Fitting methods based on distance measures of marked Markov arrival processes

Gábor Horváth[1] and Miklós Telek[2]

[1] Budapest University of Technology and Economics
Department of Networked Systems and Services
[2] MTA-BME Information Systems Research Group
Magyar Tudósok krt. 2, 1117 Budapest, Hungary
{ghorvath,telek}@hit.bme.hu

Abstract. Approximating various real world observations with stochastic processes is an essential modeling step in several fields of applied sciences. In this paper we focus on the family of Markov modulated point processes, and propose some fitting methods. The core of these methods is the computation of the distance between elements of the model family. First we introduce a methodology for computing the squared distance between the density functions of two phase-type (PH) distributions. Later we generalize this methodology for computing the distance between the joint density functions of k successive inter-arrival times of Markovian arrival processes (MAPs) and marked Markovian arrival processes (MMAPs). We also discuss the distance between the autocorrelation functions of such processes.
Based on these computable distances various versions of simple fitting procedures are introduced to approximate real world observations with the mentioned Markov modulated point processes.

1 Introduction

Phase-type (PH [17]) distributions, Markovian arrival processes (MAPs [18]), their generalizations, rational arrival processes (RAPs [1]), and their multi-type variants, MMAPs [2, 8] and MRAPs [3, 5] are versatile modeling tools in various fields of performance evaluation. If the type of an arrival event is the number of arrivals at an arrival instance then the process is commonly referred to as batch MAP (BMAP). BMAPs and MMAPs form univocally related model classes and we consider only the second one here. PH distributions represent a dense set in the field of all positive-valued distributions ([19, Theorem 8.2.8]), while MAPs represent a dense class of point processes ([2]). They are easy to work with: several important statistical properties can be expressed in a simple closed form, they exhibit many closeness properties, queues involving PH distributions and MAP/MMAP arrival and/or service processes can be solved efficiently, etc.

In the last decades considerable research effort has been spent to approximate various distributions by PH distributions, to approximate various point processes by MAPs and multi-type point processes with MMAP to take the advantage of their technical simplicity. Matching and fitting methods have been

developed to construct these Markovian modeling tools based on empirical measurement traces, or based on point processes like departure processes of queues, etc. However, the result produced by some of these procedures might not be ready for use immediately. There are situations when compactness (in terms of the number of states) and the Markovian representation is important.

In order to develop procedures to compress a PH, a MAP or a MMAP and/or to obtain a Markovian approximation of a non-Markovian representation, it is necessary to define distance functions which measure how "close" two stochastic processes of a model class are to each other. Since this distance function is evaluated repetitively in an optimization procedure, it must be reasonable efficient to evaluate.

In this paper we show that the squared distance between the density functions of two PH distributions, the joint density functions of k successive inter-arrival times of two MAPs and the joint density and type functions of k successive inter-arrival times of two MMAPs can be expressed in a closed form. Furthermore, the squared distance between the autocorrelation functions of MAPs can be expressed in a closed form as well. Based on these results a simple procedure is developed to approximate a non-Markovian representation by a Markovian one, and some further fitting procedure versions are discussed.

This paper builds on [12] and extends its applicability for the multi type version of MAPs. The rest of the paper is organized as follows. Section 2 introduces the notations and the main properties of PH distributions, MAPs and RAPs used in the paper. Section 3 presents how the distance between two PH distributions, and Section 4 how the distance between two MAPs is calculated. The most general result of the paper, the distance between marked Markovian arrival processes, is detailed in Section 5. The ME, RAP and MRAP approximation procedures are developed in Section 6. Finally, Section 7 demonstrates how the results are applied for the approximation of the departure process of a MAP/MAP/1 queue.

2 Markov modulated point processes

2.1 Phase-type distributions

A phase-type (PH) distributed random variable $\mathcal{X}$ represents the time to absorption of a transient continuous time Markov chain (CTMC). PH distributions are characterized by two parameters: the generator matrix of the transient CTMC, denoted by $\boldsymbol{B}$, and the distribution of the initial state, denoted by row vector β. The column vector of the rates to the absorbing state, can be computed as $\boldsymbol{b} = -\boldsymbol{B}\mathbb{1}$, where $\mathbb{1}$ denotes the column vector on ones. It means that the elements of $\boldsymbol{B}$, β and $\boldsymbol{b}$ comply with sign constraints. The off diagonal elements of $\boldsymbol{B}$ and all elements of β and $\boldsymbol{b}$ are non-negative, while the diagonal elements of $\boldsymbol{B}$ are negative. With these notations the density function and the moments can be expressed by

$$f(x) = \beta e^{\boldsymbol{B}x}\boldsymbol{b}, \tag{1}$$

$$E(\mathcal{X}^i) = i!\beta(-\boldsymbol{B})^{-i}\mathbb{1}. \tag{2}$$

Matrix-exponential (ME, [16]) distributions are similar to PH distributions, but the above mentioned Markovian sign constraints are relaxed. This means that vector β and matrix $\boldsymbol{B}$ can be arbitrary, the only requirement is that $f(x)$ defined by (1) must be a valid density function.

2.2 Markovian arrival processes

A Markovian Arrival Process (MAP, [18, 14]) with N phases is given by two $N \times N$ matrices, $\boldsymbol{D_0}$ and $\boldsymbol{D_1}$. The sum $\boldsymbol{D} = \boldsymbol{D_0}+\boldsymbol{D_1}$ is the generator of an irreducible CTMC with N states, which is the so called modulating or background process of the MAP. Consequently, each row sum of $\boldsymbol{D}$ is zero, that is

$$\boldsymbol{D}\mathbb{1} = (\boldsymbol{D_0} + \boldsymbol{D_1})\mathbb{1} = 0. \tag{3}$$

Matrix $\boldsymbol{D_1}$ contains the rates of those phase transitions which are accompanied by an arrival, and the off-diagonal entries of $\boldsymbol{D_0}$ are the rates of the internal phase transitions without arrival. The sign constraints for MAPs are as follows. The off diagonal elements of $\boldsymbol{D_0}$ and all elements of $\boldsymbol{D_1}$ are non-negative, while the diagonal elements of $\boldsymbol{D_0}$ are negative.

The phase process embedded at arrival instants plays an important role in the analysis of MAPs. This phase process is a discrete time Markov chain whose transition probability matrix is $\boldsymbol{P} = (-\boldsymbol{D_0})^{-1}\boldsymbol{D_1}$. The stationary probability vector of the embedded process is denoted by α, it is the unique solution to linear equations $\alpha\boldsymbol{P} = \alpha, \alpha\mathbb{1} = 1$.

The joint density function of k consecutive inter-arrival times $\mathcal{X}_1, \mathcal{X}_2, \ldots \mathcal{X}_k$ is given by

$$f_k(x_1, x_2, \ldots, x_k) = \alpha e^{\boldsymbol{D_0}x_1}\boldsymbol{D_1} \cdot e^{\boldsymbol{D_0}x_2}\boldsymbol{D_1} \cdots e^{\boldsymbol{D_0}x_k}\boldsymbol{D_1}\mathbb{1}. \tag{4}$$

For $k = 1$ we obtain that the stationary distribution of the inter-arrival time is PH distributed with α and $\boldsymbol{D_0}$.

The lag-k autocorrelation of the inter-arrival times is matrix-geometric, and can be expressed as

$$\begin{aligned}\rho_k &= \frac{E(\mathcal{X}_1\mathcal{X}_{k+1}) - E(\mathcal{X}_1)^2}{E(\mathcal{X}_1^2) - E(\mathcal{X}_1)^2} \\ &= \frac{\alpha(-\boldsymbol{D_0})^{-1}\boldsymbol{P}^k(-\boldsymbol{D_0})^{-1}\mathbb{1} - \alpha(-\boldsymbol{D_0})^{-1}\mathbb{1} \cdot \alpha(-\boldsymbol{D_0})^{-1}\mathbb{1}}{\sigma^2} \\ &= \frac{1}{\sigma^2}\alpha(-\boldsymbol{D_0})^{-1}(\boldsymbol{P} - \mathbb{1}\alpha)^k(-\boldsymbol{D_0})^{-1}\mathbb{1}\end{aligned} \tag{5}$$

for $k > 0$, and it is $\rho_0 = 1$ for $k = 0$. In (5) $\sigma^2 = E(\mathcal{X}_1^2) - E(\mathcal{X}_1)^2$ denotes the variance of the inter-arrival times, $E(\mathcal{X}_1)^i, i = 1, 2$ are obtained from (2), and $E(\mathcal{X}_1\mathcal{X}_{k+1})$ from

$$E(\mathcal{X}_1\mathcal{X}_{k+1}) = \int_{x_1} \ldots \int_{x_{k+1}} x_1 x_{k+1} f_k(x_1, x_2, \ldots, x_{k+1}) dx_{k+1} \ldots dx_1,$$

and we exploited that $\boldsymbol{P}^k - \mathbb{1}\alpha = (\boldsymbol{P} - \mathbb{1}\alpha)^k$ holds for $k > 0$ (notice however that it does not hold for $k = 0$).

Rational Arrival Processes (RAPs) are generalizations of MAPs, which do not obey the Markovian sign constraints. The $\boldsymbol{D_0}, \boldsymbol{D_1}$ matrices of RAPs can have arbitrary entries, the only restriction is that the joint density function must be non-negative.

Getting rid of the Markovian sign restrictions makes RAPs more flexible than MAPs in general, but checking that a RAP is a valid stochastic process is hard (apart from the case when the transformation to a Markovian representation is successful).

2.3 Marked Markovian arrival processes

MAPs and RAPs can be extended to generate marked point processes as well. The marked versions of these processes, that are capable of representing the arrival process of multiple types of customer, are abbreviated by MMAPs and MRAPs. If there are K arrival types, MMAPs and MRAPs are characterized by $K+1$ matrices. The interpretation of matrix $\boldsymbol{D_0}$ is the same as in the single-class case. Matrix $\boldsymbol{D_m}$, for $m = 1, \ldots, K$, holds the phase transitions that are accompanied by a type-m arrival event. The sign constraints for MMAPs are as follows. The off diagonal elements of $\boldsymbol{D_0}$ and all elements of $\boldsymbol{D_m}$, for $m = 1, \ldots, K$ are non-negative, while the diagonal elements of $\boldsymbol{D_0}$ are negative. Similar to MAPs $\boldsymbol{D} = \sum_{m=0}^{K} \boldsymbol{D_m}$ is the generator of an irreducible CTMC, which is the modulating process of the MMAP and consequently, $\boldsymbol{D}\mathbb{1} = 0$.

If $\mathcal{X}_k$ represents the kth inter-arrival time and $\mathcal{Y}_k$ the type of the kth customer, the stationary joint density function (including the type of the customers arrived) can be defined by

$$\begin{aligned} & f_k(x_1, m_1, x_2, m_2, \ldots, x_k, m_k) \\ & = \frac{d}{dx_1}\frac{d}{dx_2}\ldots\frac{d}{dx_k} P(\mathcal{X}_1 < x_1, \mathcal{Y}_1 = m_1, \mathcal{X}_2 < x_2, \mathcal{Y}_2 = m_2, \ldots, \mathcal{X}_k < x_k, \mathcal{Y}_k = m_k) \\ & = \alpha e^{\boldsymbol{D_0} x_1} \boldsymbol{D_{m_1}} \cdot e^{\boldsymbol{D_0} x_2} \boldsymbol{D_{m_2}} \cdots e^{\boldsymbol{D_0} x_k} \boldsymbol{D_{m_k}} \mathbb{1}, \end{aligned} \tag{6}$$

where α is the stationary phase distribution at arrivals, which is the solution of $\alpha(-\boldsymbol{D_0})^{-1} \sum_{m=1}^{K} \boldsymbol{D_m} = \alpha, \alpha\mathbb{1} = 1$.

3 The distance between two PH distributions

In this section we provide a simple explicit solution for the squared difference between the density functions of two PH distributions. Let us consider two PH distributed random variables, $\mathcal{A}$ and $\mathcal{B}$, with the initial probability vector, transient generator and rates to the absorbing state denoted by $(\alpha, \boldsymbol{A}, \boldsymbol{a})$ and $(\beta, \boldsymbol{B}, \boldsymbol{b})$, respectively. The squared difference between the density functions $f_{\mathcal{A}}(x)$ and

$f_{\mathcal{B}}(x)$ is defined by

$$\begin{aligned}\mathcal{D}\{\boldsymbol{\mathcal{A}},\boldsymbol{\mathcal{B}}\} &= \int_0^\infty \left(f_{\mathcal{A}}(x) - f_{\mathcal{B}}(x)\right)^2 dx = \int_0^\infty \left(\alpha e^{\boldsymbol{A}x}\boldsymbol{a} - \beta e^{\boldsymbol{B}x}\boldsymbol{b}\right)^2 dx \\ &= L(\boldsymbol{\mathcal{A}},\boldsymbol{\mathcal{A}}) - 2L(\boldsymbol{\mathcal{A}},\boldsymbol{\mathcal{B}}) + L(\boldsymbol{\mathcal{B}},\boldsymbol{\mathcal{B}}),\end{aligned} \tag{7}$$

where $L(\boldsymbol{\mathcal{A}},\boldsymbol{\mathcal{B}})$ is the integral of the product of two matrix-exponentials, i.e.

$$L(\boldsymbol{\mathcal{A}},\boldsymbol{\mathcal{B}}) = \int_0^\infty \alpha e^{\boldsymbol{A}x}\boldsymbol{a} \cdot \beta e^{\boldsymbol{B}x}\boldsymbol{b}\, dx. \tag{8}$$

The solution of this integral can be obtained by the solution of a Sylvester equation, since for compatible matrices A, B, C, $X = \int_0^\infty e^{Ax} C e^{Bx}\, dx$ satisfies $AX + BX + C = 0$ due to [15, Theorem 13.19]. Before applying this identity, we transpose the pdf of $\mathcal{B}$ in the integral. This step is not necessary here, but it will be essential in the subsequent sections. We get

$$L(\boldsymbol{\mathcal{A}},\boldsymbol{\mathcal{B}}) = \boldsymbol{b}^T \underbrace{\int_0^\infty e^{\boldsymbol{B}^T x}\beta^T\alpha e^{\boldsymbol{A}x}\, dx}_{\boldsymbol{Y}} \cdot \boldsymbol{a}, \tag{9}$$

where matrix $\boldsymbol{Y}$ is the solution to

$$-\beta^T\alpha = \boldsymbol{B}^T\boldsymbol{Y} + \boldsymbol{Y}\boldsymbol{A}. \tag{10}$$

This Sylvester equation has a unique solution if the eigenvalues of matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ have real parts in the open left half-plane ([15, Theorem 13.19]), which always holds for the transient generator of PH distributions.

By applying appropriate Kronecker operations, Sylvester equations can be transformed to traditional linear equations (of form $Ax = b$, [15, Equation 13.6]), but there are more efficient numerical solution algorithms available as well, that operate on smaller matrices by avoiding Kronecker operations. One of the fastest and most widely used direct method for solving Sylvester equations is the Hessenberg-Schur method [7], which is used by the built-in `lyap` function of Matlab at well.

4 Calculation of the distance between two MAPs

4.1 The distance between the joint density functions of two MAPs

Let us consider two MAPs, $\boldsymbol{\mathcal{A}} = (\boldsymbol{A_0}, \boldsymbol{A_1})$ and $\boldsymbol{\mathcal{B}} = (\boldsymbol{B_0}, \boldsymbol{B_1})$. The squared difference of the joint density of the inter-arrival times up to lag-k is defined similar to (7)

$$\begin{aligned}\mathcal{D}_k\{\boldsymbol{\mathcal{A}},\boldsymbol{\mathcal{B}}\} = \int_0^\infty \dots \int_0^\infty \int_0^\infty \Big(&\alpha_{\mathcal{A}} e^{\boldsymbol{A_0}x_1}\boldsymbol{A_1}\cdots e^{\boldsymbol{A_0}x_{k-1}}\boldsymbol{A_1}\cdot e^{\boldsymbol{A_0}x_k}\boldsymbol{A_1}\mathbb{1} \\ &- \alpha_{\mathcal{B}} e^{\boldsymbol{B_0}x_1}\boldsymbol{B_1}\cdots e^{\boldsymbol{B_0}x_{k-1}}\boldsymbol{B_1}\cdot e^{\boldsymbol{B_0}x_k}\boldsymbol{B_1}\mathbb{1}\Big)^2 dx_1 \dots dx_{k-1}\, dx_k,\end{aligned} \tag{11}$$

where $\alpha_{\mathcal{A}}$ and $\alpha_{\mathcal{B}}$ denote the stationary phase distributions of MAPs $\boldsymbol{\mathcal{A}}$ and $\boldsymbol{\mathcal{B}}$ at arrival instants. The square term expands to

$$\mathcal{D}_k\{\boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{B}}\} = L_k(\boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{A}}) - 2L_k(\boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{B}}) + L_k(\boldsymbol{\mathcal{B}}, \boldsymbol{\mathcal{B}}), \tag{12}$$

where $L_k(\boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{B}})$ represents the integral

$$\begin{aligned} L_k(\boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{B}}) = \int_0^\infty \dots \int_0^\infty \int_0^\infty \alpha_{\mathcal{A}} e^{\boldsymbol{A_0} x_1} \boldsymbol{A_1} \cdots e^{\boldsymbol{A_0} x_{k-1}} \boldsymbol{A_1} \cdot e^{\boldsymbol{A_0} x_k} \boldsymbol{A_1} \mathbb{1} \\ \cdot \alpha_{\mathcal{B}} e^{\boldsymbol{B_0} x_1} \boldsymbol{B_1} \cdots e^{\boldsymbol{B_0} x_{k-1}} \boldsymbol{B_1} \cdot e^{\boldsymbol{B_0} x_k} \boldsymbol{B_1} \mathbb{1} \, dx_1 \dots dx_{k-1} \, dx_k. \end{aligned} \tag{13}$$

The following theorem provides a procedure to evaluate this integral with the consecutive solutions of k Sylvester equations.

Theorem 1. *$L_k(\boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{B}})$ can be expressed by*

$$L_k(\boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{B}}) = \mathbb{1}^T {\boldsymbol{B_1}}^T \cdot \boldsymbol{Y_k} \cdot \boldsymbol{A_1} \mathbb{1}, \tag{14}$$

where matrix $\boldsymbol{Y_k}$ is defined recursively by Sylvester equations

$$\begin{cases} -{\boldsymbol{B_1}}^T \boldsymbol{Y_{k-1}} \boldsymbol{A_1} = {\boldsymbol{B_0}}^T \boldsymbol{Y_k} + \boldsymbol{Y_k} \boldsymbol{A_0} & \text{for } k > 1, \\ -\alpha_{\mathcal{B}}^T \alpha_{\mathcal{A}} = {\boldsymbol{B_0}}^T \boldsymbol{Y_1} + \boldsymbol{Y_1} \boldsymbol{A_0} & \text{for } k = 1. \end{cases} \tag{15}$$

Proof. We start by transforming (13) as

$$\begin{aligned} L_k(\boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{B}}) &= \int_0^\infty \dots \int_0^\infty \int_0^\infty \mathbb{1}^T {\boldsymbol{B_1}}^T e^{{\boldsymbol{B_0}}^T x_k} {\boldsymbol{B_1}}^T e^{{\boldsymbol{B_0}}^T x_{k-1}} \cdots {\boldsymbol{B_1}}^T e^{{\boldsymbol{B_0}}^T x_1} \alpha_{\mathcal{B}}^T \\ &\quad \cdot \alpha_{\mathcal{A}} e^{\boldsymbol{A_0} x_1} \boldsymbol{A_1} \cdots e^{\boldsymbol{A_0} x_{k-1}} \boldsymbol{A_1} \cdot e^{\boldsymbol{A_0} x_k} \boldsymbol{A_1} \mathbb{1} \, dx_1 \dots dx_{k-1} \, dx_k \\ &= \mathbb{1}^T {\boldsymbol{B_1}}^T \Bigg(\int_0^\infty \dots \int_0^\infty \int_0^\infty e^{{\boldsymbol{B_0}}^T x_k} {\boldsymbol{B_1}}^T e^{{\boldsymbol{B_0}}^T x_{k-1}} \cdots {\boldsymbol{B_1}}^T e^{{\boldsymbol{B_0}}^T x_1} \alpha_{\mathcal{B}}^T \\ &\quad \cdot \alpha_{\mathcal{A}} e^{\boldsymbol{A_0} x_1} \boldsymbol{A_1} \cdots e^{\boldsymbol{A_0} x_{k-1}} \boldsymbol{A_1} \cdot e^{\boldsymbol{A_0} x_k} \, dx_1 \dots dx_{k-1} \, dx_k \Bigg) \cdot \boldsymbol{A_1} \mathbb{1}. \end{aligned} \tag{16}$$

Let us denote the term in the parenthesis by $\boldsymbol{Y_k}$. For $k > 1$, separating the first and the last terms leads to the recursion

$$\begin{aligned} \boldsymbol{Y_k} &= \int_0^\infty e^{{\boldsymbol{B_0}}^T x_k} \cdot {\boldsymbol{B_1}}^T \Bigg(\int_0^\infty \dots \int_0^\infty e^{{\boldsymbol{B_0}}^T x_{k-1}} {\boldsymbol{B_1}}^T \cdots {\boldsymbol{B_1}}^T e^{{\boldsymbol{B_0}}^T x_1} \alpha_{\mathcal{B}}^T \\ &\quad \cdot \alpha_{\mathcal{A}} e^{\boldsymbol{A_0} x_1} \boldsymbol{A_1} \cdots e^{\boldsymbol{A_0} x_{k-1}} \boldsymbol{A_1} \, dx_1 \dots dx_{k-1} \Bigg) \boldsymbol{A_1} \cdot e^{\boldsymbol{A_0} x_k} \, dx_k \\ &= \int_0^\infty e^{{\boldsymbol{B_0}}^T x_k} {\boldsymbol{B_1}}^T \cdot \boldsymbol{Y_{k-1}} \cdot \boldsymbol{A_1} e^{\boldsymbol{A_0} x_k} \, dx_k, \end{aligned} \tag{17}$$

which is the solution of Sylvester equation $-{\boldsymbol{B_1}}^T \boldsymbol{Y_{k-1}} \boldsymbol{A_1} = {\boldsymbol{B_0}}^T \boldsymbol{Y_k} + \boldsymbol{Y_k} \boldsymbol{A_0}$. The equation for $k = 1$ is obtained similarly. □

Note that the solution of (15) is always unique as matrices $\boldsymbol{A_0}$ and $\boldsymbol{B_0}$ are subgenerators.

4.2 The distance between the lag autocorrelation functions

The squared distance between the lag autocorrelation functions of MAP $\boldsymbol{\mathcal{A}}$ and $\boldsymbol{\mathcal{B}}$ is computed by

$$\begin{aligned}\mathcal{D}_{\text{acf}}\{\boldsymbol{\mathcal{A}},\boldsymbol{\mathcal{B}}\} &= \sum_{i=0}^{\infty}(\rho_i^{(\mathcal{A})}-\rho_i^{(\mathcal{B})})^2 = \sum_{i=1}^{\infty}(\rho_i^{(\mathcal{A})}-\rho_i^{(\mathcal{B})})^2 \\ &= \sum_{i=1}^{\infty}\Big(\frac{1}{\sigma_{\mathcal{A}}^2}\alpha_{\mathcal{A}}(-\boldsymbol{A_0})^{-1}(\boldsymbol{P_{\mathcal{A}}}-\mathbb{1}\alpha_{\mathcal{A}})^i(-\boldsymbol{A_0})^{-1}\mathbb{1} \\ &\quad -\frac{1}{\sigma_{\mathcal{B}}^2}\alpha_{\mathcal{B}}(-\boldsymbol{B_0})^{-1}(\boldsymbol{P_{\mathcal{B}}}-\mathbb{1}\alpha_{\mathcal{B}})^i(-\boldsymbol{B_0})^{-1}\mathbb{1}\Big)^2,\end{aligned} \tag{18}$$

where $\sigma_{\mathcal{A}}^2$ ($\sigma_{\mathcal{B}}^2$) denotes the variance of the inter-arrival times of MAP $\boldsymbol{\mathcal{A}}$ ($\boldsymbol{\mathcal{B}}$), respectively, and we utilized that $\rho_i^{(\mathcal{A})} = \rho_i^{(\mathcal{B})} = 0$. Expanding the square term leads to

$$\begin{aligned}\mathcal{D}_{\text{acf}}\{\boldsymbol{\mathcal{A}},\boldsymbol{\mathcal{B}}\} &= \frac{1}{\sigma_{\mathcal{A}}^4}\Big(M(\boldsymbol{\mathcal{A}},\boldsymbol{\mathcal{A}}) - {m_2^{(A)}}^2/4\Big) \\ &\quad -2\frac{1}{\sigma_{\mathcal{A}}^2\sigma_{\mathcal{B}}^2}\Big(M(\boldsymbol{\mathcal{A}},\boldsymbol{\mathcal{B}}) - m_2^{(A)}m_2^{(B)}/4\Big) \\ &\quad +\frac{1}{\sigma_{\mathcal{B}}^4}\Big(M(\boldsymbol{\mathcal{B}},\boldsymbol{\mathcal{B}}) - {m_2^{(B)}}^2/4\Big),\end{aligned} \tag{19}$$

where $m_2^{(A)}$ and $m_2^{(B)}$ denote the second moment of the inter-arrival times of MAP $\boldsymbol{\mathcal{A}}$ and $\boldsymbol{\mathcal{B}}$, while matrix $M(\boldsymbol{\mathcal{A}},\boldsymbol{\mathcal{B}})$ represents the sum

$$\begin{aligned}M(\boldsymbol{\mathcal{A}},\boldsymbol{\mathcal{B}}) = \sum_{i=0}^{\infty}&\alpha_{\mathcal{A}}(-\boldsymbol{A_0})^{-1}(\boldsymbol{P_{\mathcal{A}}}-\mathbb{1}\alpha_{\mathcal{A}})^i(-\boldsymbol{A_0})^{-1}\mathbb{1}\cdot \\ &\cdot\alpha_{\mathcal{B}}(-\boldsymbol{B_0})^{-1}(\boldsymbol{P_{\mathcal{B}}}-\mathbb{1}\alpha_{\mathcal{B}})^i(-\boldsymbol{B_0})^{-1}\mathbb{1}.\end{aligned} \tag{20}$$

The terms involving the second moments in (19) are necessary since the sum goes from $i = 1$ in (18) and it goes from $i = 0$ in (20). Term 0 of $M(\boldsymbol{\mathcal{A}},\boldsymbol{\mathcal{B}})$ equals $m_2^{(A)}/2 \cdot m_2^{(B)}/2$.

The next theorem provides the solution of matrix $M(\boldsymbol{\mathcal{A}},\boldsymbol{\mathcal{B}})$.

Theorem 2. *Matrix $M(\boldsymbol{\mathcal{A}},\boldsymbol{\mathcal{B}})$ is obtained by*

$$M(\boldsymbol{\mathcal{A}},\boldsymbol{\mathcal{B}}) = \alpha_{\mathcal{A}}(-\boldsymbol{A_0})^{-1}\cdot\boldsymbol{X}\cdot(-\boldsymbol{B_0})^{-1}\mathbb{1}, \tag{21}$$

where $\boldsymbol{X}$ is the unique solution to the discrete Sylvester equation

$$(\boldsymbol{P_{\mathcal{A}}}-\mathbb{1}\alpha_{\mathcal{A}})\cdot\boldsymbol{X}\cdot(\boldsymbol{P_{\mathcal{B}}}-\mathbb{1}\alpha_{\mathcal{B}}) - \boldsymbol{X} + (-\boldsymbol{A_0})^{-1}\mathbb{1}\alpha_{\mathcal{B}}(-\boldsymbol{B_0})^{-1} = \boldsymbol{0}. \tag{22}$$

Proof. Matrices $\boldsymbol{P_{\mathcal{A}}}-\mathbb{1}\alpha_{\mathcal{A}}$ and $\boldsymbol{P_{\mathcal{B}}}-\mathbb{1}\alpha_{\mathcal{B}}$ are stable (eigenvalues are inside the unit disk), since the subtraction of $\mathbb{1}\alpha_{\mathcal{A}}$ and $\mathbb{1}\alpha_{\mathcal{B}}$ removes the eigenvalue of 1 which matrices $\boldsymbol{P_{\mathcal{A}}}$ and $\boldsymbol{P_{\mathcal{B}}}$ originally had. Hence we can utilize that the solution of the sum $X = \sum_{i=0}^{\infty}A^iCB^i$ satisfies the discrete Sylvester equation $AXB - X + C = 0$. □

5 Calculation of the distance between two MMAPs

The procedure presented in Section 4.1 can be extended for marked MAPs as well. If the number of arrival types is K, the difference between MMAPs up to lag-k is defined by

$$\begin{aligned}\mathcal{D}_k\{\boldsymbol{\mathcal{A}},\boldsymbol{\mathcal{B}}\} = \sum_{m_1=1}^{K}\cdots\sum_{m_{k-1}=1}^{K}\sum_{m_k=1}^{K}\int_0^\infty\cdots\int_0^\infty\int_0^\infty \\ \Big(\alpha_{\mathcal{A}}e^{\boldsymbol{A_0}x_1}\boldsymbol{A_{m_1}}\cdots e^{\boldsymbol{A_0}x_{k-1}}\boldsymbol{A_{m_{k-1}}}\cdot e^{\boldsymbol{A_0}x_k}\boldsymbol{A_{m_k}}\mathbb{1} \\ -\alpha_{\mathcal{B}}e^{\boldsymbol{B_0}x_1}\boldsymbol{B_{m_1}}\cdots e^{\boldsymbol{B_0}x_{k-1}}\boldsymbol{B_{m_{k-1}}}\cdot e^{\boldsymbol{B_0}x_k}\boldsymbol{B_{m_k}}\mathbb{1}\Big)^2 \\ dx_1\ldots dx_{k-1}\,dx_k,\end{aligned} \tag{23}$$

thus the squared distance is summed up for all combinations of arrival types up to lag-k. The expansion of the square term leads to a form similar to (12), but the $L_k(\boldsymbol{\mathcal{A}},\boldsymbol{\mathcal{B}})$ matrices are a bit more complicated due to the different arrival types. Hence,

$$\begin{aligned}L_k(\boldsymbol{\mathcal{A}},\boldsymbol{\mathcal{B}}) = \sum_{m_1=1}^{K}\cdots\sum_{m_{k-1}=1}^{K}\sum_{m_k=1}^{K}\int_0^\infty\cdots\int_0^\infty\int_0^\infty \\ \alpha_{\mathcal{A}}e^{\boldsymbol{A_0}x_1}\boldsymbol{A_{m_1}}\cdots e^{\boldsymbol{A_0}x_{k-1}}\boldsymbol{A_{m_{k-1}}}\cdot e^{\boldsymbol{A_0}x_k}\boldsymbol{A_{m_k}}\mathbb{1} \\ \cdot\,\alpha_{\mathcal{B}}e^{\boldsymbol{B_0}x_1}\boldsymbol{B_{m_1}}\cdots e^{\boldsymbol{B_0}x_{k-1}}\boldsymbol{B_{m_{k-1}}}\cdot e^{\boldsymbol{B_0}x_k}\boldsymbol{B_{m_k}}\mathbb{1}\,dx_1\ldots dx_{k-1}\,dx_k.\end{aligned} \tag{24}$$

The multi-type (marked) counterpart of Theorem 1 is as follows.

Theorem 3. *$L_k(\boldsymbol{\mathcal{A}},\boldsymbol{\mathcal{B}})$ can be expressed by*

$$L_k(\boldsymbol{\mathcal{A}},\boldsymbol{\mathcal{B}}) = \sum_{m=1}^{K}\mathbb{1}^T{\boldsymbol{B_m}}^T\cdot\boldsymbol{Y_k}\cdot\boldsymbol{A_m}\mathbb{1}, \tag{25}$$

where matrix $\boldsymbol{Y_k}$ is the solution of the recursive Sylvester equation

$$\begin{cases}-\sum_{m=1}^{K}{\boldsymbol{B_m}}^T\boldsymbol{Y_{k-1}}\boldsymbol{A_m} = {\boldsymbol{B_0}}^T\boldsymbol{Y_k}+\boldsymbol{Y_k}\boldsymbol{A_0} & \text{for } k>1,\\ -\alpha_{\mathcal{B}}^T\alpha_{\mathcal{A}} = {\boldsymbol{B_0}}^T\boldsymbol{Y_1}+\boldsymbol{Y_1}\boldsymbol{A_0} & \text{for } k=1.\end{cases} \tag{26}$$

Proof. The steps to prove the theorem are the same as the ones for Theorem 1. □

Note that the only difference between Theorem 1 and Theorem 3 is the summation in the Sylvester equation providing matrix $\boldsymbol{Y_k}$ over the different arrival types.

6 Application: Approximating a non-Markovian representation with a Markovian one

Having results for measuring the distance between two PH distributions, MAPs or MMAPs can be useful in many situations. In this section we use them as distance functions in an optimization problem. A simple procedure is developed to obtain a Markovian representation (PH/MAP/MMAP) that approximates the behavior of a given non-Markovian (ME/RAP/MRAP) one. Two possible applications of this procedure are as follows.

- Several matching procedures produce an ME distribution, a RAP, or a MRAP which does not have a Markovian representation, or which is not even a valid stochastic process (the joint density is negative at some points). The presented procedure returns a valid Markovian representation that closely approximates the target one.
- Several performance models involve huge PH distributions, MAPs or MMAPs which make the analysis too slow and numerically demanding. With the presented procedure it is possible to compress these large models by constructing small approximate replacements that are easier (feasible) to work with.

6.1 Approximating an ME distribution with a PH one

To approximate a possibly non-Markovian or too large ME distribution $\mathcal{A} = (\alpha, \boldsymbol{A}, \boldsymbol{a})$ with a Markovian PH distribution $\mathcal{B} = (\beta, \boldsymbol{B}, \boldsymbol{b})$, the following non-linear optimization problem has to be solved:

$$\min_{\beta, \boldsymbol{B}} \mathcal{D}\{\mathcal{A}, \mathcal{B}\}, \tag{27}$$

subject to

$$\begin{aligned} \beta &\geq 0, \\ \beta \mathbb{1} &= 1, \\ \boldsymbol{B}\mathbb{1} &\leq 0, \\ [\boldsymbol{B}]_{ij} &\geq 0, \text{for } i \neq j. \end{aligned} \tag{28}$$

As the representation of ME distributions given by the initial probability vector and transient generator is known to be redundant, this optimization problem has infinitely many global optimum (in general).

To test this method, we calculated 7 marginal moments of a measurement trace containing inter-arrival times of real data traffic[3], and created an ME

[3] We used the BC-pAug89 trace, `http://ita.ee.lbl.gov/html/contrib/BC.html`. While this is a fairly old trace, it is often used for testing fitting methods, it became like a benchmark.

distribution by moment matching based on [22]. The procedure in [21] failed to transform the resulting ME distribution to a PH representation in an exact way, thus the method introduced here became relevant.

The non-linear optimization problem has been solved with the built-in non-linear solver of MATLAB, called `fmincon`. Despite of the non-linearity of the problem and the existence of multiple optimum solutions, the optimization terminated quickly and the result turned out to be relatively independent on the initial guess.

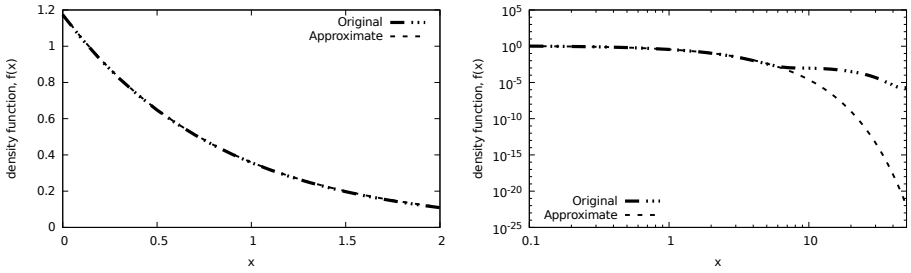

Fig. 1. Comparison of the density functions

The density functions of the original and the approximating PH distributions are depicted in Figure 1. The bodies of the density functions are very close to each other, but the log-log plot reveals a drawback of the squared distance based fitting. Namely, that the optimization sacrifices the fitting of the tiny tail densities in order to improve the distance of the body densities.

6.2 Approximating a RAP with a MAP

Throughout this section the target RAP is denoted by $\boldsymbol{\mathcal{A}} = (\boldsymbol{A_0}, \boldsymbol{A_1})$ and the approximating MAP by $\boldsymbol{\mathcal{B}} = (\boldsymbol{B_0}, \boldsymbol{B_1})$.

6.2.1 Obtaining matrix $\boldsymbol{B_1}$ given that $\alpha_{\mathcal{B}}$ and $\boldsymbol{B_0}$ are known

Given that $\alpha_{\mathcal{B}}$ and $\boldsymbol{B_0}$ are already available (see later in Section 6.2.2) matrix $\boldsymbol{B_1}$ is obtained

- either to minimize $\mathcal{D}_k\{\boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{B}}\}$ up to a given k,
- or to minimize $\mathcal{D}_{\text{acf}}\{\boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{B}}\}$.

According to the following theorem, optimizing the squared distance of the lag-1 joint density function $\mathcal{D}_2\{\boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{B}}\}$ is especially efficient.

Theorem 4. *Given that $\alpha_{\mathcal{B}}$ and $\boldsymbol{B_0}$ are available, matrix $\boldsymbol{B_1}$ minimizing $\mathcal{D}_2\{\boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{B}}\}$ is the solution of the quadratic program*

$$\min_{\boldsymbol{B_1}} \Big\{ vec\langle \boldsymbol{B_1} \rangle^T (\boldsymbol{W_{BB}} \otimes \boldsymbol{Y_{BB}}) vec\langle \boldsymbol{B_1} \rangle - 2 vec\langle \boldsymbol{A_1} \rangle^T (\boldsymbol{W_{AB}} \otimes \boldsymbol{Y_{AB}}) vec\langle \boldsymbol{B_1} \rangle \Big\} \quad (29)$$

subject to

$$\left(\boldsymbol{I} \otimes \alpha_{\mathcal{B}} (-\boldsymbol{B_0})^{-1}\right) vec\langle \boldsymbol{B_1} \rangle = \alpha_{\mathcal{A}}, \quad (30)$$

$$(\mathbb{1}^T \otimes \boldsymbol{I}) vec\langle \boldsymbol{B_1} \rangle = -\boldsymbol{B_0}\mathbb{1}. \quad (31)$$

Matrices $\boldsymbol{W_{AB}}, \boldsymbol{W_{BB}}, \boldsymbol{Y_{AB}}$ and $\boldsymbol{Y_{BB}}$ are the solutions to Sylvester equations

$$\boldsymbol{A_0} \boldsymbol{W_{AB}} + \boldsymbol{W_{AB}} \boldsymbol{B_0}^T = -\boldsymbol{A_0}\mathbb{1} \cdot \mathbb{1}^T \boldsymbol{B_0}^T, \quad (32)$$

$$\boldsymbol{B_0} \boldsymbol{W_{BB}} + \boldsymbol{W_{BB}} \boldsymbol{B_0}^T = -\boldsymbol{B_0}\mathbb{1} \cdot \mathbb{1}^T \boldsymbol{B_0}^T, \quad (33)$$

$$\boldsymbol{A_0}^T \boldsymbol{Y_{AB}} + \boldsymbol{Y_{AB}} \boldsymbol{B_0} = -\alpha_{\mathcal{A}}^T \cdot \alpha_{\mathcal{B}}, \quad (34)$$

$$\boldsymbol{B_0}^T \boldsymbol{Y_{BB}} + \boldsymbol{Y_{BB}} \boldsymbol{B_0} = -\alpha_{\mathcal{B}}^T \cdot \alpha_{\mathcal{B}}. \quad (35)$$

Proof. Let us first apply the vec$\langle\rangle$ (column stacking) operator on (14) at $k = 2$. Utilizing the identity $\text{vec}\langle AXB\rangle = (B^T \otimes A)\text{vec}\langle X\rangle$ for compatible matrices A, B, X and the identity $\text{vec}\langle u^T v\rangle = (v^T \otimes u^T)$ for row vectors u and v (see [20]). We get

$$\text{vec}\langle L_2(\boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{B}})\rangle = (\mathbb{1}^T \boldsymbol{A_0}^T \otimes \mathbb{1}^T \boldsymbol{B_0}^T) \cdot \text{vec}\langle \boldsymbol{Y_2}\rangle = \text{vec}\langle \boldsymbol{B_0}\mathbb{1} \cdot \mathbb{1}^T \boldsymbol{A_0}^T\rangle^T \cdot \text{vec}\langle \boldsymbol{Y_2}\rangle. \quad (36)$$

Applying the vec$\langle\rangle$ operator on both sides of (15) and using $\text{vec}\langle AXB\rangle = (B^T \otimes A)\text{vec}\langle X\rangle$ again leads to

$$-(\boldsymbol{I} \otimes \boldsymbol{B_1}^T \boldsymbol{Y_1})\text{vec}\langle \boldsymbol{A_1}\rangle = (\boldsymbol{I} \otimes \boldsymbol{B_0}^T)\text{vec}\langle \boldsymbol{Y_2}\rangle + (\boldsymbol{A_0}^T \otimes \boldsymbol{I})\text{vec}\langle \boldsymbol{Y_2}\rangle, \quad (37)$$

from which vec$\langle \boldsymbol{Y_2}\rangle$ is expressed by

$$\text{vec}\langle \boldsymbol{Y_2}\rangle = (-\boldsymbol{A_0}^T \oplus \boldsymbol{B_0}^T)^{-1} (\boldsymbol{I} \otimes \boldsymbol{B_1}^T)(\boldsymbol{I} \otimes \boldsymbol{Y_{AB}})\text{vec}\langle \boldsymbol{A_1}\rangle, \quad (38)$$

since $\boldsymbol{Y_1} = \boldsymbol{Y_{AB}}$. Thus we have

$$\text{vec}\langle L_2(\boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{B}})\rangle = \underbrace{\text{vec}\langle \boldsymbol{B_0}\mathbb{1} \cdot \mathbb{1}^T \boldsymbol{A_0}^T\rangle^T (-\boldsymbol{A_0}^T \oplus \boldsymbol{B_0}^T)^{-1}}_{\text{vec}\langle \boldsymbol{W_{AB}}\rangle^T} (\boldsymbol{I} \otimes \boldsymbol{B_1}^T)(\boldsymbol{I} \otimes \boldsymbol{Y_{AB}})\text{vec}\langle \boldsymbol{A_1}\rangle, \quad (39)$$

where we recognized that the transpose of vec$\langle \boldsymbol{W_{AB}}\rangle$ expressed from (32) matches the first two terms of the expression. Using the identities of the vec$\langle\rangle$ operator yields

$$\text{vec}\langle \boldsymbol{W_{AB}}\rangle^T (\boldsymbol{I} \otimes \boldsymbol{B_1}^T) = \text{vec}\langle \boldsymbol{B_1}^T \boldsymbol{W_{AB}}\rangle^T = \text{vec}\langle \boldsymbol{B_1}\rangle^T (\boldsymbol{W_{AB}} \otimes \boldsymbol{I}). \quad (40)$$

Finally, putting together (39) and (40) gives

$$\text{vec}\langle L_2(\boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{B}})\rangle = \text{vec}\langle \boldsymbol{B_1}\rangle^T (\boldsymbol{W_{AB}} \otimes \boldsymbol{Y_{AB}})\text{vec}\langle \boldsymbol{A_1}\rangle. \quad (41)$$

From the components of $\mathcal{D}_2\{\boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{B}}\}$ (see (12)) $L_2(\boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{A}})$ plays no role in the optimization as it does not depend on $\boldsymbol{B_1}$, the term $L_2(\boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{B}})$ yields the linear term in (29) according to (41), and $L_2(\boldsymbol{\mathcal{B}}, \boldsymbol{\mathcal{B}})$ introduces the quadratic term, based on (41) after replacing $\boldsymbol{\mathcal{A}}$ by $\boldsymbol{\mathcal{B}}$.

According to the first constraint (30) and the second constraint (31) the solution must satisfy $\alpha_{\mathcal{B}}(-\boldsymbol{B_0})^{-1}\boldsymbol{B_1} = \alpha_{\mathcal{B}}$ and $\boldsymbol{B_1}\mathbb{1} = -\boldsymbol{B_0}\mathbb{1}$, respectively. □

Theorem 5. *Matrix $\boldsymbol{W_{BB}} \otimes \boldsymbol{Y_{BB}}$ is positive definite, thus the quadratic optimization problem of Theorem 4 is convex.*

Proof. If $\boldsymbol{W_{BB}}$ and $\boldsymbol{Y_{BB}}$ are positive definite, then their Kronecker product is positive definite as well. First we show that matrix $\boldsymbol{Y_{BB}}$ is positive definite, thus $z\boldsymbol{Y_{BB}}z^T > 0$ holds for any non-zero row vector z. Since $\boldsymbol{Y_{BB}}$ is the solution of a Sylvester equation, we have that $\boldsymbol{Y_{BB}} = \int_0^\infty e^{\boldsymbol{B_0}^T x}\alpha_{\mathcal{B}}^T \cdot \alpha_{\mathcal{B}} e^{\boldsymbol{B_0}x}\,dx$. Hence

$$z\boldsymbol{Y_{BB}}z^T = \int_0^\infty ze^{\boldsymbol{B_0}^T x}\alpha_{\mathcal{B}}^T \cdot \alpha_{\mathcal{B}} e^{\boldsymbol{B_0}x}z^T\,dx = \int_0^\infty \left(\alpha_{\mathcal{B}} e^{\boldsymbol{B_0}x}z^T\right)^2\,dx, \qquad (42)$$

which can not be negative, furthermore, apart from a finite number of x values $\alpha_{\mathcal{B}} e^{\boldsymbol{B_0}x}z^T$ can not be zero either. Thus, the integral is always strictly positive.

The positive definiteness of matrix $\boldsymbol{W_{BB}}$ can be proven similarly. □

Being able to formalize the optimization of $\mathcal{D}_2\{\boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{B}}\}$ as a quadratic programming problem means that obtaining the optimal matrix $\boldsymbol{B_1}$ is efficient: it is fast, and there is a single optimum which is always found.

If we intend to take higher lag joint density differences also into account, the objective function is $\mathcal{D}_k\{\boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{B}}\}$, which is not quadratic for $k > 2$. However, our numerical experience indicates that the built-in non-linear optimization tool in MATLAB, called `fmincon` is able to return the solution matrix $\boldsymbol{B_1}$ quickly, independent of the initial point of the optimization. We have a strong suspicion that the returned solution is the global optimum, however we can not prove the convexity of the objective function formally.

It is also possible to use $\mathcal{D}_{acf}\{\boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{B}}\}$ as the objective function of the optimization problem, when looking for matrix $\boldsymbol{B_1}$ that minimizes the squared difference of the autocorrelation function. We found that `fmincon` is rather prone to the initial point in this case. Repeated running with different random initial points was required to obtain the best solution.

6.2.2 Approximating a RAP

The proposed procedure consists of two steps:

1. obtaining the phase-type (PH) representation of the stationary inter-arrival times, that provides vector $\alpha_{\mathcal{B}}$ and matrix $\boldsymbol{B_0}$;
2. obtaining the optimal $\boldsymbol{B_1}$ matrix which minimizes the distance of the correlation structure with the target RAP.

Section 6.2.1 describes step 2. For step 1, any phase-type fitting method can be applied. To solve this problem [6] develops a moment matching method that returns a hyper-exponential distribution of order N based on $2N-1$ moments, if it is possible. An other solution published in [13] is based on a hyper-Erlang distribution, which always succeeds if an appropriately large Erlang order is chosen.

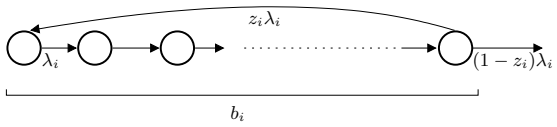

Fig. 2. Structure of a feedback Erlang block

Our method of choice, however, is a slight modification of [10], which is the generalization of the former two. It constructs PH distributions from feedback Erlang blocks (FEBs, see Figure 2), where each FEB implements an eigenvalue of the target distribution. An FEB of a single state represents areal eigenvalue. With FEBs it is possible to represent complex eigenvalues as well, as opposed to the previously mentioned methods that operate on hyper-exponential and hyper-Erlang distributions. The original method in [10] puts the FEBs in a row (as it is in Figure 3), which is not appropriate for our goals, since there is only a single absorbing state, implying that matrix $\boldsymbol{B_1}$ can have only a single non-zero row, thus no correlation can be realized. However, the original method can be modified in a straight forward way to return a hyper-FEB structure (as it is in Figure 4). A key step of [10] is the solution of a polynomial system of equations, which can have several solutions, providing several valid $\alpha_{\mathcal{B}}, \boldsymbol{B_0}$ pairs. Our RAP approximation procedure performs the optimization of matrix $\boldsymbol{B_1}$ with all of these solutions, and picks the best one among them.

6.2.3 Numerical examples

In the first numerical example we extract 7 marginal moments and 9 lag-1 joint moments from the measurement trace used in Section 6.1, and create a RAP of order 4 with the method published in [21]. The obtained matrices are as follows:

$$\boldsymbol{A_0} = \begin{bmatrix} -0.579 & -0.402 & -0.364 & -0.348 \\ -0.368 & -0.205 & -0.315 & -0.36 \\ 1.32 & -0.845 & 0.701 & 1.13 \\ -1.7 & 0.3 & -1.14 & -1.52 \end{bmatrix}, \quad \boldsymbol{A_1} = \begin{bmatrix} 0.576 & 0.262 & 0.41 & 0.446 \\ 0.168 & 0.501 & 0.313 & 0.266 \\ 0.29 & -1.69 & -0.598 & -0.302 \\ 0.292 & 1.94 & 1.03 & 0.786 \end{bmatrix}.$$

The RAP characterized by $\boldsymbol{\mathcal{A}} = (\boldsymbol{A_0}, \boldsymbol{A_1})$ is, however, not a valid stochastic process as the joint density given by (4) is negative since $f_2(0.5, 8) = -0.000357$. This RAP is the target of our approximation in this section.

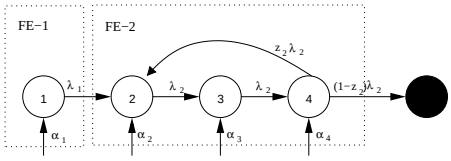

Fig. 3. PH distribution composed of serial FEBs

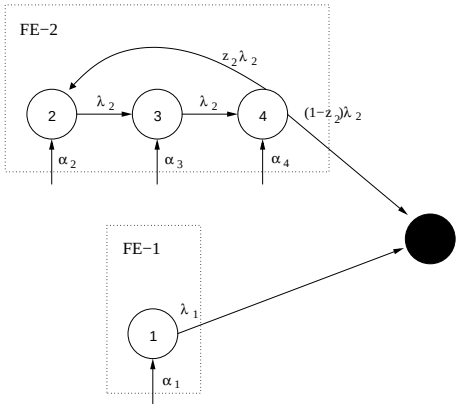

Fig. 4. PH distribution composed of parallel FEBs

Let us now construct a MAP $\boldsymbol{\mathcal{B}^{(1)}} = (\boldsymbol{B_0^{(1)}}, \boldsymbol{B_1^{(1)}})$ which minimizes the squared distance of the lag-1 joint density with $\boldsymbol{\mathcal{A}}$. The distribution of the inter-arrival times, characterized by $\alpha_{\mathcal{B}}, \boldsymbol{B_0^{(1)}}$ are obtained by the modified moment matching method of [10], and matrix $\boldsymbol{B_1^{(1)}}$ has been determined by the quadratic program provided by Theorem 4. The matrices of the MAP are

$$\boldsymbol{B_0^{(1)}} = \begin{bmatrix} -0.074 & 0 & 0 & 0 & 0 \\ 0 & -0.27 & 0.27 & 0 & 0 \\ 0 & 0 & -0.27 & 0.27 & 0 \\ 0 & 0 & 0 & -0.27 & 0 \\ 0 & 0 & 0 & 0 & -1.2 \end{bmatrix}, \boldsymbol{B_1^{(1)}} = \begin{bmatrix} 0.0065 & 0.024 & 0 & 5.5{\cdot}10^{-8} & 0.044 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0.017 & 0.086 & 0 & 0 & 0.17 \\ 0 & 0.012 & 0 & 0 & 1.2 \end{bmatrix},$$

and the squared distance in the lag-1 joint pdf is $\mathcal{D}_2\{\boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{B}^{(1)}}\} = 0.000105$. The quadratic program has been solved by MATLAB is less than a second. Next, we repeat the same procedure, but instead of focusing on the lag-1 distance, we optimize on the squared distance of the joint pdf up to lag-10. This can not be formalized as a quadratic program any more, but the optimization is still fast, lasting only 1-2 seconds. In this case the hyper-exponential distribution provided the best results ($\mathcal{D}_{11}\{\boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{B}^{(10)}}\} = 4.37 \cdot 10^{-5}$). The matrices are

$$\boldsymbol{B_0^{(10)}} = \begin{bmatrix} -0.0519 & 0 & 0 \\ 0 & -0.151 & 0 \\ 0 & 0 & -1.24 \end{bmatrix}, \quad \boldsymbol{B_1^{(10)}} = \begin{bmatrix} 10^{-6} & 0.0519 & 10^{-6} \\ 10^{-6} & 0.151 & 0.000465 \\ 0.000129 & 10^{-6} & 1.24 \end{bmatrix}.$$

To evaluate the quality of the approximation Figure 5 compares the marginal density functions of $\boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{B}^{(1)}}$ and $\boldsymbol{\mathcal{B}^{(10)}}$. The plots are close to each other, the approximation is relatively accurate. To demonstrate that the lag-1 joint densities are also accurate, Figure 6 depicts them at $x_2 = 0.5, 1$ and 1.5.

In the next experiment the objective is the squared distance of the lag-k autocorrelation function. As before, the input RAP is $\boldsymbol{\mathcal{A}}$, but now the approximation procedure has to minimize $\mathcal{D}_{\text{acf}}\{\boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{B}^{(\rho)}}\}$ which is given in a closed form

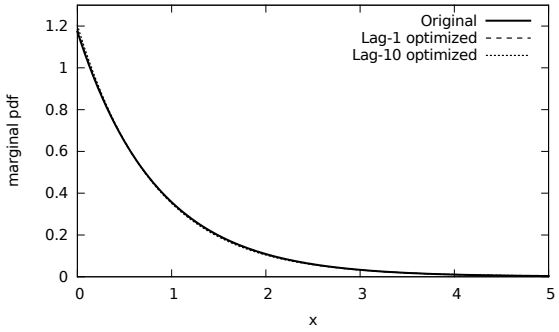

Fig. 5. Comparison of the density functions of the marginal distribution

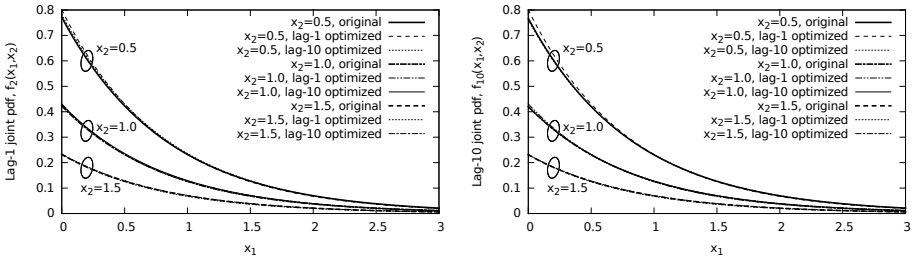

Fig. 6. Comparison of the lag-1 joint density functions

by (19) and Theorem 2. According to our experience the result of the optimization is rather prone to the initial point. The best result from 10 trials is given by matrices

$$\boldsymbol{B_0^{(\rho)}} = \begin{bmatrix} -0.0851 & 0.0851 & 0 & 0 & 0 & 0 \\ 0 & -0.0851 & 0 & 0 & 0 & 0 \\ 0 & 0 & -0.267 & 0.267 & 0 & 0 \\ 0 & 0 & 0 & -0.267 & 0.267 & 0 \\ 0 & 0 & 0 & 0 & -0.267 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1.2 \end{bmatrix}, \boldsymbol{B_1^{(\rho)}} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.0485 & 0 & 0 & 0.0366 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.0965 & 0 & 0 & 0.1705 \\ 0.0004 & 0 & 0.0117 & 0 & 0 & 1.1885 \end{bmatrix}.$$

and the corresponding autocorrelation function is depicted in Figure 7. The squared distance between the autocorrelation functions is $\mathcal{D}_{\text{acf}}\{\boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{B}^{(\rho)}}\} = 0.00237$.

6.3 Approximating a MRAP with a MMAP

If the number of arrival types K is greater than 1, the idea presented in Section 6.2.1 to formalize the approximation as a quadratic optimization problem can not be applied. To obtain the matrices $\boldsymbol{B_m}, m = 1, \ldots, K$, assuming that $\boldsymbol{B_0}$ is

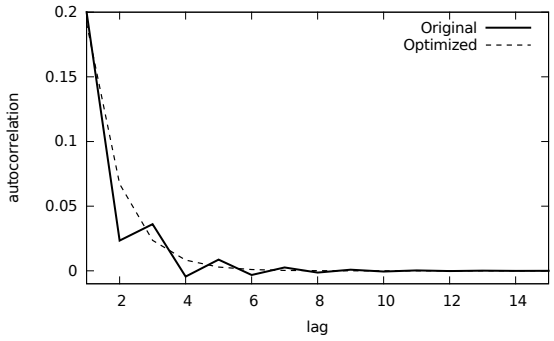

Fig. 7. Comparison of the autocorrelation functions

known, the following non-linear optimization problem has to be solved:

$$\min_{\boldsymbol{B_1},\ldots,\boldsymbol{B_K}} \mathcal{D}_k\{\mathcal{A},\mathcal{B}\}, \tag{43}$$

where $\mathcal{D}_k\{\mathcal{A},\mathcal{B}\}$ is the multi-type distance up to lag-k as defined by (23). The inequality and equality constraints (to ensure that $\sum_{m=0}^{K} \boldsymbol{B_m}\mathbb{1} = 0$ and that the stationary phase distribution at arrivals is β) are

$$\begin{aligned} \boldsymbol{B_m} &\geq \mathbf{0}, \quad \text{for } m = 1,\ldots,K, \\ \sum_{m=1}^{K} \boldsymbol{B_m}\mathbb{1} &= -\boldsymbol{B_0}\mathbb{1}, \\ \beta(-\boldsymbol{B_0})^{-1} \sum_{m=1}^{K} \boldsymbol{B_m} &= \beta. \end{aligned} \tag{44}$$

While many non-linear programming problems are difficult to solve, we had a positive numerical experience with this one. The `fmincon` method of MATLAB managed to terminate in a couple of seconds returning a valid solution that is relatively independent on the initial guess.

To demonstrate the usefulness of this procedure, the numerical example in paper [11] is revisited. In that example two MMAP[K]/PH[K]/1-FCFS queues are considered in a tandem setting, and the performance measures of the second queue are analyzed. The traffic on the link between the queues is approximated by a MMAP, which is obtained by matching the lag-1 joint moments of the departure process of the first queue. The problem in that paper is that the result of the moment matching did not define a valid stochastic process. To overcome this difficulty, [11] proposed to apply joint moment fitting instead of matching, relying on procedure [4]. Here, an alternative procedure based on Section 6.2.2 and the non-linear program above is applied to solve the same problem.

The matrices of the lag-1 joint moments of the departure process corresponding to type-1 ($\boldsymbol{N}^{(1)}$) and type-2 ($\boldsymbol{N}^{(2)}$) jobs are

$$\boldsymbol{N}^{(1)} = \begin{bmatrix} 0.7500 & 1.4959 & 11.2588 \\ 1.5001 & 4.6692 & 43.7821 \\ 11.0045 & 43.4938 & 438.3192 \end{bmatrix}, \boldsymbol{N}^{(2)} = \begin{bmatrix} 0.2500 & 0.5042 & 3.8188 \\ 0.5000 & 1.5330 & 14.2786 \\ 4.0731 & 14.8957 & 146.7858 \end{bmatrix}. \tag{45}$$

Based on these joint moments, the moment matching method returns a MRAP given by matrices

$$\boldsymbol{H_0} = \begin{bmatrix} -1.4452 & 1.7636 & -2.6186 \\ 0.0160 & -0.9806 & 0.7218 \\ -0.3493 & 0.6472 & -1.0551 \end{bmatrix}, \quad \boldsymbol{H_1} = \begin{bmatrix} 0.7311 & 0.6049 & 0.2228 \\ 0.0800 & 0.0795 & 0.0872 \\ 0.1708 & 0.1694 & 0.1630 \end{bmatrix},$$
$$\boldsymbol{H_2} = \begin{bmatrix} 0.3164 & 0.2776 & 0.1475 \\ -0.0044 & -0.0030 & 0.0034 \\ 0.0869 & 0.0873 & 0.0797 \end{bmatrix},$$

which is clearly a non-Markovian representation. Our procedure obtained an approximate MMAP based on the squared distance optimization, with matrices

$$\boldsymbol{D_0} = \begin{bmatrix} -0.18727 & 0 & 0 \\ 0 & -0.7735 & 0 \\ 0 & 0 & -2.5183 \end{bmatrix}, \quad \boldsymbol{D_1} = \begin{bmatrix} 0.089235 & 0.027462 & 0.014315 \\ 0.034411 & 0.4187 & 0.18986 \\ 0.27113 & 0.78772 & 0.52658 \end{bmatrix},$$
$$\boldsymbol{D_2} = \begin{bmatrix} 0.045767 & 1.455 \times 10^{-5} & 0.010476 \\ 0.012185 & 0.066362 & 0.051978 \\ 0.060869 & 0.65762 & 0.21443 \end{bmatrix}.$$

The joint moments of the approximate MMAP are

$$\hat{\boldsymbol{N}}^{(1)} = \begin{bmatrix} 0.74692 & 1.4461 & 10.609 \\ 1.4771 & 4.2299 & 38.164 \\ 10.763 & 38.644 & 376.84 \end{bmatrix}, \quad \hat{\boldsymbol{N}}^{(2)} = \begin{bmatrix} 0.25308 & 0.55401 & 4.4688 \\ 0.523 & 1.8518 & 18.427 \\ 4.315 & 18.093 & 188.1 \end{bmatrix},$$

which are relatively close to the original ones given by (45).

7 Application: Approximating the departure process of a MAP/MAP/1 queue by a MAP

A popular approach for the analysis of a network of MAP/MAP/1 queues is the so called traffic based decomposition, where the internal traffic in the network is modeled by MAPs. The closeness properties of MAPs over splitting and superposition make them ideal for this purpose. The key question is how to obtain a MAP that represents the departure process of a queue. Two options from the past literature which are known to perform relatively well are as follows:

- The ETAQA truncation of the queue length process in [23],
- and the joint moments based procedure presented in [9].

In the practice both methods can return a RAP instead of a MAP, thus the procedure described in Section 6 becomes relevant.

7.1 Introduction to the departure process analysis

The MAP/MAP/1 queue is a subclass of QBD queues, which are characterized by four matrices, $\boldsymbol{B}, \boldsymbol{F}, \boldsymbol{L}$ and $\boldsymbol{L_0}$. Matrices $\boldsymbol{B}$ and $\boldsymbol{F}$ consist of phase transition rates accompanied by service and arrival events, respectively, while matrices $\boldsymbol{L_0}$ and $\boldsymbol{L}$ correspond to the internal transitions when the queue is at level 0 and at level above zero. The generator matrix of the CTMC keeping track of the number of jobs in the queue and the phase of the system has a tri-diagonal structure given by

$$\boldsymbol{Q} = \begin{bmatrix} \boldsymbol{L_0} & \boldsymbol{F} & & & \\ \boldsymbol{B} & \boldsymbol{L} & \boldsymbol{F} & & \\ & \boldsymbol{B} & \boldsymbol{L} & \boldsymbol{F} & \\ & & \ddots & \ddots & \ddots \end{bmatrix}. \tag{46}$$

Separating the transitions that generate a departure leads to a MAP that captures the departure process in an exact way as

$$\boldsymbol{D_0} = \begin{bmatrix} \boldsymbol{L_0} & \boldsymbol{F} & & \\ & \boldsymbol{L} & \boldsymbol{F} & \\ & & \boldsymbol{L} & \boldsymbol{F} \\ & & & \ddots & \ddots \end{bmatrix}, \quad \boldsymbol{D_1} = \begin{bmatrix} \boldsymbol{B} & & \\ & \boldsymbol{B} & \\ & & \ddots \end{bmatrix}, \tag{47}$$

but unfortunately this representation has infinitely many states. A finite representation can be obtained by truncating the infinite model. It is proven in [23] that an appropriate truncation at level k is able to preserve the joint distribution of the departure process up to lag-$(k-1)$. The truncation at level k is done as

$$\boldsymbol{D_0}^{(k)} = \begin{bmatrix} \boldsymbol{L_0} & \boldsymbol{F} & & & \\ & \boldsymbol{L} & \boldsymbol{F} & & \\ & & \ddots & \ddots & \\ & & & & \boldsymbol{L}+\boldsymbol{F} \end{bmatrix} \begin{matrix} 0 \\ 1 \\ \vdots \\ k \end{matrix}, \quad \boldsymbol{D_1}^{(k)} = \begin{bmatrix} \boldsymbol{B} & & & \\ & \ddots & & \\ & & & \boldsymbol{B}-\boldsymbol{FG}\ \boldsymbol{FG} \end{bmatrix} \begin{matrix} 0 \\ 1 \\ \vdots \\ k \end{matrix}, \tag{48}$$

where matrix $\boldsymbol{G}$ is the minimal non-negative solution to the matrix-quadratic equation $\boldsymbol{0} = \boldsymbol{B} + \boldsymbol{LG} + \boldsymbol{FG}^2$.

Although the truncation leads to a finite model, the number of states can still be too large. The superposition operations in the queueing network increase the number of states even more, and the limits of numerical tractability are easily hit. A possible solution for the state-space explosion is provided in [9], where a compact representation is constructed while maintaining the lag-1 joint moments of the large process.

7.2 Practical problems and possible solutions

An issue with both the ETAQA departure model and the joint moment based approach is that they do not always return a Markovian representation, it is not even guaranteed that the departure model is a valid stochastic process.

Applying the RAP approximation procedure presented in Section 6 makes it possible to overcome this problem. Based on $(\boldsymbol{D_0}^{(k)}, \boldsymbol{D_1}^{(k)})$ it always returns a valid Markovian representation $(\boldsymbol{H_0}, \boldsymbol{H_1})$, and at the same time it is also able to compress the truncated departure process to a desired level.

There is, however, one issue which has to be taken account when applying the procedure of Section 6, namely that the number of marginal moments that can be used to obtain matrix $\boldsymbol{H_0}$ is limited. We are going to show that the order of the PH distribution representing the inter-departure times is finite (denoted by N_D), determined by $2N_D - 1$ moments, and using more moments during the approximation leads to a dependent moment set (see [6]).

Theorem 6. *The order of the PH distribution representing the inter-departure times of a QBD queue with block size $N > 1$ is*

$$N_D = 2N. \tag{49}$$

Proof. In [23] it is shown how an order $2N$ PH distribution is constructed that captures the inter-departure times in an exact way, thus $N_D \leq 2N$. Additionally, it is easy to find concrete matrices $\boldsymbol{B}, \boldsymbol{F}, \boldsymbol{L}$ and $\boldsymbol{L_0}$ such that the order of this PH distribution is exactly $2N$ (practically any random matrices are suitable, the order can be determined by the STAIRCASE algorithm of [5]). Consequently, we have that $N_D = 2N$. □

Surprisingly, in case of MAP/MAP/1 queues the order of the inter-departure times is lower.

Theorem 7. *([9], Theorem 2) The order of the PH distribution representing the inter-departure times of a MAP/MAP/1 queue is*

$$N_D = N_A + N_S, \tag{50}$$

where N_A denotes the size of the MAP describing the arrival process and N_S the one of the service process, assuming that $N_A + N_S > 1$.

Thus, the proposed method for producing a MAP $(\boldsymbol{B_0}, \boldsymbol{B_1})$ that approximates the departure process is as follows:

1. First the ETAQA departure model is constructed up to the desired lag k, providing matrices $(\boldsymbol{D_0}^{(k)}, \boldsymbol{D_1}^{(k)})$. The stationary phase distribution at departure instans needs to be determined as well, α_D is the unique solution to $\alpha_D(-\boldsymbol{D_0}^{(k)})^{-1}\boldsymbol{D_1}^{(k)}, \alpha_D \mathbb{1} = 1$.
2. The marginal moments of the inter-departure times are computed from α_D and $\boldsymbol{D_0}^{(k)}$. The more moments are taken into account, the larger the output of the approximation is. According to the above theorems, more than $2N_D - 1$ should not be used.
3. Matrix $\boldsymbol{B_0}$ is obtained by moment matching (see Section 6.2.2).
4. Matrix $\boldsymbol{B_1}$ is obtained such that either the squared distance of the joint density is minimized up to lag k, see 6.2.1.

7.3 Numerical example

In this example[4] we consider a simple tandem queueing network of two MAP/MAP/1 queues. The arrival process of the first station is given by matrices

$$\boldsymbol{D_0} = \begin{bmatrix} -0.542 & 0.003 & 0 \\ 0.04 & -0.23 & 0.01 \\ 0 & 0.001 & -2.269 \end{bmatrix}, \quad \boldsymbol{D_1} = \begin{bmatrix} 0.021 & 0 & 0.518 \\ 0 & 0.17 & 0.01 \\ 0.004 & 0.005 & 2.259 \end{bmatrix}, \tag{51}$$

while the matrices characterizing the service process are

$$\boldsymbol{S_0} = \begin{bmatrix} -10 & 0 \\ 0 & -2.22 \end{bmatrix}, \quad \boldsymbol{S_1} = \begin{bmatrix} 7.5 & 2.5 \\ 0.4 & 1.82 \end{bmatrix}. \tag{52}$$

With these parameters both the arrival and the service times are positively correlated ($\rho_1^{(\mathcal{A})} = 0.21$ and $\rho_1^{(\mathcal{S})} = 0.112$) and the utilization of the first queue is 0.624.

The service times of the second station are Erlang distributed with order 2 and intensity parameter 6 leading to utilization 0.685.

This queueing network is analyzed such a way, that the departure process is approximated by the ETAQA truncation and by the joint moments based methods. Next, our RAP approximation procedure (Section 6) is applied to address the issues of the approximate departure processes, namely to obtain a Markovian approximation and in case of the ETAQA truncation method, to compress the large model to a compact one.

Model of the departure process	**#states**	**E(queue len.)**
Accurate result (simulation):	n/a	2.6592
ETAQA, lag-1 truncation	18	2.3379
Our method based on 3 moments and $\mathcal{D}_2\{\}$	2	2.4266
Our method based on 5 moments and $\mathcal{D}_2\{\}$	3	2.5722
ETAQA, lag-5 truncation	42	2.5405
Our method based on 3 moments and $\mathcal{D}_2\{\}$	2	2.4266
Our method based on 5 moments and $\mathcal{D}_2\{\}$	3	2.5722
Our method based on 3 moments and $\mathcal{D}_6\{\}$	2	2.4266
Our method based on 5 moments and $\mathcal{D}_6\{\}$	3	2.6805
Joint moments based, 2 states	2	2.3255
Our method based on 3 moments and $\mathcal{D}_2\{\}$	2	2.3255
Joint moments based, 3 states	3	2.755
Our method based on 3 moments and $\mathcal{D}_2\{\}$	2	2.4266
Our method based on 5 moments and $\mathcal{D}_2\{\}$	3	2.7489

Table 1. Results of the queueing network example

[4] The implementation of the presented method and all the numerical examples can be downloaded from `http://www.hit.bme.hu/~ghorvath/software`

Table 1 depicts the mean queue length of the second station and the model size by various departure process approximations. The ETAQA truncation model has been applied with truncation levels 2 and 6, which has been compressed by our method based on either 3 or 5 marginal moments and with $\mathcal{D}_2\{\}$ or $\mathcal{D}_6\{\}$ distance optimization. The corresponding queue length distributions at the second station are compared in Figure 8. The departure process has also been approximated by the joint moments based method of [9], and an approximate Markovian representation has been constructed with our method based on 3 or 5 marginal moments and $\mathcal{D}_2\{\}$ optimization.

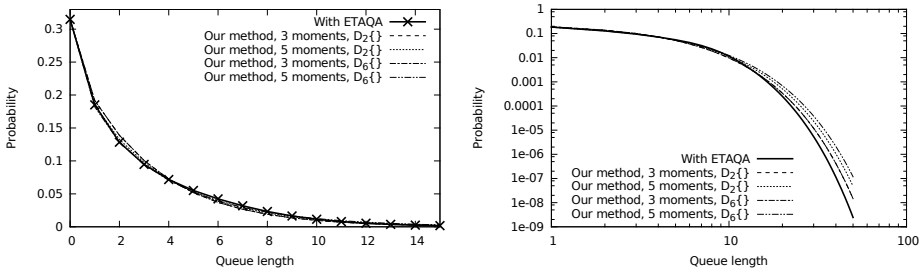

Fig. 8. Queue length distribution with the ETAQA departure model and its Markovian approximations

The results indicate that the RAP approximation and state space compression technique presented in this paper is efficient, the MAP returned is able to capture the important characteristic of the target RAP with an acceptable error.

Acknowledgment

This work was supported by the Hungarian research project OTKA K101150 and by the János Bolyai Research Scholarship of the Hungarian Academy of Sciences.

References

1. S. Asmussen and M. Bladt. Point processes with finite-dimensional conditional probabilities. *Stochastic Processes and their Application*, 82:127–142, 1999.
2. Søren Asmussen and Ger Koole. Marked point processes as limits of Markovian arrival streams. *Journal of Applied Probability*, pages 365–372, 1993.
3. N. G. Bean and B. F. Nielsen. Quasi-birth-and-death processes with rational arrival process components. *Stochastic Models*, 26(3):309–334, 2010.
4. Peter Buchholz, Peter Kemper, and Jan Kriege. Multi-class Markovian arrival processes and their parameter fitting. *Performance Evaluation*, 67(11):1092–1106, 2010.

5. Peter Buchholz and Miklós Telek. On minimal representations of rational arrival processes. *Annals of Operations Research*, 202(1):35–58, 2013.
6. Giuliano Casale, Eddy Z Zhang, and Evgenia Smirni. Trace data characterization and fitting for Markov modeling. *Performance Evaluation*, 67(2):61–79, 2010.
7. Gene Howard Golub, Stephen Nash, and Charles Van Loan. A Hessenberg-Schur method for the problem AX+XB=C. *Automatic Control, IEEE Transactions on*, 24(6):909–913, 1979.
8. Qi-Ming He and Marcel Neuts. Markov arrival processes with marked transitions. *Stochastic Processes and their Applications*, 74:37–52, 1998.
9. András Horváth, Gábor Horváth, and Miklós Telek. A joint moments based analysis of networks of MAP/MAP/1 queues. *Performance Evaluation*, 67(9):759–778, 2010.
10. Gábor Horváth. Moment matching-based distribution fitting with generalized hyper-Erlang distributions. In *Analytical and Stochastic Modeling Techniques and Applications*, pages 232–246. Springer, 2013.
11. Gábor Horváth and Benny Van Houdt. Departure process analysis of the multi-type MMAP[K]/PH[K]/1 FCFS queue. *Performance Evaluation*, 70(6):423–439, 2013.
12. Gbor Horvth. Measuring the distance between maps and some applications. In Marco Gribaudo, Daniele Manini, and Anne Remke, editors, *Analytical and Stochastic Modelling Techniques and Applications*, volume 9081 of *Lecture Notes in Computer Science*, pages 100–114. Springer International Publishing, 2015.
13. Mary A Johnson and Michael R Taaffe. Matching moments to phase distributions: Mixtures of Erlang distributions of common order. *Stochastic Models*, 5(4):711–743, 1989.
14. Guy Latouche and Vaidyanathan Ramaswami. *Introduction to matrix analytic methods in stochastic modeling*, volume 5. Society for Industrial and Applied Mathematics, 1987.
15. Alan J Laub. *Matrix analysis for scientists and engineers*. Siam, 2005.
16. Lester Lipsky. *Queueing Theory: A linear algebraic approach*. Springer Science & Business Media, 2008.
17. M. Neuts. Probability distributions of phase type. In *Liber Amicorum Prof. Emeritus H. Florin*, pages 173–206. University of Louvain, 1975.
18. M. F. Neuts. A versatile Markovian point process. *Journal of Applied Probability*, 16:764–779, 1979.
19. T Rolski, H Schmidli, V Schmidt, and J Teugels. Stochastic processes for finance and insurance. *Willey, New York*, 1999.
20. Willi-Hans Steeb. *Matrix calculus and Kronecker product with applications and C++ programs*. World Scientific, 1997.
21. Miklós Telek and Gábor Horváth. A minimal representation of Markov arrival processes and a moments matching method. *Performance Evaluation*, 64(9):1153–1168, 2007.
22. A. van de Liefvoort. The moment problem for continuous distributions. Technical report, University of Missouri, WP-CM-1990-02, Kansas City, 1990.
23. Qi Zhang, Armin Heindl, and Evgenia Smirni. Characterizing the BMAP/MAP/1 departure process via the ETAQA truncation. *Stochastic Models*, 21(2-3):821–846, 2005.