Extension of some MAP results to transient MAPs and Markovian binary trees
SI: Matrix-Analytic Methodology

Sophie Hautphenne[1a], Miklós Telek[b,c]

[a] *Department of Mathematics and Statistics, University of Melbourne, Melbourne, Australia*
[b] *Department of Telecommunications, Technical University of Budapest, Budapest, Hungary*
[c] *MTA-BME Information systems research group, Budapest, Hungary*

Abstract

In this work we extend previous results on moment-based characterization and minimal representation of stationary Markovian arrival processes (MAPs) and rational arrival processes (RAPs) to transient Markovian arrival processes (TMAPs) and Markovian binary trees (MBTs).

We show that the number of moments that characterize a TMAP of size n with full rank marginal is $n^2 + 2n - 1$, and an MBT of size n with full rank marginal is $n^3 + 2n - 1$. We provide a non-Markovian representation for both processes based on these moments.

Next, we discuss the minimal representation of TMAPs and MBTs. In both cases, the minimal representation, which is not necessarily Markovian, can be found using different adaptations of the STAIRCASE algorithm presented in an earlier work by Buchholz and Telek (2011).

Finally, we heuristically investigate possible Markovian canonical representations for TMAPs and MBTs of order 2.

Keywords: Transient Markov arrival process, Markovian binary tree, moments of inter-arrival time distribution, minimal representation, canonical representation, parameter fitting

---

[1] Corresponding author. Department of Mathematics and Statistics, University of Melbourne, Melbourne, Australia. Email: sophiemh@unimelb.edu.au

1. Introduction

Nowadays, Markovian arrival processes (MAPs) are widely used modeling tools in several application fields from traffic analysis of telecommunication systems to computer performance modeling. In the majority of these applications the main stochastic property of interest is the stationary model behavior. Consequently, the considered MAPs generate an infinite number of arrival events as time goes to infinity, and the behavior at initial time does not affect the performance measures of interest.

In several other application fields, from population dynamics to risk analysis, the focus is on the transient behavior of the model instead of its stationary behavior. The extension of MAPs to point processes with a finite number of generated events is referred to as transient Markovian arrival processes (TMAPs) [1]. The underlying Markov process of these models has an absorbing state. Therefore, they generate a finite number of arrivals before absorption, and the behavior at initial time does affect the performance measures of interest. Basic properties of TMAPs, such as the distribution of the number of generated arrivals or the time until the last arrival, are presented in [1]. Apart from the above mentioned application fields, TMAPs also find applications in biological systems, as there can, for instance, be used to model the lifetime and reproduction epochs of individuals; see for instance [2] where TMAPs are applied to women's lifetime in several countries.

In this work we investigate additional properties of TMAPs, which on the one hand help characterizing these processes, and on the other hand play some role in TMAP fitting methods. An extension of TMAPs makes the methodology developed here applicable in branching process models too. Branching processes in which individuals' lifetime and reproduction times are controlled by TMAPs are called Markovian binary trees (MBTs) [3]. The basic properties of MBTs have been investigated in [3, 4, 5, 6, 2]. MBTs find applications for instance in phylogenetics [6] and demography [2]. Similarly to TMAPs, here we carry on the analysis of MBTs with characterizing properties and representation results.

More precisely, in this paper we extend previous results on moment-based characterization of stationary MAPs [7, 8], and minimal representation of Rational Arrival processes (RAPs) [9] to TMAPs and MBTs.

We show that the number of moments that characterize a TMAP of size n (that is, whose underlying Markov process has n transient states) with full rank marginal is $n^2 + 2n - 1$, and an MBT of size n with full rank marginal is

n^3+2n-1. These values therefore correspond to the number of independent parameters of the models for a given size n. We provide a non-Markovian representation for both processes based on these moments.

Next, we discuss the minimal representation of TMAPs and MBTs. In both cases, the minimal representation, which is not necessarily Markovian, can be found using different adaptations of the STAIRCASE algorithm presented in [9]. In some cases this non Markovian representation can be transformed into an equivalent Markovian representation by adapting the algorithm developed for stationary MAPs in [7].

Finally, based on the number of independent parameters characterizing TMAPs and MBTs, we heuristically investigate the Markovian canonical representations for TMAPs and MBTs of size 2 (denoted by TMAPs(2) and MBTs(2)). It turns out that different Markovian canonical forms exist for both processes, however a formal classification of TMAPs(2) and MBTs(2) according to their canonical representation(s) seems much more challenging to find than in the stationary MAP(2) case [10], and is out of the scope of this paper.

The results obtained are particularly useful for the parameter fitting of TMAPs and MBTs. Indeed, a first estimation method used for these kinds of models is the *moment method*, which consists in the evaluation of a set of experimental moments and joint moments of the observations, followed by the computation of a TMAP or an MBT representation based on these experimental moments. The moment method for stationary MAPs is given in [7]. The results in this paper indicate the set of moments and joint moments that are needed in order to fit the parameters of a TMAP or an MBT, and we provide a fitting procedure. A major drawback of the moment method is that the procedure always generates a matrix representation, but this representation might not be Markovian. If the obtained matrix representation can be transformed into a Markovian representation, using the mentioned transformation method developed in [7], then the obtained matrix representation surely describes a valid TMAP or MBT with the proper set of moments. If the matrix representation cannot be transformed into a Markovian representation, then nothing can be said about the validity of the obtained result.

Another popular estimation procedure, based on the maximum likelihood method, is the *Expectation-Maximization (EM) algorithm*, see for instance [11]. It has been used for the parameter estimation of MAPs [12, 13, 14, 15], TMAPs [14], and MBTs [16], and provides model estimates that are optimal in terms of asymptotic variance. One drawback is that the method is computationally expensive, especially when the number of parameters to estimate is large. It is therefore of high importance to know the number of independent parameters that are to be estimated for a fixed size n, and even better, to know a canonical representation of the model, so that the EM method could be applied for these representations only. These motivated our investigation of canonical forms for TMAPs and MBTs.

The rest of the paper is divided into two main parts. Section 2 is devoted to TMAPs and Section 3 to MBTs. The structure of these two sections is similar. The first subsections summarize the previously known properties. The second subsections provide the moments based characterizations of the processes and present the number of independent parameters based on them. The following two subsections give procedures for computing a matrix representation based on a characterizing set of moments, and provide adapted versions of the STAIRCASE method for computing a minimal matrix representation based on any redundant one. In the last subsections we investigate the possible Markovian canonical representations of TMAPs(2) and MBTs(2). We conclude the paper in the final section.

2. Transient Markov arrival processes

2.1. Definitions and basic properties

A transient Markovian arrival process (TMAP) is a point process whose arrivals are modulated by an underlying Markov process with state space $S = \{0, 1, \ldots, n\}$. State 0 is absorbing, and states $1, \ldots, n$, referred to as *phases*, are transient. When the background Markov process moves to state 0, we say that the process terminates, it does not generate further arrivals and remains in state 0 forever. TMAPs are characterized by matrices and vectors of size n associated only with the transient states: $\boldsymbol{\alpha}$ is the initial probability vector, and $\mathbf{D}$ is the transient generator of the underlying Markov process. If it is not defined explicitly, all row and column vectors are of size n and all matrices are of size $n \times n$ in this section. Consequently, $\mathbf{D}\mathbb{1} \leq 0$ element-wise and $\boldsymbol{\alpha}\mathbb{1} = 1$ where $\mathbb{1}$ is the column vector of ones. We denote by $\mathbf{d} = -\mathbf{D}\mathbb{1}$ the vector of transition rates to the absorbing phase. Furthermore we define the matrices $\mathbf{D_0}$ and $\mathbf{D_1}$ such that $\mathbf{D} = \mathbf{D_0} + \mathbf{D_1}$; the off-diagonal elements of $\mathbf{D_0}$ are the transition rates of the background Markov process without an arrival, and the elements of $\mathbf{D_1}$ define the transition rates associated to an arrival event. Similarly to the analysis of phase type distributions, the overall behavior of a TMAP is completely defined by these matrices and vectors.

The arrival intensity of the TMAP at time t is

$$\lambda(t) = \boldsymbol{\alpha} e^{\mathbf{D}t} \mathbf{D_1} \mathbb{1}, \tag{1}$$

and $\lim_{t \to \infty} \lambda(t) = 0$.

Throughout this paper we assume that the eigenvalues of $\mathbf{D_0}$ have negative real part. Consequently $\mathbf{D_0}$ is non-singular. We define the transient probability matrix of the embedded process as $\mathbf{P} = (-\mathbf{D_0})^{-1}\mathbf{D_1}$. It is the matrix of phase transitions at arrival instances. From $\mathbf{D}\mathbb{1} \leq 0$ we have $\mathbf{P}\mathbb{1} \leq \mathbb{1}$ element-wise.

Let $X_0, X_1, \ldots,$ denote the inter-arrival times of the process. If the process terminates after k arrivals then $X_k = X_{k+1} = \ldots = \infty$. Consequently, X_i $(i \geq 0)$ has a defective distribution. The probability density function of X_0 is

$$f_{X_0}(x) = \boldsymbol{\alpha} e^{\mathbf{D_0}x} \mathbf{D_1} \mathbb{1}. \tag{2}$$

Its Laplace transform is given by

$$f^*_{X_0}(s) = E[e^{-sX_0}] = \boldsymbol{\alpha}(s\mathbf{I} - \mathbf{D_0})^{-1}\mathbf{D_1}\mathbb{1}. \tag{3}$$

We define the defective moments of X_0 as follows

$$\begin{aligned} E(X_0^i I_{\{X_0<\infty\}}) &= \int_{x=0}^{\infty} x^i f_{X_0}(x) dx \\ &= i!\, \boldsymbol{\alpha}(-\mathbf{D_0})^{-i-1}\mathbf{D_1}\mathbb{1} \\ &= i!\, \boldsymbol{\alpha}(-\mathbf{D_0})^{-i}\mathbf{P}\mathbb{1}, \end{aligned} \tag{4}$$

where $I_{\{X_0<\infty\}}$ is the indicator of the event $\{X_0 < \infty\}$.

The inter arrival times in TMAPs are not independent. The joint density function of the inter-arrival times $X_0, X_1, \ldots, X_k$ is given by

$$f(x_0, x_1, \ldots, x_k) = \boldsymbol{\alpha} e^{\mathbf{D_0}x_0} \mathbf{D_1} e^{\mathbf{D_0}x_1} \mathbf{D_1} \ldots e^{\mathbf{D_0}x_k} \mathbf{D_1} \mathbb{1}. \tag{5}$$

Since X_i $(i \geq 0)$ has a defective distribution, we have

$$\mathrm{P}(X_0 < \infty, \ldots, X_k < \infty) = \int_{(x_0)} \ldots \int_{(x_k)} f(x_0, x_1, \ldots, x_k) dx_0 \ldots dx_k \leq 1. \tag{6}$$

From the joint density function, the defective joint moments of the $X_0, \dots, X_k$ inter-arrival times are

$$E(X_0^{i_0} X_1^{i_1} \dots X_k^{i_k} I_{\{X_0<\infty,\dots,X_k<\infty\}})$$
$$= \int_{(x_0)} \dots \int_{(x_k)} f(x_0, x_1, \dots, x_k) \prod_{n=0}^{k} x_n^{i_n} dx_0 \dots dx_k$$
$$= \boldsymbol{\alpha} i_0!(-\mathbf{D_0})^{-i_0} \mathbf{P} i_1!(-\mathbf{D_0})^{-i_1} \mathbf{P} \dots\ i_k!(-\mathbf{D_0})^{-i_k} \mathbf{P} \mathbb{1}. \tag{7}$$

The $(\boldsymbol{\alpha}, \mathbf{D_0}, \mathbf{D_1})$ representation of a TMAP is not unique. If $\mathbf{G}$ is a nonsingular matrix such that $\mathbf{G}\mathbb{1} = \mathbb{1}$ then $(\boldsymbol{\alpha}\mathbf{G}, \mathbf{G}^{-1}\mathbf{D_0}\mathbf{G}, \mathbf{G}^{-1}\mathbf{D_1}\mathbf{G})$ is an other representation of the same TMAP, since

$$f_{(\boldsymbol{\alpha}\mathbf{G}, \mathbf{G}^{-1}\mathbf{D_0}\mathbf{G}, \mathbf{G}^{-1}\mathbf{D_1}\mathbf{G})}(x_0, x_1, \dots, x_k)$$
$$= \boldsymbol{\alpha}\mathbf{G} e^{\mathbf{G}^{-1}\mathbf{D_0}\mathbf{G}x_0} \mathbf{G}^{-1}\mathbf{D_1}\mathbf{G} e^{\mathbf{G}^{-1}\mathbf{D_0}\mathbf{G}x_1} \mathbf{G}^{-1}\mathbf{D_1}\mathbf{G} \dots e^{\mathbf{G}^{-1}\mathbf{D_0}\mathbf{G}x_k} \mathbf{G}^{-1}\mathbf{D_1}\mathbf{G}\mathbb{1}$$
$$= \boldsymbol{\alpha} e^{\mathbf{D_0}x_0} \mathbf{D_1} e^{\mathbf{D_0}x_1} \mathbf{D_1} \dots e^{\mathbf{D_0}x_k} \mathbf{D_1} \mathbb{1}$$
$$= f_{(\boldsymbol{\alpha}, \mathbf{D_0}, \mathbf{D_1})}(x_0, x_1, \dots, x_k).$$

Among the identical representations of a TMAP some representations are Markovian and some are non-Markovian according to the following definition.

Definition 1. *The* $(\boldsymbol{\alpha}, \mathbf{D_0}, \mathbf{D_1})$ *representation of a TMAP is* Markovian *if* $\boldsymbol{\alpha} \geq \mathbf{0}$ *and* $\mathbf{D_1} \geq \mathbf{0}$ *element-wise, and the elements of* $\mathbf{D_0}$ *satisfy,* $\mathbf{D_0}_{ii} < 0$ *and* $\mathbf{D_0}_{ij} \geq 0$ *for* $i \neq j$*.*

The degree of the denominator of $f^*_{X_0}(s)$ (which is a rational function of s) is called the *order* of the first inter-arrival distribution. We need the following property of TMAPs.

Definition 2. *A TMAP has* full rank marginal *when the order of its first inter-arrival time distribution is identical with the size of* $\mathbf{D_0}$ *(more precisely, with the size of the minimal representation according to Section 2.4). Otherwise the TMAP has* reduced rank marginal*.*

This property was referred to as *non-redundant* in [7, 17] and related papers, but we found that terminology confusing with respect to the size of the representation. In this paper we restrict our attention to the case of full rank marginal processes. Based on the properties of stationary MAPs with reduced rank marginal [17] we presume that the case of reduced rank marginal TMAPs is more complex too.

Remark 1. For having a full rank marginal we have the following structural conditions: $\mathbf{D_0}$ has to be such that the eigenvalues of its Jordan blocks are all different with negative real part, $\boldsymbol{\alpha}$ should not be orthogonal to the right eigenvectors and $\mathbb{1}$ to the left ones.

On the other hand if a TMAP of size n has a full rank marginal then the set of row vectors $\boldsymbol{\alpha}, \boldsymbol{\alpha}\mathbf{D_0}^{-1}, \ldots, \boldsymbol{\alpha}\mathbf{D_0}^{-n+1}$ and the set of column vectors $\mathbb{1}, \mathbf{D_0}^{-1}\mathbb{1}, \ldots, \mathbf{D_0}^{-n+1}\mathbb{1}$ are linear independent (if it was not the case then the marginal distribution has a smaller representation according to [9] and the TMAP has reduced rank marginal).

Example 1. An example of a TMAP with reduced rank marginal is the following

$$\boldsymbol{\alpha} = (1,0), \quad \mathbf{D_0} = \begin{pmatrix} -1 & 0 \\ 0 & -2 \end{pmatrix}, \quad \mathbf{D_1} = \begin{pmatrix} 0.5 & 0.2 \\ 1 & 0.5 \end{pmatrix}, \quad \mathbf{d} = \begin{pmatrix} 0.3 \\ 0.5 \end{pmatrix},$$

where

$$f^*_{X_0}(s) = \boldsymbol{\alpha}(s\mathbf{I} - \mathbf{D_0})^{-1}\,\mathbf{D_1}\mathbb{1} = \frac{7}{10(s+1)}$$

indicates that the first inter-arrival distribution is order one.

As we restrict our attention to TMAPs with full rank marginal, the following set of reduced moments and joint moments completely characterize the TMAP, as shown in Section 2.2.

$$\mu_i = \frac{E(X_0^i I_{\{X_0<\infty\}})}{i!} = \boldsymbol{\alpha}\mathbf{E}^i\mathbf{P}\mathbb{1}, \quad i \geq 0, \tag{8}$$

and

$$\eta_{ij} = \frac{E(X_0^i X_1^j I_{\{X_0<\infty, X_1<\infty\}})}{i!j!} = \boldsymbol{\alpha}\mathbf{E}^i\mathbf{P}\mathbf{E}^j\mathbf{P}\mathbb{1}, \quad i,j \geq 0, \tag{9}$$

where $\mathbf{E} = (-\mathbf{D_0})^{-1}$.

The scalar quantities from (2) to (9) are computed such that matrix expressions are multiplied with a row vector from the left and with a column vector from the right. Hereafter, we refer to the row vector as the *initial vector*, and to the column vector as the *closing vector*. In the $(\boldsymbol{\alpha}, \mathbf{D_0}, \mathbf{D_0})$ representation, the initial and closing vectors are $\boldsymbol{\alpha}$ and $\mathbb{1}$ respectively.

2.2. Moments based characterization of TMAP

Note that $\mu_0 = \mathrm{P}(X_0 < \infty) \leq 1$. This way, the first $2n$ reduced moments $\mu_0, \mu_1, \ldots, \mu_{2n-1}$ of X_0 are independent parameters and uniquely define all other moments as well as the distribution of X_0 [8].

Theorem 1. *Having the reduced moments $\{\mu_i\}_{i\geq 0}$ of X_0, the number of independent reduced joint moments η_{ij} of (X_0, X_1) is at most $n^2 - 1$.*

Proof. Let $\mathcal{N} \equiv \{0, 1, \ldots, n-1\}$. We show that, based on the reduced moments $\{\mu_i\}_{i\geq 0}$ of X_0, and the reduced joint moments $\{\eta_{ij}\}_{i,j\in\mathcal{N}^2\setminus\{n-1,n-1\}}$ of (X_0, X_1), all other joint moments can be computed.

First we present the computation of $\eta_{n-1,n-1}$. Define

$$\begin{aligned}
\mathbf{M_1} &= \begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\alpha}\mathbf{E}^0\mathbf{P} \\ \vdots \\ \boldsymbol{\alpha}\mathbf{E}^{n-1}\mathbf{P} \end{bmatrix} \left[\ \mathbb{1} \,\middle|\, \mathbf{E}^0\mathbf{P}\mathbb{1} \,\middle|\, \ldots \,\middle|\, \mathbf{E}^{n-1}\mathbf{P}\mathbb{1}\ \right] \\
&= \begin{pmatrix} 1 & \mu_0 & \cdots & \mu_{n-1} \\ \mu_0 & \eta_{0,0} & \cdots & \eta_{0,n-1} \\ \vdots & \vdots & \ddots & \vdots \\ \mu_{n-1} & \eta_{n-1,0} & \cdots & \boxed{\eta_{n-1,n-1}} \end{pmatrix}.
\end{aligned}$$

Since the matrix $\mathbf{M_1}$ is the product of an $(n+1)\times n$ matrix and an $n\times(n+1)$ matrix, the determinant of $\mathbf{M_1}$ is 0. Consequently, $\eta_{n-1,n-1}$ can be computed from the other elements of $\mathbf{M_1}$. Note that the fact that $\boldsymbol{\alpha}\mathbb{1} = 1$ is utilized in the upper right element of $\mathbf{M_1}$.

The other joint moments can be obtained from lower moments as follows. For $i \geq n$ and j fixed, define the moment matrix $\mathbf{M_2}$ as follows

$$\begin{aligned}
\mathbf{M_2} &= \begin{bmatrix} \boldsymbol{\alpha}\mathbf{E}^0 \\ \vdots \\ \boldsymbol{\alpha}\mathbf{E}^{n-1} \\ \boldsymbol{\alpha}\mathbf{E}^{n} \end{bmatrix} \left[\ \mathbf{E}^0\mathbf{P}\mathbb{1} \,\middle|\, \ldots \,\middle|\, \mathbf{E}^{n-1}\mathbf{P}\mathbb{1} \,\middle|\, \mathbf{E}^{i-n}\mathbf{P}\mathbf{E}^{j}\mathbf{P}\mathbb{1}\ \right] \\
&= \begin{pmatrix} \mu_0 & \cdots & \mu_{n-1} & \eta_{i-n,j} \\ \vdots & \ddots & \vdots & \vdots \\ \mu_{n-1} & \cdots & \mu_{2n-1} & \eta_{i-1,j} \\ \mu_n & \cdots & \mu_{2n} & \boxed{\eta_{i,j}} \end{pmatrix}.
\end{aligned}$$

Similarly to $\mathbf{M_1}$, the determinant of $\mathbf{M_2}$ is 0, and therefore $\eta_{i,j}$ with $i \geq n$ is computable from the other elements of the matrix. A similar reasoning is applicable for $\eta_{i,j}$ with $j \geq n$ and i fixed. □

Theorem 2. *Based on the reduced moments* $\{\mu_i\}_{i\in\{0,1,\ldots,2n-1\}}$ *of* X_0 *and the reduced joint moments* $\{\eta_{ij}\}_{i,j\in\mathcal{N}^2\setminus\{n-1,n-1\}}$ *of* (X_0, X_1)*, all other joint moments can be computed.*

Proof. The same recursive approach as the one in [8] is applicable here. □

Theorem 3. *The number of moments that characterize a TMAP of order* n *with full rank marginal is* $n^2 + 2n - 1$*.*

Proof. It is a consequence of the preceding two theorems. □

Remark 2. An intuitive explanation of the number of independent parameters in a TMAP is the following. The number of elements of the matrices $\mathbf{D_0}$ and $\mathbf{D_1}$ is $2n^2$. The initial probability distribution $\boldsymbol{\alpha}$ satisfies $\boldsymbol{\alpha}\mathbb{1} = 1$, from which $n-1$ elements define the initial vector. So, in total the ($\boldsymbol{\alpha}$, $\mathbf{D_0}$, $\mathbf{D_1}$) representation of a TMAP counts $2n^2 + n - 1$ elements. However, this representation of the TMAP is not unique. A similarity transformation matrix $\mathbf{G}$ contains n^2 elements and fulfills n linear constraints $\mathbf{G}\mathbb{1} = \mathbb{1}$. Consequently, the similarity transformation has n^2-n free parameters, which need to be subtracted from the total number of parameters in $(\boldsymbol{\alpha}, \mathbf{D_0}, \mathbf{D_1})$. This gives a total $(2n^2+n-1)-(n^2-n) = n^2+2n-1$ independent parameters in the TMAP.

An alternative interpretation of the number of independent parameters of a TMAP can be given based on the number of independent parameters in a stationary MAP, which is equal to n^2 [7]. Indeed in a stationary MAP, the initial probability vector $\boldsymbol{\alpha}$ is entirely determined by $\mathbf{D_0}$ and $\mathbf{D_1}$; in a TMAP this is no more the case, and $\boldsymbol{\alpha}$ contains $n-1$ parameters. A TMAP also contains n additional parameters in the vector $\mathbf{d} = -(\mathbf{D_0}+\mathbf{D_1})\,\mathbb{1}$. Therefore the number of independent parameters of a TMAP is $n^2 + (n-1) + n$.

Example 2. The TMAP with representation

$$\boldsymbol{\alpha} = (0.6, 0.4), \quad \mathbf{D_0} = \begin{pmatrix} -4 & 0 \\ 1 & -6 \end{pmatrix}, \quad \mathbf{D_1} = \begin{pmatrix} 1 & 2 \\ 3 & 0 \end{pmatrix}, \quad \mathbf{d} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}.$$

is entirely characterized by the following reduced moments of X_0 and joint moments of (X_0, X_1):

$$\mu_0 = 0.700, \quad \mu_1 = 0.1667, \quad \mu_2 = 0.0201, \quad \mu_3 = 0.0016,$$

$$\eta_{00} = 0.4833, \quad \eta_{01} = 0.1139, \quad \eta_{10} = 0.1139.$$

2.3. Finding a TMAP representation based on the moments

Based on the reduced moments $\{\mu_i\}_{i\in\{0,1,\dots,2n-1\}}$ of X_0, using the method of Appie van de Liefvoort [18] and its implementation from [7], we can obtain a vector-matrix pair $(\boldsymbol{\phi}, \mathbf{K})$ such that

$$\mu_i = \mu_0 \boldsymbol{\phi} \mathbf{K}^i \mathbb{1}, \quad i = 0, 1, \dots, 2n-1.$$

In this representation, $\boldsymbol{\phi} = (1/n, \dots, 1/n)$. From

$$\mu_i = \boldsymbol{\alpha} \mathbf{E}^i \mathbf{P} \mathbb{1} = \boldsymbol{\phi} \mathbf{K}^i \mu_0 \mathbb{1}$$

and [7, Theorem 1] there exists a non singular matrix $\mathbf{H}$ such that $\boldsymbol{\phi} = \boldsymbol{\alpha}\mathbf{H}$, $\mathbf{K} = \mathbf{H}^{-1}\mathbf{E}\mathbf{H}$ and $\mu_0\mathbb{1} = \mathbf{H}^{-1}\mathbf{P}\mathbb{1}$. Using this, from (9) we have

$$\begin{aligned} \eta_{ij} &= \underbrace{\boldsymbol{\alpha}\mathbf{H}}_{\boldsymbol{\phi}} \underbrace{\mathbf{H}^{-1}\mathbf{E}^i\mathbf{H}}_{\mathbf{K}^i} \underbrace{\mathbf{H}^{-1}\mathbf{P}\mathbf{H}}_{\hat{\mathbf{P}}} \underbrace{\mathbf{H}^{-1}\mathbf{E}^j\mathbf{H}}_{\mathbf{K}^j} \underbrace{\mathbf{H}^{-1}\mathbf{P}\mathbb{1}}_{\mu_0\mathbb{1}}, \\ &= \boldsymbol{\phi} \quad \mathbf{K}^i \quad \hat{\mathbf{P}} \quad \mathbf{K}^j \quad \underbrace{\mu_0\mathbb{1}}_{\hat{\mathbf{P}}\boldsymbol{v}}, \end{aligned} \tag{10}$$

where $\hat{\mathbf{P}} = \mathbf{H}^{-1}\mathbf{P}\mathbf{H}$. This computation of the joint moments already indicates that the previously used representation of TMAP with $(\boldsymbol{\alpha}, \mathbf{D_0}, \mathbf{D_1})$ needs to be extended occasionally to a different representation which explicitly defines the closing vector on the right, because the notation $(\boldsymbol{\alpha}, \mathbf{D_0}, \mathbf{D_1})$ implicitly assumes that the closing vector is $\mathbb{1}$. The extended notation of the TMAP $(\boldsymbol{\alpha}, \mathbf{D_0}, \mathbf{D_1})$ which includes the closing vector is $[\boldsymbol{\alpha}, \mathbf{E}, \mathbf{P}, \mathbb{1}]$, where recall that $\mathbf{E} = (-\mathbf{D_0})^{-1}$ and $\mathbf{P} = \mathbf{E}\,\mathbf{D_1}$. In this extended representation we replaced $\mathbf{D_0}$ and $\mathbf{D_1}$ by $\mathbf{E}$ and $\mathbf{P}$ for notational convenience. The extended representation of the TMAP in (10) is thus $[\boldsymbol{\phi}, \mathbf{K}, \hat{\mathbf{P}}, \boldsymbol{v}]$ where the closing vector $\boldsymbol{v}$ is such that $\hat{\mathbf{P}}\boldsymbol{v} = \mu_0\mathbb{1}$.

Matrices $\boldsymbol{\Phi}$, $\boldsymbol{\Lambda}$ and $\boldsymbol{\Gamma}$ of size $n \times n$ are defined as follows. The ith row of $\boldsymbol{\Phi}$ is $\boldsymbol{\phi}\mathbf{K}^i$, the ith column of $\boldsymbol{\Lambda}$ is $\mathbf{K}^i\mu_0\mathbb{1}$ and the i, j element of $\boldsymbol{\Gamma}$ is η_{ij}, where rows and columns are numbered from 0 to $n-1$. With these notations (10) can be written in matrix form as

$$\boldsymbol{\Gamma} = \boldsymbol{\Phi}\hat{\mathbf{P}}\boldsymbol{\Lambda}, \tag{11}$$

from which

$$\hat{\mathbf{P}} = \mathbf{\Phi}^{-1}\mathbf{\Gamma}\mathbf{\Lambda}^{-1}, \tag{12}$$

where $\mathbf{\Phi}$ and $\mathbf{\Lambda}$ are nonsingular according to Remark 1. Taking any vector $\boldsymbol{v}$ which satisfies $\hat{\mathbf{P}}\boldsymbol{v} = \mu_0 \mathbb{1}$, the associated reduced moments and joint moments satisfy

$$\mu_i = \boldsymbol{\phi}\mathbf{K}^i\hat{\mathbf{P}}\boldsymbol{v}, \quad i = 0, 1, \ldots, 2n-1$$

and

$$\eta_{ij} = \boldsymbol{\phi}\mathbf{K}^i\hat{\mathbf{P}}\mathbf{K}^j\hat{\mathbf{P}}\boldsymbol{v}, \quad i, j = 0, 1, \ldots, n-1.$$

That is, $[\boldsymbol{\phi}, \mathbf{K}, \hat{\mathbf{P}}, \boldsymbol{v}]$ is an extended representation of the TMAP whose reduced moments and joint moments are μ_i and η_{ij}. The only remaining step is to transform the representation into another one with closing vector $\mathbb{1}$.

Let $\mathbf{C}$ be a non-singular matrix such that $\mathbf{C}\mathbb{1} = \boldsymbol{v}$. For example $C = \text{diag}\langle v \rangle$ and in case of zero diagonal element add 1 to the diagonal and subtract 1 from the first element of the row. Then

$$[\boldsymbol{\phi}\mathbf{C}, \mathbf{C}^{-1}\mathbf{K}\mathbf{C}, \mathbf{C}^{-1}\hat{\mathbf{P}}\mathbf{C}, \mathbf{C}^{-1}\boldsymbol{v}]$$

with closing vector $\mathbf{C}^{-1}\boldsymbol{v} = \mathbb{1}$, is another equivalent extended representation of the same TMAP. The initial vector of this representation satisfies $\boldsymbol{\phi}\mathbf{C}\mathbb{1} = \boldsymbol{\phi}\boldsymbol{v} = 1$. From this representation, the standard (3 elements) representation of the TMAP is $(\boldsymbol{\phi}\mathbf{C}, -(\mathbf{C}^{-1}\mathbf{K}\mathbf{C})^{-1}, -\mathbf{C}^{-1}\mathbf{K}^{-1}\hat{\mathbf{P}}\mathbf{C})$.

Remark 3. The $(\boldsymbol{\phi}\mathbf{C}, -(\mathbf{C}^{-1}\mathbf{K}\mathbf{C})^{-1}, -\mathbf{C}^{-1}\mathbf{K}^{-1}\hat{\mathbf{P}}\mathbf{C})$ representation is generally non-Markovian due to the fact that the method of [18] results in a non-Markovian representation of the distribution of X_0. However, in some cases the algorithm developed in [7] can be adapted to the TMAP case in order to transform any non Markovian representation into an equivalent Markovian representation. This procedure is implemented in Mathematica and Matlab and publicly available as part of the special processes component of the BuTools program package [19].

Example 3. A non-Markovian representation based on the reduced moments and joint moments in Example 2 is given by

$$\boldsymbol{\alpha} = (0.51, 0.49), \quad \mathbf{D_0} = \begin{pmatrix} -7.8980 & 3.3884 \\ -2.1857 & -2.1000 \end{pmatrix},$$
$$\mathbf{D_1} = \begin{pmatrix} 0.5500 & 2.4499 \\ 2.5502 & 0.4497 \end{pmatrix}, \quad \mathbf{d} = \begin{pmatrix} 1.5096 \\ 1.2857 \end{pmatrix},$$

which can be transformed into the following equivalent Markovian representation

$$\boldsymbol{\alpha} = (0.5331, 0.4669), \quad \mathbf{D_0} = \begin{pmatrix} -5.7918 & 0.8345 \\ 0.4429 & -4.2061 \end{pmatrix},$$
$$\mathbf{D_1} = \begin{pmatrix} 0.6655 & 2.3344 \\ 2.6658 & 0.3343 \end{pmatrix}, \quad \mathbf{d} = \begin{pmatrix} 1.9575 \\ 0.7632 \end{pmatrix}.$$

This Markovian representation is therefore an equivalent representation to the one given in Example 2. The presented representations are computed by the TRAPFromMoments and the TRAPToTMAP functions of the BuTools package, where TRAP refers to transient rational arrival process, which is the analogue of TMAP with non-Markovian representation.

2.4. Minimal representation of Transient MAPs

It may happen that a TMAP with representation $(\boldsymbol{\alpha}, \mathbf{D_0}, \mathbf{D_1})$ of size m has an equivalent represention of smaller size $n < m$. In this section we investigate the representation of a TMAP with minimal size.

Definition 3. *The* $(\boldsymbol{\alpha}, \mathbf{D_0}, \mathbf{D_1})$ *representation of a TMAP is* minimal *if the same TMAP does not have any representation with smaller size. In this case the size of the representation is referred to as the* rank *of the TMAP.*

Section 2.2 provides a moments based characterization of TMAPs with full rank marginal. Based on that characterization, the rank of a TMAP with full rank marginal can be obtained as the size of appropriate moment and joint moment matrices with non-zero determinant (see for instance matrices $\mathbf{M_1}$ and $\mathbf{M_2}$ in the proof of Theorem 1). Here we present general results which are applicable for both, TMAPs with full and with reduced rank marginal.

We adopt the same approach as the one in [9]. If the size of the representation $(\boldsymbol{\alpha}, \mathbf{D_0}, \mathbf{D_1})$ is n, then the generalized controllability matrices are

$$\mathcal{C}_{\mathbf{D_0}} = \mathcal{C}_{\mathbf{D_0}}(0) = (\mathbf{P}\mathbb{1}, \mathbf{E}\mathbf{P}\mathbb{1}, \ldots, \mathbf{E}^{n-1}\mathbf{P}\mathbb{1}),$$
$$\mathcal{C}_{\mathbf{D_0},\mathbf{D_1}} = \mathcal{C}_{\mathbf{D_0},\mathbf{D_1}}(0) = (\mathcal{C}_{\mathbf{D_0}}, \mathbf{P}\mathcal{C}_{\mathbf{D_0}}, \mathbf{P}^2\mathcal{C}_{\mathbf{D_0}}, \ldots, \mathbf{P}^{n-1}\mathcal{C}_{\mathbf{D_0}}),$$
$$\mathcal{C}_{\mathbf{D_0}}(\ell{+}1) = (\mathcal{C}_{\mathbf{D_0},\mathbf{D_1}}(\ell), \mathbf{E}\mathcal{C}_{\mathbf{D_0},\mathbf{D_1}}(\ell), \ldots, \mathbf{E}^{n-1}\mathcal{C}_{\mathbf{D_0},\mathbf{D_1}}(\ell)), \ \ell \geq 0$$
$$\mathcal{C}_{\mathbf{D_0},\mathbf{D_1}}(\ell{+}1) = (\mathcal{C}_{\mathbf{D_0}}(\ell{+}1), \mathbf{P}\mathcal{C}_{\mathbf{D_0}}(\ell{+}1), \mathbf{P}^2\mathcal{C}_{\mathbf{D_0}}(\ell{+}1), \ldots, \mathbf{P}^{n-1}\mathcal{C}_{\mathbf{D_0}}(\ell{+}1)), \ \ell \geq 0.$$

$\mathcal{C}_{\mathbf{D_0}}(\ell)$ is an $n \times n^{2\ell+1}$ matrix and $\mathcal{C}_{\mathbf{D_0},\mathbf{D_1}}(\ell)$ is an $n \times n^{2\ell+2}$ matrix. Based on these matrices, the controllability rank is $r_C = rank(\mathcal{C}_{\mathbf{D_0},\mathbf{D_1}}(n))$.

The generalized observability matrices are

$$\mathcal{O}_{\mathbf{D_0}} = \mathcal{O}_{\mathbf{D_0}}(0) = \begin{pmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\alpha}\mathbf{E} \\ \vdots \\ \boldsymbol{\alpha}\mathbf{E}^{n-1} \end{pmatrix}, \quad \mathcal{O}_{\mathbf{D_0},\mathbf{D_1}} = \mathcal{O}_{\mathbf{D_0},\mathbf{D_1}}(0) = \begin{pmatrix} \mathcal{O}_{\mathbf{D_0}} \\ \mathcal{O}_{\mathbf{D_0}}\mathbf{P} \\ \vdots \\ \mathcal{O}_{\mathbf{D_0}}\mathbf{P}^{n-1} \end{pmatrix},$$

$$\mathcal{O}_{\mathbf{D_0}}(\ell+1) = \begin{pmatrix} \mathcal{O}_{\mathbf{D_0},\mathbf{D_1}}(\ell) \\ \mathcal{O}_{\mathbf{D_0},\mathbf{D_1}}(\ell)\mathbf{E} \\ \vdots \\ \mathcal{O}_{\mathbf{D_0},\mathbf{D_1}}(\ell)\mathbf{E}^{n-1} \end{pmatrix}, \quad \mathcal{O}_{\mathbf{D_0},\mathbf{D_1}}(\ell+1) = \begin{pmatrix} \mathcal{O}_{\mathbf{D_0}}(\ell+1) \\ \mathcal{O}_{\mathbf{D_0}}(\ell+1)\mathbf{P} \\ \vdots \\ \mathcal{O}_{\mathbf{D_0}}(\ell+1)\mathbf{P}^{n-1} \end{pmatrix}, \quad \ell \geq 0.$$

The size of $\mathcal{O}_{\mathbf{D_0}}(\ell)$ is $n^{2\ell+1} \times n$ and the size of $\mathcal{O}_{\mathbf{D_0},\mathbf{D_1}}(\ell)$ is $n^{2\ell+2} \times n$. The observability rank is $r_O = rank(\mathcal{O}_{\mathbf{D_0},\mathbf{D_1}}(n))$.

Theorem 4. *The $(\boldsymbol{\alpha}, \mathbf{D_0}, \mathbf{D_1})$ representation of size n is minimal if and only if $r_O = r_C = n$.*

Proof. The same approach is applicable as in [9], here we only provide an intuitive explanation. Having $r_C = n$ means that the repeated multiplication of the closing vector with $\mathbf{D_0}$ and $\mathbf{D_1}$ (or with $\mathbf{E}$ and $\mathbf{P}$) spans an n-dimensional space. Similarly, $r_O = n$ means that the repeated multiplication of the initial vector with $\mathbf{D_0}$ and $\mathbf{D_1}$ (or with $\mathbf{E}$ and $\mathbf{P}$) spans an n-dimensional space. Consequently, if the size of the representation is n then $r_C = n$ means that $\mathcal{C}_{\mathbf{D_0},\mathbf{D_1}}(n)$ has full rank and $r_O = n$ means that $\mathcal{O}_{\mathbf{D_0},\mathbf{D_1}}(n)$ has full rank.

If, for example $r_C = n$ but $r_O < n$ then the representation is at least size n (due to $r_C = n$), but the rank of the observability matrix is at most $r_O < n$. The moment matrices (see for instance matrices $\mathbf{M_1}$ and $\mathbf{M_2}$), are the product of an observability matrix and a controllability matrix, consequently their rank is at most $r_O < n$. Finally, if the rank of the moment matrices composed as the product of an observability matrix and a controllability matrix is at most $r_O < n$, then there is a size r_O representation of the process. Indeed, in case of TMAP with full rank marginal the procedure in Section 2.3 can be used to construct a representation of size r_O. □

Due to the close similarities between stationary MAPs and TMAPs, the same numerical procedure as the one introduced in [9], and referred to as the

STAIRCASE method, is applicable for the computation of the minimal representation of TMAPs. The only differences are in the definition of the initial and closing vectors. The definition of TMAPs contains the initial vector, while in case of stationary MAPs, the initial vector is the stationary vector of the stochastic matrix $\mathbf{P}$. The closing vector a stationary MAP is $\mathbb{1}$ while the closing vector of a TMAP here is $\mathbf{P}\mathbb{1}$. Consequently the STAIRCASE method called with $(\mathbf{D_0}, \mathbf{D_1}, \mathbf{P}\mathbb{1})$ provides a representation with minimal controllability rank, and the same method called with $(\mathbf{D_0}', \mathbf{D_1}', \boldsymbol{\alpha}')$, where $'$ denotes the transpose, provides a representation with minimal observability rank.

If we have an arbitrary representation $(\boldsymbol{\alpha}, \mathbf{D_0}, \mathbf{D_1})$ of a TMAP, we can obtain a minimal representation as follows: first, we reduce the size of the TMAP by eliminating the redundancy due to the closing vector using the STAIRCASE method with $(\mathbf{D_0}, \mathbf{D_1}, \mathbf{P}\mathbb{1})$, and second, we reduce the size by eliminating the redundancy due to the initial vector using the STAIRCASE method with $(\mathbf{D_0^*}', \mathbf{D_1^*}', \boldsymbol{\alpha}^{*\prime})$, where $(\boldsymbol{\alpha}^*, \mathbf{D_0^*}, \mathbf{D_1^*})$ is the representation obtained after the first step.

Example 4. A TMAP which is non-minimal due to the closing vector is the following:

$$\boldsymbol{\alpha} = (0.1, 0.2, 0.7), \quad \mathbf{D_0} = \begin{pmatrix} -20 & 8 & 1 \\ 4 & -16 & 1 \\ 1 & 2 & -11 \end{pmatrix},$$

$$\mathbf{D_1} = \begin{pmatrix} 2 & 5 & 3 \\ 3 & 4 & 3 \\ 3 & 2 & 1 \end{pmatrix}, \quad \mathbf{d} = \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix}.$$

This TMAP is lumpable into the equivalent representation of size 2

$$\boldsymbol{\alpha} = (0.3, 0.7), \quad \mathbf{D_0} = \begin{pmatrix} -12 & 1 \\ 3 & -11 \end{pmatrix}, \quad \mathbf{D_1} = \begin{pmatrix} 7 & 3 \\ 5 & 1 \end{pmatrix}, \quad \mathbf{d} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}.$$

The STAIRCASE method of the BuTools package [19], namely the TRAPMinimize function, finds the following equivalent minimal representation

$$\boldsymbol{\alpha} = (1.77782, -0.777817), \quad \mathbf{D_0} = \begin{pmatrix} -9.29289 & -0.707107 \\ 2.29289 & -13.7071 \end{pmatrix},$$

$$\mathbf{D_1} = \begin{pmatrix} 7.48816 & 1.17851 \\ 10.0404 & 0.51184 \end{pmatrix}, \quad \mathbf{d} = \begin{pmatrix} 1.33333 \\ 0.861929 \end{pmatrix}.$$

Example 5. A TMAP which is non-minimal due to the initial vector is the following:

$$\boldsymbol{\alpha} = (0.25, 0.25, 0.5), \quad \mathbf{D_0} = \begin{pmatrix} -20 & 4 & 3 \\ 4 & -20 & 1 \\ 1 & 1 & -11 \end{pmatrix},$$

$$\mathbf{D_1} = \begin{pmatrix} 2 & 3 & 1 \\ 4 & 3 & 5 \\ 1 & 1 & 2 \end{pmatrix}, \quad \mathbf{d} = \begin{pmatrix} 7 \\ 3 \\ 5 \end{pmatrix}.$$

This TMAP is weakly-lumpable into the equivalent representation of size 2

$$\boldsymbol{\alpha} = (0.5, 0.5), \quad \mathbf{D_0} = \begin{pmatrix} -16 & 2 \\ 2 & -11 \end{pmatrix}, \quad \mathbf{D_1} = \begin{pmatrix} 6 & 3 \\ 2 & 2 \end{pmatrix}, \quad \mathbf{d} = \begin{pmatrix} 5 \\ 5 \end{pmatrix}.$$

The TRAPMinimize function of the BuTools package [19] finds the following minimal representation

$$\boldsymbol{\alpha} = (1, 0), \quad \mathbf{D_0} = \begin{pmatrix} -10.666667 & -0.83333333 \\ -2.666667 & -16.333333 \end{pmatrix},$$

$$\mathbf{D_1} = \begin{pmatrix} 6 & 0.5 \\ 12 & 2 \end{pmatrix}, \quad \mathbf{d} = \begin{pmatrix} 5 \\ 5 \end{pmatrix}.$$

2.5. Canonical representation of transient MAPs of order 2

A canonical representation of a TMAP of size n is a unique representation which is conveniently chosen from the infinitely many equivalent representations, and the method for choosing the unique representation is applicable for the whole TMAP class of size n. In this paper we assume that the convenient canonical representation is Markovian and contains minimal number of nonzero elements. This canonical representation is a representation $(\boldsymbol{\alpha}, \mathbf{D_0}, \mathbf{D_1})$ of size n where the total number of nonzero entries in the vector $\boldsymbol{\alpha}$ and the matrices $\mathbf{D_0}$ and $\mathbf{D_1}$ corresponds exactly to the maximal number of independent parameters $n^2 + 2n - 1$. It should be noted that the uniqueness of the canonical representation is not ensured by this assumption yet, but the main focus of our investigation is the completeness of the considered structures. Especially, the purpose of this section is to investigate the possible Markovian canonical representations of TMAPs of order 2, denoted by

TMAPs(2), and to find the minimal number of structures which covers the whole set TMAPs(2).

In [10] the authors show that there are two possible Markovian canonical forms for the stationary MAPs(2), and they give a classification of these processes in terms of the correlation parameter γ, which is the second eigenvalue of $\mathbf{P} = (-\mathbf{D_0})^{-1}\mathbf{D_1}$. We specify below the two zero entries in each of the possible canonical representations, corresponding to $\gamma < 0$ or $\gamma \geq 0$.

Structure No.	Zero elements	Use in [10]
1	$\mathbf{D_0}_{12}$, $\mathbf{D_1}_{22}$	$\gamma < 0$
2	$\mathbf{D_0}_{12}$, $\mathbf{D_1}_{21}$	$\gamma \geq 0$

Before investigating the possible canonical form of a TMAP(2), let us examine those of a non-stationary MAPs(2). First note that, since the number of independent parameters of a stationary MAP of size n is n^2 [7], the number of independent parameters of a non-stationary MAP is n^2+n-1, where the additional $n-1$ elements come from the initial vector $\boldsymbol{\alpha}$ (which is not determined by $\mathbf{D_0}$ and $\mathbf{D_1}$ anymore). There are therefore five independent parameters in a non-stationary MAP(2).

A non-stationary MAP(2) with representation $(\boldsymbol{\alpha}, \mathbf{D_0}, \mathbf{D_1})$ is determined by the $2+4+4 = 10$ elements of the representation. Out of these 10 elements 7 are independent because $\boldsymbol{\alpha}\mathbb{1} = 1$ and $\mathbf{D_0}\mathbb{1} + \mathbf{D_1}\mathbb{1} = \mathbb{1}$. Due to the fact that the total number of parameters of a non-stationary MAP(2) is seven and the number of independent parameters is five, we can set two elements of the representation to zero. A canonical representation should therefore have five nonzero entries and two zero entries. By considering symmetric representations (to a permutation of the two phases) only once, there is thus a total of 15 possible structures with two zero entries.

In order to find the Markovian structures with two zero entries which describe the whole non-stationary MAPs(2) set, randomly generated a large number of random non-stationary MAPs(2), and we tested all their possible Markovian canonical representations. Then, we determined the smallest subsets of canonical forms covering all random examples.

We intended to randomly generate the non-stationary MAPs(2) in such a way that we cover the whole set of non-stationary MAPs(2) as well as possible, by proceeding as follows. We generated a first set of 1000 random non-stationary MAPs(2), whose seven independent entries are *i.i.d.* according to a discrete uniform random variable $U[0, 100]$. Then, we generated a second set of 1000 random non-stationary MAPs(2) of which each of the seven entries is either 0 with probability $p = 0.3$, or is distributed according to a discrete uniform random variable $U[0, 100]$ with probability $1 - p$. Finally, we generated a last set of 1500 random non-stationary MAPs(2) whose representation corresponds to one of the 15 possible canonical forms, and whose nonzero entries are generated according to the same procedure as for the second set, but with $p = 0.2$.

We found that at least five structures are required to cover all randomly generated MAPs(2). A set of five structures which is complete according to our numerical investigations is the following.

Structure No.	Zero elements
1	$\mathbf{D_0}_{12}, \mathbf{D_1}_{21}$
2	$\boldsymbol{\alpha}_1, \mathbf{D_0}_{12}$
3	$\boldsymbol{\alpha}_1, \mathbf{D_1}_{12}$
4	$\mathbf{D_0}_{12}, \mathbf{D_1}_{11}$
5	$\boldsymbol{\alpha}_1, \mathbf{D_1}_{11}$

Apart from this set of five structures (denoted as Set 1), three other sets of five structures covered all the generated MAPs(2). In the table, the symbol "-" means "the same as in Set 1".

Structure No.	Set 1	Set 2	Set 3	Set 4
1	$\mathbf{D_0}_{12}, \mathbf{D_1}_{21}$	-	-	-
2	$\boldsymbol{\alpha}_1, \mathbf{D_0}_{12}$	-	-	-
3	$\boldsymbol{\alpha}_1, \mathbf{D_1}_{12}$	-	-	-
4	$\mathbf{D_0}_{12}, \mathbf{D_1}_{11}$	$\mathbf{D_0}_{12}, \mathbf{D_1}_{11}$	$\mathbf{D_0}_{12}, \mathbf{D_1}_{22}$	$\mathbf{D_0}_{12}, \mathbf{D_1}_{22}$
5	$\boldsymbol{\alpha}_1, \mathbf{D_1}_{11}$	$\boldsymbol{\alpha}_1, \mathbf{D_1}_{22}$	$\boldsymbol{\alpha}_1, \mathbf{D_1}_{11}$	$\boldsymbol{\alpha}_1, \mathbf{D_1}_{22}$

Note that moving from stationary to non-stationary MAPs(2) already increases the number of possible Markovian canonical forms from two to five to cover the whole class.

We used exactly the same method for characterizing the Markovian canonical forms of TMAPs(2). A TMAP(2) with representation $(\boldsymbol{\alpha}, \mathbf{D_0}, \mathbf{D_1})$ is also determined by the $2+4+4 = 10$ elements of the representation, but in case of TMAP(2) the number of independent elements is 9 because only the $\boldsymbol{\alpha}\mathbb{1} = 1$ constraint applies for TMAPs. This way we have 9 independent elements in the representation of a TMAP(2), while by Theorem 3 the number of independent parameters is seven, consequently we set two elements of the representation to zero. There is a total of 23 possible structures with two zeros in the representation. We proceeded as in the non-stationary MAP case, and we found that there are three possible sets of five Markovian canonical forms.

Structure No.	Set 1	Set 2	Set 3
1	$\mathbf{D_0}_{12}, \mathbf{D_1}_{21}$	-	-
2	$\boldsymbol{\alpha}_1, \mathbf{D_1}_{22}$	-	-
3	$\mathbf{d}_1, \mathbf{D_0}_{21}$	$\mathbf{d}_1, \mathbf{D_0}_{21}$	$\mathbf{d}_1, \mathbf{D_1}_{21}$
4	$\mathbf{D_0}_{12}, \mathbf{D_1}_{22}$	$\mathbf{D_0}_{12}, \mathbf{D_1}_{22}$	$\mathbf{D_0}_{12}, \mathbf{D_1}_{11}$
5	$\boldsymbol{\alpha}_1, \mathbf{D_1}_{12}$	$\mathbf{d}_1, \mathbf{D_1}_{21}$	$\boldsymbol{\alpha}_1, \mathbf{D_0}_{12}$

To ensure the uniqueness of the canonical form, a classification of non-stationary MAP(2) and TMAPs(2) into one of the possible canonical representations of a given set is still an open question.

3. Markovian binary trees

3.1. Definitions and basic properties

When the arrivals (observable events) of a TMAP give birth to a child TMAP which is stochastically identical to the parent process, then the overall process of births and terminations of TMAPs is referred to as a Markovian binary tree (MBT).

The instances of TMAPs are defined as in Section 2, with absorbing phase 0, transient phases $1, \ldots, n$, transition rate matrices $\mathbf{D_0}$ and $\mathbf{D_1}$ of size $n \times n$, and absorption rate column vector $\mathbf{d}$ of size n. An MBT is initiated at time 0 with a single TMAP with known initial probability vector $\boldsymbol{\alpha}$ of size n (where $\boldsymbol{\alpha}\mathbb{1} = 1$). The initial phase of a child TMAP is determined by the phase transition of its parent at the time of the observable event of the birth. It is define by the elements of the birth rate matrix $\mathbf{B}$ of size $n \times n^2$. The $\mathbf{B}_{i,jk}$ element (which is the short notation for the element $\mathbf{B}_{i,(j-1)n+k}$) is the rate at which a parent in phase i experiences an observable event and moves to phase k by giving birth to a new TMAP with initial phase j. The matrix $\mathbf{D_1}$ can be obtained from $\mathbf{B}$ as $\mathbf{D_1} = \mathbf{B}(\mathbb{1} \otimes \mathbf{I})$, where $\mathbf{I}$ is the identity matrix and $\otimes$ stands for Kronecker product. Consequently, the triple $(\boldsymbol{\alpha}, \mathbf{D_0}, \mathbf{B})$ of size $(n, n \times n, n \times n^2)$ completely defines an MBT.

Let $X_0, X_1, \ldots,$ denote the inter-arrival times in the initial TMAP, and $Y_0, Y_1, \ldots,$ denote the inter-arrival times of the first child TMAP (which starts at time X_0). The random variables X_i and Y_i $(i \geq 0)$ have a defective distribution. As in Section 2.1, the probability density function of X_0 is

$$f_{X_0}(x) = \boldsymbol{\alpha} e^{\mathbf{D_0} x} \mathbf{D_1} \mathbb{1} = \boldsymbol{\alpha} e^{\mathbf{D_0} x} \mathbf{B} \mathbb{1}\!\mathbb{1}. \tag{13}$$

where $\mathbb{1}\!\mathbb{1} = \mathbb{1} \otimes \mathbb{1}$ explicitly refers to the size of the column vector of ones, which is n^2. Similarly to TMAPs, we refer to the row vector $\boldsymbol{\alpha}$ as the initial vector, and to the column vector $\mathbb{1}\!\mathbb{1}$ as the closing vector. The Laplace transform of X_0 is given by

$$f^*_{X_0}(s) = E[e^{-sX_0}] = \boldsymbol{\alpha}(s\mathbf{I} - \mathbf{D_0})^{-1} \mathbf{B} \mathbb{1}\!\mathbb{1}. \tag{14}$$

Let $\mathbf{R} = (-\mathbf{D_0})^{-1} \mathbf{B}$ be the transient probability matrix of size $n \times n^2$ of the embedded process at observable events of births. From $\mathbf{D} \mathbb{1} \leq 0$ we have $\mathbf{R} \mathbb{1}\!\mathbb{1} \leq \mathbb{1}$ element-wise.

The defective moments of X_0 are as follows

$$\begin{aligned} E(X_0^i I_{\{X_0 < \infty\}}) &= \int_{x=0}^{\infty} x^i f_{X_0}(x) dx \\ &= i! \, \boldsymbol{\alpha}(-\mathbf{D_0})^{-i-1} \mathbf{B} \mathbb{1}\!\mathbb{1} \\ &= i! \, \boldsymbol{\alpha} \mathbf{E}^i \mathbf{R} \mathbb{1}\!\mathbb{1}. \end{aligned} \tag{15}$$

The joint density function of the inter-arrival times (X_0, X_1, Y_0) is given by

$$f(x_0, x_1, y_0) = \boldsymbol{\alpha} e^{\mathbf{D_0} x_0} \mathbf{B} \left(e^{\mathbf{D_0} x_1} \mathbf{B} \mathbb{1}\!\mathbb{1} \otimes e^{\mathbf{D_0} y_0} \mathbf{B} \mathbb{1}\!\mathbb{1} \right). \tag{16}$$

Since $X_0, X_1,$ and Y_0 have defective distributions, we have

$$\mathrm{P}(X_0 < \infty, X_1 < \infty, Y_0 < \infty) = \int_{(x_0)} \int_{(x_1)} \int_{(y_0)} f(x_0, x_1, y_0) dx_0 dx_1 dy_0 \leq 1. \tag{17}$$

From the joint density function, the defective joint moments of the inter-arrival times (X_0, X_1, Y_0) are

$$\begin{aligned} & E(X_0^{i_0} X_1^{i_1} Y_0^{j_0} I_{\{X_0 < \infty, X_1 < \infty, Y_0 < \infty\}}) \\ &= \int_{(x_0)} \int_{(x_1)} \int_{(y_0)} x_0^{i_0} x_1^{i_1} y_0^{j_0} f(x_0, x_1, y_0) dx_0 dx_1 dy_0 \\ &= i_0! i_1! j_0! \; \boldsymbol{\alpha}(-\mathbf{D_0})^{-i_0-1} \mathbf{B} \left((-\mathbf{D_0})^{-i_1-1} \mathbf{B} \mathbb{1}\!\mathbb{1} \otimes (-\mathbf{D_0})^{-j_0-1} \mathbf{B} \mathbb{1}\!\mathbb{1} \right) \\ &= i_0! i_1! j_0! \; \boldsymbol{\alpha} \mathbf{E}^{i_0} \mathbf{R} \left(\mathbf{E}^{i_1} \mathbf{R} \mathbb{1}\!\mathbb{1} \otimes \mathbf{E}^{j_0} \mathbf{R} \mathbb{1}\!\mathbb{1} \right). \end{aligned} \tag{18}$$

Similarly to TMAPs, the $(\boldsymbol{\alpha}, \mathbf{D_0}, \mathbf{B})$ representation of an MBT is not unique. For any nonsingular matrix $\mathbf{G}$ such that $\mathbf{G}\mathbb{1} = \mathbb{1}$,

$$(\boldsymbol{\alpha}\mathbf{G}, \mathbf{G}^{-1}\mathbf{D_0}\mathbf{G}, \mathbf{G}^{-1}\mathbf{B}(\mathbf{G} \otimes \mathbf{G}))$$

is an other equivalent representation of the same MBT. Indeed, we have in particular

$$\begin{aligned}
&f_{(\boldsymbol{\alpha}\mathbf{G}, \mathbf{G}^{-1}\mathbf{D_0}\mathbf{G}, \mathbf{G}^{-1}\mathbf{B}\mathbf{G}\otimes\mathbf{G})}(x_0, x_1, y_0) \\
&= \boldsymbol{\alpha}\mathbf{G}e^{\mathbf{G}^{-1}\mathbf{D_0}\mathbf{G}x_0}\mathbf{G}^{-1}\mathbf{B}(\mathbf{G} \otimes \mathbf{G}) \cdot \\
&\quad \cdot \left(e^{\mathbf{G}^{-1}\mathbf{D_0}\mathbf{G}x_1}\mathbf{G}^{-1}\mathbf{B}(\mathbf{G} \otimes \mathbf{G})\mathbb{1} \otimes e^{\mathbf{G}^{-1}\mathbf{D_0}\mathbf{G}y_0}\mathbf{G}^{-1}\mathbf{B}(\mathbf{G} \otimes \mathbf{G})\mathbb{1}\right) \\
&= \boldsymbol{\alpha}e^{\mathbf{D_0}x_0}\mathbf{B}\left(e^{\mathbf{D_0}x_1}\mathbf{B}\mathbb{1} \otimes e^{\mathbf{D_0}y_0}\mathbf{B}\mathbb{1}\right) \\
&= f_{(\boldsymbol{\alpha}, \mathbf{D_0}, \mathbf{B})}(x_0, x_1, y_0).
\end{aligned}$$

Definition 4. *The* $(\boldsymbol{\alpha}, \mathbf{D_0}, \mathbf{B})$ *representation of an MBT is* Markovian *if* $\boldsymbol{\alpha} \geq \mathbf{0}$ *and* $\mathbf{B} \geq \mathbf{0}$ *element-wise, and the elements of* $\mathbf{D_0}$ *satisfy,* $\mathbf{D_0}_{ii} < 0$ *and* $\mathbf{D_0}_{ij} \geq 0$ *for* $i \neq j$*.*

The marginal rank property is also crucial for MBTs.

Definition 5. *An MBT has* full rank marginal *when the order of its first inter-arrival time distribution* X_0 *is identical with the size of* $\mathbf{D_0}$ *(more precisely, with the size of the minimal representation according to Section 3.4). Otherwise the MBT has* reduced rank marginal*.*

Example 6. An example of an MBT with reduced rank marginal is the following

$$\boldsymbol{\alpha} = (1, 0), \quad \mathbf{D_0} = \begin{pmatrix} -1 & 0 \\ 0 & -2 \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} 0.5 & 0.1 & 0.2 & 0.1 \\ 1 & 0.1 & 0.2 & 0.2 \end{pmatrix},$$

where

$$f^*_{X_0}(s) = \boldsymbol{\alpha}(s\mathbf{I} - \mathbf{D_0})^{-1}\,\mathbf{B}\mathbb{1} = \frac{9}{10(s+1)}$$

indicates that the first inter-arrival time has an order one distribution.

In this work we assume that MBTs have full rank marginal, and we investigate the properties of the following set of reduced moments and joint moments

$$\mu_i = \frac{E(X_0^i I_{\{X_0<\infty\}})}{i!} = \boldsymbol{\alpha}\mathbf{E}^i\mathbf{R}1\!\!1, \quad i \geq 0 \tag{19}$$

$$\gamma_{ijk} = \frac{E(X_0^i X_1^j Y_0^k I_{\{X_0<\infty, X_1<\infty, Y_0<\infty\}})}{i!j!k!} = \boldsymbol{\alpha}\mathbf{E}^i\mathbf{R}\left(\mathbf{E}^j\mathbf{R}1\!\!1 \otimes \mathbf{E}^k\mathbf{R}1\!\!1\right), \quad i,j,k \geq 0 \tag{20}$$

where recall that $\mathbf{E} = (-\mathbf{D_0})^{-1}$.

3.2. Moments based characterization of MBT

Recall from Section 2.2 that all reduced moments $\{\mu_i\}_{i\geq 0}$ of X_0 can be obtained from the first $2n$ reduced moments $\{\mu_i\}_{i=0,1,\dots,2n-1}$. We now investigate how many additional joint moments of (X_0, X_1, Y_0) are needed to obtain the reduced joint moment $\gamma_{i,j,k}$ for any $i,j,k \geq 0$. The next result is the analogue of Theorem 1 for MBTs.

Theorem 5. *Having the reduced moments $\{\mu_i\}_{i\geq 0}$ of X_0, the number of independent reduced joint moments $\gamma_{i,j,k}$ is at most $n^3 - 1$.*

Proof. Recall that $\mathcal{N} = \{0, 1, \dots, n-1\}$. Using the same type of argument as in the proof of Theorem 1, we show that, based on the reduced moments $\{\mu_i\}_{i\geq 0}$ of X_0 and the reduced joint moments $\{\gamma_{i,j,k}\}_{i,j,k\in\mathcal{N}^3\setminus\{n-1,n-1,n-1\}}$ of (X_0, X_1, Y_0), all other joint moments can be computed.

We first show that $\gamma_{i,j,k}$, for $i \geq n$ and j,k fixed, can be computed from lower moments. Indeed, define the moments matrix $\mathbf{M_3}$ of size $(n+1)\times(n+1)$ as

$$\mathbf{M_3} = \begin{bmatrix} \boldsymbol{\alpha}\mathbf{E}^{i-n} \\ \vdots \\ \boldsymbol{\alpha}\mathbf{E}^{i-1} \\ \boldsymbol{\alpha}\mathbf{E}^{i} \end{bmatrix} \left[\ \mathbf{E}^0\mathbf{R}1\!\!1 \,\middle|\, \cdots \,\middle|\, \mathbf{E}^{n-1}\mathbf{R}1\!\!1 \,\middle|\, \mathbf{R}(\mathbf{E}^j\mathbf{R}1\!\!1 \otimes \mathbf{E}^k\mathbf{R}1\!\!1)\ \right]$$

$$= \begin{pmatrix} \mu_{i-n} & \cdots & \mu_{i-1} & \gamma_{i-n,j,k} \\ \vdots & \ddots & \vdots & \vdots \\ \mu_{i-1} & \cdots & \mu_{i+n-2} & \gamma_{i-1,j,k} \\ \mu_i & \cdots & \mu_{i+n-1} & \gamma_{i,j,k} \end{pmatrix}.$$

Since $\mathbf{M_3}$ is the product of an $(n+1)\times n$ matrix and an $n\times(n+1)$ matrix, its determinant is 0, and its lower right element $\gamma_{i,j,k}$ is determined by the other elements.

We show similarly that $\gamma_{i,j,k}$, for $j\geq n$ and i,k fixed, respectively for $k\geq n$ and i,j fixed, can be computed from lower moments. Indeed, define the moments matrices $\mathbf{M_4}$ and $\mathbf{M_5}$ of size $(n+1)\times(n+1)$ as follows

$$\mathbf{M_4}=\begin{bmatrix}\boldsymbol{\alpha}\mathbf{E}^0\otimes\boldsymbol{\alpha}\\ \vdots\\ \boldsymbol{\alpha}\mathbf{E}^{n-1}\otimes\boldsymbol{\alpha}\\ \boldsymbol{\alpha}\mathbf{E}^i\mathbf{R}\end{bmatrix}\times$$
$$\times\left[\ \mathbf{E}^{j-n}\mathbf{R}\mathbb{1}\otimes\mathbf{E}^k\mathbf{R}\mathbb{1}\ \middle|\ \cdots\ \middle|\ \mathbf{E}^{j-1}\mathbf{R}\mathbb{1}\otimes\mathbf{E}^k\mathbf{R}\mathbb{1}\ \middle|\ \mathbf{E}^j\mathbf{R}\mathbb{1}\otimes\mathbf{E}^k\mathbf{R}\mathbb{1}\ \right]$$
$$=\begin{pmatrix}\mu_{j-n}\mu_k & \cdots & \mu_{j-1}\mu_k & \mu_j\mu_k\\ \vdots & \ddots & \vdots & \vdots\\ \mu_{j-1}\mu_k & \cdots & \mu_{n+j-2}\mu_k & \mu_{n+j-1}\mu_k\\ \gamma_{i,j-n,k} & \cdots & \gamma_{i,j-1,k} & \boxed{\gamma_{i,j,k}}\end{pmatrix}.$$

$$\mathbf{M_5}=\begin{bmatrix}\boldsymbol{\alpha}\mathbf{E}^0\otimes\boldsymbol{\alpha}\\ \vdots\\ \boldsymbol{\alpha}\mathbf{E}^{n-1}\otimes\boldsymbol{\alpha}\\ \boldsymbol{\alpha}\mathbf{E}^i\mathbf{R}\end{bmatrix}\times$$
$$\times\left[\ \mathbf{E}^{j}\mathbf{R}\mathbb{1}\otimes\mathbf{E}^{k-n}\mathbf{R}\mathbb{1}\ \middle|\ \cdots\ \middle|\ \mathbf{E}^{j}\mathbf{R}\mathbb{1}\otimes\mathbf{E}^{k-1}\mathbf{R}\mathbb{1}\ \middle|\ \mathbf{E}^j\mathbf{R}\mathbb{1}\otimes\mathbf{E}^k\mathbf{R}\mathbb{1}\ \right]$$
$$=\begin{pmatrix}\mu_j\mu_{k-n} & \cdots & \mu_j\mu_{k-1} & \mu_j\mu_k\\ \vdots & \ddots & \vdots & \vdots\\ \mu_{n-1+j}\mu_{k-n} & \cdots & \mu_{n-1+j}\mu_{k-1} & \mu_{n-1+j}\mu_k\\ \gamma_{i,j,k-n} & \cdots & \gamma_{i,j,k-1} & \boxed{\gamma_{i,j,k}}\end{pmatrix}.$$

The matrices $\mathbf{M_4}$ and $\mathbf{M_5}$ are the product of an $(n+1)\times n^2$ matrix and an $n^2\times(n+1)$ matrix. If the rank of the $n^2\times(n+1)$ matrices is n, then the determinant of $\mathbf{M_4}$ and $\mathbf{M_5}$ is 0, and therefore their lower right element is determined by the other elements.

To show that the two $n^2\times(n+1)$ matrices in the composition of $\mathbf{M_4}$ and $\mathbf{M_5}$ are indeed of rank n, first observe that the $n^2\times 1$ vector $\mathbf{z}=\mathbf{v}\otimes\mathbf{w}$ obtained as the Kronecker product of two $n\times 1$ vectors $\mathbf{v}$ and $\mathbf{w}$ satsifies the following property: for each $0 \le j \le n-1$, the ratio z_{jn+k_1}/z_{jn+k_2} is independent of $\mathbf{v}$ for any $k_1, k_2 \in \{1, 2, \ldots, n\}$, and similarly, for each $1 \le k \le n$, the ratio $z_{j_1 n+k}/z_{j_2 n+k}$ is independent of $\mathbf{w}$ for any $j_1, j_2 \in \{0, 1, \ldots, n-1\}$. As a consequence, in the $n^2 \times (n+1)$ matrix composing $\mathbf{M_4}$, for each $0 \le j \le n-1$, the rows $jn+k$ for $k = 1, \ldots, n$ are linearly dependent, and in the $n^2 \times n+1$ matrix composing $\mathbf{M_5}$, for each $1 \le k \le n$, the rows $jn+k$ for $j = 0, \ldots, n-1$ are linearly dependent. So the maximum number of independent rows is n in both matrices, and therefore they are of rank n.

Using matrices $\mathbf{M_3}$, $\mathbf{M_4}$ and $\mathbf{M_5}$, we can therefore compute all the reduced joint moments $\gamma_{i,j,k}$ based on the n^3 joint moments $\{\gamma_{i,j,k}\}_{i,j,k \in \mathcal{N}^3}$. The number of independent moments further reduces by one using the following relation among this set of moments: define the moments matrix $\mathbf{M_6}$ of size $(n^2+1) \times (n^2+1)$ as

$$\mathbf{M_6} = \begin{bmatrix} \boldsymbol{\alpha} \otimes \boldsymbol{\alpha} \\ \boldsymbol{\alpha}\mathbf{E}^0\mathbf{R} \\ \vdots \\ \boldsymbol{\alpha}\mathbf{E}^{n-1}\mathbf{R} \\ \boldsymbol{\alpha}\mathbf{E}^{n}\mathbf{R} \\ \vdots \\ \boldsymbol{\alpha}\mathbf{E}^{n^2-1}\mathbf{R} \end{bmatrix} \times$$

$$\times \left[\ \mathbb{1} \mid \mathbf{E}^0\mathbf{R}\mathbb{1} \otimes \mathbf{E}^0\mathbf{R}\mathbb{1} \mid \mathbf{E}^0\mathbf{R}\mathbb{1} \otimes \mathbf{E}^1\mathbf{R}\mathbb{1} \mid \cdots \mid \mathbf{E}^{n-1}\mathbf{R}\mathbb{1} \otimes \mathbf{E}^{n-1}\mathbf{R}\mathbb{1}\ \right]$$

$$= \left(\begin{array}{ccccc} 1 & \mu_0\mu_0 & \mu_0\mu_1 & \cdots & \mu_{n-1}\mu_{n-1} \\ \mu_0 & \gamma_{0,0,0} & \gamma_{0,0,1} & \cdots & \gamma_{0,n-1,n-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mu_{n-1} & \gamma_{n-1,0,0} & \gamma_{n-1,0,1} & \cdots & \boxed{\gamma_{n-1,n-1,n-1}} \\ \hline \mu_n & \gamma_{n,0,0} & \gamma_{n,0,1} & \cdots & \gamma_{n,n-1,n-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mu_{n^2-1} & \gamma_{n^2-1,0,0} & \gamma_{n^2-1,0,1} & \cdots & \gamma_{n^2-1,n-1,n-1} \end{array} \right).$$

The matrix $\mathbf{M_6}$ is the product of an $(n^2+1) \times n^2$ matrix and an $n^2 \times (n^2+1)$ matrix, and therefore its determinant is 0, and the element $\gamma_{n-1,n-1,n-1}$ can be determined by the other elements. Moreover, according to $\mathbf{M_3}$, the moments below the horizontal line can be obtained from the moments above the line. Consequently, the element $\gamma_{n-1,n-1,n-1}$ can be obtained from the moments above the line too, which reduces the number of independent $\gamma_{i,j,k}$ moments to $n^3 - 1$.

□

Theorem 6. *Based on the reduced moments $\{\mu_i\}_{i\in\{0,1,\ldots,2n-1\}}$ of X_0, and the reduced joint moments $\{\gamma_{i,j,k}\}_{i,j,k\in\mathcal{N}^3\setminus\{n-1,n-1,n-1\}}$ of (X_0, X_1, Y_0), all other joint moments can be computed.*

Proof. The same recursive approach as the one in [8] is applicable here. □

Theorem 7. *The number of moments that characterize an MBT of size n with full rank marginal is $n^3 + 2n - 1$.*

Proof. It is a consequence of the preceding two theorems. □

Remark 4. As for TMAPs, we can give two intuitive interpretations for the number of independent parameters in an MBT. First, note that the total number of elements in the vectors and matrices of the $(\boldsymbol{\alpha}, \mathbf{D_0}, \mathbf{B})$ representation of an MBT is $(n-1) + n^2 + n^3$. As the $(\boldsymbol{\alpha}, \mathbf{D_0}, \mathbf{B})$ representation is not unique and a similarity transformation matrix $\mathbf{G}$ contains $n^2 - n$ free parameters, the total number of independent parameters is reduced to $(n-1) + n^2 + n^3 - (n^2 - n) = n^3 + 2n - 1$.

Alternatively, an MBT $(\boldsymbol{\alpha}, \mathbf{D_0}, \mathbf{B})$ is entirely characterized by a TMAP $(\boldsymbol{\alpha}, \mathbf{D_0}, \mathbf{D_1})$ which governs the lifetime and reproduction of individuals and contains $n^2 + 2n - 1$ independent parameters, and by a stochastic $n^2 \times n$ matrix $\mathbf{S}$ whose entries $\mathbf{S}_{ik,j}$ are such that $\mathbf{B}_{i,jk} = \mathbf{D_1}_{ik}\, \mathbf{S}_{ik,j}$. As the matrix $\mathbf{S}$ contains $n^2(n-1)$ independent parameters, the total number of independent parameters of an MBT is $(n^2 + 2n - 1) + n^2(n-1) = n^3 + 2n - 1$.

Example 7. The MBT with representation

$$\boldsymbol{\alpha} = (0.6, 0.4), \quad \mathbf{D_0} = \begin{pmatrix} -4 & 0 \\ 1 & -6 \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} 1 & 1 & 0 & 1 \\ 3 & 0 & 1 & 0 \end{pmatrix}, \quad \mathbf{d} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

is entirely characterized by the following reduced moments of X_0 and joint moments of (X_0, X_1, Y_0)

$$\mu_0 = 0.7667, \quad \mu_1 = 0.0421, \quad \mu_2 = 0.0211, \quad \mu_3 = 0.0017,$$

$$\gamma_{0,0,0} = 0.4492, \quad \gamma_{0,0,1} = 0.1034, \quad \gamma_{0,1,0} = 0.1060, \quad \gamma_{0,1,1} = 0.0245,$$
$$\gamma_{1,0,0} = 0.1046, \quad \gamma_{1,0,1} = 0.0238, \quad \gamma_{1,1,0} = 0.0246.$$

3.3. Finding an MBT representation based on the moments

Similarly to Section 2.3, based on the reduced moments $\{\mu_i\}_{i\in\{0,1,\dots,2n-1\}}$ of X_0, we can obtain a vector-matrix pair $(\boldsymbol{\phi}, \mathbf{K})$ such that

$$\mu_i = \mu_0 \boldsymbol{\phi} \mathbf{K}^i \mathbb{1}, \quad i = 0, 1, \dots, n-1,$$

where $\boldsymbol{\phi} = (1/n, \dots, 1/n)$. From

$$\mu_i = \boldsymbol{\alpha} \mathbf{E}^i \mathbf{R} \mathbb{1\!\!1} = \boldsymbol{\phi} \mathbf{K}^i \mu_0 \mathbb{1} \tag{21}$$

and [7, Theorem 1] there exists a non singular matrix $\mathbf{H}$ such that $\boldsymbol{\phi} = \boldsymbol{\alpha}\mathbf{H}$, $\mathbf{K} = \mathbf{H}^{-1}\mathbf{E}\mathbf{H}$ and $\mu_0 \mathbb{1} = \mathbf{H}^{-1}\mathbf{R}\mathbb{1\!\!1}$. Using this, from (20) we have

$$\gamma_{ijk} = \underbrace{\boldsymbol{\alpha}\mathbf{H}}_{\boldsymbol{\phi}} \underbrace{\mathbf{H}^{-1}\mathbf{E}^i\mathbf{H}}_{\mathbf{K}^i} \underbrace{\mathbf{H}^{-1}\mathbf{R}(\mathbf{H}\otimes\mathbf{H})}_{\hat{\mathbf{R}}} \Big(\underbrace{\mathbf{H}^{-1}\mathbf{E}^j\mathbf{H}}_{\mathbf{K}^j} \underbrace{\mathbf{H}^{-1}\mathbf{R}\mathbb{1\!\!1}}_{\mu_0\mathbb{1}} \otimes \underbrace{\mathbf{H}^{-1}\mathbf{E}^k\mathbf{H}}_{\mathbf{K}^k} \underbrace{\mathbf{H}^{-1}\mathbf{R}\mathbb{1\!\!1}}_{\mu_0\mathbb{1}} \Big) = \boldsymbol{\phi}\mathbf{K}^i\hat{\mathbf{R}}\left(\mathbf{K}^j\mu_0\mathbb{1} \otimes \mathbf{K}^k\mu_0\mathbb{1}\right), \tag{22}$$

where $\hat{\mathbf{R}} = \mathbf{H}^{-1}\mathbf{R}\mathbf{H}$ is of size $n \times n^2$. The same approach as in Section 2.3 is applicable here: we define the matrices $\boldsymbol{\Phi}$ of size $n \times n$, $\boldsymbol{\Lambda}$ of size $n^2 \times n^2$ and $\boldsymbol{\Gamma}$ of size $n \times n^2$, such that the ith row of $\boldsymbol{\Phi}$ is $\boldsymbol{\phi}\mathbf{K}^i$, the $jn+k$th column of $\boldsymbol{\Lambda}$ is $\mathbf{K}^j\mu_0\mathbb{1} \otimes \mathbf{K}^k\mu_0\mathbb{1}$ and the $(i, jn+k)$ element of $\boldsymbol{\Gamma}$ is $\gamma_{i,j,k}$, where rows and columns are numbered from 0 to $n-1$ or to n^2-1. With these notations (20) can be written in matrix form as

$$\boldsymbol{\Gamma} = \boldsymbol{\Phi}\hat{\mathbf{R}}\boldsymbol{\Lambda}, \tag{23}$$

from which

$$\hat{\mathbf{R}} = \boldsymbol{\Phi}^{-1}\boldsymbol{\Gamma}\boldsymbol{\Lambda}^{-1}, \tag{24}$$

where $\boldsymbol{\Phi}$ and $\boldsymbol{\Lambda}$ are nonsingular according to Remark 1. Now, $(\boldsymbol{\alpha}, \mathbf{E}, \mathbf{R})$ with closing vector $\mathbb{1\!\!1} = \mathbb{1} \otimes \mathbb{1}$, and $[\boldsymbol{\phi}, \mathbf{K}, \hat{\mathbf{R}}, \boldsymbol{v} \otimes \boldsymbol{v}]$ with closing vector $\boldsymbol{v} \otimes \boldsymbol{v}$, are two equivalent representations of the same MBT. From

$$\mu_i = \boldsymbol{\phi}\mathbf{K}^i\mu_0\mathbb{1} = \boldsymbol{\phi}\mathbf{K}^i\hat{\mathbf{R}}(\boldsymbol{v}\otimes\boldsymbol{v}), \quad i = 0, 1, \dots 2n-1,$$

we have that $\boldsymbol{v}$ must satisfy $\mu_0\mathbb{1} = \hat{\mathbf{R}}(\boldsymbol{v}\otimes\boldsymbol{v})$ and $\boldsymbol{\phi}\boldsymbol{v} = 1$. Relations (21) and (22) indicate that any $\boldsymbol{v}$ satisfying this conditions is a proper closing vector for the $[\boldsymbol{\phi}, \mathbf{K}, \hat{\mathbf{R}}, \boldsymbol{v} \otimes \boldsymbol{v}]$ representation.

Finally, we need to transform this representation such that the closing vector becomes $\mathbb{1}$. Let $\mathbf{C}$ be a non-singular matrix such that $\mathbf{C}\mathbb{1} = \boldsymbol{v}$ and $\boldsymbol{\phi}\mathbf{C}\mathbb{1} = 1$. Then

$$(\boldsymbol{\phi}\mathbf{C}, \mathbf{C}^{-1}\mathbf{K}\mathbf{C}, \mathbf{C}^{-1}\hat{\mathbf{R}}\,(\mathbf{C}\otimes\mathbf{C}))$$

with closing vector $\mathbf{C}^{-1}\boldsymbol{v} = \mathbb{1}$ is another equivalent representation of the same MBT.

Remark 5. As in the TMAP case, the representation found with this method is generally non-Markovian, but in some cases the algorithm developed in [7] can be adapted to the MBT case too in order to transform any non Markovian representation into an equivalent Markovian representation [19].

Example 8. Using the RBTFromMoments function of the BuTools package we can compute the following non-Markovian representation based on the reduced moments and joint moments in Example 7

$$\boldsymbol{\alpha} = (0.4969, 0.5031),$$

$$\mathbf{D_0} = \begin{pmatrix} -7.8437 & 3.3280 \\ -2.1295 & -2.1562 \end{pmatrix}, \ \mathbf{B} = \begin{pmatrix} 3.9473 & -3.9669 & -4.1036 & 7.6388 \\ 5.4557 & -5.1895 & -3.3262 & 6.3458 \end{pmatrix}.$$

This representation can be transformed into the following equivalent Markovian representation by the RBTToMBT function of the BuTools package

$$\boldsymbol{\alpha} = (0.4538, 0.5462),$$

$$\mathbf{D_0} = \begin{pmatrix} -5.9486 & 0.1193 \\ 0.8390 & -4.0514 \end{pmatrix}, \ \mathbf{B} = \begin{pmatrix} 0.2406 & 1.5950 & 0.5791 & 2.4146 \\ 0.9876 & 0.0603 & 1.0444 & 0.1199 \end{pmatrix}.$$

3.4. Minimal representation of MBTs

In this section, we present a procedure to compute a minimal representation of any general MBT. We note again that the minimal representation is not necessarily Markovian. For the analysis of MBTs we need to generalize the approach used in [9] and in Section 2.4.

We start with the relation between the minimal size of MBTs and the minimal size of the related transient marked MAPs (TMMAPs), mainly because the later ones can be analyzed by the marked version of the STAIRCASE method (denoted as MSTAIRCASE hereafter) [9]. The class of TMMAPs is obtained by the extension of marked MAPs (MMAPs) with an absorbing state in a similar way as TMAPs are obtained from MAPs. The minimal representation of MMAPs is provided in [9].

Let $\mathbf{D_1} = \mathbf{B}(\mathbb{1} \otimes \mathbf{I})$ and $\mathbf{D_2} = \mathbf{B}(\mathbf{I} \otimes \mathbb{1})$. The next theorem presents a sufficient condition for the minimality of MBTs which can be checked with the MSTAIRCASE method [9]

Theorem 8. *If the TMMAP*$(\boldsymbol{\alpha}, \mathbf{D_0}, \mathbf{D_1}, \mathbf{D_2})$ *is minimal, then the MBT with representation* $(\boldsymbol{\alpha}, \mathbf{D_0}, \mathbf{B})$ *is also minimal.*

Proof. First we consider the minimality of the representations with respect to the initial vector.

A TMMAP$(\boldsymbol{\alpha}, \mathbf{D_0}, \mathbf{D_1}, \mathbf{D_2})$ of size $(1 \times n, n \times n, n \times n, n \times n)$ is minimal with respect to the initial vector if there is no size $s < n$ and non-singular similarity matrix of size $n \times n$ with unit row sum, $\mathbf{W}$, such that $(\boldsymbol{\alpha}\mathbf{W})_\ell = 0$ for $\ell = s+1, s+2, \ldots, n$, and $(\mathbf{W}^{-1}\mathbf{D_i}\mathbf{W})_{k\ell} = 0$ for $i = 0, 1, 2$, $k = 1, 2, \ldots, s$, and $\ell = s+1, s+2, \ldots, n$.

Since $\mathbf{D_1} = \mathbf{B}(\mathbb{1} \otimes \mathbf{I})$ and $\mathbf{D_2} = \mathbf{B}(\mathbf{I} \otimes \mathbb{1})$, it also means that there is no size $s < n$ and non-singular similarity matrix with unit row sum $\mathbf{W}$ such that $(\boldsymbol{\alpha}\mathbf{W})_\ell = 0$, $(\mathbf{W}^{-1}\mathbf{D_0}\mathbf{W})_{k,\ell} = 0$ for $k = 1, 2, \ldots, s$, and $\ell = s+1, s+2, \ldots, n$, and $(\mathbf{W}^{-1}\mathbf{B}(\mathbf{W} \otimes \mathbf{W}))_{k,(\ell-1)n+m}$ for $k = 1, 2, \ldots, s$, $\ell = s+1, s+2, \ldots, n$, and $m = 1, 2, \ldots, n$, and for $k = 1, 2, \ldots, s$, $\ell = 1, 2, \ldots, n$, and $m = s + 1, s + 2, \ldots, n$. If it was the case that such s and $\mathbf{W}$ exist, then

$$\mathbf{W}^{-1}\mathbf{B}(\mathbf{W} \otimes \mathbf{W})(\mathbb{1} \otimes \mathbf{I}) = \mathbf{W}^{-1}\mathbf{B}(\mathbf{W}\mathbb{1} \otimes \mathbf{W}\mathbf{I}) =$$
$$\mathbf{W}^{-1}\mathbf{B}(\mathbb{1} \otimes \mathbf{W}) = \mathbf{W}^{-1}\mathbf{B}(\mathbb{1} \otimes \mathbf{I})\mathbf{W} = \mathbf{W}^{-1}\mathbf{D_1}\mathbf{W},$$

where $\mathbf{W}\mathbb{1} = \mathbb{1}$ due to the unit row sum of $\mathbf{W}$. Similarly,

$$\mathbf{W}^{-1}\mathbf{B}(\mathbf{W} \otimes \mathbf{W})(\mathbf{I} \otimes \mathbb{1}) = \mathbf{W}^{-1}\mathbf{D_2}\mathbf{W}$$

would contain zero elements at positions (k, ℓ) for $k = 1, 2, \ldots, s$, and $\ell = s+1, s+2, \ldots, n$, which contradicts the fact that the TMMAP$(\boldsymbol{\alpha}, \mathbf{D_0}, \mathbf{D_1}, \mathbf{D_2})$ is minimal with respect to the initial vector.

A similar argument with respect to the closing vector completes the proof. □

Unfortunately, the opposite of Theorem 8 does not hold in general, because in case of non-Markovian representations it might happen that $\mathbf{W}^{-1}\mathbf{D_1}\mathbf{W}$ and $\mathbf{W}^{-1}\mathbf{D_2}\mathbf{W}$ contain only zero elements at positions (k, ℓ) for $k = 1, 2, \ldots, s$, and $\ell = s+1, s+2, \ldots, n$, but there are nonzero elements among $(\mathbf{W}^{-1}\mathbf{B}(\mathbf{W} \otimes \mathbf{W}))_{k,(\ell-1)n+m}$ for $k = 1, 2, \ldots, s$, $\ell = s + 1, s + 2, \ldots, n$, and $m = 1, 2, \ldots, n$, or for $k = 1, 2, \ldots, s$, $\ell = 1, 2, \ldots, n$, $m = s + 1, s + 2, \ldots, n$.

Theorem 9. *Suppose that for an MBT($\boldsymbol{\alpha}, \mathbf{D_0}, \mathbf{B}$) of size m there exists a non singular $m \times m$ transformation matrix $\mathbf{W}$ such that*

$$\mathbf{W}^{-1}\mathbf{B}(\mathbf{W} \otimes \mathbf{W}) = \left(\underbrace{\left| \begin{matrix} \hat{\mathbf{B}}_\mathbf{1} & \star \\ \mathbf{0} & \star \end{matrix} \right| \dots \left| \begin{matrix} \hat{\mathbf{B}}_\mathbf{n} & \star \\ \mathbf{0} & \star \end{matrix} \right|}_{n} \underbrace{\left| \begin{matrix} \star & \star \\ \star & \star \end{matrix} \right| \dots \left| \begin{matrix} \star & \star \\ \star & \star \end{matrix} \right|}_{m-n} \right),$$

$$\mathbf{W}^{-1}\mathbf{D_0}\mathbf{W} = \begin{pmatrix} \hat{\mathbf{D}}_\mathbf{0} & \star \\ \mathbf{0} & \star \end{pmatrix} \text{ and } \mathbf{W}^{-1}\mathbb{1}_m = \begin{pmatrix} \mathbb{1}_n \\ \mathbf{0} \end{pmatrix},$$

where the matrix blocks (between vertical lines) are of size $m \times m$ and their sub-blocks from left to right and from top to bottom are $n \times n$, $n \times (m-n)$, $(m-n) \times n$ and $(m-n) \times (m-n)$, and the $\star$ symbols indicate possibly non-zero (irrelevant) matrix blocks. Then $(\hat{\boldsymbol{\alpha}}, \hat{\mathbf{D}}_\mathbf{0}, \hat{\mathbf{B}})$ is an equivalent representation of the same MBT of size $n < m$, where $\boldsymbol{\alpha}\mathbf{W} = \begin{pmatrix} \hat{\boldsymbol{\alpha}} & \star \end{pmatrix}$ and matrix $\hat{\mathbf{B}}$ of size $n \times n^2$ is composed by $\hat{\mathbf{B}}_\mathbf{1}, \dots, \hat{\mathbf{B}}_\mathbf{n}$ of size $n \times n$.

Proof. None of the $\star$ blocks play role in the behavior of the MBT($\boldsymbol{\alpha}\mathbf{W}, \mathbf{W}^{-1}\mathbf{D_0}\mathbf{W}, \mathbf{W}^{-1}\mathbf{B}(\mathbf{W} \otimes \mathbf{W})$), because all of them are multiplied by a zero element of the closing vector, see for instance (16). □

The following modified version of the STAIRCASE method computes the minimal representation of MBTs with respect to the closing vector. The implementation of the method is available in the BuTools package [19].

STAIRCASE_MBT_Cont($\mathbf{X}$, $\mathbf{Y}$, $\mathbf{Z}$)

$i = 0$; $\{m, j\}$ =SIZE($\mathbf{X}$); $\mathbf{U}^* = \mathbf{I}$; $r_0 = m$;

REPEAT

$i = i + 1$; $r_i = rank(\mathbf{Z})$; $\{\mathbf{U}_i, \mathbf{S}_i, \mathbf{T}_i\}$ =SVD($\mathbf{Z}$) ;

$\begin{pmatrix} \mathbf{Z}_1 \\ \mathbf{0} \end{pmatrix} = \mathbf{U}_i^*\mathbf{Z}$; $\begin{pmatrix} \mathbf{X}_1 & \mathbf{X}_2 \\ \mathbf{X}_3 & \mathbf{X}_4 \end{pmatrix} = \mathbf{U}_i^*\mathbf{X}\mathbf{U}_i$;

/* $\mathbf{X}_1$ is of size $r_i \times r_i$ */

$$\left(\underbrace{\left| \begin{matrix} \mathbf{Y_1}_1 & \mathbf{Y_2}_1 \\ \mathbf{Y_3}_1 & \mathbf{Y_4}_1 \end{matrix} \right| \dots \left| \begin{matrix} \mathbf{Y_1}_{r_i} & \mathbf{Y_2}_{r_i} \\ \mathbf{Y_3}_{r_i} & \mathbf{Y_4}_{r_i} \end{matrix} \right|}_{r_i} \underbrace{\left| \begin{matrix} \mathbf{Y_5}_1 & \mathbf{Y_6}_1 \\ \mathbf{Y_7}_1 & \mathbf{Y_8}_1 \end{matrix} \right| \dots \left| \begin{matrix} \mathbf{Y_5}_{r_{i-1}-r_i} & \mathbf{Y_6}_{r_{i-1}-r_i} \\ \mathbf{Y_7}_{r_{i-1}-r_i} & \mathbf{Y_8}_{r_{i-1}-r_i} \end{matrix} \right|}_{r_{i-1}-r_i} \right)$$

$= \mathbf{U}_i^*\mathbf{Y}(\mathbf{U}_i \otimes \mathbf{U}_i)$;

/* $\mathbf{Y}_1$ is of size $r_i \times r_i^2$ */

7. $\mathbf{U}^* = \begin{pmatrix} \mathbf{I}_{\sum_{j=1}^{i-1} r_j} & \mathbf{0} \\ \mathbf{0} & \mathbf{U}_i^* \end{pmatrix} \mathbf{U}^*$;

8. $\mathbf{X} = \mathbf{X}_4$; $\mathbf{Y} = \mathbf{Y}_8$; $\mathbf{Z} = (\mathbf{X}_3 \ \mathbf{Y}_3)$;

9. UNTIL $rank(\mathbf{Z}) = m - \sum_{j=1}^{i} r_j$ or $\mathbf{Z} = \mathbf{0}$;

10. IF ($\mathbf{Z} = \mathbf{0}$) THEN

11. $n = \sum_{j=1}^{i} r_j$;

12. $\begin{pmatrix} \mathbf{x} \\ \mathbf{0}_{m-n} \end{pmatrix} = \mathbf{U}^* \mathbb{1}_m$;

13. IF ($\mathbf{x} \neq \mathbf{0}$) THEN $\mathbf{R} = \mathbf{I}$ ELSE $\mathbf{R}$ = non-singular matrix such that $\mathbf{R}\mathbf{x} \neq \mathbf{0}$

/* element-wise */

14. $\mathbf{y} = \mathbf{R}\mathbf{x}$; $\mathbf{\Gamma} = diag(\mathbf{y})$; $\mathbf{W} = \left[\begin{pmatrix} \mathbf{\Gamma}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{m-n} \end{pmatrix} \begin{pmatrix} \mathbf{R} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{m-n} \end{pmatrix} \mathbf{U}^* \right]^{-1}$;

15. ELSE $n = m$; $\mathbf{W} = \mathbf{I}$; /* no reduction is possible */

16. RETURN($n, \mathbf{W}$) ;

There is a kind of symmetry in the roles of the initial and closing vectors of traditional or transient MAPs, and rational arrival processes (RAPs, the analogue of MAPs with non-Markovian representation). Unfortunately, this symmetry neither holds for MBTs nor for rational binary trees (RBTs, the analogue of MBTs with non-Markovian representation). As a consequence an essentially different method reduces the size of MBTs which are non-minimal due to the initial vector.

Theorem 10. *If for an MBT($\boldsymbol{\alpha}, \mathbf{D_0}, \mathbf{B}$) of size m there exists a non singular $m \times m$ transformation matrix $\mathbf{W}$ such that*

$$\mathbf{W}^{-1}\mathbf{B}(\mathbf{W} \otimes \mathbf{W}) = \left(\underbrace{\left. \begin{array}{cc} \hat{\mathbf{B}}_\mathbf{1} & \mathbf{0} \\ \star & \star \end{array} \right| \dots \left| \begin{array}{cc} \hat{\mathbf{B}}_\mathbf{n} & \mathbf{0} \\ \star & \star \end{array} \right|}_{n} \left| \underbrace{\begin{array}{cc} \mathbf{0} & \mathbf{0} \\ \star & \star \end{array} \right| \dots \left| \begin{array}{cc} \mathbf{0} & \mathbf{0} \\ \star & \star \end{array}}_{m-n} \right) ,$$

$$\mathbf{W}^{-1}\mathbf{D_0}\mathbf{W} = \begin{pmatrix} \hat{\mathbf{D}}_\mathbf{0} & \mathbf{0} \\ \star & \star \end{pmatrix} \text{ and } \alpha\mathbf{W} = \begin{pmatrix} \hat{\alpha} & \mathbf{0} \end{pmatrix},$$

then ($\hat{\boldsymbol{\alpha}}, \hat{\mathbf{D}}_\mathbf{0}, \hat{\mathbf{B}}$) is an equivalent representation of size $n < m$, where matrix $\hat{\mathbf{B}}$ of size $n \times n^2$ is composed by $\hat{\mathbf{B}}_\mathbf{1}, \dots, \hat{\mathbf{B}}_\mathbf{n}$ of size $n \times n$.

Proof. None of the $\star$ elements plays role in the behavior of the MBT($\boldsymbol{\alpha}\mathbf{W}, \mathbf{W}^{-1}\mathbf{D_0}\mathbf{W}, \mathbf{W}^{-1}\mathbf{B}(\mathbf{W} \otimes \mathbf{W})$) when the matrix blocks indicated by $\mathbf{0}$ contain only zero elements. □

The decomposition of the matrix $\mathbf{W}^{-1}\mathbf{B}(\mathbf{W} \otimes \mathbf{W})$ suggests a simple implementation for finding such matrix $\mathbf{W}$ based on the MSTAIRCASE method available for TMMAPs. Let the matrix $\mathbf{B_i}$ of size $m \times m$ be the ith block of matrix $\mathbf{B}$ such that $\mathbf{B} = (\mathbf{B_1}, \ldots, \mathbf{B_m})$.

Theorem 11. *The MBT($\boldsymbol{\alpha}, \mathbf{D_0}, \mathbf{B}$) and the TMMAP($\boldsymbol{\alpha}, \mathbf{D_0}, \mathbf{B_1}, \ldots, \mathbf{B_m}$) are minimal with respect to the initial vector at the same time.*

Proof. Practically the same structural properties are required for the two processes to be non-minimal. If there is a matrix $\mathbf{W}$ which fulfils the conditions of Theorem 10, then $\mathbf{W}$ also transforms the TMMAP($\boldsymbol{\alpha}, \mathbf{D_0}, \mathbf{B_1}, \ldots, \mathbf{B_m}$) into a form which fulfils the structural conditions for being non-minimal [9]. □

As a consequence of this theorem, the MSTAIRCASE method applied for finding the minimal representation of the TMMAP($\boldsymbol{\alpha}, \mathbf{D_0}, \mathbf{B_1}, \ldots, \mathbf{B_m}$) with respect to the initial vector computes an appropriate matrix $\mathbf{W}$ for finding the minimal representation of the MBT($\boldsymbol{\alpha}, \mathbf{D_0}, \mathbf{B}$) with respect to the initial vector. This procedure is also implemented in the BuTools package [19].

Example 9. An example of an MBT which is non-minimal due to the closing vector is the following:

$$\boldsymbol{\alpha} = (0.1, 0.2, 0.7), \quad \mathbf{D_0} = \begin{pmatrix} -24 & 8 & 1 \\ 4 & -20 & 1 \\ 1 & 2 & -21 \end{pmatrix},$$

$$\mathbf{B} = \begin{pmatrix} 1 & 2 & 2 & 1 & 1 & 2 & 0 & 3 & 2 \\ 2 & 0 & 1 & 1 & 2 & 3 & 1 & 2 & 2 \\ 1 & 1 & 4 & 1 & 1 & 2 & 2 & 0 & 4 \end{pmatrix}, \quad \mathbf{d} = \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix}.$$

This MBT is lumpable into the equivalent representation of size 2

$$\boldsymbol{\alpha} = (0.3, 0.7), \quad \mathbf{D_0} = \begin{pmatrix} -16 & 1 \\ 3 & -21 \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} 5 & 4 & 3 & 2 \\ 4 & 6 & 2 & 4 \end{pmatrix}, \quad \mathbf{d} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}.$$

RBTMinimize function of the BuTools package gives the minimal representation

$$\boldsymbol{\alpha} = (0.222183, 0.777817), \quad \mathbf{D_0} = \begin{pmatrix} -13.8787 & -2.1213 \\ 5.70711 & -23.1213 \end{pmatrix},$$

$$\mathbf{B} = \begin{pmatrix} 8.87174 & 4.85212 & 0.609476 & 0.333333 \\ 7.32639 & 7.00694 & -0.235702 & 1.51184 \end{pmatrix}, \quad \mathbf{d} = \begin{pmatrix} 1.33333 \\ 1.80474 \end{pmatrix}.$$

Example 10. An example of an MBT which is non-minimal due to the initial vector is the following:

$$\boldsymbol{\alpha} = (0.25, 0.25, 0.5), \quad \mathbf{D_0} = \begin{pmatrix} -25 & 4 & 3 \\ 4 & -25 & 1 \\ 1 & 1 & -31 \end{pmatrix},$$

$$\mathbf{B} = \begin{pmatrix} 1 & 2 & 3 & 0 & 1.5 & 0 & 2 & 1 & 0 \\ 2 & 1 & 1 & 3 & 1.5 & 4 & 0 & 1 & 2 \\ 3 & 3 & 4 & 3 & 3 & 4 & 2 & 2 & 1 \end{pmatrix}, \quad \mathbf{d} = \begin{pmatrix} 7.5 \\ 4.5 \\ 4 \end{pmatrix}.$$

This MBT is weakly-lumpable into the equivalent representation of size 2

$$\boldsymbol{\alpha} = (0.5, 0.5), \quad \mathbf{D_0} = \begin{pmatrix} -21 & 2 \\ 2 & -31 \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} 6 & 4 & 2 & 1 \\ 12 & 8 & 4 & 1 \end{pmatrix}, \quad \mathbf{d} = \begin{pmatrix} 6 \\ 4 \end{pmatrix}.$$

The RBTMinimize function gives the minimal representation

$$\boldsymbol{\alpha} = (1, 0), \quad \mathbf{D_0} = \begin{pmatrix} -25.6667 & 1.66667 \\ 17.3333 & -26.3333 \end{pmatrix},$$
$$\mathbf{B} = \begin{pmatrix} 13.7778 & 1.55556 & 3.55556 & 0.111111 \\ 1.77778 & -0.444444 & -0.444444 & 0.111111 \end{pmatrix}, \quad \mathbf{d} = \begin{pmatrix} 5 \\ 8 \end{pmatrix}.$$

3.5. Canonical representation of MBT of order 2

To investigate the possible Markovian canonical forms of an MBT of order 2 (MBT(2)), we proceeded exactly in the same way as for non-stationary MAPs(2) and TMAPs(2), as described in Section 2.5.

The total number of parameters in an arbitrary representation $(\boldsymbol{\alpha}, \mathbf{D_0}, \mathbf{B})$ of an MBT(2) is 13 (which comes from the number of elements in the representation, $2 + 4 + 8 = 14$, reduced by one due to the constraint $\boldsymbol{\alpha}\mathbb{1} = 1$), while the number of independent parameters is 11 by Theorem 7. A canonical representation should thus have two zero entries. By considering symmetric representations only once, there is thus a total of 47 possible canonical forms.

We randomly generated about 3500 random MBTs(2) with the same method as for non-stationary MAPs(2) and TMAPs(2). We found out that there are 12 sets of 13 possible Markovian canonical forms covering all random examples. We show here only two of these sets, and we specify the zero entries of their 13 canonical forms (where, as before, the symbol "-" means "the same as in Set 1").

Structure No.	Set 1	Set 2
1	$\mathbf{d}_1, \mathbf{B}_{2,11}$	-
2	$\mathbf{d}_1, \mathbf{B}_{2,12}$	-
3	$\mathbf{d}_1, \mathbf{B}_{2,21}$	-
4	$\mathbf{d}_1, \mathbf{B}_{2,22}$	-
5	$\mathbf{d}_1, \mathbf{D}_{\mathbf{0}21}$	-
6	$\mathbf{D}_{\mathbf{0}12}, \mathbf{B}_{2,11}$	-
7	$\mathbf{D}_{\mathbf{0}12}, \mathbf{B}_{2,12}$	-
8	$\mathbf{D}_{\mathbf{0}12}, \mathbf{B}_{2,21}$	-
9	$\mathbf{D}_{\mathbf{0}12}, \mathbf{B}_{2,22}$	-
10	$\mathbf{B}_{1,11}, \mathbf{B}_{2,22}$	$\mathbf{B}_{1,11}, \mathbf{B}_{1,12}$
11	$\mathbf{B}_{1,12}, \mathbf{B}_{2,11}$	$\mathbf{B}_{1,12}, \mathbf{B}_{2,11}$
12	$\mathbf{B}_{1,12}, \mathbf{B}_{2,21}$	$\mathbf{B}_{1,21}, \mathbf{B}_{2,12}$
13	$\mathbf{B}_{1,21}, \mathbf{B}_{2,12}$	$\boldsymbol{\alpha}_1, \mathbf{B}_{1,11}.$

Compared to the case of TMAPs(2) it is an even more challenging open question therefore to find a classification of MBTs(2) into one of the 13 possible canonical forms of a given set.

4. Conclusion

Markov modulated stochastic processes have several interesting analytical and computational properties, and practical applications; however, the parameters estimation of such processes is still an open problem and a current research topic. For efficient parameter estimation we need a proper understanding of the structural properties of these processes.

MAPs are studied for several years, but the properties of other Markov modulated processes, more recently introduced, are not deeply studied yet. In this work we investigated some basic properties of TMAPs and MBTs, and we found that there are several similarities between their properties and the ones of stationary MAPs. Based on these similarities, we collected a set of structural properties, including the number of independent parameters, the set of characterizing moments, and some procedures for computing matrix representations based on the characterizing moments sets, and representations of minimal size. We also investigated the possible canonical forms of TMAPs and MBTs of order 2, and found that several different forms needs to be considered in order to represent the whole TMAP(2) and MBT(2) classes.

Acknowledgements

The first author would like to acknowledge the financial support of the Australian Research Council, grant no. DP110101663, the second author is supported by Hungarian Research Found OTKA grant no. K101150.

References

[1] G. Latouche, M.-A. Remiche, P. Taylor, Transient Markov arrival processes, Annals of Applied Probability 13 (2003) 628–640.

[2] S. Hautphenne, G. Latouche, The Markovian binary tree applied to demography, Journal of Mathematical Biology 64 (2012) 1109–1135.

[3] N. Bean, N. Kontoleon, P. Taylor, Markovian trees: properties and algorithms, Annals of Operations Research 160 (2008) 31–50.

[4] S. Hautphenne, G. Latouche, M.-A. Remiche, Transient features for Markovian binary trees, in: Proceedings of the Fourth International ICST Conference on Performance Evaluation Methodologies and Tools, Berlin, Germany.

[5] S. Hautphenne, G. Latouche, M.-A. Remiche, Algorithmic approach to the extinction probability of branching processes, Methodology and Computing in Applied Probability 13 (2011) 172–192.

[6] N. Kontoleon, The Markovian binary tree: a model of the macroevolutionary process, Technical Report, Phd Thesis, The University of Adelaide, 2005.

[7] M. Telek, G. Horváth, A minimal representation of Markov arrival processes and a moments matching method, Performance Evaluation 64 (2007) 1153–1168.

[8] L. Bodrog, A. Horváth, M. Telek, Moment characterization of matrix exponential and Markovian arrival processes, Annals of Operations Research 160 (2008) 51–68.

[9] P. Buchholz, M. Telek, On minimal representation of rational arrival processes, Annals of Operations Research (2011). DOI 10.1007/s10479-011-1001-5.

[10] L. Bodrog, A. Heindl, G. Horvath, M. Telek, A Markovian canonical form of second-order matrix-exponential processes, European journal of operational research 190 (2008) 459–477.

[11] S. Asmussen, O. Nerman, M. Olsson, Fitting phase-type distributions via the EM algorithm, Scandinavian Journal of Statistics 23 (1996) 419–441.

[12] L. Breuer, Parameter estimation for a class of BMAPs, in: Advances in Algorithmic Methods for Stochastic Models, Notable Publications, New Jersey, 2000, pp. 87–97.

[13] L. Breuer, A. Kume, An EM algorithm for Markovian arrival processes observed at discrete times, in: Measurement, Modelling, and Evaluation of Computing Systems and Dependability and Fault Tolerance, Lecture Notes in Computer Science, 2010, pp. 242–258.

[14] L. Breuer, G. Latouche, M.-A. Remiche, An EM algorithm for the model fitting of a class of transient Markovian arrival processes, Technical Report, Department of Informatique, Universitite libre de Bruxelles, Belgium, 2003.

[15] L. Breuer, An EM algorithm for batch Markovian arrival processes and its comparison to a simpler estimation procedure, Annals of Operations Research 112 (2002) 123–138.

[16] S. Hautphenne, M. Fackrell, An EM algorithm for the model fitting of Markovian binary trees (2012). Submitted for publication.

[17] M. Telek, Some structural properties of Markov and rational arrival processes, Stochastic Models 27 (2011) 723–746.

[18] A. van de Liefvoort, The moment problem for continuous distributions, Technical Report WP-CM-1990-02, School of Computing and Engineering, University of Missouri – Kansas City, USA, 1990.

[19] L. Bodrog, P. Buchholz, A. Heindl, A. Horváth, G. Horváth, I. Kolossváry, Z. Németh, P. Reinecke, M. Telek, M. Vécsei, Butools: Program packages for computations with PH, ME distributions and MAP, RAP processes, http://webspn.hit.bme.hu/~butools, 2011.