ON EQUAL-INPUT AND MONOTONE MARKOV MATRICES

MICHAEL BAAKE,[*] *Bielefeld University*
JEREMY SUMNER,[**] *University of Tasmania*

Abstract

The practically important classes of equal-input and of monotone Markov matrices are revisited, with special focus on embeddability, infinite divisibility, and mutual relations. Several uniqueness results for the classic Markov embedding problem are obtained in the process. To achieve our results, we need to employ various algebraic and geometric tools, including commutativity, permutation invariance, and convexity. Of particular relevance in several demarcation results are Markov matrices that are idempotents.

Keywords: Markov matrix; embedding problem; monotone partial order; convexity

2020 Mathematics Subject Classification: Primary 60J10

1. Introduction

Let $\mathcal{M}_d$ denote the set of d-dimensional *Markov* (or stochastic) matrices, which are the elements of $\mathrm{Mat}(d, \mathbb{R})$ with nonnegative entries and all row sums equal to 1. Clearly, $\mathcal{M}_d$ is a compact convex set, with the d^d Markov matrices with entries in $\{0, 1\}$ being its extremal points (or elements); see [24, Sec. II.1] for a summary. Another classic example is the subset of $\mathcal{M}_d$ of *doubly stochastic* matrices, where both the matrix and its transpose are Markov. Here, the extremal elements are the $d!$ permutation matrices, which is known as Birkhoff's theorem. Concepts and methods from convexity will be needed and employed throughout the manuscript; see [30] for general background on convex structures.

A Markov matrix M is called *embeddable* [2, 5, 7, 20] if it can be written as $M = \mathrm{e}^{Q}$ with a *rate matrix* Q, which is a matrix with nonnegative off-diagonal elements and vanishing row sums. A rate matrix Q is also called a *Markov generator*, or simply *generator*, because $\{\mathrm{e}^{tQ} : t \geqslant 0\}$ is a semigroup of Markov matrices with unit, and is thus a monoid; see [25] for general background on Markov chains in continuous time. The set of embeddable matrices from $\mathcal{M}_d$ is denoted by $\mathcal{M}_d^{\mathrm{E}}$. A Markov matrix is called *infinitely divisible* [8, 11] if it has a Markov nth root for every $n \in \mathbb{N}$, where $\mathbb{N}$ is the set of positive integers. It is a well-known fact [20, Prop. 7] that a Markov matrix is embeddable if and only if it is nonsingular and infinitely divisible; see [14, Sec. 2.3] as well as [6, 12] for related material. In the other direction of taking powers, whenever Q is a generator, $M_{\infty} := \lim_{t \to \infty} \mathrm{e}^{tQ}$ exists, and is an idempotent Markov matrix by [1, Prop. 2.3(3)], so $M_{\infty}^2 = M_{\infty}$.

The embedding problem goes back to Elfving [7] and became prominent through the foundational paper by Kingman [20]. In the two decades following it, many abstract characterisations of embeddable matrices were found and investigated; see [1, 5] and references cited there for some of the literature, as well as [14, 17] for various related questions in matrix analysis. However, beyond $d = 3$, concrete criteria suitable for real-world applications remained elusive, and the interest in the problem diminished somewhat. New impetus then came from theoretical economics and, more recently and quite vigorously, from bioinformatics, where precisely the embedding problem for $d = 4$ is relevant in phylogenetics, with significant progress still being rather fresh; see [1, 2, 11] and references therein.

In this context, the inference problem of discretely-observed, continuous-time Markov chains is the starting point; compare [31], where natural consistency considerations led to several new results (see [9] and references therein). Of particular relevance, both theoretically and practically, are equal-input matrices [27], which possess a powerful algebraic structure and paved the way to some progress, also beyond this class. In fact, it is precisely the systematic use of some standard (and some perhaps not quite so standard) tools from (linear) algebra that unlocks the somewhat stuck embedding problem for progress beyond $d = 2$ as needed in the applications. It is one goal of this paper to explain some of these techniques in action, by applying them to two particularly important classes of Markov chains. This way, we also attempt to convince the reader that stepping a little into algebraic methods can be more than a little profitable.

To avoid trivial statements, we will generally assume $d \geqslant 2$. Let $C \in \mathrm{Mat}(d, \mathbb{R})$ have equal rows, each being $(c_1, \ldots, c_d)$, and define $c = c_1 + \ldots + c_d$ as its *parameter sum*. Such a matrix C is Markov precisely when we have $c = 1$ together with $c_i \geqslant 0$ for all i. However, it then has rank 1, so $\det(C) = 0$ for $d > 1$. For this reason, such matrices C are often of limited interest, for instance in the context of embeddability. Instead, consider $M_C = (1 - c)\mathbb{1} + C$, which is a matrix with each row summing to 1. Here and below, $\mathbb{1}$ denotes the identity matrix. Cleary, M_C is Markov if and only if $c_i \geqslant 0$ and $c \leqslant 1 + c_i$ for all i. Following the much-cited monograph [27], we call such Markov matrices *equal-input*, since they describe Markov chains where the probability of a transition $i \to j$, for $i \neq j$, depends on j only. As detailed in [27] (see also [1] and references therein), they constitute an important model class in bioinformatics and phylogenetics. All such matrices are of the above form, and the underlying c is called its *summatory parameter*. For a given d, they form another convex set, which we denote by $\mathcal{C}_d$; see Lemma 2.8 for more.

A seemingly unrelated concept, at least at first sight, is the following. Consider the standard simplex of d-dimensional probability vectors,

$$\mathbb{P}_d := \big\{ (x_1, \ldots, x_d) : \text{all } x_i \geqslant 0 \text{ and } x_1 + \ldots + x_d = 1 \big\},$$

which is a compact convex set. Its extremal elements are the standard (row) basis vectors e_i with $i \in [d] := \{1, \ldots, d\}$. One can introduce the partial order of *stochastic monotonicity* on $\mathbb{P}_d$ by saying that x is *dominated* by y, written as $x \preccurlyeq y$, when $\sum_{i=m}^{d} x_i \leqslant \sum_{i=m}^{d} y_i$ holds for all $m \in [d]$; see Equation (3.2) below for an alternative formulation. The corresponding partial order is well defined also on the positive multiples of $\mathbb{P}_d$, called *level sets*, hence extendable to convex combinations. Further, it is consistent on the entire positive cone, where two vectors can at most be compared when they lie in the same level set. This notion has its origin in an important class of stochastic processes [4] that show up in many places in probability theory and its applications [18, 19, 22]. Though practically perhaps most relevant in various

economic contexts, stochastic monotonicity induces another class of matrices with a lot of internal structure, which is relevant in the embedding context, as we shall explain below.

A Markov matrix $M = (m_{ij})_{1 \leqslant i, j \leqslant d}$ is called *stochastically monotone*, or *monotone* for short, when the mapping $x \mapsto xM$ preserves stochastic monotonicity; compare [4, 18]. It is well-known that M is monotone if and only if its rows $m_i = (m_{i1}, \ldots, m_{id}) = e_i M$ are ordered accordingly with increasing row numbers, meaning $m_i \preccurlyeq m_j$ for all $i \leqslant j$. More generally, the concept of being monotone is well defined for all nonnegative matrices with equal row sums, which are simply multiples of Markov matrices. Then, preserving the partial order means that an inequality in one level set is turned into the corresponding one in another. The monotone Markov matrices in $\mathcal{M}_d$ form a closed convex set, which we denote by $\mathcal{M}_{d, \preccurlyeq}$. All elements of $\mathcal{M}_{d, \preccurlyeq}$ have trace $\geqslant 1$, and the extremal points of $\mathcal{M}_{d, \preccurlyeq}$, as detailed in Lemma 3.4 below, are the $\binom{2d-1}{d}$ monotone Markov matrices with entries in $\{0, 1\}$. Monotone Markov matrices appear in many contexts; see [18, 22] as well as [19, Ch. 3] for examples.

A *stationary vector* of $M \in \mathcal{M}_d$ is any $x \in \mathbb{P}_d$ with $xM = x$. Given M, the set of all stationary vectors is convex. In fact, it is a subsimplex of $\mathbb{P}_d$ that can be a singleton set (as for all irreducible M) or larger, up to $\mathbb{P}_d$ itself (for $M = \mathbb{1}$). A Markov matrix M is an *idempotent* when $M^2 = M$. This means that M maps any $x \in \mathbb{P}_d$ to a stationary vector in one step. Except for $M = \mathbb{1}$, any idempotent in $\mathcal{M}_d$ has minimal polynomial $x(x - 1)$. In particular, all idempotents in $\mathcal{M}_d$ are diagonalisable, and all but $M = \mathbb{1}$ are singular. Markov idempotents constitute an interesting subset of the Markov matrices in their own right (compare [16, Sec. 1.6]), but also provide a natural link between equal-input and monotone matrices. Indeed, the above two matrix classes are intimately connected, and several of their properties can be derived and understood interactively. In our presentation below, idempotents will play an important role, which is an interesting point that does not seem to have attracted enough attention in the past.

The structure of the paper is as follows. Since we build on terminology, methods, and results from [1], we feel that a renewed section on preliminaries is unnecessary. While we try to make the paper as self-contained as possible, some reference to [1] is inevitable to avoid duplication. Instead, we proceed by recalling and extending some important properties of equal-input matrices in Section 2 and then deriving new results on their graded semigroup structure (Proposition 2.6), which provides relevant insight into the properties of such Markov matrices. We then determine their embeddability (Proposition 2.12) and their multiplicative structure in exponential form (Theorem 2.15).

The latter can be viewed as a nontrivial, explicit version of the Baker–Campbell–Hausdorff (BCH) formula in this case; see the Wikipedia entry on the BCH formula for a good summary of this tool, which is particularly useful in the context of inhomogeneous Markov chains. Beyond the commuting case, where the BCH formula is trivial, we are not aware of many other matrix classes with such a favourable structure, and a better understanding of multiplicatively closed matrix classes is needed; compare [28, 29] as well as [27, Ch. 7].

In Section 3, we turn to the monotone matrices, first recalling some of their elementary properties and then continuing with results on embeddability (Theorem 3.8) and monotone generators (Proposition 3.11). Throughout the discussion, idempotent matrices will naturally appear, which can be explained by the intrinsic (pseudo-)Poissonian structure of infinitely divisible Markov matrices (Proposition 3.14 and Theorem 3.17). We consider the case $d = 3$ in more detail in Section 4, where the embeddability can be decided completely (Proposition 4.3

and Theorem 4.4), and close with a general uniqueness result (Theorem 5.3 and Corollary 5.5) and some comments on Markov roots and limiting cases in Section 5.

2. Equal-input matrices and some of their properties

From now on, we always use C to denote a nonnegative matrix with equal rows and the parameter sum $c \geqslant 0$, where $c = 0$ then implies $C = \mathbb{0}$. For any matrix C of this type, one has $MC = C$ for all $M \in \mathcal{M}_d$.

2.1. Equal-input Markov matrices

Given such a matrix C, the corresponding matrix

$$M_C := (1 - c)\mathbb{1} + C \tag{2.1}$$

is Markov precisely when $0 \leqslant c_i$ for all i and $c \leqslant 1 + \min_i c_i$. It is then called an *equal-input* Markov matrix, and the set of all such matrices for a fixed d is denoted by $\mathcal{C}_d$. As $C \neq \mathbb{0}$ has eigenvalues c and 0, the latter with multiplicity $d - 1$, it is clear that

$$\det(M_C) = (1 - c)^{d-1}. \tag{2.2}$$

A Markov generator Q is called *equal-input* if it is of the form $Q = Q_C = C - c\mathbb{1}$, with C as above and $c \geqslant 0$ without further restrictions. Clearly, $c = 0$ means $Q = \mathbb{0}$. Also here, we call c the *summatory parameter*, as it will always be clear from the context whether we are referring to a Markov matrix or to a generator. Since $C^2 = cC$, one gets $Q_C^2 = -cQ_C$, so any equal-input generator is diagonalisable. When $c > 0$, the matrix $\frac{1}{c}C$ is both Markov and an idempotent. As we shall see, the relation $Q_C = M_C - \mathbb{1}$ will become particularly important.

Fact 2.1. *If $M \in \mathcal{C}_d$, with $d \geqslant 2$, its summatory parameter satisfies $c \in \left[0, \frac{d}{d-1}\right]$. Moreover, the matrix $M \in \mathcal{C}_d$ is an idempotent if and only if $c = 1$ or $c = 0$, where the latter case means $M = \mathbb{1}$, while all other idempotents are singular.*

Proof. Assume that $M = M_C$ is Markov. Since $0 \leqslant c = c_1 + \ldots + c_d \leqslant 1 + \min_i c_i$, the maximal value of c is realised for $c_1 = \ldots = c_d = \frac{c}{d}$, which gives the first claim. The second follows from considering the equation

$$(1 - c)\mathbb{1} + C = M_C = M_C^2 = (1 - c)^2\mathbb{1} + (2 - c)C,$$

which is solved by $C = \mathbb{0}$ with $c = 0$ or, for any $c > 0$, implies $(1 - c)C = \mathbb{0}$ and hence $c = 1$. This leads to the two cases stated, where $\mathbb{1}$ is the only nonsingular idempotent by (2.2). $\qquad\square$

Let us next look at some asymptotic properties. Here, one has

$$M_C^n = (1 - c)^n\mathbb{1} + \frac{1 - (1 - c)^n}{c} C$$

for $n \in \mathbb{N}_0$ and $c > 0$ (so $C \neq \mathbb{0}$). If $|1 - c| < 1$, we see that M_C^n, as $n \to \infty$, converges to the Markov matrix $\frac{1}{c}C$. Adding the case with $c = 0$, but excluding from consideration $c = 2$ (where convergence fails, occurring only for $d = 2$), one can summarise as follows.

Fact 2.2. *Let $d \geqslant 2$ and let C be a nonnegative matrix with equal rows and parameter sum $0 \leqslant c < 2$. Then, if the matrix M_C from* (2.1) *is Markov, one has*

$$\lim_{n \to \infty} M_C^n = \begin{cases} \mathbb{1}, & \text{if } c = 0, \\ \dfrac{1}{c} C, & \text{otherwise}, \end{cases}$$

where all limits are idempotents. Here, the summatory parameter of M_C^n is $1 - (1 - c)^n$, which is 0 for $c = 0$, or otherwise converges to 1 as $n \to \infty$. $\quad\square$

Since idempotents will show up repeatedly below, we recall the following well-known property of Markov matrices, which we also prove for the reader's convenience.

Lemma 2.3. *For $M \in \mathcal{M}_d$, the following properties are equivalent:*

1. *M is a nonsingular idempotent.*

2. *1 is the only eigenvalue of M.*

3. *$M = \mathbb{1}$.*

4. *M has minimal polynomial $q(x) = x - 1$.*

Proof. Clearly, M has 1 as an eigenvalue, because it is Markov. When $M^2 = M$, the only possible other eigenvalue is 0, and (1) $\Rightarrow$ (2) is clear. The implications (3) $\Rightarrow$ (4) $\Rightarrow$ (1) are immediate, and it remains to show (2) $\Rightarrow$ (3).

By [10, Thm. 13.10], we know that the algebraic multiplicity of the eigenvalue 1 agrees with the geometric one. When no other eigenvalue exists, this means $M = \mathbb{1}$. $\quad\square$

The set $\mathcal{C}_d$ of equal-input matrices is important in many applications (see [1, 27] and references therein) and has interesting and revealing algebraic properties as follows.

Fact 2.4. *Let C and C' be two nonnegative, equal-row matrices, with parameter sums c and c', such that M_C and $M_{C'}$ are Markov matrices, so both lie in $\mathcal{C}_d$. Then, one also obtains that $M = M_C M_{C'} \in \mathcal{C}_d$, where one has $M = M_{C''}$ with $C'' = (1 - c')C + C'$ and parameter sum $c'' = c + c' - cc'$.* $\quad\square$

The Markov property of $M = M_C M_{C'}$ implies $0 \leqslant c_i'' \leqslant c'' \leqslant 1 + c_i''$ for all i, which may not be obvious from the formula for C''. In fact, the relation between the summatory parameters can be further analysed as follows, where we refer to [21] for the grading notion. We will need the group C_2 with two elements, here written as $\{1, -1\}$ with ordinary multiplication.

Lemma 2.5. *Consider $f(a, b) = a + b - ab$ for $a, b \in X := [0, 1) \cup (1, 2]$. Then, one of the following three cases applies:*

1. *If $\max(a, b) < 1$, one has $0 \leqslant \max(a, b) \leqslant f(a, b) < 1$.*

2. *If $\min(a, b) > 1$, one has $0 \leqslant 2 - \min(a, b) \leqslant f(a, b) < 1$.*

3. *Otherwise, one has $1 < f(a, b) \leqslant \max(a, b) \leqslant 2$.*

In particular, the mapping $(a, b) \mapsto f(a, b)$ turns X into a C_2-graded, commutative monoid, with 0 as the neutral element of X, and with the grading being induced by the two connected components of X, for instance via the mapping $x \mapsto \text{sgn}(1 - x)$ for $x \in X$

Proof. Without loss of generality, we may assume $a \leqslant b$. Also, observe that the function $x \mapsto x(2 - x)$, on $[0, 2]$, has a unique maximum at $x = 1$, with value 1, so $x(x - 1) < 1$ holds for all $x \in [0, 1) \cup (1, 2]$. Now, we can look at the three cases as follows.

When $0 \leqslant a \leqslant b < 1$, where $1 - b$ is positive, we obtain the estimate

$$0 \leqslant b \leqslant b + (1 - b)a = f(a, b) < b + (1 - b) = 1,$$

while $1 < a \leqslant b \leqslant 2$, where $1 - a$ is negative, leads to

$$0 \leqslant 2 - a = a + 2(1 - a) \leqslant a + b(1 - a) = f(a, b) \leqslant a + a(1 - a) = a(2 - a) < 1.$$

For the remaining case, it suffices to consider $0 \leqslant a < 1 < b \leqslant 2$, which gives

$$1 = b + (1 - b) < b + a(1 - b) = f(a, b) \leqslant a + b - a = b \leqslant 2,$$

from which the claims (1)–(3) follow.

Since $f(f(a, b), c) = a + b + c - ab - ac - bc + abc = f(a, f(b, c))$, associativity of the mapping $(a, b) \mapsto f(a, b)$ is clear, and the C_2-graded monoid structure is now obvious. $\square$

Proposition 2.6. *The set $\mathcal{C}_d$ is a monoid under matrix multiplication, with the subset of non-singular elements forming a submonoid. The latter is C_2-graded by $\mathrm{sgn}(1 - c)$, where c is the summatory parameter, which matches with the grading of X from Lemma 2.5. When d is even, the same grading emerges from the sign of the determinant.*

Proof. The semigroup property follows from Fact 2.4, and $\mathbb{1} \in \mathcal{C}_d$ shows that $\mathcal{C}_d$ is a monoid. The nonsingular matrices, which include $\mathbb{1}$, are closed under multiplication.

The formula for the summatory parameter of a product from Fact 2.4, in conjunction with Lemma 2.5, implies $c'' \in [0, 1)$ when c and c' are either both in $[0, 1)$ or both in $\left(1, \frac{d}{d-1}\right]$, where the ranges follow from Fact 2.1. Likewise, $c'' > 1$ if and only if $c < 1 < c'$ or $c' < 1 < c$. Together, these observations provide the claimed C_2-grading.

For even d, by Equation (2.2), the sign of $1 - c$ matches the sign of the determinant. $\square$

Remark 2.7. Given $d \geqslant 2$, one can consider $X_d := \left\{x \in X : x \leqslant \frac{d}{d-1}\right\}$, which defines a submonoid of X which is again C_2-graded. We then have two successive monoid homomorphisms, namely

$$\mathcal{C}_d \longrightarrow X_d \longrightarrow C_2,$$

which summarises the grading structure. Let us mention in passing that the grading can be extended to include the singular matrices (that is, those with $c = 1$), and thus cover all of $\mathcal{C}_d$, by employing the semigroup $\{-1, 0, 1\}$ instead of C_2.

Next, we consider the set $\mathcal{C}_d$ in a little more detail. We begin with the closed subset (and semigroup) of nontrivial idempotents,

$$\mathcal{C}_{d,1} := \left\{M \in \mathcal{C}_d : c = 1\right\} = \left\{M \in \mathcal{C}_d : M^2 = M \neq \mathbb{1}\right\} = \left\{M \in \mathcal{C}_d : \det(M) = 0\right\},$$

where the alternative characterisations immediately follow from Fact 2.1 and Equation (2.2). Now, let $E_i \in \mathrm{Mat}(d, \mathbb{R})$ denote the matrix with 1 in all positions of column i and 0 everywhere else, which is an idempotent Markov matrix. In fact, these matrices satisfy

$$E_i E_j = E_j \quad \text{for all } 1 \leqslant i, j \leqslant d, \tag{2.3}$$

and it is easy to see that any convex combination $M = \sum_{i=1}^{d} \beta_i E_i$, where all $\beta_i \geqslant 0$ subject to $\beta_1 + \ldots + \beta_d = 1$, is an idempotent as well. For $d \geqslant 2$, we also define the rational matrix

$$G_d = \frac{1}{d-1} \left(\begin{pmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{pmatrix} - \mathbb{1} \right) = \frac{1}{d-1} (E_1 + \ldots + E_d - \mathbb{1}), \qquad (2.4)$$

which is the unique element of $\mathcal{C}_d$ with maximal summatory parameter, $c = \frac{d}{d-1}$.

Lemma 2.8. *The sets $\mathcal{C}_d$ and $\mathcal{C}_{d,1}$ are convex. When $d \geqslant 2$, the extremal elements of $\mathcal{C}_{d,1}$ are the idempotent matrices $E_1, \ldots, E_d$, which remain extremal in $\mathcal{C}_d$. There are two further extremal elements in $\mathcal{C}_d$, namely $\mathbb{1}$ and G_d.*

Proof. The convexity of $\mathcal{C}_d$ follows from

$$\alpha M_C + (1-\alpha) M_{C'} = \big(1 - (\alpha c + (1-\alpha)c')\big)\mathbb{1} + \alpha C + (1-\alpha)C'$$

for $\alpha \in [0, 1]$ together with the linearity of the summatory parameter. The convexity of the subset $\mathcal{C}_{d,1}$ is then obvious because these are the elements with $c = 1$. Clearly, the convex combinations of the d matrices $E_1, \ldots, E_d$ span $\mathcal{C}_{d,1}$. Since they are linearly independent, they must be extremal.

Let $M \in \mathcal{C}_d$, so $M = M_C$, where C has equal rows $(c_1, \ldots, c_d)$ with $c_i \geqslant 0$ and parameter sum $c \in \left[0, \frac{d}{d-1}\right]$, subject to the condition $c \leqslant 1 + c_{\min}$ with $c_{\min} = \min_i c_i$. We will now show that M is of the form $r\mathbb{1} + sG_d + \sum_i t_i E_i$ for some $r, s, t_i \geqslant 0$ that sum to 1.

If $c \in [0, 1]$, simply choose $s = 0$, $t_i = c_i$, and $r = 1 - c$, which does the job. If $c > 1$, we have $c_{\min} > 0$ in our setting. Choose $s = (d-1)c_{\min}$ and $t_i = c_i - c_{\min}$, where $s \geqslant 0$ and all $t_i \geqslant 0$ by construction. Then $s + \sum_i t_i = c - c_{\min} \leqslant 1$, so we can complete this with the choice $r = 1 - s - \sum_i t_i \geqslant 0$. It is easy to check that this gives a convex combination with summatory parameter c, and also that the sum equals M.

Consequently, the compact set $\mathcal{C}_d$ is the convex hull of $\{\mathbb{1}, G_d, E_1, \ldots, E_d\}$, and hence is the smallest convex set that contains these $d + 2$ matrices. In view of the classic Krein–Milman theorem [30, Thm. 2.6.16], it remains to show that they all are extremal. This is clear for $\mathbb{1}$, where $c = 0$, and for G_d, which is the only matrix in $\mathcal{C}_d$ with $c = \frac{d}{d-1}$. Since the E_i are linearly independent, but all have $c = 1$, none can be replaced by a convex combination of the other matrices (including $\mathbb{1}$ and G_d), which completes the argument. $\square$

Example 2.9. For $d = 2$, all Markov matrices are of equal-input type. The four extremal elements of $\mathcal{C}_2 = \mathcal{M}_2$ are given by

c	0	1	1	2
C	$\mathbb{0}$	$\begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}$	$\begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$
M_C	$\mathbb{1}$	$\begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}$	$\begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$

where the two Markov matrices with $c = 1$ span $\mathcal{C}_{2,1}$; see [1, Fig. 1] for an illustration.

The situation is a little more interesting for $d = 3$, where $\mathcal{C}_3 \subsetneq \mathcal{M}_3$, and one has

c	0	1	1	1	$\frac{3}{2}$
M_C	$\mathbb{1}$	$\begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \end{pmatrix}$	$\begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}$	$\frac{1}{2}\begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$

part of which will reappear in Table 1 below.

2.2. Equal-input generators and embeddability

If Q is an equal-input generator with summatory parameter c, so $Q = C - c\mathbb{1}$, its exponential is

$$e^Q = \mathbb{1} + \frac{1 - e^{-c}}{c}Q = \frac{1 - e^{-c}}{c}C + e^{-c}\mathbb{1}, \tag{2.5}$$

with $C = \mathbb{0}$ for $c = 0$, so the summatory parameter of e^Q is always given by $\tilde{c} = 1 - e^{-c}$. For embeddability, one has the following well-known result; see [5, 20] for background.

Lemma 2.10. *The Markov matrix $M = \begin{pmatrix} 1 - a & a \\ b & 1 - b \end{pmatrix}$ with $a, b \in [0, 1]$ is embeddable if and only if $\det(M) > 0$, which is equivalent to the condition $0 \leqslant a + b < 1$. In this case, there is precisely one generator Q such that $M = e^Q$, namely*

$$Q = -\frac{\log{(1 - a - b)}}{a + b}(M - \mathbb{1}),$$

which is an equal-input generator.

Proof. The first statement is Kendall's theorem; see [1, Thm. 3.1] for a complete formulation. The uniqueness claim is established in [1, Eq. (5) and Cor. 3.3]. $\square$

Put differently, since $\mathcal{C}_2 = \mathcal{M}_2$, a matrix $M \in \mathcal{M}_2$ is embeddable if and only if its summatory parameter satisfies $0 \leqslant c < 1$. The closure of the set of embeddable matrices in $\mathcal{M}_2$ consists of all infinitely divisible elements of $\mathcal{M}_2$, as we shall discuss in more detail later, in Theorem 3.8 and Example 3.15.

Remark 2.11. The equation $e^x = 1$ has precisely one solution $x \in \mathbb{R}$, namely $x = 0$. In contrast, $e^A = \mathbb{1}$ with $A \in \mathrm{Mat}(2, \mathbb{R})$ has already infinitely many solutions, including $A = n\begin{pmatrix} 0 & -2\pi \\ 2\pi & 0 \end{pmatrix}$ with $n \in \mathbb{Z}$. Restricting A to real matrices with zero row sums restores uniqueness, because one eigenvalue of A is then 0, hence also the second, by the *spectral mapping theorem* (SMT), as A is real. Since A must be diagonalisable by [1, Fact 2.15], $A = \mathbb{0}$ is the only solution.

Whenever $A = (a_{ij})_{1 \leqslant i, j \leqslant d}$ is a Markov generator with $e^A = \mathbb{1}$, for arbitrary d, one has $1 = \det(e^A) = e^{\mathrm{tr}(A)}$ and thus $0 = \mathrm{tr}(A) = -\sum_{i \neq j} a_{ij}$. With $a_{ij} \geqslant 0$ for all $i \neq j$ by the generator property, this gives $a_{ij} = 0$ for all $i \neq j$, hence also $a_{ii} = 0$ for all i. Consequently, $A = \mathbb{0}$ is the only generator with $e^A = \mathbb{1}$. However, already for $d = 3$, there are further solutions of $e^A = \mathbb{1}$ among the real matrices with zero row sums, which is one reason why the embedding problem becomes significantly more complicated for $d \geqslant 3$.

In general, if Q is an equal-input generator, then so is $\frac{1}{n}Q$, for every $n \in \mathbb{N}$. Now, we can reformulate results from [1] and combine them with Kingman's characterisation of embeddability via regularity in conjunction with infinite divisibility [20, Prop. 7]. As this is compatible with the equal-input structure, we can summarise the general situation as follows.

Proposition 2.12. *When d is even, $M \in \mathcal{C}_d$ is embeddable if and only if $0 \leqslant c < 1$. When $d \geqslant 3$ is odd, there are further embeddable cases with $c > 1$.*

For arbitrary d and $M \in \mathcal{C}_d$, the following properties are equivalent:

1. *M has positive spectrum.*

2. *M is embeddable via an equal-input generator.*

3. *M is nonsingular and infinitely divisible within $\mathcal{C}_d$.*

4. *The summatory parameter of M satisfies $0 \leqslant c < 1$.* □

Note that, for $M \in \mathcal{C}_d$ with summatory parameter c, the even/odd dichotomy with the dimension emerges from Equation (2.2). A concrete example of an embeddable matrix $M \in \mathcal{C}_3$ with $c > 1$ is discussed in [5, Ex. 16] and [1, Ex. 4.3]. This M is infinitely divisible within $\mathcal{M}_3$, but not within $\mathcal{C}_3$. In fact, since $\sqrt[n]{M}$ has spectrum $\left\{ 1, \exp\left(\frac{-\pi\sqrt{3}}{n} \pm \frac{\mathrm{i}\pi}{n} \right) \right\}$ in this case, M does not possess an nth root of equal-input type for any $n \geqslant 2$.

Example 2.13. Within $\mathcal{C}_d$ lies the submonoid of *constant-input* matrices [1, Rem. 4.8], which all are of the form $M_c := \mathbb{1} + c J_d$ with $0 \leqslant c \leqslant \frac{d}{d-1}$, where $J_d = \frac{d-1}{d}(G_d - \mathbb{1})$ with G_d from (2.4) is a constant-input generator with summatory parameter 1, and hence $J_d^2 = -J_d$. Clearly J_d, as well as every constant-input matrix, is diagonalisable. If $c \in [0, 1)$, the spectral radius of $c J_d$ is $c < 1$, and a simple calculation with $\log(\mathbb{1} + c J_d)$ gives

$$M_c := \mathbb{1} + c J_d = \exp\bigl(-\log(1-c) J_d\bigr).$$

For d even, by Proposition 2.12, no constant-input Markov matrix with $c > 1$ can be embeddable, while this changes for $d \geqslant 3$ odd.

Assume that d is odd and M_c with $c > 1$ is embeddable, so $M = \mathrm{e}^Q$, where we also have $[J_d, Q] = \mathbb{0}$. So, by [1, Lemma 4.10 and Fact 2.15], Q is doubly stochastic and diagonalisable. As the eigenvalues of M_c are 1 and $1 - c < 0$, the latter with multiplicity $d - 1$, the spectrum of Q cannot be real. In particular, Q is not symmetric. Still, for any $a \in [0, 1)$, we get

$$M_a M_c = \mathrm{e}^{-\log(1-a) J_d} \mathrm{e}^Q = \mathrm{e}^{Q - \log(1-a) J_d},$$

and $M_{f(a,c)}$ is embeddable as well, where f is the function from Lemma 2.5.

When $c > 1$ is fixed and a varies in $[0, 1)$, $f(a, c)$ runs through $(1, c]$. Now, the infinitely divisible elements of $\mathcal{M}_d$ form a closed subset, as follows by a standard compactness argument via convergent subsequences. Since M_c with $c > 1$ is nonsingular, we see that there is a number $c_{\max} \in (1, 2]$ such that M_c is embeddable precisely for all $c \in [0, 1) \cup (1, c_{\max}]$. These constant-input matrices form a monoid that inherits the C_2-grading from Proposition 2.6. For the case $d = 3$, we know from [1, Cor. 6.6] that $c_{\max} = 1 + \mathrm{e}^{-\pi\sqrt{3}}$. The determination of $c_{\max} > 1$ for $d = 2m + 1$ with $m \geqslant 2$ is an interesting open question.

When $d \geqslant 3$, the embedding of $M \in \mathcal{C}_d$ need no longer be unique as for $d = 2$, but one still has the following property.

Lemma 2.14. *Let $d \geqslant 2$ and let $M \in \mathcal{C}_d$ be embeddable. If M admits a representation of the form $M = e^Q$ with a generator Q of equal-input type, the latter is unique in the sense that no other embedding can have an equal-input generator.*

Proof. The claim is obvious for $M = \mathbb{1}$, where $Q = \mathbb{0}$ is the only generator that solves $\mathbb{1} = e^Q$; compare Remark 2.11. Next, let Q and Q' be equal-input generators, with summatory parameters c and c', where we may now assume that $cc' > 0$. If $e^Q = e^{Q'}$, Equation (2.5) implies

$$e^{-c}\mathbb{1} + \frac{1 - e^{-c}}{c}\, C = e^{-c'}\mathbb{1} + \frac{1 - e^{-c'}}{c'}\, C'.$$

As both C and C' are equal-row matrices, this can only hold when $e^{-c} = e^{-c'}$; hence $c = c'$, which in turn forces $C = C'$ and thus $Q = Q'$ as claimed. $\qquad\square$

When $M \in \mathcal{C}_d$ is equal-input embeddable, so $M = M_C$ for some nonnegative matrix C with parameter sum $0 \leqslant c < 1$ by Proposition 2.12, the unique generator from Lemma 2.14 is given by

$$Q = -\frac{\log(1 - c)}{c}\, (M_C - \mathbb{1}), \tag{2.6}$$

meaning $Q = \mathbb{0}$ for $c = 0$, which is a nice extension of Lemma 2.10. The derivation rests on the observation that $c < 1$ is the spectral radius of $M_C - \mathbb{1}$, which permits the use of the standard branch of the matrix logarithm and its power series.

Let us next expand on an observation made in [1], in the context of an effective BCH formula for embeddable equal-input matrices. Here, one considers products of exponentials of equal-input generators, including the complicated case of noncommuting ones.

Theorem 2.15. *Let Q and Q' be equal-input generators, with summatory parameters c and c', respectively. Then, one has $e^Q e^{Q'} = e^{Q''}$ with*

$$Q'' = \frac{c + c'}{c\left(1 - e^{-(c+c')}\right)}\left(e^{-c'}\left(1 - e^{-c}\right)Q + \frac{c}{c'}\left(1 - e^{-c'}\right)Q'\right),$$

interpreted appropriately for $c = 0$ or $c' = 0$, where Q'' is again an equal-input generator. In particular, when $[Q, Q'] = \mathbb{0}$, the formula simplifies to $Q'' = Q + Q'$.

Proof. Let C and C' be the constant-row matrices underneath Q and Q'. Using the second identity from Equation (2.5) in conjunction with the relation $CC' = c\, C'$, one finds

$$e^Q e^{Q'} = e^{-(c+c')}\mathbb{1} + \frac{e^{-c'}(1 - e^{-c})}{c}\, C + \frac{1 - e^{-c'}}{c'}\, C'. \tag{2.7}$$

Since the summatory parameters of e^Q and $e^{Q'}$ are $\tilde{c} = 1 - e^{-c}$ and $\tilde{c}' = 1 - e^{-c'}$, which both lie in $[0, 1)$, the product $e^Q e^{Q'}$ is an equal-input matrix that has the summatory parameter $\tilde{c}'' \in [0, 1)$ by Lemma 2.5. As such, it is equal-input embeddable by Proposition 2.12; see also [1, Thm. 4.6]. Consequently, there exists an equal-input generator $Q'' = C'' - c''\mathbb{1}$ with

$$e^Q e^{Q'} = e^{Q''} = e^{-c''}\mathbb{1} + \frac{1 - e^{-c''}}{c''}\, C'', \tag{2.8}$$

where the last step follows once more from (2.5).

A comparison of (2.7) and (2.8) reveals that equality can only hold when the summatory parameters satisfy $c'' = c + c'$, which then gives

$$C'' = \frac{c + c'}{1 - e^{-(c+c')}} \left(\frac{e^{-c'}(1 - e^{-c})}{c} C + \frac{1 - e^{-c'}}{c'} C' \right).$$

Inserting $C = Q + c\mathbb{1}$ and the analogous terms for C' and C'' leads to the formula stated.

The condition $[Q, Q'] = \mathbb{0}$, which includes the case that one generator is $\mathbb{0}$, is equivalent to $c'C = c\, C'$, which also gives $c'Q = c\, Q'$. Inserting this into the formula for Q'' produces the claimed simplification after a short calculation. $\qquad\square$

In Theorem 2.15, the summatory parameters of e^Q, $e^{Q'}$, and $e^{Q''}$ are always related by

$$\tilde{c}'' = \tilde{c} + \tilde{c}' - \tilde{c}\tilde{c}' = f(\tilde{c}, \tilde{c}') = 1 - e^{-(c+c')} \,<\, 1,$$

in accordance with Fact 2.4 and Lemma 2.5. Various other aspects of equal- and constant-input Markov matrices have been discussed in [1], although without consideration of their connection with idempotents. We rectify this omission now by exploring some ideas in this direction that will prove useful later.

2.3. Markov idempotents and equal-input matrices

When $d = 2$, the only Markov idempotents are $\mathbb{1}$ and all the equal-input matrices $\alpha E_1 + (1 - \alpha)E_2$ with $\alpha \in [0, 1]$. This situation is deceptively simple, as one realises already for $d = 3$. Still, some general considerations are possible, some of which can be considered as a refinement of Fact 2.1.

Lemma 2.16. *Let $M = (m_{ij})_{1 \leqslant i, j \leqslant d}$ be a Markov idempotent that is also a positive matrix, so $m_{ij} > 0$ for all $i, j \in [d]$. Then, M is equal-input with $M = C$ for some positive equal-row matrix C with parameter sum $c = 1$.*

Proof. Let m_{ij} be a maximal element in column j of M, so $m_{kj} \leqslant m_{ij}$ holds for all $k \in [d]$. Then, with $M^2 = M$, we get the inequality

$$m_{ij} = \left(M^2\right)_{ij} = \sum_{k=1}^{d} m_{ik}\, m_{kj} \,\leqslant\, m_{ij} \sum_{k=1}^{d} m_{ik} = m_{ij}, \qquad (2.9)$$

where we actually have equality. When all matrix elements are positive, this is only possible if $m_{kj} = m_{ij}$ holds for all $k \in [d]$, which means that column j of M is constant.

Since $j \in [d]$ was arbitrary, the above argument applies to any column of M, and we obtain $M = c_1 E_1 + \ldots + c_d E_d$, where all $c_i > 0$ with $c = c_1 + \ldots + c_d = 1$ because M is Markov. $\square$

The statement of Lemma 2.16 can also be understood via the Perron–Frobenius theorem, as M under the assumed conditions is primitive. Then, $M_\infty = \lim_{n \to \infty} M^n = M$ is the projector to the unique equilibrium vector of M, which is $(c_1, \ldots, c_d)$. Note that being idempotent implies that primitivity of M becomes equivalent with positivity of M.

When M fails to be positive, there are further cases. These are driven by the added possibility of having equality in (2.9) due to the presence of vanishing matrix elements.

Example 2.17. Let us look at Markov idempotents for $d = 3$. When $M^2 = M$ is positive, the complete answer is provided by Lemma 2.16, so we only need to analyse cases with zero

entries. Since the set of idempotents within $\mathcal{M}_d$ is closed, it clearly contains the simplex

$$\{c_1 E_1 + c_2 E_2 + c_3 E_3 : c_i \geqslant 0 \text{ and } c_1 + c_2 + c_3 = 1\},$$

in line with Fact 2.1. This simplex includes the three $\{0, 1\}$ matrices E_1, E_2, and E_3, which are its extremal elements. Each matrix in this simplex has a unique equilibrium vector.

Further, one finds that the matrices

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & a & 1-a \\ 0 & a & 1-a \end{pmatrix}, \quad \begin{pmatrix} a & 0 & 1-a \\ 0 & 1 & 0 \\ a & 0 & 1-a \end{pmatrix}, \quad \text{and} \quad \begin{pmatrix} a & 1-a & 0 \\ a & 1-a & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

are idempotents for all $a \in [0, 1]$. For each matrix, its equilibrium vectors form a 1-simplex, hence with two extremal vectors. By choosing $a = 0$ or $a = 1$, we obtain six further $\{0, 1\}$ matrices. So far, all nine of them are equal-input or blockwise equal-input, possibly after a state permutation, and the only missing $\{0, 1\}$ Markov idempotent is $M = \mathbb{1}$.

Next, again for any $a \in [0, 1]$, the matrices

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ a & 1-a & 0 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 & 0 \\ a & 0 & 1-a \\ 0 & 0 & 1 \end{pmatrix}, \quad \text{and} \quad \begin{pmatrix} 0 & a & 1-a \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

are Markov idempotents. They do not produce new $\{0, 1\}$ matrices. Also, for $0 < a < 1$, they are not of equal-input form, not even blockwise, which means that more complicated cases do occur. We leave it as an exercise to the interested reader to verify that we have covered all cases for $d = 3$, and to analyse them further.

At this point, it seems worthwhile to recall the structure of general Markov idempotents, where we follow [16, Sec. 1.6]. Given any idempotent $M \in \mathcal{M}_d$ of rank r, where we must have $1 \leqslant r \leqslant d$, there is a partition of $[d] = \{1, 2, \ldots, d\}$ of the form

$$[d] = Z \dot{\cup} K_1 \dot{\cup} \ldots \dot{\cup} K_r \tag{2.10}$$

with $Z = \{\ell : \text{the } \ell\text{th column of } M \text{ is } 0\}$. Without any further specification of the K_i, we know from the definition of Z that the matrix elements of M satisfy $m_{ij} = 0$ for all $i \in [d]$ and every $j \in Z$. Further, given some subset $K \subseteq [d]$, we follow standard notation and use $M\big|_{K \times K}$ for the restriction of M to indices from K. Now, we can reformulate [16, Thm. 1.16] as follows.

Theorem 2.18. *Let $M = (m_{ij})_{1 \leqslant i,j \leqslant d} \in \mathcal{M}_d$ be an idempotent of rank r. Then, there is a partition of $[d]$ as in (2.10) with the following properties:*

1. *For all $s \in [r]$, one has $m_{ij} = 0$ for all $i \in K_s$ and $j \in [d] \setminus K_s$.*

2. *For all $s \in [r]$, the restriction $M\big|_{K_s \times K_s}$ is a Markov matrix with equal, positive rows.*

3. *For all $i \in Z$ and $s \in [r]$, and then every $k, \ell \in K_s$, one has $m_{ik} m_{k\ell} = m_{i\ell} m_{kk}$, where $m_{kk} m_{k\ell} \neq 0$ by (2).*

Conversely, every Markov matrix $M \in \mathcal{M}_d$ with a partition of $[d]$ as in (2.10) with these three properties is an idempotent. □

Specialising this classification result to $\{0, 1\}$ Markov matrices gives the following consequence, the explicit derivation of which we leave to the interested reader. It can also be derived directly by a careful analysis of $\{0, 1\}$ Markov matrices.

Corollary 2.19. *Possibly after an appropriate state permutation, any idempotent $\{0, 1\}$ Markov matrix appears in a block form where each block is an equal-input matrix with a single column of 1s.*   □

Simple examples, with an indication of the blocks according to Theorem 18(2), include

$$
\begin{pmatrix} \boxed{1} & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & \boxed{1} \end{pmatrix}, \quad
\begin{pmatrix} 0 & 1 & 0 \\ 0 & \boxed{1} & 0 \\ 0 & 0 & \boxed{1} \end{pmatrix} \quad \text{or} \quad
\begin{pmatrix} \boxed{1} & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & \boxed{1} & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix} \sim
\begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix},
$$

where the similarity in the third case is under the obvious state permutation. On the other hand, the matrices

$$
\begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \quad \text{and} \quad
\begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}
$$

fail to be idempotent, for instance. Let us now analyse the second class of matrices mentioned in the introduction before we return to this type of structure.

3. Monotone Markov matrices and embeddability

Let us begin this section with a formalisation of some of our previous recollections. To this end, we follow [18] and employ the lower-triangular matrix $T \in \mathrm{Mat}(d, \mathbb{R})$ given by

$$
T = \begin{pmatrix} 1 & & \mathbf{0} \\ \vdots & \ddots & \\ 1 & \cdots & 1 \end{pmatrix}, \tag{3.1}
$$

together with its inverse, T^{-1}, which has entries 1 on the diagonal, -1 on the first subdiagonal, and 0 everywhere else. With T and vectors $x, y \in \mathbb{P}_d$, one has the equivalence

$$
x \preccurlyeq y \quad \Longleftrightarrow \quad xT \leqslant yT, \tag{3.2}
$$

where the inequality on the right means that it is satisfied element-wise.

3.1. Monotone Markov matrices

If $E_{(i,j)} \in \mathrm{Mat}(d, \mathbb{R})$ denotes the elementary matrix with a single 1 in position (i, j) and 0 everywhere else, which results in

$$
E_{(k,\ell)} E_{(m,n)} = \delta_{\ell,m} E_{(k,n)}, \tag{3.3}
$$

one obtains the relations

$$
E_{(i,j)} T = E_{(i,1)} + \ldots + E_{(i,j)} \quad \text{and} \quad T^{-1} E_{(i,j)} = E_{(i,j)} - E_{(i+1,j)},
$$

where $E_{(d+1,j)} := \mathbb{0}$. Furthermore, we call a column vector $v = (v_1, \ldots, v_d)^T$ *nondecreasing* if $v_i \leqslant v_{i+1}$ holds for all $i \in [d-1]$. Now, we can characterise monotone matrices as follows.

Fact 3.1. *For a Markov matrix $M \in \mathcal{M}_d$, the following statements are equivalent:*

1. *The matrix M is monotone.*

2. *The mapping $x \mapsto xM$ preserves the partial order $\preccurlyeq$ on the positive cone.*

3. *One has $T^{-1}MT \geqslant \mathbb{0}$, understood element-wise, with T as in Equation (3.1).*

4. *Whenever v is a nondecreasing vector, Mv is also nondecreasing.*

Further, the same equivalences hold for any nonnegative $B \in \mathrm{Mat}(d, \mathbb{R})$ with equal row sums.

Proof. By our definition, (1) is equivalent to preserving $\preccurlyeq$ on $\mathbb{P}_d$, which clearly extends to all level sets $\alpha\mathbb{P}_d$ with $\alpha > 0$, so (1) $\Longleftrightarrow$ (2) is clear. The equivalences (2) $\Longleftrightarrow$ (3) $\Longleftrightarrow$ (4) are now an immediate consequence of [18, Thm. 1.1].

The case $B = \mathbb{0}$ is trivial. When $B \neq \mathbb{0}$, with elements b_{ij}, is a nonnegative matrix with equal row sums, meaning that $\sum_{j=1}^{d} b_{ij} = b > 0$ for all $i \in [d]$, the matrix $M = \frac{1}{b}B$ is Markov, and the final claim follows from the compatibility of the partial order on the positive cone with scaling by b and the fact that the conditions in (3) and (4) are linear in M. $\square$

Example 3.2. Let $M = (m_{ij})_{1 \leqslant i,j \leqslant d}$ be Markov. When $d = 2$, the nonnegativity of $T^{-1}MT$ is equivalent to the single condition $\mathrm{tr}(M) \geqslant 1$, or alternatively to

$$m_{22} \geqslant m_{12}.$$

Likewise, for $d = 3$, the original monotonicity condition for M, or equivalently the nonnegativity condition for $T^{-1}MT$, boils down to

$$m_{33} \geqslant m_{23} \geqslant m_{13} \quad \text{and} \quad m_{11} \geqslant m_{21} \geqslant m_{31},$$

where it was used that all rows of M sum to 1.

Let $E_{\ell_1,\ldots,\ell_d}$ denote the $\{0, 1\}$ Markov matrix with the row vector e_{ℓ_i} as row i, for $i \in [d]$, so

$$E_{\ell_1,\ldots,\ell_d} = E_{(1,\ell_1)} + \ldots + E_{(d,\ell_d)}.$$

There exist d^d such matrices, which are the extremal elements of $\mathcal{M}_d$. Since $e_i \preccurlyeq e_j$ if and only if $i \leqslant j$, it is clear that $E_{\ell_1,\ldots,\ell_d}$ is monotone precisely when $\ell_1 \leqslant \ell_2 \leqslant \cdots \leqslant \ell_d$. Using (3.3), or alternatively tracing the images of the basis vectors e_i, one verifies the multiplication rule

$$E_{k_1,\ldots,k_d}E_{\ell_1,\ldots,\ell_d} = E_{\ell_{k_1},\ldots,\ell_{k_d}}. \tag{3.4}$$

Owing to the existence of singular idempotents among these matrices (when $d \geqslant 2$), one thus obtains the following simple, but helpful, structure result.

Fact 3.3. *For $d \geqslant 2$, the set of $\{0, 1\}$ Markov matrices, under matrix multiplication, is a monoid, but not a group. The same property holds for the subset of monotone $\{0, 1\}$ matrices.*

A $\{0, 1\}$ Markov matrix in $\mathcal{M}_d$ is nonsingular if and only if it is a permutation matrix. The subset of the $d!$ permutation matrices is isomorphic with the symmetric group S_d.

Now, we turn to the convexity structure of $\mathcal{M}_{d,\preccurlyeq}$. While this is certainly known, we are not aware of a source with a proof, whence we include one for convenience.

Lemma 3.4. *The set $\mathcal{M}_{d,\preccurlyeq}$ is convex. It has $\binom{2d-1}{d}$ extremal points, which are the monotone Markov matrices with entries in $\{0, 1\}$, that is, the $E_{\ell_1,\ldots,\ell_d}$ with $1 \leqslant \ell_1 \leqslant \ell_2 \leqslant \cdots \leqslant \ell_d \leqslant d$.*

Proof. Convexity is clear, for instance via Fact 3.1 (3). Consequently, any convex combination of monotone $\{0, 1\}$ Markov matrices must lie in $\mathcal{M}_{d,\preccurlyeq}$. Thus, we first have to show that such convex combinations exhaust $\mathcal{M}_{d,\preccurlyeq}$. This follows from a greedy algorithm that is based on the following reduction argument.

Consider a nonnegative matrix $B \neq \mathbb{0}$ with equal row sums, b say, where $b > 0$. Assume that B is monotone. Such a matrix, because of the monotonicity condition, appears in a (non-reduced) row-echelon form. This is captured in a set of integer pairs $\big((i_1, j_1), \ldots, (i_r, j_r)\big)$, where j_1 is the position of the first (or leftmost) nonzero column of B, with i_1 the lowest position of a positive element in it, j_2 then is the leftmost position of a column that is nonzero below row i_1, with i_2 the lowest position of a positive element in column j_2, and so on. Clearly, $r \geqslant 1$ and $i_r = d$ since $b > 0$ and B is monotone.

For instance, B may have the row-echelon form

$$
B = \begin{pmatrix}
0 & * & \cdot & \cdot & \cdot & \cdot \\
0 & * & \cdot & \cdot & \cdot & \cdot \\
0 & \bullet & \cdot & \cdot & \cdot & \cdot \\
0 & 0 & \bullet & \cdot & \cdot & \cdot \\
0 & 0 & 0 & 0 & * & \cdot \\
0 & 0 & 0 & 0 & \bullet & \cdot
\end{pmatrix},
$$

here with $r = 3$ and integer pairs $((3, 2), (4, 3), (6, 5))$. A symbol $\bullet$ marks the lowest positive element in a column, and $*$ any element in the same column (above $\bullet$) that cannot be smaller (as a consequence of monotonicity). In each row, there is thus either one $*$ or one $\bullet$ by this rule. The total number of symbols of type $\bullet$ or $*$ is d, so 6 in this particular case. To the left of them, all elements are 0, while the remaining elements of B are left unspecified, as they play no role at this stage.

Now, let $\alpha > 0$ be the minimal element in the $\bullet$ positions, which are the (i_k, j_k), and let E denote the matrix that has a 1 in every $\bullet$ and in every $*$ position and a 0 anywhere else, which obviously is a monotone $\{0, 1\}$ Markov matrix, namely the $E_{\ell_1,\ldots,\ell_d}$ where ℓ_i is the unique position of $*$ or $\bullet$ in row i, for $i \in [d]$. Now, set $B' = B - \alpha E$, which is still monotone and has constant row sums $b' = b - \alpha \geqslant 0$, but one $\bullet$ is now replaced by a 0, which means that this $\bullet$ is gone or has moved up or right (or both) in the matrix. Unless $B' = \mathbb{0}$, we repeat the procedure with the new row-echelon form, which terminates after finitely many steps. The result is a decomposition of B as a sum of monotone $\{0, 1\}$ Markov matrices with positive weight factors. If we start with M, which has equal row sums $b = 1$, it is clear that we end up with a convex combination.

TABLE 1. The 10 extremal elements E_{ℓ_1,ℓ_2,ℓ_3} of $\mathcal{M}_{3,\preccurlyeq}$ with some of their properties. Here, $\sigma(M)$ is the spectrum of M with multiplicities, while p and q are, up to an overall sign, the characteristic and the minimal polynomial of M. Note that $p \neq q$ precisely when M is an idempotent. The last column gives M^2 in terms of its index parameters.

(ℓ_1, ℓ_2, ℓ_3)	M	$\sigma(M)$	p	q	M^2
$(1,1,1)$	$\left(\begin{smallmatrix} 1&0&0 \\ 1&0&0 \\ 1&0&0 \end{smallmatrix}\right)$	$\{1,0,0\}$	$x^2(x-1)$	$x(x-1)$	$(1,1,1)$
$(1,1,2)$	$\left(\begin{smallmatrix} 1&0&0 \\ 1&0&0 \\ 0&1&0 \end{smallmatrix}\right)$	$\{1,0,0\}$	$x^2(x-1)$	$p(x)$	$(1,1,1)$
$(1,1,3)$	$\left(\begin{smallmatrix} 1&0&0 \\ 1&0&0 \\ 0&0&1 \end{smallmatrix}\right)$	$\{1,1,0\}$	$x(x-1)^2$	$x(x-1)$	$(1,1,3)$
$(1,2,2)$	$\left(\begin{smallmatrix} 1&0&0 \\ 0&1&0 \\ 0&1&0 \end{smallmatrix}\right)$	$\{1,1,0\}$	$x(x-1)^2$	$x(x-1)$	$(1,2,2)$
$(1,2,3)$	$\left(\begin{smallmatrix} 1&0&0 \\ 0&1&0 \\ 0&0&1 \end{smallmatrix}\right)$	$\{1,1,1\}$	$(x-1)^3$	$x-1$	$(1,2,3)$
$(1,3,3)$	$\left(\begin{smallmatrix} 1&0&0 \\ 0&0&1 \\ 0&0&1 \end{smallmatrix}\right)$	$\{1,1,0\}$	$x(x-1)^2$	$x(x-1)$	$(1,3,3)$
$(2,2,2)$	$\left(\begin{smallmatrix} 0&1&0 \\ 0&1&0 \\ 0&1&0 \end{smallmatrix}\right)$	$\{1,0,0\}$	$x^2(x-1)$	$x(x-1)$	$(2,2,2)$
$(2,2,3)$	$\left(\begin{smallmatrix} 0&1&0 \\ 0&1&0 \\ 0&0&1 \end{smallmatrix}\right)$	$\{1,1,0\}$	$x(x-1)^2$	$x(x-1)$	$(2,2,3)$
$(2,3,3)$	$\left(\begin{smallmatrix} 0&1&0 \\ 0&0&1 \\ 0&0&1 \end{smallmatrix}\right)$	$\{1,0,0\}$	$x^2(x-1)$	$p(x)$	$(3,3,3)$
$(3,3,3)$	$\left(\begin{smallmatrix} 0&0&1 \\ 0&0&1 \\ 0&0&1 \end{smallmatrix}\right)$	$\{1,0,0\}$	$x^2(x-1)$	$x(x-1)$	$(3,3,3)$

No monotone $\{0,1\}$ Markov matrix can be written as a convex combination of the other ones, so their extremality is clear. Now, given d, the monotone $\{0,1\}$ Markov matrices are in obvious bijection with the possibilities to distribute d indistinguishable balls (the 1s) to d distinguishable boxes (the columns of the matrix), where an outcome $(n_1, \ldots, n_d)$, with $n_1 + \ldots + n_d = d$, parameterises the matrix M with the row vector e_1 in the first n_1 rows, then e_2 in the next n_2 rows, and so on. The total number of possible outcomes is well known to be $\binom{2d-1}{d-1} = \binom{2d-1}{d}$, as this is the number of choices to place $d-1$ separating walls between the d balls on the altogether $2d-1$ positions; see [26, A001700] for details. $\quad\square$

The case $d = 3$ is summarised in Table 1. By Lemma 3.4, every monotone Markov matrix $M \in \mathcal{M}_d$ can be expressed as a convex combination of the form

$$M = \sum_{1 \leqslant \ell_1 \leqslant \cdots \leqslant \ell_d \leqslant d} \alpha_{\ell_1,\ldots,\ell_d} E_{\ell_1,\ldots,\ell_d} \tag{3.5}$$

with all coefficients $\alpha_{\ell_1,\ldots,\ell_d} \geqslant 0$, their sum being 1, and $E_{\ell_1,\ldots,\ell_d}$ as above. Observing that $\operatorname{tr}(E_{\ell_1,\ldots,\ell_d}) \geqslant 1$ whenever $\ell_1 \leqslant \ell_2 \leqslant \cdots \leqslant \ell_d$, one finds from (3.5) that $\operatorname{tr}(M) \geqslant 1$ holds for all monotone Markov matrices, which easily generalises as follows.

Corollary 3.5. *Let $B \in \operatorname{Mat}(d, \mathbb{R})$ be a nonnegative matrix with equal row sums, $b \geqslant 0$. If B is also monotone, one has $\operatorname{tr}(B) \geqslant b$.* $\quad\square$

3.2. Monotonicity and embeddability

As in [1], we use $\mathcal{E}_d$ to denote the semigroup generated by the embeddable Markov matrices of dimension d. For $d = 2$, every element of $\mathcal{E}_2$ is itself embeddable (so $\mathcal{E}_2 = \mathcal{M}_2^E$, which is no longer true for $d \geqslant 3$), and the set of monotone Markov matrices agrees with the closure of $\mathcal{E}_2$. In fact, $\mathcal{M}_{2,\preccurlyeq}$ is the closed triangle in $\mathcal{M}_2$ with the vertices being $\mathbb{1}_2$, $E_1 = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}$, and $E_2 = \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}$. Only the line $\{\alpha E_1 + (1 - \alpha)E_2 : 0 \leqslant \alpha \leqslant 1\}$ does not belong to $\mathcal{E}_2$, because it consists of singular idempotents. The only other idempotent in $\mathcal{M}_2$ is $\mathbb{1}$, the trivial case. This leads to the following result.

Proposition 3.6. *An element $M \in \mathcal{M}_2$ is monotone if and only if $\mathrm{tr}(M) \geqslant 1$. Thus, being monotone is equivalent to either being embeddable or being a nontrivial idempotent.*

Proof. Observe that $\mathrm{tr}(M) \leqslant 2$ holds for all $M \in \mathcal{M}_2$. Since $M = \begin{pmatrix} 1 - a & a \\ b & 1 - b \end{pmatrix}$ with numbers $a, b \in [0, 1]$ is monotone if and only if $1 \geqslant a + b$ (compare Example 3.2), the first claim is immediate. By Lemma 2.10, $\mathrm{tr}(M) = 1$ means M is monotone, but not embeddable.

For the second claim, recall that $M \in \mathcal{M}_2$ is an idempotent if and only if $M^2 = M$, which implies $\sigma(M) \subseteq \{0, 1\}$. Since $\lambda = 1$ is always an eigenvalue, being an idempotent means either that the second eigenvalue is also 1, hence $M = \mathbb{1}$ by Lemma 2.3, or that $\det(M) = 0$, which gives the line from E_1 to E_2 discussed above. $\qquad\square$

Corollary 3.7. *Any $M \in \mathcal{M}_{2,\preccurlyeq}$ is infinitely divisible within $\mathcal{M}_{2,\preccurlyeq}$. In fact, $M \in \mathcal{M}_2$ is infinitely divisible if and only if it is monotone.*

Proof. Let $M \in \mathcal{M}_2$ be monotone. By Proposition 3.6, the case $\det(M) = 0$ means $M^2 = M$, so also $M^n = M$ for all $n \in \mathbb{N}$ by induction, and M is a monotone nth root of itself. Clearly, the latter statement also applies to $M = \mathbb{1}$.

When $M = \begin{pmatrix} 1 - a & a \\ b & 1 - b \end{pmatrix} \neq \mathbb{1}$ is embeddable, we have $a + b > 0$ and $M = \mathrm{e}^Q$ with the unique generator Q from Lemma 2.10. Then, for any $n \in \mathbb{N}$, a Markov nth root of M is given by

$$\exp\left(\tfrac{1}{n}Q\right) = \begin{pmatrix} 1 - \epsilon a & \epsilon a \\ \epsilon b & 1 - \epsilon b \end{pmatrix} \quad \text{with } \epsilon = \frac{1 - \sqrt[n]{1 - a - b}}{a + b},$$

as follows from the same standard calculation with the matrix exponential that was used to derive (2.6). Now, $\exp\left(\tfrac{1}{n}Q\right)$ is monotone if and only if $1 \geqslant \epsilon(a + b)$, by an application of the criterion from Example 3.2. But this estimate follows from $0 < a + b < 1$ because $\epsilon \in [0, 1]$.

It remains to show that infinite divisibility of $M \in \mathcal{M}_2$ implies its monotonicity, which can be derived from the spectrum as follows. If 1 is the only eigenvalue of M, we have $M = \mathbb{1}$ by Lemma 2.3, which is embeddable. Otherwise, one has $\sigma(M) = \{1, \lambda\}$ where $\lambda \neq 1$ must be real, with $|\lambda| \leqslant 1$. Since M has a Markov square root by assumption, we get $\lambda \in [0, 1)$. Now, $\lambda = 0$ means that M is an idempotent, while $\lambda > 0$ implies $\det(M) > 0$, so M is embeddable by Lemma 2.10. Monotonicity of M now follows from Proposition 3.6. $\qquad\square$

Our next goal is a better understanding of the connection between $\mathcal{M}_{d,\preccurlyeq}$ and $\mathcal{C}_d$, aiming at generalisations of Corollary 3.7 to general d. To this end, we once more consider a nonnegative

matrix C with equal rows and parameter sum c, with $C = \mathbb{0}$ only when $c = 0$. For $c > 0$, the matrix $\frac{1}{c}C$ is both Markov and monotone. Consequently, the Markov matrix M_C from Equation (2.1), for any $c \in [0, 1]$, is a convex combination of $\mathbb{1}$ and $\frac{1}{c}C$, hence monotone as well.

When $c = 0$, which means $M_C = \mathbb{1}$, or when $c = 1$, where $M_C = C$, the matrix M_C is a monotone idempotent. When $c \in (0, 1)$, Equation (2.6) implies

$$M_C = \mathrm{e}^Q \quad \text{with } Q = -\frac{\log (1 - c)}{c} Q_C, \tag{3.6}$$

where $Q_C = M_C - \mathbb{1}$ as before, which is an equal-input generator. Now, for arbitrary $n \in \mathbb{N}$, a standard calculation with the exponential series gives the formula

$$\exp\left(\tfrac{1}{n}Q\right) = \mathbb{1} + \frac{1 - \sqrt[n]{1 - c}}{c} Q_C = \sqrt[n]{1 - c}\,\mathbb{1} + \left(1 - \sqrt[n]{1 - c}\right)\tfrac{1}{c}C. \tag{3.7}$$

Since $\sqrt[n]{1 - c} \in (0, 1)$ under our assumptions, this is a convex combination of two monotone Markov matrices, hence monotone and Markov itself. We have thus proved the following generalisation of our previous statements, assuming $d \geqslant 2$ as usual. In particular, the idempotent elements play a similar role as in the two-dimensional case.

Theorem 3.8. *Let C, Q_C, and M_C be as above, and let $c = c_1 + \ldots + c_d$ be the corresponding parameter sum. If $c \in [0, 1]$, M_C is Markov and monotone, with the following properties:*

1. *M_C is an idempotent if and only if $c \in \{0, 1\}$, where $c = 0$ means $M_C = \mathbb{1}$.*

2. *M_C is embeddable if and only if $c \in [0, 1)$, with $Q = \mathbb{0}$ for $c = 0$ or the generator Q from (3.6) otherwise.*

In particular, $M_C = \mathbb{1}$ is the only embeddable idempotent. Further, for all $c \in [0, 1]$ and for every $n \in \mathbb{N}$, M_C has a Markov nth root that is both equal-input and monotone, which is to say that M_C is infinitely divisible within $\mathcal{M}_{d, \preccurlyeq} \cap \mathcal{C}_d$.

Note that, also in generalisation of the case $d = 2$, the set of monotone Markov matrices of type M_C with summatory parameter $c \in [0, 1]$ is the closure of the set of equal-input Markov matrices that are embeddable with an equal-input generator, with all non-embeddable boundary cases being nontrivial idempotents. In fact, one has more as follows.

Corollary 3.9. *A Markov matrix $M \in \mathcal{C}_d$ is monotone if and only if its summatory parameter satisfies $c \in [0, 1]$. So, one obtains the convex set*

$$\mathcal{C}_{d, \preccurlyeq} := \mathcal{C}_d \cap \mathcal{M}_{d, \preccurlyeq} = \big\{ M \in \mathcal{C}_d : c \in [0, 1] \big\},$$

with the $d + 1$ extremal elements $E_1, \ldots, E_d$ and $\mathbb{1}$.

Further, $\mathcal{C}_{d, \preccurlyeq}$ is the disjoint union of the set of equal-input embeddable elements from $\mathcal{C}_d$ with the set $\mathcal{C}_{d, 1}$ of nontrivial idempotents in $\mathcal{C}_d$. The eigenvalues of any $M \in \mathcal{C}_{d, \preccurlyeq}$ are real and nonnegative, and they are positive precisely for the embeddable cases.

Proof. It is clear from Theorem 3.8 that all $M \in \mathcal{C}_d$ with $c \in [0, 1]$ are monotone, so we need to show that no further element of $\mathcal{C}_d$ is. To this end, consider a Markov matrix of the form $M = (1 - c)\mathbb{1} + C$ with $c > 1$, which implies that all $c_i > 0$. Then it is easy to check that $T^{-1}MT$ fails to be a nonnegative matrix, where T is the matrix from (3.1), and M fails to be monotone by Fact 3.1 (3).

When $c \in [0, 1]$, we have $0 \leqslant c_i \leqslant c_1 + \ldots + c_d = c \leqslant 1$ for all $1 \leqslant i \leqslant d$, and

$$M_C = (1 - c)\mathbb{1} + C = (1 - c)\mathbb{1} + \sum_{i=1}^{d} c_i E_i$$

is a convex combination, where the extremality of $E_1, \ldots, E_d$ and $\mathbb{1}$ is clear.

Another application of Theorem 3.8 gives the decomposition claimed, while the statement on the spectrum is clear because the eigenvalues of M_C are 1 and $1 - c$. □

Geometrically, the situation is that the simplex $\mathcal{C}_{d,\preccurlyeq}$ separates the compact set $\mathcal{C}_d$ into the subset with $c \in [0, 1)$, which are the 'good' cases for embeddability, and the subset with $c \in \left(1, \frac{d}{d-1}\right]$, where embeddability requires d even and further conditions, but is never possible with an equal-input generator. For $d = 2$, we refer to [1, Fig. 1] for an illustration.

One can view $\mathcal{C}_{d,\preccurlyeq}$ differently when starting in $\mathcal{M}_{d,\preccurlyeq}$. Let S_d be the symmetric (or permutation) group of d elements, and P_π for $\pi \in S_d$ the standard permutation matrix that represents the linear mapping $e_i \mapsto e_{\pi(i)}$ under multiplication to the right. P_π has elements $\delta_{i,\pi(j)}$ and satisfies $P_\pi^{-1} = P_{\pi^{-1}}$. There are $d!$ such matrices, which are the extremal elements among the doubly stochastic matrices mentioned earlier. The conjugation action by such a matrix gives

$$P_\pi E_{(k,\ell)} P_\pi^{-1} = E_{(\pi(k),\pi(\ell))} \quad \text{and} \quad P_\pi E_{\ell_1,\ldots,\ell_d} P_\pi^{-1} = E_{\pi\left(\ell_{\pi^{-1}(1)}\right),\ldots,\pi\left(\ell_{\pi^{-1}(d)}\right)},$$

as follows from a simple calculation with $E_{\ell_1,\ldots,\ell_d} = E_{(1,\ell_1)} + \ldots + E_{(d,\ell_d)}$, or, alternatively, from tracing the images of the basis vectors e_i for $1 \leqslant i \leqslant d$.

Now, a set $\mathcal{F} \subseteq \mathcal{M}_d$ of Markov matrices is called *permutation-invariant* if $P_\pi \mathcal{F} P_\pi^{-1} = \mathcal{F}$ holds for all $\pi \in S_d$. Clearly, $\mathcal{M}_d$ itself is such a set, as is $\mathcal{C}_d$ or its subset of constant-input matrices. The latter are also *individually* permutation-invariant (which is also called *exchangeable* in probability theory [8]); that is, $P_\pi M P_\pi^{-1} = M$ holds for every constant-input matrix M and all $\pi \in S_d$. In fact, the Markov matrices that are individually permutation-invariant are precisely the constant-input ones, without restriction on the summatory parameter c.

The set of all $\{0, 1\}$ Markov matrices is permutation-invariant as well, and partitions into S_d-orbits of the form $\mathcal{O}_{S_d}(M) = \{P_\pi M P_\pi^{-1} : \pi \in S_d\}$. Two such orbits are

$$\mathcal{O}_{S_d}(\mathbb{1}) = \{\mathbb{1}\} \quad \text{and} \quad \mathcal{O}_{S_d}(E_1) = \{E_1, \ldots, E_d\},$$

which both consist of monotone matrices only. One can check that no other orbit of the decomposition has this property, which implies the following characterisation.

Fact 3.10. *The convex set $\mathcal{C}_{d,\preccurlyeq}$ is the maximal subset of $\mathcal{M}_{d,\preccurlyeq}$ that is permutation-invariant. The elements of $\mathcal{C}_{d,\preccurlyeq}$ that are individually permutation-invariant are the constant-input matrices with $c \in [0, 1]$.* □

Let us turn to Markov semigroups. Recall from [18] that a (homogeneous) Markov semigroup $\{e^{tQ} : t \geqslant 0\}$, with generator Q, is called *monotone* when e^{tQ} is monotone for every $t \geqslant 0$. Moreover, a generator Q is called *monotone* if all off-diagonal elements of $T^{-1} Q T$ are nonnegative, where T is the matrix from Equation (3.1). This concept is motivated by the following connection, which is a minor variant of [18, Thm. 2.1]. Because of its importance, we include a short proof that is tailored to our later needs in this context.

Proposition 3.11. *If Q is a Markov generator, the following properties are equivalent:*

1. *The semigroup $\{\mathrm{e}^{tQ} : t \geqslant 0\}$ is monotone.*

2. *The generator Q is monotone.*

Proof. For $(1) \Rightarrow (2)$, observe that $T^{-1}\mathrm{e}^{tQ}T \geqslant \mathbb{0}$ implies $\left(T^{-1}\frac{1}{t}(\mathrm{e}^{tQ} - \mathbb{1})T\right)_{ij} \geqslant 0$ for $t > 0$ and all $i \neq j$. Then, taking $t \searrow 0$ establishes this direction.

For $(2) \Rightarrow (1)$, is it clear that $T^{-1}QT + \alpha\mathbb{1} \geqslant \mathbb{0}$ holds for any sufficiently large $\alpha > 0$. Choose α also large enough so that $M_\alpha := \mathbb{1} + \alpha^{-1}Q$ is Markov, which is clearly possible. Then $T^{-1}M_\alpha T \geqslant \mathbb{0}$, and we get $T^{-1}M_\alpha^m T = \left(T^{-1}M_\alpha T\right)^m \geqslant \mathbb{0}$ for all integers $m \geqslant 0$. Now, observe

$$\mathrm{e}^{tQ} = \mathrm{e}^{-\alpha t}\mathrm{e}^{\alpha t M_\alpha} = \sum_{m=0}^{\infty} \mathrm{e}^{-\alpha t}\frac{(\alpha t)^m}{m!}\, M_\alpha^m,$$

which, for all $t \geqslant 0$, constitutes a convergent sum that is a convex combination of monotone Markov matrices. Consequently, e^{tQ} is monotone as well. $\square$

Example 3.12. Let $Q = \left(q_{ij}\right)_{1 \leqslant i,j \leqslant d}$ be a Markov generator. When $d = 2$, it is always monotone, that is, no extra condition emerges; compare Proposition 3.6. When $d = 3$, being monotone is equivalent to the two conditions

$$q_{23} \geqslant q_{13} \quad \text{and} \quad q_{21} \geqslant q_{31},$$

which provide a surprisingly simple criterion for monotonicity in this case.

The above considerations have the following consequence.

Corollary 3.13. *For $M \in \mathcal{M}_{d,\preccurlyeq}$, the following properties are equivalent:*

1. *M is embeddable via a monotone Markov generator.*

2. *M is nonsingular and infinitely divisible within $\mathcal{M}_{d,\preccurlyeq}$.*

Proof. $(1) \Rightarrow (2)$: Let $M = \mathrm{e}^Q$ with Q a monotone generator, so $\det(M) = \mathrm{e}^{\mathrm{tr}(Q)} > 0$. Now, $\frac{1}{n}Q$ is still a monotone generator, for any $n \in \mathbb{N}$, and $\exp\left(\frac{1}{n}Q\right)$ is a monotone Markov matrix (by Proposition 3.11) that is also an nth root of M.

$(2) \Rightarrow (1)$: By Kingman's characterisation [20, Prop. 7], M is embeddable, so $M = \mathrm{e}^Q$ with some generator Q. We need to show that Q can be chosen to be monotone. Let R_n be an nth root of M that is Markov and monotone, which exists and implies that $A_n := n(R_n - \mathbb{1})$ is a monotone generator. By a standard compactness argument, there is a subsequence $\left(n_i\right)_{i \in \mathbb{N}}$ of increasing integers such that $Q' = \lim_{i \to \infty} A_{n_i}$ is a monotone generator as well. From here on, we can employ Kingman's original proof to conclude that

$$M = \left(\mathbb{1} + \frac{Q'}{n_i}\right)^{n_i} + o(1) \qquad \text{as } i \to \infty,$$

which gives $M = \mathrm{e}^{Q'}$ as claimed. $\square$

3.3. Idempotents and infinite divisibility

At this point, it seems worthwhile to take a closer look at infinite divisibility in general. In this context, we refer to [8, Sec. X.9] for the underlying (pseudo-)Poissonian structures.

Proposition 3.14. *Let P_0, $P \in \mathcal{M}_d$ be chosen such that $P_0^2 = P_0$, so P_0 is an idempotent, and that $P_0 P = PP_0 = P$. Then, the matrix family $\{M(t) : t \geqslant 0\}$ with*

$$M(t) := \mathrm{e}^{-t}\left(P_0 + \sum_{m=1}^{\infty} \frac{t^m}{m!} P^m \right) = \mathrm{e}^{-t}\left(P_0 - \mathbb{1} + \mathrm{e}^{tP} \right)$$

satisfies the following properties, where $A := P - \mathbb{1}$ is a Markov generator:

1. *The mapping $t \mapsto M(t)$ is continuous, with $M(0) = P_0$.*

2. *$M(t)M(s) = M(t+s)$ holds for all $t, s \geqslant 0$.*

3. *$M(t)$ is Markov, for all $t \geqslant 0$.*

4. *$M(t) = \mathrm{e}^{-t}\left(P_0 - \mathbb{1} \right) + \mathrm{e}^{tA} = P_0 \mathrm{e}^{tA}$ for all $t \geqslant 0$.*

5. *$P = P_0$ if and only if $M(t) = P_0$ holds for all $t \geqslant 0$.*

6. *For $t \geqslant 0$, $M(t)$ is idempotent if and only if $M(t) = P_0$.*

In particular, $\{M(t) : t \geqslant 0\}$ always constitutes a continuous monoid, with P_0 as its neutral element, while it is a homogeneous Markov semigroup if and only if $P_0 = \mathbb{1}$.

   Proof. (1) is obvious, while (2) follows from a standard calculation with the convergent series. Both P_0 and P are Markov, and so is P^m for all $m \in \mathbb{N}$. Now, for any $t \geqslant 0$, $M(t)$ is a convergent, convex combination of Markov matrices, hence Markov as well, which shows (3).

   Next, (4) and the easy direction of (5) follow from elementary calculations, using the expansion $P_0 \mathrm{e}^{tP} = P_0 + \sum_{m \geqslant 1} \frac{t^m}{m!} P^m$ together with $\mathrm{e}^{tP} = \mathrm{e}^t \mathrm{e}^{tA}$. When $M(t) = P_0$ for all $t \geqslant 0$, one obtains $P_0 + \mathrm{e}^{tP} = \mathbb{1} + \mathrm{e}^t P_0$. But, observing $\mathrm{e}^{tP_0} = \mathbb{1} - P_0 + \mathrm{e}^t P_0$, this implies $\mathrm{e}^{tP} = \mathrm{e}^{tP_0}$ for all $t \geqslant 0$ and thus $P = P_0$.

   For (6), one direction is trivial. The other follows from (4) upon the observation that $M(t)$ idempotent implies $P_0 \mathrm{e}^{tA} = M(t) = M(t)^2 = M(2t) = P_0 \mathrm{e}^{2tA}$ and hence gives the relation $P_0 = P_0 \mathrm{e}^{tA} = M(t)$ as claimed.

   Finally, while the (abstract) semigroup and monoid properties are clear, the family can only be a homogeneous Markov semigroup when $M(t) = \mathrm{e}^{tQ}$ for some generator Q and all $t \geqslant 0$, so $M(0) = \mathbb{1}$, which is the only nonsingular idempotent in $\mathcal{M}_d$ by Lemma 2.3. We thus have $P_0 = \mathbb{1}$ in this case, and (4) implies $M(t) = \mathrm{e}^{tA}$ as claimed. $\qquad \square$

Example 3.15. Let us analyse the meaning of Proposition 3.14 for $d \geqslant 2$. If $P_0 = \mathbb{1}$, there is no further restriction on P, and $M(t) = \mathrm{e}^{tA}$ with $A = P - \mathbb{1}$, which can thus be any generator with diagonal elements in $[-1, 0]$. This way, possibly after rescaling t, all embeddable matrices are covered. For $P \in \mathcal{C}_d$ with $c \in [0, 1)$, we see that $A = P - \mathbb{1}$ is a generator of equal-input type, in line with Lemma 2.10 and Proposition 2.12.

   If $P_0 \in \mathcal{C}_{d,1}$, we get $P_0 = \sum_{i=1}^{d} \beta_i E_i$ with $\beta_i \geqslant 0$ and $\beta_1 + \ldots + \beta_d = 1$; hence $P_0 = PP_0$ for any $P \in \mathcal{M}_d$. Assuming $PP_0 = P_0 P = P$, we find $P = P_0$, which then gives $M(t) \equiv P_0$ by Proposition 3.14 (5).

   For $d = 2$, where $\mathcal{M}_2 = \mathcal{C}_2$, this exhausts all cases because no further idempotents exist. Consequently, $M \in \mathcal{M}_2$ is infinitely divisible if and only if it is embeddable or an idempotent, with $M = \mathbb{1}$ being the only case that is both; compare Proposition 3.6 and Corollary 3.7.

When $d \geqslant 3$, one obtains mixtures via direct sums, where $\mathbb{1} \oplus P_0$ and $(\mathbb{1} + A) \oplus P_0$ lead to $M(t) = \mathrm{e}^{tA} \oplus P_0$. There are further examples for $d = 3$ (compare Example 2.17), such as

$$P_0 = \alpha E_{1,1,3} + (1 - \alpha) E_{1,3,3} = \begin{pmatrix} 1 & 0 & 0 \\ \alpha & 0 & 1 - \alpha \\ 0 & 0 & 1 \end{pmatrix}, \quad \text{with } \det(P_0) = 0,$$

which is idempotent for any $\alpha \in [0, 1]$. Then, $P \in \mathcal{M}_3$ with $PP_0 = P_0 P = P$ leads to

$$P = \begin{pmatrix} a & 0 & 1 - a \\ c & 0 & 1 - c \\ 1 - b & 0 & b \end{pmatrix}$$

with $a, b \in [0, 1]$ and $c = \alpha a + (1 - \alpha)(1 - b)$. Here, $M(t) = P_0 \mathrm{e}^{tA}$ is singular for all $t \geqslant 0$ and thus never embeddable. Moreover, for $P \neq P_0$, the matrix $M(t)$ can only be idempotent for $t = 0$ and possibly for *isolated* further values of $t > 0$. This is so because $M(t) = M(t)^2 = M(2t)$ implies $\mathrm{e}^{tP} = \mathrm{e}^{tP_0}$ via an elementary calculation. When this holds for t from a set with an accumulation point, t_0 say, standard arguments imply $P = P_0$. So, more interesting as well as more complicated cases emerge in $\mathcal{M}_d \backslash \mathcal{C}_d$ as d grows.

Lemma 3.16. *With P_0, P, and $M(t)$ as in Proposition 3.14, one has either $\det(M(t)) > 0$ for all $t \geqslant 0$, which happens if and only if $P_0 = \mathbb{1}$, or $\det(M(t)) = 0$ for all $t \geqslant 0$, which is true whenever $P_0 \neq \mathbb{1}$, or equivalently whenever P_0 is singular.*

Proof. Recall via Lemma 2.3 that $P_0^2 = P_0$ either means $\det(P_0) = 1$, which forces $P_0 = \mathbb{1}$, or $\det(P_0) = 0$. Now, observe that $P_0 M(t) = M(t)$ holds for all $t \geqslant 0$, so

$$\det(P_0) \det(M(t)) = \det(M(t)),$$

and $\det(M(t)) \equiv 0$ for singular P_0 is immediate.

The only remaining case is $P_0 = \mathbb{1}$. Here, Proposition 3.14 (3) implies $M(t) = \mathrm{e}^{tA}$ with $A = P - \mathbb{1}$ and hence $\det(M(t)) = \mathrm{e}^{\mathrm{tr}(tA)} > 0$. $\square$

A semigroup as in Proposition 3.14 is called *Poissonian* if $P_0 = \mathbb{1}$, and *pseudo-Poissonian* otherwise [8, Sec. X.1]. We can now recall the central classification result on infinitely divisible, finite-dimensional Markov matrices from [11] as follows. It seems a bit hidden in the literature, but nicely underpins the role of idempotents in the embedding problem.

Theorem 3.17. *A Markov matrix $M \in \mathcal{M}_d$ is infinitely divisible if and only if there are Markov matrices P_0, $P \in \mathcal{M}_d$, with $P_0^2 = P_0$ and $P_0 P = P P_0 = P$, and some $s \geqslant 0$ such that*

$$M = \mathrm{e}^{-s} \left(P_0 + \sum_{m=1}^{\infty} \frac{s^m}{m!} P^m \right).$$

Moreover, M is embeddable if and only if one also has $P_0 = \mathbb{1}$.

Proof. For the proof of the first claim, we refer to [11].

For the second claim, we know that M embeddable implies $\det(M) > 0$, and we are in the case with $P_0 = \mathbb{1}$ by Lemma 3.16. Conversely, when $P_0 = \mathbb{1}$, Proposition 3.14 (4) gives $M = \mathrm{e}^{sA}$ with the generator $A = P - \mathbb{1}$. $\square$

Note that the parameter s cannot be avoided in this formulation, because a generator Q can have diagonal entries of arbitrarily large negative value, whence $\mathbb{1} + Q$ need not be Markov, while $\mathbb{1} + sQ$, for all suitably small $s > 0$, will be; compare Example 3.15 and Proposition 3.14.

Further consequences can be derived from $[P_0, P] = \mathbb{0}$ when P is *cyclic*, which means that the minimal and characteristic polynomial of P agree. In particular, this is the case when P is simple; see [1, Fact 2.10] for a systematic characterisation of cyclic matrices. Whenever $P \in \mathcal{M}_d$ is cyclic, its *centraliser* is the abelian ring

$$\mathrm{cent}(P) = \{B \in \mathrm{Mat}(d, \mathbb{R}) : [P, B] = \mathbb{0}\} = \mathbb{R}[P],$$

where each element of this ring is of the form $\sum_{n=0}^{d-1} \alpha_n P^n$ with all $\alpha_n \in \mathbb{R}$, as a consequence of the Cayley–Hamilton theorem. In particular, P_0 is then an idempotent from this ring. We leave further details to the interested reader.

4. Monotone Markov matrices in three dimensions

Let us now look at $d = 3$ in more detail, stating the following simple and certainly well-known property, which we also prove owing to lack of reference.

Proposition 4.1. *The eigenvalues of any $M \in \mathcal{M}_{3, \preccurlyeq}$ are real. Moreover, at most one eigenvalue of M can be negative, which happens if and only if* $\det(M) < 0$.

Further, if $d = 3$ and Q is a monotone Markov generator, its eigenvalues are nonpositive real numbers.

Proof. First, $M \in \mathcal{M}_{3, \preccurlyeq} \subset \mathcal{M}_3$ implies $1 \in \sigma(M)$. As $d = 3$, the characteristic polynomial of M then is $(1 - x)\big(x^2 - (\mathrm{tr}(M) - 1)x + \det(M)\big)$, and the remaining two eigenvalues are

$$\lambda_{\pm} = \frac{1}{2}\Big(\mathrm{tr}(M) - 1 \pm \sqrt{\Delta}\,\Big), \tag{4.1}$$

with the discriminant $\Delta = (\mathrm{tr}(M) - 1)^2 - 4 \det(M)$. By an explicit calculation, where one first eliminates m_{22} and later also m_{11} via the row sum condition, one verifies that

$$\Delta = \big(m_{11} - m_{21} + m_{23} - m_{33}\big)^2 + 4\big(m_{23} - m_{13}\big)\big(m_{21} - m_{31}\big) \geqslant 0,$$

where the inequality follows from the monotonicity of M via Example 3.2.

This implies $\sigma(M) \subset \mathbb{R}$, and the formula for $\lambda_{\pm}$ from Equation (4.1) shows that at most λ_{-} can be negative, because $\mathrm{tr}(M) \geqslant 1$ by Corollary 3.5. Also, when $\det(M) = 0$, the spectrum is $\sigma(M) = \{1, \mathrm{tr}(M) - 1, 0\}$, which is nonnegative. This establishes the claims on M.

If $Q = \big(q_{ij}\big)_{1 \leqslant i,j \leqslant 3}$ is a Markov generator, its spectrum contains 0, while the other two eigenvalues are given by

$$\mu_{\pm} = \frac{1}{2}\Big(\mathrm{tr}(Q) \pm \sqrt{D}\,\Big), \tag{4.2}$$

where, in analogy to above, one finds

$$D = \big(q_{11} - q_{21} + q_{23} - q_{33}\big)^2 + 4\big(q_{23} - q_{13}\big)\big(q_{21} - q_{31}\big) \geqslant 0.$$

Here, the inequality follows via the monotonicity criterion from Example 3.12.

Consequently, all eigenvalues are real. They are then nonpositive because all eigenvalues of Markov generators have real part $\leqslant 0$ (compare [1, Prop. 2.3(1)]), as can also be checked explicitly from (4.2), where $\mathrm{tr}(Q) \leqslant 0 \leqslant \sqrt{D} \leqslant |\mathrm{tr}(Q)|$. □

Considering the convex combinations

$$M(\alpha) := \alpha E_{1,1,2} + (1-\alpha)E_{2,3,3} = \begin{pmatrix} \alpha & 1-\alpha & 0 \\ \alpha & 0 & 1-\alpha \\ 0 & \alpha & 1-\alpha \end{pmatrix} \quad \text{with } \alpha \in (0, 1),$$

in the notation of Table 1, one finds $\mathrm{tr}(M(\alpha)) = 1$ and $\det(M(\alpha)) = -\alpha(1-\alpha) < 0$. This shows that cases with a simple negative eigenvalue exist. On the other hand, Proposition 4.1 also means that the spectrum of a matrix $M \in \mathcal{M}_{3,\preccurlyeq}$ is positive if and only if $\det(M) > 0$.

Corollary 4.2. *Consider any $M \in \mathcal{M}_{3,\preccurlyeq}$ with $\det(M) < 0$. Then, M is neither embeddable nor can it have a monotone nth Markov root for n even. In fact, M has no Markov or, indeed, real square root at all, and M is not infinitely divisible.*

Proof. Since $\det(M) < 0$, embeddability is ruled out immediately, and so is the existence of a real square root because M has only a simple negative eigenvalue; compare [14, Thm. 6.6] or [15]. Consequently, M cannot be infinitely divisible.

Further, $B \in \mathcal{M}_{3,\preccurlyeq}$ with $B^{2m} = M$ for any $m \in \mathbb{N}$ implies nonnegative spectrum for M since $\sigma(B) \subset \mathbb{R}$, which contradicts $\det(M) < 0$. $\qquad\square$

Let us now analyse when a matrix $M \in \mathcal{M}_{3,\preccurlyeq}$ is embeddable. For this, M must be non-singular and thus, by Corollary 4.2, have positive spectrum. So, all eigenvalues of M must satisfy $0 < \lambda \leqslant 1$. Further, $A = M - \mathbb{1}$ is a generator that inherits monotonicity from M as all off-diagonal elements of $T^{-1}AT$ are nonnegative. Since the spectral radius of A is $\varrho_A < 1$,

$$Q := \log(\mathbb{1} + A) = \sum_{m=1}^{\infty} \frac{(-1)^{m-1}}{m} A^m \tag{4.3}$$

converges and defines a real matrix with zero row sums and $M = e^Q$. As we are not interested in other types of solutions, we now introduce the non-unital, real algebra

$$\mathcal{A}_0^{(3)} := \{B \in \mathrm{Mat}(3, \mathbb{R}) : \text{all row sums of } B \text{ are } 0\}, \tag{4.4}$$

which certainly contains the Q from (4.3). Since positive spectrum of M means $\sigma(M) \subset (0, 1]$, all eigenvalues of Q are nonpositive real numbers by the SMT. It remains to analyse the generator property and potential uniqueness of Q.

Let q be the minimal polynomial of A. Its degree satisfies $\deg(q) \in \{1, 2, 3\}$ and equals the degree of the minimal polynomial of M. Further, let

$$\mathrm{alg}(A) := \langle A^m : m \in \mathbb{N} \rangle_{\mathbb{R}}$$

be the real algebra spanned by the positive powers of A, which does not contain $\mathbb{1}$, as one can check easily. In fact, one has $\mathrm{alg}(A) = \langle A, A^2 \rangle_{\mathbb{R}}$ by the Cayley–Hamilton theorem, so $\mathrm{alg}(A)$ is a subalgebra of $\mathcal{A}_0^{(3)}$ of dimension $\leqslant 2$. Indeed, we clearly get

$$\dim\big(\mathrm{alg}(A)\big) = \deg(q) - 1 \in \{0, 1, 2\} \tag{4.5}$$

together with $Q \in \mathrm{alg}(A)$ by (4.3), which leads to different situations as follows.

Proposition 4.3. *Let the matrix $M \in \mathcal{M}_{3,\preccurlyeq}$ have a minimal polynomial of degree $\leqslant 2$. Then, the following properties are equivalent:*

1. *The spectrum of M is positive.*

2. *One has $\det(M) > 0$.*

3. *M is embeddable.*

4. *M is embeddable with a monotone generator.*

If M is embeddable, there is precisely one monotone generator Q with $M = e^Q$, namely the one given in Equation (4.3).

Proof. The implications $(4) \Rightarrow (3) \Rightarrow (2)$ are clear, while $(2) \Longleftrightarrow (1)$ follows from Proposition 4.1. It remains to show $(1) \Rightarrow (4)$, so assume $\sigma(M) \subset \mathbb{R}_+ := \{x \in \mathbb{R} : x > 0\}$. As above, let q be the minimal polynomial of $A = M - \mathbb{1}$, which has the same degree as that of M.

When $\deg(q) = 1$, since 0 is always an eigenvalue of A, the only possibility is $q(x) = x$; hence $A = \mathbb{0}$ and thus also $Q = \mathbb{0}$ from (4.3), which gives the trivial case $\mathbb{1} = \exp(\mathbb{0})$, where $\mathbb{1}$ is the only matrix in $\mathcal{M}_{3,\preccurlyeq}$ with a minimal polynomial of degree 1. By Remark 2.11, $Q = \mathbb{0}$ is the only generator with $e^Q = \mathbb{1}$, which is trivially monotone.

When $\deg(q) = 2$, we have $A \neq \mathbb{0}$, and hence $\text{tr}(A) < 0$, so by (4.5) we get $A^2 = -\alpha A$ for some $\alpha \in \mathbb{R}$, where $\text{tr}(A^2) > 0$ implies $\alpha > 0$. Here, $A = M - \mathbb{1}$ is diagonalisable, with eigenvalues 0 and $-\alpha > -1$, in line with $\sigma(M) \subset \mathbb{R}_+$. Then (4.3) simplifies to

$$Q = -\frac{\log(1-\alpha)}{\alpha} A,$$

which is a positive multiple of A and hence a monotone generator, so M is embeddable as $M = e^Q$.

To establish the uniqueness claim, consider any $Q' \in \mathcal{A}_0^{(3)}$ such that $M = e^{Q'}$, where M is diagonalisable by assumption, hence also Q' by [1, Fact 2.15]. We then still have $[Q', A] = \mathbb{0}$, but not necessarily $Q' \in \text{alg}(A)$. Now, by [1, Lemma 6.1], there are two possibilities, namely $\dim(\text{alg}(Q')) \in \{1, 2\}$. Here, if the dimension is 2, Q' must be simple, which is only possible if Q' has a complex conjugate pair of (non-real) eigenvalues. But then, Q' cannot be monotone, by Proposition 4.1. It remains to consider $\dim(\text{alg}(Q')) = 1$, where we get $\text{alg}(Q') = \text{alg}(A)$ from [1, Lemma 6.1(1)], and hence $Q' = aA$ for some $a > 0$. By taking the determinant on both ends of $e^{Q'} = M = e^Q$, which gives a positive number, one finds

$$a\,\text{tr}(A) = -\frac{\log(1-\alpha)}{\alpha}\,\text{tr}(A).$$

As $\text{tr}(A) \neq 0$, this implies $a = -\frac{\log(1-\alpha)}{\alpha}$ and thus $Q' = Q$. $\qquad\square$

Let us pause to state the asymptotic behaviour of M^n for the embeddable matrices covered by Proposition 4.3. In the above notation, one trivially has $M^n = \mathbb{1}$ for all n when $\deg(q) = 1$, while a simple calculation gives

$$M_\infty := \lim_{n \to \infty} M^n = \mathbb{1} + \lim_{n \to \infty} \frac{1-(1-\alpha)^n}{\alpha} A = \mathbb{1} + \frac{1}{\alpha} A$$

for the more interesting case that $\deg(q) = 2$.

When $\deg(q) = 3$, the situation becomes a little more complex. Here, $A = M - \mathbb{1}$ is always cyclic, with 0 being a simple eigenvalue. Then, we obtain $A^3 = rA + sA^2$ together with $r = \operatorname{tr}(M) - \det(M) - 2$ and $s = \operatorname{tr}(A)$. Since $\sigma(A) = \{0, \mu_+, \mu_-\}$, where $\mu_\pm$ are negative numbers by Proposition 4.1, we get $r = -\mu_+\mu_- < 0$ and $s = \mu_+ + \mu_- < 0$. This remains correct when $\mu_+ = \mu_-$, where A has a nontrivial Jordan normal form (as it is cyclic). Note that, although A is a *real* matrix, it is more convenient, and also completely consistent, to always employ the *complex* Jordan normal form of A in our arguments.

Let us first consider the case that A is *simple* (and hence also diagonalisable). Here, we have $-1 < \mu_- < \mu_+ < 0$ together with $\sigma(M) = \{1, 1 + \mu_+, 1 + \mu_-\}$. As A is cyclic, any matrix $Q \in \mathcal{A}_0^{(3)}$ with $M = \mathrm{e}^Q$ must lie in $\operatorname{alg}(A) = \mathbb{R}[A] \cap \mathcal{A}_0^{(3)}$, so $Q = \alpha A + \beta A^2$ for some $\alpha, \beta \in \mathbb{R}$, again by Frobenius's theorem. Then, the SMT implies

$$\alpha\mu_\pm + \beta\mu_\pm^2 = \log\left(1 + \mu_\pm\right) \in \mathbb{R},$$

which is an inhomogeneous system of linear equations for α and β. As

$$\det\begin{pmatrix} \mu_+ & \mu_+^2 \\ \mu_- & \mu_-^2 \end{pmatrix} = \mu_+\mu_-\left(\mu_- - \mu_+\right) < 0,$$

we get a unique solution, which is given by

$$\alpha = \frac{\mu_-^2 \log\left(1 + \mu_+\right) - \mu_+^2 \log\left(1 + \mu_-\right)}{\mu_+\mu_-\left(\mu_- - \mu_+\right)}, \quad \beta = \frac{\mu_+ \log\left(1 + \mu_-\right) - \mu_- \log\left(1 + \mu_+\right)}{\mu_+\mu_-\left(\mu_- - \mu_+\right)}.$$
(4.6)

This shows that $M = \mathrm{e}^Q$ has precisely one solution with $Q \in \mathcal{A}_0^{(3)}$, which must be the one from (4.3). One can check via the Taylor series that $\alpha > 0$ and $\beta < 0$, though this is not sufficient to guarantee the generator or the monotonicity property of Q. So far, we have derived the following constructive result.

Theorem 4.4. *Let $M \in \mathcal{M}_{3,\preccurlyeq}$ have simple spectrum with $\det(M) > 0$, and set $A = M - \mathbb{1}$, with $\sigma(A) = \{0, \mu_+, \mu_-\}$ as above. Then, there is precisely one $Q \in \mathcal{A}_0^{(3)}$ such that $M = \mathrm{e}^Q$, namely the matrix Q from (4.3), which also satisfies $Q = \alpha A + \beta A^2$ with α, β from (4.6).*

Further, Q is a generator, and $M = \mathrm{e}^Q$, precisely when $\left(\alpha A + \beta A^2\right)_{ij} \geqslant 0$ holds for all $i \neq j$, and Q is also monotone when the criterion from Example 3.12 is satisfied by $\alpha A + \beta A^2$.

It remains to consider the case that A is *cyclic*, but not simple. Here, its eigenvalues are 0 and $-1 < \mu < 0$, the latter twice; also, any solution of $M = \mathrm{e}^Q$ with $Q \in \mathcal{A}_0^{(3)}$ must be of the form $Q = \alpha A + \beta A^2$, which implies the condition $\alpha\mu + \beta\mu^2 = \log\left(1 + \mu\right)$ by the SMT. Using the standard Jordan normal form of A, which must comprise the Jordan block $\begin{pmatrix} \mu & 1 \\ 0 & \mu \end{pmatrix}$ because of our assumption, one obtains another condition from $\mathrm{e}^Q = \mathbb{1} + A$, this time from the superdiagonal element of the Jordan block, namely $(1 + \mu)(\alpha + 2\beta\mu) = 1$. This results in the unique solution

$$\alpha = 2\frac{\log\left(1 + \mu\right)}{\mu} - \frac{1}{1 + \mu} \quad \text{and} \quad \beta = \frac{1}{\mu(1 + \mu)} - \frac{\log\left(1 + \mu\right)}{\mu^2}.$$
(4.7)

Note that (4.7) also follows from (4.6) by an approximation argument of de L'Hospital type, via setting $\mu_- = \mu = \mu_+ + x$ and letting $x \to 0$. So, the Q from (4.3) is once more the only

solution for $M = \mathrm{e}^Q$ with $Q \in \mathcal{A}_0^{(3)}$. Here, $\alpha > 0$ for μ sufficiently large (above approximately -0.7) and $\beta < 0$, which is more complicated than in the previous case. Nevertheless, we have the following result.

Corollary 4.5. *Let $M \in \mathcal{M}_{3,\preccurlyeq}$ be cyclic, but not simple. Then, the matrix $A = M - \mathbb{1}$ has spectrum $\sigma(A) = \{0, \mu\}$ with $-1 < \mu < 0$, where μ has algebraic multiplicity 2, but geometric multiplicity 1. Further, all statements of Theorem* 4.4 *remain true, this time with the coefficients α, β from* (4.7).

Not all $M \in \mathcal{M}_{3,\preccurlyeq}$ with positive determinant (and hence spectrum) can be embeddable, as there are cases with a single 0 in one position; see [1, 5, 12] and references therein for further examples. For $d \geqslant 4$, the possibility of complex conjugate pairs of eigenvalues increases the complexity of the treatment, which is nevertheless possible with the recent results from [2].

5. Uniqueness of embedding and further directions

The explicit treatment of $\mathcal{M}_{3,\preccurlyeq}$ in the previous section shows that some useful sufficient criteria for unique embeddability should be in store, such as the one stated in [14, Sec. 2.3] for Markov matrices with distinct positive eigenvalues. Let us first recall a classic result on the existence of a real logarithm of a given matrix, which can be found in several places, for instance in [10, Sec. 8.8.2] or as [3, Thm. 1].

Fact 5.1 *For $B \in \mathrm{GL}(d, \mathbb{R})$, the equation $B = \mathrm{e}^R$ has a solution $R \in \mathrm{Mat}(d, \mathbb{R})$ if and only if every elementary Jordan block of B with an eigenvalue on the negative real axis occurs with even multiplicity. When B is diagonalisable, this simplifies to the condition that each eigenvalue of B on the negative real axis has even algebraic multiplicity.* $\square$

Any matrix $R \in \mathrm{Mat}(d, \mathbb{R})$ that solves $B = \mathrm{e}^R$ is called a *real logarithm* of B. When considering a nonsingular Markov matrix M, we are only interested in real logarithms of M with zero row sums, that is, in elements from the subalgebra $\mathcal{A}_0^{(d)} \subset \mathrm{Mat}(d, \mathbb{R})$. This is justified by the following observation. Suppose e^R has unit row sums with a real matrix R that fails to have zero row sums. Then the set of matrices e^{tR} with unit row sums and $t \in \mathbb{R}$ forms a discrete subgroup of $\{\mathrm{e}^{tR} : t \in \mathbb{R}\} \simeq \mathbb{R}$. This is so because the existence of an accumulation point with unit row sums, t_0 say, would result in $(1, \ldots, 1)^T$ being an eigenvector of R with eigenvalue 0, which is a contradiction.

In analogy to the previous case with $d = 3$, now for $A \in \mathcal{A}_0^{(d)}$ with arbitrary $d \geqslant 2$, we define the non-unital algebra

$$\mathrm{alg}(A) := \left\langle A^m : m \in \mathbb{N} \right\rangle_{\mathbb{R}} = \left\langle A, A^2, \ldots, A^{d-1} \right\rangle_{\mathbb{R}} \subset \mathcal{A}_0^{(d)},$$

where the second formulation follows from the Cayley–Hamilton theorem in conjunction with the fact that $\mathcal{A}_0^{(d)}$ is non-unital.

Lemma 5.2. *Let $M \in \mathcal{M}_d$ be cyclic and nonsingular, and assume that M possesses at least one real logarithm, according to Fact* 5.1. *Then, with $A = M - \mathbb{1}$, any real logarithm R of M satisfies $R \in \mathrm{alg}(A)$.*

Proof. Clearly, we have $[R, M] = [R, A] = \mathbb{0}$, so M cyclic implies $R \in \mathbb{R}[A]$ by Frobenius's theorem, and thus $R = \alpha_0 \mathbb{1} + \sum_{n=1}^{d-1} \alpha_n A^n$ for some $\alpha_0, \ldots, \alpha_{d-1} \in \mathbb{R}$. Hence, $R = \alpha_0 \mathbb{1} + X$ with $X \in \mathcal{A}_0^{(d)}$, where e^X then has unit row sums. Consequently, all row sums of $\mathrm{e}^R = \mathrm{e}^{\alpha_0} \mathrm{e}^X$ equal e^{α_0}, which must be 1. So we get $\alpha_0 = 0$ and $R \in \mathrm{alg}(A) \subset \mathcal{A}_0^{(d)}$ as claimed. $\square$

Now, we can extend the uniqueness result mentioned earlier to cyclic matrices. It is a variant of [3, Thm. 2], but we give a different and constructive proof that later leads to an effective (and numerically stable) criterion for embeddability. It generalises what we saw in Theorem 4.4 and Corollary 4.5, and also differs from the approach used in [2].

Theorem 5.3. *Suppose $M \in \mathcal{M}_d$ is cyclic and has real spectrum. Then, M possesses a real logarithm R, so $M = e^R$, if and only if the spectrum of M is positive. In this case, R is unique, and is always an element of $\mathrm{alg}(A) \subset \mathcal{A}_0^{(d)}$, where $A = M - \mathbb{1}$.*

Proof. When M is cyclic, no elementary Jordan block can occur twice, and the first implication follows from Fact 5.1. When $\sigma(M) \subset \mathbb{R}_+$, by Fact 5.1 and Lemma 5.2, all real logarithms of M must lie in $\mathrm{alg}(A)$, and there is at least one $R \in \mathrm{alg}(A)$ with $e^R = M$, so we have $R = \sum_{i=1}^{d-1} \alpha_i A^i$ for some $\alpha_1, \ldots, \alpha_{d-1} \in \mathbb{R}$. It remains to establish uniqueness.

First, assume that A is *simple*. As A is a generator, this means $\sigma(A) = \{0, \mu_1, \ldots, \mu_{d-1}\}$, with distinct $\mu_i \in (-1, 0)$ because of our assumptions. Then, since all α_i and μ_j are real, the SMT implies the $d-1$ identities

$$\sum_{\ell=1}^{d-1} \alpha_\ell \, \mu_i^\ell = \log\left(1 + \mu_i\right), \qquad 1 \leqslant i \leqslant d-1. \tag{5.1}$$

They constitute an inhomogeneous system of linear equations for the α_i with the matrix

$$B = \begin{pmatrix} \mu_1 & \mu_1^2 & \cdots & \mu_1^{d-1} \\ \mu_2 & \mu_2^2 & \cdots & \mu_2^{d-1} \\ \vdots & \vdots & & \vdots \\ \mu_{d-1} & \mu_{d-1}^2 & \cdots & \mu_{d-1}^{d-1} \end{pmatrix}. \tag{5.2}$$

Since $\det(B) = \left(\prod_i \mu_i\right) \prod_{k > \ell} (\mu_k - \mu_\ell)$ by an obvious variant of the standard Vandermonde determinant formula, B is invertible when A is simple, and (5.1) has a unique real solution.

When A is *cyclic*, but not simple, the appearance of nontrivial Jordan blocks necessitates a more refined argument. Clearly, as A is a generator and also cyclic, 0 is a simple eigenvalue of A by [1, Prop. 2.3(2)]. Let $\mu \in (-1, 0)$ be any of the other eigenvalues, say with algebraic multiplicity m. When $m = 1$, we get one condition from the SMT, and nothing else is needed. So, assume $m \geqslant 2$. As A is cyclic, the geometric multiplicity of μ is 1, and the corresponding Jordan block in standard form is $\mathbb{J}_\mu = \mu \mathbb{1}_m + N_m$, where N_m is the nilpotent matrix with entries 1 on the first superdiagonal and 0 everywhere else. It satisfies $N_m^m = \mathbb{0}$, while N_m^k, for $1 \leqslant k < m$, has entries 1 on the kth superdiagonal and 0 elsewhere. In this case, we get only *one* condition from the SMT, namely

$$\sum_{\ell=1}^{d-1} \alpha_\ell \, \mu^\ell = \log(1 + \mu), \tag{5.3}$$

as in (5.1), while we need $m - 1$ independent further ones. They will come from derivatives of (5.3), which needs a justification as follows.

First, from $e^R = \mathbb{1} + A$, one concludes that we must have

$$\exp\left(\sum_{\ell=1}^{d-1} \alpha_\ell \, \mathbb{J}_\mu^\ell\right) = \mathbb{1}_m + \mathbb{J}_\mu = (1 + \mu)\mathbb{1}_m + N_m. \tag{5.4}$$

Now, defining the polynomial $\phi(x) = \sum_{\ell=1}^{d-1} \alpha_\ell x^\ell$ and the function $\psi(x) = \mathrm{e}^{\phi(x)}$, one can employ the standard method from [14, Sec. 1.2.1] to calculate the exponential in (5.4) as

$$\psi(\mathbb{J}_\mu) = \psi(\mu)\,\mathbb{1}_m + \sum_{k=1}^{m-1} \frac{\psi^{(k)}(\mu)}{k!} N_m^k,$$

where $\psi^{(k)}$ denotes the kth derivative of ψ. As $\psi(\mu) = 1 + \mu$ by (5.3), a comparison with (5.4) leads to the conditions $\psi^{(k)}(\mu) = \delta_{k,1}$ for $1 \leqslant k \leqslant m - 1$, noting that $1 + \mu \neq 0$. Iterating the product rule on $\psi = \mathrm{e}^\phi$ and inserting (5.3) then results in

$$\frac{\mathrm{d}^k}{\mathrm{d}\mu^k} \sum_{\ell=k}^{d-1} \alpha_\ell\,\mu^\ell = \phi^{(k)}(\mu) = \frac{(-1)^{k-1}(k-1)!}{(1+\mu)^k} = \frac{\mathrm{d}^k}{\mathrm{d}\mu^k} \log(1+\mu), \tag{5.5}$$

which shows that the additional conditions on the α_ℓ emerge via derivatives of the fundamental relation (5.3). It remains to check when we obtain a *unique* solution $\alpha_1, \ldots, \alpha_{d-1}$ this way.

Here, the matrix B that generalises (5.2) is specified by an n-tuple

$$\big((\mu_1, m_1), \ldots, (\mu_n, m_n)\big)$$

of distinct nonzero eigenvalues μ_i of A with their algebraic multiplicities m_i, where $n \leqslant d - 1$ and $m_1 + \ldots + m_n = d - 1$. A pair (μ_i, m_i) is responsible for m_i rows of B of length $d - 1$, with the first derived from (5.3), followed by $m_i - 1$ rows induced by (5.5). Here, each new row emerges from the previous one by differentiation with respect to μ_i. The resulting matrix is a variant of the *confluent Vandermonde matrix* (see [13, Sec. 22.2] or [23]), which is known from Hermite interpolation. It is invertible if and only if the μ_i are distinct.

To substantiate the latter claim, define the sequence $(\gamma_n)_{n \in \mathbb{N}}$ by

$$\gamma_n = \det \begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & 2 & 4 & \cdots & 2^{n-1} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & n & n^2 & \cdots & n^{n-1} \end{pmatrix},$$

which starts as 1, 1, 2, 12, 288, 34560; compare [26, A000178]. Then one finds

$$\det(B) = \left(\prod_{i=1}^n \mu_i^{m_i} \gamma_{m_i} \right) \prod_{k > \ell} (\mu_k - \mu_\ell)^{m_k m_\ell}$$

(compare [13, Exc. 22.6]), which reduces to the determinant formula stated previously for the special case $m_1 = \ldots = m_n = 1$. Since the μ_i are distinct, $\det(B) \neq 0$ is clear, and the claimed uniqueness follows. $\square$

The benefit of this approach is that one can calculate B^{-1} and thus determine the coefficients purely from the eigenvalues of A. In particular, the unique R is a generator if and only if $\sum_{i=1}^{d-1} \alpha_i A^i$ satisfies the corresponding conditions.

Remark 5.4. The Vandermonde matrix and its inverse are well known from Lagrange interpolation theory. Therefore, in minor modification of the results from [17, Sec. 0.9.11], one finds the matrix elements of the inverse of B from (5.2) as

$$\left(B^{-1}\right)_{ij} = \frac{(-1)^{i-1} S_{d-1-i}^{(d-2)}\left(\{\mu_1, \ldots, \mu_{d-1}\} \setminus \{\mu_j\}\right)}{\mu_j \prod_{k \neq j} \left(\mu_k - \mu_j\right)},$$

where $S_m^{(n)}(\{a_1, \ldots, a_n\})$ is the elementary symmetric polynomial as defined by $S_0^{(n)} \equiv 1$, then $S_1^{(n)}(\{a_1, \ldots, a_n\}) = a_1 + \ldots + a_n$, followed by $S_2^{(n)}(\{a_1, \ldots, a_n\}) = \sum_{i<j} a_i a_j$, and so on, up to the final one, which is $S_n^{(n)}(\{a_1, \ldots, a_n\}) = \prod_i a_i$.

With a little more effort, this formula can be extended to the cyclic situation as well; see [23] for a constructive approach to B^{-1} in this case.

The above result leads to the following sufficient criterion for unique embeddability.

Corollary 5.5. *Let $M \in \mathcal{M}_d$ be cyclic and have real spectrum, $\sigma(M) \subset \mathbb{R}$. Then, M has a real logarithm if and only if $\sigma(M) \subset \mathbb{R}_+$.*

In this case, the spectral radius of $A = M - \mathbb{1}$ is $\varrho_A < 1$, and there is at most one Markov generator Q such that $M = e^Q$. The only choice is $Q = \log(\mathbb{1} + A) \in \text{alg}(A) \subset \mathcal{A}_0^{(d)}$, calculated with the standard branch of the matrix logarithm as a convergent series. In particular, M is embeddable if and only if the matrix $\log(\mathbb{1} + A)$ is a generator.

Proof. The first claim follows from Theorem 5.3. If $\sigma(M) \subset \mathbb{R}_+$, all eigenvalues of A lie in the half-open interval $(-1, 0]$, so $\varrho_A < 1$ is clear. By Theorem 5.3, as M is cyclic with $\sigma(M) \subset (0, 1]$, there is precisely one real matrix R with $M = e^R$. Since $\varrho_A < 1$, the series

$$\log(\mathbb{1} + A) = \sum_{m=1}^{\infty} \frac{(-1)^{m-1}}{m} A^m$$

converges. The limit is then an element of $\text{alg}(A)$, because this algebra is a closed subset of $\text{Mat}(d, \mathbb{R})$. So we get $R = \log(\mathbb{1} + A) \in \text{alg}(A)$ in this case, which has zero row sums, but need not be a generator. $\square$

One can go beyond this result, but care is required with the branches of the complex logarithm; see [14, Ch. 11] for background and [2] for recent progress in this direction.

Remark 5.6. The results from Theorem 4.4 and Corollary 4.5 in particular apply to all cyclic matrices $M \in \mathcal{M}_3$ with positive spectrum, in that the embeddability of M can most easily be verified via testing whether $\alpha A + \beta A^2$ is a generator, where $A = M - \mathbb{1}$ and $\alpha, \beta \in \mathbb{R}$ are the numbers from Equation (4.6) if M is simple, or from Equation (4.7) otherwise.

Clearly, uniqueness results have interesting consequences on the structure of Markov roots, as can be seen in the following refinement of [1, Ex. 3.9].

Example 5.7. The two-dimensional Markov matrix

$$M = \begin{pmatrix} \frac{3}{4} & \frac{1}{4} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}$$

is uniquely embeddable by Lemma 2.10, so $M = \mathrm{e}^Q$ with a unique generator Q. Nevertheless, as follows from a simple calculation, it has precisely two Markov square roots, namely

$$M_1 = \begin{pmatrix} \frac{5}{6} & \frac{1}{6} \\ \frac{1}{3} & \frac{2}{3} \end{pmatrix} \quad \text{and} \quad M_2 = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ 1 & 0 \end{pmatrix}.$$

Of these, $M_1 = \exp\left(\frac{1}{2}Q\right)$ is embeddable, while M_2 is not. So, in the embeddable case, there is always at least one Markov nth root for every $n \in \mathbb{N}$ of the form $\exp\left(\frac{1}{n}Q\right)$, but there can still be others. A uniqueness result for the embedding of M then means that, among all Markov roots, there is precisely one sequence of embeddable Markov nth roots of M.

The set $\mathcal{M}_d^{\mathrm{E}}$ of embeddable matrices is a relatively closed subset of $\{M \in \mathcal{M}_d : \det(M) > 0\}$ by [20, Prop. 3], but (for $d \geqslant 2$) it is *not* a closed subset of $\mathcal{M}_d$. The closure of $\mathcal{M}_d^{\mathrm{E}}$ is still a subset of the infinitely divisible elements of $\mathcal{M}_d$, and it is a natural question which matrices lie on the boundary, which we denote by $\partial \mathcal{M}_d^{\mathrm{E}}$. Clearly, there can be embeddable cases, such as $\begin{pmatrix} 1 & 0 \\ 1-\alpha & \alpha \end{pmatrix}$ or $\begin{pmatrix} \alpha & 1-\alpha \\ 0 & 1 \end{pmatrix}$ for $d = 2$ and $0 < \alpha \leqslant 1$, as well as (singular) idempotent ones, such as $M_\infty = \lim_{t \to \infty} \mathrm{e}^{tQ}$ for any generator $Q \neq \mathbb{0}$. For $d > 2$, there are further possibilities. Any $M \in \partial \mathcal{M}_d^{\mathrm{E}}$ satisfies $M = \lim_{n \to \infty} \mathrm{e}^{Q_n}$ for some sequence $(Q_n)_{n \in \mathbb{N}}$ of generators, which implies $\lim_{n \to \infty} \mathrm{e}^{\mathrm{tr}(Q_n)} = \det(M) \geqslant 0$.

When $M \in \partial \mathcal{M}_d^{\mathrm{E}}$ has $\det(M) > 0$, it is embeddable by Kingman's infinite divisibility criterion. Alternatively, a positive determinant implies that the sequence $\left(\mathrm{tr}(Q_n)\right)_{n \in \mathbb{N}}$ converges and is thus bounded. Since all diagonal elements of a generator are nonpositive, they are bounded as well, hence also all elements of the Q_n thanks to the vanishing row sums. By a standard compactness argument, there is thus a subsequence $(Q_{n_i})_{i \in \mathbb{N}}$ such that $\lim_{i \to \infty} Q_{n_i} = Q$ is a generator with $M = \mathrm{e}^Q$ as expected, and M lies in the Markov semigroup $\left\{\mathrm{e}^{tQ} : t \geqslant 0\right\}$.

When $\det(M) = 0$, the sequence $(\mathrm{tr}(Q_n))_{n \in \mathbb{N}}$ must be (negatively) unbounded, and we may assume that, at least at one off-diagonal position, the entries of the Q_n are (positively) unbounded. When $d = 2$, this suffices to show that M is a singular idempotent. Already for $d = 3$, the situation becomes more complex, since one can have a limiting M with $\det(M) = 0$ that is not an idempotent, by considering

$$\exp \begin{pmatrix} -a-b & a & b \\ c & -c-d & d \\ e & f & -e-f \end{pmatrix}$$

for $a, b, \ldots, f \geqslant 0$. Then, fixing $b, \ldots, f$ at generic values and letting $a \to \infty$ produces such examples, and similarly for various other choices. When $d = 4$, one can have mixtures in block matrix form, such as

$$M = \begin{pmatrix} 1 & 0 \\ 1-\alpha & \alpha \end{pmatrix} \oplus \begin{pmatrix} \beta & 1-\beta \\ \beta & 1-\beta \end{pmatrix}$$

for $\alpha \in (0, 1)$ and $\beta \in [0, 1]$, which is singular but not idempotent.

It seems worthwhile to characterise the boundary more completely, for instance by relating semigroups with reducible generators to properties of the boundary, which we leave as an open problem at this point. It is clear, though, that Markov idempotents and their connection with blockwise equal-input matrices will be important here. More generally, a simplified

systematic treatment of the embeddability problem for $d \leqslant 4$ would be helpful in view of the applications in phylogeny; see [2] for recent progress in this direction. Finally, even some of the elementary questions become much harder in the situation of countable states, where many new phenomena occur; see [6] and references therein for some recent results.

Acknowledgements

Funding Information

Competing Interests

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

References

[1] BAAKE, M. AND SUMNER, J. (2020). Notes on Markov embedding. *Linear Algebra Appl.* **594**, 262–299.

[2] CASANELLAS, M., FERNÁNDEZ-SÁNCHEZ, J. AND ROCA-LACOSTENA, J. (2020). The embedding problem for Markov matrices. Preprint. Available at https://arxiv.org/abs/2005.00818.

[3] CULVER, W. J. (1966). On the existence and uniqueness of the real logarithm of a matrix. *Proc. Amer. Math. Soc.* **17**, 1146–1151.

[4] DALEY, D. J. (1968). Stochastically monotone Markov chains. *Z. Wahrscheinlichkeitsth.* **10**, 305–317.

[5] DAVIES, E. B. (2010). Embeddable Markov matrices. *Electron. J. Prob.* **15**, 1474–1486.

[6] EISNER, T. AND RADL, A. (2021). Embeddability of real and positive operators. To appear in *Linear Multilinear Algebra*. Available at https://arxiv.org/abs/2003.08186.

[7] ELFVING, G. (1937). Zur Theorie der Markoffschen Ketten. *Acta Soc. Sci. Fennicae* **A2**, 1–17.

[8] FELLER, W. (1971). *An Introduction to Probability Theory and Its Applications*, Vol. **II**, 2nd edn. John Wiley, New York.

[9] FERNÁNDEZ-SÁNCHEZ, J., SUMNER, J. G., JARVIS, P. D. AND WOODHAMS, M. D. (2015). Lie Markov models with purine/pyrimidine symmetry. *J. Math. Biol.* **70**, 855–891.

[10] GANTMACHER, F. R. (1986). *Matrizentheorie*. Springer, Berlin.

[11] GÖNDÖCS, F., MICHALETZKY, G., MÓRI, T. F. AND SZÉKELY, G. J. (1985). A characterization of infinitely divisible Markov chains with finite state space. *Ann. Univ. Sci. Budapest R. Eötvös Nom.* **27**, 137–141.

[12] GUERRY, M.-A. (2019). Sufficient embedding conditions for three-state discrete-time Markov chains with real eigenvalues. *Linear Multilinear Algebra* **67**, 106–120.

[13] HIGHAM, N. J. (2002). *Accuracy and Stability of Numerical Algorithms*, 2nd edn. Society for Industrial and Applied Mathematics, Philadelphia.

[14] HIGHAM, N. J. (2008). *Functions of Matrices: Theory and Computation*. Society for Industrial and Applied Mathematics, Philadelphia.

[15] HIGHAM, N. J. AND LIN, L. (2011). On pth roots of stochastic matrices. *Linear Algebra Appl.* **435**, 448–463.

[16] HÖGNÄS, G. AND MUKHERJEA, A. (2011). *Probability Measures on Semigroups: Convolution Products, Random Walks and Random Matrices*, 2nd edn. Springer, New York.

[17] HORN, R. A. AND JOHNSON, C. R. (2013). *Matrix Analysis*, 2nd edn. Cambridge University Press.

[18] KEILSON, J. AND KESTER, A. (1977). Monotone Markov matrices and monotone Markov processes. *Stoch. Process. Appl.* **5**, 231–241.

[19] KIJIMA, M. (1997). *Markov Processes for Stochastic Modeling*. Springer, Boston.

[20] KINGMAN, J. F. C. (1962). The imbedding problem for finite Markov chains. *Z. Wahrscheinlichkeitsth.* **1**, 14–24.

[21] LANG, S. (1993). *Algebra*, 3rd edn. Addison-Wesley, Reading, MA.

[22] LINDQVIST, B. H. (1987). Monotone and associated Markov chains, with applications to reliability theory. *J. Appl. Prob.* **24**, 679–695.

[23] LUTHER, U. AND ROST, K. (2004). Matrix exponentials and inversion of confluent Vandermonde matrices. *Electron. Trans. Numer. Anal.* **18**, 91–100.

[24] MARCUS, M. AND MINC, H. (1992). *A Survey of Matrix Theory and Matrix Inequalities*. Dover, New York.

[25] NORRIS, J. R. (2005). *Markov Chains*. Cambridge University Press.

[26] SLOANE, N. J. A. (1964). *The On-Line Encyclopedia of Integer Sequences*. Available at https://oeis.org.

[27] STEEL, M. (2016). *Phylogeny—Discrete and Random Processes in Evolution*. Society for Industrial and Applied Mathematics, Philadelphia.

[28] SUMNER, J. (2017). Multiplicatively closed Markov models must form Lie algebras. *ANZIAM J.* **59**, 240–246.

[29] SUMNER, J. G., FERNÁNDEZ-SÁNCHEZ, J. AND JARVIS, P. D. (2012). Lie Markov models. *J. Theoret. Biol.* **298**, 16–31.

[30] WEBSTER, R. (1994). *Convexity*. Oxford University Press.

[31] YANG, Z. AND RANALA, B. (2012). Molecular phylogenetics: principles and practice. *Nature Rev. Genet.* **13**, 303–314.