Transformation of Acyclic Phase Type Distributions for Correlation Fitting

Peter Buchholz, Iryna Felko, Jan Kriege
Department of Computer Science,
TU Dortmund
{peter.buchholz,iryna.felko,jan.kriege}@udo.edu

Author prepared version of a paper published in

Analytical and Stochastic Modeling Techniques and Applications

Proceedings of the ASMTA 2013

Lecture Notes in Computer Science volume 7984, pages 96-111, Springer Berlin Heidelberg

The original publication is available at www.springerlink.com

http://dx.doi.org/10.1007/978-3-642-39408-9_8

Transformation of Acyclic Phase Type Distributions for Correlation Fitting

Peter Buchholz, Iryna Felko, Jan Kriege
Department of Computer Science,
TU Dortmund
{peter.buchholz,iryna.felko,jan.kriege}@udo.edu

Abstract

In this paper similarity transformations for *Acyclic Phase Type Distributions* (APHs) are considered, and representations maximizing the first joint moment that can be reached when the distribution is expanded into a *Markovian Arrival Process* (MAP) are investigated. For the acyclic case the optimal representation corresponds to a hyperexponential representation, which is optimal among all possible representations that can be reached by similarity transformations. The parameterization aspect for the possible transformation of APHs into a hyperexponential form is revealed, together with corresponding transformation rules. For the case when APHs cannot be transformed into a hyperexponential representation a heuristic optimization method is presented to obtain good representations, while transformation methods to increase the first joint moment by adding additional phases are derived.

1 Introduction

In stochastic modeling, the appropriate representation of dynamic processes behavior is of major importance to built realistic models. Phase type distributions (PHDs) and Markovian arrival processes (MAPs) have developed to versatile modeling tools which can be applied to capture a wide range of different stochastic behaviors and can be used in queuing network models [18], renewal theory [15], reliability, and healthcare modeling [9]. PHDs and MAPs allow for applying efficient matrix analytic methods, can approximate any continuous distribution and can be used to model empirical data containing a wide range of stochastic behaviors.

To capture real behavior the parameters of a PHD or MAP have to be estimated according to some observations measured in a real system [13]. Parameter estimation for PHDs and MAPs is a nontrivial task resulting in a nonlinear optimization problem which becomes even more complex since the matrix representation of a PHD or MAP is redundant [20]. To reduce the complexity fitting algorithms for PHDs are often tailored to specific subclasses and canonical representations such that the number of free parameters in the initial distribution vector $\mathbf{p}$ and the transition rates matrix $\mathbf{D}_0$ is minimized [3, 22, 10]. For the same reason MAP fitting is often performed in two steps [3, 11]. In a first step a PHD $(\mathbf{p}, \mathbf{D}_0)$ is fitted and in a second step a matrix $\mathbf{D}_1$ is constructed to model the correlation such that the resulting MAP $(\mathbf{D}_0, \mathbf{D}_1)$ has the stationary distribution $\mathbf{p}$. Obviously, the entries in $(\mathbf{p}, \mathbf{D}_0)$ have a large influence when fitting $\mathbf{D}_1$, since the entries put constraints on the possible values of entries of $\mathbf{D}_1$, thereby limiting the possible range of autocorrelation one can achieve. Since the representation of a PH distribution is non-unique [20] one tries to find an equivalent representation of $(\mathbf{p}, \mathbf{D}_0)$ by transforming the distribution that allows for the largest flexibility before fitting $\mathbf{D}_1$. However, only little is known about which representation of a PHD is favorable when used in a two-step MAP fitting approach. Obviously, the canonical form of a PHD that is often used for PH fitting is not suitable since it has only a single exit state and does not allow for any flexibility when fitting $\mathbf{D}_1$. The same holds for representations that only have a single entry state. Consequently, existing transformations [2, 3, 11] aim at increasing the number of entry and exit states. However, in an empirical study it has been shown recently [16] that in general these transformations do not find the optimal representation that allows for the maximal flexibility. In this paper we present a similarity transformation approach in the field of equivalent representations for PHDs and we try to shed some light on which representation of a PHD is favorable for a subsequent MAP fitting step. In particular, we show which representation of a PHD allows for the expansion into a MAP with maximal first order joint moment and present transformations to obtain this representation.

The paper is organized as follows: In Sect. 2 we give a short introduction to PHDs and MAPs. In Sect. 3 we describe the transformation algorithm for the case with two phases in detail. Sect. 4 contains descriptions for the general case. In Sect. 5 a technique for state space expansion with the aim of finding a representation allowing larger flexibility is presented. In Sect. 6 we present some examples to demonstrate the effectiveness of the approach. The paper ends with the conclusions.

2 Background

We first introduce the basic notation and define PHDs and MAPs, then we present the basic results for similarity transformations.

2.1 Basic Definitions for PH Distributions and MAPs

We consider acyclic phase type distributions (APHs) [8] which are defined by a vector matrix pair $(\mathbf{p}, \mathbf{D}_0)$ where $\mathbf{p}$ is a n dimensional row vector containing the initial probability distribution (i.e., $\mathbf{p} \geq \mathbf{0}$, $\mathbf{p}\mathbb{1} = 1$ where $\mathbb{1} = (1, \ldots, 1)^T$ of length n). $\mathbf{D}_0$ is an upper triangular matrix of order n with negative diagonal elements, non-negative elements above the diagonal and row sums ≤ 0. We denote the diagonal entries by $-\lambda_i < 0$ and the non-diagonal entries by $\lambda_{i,j} \geq 0$. Since the matrix is upper triangular, we have $\lambda_{i,j} = 0$ for $i > j$. Furthermore, we denote by $\mathbf{\Lambda}$ a column vector of length n that contains in position i the value λ_i. We use the terms state and phase interchangeably for APHs.

$(\mathbf{p}, \mathbf{D}_0)$ defines an absorbing CTMC and the time to absorption is phase type distributed. The density function, distribution function and moments are given by

$$f(x) = \mathbf{p}e^{\mathbf{D}_0 x}\mathbf{d}_1, \qquad F(x) = 1 - \mathbf{p}e^{\mathbf{D}_0 x}\mathbb{1} \qquad \text{and} \qquad \mu_k = k!\mathbf{p}\mathbf{M}^k\mathbb{1},$$

respectively, where $\mathbf{d}_1 = -\mathbf{D}_0\mathbb{I}$ is the vector of exit rates, $x \geq 0$, $\mathbf{M} = (-\mathbf{D}_0)^{-1}$ is the moment matrix and $\mathbf{m}^k = \mathbf{M}^k\mathbb{I}$ is the vector of the conditional kth moments. For $\mathbf{m}^1$ we use the notation $\mathbf{m}$.

APHs can be expanded to Markovian Arrival Processes (MAPs) [17] which are defined by a pair of two matrices $(\mathbf{D}_0, \mathbf{D}_1)$ such that $\mathbf{D}_0 + \mathbf{D}_1$ is a generator of an irreducible CTMC. A MAP $(\mathbf{D}_0, \mathbf{D}_1)$ is an expansion of an APH $(\mathbf{p}, \mathbf{D'}_0)$ if $\mathbf{D}_0 = \mathbf{D'}_0$ and the embedded distributions at event times are identical which means $\mathbf{p} = \mathbf{pMD}_1$ holds. The expansion of APHs to MAPs is used in so called two phase fitting approaches [6, 3, 11] where in a first step an APH is fitted and afterwards correlation is inserted by finding an appropriate matrix $\mathbf{D}_1$. This approach is more efficient than the combined fitting of both matrices but has the disadvantage that the generated APH might not be flexible enough.

The joint density and the joint moments of a MAP are given by

$$f(x_1, \ldots, x_m) = \mathbf{p}\left(\prod_{i=1}^{m} e^{\mathbf{D}_0 x_i}\mathbf{D}_1\right)\mathbb{I} \qquad \text{and} \qquad \mu_{k,l} = k!l!\mathbf{pM}^k\mathbf{MD}_1\mathbf{M}^l\mathbb{I}$$

for $x_i \geq 0$. The lag k autocorrelation coefficients and (as special case) the lag 1 autocorrelation are computed as

$$\rho_k = \frac{\mu_1^{-2}\mathbf{pM}\left(\mathbf{MD}_1\right)^k\mathbf{M}\mathbb{I} - 1}{2\mu_1^{-2}\mathbf{pM}^2\mathbb{I} - 1} \qquad \text{and} \qquad \rho_1 = \frac{\mu_1^{-2}\mu_{1,1} - 1}{\mu_1^{-2}\mu_2 - 1}. \tag{1}$$

For a given APH ρ_1 of an expanded MAP is directly proportional to the first joint moment $\mu_{1,1}$ since the other quantities in (1) are determined by the distribution. Observe that if $(\mathbf{p}, \mathbf{D}_0)$ is an APH and $(\mathbf{D}_0, \mathbf{D}_1)$ is a MAP with first joint moment $\mu_{1,1}$ that is expanded from the APH, then $(\mathbf{D}_0, \alpha\mathbf{D}_1 + (1-\alpha)\mathbf{d}_1\mathbf{p})$ for $\alpha \in [0, 1]$ is a MAP with first joint moment $\alpha\mu_{1,1} + (1-\alpha)\mu_1\mu_1$ which implies that for every first joint moment between $\mu_{1,1}$ and $\mu_1\mu_1$ a MAP can be expanded from the APH. Consequently, it is important to know the MAP with the maximal or minimal first joint moment that can be expanded from an APH. The maximum case is usually more important since positive correlation is more common in practice than negative correlation. Therefore we consider in the following MAPs with a maximal first joint moment that can be expanded from APHs. However, approaches to derive MAPs with a minimal first joint moment are similar.

2.2 Similarity Transformation Results

As already mentioned the representation of a PHD is not unique and different representations can be transformed into one another. In principle one can use every non-singular matrix $\mathbf{C}$ with unit row sum and generate a new representation $(\mathbf{pC}, \mathbf{C}^{-1}\mathbf{D}_0\mathbf{C})$ [20]. This representation usually is not an APH and it is often not even a Markov representation. A restricted set of similarity transformations is used in [8] to transform an APH in canonical form. Several canonical forms have been defined by Cumani [8] which have only a single exit state (i.e, $\mathbf{d}_1(i) > 0$ for exactly one i) or single entry state (i.e., $\mathbf{p}(i) = 1$ for exactly one i). Both representations are not suited for the expansion of the APH to a MAP since a single entry or exit state implies that the MAP's correlation matrix $\mathbf{D}_1$ is completely defined by the APH and the MAP has no correlation. However, the transformation rules of Cumani can be used in a more general way which has been done in [3, 11] to transform an APH in canonical representation into a representation that is more appropriate for expansion into a MAP. The used methods are heuristics which apply a single transformation step that depends on a free parameter for all rows of the matrix. Experiments show that the transformation often does not work appropriately, in particular for APHs with a larger number of states.

This has recently been investigated systematically in [16] where it could be shown that in many cases the transformation is not satisfactory and it is unclear how to set the free parameter. However, the results in [16] also show that an APH which can be expanded to MAPs with a wide range of joint moments $\mu_{1,1}$, in general also implies that the resulting MAPs are flexible in reaching other correlation quantities like higher order joint moments or lag k autocorrelations for $k > 1$. Thus, the optimization of the representation of APHs is important for a good approximation of general traffic processes including correlation.

We first consider the basic transformation step to modify APHs which is based on the ideas presented in [8]. The approach uses the representation of an exponential distribution by a Coxian distribution with 2 states. Consider an exponential distribution with rate λ_1, then this distribution can be represented by the following APH with 2 states.

$$\mathbf{p} = (1 - \lambda_1/\lambda_2, \lambda_1/\lambda_2)\,,\ \mathbf{D}_0 = \begin{pmatrix} -\lambda_1 & \lambda_1 \\ 0 & -\lambda_2 \end{pmatrix} \tag{2}$$

with $\lambda_2 \geq \lambda_1$. The resulting representation has two entry and one exit state. A variant of the transformation with one entry and two exit states has been proposed in [12]. Based on (2), the following transformations can be defined which modify the representation but not the distribution and still result in an APH. The following equations apply the transformation to two states $i < j$. We assume that for $i < j$ $\lambda_i \leq \lambda_j$ which can be always achieved by elementary transformations following from [8] that are implemented in the BuTools [21]. $\mathbf{p}^\delta$ and $\mathbf{D}_0^\delta$ are the vector and matrix after the transformation with parameter δ has been applied. Proofs for the validity of the transformation steps can be found in [2].

$$\mathbf{p}^\delta(k) = \begin{cases} \mathbf{p}(i) + \delta & \text{for } k = i \\ \mathbf{p}(j) - \delta & \text{for } k = j \\ \mathbf{p}(k) & \text{otherwise} \end{cases} \qquad \lambda_{k,l} = \begin{cases} \lambda_{i,j}\frac{\mathbf{p}(i)}{\mathbf{p}(i)+\delta} - \frac{(\lambda_j - \lambda_i)\delta}{\mathbf{p}(i)+\delta} & \text{for } k = i \text{ and } l = j \\ \lambda_{i,l}\frac{\mathbf{p}(i)}{\mathbf{p}(i)+\delta} & \text{for } k = i \text{ and } l > i \wedge l < j \\ \lambda_{i,l}\frac{\mathbf{p}(i)}{\mathbf{p}(i)+\delta} + \lambda_{j,l}\frac{\delta}{\mathbf{p}(i)+\delta} & \text{for } k = i \text{ and } l > j \\ \lambda_{k,i}\frac{\mathbf{p}(i)+\delta}{\mathbf{p}(i)} & \text{for } k < i \text{ and } l = i \\ \lambda_{k,j} - \lambda_{k,i}\frac{\delta}{\mathbf{p}(i)} & \text{for } k < i \text{ and } l = j \\ \lambda_{k,l} & \text{otherwise} \end{cases} \tag{3}$$

The diagonal entries λ_k are not modified by the transformation. For the exit vector $\mathbf{d}_1$, $\mathbf{d}_1^\delta(k) = \mathbf{d}_1(k)$ for $k \neq i$ and

$$\mathbf{d}_1^\delta(i) = \frac{\mathbf{p}(i)\mathbf{d}_1(i) + \delta\mathbf{d}_1(j)}{\mathbf{p}(i) + \delta} = \mathbf{d}_1(i) + \delta\frac{\mathbf{d}_1(j) - \mathbf{d}_1(i)}{\mathbf{p}(i) + \delta}. \tag{4}$$

To compute a valid APH, parameter δ has to be chosen from the following interval to assure that the rates and probabilities remain non-negative.

$$\left[\max\left(-\mathbf{p}(i), \min_{l>j,\lambda_{j,l}>0}\left(-\frac{\mathbf{p}(i)\lambda_{i,l}}{\lambda_{j,l}}\right), -\frac{\mathbf{p}(i)\mathbf{d}_1(i)}{\mathbf{d}_1(j)}\right),\ \min\left(\mathbf{p}(j), \min_{k<i,\lambda_{k,i}>0}\left(\frac{\mathbf{p}(i)\lambda_{k,j}}{\lambda_{k,i}}\right), \frac{\mathbf{p}(i)\lambda_{i,j}}{\lambda_j - \lambda_i}\right)\right] \tag{5}$$

If $\mathbf{d}_1(j) = 0$, then the last term for the lower bound becomes $-\infty$ and is not used. If $\lambda_j = \lambda_i$, then the last term in the upper bound evaluates to ∞. Furthermore, if δ is set to $-\mathbf{p}(i)$, then the rates of the new APH become infinite.

3 The Two Phase Case

In this section we consider APHs with two states and compute a representation that allows the expansion to a MAP with a maximal first joint moment $\mu_{1,1}$. The APH is given by

$$\mathbf{p} = (\pi, 1-\pi)\,,\ \mathbf{D}_0 = \begin{pmatrix} -\lambda_1 & \lambda_{1,2} \\ 0 & -\lambda_2 \end{pmatrix}.$$

We assume $\lambda_2 \geq \lambda_1$. If this is not the case, we use the representation

$$\mathbf{p} = \left(\frac{(1-\pi)\lambda_1 + \pi\lambda_{1,2}}{\lambda_1}, 1 - \frac{(1-\pi)\lambda_1 + \pi\lambda_{1,2}}{\lambda_1}\right),\ \mathbf{D}_0 = \begin{pmatrix} -\lambda_2 & \frac{\pi\lambda_2\lambda_{1,2}}{(1-\pi)\lambda_1+\pi\lambda_{1,2}} \\ 0 & -\lambda_1 \end{pmatrix}$$

Equivalence of both representations can be proved by a comparison of their elementary series [2, 8]. Furthermore, the time scale can be modified such that λ_1 becomes 1 and we obtain a representation

$$\mathbf{p} = (\pi, 1-\pi)\,,\ \mathbf{D}_0 = \begin{pmatrix} -1 & \lambda_{1,2} \\ 0 & -\lambda_2 \end{pmatrix} \text{ where } 0 \leq \pi, \lambda_{1,2} \leq 1 \text{ and } \lambda_2 \geq 1. \tag{6}$$

The transformations (3) simplify in the 2 state case (6) to

$$\mathbf{p}^\delta = \left(\pi^\delta, 1-\pi^\delta\right) = (\pi+\delta, 1-\pi-\delta)\,,\ \lambda_{1,2}^\delta = \frac{\pi\lambda_{1,2}}{\pi+\delta} - \frac{(\lambda_2-1)\delta}{\pi+\delta} \tag{7}$$

where $\delta \in \left[\max\left(-\pi, -\frac{\pi(1-\lambda_{1,2})}{\lambda_2}\right), \min\left(1-\pi, \frac{\pi\lambda_{1,2}}{\lambda_2-1}\right)\right]$. We denote the boundaries of the interval as δ^- and δ^+, respectively. APH(δ) is the APH ($\mathbf{p}^\delta, \mathbf{D}_0^\delta$) with the parameters in (7) computed for δ and diagonal entries $-1, -\lambda_2$ in matrix $\mathbf{D}_0^\delta$.

Let $\mu_{1,1}^*(\delta)$ be the maximal joint moment $\mu_{1,1}$ that can be reached with a MAP that is expanded from APH(δ). This raises two questions, namely, what is the best δ, i.e., $\delta^* = \arg\max_{\delta\in[\delta^-,\delta^+]}\left(\mu_{1,1}^*(\delta)\right)$, and what are the parameters of the MAP that results in $\mu_{1,1}^*(\delta)$ for a given δ?

Before we answer the questions, a slightly different representation for $\mu_{1,1}$ will be introduced. In vector matrix representation, the first joint moment is given by

$$\mu_{1,1} = \mathbf{pMMD}_1\mathbf{M}\mathbb{1}$$

where $\mathbf{M} = (-\mathbf{D}_0)^{-1}$. Let $a_{i,j}^\delta$ be the probability that the APH starts in state i and j is the last state before the event occurs (i.e., before absorption if $\mathbf{D}_0$ is interpreted as generator matrix of an absorbing CTMC). We have

$$a_{1,1}^\delta = \pi^\delta(1-\lambda_{1,2}^\delta),\ a_{1,2}^\delta = \pi^\delta\lambda_{1,2}^\delta,\ a_{2,1}^\delta = 0, \text{ and } a_{2,2}^\delta = 1-\pi^\delta.$$

Let furthermore $\nu_{i,j}$ be the mean duration if the process starts in i and j is the last state before the event occurs, then

$$\nu_{1,1} = 1, \nu_{1,2} = 1 + (\lambda_2)^{-1}, \text{ and } \nu_{2,2} = (\lambda_2)^{-1}.$$

$\nu_{2,1}$ is not available since an APH can not start in state 2 and visit state 1 before an event occurs. To define a MAP we use the probabilities b_i which describe the probability that the MAP starts in state 1 after i was the last state before absorption. Matrix $\mathbf{D}_1$ is then given by

$$\mathbf{D}_1 = \begin{pmatrix} 1-\lambda_{1,2}^\delta & 0 \\ 0 & \lambda_2 \end{pmatrix}\begin{pmatrix} b_1 & 1-b_1 \\ b_2 & 1-b_2 \end{pmatrix}$$

To observe $\mathbf{pMD}_1 = \mathbf{p}$, the following relation has to hold

$$\pi^\delta = a^\delta_{1,1} b_1 + (a^\delta_{1,2} + a^\delta_{2,2}) b_2 = \pi^\delta (1 - \lambda^\delta_{1,2}) b_1 + (\pi^\delta \lambda^\delta_{1,2} + 1 - \pi^\delta) b_2 \tag{8}$$

The equation shows that only one value b_i can be set, the other one is fixed to observe the initial probability. If we fix b_2, it can be chosen from

$$\left[\frac{\pi^\delta \lambda^\delta_{1,2}}{1 - \pi^\delta (1 - \lambda^\delta_{1,2})}, \min\left(1, \frac{\pi^\delta}{1 - \pi^\delta (1 - \lambda^\delta_{1,2})}\right)\right]$$

to assure that $b_1 \in [0, 1]$. With these notations we can write down an equation for $\mu_{1,1}(\delta)$

$$\begin{aligned} \mu_{1,1}(\delta) \quad &= \quad b_1 a^\delta_{1,1} \nu_{1,1} (a^\delta_{1,1} \nu_{1,1} + a^\delta_{1,2} \nu_{1,2}) / \pi^\delta + (1 - b_1) a^\delta_{1,1} \nu_{1,1} \nu_{2,2} + \\ &\quad\quad b_2 (a^\delta_{1,2} \nu_{1,2} + a^\delta_{2,2} \nu_{2,2})(a^\delta_{1,1} \nu_{1,1} + a^\delta_{1,2} \nu_{1,2}) / \pi^\delta + \\ &\quad\quad (1 - b_2)(a^\delta_{1,2} \nu_{1,2} + a^\delta_{2,2} \nu_{2,2}) \nu_{2,2} \end{aligned}$$

The equations results from considering all possible sequences of states with the corresponding probabilities and mean durations in the MAP.

After some elementary but still lengthy substitutions and transformations we obtain the following representation for $\mu_{1,1}(\delta)$ in terms of δ and b_2.

$$\begin{aligned} \mu_{1,1}(\delta) \quad &= \quad \frac{y_1 + b_2 \left(y_{21} + \delta y_{22} + \frac{1}{\pi+\delta} y_{23} + \frac{\delta}{\pi+\delta} y_{24} + \frac{\delta^2}{\pi+\delta} y_{25}\right)}{\lambda_2^2} \\ \text{where} \quad & \quad y_1 = 1 + \pi(\lambda_2^2 + (\lambda_2 + 1)\lambda_{1,2} - 1) \\ & \quad y_{21} = \pi((\lambda_2 - 1)^2 + 2\lambda_{1,2}(\lambda_2 - 1)) - (\lambda_2 - 1)^2 \\ & \quad y_{22} = -(\lambda_2 - 1)^2 \qquad y_{23} = \pi \lambda_{1,2}(1 + \pi \lambda_{1,2} - \lambda_2) \\ & \quad y_{24} = (\lambda_2 - 1 - 2\pi\lambda_{1,2})(\lambda_2 - 1) \qquad y_{25} = (\lambda_2 - 1)^2 \end{aligned}$$

Observe that the coefficients y_{xy} do not contain δ or b_i and are defined for the original APH without modifications due to δ. Define

$$y_2(\delta) = y_{21} + \delta y_{22} + \frac{1}{\pi + \delta} y_{23} + \frac{\delta}{\pi + \delta} y_{24} + \frac{\delta^2}{\pi + \delta} y_{25} \tag{9}$$

The representation implies that for $y_2(\delta) < 0$ b_2 should be as small as possible to maximize $\mu_{1,1}(\delta)$ (i.e., find $\mu^*_{1,1}(\delta)$) and for $y_2(\delta) > 0$ b_2 should be as large as possible for $\mu^*_{1,1}(\delta)$. $y_2(\delta) = 0$ implies that the APH has no flexibility in adopting joint moments.

Now consider the case $\delta = 0$ such that $y_2(0) = y_{21} + y_{23}/\pi$. $y_2(0) = 0$ for $\pi = \frac{\lambda_2 - 1}{\lambda_2 + \lambda_{1,2} - 1}$, $y_2(0) < 0$ for $\pi < \frac{\lambda_2 - 1}{\lambda_2 + \lambda_{1,2} - 1}$ and $y_2(0) > 0$ for $\pi > \frac{\lambda_2 - 1}{\lambda_2 + \lambda_{1,2} - 1}$. Interestingly, the relations remain if we substitute π by π^δ and $\lambda_{1,2}$ by $\lambda^\delta_{1,2}$ since then

$$\pi + \delta = \frac{(\lambda_2 - 1)(\pi + \delta)}{\lambda_2(\pi + \delta) + \pi\lambda_{1,2} - (\lambda_2 - 1)\delta - (\pi + \delta)} = \frac{(\pi + \delta)(\lambda_2 - 1)}{\pi(\lambda_2 + \lambda_{1,2} - 1)}.$$

The latter term equals $\pi + \delta$ if π observes the above equality. This implies that an APH belongs to one of three classes depending on the relation between π and $\frac{\lambda_2 - 1}{\lambda_2 + \lambda_{1,2} - 1}$. The case with equality is not interesting since it allows no flexibility. However, it may be handled by expanding the state space as shown in Section 5. Now we consider the remaining two cases in some detail.

We begin with **Case 1** $\pi < \frac{\lambda_2-1}{\lambda_2+\lambda_{1,2}-1}$. To obtain $\mu^*_{1,1}(0)$ b_2 is set to the minimum, i.e., $b_2 = \frac{\pi\lambda_{1,2}}{1-\pi(1-\lambda_{1,2})}$ and $b_1 = 1$ which follows from substituting b_2 into (8). Denote by b_2^δ the value of b_2 for varying δ which equals

$$b_2^\delta = \frac{(\pi\lambda_{1,2} - (\lambda_2 - 1)\delta)}{1 - \pi + \pi\lambda_{1,2} - \lambda_2\delta} \geq 0.$$

The derivative with respect to δ equals

$$\frac{db_2^\delta}{d\delta} = \frac{\left(\pi - \frac{\lambda_2-1}{\lambda_2+\lambda_{1,2}-1}\right)(\lambda_2 - 1 + \lambda_{1,2})}{(\pi\lambda_{1,2} - \pi + 1 - \lambda_2\delta)^2} < 0$$

The denominator is always positive and the numerator is negative for *Case 1* as we will show. The first term of the numerator is obviously negative by the assumption for *Case 1*. For the second term $\lambda_2 + \lambda_{1,2} - 1 \geq 0$ holds since $\lambda_2 \geq 1$ is assumed. The derivative of $y_2(\delta)$ equals

$$\begin{aligned} \frac{dy_2(\delta)}{d\delta} &= -(\lambda_2-1)^2 + \frac{-\pi\lambda_{1,2}(1+\pi\lambda_{1,2}-\lambda_2)+\pi(\lambda_2-1-2\pi\lambda_{1,2})(\lambda_2-1)+(2\delta\pi+\delta^2)(\lambda_2-1)^2}{(\pi+\delta)^2} \\ &= \frac{\pi\big((\lambda_2+\lambda_{1,2}-1)\big((1-\pi)(\lambda_2-1)-\pi\lambda_{1,2}\big)\big)}{(\pi+\delta)^2} > 0 \end{aligned} \tag{10}$$

Since the denominator and the first part of the numerator are positive, we only have to consider the last term, namely $(1-\pi)(\lambda_2-1)-\pi\lambda_{1,2}$. Since *Case 1* implies $\lambda_{1,2} < \frac{(\lambda_2-1)(1-\pi)}{\pi}$, we have $(1-\pi)(\lambda_2-1)-\pi\lambda_{1,2} > (1-\pi)(\lambda_2-1)-(1-\pi)(\lambda_2-1) = 0$. Then

$$\frac{d\mu_{1,1}(\delta)}{d\delta} = \frac{db_2^\delta}{\delta}\frac{y_2(\delta)}{\lambda_2^2} + \frac{b_2^\delta}{\lambda_2^2}\frac{dy_2(\delta)}{\delta} > 0$$

since the first product consists of two negative values and the second of two positive values such that the sum is positive and δ should be chosen as the maximum value. We set $\delta = \frac{\lambda_{1,2}\pi}{\lambda_2-1}$ which implies $1 - \pi > \delta$ since $1 - \pi > \frac{\lambda_{1,2}\pi}{\lambda_2-1} \Leftrightarrow \frac{(1-\pi)(\lambda_2-1)}{\pi} > \lambda_{1,2}$ which holds by assumption in *Case 1*. Furthermore,

$$\lambda_{1,2}^\delta = \frac{\pi\lambda_{1,2} - (\lambda_2 - 1)\frac{\lambda_{1,2}\pi}{\lambda_2-1}}{\pi + \frac{\lambda_{1,2}\pi}{\lambda_2-1}} = 0. \tag{11}$$

This implies that the distribution where the maximum value for $\mu_{1,1}(\delta)$ can be reached is a hyperexponential distribution.

In **Case 2** we have $\pi > \frac{\lambda_2-1}{\lambda_2+\lambda_{1,2}-1}$. To obtain $\mu^*_{1,1}(0)$ b_2 is set to the maximum, i.e., $b_2 = \min\left(1, \frac{\pi}{1-\pi(1-\lambda_{1,2})}\right)$. Two cases have to be distinguished depending whether b_2^δ is bounded by 1 or by the second term. If $b_2^\delta = 1$, then the derivative of b_2^δ with respect to δ is 0. This implies that

$$\frac{d\mu_{1,1}(\delta)}{d\delta} = b_2^\delta\frac{dy_2(\delta)}{\delta} < 0$$

since $b_2^\delta > 0$ and $\lambda_{1,2} > \frac{(\lambda_2-1)(1-\pi)}{\pi}$ holds such that the > in (11) changes to a <.

In the second case

$$b_2^\delta = \frac{\pi + \delta}{1 - \pi + \pi\lambda_{1,2} - \lambda_2\delta} \quad \text{and} \quad \frac{db_2^\delta}{d\delta} = \frac{1 + \pi(\lambda_2 + \lambda_{1,2} - 1)}{(1 - \pi + \lambda_{1,2}\pi - \lambda_2\delta)^2}.$$

$y_2(\delta)$ and its derivative are as given in (9) and (10), respectively. Thus, we obtain after some simplifications

$$\frac{d\mu_{1,1}(\delta)}{d\delta} = \frac{db_2^\delta}{\delta}\frac{y_2(\delta)}{\lambda_2^2} + \frac{b_2^\delta}{\lambda_2^2}\frac{dy_2(\delta)}{\delta} = \frac{\pi\lambda_2\big((\pi-1)\big(\lambda_2^2-2\lambda_2+1\big)+\pi\lambda_{1,2}\big(2\lambda_2-2+\lambda_{1,2}\big)+\lambda_{1,2}(1-\lambda_2)\big)}{\lambda_2^2\big(1-\pi+\pi\lambda_{1,2}-\delta\lambda_2\big)^2}$$

The value of the derivative is positive, as we will show now. First, the denominator and the first term in the numerator are positive such that we have to show that the second term of the numerator (i.e., the term in the brackets) is positive. Observe that π is multiplied in this term with non-negative factors because $\lambda_2 > 1$ such that we can substitute π by the lower bound to obtain a lower bound for the whole term.

$$\begin{array}{ll} (\pi-1)\left(\lambda_2^2 - 2\lambda_2 + 1\right) + \pi\lambda_{1,2}\left(2\lambda_2 - 2 + \lambda_{1,2}\right) + \lambda_{1,2}(1-\lambda_2) & > \\ \frac{-\lambda_{1,2}}{\lambda_2+\lambda_{1,2}-1}\left(\lambda_2^2 - 2\lambda_2 + 1\right) + \frac{(\lambda_2-1)\lambda_{1,2}}{\lambda_2+\lambda_{1,2}-1}\left(2\lambda_2 - 2 + \lambda_{1,2}\right) + \lambda_{1,2}(1-\lambda_2) = \frac{\lambda_2\lambda_{1,2}}{\lambda_2+\lambda_{1,2}-1} & \geq 0 \end{array}$$

Thus, the behavior is as follows: As long as $\pi^\delta/(1-\pi^\delta(1-\lambda_{1,2}^\delta)) < 1$, the derivative with respect to δ is positive such that δ has to be increased to increase $\mu_{1,1}(\delta)$. If $\pi^\delta/(1-\pi^\delta(1-\lambda_{1,2}^\delta)) > 1$, the derivative is negative and δ should decreased to increase $\mu_{1,1}(\delta)$. At $\pi^\delta/(1-\pi^\delta(1-\lambda_{1,2}^\delta)) = 1$, the derivative from the left is positive and from the right is negative such that this point is optimal which implies $\delta = (\lambda_{1,2}\pi + 1 - 2\pi)/(1+\lambda_2)$.

Example: We consider three simple examples. The first belongs to *Case 1* with $\delta^* = 0.3$ and is given by the following $\mathbf{p}$ and $\mathbf{D}_0$ and the resulting $\mathbf{p}^{\delta^*}$ and $\mathbf{D}_0^{\delta^*}$

$$\mathbf{p} = (0.6, 0.4)\,,\ \mathbf{D}_0 = \begin{pmatrix} -1.0 & 0.5 \\ 0 & -2.0 \end{pmatrix}, \qquad \mathbf{p}^{\delta^*} = (0.9, 0.1)\,,\ \mathbf{D}_0^{\delta^*} = \begin{pmatrix} -1.0 & 0 \\ 0 & -2.0 \end{pmatrix}.$$

The second example belongs to *Case 2* with $\delta^* = 0.112$ and the original and transformed matrices are as follows:

$$\mathbf{p} = (0.6, 0.4)\,,\ \mathbf{D}_0 = \begin{pmatrix} -1.0 & 0.8 \\ 0 & -1.5 \end{pmatrix}, \qquad \mathbf{p}^{\delta^*} = (0.712, 0.288)\,,\ \mathbf{D}_0^{\delta^*} = \begin{pmatrix} -1.0 & 0.5955 \\ 0 & -2.0 \end{pmatrix}.$$

The third example is given by

$$\mathbf{p} = \left(\frac{2}{3}, \frac{1}{3}\right),\ \mathbf{D}_0 = \begin{pmatrix} -1.0 & 0.25 \\ 0 & -1.5 \end{pmatrix}.$$

In this example $\pi = \frac{\lambda_2-1}{\lambda_2+\lambda_{2,1}-1}$ such that $y_2(\delta) = 0$ and the APH has no flexibility according to the expansion to a MAP. The last example, in fact, describes an exponential distribution.

4 The General Case

It is very hard to extend the detailed analysis we presented in the previous section to APHs with an arbitrary number of states. We were only able to prove one of the two cases from the previous section for n states. For the second case, we present below a heuristic optimization algorithm.

Theorem 4.1 *If an APH can be transformed into an hyperexponential representation with* $\mathbf{p}(i) > 0$ *using similarity transformations (3)-(4), then this representation results in the maximal value* $\mu_{1,1}^*$.

Proof Note, that for some APH with representation $(\tilde{\mathbf{p}}, \tilde{\mathbf{D}}_0, \tilde{\mathbf{d}}_1)$ which results from the hyperexponential distribution $(\mathbf{p}, \mathbf{D}_0, \mathbf{d}_1)$ using the similarity transformation defined in (3)-(4) these transformations can be collected in a non-singular transformation matrix $\mathbf{C}$ with $\mathbf{C}\mathbb{I} = \mathbb{I}$ such that $\tilde{\mathbf{p}} = \mathbf{p}\mathbf{C}$, $\tilde{\mathbf{D}}_0 = \mathbf{C}^{-1}\mathbf{D}_0\mathbf{C}$ and $\tilde{\mathbf{d}}_1 = \mathbf{C}^{-1}\mathbf{d}_1$. The key idea of the proof is to formulate a linear program that maximizes the first joint moment of $(\tilde{\mathbf{p}}, \tilde{\mathbf{D}}_0, \tilde{\mathbf{d}}_1)$ and from which it becomes visible that the maximal value equals $\mu_{1,1}^*$. The complete proof can be found in the online companion [5].

The proof of the theorem indicates that the hyperexponential representation is not only optimal for the acyclic case, it is optimal among all possible representations that can be reached by similarity transformations. For a given APH the hyperexponential distribution can be generated by a repeated use of the transformation steps (3)-(4) where parameter δ is chosen such that $\lambda_{i,j}$ becomes zero. If a transformation step is not possible because the interval for δ is too small, the distribution cannot be transformed into a hyperexponential form. Formally, the theorem shows the optimality of a hyperexponential representation with $\mathbf{p}(i) > 0$ for all i. However, since $\mu_{1,1}$ is continuous in the transformations (3)-(4), the limiting cases with $\mathbf{p}(i) = 0$ are optimal too and allow the generation of a representation with less states because a hyperexponential phase with initial probability 0 is redundant.

For the case that the APH cannot be transformed into a hyperexponential form, we were not yet able to extend the 2-state case. Our first and natural idea to apply the results for 2 states also in the general case fails in most cases. The transformation often stucks in local maxima that may be far from the global maximum.

Therefore, we used, similar to [16], a heuristic approach to find an optimal APH representation. In our case, simulated annealing [7] is applied for optimization by choosing random indices i, j with $i < j$ and $\lambda_{i,j}$, choose randomly a δ from the interval $[\delta^-, \delta^+]$, perform the transformation, compute the new value for $\mu_{1,1}$ which is kept if it is larger than the previous value or, if it smaller, it is kept with some probability according to the simulated annealing approach. With this approach we usually obtain good representations which, however, needs not be optimal.

Examples: Consider the following APH in canonical form and the corresponding hyperexponential representation that results by applying the transformation steps:

$$\mathbf{p} = \left(\frac{3}{32}, \frac{15}{32}, \frac{7}{16}\right), \ \mathbf{D}_0 = \begin{pmatrix} -0.5 & 0.5 & 0 \\ 0 & -1 & 1 \\ 0 & 0 & -4 \end{pmatrix} \quad \mathbf{p}' = \left(\frac{3}{14}, \frac{1}{2}, \frac{4}{14}\right), \ \mathbf{D}_0' = \begin{pmatrix} -0.5 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -4 \end{pmatrix}. \tag{12}$$

The transformed hyperexponential representation has a maximal lag 1 autocorrelation coefficient of 0.21429.

As a second example we consider the following APH $(\mathbf{p}, \mathbf{D}_0)$ in canonical form which has no hyperexponential representation. If we apply the transformations towards a hyperexponential representation we obtain the representation $(\mathbf{p}', \mathbf{D}_0')$:

$$\mathbf{p} = \left(\frac{2}{5}, \frac{19}{40}, \frac{1}{8}\right), \ \mathbf{D}_0 = \begin{pmatrix} -1 & 1 & 0 \\ 0 & -2 & 2 \\ 0 & 0 & -4 \end{pmatrix} \quad \mathbf{p}' = (0.925, 0.075, 0)\,, \ \mathbf{D}_0' = \begin{pmatrix} -0.5 & 0 & 0.45946 \\ 0 & -2 & 1.4375 \\ 0 & 0 & -4 \end{pmatrix}.$$

The maximal lag 1 autocorrelation which can be achieved with a MAP generated from this representation is 0.0024. With the simulated annealing approach the following representation is generated.

$$\mathbf{p} = (0.783847, 0.182691, 0.033463)\,, \ \mathbf{D}_0 = \begin{pmatrix} -1.00000 & 0.04171 & 0.95732 \\ 0.00000 & -2.00000 & 0.00000 \\ 0.00000 & 0.00000 & -4.00000 \end{pmatrix}. \tag{13}$$

This representation can be expanded to a MAP with lag 1 autocorrelation 0.0751 which is still small but more than with the APH resulting from the previous transformation.

5 State Space Expansion

If the autocorrelation which can be achieved with a given APH is not sufficient, then it is possible to enlarge the state space to find another representation of the APH with a larger flexibility. The extension of the state space is done by cloning single states. The approach uses an equivalence relation between APHs which has been defined in a more general context in [4]. Let $(\mathbf{p}, \mathbf{D}_0)$ and $(\mathbf{p}', \mathbf{D}_0')$ be two APHs of order n and $n+1$, respectively. $\mathbf{V}$ is an $(n+1)\times n$ matrix with unit row sums (i.e., $\mathbf{V}\mathbb{1} = \mathbb{1}$). If $\mathbf{p}'\mathbf{V} = \mathbf{p}$ and $\mathbf{D}_0'\mathbf{V} = \mathbf{V}\mathbf{D}_0$, then both APHs are equivalent, i.e., are different representations of the same distribution [4].

We use one specific matrix $\mathbf{V}$ to realize the expanded APH. Let $\mathbf{V} = \begin{pmatrix} \mathbf{I} \\ \mathbf{0}\ 1 \end{pmatrix}$ be the $(n+1) \times n$ transformation matrix describing the equivalence. $(\mathbf{p}, \mathbf{D}_0)$ is the original n-dimensional APH. We define a $n+1$ dimensional APH $(\mathbf{p}', \mathbf{D}_0')$ with

$$\mathbf{p}'(i) = \begin{cases} \mathbf{p}(i) & \text{if } i < n \\ 0 & \text{if } i = n \\ \mathbf{p}(n) & \text{if } i = n+1 \end{cases} \quad \text{and } \mathbf{D}_0'(i,j) = \begin{cases} \mathbf{D}_0(i,j) & \text{if } i < n \text{ and } j < n \\ \mathbf{a}(i)\mathbf{D}_0(i,n) & \text{if } i < n \text{ and } j = n \\ (1-\mathbf{a}(i))\mathbf{D}_0(i,n) & \text{if } i < n \text{ and } j = n+1 \\ \mathbf{D}_0(n,n) & \text{if } i,j \in \{n, n+1\} \text{ and } i = j \\ 0 & \text{if } i = n \text{ and } j = n+1 \end{cases} \tag{14}$$

for some vector $\mathbf{a}$ of length n with elements out of $[0, 1]$. It is easy to show that the required relation between the two APHs holds. Although the above transformation works for all vectors $\mathbf{a}$ with elements from $[0, 1]$ we assume in the sequel that $\mathbf{a} = \mathbb{1}$.

Let $(\mathbf{D}_0, \mathbf{D}_1)$ be a MAP expanded from $(\mathbf{p}, \mathbf{D}_0)$ with a maximal first joined moments $\mu^*_{1,1}$. We define a MAP $(\mathbf{D}_0', \mathbf{D}_1')$ with matrix $\mathbf{D}_0'$ as in (14) with vector $\mathbf{a} = \mathbb{1}$ and

$$\mathbf{D}_1'(i,j) = \begin{cases} \mathbf{D}_1(i,j) + \frac{\mathbf{D}_1(n,j)}{\sum_{k=1}^{n-1}\mathbf{D}_1(n,k)}\mathbf{D}_1(i,n) & \text{if } i < n \text{ and } j < n \\ 0 & \text{if } i < n \text{ and } j \geq n \\ \frac{\sum_{k=1}^{n}\mathbf{D}_1(n,k)}{\sum_{l=1}^{n-1}\mathbf{D}_1(n,l)}\mathbf{D}_1(n,j) & \text{if } i = n \text{ and } j < n \\ 0 & \text{if } i = j = n \\ 0 & \text{if } i \leq n \text{ and } j = n+1 \\ \lambda_n & \text{if } i = j = n+1 \end{cases}$$

Observe that if the denominator in the first case becomes zero, then also the numerator is zero and the second term of the sum becomes zero. We assume that $\mathbf{D}_1(n, j) > 0$ for at least one $j < n$. If this is not the case, then the transformation cannot be applied and one should use the alternative expansion described below.

In the online companion [5] we present the complete proofs which show that $(\mathbf{D}_0', \mathbf{D}_1')$ describes a valid MAP and that $\mu^*_{1,1} \leq \mu'_{1,1}$ holds.

The above transformation results in a representation with a first joint moment which is not smaller than the first joint moment of the original APH. The transformation cannot be applied if $\mathbf{D}_1(n, j) = 0$ for all $j < n$ or $\mathbf{D}_1(i, n) = 0$ for all $i < n$. This is for example the case for hyperexponential APHs. In this case we introduce a second expansion approach that adds two states by first representing one of the exponential phases by a two state representation (2) and afterwards cloning the second state. This results in the following representation of an exponential distribution with three states.

$$\mathbf{p}' = \left(\frac{\lambda_2 - \lambda_1}{\lambda_2}, 0, \frac{\lambda_1}{\lambda_2}\right) \text{ and } \mathbf{D}'_0 = \begin{pmatrix} -\lambda_1 & \lambda_1 & 0 \\ 0 & -\lambda_2 & 0 \\ 0 & 0 & -\lambda_2 \end{pmatrix} \tag{15}$$

By simple calculations it can be shown that the maximum first joined moment is reached by a MAP with

$$\mathbf{D}'_1 = \begin{pmatrix} 0 & 0 & 0 \\ \lambda_2 & 0 & 0 \\ 0 & 0 & \lambda_2 \end{pmatrix}$$

for $\lambda_2 = 2\lambda_1$. In this case, the lag 1 autocorrelation coefficient equals 0.25. The above construction can, of course, be extended, by representing the last exponential phase by a three state representation with rate $\lambda_3 = 2\lambda_2$ such that the overall distribution has 5 states and a lag 1 autocorrelation of 0.28125. This representation of correlated exponentially distributed random variables using a MAP is an alternative to the approach proposed in [1] which uses bivariate APHs.

Representation (15) can be used to substitute an arbitrary state in a APH. If the last state is substituted then the ordering of states with increasing rates is kept. If another state is substituted, it might be necessary to compute for the resulting APH the canonical representation, because the rates are no longer increasing, and start the transformation process afterwards again. This is easy for hyperexponential distributions but requires the use of the simulated annealing based optimization in general. A state which is substituted by the 3 state representation should have a larger initial probability to assure an effect on the flexibility of the representation.

Example: We apply the transformations to the APHs (12) and (13). Example (12) has a hyperexponential representation with 3 states and can be extended to a MAP with lag 1 autocorrelation coefficient $\rho_1 = 0.21429$. Due to the hyperexponential structure the first transformation cannot be applied. If we substitute the third state for the second transformation, we obtain a representation with 5 states which can be expanded to a MAP with $\rho_1 = 0.21684$ which is only a minor expansion. If the first state is substituted, the resulting APH can be expanded to MAP with $\rho_1 = 0.33641$. However, in this case the representation resulting from the expansion has to be first transformed to canonical form and then the simulated annealing based optimization has to be applied to find the right representation.

Example (13) cannot be transformed to an hyperexponential representation and MAP which can be expanded from the representation has $\rho_1 = 0.0751$. The first transformation results in an APH with 4 states which can be expanded to a MAP with $\rho_1 = 0.0889$. The second transformation cannot be applied to the last state since the initial probability of this state is 0. Expansion of the first state results in an APH with 5 states which can be transformed to a representation which is expanded to a MAP with $\rho_1 = 0.18079$. Again the representation resulting from the transformation first has to be transformed to the canonical form and then the simulated annealing based optimization is applied.

6 Examples

To demonstrate the effect of different PH transformations on the parameter estimation of MAPs we fitted various PHDs to real-world traffic data, transformed the resulting distributions and used it as input for a MAP fitting approach. In particular, the following experiment setup was used: We chose two different traces with interarrival times of network packets, i.e. the well-known benchmark trace *LBL-TCP-3* [19] from the Internet Traffic Archive (http://ita.ee.lbl.gov/) and the newer trace (*TUDo*) that was recorded at the computer science department at TU Dortmund [14]. Acyclic PHDs were fitted to the two traces using the approach from [3] that estimates the parameters of the distribution according to the empirical moments from the trace. The approach results in a PHD in series canonical form, such that no correlation can be modeled with this representation. Consequently, we applied different transformations to the representation and used the resulting distributions as input for a two-step MAP fitting approach that constructs matrix $\mathbf{D}_1$ according to the empirical autocorrelation coefficients of the trace file [14]. As a measure to assess the fitting quality we present plots of the autocorrelation coefficients and the likelihood values of the resulting MAPs. The transformations applied to the PHD are the ones described in Sect. 4, i.e. simulated annealing (denoted by `Rand` in the plots) and the transformation into a hyperexponential distribution (denoted by `HExp`). If necessary, we expanded the state space as described in Sect. 5 (denoted by an additional + in the plots). For comparison we also used the transformation from [3] (denoted by `QEST`) that chooses two phases i and j in each iteration and sets $\delta = 0.9 \cdot \delta^*$ where δ^* is the upper bound of the interval in Eq. 5. It was shown in [16] that this transformation usually does not find the optimal representation.

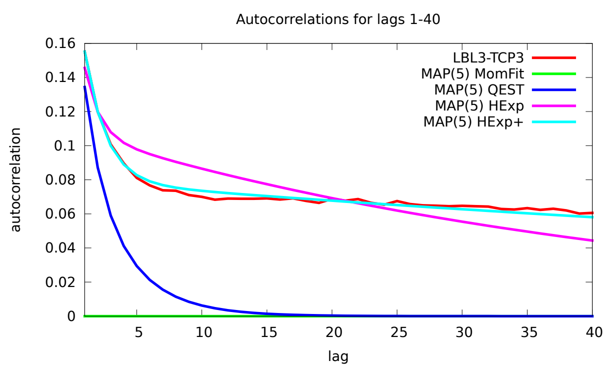
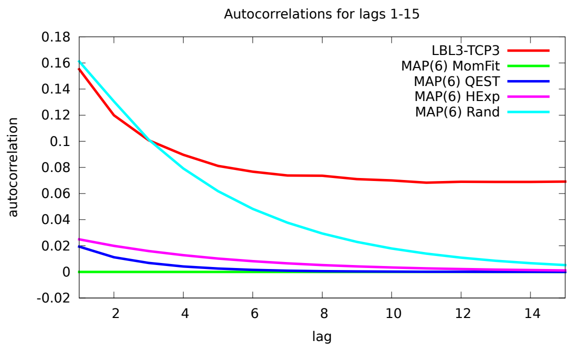


Figure 1: Autocorrelation coefficients for the Trace *LBL-TCP-3* and PHDs of order 5

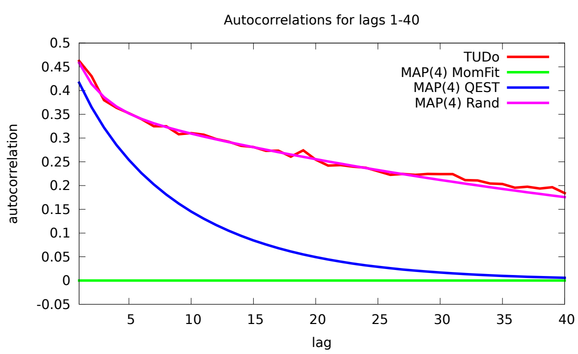
Fig. 1 shows the results for a PHD of order 5 fitted to the trace *LBL-TCP-3*. The plot contains the autocorrelation for the trace and 4 MAPs that resulted from expanding the transformed PH representation into a MAP. The MAPs are sorted in increasing order according to the maximal lag 1 autocorrelation that could be modeled with the corresponding PHD. For the original untransformed representation (labeled `MomFit`) of course no autocorrelation could be modeled. The transformation from [3] (`QEST`) and the transformation into a hyperexponential distribution (`HExp`) both resulted in a representation that allows for modeling a lag 1 correlation that is (slightly) larger than the one of the trace, albeit `HExp` resulted in the larger of the two correlation values. Consequently, the hyperexponential representation allowed for a better estimation of the correlation from the trace. Since the transformation into hyperexponential representation worked for this distribution we omitted the transformation using simulated annealing and instead expanded the state space by adding additional phases, which further increased the maximal lag 1 correlation and allowed for the best estimation of the autocorrelation coefficients.



MAP	Likelihood
Momfit	-1.69885e+06
QEST	-1.69179e+06
HExp	-1.68769e+06
Rand	-1.65226e+06

Figure 2: Autocorrelation coefficients and likelihood values for the Trace *LBL-TCP-3* and PHDs of order 6

Fig. 2 shows the autocorrelation coefficients and the likelihood values for the trace *LBL-TCP-3* and models of order 6. The transformation from [3] and the transformation into the hyperexponential distribution failed to provide a representation that is adequate for MAP fitting. Since the PHD could not be transformed into a hyperexponential representation we applied the transformation using simulated annealing that resulted in the representation with the largest maximal lag 1 autocorrelation that was suited best for MAP fitting. Fig. 2 also shows the likelihood values which confirm the observations from the plot of the autocorrelation coefficients.



Model	(max) lag 1 corr.
Trace *TUDo*	0.46254
Momfit	0.0
QEST	0.4312
Rand	0.4997

Figure 3: Autocorrelation coefficients and maximal lag 1 correlation for the Trace *TUDo* and PHDs of order 4

For the last example shown in Fig. 3 the trace *TUDo* was used. Since the original distribution could not be transformed into hyperexponential representation we omitted its curve in the plot. Fig. 3 shows the maximal lag 1 correlation that could be achieved with the representations resulting from the different transformation approaches. As one can see the maximal correlation resulting from the simulated annealing approach is a little larger than the correlation from the trace, while the maximal correlation from the transformation from [3] is slightly lower. However, this has a large impact when fitting the autocorrelation coefficients as shown in the plot.

The experiments show that PH transformations can help to find an adequate representation of a PHD for a subsequent MAP fitting step. Moreover, the results suggest that a larger value for the maximal lag 1 correlation also allows for a better fitting of higher lags, i.e. the representations appear to be more flexible with increasing maximal correlation value.

7 Conclusions

In this article we considered different transformations and representations for acyclic PHDs to compute a representation that allows the expansion to a MAP with a maximal first joint moment. For PHDs of order 2 we distinguished two cases depending on the values of the initial probabilities and transition rates and presented optimal representations for both cases.

For acyclic PHDs with an arbitrary number of states we showed that the hyperexponential representation results in the maximal first joint moment for a subsequent MAP fitting step. The optimal representation for PHDs that cannot be transformed into a hyperexponential distribution remains subject to future research. However, we presented a heuristic approach to increase the maximal first joint moment in these cases that showed good experimental results. Moreover, we presented methods to increase the maximal first joint moment by adding additional states that can be used in cases where the transformations do not result in a sufficiently large first joint moment.

Our considerations focused on the first joint moment which is related to the lag 1 autocorrelation coefficient. Although our experiments suggest that a representation that allows for expansion into a MAP with a maximal lag 1 correlation coefficient is favorable in general, i.e. it also provides a high flexibility to capture correlation coefficients with a higher lag, this of course needs further investigation.

References

[1] M. Bladt and B. F. Nielson. On the construction of bivariate exponential distributions with an arbitrary correlation coefficient. *Stochastic Models*, 26:295–308, 2010.

[2] P. Buchholz and J. Kriege. Equivalence Transformations for Acyclic Phase Type Distributions. Technical Report 827, Dep. of Informatics, TU Dortmund, 2009.

[3] P. Buchholz and J. Kriege. A heuristic approach for fitting MAPs to moments and joint moments. In *QEST*, pages 53–62, 2009.

[4] P. Buchholz and M. Telek. Stochastic Petri nets with matrix exponentially distributed firing times. *Performance Evaluation*, 67(12):1373–1385, 2010.

[5] Peter Buchholz, Iryna Felko, and Jan Kriege. Online Companion to the Paper Transformation of Acyclic Phase Type Distributions for Correlation Fitting. Available online under `http://www4.cs.tu-dortmund.de/download/kriege/publications/asmta2013_online_companion.pdf`, 2013.

[6] Peter Buchholz and Andriy Panchenko. A two-step EM algorithm for MAP fitting. In *19th International Symposium on Computer and Information Sciences (ISCIS)*, pages 217–227, 2004.

[7] A. Corana, M. Marchesi, C. Martini, and S. Ridella. Minimizing multimodal functions of continuous variables with the "simulated annealing" algorithm. *ACM Trans. Math. Softw.*, 13(3):262–280, 1987.

[8] A. Cumani. On the canonical representation of homogeneous Markov processes modeling failure-time distributions. *Micorelectronics and Reliability*, 22(3):583–602, 1982.

[9] Mark Fackrell. Modelling healthcare systems with phase-type distributions. *Health Care Management Science*, 12(1):11–26, 2009.

[10] A. Horváth and M. Telek. Phfit: A general purpose phase type fitting tool. In *Performance Tools 2002*, volume 2324 of *LNCS*, pages 82–91. Springer, 2002.

[11] G. Horváth, M. Telek, and P. Buchholz. A MAP fitting approach with independent approximation of the inter-arrival time distribution and the lag-correlation. In *QEST*, 2005.

[12] V. B. Iversen and B. F. Nielsen. Some properties of Coxian distributions with applications. In *Modeling Techniques and Tools for Performance Analysis*, pages 61–66. Elsevier, 1986.

[13] A. Klemm, C. Lindemann, and M. Lohmann. Modeling IP traffic using the batch Markovian arrival process. *Perform. Eval.*, 54(2):149–173, 2003.

[14] Jan Kriege and Peter Buchholz. An Empirical Comparison of MAP Fitting Algorithms. In *Proceedings of MMB & DFT 2010*, volume 5987 of *LNCS*, pages 259–273. Springer, 2010.

[15] L. Lipsky. *Queueing Theory: A Linear Algebraic Approach*. Springer, 2010.

[16] A. Mészáros and M. Telek. A two-phase MAP fitting method with APH interarrival time distribution. In *Winter Simulation Conference*. ACM, 2012.

[17] M. F. Neuts. A versatile Markovian point process. *Journ. of Applied Probability*, 16, 1979.

[18] M. F. Neuts. *Matrix-geometric solutions in stochastic models*. Johns Hopkins University Press, 1981.

[19] V. Paxson and S. Floyd. Wide-area traffic: The failure of Poisson modeling. *IEEE/ACM Transactions on Networking*, 3:226–244, 1995.

[20] M. Telek and G. Horváth. A minimal representation of Markov arrival processes and a moments matching method. *Perform. Eval.*, 64(9-12):1153–1168, 2007.

[21] M. Telek et al. Butools - program packages. webspn.hit.bme.hu/~butools/.

[22] A. Thümmler, P. Buchholz, and M. Telek. A novel approach for phase-type fitting with the EM algorithm. *IEEE Trans. Dep. Sec. Comput.*, 3(3):245–258, 2006.