On the gap between deterministic and probabilistic joint spectral radii for discrete-time linear systems*

Yacine Chitour[†], Guilherme Mazanti[†‡], Mario Sigalotti[‡§]

November 16, 2021

Abstract

Given a discrete-time linear switched system $\Sigma(\mathcal{A})$ associated with a finite set $\mathcal{A}$ of matrices, we consider the measures of its asymptotic behavior given by, on the one hand, its deterministic joint spectral radius $\rho_{\mathrm{d}}(\mathcal{A})$ and, on the other hand, its probabilistic joint spectral radii $\rho_{\mathrm{p}}(\nu, P, \mathcal{A})$ for Markov random switching signals with transition matrix P and a corresponding invariant probability ν. Note that $\rho_{\mathrm{d}}(\mathcal{A})$ is larger than or equal to $\rho_{\mathrm{p}}(\nu, P, \mathcal{A})$ for every pair (ν, P). In this paper, we investigate the cases of equality of $\rho_{\mathrm{d}}(\mathcal{A})$ with either a single $\rho_{\mathrm{p}}(\nu, P, \mathcal{A})$ or with the supremum of $\rho_{\mathrm{p}}(\nu, P, \mathcal{A})$ over (ν, P) and we aim at characterizing the sets $\mathcal{A}$ for which such equalities may occur.

Keywords. Linear switched systems, discrete time, joint spectral radius, Markov process.
2020 Mathematics Subject Classification. 93C30, 93C55, 37H15.

This paper was first published in Linear Algebra and its Applications*, 613:24–45, 2021. With respect to the published version, this version provides an additional remark (Remark 2.10) and a more precise proof of Lemma 3.2. All modifications with respect to the published version are given in blue. The authors are very grateful to Matteo Della Rossa for pointing out the imprecisions in the previous version of the proof of Lemma 3.2.*

1 Introduction

In this paper, we consider discrete-time switched linear systems of the form

$$\Sigma(\mathcal{A}): \qquad x_{k+1} = A_{\sigma(k)} x_k, \qquad \sigma \in \mathfrak{S}, \quad k \in \mathbb{N}, \tag{1.1}$$

*This research was partially supported by the iCODE Institute, research project of the IDEX Paris-Saclay. The second author was partially supported by the public grant number ANR-10-CAMP-0151-02-FMJH as part of the "Programme des Investissements d'Avenir".

†Université Paris-Saclay, CNRS, CentraleSupélec, Laboratoire des signaux et systèmes, 91190, Gif-sur-Yvette, France.

‡Inria, France.

§Sorbonne Université, Université de Paris, CNRS, Inria, Laboratoire Jacques-Louis Lions, 75005 Paris, France.

where d and N are positive integers, $x_k \in \mathbb{R}^d$, $\mathfrak{S}$ is the set of the set of all maps $\sigma : \mathbb{N} \to \{1, \dots, N\}$, and $\mathcal{A} = (A_1, \dots, A_N)$ is an N-tuple of $d \times d$ matrices with real coefficients.

Switched systems model the behavior of a continuous variable x whose dynamics may change over time according to the value of a discrete variable σ. These models are useful for several applications, ranging from air traffic control, electronic circuits, and automotive engines to chemical processes and population models in biology. This wide field of applications, together with the interesting mathematical questions arising from their analysis, justify the extensive literature on switched systems, which have been studied from the point of view of both deterministic and random switching [6, 7, 22, 23, 28, 29]. A commonly used point of view on the switching signal σ, which we adopt in this paper, is to consider it as an uncertainty or perturbation acting on the system, the goal being thus to provide properties of the system independent of a particular choice of σ.

We are interested in describing the asymptotic behavior of $\Sigma(\mathcal{A})$. For a given $\sigma \in \mathfrak{S}$, the asymptotic behavior of the corresponding non-autonomous linear system is measured by the quantity $\rho(\sigma)$ defined by

$$\rho(\sigma) = \limsup_{n \to \infty} \|A_{\sigma(n)} \cdots A_{\sigma(1)}\|^{1/n}. \tag{1.2}$$

Indeed, $\rho(\sigma) < 1$ if and only if all trajectories of the non-autonomous system $x_{k+1} = A_{\sigma(k)} x_k$ converge exponentially to the origin.

In order to capture the asymptotic behavior of $\Sigma(\mathcal{A})$, we must formulate some condition which is independent of the choice of $\sigma \in \mathfrak{S}$. There exist two main approaches to proceed. The first one is deterministic and consists in considering the *joint spectral radius* $\rho_{\mathrm{d}}(\mathcal{A})$ of $\mathcal{A}$, defined as the supremum of $\rho(\sigma)$ over all $\sigma \in \mathfrak{S}$. Since its introduction in [26] and after the seminal paper [13], it has been extensively studied in the computer science and control theory communities (see, e.g., the monograph [19]).

The other approach to handle the asymptotic behavior of $\Sigma(\mathcal{A})$ is probabilistic and amounts to considering a probability measure μ on $\mathfrak{S}$ and hence $\sigma \mapsto \rho(\sigma)$ as a random variable. We may then consider as a *probabilistic joint spectral radius* the expected value of $\rho(\sigma)$ with respect to the probability law μ, which we denote by $\rho_{\mathrm{p}}(\mu, \mathcal{A})$. There exists a vast literature devoted to the properties of products of random matrices, and we refer the reader to [1, 5, 8] for more details. A major result in this field has been obtained in [16] and provides general conditions on μ under which $\rho(\sigma) = \rho_{\mathrm{p}}(\mu, \mathcal{A})$ on a set of μ probability 1.

The interest in considering $\rho_{\mathrm{d}}(\mathcal{A})$ and $\rho_{\mathrm{p}}(\mu, \mathcal{A})$ comes from the stability analysis of (1.1). Indeed, $\rho_{\mathrm{d}}(\mathcal{A}) < 1$ if and only if (1.1) is uniformly exponentially stable [19], whereas, under the conditions of [16], $\rho_{\mathrm{p}}(\mu, \mathcal{A}) < 1$ if and only if μ-almost every trajectory of (1.1) converges exponentially to the origin.

In this paper, we aim at understanding the relations between the deterministic and the probabilistic approaches. The deterministic measure of stability $\rho_{\mathrm{d}}(\mathcal{A})$ characterizes the worst possible behavior over all $\sigma \in \mathfrak{S}$, while the probabilistic counterpart $\rho_{\mathrm{p}}(\mu, \mathcal{A})$ provides the average behavior for $\sigma \in \mathfrak{S}$ corresponding to the probability measure μ. As a consequence, the deterministic approach provides a more conservative estimate of the asymptotic behavior of the system than the probabilistic one, in the sense that

$$\rho_{\mathrm{p}}(\mu, \mathcal{A}) \leq \rho_{\mathrm{d}}(\mathcal{A}). \tag{1.3}$$

A natural question is then to understand under which conditions on $\mathcal{A}$ and μ the inequality in (1.3) is strict. Furthermore, for practical and modeling purposes, it is important to understand whether, given a family of probability measures $\{\mu_\ell\}_{\ell\in\mathfrak{I}}$, the strict inequality $\sup_{\ell\in\mathfrak{I}} \rho_{\mathrm{p}}(\mu_\ell,\mathcal{A}) < \rho_{\mathrm{d}}(\mathcal{A})$ holds true. Regarding the first question, it is known that there always exists a measure μ such that equality holds in (1.3) (see, for instance, [24], where such measures are referred to as *maximizing measures*). At such a level of generality, a handy characterization of maximizing measures cannot be expected. This is why we restrict our attention to the family $\mathfrak{M}$ of probability measures on $\mathfrak{S}$ obtained from discrete-time shift-invariant Markov chains and reformulate the previous two questions as follows: under which conditions on $\mathcal{A}$ do we have

(Q1) equality between $\rho_{\mathrm{p}}(\mu,\mathcal{A})$ and $\rho_{\mathrm{d}}(\mathcal{A})$ for a given $\mu\in\mathfrak{M}$?

(Q2) equality between $\sup_{\mu\in\mathfrak{M}} \rho_{\mathrm{p}}(\mu,\mathcal{A})$ and $\rho_{\mathrm{d}}(\mathcal{A})$?

Notice that the condition $\sup_{\mu\in\mathfrak{M}} \rho_{\mathrm{p}}(\mu,\mathcal{A}) < 1$ is related to the almost sure stability of the system uniformly with respect to the Markov process, a stability property first considered in [18] in the case of Markov chains with positive transition probabilities. Other stability notions have also been considered for (1.1), such as periodic stability, meaning stability for all *periodic* signals $\sigma\in\mathfrak{S}$, or mean square stability. Several works explore relations between these different notions, see, e.g., [6, 9, 12, 14, 15, 18]. In particular, [12] establishes a probabilistic version of the finiteness conjecture, i.e., if (1.1) is periodically stable, then $\rho_{\mathrm{p}}(\mu,\mathcal{A}) < 1$ for every $\mu\in\mathfrak{M}$.

Another interesting fact is that the quantities $\rho_{\mathrm{d}}(\mathcal{A})$ and of $\rho_{\mathrm{p}}(\mu,\mathcal{A})$ for $\mu\in\mathfrak{M}$ could be equivalently computed by replacing the norm $\|\cdot\|$ in (1.2) by the spectral radius. In the deterministic case, this result is known as the Berger–Wang formula or also as the Joint Spectral Radius Theorem [19], and it has been extended to the Markovian setting in [10, 20].

In order to describe the main results of our paper, let us identify a measure $\mu\in\mathfrak{M}$ with the pair (ν,P), where P is the transition matrix of the Markov chain corresponding to μ and ν is its (invariant) initial probability. In particular, we write $\rho_{\mathrm{p}}(\nu,P,\mathcal{A})$ for $\rho_{\mathrm{p}}(\mu,\mathcal{A})$. Our main result concerning (Q1) (see Theorem 3.1) establishes that a necessary and sufficient condition for equality is that $\rho_{\mathrm{d}}(\mathcal{A}) = \rho(A_{i_k}\cdots A_{i_1})^{1/k}$ for every $(i_1,\dots,i_k)$ that corresponds to a cycle in the directed weighted graph determined by P such that $\nu_{i_1} > 0$. The necessity follows from results provided in [24], whereas, for sufficiency, we consider first the particular case where $\mathcal{A}$ is irreducible and P is strongly connected (see Lemma 3.3). Irreducibility implies in particular the existence of a Barabanov norm for $\mathcal{A}$ (see Definition 2.1), which is an important tool in our proof. We then generalize the result to the case of reducible $\mathcal{A}$ (see Lemma 3.5) by a suitable block decomposition of the matrices in $\mathcal{A}$ and the fact that $\rho_{\mathrm{p}}(\nu,P,\mathcal{A})$ and $\rho_{\mathrm{d}}(\mathcal{A})$ can be read on the diagonal blocks of the decomposed matrix (cf. [17, 19]). Finally, the general case for P can be obtained by using a classical block decomposition of stochastic matrices.

The equivalence established in Theorem 3.1 can be further characterized in terms of simultaneous similarity of the matrices $\rho_{\mathrm{d}}(\mathcal{A})^{-1}A_i$, $i\in\{1,\dots,N\}$, to orthogonal matrices, under some additional assumptions on $\mathcal{A}$ and P (Proposition 3.9). The latter characterization is based on the description of matrix semigroups with constant spectral radius from [25].

Our next main result, Theorem 3.13, concerns (Q2) and states that equality is equivalent to the existence of a family of pairwise distinct indices $i_1,\dots,i_k \in \{1,\dots,N\}$ such that $\rho_{\mathrm{d}}(\mathcal{A}) = \rho(A_{i_1}\cdots A_{i_k})^{1/k}$. This corresponds to the case where the worst behavior of the system is attained by a periodic σ with no repetition of indices on a period. This property is reminiscent of the *finiteness property*, except for the fact that, in the finiteness property, repetition of indices is allowed. We recall that the finiteness property is known to hold only for a proper subclass of N-tuples $\mathcal{A}$ [3,4], contrarily to what had been earlier conjectured [21]. By applying a standard lifting argument of Markov chains of higher order to Markov chains of order one, we generalize the equivalence stated in Theorem 3.13 by providing the following characterization of the finiteness property: a N-tuple $\mathcal{A}$ satisfies the finiteness property if and only if there exist $m \geq 1$ and a Markov chain of order m whose corresponding probabilistic Lyapunov exponent is equal to $\rho_{\mathrm{d}}(\mathcal{A})$ (see Corollary 4.4). This, in turns, is equivalent to say that the finiteness property holds if and only if the set of maximizing measures contains the measure induced by some Markov chain of arbitrary order.

Acknowledgements. The authors are indebted with D. Chafaï for helpful discussions. They are also grateful to the anonymous reviewers of a preceding version of the manuscript for providing helpful comments and pointing out relevant literature.

2 Definitions, notations, and basic facts

Throughout the paper, d and N belong to $\mathbb{N}$, which is used to denote the set of positive integers. If a and b are positive integers, $[\![a,b]\!]$ denotes the set of integers j such that $a \leq j \leq b$. For $x \in \mathbb{R}$, $\lceil x \rceil$ denotes the smallest integer greater than or equal to x, and we extend this notation componentwise to vectors and matrices. We use $\|\cdot\|$ to denote a norm in $\mathbb{R}^d$ as well as the corresponding induced norm on the space $\mathcal{M}_d(\mathbb{R})$ of $d \times d$ matrices with real coefficients. We only consider in this paper norms in $\mathcal{M}_d(\mathbb{R})$ obtained as induced norms from $\mathbb{R}^d$. An N-tuple $\mathcal{A} = (A_1,\dots,A_N) \in \mathcal{M}_d(\mathbb{R})^N$ is said to be *irreducible* if the only subspaces of $\mathbb{R}^d$ invariant under all the matrices $A_1,\dots,A_N$ are $\{0\}$ and $\mathbb{R}^d$.

2.1 Deterministic joint spectral radius

Let $\Sigma(\mathcal{A})$ be the discrete-time switched system defined in (1.1). The *deterministic joint spectral radius* $\rho_{\mathrm{d}}(\mathcal{A})$ of $\Sigma(\mathcal{A})$, introduced in [26], is defined by

$$\rho_{\mathrm{d}}(\mathcal{A}) = \limsup_{n\to\infty} \max_{(i_1,\dots,i_n)\in[\![1,N]\!]^n} \|A_{i_n}\cdots A_{i_1}\|^{1/n}. \tag{2.1}$$

Since all norms in $\mathcal{M}_d(\mathbb{R})$ induced by norms in $\mathbb{R}^d$ are submultiplicative and equivalent to each other, it immediately follows that $\rho_{\mathrm{d}}(\mathcal{A})$ does not depend on a specific choice of such a norm and that

$$\rho_{\mathrm{d}}(\mathcal{A}) = \lim_{n\to\infty} \max_{(i_1,\dots,i_n)\in[\![1,N]\!]^n} \|A_{i_n}\cdots A_{i_1}\|^{1/n} = \inf_{n\in\mathbb{N}} \max_{(i_1,\dots,i_n)\in[\![1,N]\!]^n} \|A_{i_n}\cdots A_{i_1}\|^{1/n}.$$

Notice that, for every $n \in \mathbb{N}$ and $(i_1,\dots,i_n) \in [\![1,N]\!]^n$, we have

$$\rho(A_{i_n}\cdots A_{i_1})^{1/n} \leq \rho_{\mathrm{d}}(\mathcal{A}), \tag{2.2}$$

where we use the definition of $\rho_{\mathrm{d}}(\mathcal{A})$ and the fact that $\rho(M) = \rho(M^k)^{1/k} \leq \|M^k\|^{1/k}$ for every square matrix M and $k \in \mathbb{N}$.

Definition 2.1 (Barabanov norm)**.** Let $\mathcal{A} = (A_1, \dots, A_N)$ be an N-tuple of $d \times d$ matrices with real coefficients. A norm $\|\cdot\|_{\mathrm{B}}$ on $\mathbb{R}^d$ is said to be a *Barabanov norm* for $\mathcal{A}$ if the following two conditions hold.

(a) For every $i \in [\![1, N]\!]$, $\|A_i\|_{\mathrm{B}} \leq \rho_{\mathrm{d}}(\mathcal{A})$.

(b) For every $x \in \mathbb{R}^d$ and $k \in \mathbb{N}$, there exists $\sigma \in \mathfrak{S}$ such that $\|A_{\sigma(k)} \cdots A_{\sigma(1)} x\|_{\mathrm{B}} = \rho_{\mathrm{d}}(\mathcal{A})^k \|x\|_{\mathrm{B}}$.

The following basic result on Barabonov norms was proved in [2].

Proposition 2.2. *Let $\mathcal{A}$ be an N-tuple of $d \times d$ matrices with real coefficients. If $\mathcal{A}$ is irreducible, then it admits a Barabanov norm.*

2.2 Probabilistic joint spectral radius

We now provide a probabilistic counterpart to $\rho_{\mathrm{d}}(\mathcal{A})$. For that purpose, we collect some basic notions concerning transition matrices of Markov chains.

Definition 2.3. Let $P = (p_{ij})_{1 \leq i,j \leq N}$ be an $N \times N$ matrix with nonnegative coefficients.

(a) P is said to be *stochastic* if, for every $i \in [\![1, N]\!]$, $\sum_{j=1}^N p_{ij} = 1$.

(b) P is said to be *strongly connected* if it is not similar via a permutation to an upper block triangular matrix.

(c) For $k \in \mathbb{N}$ and $i_1, \dots, i_k \in [\![1, N]\!]$, we say that $(i_1, \dots, i_k)$ is a *P-word* if $p_{i_1 i_2} p_{i_2 i_3} \cdots p_{i_{k-1} i_k} > 0$. The integer k is called the *length* of the P-word $(i_1, \dots, i_k)$. We say that $(i_1, \dots, i_k)$ is a *P-cycle* if $p_{i_1 i_2} p_{i_2 i_3} \cdots p_{i_{k-1} i_k} p_{i_k i_1} > 0$. The index i_1 is called the *starting index* of the P-cycle $(i_1, \dots, i_k)$.

(d) Let ν be a vector in $\mathbb{R}^N$ with nonnegative coefficients. We say that $(i_1, \dots, i_k)$ is a *(ν, P)-word* (respectively, *(ν, P)-cycle*) if it is a P-word (respectively, P-cycle) and $\nu_{i_1} > 0$.

(e) If P is stochastic, a row vector $\nu = (\nu_1, \dots, \nu_N) \in \mathbb{R}^N$ is said to be an *invariant probability* for P if $\nu_i \geq 0$ for every $i \in [\![1, N]\!]$, $\sum_{i=1}^N \nu_i = 1$, and $\nu = \nu P$.

Remark 2.4. In the context of discrete-time Markov chains in a finite state space with N states, the *transition matrix* is the stochastic matrix $P = (p_{ij})_{1 \leq i,j \leq N}$ where p_{ij} represents the probability to switch from the state i to the state j. Notice that P is strongly connected if and only if its associated oriented graph G_P is strongly connected. In the stochastic processes literature, the strong connectedness of P is more often referred to as *irreducibility*. We choose to stick with the former to avoid ambiguities with the homonymous notion for N-tuples of matrices. Notice also that the notions of strong connectedness, P-cycles, and P-words only depend on the adjacency matrix $\lceil P \rceil$ of the graph G_P, while (ν, P)-cycles and (ν, P)-words depend on $\lceil P \rceil$ and $\lceil \nu \rceil$.

Remark 2.5. Recall that, by the Perron–Frobenius Theorem, a stochastic matrix P always admits an invariant probability, which is unique and has positive entries if P is strongly connected. In the latter case, the definitions of P-word and (ν,P)-word coincide, as well as those of P-cycle and (ν,P)-cycle.

We have the following classical decomposition result for stochastic matrices [27, §§1.2 and 4.2].

Proposition 2.6. *Let $P \in \mathcal{M}_N(\mathbb{R})$ be a stochastic matrix. Then, up to a permutation in the set of indices $[\![1,N]\!]$, P is given by*

$$P = \begin{pmatrix} P_1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & P_2 & 0 & \cdots & 0 & 0 \\ \vdots & 0 & \ddots & \ddots & \vdots & \vdots \\ 0 & \vdots & \ddots & \ddots & 0 & 0 \\ 0 & 0 & \cdots & 0 & P_R & 0 \\ * & * & \cdots & * & * & Q \end{pmatrix}, \tag{2.3}$$

where $\rho(Q) < 1$ and, for $i \in [\![1,R]\!]$, $P_i \in \mathcal{M}_{n_i}(\mathbb{R})$ is a stochastic and strongly connected matrix.

Moreover, for $i \in [\![1,R]\!]$, let $\nu^{[i]}$ be the unique invariant probability for P_i and denote by the same symbol its canonical extension as a vector in $\mathbb{R}^N$ according to the decomposition (2.3)*. Then every invariant probability $\nu \in \mathbb{R}^N$ can be uniquely decomposed as*

$$\nu = \sum_{i=1}^{R} \alpha_i \nu^{[i]}, \tag{2.4}$$

where $\alpha_1,\dots,\alpha_R \in [0,1]$ and $\sum_{i=1}^R \alpha_i = 1$.

The next lemma, useful in the proof of some of our results, uses the previous decomposition to obtain that any (ν,P)-cycle has all its indices corresponding to a same diagonal block P_i in (2.3).

Lemma 2.7. *Let $P \in \mathcal{M}_N(\mathbb{R})$ be a stochastic matrix decomposed according to Proposition 2.6. For $i \in [\![1,R]\!]$, let*

$$\mathcal{I}_i = \left[\!\!\left[1 + \sum_{j=1}^{i-1} n_j, \sum_{j=1}^{i} n_j \right]\!\!\right],$$

i.e., $\mathcal{I}_i$ is the set of indices corresponding to the diagonal block P_i in (2.3)*. Let ν be an invariant probability for P. Then, for every (ν,P)-cycle $(i_1,\dots,i_n)$, there exists $j \in [\![1,R]\!]$ such that $i_1,\dots,i_n$ are in $\mathcal{I}_j$.*

Proof. Notice that, by (2.4), $\nu_i = 0$ if $i \notin \bigcup_{j\in[\![1,R]\!]} \mathcal{I}_j$. Hence, since $\nu_{i_1} > 0$, there exists $j \in [\![1,R]\!]$ such that $i_1 \in \mathcal{I}_j$. Since $p_{i_1 i_2} > 0$, it follows by the block decomposition (2.3) that $i_2 \in \mathcal{I}_j$. The conclusion follows by an immediate inductive argument. $\square$

We also introduce the following notation.

Definition 2.8. Let P be a stochastic matrix and $\mathcal{A} = (A_1,\dots,A_N)$ be an N-tuple of $d \times d$ matrices with real coefficients.

(a) For every P-word $(i_1,\dots,i_k)$, we use $A(i_1,\dots,i_k)$ to denote the matrix product $A_{i_k}\cdots A_{i_1}$.

(b) For every $s\in[\![1,N]\!]$, let $C(P,s)$ be the matrix semigroup made of all matrix products associated with P-cycles with starting index s, i.e.,

$$C(P,s)=\{A(i_1,\dots,i_k)\mid (i_1,\dots,i_k)\text{ is a }P\text{-cycle and }i_1=s\}.$$

We also set

$$C(P)=\bigcup_{s\in[\![1,N]\!]}C(P,s).$$

We finally provide the definition of the probabilistic counterpart of $\rho_{\mathrm{d}}(\mathcal{A})$ for $\Sigma(\mathcal{A})$. Let $P=(p_{ij})_{1\le i,j\le N}$ be a stochastic matrix, $\nu=(\nu_1,\dots,\nu_N)$ be an invariant probability for P, and $\mathcal{A}=(A_1,\dots,A_N)$ an N-tuple in $\mathcal{M}_d(\mathbb{R})$. The *probabilistic joint spectral radius* $\rho_{\mathrm{p}}(\nu,P,\mathcal{A})$ is defined as

$$\rho_{\mathrm{p}}(\nu,P,\mathcal{A})=\limsup_{n\to\infty}\mathbb{E}_{(\nu,P)}\left[\|A_{i_n}\cdots A_{i_1}\|^{1/n}\right],\tag{2.5}$$

where

$$\mathbb{E}_{(\nu,P)}\left[\|A_{i_n}\cdots A_{i_1}\|^{1/n}\right]=\sum_{(i_1,\dots,i_n)\in[\![1,N]\!]^n}\nu_{i_1}p_{i_1i_2}\cdots p_{i_{n-1}i_n}\|A_{i_n}\cdots A_{i_1}\|^{1/n}.\tag{2.6}$$

As in the deterministic case, $\rho_{\mathrm{p}}(\nu,P,\mathcal{A})$ does not depend on the specific choice of the norm $\|\cdot\|$ and we have

$$\rho_{\mathrm{p}}(\nu,P,\mathcal{A})=\lim_{n\to\infty}\mathbb{E}_{(\nu,P)}\left[\|A_{i_n}\cdots A_{i_1}\|^{1/n}\right]=\inf_{n\in\mathbb{N}}\mathbb{E}_{(\nu,P)}\left[\|A_{i_n}\cdots A_{i_1}\|^{1/n}\right].\tag{2.7}$$

Remark 2.9. The expectation in (2.5) is taken with respect to the random variable $(i_1,\dots,i_n)\in[\![1,N]\!]^n$. The definition of probabilistic joint spectral radius provided here is a particular instance of a more general and comprehensive formulation based on symbolic dynamics; see, for instance, [11, 12, 24]. Notice also that it follows from the definition of (ν,P)-word that the summation in (2.6) can be restricted to (ν,P)-words of length n.

Remark 2.10. When dealing with probabilistic switching phenomena in discrete time, several works, such as [1, 8, 11, 16, 17, 24], deal with the *probabilistic Lyapunov exponent* $\lambda_{\mathrm{p}}(\nu,P,\mathcal{A})$ defined by

$$\lambda_{\mathrm{p}}(\nu,P,\mathcal{A})=\limsup_{n\to\infty}\frac{1}{n}\mathbb{E}_{(\nu,P)}[\log\|A_{i_n}\cdots A_{i_1}\|].$$

Our choice to use $\rho_{\mathrm{p}}(\nu,P,\mathcal{A})$ instead is motivated by the fact that the main goal of our paper is to compare the probabilistic behavior of (1.1) with the worst deterministic behavior provided by the classical joint spectral radius $\rho_{\mathrm{d}}(\mathcal{A})$, whose definition in discrete-time (2.1) does not involve taking the logarithm of the norm of the matrix product. Working with $\rho_{\mathrm{p}}(\nu,P,\mathcal{A})$ also has the additional advantage of being able to handle the case $\rho_{\mathrm{p}}(\nu,P,\mathcal{A})=0$ without dealing with the singularity of the logarithm at 0.

Clearly, by Jensen's inequality, we have $e^{\lambda_{\mathrm{p}}(\nu,P,\mathcal{A})} \leq \rho_{\mathrm{p}}(\nu,P,\mathcal{A})$, but this inequality may be strict in some cases. Indeed, for $d = 1$, $N = 2$, $P = \mathrm{Id}_2$, $\nu = (\frac{1}{2},\frac{1}{2})$, and $\mathcal{A} = (A_1,A_2) \in \mathcal{M}_1(\mathbb{R})^2 \simeq \mathbb{R}^2$, we easily compute that $\rho_{\mathrm{p}}(\nu,P,\mathcal{A}) = \frac{1}{2}(A_1+A_2)$ and $e^{\lambda_{\mathrm{p}}}(\nu,P,\mathcal{A}) = \sqrt{A_1A_2}$.

We do have equality between $e^{\lambda_{\mathrm{p}}(\nu,P,\mathcal{A})}$ and $\rho_{\mathrm{p}}(\nu,P,\mathcal{A})$, however, under the assumption that (ν,P) defines an *ergodic* Markov chain, i.e., $\nu = \nu^{[i]}$ for some $i \in [\![1,R]\!]$ in the decompositions (2.3) and (2.4) in Proposition 2.6. Indeed, in this case, the main result of [16] implies that

$$\lambda_{\mathrm{p}}(\nu,P,\mathcal{A}) = \lim_{n\to\infty} \frac{1}{n} \log\left\|A_{\sigma(n)}\cdots A_{\sigma(1)}\right\| \qquad \text{for } \mathbb{P}_{(\nu,P)}\text{-almost every } \sigma \in \mathfrak{S},$$

where $\mathbb{P}_{(\nu,P)}$ denotes the probability measure on $\mathfrak{S}$ associated canonically with the transition matrix P and the invariant probability ν. Using this fact, one deduces that $e^{\lambda_{\mathrm{p}}(\nu,P,\mathcal{A})} = \rho_{\mathrm{p}}(\nu,P,\mathcal{A})$, and, in addition, we also have the equality

$$\rho_{\mathrm{p}}(\nu,P,\mathcal{A}) = \lim_{n\to\infty} \left\|A_{\sigma(n)}\cdots A_{\sigma(1)}\right\|^{1/n} \qquad \text{for } \mathbb{P}_{(\nu,P)}\text{-almost every } \sigma \in \mathfrak{S}. \tag{2.8}$$

Notice that, in particular, (ν,P) is ergodic when P is strongly connected and ν is its unique invariant measure.

Remark 2.11. The deterministic joint spectral radius $\rho_{\mathrm{d}}(\mathcal{A})$ provides the worst asymptotic behavior of $\Sigma(\mathcal{A})$ with respect to $\sigma \in \mathfrak{S}$. By introducing the probability measure $\mathbb{P}_{(\nu,P)}$ on $\mathfrak{S}$ associated canonically with the transition matrix P and the invariant probability ν, the quantity $\rho_{\mathrm{p}}(\nu,P,\mathcal{A})$ defined in (2.5) can be interpreted as an asymptotic behavior averaged by $\mathbb{P}_{(\nu,P)}$. When (ν,P) is ergodic, thanks to (2.8), we have the stronger interpretation of $\rho_{\mathrm{p}}(\nu,P,\mathcal{A})$ as the $\mathbb{P}_{(\nu,P)}$-almost sure asymptotic behavior of $\Sigma(\mathcal{A})$.

It is immediate to see that, for every $(\nu,P,\mathcal{A})$ as above, we have $\rho_{\mathrm{p}}(\nu,P,\mathcal{A}) \leq \rho_{\mathrm{d}}(\mathcal{A})$, and then

$$\rho_{\mathrm{p}}(\nu,P,\mathcal{A}) \leq \sup_{\nu'} \rho_{\mathrm{p}}(\nu',P,\mathcal{A}) \leq \sup_{(\nu',P')} \rho_{\mathrm{p}}(\nu',P',\mathcal{A}) \leq \rho_{\mathrm{d}}(\mathcal{A}), \tag{2.9}$$

where the first supremum is taken over all invariant probabilities ν' for P and the second one over the pairs (ν',P') made of an $N \times N$ stochastic matrix P' and an invariant probability ν' for P'. We find it useful to introduce the notation

$$\rho_{\mathrm{p}}(P,\mathcal{A}) = \sup_{\nu'} \rho_{\mathrm{p}}(\nu',P,\mathcal{A}), \qquad \rho_{\mathrm{p}}(\mathcal{A}) = \sup_{(\nu',P')} \rho_{\mathrm{p}}(\nu',P',\mathcal{A}). \tag{2.10}$$

Remark 2.12. It follows from (2.7) that $(\nu',P') \mapsto \rho_{\mathrm{p}}(\nu',P',\mathcal{A})$ is upper semicontinuous. Moreover, the set of pairs (ν',P') consisting of an $N \times N$ stochastic matrix P' and an invariant probability ν' for P' is compact. As a consequence, the suprema in (2.10) can be replaced by maxima.

3 Equality between deterministic and probabilistic joint spectral radii

3.1 Equality between $\rho_{\mathrm{d}}(\mathcal{A})$ and $\rho_{\mathrm{p}}(\nu,P,\mathcal{A})$

The goal of this section is to prove the following result characterizing equality between $\rho_{\mathrm{d}}(\mathcal{A})$ and $\rho_{\mathrm{p}}(\nu,P,\mathcal{A})$.

Theorem 3.1. *Let $P \in \mathcal{M}_N(\mathbb{R})$ be a stochastic matrix, $\nu \in \mathbb{R}^N$ be an invariant probability measure for P, and $\mathcal{A} = (A_1,\dots,A_N) \in \mathcal{M}_d(\mathbb{R})^N$. Then the following statements are equivalent:*

(a) $\rho_{\mathrm{d}}(\mathcal{A}) = \rho_{\mathrm{p}}(\nu,P,\mathcal{A})$.

(b) $\rho(A_{i_k}\cdots A_{i_1})^{1/k} = \rho_{\mathrm{d}}(\mathcal{A})$ *for every (ν,P)-cycle $(i_1,\dots,i_k)$.*

The fact that (a) implies (b) follows from the results in [24], as detailed in the following lemma.

Lemma 3.2. *Let $P \in \mathcal{M}_N(\mathbb{R})$ be a stochastic matrix, $\nu \in \mathbb{R}^N$ be an invariant probability measure for P, and $\mathcal{A} = (A_1,\dots,A_N) \in \mathcal{M}_d(\mathbb{R})^N$. If $\rho_{\mathrm{d}}(\mathcal{A}) = \rho_{\mathrm{p}}(\nu,P,\mathcal{A})$, then $\rho(A_{i_k}\cdots A_{i_1})^{1/k} = \rho_{\mathrm{d}}(\mathcal{A})$ for every (ν,P)-cycle $(i_1,\dots,i_k)$.*

Proof. If $\rho_{\mathrm{d}}(\mathcal{A}) = 0$, the result follows trivially from (2.2). We then assume $\rho_{\mathrm{d}}(\mathcal{A}) > 0$, we decompose P and ν according to Proposition 2.6, and we use in the sequel the same notations as in its statement. We also let $\mathcal{I}_1,\dots,\mathcal{I}_R$ be defined as in the statement of Lemma 2.7. Thanks to (2.4), (2.5), and (2.6), we have

$$\rho_{\mathrm{p}}(\nu,P,\mathcal{A}) = \sum_{j=1}^{R} \alpha_j \rho_{\mathrm{p}}(\nu^{[j]},P,\mathcal{A}). \tag{3.1}$$

By (2.9), we have $\rho_{\mathrm{p}}(\nu^{[j]},P,\mathcal{A}) \leq \rho_{\mathrm{d}}(\mathcal{A})$ for every $j \in [\![1,R]\!]$ and, since $\alpha_j \in [0,1]$ for every $j \in [\![1,R]\!]$ and $\sum_{j=1}^{R} \alpha_j = 1$, we deduce from (3.1) and the equality $\rho_{\mathrm{d}}(\mathcal{A}) = \rho_{\mathrm{p}}(\nu,P,\mathcal{A})$ that $\rho_{\mathrm{p}}(\nu^{[j]},P,\mathcal{A}) = \rho_{\mathrm{d}}(\mathcal{A})$ for every $j \in [\![1,R]\!]$ such that $\alpha_j > 0$.

Let $(i_1,\dots,i_k)$ be a (ν,P)-cycle and note that, by Lemma 2.7, there exists $r \in [\![1,R]\!]$ such that $i_1,\dots,i_k$ are in $\mathcal{I}_r$, and thus, in particular, $(i_1,\dots,i_k)$ is also a $(\nu^{[r]},P)$-cycle. Moreover, such a r necessarily satisfies $\alpha_r > 0$. Consider the k-periodic switching signal $\sigma \in \mathfrak{S}$ corresponding to $(i_1,\dots,i_k)$, defined by $\sigma(j+\ell k) = i_j$ for all integers $j \in [\![1,k]\!]$ and $\ell \geq 0$. Endow $\mathfrak{S}$ with its usual product topology and denote by $\mathbb{P}_r$ the Borel probability measure on $\mathfrak{S}$ corresponding to the Markov chain defined by $(\nu^{[r]},P)$. Note that, since $(i_1,\dots,i_k)$ is a $(\nu^{[r]},P)$-cycle, for every $n \in \mathbb{N}$ the set $\{\widetilde{\sigma} \in \mathfrak{S} \mid \widetilde{\sigma}(i) = \sigma(i) \text{ for every } i \in [\![1,n]\!]\}$ has positive $\mathbb{P}_r$ measure, and thus σ is in the support of $\mathbb{P}_r$. Moreover, using also Remark 2.10, we have $\rho_{\mathrm{d}}(\mathcal{A}) = \rho_{\mathrm{p}}(\nu^{[r]},P,\mathcal{A}) = e^{\lambda_{\mathrm{p}}(\nu^{[r]},P,\mathcal{A})}$, and then $\mathbb{P}_r$ is a maximizing measure of $\mathcal{A}$ in the sense of [24], where the set of maximizing measures of $\mathcal{A}$ is defined as the set of all Borel probability measures on $\mathfrak{S}$ invariant under the usual time shift and such that the corresponding probabilistic Lyapunov exponent coincides with $\log\rho_{\mathrm{d}}(\mathcal{A})$. Hence σ belongs to the Mather set of $\mathcal{A}$ (see [24, Theorem 2.3], where the Mather set of $\mathcal{A}$ is defined as the union of the supports of all maximizing measures of $\mathcal{A}$), and thus, by [24, Theorem 2.3(3)], we get

$$\limsup_{n\to\infty} \rho_{\mathrm{d}}(\mathcal{A})^{-n}\rho(A_{\sigma(n)}\cdots A_{\sigma(1)}) = 1. \tag{3.2}$$

Set $M = \rho_{\mathrm{d}}(\mathcal{A})^{-k}A_{i_k}\cdots A_{i_1}$. By (2.2), we have that $\rho(M) \leq 1$. For every $n \geq 1$, there exist integers $\ell \geq 0$ and $j \in [\![0, k-1]\!]$ such that $n = j + \ell k$. Since σ is k-periodic, we have that

$$\rho_{\mathrm{d}}(\mathcal{A})^{-n}\rho(A_{\sigma(n)}\cdots A_{\sigma(1)}) = \rho_{\mathrm{d}}(\mathcal{A})^{-j}\rho(A_{i_j}\cdots A_{i_1}M^{\ell}).$$

If $\rho(M) < 1$, then the right-hand side of the above inequality tends to 0 as $\ell \to \infty$, contradicting (3.2). Hence, we have necessarily $\rho(M) = 1$. □

The proof that (b) implies (a) in Theorem 3.1 is decomposed in three steps. We first establish the result under the extra assumptions that $\mathcal{A}$ is irreducible and P is strongly connected (Lemma 3.3). We then obtain the conclusion under the sole additional assumption that P is strongly connected (Lemma 3.5). Finally, we consider the general case in the third step.

Lemma 3.3. *Let $P \in \mathcal{M}_d(\mathbb{R})$ be a stochastic strongly connected matrix, $\mathcal{A} = (A_1, \ldots, A_N) \in \mathcal{M}_d(\mathbb{R})^N$ be irreducible, and $\|\cdot\|_{\mathrm{B}}$ be a Barabanov norm for $\mathcal{A}$. Then the following statements are equivalent:*

(a) $\rho_{\mathrm{d}}(\mathcal{A}) = \rho_{\mathrm{p}}(P, \mathcal{A})$.

(b) $\rho(A_{i_k}\cdots A_{i_1})^{1/k} = \rho_{\mathrm{d}}(\mathcal{A})$ *for every P-cycle* $(i_1, \ldots, i_k)$.

(c) $\|A_{i_k}\cdots A_{i_1}\|_{\mathrm{B}}^{1/k} = \rho_{\mathrm{d}}(\mathcal{A})$ *for every P-word* $(i_1, \ldots, i_k)$.

Proof. The fact that (a) implies (b) is a particular case of Lemma 3.2. Moreover, it is immediate that (c) implies (a) thanks to (2.5), (2.6), and Remark 2.9. We are then left to prove that (b) implies (c).

Assume that (b) holds. Fix a P-word $(i_1, \ldots, i_k)$. Since P is strongly connected, there exist $r \in \mathbb{N}$ and $i_{k+1}, \ldots, i_r \in [\![1, N]\!]$ (obtained by connecting i_k to i_1) such that $(i_1, \ldots, i_r)$ is a P-cycle. Then, by (b),

$$\rho(A_{i_r}\cdots A_{i_1}) = \rho_{\mathrm{d}}(\mathcal{A})^r.$$

Since the spectral radius is a lower bound for any induced norm of a matrix, we obtain that

$$\rho_{\mathrm{d}}(\mathcal{A})^r \leq \|A_{i_r}\cdots A_{i_1}\|_{\mathrm{B}} \leq \left\|A_{i_r}\cdots A_{i_{k+1}}\right\|_{\mathrm{B}}\|A_{i_k}\cdots A_{i_1}\|_{\mathrm{B}}.$$

Using the fact that $\|\cdot\|_{\mathrm{B}}$ is a Barabanov norm, we also have that

$$\left\|A_{i_r}\cdots A_{i_{k+1}}\right\|_{\mathrm{B}}\|A_{i_k}\cdots A_{i_1}\|_{\mathrm{B}} \leq \rho_{\mathrm{d}}(\mathcal{A})^{r-k}\rho_{\mathrm{d}}(\mathcal{A})^k = \rho_{\mathrm{d}}(\mathcal{A})^r.$$

By combining the previous inequalities, it follows that $\|A_{i_k}\cdots A_{i_1}\|_{\mathrm{B}} = \rho_{\mathrm{d}}(\mathcal{A})^k$. □

Remark 3.4. The proof of Lemma 3.3 only uses that $\|\cdot\|_{\mathrm{B}}$ is an extremal norm, i.e., it satisfies (a) in Definition 2.1. The irreducibility assumption on $\mathcal{A}$ could then be replaced by its nondefectiveness (we refer the reader to [19, Section 2.1.2] for details). However, we prefer to state Lemma 3.3 in terms of irreducibility since this condition is easier to handle: it can be checked more directly and, up to a linear change of coordinates, a reducible $\mathcal{A}$ can be put into block-triangular form with irreducible diagonal blocks. This block decomposition is a key argument in the proof of Lemma 3.5.

We now consider the case where $\mathcal{A}$ is not necessarily irreducible. Here, a Barabanov norm for $\mathcal{A}$ in general does not exist, and hence item (c) from Lemma 3.3 cannot be expected.

Lemma 3.5. *Let $P \in \mathcal{M}_N(\mathbb{R})$ be a stochastic strongly connected matrix and $\mathcal{A} = (A_1, \dots, A_N) \in \mathcal{M}_d(\mathbb{R})^N$. Then the following statements are equivalent:*

(a) $\rho_{\mathrm{d}}(\mathcal{A}) = \rho_{\mathrm{p}}(P, \mathcal{A})$.

(b) $\rho(A_{i_k} \cdots A_{i_1})^{1/k} = \rho_{\mathrm{d}}(\mathcal{A})$ *for every P-cycle $(i_1, \dots, i_k)$.*

Proof. Before giving the core of the argument, we start with a set of remarks. First, up to a linear change of coordinates, $A_1, \dots, A_N$ can be presented in block-triangular form as

$$A_j = \begin{pmatrix} A_j^{(1)} & * & * & \cdots & * \\ 0 & A_j^{(2)} & * & \cdots & * \\ 0 & 0 & A_j^{(3)} & \cdots & * \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & A_j^{(R)} \end{pmatrix}, \qquad j \in [\![1, N]\!],$$

with $\mathcal{A}^{(r)} = (A_1^{(r)}, \dots, A_N^{(r)})$ irreducible for every $r \in [\![1, R]\!]$. Remark that, on the one hand, according to [19, Proposition 1.5], we have $\rho_{\mathrm{d}}(\mathcal{A}) = \max_{r \in [\![1,R]\!]} \rho_{\mathrm{d}}(\mathcal{A}^{(r)})$ and, on the other hand, it follows from [17, Theorem 1.1] and the strong connectedness of P that $\rho_{\mathrm{p}}(P, \mathcal{A}) = \max_{r \in [\![1,R]\!]} \rho_{\mathrm{p}}(P, \mathcal{A}^{(r)})$. Moreover, for every P-cycle $(i_1, \dots, i_k)$, we have

$$\rho_{\mathrm{d}}(\mathcal{A}) \geq \rho(A_{i_k} \cdots A_{i_1})^{1/k} = \max_{r \in [\![1,R]\!]} \rho\left(A_{i_k}^{(r)} \cdots A_{i_1}^{(r)}\right)^{1/k}, \tag{3.3}$$

where the inequality comes from (2.2) and the equality results from the simple fact that the spectral radius of a block-triangular matrix is equal to the maximum of the spectral radii over the diagonal blocks.

Since (a) implies (b) by Lemma 3.2, we are left to prove the converse implication. Assume that (b) holds true. Then (a) holds trivially if $\rho_{\mathrm{d}}(\mathcal{A}) = 0$. Otherwise, we can assume, with no loss of generality, that $\rho_{\mathrm{d}}(\mathcal{A}) = 1$ up to replacing $\mathcal{A}$ by $\rho_{\mathrm{d}}(\mathcal{A})^{-1}\mathcal{A}$. By assumption and (3.3), for every P-cycle $(i_1, \dots, i_k)$, there exists $r \in [\![1, R]\!]$ such that $\rho\left(A_{i_k}^{(r)} \cdots A_{i_1}^{(r)}\right) = 1$.

We claim that r can be chosen independently of the P-cycle. We argue by contradiction, i.e., we assume that, for every $r \in [\![1, R]\!]$, there exists a P-cycle $i^r = (i_1^r, \dots, i_{\ell_r}^r)$ such that $\rho(A^{(r)}(i^r)) < 1$. Let $j^r = (j_1^r, \dots, j_{k_r}^r)$ be a P-word such that $j_1^r = i_1^r$ and $p_{j_{k_r}^r i_1^{r+1}} > 0$ (with the convention that $i_1^{R+1} = i_1^1$). Then, for every $n \in \mathbb{N}$,

$$A(j^R)A(i^R)^n \cdots A(j^2)A(i^2)^n A(j^1)A(i^1)^n \in C(P).$$

For every n, we apply (b) to the above product, and we deduce from (3.3) that there exists $r_n \in [\![1, R]\!]$ such that

$$\begin{aligned} &\rho\left(A^{(r_n)}(j^R)A^{(r_n)}(i^R)^n \cdots A^{(r_n)}(j^2)A^{(r_n)}(i^2)^n A^{(r_n)}(j^1)A^{(r_n)}(i^1)^n\right) \\ &\qquad\qquad = \rho\left(A(j^R)A(i^R)^n \cdots A(j^2)A(i^2)^n A(j^1)A(i^1)^n\right) = 1. \end{aligned}$$

Let $(n_q)_{q\in\mathbb{N}}$ be an increasing sequence such that there exists $\bar{r}\in [\![1,R]\!]$ with $r_{n_q}=\bar{r}$ for every $q\in\mathbb{N}$. Since $\mathcal{A}^{(\bar{r})}$ is irreducible, there exists a Barabanov norm $\|\cdot\|_{\bar{r}}$ for $\mathcal{A}^{(\bar{r})}$. Then, for every $q\in\mathbb{N}$, we have

$$\begin{aligned}
1 &= \rho\left(A^{(\bar{r})}(j^R)A^{(\bar{r})}(i^R)^{n_q}\cdots A^{(\bar{r})}(j^2)A^{(\bar{r})}(i^2)^{n_q}A^{(\bar{r})}(j^1)A^{(\bar{r})}(i^1)^{n_q}\right)\\
&\leq \left\|A^{(\bar{r})}(j^R)A^{(\bar{r})}(i^R)^{n_q}\cdots A^{(\bar{r})}(j^2)A^{(\bar{r})}(i^2)^{n_q}A^{(\bar{r})}(j^1)A^{(\bar{r})}(i^1)^{n_q}\right\|_{\bar{r}}\\
&\leq \left\|A^{(\bar{r})}(i^{\bar{r}})^{n_q}\right\|_{\bar{r}},
\end{aligned}$$

where the last inequality follows from the fact that $\|\cdot\|_{\bar{r}}$ is a Barabanov norm. Since $\rho(A^{(\bar{r})}(i^{\bar{r}}))<1$, we have that $\left\|A^{(\bar{r})}(i^{\bar{r}})^{n_q}\right\|_{\bar{r}} \xrightarrow[q\to\infty]{} 0$, hence the contradiction.

We thus have proved that there exists $\bar{r}\in[\![1,R]\!]$ such that, for every P-cycle $(i_1,\dots,i_k)$,

$$\rho\left(A^{(\bar{r})}_{i_k}\cdots A^{(\bar{r})}_{i_1}\right) = 1 = \rho_{\mathrm{d}}(\mathcal{A}).$$

On the other hand, by (2.2), we have $\rho\left(A^{(\bar{r})}_{i_k}\cdots A^{(\bar{r})}_{i_1}\right)\leq \rho_{\mathrm{d}}(\mathcal{A}^{(\bar{r})})$. Since $\rho_{\mathrm{d}}(\mathcal{A}^{(\bar{r})})\leq\rho_{\mathrm{d}}(\mathcal{A})$, we deduce that

$$\rho\left(A^{(\bar{r})}_{i_k}\cdots A^{(\bar{r})}_{i_1}\right) = \rho_{\mathrm{d}}(\mathcal{A}^{(\bar{r})}) = \rho_{\mathrm{d}}(\mathcal{A})$$

for every P-cycle $(i_1,\dots,i_k)$. Then, using Lemma 3.3, we obtain that

$$\rho_{\mathrm{p}}(P,\mathcal{A})\geq\rho_{\mathrm{p}}(P,\mathcal{A}^{(\bar{r})}) = \rho_{\mathrm{d}}(\mathcal{A}^{(\bar{r})}) = \rho_{\mathrm{d}}(\mathcal{A}),$$

and then (a) holds thanks to (2.9). □

We can conclude now the proof of Theorem 3.1.

Proof of Theorem 3.1. Recall that, thanks to Lemma 3.2, we are only left to prove that (b) implies (a). We first decompose P and ν according to Proposition 2.6 and use in the sequel the same notations as in its statement. Thanks to (2.5) and (2.6), we have

$$\rho_{\mathrm{p}}(\nu,P,\mathcal{A}) = \sum_{j=1}^{R}\alpha_j\rho_{\mathrm{p}}(\nu^{[j]},P,\mathcal{A}). \tag{3.4}$$

For $j\in[\![1,R]\!]$, let $\mathcal{A}^{[j]}$ be the ordered n_j-tuple made of the matrices A_ℓ such that $\nu^{[j]}_\ell>0$. Notice that $\rho_{\mathrm{p}}(\nu^{[j]},P_j,\mathcal{A}^{[j]}) = \rho_{\mathrm{p}}(\nu^{[j]},P,\mathcal{A})$ for every $j\in[\![1,R]\!]$. Using (2.9) and the fact that $\mathcal{A}^{[j]}$ is made of matrices from $\mathcal{A}$, we obtain that, for every $j\in[\![1,R]\!]$,

$$\rho_{\mathrm{p}}(\nu^{[j]},P_j,\mathcal{A}^{[j]})\leq\rho_{\mathrm{d}}(\mathcal{A}^{[j]})\leq\rho_{\mathrm{d}}(\mathcal{A}). \tag{3.5}$$

Let $\mathcal{I}_i$ be defined for $i\in[\![1,R]\!]$ as in Lemma 2.7 and let $j\in[\![1,R]\!]$ be such that $\alpha_j>0$. Thanks to Lemma 2.7, there exists a (ν,P)-cycle $(i_1,\dots,i_k)$ with $i_1,\dots,i_k$ in $\mathcal{I}_j$. Then, by (2.2), (3.5), and (b), we have

$$\rho_{\mathrm{d}}(\mathcal{A}^{[j]})\leq\rho_{\mathrm{d}}(\mathcal{A}) = \rho(A_{i_k}\cdots A_{i_1})^{1/k}\leq\rho_{\mathrm{d}}(\mathcal{A}^{[j]}).$$

In particular, $\rho_{\mathrm{d}}(\mathcal{A}) = \rho_{\mathrm{d}}(\mathcal{A}^{[j]})$ and $\rho(A_{i_k}\cdots A_{i_1})^{1/k} = \rho_{\mathrm{d}}(\mathcal{A}^{[j]})$. Lemma 3.5 applied to P_j and $\mathcal{A}^{[j]}$ yields that $\rho_{\mathrm{p}}(\nu^{[j]},P_j,\mathcal{A}^{[j]}) = \rho_{\mathrm{d}}(\mathcal{A}^{[j]})$. Hence $\rho_{\mathrm{p}}(\nu^{[j]},P_j,\mathcal{A}^{[j]}) = \rho_{\mathrm{d}}(\mathcal{A})$, and, since this holds for every $j\in[\![1,R]\!]$ such that $\alpha_j>0$, it follows from (3.4) that $\rho_{\mathrm{p}}(\nu,P,\mathcal{A}) = \rho_{\mathrm{d}}(\mathcal{A})$, as required. □

Remark 3.6. Theorem 3.1 and Lemmas 3.3 and 3.5 characterize equality between deterministic and probabilistic joint spectral radii in terms of P-cycles and (ν,P)-cycles only, and hence only on $\lceil P \rceil$ and $\lceil \nu \rceil$ (see Remark 2.4). In other words, equality in Theorem 3.1(a) depends only on the graph associated with the Markov chain and the possible choices of initial states, but not on the precise values of the non-zero initial and transition probabilities.

3.2 Geometric characterization of equality between $\rho_{\mathrm{d}}(\mathcal{A})$ and $\rho_{\mathrm{p}}(P, \mathcal{A})$

It is clear from Theorem 3.1 that equality between $\rho_{\mathrm{d}}(\mathcal{A})$ and $\rho_{\mathrm{p}}(P,\mathcal{A})$ is possible only for restricted choices of $\mathcal{A}$. The goal of this section is to provide a more precise description of such choices of $\mathcal{A}$ using results from [25], where the authors classify matrix semigroups of constant spectral radius. We start with the following proposition.

Proposition 3.7. *Let $P \in \mathcal{M}_N(\mathbb{R})$ be a stochastic strongly connected matrix and $\mathcal{A} = (A_1,\dots,A_N) \in \mathcal{M}_d(\mathbb{R})^N$ be such that $\rho_{\mathrm{d}}(\mathcal{A}) = \rho_{\mathrm{p}}(P,\mathcal{A})$. Assume that there exists $s \in [\![1,N]\!]$ such that $C(P,s)$ is irreducible. Then there exists an invertible matrix $G \in \mathcal{M}_d(\mathbb{R})$ such that, for every P-cycle i starting at s, either $A(i)$ is singular or $\rho_{\mathrm{d}}(\mathcal{A})^{-k} G A(i) G^{-1}$ is orthogonal, where k is the length of i.*

Proof. We only have to provide an argument if there exists a P-cycle i_* starting at s such that $A(i_*)$ is invertible. In that case, from (2.2), $\rho_{\mathrm{d}}(\mathcal{A}) \geq \rho(A(i_*))^{1/k_*} > 0$, where k_* is the length of i_*. From Lemma 3.5, the set

$$\{\rho_{\mathrm{d}}(\mathcal{A})^{-k} A(i) \mid k \in \mathbb{N},\ i \text{ is a } P\text{-cycle starting at } s \text{ of length } k\}$$

is a matrix semigroup with constant spectral radius. Since, moreover, this semigroup is also irreducible, the conclusion follows from [25, Theorem 2]. $\square$

Remark 3.8. As remarked in [25], the problem of classifying matrix semigroups with constant spectral radius is highly nontrivial when the semigroup contains singular matrices. By using additional results from [25], we may obtain, under the assumptions of Proposition 3.7, properties on $\rho_{\mathrm{d}}(\mathcal{A})^{-k} G A(i) G^{-1}$ that are weaker than orthogonality but apply to all matrices $A(i) \in C(P,s)$, and not only nonsingular ones. We refer the interested reader to [25, Theorem 3 and Corollary 6].

A limitation of Proposition 3.7 lies in the fact that, in general, given a stochastic and strongly connected matrix P, it is a nontrivial task to verify the existence of an index s such that $C(P,s)$ is irreducible, even if $\mathcal{A}$ is itself irreducible. However, this is true if one assumes in addition that $\mathcal{A}$ contains only invertible matrices and that all diagonal elements of P are positive, in which case we have the following proposition.

Proposition 3.9. *Let $P \in \mathcal{M}_d(\mathbb{R})$ be a stochastic strongly connected matrix with positive diagonal entries and $\mathcal{A} = (A_1,\dots,A_N) \in \mathcal{M}_d(\mathbb{R})^N$ be irreducible with $A_1,\dots,A_N$ invertible. Then, for every $s \in [\![1,N]\!]$, $C(P,s)$ is irreducible. Moreover, $\rho_{\mathrm{d}}(\mathcal{A}) = \rho_{\mathrm{p}}(P,\mathcal{A})$ if and only if there exists an invertible matrix $G \in \mathcal{M}_d(\mathbb{R})$ such that, for every $i \in [\![1,N]\!]$, $\rho_{\mathrm{d}}(\mathcal{A})^{-1} G A_i G^{-1}$ is orthogonal.*

Proof. Let $s \in [\![1,N]\!]$ and consider the group $\widetilde{C}(P,s)$ generated by $C(P,s)$. We claim that $A_1,\dots,A_N \in \widetilde{C}(P,s)$. Indeed, since P is strongly connected, there exists a P-cycle $i = (i_1,\dots,i_k)$ starting at s such that $\{i_1,\dots,i_k\} = [\![1,N]\!]$. Since $p_{i_k i_k} > 0$, then $A_{i_k}^2 A_{i_{k-1}} \cdots A_{i_1} \in C(P,s)$ and

$$A_{i_k} = \left(A_{i_k}^2 A_{i_{k-1}} \cdots A_{i_1}\right)\left(A_{i_k} \cdots A_{i_1}\right)^{-1} \in \widetilde{C}(P,s).$$

Similarly, since $p_{i_{k-1}i_{k-1}} > 0$, then $A_{i_k} A_{i_{k-1}}^2 A_{i_{k-2}} \cdots A_{i_1} \in C(P,s)$ and

$$A_{i_{k-1}} = A_{i_k}^{-1}\left(A_{i_k} A_{i_{k-1}}^2 A_{i_{k-2}} \cdots A_{i_1}\right)\left(A_{i_k} A_{i_{k-1}} \cdots A_{i_1}\right)^{-1} A_{i_k} \in \widetilde{C}(P,s).$$

An inductive reasoning based on the identity

$$A_{i_j} = \left(A_{i_k} \cdots A_{i_{j+1}}\right)^{-1}\left(A_{i_k} \cdots A_{i_{j+1}} A_{i_j}^2 A_{i_{j-1}} \cdots A_{i_1}\right)\left(A_{i_k} \cdots A_{i_1}\right)^{-1}\left(A_{i_k} \cdots A_{i_{j+1}}\right) \tag{3.6}$$

yields that $A_{i_j} \in \widetilde{C}(P,s)$ for $j \in [\![1,k]\!]$, as required.

To prove that $C(P,s)$ is irreducible for every s, assume by contradiction that there exists $s \in [\![1,N]\!]$ such that $C(P,s)$ is reducible. Then the group $\widetilde{C}(P,s)$ is also reducible, however, since it contains $A_1,\dots,A_N$, this contradicts the irreducibility of $\mathcal{A}$.

Since $A_1,\dots,A_N$ are invertible matrices, $\rho_{\mathrm{d}}(\mathcal{A})$ is positive and, with no loss of generality, we can assume that $\rho_{\mathrm{d}}(\mathcal{A}) = 1$. If $\rho_{\mathrm{d}}(\mathcal{A}) = \rho_{\mathrm{p}}(P,\mathcal{A})$, then, applying Proposition 3.7 to $C(P,1)$, there exists a basis in which every $M \in C(P,1)$ is orthogonal. Hence, in this same basis, $\widetilde{C}(P,1)$ is also made of orthogonal matrices, yielding the conclusion. On the other hand, if there exists a basis in which $A_1,\dots,A_N$ are orthogonal, then $\rho(A(i)) = 1$ for every P-word i, and the conclusion follows by Lemma 3.5. □

Remark 3.10. Notice that, to obtain the second part of the conclusion of Proposition 3.9, it is enough that there exists $s \in [\![1,N]\!]$ such that $C(P,s)$ is irreducible and the generated group $\widetilde{C}(P,s)$ contains all matrices $A_1,\dots,A_N$. The assumption that P has positive diagonal entries is used to guarantee the latter, and therefore it can be replaced by any other condition ensuring that $A_1,\dots,A_N$ belong to $\widetilde{C}(P,s)$ for some $s \in [\![1,N]\!]$. For instance, assume that $p_{11} = 0$ and $p_{jj} > 0$ for $j \in [\![2,N]\!]$. For every P-cycle $(i_1,\dots,i_k)$ with $i_1 = 1$ and $i_j \neq 1$ for every $j \in [\![2,k]\!]$, we can proceed as in the proof of the proposition to obtain that $A_{i_j} \in \widetilde{C}(P,1)$ for every $j \in [\![2,k]\!]$ and use the identity

$$A_{i_1} = (A_{i_k} \cdots A_{i_2})^{-1}(A_{i_k} \cdots A_{i_1})$$

to obtain that $A_{i_1} \in \widetilde{C}(P,1)$. Since P is strongly connected, every matrix A_i, $i \in [\![1,N]\!]$, belongs to such a P-cycle, hence the conclusion.

At the light of Remark 3.10, we may wonder whether the second part of the conclusion of Proposition 3.9 can be obtained under even weaker assumptions on the matrix P, allowing for instance the presence of more than one diagonal element equal to zero, but requiring at least one non-zero element in the diagonal. The example given below shows that this is not possible.

Example 3.11. Consider the case $d = 2$, $N = 3$,

$$A_1 = \mathrm{Id}, \quad A_2 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad A_3 = \begin{pmatrix} 0 & -\frac{1}{2} \\ 1 & 0 \end{pmatrix}, \quad P = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}.$$

Note that, in this case,

$$A_3A_2 = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & 1 \end{pmatrix}, \quad A_2A_3 = \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{2} \end{pmatrix}.$$

The matrix P is stochastic, strongly connected, and its unique invariant probability is $\nu = (\frac{1}{2}, \frac{1}{4}, \frac{1}{4})$. Moreover, $\mathcal{A} = (A_1, A_2, A_3)$ is irreducible and A_1, A_2, A_3 are invertible. Denoting by $\|\cdot\|$ the Euclidean norm in $\mathbb{R}^2$, we have $\|A_1\| = \|A_2\| = \|A_3\| = 1$, yielding that $\rho_{\mathrm{d}}(\mathcal{A}) \leq 1$, and we easily check that $\rho_{\mathrm{d}}(\mathcal{A}) = 1$ by considering $\sigma \in \mathfrak{S}$ given by $\sigma(i) = 1$ for every $i \in \mathbb{N}$. Moreover, for any (ν, P)-word $(i_1, \dots, i_k)$, there exist an integer $\ell \geq 0$ and $a, b \in \{0, 1\}$ such that $\|A_{i_k} \cdots A_{i_1}\| = \|A_2^b (A_3A_2)^\ell A_3^a\|$. Setting $x = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ in the case $a = 1$ and $x = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ in the case $a = 0$, it is immediate to verify that $\|A_2^b (A_3A_2)^\ell A_3^a x\| = 1$, yielding that $\|A_{i_k} \cdots A_{i_1}\| = 1$ for every (ν, P)-word $(i_1, \dots, i_k)$, and thus $\rho_{\mathrm{p}}(\nu, P, \mathcal{A}) = \rho_{\mathrm{p}}(P, A) = 1$. However, A_3 is not similar to an orthogonal matrix, and hence the second conclusion of Proposition 3.9 does not hold. Notice moreover that, in this case, $C(P, 1) = \{(A_3A_2)^n \mid n \in \mathbb{N} \cup \{0\}\}$, $C(P, 2) = \{(A_3A_2)^n \mid n \in \mathbb{N}\}$, and $C(P, 3) = \{(A_2A_3)^n \mid n \in \mathbb{N}\}$, and thus $C(P, s)$ is reducible for every $s \in \{1, 2, 3\}$.

Remark 3.12. We now provide a description of all cases where equality holds between $\rho_{\mathrm{d}}(\mathcal{A})$ and $\rho_{\mathrm{p}}(\nu, P, \mathcal{A})$ under the assumption that $\mathcal{A}$ is irreducible and made of two invertible matrices.

(a) If $P = \begin{pmatrix} p & 1-p \\ 1-q & q \end{pmatrix}$ for $p, q \in [0, 1)$ with $p + q > 0$, by Remark 3.10, equality occurs if and only if there exists an invertible matrix $G \in \mathcal{M}_d(\mathbb{R})$ such that $\rho_{\mathrm{d}}(\mathcal{A})^{-1} G A_1 G^{-1}$ and $\rho_{\mathrm{d}}(\mathcal{A})^{-1} G A_2 G^{-1}$ are orthogonal.

(b) If $P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, equality occurs if and only if $\rho(A_1A_2) = \rho(A_2A_1) = \rho_{\mathrm{d}}(\mathcal{A})^2$.

(c) If $P = \mathrm{Id}_2$, equality occurs if and only if $\rho(A_i) = \rho_{\mathrm{d}}(\mathcal{A})$ whenever $\nu_i > 0$, $i \in \{1, 2\}$.

(d) If $P = \begin{pmatrix} 1 & 0 \\ 1-p & p \end{pmatrix}$ for some $p \in [0, 1)$, then equality is equivalent to $\rho(A_1) = \rho_{\mathrm{d}}(\mathcal{A})$.

(e) If $P = \begin{pmatrix} p & 1-p \\ 0 & 1 \end{pmatrix}$ for some $p \in [0, 1)$, then equality is equivalent to $\rho(A_2) = \rho_{\mathrm{d}}(\mathcal{A})$.

3.3 Equality between $\rho_{\mathrm{d}}(\mathcal{A})$ and $\rho_{\mathrm{p}}(\mathcal{A})$

Based on the results obtained previously, we can now address the issue of characterizing the equality between $\rho_{\mathrm{d}}(\mathcal{A})$ and $\rho_{\mathrm{p}}(\mathcal{A})$. Recall that the latter is defined as the maximum of $\rho_{\mathrm{p}}(\nu, P, \mathcal{A})$ over all pairs (ν, P).

Theorem 3.13. *Let $\mathcal{A} = (A_1, \dots, A_N) \in \mathcal{M}_d(\mathbb{R})^N$. Then the following statements are equivalent:*

(a) $\rho_{\mathrm{d}}(\mathcal{A}) = \rho_{\mathrm{p}}(\mathcal{A})$.

(b) *There exist $i_1, \ldots, i_k \in [\![1,N]\!]$ pairwise distinct such that*

$$\rho_{\mathrm{d}}(\mathcal{A}) = \rho(A_{i_k} \cdots A_{i_1})^{1/k}. \tag{3.7}$$

Proof. We start by proving that (a) implies (b). Recall that, by Remark 2.12, there exist a stochastic matrix P and an invariant probability ν for P such that $\rho_{\mathrm{p}}(\nu, P, \mathcal{A}) = \rho_{\mathrm{p}}(\mathcal{A})$. Using (a), we deduce that $\rho_{\mathrm{p}}(\nu, P, \mathcal{A}) = \rho_{\mathrm{d}}(\mathcal{A})$. It is clear that there exists a (ν, P)-cycle $(i_1, \ldots, i_k)$ such that $i_1, \ldots, i_k$ are pairwise distinct, and the conclusion follows from Theorem 3.1.

To prove that (b) implies (a), let $P = (p_{ij})$ be a stochastic matrix with $p_{i_{j-1} i_j} = 1$ for $j \in [\![2,k]\!]$ and $p_{i_k i_1} = 1$. Set $\nu \in \mathbb{R}^N$ as the probability vector such that $\nu_{i_j} = \frac{1}{k}$ for $j \in [\![1,k]\!]$. Then ν is invariant under P and the set of (ν, P)-cycles is made of the shifts of $(i_1, \ldots, i_k)$ and their powers. Moreover, for every such (ν, P)-cycle $(j_1, \ldots, j_s)$, we have

$$\rho(A_{j_s} \cdots A_{j_1})^{1/s} = \rho(A_{i_k} \cdots A_{i_1})^{1/k} = \rho_{\mathrm{d}}(\mathcal{A}).$$

Indeed, this follows from the fact that $\rho(M_1 M_2) = \rho(M_2 M_1)$ for every $M_1, M_2 \in \mathcal{M}_d(\mathbb{R})$. Then Theorem 3.1(b) holds, hence $\rho_{\mathrm{p}}(\nu, P, \mathcal{A}) = \rho_{\mathrm{d}}(\mathcal{A})$, and the conclusion follows from (2.9). $\square$

Remark 3.14. It follows from (2.9) that, if $\rho_{\mathrm{d}}(\mathcal{A}) > 0$, the ratio $\frac{\rho_{\mathrm{p}}(\mathcal{A})}{\rho_{\mathrm{d}}(\mathcal{A})}$ belongs to $[0,1]$ and Theorem 3.13 addresses the case where it is equal to 1. We provide next an example where it is equal to 0, proving that it is not possible to find a uniform positive lower bound for this ratio. Indeed, considering

$$A_1 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \qquad A_2 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix},$$

an immediate computation yields

$$A_1^2 A_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \; A_1 A_2 A_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \; A_2 A_1^2 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

and $A_1^3 = A_1 A_2^2 = A_2 A_1 A_2 = A_2^2 A_1 = A_2^3 = 0$. Let $\|\cdot\|_1$ denote the matrix norm induced by the ℓ^1 norm in $\mathbb{R}^3$. Define

$$\mathcal{E} = \{(2,1,1,2,1,1,\ldots), (1,2,1,1,2,1,\ldots), (1,1,2,1,1,2,\ldots)\}$$

and, for $k \in \mathbb{N}$, let $\mathcal{E}_k$ be the set made of the three words of length k obtained by taking the first k entries of each element of $\mathcal{E}$. By an easy computation, we get that, for every $k \in \mathbb{N}$ and $(i_1, \ldots, i_k) \in [\![1,N]\!]^k$,

$$\|A_{i_k} \cdots A_{i_1}\|_1 = \begin{cases} 1, & \text{if } (i_1, \ldots, i_k) \in \mathcal{E}_k, \\ 0, & \text{otherwise.} \end{cases}$$

We then obtain that $\rho_{\mathrm{d}}(\mathcal{A}) = 1$. On the other hand, for every stochastic matrix $P \in \mathcal{M}_2(\mathbb{R})$ and every invariant probability vector ν for P, we have $\mathbb{P}_{(\nu,P)}(\mathcal{E}) = 0$. Hence

$$\lim_{n \to \infty} \|A_{i_n} \cdots A_{i_1}\|_1^{1/n} = 0 \quad \mathbb{P}_{(\nu,P)}\text{-a.s.},$$

proving that $\rho_{\mathrm{p}}(\nu, P, \mathcal{A}) = 0$. Then $\rho_{\mathrm{p}}(\mathcal{A}) = 0$.

4 Markov chains of higher order

In this section, we extend the previous results to probability measures on $\mathfrak{S}$ obtained from discrete-time shift-invariant Markov chains of order $m \geq 1$. Any such probability measure μ can be described by a pair (ν, P) of tensors of orders m and $m+1$, respectively, where the non-negative scalar $P_{i_1 \dots i_m i_{m+1}}$ represents the probability to switch from the state i_m to the state i_{m+1} when the previous m states of the chain are $(i_1, \dots, i_m)$, and $\nu_{i_1 \dots i_m}$ represents the probability of the first m states being $(i_1, \dots, i_m)$. In particular, for every $(i_1, \dots, i_m) \in [\![1, N]\!]^m$, we have that

$$\sum_{i_{m+1}=1}^{N} P_{i_1 \dots i_m i_{m+1}} = 1$$

and ν satisfies

$$\sum_{(i_1, \dots, i_m) \in [\![1,N]\!]^m} \nu_{i_1 \dots i_m} = 1.$$

We refer to such ν and P as a *probability tensor of order m* and a *stochastic tensor of order $m+1$*, respectively. The shift-invariance property now reads

$$\sum_{i_1=1}^{N} \nu_{i_1 \dots i_m} P_{i_1 \dots i_{m+1}} = \nu_{i_2 \dots i_{m+1}}, \qquad \text{for every } (i_2, \dots, i_{m+1}) \in [\![1,N]\!]^m,$$

and any probability tensor ν satisfying the above shift-invariant property is said to be *invariant* under P. The probabilistic joint spectral radius $\rho_{\mathrm{p}}(\nu, P, \mathcal{A})$ associated with (ν, P) is still defined by (2.5), where the expectation $\mathbb{E}_{(\nu,P)}$ corresponds to the probability measure on $\mathfrak{S}$ defined above.

Markov chains of order $m \geq 1$ can be canonically transformed into Markov chains of order 1 by considering as state space the set $[\![1,N]\!]^m$ and defining a pair $(\widehat{\nu}, \widehat{P})$ from (ν, P) by $\widehat{\nu}_{(i_1, \dots, i_m)} = \nu_{i_1 \dots i_m}$ and

$$\widehat{P}_{(i_1, \dots, i_m), (j_1, \dots, j_m)} = \begin{cases} P_{i_1 \dots i_m j_m} & \text{if } (i_2, \dots, i_m) = (j_1, \dots, j_{m-1}), \\ 0 & \text{otherwise}, \end{cases}$$

for every $(i_1, \dots, i_m)$ and $(j_1, \dots, j_m)$ in $[\![1,N]\!]^m$. It is immediate from the definitions and the shift-invariance property that

$$\rho_{\mathrm{p}}(\nu, P, \mathcal{A}) = \rho_{\mathrm{p}}(\widehat{\nu}, \widehat{P}, \widehat{\mathcal{A}}),$$

where $\widehat{\mathcal{A}} = (\widehat{A}_{i_1 \dots i_m})_{(i_1, \dots, i_m) \in [\![1,N]\!]^m}$ and $\widehat{A}_{i_1 \dots i_m} = A_{i_m}$ for every $(i_1, \dots, i_m) \in [\![1,N]\!]^m$.

For every positive integer k, we say that $(i_1, \dots, i_k)$ is a (ν, P)*-cycle* if

$$\big((i_{-m+2}, \dots, i_0, i_1), \dots, (i_{k-m+1}, \dots, i_k)\big)$$

is a $(\widehat{\nu}, \widehat{P})$-cycle, where $z \mapsto i_z$ is extended to $\mathbb{Z}$ by k-periodicity.

Applying Theorem 3.1 to $(\widehat{\nu}, \widehat{P})$ and $\widehat{\mathcal{A}}$, we deduce at once the following.

Theorem 4.1. *Let m be a positive integer, P be a stochastic tensor of order $m+1$, ν be an invariant probability tensor for P, and $\mathcal{A} = (A_1, \dots, A_N) \in \mathcal{M}_d(\mathbb{R})^N$. Then the following statements are equivalent:*

(a) $\rho_{\mathrm{d}}(\mathcal{A}) = \rho_{\mathrm{p}}(\nu, P, \mathcal{A})$.

(b) $\rho(A_{i_k} \cdots A_{i_1})^{1/k} = \rho_{\mathrm{d}}(\mathcal{A})$ *for every* (ν, P)*-cycle* $(i_1, \dots, i_k)$.

Recall that (1.1) is said to be *periodically stable* if $\rho(\sigma) < 1$ for all periodic signals $\sigma \in \mathfrak{S}$. It has been shown in [12] that this property implies $\rho_{\mathrm{p}}(\nu, P, \mathcal{A}) < 1$ for every strongly connected stochastic matrix $P \in \mathcal{M}_N(\mathbb{R})$, where $\nu \in \mathbb{R}^N$ is the unique invariant probability vector for P. A slightly improved version of this result can be obtained as a consequence of Theorem 4.1 as stated in the following corollary.

Corollary 4.2. *Assume that* (1.1) *is periodically stable. Then, for every $m \in \mathbb{N}$, every stochastic tensor P of order $m+1$, and every invariant probability tensor ν for P, we have* $\rho_{\mathrm{p}}(\nu, P, \mathcal{A}) < 1$.

Proof. By the Joint Spectral Radius Theorem (see, e.g., [19, Theorem 2.3]), periodic stability implies that $\rho_{\mathrm{d}}(\mathcal{A}) \leq 1$. In the case $\rho_{\mathrm{d}}(\mathcal{A}) < 1$, the conclusion follows immediately. Otherwise, when $\rho_{\mathrm{d}}(\mathcal{A}) = 1$, the periodic stability assumption implies that assertion (b) from Theorem 4.1 does not hold, which proves that $\rho_{\mathrm{p}}(\nu, P, \mathcal{A}) < \rho_{\mathrm{d}}(\mathcal{A}) = 1$, yielding the conclusion. $\square$

Similarly as for Theorem 4.1, we deduce by applying Theorem 3.13 to $(\widehat{\nu}, \widehat{P})$ and $\widehat{\mathcal{A}}$ the following.

Theorem 4.3. *Let m be a positive integer and $\mathcal{A} = (A_1, \dots, A_N) \in \mathcal{M}_d(\mathbb{R})^N$. Then the following statements are equivalent:*

(a) $\rho_{\mathrm{d}}(\mathcal{A}) = \rho_{\mathrm{p}}(m, \mathcal{A})$*, where $\rho_{\mathrm{p}}(m, \mathcal{A})$ is the supremum of $\rho_{\mathrm{p}}(\nu, P, \mathcal{A})$ over all pairs (ν, P) with P a stochastic tensor of order $m+1$ and ν an invariant probability tensor for P.*

(b) *There exist $i_1, \dots, i_k \in [\![1, N]\!]$ such that*

$$\rho_{\mathrm{d}}(\mathcal{A}) = \rho(A_{i_k} \cdots A_{i_1})^{1/k}$$

and $(i_{j_1}, \dots, i_{j_1+m-1}) \neq (i_{j_2}, \dots, i_{j_2+m-1})$ whenever $j_1, j_2 \in [\![1, k]\!]$ with $j_1 \neq j_2$, where $z \mapsto i_z$ is extended to $\mathbb{Z}$ by k-periodicity.

As a consequence of Theorem 4.3, we have the following corollary. To state it, recall that $\mathcal{A}$ is said to have the *finiteness property* if there exist $i_1, \dots, i_k \in [\![1, N]\!]$ such that $\rho_{\mathrm{d}}(\mathcal{A}) = \rho(A_{i_k} \cdots A_{i_1})^{1/k}$.

Corollary 4.4. *Let $\mathcal{A} = (A_1, \dots, A_N)$. Then $\mathcal{A}$ has the finiteness property if and only if there exists $m \in \mathbb{N}$ such that $\rho_{\mathrm{d}}(\mathcal{A}) = \rho_{\mathrm{p}}(m, \mathcal{A})$.*

Proof. If there exists m such that $\rho_{\mathrm{d}}(\mathcal{A}) = \rho_{\mathrm{p}}(m, \mathcal{A})$, then the finiteness property of $\mathcal{A}$ follows immediately from Theorem 4.3. Assume now that $\mathcal{A}$ has the finiteness property and let $i_1, \dots, i_k \in [\![1, N]\!]$ be such that $\rho_{\mathrm{d}}(\mathcal{A}) = \rho(A_{i_k} \cdots A_{i_1})^{1/k}$. Extend $z \mapsto i_z$ over $\mathbb{Z}$ by k-periodicity and let k' be the minimal period of $z \mapsto i_z$. Without loss of generality, we can assume that $k = k'$. We claim that property (b) of Theorem 4.3 holds with $m = k$. Indeed, let $j_1, j_2 \in [\![1, k]\!]$ be such that $(i_{j_1}, \dots, i_{j_1+k-1}) = (i_{j_2}, \dots, i_{j_2+k-1})$ and assume, to obtain a contradiction, that $j_1 \neq j_2$. Without loss of generality, $j_1 < j_2$. Set $k'' = j_2 - j_1$

and notice that $0 < k'' < k$ and $i_{j_1+\ell} = i_{j_1+k''+\ell}$ for every $\ell \in [\![0, k-1]\!]$. Since $z \mapsto i_z$ is k-periodic, the previous equality holds for every $\ell \in \mathbb{Z}$, proving that $z \mapsto i_z$ is k''-periodic, contradicting the minimality of k as period of $z \mapsto i_z$. Hence property (a) of Theorem 4.3 holds, as required. $\square$

Remark 4.5. Given $\mathcal{A} = (A_1, \dots, A_N)$, $\ell \in \mathbb{N}$, and a word $w = (i_1, \dots, i_\ell) \in [\![1, N]\!]^\ell$, set $A(w) = A_{i_\ell} \cdots A_{i_1}$ and let $|w| = \ell$ be the length of w. Notice that, by proceeding similarly to the second part of the proof of Theorem 3.13, we can construct, for every word w of finite length, a Markov chain of order $|w|$ with tensors ν_w, P_w such that $\rho(A(w))^{1/|w|} = \rho_{\mathrm{p}}(\nu_w, P_w, \mathcal{A})$. We deduce that

$$\rho_{\mathrm{d}}(\mathcal{A}) = \sup_{w \text{ word of finite length}} \rho(A(w))^{1/|w|} \leq \sup_{m \in \mathbb{N}} \rho_{\mathrm{p}}(m, \mathcal{A}),$$

where the equality is a consequence of the Joint Spectral Radius Theorem (see, e.g., [19]). Since, moreover, $\rho_{\mathrm{p}}(m, \mathcal{A}) \leq \rho_{\mathrm{d}}(\mathcal{A})$ for every m, it follows that $\rho_{\mathrm{d}}(\mathcal{A}) = \sup_{m \in \mathbb{N}} \rho_{\mathrm{p}}(m, \mathcal{A})$.

A further characterization of the equivalence in Corollary 4.4 can then be stated as follows: an N-tuple of matrices $\mathcal{A} = (A_1, \dots, A_N)$ satisfies the finiteness property if and only if

$$\sup_{m, \nu, P} \rho_{\mathrm{p}}(\nu, P, \mathcal{A})$$

is attained at some (m, ν, P), where the supremum is taken over all (m, ν, P) with $m \in \mathbb{N}$, P a stochastic tensor of order $m+1$, and ν an invariant probability tensor for P.

References

[1] L. Arnold. *Random dynamical systems*. Springer Monographs in Mathematics. Springer-Verlag, Berlin, 1998.

[2] N. E. Barabanov. Lyapunov indicator of discrete inclusions. I–III. *Autom. Remote Control*, 49:152–157, 283–287, 558–565, 1988.

[3] V. D. Blondel, J. Theys, and A. A. Vladimirov. An elementary counterexample to the finiteness conjecture. *SIAM J. Matrix Anal. Appl.*, 24(4):963–970, 2003.

[4] T. Bousch and J. Mairesse. Asymptotic height optimization for topical IFS, Tetris heaps, and the finiteness conjecture. *J. Amer. Math. Soc.*, 15(1):77–111, 2002.

[5] F. Colonius and W. Kliemann. *Dynamical systems and linear algebra*, volume 158 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2014.

[6] O. L. V. Costa, M. D. Fragoso, and R. P. Marques. *Discrete-time Markov jump linear systems*. Probability and its Applications (New York). Springer-Verlag London, Ltd., London, 2005.

[7] O. L. V. Costa, M. D. Fragoso, and M. G. Todorov. *Continuous-time Markov jump linear systems*. Probability and its Applications (New York). Springer, Heidelberg, 2013.

[8] A. Crisanti, G. Paladin, and A. Vulpiani. *Products of random matrices in statistical physics*, volume 104 of *Springer Series in Solid-State Sciences*. Springer-Verlag, Berlin, 1993. With a foreword by Giorgio Parisi.

[9] X. Dai. Weakly Birkhoff recurrent switching signals, almost sure and partial stability of linear switched dynamical systems. *J. Differential Equations*, 250(9):3584–3629, 2011.

[10] X. Dai. Robust periodic stability implies uniform exponential stability of Markovian jump linear systems and random linear ordinary differential equations. *J. Franklin Inst.*, 351(5):2910–2937, 2014.

[11] X. Dai, Y. Huang, and M. Xiao. Almost sure stability of discrete-time switched linear systems: a topological point of view. *SIAM J. Control Optim.*, 47(4):2137–2156, 2008.

[12] X. Dai, Y. Huang, and M. Xiao. Periodically switched stability induces exponential stability of discrete-time linear switched systems in the sense of Markovian probabilities. *Automatica J. IFAC*, 47(7):1512–1519, 2011.

[13] I. Daubechies and J. C. Lagarias. Sets of matrices all infinite products of which converge. *Linear Algebra Appl.*, 161:227–263, 1992.

[14] Y. G. Fang, K. A. Loparo, and X. Feng. Almost sure and δ-moment stability of jump linear systems. *Internat. J. Control*, 59(5):1281–1307, 1994.

[15] X. Feng, K. A. Loparo, Y. Ji, and H. J. Chizeck. Stochastic stability properties of jump linear systems. *IEEE Trans. Automat. Control*, 37(1):38–53, 1992.

[16] H. Furstenberg and H. Kesten. Products of random matrices. *Ann. Math. Statist.*, 31:457–469, 1960.

[17] L. Gerencsér, G. Michaletzky, and Z. Orlovits. Stability of block-triangular stationary random matrices. *Systems Control Lett.*, 57(8):620–625, 2008.

[18] L. Gurvits. Stability of discrete linear inclusion. *Linear Algebra Appl.*, 231:47–85, 1995.

[19] R. Jungers. *The joint spectral radius. Theory and applications*, volume 385 of *Lecture Notes in Control and Information Sciences*. Springer-Verlag, Berlin, 2009.

[20] V. Kozyakin. The Berger-Wang formula for the Markovian joint spectral radius. *Linear Algebra Appl.*, 448:315–328, 2014.

[21] J. C. Lagarias and Y. Wang. The finiteness conjecture for the generalized spectral radius of a set of matrices. *Linear Algebra Appl.*, 214:17–42, 1995.

[22] D. Liberzon. *Switching in systems and control*. Systems & Control: Foundations & Applications. Birkhäuser Boston, Inc., Boston, MA, 2003.

[23] H. Lin and P. J. Antsaklis. Stability and stabilizability of switched linear systems: a survey of recent results. *IEEE Trans. Automat. Control*, 54(2):308–322, 2009.

[24] I. D. Morris. Mather sets for sequences of matrices and applications to the study of joint spectral radii. *Proc. Lond. Math. Soc. (3)*, 107(1):121–150, 2013.

[25] V. Y. Protasov and A. S. Voynov. Matrix semigroups with constant spectral radius. *Linear Algebra Appl.*, 513:376–408, 2017.

[26] G.-C. Rota and G. Strang. A note on the joint spectral radius. *Nederl. Akad. Wetensch. Proc. Ser. A 63 = Indag. Math.*, 22:379–381, 1960.

[27] E. Seneta. *Non-Negative Matrices and Markov Chains*. Springer Series in Statistics. Springer, New York, 2006. Revised reprint of the second edition [Springer-Verlag, New York, 1981].

[28] R. Shorten, F. Wirth, O. Mason, K. Wulff, and C. King. Stability criteria for switched and hybrid systems. *SIAM Rev.*, 49(4):545–592, 2007.

[29] Z. Sun and S. S. Ge. *Switched Linear Systems: Control and Design*. Communications and Control Engineering. Springer-Verlag, London, 2005.