EM based parameter estimation for Markov modulated Fluid Arrival Processes*

Salah Al-Deen Almousa[1[0000−0002−2648−4133]], Gábor Horváth[1[0000−0003−3097−1273]], and Miklós Telek[1,2[0000−0001−9600−6084]]

[1] Department of Networked Systems and Services, Budapest University of Technology and Economics, Hungary {almousa, ghorvath,telek}@hit.bme.hu
[2] MTA-BME Information Systems Research Group, ELKH, Hungary

Abstract. Markov modulated discrete arrival processes have a wide literature, including parameter estimation methods based on expectation-maximization (EM). In this paper, we investigate the adaptation of these EM based methods to Markov modulated fluid arrival processes (MMFAP), and conclude that only some parameters of MMFAPs can be approximated this way.
Keywords: Markov modulated fluid arrival processes, expectation-maximization method, parameter estimation.

1 Introduction

Markovian queueing systems with discrete customers are widely used in stochastic modeling. Markovian Arrival Process (MAP, [11]), that are able to characterize a wide class of point processes, play an important role in these systems. The properties of MAPs have been studied exhaustively, using queueing models involving MAPs nowadays has become common, queueing networks with MAP traffic have also been investigated. Several methods exist to create a MAP approximating real, empirical data. Some of them aim to match statistical quantities like marginal moments, joint moments, auto-correlation, etc. [9, 15] An other approach to create MAPs from measurement data is based on likelihood maximization, which is often performed by Expectation-Maximization (EM) [12]. Several EM-based fitting methods have been published for MAPs, based on randomization [5], based on special structures [14, 7], methods that support batch arrivals [4] and those that are able to work with group data [13].

In many systems the workload can be represented easier with continuous, fluid-like models rather than discrete demands [1]. Basic Markovian fluid models have been introduced and analyzed in [10, 2], later on several model variants appeared and were investigated. Despite of their practical relevance, the "ecosystem" around fluid models is far less complete than in the discrete case. In particular, fitting methods for Markov modulated fluid arrival processes (MMFAP) are available only to some very restricted cases like on-off models, motivated

* This work is partially supported by the OTKA K-123914 project and the Artificial Intelligence National Laboratory Programme of Hungary.

by telecommunication applications. To the best of our knowledge, fitting methods for the general class of MMFAPs based on likelihood maximization has not been investigated in the past. At first glance adapting the methods available for MAPs might seem feasible, since fluid models can be treated as a limit of a discrete model generating infinitesimally small fluid drops. In fact, adapting the algorithms for MAPs to MMFAPs is not straight forward at all, fitting MMFAPs is a qualitatively different problem.

The rest of the paper is organized as follows. Section 2 introduces the mathematical model and the parameter estimation problem. The next section discusses the applicability of the EM method for MMFAPs. Some implementation details associated with the EM method for MMFAPs are provided in Section 4. Finally, Section 5 provides numerical experiments about the properties of the proposed method and Section 6 concludes the paper.

2 Problem definition

2.1 Fluid arrival process

The fluid arrival process $\{\mathcal{Z}(t) = \{\mathcal{J}(t), \mathcal{X}(t)\}, t > 0\}$ consists of a background continuous time Markov chain $\{\mathcal{J}(t), t > 0\}$ which modulates the arrival process of the fluid $\{\mathcal{X}(t), t > 0\}$. When the Markov chain stays in state i for a Δ long interval a normal distributed amount of fluid arrives with mean $r_i\Delta$ and variance $\sigma_i^2\Delta$, that is, when $\mathcal{J}(\tau) = i, \forall\tau \in (t, t+\Delta)$

$$\frac{d}{dx} Pr(\mathcal{X}(t+\Delta) - \mathcal{X}(t) < x) = \mathcal{N}(r_i\Delta, \sigma_i^2\Delta, x), \tag{1}$$

where $\mathcal{N}(\mu, \sigma^2, x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ is the Gaussian density function. We note that our proposed analysis approach allows negative fluid rates as well. Since the normal distribution has infinite support also in case of strictly positive fluid rates Section 4.2 discusses a numerical approach to handle negative fluid samples.

The generator matrix and the initial probability vector of the N-state background continuous time Markov chain (CTMC) are $\mathbf{Q}$ and $\underline{\alpha}$, and the diagonal matrix of the fluid rates and the fluid variances are given by matrix $\mathbf{R}$ with $\mathbf{R}_{i,j} = \delta_{i,j} r_i$, and matrix $\mathbf{\Sigma}$ with $\mathbf{\Sigma}_{i,j} = \delta_{i,j}\sigma_i^2$, where $\delta_{i,j}$ denotes the Kronecker delta.

Assuming $\mathcal{X}(0) = 0$, the amount of fluid arriving in the $(0, t)$ interval is $\mathcal{X}(t)$, with density matrix defined by

$$[\mathbf{N}(t, x)]_{i,j} = \frac{\partial}{\partial x} \Pr\left(\mathcal{X}(t) < x, \mathcal{J}(t) = j | \mathcal{J}(0) = i\right) \tag{2}$$

The double sided Laplace transform of this quantity regarding the amount of fluid arrived can be expressed as [6]

$$\mathbf{N}^*(t, v) = \int_{x=-\infty}^{\infty} e^{-xv} \mathbf{N}(t, x) dx = e^{(\mathbf{Q} - v\mathbf{R} - v^2\mathbf{\Sigma}/2)t}. \tag{3}$$

The stationary distribution of the CTMC is denoted by vector $\underline{\pi}$, which is the solution of $\underline{\pi}\mathbf{Q} = \underline{0}, \underline{\pi}\mathbb{1} = \mathbb{1}$. In this work, we are interested in the stationary fluid arrival process and assume that the initial probability vector of the background CTMC is $\underline{\alpha} = \underline{\pi}$.

2.2 Measurement data to fit

We assume that the data to fit is given by a series of pairs $\mathcal{D} = \{(t_k, x_k); k = 1, \ldots, K\}$, where t_k is the time since the last measurement instant and x_k is the amount of fluid arrived since the last measurement instant (which can be negative as well). That is, the measurement instances are $T_k = \sum_{\ell=1}^{k} t_\ell$ for $k \in \{1, \ldots, K\}$.

The likelihood of the data is defined as

$$\mathcal{L} = \underline{\alpha} \prod_{k=1}^{K} \mathbf{N}(t_k, x_k)\mathbb{1}. \tag{4}$$

Our goal is to find $\mathbf{Q}$, $\mathbf{R}$ and $\mathbf{S}$ which maximize the likelihood.

3 The EM algorithm

The EM algorithm is based on the observation that the likelihood would be easier to maximize when certain unobserved, hidden variables were known. In our case the hidden variables are related to the trajectory of the hidden Markov chain, specifically

- $J_n^{(k)}$ is the nth state visited by the Markov chain in the kth measurement interval,
- $\theta_n^{(k)}$ is the sojourn time of the nth sojourn of the Markov chain (which is in state $J_n^{(k)}$) in the kth measurement interval,
- $f_n^{(k)}$ is the fluid accumulated during the nth sojourn in the kth measurement interval,
- $n^{(k)}$ is the number of sojourns in the kth measurement interval.

Based on these hidden variables the logarithm of the likelihood is computed in the next section.

3.1 Log-likelihood as a function of the hidden variables

With the hidden variables defined above, the likelihood $\mathcal{L}$ can be expressed as

$$\mathcal{L} = \prod_{k=1}^{K} e^{-\theta_1^{(k)} q_{J_1^{(k)}}} \mathcal{N}\left(\theta_1^{(k)} r_{J_1^{(k)}}, \theta_1^{(k)} \sigma^2_{J_1^{(k)}}, f_1^{(k)}\right) q_{J_1^{(k)} J_2^{(k)}}$$
$$\cdot\, e^{-\theta_2^{(k)} q_{J_2^{(k)}}} \mathcal{N}\left(\theta_2^{(k)} r_{J_2^{(k)}}, \theta_2^{(k)} \sigma^2_{J_2^{(k)}}, f_2^{(k)}\right) q_{J_2^{(k)} J_3^{(k)}} \cdot \ldots$$

$$
\begin{aligned}
&\cdot e^{-\theta^{(k)}_{n^{(k)}} q_{J^{(k)}_{n^{(k)}}}} \mathcal{N}\left(\theta^{(k)}_{n^{(k)}} r_{J^{(k)}_{n^{(k)}}}, \theta^{(k)}_{n^{(k)}} \sigma^2_{J^{(k)}_{n^{(k)}}}, f^{(k)}_{n^{(k)}}\right) \\
=&\prod_{k=1}^{K} \prod_{n=1}^{n^{(k)}-1} e^{-\theta^{(k)}_n q_{J^{(k)}_n}} \mathcal{N}\left(\theta^{(k)}_n r_{J^{(k)}_n}, \theta^{(k)}_n \sigma^2_{J^{(k)}_n}, f^{(k)}_n\right) q_{J^{(k)}_n J^{(k)}_{n+1}} \\
&\cdot e^{-\theta^{(k)}_{n^{(k)}} q_{J^{(k)}_{n^{(k)}}}} \mathcal{N}\left(\theta^{(k)}_{n^{(k)}} r_{J^{(k)}_{n^{(k)}}}, \theta^{(k)}_{n^{(k)}} \sigma^2_{J^{(k)}_{n^{(k)}}}, f^{(k)}_{n^{(k)}}\right),
\end{aligned}
$$

where $\mathcal{N}(\mu, \sigma^2, x)$ is the Gaussian density function and $q_i = \sum_{j,j\neq i} q_{ij}$ is the departure rate of state i of the CTMC. Using $\log \mathcal{N}(\mu, \sigma^2, x) = -\frac{c}{2} - \frac{\log \sigma^2}{2} - \frac{(x-\mu)^2}{2\sigma^2}$ with $c = \log 2\pi$ we have

$$
\begin{aligned}
&\log\left(\mathcal{N}\left(\theta^{(k)}_n r_{J^{(k)}_n}, \theta^{(k)}_n \sigma^2_{J^{(k)}_n}, f^{(k)}_n\right)\right) \\
&= -\frac{c}{2} - \frac{\log(\theta^{(k)}_n \sigma^2_{J^{(k)}_n})}{2} - \frac{(f^{(k)}_n - \theta^{(k)}_n r_{J^{(k)}_n})^2}{2\theta^{(k)}_n \sigma^2_{J^{(k)}_n}} \\
&= -\frac{c}{2} - \frac{\log(\theta^{(k)}_n) + \log(\sigma^2_{J^{(k)}_n})}{2} - \frac{{f^{(k)}_n}^2 - 2 f^{(k)}_n \theta^{(k)}_n r_{J^{(k)}_n} + {\theta^{(k)}_n}^2 r^2_{J^{(k)}_n}}{2\theta^{(k)}_n \sigma^2_{J^{(k)}_n}} \\
&= -\frac{c}{2} - \frac{\log \theta^{(k)}_n}{2} - \frac{\log(\sigma^2_{J^{(k)}_n})}{2} - \frac{{f^{(k)}_n}^2}{2\theta^{(k)}_n \sigma^2_{J^{(k)}_n}} + \frac{f^{(k)}_n r_{J^{(k)}_n}}{\sigma^2_{J^{(k)}_n}} - \frac{\theta^{(k)}_n r^2_{J^{(k)}_n}}{2\sigma^2_{J^{(k)}_n}},
\end{aligned}
$$

and the log-likelihood is

$$
\begin{aligned}
\log \mathcal{L} = &\sum_{k=1}^{K} \sum_{n=1}^{n^{(k)}-1} -\theta^{(k)}_n q_{J^{(k)}_n} + \log\left(\mathcal{N}\left(\theta^{(k)}_n r_{J^{(k)}_n}, \theta^{(k)}_n \sigma^2_{J^{(k)}_n}, f^{(k)}_n\right)\right) + \log q_{J^{(k)}_n J^{(k)}_{n+1}} \\
&- \theta^{(k)}_{n^{(k)}} q_{J^{(k)}_{n^{(k)}}} + \log\left(\mathcal{N}\left(\theta^{(k)}_{n^{(k)}} r_{J^{(k)}_{n^{(k)}}}, \theta^{(k)}_{n^{(k)}} \sigma^2_{J^{(k)}_{n^{(k)}}}, f^{(k)}_{n^{(k)}}\right)\right) \\
= &\sum_{k=1}^{K} \sum_{n=1}^{n^{(k)}-1} -\theta^{(k)}_n q_{J^{(k)}_n} - \frac{c}{2} - \frac{\log \theta^{(k)}_n}{2} - \frac{\log(\sigma^2_{J^{(k)}_n})}{2} - \frac{{f^{(k)}_n}^2}{2\theta^{(k)}_n \sigma^2_{J^{(k)}_n}} + \frac{f^{(k)}_n r_{J^{(k)}_n}}{\sigma^2_{J^{(k)}_n}} \\
&- \frac{\theta^{(k)}_n r^2_{J^{(k)}_n}}{2\sigma^2_{J^{(k)}_n}} + \log q_{J^{(k)}_n J^{(k)}_{n+1}} - \theta^{(k)}_{n^{(k)}} q_{J^{(k)}_{n^{(k)}}} - \frac{c}{2} - \frac{\log \theta^{(k)}_{n^{(k)}}}{2} - \frac{\log(\sigma^2_{J^{(k)}_{n^{(k)}}})}{2} \\
&- \frac{{f^{(k)}_{n^{(k)}}}^2}{2\theta^{(k)}_{n^{(k)}} \sigma^2_{J^{(k)}_{n^{(k)}}}} + \frac{f^{(k)}_{n^{(k)}} r_{J^{(k)}_{n^{(k)}}}}{\sigma^2_{J^{(k)}_{n^{(k)}}}} - \frac{\theta^{(k)}_{n^{(k)}} r^2_{J^{(k)}_{n^{(k)}}}}{2\sigma^2_{J^{(k)}_{n^{(k)}}}}.
\end{aligned}
$$

Observe that knowing each individual hidden variable is not necessary to express the log-likelihood. It is enough to introduce the following aggregated measures to fully characterize interval k:

- $\Theta_i^{(k)} = \sum_{n=1}^{n^{(k)}} \theta_n^{(k)} \mathcal{I}_{\left\{J_n^{(k)}=i\right\}}$ is the total time spent in state i,
- $F_i^{(k)} = \sum_{n=1}^{n^{(k)}} f_n^{(k)} \mathcal{I}_{\left\{J_n^{(k)}=i\right\}}$ is the total amount of fluid arriving during a visit in state i,
- $M_i^{(k)} = \sum_{n=1}^{n^{(k)}} \mathcal{I}_{\left\{J_n^{(k)}=i\right\}}$ the number of visits to state i,
- $M_{i,j}^{(k)} = \sum_{n=1}^{n^{(k)}-1} \mathcal{I}_{\left\{J_n^{(k)}=i, J_{n+1}^{(k)}=j\right\}}$ the number of state transitions from state i to state j, additionally
- $L\Theta_i^{(k)} = \sum_{n=1}^{n^{(k)}} \log \theta_n^{(k)} \mathcal{I}_{\left\{J_n^{(k)}=i\right\}}$ is the total time spent in state i,
- $F\Theta_i^{(k)} = \sum_{n=1}^{n^{(k)}} \frac{f_n^{(k)2}}{\theta_n^{(k)}} \mathcal{I}_{\left\{J_n^{(k)}=i\right\}}$ is the total amount of fluid arriving during a visit in state i.

With these aggregate measures the log-likelihood simplifies to

$$\begin{aligned}
\log \mathcal{L} = \sum_{k=1}^{K} \sum_{n=1}^{n^{(k)}-1} & -\theta_n^{(k)} q_{J_n^{(k)}} - \frac{c}{2} - \frac{\log \theta_n^{(k)}}{2} - \frac{\log(\sigma^2_{J_n^{(k)}})}{2} - \frac{{f_n^{(k)}}^2}{2\theta_n^{(k)} \sigma^2_{J_n^{(k)}}} + \frac{f_n^{(k)} r_{J_n^{(k)}}}{\sigma^2_{J_n^{(k)}}} \\
& - \frac{\theta_n^{(k)} r^2_{J_n^{(k)}}}{2\sigma^2_{J_n^{(k)}}} + \log q_{J_n^{(k)} J_{n+1}^{(k)}} - \theta_{n^{(k)}}^{(k)} q_{J_{n^{(k)}}^{(k)}} - \frac{c}{2} - \frac{\log \theta_{n^{(k)}}^{(k)}}{2} - \frac{\log(\sigma^2_{J_{n^{(k)}}^{(k)}})}{2} \\
& - \frac{{f_{n^{(k)}}^{(k)}}^2}{2\theta_{n^{(k)}}^{(k)} \sigma^2_{J_{n^{(k)}}^{(k)}}} + \frac{f_{n^{(k)}}^{(k)} r_{J_{n^{(k)}}^{(k)}}}{\sigma^2_{J_{n^{(k)}}^{(k)}}} - \frac{\theta_{n^{(k)}}^{(k)} r^2_{J_{n^{(k)}}^{(k)}}}{2\sigma^2_{J_{n^{(k)}}^{(k)}}} \\
= \sum_{k=1}^{K} - \frac{cn^{(k)}}{2} + \sum_i \Bigg(& - \Theta_i^{(k)} \left(q_i + \frac{r_i^2}{2\sigma_i^2} \right) + F_i^{(k)} \frac{r_i}{\sigma_i^2} - \frac{L\Theta_i^{(k)}}{2} - \frac{F\Theta_i^{(k)}}{2\sigma_i^2} \\
& - M_i^{(k)} \frac{\log \sigma_i^2}{2} + \sum_{j, j \neq i} M_{i,j}^{(k)} \log q_{i,j} \Bigg).
\end{aligned}$$

3.2 The maximization step of the EM method

The maximization step of the EM method aims to find the optimal value of the model parameters based on the hidden variables. They are obtained from the partial derivatives of the log-likelihood as detailed in Appendix A.

Summarizing the results, the model parameter value which maximizes the log-likelihood based on the hidden variables are

$$q_{i,j} = \frac{\sum_{k=1}^{K} M_{i,j}^{(k)}}{\sum_{k=1}^{K} \Theta_i^{(k)}}, r_i = \frac{\sum_{k=1}^{K} F_i^{(k)}}{\sum_{k=1}^{K} \Theta_i^{(k)}}, \text{ and } \sigma_i^2 = \frac{\sum_{k=1}^{K} \Theta_i^{(k)} r_i^2 - 2F_i^{(k)} r_i + F\Theta_i^{(k)}}{\sum_{k=1}^{K} M_i^{(k)}}.$$

That is, $\sum_{k=1}^{K} \Theta_i^{(k)}$, and $\sum_{k=1}^{K} M_{i,j}^{(k)}$ are needed for computing the optimal $q_{i,j}$ parameters and additionally, $\sum_{k=1}^{K} F_i^{(k)}$, $\sum_{k=1}^{K} M_i$ and $\sum_{k=1}^{K} F\Theta_i^{(k)}$ are needed for the optimal r_i and σ_i^2 parameters.

3.3 The expectation step of the EM method

In the expectation step of the EM method the expected values of the hidden variables has to be evaluated based on the samples. Appendix B provides the analysis of those expectations, resulting $E(F_i^{(k)}) = r_i E(\Theta_i^{(k)})$ and $\mathrm{E}\left(F\Theta_i^{(k)}\right) = \mathrm{E}\left(M_i^{(k)}\right)\sigma_i^2 + \mathrm{E}\left(\Theta_i^{(k)}\right) r_i^2$, from which the zth iteration of the EM method updates the fluid rate and variance parameters as

$$r_i(z+1) = \frac{\sum_{k=1}^{K} \mathrm{E}\left(F_i^{(k)}\right)}{\sum_{k=1}^{K} \mathrm{E}\left(\Theta_i^{(k)}\right)} = r_i(z) \tag{5}$$

and

$$\begin{aligned}
\sigma_i^2(z+1) &= \frac{\sum_{k=1}^{K} \mathrm{E}\left(\Theta_i^{(k)} r_i^2(z) - 2F_i^{(k)} r_i(z) + F\Theta_i^{(k)}\right)}{\sum_{k=1}^{K} \mathrm{E}\left(M_i^{(k)}\right)} \\
&= \frac{\sum_{k=1}^{K} \mathrm{E}\left(\Theta_i^{(k)} r_i^2(z)\right) - 2\mathrm{E}\left(F_i^{(k)}\right) r_i(z) + \mathrm{E}\left(F\Theta_i^{(k)}\right)}{\sum_{k=1}^{K} \mathrm{E}\left(M_i^{(k)}\right)} \\
&= \frac{\sum_{k=1}^{K} \mathrm{E}\left(M_i^{(k)}\right) \sigma_i^2(z)}{\sum_{k=1}^{K} \mathrm{E}\left(M_i^{(k)}\right)} = \sigma_i^2(z).
\end{aligned} \tag{6}$$

Consequently, the fluid rate and variance parameters remain untouched by the EM method.

Remark 1. This result is in line with the results obtained for discrete arrival processes in [13] considering the special features of the fluid model. That is, we consider the MMPP arrival process, since there is no state transition at the fluid drop arrival, and fluid drops are assumed to be infinitesimal, hence for a finite amount of time there is an unbounded number of fluid drop arrivals. Using these features, equations (21) and (23) of [13] take the form

$$\mathrm{E}\left(Z_i^{[k]}\right) = \sum_{l=0}^{x_k} \int_0^{t_k} [f_k(l,\tau)]_i [b_k(x_k - l, t_k - \tau)]_i d\tau$$

$$\mathrm{E}\left(Y_{ii}^{[k]}\right) = \lambda_{ii} \sum_{l=0}^{x_k-1} \int_0^{t_k} [f_k(l,\tau)]_i [b_k(x_k - l, t_k - \tau)]_i d\tau.$$

Assuming x_k is large, the update of λ_{ii} in the zth step of the iteration is ((12) of [13])

$$\lambda_{ii}(z+1) = \frac{\sum_{k=1}^{K} \mathrm{E}\left(Y_{ii}^{[k]}\right)}{\sum_{k=1}^{K} \mathrm{E}\left(Z_i^{[k]}\right)}$$

$$= \lambda_{ii}(z) \frac{\sum_{k=1}^{K} \sum_{l=0}^{x_k-1} \int_0^{t_k} [f_k(l,\tau)]_i [b_k(x_k-l, t_k-\tau)]_i d\tau}{\sum_{k=1}^{K} \sum_{l=0}^{x_k} \int_0^{t_k} [f_k(l,\tau)]_i [b_k(x_k-l, t_k-\tau)]_i d\tau} \approx \lambda_{ii}(z).$$

The transition rate parameters are updated by the EM method as

$$q_{i,j} = \frac{\sum_{k=1}^{K} \mathrm{E}\left(M_{i,j}^{(k)}\right)}{\sum_{k=1}^{K} \mathrm{E}\left(\Theta_i^{(k)}\right)}. \tag{7}$$

The computation of $\mathrm{E}\left(M_{i,j}^{(k)}\right)$ and $\mathrm{E}\left(\Theta_i^{(k)}\right)$ are detailed in Appendix C and the results are summarized in (22) and (23).

4 Implementation details

The implementation of the EM based parameter estimation method contains some intricate elements which influence the computational complexity and the accuracy of the computations. This section summarizes our proposal for those elements.

4.1 Structural restrictions of MMFAP models

In case of many discrete Markov modulated arrival processes (e.g. MAP, BMAP) the representation is not unique, and starting form a given representation of an arrival process infinitely many different, but stochastically equivalent representations of the same process can be generated with similarity transformation. Based on past experience it is known that optimizing non-unique representations should be avoided, since most computational effort of the optimizers is wasted on going back and forth between almost equivalent representations having significantly different parameters. The usual solution to address this issue is to apply some structural restrictions (e.g. the Jordan representation of some of the matrices), which can make the representation unique [15].

In this work, we also apply a structural restriction to make the optimization of MMFAPs more efficient (by making the path to the optimum more straight): We restrict matrix $\mathbf{R}$ to be diagonal such that the diagonal elements of $\mathbf{R}$ are non-decreasing, which makes the representation of an MMFAP unique except for the ordering of states with identical fluid arrival rates.

4.2 Computation of the double sided inverse Laplace transform

A crucial step of the algorithm both in terms of execution speed and numerical accuracy is that to compute the numerical inverse Laplace transformation (NILT) of the expression in (3). There are many efficient numerical inverse transformation methods for single sided functions [8]. However, in our case the function we have is double sided (as Gauss distributions can be negative, too), and numerical inverse transformation of double sided Laplace transforms are rather limited.

If $f(t)$ is the density of a positive random variable then $\int_{-\infty}^{\infty} e^{-st} f(t)dt = \int_0^{\infty} e^{-st} f(t)dt$ and the single and double sided Laplace transforms of $f(t)$ are identical. If $f(t)$ is the density of a random variable which is positive with a high probability then $\int_{-\infty}^{\infty} e^{-st} f(t)dt \approx \int_0^{\infty} e^{-st} f(t)dt$. Based on this approximation one can apply single sided numerical inverse Laplace transformation for density functions of dominantly positive random variables.

For a general MMFAP the non-negativity of the fluid increase samples in $T = \{(t_k, x_k); k = 1, \dots, K\}$ can not be assumed. To make the single sided numerical inverse Laplace transformation appropriately accurate also in this case we apply the following model transformation

$$\mathcal{L}_{\mathbf{Q},\mathbf{R},\mathbf{S}}(\{(t_k, x_k); k = 1, \dots, K\}) = \mathcal{L}_{\mathbf{Q},\mathbf{R}+c\mathbf{I},\mathbf{S}}(\{(t_k, x_k + ct_k); k = 1, \dots, K\}),$$

where $\mathcal{L}_{\mathbf{Q},\mathbf{R},\mathbf{S}}(\{(t_1, x_1); (t_2, x_2); \dots; (t_K, x_K)\}) = \underline{\alpha} \prod_{k=1}^{K} \mathbf{N}(t_k, x_k)\mathbb{1}$ is the likelihood of the samples when $\mathbf{N}(t_k, x_k)$ is computed with $\mathbf{Q}, \mathbf{R}, \mathbf{S}$ according to (3) and c is an appropriate constant. If c is large enough, the relative difference of the fluid increase samples reduces and the likelihood function gets less sensitive to the changes of the model parameters. If c is small, fluid increase samples might become close to zero and the single sided numerical inverse Laplace transformation might cause numerical issues.

4.3 Reducing computational cost for equidistant measurement intervals

For computing the likelihood function, the numerical inverse Laplace transformation of matrix $\mathbf{N}^*(t, v)$ needs to be performed once for each data point, i.e. K times, which might be computationally expensive.

In the special case when the samples are from identical time intervals, that is $t_1 = \dots = t_K = \bar{t}$, we apply the following approximate approach to reduce the computational complexity to M ($M << K$) numerical inverse Laplace transformation of matrix $\mathbf{N}^*(t, v)$.

- Let $x_{min} = \min(x_1, \dots, x_K)$, $x_{max} = \max(x_1, \dots, x_K)$ and $\Delta = (x_{max} - x_{min})/M$.
- Compute $\mathbf{N}(\bar{t}, x_{min} + (m - 0.5)\Delta)$ for $m = 1, \dots, M$ by NILT of $\mathbf{N}^*(t, v)$.
- For $x \in (x_{min} + (m-1)\Delta, x_{min} + m\Delta)$, apply $\mathbf{N}(\bar{t}, x) \approx \mathbf{N}(\bar{t}, x_{min} + (m - 0.5)\Delta)$.

This way the $[x_{min}, x_{max}]$ range is divided to M equidistant intervals and the ILT is performed once for each. The higher the parameter M, the higher the accuracy, but the slower the procedure.

4.4 Computation of $E(\Theta_i^{(k)})$ and $E(M_{i,j}^{(k)})$

Let us introduce the forward and backward likelihood vectors for the beginning and the end of the kth observation period

$$\underline{\hat{f}}_k = \underline{\alpha} \left(\prod_{\ell=1}^{k-1} \mathbf{N}(t_\ell, x_\ell) \right) = \underline{\alpha} \prod_{\ell=1}^{k-1} \mathrm{ILT}_{v \to x_\ell} \mathbf{N}^*(t_\ell, v), \tag{8}$$

$$\underline{\hat{b}}_k = \left(\prod_{\ell=k+1}^{K} \mathbf{N}(t_\ell, x_\ell) \right) \mathbb{1} = \prod_{\ell=k+1}^{K} \mathrm{ILT}_{v \to x_\ell} \mathbf{N}^*(t_\ell, v) \mathbb{1}. \tag{9}$$

and the forward and backward likelihood vectors for an internal point in the kth observation period as

$$\underline{f}_k(t,x) = \underline{\alpha} \left(\prod_{\ell=1}^{k-1} \mathbf{N}(t_\ell, x_\ell) \right) \mathbf{N}(t,x), \tag{10}$$

$$\underline{b}_k(t,x) = \mathbf{N}(t,x) \left(\prod_{\ell=k+1}^{K} \mathbf{N}(t_\ell, x_\ell) \right) \mathbb{1}. \tag{11}$$

We note that, using $\underline{\hat{f}}_k$ and $\underline{\hat{b}}_k$, the likelihood can be expressed as

$$\begin{aligned} \mathcal{L} &= \underline{\alpha} \cdot \underline{b}_1(t_1, x_1) = \underline{f}_{k-1}(t_{k-1}, x_{k-1}) \cdot \underline{b}_k(t_k, x_k) = \underline{f}_K(t_K, x_K) \mathbb{1} \\ &= \underline{\alpha} \cdot \underline{\hat{b}}_0 = \underline{\hat{f}}_\ell \cdot \underline{\hat{b}}_{\ell-1} = \underline{\hat{f}}_{K+1} \mathbb{1}, \end{aligned}$$

for any $k = 2, \ldots, K-1$ and $\ell = 1, \ldots, K$.

To compute the expected value of $\Theta_i^{(k)}$, the integrals of the forward and backward likelihood vectors have to be evaluated. The special form of the integrals allows for simplifications as

$$\begin{aligned} E(\Theta_i^{(k)}) &= \int_{x=0}^{x_k} \int_{t=0}^{t_k} [\underline{f}_k(t,x)]_i \cdot [\underline{b}_k(t_k - t, x_k - x)]_i \, dt \, dx \\ &= \underline{\hat{f}}_k \left(\int_{x=0}^{x_k} \int_{t=0}^{t_k} \mathbf{N}(t,x) \underline{e_i} \cdot \underline{e_i}^T \mathbf{N}(t_k - t, x_k - x) \, dt \, dx \right) \underline{\hat{b}}_k \\ &= \underline{\hat{f}}_k \ \mathrm{ILT}_{v \to x_k} \left(\int_{t=0}^{t_k} \mathbf{N}^*(t,v) \underline{e_i} \cdot \underline{e_i}^T \mathbf{N}^*(t_k - t, v) \, dt \right) \underline{\hat{b}}_k \\ &= \underline{\hat{f}}_k \ \mathrm{ILT}_{v \to x_k} \left(\int_{t=0}^{t_k} e^{(\mathbf{Q} - v\mathbf{R} - v^2 \mathbf{\Sigma}/2)t} \underline{e_i} \cdot \underline{e_i}^T e^{(\mathbf{Q} - v\mathbf{R} - v^2 \mathbf{\Sigma}/2)(t_k - t)} \, dt \right) \underline{\hat{b}}_k \\ &= \underline{\hat{f}}_k \ \mathrm{ILT}_{v \to x_k} \left(\begin{bmatrix} \mathbf{0} \ \mathbf{I} \end{bmatrix} e^{\begin{bmatrix} \mathbf{Q} - v\mathbf{R} - v^2 \mathbf{\Sigma}/2 & \underline{e_i} \cdot \underline{e_i}^T \\ \mathbf{0} & \mathbf{Q} - v\mathbf{R} - v^2 \mathbf{\Sigma}/2 \end{bmatrix} t_k} \begin{bmatrix} \mathbf{I} \\ \mathbf{0} \end{bmatrix} \right) \underline{\hat{b}}_k. \end{aligned}$$

That is, the convolution integral is replaced by the evaluation of a matrix exponential of double size [16]. In a similar manner, the expected value of $M_{i,j}^{(k)}$

is

$$E(M_{i,j}^{(k)}) = \int_{x=0}^{x_k}\int_{t=0}^{t_k} [\underline{f}_k(t,x)]_i \cdot q_{ij} \cdot [\underline{b}_k(t_k - t, x_k - x)]_j \, dt \, dx$$
$$= q_{ij} \ \hat{\underline{f}}_k \ \mathrm{ILT}_{v\to x_k}\left(\begin{bmatrix}\mathbf{0} \ \mathbf{I}\end{bmatrix} e^{\begin{bmatrix}\mathbf{Q} - v\mathbf{R} - v^2\mathbf{\Sigma}/2 & \underline{e_i}\cdot\underline{e_j}^T \\ \mathbf{0} & \mathbf{Q} - v\mathbf{R} - v^2\mathbf{\Sigma}/2\end{bmatrix} t_k} \begin{bmatrix}\mathbf{I}\\ \mathbf{0}\end{bmatrix}\right)\hat{\underline{b}}_k.$$

We note that these expressions give an interpretation for the zth iteration of the EM method for $q_{i,j}$

$$q_{i,j}(z+1) = \frac{\sum_{k=1}^{K} \mathrm{E}\left(M_{i,j}^{(k)}\right)}{\sum_{k=1}^{K} \mathrm{E}\left(\Theta_i^{(k)}\right)} =$$
$$q_{i,j}(z) \ \frac{\hat{\underline{f}}_k \ \mathrm{ILT}_{v\to x_k}\left(\begin{bmatrix}\mathbf{0} \ \mathbf{I}\end{bmatrix} e^{\begin{bmatrix}\mathbf{Q} - v\mathbf{R} - v^2\mathbf{\Sigma}/2 & \underline{e_i}\cdot\underline{e_j}^T \\ \mathbf{0} & \mathbf{Q} - v\mathbf{R} - v^2\mathbf{\Sigma}/2\end{bmatrix} t_k} \begin{bmatrix}\mathbf{I}\\ \mathbf{0}\end{bmatrix}\right)\hat{\underline{b}}_k}{\hat{\underline{f}}_k \ \mathrm{ILT}_{v\to x_k}\left(\begin{bmatrix}\mathbf{0} \ \mathbf{I}\end{bmatrix} e^{\begin{bmatrix}\mathbf{Q} - v\mathbf{R} - v^2\mathbf{\Sigma}/2 & \underline{e_i}\cdot\underline{e_i}^T \\ \mathbf{0} & \mathbf{Q} - v\mathbf{R} - v^2\mathbf{\Sigma}/2\end{bmatrix} t_k} \begin{bmatrix}\mathbf{I}\\ \mathbf{0}\end{bmatrix}\right)\hat{\underline{b}}_k},$$

that is, $q_{i,j}(z+1)$ is the product of $q_{i,j}(z)$ and an actual guess dependent value.

4.5 Computation of $\hat{\underline{f}}_k$ and $\hat{\underline{b}}_k$

The computation of $\hat{\underline{f}}_k$ and $\hat{\underline{b}}_k$ follows a similar pattern and contains the same difficulties, except that $\hat{\underline{f}}_k$ is computed from $k = 0$ onward and $\hat{\underline{b}}_k$ is computed from $k = K$ downward. The main implementation issue with the computation of $\hat{\underline{f}}_k$ and $\hat{\underline{b}}_k$, is to avoid under-/overflow during the computation. We adopted the under-/overflow avoiding method proposed in [3].

5 Numerical examples

5.1 MMFAP simulator

For the numerical evaluation of the proposed method we developed a simulator which generates the required number (K) of traffic samples based on matrices **Q**, **R** and **S**. In each step of the simulation, the program samples the next state transition of the Markov chain and checks if it occurs before or after the next measurement instance. In the first case it samples the accumulated fluid until the next state transition and performs the state transition, in the second case it samples the accumulated fluid until the next measurement instance and maintains the state of the Markov chain.

To reduce the computational time of the fitting procedure by applying the approximate approach introduced in Section 4.3, the simulator generates the data samples such that $t_1 = \ldots = t_K = 1$.

For the MMFAP with

$$\mathbf{Q}_{slow} = \begin{bmatrix} -0.8 & 0.5 & 0.3 \\ 0.6 & -0.7 & 0.1 \\ 0.2 & 0.3 & -0.5 \end{bmatrix}, \ \mathbf{R} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 8 \end{bmatrix}, \ \mathbf{S} = \begin{bmatrix} 0.01 & 0 & 0 \\ 0 & 0.02 & 0 \\ 0 & 0 & 0.04 \end{bmatrix} \tag{12}$$

the histogram of the samples is presented in Figure 1a. The histogram indicates that the Markov chain is "slow" in this case, i.e., it stays in a single state (e.g. state i) during the measurement interval of length 1 with high probability and accumulates $\mathcal{N}(r_i, \sigma_i^2)$ distributed amount of fluid during this interval. That is the explanation of the peaks at around $r_1 = 2$, $r_2 = 4$ and $r_3 = 8$. It is also visible that the transitions between state 1 and 2 are faster than the transitions to and from state 3 and consequently, the histogram indicates fluid samples in the $x \in (2, 4)$ interval. These samples might come from measurement intervals starting from state 1 with $r_1 = 2$ and moving to state to with $r_2 = 4$, or vice versa.

To indicate the effect of the "speed" of the Markov chain on the histogram of the generated samples, Figure 1b depicts the histogram when the Markov chain is "fast", namely 10 times faster, $\mathbf{Q}_{fast} = 10\mathbf{Q}_{slow}$. The "fast" Markov chain experiences state transitions during the measurement interval with very high probability and the amount of fluid accumulated during the interval gets to be less dependent on the state of the Markov chain at the beginning of the measurement interval.

When the variance is low, as it is in this example, one can easily predict the values of the $\mathbf{R}$ matrix with the "slow" Markov chain, while for the "fast" Markov chain the values of the $\mathbf{R}$ matrix is not possible to guess based on the histogram. Still, the minimal and the maximal sample values allows to estimate the minimal and the maximal fluid rates of matrix $\mathbf{R}$.

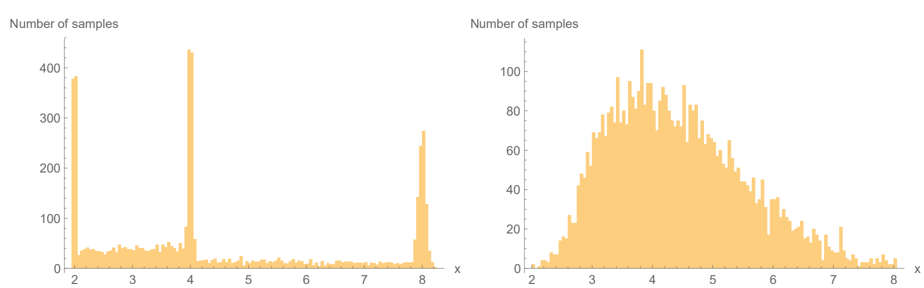

Fig. 1: Histogram of 5000 generated samples with $\mathbf{Q}_{slow}$, $\mathbf{R}$ and $\mathbf{S}$ and $\mathbf{Q}_{fast}$, $\mathbf{R}$ and $\mathbf{S}$ defined in (12).

5.2 Approximating Q with the EM method

Based on 300 samples of the MMFAP with

$$\bar{\mathbf{Q}} = \begin{bmatrix} -2 & 2 \\ 4 & -4 \end{bmatrix}, \ \bar{\mathbf{R}} = \begin{bmatrix} 4 & 0 \\ 0 & 8 \end{bmatrix}, \ \bar{\mathbf{S}} = \begin{bmatrix} 0.01 & 0 \\ 0 & 0.02 \end{bmatrix}, \tag{13}$$

we approximate the MMFAP starting from $\bar{\mathbf{Q}}_\mathbf{0} = \begin{bmatrix} -1 & 1 \\ 2.5 & -2.5 \end{bmatrix}$, $\bar{\mathbf{R}}_\mathbf{0} = \bar{\mathbf{R}}$, $\bar{\mathbf{S}}_\mathbf{0} = \bar{\mathbf{S}}$ with the EM method. The evolution of the log-likelihood value and the transition rates of the Markov chain are depicted in Figure 2a and 2b respectively, where the dotted horizontal lines refer to the MMFAP according to (13), which was used to generate the samples. The figure indicates that the obtained transition rates provide a bit higher log-likelihood than the ones in (13). Additionally, the figures report convergence after ~ 25 iterations of the EM method.

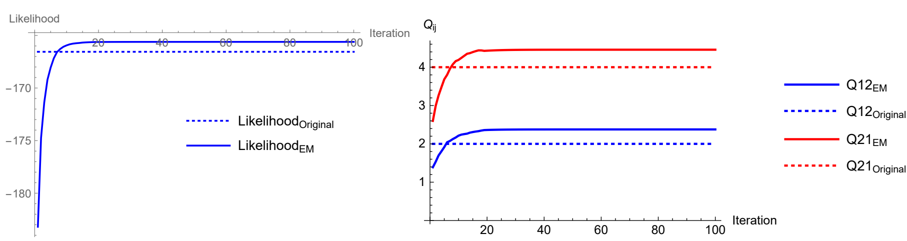

Fig. 2: Behaviour of the EM method based on 300 samples generated from $\bar{\mathbf{Q}}$, $\bar{\mathbf{R}}$ and $\bar{\mathbf{S}}$ in (13) with initial guess $\bar{\mathbf{Q}}_\mathbf{0}$, $\bar{\mathbf{R}}_\mathbf{0}$ and $\bar{\mathbf{S}}_\mathbf{0}$. According to (5) and (6), $\bar{\mathbf{R}}_\mathbf{0}$ and $\bar{\mathbf{S}}_\mathbf{0}$ remained unchanged during the EM iterations.

Similarly, we evaluated the EM based approximation of the MMFAP defined in (12) based on 1000 samples starting from $\mathbf{Q}_\mathbf{0} = \begin{bmatrix} -0.8 & 0.5 & 0.3 \\ 0.6 & -0.7 & 0.1 \\ 0.2 & 0.3 & -0.5 \end{bmatrix}$, $\mathbf{R}_\mathbf{0} = \mathbf{R}$, $\mathbf{S}_\mathbf{0} = \mathbf{S}$.

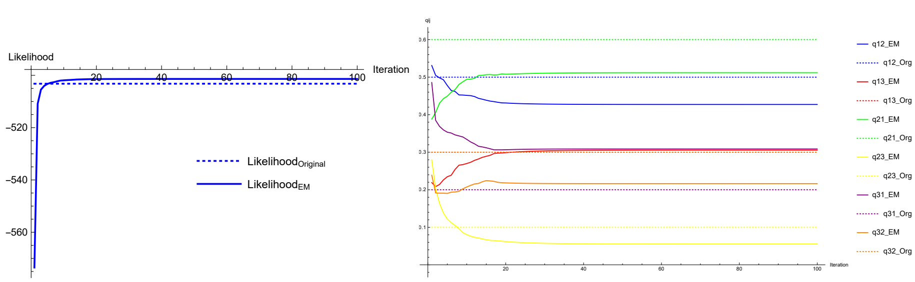

Fig. 3: Behaviour of the EM method based on 1000 samples generated from $\mathbf{Q}$, $\mathbf{R}$ and $\mathbf{S}$ in (12) with initial guess $\mathbf{Q}_\mathbf{0}$, $\mathbf{R}_\mathbf{0}$ and $\mathbf{S}_\mathbf{0}$.

The evolution of the log-likelihood value and the transition rates of the Markov chain along the EM iterations are depicted in Figure 3a and 3b respectively. Figure 3a indicates that similar to the 2×2 example in Figure 2a the likelihood value increased above the one obtained from original MMFAP. At the same time, the transition rate values in Figure 3b differ more significantly from ones of the original MMFAP than in Figure 2b, which might be a consequence of a looser relation between the transition rates and the likelihood value in higher dimensions.

6 Conclusions

The EM method is commonly applied for parameter estimation of Markov modulated models. In this paper we consider the fitting of MMFAP and recognize that the EM method is not applicable for optimizing the fluid rate and variance parameters. As a result, we investigated the properties of the EM based MMFAP method for fitting the parameters of the governing Markov chain via numerical experiments.

References

1. Anick, D., Mitra, D., Sondhi, M.M.: Stochastic theory of a data-handling system with multiple sources. Bell System technical journal **61**(8), 1871–1894 (1982)
2. Asmussen, S.: Stationary distributions for fluid flow models with or without Brownian noise. Communications in Statistics. Stochastic Models **11**(1), 21–49 (1995)
3. Bražėnas, M., Horváth, G., Telek, M.: Parallel algorithms for fitting Markov Arrival Processes. Performance Evaluation **123-124**, 50 – 67 (2018)
4. Breuer, L.: An EM algorithm for batch Markovian arrival processes and its comparison to a simpler estimation procedure. Annals of Operations Research **112**(1), 123–138 (2002)
5. Buchholz, P.: An EM-algorithm for MAP fitting from real traffic data. In: International Conference on Modelling Techniques and Tools for Computer Performance Evaluation. pp. 218–236. Springer (2003)
6. Horváth, G., Rácz, S., Telek, M.: Analysis of second-order Markov reward models. In: The International Conference on Dependable Systems and Networks, DSN/PDS 2004. pp. 845–854. IEEE CS Press, Florence, Italy (June 2004)
7. Horváth, G., Okamura, H.: A fast EM algorithm for fitting marked Markovian Arrival Processes with a new special structure. In: Computer Performance Engineering. pp. 119–133. Springer Berlin Heidelberg (2013)
8. Horváth, I., Horváth, G., Almousa, S.A.D., Telek, M.: Numerical inverse Laplace transformation using concentrated matrix exponential distributions. Performance Evaluation (2019)
9. Mitchell, K., van de Liefvoort, A.: Approximation models of feed-forward G/G/1/N queueing networks with correlated arrivals. Performance Evaluation **51**(2), 137–152 (2003)
10. Mitra, D.: Stochastic theory of a fluid model of producers and consumers coupled by a buffer. Advances in Applied Probability **20**(3), 646–676 (1988)

11. Neuts, M.F.: A versatile Markovian point process. Journal of Applied Probability **16**, 764–779 (1979)
12. Okamura, H., Dohi, T.: Faster maximum likelihood estimation algorithms for Markovian arrival processes. In: 2009 Sixth international conference on the quantitative evaluation of systems. pp. 73–82. IEEE (2009)
13. Okamura, H., Dohi, T., Trivedi, K.: Markovian arrival process parameter estimation with group data. IEEE/ACM Trans. Netw. **17**, 1326–1339 (08 2009)
14. Rydén, T.: An EM algorithm for estimation in Markov-modulated Poisson processes. Computational Statistics & Data Analysis **21**(4), 431–447 (1996)
15. Telek, M., Horváth, G.: A minimal representation of Markov arrival processes and a moments matching method. Performance Evaluation **64**(9-12), 1153–1168 (Aug 2007)
16. Van Loan, C.: Computing integrals involving the matrix exponential. IEEE Transactions on Automatic Control **23**(3), 395–404 (1978)

Appendix

A Maximizing the model parameters

Assuming $q_i = \sum_{j,j\neq i} q_{i,j}$, the derivatives of $\log \mathcal{L}$ are as follows:

$$\begin{aligned}\frac{\partial}{\partial q_{i,j}} \log \mathcal{L} &= \frac{\partial}{\partial q_{i,j}} \sum_{k=1}^{K} \left(-\Theta_i^{(k)} q_{i,j} + M_{i,j}^{(k)} \log q_{i,j}\right) \\ &= \sum_{k=1}^{K} \left(-\Theta_i^{(k)} + M_{i,j}^{(k)} \frac{1}{q_{i,j}}\right),\end{aligned}$$

$$\begin{aligned}\frac{\partial}{\partial r_i} \log \mathcal{L} &= \frac{\partial}{\partial r_i} \sum_{k=1}^{K} \left(-\Theta_i^{(k)} \frac{r_i^2}{2\sigma_i^2} + F_i^{(k)} \frac{r_i}{\sigma_i^2}\right) \\ &= \sum_{k=1}^{K} \left(-\Theta_i^{(k)} \frac{r_i}{\sigma_i^2} + F_i^{(k)} \frac{1}{\sigma_i^2}\right),\end{aligned}$$

$$\begin{aligned}\frac{\partial}{\partial \sigma_i^2} \log \mathcal{L} &= \frac{\partial}{\partial \sigma_i^2} \sum_{k=1}^{K} \left(-\Theta_i^{(k)} \frac{r_i^2}{2\sigma_i^2} + F_i^{(k)} \frac{r_i}{\sigma_i^2} - \frac{1}{2\sigma_i^2} F\Theta_i^{(k)} - M_i^{(k)} \frac{\log \sigma_i^2}{2}\right) \\ &= \frac{\partial}{\partial \sigma_i^2} \sum_{k=1}^{K} \left(\left(-\Theta_i^{(k)} r_i^2 + 2F_i^{(k)} r_i - F\Theta_i^{(k)}\right) \frac{1}{2\sigma_i^2} - M_i^{(k)} \frac{\log \sigma_i^2}{2}\right) \\ &= \sum_{k=1}^{K} \left(\left(-\Theta_i^{(k)} r_i^2 + 2F_i^{(k)} r_i - F\Theta_i^{(k)}\right) \frac{-1}{2(\sigma_i^2)^2} - M_i^{(k)} \frac{1}{2\sigma_i^2}\right).\end{aligned}$$

The optimal parameter values are obtained where the derivative is zero:

$$0 = \sum_{k=1}^{K} \Theta_i^{(k)} - M_{i,j}^{(k)}/q_{i,j} \longrightarrow q_{i,j} = \frac{\sum_{k=1}^{K} M_{i,j}^{(k)}}{\sum_{k=1}^{K} \Theta_i^{(k)}}. \tag{14}$$

$$0 = \sum_{k=1}^{K} \frac{\Theta_i^{(k)} r_i - F_i^{(k)}}{\sigma_i^2} \longrightarrow r_i = \frac{\sum_{k=1}^{K} F_i^{(k)}}{\sum_{k=1}^{K} \Theta_i^{(k)}}. \tag{15}$$

$$0 = \sum_{k=1}^{K} \left(\Theta_i^{(k)} r_i^2 - 2F_i^{(k)} r_i + F\Theta_i^{(k)} \right) \frac{1}{\sigma_i^2} - M_i^{(k)}$$

$$\longrightarrow \sigma_i^2 = \frac{\sum_{k=1}^{K} \Theta_i^{(k)} r_i^2 - 2F_i^{(k)} r_i + F\Theta_i^{(k)}}{\sum_{k=1}^{K} M_i^{(k)}}. \tag{16}$$

B Expected values of the hidden parameters

For $E(F_i^{(k)})$ we have

$$\mathrm{E}\left(F_i^{(k)}\right) = E_{\Theta_i^{(k)}} E_{F_i^{(k)}|\Theta_i^{(k)}} (F_i^{(k)}) = E_{\Theta_i^{(k)}} r_i \Theta_i^{(k)} = r_i E(\Theta_i^{(k)}). \tag{17}$$

For $F\Theta_i^{(k)} = \sum_{n=1}^{n^{(k)}} \frac{{f_n^{(k)}}^2}{\theta_n^{(k)}} \mathcal{I}_{\left\{J_n^{(k)}=i\right\}}$, we have

$$\mathrm{E}\left(F\Theta_i^{(k)}\right) = \mathrm{E}\left(\sum_{n=1}^{n^{(k)}} \frac{{f_n^{(k)}}^2}{\theta_n^{(k)}} \mathcal{I}_{\left\{J_n^{(k)}=i\right\}} \right)$$

$$= \mathrm{E}_{\{n^{(k)},\theta_1^{(k)},\ldots,\theta_{n^{(k)}}^{(k)}\}} \left(\sum_{n=1}^{n^{(k)}} \mathrm{E}_{\{f_n^{(k)}|\theta_n^{(k)}\}} \left(\frac{{f_n^{(k)}}^2}{\theta_n^{(k)}} \mathcal{I}_{\left\{J_n^{(k)}=i\right\}} \right) \right). \tag{18}$$

Let $\mathcal{X}(\mu, \sigma^2)$ denote a normal distributed random variable with mean μ and variance σ^2. Its second moment is $\mathrm{E}\left(\mathcal{X}^2(\mu,\sigma^2)\right) = \sigma^2 + \mu^2$. When $\theta_n^{(k)} = x$ then $f_n^{(k)}$ is $\mathcal{X}(xr_i, x\sigma_i^2)$ distributed and $\mathrm{E}\left({f_n^{(k)}}^2\right) = \mathrm{E}\left(\mathcal{X}^2(xr_i, x\sigma_i^2)\right) = x\sigma_i^2 + x^2 r_i^2$, that is

$$\mathrm{E}_{\{f_n^{(k)}|\theta_n^{(k)}\}} \left(\frac{{f_n^{(k)}}^2}{\theta_n^{(k)}} \mathcal{I}_{\left\{J_n^{(k)}=i\right\}} \right) = \frac{\mathrm{E}_{\{f_n^{(k)}|\theta_n^{(k)}\}} \left({f_n^{(k)}}^2\right)}{\theta_n^{(k)}} \mathcal{I}_{\left\{J_n^{(k)}=i\right\}}$$

$$= \frac{\theta_n^{(k)} \sigma_i^2 + {\theta_n^{(k)}}^2 r_i^2}{\theta_n^{(k)}} \mathcal{I}_{\left\{J_n^{(k)}=i\right\}} = \left(\sigma_i^2 + \theta_n^{(k)} r_i^2\right) \mathcal{I}_{\left\{J_n^{(k)}=i\right\}}. \tag{19}$$

Substituting (19) into (18) results

$$\mathrm{E}\left(F\Theta_i^{(k)}\right) = \mathrm{E}\left(\sum_{n=1}^{n^{(k)}} \frac{{f_n^{(k)}}^2}{\theta_n^{(k)}} \mathcal{I}_{\left\{J_n^{(k)}=i\right\}} \right)$$

$$= \mathrm{E}_{\{n^{(k)},\theta_1^{(k)},\dots,\theta_{n^{(k)}}^{(k)}\}} \left(\sum_{n=1}^{n^{(k)}} \mathrm{E}_{\{f_n^{(k)}|\theta_n^{(k)}\}} \left(\frac{{f_n^{(k)}}^2}{\theta_n^{(k)}} \mathcal{I}_{\left\{J_n^{(k)}=i\right\}} \right) \right)$$

$$= \mathrm{E}_{\{n^{(k)},\theta_1^{(k)},\dots,\theta_{n^{(k)}}^{(k)}\}} \left(\sum_{n=1}^{n^{(k)}} \left(\sigma_i^2 + \theta_n^{(k)} r_i^2 \right) \mathcal{I}_{\left\{J_n^{(k)}=i\right\}} \right)$$

$$= \mathrm{E}\left(M_i^{(k)}\right)\sigma_i^2 + \mathrm{E}\left(\Theta_i^{(k)}\right) r_i^2.$$

C Numerical computation of the expected value of the hidden parameters

In the E-step we compute the expected value of the hidden parameters for given $\underline{\alpha}$, $\mathbf{Q}$, $\mathbf{R}$, $\mathbf{S}$ and observed data (t_k, x_k) for $k = 1, \dots, K$. For the expected values of $\Theta_i^{(k)}$ we have

$$\mathrm{E}\left(\Theta_i^{(k)}|t_k,x_k\right) = \mathrm{E}\left(\sum_{n=1}^{n^{(k)}} \theta_n^{(k)} \mathcal{I}_{\left\{J_n^{(k)}=i\right\}} |t_k,x_k \right) = \mathrm{E}\left(\int_{t=0}^{t_k} \mathcal{I}_{\{\mathcal{J}(t)=i|x_k\}} dt \right) \tag{20}$$

$$= \int_{t=0}^{t_k} \mathrm{E}\left(\mathcal{I}_{\{\mathcal{J}(t)=i|x_k\}}\right) dt = \int_{t=0}^{t_k} \Pr\left(\mathcal{J}(t)=i|x_k\right) dt$$

$$= \sum_k \sum_\ell \int_{t=0}^{t_k} \Pr\left(\mathcal{J}(0)=k, \mathcal{J}(t)=i, \mathcal{J}(t_k)=\ell|x_k\right) dt$$

$$= \sum_k \sum_\ell \int_{t=0}^{t_k} \Pr\left(\mathcal{J}(0)=k\right)$$

$$\int_{x=0}^{x_k} \lim_{\Delta\to 0} \frac{1}{\Delta} \Pr\left(x \le \mathcal{X}(t) < x+\Delta, \mathcal{J}(t)=i|\mathcal{J}(0)=k, \mathcal{X}(0)=0\right)$$

$$\lim_{\Delta\to 0} \frac{1}{\Delta} \Pr\left(x_k \le \mathcal{X}(t_k) < x_k+\Delta, \mathcal{J}(t)=\ell|\mathcal{J}(t)=i, \mathcal{X}(t)=x\right) dx dt$$

$$= \underline{\alpha}_k \int_{t=0}^{t_k} \int_{x=0}^{x_k} \mathbf{N}(t,x) \underline{e_i} \underline{e_i}^T \mathbf{N}(t_k-t, x_k-x) \mathbb{1} dx dt \tag{21}$$

where the jth element of vector $\underline{\alpha}_k$ is $\Pr\left(\mathcal{J}(0)=j\right)$ and $\underline{e_i}$ is the ith unit column vector.

According to (21), (10) and (11) , the expected value of $\Theta_i^{(k)}$ is

$$E(\Theta_i^{(k)}) = \int_{x=0}^{x_k} \int_{t=0}^{t_k} [\underline{f}_k(t,x)]_i \cdot [\underline{b}_k(t_k-t, x_k-x)]_i \, dt \, dx. \tag{22}$$

In a similar manner, the expected value of $M_{i,j}^{(k)}$ is

$$E(M_{i,j}^{(k)}) = \int_{x=0}^{x_k} \int_{t=0}^{t_k} [\underline{f}_k(t,x)]_i \cdot q_{i,j} \cdot [\underline{b}_k(t_k-t, x_k-x)]_j \, dt \, dx. \tag{23}$$