On the Representation of Correlated Exponential Distributions by Phase Type Distributions

Peter Buchholz
Informatik IV, TU Dortmund
peter.buchholz@tu-dortmund.de

ABSTRACT

In this paper we present results for bivariate exponential distributions which are represented by phase type distributions. The paper extends results from previous publications [3, 11] on this topic by introducing new representations that require a smaller number of phases to reach some correlation coefficient and introduces different ways to describe correlation between exponentially distributed random variables.

1. INTRODUCTION

Phase type distributions (PHDs) are used to describe non-exponential distributions in stochastic models that can be mapped on *Continuous Time Markov Chains* (CTMCs) and solved by simulation or, preferably, by numerical techniques [15, 16]. A large number of papers on PHDs, their properties, the estimation of parameters, and their application in stochastic modeling exists [1, 6]. PHDs describe uncorrelated event streams whereas their extension *Markovian Arrival Processes* (MAPs) are applied to model autocorrelated sequences. In many stochastic models, correlation does not only occur between the events of one stream, instead the times of different events are correlated. E.g., inter-arrival times and service times in a queuing system are correlated [7], failures times of components are correlated [9] or processing times of parallel jobs are correlated [14]. In these cases, correlation between different PHDs has to be modeled. To describe the required correlation, PHDs that run in parallel or sequentially have to be defined to realize correlation of processing times of parallel jobs or of sequential steps of one job.

In this paper, we consider correlated exponential distributions which are the base for PHDs and MAPs. Results from two previous papers on bivariate exponential distributions [3, 11] are extended. New representations for bivariate exponential distributions described by PHDs are defined that require fewer states to represent a given coefficient of correlation than the technqiues presented in [3, 11].

The paper is structured as follows. In the next section PHDs and phase type representations of exponential distributions are introduced. Section 3 reviews related work. Then, a step-wise approach to generate phase type representations of exponential distributions with an increasing number of phases is defined. The following two sections describe how a maximal and minimal coefficient of correlation can be achieved by phase type representations of exponential distributions. Then a small example is presented and the paper is concluded. A more detailed version of this paper including the proofs can be found online [4].

2. BASIC MODEL

A PHD is described by an initial distribution vector $\boldsymbol{\pi}$ and a sub-generator $\boldsymbol{D}$ with only transient states [6, 15]. Consequently, the real part of all eigenvalues of matrix $\boldsymbol{D}$ is negative implying that the matrix is non-singular and the inverse matrix is non-positive. Furthermore, we assume $\boldsymbol{\pi}\mathbb{1} = 1$ and $\boldsymbol{d} = -\boldsymbol{D}\mathbb{1}$ where $\mathbb{1}$ is a column vector of all 1s. Let n be the number of phases of the PHD.

For some PHD $(\boldsymbol{\pi}, \boldsymbol{D})$, matrix $\boldsymbol{M} = -\boldsymbol{D}^{-1} \geq \boldsymbol{0}$ exists, $\boldsymbol{M}(i,j)$ is the mean time the PHD stays in state j before absorption if the current state is i and $\boldsymbol{m} = \boldsymbol{M}\mathbb{1}$ is a vector containing the first moments of the time to absorption conditioned on the initial state. Let $\boldsymbol{\psi}$ be the vector of exit probabilities from the states, containing the probabilities that the PHD is left from a specific state. The vector can be computed as follows [13, Parag. 3.5].

$$\boldsymbol{B} = \boldsymbol{M}\,\mathrm{diag}\,(\boldsymbol{d}) \text{ and } \boldsymbol{\psi} = \boldsymbol{\pi}\boldsymbol{B}. \tag{1}$$

Another quantity which is of interest is the expected absorption time conditioned on the exit state, i.e., the last transient state before absorption. Let $\boldsymbol{a}$ be the vector containing the conditional absorption times. The vector can be computed as

$$\boldsymbol{a}(i) = \frac{1}{\boldsymbol{\psi}(i)}\boldsymbol{\pi}\boldsymbol{M}\boldsymbol{B}(:,i) \text{ for } \boldsymbol{\psi}(i) > 0 \tag{2}$$

where $\boldsymbol{B}(:,i)$ is the ith column of matrix $\boldsymbol{B}$.

We consider *Acyclic Phase Type Distributions* (APHDs) which are described by the initial distribution $\boldsymbol{\pi}$ and a matrix $\boldsymbol{D}$ that can be permuted to an upper triangular matrix. In contrast to general PHDs, APHDs can be transformed to some canonical form by equivalence transformations. We consider in the following two canonical forms [8]. The first is given by

$$\boldsymbol{\pi} = (\pi_1, \ldots, \pi_n),$$
$$\boldsymbol{D} = \begin{pmatrix} -\mu_1 & \mu_1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \vdots & \vdots & \vdots & -\mu_{n-1} & \mu_{n-1} \\ 0 & \cdots & \cdots & 0 & -\mu_n \end{pmatrix} \tag{3}$$

with $\mu_1 \leq \mu_2 \leq \ldots \leq \mu_n$. The second canonical form equals

$$\boldsymbol{\pi} = (1, 0, \ldots, 0)\,,$$
$$\boldsymbol{D} = \begin{pmatrix} -\mu_1 & \mu_{1,2} & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \vdots & \vdots & \vdots & -\mu_{n-1} & \mu_{n-1,n} \\ 0 & \cdots & \cdots & 0 & -\mu_n \end{pmatrix} \quad (4)$$

where $\mu_1 \geq \mu_2 \geq \ldots \geq \mu_n$ and $\mu_i \geq \mu_{i,i+1}$. The Laplace transform of the distribution described by the first or second canonical form is given by

$$L_1(s) = \sum_{i=1}^{n} \pi_i \prod_{j=i}^{n} \frac{\mu_j}{\mu_j + s} \quad \text{and}$$
$$L_2(s) = \sum_{i=1}^{n} \left(1 - \frac{\mu_{i,i+1}}{\mu_i}\right) \left(\prod_{j=1}^{i-1} \frac{\mu_{j,j+1}}{\mu_j}\right) \left(\prod_{k=1}^{i} \frac{\mu_k}{\mu_k + s}\right) \quad (5)$$

where $\mu_{n,n+1} = 0$. To represent some exponential distribution with rate λ by an APHD the Laplace transform of the corresponding representation has to equal $\lambda/(\lambda+s)$. We denote a representation as normalized if $\lambda = 1$ which implies $E(X) = 1$. By multiplication of matrix $\boldsymbol{D}$ with $\lambda^{-1} > 0$ we obtain a normalized representation. Thus, it is sufficient to consider only normalized representations in the sequel.

If we assume that a PHD is immediately started again with vector $\boldsymbol{\pi}$ after an absorption, we obtain the stationary vector at random times $\boldsymbol{\phi}$ as the solution of the following set of equations.

$$\boldsymbol{\phi}\,(\boldsymbol{D} + \boldsymbol{d}\boldsymbol{\pi}) = \boldsymbol{0} \text{ and } \boldsymbol{\phi}\mathbb{1} = 1. \quad (6)$$

For some random variable X we denote by $(\boldsymbol{\pi}_X, \boldsymbol{D}_X)$ a PHD representation. If two PHDs $(\boldsymbol{\pi}_X, \boldsymbol{D}_X)$ and $(\boldsymbol{\pi'}_X, \boldsymbol{D'}_X)$ are different representations for the same random variable we use the notation $(\boldsymbol{\pi}_X, \boldsymbol{D}_X) \sim (\boldsymbol{\pi'}_X, \boldsymbol{D'}_X)$. For X with representation $(\boldsymbol{\pi}_X, \boldsymbol{D}_X)$, the relations $E(X) = (\boldsymbol{\phi}_X \boldsymbol{d}_X)^{-1} = \boldsymbol{\pi}_X \boldsymbol{m}_X$ hold.

Correlated Random Variables

For two random variables X and Y, the correlation and coefficient of correlation are defined as

$$C_{X,Y} = E(XY) - E(X)E(Y) \text{ and } \rho_{X,Y} = \frac{C_{X,Y}}{\sigma_X \sigma_Y} \quad (7)$$

where σ_X, σ_Y are the standard deviations. For the exponential distribution $\sigma_X = E(X)$ which implies for two normalized representations $\rho_{X,Y} = E(XY) - 1$. Let Y and X be two exponentially distributed random variables. We assume that $(\boldsymbol{\pi}_X, \boldsymbol{D}_X)$ and $(\boldsymbol{\pi}_Y, \boldsymbol{D}_Y)$ are APHDs representing normalized random variables. By multiplying the matrices $\boldsymbol{D}_X$ and $\boldsymbol{D}_Y$ with λ_X and λ_Y, respectively, the rates of the distributions can be shifted.

We describe two different possibilities to combine PHDs. The bivariate representation used in [3, 11] can be interpreted as a sequential composition of PHDs. Let $(\boldsymbol{\pi}_X, \boldsymbol{D}_X)$ and $(\boldsymbol{\pi}_Y, \boldsymbol{D}_Y)$ be two PHDs of order n_X and n_Y describing random variables X and Y, then the following absorbing CTMC defines the sequential combination.

$$(\boldsymbol{\pi}_X, \boldsymbol{0}, 0)\,, \begin{pmatrix} \boldsymbol{D}_X & \mathrm{diag}(\boldsymbol{d}_X)\boldsymbol{\Psi}_{X,Y} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{D}_Y & \boldsymbol{d}_Y \\ \boldsymbol{0} & \boldsymbol{0} & 0 \end{pmatrix} \quad (8)$$

where $\boldsymbol{\Psi}_{X,Y}$ is a non-negative $n_X \times n_Y$ matrix with $\boldsymbol{\Psi}_{X,Y}\mathbb{1} = \mathbb{1}$ and $\boldsymbol{\psi}_X \boldsymbol{\Psi}_{X,Y} = \boldsymbol{\pi}_Y$. The time of the transition from the first block into the second block determines the value of the first random variable, the exit state from where the first block is left defines the initial distribution for the second random variable, and the time of the transition into the absorbing state determines the sum of both random variables and thus the value of the second random variable. The coefficient of correlation of X and Y is then given by

$$\rho_{X,Y} = \frac{\left(\sum_{i=1}^{n_X} \boldsymbol{\psi}_X(i)\boldsymbol{a}_X(i) \sum_{j=1}^{n_Y} \boldsymbol{\Psi}_{X,Y}(i,j)\boldsymbol{m}_Y(j)\right) - E(X)E(Y)}{\sigma_X \sigma_Y} \quad (9)$$

The second composition is a parallel composition. In this case, two (or more) PHDs are started jointly. The combined PHDs built again a PHD described by the following absorbing CTMC.

$$(\boldsymbol{\pi}_{X,Y}, \boldsymbol{0}, \boldsymbol{0}, 0)\,,$$
$$\begin{pmatrix} \boldsymbol{D}_X \oplus \boldsymbol{D}_Y & \boldsymbol{d}_X \otimes \boldsymbol{I}_{n_Y} & \boldsymbol{I}_{n_X} \otimes \boldsymbol{d}_Y & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{D}_Y & \boldsymbol{0} & \boldsymbol{d}_Y \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{D}_X & \boldsymbol{d}_X \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & 0 \end{pmatrix} \quad (10)$$

$\boldsymbol{\pi}_{X,Y}$ is a probability distribution which observes

$$\boldsymbol{\pi}_{X,Y}\,(\boldsymbol{I}_{n_X} \otimes \mathbb{1}_{n_Y}) = \boldsymbol{\pi}_X \quad \text{and} \quad \boldsymbol{\pi}_{X,Y}\,(\mathbb{1}_{n_X} \otimes \boldsymbol{I}_{n_Y}) = \boldsymbol{\pi}_Y,$$

where $\boldsymbol{I}_n$ is the $n \times n$ identity matrix and $\mathbb{1}_n$ a column vector of 1s of length n. The time of a transition from the first into the second block describes the value of X if it is smaller than the value of Y, a transition from the first into the third block describes the value of Y if it is smaller than X and a transition into the absorbing state defines the maximum of both random variables. The coefficient of correlation equals

$$\rho_{X,Y} = \frac{\left(\sum_{i=1}^{n_X} \sum_{j=1}^{n_Y} \boldsymbol{\pi}_{X,Y}((i-1)*n_Y + j)\boldsymbol{m}_X(i)\boldsymbol{m}_Y(j)\right) - E(X)E(Y)}{\sigma_X \sigma_Y} \quad (11)$$

Obviously, the coefficients of correlation that can be achieved for two random variables X and Y that are composed sequentially or in parallel depend on the choice of matrix $\boldsymbol{\Psi}_{X,Y}$ and vector $\boldsymbol{\pi}_{X,Y}$, respectively. Additionally, the value depends also on the representation of the random variables by PHDs. Finding an appropriate representation to reach a given coefficient of correlation with a small dimension of the PHDs is a challenging problem that will be investigated in this paper (for further details see [4]).

3. RELATED WORK

Based on the model for multivariate PHDs, [3] developed a model of bivariate exponential PHDs using the representation (8). To generate correlated exponential distributions, matrix $\boldsymbol{D}_X$ is of the second canonical form and $\boldsymbol{D}_Y$ of the first canonical form. Furthermore, $n_X = n_Y$ and $\mu_i = i$ $(i = 1, \ldots, n_i)$. In this case the minimal coefficient of correlation $\rho^{-(n)}$ is reached for $\boldsymbol{\Psi}^-_{X,Y} = \boldsymbol{I}$ and the maximal coefficient of correlation $\rho^{+(n)}$ is achieved for matrix $\boldsymbol{\Psi}^+_{X,Y}(i,j) = 1$ if $j = n_X - i + 1$ and 0 otherwise. For positive correlation, the maximal coefficient of correlation for order n PHDs equals then $\rho^{+(n)} = 1 - \frac{1}{n}\sum_{i=1}^{n} \frac{1}{i}$ and the minimal coefficient of correlation equals $1 - \sum_{i=1}^{n} \frac{1}{i^2}$.

Thus, $\lim_{n\to\infty}\rho^{+(n)} = 1$ and $\lim_{n\to\infty}\rho^{-(n)} = 1 - \frac{\pi^2}{6}$, the minimal correlation coefficient for bivariate exponential distributions. Any coefficient of correlation between $\rho^{-(n)}$ and $\rho^{+(n)}$ can be reached with two APHDs of order n by mixing $\boldsymbol{\Psi}^-_{X,Y}$ and $\boldsymbol{\Psi}^+_{X,Y}$ [3].

The results of [3] are extended in [11]. The APHDs used in [3] to represent exponential distributions according to (8) have the following properties.

i) $\boldsymbol{D}_X\mathbb{1} = -\mathbb{1}$,

ii) $\mathbb{1}^T\boldsymbol{D}_Y = -\mathbb{1}^T$,

iii) $\mathbb{1}^T\boldsymbol{D}_X \le \boldsymbol{0}$,

iv) $\mathbb{1}^T\boldsymbol{\Psi}_{X,Y} = \mathbb{1}^T$.

[11] shows that among all phase type representations of a fixed order n that observe i)-iv), the representations proposed in [3] have the smallest, respectively, largest coefficient of correlation. However, the properties are not mandatory to represent a normalized exponential distribution by a PHD or an APHD. In this paper, we show that relaxing some of the properties allows us to generate representations which can be used to reach a larger coefficient of variation with a given number of phases.

4. CONSECUTIVE GENERATION

To represent an exponential distribution by an APHD in canonical form, the Laplace transforms of both have to coincide. The following theorem shows that this determines the rate of the first phase of an APHD in the first canonical form.

THEOREM 1. *If a normalized exponential distribution is represented by an APHD in canonical form (3), then $\mu_1 = 1$ has to hold.*

The proof of the theorem shows that $\boldsymbol{\pi}_t(i) = \boldsymbol{\pi}(i)e^{-t}$ holds for all states of the APHD in the first canonical form which represents an exponential distribution. This also implies that $\boldsymbol{\phi} = \boldsymbol{\pi}$ holds in this case.

Obviously, the coefficient of correlation of two exponential distributions composed as in (8) depends on matrix $\boldsymbol{\Psi}_{X,Y}$ which can be computed from a system of linear equations, if both PHDs and $\rho_{X,Y}$ are known. For given phase type representations the minmal or maximal coefficient of correlation can be computed from the following linear program.

$$\begin{aligned}&\rho^{\pm}_{(\boldsymbol{\pi}_X,\boldsymbol{D}_X)(\boldsymbol{\pi}_Y,\boldsymbol{D}_Y)} = \\ &\quad \min/\max \sum_{i=1}^{n_X}\sum_{j=1}^{n_Y} \boldsymbol{\Psi}_{X,Y}(i,j)\boldsymbol{\psi}_X(i)\boldsymbol{a}_X(i)\boldsymbol{m}_Y(j) - 1\\ &\text{s.t. } \boldsymbol{\psi}_X\boldsymbol{\Psi}_{X,Y} = \boldsymbol{\pi}_Y, \boldsymbol{\Psi}_{X,Y} \ge \boldsymbol{0} \text{ and } \boldsymbol{\Psi}_{X,Y}\mathbb{1} = \mathbb{1}.\end{aligned} \tag{12}$$

A similar LP can be derived for the composition (10) to compute initial vector $\boldsymbol{\pi}_{X,Y}$ that minimizes or maximizes the coefficient of correlation.

$$\begin{aligned}&\varrho^{\pm}_{(\boldsymbol{\pi}_X,\boldsymbol{D}_X)(\boldsymbol{\pi}_Y,\boldsymbol{D}_Y)} = \\ &\min/\max \sum_{i=1}^{n_X}\sum_{j=1}^{n_Y} \boldsymbol{\pi}_{X,Y}(i,j)\boldsymbol{m}_X(i)\boldsymbol{m}_Y(j) - 1\\ &\text{s.t. } \sum_{h=1}^{n_Y}\boldsymbol{\pi}_{X,Y}(i,h) = \boldsymbol{\pi}_X(i) \text{ and } \sum_{h=1}^{n_Y}\boldsymbol{\pi}(h,j) = \boldsymbol{\pi}_Y(j)\end{aligned} \tag{13}$$

Both LPs depend on different quantities of the APHDs. In the first LP (12), the spread of the absorption times depending on the exit state of the first random variable and the absorption times depending on the entry state are relevant, whereas the second LP (13) considers only absorption times depending on the entry state.

Much more challenging is the generations of appropriate PHDs that allow one to minimize or maximize the correlation for given orders of the PHDs. Since the first canonical representation has only a single output state, $\boldsymbol{a}_X(n_X) = E(X)$, and the second canonical form has a single input state such that $m_X(1) = E(X)$, for a sequential composition an APHD in the second canonical form has to be combined with a PHD in the first canonical form and for a parallel composition two APHDs of the first canonical form have to be composed to achieve positive or negative correlation for the bivariate distribution.

THEOREM 2. *For some PHD $(\boldsymbol{\pi}, \boldsymbol{D})$ with the vectors $\boldsymbol{m}$, $\boldsymbol{\psi}$ and $\boldsymbol{a}$ an PHD $(\boldsymbol{\pi}', \boldsymbol{D}')$ can be generated by setting $\boldsymbol{\pi}'(i) = \boldsymbol{\psi}(i)$, $\mu'_i = \mu_i$ and $\mu'_{i,j} = \mu_{j,i}\frac{\phi(j)}{\phi(i)}$, then $(\boldsymbol{\pi}, \boldsymbol{D}) \sim (\boldsymbol{\pi}', \boldsymbol{D}')$ and the following relations hold*

$$\boldsymbol{m}'(i) = \boldsymbol{a}(i),\ \boldsymbol{\psi}'(i) = \boldsymbol{\pi}(i) \text{ and } \boldsymbol{a}'(i) = \boldsymbol{m}(i).$$

If one applies the theorem to an APHD with an upper triangular matrix $\boldsymbol{D}$, the resulting matrix $\boldsymbol{D}'$ is lower triangular but can be easily transformed into an APHD with an upper triangular matrix by swapping the states which implies that the indices change, i.e., state i in the original APHD corresponds to state $n - i + 1$ in the transformed APHD. The following theorem shows that Theorem 2 transforms one canonical form in the other.

THEOREM 3. *If Theorem 2 is applied to an APHD $(\boldsymbol{\pi}, \boldsymbol{D})$ in canonical form (3), then the resulting APHD $(\boldsymbol{\pi}', \boldsymbol{D}')$ is in canonical form (4) after reversing the order of the states.*

The theorems imply that if an optimal representation for two PHDs is available, in the sense that it maximizes/minimizes the objective function in (12) or (13), then the representation is also optimal for the other LP after transforming the representation for X using Theorem 2.

Expansion of APHDs

We now consider the consecutive generation of APHDs describing exponential distributions. To generate an APHD representation of an exponential distribution with dimension $n + 1$ from an exponential distribution of dimension n, we append a single phase resulting in a representation $(\boldsymbol{\pi}', \boldsymbol{D}')$. The new APHD is then generated as follows.

$$\boldsymbol{\pi}' = ((1-p)\boldsymbol{\pi}, p)\,,\ \boldsymbol{D}' = \begin{pmatrix} \boldsymbol{D} & \boldsymbol{f} \\ \boldsymbol{0} & -\mu \end{pmatrix} \quad \text{where } \boldsymbol{d} \ge \boldsymbol{f}, p \in [0, 1). \tag{14}$$

Let $\boldsymbol{d}' = -\boldsymbol{D}'\mathbb{1}$ and $\boldsymbol{\pi}'_t = \boldsymbol{\pi}'e^{-\boldsymbol{D}'t}$. $(\boldsymbol{\pi}', \boldsymbol{D}')$ represents an exponential distribution with rate λ if and only if $\frac{\boldsymbol{\pi}'_t\boldsymbol{d}'}{\boldsymbol{\pi}'_t\mathbb{1}} = \lambda$ for all $t \ge 0$. The following theorem shows that the choice of $\boldsymbol{f}$ and μ suffers severe restrictions.

THEOREM 4. *For the generation of an exponential distribution of dimension $n + 1$ with rate 1 from an APHD of an exponential distribution of order n according to (14), $\boldsymbol{f} = (1 - q)\boldsymbol{d}$ for $q \in [0, 1]$ and $\mu = \frac{1-q+pq}{p}$ are required.*

Theorem 4 describes an approach to generate the first canonical form as a corner case, where $q = 0$ in each step. We will show that this case is optimal if the correlation should be maximized or minimized. Closed form expressions for the vectors $\boldsymbol{\pi'}$, $\boldsymbol{\psi'}$, $\boldsymbol{m'}$ and $\boldsymbol{a'}$ and a method to expand an APHD according to the second canonical form can be found in [4].

5. POSITIVE CORRELATION

We consider now APHDs that have been generated according to Theorem 4 and maximize the correlation $\rho^+_{(\boldsymbol{\pi},\boldsymbol{D})}$. Now let $\rho^{+(n)}$ be the maximal correlation which can be obtained by some APHD $(\boldsymbol{\pi},\boldsymbol{D})$ with n phases using the step-wise approach based on Theorem 4. For a fixed representation the optimal solution results in $\sum_{i=1}^n \boldsymbol{\pi}(i)\,(\boldsymbol{m}(i))^2 - 1$ where $\boldsymbol{m} = (-\boldsymbol{D})^{-1}\,\mathbb{1}$. Thus, representations have to be found to maximize the sum.

We can use the normalized representation with rate $\lambda = 1$. Obviously $\rho^{+(1)} = 0$ because $\boldsymbol{D} = (-1)$ and $\boldsymbol{\pi} = (1)$ in this case. This representation is unique and therefor optimal. Let $\rho^{+(n)}$ be the maximal coefficient of correlation which can be achieved by the step-wise construction of APHDs and let $(\boldsymbol{\pi},\boldsymbol{D})$ be the corresponding representation. Now we compute a representation $(\boldsymbol{\pi'},\boldsymbol{D'})$ that achieves $\rho^{+(n+1)}$. According to (14) and Theorem 4 we have

$$\begin{array}{ll}\boldsymbol{\pi'} = ((1-p)\boldsymbol{\pi}, p)\,, & \boldsymbol{D'} = \begin{pmatrix} \boldsymbol{D} & (1-q)\boldsymbol{d} \\ \boldsymbol{0} & -\mu \end{pmatrix} \\ \Rightarrow \boldsymbol{M'} = \begin{pmatrix} \boldsymbol{M} & \frac{1-q}{\mu}\mathbb{1} \\ \boldsymbol{0} & \frac{1}{\mu} \end{pmatrix}, & \boldsymbol{m'} = \begin{pmatrix} \boldsymbol{m} + \frac{1-q}{\mu}\mathbb{1} \\ \frac{1}{\mu} \end{pmatrix}\end{array} \tag{15}$$

where $\mu = \frac{1-q+pq}{p}$. This results in the following representation of the coefficient of correlation.

$$\begin{array}{rcl}\rho^{(n+1)} & = & \sum_{i=1}^{n} (1-p)\boldsymbol{\pi}(i)\left(\boldsymbol{m}(i) + \frac{1-q}{\mu}\right)^2 + \frac{p}{\mu^2} - 1 \\ & = & (1-p)\rho^{+(n)} + p\frac{p(1-p)(1-q)^2}{(1-q+pq)^2}\end{array} \tag{16}$$

The equation results from exploitation of the relations $\rho^{+(n)} = \sum_{i=1}^n \boldsymbol{\pi}(i)\boldsymbol{m}(i)^2 - 1$ and $\sum_{i=1}^n \boldsymbol{\pi}(i)\boldsymbol{m}(i) = \sum_{i=1}^n \boldsymbol{\pi}(i) = 1$ and from substituting the representation of μ from Theorem 4. Now consider the second term in the above sum.

$$f(p,q) = \frac{p(1-p)(1-q)^2}{(1-q+pq)^2} \;\Rightarrow\; \frac{df}{dq} = -\frac{2p^2(1-p)(1-q)}{(1-q+pq)^3} \tag{17}$$

For a fixed $p \in (0,1)$ and $q \in [0,1]$ the first derivative is negative for $q \in [0,1)$ which means that for $q = 0$ the maximum is reached for any p. Since q does not appear in the first term this also holds for $\rho^{(n+1)}$. Thus, we have

$$\begin{array}{l}\rho^{+(n+1)} = \max_{p\in(0,1)}\left((1-p)(\rho^{+(n)} + p)\right) \\ p = \frac{1-\rho^{+(n)}}{2} = \arg\max_{p\in(0,1)}\left((1-p)(\rho^{+(n)} + p)\right)\end{array} \tag{18}$$

The resulting APHD is in canonical form (3) and $\mu_n = 2/(1-\rho^{+(n)})$. Furthermore, $\rho^{+(n+1)} = \rho^{+(n)} + 0.25\left(1-\rho^{+(n)}\right)^2$ and $\lim_{n\to\infty}\rho^{+(n)} = 1$. The partial derivatives of $\rho^{(n)}$ with respect to μ_i^{-1} are given by

$$\begin{array}{l}\frac{d}{d\mu_i^{-1}} = \\ \boldsymbol{\pi}(i)\left(\mu_i\boldsymbol{m}(i) + 2\right)\boldsymbol{m}(i) + \sum_{j=1}^{i-1}\boldsymbol{\pi}(j)\boldsymbol{m}(j)\left(2 - \frac{\boldsymbol{m}(j)\mu_i}{\mu_i - 1}\right)\end{array} \tag{19}$$

If we plug in the values for μ_i resulting from (18) into (19), the derivatives for $i = 2,\ldots,n$ become 0 such that the necessary conditions for optimality are observed.

We can compare the generated APHDs with the representation used in [3] with $\mu_i = i$. Figure 1 shows the resulting values for ρ depending on n and the values for μ_n. For both representations $\lim_{n\to\infty}\rho^{+(n)} = 1$ holds. However, the representation from (18) converges slightly faster.

6. NEGATIVE CORRELATION

We now introduce a similar approach to model negative rather than positive correlation between two exponential distributions with normalized APHD representation. In contrast to positive correlation, where for the joint initial vector of two identical exponential APHs $(\boldsymbol{\pi}_X, \boldsymbol{D}_X)$ $\boldsymbol{\pi}(i,i) = \boldsymbol{\pi}_X(i)$ holds, it is not obvious which joint initial vector minimizes the correlation. It is known that, if $\boldsymbol{\pi}(i) = \boldsymbol{\pi}(n-i+1)$ for all $i = 1,\ldots,n$, then $\boldsymbol{\pi}(i,j) = \boldsymbol{\pi}(i)$ for $j = n-i+1$ and 0 otherwise results in the minimal coefficient of correlation because $\boldsymbol{m}(i) > \boldsymbol{m}(i+1)$. However, the conditions $\boldsymbol{\pi}(i) = \boldsymbol{\pi}(n-i+1)$ puts additional constraints on the class of APHDs and the representation computed for the maximal coefficient of correlation in the previous section does not belong to this class. If the step-wise approach (14) with $\mu = \frac{1-q+pq}{p}$ is applied starting with an exponential APHD of order n, we obtain the following non-linear program to compute the parameters p and q according to Theorem 4.

$$\begin{array}{l}\min_{p,q,\boldsymbol{\pi}(i,j)} \sum_{i=1}^{n+1}\sum_{j=1}^{n+1} \boldsymbol{\pi'}(i,j)\boldsymbol{m'}(i)\boldsymbol{m'}(j) \\ s.t.\quad \boldsymbol{m'}(i) = \begin{cases} \boldsymbol{m}(i) + \frac{p-pq}{1-q+pq} & \text{if } i \le n, \\ \frac{p}{1-q+pq} & \text{else,} \end{cases} \\ \qquad p \in (0,1),\; q \in [0,1] \\ \qquad \sum_{i=1}^{n+1} \boldsymbol{\pi'}(i,j) = \begin{cases} (1-p)\boldsymbol{\pi}(j) & \text{if } i \le n, \\ p & \text{else,} \end{cases} \\ \qquad \sum_{j=1}^{n+1} \boldsymbol{\pi'}(i,j) = \begin{cases} (1-p)\boldsymbol{\pi}(i) & \text{if } i \le n, \\ p & \text{else,} \end{cases}\end{array} \tag{20}$$

If we fix p and q the problem becomes linear. Solving this problem consecutively for $n = 1,2,3,\ldots$ we obtain for the optimal solution $q = 0$ and $p = 1/(n+1)$ which results in the APHD proposed in [3].

To show that the resulting APHD is not globally optimal, in the sense that we cannot find an APHD representation with n phases that has a larger input flexibility, we now construct an APHD $(\boldsymbol{\pi},\boldsymbol{D})$ with 3 phases and the additional restrictions that $\boldsymbol{\pi}(1) = \boldsymbol{\pi}(3)$, $\mu_1 = 1$ and the APHD is in canonical form (3). For $\mu_3 = 3.09529$, $\mu_2 = 1.912996$ we have $\rho^-_{(\boldsymbol{\pi},\boldsymbol{D})} = -0.36154$ which is slightly smaller than $\rho^{-(3)} = -0.36111$ which results form $\mu_3 = 3$ and $\mu_2 = 2$.

7. AN EXAMPLE

Single server queues where inter arrival and service times are correlated can be modeled as $PH[n]/PH[n]/1$ queues, where n equals the number of phases of the PHD to describe the arrival and service process, respectively. The queue can be solved with matrix geometric methods using the approach proposed in [5, 10, 12]. To model an M/M/1 queue with correlated arrival and service times and correlation coefficient $\rho > 0$ we first have to find the minimal n such that $\rho \le \rho^{+(n)}$. Then the corresponding APHD of the first canonical form is generated and used to represent

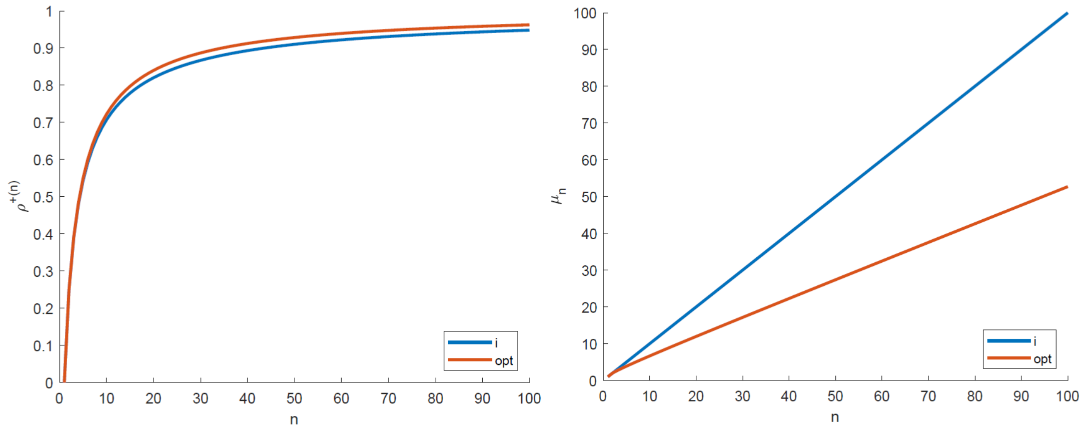

Figure 1: Value of $\rho^{+(n)}$ and μ_n for the representation from (18) and the representations with phase rates i.

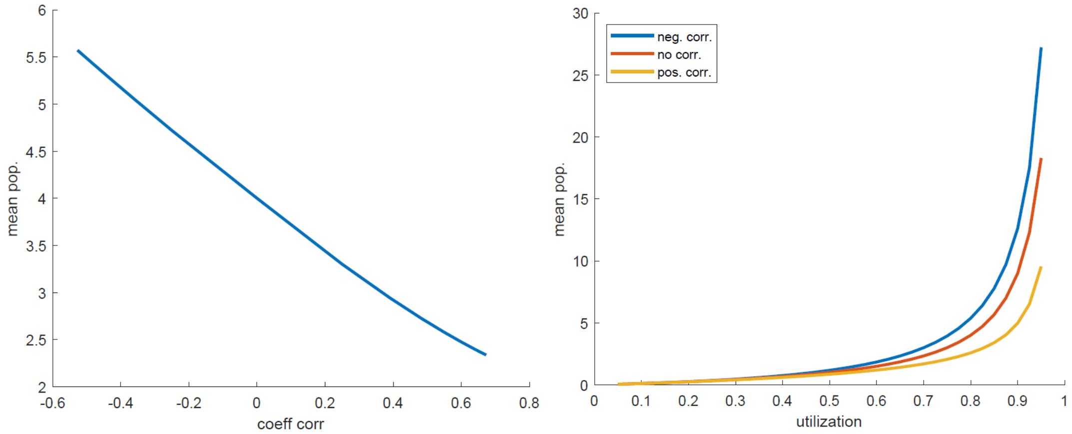

Figure 2: Mean population including the job in service of an M/M/1 queue with correlation between arrival and service times. Left side mean queue length depending on the coefficient of correlation for utilization 0.8**. Right side mean population including the job in service for varying utilization and negative correlation (coefficient of correlation** -0.4636**), no correlation and positive correlation (coefficient of correlation** 0.5502**)**

the service time, the distribution is also transformed to the second canonical form to represent the arrivals. Both distributions are multiplied with the corresponding rates to have the correct arrival and service rates. A customer that arrives from phase k of the arrival time distribution is assigned to class k ($\in \{1, \ldots, n\}$), the initial vector of class k is chosen as $\rho/\rho^{+(n)}\boldsymbol{e}_k + (1 - \rho/\rho^{+(n)})\boldsymbol{\pi}$ where $\boldsymbol{\pi}$ is the initial vector of the computed APHD in the first canonical form and $\boldsymbol{e}_k$ is a vector of length n with 1 in position k and 0 elsewhere. Negative correlations can be represented similarly. The resulting queue can be analyzed with the algorithms available in [2].

We present some results for an M/M/1 queue where inter arrival and service times are correlated. Figure 2 shows the mean population including the job in service for different versions of the queue. It can be seen that a negative correlation between both values results in a significant larger population, in particular for a larger utilization, and that the population depends for a given utilization almost linearly on the coefficient of correlation. Some more detailed results can be found in Figure 3. It can be noticed that for the tail of the queue length differences between the different versions of the queue increase and that the sojourn time of a specific job depends heavily on the inter arrival time.

	U				
ρ	0.2	0.5	0.8	0.9	0.95
-0.5498	828.9	1733.5	4372.1	8263.5	15609
-0.4636	0.95	1.82	5.48	10.46	20.51
-0.2500	0.00	0.01	0.03	0.03	0.07
0.2500	0.00	0.00	0.01	0.02	0.03
0.6008	0.71	1.31	2.24	4.07	6.90
0.7222	622.2	994.2	1608.9	2363.9	3862.7

Table 1: CPU times in seconds to solve the M/M/1 queue with varying coefficient of correlation ρ and utilization U.

The solution algorithm required to solve the M/M/1 queue is based on the solver for the solution of SM[K]/PH[K]/1/FCFS queues from [10], a MATLAB implementation of this algorithm is available in [2]. Unfortunately, the block size of the matrix blocks solved in the matrix geometric solution is in $O(n^3)$ where n is the dimension of the APHD representation of the exponential distributions. Thus, for small and large coefficients of correlation, the effort grows quickly since the number of phases grows quickly as shown in Figure 1. Table 1 includes the CPU times required to solve the queue on a standard PC with Intel(R) Core(TM) i5-9400 CPU @ 2.90GHz with 6 cores and 8 GB of main memory. The

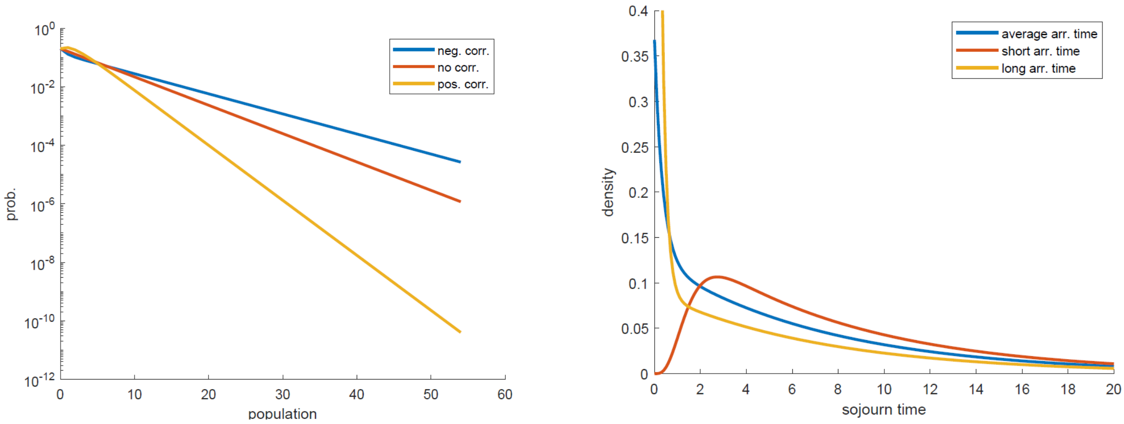

Figure 3: M/M/1 queue with utilization 0.8, mean service time 1. Left side queue length distribution for negative correlation, positive correlation or no correlation. Right side density of the sojourn time for the queue with negative correlation (coeff. of corr. -0.4636). Density for average customers, customers departing from the first phase of arrival time distribution and arrivals from the last phase of the arrival time distribution.

wall clock time is shorter due to MATLAB's internal parallelization. It can be noticed that the solution time mainly depends on the number of phases. In the example we used 2, 5 and 10 phases. Since the block size for the matrix geometric solution grows with $O(n^3)$, we see that it becomes very costly to increase or decrease the coefficient of ocrrelation further. Additionally, solution time grows with the utilization of the queue because the solver requires more iterations to converge and solution takes longer for negative than for positive correlation with the same number of phases because the rates are larger in the negative case.

8. CONCLUSIONS

The paper introduces phase type representations of exponential distributions that allow one to model correlated exponential distributions by starting the distribution in some joint states or route the output of one distribution to the next one. The proposed results extend previously published phase type representations for exponential distributions. The representations of the exponential distribution can be used a building blocks for phase type distributions and Markovian arrival processes [4].

9. REFERENCES

[1] J. R. Artalejo, A. Gomez-Corral, and Q. M. He. Markovian arrivals in stochastic modelling: a survey and some new results. *SORT: Statistics and Operations Research Transactions*, 34(2):101–156, 2010.

[2] Dario Bini, Beatrice Meini, S. Steffé, Juan F. Pérez, and Benny Van Houdt. SMCSolver and Q-MAM: tools for matrix-analytic methods. *SIGMETRICS Perform. Evaluation Rev.*, 39(4):46, 2012.

[3] Mogens Bladt and Bo Friis Nielsen. On the construction of bivariate exponential distributions with an arbitrary correlation coefficient. *Stochastic Models*, 26(3):295–308, 2010.

[4] Peter Buchholz. On the representation of correlated exponential distributions by phase type distributions. *CoRR*, abs/2108.12223, 2021.

[5] Peter Buchholz and Jan Kriege. Fitting correlated arrival and service times and related queueing performance. *Queueing Syst. Theory Appl.*, 85(3-4):337–359, 2017.

[6] Peter Buchholz, Jan Kriege, and Iryna Felko. *Input Modeling with Phase-Type Distributions and Markov Models - Theory and Applications*. Springer, 2014.

[7] Ismail Civelek, Bahar Biller, and Alan Scheller-Wolf. Impact of dependence on single-server queueing systems. *European Journal of Operational Research*, 290(3):1031–1045, 2021.

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

[9] Belén García-Mora, Cristina Santamaría, and Gregorio Rubio. Modeling dependence in the inter-failure times. an analysis in reliability models by markovian arrival processes. *J. Comput. Appl. Math.*, 343:762–770, 2018.

[10] Qi-Ming He. Analysis of a continuous time SM[K]/PH[K]/1/FCFS queue: age process, sojourn times, and queue lengths. *J. Syst. Sci. Complex*, 25(1):133–155, 2012.

[11] Qi-Ming He, Hanqin Zhang, and Juan Vera. On some properties of bivariate exponential distributions. *Stochastic Models*, 28(2):187–206, 2012.

[12] Benny Van Houdt. A matrix geometric representation for the queue length distribution of multitype semi-markovian queues. *Perform. Evaluation*, 69(7-8):299–314, 2012.

[13] J. G. Kemeny and J. L. Snell. *Finite Markov chains.* University series in undergraduate mathematics. VanNostrand, New York, repr edition, 1969.

[14] Yang-Kuei Lin, Michele E. Pfund, and John W. Fowler. Processing time generation schemes for parallel machine scheduling problems with various correlation structures. *J. Sched.*, 17(6):569–586, 2014.

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

[16] W. J. Stewart. *Probability, Markov Chains, Queues, and Simulation*. Princeton University Press, 2009.