

Coxian Approximations of Matrix-Exponential Distributions

Qi-Ming He* and Hanqin Zhang[†]

April 17, 2007

Abstract

In this paper, we study the approximation of matrix-exponential distributions by Coxian distributions. Based on the spectral polynomial algorithm, we develop an algorithm for computing Coxian representations of Coxian distributions that are approximations of matrix-exponential distributions. As a specialization, we show that phase-type (PH) distributions can be approximated by Coxian distributions. We also show that any phase-type generator with only real eigenvalues is PH -majorized by some ordered Coxian generators. Consequently, the algorithm is modified for computing ordered Coxian representations of any phase-type distribution whose Laplace-Stieltjes transform has only real poles. Numerical examples are presented to show the efficiency of the algorithm and the accuracy of the Coxian approximations.

Key words. Matrix-exponential distribution, Coxian distribution, phase-type distribution, matrix analytic methods, Perron-Frobenius theory.

Mathematics subject classification. Primary 60A99, Secondary 15A18.

*Department of Industrial Engineering, Dalhousie University, Halifax, N.S., Canada B3J 2X4.
<Qi-Ming.HE@DAL.CA>

[†]Institute of Applied Mathematics, Academy of Mathematics and Systems Science, Academia Sinica, Beijing, 100080, China. <hanqin@amt.ac.cn>

1 Introduction

This paper focuses on the approximation of matrix-exponential (ME) distributions by Coxian distributions. An algorithm is developed for computing Coxian representations of Coxian approximations of ME -distributions. As a specialization, the problem of Coxian approximations of phase-type (PH) distributions is resolved. Moreover, the algorithm developed in this paper is modified for computing ordered Coxian representations for PH -distributions whose Laplace-Stieltjes transform has only real poles.

Neuts [22] introduced the PH -distribution as the distribution of the absorption time of a finite-state Markov process. Since the class of PH -distributions is dense in the class of all probability distributions with nonnegative support, the introduction of PH -distributions made it possible to study complicated queueing models such as the $PH/PH/c$ queue analytically and numerically (Takahashi [33]). ME -distributions are generalizations of PH -distributions and have been used in the study of queueing models (Lipsky [19]). Asmussen and Bladt [4] studied ME -distributions and their corresponding renewal processes and queueing models. Today, both PH -distributions and ME -distributions have been widely used in the study of queueing networks, reliability models, supply chain models, insurance and risk models, and telecommunications systems (Asmussen [2, 3], Latouche and Ramaswami [18], Neuts [23, 24], and references therein).

It is well known that the representation of a PH -distribution is not unique. To reduce the time complexity of algorithms involving PH -distributions, it is useful to find representations with the minimal number of phases. This is known as the minimal PH -representation problem (Commault and Mocanu [7], Mocanu and Commault [21], Neuts [22, 25], O’Cinneide [26, 27, 28, 29, 30]). Commault and Mocanu [7] and O’Cinneide [30] reviewed the literature on PH -distributions in recent years. While the problem of finding the minimal PH -representation of a PH -distribution is still open, the problem of finding simpler representations for PH -distributions or other probability distributions has been investigated extensively in recent years ([4, 7, 10, 11, 21, 26, 27, 28, 29]) with the focus on finding PH -representations with a simple structure for PH -distributions that approximate probability distributions (Altioik [1], Asmussen [2], and Bobbio, Horváth, Scarpa, and Telek [6]). The Coxian representation is one of the simpler PH -representations that have been investigated (Cumani [10], Dehon and Latouche [11], and O’Cinneide [26, 28, 29]).

Erlang [12] introduced the idea of phase into the study of telephone systems, which led to the introduction of Erlang distributions in probability and statistics. Cox [8, 9] generalized the class of Erlang distributions and systematically studied probability distributions as mixtures of Erlang distributions. Consequently, Cox [8, 9] gave the first definition of Coxian distributions that have been used in many branches of science and engineering. In O’Cinneide [28], it was shown that Coxian distributions with a positive density function on

positive real numbers have an ordered Coxian representation, which is a special bi-diagonal PH -representation. Coxian representations have many advantages in numerical computations. For instance, the eigenvalues of their Coxian generators are the diagonal elements and can be used directly without further numerical computation. Therefore, it is computationally attractive to replace ME -representations or PH -representations with Coxian representations in the study of stochastic models. For that purpose, we need to find Coxian representations of Coxian distributions that approximate ME -distributions or PH -distributions. The objective of this paper is to introduce an algorithm for computing Coxian representations of Coxian distributions as approximations of ME -distributions and PH -distributions.

Recently, HE and Zhang [15] (also see HE and Zhang [14, 16]) developed a spectral polynomial approach for computing bi-diagonal representations of PH -distributions. The idea of the spectral polynomial approach is to treat a PH -generator as a linear mapping and to generate some invariant polytopes of that linear mapping in order to calculate new ME -representations. The theory developed in [15] provided the basis for developing an algorithm for computing Coxian representations and Coxian approximations. In HE and Zhang [16], based on the spectral polynomial approach, an algorithm for computing minimal Coxian representations for Coxian distributions was developed. The algorithm in HE and Zhang [16] dealt with individual Coxian distributions, while the algorithm in this paper focuses on PH -generators. Thus, the algorithm developed in [16] complements the algorithm of this paper. In van de Liefvoort and Heindl [34], computational methods were developed for evaluating ME -distributions with a given representation. That paper considered the numerical evaluation of ME -distributions, while this paper finds Coxian representations of Coxian approximations of ME -distributions.

In this paper, we first concentrate on the class of ME -distributions and develop an algorithm that can be used to find some special Coxian distributions as approximations of ME -distributions. By taking into account the structure of PH -generators, the algorithm is modified to find ordered Coxian representations of Coxian distributions that approximate PH -distributions. We show that any PH -generator with only real eigenvalues is PH -majorized by some special ordered Coxian generator, which is consistent with Theorem 4.1 in O’Cinneide [28] and Theorem 5.2 in O’Cinneide [29]. In fact, the construction process of the special ordered Coxian generator bears some similarities to the proof of Theorem 4.1 in O’Cinneide [28]. The algorithm is then modified for computing such ordered Coxian generators from the original PH -generator.

The remainder of this paper is organized as follows. Section 2 gives preliminary results for the development of the theory and the algorithm in this paper. In Section 3, we show that, under some mild conditions, every matrix-exponential distribution can be approximated by Coxian distributions. An algorithm is developed for computing Coxian representations of

such approximations. In Section 4, we show how PH -distributions can be approximated by Coxian distributions. The algorithm developed in Section 3 is modified for finding Coxian approximations for the class of PH -distributions. Section 5 modifies the algorithm in Section 3 for computing Coxian representations of PH -distributions whose PH -generator has only real eigenvalues. Numerical examples are presented in Section 6 to demonstrate the efficiency of the algorithm and the accuracy of Coxian approximations. Section 7 concludes this paper.

2 PH , ME , and Coxian Distributions

A square matrix T with negative diagonal elements, nonnegative off-diagonal elements, and non-positive row sums with at least one negative row sum, is called a *sub-generator* in the general literature of Markov processes. We shall call a sub-generator T of finite size a PH -generator. Define an infinitesimal generator for a continuous-time Markov chain with $m + 1$ states

$$\begin{pmatrix} T & -T\mathbf{e} \\ 0 & 0 \end{pmatrix} \quad (2.1)$$

where the state $m + 1$ is an absorption state and \mathbf{e} is the column vector with all elements being one. The matrix T is an $m \times m$ PH -generator. We assume that states $\{1, 2, \dots, m\}$ are transient, which is equivalent to assuming that T is invertible. Let $\boldsymbol{\alpha}$ be a nonnegative 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 2-tuple $(\boldsymbol{\alpha}, T)$ a PH -representation of the PH -distribution. The number m is the order of the PH -representation $(\boldsymbol{\alpha}, T)$. We refer to Chapter 2 in Neuts [23] for basic properties of PH -distributions. The probability distribution function of the PH -distribution is given as $1 - \boldsymbol{\alpha} \exp\{Tt\}\mathbf{e}$ for $t \geq 0$, and the density function is given as $1 - \boldsymbol{\alpha} \exp\{Tt\}T\mathbf{e}$ for $t > 0$. 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 $\boldsymbol{\alpha}\mathbf{e} \neq 0$, 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, the study of the representations of $(\boldsymbol{\alpha}, T)$ is equivalent to that of $(\boldsymbol{\alpha}/(\boldsymbol{\alpha}\mathbf{e}), T)$. Throughout this paper, we shall assume that $\boldsymbol{\alpha}$ is a vector for which the sum of all its elements is one. This assumption implies that all probability distributions considered in this paper have a zero mass at $t = 0$. In O’Cinneide [27], the following fundamental characterization of PH -distributions was proved.

Theorem 2.1. (Theorem 1.1 in O’Cinneide [27]) *A probability distribution with nonnegative support and a rational Laplace-Stieltjes transform is a PH -distribution if and only if*

- *it is either the point mass at zero, or*

- it has a continuous positive density on the positive real numbers, and its Laplace-Stieltjes transform has a unique pole of maximal real part. \square

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* (*ME-representation*) of a *matrix-exponential distribution* (*ME-distribution*). For the aforementioned reason, we assume that $\boldsymbol{\alpha} \mathbf{u} = 1$ so that the *ME-distribution* has a zero mass at $t = 0$. We refer to Asmussen and Bladt [4] for more details of *ME-distributions*. The class of *PH-distributions* is a subset of the class of *ME-distributions*. Throughout this paper, if $(\boldsymbol{\alpha}, T)$ is used, it signifies that $(\boldsymbol{\alpha}, T)$ is a *PH-representation* of a *PH-distribution*, where $\boldsymbol{\alpha}$ is nonnegative and T is a *PH-generator*. If $(\boldsymbol{\alpha}, T, \mathbf{u})$ is used, $\boldsymbol{\alpha}$ may not be nonnegative, T may not be a *PH-generator*, and it represents an *ME-distribution*.

It is well known that the class of *PH-distributions* is dense in the class of probability distributions with nonnegative support. This leads to the issue of finding *PH-approximations* for probability distributions. We address this issue by considering approximations related to the class of Coxian distributions. For $\mathbf{x} = (x_1, x_2, \dots, x_N)$, where N is a positive integer, a bi-diagonal matrix $S(\mathbf{x})$ is defined as

$$S(\mathbf{x}) = \begin{pmatrix} -x_1 & 0 & \cdots & \cdots & 0 \\ x_2 & -x_2 & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & x_{N-1} & -x_{N-1} & 0 \\ 0 & \ddots & 0 & x_N & -x_N \end{pmatrix} \quad (2.2)$$

If $\{x_1, x_2, \dots, x_N\}$ are all real and positive and $\boldsymbol{\beta}$ is a probability vector (i.e., $\boldsymbol{\beta} \geq 0$ and $\boldsymbol{\beta} \mathbf{e} = 1$), then $(\boldsymbol{\beta}, S(\mathbf{x}))$ is called a *Coxian representation*, which represents a *Coxian distribution*. The class of Coxian distributions is a subset of the class of *PH-distributions*. Further, if $x_1 \geq x_2 \geq \dots \geq x_N > 0$, then $(\boldsymbol{\beta}, S(\mathbf{x}))$ is called an ordered Coxian representation. In this paper, we are interested in a special class of ordered Coxian representations with \mathbf{x} being of the form given in equation (3.9). If $x_1 = x_2 = \dots = x_N > 0$, then $(\boldsymbol{\beta}, S(\mathbf{x}))$ represents a *generalized Erlang distributions* (a mixture of Erlang distributions). The following theorem that characterizes the class of Coxian distributions is also due to O’Cinneide.

Theorem 2.2. (Theorem 4.1 in O’Cinneide [28] and Theorem 5.2 in O’Cinneide [29]) *Every PH-distribution whose Laplace-Stieltjes transform has only real poles is a Coxian distribution and has an ordered Coxian representation.* \square

Like the *PH-representation* of a *PH-distribution*, the ordered Coxian representation of a Coxian distribution is not unique. We call an ordered Coxian representation with the minimal

number of phases a *minimal ordered Coxian representation*. The number of phases of a minimal ordered Coxian representation is called the *triangular order* of the corresponding *PH*-distribution (O’Cinneide [29]). Theorem 6.2 in O’Cinneide [30] gave a necessary and sufficient condition for an ordered Coxian representation to be minimal. In HE and Zhang [16], a set of necessary and sufficient conditions for an ordered Coxian representation to be minimal was identified, which led to an algorithm for computing minimal Coxian representations of Coxian distributions. The following results show some relationships between Coxian representations and *PH*-representations.

Theorem 2.3. 1) (Theorem 2 in Cumani [10]) *Any PH-representation with a triangular PH-generator has an equivalent ordered Coxian representation of the same order;*
 2) (Theorem 4.2 in HE and Zhang [15]) *Any PH-representation with a symmetric PH-generator has an equivalent ordered Coxian representation of the same order.* \square

In general, a *PH*-representation that represents a Coxian distribution has equivalent Coxian representations, but the orders of the equivalent Coxian representations may be greater than that of the *PH*-representation (see Theorem 4.5 in HE and Zhang [15] and Example 6.1 in HE and Zhang [16]).

If $\{x_1, x_2, \dots, x_N\}$ are all real and positive and β may or may not be nonnegative, we call $(\beta, S(\mathbf{x}), \mathbf{e})$ a *Coxian function representation* that represents *Coxian function* $1 - \beta \exp\{S(\mathbf{x})t\}\mathbf{e}$, which may or may not be a probability distribution. If β is not nonnegative but the function $1 - \beta \exp\{S(\mathbf{x})t\}\mathbf{e}$ is a probability distribution, then $(\beta, S(\mathbf{x}), \mathbf{e})$ is an *ME*-representation of a Coxian distribution. For such a case, the algorithm developed in HE and Zhang [16] can be used for computing an ordered Coxian representation of the Coxian distribution.

PH-majorization was introduced and studied in O’Cinneide [26]. For a given *PH*-generator T , we denote by $PH(T)$ the set of all *PH*-distributions with a *PH*-representation (α, T) . For two *PH*-generators T and S , S is said to *PH-majorize* T if $PH(T) \subseteq PH(S)$. For a *PH*-generator T , if *PH*-representations (α, T) and (β, T) represent two different distributions for any different α and β , then T is called *PH-simple*. It was shown in O’Cinneide [26] that a *PH*-simple S *PH-majorizes* T if and only if there exists a nonnegative matrix P with unit row sums for which $TP = PS$. If S *PH-majorizes* T , then (α, T) and $(\alpha P, S)$ represent the same *PH*-distribution. Using the notion of *PH*-majorization, Theorem 2.3 can be improved to a stronger form: 1) Any triangular *PH*-generator is *PH*-majorized by an ordered Coxian generator of the same order; 2) Any symmetric *PH*-generator is *PH*-majorized by an ordered Coxian generator of the same order. In Theorem 5.1 of this paper, we shall show that a *PH*-generator with only real eigenvalues is *PH*-majorized by an ordered Coxian generator of the same or a higher order.

Similar to PH -majorization, ME -majorization can be defined for ME -distributions. For a pair $\{T, \mathbf{u}\}$, we denote by $ME(T, \mathbf{u})$ the set of all ME -distributions 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 $ME(T, \mathbf{u}) \subseteq ME(S, \mathbf{v})$. According to Proposition 2.1 in HE and Zhang [15], if there exists a matrix P such that $TP = PS$ and $\mathbf{u} = P\mathbf{v}$, then $\{S, \mathbf{v}\}$ *ME-majorizes* $\{T, \mathbf{u}\}$, i.e., if $(\boldsymbol{\alpha}, T, \mathbf{u})$ is an ME -distribution, then $(\boldsymbol{\alpha}P, S, \mathbf{v})$ represents the same ME -distribution.

The following spectral polynomial algorithm introduced in HE and Zhang [15] is useful for computing bi-diagonal representations of ME -distributions and PH -distributions. Suppose that $(\boldsymbol{\alpha}, T, \mathbf{u})$ is an ME -representation of order m . Denote by $\mathbf{x} = (x_1, x_2, \dots, x_N)$, a set of nonzero complex numbers, where N is a positive integer. Define

$$\begin{aligned} \mathbf{p}_1 &= -T\mathbf{u}/x_1; \\ \mathbf{p}_n &= (x_{n-1}I + T)\mathbf{p}_{n-1}/x_n, \quad 2 \leq n \leq N; \\ \mathbf{p}_{N+1} &= (x_N I + T)\mathbf{p}_N. \end{aligned} \tag{2.3}$$

Let $P = (\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_N)$, which is an $m \times N$ matrix. If $\mathbf{p}_{N+1} = 0$, equation (2.3) can be rewritten as $TP = PS(\mathbf{x})$ and it can be shown that $Pe = \mathbf{u}$. Then $(\boldsymbol{\alpha}P, S(\mathbf{x}), e)$ represents the same ME -distribution as $(\boldsymbol{\alpha}, T, \mathbf{u})$. We also say that $(\boldsymbol{\alpha}P, S(\mathbf{x}), e)$ and $(\boldsymbol{\alpha}, T, \mathbf{u})$ are *equivalent* representations. Similar to the proof of Propositions 2.1 and 3.1 in HE and Zhang [15], the following results can be proved.

Theorem 2.4. *If $\mathbf{p}_{N+1} = 0$, then $\{S(\mathbf{x}), e\}$ *ME-majorizes* $\{T, \mathbf{u}\}$, and the ME -distribution $(\boldsymbol{\alpha}, T, \mathbf{u})$ has a bi-diagonal ME -representation $(\boldsymbol{\beta}, S(\mathbf{x}), e)$ of order N with $\boldsymbol{\beta}e = \boldsymbol{\alpha}\mathbf{u} = 1$ and $\boldsymbol{\beta} = \boldsymbol{\alpha}P$. \square*

A particular choice of \mathbf{x} was given in HE and Zhang [15]. Denote by $\{-\lambda_1, -\lambda_2, \dots, -\lambda_m\}$ the spectrum of T (counting multiplicities). If $\{\lambda_1, \lambda_2, \dots, \lambda_m\}$ is a subset of \mathbf{x} , by the Cayley-Hamilton theorem (Lancaster and Tismenetsky [17]), we have $\mathbf{p}_{N+1} = 0$. If $x_1 \geq x_2 \geq \dots \geq x_N > 0$ and $\boldsymbol{\alpha}P$ is nonnegative, $(\boldsymbol{\alpha}P, S(\mathbf{x}))$ is an ordered Coxian representation. For this case, we find an equivalent ordered Coxian representation for $(\boldsymbol{\alpha}, T, \mathbf{u})$.

3 Approximating ME -Distributions by Coxian Distributions

Theorem 2.4 indicates that any ME -distribution has a bi-diagonal ME -representation, provided all poles of its Laplace-Stieltjes transform are nonzero. However, the bi-diagonal ME -representation $(\boldsymbol{\alpha}P, S(\mathbf{x}), e)$ may be neither a PH -representation nor a Coxian representation. Since the class of Coxian distributions is dense in the class of probability distributions with nonnegative support, for any ME -distribution, there are Coxian distributions that are

arbitrarily close to it. The objective of this section is to propose an algorithm that can find Coxian representations of Coxian distributions that approximate an ME -distribution. To that end, we study the relationship between $(\boldsymbol{\alpha}, T, \mathbf{u})$ and the representation $(\boldsymbol{\alpha}P, S(\mathbf{x}), \mathbf{e})$ obtained by the spectral polynomial algorithm first.

Denote by $F_{(\boldsymbol{\alpha}, T, \mathbf{u})}(t) = 1 - \boldsymbol{\alpha} \exp\{Tt\} \mathbf{u}$, $t \geq 0$, for $(\boldsymbol{\alpha}, T, \mathbf{u})$. For given \mathbf{x} , since $\mathbf{p}_{N+1} = 0$ may not be true, the Coxian function representation $(\boldsymbol{\alpha}P, S(\mathbf{x}), \mathbf{e})$ may not be an equivalent representation of $(\boldsymbol{\alpha}, T, \mathbf{u})$. Denote by $F_{(\boldsymbol{\alpha}P, S(\mathbf{x}), \mathbf{e})}(t) = 1 - \boldsymbol{\alpha}P \exp\{S(\mathbf{x})t\} \mathbf{e}$, $t \geq 0$, which may or may not represent a probability distribution.

Lemma 3.1. *Assume that all elements of \mathbf{x} are nonzero and T is invertible. For $(\boldsymbol{\alpha}, T, \mathbf{u})$ and the representation $(\boldsymbol{\alpha}P, S(\mathbf{x}), \mathbf{e})$ obtained by the spectral polynomial algorithm, we have*

$$F_{(\boldsymbol{\alpha}P, S(\mathbf{x}), \mathbf{e})}(t) - F_{(\boldsymbol{\alpha}, T, \mathbf{u})}(t) = \varepsilon_1(t) + \varepsilon_2(t), \quad t \geq 0, \quad (3.1)$$

where

$$\begin{aligned} \varepsilon_1(t) &= \sum_{k=0}^{\infty} (\boldsymbol{\alpha} T^k \mathbf{p}_{N+1}) \sum_{n=0}^{\infty} \frac{t^{n+k+1}}{(n+k+1)!} \left(\mathbf{e}_N(S(\mathbf{x}))^n \mathbf{e} \right), \\ \varepsilon_2(t) &= -\boldsymbol{\alpha} \exp\{Tt\} T^{-1} \mathbf{p}_{N+1}, \end{aligned} \quad (3.2)$$

and $\mathbf{e}_N = (0, \dots, 0, 1)$, a row vector of size N .

Proof. Since all elements of \mathbf{x} are nonzero, the matrix P is well defined. First note that equation (2.3) can be rewritten as

$$TP = PS(\mathbf{x}) + (0, \dots, 0, \mathbf{p}_{N+1}), \quad (3.3)$$

where $(0, \dots, 0, \mathbf{p}_{N+1})$ is an $m \times N$ matrix. By (3.3), it is easy to show that, for $n \geq 0$,

$$T^n P = P(S(\mathbf{x}))^n + \sum_{k=0}^{n-1} T^k (0, \dots, 0, \mathbf{p}_{N+1}) (S(\mathbf{x}))^{n-1-k}. \quad (3.4)$$

Note that, for $n = 0$, both sides of equation (3.4) are reduced to P . Pre-multiplying by $\boldsymbol{\alpha}$, post-multiplying by $t^n \mathbf{e}/n!$ on both sides of equation (3.4), and summing them up, for $n \geq 0$, yield, for $t \geq 0$,

$$\begin{aligned} \boldsymbol{\alpha} \exp\{Tt\} P \mathbf{e} &= \boldsymbol{\alpha} P \exp\{S(\mathbf{x})t\} \mathbf{e} + \sum_{n=1}^{\infty} \frac{t^n}{n!} \sum_{k=0}^{n-1} \left(\boldsymbol{\alpha} T^k \mathbf{p}_{N+1} \right) \left(\mathbf{e}_N(S(\mathbf{x}))^{n-1-k} \mathbf{e} \right) \\ &= \boldsymbol{\alpha} P \exp\{S(\mathbf{x})t\} \mathbf{e} + \sum_{k=0}^{\infty} \left(\boldsymbol{\alpha} T^k \mathbf{p}_{N+1} \right) \sum_{n=0}^{\infty} \frac{t^{n+k+1}}{(n+k+1)!} \left(\mathbf{e}_N(S(\mathbf{x}))^n \mathbf{e} \right) \\ &= \boldsymbol{\alpha} P \exp\{S(\mathbf{x})t\} \mathbf{e} + \varepsilon_1(t). \end{aligned} \quad (3.5)$$

From equation (2.3), we have

$$\mathbf{p}_{N+1} = T(\mathbf{p}_N + \mathbf{p}_{N-1} + \dots + \mathbf{p}_2 + \mathbf{p}_1) + x_1 \mathbf{p}_1 = TP\mathbf{e} - T\mathbf{u}, \quad (3.6)$$

which leads to $P\mathbf{e} = \mathbf{u} + T^{-1}\mathbf{p}_{N+1}$ (Note that T is invertible). Equation (3.1) is obtained from equation (3.5) and $P\mathbf{e} = \mathbf{u} + T^{-1}\mathbf{p}_{N+1}$. This completes the proof of Lemma 3.1. \square

The Coxian function $F_{(\boldsymbol{\alpha}_P, S(\mathbf{x}), \mathbf{e})}(t)$ may not be a probability distribution function (see Example 6.1). Nonetheless, equation (3.1) indicates that $F_{(\boldsymbol{\alpha}_P, S(\mathbf{x}), \mathbf{e})}(t)$ can be a satisfactory approximation of $F_{(\boldsymbol{\alpha}_T, \mathbf{u})}(t)$ if $\varepsilon_1(t)$ and $\varepsilon_2(t)$ are small enough for all t . By equation (3.2), it is clear that $F_{(\boldsymbol{\alpha}_P, S(\mathbf{x}), \mathbf{e})}(t) = F_{(\boldsymbol{\alpha}_T, \mathbf{u})}(t)$, for $t \geq 0$, if $\mathbf{p}_{N+1} = 0$. The condition $\mathbf{p}_{N+1} = 0$ is sufficient but not necessary for $F_{(\boldsymbol{\alpha}_P, S(\mathbf{x}), \mathbf{e})}(t) = F_{(\boldsymbol{\alpha}_T, \mathbf{u})}(t)$, for $t \geq 0$, and, for some cases, is not possible. It can be shown, by comparing the coefficients of t^n on both sides, $F_{(\boldsymbol{\alpha}_P, S(\mathbf{x}), \mathbf{e})}(t) = F_{(\boldsymbol{\alpha}_T, \mathbf{u})}(t)$, for $t \geq 0$, if and only if $\boldsymbol{\alpha}T^n\mathbf{p}_{N+1} = 0$ for $n \geq -1$. If the condition $\{\boldsymbol{\alpha}T^n\mathbf{p}_{N+1} = 0, \text{ for } n \geq -1\}$ is used to find approximations or new representations, the results will depend on the vector $\boldsymbol{\alpha}$. A study in that direction is beyond the scope of this paper. In this paper, we shall focus on new representations for which the Coxian generator $S(\mathbf{x})$ is independent of $\boldsymbol{\alpha}$. Basically, we look for \mathbf{x} such that \mathbf{p}_{N+1} is small.

Define

$$\zeta(k, t) = \sum_{n=0}^{\infty} \frac{t^{n+k}}{(n+k)!} \left(\mathbf{e}_N (S(\mathbf{x}))^n \mathbf{e} \right), \quad k \geq 0, t \geq 0. \quad (3.7)$$

Lemma 3.2. *Assume that all elements of \mathbf{x} are positive. Then $0 < \zeta(k, t) \leq t^k/k!$, $k \geq 0, t \geq 0$.*

Proof. Denote by $\zeta^{(j)}(k, t)$ the j -th derivative of $\zeta(k, t)$ with respect to t . It is easy to verify that $\zeta^{(j)}(k, t) = \zeta(k-j, t)$, $0 \leq j \leq k, t \geq 0$ and $\zeta^{(j)}(k, t) = \zeta(0, t) = \mathbf{e}_N \exp\{S(\mathbf{x})t\}\mathbf{e} \geq 0, k \geq 0$. Since $\zeta(k, 0) = 0, k \geq 1$, by induction, it can be proved that $\zeta(k, t)$ is nonnegative and non-decreasing in t . Since all elements of \mathbf{x} are positive, $S(\mathbf{x})$ is a PH -generator and $(\mathbf{e}_N, S(\mathbf{x}))$ represents a PH -distribution. Then $1 - \zeta(0, t)$ is a probability distribution function. Therefore, $0 < \zeta(0, t) \leq 1$, for $t \geq 0$. Again, by induction, $\zeta(k, t)$ is positive and

$$\zeta(k, t) = \int_0^t \zeta(k-1, z) dz \leq \int_0^t \frac{z^{k-1}}{(k-1)!} dz \leq \frac{t^k}{k!}, \quad t \geq 0, k \geq 1. \quad (3.8)$$

This completes the proof of Lemma 3.2. \square

Now, we are ready to state and prove the main result of this section. We show that $|\varepsilon_1(t) + \varepsilon_2(t)|$ can be arbitrarily small if \mathbf{x} is chosen properly.

Theorem 3.3. *Assume that all the eigenvalues of T have negative real parts. We choose an N -dimensional vector*

$$\mathbf{x} = (\lambda, \dots, \lambda, y_1, \dots, y_{L-1}, y_L), \quad (3.9)$$

where L is a positive integer, λ is repeated in \mathbf{x} $N - L$ times, and $y_1 \geq y_2 \geq \dots \geq y_{L-1} \geq y_L > 0$. For any given positive ε^* , there exists a positive N^* such that, for any $N \geq N^*$,

if λ is large enough, we have $|\varepsilon_1(t) + \varepsilon_2(t)| < \varepsilon^*$ for $t \geq 0$. Consequently, there is always a Coxian function approximation $(\alpha P, S(\mathbf{x}), \mathbf{e})$ to (α, T, \mathbf{u}) for any given error level ε^* . In addition, the Coxian function $(\alpha P/(\alpha P \mathbf{e}), S(\mathbf{x}), \mathbf{e})$ also approximates (α, T, \mathbf{u}) if N is large enough.

Proof. The proof consists of two parts. First, we show the theorem for $t > t^*$, where t^* is a large positive number (to be determined). Second, we prove the theorem for $0 \leq t \leq t^*$.

Let $\text{re}(\lambda_j)$ and $\text{imag}(\lambda_j)$ denote the real and imaginary parts of the complex number λ_j , respectively, where $-\lambda_j$ is an eigenvalue of T . Since the real part of each eigenvalue of T is negative, we know that $\text{re}(\lambda_j)$ is positive. Thus, if λ is large enough, $(\text{re}(\lambda_j))^2 + (\text{imag}(\lambda_j))^2 < 2\text{re}(\lambda_j)\lambda$, which is equivalent to $\|1 - \lambda_j/\lambda\| < 1$, where $\|1 - \lambda_j/\lambda\| = [(1 - \text{re}(\lambda_j)/\lambda)^2 + (\text{imag}(\lambda_j)/\lambda)^2]^{1/2}$. Let η be a real number such that $\max\{\|1 - \lambda_j/\lambda\|, 1 \leq j \leq m\} < \eta < 1$. Since $\lambda \geq y_1 \geq y_2 \geq \dots \geq y_{L-1} \geq y_L > 0$, $S(\mathbf{x})$ is a *PH*-generator. Thus, every element of the vector $\exp\{S(\mathbf{x})t\}\mathbf{e}$ is less than or equal to one for $t \geq 0$. By definition, we have $\mathbf{p}_1 = -T\mathbf{u}/\lambda$ and

$$\begin{aligned} \mathbf{p}_n &= \left(I + \frac{T}{\lambda}\right)^{n-1} \mathbf{p}_1, \quad 2 \leq n \leq N-L; \\ \mathbf{p}_{N-L+1} &= \frac{\lambda}{y_1} \left(I + \frac{T}{\lambda}\right)^{N-L} \mathbf{p}_1, \\ \mathbf{p}_{N-L+j} &= \frac{\lambda}{y_j \cdots y_1} (y_{j-1}I + T) \cdots (y_1I + T) \left(I + \frac{T}{\lambda}\right)^{N-L} \mathbf{p}_1, \quad 2 \leq j \leq L; \\ \mathbf{p}_{N+1} &= \frac{\lambda}{y_L \cdots y_1} (y_L I + T) \cdots (y_1 I + T) \left(I + \frac{T}{\lambda}\right)^{N-L} \mathbf{p}_1. \end{aligned} \tag{3.10}$$

Note that the spectrum of $I + T/\lambda$ is $\{1 - \lambda_j/\lambda, 1 \leq j \leq m\}$. Using the Jordan canonical form of $I + T/\lambda$ (Lancaster and Tismenetsky [17]), we have $\mathbf{p}_n = o(\eta^{n-m-L})\mathbf{e}$ for large n . Then we have

$$\begin{aligned} \left| \alpha P \exp\{S(\mathbf{x})t\}\mathbf{e} \right| &= \left| \sum_{n=1}^N (\alpha \mathbf{p}_n) \left(\exp\{S(\mathbf{x})t\}\mathbf{e} \right) \right| \\ &\leq \left| \sum_{n=1}^{N_1} (\alpha \mathbf{p}_n) \left(\exp\{S(\mathbf{x})t\}\mathbf{e} \right)_n \right| + c_1 \sum_{n=N_1+1}^N \eta^{n-m-L} \left| \left(\exp\{S(\mathbf{x})t\}\mathbf{e} \right)_n \right| \\ &\leq \left| \sum_{n=1}^{N_1} (\alpha \mathbf{p}_n) \left(\exp\{S(\mathbf{x})t\}\mathbf{e} \right)_n \right| + c_1 \frac{\eta^{N_1+1-m-L}}{1-\eta}, \end{aligned} \tag{3.11}$$

where c_1 is a positive constant and N_1 is a fixed large integer. If N_1 is large enough, the second part of the right-hand side of the last line in equation (3.11) can be made smaller than $\varepsilon^*/4$. For fixed N_1 , if t is large enough, the first part of the right-hand side of the last line in equation (3.11) can be made smaller than $\varepsilon^*/4$. Therefore, if t is large enough, we have $|1 - F(\alpha P, S(\mathbf{x}), \mathbf{e})(t)| = |\alpha P \exp\{S(\mathbf{x})t\}\mathbf{e}| < \varepsilon^*/2$.

Since $(\boldsymbol{\alpha}, T, \mathbf{u})$ is a probability distribution, we must have $1 - F_{(\boldsymbol{\alpha}, T, \mathbf{u})}(t) < \varepsilon^*/2$ if t is large enough. Combining the above two results, we conclude that there exists t^* such that for $t > t^*$ and any N ,

$$|\varepsilon_1(t) + \varepsilon_2(t)| \leq |1 - F_{(\boldsymbol{\alpha}, T, \mathbf{u})}(t)| + |1 - F_{(\boldsymbol{\alpha}P, S(\mathbf{x}), \mathbf{e})}(t)| < \frac{\varepsilon^*}{2} + \frac{\varepsilon^*}{2} = \varepsilon^*. \quad (3.12)$$

Let $|\boldsymbol{\alpha}| = (|\alpha_j|)$ and $|T| = (|t_{i,j}|)$. By equation (3.2) and Lemma 3.2, we have

$$\begin{aligned} |\varepsilon_1(t)| &\leq \sum_{k=0}^{\infty} \left| \boldsymbol{\alpha} T^k \mathbf{p}_{N+1} \right| \frac{t^{k+1}}{(k+1)!} \\ &\leq t |\boldsymbol{\alpha}| \exp\{|T|t\} |\mathbf{p}_{N+1}| \\ &\leq c_2 \cdot (t |\boldsymbol{\alpha}| \exp\{|T|t\} \mathbf{e}) \eta^{N-m-L}; \\ \varepsilon_2(t) &= \left| \boldsymbol{\alpha} \exp\{Tt\} T^{-1} \mathbf{p}_{N+1} \right| \leq c_2 \cdot (|\boldsymbol{\alpha}| \exp\{|T|t\} |T^{-1}| \mathbf{e}) \eta^{N-m-L}, \end{aligned} \quad (3.13)$$

where c_2 is a positive constant. Therefore, for $0 \leq t \leq t^*$, if N is large enough, equation (3.13) leads to $|\varepsilon_1(t) + \varepsilon_2(t)| \leq |\varepsilon_1(t)| + |\varepsilon_2(t)| < \varepsilon^*$.

In summary, we have shown that if N is large enough, $|\varepsilon_1(t) + \varepsilon_2(t)| < \varepsilon^*$ for $t \geq 0$.

To prove the last part of the theorem, note that equation (3.6) leads to $\boldsymbol{\alpha} P \mathbf{e} = \boldsymbol{\alpha} \mathbf{u} + \boldsymbol{\alpha} T^{-1} \mathbf{p}_{N+1} = 1 + o(\eta^N)$. Thus, $\boldsymbol{\alpha} P - \boldsymbol{\alpha} P / (\boldsymbol{\alpha} P \mathbf{e}) = (\boldsymbol{\alpha} P) o(\eta^N)$. Therefore, $(\boldsymbol{\alpha} P / (\boldsymbol{\alpha} P \mathbf{e}), S(\mathbf{x}), \mathbf{e})$ approximates $(\boldsymbol{\alpha} P, S(\mathbf{x}), \mathbf{e})$. Together with the first part, we conclude that $(\boldsymbol{\alpha} P / (\boldsymbol{\alpha} P \mathbf{e}), S(\mathbf{x}), \mathbf{e})$ is a satisfactory approximation of $(\boldsymbol{\alpha}, T)$ if N is large enough. This completes the proof of Theorem 3.3. \square

By the well-known fact that the class of Coxian distributions is dense in the class of all probability distributions with nonnegative support, Theorem 3.3 is not surprising. The construction process of $\{\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_N\}$ bears some similarities to the construction process of an invariant polytope in O'Connell [27]. Since our construction is realized in a linear space, instead of a measure space, more analytic results can be obtained. More importantly, Theorem 3.3 leads to simple algorithms to find satisfactory Coxian approximations of *ME*-distributions and *PH*-distributions. In Theorem 3.3, the condition that every eigenvalue of T has a negative real part is not restrictive since the poles of the Laplace-Stieltjes transforms of probability distribution functions must have negative real parts.

The Coxian function representation $(\boldsymbol{\alpha} P, S(\mathbf{x}), \mathbf{e})$ obtained in Theorem 3.3 may not be a desired solution for several reasons: 1) $\boldsymbol{\alpha} P$ may not be real; 2) The representation may not represent a Coxian distribution (see Example 6.1); and 3) $\boldsymbol{\alpha} P$ may not be nonnegative. By Theorem 3.3, the first issue can be resolved by simply removing the imaginary part of $\boldsymbol{\alpha} P$. The second issue can be dealt with by using Theorem 3.4 (to be proved next). The last reason implies that, even if the representation does represent a Coxian distribution, it may not be a Coxian representation. This issue can be dealt with by using Theorem 3.4 and an algorithm developed in HE and Zhang [16] (see the discussion right after the proof of Theorem 3.4).

Next, from $(\alpha P, S(\mathbf{x}), \mathbf{e})$, we construct a Coxian distribution that approximates the original ME-distribution.

Define $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_N) = \alpha P$. By the definition given in equation (3.10), it is easy to see that β_n remains the same for $N > n + L$. Thus, the dependency of β_n on N is not shown explicitly. Let $\boldsymbol{\beta}_R$ be the real part of $\boldsymbol{\beta}$. Define (componentwise)

$$\boldsymbol{\beta}_R^+ = \max\{0, \boldsymbol{\beta}_R\} \quad \text{and} \quad \boldsymbol{\beta}_R^- = \max\{0, -\boldsymbol{\beta}_R\}. \quad (3.14)$$

Then it is easy to see $\boldsymbol{\beta}_R = \boldsymbol{\beta}_R^+ - \boldsymbol{\beta}_R^-$. Define, for $1 \leq N_0 \leq N$,

$$\boldsymbol{\beta}(N_0, N) = \boldsymbol{\beta}_R^+ - (\beta_{R,1}^-, \beta_{R,2}^-, \dots, \beta_{R,N_0}^-, 0, \dots, 0). \quad (3.15)$$

Theorem 3.4. *Assume that all conditions in Theorem 3.3 hold. In addition, we assume that the density function of $(\boldsymbol{\alpha}, T, \mathbf{u})$ is positive on the positive real numbers. If N is large enough, there exists N_0 such that $(\boldsymbol{\beta}(N_0, N)/(\boldsymbol{\beta}(N_0, N)\mathbf{e}), S(\mathbf{x}), \mathbf{e})$ represents a Coxian distribution that approximates the ME-distribution $(\boldsymbol{\alpha}, T, \mathbf{u})$.*

Proof. To show that $(\boldsymbol{\beta}(N_0, N)/(\boldsymbol{\beta}(N_0, N)\mathbf{e}), S(\mathbf{x}), \mathbf{e})$ represents a Coxian distribution, we need to prove that the corresponding derivative function is nonnegative. For $1 \leq n \leq N$, define

$$\begin{aligned} G_n(t) &= 1 - \mathbf{e}_n \exp\{S(\mathbf{x})t\}\mathbf{e}, \quad t \geq 0; \\ G_n^{(1)}(t) &= -\mathbf{e}_n \exp\{S(\mathbf{x})t\}S(\mathbf{x})\mathbf{e}, \quad t \geq 0, \end{aligned} \quad (3.16)$$

where \mathbf{e}_n is a row vector with all elements being zero except that the n -th element is one and $G_n^{(1)}(t)$ is the first derivative of $G_n(t)$. For \mathbf{x} given in equation (3.9), $G_n(t)$ and $G_n^{(1)}(t)$ are the distribution function and the density function of a generalized Erlang distribution of order n , respectively. If $n \leq N - L$, $G_n(t)$ and $G_n^{(1)}(t) = \lambda e^{-\lambda t} (\lambda t)^{n-1} / (n-1)!$ are the distribution function and density function of an Erlang distribution of order n , respectively. By Theorem 3.3, $F(\boldsymbol{\alpha}_P, S(\mathbf{x}), \mathbf{e})(t)$ converges to $F(\boldsymbol{\alpha}, T, \mathbf{u})(t)$ uniformly as N goes to infinity. Since $|\beta_{R,n}| = |\boldsymbol{\alpha}_P \mathbf{p}_n| = o(\eta^n)$, we have

$$F(\boldsymbol{\alpha}, T, \mathbf{u})(t) = \sum_{n=1}^{\infty} \beta_{R,n} G_n(t) = \sum_{n=1}^{\infty} \beta_{R,n}^+ G_n(t) - \sum_{n=1}^{\infty} \beta_{R,n}^- G_n(t). \quad (3.17)$$

It is straightforward to verify, for $t > 0$,

$$\begin{aligned} F_{(\boldsymbol{\alpha}, T, \mathbf{u})}^{(1)}(t) &= \lim_{\delta \rightarrow 0} \frac{F(\boldsymbol{\alpha}, T, \mathbf{u})(t + \delta) - F(\boldsymbol{\alpha}, T, \mathbf{u})(t)}{\delta} \\ &= \lim_{\delta \rightarrow 0} \left(\sum_{n=1}^{\infty} \beta_{R,n}^+ \frac{G_n(t + \delta) - G_n(t)}{\delta} - \sum_{n=1}^{\infty} \beta_{R,n}^- \frac{G_n(t + \delta) - G_n(t)}{\delta} \right) \\ &= \sum_{n=1}^{\infty} \beta_{R,n}^+ \lim_{\delta \rightarrow 0} \frac{G_n(t + \delta) - G_n(t)}{\delta} - \sum_{n=1}^{\infty} \beta_{R,n}^- \lim_{\delta \rightarrow 0} \frac{G_n(t + \delta) - G_n(t)}{\delta} \\ &= \sum_{n=1}^{\infty} \beta_{R,n}^+ G_n^{(1)}(t) - \sum_{n=1}^{\infty} \beta_{R,n}^- G_n^{(1)}(t). \end{aligned} \quad (3.18)$$

The exchange of limits in the third equality in equation (3.18) is valid since the derivative of $G_n(t)$ is uniformly bounded for all $n > 0$ and $t \geq 0$ and $|\beta_{R,n}| = |\alpha \mathbf{p}_n| = o(\eta^n)$.

For any given error level, we can choose N_0 such that, for any $N \geq N_0$, the function $F_{(N_0,N)}(t) = \sum_{n=1}^N \beta_{R,n}^+ G_n(t) - \sum_{n=1}^{N_0} \beta_{R,n}^- G_n(t)$ approximates $F_{(\alpha,T,\mathbf{u})}(t)$. It is readily seen that $F_{(N_0,N)}(t)$ has a Coxian function representation $(\beta(N_0, N), S(\mathbf{x}), \mathbf{e})$. Since $F_{(\alpha,T,\mathbf{u})}^{(1)}(t)$ is positive in $(0, \infty)$, the function $\sum_{n=1}^{\infty} \beta_{R,n}^+ G_n^{(1)}(t) - \sum_{n=1}^{N_0} \beta_{R,n}^- G_n^{(1)}(t)$ is positive in $(0, \infty)$. Next, we show that there exists a finite N such that the derivative function $F_{(N,N_0)}^{(1)}(t) = \sum_{n=1}^N \beta_{R,n}^+ G_n^{(1)}(t) - \sum_{n=1}^{N_0} \beta_{R,n}^- G_n^{(1)}(t)$ is nonnegative. We consider three cases: 1) t is very large; 2) t is close to zero; and 3) others.

First, we consider case 1). For $N_0 < n < N - L$, we have

$$\frac{G_{N_0}^{(1)}(t)}{G_n^{(1)}(t)} = \frac{(n-1)!}{(N_0-1)!(\lambda t)^{n-N_0}} \rightarrow 0, \quad t \rightarrow \infty. \quad (3.19)$$

If $\beta_{R,n}^+ = 0$ for all $n > N_0$, then $F_{(N_0,N)}^{(1)}(t) \geq F_{(\alpha,T,\mathbf{u})}^{(1)}(t) > 0$ for $t > 0$. Otherwise, there exists $n_1 > N_0$ such that $\beta_{R,n_1}^+ > 0$. By equation (3.19), there exists t_0 such that, for $t > t_0$, the function $F_{(N_0,N)}^{(1)}(t)$ is positive for $t > 0$ and $N \geq n_1$.

For case 2), first note that $F_{(\alpha,T,\mathbf{u})}^{(n)}(0) = -\alpha T^n \mathbf{u}$, $n \geq 1$, where $F_{(\alpha,T,\mathbf{u})}^{(n)}(0)$ denotes the n -th derivative of the function $F_{(\alpha,T,\mathbf{u})}(t)$ at $t = 0$. Let $n_2 = \min\{n : \beta_n \neq 0, n \geq 1\}$. Since $\beta_1 = \alpha \mathbf{p}_1 = -\alpha T \mathbf{u} / \lambda$, it can be shown that $F_{(\alpha,T,\mathbf{u})}^{(n)}(0) = 0$, $1 \leq n \leq n_2$, and $F_{(\alpha,T,\mathbf{u})}^{(n_2)}(0) = -\alpha T^{n_2} \mathbf{u} = \lambda^{n_2} \beta_{n_2}$. Since $F_{(\alpha,T,\mathbf{u})}(t)$ is nonnegative, we must have $F_{(\alpha,T,\mathbf{u})}^{(n_2)}(0) > 0$. Consequently, we have $\beta_{n_2} > 0$. Since $G_n^{(1)}(t)$ is the density function of the sum of n exponential random variables, it can be shown that

$$\begin{aligned} \left| \frac{\sum_{n=n_2+1}^{\infty} \beta_n G_n^{(1)}(t)}{\beta_{n_2} G_{n_2}^{(1)}(t)} \right| &\leq \frac{\sum_{n=n_2+1}^{\infty} |\beta_n| G_n^{(1)}(t)}{\beta_{n_2} G_{n_2}^{(1)}(t)} \\ &\leq ct \sum_{n=n_2+1}^{\infty} \frac{(n_2-1)!(\lambda t)^{n-n_2-1}}{(n-1)!} \leq cte^{\lambda t} \rightarrow 0, \end{aligned} \quad (3.20)$$

as $t \rightarrow 0$, where c is a constant. Then we have

$$\begin{aligned} F_{(N_0,N)}^{(1)}(t) &= \beta_{n_2} G_{n_2}^{(1)}(t) \left(1 + \sum_{n=n_2+1}^N \frac{\beta_{R,n}^+ G_n^{(1)}(t)}{\beta_{R,n_2} G_{n_2}^{(1)}(t)} - \sum_{n=n_2+1}^{N_0} \frac{\beta_{R,n}^- G_n^{(1)}(t)}{\beta_{R,n_2} G_{n_2}^{(1)}(t)} \right) \\ &\geq \beta_{n_2} G_{n_2}^{(1)}(t) \left(1 - \left| \sum_{n=n_2+1}^N \frac{\beta_{R,n}^+ G_n^{(1)}(t)}{\beta_{R,n_2} G_{n_2}^{(1)}(t)} + \sum_{n=n_2+1}^{N_0} \frac{\beta_{R,n}^- G_n^{(1)}(t)}{\beta_{R,n_2} G_{n_2}^{(1)}(t)} \right| \right) \\ &\geq \beta_{n_2} G_{n_2}^{(1)}(t) (1 - cte^{\lambda t}). \end{aligned} \quad (3.21)$$

Therefore, we have $F_{(N_0,N)}^{(1)}(0) \geq F_{(\alpha,T,\mathbf{u})}^{(1)}(0) \geq 0$ and $F_{(N_0,N)}^{(1)}(t) > 0$ if t is close to zero, say, $0 < t < \delta$, for some positive δ , for all $N > N_0$.

For case 3), since $F_{(\boldsymbol{\alpha}, T, \mathbf{u})}^{(1)}(t)$ is positive for $\delta \leq t \leq t_0$, by equation (3.18), it is easy to see that we can choose N large enough so that $F_{(N_0, N)}^{(1)}(t)$ is positive.

Combining the above three cases, we conclude that, if N is large enough, the function $F_{(N_0, N)}(t)$ with a representation $(\boldsymbol{\beta}(N_0, N), S(\mathbf{x}), \mathbf{e})$ has a nonnegative derivative.

Since $\boldsymbol{\beta}(N_0, N)\mathbf{e} \approx 1$, the Coxian distribution represented by $(\boldsymbol{\beta}(N_0, N)/(\boldsymbol{\beta}(N_0, N)\mathbf{e}), S(\mathbf{x}), \mathbf{e})$ approximates the Coxian function represented by $(\boldsymbol{\beta}(N_0, N), S(\mathbf{x}), \mathbf{e})$. Since the Coxian function $(\boldsymbol{\beta}(N_0, N), S(\mathbf{x}), \mathbf{e})$ approximates the *ME*-distribution $(\boldsymbol{\alpha}, T, \mathbf{u})$, the Coxian distribution $(\boldsymbol{\beta}(N_0, N)/(\boldsymbol{\beta}(N_0, N)\mathbf{e}), S(\mathbf{x}), \mathbf{e})$ approximates the *ME*-distribution $(\boldsymbol{\alpha}, T, \mathbf{u})$. This completes the proof of Theorem 3.4. \square

If $\boldsymbol{\beta}(N_0, N)$ is nonnegative, $(\boldsymbol{\beta}(N_0, N)/(\boldsymbol{\beta}(N_0, N)\mathbf{e}), S(\mathbf{x}))$ is a Coxian representation of a Coxian distribution that approximates $(\boldsymbol{\alpha}, T, \mathbf{u})$. Otherwise, $(\boldsymbol{\beta}(N_0, N)/(\boldsymbol{\beta}(N_0, N)\mathbf{e}), S(\mathbf{x}), \mathbf{e})$ is an *ME*-representation that represents a Coxian distribution. An algorithm developed in HE and Zhang [16] can be used for computing a Coxian representation from $(\boldsymbol{\beta}(N_0, N)/(\boldsymbol{\beta}(N_0, N)\mathbf{e}), S(\mathbf{x}), \mathbf{e})$ for that Coxian distribution. We refer readers to HE and Zhang [16] for details of the algorithm for computing a minimal Coxian representation of a Coxian distribution.

It is clear from Theorem 3.3 that \mathbf{x} in the representation $(\boldsymbol{\alpha}P, S(\mathbf{x}), \mathbf{e})$ can be chosen differently. It is also clear that, to find a satisfactory approximation, we want to choose \mathbf{x} such that \mathbf{p}_{N+1} is small and P is nonnegative for N that is not large. In general, it is a complicated issue to choose a proper \mathbf{x} and we shall address this issue for some subsets of *ME*-distributions in Sections 4 and 5. In the meantime, we have the following observations that may help us choose \mathbf{x} .

According to Theorem 3.3, most of the elements of \mathbf{x} should be large enough so that η (defined in the proof of Theorem 3.3) is as small as possible. By routine calculations, it can be proved that $\|1 - \lambda_j/\lambda\|$, as a function of λ , is minimized at $\text{re}(\lambda_j) + (\text{imag}(\lambda_j))^2/\text{re}(\lambda_j)$. Therefore, we shall choose λ so that

$$\lambda \geq \max_{1 \leq j \leq m} \left\{ \text{re}(\lambda_j) + \frac{(\text{imag}(\lambda_j))^2}{\text{re}(\lambda_j)} \right\}. \quad (3.22)$$

According to the Cayley-Hamilton theorem, for \mathbf{p}_{N+1} to be small, some elements of \mathbf{x} should be close to the eigenvalues of T . Therefore, we should include numbers such as

$$\{|\text{re}(\lambda_1) + \text{imag}(\lambda_1)|, |\text{re}(\lambda_2) + \text{imag}(\lambda_2)|, \dots, |\text{re}(\lambda_m) + \text{imag}(\lambda_m)|\}$$

or

$$\{|\text{re}(\lambda_1)|, |\text{re}(\lambda_2)|, \dots, |\text{re}(\lambda_m)|\}$$

in \mathbf{x} .

Based on Theorem 3.3, Theorem 3.4, and the above observations, we propose an algorithm that produces a Coxian distribution as a satisfactory approximation of an *ME*-distribution.

Coxian Approximation of ME -distribution (CAMED) Algorithm We consider an ME -representation (α, T, \mathbf{u}) satisfying conditions given in Theorem 3.3 and Theorem 3.4.

Step 1: Find the spectrum $\{-\lambda_1, -\lambda_2, \dots, -\lambda_m\}$ of T and arrange the eigenvalues such that $|\operatorname{re}(\lambda_1) + \operatorname{imag}(\lambda_1)| \geq |\operatorname{re}(\lambda_2) + \operatorname{imag}(\lambda_2)| \geq \dots \geq |\operatorname{re}(\lambda_m) + \operatorname{imag}(\lambda_m)|$. Denote by ε^* a small positive number. Let $N = m$ and choose using equation (3.22).

Step 2: Let $\mathbf{x} = (\lambda, \dots, \lambda, |\operatorname{re}(\lambda_1) + \operatorname{imag}(\lambda_1)|, |\operatorname{re}(\lambda_2) + \operatorname{imag}(\lambda_2)|, \dots, |\operatorname{re}(\lambda_m) + \operatorname{imag}(\lambda_m)|)$, where λ is repeated in \mathbf{x} $N - m$ times.

Step 3: Use the spectral polynomial algorithm to compute the matrix P and the vector \mathbf{p}_{N+1} . Compute $\varepsilon_{\max} = \max_{1 \leq i \leq m} \{ |(\mathbf{p}_{N+1})_i| \}$.

Step 4: If $\max \varepsilon_{\max} \leq \varepsilon^*$, go to Step 5. Otherwise, set $N = N + 1$ and go back to Step 2.

Step 5: Calculate αP and construct $S(\mathbf{x})$. Find the greatest $N_0 \leq N$ such that $F_{(N_0, N)}^{(1)}(t)$ is positive on the positive real numbers. Construct $\beta(N_0, N)$ from P . If $(\beta(N_0, N)/(\beta(N_0, N)\mathbf{e}), S(\mathbf{x}), \mathbf{e})$ represents a Coxian distribution that is a satisfactory approximation of the ME -distribution (α, T, \mathbf{u}) , go to Step 6. Otherwise, reset ε^* to be $\varepsilon^*/2$, go back to Step 2.

Step 6: If $\beta(N_0, N)$ is nonnegative, then $(\beta(N_0, N)/(\beta(N_0, N)\mathbf{e}), S(\mathbf{x}))$ is a desired solution. Otherwise, use the algorithm developed in HE and Zhang [16] to find an ordered Coxian representation from $(\beta(N_0, N)/(\beta(N_0, N)\mathbf{e}), S(\mathbf{x}), \mathbf{e})$.

Note that ε^* used in the above algorithm does not measure the difference between $F_{(\alpha, T, \mathbf{u})}(t)$ and $F_{(\alpha P, S(\mathbf{x}), \mathbf{e})}(t)$ directly. Instead, ε^* measures how small the vector \mathbf{p}_{N+1} is, which ensures a small difference between $F_{(\alpha, T, \mathbf{u})}(t)$ and the Coxian approximation.

4 Approximating PH -distributions by Coxian Distributions

In this section, we consider Coxian approximations of PH -distributions with a PH -representation (α, T) . First, we present some results directly obtained from Theorem 3.3 for this special case.

Proposition 4.1. *Assume that the matrix $T = (t_{i,j})$ is a PH -generator. We choose $\mathbf{x} = (\lambda, \dots, \lambda, y_1, \dots, y_{L-1}, y_L)$ of order N , where L is a positive integer, is repeated in \mathbf{x} $N - L$ times, and $\lambda \geq y_1 \geq y_2 \geq \dots \geq y_{L-1} \geq y_L > 0$. For any given positive ε^* , there exists N^* such that, for $N \geq N^*$ and λ large enough, we have $|\varepsilon_1(t) + \varepsilon_2(t)| < \varepsilon^*$ for $t \geq 0$. If every element of \mathbf{x} is greater than $\max\{-t_{1,1}, -t_{2,2}, \dots, -t_{m,m}\}$, the matrix P is nonnegative for all $N \geq L$. Any PH -distribution with a PH -representation (α, T) can be approximated with a Coxian distribution with Coxian representation $(\alpha P/(\alpha P\mathbf{e}), S(\mathbf{x}))$. \square*

Proposition 4.1 implies that, if T is a PH -generator, $(\alpha P/(\alpha P\mathbf{e}), S(\mathbf{x}), \mathbf{e})$ can be an ordered

Coxian representation that represents an approximation of a PH -distribution with a PH -representation $(\boldsymbol{\alpha}, T)$. However, to ensure that $(\boldsymbol{\alpha}P/(\boldsymbol{\alpha}P\mathbf{e}), S(\mathbf{x}), \mathbf{e})$ is an ordered Coxian representation, we choose \mathbf{x} such that all elements of \mathbf{x} are greater than $\max\{-t_{1,1}, -t_{2,2}, \dots, -t_{m,m}\}$. One consequence of such a selection of \mathbf{x} is that the number of phases (i.e., the integer N) may have to be large for a satisfactory approximation. To reduce the numbers of phases in satisfactory Coxian approximations, we exploit the structure of the PH -generator T for a better choice of \mathbf{x} .

It is readily seen that the PH -generator T is an M -matrix (Berman and Plemmons [5]). Recall that $\{-\lambda_1, -\lambda_2, \dots, -\lambda_m\}$ is the spectrum of T . Assume that $-\lambda_m$ is the Perron-Frobenius eigenvalue of T (i.e., the eigenvalue of T with the largest real part).

Theorem 4.2. *Consider a PH -generator T . We choose $\mathbf{x} = (\lambda, \dots, \lambda, y_1, \dots, y_{L-1}, y_L)$ of order N , where λ is large enough, $\lambda > y_1 \geq y_2 \geq \dots \geq y_{L-1} \geq y_L > 0$, and $y_{L-1} > \lambda_m$. If N is large enough, then the matrix P is nonnegative. Consequently, if $\boldsymbol{\alpha}$ is nonnegative, $(\boldsymbol{\alpha}P/(\boldsymbol{\alpha}P\mathbf{e}), S(\mathbf{x}))$ is an ordered Coxian representation that represents an approximation of the PH -distribution $(\boldsymbol{\alpha}, T)$.*

Proof. A complete proof of Theorem 4.2 is tedious and is presented in the Appendix. To help readers understand the proof in the Appendix, we first prove the theorem by assuming that T is irreducible. We choose

$$\lambda > \max \left\{ \max_{1 \leq i \leq m} \{-t_{i,i}\}, \max_{1 \leq j \leq m} \left\{ \operatorname{re}(\lambda_j) + \frac{(\operatorname{imag}(\lambda_j))^2}{\operatorname{re}(\lambda_j)} \right\} \right\}. \quad (4.1)$$

Then the matrix $\lambda I + T$ is nonnegative, irreducible, and aperiodic (see Berman and Plemmons [5], Minc [20] and Seneta [32] for more about nonnegative matrices and M -matrices). It is readily seen that $\lambda - \lambda_m$ is the Perron-Frobenius eigenvalue of $\lambda I + T$ with algebraic multiplicity being one (i.e., the eigenvalue with the largest modulus). Let \mathbf{u} and \mathbf{v} be the left and right eigenvectors corresponding to the eigenvalue $\lambda - \lambda_m$, respectively. Since $\lambda I + T$ is irreducible, by the Perron-Frobenius theory, the vectors \mathbf{u} and \mathbf{v} can be chosen to be positive (componentwise) and $\mathbf{u}\mathbf{v} = 1$ and $\mathbf{u}\mathbf{e} = 1$. It is well known that the spectral radius of $\lambda I + T$ equals its Perron-Frobenius eigenvalue $\lambda - \lambda_m$ and

$$(\lambda I + T)^n = (\lambda - \lambda_m)^n \mathbf{v}\mathbf{u} + o\left((\lambda - \lambda_m)^n\right) \mathbf{v}\mathbf{u}. \quad (4.2)$$

Post-multiplying by $-T\mathbf{e}$ on both sides of equation (4.2), yields

$$\begin{aligned} (\lambda I + T)^n (-T\mathbf{e}) &= (-\mathbf{u}T\mathbf{e})(\lambda - \lambda_m)^n \mathbf{v} + o\left((\lambda - \lambda_m)^n\right) \mathbf{v} \\ &= \lambda_m (\lambda - \lambda_m)^n \mathbf{v} + o\left((\lambda - \lambda_m)^n\right) \mathbf{v}. \end{aligned} \quad (4.3)$$

Note that we use $\mathbf{u}T = -\lambda_m \mathbf{u}$ and $\mathbf{u}\mathbf{e} = 1$ in equation (4.3). By equation (4.3), we have

$$(y_1 I + T)(\lambda I + T)^n(-T\mathbf{e}) = (y_1 - \lambda_m)\lambda_m(\lambda - \lambda_m)^n \mathbf{v} + \left((\lambda - \lambda_m)^n \right) \mathbf{v}. \quad (4.4)$$

Using equation (4.4) and induction, it can be shown, for $1 \leq j \leq L - 1$,

$$\begin{aligned} & (y_j I + T) \cdots (y_1 I + T)(\lambda I + T)^n(-T\mathbf{e}) \\ &= (y_j - \lambda_m) \cdots (y_1 - \lambda_m)\lambda_m(\lambda - \lambda_m)^n \mathbf{v} + \left((\lambda - \lambda_m)^n \right) \mathbf{v}. \end{aligned} \quad (4.5)$$

By the assumption that $y_j > \lambda_m$, for $1 \leq j \leq L - 1$, the expression in equation (4.5) becomes nonnegative if n is large enough. Since the columns of the matrix P are defined as (see equation (2.3))

$$\begin{aligned} \mathbf{p}_1 &= \frac{-T\mathbf{e}}{\lambda}; \\ \mathbf{p}_j &= \frac{1}{\lambda^j}(\lambda I + T)^{j-1}(-T\mathbf{e}), \quad 2 \leq j \leq N - L; \\ \mathbf{p}_{N-L+1} &= \frac{1}{y_1 \lambda^{N-L}}(\lambda I + T)^{N-L}(-T\mathbf{e}); \\ \mathbf{p}_{N-L+j} &= \frac{1}{y_j \cdots y_1 \lambda^{N-L}}(y_{j-1} I + T) \cdots (y_1 I + T)(\lambda I + T)^{N-L}(-T\mathbf{e}), \quad 2 \leq j \leq L, \end{aligned} \quad (4.6)$$

equations (4.3), (4.5), and (4.6) imply that the matrix P is nonnegative if N is large enough. The rest of the results is obtained by Theorem 3.3. This completes the proof of Theorem 4.2. \square

Based on Theorem 4.2, we modify the CAMED algorithm for computing $S(\mathbf{x})$ and an ordered Coxian representation $(\boldsymbol{\alpha}P/(\boldsymbol{\alpha}P\mathbf{e}), S(\mathbf{x}))$ of an approximation of the PH -distribution $(\boldsymbol{\alpha}, T)$.

Coxian Approximation of PH -distribution (CAPHD) Algorithm We consider a PH -representation $(\boldsymbol{\alpha}, T)$.

Step 1, Step 2, and Step 3: They are the same as Steps 1, 2, and 3 in the CAMED algorithm given in Section 3. We choose \mathbf{x} according to Theorem 4.2.

Step 4: Calculate $p_{\min} = \min_{1 \leq i \leq m, 1 \leq j \leq N} \{(P)_{i,j}\}$. If $\varepsilon_{\max} \leq \varepsilon^*$ and $p_{\min} \geq 0$ (the matrix P becomes nonnegative), go to Step 5. Otherwise, set $N =: N + 1$ and go back to Step 2.

Step 5: Calculate $\boldsymbol{\alpha}P$ and construct $S(\mathbf{x})$, then $(\boldsymbol{\alpha}P/(\boldsymbol{\alpha}P\mathbf{e}), S(\mathbf{x}))$ is a Coxian representation that represents a Coxian distribution approximating the PH -distribution $(\boldsymbol{\alpha}, T)$.

It is possible that $\boldsymbol{\alpha}P$ is nonnegative even if $\boldsymbol{\alpha}$ is not nonnegative. Thus, Theorem 4.2 and the CAPHD algorithm can be applied to ME -distributions with an ME -representation $(\boldsymbol{\alpha}, T, \mathbf{e})$, where T is a PH -generator.

5 Coxian Representations of Coxian Distributions

Proposition 4.1 and Theorem 4.2 improve Theorem 3.3 in the sense that, in addition to a small \mathbf{p}_{N+1} , P can be nonnegative and $(\boldsymbol{\alpha}P/(\boldsymbol{\alpha}P\mathbf{e}), S(\mathbf{x}))$ is a Coxian representation of a Coxian distribution. Unfortunately, if T has complex eigenvalues, \mathbf{p}_{N+1} may never be zero no matter how large N is. On the other hand, if T is a PH -generator and has only real eigenvalues, the PH -representation $(\boldsymbol{\alpha}, T)$ represents a Coxian distribution and has Coxian representations (see Theorem 2.2). For this case, we extend Theorem 4.2 to find equivalent ordered Coxian representations for $(\boldsymbol{\alpha}, T)$. This gives an alternative proof to Theorem 2.2.

Theorem 5.1. *Assume that all eigenvalues $\{-\lambda_1, -\lambda_2, \dots, -\lambda_m\}$ of PH -generator T (counting multiplicities) are real and $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m$. We choose \mathbf{x} as*

$$\mathbf{x} = (\lambda, \dots, \lambda, \lambda_1, \dots, \lambda_{m-1}, \lambda_m), \quad (5.1)$$

where λ is repeated in \mathbf{x} $N - m$ times and $\lambda > \max\{\lambda_1, \max\{-t_{1,1}, -t_{2,2}, \dots, -t_{m,m}\}\}$. Then we have $TP = PS(\mathbf{x})$ and $P\mathbf{e} = \mathbf{e}$ for $N \geq m$. If N is large enough, P becomes nonnegative, i.e., $S(\mathbf{x})$ PH -majorizes T . Consequently, for any probability vector $\boldsymbol{\alpha}$, $(\boldsymbol{\alpha}P, S(\mathbf{x}))$ is an equivalent ordered Coxian representation of the PH -representation $(\boldsymbol{\alpha}, T)$.

Proof. A complete proof of this theorem is given in the Appendix. In this proof, similar to that of Theorem 4.2, we assume that T is irreducible. Since $\{\lambda_1, \lambda_2, \dots, \lambda_m\}$ is a subset of \mathbf{x} (see equation (5.1)), by the Cayley-Hamilton theorem, $\mathbf{p}_{N+1} = 0$ for any $N \geq m$. Then we have $TP = PS(\mathbf{x})$, $P\mathbf{e} = \mathbf{e}$, and $P\mathbf{e} = 1$. Thus, $(\boldsymbol{\alpha}, T)$ and $(\boldsymbol{\alpha}P, S(\mathbf{x}), \mathbf{e})$ represent the same distribution for any $N \geq m$. Next, we show the nonnegativity of the matrix P for large N . For our choice of \mathbf{x} , equation (4.5) becomes, for $1 \leq j \leq m - 1$,

$$\begin{aligned} & (\lambda_j I + T) \cdots (\lambda_1 I + T) (\lambda I + T)^n (-T\mathbf{e}) \\ &= (\lambda_j - \lambda_m) \cdots (\lambda_1 - \lambda_m) \lambda_m (\lambda - \lambda_m)^n \mathbf{v} + o\left((\lambda - \lambda_m)^n\right) \mathbf{v}. \end{aligned} \quad (5.2)$$

Since the matrix T is irreducible, we have $\lambda_{m-1} > \lambda_m$. Therefore, the expression in equation (5.2) becomes nonnegative if n is large enough. Combining equations (4.6) and (5.2), we conclude that the matrix P is nonnegative if N is large enough.

For large enough N , P becomes nonnegative. Since $TP = PS(\mathbf{x})$, and $P\mathbf{e} = \mathbf{e}$, according to Theorem 2 in O'Kinneide [26], $S(\mathbf{x})$ PH -majorizes T .

Finally, if P is nonnegative, $(\boldsymbol{\alpha}P, S(\mathbf{x}))$ is an ordered Coxian representation representing the same probability distribution as that of the PH -representation $(\boldsymbol{\alpha}, T)$. This completes the proof of Theorem 5.1. \square

Note that for some Coxian distributions $(\boldsymbol{\alpha}, T, \mathbf{e})$ for which $\boldsymbol{\alpha}$ is not nonnegative, it is possible that $(\boldsymbol{\alpha}P, S(\mathbf{x}))$ is an ordered Coxian representation. Thus, the application of

Theorem 5.1 is not limited to PH -representations (see Example 6.3). Based on Theorem 5.1, we modify the CAPHD algorithm to compute equivalent ordered Coxian representations for Coxian distributions.

Coxian Representation of Coxian Distribution (CRCD) Algorithm Assume that T is a PH -generator with spectrum $\{-\lambda_1, -\lambda_2, \dots, -\lambda_m\}$ ordered as $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m$. Let $\lambda > \max\{\lambda_1, \max_{1 \leq i \leq m} \{-t_{i,i}\}\}$ and $N = m$.

Step 1: Use equation (5.1) to define \mathbf{x} .

Step 2: Use the spectral polynomial algorithm to compute the matrix P . Let $p_{\min} = \min_{1 \leq i \leq m, 1 \leq j \leq N} \{(P)_{i,j}\}$.

Step 3: If $p_{\min} \geq 0$, go to Step 4; Otherwise, set $N = N + 1$ and go back to Step 1.

Step 4: Calculate αP and construct $S(\mathbf{x})$. Then $(\alpha P, S(\mathbf{x}), \mathbf{e})$ is an equivalent bi-diagonal ME -representation of (α, T, \mathbf{e}) . If P is nonnegative, $(\alpha P, S(\mathbf{x}))$ is an ordered Coxian representation.

6 Numerical Examples

In Sections 3, 4, and 5, we developed an algorithm for computing Coxian representations of Coxian distributions as approximations of ME -distributions or PH -distributions. The algorithm also finds equivalent Coxian representations for PH -representations that represent Coxian distributions. In this section, we present three numerical examples to show the accuracy of such approximations and the number of phases needed for a satisfactory approximation.

Example 6.1 Consider Example 2.1 in Asmussen and Bladt [4]. In that example, an ME -representation (α, T, \mathbf{u}) is given as $\alpha = (1, 0, 0)$, $\mathbf{u} = \mathbf{e}$, and T in equation (6.1). The eigenvalues of T are $\{-1, -1 + 2\pi\sqrt{-1}, -1 - 2\pi\sqrt{-1}\}$. For brevity, we only look at $(\alpha P, S(\mathbf{x}), \mathbf{e})$ and $(\beta(N_0, N), S(\mathbf{x}), \mathbf{e})$.

$$T = \begin{pmatrix} 0 & -1 - 4\pi^2 & 1 + 4\pi^2 \\ 3 & 2 & -6 \\ 2 & 2 & -5 \end{pmatrix}. \quad (6.1)$$

By using the CAMED algorithm, for $\lambda = 40$ (chosen by using equation (3.22)), we obtained approximations for $N = 10, 30, 100, 200, 300$, and 400 . The density functions of the original matrix-exponential distribution and all the approximations are plotted in Figure 6.1 ($N=10, 30$, and 100 in (a) and $N=200, 300$, and 400 in (b)). The corresponding errors of the approximations are given in Table 6.1.

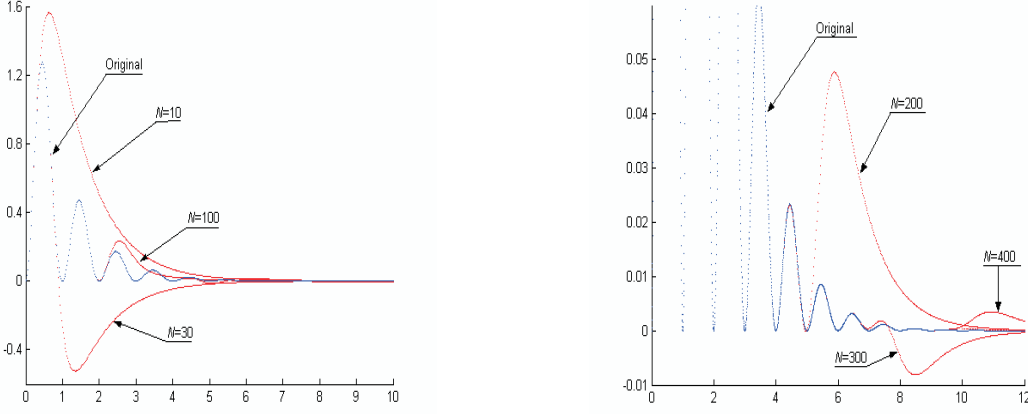


Figure 6.1 The functions $-\alpha \exp\{Tt\}T\mathbf{u}$ and $-\alpha P \exp\{S(\mathbf{x})t\}S(\mathbf{x})\mathbf{e}$

Figure 6.1 shows that if $N = 10$, the function $-\alpha P \exp\{S(\mathbf{x})t\}S(\mathbf{x})\mathbf{e}$ is close to the original density function $-\alpha \exp\{Tt\}T\mathbf{u}$ for $0 \leq t \leq 0.5$; if $N = 30$, the approximation is close to the original density function for $0 \leq t \leq 1$; if $N = 100$, the approximation is close to the original density function for $0 \leq t \leq 2.5$; if $N = 200$, the approximation is close to the original density function for $0 \leq t \leq 5$; if $N = 300$, the approximation is close to the original density function for $0 \leq t \leq 7.5$; and finally, if $N = 400$, the approximation and the original density function are almost the same. It is clear that if we put more phases into the ordered Coxian generator $S(\mathbf{x})$, a better approximation can be obtained.

Table 6.1 β_{\min} and ε_{\max} of the six approximations in Example 6.1

	$N = 10$	$N = 30$	$N = 100$	$N = 200$	$N = 300$	$N = 400$
β_{\min}	0	-0.8937	-0.0062	-0.0615	-0.0208	-0.0062
ε_{\max}	4.8589	3.8001	3.3621	0.9357	0.2414	0.0567

Note that, in Table 6.1, β_{\min} is defined as $\beta_{\min} = \min_{1 \leq i \leq N} \{(\alpha P)_i\}$, which is real since \mathbf{x} (given in the CAMED algorithm given in Section 3) is real, and ε_{\max} is defined in the CAMED algorithm. In Table 6.1, β_{\min} and ε_{\max} are presented for the six approximations. Table 6.1 shows that all the representations of the six approximations are not Coxian representations (since $\beta_{\min} < 0$), except for $N = 10$. Furthermore, some of the approximations to the probability distribution functions are monotone functions (e.g., $N = 10$ and 100), while others are not (e.g., $N = 30, 200, 300$, and 400) (see Figure 6.1).

For cases with $N = 10$ and 100, after normalizing the vector β , the corresponding functions are Coxian distribution functions that approximate the original distribution function. For cases with $N = 30, 200, 300$, and 400, by using Theorem 3.4, we can find Coxian distributions that approximate the original distributions. When $N=30$, all elements of β are positive except β_{30} . Then we choose $N_0 = 29$ and the function $F_{(29,30)}(t)$ is monotone. The Coxian representation $(\beta(29, 30)/(\beta(29, 30)\mathbf{e}), S(\mathbf{x}))$ represents a Coxian distribution. When $N = 200$, we choose $N_0 = 197$ and the function $F_{(197,200)}(t)$ is monotone. Then

the ME -representation $(\beta(197, 200)/(\beta(197, 200)e), S(\mathbf{x}), e)$ represents a Coxian distribution. When $N = 300$, we choose $N_0 = 299$ and the function $F_{(299,300)}(t)$ is monotone. Then $(\beta(299, 300)/(\beta(299, 300)e), S(\mathbf{x}), e)$ represents a Coxian distribution. When $N = 400$, we choose $N_0 = 398$ and the function $F_{(398,400)}(t)$ is monotone. Then $(\beta(398, 400)/(\beta(398, 400)e), S(\mathbf{x}), e)$ represents a Coxian distribution. For the last three cases, an equivalent Coxian representation can be obtained by using the algorithm developed in HE and Zhang [16].

It is reasonable to say that, by adding more phases, the corresponding Coxian approximation is better and it is more likely that the resulting representation is a Coxian representation. Unfortunately, when the approximation becomes satisfactory, the corresponding representation may have too many phases (compared to that of the original ME -representation). Our numerical experiments demonstrate that if T is a PH -generator, the number of phases needed for a good Coxian approximation does not have to be large when compared to the number of phases of the original PH -representation. One of such examples is shown next.

Example 6.2 We consider T of order 10 given as

$$T = \begin{pmatrix} -4 & 0 & 0 & 0.1 & 0 & 0 & 0 & 0 & 0.1 & 0.1 \\ 4 & -4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 4 & -4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.1 & 0 & 3.5 & -4 & 0 & 0 & 0 & 0 & 0 & 0.1 \\ 0.1 & 0 & 0 & 3.8 & -4 & 0 & 0 & 0 & 0 & 0.1 \\ 0 & 0 & 0 & 0.9 & 0 & -1 & 0 & 0 & 0.1 & 0 \\ 0 & 0.1 & 0 & 0 & 0 & 0.9 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0.1 & 0 & 0 & 0 & 0 & 0 & 0 & 0.9 & -1 \end{pmatrix}. \quad (6.2)$$

The eigenvalues of T are $\{-0.3399, -2.4415, -4.000, -5.5392, -0.8539 - 0.6100\sqrt{-1}, -0.8539 + 0.6100\sqrt{-1}, -1.4832 - 0.2856\sqrt{-1}, -1.4832 + 0.2856\sqrt{-1}, -4.0027 - 1.5382\sqrt{-1}, -4.0027 + 1.5382\sqrt{-1}\}$. We choose $\mathbf{x} = (\lambda, \dots, \lambda, y_1, \dots, y_{10})$, where $\lambda = 6$ and $y_i = \operatorname{re}(\lambda_i)$, $1 \leq i \leq 10$. For $N=13, 15, 20, 30, 100$, and 200 , by using the spectral polynomial algorithm and the CAPHD algorithm, we calculate the corresponding matrix P and vector \mathbf{p}_{N+1} . We consider a PH -distribution $(\boldsymbol{\alpha}, T)$ with $\boldsymbol{\alpha} = (0, 0, 0, 0, 0.2, 0, 0, 0, 0, 0.8)$. The corresponding p_{min} , β_{min} , and ε_{max} are presented in the following table.

Table 6.2 p_{min}, β_{min} and ε_{max} of the six approximations in Example 6.2

	$N = 13$	$N = 15$	$N = 20$	$N = 30$	$N = 100$	$N = 200$
p_{min}	-8.1562	-3.3428	0	0	0	0
β_{min}	0	-0.1421	0	0	0	0
ε_{max}	34.3317	10.9335	0.1779	0.0169	10^{-6}	9.9×10^{-11}

Table 6.3 shows that if \mathbf{x} is chosen properly, $(\alpha P, S(\mathbf{x}), \mathbf{e})$ is an ordered Coxian representation that may represent a satisfactory approximation for a moderate size $N(\approx 20)$. Note that, since some of the eigenvalues of T are not real, then $(\alpha P, S(\mathbf{x}), \mathbf{e})$ always represents an approximation. Table 6.2 also shows that the elements of P become nonnegative faster than the elements of \mathbf{p}_{N+1} become small. The reason is that more than half of the eigenvalues of T are not real. Thus, we need to increase N to achieve a small \mathbf{p}_{N+1} .

Figure 6.2 plots the density functions of (α, T) and its approximations with $N=13, 15,$ and 20 . If N is 30 or larger, the approximations match the original density function extremely well. That indicates that the approximation can be very good, even if the corresponding ε_{\max} is not small (see Table 6.2). Since α is nonnegative, the approximations are Coxian distributions, if N is 20 or greater. For $N \geq 20$, the matrix P becomes nonnegative. Then $(\alpha P/(\alpha P \mathbf{e}), S(\mathbf{x}))$ is a Coxian representation that represents a Coxian distribution.

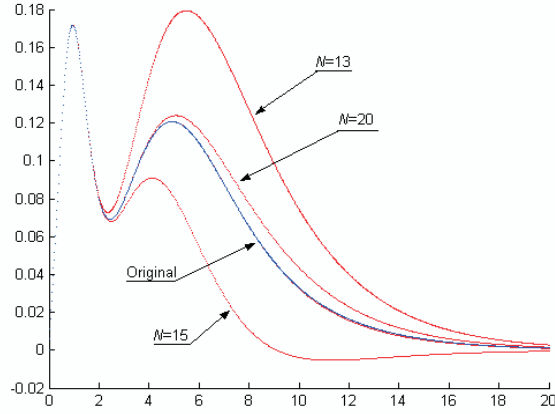


Figure 6.2 Density functions for $N = 13, 15,$ and 20 and the original

In this example, we included $\{\text{re}(\lambda_i), 1 \leq i \leq 10\}$ in the definition of \mathbf{x} given in Theorem 4.2, which is better than the other choice $\{|\text{re}(\lambda_i) + \text{imag}(\lambda_i)|, 1 \leq i \leq 10\}$ with respect to the number of phases N needed for a satisfactory approximation. However, according to our numerical experiments with a large number of ME -distributions for $m = 10$ to 50 , the choice $\{|\text{re}(\lambda_i) + \text{imag}(\lambda_i)|, 1 \leq i \leq m\}$ is better than $\{\text{re}(\lambda_i), 1 \leq i \leq m\}$ for some cases.

If the PH -generator has only real eigenvalues, then it is possible to find an equivalent ordered Coxian representation, instead of approximations. As indicated by the following example, the size of the ordered Coxian representation does not have to be large when compared to that of the original PH -generator.

Example 6.3 Consider a PH -generator T defined as

$$T = \begin{pmatrix} -10 & 0 & 1 & 0 & 0.5 \\ 5 & -5 & 0 & 0 & 0 \\ 1 & 4 & -5 & 0 & 0 \\ 0 & 0 & 0.5 & -0.5 & 0 \\ 0 & 0 & 0 & 0.5 & -0.5 \end{pmatrix}. \quad (6.3)$$

Eigenvalues of T are $\{-8.9960, -7.7986, -3.1946, -0.6408, -0.3699\}$. Using the Coxian representation algorithm, we compute Coxian representations for a PH -distribution with a ME -representation $(\boldsymbol{\alpha}, T, \mathbf{e})$, where $\boldsymbol{\alpha} = (0, 1, 0, -0.1, 0.1)$. We choose $\mathbf{x} = (10, \dots, 10, 8.9960, 7.7986, 3.1946, 0.6408, 0.3699)$. Results related to representations $(\boldsymbol{\alpha}P, S(\mathbf{x}), \mathbf{e})$ are given in Table 6.3.

Table 6.3 p_{\min} and β_{\min} of four representations for Example 6.3

	$N = 5$	$N = 8$	$N = 11$	$N \geq 18$
p_{\min}	-0.1216	-0.0052	0	0
β_{\min}	-0.0798	-0.0519	-0.0304	0

Note that, for this example, $\varepsilon_{\max} = 0$ for all $N \geq 5$. Table 6.3 shows that, if $N \geq 11$, the matrix P is nonnegative and, if $N \geq 18$, the vector $\boldsymbol{\alpha}P$ becomes nonnegative, i.e., $(\boldsymbol{\alpha}P, S(\mathbf{x}))$ becomes an equivalent ordered Coxian representation of $(\boldsymbol{\alpha}, T, \mathbf{e})$.

Suppose that $\boldsymbol{\alpha} = (0, 1, -0.5, 0, 0.5)$. If we still choose $\mathbf{x} = (10, \dots, 10, 8.9960, 7.7986, 3.1946, 0.6408, 0.3699)$, $\boldsymbol{\alpha}P$ is not nonnegative for all $N \geq 5$. In fact, for $N \geq 11$, we always have $\beta_{\min} = -0.0061$. However, if we choose $\mathbf{x} = (30, \dots, 30, 8.9960, 7.7986, 3.1946, 0.6408, 0.3699)$, then $\boldsymbol{\alpha}P$ becomes nonnegative for $N \geq 25$. This example indicates that the choice of \mathbf{x} is an important factor for finding an ordered Coxian representation successfully, which is an interesting issue for future research.

7 Conclusions and Discussion

In this paper, based on the spectral polynomial algorithm, we developed an algorithm for computing 1) exact Coxian representations for PH -representations of Coxian distributions; 2) Coxian representations for approximate Coxian distributions of PH -distributions; and 3) Coxian representations for approximate Coxian distributions of ME -distributions. The algorithm can be implemented in a straightforward manner. The form of the Coxian representations of the approximations is simple and convenient for theoretical and numerical studies of stochastic models.

A direction for further research is related to the number of phases in the Coxian representations of the approximations. This issue is sensitive to the choice of vector \mathbf{x} (see the

CAMED algorithm in Section 3), where it is not clear how to choose \mathbf{x} in an optimal way in a practical case. For the approximations of *ME*-distributions and *PH*-distributions, an interesting problem is to reduce the number of phases in their Coxian representations. For the Coxian representations of Coxian distributions, an open question is the number of phases needed by their Coxian representations.

A Appendix: Proofs of Theorem 4.2 and Theorem 5.1

To give complete proofs of Theorem 4.2 and Theorem 5.1, we need to show that the matrix P is nonnegative. Our proofs are based on the extension of the Perron-Frobenius theorem for reducible nonnegative matrices developed in Gantmacher [13] and Rothblum [31], and the Jordan canonical form of matrices. We use the notation introduced in Berman and Plemmons [5] (see Section 2.3 in [5]). First, we need to introduce a number of concepts related to nonnegative matrices.

Assume that A is a nonnegative matrix of order m . Denote by $A^n = (a_{i,j}^{(n)})$. We say that phase j is accessible from i if $a_{i,j}^{(n)} > 0$ for some n . A class of phases consists of phases for which any two phases are accessible from each other. Class φ is accessible from class ϕ , if any phase in ϕ has access to any phase in φ . A class is final if it has access to no other class. Denote by $\rho(A)$ the Perron-Frobenius eigenvalue of the nonnegative matrix A , which is the spectral radius of A since A is nonnegative. Class φ (a subset of $\{1, 2, \dots, m\}$) is basic if $\rho(A[\varphi]) = \rho(A)$, where $A[\varphi]$ is the submatrix of A based on the phases in φ , and non-basic if $\rho(A[\varphi]) < \rho(A)$.

A collection of basic classes $\{\varphi_1, \varphi_2, \dots, \varphi_n\}$ is a chain from φ_1 to φ_n , if φ_k has access to φ_{k+1} , $k = 1, 2, \dots, n-1$. The length of a chain is the number of basic classes it contains. A class φ has access to a class ϕ in j steps if j is the length of the longest chain from φ to ϕ . The height of a class φ is the length of the longest chain of classes that terminates in φ .

The degree $v(A)$ of $\rho(A)$ is the dimension of the largest Jordan block corresponding to $\rho(A)$. The null space $\mathbf{N}((\rho(A)I - A)^{v(A)})$ is called *the algebraic eigenspace* of A and its elements are called generalized eigenvectors. The following theorem is given in Berman and Plemmons [5].

Theorem A.1. (Theorem 3.20, Berman and Plemmons [5]) *Assume that A has spectral radius $\rho(A)$ and n basic classes $\varphi_1, \varphi_2, \dots$, and φ_n . Then the algebraic eigenspace of A contains nonnegative vectors, $\mathbf{u}^{(1)}, \mathbf{u}^{(2)}, \dots$, and $\mathbf{u}^{(n)}$, such that the i -th element $(\mathbf{u}^{(j)})_i$ of $\mathbf{u}^{(j)}$ is positive if and only if the phase i has access to the class j , and any such collection is a basis of the algebraic eigenspace of A .*

It is also known that the degree of an irreducible matrix is one. An immediate consequence of Theorem A.1 is that any Jordan chain of $\rho(A)$ corresponds to a set of basic classes

(Lancaster and Tismenetsky [17]). The length of the Jordan chain equals the number of basic classes involved. The vectors of the Jordan chain can be chosen to be nonnegative with positive elements corresponding to the related basic classes.

Now we consider a Jordan chain $\{\mathbf{u}^{(1)}, \mathbf{u}^{(2)}, \dots, \mathbf{u}^{(K)}\}$ of $\rho(A)$, i.e.,

$$(\rho(A)I - A)^k \mathbf{u}^{(k)} = 0, \quad 1 \leq k \leq K. \quad (\text{A.1})$$

Or equivalently, we have $A\mathbf{u}^{(1)} = \rho(A)\mathbf{u}^{(1)}$ and $A\mathbf{u}^{(k)} = \rho(A)\mathbf{u}^{(k)} + \mathbf{u}^{(k-1)}, 2 \leq k \leq K$. We denote the corresponding chain of basic classes as $\varphi_1, \varphi_2, \dots$, and φ_K . The length of the chain is K and the chain terminates at the class φ_K . Since any phase that has access to the basic class φ_j also has access to the basic class φ_{j+1} , we have $(\mathbf{u}^{(j+1)})_i > 0$ if $(\mathbf{u}^{(j)})_i > 0$.

We divide the phases $\{1, 2, \dots, m\}$ into two subsets: the phases that have access to φ_K and the phases that have no access to φ_K . By reordering the phases, A can be rewritten as

$$A = \begin{pmatrix} A_0 & 0 \\ A_{1,0} & A_1 \end{pmatrix}, \quad (\text{A.2})$$

where phases corresponding to A_1 have access to φ_K and phases corresponding to A_0 have no access to φ_K . The matrix A_1 contains not only all the submatrices corresponding to the basic classes $\varphi_1, \varphi_2, \dots$, and φ_K , but also all phases that have access to φ_K . Some basic classes may be in the matrix A_0 , but such basic classes have no access to φ_K and vice versa. According to Theorem A.1, we have

$$\mathbf{u}^{(k)} = \begin{pmatrix} 0 \\ \mathbf{u}_1^{(k)} \end{pmatrix}, \quad 1 \leq k \leq K, \quad (\text{A.3})$$

and the vector $\mathbf{u}_1^{(K)}$ is positive since φ_K is accessible from all phases related to A_1 . By equation (A.1), the vectors $\{\mathbf{u}_1^{(1)}, \mathbf{u}_1^{(2)}, \dots, \mathbf{u}_1^{(K)}\}$ satisfy $A_1 \mathbf{u}_1^{(1)} = \rho(A) \mathbf{u}_1^{(1)}$ and $A_1 \mathbf{u}_1^{(k)} = \rho(A) \mathbf{u}_1^{(k)} + \mathbf{u}_1^{(k-1)}, 2 \leq k \leq K$.

By Theorem A.1, we can extend $\{\mathbf{u}^{(1)}, \mathbf{u}^{(2)}, \dots, \mathbf{u}^{(K)}\}$ to a complete set of Jordan chains $\{\mathbf{u}^{(1)}, \mathbf{u}^{(2)}, \dots, \mathbf{u}^{(K)}, \mathbf{u}^{(K+1)}, \dots, \mathbf{u}^{(m)}\}$ of the matrix A . Let $U = (\mathbf{u}^{(1)}, \mathbf{u}^{(2)}, \dots, \mathbf{u}^{(m)})$. By definition, U is invertible. Denote by V the inverse matrix of U and by $\{\mathbf{v}^{(1)}, \mathbf{v}^{(2)}, \dots, \mathbf{v}^{(m)}\}$ the 1-st, 2-nd, \dots , m -th rows of V , respectively. By definition of the Jordan chain, it is clear that $\mathbf{v}^{(K)} \mathbf{u}^{(K)} = 1$ and $\mathbf{v}^{(K)} A = \rho(A) \mathbf{v}^{(K)}$. Since $\mathbf{u}^{(K)}$ is chosen to be nonnegative, the vector $\mathbf{v}^{(K)}$ is nonnegative and $\mathbf{v}^{(K)} \mathbf{e}$ is positive.

By the Jordan canonical form and routine calculations, we obtain

$$A^n = \sum_{k=1}^{\min\{K, n+1\}} \mathbf{u}^{(k)} \left(\sum_{j=k}^{\min\{n+1, K\}} \frac{n! (\rho(A))^{n+k-j}}{(n+k-j)!} \mathbf{v}^{(j)} \right) + R(n), \quad (\text{A.4})$$

where $R(n)$ includes all the remaining items in the expansion of A^n .

Proof of Theorem 4.2

Let $A = \lambda I + T$, which is nonnegative for λ given by equation (4.1). We rewrite T into the following form according to that of A given in equation (A.2):

$$T = \begin{pmatrix} T_0 & 0 \\ T_{1,0} & T_1 \end{pmatrix}. \quad (\text{A.5})$$

The spectrum of A is $\{\lambda - \lambda_1, \dots, \lambda - \lambda_m\}$. In Section 4, we assumed that $-\lambda_m$ is the Perron-Frobenius eigenvalue of T . Then it is easy to see $\rho(A) = \lambda - \lambda_m$. We choose $n > K$. Post-multiplying by $-T\mathbf{e}$ on both sides of equation (A.4), we obtain, since $\mathbf{v}^{(K)}\mathbf{e}$ is positive,

$$\begin{aligned} & (\lambda I + T)^n (-T)\mathbf{e} \quad (\text{A.6}) \\ &= \sum_{k=1}^{\min\{K, n+1\}} \mathbf{u}^{(k)} \left(\sum_{j=k}^{\min\{n+1, K\}} \frac{n!(\rho(A))^{n+k-j}}{(n+k-j)!} \mathbf{v}^{(j)} (-T)\mathbf{e} \right) + R(n)(-T)\mathbf{e} \\ &= \sum_{k=1}^{\min\{K, n+1\}} \mathbf{u}^{(k)} \left(\frac{n!(\rho(A))^{n+k-K+1}}{(n+k-K)!} \mathbf{v}^{(K)}\mathbf{e} + \sum_{j=k}^{\min\{n, K-1\}} \frac{n!(\rho(A))^{n+k-j}}{(n+k-j)!} \mathbf{v}^{(j)} (-T)\mathbf{e} \right) \\ &\quad + R(n)(-T)\mathbf{e} \\ &= \sum_{k=1}^{\min\{K, n+1\}} \begin{pmatrix} 0 \\ \mathbf{u}_1^{(k)} \end{pmatrix} \left(\frac{n!(\rho(A))^{n+k-K+1}}{(n+k-K)!} \mathbf{v}^{(K)}\mathbf{e} + o\left(\frac{n!(\rho(A))^{n+k-K+1}}{(n+k-K)!}\right)\mathbf{e} \right) + R(n)(-T)\mathbf{e}. \end{aligned}$$

Note that, in equation (A.6), the notation “ o ” is respect to n . Since the phases of any Jordan chain associated with $\rho(A)$ has no access to the phases of any other Jordan chain associated with $\rho(A)$ (otherwise, they form a single Jordan chain), equations (A.2)-(A.6) imply

$$\begin{aligned} & (\lambda I + T)^n (-T)\mathbf{e} \quad (\text{A.7}) \\ &= \begin{pmatrix} (\lambda I + T_0)^n (-T_0)\mathbf{e} \\ \left(\sum_{k=1}^{\min\{K, n+1\}} \mathbf{u}_1^{(k)} \left(\frac{n!(\rho(A))^{n+k-K+1}}{(n+k-K)!} \mathbf{v}^{(K)}\mathbf{e} + o\left(\frac{n!(\rho(A))^{n+k-K+1}}{(n+k-K)!}\right)\mathbf{e} \right) + o\left(\left(\rho(A)\right)^n\right)\mathbf{e} \right) \end{pmatrix}. \end{aligned}$$

Since $-T\mathbf{e}$ is nonnegative, $(\lambda I + T)^n (-T)\mathbf{e}$ is nonnegative. Furthermore, since $y_1 > \lambda_m$, we have

$$\begin{aligned} & (y_1 I + T)(\lambda I + T)^n (-T)\mathbf{e} \quad (\text{A.8}) \\ &= \begin{pmatrix} (y_1 I + T_0)(\lambda I + T_0)^n (-T_0)\mathbf{e} \\ \sum_{k=1}^{\min\{K, n+1\}} (y_1 I + T_1) \mathbf{u}_1^{(k)} \left(\frac{n!(\rho(A))^{n+k-K+1}}{(n+k-K)!} \mathbf{v}^{(K)}\mathbf{e} + o\left(\frac{n!(\rho(A))^{n+k-K+1}}{(n+k-K)!}\right)\mathbf{e} \right) \end{pmatrix} \\ &\quad + \begin{pmatrix} 0 \\ o\left(\left(\rho(A)\right)^n\right)\mathbf{e} \end{pmatrix} \end{aligned}$$

$$\begin{aligned}
&= \left(\begin{array}{c} (y_1 I + T_0)(\lambda I + T_0)^n (-T_0) \mathbf{e} \\ \sum_{k=1}^{\min\{K, n+1\}} \left[(y_1 - \lambda_m) \mathbf{u}_1^{(k)} + \mathbf{u}_1^{(k-1)} \right] \left(\frac{n! (\rho(A))^{n+k-K+1}}{(n+k-K)!} \mathbf{v}^{(K)} \mathbf{e} + o\left(\frac{n! (\rho(A))^{n+k-K+1}}{(n+k-K)!} \right) \mathbf{e} \right) \end{array} \right) \\
&\quad + \left(\begin{array}{c} 0 \\ o\left((\rho(A))^n \right) \mathbf{e} \end{array} \right) \\
&= \left(\begin{array}{c} (y_1 I + T_0)(\lambda I + T_0)^n (-T_0) \mathbf{e} \\ \sum_{k=1}^{\min\{K, n+1\}} \mathbf{u}_1^{(K)} \left(\frac{n! (y_1 - \lambda_m) (\rho(A))^{n+k-K+1}}{(n+k-K)!} \mathbf{v}^{(K)} \mathbf{e} + o\left(\frac{n! (\rho(A))^{n+k-K+1}}{(n+k-K)!} \right) \mathbf{e} \right) \end{array} \right) \\
&\quad + \left(\begin{array}{c} 0 \\ o\left((\rho(A))^n \right) \mathbf{e} \end{array} \right).
\end{aligned}$$

By induction and equation (A.6) and $y_j > \lambda_m$, we obtain, for $j < L$,

$$\begin{aligned}
&(y_j I + T) \cdots (y_1 I + T)(\lambda I + T)^n (-T) \mathbf{e} \\
&= \left(\begin{array}{c} (y_j I + T_0) \cdots (y_1 I + T_0)(\lambda I + T_0)^n (-T_0) \mathbf{e} \\ \sum_{k=1}^{\min\{K, n+1\}} \mathbf{u}_1^{(K)} \left(\frac{n! (y_j - \lambda_m) \cdots (y_1 - \lambda_m) (\rho(A))^{n+k-K+1}}{(n+k-K)!} \mathbf{v}^{(K)} \mathbf{e} + o\left(\frac{n! (\rho(A))^{n+k-K+1}}{(n+k-K)!} \right) \mathbf{e} \right) \end{array} \right) \\
&\quad + \left(\begin{array}{c} 0 \\ o\left((\rho(A))^n \right) \mathbf{e} \end{array} \right) \\
&\equiv \left(\begin{array}{c} (y_j I + T_0) \cdots (y_1 I + T_0)(\lambda I + T_0)^n (-T_0) \mathbf{e} \\ \mathbf{w}(n) \end{array} \right) \\
&\quad + \left(\begin{array}{c} 0 \\ o\left((\rho(A))^n \right) \mathbf{e} \end{array} \right) \\
&= \left(\begin{array}{c} (y_j I + T_0) \cdots (y_1 I + T_0)(\lambda I + T_0)^n (-T_0) \mathbf{e} \\ \mathbf{w}(n) + o\left((\rho(A))^n \right) \mathbf{e} \end{array} \right). \tag{A.9}
\end{aligned}$$

Since $\mathbf{v}^{(K)} \mathbf{e}$ is positive and $\mathbf{u}_1^{(1)} + \cdots + \mathbf{u}_1^{(K)}$ is positive, we have $w_j(n) \geq O((\rho(A))^n)$ for all possible indices j for the vector $\mathbf{w}(n)$. Then we have $w_j(n) + o((\rho(A))^n) \geq O((\rho(A))^n)$. Thus, equation (A.9) indicates that the vector $(y_j I + T) \cdots (y_1 I + T)(\lambda I + T)^n (-T) \mathbf{e}$ becomes nonnegative if n is large enough and $(y_j I + T_0) \cdots (y_1 I + T_0)(\lambda I + T_0)^n (-T_0) \mathbf{e}$ is nonnegative. The non-negativity of the vector $(y_j I + T_0) \cdots (y_1 I + T_0)(\lambda I + T_0)^n (-T_0) \mathbf{e}$ can be proved by repeating the above proof on T_0 .

Equations (A.7) and (A.9) show the non-negativity of matrices $\{(\lambda I + T)^n (-T), (y_j I + T) \cdots (y_1 I + T)(\lambda I + T)^n (-T) \mathbf{e}, 1 \leq j \leq L - 1\}$ if n is large enough, which implies the non-negativity of the matrix P . This completes the proof of Theorem 4.2. \square

Proof of Theorem 5.1

The proof of Theorem 5.1 is similar to that of Theorem 4.2. For this case, we have $L = m$ and $y_j = \lambda_j, 1 \leq j \leq m$. The major difference between the proof given in Section 5 and this proof is that $\lambda_j = \lambda_m$ can be true for $j < m$. Because of that, in equation (A.9), it is now possible that $(\lambda_j - \lambda_m) \cdots (\lambda_1 - \lambda_m)$ becomes zero for $j < m$.

To show that the second part of the vector in equation (A.9) is nonnegative, we note that $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m > 0$. We consider two cases. First, if $\lambda_j > \lambda_m$, equation (A.9) shows that the second part of the vector is nonnegative if n is large enough. Second, if $\lambda_j = \lambda_m$, then for all other eigenvalues $-\lambda_i$ with $\lambda_i > \lambda_m$, the term $\lambda_i I + T$ is in the product $(\lambda_j I + T) \cdots (\lambda_1 I + T)(\lambda I + T)^n (-T)\mathbf{e}$ and repeats itself in the product for $\tau(-\lambda_i)$ times, where $\tau(-\lambda_i)$ is the algebraic multiplicity of $-\lambda_i$. That implies that the Jordan chains associated with $-\lambda_i$ are not in the expansion of the product $(\lambda_j I + T) \cdots (\lambda_1 I + T)(\lambda I + T)^n (-T)\mathbf{e}$. One of the consequences is that the residual term $-R(n)T\mathbf{e}$ disappears. Thus, equation (A.4) becomes

$$\begin{aligned} & (\lambda_j I + T) \cdots (\lambda_1 I + T)(\lambda I + T)^n (-T)\mathbf{e} \\ &= \left(\begin{array}{c} 0 \\ \sum_{k=1}^{\min\{K, n+1\}} (\lambda_j I + T_1) \cdots (\lambda_1 I + T_1) \mathbf{u}_1^{(k)} \left(\frac{n! (\rho(A))^{n+k-K+1}}{(n+k-K)!} \mathbf{v}^{(K)} \mathbf{e} + o\left(\frac{n! (\rho(A))^{n+k-K+1}}{(n+k-K)!} \right) \mathbf{e} \right) \end{array} \right). \end{aligned} \quad (\text{A.10})$$

Suppose that $n_0 = \min\{n : \lambda_n = \lambda_m\}$. For $j \geq n_0$, equation (A.10) becomes

$$\begin{aligned} & (\lambda_j I + T) \cdots (\lambda_1 I + T)(\lambda I + T)^n (-T)\mathbf{e} \\ &= \left(\begin{array}{c} 0 \\ \sum_{k=1}^{\min\{K, n+1\}} (\lambda_m I + T_1)^{j-n_0+1} \mathbf{u}_1^{(k)} \left(\frac{n! \left(\prod_{i=1}^{n_0-1} (\lambda_i - \lambda_m) \right) (\rho(A))^{n+k-K+1}}{(n+k-K)!} \mathbf{v}^{(K)} \mathbf{e} \right) \end{array} \right) \\ &+ \left(\begin{array}{c} 0 \\ \sum_{k=1}^{\min\{K, n+1\}} (\lambda_m I + T_1)^{j-n_0+1} \mathbf{u}_1^{(k)} \left(o\left(\frac{n! (\rho(A))^{n+k-K+1}}{(n+k-K)!} \right) \right) \end{array} \right) \\ &= \left(\begin{array}{c} 0 \\ \sum_{k=j-n_0+2}^{\min\{K, n+1\}} \mathbf{u}_1^{(k-j+n_0-1)} \left(\frac{n! \left(\prod_{i=1}^{n_0-1} (\lambda_i - \lambda_m) \right) (\rho(A))^{n+k-K+1}}{(n+k-K)!} \mathbf{v}^{(K)} \mathbf{e} + o\left(\frac{n! (\rho(A))^{n+k-K+1}}{(n+k-K)!} \right) \mathbf{e} \right) \end{array} \right). \end{aligned} \quad (\text{A.11})$$

where we used $(\lambda_m I + T_1) \mathbf{u}_1^{(k)} = \mathbf{u}_1^{(k-1)}$ and we assumed that K is the dimension of the largest Jordan block associated with $-\lambda_m$. The second part of the vector in the last line of equation (A.11) is either zero or nonnegative if n is large enough. If two or more than two Jordan chains associated with $-\lambda_m$ have the same length K , then these chains are not accessible from each other and can be dealt with separately. Thus, the vector $(\lambda_j I + T) \cdots (\lambda_1 I + T)(\lambda I + T)^n (-T)\mathbf{e}$

is nonnegative if n is large enough. Therefore, the matrix P is nonnegative. This completes the proof of Theorem 5.1. \square

Acknowledgements: The authors would like to thank two anonymous referees for their valuable comments and suggestions. Theorem 3.4 is added by suggestions from a referee. This research project was financially supported by NSERC (Canada) and the Chinese Academy of Sciences (P.R. China).

References

- [1] Altioik, T. (1985), On the phase-type approximations of general distributions, *IIE Transactions*, Vol **17**, 110-116.
- [2] Asmussen, S. (2000), *Ruin Probabilities*, World Scientific, Singapore.
- [3] Asmussen, S. (2003), *Applied Probability and Queues*, 2nd Edition, Springer-Verlag, New York.
- [4] Asmussen, S. and M. Bladt (1996), Renewal theory and queueing algorithms for matrix-exponential distributions, *Proceedings of the First International Conference on Matrix Analytic Methods in Stochastic Models*, 313-341, Editors: A.S. Alfa and S. Chakravathy, Marcel Dekker, New York.
- [5] Berman, A. and R.J. Plemmons (1979), *Nonnegative Matrices in the Mathematical Sciences*, Academic Press, New York.
- [6] Bobbio, A., M. Horvath, M. Scarpa, and M. Telek (2002), A cyclic discrete phase type distribution: properties and a parameter estimation algorithm, *Performance Evaluation*, Vol **54**, 1-32.
- [7] Commault, C. and S. Mocanu (2003), Phase-type distributions and representations: some results and open problems for system theory, *International Journal of Control*, Vol **76**, 566-580.
- [8] Cox, D.R. (1955), On the use of complex probabilities in the theory of stochastic processes, *Proc. Camb. Phil. Soc.*, Vol **51**, 313-319.
- [9] Cox, D.R. (1955), The analysis of non-Markovian stochastic processes by the inclusion of supplementary variables, *Proc. Camb. Phil. Soc.*, Vol **51**, 433-441.
- [10] Cumani, A. (1982), On the canonical representation of Markov processes modeling failure time distributions, *Microelectronics and Reliability*, Vol **22**, 583-602.
- [11] Dehon, M. and G. Latouche (1982), A geometric interpretation of the relations between the exponential and the generalized Erlang distributions, *Advances in Applied Probability*, Vol **14**, 885-897.
- [12] Erlang, A.K. (1917), Solution of some problems in the theory of probabilities of significance in automatic telephone exchange, *The Post Office Electrical Engineer's Journal*, Vol **10**, 189-197.
- [13] Gantmacher, F.R. (1959), *The Theory of Matrices*, Vols. I and II, Chelsea, New York.
- [14] HE, Q. and H. Zhang (2005), A note on unicyclic representations of phase type distributions. *Stochastic Models*, Vol **21**, 465-483.
- [15] HE, Q. and H. Zhang (2006), Spectral polynomial algorithms for computing bi-diagonal representations for phase-type distributions and matrix-exponential distributions, *Stochastic Models*, Vol **22**, 289-317.
- [16] HE, Q. and H. Zhang (2007), An algorithm for computing minimal Coxian representations, *INFORMS Journal on Computing* (accepted) (Working paper #05-03, Department of Industrial Engineering, Dalhousie University.)

- [17] Lancaster, P. and M. Tismenetsky (1985), *The Theory of Matrices*, Academic Press, New York.
- [18] Latouche, G. and V. Ramaswami (1999), *Introduction to Matrix Analytic Methods in Stochastic Modelling*, ASA & SIAM, Philadelphia, USA.
- [19] Lipsky, L. (1992), *Queueing Theory: A Linear Algebraic Approach*, McMillan and Company, New York.
- [20] Minc, H. (1988), *Non-negative Matrix*, John Wiley & Sons, New York.
- [21] Mocanu, S. and C. Commault (1999), Sparse representations of phase type distributions, *Stochastic Models*, Vol **15**, 759-778.
- [22] Neuts, M.F. (1975), Probability distributions of phase type, *In Liber Amicorum Prof. Emeritus H. Florin*, Department of Mathematics, Belgium: University of Louvain, 173-206.
- [23] Neuts, M. F. (1981), *Matrix-Geometric Solutions in Stochastic Models – An Algorithmic Approach*, The Johns Hopkins University Press, Baltimore.
- [24] Neuts, M.F. (1989), *Structured Stochastic Matrices of M/G/1 Type and their Applications*, Marcel Dekker, New York.
- [25] Neuts, M. F. (1992), Two further closure-properties of *PH*-distributions, *Asia-Pacific Journal of Operational Research*, Vol **9**, 77-85.
- [26] O’Cinneide, C.A. (1989), On non-uniqueness of representations of phase-type distributions, *Stochastic Models*, Vol **5**, 247-259.
- [27] O’Cinneide, C.A. (1990), Characterization of phase-type distributions, *Stochastic Models*, Vol **6**, 1-57.
- [28] O’Cinneide, C.A. (1991), Phase-type distributions and invariant polytope, *Advances in Applied Probability*, Vol **23**, 515-535.
- [29] O’Cinneide, C.A. (1993), Triangular order of triangular phase-type distributions, *Stochastic Models*, Vol **9**, 507-529.
- [30] O’Cinneide, C. A. (1999), Phase-type distributions: open problems and a few properties, *Stochastic Models*, Vol **15**, 731-757.
- [31] Rothblum, U.G. (1975), Algebraic eigenspaces of nonnegative matrices, *Linear Algebra and its Applications*, Vol **12**, 281-292.
- [32] Seneta, E. (1973), *Non-negative Matrices: An Introduction to Theory and Applications*, John Wiley & Sons, New York.
- [33] Takahashi, Y. (1981), Asymptotic exponentiality of the tail of the waiting time distribution in a *PH/PH/c* queue, *Advances in Applied Probability*, Vol **13**, 619-630.
- [34] van de Liefvoort, A. and A. Heindl (2005), Approximating matrix-exponential distributions by global randomization, *Stochastic Models*, Vol **21**, 669-694.