J. Appl. Probab. 1–23 (2024)
*doi:*10.1017/jpr.2024.96

THE MARKOV CHAIN EMBEDDING PROBLEM IN A ONE-JUMP SETTING

PHILIPPE CARETTE,* *Universiteit Gent*
MARIE-ANNE GUERRY,** *Vrije Universiteit Brussel*

Abstract

The embedding problem of Markov chains examines whether a stochastic matrix $\mathbf{P}$ can arise as the transition matrix from time 0 to time 1 of a continuous-time Markov chain. When the chain is homogeneous, it checks if $\mathbf{P} = \exp \mathbf{Q}$ for a rate matrix $\mathbf{Q}$ with zero row sums and non-negative off-diagonal elements, called a Markov generator. It is known that a Markov generator may not always exist or be unique. This paper addresses finding $\mathbf{Q}$, assuming that the process has at most one jump per unit time interval, and focuses on the problem of aligning the conditional one-jump transition matrix from time 0 to time 1 with $\mathbf{P}$. We derive a formula for this matrix in terms of $\mathbf{Q}$ and establish that for any $\mathbf{P}$ with non-zero diagonal entries, a unique $\mathbf{Q}$, called the $\mathbb{1}$-generator, exists. We compare the $\mathbb{1}$-generator with the one-jump rate matrix from Jarrow, Lando, and Turnbull (1997), showing which is a better approximate Markov generator of $\mathbf{P}$ in some practical cases.

Keywords: Markov chain; embedding problem; Markov generator; matrix exponential

2020 Mathematics Subject Classification: Primary 60J27; 60J10
Secondary 15A16

1. Introduction

The embedding problem of Markov chains is a long-standing challenge. It involves investigating whether a given stochastic matrix $\mathbf{P}$ can be considered as the matrix of transition probabilities from time 0 to time 1 of a continuous-time Markov chain. The problem was first proposed by Elfving [12] with a first in-depth study by Kingman [24]. If the Markov chain is homogeneous, this problem boils down to checking whether $\mathbf{P}$ is the exponential of some matrix $\mathbf{Q}$ having all non-negative off-diagonal entries and zero row-sums, called a rate matrix. This matrix $\mathbf{Q}$ then represents the transition rates of the underlying continuous-time homogeneous Markov chain (CTHMC). If $\exp \mathbf{Q} = \mathbf{P}$ for some rate matrix $\mathbf{Q}$, it is said that $\mathbf{P}$ is embeddable and that $\mathbf{Q}$ is a Markov generator of $\mathbf{P}$. For the non-homogeneous case, it has been proved that if $\mathbf{P}$ belongs to the interior of the set of embeddable matrices, then $\mathbf{P}$ is equal to a finite product of exponentials of extreme rate matrices, i.e. rate matrices with only one off-diagonal element strictly positive; see [20]. This characterization is known as a Bang-Bang representation.

Received 11 July 2023; accepted 12 September 2024.
* Postal address: Department of Economics, Ghent University, Sint-Pietersplein 5, B-9000 Gent, Belgium. Email address: Philippe.Carette@ugent.be
** Postal address: Department of Business Technology and Operations, Vrije Universiteit Brussel, Pleinlaan 2, B-1050 Brussels, Belgium. Email address: Marie-Anne.Guerry@vub.be

The characterizations described above are general and theoretically significant, yet they are difficult to verify practically. As of now, a complete solution to the embedding problem in terms of the matrix elements remains to be found. Currently, embeddability criteria have been established in the homogeneous case for chains up to four states; see [4], [5], [7], and [21]. In the non-homogeneous case, research has focused on the Bang-Bang representation of embeddable stochastic matrices, leading to more concrete results for the 3×3 case; see [13] and [22]. In this paper we focus on the homogeneous embedding problem and will therefore omit the word 'homogeneous'.

The issue of embeddability presents significant challenges in various aspects. Even if $\mathbf{P}$ is confirmed as embeddable, there could be more than one Markov generator associated with it. Using a CTHMC as a foundational model necessitates decisions on handling non-embeddable transition matrices or multiple Markov generators. In the case of non-embeddability, a regularization algorithm is proposed to obtain a rate matrix $\mathbf{Q}$ that minimizes $\|\mathbf{P} - \exp \mathbf{Q}\|$; see [10], [17], and [25]. When there are multiple Markov generators, the practitioner can try to identify and select the rate matrix reflecting the nature of the system under study; see [28].

Rather than accounting for the system's nature when determining an appropriate Markov generator for $\mathbf{P}$, model-specific embedding problems have recently been considered for certain classes of stochastic matrices, leading to further results. For example, the elements of the transition matrix can be limited to certain values reflecting the characteristics of the system under study; see [1], [2], [3], [19], and [27].

An alternative train of thought is to derive a rate matrix from the empirical transition matrix based on the assumption that there is a maximum of one transition per unit of time. Considering that a discrete-time Markov chain allows only one transition or jump per time step, it seems reasonable to view the underlying continuous-time process through a similar perspective. Furthermore, this one-jump hypothesis has not only theoretical but also practical grounds. In financial risk management, credit rating migration models evaluate how borrowers move between credit rating categories over time. These changes in credit ratings usually occur gradually and infrequently. Similarly, in healthcare, transitions between health states occur slowly over extended periods, resulting in infrequent state changes within short time intervals. In their study of credit risk migration, Jarrow, Lando, and Turnbull (JLT) assume that each firm has made zero or one transition per year; see [18]. Fundamentally, their approach designates the empirical transition probability p_{ij} to represent the probability of jumping from state i to state j by or before $t = 1$. This facilitates the formulation of an approximate Markov generator for the empirical transition matrix in closed form.

In this study, we explore the one-jump hypothesis through a data-centric approach. Our only assumption is that the sampled entities or individuals adhere to the one-jump hypothesis; we do not presume any specific number of jumps for the entire population under study. Considering that a CTHMC model allows for several jumps in any given period, this approach appears to be more fitting if the goal is to employ the CTHMC to explore population dynamics. Hence we study a conditional version of the embedding problem by aligning the empirical transition matrix with the *conditional* transition matrix over a single unit of time of the CTHMC *given the event* that at most one jump has occurred during a unit length time interval. The main result of this paper is that, regardless of the number of states, precisely one rate matrix solves our conditional embedding problem in the case where the empirical transition matrix has non-zero diagonal elements; see Theorem 3.2. We also describe how to obtain this rate matrix from the empirical transition matrix. In summary, the conditional embedding problem, when considered in a one-jump context, yields a solution that avoids the regularization and identification challenges associated with the original embedding problem.

Our paper is structured as follows. In Section 2 we formalize and define the notion of a conditional one-step probability matrix given an event of interest. We derive an explicit expression for the conditional transition probabilities for the event that the number of jumps between t and $t+1$ is zero or one. In Section 3 we dive into the conditional one-jump version of the embedding problem, introducing and focusing on the concepts of $\mathbb{1}$-embeddability and $\mathbb{1}$-generator. Our key finding in this section asserts that any stochastic matrix of order n with non-zero diagonal elements is $\mathbb{1}$-embeddable and possesses a unique $\mathbb{1}$-generator. Section 4 highlights the difference in one-jump setting approaches leading to the $\mathbb{1}$-generator and the rate matrix $\mathbf{Q}_{\mathrm{JLT}}$ derived by Jarrow *et al.* [18]. We also show which of these two matrices serves as a better approximate Markov generator of the empirical transition matrix for a number of specific cases. We conclude with a discussion in Section 5. The supporting lemmas and their proofs are gathered in the Appendix.

2. Conditional transition probabilities

To state the conditional embedding problem, we first introduce the concept of conditional transition probability.

Consider a continuous-time homogeneous Markov chain ('CTHMC') $(X_t)_{t\geq 0}$ on a probability triple $(\Omega, \mathscr{F}, \mathbb{P})$ with state space $\mathcal{S} = \{1, 2, \ldots, n\}$.

Definition 2.1. Let $t \geq 0$ and $\mathcal{E}_t \in \mathscr{F}$. We call the matrix $\mathbf{P}^{\mathcal{E}_t}(t)$ with elements

$$p_{ij}^{\mathcal{E}_t}(t) = \mathbb{P}(X_{t+1} = j \mid X_t = i,\ \mathcal{E}_t), \quad i, j \in \mathcal{S},$$

the *conditional transition probability matrix from time t to time* $t+1$ *given the event* $\mathcal{E}_t$ of the CTHMC.

Remark 2.1. The (classical) transition probabilities $p_{ij}(t) = \mathbb{P}(X_{t+1} = j \mid X_t = i)$ can be obtained using $\mathcal{E}_t := \Omega$, i.e. $p_{ij}(t) = p_{ij}^{\Omega}(t)$.

In this paper, for given t, the event of interest is the occurrence of at most one jump of the CTHMC in the interval $(t, t+1]$, i.e. $\mathcal{E}_t = \{N_{t+1} - N_t \leq 1\}$, where N_s represents the number of state changes ('jumps') of the CTHMC up to and including time $s \geq 0$. For simplicity, we denote $\mathbf{P}^{\mathbb{1}}(t) := \mathbf{P}^{\mathcal{E}_t}(t)$.

For a CTHMC with rate matrix $\mathbf{Q}$, the connection between $\mathbf{P}^{\mathbb{1}}(t) = (p_{ij}^{\mathbb{1}}(t))$ and $\mathbf{Q}$ is established by the following proposition.

Proposition 2.1. *For a CTHMC with rate matrix* $\mathbf{Q} = (q_{ij})$*, it holds that*

$$p_{ij}^{\mathbb{1}}(t) = \frac{p_{ij}^{*}}{\sum_{k\in\mathcal{S}} p_{ik}^{*}} \quad \textit{for all } i, j \in \mathcal{S} \textit{ and } t \geq 0,$$

where

$$p_{ij}^{*} = \begin{cases} q_{ij}\,\tau(q_{ii}, q_{jj}) & \text{if } i \neq j, \\ \tau(q_{ii}, q_{ii}) & \text{if } i = j, \end{cases} \tag{2.1}$$

https://doi.org/10.1017/jpr.2024.96 Published online by Cambridge University Press

and where the function $\tau : \mathbb{R}^2 \to \mathbb{R}$ *is defined as*

$$\tau(x, y) = \int_0^1 \mathrm{e}^{ux+(1-u)y}\,\mathrm{d}u = \begin{cases} \dfrac{\mathrm{e}^x - \mathrm{e}^y}{x - y} & \textit{if } x \neq y, \\ \mathrm{e}^x & \textit{if } x = y. \end{cases} \tag{2.2}$$

Proof. Using the definition of conditional probability, we have

$$\mathbb{P}(A \mid B \cap C) = \frac{\mathbb{P}(A \cap B \mid C)}{\mathbb{P}(B \mid C)}, \quad \text{if } \mathbb{P}(B \mid C) > 0,$$

whence

$$p^{\mathbb{1}}_{ij}(t) = \mathbb{P}(X_{t+1} = j \mid X_t = i,\ \mathcal{E}_t) = \frac{\mathbb{P}(X_{t+1} = j,\ \mathcal{E}_t \mid X_t = i)}{\mathbb{P}(\mathcal{E}_t \mid X_t = i)}.$$

Let us denote

$$p^*_{ij}(t) := \mathbb{P}(X_{t+1} = j,\ \mathcal{E}_t \mid X_t = i). \tag{2.3}$$

By the sum rule for disjoint events, we have

$$p^{\mathbb{1}}_{ij}(t) = \frac{p^*_{ij}(t)}{\sum_{k\in\mathcal{S}} p^*_{ik}(t)},$$

so it remains to be shown that $p^*_{ij}(t) = p^*_{ij}$ for all $i, j \in \mathcal{S}$, where p^*_{ij} is given by (2.1).

Denote $\mathcal{S}$ as the disjoint union of $\mathcal{S}^{\mathrm{T}} := \{i \in \mathcal{S} \colon q_{ii} < 0\}$ (the subset of transient states) and $\mathcal{S}^{\mathrm{A}} := \{i \in \mathcal{S} \colon q_{ii} = 0\}$ (the subset of absorbing states). Let $Y_t := \inf\{s > 0 \colon X_{t+s} \neq X_t\}$ be the residual time in state X_t until the first jump after t. In a CTHMC, the holding times are memoryless, hence the conditional distribution of Y_t, given state X_t, does not depend on time t. For $k \in \mathcal{S}^{\mathrm{T}}$, let f_k denote the conditional probability density function of Y_t given $X_t = k$ and F_k its associated cumulative distribution function. Lastly, we shall write s_{ij} to denote the transition probability from state i to state j conditional on transitioning out of state i.

We start with the case $i, j \in \mathcal{S}^{\mathrm{T}}$, $i \neq j$. Then (2.3) entails

$$p^*_{ij}(t) = \mathbb{P}(Y_t \leq 1,\, X_{t+Y_t} = j,\, Y_{t+Y_t} > 1 - Y_t \mid X_t = i),$$

and hence, by marginalizing on Y_t, we obtain

$$\begin{aligned} p^*_{ij}(t) &= \int_0^1 \mathbb{P}(X_{t+u} = j,\, Y_{t+u} > 1 - u \mid Y_t = u,\, X_t = i) f_i(u)\,\mathrm{d}u \\ &= \int_0^1 \mathbb{P}(Y_{t+u} > 1 - u \mid X_{t+u} = j,\, Y_t = u,\, X_t = i)\,\mathbb{P}(X_{t+u} = j \mid Y_t = u,\, X_t = i) f_i(u)\,\mathrm{d}u \\ &= \int_0^1 (1 - F_j(1-u))\, s_{ij}\, f_i(u)\,\mathrm{d}u. \end{aligned}$$

Since both residual and holding times in a transient state k have an exponential distribution with rate parameter $-q_{kk}$ and since $s_{ij} = q_{ij}/{-q_{ii}}$, we arrive at

$$p^*_{ij}(t) = \int_0^1 \mathrm{e}^{q_{jj}(1-u)}\,\frac{q_{ij}}{-q_{ii}}\,(-q_{ii})\mathrm{e}^{q_{ii}u}\,\mathrm{d}u = q_{ij}\int_0^1 \mathrm{e}^{q_{ii}u+q_{jj}(1-u)}\,\mathrm{d}u = q_{ij}\,\tau(q_{ii}, q_{jj}) = p^*_{ij},$$

where p^*_{ij} and τ are given by (2.1) and (2.2).

https://doi.org/10.1017/jpr.2024.96 Published online by Cambridge University Press

We now turn to the case $i \in \mathcal{S}^{\mathrm{T}}$ and $j \in \mathcal{S}^{\mathrm{A}}$. Then, using (2.3) and marginalizing on Y_t, we get

$$\begin{aligned} p^*_{ij}(t) &= \mathbb{P}(Y_t \le 1, X_{t+Y_t} = j \mid X_t = i) \\ &= \int_0^1 \mathbb{P}(X_{t+u} = j \mid Y_t = u, X_t = i) f_i(u)\,\mathrm{d}u \\ &= \int_0^1 s_{ij} f_i(u)\,\mathrm{d}u \\ &= \int_0^1 \frac{q_{ij}}{-q_{ii}} (-q_{ii}) \mathrm{e}^{q_{ii}u}\,\mathrm{d}u \\ &= q_{ij} \int_0^1 \mathrm{e}^{q_{ii}u}\,\mathrm{d}u = q_{ij}\tau(q_{ii}, 0). \end{aligned}$$

Consequently, $p^*_{ij}(t) = q_{ij}\tau(q_{ii}, q_{jj}) = p^*_{ij}$, as $q_{jj} = 0$.

When $i \in \mathcal{S}^{\mathrm{A}}$ and $j \in \mathcal{S}$, $j \neq i$, we obviously have $p^*_{ij}(t) = 0$. Hence

$$p^*_{ij}(t) = q_{ij}\tau(q_{ii}, q_{jj}) = p^*_{ij},$$

as $q_{ij} = 0$ under these conditions.

Finally, if $i \in \mathcal{S}^{\mathrm{T}}$, (2.3) yields

$$p^*_{ii}(t) = \mathbb{P}(Y_t > 1 \mid X_t = i) = 1 - F_i(1) = \mathrm{e}^{q_{ii}} = \tau(q_{ii}, q_{ii}) = p^*_{ii},$$

and if $i \in \mathcal{S}^{\mathrm{A}}$, then

$$p^*_{ii}(t) = 1 = \tau(0, 0) = \tau(q_{ii}, q_{ii}) = p^*_{ii}. \qquad \square$$

We note that Minin and Suchard [26] arrive at the same formula for p^*_{ij}, using a matrix differential equation involving the matrix with elements $q_{ij}(k, t)$ representing the joint probability of transitioning in k jumps from state i to state j within a time-interval of length t.

According to Proposition 2.1, the matrix $\mathbf{P}^{\mathbb{1}}(t)$ does not depend on t. We will therefore write $\mathbf{P}^{\mathbb{1}} := \mathbf{P}^{\mathbb{1}}(t)$ and refer to it as a *conditional one-jump transition matrix*.

Corollary 2.1. *For all $i \in \mathcal{S}$, we have $p^{\mathbb{1}}_{ii} > 0$.*

Proof. This follows from the fact that $p^*_{ii} = \tau(q_{ii}, q_{ii}) = \mathrm{e}^{q_{ii}} > 0$. $\square$

Remark 2.2. Corollary 2.1 can also be justified as follows. The only way to go from state i at time t to state i at time $t + 1$, given the event $\{N_{t+1} - N_t \le 1\}$, is to stay in that state for the entire time interval. The probability of this event is non-zero due to the exponential distribution governing the holding time in a state.

According to Proposition 2.1, the conditional one-jump transition matrix $\mathbf{P}^{\mathbb{1}}$ is fully specified by the rate matrix $\mathbf{Q}$ of the CTHMC. In what follows, and when needed, we explicitly indicate this dependence using the notation $\mathbf{P}^{\mathbb{1}}(\mathbf{Q})$.

3. Conditional embedding problem

The following question then naturally arises. Given a stochastic matrix $\mathbf{P}$, is there a rate matrix $\mathbf{Q}$ such that $\mathbf{P}^{\mathbb{1}}(\mathbf{Q}) = \mathbf{P}$? And if so, is this rate matrix $\mathbf{Q}$ unique?

It will be helpful to introduce some terminology before proceeding.

https://doi.org/10.1017/jpr.2024.96 Published online by Cambridge University Press

Definition 3.1. A stochastic matrix $\mathbf{P}$ is called $\mathbb{1}$-embeddable if and only if there exists a CTHMC with rate matrix $\mathbf{Q}$ satisfying $\mathbf{P}=\mathbf{P}^{\mathbb{1}}(\mathbf{Q})$. This rate matrix is called a $\mathbb{1}$-generator of $\mathbf{P}$.

For a transition matrix $\mathbf{P}$ that is not embeddable, a $\mathbb{1}$-*generator* can be seen as a solution to a conditional version of the embedding problem where the rate matrix $\mathbf{Q}$ satisfies $\mathbf{P}^{\mathbb{1}}(\mathbf{Q})=\mathbf{P}$. Corollary 2.1 establishes that a $\mathbb{1}$-embeddable stochastic matrix must not contain a zero on its main diagonal. Consequently, our subsequent analysis will focus on stochastic matrices $\mathbf{P}=(p_{ij})$ where $p_{ii}>0$ for all i. Note that this necessary condition for $\mathbb{1}$-embeddability is not more restrictive than the necessary embedding condition $\prod_{i=1}^{n} p_{ii} \geq \det P > 0$ as formulated by Goodman in [16].

Interestingly, the off-diagonal elements of a $\mathbb{1}$-generator of $\mathbf{P}$ are uniquely defined by its diagonal elements and the elements of $\mathbf{P}$. To express this relationship, we introduce the function $\rho\colon \mathbb{R}_+^2 \to \mathbb{R}_+$, defined as follows:

$$\rho(x, y)=\frac{\mathrm{e}}{\tau(1-\ln x,\, 1-\ln y)}=\begin{cases} xy\dfrac{\ln x-\ln y}{x-y} & \text{if } x\neq y, \\ x & \text{if } x=y. \end{cases} \tag{3.1}$$

Proposition 3.1. *Suppose* $\mathbf{P}=(p_{ij})$ *is an* $n\times n$ *stochastic matrix that satisfies* $p_{ii}>0$ *for all* i*. If* $\mathbf{Q}=(q_{ij})$ *is a* $\mathbb{1}$*-generator of* $\mathbf{P}$*, then*

$$q_{ij}=\frac{\rho(\theta_i, \theta_j)p_{ij}}{\theta_i p_{ii}} \quad \textit{for all } i\neq j,$$

where $\theta_i=\mathrm{e}^{1-q_{ii}}$ *for all* i *and the function* $\rho\colon \mathbb{R}_+^2\to\mathbb{R}_+$ *is given by* (3.1).

Proof. Suppose $\mathbf{Q}$ is a $\mathbb{1}$-generator of $\mathbf{P}=(p_{ij})$ and let $i\neq j$. Then, according to Proposition 2.1, we have

$$\frac{p_{ij}}{p_{ii}}=\frac{p_{ij}^{\mathbb{1}}}{p_{ii}^{\mathbb{1}}}=\frac{p_{ij}^{*}}{p_{ii}^{*}}=\frac{q_{ij}\tau(q_{ii}, q_{jj})}{\tau(q_{ii}, q_{ii})}.$$

Consequently, since $\tau(q_{ii}, q_{ii})=\mathrm{e}^{q_{ii}}=\mathrm{e}/\theta_i$ and $\tau(q_{ii}, q_{jj})=\tau(1-\ln\theta_i,\, 1-\ln\theta_j)=\mathrm{e}/\rho(\theta_i, \theta_j)$, we get

$$q_{ij}=\frac{\tau(q_{ii}, q_{ii})p_{ij}}{\tau(q_{ii}, q_{jj})p_{ii}}=\frac{(\mathrm{e}/\theta_i)p_{ij}}{(\mathrm{e}/\rho(\theta_i, \theta_j))p_{ii}}=\frac{\rho(\theta_i, \theta_j)p_{ij}}{\theta_i p_{ii}} \quad \text{for all } i\neq j.$$

□

The outcome of Proposition 3.2 provides a requirement concerning the diagonal elements of any $\mathbb{1}$-generator of $\mathbf{P}$.

Proposition 3.2. *Let* $\mathbf{P}=(p_{ij})$ *be an* $n\times n$ *stochastic matrix that satisfies* $p_{ii}>0$ *for all* i*. If* $\mathbf{Q}=(q_{ij})$ *is a* $\mathbb{1}$*-generator of* $\mathbf{P}$*, then the n-tuple* $(\mathrm{e}^{1-q_{11}}, \ldots, \mathrm{e}^{1-q_{nn}})$ *is a fixed point of the vector function* $\mathbf{T}=(T_1, \ldots, T_n)\colon \mathbb{R}_+^n\to\mathbb{R}_+^n$*, which is defined as*

$$T_i(x_1, \ldots, x_n)=\exp W_0\left(\frac{1}{p_{ii}}\sum_{j\in\mathcal{S}} p_{ij}\rho(x_i, x_j)\right) \quad \textit{for all } i\in\mathcal{S}, \tag{3.2}$$

where W_0 *represents the principal branch of the Lambert W function, and where the function* $\rho\colon \mathbb{R}_+^2\to\mathbb{R}_+$ *is defined as in* (3.1).

https://doi.org/10.1017/jpr.2024.96 Published online by Cambridge University Press

Proof. Denote $\theta_i = \mathrm{e}^{1-q_{ii}}$ for all $i \in \mathcal{S}$. Then $\theta_i > 0$ and $q_{ii} = 1 - \ln \theta_i$ for all i. Using Proposition 3.1 and the fact that $\mathbf{Q}$ is a rate matrix, we then have

$$-1 + \ln \theta_i = -q_{ii} = \sum_{j\in\mathcal{S}\setminus\{i\}} q_{ij} = \sum_{j\in\mathcal{S}\setminus\{i\}} \frac{\rho(\theta_i, \theta_j)p_{ij}}{\theta_i p_{ii}} \quad \text{for all } i \in \mathcal{S},$$

which can be rewritten, using the fact that $\rho(\theta_i, \theta_i) = \theta_i$, as

$$\theta_i \ln \theta_i = \frac{1}{p_{ii}} \sum_{j\in\mathcal{S}} p_{ij}\rho(\theta_i, \theta_j) \quad \text{for all } i \in \mathcal{S}. \tag{3.3}$$

Using the principal branch W_0 of the Lambert W function (which is the multi-valued inverse of the function $w \mapsto w\mathrm{e}^w$ ($w \in \mathbb{C}$); see [9]), we find that

$$\ln \theta_i = W_0\left(\frac{1}{p_{ii}} \sum_{j\in\mathcal{S}} p_{ij}\rho(\theta_i, \theta_j)\right) \quad \text{for all } i \in \mathcal{S},$$

which proves the result. □

Propositions 3.1 and 3.2 imply that a $\mathbb{1}$-generator of $\mathbf{P}$ defines a fixed point of the vector function $\mathbf{T}$. The reverse is also true, as stated in the following proposition.

Proposition 3.3. *Suppose $\mathbf{P} = (p_{ij})$ is a stochastic matrix that satisfies $p_{ii} > 0$ for all $i \in \mathcal{S}$. Let $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n)$ be a fixed point of the vector function $\mathbf{T}\colon \mathbb{R}_+^n \to \mathbb{R}_+^n$, defined in (3.2). Then the matrix $\mathbf{Q} = (q_{ij})$ with elements*

$$q_{ii} = 1 - \ln \theta_i, \quad q_{ij} = \frac{\rho(\theta_i, \theta_j)p_{ij}}{\theta_i p_{ii}}, \quad i \neq j, \tag{3.4}$$

where ρ is defined by (3.1), is a $\mathbb{1}$-generator of $\mathbf{P}$.

Proof. Let $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_n) \in \mathbb{R}_+^n$ be a fixed point of $\mathbf{T}$ and let the matrix $\mathbf{Q}$ be defined by (3.4). We first show that $\mathbf{Q}$ is a rate matrix. By (3.4), all off-diagonal elements of $\mathbf{Q}$ are non-negative. Since $T_i(\boldsymbol{\theta}) = \theta_i$, we have

$$\ln \theta_i = W_0\left(\frac{1}{p_{ii}} \sum_{j\in\mathcal{S}} p_{ij}\rho(\theta_i, \theta_j)\right),$$

yielding

$$\theta_i \ln \theta_i = \frac{1}{p_{ii}} \sum_{j\in\mathcal{S}} p_{ij}\rho(\theta_i, \theta_j),$$

by definition of the Lambert W_0 function. Using (3.4) and since $\rho(\theta_i, \theta_i) = \theta_i$, we can rewrite this equation as

$$\theta_i(1 - q_{ii}) = \theta_i + \frac{1}{p_{ii}} \sum_{j\in\mathcal{S}\setminus\{i\}} q_{ij}\theta_i p_{ii}.$$

After simplification, we get $q_{ii} = -\sum_{j\in\mathcal{S}\setminus\{i\}} q_{ij}$. Thus $\mathbf{Q}$ has zero row sums. Consequently, $\mathbf{Q}$ is a rate matrix.

https://doi.org/10.1017/jpr.2024.96 Published online by Cambridge University Press

It remains to be proved that $p^{\mathbb{1}}_{ij}(\mathbf{Q}) = p_{ij}$ for all i and j. Let p^*_{ij} be defined by (2.1) and (2.2). Using (3.4) and (3.1), we have

$$p^*_{ij} = q_{ij}\tau(q_{ii}, q_{jj}) = \frac{\rho(\theta_i, \theta_j)p_{ij}}{\theta_i p_{ii}}\tau(1-\ln\theta_i, 1-\ln\theta_j) = \frac{\mathrm{e}\, p_{ij}}{\theta_i p_{ii}}, \quad \text{if } i \neq j,$$

and

$$p^*_{ii} = \tau(q_{ii}, q_{ii}) = \mathrm{e}^{q_{ii}} = \mathrm{e}^{1-\ln\theta_i} = \frac{\mathrm{e}}{\theta_i} \quad \text{for all } i,$$

hence

$$p^*_{ij} = \frac{\mathrm{e}\, p_{ij}}{\theta_i p_{ii}} \quad \text{for all } i \text{ and } j.$$

Consequently, by Proposition 2.1 and since $\mathbf{P}$ is a stochastic matrix, we now have

$$p^{\mathbb{1}}_{ij}(\mathbf{Q}) = \frac{p^*_{ij}}{\sum_{k\in\mathcal{S}} p^*_{ik}} = \frac{\mathrm{e}\, p_{ij}}{\theta_i p_{ii}} \Bigg/ \frac{\mathrm{e}}{\theta_i p_{ii}} = p_{ij},$$

which concludes the proof. □

By combining Propositions 3.1, 3.2, and 3.3, we establish a one-to-one relationship between the $\mathbb{1}$-generators of $\mathbf{P}$ and the fixed points of the vector function $\mathbf{T}$. The subsequent lemma describes the properties of the vector function $\mathbf{T}$, which are crucial to determining its number of fixed points.

Lemma 3.1. *Let* $\mathbf{P} = (p_{ij})$ *be an* $n \times n$ *stochastic matrix. Let* $\delta = \min\{p_{ii},\ i \in \mathcal{S}\}$ *and* $\Delta = \max\{p_{ii},\ i \in \mathcal{S}\}$. *Suppose* $p_{ii} > 0$ *for all i. Consider the vector function* $\mathbf{T}\colon \mathbb{R}^n_+ \to \mathbb{R}^n_+$, *defined as in* (3.2), *and the set*

$$\mathcal{X} = \{(x_1, \ldots, x_n) \in \mathbb{R}^n \colon \mathrm{e}^{1/\Delta} \le x_i \le \mathrm{e}^{1/\delta}\ \forall i\}. \tag{3.5}$$

Then

(i) *every fixed point of* $\mathbf{T}$ *belongs to* $\mathcal{X}$.

(ii) $\mathbf{T}$ *maps* $\mathcal{X}$ *into* $\mathcal{X}$.

Proof. (i) Let $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n) \in \mathbb{R}^n_+$ be a fixed point of $\mathbf{T}$. Let $m = \min\{\theta_1, \ldots, \theta_n\}$ and $M = \max\{\theta_1, \ldots, \theta_n\}$. We shall prove that $m \ge \mathrm{e}^{1/\Delta}$ and that $M \le \mathrm{e}^{1/\delta}$.

Let r be an index such that $\theta_r = m$. Then, by Lemma A.2(iv), we have $\rho(\theta_r, \theta_j) \ge m$ for all j. Since $T_r(\boldsymbol{\theta}) = \theta_r$, we have

$$\ln\theta_r = W_0\left(\frac{1}{p_{rr}}\sum_{j\in\mathcal{S}} p_{rj}\rho(\theta_r, \theta_j)\right),$$

yielding

$$\theta_r \ln\theta_r = \frac{1}{p_{rr}}\sum_{j\in\mathcal{S}} p_{rj}\rho(\theta_r, \theta_j)$$

by definition of the Lambert W_0 function. Using the fact that $\sum_{j\in\mathcal{S}} p_{rj} = 1$, we then obtain

$$m\ln m = \theta_r\ln\theta_r = \frac{1}{p_{rr}}\sum_{j\in\mathcal{S}} p_{rj}\rho(\theta_r, \theta_j) \ge \frac{m}{p_{rr}} \ge \frac{m}{\Delta},$$

which implies $\ln m \geq 1/\Delta$ and thus $m \geq \mathrm{e}^{1/\Delta}$. To prove the second inequality, let s be an index such that $\theta_s = M$. Then, by Lemma A.2(iv), we have $\rho(\theta_s, \theta_j) \leq M$ for all j. Hence, by $T_s(\theta) = \theta_s$ and the unit row sum property of $\mathbf{P}$,

$$M \ln M = \theta_s \ln \theta_s = \frac{1}{p_{ss}} \sum_{j \in \mathcal{S}} p_{sj} \rho(\theta_s, \theta_j) \leq \frac{M}{p_{ss}} \leq \frac{M}{\delta},$$

which yields $\ln M \leq 1/\delta$ and thus $M \leq \mathrm{e}^{1/\delta}$.

(ii) Let $(x_1, \ldots, x_n) \in \mathcal{X}$. By Lemma A.2(iv),

$$\mathrm{e}^{1/\Delta} \leq \min\{x_i, x_j\} \leq \rho(x_i, x_j) \leq \max\{x_i, x_j\} \leq \mathrm{e}^{1/\delta} \quad \text{for all } i \text{ and } j.$$

Then, since $\mathbf{P}$ has unit row sums, we have for all i

$$\frac{\mathrm{e}^{1/\Delta}}{\Delta} \leq \frac{\mathrm{e}^{1/\Delta}}{p_{ii}} \leq \frac{1}{p_{ii}} \sum_{j \in \mathcal{S}} p_{ij} \rho(x_i, x_j) \leq \frac{\mathrm{e}^{1/\delta}}{p_{ii}} \leq \frac{\mathrm{e}^{1/\delta}}{\delta}.$$

Now, W_0 and exp are increasing functions, and therefore

$$\exp W_0\left(\frac{1}{\Delta}\mathrm{e}^{1/\Delta}\right) \leq T_i(x_1, \ldots, x_n) \leq \exp W_0\left(\frac{1}{\delta}\mathrm{e}^{1/\delta}\right) \quad \text{for all } i.$$

The result now follows by applying the property $W_0(x\mathrm{e}^x) = x$ for $x > 0$. □

Lemma 3.1 implies that the maximal and minimal diagonal elements of $\mathbf{P}$ impose a lower and upper bound on the diagonal elements of the $\mathbb{1}$-generators of $\mathbf{P}$.

Proposition 3.4. *Let $\mathbf{P} = (p_{ij})$ be an $n \times n$ stochastic matrix. Let $\Delta = \max\{p_{ii},\ i \in \mathcal{S}\}$ and $\delta = \min\{p_{ii},\ i \in \mathcal{S}\}$. Suppose $p_{ii} > 0$ for all i. Then, if $\mathbf{Q} = (q_{ij})$ is a $\mathbb{1}$-generator of $\mathbf{P}$, we have*

$$1 - \frac{1}{\delta} \leq q_{ii} \leq 1 - \frac{1}{\Delta} \quad \textit{for all } i.$$

Proof. If $\mathbf{Q} = (q_{ij})$ is a $\mathbb{1}$-generator of $\mathbf{P}$, we have by Proposition 3.2 that the vector $(\theta_1, \ldots, \theta_n)$, where $\theta_i = \mathrm{e}^{1-q_{ii}}$ for all i, is a fixed point of $\mathbf{T}$. Applying Lemma 3.1(i), we have $\mathrm{e}^{1/\Delta} \leq \theta_i \leq \mathrm{e}^{1/\delta}$ for all i, from which the result now follows. □

With regard to the number of fixed points of $\mathbf{T}$, we now prove the following important result.

Theorem 3.1. *Suppose $\mathbf{P} = (p_{ij})$ is an $n \times n$ stochastic matrix that satisfies $p_{ii} > 0$ for all i. Then the vector function $\mathbf{T}\colon \mathbb{R}_+^n \to \mathbb{R}_+^n$, defined as in (3.2), has a unique fixed point.*

Proof. According to Lemma 3.1(ii), $\mathbf{T}$ maps the compact convex set $\mathcal{X} \subset \mathbb{R}_+^n$, defined by (3.5), into itself. In addition, $\mathbf{T}$ is continuous as the function ρ, defined by (3.1), is continuous (Lemma A.2(ii)) and continuity is preserved by linear combination and composition of continuous functions. Hence, according to the Brouwer fixed point theorem, $\mathbf{T}$ has a fixed point. By the definition of $\mathbf{T}$, this fixed point must have all positive components. We now show that the function $\mathbf{g}\colon \mathbb{R}_+^n \to \mathbb{R}^n$ defined as $\mathbf{g} = \mathbf{T} - \mathbf{Id}$, where $\mathbf{Id}\colon \mathbb{R}_+^n \to \mathbb{R}_+^n$ is the identity mapping, satisfies all conditions of Theorem 3.1 in [23]. This theorem presents sufficient conditions for the function $\mathbf{g}$ to have at most one vector $\mathbf{x} \in \mathbb{R}_+^n$ with $\mathbf{g}(\mathbf{x}) = (0, \ldots, 0)$. These conditions

https://doi.org/10.1017/jpr.2024.96 Published online by Cambridge University Press

are that (a) $\mathbf{g}$ is quasi-increasing and (b) $\mathbf{g}$ is strictly R-concave. Both (a) and (b) are proved in the Appendix; see Lemma A.4. So, we have established that $\mathbf{T}$ has precisely one fixed point. $\square$

We are now ready to state and prove our main result in the following theorem.

Theorem 3.2. *Suppose $\mathbf{P} = (p_{ij})$ is an $n \times n$ stochastic matrix that satisfies $p_{ii} > 0$ for all i. Then $\mathbf{P}$ has precisely one $\mathbb{1}$-generator. Moreover, this $\mathbb{1}$-generator $\mathbf{Q} = (q_{ij})$ is given by*

$$q_{ii} = 1 - \ln \theta_i, \quad q_{ij} = \frac{\rho(\theta_i, \theta_j) p_{ij}}{\theta_i p_{ii}}, \quad i \neq j, \tag{3.6}$$

where the scalar function $\rho\colon \mathbb{R}_+^2 \to \mathbb{R}_+$ is defined by (3.1), and $(\theta_1, \ldots, \theta_n)$ is the unique fixed point of the vector function $\mathbf{T}\colon \mathbb{R}_+^n \to \mathbb{R}_+^n$ defined by (3.2).

Proof. We first prove that $\mathbf{P}$ has a $\mathbb{1}$-generator. By Theorem 3.1, the vector function $\mathbf{T}$ has a unique fixed point $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_n) \in \mathbb{R}_+^n$. Starting from $\boldsymbol{\theta}$, construct the matrix $\mathbf{Q} = (q_{ij})$ according to (3.6). Then $\mathbf{Q}$ is a $\mathbb{1}$-generator of $\mathbf{P}$, by Proposition 3.3.

To prove the uniqueness of the $\mathbb{1}$-generator, suppose that $\mathbf{P}$ has $\mathbb{1}$-generators $\mathbf{R} = (r_{ij})$ and $\mathbf{S} = (s_{ij})$. Then, by Proposition 3.2, the vectors $\theta_{\mathbf{R}} = (\mathrm{e}^{1-r_{11}}, \ldots, \mathrm{e}^{1-r_{nn}})$ and $\theta_{\mathbf{S}} = (\mathrm{e}^{1-s_{11}}, \ldots, \mathrm{e}^{1-s_{nn}})$ are fixed points of the vector function $\mathbf{T}$. By Theorem 3.1, $\theta_{\mathbf{R}} = \theta_{\mathbf{S}}$. Hence, by Proposition 3.1, we must have $\mathbf{R} = \mathbf{S}$.

Finally, the requirement that the elements of a $\mathbb{1}$-generator comply with (3.6) is inferred from Propositions 3.1 and 3.2. The proof is now complete. $\square$

Since $p_{ii} > 0$ for all i is a necessary condition for embeddability (see [16]), Theorem 3.2 immediately establishes a link between embeddability and $\mathbb{1}$-embeddability.

Corollary 3.1. *An embeddable stochastic matrix is $\mathbb{1}$-embeddable.*

According to Theorem 3.2, the unique $\mathbb{1}$-generator is entirely characterized by the fixed point of the function $\mathbf{T}\colon \mathcal{X} \to \mathcal{X}$. When $\mathbf{T}$ acts as a contraction on $\mathcal{X}$, the Banach fixed point theorem ensures the convergence of the fixed point iteration method, which can then be used to estimate $(\theta_1, \ldots, \theta_n)$. In the case when not all diagonal elements of $\mathbf{P}$ are equal, Proposition 3.5 produces a necessary condition for $\mathbf{T}$ to behave as a contraction under the infinity vector norm. This condition is based on the largest and smallest diagonal elements of the matrix $\mathbf{P}$.

Proposition 3.5. *Let $\mathbf{P} = (p_{ij})$ be an $n \times n$ stochastic matrix such that $0 < \delta < \Delta$, where $\delta = \min\{p_{ii},\ i \in \mathcal{S}\}$ and $\Delta = \max\{p_{ii},\ i \in \mathcal{S}\}$. Let the set $\mathcal{X}$ be defined by (3.5) and the function $\mathbf{T}$ as in (3.2). Then, with respect to the infinity norm, the function $\mathbf{T}$ is Lipschitz-continuous on $\mathcal{X}$ with Lipschitz constant*

$$K = \frac{1 + (1/\delta - 1)C(\alpha)}{1 + 1/\Delta},$$

where

$$C(\alpha) = -1 + \frac{\alpha + 1}{\alpha - 1} \ln \alpha \quad \text{and} \quad \alpha = \mathrm{e}^{1/\delta - 1/\Delta}.$$

Proof. Using Lemmas A.6 and A.7, we can prove that

$$|T_i(\mathbf{x}) - T_i(\mathbf{y})| \leq K \|\mathbf{x} - \mathbf{y}\|_\infty \quad \text{for all } \mathbf{x}, \mathbf{y} \in \mathcal{X} \text{ and for all } i,$$

https://doi.org/10.1017/jpr.2024.96 Published online by Cambridge University Press

where

$$K = \frac{1 + (1/\delta - 1)C(\alpha)}{1 + 1/\Delta}, \quad C(\alpha) = -1 + \frac{\alpha + 1}{\alpha - 1} \ln \alpha \quad \text{and} \quad \alpha = \mathrm{e}^{1/\delta - 1/\Delta}.$$

For more details we refer the reader to the Appendix. □

If all diagonal elements of $\mathbf{P}$ are identical, there is a closed-form expression for the unique $\mathbb{1}$-generator, which is detailed in the following proposition. Such matrices occur as substitution matrices in Markov models for DNA evolution, among others. See [6] for a discussion of the embeddability of such matrices in the context of the so-called JC69, K80, and K81 models.

Proposition 3.6. *Suppose $\mathbf{P} = (p_{ij})$ is an $n \times n$ stochastic matrix satisfying $p_{ii} = p > 0$ for all i. Then its unique $\mathbb{1}$-generator $\mathbf{Q}$ is given by $\mathbf{Q} = \frac{1}{p}(\mathbf{P} - \mathbf{I})$, where $\mathbf{I}$ is the $n \times n$ identity matrix.*

Proof. Let $\mathbf{Q} = (q_{ij})$ be a $\mathbb{1}$-generator of $\mathbf{P}$. It follows from Proposition 3.4 that $q_{ii} = 1 - 1/p$ for all i. Moreover, if $i \neq j$, Theorem 3.2 and equation (3.1) imply that

$$q_{ij} = \frac{\rho(\mathrm{e}^{1-q_{ii}}, \mathrm{e}^{1-q_{jj}})p_{ij}}{\mathrm{e}^{1-q_{ii}}p_{ii}} = \frac{\rho(\mathrm{e}^{1/p}, \mathrm{e}^{1/p})p_{ij}}{\mathrm{e}^{1/p}p} = \frac{p_{ij}}{p}.$$

In summary, we have

$$q_{ii} = 1 - \frac{1}{p} = \frac{1}{p}(p_{ii} - 1), \quad q_{ij} = \frac{1}{p}p_{ij}, \quad i \neq j,$$

concluding the proof. □

4. Comparison of alternative one-jump approaches and illustrations

Consider a continuous-time Markov chain that is discretely observed at times $t = 0, 1, 2, \ldots, T$ $(T \geq 2)$. Let $\mathbf{P}$ be the empirical transition matrix estimated from these observations. If $\mathbf{P}$ is not embeddable, there is no Markov generator of $\mathbf{P}$, making it a challenge to find an approximate generator. However, if one assumes that the likelihood of more than one transition occurring between two consecutive observations is very low – a scenario referred to as a one-jump setting – it is possible to find a unique rate matrix serving as an approximate generator of $\mathbf{P}$.

Let us adopt the notation $\mathbf{Q}^{\mathrm{JLT}}$, employed in [17], to denote the one-jump rate matrix of Jarrow *et al.* [18]. Furthermore, we denote $\mathbf{Q}^{\mathbb{1}}$ for the $\mathbb{1}$-generator of $\mathbf{P}$. In this section we discuss the specifics of both approaches and compare the resulting rate matrices $\mathbf{Q}^{\mathrm{JLT}}$ and $\mathbf{Q}^{\mathbb{1}}$ as approximate generators of $\mathbf{P}$. To effectively represent the distinction between the two approaches, we use the recursively defined jump times $\tau_0 := 0$ and $\tau_k = \inf\{t \geq \tau_{k-1} : X_t \neq X_{\tau_{k-1}}\}$ $(k = 1, 2, \ldots)$ of the associated CTHMC $(X_t)_{t \geq 0}$.

In their interpretation of the one-jump scenario, Jarrow *et al.* [18] deduce the rate matrix $\mathbf{Q}^{\mathrm{JLT}} = (q_{ij}^{\mathrm{JLT}})$ directly from $\mathbf{P} = (p_{ij})$ through

$$q_{ii}^{\mathrm{JLT}} = \ln p_{ii}, \quad q_{ij}^{\mathrm{JLT}} = \begin{cases} \dfrac{p_{ij} \ln p_{ii}}{p_{ii} - 1} & \text{if } 0 < p_{ii} < 1, \\ 0 & \text{if } p_{ii} = 1, \end{cases} \quad i \neq j, \tag{4.1}$$

assuming $0 < p_{ii}$ for all i. From a probabilistic perspective considering jump times, the elements q_{ij}^{JLT} are defined as the solutions to the equations

$$\mathbb{P}(\tau_1 > 1 \mid X_0 = i) = p_{ii}, \quad \mathbb{P}(\tau_1 \leq 1, X_{\tau_1} = j \mid X_0 = i) = p_{ij}, \quad i \neq j. \tag{4.2}$$

Essentially, their one-jump approach designates p_{ij} to represent the probability of jumping from state i to state j at or before time $t = 1$.

In this paper we derive the $\mathbb{1}$-generator via an alternative approach, as expressed by

$$\mathbb{P}(X_1 = j \mid X_0 = i,\, \tau_2 > 1) = p_{ij} \quad \text{for all } i \text{ and } j, \tag{4.3}$$

which implies that p_{ij} represents the likelihood of going from state i at $t = 0$ to state j at $t = 1$ via at most one jump before or at $t = 1$. Hence the difference between the methods (4.2) and (4.3) depends on how the empirical transition probability matrix $\mathbf{P}$ is viewed in the scenario of a single jump. Both approaches have the merit of bypassing the identification and regularization stages (see [10], [17], and [25]) associated with the embedding problem.

In regularization problems where the transition matrix $\mathbf{P}$ is not embeddable, the goal is to find an approximate Markov generator $\mathbf{Q}$ for $\mathbf{P}$ by minimizing the discrepancy between $\exp \mathbf{Q}$ and $\mathbf{P}$. The magnitude of this difference $\|\mathbf{P} - \exp \mathbf{Q}\|$ depends on the norm used. Davies [10] explored the use of the infinity norm in addressing the regularization problem. For an $n \times n$ matrix $\mathbf{M} = (m_{ij})$, the infinity norm is given by $\|\mathbf{M}\|_\infty = \max_{1 \le i \le n} \sum_{j=1}^{n} |m_{ij}|$. This norm proves useful in comparing transition matrices, since both $\mathbf{P}$ and $\exp \mathbf{Q}$ consist of rows that are stochastic vectors. The Manhattan or Taxicab distance between the corresponding rows, the ith rows in particular, i.e. $\sum_{j=1}^{n} |\mathbf{P}_{ij} - (\exp \mathbf{Q})_{ij}|$, is a simple indicator of the discrepancy between the transition probability distributions from state i. Consequently, the maximum absolute row sum of $\mathbf{P} - \exp \mathbf{Q}$ offers a distinct measure to evaluate the difference between $\mathbf{P}$ and $\exp \mathbf{Q}$. The infinity norm will be used consistently in the remainder of this paper.

An interesting question arises as to whether systematically $\exp \mathbf{Q}^{\mathrm{JLT}}$ or $\exp \mathbf{Q}^{\mathbb{1}}$ is the best approximation to $\mathbf{P}$, or whether this answer depends on the transition matrix $\mathbf{P}$. In the following sections, we analyse the discrepancies $\|\mathbf{P} - \exp \mathbf{Q}^{\mathrm{JLT}}\|_\infty$ and $\|\mathbf{P} - \exp \mathbf{Q}^{\mathbb{1}}\|_\infty$ for specific categories of the transition matrix $\mathbf{P}$.

4.1. Credit rating transition matrix

Consider the empirical transition matrix

$$\mathbf{P} = \begin{bmatrix} 0.8910 & 0.0963 & 0.0078 & 0.0019 & 0.0030 & 0.0000 & 0.0000 & 0.0000 \\ 0.0086 & 0.9010 & 0.0747 & 0.0099 & 0.0029 & 0.0029 & 0.0000 & 0.0000 \\ 0.0009 & 0.0291 & 0.8896 & 0.0649 & 0.0101 & 0.0045 & 0.0000 & 0.0009 \\ 0.0006 & 0.0043 & 0.0656 & 0.8428 & 0.0644 & 0.0160 & 0.0018 & 0.0045 \\ 0.0004 & 0.0022 & 0.0079 & 0.0719 & 0.7765 & 0.1043 & 0.0127 & 0.0241 \\ 0.0000 & 0.0019 & 0.0031 & 0.0066 & 0.0517 & 0.8247 & 0.0435 & 0.0685 \\ 0.0000 & 0.0000 & 0.0116 & 0.0116 & 0.0203 & 0.0754 & 0.6492 & 0.2319 \\ 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 1.0000 \end{bmatrix},$$

which is derived from Table 3 in Jarrow *et al.* [18, p. 506]. We have made minor adjustments to five entries on the main diagonal to ensure that all row sums are equal to one, as some rows in Table 3 did not add up to one because of rounding. For this matrix, the vector function $\mathbf{T} \colon \mathcal{X} \to \mathcal{X}$, defined by (3.2), is a contraction mapping with Lipschitz constant $K < 1$ ($K \approx 0.7833$); see

Proposition 3.5. Using fixed-point iteration and (3.6), we find that the unique $\mathbb{1}$-generator, truncated to four decimal places, is given by

$$\mathbf{Q}^{\mathbb{1}} = \begin{bmatrix} -0.1221 & 0.1075 & 0.0088 & 0.0022 & 0.0036 & 0.0000 & 0.0000 & 0.0000 \\ 0.0096 & -0.1114 & 0.0836 & 0.0114 & 0.0035 & 0.0034 & 0.0000 & 0.0000 \\ 0.0010 & 0.0325 & -0.1271 & 0.0752 & 0.0122 & 0.0053 & 0.0000 & 0.0009 \\ 0.0007 & 0.0049 & 0.0755 & -0.1874 & 0.0798 & 0.0192 & 0.0024 & 0.0049 \\ 0.0005 & 0.0026 & 0.0094 & 0.0886 & -0.2759 & 0.1301 & 0.0178 & 0.0270 \\ 0.0000 & 0.0022 & 0.0036 & 0.0079 & 0.0647 & -0.2121 & 0.0592 & 0.0746 \\ 0.0000 & 0.0000 & 0.0152 & 0.0157 & 0.0287 & 0.1031 & -0.4460 & 0.2834 \\ 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \end{bmatrix}.$$

In contrast, the rate matrix $\mathbf{Q}^{\mathrm{JLT}}$, published in Jarrow *et al.* [18] and defined by equation (4.1), is

$$\mathbf{Q}^{\mathrm{JLT}} = \begin{bmatrix} -0.1154 & 0.1020 & 0.0083 & 0.0020 & 0.0032 & 0.0000 & 0.0000 & 0.0000 \\ 0.0091 & -0.1043 & 0.0787 & 0.0104 & 0.0031 & 0.0031 & 0.0000 & 0.0000 \\ 0.0010 & 0.0308 & -0.1170 & 0.0688 & 0.0107 & 0.0048 & 0.0000 & 0.0010 \\ 0.0007 & 0.0047 & 0.0714 & -0.1710 & 0.0701 & 0.0174 & 0.0020 & 0.0049 \\ 0.0005 & 0.0025 & 0.0089 & 0.0814 & -0.2530 & 0.1180 & 0.0144 & 0.0273 \\ 0.0000 & 0.0021 & 0.0034 & 0.0073 & 0.0568 & -0.1927 & 0.0478 & 0.0753 \\ 0.0000 & 0.0000 & 0.0143 & 0.0143 & 0.0250 & 0.0929 & -0.4320 & 0.2856 \\ 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \end{bmatrix}.$$

We find $\|\mathbf{P} - \exp \mathbf{Q}^{\mathbb{1}}\|_\infty \approx 0.0247$ and $\|\mathbf{P} - \exp \mathbf{Q}^{\mathrm{JLT}}\|_\infty \approx 0.0277$ (accurate to four decimal places). Hence it appears that $\mathbf{Q}^{\mathbb{1}}$ is a better approximate generator of $\mathbf{P}$ than $\mathbf{Q}^{\mathrm{JLT}}$, when using the infinity norm.

Observe that the magnitudes of the diagonal entries in $\mathbf{Q}^{\mathrm{JLT}}$ are smaller compared to those in the $\mathbb{1}$-generator. This observation persists across all matrices $\mathbf{P}$ with non-zero diagonal entries, as demonstrated by the following proposition.

Proposition 4.1. *Suppose $\mathbf{P} = (p_{ij})$ is an $n \times n$ stochastic matrix that satisfies $p_{ii} > 0$ for all i. Then $q_{ii}^{\mathbb{1}} \leq q_{ii}^{\mathrm{JLT}}$ for all i.*

Proof. By Theorem 3.2, $q_{ii}^{\mathbb{1}} = 1 - \ln \theta_i$ for all i, where $\theta = (\theta_1, \ldots, \theta_n)$ is the unique fixed point of the vector function $\mathbf{T} = (T_1, \ldots, T_n)$, defined by (3.2). For $i \in \mathcal{S}$, we have by (3.3) that

$$\theta_i \ln \theta_i = \frac{1}{p_{ii}} \sum_{j \in \mathcal{S}} p_{ij} \rho(\theta_i, \theta_j).$$

Using the definition of the function ρ in (3.1), the above equation can be rewritten as

$$\left(\frac{\theta_i}{\mathrm{e}} - 1\right)\rho(\theta_i, \mathrm{e}) = \theta_i \ln \theta_i - \theta_i = \frac{1}{p_{ii}} \sum_{j \in \mathcal{S}\setminus\{i\}} p_{ij} \rho(\theta_i, \theta_j). \tag{4.4}$$

Now, since $1 - \ln \theta_j = q_{jj}^{\mathbb{1}} \leq 0$, we have $\theta_j \geq \mathrm{e}$ for all j. By Lemma A.2(iii), $\rho(\theta_i, \theta_j) \geq \rho(\theta_i, \mathrm{e})$ for all j. Hence, from (4.4), it follows that

$$\left(\frac{\theta_i}{\mathrm{e}} - 1\right)\rho(\theta_i, \mathrm{e}) \geq \frac{1}{p_{ii}} \sum_{j \in \mathcal{S}\setminus\{i\}} p_{ij} \rho(\theta_i, \mathrm{e}) = \left(\frac{1}{p_{ii}} - 1\right)\rho(\theta_i, \mathrm{e}),$$

https://doi.org/10.1017/jpr.2024.96 Published online by Cambridge University Press

which simplifies to $\theta_i \geq \mathrm{e}/p_{ii}$. Taking logarithms from both sides of this inequality and using $q_{ii}^{\mathbb{1}} = 1 - \ln \theta_i$, we arrive at $q_{ii}^{\mathbb{1}} \leq \ln p_{ii} = q_{ii}^{\mathrm{JLT}}$. □

4.2. Transition matrices with identical diagonal entries

When the stochastic matrix $\mathbf{P}$ has identical diagonal entries $p_{ii} = p > 0$, for all i, Proposition 3.6 indicates that the $\mathbb{1}$-generator can be explicitly represented as

$$\mathbf{Q}^{\mathbb{1}} = \frac{1}{p}(\mathbf{P} - \mathbf{I}),$$

with $\mathbf{I}$ denoting the $n \times n$ identity matrix. Furthermore, by (4.1), we have

$$\mathbf{Q}^{\mathrm{JLT}} = \frac{\ln p}{p-1}(\mathbf{P} - \mathbf{I}).$$

Both $\mathbf{Q}^{\mathbb{1}}$ and $\mathbf{Q}^{\mathrm{JLT}}$ share a similar structure; indeed, they can be expressed as

$$\mathbf{Q}^{\mathbb{1}} = \mathbf{Q}\left(\frac{1}{p}\right) \quad \text{and} \quad \mathbf{Q}^{\mathrm{JLT}} = \mathbf{Q}\left(\frac{\ln p}{p-1}\right),$$

where

$$\mathbf{Q}(k) = k(\mathbf{P} - \mathbf{I}), \qquad k \geq 0. \tag{4.5}$$

In this section, we explore which matrix, either $\mathbf{Q}^{\mathbb{1}}$ or $\mathbf{Q}^{\mathrm{JLT}}$, is the better approximate generator of $\mathbf{P}$ according to the infinity norm.

For the 2×2 case,

$$\mathbf{P} = \begin{bmatrix} p & 1-p \\ 1-p & p \end{bmatrix}, \quad 0 < p < 1,$$

it is known from [24] that $\mathbf{P}$ is embeddable if and only if its trace exceeds 1, i.e. $p > \frac{1}{2}$. The unique Markov generator of $\mathbf{P}$ is then given by

$$\mathbf{Q}\left(\frac{\ln(2p-1)}{2(p-1)}\right);$$

see [28, p. 392]. If $\mathbf{P}$ is not embeddable ($p \leq \frac{1}{2}$), it turns out that $\|\mathbf{P} - \exp \mathbf{Q}^{\mathbb{1}}\|_\infty < \|\mathbf{P} - \exp \mathbf{Q}^{\mathrm{JLT}}\|_\infty$, an inequality that even holds for all $p \in (0, 1)$. This is proved in Lemma A.8 in the Appendix.

For the $n \times n$ case where $n \geq 3$, the answer to our question seems no longer clear-cut. For instance, if

$$\mathbf{P} = \begin{bmatrix} p & 1-p & 0 \\ \frac{1}{2}(1-p) & p & \frac{1}{2}(1-p) \\ 0 & 1-p & p \end{bmatrix}, \quad 0 < p < 1, \tag{4.6}$$

we have $\|\mathbf{P} - \exp \mathbf{Q}^{\mathbb{1}}\|_\infty < \|\mathbf{P} - \exp \mathbf{Q}^{\mathrm{JLT}}\|_\infty$; see Lemma A.9. However, if

$$\mathbf{P} = \begin{bmatrix} p & 0 & 1-p \\ 0 & p & 1-p \\ 0 & 1-p & p \end{bmatrix}, \quad 0 < p < 1, \tag{4.7}$$

https://doi.org/10.1017/jpr.2024.96 Published online by Cambridge University Press

we have the reversed inequality, i.e. $\|\mathbf{P} - \exp \mathbf{Q}^{\mathbb{1}}\|_\infty > \|\mathbf{P} - \exp \mathbf{Q}^{\mathrm{JLT}}\|_\infty$; see Lemma A.10. We note that the transition matrices of (4.6) and (4.7) are not embeddable, since for some $i \neq j$ we have $\mathbf{P}_{ij} = 0$ but $(\mathbf{P}^2)_{ij} > 0$; see [8, Theorem 5, p. 126], highlighting the importance of investigating $\mathbf{Q}^{\mathbb{1}}$ and $\mathbf{Q}^{\mathrm{JLT}}$ as approximate Markov generators.

5. Discussion and future research avenues

From a one-jump perspective, we obtain a modified version of the Markov embedding problem, producing for each $n \times n$ transition matrix $\mathbf{P}$ having non-zero diagonal elements a unique rate matrix $\mathbf{Q}^{\mathbb{1}}$. If $\mathbf{P}$ is embeddable, its Markov generator does not necessarily equal $\mathbf{Q}^{\mathbb{1}}$. However, if the one-jump assumption is acceptable, we might expect the Markov generator to be close to $\mathbf{Q}^{\mathbb{1}}$ in some sense.

The existence and uniqueness of a $\mathbb{1}$-generator is an advantage of our conditional embedding approach because neither regularization nor identification is needed. For some types of transition matrix (as in Proposition 3.6) there exists a closed form for the $\mathbb{1}$-generator; in other situations the $\mathbb{1}$-generator can be determined based on Theorem 3.2.

Comparing our conditional one-jump approach with that of Jarrow *et al.* in [18], we observe reduced discrepancies $\|\mathbf{P} - \exp \mathbf{Q}\|_\infty$ for certain transition matrices $\mathbf{P}$ when using $\mathbf{Q} = \mathbf{Q}^{\mathbb{1}}$ over $\mathbf{Q} = \mathbf{Q}^{\mathrm{JLT}}$. Future research could focus on quantifying the improvements achieved by using $\mathbf{Q}^{\mathbb{1}}$ over $\mathbf{Q}^{\mathrm{JLT}}$. It would also be valuable to determine which transition matrices $\mathbf{P}$ are better approximately generated by $\mathbf{Q}^{\mathbb{1}}$ compared to $\mathbf{Q}^{\mathrm{JLT}}$. Furthermore, exploring how the deviations of $\exp(\mathbf{Q}^{\mathbb{1}})$ and $\exp(\mathbf{Q}^{\mathrm{JLT}})$ from $\mathbf{P}$ vary with different matrix norms beyond the infinity norm used in this study could provide deeper insights. In addition to the comparison based on the discrepancies $\|\mathbf{P} - \exp \mathbf{Q}\|_\infty$, it would also be interesting to consider other criteria, such as the extent to which $\mathbf{Q}^{\mathbb{1}}$ and $\mathbf{Q}^{\mathrm{JLT}}$ are in line with the specificity of the model. Moreover, for an embeddable transition matrix $\mathbf{P}$, a comparison of $\mathbf{Q}^{\mathbb{1}}$ and $\mathbf{Q}^{\mathrm{JLT}}$ with the Markov generators of $\mathbf{P}$ could be of interest.

To study the dynamics of CTHMCs with Markov generators $\mathbf{Q}^{\mathbb{1}}$ and $\mathbf{Q}^{\mathrm{JLT}}$, we can evaluate the mobility index $M(\mathbf{Q}) = -\frac{1}{n}\mathrm{Tr}\mathbf{Q}$ (see [15]), where $\mathbf{Q}$ is an $n \times n$ rate matrix and $\mathrm{Tr}\mathbf{Q}$ its trace. Proposition 4.1 indicates $M(\mathbf{Q}^{\mathbb{1}}) > M(\mathbf{Q}^{\mathrm{JLT}})$, suggesting more dynamic transitions in the CTHMC governed by $\mathbf{Q}^{\mathbb{1}}$. More comparative research on $\mathbf{Q}^{\mathbb{1}}$ and $\mathbf{Q}^{\mathrm{JLT}}$ using other mobility indices would be valuable.

Appendix. Lemmas and proofs

Lemma A.1. *The function f with $f(t) = \mathrm{e}^{W_0(t)}$, $t \geq 0$, is strictly concave.*

Proof. By taking second-order derivatives and since

$$W_0'(t) = \frac{W_0(t)}{t(1 + W_0(t))} \quad \text{and} \quad W_0''(t) = \frac{-2W_0(t)^2 - W_0(t)^3}{t^2(1 + W_0(t))^3}$$

(see e.g. [11]), we find

$$f''(t) = f(t)\big(W_0'(t)^2 + W_0''(t)\big) = f(t)\frac{-W_0(t)^2}{t^2(1 + W_0(t))^3},$$

which is negative for all $t > 0$ since $W_0(t) > 0$ if $t > 0$. □

https://doi.org/10.1017/jpr.2024.96 Published online by Cambridge University Press

Let $\mathbf{o} = (0, \ldots, 0) \in \mathbb{R}^n$. In what follows, we consider the partial ordering of $\mathbb{R}^n$ induced by componentwise ordering. That is, if $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$, and $\mathbf{y} = (y_1, \ldots, y_n) \in \mathbb{R}^n$, we write $\mathbf{x} \preceq \mathbf{y}$ if and only if $x_i \le y_i$ for all i. Likewise, we write $\mathbf{x} \succ \mathbf{o}$ if and only if $x_i > 0$ for all i.

Lemma A.2. *The function $\rho \colon \mathbb{R}^2_+ \to \mathbb{R}_+$, defined as in (3.1), satisfies the following properties:*

(i) *ρ is linearly homogeneous, i.e. $\rho(\lambda \mathbf{x}) = \lambda \rho(\mathbf{x})$ for all $\mathbf{x} \in \mathbb{R}^2_+$ and $\lambda > 0$,*

(ii) *ρ is continuous on $\mathbb{R}^2_+$,*

(iii) *ρ is increasing, i.e. $\rho(\mathbf{x}) \le \rho(\mathbf{y})$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^2_+$ with $\mathbf{x} \preceq \mathbf{y}$,*

(iv) *$\min\{x, y\} \le \rho(x, y) \le \max\{x, y\}$ for all $(x, y) \in \mathbb{R}^2_+$,*

(v) *ρ is concave.*

Proof. Let $\mathbf{u} = (u_1, u_2) \in \mathbb{R}^2_+$. It is easy to see that $\rho(\mathbf{u}) = u_2 f(u_1/u_2) = u_1 f(u_2/u_1)$, where f is the continuous function defined by

$$f(t) = \begin{cases} \dfrac{t \ln t}{t-1} & \text{if } t > 0 \text{ and } t \neq 1, \\ 1 & \text{if } t = 1. \end{cases}$$

Property (i) follows directly from the above. Property (ii) is a direct consequence of the above. For property (iii), by standard calculus, we have

$$f'(t) = \begin{cases} \dfrac{t - 1 - \ln t}{(t-1)^2} & \text{if } t > 0 \text{ and } t \neq 1, \\ 1/2 & \text{if } t = 1, \end{cases}$$

hence f is (strictly) increasing on the positive real half-line since $\ln t < t - 1$ for all $t > 0$ with $t \neq 1$. Now, take $\mathbf{x} = (x_1, x_2)$ and $\mathbf{y} = (y_1, y_2)$, so that $\mathbf{o} \prec \mathbf{x} \preceq \mathbf{y}$. Then, as f is increasing, $\rho(\mathbf{x}) = x_2 f(x_1/x_2) \le x_2 f(y_1/x_2) = y_1 f(x_2/y_1) \le y_1 f(y_2/y_1) = \rho(\mathbf{y})$.

(iv) Consider the case $0 < x \le y$. Since f is increasing on $\mathbb{R}_+$, we have

$$x = x f(1) \le x f(y/x) = \rho(x, y) = y f(x/y) \le y f(1) = y.$$

The result for the case $0 < y \le x$ is proved analogously.

(v) Since $\rho(x, y) = x \rho(1, y/x)$ by the homogeneity property (i) from Lemma A.2, ρ is the perspective of the continuous function $t \mapsto f(t) = \rho(1, t)$. A straightforward computation reveals

$$f''(t) = \frac{g(t)}{(t-1)^3},$$

where $g(t) = 1/t - t + 2 \ln t$. Since

$$g'(t) = -\frac{(t-1)^2}{t^2} < 0 \quad \text{for all } t > 0,$$

we have $g(t) > g(1) = 0$ for all $0 < t < 1$ and $g(t) < g(1) = 0$ for all $t > 1$. Hence $f''(t) < 0$ for all $t > 0$, $t \neq 1$. Therefore f is concave, so ρ is concave as the perspective of a concave function. $\square$

https://doi.org/10.1017/jpr.2024.96 Published online by Cambridge University Press

Lemma A.3. *The vector function* $\mathbf{T}\colon \mathbb{R}_+^n \to \mathbb{R}_+^n$ *from Proposition 3.2 is increasing, i.e.* $\mathbf{T}(\mathbf{x}) \preceq \mathbf{T}(\mathbf{y})$ *for all* $\mathbf{x}, \mathbf{y} \in \mathbb{R}_+^n$ *with* $\mathbf{x} \preceq \mathbf{y}$.

Proof. Let $i \in \{1, \ldots, n\}$ and take $\mathbf{x}, \mathbf{y} \in \mathbb{R}_+^n$ so that $\mathbf{x} \preceq \mathbf{y}$. Denote

$$F_i(\mathbf{u}) = \frac{1}{p_{ii}}\left(\sum_{j=1}^{n} p_{ij}\rho(u_i, u_j)\right) \quad \text{for } \mathbf{u} = (u_1, \ldots, u_n).$$

By Lemma A.2(iii) we have $F_i(\mathbf{x}) \le F_i(\mathbf{y})$, which yields $T_i(\mathbf{x}) = \mathrm{e}^{W_0(F_i(\mathbf{x}))} \le \mathrm{e}^{W_0(F_i(\mathbf{y}))} = T_i(\mathbf{y})$ since the principal branch W_0 of the Lambert W function is increasing (see e.g. [9]). □

Lemma A.4. *Let the vector function* $\mathbf{g}\colon \mathbb{R}_+^n \to \mathbb{R}^n$ *given by* $\mathbf{g}(\mathbf{x}) = \mathbf{T}(\mathbf{x}) - \mathbf{x}$ *for all* $\mathbf{x} \succ \mathbf{o}$, *where the vector function* $\mathbf{T}\colon \mathbb{R}_+^n \to \mathbb{R}_+^n$ *is defined as in (3.2). Then* $\mathbf{g} = (g_1, \ldots, g_n)$ *is*

(i) *quasi-increasing, i.e. for all i it holds that* $\mathbf{o} \prec \mathbf{x} \preceq \mathbf{y}$ *and* $x_i = y_i$ *imply* $g_i(\mathbf{x}) \le g_i(\mathbf{y})$,

(ii) *strictly R-concave, i.e. if* $\mathbf{x} \succ \mathbf{o}$, $\mathbf{g}(\mathbf{x}) = \mathbf{o}$ *and* $0 < \lambda < 1$, *then* $\mathbf{g}(\lambda\mathbf{x}) \succ \mathbf{o}$.

Proof. (i) Take $i \in \{1, \ldots, n\}$ and suppose that $\mathbf{o} \prec \mathbf{x} \preceq \mathbf{y}$ with $x_i = y_i$. Then $g_i(\mathbf{x}) \le g_i(\mathbf{y})$, since $x_i = y_i$ and $T_i(\mathbf{x}) \le T_i(\mathbf{y})$ (Lemma A.3).

(ii) Let $\mathbf{x} = (x_1, \ldots, x_n) \succ \mathbf{o}$ be such that $\mathbf{g}(\mathbf{x}) = \mathbf{o}$. Let $0 < \lambda < 1$ and take $i \in \{1, \ldots, n\}$. Denote

$$F_i(\mathbf{x}) = \frac{1}{p_{ii}}\left(\sum_{j=1}^{n} p_{ij}\rho(x_i, x_j)\right).$$

By Lemma A.2(i), $F_i(\lambda\mathbf{x}) = \lambda F_i(\mathbf{x})$. Hence

$$T_i(\lambda\mathbf{x}) = \mathrm{e}^{W_0(F_i(\lambda\mathbf{x}))} = \mathrm{e}^{W_0(\lambda F_i(\mathbf{x}))} > \lambda \mathrm{e}^{W_0(F_i(\mathbf{x}))} = \lambda T_i(\mathbf{x}),$$

where the inequality follows from the fact that the function $t \mapsto \mathrm{e}^{W_0(t)}$ is strictly concave (Lemma A.1) and $W_0(0) = 0$. Consequently, for all i,

$$g_i(\lambda\mathbf{x}) = T_i(\lambda\mathbf{x}) - \lambda x_i > \lambda(T_i(\mathbf{x}) - x_i) = \lambda g_i(\mathbf{x}) = 0,$$

i.e. $\mathbf{g}(\lambda\mathbf{x}) \succ \mathbf{o}$. □

Lemma A.5. *For all* $0 < p < 1$, *it holds that*

(i) $1 + \mathrm{e}^{2-2/p} - 2p > 0$,

(ii) $1 - \mathrm{e}^{1-1/p} < \frac{4}{3}(1-p)$.

Proof. (i) Let $f(t) = 1 + \mathrm{e}^{2-2/t} - 2t$, which is continuous on the half-open interval (0, 1]. A straightforward calculation reveals $f'(t) = 2t^{-2}(\mathrm{e}^{2-2/t} - t^2)$ and $f''(t) = 4t^{-4}(1-t)\mathrm{e}^{2-2/t}$. So, $f''(t) > 0$ for all $t \in (0, 1)$. Consequently, $f'(t) < 0$ for all $t \in (0, 1)$ because f' increases monotonically on (0,1) and $f'(1) = 0$. Hence f is monotonically decreasing on (0,1). The result now follows from the fact that $f(1) = 0$.

(ii) Consider the function $f(t) = \frac{4}{3}(1-t) - 1 + \mathrm{e}^{1-1/t}$ which is differentiable on $\mathbb{R}_+ \setminus \{0\}$. Let p_0 be a critical point of f, then $f'(p_0) = -\frac{4}{3} + \mathrm{e}^{1-1/p_0}p_0^{-2} = 0$, yielding $\mathrm{e}^{1-1/p_0} = \frac{4}{3}p_0^2$. Clearly $p_0 \ne \frac{1}{2}$, and hence

$$f(p_0) = \frac{4}{3}(1-p_0) - 1 + \frac{4}{3}p_0^2 = \frac{4}{3}\left(\frac{1}{4} - p_0 + p_0^2\right) = \frac{4}{3}\left(\frac{1}{2} - p_0\right)^2 > 0.$$

https://doi.org/10.1017/jpr.2024.96 Published online by Cambridge University Press

So, all critical points of f have positive function values. In addition, we have $f(1)=0$ and $\lim_{t\to 0^+} f(t)=1/3>0$. Therefore $f(t)>0$ for all $t\in(0,1)$. $\square$

Lemma A.6. *For all $x, y\geq 0$, $x\neq y$, it holds that*

$$|\mathrm{e}^{W_0(x)}-\mathrm{e}^{W_0(y)}|<\frac{|x-y|}{1+\min\{W_0(x), W_0(y)\}}.$$

Proof. Assume $x, y\geq 0$ and $x\neq y$. To simplify, consider the function $f(t)=\mathrm{e}^{W_0(t)}$ where $t\geq 0$. Given that W_0 is a monotonically increasing function, it follows that f is also increasing. Hence, according to the mean value theorem, there is a value c in the interval between x and y such that

$$|f(x)-f(y)|=f'(c)\cdot|x-y|.$$

It remains to be proved that

$$f'(c)<\frac{1}{1+W_0(\min\{x, y\})}.$$

Since

$$W_0'(t)=\frac{W_0(t)}{t(1+W_0(t))}$$

(see [9]), we have

$$f'(t)=\mathrm{e}^{W_0(t)}\frac{W_0(t)}{t(1+W_0(t))}=\frac{1}{1+W_0(t)},$$

where the last equality follows from the identity $W_0(t)\mathrm{e}^{W_0(t)}=t$. Hence, because W_0 is an increasing function, we have

$$f'(c)=\frac{1}{1+W_0(c)}<\frac{1}{1+W_0(\min\{x, y\})}=\frac{1}{1+\min\{W_0(x), W_0(y)\}}.$$ $\square$

Lemma A.7. *Let $\mathbf{P}=(p_{ij})$ be an $n\times n$ stochastic matrix such that $p_{ii}>0$ for all i. Let $\delta=\min\{p_{ii},\ i\in\mathcal{S}\}$. For all i and $\mathbf{x}=(x_1,\ldots,x_n)\in\mathbb{R}_+^n$, denote*

$$F_i(\mathbf{x})=\frac{1}{p_{ii}}\sum_{j=1}^{n}p_{ij}\rho(x_i, x_j). \tag{A.1}$$

For $m>0$ and $\alpha>1$, define the set

$$\mathcal{X}_{m,\alpha}=\{(x_1,\ldots,x_n)\in\mathbb{R}^n : m\leq x_i\leq m\alpha\ \forall i\}. \tag{A.2}$$

Then

$$|F_i(\mathbf{x})-F_i(\mathbf{y})|\leq K\|\mathbf{x}-\mathbf{y}\|_\infty \quad \textit{for all } \mathbf{x}, \mathbf{y}\in\mathcal{X}_{m,\alpha} \textit{ and for all } i,$$

where

$$K=1+\left(\frac{1}{\delta}-1\right)C(\alpha) \quad \textit{and} \quad C(\alpha)=-1+\frac{\alpha+1}{\alpha-1}\ln\alpha.$$

https://doi.org/10.1017/jpr.2024.96 Published online by Cambridge University Press

Proof. Take $\mathbf{x}, \mathbf{y} \in \mathcal{X}_{m,\alpha}$ and take $i, j \in \{1, \ldots, n\}$. If $\rho(x_i, x_j) \leq \rho(y_i, y_j)$, then, by concavity of the function $\rho(x, y)$ and the fact that its partial derivatives are non-negative (as stated in Lemma A.2, items (v) and (iii)), it follows that

$$\begin{aligned}|\rho(x_i, x_j) - \rho(y_i, y_j)| &= \rho(y_i, y_j) - \rho(x_i, x_j) \\ &\leq \frac{\partial \rho}{\partial x}(x_i, x_j)(y_i - x_i) + \frac{\partial \rho}{\partial y}(x_i, x_j)(y_j - x_j) \\ &\leq C\|\mathbf{x} - \mathbf{y}\|_\infty,\end{aligned}$$

where

$$C = \max\left\{\frac{\partial \rho}{\partial x}(u, v) + \frac{\partial \rho}{\partial y}(u, v) \colon m \leq u, v \leq m\alpha\right\}. \tag{A.3}$$

Similarly, if $\rho(x_i, x_j) > \rho(y_i, y_j)$, the concave nature of the function ρ along with the non-negativity of its partial derivatives imply

$$\begin{aligned}|\rho(x_i, x_j) - \rho(y_i, y_j)| &= \rho(x_i, x_j) - \rho(y_i, y_j) \\ &\leq \frac{\partial \rho}{\partial x}(y_i, y_j)(x_i - y_i) + \frac{\partial \rho}{\partial y}(y_i, y_j)(x_j - y_j) \\ &\leq C\|\mathbf{x} - \mathbf{y}\|_\infty.\end{aligned}$$

Hence, since $\mathbf{P}$ is a stochastic matrix,

$$\begin{aligned}|F_i(\mathbf{x}) - F_i(\mathbf{y})| &\leq \frac{1}{p_{ii}} \sum_{j=1}^{n} p_{ij}|\rho(x_i, x_j) - \rho(y_i, y_j)| \\ &\leq \frac{1}{p_{ii}}\left(p_{ii}|x_i - y_i| + \sum_{j=1, j\neq i}^{n} p_{ij}|\rho(x_i, x_j) - \rho(y_i, y_j)|\right) \\ &\leq \|\mathbf{x} - \mathbf{y}\|_\infty + \frac{1 - p_{ii}}{p_{ii}} C\|\mathbf{x} - \mathbf{y}\|_\infty \\ &= \left(1 + \left(\frac{1}{p_{ii}} - 1\right)C\right)\|\mathbf{x} - \mathbf{y}\|_\infty \\ &\leq \left(1 + \left(\frac{1}{\delta} - 1\right)C\right)\|\mathbf{x} - \mathbf{y}\|_\infty.\end{aligned}$$

It remains to be shown that

$$C = C(\alpha) = -1 + \frac{\alpha + 1}{\alpha - 1}\ln \alpha.$$

For $u > 0$ and $v > 0$, the homogeneity of the function ρ (see Lemma A.2(i)) and a simple calculation of its partial derivatives entail

$$\frac{\partial \rho}{\partial x}(u, v) + \frac{\partial \rho}{\partial y}(u, v) = -1 + \frac{u + v}{uv}\rho(u, v) = -1 + \left(1 + \frac{v}{u}\right)\rho\left(\frac{u}{v}, 1\right).$$

Denoting $t = u/v$ and $f(t) = (1 + \frac{1}{t})\rho(t, 1)$, it then follows from (A.3) that

$$C = -1 + \max\left\{ f(t) \colon \frac{1}{\alpha} \le t \le \alpha \right\}$$

Using (3.1), we obtain

$$f(t) = \begin{cases} \dfrac{t+1}{t-1} \ln t & \text{if } t \ne 1, \\ 2 & \text{if } t = 1, \end{cases}$$

and by standard calculus, we find that $\max\{f(t) \colon 1/\alpha \le t \le \alpha\} = f(\alpha)$. The proof is complete. □

Proof of Proposition 3.5. Let $i \in \{1, \dots, n\}$ and $\mathbf{x}, \mathbf{y} \in \mathcal{X}$. Using (3.2) and (A.1), we have $T_i(\mathbf{u}) = \exp W_0(F_i(\mathbf{u}))$ for all $\mathbf{u} \in \mathbb{R}^n_+$. Since $\mathbf{T}$ maps $\mathcal{X}$ into itself (see Lemma 3.1(ii)), we have $\min\{T_i(\mathbf{x}), T_i(\mathbf{y})\} \ge \mathrm{e}^{1/\Delta}$, hence $\min\{W_0(F_i(\mathbf{x})), W_0(F_i(\mathbf{y}))\} \ge 1/\Delta$. Therefore, by Lemma A.6,

$$|T_i(\mathbf{x}) - T_i(\mathbf{y})| \le \frac{|F_i(\mathbf{x}) - F_i(\mathbf{y})|}{1 + \min\{W_0(F_i(\mathbf{x})), W_0(F_i(\mathbf{y}))\}} \le \frac{|F_i(\mathbf{x}) - F_i(\mathbf{y})|}{1 + 1/\Delta}.$$

Denote $m = \mathrm{e}^{1/\Delta}$ and $\alpha = \mathrm{e}^{1/\delta - 1/\Delta}$, then by (A.2), $\mathcal{X} = \mathcal{X}_{m,\alpha}$. The proposition now follows from Lemma A.7. □

Lemma A.8. *Let*

$$\mathbf{P} = \begin{bmatrix} p & 1-p \\ 1-p & p \end{bmatrix}, \quad 0 < p < 1,$$

and let $\mathbf{Q}(k)$ *be defined as in* (4.5). *Then*

$$\left\| \mathbf{P} - \exp \mathbf{Q}\left(\frac{1}{p}\right) \right\|_\infty < \left\| \mathbf{P} - \exp \mathbf{Q}\left(\frac{\ln p}{p-1}\right) \right\|_\infty.$$

Proof. It can be shown (e.g. by using Lagrange–Sylvester interpolation to compute functions of matrices; see [14]), that

$$\mathbf{P} - \exp \mathbf{Q}(k) = \left(\frac{1}{2} + \frac{\mathrm{e}^{2k(p-1)}}{2} - p \right) \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix},$$

so that

$$\|\mathbf{P} - \exp \mathbf{Q}(k)\|_\infty = |f(k)|,$$

where $f(k) = 1 + \mathrm{e}^{2k(p-1)} - 2p$. Since $p < 1$, the function f is strictly decreasing, hence

$$f\left(\frac{\ln p}{p-1}\right) > f\left(\frac{1}{p}\right).$$

Furthermore, $f(1/p) > 0$ by Lemma A.5(i). Consequently,

$$\left\| \mathbf{P} - \exp \mathbf{Q}\left(\frac{1}{p}\right) \right\|_\infty = f\left(\frac{1}{p}\right) < f\left(\frac{\ln p}{p-1}\right) = \left\| \mathbf{P} - \exp \mathbf{Q}\left(\frac{\ln p}{p-1}\right) \right\|_\infty. \qquad \square$$

Lemma A.9. *For*

$$\mathbf{P} = \begin{bmatrix} p & 1-p & 0 \\ \frac{1}{2}(1-p) & p & \frac{1}{2}(1-p) \\ 0 & 1-p & p \end{bmatrix}, \quad 0 < p < 1,$$

https://doi.org/10.1017/jpr.2024.96 Published online by Cambridge University Press

and $\mathbf{Q}(k)$ *defined as in* (4.5), *it holds that*

$$\left\|\mathbf{P} - \exp \mathbf{Q}\left(\frac{1}{p}\right)\right\|_\infty < \left\|\mathbf{P} - \exp \mathbf{Q}\left(\frac{\ln p}{p-1}\right)\right\|_\infty.$$

Proof. It can be shown (e.g. by applying Sylvester's theorem for computing functions of a matrix) that

$$\mathbf{P} - \exp \mathbf{Q}(k) = \begin{bmatrix} -\alpha(k) & \beta(k) & \alpha(k)-\beta(k) \\ \frac{1}{2}\beta(k) & -\beta(k) & \frac{1}{2}\beta(k) \\ \alpha(k)-\beta(k) & \beta(k) & -\alpha(k) \end{bmatrix},$$

where

$$\alpha(k) = \frac{1}{4}\gamma(k)^2 + \frac{1}{2}\gamma(k) + \frac{1}{4} - p, \qquad \beta(k) = \frac{1}{2}\gamma(k)^2 + \frac{1}{2} - p, \qquad \gamma(k) = \mathrm{e}^{k(p-1)}. \tag{A.4}$$

A simple calculation reveals

$$\alpha(k) - \beta(k) = -\frac{1}{4}(1-\gamma(k)^2), \tag{A.5}$$

hence $\alpha(k) - \beta(k) \leq 0$. Therefore, by the triangle inequality,

$$\|\mathbf{P} - \exp \mathbf{Q}(k)\|_\infty = |\alpha(k)| + |\beta(k)| + \beta(k) - \alpha(k). \tag{A.6}$$

Since

$$\gamma\left(\frac{\ln p}{p-1}\right) = p,$$

we have

$$\alpha\left(\frac{\ln p}{p-1}\right) > 0 \quad \text{and} \quad \beta\left(\frac{\ln p}{p-1}\right) > 0,$$

so that (A.6) leads to

$$\left\|\mathbf{P} - \exp \mathbf{Q}\left(\frac{\ln p}{p-1}\right)\right\|_\infty = 2\beta\left(\frac{\ln p}{p-1}\right) = (1-p)^2. \tag{A.7}$$

Furthermore, Lemma A.5(i) entails that $\beta(1/p) > 0$. We now need to consider two cases related to $\alpha(1/p)$. If $\alpha(1/p) \geq 0$, we have by (A.6), (A.7) and the fact that β is strictly decreasing (because $p < 1$) that

$$\left\|\mathbf{P} - \exp \mathbf{Q}\left(\frac{1}{p}\right)\right\|_\infty = 2\beta\left(\frac{1}{p}\right) < 2\beta\left(\frac{\ln p}{p-1}\right) = \left\|\mathbf{P} - \exp \mathbf{Q}\left(\frac{\ln p}{p-1}\right)\right\|_\infty.$$

When $\alpha(1/p) < 0$, we have by (A.6) and (A.5)

$$\left\|\mathbf{P} - \exp \mathbf{Q}\left(\frac{1}{p}\right)\right\|_\infty = 2\beta\left(\frac{1}{p}\right) - 2\alpha\left(\frac{1}{p}\right) = \frac{1}{2}\left(1 - \gamma\left(\frac{1}{p}\right)\right)^2.$$

By Lemma A.5(ii) and the assumption $0 < p < 1$, we have $0 < 1 - \gamma(1/p) < \frac{4}{3}(1-p)$. Hence

$$\left\|\mathbf{P} - \exp \mathbf{Q}\left(\frac{1}{p}\right)\right\|_\infty < \frac{8}{9}(1-p)^2 < (1-p)^2 = \left\|\mathbf{P} - \exp \mathbf{Q}\left(\frac{\ln p}{p-1}\right)\right\|_\infty.$$

In either case, we have proved the result. □

https://doi.org/10.1017/jpr.2024.96 Published online by Cambridge University Press

Lemma A.10. *For*

$$\mathbf{P} = \begin{bmatrix} p & 0 & 1-p \\ 0 & p & 1-p \\ 0 & 1-p & p \end{bmatrix}, \quad 0 < p < 1,$$

and $\mathbf{Q}(k)$ *defined as in* (4.5)*, it holds that*

$$\left\| \mathbf{P} - \exp \mathbf{Q}\left(\frac{1}{p}\right) \right\|_\infty > \left\| \mathbf{P} - \exp \mathbf{Q}\left(\frac{\ln p}{p-1}\right) \right\|_\infty .$$

Proof. It can be shown (e.g. by applying Sylvester's theorem for computing functions of a matrix) that

$$\mathbf{P} - \exp \mathbf{Q}(k) = \begin{bmatrix} p - \gamma(k) & \gamma(k) - p - \beta(k) & \beta(k) \\ 0 & -\beta(k) & \beta(k) \\ 0 & \beta(k) & -\beta(k) \end{bmatrix},$$

where $\beta(k) = \frac{1}{2}\,\gamma(k)^2 + \frac{1}{2} - p$ and $\gamma(k) = \mathrm{e}^{k(p-1)}$. Therefore, by the triangle inequality,

$$\|\mathbf{P} - \exp \mathbf{Q}(k)\|_\infty = |p - \gamma(k)| + |p - \gamma(k) + \beta(k)| + |\beta(k)|. \tag{A.8}$$

Let $\delta(k) = (\gamma(k) - 1)^2$. Since

$$\gamma\left(\frac{\ln p}{p-1}\right) = p,$$

we have

$$\beta\left(\frac{\ln p}{p-1}\right) = \frac{1}{2}(p-1)^2 > 0,$$

so that (A.8) leads to

$$\left\| \mathbf{P} - \exp \mathbf{Q}\left(\frac{\ln p}{p-1}\right) \right\|_\infty = 2\beta\left(\frac{\ln p}{p-1}\right) = (p-1)^2 = \delta\left(\frac{\ln p}{p-1}\right). \tag{A.9}$$

Moreover, Lemma A.5(i) shows that $\beta(1/p) > 0$. Also,

$$\gamma\left(\frac{1}{p}\right) < \gamma\left(\frac{\ln p}{p-1}\right) = p,$$

because γ is a strictly decreasing function of k. Thus it follows from (A.8) that

$$\left\| \mathbf{P} - \exp \mathbf{Q}\left(\frac{1}{p}\right) \right\|_\infty = 2\left(p - \gamma\left(\frac{1}{p}\right) + \beta\left(\frac{1}{p}\right)\right) = \delta\left(\frac{1}{p}\right).$$

The proof is now complete because δ is a strictly increasing function of k, leading to

$$\delta\left(\frac{\ln p}{p-1}\right) < \delta\left(\frac{1}{p}\right). \qquad \square$$

Funding information

There are no funding bodies to thank relating to the creation of this article.

https://doi.org/10.1017/jpr.2024.96 Published online by Cambridge University Press

Competing interests

There were no competing interests to declare which arose during the preparation or publication process of this article.

References

[1] ARDIYANSYAH, M., KOSTA, D. AND KUBJAS, K. (2021). The model-specific Markov embedding problem for symmetric group-based models. *J. Math. Biol.* **83**, 1–26.
[2] BAAKE, M. AND SUMNER, J. (2020). Notes on Markov embedding. *Linear Algebra Appl.* **594**, 262–299.
[3] BAAKE, M. AND SUMNER, J. (2022). On equal-input and monotone Markov matrices. *Adv. Appl. Prob.* **54**, 460–492.
[4] BAAKE, M. AND SUMNER, J. (2024). Embedding of Markov matrices for $d \leq 4$. *J. Math. Biol.* **89**, 23:1–45.
[5] CARETTE, P. (1995). Characterizations of embeddable 3×3 stochastic matrices with a negative eigenvalue. *New York J. Math* **1**, 120–129.
[6] CASANELLAS, M., FERNÁNDEZ-SÁNCHEZ, J. AND ROCA-LACOSTENA, J. (2020). Embeddability and rate identifiability of Kimura 2-parameter matrices. *J. Math. Biol.* **80**, 995–1019.
[7] CASANELLAS, M., FERNÁNDEZ-SÁNCHEZ, J. AND ROCA-LACOSTENA, J. (2023). The embedding problem for Markov matrices. *Publicacions Matemàtiques* **67**, 411–445.
[8] CHUNG, K. L. (1967). *Markov Chains with Stationary Transition Probabilities*, 2nd edn. Springer.
[9] CORLESS, R. M., GONNET, G. H., HARE, D. E., JEFFREY, D. J. AND KNUTH, D. E. (1996). On the Lambert W function. *Adv. Comput. Math.* **5**, 329–359.
[10] DAVIES, B. (2010). Embeddable Markov matrices. *Electron. J. Prob.* **15**, 1474–1486.
[11] DENCE, T. P. (2013). A brief look into the Lambert W function. *Appl. Math.* **4**, 887–892.
[12] ELFVING, G. (1937). Zur theorie der Markoffschen Ketten. *Acta Social Sci. Fennicae* **A.8**, 1–17.
[13] FRYDMAN, H. (1980). A structure of the Bang-Bang representation for 3×3 embeddable matrices. *Z. Wahrscheinlichkeitsth.* **53**, 305–316.
[14] GANTMACHER, F. R. (1960). *The Theory of Matrices*, Vol. 1. Chelsea Publishing Company.
[15] GEWEKE, J., MARSHALL, R. C. AND ZARKIN, G. A. (1986). Mobility indices in continuous time Markov chains. *Econometrica* **54**, 1407–1423.
[16] GOODMAN, G. S. (1970). An intrinsic time for non-stationary finite Markov chains. *Z. Wahrscheinlichkeitsth.* **16**, 165–180.
[17] ISRAEL, R. B., ROSENTHAL, J. S. AND WEI, J. Z. (2001). Finding generators for Markov chains via empirical transition matrices, with applications to credit ratings. *Math. Finance* **11**, 245–265.
[18] JARROW, R. A., LANDO, D. AND TURNBULL, S. M. (1997). A Markov model for the term structure of credit risk spreads. *Rev. Financial Studies* **10**, 481–523.
[19] JIA, C. (2016). A solution to the reversible embedding problem for finite Markov chains. *Statist. Prob. Lett.* **116**, 122–130.
[20] JOHANSEN, S. (1973). The bang-bang problem for stochastic matrices. *Z. Wahrscheinlichkeitsth.* **26**, 191–195.
[21] JOHANSEN, S. (1974). Some results on the imbedding problem for finite Markov chains. *J. London Math. Soc.* **2**, 345–351.
[22] JOHANSEN, S. AND RAMSEY, F. L. (1979). A bang-bang representation for 3×3 embeddable stochastic matrices. *Z. Wahrscheinlichkeitsth.* **47**, 107–118.
[23] KENNAN, J. (2001). Uniqueness of positive fixed points for increasing concave functions on R^n: An elementary result. *Rev. Economic Dynamics* **4**, 893–899.
[24] KINGMAN, J. F. C. (1962). The imbedding problem for finite Markov chains. *Z. Wahrscheinlichkeitsth.* **1**, 14–24.
[25] KREININ, A. AND SIDELNIKOVA, M. (2001). Regularization algorithms for transition matrices. *Algo Research Quarterly* **4**, 23–40.
[26] MININ, V. N. AND SUCHARD, M. A. (2008). Counting labeled transitions in continuous-time Markov models of evolution. *J. Math. Biol.* **56**, 391–412.
[27] ROCA-LACOSTENA, J. AND FERNÁNDEZ-SÁNCHEZ, J. (2018). Embeddability of Kimura 3ST Markov matrices. *J. Theoret. Biol.* **445**, 128–135.
[28] SINGER, B. AND SPILERMAN, S. (1973/1974). Social mobility models for heterogeneous populations. *Sociol. Methodol.* **5**, 356–401.