Proceedings of "The 6th SEAMS-UGM Conference 2011"
pp. 65–78.

THE ORDER OF PHASE-TYPE DISTRIBUTIONS

Reza Pulungan

Abstract. This paper lays out the past and the future of one of the most interesting research problems in the area of phase-type distributions: the problem of their minimal representations. We will chronologically present contemporary results, including our own contributions to the problem, and provide several pointers and possible approaches in attempting to solve the problem in future work.
Keywords and Phrases: Phase-type distributions, Markov chain, minimal representations, order.

1. INTRODUCTION

The problem of minimal representations remains one of the open problems in the research area of phase-type (PH) distributions [16, 18]. Given a phase-type distribution, a minimal representation is an absorbing Markov chain with the fewest number of states, whose distribution of time to absorption is governed by the same phase-type distribution. Obtaining minimal representations is important in various circumstances, including, but not limited to, modeling formalisms that support compositionality [13, 2, 12]. In such circumstance, models are constructed by composing smaller components via various operations that usually result in exponential blowups of the state space. Ensuring that all components and all intermediate results of the composition come in minimal representations will significantly reduce these blowups.

Previous researches [1, 19, 20, 3, 4, 15, 5, 6] have produced several techniques to obtain these minimal representations. However, the frontier is still limited to acyclic phase-type distributions [9, 10, 11, 22], namely those phase-type distributions having at least one Markovian representation that contains no cycle. Even in this case, the resulting algorithm is not yet satisfactory, for it contains non-linear programming, which can be inefficient in many cases.

2010 Mathematics Subject Classification: 60J27, 60J28.

This paper lays out the past and the future of one of the most interesting research problems in the area of phase-type distributions: the problem of their minimal representations. We will chronologically present contemporary results, including our own contributions to the problem, and provide several pointers and possible approaches in attempting to solve the problem in future work.

The paper is organized as follows: Section 2 introduces phase-type distributions and other concepts required throughout the paper. In this section, we also formulate the problem of the order of phase-type distributions. Section 3 lays out previous partial solutions to the problem. In Section 4, we describe our contribution in solving the problem by proposing an algorithm to reduce the size of acyclic phase-type representations. The paper is concluded in Section 5.

2. PRELIMINARIES

2.1. **Phase-Type Distributions.** Let the stochastic process $\{X(t) \in \mathcal{S} \mid t \in \mathbb{R}^+\}$ be a homogeneous Markov process defined on a discrete and finite state space

$$\mathcal{S} = \{s_1, s_2, \cdots, s_n, s_{n+1}\}$$

and with time parameter $t \in \mathbb{R}^+ := [0, \infty)$. The Markov process is a finite continuous-time Markov chain (CTMC). We view the structure of such a CTMC as a tuple $\mathcal{M} = (\mathcal{S}, \mathbf{R})$ where $\mathbf{R}$ a rate matrix $\mathbf{R} : \mathcal{S} \times \mathcal{S} \to \mathbb{R}^+$. The rate matrix $\mathbf{R}$ is related to the corresponding infinitesimal generator matrix by: $\mathbf{Q}(s, s') = \mathbf{R}(s, s')$ if $s \neq s'$ else $\mathbf{Q}(s, s) = -\sum_{s' \neq s} \mathbf{R}(s, s')$ for all $s, s' \in S$. If state s_{n+1} is absorbing (*i.e.*, $\mathbf{Q}(s_{n+1}, s_{n+1}) = 0$) and all other states s_i are transient (*i.e.*, there is a nonzero probability that the state will never be visited once it is left, or equivalently, there exists at least one path from the state to the absorbing state), the infinitesimal generator matrix of the Markov chain can be written as:

$$\mathbf{Q} = \begin{bmatrix} \mathbf{A} & \vec{A} \\ \vec{0} & 0 \end{bmatrix}.$$

Matrix $\mathbf{A}$ is called a *PH-generator* and it is non-singular because the first n states in the Markov chain are transient. Vector $\vec{A}$ is a column vector where its component $\vec{A}_i$ for $i = 1, \cdots, n$ represents the transition rate from state s_i to the absorbing state. The Markov chain is fully specified by the generator matrix $\mathbf{Q}$ and the initial probability vector $(\vec{\alpha}, \alpha_{n+1})$, where $\vec{\alpha}$ is an n-dimensional row vector corresponding to the initial probabilities of the transient states and α_{n+1} is the initial probability to be immediately in the absorbing state. Therefore $\vec{\alpha}\vec{1} + \alpha_{n+1} = 1$, where $\vec{1}$ is an n-dimensional column vector whose components are all equal to 1.

Definition 2.1 (Phase-Type Distribution [16])**.** *A probability distribution on* $\mathbb{R}^+$ *is a* phase-type (PH) distribution *if and only if it is the distribution of the time until absorption in a Markov process of the type described above.*

The pair $(\vec{\alpha}, \mathbf{A})$ is called the *representation* of the PH distribution and $PH(\vec{\alpha}, \mathbf{A})$ is used to denote the *PH distribution* with representation $(\vec{\alpha}, \mathbf{A})$.

The probability distribution of the time until absorption in the Markov chain (hence of PH distribution) is given by:

$$F(t) = 1 - \vec{\alpha}\exp(\mathbf{A}t)\vec{1}, \quad \text{for } t \geq 0. \tag{1}$$

The Laplace-Stieltjes transform (LST) of the PH distribution is given by:

$$\tilde{f}(s) = \int_{-\infty}^{\infty} \exp(-st)dF(t) = \vec{\alpha}(s\mathbf{I} - \mathbf{A})^{-1}\vec{A} + \alpha_{n+1}, \tag{2}$$

where $s \in \mathbb{R}^+$ and $\mathbf{I}$ is the n-dimensional identity matrix. Consider the LST of the PH distribution in (2). This transform is a rational function, namely:

$$\tilde{f}(s) = \vec{\alpha}(s\mathbf{I} - \mathbf{A})^{-1}\vec{A} + \alpha_{n+1} = \frac{P(s)}{Q(s)},$$

for some polynomials $P(s)$ and $Q(s) \neq 0$.

2.2. **Acyclic Phase-Type Distributions.** An interesting subset of the family of PH distributions is the family of *acyclic PH distributions*. The family can be identified by the fact that they have triangular representations. A *triangular representation* is a representation $(\vec{\alpha}, \mathbf{A})$ where matrix $\mathbf{A}$, under some permutation of its components, is an upper triangular matrix.

In [19], O'Cinneide proved the following theorem, which characterizes acyclic PH distributions in terms of the properties of their density functions and their LSTs.

Theorem 2.1 ([19]). *A probability distribution defined on $\mathbb{R}^+$, which is not the point mass at zero, is an* acyclic PH distribution *if and only if (1) its density function is strictly positive on $(0, \infty)$, and (2) its LST is rational and has only real poles.*

Thus, any general PH representation—possibly containing cycles—represents an acyclic PH distribution (and hence has an acyclic representation) whenever the poles of its LST are all real numbers.

2.3. **Ordered Bidiagonal Representations.** Let PH-generator:

$$\mathbf{Bi}(\lambda_1, \lambda_2, \cdots, \lambda_n) = \begin{bmatrix} -\lambda_1 & \lambda_1 & 0 & \cdots & 0 \\ 0 & -\lambda_2 & \lambda_2 & \cdots & 0 \\ 0 & 0 & -\lambda_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & -\lambda_n \end{bmatrix}.$$

If $\lambda_n \geq \lambda_{n-1} \geq \cdots \geq \lambda_1 > 0$ and, then a PH representation $(\vec{\beta}, \mathbf{Bi}(\lambda_1, \lambda_2, \cdots, \lambda_n))$ is called an *ordered bidiagonal representation*.

2.4. **Size of Representation, Algebraic Degree, and Order.** All PH representations we are dealing with in this paper are assumed to be irreducible. A representation is *irreducible* if for the specified initial distribution any transient state is visited with non-zero probability.

For a PH distribution with an irreducible representation $(\vec{\alpha}, \mathbf{A})$, the *size of the representation* is defined as the dimension of matrix $\mathbf{A}$. The degree of the denominator

polynomial of its LST expressed in irreducible ratio is called the *algebraic degree*—or simply the *degree*—of the distribution.

It is known [16, 18] that a given PH distribution has more than one irreducible representation. The size of a minimal irreducible representation, namely a representation with the fewest possible number of states, is referred to as the *order of the PH distribution*. O'Cinneide in [18] showed that the order of a PH distribution may be different from, but at least as great as, its algebraic degree. Therefore the following lemma is straightforward.

Theorem 2.2. *Let n be the size of a PH representation whose size is equal to the algebraic degree of its PH distribution. The the order of the PH distribution is n.*

2.5. **The Problem of the Order of Phase-Type Distributions.** The main problem addressed in this paper is the problem of the order of PH distributions, namely: given a PH distribution, what is its order? Stated differently, we would like to find the minimal number of states required to represent a given PH distribution as an absorbing CTMC. The PH distribution can be given in various ways: as a probability distribution in mathematical formulas, as a Laplace-Stieltjes transform, or even as a PH representation of a certain size.

As a byproduct, of course, it would be advantageous to also be able to devise methods to compute a minimal representation—namely a PH representation whose size is equal to the order—of the given PH distribution.

3. PREVIOUS RESULTS

In this section, previous results on the partial solutions to the problem of the order of PH distributions are presented. The first early result is given in Theorem 2.2. This theorem is a restatement of lemmas found in [16, 18]. The theorem basically establishes that the lower bound of the order of PH distributions is their respective algebraic degree.

In the following subsections, we present further partial results, starting in acyclic PH distributions, the general PH distributions, the relationship between simplicity and order, and, in the end, an attempt to find non-minimal but nonetheless sparse representations.

3.1. **Acyclic Phase-Type Distributions.** Cumani in [7] presented three canonical forms of acyclic PH representations. Of particular interest to us, he proved Theorem 3.1. Aside from the ordered bidiagonal representation, he also provided two other canonical forms and straightforward procedures to transform one to others. A similar theorem was proved by O'Cinneide in [19].

Theorem 3.1 ([7])**.** *Any PH distribution with an acyclic PH representation of a certain size has an ordered bidiagonal representation of equal to or less size.*

HE and Zhang in [9] provided an algorithm, called the *spectral polynomial algorithm*, to obtain the ordered bidiagonal representation of any given acyclic PH representation. The spectral polynomial algorithm is of complexity $\mathcal{O}(n^3)$ where n is the size of the given acyclic PH representation.

In [19], O'Cinneide formally characterized acyclic PH distributions by proving Theorem 2.1. The characterization basically relates acyclic PH distributions to the shape of their density functions and LSTs. The theorem maintains that the LST of any acyclic PH distribution is a rational function and all of its poles are real. Hence, a PH representation could be cyclic; but as long as its LST has only real poles, there must exist an acyclic PH representation that has the same PH distribution.

The following three theorems by Commault and Chemla in [4] specify certain conditions for acyclic PH representations to be minimal, namely to have their size be equal to their respective order.

Theorem 3.2 ([4])**.** *The order of a PH distribution with LST $\tilde{f}(s) = P(s)/Q(s)$, where $P(S)$ and $Q(s)$ are co-prime polynomials, such that $Q(s)$ has degree n with n real roots and $P(s)$ has degree less than or equal to one, is n.*

Theorem 3.2 establishes that the convolution of several exponential distributions always produces minimal PH representations. This means that Erlang representations—formed by a convolution of several exponential distributions of the same rate—and hypoexponential representations—formed by a convolution of several exponential distributions of possibly different rates—are always minimal.

Theorem 3.3 ([4])**.** *Consider a PH distribution with LST $\tilde{f}(s) = P(s)/Q(s)$, where $P(S)$ and $Q(s)$ are co-prime polynomials with real roots, such that:*

$$P(s) = \left(\frac{s+\mu_1}{\mu_1}\right)\left(\frac{s+\mu_2}{\mu_2}\right), \qquad \mu_1 \geq \mu_2 > 0, \text{ and}$$

$$Q(s) = \prod_{i=1}^{n}\left(\frac{s+\lambda_i}{\lambda_i}\right), \qquad \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n > 0.$$

If $\mu_2 \geq \lambda_n$ and $(\mu_1 + \mu_2) \geq (\lambda_{n-1} + \lambda_n)$, then the order of the distribution is n.

Theorem 3.4 ([4])**.** *Consider a PH distribution with LST $\tilde{f}(s) = P(s)/Q(s)$, where $P(S)$ and $Q(s)$ are co-prime polynomials with real roots, such that:*

$$P(s) = \prod_{i=1}^{m}\left(\frac{s+\mu_i}{\mu_i}\right), \qquad \mu_1 \geq \mu_2 \geq \cdots \mu_m > 0, \text{ and}$$

$$Q(s) = \prod_{i=1}^{n}\left(\frac{s+\lambda_i}{\lambda_i}\right), \qquad \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n > 0, n > m.$$

If $\mu_m \geq \lambda_n, \mu_{m-1} \geq \lambda_{n-1}, \cdots, \mu_1 \geq \lambda_{n-m+1}$, then the order of such PH distribution is n.

Theorems 3.3 and 3.4 provides several conditions for the convolution of several exponential distributions of possibly different rates, which starts not only from the first state, to be minimal.

So far, the partial results only provide conditions for the order of acyclic PH distributions to be equal to the size of the representations. This, in itself, is important, since it provides a means to determine whether an existing PH representation is already minimal, or we should first try to find a smaller or even a minimal representation before proceeding to use it. However, an algorithmic results would be useful. Such results will allow us to obtain not only the order but also the minimal PH representations themselves. We shall return to this issue in Section 4, where such algorithmic methods are described.

3.2. **General Phase-Type Distributions.** For general PH distributions, namely cyclic and acyclic, the following two theorems provide the lower bound of the number of the state required to represent PH distributions.

Let $m(\mu)$ be the mean of distribution μ and $\sigma(\mu)$ be its standard deviation. Then the coefficient of variation of the distribution is defined by:

$$Cv(\mu) = \frac{\sigma(\mu)}{m(\mu)}.$$

Theorem 3.5 ([1])**.** *Consider a PH representation of size n and let μ be its PH distribution, then:*

$$Cv(\mu) \geq \frac{1}{\sqrt{n}}.$$

Moreover, the equality holds only in the case of n-state Erlang representations.

Theorem 3.5 establishes that it requires n—where $n \geq \frac{1}{Cv(\mu)^2}$—states to represent PH distributions with coefficient of variation $Cv(\mu)$. Hence, in order to obtain a low coefficient of variation, bigger PH representations are needed.

Theorem 3.6 ([8] in [6])**.** *Let $\mathbf{A}$ be a PH-generator of size n. Let $-\lambda_1$, $\lambda > 0$, be its eigenvalue with maximal real part and $-\lambda_2 \pm i\theta$, $\lambda_2 > 0$ and $\theta > 0$, be any pair of its complex eigenvalues. The following relation is satisfied:*

$$\frac{\theta}{\lambda_2 - \lambda_1} \leq \cot \frac{\pi}{n}.$$

Theorem 3.6 establishes that it requires n—where:

$$n \geq \frac{\pi}{\arctan\left(\frac{\lambda_2 - \lambda_1}{\theta}\right)},$$

—states to represent such PH distributions. Since the poles of the LST of a PH distribution are eigenvalues of the PH-generator of any of its representation, the order of the representation increases when the angle between the position of any complex poles and the vertical line passing through the real dominating pole decreases [6]. This theorem assures us that finding a PH representation of a size that is exactly equal to the algebraic degree of its PH distribution is not always possible.

3.3. **PH-Simplicity and Order.** Let $\{X_t \mid t \in \mathbb{R}^{\geq 0}\}$ be an absorbing Markov process representing a PH distribution and let τ be a random variable denoting its absorption time.

Definition 3.1 ([14])**.** *The* dual *or the* time-reversal *representation of the absorbing Markov process* $\{X_t \mid t \in \mathbb{R}^{\geq 0}\}$ *is given by an absorbing Markov process* $\{X_{\tau-t} \mid t \in \mathbb{R}^{\geq 0}\}$.

The relationship between the two processes can be described intuitively as follows: the probability of being in state s at time t in one Markov process is equal to the probability of being in state s at time $\tau - t$ in the time-reversal Markov process and vice versa.

Lemma 3.1 ([3, 5])**.** *Given a PH representation* $(\vec{\alpha}, \mathbf{A})$*, then its dual representation is* $(\vec{\beta}, \mathbf{B})$ *such that:*

$$\vec{\beta} = \vec{A}^{\mathrm{T}}\mathbf{M} \quad \textit{and} \quad \mathbf{B} = \mathbf{M}^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{M},$$

where $\mathbf{M} = \mathrm{diag}(\vec{m})$ *is a diagonal matrix whose diagonal components are formed by the components of vector* $\vec{m} = -\vec{\alpha}\mathbf{A}^{-1}$.

Lemma 3.1 provides a recipe to obtain the dual representation of a given PH representation. It is important to note that the size of both PH representation and its dual are equal.

The notion of PH-simplicity, on the other hand, was first formalized in [17] and it is closely related to the notion of simplicity in convex analysis.

Definition 3.2. *A PH-generator* $\mathbf{A}$ *(of dimension* n*) is* PH-simple *if and only if for any two* n*-dimensional substochastic vectors* $\vec{\alpha}_1$ *and* $\vec{\alpha}_2$*, where* $\vec{\alpha}_1 \neq \vec{\alpha}_2$*,* $PH(\vec{\alpha}_1, \mathbf{A}) \neq PH(\vec{\alpha}_2, \mathbf{A})$.

Theorem 3.7 ([3])**.** *Given a PH representation of size* n*. If both PH-generators of the representation and its dual representation are PH-simple, then the algebraic degree of the associated PH distribution is* n.

Theorem 3.7 establishes the relationship between PH-simplicity and the order of PH distributions. In particular, the theorem maintains that if PH-generators of a PH representation and of its dual representation are both PH-simple, than, no matter their initial probability distributions, both representations are minimal and the order of the associated PH distribution is equal to the size of the representations.

3.4. **Mixture of Monocyclic Erlang.** Figure 1 depicts an example of a monocyclic Erlang representation in graph form.

The representation has n states and ends in an absorbing state, depicted by the black circle. The representation is basically formed by a convolution of n exponential distributions of the same rate λ—hence, Erlang—but with a single cycle from the last to the first state—hence, monocyclic—with rate $\mu < \lambda$.

Mocanu and Commault in [15] show that a conjugate pair of complex poles in the LST of a PH distribution can be represented by a single monocyclic Erlang, and they proceeded to prove Theorem 3.8.

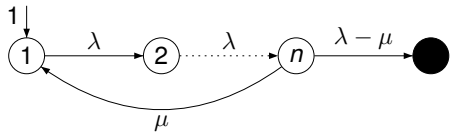

FIGURE 1. A Monocyclic Erlang Representation

Theorem 3.8 ([15])**.** *Every PH distribution has a PH representation, which is a* mixture of monocyclic Erlangs *(MME).*

Theorem 3.8 establishes that any PH distribution can be represented by a mixture of monocyclic Erlang representation. This is done by constructing an MME PH-generator based on the poles of the LST of the given PH distribution. Then a "representation"—that is not necessarily Markovian, namely whose initial probability vector is not substochastic—is formed by using the obtained PH-generator. A proper PH representation is then looked for by repeatedly constructing Euler approximants in a suitable space of probability distributions until a representation with a substochastic vector is obtained. Each approximation adds a new state to the existing, intermediate "representation" [15].

The procedure to obtain the mixture of monocyclic Erlang representation of a PH distribution is not guaranteed to end with a minimal PH representation. Hence, the order of the PH distribution still cannot be determined. However, the resulting representation is sparse, in the sense that, even though it contains more states, it contains only a small number of transitions.

In this section, we have described several partial results on the solution to the problem of the order of PH distributions. In the field of acyclic PH distributions, aside from the conditions described in the previous section, a complete solution has been found as will be explained further in the next section.

In the field of general (cyclic) PH distributions, on the other hand, the partial results are rather limited. Several lower bounds on the order of PH distributions have been discovered (*cf.* Theorems 3.5 and 3.6). A condition specifying when a (cyclic) PH representation is minimal, because its size is equal to the degree of its PH distribution, has been provided (*cf.* Theorem 3.7). An complete algorithmic solution to the problem of the order in the field of general PH distributions, however, does not exist yet. The frontier in the algorithmic solution is provided in [15] (*cf.* Theorem 3.8). The proposed algorithm, nevertheless, only produces PH representations that are sparse but not necessarily minimal. Hence, there is yet no way to determine the order of the associated PH distributions.

4. OUR CONTRIBUTION

In this section, we will explore further on the algorithmic solution to the problem in the field of acyclic PH distributions. In [10], HE and Zhang provided an algorithm

for computing minimal ordered bidiagonal representations of acyclic PH distributions. The algorithm of HE and Zhang starts by immediately transforming a given acyclic PH distribution to a representation that only contains states that represent the poles of the LST of the distribution. This representation is not necessarily a PH distribution, but certainly a matrix-exponential distribution. If this is the case, another state and its total outgoing rate are determined and appended to the representation. This is performed one by one until a PH representation is obtained. The first PH representation found is a minimal representation. The algorithm involves solving systems of non-linear equations when additional states and their total outgoing rates are to be determined. Since non-linear programming is difficult, the practicality of this algorithm for large models is not obvious, and has not been investigated so far.

In the following, we will describe our contribution to the field, namely an algorithm to reduce the size of acyclic PH representations. The algorithm is of cubic complexity in the size of the state space, and only involves standard numerical computations. The goal is to reduce the state space of the original representation one state by one state. The algorithm returns a smaller or equal size representation than the original one. However, unlike the algorithm of HE and Zhang, the result is not guaranteed to be minimal. The algorithm starts by transforming a given acyclic PH representation to its ordered bidiagonal representation. This transformation does not increase the number of states. It then proceeds by removing "unnecessary" states while maintaining the resulting representation to be phase-type. The removal of a state involves solving a system of linear equations. This removal is repeated until no more removal is possible. The algorithm is easy to implement and straightforward to parallelize. It only consists of vector-matrix multiplications and the solutions of well-conditioned systems of linear equations. Furthermore, because we are dealing with bidiagonal representations, these operations can be carried out even more efficiently.

The exposition in the rest this section is based mainly on [22]. In the following, we discuss a procedure to reduce the size of acyclic PH representations. The procedure is roughly as follows: (1) Given an acyclic PH distribution with representation $(\vec{\alpha}, \mathbf{A})$, it is transformed into an ordered bidiagonal representation $(\vec{\beta}, \mathbf{Bi}(\lambda_1, \lambda_2, \cdots, \lambda_n))$ by using the spectral polynomial algorithm[9], without increasing its size. (2) A smaller representation is obtained by eliminating unnecessary states from the ordered bidiagonal representation. If successful, the resulting representation is also an ordered bidiagonal representation with fewer states.

L-terms. The LST of an exponential distribution with rate λ is given by $\tilde{f}(s) = \frac{\lambda}{s+\lambda}$. Let $L(\lambda) = \frac{s+\lambda}{\lambda}$, *i.e.*, the reciprocal of the LST. We call a single expression of $L(\cdot)$ an *L-term*. The LST of an ordered bidiagonal representation $(\vec{\beta}, \mathbf{Bi}(\lambda_1, \lambda_2, \cdots, \lambda_n))$ can be written as:

$$\begin{aligned}\tilde{f}(s) &= \frac{\beta_1}{L(\lambda_1)\cdots L(\lambda_n)} + \frac{\beta_2}{L(\lambda_2)\cdots L(\lambda_n)} + \cdots + \frac{\beta_n}{L(\lambda_n)}, \\ &= \frac{\beta_1 + \beta_2 L(\lambda_1) + \cdots + \beta_n L(\lambda_1)\cdots L(\lambda_{n-1})}{L(\lambda_1)L(\lambda_2)\cdots L(\lambda_n)}, \end{aligned} \tag{3}$$

but this may not be in irreducible ratio form. Here the denominator polynomial corresponds exactly to the sequence of the transition rates of the ordered bidiagonal representation, and thus its degree is equal to the size of the ordered bidiagonal representation.

Reduction. Observing (3), we see that in order to remove a state from the ordered bidiagonal representation, we have to find a common L-term in both the numerator and denominator polynomials. If we find that, we might be able to drop a state from the representation. But removing a common L-term from the numerator and denominator involves redistributing the initial probability distribution. This may not be possible, because the resulting vector $\vec{\delta}$ may not be substochastic (a vector $\vec{\delta}$ is substochastic if $\delta_i \geq 0$ and $\vec{\delta}\vec{1} \leq 1$). Otherwise, a state can be removed. The procedure of identifying and properly removing a state from an ordered bidiagonal representation is based on Lemma 4.1 (see [21] for proof).

Let $ME(\vec{\alpha}, \mathbf{A}, \vec{\omega})$ denote the matrix-exponential (ME) distribution of representation $(\vec{\alpha}, \mathbf{A}, \vec{\omega})$. The set of PH distributions is a subset of ME distributions. In particular, the initial distribution vector $\vec{\alpha}$ in an ME representation is allowed to be non-stochastic vector, as long as $0 < \sum_i \vec{\alpha}_i \leq 1$. Let $\vec{1}|_x$ be a vector of dimension x whose components are all equal to 1.

Lemma 4.1. *If for some $1 \leq i \leq n$, $\beta_1 + \beta_2 L(\lambda_1) + \cdots + \beta_i L(\lambda_1) \cdots L(\lambda_{i-1})$ is divisible by $L(\lambda_i)$ then there exists a unique vector $\vec{\delta}$ such that:*

$$PH(\vec{\beta}, \mathbf{Bi}(\lambda_1, \cdots, \lambda_n)) = ME(\vec{\delta}, \mathbf{Bi}(\lambda_1, \cdots, \lambda_{i-1}, \lambda_{i+1}, \cdots, \lambda_n), \vec{1}|_{n-1}).$$

If vector $\vec{\delta}$ is substochastic, then

$$PH(\vec{\beta}, \mathbf{Bi}(\lambda_1, \cdots, \lambda_n)) = PH(\vec{\delta}, \mathbf{Bi}(\lambda_1, \cdots, \lambda_{i-1}, \lambda_{i+1}, \cdots, \lambda_n)).$$

If both conditions are fulfilled, then switching from the given representation $(\vec{\beta}, \mathbf{Bi}(\lambda_1, \cdots, \lambda_n))$ to the smaller representation $(\vec{\delta}, \mathbf{Bi}(\lambda_1, \cdots, \lambda_{i-1}, \lambda_{i+1}, \cdots, \lambda_n))$ means reducing the size from n to $n-1$. Algorithmically, we investigate the two conditions for a given λ_i. The divisibility of the numerator polynomial is obtained by checking whether $R(-\lambda_i) = 0$, where $R(s)$ is the numerator polynomial in (3). The substochasticity (*i.e.*, the absence of nonnegative components) of $\vec{\delta}$ is checked while computing it, as explained below.

Let $\mathbf{Bi}_1 := \mathbf{Bi}(\lambda_1, \cdots, \lambda_i)$, $\mathbf{Bi}_2 := \mathbf{Bi}(\lambda_1, \cdots, \lambda_{i-1})$. Lemma 4.2 (see [21] for proof) states that we can simply ignore the last $n - i$ states in both bidiagonal chains.

Lemma 4.2. *If $\delta_j = \beta_{j+1}$, for $i \leq j \leq n-1$, then:*

$$PH(\vec{\beta}, \mathbf{Bi}(\lambda_1, \cdots, \lambda_n)) = ME(\vec{\delta}, \mathbf{Bi}(\lambda_1, \cdots, \lambda_{i-1}, \lambda_{i+1}, \cdots, \lambda_n), \vec{1}|_{n-1})$$

implies:

$$PH([\beta_1, \cdots, \beta_i], \mathbf{Bi}_1) = ME([\delta_1, \cdots, \delta_{i-1}], \mathbf{Bi}_2, \vec{1}|_{i-1}). \tag{4}$$

From (4), we obtain:

$$[\beta_1, \cdots, \beta_i] \exp(\mathbf{Bi}_1 t)\vec{1}|_i = [\delta_1, \cdots, \delta_{i-1}] \exp(\mathbf{Bi}_2 t)\vec{1}|_{i-1}. \tag{5}$$

Therefore, to compute $[\delta_1, \cdots, \delta_{i-1}]$ from $[\beta_1, \cdots, \beta_i]$, $i-1$ equations relating their components are needed. Equation (5) can be evaluated at $i-1$ different t values to obtain such a required system of equations. However, such function evaluations in practice are costly, because they involve matrix exponentiations. We proceed differently.

For a PH representation $(\vec{\alpha}, \mathbf{A})$, the j-th derivative of its distribution function, for $j \in \mathbb{N}^0$, is:

$$\frac{d^j}{dt^j} F(t) = -\vec{\alpha}\mathbf{A}^j \exp(\mathbf{A}t)\vec{1}.$$

Evaluating these derivatives at $t = 0$ allows us to avoid computing the exponential of matrices. Hence, the components of vector $[\delta_1, \cdots, \delta_{i-1}]$ can be computed by solving:

$$\begin{aligned} &[\delta_1, \cdots, \delta_{i-1}]\mathbf{Bi}(\lambda_1, \cdots, \lambda_{i-1})^j \vec{1}|_{i-1} \\ &\quad = [\beta_1, \cdots, \beta_i]\mathbf{Bi}(\lambda_1, \cdots, \lambda_i)^j \vec{1}|_i, \ 0 \le j \le i-2. \end{aligned} \tag{6}$$

Once the system of equations (6) is solved, the substochasticity of vector $[\delta_1, \cdots, \delta_{i-1}]$ can be determined simply by verifying that all of its entries are nonnegative real numbers.

However, we observe that for any bidiagonal PH-generator of dimension d,

$$\mathbf{Bi}[s_i, s_i] = -\mathbf{Bi}[s_i, s_{i+1}],$$

for $1 \le i < d$. Since both PH-generators in the system of equations are bidiagonal, we can prove the following lemma [21].

Lemma 4.3. *Equations* (6) *can be transformed into:*

$$\mathbf{A}[\delta_1, \cdots, \delta_{i-1}]^\top = \vec{b}, \tag{7}$$

where $\mathbf{A}$ *is an upper triangular matrix of dimension* $i-1$.

This transformation requires $\mathcal{O}(i^2)$ multiplications and $\mathcal{O}(i^2)$ additions.

Algorithm. Lemma 4.1 can thus be turned into an algorithm that reduces the size of a given APH representation $(\alpha, \mathbf{A})$, which we give here in an intuitive form.

(1) Use SPA to turn $(\vec{\alpha}, \mathbf{A})$ into $(\vec{\beta}, \mathbf{Bi}(\lambda_1, \cdots, \lambda_n))$, which takes $\mathcal{O}(n^3)$ time.
(2) Set i to 2.
(3) While $i \le n$:
   (a) Check divisibility *w.r.t.* λ_i (*i.e.*, $R(\lambda_i) = 0$), which takes $\mathcal{O}(n)$ time.
   (b) If not divisible (*i.e.*, $R(\lambda_i) \neq 0$), continue the while-loop with i is set to $i+1$.
   Otherwise, construct (7), and then solve it by backward substitution. This takes $\mathcal{O}(n^2)$ time, and produces $(\vec{\delta}, \mathbf{Bi}(\lambda_1, \cdots, \lambda_{i-1}, \lambda_{i+1}, \cdots, \lambda_n))$. If vector $\vec{\delta}$ is substochastic (which takes $\mathcal{O}(n)$ time to check), continue with the PH representation $(\vec{\delta}, \mathbf{Bi}(\lambda_1, \cdots, \lambda_{i-1}, \lambda_{i+1}, \cdots, \lambda_n))$ and then decrease n to $n-1$; otherwise continue the while-loop with $(\vec{\beta}, \mathbf{Bi}(\lambda_1, \cdots, \lambda_n))$ and i is set to $i+1$.
(4) Return $(\vec{\beta}, \mathbf{Bi}(\lambda_1, \cdots, \lambda_n))$.

In each iteration of the while-loop, either n is decreased, or i is increased. As a consequence, the reduction algorithm terminates in $\mathcal{O}(n^3)$ time, and produces a reduced representation of the original input $(\alpha, \mathbf{A})$. We refer to [21] for a more exhaustive discussion.

Properties of the Algorithm. In the following, we discuss several properties of the proposed algorithm:

(1) Non-minimality for the general case. Even though the algorithm reduces the size of given acyclic PH representations, it does not always produce minimal representations. An example is provided in [22] to show why this is the case.
(2) Minimality for triangular ideal PH distributions. A PH distribution is called *triangular ideal* if it has acyclic PH representations whose size is equal to the degree of the PH distribution. Theorems 3.5, 3.6, and 3.7 establish conditions under which an acyclic PH distribution is triangular ideal. We may encounter such an acyclic PH distribution, however, in a representation having strictly larger size than the order of its distribution. In this case, verifying the conditions will be difficult. Even more so if we wish to build a representation of the same size as the order of the distribution. Given an acyclic PH representation whose PH distribution is triangular ideal—no matter how large the size of the representation is—the proposed algorithm is certain to produce a minimal representation [22].
(3) A realistic case study has demonstrated the use of the proposed algorithm in [22]. A further case study in [21] shows the feasibility of using the proposed algorithm to reduce the size of acyclic PH representations from a trillion of states to thousands of states.

5. CONCLUDING REMARKS

This paper has described one of the most interesting research problems in the area of phase-type distributions: the problem of their order and hence their minimal representations. Several partial solutions to the problem have been discussed. For acyclic phase-type distributions, the problem has basically been solved. An algorithm that is guaranteed to transform any given acyclic phase-type representation to its minimal representation has been proposed in [10]. Although the algorithm involves solving non-linear programming, which can be difficult, highly unstable and prone to numerical errors, this algorithm is an excellent basis for further developments and improvements. Based on our own proposed algorithm to reduce the size of acyclic phase-type representations, we think that to achieve minimality, non-linearity seems to be unavoidable.

For the general phase-type distributions, the problem is still open. Currently available partial results are restricted to conditions for minimality without algorithmic possibilities. The only algorithmic method that we are aware of is the algorithm proposed in [15]. However, this algorithm only strives for obtaining sparse representations,

not minimal ones. Nevertheless, we think that this algorithm is also an excellent basis for further developments and improvements towards an algorithm that can produce minimal representations, since the output of the algorithm is quite similar to ordered bidiagonal representations.

References

[1] Aldous, D. and Shepp, L., The least variable phase-type distribution is erlang, *Communications in Statistics: Stochastic Models*, **3**, 467-473, 1987.

[2] Bernardo, M. and Gorrieri, R., Extended Markovian process algebra, *In CONCUR 96, Concurrency Theory, 7th International Conference, Pisa, Italy, August 26-29, 1996, Proceedings, volume 1119 of Lecture Notes in Computer Science*, 315-339, Springer, 1996.

[3] Commault, C. and Chemla, J.-P., On dual and minimal phase-type representations, *Communications in Statistics: Stochastic Models*, **9**(3), 421-434, 1993.

[4] Commault, C. and Chemla, J.-P., An invariant of representations of phase-type distributions and some applications, *Journal of Applied Probability*, **33**, 368-381, 1996.

[5] Commault, C. and Mocanu, S., A generic property of phase-type representations, *Journal of Applied Probability*, **39**, 775-785, 2002.

[6] Commault, C. and Mocanu, S., Phase-type distributions and representations: Some results and open problems for system theory, *International Journal of Control*, **76**(6), 566-580, 2003.

[7] Cumani, A., On the canonical representation of homogeneous Markov processes modelling failure-time distributions, *Microelectronics and Reliability*, **22**, 583-602, 1982.

[8] Dmitriev, N. and Dynkin, E.B., On the characteristic numbers of a stochastic matrix. *Comptes Rendus (Doklady) de lAcadémie de Sciences de lURSS (Nouvelle Série)*, **49**, 159162, 1945.

[9] He, Q.-M. and Zhang, H., Spectral polynomial algorithms for computing bi-diagonal representations for phase-type distributions and matrix-exponential distributions, *Stochastic Models*, **2**(2), 289-317, 2006.

[10] He, Q.-M. and Zhang, H., An Algorithm for Computing Minimal Coxian Representations, *INFORMS Journal on Computing*, **ijoc.1070.0228**, 2007.

[11] He, Q.-M. and Zhang, H., On matrix exponential distributions, *Advances in Applied Probability*, **39**(1), 271-292, 2007.

[12] Hermanns, H., Interactive Markov Chains: The Quest for Quantified Quality, *Lecture Notes in Computer Science*, **2428**, Springer, 2002.

[13] Hillston, J., *A compositional approach to performance modelling*, Cambridge University Press, 1996.

[14] Kelly, F.P., *Reversibility and Stochastic Networks*, Wiley, 1979.

[15] Mocanu, S. and Commault, C., Sparse representation of phase-type distributions, *Communications in Statistics: Stochastic Models*, **15**(4), 759-778, 1999.

[16] Neuts, M. F., *Matrix-Geometric Solutions in Stochastic Models: An Algorithmic Approach*, Dover, 1981.

[17] O'Cinneide, C. A., On non-uniqueness of representations of phase-type distributions, *Communications in Statistics: Stochastic Models*, **5**(2), 247-259, 1989.

[18] O'Cinneide, C. A., Characterization of phase-type distributions, *Communications in Statistics: Stochastic Models*, **6**(1), 1-57, 1990.

[19] O'Cinneide, C. A., Phase-type distributions and invariant polytopes, *Advances in Applied Probability*, **23**(43), 515-535, 1991.

[20] O'Cinneide, C. A., Triangular order of triangular phase-type distributions, *Communications in Statistics: Stochastic Models*, **9**(4), 507-529, 1993.

[21] Pulungan, R., Reduction of acyclic phase-type representations, Ph.D. Dissertation, Saarland University, Saarbrücken, Germany, 2009.

[22] PULUNGAN, R. AND HERMANNS, H., Acyclic minimality by construction—almost, *Sixth International Conference on Quantitative Evaluation of Systems, QEST 2009*, 61-72, IEEE Computer Society, 2009.

REZA PULUNGAN
Department of Computer Science and Electronics,
Faculty of Mathematics and Natural Sciences,
Universitas Gadjah Mada, Yogyakarta, Indonesia.
e-mail : pulungan@ugm.ac.id