Survey

Open Access

Hung V. Le and M. J. Tsatsomeros*

Matrix Analysis for Continuous-Time Markov Chains

<https://doi.org/10.1515/spma-2021-0157>

Received August 27, 2021; accepted January 5, 2022

Abstract: Continuous-time Markov chains have transition matrices that vary continuously in time. Classical theory of nonnegative matrices, M-matrices and matrix exponentials is used in the literature to study their dynamics, probability distributions and other stochastic properties. For the benefit of Perron-Frobenius cognoscentes, this theory is surveyed and further adapted to study continuous-time Markov chains on finite state spaces.

Keywords: Continuous-time Markov chain; Stochastic matrix; M-matrix; Matrix exponential; Exponential nonnegativity; Irreducible matrix; Primitive matrix; Group inverse

MSC: 15B51, 15A48, 15A18, 15A09, 60J10, 92D40

1 Introduction

The concept and fundamental theory of continuous-time Markov chains are rarely encountered as applications in matrix theoretic textbooks, which is a motivating factor for this article. Similarly to classical Markov chains, the analysis involves stochastic matrices, M-matrices, Perron-Frobenius theory, generalized inverses and graph theory, along with enhancements that allow for time-dependence and non-stationary dynamics. For the benefit of linear algebraists who have not been exposed to this elegant area, we offer a survey of continuous-time Markov chain theory and some further developments, as well as directions and connections that point to future possible contributions.

Markov chains are indeed popular and well-established stochastic modeling tools, where it is often assumed that the transitions among states take place in fixed time intervals (steps) with transition probabilities that are also fixed in time. These assumptions are not always justified, as e.g., ecological systems rarely remain stable under dynamic environmental conditions. Relaxing them can therefore provide a powerful tool for studying fundamental problems of non-stationary dynamics, such as environmental effects of climatic changes.

The effort herein is specifically motivated by stochastic modeling, where the time inhomogeneity of the underlying Markov chain manifests itself in a continuous time-dependent transition matrix $P(t) = [p_{ij}(t)]$, which is row-stochastic for all $t \geq 0$ and comprises the transition probabilities $p_{ij}(t)$ among a finite number of states. A specific instance of such an ecological model is proposed in [10] for forest path dynamics among a finite number of biomass stages, where the functional form of the time-dependent probabilities are estimated via sampling and empirical data.

When the probabilistic model above presumes a continuous version of the Markov property (memorylessness), the resulting stochastic process is (frequently but not exclusively) referred to as a *continuous-time Markov chain* (ctMC). Such chains have been extensively studied; see e.g., [5, 13, 14]. Certain aspects of the

Hung V. Le: Mathematics and Statistics, Washington State University, Pullman, WA 99164, E-mail: hung.v.le@wsu.edu

***Corresponding Author: M. J. Tsatsomeros:** Mathematics and Statistics, Washington State University, Pullman, WA 99164, E-mail: tsat@wsu.edu

analysis of ctMC's on finite state spaces resemble the matrix theoretic approach used in discrete-time Markov chains. However, the Chapman-Kolmogorov conditions apply, giving rise to a semigroup, matrix exponential dependence of the probability distributions of ctMC's. In summary, one can postulate the behavior of $P(h)$ for small $h > 0$ (i.e., specify the derivatives of the entries of $P(t)$ at $t = 0$) and then use Markov properties to determine $P(t)$ for all $t \geq 0$ as the solution to an initial value differential problem $P'(t) = P(t)P'(0)$, which is indeed a matrix exponential function of $P'(0)$.

The presentation proceeds as follows. We assume that the reader is familiar with the basic theory of (entrywise) nonnegative matrices, stochastic matrices, as well as Markov chains and their association to matrices. Our terminology and notation generally follow [3]. However, basic notation, background and concepts are reviewed in Section 2 and within the rest of the text, as needed. In Section 3, the classical continuous-time Markov chain model is set up based on the exposition in [14], along with the underlying assumptions, principles and main consequences. An entrywise study of the transition matrix of a ctMC is pursued in Section 4; new results include Theorem 4.3, Corollary 4.4 and Theorem 4.5 from [9]. The entrywise dynamics as functions of time are further examined in Section 5, where some new observations are contained [9]: Theorem 5.1, Corollary 5.2 and Theorem 5.3. In Section 6 a study of a ctMC via generalized matrix inverses is initiated [9]. We conclude with a summary and possible future work on the subject.

2 General notations, definitions and preliminaries

The following definitions, facts and concepts are used in the sequel.

For an $n \times n$ real (complex) matrix $A \in \mathcal{M}_n(\mathbb{R})$ ($A \in \mathcal{M}_n(\mathbb{C})$), the spectrum of A is denoted by $\sigma(A)$ and its spectral radius by $\rho(A) = \max\{|\lambda| : \lambda \in \sigma(A)\}$.

Matrix $A = [a_{ij}]$ is called *nonnegative* (*positive*) if $a_{ij} \geq 0$ ($a_{ij} > 0$) for all $i, j \in \{1, 2, \dots, n\}$; this is denoted by $A \geq 0$ ($A > 0$).

Digraphs, primitivity and irreducibility

Consider a *digraph* (directed graph) $\Gamma = (V, E)$ with vertex set V and directed edge set E . A *walk of length* k from vertex u to vertex v in Γ is a sequence of k edges

$$u = i_1 \rightarrow i_2 \rightarrow \cdots \rightarrow i_{k+1} = v.$$

When the vertices are distinct, we refer to the walk as a *path* from u to v . A *cycle of length* k in Γ is a path

$$i_1 \rightarrow i_2 \rightarrow \cdots \rightarrow i_k \rightarrow i_1,$$

where the vertices $i_1, i_2, \dots, i_k$ are distinct. We say that vertex $u \in E$ has *access to* vertex $v \in E$ if there exists a path in Γ from u to v or $u = v$. If u has access to v and v has access to u , then u and v are *access equivalent*; an equivalence class under the relation of access is called an *access class*. We say access class W has *access to* vertex u (u has *access to* W) if there is a path from some vertex $w \in W$ to u (there is a path to u from some vertex $w \in W$). The subdigraphs induced by each of the access classes of Γ are the *strongly connected components* of Γ . A strongly connected component of Γ is *trivial* if it has no edges (i.e., it comprises a single vertex that does not have a loop). Γ is called *strongly connected* if it has exactly one strongly connected component and that component is not a trivial one.

The *transitive closure* of a digraph $\Gamma = (V, E)$ is the digraph $\bar{\Gamma} = (V, \bar{E})$, where a directed edge from i to j belongs to $\bar{E}$ if and only if there is path from i to j in Γ .

A digraph Γ is called *primitive* if it is strongly connected and the greatest common divisor of the lengths of its cycles is 1. Equivalently, Γ is primitive if and only if there exists a positive integer k such that for all vertices i, j of Γ , there is a walk of length k from i to j .

Given $A = [a_{ij}] \in \mathcal{M}_n(\mathbb{C})$, the digraph of A is

$$\Gamma(A) = (\{1, 2, \dots, n\}, \{(i, j) : a_{ij} \neq 0\}).$$

For a nonnegative matrix $A \in \mathcal{M}_n(\mathbb{R})$ and a positive integer k , the (i, j) -th entry of A^k is positive if and only if there exists a walk in $\Gamma(A)$ from i to j of length k .

The digraph $\Gamma(A)$ is primitive if and only if there exists m such that $A^k > 0$ for all $k \geq m$. We refer to this condition as the *Frobenius test for primitivity*; see e.g., [3, Chapter 2], where this condition serves as the definition of a nonnegative primitive matrix. We say A is *primitive* if $\Gamma(A)$ is primitive.

Next, recall that $A \in \mathcal{M}_n(\mathbb{C})$ is called *reducible* if there exists a permutation matrix P such that

$$PAP^T = \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix},$$

where A_{11} and A_{22} are square, non-vacuous matrices. Otherwise, A is called *irreducible*. Note that $[0]$ is considered reducible. It is well known that A is irreducible if and only if $\Gamma(A)$ is strongly connected.

Associated with the access classes of $\Gamma(A)$ is a standard form for a reducible matrix $A \in \mathcal{M}_n(\mathbb{C})$; it is obtained by repeated application of the reduction to block triangular form in the definition of matrix reducibility above. It is known as the *Frobenius Normal Form* of A : Given $A \in \mathcal{M}_n(\mathbb{C})$, there exist a permutation matrix P and positive integer $k \leq n$ such that

$$PAP^T = \begin{bmatrix} A_{11} & A_{12} & \cdots & \cdots & A_{1k} \\ 0 & A_{22} & A_{23} & \cdots & A_{2k} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & A_{k-1,k-1} & A_{k-1,k} \\ 0 & 0 & 0 & 0 & A_{kk} \end{bmatrix},$$

where $A_{11}, A_{22}, \dots, A_{kk}$ are either irreducible matrices or 1×1 zero matrices. Note that A is irreducible if and only if $k = 1$. Finally, an irreducible matrix $A \in \mathcal{M}_n(\mathbb{C})$ is called *cyclic* if there exist a permutation matrix P and positive integer $h \leq n$ such that

$$PAP^T = \begin{bmatrix} 0 & A_{12} & \cdots & \cdots & 0 \\ 0 & 0 & A_{23} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & 0 & A_{h-1,h} \\ A_{h1} & 0 & 0 & 0 & 0 \end{bmatrix},$$

where the zero diagonal blocks are square matrices. The integer h is called the *index of cyclicity* of A . A non-negative, irreducible matrix $A \in \mathcal{M}_n(\mathbb{R})$ with index of cyclicity h has exactly h eigenvalues of modulus $\rho(A)$, which are the distinct roots of $\lambda^h - \rho(A) = 0$. Such matrix A is primitive if and only if $h = 1$.

Perron-Frobenius and M-matrices

For general theory and background on the theory of nonnegative matrices (Perron-Frobenius theory) the reader is directed to [3] and [6]. The following variants of the Perron-Frobenius Theorem will specifically be quoted in the sequel.

Theorem 2.1. Let $A \geq 0$ be a square matrix. Then $\rho(A) \in \sigma(A)$ and $\rho(A)$ has a corresponding nonnegative eigenvector. Moreover, if A is irreducible, then $\rho(A)$ is a simple eigenvalue of A with a corresponding positive eigenvector x , and any other positive eigenvector of A is a scalar multiple of x .

Theorem 2.2. Let $A \geq 0$ be a primitive matrix. Then $\rho(A)$ is a simple eigenvalue of A greater than the modulus of any other eigenvalue of A .

We proceed with more relevant facts.

- (a) Nonnegative matrix $A \in \mathcal{M}_n(\mathbb{R})$ is irreducible if and only if $(I + A)^{n-1} > 0$ [3, Theorem 1.3].
- (b) Every nonnegative primitive matrix is irreducible. Also if $A \geq 0$ is irreducible and has positive trace, then A is primitive.
- (c) Matrix $T \in \mathcal{M}_n(\mathbb{C})$ is called *semiconvergent* whenever $\lim_{j \rightarrow \infty} T^j$ exists. It is known that T is semiconvergent if and only if $\rho(T) \leq 1$, any eigenvalue of T of modulus 1 is equal to 1, and the size of any Jordan block of T corresponding to 1 is 1×1 .

- (d) Matrix $M \in \mathcal{M}_n(\mathbb{R})$ of the form $M = sI - A$ where $s \geq \rho(A)$, $A \geq 0$ is called an *M-matrix*. An M-matrix M is nonsingular if and only if $s > \rho(A)$.
- (e) An M-matrix $M = sI - A$, $s \geq \rho(A)$, $A \geq 0$ is said to have *property c* if the matrix A/s is semiconvergent. We will work with graph inverses of M-matrices for which [8] is a comprehensive reference.

Stochastic matrices and Markov chains

- Recall that a nonnegative matrix $P = [p_{ij}] \in \mathcal{M}_n(\mathbb{R})$ is called (row) *stochastic* if $0 \leq p_{ij} \leq 1$ for all $i, j = 1, 2, \dots, n$ and $\sum_{j=1}^n p_{ij} = 1$ for all $i = 1, 2, \dots, n$. Equivalently, $P \in \mathcal{M}_n(\mathbb{R})$ is stochastic if $P \geq 0$ and $Pe = e$, where e denotes the all-ones column vector (in $\mathbb{R}^n$).
- For a stochastic matrix P , $\rho(P) = 1 \in \sigma(P)$ with e as a corresponding eigenvector.
- The entries of a stochastic matrix in $P = [p_{ij}] \in \mathcal{M}_n(\mathbb{R})$ are typically associated with the transition probabilities among the n states, $s_1, s_2, \dots, s_n$ of a Markov chain. The entry p_{ij} represents the probability that the process will next move to state s_j , given that it is currently in state s_i . The nonnegative k -th *distribution vector* is the vector $\pi^{(k)} = [\pi_1^{(k)}, \pi_2^{(k)}, \dots, \pi_n^{(k)}]^T$, where for each $k = 0, 1, \dots$, $e^T \pi^{(k)} = 1$. The stochastic modeling assumptions dictate that the distribution vectors and the *transition matrix* P satisfy the iteration

$$\pi^{(k)} = P^T \pi^{(k-1)} = (P^T)^k \pi^{(0)}, \quad k = 1, 2, \dots$$

The *initial distribution* is $\pi^{(0)}$. Each entry $\pi_i^{(k)}$ measures the probability that the process is in state s_i after k steps. If

$$\pi^{(0)} = \pi^{(k)}, \quad k = 1, 2, \dots,$$

we refer to $\pi^{(0)}$ as a *stationary distribution* of the Markov chain. Every Markov chain has a stationary distribution.

- A Markov chain with transition matrix $P \in \mathcal{M}_n(\mathbb{R})$ and its states can be classified according to the digraph of P , $\Gamma(P)$ as follows.
The vertices of $\Gamma(P)$ represent and are labeled by the states $s_1, s_2, \dots, s_n$. We say s_i has *access* to s_j if there is path from s_i to s_j . When s_i has access to s_j and s_j has access to s_i , we say these two states *communicate*. Communication is an equivalence relation that partitions the states in equivalence (access) classes that correspond to the irreducible components (diagonal blocks) of the Frobenius normal form of P .
A state s_i is called *transient* if it has access to a state s_j and s_j does not have access to s_i ; otherwise s_i is called *ergodic*. An equivalence class of states is called *ergodic* if every state in it is ergodic. A Markov chain is called *ergodic* if its states form a single ergodic class. A Markov chain is called *regular* if at some fixed step k there is a positive probability to move to every state. A Markov chain is called *periodic* if it is ergodic but not regular. As a consequence, a Markov chain with transition matrix A is
-ergodic if and only if P is irreducible;
-regular if and only if P is primitive;
-periodic if and only if P is irreducible and cyclic.

The matrix exponential

The matrix exponential function plays a central role in linear differential systems and, as we shall see, in the analysis of continuous-time Markov chains. We review some of its basic properties below.

The *matrix exponential* of $A \in \mathcal{M}_n(\mathbb{C})$ is denoted by e^A and defined by the power series

$$e^A = \sum_{k=0}^{\infty} \frac{1}{k!} A^k.$$

We will frequently consider the function e^{tA} , where $t \geq 0$ will be viewed as a variable representing time. It is well-known that the initial value problem

$$\frac{dx(t)}{dt} = Ax(t), \quad A \in \mathcal{M}_n(\mathbb{C}), \quad x(t) \in \mathbb{C}^n, \quad x_0 = x(0), \quad t \geq 0$$

has a unique solution given by

$$x(t) = e^{tA}x(0).$$

Let $A, B \in \mathcal{M}_n(\mathbb{C})$ and let $s, t \in \mathbb{C}$. Then the matrix exponential satisfies the following well-known properties:

$$e^0 = I, \quad e^{sA}e^{tA} = e^{(s+t)A}, \quad \frac{d}{dt}e^{tA} = Ae^{tA}, \quad (e^{At})^{-1} = e^{-At}.$$

Moreover, if $AB = BA$, then

$$Be^{At} = e^{At}B \quad \text{and} \quad e^{At}e^{Bt} = e^{(A+B)t} = e^{Bt}e^{At}.$$

Finally, if x is an eigenvector of $A \in \mathcal{M}_n(\mathbb{C})$ corresponding to $\lambda \in \sigma(A)$, then it is well-known that for each $t \geq 0$, x is an eigenvector of e^{tA} corresponding to $e^{t\lambda} \in \sigma(e^{tA})$.

The following classical result (see [15]) is fundamental in the continuous-time Markov chain stochastic model to be subsequently developed.

Observation 2.3. Let $B = [b_{ij}] \in \mathcal{M}_n(\mathbb{R})$. Then $e^{tB} \geq 0$ for all $t \geq 0$ if and only if $b_{ij} \geq 0$ for all $i \neq j$.

Proof. First, suppose that $b_{ij} \geq 0$ for all $i \neq j$. Then, there exists $s \geq 0$ such that $B + sI \geq 0$. It follows that

$$e^{tB} = e^{-ts}e^{tB+stI} = e^{-ts} \sum_{k=0}^n \frac{t^k}{k!} (B + sI)^k \geq 0.$$

For the converse, assume $e^{tB} \geq 0$ for all $t \geq 0$ and by way of contradiction $b_{ij} < 0$ for some $i \neq j$. Then, as $t \rightarrow 0^+$, the (i, j) entry of e^{tB} is dominated by b_{ij} and is thus negative, a contradiction to the assumption that $e^{tB} \geq 0$ for all $t \geq 0$. ■

3 The continuous-time Markov chain stochastic model

In this section, we follow closely the exposition in [14, Chapter VI, Section 6] (see also [4, Chapter VII, Section 2]) by considering a continuous-time Markov chain whose state at time $t \geq 0$ is $X(t) \in \{s_1, s_2, \dots, s_n\}$. The transition probabilities of this stochastic process from state to state comprise a time-dependent matrix $P(t) = [p_{ij}(t)] \in \mathcal{M}_n(\mathbb{R})$. It is assumed that for all $i, j \in \{1, 2, \dots, n\}$ and all $s \geq 0$,

$$p_{ij}(t) = \Pr(X(t+s) = s_j \mid X(s) = s_i) \in [0, 1] \quad \text{for all } t \geq 0;$$

that is, the probability of transition from state s_i to state s_j in t units of time is independent of the initial time s (stationary).

Furthermore, we assume that the stochastic model satisfies the following conditions for all $s, t \geq 0$ and all $i, j \in \{1, 2, \dots, n\}$:

$$\begin{aligned} \sum_{j=1}^n p_{ij}(t) &= 1; \\ p_{ij}(s+t) &= \sum_{k=1}^n p_{ik}(s)p_{kj}(t); \\ \lim_{t \rightarrow 0^+} p_{ij}(t) &= \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j. \end{cases} \end{aligned} \tag{3.1}$$

In matrix terms, we thus have that for all $s, t \geq 0$, $P(t)$ satisfies

$$P(t)e = e, \quad \text{where } e \in \mathbb{R}^n \text{ is the all-ones vector;} \tag{3.2}$$

$$P(t+s) = P(t)P(s) \quad (\text{Chapman-Kolmogorov condition}); \quad (3.3)$$

$$P(0) = I. \quad (3.4)$$

The Chapman-Kolmogorov equation (3.3) is a natural assertion of the fundamental Markov property that given the value of $X(t)$, the value of $X(t+s)$ for $s > 0$ only depends on $X(t)$ and not on the value of $X(u)$ for any $u < t$.

Equation (3.4) follows from (3.3) for $t = s = 0$ and along with (3.1) it asserts that $P(t)$ is continuous at $t = 0$. It can then be shown (for details, see [14, Chapter VI, Section 6]) that $P(t)$ is an entrywise continuous function of all $t > 0$. Indeed,

$$\lim_{h \rightarrow 0^+} P(t+h) = P(t) \lim_{h \rightarrow 0^+} P(h) = P(t)I = P(t).$$

Also for $0 < h < t$, we have

$$P(t) = P(t-h)P(h).$$

But $P(h)$ is near the identity matrix I for sufficiently small h , so $(P(h))^{-1}$ exists and also approaches the identity matrix I . Therefore,

$$\lim_{h \rightarrow 0^+} P(t-h) = P(t) \lim_{h \rightarrow 0^+} (P(h))^{-1} = P(t).$$

In fact, $P(t)$ is differentiable at $t = 0$ in the sense that

$$B := \lim_{h \rightarrow 0^+} \frac{P(h) - I}{h} \quad (3.5)$$

exists. Since

$$\frac{P(t+h) - P(t)}{h} = \frac{P(t)P(h) - P(t)}{h} = P(t) \frac{P(h) - I}{h},$$

it follows that $P(t)$ satisfies the matrix differential equation

$$P'(t) = P(t)B = BP(t) \quad (\text{Kolmogorov equation}) \quad (3.6)$$

with initial condition $P(0) = I$, where $P'(t) = [\frac{d}{dt} p_{ij}(t)]$. The solution to (3.6) thus is

$$P(t) = e^{tB}, \quad \text{where } B = P'(0). \quad (3.7)$$

The dynamics of the continuous-time Markov chain as described above are therefore governed by the Kolmogorov equation (3.6) and its explicit solution in (3.7). Indeed, let $z(t) = [z_j(t)] \in \mathbb{R}^n$ denote the state probability vector at time t , i.e.,

$$z_j(t) = \Pr(X(t) = s_j), \quad j = 1, 2, \dots, n. \quad (3.8)$$

Then

$$z^T(t) = z^T(0)P(t) = z^T(0)e^{tB}. \quad (3.9)$$

Remark 3.1. We note here that the continuous-time Markov process described above also admits a sojourn description as follows: Let $B = P'(0) = [b_{ij}]$ be the matrix in (3.7). Starting in state s_i , the process sojourns there for a duration of time that is exponentially distributed with parameter $-b_{ii}$. Then the process jumps to state s_j ($j \neq i$) with probability $-b_{ij}/b_{ii}$. The sojourn time in state s_j is exponentially distributed with parameter $-b_{jj}$ before it jumps to another state, etc. The sequence of states visited in this process forms a discrete-time Markov chain called the *embedded jump chain*.

Remark 3.2. For a continuous-time Markov chain as described above, we adopt a classification and terminology for its states at any given time $t \geq 0$ based on the digraph of $P(t)$. In particular, we call the continuous-time Markov chain *regular* if $P(t)$ is primitive for all $t > 0$.

4 The transition matrix of a continuous-time Markov chain

We proceed using the notation established in Section 3.

It is of interest to examine the time-dependent state, as well as the asymptotic (as $t \rightarrow \infty$) behavior of $z(t)$ and $P(t)$ of a continuous-time Markov chain as functions of $B = P'(0)$ (the infinitesimal transitions rates). This task is accomplished by adapting the classical matrix-theoretic analysis of homogeneous Markov chains [3, Chapter 8].

We first observe the special nature of B in (3.7). Recall that $e \in \mathbb{R}^n$ denotes the all-ones vector.

Lemma 4.1. *Let $B = P'(0) = [b_{ij}] \in \mathcal{M}_n(\mathbb{R})$ as in (3.7). Then $-B$ is a singular M -matrix and $Be = 0$. Moreover, there exists $h > 0$ such that $-1 \leq h b_{ii} < 0$ and $0 \leq h b_{ij} \leq 1$ for all $i \neq j$, and $-b_{ii} \geq b_{ij}$ for all $i \neq j$.*

Proof. By (3.2), for all $h \geq 0$, we have

$$P(h)e = e \Rightarrow [P(h) - I]e = 0 \Rightarrow \lim_{h \rightarrow 0^+} \left[\frac{P(h) - I}{h} \right]e = 0.$$

Thus, $Be = 0$, which implies that B is singular. Moreover, we know

$$-B = \lim_{h \rightarrow 0^+} \frac{I - P(h)}{h} = \lim_{h \rightarrow 0^+} \frac{I}{h} - \lim_{h \rightarrow 0^+} \frac{P(h)}{h}. \quad (4.10)$$

Since $P(h) \geq 0$, we have $\frac{P(h)}{h} \geq 0$ for all $h > 0$. For sufficiently small $h = \hat{h} > 0$, we have $\frac{1}{\hat{h}} > 1$ and so by (4.10), $-\hat{h}B = I - A$ for some nonnegative matrix $A = [a_{ij}]$, which must therefore be stochastic. Writing now $\hat{h}B = A - I$ with $A \geq 0$, since each row of B sums to 0 and $-\hat{h}b_{ii} = 1 - a_{ii} > 0$, we have that for each $i = 1, 2, \dots, n$,

$$0 < -\hat{h}b_{ii} = \sum_{j \neq i} \hat{h}b_{ij} = \sum_{j \neq i} a_{ij} = 1 - a_{ii} \leq 1 \Rightarrow -1 \leq \hat{h}b_{ii} < 0.$$

Also, for $i \neq j$, we have $\hat{h}b_{ij} = a_{ij} \in [0, 1]$. Moreover, for each $i \neq j$,

$$-b_{ii} = \sum_{i \neq j} b_{ij} \geq b_{ij}. \quad \blacksquare$$

The following result describes some properties of the digraph of $P(t)$ and how they relate to the digraph of B . Note that if we write $hB = A - I$ as in the proof of Lemma 4.1, then $\overline{\Gamma(B)} = \overline{\Gamma(A)}$, and $\Gamma(B)$ is irreducible if and only if $\Gamma(A)$ is irreducible.

Lemma 4.2. *Consider $P(t) = e^{tB}$ as in (3.7). The following hold*

- (i) *For each $t \geq 0$, $P(t) \geq 0$ is a stochastic matrix.*
- (ii) *For each $t > 0$, the digraph $\Gamma(P(t))$ is the transitive closure of $\Gamma(B)$.*
- (iii) *$P(t)$ is irreducible for some $t > 0$ if and only if B is irreducible.*
- (iv) *If $P(t)$ is irreducible for some $t > 0$, then $P(t)$ is irreducible for all $t > 0$.*
- (v) *$P(t) > 0$ (and hence $P(t)$ is primitive) for some $t > 0$ if and only if B is irreducible.*
- (vi) *If $P(t)$ is primitive for some $t > 0$, then it is primitive for all $t > 0$.*

Proof.

As in the proof of Lemma 4.1, we have that for some $h > 0$, $hB = A - I$, where $A \geq 0$ is stochastic.

(i) For each $t = h\hat{t} \geq 0$, we have

$$P(t) = e^{tB} = e^{\hat{t}hB} = e^{-\hat{t}} \sum_{k=0}^{\infty} \frac{\hat{t}^k A^k}{k!} \geq 0. \quad (4.11)$$

Thus, since $Be = 0$, we have $P(t)e = e$ and therefore $P(t)$ is a stochastic matrix.

(ii) Since $A \geq 0$, if there is a walk from i to j in $\Gamma(A)$, then there exists a positive integer k such that the (i, j) entry of A^k is positive. For each $t > 0$, $P(t)$ is an infinite series of positive multiples of all the powers of A , and so it follows that $\Gamma(P(t))$ is the transitive closure of $\Gamma(A)$, which in turn coincides with the transitive closure of $\Gamma(B)$.

(iii) Since $A \geq 0$, A is irreducible if and only if its powers are irreducible (there are no accidental cancellations when the powers are computed). By (4.11), $P(t)$ is irreducible for some $t > 0$ if and only if A , and thus B , are irreducible matrices.

(iv) If $P(t)$ is irreducible for some $t > 0$, then A and $hB = A - I$ are irreducible by (iii), in which case it follows by (4.11), as in the proof of (iii), that $P(t)$ is irreducible for any $t > 0$.

(v) & (vi) B is irreducible if and only if there is a walk from any vertex i to any other vertex j in $\Gamma(A)$, or equivalently, there exists k such that the (i, j) entry of A^k is positive. Thus, by (4.11), $P(t) > 0$ for some (and hence all) $t > 0$ is equivalent to B being irreducible. ■

We will now proceed under the **assumption that B is irreducible**, in which case, by Lemma 4.2, $P(t)$ is a primitive stochastic matrix for all $t > 0$. In particular, by Theorem 2.2, $\rho(P(t)) = 1$ is a simple and the sole eigenvalue of $P(t)$ lying on the unit circle; the corresponding eigenspace is spanned by e .

Notice also that since $P(t) = e^{tB}$, where $B = P'(0)$ is fixed, and since e^{tB} and B share left eigenvectors, the left eigenvectors of $P(t)$ corresponding to 1 coincide with the left eigenvectors of B .

As a consequence, because of the primitivity of $P(t)$, for all $t > 0$, the left eigenspace of $P(t)$ corresponding to $\rho(t) = 1$ is spanned by some constant, common (for all $t \geq 0$) vector, $\pi \in \mathbb{R}^n$. As the transpose of $P(t)$ is also a primitive matrix, we can take π to be entrywise positive and normalize it to satisfy $\pi^T e = 1$.

Note that the above observations are intuitively in agreement with the notion of the embedded jump chain (see Remark 3.1), which suggests that under certain conditions (primitivity), there exists a unique stationary probability distribution $\pi \in \mathbb{R}^n$ such that

$$P^T(t)\pi = \pi = \lim_{t \rightarrow \infty} z(t) = \lim_{t \rightarrow \infty} P^T(t)z(0). \quad (4.12)$$

In conclusion, we have that π in (4.12) is thus the unique solution to the system

$$\pi^T B = 0, \quad \pi^T e = 1, \quad (4.13)$$

where $z(t)$ is the state probability vector introduced in (3.8) and (3.9). Moreover, under the assumption of irreducibility of B , and since then e^{tB} is primitive for all $t > 0$, we have that (see [6, Theorem 8.2.8])

$$\lim_{t \rightarrow \infty} e^{tB} = \lim_{m \rightarrow \infty} (e^B)^m = e\pi^T.$$

The properties and behavior of homogeneous Markov chains with fixed transition probabilities $P \in \mathcal{M}_n(\mathbb{R})$ are intimately related to the singular M-matrix $A = I - P$; see [3, Chapter 6]. In keeping with this approach, we now consider the continuous-time model described above and define a time-dependent M-matrix function $A(t)$ as follows.

$$A(t) = I - P(t), \quad P(t) = e^{tB} \geq 0, \quad B = P'(0). \quad (4.14)$$

Under the assumption that B is irreducible, we have that for all $t > 0$, $P(t)$ is a primitive matrix satisfying

$$\rho(P(t)) = 1, \quad P(t)e = e, \quad P^T(t)\pi = \pi, \quad e^T \pi = 1, \quad (4.15)$$

where e is the all-ones vector, π is a positive vector, $\dim \text{Nul}(I - P(t)) = 1$, and 1 is the sole eigenvalue of $P(t)$ on the unit circle, for all $t > 0$.

We have seen in (4.12) that $\pi = \lim_{t \rightarrow \infty} P^T(t)z(0)$. In what follows we directly connect π to the adjoint of $I - P(t)$ and its trace, thus obtaining an explicit relation of π in terms of $P(t)$ for every $t > 0$.

Recall that the (i, j) entry of the adjoint of a matrix $X \in \mathcal{M}_n(\mathbb{R})$, denoted by $\text{adj}(X)$, is the (i, j) -th co-factor of X , i.e.,

$$(-1)^{i+j} \det X_{ji},$$

where X_{ji} denotes the submatrix of X obtained when row j and column i are deleted. It is well-known that $\text{adj}(X)X = X\text{adj}(X) = \det(X)I$.

Theorem 4.3. Let $P(t) \in \mathcal{M}_n(\mathbb{R})$, $A(t) = I - P(t)$ as in (4.14) and (4.15) under the assumption that $B = P'(0)$ is irreducible. Let $C(t) = \text{adj}(A(t))$ for $t > 0$. Then there exists scalar function $\delta(t) > 0$ such that for all $t > 0$,

$$C(t) = \delta(t) e \pi^T.$$

Proof. Let $t > 0$. Since $P(t)$ has the simple eigenvalue $\rho(P(t)) = 1$, $A(t) = I - P(t)$ has nullity one and rank $n - 1$. By the monotonicity of the spectral radius as a function of the entries of a nonnegative matrix [6, Corollary 8.1.19], the spectral radius of every $(n - 1) \times (n - 1)$ principal submatrix of $P(t)$ is strictly less than 1. Thus every $(n - 1) \times (n - 1)$ principal submatrix of $A(t)$ is invertible, which implies that $C(t) = \text{adj}(A(t))$ is nonzero. Also, since $C(t)A(t) = A(t)C(t) = \det(A(t))I = 0$ and $\text{rank}(A(t)) = n - 1$, it follows that nullity of $C(t) \geq n - 1$, which implies that $\text{rank}(C(t)) \leq 1$. Thus, $\text{rank}(C(t)) = 1$.

The null space of $A(t)$ is indeed spanned by e and the left null space of $A(t)$ is spanned by π . Thus the columns of $C(t)$ are multiples of e and its rows multiples of π . It follows that $C(t) = \delta(t)e\pi^T$ for some $\delta(t) \in \mathbb{R}$. To prove that $\delta(t)$ is positive, note that as $A(t) = I - P(t)$ is an M-matrix, its principal minors are nonnegative. In particular, the diagonal entries of $C(t)$ are positive. Since e and π are positive vectors, it must then also be that $\delta(t) > 0$. ■

Corollary 4.4. In the context of Theorem 4.3, let $\sigma(P(t)) = \{\lambda_1(t) = 1, \lambda_2(t), \dots, \lambda_n(t)\}$ be the spectrum of $P(t)$. Then for each $t > 0$,

$$\delta(t) = \text{trace}(C(t)) = \prod_{j=2}^n (1 - \lambda_j(t)).$$

Proof.

Notice that

$$\text{trace}(C(t)) = \delta(t) \text{trace}(\pi e^T) = \delta(t) \text{trace}(e^T \pi) = \delta(t).$$

Also

$$\text{trace}(C(t)) = (-1)^{i+i} \det(A(t))_{ii},$$

which is equal to the sum of all the $(n - 1)$ -fold products (i.e., the $(n - 1)$ -st elementary symmetric function) of the n eigenvalues of $A(t) = I - P(t)$ [6]. Thus,

$$\delta(t) = \sum_{k=1}^n \prod_{j \neq k} (1 - \lambda_j(t)) = \prod_{j=2}^n (1 - \lambda_j(t)).$$

We can now use Theorem 4.3 to explicitly express π in terms of $P(t)$. ■

Theorem 4.5. Let $P(t) \in \mathcal{M}_n(\mathbb{R})$, $A(t) = I - P(t)$ and $C(t) = \text{adj}(A(t)) = [c_{ij}(t)]$ as in Theorem 4.3. Let $\alpha(t) = \sum_{j=1}^n c_{1j}(t)$. Then for each $t > 0$,

$$\pi = \frac{1}{\alpha(t)} [c_{11}(t) \ c_{12}(t) \ \dots \ c_{1n}(t)]^T,$$

where $c_{1j}(t) = (-1)^{1+j} \det A_{1j}(t)$.

Proof. By Theorem 4.3, each row of $C(t)$ is equal to a positive multiple of π^T . Since π is the unique normalized left null vector of B , each row of $C(t)$ normalized to have sum equal to one, must equal π . ■

Remark 4.6. From Lemma 4.2, it is evident that when B is irreducible, $P(t)$ has to be primitive for all $t > 0$ and so the continuous-time Markov chain must have a unique stationary distribution. Also, all the entries of $P(t) = e^{tB}$ are positive for all $t > 0$, that is, the probability for the process leaving its current state is always

positive. This means that under the assumption of irreducibility of B , by Lemma 4.2 (v), the continuous-time Markov chain is regular.

Also note that from Lemma 4.2, when $B = P'(0)$ is reducible, then $P(t) = e^{tB}$ is reducible for all $t \geq 0$. Without loss of generality, let B be symmetrically permuted into Frobenius normal form with irreducible diagonal blocks B_{jj} ($j = 1, 2, \dots, k, k \geq 2$). By Lemma 4.2, for each $t > 0$, $\Gamma(P(t))$ is the transitive closure of $\Gamma(B)$ and thus its Frobenius normal form has irreducible diagonal blocks equal to $P_{jj}(t) = e^{tB_{jj}}$ ($j = 1, 2, \dots, k$). A study can then ensue that ties the k irreducible components of B to the limiting behavior of $z(t)$ in (3.9). It is apparent that in this case, the stationary distribution vector is not in general unique anymore as the Markov chain contains transient states. The study of $P(t)$ pursued in Section 5 applies to reducible case of B as no assumption of irreducibility is made.

5 More on the dynamic behavior of $P(t)$

Recalling the stochastic process of a continuous-time Markov chain as described in (3.8) and (3.9), we will examine more closely the dynamic behavior of (the entries of) $P(t)$, where $P(t)$ is as described in (3.7). The results that follow offer a better understanding of monotonic and relative behavior of the transition probabilities as functions of t .

Given the dynamical system in (3.9), observe that

$$z'(t) = B^T e^{tB^T} z(0) = e^{tB^T} (B^T z(0)),$$

so the monotonicity of $z(t)$ depends on $B^T z(0)$. As the entries of $z(t)$ are nonnegative and add up to 1, at no time $t \geq 0$ can all entries of $B^T z(t)$ be positive. In particular, taking $z(0)$ to be the l -th column of the identity, it follows that the l -th row of $P(t) = e^{tB}$ behaves monotonically according to the sign pattern of $z^T(0)B$. The columns of $P(t)$ also obey an orderly monotonic pattern, as explained in the following results.

Theorem 5.1. *Let $P(t) = [p_{ij}(t)] \geq 0$ be the transition matrix in (3.7). For any $t > 0$, let $i, j \in \{1, 2, \dots, n\}$ be such that $p_{il}(t)$ and $p_{jl}(t)$ are maximal and minimal entries in column l of $P(t)$, respectively. Then, $[p_{il}(t)]' \leq 0$ and $[p_{jl}(t)]' \geq 0$.*

Proof. Since $P'(t) = BP(t)$, $p_{il}(t) \geq p_{ml}(t) \geq 0$ for every m , and $b_{im} \geq 0$ for all $i \neq m$, we have

$$[p_{jl}(t)]' = \sum_{m=1}^n b_{jm} p_{ml}(t) = b_{jj} p_{jl}(t) + \sum_{m \neq j} b_{jm} p_{ml}(t) \geq b_{jj} p_{jl}(t) + \sum_{m \neq j} b_{jm} p_{jl}(t) = \sum_{m=1}^n b_{jm} p_{jl}(t).$$

Thus,

$$[p_{jl}(t)]' \geq \sum_{m=1}^n b_{jm} p_{jl}(t) = p_{jl}(t) \sum_{m=1}^n b_{jm} = p_{jl}(t) 0 = 0 \Rightarrow [p_{jl}(t)]' \geq 0.$$

Analogously, we have $[p_{il}(t)]' \leq 0$. ■

Corollary 5.2. *Let $P(t) = [p_{ij}(t)] \geq 0$ be the transition matrix in (3.7). For any $t > 0$, let $i, j \in \{1, 2, \dots, n\}$ be such that $p_{il}(t) = M(t)$, $p_{jl}(t) = m(t)$ are maximal and minimal entries, respectively, in column l of $P(t)$. Let $p_{kl}(t)$ be an arbitrary entry of column l of $P(t)$. Then*

$$[M(t) - m(t)] b_{ii} \leq [p_{il}(t)]' \leq 0,$$

$$0 \leq [p_{jl}(t)]' \leq [m(t) - M(t)] b_{jj}$$

and

$$[p_{kl}(t) - m(t)] b_{kk} \leq [p_{kl}(t)]' \leq [p_{kl}(t) - M(t)] b_{kk}.$$

Proof. Let $x(t)$ denote column l of $P(t)$, and let $x_i(t)$ and $x_j(t)$ be maximal and minimal entries of $x(t)$, respectively. By Theorem 5.1, we have $(Bx(t))_i \leq 0$ and $(Bx(t))_j \geq 0$. Let $\tilde{x}(t)$ be the vector generated by replacing $x_i(t)$ by $m(t)$ in $x(t)$. Then $(\tilde{x}(t))_i$ is a minimal entry of $\tilde{x}(t)$. Thus,

$$\begin{aligned}(Bx(t))_i - [x_i(t) - m(t)] b_{ii} &= (B\tilde{x}(t))_i \geq 0 \\ \Rightarrow [M(t) - m(t)] b_{ii} &\leq (Bx(t))_i \leq 0 \\ \Rightarrow [M(t) - m(t)] b_{ii} &\leq [p_{ii}(t)]' \leq 0.\end{aligned}$$

Similarly, we have

$$0 \leq [p_{jj}(t)]' \leq [m(t) - M(t)] b_{jj}.$$

Now let $x_k(t)$ be an arbitrary entry of $x(t)$. Let $y(t)$ be the vector generated from $x(t)$ by replacing $x_k(t)$ by $M(t)$ and let $z(t)$ be the vector generated from $x(t)$ by replacing $x_k(t)$ by $m(t)$. Then we have

$$\begin{aligned}(Bz(t))_k + [x_k(t) - m(t)] b_{kk} &= (Bx(t))_k = (By(t))_k - [M(t) - x_k(t)] b_{kk} \\ \Rightarrow [x_k(t) - m(t)] b_{kk} &\leq (Bx(t))_k \leq [x_k(t) - M(t)] b_{kk}\end{aligned}$$

since $(By(t))_k \leq 0 \leq (Bz(t))_k$.

Thus,

$$[p_{kl}(t) - m(t)] b_{kk} \leq [p_{kl}(t)]' \leq [p_{kl}(t) - M(t)] b_{kk}.$$

The following theorem is analogous to and implicit in the proof of convergence to a stationary distribution of a discrete Markov process in [1, Eq. (9), p. 268]. ■

Theorem 5.3. Let $P(t) = [p_{ij}(t)] = e^{tB} \geq 0$ be the transition matrix in (3.7). For any $t > 0$, let $\epsilon(t)$ be the smallest value among the entries of $P(t)$. Let m_0 and M_0 be the minimum and maximum values among the entries of a fixed real vector x , respectively. Also, let $m_1(t) = (P(t)x)_l$ and $M_1(t) = (P(t)x)_s$ be the minimum and maximum values among the entries of $P(t)x$, respectively, for some $l, s \in \{1, 2, \dots, n\}$. Then $m_0 \leq m_1(t) \leq M_1(t) \leq M_0$, and

$$M_1(t) - m_1(t) \leq (1 - 2\epsilon(t))(M_0 - m_0).$$

Proof. For some integer l , we have

$$m_1(t) = \sum_{k=1}^n p_{lk}(t) x_k \geq \sum_{k=1}^n p_{lk}(t) m_0 = m_0$$

since $P(t) = [p_{ij}(t)] \geq 0$ and $\sum_{k=1}^n p_{lk}(t) = 1$. Similarly, for some integer s , we have

$$M_1(t) = \sum_{k=1}^n p_{sk}(t) x_k \leq \sum_{k=1}^n p_{sk}(t) M_0 = M_0.$$

Thus, $m_0 \leq m_1(t) \leq M_1(t) \leq M_0$.

Now assume that $x_i = m_0$ and $x_j = M_0$ are a minimal entry and a maximal entry of a fixed real vector x , respectively. Let y be the vector obtained from x by replacing all entries of x by M_0 except for the entry x_i . Then, $x \leq y$ and each entry of $P(t)y$ satisfies

$$(P(t)y)_l = \sum_{k=1}^n p_{lk}(t) y_k = p_{li} m_0 + (1 - p_{li}) M_0 = M_0 - p_{li}(M_0 - m_0) \leq M_0 - \epsilon(t)(M_0 - m_0)$$

since $\epsilon(t)$ is the smallest entry of $P(t)$. Moreover, since $x \leq y$, for some $l \in \{1, 2, \dots, n\}$ we have

$$M_1 \leq (P(t)y)_l \leq M_0 - \epsilon(t)(M_0 - m_0). \quad (1)$$

Negating x , we see that $-m_0$ is maximal entry and $-M_0$ is minimal among the entries of $-x$. As above, we now get

$$-m_1 \leq -m_0 - \epsilon(t)(M_0 - m_0). \quad (2)$$

From (1) and 2, we obtain

$$M_1 - m_1 \leq (1 - 2\epsilon(t))(M_0 - m_0).$$

■

Remark 5.4. We note that the results in this section hold for irreducible as well as reducible B . However, if B is irreducible, in which case $\lim_{t \rightarrow \infty} P(t) = e\pi^T$, then the observations herein take a special meaning. For example, at any $t > 0$,

- the entries in the l -th column of $P(t)$ approach in value the l -th entry of π in a fashion described in Theorem 5.1: maximal (minimal) elements in the column do not further increase (decrease).
- the entries in every row of $P(t)$ tend toward π as $t \rightarrow \infty$ in a mixed monotonic fashion within $[0, 1]$.

6 The group generalized inverse of $A(t) = I - P(t)$

Recall the notion of the *group (generalized) inverse* of $A \in \mathcal{M}_n(\mathbb{R})$, which, if it exists, is the unique solution X to the matrix equations $AXA = A$, $XAX = X$ and $XA = AX$. When the group inverse exists, it is denoted by $A^\sharp$. A necessary and sufficient condition for $A^\sharp$ to exist is that $\text{rank}(A^2) = \text{rank}(A)$; see [2] for details.

In this section, we continue to consider the continuous-time Markov chain $P(t) = e^{tB}$ with irreducible B and its association with the M-matrix function

$$A(t) = I - P(t), \quad t \geq 0,$$

which is indeed a singular matrix for each $t \geq 0$.

We will consider the matrices

$$A(t)^\sharp = (I - P(t))^\sharp \quad \text{and} \quad L(t) = I - A(t)A(t)^\sharp,$$

where $A(t)^\sharp$ is the group inverse of $A(t)$, whose existence is guaranteed by Theorem 6.1 below. In the homogeneous case, these matrices contain critical information about a Markov chain, as well as provide computationally convenient methods to obtain this information (see [3, Chapter 8, Section 4]). We will pursue an analogous study for continuous-time Markov chains based on $A(t)$. For clarity, we will focus on the case where B is irreducible, so by Lemma 4.2, $P(t)$ is primitive and $\rho(t) = 1$ is a simple eigenvalue of $P(t)$ for each $t > 0$ and all other eigenvalues have modulus less than 1.

Recalling the notions of semiconvergence and "property c" from Section 2 (fact (e) following Theorem 2.2), we have the following basic observation, which is based on the fact that an M-matrix A has "property c" if and only if $\text{rank}(A) = \text{rank}(A^2)$.

Theorem 6.1. *Let $B = P'(0) \in \mathcal{M}_n(\mathbb{R})$ be irreducible and $P(t) = e^{tB}$ be the transition matrix for a continuous-time Markov chain in (3.7). Then for each $t \geq 0$,*

$$A(t) = I - P(t)$$

is an M-matrix with "property c".

Proof. Since $P(t) \geq 0$ and $\rho(P(t)) = 1$, $A(t)$ is a singular M-matrix for every fixed $t \geq 0$. Now we only need to show that $A(t)$ has "property c" for all $t \geq 0$. When $t = 0$, $P(0) = I$ and $A(0) = 0$ which is indeed an M-matrix with "property c". Let $t > 0$ be arbitrary but fixed. We need to show that $\text{rank}(A(t)) = \text{rank}(A^2(t))$. Since B is irreducible, $\rho(P(t)) = 1$ is a simple eigenvalue of $P(t)$ and all other eigenvalues have modulus less

than 1. Therefore, invoking the Jordan canonical form of $P(t)$, there exists a nonsingular matrix $S(t)$ such that $P(t) = S(t)J(t)S(t)^{-1}$, where

$$J(t) = \begin{bmatrix} 1 & 0 \\ 0 & N(t) \end{bmatrix}$$

and $\rho(N(t)) < 1$. Then $A(t) = I - P(t) = S(t)(I - J(t))S(t)^{-1}$, which implies that $\text{rank}(A(t)) = \text{rank}(I - J(t))$. Since

$$I - J(t) = \begin{bmatrix} 0 & 0 \\ 0 & I - N(t) \end{bmatrix}$$

and $I - N(t)$ is nonsingular, we can see that $\text{rank}(A(t)) = \text{rank}(I - J(t)) = \text{rank}(I - J(t))^2 = \text{rank}(A^2(t))$. Therefore, $A(t) = I - P(t)$ is an M-matrix with "property c". ■

Remark 6.2. An analogue to Theorem 6.1 for reducible B can be based on the Frobenius normal form of B .

The condition $\text{rank}(A(t)) = \text{rank}(A^2(t))$ in the proof of Theorem 6.1 is indeed equivalent to the existence of the group generalized inverse $A(t)^\sharp$, which, in turn, is also characterized by its action on $x \in \mathbb{R}^n$ as follows:

$$A(t)^\sharp x = y \quad \text{if } x \in R(A(t)) \text{ and } A(t)y = x$$

and

$$A(t)^\sharp x = 0 \quad \text{if } x \in N(A(t)).$$

Thus, as an immediate consequence of Theorem 6.1 we have the following corollary.

Corollary 6.3. Let $B = P'(0) \in \mathcal{M}_n(\mathbb{R})$ be irreducible and $P(t) = e^{tB}$ be the transition matrix for a continuous-time Markov chain as in (3.7). Let $A(t) = I - P(t)$. Then $A(t)^\sharp$ exists for all $t \geq 0$.

Theorem 6.4. Let $B = P'(0) \in \mathcal{M}_n(\mathbb{R})$ be irreducible and $P(t) = e^{tB}$ be the transition matrix for a continuous-time Markov chain as in (3.7). Let $A(t) = I - P(t)$ and define $L(t) = I - A(t)A(t)^\sharp$. Then for all $t \geq 0$,

$$\lim_{j \rightarrow \infty} P(t)^j = L(t).$$

Proof. As in the proof of Theorem 6.1, for each $t \geq 0$, there exists a nonsingular matrix $S(t)$ such that $P(t) = S(t)J(t)S(t)^{-1}$, where

$$J(t) = \begin{bmatrix} 1 & 0 \\ 0 & N(t) \end{bmatrix}$$

and $\rho(N(t)) < 1$. Thus,

$$\lim_{j \rightarrow \infty} P(t)^j = \lim_{j \rightarrow \infty} S(t)J(t)^j S(t)^{-1} = \lim_{j \rightarrow \infty} S(t) \begin{bmatrix} 1 & 0 \\ 0 & N(t)^j \end{bmatrix} S(t)^{-1} = S(t)J(t)^j S(t)^{-1} = S(t) \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} S(t)^{-1}$$

since $\rho(N(t)) < 1$. Also, since

$$A(t)^\sharp x = y \quad \text{if } x \in R(A(t)) \text{ and } A(t)y = x$$

and

$$A(t)^\sharp x = 0 \quad \text{if } x \in N(A(t)),$$

it follows that

$$S(t) \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} S(t)^{-1} = I - A(t)A(t)^\sharp = L(t).$$

Next, we will establish the relationship between $A(t) = I - P(t)$ and the left and right Perron vectors of $P(t)$. ■

Theorem 6.5. Let $B = P'(0) \in \mathcal{M}_n(\mathbb{R})$ be irreducible and $P(t) = e^{tB}$ be the transition matrix for a continuous-time Markov chain as in (3.7). Let $A(t) = I - P(t)$ and $L(t) = I - A(t)A(t)^\sharp$. Then for all $t \geq 0$,

$$L(t) = e\pi^T,$$

where π is the stationary probability distribution vector associated with the stochastic model and e is the all-ones vector.

Proof. For each $t \geq 0$, we have

$$L(t)A(t) = (I - A(t)A(t)^\sharp)A(t) = A(t) - A(t) = 0,$$

since $A(t)A(t)^\sharp A(t) = A(t)$. Thus

$$L(t)P(t) = L(t).$$

Also, since a Markov chain is ergodic if and only if it has a unique positive stationary distribution vector π (see, Section 2 and Lemma 4.2 (iv), or [5, Theorem 21, p. 261]), each row of $L(t)$ is a multiple of π . By Theorem 6.4, $L(t)$ is also a stochastic matrix as the limit of stochastic matrices, which implies that

$$L(t) = \begin{bmatrix} \pi^T \\ \pi^T \\ \vdots \\ \pi^T \end{bmatrix} = e\pi^T.$$

The following remarks provide context and exemplify the potential of the approach developed in this section. ■

Remark 6.6.

- By Theorems 6.4 and 6.5, we confirm our findings that when $B = P'(0) \in \mathcal{M}_n(\mathbb{R})$ is irreducible and $P(t) = e^{tB}$ is the transition matrix for a continuous-time Markov chain as in (3.7), then

$$\lim_{t \rightarrow \infty} P(t) = L(t) = e\pi^T.$$

- We note that in Theorem 6.5, $L(t)$ is a nonnegative, rank 1, stochastic matrix. For an arbitrary chain, $L(t)$ is a nonnegative, rank r , stochastic matrix, where r is the number of ergodic classes associated with the chain.
- For a general chain, the entries of $L(t)$ can be used to obtain the probabilities of eventual absorption into any one particular ergodic class, for each starting state. Moreover, if $P(t)$ is the transition matrix for an absorbing chain, then we have the following:
 - (a) If s_j is an absorbing state, then $l_{ij}(t)$ equals that probability of eventual absorption of the process from state s_i to the class $[s_j]$.
 - (b) If s_i and s_j are transient states, then $a_{ij}^\sharp(t)$ equals the expected number of times the chain will be in s_j when it is initially in s_i .

7 Summary, conclusions and future work

The stochastic process we identified as a continuous-time Markov chain (ctMC) is described in equations (3.6) - (3.9) and satisfies the following:

- The transition probabilities $P(t)$ of the ctMC are entirely determined by $B = P'(0)$.
- ctMC possesses at least one stationary distribution.
- If $B = P'(0)$ is irreducible, then
 - (i) ctMC possesses a unique stationary distribution π ;
 - (ii) π is the unique nonnegative solution to $B^T \pi = 0$, $e^T \pi = 1$;

(iii) ctMC is regular, i.e., at all times $t > 0$, the probability of the process of transitioning to any state is positive.

- The entries of $P(t)$ ensue in a particular balanced manner in which the largest (resp., smallest) transition probabilities to any given state trend downward (resp., upward) at any given $t > 0$.

The particular way the mathematical model of a ctMC arises, points to the need for a specific type of simulation and empirical data collection required to pursue real-life applications. Currently, inhomogeneous stochastic models in use focus on (1) making assumptions about the functional and parametric form of the entries of the transition matrix and (2) collecting or simulating time-sensitive data that inform the choice of the parameters. The model pursued herein indicates that an analysis can be pursued by estimating the rates of growth of the transitional probabilities at an initial time, namely the matrix B .

Our intention in this article, which is based on [9], is to provide a primer for a deeper matrix-theoretic analysis of continuous-time Markov chains. As such, we foresee several possible lines of future research:

- (1) Discover and interpret more dynamic relations among the entries of $P(t)$.
- (2) Relate properties of the ctMC directly to $B = P'(0)$.
- (3) When B is reducible, analyze the asymptotic behavior of the probability distributions in terms of the Frobenius normal form and the reduced graph of B [12].
- (4) Extend and expand the use of group inverses of M -matrices to general continuous-time Markov chains, analogously to the homogeneous case.
- (5) Study initial distributions $z(0)$ in (3.9) that lead to specific monotonic behavior of the state probabilities; this is equivalent to studying the image of $\mathbb{R}_+^n$ under B .

Data Availability Statement: Data sharing is not applicable to this article as no datasets were generated or analyzed during the current study.

References

- [1] R. Bellman, *Introduction to Matrix Analysis*, Second Edition, SIAM, Philadelphia, 1997.
- [2] A. Ben-Israel and T. N. E. Greville, *Generalized Inverses - Theory and Applications*, vol. 15, Springer-Verlag, New York, 2003.
- [3] A. Berman and R. J. Plemmons, *Nonnegative Matrices in the Mathematical Sciences*, SIAM, Philadelphia, 1994.
- [4] I. I. Gikhman and A. V. Skorokhod, *Introduction to the Theory of Random Processes*, W.B. Saunders, Philadelphia, 1969.
- [5] G. Grimmett and D. Stirzaker, *Probability and Random Processes*, Oxford University Press, 2009.
- [6] R.A. Horn and C.R. Johnson, *Matrix Analysis*, Cambridge University Press, 1985.
- [7] R.A. Horn and C.R. Johnson, *Topics in Matrix Analysis*, Cambridge University Press, 1991.
- [8] S. Kirkland and M. Neumann, *Group Inverses of M-Matrices and Their Applications*, CRC Press, 2012.
- [9] Hung Van Le, *Matrix Analysis Primer for Inhomogeneous Markov Chains*, Ph.D dissertation, Washington State University, December 2020.
- [10] J. Lienard, N.S. Strigul, Modeling of hardwood forest in Quebec under dynamic disturbance regimes: a time-inhomogeneous Markov chain approach, *Journal of Ecology*, 104(3):806-816, 2016.
- [11] William S. Stewart, *Introduction to the Numerical Solution of Markov Chains*, Princeton University Press, Princeton, 1994.
- [12] H. Schneider, The influence of the marked reduced graph of a nonnegative matrix on the Jordan form and on related properties: A survey, *Linear Algebra and its Applications*, 84:161-189, 1986.
- [13] Malgorzata Pulka, On the mixing property and the ergodic principle for nonhomogeneous Markov chains, *Linear Algebra and its Applications*, 434:1475-1488, 2011.
- [14] H. M. Taylor and S. Karlin, *An Introduction To Stochastic Modeling*, Academic Press, 1998.
- [15] Richard Varga, *Matrix Iterative Analysis*, Springer, 1963.