Stochastic Models, 22:289–317, 2006

ISSN: 1532-6349 print/1532-4214 online
DOI: 10.1080/15326340600649045

Taylor & Francis
Taylor & Francis Group

SPECTRAL POLYNOMIAL ALGORITHMS FOR COMPUTING BI-DIAGONAL REPRESENTATIONS FOR PHASE TYPE DISTRIBUTIONS AND MATRIX-EXPONENTIAL DISTRIBUTIONS

Qi-Ming He □ *Department of Industrial Engineering, Dalhousie University, Halifax, N.S., Canada*

Hanqin Zhang □ *Academy of Mathematics and Systems Sciences, Chinese Academy of Sciences, Beijing, P.R. China*

□ *In this paper, we develop two spectral polynomial algorithms for computing bi-diagonal representations of matrix-exponential distributions and phase type (PH) distributions. The algorithms only use information about the spectrum of the original representation and, consequently, are efficient and easy to implement. For PH-representations with only real eigenvalues, some conditions are identified for the bi-diagonal representations to be ordered Coxian representations. It is shown that every PH-representation with a symmetric PH-generator has an equivalent ordered Coxian representation of the same or a smaller order. An upper bound of the PH-order of a PH-distribution with a triangular or symmetric PH-generator is obtained.*

Keywords Coxian distribution; Invariant polytope; Matrix analytic methods; Matrix-exponential distribution; *PH*-distribution.

Mathematics Subject Classification Primary 60J27; Secondary 15A18.

1. INTRODUCTION

Phase type distributions (*PH*-distribution) were introduced by Neuts[32] as the probability distribution of the absorption time of a finite state Markov process. *PH*-distributions possess the so-called partial memoryless property, since a phase variable can be used to keep track of the state of the underlying Markov process. *PH*-distributions can approximate any non-negative probability distributions. Stochastic models with *PH*-distributions

Received May 2004; Accepted May 2005
Address correspondence to Qi-Ming He, Department of Industrial Engineering, Dalhousie University, Halifax, N.S., B3J 2X4 Canada; E-mail: Qi-Ming.he@dal.ca

are usually analytically and numerically tractable. Because of these properties, *PH*-distributions have been widely used in stochastic modeling. For instance, *PH*-distributions are used in queueing theory and reliability theory to model random variables such as service times, customer interarrival times and component life times. General references on the theory and applications of *PH*-distributions can be found in Alfa and Chakravarthy[2], Asmussen[3], Chakravarthy and Alfa[9], Commault and Mocanu[14], Latouche and Ramaswami[23], Latouche and Taylor[24,25], Neuts[33], and O'Cinneide[40].

For more effective and efficient use of *PH*-distributions in science and engineering, a number of studies on *PH*-distributions have been carried out in the past. One of these studies investigated the determination of the minimal number of phases for a given *PH*-distribution, which is known as the minimal *PH*-representation problem. This problem is important for the practical use of *PH*-distributions since a smaller representation may lead to shorter computational time and higher accuracy in parameter estimation and in performance analysis.

Soon after the introduction of *PH*-distributions, several researchers explored the basic properties of *PH*-distributions (Cumani[16]; Dehon and Latouche[17]; Neuts[32,33]). Cumani[16] showed that any *PH*-representation with a triangular *PH*-generator can be reduced to a *PH*-representation with a bi-diagonal *PH*-generator of the same or a smaller order. He demonstrated for the first time that the *PH*-representation of a *PH*-distribution can be drastically simplified, which has great significance for practical applications. Dehon and Latouche[17] established a relationship between mixtures of exponential distributions and polytopes of probability measures, which can be useful in the theoretical study of *PH*-distributions. In the late 1980's and early 1990's, more studies on the characterization of *PH*-distributions were carried out (Aldous and Shepp[1]; Asmussen[4]; Botta et al.[8]; Commault and Chemla[11]; Maier[26]; Maier and O'Cinneide[27]; Neuts[34]; and O'Cinneide[35–39]). By taking a martingale approach, Aldous and Shepp[1] found a lower bound for the *PH*-order of a *PH*-distribution in terms of its coefficient of variation. O'Cinneide[36] developed a fundamental characterization theorem for *PH*-distributions. He also introduced the concepts of *PH*-simplicity and *PH*-majorization (Refs.[35,38]), which are useful tools in the study of *PH*-distributions. In all of his publications on *PH*-distributions, O'Cinneide showed that Coxian representations are important in providing counterexamples and in determining the minimal triangular representation. In the late 1990's and early in this century, the study of *PH*-distributions advanced from focusing on a few explicitly defined problems to a greater variety of investigations. The main attention shifted from the minimal representation problem to specially structured *PH*-representations including the triangular, bi-diagonal, monocyclic, and unicyclic *PH*-representations (Commault[10]; Commault and Chemla[12]; Commault and Mocanu[13,14]; Mocanu and Commault[31]; O'Cinneide[40]; Telek[44]; and Yao[45]). Asmussen and Bladt[5] introduced matrix-exponential distributions, a class of probability distributions larger than the class of *PH*-distributions. They completely resolved the minimal representation problem of this class of probability distributions. Commault and Mocanu[14] and O'Cinneide[40] provided surveys on the current status of the study on *PH*-distributions. They recommended more studies on problems related to the minimal representations of *PH*-distributions. Commault and Mocanu[14] also proved the equivalence between the minimal *PH*-representation problem and a fundamental representation problem in control theory. Therefore, the minimal *PH*-representation problem is a problem of wide interest in stochastic modeling, statistics, and control theory.

Today, the minimal *PH*-representation problem has evolved into an area of study on problems such as finding simpler, smaller, and specially structured *PH*-representations. A number of theoretical results have been obtained, particularly on sparse representations of *PH*-distributions. However, there has been no systematic study on algorithms for computing sparse representations for *PH*-distributions or matrix-exponential distributions. The goal of this paper is to develop some algorithms that can be used to find bi-diagonal representations for *PH*-representations and matrix-exponential representations.

Finding simpler or smaller representations for a matrix-exponential representation $(\boldsymbol{\alpha}, T, \mathbf{u})$ (Asmussen and Bladt[5]) has much to do with solving the equation $TP = PS$ for matrices S and P, which are related to *PH*-majorization. It is well understood now that S can take many forms. The most interesting form of S is the bi-diagonal form. Among the bi-diagonal representations, the Coxian representation (Cox[15] and O'Cinneide[35]) and the generalized Erlang representation are of particular importance (Commault and Mocanu[14] and O'Cinneide[40]). We generalize the concept of *PH*-majorization introduced by O'Cinneide[35] and explore the relationship $TP = PS$ further when S is bi-diagonal. An explicit connection between spectral polynomials of T and the matrix P is established in this paper. Based on that relationship, algorithms are developed for computing the matrix P for bi-diagonal S and, consequently, for finding a new equivalent representation of $(\boldsymbol{\alpha}, T, \mathbf{u})$. The new representation may not be a *PH*-representation even if $(\boldsymbol{\alpha}, T, \mathbf{u})$ is a *PH*-representation, but it is always a matrix-exponential representation. In addition, the spectral polynomial approach is useful for theoretical studies. For instance, in this paper, we show that every *PH*-representation with a symmetric *PH*-generator has an ordered Coxian representation of the same or a smaller order. We also show that every *PH*-representation of order 3 with only real eigenvalues has an equivalent ordered Coxian representation of order 4 or a smaller order. A sufficient condition is identified for two matrix-exponential distributions with only real eigenvalues to be ordered according to the stochastically larger order.

The minimal *PH*-representation problem is closely related to the identifiability problem of functions of Markov processes (Blackwell and Koopmans[7]; Ito et al.[21]; and Ryden[42]). Identifiability of functions of Markov processes is mainly concerned with the relationship of two hidden processes. Ryden[42] showed that the equivalence of two *PH*-representations is a special case of the identifiability problem. Ito et al.[21] and Ryden[42] identified a set of necessary and sufficient conditions for two *PH*-representations to be equivalent. That condition can be useful in the study of the minimal representation problem of *PH*-distributions. The relationship between *PH*-generators obtained in this paper is a special case of that obtained by Ito et al.[21].

The main contribution of this paper is the introduction of two spectral polynomial algorithms for computing bi-diagonal representations of matrix-exponential distributions and for computing Coxian representations of *PH*-distributions with real eigenvalues. The algorithms only use information about the spectrum of the original representation and, consequently, are efficient and easy to implement. If information about the Jordan chains associated with the original representation is available, the algorithms can be modified to find bi-diagonal representations of the same or a smaller order. The algorithms can be used for computing simpler representations and can serve as tools to do numerical experimentations for theoretical explorations on the minimal representation and the minimal bi-diagonal representation of *PH*-distributions.

The rest of the paper is organized as follows. In Section 2, some basic concepts on matrix-exponential distributions and *PH*-distributions are introduced. The concept of *PH*-majorization for *PH*-distributions is generalized to *ME*-majorization for matrix-exponential distributions. In Section 3, the Post-T spectral polynomial algorithm is developed. In Section 4, we limit our attention to *PH*-representations with only real eigenvalues. We show that every symmetric *PH*-representation has an equivalent ordered Coxian representation of the same or a smaller order. Section 5 presents a collection of results related to Coxian representations and the Post-T spectral polynomial algorithm. The Pre-T spectral polynomial algorithm is introduced in Section 6. We demonstrate that if information about Jordan chains is available, the spectral polynomial algorithm can be modified to find bi-diagonal representations of a smaller order. Section 7 shows that all results obtained in this paper are valid for discrete time *PH*-distributions. Some comments on future research are offered to conclude this paper.

2. PRELIMINARIES

For a given $m \times m$ matrix T, its characteristic polynomial is defined as $f(\lambda) = \det(\lambda I - T)$ (i.e., the determinant of matrix $\lambda I - T$), where I is the identity matrix and m is a finite positive integer. Let spectrum $(T) = \{-\lambda_i, 1 \le i \le m\}$ be the spectrum of T (counting multiplicities), which are all the roots of $f(\lambda)$. The set spectrum(T) includes all eigenvalues of T.

An $m \times m$ matrix T with negative diagonal, non-negative off-diagonal elements, nonpositive row sums, and at least one negative row sum is called a *subgenerator* in the general literature of Markov processes. We shall call a subgenerator T a *PH-generator* (if m is finite). Consider a continuous time Markov chain with $m + 1$ states and an infinitesimal generator

$$\begin{pmatrix} T & -T\mathbf{e} \\ 0 & 0 \end{pmatrix}, \tag{2.1}$$

where the $(m+1)$st state is an absorption state and $\mathbf{e}$ is the column vector with all elements being one. We assume that states $\{1, 2, \ldots, m\}$ are transient. Let $\boldsymbol{\alpha}$ be a non-negative vector of size m for which the sum of its elements is less than or equal to one. We call the distribution of the absorption time of the Markov chain to state $m + 1$, with initial distribution $(\boldsymbol{\alpha}, 1 - \boldsymbol{\alpha}\mathbf{e})$, a *phase type distribution* (*PH*-distribution). We call the 3-tuple $(\boldsymbol{\alpha}, T, \mathbf{e})$ a *PH-representation* of that *PH*-distribution. The number m is the order of the *PH*-representation $(\boldsymbol{\alpha}, T, \mathbf{e})$. We refer to Chapter 2 in Neuts[33] for basic properties about *PH*-distributions. The probability distribution function of the *PH*-distribution is given as $1 - \boldsymbol{\alpha}\exp\{Tt\}\mathbf{e}$ for $t \ge 0$, and the density function is given as $-\boldsymbol{\alpha}\exp\{Tt\}T\mathbf{e}$ for $t \ge 0$. If $\boldsymbol{\alpha}\mathbf{e} = 0$, the distribution has a unit mass at time 0. There is no need for a *PH*-representation for such a distribution. If $\boldsymbol{\alpha}\mathbf{e} \ne 0$, the expression $\boldsymbol{\alpha}\exp\{Tt\}\mathbf{e}$ can be written as $(\boldsymbol{\alpha}\mathbf{e})(\boldsymbol{\alpha}/(\boldsymbol{\alpha}\mathbf{e}))\exp\{Tt\}\mathbf{e}$. Thus, if $\boldsymbol{\alpha}\mathbf{e} \ne 0$, a study on the representations of the *PH*-distribution $(\boldsymbol{\alpha}, T, \mathbf{e})$ is equivalent to that of $(\boldsymbol{\alpha}/(\boldsymbol{\alpha}\mathbf{e}), T, \mathbf{e})$. Without loss of generality, we shall assume that $\boldsymbol{\alpha}$ is a *probability vector* (a non-negative vector for which the sum of all its elements is one) or has a unit sum. Throughout this paper, we assume that all probability distributions have a zero mass at $t = 0$.

It is possible that $1 - \boldsymbol{\alpha}\exp\{Tt\}\mathbf{u}$ is a probability distribution function for a row vector $\boldsymbol{\alpha}$ of size m, an $m \times m$ matrix T, and a column vector $\mathbf{u}$ of size m, where the elements of $\boldsymbol{\alpha}$, T, and $\mathbf{u}$ can be complex numbers. For this case, the 3-tuple $(\boldsymbol{\alpha}, T, \mathbf{u})$ is called a *matrix-exponential representation* of a *matrix-exponential distribution*. Without loss of generality, we assume that $\boldsymbol{\alpha}\mathbf{u} = 1$ (i.e., zero mass at $t = 0$). *PH*-representations are special matrix-exponential representations. We refer to Asmussen and Bladt[5] for more details about matrix-exponential distributions.

The matrix T can be considered as a linear mapping. A polytope conv $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$ is defined as the convex combinations of the column vectors $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$ of size m (Rockafellar[41]). The polytope $\mathrm{conv}\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$ is *invariant under* T if

$$T\mathbf{x}_i = \sum_{j=1}^{N} \mathbf{x}_j s_{j,i}, \quad 1 \le i \le N. \tag{2.2}$$

If the $N \times N$ matrix $S = (s_{i,j})$ is a *PH*-generator, the polytope $\mathrm{conv}\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$ is *PH*-invariant under T (*PH*-invariant polytope). Let P be an $m \times N$ matrix with columns $\{\mathbf{x}_i, 1 \le i \le N\}$, i.e., with $\mathbf{x}_1$ as its first column, $\mathbf{x}_2$ the second column,..., and $\mathbf{x}_N$ the Nth column. Then equation (2.2) becomes $TP = PS$. Let $\boldsymbol{\beta} = \boldsymbol{\alpha} P$. If there exists a vector $\mathbf{v}$ such that $\mathbf{u} = P\mathbf{v}$, then $(\boldsymbol{\alpha}, T, \mathbf{u})$ finds an equivalent representation $(\boldsymbol{\beta}, S, \mathbf{v})$ (Proposition 2.1). The objective of this paper is to find $(\boldsymbol{\beta}, S, \mathbf{v})$ for $(\boldsymbol{\alpha}, T, \mathbf{u})$, where S is a bi-diagonal matrix.

Similarly, for row vectors $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$, the polytope $\mathrm{conv}\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$ is invariant under T if

$$\mathbf{x}_i T = \sum_{j=1}^{N} s_{i,j} \mathbf{x}_j, \quad 1 \le i \le N. \tag{2.3}$$

Let $\mathbf{e}_i$ be a row vector of size m with the ith element being one and all others zero, $1 \le i \le m$. Apparently, the polytope $\mathrm{conv}\{\mathbf{e}_i, 1 \le i \le m\}$ is *PH*-invariant under T (i.e., $IT = TI$). In this paper, we use equation (2.2) for the definition of invariant polytope when column vectors are involved and equation (2.3) when row vectors are involved.

For $\mathbf{x} = (x_1, x_2 \ldots, x_N)$, a bi-diagonal matrix $S(\mathbf{x})$ is defined as

$$S(\mathbf{x}) = \begin{pmatrix} -x_1 & 0 & \cdots & \cdots & 0 \\ x_2 & -x_2 & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & x_{N-1} & -x_{N-1} & 0 \\ 0 & \cdots & 0 & x_N & -x_N \end{pmatrix}. \tag{2.4}$$

If $\{x_1, x_2 \ldots, x_N\}$ are all real and positive and $\boldsymbol{\beta}$ is a probability vector, then $S(\mathbf{x})$ is called a *Coxian generator* and $(\boldsymbol{\beta}, S(\mathbf{x}), \mathbf{e})$ is called a *Coxian representation* that represents a *Coxian distribution* (Cox[15] and O'Cinneide[35]). Furthermore, if $x_1 \ge x_2 \ge \cdots \ge x_N > 0$, then $(\boldsymbol{\beta}, S(\mathbf{x}), \mathbf{e})$ is called an *ordered Coxian representation.*

The relationship between two *PH*-generators satisfying the equation $TP = PS$ for a non-negative matrix P with unit row sums was first investigated by O'Cinneide[35]. For a given *PH*-generator T, let $PH(T)$

denote the set of all *PH*-distributions with a *PH*-representation of the form $(\boldsymbol{\alpha}, T, \mathbf{e})$. For two *PH*-generators T and S, S is said to *PH-majorize* T if $PH(T) \subseteq PH(S)$. O'Cinneide[35] showed that S *PH*-majorizes T if and only if there exists a non-negative matrix P with unit row sums such that $TP = PS$. We extend the concept of *PH*-majorization to *ME*-majorization. Let $ME(T, \mathbf{u})$ denote the set of all probability distributions with a matrix-exponential representation of the form $(\boldsymbol{\alpha}, T, \mathbf{u})$. For two pairs $\{T, \mathbf{u}\}$ and $\{S, \mathbf{v}\}$, $\{S, \mathbf{v}\}$ *ME-majorizes* $\{T, \mathbf{u}\}$ if $ME(T, \mathbf{u}) \subseteq ME(S, \mathbf{v})$.

Proposition 2.1. *Assume that $\{T, \mathbf{u}\}$ and $\{S, \mathbf{v}\}$ are of orders m and N, respectively. If there exists an $m \times N$ matrix P such that $TP = PS$ and $\mathbf{u} = P\mathbf{v}$, then $(\boldsymbol{\alpha}P,\ S,\ \mathbf{v})$ is an equivalent matrix-exponential representation of matrix-exponential representation $(\boldsymbol{\alpha},\ T,\ \mathbf{u})$. Consequently, $\{S, \mathbf{v}\}$ ME-majorizes $\{T, \mathbf{u}\}$.*

Proof. Since $TP = PS$, it is easy to verify that $T^nP = PS^n$ for $n \geq 0$. Then we have, for $t \geq 0$,

$$\exp\{Tt\}\mathbf{u} = \exp\{Tt\}P\mathbf{v} = \sum_{n=0}^{\infty} \frac{t^n}{n!} T^n P\mathbf{v} = \sum_{n=0}^{\infty} \frac{t^n}{n!} PS^n\mathbf{v} = P\exp\{St\}\mathbf{v}, \quad (2.5)$$

which implies that $(\boldsymbol{\alpha}P,\ S,\ \mathbf{v})$ and $(\boldsymbol{\alpha},\ T,\ \mathbf{u})$ have the same distribution for any $\boldsymbol{\alpha}$ such that $(\boldsymbol{\alpha},\ T,\ \mathbf{u})$ is a matrix-exponential distribution. Thus, $ME(T, \mathbf{u}) \subseteq ME(S, \mathbf{v})$, i.e., $\{S, \mathbf{v}\}$ *ME*-majorizes $\{T, \mathbf{u}\}$. This completes the proof of Proposition 2.1.

Note 2.1. Note that the condition $\boldsymbol{\alpha}\mathbf{u} = 1$ is not used in the proof of Proposition 2.1. Hence, the proposition also holds without that condition.

Note 2.2. The conditions in Proposition 2.1 are sufficient but not necessary for *ME*-majorization. For example, consider the following $\{T, \mathbf{u}\}$ and $\{S, \mathbf{v}\}$:

$$T = \begin{pmatrix} \sqrt{-1} & 0 \\ 0 & -1 \end{pmatrix}, \quad \mathbf{u} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad S = (-1), \quad \text{and} \quad \mathbf{v} = (1). \quad (2.6)$$

It is easy to show that the matrix-exponential distributions with a representation $(\boldsymbol{\alpha},\ T,\ \mathbf{u})$ must have $\boldsymbol{\alpha} = (0, \alpha_2)$, where $0 \leq \alpha_2 \leq 1$. Therefore, $ME(T, \mathbf{u}) \subseteq ME(S, \mathbf{v})$. However, there is no matrix P satisfying $TP = PS$ and $\mathbf{u} = P\mathbf{v}$.

3. POST-*T* SPECTRAL POLYNOMIAL ALGORITHM

In this section, we develop an algorithm for computing bi-diagonal representations of a matrix-exponential representation $(\boldsymbol{\alpha},\ T,\ \mathbf{u})$ of order m. The basic idea is to use the *spectral polynomials* of matrix T to construct an invariant polytope under T.

We assume that $\{x_1, x_2, \dots, x_N\}$ are nonzero complex numbers, where N is a positive integer. Let $\mathbf{x} = (x_1, x_2, \dots, x_N)$. For a given column vector $\mathbf{p}_1$ of size m, we define

$$\begin{cases} \mathbf{p}_n = (x_{n-1}I + T)\mathbf{p}_{n-1}/x_n, \quad 2 \le n \le N; \\ \mathbf{p}_{N+1} = (x_N I + T)\mathbf{p}_N. \end{cases} \tag{3.1}$$

If $\mathbf{p}_{N+1} = 0$, it is easy to see that $T\mathbf{p}_n = -x_n\mathbf{p}_n + x_{n+1}\mathbf{p}_{n+1}$, for $1 \le n \le N-1$, and $T\mathbf{p}_N = -x_N\mathbf{p}_N$, which can be written into the following matrix form

$$TP = PS(\mathbf{x}), \tag{3.2}$$

where P is an $m \times N$ matrix whose columns are $\{\mathbf{p}_i, 1 \le i \le N\}$ and $S(\mathbf{x})$ is the bi-diagonal matrix defined in equation (2.4). By the definition given in equation (2.2), equation (3.2) implies that conv$\{\mathbf{p}_i, 1 \le i \le N\}$ is an invariant polytope under T. It is readily seen that, if the vectors $\mathbf{x}$ and $\mathbf{p}_1$ are properly chosen, we may have $P\mathbf{v} = \mathbf{u}$ for some $\mathbf{v}$ and, consequently, $\{S(\mathbf{x}), \mathbf{v}\}$ *ME*-majorizes $\{T, \mathbf{u}\}$ (Proposition 2.1). Thus, the issue of interest becomes how to choose $\mathbf{x}$ and $\mathbf{p}_1$ for the given pair $\{T, \mathbf{u}\}$ so that equation (3.2) and $P\mathbf{v} = \mathbf{u}$ hold. It turns out that there can be many solutions to the problem and we present one here.

Since we expect $TP = PS(\mathbf{x})$, the eigenvalues of T should be included in the set $\{-x_1, -x_2, \dots, -x_N\}$, which is the spectrum of $S(\mathbf{x})$. That leads to a specific selection of $\mathbf{x}$ and $N : \mathbf{x} = \boldsymbol{\lambda} = (\lambda_1, \lambda_2, \dots, \lambda_m)$ and $N = m$, where $-\boldsymbol{\lambda}$ is the spectrum of T. For this choice of $\mathbf{x}$ and N, if $\lambda_i \neq 0$, for $1 \le i \le m$, equation (3.1) becomes,

$$\mathbf{p}_n = \begin{cases} \dfrac{1}{\lambda_n \cdots \lambda_2}(\lambda_{n-1}I + T) \cdots (\lambda_1 I + T)\mathbf{p}_1, \quad 2 \le n \le m; \\ \dfrac{1}{\lambda_m \cdots \lambda_2}(\lambda_m I + T) \cdots (\lambda_1 I + T)\mathbf{p}_1, \qquad n = m+1. \end{cases} \tag{3.3}$$

The matrices $\{(\lambda_{n-1}I + T) \cdots (\lambda_1 I + T), 2 \le n \le m+1\}$ appeared in equation (3.3) are called the *spectral polynomials* of T. If $n = m+1$, by the Cayley–Hamilton theorem (see Lancaster and Tismenetsky[22]), $f(T) = (\lambda_m I + T) \cdots (\lambda_1 I + T) = 0$. Consequently, we have $\mathbf{p}_{m+1} = 0$ and $TP = PS(\boldsymbol{\lambda})$. Next, we choose $\mathbf{p}_1$ so that the matrix P satisfies $P\mathbf{v} = \mathbf{u}$ for some $\mathbf{v}$. For that purpose, we use the following identity derived from the equality $f(T) = 0$:

$$\left(\prod_{j=1}^{m} \lambda_j\right) I + \left(\prod_{j=2}^{m} \lambda_j\right) T + \sum_{n=1}^{m-1} \left(\prod_{j=n+2}^{m} \lambda_j\right)\left(\prod_{j=1}^{n} (\lambda_j I + T)\right) T = 0. \tag{3.4}$$

Suppose that $\mathbf{p}_1 = -T\mathbf{u}/\lambda_1$. Post multiplying $\mathbf{u}/(\lambda_1\lambda_2\cdots\lambda_m)$ on both sides of equation (3.4) yields $\mathbf{u} - \mathbf{p}_1 - \mathbf{p}_2 - \cdots - \mathbf{p}_m = 0$. Therefore, we obtain $P\mathbf{e} = \mathbf{u}$. According to Proposition 2.1, there exists an equivalent bi-diagonal representation $(\boldsymbol{\beta}, S(\boldsymbol{\lambda}), \mathbf{e})$ for any matrix-exponential representation $(\boldsymbol{\alpha}, T, \mathbf{u})$, where $\boldsymbol{\beta} = \alpha P$. For the two representations, we have $\boldsymbol{\beta}\mathbf{e} = \boldsymbol{\alpha}\mathbf{u} = 1$.

In summary, we propose the following algorithm for computing $(\boldsymbol{\beta}, S(\boldsymbol{\lambda}), \mathbf{e})$.

Post-*T* Spectral Polynomial Algorithm

Consider a matrix-exponential representation $(\boldsymbol{\alpha}, T, \mathbf{u})$ of order m. We assume that all eigenvalues of T are nonzero.

Step 1: Find the spectrum $\{-\lambda_1, -\lambda_2, \ldots, -\lambda_m\}$ of T.
Step 2: Compute $\mathbf{p}_1 = -T\mathbf{u}/\lambda_1$, and $\{\mathbf{p}_n, 2 \le n \le m\}$ by equation (3.3). Set $N = \min\{n : \mathbf{p}_n = 0\} - 1$.
Step 3: Construct the bi-diagonal matrix $S(\boldsymbol{\lambda})$ of size N for $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_N)$. Let $P = (\mathbf{p}_1, \ldots, \mathbf{p}_N)$ and compute $\boldsymbol{\beta} = \boldsymbol{\alpha}P$. Then $(\boldsymbol{\beta}, S(\boldsymbol{\lambda}), \mathbf{e})$ gives an equivalent bi-diagonal representation to $(\boldsymbol{\alpha}, T, \mathbf{u})$.

Note 3.1. In the Post-T spectral polynomial algorithm, we have chosen $\mathbf{p}_1 = -T\mathbf{u}/\lambda_1$. That choice of $\mathbf{p}_1$ is unique since we require $P\mathbf{e} = \mathbf{u}$. In fact, post-multiplying $\mathbf{e}$ on both sides of $TP = PS(\boldsymbol{\lambda})$, we obtain $TP\mathbf{e} = -\lambda_1\mathbf{p}_1$. Since all eigenvalues of T are nonzero, T is invertible. Thus, $\mathbf{p}_1 = -T\mathbf{u}/\lambda_1$ if and only if $P\mathbf{e} = \mathbf{u}$.

Note 3.2. We like to point out that the Jordan canonical form of T leads to a bi-diagonal matrix-exponential representation. However, finding such a bi-diagonal representation needs information about the Jordan chains and the Jordan canonical form of T, which can be difficult to obtain even for m as large as 5.

It is readily seen that all matrix-exponential representations with nonzero eigenvalues have bi-diagonal representations of the same or a smaller order. If $\{\lambda_1, -\lambda_2, \ldots, -\lambda_m\}$ are ordered differently, the matrix P and the corresponding $S(\boldsymbol{\lambda})$ can be different. For a given $\boldsymbol{\alpha}$, the vector $\boldsymbol{\beta} = \boldsymbol{\alpha}P$ can be different as well. Denote by $f_T(\lambda) = (\lambda + \mu_1)(\lambda + \mu_2)\ldots(\lambda + \mu_K)$ the minimal polynomial of T, where K is the degree of the minimal polynomial. It is well known that $f_T(T) = 0$, $\{-\mu_1, -\mu_2, \ldots, -\mu_K\}$ is a subset of spectrum(T), and $K \le m$. (It is well known that $f_T(\lambda)$ divides all annihilating polynomials of T (Theorem 1, page 224, Lancaster and Tismenetsky[22])). Thus, if $K < m$, the corresponding representation

$(\boldsymbol{\beta}, S(\boldsymbol{\lambda}), \mathbf{e})$ has an order lower than that of $(\boldsymbol{\alpha}, T, \mathbf{u})$. That is why the constant N was introduced in the Post-T spectral polynomial algorithm. Note that, depending on the vector $\mathbf{u}$, it is possible that $N < K$ (see Example 6.1). In the next proposition, we summarize the above results and give an alternative expression for the vector $\boldsymbol{\beta}$.

Proposition 3.1. *Assume that T has no zero eigenvalue. Then $\{S(\boldsymbol{\lambda}), \mathbf{e}\}$ ME-majorizes $\{T, \mathbf{u}\}$. A matrix-exponential representation $(\boldsymbol{\alpha}, T, \mathbf{u})$ has an equivalent bi-diagonal representation $(\boldsymbol{\beta}, S(\boldsymbol{\lambda}), \mathbf{e})$ of the same or a smaller order with $\boldsymbol{\beta}\mathbf{e} = \boldsymbol{\alpha}\mathbf{u} = 1$ and $\boldsymbol{\beta} = \boldsymbol{\alpha}P$ or*

$$\beta_n = \begin{cases} \dfrac{1}{\lambda_1} F^{(1)}(0), & n = 1; \\[2ex] \dfrac{1}{\lambda_1\lambda_2\cdots\lambda_n} \displaystyle\sum_{k=1}^{n} \left(F^{(k)}(0) \sum_{\substack{\{j_1, j_2, \ldots, j_{n-k}\}:\text{subset of } \{1,2,\ldots,n-1\} \\ j_1 < j_2 < \cdots < j_{n-k}}} \lambda_{j_1}\lambda_{j_2}\cdots\lambda_{j_{n-k}} \right), & 2 \le n \le N, \end{cases} \tag{3.5}$$

where $F^{(k)}(0)$ is the kth derivative of the probability distribution function $F(t) = 1 - \boldsymbol{\alpha}\exp\{Tt\}\mathbf{u}$ at $t = 0$, which is given by $F^{(k)}(0) = -\boldsymbol{\alpha}T^k\mathbf{u}, k \ge 1$.

Proof. We only need to show equation (3.5). By equation (3.3), we have

$$\beta_n = \begin{cases} \boldsymbol{\alpha}\mathbf{p}_n = \dfrac{-1}{\lambda_1}\boldsymbol{\alpha}T\mathbf{u}, & n = 1; \\[2ex] \boldsymbol{\alpha}\mathbf{p}_n = \dfrac{-1}{\lambda_1\lambda_2\cdots\lambda_n}\boldsymbol{\alpha}(\lambda_{n-1}I + T)\cdots(\lambda_1 I + T)T\mathbf{u}, & n \ge 2, \end{cases} \tag{3.6}$$

which leads to equation (3.5) directly. This completes the proof of Proposition 3.1.

Equation (3.5) indicates that a bi-diagonal representation of a matrix-exponential distribution $F(t)$ can be found from the poles of its Laplace-Stieltjes transform and $\{F^{(k)}(0), k \ge 1\}$. That may lead to new methods for parameter estimation of matrix-exponential distributions and *PH*-distributions (Asmussen et al.[6]). However, a study in that direction is beyond the scope of this paper and we shall not explore this direction further. In He and Zhang[19], equation (3.5) is utilized in finding a minimal ordered Coxian representation for *PH*-distributions whose Laplace-Stieltjes transforms have only real poles.

Next, we present two numerical examples and discuss the possible outcomes from the Post-T spectral polynomial algorithm.

Example 3.1. Consider Example 2.1 in Asmussen and Bladt[5] (see Cox[15]). In that example, the matrix-exponential representation $(\boldsymbol{\alpha}, T, \mathbf{u})$ is given as $\boldsymbol{\alpha} = (1, 0, 0), \mathbf{u} = \mathbf{e}$, and T in equation (3.7). The eigenvalues of T are $\{-1 + 6.2832\sqrt{-1}, -1 - 6.2832\sqrt{-1}, -1\}$. By using the Post-$T$ spectral polynomial algorithm, the matrix P and $S(\boldsymbol{\lambda})$ are obtained and presented in equations (3.8) and (3.9), respectively.

$$T = \begin{pmatrix} 0 & -1-4\pi^2 & 1+4\pi^2 \\ 3 & 2 & -6 \\ 2 & 2 & -5 \end{pmatrix}. \tag{3.7}$$

$$P = \begin{pmatrix} 0 & 0 & 1 \\ 0.0247 + 0.1552\sqrt{-1} & -0.0741 - 0.1552\sqrt{-1} & 1.0494 \\ 0.0247 + 0.1552\sqrt{-1} & -0.0494 - 0.1552\sqrt{-1} & 1.0247 \end{pmatrix}. \tag{3.8}$$

$$S(\boldsymbol{\lambda}) = \begin{pmatrix} -1 + 6.2832\sqrt{-1} & 0 & 0 \\ 1 + 6.2832\sqrt{-1} & -1 - 6.2832\sqrt{-1} & 0 \\ 0 & 1 & -1 \end{pmatrix}. \tag{3.9}$$

An alternative matrix-exponential representation for $(\boldsymbol{\alpha}, T, \mathbf{u})$ is given as $(\boldsymbol{\beta}, S(\boldsymbol{\lambda}), \mathbf{e})$ with $\boldsymbol{\beta} = (0, 0, 1)$.

Example 3.2. Consider a *PH*-representation $(\boldsymbol{\alpha}, T, \mathbf{e})$ with T given in equation (3.10). The eigenvalues of T are $\{-6.1830 + 2.4352\sqrt{-1}, -6.1830 - 2.4352\sqrt{-1}, -0.6341\}$. Using the Post-$T$ spectral polynomial algorithm, matrices P and $S(\boldsymbol{\lambda})$ can be found and are presented in equations (3.11) and (3.12), respectively. If $\boldsymbol{\alpha} = (1, 0, 0)$, $(\boldsymbol{\alpha}, T, \mathbf{e})$ has an equivalent bi-diagonal representation $(\boldsymbol{\beta}, S(\boldsymbol{\lambda}), \mathbf{e})$, where $\boldsymbol{\beta} = (0.28 + 0.1103\sqrt{-1}, 0.0989 - 0.1103\sqrt{-1}, 0.6211)$. If $\boldsymbol{\alpha} = (0, 1, 0)$, $(\boldsymbol{\alpha}, T, \mathbf{e})$ has an equivalent bi-diagonal representation $(\boldsymbol{\beta}, S(\boldsymbol{\lambda}), \mathbf{e})$, where $\boldsymbol{\beta} = (0, 0.0906, 0.9094)$.

$$T = \begin{pmatrix} -6 & 0 & 4 \\ 2 & -2 & 0 \\ 0 & 4 & -5 \end{pmatrix}. \tag{3.10}$$

$$P = \begin{pmatrix} 0.28 + 0.1103\sqrt{-1} & 0.0989 - 0.1103\sqrt{-1} & 0.6211 \\ 0 & 0.0906 & 0.9094 \\ 0.14 + 0.0551\sqrt{-1} & 0.0268 - 0.0551\sqrt{-1} & 0.8332 \end{pmatrix}. \tag{3.11}$$

$$S(\boldsymbol{\lambda}) = \begin{pmatrix} -6.183 + 2.4352\sqrt{-1} & 0 & 0 \\ 6.183 + 2.4352\sqrt{-1} & -6.183 - 2.4352\sqrt{-1} & 0 \\ 0 & 0.6341 & -0.6341 \end{pmatrix}. \tag{3.12}$$

Since not all eigenvalues of T are real, a *PH*-representation $(\boldsymbol{\alpha}, T, \mathbf{e})$ may not have an equivalent bi-diagonal *PH*-representation, except for some special probability vector $\boldsymbol{\alpha}$. For example, if $\boldsymbol{\alpha}$ is the non-negative eigenvector (normalized to have a unit sum) corresponding to the eigenvalue -0.6341, i.e., $\boldsymbol{\alpha}$ satisfies $\boldsymbol{\alpha}T = -0.6341\boldsymbol{\alpha}$, then $(\boldsymbol{\alpha}, T, \mathbf{e})$ represents an exponential distribution with parameter 0.6341. An extension of the Post-T spectral polynomial algorithm discussed in Section 6 can be used to deal with such cases.

To end this section, we briefly discuss the time and space complexity of the Post-T spectral polynomial algorithm. First, the algorithm is simple and only depends on the spectrum of T. Thus, it is quite feasible to implement the algorithm. Given that the spectrum of T is available, the space complexity of the algorithm is $O(m^2)$ and the time complexity is $O(m^3)$. By using MatLab, a straightforward implementation of the algorithm performs properly for m up to 50. If $m > 50$, the algorithm becomes instable since cumulated machine errors may become significant for such cases.

Note 3.3. For a given matrix-exponential representation $(\boldsymbol{\alpha}, T, \mathbf{u})$ with spectrum $-\boldsymbol{\lambda}$, finding the matrix P is equivalent to solving a linear system: $TP = PS(\boldsymbol{\lambda})$ and $P\mathbf{e} = \mathbf{u}$. The linear system is equivalent to

$$\phi(P)(T' \otimes I - I \otimes S(\boldsymbol{\lambda}), I \otimes \mathbf{e}) = (0, \mathbf{u}'), \tag{3.13}$$

where $\phi(P)$ is the direct-sum of P, $\otimes$ is for Kronecker product of matrices, and T' (or $\mathbf{u}'$) is the transpose of T (or $\mathbf{u}$). For a straightforward implementation, the space complexity for solving the linear equation (3.13) is $O(m^4)$ and the time complexity is $O(m^6)$, which are significantly larger than that of the Post-T spectral polynomial algorithm. Note that a linear system approach for computing the new representation was proposed in O'Cinneide[35] for *PH*-representations with a *PH*-simple triangular generator.

4. *PH*-REPRESENTATIONS WITH ONLY REAL EIGENVALUES

In this section, we focus on finding bi-diagonal *PH*-representations of a *PH*-representation $(\boldsymbol{\alpha}, T, \mathbf{e})$ for which all eigenvalues of T are real. First, we summarize some conclusions related to the Post-T spectral polynomial algorithm for *PH*-representations.

Proposition 4.1. *For a given PH-representation* $(\boldsymbol{\alpha}, T, \mathbf{e})$ *with only real eigenvalues, the following conclusions hold.*

4.1.1) P *is a matrix with real elements and unit row sums, i.e.,* $P\mathbf{e} = \mathbf{e}$. *The first column of* P *– the vector* $\mathbf{p}_1$ *– is always non-negative. Further, if* $\lambda_N = \min\{\lambda_1, \lambda_2, \ldots, \lambda_m\}$, *then* $\mathbf{p}_N$ *is non-negative.*
4.1.2) $S(\boldsymbol{\lambda})$ *is a Coxian generator. The matrix-exponential representation* $(\boldsymbol{\beta}, S(\boldsymbol{\lambda}), \mathbf{e})$ *with* $\boldsymbol{\beta} = \boldsymbol{\alpha}P$ *has the same distribution as* $(\boldsymbol{\alpha}, T, \mathbf{e})$.
4.1.3) $\{S(\boldsymbol{\lambda}), \mathbf{e}\}$ *ME-majorizes* $\{T, \mathbf{e}\}$. *Furthermore,* $S(\boldsymbol{\lambda})$ *PH-majorizes* T *if and only if the matrix* P *is non-negative.*
4.1.4) *If* $\boldsymbol{\beta} = \boldsymbol{\alpha}P$ *is non-negative, then* $(\boldsymbol{\beta}, S(\boldsymbol{\lambda}), \mathbf{e})$ *is an equivalent Coxian representation of* $(\boldsymbol{\alpha}, T, \mathbf{e})$. *If* $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N$ *and* $\boldsymbol{\beta} = \boldsymbol{\alpha}P$ *is non-negative, then* $(\boldsymbol{\beta}, S(\boldsymbol{\lambda}), \mathbf{e})$ *is an equivalent ordered Coxian representation of* $(\boldsymbol{\alpha}, T, \mathbf{e})$.

Here the matrices P and $S(\boldsymbol{\lambda})$ are given by the Post-T spectral polynomial algorithm introduced in Section 3.

Proof. Since T is a *PH*-generator, the assumption that all its eigenvalues are real implies that all its eigenvalues are negative. Then $S(\boldsymbol{\lambda})$ is a Coxian generator.

First note that $-T\mathbf{e}$ is non-negative since T is a *PH*-generator, which implies that $\mathbf{p}_1$ is non-negative. The last part of 4.1.1 is obtained from the fact that $-(\lambda_{N-1}I + T)\cdots(\lambda_1 I + T)T\mathbf{e}$ is an eigenvector of T corresponding to the eigenvalue $-\lambda_N$. Since $\lambda_N = \min\{\lambda_1, \lambda_2, \ldots, \lambda_m\}$, $-\lambda_N$ is the eigenvalue of T with the largest real part. By the Perron-Probenius theory (Minc[30]), we must have either

$$-(\lambda_{N-1}I + T)\cdots(\lambda_1 I + T)T\mathbf{e} \leq 0 \quad \text{or} \quad -(\lambda_{N-1}I + T)\cdots(\lambda_1 I + T)T\mathbf{e} \geq 0. \tag{4.1}$$

If $-(\lambda_{N-1}I + T)\cdots(\lambda_1 I + T)T\mathbf{e} \leq 0$ with at least one negative element, let $\mathbf{y}$ denote the nonzero and non-negative *left* eigenvector of T corresponding to eigenvalue $-\lambda_N$ and corresponding to the *right* eigenvector $-(\lambda_{N-1}I + T)\cdots(\lambda_1 I + T)T\mathbf{e}$. Then the vector $\mathbf{y}$ can be so chosen that $\mathbf{y}$ and $-(\lambda_{N-1}I + T)\cdots(\lambda_1 I + T)T\mathbf{e}$ are not orthogonal to each other. Consequently, we have $-\mathbf{y}(\lambda_{N-1}I + T)\cdots(\lambda_1 I + T)T\mathbf{e} < 0$. However, since $\lambda_N = \min\{\lambda_1, \lambda_2, \ldots, \lambda_m\} > 0$, we have

$$-\mathbf{y}(\lambda_{N-1}I + T)\cdots(\lambda_1 I + T)T\mathbf{e} = (\lambda_{N-1} - \lambda_N)\cdots(\lambda_1 - \lambda_N)\lambda_N(\mathbf{ye}) \geq 0, \tag{4.2}$$

which is a contradiction. Therefore, $-(\lambda_{N-1}I + T)\cdots(\lambda_1 I + T)T\mathbf{e}$ is non-negative, i.e., $\mathbf{p}_N$ is non-negative if $\min\{\lambda_1, \lambda_2, \ldots, \lambda_m\} = \lambda_N$. This proves 4.1.1.

4.1.2 and 4.1.3 are immediate from Proposition 3.1. 4.1.4 is obtained by definitions and the fact that $S(\boldsymbol{\lambda})$ is a Coxian or an ordered Coxian generator. This completes the proof of Proposition 4.1.

Note 4.1. Dehon and Latouche[17] showed that the *PH*-invariant polytope $\text{conv}\{\mathbf{e}_i, 1 \leq i \leq m\}$ is located in the intersection of two half spaces. Geometrically, that means that at least two coordinates of the vector $\boldsymbol{\beta} = \boldsymbol{\alpha}P$ are always non-negative for every $\boldsymbol{\alpha}$ in $\text{conv}\{\mathbf{e}_i, 1 \leq i \leq m\}$. That implies that at least two columns of P are non-negative, which gives a geometric interpretation to Proposition 4.1.1.

Next, let us have a look at the following example.

Example 4.1. We consider two *PH*-representations $(\boldsymbol{\alpha}, T_1, \mathbf{e})$ and $(\boldsymbol{\alpha}, T_2, \mathbf{e})$ with

$$T_1 = \begin{pmatrix} -5 & 0 & 0.5 \\ 1.8 & -2 & 0 \\ 0 & 2 & -3 \end{pmatrix} \quad \text{and} \quad T_2 = \begin{pmatrix} -5 & 3 & 0.5 \\ 1 & -7 & 1 \\ 1 & 1.8 & -3 \end{pmatrix}. \tag{4.3}$$

The matrix T_1 has eigenvalues $-\boldsymbol{\lambda}_1 = (-4.5397, -3.8446, -1.6158)$. The matrix T_2 has eigenvalues $-\boldsymbol{\lambda}_2 = (-8.1071, -4.8815, -2.0114)$. By using the Post-$T$ spectral polynomial algorithm, the corresponding matrix P for the case with $\lambda_1 > \lambda_2 > \lambda_3$ can be obtained as

$$P_1 = \begin{pmatrix} 0.9913 & -0.0900 & 0.0988 \\ 0.0441 & 0.4932 & 0.4627 \\ 0.2203 & 0.1111 & 0.6686 \end{pmatrix} \quad \text{and} \quad P_2 = \begin{pmatrix} 0.1850 & 0.4993 & 0.3157 \\ 0.6167 & 0.1828 & 0.2004 \\ 0.0247 & 0.2911 & 0.6842 \end{pmatrix}, \tag{4.4}$$

for T_1 and T_2, respectively. For T_1, P_1 is not non-negative. Therefore, some *PH*-distributions $(\boldsymbol{\alpha}, T_1, \mathbf{e})$ have a Coxian representation $(\boldsymbol{\beta}, S(\boldsymbol{\lambda}_1), \mathbf{e})$ while others do not. For instance, if $\boldsymbol{\alpha} = (1, 0, 0)$, the corresponding $\boldsymbol{\beta} = (0.9913, -0.0900, 0.0988)$ is not non-negative. If $\boldsymbol{\alpha} = (0, 1, 0)$, the corresponding $\boldsymbol{\beta} = (0.0441, 0.4932, 0.4627)$ is non-negative. Note that T_1 has a structure close to the unicyclic representation defined in O'Cinneide[40], which explains partially why for this case we do not have $PH(T_1) \subseteq PH(S(\boldsymbol{\lambda}_1))$. For T_2, we have $PH(T_2) \subseteq PH(S(\boldsymbol{\lambda}_2))$ since P_2 is non-negative (O'Cinneide[35]). A geometric interpretation of these results can be found in Figure 1.

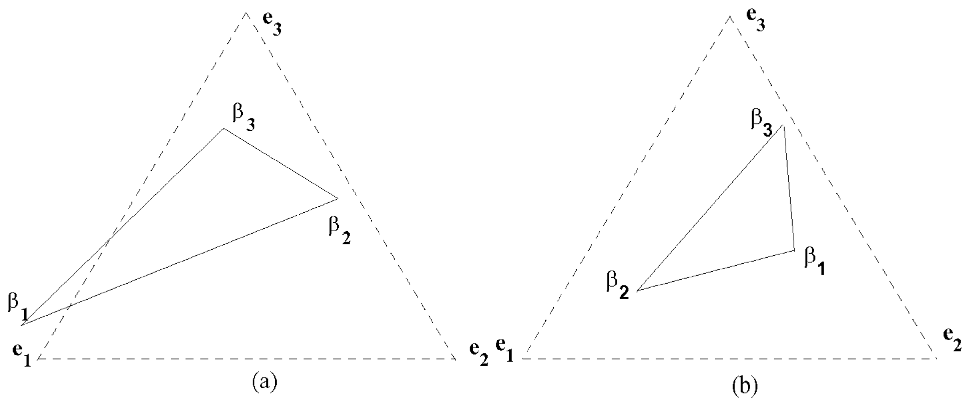

FIGURE 1 *PH*-invariant polytopes for Example 4.1.

If we choose $\boldsymbol{\alpha} = (0, 1.01, -0.01)$, then both $(\boldsymbol{\alpha}, T_1, \mathbf{e})$ and $(\boldsymbol{\alpha}, T_2, \mathbf{e})$ have an equivalent Coxian representation. That implies that $(\boldsymbol{\alpha}, T_1, \mathbf{e})$ and $(\boldsymbol{\alpha}, T_2, \mathbf{e})$ represent Coxian distributions. Therefore, the Post-T spectral polynomial algorithm can find Coxian representations for matrix-exponential representations $(\boldsymbol{\alpha}, T_1, \mathbf{e})$ and $(\boldsymbol{\alpha}, T_2, \mathbf{e})$ for which $\boldsymbol{\alpha}$ is not non-negative.

Let $\boldsymbol{\beta}_k$ be the kth row of P_1 or P_2, $1 \leq k \leq 3$. By Definition (2.3), the polytope conv$\{\boldsymbol{\beta}_k, 1 \leq k \leq 3\}$ is *PH*-invariant under $S(\boldsymbol{\lambda}_1)$ or $S(\boldsymbol{\lambda}_2)$. The polytope conv$\{\mathbf{e}_k, 1 \leq k \leq 3\}$ is *PH*-invariant under T_1 and T_2. Note that $(\mathbf{e}_k, T_i, \mathbf{e})$ and $(\boldsymbol{\beta}_k, S(\boldsymbol{\lambda}_i), \mathbf{e})$ have the same distribution, for $k = 1, 2, 3$ and $i = 1, 2$. For T_1 and T_2, we plot the two polytopes conv$\{\boldsymbol{\beta}_k, 1 \leq k \leq 3\}$ and conv$\{\mathbf{e}_k, 1 \leq k \leq 3\}$ in Figure 1. Part (a) of Figure 1 shows that a small part of the polytope conv$\{\boldsymbol{\beta}_k, 1 \leq k \leq 3\}$ (the triangle with solid lines) is outside of the polytope conv$\{\mathbf{e}_k, 1 \leq k \leq 3\}$ (the triangle with dashed lines) for T_1. Thus, for every $\boldsymbol{\alpha}$ with $\boldsymbol{\alpha}P$ in that area, $(\boldsymbol{\alpha}P, S(\boldsymbol{\lambda}_1), \mathbf{e})$ is not a Coxian representation. For T_2, from part (b) of Figure 1, the polytope conv$\{\boldsymbol{\beta}_k, 1 \leq k \leq 3\}$ is located completely inside conv$\{\mathbf{e}_k, 1 \leq k \leq 3\}$. Thus, for every probability vector $\boldsymbol{\alpha}$, $(\boldsymbol{\alpha}P, S(\boldsymbol{\lambda}_2), \mathbf{e})$ is a Coxian representation.

Example 4.1 indicates that for *PH*-generator T with only real eigenvalues, it is possible that every *PH*-representation $(\boldsymbol{\alpha}, T, \mathbf{e})$ has a Coxian representation of the same or a smaller order. This gives an interpretation to the observation in Section 5.8 in Asmussen et al.[6]: "*In most of our experimental work, we found a Coxian distribution to provide almost as good a fit as a general phase-type distribution with the same m; for one exception, see the Erlang distribution with feedback in Section* 5.3." Example 4.1 also shows that the Post-T spectral polynomial algorithm may fail to find a Coxian representation of the same or a smaller order for some $(\boldsymbol{\alpha}, T, \mathbf{e})$ (since such a representation may not exist). It is interesting to determine what kind of *PH*-representations has a bi-diagonal *PH*-representation of the same or a smaller order. It is also interesting to know, for a given T with only real eigenvalues, what kind of $\boldsymbol{\alpha}$ corresponds to a bi-diagonal *PH*-representation of the same or a smaller order. Proposition 4.1 provides some general conditions for a *PH*-representation to have a bi-diagonal *PH*-representation. For some special classes of *PH*-representations, further results can be obtained. We summarize them in the following theorems.

Theorem 4.2. *Assume that the PH-generator T is symmetric. If $-\boldsymbol{\lambda} = (-\lambda_1, -\lambda_2, \ldots, -\lambda_m)$, ordered as $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m$, is the spectrum of T, then $S(\boldsymbol{\lambda})$ PH-majorizes T, i.e., $PH(T) \subseteq PH(S(\boldsymbol{\lambda}))$.*

Proof. According to Micchelli and Willoughby[29], the spectral polynomials are non-negative matrices for any non-negative and symmetric matrix. It is also well-known that $-\lambda_1 \leq \min\{(T)_{j,j}, 1 \leq j \leq m\}$ for symmetric matrix T (Ref.[29]). Thus, $\lambda_1 I + T$ is a non-negative symmetric matrix. The eigenvalues of $\lambda_1 I + T$ are $\{\lambda_1 - \lambda_m, \lambda_1 - \lambda_{m-1}, \ldots, \lambda_1 - \lambda_2, 0\}$ ordered in nonincreasing order. Applying Theorem 3.2 from Micchelli and Willoughby[29] to $\lambda_1 I + T$, we obtain that $\{(\lambda_n I + T) \cdots (\lambda_1 I + T), 1 \leq n \leq m\}$ are non-negative matrices. Since $\mathbf{p}_1 \geq 0$, by equation (3.3), $\mathbf{p}_n \geq 0$ for $1 \leq n \leq m$. Thus, the matrix P is non-negative and has unit row sums. By Proposition 4.1 or Theorem 3 in O'Cinneide[35], $PH(T) \subseteq PH(S(\boldsymbol{\lambda}))$. This completes the proof of Theorem 4.2.

Theorem 4.3 (Cumani[16]). *Assume that the PH-generator T is triangular. If $-\boldsymbol{\lambda} = (-\lambda_1, -\lambda_2, \ldots, -\lambda_m)$, ordered as $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m$, is the spectrum of T, then $S(\boldsymbol{\lambda})$ PH-majorizes T, i.e., $PH(T) \subseteq PH(S(\boldsymbol{\lambda}))$.*

Proof. This result has been proved in Cumani[16] (also see Dehon and Latouche[17] and O'Cinneide[35]). Here we give a new proof and an explicit expression for the matrix P. Without loss of generality, we assume that T is lower triangular. All we need to show is $\mathbf{p}_n \geq 0$, for $1 \leq n \leq m$. In Appendix A, we show that the spectral polynomials $\{(\lambda_n I + T) \cdots (\lambda_1 I + T), 1 \leq n \leq m\}$ are non-negative matrices. Thus, by equation (3.3), $\mathbf{p}_n \geq 0$, for $1 \leq n \leq m$. Therefore, the matrix P is non-negative, i.e., $PH(T) \subseteq PH(S(\boldsymbol{\lambda}))$. This completes the proof of Theorem 4.3.

If T does not have a special structure, the spectral polynomial method can still lead to a complete solution to cases with $m = 2$ and $m = 3$.

Theorem 4.4. *Every PH-generator T of order 2 is PH-majorized by a Coxian generator of order 2. That is: if the two eigenvalues $\{-\lambda_1, -\lambda_2\}$ of T are ordered as $\lambda_1 \geq \lambda_2$, we have $PH(T) \subseteq PH(S(\boldsymbol{\lambda}))$.*

Proof. This result is well known (see O'Cinneide[35]). Our spectral polynomial approach provides an alternative proof. Note that, if $m = 2$, the two eigenvalues of T are real. If $\lambda_1 \geq \lambda_2$, by 4.1.1 in Proposition 4.1, the matrix P is non-negative. Therefore, we have $PH(T) \subseteq PH(S(\boldsymbol{\lambda}))$. This completes the proof of Theorem 4.4.

Theorem 4.5. *Consider a PH-generator T of order 3. If all eigenvalues $\{-\lambda_1, -\lambda_2, -\lambda_3\}$ of T are real (counting multiplicities), then T is PH-majorized by an ordered Coxian generator of order 4 or a smaller order. Furthermore, if $\min\{-\lambda_1, -\lambda_2, -\lambda_3\} \leq \min\{(T)_{1,1}, (T)_{2,2}, (T)_{3,3}\}$, T is PH-majorized by an ordered Coxian generator of order 3 or a smaller order.*

Proof. See Appendix B. Note that Appendix B not only proves the existence of an ordered Coxian generator of order 4 or a smaller order that *PH*-majorizes T, but also provides a method for computing such a Coxian generator and the corresponding matrix P.

Note 4.2. For Theorems 4.2 to 4.5, $\{\lambda_1, \lambda_2, \ldots, \lambda_m\}$ are required to be in non-increasing order to ensure P to be non-negative. If $\{\lambda_1, \lambda_2, \ldots, \lambda_m\}$ is not ordered that way, the result can be different. For example, the *PH*-generator T given in equation (4.5) is symmetric with eigenvalues $\{-11.8553, -5.6190, -2.5257\}$.

$$T = \begin{pmatrix} -10 & 3 & 0.5 \\ 3 & -7 & 1 \\ 0.5 & 1 & -3 \end{pmatrix}. \tag{4.5}$$

For $\boldsymbol{\lambda}_1 = (11.8553, 5.6190, 2.5257)$ and $\boldsymbol{\lambda}_2 = (2.5257, 5.6190, 11.8553)$, the corresponding matrices P are denoted as P_1 and P_2, respectively, and are given in equation (4.6):

$$P_1 = \begin{pmatrix} 0.5483 & 0.3274 & 0.1243 \\ 0.2531 & 0.5339 & 0.2130 \\ 0.1265 & 0.2932 & 0.5803 \end{pmatrix} \quad \text{and} \quad P_2 = \begin{pmatrix} 2.5735 & -2.7363 & 1.1627 \\ 1.1878 & 0.5339 & -0.7217 \\ 0.5939 & 0.3903 & 0.0158 \end{pmatrix}. \tag{4.6}$$

For $\boldsymbol{\lambda}_1$ with $\lambda_1 > \lambda_2 > \lambda_3$, P_1 is non-negative. Thus, we have $PH(T) \subseteq PH(S(\boldsymbol{\lambda}_1))$. However, for $\boldsymbol{\lambda}_2$, P_2 is not non-negative, which implies that $PH(T) \subseteq PH(S(\boldsymbol{\lambda}_2))$ does not hold. Proposition 5.1 explains partially why $\{\lambda_1, \lambda_2, \ldots, \lambda_m\}$ should be ordered in non-increasing order. See He and Zhang[18] for a comprehensive geometric interpretation on this issue.

5. PROPERTIES AND APPLICATIONS

Results in Section 4 indicate a strong relationship between the spectral polynomial algorithm and the ordered Coxian representation. In this section, we further explore that relationship and show some applications.

Proposition 5.1. *Assume that* $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_m)$ *are positive numbers ordered as* $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m$. *Let* $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \ldots, \gamma_N)$. *If* γ *is a subset of* $\boldsymbol{\lambda}$ *(i.e., the set of elements of* γ *is a subset of the set of elements of* $\boldsymbol{\lambda}$*), then* $S(\boldsymbol{\lambda})$ *PH-majorizes* $S(\boldsymbol{\gamma})$.

Proof. Since $\boldsymbol{\gamma}$ is a subset of $\boldsymbol{\lambda}$, we must have $m \geq N$. We introduce a new set $\boldsymbol{\eta}$ by $\eta_1 = \gamma_1, \ldots, \eta_N = \gamma_N$, and $\{\eta_{N+1}, \ldots, \eta_m\} = \{\lambda_j : \lambda_j \in \boldsymbol{\lambda}$ and $\lambda_j \notin \boldsymbol{\gamma}\}$. Then $\boldsymbol{\lambda}$ is an ordered set of $\boldsymbol{\eta}$. Since the *PH*-generator $S(\boldsymbol{\eta})$ is lower triangular, and $S(\boldsymbol{\eta})$ and $S(\boldsymbol{\lambda})$ have the same spectrum, by Theorem 4.3, there exists an $m \times m$ stochastic matrix L such that $S(\boldsymbol{\eta})L = LS(\boldsymbol{\lambda})$. We decompose L and $S(\boldsymbol{\eta})$ as follows

$$L = \begin{pmatrix} L_1 \\ L_2 \end{pmatrix} \quad \text{and} \quad S(\boldsymbol{\eta}) = \begin{pmatrix} S(\boldsymbol{\gamma}) & 0 \\ S_{21} & S_2 \end{pmatrix} \tag{5.1}$$

where L_1 is an $N \times m$ matrix and L_2 is an $(m - N) \times m$ matrix. It can be verified that $S(\boldsymbol{\gamma})L_1 = L_1 S(\boldsymbol{\lambda})$. It is readily seen that L_1 is non-negative and has unit row sums. Therefore, $S(\boldsymbol{\lambda})$ *PH*-majorizes $S(\boldsymbol{\gamma})$. This completes the proof of Proposition 5.1.

For a given *PH*-representation $(\boldsymbol{\alpha}, T, \mathbf{e})$, if it has a Coxian representation $(\boldsymbol{\beta}, S(\boldsymbol{\gamma}), \mathbf{e})$ ($\boldsymbol{\beta}$ is a probability vector), then it has an ordered Coxian representation $(\boldsymbol{\beta}', S(\boldsymbol{\lambda}), \mathbf{e})$ when $\boldsymbol{\gamma} \subseteq \boldsymbol{\lambda}$ and $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m$. In fact, the conclusion can be obtained directly since $\boldsymbol{\beta}' = \boldsymbol{\alpha}PL_1$ is non-negative if $\boldsymbol{\beta} = \boldsymbol{\alpha}P$ is non-negative, where P satisfies $TP = PS(\boldsymbol{\gamma})$ and $P\mathbf{e} = \mathbf{e}$. Furthermore, it is possible that $\boldsymbol{\beta}'$ is non-negative while $\boldsymbol{\beta}$ is not non-negative.

Assume that information about the minimal polynomial $f_T(\lambda)$ of T is available and we can find the roots $\{-\mu_1, -\mu_2, \ldots, -\mu_K\}$ of $f_T(\lambda)$. Then some results on the minimal *PH*-representation of a *PH*-distribution can be obtained. Define the *PH*-order of a *PH*-distribution as the minimal number of phases required by any *PH*-representation of that *PH*-distribution. We give two such examples in the following proposition.

Proposition 5.2. *Assume that* T *is lower triangular or symmetric and* $\mu_1 \geq \mu_2 \geq \cdots \geq \mu_K$. *Then we have* $PH(T) \subseteq PH(S(\boldsymbol{\mu}))$, *i.e.,* $S(\boldsymbol{\mu})$ *PH-majorizes* T. *Consequently, the PH-distribution* $(\boldsymbol{\alpha}, T, \mathbf{e})$ *has a PH-order no larger than* K.

Proof. According to the proofs of Theorems 4.2 and 4.3, for both cases, the spectral polynomials $(\mu_n I + T)\cdots(\mu_1 I + T)$ corresponding to the eigenvalue set $\boldsymbol{\mu}$ are non-negative. For the symmetric case, K is the number of distinct eigenvalues. From Theorems 4.2 and 4.3 and Proposition 5.1, it is readily seen that an ordered Coxian representation of order K exists for $(\boldsymbol{\alpha}, T, \mathbf{e})$. Therefore, the *PH*-order of *PH*-distribution $(\boldsymbol{\alpha}, T, \mathbf{e})$ is not larger than K. This completes the proof of Proposition 5.2.

We can also use the Post-T spectral polynomial algorithm and the ordered Coxian distributions to explore the stochastically larger order for matrix-exponential distributions (with only real eigenvalues). Denote by $F_1(t)$ and $F_2(t)$ the distribution functions of $(\boldsymbol{\alpha}_1, T, \mathbf{u})$ and $(\boldsymbol{\alpha}_2, T, \mathbf{u})$, respectively. We say that $(\boldsymbol{\alpha}_1, T, \mathbf{u})$ is *stochastically larger* than $(\boldsymbol{\alpha}_2, T, \mathbf{u})$ if $F_1(t) \le F_2(t)$ for $t \ge 0$. For two row vectors $\boldsymbol{\beta}_i = (\beta_{i,1}, \beta_{i,2}, \ldots, \beta_{i,m})$, $i = 1, 2$, if $\beta_{1,1} + \beta_{1,2} + \cdots + \beta_{1,n} \le \beta_{2,1} + \beta_{2,2} + \cdots + \beta_{2,n}$, for $1 \le n \le m$, and equality holds for $n = m$, then we say that $\boldsymbol{\beta}_2$ *majorizes* $\boldsymbol{\beta}_1$ (Marshall and Olkin[28]). Note that $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ do not have to be non-negative.

Lemma 5.1. *For matrix-exponential distributions* $(\boldsymbol{\beta}_1, S(\boldsymbol{\lambda}), \mathbf{e})$ *and* $(\boldsymbol{\beta}_2, S(\boldsymbol{\lambda}), \mathbf{e})$ *with positive* $\boldsymbol{\lambda}$ *ordered as* $\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_m > 0$, $(\boldsymbol{\beta}_1, S(\boldsymbol{\lambda}), \mathbf{e})$ *is stochastically larger than* $(\boldsymbol{\beta}_2, S(\boldsymbol{\lambda}), \mathbf{e})$ *if* $\boldsymbol{\beta}_2$ *majorizes* $\boldsymbol{\beta}_1$.

Proof. Let $F_1(t)$ and $F_2(t)$ be the distribution functions of the two distributions $(\boldsymbol{\beta}_1, S(\boldsymbol{\lambda}), \mathbf{e})$ and $(\boldsymbol{\beta}_2, S(\boldsymbol{\lambda}), \mathbf{e})$, respectively. Suppose that $\boldsymbol{\beta}_i = (\beta_{i,1}, \beta_{i,2}, \ldots, \beta_{i,m})$, $i = 1, 2$. By definition, we have

$$F_i(t) = \sum_{j=1}^{m} \beta_{i,j} G_j(t), \quad t \ge 0, \ \ i = 1, 2, \tag{5.2}$$

where $G_j(t)$ is the distribution function of $(\mathbf{e}_j, S(\boldsymbol{\lambda}), \mathbf{e}), 1 \le j \le m$. Note that $G_j(t)$ is the distribution function of the sum of j independent exponential random variables with parameters $\{\lambda_1, \lambda_2, \ldots, \lambda_j\}$. Then it is easy to see that $G_1(t) \ge G_2(t) \ge \cdots \ge G_m(t)$, $t \ge 0$. By equation (5.2), it is easy to obtain

$$F_i(t) = \sum_{j=1}^{m-1} \left(\sum_{k=1}^{j} \beta_{i,k} \right) (G_j(t) - G_{j+1}(t)) + G_m(t), \quad t \ge 0, \ \ i = 1, 2. \tag{5.3}$$

Note that $\boldsymbol{\beta}_1 \mathbf{e} = \boldsymbol{\beta}_2 \mathbf{e} = 1$. If $\boldsymbol{\beta}_2$ majorizes $\boldsymbol{\beta}_1$, by equation (5.2), we have $F_1(t) \le F_2(t)$ for $t \ge 0$. Consequently, $(\boldsymbol{\beta}_1, S(\boldsymbol{\lambda}), \mathbf{e})$ is stochastically larger than $(\boldsymbol{\beta}_2, S(\boldsymbol{\lambda}), \mathbf{e})$. This completes the proof of Lemma 5.1.

Combining with the Post-T spectral polynomial algorithm, Lemma 5.1 leads to the following stochastic comparison result for matrix-exponential distributions.

Proposition 5.3. *Consider two matrix-exponential distributions* $(\boldsymbol{\alpha}_1, T, \mathbf{u})$ *and* $(\boldsymbol{\alpha}_2, T, \mathbf{u})$. *Then* $(\boldsymbol{\alpha}_1, T, \mathbf{u})$ *is stochastically larger than* $(\boldsymbol{\alpha}_2, T, \mathbf{u})$ *if there exists positive* $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_m)$, *ordered as* $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m$, *such that* $TP = PS(\boldsymbol{\lambda})$, $P\mathbf{e} = \mathbf{u}$, *and* $\boldsymbol{\alpha}_2 P$ *majorizes* $\boldsymbol{\alpha}_1 P$.

Proof. By Proposition 3.1, the matrix exponential distribution $(\boldsymbol{\alpha}_i, T, \mathbf{u})$ has a bi-diagonal representation $(\boldsymbol{\alpha}_i P, S(\boldsymbol{\lambda}), \mathbf{e})$, $i = 1, 2$. Then we apply Lemma 5.1 to obtain the desired result. This completes the proof of Proposition 5.3.

6. EXTENSIONS OF THE POST-*T* SPECTRAL POLYNOMIAL ALGORITHM

First, we propose a dual algorithm to the Post-T spectral polynomial algorithm. As in Section 3, we consider a matrix-exponential representation $(\boldsymbol{\alpha}, T, \mathbf{u})$ of order m. For the Post-T spectral polynomial algorithm, we find $\{P, S(\boldsymbol{\lambda})\}$ from $\{T, \mathbf{u}\}$, while $\boldsymbol{\alpha}$ is a free vector variable. For the Pre-T spectral polynomial algorithm to be introduced next, we shall find $\{Q, S(\boldsymbol{\lambda})\}$ from $\{\boldsymbol{\alpha}, T\}$ and leave $\mathbf{u}$ as a free vector variable.

For $\{x_1, x_2, \ldots, x_N\}$ with nonzero elements, and a given row vector $\mathbf{q}_1$ of size m, we define

$$\mathbf{q}_n = \mathbf{q}_{n-1}(x_{n-1} I + T)/x_n, \quad 2 \leq n \leq N+1, \tag{6.1}$$

where $x_{N+1} = 1$. If $\mathbf{q}_{N+1} = 0$, it is readily seen that $\mathbf{q}_n T = -x_n \mathbf{q}_n + x_{n+1}\mathbf{q}_{n+1}$, for $1 \leq n \leq N-1$, and $\mathbf{q}_N T = -x_N \mathbf{q}_N$, which is equivalent to the following matrix equation

$$QT = (S(\mathbf{x}))' Q, \tag{6.2}$$

where Q is an $N \times m$ matrix with rows $\{\mathbf{q}_n, 1 \leq n \leq N\}$. Equation (6.2) shows that the polytope conv $\{\mathbf{q}_n, 1 \leq n \leq N\}$ is invariant under T. Similar to the Post-T case, we choose $-\mathbf{x}$ as the spectrum of $T : \mathbf{x} = \boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_m)$ and $N = m$. For this choice of $\mathbf{x}$, equation (6.1) becomes

$$\mathbf{q}_n = \frac{1}{\lambda_2 \ldots \lambda_n} \mathbf{q}_1 (\lambda_1 I + T) \cdots (\lambda_{n-1} I + T), \quad 2 \leq n \leq m+1, \tag{6.3}$$

where $\lambda_{m+1} = 1$. Then $\mathbf{q}_{m+1} = 0$ holds by the Cayley–Hamilton theorem. Let $\mathbf{q}_1 = -\boldsymbol{\alpha} T/\lambda_1$. Then we have $\boldsymbol{\alpha} = \mathbf{e}' Q$ by equation (3.4). Let $\mathbf{v} = Q\mathbf{u}$. In view of zero mass at $t = 0$, we have $\mathbf{v}'\mathbf{e} = \boldsymbol{\alpha}\mathbf{u} = 1$.

Pre-*T* Spectral Polynomial Algorithm

For a given matrix-exponential representation $(\boldsymbol{\alpha}, T, \mathbf{u})$, we assume that all eigenvalues of T are nonzero.

Step 1: Find the spectrum $\{-\lambda_1, -\lambda_2, \ldots, -\lambda_m\}$ of T.
Step 2: Compute $\mathbf{q}_1 = -\boldsymbol{\alpha} T/\lambda_1$ and $\{\mathbf{q}_n, 2 \leq n \leq m\}$ by equation (6.3). Set $N = \max\{n : \mathbf{q}_n = 0\} - 1$.
Step 3: Construct the bi-diagonal matrix $S(\boldsymbol{\lambda})$ of size N with $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_N)$. Define a matrix Q with row vectors $\{\mathbf{q}_n, 1 \leq n \leq N\}$. Compute $\mathbf{v} = Q\mathbf{u}$. Then $(\mathbf{e}', (S(\boldsymbol{\lambda}))', \mathbf{v})$ (or $(\mathbf{v}', S(\boldsymbol{\lambda}), \mathbf{e})$) of order N represents the same distribution as $(\boldsymbol{\alpha}, T, \mathbf{u})$.

In essence, the Pre-T and the Post-T spectral polynomial algorithms are the same. The reason is that $(\boldsymbol{\alpha}, T, \mathbf{u})$ and $(\mathbf{u}', T', \boldsymbol{\alpha}')$ represent the same matrix-exponential distribution. Thus, applying the Pre-T spectral polynomial algorithm on $(\boldsymbol{\alpha}, T, \mathbf{u})$ is equivalent to applying the Post-T spectral polynomial algorithm on $(\mathbf{u}', T', \boldsymbol{\alpha}')$.

We like to point out that, even though the spectral polynomial algorithms can be efficient in finding bi-diagonal representations, they may not be efficient in reducing the orders of representations. For instance, if $T\mathbf{e} = -\lambda_m \mathbf{e}$, by direct calculations, it can be shown that $(\boldsymbol{\alpha}, T, \mathbf{e})$ is an exponential distribution with parameter λ_m for any vector $\boldsymbol{\alpha}$ with a unit sum. That implies that the *PH*-order of distribution $(\boldsymbol{\alpha}, T, \mathbf{e})$ is 1. Assume that $\lambda_1 > \lambda_2 > \cdots > \lambda_m > 0$. Let

$$c_i = \frac{\lambda_m}{\lambda_i} \prod_{j=1}^{i-1} \left(1 - \frac{\lambda_m}{\lambda_j}\right), \quad 1 \leq i \leq m. \tag{6.4}$$

By equation (3.3), we have $\mathbf{p}_i = c_i \mathbf{e}$, $1 \leq i \leq m$, and $P = \mathbf{e}(c_1, \ldots, c_m) \equiv \mathbf{ec}$. Note that $\mathbf{c}$ is a probability vector. Thus, any matrix-exponential representation $(\boldsymbol{\alpha}, T, \mathbf{e})$ has an ordered Coxian representation $(\mathbf{c}, S(\boldsymbol{\lambda}), \mathbf{e})$ of order m. This example shows that the Post-T spectral polynomial algorithm is not able to reduce the order of the representation to 1.

Nevertheless, if information about Jordan chains of T is available, the spectral polynomial algorithm can be extended to find the smallest matrix-exponential representation in a bi-diagonal form. For instance, for the above example, if we know that $T\mathbf{e} = -\lambda_m \mathbf{e}$, we choose $\boldsymbol{\lambda} = (\lambda_m)$ and $N = 1$. Then we find the exponential distribution immediately by using the Post-T spectral polynomial algorithm.

In general, we consider a matrix-exponential distribution $(\boldsymbol{\alpha}, T, \mathbf{u})$. Suppose that $\boldsymbol{\alpha}$ is located in the linear subspace generated by Jordan chains $\{\boldsymbol{\alpha}_i, 1 \leq i \leq N\}$ corresponding to eigenvalues $\{-\lambda_1, -\lambda_2, \ldots, -\lambda_N\}$ (counting multiplicities), i.e., $\boldsymbol{\alpha} = \sum_{i=1}^{N} x_i \boldsymbol{\alpha}_i$. We assume that, if an eigenvalue $-\lambda$ is in $\{-\lambda_1, -\lambda_2, \ldots, -\lambda_N\}$ and its corresponding Jordan block is of order n, then $-\lambda$ repeats n times in $\{-\lambda_1, -\lambda_2, \ldots, -\lambda_N\}$. Based on that assumption, we have $\boldsymbol{\alpha}_i(\lambda_1 I + T)\cdots(\lambda_N I + T) = 0$ for $1 \leq i \leq N$ (see Lancaster and Tismenetsky[22]). Thus, we have

$$\begin{aligned}
\boldsymbol{\alpha}\mathbf{q}_{N+1} &= \frac{1}{\lambda_1\lambda_2\ldots\lambda_{N+1}}\left(\sum_{i=1}^{N} x_i\boldsymbol{\alpha}_i\right)(\lambda_1 I + T)(\lambda_2 I + T)\cdots(\lambda_N I + T)T \\
&= \frac{1}{\lambda_1\lambda_2\ldots\lambda_{N+1}}\left(\sum_{i=1}^{N} x_i\boldsymbol{\alpha}_i(\lambda_1 I + T)(\lambda_2 I + T)\cdots(\lambda_N I + T)T\right) \\
&= 0, \qquad (6.5)
\end{aligned}$$

where $\mathbf{q}_{N+1}$ is defined in equation (6.3). Then the Pre-T spectral polynomial algorithm produces an equivalent bi-diagonal representation $(\mathbf{v}', S(\boldsymbol{\lambda}), \mathbf{e})$ of order N for $(\boldsymbol{\alpha}, T, \mathbf{u})$. More systematic studies on order reduction can be found in He and Zhang[18,19].

Example 6.1. Consider $(\boldsymbol{\alpha}, T, \mathbf{e})$ with T given as

$$T = \begin{pmatrix} -15 & 0 & 1 & 1 & 2 \\ 1 & -5 & 0.5 & 1 & 0.5 \\ 2 & 0 & -4 & 0 & 1 \\ 0 & 2 & 1 & -7 & 1 \\ 1 & 0.5 & 0 & 2 & -5 \end{pmatrix} \qquad (6.6)$$

The matrix T has five distinct eigenvalues: -15.3197, -7.9712, -2.5075, $-5.1008 + 0.1706\sqrt{-1}$, and $-5.1008 - 0.1706\sqrt{-1}$. The left eigenvectors corresponding to the eigenvalues are, respectively,

$$\begin{aligned}
\boldsymbol{\alpha}_1 &= (-0.9987, -0.0242, -0.0822, -0.0799, 0.1790); \\
\boldsymbol{\alpha}_2 &= (0.1456, 0.6685, 0.1346, -1.0146, 0.0856); \\
\boldsymbol{\alpha}_3 &= (0.1670, 0.4394, 0.5266, 0.3992, 0.5936); \\
\boldsymbol{\alpha}_4 &= (0.0623 - 0.1325\sqrt{-1}, -0.1051 + 3.8057\sqrt{-1}, \\
&\quad 0.4935 - 1.7255\sqrt{-1}, -0.2587 + 0.2133\sqrt{-1}, \qquad (6.7) \\
&\quad - 0.2424 - 1.6561\sqrt{-1}); \\
\boldsymbol{\alpha}_5 &= (0.0623 + 0.1325\sqrt{-1}, -0.1051 - 3.8057\sqrt{-1}, \\
&\quad 0.4935 + 1.7255\sqrt{-1}, -0.2587 - 0.2133\sqrt{-1}, \\
&\quad - 0.2424 + 1.6561\sqrt{-1}).
\end{aligned}$$

Although the *PH*-generator T has complex eigenvalues, $(\boldsymbol{\alpha}, T, \mathbf{e})$ may have an ordered Coxian representation. For example, suppose that $\boldsymbol{\alpha} = (0.0324, 0.3082, 0.2836, 0.0594, 0.3164) = (0.1\boldsymbol{\alpha}_1 + 0.2\boldsymbol{\alpha}_2 + 0.7\boldsymbol{\alpha}_3)/r$, where

$r = (0.1\boldsymbol{\alpha}_1 + 0.2\boldsymbol{\alpha}_2 + 0.7\boldsymbol{\alpha}_3)\mathbf{e}$, i.e., $\boldsymbol{\alpha}$ is in the subspace generated by three Jordan chains $\{\boldsymbol{\alpha}_1\}$, $\{\boldsymbol{\alpha}_2\}$, and $\{\boldsymbol{\alpha}_3\}$ corresponding to eigenvalues $\{-15.3197, -7.9712, -2.5075\}$. Then we choose $\boldsymbol{\lambda} = (15.3197, 7.9712, 2.5075)$ and apply the Pre-T spectral polynomial algorithm. It can be found that $(\boldsymbol{\alpha}, T, \mathbf{e})$ has an ordered Coxian representation $(\mathbf{v}', S(\boldsymbol{\lambda}), \mathbf{e})$, where

$$\mathbf{v}' = (0.1246, 0.2763, 0.5991),$$
$$S(\boldsymbol{\lambda}) = \begin{pmatrix} -15.3197 & 0 & 0 \\ 7.9712 & -7.9712 & 0 \\ 0 & 2.5075 & -2.5075 \end{pmatrix}. \tag{6.8}$$

Note 6.1. If the vector $\mathbf{e}$ is located in a subspace generated by some Jordan chains of T, it can be shown in a similar way that a bi-diagonal representation of a smaller order can be found for $(\boldsymbol{\alpha}, T, \mathbf{e})$.

7. CONCLUDING REMARKS

The results obtained in this paper hold for discrete time *PH*-distributions, which are defined as the distributions of the absorption times of discrete time Markov chains with a finite number of states. The main reason is the following relationship between discrete time *PH*-distributions and continuous time *PH*-distributions. For a continuous time *PH*-distribution $(\boldsymbol{\alpha}, T, \mathbf{e})$, define $S = I + T/v$ for $v > \max\{-(T)_{j,j}\}$. The matrix S is a substochastic matrix and

$$\boldsymbol{\alpha} \exp\{Tt\}\mathbf{e} = \boldsymbol{\alpha} e^{-vt} \exp\{Svt\}\mathbf{e} = e^{-vt} \sum_{n=0}^{\infty} \frac{(vt)^n}{n!} \boldsymbol{\alpha} S^n \mathbf{e}. \tag{7.1}$$

Equation (7.1) establishes a relationship between $(\boldsymbol{\alpha}, T, \mathbf{e})$ and the discrete time *PH*-distribution $(\boldsymbol{\alpha}, S, \mathbf{e})$. For a given v (large enough), the distributions of $(\boldsymbol{\alpha}, T, \mathbf{e})$ and $(\boldsymbol{\alpha}, S, \mathbf{e})$ determine each other uniquely. Thus, $(\boldsymbol{\alpha}, T, \mathbf{e})$ has an equivalent *PH*-representation $(\boldsymbol{\beta}, T_1, \mathbf{e})$ if and only if $(\boldsymbol{\alpha}, S, \mathbf{e})$ has an equivalent *PH*-representation $(\boldsymbol{\beta}, S_1, \mathbf{e})$, where T_1 and S_1 have the relationship $S_1 = I + T_1/v$. By equation (7.1), we can translate a *PH*-representation problem of discrete time *PH*-distributions into a continuous time one. Therefore, all results obtained in this paper apply to the discrete time case.

This paper does not consider the minimal *PH*-representation problem directly, although the results in this paper can be used in the study of that problem. The results obtained in this paper indicate that finding the minimal *PH*-representation has much to do with *PH*-invariant polytopes with the minimum number of extreme points. Therefore, how to construct *PH*-invariant polytopes with a smaller number of extreme points can be a key to solving the problem. Research in this direction is ongoing (He and Zhang[18,19]). In Ref.[18], we construct *PH*-invariant polytopes associated with bi-diagonal representations of *PH*-distributions. Geometric and probabilistic interpretations to the spectral polynomial algorithms are offered. In Ref.[19], we develop an algorithm for computing a minimal ordered Coxian representation for *PH*-distributions whose Laplace-Stieltjes transforms have only real poles.

APPENDIX A. NONNEGATIVITY OF SPECTRAL POLYNOMIALS OF TRIANGULAR MATRICES

This appendix gives a proof to the non-negativity of the spectral polynomials of lower triangular matrices with non-negative off-diagonal elements. Assume that T is a lower triangular matrix with non-negative off-diagonal elements. Suppose that $\{-\lambda_1, -\lambda_2, \ldots, -\lambda_m\}$ are the diagonal elements of T and are ordered as $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m$. We show that the spectral polynomials $\{(\lambda_n I + T)\cdots(\lambda_1 I + T), 1 \leq n \leq m\}$ are non-negative matrices. Note that $T = (t_{i,j})$.

Let $T^{(n)} = \big(t_{i,j}^{(n)}\big) = \big(\lambda_n I + T\big)\cdots\big(\lambda_1 I + T\big)$, $1 \leq n \leq m$. Since T is lower triangular, $t_{i,j}^{(n)} = 0$ if $j > i$. It is easy to see that

$$t_{i,i}^{(n)} = (\lambda_1 + t_{i,i})\cdots(\lambda_n + t_{i,i}), \quad 1 \leq i \leq m. \tag{A.1}$$

Thus, $t_{i,i}^{(n)} \geq 0$, since either $-t_{i,i} \in \{\lambda_1, \lambda_2, \ldots, \lambda_n\}$ or $-t_{i,i} < \lambda_n$, $1 \leq i \leq m$. For $j < i$, we have

$$\begin{aligned} t_{i,j}^{(n)} &= t_{i,j}\sum_{d=1}^{n}\left(\prod_{s=1}^{d-1}\big(\lambda_s + t_{i,i}\big)\right)\left(\prod_{s=d+1}^{n}\big(\lambda_s + t_{j,j}\big)\right) \\ &\quad + \sum_{k=1}^{j-i-1}\ \sum_{i<i_1<\cdots<i_k<j} t_{i,i_1}t_{i_1,i_2}\ldots t_{i_k,j} \sum_{1\leq d_1<d_2<\cdots<d_k\leq n}\left(\prod_{s=1}^{d_1-1}(\lambda_s + t_{i,i})\right) \\ &\quad \times\left(\prod_{s=d_1+1}^{d_2-1}(\lambda_s + t_{i_1,i_1})\right)\cdots\left(\prod_{s=d_{k-1}+1}^{d_k-1}\big(\lambda_s + t_{i_k,i_k}\big)\right)\left(\prod_{s=d_k+1}^{n}\big(\lambda_s + t_{j,j}\big)\right). \end{aligned} \tag{A.2}$$

Intuitively, the above expression shows that $t_{i,j}^{(n)}$ can be interpreted as the transition "probability" from state i to state j in n steps with transition matrices $\{\lambda_n I + T, \ldots, \lambda_1 I + T\}$. To prove the non-negativity of $t_{i,j}^{(n)}$, we introduce the function $g(x) = (\lambda_n + x)\cdots(\lambda_1 + x)$. Define the *divided differences* of $g(x)$ as follows (Horn and Johnson[20]), $g[x] = g(x)$, and

$$g[x_1, x_2] = \frac{g[x_1] - g[x_2]}{x_1 - x_2};$$
$$g[x_1, x_2, \ldots, x_n, x_{n+1}] = \frac{g[x_1, x_2, \ldots, x_{n-1}, x_{n+1}] - g[x_1, x_2, \ldots, x_{n-1}, x_n]}{x_{n+1} - x_n}. \tag{A.3}$$

It can be shown that $g[x_1, \ldots, x_k]$ is permutation invariant. In fact, it can be verified that, for $2 \le k \le n$,

$$g[x_1, \ldots, x_k] = \sum_{1 \le d_1 < d_2 < \cdots < d_{k-1} \le n} \left(\prod_{s=1}^{d_1 - 1} (\lambda_s + x_1) \right) \cdots \left(\prod_{s=d_{k-1}+1}^{n} (\lambda_s + x_n) \right). \tag{A.4}$$

By using the divided differences of $g(x)$, equation (A.2) can be rewritten as

$$t_{i,j}^{(n)} = t_{i,j} g\big[t_{i,i}, t_{j,j}\big] + \sum_{k=1}^{j-i-1} \sum_{i<i_1<\cdots<i_k<j} t_{i,i_1} t_{i_1,i_2} \ldots t_{i_k,j} g\big[t_{i,i}, t_{i_1,i_1}, \ldots, t_{i_k,i_k}, t_{j,j}\big]. \tag{A.5}$$

We can arrange $\{t_{i,i}, t_{i_1,i_1}, \ldots, t_{i_k,i_k}, t_{j,j}\}$ in increasing order and the divided differences remain unchanged. Next, we show that $g[t_{i,i}, t_{i_1,i_1}, \ldots, t_{i_k,i_k}, t_{j,j}] \ge 0$. To simply the notation, we use $g[x_1, \ldots, x_N]$, where $\{x_1, \ldots, x_N\} \subseteq \{-\lambda_1, \ldots, -\lambda_m\}$. If $\min\{x_1, \ldots, x_N\} \ge \max\{-\lambda_1, \ldots, -\lambda_n\}$, equation (A.4) implies that $g[x_1, \ldots, x_N] \ge 0$. Otherwise, some x_i is in $\{-\lambda_1, \ldots, -\lambda_n\}$. Assume that $x_1 \le x_2 \le \cdots \le x_N$, $x_K \in \{-\lambda_1, \ldots, -\lambda_n\}$, and $x_{K+1} \notin \{-\lambda_1, \ldots, -\lambda_n\}$. We rearrange the function $g(x)$ in the following way:

$$g(x) = \left(\prod_{i \le n:\, \lambda_i \in \{x_1, \ldots, x_N\}} (x + \lambda_i) \right) \left(\prod_{i \le n:\, \lambda_i \notin \{x_1, \ldots, x_N\}} (x + \lambda_i) \right) \equiv g_1(x)\, g_2(x). \tag{A.6}$$

By Steffenson's product rule (Ref.[29]), we have

$$g[x_1, x_2, \ldots, x_N] = \sum_{s=1}^{N} g_1[x_1, x_2, \ldots, x_s]\, g_2[x_s, \ldots, x_N]. \tag{A.7}$$

If $s \le K$, we have $g_1[x_1, \ldots, x_s] = 0$ since $g_1(x_i) = 0$, $1 \le i \le s$. If $s > K$, $g_1[x_1, \ldots, x_s] = 1$ and $g_2[x_s, \ldots, x_N] \ge 0$ since $x_i > \max\{-\lambda_1, \ldots, -\lambda_n\}$, $K + 1 \le i \le N$. Equation (A.7) shows that $g[x_1, \ldots, x_s]$ is non-negative.

Thus, every $g[t_{i,i}, t_{i_1,i_1}, \ldots, t_{i_k,i_k}, t_{j,j}]$ is *non-negative.* Equation (A.5) implies that the spectral polynomials are all non-negative. This completes the proof of Appendix A.

APPENDIX B. A PROOF OF THEOREM 4.5

First, we note that results about M-matrix and non-negative matrix can be found in Minc[30] and Seneta[43]. Suppose that $\lambda_1 \geq \lambda_2 \geq \lambda_3$. By 4.1.1 of Proposition 4.1, $\mathbf{p}_1$ and $\mathbf{p}_3$ are non-negative. Thus, to prove the second result, we only need to show that $\mathbf{p}_2$ is non-negative. By equation (3.3), we have $\mathbf{p}_2 = (\lambda_1 I + T)(-T\mathbf{e})/(\lambda_1\lambda_2)$. If $-\lambda_1 = \min\{-\lambda_1, -\lambda_2, -\lambda_3\} \leq \min\{t_{1,1}, t_{2,2}, t_{3,3}\}$, $\lambda_1 I + T$ is non-negative. Note that $T = (t_{i,j})$. Therefore, $\mathbf{p}_2$ is non-negative, which implies that P is non-negative. Thus, under that condition, $(\boldsymbol{\alpha}, T, \mathbf{e})$ has an ordered Coxian representation of order 3 or a smaller order. This proves the second part of Theorem 4.5.

To show the general result, we order $\{t_{1,1}, t_{2,2}, t_{3,3}\}$ in increasing order as $t_{(1,1)} \leq t_{(2,2)} \leq t_{(3,3)}$. We first show that $-\lambda_1 \leq t_{(2,2)}$. It is well known that $-(\lambda_1 + \lambda_2 + \lambda_3) = t_{1,1} + t_{2,2} + t_{3,3} = t_{(1,1)} + t_{(2,2)} + t_{(3,3)}$. It is also known from Perron-Frobenius theory that $t_{(3,3)} \leq -\lambda_3$. If $-\lambda_1 > t_{(2,2)}$, then $-(\lambda_1 + \lambda_2 + \lambda_3) > 2t_{(2,2)} + t_{(3,3)} \geq t_{(1,1)} + t_{(2,2)} + t_{(3,3)} = t_{1,1} + t_{2,2} + t_{3,3}$, which is a contradiction. An immediate implication of $-\lambda_1 \leq t_{(2,2)}$ is that at most one element of $\mathbf{p}_2$ can be negative, since $\mathbf{p}_2 = (\lambda_1 I + T)(-T\mathbf{e})/(\lambda_1\lambda_2)$. To prove the general result, we need to consider three cases.

First, if $\mathbf{p}_2$ is non-negative, the conclusion is obvious.

Second, if $\mathbf{p}_2$ is not non-negative and $\{\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3\}$ are dependent, then $\{T\mathbf{e}, T^2\mathbf{e}, T^3\mathbf{e}\}$ are dependent, which implies that $\{\mathbf{e}, T\mathbf{e}, T^2\mathbf{e}\}$ are dependent (since $m = 3$). That further implies that $(\mu_1 I + T)(\mu_2 I + T)\mathbf{e} = 0$ for some μ_1 and μ_2. Suppose that $\mathbf{u}$ is a left eigenvector corresponding to $-\lambda_3$. Since the eigenvector corresponding to $-\lambda_3$ can be chosen to be non-negative (and must be nonzero), we choose $\mathbf{u}$ to be non-negative and normalize $\mathbf{u}$ by $\mathbf{ue} = 1$. Pre-multiplying $\mathbf{u}$ on both sides of $(\mu_1 I + T)(\mu_2 I + T)\mathbf{e} = 0$ yields $(\mu_1 - \lambda_3)(\mu_2 - \lambda_3) = 0$. Thus, we must have (say) $\mu_2 = \lambda_3$. Then it is easy to see that either $T\mathbf{e} = -\lambda_3\mathbf{e}$ or $\mu_1 \in \{\lambda_1, \lambda_2\}$. For the first case, $(\boldsymbol{\alpha}, T, \mathbf{e})$ is an exponential distribution for any probability vector $\boldsymbol{\alpha}$. For the second case, we apply the Post-T spectral polynomial algorithm with $\{\mu_1, \mu_2\}$. Then the new vector $\mathbf{p}_2$ is non-negative since it is an eigenvector corresponding to $-\lambda_3$ (see the proof of Proposition 4.1.1). Thus, the new matrix P is non-negative for this case as well. Therefore, if $\{\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3\}$ are dependent, P is non-negative and an ordered Coxian representation of order 1 or 2 can be found for $(\boldsymbol{\alpha}, T, \mathbf{e})$ for any probability vector $\boldsymbol{\alpha}$.

Lastly, if $\mathbf{p}_2$ is not non-negative and $\{\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3\}$ are independent, let $P = (\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3) = (p_{i,j})$. Without loss of generality, suppose that the first element of $\mathbf{p}_2$ is negative. Then it is clear that the *PH*-representation $(\mathbf{e}_1, T, \mathbf{e})$ has a representation $((p_{1,1}, p_{1,2}, p_{1,3}),\ S(\boldsymbol{\lambda}), \mathbf{e})$, which is not a *PH*-representation since $p_{1,2}$ is negative. Let $Q = P^{-1}$. Denote by $\{\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3\}$ the first row, second row, and third row of Q, respectively. Then we have $QT = S(\boldsymbol{\lambda})Q$, i.e., $\text{conv}\{\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3\}$ is a *PH*-invariant polytope under T. That *PH*-invariant polytope covers probability vector $\mathbf{e}_2$ and $\mathbf{e}_3$, but not $\mathbf{e}_1$. In order to construct a *PH*-invariant polytope that covers all probability vectors, we consider the polytope $\text{conv}\{\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3, \mathbf{e}_1\}$. More specifically, by using $PQ = I$, we have

$$\begin{aligned}\mathbf{e}_1 T &= t_{1,1}\mathbf{e}_1 + t_{1,2}\mathbf{e}_2 + t_{1,3}\mathbf{e}_3 \\ &= t_{1,1}\mathbf{e}_1 + \sum_{i=1}^{3} \big(t_{1,2}p_{2,i} + t_{1,3}p_{3,i}\big)\mathbf{q}_i \equiv t_{1,1}\mathbf{e}_1 + \sum_{i=1}^{3} w_i\mathbf{q}_i. \end{aligned} \tag{B.1}$$

By the definition of $\{w_1, w_2, w_3\}$, we have

$$\begin{aligned} w_1 + w_2 + w_3 + t_{1,1} &= \sum_{i=1}^{3} \big(t_{1,2}p_{2,i} + t_{1,3}p_{3,i}\big) + t_{1,1} \\ &= t_{1,2}\sum_{i=1}^{3} p_{2,i} + t_{1,3}\sum_{i=1}^{3} p_{3,i} + t_{1,1} \\ &= t_{1,2} + t_{1,3} + t_{1,1} \le 0. \end{aligned} \tag{B.2}$$

Let

$$Q_1 = \begin{pmatrix} Q \\ \mathbf{e}_1 \end{pmatrix} \quad \text{and} \quad H = \begin{pmatrix} S(\boldsymbol{\lambda}) & 0 \\ (w_1, w_2, w_3) & t_{1,1} \end{pmatrix} \tag{B.3}$$

By equations (B.1) and (B.2), it is easy to verify $Q_1 T = HQ_1$, $Q_1\mathbf{e} = \mathbf{e}$, and H is a *PH*-generator. By the definition of Q_1, it is easy to see that the polytope $\text{conv}\{\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3, \mathbf{e}_1\}$ is *PH*-invariant under T and it covers all probability vectors $\boldsymbol{\alpha}$ of order 3. Since H is a lower triangle *PH*-generator, by Theorem 4.3, there exists a stochastic matrix P_1 and an ordered Coxian generator S_1 with spectrum $\{-\lambda_1, -\lambda_2, -\lambda_3, t_{1,1}\}$ such that $HP_1 = P_1 S_1$.

For any probability vector $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \alpha_3)$, we have

$$\begin{aligned}\boldsymbol{\alpha} &= \alpha_1\mathbf{e}_1 + \alpha_2\mathbf{e}_2 + \alpha_3\mathbf{e}_3 = \alpha_1\mathbf{e}_1 + \alpha_2\sum_{i=1}^{3} p_{2,i}\mathbf{q}_i + \alpha_3\sum_{i=1}^{3} p_{3,i}\mathbf{q}_i \\ &= \alpha_1\mathbf{e}_1 + \sum_{i=1}^{3}(\alpha_2 p_{2,i} + \alpha_3 p_{3,i})\mathbf{q}_i \equiv \sum_{i=1}^{3} x_i\mathbf{q}_i + x_4\mathbf{e}_1. \end{aligned} \tag{B.4}$$

It is easy to verify that $\{x_1, x_2, x_3, x_4\}$ are non-negative and $x_1 + x_2 + x_3 + x_4 = 1$. Thus, there exists a probability vector $\mathbf{x} = (x_1, x_2, x_3, x_4)$ such that $\boldsymbol{\alpha} = \mathbf{x}Q_1$. Then $(\boldsymbol{\alpha}, T, \mathbf{e})$ and $(\mathbf{x}, H, \mathbf{e})$ represent the same probability distribution. Let $\boldsymbol{\beta} = \mathbf{x}P_1$, which is a probability vector. Then $(\boldsymbol{\beta}, S_1, \mathbf{e})$ is an ordered Coxian representation of order 4, which represents the same probability distribution as $(\mathbf{x}, H, \mathbf{e})$ and $(\boldsymbol{\alpha}, T, \mathbf{e})$. This completes the proof of Theorem 4.5.

ACKNOWLEDGMENTS

The authors would like to thank an anonymous referee and Dr. O'Cinneide for valuable comments and suggestions on the paper. The authors would also like to thank Dr. Marvin Silver for proofreading the paper. This research project was supported in part by a NSERC research grant and a research grant from the Chinese Academy of Sciences.

REFERENCES

1. Aldous, D.; Shepp, L. The least variable phase type distribution is Erlang. Stochastic Models. *3* (3), 467–473.
2. Alfa, A.S.; Chakravarthy, S.R., Eds. *Advances in Matrix Analytic Methods for Stochastic Models*; Notable Publications: New Jersey, 1998.
3. Asmussen, S. *Applied Probability and Queues*; Springer: New York, 2003.
4. Asmussen, S. Exponential families generated by phase-type distributions and other Markov lifetimes. Scand. J. Statist. **1989**, *16*, 319–334.
5. Asmussen, S.; Bladt, M. Renewal theory and queueing algorithms for matrix-exponential distributions. In *Proceedings of the First International Conference on Matrix Analytic Methods in Stochastic Models*; Alfa, A.S.; Chakravarthy, S., Eds.; Marcel Dekker: New York, 1996.
6. Asmussen, S.; Nerman, O.; Olsson, M. Fitting phase-type distributions via the EM algorithm. Scand. J. Statist. **1996**, *23*, 419–441.
7. Blackwell, D.; Koopmans, L. On the identifiability problem for functions of finite Markov chains. Ann. Math. Statist. **1957**, *28*, 1011–1015.
8. Botta, R.F.; Harris, C.M.; Marchal, W.G. Characterizations of generalized hyperexponential distribution functions. Stochastic Models **1987**, *3* (1), 115–148.
9. Chakravarthy, S.R.; Alfa, A.S. *Matrix-Analytic Methods in Stochastic Models*; Marcel Dekker: New York, 1997.
10. Commault, C. Linear positive systems and phase-type representations. Positive Systems, Proceedings **2003**, *294*, 281–288.
11. Commault, C.; Chemla, J.P. On dual and minimal phase-type representations. Stochastic Models **1993**, *9*, 421–434.
12. Commault, C.; Chemla, J.P. An invariant of representations of phase-type distributions and some applications. Journal of Applied Probability **1996**, *33*, 368–381.
13. Commault, C.; Mocanu, S. A generic property of phase-type representations. Journal of Applied Probability **2002**, *39*, 775–785.
14. Commault, C.; Mocanu, S. Phase-type distributions and representations: some results and open problems for system theory. International Journal of Control **2003**, *76*, 566–580.
15. Cox, D.R. On the use of complex probabilities in the theory of stochastic processes. Proc. Camb. Phil. Soc. **1955**, *51*, 313–319.
16. Cumani, A. On the canonical representation of Markov processes modeling failure time distributions. Microelectronics and Reliability **1982**, *22* (3), 583–602.
17. Dehon, M.; Latouche, G. A geometric interpretation of the relations between the exponential and the generalized Erlang distributions. Adv. Appl. Probab. **1982**, *14*, 885–897.

18. He, Qi-Ming; Zhang, H. *PH*-invariant polytopes and coxian representations of *PH*-representations (in press).
19. He, Qi-Ming; Zhang, H. An algorithm for computing minimal coxian representations (submitted for publication).
20. Horn, R.A.; Johnson, C.R. *Matrix Analysis*; Cambridge University Press: Cambridge, 1991.
21. Ito, H.; Amari, S.I.; Kobayashi, K. Identifiability of hidden Markov information sources and their minimum degrees of freedom. IEEE Trans. Inf. Theory **1992**, *38*, 324–333.
22. Lancaster, P.; Tismenetsky, M. *The Theory of Matrices*; Academic Press: New York, 1985.
23. Latouche, G.; Ramaswami, V. *Introduction to Matrix Analytic Methods in Stochastic Modelling*; ASA & SIAM: Philadelphia, USA.
24. Latouche, G.; Taylor, P., Eds. *Advances in Algorithmic Methods for Stochastic Models*; Notable Publications: New Jersey.
25. Latouche, G.; Taylor, P., Eds. *Matrix-Analytic Methods: Theory and Applications*; World Scientific: New Jersey.
26. Maier, R.S. The algebraic construction of phase-type distributions. Stochastic Models **1991**, *7* (4), 573–602.
27. Maier, R.S.; O'Cinneide, C.A. A closure characterization of phase-type distributions. Journal of Applied Probability **1992**, *29*, 92–103.
28. Marshall, A.W.; Olkin, I. *Inequalities: Theory of Majorization and its Applications*; Academic Press: New York, 1979.
29. Micchelli, C.A.; Willoughby, R.A. On functions which preserve the class of Stieltjes matrices. Linear Algebra and its Applications **1979**, *23*, 141–156.
30. Minc, H. *Non-negative Matrix*; John Wiley & Sons: New York, 1988.
31. Mocanu, S.; Commault, C. Sparse representations of phase type distributions. Stochastic Models **1999**, *15*, 759–778.
32. Neuts, M.F. Probability distributions of phase type. In: *Liber Amicorum Prof. Emeritus H. Florin*; University of Louvain: Louvain, 1975; 173–206.
33. Neuts, M.F. *Matrix-Geometric Solutions in Stochastic Models – An Algorithmic Approach*; The Johns Hopkins University Press: Baltimore, 1981.
34. Neuts, M.F. Two further closure-properties of *PH*-distributions. Asia-Pacific Journal of Operational Research **1992**, *9*, 77–85.
35. O'Cinneide, C.A. On non-uniqueness of representations of phase-type distributions. Stochastic Models **1989**, *5* (2), 247–259.
36. O'Cinneide, C.A. Characterization of phase-type distributions. Stochastic Models **1990**, *6* (1), 1–57.
37. O'Cinneide, C.A. Phase-type distributions and invariant polytope. Adv. Appl. Probab. **1991**, *23*, 515–535.
38. O'Cinneide, C.A. Phase-type distributions and majorization. Annals of Applied Probability **1991**, *1* (2), 219–227.
39. O'Cinneide, C.A. Triangular order of triangular phase-type distributions. Stochastic Models **1993**, *9* (4), 507–529.
40. O'Cinneide, C.A. Phase-type distributions: open problems and a few properties. Stochastic Models **1999** *15* (4), 731–757.
41. Rockafellar, R.T. *Convex Analysis*; Princeton University Press: New Jersey.
42. Ryden, T. On identifiability and order of continuous-time aggregated markov chains, Markov-modulated Poisson processes, and phase-type distributions. Journal of Applied Probability **1996**, *33*, 640–653.
43. Seneta, E. *Non-Negative Matrices: An Introduction to Theory and Applications*; John Wiley & Sons: New York, 1973.
44. Telek, M. The minimal coefficient of variation of discrete phase type distributions. In *Proceedings of the First International Conference on Matrix Analytic Methods in Stochastic Models*; Latouche, G.; Taylor, P.G., Eds.; Notable Publications, Inc.: New Jersey, 2000.
45. Yao, R. A. Proof of the steepest increase conjecture of a phase-type density. Stochastic Models **2002**, *18* (1), 1–6.