

The Markov chain embedding problem in a low jump frequency context

Philippe Carette

Department of Economics
Ghent University, Sint-Pietersplein 5, B-9000 Ghent, Belgium
`philippe.carette@ugent.be`

Marie-Anne Guerry

Department Business Technology and Operations
Vrije Universiteit Brussel, Pleinlaan 2, B-1050 Brussels, Belgium
`marie-anne.guerry@vub.be`

June 1, 2023

Abstract

We consider the problem of finding the transition rates of a continuous-time homogeneous Markov chain under the empirical condition that the state changes at most once during a time interval of unit length. It is proven that this conditional embedding approach results in a unique intensity matrix for a transition matrix with non-zero diagonal entries. Hence, the presented conditional embedding approach has the merit to avoid the identification phase as well as regularization for the embedding problem. The resulting intensity matrix is compared to the approximation for the Markov generator found by Jarrow in [1].

1 Introduction

The embedding problem of Markov chains is a long standing problem where a given stochastic matrix is examined as the 1-step transition matrix of some continuous-time homogeneous Markov chain (CTHMC) ([2, 3]). This problem boils down to characterizing the empirical transition matrix $\hat{\mathbf{P}}$ as the exponential of some matrix \mathbf{Q} with all non-negative off-diagonal entries and zero row-sums, called an intensity matrix. This matrix \mathbf{Q} represents the transition rates of the underlying CTHMC. If such a \mathbf{Q} exists, $\hat{\mathbf{P}}$ is said to be embeddable. It turns out that the embedding problem is a formidable one in a number of respects. First, $\hat{\mathbf{P}}$ may not be embeddable. In that case, a regularization algorithm can be used to find an intensity matrix \mathbf{Q} for which $\|\hat{\mathbf{P}} - \exp(\mathbf{Q})\|$ is minimized

([4, 5, 6]). Next, no embeddability criteria in terms of the matrix elements, which are easily verifiable in practice, seem at hand when the number of states exceeds 3. Lastly, for an embeddable $\widehat{\mathbf{P}}$, there may not be a unique solution to the equation $\exp(\mathbf{Q}) = \widehat{\mathbf{P}}$ in the set of intensity matrices. The identification aspect of the embedding problem deals with the selection of the suitable intensity matrix reflecting the nature of the system under study ([7]).

More recently, model specific embedding problems are studied for specific subcategories of transition matrices ([8, 9, 10, 1]). In modeling a specific context, the transition matrix as well as the generator matrix are expected to reflect the characteristics of the system under study. Hence, the transition matrix is subjected to constraints and, therefore, belongs to a specific subset of stochastic matrices, and similar, the generator matrix is expected to be an element of a specific subset of intensity matrices. Whereas model specific embedding problems are characterized by setting model assumptions and restrictions on the transition matrix, this paper presents an embedding approach that incorporates empirical assumptions.

More specifically, we propose the conditional embedding approach where the empirical 1-step transition matrix $\widehat{\mathbf{P}}$ corresponds with the conditional 1-step transition matrix of the CTHMC given the event that at most one jump has occurred during a time interval of unit length. For a Markov model the unit time interval can be defined in such a way that the empirical 1-step transition matrix meets this condition. Moreover, this condition is inherent in some applications. For example, in credit rating migration models the credit ratings are typically based on slowly varying characteristics, such that they do not tend to change more than once within the baseline time interval (e.g. a quarter).

We found that, regardless the number of states, exactly one intensity matrix solves this conditional embedding problem when $p_{ii} > 0$ for all i . Our approach results in an easy embeddability criterium and does not require identification neither regularization. Moreover, the presented conditional embedding approach and its proven properties, result in an embeddability roadmap reflecting that the conditional embedding approach is atmost useful in case either the transition matrix is not embeddable or no unique Markov generator can be identified based on the context of the system.

2 Conditional transition probabilities

In order 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 $(\Omega, \mathcal{F}, \mathbb{P})$ with state space $\mathcal{S} = \{1, 2, \dots, n\}$.

Definition 1. *Let $E \in \mathcal{F}$. We call the matrix \mathbf{P}^E with elements*

$$p_{ij}^E = \mathbb{P}(X_1 = j \mid X_0 = i, E), \quad i, j \in \mathcal{S},$$

the conditional one-step transition probability matrix given the event E of the chain $(X_t)_{t \geq 0}$.

Remark. The usual (unconditional) one-step transition probabilities $p_{ij} = \mathbb{P}(X_1 = j \mid X_0 = i)$ can be obtained by setting $E = \Omega$, that is, $p_{ij} = p_{ij}^\Omega$.

In the remainder of this paper, we are interested in the event $E = \{N_J \leq 1\}$, where N_J is the random variable counting the state changes or jumps of the CTHMC up to time $t = 1$.

The relationship between the conditional transition matrix $\mathbf{P}^{\{N_J \leq 1\}}$ and the transition rate matrix \mathbf{Q} of the CTHMC is given by the following proposition.

Proposition 1. *For a CTHMC with transition rate matrix $\mathbf{Q} = (q_{ij})$, it holds that*

$$p_{ij}^{\{N_J \leq 1\}} = \frac{p_{ij}^*}{\sum_{k=1}^n p_{ik}^*} \quad \text{for all } i \text{ and } j,$$

where

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

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

$$\tau(x, y) = \int_0^1 e^{ux + (1-u)y} du = \begin{cases} \frac{e^x - e^y}{x - y} & \text{if } x \neq y \\ e^x & \text{if } x = y \end{cases}. \quad (1)$$

Proof. Using the definition of conditional probability,

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

Hence,

$$p_{ij}^{\{N_J \leq 1\}} = \mathbb{P}(X_1 = j \mid X_0 = i, N_J \leq 1) = \frac{\mathbb{P}(X_1 = j, N_J \leq 1 \mid X_0 = i)}{\mathbb{P}(N_J \leq 1 \mid X_0 = i)}. \quad (2)$$

Let us denote $p_{ij}^* = \mathbb{P}(X_1 = j, N_J \leq 1 \mid X_0 = i)$. Using the sum rule for disjoint events, we then have

$$p_{ij}^{\{N_J \leq 1\}} = \frac{p_{ij}^*}{\sum_{k=1}^n p_{ik}^*}.$$

Let us now calculate p_{ij}^* , which is the joint probability of being in state j at $t = 1$ in at most one jump, starting from state i at $t = 0$. For $k \in \{1, \dots, n\}$, let f_k be the density function of the holding time H_k in state k and F_k the associated cumulative distribution function. For $i \neq j$, denote by s_{ij} the transition probability from state i to

state j conditional on transitioning out of state i . Marginalising on H_i , we find for $i \neq j$

$$\begin{aligned}
p_{ij}^* &= \mathbb{P}(H_i < 1, X_{H_i} = j, H_j > 1 - H_i \mid X_0 = i) \\
&= \int_0^1 \mathbb{P}(X_u = j, H_j > 1 - u \mid H_i = u, X_0 = i) f_i(u) \, du \\
&= \int_0^1 \mathbb{P}(H_j > 1 - u \mid X_u = j, H_i = u, X_0 = i) \mathbb{P}(X_u = j \mid H_i = u, X_0 = i) f_i(u) \, du \\
&= \int_0^1 (1 - F_j(1 - u)) s_{ij} f_i(u) \, du
\end{aligned}$$

and

$$p_{ii}^* = \mathbb{P}(H_i > 1 \mid X_0 = i) = 1 - F_i(1).$$

Since H_k has an exponential distribution with rate parameter $-q_{kk}$ and since $s_{ij} = \frac{q_{ij}}{-q_{ii}}$ ($i \neq j$), we have for $i \neq j$

$$p_{ij}^* = \int_0^1 e^{q_{jj}(1-u)} \frac{q_{ij}}{-q_{ii}} (-q_{ii}) e^{q_{ii}u} \, du = q_{ij} \int_0^1 e^{q_{ii}u + q_{jj}(1-u)} \, du = q_{ij} \tau(q_{ii}, q_{jj})$$

where the function τ is defined as in (1). Finally, $p_{ii}^* = 1 - F_i(1) = e^{q_{ii}} = \tau(q_{ii}, q_{ii})$. \square

One can remark that Minin et al. [11] arrive at the same result for p_{ij}^* using a recursive relation for the joint probabilities $\mathbb{P}(X_1 = j \mid X_0 = i, N_J = n)$, $i, j \in \mathcal{S}$.

Corollary 1. *For all i , we have that $p_{ii}^{\{N_J \leq 1\}} > 0$.*

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

Remark. An alternative argument for corollary 1 goes as follows. Given the event $\{N_J \leq 1\}$, the only way of going from state i at time $t = 0$ to state i at time $t = 1$, is to remain in that state throughout the entire time interval from $t = 0$ to $t = 1$. The probability of this event is non-zero, since the holding time in a state has an exponential distribution.

According to proposition 1, the conditional transition matrix $\mathbf{P}^{\{N_J \leq 1\}}$ depends on the transition rate matrix \mathbf{Q} of the CTHMC involved. In what follows, and when needed, we explicitly indicate this dependency using the notation $\mathbf{P}^{\{N_J \leq 1\}}(\mathbf{Q})$.

3 Conditional embedding problem

When building a discrete time Markov model, the choice of the time unit and time interval is important to end up with a valid model ([12]). In this respect, an appropriate choice can be made by comparing for diverse values of the time unit the internal validity of the corresponding models. The internal validity of a model is determined by the discrepancy between the observed stock vectors and the stock vectors that are estimated by the model. Based on goodness of fit tests a time unit can be selected that results in a

model for which the discrepancy between observed and estimated stock vectors is limited. ([13]). For an appropriate time unit it is acceptable to assume that there is at most 1 jump in between $t = 0$ and $t = 1$. Indeed, more than 1 jump during a one-unit time interval would result in a situation where the transitions to and from the intermediate state are not captured by the discrete time Markov model.

A question that then naturally arises is whether, for a given stochastic matrix \mathbf{P} , there does exist an intensity matrix \mathbf{Q} such that $\mathbf{P}^{\{N_J \leq 1\}}(\mathbf{Q}) = \mathbf{P}$. And if so, whether such an intensity matrix \mathbf{Q} is unique.

It will be helpful to introduce some terminology before proceeding.

Definition 2. A stochastic matrix \mathbf{P} is called J_1 -embeddable iff there exist a CTHMC with transition rate matrix \mathbf{Q} satisfying $\mathbf{P} = \mathbf{P}^E(\mathbf{Q})$, where E is the event that the CTHMC changes state at most once between $t = 0$ and $t = 1$. Such a transition rate matrix is called a J_1 -generator of \mathbf{P} .

For a transition matrix \mathbf{P} that is not embeddable a J_1 -generator can be seen as a solution to the generalization problem where the intensity matrix \mathbf{Q} satisfies $\mathbf{P}^{\{N_J \leq 1\}}(\mathbf{Q}) = \mathbf{P}$. For a transition matrix \mathbf{P} that is embeddable, its Markov generator not necessarily equals the J_1 -generator. In fact, the solution to the conditional embedding problem is generally different from the solution to the (general) embedding problem if the latter exists. However, if the time-scale is chosen such that no more than one transition occurs in the system during the unit time interval, we might expect Markov generator to be close to J_1 -generator in some sense.

A matrix \mathbf{P} that is embeddable satisfies the necessary condition for embeddability formulated by Goodman in [14]: $\prod_{i=1}^n p_{ii} \geq \det P > 0$. Hence, all diagonal entries of such an embeddable matrix \mathbf{P} are non-zero. Consequently, a matrix \mathbf{P} with $p_{ii} = 0$, for some i , is neither embeddable nor J_1 -embeddable, according to corollary 1. For this reason, and without loss of generality, we examine in the remainder of the paper stochastic matrices $\mathbf{P} = (p_{ij})$ satisfying $p_{ii} > 0$ for all i .

It turns out that the off-diagonal elements of a J_1 -generator of \mathbf{P} are uniquely determined by its diagonal elements and the elements of \mathbf{P} . To formulate this relationship, we introduce the function $\rho : \mathbb{R}_+^2 \rightarrow \mathbb{R}_+$, defined as follows:

$$\rho(x, y) = \frac{e}{\tau(1 - \ln x, 1 - \ln y)} = \begin{cases} xy^{\frac{\ln x - \ln y}{x - y}} & \text{if } x \neq y \\ x & \text{if } x = y. \end{cases} \quad (3)$$

Proposition 2. Suppose $\mathbf{P} = (p_{ij})$ is a $n \times n$ stochastic matrix satisfying $p_{ii} > 0$ for all i . If $\mathbf{Q} = (q_{ij})$ is a J_1 -generator of \mathbf{P} , then

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

where $\theta_i = e^{1 - q_{ii}}$ for all i and the function $\rho : \mathbb{R}_+^2 \rightarrow \mathbb{R}_+$ is given by (3).

Proof. Suppose \mathbf{Q} is a J_1 -generator of $\mathbf{P} = (p_{ij})$ and let $i \neq j$. Then, according to proposition 1, we have

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

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

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

The result of proposition 3 yields a condition on the diagonal elements of any J_1 -generator of \mathbf{P} .

Proposition 3. *Suppose $\mathbf{P} = (p_{ij})$ is a $n \times n$ stochastic matrix satisfying $p_{ii} > 0$ for all i . Then, if $\mathbf{Q} = (q_{ij})$ is a J_1 -generator of \mathbf{P} , the n -tuple $(e^{1-q_{11}}, \dots, e^{1-q_{nn}})$ is a fixed point of the vector function $\mathbf{T} = (T_1, \dots, T_n) : \mathbb{R}_+^n \rightarrow \mathbb{R}_+^n$ defined as follows*

$$T_i(x_1, \dots, x_n) = \exp W_0 \left(\frac{1}{p_{ii}} \sum_{j=1}^n p_{ij} \rho(x_i, x_j) \right) \quad \text{for all } i, \quad (4)$$

where W_0 denotes the principal branch of the Lambert W function and where the function $\rho : \mathbb{R}_+^2 \rightarrow \mathbb{R}_+$ is defined as in (3).

Proof. Denote $\theta_i = e^{1-q_{ii}}$ for all i . Then $\theta_i > 0$ and $q_{ii} = 1 - \ln \theta_i$ for all i . Using proposition 2 and the fact that \mathbf{Q} is an intensity matrix, we then have

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

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

$$\theta_i \ln \theta_i = \frac{1}{p_{ii}} \sum_{j=1}^n p_{ij} \rho(\theta_i, \theta_j) \quad \text{for all } i. \quad (5)$$

Using the principal branch W_0 of the Lambert W function (which is the multi-valued inverse of the function $w \mapsto we^w$ ($w \in \mathbb{C}$), see [15]), we find that

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

which proves the result. \square

Proposition 2 and proposition 3 entail that a J_1 -generator of \mathbf{P} defines a fixed point of the vector function \mathbf{T} . The converse is also true, as is stated in the following proposition.

Proposition 4. *Let the stochastic $n \times n$ matrix $\mathbf{P} = (p_{ij})$ be such that $p_{ii} > 0$ for all i . Suppose $\theta = (\theta_1, \dots, \theta_n)$ is a fixed point of the vector function $\mathbf{T} : \mathbb{R}_+^n \rightarrow \mathbb{R}_+^n$, defined in (4). Then, the matrix $\mathbf{Q} = (q_{ij})$ with elements*

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

where ρ is defined by (3), is a J_1 -generator of \mathbf{P} .

Proof. Let $\theta = (\theta_1, \theta_2, \dots, \theta_n) \in \mathbb{R}_+^n$ be a fixed point of \mathbf{T} and let the matrix \mathbf{Q} be constructed as stated above. We first show that \mathbf{Q} is an intensity matrix. By definition, all off-diagonal elements of \mathbf{Q} are non-negative. Since $T_i(\theta) = \theta_i$, we have $\ln \theta_i = W_0(\frac{1}{p_{ii}} \sum_j p_{ij} \rho(\theta_i, \theta_j))$ yielding $\theta_i \ln \theta_i = \frac{1}{p_{ii}} \sum_j p_{ij} \rho(\theta_i, \theta_j)$, by definition of the Lambert W_0 -function. Using $q_{ij} = \frac{\rho(\theta_i, \theta_j) p_{ij}}{\theta_i p_{ii}}$ ($i \neq j$) and $\rho(\theta_i, \theta_i) = \theta_i$, we can rewrite this equation as $\theta_i(1 - q_{ii}) = \theta_i + \frac{1}{p_{ii}} \sum_{j:j \neq i} q_{ij} \theta_i p_{ii}$. After simplification, we get $q_{ii} = -\sum_{j:j \neq i} q_{ij}$. Thus \mathbf{Q} has zero row-sums. Consequently, \mathbf{Q} is an intensity matrix.

It remains to be shown that $p_{ij}^{\{N_J \leq 1\}}(\mathbf{Q}) = p_{ij}$ for all i and j . By proposition 1, we have $p_{ij}^{\{N_J \leq 1\}}(\mathbf{Q}) = \frac{p_{ij}^*}{\sum_k p_{ik}^*}$, where $p_{ik}^* = q_{ik} \tau(q_{ii}, q_{kk})$ if $i \neq k$ and $p_{ii}^* = \tau(q_{ii}, q_{ii})$ and where the function $\tau : \mathbb{R}^2 \rightarrow \mathbb{R}$ is defined by (1). Using the definition of \mathbf{Q} and (3), we then have that

$$p_{ik}^* = \frac{\rho(\theta_i, \theta_k) p_{ik}}{\theta_i p_{ii}} \tau(1 - \ln \theta_i, 1 - \ln \theta_k) = \frac{e p_{ik}}{\theta_i p_{ii}}, \quad i \neq k$$

and

$$p_{ii}^* = \tau(q_{ii}, q_{ii}) = e^{q_{ii}} = e^{1 - \ln \theta_i} = \frac{e}{\theta_i}.$$

Thus, $p_{ik}^* = \frac{e p_{ik}}{\theta_i p_{ii}}$ for all i and k , which yields

$$\sum_k p_{ik}^* = \frac{e}{\theta_i p_{ii}} \sum_k p_{ik} = \frac{e}{\theta_i p_{ii}},$$

since \mathbf{P} has all row sums equal to 1. Consequently, for all i and j , we get

$$p_{ij}^{\{N_J \leq 1\}}(\mathbf{Q}) = \frac{p_{ij}^*}{\sum_k p_{ik}^*} = \frac{e p_{ij}}{\theta_i p_{ii}} \bigg/ \frac{e}{\theta_i p_{ii}} = p_{ij},$$

which concludes the proof. \square

The following lemma states some properties of the vector function \mathbf{T} , which will play a crucial role in its number of fixed points.

Lemma 1. *Let $\mathbf{P} = (p_{ij})$ be a $n \times n$ stochastic matrix. Let $\Delta = \max\{p_{11}, \dots, p_{nn}\}$ and $\delta = \min\{p_{11}, \dots, p_{nn}\}$. Suppose $\delta > 0$. Consider the vector function $\mathbf{T} : \mathbb{R}_+^n \rightarrow \mathbb{R}_+^n$, defined as in (4), and the set*

$$\mathcal{X} = \{\mathbf{x} = (x_1, \dots, x_n) \in \mathbb{R}^n \mid \forall i : e^{1/\Delta} \leq x_i \leq e^{1/\delta}\}. \quad (6)$$

Then,

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

(2) \mathbf{T} maps \mathcal{X} into \mathcal{X} .

Proof.

- (1) Let $\theta = (\theta_1, \dots, \theta_n) \in \mathbb{R}_+^n$ be a fixed point of \mathbf{T} . Let $m = \min\{\theta_1, \dots, \theta_n\}$ and $M = \max\{\theta_1, \dots, \theta_n\}$. We shall prove that $m \geq e^{1/\Delta}$ and that $M \leq e^{1/\delta}$.

Let r be an index such that $\theta_r = m$. Then, by lemma 5(4), we have $\rho(\theta_r, \theta_j) \geq m$ for all j . Since $T_r(\theta) = \theta_r$, we have $\ln \theta_r = W_0(\frac{1}{p_{rr}} \sum_j p_{rj} \rho(\theta_r, \theta_j))$ yielding $\theta_r \ln \theta_r = \frac{1}{p_{rr}} \sum_j p_{rj} \rho(\theta_r, \theta_j)$ by definition of the Lambert W_0 -function. Using the fact that $\sum_{j=1}^n p_{rj} = 1$, we then obtain

$$m \ln m = \theta_r \ln \theta_r = \frac{1}{p_{rr}} \sum_{j=1}^n p_{rj} \rho(\theta_r, \theta_j) \geq \frac{m}{p_{rr}} \geq \frac{m}{\Delta},$$

which implies $\ln m \geq \frac{1}{\Delta}$ and thus $m \geq e^{1/\Delta}$. To prove the second inequality, let s be an index such that $\theta_s = M$. Then, by lemma 5(4), we have $\rho(\theta_s, \theta_j) \leq M$ for all j . Hence, by $T_s(\theta) = \theta_s$ and the unit row-sums property of \mathbf{P} ,

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

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

- (2) Let $(x_1, \dots, x_n) \in \mathcal{X}$. By lemma 5(4),

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

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

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

Now, W_0 and \exp are increasing functions, therefore

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

Finally, using the property $W_0(xe^x) = x$ for $x > 0$, we conclude the proof. \square

Lemma 1 entails that the diagonal elements of \mathbf{P} bound the diagonal elements of the J_1 -generators of \mathbf{P} .

Corollary 2. Let $\mathbf{P} = (p_{ij})$ be a $n \times n$ stochastic matrix. Let $\Delta = \max\{p_{11}, \dots, p_{nn}\}$ and $\delta = \min\{p_{11}, \dots, p_{nn}\}$. Suppose $\delta > 0$. Then, if $\mathbf{Q} = (q_{ij})$ is a J_1 -generator of \mathbf{P} , we have

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

Proof. If $\mathbf{Q} = (q_{ij})$ is a J_1 -generator of \mathbf{P} , we have by proposition 3 that the vector $(\theta_1, \dots, \theta_n)$, where $\theta_i = e^{1-q_{ii}}$ for all i , is a fixed point of \mathbf{T} . Applying lemma 1(1), we have $e^{1/\Delta} \leq \theta_i \leq e^{1/\delta}$ for all i , from which the result follows readily. \square

By combining propositions (2), (3) and (4), it turns out that there is a one-to-one correspondence between the possible J_1 -generators of \mathbf{P} and the fixed points of the vector function \mathbf{T} . Regarding these fixed points, we now prove the following important result.

Theorem 1. *Let the stochastic $n \times n$ matrix $\mathbf{P} = (p_{ij})$ be such that $p_{ii} > 0$ for all i . Then, the vector function $\mathbf{T} : \mathbb{R}_+^n \rightarrow \mathbb{R}_+^n$, defined as in (4), has a unique fixed point.*

Proof. From lemma 1(2), we know that \mathbf{T} maps the compact convex set $\mathcal{X} \subset \mathbb{R}_+^n$, defined by (6), into itself. Also, \mathbf{T} is continuous as the function ρ , defined by (3), is continuous (lemma 5(2)) and continuity is preserved by linear combination and composition of continuous functions. Hence, by the Brouwer fixed-point theorem, \mathbf{T} has a fixed point. By definition of \mathbf{T} , this fixed point must have all positive components. We now show that the function $\mathbf{g} : \mathbb{R}_+^n \rightarrow \mathbb{R}^n$ defined as $\mathbf{g} = \mathbf{T} - \mathbf{Id}$, where $\mathbf{Id} : \mathbb{R}_+^n \rightarrow \mathbb{R}_+^n$ is the identity mapping, satisfies all conditions of Theorem 3.1 in [16]. This theorem states sufficient conditions in order for the function \mathbf{g} to have at most one vector $\mathbf{x} \in \mathbb{R}_+^n$ with $\mathbf{g}(\mathbf{x}) = \mathbf{o}$. These conditions are (a) \mathbf{g} is quasi-increasing and (b) \mathbf{g} is strictly R -concave. Both (a) and (b) are proven in this paper, see lemma 7. So, we have established that \mathbf{T} has exactly one fixed point. \square

We are now in the position to formulate and prove our main theorem.

Theorem 2. *Let the stochastic $n \times n$ matrix $\mathbf{P} = (p_{ij})$ be such that $p_{ii} > 0$ for all i . Then, \mathbf{P} has exactly one J_1 -generator. Moreover, this J_1 -generator $\mathbf{Q} = (q_{ij})$ has elements given by*

$$q_{ii} = 1 - \ln \theta_i, \quad q_{ij} = \frac{\rho(\theta_i, \theta_j) p_{ij}}{\theta_i p_{ii}} \quad (i \neq j) \quad (7)$$

where the scalar function $\rho : \mathbb{R}_+^2 \rightarrow \mathbb{R}_+$ is defined by (3) and $(\theta_1, \dots, \theta_n)$ is the unique fixed point of the vector function $\mathbf{T} : \mathbb{R}_+^n \rightarrow \mathbb{R}_+^n$ defined by (4).

Proof. We first prove that \mathbf{P} has a J_1 -generator. By theorem 1, the vector function \mathbf{T} has a unique fixed point $\theta = (\theta_1, \theta_2, \dots, \theta_n) \in \mathbb{R}_+^n$. Starting from θ , construct the matrix $\mathbf{Q} = (q_{ij})$ according to (7). Then, \mathbf{Q} is a J_1 -generator of \mathbf{P} , by proposition 4.

To prove the uniqueness of the J_1 -generator, suppose that \mathbf{P} has J_1 -generators $\mathbf{R} = (r_{ij})$ and $\mathbf{S} = (s_{ij})$. Then, by proposition 3, the vectors $\theta_{\mathbf{R}} = (e^{1-r_{11}}, \dots, e^{1-r_{nn}})$ and $\theta_{\mathbf{S}} = (e^{1-s_{11}}, \dots, e^{1-s_{nn}})$ are fixed points of the vector function \mathbf{T} . By theorem 1, $\theta_{\mathbf{R}} = \theta_{\mathbf{S}}$. Hence, by proposition 2, we must have $\mathbf{R} = \mathbf{S}$.

Finally, the fact that a J_1 -generator assumes the form 7 is a consequence of proposition 2 and proposition 3. The proof is now complete. \square

According to theorem 2, the unique J_1 -generator is completely determined by the fixed point of the function \mathbf{T} . Besides, under the conditions of lemma 9, the function \mathbf{T} is a contraction. Hence, under these conditions, an algorithm based on the fixed point iteration approach guarantees an appropriate estimation for the fixed point $(\theta_1, \dots, \theta_n)$ as outcome.

For a transition matrix \mathbf{P} with identical positive diagonal elements, a closed-form formula for its unique J_1 -generator can be given by virtue of corollary 2. This type of transition matrices appear in e.g. models of DNA sequence evolution [17].

Corollary 3. *Suppose $\mathbf{P} = (p_{ij})$ is a $n \times n$ stochastic matrix satisfying $p_{ii} = p > 0$ for all i . Then, its unique J_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 J_1 -generator of \mathbf{P} . It follows from corollary 2 that $q_{ii} = 1 - 1/p$ for all i . Moreover, if $i \neq j$, theorem 2 and equation (3) imply that

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

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 Illustrations

The aim of this section is twofold, namely (1) to illustrate the conditional embedding approach for some concrete transition matrices and (2) to compare the new approach with alternative low jump frequency approaches for embedding problems. In [1] Jarow, Lando and Turnbull found an approximation for the Markov generator in closed form under the model assumption that the probability of more than one jump per year is negligible. Their Markov generator $\mathbf{Q}_{\text{JLT}} = (q_{ij}^{\text{JLT}})$ is a product of the model assumption

$$\mathbb{P}(H_i \geq 1 \mid X_0 = i) = p_{ii} \quad , \quad \mathbb{P}(H_i < 1, X_{H_i} = j \mid X_0 = i) = p_{ij} \quad (i \neq j), \quad (8)$$

where H_i is the holding time in state i . This system of equations in the unknowns q_{ij}^{JLT} can be solved explicitly to obtain

$$q_{ii}^{\text{JLT}} = \ln p_{ii} \quad , \quad q_{ij}^{\text{JLT}} = \frac{p_{ij} \ln p_{ii}}{p_{ii} - 1} \quad (i \neq j). \quad (9)$$

In this paper we study the embedding problem under the assumption

$$\mathbb{P}[X_1 = j \mid X_0 = i, N_J \leq 1] = p_{ij} \quad \text{for all } i \text{ and } j. \quad (10)$$

In contrast to assumption (8), assumption (10) is about the data and not about the process. In fact, this paper does not preclude the model from having multiple transitions between $t = 0$ and $t = 1$. It does, however suppose that the data at hand are realisations of the underlying process with no more than one transition between $t = 0$ and $t = 1$. Hence, the estimated one step transition matrix is based solely on observations that did not jump more than once between $t = 0$ and $t = 1$.

Both approaches have the merit to make the identification phase as well as regularization redundant in the embedding problem.

To avoid confusion, let us denote for the transition matrix \mathbf{P} , Jarrow's generator as \mathbf{Q}_{JLT} and the J_1 -generator as \mathbf{Q}_{J_1} . Then, an interesting question emerges of which of the matrices $\exp(\mathbf{Q}_{\text{JLT}})$ and $\exp(\mathbf{Q}_{J_1})$ is the best approximation to \mathbf{P} ? The following section compares both approach for some interesting illustrations. As in [4], throughout this paper the maximum absolute row sum is used as matrix norm, i.e. for an $n \times n$ matrix $\mathbf{M} = (m_{ij})$, $\|\mathbf{M}\|_\infty = \max_{1 \leq i \leq n} \sum_{j=1}^n m_{ij}$.

4.1 Credit rating transition matrix

Consider the empirical transition matrix

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

based¹ on Table 3 in Jarrow et al. [1, p. 506]. For this matrix, it can be proven that the vector function $\mathbf{T} : \mathbb{R}_+^8 \rightarrow \mathbb{R}_+^8$, defined by (4), is a contraction mapping (according to 9). Using fixed-point iteration and (7), we find that the unique J_1 -generator, truncated to 4 decimal places, is:

$$\mathbf{Q}_{J_1} = \begin{bmatrix} -0.1221 & 0.1075 & 0.0088 & 0.0022 & 0.0036 & 0.0000 & 0.0000 & 0.0000 \\ 0.0096 & -0.1114 & 0.0836 & 0.0114 & 0.0035 & 0.0034 & 0.0000 & 0.0000 \\ 0.0010 & 0.0325 & -0.1271 & 0.0752 & 0.0122 & 0.0053 & 0.0000 & 0.0009 \\ 0.0007 & 0.0049 & 0.0755 & -0.1874 & 0.0798 & 0.0192 & 0.0024 & 0.0049 \\ 0.0005 & 0.0026 & 0.0094 & 0.0886 & -0.2759 & 0.1301 & 0.0178 & 0.0270 \\ 0.0000 & 0.0022 & 0.0036 & 0.0079 & 0.0647 & -0.2121 & 0.0592 & 0.0746 \\ 0.0000 & 0.0000 & 0.0152 & 0.0157 & 0.0287 & 0.1031 & -0.4460 & 0.2834 \\ 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \end{bmatrix}$$

In contrast, the rate matrix \mathbf{Q}_{JLT} , published in Jarrow et al. [1] and defined by

¹We have adjusted five entries on the main diagonal to ensure all rows sum up to one.

equation (9), is

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

Notice that the elements on the main diagonal of \mathbf{Q}_{JLT} are in absolute value smaller than their counterparts in the J_1 -generator. This property remains true for all matrices \mathbf{P} having non-zero diagonal elements and follows from lemma 8. It can be shown that $\|\mathbf{P} - \exp \mathbf{Q}_{J_1}\|_\infty < \|\mathbf{P} - \exp \mathbf{Q}_{\text{JLT}}\|_\infty$.

4.2 Transition matrices with same diagonal entries

In case the stochastic matrix \mathbf{P} has coinciding diagonal elements equal to $p > 0$, according to corollary 3 a closed-form solution to the equation $\mathbf{P}^{\{N_J \leq 1\}} = \mathbf{P}$ exists: $\mathbf{Q}_{J_1} = \frac{1}{p}(\mathbf{P} - \mathbf{I})$, where \mathbf{I} is the $n \times n$ identity matrix. Besides, Jarrow's solution (see [1, eqs. 30a & 30b]) is $\mathbf{Q}_{\text{JLT}} = \frac{\ln p}{p-1}(\mathbf{P} - \mathbf{I})$. It is interesting to investigate to what extent $\exp \mathbf{Q}_{J_1}$ and $\exp \mathbf{Q}_{\text{JLT}}$ differ from \mathbf{P} . Note that both matrices \mathbf{Q}_{J_1} and \mathbf{Q}_{JLT} are of the form $\mathbf{Q}(k) := k(\mathbf{P} - \mathbf{I})$ with k constant, and equal to $\frac{1}{p}$ and $\frac{\ln p}{p-1}$ respectively.

For the 2×2 case $\mathbf{P} = \begin{bmatrix} p & 1-p \\ 1-p & p \end{bmatrix}$ it is known that \mathbf{P} is embeddable, with unique generator $\mathbf{Q} = \frac{\ln(2p-1)}{2(p-1)}(\mathbf{P} - \mathbf{I})$, if and only if $p > \frac{1}{2}$ ([7]). For transition matrices \mathbf{P} with $p \leq \frac{1}{2}$ the conditional embedding approach results in a unique J_1 -generator \mathbf{Q}_{J_1} for which $\exp \mathbf{Q}_{J_1}$ is a better approximation to \mathbf{P} than $\exp \mathbf{Q}_{\text{JLT}}$:

Lemma 2. Let $\mathbf{P} = \begin{bmatrix} p & 1-p \\ 1-p & p \end{bmatrix}$, where $0 < p < 1$ and define $\mathbf{Q}(k) = k(\mathbf{P} - \mathbf{I})$, where \mathbf{I} is the 2×2 identity matrix. Then,

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

Proof. It can be shown that

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

so that we have

$$\|\mathbf{P} - \exp \mathbf{Q}(k)\|_\infty = |1 + e^{2k(p-1)} - 2p|.$$

Let $f(k) = 1 + e^{2k(p-1)} - 2p$. Since $p < 1$, the function f is strictly decreasing. Also, since $\ln x < x - 1$ for all $x > 0$ and $x \neq 1$, we have that $\ln \frac{1}{p} < \frac{1}{p} - 1 = \frac{1-p}{p}$ yielding $\frac{\ln p}{p-1} < \frac{1}{p}$. Hence, $f(\frac{\ln p}{p-1}) > f(\frac{1}{p})$. Furthermore, $f(\frac{1}{p}) > 0$ (lemma 10, 1.). Consequently,

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

Hence, for all (2×2) transition matrices with same diagonal entries, it is proven that $\|\mathbf{P} - \exp \mathbf{Q}_{J_1}\|_\infty < \|\mathbf{P} - \exp \mathbf{Q}_{JLT}\|_\infty$. For the (3×3) case, we investigate the transition matrices as introduced in lemma 3. Those transition matrices are not embeddable, since $p_{13} = 0$ but $p_{13}^{(2)} = (1-p)^2/2 > 0$ [18, Theorem 5, p. 126]. Since no generator does exist, it is worth to investigate the J_1 -generator and Jarrow's generator. Lemma 3 proves that $\exp \mathbf{Q}_{J_1}$ better approximates \mathbf{P} than $\exp \mathbf{Q}_{JLT}$, i.e. $\|\mathbf{P} - \exp \mathbf{Q}_{J_1}\|_\infty < \|\mathbf{P} - \exp \mathbf{Q}_{JLT}\|_\infty$.

Lemma 3. Let $\mathbf{P} = \begin{bmatrix} p & 1-p & 0 \\ \frac{1}{2}(1-p) & p & \frac{1}{2}(1-p) \\ 0 & 1-p & p \end{bmatrix}$, where $0 < p < 1$ and define $\mathbf{Q}(k) = k(\mathbf{P} - \mathbf{I})$, where \mathbf{I} is the 3×3 identity matrix. Then,

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

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

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

where

$$\alpha(k) = \frac{1}{4}e^{2k(p-1)} + \frac{1}{2}e^{k(p-1)} + \frac{1}{4} - p \quad \text{and} \quad \beta(k) = \frac{1}{2}e^{2k(p-1)} + \frac{1}{2} - p.$$

It holds that

$$\alpha(k) - \beta(k) = -\frac{1}{4}(1 - e^{k(p-1)})^2 \leq 0, \tag{11}$$

whence,

$$\begin{aligned} \|\mathbf{P} - \exp \mathbf{Q}(k)\|_\infty &= |\alpha(k)| + |\beta(k)| + |\alpha(k) - \beta(k)| \\ &= |\alpha(k)| + |\beta(k)| + \beta(k) - \alpha(k). \end{aligned} \tag{12}$$

Note that α and β are both strictly decreasing in k since $p < 1$. Also, we have

$$\alpha(\frac{\ln p}{p-1}) = \frac{1}{4}(1-p)^2 > 0, \tag{13}$$

$$\beta(\frac{\ln p}{p-1}) = \frac{1}{2}(1-p)^2 > 0, \tag{14}$$

$$\beta(\frac{1}{p}) = \frac{1}{2}(e^{2-2/p} + 1 - 2p) > 0 \quad \text{by lemma 10 (1.).} \tag{15}$$

Hence, by (12), (13) and (14),

$$\|\mathbf{P} - \exp \mathbf{Q}(\frac{\ln p}{p-1})\|_\infty = \alpha(\frac{\ln p}{p-1}) + \beta(\frac{\ln p}{p-1}) + \beta(\frac{\ln p}{p-1}) - \alpha(\frac{\ln p}{p-1}) = 2\beta(\frac{\ln p}{p-1}). \quad (16)$$

In case $\alpha(\frac{1}{p}) \geq 0$, then by (12) and (15),

$$\|\mathbf{P} - \exp \mathbf{Q}(\frac{1}{p})\|_\infty = \alpha(\frac{1}{p}) + \beta(\frac{1}{p}) + \beta(\frac{1}{p}) - \alpha(\frac{1}{p}) = 2\beta(\frac{1}{p}),$$

which yields $\|\mathbf{P} - \exp \mathbf{Q}(\frac{1}{p})\|_\infty < \|\mathbf{P} - \exp \mathbf{Q}(\frac{\ln p}{p-1})\|_\infty$, using (16) and because $\frac{\ln p}{p-1} < \frac{1}{p}$ and β is strictly decreasing.

In case $\alpha(\frac{1}{p}) < 0$, we have by (12) and (11),

$$\|\mathbf{P} - \exp \mathbf{Q}(\frac{1}{p})\|_\infty = -\alpha(\frac{1}{p}) + \beta(\frac{1}{p}) + \beta(\frac{1}{p}) - \alpha(\frac{1}{p}) = \frac{1}{2}(1 - e^{1-1/p})^2.$$

By lemma 10 (2.) and the assumption $0 < p < 1$, it holds that $0 < 1 - e^{1-1/p} < \frac{4}{3}(1-p)$. Hence, using (14) and (16),

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

In either case, we have proven the result. \square

Appendix: Lemma's and proofs

Lemma 4. *The function f with $f(t) = e^{W_0(t)}$, $t \geq 0$, is strictly concave.*

Proof. By taking second order derivatives and since $W_0'(t) = \frac{W_0(t)}{t(1+W_0(t))}$ and $W_0''(t) = \frac{-2W_0(t)^2 - W_0(t)^3}{t^2(1+W_0(t))^3}$ (see e.g. [19]), we find

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

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

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

Lemma 5. *The function $\rho : \mathbb{R}_+^2 \rightarrow \mathbb{R}_+$, defined as in (3), satisfies the following properties:*

- (1) *It is linearly homogeneous, i.e. $\rho(\lambda \mathbf{x}) = \lambda \rho(\mathbf{x})$ for all $\mathbf{x} \in \mathbb{R}_+^2$ and $\lambda > 0$.*
- (2) *It is continuous on \mathbb{R}_+^2 .*
- (3) *It is increasing, i.e. $\rho(\mathbf{x}) \leq \rho(\mathbf{y})$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}_+^2$ with $\mathbf{x} \preceq \mathbf{y}$.*

(4) $\min\{x, y\} \leq \rho(x, y) \leq \max\{x, y\}$ for all $(x, y) \in \mathbb{R}_+^2$.

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

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

(1) Follows directly from the above.

(2) A direct consequence of the above.

(3) By standard calculus, we have

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

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, $\rho(\mathbf{x}) = x_2 f(x_1/x_2) \leq x_2 f(y_1/x_2) = y_1 f(x_2/y_1) \leq y_1 f(y_2/y_1) = \rho(\mathbf{y})$.

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

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

The result for the case $0 < y \leq x$ is proven analogously. \square

Lemma 6. *The vector function $\mathbf{T} : \mathbb{R}_+^n \rightarrow \mathbb{R}_+^n$ from proposition 3 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 $i \in \{1, \dots, n\}$ and take $\mathbf{x}, \mathbf{y} \in \mathbb{R}_+^n$ so that $\mathbf{x} \preceq \mathbf{y}$. Denote, for $\mathbf{u} = (u_1, \dots, u_n)$, $F_i(\mathbf{u}) = \frac{1}{p_{ii}} \left(\sum_{j=1}^n p_{ij} \rho(u_i, u_j) \right)$. By lemma 5(3), we have $F_i(\mathbf{x}) \leq F_i(\mathbf{y})$, which yields $T_i(\mathbf{x}) = e^{W_0(F_i(\mathbf{x}))} \leq e^{W_0(F_i(\mathbf{y}))} = T_i(\mathbf{y})$ since the principal branch W_0 of the Lambert W function is increasing (see e.g. [15]). \square

Lemma 7. *Let the vector function $\mathbf{g} : \mathbb{R}_+^n \rightarrow \mathbb{R}^n$ given by $\mathbf{g}(\mathbf{x}) = \mathbf{T}(\mathbf{x}) - \mathbf{x}$ for all $\mathbf{x} \succ \mathbf{o}$, where the vector function $\mathbf{T} : \mathbb{R}_+^n \rightarrow \mathbb{R}_+^n$ is defined as in (4). Then, $\mathbf{g} = (g_1, \dots, g_n)$ is*

- (1) quasi-increasing, i.e. for all i holds that $\mathbf{o} \prec \mathbf{x} \preceq \mathbf{y}$ and $x_i = y_i$ imply $g_i(\mathbf{x}) \leq g_i(\mathbf{y})$,
- (2) strictly R -concave, i.e. if $\mathbf{x} \succ \mathbf{o}$ and $\mathbf{g}(\mathbf{x}) = \mathbf{o}$ and $0 < \lambda < 1$, then $\mathbf{g}(\lambda \mathbf{x}) \succ \mathbf{o}$.

Proof.

- (1) Take $i \in \{1, \dots, n\}$ and suppose $\mathbf{o} \prec \mathbf{x} \preceq \mathbf{y}$ with $x_i = y_i$. Then, $g_i(\mathbf{x}) \leq g_i(\mathbf{y})$, since $x_i = y_i$ and $T_i(\mathbf{x}) \leq T_i(\mathbf{y})$ (lemma 6).

- (2) Let $\mathbf{x} = (x_1, \dots, x_n) \succ \mathbf{o}$ so that $\mathbf{g}(\mathbf{x}) = \mathbf{o}$. Let $0 < \lambda < 1$ and take $i \in \{1, \dots, n\}$. Denote $F_i(\mathbf{x}) = \frac{1}{p_{ii}} \left(\sum_{j=1}^n p_{ij} \rho(x_i, x_j) \right)$. By lemma 5(1), $F_i(\lambda \mathbf{x}) = \lambda F_i(\mathbf{x})$. Hence,

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

where the inequality follows from the fact that the function $t \mapsto e^{W_0(t)}$ is strictly concave (lemma 4) 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 8. *Let the stochastic $n \times n$ matrix $\mathbf{P} = (p_{ij})$ be such that $p_{ii} > 0$ for all i . Let $\mathbf{Q} = (q_{ij})$ be the J_1 -generator of \mathbf{P} . Then, $q_{ii} \leq \ln p_{ii}$ for all i .*

Proof. By theorem 2, $q_{ii} = 1 - \ln \theta_i$ for all i , where $\theta = (\theta_1, \dots, \theta_n)$ is the unique fixed point of the vector function $\mathbf{T} = (T_1, \dots, T_n)$, defined by (4). Take $i \in \{1, \dots, n\}$. Thus,

$$\theta_i = T_i(\theta) = \exp W_0\left(\frac{1}{p_{ii}} \sum_j p_{ij} \rho(\theta_i, \theta_j)\right),$$

which yields

$$\theta_i \ln \theta_i = \frac{1}{p_{ii}} \sum_j p_{ij} \rho(\theta_i, \theta_j).$$

Using the definition of the function ρ in (3), the above equation can be rewritten as

$$\left(\frac{\theta_i}{e} - 1\right) \rho(\theta_i, e) = \theta_i \ln \theta_i - \theta_i = \frac{1}{p_{ii}} \sum_{j:j \neq i} p_{ij} \rho(\theta_i, \theta_j). \quad (17)$$

Now, since $1 - \ln \theta_j = q_{jj} \leq 0$, we have that $\theta_j \geq e$ for all j . By lemma 5(3), $\rho(\theta_i, \theta_j) \geq \rho(\theta_i, e)$ for all j . Hence, it follows from (17) that

$$\left(\frac{\theta_i}{e} - 1\right) \rho(\theta_i, e) \geq \frac{1}{p_{ii}} \sum_{j:j \neq i} p_{ij} \rho(\theta_i, e) = \frac{1 - p_{ii}}{p_{ii}} \rho(\theta_i, e),$$

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

Lemma 9. *Let $\mathbf{P} = (p_{ij})$ be a $n \times n$ stochastic matrix. Let $\Delta = \max\{p_{11}, \dots, p_{nn}\}$ and $\delta = \min\{p_{11}, \dots, p_{nn}\}$. Suppose $\delta > 0$. Then, with respect to the max norm $\|\cdot\|_\infty$, the function $\mathbf{T} : \mathcal{X} \rightarrow \mathcal{X}$, defined as in (4), is Lipschitz continuous with Lipschitz constant $K = \frac{1 + (\frac{1}{\delta} - 1)C(\alpha)}{1 + \frac{1}{\Delta}}$ where $C(\alpha) = -1 + \frac{\alpha + 1}{\alpha - 1} \ln \alpha$ and $\alpha = e^{\frac{1}{\delta} - \frac{1}{\Delta}}$.*

Lemma 10. *For all $0 < p < 1$, it holds that*

1. $1 + e^{2-2/p} - 2p > 0$,
2. $1 - e^{1-1/p} < \frac{4}{3}(1 - p)$.

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

To prove the second inequality, consider the function $f(t) = \frac{4}{3}(1 - t) - 1 + e^{1-1/t}$ which is differentiable on $\{t \in \mathbb{R} \mid t > 0\}$. Let p_0 be a critical point of f , then $f'(p_0) = -\frac{4}{3} + e^{1-1/p_0}p_0^{-2} = 0$, yielding $e^{1-1/p_0} = \frac{4}{3}p_0^2$. Clearly, $p_0 \neq \frac{1}{2}$, hence,

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

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

References

- [1] Robert A Jarrow, David Lando, and Stuart M Turnbull. A markov model for the term structure of credit risk spreads. *The review of financial studies*, 10(2):481–523, 1997.
- [2] G Elfving. Zur theorie der markoffschen ketten. *Acta Social Sci. Fennicae*, A.8(2):1–17, 1937.
- [3] J.F.C Kingman. The imbedding problem for finite markov chains. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 1(1):14–24, 1962.
- [4] E Davies. Embeddable markov matrices. *Electronic Journal of Probability*, 15:1474–1486, 2010.
- [5] Alexander Kreinin and Marina Sidelnikova. Regularization algorithms for transition matrices. *Algo Research Quarterly*, 4(1/2):23–40, 2001.
- [6] Robert B Israel, Jeffrey S Rosenthal, and Jason Z Wei. Finding generators for markov chains via empirical transition matrices, with applications to credit ratings. *Mathematical finance*, 11(2):245–265, 2001.
- [7] B. Singer and S. Spilerman. Social mobility models for heterogeneous populations. *Sociological Methodology*, 5:356–401, 1973/1974.
- [8] Muhammad Ardiyansyah, Dimitra Kosta, and Kaie Kubjas. The model-specific markov embedding problem for symmetric group-based models. *Journal of Mathematical Biology*, 83(3):1–26, 2021.

- [9] Michael Baake and Jeremy Sumner. Notes on markov embedding. *Linear Algebra and its Applications*, 594:262–299, 2020.
- [10] Jordi Roca-Lacostena and Jesús Fernández-Sánchez. Embeddability of Kimura 3ST Markov matrices. *Journal of theoretical biology*, 445:128–135, 2018.
- [11] Vladimir N Minin and Marc A Suchard. Counting labeled transitions in continuous-time markov models of evolution. *Journal of mathematical biology*, 56(3):391–412, 2008.
- [12] Manudeep Bhuller, Christian N. Brinch, and Sebastian Königs. Time aggregation and state dependence in welfare receipt. *The Economic Journal*, 127(604):1833–1873, 2017.
- [13] Pauline Sales. The validity of the markov chain model for a class of the civil service. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 20(1):85–110, 1971.
- [14] Gerald S. Goodman. An intrinsic time for non-stationary finite markov chains. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 16(3):165–180, 1970.
- [15] Robert M Corless, Gaston H Gonnet, David EG Hare, David J Jeffrey, and Donald E Knuth. On the Lambert W function. *Advances in Computational Mathematics*, 5(1):329–359, 1996.
- [16] John Kennan. Uniqueness of positive fixed points for increasing concave functions on \mathbb{R}_+ : An elementary result. *Review of Economic Dynamics*, 4(4):893–899, 2001.
- [17] Marta Casanellas, Jesús Fernández-Sánchez, and Jordi Roca-Lacostena. Embeddability and rate identifiability of kimura 2-parameter matrices. *Journal of mathematical biology*, 80(4):995–1019, 2020.
- [18] Kai Lai Chung. Markov chains. *Springer-Verlag, New York*, 1967.
- [19] Thomas P Dence. A brief look into the Lambert W function. *Applied Mathematics*, 4(6):887–892, 2013.