Robust filtering for the Markov modulated Poisson Process

James Ledoux

INSA & IRMAR
20 avenue des Buttes de Coësmes, CS 14315,
35043 Rennes Cedex, France
(e-mail: james.ledoux@insa-rennes.fr*)*

Abstract

In this note, we deal with filters needed in the parameter estimation of the so-called Markov modulated Poisson process by the EM-algorithm. Specifically, using a Clark's transform, a Skorokhod-continuous version of the unnormalized filter of various statistics of Markov Modulated Poisson Processes is given.

Key words: Hidden Markov process, Zakai's filter

1 Introduction

In this note, we focus on a class of Markovian models known as the Markov Modulated Poisson Process (MMPP). An MMPP is formally defined as a bivariate Markov process $\{(N_t, X_t)\}_{t\geq 0}$ such that

- $\{N_t\}_{t\geq 0}$ is a counting process of "events";
- $\{X_t\}_{t\geq 0}$ is a Markov process with a finite state space, say $\{1, \ldots, n\}$, and generator Q;
- listing the state space $\mathbb{N} \times \{1, \ldots, n\}$ in lexicographic order, the generator of $\{(N_t, X_t)\}_{t\geq 0}$ has the form

$$A = \begin{pmatrix} Q - \Lambda & \mathbf{0} & \cdots \\ \Lambda & Q - \Lambda & \ddots \\ \mathbf{0} & \Lambda & \ddots \\ \vdots & \ddots & \ddots \end{pmatrix} \tag{1.1}$$

where Λ is a diagonal matrix with ith diagonal entry $\lambda(i)$.

An MMPP is an instance of a Markovian Arrival Process as popularized by Neuts (e.g. see [12]). It is well-known that the intensity of the counting process $\{N_t\}_{t\geq 0}$ (wrt to the filtration generated by $\{N_t, X_t\}_{t\geq 0}$) is

$$\lambda(X_t).$$

In this model, the events occur according to a Poisson process with intensity $\lambda(i)$ during a sojourn time of $\{X_t\}_{t\geq 0}$ in state i. Thus, an MMPP may be thought of as a Poisson process with an intensity driven by the Markov process $\{X_t\}_{t\geq 0}$ (e.g. see [16]). In statistical literature, it is known as an instance of a Hidden Markov Process because $\{N_t\}_{t\geq 0}$ has conditionally independent increments given the Markov process $\{X_t\}_{t\geq 0}$ [2].

The MMPPs have gained widespread use in stochastic modeling, for example in engineering [16], in computer science [13, and the reference therein], in finance (see [11, and the reference therein]. Our initial motivation for dealing with MMPP originates in the software reliability modeling using an "architecture-based" approach (see [6]). Thus, in this context, $\{N_t\}_{t\geq 0}$ is the counting process of failures and $\{X_t\}_{t\geq 0}$ is interpreted to be a Markovian model of the flow of control between the modules of a software.

Since an MMPP is specified by the matrices Q, Λ, the fitting of MMPPs to data requires the estimation of the non-negative parameter vector

$$\theta = \{\lambda(i), \quad Q(j,i) \quad i,j = 1,\ldots,n\}. \tag{1.2}$$

The only available data is the observation of events. In that perspective, the process $\{(N_t, X_t)\}_{t\geq 0}$ can be thought of as a partially observed Markov process or a hidden Markov process. The observed process is the counting process of events $\{N_t\}_{t\geq 0}$ and the state or hidden process is the finite Markov process $\{X_t\}_{t\geq 0}$. The EM-algorithm is a standard way to estimate the parameters of hidden Markov processes and has been used for MMPPs [15,14, and references therein]. All these works use the standard forward-backward principle which is based on data processing in batch. Due to the backward pass through the data, the storage cost is linear in the number of observations. Elliott has proposed a filter-based EM-algorithm where the standard forward-backward form of the E-step of the algorithm is replaced by a single forward pass procedure [4]. The main advantages are that "on-line" estimation is allowed and the storage cost does not depend on the number of observations so that very large data sets can be processed. To implement the filter-based approach, finite-dimensional recursive unnormalized or normalized filters for various statistics related to MMPPs are used [5,8]. The aim of this note is to provide finite-dimensional filters which are robust in the sense of [5]. That is, they are continuous functionals of the observations , i.e. of paths of the counting process $\{N_t\}_{t\geq 0}$. Our work is a natural continuation of that in [10,5]. Our results are counterparts for MMPPs of those obtained in [3,7] for a finite Markov process observed through an independent Brownian noise.

The paper is organized as follows. In Section 2, the basic material on the statistics needed in the estimation of θ by the EM-algorithm is introduced. Linear stochastic differential equations satisfied by the so-called Zakai filters are also recalled. In Section 3, the Clark's transform is introduced and a continuous-Skorokhod version of the unnormalized filters of MMPPs is derived.

Main notation and convention

- Vectors are column vectors. The ith component of any vector v is denoted by $v(i)$. For any $u, v \in \mathbb{R}^n$, $\langle u, v\rangle = \sum_i u(i)v(i)$, is the usual scalar product in $\mathbb{R}^n$. $\mathbf{1}$ is a n-dimensional vector with each entry equals to one.
- For any right-continuous with left-hand limits (rcll) function $t \mapsto f_t$, the left-hand limit at t of f is denoted by f_{t-} and $\Delta f_t := f_t - f_{t-}$ for $t > 0$ is the jump of the function at time t. We set $\Delta f_0 := f_0$.
- The state space of the Markov process $\{X_t\}_{t\geq 0}$ is assumed to be $\mathscr{X} := \{e_i, i = 1, \ldots, n\}$, where e_i is the ith vector of the canonical basis of $\mathbb{R}^n$. With this convention, the indicator function $1_{\{X_t=e_i\}}$ of the set $\{X_t = e_i\}$ is the ith component of vector X_t, that is $\langle X_t, e_i\rangle$. In other words, for any time t, the vector X_t is the n-dimensional vector $X_t = (1_{\{X_t=e_i\}})_{i=1}^n$. Therefore, the sum of the components of vector X_t, $\langle \mathbf{1}, X_t\rangle$, is equal to 1.
- All processes are assumed to be defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The internal filtrations of processes $\{N_t\}_{t\geq 0}$ and $\{(N_t, X_t)\}_{t\geq 0}$ are denoted by $\mathbb{F}^N := \{\mathbb{F}^N_t\}_{t\geq 0}$ and $\mathbb{F} := \{\mathbb{F}_t\}_{t\geq 0}$ respectively, where $\mathbb{F}^N_t := \sigma(N_s,\ s \leq t)$ and $\mathbb{F}_t := \sigma(N_s, X_s\ s \leq t)$. These filtrations are assumed to be complete, that is $\mathbb{F}_0, \mathbb{F}^N_0$ contain all the sets of $\mathbb{P}$-probability zero of $\mathcal{F}$.
- For any integrable $\mathbb{F}$-adapted random process $\{Z_t\}_{t\geq 0}$, the conditional expectation $\mathbb{E}[Z_t \mid \mathbb{F}^N_t]$ is denoted by $\widehat{Z}_t$ and $\{\widehat{Z}_t\}_{t\geq 0}$ is called the *filter* associated with the process $\{Z_t\}_{t\geq 0}$.

2 Basic material on the observed/hidden processes

In this note, we are concerned with the following basic statistics of an MMPP

$$\begin{aligned}
N^{X,ji}_t &:= \sum_{0<s\leq t} (1 - \Delta N_s)\langle X_s, e_j\rangle\langle X_{s-}, e_i\rangle \quad j \neq i \\
\mathcal{O}^{(i)}_t &:= \int_0^t \langle X_{s-}, e_i\rangle \,\mathrm{d}s \qquad (2.1)\\
G^i_t &:= \sum_{0<s\leq t} \Delta N_s\langle X_s, e_j\rangle\langle X_{s-}, e_i\rangle = \int_0^t \langle X_s, e_j\rangle\langle X_{s-}, e_i\rangle \,\mathrm{d}N_s.
\end{aligned}$$

The first statistic is the number of jumps of $\{X_t\}_{t\geq 0}$ from state e_i to state e_j $(i \neq j)$ over the interval $]0,t]$. The second statistic is the sojourn time of $\{X_t\}_{t\geq 0}$ in the state e_i in the interval $[0,t]$. G^i_t is the number of events observed during the sojourn of $\{X_t\}_{t\geq 0}$ in state e_i over the interval $[0,t]$. This family of statistics is introduced because their filters appear in the re-estimation formulas used for estimating the parameter $\theta := \{\lambda(i), Q(j,i), i,j = 1,\ldots,n\}$ of an MMPP with the EM-algorithm (e.g. see [15]): for $i,j = 1,\ldots,n$

$$\widetilde{\lambda}(i) = \frac{\widehat{G^i}_t}{\widehat{\mathcal{O}^{(i)}}_t} \quad \text{and} \quad \widetilde{Q}(j,i) = \frac{\widehat{N^{X,ji}}_t}{\widehat{\mathcal{O}^{(i)}}_t} \quad \text{with } i \neq j. \tag{2.2}$$

.

The basic material below is found in [1] for instance. We report here a $\mathbb{F}$-semi-martingale (or Doob-Meyer here) representation of the counting processes $\{N_t\}_{t\geq 0}$, $\{N^{ji}_t\}_{t\geq 0}$, $\{G^i_t\}_{t\geq 0}$. It is clear from their definition that $\{N_t\}_{t\geq 0}$, $\{N^{ji}_t\}_{t\geq 0}$, $\{G^i_t\}_{t\geq 0}$ are counters of specific transitions in $\{(N_t, X_t)\}_{t\geq 0}$. Then, it is easily deduced from the special structure of A (see (1.1)) that the $\mathbb{F}$-semi-martingale decomposition of the counting processes above are

$$N^{X,ji}_t := \int_0^t Q(j,i)\langle X_{s-}, e_i\rangle \,\mathrm{d}s + \mathcal{M}^{N^{X,ji}}_t \tag{2.3}$$

$$G^i_t = \int_0^t \lambda(i)\langle X_{s-}, e_i\rangle \,\mathrm{d}s + \mathcal{M}^{G^i}_t \quad k = 0,1 \tag{2.4}$$

$$N_t = \int_0^t \lambda_s \,\mathrm{d}s + \mathcal{M}_t \quad \text{with } \lambda_s := \langle \mathbf{1}, \Lambda X_{s-}\rangle \tag{2.5}$$

where $\{\mathcal{M}^{N^{X,ji}}\}_{t\geq 0}, \{\mathcal{M}^{G^i}\}_{t\geq 0}, \{\mathcal{M}\}_{t\geq 0}$ are $\mathbb{F}$-martingales. The process $\{\lambda_t\}_{t\geq 0}$ is the so-called *stochastic intensity* of $\{N_t\}_{t\geq 0}$ with respect to the filtration $\mathbb{F}$.

The basic $\mathbb{F}$-semi-martingale decomposition of the Markov process $\{X_t\}_{t\geq 0}$ is

$$X_t = X_0 + \int_0^t QX_{s-}\,\mathrm{d}s + \mathcal{M}^X_t \tag{2.6}$$

where $\{\mathcal{M}^X_t\}_{t\geq 0}$ is a $\mathbb{F}$-martingale (and a martingale with respect to the internal filtration of $\{X_t\}_{t\geq 0}$ as well). We recognize in (2.5) and (2.6) a standard representation of a continuous-time hidden Markov process, with $\{X_t\}_{t\geq 0}$ as state process and $\{N_t\}_{t\geq 0}$ as observed process. The observation and state "noises" $\{\mathcal{M}_t\}_{t\geq 0}, \{\mathcal{M}^X_t\}_{t\geq 0}$ have no common jumps:

$$\sum_{0<s\leq t} \Delta\mathcal{M}_s \Delta\mathcal{M}^X_s = \sum_{0<s\leq t} \Delta N_s \Delta X_s = 0.$$

The following facts will be used throughout the paper.

- Since any rcll process $\{Z_t\}_{t\geq 0}$ has finitely many jumps on finite intervals for almost all $\omega \in \Omega$, the following standard Lebesgue integrals

$$\int_0^t Z_s \mathrm{d}s \quad \text{or} \quad \int_0^t Z_{s-} \mathrm{d}s$$

are interchangeably used.
- (PR) The stochastic integrals in this paper are Lebesgue-Stieltjes integrals. We report here the product rule for two rcll processes $\{Z_t^{(1)}\}_{t\geq 0}, \{Z_t^{(2)}\}_{t\geq 0}$ having finite variations on the bounded intervals, *i.e.* having locally finite variations,

$$Z_t^{(1)} Z_t^{(2)} = Z_0^{(1)} Z_0^{(2)} + \int_0^t Z_{s-}^{(1)} \mathrm{d}Z_s^{(2)} + \int_0^t Z_{s-}^{(1)} \mathrm{d}Z_s^{(2)} + \sum_{0<s\leq t} \Delta Z_s^{(1)} \Delta Z_s^{(2)}.$$

Zakai filters

Now, we discuss the so-called Zakai filters associated with the statistics in (2.1). Their derivation is based on a technique of change of measure. We recall some basic facts which are borrowed from [1, Chap VI] for instance. Assume that

$$\lambda(i) > 0, \quad i = 1, \ldots, n. \tag{2.7}$$

Then, we have

$$0 < \min_i \lambda(i) \leq \lambda_t = \langle \mathbf{1}, \Lambda X_{t-} \rangle \leq \sum_i \lambda(i)$$

so that

$$\frac{1}{\sum_i \lambda(i)} \leq \frac{1}{\lambda_t} \leq \frac{1}{\min_i \lambda(i)}. \tag{2.8}$$

Let us consider the likelihood ratio

$$\overline{L}_t := \exp\Big(\int_0^t \ln \frac{1}{\lambda_s} \mathrm{d}N_s + \int_0^t (\lambda_s - 1)\, \mathrm{d}s\Big) = \prod_{0<s\leq t} (\frac{1}{\lambda_s})^{\Delta N_s} \exp\Big(\int_0^t (\lambda_s - 1)\, \mathrm{d}s\Big)$$

which is a solution of the equation

$$\overline{L}_t = 1 + \int_0^t \overline{L}_{s-} (\frac{1}{\lambda_s} - 1)\, (\mathrm{d}N_s - \lambda_s \mathrm{d}s).$$

Then, $\{\overline{L}_t\}_{t\geq 0}$ is a $(\mathbb{P}, \mathbb{F})$-martingale. A new probability measure $\mathbb{P}_0$ on $(\Omega, \vee_{t\geq 0} \mathbb{F}_t)$ is defined by

$$\frac{d\mathbb{P}_0}{d\mathbb{P}}\Big|_{\mathbb{F}_t} := \overline{L}_t.$$

It results from Girsanov's theorem that, under $\mathbb{P}_0$, $\{N_t\}_{t\geq 0}$ is a $\mathbb{F}$-homogeneous Poisson process with intensity $\overline{\lambda}_s \equiv 1$. Therefore, the process $\{n_t\}_{t\geq 0}$ defined by

$$n_t := N_t - t$$

is a $(\mathbb{P}_0, \mathbb{F})$-martingale and must be thought of as the innovation martingale associated with $\{N_t\}_{t\geq 0}$ under the new probability measure $\mathbb{P}_0$.

Next, set

$$L_t := \frac{1}{\overline{L}_t}.$$

$\{L_t\}_{t\geq 0}$ is a solution of the equation

$$L_t = 1 + \int_0^t L_{s-}(\lambda_s - 1)\,\mathrm{d}n_s \tag{2.9}$$

and is a $(\mathbb{P}_0, \mathbb{F})$-martingale. We know that $\mathbb{P} \ll \mathbb{P}_0$ and

$$\left.\frac{d\mathbb{P}}{d\mathbb{P}_0}\right|_{\mathbb{F}_t} = L_t.$$

Under the mild condition stated at the beginning of this subsection, $\mathbb{P}$ and $\mathbb{P}_0$ are equivalent probability measures. The expectation with respect to $\mathbb{P}_0$ is denoted by $\mathbb{E}_0[\cdot]$. Any $\mathbb{F}^N$-conditional expectation, under $\mathbb{P}$, of an integrable $\mathbb{F}$-measurable random variable, may be computed from Bayes formula (e.g. see [4])

$$\mathbb{E}[Z \mid \mathbb{F}_t^N] = \frac{\mathbb{E}_0[ZL_t \mid \mathbb{F}_t^N]}{\mathbb{E}_0[L_t \mid \mathbb{F}_t^N]} \quad \mathbb{P}\text{ a.s.} \tag{2.10}$$

The above equation is the basic one for deriving the Zakai filters $\mathbb{E}_0[ZL_t \mid \mathbb{F}_t^N]$.

For any $\mathbb{F}$-adapted integrable process $\{Z_t\}_{t\geq 0}$, $\sigma(Z_t)$ denotes the conditional expectation $\mathbb{E}_0[Z_t L_t \mid \mathbb{F}_t^N]$. Then, we known from [8] that: for any $t \geq 0$

$$\sigma(X_t) = \widehat{X}_0 + \int_0^t Q\sigma(X_{s-})\,\mathrm{d}s + \int_0^t (\Lambda - I)\sigma(X_{s-})\,\mathrm{d}n_s. \tag{2.11a}$$

$$\begin{aligned}\sigma(N_t^{X,ji}X_t) = &\int_0^t \Big[Q\sigma(N_s^{X,ji}X_s) + Q(j,i)\langle\sigma(X_s), e_i\rangle\, e_j\Big]\,\mathrm{d}s \\ &+ \int_0^t \Big((\Lambda - I)\sigma(N_{s-}^{X,ji}X_{s-}) + \Lambda(j,i)\langle\sigma(X_{s-}), e_i\rangle\, e_j\Big)\,\mathrm{d}n_s.\end{aligned} \tag{2.11b}$$

$$\begin{aligned}\sigma(\mathcal{O}_t^{(i)}X_t) = &\int_0^t \Big[Q\sigma(\mathcal{O}_{s-}^{(i)}X_{s-}) + \langle\sigma(X_{s-}), e_i\rangle\, e_i\Big]\,\mathrm{d}s \\ &+ \int_0^t (\Lambda - I)\sigma(\mathcal{O}_{s-}^{(i)}X_{s-})\,\mathrm{d}n_s.\end{aligned} \tag{2.11c}$$

$$\begin{aligned}\sigma(G_t^i X_t) = &\int_0^t Q\sigma(G_{s-}^i X_{s-})\,\mathrm{d}s + \int_0^t \lambda(i)\langle\sigma(X_{s-}), e_i\rangle\, e_i \mathrm{d}N_s \\ &+ \int_0^t (\Lambda - I)\sigma(G_{s-}^i X_{s-})\,\mathrm{d}n_s.\end{aligned} \tag{2.11d}$$

These SDEs are standard linear ordinary differential equations (ODE) between two jumps of $\{N_t\}_{t\geq 0}$. Therefore, a basic way to deal with the equations (2.11a-2.11d) is to integrate the linear ODEs over the interval of time between two jumps and to update the solution at the endpoint of the interval. For instance, the state filter $\sigma(X_t)$ is solution of

$$\frac{\mathrm{d}}{\mathrm{d}t}\, q_t = (Q - \Lambda + I)q_t = (\Lambda + I)q_t, \quad q_{t_{l-1}} := \sigma(X_{t_{l-1}})$$

in the interval $[t_{l-1}, t_l[$, where $(t_l)_{l\in\mathbb{N}}$ is the sequence of times of jump of $\{N_t\}_{t\geq 0}$ $(t_0 := 0)$. Then, the solution at time of jump t_l is updated as follows

$$\sigma(X_{t_l}) - \sigma(X_{t_l -}) = (\Lambda - I)\sigma(X_{t_l -}) \Longrightarrow \sigma(X_{t_l}) = \Lambda\sigma(X_{t_l -}).$$

In this special case, it is easily seen that (2.11a) has the explicit solution given for $t > 0$ by

$$\sigma(X_t) = \exp(t)\exp\Big(\Lambda(t - t_{N_t})\Big)\Lambda\exp\Big(\Lambda(t_{N_t} - t_{N_t - 1})\Big)\cdots\Lambda\exp\Big(\Lambda t_1\Big)\widehat{X}_0$$

and the vector of conditional probabilities $\widehat{X}_t$ is for $t > 0$

$$\widehat{X}_t = \frac{\exp\Big(\Lambda(t - t_{N_t})\Big)\Lambda\exp\Big(\Lambda(t_{N_t} - t_{N_t-1})\Big)\cdots\Lambda\exp(\Lambda t_1)\widehat{X}_0}{\mathbf{1}^\top\exp\Big(\Lambda(t - t_{N_t})\Big)\Lambda\exp\Big(\Lambda(t_{N_t} - t_{N_t-1})\Big)\cdots\Lambda\exp(\Lambda t_1)\widehat{X}_0}$$

where the row vector $\mathbf{1}^\top$ is the transpose of vector $\mathbf{1}$.

In the next section, we derive a robust-version of the filters.

3 Conditional Poisson observations

Here, we use the ideas developed in [3,7] for the analysis of filtering equations for a finite Markov process observed through the continuous-valued observation process $y_t = \int_0^t \langle \boldsymbol{\lambda}, X_s\rangle\,\mathrm{d}s + W_t$, where $\{W_t\}_{t\geq 0}$ is an Brownian motion that is independent of $\{X_t\}_{t\geq 0}$ and $\boldsymbol{\lambda}$ is a n-vector of scalars. This discretization procedure is based on a transformation of the filtering equation for the state proposed by [3]. One of the main interests in this transformation was to offer a robust form of the filter, that is the "transformed filter" has a continuous dependence on the observation paths. Here, our observation process is discrete-valued, so that the uniform metric is replaced by a suitable metric for rcll processes, a Skorokhod metric. Two papers that uses this approach for state estimation of an MMPP have been published very recently [10,5]. So, we will take benefit of their results for the state estimation problem. Since we are concerned with other filters than the state filter, we just prove our results for the new estimates. Thus, a complete counterpart of the James-Krishnamurthy-Le Gland approach is obtained for MMPPs.

3.1 Clark's transformation and filtering.

We recall that the diagonal matrix Λ is non-singular from (2.7). Let us introduce the diagonal random matrix

$$C_t := \Lambda^{N_t} \exp\Big((I-\Lambda)t\Big) = \mathrm{Diag}\Big(\lambda(i)^{N_t} \exp\Big((1-\lambda(i))t\Big)\Big). \tag{3.1}$$

Note that the matrices C_t and C_t^{-1} commute with Λ. The product rule (PR) allows us to show that

$$C_t^{-1} = \int_0^t C_{s-}^{-1}(\Lambda^{-1}-I)\,\mathrm{d}N_s + \int_0^t (\Lambda - I)C_{s-}^{-1}\,\mathrm{d}s. \tag{3.2}$$

In particular, we have for the jumps

$$\Delta C_s^{-1} = C_{s-}^{-1}(\Lambda^{-1}-I)\Delta N_s. \tag{3.3}$$

Theorem 1 *For any integrable $\mathbb{F}$-adapted process $\{Z_t\}_{t\geq 0}$, $\sigma(Z_t)$ denotes its Zakai's filter. Let us denote $C_t^{-1}\sigma(Z_t)$ by $\overline{\sigma}(Z_t)$. The following relations hold for every $t \geq 0$:*

$$\overline{\sigma}(X_t) = \overline{\sigma}(X_0) + \int_0^t C_s^{-1} Q C_s \overline{\sigma}(X_s)\,\mathrm{d}s \tag{3.4a}$$

$$\overline{\sigma}(\mathcal{O}_t^{(i)} X_t) = \int_0^t \Big[C_s^{-1} Q C_s \overline{\sigma}(\mathcal{O}_s^{(i)} X_s) + \langle \overline{\sigma}(X_s), e_i\rangle\, e_i\Big]\,\mathrm{d}s \tag{3.4b}$$

$$\overline{\sigma}(N_t^{ji} X_t) = \int_0^t \Big[C_s^{-1} Q C_s \overline{\sigma}(N_s^{X,ji} X_s) + \langle \overline{\sigma}(X_s), e_i\rangle \langle C_s^{-1} Q C_s e_i, e_j\rangle\, e_j\Big]\,\mathrm{d}s \quad j \neq i \tag{3.4c}$$

$$\overline{\sigma}(G_t^i X_t) = \int_0^t C_s^{-1} Q C_s \overline{\sigma}(G_s^i X_s)\,\mathrm{d}s + \int_0^t \langle \overline{\sigma}(X_{s-}), e_i\rangle\, e_i\,\mathrm{d}N_s. \tag{3.4d}$$

The last stochastic integral in (3.4d) may be rewritten as follows using the product rule (PR)

$$\begin{aligned}\int_0^t \langle \overline{\sigma}(X_{s-}), e_i\rangle\, e_i\,\mathrm{d}N_s &= \langle \overline{\sigma}(X_t), e_i\rangle N_t e_i - \int_0^t N_{s-}\langle \mathrm{d}\overline{\sigma}(X_s), e_i\rangle\, e_i \\ &= \langle \overline{\sigma}(X_t), e_i\rangle N_t e_i - \int_0^t N_{s-}\langle C_s^{-1} Q C_s \overline{\sigma}(X_s), e_i\rangle\, e_i\,\mathrm{d}s\end{aligned}$$

from (3.4a). Equation (3.4d) has the final form

$$\begin{aligned}\overline{\sigma}(G_t^i X_t) = &\int_0^t \Big[C_s^{-1}QC_s\overline{\sigma}(G_s^i X_s) - N_{s-}\langle C_s^{-1}QC_s^{-1}\overline{\sigma}(X_s), e_i\rangle\, e_i\Big]\,\mathrm{d}s \\ &+ \langle\overline{\sigma}(X_t), e_i\rangle N_t e_i.\end{aligned} \tag{3.5}$$

Proof. First, we find from the diagonal structure of C_t and C_t^{-1} that

$$\begin{aligned}\langle\sigma(X_t), e_i\rangle C_t^{-1}e_i &= \langle\overline{\sigma}(X_t), e_i\rangle e_i \\ \text{and}\quad Q(j,i)\langle\sigma(X_t), e_i\rangle C_t^{-1}e_j &= \langle\overline{\sigma}(X_t), e_i\rangle\langle C_t^{-1}QC_t e_i, e_j\rangle\, e_j.\end{aligned}$$

Relation (3.4a) is stated in [10]. The authors refer to [9] for a proof. A short proof is reported here for the convenience of the reader. The product rule (PR) gives

$$\begin{aligned}C_t^{-1}\sigma(X_t) = &\, C_0^{-1}\sigma(X_0) + \int_0^t C_{s-}^{-1}\mathrm{d}\sigma(X_s)\,\mathrm{d}s + \int_0^t \mathrm{d}C_s^{-1}\sigma(X_{s-}) \\ &+ \sum_{0<s\le t}\Delta C_s^{-1}\Delta\sigma(X_s) \\ = &\,\sigma(X_0) + \int_0^t C_{s-}^{-1}\Big[Q\sigma(X_s)\,\mathrm{d}s + (\Lambda - I)\sigma(X_{s-})\,\mathrm{d}n_s\Big] \\ &+ \int_0^t C_{s-}^{-1}(\Lambda^{-1} - I)\sigma(X_{s-})\,\mathrm{d}N_s + \int_0^t C_{s-}^{-1}(\Lambda - I)C_{s-}^{-1}\sigma(X_{s-})\,\mathrm{d}s \\ &+ \sum_{0<s\le t} C_{s-}^{-1}(\Lambda^{-1} - I)(\Lambda - I)\sigma(X_{s-})\,\Delta N_s\end{aligned}$$

the last equality is deduced from (2.11a), (3.2), (3.3) respectively. Next, replace $\mathrm{d}n_s$ by $\mathrm{d}N_s - \mathrm{d}s$ and we obtain

$$\begin{aligned}\overline{\sigma}(X_t) = &\,\overline{\sigma}(X_0) + \int_0^t Q\overline{\sigma}(X_s)\,\mathrm{d}s + \int_0^t(\Lambda - I)\overline{\sigma}(X_{s-})\,\mathrm{d}N_s \\ &+ \int_0^t(\Lambda^{-1} - I)\overline{\sigma}(X_{s-})\,\mathrm{d}N_s \\ &+ \int_0^t(\Lambda^{-1} - I)(\Lambda - I)\overline{\sigma}(X_{s-})\,\mathrm{d}N_s.\end{aligned}$$

In the right hand side of the equality above, all terms cancel except the two former and Formula (3.4a) follows.

Now, if $\{\overline{\sigma}_t\}_{t\ge 0}$ is a solution of

$$\overline{\sigma}_t = \overline{\sigma}_0 + \int_0^t C_s^{-1}QC_s\overline{\sigma}_s\,\mathrm{d}s$$

then $(C_t\overline{\sigma}_t)$ is a solution of (2.11a) since

$$
\begin{aligned}
C_t\overline{\sigma}_t &= C_0\overline{\sigma}_0 + \int_0^t C_{s-}\,\mathrm{d}\overline{\sigma}_s + \int_0^t \mathrm{d}C_s\,\overline{\sigma}_{s-}\\
&= C_0\overline{\sigma}_0 + \int_0^t C_s C_s^{-1} Q C_s\overline{\sigma}_s\,\mathrm{d}s + \int_0^t (\Lambda - I)C_{s-}\overline{\sigma}_{s-}\mathrm{d}N_s + \int_0^t C_s(I-\Lambda)\overline{\sigma}_s\,\mathrm{d}s\\
&= C_0\overline{\sigma}_0 + \int_0^t QC_s\sigma(X_s)\,\mathrm{d}s + \int_0^t (\Lambda - I)C_{s-}\sigma(X_{s-})\,\mathrm{d}n_s.
\end{aligned}
$$

Let us check that (3.4d) is valid. We deduce from (3.3) and (2.11d) that

$$
\sum_{0<s\leq t} \Delta C_s^{-1}\Delta\sigma(G_s^iX_s) = \int_0^t C_{s-}^{-1}(\Lambda^{-1}-I)\Big[\lambda(i)\langle\sigma(X_{s-}),e_i\rangle e_i + (\Lambda - I)\sigma(G_{s-}^iX_{s-})\Big]\,\mathrm{d}N_s.
$$

Then, the product rule gives

$$
\begin{aligned}
\overline{\sigma}(G_t^iX_t) &= \int_0^t C_{s-}^{-1}\Big[Q\sigma(G_s^iX_s)\,\mathrm{d}s + \lambda(i)\langle\sigma(X_{s-}),e_i\rangle e_i\,\mathrm{d}N_s +\\
&\qquad\qquad (\Lambda - I)\sigma(G_{s-}^iX_{s-})\,\mathrm{d}n_s]\\
&\quad + \int_0^t C_{s-}^{-1}(\Lambda^{-1}-I)\,\mathrm{d}N_s + (\Lambda - I)C_{s-}^{-1}\,\mathrm{d}s]\,\sigma(G_{s-}^iX_{s-})\\
&\quad + \int_0^t C_{s-}^{-1}(\Lambda^{-1}-I)\lambda(i)\langle\sigma(X_{s-}),e_i\rangle\, e_i\,\mathrm{d}N_s\\
&\quad + \int_0^t C_{s-}^{-1}(\Lambda^{-1}-I)(\Lambda - I)\sigma(G_{s-}^iX_{s-})\,\mathrm{d}N_s.
\end{aligned}
$$

Replacing $\mathrm{d}n_s$ by $\mathrm{d}N_s - \mathrm{d}s$, several terms cancel and formula (3.4d) follows. Its is easily seen that if $\{\overline{\sigma}_t\}_{t\geq 0}$ is a solution of

$$
\overline{\sigma}_t = \int_0^t C_s^{-1}QC_s\overline{\sigma}_s\,\mathrm{d}s + \int_0^t \lambda(i)\langle\sigma(X_{s-}),e_i\rangle\, C_{s-}^{-1}\Lambda^{-1}e_i\,\mathrm{d}N_s,
$$

then $(C_t\overline{\sigma}_t)$ is a solution of (2.11d). Indeed, we have

$$
\begin{aligned}
\mathcal{C}_t\overline{\sigma}_t &= \int_0^t C_{s-}\,\mathrm{d}\overline{\sigma}_s + \int_0^t \mathrm{d}C_s\,\overline{\sigma}_{s-} + \sum_{0<s\leq t}\Delta C_s\Delta\overline{\sigma}_s\\
&= \int_0^t C_sC_s^{-1}QC_s\overline{\sigma}_s\,\mathrm{d}s + \int_0^t \lambda(i)\langle\sigma(X_{s-}),e_i\rangle\Lambda^{-1}e_i\mathrm{d}N_s\\
&\quad + \int_0^t (\Lambda - I)C_{s-}\overline{\sigma}_{s-}\mathrm{d}N_s + \int_0^t C_s(I-\Lambda)\overline{\sigma}_s\,\mathrm{d}s\\
&\quad + \int_0^t \lambda(i)\langle\sigma(X_{s-}),e_i\rangle(I-\Lambda^{-1})e_i\,\mathrm{d}N_s\\
&= \int_0^t QC_s\overline{\sigma}_s\,\mathrm{d}s + \int_0^t (\Lambda - I)C_s\overline{\sigma}_s\,\mathrm{d}n_s + \int_0^t \lambda(i)\langle\sigma(X_{s-}),e_i\rangle e_i\,\mathrm{d}N_s.
\end{aligned}
$$

The proofs of (3.4b,3.4c) are quite similar so that they are omitted. ∎

In the spirit of the works of [3,7], a discretization scheme of (3.4a-3.4c) and (3.4d),(3.5) is as follows. Consider a regular partition $0 = t_0 < t_1 < \cdots t_l < \cdots$, with a constant mesh h. A basic approximation of (3.4a) is a

$$\begin{aligned}\overline{\sigma}(X_{t_l}) &= \overline{\sigma}(X_{t_{l-1}}) + \int_{t_{l-1}}^{t_l} C_s^{-1} Q C_s \overline{\sigma}(X_s)\,\mathrm{d}s \\ &\simeq \overline{\sigma}(X_{t_{l-1}}) + h C_{t_{l-1}}^{-1} Q C_{t_{l-1}} \overline{\sigma}(X_{t_{l-1}}) = \Big[I + h C_{t_{k-1}}^{-1} Q C_{t_{k-1}}\Big]\overline{\sigma}(X_{t_{k-1}})\end{aligned}$$

The following explicit approximations are obtained in a similar way

$$\begin{aligned}\overline{\sigma}(\mathcal{O}^{(i)}_{t_l} X_{t_l}) &\simeq \Big(I + h C_{t_{l-1}}^{-1} Q C_{t_{l-1}}\Big)\overline{\sigma}(\mathcal{O}^{(i)}_{t_{l-1}} X_{t_{l-1}}) + h\langle \overline{\sigma}(X_{t_{l-1}}), e_i\rangle e_i \\ \overline{\sigma}(N^{ji}_{t_l} X_{t_l}) &\simeq \Big(I + h C_{t_{l-1}}^{-1} Q C_{t_{l-1}}\Big)\overline{\sigma}(N^{ji}_{t_{l-1}} X_{t_{l-1}}) \\ &\qquad + h\langle \overline{\sigma}(X_{t_{l-1}}), e_i\rangle \langle C_{t_{l-1}}^{-1} Q C_{t_{l-1}} e_i, e_j\rangle\, e_j \\ \overline{\sigma}(G^{i}_{t_l} X_{t_l}) &\simeq \Big(I + h C_{t_{l-1}}^{-1} Q C_{t_{l-1}}\Big)\overline{\sigma}(G^{i}_{t_{l-1}} X_{t_{l-1}}) \\ &\qquad + \Big\langle (I + h C_{t_{l-1}}^{-1} Q C_{t_{l-1}})\overline{\sigma}(X_{t_{l-1}}), e_i\Big\rangle (N_{t_l} - N_{t_{l-1}})\, e_i.\end{aligned}$$

Multiplying both sides of the previous equations by C_{t_l}, we obtain the following scheme for approximating the solutions of equations (2.11a-2.11d)

$$\begin{aligned}\sigma(X_{t_l}) &\simeq \Psi_{l,l-1}\Big(I + hQ\Big)\sigma(X_{t_{l-1}}) \\ \sigma(\mathcal{O}^{(i)}_{t_l} X_{t_l}) &\simeq \Psi_{l,l-1}\Big(I + hQ\Big)\sigma(\mathcal{O}^{(i)}_{t_{l-1}} X_{t_{l-1}}) + h\langle \sigma(X_{t_{l-1}}), e_i\rangle\, \Psi_{l,l-1} e_i \\ \sigma(N^{ji}_{t_l} X_{t_l}) &\simeq \Psi_{l,l-1}\Big(I + hQ\Big)\sigma(N^{ji}_{t_{l-1}} X_{t_{l-1}}) \\ &\qquad + hQ(j,i)\langle \sigma(X_{t_{l-1}}), e_i\rangle\, \Psi_{l,l-1} e_j \\ \sigma(G^{i}_{t_l} X_{t_l}) &\simeq \Psi_{l,l-1}\Big(I + hQ\Big)\sigma(G^{i}_{t_{l-1}} X_{t_{l-1}}) \\ &\qquad + \Big\langle (I + hQ)\sigma(X_{t_{l-1}}), e_i\Big\rangle\, \Psi_{l,l-1} e_i\end{aligned} \tag{3.6}$$

where $\Psi_{l,l-1} := C_{t_l} C_{t_{l-1}}^{-1}$ is a diagonal matrix. If the mesh h is less than $1/\max_i(|Q(i,i)|)$, it has been shown [10, Th 5.2] that the recursive equation (3.6) produces values of $\sigma(X_t)$ that can not be negative as it may happen with a standard discrete-time approximation as Euler-Maruyama's one. We refer to [10] for details.

3.2 Skorokhod-continuity.

A measure of the discrepancy between two paths $N(\omega_1) := \{N_t(\omega_1)\}_{[0,T]}$ and $N(\omega_1) := \{N_t(\omega_2)\}_{[0,T]}$ of the counting process $\{N_t\}_{t\geq 0}$ over the interval $[0,T]$

is given by

$$d_S\Big(N(\omega_1), N(\omega_2)\Big) = \inf_b \Big\{ \sup_{0\le t\le T} |b_t - t| \vee \sup_{0\le t\le T} |N_t(\omega_1) - N_{b_t}(\omega_2)| \Big\}$$

where $t \mapsto b_t$ is an increasing, one to one, function from $[0, T]$ into $[0, T]$ such that $b_0 = 0$ and $b_T = T$. Such a function is called a change of time and (N_{b_t}) is the so-called time change of $\{N_t\}_{t\ge 0}$ associated with $\{b_t\}_{[0,T]}$. d_S is one of the metrics for the Skorokhod topology of the set of rcll processes $\mathbb{D}[0, +\infty]$. Note that the space $\mathbb{D}[0, +\infty]$ is separable but not complete for the metric above. However d_S is an equivalent metric to the standard one for which the space $\mathbb{D}[0, +\infty]$ is complete. The following theorem extents Lemma 3.8 in [10] to our family of filters. It states that if the observation paths are "close", so are the filters. The term "robust filters" relies on this property.

Theorem 2 *The notations in Theorem 1 are used. Let us define $\pi(Z_t X_t)$ by*

$$\pi(Z_t X_t) := \frac{C_t \overline{\sigma}(Z_t X_t)}{\langle C_t \overline{\sigma}(X_t), \mathbf{1}\rangle}. \tag{3.7}$$

Then, $\pi(Z_t X_t)$ is a Skorokhod-continuous version of $\mathbb{E}[Z_t X_t \mid \mathbb{F}_t^N]$.

Proof. It has been shown in the previous theorem that $C_t\overline{\sigma}(Z_t X_t)$ is a solution of the associated equation in (2.11a-2.11d). Since these equations have an unique solution, $C_t\overline{\sigma}(Z_t X_t) = \sigma(Z_t X_t)$ a.s. Thus, it follows from $\langle \sigma(X_t), \mathbf{1}\rangle = \mathbb{E}_0[L_t \mid \mathbb{F}_t^N]$ and (2.10) that $\mathbb{E}[Z_t X_t \mid \mathbb{F}_t^N] = \pi(Z_t X_t)$ a.s.

We restrict ourselves to paths for which $d_S\Big(N(\omega_1), N(\omega_2)\Big) < 1$. In such a case, we note that $N_T(\omega_1) = N_T(\omega_2)$, i.e. we have observed the same number of jumps at time T. Assume that this common value is K. Let $t_1, \dots, t_K$ and $s_1, \dots, s_K$ be the sequence of times of jump. Then,

$$d_S\Big(N(\omega_1), N(\omega_2)\Big) = \max_{1\le i\le K} |t_i - s_i|.$$

We know from [10] that $\pi(X_t)$ defined in (3.7) is Skorokhod-continuous. In particular, the authors show that

$$\max_{0\le t\le T} \|\overline{\sigma}(X_t)(\omega_1) - \overline{\sigma}(X_t)(\omega_2)\| \le K_T d_S\Big(N(\omega_1), N(\omega_2)\Big) \tag{3.8}$$

Now we give a proof of Skorokhod-continuity of $\pi(\mathcal{O}_t^{(i)} X_t)$. We have from (3.4b)

$$\begin{aligned}
&\|\overline{\sigma}(\mathcal{O}_t^{(i)}X_t)(\omega_1) - \overline{\sigma}(\mathcal{O}_t^{(i)}X_t)(\omega_2)\| \\
&\leq \int_0^t \left\| C_s^{-1}QC_s(\omega_1)\overline{\sigma}(\mathcal{O}_s^{(i)}X_s)(\omega_1) - C_s^{-1}QC_s(\omega_2)\overline{\sigma}(\mathcal{O}_s^{(i)}X_s)(\omega_2)\right\| \mathrm{d}s \quad (3.9)\\
&\quad + \int_0^t \left\| \langle \overline{\sigma}(X_s)(\omega_1) - \overline{\sigma}(X_s)(\omega_2), e_i\rangle \, e_i \right\| \mathrm{d}s. \quad (3.10)
\end{aligned}$$

First, observe that $C_s^{-1}QC_s(\omega_1) = C_s^{-1}QC_s(\omega_2)$ except when $N_s(\omega_1) \neq N_s(\omega_2)$. Since
$\|C_s^{-1}QC_s(\omega_1)\| \leq \kappa_T$, it follows that the term in (3.9) is bounded from above by

$$\begin{aligned}
&\kappa_T \int_0^t \|\overline{\sigma}(\mathcal{O}_s^{(i)}X_s)(\omega_1) - \overline{\sigma}(\mathcal{O}_s^{(i)}X_s)(\omega_2)\| \,\mathrm{d}s + D_T\, K \, \max_i |t_i - s_i| \\
&\leq \kappa_T \int_0^t \|\overline{\sigma}(\mathcal{O}_s^{(i)}X_s)(\omega_1) - \overline{\sigma}(\mathcal{O}_s^{(i)}X_s)(\omega_2)\| \,\mathrm{d}s + D_T\, K \, d_S\Big(N(\omega_1), N(\omega_2)\Big).
\end{aligned}$$

In the sequel, κ_T will denote a generic constant. It is easily seen from (3.8) that the term in (3.10) is bounded from above by

$$\kappa_T d_S\Big(N(\omega_1), N(\omega_2)\Big).$$

Combining the bounds obtained for the two terms in (3.9-3.10), we obtain for $t \leq T$

$$\Phi(t) \leq \kappa_T \left(\int_0^t \Phi(s)\,\mathrm{d}s + d_S\Big(N(\omega_1), N(\omega_2)\Big)\right)$$

where $\Phi(t) := \max_{0\leq s\leq t} \|\overline{\sigma}(\mathcal{O}_s^{(i)}X_s)(\omega_1) - \overline{\sigma}(\mathcal{O}_s^{(i)}X_s)(\omega_2)\|$. Then, Gronwall's lemma allows us to conclude to

$$\max_{0\leq t\leq T} \|\overline{\sigma}(\mathcal{O}_t^{(i)}X_t)(\omega_1) - \overline{\sigma}(\mathcal{O}_t^{(i)}X_t)(\omega_2)\| \leq \kappa_T d_S\Big(N(\omega_1), N(\omega_2)\Big).$$

Using the equality $\sigma(X_t) = C_t\overline{\sigma}(\mathcal{O}_t^{(i)}X_t)$, it is easily seen from the inequality above that

$$\max_{0\leq t\leq T} \|\sigma(\mathcal{O}_t^{(i)}X_t)(\omega_1) - \sigma(\mathcal{O}_t^{(i)}X_t)(\omega_2)\| \leq \kappa_T d_S\Big(N(\omega_1), N(\omega_2)\Big).$$

The Skorokhod continuity of $\pi(Z_tX_t)$ with $Z_t := N_t^{ji}, G_t^i$ is shown in a quite similar way. ▪

References

[1] P. Bremaud. *Point Processes and Queues.* Springer Series in Statistics. Springer, 1981.

[2] O. Cappé, E. Moulines, and T. Rydén. *Inference in Hidden Markov Models.* Springer, 2005.

[3] J. M. C. Clark. The design of robust approximations to the stochastic differential equations of nonlinear filtering. In Skwirzynski, J.K., editor, *Communication Systems and Random Processes Theory (Darlington 1977)*, pages 721–734. Alphen aan den Rijn, The Netherlands, 1978.

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

[5] R. J. Elliott and W. P. Malcolm. General smoothing formulas for Markov-modulated poisson observations. *IEEE Trans. Automat. Control*, 50(8):1123–1134, 2005.

[6] K. Goseva-Popstojanova and K. S. Trivedi. Architecture-based approach to reliability assessment of software systems. *Performance Evaluation*, 45:179–204, 2001.

[7] M. R. James, V. Krishnamurthy, and F. Le Gland. Time discretization of continuous-time filters and smoothers for HMM parameter estimation. *IEEE Trans. Inform. Theory*, 42:593–605, 1996.

[8] J. Ledoux. Filtering and the EM-algorithm for the Markovian Arrival Process. *Comm. Statist. - Theory and Methods*, 2006. To appear.

[9] W. P. Malcolm. *Robust filtering and estimation with Poisson observations.* PhD thesis, The Australian National University, Camberra, 1999.

[10] W. P. Malcolm, R. J. Elliott, and J. van der Hoek. A deterministic discretization-step upper bound for state estimation via Clark transformations. *J. Appl. Math. Stoch. Anal.*, 17(4):371–384, 2004.

[11] D. W.-C. Miao. *Markov Modulated Poisson Processes in Credit Risk Modeling.* PhD thesis, University of Oxford, 2005.

[12] M. F. Neuts. *Structured Stochastic Matrices of M/G/1 Type and Their Applications.* Marcel Dekker Inc., New-York and Basel, 1989.

[13] M. F. Neuts. Chapter 10: Matrix-analytic methods in queueing theory. In J. H. Dshalalow, editor, *Advances in Queueing*, Probability and Stochastic Series, pages 265–292. CRC Press, 1995.

[14] W. J. Roberts, Y. Ephraim, and E. Dieguez. On Rydén's EM algorithm for estimating MMPPs. *IEEE Signal Processing Letters*, 13:373–376, 2006.

[15] T. Rydén. An EM-algorithm for estimation in Markov-modulated Poisson process. *Computational Statistics and Data Analysis*, 21:431–447, 1996.

[16] D. L. Snyder and M. I. Miller. *Random Point Processes in Time and Space.* Springer, 1991.