to be published on Stochastic Model,

Datomi Dic/98

SPARSE REPRESENTATIONS OF PHASE-TYPE DISTRIBUTIONS

Ştefăniţă Mocanu and Christian Commault

Laboratoire d'Automatique de Grenoble
ENSIEG B.P. 46
38402 St. Martin d'Hères
France
e-mail: Stefan.Mocanu@lag.ensieg.inpg.fr
Christian.Commault@lag.ensieg.inpg.fr

Abstract

A phase-type distribution is the distribution of hitting time in a finite-state Markov chain. A phase-type distribution is triangular if there exist an upper triangular markovian representation. We introduce in this paper an extension of the triangular phase-type distributions, which we call monocyclic distributions. They are convolutions of Erlang and feedback Erlang distributions. We will show that any phase-type distribution can be represented as a mixture of these simple sparse distributions.

Key words. phase-type distributions, Markov chain, monocyclic phase-type distribution.

1 Introduction

A phase-type distribution is the distribution of absorption time in a finite-state Markov chain. The first unified approach to the standard parametrization of phase-type distributions, and some closure properties, were presented in [8]. Some other closure properties are given in ([2, 7]). A caracterization theorem was formulated in [11] and a remarkable property of the coefficient of variation was given by Aldous and Shepp [1]. However, detailed characterizations of phase type distributions are given only in the particular case of phase-type distributions whose transforms have only real poles, also called Coxian or triangular distributions ([4, 5, 6, 12, 13]). Some open problems concerning phase-type distributions are described in [14].

Our goal is to find a simple extension of the Coxian distribution which can parametrize the entire class of phase-type distributions and preserves most of the remarkable properties of the Coxian distributions.

In Section 2 below we review some basic notions concerning phase-type distributions. In Section 3 we define the extension of the triangular distributions that we call monocyclic distributions. Section 4 is concerned with the representation theorem that will show that any phase-type distribution can be expressed as a mixture of such highly sparse distributions. The last section contains some applications of the representation theorem, and we discuss some conjectures from [14]

2 Basic notions and definitions

Definition 1 *A probability distribution $F(\cdot)$ over $[0, \infty)$ is a phase-type distribution (PH-distribution) if it is the distribution of the time to absorption in a finite-state Markov chain, characterized by the vector of initial probability distribution $\underline{\alpha}$ and the transition matrix T. The pair $(\underline{\alpha}, T)$ is called a representation of $F(\cdot)$.*

The order of a representation $(\underline{\alpha}, T)$ is defined as the order of the generator T. We use the notation $PH(\underline{\alpha}, T)$ to denote the phase-type distribution whose representation is $(\underline{\alpha}, T)$. By $PH(T)$ we denote the set of all phase-type distributions having a representation whose generator is T. We denote by $\underline{e}$ a vector of adequate dimension, whose entries are all equal to one.

A PH distribution assigns a mass $1 - \underline{\alpha}\ \underline{e}$ to 0 and the absolutely continuous part of its density and its Laplace-Stieljes transform are given by:

$$f(t) = \underline{\alpha} e^{tT}(-T\underline{e}), \quad \tilde{f}(s) = \underline{\alpha}(sI - T)^{-1}(-T\underline{e}), \tag{1}$$

where I is the identity matrix.

The degree of the denominator of the irreducible form of the Laplace-Stieljes transform is called *the algebraic degree* of the distribution.

We are concerned with signed Borel measures on $[0,\infty)$ with finite total variation [16] and rational Laplace-Stieljes transform ($RLST$). Let Z be the vector space of such measures and let μ be an element of $RLST \subset Z$. Additionally we suppose that $\mu \in PH$, so μ satisfies the necessary and sufficient conditions of the characterization theorem [11]

Theorem 1 *A probability distribution on $[0,\infty)$ wich is not the point mass at zero is of phase-type if and only if*

a) it has a rational Laplace-Stieljes transform with a unique pole of maximal real part, and

b) the continuous density of its absolutely continuous part is positive everywhere on $[0,\infty)$.

Some important sets will be used in the following. The terminology is the same as in [11].

POS denotes the set of elements of Z wich assign no mass to zero.

PM denotes the set of elements of Z wich are probability measures.

UTM denotes the set of elements of Z with unit total mass.

PH denotes the set of phase-type distributions.

We define, following [11], the *residual life operators* R_t, $t \geq 0$ on Z:

$$R_t\mu(\{0\}) = \mu([0,t]), R_t\mu(E) = \mu(E+t). \tag{2}$$

where E is a Borel subset of $[0,\infty)$ and $E+t$ denotes its translate by t units to the right. Thus if μ has the absolutely continuous part of its density equal to $f(x)$, the absolutely continuous part of $R_t\mu$ has density $f(x+t)$ and the mass assigned to zero will be:

$$\mu(\{0\}) + \int_0^t f(x)dx.$$

We define the space generated by the measures $R_t\mu$, $t \geq 0$:

$$\begin{aligned} Span(\mu) &= span\{R_t\mu : t \geq 0\} \\ &= \left\{ \sum_{i=1}^{l} a_i R_{t_i}\mu : a_i \in (-\infty,\infty), t_i \in [0,\infty), i = 1,2,\ldots,l = 1,2,\ldots \right\}. \end{aligned} \tag{3}$$

It is easy to check the relation $R_sR_t = R_{s+t}$, so the family $\{R_t : t \geq 0\}$ has a semi-group structure and so it possesses a generator Γ defined by:

$$\Gamma\mu = \lim_{t\downarrow 0} \frac{R_t\mu - \mu}{t}, \tag{4}$$

with $R_t = e^{t\Gamma}$ on $Span(\mu)$. The semi-group property of R_t involves that:

$$\frac{d}{dt} R_t \mu = \Gamma R_t \mu. \tag{5}$$

If $(\underline{\alpha}, G)$ is a representation of a phase-type distribution, then the application of the operators R_t and Γ is described by:

$$R_t PH(\underline{\alpha}, G) = PH(\underline{\alpha} e^{tG}, G), \ t \geq 0, \tag{6}$$

$$\Gamma PH(\underline{\alpha}, G) = DIST(\underline{\alpha} G, G), \tag{7}$$

where $DIST(v, G)$ is an element in UTM with v not necessarily sub-stochastic (as $\underline{\alpha} G$ is generally not sub-stochastic).

For a detailed description and properties of the operators R_t and Γ see [11, 12, 13].

The goal of this paper is to find a representation $(\underline{\alpha}, \Lambda)$ for a phase-type probability measure μ, with a sparse generator Λ having a quasi-bi-diagonal structure. A certain number of elements in the lower triangular part are non-zero; equivalently, the associated Markov chain graph has a certain number of backward transitions.

The idea is to establish a link between the number of cycles and the number of complex pairs of poles. This fact is made explicit in Section 3.

The construction technique for this kind of representations is as follows: we define a generator Θ with such a quasi-bi-diagonal structure, in order that the poles of the Laplace-Stieljes transform of μ are among the eigenvalues of this generator Θ:

$$poles(\tilde{\mu}(s)) \subseteq \sigma(\Theta),$$

where $\sigma(\cdot)$ is the spectrum of a matrix. Then, we prove that we can always find a representation (in the general case non-Markovian) for μ:

$$\mu = DIST(\underline{v}, \Theta), \quad \underline{v} \in \mathbb{R}^n. \tag{8}$$

Then we apply the O'Cinneide's invariant polytope technique [11, 12] and we augment the polytope $PH(\Theta)$ such that we obtain a certain R-invariant polytope containing μ. Then we deduce that there is a generator Λ with the required structure, such that:

$$\mu \in PH(\Lambda).$$

The details of this method are given in Section 4.

We recall some notions from convex analysis, required in the next sections. The following definitions are detailed in [15].

In the following the n-dimensional Euclidian space is denoted by $\mathbb{E}^n$. We will denote the convex hull of $\mathcal{S} \subset \mathbb{E}^n$ by $co(\mathcal{S})$.

Definition 2 *The affine hull of a subset $\mathcal{S} \subset \mathbb{E}^n$ is the intersection of all affine sets in $\mathbb{E}^n$ containing $\mathcal{S}$.*

Definition 3 *A set that is the convex hull of a finite number of points $\{b_0, b_1, \ldots, b_n\}$ is called a* **polytope**. $\nabla\nabla\nabla$

If the set $\{b_0, b_1, \ldots, b_n\}$ is affinely independent then its convex hull is called an *n-dimensional simplex.*

Definition 4 *The relative interior $ri(\mathcal{S})$ of a convex set $\mathcal{S} \subset \mathbb{E}^n$ is the interior of $\mathcal{S}$ relative to the affine hull of $\mathcal{S}$.* $\nabla\nabla\nabla$

3 Cyclic representations

The elementary fact, that the presence of complex poles in the Laplace-Stieljes transform implies the presence of backward transitions in the associated Markov chain suggests the search for sparse cyclic representations for general phase-type distributions. Some workers were already interested in the study of representations that present a single backward transition (the feedback Erlang distributions [14, 3], or with multiple backward transitions from the last state only [14]. We propose another particular case of cyclic representations, called **monocyclic** representations. They are characterized by the fact that every state can belong at most to one cycle of the representation. The idea is to associate one cycle of the representation to each pair of complex poles.

However, there is an indeterminacy in the choice of the parameters for each cycle. The fact that, in the general case, we cannot find a representation of dimension equal to the algebraic degree ([12]), such that the set of complex poles of the Laplace-Stieljes transform equals the set of eigenvalues of the generator, makes difficult the use of general monocyclic representations. To see that consider the simple example of a phase-type distribution of algebraic degree three with a single pair of complex poles

$$f(s) = \frac{1}{\left(\frac{s}{\lambda}+1\right)\left(\frac{s}{a+ib}+1\right)\left(\frac{s}{a-ib}+1\right)}.$$

It was shown [12, 3] that if and only if the poles satisfy the condition

$$\frac{b}{a-\lambda} \leq \cot\frac{\pi}{3}, \tag{9}$$

one can find an order three representation $(\underline{e}_1, T_3)$, with

$$T_3 = \begin{pmatrix} -\lambda_1 & \lambda_1 & 0 \\ 0 & -\lambda_2 & \lambda_2 \\ p\lambda_3 & 0 & -\lambda_3 \end{pmatrix}, \tag{10}$$

where

$$\begin{aligned} \lambda_1 &\in \left[\tfrac{1}{3}\left(2a+\lambda-2\sqrt{(a-\lambda)^2-3b^2}\right), \tfrac{1}{3}\left(2a+\lambda+2\sqrt{(a-\lambda)^2-3b^2}\right)\right], \\ \lambda_2 &= \tfrac{1}{2}\left(2a+\lambda-\lambda_1-\sqrt{-3\lambda_1^2+2\lambda_1(2a+\lambda)\lambda(4a-\lambda)-4b^2}\right), \\ \lambda_3 &= \tfrac{1}{2}\left(2a+\lambda-\lambda_1+\sqrt{-3\lambda_1^2+2\lambda_1(2a+\lambda)\lambda(4a-\lambda)-4b^2}\right), \\ p &= 1-\frac{(a^2+b^2)\lambda}{\lambda_1(a^2+b^2+(\lambda_1-\lambda)(\lambda_1-2a))} \in (0,1). \end{aligned} \tag{11}$$

Clearly there is an infinity of solutions as λ_1 ranges between its minimal and maximal value. To avoid this kind of indeterminacy in the choice of parameters we will refine the definition of monocyclic representation and finally we will define a canonical basis of the representation.

Giving a closer look at equations (11) we see that the solution of the system is uniquely determined if the quantity $(a-\lambda)^2-3b^2$ is equal to 0. Then $\lambda_1 = \lambda_2 = \lambda_3 = \frac{1}{2}(2a+\lambda)$, and the representation is unique. This fact corresponds to equality in (9). Indeed the equality occurs in (9) only if the elements on the first diagonal of T_3 are equal. This simple observation is a consequence of the extremality of the feedback Erlang distribution with respect to the result of Theorem 3.1 of [12].

One hint to find a uniquely determined representation of the given phase-type distribution is to consider a non-minimal fourth order representation $(\underline{\alpha}, T')$, where:

$$T' = \begin{pmatrix} -\lambda & \lambda & 0 & 0 \\ 0 & -\lambda_1 & \lambda_1 & 0 \\ 0 & 0 & -\lambda_1 & \lambda_1 \\ 0 & p\lambda_1 & 0 & -\lambda_1 \end{pmatrix}, \tag{12}$$

$$\underline{\alpha} = (\alpha_1\ \alpha_2\ 0\ 0).$$

with $\lambda_1 = a + b\sqrt{3}$ and $\alpha_2 = \frac{\lambda}{\lambda_1}$, $\alpha_1 = 1 - \alpha_2$.

This fact suggests the definition of a canonical representation for phase-type distributions, such that the parameters of the representation are uniquely determined. Even if such a representation is not minimal by construction, in the sense of the order of the generator, the representation is sparse and the number of parameters is low. In the previous example for both representations we have the same number of parameters.

A closer study of the eigenvalues localizations is needed for the definition of the canonical representation basis.

3.1 Eigenvalues localization.

Definition 5 *We define the relative order of a complex pair of poles* $\lambda_{i,i+1} = a \pm ib$*, w.r.t the maximal pole* $-\lambda_1$*, as the smallest integer* m *such that:*

$$a - \lambda_1 > b \tan \frac{\pi}{m}.$$

An example of the geometric interpretation of the relative order is depicted in Figure 1

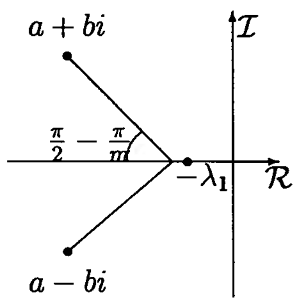

Figure 1: The positioning of a complex pair of poles of relative order m

We consider a particular case of monocyclic representations, the feedback

Erlang distributions (Figure 2):

$$\Theta = \begin{pmatrix} -\lambda & \lambda & 0 & \dots & 0 & 0 \\ 0 & -\lambda & \lambda & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & -\lambda & \lambda \\ z\lambda & 0 & 0 & \dots & 0 & \lambda \end{pmatrix}. \tag{13}$$

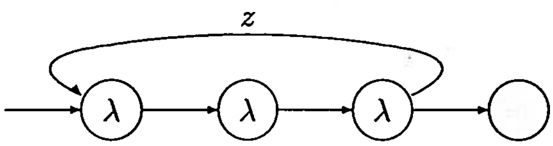

Figure 2: Feedback Erlang distribution

The eigenvalues of such a generator can be obtained by a simple transformation on the n-order roots of unity.

$$\lambda_k = -\left(1 - z^{\left(\frac{1}{n}\right)} \cos\frac{2(k-1)\pi}{n}\right)\lambda + i\left(z^{\left(\frac{1}{n}\right)} \sin\frac{2(k-1)\pi}{n}\right)\lambda.$$

Proposition 1 *Given a complex pair of poles $a \pm ib$ there is a unique feedback Erlang distribution, of a given order n, such that the complex pair of poles is an eigenvalues pair of the generator of the representation. The parameters of that representation are uniquely determined by:*

$$\lambda = \frac{1}{2}\left(2a - b\tan\frac{\pi}{n} + b\cot\frac{\pi}{n}\right) \tag{14}$$

$$1 - z^{\frac{1}{n}} = \frac{2(a - b\tan\frac{\pi}{n})}{2a - b\tan\frac{\pi}{n} + b\cot\frac{\pi}{n}}. \tag{15}$$

$\nabla\nabla\nabla$

Proof. From a geometric point of view, the transition rate λ is the center of the circle circumscribed about the regular polygon with the three rightmost vertices given by the two complex poles and the point λ_1' where $\lambda_1' = a - b\tan\frac{\pi}{n}$. The center of this circle is uniquely determined by (14). The nth root of the probability of the backward transition z is the radius of the same circle (Figure 3). It will be uniquely determined by (15). The feedback Erlang distribution of order n is uniquely determined by these two parameters. $\square$

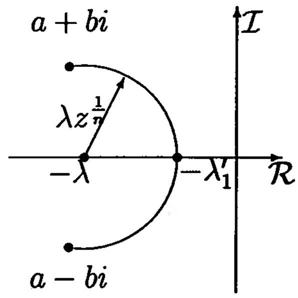

Figure 3: Geometric relations between the poles and the parameters of the feedback Erlang distribution

The eigenvalue λ_1' of maximal real part of the feedback Erlang generator is equal to the product $-\lambda(1 - z^{\frac{1}{n}})$.

Remark 1 *If* $n = 3$, $\lambda < a$; *if* $n = 4$ *then* $\lambda = a$ *and for* $n > 4$, $\lambda > a$ *(see Figure 4).*

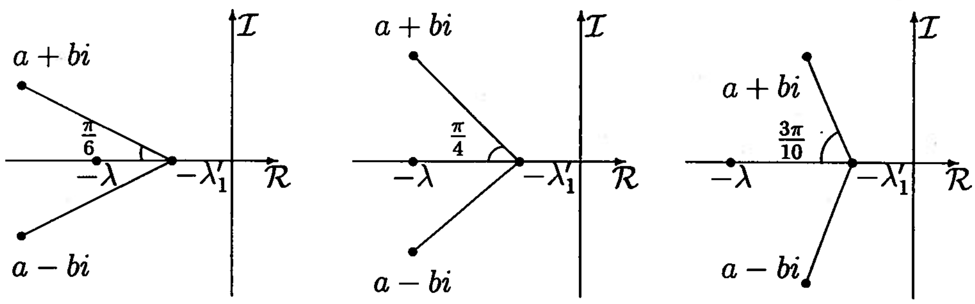

Figure 4: λ's positioning for $n = 3,\ 4,\ 5$

3.2 The canonical basis

In the following we define the canonical representation basis. The goal is to provide a simple representation basis for general phase-type distributions whose elements are formed by convolutions of Erlang and feedback Erlang distributions (13).

We will provide a method to compute the transition rates and the initial probability vector.

We consider the set of poles of the Laplace-Stieljes transform of a phase-type distribution. The idea is to identify each complex pair of poles with an eigenvalue pair of a feedback Erlang representation.

Let $\mathcal{A} = \{-\lambda_i\}_{i=1\ldots n} \subset \mathbb{C}$ be the set of the n (real and complex conjugate) poles of the Laplace-Stieljes transform of a phase-type distribution. Let $-\lambda_1 \in \mathbb{R}$ be the element of maximal real part from $\mathcal{A}$. For every complex conjugate pair of poles $\left(\lambda_j,\ \overline{\lambda}_j\right)$ we define a feedback Erlang representation for which they are a conjugate pair of eigenvalues. The order of the feedback Erlang representation is the relative order of the pair.

Following Proposition 1, each such representation is uniquely determined.

For the remaining real poles we consider the corresponding exponential and Erlang distributions.

Eventually, we obtain the generator by the convolution of all the above defined generators. For a uniqueness purpose we order the components of the convolution in order of increasing Perron-Frobenius eigenvalues. If these eigenvalues are equal then we consider the order of the transition rates and, if the equality occurs also for transition rates, we consider the orders of the transition matrices. Then the set of the distributions is described by a set of ordered triplets:

$$\mathcal{A}' = \{(\tilde{\lambda}_i, z_i, n_i)\}. \tag{16}$$

where $\tilde{\lambda}_i$ is the transition rate, z_i is the weight of the backward transition ($z = 0$ for the exponential and Erlang distributions) and n_i is the order of the distribution.

After convolution we obtain a monocyclic generator Θ. The order of this generator is the sum of relative orders of all the complex pairs of poles and the number of real poles. Note that, by construction, an Erlang component of order k is associated to the pole of maximal real part λ_1, where k is the multiplicity of that pole, and no other component has λ_1 as eigenvalue. The canonical basis is the set of the distributions $(\underline{e}_i, \Theta)$, where $\underline{e}_i$ are the vectors having the ith entry equal to 1, all the others being 0. Notice also that Θ is simple as defined in [10].

Example.

$$\mathcal{A} = \{\lambda_1,\ \lambda_2,\ a+ib,\ a-ib,\ c+id,\ c-id\}$$

We assume that both complex pairs of poles are of relative order 3.

We suppose that after the construction we obtain the ordered set of parameters for the convolving distributions:

$$\mathcal{A}' = \left\{(\lambda_1, 0, 1), (\tilde{\lambda}_a, z_a, 3), (\tilde{\lambda}_c, z_c, 3), (\lambda_2, 0, 1)\right\}$$

The corresponding generator is:

$$\Theta = \begin{bmatrix} -\lambda_1 & \lambda_1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -\lambda_a & \lambda_a & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -\lambda_a & \lambda_a & 0 & 0 & 0 & 0 \\ 0 & z_a\lambda_a & 0 & -\lambda_a & (1-z_a)\lambda_c & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -\lambda_c & \lambda_c & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -\lambda_c & \lambda_c & 0 \\ 0 & 0 & 0 & 0 & z_c\lambda_c & 0 & -\lambda_c & (1-z_c)\lambda_c \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -\lambda_2 \end{bmatrix} \tag{17}$$

For convenience, the distributions $(\underline{e}_i, \Theta)$ will be called **monocyclic generalized Erlang distributions**, in what follows.

In Figure 5 we depict the cyclic graph corresponding to a monocyclic Erlang distribution.

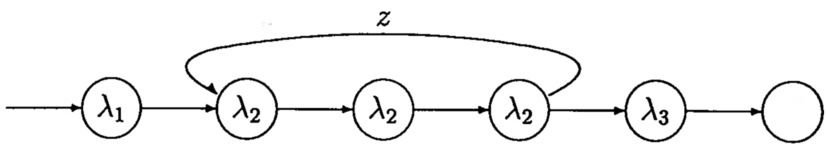

Figure 5: Monocyclic generalized Erlang distribution

Two particular classes of phase-type distributions naturally arise from the definition of monocyclic generalized Erlang distributions.

Definition 6 *We define the class MME (mixture of monocyclic generalized Erlang) of PH distributions, the phase-type distributions that can be represented as $(\underline{\alpha}, \Theta)$ where Θ is a monocyclic generator and $\underline{\alpha}$ is a sub-stochastic vector.* $\nabla\nabla\nabla$

The graph of the generator of a MME distribution has the same generator as the monocyclic generalized Erlang distribution, but the initial probability distribution vector may have more than one non-zero entry.

The reverse time representations of the MME distributions are the analogous of the Coxian distributions (see Figure 6).

Definition 7 *We define the class MCox (monocyclic Coxian) of PH distributions, the phase-type distributions that have a monocyclic bidiagonal representation where the output rates of the stages in the same cycle are not necessary are.*

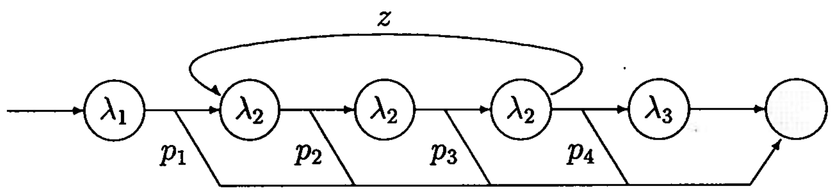

Figure 6: Bidiagonal monocyclic representation

It is useful to remark that if we set $n_i = 1$ for all the cycles then we find the Erlang and Cox distributions. In that sense the monocyclic representations are a generalization of the case of real poles.

It can be shown that every phase-type distribution has a monocyclic (in the general with a non-stochastic initial distribution vector) representation.

Proposition 2 *Let μ be a probability measure of phase-type . We can always find a representation for μ $DIST(\underline{v}, \Theta)$, $\underline{v} \in \mathbb{R}^n$ with Θ a monocyclic generator.* $\nabla\nabla\nabla$

The proof is given in appendix.

4 The representation theorem

Theorem 2 *Every phase-type distribution has a MME representation.*

Proof. The proof follows the same steps as the proof of Theorem 4.1 in [12]. For conformity we will preserve, if possible, the same notations as in [12].

Let μ be a probability measure of phase-type and suppose that μ does not assign a mass to 0, being absolutely continuous. Let $\mathcal{A} = \{-\lambda_i\}_{i=1\dots n} \subset \mathbb{C}$ be the pole set of $\tilde{\mu}$ and let Θ be the monocyclic generator defined over $\mathcal{A}'$.

We will construct a certain R-invariant polytope containing μ, and we apply Lemma 2.1 from [12] to deduce that there is a generator Λ such that the polytope is included in $PH(\Lambda)$. The polytope is constructed by augmenting the polytope $PH(\Theta)$ with the convex hull of the measures obtained by solving the differential equations system (5) by Euler's method. The final polytope will have all the extreme points absolutely continuous except for δ_0. Choosing the increments sufficiently small in Euler's method will ensure that the polytope is R-invariant. We construct a first series of Euler's approximants from μ to $\mu_{c+1} \in ri(Span(\mu) \cap PM)$.

Following Lemma 5.4B [11] a measure $\nu \in Span(\mu) \cap PM$ is in $ri(Span(\mu) \cap PM)$ if and only if it satisfies the conditions:

(1) ν gives positive mass to 0,

(2) ν has a continuous positive density on $[0, \infty)$,

(3) $\tilde{\nu}$ has an order k pole at λ_1.

The probability measure μ is in $Span(\mu) \cap PM$, but is not an interior point, since $\mu(0) = 0$.

Using Lemma 5.5 (B and C) [11] if $\nu \in Span(\mu) \cap PM$ and if it satisfies the conditions

(i) ν has a continuous positive density on $(0, \infty)$,

(ii) $\tilde{\nu}$ has an order k pole at λ_1,

then for all ε sufficiently small $(I + \varepsilon\Gamma)\nu = \nu + \varepsilon\Gamma\nu$ is in $Span(\mu) \cap PM$ and satisfies (i) and (ii). To bring μ to satisfy the conditions $(1) - (3)$, and then to be in $ri(Span(\mu) \cap PM)$, μ must assign a positive mass to 0. The effect of the Γ operator is such that if f is the density of the measure μ, the density of $\Gamma\mu$ will be f'. Then the absolutely continuous part of $(I + \varepsilon\Gamma)\mu$ is $f + \varepsilon f'$ and, for some positive ε, we can reduce the order of the zero of f at 0 by 1. Let c be the order of the zero of f at 0. One can choose $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_{c+1}$ such that the measures:

$$\mu_1 = (I + \varepsilon_1 \Gamma)\mu, \ \ \mu_2 = (I + \varepsilon_2 \Gamma)\mu_1, \ldots, \mu_{c+1} = (I + \varepsilon_1 \Gamma)\mu_c, \tag{18}$$

are in $Span(\mu) \cap PM$ and also satisfy the conditions (i), (ii). The absolutely continuous part of μ_c will be strictly positive at 0, as the order of the zero of f_μ reduces by one unit at each step. We again use the description of the Γ operator, and it follows that μ_{c+1} gives a positive mass to 0. So μ_{c+1} satisfies the conditions $(1)-(3)$ and thus is in $ri(Span(\mu) \cap PM)$.

We will show now that $R_T\mu_{c+1} \in ri(PH(\Theta))$ for T sufficiently large. As $\mathcal{A} \subset spec(\Theta)$, $Span(\mu) \cap UTM \subset DIST(\underline{v}, \Theta)$, $\underline{v} \in \mathbb{R}^n$ (Proposition 2). More precisely:

$$Span(\mu) \cap UTM = \{DIST(\underline{v}', \Theta) \,|\, (spec(\Theta) - \mathcal{A}) \subset zero(LST(DIST(\underline{v}', \Theta)), \underline{v}' \in \mathbb{R}^n\}.$$

where $LST(\cdot)$ means the Laplace-Stieljes transform and $zero(p(s))$ is the set of zeros of $p(s)$.

Let $DIST(\underline{v}', \Theta)$ be a (non-Markovian) representation of μ_{c+1}. In order to prove that $R_T\mu_{c+1} = DIST(\underline{v}'(T), \Theta) \in ri(PH(\Theta))$ for T sufficiently large, it is enough to prove that $\underline{v}'(T)$ becomes sub-stochastic for such a T and that is equivalent, following (6), to proving that $\underline{v}'(T) = \underline{v}' e^{T\Theta} > 0$.

That can be proved from computation of the matrix $(sI - \Theta)^{-1}$ by cofactor expansion.

In order to illustrate the sense of the generalization of the triangular representations, we will provide an alternate probabilistic proof.

We will examine the matrix $e^{T\Theta}$. As follows from the construction algorithm we can partition Θ as

$$\Theta = \left(\begin{array}{c|c} A_1 & A_{12} \\ \hline 0 & A_2 \end{array} \right) \tag{19}$$

where A_1 is an Erlang generator with transition rate λ_1 and whose dimension k is the multiplicity of λ_1. A_2 is a monocyclic generator of dimension $n-k$ whose eigenvalue of maximal real part is strictly less than $-\lambda_1$. The connexion matrix A_{12} has an unique non-zero element $A_{12[k,1]} = \lambda_1$.

Clearly the structure of $e^{t\Theta}$ is similar to that of Θ:

$$e^{t\Theta} = \left(\begin{array}{c|c} e^{tA_1} & B \\ \hline 0 & e^{tA_2} \end{array} \right). \tag{20}$$

It follows that on the columns of $e^{t\Theta}$ the elements (i,j), $i > k$ will decay with faster rates than λ_1. In order to determine the structure of the elements (i,j), $i \leq k$ we invoke some probabilistic considerations. Let X_i, $i = 1, 2, \ldots n$ be the exponential random variables which describe the sojourn time in the states of the Markov chain whose generator is Θ. They may be classified in communicating classes. We write Y_l, $l = 1 \ldots L$ for the communicating classes, where L is the number of cycles. Then, if $X_{i_1} \in Y_{l_1}$ and $X_{i_2} \in Y_{l_2}$, $l_1 \neq l_2$, X_{i_1} and X_{i_1} are independent. We denote by $P(X_i|Y_l > t)$ the conditional probability $P(X_i|X_{i_a} \ldots X_{i_b} > t)$ with $\{X_i, X_{i_a} \ldots X_{i_b}\} = Y_l$. Of course $P(X_i|Y_l > t) = P(X_i > t)$ if $Y_l = \{X_i\}$. Then the process stays in state i of class Y_{l_i} for a time $X_i|Y_{l_i}$ before moving to the state $i+1$ which either is in the same class or is the first state of the next class. State $n+1$ is absorbing.

The functioning of the process is similar to the triangular case but instead of moving from a state to another in a sequence it will move from a class of states to another one.

With these conventions the entry (i,j), $i \leq k$, $j > i$ of $e^{t\Theta}$ is $P(X_i + \ldots + X_j|Y_l > t) - P(X_i + \ldots + X_{j-1}|Y_{l'} > t)$. Then using the fact that the first k random variables X_i are independent and of the same transition rate λ_1 it follows that apart from a positive constant factor the (i,j) entries, $i \leq k$, $j > i$, of $e^{t\Theta}$ behave as $t^{n-i}e^{-\lambda_1 t}$ for $j \geq k$ and as $t^{n-k-i+j}e^{-\lambda_1 t}$ for $j < k$.

It follows that all the rows of $e^{t\Theta}$ decay to 0 faster than the first row, as $t \to \infty$. This means that in $\mu_{c+1} = DIST(\underline{v}, \Theta)$ the first element of $\underline{v}'$ must be strictly positive. It cannot be strictly negative as μ_{c+1} is a probability measure and cannot be 0 as $\tilde{\mu}_{c+1}(s)$ has a pole of order k at $-\lambda_1$. The positivity of the first element of $\underline{v}'$ and the asymptotic behavior of $e^{t\Theta}$ imply that for T sufficiently large, $\underline{v}'e^{T\Theta}$ becomes sub-stochastic and then $R_T\mu_{c+1} = DIST(\underline{v}'(T), \Theta) \in ri(PH(\Theta))$. Let T be chosen this way.

With μ_{c+1} in $ri(Span(\mu) \cap PM)$, it is easy to prove that $R_t\mu_{c+1}$ is also in $ri(Span(\mu) \cap PM)$ for $t \geq 0$. We construct the Euler's approximants to the solution of the differential equation:

$$\frac{d}{dt}R_t\mu_{c+1} = \Gamma R_t\mu_{c+1} \tag{21}$$

on the interval $[0,T]$ and with an integration step $\varepsilon = T/N > 0$. We obtain the approximants

$$\mu_j^* = (I + \varepsilon\Gamma)\mu_{j-1}^* = (I + \varepsilon\Gamma)^j \mu_{c+1}^*, \;\; j = 1, 2, \ldots, N \tag{22}$$

to the exact values $R_{j\varepsilon}$. The trajectory $t \to R_t\mu_{c+1}$ and its approximants lie all in $Span(\mu) \cap PM$. For ε sufficiently small the approximants are uniformly close to the trajectory and so we can ensure that they lie in $ri(Span(\mu) \cap PM)$. We have also proved that $R_T\mu_{c+1}$ is in $ri(PH(\Theta))$ which is open in $Span(\mu) \cap PM$. With ε sufficiently small we can also ensure that $\mu_N^* = (I + \varepsilon\Gamma)^N \mu_{c+1}^*$ is in $ri(PH(\theta))$.

Then we construct the polytope:

$$P = co\{\delta_0, \mu, \mu_1, \ldots, \mu_c, \mu_{c+1}, \overline{\mu}_1^*, \ldots, \overline{\mu}_N^*, PH(\underline{e}_1, \Theta), \ldots, PH(\underline{e}_n, \Theta)\}, \tag{23}$$

where the measures $\overline{\mu}$ are formed from μ by removing the mass at 0 and rescaling so that $\overline{\mu}$ is in UTM:

$$\overline{\mu} = \frac{\mu - \mu(\{0\})\delta_0}{1 - \mu(\{0\})}.$$

Except for δ_0 all the extreme points of P are absolutely continuous. In order to apply the invariant polytope Lemma (Lemma 2.1 [12]) we will prove that P is R-invariant. For that we show that Γ points inward to P at each of its extreme points and then apply Lemma 2.2 [12]. Obviously Γ points inward to P at δ_0. That is also true for $\mu_1, \ldots, \mu_{N-1}^*$ as the left-hand sides of (18) and (22) are in P. From Lemma 2.2 [12] follows the same property for $PH(\underline{e}_i, \Theta)$, $i = 1 \ldots n$, $PH(\Theta)$ being R-invariant and $PH(\Theta) \subseteq P$. μ_N^* is an interior point of $PH(\Theta)$ so it does not need to be checked. It follows from Lemma 2.2 [12] that P is R-invariant. Lemma 2.2 [12] also implies that there exists a generator Λ such that $P = PH(\Lambda)$. Consequently from the same Lemma 2.2 [12] the relation between the elements of Λ and the extreme points ν_i of P are given by:

$$\Gamma\nu_i = \sum_{j \neq i} g_{ij}(\nu_j - \nu_i) + g_{i0}(\delta_0 - \nu_i). \tag{24}$$

We will write down explicitly the generator Λ in that manner.

With the notation f_i for the continuous part of the density of μ_i,$i = 0, 1 \ldots c$ and $\overline{f}_i^*$ for the density of $\overline{\mu}_i^*$, $i = 0, 1 \ldots N-1$, let $(\underline{\alpha}, \Theta)$, $\underline{\alpha} = (\alpha_1, \; \alpha_2, \ldots, \alpha_n)$ be the representation of μ_N^* in $PH(\Theta)$.

With the remark that for $\nu \neq \delta_0$ in $RLST \cap PM$, and $\nu_1 = \nu + \eta\Gamma\nu$, we have:

$$\Gamma\overline{\nu} = \left(\frac{1}{\eta} - \overline{g}(0)\right)(\overline{\nu}_1 - \overline{\nu}) + \overline{g}(0)(\delta_0 - \overline{\nu}), \tag{25}$$

where $\overline{g}$ is the continuous density of $\overline{\nu}$. We use this for the construction of the generator as in (24). We obtain:

$$\begin{aligned}
\Gamma\mu_i &= \frac{1}{\varepsilon_{i+1}}(\mu_{i+1} - \mu_i), \quad i = 0, 1, \ldots, c-1;\\
\Gamma\mu_c &= \left(\tfrac{1}{\varepsilon_{c+1}} - f_c(0)\right)(\overline{\mu}_0^* - \mu_c) + f_c(\delta_0 - \mu_c);\\
\Gamma\overline{\mu}_i^* &= \left(\tfrac{1}{\varepsilon} - \overline{f}_i^*(0)\right)(\overline{\mu}_{i+1}^* - \overline{\mu}_i^*) + \overline{f}_i^*(0)(\delta_0 - \overline{\mu}_i^*), \quad i = 0, 1, \ldots, N-1;\\
\Gamma\overline{\mu}_{N-1}^* &= \left(\tfrac{1}{\varepsilon} - \overline{f}_{N-1}^*(0)\right)(PH(\underline{\alpha}, \Theta) - \overline{\mu}_{N-1}^*) + \overline{f}_{N-1}^*(0)(\delta_0 - \overline{\mu}_{N-1}^*);\\
\Gamma PH(\underline{e}_i, \Theta) &= \lambda_i(PH(\underline{e}_{i+1}, \Theta) - PH(\underline{e}_i, \Theta)), \text{ if } i \text{ is not the last state of a cycle;}\\
\Gamma PH(\underline{e}_i, \Theta) &= \lambda_i((1 - z_j)PH(\underline{e}_{i+1}, \Theta) + z_j PH(\underline{e}_{i-j+1}, \Theta) - PH(\underline{e}_i, \Theta)),\\
&\quad \text{if } i \text{ is the last state of a } j\text{-order cycle;}\\
&\quad i = 1, 2, \ldots, n-1;\\
\Gamma PH(\underline{e}_n, \Theta) &= \lambda_n(\delta_0 - PH(\underline{e}_n, \Theta)).
\end{aligned}$$

We write down the corresponding generator following (24):

$$\begin{pmatrix}
-\varepsilon_1^{-1} & \varepsilon_1^{-1} & & & & & & & & \\
 & -\varepsilon_2^{-1} & \varepsilon_2^{-1} & & & & & & & \\
 & & \cdots & & & & & & & \\
 & & & -\varepsilon_{c+1}^{-1} & \varepsilon_c^{-1} - f_c(0) & & & & & \\
 & & & & -\varepsilon^{-1} & \varepsilon^{-1} - \overline{f}_0^*(0) & & & & \\
 & & & & & -\varepsilon^{-1} & \varepsilon^{-1} - \overline{f}_1^*(0) & & & \\
 & & & & & & \cdots & & & \\
 & & & & & & & -\varepsilon^{-1} & (\varepsilon^{-1} - \overline{f}_{N-1}^*(0)\underline{\alpha}) & \\
 & & & & & & & & \Theta &
\end{pmatrix}$$

Now we use the properties of the TPH generators and we replace the upper part of Λ by a bidiagonal block. Then μ is MME. $\square$

5 Final comments and perspectives

The representation theorem is a generalization of the result concerning the case of real poles and then offers the possibility to use sparse representations for modeling general phase-type distributions. Qualitative analysis of their properties is significantly simplified.

An important property is the characterization of extremal phase-type distributions. Because every phase-type distribution can be expressed as a mixture of monocyclic Erlang distributions, then the last ones are the extremal distributions. This remark is the equivalent of the Conjecture 2 in [14].

The representation theorem seems also to be of some help in the proof of other properties. One of these properties of interest is the Conjecture 1 in [14].

Conjecture 1 *Let f be the density of a phase-type distribution of order n with mean $\mu > 0$ and coefficient of variation c. Then with $\lambda \equiv n/\mu$*

$$f(t) \geq e^{-(nc^2-1)} \frac{\lambda^n t^{n-1}}{(n-1)!} e^{(-\lambda t)} = e^{-(nc^2-1)} \mathcal{E}_{n,\lambda}(t), \quad t \geq 0. \tag{26}$$

Here $\mathcal{E}_{n,\lambda}(t)$ is the Erlang distribution of order n and transition rate λ.

The conjecture in true in the case of real poles (see [14] for proof). Looking at the comments following the conjecture in [14] the conjecture seems to be true also for the feedback Erlang distribution. If it is the case, then, as every phase-type distribution can be represented as MME, the representation theorem can be of use for the proof in the case of complex poles.

Another problem of interest for which the representation theorem could be of some help is *the step increasing conjecture* [14]:

Conjecture 2 *For any phase-type density of order n, $f(t)/t^{n-1}$ is non-increasing for $t > 0$.*

Because the conjecture is true again in the case of real poles, and a phase-type distribution with real pole has a bidiagonal representation we can use the following stochastic argument in favor of the conjecture: adding backward transitions on the graph of a Markov chain will slow down the absorption process. Then the graph of the distribution function will be smother than the "original" function.

A question of interest concerning the monocyclic representations is their dimensions. Even if the state number of such a representation can be very large compared with the algebraic degree, the number of parameters is nearly the same. However, in some applications, the dimension of the Markov chain could be important. There are some possible ways to reduce the number of states of a monocyclic representation.

One way is to relax the strict inequality in the definition of relative order, and also to consider the possibility to have more than one pair of complex poles of the Laplace-Stieljes transform among the eigenvalues of the same feedback Erlang distribution (note that the actual construction algorithm provides a bijection between the number of complex pairs of poles and the number of cycles). That will give the possibility to decrease the number of cycles. On the other hand, the relaxation of the inequality in Definition 5 gives the possibility to reduce the dimensions of the generated feedback Erlang distributions. An immediate implication of that fact is that we will "distribute" the maximal eigenvalue λ_1 in some monocyclic blocks. Then the first Erlang block of transition rate λ_1 can disappear. In such a case the proof of representation theorem becomes more complicated in that the asymptotic characterization of the trajectories $R_t\mu_{c+1}$ is more difficult. Authors have proved this characterization using a property of semi-stable matrices ([9],[17]). The proof is quite intricate and is not provided here.

Another direction for further study that seems to be of interest is to define the "monocyclic order" previously suggested by the definition of the monocyclic Coxian distributions ($MCox$). It seems to be a natural extension of the triangular order. We can look at the Coxian distributions as a particular case of monocyclic Coxian distributions, when all the diagonal blocks (cycles) are of dimension 1.

The study of such "block-triangular" distributions is justified by the fact that the MME distributions are in the general case non-minimal by construction even if we accept the above relaxations on the definition of the relative order and on the construction method.

Then the study of the monocyclic Coxian distributions can provide an issue against the (possible) huge dimension of MME representations. It could provide a smaller dimension representations, preserving the tractability offered by the structural properties. The authors investigations in the class of monocyclic triangular and monocyclic bidiagonal distributions show that they look to be an issue between preserving the tractability of the triangular representations and the generality of use of the phase-type distributions with complex poles.

Appendix.

Proof of Proposition 2.

Let

$$\tilde{\mu}(s) = \frac{\mathcal{P}(s)}{\mathcal{Q}(s)}, \tag{A.1}$$

be the Laplace-Stieljes transform of μ and let $A \in \mathbb{C}$ be the set of poles of $\tilde{\mu}(s)$. We construct the canonical representation basis as in Section 2.3 and we write the nth order monocyclic generator. Then we try to solve

$$\frac{\mathcal{P}(s)}{\mathcal{Q}(s)} = \underline{v}(sI - \Theta)^{-1}(-\Theta\underline{e}). \tag{A.2}$$

If the system (A.2) has a solution, it will be an unit element sum vector as the left and right members in (A.2) are in UTM

We compute the vector:

$$\underline{X}(s) = (sI - \Theta)^{-1}(-\Theta\underline{e}). \tag{A.3}$$

Obviously, each element $x_i(s)$ of $\underline{X}(s)$ is in fact the Laplace-Stieljes transform of the distribution $(\underline{e}_i, \Theta)$.

The difference between the algebraic degree of $x_i(s)$ and the degree of its numerator equals $n - i + 1$. This is a consequence of the following theorem [3]:

Theorem 3 *In any representation of a phase-type distribution the minimal number of transient states which are visited before absorption is equal to the difference between the degrees of the numerator and the denominator of the Laplace-Stieljes transform of the distribution.*

The first element $x_1(s)$ of $\underline{X}(s)$ is equal to:

$$x_1(s) = \frac{y_0}{det(sI - \Theta)}, \tag{A.4}$$

with y_0 some constant equal to the product of eigenvalues of Θ.

For each other distribution $(\underline{e}_i, \Theta)$ the corresponding element of of $\underline{X}(s)$ is:

$$x_i(s) = \frac{p_i(s)}{det(sI - \Theta)} \tag{A.5}$$

with $p_i(s)$ some polynomial of degree $i - 1$. Finally we obtain that:

$$\underline{X}(s) = \frac{1}{det(sI - \Theta)}\underline{Y}(s) \tag{A.6}$$

where $\underline{Y}(s)$ is a vector of polynomials with degrees strictly increasing from 0 to $n-1$.

As the poles of $\tilde{\mu}(s)$ are among the eigenvalues of Θ, $\mathcal{Q}(s)$ divides $det(sI - \Theta)$. Let $\mathcal{R}(s)$ be the quotient of the division of $det(sI - \Theta)$ by $\mathcal{Q}(s)$.

Then the equation (A.2) becomes:

$$\mathcal{P}(s)\mathcal{R}(s) = \underline{v}\,\underline{Y}(s), \tag{A.7}$$

with the observation that because $deg(\mathcal{P}(s))$ is strictly lower than the degree of $\mathcal{Q}(s)$ and $\mathcal{R}(s)$ is of degree $n - deg(\mathcal{Q}(s))$, then the product $\mathcal{P}(s)\mathcal{R}(s)$ is at most of degree $n-1$.

Using the notation a_k^j for the coefficient of s^k in $y_j(s)$ and b_k for the coefficient of s^k in the product $\mathcal{P}(s)\mathcal{R}(s)$ we can rewrite the system (A.7) as

$$\underline{v}\begin{vmatrix} a_0^1 & & & \\ a_0^2 & a_1^2 & & \\ \vdots & & \ddots & \\ a_0^n & & & a_{n-1}^n \end{vmatrix} = \begin{vmatrix} b_0 \\ b_1 \\ \vdots \\ b_{n-1} \end{vmatrix}. \tag{A.8}$$

The system is equivalent to

$$\underline{v}A = \underline{b}$$

with A lower triangular with non-zero diagonal entries. It follows that there exists a unique solution $v \in \mathbb{R}^n$ to that system, equivalently, we can always express μ as $DIST(\underline{v}, \Theta)$.

References

[1] David Aldous and Larry Shepp. The least variable phase-type distribution is Erlang. *Commun. stat., Sthochastic Models*, 3(1):467–473, 1987.

[2] David Assaf and Benny Levikson. Closure of phase-type distributions under operations arising in reliability theory. *Annals of Probability*, 10(1):265–269, 1982.

[3] Christian Commault and Jean-Paul Chemla. An invariant of representations of phase-type distributions and some applications. *Journal of Applied Probability*, 33(2):368–381, 1996.

[4] D.R. Cox. A use of complex probabilities in the theory of stochastic processes. *Proc. Camb. Phil. Soc.*, 51:313–319, 1955.

[5] Aldo Cumani. On the canonical representation of homogenous markov processes modelling failure-time distributions. *Microelectronics and reliability*, 22(3):583–602, 1982.

[6] Michel Dehon and Guy Latouche. A geometric interpretation of the relations between the exponential and generalized Erlang distributions. *Adv. Appl. Prob.*, 14:885–897, 1985.

[7] Robert S. Maier. The algebraic construction of phase-type distributions. *Commun. Stat., Stochastic Models*, 5(2):573–602, 1991.

[8] Marcel Neuts. *Matrix-Geometric Solutions in Stochastic Models. An Algorithmic Approach.* The John Hopkins University Press, Baltimore and London, 1981.

[9] Marcel Neuts. *Algorithmic Probability.* Stochastic Modeling Series. Chapman & Hall, 1995.

[10] Colm Art O'Cinneide. On non-uniqueness of representations of phase-type distributions. *Commun. Stat., Stochastic Models*, 5(2):247–259, 1989.

[11] Colm Art O'Cinneide. Characterization of phase-type distributions. *Commun. Stat., Stochastic Models*, 6(1):1–57, 1990.

[12] Colm Art O'Cinneide. Phase-type distributions and invariant polytopes. *Adv. Appl. Probab.*, 23(3):515–535, 1991.

[13] Colm Art O'Cinneide. Triangular order of triangular phase-type distributions. *Commun. Stat., Stochastic Models*, 9(4):507–529, 1993.

[14] Colm Art O'Cinneide. Phase-type distributions: open problems and a few properties. Technical report, 1998.

[15] Tyrrell R. Rockafellar. *Convex Analysis.* McGraw-Hill, New York, 1970.

[16] Walter Rudin. *Real and Complex Analysis.* McGraw-Hill, New York, 1987.

[17] Eugene Seneta. *Nonnegative Matrices.* George Allen an Unwin, London, 1973.