Markovian Arrival Processes in Multi-Dimensions

Andreas Blume[1], Peter Buchholz[1], and Clara Scherbaum[1]

Informatik IV, Technical University of Dortmund,
D-44221 Dortmund, Germany
{andreas.blume,peter.buchholz,clara.scherbaum}@cs.tu-dortmund.de

Abstract. Phase Type Distributions (PHDs) and Markovian Arrival Processes (MAPs) are established models in computational probability to describe random processes in stochastic models. In this paper we extend MAPs to Multi-Dimensional MAPs (MDMAPs) which are a model for random vectors that may be correlated in different dimensions. The computation of different quantities like joint moments or conditional densities is introduced and a first approach to compute parameters with respect to measured data is presented.
Keywords. Input Modeling, Markovian Arrival Process, High-Dimensional Data, Stochastic Processes.

1 Introduction

In many application areas like computer networks [15], supply chains [27] or dependable systems [18], high dimensional data plays an important role in understanding, analyzing and improving the behavior of contemporary systems. Currently available data is mainly analyzed offline to monitor or predict the behavior of complex systems. However, it is known that model-based approaches are often necessary to understand and analyze large systems. In simulation models [24] and also in models based on Markov chains [12], multi-dimensional data is usually described by independent data streams, where at most the elements in one stream are correlated. In practice, multi-dimensional data is correlated in several dimensions and this correlation cannot be neglected in realistic models. The necessary models to describe such a behavior are denoted as multivariate input models.

In multivariate input models one usually distinguishes between random vectors which describe K-dimensional vectors of random variables that are correlated. Subsequent vectors are assumed to be independent. Random variables that are correlated over time are described by stochastic processes. The combination of both results in multivariate time series. Multivariate input models are mainly considered in simulation, for an overview of available approaches see [5]. Although different approaches for multivariate stochastic processes exist their practical applicability is limited mainly due to very specific structures that capture only parts of the observed behavior, complex methods for parameter

fitting, complex methods to generate random variates and the limiting possibilities to perform numerical or analytical analysis of the models. Most promising approaches seem to be VARTA processes [6] and copula-based models [3]. Both approaches are restricted to specific marginal distributions, like normal distributions or distributions of the Johnson-type. The VARTA approach has been extended to phase type distributions in [21].

In computational probability [29, 31, 34] input models based on Markov processes like phase type distributions (PHDs) or Markovian arrival processes (MAPs) are very popular because they allow one to model a wide variety of behaviors and they can also be analyzed by numerical techniques and not only by simulation. MAPs are a model to describe correlated univariate processes and are therefore an alternative to time series but they cannot be applied to describe multivariate processes. In this paper we extend Markov models like PHDs and MAPs to the multivariate case. This results in a new stochastic model which is an alternative to VARTA processes and similar models. Some older approaches to extend phase type distributions to multivariate phase type distributions exist [1, 23]. However, the models defined in these papers differ from our model in that they describe an absorbing Markov process with a multi-dimensional reward structure to generate random vectors. Here, we consider parallel running absorbing Markov processes which are coupled by the initial distributions. This model allows us to generate random vectors with correlated components and correlation between subsequent realizations.

The structure of the paper is as follows. In the following section we introduce the notation and Markov input models. Afterwards, in Section 3, we define the multi-dimensional stochastic model and define afterwards multi-dimensional Markovian arrival processes (MDMAPs). In Section 4, the analysis of MDMAPs is presented, followed by a first approach to fit the parameters according to some quantities like joint moments or values of the conditional probability distribution function. In Section 6 first examples are presented and then the paper is concluded. Proofs of the theorems and major equations can be found in an online appendix.

2 Background

We first introduce some notation and define afterwards the basic models used in this paper.

2.1 Basic Notation

Matrices and vectors are denoted by bold face small and capital letters. $\mathbb{1}$ is a column vector of 1s, all other vectors are row vectors. $\mathbf{0}$ is a matrix or vector containing only 0 elements, $\boldsymbol{I}$ is the identity matrix, $\boldsymbol{e}_i$ is the ith unit row vector. $\boldsymbol{a}^T$ describes the transposed of vector $\boldsymbol{a}$ and $\mathrm{diag}(\boldsymbol{a})$ denotes a diagonal matrix with elements $\boldsymbol{a}(i)$ on the main diagonal. $\mathbb{R}^{n\times n}$ is the set of $n \times n$ matrices. $\boldsymbol{Q}(i\bullet)$ and $\boldsymbol{Q}(\bullet i)$ describe the ith row and column of matrix $\boldsymbol{Q}$. A generator is a

matrix $\boldsymbol{Q} \in \mathbb{R}^{N \times N}$ with row sum zero (i.e., $\boldsymbol{Q}\mathbb{1} = \boldsymbol{0}$) and $\boldsymbol{Q}(i,j) \geq 0$ for $i \neq j$. $\boldsymbol{Q}$ is a sub-generator if $\boldsymbol{Q}\mathbb{1} \leq \boldsymbol{0}$ and some $i \in \{0, \ldots, n-1\}$[1] exists such that $\boldsymbol{Q}(i\bullet)\mathbb{1} < 0$. In the sequel we assume that all sub-generators we consider in this paper are non-singular which means that the inverse exists and is non-positive. $\boldsymbol{Q}$ is irreducible if between every pair of states i, j a path $i = i_0, i_1, \ldots, i_k = j$ exists such that $\boldsymbol{Q}(i_{h-1}, i_h) > 0$ for $h = 1, \ldots, k$.

2.2 Markov Input Models

In stochastic modeling *input modeling* describes the generation of appropriate, usually stochastic, models to represent the input parameters based on measured data from some real process [24]. In simulation, traditionally standard distributions or stochastic processes have been used for this purpose. More recently input models based on Markov processes like phase type distributions and Markovian arrival processes gained much attention. These models are flexible and can be used in simulation as well as in combination with numerical analysis techniques. We use the following traditional definitions [12, 29].

Definition 1. *A Phase Type Distribution (PHD) is defined by* $(\boldsymbol{\pi}, \boldsymbol{D})$ *where* $\boldsymbol{\pi}$ *is the initial distribution and* $\boldsymbol{D}$ *is a sub-generator of an absorbing Markov chain.*

A PHD is characterized by the time to absorption of the absorbing Markov chain described by $(\boldsymbol{\pi}, \boldsymbol{D})$. A Markovian Arrival Process [28, 26] is an extension of a PHD.

Definition 2. *A Markovian Arrival Process (MAP) is described by two matrices* $(\boldsymbol{D}, \boldsymbol{C})$[2] *where* $\boldsymbol{D}$ *is a sub-generator,* $\boldsymbol{C} \geq 0$ *and* $\boldsymbol{Q} = \boldsymbol{D} + \boldsymbol{C}$ *is an irreducible generator.*

The interpretation of the behavior of a MAP is as follows. The process performs transitions as described by the matrices $\boldsymbol{D}$ and $\boldsymbol{C}$ and whenever a transition from $\boldsymbol{C}$ occurs, an event is triggered. Let $\boldsymbol{d} = \boldsymbol{C}\mathbb{1} = -\boldsymbol{D}\mathbb{1}$ and $\boldsymbol{P} = (-\boldsymbol{D})^{-1}\boldsymbol{C}$ be the transition matrix of the embedded process at event times. Since $\boldsymbol{Q}$ is irreducible, it has a unique stationary vector observing $\boldsymbol{\phi}\boldsymbol{Q} = \boldsymbol{0}$ and $\boldsymbol{\phi}\mathbb{1} = 1$. Then $\boldsymbol{\pi} = \boldsymbol{\phi}\boldsymbol{C}/(\boldsymbol{\phi}\boldsymbol{C}\mathbb{1})$, $\boldsymbol{\pi}\boldsymbol{P} = \boldsymbol{\pi}$ and $\boldsymbol{\pi}$ describes the stationary vector of the MAP immediately after an event. $(\boldsymbol{\pi}, \boldsymbol{D})$ is the embedded PHD of the MAP. We define the following two sets

$$inp(\boldsymbol{\pi}, \boldsymbol{D}) = inp(\boldsymbol{\pi}) = \{i | \boldsymbol{\pi}(i) > 0\}, outp(\boldsymbol{\pi}, \boldsymbol{D}) = outp(\boldsymbol{D}) = \{i | \boldsymbol{d}(i) > 0\} \tag{1}$$

[1] In an n-dimensional space elements are always numbered from 0 through $n-1$ because this numbering is more appropriate for mapping multi-dimensional spaces into a single space.

[2] We use the names $\boldsymbol{D}$ and $\boldsymbol{C}$ rather than $\boldsymbol{D}_0$ and $\boldsymbol{D}_1$ for the matrices of a MAP because the numbers in the postfix are later used to denote matrices of different MAPs or PHDs.

of input and output states. $n^c = |inp(\boldsymbol{\pi}, \boldsymbol{D})|$ and $n^r = |outp(\boldsymbol{\pi}, \boldsymbol{D})|$ are the cardinalities of the sets. A PHD is *input flexible* if $n^c > 1$ and it is output flexible if $n^r > 1$. If we assume that each input and output state describes an individual stochastic behavior, then an input flexible PHD allows one to choose a specific behavior by selecting the input state, an output flexible PHD allows one to interpret the previous behavior by considering the output state. To expand a PHD to a MAP (see [12, 19] for details), the PHD has to be input and output flexible to specify correlation. A MAP $(\boldsymbol{D}, \boldsymbol{C})$ can also be represented as $(\boldsymbol{D}, \boldsymbol{G})$ where $\boldsymbol{C} = \text{diag}(\boldsymbol{d})\boldsymbol{G}$ and $\boldsymbol{G}$ is a matrix with $\boldsymbol{G}(i\bullet)\mathbb{1} = 1$ for $i \in outp(\boldsymbol{D})$ and 0 otherwise.

PHDs and MAPs can be easily analyzed according to several quantities including moments, probability density and, in case of MAPs, joint moments or joint densities. For details we refer to the literature [12]. Parameter fitting for a stochastic model describes the process of finding good or optimal parameters such that the stochastic model mimics the behavior of a real process for which data is available. Parameter fitting for PHDs or MAPs is more complex than for many other stochastic models because both models have a highly redundant representation [32]. In principle two approaches can be applied. First, some derived measures can be computed, like moments or joint moments and a least squares approach is used to fit the parameters in such a way that the quantities of the measured data are approximated by the PHD or MAP. Alternatively, maximum likelihood estimators for the parameters can be used which are usually based on the EM algorithm. For details about the corresponding algorithms we refer to the literature [12].

3 Multi-Dimensional Data and Stochastic Models

We first introduce the basic setting for multivariate distributions and random vectors. Afterwards we present a Markov model to describe those quantities.

3.1 Multi-Dimensional Data

Let $\boldsymbol{X} = (X_1, \ldots, X_K)$ be a random vector where each X_k is a random variable. We assume that all random variables are non-negative and the underlying distribution functions have an infinite support. We denote by K the number of dimensions or components of the random vector. If the random variables are mutually independent, each X_k can be modeled by a MAP, if subsequent realizations of X_k are also independent, a PHD is sufficient. Here we consider the case that various dependencies exist between the random variables and subsequent realizations. Thus, $\boldsymbol{X}^{(t)}$ is the vector observed at time t $(= 1, 2, \ldots)$ and $\boldsymbol{X}^{(t+h)}$ is the vector h steps later. In general $X_k^{(t)}$ and $X_l^{(t)}$ as well as $X_k^{(t)}$ and $X_l^{(t+h)}$ may be correlated.

We assume that the stochastic structure of $\boldsymbol{X}$ is unknown but we can observe realizations of $\boldsymbol{X}$. Let $\boldsymbol{x}^{(i)} = \left(x_1^{(i)}, \ldots, x_K^{(i)}\right)$ be the ith realization of $\boldsymbol{X}$ and

$x_k^{(i)}$ is the ith realization of X_k. From a sequence of observations $\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(L)}$ various quantities can be estimated.

$$\hat{X}_k^j = \frac{1}{L} \sum_{i=1}^{L} \left(x_k^{(i)} \right)^j \text{ and } \hat{\sigma}_k^2 = \frac{1}{L-1} \sum_{i=1}^{L} \left(x_k^{(i)} - \hat{X}_k^1 \right)^2 \tag{2}$$

are estimates for the jth moments and the variance of the random variables X_k. We denote by $\hat{\boldsymbol{R}}_h$ the correlation matrix of elements h steps apart which contains the correlation coefficients. Elements of the correlation matrix are estimated by

$$\hat{\boldsymbol{R}}_h(k,l) = \frac{1}{(L-h-1)\hat{\sigma}_k \hat{\sigma}_l} \sum_{i=1}^{L-h} \left(x_k^{(i)} - \hat{X}_k^1 \right) \left(x_l^{(i+h)} - \hat{X}_l^1 \right). \tag{3}$$

The definition can be extended to higher order joint moments as follows.

$$\hat{\boldsymbol{J}}_h^{m,n}(k,l) = \frac{1}{(L-h)} \sum_{i=1}^{L-h} \left(x_k^{(i)} \right)^m \left(x_l^{(i+h)} \right)^n \tag{4}$$

where $k, l \in \{1, \ldots, K\}$ and $n, m \geq 1$. Similarly, the distribution function for one or some dimensions of the random vector can be estimated. All presented estimators are consistent. Like for joint moments in (4) we consider especially dependencies between two components k and l which are described in the following joint dependencies.

$$\hat{\boldsymbol{F}}_h^{y,z}(k,l) = \frac{1}{L-h} \sum_{i=1}^{L-h} \delta \left(x_k^{(i)} \leq y \right) \delta \left(x_l^{(i+h)} \leq z \right) \tag{5}$$

3.2 Multi-Dimensional Markov Models

The available Markov models are not able to describe multi-dimensional data. Therefore we propose an extended model which consists of K absorbing Markov chains that run in parallel, the absorption time of the kth chain determines the value of the kth random variable. After absorption of all chains, they are restarted according to a joint probability distribution which depends on the states immediately before absorption. The later concept is a direct extension of the idea that is used in MAPs to describe correlation. The following definition formalizes the model.

Definition 3. *A Multi-Dimensional Markovian Arrival Process (MDMAP) is defined by K sub-generators $\boldsymbol{D}_k$ of order n_k ($k = 1, \ldots, K$) and a coupling matrix $\boldsymbol{G}$.*

K is the dimension of the MDMAP. Matrix $\boldsymbol{G}$ is an $n_{1:K} \times n_{1:K}$ matrix ($n_{1:K} = \prod_{k=1}^{K} n_k$) matrix where state vector $(i_1, \ldots, i_K)$ ($i_k \in \{0, \ldots, n_k - 1\}$) is mapped onto index $i_{1:K} = \sum_{k=1}^{K} i_k \cdot n_{k+1:K}$ (where $n_{l:k} = \prod_{i=l}^{k} n_k$ for $k \geq l$ and 1

for $l > k$). $\boldsymbol{G} \geq \boldsymbol{0}$, $G(i_{1:K}\bullet) = \boldsymbol{0}$ if $i_k \notin outp(\boldsymbol{D}_k)$ for some k and $\boldsymbol{G}(i_{1:K}\bullet)\mathbb{1} = 1$ otherwise. $\boldsymbol{G}(i_{1:K}, j_{1:K}) > 0$ implies $i_k \in outp(\boldsymbol{D}_k)$ and $j_k \in inp(\boldsymbol{\pi}_k)$ for all $k = 1, \ldots, K$ which is denoted as $i_{1:K} \in outp_{1:K}$ and $j_{1:K} \in inp_{1:K}$, respectively. The notations may be restricted to subsets of indices $k : l$ for $k \leq l$ or subset $\mathcal{K} \subseteq \{1, \ldots, K\}$. For the cardinalities of the sets we use the following notations $n^r_{\mathcal{K}} = |outp_{\mathcal{K}}|$ and $n^c_{\mathcal{K}} = |inp_{\mathcal{K}}|$. Let $\mathbb{1}^{inp}_{n_{1:K}}$ be a vector of length $n_{1:K}$ where $\mathbb{1}^{inp}_{n_{1:K}} = 1$ if $i \in inp_{1:K}$ and 0 otherwise. Similarly $\mathbb{1}^{outp}_{n_{1:K}}$ is defined. Then $\boldsymbol{G}\mathbb{1}_{n_{1:K}} = \boldsymbol{G}\mathbb{1}^{inp}_{n_{1:K}} = \mathbb{1}^{outp}_{n_{1:K}}$.

The behavior of an MDMAP is as follows. Each of the K absorbing Markov chains generates a non-negative value, the exit states i_k are kept and finally row $\boldsymbol{G}(i_{1:K}\bullet)$ defines a probability distribution over the input states of each chain. Dependencies between successive events of one chain and between chains are realized by the relation between input and output states.

Let $\boldsymbol{\pi}$ be a vector of length $n_{1:K}$ which contains the distribution immediately before the next event starts. Let $\boldsymbol{d}_k = -\boldsymbol{D}_k\mathbb{1}$ and the stochastic matrix $\boldsymbol{H}_k = (-\boldsymbol{D}_k)^{-1}\operatorname{diag}(\boldsymbol{d}_k)$. Observe that $\boldsymbol{H}_k(i_k, j_k) > 0$ implies $j_k \in outp_k$. Vector $\boldsymbol{\pi}$ can be computed from the following set of linear equations.

$$\boldsymbol{\pi}\left(\left(\bigotimes_{k=1}^{K}\boldsymbol{H}_k\right)\boldsymbol{G}\right) = \boldsymbol{\pi} \text{ and } \boldsymbol{\pi}\mathbb{1} = 1, \tag{6}$$

if the matrix in brackets contains a single irreducible subset of states which will be assumed for the moment. For some vector $\boldsymbol{\pi} \in \mathbb{R}^{n_{1:K}}$ we define the mapping onto the kth dimension as vector $\boldsymbol{\pi}_k \in \mathbb{R}^{n_k}$ with

$$\boldsymbol{\pi}_k(i_k) = \sum_{i_1=0}^{n_1-1}\cdots\sum_{i_{k-1}=0}^{n_{k-1}-1}\sum_{i_{k+1}=0}^{n_{k+1}-1}\cdots\sum_{i_K=0}^{n_K-1}\boldsymbol{\pi}(i_{1:K}) \tag{7}$$

This mapping can be computed by right multiplication of $\boldsymbol{\pi}$ with the following matrix.

$$\boldsymbol{\pi}_k = \boldsymbol{\pi}\boldsymbol{V}_k \text{ where } \boldsymbol{V}_k = \bigotimes_{l=1}^{K}\boldsymbol{id}_l \text{ and } \boldsymbol{id}_l = \begin{cases}\mathbb{1}_{n_l} & \text{if } l \neq k,\\ \boldsymbol{I}_{n_k} & \text{if } k = l\end{cases} \tag{8}$$

where $\otimes$ is the Kronecker product. Obviously $\boldsymbol{V}_k = \mathbb{1}_{n_{1:k-1}} \otimes \boldsymbol{I}_{n_k} \otimes \mathbb{1}_{n_{k+1:K}}$. The mapping can be extended to subsets of components. Let $\mathcal{K} \subseteq \{1, \ldots, K\}$ and

$$\boldsymbol{V}_{\mathcal{K}} = \bigotimes_{l=1}^{K}\boldsymbol{id}_l \text{ where } \boldsymbol{id}_l = \begin{cases}\mathbb{1}_{n_l} & \text{if } l \notin \mathcal{K},\\ \boldsymbol{I}_{n_l} & \text{if } l \in \mathcal{K}.\end{cases} \tag{9}$$

$\boldsymbol{\pi}_{\mathcal{K}} = \boldsymbol{\pi}\boldsymbol{V}_{\mathcal{K}}$ is the embedded initial vector mapped onto the subset $\mathcal{K}$. For notational convenience we write $\boldsymbol{V}_{kl}$ for $\boldsymbol{V}_{\{k,l\}}$.

For an initial vector $\boldsymbol{\pi}$, the exit vector $\boldsymbol{\psi}$ is given by

$$\boldsymbol{\psi} = \boldsymbol{\pi}\left(\bigotimes_{k=1}^{K}\boldsymbol{H}_k\right) \text{ and } \boldsymbol{\psi}_k = \boldsymbol{\pi}_k\boldsymbol{H}_k. \tag{10}$$

$\boldsymbol{\psi}$ and $\boldsymbol{\psi}_k$ contain the probabilities of absorption from state $i_{1:k}$ and i_k, respectively. Obviously, $\boldsymbol{\psi}_k(i_k) = 0$ for $i_k \notin outp_k$. We assume that $\boldsymbol{\psi}_k(i_k) > 0$ for $i_k \in outp_k$ otherwise the corresponding state would not be reachable from an initial state and can therefore removed from the PHD.

The mapping of matrix $\boldsymbol{G}$ on the state space of some components is defined according to some exit vector $\boldsymbol{\psi}$ using the following matrix

$$\boldsymbol{W}_{\mathcal{K}}[\boldsymbol{\psi}] = \operatorname{diag}\left(\boldsymbol{\psi}\boldsymbol{V}_{\mathcal{K}}\right)^{+} \boldsymbol{V}_{\mathcal{K}}^{T} \operatorname{diag}(\boldsymbol{\psi}) \text{ and } \boldsymbol{G}_{\mathcal{K}}[\boldsymbol{\psi}] = \boldsymbol{W}_{\mathcal{K}}[\boldsymbol{\psi}]\boldsymbol{G}\boldsymbol{V}_{\mathcal{K}} \tag{11}$$

where $\boldsymbol{A}^{+}$ is the pseudo-inverse of matrix $\boldsymbol{A}$ which can be computed for the diagonal matrix $\operatorname{diag}\left(\boldsymbol{\psi}\boldsymbol{V}_{\mathcal{K}}\right)$ by substituting non-zero diagonal elements by the inverse and leaving zero diagonal elements unchanged. This is the usual way of aggregation in multi-dimensional Markov models (see e.g. [13] for details).

Theorem 1. *For some MDMAP with K components, coupling matrix $\boldsymbol{G}$, initial vector $\boldsymbol{\pi}$ and any subset $\mathcal{K} \subseteq \{1, \ldots, K\}$, the initial vector $\boldsymbol{\pi}_{\mathcal{K}}$ of the MDMAP restricted to the components from $\mathcal{K}$ is the solution of*

$$\boldsymbol{\pi}_{\mathcal{K}} \left(\left(\bigotimes_{k \in \mathcal{K}} \boldsymbol{H_k} \right) \boldsymbol{G}_{\mathcal{K}}[\boldsymbol{\psi}] \right) = \boldsymbol{\pi}_{\mathcal{K}} \text{ and } \boldsymbol{\pi}_{\mathcal{K}} \mathbb{1} = 1$$

where $\boldsymbol{\psi} = \boldsymbol{\pi} \otimes_{k=1}^{K} \boldsymbol{H}_k$.

The number of parameters to represent all matrices $\boldsymbol{D}_k$ only linearly with K and quadratic with n_k. This does not hold for the number of entries in $\boldsymbol{G}$ which may grow with $n^c_{1:K} n^r_{1:K}$. Therefore we consider MDMAPs of *rank* R with product form, that can be represented as follows

$$\boldsymbol{G} = \sum_{r=1}^{R} \lambda^{(r)} \bigotimes_{k=1}^{K} \boldsymbol{G}_k^{(r)} \tag{12}$$

where $\lambda^{(r)} > 0$, $\sum_{r=1}^{R} \lambda^{(r)} = 1$, $\boldsymbol{G}_k^{(r)} \geq 0$ and $\boldsymbol{G}_k^{(r)} \mathbb{1}_{n_k} = \mathbb{1}_{n_k}^{outp}$.

Theorem 2. *For an MDMAP of rank R with product form and some set $\mathcal{K} \subseteq \{1, \ldots, K\}$ vector $\boldsymbol{\pi}_{\mathcal{K}} = \boldsymbol{\pi}\boldsymbol{V}_{\mathcal{K}}$ is the solution of*

$$\boldsymbol{\pi}_{\mathcal{K}} \left(\bigotimes_{k \in \mathcal{K}} \boldsymbol{H}_k \right) \left(\sum_{r=1}^{R} \lambda^{(r)} \bigotimes_{k \in \mathcal{K}} \boldsymbol{G}_k^{(r)} \right) = \boldsymbol{\pi}_{\mathcal{K}} \text{ and } \boldsymbol{\pi}_{\mathcal{K}} \mathbb{1}_{n_{\mathcal{K}}} = 1.$$

Theorem 2 holds in particular for sets $\mathcal{K} = \{k\}$. This implies that in a product form MDMAP each component behaves locally like a MAP$(\boldsymbol{D}_k, \sum_{r=1}^{R} \lambda^{(r)} \boldsymbol{G}_k^{(r)})$.

4 Analysis of MDMAPs

Analysis of MDMAPs can be performed according to one dimension of the random vector or according to the joint distribution.

4.1 Analysis of a Single Vector Component

An MDMAP can be easily mapped on a MAP $(\boldsymbol{D}_k, \boldsymbol{G}_k)$ for one vector component k that neglects all other dimensions. If the MDMAP is of rank R and product form, then $\boldsymbol{G}_k = \sum_{r=1}^{R} \lambda^{(h)} \boldsymbol{G}_k^{(r)}$, otherwise $\boldsymbol{G}_k = \boldsymbol{W}_k[\psi]\boldsymbol{G}\boldsymbol{V}_k$. The resulting MAP can then be analyzed with the available methods (see e.g. [12, Sect. 4]).

4.2 Analysis of Joint Measures

In the following we consider mainly dependencies between two dimensions k and l. In the equations we assume $k < l$, $k > l$ requires a different ordering of the matrices in the equations, but does, of course, not change the general structure. The case $k = l$ describes a single component and is mentioned above. Equations are formulated for MDMAPs of rank R with product form, matrices $\boldsymbol{G}$ for the general case are written underneath the rank R representation. Let $\boldsymbol{J}_1^{m,n}(k,l) = E\left[\left(X_k^{(t)}\right)^m, \left(X_l^{(t+1)}\right)^n\right]$, the joint moment of order m, n for dimension k and dimension l, h steps apart. For $\boldsymbol{J}_0^{m,n}(k,l)$ we obtain

$$\boldsymbol{J}_0^{m,n}(k,l) = m!n!\boldsymbol{\pi}_{kl}\left((\boldsymbol{M}_k)^m \otimes (\boldsymbol{M}_l)^n\right)\mathbb{1}_{n_k n_l} \tag{13}$$

where $\boldsymbol{M}_k = (-\boldsymbol{D}_k)^{-1}$. The joint moment for $h = 1$ and MDMAPs for rank R with product form is given by

$$\boldsymbol{J}_1^{m,n}(k,l) = m!n!\boldsymbol{\pi}_{k,l}\left((\boldsymbol{M}_k)^m \otimes \boldsymbol{H}_l\right)\underbrace{\sum_{r=1}^{R}\lambda^{(r)}\left(\mathbb{1}_{n_k} \otimes \boldsymbol{G}_l^{(r)}\right)}_{\boldsymbol{G}_{kl}\left(\mathbb{1}_{n_l}\otimes \boldsymbol{I}_{n_l}\right)}(\boldsymbol{M}_l)^n\,\mathbb{1}_{n_l} \tag{14}$$

For two components k and l the joint distribution is given by

$$\boldsymbol{F}_0^{x,y}(k,l) = \boldsymbol{\pi}_{kl}\left(\int_0^x e^{\tau \boldsymbol{D}_k}\boldsymbol{d}_k d\tau \otimes \int_0^y e^{\tau \boldsymbol{D}_l}\boldsymbol{d}_l d\tau\right) \tag{15}$$

For successive observations. $\boldsymbol{F}_1^{x,y}(k,l)$ denotes the conditional probability that we observe values $\leq x$ for k and for the next observation z_l of dimension l $z_l \leq y$ holds. The function can be computed using the following equation.

$$\boldsymbol{F}_1^{x,y}(k,l) = \boldsymbol{\pi}_{kl}\left(\int_0^x e^{\tau \boldsymbol{D}_k}\,\mathrm{diag}\,(\boldsymbol{d}_k)\,d\tau \otimes \boldsymbol{H}_l\right)\underbrace{\sum_{r=1}^{R}\lambda^{(r)}\mathbb{1}_{n_k}^{outp} \otimes \boldsymbol{G}_l^{(r)}}_{\boldsymbol{G}_{kl}\left(\mathbb{1}_{n_l}\otimes \boldsymbol{I}_{n_l}\right)}\int_0^y e^{\tau \boldsymbol{D}_l}\boldsymbol{d}_l d\tau \tag{16}$$

Observe that (13) and (15) as well as (14) and (16) are of an identical structure. Therefore we define a common notation which allows us to handle joint moments and values of the distributions functions in a common framework. We denote these measures as zero or first order quantities, respectively. For zero order joint moments and distribution functions we have

$$\boldsymbol{\Theta}_0^{\alpha,\beta}(k,l) = \boldsymbol{\pi}_{kl}\left(\boldsymbol{\phi}_k^{\alpha} \otimes \boldsymbol{\phi}_l^{\beta}\right) \tag{17}$$

where $\boldsymbol{\phi}_l^{\alpha}$ equals $\alpha!\boldsymbol{M}_l^{\alpha}\mathbb{1}_{n_l}$ or $\int_0^{\alpha} e^{\tau \boldsymbol{D}_0}\boldsymbol{d}_l d\tau$ and $\boldsymbol{\Theta}_0^{\alpha,\beta}(k,l)$ equals $\boldsymbol{J}_0^{\alpha,\beta}(k,l)$ or $\boldsymbol{F}_0^{x,y}(k,l)$, $\alpha,\beta \in \mathbb{N}$ for joint moments and $\alpha,\beta \in \mathbb{R}_{>0}$ for joint densities. $\hat{\boldsymbol{\Theta}}_0^{\alpha,\beta}(k,l)$ is then the estimated value for $\boldsymbol{\Theta}_0^{\alpha,\beta}(k,l)$. Similarly we can define a common description of first order joint moments or values of the distribution function.

$$\boldsymbol{\Theta}_1^{\alpha,\beta}(k,l) = \boldsymbol{\pi}_{kl}\left(\boldsymbol{\Xi}_k^{\alpha} \otimes \boldsymbol{H}_l\right)\boldsymbol{G}_{k,l}\left(\mathbb{1}_{n_l} \otimes \boldsymbol{\phi}_l^{\beta}\right) \tag{18}$$

where $\boldsymbol{\Theta}_0^{\alpha,\beta}(k,l)$ equals $\boldsymbol{J}_1^{\alpha,\beta}(k,l)$ or $\boldsymbol{F}_1^{x,y}(k,l)$, $\hat{\boldsymbol{\Theta}}_1^{\alpha,\beta}(k,l)$ is the corresponding estimate and $\boldsymbol{\Xi}_k$ equals $\alpha!\boldsymbol{M}_k^{\alpha}\boldsymbol{H}_k$ or $\int_0^{\alpha} e^{\tau \boldsymbol{D}_k}\,\mathrm{diag}(\boldsymbol{d}_k)d\tau$.

5 Moment-Based Parameter Fitting

We consider different approaches to determine the parameters of an MDMAP based on moments, joint moments and joint values of the probability densities. For the methods we distinguish between general MDMAPs and MDMAPs of rank R with product form. The approaches are based on algorithms that have been proposed for MAPs and MMAPs [12, 9, 19]. In all cases we start with the computation of a PHD $(\boldsymbol{\pi}_k, \boldsymbol{D}_k)$ from the observations $x_k^{(l)}$ $(l = 1, \ldots, L)$. For this purpose any algorithm for parameter fitting of PHDs can be applied, the resulting PHD can be further transformed to increase the number of input and output states using equivalence transformations [8]. The corresponding approach is denoted as two-phase fitting approach [19] and sometimes becomes a three-phase approach in this paper. The computation in different phases allows one to formulate the resulting optimization problems as non-negative least squares problem with linear constraints that can be solved efficiently. Furthermore, it is a common approach used for the parameter fitting of multivariate distributions in general [5]. Most fitting methods for multivariate distributions use $\hat{\boldsymbol{J}}_0^{1,1}(k,l)$ and $\hat{\boldsymbol{J}}_1^{1,1}(k,l)$ as measures to be matched by the multivariate distribution which is often a multivariate normal distribution. This is sometimes criticized [3, 4] and other measures like the joint tail behavior of two components are considered. In the following approaches measures such as $\hat{\boldsymbol{J}}_0^{m,n}(k,l)$, $\hat{\boldsymbol{F}}_0^{x,y}(k,l)$ (i.e., $\hat{\boldsymbol{\Theta}}_0^{\alpha,\beta}(k,l)$) and $\hat{\boldsymbol{J}}_1^{m,n}(k,l)$, $\hat{\boldsymbol{F}}_1^{x,y}(k,l)$ (i.e., $\hat{\boldsymbol{\Theta}}_1^{\alpha,\beta}(k,l)$) are incorporated in the fitting process. We do not consider dependencies of lags larger than 1 like $\hat{\boldsymbol{J}}_p^{1,1}(k,l)$ $(p > 1)$ which are used in VARTA processes [6, 21].

5.1 Dependencies in a Single Component

For a single component the zero and first order quantities are given by

$$\boldsymbol{\Theta}_0^{\alpha}(k,k) = \boldsymbol{\pi}_k \boldsymbol{\phi}_k^{\alpha} \text{ and } \boldsymbol{\Theta}_1^{\alpha,\beta}(k,k) = \boldsymbol{\xi}_k^{\alpha} \boldsymbol{G}_k \boldsymbol{\phi}_k^{\beta} \tag{19}$$

where $\boldsymbol{\xi}_k^{\alpha} = \boldsymbol{\pi}_k \boldsymbol{\Xi}_k^{\alpha}$. Observe that $\boldsymbol{\Theta}_0^{\alpha}(k,k)$ has only a single parameter α and is completely determined by the PHD $(\boldsymbol{\pi}_k, \boldsymbol{D}_k)$. To describe $\boldsymbol{\Theta}_1^{\alpha,\beta}(k,k)$ we expand the PHD into a MAP $(\boldsymbol{D}_k, \boldsymbol{G}_k)$ (see [12, 19]). Now assume that we have H_k estimates $\hat{\boldsymbol{\Theta}}_1^{\alpha_i,\beta_i}(k,k)$ which should be approximated by $(\boldsymbol{D}_k, \boldsymbol{G}_k)$. Computation of matrix $\boldsymbol{G}_k$ results then in the following *Non-Negative Least Squares Problem with Linear Constraints* (NNLSPLC) [25].

$$\begin{aligned} &\min_{\boldsymbol{G}_k \geq \mathbf{0}} \left(\sum_{i=1}^{H_k} \mu_i \left(\hat{\boldsymbol{\Theta}}_1^{\alpha_i,\beta_i}(k,k) - \boldsymbol{\xi}_k^{\alpha_i} \boldsymbol{G}_k \boldsymbol{\phi}_k^{\beta_i} \right)^2 \right) \\ &\text{subject to } \boldsymbol{G}_k \mathbb{1}_{n_k}^{inp} = \mathbb{1}_{n_k}^{outp}, \boldsymbol{\psi}_k \boldsymbol{G}_k = \boldsymbol{\pi}_k \end{aligned} \tag{20}$$

where μ_i are non-negative weights for the different joint moments/densities. The problem has $n_k^c n_k^r$ variables and $n_k^r + n_k^c$ constraints. The number of non-zero elements in $\boldsymbol{G}_k$ is at most $n_k^r n_k^c$ but often the optimal solution describes a corner case with less non-zero elements. In some situations, it is better to have some more non-zero elements to allow more flexibility for following optimization steps. This can be achieved by adding a penalty term $\lambda \|\boldsymbol{G}_k\|_2$ to the objective function. In this case a matrix with more and smaller non-zero elements results in a smaller two norm. This step can be applied in all NNSPLCs we present in the following paragraphs.

5.2 Joint Moment Fitting for General MDMAPs

For general MDMAPs we put no restriction on matrix $\boldsymbol{G}$ which means that $n_{1:K}^c n_{1:K}^r$ variables are available. However, some constraints exist. First, for the row sums $\boldsymbol{G}(i_{1:K}\bullet)\mathbb{1} = 1$ for $i_{1:K} \in out_{1:K}$ and 0 otherwise has to hold. Furthermore, $\boldsymbol{G}$ determines $\boldsymbol{\pi}$ because (6) has to hold for given matrices $\boldsymbol{H}_k$. Additionally, the availability of the distributions $(\boldsymbol{\pi}_k, \boldsymbol{D}_k)$ implies that $\boldsymbol{\pi}\boldsymbol{V}_k = \boldsymbol{\pi}_k$ has to hold.

The parameter fitting is done in two steps. First, vector $\boldsymbol{\pi}$ is determined to approximate quantities $\hat{\boldsymbol{\Theta}}_0^{\alpha_i,\beta_i}(k_i,l_i)$ $(i = 1, \ldots, I_0)$. Then an appropriate matrix $\boldsymbol{G}$ is determined to approximate additional values $\hat{\boldsymbol{\Theta}}_1^{\alpha_i,\beta_i}(k_i,l_i)$ $(i = 1, \ldots, I_1)$.

We begin with the computation of $\boldsymbol{\pi}$ from the zero order quantities. Let $\boldsymbol{u}^{(i)} = \boldsymbol{V}_{kl} \left(\boldsymbol{\phi}_{k_i}^{\alpha_i} \otimes \boldsymbol{\phi}_{l_i}^{\beta_i} \right)$. E.g., if all first joint moments $\boldsymbol{J}_0^{1,1}(k,l)$ are considered, then $(K-1)K/2$ vectors $\boldsymbol{V}_{kl} \left(\boldsymbol{m}_k^1 \otimes \boldsymbol{m}_l^1 \right)$ $(k < l)$ are used. Then $\boldsymbol{\Theta}_0^{\alpha_i,\beta_i} = \boldsymbol{\pi}\boldsymbol{u}^{(i)}$. With these notations we can set up the following NNLSPLC.

$$\begin{aligned} &\min_{\boldsymbol{\pi} \geq \mathbf{0}} \left(\sum_{i=1}^{I_0} \mu_i \left(\hat{\boldsymbol{\Theta}}_0^{\alpha_i,\beta_i}(k_i,l_i) - \boldsymbol{\pi}\boldsymbol{u}^{(i)} \right)^2 \right) \\ &\text{subject to } \boldsymbol{\pi}\mathbb{1} = 1, \boldsymbol{\pi} \geq \mathbf{0} \text{ and } \boldsymbol{\pi}\boldsymbol{V}_k = \boldsymbol{\pi}_k \text{ for all } k = 1, \ldots, K \end{aligned} \tag{21}$$

Again μ_i are appropriate non-negative weights. If the minimum of the objective function becomes 0, then all joint moments and conditional values of the distribution function are matched exactly. The result is a set of PHDs coupled via initial vector $\boldsymbol{\pi}$ that generate random vectors. Vector $\boldsymbol{\pi}$ contains $n_{1:K}$ elements of which are at most $n^r_{1:K}$ are non-zero.

To match estimated values $\hat{\boldsymbol{\Theta}}_1^{\alpha_i,\beta_i}(k_i,l_i)$ $(i=1,\ldots,I_1)$, we assume that the vector $\boldsymbol{\pi}$ is available (e.g. from (21)). This implies that $\boldsymbol{\psi} = \boldsymbol{\pi} \otimes_{k=1}^{K} \boldsymbol{H}_k$ is also available. The optimization problem for general matrices $\boldsymbol{G}$ results in the following NNLSPLC.

$$\begin{aligned} &\min_{\boldsymbol{G}\geq\boldsymbol{0}} \left(\sum_{i=1}^{I_1} \mu_i \left(\hat{\boldsymbol{\Theta}}_1^{\alpha_i,\beta_i}(k_i,l_i) - \boldsymbol{w}^{(i)} \boldsymbol{G} \boldsymbol{v}^{(i)} \right)^2 \right) \\ &\text{subject to } \boldsymbol{G}\mathbb{1}^{inp}_{n_{1:k}} = \mathbb{1}^{outp}_{n_{1:K}} \text{ and } \boldsymbol{\psi}\boldsymbol{G} = \boldsymbol{\pi} \end{aligned} \tag{22}$$

where

$$\begin{aligned} \boldsymbol{w}^{(i)} &= \boldsymbol{\pi} \left(\boldsymbol{I}_{n_{1:k_i-1}} \otimes \boldsymbol{\Xi}^{\alpha_i}_{k_i} \otimes \boldsymbol{I}_{n_{k_i+1:K}} \right) \prod_{j=1,j\neq k_i}^{K} \left(\boldsymbol{I}_{n_{1:j-1}} \otimes \boldsymbol{H}_j \otimes \boldsymbol{I}_{n_{j+1:K}} \right) \\ \boldsymbol{v}^{(i)} &= \mathbb{1}_{n_{1:l_i-1}} \otimes \boldsymbol{\phi}^{\beta_i}_{l_i} \otimes \mathbb{1}_{n_{l_i+1:K}} \end{aligned}$$

The problem contains $n^c_{1:K} n^r_{1:K}$ non-zero variables, after removing zero elements from $\boldsymbol{G}$, but has relatively simple equality constraints.

5.3 Joint Moment Fitting for MDMAPs of Rank R with Product Form

We begin with the generation of product form MDMAPs of rank R. As long as we consider the approximation of quantities $\hat{\boldsymbol{\Theta}}_0^{\alpha,\beta}(k,l)$ Theorem 2 applies and allows us to compute the distribution $\boldsymbol{\pi}_{kl}$ from the matrices for components k and l, independently of the remaining components. Unfortunately, the joint computation of the matrices $\boldsymbol{G}_k^{(r)}$ and $\boldsymbol{G}_l^{(r)}$ results in a non-linear optimization problem which is hard to solve. To keep the optimization manageable, we use *Alternating Least Squares* (ALS) [22] which is a common approach applied in many areas including the solution of partial differential equations [16, 17] or performance models [10]. The basic idea of the approach is fairly simple. It is assumed that matrices $\boldsymbol{G}_l^{(r)}$ $(l \in \{1,\ldots,K\}\setminus\{k\}, r=1,\ldots,R)$ are known when matrices $\boldsymbol{G}_k^{(r)}$ are computed. Then new matrices are computed for $k=1,\ldots,K$ and the iteration is repeated until convergence is observed. Some results about local convergence of the approach exist [30] and also hold in our setting, but will not be further analyzed.

To start with the computation we assume that initial matrices $\boldsymbol{G}_k^{(r)}$ are available. Matrices $\boldsymbol{G}_k$ result from the solution of (20) or are initialized as $\mathbb{1}^{outp}_{n_k}\boldsymbol{\pi}_k$. Then a random distribution $(\lambda^{(1)},\ldots,\lambda^{(R)})$ with $0<\lambda^{(r)}<1$ and $\sum_{r=1}^{R}\lambda^{(r)}=1$ is generated. The result is an MDMAP of rank R with product form but different components are uncorrelated.

To introduce correlation between two components k and l, we consider quantities $\hat{\boldsymbol{\Theta}}_0^{\alpha_i,\beta_i}(k,l)$ $(i=1,\ldots,I_0^{kl})$ which results in the following NNLSPLC.

$$\begin{aligned}&\min_{\boldsymbol{\pi}_{kl}\geq\mathbf{0}}\left(\sum_{i=1}^{I_0^{kl}}\mu_i\left(\hat{\boldsymbol{\Theta}}_0^{\alpha_i,\beta_i}(k,l)-\boldsymbol{\pi}_{kl}\left(\boldsymbol{\phi}_k^{\alpha_i}\otimes\boldsymbol{\phi}_l^{\beta_i}\right)\right)^2\right)\\&\text{subject to } \boldsymbol{\pi}_{kl}\mathbb{1}=1,\boldsymbol{\pi}_{kl}\left(\mathbb{1}_{n_k}\otimes\boldsymbol{I}_{n_l}\right)=\boldsymbol{\pi}_l \text{ and } \boldsymbol{\pi}_{kl}\left(\boldsymbol{I}_{n_k}\otimes\mathbb{1}_{n_k}\right)=\boldsymbol{\pi}_l\end{aligned}\tag{23}$$

Up to $K(K-1)/2$ NNLSPLCs of the above type have to be solved. From the resulting vectors $\boldsymbol{\pi}_{kl}$ the vectors $\boldsymbol{\psi}_{kl}=\boldsymbol{\pi}_{kl}\left(\boldsymbol{H}_k\otimes\boldsymbol{H}_l\right)$ can be computed.

In the next step matrices $\boldsymbol{G}_k^{(r)}$ have to be found such that the vectors computed in (23) are the embedded stationary vectors of the two components. Due to the product form it is sufficient to consider only the relation between two components if we restrict dependencies to joint moments or densities between two components. Let $\bar{\boldsymbol{G}}_k^{(r)}=\lambda^{(r)}\boldsymbol{G}_k^{(r)}$. If we consider the local optimization problem, where matrices $\bar{\boldsymbol{G}}_k^{(r)}$ are unknown and matrices $\boldsymbol{G}_l^{(r)}$ $(l\neq k)$ are known, we have to find matrices such that

$$\boldsymbol{\psi}_{kl}\left(\sum_{r=1}^{R}\bar{\boldsymbol{G}}_k^{(r)}\otimes\boldsymbol{G}_l^{(r)}\right)=\boldsymbol{\pi}_{kl}\ (k<l) \text{ and } \boldsymbol{\psi}_{lk}\left(\sum_{r=1}^{R}\boldsymbol{G}_l^{(r)}\otimes\bar{\boldsymbol{G}}_k^{(r)}\right)=\boldsymbol{\pi}_{kl}\ (k>l)\tag{24}$$

This can be describes in the following NNLSPLC.

$$\begin{aligned}\min_{\bar{\boldsymbol{G}}_k^{(1)},\ldots,\bar{\boldsymbol{G}}_k^{(R)},\lambda^{(1)},\ldots,\lambda^{(R)}\geq 0}&\left(\sum_{l=1}^{k-1}\left\|\boldsymbol{\pi}_{lk}-\boldsymbol{\psi}_{lk}\sum_{r=1}^{R}\left(\boldsymbol{G}_l^{(r)}\otimes\bar{\boldsymbol{G}}_k^{(r)}\right)\right\|_2^2\right.\\&+\sum_{l=k+1}^{K}\left\|\boldsymbol{\pi}_{kl}-\boldsymbol{\psi}_{kl}\sum_{r=1}^{R}\left(\bar{\boldsymbol{G}}_k^{(r)}\otimes\boldsymbol{G}_l^{(r)}\right)\right\|_2^2\\&\left.+\sum_{(h,l),h<l,h,l\neq k}\left\|\boldsymbol{\pi}_{hl}-\boldsymbol{\psi}_{hl}\sum_{r=1}^{R}\lambda^{(r)}\left(\boldsymbol{G}_h^{(r)}\otimes\boldsymbol{G}_l^{(r)}\right)\right\|_2^2\right)\\\text{subject to }\sum_{r=1}^{R}\bar{\boldsymbol{G}}_k^{(r)}\mathbb{1}=\mathbb{1},\boldsymbol{\psi}_k\sum_{r=1}^{R}\bar{\boldsymbol{G}}_k^{(r)}=\boldsymbol{\pi}_k,&\text{ for all } i:\bar{\boldsymbol{G}}_k^{(r)}(i\bullet)\mathbb{1}=\lambda^{(r)}\end{aligned}\tag{25}$$

If matrices $\boldsymbol{G}_k$ are available from (20), then the second set of constraints can be substituted by $\sum_{r=1}^{R}\boldsymbol{G}_k^{(r)}=\boldsymbol{G}_k$. In this case, a solution assures that values $\boldsymbol{J}_1^{m,n}(k,k)$ are kept by the resulting MDMAP. The optimization problem is solved for $k=1,\ldots,K$ and this process is iterated until the objective function becomes 0 for all components or no progress is made any more. Observe that a solution of (25) cannot increase the overall error defined as

$$\sum_{k=1}^{K}\sum_{l=k+1}^{K}\left\|\boldsymbol{\pi}_{kl}-\boldsymbol{\psi}_{kl}\left(\sum_{r=1}^{R}\lambda^{(r)}\left(\boldsymbol{G}_k^{(r)}\otimes\boldsymbol{G}_l^{(r)}\right)\right)\right\|_2^2.\tag{26}$$

If the global error cannot be reduced to 0, then vectors $\boldsymbol{\pi}_{kl}$ have to be computed for the resulting MDMAP from which the joint moments and joint densities can be recomputed.

6 Examples

In the following we consider different examples for MDMAPs. First, random vectors where the components of one vector are correlated and subsequent vectors are independent are considered, then random vectors which with correlation between components of one vector and of subsequent vectors are analyzed.

6.1 Independent Random Vectors

We begin with random vectors with correlated components. A first simple example are two correlated exponential distributions. We consider exponential distributions with rate 1 and correlation coefficient $\boldsymbol{R}_1(1,2) = \boldsymbol{R}_1(2,1) = 0.5$. To build correlated exponential distributions, the following representation as PHD with n phases is used [7].

$$\boldsymbol{\pi} = \left(n^{-1}, \ldots, n^{-1}\right) \quad \boldsymbol{D} = \begin{pmatrix} -1 & 1 & & \\ & -2 & 2 & \\ & & \ddots & \ddots \\ & & & -n \end{pmatrix} \tag{27}$$

To obtain a coefficient of correlation of 0.5 at least 5 phases are needed. Observe that in the representation (27) the expected time to absorption is decreasing when entering the distribution at a state with larger index. To obtain a positive correlation if two distributions are coupled, both distributions have to start with a higher probability in the same state. In an MDMAP with two exponential PHDs of order 5, the joint initial vector has 25 entries. Let $\boldsymbol{\pi}(i,j)$ be the probability that the MDMAP starts in phase i of the first and phase j of the second MDMAP. For independent PHDs the probability is n^{-2} for each state, the coefficient of correlation is 0 in this case. By solving (21) we obtain an initial vector with only 7 non-zero entries, namely $\boldsymbol{\pi}(1,1) = \boldsymbol{\pi}(5,5) = 0.1902$, $\boldsymbol{\pi}(2,2) = \boldsymbol{\pi}(3,3) = \boldsymbol{\pi}(4,4) = 0.2$ and $\boldsymbol{\pi}(1,5) = \boldsymbol{\pi}(5,1) = 0.00998$. The resulting MDMAP describes two exponential distribution with rates 1 and correlation coefficient 0.5.

In exactly the same way random vectors with several correlated exponentially distributed components can be generated. It should be mentioned that even in simulation the generation of high dimensional random vectors of correlated exponential distributions is non-trivial. We applied the method from [6, 14] which transforms correlated standard normal distributed random vectors into exponential distributions. We consider the case of three correlated exponential

distributions all with mean 1 and the following two correlation matrices.

$$\boldsymbol{R}_0 = \begin{pmatrix} 1 & 0.5 & 0.5 \\ 0.5 & 1 & 0.5 \\ 0.5 & 0.5 & 1 \end{pmatrix} \text{ and } \boldsymbol{R}_0' = \begin{pmatrix} 1 & 0.5 & 0.1 \\ 0.5 & 1 & 0.3 \\ 0.1 & 0.3 & 1 \end{pmatrix}$$

Again we use the above representation with 5 states for the exponential distribution. Thus, the joint state space contains 125 states. For the first correlation matrix the algorithm generates an initial vector with 64 non-zero entries that exactly matches the correlation structure. For the second correlation matrix the function *lsqlin* of *octave* or *matlab* generates an MDMAP with the following correlation matrix.

$$\hat{\boldsymbol{R}}_0 = \begin{pmatrix} 1.00000 & 0.4965 & 0.1011 \\ 0.4965 & 1.00000 & 0.2987 \\ 0.1011 & 0.2987 & 1.00000 \end{pmatrix}$$

This is very near to the required correlation but not exactly the same. Interestingly in the resulting initial vector only 14 of the 125 entries are non-zero (if we set values smaller than $1.0e-8$ to zero).

If we use a product form approximation, then a rank of 4 is required to approximate matrix $\boldsymbol{R}_0$ with a relative error of less than 1%. The correlation described by $\boldsymbol{R}_0'$ could not be approximated with a small approximation error using a product form representation.

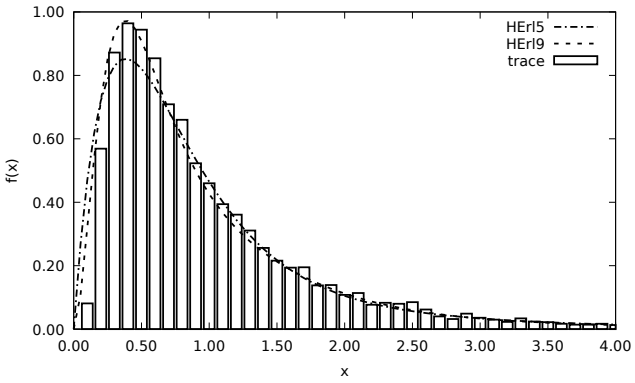

Fig. 1. Densities of the log-normal and the fitted Hyper-Erlang distributions.

As a second example we consider log-normal distribution with means and standard deviation 1. In a first step $10,000$ samples are drawn from the distribution and are used to fit a Hype-Erlang distribution using the software *gfit* [33]. Figure 1 show the empirical density of the trace and the densities of hyper-Erlang distributions with 5 and 9 states. It can be noticed that both Hyper-Erlang distributions provide a good matching of the empirical density. The 5 state Hype-Erlang distribution consists of 3 Erlang branches, one with 1 phase and the other two with 2 phases. The 9 state Hyper-Erlang distribution contains

4 branches, one with 1 phase, 1 with 2 phases and 1 with 3 phases. Combining two Hyper-Erlang distributions with 5 phases allows us to express a correlation coefficient up to 0.3 whereas 9 phases allow one to model correlation coefficients up to 0.4. If we combine the Hyper-Erlang distribution with 9 phases, modeling the log-normal distribution and the PHD with 5 states representing the exponential distribution, coefficients of correlation between -0.29 and 0.34 can be achieved. To obtain larger coefficient of correlation additional phases have to be added.

If we consider product from representations for the MDMAP with two 9-state hyper-Erlang distributions, then for $\boldsymbol{R}_0^{1,1}(1,2) = 0.1$ and 0.2 rank 5 representations are computed, whereas for $\boldsymbol{R}_0^{1,1}(1,2) = 0.3$ a rank 3 representation is generated.

6.2 Random Processes

We consider again correlated exponential distributions. With the representation of (27) it is not possible to model correlations between subsequent realization because the representation has only a single output state. If we enlarge the number of phases by using a Hyper-Erlang representation where each path starting in phase i $(= 1, \ldots, n)$ and ending in phase n is modeled by a single Erlang branch, we obtain a distribution with $n(n+1)/2$ phases, n input and n output states. We analyze an MDMAP with 2 distributions with $n = 5$. The correlation coefficient reachable by these distribution ranges between -0.46 and 0.54.

	$\max(R_1(1,1)$ $\max(R_1(1,2))$		$\min(R_1(1,1))$ $\min(R_1(1,2))$		$\max(R_1(1,1))$ $\min(R_1(1,2))$		$\min(R_1(1,1))$ $\max(R_1(1,2))$	
$R_0(1,2)$	$R_1(1,1)$	$R_1(1,2)$	$R_1(1,1)$	$R_1(1,2)$	$R_1(1,1)$	$R_1(1,2)$	$R_1(1,1)$	$R_1(1,2)$
-0.40	0.17	0.17	-0.17	-0.17	0.53	-0.44	-0.44	0.53
-0.30	0.23	0.23	-0.24	-0.24	0.50	-0.40	-0.40	0.50
-0.20	0.28	0.28	-0.28	-0.28	0.46	-0.38	-0.38	0.46
-0.10	0.32	0.32	-0.31	-0.31	0.43	-0.34	-0.34	0.43
0.00	0.36	0.36	-0.34	-0.34	0.40	-0.30	-0.30	0.40
0.10	0.40	0.40	-0.36	-0.36	0.36	-0.26	-0.26	0.36
0.20	0.43	0.43	-0.38	-0.38	0.32	-0.22	-0.22	0.32
0.30	0.47	0.47	-0.40	-0.40	0.27	-0.17	-0.17	0.27
0.40	0.49	0.49	-0.42	-0.42	0.21	-0.11	-0.11	0.21
0.54	0.54	0.54	-0.46	-0.46	0.05	0.04	0.04	0.05

Table 1. Maximal and minimal reachable correlation coefficients.

To analyze the flexibility of the representation, we first select some value for $R_0(1,2)$ and compute the corresponding vector $\boldsymbol{\pi}$ by solving (21). Then we try to maximize/minimize $R_1(1,1)$ and $R_1(1,2)$. If we maximize/minimize only one of the two values, then independently of $R_0(1,2)$ the values for $R_1(1,1)$ and $R_1(1,2)$ can range between -0.45 and 0.54, the maximum range reachable by this distribution. If we try to jointly maximize/minimize $R_1(1,1)$ and $R_1(1,2)$, then the range shrinks and depends on $R_0(1,2)$. Results are shown in Table 1 and indicate that there is still a lot flexibility in the representation. The solution

of the NNLSPLC problems for this examples requires with *matlab* on a standard PC less than a second.

Finally we consider an MDMAP with three PHDs of the mentioned type. The joint state space of this process contains 15^3 states, 125 input and 125 output states. We define the following two matrices for the correlation of lag 0 and 1

$$\boldsymbol{R}_0 = \begin{pmatrix} 1 & 0.5 & 0.1 \\ 0.5 & 1 & 0.3 \\ 0.1 & 0.3 & 1 \end{pmatrix} \text{ and } \boldsymbol{R}_1 = \begin{pmatrix} 0.3 & 0.2 & 0.1 \\ 0.2 & 0.3 & 0.2 \\ 0.1 & 0.2 & 0.3 \end{pmatrix}$$

In a first step the initial vector is computed and then matrix $\boldsymbol{G}$ is determined resulting in an MDMAP with the following matrices $\hat{\boldsymbol{R}}_0$ (which is already shown above) and $\hat{\boldsymbol{R}}_1$.

$$\hat{\boldsymbol{R}}_0 = \begin{pmatrix} 1.00000 & 0.4965 & 0.1011 \\ 0.4965 & 1.00000 & 0.2987 \\ 0.1011 & 0.2987 & 1.00000 \end{pmatrix} \text{ and } \hat{\boldsymbol{R}}_1 = \begin{pmatrix} 0.2636 & 0.2341 & 0.0912 \\ 0.2341 & 0.2692 & 0.2072 \\ 0.0912 & 0.2072 & 0.2987 \end{pmatrix}$$

It can be seen that the correlation structure is approximated with small approximation errors. The computation of the initial vector requires negligible time, whereas the solution of the second NNLSPLC problem to determine matrix $\boldsymbol{G}$ requires about an hour of CPU time.

7 Conclusion

In this paper we present MDMAPs, a Markov model for random vectors that may be correlated in different dimensions and extends *Phase Type Distributions* and *Markovian Arrival Processes* to the multi-dimensional case. It is shown how MDMAPs can be analyzed and algorithms are presented to fit the parameters of MDMAPs according to joint moments or some values of the conditional distribution function. The proposed model can be applied in input modeling for simulation models where it is an alternative for models that are based on transformed correlated normal distributions. These models usually only use the correlation coefficient to describe dependencies whereas MDMAPs can also use higher order joint moments or values of the conditional distribution function which introduces additional flexibility when real data has to be represented by a stochastic model. Since MDMAPs are Markov models they can be analyzed numerically and can also be used as a stochastic model for correlated failures in dependability models or to represent correlated arrivals and service times in queues with PHD arrivals and services as in [11].

We currently have a first prototype matlab implementation of the algorithms proposed in the paper. This representation will be further improved and then made publically available in the tool ProFiDo [2]. Apart from parameter fitting with respect to moments and joint moments also an EM algorithm for MDMAPs will be considered in future research.

References

1. D. Assaf, N. A. Langberg, T. H. Savits, and M. Shaked. Multivariate phase-type distributions. *Oper. Res.*, 32(3):688–702, 1984.
2. F. Bause, P. Buchholz, and J. Kriege. ProFiDo - the processes fitting toolkit dortmund. In *QEST 2010, Seventh International Conference on the Quantitative Evaluation of Systems, Williamsburg, Virginia, USA, 15-18 September 2010*, pages 87–96. IEEE Computer Society, 2010.
3. B. Biller. Copula-based multivariate input models for stochastic simulation. *Operations Research*, 57(4):878–892, 2009.
4. B. Biller. Copula-based multivariate input modeling. *Surveys in Operations Research and Management Sciences*, 17:69–84, 2012.
5. B. Biller and S. Ghosh. Multivariate input processes. In S. G. Henderson and B. L. Nelson, editors, *Handbook of OR & MS*, volume 13, pages 123–152. Elsevier, 2006.
6. B. Biller and B. L. Nelson. Modeling and generating multivariate time-series input processes using a vector autoregressive technique. *ACM Trans. Model. Comput. Simul.*, 13(3):211–237, 2003.
7. M. Bladt and B. F. Nielsen. On the constructionof bivariate exponential distributions with an arbitrary correlation coefficient. *Stochastic Models*, 26(3):295–308, 2010.
8. P. Buchholz, I. Felko, and J. Kriege. Transformation of acyclic phase type distributions for correlation fitting. In A. N. Dudin and K. D. Turck, editors, *Analytical and Stochastic Modelling Techniques and Applications - 20th International Conference, ASMTA 2013, Ghent, Belgium, July 8-10, 2013. Proceedings*, volume 7984 of *Lecture Notes in Computer Science*, pages 96–111. Springer, 2013.
9. P. Buchholz, P. Kemper, and J. Kriege. Multi-class Markovian arrival processes and their parameter fitting. *Perform. Eval.*, 67(11):1092–1106, 2010.
10. P. Buchholz and J. Kriege. Approximate aggregation of markovian models using alternating least squares. *Perform. Evaluation*, 73:73–90, 2014.
11. P. Buchholz and J. Kriege. Fitting correlated arrival and service times and related queueing performance. *Queueing Syst.*, 85(3-4):337–359, 2017.
12. P. Buchholz, J. Kriege, and I. Felko. *Input Modeling with Phase-Type Distributions and Markov Models - Theory and Applications*. Springer, 2014.
13. P. Buchholz and M. Telek. Rational processes related to communicating Markov processes. *J. Applied Probability*, 49(1):40–59, 2012.
14. M. C. Cario and B. L. Nelson. Numerical methods for fitting and simulating autoregressive-to-anything processes. *INFORMS J. Comput.*, 10(1):72–81, 1998.
15. A. D'Alconzo, I. Drago, A. Morichetta, M. Mellia, and P. Casas. A survey on big data for network traffic monitoring and analysis. *IEEE Trans. Network and Service Management*, 16(3):800–813, 2019.
16. S. Dolgov and D. V. Savostyanov. Alternating minimal energy methods for linear systems in higher dimensions. *SIAM J. Scientific Computing*, 36(5), 2014.
17. S. Holtz, T. Rohwedder, and R. Schneider. The alternating linear scheme for tensor optimization in the tensor train format. *SIAM J. Scientific Computing*, 34(2), 2012.
18. Y. Hong, M. Zhang, and W. Q. Meeker. Big data and reliability applications: The complexity dimension. *Journal of Quality Technology*, 50(2):135–149, 2018.
19. G. Horváth, M. Telek, and P. Buchholz. A MAP fitting approach with independent approximation of the inter-arrival time distribution and the lag-correlation. In *QEST*, pages 124–133. IEEE CS Press, 2005.

20. J. G. Kemeny and J. L. Snell. Finite continuous time Markov chains. *Theory of Probability & Its Applications*, 6(1):101–105, 1961.
21. J. Kriege and P. Buchholz. Traffic modeling with phase-type distributions and VARMA processes. In G. Agha and B. V. Houdt, editors, *Quantitative Evaluation of Systems - 13th International Conference, QEST 2016, Quebec City, QC, Canada, August 23-25, 2016, Proceedings*, volume 9826 of *Lecture Notes in Computer Science*, pages 295–310. Springer, 2016.
22. P. M. Kroonenberg and J. D. Leeuw. Principal component analysis of three-mode data by means of alternating least squares algorithms. *Psychometrika*, 45:69–97, 1980.
23. V. G. Kulkarni. A new class of multivariate phase type distributions. *Oper. Res.*, 37(1):151–158, 1989.
24. A. M. Law. *Simulation Modeling and Analysis*. Mc Graw Hill, 5 edition, 2013.
25. C. L. Lawson and R. J. Hanson. *Solving Least Squares Problems*. Classics in Applied Mathematics. SIAM, 1995.
26. D. M. Lucantoni. New results on the single server queue with a batch Markovian arrival process. *Stochastic Models*, 7(1):1–46, 1991.
27. D. Mishra, A. Gunasekaran, T. Papadopoulos, and S. J. Childe. Big data and supply chain management: a review and bibliometric analysis. *Annals OR*, 270(1-2):313–336, 2018.
28. M. F. Neuts. A versatile Markovian point process. *Journal of Applied Probability*, 16:764–779, 1979.
29. M. F. Neuts. *Matrix-geometric solutions in stochastic models - an algorithmic approach*. Johns Hopkins University Press, 1981.
30. T. Rohwedder and A. Uschmajew. On local convergence of alternating schemes for optimization of convex problems in the tensor train format. *SIAM J. Numer. Anal.*, 51(2):1134–1162, 2013.
31. W. J. Stewart. *Probability, Markov Chains, Queues, and Simulation: The Mathematical Basis of Performance Modeling*. Princeton University Press, 2009.
32. M. Telek and G. Horváth. A minimal representation of Markov arrival processes and a moments matching method. *Perform. Eval.*, 64(9-12):1153–1168, 2007.
33. A. Thümmler, P. Buchholz, and M. Telek. A novel approach for phase-type fitting with the EM algorithm. *IEEE Trans. Dep. Sec. Comput.*, 3(3):245–258, 2006.
34. K. S. Trivedi. *Probability and Statistics with Reliability, Queuing and Computer Science Applications, Second Edition*. Wiley, 2 edition, 2016.

A Proofs of the Theorems

We use the following vectors $\mathbb{1}^{inp}_{n_\mathcal{K}}$ where $\mathbb{1}^{inp}_{n_\mathcal{K}}(i_\mathcal{K}) = 1$ for $i_\mathcal{K} \in inp_\mathcal{K}$ and 0 otherwise. Similarly, $\mathbb{1}^{outp}_{n_\mathcal{K}}$ where $\mathbb{1}^{outp}_{n_\mathcal{K}}(i_\mathcal{K}) = 1$ for $i_\mathcal{K} \in outp_\mathcal{K}$ and 0 otherwise. Observe that $\boldsymbol{\pi}_\mathcal{K}(i_\mathcal{K}) > 0 \Leftrightarrow \mathbb{1}^{inp}_{n_\mathcal{K}}(i_\mathcal{K}) = 1$ and $\boldsymbol{d}_\mathcal{K}(i_\mathcal{K}) > 0 \Leftrightarrow \boldsymbol{\psi}_\mathcal{K}(i_\mathcal{K}) > 0 \Leftrightarrow \mathbb{1}^{outp}_{n_\mathcal{K}}(i_\mathcal{K}) = 1$. Furthermore the following results for the matrices $\boldsymbol{G}$ and $\boldsymbol{H}$ hold.

$\boldsymbol{H}_\mathcal{K}(i_\mathcal{K}, j_\mathcal{K}) > 0 \Rightarrow j_\mathcal{K} \in outp_\mathcal{K}$,
$\boldsymbol{H}_\mathcal{K}\mathbb{1}_{n_\mathcal{K}} = \boldsymbol{H}_\mathcal{K}\mathbb{1}^{outp}_{n_\mathcal{K}} = \mathbb{1}_{n_\mathcal{K}}, \boldsymbol{G}_\mathcal{K}\mathbb{1}_{n_\mathcal{K}} = \boldsymbol{G}_\mathcal{K}\mathbb{1}^{inp}_{n_\mathcal{K}} = \mathbb{1}^{outp}_{n_\mathcal{K}} \Rightarrow \boldsymbol{H}_\mathcal{K}\boldsymbol{G}_\mathcal{K}\mathbb{1}_{n_\mathcal{K}} = \mathbb{1}_{n_\mathcal{K}}$

Proof of Theorem 1

We have to show

$$\boldsymbol{\pi}\boldsymbol{V}_\mathcal{K}\left(\left(\bigotimes_{k\in\mathcal{K}}\boldsymbol{H}_k\right)\boldsymbol{G}_\mathcal{K}[\boldsymbol{\psi}]\right) = \boldsymbol{\pi}_\mathcal{K}.$$

First oberseve that

$$\boldsymbol{\pi}\boldsymbol{V}_\mathcal{K}\left(\bigotimes_{k\in\mathcal{K}}\boldsymbol{H}_k\right) = \boldsymbol{\pi}\left(\bigotimes_{k=1}^{K}\boldsymbol{H}_k\right)\boldsymbol{V}_\mathcal{K} = \boldsymbol{\psi}\boldsymbol{V}_\mathcal{K}$$

because $\boldsymbol{V}_\mathcal{K} = \otimes_{k=1}^{K}\boldsymbol{id}_k$ with $\boldsymbol{id}_k = \mathbb{1}_{n_k}$ for $k \notin \mathcal{K}$ and $\boldsymbol{id}_k = \boldsymbol{I}_{n_k}$ for $k \in \mathcal{K}$, $\boldsymbol{H}_k\mathbb{1}_{n_k} = \mathbb{1}_{n_k}$ and $\boldsymbol{H}_k\boldsymbol{I}_{n_k} = \boldsymbol{I}_{n_k}\boldsymbol{H}_k$. Thus, it remains to show

$$\boldsymbol{\psi}\boldsymbol{V}_\mathcal{K}\boldsymbol{G}_\mathcal{K} = \boldsymbol{\psi}\boldsymbol{V}_\mathcal{K}\operatorname{diag}(\boldsymbol{\psi}\boldsymbol{V}_\mathcal{K})^{+}\boldsymbol{V}_\mathcal{K}^{T}\operatorname{diag}(\boldsymbol{\psi})\boldsymbol{G}\boldsymbol{V}_\mathcal{K} = \boldsymbol{\pi}_\mathcal{K}.$$

The equation holds since $\boldsymbol{\psi}\boldsymbol{V}_\mathcal{K} = \boldsymbol{\psi}_\mathcal{K}$, $\boldsymbol{\psi}\operatorname{diag}(\boldsymbol{\psi})^{+} = \left(\mathbb{1}^{outp}_{n_\mathcal{K}}\right)^{T}$, $\left(\mathbb{1}^{outp}_{n_\mathcal{K}}\right)^{T}\boldsymbol{V}_\mathcal{K}^{T} = \boldsymbol{x}$ where $\left(\mathbb{1}^{outp}_{n_{1:K}}\right)^{T} \leq \boldsymbol{x} \leq \left(\mathbb{1}_{n_{1:K}}\right)^{T}$ such that

$$\boldsymbol{\psi} = \left(\mathbb{1}^{outp}_{n_{1:K}}\right)^{T}\operatorname{diag}(\boldsymbol{\psi}) \leq \boldsymbol{x}\operatorname{diag}(\boldsymbol{\psi}) \leq \left(\mathbb{1}_{n_{1:K}}\right)^{T}\operatorname{diag}(\boldsymbol{\psi}) = \boldsymbol{\psi}$$

which completes the proof.

Proof of Theorem 2

The mapping of vector $\boldsymbol{\pi}$ onto $\boldsymbol{\pi}_\mathcal{K}$ is realized by multiplication with matrix $\boldsymbol{V}_\mathcal{K} = \otimes_{k=1}^{K}\boldsymbol{id}_k$ where $\boldsymbol{id}_k = \mathbb{1}_{n_k}$ for $k \notin \mathcal{K}$ and $\boldsymbol{I}_{n_k}$ for $k \in \mathcal{K}$. Furthermore, the following relations hold.

$$\boldsymbol{H}_k\left(\sum_{r=1}^{R}\lambda^{(r)}\boldsymbol{G}_k^{(r)}\right)\mathbb{1}_{n_k} = \boldsymbol{H}_k\mathbb{1}^{outp}_{n_k} = \mathbb{1}_{n_k} \qquad \text{and}$$
$$\boldsymbol{H}_k\left(\sum_{r=1}^{R}\lambda^{(r)}\boldsymbol{G}_k^{(r)}\right)\boldsymbol{I}_{n_k} = \boldsymbol{I}_{n_k}\boldsymbol{H}_k\left(\sum_{r=1}^{R}\lambda^{(r)}\boldsymbol{G}_k^{(r)}\boldsymbol{I}_{n_k}\right)$$

It follows by basic properties of Kronecker products

$$\begin{aligned}
\boldsymbol{\pi}_k &= \boldsymbol{\pi} \left(\bigotimes_{l=1}^{K} \boldsymbol{H}_l \left(\sum_{r=1}^{R} \lambda^{(r)} \left(\bigotimes_{l=1}^{K} \boldsymbol{G}_l^{(r)} \right) \right) \right) \boldsymbol{V}_k \\
&= \boldsymbol{\pi} \left(\bigotimes_{l=1}^{K} \boldsymbol{H}_l \left(\sum_{r=1}^{R} \lambda^{(r)} \left(\bigotimes_{l=1}^{K} \boldsymbol{G}_l^{(r)} \right) \right) \boldsymbol{id}_k \right) \\
&= \boldsymbol{\pi} \left(\bigotimes_{k=1}^{K} \boldsymbol{id}_k \right) \left(\bigotimes_{l=1}^{K} \boldsymbol{H}_l \left(\sum_{r=1}^{R} \lambda^{(r)} \left(\bigotimes_{l=1}^{K} \boldsymbol{G}_l^{(r)} \right) \right) \right) \\
&= \boldsymbol{\pi}_k \left(\bigotimes_{l=1}^{K} \boldsymbol{H}_l \left(\sum_{r=1}^{R} \lambda^{(r)} \left(\bigotimes_{l=1}^{K} \boldsymbol{G}_l^{(r)} \right) \right) \right)
\end{aligned} \tag{28}$$

B Proofs of basic Equations

Proof of Equation (6)

Assume that the kth component is started with initial vector $\boldsymbol{\pi}_k$ and is described by sub-generator $\boldsymbol{D}_k$. Then

$$\begin{pmatrix} \boldsymbol{D}_k & \mathrm{diag}(\boldsymbol{d}_k) \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix}$$

describes the absorbing behavior of the process. The process will eventually end up in a state in the second block and it ends in the ith state of the second block, if state i of the first block is the exit from the first block (i.e., the state immediately before absorption). Thus, matrix $\boldsymbol{H}_k = (-\boldsymbol{D}_k)^{-1} \mathrm{diag}(\boldsymbol{d}_k)$ is a stochastic matrix that includes in position $\boldsymbol{H}_k(i,j)$ the conditional probability that the process, if started in state i, will exit from state j. This follows from standard results for absorbing continuous time Markov chains (CTMCs) [20].

Since all K absorbing processes are running independently and in parallel after being started, the matrix

$$\boldsymbol{H} = \bigotimes_{k=1}^{K} \boldsymbol{H}_k$$

describes the joint behavior, i.e., it maps initial state $i_{1:K}$ onto a joint state $j_{1:K}$ where j_k is the state in component k just before absorption. Observe that $\boldsymbol{H}$ is a stochastic matrix and $\boldsymbol{H}(i_{1:K}, j_{1:K}) > 0$ implies $j_{1:K} \in outp_{1:K}$. Then $\boldsymbol{HG}$ is a matrix that describes the embedded process which is observed immediately before a new start of the absorbing Markov chain. Consequently, $\boldsymbol{HG}(i_{1:K}, j_{1:K}) > 0$ implies $i_{1:K} \in inp_{1:K}$ and $j_{1:K} \in outp_{1:K}$ and $\sum_{j_{1:K} \in outp_{1:K}} \boldsymbol{HG}(i_{1:K}, j_{1:K}) =$

1 for $i_{1:K} \in inp_{1:K}$ and 0 if $i_{1:K} \notin inp_{1:K}$. If matrix $\boldsymbol{HG}$ contains a single irreducible subset of states, then its stationary vector describes the embedded stationary distribution immediately before the start of the absorbing Markov chains.

Proof of Equation (10)

By definition $\boldsymbol{\psi V}_k = \boldsymbol{\psi}_k$. Then we have

$$\begin{aligned}\boldsymbol{\pi}\left(\bigotimes_{l=1}^{K}\boldsymbol{H}_l\right)\boldsymbol{V}_k &= \boldsymbol{\pi}\left(\bigotimes_{l=1}^{K}\boldsymbol{H}_l\right)\left(\mathbb{1}_{n_{1:k-1}}\otimes\boldsymbol{I}_{n_k}\otimes\mathbb{1}_{n_{k+1:K}}\right)\\ &= \boldsymbol{\pi}\left(\mathbb{1}_{n_{1:k-1}}\times\boldsymbol{I}_{n_k}\otimes\mathbb{1}_{n_{k+1:K}}\right)\boldsymbol{H}_k \qquad = \boldsymbol{\pi}_k\boldsymbol{H}_k\end{aligned}$$

which follows because all matrices $\boldsymbol{H}_l$ are stochastic.

Proof of Equation (13)

For an absorbing CTMC with sub-generator $\boldsymbol{D}$ and initial distribution $\boldsymbol{\pi}$, the mth moment of absorption is given by $m!\boldsymbol{\pi}(-\boldsymbol{D})^{-m}\mathbb{1}$ (see [12, 20]). In our case we have two absorbing CTMCs and consider the product of the mth and n moments of absorption time. Let i_k, i_l be the initial state, then the conditional absorption times are given by $m!(\boldsymbol{M}_k)^m(i_k\bullet)\mathbb{1}_{n_k}$ and $n!(\boldsymbol{M}_l)^n(i_l\bullet)\mathbb{1}_{n_l}$. The initial probability of this state is given by $\boldsymbol{\pi}_{kl}(i_k, i_l)$[3]. Thus,

$$\boldsymbol{J}_0^{m,n}(k,l) = \sum_{i=0}^{n_k-1}\sum_{j=0}^{n_l-1}\boldsymbol{\pi}_{kl}(i,j)m!n!(\boldsymbol{M}_k)^m(i\bullet)\mathbb{1}_{n_k}(\boldsymbol{M}_l)^n(j\bullet)\mathbb{1}_{n_l}$$

which is just another representation of (9).

Proof of Equation (14)

The equation follows since $\boldsymbol{G}_k^{(r)}\mathbb{1}_{n_k} = \mathbb{1}_{n_k}^{outp}$ and $\boldsymbol{H}_k\mathbb{1}_{n_k}^{outp} = \mathbb{1}_{n_k}$.

Proof of Equation (16)

In the first part the distribution of exit states up to time x is considered for k whereas no restriction is put on component l. This implies that the integral

[3] For notational convenience we use the two dimensional numbering of states which can be mapped on $i_k n_l + i_l$ if $k < l$ or $i_l n_k + i_l$ if $k > l$.

defines the exit distribution of k and the exit distribution of l is given by multiplication with $\boldsymbol{H}_l$. Let $\boldsymbol{\pi}'_{kl}$ be the resulting vector. Then we have

$$\boldsymbol{\pi}'_{kl} \underbrace{\left(\sum_{r=1}^{R} \lambda^{(r)} \boldsymbol{G}_k^{(r)} \otimes \boldsymbol{G}_l^{(r)} \right)}_{\boldsymbol{G}_{kl}} \left(\boldsymbol{H}_k \otimes \int_0^y e^{\tau \boldsymbol{D}_l} \operatorname{diag}(\boldsymbol{d}_l)\, d\tau \right) \mathbb{1}_{n_k n_l} =$$

$$\boldsymbol{\pi}'_{kl} \underbrace{\left(\sum_{r=1}^{R} \lambda^{(r)} \mathbb{1}_{n_k}^{outp} \otimes \boldsymbol{G}_l^{(r)} \right)}_{\boldsymbol{G}_{kl}\left(\mathbb{1}_{n_l} \otimes \boldsymbol{I}_{n_l}\right)} \left(\int_0^y e^{\tau \boldsymbol{D}_l} \boldsymbol{d}_l d\tau \right)$$