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

Yacine Chitour, Guilherme Mazanti, Mario Sigalotti

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

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

December 20, 2018

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.

1 Introduction

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

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

where d and N are positive integers, $x_k \in \mathbb{R}^d$, $\mathfrak{S}_N$ 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

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

†Laboratoire des Signaux et Systèmes (L2S, UMR CNRS 8506), CNRS — CentraleSupelec — Université Paris-Sud, 3, rue Joliot Curie, 91192, Gif-sur-Yvette, France.

‡Laboratoire de Mathématiques d'Orsay, Univ. Paris-Sud, CNRS, Université Paris-Saclay, 91405 Orsay, France.

§Inria Paris & Laboratoire Jacques-Louis Lions, Sorbonne Université, Université Paris-Diderot SPC, CNRS, Inria, 75005 Paris, France.

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, 14, 15, 19, 20]. 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 $\sigma \in \mathfrak{S}_N$, one can measure the asymptotic behavior of the corresponding non-autonomous linear system by the quantity $\rho(\sigma)$ defined by

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

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 $\Sigma(\mathcal{A})$, one must formulate some condition which is independent of the choice of $\sigma \in \mathfrak{S}_N$. 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 $\sigma \in \mathfrak{S}_N$. Since its introduction in [17] and after the seminal paper [9], it has been extensively studied in the computer science and control theory communities (see, e.g., the monograph [12]).

The other approach to handle the asymptotic behavior of $\Sigma(\mathcal{A})$ is probabilistic and amounts to considering $\sigma \in \mathfrak{S}_N$ as a random process, and hence $\rho(\sigma)$ as a random variable. One may then consider as a *probabilistic joint spectral radius* the expected value of $\rho(\sigma)$ with respect to the probability law of σ. In this paper, we consider the case where σ is generated by a discrete-time Markov chain on the state space $\{1, \dots, N\}$ with a transition matrix P and a corresponding invariant probability ν, and we denote by $\rho_{\mathrm{p}}(\nu, P, \mathcal{A})$ the corresponding expected value of $\rho(\sigma)$.

A major result in this direction has been obtained in [10] and implies that, under a generic condition on P, $\rho(\sigma)$ is constant on a set of probability 1. There exists a vast literature on studying the properties of products of random matrices, and we refer the reader to [1, 5, 8] for more details. In the setting of the Markov chains considered in this paper, the above genericity assumption is satisfied if, for instance, the transition matrix P is strongly connected (see Definition 2.3(b) and Remark 2.4), in which case $\rho_{\mathrm{p}}(\nu, P, \mathcal{A})$ represents the typical behavior of $\Sigma(\mathcal{A})$ (i.e., $\rho(\sigma) = \rho_{\mathrm{p}}(\nu, P, \mathcal{A})$ almost surely).

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 $\sigma \in \mathfrak{S}_N$, while the probabilistic counterpart $\rho_{\mathrm{p}}(\nu, P, \mathcal{A})$ provides the average behavior for $\sigma \in \mathfrak{S}_N$ corresponding to the stochastic process defined by ν and P. Clearly, the deterministic approach provides a more conservative estimate of the asymptotic behavior of the system than the probabilistic one, in the sense that

$$\rho_{\mathrm{p}}(\nu, P, \mathcal{A}) \leq \rho_{\mathrm{d}}(\mathcal{A}). \tag{1.2}$$

The objective of this paper is to characterize the N-tuples $\mathcal{A}$ for which the average behavior corresponding to ν and P coincides with the worst behavior, i.e., one has equality in (1.2).

Our main result concerning equality cases in (1.2) (see Theorem 3.1) establishes that equality occurs if and only if $\rho_{\mathrm{d}}(\mathcal{A}) = \rho(A_{i_k} \cdots A_{i_1})^{1/k}$ for every $(i_1, \dots, i_k)$ that corresponds to a cycle in the directed weighted graph determined by P such that $\nu_{i_1} > 0$. This equivalence 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 3.7). The latter characterization is based on the description of matrix semigroups with constant spectral radius from [16].

Our strategy to prove Theorem 3.1 consists in considering first the particular case where $\mathcal{A}$ is irreducible and P is strongly connected (see Proposition 3.2). 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 Proposition 3.4) 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. [11]). Finally, the general case for P can be obtained by using a classical block decomposition of stochastic matrices.

In some cases, due to incomplete information on ν or P, one may be interested in providing an estimate of the probabilistic joint spectral radius which is robust with respect to (ν,P). A natural candidate is the quantity $\rho_{\mathrm{p}}(\mathcal{A}) = \sup_{(\nu,P)} \rho_{\mathrm{p}}(\nu,P,\mathcal{A})$, which corresponds to the worst possible average behavior with respect to all possible choices of (ν,P). It is therefore natural to determine also for which N-tuples $\mathcal{A}$ equality holds in the stronger inequality

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

One of our main results (see Theorem 3.10) states that equality holds in (1.3) if and only if there exists 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. It was conjectured in [13] that the finiteness property would hold for every N-tuple $\mathcal{A}$, but this has been disproved in [3,4].

The paper is organized as follows. Section 2 contains the main definitions used in the paper and some elementary properties of the deterministic and probabilistic spectral radii. Section 3 contains our main results, starting in Section 3.1 with the characterization of equality in (1.2) and followed by a geometric point of view on this problem in Section 3.2. We conclude the paper in Section 3.3 by providing a characterization of equality in (1.3).

Acknowledgements. The authors are indebted with D. Chafaï for helpful discussions.

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 $a \leq j \leq b$. 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. 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 invariant subspaces by all A_i 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 [17], is defined by

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

Since all norms in $\mathcal{M}_d(\mathbb{R})$ are equivalent, it immediately follows that $\rho_{\mathrm{d}}(\mathcal{A})$ does not depend on the specific choice of $\|\cdot\|$. Since $\|\cdot\|$ is submultiplicative, one also has 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 $(i_1,\dots,i_n)\in[\![1,N]\!]^n$, one has

$$\rho(A_{i_n}\cdots A_{i_1})^{1/n} \leq \rho_{\mathrm{d}}(\mathcal{A}), \tag{2.1}$$

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

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 $\|\cdot\|_{\mathrm{B}}$ is said to be a *Barabanov norm* for $\mathcal{A}$ if the following two conditions hold.

(a) For every $\sigma\in\mathfrak{S}_N$ and $k\in\mathbb{N}$, $\|A_{\sigma(k)}\cdots A_{\sigma(1)}\|_{\mathrm{B}} \leq \rho_{\mathrm{d}}(\mathcal{A})^k$.

(b) For every $x\in\mathbb{R}^d$ and $k\in\mathbb{N}$, there exists $\sigma\in\mathfrak{S}_N$ 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 a block upper 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 $p_{i_1i_2}p_{i_2i_3}\cdots p_{i_{k-1}i_k} > 0$. 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 $p_{i_1i_2}p_{i_2i_3}\cdots p_{i_{k-1}i_k}p_{i_ki_1} > 0$. 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 is strongly connected. In the stochastic processes literature, one more often uses *irreducibility* to refer to strong connectedness of P. We choose to stick with the latter to avoid ambiguities with the homonymous notion for N-tuples of matrices.

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 [18, §§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*

$$P = \begin{pmatrix} P_1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & P_2 & 0 & \cdots & 0 & 0 \\ \vdots & 0 & \ddots & \ddots & \vdots & \vdots \\ 0 & \vdots & \ddots & \ddots & 0 & 0 \\ 0 & 0 & \cdots & 0 & P_R & 0 \\ * & * & \cdots & * & * & Q \end{pmatrix}, \tag{2.2}$$

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 (2.2). Then every invariant probability $\nu \in \mathbb{R}^N$ can be uniquely decomposed as

$$\nu = \sum_{i=1}^{R} \alpha_i \nu^{[i]}, \tag{2.3}$$

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

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 (2.2). 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 (2.3), $\nu_i = 0$ 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 (2.2) that $i_2 \in \mathcal{I}_j$. One gets the conclusion by an immediate inductive argument. □

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 vertex 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

$$\rho_{\mathrm{p}}(\nu,P,\mathcal{A}) = \limsup_{n\to\infty} \mathbb{E}_{(\nu,P)}\left[\|A_{i_n}\cdots A_{i_1}\|^{1/n}\right], \tag{2.4}$$

where

$$\mathbb{E}_{(\nu,P)}\left[\|A_{i_n}\cdots A_{i_1}\|^{1/n}\right] = \sum_{(i_1,\dots,i_n)\in[\![1,N]\!]^n} \nu_{i_1} p_{i_1 i_2}\cdots p_{i_{n-1} i_n} \|A_{i_n}\cdots A_{i_1}\|^{1/n}. \tag{2.5}$$

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

$$\rho_{\mathrm{p}}(\nu,P,\mathcal{A}) = \lim_{n\to\infty} \mathbb{E}_{(\nu,P)}\left[\|A_{i_n}\cdots A_{i_1}\|^{1/n}\right] = \inf_{n\in\mathbb{N}} \mathbb{E}_{(\nu,P)}\left[\|A_{i_n}\cdots A_{i_1}\|^{1/n}\right]. \tag{2.6}$$

Remark 2.9. It follows from the definition of (ν,P)-word that one can restrict the summation in (2.5) to (ν,P)-words of length n.

Remark 2.10. The deterministic joint spectral radius $\rho_{\mathrm{d}}(\mathcal{A})$ provides the worst asymptotic behavior of $\Sigma(\mathcal{A})$ with respect to $\sigma \in \mathfrak{S}_N$. By introducing the probability measure $\mathbb{P}_{(\nu,P)}$ on $\mathfrak{S}_N$ associated canonically with the transition matrix P and the invariant probability ν, one can interpret $\rho_{\mathrm{p}}(\nu,P,\mathcal{A})$ defined in (2.4) as an asymptotic behavior averaged by $\mathbb{P}_{(\nu,P)}$. Thanks to a classical result by Furstenberg and Kesten [10], under some additional assumptions on (ν,P), one has the stronger interpretation of $\rho_{\mathrm{p}}(\nu,P,\mathcal{A})$ as the $\mathbb{P}_{(\nu,P)}$-almost sure asymptotic behavior of $\Sigma(\mathcal{A})$. More precisely, if, in the decompositions (2.2) and (2.3) in Proposition 2.6, $\nu = \nu^{[i]}$ for some $i \in [\![1,R]\!]$, then the main result of [10] implies that, for $\mathbb{P}_{(\nu,P)}$-almost every $\sigma \in \mathfrak{S}_N$,

$$\rho_{\mathrm{p}}(\nu,P,\mathcal{A}) = \lim_{n\to\infty} \left\| A_{\sigma(n)} \cdots A_{\sigma(1)} \right\|^{1/n}.$$

Notice that the above assumption is satisfied when P is strongly connected.

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

$$\rho_{\mathrm{p}}(\nu,P,\mathcal{A}) \leq \sup_{\nu'} \rho_{\mathrm{p}}(\nu',P,\mathcal{A}) \leq \sup_{(\nu',P')} \rho_{\mathrm{p}}(\nu',P',\mathcal{A}) \leq \rho_{\mathrm{d}}(\mathcal{A}), \tag{2.7}$$

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

$$\rho_{\mathrm{p}}(P,\mathcal{A}) = \sup_{\nu'} \rho_{\mathrm{p}}(\nu',P,\mathcal{A}), \qquad \rho_{\mathrm{p}}(\mathcal{A}) = \sup_{(\nu',P')} \rho_{\mathrm{p}}(\nu',P',\mathcal{A}). \tag{2.8}$$

Remark 2.11. It follows from (2.6) that $(\nu',P') \mapsto \rho_{\mathrm{p}}(\nu',P',\mathcal{A})$ is upper semicontinuous. Moreover, the set of pairs (ν',P') made of an $N \times N$ stochastic matrix P' and an invariant probability ν' for P' is compact. As a consequence, one can replace the suprema in (2.8) 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 proof of 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 (Proposition 3.2). We then obtain the conclusion under the sole additional assumption that P is strongly connected (Proposition 3.4). Finally, we consider the general case in the third step.

Proposition 3.2. *Let $P \in \mathcal{M}_d(\mathbb{R})$ be a stochastic strongly connected matrix, $\mathcal{A} = (A_1, \dots, A_N) \in \mathcal{M}_d(\mathbb{R})^N$ 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) $\|A_{i_k} \cdots A_{i_1}\|_{\mathrm{B}}^{1/k} = \rho_{\mathrm{d}}(\mathcal{A})$ *for every P-word $(i_1, \dots, i_k)$.*

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

Proof. The argument goes as follows. We first prove that (a) is equivalent to (b) and then that (b) is equivalent to (c). By Remark 2.5, P admits a unique invariant probability ν, and hence $\rho_{\mathrm{p}}(P, \mathcal{A}) = \rho_{\mathrm{p}}(\nu, P, \mathcal{A})$.

It is immediate that (b) implies (a) thanks to (2.4), (2.5), and Remark 2.9. We prove the converse by contrapositive. Notice that, by definition of Barabanov norm, one has $\|A_{i_n} \cdots A_{i_1}\|_{\mathrm{B}} \leq \rho_{\mathrm{d}}(\mathcal{A})^n$ for every $n \in \mathbb{N}$ and every P-word $(i_1, \dots, i_n)$. Assuming that (b) does not hold, there exist $n \in \mathbb{N}$ and a P-word $(i_1, \dots, i_n)$ such that

$$\|A_{i_n} \cdots A_{i_1}\|_{\mathrm{B}}^{1/n} < \rho_{\mathrm{d}}(\mathcal{A}).$$

Then, by (2.5) and Remark 2.5,

$$\mathbb{E}_{(\nu,P)}\left[\|A_{i_n} \cdots A_{i_1}\|_{\mathrm{B}}^{1/n}\right] < \rho_{\mathrm{d}}(\mathcal{A}),$$

which implies that $\rho_{\mathrm{p}}(P, \mathcal{A}) < \rho_{\mathrm{d}}(\mathcal{A})$ according to (2.6). This concludes the proof of the equivalence between (a) and (b).

We now establish the equivalence between (b) and (c). Assume first that (b) holds. If $(i_1, \dots, i_k)$ is a P-cycle, we define $\sigma \in \mathfrak{S}_N$ by setting $\sigma(j) = i_\ell$ when $j \equiv \ell \mod k$. Then $(\sigma(1), \dots, \sigma(rk))$ is a P-word of length rk for every $r \in \mathbb{N}$ and, by hypothesis, one has

$$\|(A_{i_k} \cdots A_{i_1})^r\|_{\mathrm{B}}^{1/rk} = \rho_{\mathrm{d}}(\mathcal{A}).$$

Letting $r \to \infty$ and using Gelfand's formula, one concludes that

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

which yields (c).

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

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

Since the spectral radius is a lower bound for any norm of a matrix, one obtains that

$$\|A_{i_r} \cdots A_{i_1}\|_{\mathrm{B}} \geq \rho_{\mathrm{d}}(\mathcal{A})^r.$$

Then

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

Using the fact that $\|\cdot\|_{\mathrm{B}}$ is a Barabanov norm, one also has 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 two previous inequalities, it follows that $\|A_{i_k} \cdots A_{i_1}\|_{\mathrm{B}} = \rho_{\mathrm{d}}(\mathcal{A})^k$. □

Remark 3.3. The proof of Proposition 3.2 only uses that $\|\cdot\|_{\mathrm{B}}$ is an extremal norm, i.e., it satisfies (a) in Definition 2.1. One could then replace the irreducibility assumption on $\mathcal{A}$ by its nondefectiveness (we refer the reader to [12, Section 2.1.2] for details). However, we prefer to state Proposition 3.2 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 Proposition 3.4.

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 (b) from Proposition 3.2 cannot be expected.

Proposition 3.4. *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

$$A_j = \begin{pmatrix} A_j^{(1)} & * & * & \cdots & * \\ 0 & A_j^{(2)} & * & \cdots & * \\ 0 & 0 & A_j^{(3)} & \cdots & * \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & A_j^{(R)} \end{pmatrix}, \qquad j \in [\![1, N]\!], \tag{3.1}$$

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 [12, Proposition 1.5], one has $\rho_{\mathrm{d}}(\mathcal{A}) = \max_{r \in [\![1,R]\!]} \rho_{\mathrm{d}}(\mathcal{A}^{(r)})$ and, on the other hand, it follows from [11, 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 $(i_1, \dots, i_k)$, one has

$$\rho_{\mathrm{d}}(\mathcal{A}) \geq \rho(A_{i_k} \cdots A_{i_1})^{1/k} = \max_{r \in [\![1,R]\!]} \rho\left(A_{i_k}^{(r)} \cdots A_{i_1}^{(r)}\right)^{1/k}, \tag{3.2}$$

where the inequality comes from (2.1) 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.

We start by proving that (a) implies (b). Let $\bar{r} \in [\![1, N]\!]$ be such that $\rho_{\mathrm{p}}(P, \mathcal{A}^{(\bar{r})}) = \max_{r \in [\![1,N]\!]} \rho_{\mathrm{p}}(P, \mathcal{A}^{(r)}) = \rho_{\mathrm{p}}(P, \mathcal{A})$. Then, by (a) and the previous remarks, one has

$$\rho_{\mathrm{p}}(P, \mathcal{A}^{(\bar{r})}) = \rho_{\mathrm{d}}(\mathcal{A}) = \max_{r \in [\![1,N]\!]} \rho_{\mathrm{d}}(\mathcal{A}^{(r)}) \geq \rho_{\mathrm{d}}(\mathcal{A}^{(\bar{r})}),$$

and one obtains from (2.7) that $\rho_{\mathrm{p}}(P, \mathcal{A}^{(\bar{r})}) = \rho_{\mathrm{d}}(\mathcal{A}^{(\bar{r})}) = \rho_{\mathrm{d}}(\mathcal{A})$. By Proposition 3.2, for every P-cycle $(i_1, \dots, i_k)$, one has, using (3.2),

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

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

Assume now that (b) holds. Then (a) holds trivially if $\rho_{\mathrm{d}}(\mathcal{A}) = 0$. Otherwise, one can assume, with no loss of generality, that $\rho_{\mathrm{d}}(\mathcal{A}) = 1$ up to replacing $\mathcal{A}$ by $\mathcal{A}/\rho_{\mathrm{d}}(\mathcal{A})$. By assumption and (3.2), for every P-cycle $(i_1,\dots,i_k)$, there exists $r \in [\![1,R]\!]$ such that $\rho\left(A^{(r)}_{i_k}\cdots A^{(r)}_{i_1}\right) = 1$.

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 $i^r = (i^r_1,\dots,i^r_{\ell_r})$ such that $\rho(A(i^r)) < 1$. Let $j^r = (j^r_1,\dots,j^r_{k_r})$ be a P-word such that $j^r_1 = i^r_1$ and $p_{j^r_{k_r} i^{r+1}_1} > 0$ (with the convention that $i^{R+1}_1 = i^1_1$). 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 (3.2) that there exists $r_n \in [\![1,R]\!]$ such that

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

Pick $\bar r \in [\![1,R]\!]$ and an increasing sequence $(n_q)_{q\in\mathbb{N}}$ such that $r_{n_q} = \bar r$ 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 $q \in \mathbb{N}$, one has

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

where the last inequality follows from the fact that $\|\cdot\|_{\bar r}$ is a Barabanov norm. Since $\rho(A^{(\bar r)}(i^{\bar r})) \le \rho(A(i^{\bar r})) < 1$, one has that $\Big\|A^{(\bar r)}(i^{\bar r})^{n_q}\Big\|_{\bar r} \xrightarrow[q\to\infty]{} 0$, 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 (2.1), one has $\rho\left(A^{(\bar r)}_{i_k}\cdots A^{(\bar r)}_{i_1}\right) \le \rho_{\mathrm{d}}(\mathcal{A}^{(\bar r)})$. Since $\rho_{\mathrm{d}}(\mathcal{A}^{(\bar r)}) \le \rho_{\mathrm{d}}(\mathcal{A})$, one deduces 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 Proposition 3.2, one obtains that

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

and then (a) holds thanks to (2.7). □

We can conclude now the proof of Theorem 3.1.

Proof of Theorem 3.1. Before proving the equivalence between (a) and (b), we first decompose P and ν according to Proposition 2.6. Using the notations of (2.2) and (2.3), notice that, thanks to (2.4) and (2.5), one has

$$\rho_{\mathrm{p}}(\nu, P, \mathcal{A}) = \sum_{j=1}^{R} \alpha_j \rho_{\mathrm{p}}(\nu^{[j]}, P, \mathcal{A}). \tag{3.3}$$

For $j \in [\![1, R]\!]$, let $\mathcal{A}^{[j]}$ be the ordered n_j-tuple made of the matrices A_ℓ such that $\nu_\ell^{[j]} > 0$. 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 (2.7) and the fact that $\mathcal{A}^{[j]}$ is made of matrices from $\mathcal{A}$, one obtains that, for every $j \in [\![1, R]\!]$,

$$\rho_{\mathrm{p}}(\nu^{[j]}, P_j, \mathcal{A}^{[j]}) \leq \rho_{\mathrm{d}}(\mathcal{A}^{[j]}) \leq \rho_{\mathrm{d}}(\mathcal{A}). \tag{3.4}$$

We now prove that (a) implies (b). Using (3.3), (3.4), and (a), one deduces that

$$\rho_{\mathrm{p}}(\nu^{[j]}, P_j, \mathcal{A}^{[j]}) = \rho_{\mathrm{d}}(\mathcal{A}^{[j]}) = \rho_{\mathrm{d}}(\mathcal{A}) \qquad \text{for every } j \in [\![1, R]\!] \text{ such that } \alpha_j > 0.$$

Let $(i_1, \dots, i_k)$ be a (ν, P)-cycle. Then, letting $\mathcal{I}_1, \dots, \mathcal{I}_R$ be as in the statement of Lemma 2.7, one obtains that $i_1, \dots, i_k$ are in $\mathcal{I}_\ell$ for some $\ell \in [\![1, R]\!]$. Moreover, since $\nu_{i_1} > 0$, one obtains that $\alpha_\ell > 0$. Hence, since $\rho_{\mathrm{p}}(\nu^{[\ell]}, P_\ell, \mathcal{A}^{[\ell]}) = \rho_{\mathrm{d}}(\mathcal{A}^{[\ell]})$, one concludes from Proposition 3.4 that $\rho(A_{i_k} \cdots A_{i_1})^{1/k} = \rho_{\mathrm{d}}(\mathcal{A})$, as required.

We finally prove that (b) implies (a). Let $j \in [\![1, R]\!]$ be such that $\alpha_j > 0$ and take a (ν, P)-cycle $(i_1, \dots, i_k)$ with $i_1, \dots, i_k$ in $\mathcal{I}_j$. Then, by (2.1), (3.4), and (b), one has

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

In particular, $\rho_{\mathrm{d}}(\mathcal{A}) = \rho_{\mathrm{d}}(\mathcal{A}^{[j]})$ and $\rho(A_{i_k} \cdots A_{i_1})^{1/k} = \rho_{\mathrm{d}}(\mathcal{A}^{[j]})$. Applying Proposition 3.4 to P_j and $\mathcal{A}^{[j]}$, one then gets 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 (3.3) that $\rho_{\mathrm{p}}(\nu, P, \mathcal{A}) = \rho_{\mathrm{d}}(\mathcal{A})$, as required. □

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 [16], where the authors classify matrix semigroups of constant spectral radius. We start with the following proposition.

Proposition 3.5. *Let $P \in \mathcal{M}_N(\mathbb{R})$ be a stochastic strongly connected matrix and $\mathcal{A} = (A_1, \dots, A_N) \in \mathcal{M}_d(\mathbb{R})^N$ 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 (2.1), $\rho_{\mathrm{d}}(\mathcal{A}) \geq \rho(A(i_*))^{1/k_*} > 0$, where k_* is the length of i_*. From Proposition 3.4, the set

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

is a matrix semigroup with constant spectral radius. Since, moreover, the latter is also irreducible, the conclusion follows from [16, Theorem 2]. □

Remark 3.6. As remarked in [16], the problem of classifying matrix semigroups with constant spectral radius is highly nontrivial when the semigroup contains singular matrices. By using additional results from [16], one may obtain, under the assumptions of Proposition 3.5, 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 [16, Theorem 3 and Corollary 6].

A limitation of Proposition 3.5 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 a vertex 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 $\mathcal{A}$ is made of invertible matrices and that all diagonal elements of P are positive, in which case one has the following proposition.

Proposition 3.7. *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 and made of invertible matrices. 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

$$A_{i_j} = \left(A_{i_k} \cdots A_{i_{j+1}}\right)^{-1} \left(A_{i_k} \cdots A_{i_{j+1}} A_{i_j}^2 A_{i_{j-1}} \cdots A_{i_1}\right) \left(A_{i_k} \cdots A_{i_1}\right)^{-1} \left(A_{i_k} \cdots A_{i_{j+1}}\right) \tag{3.5}$$

allows one to deduce 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 $\mathcal{A}$ is made of 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 3.5 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 Proposition 3.4. □

Remark 3.8. Notice that, to obtain the second part of the conclusion of Proposition 3.7, 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 $j \in [\![2,k]\!]$, one 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.

Remark 3.9. 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 3.8, equality occurs if and only if there exists an invertible matrix $G \in \mathcal{M}_d(\mathbb{R})$ such that $\rho_{\mathrm{d}}(\mathcal{A})^{-1}GA_1G^{-1}$ and $\rho_{\mathrm{d}}(\mathcal{A})^{-1}GA_2G^{-1}$ 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 3.10. *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 $i_1,\dots,i_k \in [\![1,N]\!]$ pairwise distinct such that*

$$\rho_{\mathrm{d}}(\mathcal{A}) = \rho(A_{i_k}\cdots A_{i_1})^{1/k}. \tag{3.6}$$

Proof. We start by proving that (a) implies (b). Recall that, by Remark 2.11, there exist a stochastic matrix P and an invariant probability ν for P such that $\rho_{\rm p}(\nu,P,\mathcal{A}) = \rho_{\rm p}(\mathcal{A})$. Using (a), one deduces that $\rho_{\rm p}(\nu,P,\mathcal{A}) = \rho_{\rm d}(\mathcal{A})$. It is clear that there exists a (ν,P)-cycle $(i_1,\dots,i_k)$ such that $i_1,\dots,i_k$ 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 $(i_1,\dots,i_k)$ and their powers. Moreover, for every such (ν,P)-cycle $(j_1,\dots,j_s)$, one has

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

Indeed, this follows from the fact that $\rho(M_1M_2) = \rho(M_2M_1)$ for every $M_1,M_2 \in \mathcal{M}_d(\mathbb{R})$. Then Theorem 3.1(b) holds, hence $\rho_{\rm p}(\nu,P,\mathcal{A}) = \rho_{\rm d}(\mathcal{A})$. One concludes by (2.7). □

Remark 3.11. It follows from (2.7) that, if $\rho_{\rm d}(\mathcal{A}) > 0$, the ratio $\frac{\rho_{\rm p}(\mathcal{A})}{\rho_{\rm d}(\mathcal{A})}$ belongs to $[0,1]$ and Theorem 3.10 addresses the case where it is equal to 1. We provide next an example where it is equal to 0, proving that one cannot expect a positive lower bound for this ratio, uniformly with respect to $\mathcal{A}$.

Consider

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

One computes

$$A_1^2A_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, A_1A_2A_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, A_2A_1^2 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

and $A_1^3 = A_1A_2^2 = A_2A_1A_2 = A_2^2A_1 = A_2^3 = 0$. Let $\|\cdot\|_1$ denote the matrix norm induced by the ℓ^1 norm in $\mathbb{R}^3$. Define

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

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, one gets that, for every $k \in \mathbb{N}$ and $(i_1,\dots,i_k) \in [\![1,N]\!]^k$,

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

One then obtains that $\rho_{\rm d}(\mathcal{A}) = 1$. On the other hand, for every stochastic matrix $P \in \mathcal{M}_2(\mathbb{R})$ and every invariant probability vector ν for P, one has $\mathbb{P}_{(\nu,P)}(\mathcal{E}) = 0$. 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 $\rho_{\rm p}(\nu,P,\mathcal{A}) = 0$. Then $\rho_{\rm p}(\mathcal{A}) = 0$.

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

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

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

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

[13] 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.

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

[15] 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.

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

[17] 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.

[18] 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].

[19] 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.

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