \rangle N_{t_{k-1}} e_i = \Gamma_{k-1}^{-1} \langle q_{t_{k-1}}, e_i \rangle (N_k - N_{k-1}) e_i + \Delta_{(k-1,k)} \Gamma_{k-1}^{-1} \langle A \tilde{q}_{t_{k-1}} e_i \rangle N_{t_{k-1}} e_i\]
(3.47)

we see that
\[\tilde{\sigma}(G_k X_k) \approx \tilde{\sigma}(G_{k-1} X_{k-1}) + \Gamma_{k-1}^{-1} A \Gamma_{k-1} \tilde{\eta}(G_{k-1} X_{k-1}) \Delta_{(k-1,k)} + \Gamma_{k-1}^{-1} \langle \tilde{q}_{t_{k-1}}, e_i \rangle (N_k - N_{k-1}) e_i + \Delta_{(k-1,k)} \Gamma_{k-1}^{-1} \langle A \tilde{q}_{t_{k-1}} e_i \rangle N_{t_{k-1}} e_i - N_{t_{k-1}}^{-1} \langle A \tilde{q}_{t_{k-1}} e_i \rangle \Delta_{(k-1,k)} e_i\]
(3.48)

Now, by multiplying both sides of (3.48) on the left-hand side by the matrix \(\Gamma_k\), we get
\[\sigma(G_k X_k) \approx \sigma(G_{k-1} X_{k-1}) + \Gamma_k \Gamma_{k-1}^{-1} A \sigma(G_{k-1} X_{k-1}) \Delta_{(k-1,k)} + \Gamma_k \Gamma_{k-1}^{-1} \langle \tilde{q}_{t_{k-1}}, e_i \rangle (N_k - N_{k-1}) e_i + \Gamma_k \Gamma_{k-1}^{-1} \langle A \tilde{q}_{t_{k-1}} e_i \rangle (N_k - N_{k-1}) e_i\]
(3.49)

Our estimator of the quantity \(\sigma(G_k X_k)\) has dynamics
\[\hat{\sigma}(G_k X_k) \hat{=} \Gamma_k \Gamma_{k-1}^{-1} [I + \Delta_{(k-1,k)} A] \hat{\sigma}(G_{k-1} X_{k-1}) + \Gamma_k \Gamma_{k-1}^{-1} \langle \tilde{q}_{t_{k-1}}, e_i \rangle + \Delta_{(k-1,k)} \langle A \tilde{q}_{t_{k-1}} e_i \rangle \times (N_k - N_{k-1}) e_i\]
(3.50)

After similar calculations, the remaining discretized filters read
\[\hat{\sigma}(N_{(j,i)} X_k) = \Gamma_k \Gamma_{k-1}^{-1} [I + \Delta_{(k-1,k)} A] \hat{\sigma}(N_{(j,i)} X_{k-1}) + \Gamma_k \Gamma_{k-1}^{-1} \langle \tilde{q}_{t_{k-1}}, e_i \rangle \langle A e_i, e_j \rangle \Delta_{(k-1,k)} e_i\]
(3.15)

\[\hat{\sigma}(J_k^j X_k) = \Gamma_k \Gamma_{k-1}^{-1} [I + \Delta_{(k-1,k)} A] \hat{\sigma}(J_{k-1}^j X_{k-1}) + \Gamma_k \Gamma_{k-1}^{-1} \langle \tilde{q}_{t_{k-1}}, e_i \rangle e_i\]
(3.15)

\(^1\)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

\[\text{Initialization:} \quad \forall (i, j) \in \{(1, 1), (1, 2), \ldots, (n, n)\}, \quad \text{Choose} \quad [\hat{A}_0]_{i,j}, \quad \text{for each} \quad i \in \{1, 2, \ldots, n\} \quad \text{choose} \quad (\hat{\lambda}, e_i).
\]

Step 1 Using (3.26) and (3.27), compute the MLEs. [\hat{A}_{i+1}, e_{i}], and \(\hat{\lambda}_{i+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 \(E[X_t | Y_{0,t}]\), where \(0 \leq t \leq T\). By Bayes’ rule [3], we have
\[E[X_t | Y_{0,t}] = E[|A_{0,T} X_t | Y_{0,t}] E[|A_{0,T} | Y_{0,t}].\]
(4.53)

Consider the numerator of (4.53)
\[r_t \overset{\Delta}{=} E[|A_{0,T} X_t | Y_{0,t}] = E[|A_{0,t} A_{t,T} X_t | Y_{0,t}] = E[|E[|A_{0,t} X_t | Y_{0,t} \cup F_t]|Y_{0,t}] = E[|A_{0,t} X_t | E[|A_{t,T} | Y_{0,t} \cup F_t]|Y_{0,t}].\]
(4.54)

Under the measure \(P^t\), \(X\) is a Markov process, so the inner expectation in the previous line of (4.54) is
\[E[|A_{t,T} | Y_{0,t} \cup F_t] = E[|A_{t,T} | Y_{0,t} \cup \sigma(X_t)].\]
(4.55)

Write
\[v_{t,T}^t \overset{\Delta}{=} E[|A_{t,T} | Y_{0,t} \cup X_t = e_i].\]
(4.56)

Omitting further calculations, it can be shown [14] that
\[r_t = \langle q_t, e_1 \rangle v_{t,T} | v_{t,T} , e_1 \rangle e_1 + \langle q_t, e_2 \rangle v_{t,T} | v_{t,T} , e_2 \rangle e_2 + \cdots + \langle q_t, e_m \rangle v_{t,T} | v_{t,T} , e_m \rangle e_m \in \mathbb{R}^n.\]
(4.57)

The normalized smoothed-state estimate of \(X\) is then
\[E[X_t | Y_{0,t}] = \langle r_t, 1 \rangle.\]
(4.58)
Note that
\[ \langle r_t, 1 \rangle = \langle q_t, v_t, T \rangle \]
\[ = E'[\Lambda_0, T (X_t, 1)|Y_0, T] \]
\[ = E'[\Lambda_0, T |Y_0, T] \]  
(4.59)

is independent of \( t \). Therefore

\[
\frac{d}{dt} \langle r_t, 1 \rangle = \frac{d}{dt} \langle q_t, v_t, T \rangle \\
= \langle \partial q_t, v_t, T \rangle + \langle q_t, d\overline{r}_t, T \rangle \\
= \langle \Gamma_t^{-1} A\Gamma_t \overline{q}_t, v_t, T \rangle + \langle \overline{q}_t, d\overline{r}_t, T \rangle \\
= 0. 
\]  
(4.60)

The vector \( v_{t,T} = \langle v_{t,T}, e_1 \rangle, v_{t,T}, e_2 \rangle, \ldots, v_{t,T}, e_n \rangle \) 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

\[ \langle \overline{q}_t, \overline{v}_t, T \rangle = \langle \Gamma_t^{-1} \overline{q}_t, \Gamma_t v_t, T \rangle = \langle q_t, v_t, T \rangle, \]

for all \( t \in [0, T] \).  
(4.61)

That is, \( \overline{v}_{t,T} \overset{\Delta}{=} \Gamma_t v_t, T \). Using (4.60), one can show that

\[
\frac{d\overline{v}_{t,T}}{dt} = -\Gamma_t A\Gamma_t^{-1} \overline{v}_t, T, \quad \overline{v}_{t,T} = v_{t,T} = 1 
\]  
(4.62)

so that

\[ \overline{v}_{t,T} = 1 + \int_t^T \Gamma_u A\Gamma_u^{-1} \overline{v}_{u,T} du. \]  
(4.63)

Further, using the time discretization of (3.40)

\[ \overline{v}_{k-1,T} = \overline{v}_k, T + \int_{t_{k-1}}^{t_k} \Gamma_u A\Gamma_u^{-1} \overline{v}_{u,T} du \\
\approx \overline{v}_k, T + \Gamma_k A\Gamma_k^{-1} \overline{v}_{k,T} \Delta(t_{k-1:k}), \]  
(4.64)

so, our suboptimal estimator \( \overline{v} \approx v \) has dynamics

\[ \overline{v}_{k-1,T} \overset{\Delta}{=} \Gamma_{k-1} \Gamma_k \left[ I + \Delta(t_{k-1:k}) A \right] \overline{v}_k, T. \]  
(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

\[ \langle \sigma(G_t^i X_t), v_{t,T} \rangle = \langle \sigma_t, 1 \rangle \sigma(G_t^i X_t), \Gamma v_{t,T} \rangle \]
\[ = \langle \overline{\sigma}(G_t^i X_t), \overline{v}_{t,T} \rangle. \]  
(4.66)

Now, define

\[ \overline{\sigma}(G_t^i X_t) \overset{\Delta}{=} \sigma(G_t^i X_t) - \langle \overline{q}_t, e_i \rangle N_t e_i. \]  
(4.67)

Then

\[
d\overline{\sigma}(G_t^i X_t) = \Gamma_t^{-1} A\Gamma_t \overline{\sigma}(G_t^i X_t) dt \\
- N_t \langle \Gamma_t^{-1} A\Gamma_t \overline{q}_t, e_i \rangle e_i dt. 
\]  
(4.68)

Now,

\[ \langle \overline{\sigma}(G_t^i X_t), \overline{v}_{t,T} \rangle = \langle \overline{\sigma}(G_t^i X_t), \overline{v}_{t,T} \rangle + N_t \langle \overline{q}_t, e_i \rangle \langle v_{t,T}, e_i \rangle \]  
(4.69)

and

\[ N_t \langle \overline{q}_t, e_i \rangle \langle v_{t,T}, e_i \rangle = N_t \langle \overline{q}_t, e_i \rangle \langle v_{t,T}, e_i \rangle. \]  
(4.70)

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

\[
d\langle \overline{\sigma}(G_t^i X_t), \overline{v}_{t,T} \rangle \\
= \langle \Gamma_t^{-1} A\Gamma_t \overline{\sigma}(G_t^i X_t), \overline{v}_{t,T} \rangle dt \\
- N_t \langle \Gamma_t^{-1} A\Gamma_t \overline{q}_t, e_i \rangle \langle e_i, \overline{v}_{t,T} \rangle dt \\
- \langle \overline{\sigma}(G_t^i X_t), \Gamma_t \Gamma_t^{-1} \overline{v}_{t,T} \rangle dt \\
= \langle \Gamma_t^{-1} A\Gamma_t \overline{\sigma}(G_t^i X_t), \overline{v}_{t,T} \rangle dt \\
- N_t \langle \Gamma_t^{-1} A\Gamma_t \overline{q}_t, e_i \rangle \langle e_i, \overline{v}_{t,T} \rangle dt \\
- \langle \overline{\sigma}(G_t^i X_t) - \langle \overline{q}_t, e_i \rangle, N_t \langle e_i, \Gamma_t \overline{v}_{t,T} \rangle \rangle dt \\
= -N_t \langle \Gamma_t^{-1} A\Gamma_t \overline{q}_t, e_i \rangle \langle e_i, \overline{v}_{t,T} \rangle dt \\
+ N_t \langle \overline{q}_t, e_i \rangle \langle \Gamma_t^{-1} A\Gamma_t e_i, \overline{v}_{t,T} \rangle dt 
\]  
(4.71)

i.e.,

\[ \langle \overline{\sigma}(G_t^i X_t), \overline{v}_{t,T} \rangle = \langle \overline{\sigma}(G_t^i X_t), \overline{v}_{t,T} \rangle \]
\[ = -\int_0^T N_u \langle \Gamma_u^{-1} A\Gamma_u \overline{q}_u, e_i \rangle \langle e_i, \overline{v}_{u,T} \rangle du \\
+ \int_0^T N_u \langle \overline{q}_u, e_i \rangle \langle \Gamma_u^{-1} A\Gamma_u e_i, \overline{v}_{u,T} \rangle du. \]  
(4.72)

Therefore,

\[ \langle \sigma(G_t^i X_t), v_{t,T} \rangle = \langle \sigma(G_t^i X_t), \overline{v}_{t,T} \rangle \\
= \langle \sigma(G_t^i X_t), v_{t,T} \rangle + N_T \langle \overline{q}_T, e_i \rangle \langle v_{t,T}, e_i \rangle \\
= -\int_0^T N_u \langle \Gamma_u^{-1} A\Gamma_u \overline{q}_u, e_i \rangle \langle e_u, \overline{v}_{u,T} \rangle du \\
+ \int_0^T N_u \langle \overline{q}_u, e_i \rangle \langle \Gamma_u^{-1} A\Gamma_u e_i, \overline{v}_{u,T} \rangle du \\
+ N_T \langle \overline{q}_T, e_i \rangle \langle v_{t,T}, e_i \rangle. \]  
(4.73)

By using similar calculations, one can also show that

\[ \langle \overline{\sigma}(J_t^i X_t), \overline{v}_{t,T} \rangle = \int_0^T \langle q_u, e_i \rangle \langle v_{u,T}, e_i \rangle du. \]  
(4.74)

and

\[ \langle \overline{\sigma}(N_t^{(j,i)} X_t), \overline{v}_{t,T} \rangle = \int_0^T \langle A e_i, e_j \rangle \langle q_u, e_i \rangle \langle v_{u,T}, e_j \rangle du. \]  
(4.75)
C. Smoother-Based EM Algorithm

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

\[
\tilde{A}_{t+1} = [A_t]_{(i,j)} = \frac{\int_0^T \langle q_u, e_i \rangle \langle v_{u,T}, e_j \rangle du}{\int_0^T \langle q_u, e_i \rangle \langle v_{u,T}, e_j \rangle du} \tag{4.76}
\]

and

\[
\langle \tilde{A}_{t+1}, e_i \rangle = \frac{\int_0^T N_u \langle q_u, e_i \rangle \langle e_i, A_{v_{u,T}} \rangle du}{\int_0^T \langle q_u, e_i \rangle \langle v_{u,T}, e_i \rangle du} - \frac{\int_0^T N_u \langle A_{q_u}, e_i \rangle \langle e_i, v_{u,T} \rangle du}{\int_0^T \langle q_u, e_i \rangle \langle v_{u,T}, e_i \rangle du} + \frac{N_T \langle q_T, e_i \rangle \langle v_T, e_i \rangle}{\int_0^T \langle q_u, e_i \rangle \langle v_{u,T}, e_i \rangle du}. \tag{4.77}
\]

V. DISCRETE-TIME SMOOHERS

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] ∪ [T, T’], that is,

\[0 = t_0 < t_1 < \ldots < t_K = T < t_0’ < t_1’ < \ldots < t_K’ = T’.\]

Here, we denote this augmented partition by \(\Pi_{[0,T]}^{(K)} \cup \Pi_{[T,T’]}^{(K)}\), where \(K \in \mathbb{N}\) and

\[\Pi_{[T,T’]}^{(K)} = \{T = t_0’, t_1’, \ldots, t_K’ = T’\}. \tag{5.78}\]

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

\[
\hat{v}_{k-1,T} = \Gamma_{k-1}^{-1} \Gamma_k \left[ I + \Delta(k-1)A \right] \Gamma_{k-1}^{-1} \Gamma_k \hat{v}_{k,T} = \Gamma_{k-1}^{-1} \Gamma_k \left[ I + \Delta(k-1)A \right] \Gamma_{k-1}^{-1} \Gamma_k \left[ I + \Delta(k+1)A \right] \hat{v}_{k,T}, \tag{5.79}\]

Recall here that \(K = t_K = T\).

Write

\[
\Psi_{k-1,T} = \Gamma_{k-1}^{-1} \Gamma_k \left[ I + \Delta(k-1)A \right] \Gamma_{k-1}^{-1} \Gamma_k \left[ I + \Delta(k+1)A \right] \in \mathbb{R}^{n \times n}. \tag{5.80}\]

Further, for two epochs \(T\) and \(T’\), where \(T < T’\), it follows that

\[
\Psi_{k-1,T} \hat{v}_{k-1,T} = \Psi_{k-1,T} \hat{v}_{T,T’}, \quad k \in \{0, 1, 2, \ldots, T\}. \tag{5.81}\]

At a boundary \(T’\), we set \(\Psi_{T,T’} = \Delta \text{diag} \{1, 1, \ldots, 1\} \in \mathbb{R}^{n \times n}\).

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

Using the matrix \(\Psi\), the backward recursion for \(v_{k-1,T}\) may be written in the following compact form:

\[
v_{k-1,T} = \Psi_{k-1,T} 1, \quad 1 = (1, 1, \ldots, 1)’ \in \mathbb{R}^n. \tag{5.82}\]

Equation (5.82) and the transitivity property of \(\Psi\) 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 \(\Psi\) 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 \(1\). 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

\[
v_{k-1,T’} = \Psi_{k-1,T} T’ 1, \quad 1 = (1, 1, \ldots, 1)’ \in \mathbb{R}^n. \tag{5.83}\]

Consider, for example, the following smoothing problem. Suppose one first observes data on [0, T] and computes the smoothed estimate \(P(X_t = e_i | Y_{0,T})\) for some \(t \in [0, T]\). Using \(\Psi_T\), this estimation can be written as

\[
P(X_t = e_i | Y_{0,T}) = \langle q_t, e_i \rangle \langle v_{T,T}, e_i \rangle \sum_{t=1}^n \langle q_t, e_i \rangle \langle v_{T,T}, e_i \rangle = \frac{\langle q_t, e_i \rangle \langle e_i | \Psi_T \rangle \langle e_i | \Psi_T, T’T \rangle}{\sum_{t=1}^n \langle q_t, e_i \rangle \langle e_i | \Psi_T \rangle \langle e_i | \Psi_T, T’T \rangle}. \tag{5.84}\]

Now, suppose subsequent data are received on [T, T’] and we wish to compute \(P(X_t = e_i | Y_{0,T})\). Using \(\Psi_T, T’T\), this estimate may be computed by

\[
P(X_t = e_i | Y_{0,T}) = \langle q_t, e_i \rangle \langle v_{T,T}, e_i \rangle \sum_{t=1}^n \langle q_t, e_i \rangle \langle v_{T,T}, e_t \rangle = \frac{\langle q_t, e_i \rangle \langle e_i | \Psi_T \rangle \langle e_i | \Psi_T, T’T \rangle}{\sum_{t=1}^n \langle q_t, e_i \rangle \langle e_i | \Psi_T \rangle \langle e_i | \Psi_T, T’T \rangle}. \tag{5.85}\]

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 expressions by the Trapezoidal rule. These approximations can also be written in a data-recursive form. To approximate \(\tilde{A}_{t+1} = [A_t]_{(i,j)}\) on the interval [0, T], we write (5.86), as shown at the bottom of 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
The update formulas incorporating the information on $[T, T']$ in the estimates (4.76) and (4.77) are, respectively,

$$[\hat{A}_{r+1}]_{(i,j)} = [\hat{A}_r]_{(i,j)}$$

and

$$\langle \hat{\lambda}_{r+1}, e_i \rangle = \left[ \sum_{t=1}^{K} \frac{1}{2} \Delta_{(t-1,t)} (N_{t-1} \langle \hat{q}_{t-1}, e_i \rangle (e_i' A' \Psi_{t-1,T}) + N_t (\hat{q}_t, e_i) (e_i' A' \Psi_{t-1,T}) \right] 1$$

$$+ \left( \sum_{t=K+1}^{K} \frac{1}{2} \Delta_{(t-1,t)} (\hat{q}_{t-1}, e_i) (e_i' \Psi_{t-1,T}) + N_t (\hat{q}_t, e_i) (e_i' \Psi_{t-1,T}) \right) 1$$

$$(5.88)$$

$$\langle \hat{\lambda}_{r+1}, e_i \rangle = \left[ \sum_{t=1}^{K} \frac{1}{2} \Delta_{(t-1,t)} (N_{t-1} \langle \hat{q}_{t-1}, e_i \rangle (e_i' A' \Psi_{t-1,T}) + N_t (\hat{q}_t, e_i) (e_i' A' \Psi_{t-1,T}) \right] 1$$

$$+ \left( \sum_{t=K+1}^{K} \frac{1}{2} \Delta_{(t-1,t)} (\hat{q}_{t-1}, e_i) (e_i' \Psi_{t-1,T}) + N_t (\hat{q}_t, e_i) (e_i' \Psi_{t-1,T}) \right) 1$$

$$(5.89)$$

$$\langle \hat{\lambda}_{r+1}, e_i \rangle = \left[ \sum_{t=1}^{K} \frac{1}{2} \Delta_{(t-1,t)} (N_{t-1} \langle \hat{q}_{t-1}, e_i \rangle (e_i' A' \Psi_{t-1,T}) + N_t (\hat{q}_t, e_i) (e_i' A' \Psi_{t-1,T}) \right] 1$$

$$+ \left( \sum_{t=K+1}^{K} \frac{1}{2} \Delta_{(t-1,t)} (\hat{q}_{t-1}, e_i) (e_i' \Psi_{t-1,T}) + N_t (\hat{q}_t, e_i) (e_i' \Psi_{t-1,T}) \right) 1$$

$$(5.90)$$
To complete the proof, we note that satisfies the Markov property and that its rate matrix is again of Bayes’ rule, we see that, for

\[ E[X_t|Y_{0,t}] = E^*[\Lambda_{0,t}X_t|Y_{0,t}] \]  
\[ E[X_{t+1}|Y_{0,t+1}] = E^*[\Lambda_{0,t+1}X_{t+1}|Y_{0,t+1}] \]


\[ E[X_{t+1}|Y_{0,t+1}] = E^*[\Lambda_{0,t+1}X_{t+1}|Y_{0,t+1}] \]  
\[ E[X_{t+1}|Y_{0,t+1}] = E^*[\Lambda_{0,t+1}X_{t+1}|Y_{0,t+1}] \]

Therefore,

\[ E\left[M_t - M_s|\mathcal{F}_s\right] = 0 \in \mathbb{R}_+^n. \]  
\[ E\left[M_t - M_s|\mathcal{F}_s\right] = 0 \in \mathbb{R}_+^n. \]

So,

\[ E\left[X_t - X_s - \int_s^t AX_u du|\mathcal{F}_s\right] = 0 \in \mathbb{R}_+^n. \]  
\[ E\left[X_t - X_s - \int_s^t AX_u du|\mathcal{F}_s\right] = 0 \in \mathbb{R}_+^n. \]

Then,

\[ Z_t \triangleq E\left[X_t|\mathcal{F}_s\right] = X_s + \int_s^t AE\left[X_u|\mathcal{F}_s\right] du \]  
\[ Z_t \triangleq E\left[X_t|\mathcal{F}_s\right] = X_s + \int_s^t AZ_u du. \]

Therefore,

\[ Z_t = \exp\left(A(t - s)\right)X_s = E\left[X_t|\mathcal{F}_s\right] = E\left[X_t|X_s\right]. \]  
\[ Z_t = \exp\left(A(t - s)\right)X_s = E\left[X_t|\mathcal{F}_s\right] = E\left[X_t|X_s\right]. \]

Equation (A6) shows that, under the measure \( P^1 \), the process \( X \) satisfies the Markov property and that its rate matrix is again \( A \).

To complete the proof, we note that

\[ E\left[X_0|\{\Omega, \emptyset\}\right] = \frac{E^*[\Lambda_{0}^{-1}X_0|\{\Omega, \emptyset\}]}{E^*[\Lambda_{0}^{-1}|\{\Omega, \emptyset\}]} = E\left[X_0\right] = p_0. \]  
\[ E\left[X_0|\{\Omega, \emptyset\}\right] = \frac{E^*[\Lambda_{0}^{-1}X_0|\{\Omega, \emptyset\}]}{E^*[\Lambda_{0}^{-1}|\{\Omega, \emptyset\}]} = E\left[X_0\right] = p_0. \]
The result follows by conditioning both sides of (C1) on \( \gamma_{0,t} \), using a version of Fubini’s theorem [27] and noting that, under the measure \( P^i \), the process \( N \) is a standard Poisson process. Consequently we see that

\[
E^\dagger[E^\dagger[G^i_t X_i|\gamma_{0,t}]] = \int_0^t E^\dagger[E^\dagger[\Lambda_u - X_u (X_u, e_t)]|\gamma_{0,s}] dN_u
\]

\[
+ \int_0^t E^\dagger[E^\dagger[\Lambda_u G^i_t AX_u|\gamma_{0,s}] dM_u
\]

\[
+ \int_0^t E^\dagger[E^\dagger[G^i_t X_u \Lambda_u - (X_u, \lambda) - 1]|\gamma_{0,s}] (dN_u - dU)
\]

\[
+ \int_0^t E^\dagger[X_u (X_u, e_t)] (X_u, \lambda) - 1]|\gamma_{0,s}] (dN_u - dU)
\]

\[
= \int_0^t A \sigma (G^i_t X_u) dU
\]

\[
+ \int_0^t \text{diag} \{\lambda, e_t\} (X_u, \lambda - 1) \sigma (G^i_t X_u) (dN_u - dU)
\]

\[
+ \int_0^t \{q_u, e_t\} (X_u, \lambda) dN_u, e_t.
\]

(C2)

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


Robert J. Elliott received the Bachelor’s and Master’s degrees from the University of Oxford, 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 also 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 Flanders 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.