Int. J. Appl. Math. Comput. Sci., , Vol. , No. , –
DOI:

FITTING TRAFFIC TRACES WITH DISCRETE CANONICAL PHASE TYPE DISTRIBUTIONS AND MARKOV ARRIVAL PROCESSES

ANDRÁS MÉSZÁROS [*,***], JÁNOS PAPP [*], MIKLÓS TELEK [*,**]

[*] Budapest University of Techonolgy and Economics

[**]MTA-BME Information systems research group

[***]Inter-University Center for Telecommunications and Informatics Debrecen

e-mail: (meszarosa,telek)@hit.bme.hu,papp.janos.90@gmail.com

Recent developments of matrix analytic methods make phase type distributions (PHs) and Markov arrival processes (MAPs) promising stochastic model candidates for capturing traffic trace behaviour and for efficient usage in queueing analysis. After introducing the basics of these sets of stochastic models the paper discusses the following subjects in details:

1. PHs and MAPs have different representations. For the efficient use of these models sparse (defined by minimal number of parameters) and unique representations of discrete time PHs and MAPs are needed, which are commonly referred to as *canonical representations*. The paper presents new results on the canonical representation of discrete PHs and MAPs.
2. The canonical representation allows a direct mapping between experimental moments and the stochastic models, referred to as moments *matching*. Explicit procedures are provided for this mapping.
3. Moments matching is not always the best way to model the behavior of traffic traces. Model *fitting* based on appropriately chosen distance measures might result better performing stochastic models. We also demonstrate the efficiency of fitting procedures with experimental results.

Keywords: Fitting traffic traces, discrete phase type distribution, discrete Markov arrival process, canonical representation.

1. Introduction

Stochastic models with underlying Markov chains are known for being flexible in modelling general stochastic behaviour and for allowing efficient numerical analysis through matrix analytic methods (Neuts, 1981). These nice properties make phase type distributions (PHs) and Markov arrival processes (MAPs) promising candidates for modelling the traffic load of computer and communication systems.

For a period of time continuous time stochastic models were more often applied in performance modelling of computer and communication systems. Later on, with the rise of slotted time telecommunication protocols (e.g. ATM) discrete time models became primary modelling tools (for a recent surveys see (Alfa, 2002; Lakatos *et al*., 2013)). In this paper we focus on discrete time models and present some results whose continuous time counterparts are already available. It turns out that discrete time models with strictly positive eigenvalues are practically identical with their the continuous time counterparts, but the discrete time models containing also negative eigenvalues pose new problems.

One main problem of PHs and MAPs is the non-uniqueness and over-parametrization of their general matrix form (see e.g. (Telek and Horváth, 2007a) for more details). Specifically, there are descriptions with minimal number of parameters for describing these processes, but those descriptions are hard to use in practice because they do not indicate the feasibility of the associated stochastic model (for example the moments of a random variable of a given class might define the random variable fully, but it is not easy to check if a set of moments is feasible, i.e., if there exists a random variable in the given class with those moments). On the other hand over-parametrised matrix descriptions give a direct mapping to Markov chains, which ensures the feasibility of the model, however the over-parametrization causes significant problems in fitting methods. The above obstacle can be eliminated by finding unique matrix representations with minimal number of parameters. These representations are referred to as canonical representations. Apart from the benefits in fitting methods canonical representations also enable parameter matching, i.e., a direct mapping of important traffic parameters (moments, autocorrelation) to these models.

In this paper we present new results on the canonical representation of order 2 and order 3 discrete PHs (DPH(2) and DPH(3)) as well as on order 2 discrete MAPs (DMAP(2)). We provide explicit formulas for parameter matching using these canonical forms, give moments and correlation bounds for these models, and show their efficiency in fitting through numerical examples.

The rest of the paper is organized as follows. In Section 2 we survey the necessary definitions and essential properties of existing Markov chain driven stochastic processes and their non-Markovian generalizations. The discussion of canonical forms for DPH(2), DPH(3) and DMAP(2) can be found in Section 3, 4, and 5 respectively. Section 2 gives formulas for parameter matching. Section 6 presents moments based matching methods for approximating discrete PH and MAP. The numerical examples for trace fitting are presented in Section 7. Section 8 concludes the paper.

2. Markov chain driven point processes and their non-Markovian generalizations

The following subsections summarize the main properties of simple stochastic models with a background discrete state Markov chain and their non-Markovian generalizations. If the background chain is a discrete time Markov chain we obtain discrete (time) stochastic models and if it is a continuous time Markov chain we obtain continuous (time) stochastic models. The main focus of the paper is on discrete models, but some results are related to their continuous counterparts, thus we introduce both of them.

2.1. Discrete phase type and matrix geometric distributions. The following stochastic models define discrete distributions on the positive integers.

Definition 1. Let $\mathcal{X}$ be a discrete random variable on $\mathbb{N}^+$ with probability mass function (pmf)

$$P_{\mathcal{X}}(i) = Pr(\mathcal{X} = i) = \alpha \boldsymbol{A}^{i-1}(\mathbb{1} - \boldsymbol{A}\mathbb{1}) \ \forall i \in \mathbb{N}^+, \quad (1)$$

where α is a row vector of size n, $\boldsymbol{A}$ is a square matrix of size $n \times n$, and $\mathbb{1}$ is the column vector of ones of size n. If the pmf has this matrix geometric form, then we say that $\mathcal{X}$ is matrix geometrically distributed with representation $(\alpha, \boldsymbol{A})$, or shortly, MG$(\alpha, \boldsymbol{A})$ distributed.

The size of $\boldsymbol{A}$ is also referred to as the order of the associated distribution. In this and the subsequent models scalar quantities are obtained as a product of a row vector, a given number of square matrices and a column vector. In the sequel we refer to the row vector as initial vector and to the column vector as closing vector. It is an important consequence of Definition 1 that α and $\boldsymbol{A}$ have to be such that (1) is non-negative.

Definition 2. If $\mathcal{X}$ is an MG$(\alpha, \boldsymbol{A})$ distributed random variable, where α and $\boldsymbol{A}$ have the following properties:

- $\alpha_i \geq 0$,
- $A_{ij} \geq 0$, $\boldsymbol{A}\mathbb{1} \leq \mathbb{1}$,

then we say that $\mathcal{X}$ is discrete phase type distributed with representation $(\alpha, \boldsymbol{A})$, or shortly, DPH$(\alpha, \boldsymbol{A})$ distributed.

The vector-matrix representations satisfying the conditions of Definition 2 are called Markovian.

In this paper we focus on distributions on the positive integers, consequently, $\alpha\mathbb{1} = 1$. The cumulative density function (cdf), the moment generating function, and the factorial moments of $\mathcal{X}$ are

$$F_{\mathcal{X}}(i) = Pr(\mathcal{X} \leq i) = 1 - \alpha \boldsymbol{A}^i \mathbb{1}, \quad (2)$$

$$\begin{aligned} f_n &= E(\mathcal{X}(\mathcal{X}-1)\dots(\mathcal{X}-n+1)) \\ &= n!\alpha(\boldsymbol{I} - \boldsymbol{A})^{-n}\boldsymbol{A}^{n-1}\mathbb{1}. \end{aligned} \quad (3)$$

A DPH has infinitely many different Markovian and non-Markovian representations (matrix-vector pairs, that fulfill (1)). One way to get a different representation of a DPH$(\boldsymbol{\alpha}, \boldsymbol{A})$ with the same size is the application of the similarity transformation

$$\boldsymbol{B} = \boldsymbol{T}^{-1}\boldsymbol{A}\boldsymbol{T}, \quad \beta = \alpha \boldsymbol{T}, \quad (4)$$

where $\boldsymbol{T}$ is an arbitrary non-singular matrix for which $\boldsymbol{T}\mathbb{1} = \mathbb{1}$. If a DPH has an $(\alpha, \boldsymbol{A})$ Markovian representation, for which $\boldsymbol{A}$ is upper triangular, we call the distribution acyclic DPH (shortly ADPH) distribution, and the specific representation an ADPH representation.

2.2. Discrete Markov arrival processes and discrete rational arrival processes. Let $\mathcal{X}(t)$ be a point process on $\mathbb{N}^+$ with joint probability mass function of inter-event times $P_{\mathcal{X}(t)}(x_0, x_1, \dots, x_k)$ for $k = 1, 2, \dots$ and $x_0, \dots, x_k \in \mathbb{N}^+$.

Definition 3. $\mathcal{X}(t)$ is called a rational arrival process if there exists a finite $(\boldsymbol{H_0}, \boldsymbol{H_1})$ square matrix pair such that $(\boldsymbol{H_0} + \boldsymbol{H_1})\mathbb{1} = \mathbb{1}$,

$$\underline{\pi}(\boldsymbol{I} - \boldsymbol{H_0})^{-1}\boldsymbol{H_1} = \underline{\pi}, \ \ \underline{\pi}\mathbb{1} = 1 \quad (5)$$

has a unique solution, and

$$\begin{aligned} &P_{\mathcal{X}(t)}(x_0, x_1, \dots, x_k) = \\ &= \underline{\pi}\boldsymbol{H_0}^{x_0-1}\boldsymbol{H_1}\boldsymbol{H_0}^{x_1-1}\boldsymbol{H_1}\dots\boldsymbol{H_0}^{x_k-1}\boldsymbol{H_1}\mathbb{1}, \end{aligned} \quad (6)$$

In this case we say that $\mathcal{X}(t)$ is a discrete rational arrival process with representation $(\boldsymbol{H_0}, \boldsymbol{H_1})$, or shortly, DRAP$(\boldsymbol{H_0}, \boldsymbol{H_1})$.

The size of the $\boldsymbol{H_0}$ and $\boldsymbol{H_1}$ matrices is also referred to as the order of the associated process. For the sake of conciseness we will denote order n MGs, DPHs, DRAPs etc. by MG(n), DPH(n), DRAP(n) etc. respectively.

An important consequence of Definition 3 is that $\boldsymbol{H_0}$ and $\boldsymbol{H_1}$ have to be such that (6) is always non-negative.

Definition 4. If $\mathcal{X}(t)$ is a DRAP$(\boldsymbol{H_0}, \boldsymbol{H_1})$, where $\boldsymbol{H_0}$ and $\boldsymbol{H_1}$ are non-negative, we say that $\mathcal{X}(t)$ is a Discrete Markov arrival process with representation $(\boldsymbol{H_0}, \boldsymbol{H_1})$, or shortly, DMAP$(\boldsymbol{H_0}, \boldsymbol{H_1})$.

The matrix pairs satisfying the conditions of Definition 4 are called Markovian and the matrix pairs violating Definition 4 are called non-Markovian.

Definition 5. The correlation parameter γ of a DRAP$(\boldsymbol{H_0}, \boldsymbol{H_1})$ is the eigenvalue of $(\boldsymbol{I}-\boldsymbol{H_0})^{-1}\boldsymbol{H_1}$ with the second largest absolute value.
One of the eigenvalues of $(\boldsymbol{I}-\boldsymbol{H_0})^{-1}\boldsymbol{H_1}$ is 1, because $(\boldsymbol{H_0}+\boldsymbol{H_1})\mathbb{1} = \mathbb{1}$, and the other eigenvalues are on the unit disk. If γ is real, it is between -1 and 1. This parameter is especially important in case of order 2 DRAPs, as their ρ_k lag-k autocorrelation coefficient can be given as $\rho_k = \gamma^k c_0$, where c_0 depends only on the stationary inter-arrival time distribution of the process.

Similar to DPHs a DMAP has infinitely many different Markovian and non-Markovian representations (matrix pairs that fulfil (6)). One way to get a different representation of a DMAP$(\boldsymbol{D_0}, \boldsymbol{D_1})$ with the same size is the application of the similarity transformation

$$\boldsymbol{H_0} = \boldsymbol{T}^{-1}\boldsymbol{D_0}\boldsymbol{T}, \quad \boldsymbol{H_1} = \boldsymbol{T}^{-1}\boldsymbol{D_1}\boldsymbol{T}, \tag{7}$$

where $\boldsymbol{T}$ is an arbitrary non-singular matrix for which $\boldsymbol{T}\mathbb{1} = \mathbb{1}$.

The (stationary) marginal distribution of the inter-event time of DRAP$(\boldsymbol{H_0}, \boldsymbol{H_1})$ is MG$(\pi, \boldsymbol{H_0})$, where π is the unique solution of (5). Similarly the (stationary) marginal distribution of the inter-event time of DMAP$(\boldsymbol{H_0}, \boldsymbol{H_1})$ is DPH$(\pi, \boldsymbol{H_0})$, where π is the unique solution of (5).

2.3. Continuous phase type and matrix exponential distributions.

The continuous counterparts of the above introduced models are defined as follows.

Definition 6. Let $\mathcal{X}$ be a continuous random variable with support on $\mathbb{R}^+$ and cumulative distribution function (cdf)

$$F_X(x) = Pr(\mathcal{X} < x) = 1 - \alpha e^{\boldsymbol{A}x}\mathbb{1}, \tag{8}$$

where α is a row vector of size n, $\boldsymbol{A}$ is a square matrix of size $n \times n$, and $\mathbb{1}$ is the column vector of ones of size n. In this case, we say that $\mathcal{X}$ is matrix exponentially distributed with representation $(\alpha, \boldsymbol{A})$, or shortly, ME$(\alpha, \boldsymbol{A})$ distributed.

Definition 7. If $\mathcal{X}$ is an ME$(\alpha, \boldsymbol{A})$ distributed random variable, where α and $\boldsymbol{A}$ have the following properties:

- $\alpha_i \geq 0$, $\alpha\mathbb{1} = 1$ (there is no probability mass at $x = 0$),
- $A_{ii} < 0$, $A_{ij} \geq 0$ for $i \neq j$, $\boldsymbol{A}\mathbb{1} \leq 0$,

we say that $\mathcal{X}$ is phase type distributed with representation $(\alpha, \boldsymbol{A})$, or shortly, CPH$(\alpha, \boldsymbol{A})$ distributed.

The vector-matrix representations satisfying the conditions of Definition 7 are called Markovian.

The probability density function (pdf), the Laplace transform, and the moments of $\mathcal{X}$ are

$$f_{\mathcal{X}}(x) = -\alpha e^{\boldsymbol{A}x}\boldsymbol{A}\mathbb{1}, \tag{9}$$

$$\mu_n = E(\mathcal{X}^n) = n!\alpha(-\boldsymbol{A})^{-n}\mathbb{1}. \tag{10}$$

2.4. Continuous Markov arrival process and continuous rational arrival process.

Let $\mathcal{X}(t)$ be a point process on $\mathbb{R}^+$ with joint probability density function of inter-event times $f(x_0, x_1, \ldots, x_k)$ for $k = 1, 2, \ldots$.

Definition 8. $\mathcal{X}(t)$ is called a rational arrival process if there exists a finite $(\boldsymbol{H_0}, \boldsymbol{H_1})$ square matrix pair such that $(\boldsymbol{H_0}+\boldsymbol{H_1})\mathbb{1} = 0$,

$$\underline{\pi}(-\boldsymbol{H_0})^{-1}\boldsymbol{H_1} = \underline{\pi}, \ \ \underline{\pi}\mathbb{1} = 1\ , \tag{11}$$

has a unique solution, and

$$f(x_0, x_1, \ldots, x_k) = \\ = \underline{\pi} e^{\boldsymbol{H_0}x_0}\boldsymbol{H_1} e^{\boldsymbol{H_0}x_1}\boldsymbol{H_1} \ldots e^{\boldsymbol{H_0}x_k}\boldsymbol{H_1}\mathbb{1}. \tag{12}$$

In this case we say that $\mathcal{X}(t)$ is a rational arrival process with representation $(\boldsymbol{H_0}, \boldsymbol{H_1})$, or shortly, RAP$(\boldsymbol{H_0}, \boldsymbol{H_1})$.

Definition 9. If $\mathcal{X}(t)$ is a RAP$(\boldsymbol{H_0}, \boldsymbol{H_1})$, where $\boldsymbol{H_0}$ and $\boldsymbol{H_1}$ have the following properties:

- $\boldsymbol{H_1}_{ij} \geq 0$,
- $\boldsymbol{H_0}_{ii} < 0$, $\boldsymbol{H_0}_{ij} \geq 0$ for $i \neq j$, $\boldsymbol{H_0}\mathbb{1} \leq 0$,

we say that $\mathcal{X}(t)$ is a Markov arrival process with representation $(\boldsymbol{H_0}, \boldsymbol{H_1})$, or shortly, MAP$(\boldsymbol{H_0}, \boldsymbol{H_1})$.

Similar to the discrete case, the representations satisfying the conditions of Definition 9 are called Markovian, and similarity transformations generate different representations of the same process.

3. Canonical form of order 2 DPH distributions

In this section we provide a canonical form for DPH(2) distributions. We start with characterizing the properties of all possible MG(2) distributions, i.e., distributions of form (1), where $\boldsymbol{A}$ is a 2×2 matrix. Using this characterization we prove that all MG(2) distributions (thus all order 2 DPH distributions) have a Markovian canonical form. After that we present the exact transformation method.

3.1. Canonical form of DPH(2).

Theorem 1. *An MG(2) distribution has one of the following two forms*

- *different eigenvalues:*

$$p_i = Pr(\mathcal{X} = i) = a_1 s_1^{i-1} + a_2 s_2^{i-1}, \qquad (13)$$

where s_1, s_2 are the eigenvalues of $\boldsymbol{A}$. These eigenvalues are real with $0 < s_1 < 1$, $s_1 > |s_2|$. Moreover a_1, a_2 are such that $a_1 \le \frac{(1-s_1)(1-s_2)}{s_1-s_2}$ and $a_2 = (1-s_2)\left(1 - \frac{a_1}{1-s_1}\right)$, furthermore $a_1 > 0$ if $s_2 \ge 0$ and $a_1 \ge \frac{s_2(1-s_1)(1-s_2)}{s_2(1-s_2)-s_1(1-s_1)}$ if $s_2 < 0$;

- *identical eigenvalues:*

$$p_i = Pr(\mathcal{X} = i) = (a_1(i-1) + a_2)s^{i-1}, \qquad (14)$$

where s is the double eigenvalue of $\boldsymbol{A}$. This eigenvalue is real with $0 < s < 1$. Furthermore a_1, a_2 are such that $0 < a_1 \le \frac{(1-s)^2}{s}$ and $a_2 = \frac{(1-s)^2 - a_1 s}{1-s}$.

A vector matrix representation of the first form is

$$\alpha = \left[\frac{a_1}{1-s_1}, \frac{a_2}{1-s_2}\right], \boldsymbol{A} = \begin{bmatrix} s_1 & 0 \\ 0 & s_2 \end{bmatrix}, \qquad (15)$$

and of the second form is

$$\alpha = \left[\frac{a_1}{1-s}, \frac{a_2(1-s) - a_1(1-2s)}{(1-s)^2}\right], \boldsymbol{A} = \begin{bmatrix} s & s \\ 0 & s \end{bmatrix}. \qquad (16)$$

Proof. The first form covers the cases when the s_1, s_2 eigenvalues of $\boldsymbol{A}$ are different and the second one when the eigenvalues are identical ($s_1 = s_2 = s$). We discuss these cases separately.

- different eigenvalues:

  First we show that the eigenvalues are real. Assume, that $\boldsymbol{A}$ has a complex eigenvalue. In this case the other eigenvalue has to be its complex conjugate and a_1 and a_2 must be conjugates too to obtain real $p_i = a_1 s_1^{i-1} + a_2 s_2^{i-1}$ values. Let φ be the argument of a_1 ($a_1 = |a_1|e^{i\varphi}$), and ψ the argument of s_1. Moreover assume that $\psi \in (0, \pi)$. From $i = 1$ we get that $\varphi \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$. Now consider the case $i = \lceil \frac{\pi}{\psi} \rceil + 1$. The argument of $a_1 s_1^{i-1}$ is $\varphi + (i-1)\psi$, and it is in $\left[\frac{\pi}{2}, \frac{3\pi}{2}\right]$. This means that p_i is negative since $a_1 s_1^{i-1}$ and $a_2 s_2^{i-1}$ are conjugates. Thus we get that the eigenvalues are real.

  The two real eigenvalues have to be such that the one with the larger absolute value (s_1) is positive, because it becomes dominant for large i values and p_i would become negative for large i values with negative dominant eigenvalue. Additionally the dominant eigenvalue has to be less than one to ensure that the p_i series has finite sum.

  The relation of the a_1, a_2 coefficients is obtained from $\sum_i p_i = 1$. The $a_1 > 0$ bound of a_1 for the $s_2 \ge 0$ case comes from the fact that $p_i \sim a_1 s_1^{i-1}$ for large i, where s_1 is positive. A negative a_1 would result in negative p_i for large i. If $s_2 < 0$ this is not enough, since p_i can still be negative for smaller i values, if a_2 is sufficiently large. In this case the lower bound for a_1 comes from $p_2 \ge 0$, as

$$\begin{aligned} 0 &\le p_2 = a_1 s_1 + a_2 s_2 \\ 0 &\le a_1 s_1 + (1-s_2)\left(1 - \frac{a_1}{1-s_1}\right) s_2 \\ 0 &\le a_1 \frac{s_2(1-s_2) - s_1(1-s_1)}{1-s_1} + s_2(1-s_2) \\ a_1 &\ge \frac{s_2(1-s_1)(1-s_2)}{s_1(1-s_1) - s_2(1-s_2)} \end{aligned} \qquad (17)$$

  The upper bound of a_1 can be derived from $p_1 \ge 0$, since

$$\begin{aligned} 0 &\le p_1 = a_1 + a_2 \\ 0 &\le a_1 + (1-s_2)\left(1 - \frac{a_1}{1-s_1}\right) \\ 0 &\le a_1 \frac{s_2 - s_1}{1-s_1} + (1-s_2) \\ a_1 &\le \frac{(1-s_1)(1-s_2)}{s_1 - s_2} \end{aligned} \qquad (18)$$

- identical eigenvalues:

  First we show that the eigenvalue is real and non-negative. If s is complex or negative in (14) then $p_i \sim a_1(i-1)s^{i-1}$ for large i, which becomes complex or negative, respectively, for any a_1 in case of two consecutive large i values.

  The inequality $s < 1$ comes from the fact that the p_i series has finite sum.

  Similar to the previous case, the relation of the a_1, a_2 coefficients is obtained from $\sum_i p_i = 1$ and the $a_1 >$

0 bound comes from the fact that $p_i \sim a_1(i-1)s^{i-1}$ for large i, where s is positive. A negative a_1 would result in negative p_i for large i. The upper bound of a_1 comes from $p_1 \geq 0$, since

$$0 \leq p_1 = a_2 \tag{19}$$

$$0 \leq \frac{(1-s)^2 - a_1 s}{1-s} \tag{20}$$

$$a_1 \leq \frac{(1-s)^2}{s} \tag{21}$$

■

Theorem 2. *If $\mathcal{X}$ is $MG(2)$ distributed with two distinct positive eigenvalues ($0 < s_2 < s_1 < 1$) then it can be represented as ADPH($\alpha, \boldsymbol{A}$), where*

$$\alpha=\left[\frac{a_1(s_1-s_2)}{(1-s_1)(1-s_2)}, \frac{a_1+a_2}{1-s_2}\right], \boldsymbol{A}=\begin{bmatrix} s_1 & 1-s_1 \\ 0 & s_2 \end{bmatrix}.$$

Proof. The $(\alpha, \boldsymbol{A})$ vector-matrix pair is such that $p_i = \alpha \boldsymbol{A}^{i-1}(\mathbb{1}-\boldsymbol{A}\mathbb{1}) = a_1 s_1^{i-1} + a_2 s_2^{i-1}$. Matrix $\boldsymbol{A}$ obviously satisfies the conditions of Definition 2 when $0 < s_2 < s_1 < 1$. It remains to be shown that α is non-negative when $0 < s_2 < s_1 < 1$, $0 < a_1$, and $p_1 \geq 0$. In the first element of α we have $a_1 > 0$, $s_1 - s_2 > 0$, $s_1 - 1 < 0$, $s_2 - 1 < 0$, from which it is positive. In the second element we have $a1 + a2 = p_1 \geq 0$ and $1 - s_2 > 0$. Note that $\alpha\mathbb{1} = 1$ when $a_2 = (1-s_2)\left(1 - \frac{a_1}{1-s_1}\right)$. ■

Theorem 3. *If $\mathcal{X}$ is $MG(2)$ distributed with a dominant positive and a negative eigenvalue ($s_2 < 0 < s_1 < 1$ and $s_1 + s_2 > 0$), then it can be represented as DPH($\alpha, \boldsymbol{A}$), where*

$$\alpha = \left[\frac{a_1 s_1 + a_2 s_2}{(1-s_1)(1-s_2)}, \frac{(a_1+a_2)(1-s_1-s_2)}{(1-s_1)(1-s_2)}\right],$$

$$\boldsymbol{A} = \begin{bmatrix} 1-\beta_1 & \beta_1 \\ \beta_2 & 0 \end{bmatrix},$$

$\beta_1 = 1 - s_1 - s_2$ *and* $\beta_2 = \frac{s_1 s_2}{s_1+s_2-1}$.

Proof. The eigenvalues of $\boldsymbol{A}$ are s_1, s_2 and the $(\alpha, \boldsymbol{A})$ pair is such that $p_i = \alpha\boldsymbol{A}^{i-1}(\mathbb{1} - \boldsymbol{A}\mathbb{1}) = a_1 s_1^{i-1} + a_2 s_2^{i-1}$.

Parameters β_1 and β_2 are positive and less than 1 from which matrix $\boldsymbol{A}$ satisfies the conditions of Definition 2.

It remains to show that α is non-negative when $s_2 < 0 < s_1 < 1$, $1 > s_1 > s_1 + s_2 > 0$ and $p_1, p_2 \geq 0$. For the first element of α we have $a_1 s_1 + a_2 s_2 = p_2 \geq 0$, $s_1 - 1 < 0$, $s_2 - 1 < 0$, from which it is positive and for the numerator of the second element we have $a_1 + a_2 = p_1 \geq 0$ and $1 - s_1 - s_2 > 0$. The denominator of the second element is the same as that of the first one, thus the second element of α is also non-negative. ■

Theorem 4. *If $\mathcal{X}$ is $MG(2)$ distributed with two identical eigenvalues ($0 < s = s_2 = s_1 < 1$) then it can be represented as ADPH($\alpha, \boldsymbol{A}$), where*

$$\alpha = \left[\frac{a_1 s}{(1-s)^2}, \frac{a_2}{1-s}\right], \quad \boldsymbol{A} = \begin{bmatrix} s & 1-s \\ 0 & s \end{bmatrix}.$$

Proof. The $(\alpha, \boldsymbol{A})$ vector-matrix pair is such that $p_i = \alpha\boldsymbol{A}^{i-1}(\mathbb{1} - \boldsymbol{A}\mathbb{1}) = (a_1(i-1) + a_2)s^{i-1}$, and matrix $\boldsymbol{A}$ satisfies the conditions of Definition 2 when $0 < s < 1$.

It remains to show that α is non-negative when $0 < s < 1$, $0 < a_1$ and $p_1 \geq 0$. All terms of the elements of α are non-negative since $a_2 = p_1 \geq 0$. ■

Theorem 2 – 4 have the following consequences.

Corollary 1. *The vector-matrix representations in Theorem 2 – 4 can be used as canonical representations of DPH(2) and MG(2) distributions.*

Corollary 2.

order 2 DPH $\equiv$ order 2 MG

order 2 ADPH $\equiv$ order 2 MG with positive eigenvalues

Corollary 3. *If the eigenvalues of the order 2 MG($\gamma, \boldsymbol{G}$) are positive and its canonical representation is ADPH($\alpha, \boldsymbol{A}$), then ME($\gamma, \boldsymbol{G} - \boldsymbol{I}$) is a matrix exponential distribution whose canonical ACPH representation (Cumani's canonical form) is ACPH($\alpha, \boldsymbol{A} - \boldsymbol{I}$).*

Proof. The matrix of the canonical representation ADPH($\alpha, \boldsymbol{A}$) has the form $\begin{bmatrix} s_1 & 1-s_1 \\ 0 & s_2 \end{bmatrix}$, where $1 > s_1 \geq s_2 > 0$. Consequently $\boldsymbol{A} - \boldsymbol{I}$ is a matrix of an ACPH distribution in Cumani's canonical form with eigenvalues $0 > s_1 - 1 \geq s_2 - 1 > -1$.

Furthermore, due to the fact that ME($\gamma, \boldsymbol{G} - \boldsymbol{I}$) and ACPH($\alpha, \boldsymbol{A} - \boldsymbol{I}$) represent the same distribution ME($\gamma, \boldsymbol{G} - \boldsymbol{I}$) is a valid ME distribution. ■

3.2. Transformation of DPH(2) to canonical form.

The introduced canonical representations can be obtained from a general vector-matrix representation with the following similarity transformation.

Corollary 4. *If the eigenvalues of the order 2 MG($\gamma, \boldsymbol{G}$) are $0 < s_2 < s_1 < 1$, then its canonical representation is ADPH($\alpha = \gamma\boldsymbol{B}, \boldsymbol{A} = \boldsymbol{B}^{-1}\boldsymbol{G}\boldsymbol{B}$), where matrix $\boldsymbol{B}$ is composed by column vectors $b_1 = \mathbb{1} - b_2$ and $b_2 = \frac{1}{1-s_2}(\mathbb{1} - \boldsymbol{G}\mathbb{1})$.*

Proof. Matrix $\boldsymbol{B}$ is obtained as the solution of $\boldsymbol{B}\mathbb{1} = b_1 + b_2 = \mathbb{1}$ and $\boldsymbol{G}\boldsymbol{B} = \boldsymbol{B}\begin{bmatrix} s_1 & 1-s_1 \\ 0 & s_2 \end{bmatrix}$, whose column vector form is $\boldsymbol{G}b_1 = s_1 b_1$ and $\boldsymbol{G}b_2 = (1-s_1)b_1 + s_2 b_2$. Consequently, $\boldsymbol{A} = \begin{bmatrix} s_1 & 1-s_1 \\ 0 & s_2 \end{bmatrix}$. ■

The proofs for the subsequent corollaries in this section follow the same pattern and are omitted.

Corollary 5. *If the eigenvalues of the order 2 MG($\gamma, \boldsymbol{G}$) are $s_2 < 0 < s_1 < 1$, then its canonical representation is ADPH($\gamma\boldsymbol{B}$, $\begin{bmatrix} s_1+s_2 & 1-s_1-s_2 \\ \frac{s_1 s_2}{s_1+s_2-1} & 0 \end{bmatrix}$), where matrix $\boldsymbol{B}$ is composed of column vectors $b_1 = \mathbb{1} - b_2$ and $b_2 = \frac{1-s_1-s_2}{(1-s_1)(1-s_2)}(\mathbb{1} - \boldsymbol{G}\mathbb{1})$.*

Corollary 6. *If the eigenvalues of the order 2 MG($\gamma, \boldsymbol{G}$) are $s = s_1 = s_2 < 1$ then its canonical representation is ADPH($\gamma\boldsymbol{B}$, $\begin{bmatrix} s & 1-s \\ 0 & s \end{bmatrix}$), where matrix $\boldsymbol{B}$ is composed of column vectors $b_1 = \mathbb{1} - b_2$ and $b_2 = \frac{1}{1-s}(\mathbb{1} - \boldsymbol{G}\mathbb{1})$.*

The presented similarity transformations can be used as transformation methods to compute the canonical representation from a general (Markovian or non-Markovian) vector matrix representation. As an example a simple implementation of Corollary 4 is presented in Figure 1.

```
procedure CanonicalDPH-PP(γ, G)
    [s1, s2] = eig(G);
    e = [1, 1];
    b2 = 1/(1 - s2) * (e - G * e);
    b1 = e - b2;
    return (γ * [b1, b2], [s1  1 - s1; 0  s2])
end procedure
```

Fig. 1. Canonical order 2 DPH representation based on Corollary 4

4. Canonical form of order 3 DPH distributions

In the previous section we proved that the whole MG(2) class can be represented with Markovian vector matrix pairs. That is why we started with the characterization of the order 2 MG class. For order 3 distributions the same does not hold, that is DPH(3) $\not\equiv$ MG(3). Due to this difference we follow a different approach here and show only that transformation with a given similarity matrix results in a Markovian canonical form for all DPH(3).

Similar to the order 2 case the canonical representations of DPH(3) distributions are classified according to the eigenvalue structure of the distribution. We encode the eigenvalues in decreasing absolute value and denote the ones with negative real part by N and the ones with non-negative real part by P. For example PNP means that $1 \geq |s_1| \geq |s_2| \geq |s_3|$ and $\mathrm{Re}(s_1) \geq \mathrm{Re}(s_3) \geq 0 > \mathrm{Re}(s_2)$, where s_i, $i = 1,2,3$ denote the eigenvalues. Due to the fact that the eigenvalue with the largest absolute value (dominant eigenvalue) has to be real and positive (to ensure positive probabilities in (1) for large i) we have the following cases: PPP, PPN, PNP, PNN. Complex (conjugate) eigenvalues can occur only in case of PPP and PNN.

4.1. Case PPP.

Following the pattern of Corollary 3 we define the canonical form in the PPP case based on the canonical representation of CPH(3) distributions.

Theorem 5. *If the eigenvalues of the order 3 DPH($\gamma, \boldsymbol{G}$) are all non-negative we define the canonical form as follows. The vector-matrix pair ($\gamma, \boldsymbol{G} - \boldsymbol{I}$) defines a CPH(3). Let ($\alpha, \boldsymbol{A}$) be the canonical representation of CPH($\gamma, \boldsymbol{G} - \boldsymbol{I}$) as defined in (Horváth and Telek, 2009). The canonical representation of DPH($\gamma, \boldsymbol{G}$) is ($\alpha, \boldsymbol{A}+\boldsymbol{I}$).*

Proof. The complete proof of the theorem requires the introduction of the procedure defined in (Horváth and Telek, 2009). Here we only demonstrate the result for the case when the canonical representation of CPH($\gamma, \boldsymbol{G} - \boldsymbol{I}$) is acyclic. When the eigenvalues of $\boldsymbol{G}$ are $1 > s_1 \geq s_2 \geq s_3 > 0$ the eigenvalues of $\boldsymbol{G} - \boldsymbol{I}$ are $0 > s_1 - 1 \geq s_2 - 1 \geq s_3 - 1 > -1$. In this case the matrix of the acyclic canonical form of CPH($\gamma, \boldsymbol{G} - \boldsymbol{I}$) is $\boldsymbol{A} = \begin{bmatrix} s_3 - 1 & 0 & s^* = 0 \\ 1 - s_2 & s_2 - 1 & 0 \\ 0 & 1 - s_1 & s_1 - 1 \end{bmatrix}$ and the associated vector α is non-negative. Finally, $\boldsymbol{A} + \boldsymbol{I}$ is non-negative and the associated exit probability vector, $\mathbb{1} - \boldsymbol{A}\mathbb{1} = [1 - s_3, 0, 0]^T$, is non-negative as well.

In the general case s^* might be positive and $s_i - 1$, $i = 1,2,3$ are not the eigenvalues of $\boldsymbol{A}$, but also in that case it holds that the elements of $\boldsymbol{A} + \boldsymbol{I}$ and $\mathbb{1} - \boldsymbol{A}\mathbb{1}$ are non-negative. ■

The rest of the cases require the introduction of new canonical structures.

4.2. Case PPN.

Theorem 6. *If the eigenvalues of the order 3 DPH($\gamma, \boldsymbol{G}$) are $1 > |s_1| \geq |s_2| \geq |s_3|$ and $\mathrm{Re}(s_1) \geq \mathrm{Re}(s_2) > 0 > \mathrm{Re}(s_3)$, then its canonical representation is DPH($\gamma\boldsymbol{B}, \boldsymbol{A}$), where*

$$\boldsymbol{A} = \begin{bmatrix} x_1 & 1-x_1 & 0 \\ 0 & x_2 & 1-x_2 \\ 0 & x_3 & 0 \end{bmatrix},$$

$x_1 = s_1$, $x_2 = s_2 + s_3$, $x_3 = \frac{-s_2 s_3}{1-s_2-s_3}$, and matrix $\boldsymbol{B}$ is composed of column vectors $b_1 = \mathbb{1} - b_2 - b_3$, $b_2 = \frac{1}{(1-x_2)(1-x_3)}\boldsymbol{G}(\mathbb{1} - \boldsymbol{G}\mathbb{1})$, $b_3 = \frac{1}{1-x_3}(\mathbb{1} - \boldsymbol{G}\mathbb{1})$.

Proof. The eigenvalues of the canonical matrix are s_1, s_2, s_3. We need to prove that $0 \le x_i < 1$ and $\gamma b_i \ge 0$ for $i = 1, 2, 3$. Based on the eigenvalue conditions of the PPN case the validity of x_1 and x_2 are immediate. For x_3 it is easy to see that $x_3 > 0$. For the other limit we have

$$\frac{-s_2 s_3}{1-s_2-s_3} < 1 \tag{22}$$

$$0 < 1 - s_2 - s_3 + s_2 s_3 \tag{23}$$

$$0 < \underbrace{(1-s_2)}_{>0}\underbrace{(1-s_3)}_{>0}. \tag{24}$$

The elements of b_2 and b_3 are non-negative, because $(\mathbb{1} - \boldsymbol{G}\mathbb{1})$ and $\boldsymbol{G}(\mathbb{1} - \boldsymbol{G}\mathbb{1})$ are the one and two steps exit probability vectors of DPH($\gamma, \boldsymbol{G}$) and $0 \le x_2, x_3 < 1$.

All that is left is to prove that b_1 is non-negative. By substituting into $b_1 = \mathbb{1} - (b_2 + b_3)$ we get

$$b_2 + b_3 = \left(\frac{1}{1-x_2}\boldsymbol{G} + \boldsymbol{I}\right)\frac{1}{1-x_3}(\boldsymbol{I} - \boldsymbol{G})\mathbb{1} = \boldsymbol{M}\mathbb{1}, \tag{25}$$

which is the product of a matrix (denoted by $\boldsymbol{M}$ above) and vector $\mathbb{1}$. Let us examine the $\sigma_i,\ i = 1, 2, 3$ eigenvalues of $\boldsymbol{M}$. Matrix $\boldsymbol{M}$ is a polynomial function of $\boldsymbol{G}$, therefore its eigenvalues can be calculated using (25) as

$$\sigma_i = \left(\frac{1}{1-x_2}s_i + 1\right)\frac{1}{1-x_3}(1 - s_i)\mathbb{1}. \tag{26}$$

First note that $\sigma_i \ge 0,\ i = 1, 2, 3$, as x_2 and x_3 are < 1. Substituting into x_2 and x_3 for $i = 1$ we get

$$\begin{aligned}
\sigma_1 &= \left(\frac{1}{1-x_2}s_1 + 1\right)\frac{1}{1-x_3}(1-s_1) \\
&= \left(\frac{1}{1-s_2-s_3}s_1 + 1\right)\frac{1-s_2-s_3}{(1-s_2)(1-s_3)}(1-s_1) \\
&= (s_1 + 1 - s_2 - s_3)\frac{1}{(1-s_2)(1-s_3)}(1-s_1) \\
&= \frac{1-s_1}{1-s_2}\left(1 + \frac{s_1 - s_2}{1-s_3}\right) \le \frac{1-s_1}{1-s_2}(1 + s_1 - s_2),
\end{aligned} \tag{27}$$

which is ≤ 1 as

$$\frac{1-s_1}{1-s_2}(1 + s_1 - s_2) \le 1 \tag{28}$$

$$(1-s_1)(1+s_1-s_2) \le (1-s_2) \tag{29}$$

$$-s1^2 + s_1 s_2 \le 0, \tag{30}$$

holds. For $i = 2$

$$\begin{aligned}
\sigma_i &= \left(\frac{1}{1-x_2}s_i + 1\right)\frac{1}{1-x_3}(1-s_i) \\
&= \left(\frac{1}{1-s_2-s_3}s_2 + 1\right)\frac{1-s_2-s_3}{(1-s_2)(1-s_3)}(1-s_2) \\
&= (s_2 + 1 - s_2 - s_3)\frac{1}{(1-s_2)(1-s_3)}(1-s_2) = 1.
\end{aligned} \tag{31}$$

Identically $\sigma_3 = 1$ can be derived. Thus the eigenvalues of $\boldsymbol{M}$ are between 0 and 1. This means that the $\boldsymbol{M}\mathbb{1}$ transformation cannot increase the length of $\mathbb{1}$, i.e., the smallest element of $b_2 + b_3 = \boldsymbol{M}\mathbb{1}$ is smaller than 1, in other words at least one of the elements of $b_1 = \mathbb{1} - (b_2 + b_3)$ is positive. However, from the first column of the matrix equation $\boldsymbol{G}\boldsymbol{B} = \boldsymbol{B}\boldsymbol{A}$ we have another expression for b_1, $x_1 b_1 = \boldsymbol{G} b_1$. That is, $x_1 = s_1$ is the largest eigenvalue of $\boldsymbol{G}$, and b_1 is the associated eigenvector, which is either strictly positive or strictly negative according to the Perron-Frobenius theorem, consequently b_1 is strictly positive. The elements of γ are non-negative, therefore $\gamma b_i,\ i = 1, 2, 3$ is non-negative as well. This completes the proof. ■

4.3. Case PNP.

The PNP case exhibits the widest set of representations. In this case the eigenvalues are real and such that $0 < s_3 < -s_2 < s_1 < 1$. Let the eigenvalue representation of the distribution be $p_i = \gamma \boldsymbol{G}^{i-1}(\mathbb{1} - \boldsymbol{G}\mathbb{1}) = \sigma_1 s_1^{i-1} + \sigma_2 s_2^{i-1} + \sigma_3 s_3^{i-1}$. Using these notations we first define the required representations.

Definition 10. The PNP representation of the distribution is

$$\alpha = \gamma \boldsymbol{B},\ \boldsymbol{A} = \begin{bmatrix} x_1 & 1-x_1 & 0 \\ x_2 & 0 & 1-x_2 \\ 0 & x_3 & 0 \end{bmatrix},$$

where $x_1 = -a_2$, $x_2 = \frac{a_0 - a_1 a_2}{a_2(1+a_2)}$, $x_3 = \frac{a_0(1+a_2)}{a_0 - a_2 - a_1 a_2 - a_2^2}$, a_0, a_1, and a_2 are the coefficients of the characteristic polynomial of $\boldsymbol{G}$, i.e., $a_0 = -s_1 s_2 s_3$, $a_1 = s_1 s_2 + s_1 s_3 + s_2 s_3$, $a_2 = -s_1 - s_2 - s_3$; matrix $\boldsymbol{B}$ is composed of column vectors $b_1 = \mathbb{1} - b_2 - b_3$, $b_2 = \frac{1}{(1-x_2)(1-x_3)}\boldsymbol{G}(\mathbb{1} - \boldsymbol{G}\mathbb{1})$, $b_3 = \frac{1}{1-x_3}(\mathbb{1} - \boldsymbol{G}\mathbb{1})$.

Definition 11. The PNP+ representation of the distribution is

$$\alpha = \left[\frac{\sigma_3}{1-s_3}, \frac{\sigma_1 s_1 + \sigma_2 s_2}{(1-s_1)(1-s_2)}, \frac{(\sigma_1+\sigma_2)(1-s_1-s_2)}{(1-s_1)(1-s_2)}\right],$$

$$\boldsymbol{A} = \begin{bmatrix} x_1 & 0 & 0 \\ 0 & x_2 & 1-x_2 \\ 0 & x_3 & 0 \end{bmatrix},$$

$x_1 = s_3$, $x_2 = s_1 + s_2$, $x_3 = \frac{-s_1 s_2}{1-s_1-s_2}$.

Definition 12. The PNP++ representation of the distribution is

$$\alpha = \left[\tfrac{\sigma_1+\sigma_2+\sigma_3}{1-s_3}, \tfrac{\sigma_1 s_1(s_1-s_3)+\sigma_2 s_2(s_2-s_3)}{(1-s_1)(1-s_2)(1-s_3)}, \tfrac{(1-s_1-s_2)(\sigma_1 s_1+\sigma_2 s_2-(\sigma_1+\sigma_2)s_3)}{(1-s_1)(1-s_2)(1-s_3)}\right]$$

$$\boldsymbol{A} = \begin{bmatrix} x_1 & 0 & 0 \\ 0 & x_2 & 1-x_2 \\ 1-x_3 & x_3 & 0 \end{bmatrix},$$

$x_1 = s_3$, $x_2 = s_1 + s_2$, $x_3 = \frac{-s_1 s_2}{1-s_1-s_2}$.

Theorem 7. *When the eigenvalues are such that $0 < s_3 < -s_2 < s_1 < 1$ the generator matrices of the PNP, the PNP+ and the PNP representations are Markovian.*

Proof. PNP representation: Let $\lambda_i = -s_i$ for $i = 1, 2, 3$. In this case λ_2 is strictly positive and so λ_1 is strictly negative, while λ_3 is non-positive. Consequently $a_0 = \lambda_1\lambda_2\lambda_3 \geq 0$. The positivity of $x_1 = -a_2$ follows from the fact that the sum of the eigenvalues of $\boldsymbol{G}$ is positive.

$$1 + a_2 = \underbrace{1+\lambda_1}_{>0} + \underbrace{\lambda_2+\lambda_3}_{\geq 0} > 0 \tag{32}$$

$$1 > -a_2 \tag{33}$$

$$1 > x_1. \tag{34}$$

The first inequality follows from $-1 < \lambda_1$ and $|\lambda_3| \leq |\lambda_2|$. The next inequality also follows from $-1 < \lambda_1, \lambda_3$ and $0 < \lambda_2$.

$$1 + a_0 + a_1 + a_2 = (1+\lambda_1)(1+\lambda_2)(1+\lambda_3) > 0. \tag{35}$$

In the following we use that $-a_2 < 1$. From that we get $a_0 \geq -a_2 a_0$.

The denominator of x_3 is

$$a_0 - a_2 - a_1 a_2 - a_2^2 \geq - \underbrace{a_2}_{<0}\underbrace{(1 + a_1 + a_2 + a_0)}_{>0} > 0. \tag{36}$$

In the nominator of x_3 a_0 is non-negative and $1 + a_2$ is positive, therefore x_3 is non-negative. We need to show that $x_3 < 1$:

$$x_3 < 1 \tag{37}$$

$$a_0 + a_0 a_2 < a_0 - a_2 - a_1 a_2 - a_2^2 \tag{38}$$

$$0 < -a_2(1 + a_0 + a_1 + a_2), \tag{39}$$

which was proven in (36). Finally, let us consider x_2:

$$x_2 < 1 \tag{40}$$

$$a_0 - a_1 a_2 > a_2(1 + a_2) \tag{41}$$

$$a_0 - a_2 - a_1 a_2 - a_2^2 > 0. \tag{42}$$

We use here that the eigenvalues of λ_i are decreasing and only λ_2 is positive:

$$x_2 = \frac{-\overbrace{(\lambda_1+\lambda_2)}^{\leq 0}\overbrace{(\lambda_1+\lambda_3)}^{\leq 0}\overbrace{(\lambda_2+\lambda_3)}^{\geq 0}}{-\underbrace{x_1}_{>0}\underbrace{(1-x_1)}_{>0}} \geq 0 \tag{43}$$

PNP+ and PNP++ representations: In these cases the properties of x_i are easy to read from the eigenvalue conditions and we have that $0 < x_1, x_2, x_3 < 1$. ■

Conjecture 1 *One of the PNP, the PNP+ and the PNP++ representations of a DPH(3) with PNP eigenvalues is Markovian.*

Proof. We could analytically treat several special cases of the DPH(3) PNP class, but we do not have formal proof which covers the whole class. Apart from the analytical treatment of the special cases we also completed an exhaustive numerical investigation and have not found any counterexample yet. ■

According to our numerical investigations the PNP++ representation covers (transforms to Markovian representation) the largest set of randomly generated DPH(3)s. The second one is the PNP representation, and the PNP+ representation covers the least among our randomly generated DPH(3)s. Among 400000 DPH(3)s with PNP eigenvalues there are ~300 ones whose PNP++ and PNP representations are non-Markovian and PNP+ representation is Markovian.

4.4. Case PNN.

Theorem 8. *If the eigenvalues of the order 3 DPH($\gamma, \boldsymbol{G}$) are $1 > |s_1| \geq |s_2| \geq |s_3|$, $\mathrm{Re}(s_1) > 0 > \mathrm{Re}(s_3) \geq \mathrm{Re}(s_2)$ and $|s_2|^2 \leq 2s_1(-\mathrm{Re}(s_2))$ then its canonical representation is DPH($\gamma\boldsymbol{B}, \boldsymbol{A}$), where*

$$\boldsymbol{A} = \begin{bmatrix} x_1 & 1-x_1 & 0 \\ x_2 & 0 & 1-x_2 \\ x_3 & 0 & 0 \end{bmatrix},$$

$x_1 = -a_2$, $x_2 = \frac{-a_1}{1+a_2}$, $x_3 = \frac{-a_0}{1+a_1+a_2}$, *the matrix elements are defined based on the coefficients of the characteristic polynomial of $\boldsymbol{G}$, $a_0 = -s_1 s_2 s_3$, $a_1 =$*

$s_1s_2 + s_1s_3 + s_2s_3$, $a_2 = -s_1 - s_2 - s_3$. *and matrix* $\boldsymbol{B}$ *is composed of column vectors* $b_1 = \mathbb{1} - b_2 - b_3$, $b_2 = \frac{1}{(1-x_2)(1-x_3)}\boldsymbol{G}(\mathbb{1} - \boldsymbol{G}\mathbb{1})$, $b_3 = \frac{1}{1-x_3}(\mathbb{1} - \boldsymbol{G}\mathbb{1})$.

Proof. The eigenvalues of the canonical matrix are s_1, s_2, s_3. We need to prove that $0 \le x_i < 1$ and $\gamma b_i \ge 0$ for $i = 1, 2, 3$.

Let $\lambda_i = -s_i$ for $i = 1, 2, 3$. The statements about a_2 in the PNP case are also valid for this case. The trace of matrix $\boldsymbol{G}$ (the sum of its diagonal elements) equals to the sum of its eigenvalues, and so the sum of the eigenvalues as well as $-a_2$ are non-negative. Consequently, $0 \le x_1 < 1$. Now we consider x_2. $(1 + a_2)$ is positive, thus we need to show that a_1 is non-positive.

If the eigenvalues are all real, then we can write

$$a_1 = \underbrace{s_1s_2}_{<0} + \underbrace{s_3}_{<0}\underbrace{(s_1 + s_2)}_{\ge 0}, \tag{44}$$

that is the sum of a negative and a non-positive number, as a consequence the result will be negative as well.

If s_2 and s_3 are complex conjugates, we can write them as $s_2 = -u + \mathrm{i}v$ and $s_3 = -u - \mathrm{i}v$, where u, v are positive reals. With these notations:

$$\begin{aligned} a_1 &= s_1(-u + \mathrm{i}v) + s_1(-u - \mathrm{i}v) + (u^2 + v^2) \\ &= u^2 + v^2 - 2s_1u \le 0, \end{aligned} \tag{45}$$

where the last inequality comes from $|s_2|^2 \le 2s_1(-\mathrm{Re}(s_2))$.

Now we show that x_2 is less than 1:

$$\begin{aligned} x_2 &< 1 \\ -a_1 &< 1 + a_2 \\ 0 &< 1 + a_1 + a_2 \end{aligned} \tag{46}$$

The last inequality can be proven by writing $1 + a_1 + a_2$ in the following way:

$$1 + a_1 + a_2 = \underbrace{(1 + \lambda_1)(1 + \lambda_2)(1 + \lambda_3)}_{>0} - \underbrace{\lambda_1\lambda_2\lambda_3}_{<0} > 0 \tag{47}$$

$\lambda_1\lambda_2\lambda_3$ is a_0, thus we also get that x_3 is positive:

$$x_3 = \frac{-\overbrace{a_0}^{<0}}{\underbrace{1 + a_1 + a_2}_{>0}} > 0$$

Similarly for the upper bound of x_3:

$$\begin{aligned} x_3 &< 1 \\ -a_0 &< 1 + a_1 + a_2 \\ 0 &< 1 + a_0 + a_1 + a_2 \\ 0 &< (1 + \lambda_1)(1 + \lambda_2)(1 + \lambda_3) \end{aligned} \tag{48}$$

The b_2 and b_3 vectors are non-negative, because $(\mathbb{1} - \boldsymbol{G}\mathbb{1})$ and $\boldsymbol{G}(\mathbb{1} - \boldsymbol{G}\mathbb{1})$ are the one and two steps exit probability vector of DPH($\gamma, \boldsymbol{G}$), and $0 \le x_2, x_3 < 1$.

Finally, from the matrix equation $\boldsymbol{GB} = \boldsymbol{BA}$ we have an explicit expression for b_1, $b_1 = \frac{1}{(1-x_1)(1-x_2)(1-x_3)}\boldsymbol{G}^2(\mathbb{1} - \boldsymbol{G}\mathbb{1})$. That is, b_1 is the three steps exit probability vector multiplied with a positive constant. ■

Theorem 8 does not cover the case when $|s_2|^2 > 2s_1(-\mathrm{Re}(s_2))$. This can occur only when s_2 and s_3 are complex conjugate eigenvalues. The following theorem applies in this case.

Theorem 9. *If the eigenvalues of the order 3 DPH($\gamma, \boldsymbol{G}$) are* $1 \ge |s_1| \ge |s_2| \ge |s_3|$, $\mathrm{Re}(s_1) > 0 > \mathrm{Re}(s_3) \ge \mathrm{Re}(s_2)$, *and* $|s_2|^2 > 2s_1(-\mathrm{Re}(s_2))$, *then we use the same canonical form as in case of PPP in Theorem 5.*

Proof. Similar to the proof of Theorem 5 we need to introduce the procedure of (Horváth and Telek, 2009) in order to prove the theorem, which we omit here. ■

5. Canonical representation of order 2 DMAP processes

In this section we give a canonical form for DMAP(2) processes.

We use a similar approach to that in Section 3, i.e. we prove that every DRAP(2) can be transformed to the introduced Markovian canonical form. We do this by choosing a set of the bounds of DRAP(2) and show that they are the tight bounds of the introduced DMAP(2) canonical form, which means that DRAP(2)$\subseteq$canonical DMAP(2), but by definition canonical DMAP(2)$\subseteq$DRAP(2), consequently DRAP(2)$\equiv$canonical DMAP(2).

The DRAP(2) processes are defined by 4 parameters (Telek and Horváth, 2007b), e.g. the first 3 factorial moments of the stationary inter-arrival time distribution (f_1, f_2, f_3), and the correlation parameter (γ). $\boldsymbol{D_0}$ and $\boldsymbol{D_1}$ of size 2×2 have a total of 8 elements (free parameters). The $(\boldsymbol{D_0} + \boldsymbol{D_1})\mathbb{1} = \mathbb{1}$ constraint reduces the number of free parameters to 6. If, additionally, two elements of the representation are set to 0, then the obtained (canonical) representation characterizes the process exactly with 4 parameters.

5.1. Canonical forms of CMAP(2).

Theorem 5 uses the relation of discrete and continuous distributions. We are going to utilize a similar relation between DMAP(2) and CMAP(2). To this end we summarize the canonical representation of CMAP(2) from (Bodrog *et al.*, 2008).

Theorem 10. *(Bodrog et al., 2008) If the correlation parameter of the order 2 CRAP(*$\boldsymbol{H_0}, \boldsymbol{H_1}$*) is*

- *non-negative, then it can be represented in the following Markovian canonical form*

$$D_0=\begin{bmatrix}-\lambda_1 & (1-a)\lambda_1\\ 0 & -\lambda_2\end{bmatrix},\ D_1=\begin{bmatrix}a\lambda_1 & 0\\ (1-b)\lambda_2 & b\lambda_2\end{bmatrix}. \tag{49}$$

where $0<\lambda_1\le\lambda_2$, $0\le a,b\le 1$, $min\{a,b\}\neq 1$, $\gamma=ab$, *and the associated embedded stationary vector is* $\boldsymbol{\pi}=\begin{bmatrix}\frac{1-b}{1-ab} & \frac{b-ab}{1-ab}\end{bmatrix}$,

- *negative, then it can be represented in the following Markovian canonical form*

$$D_0=\begin{bmatrix}-\lambda_1 & (1-a)\lambda_1\\ 0 & -\lambda_2\end{bmatrix},\ D_1=\begin{bmatrix}0 & a\lambda_1\\ b\lambda_2 & (1-b)\lambda_2\end{bmatrix}, \tag{50}$$

where $0<\lambda_1\le\lambda_2$, $0\le a\le 1$, $0<b\le 1$, $\gamma=-ab$, *and the associated embedded stationary vector is* $\boldsymbol{\pi}=\begin{bmatrix}\frac{b}{1+ab} & 1-\frac{b}{1+ab}\end{bmatrix}$.

5.2. Canonical forms of DMAP(2) with positive eigenvalues.

Theorem 11. *If the eigenvalues of* H_0 *are positive and the correlation parameter of the order 2 DRAP(*H_0,H_1*) is*

- *non-negative, then it can be represented in the following Markovian canonical form*

$$D_0=\begin{bmatrix}1-\lambda_1 & (1-a)\lambda_1\\ 0 & 1-\lambda_2\end{bmatrix},\ D_1=\begin{bmatrix}a\lambda_1 & 0\\ (1-b)\lambda_2 & b\lambda_2\end{bmatrix}. \tag{51}$$

where $0<\lambda_1\le\lambda_2$, $0\le a,b<1$, $\gamma=ab$, *and the associated embedded stationary vector is* $\boldsymbol{\pi}=\begin{bmatrix}\frac{1-b}{1-ab} & \frac{b-ab}{1-ab}\end{bmatrix}$,

- *negative, then it can be represented in the following Markovian canonical form*

$$D_0=\begin{bmatrix}1-\lambda_1 & (1-a)\lambda_1\\ 0 & 1-\lambda_2\end{bmatrix},\ D_1=\begin{bmatrix}0 & a\lambda_1\\ b\lambda_2 & (1-b)\lambda_2\end{bmatrix}, \tag{52}$$

where $0<\lambda_1\le\lambda_2$, $s_1=1-\lambda_1$, $s_2=1-\lambda_2$, $0\le a\le 1$, $0<b\le 1$, $\gamma=-ab$, *and the associated embedded stationary vector is* $\boldsymbol{\pi}=\begin{bmatrix}\frac{b}{1+ab} & 1-\frac{b}{1+ab}\end{bmatrix}$.

Proof. Practically the same approach is applied here as in Theorem 5. First note that if (H_0,H_1) is a DRAP(2), then (H_0-I,H_1) is a CRAP(2). Using this

$$\text{DRAP}(H_0,H_1)\overset{D\Rightarrow C}{\Rightarrow}\text{CRAP}(H_0-I,H_1)\equiv$$
$$\equiv\text{CMAP}(T^{-1}(H_0-I)T,T^{-1}(H_1)T) \tag{53}$$

proves the theorem. The steps are self-explanatory, except for the equivalence in the above expression, which is based on Theorem 10 in (Bodrog *et al.*, 2008). ■

5.3. Canonical forms of DMAP(2) with a negative eigenvalue.

Theorem 12. *If one eigenvalue of* H_0 *is negative and the correlation parameter of the order 2 DRAP(*H_0,H_1*) is*

- *non-negative, then it can be represented in the following Markovian canonical form*

$$\begin{aligned}D_0&=\begin{bmatrix}1-\beta_1 & a\beta_1\\ \frac{1}{a}\beta_2 & 0\end{bmatrix},\\ D_1&=\begin{bmatrix}(1-a)\beta_1 & 0\\ (1-\frac{1}{a}\beta_2)b & (1-\frac{1}{a}\beta_2)(1-b)\end{bmatrix},\end{aligned} \tag{54}$$

- *negative, then it can be represented in the following Markovian canonical form*

$$\begin{aligned}D_0&=\begin{bmatrix}1-\beta_1 & a\beta_1\\ \frac{1}{a}\beta_2 & 0\end{bmatrix},\\ D_1&=\begin{bmatrix}0 & (1-a)\beta_1\\ (1-\frac{1}{a}\beta_2)b & (1-\frac{1}{a}\beta_2)(1-b)\end{bmatrix},\end{aligned} \tag{55}$$

where the eigenvalues are such that $s_2<0<s_1<1$, $s_1+s_2>0$, *the relation of the parameters and the eigenvalues is* $\beta_1=1-s_1-s_2$, $\beta_2=\frac{s_1s_2}{s_1+s_2-1}$, $0\le b<1$, *and* $\beta_2\le a\le\min\left(1,b\frac{1-s_2}{1-s_1}\right)$ *in case of* $\gamma\ge 0$ *or* $\beta_2\le a\le 1$ *in case of* $\gamma<0$.

The correlation parameter and the first coordinate of the embedded stationary probability vectors (the unique solution of (5))

- of (54) are

$$\gamma=(1-a)(1-b)\left(1+\frac{1-a}{a}\frac{s_1s_2}{1-s_1-s_2+s_1s_2}\right), \tag{56}$$

$$\pi_1=\frac{1-\frac{1}{1-a}\gamma}{1-\gamma}, \tag{57}$$

- of (55) are

$$\gamma = -(1-a)b\left(1+\frac{1-a}{a}\frac{s_1s_2}{1-s_1-s_2+s_1s_2}\right), \tag{58}$$

$$\pi_1 = 1-\frac{1+\frac{a}{1-a}\gamma}{1-\gamma}. \tag{59}$$

We prove the theorem by considering the full flexibility of the DRAP(2) class with a negative eigenvalue and showing that the canonical forms of Theorem 12 cover this whole set of processes. To this end we first investigate the flexibility of the DRAP(2) class.

5.3.1. Constraints of the DRAP(2) class. We investigate the flexibility of the DRAP(2) class based on the following representation

$$\boldsymbol{H_0} = \begin{bmatrix} s_1 & 0 \\ 0 & s_2 \end{bmatrix},$$

$$\boldsymbol{H_1} = \begin{bmatrix} a_1+(1-a_1-s_1)\gamma & (1-a_1-s_1)(1-\gamma) \\ \frac{a_1(1-s_2)(1-\gamma)}{1-s_1} & \frac{(1-s_2)(1-a_1-s_1+a_1\gamma)}{1-s_1} \end{bmatrix}, \tag{60}$$

where s_1 is the positive, s_2 is the negative eigenvalue, γ is the correlation parameter, and a_1 is the parameter that characterizes the stationary inter-arrival distribution together with the eigenvalues according to (13). With this representation the first coordinate of the embedded stationary vector is $\pi_1 = \frac{a_1}{1-s_1}$.

For a given pair of eigenvalues, $s_1 > 0$ and $s_2 < 0$, Theorem 1 defines the limits of a_1. According to these limits the first coordinate of any embedded vector of DRAP($\boldsymbol{H_0}, \boldsymbol{H_1}$) should be bounded by

$$\frac{(1-s_2)s_2}{(1-s_2)s_2-(1-s_1)s_1} \le x \le \frac{(1-s_2)(1-s_2)}{s_1-s_2}. \tag{61}$$

Function $U_n(x)$ describes the effect of an n long inter-arrival period on the first coordinate of the embedded vector.

$$U_n(x) = \frac{(x, 1-x)\boldsymbol{H_0}^{n-1}\boldsymbol{H_1}}{(x,1-x)\boldsymbol{H_0}^{n-1}\boldsymbol{H_1}\mathbb{1}}(1,0)^T. \tag{62}$$

If the embedded vector is $(x, 1-x)$ at an arrival instance and the next inter-arrival is n time unit long, the embedded vector is going to be $(U_n(x), 1-U_n(x))$ at the next arrival instance. In case of DMAPs the embedded vector represents the probability distribution of the background Markov chain at arrivals, but in case of DRAPs it does not have any probabilistic interpretations. $\boldsymbol{H_0}$ and $\boldsymbol{H_1}$ has to be such that starting from the stationary embedded vector π for any series of inter-arrival times the first coordinate of the embedded vector satisfies (61). Based on this property we define simple constraints.

- *long series of* 1 *time unit long inter-arrivals:* $U_1(x) = x$ has to have a real solution between the bounds in (61), because if the solution was complex or larger (smaller) than the respective bound, then a series of one time unit long inter-arrivals would increase (decrease) the first coordinate above the upper (below the lower) limit (cf. Figure 2). This constraint results in

$$\gamma \le \frac{(\sqrt{c_1}-\sqrt{c_2})^2}{(c_3-a_1s_2)^2}. \tag{63}$$

- *a long series of* 1 *time unit long inter-arrivals, then a* 2 *time unit long inter-arrival:* If $\gamma > 0$, then $U_1(x)$ is a shifted negative hyperbolic function which increases monotonously between the bounds in (61). If $U_1(x) = x$ has two solutions, w_1, w_2 ($w_1 < w_2$), then w_1 is stable and w_2 is unstable, which means that starting from $x < w_1$ or $w_1 < x < w_2$ and having a long series of 1 time unit long inter-arrivals the first coordinate converges to w_1, while starting from $x > w_2$ and having a long series of 1 time unit long inter-arrivals the first coordinate diverges. Consequently a long series of 1 time unit long inter-arrivals and a 2 time unit long inter-arrival keep the first coordinate between the bounds if $U_2(w_1) \le w_2$ holds. This constraint results in

$$\gamma \le \frac{s_1s_2c_2 - c_1(1-s_1-s_2)}{c_4c_5} - \\ - \frac{\sqrt{s_1s_2c_1c_2(s_1+s_2)^2}}{c_4c_5}. \tag{64}$$

- *long series of* 2 *time unit long inter-arrivals:* Similar to the first constraint $U_2(x) = x$ has to have a real solution which results in

$$\gamma \ge \frac{\sqrt{s_1s_2c_2}+\sqrt{c_6})^2}{{c_4}^2}. \tag{65}$$

- *a long series of* 1 *time unit long inter-arrivals:* If $\gamma < 0$, then $U_1(x)$ is a shifted hyperbolic function which decreases monotonously between the bounds in (61). $U_1(x) = x$ has to have a stable real solution (w_1) between the bounds in (61), which holds if $\frac{d}{dx}U_1(x)|_{x=w_1} > -1$ (cf. Figure 3) (in case of a long series of 1 time unit long inter-arrivals the first coordinate converges to w_1). This constraint results in

$$\gamma \ge \frac{s_2(1-a_1-s_1)+a_1s_1}{(c_3-a_1s_1)^2}. \tag{66}$$

In the above expressions the auxiliary variables are

$$\begin{aligned}
&c_1 = -a_1(s_1-s_2)^2(1-a_1-s_1),\\
&c_2 = (1-s_1)^3(1-s_2),\\
&c_3 = 1-s_1(2-a_1-s_1),\\
&c_4 = s_1(1-s1)(1-a_1-s_1)+a_1s_2(1-s_2),\\
&c_5 = (a_1(s_1-s_2)+s_2(1-s_1)^2),\\
&c_6 = -a_1(1-a_1-s_1)(s_1(1-s_1)-s_2(1-s_2))^2.
\end{aligned} \tag{67}$$

We summarize the results of this subsection in the following theorem.

Theorem 13. *For DRAP($\boldsymbol{H_0}, \boldsymbol{H_1}$) defined in* (60) *with* $0 < s_1 < 1$, $-s_1 < s_2 < 0$ *and* a_1 *satisfying Theorem 1 the correlation parameter satisfies the inequalities* (63) - (66).

Theorem 13 defines only some bounds of the set of DRAP(2) processes, but the subsequent analysis of the canonical DMAP(2) proves that these bounds are tight.

5.3.2. Constraints of the set of canonical DMAP(2) processes. Having the bounds of the DRAP(2) class from Theorem 13 we are ready to prove Theorem 12.

Proof. (Theorem 7) First we need to relate the variables of the canonical representation with the parameters used for characterizing the DMAP(2) processes. The relation of β_1, β_2 with s_1, s_2 is

$$s_{1,2} = \frac{1}{2}\left(1-\beta_1 \pm \sqrt{(1-\beta_1)^2+4\beta_1\beta_2}\right) \tag{68}$$

The relation of s_1, s_2, a_1, γ with a and b can be obtained from (56) and (57) for the first canonical form and from (58) and (59) for the second canonical form.

If $\gamma > 0$, then

$$\begin{aligned}
a &= \frac{g_1+\sqrt{g_1^2-g_2}}{2e_1},\\
b &= 1-\frac{a\gamma(1-s_1-s_2+s_1s_2)}{(1-a)(a(1-s_1-s_2)+s_1s_2)},
\end{aligned} \tag{69}$$

where

$$\begin{aligned}
&e_1 = (1-s_1)(1-s_1-s_2)^2,\\
&e_2 = (1-s_1-s_2)(a_1(s_1-s_2)(1-\gamma)-s_1(1-s_1)),\\
&e_3 = \gamma(1-s_1)^2,\\
&g_1 = e_1+e_2-e_3(1-s_1-s_2),\\
&g_2 = 4e_1(e_2+e_3s_1)
\end{aligned} \tag{70}$$

and if $\gamma < 0$, then

$$\begin{aligned}
a &= \frac{g_3-\sqrt{g_3^2+g_4}}{g_5},\\
b &= 1-\frac{a\gamma(1-s_1-s_2-s_1s_2)}{(1-a)(a(1-s_1-s_2)+s_1s_2)},
\end{aligned} \tag{71}$$

where

$$\begin{aligned}
e_6 &= a_1(s1-s2)(1-\gamma),\\
e_7 &= (1-s_1)(s_2(1-\gamma)-(1-s_1-s_2)\gamma),\\
e_8 &= (1-s_1-s_2)(1-s_1)s_2,\\
g_3 &= -(1-s_1-s_2)e_6+e_7s_1-e_8,\\
g_4 &= 4(e_6+e_7)e_8s_1,\\
g_5 &= -2(1-s_1-s_2)(e_6+e_7).
\end{aligned} \tag{72}$$

Based on these relations the constraints of the canonical DMAP(2) processes can be obtained using the fact that all the elements of $\boldsymbol{D_0}$ and $\boldsymbol{D_1}$ have to be non-negative real numbers. That is, a is real, $\beta_2 \le a \le 1$ and $0 \le b \le 1$. Parameter a is real when the expression under the square root sign in (69) for $\gamma > 0$ and in (71) for $\gamma < 0$ is non-negative. All together these constrains result in 5 inequalities for $\gamma > 0$ and 5 for $\gamma < 0$. Out of these the following ones are relevant.

- Case $\gamma > 0$:
  - a is real when $g_1^2 - g_2 \ge 0$, which translates to (63),
  - the inequality $b \le 1$ translates to (64),
- Case $\gamma < 0$:
  - a is real when $g_3^2 + g_4 \ge 0$, which translates to (65),
  - the inequality $b \ge 0$ translates to (66).

Appendix A provides a detailed derivation of (63) based on $g_1^2 - g_2 \ge 0$. We neglect the details of the other derivations. ■

6. Explicit moments and correlation matching with the canonical forms

One of the most important applications of the introduced canonical forms is the factorial moments matching for DPH(2) and DPH(3) distributions and the factorial moments and correlation matching of DMAP(2) processes.

In the second part of this section we give explicit factorial moment and correlation matching formulas for order 2 models. While such formulas cannot be provided for DPH(3), the canonical form still makes moment matching possible. In the first part of this section we discuss this matching procedure for DPH distributions in general.

6.1. Moment matching with DPH. To obtain formulas for moments matching the inverse of (3) is required, that is, we need to find a vector-matrix pair based on a given set of factorial moments. For the full characterization of a DPH(n) we need the first $2n-1$ factorial moments ($f_1, f_2, \ldots, f_{2n-1}$). We find an appropriate vector-matrix pair exhibiting a given set of factorial moments

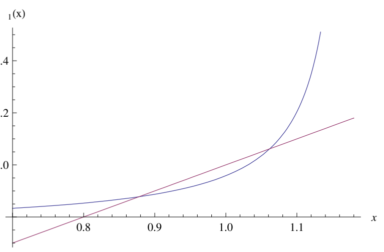

Fig. 2. The $U_1(x)$ function when $s_1 = 0.8,\ s_2 = -0.3,\ a_1 = 0.19,\ \gamma = 0.17$.

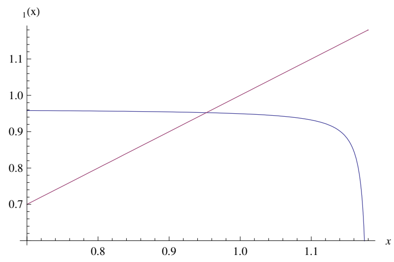

Fig. 3. The $U_1(x)$ function when $s_1 = 0.8,\ s_2 = -0.3,\ a_1 = 0.19,\ \gamma = -0.012$.

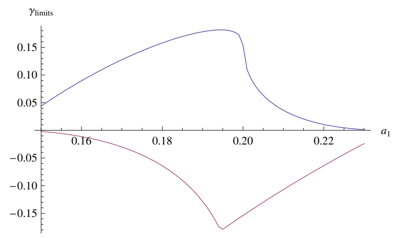

Fig. 4. The upper and lower γ limits as a function of a_1 when $s_1 = 0.8,\ s_2 = -0.3$

using the procedure available for CPH moments matching in (Horváth and Telek, 2007). In spite of the fact that (10) and (3) look similar, we cannot directly use the CPH moments matching method for DPH moments matching, because of the extra term in (3). That is why we first transform the factorial moments such that they exhibit an expression similar to (10).

Let us define $\boldsymbol{G} = -(\boldsymbol{I} - \boldsymbol{A})^{-1}\boldsymbol{A}$, then $\boldsymbol{A}^{-1} = \boldsymbol{I} - \boldsymbol{G}^{-1}$. Substituting this into (3) we get

$$\begin{aligned}\frac{f_i}{i!} &= \alpha \left((\boldsymbol{I} - \boldsymbol{A})^{-1}\boldsymbol{A}\right)^i \boldsymbol{A}^{-1}\mathbb{1} = \\ &= \alpha(-\boldsymbol{G})^i(\boldsymbol{I} - \boldsymbol{G}^{-1})\mathbb{1} \\ &= (-1)^i\, \alpha(\boldsymbol{G}^i - \boldsymbol{G}^{i-1})\mathbb{1}. \end{aligned} \tag{73}$$

Assuming $f_0 = 1$ and $\alpha\mathbb{1} = 1$, from (73) we have

$$\sum_{i=0}^{k}(-1)^i \frac{f_i}{i!} = \alpha\boldsymbol{G}^k\mathbb{1} \tag{74}$$

Multiplying both sides with $k!$ we obtain

$$\hat{\mu}_k \triangleq k! \sum_{i=0}^{k}(-1)^i \frac{f_i}{i!} = k!\, \alpha\boldsymbol{G}^k\mathbb{1}, \tag{75}$$

which has the same form as (10). Applying the CPH moments matching procedure with $\hat{\mu}_k$ results in α and $\boldsymbol{G}$ which satisfy (75). Finally, matrix $\boldsymbol{A}$ is obtained from $\boldsymbol{A} = (\boldsymbol{G} - \boldsymbol{I})^{-1}\boldsymbol{G}$. This procedure commonly generates a non-Markovian matrix $\boldsymbol{A}$.

6.1.1. Moment matching with canonical DPH(2) and DPH(3). Applying the general DPH moments matching procedure of the previous subsection we attain an $(\alpha, \boldsymbol{A})$ MG(2) or MG(3) representation based on (f_1, f_2, f_3) or $(f_1, f_2, \ldots, f_5)$. By determining the eigenvalues of $\boldsymbol{A}$ the appropriate type of canonical form can be decided and its elements can be calculated according to Section 3 or 4. If the resulting representation is Markovian, then the given set of factorial moments can be matched with a DPH(2) or a DPH(3). Otherwise it is not possible.

6.2. Parameter matching with DMAP(2). For DMAP(2) processes the previously mentioned inverse characterization is possible, that is, the first 3 moments (f_1, f_2, f_3) and the correlation parameter (γ) can be used to give explicit formulas for β_1, β_2, a, b of Theorem 11 and 12.

Our matching method is composed of two steps. The first step is moment matching with a DPH(2). The result of this phase is an $(\alpha, \boldsymbol{A})$ canonical DPH(2) representation. The second step is the matching of γ. This means the calculation of a and b of Theorem 11 and 12 from α, $\boldsymbol{A}$, and γ.

6.2.1. Bounds of DMAP(2) processes. For exact parameter matching first it has to be decided if a DMAP(2) exists with a given set of f_1, f_2, f_3, γ moments and correlation parameter set, and if the matching is possible, it has to be determined if one of the eigenvalues of $\boldsymbol{D_0}$ is negative, as this affects the formulas for the elements of the canonical form. To this end moment and correlation bounds have to be established.

It can be easily proven that the class of DPH(2) distributions can be defined as the stationary inter-arrival time distribution of DMAP(2) processes, thus their moment bounds are identical. These bounds can be derived from the Markovian constraints on the canonical form of DPH(2) distributions (i.e. the elements of α and $\boldsymbol{A}$ in Theorem 2, 3, and 4 have to be between 0 and 1). For $\boldsymbol{A}$ with two positive eigenvalues the constraints are already given in (Telek and Heindl, 2002). These results are summarized in Table 1, where

$$\begin{aligned} j_1 = \frac{6}{(2\,f_1 + \sqrt{2\,j_2})^3} \cdot \Big(& f_1\,(2\,f_1 + \sqrt{2\,j_2})(3\,f_2 + 2\,f_1) \\ & \cdot(f_2 - 2\,f_1 + 2) - 2\,{f_2}^2\,(f_2 - \sqrt{2\,j_2})\Big)\ , \end{aligned} \tag{76}$$

and $j_2 = 2\,{f_1}^2 - 2\,f_1 - f_2$.

For the negative eigenvalue case we have derived similar constraints as shown in Table 2, where

$$j_3 = \frac{3\sqrt{(f_2-2f_1(f_1-5)-8)(f_2-2f_1(f_1-1))^3}}{4(f_1-1)} + \\ + \frac{3(-4f_1(f_1-2)(f_1-1)^2+8f_2+4f_1f_2(f_1-3))+f_2^2}{4(f_1-1)} \qquad (77)$$

In the following we present formulas for β_1, β_2, p. Substituting them into equations (63) – (66) exact γ bounds can be easily derived. However, the resulting expressions are rather long, therefore we do not show them.

6.2.2. Transformation to DMAP(2) canonical form with positive eigenvalues. If the f_1, f_2, f_3 moments are in the bounds described by Table 1, they can be matched with a DPH(2) with positive eigenvalues. In this case the first step is based on Table 3 in (Telek and Heindl, 2002). The s_1 and s_2 elements of matrix $\boldsymbol{A}$ and vector α can be calculated as

$$\begin{aligned} \alpha &= [p, 1-p]\,,\ p = \frac{-z(h_3-6f_1h_1)+\sqrt{h_4}}{zh_3+\sqrt{h_4}}, \\ s_1 &= 1-\frac{h_3-z\sqrt{h_4}}{h_2},\ s_2 = 1-\frac{h_3+z\sqrt{h_4}}{h_2}, \end{aligned}$$

where

$$\begin{aligned} h_1 &= 2{f_1}^2-2f_1-f_2,\ h_2 = 3{f_2}^2-2f_1f_3, \\ h_3 &= 3f_1f_2-6(f_1+f_2-f_1^2)-f_3, \\ h_4 &= h_3^2-6h_1h_2, \\ z &= \frac{h_2}{|h_2|}. \end{aligned} \qquad (78)$$

The second step is the calculation of a, b of Theorem 11. If $\gamma = 0$, then $a = 1,\ b = 0$. If $\gamma > 0$, then a and b can be computed using

$$a = \frac{d_1-\sqrt{d_2}}{2(1-s_1)},\quad b = \frac{d_1+\sqrt{d_2}}{2(1-s_2)},$$

with

$$\begin{aligned} d_1 &= 1-s_2-p(1-s_2)(1-\gamma)+(1-s_1)\gamma, \\ d_2 &= d_1^2-4(1-s_1)(1-s_2)\gamma. \end{aligned}$$

If $\gamma \le 0$, then

$$\begin{aligned} a &= \frac{-\gamma(1-s_2)}{p(1-s_2)(1-\gamma)-\gamma(1-s_1)}, \\ b &= \frac{p(1-s_2)(1-\gamma)-\gamma(1-s_1)}{1-s_2}. \end{aligned} \qquad (79)$$

6.2.3. Transformation to canonical form with a negative eigenvalue. If the f_1, f_2, f_3 moments are in the bounds described by Table 2, they can be matched with a DPH(2) with a positive and a negative eigenvalue. In this case the $\beta_1,\ \beta_2$ parameters and the α vector can be calculated using

$$\begin{aligned} \beta_1 &= \frac{12f_1^2-3f_2(4+f_2)-2f_3+2f_1(-6+3f_2+f_3)}{(3f_2^2-2f_1f_3)} \\ \beta_2 &= \frac{-3f_2(2-2f_1+f_2)+2(-1+f_1)f_3}{12f_1^2-3f_2(4+f_2)-2f_3+2f_1(-6+3f_2+f_3)} \\ p &= \frac{\beta_1-f_1\beta_1+\beta_2+f_1\beta_1\beta_2}{-1+\beta_2},\ \alpha = [p, 1-p]\,. \end{aligned}$$

From β_1 and β_2 the eigenvalues s_1 and s_2 are obtained by (68). In the second step a, b of Theorem 12 are calculated. If $\gamma = 0$, then $a = 1,\ b = 0$ stands again. Otherwise if $\gamma > 0$, then

$$\begin{aligned} a &= \frac{k_1+\sqrt{k_1^2-k_2}}{2\beta_1}, \\ b &= 1-\frac{a\gamma(1-\beta_2)}{(1-a)(a-\beta_2)}, \end{aligned} \qquad (80)$$

if $\gamma < 0$, then

$$\begin{aligned} a &= \frac{k_3+\sqrt{k_3^2+4\beta_2k_4}}{2k_4}, \\ b &= -\frac{a\gamma(1-\beta_2)}{(1-a)(a-\beta_2)}, \end{aligned} \qquad (81)$$

where

$$\begin{aligned} k_1 &= (1-\gamma)(p+\beta_1+\beta_2-p\beta_2)-1+\beta_1, \\ k_2 &= 4\beta_1(k_1-\beta_1+\gamma-\beta_2\gamma), \\ k_3 &= (1-\gamma)(-p(1-\beta_2)-2\beta_2)-\gamma(1-\beta_1), \\ k_4 &= k_3+\beta_2+\gamma-\beta_2\gamma. \end{aligned} \qquad (82)$$

If the f_1, f_2, f_3 moments are out of the bounds described by both Table 1 and 2, then exact matching is not possible.

7. Fitting using canonical forms

In some cases fitting based on a well chosen distance measure might capture the important characteristics of traffic traces better than moment matching. Employing canonical forms is beneficial in this case as well.

The main advantage of using canonical forms in model fitting compared to the corresponding general form is that, while the canonical forms have the full flexibility of the given class, the number of parameters that has to be optimized is lower. When fitting with DPH(2) the canonical form has 3 parameters instead of the 5 of the general form (a DPH(2) has 6 elements and the $\alpha\mathbb{1} = 1$ equation gives one constraint). The canonical form of DPH(3)

condition	bounds	DPH(2)
	$1 \leq f_1 < \infty$	-
$1 \leq f_1 < 2$	$2(f_1 - 1) \leq f_2 < \infty$	-
$2 \leq f_1$	$\frac{f_1(3f_1-4)}{2} \leq f_2 < \infty$	-
	$1 \leq f_1 < 2$	
$2(f_1 - 1) \leq f_2$	$j_1 \leq f_3$	$\beta_1 = \beta_2$
$f_2 < 2f_1(f_1 - 1)$	$f_3 \leq \frac{3\,f_2(f_2 - 2\,f_1 + 2)}{2\,(f_1 - 1)}$	$\beta_2 = 1$
	$2 \leq f_1$	
$\frac{f_1(3f_1-4)}{2} \leq f_2$	$j_1 \leq f_3$	$\beta_1 = \beta_2$
$f_2 < 2(f_1 - 1)$	$f_3 \leq 6(f_1 - 1)(f_2 - f_1^2 + f_1)$	$p = 1$
$2(f_1 - 1) \leq f_2$	$j_1 \leq f_3$	$\beta_1 = \beta_2$
$f_2 < 1 - \frac{1}{f_1}$	$f_3 \leq \frac{3\,f_2(f_2 - 2\,f_1 + 2)}{2\,(f_1 - 1)}$	$\beta_2 = 1$
	$1 \leq f_1$	
$2f_1(f_1 - 1) \leq f_2$	$\frac{3\,f_2(f_2 - 2\,f_1 + 2)}{2\,(f_1 - 1)} \leq f_3$	$\beta_2 = 1$

Table 1. Bounds for the first three moments of DPH(2) distributions with positive eigenvalues

condition	bounds	DPH(2)
	$1 \leq f_1 < \infty$	-
$1 \leq f_1 < 2$	$2(f_1 - 1) \leq f_2 < \infty$	-
$2 \leq f_1$	$2(f_1 - 1)^2 \leq f_2 < \infty$	-
	$1 \leq f_1 < 2$	
$2(f_1 - 1) \leq f_2$	$\frac{3(f_2-2f_1+2)(f_1+f_2)}{(2f_1-1)} \leq f_3$	$\beta_2 = 0$
$f_2 < 2f_1(f_1 - 1)$	$f_3 \leq \frac{3\,f_2(f_2 - 2\,f_1 + 2)}{2\,(f_1 - 1)}$	$\beta_1 = 1$
	$2 \leq f_1$	
$2(f_1 - 1)^2 \leq f_2$	$\frac{3\,f_2(f_2-2\,f_1+2)}{2\,(f_1-1)} \leq f_3$	$\beta_2 = 0$
$f_2 < f_1(2f_1 - 3)$	$f_3 \leq 6(f_1 - 1)(f_2 - f_1^2 + f_1)$	$p = 1$
$f_1(2f_1 - 3) \leq f_2$	$\frac{3\,f_2(f_2-2\,f_1+2)}{2\,(f_1-1)} \leq f_3$	$\beta_2 = 0$
$f_2 < 2f_1(f1 - 1)$	$f_3 \leq \frac{3(f_2 - 2f_1 + 2)(f_1 + f_2)}{(2f_1 - 1)}$	$\beta_1 = 1$
	$1 \leq f_1$	
$1 - 2f_1(f1 - 1) \leq f_2$	$j_3 \leq f_3$	$p = 0$
	$f_3 \leq \frac{3\,f_2(f_2 - 2\,f_1 + 2)}{2\,(f_1 - 1)}$	$\beta_2 = 0$

Table 2. Bounds for the first three moments of DPH(2) distributions with a negative eigenvalue

has 5 parameters instead of the 8 of the general case (a DPH(3) has 9 elements and $\alpha \mathbb{1} = 1$ gives a single constraint again). Finally, a canonical DMAP(2) has 4 parameters instead of 6 (a DMAP(2) has 8 elements, but the $(\boldsymbol{D_0} + \boldsymbol{D_1})\mathbb{1} = \mathbb{1}$ equation means 2 constraints). Having fewer parameters results in a faster and better fitting in general (for the chosen distance measure).

In this section we provide numerical examples to demonstrate the advantages of using canonical forms. We use DPH(3) fitting as an illustration. Our choice is motivated by the fact that DPH(3)s are significantly more complex than DPH(2)s, however we can use a very straightforward fitting method for them with relative entropy as a distance measure, which makes the demonstration simpler than it would be with DMAP(2) fitting.

As mentioned above, we use relative entropy as distance function in our examples. Having the $\mathcal{X}$ and $\mathcal{Y}$ discrete distributions on $\mathbb{N}^+$ with pmfs $f(i)$ and $g(i)$, we can calculate their H relative entropy (or Kullback-Leibler divergence) as

$$H(\mathcal{X}, \mathcal{Y}) = -\sum_{i=1}^{\infty} f(i) \ln \left(\frac{g(i)}{f(i)} \right). \qquad (83)$$

If $f(i)$ is zero for a given i, that part of the expression is considered zero. The relative entropy of two distributions is strictly non-negative and is only zero if $f(i) = g(i)$. Intuitively, higher H means a bigger difference between the two distributions and a worse fitting in our case.

In the following we present the results of fitting to three different distributions. The first one is the discrete uniform distribution on 1 to 50 (i.e. $f(i) = 0.02$ if $i = 1 \dots 50$ and $f(i) = 0$ otherwise). The second one is the DPH(4) with

$$\alpha = [0.5, 0.2, 0.1, 0.2]],$$

$$\boldsymbol{A} = \begin{bmatrix} 0.6 & 0.1 & 0.07 & 0.03 \\ 0.3 & 0.06 & 0.22 & 0.36 \\ 0.14 & 0.4 & 0.1 & 0.2 \\ 0.3 & 0.1 & 0.2 & 0.05 \end{bmatrix},$$

Distribution	General form		Canonical form	
	Distance	Time	Distance	Time
Uniform	4.55	473 s	0.355	168 s
DPH(4)	0.00256	511 s	$3.29{\times}10^{-4}$	319 s
DPH(3)	6.32	4859 s	0.025	1571 s

Table 3. Fitting of distributions with general and canonical DPH(3) form

which has a monotonically decreasing pmf, and the third one is the DPH(3) with

$$\alpha = [0.3, 0.1, 0.6]\,,\; \boldsymbol{A} = \begin{bmatrix} 0.2 & 0.75 & 0.05 \\ 0.5 & 0.1 & 0.4 \\ 0.1 & 0.7 & 0.07 \end{bmatrix},$$

which has a fluctuating pmf with a slow decay. We made the fitting using the built-in optimization function of Wolfram Mathematica (called NMinimize). For the general form we had to consider only one type of representation. In case of the canonical form, we ran the fitting algorithm for all the different types of representations and chose the best one. When fitting the uniform distribution we took the theoretical pmf values. In the other two cases we simulated $100,000$ inter-arrival times using the respective $(\alpha, \boldsymbol{A})$ and fitted using the empirical pmf of the traces. The results are summarized in Table 3. They clearly show that canonical forms perform better than the general form in fitting. The intuitive explanation is that the canonical forms have less parameters, consequently the optimization is a simpler task than in the general case. Furthermore, as different representations describe the same distribution in the general form, they have the same distance from the fitted trace. This suggests that the relative entropy is a very bumpy function of the parameters for the general case, which also makes the optimization harder.

The uniform distribution was hard to fit for both the canonical and the general form, however the first one still gave a much better result in terms of both distance and running time. Similarly, both the canonical and the general form was able to fit the trace of the DPH(4), but the canonical fitting was faster again. Probably the most interesting example is the fitting of the DPH(3) trace. The pmfs of the fitted DPHs can be seen in Figure 5. While, in theory, a perfect fit would have been possible, the general form provided a poor solution. Using the canonical form resulted in a good fitting, although it took a long time. This is due to the slow decay of the distribution, because it makes the goal function much more complex, as it has more elements than in the previous cases.

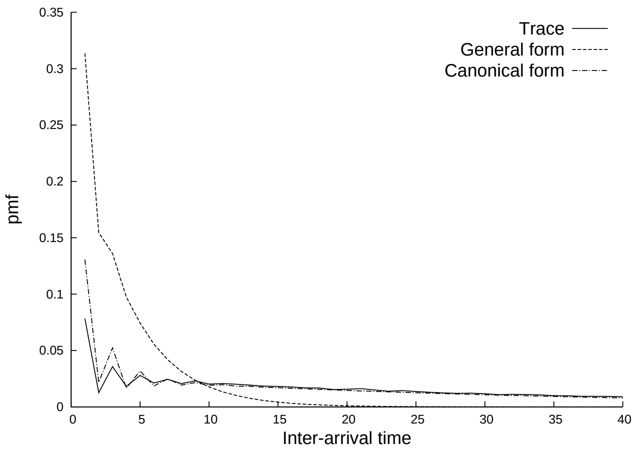

Fig. 5. Fitting of DPH(3) distribution with general and canonical form

8. Conclusions

In this paper we presented canonical representations for order 2 and 3 DPH distributions and order 2 DMAPs. We provided a detailed proof for the validity of these canonical representations, gave explicit methods to obtain these representations, and proved that the order 2 Markovian models are equivalent to their non-Markovian counterparts.

We demonstrated the benefits of these canonical forms in parameter matching and trace fitting. Using them, we derived the moment and correlation bounds of order 2 DMAPs (and DPHs) and presented explicit matching formulas for these parameters. For order 3 DPH distributions we provided a simple procedure that can be used for moment matching.

We illustrated the advantages of fitting with canonical forms instead of the general form through numerical examples. The results confirmed that with canonical forms a substantially better performance can be achieved in both running time and fitting quality than with using general Markovian forms.

Acknowledgment

The authors gratefully acknowledge the support of the TÁMOP-4.2.2C- 11/1/KONV-2012-0001 and the OTKA K101150 projects. They would also like to thank Illés Horváth for his assistance in the proofs.

References

Alfa, A. (2002). Discrete time queues and matrix-analytic methods, *Top* **10**: 147–185.

Bodrog, L., Heindl, A., Horváth, G. and Telek, M. (2008). A markovian canonical form of second-order matrix-exponential processes, *European Journal of Operation Research* **190**: 459–477.

Horváth, G. and Telek, M. (2007). A canonical representation of order 3 phase type distributions, *Formal Methods and Stochastic Models for Performance Evaluation* pp. 48–62.

Horváth, G. and Telek, M. (2009). On the canonical representation of phase type distributions, *Performance Evaluation* **66**(8): 396 – 409.

Lakatos, L., Szeidl, L. and Telek, M. (2013). *Introduction to Queueing Systems with Telecommunication Applications*, Springer.

Neuts, M. (1981). *Matrix Geometric Solutions in Stochastic Models*, Johns Hopkins University Press, Baltimore.

Telek, M. and Heindl, A. (2002). Matching moments for acyclic discrete and continuous phase-type distributions of second order, *International Journal of Simulation Systems, Science & Technology* **3**(3-4): 47–57. Special Issue on: Analytical & Stochastic Modelling Techniques.

Telek, M. and Horváth, G. (2007a). A minimal representation of markov arrival processes and a moments matching method, *Performance Evaluation* **64**(9): 1153–1168.

Telek, M. and Horváth, G. (2007b). A minimal representation of Markov arrival processes and a moments matching method, *Performance Evaluation* **64**(9-12): 1153–1168.

András Mészáros is a Ph.D. student at the Department of Networked Systems and Services at Budapest University of Techonology and Economics. He received the M. Sc. degree in Electrical Engineering from the Technical University of Budapest in 2011. He carries out research in Markovian modeling of complex stochastic systems.

János Papp is a M.Sc. student at Department of Mathematics of the Technical University of Budapest. He received the B. Sc. degree in Mathematics from the Technical University of Budapest in 2012. His research interest includes the canonical representation of PH distribution.

Miklós Telek received the M. Sc. degree in Electrical Engineering from the Technical University of Budapest in 1987. In the same year he joined the Hungarian Post Research Institute where he studied the modelling, analysis and planning aspects of communication networks. Since 1990 he has been with the Department of Networked Systems and Services of the Technical University of Budapest, where he is a full professor now. He received the C.Sc. and D.Sc. degree from the Hungarian Academy of Science in 1995 and 2004, respectively. His current research interest includes stochastic performance modelling and analysis of computer and communication systems.

Appendix A

The constraint for the correlation coefficient is the solution of $g_1^2(\gamma) - g_2(\gamma) = 0$ for γ. By substituting (70) into g_1 and g_2, regrouping the terms according to the different powers of γ, and simplifying the expression we get

$$
\begin{aligned}
&g_1^2 - g_2 = \\
&(1-s_1-s_2)^2(1-s_1(2-a_1-s_1)-a_1-s_2)^2\gamma^2- \\
&-2(1-s1-s_2)^2\Big[1+s_1^2\big(3-a_1(1-a_1+2s_2)-3s_2\big)- \\
&-s_1^3(1-a_1-s_2)-s_2\big(1+(1-a_1)a_1s_2\big)- \\
&-s_1\Big(3-s_2\big(3+a_1(2-2a_1+s_2)\big)\Big)\Big]\gamma+ \\
&+(1-s_1-s_2)^2(1-s_1(1+a_1-s_2)-s_2+a_1s_2)^2.
\end{aligned}
$$

By solving the equation $g_1^2(\gamma)-g_2(\gamma)=0$ and taking the smaller solution and simplifying the result we get

$$
\begin{aligned}
&\gamma = \\
&\frac{a_1^2(s_1-s_2)^2-a_1(1-s_1)(s_1-s_2)^2+(1-s_1)^3(1-s_2)}{(1-s_1(2-a_1-s_1)-a_1s_2)^2}- \\
&-\frac{2\sqrt{-a_1(1-s1)^3(1-a_1-s_1)(s_1-s_2)^2(1-s_2)}}{(1-s_1(2-a_1-s_1)-a_1s_2)^2}.
\end{aligned}
$$

From this last expression, one can see that the numerator is the square of $\sqrt{-a_1(s_1-s_2)^2(1-a_1-s_1)} - \sqrt{(1-s_1)^3(1-s_2)}$. The constraint will be this smaller solution. Finally, we get (63) by substituting the c_i formulas from (67).

Received: