



**HAL**  
open science

# On the gap between deterministic and probabilistic Lyapunov exponents for continuous-time linear systems

Yacine Chitour, Guilherme Mazanti, Pierre Monmarché, Mario Sigalotti

## ► To cite this version:

Yacine Chitour, Guilherme Mazanti, Pierre Monmarché, Mario Sigalotti. On the gap between deterministic and probabilistic Lyapunov exponents for continuous-time linear systems. *Electronic Journal of Probability*, 2023, 28, Paper No. 43, 39 pp. 10.1214/23-EJP932 . hal-03478271v2

**HAL Id: hal-03478271**

**<https://hal.science/hal-03478271v2>**

Submitted on 21 Nov 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# On the gap between deterministic and probabilistic Lyapunov exponents for continuous-time linear systems

Yacine Chitour\*    Guilherme Mazanti†    Pierre Monmarché‡  
Mario Sigalotti§

November 21, 2022

## Abstract

Consider a non-autonomous continuous-time linear system in which the time-dependent matrix determining the dynamics is piecewise constant and takes finitely many values  $A_1, \dots, A_N$ . This paper studies the equality cases between the maximal Lyapunov exponent associated with the set of matrices  $\{A_1, \dots, A_N\}$ , on the one hand, and the corresponding ones for piecewise deterministic Markov processes with modes  $A_1, \dots, A_N$ , on the other hand. A fundamental step in this study consists in establishing a result of independent interest, namely, that any sequence of Markov processes associated with the matrices  $A_1, \dots, A_N$  converges, up to extracting a subsequence, to a Markov process associated with a suitable convex combination of those matrices.

**Keywords.** Linear switched systems, continuous-time Markov processes, piecewise deterministic Markov processes, convexified Markov processes, Lyapunov exponents  
**2020 Mathematics Subject Classification.** 60J25, 34A38, 34D08

## Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Definitions, notations, and basic facts</b>	<b>5</b>
2.1	Deterministic Lyapunov exponent . . . . .	5
2.2	Continuous-time Markov processes . . . . .	6
2.3	Probabilistic Lyapunov exponent . . . . .	7

---

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

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

‡Sorbonne Université, CNRS, Laboratoire Jacques-Louis Lions (LJLL), Laboratoire de Chimie Théorique (LCT), F-75005 Paris, France

§Sorbonne Université, Inria, CNRS, Laboratoire Jacques-Louis Lions (LJLL), F-75005 Paris, France

<b>3</b>	<b>Statements of the main results</b>	<b>9</b>
<b>4</b>	<b>Characterization of equality between <math>\lambda_d(A)</math> and <math>\lambda_p(v, \mu, P, A)</math></b>	<b>10</b>
<b>5</b>	<b>Compactification of the space of Markov processes</b>	<b>14</b>
5.1	Convexified Markov processes as limits of Markov processes . . . . .	14
5.2	Convexified Markov processes compactify the space of Markov processes	17
<b>6</b>	<b>On the equality between <math>\lambda_d(A)</math> and <math>\lambda_p^{\text{sup}}(A)</math></b>	<b>31</b>
<b>A</b>	<b>Decomposition of a Markov chain</b>	<b>34</b>
<b>B</b>	<b>Technical results</b>	<b>38</b>
B.1	Proof of Proposition 6.5 . . . . .	38
B.2	Proof of Lemma 6.7 . . . . .	42

# 1 Introduction

In this paper, we consider the family of non-autonomous continuous-time linear systems

$$\dot{x}(t) = A_{\sigma(t)}x(t), \tag{1.1}$$

where  $x(\cdot)$  takes values in  $\mathbb{R}^d$ ,  $\sigma(\cdot)$  is piecewise constant and takes values in the finite set of indices  $\{1, \dots, N\}$ , and we set  $A$  to be the  $N$ -tuple made of  $d \times d$  matrices with real coefficients  $A_1, \dots, A_N$  (also called the *modes* of (1.1)). Each *signal*  $\sigma$  corresponds to a possible evolution in time of a discrete parameter affecting the dynamics. This class of systems can be used to describe phenomena where different dynamical modes operate and the order in which they are active is not precisely known. In the engineering literature, such systems and their discrete-time counterparts bear the name of *switched systems* [20] and they have been widely studied in the mathematical community since [8].

One of the major issues regarding these systems concerns their asymptotic stability, uniformly with respect to the signal  $\sigma$ . Indeed, the fact that each individual mode is asymptotically stable does not imply that the trajectories of the corresponding switched system converge to 0: it is easy to find two positive times  $t_1, t_2$  and two matrices  $A_1, A_2$  whose eigenvalues have negative real part such that the spectral radius of  $e^{A_1 t_1} e^{A_2 t_2}$  is larger than 1, as illustrated, for instance, in [20]. The measure of stability of a switching system with respect to *all* possible signals  $\sigma$  is characterized by its *deterministic maximal Lyapunov exponent*  $\lambda_d(A)$ , measuring the maximal asymptotic exponential rate of (1.1) (see (2.1) below).

A difficulty in dealing with  $\lambda_d(A)$  is that the characterization of the asymptotic behavior of (1.1) based on its value is in general conservative from a practical viewpoint, since it corresponds to a maximization with respect to all possible signals. In the (few) cases where maximizing signals (or maximizing sequences of signals) are known, they happen to have a very specific structure, for instance switching between modes at precise times or at a fast rate [1]. In many situations, one disposes of additional information on the signal  $\sigma$  implying that such specific structures occur rarely, at least in a probabilistic sense. This motivates addressing the measure of stability of switched systems within a probabilistic

framework, a question which has been considered in the literature, for instance in [4] for systems in dimension  $d = 2$ . One may naturally expect that, except for some very particular situations, such a probabilistic framework gives rise to less conservative measures of stability of (1.1), as it is indeed observed for the two-dimensional systems considered in [4].

An important class of switched systems with random switching is that of piecewise deterministic Markov processes (PDMPs) introduced in [9], which provide a natural modeling framework for phenomena of random switching without memory and corresponds to considering  $\sigma$  as a continuous-time Markov process. In that case, the asymptotic behavior of (1.1) can be studied through a *probabilistic Lyapunov exponent*  $\lambda_p(\nu, \mu, P, A)$ , which consists of the expected value with respect to  $\sigma$  of the asymptotic exponential rate of (1.1), where  $(\nu, \mu, P)$  are the parameters of the Markov process as described in Section 2.3 below. Recall that, due to the classical result in [14], under generic assumptions on the Markov process, the asymptotic exponential rate for a given  $\sigma$  is equal to  $\lambda_p(\nu, \mu, P, A)$  almost surely.

In applications, the quantity  $\lambda_p(\nu, \mu, P, A)$  is a suitable measure of asymptotic behavior if the parameters  $(\nu, \mu, P)$  of the Markov process are fixed. However, it is also important to consider situations in which such parameters are not known exactly, and, for that purpose, we introduce the quantity  $\lambda_p^{\text{sup}}(A)$  as the supremum of  $\lambda_p(\nu, \mu, P, A)$  with respect to all Markov processes  $(\nu, \mu, P)$ , which corresponds to a worst-case scenario.

Clearly, for every Markov process  $(\nu, \mu, P)$  and every  $N$ -tuple of matrices  $A$ , one has  $\lambda_p(\nu, \mu, P, A) \leq \lambda_d(A)$ , and hence one also has  $\lambda_p^{\text{sup}}(A) \leq \lambda_d(A)$ . The goal of this paper is to investigate under which conditions on  $A$  the probabilistic point of view is strictly less restrictive than the deterministic one, i.e., to characterize in terms of  $A$  the strict inequality (or, equivalently, the equality) between  $\lambda_p(\nu, \mu, P, A)$  and  $\lambda_d(A)$  for a fixed  $(\nu, \mu, P)$  and also between  $\lambda_p^{\text{sup}}(A)$  and  $\lambda_d(A)$ . Hence, this paper is the continuous-time counterpart of [7], in which similar issues have been addressed for discrete-time systems.

Our main results in that sense, Theorems 3.1 and 3.4, show that probabilistic measures of stability of (1.1) are indeed less restrictive than the deterministic ones, except for some particular situations that we characterize. For instance, under additional irreducibility and strong connectedness assumptions, Theorem 3.1 reduces to Proposition 4.1, which states that  $\lambda_p(\nu, \mu, P, A)$  is strictly smaller than  $\lambda_d(A)$  except for the particular situation in which the matrices  $A_1, \dots, A_N$  are skew-symmetric up to a common translation by a multiple of the identity matrix and a common change of basis. Up to a technical assumption which is always satisfied in dimension  $d \leq 3$ , Theorem 3.4 states that the probabilistic measure of stability  $\lambda_p^{\text{sup}}(A)$  is strictly smaller than  $\lambda_d(A)$  except for the particular situation where the worst possible deterministic behavior is attained by a matrix in the convex hull of  $A_1, \dots, A_N$ . In particular, these results confirm that probabilistic measures of stability are most often less conservative than their deterministic counterparts, as previously observed in [4] for two-dimensional systems.

The characterization of equality between  $\lambda_p(\nu, \mu, P, A)$  and  $\lambda_d(A)$  follows essentially the same lines as the corresponding problem in discrete time addressed in [7]. We first consider the case where the  $N$ -tuple  $A$  is irreducible and the matrix  $P$  is strongly connected (as defined in Section 2): it is shown that equality occurs if and only if the matrices  $A_i - \lambda_d(A)\text{Id}$  are skew-symmetric (up to a common change of basis), a result that relies on the use of an extremal norm for  $A$  (Definition 2.1) and the characterization of semigroups

with constant spectral radius from [21]. We then treat the general case by decomposing  $A$  into irreducible blocks and  $P$  into strongly connected blocks.

As regards the question of equality between  $\lambda_p^{\text{sup}}(A)$  and  $\lambda_d(A)$ , we show that it implies that  $\lambda_d(A)$  is equal to the maximum of the real part of the eigenvalues of some matrix  $M$  belonging to the convex hull of  $A$ , and that the converse is true under a technical assumption on  $A = (A_1, \dots, A_N)$  and  $M$ , cf. Definition 6.3, which is always satisfied in dimension  $d \leq 3$ . We conjecture that this technical assumption is not necessary to get the converse implication. The difficulty in removing the technical assumption consists in proving the convergence of the probabilistic Lyapunov exponent when the jump rate goes to  $+\infty$ .

Indeed, our analysis relies on the investigation of the behavior of maximizing sequences of Markov processes for  $\lambda_p^{\text{sup}}(A)$ . In the discrete-time setting considered in [7], the issue is easily handled thanks to the compactness of the space of discrete-time Markov processes. This is not anymore the case in the continuous-time setting, where the situation is much more delicate since switching between modes can occur arbitrarily fast. When the transition matrix  $P$  is fixed and strongly connected and the jump rate  $\mu$  goes to infinity, it is well-known that high-frequency jumps lead to deterministic averaging [4, 5], namely the Markov process converges to a deterministic motion  $\dot{x} = Mx$ , where  $M$  belongs to the convex hull of  $A$ . More generally (if  $P$  is not fixed as  $\mu \rightarrow +\infty$ ), in case of a combination of fast and slow jumps, one can expect the convergence towards a Markov process on convex combinations of matrices of  $A$ .

We rigorously handle such a decomposition of Markov processes in different time-scales by relying on results from various works by Landim and collaborators, in particular [19], although these works are primarily interested in metastability phenomena, i.e., Markov chains for which the different timescales are all slow, instead of fast as in our case. As a consequence, we prove that we can extract from any sequence of Markov processes with modes in  $A$  a subsequence that converges in law to a Markov process associated with a suitable convex combinations of the original matrices in  $A$ . This compactification result, Theorem 3.3, is one of our main results.

Since the convergence in law obtained in Theorem 3.3 is not uniform in time, it is not sufficient to deduce convergence of the Lyapunov exponents of the sequence of Markov processes to the Lyapunov exponent of the limit process. Such a convergence property is interesting in itself and has already been addressed in particular cases, e.g. [4, Section 2.5] (where  $N = d = 2$  and the matrices are Hurwitz) or [5, Corollary 2.15] (where the matrix  $M$  appearing in the limit is Metzler and strongly connected). We prove in Proposition 6.4 a result in that direction under the already mentioned additional technical assumption on  $A$  and  $M$ . Establishing such a result unconditionally would provide a complete characterization of the equality between  $\lambda_d(A)$  and  $\lambda_p^{\text{sup}}(A)$ .

As a conclusion, our results show that the equality between the deterministic maximal Lyapunov exponent and the probabilistic ones (either for a fixed Markov chain or for the worst probabilistic case) only occurs in very specific cases. This shows that, in most cases, working within a probabilistic framework yields a less conservative estimate, closer to the stability properties most commonly observed in practice.

The paper is organized as follows. Section 2 collects definitions, notations, and basic facts relative to deterministic and probabilistic Lyapunov exponents as well as continuous-time Markov processes. The statements of the main results proved in the paper are

presented in Section 3. Section 4 addresses the characterization of equality between  $\lambda_p(v, \mu, P, A)$  and  $\lambda_d(A)$ . We describe in Section 5 the compactification of the space of Markov processes, which is used in Section 6 to study the case of equality between  $\lambda_p^{\text{sup}}(A)$  and  $\lambda_d(A)$ . The paper is completed by two appendices. In Appendix A, we prove a general result of decomposition into different timescales for a sequence of Markov chains on a finite state space, adapted from [19], which is a central tool in the proofs of Section 5. Appendix B provides the proofs of some linear-algebraic technical results used in Section 6.

## 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 integers,  $\llbracket a, b \rrbracket$  denotes the set of integers  $j$  such that  $a \leq j \leq b$ . We use  $|\cdot|$  to denote a norm in  $\mathbb{R}^d$  and  $\|\cdot\|$  to denote the corresponding induced norm on the space  $\mathcal{M}_d(\mathbb{R})$  of  $d \times d$  matrices with real coefficients. The identity matrix in  $\mathcal{M}_d(\mathbb{R})$  is denoted by  $\text{Id}$ . The spectral radius of a square matrix  $M$  is denoted by  $\rho(M)$ , and its spectral abscissa, defined as the maximum of the real parts of its eigenvalues, is denoted by  $\lambda(M)$ . An  $N$ -tuple  $A = (A_1, \dots, A_N) \in \mathcal{M}_d(\mathbb{R})^N$  is said to be *irreducible* if the only invariant subspaces by all  $A_i$  are  $\{0\}$  and  $\mathbb{R}^d$ . If  $\mathcal{A} \subset \mathcal{M}_d(\mathbb{R})$ , we use  $\text{co}(\mathcal{A})$  to denote the convex hull of  $\mathcal{A}$ .

Denote by  $\Sigma$  the set of all piecewise constant right-continuous functions defined on  $[0, \infty)$  and taking values in  $\llbracket 1, N \rrbracket$ . Given  $\sigma \in \Sigma$ , we use  $t \mapsto \Phi_\sigma(t)$  to denote the flow of

$$\dot{x} = A_{\sigma(t)}x$$

with  $\Phi_\sigma(0) = \text{Id}$ . In particular,

$$\Phi_\sigma(t_n) = e^{A_{\sigma(t_{n-1})}(t_n - t_{n-1})} \dots e^{A_{\sigma(0)}t_1},$$

where  $(t_i)_{i \in \mathbb{N}}$  is an increasing sequence containing all discontinuity times of  $\sigma$ .

### 2.1 Deterministic Lyapunov exponent

Let  $A = (A_1, \dots, A_N) \in \mathcal{M}_d(\mathbb{R})^N$ . The *deterministic Lyapunov exponent*  $\lambda_d(A)$  associated with  $A$  is defined as

$$\lambda_d(A) = \limsup_{t \rightarrow \infty} \frac{1}{t} \sup_{\sigma \in \Sigma} \log \|\Phi_\sigma(t)\|. \quad (2.1)$$

Since all norms in  $\mathbb{R}^d$  are equivalent, it immediately follows that  $\lambda_d(A)$  does not depend on the specific choice of  $\|\cdot\|$ . It turns out (see, e.g., [16, Lemma 1.2]) that, since  $\|\cdot\|$  is submultiplicative on  $\mathcal{M}_d(\mathbb{R})$ , one has

$$\lambda_d(A) = \lim_{t \rightarrow \infty} \frac{1}{t} \sup_{\sigma \in \Sigma} \log \|\Phi_\sigma(t)\| = \inf_{t > 0} \frac{1}{t} \sup_{\sigma \in \Sigma} \log \|\Phi_\sigma(t)\|. \quad (2.2)$$

Moreover, for every  $\sigma \in \Sigma$  and  $t > 0$ , one has

$$\frac{1}{t} \log \rho(\Phi_\sigma(t)) \leq \lambda_d(A). \quad (2.3)$$

Indeed, let  $\hat{\sigma} \in \Sigma$  be the  $t$ -periodic signal coinciding with  $\sigma$  on the interval  $[0, t)$ . Then, for every  $k \in \mathbb{N}$ ,

$$\begin{aligned} \frac{1}{t} \log \rho(\Phi_\sigma(t)) &= \frac{1}{kt} \log \rho(\Phi_{\hat{\sigma}}(kt)) \leq \frac{1}{kt} \log \|\Phi_{\hat{\sigma}}(kt)\| \\ &\leq \frac{1}{kt} \sup_{\zeta \in \Sigma} \log \|\Phi_\zeta(kt)\|, \end{aligned}$$

and we conclude from (2.1) by taking the lim sup as  $k \rightarrow +\infty$ .

**Definition 2.1** (Extremal norm). Let  $A = (A_1, \dots, A_N) \in \mathcal{M}_d(\mathbb{R})^N$ . A norm  $\|\cdot\|_e$  in  $\mathcal{M}_d(\mathbb{R})$  is said to be *extremal for A* if, for every  $\sigma \in \Sigma$  and  $t \geq 0$ , it holds  $\|\Phi_\sigma(t)\|_e \leq e^{\lambda_d(A)t}$ .

**Remark 2.2.** A necessary and sufficient condition for the existence of an extremal norm for a given  $A = (A_1, \dots, A_N)$  is the *nondefectiveness* of  $A$ , i.e., the existence of  $C > 0$  such that  $\|\Phi_\sigma(t)\| \leq Ce^{\lambda_d(A)t}$  for every  $t \geq 0$  and  $\sigma \in \Sigma$  (see, e.g., [16, Theorem 2.2] for the discrete-time case, which extends readily to the continuous-time setting).

Note that, since the computation of  $\lambda_d(A)$  is intractable in general (cf. [16]), nondefectiveness turns out to be also difficult to check. This motivates the search for simpler conditions implying the nondefectiveness of a family of matrices  $A$ . One such condition is the irreducibility of  $A$  (see, e.g., [24], where it is shown that irreducibility actually implies the existence of a so-called Barabanov norm, which is an extremal norm satisfying some additional properties).

## 2.2 Continuous-time Markov processes

In this paper, we consider continuous-time Markov processes in  $\llbracket 1, N \rrbracket$  defined by triples  $(\mathbf{v}, \mu, P)$ , where  $P = (p_{ij})_{i,j=1}^N \in \mathcal{M}_N(\mathbb{R})$  is a stochastic matrix,  $\mu > 0$ , and  $\mathbf{v} \in \mathbb{R}^N$  is a probability vector, seen as a row vector, i.e., as a  $1 \times N$  matrix. The Markov process corresponding to  $(\mathbf{v}, \mu, P)$ , denoted by  $\sigma$ , is the continuous-time Markov chain on  $\llbracket 1, N \rrbracket$  with initial law  $\mathbf{v}$ , transition matrix  $P$ , and jump rate  $\mu$ . Hence, if  $\sigma(t) = i \in \llbracket 1, N \rrbracket$  and  $t'$  is the next jump time, then  $t' - t$  follows an exponential law of parameter  $\mu$  and  $\sigma(t') = j \in \llbracket 1, N \rrbracket$  with probability  $p_{ij}$ . Note that trivial jumps (i.e., from a state to itself) are allowed and that, for every  $\alpha \in (0, 1]$ , both triples of parameters  $(\mathbf{v}, \mu/\alpha, \text{Id} + \alpha(P - \text{Id}))$  and  $(\mathbf{v}, \mu, P)$  determine Markov processes with the same law.

When  $\sigma(\cdot)$  and  $\mathbf{v}$  are, respectively, a Markov chain and a probability vector on  $\llbracket 1, N \rrbracket$ , we occasionally denote by  $\mathbb{P}_\mathbf{v}$  and  $\mathbb{E}_\mathbf{v}$  probabilities and expectations to indicate that the law of the initial condition  $\sigma(0)$  is  $\mathbf{v}$ . If  $\mathbf{v} = \delta_i$  for some  $i \in \llbracket 1, N \rrbracket$ , we simply write  $\mathbb{P}_i$  and  $\mathbb{E}_i$ .

Given  $A = (A_1, \dots, A_N) \in \mathcal{M}_d(\mathbb{R})^N$  and  $x_0 \in \mathbb{R}^d$ , the above Markov process  $\sigma$  in  $\llbracket 1, N \rrbracket$  induces the stochastic processes  $A_{\sigma(\cdot)}$  in  $\mathcal{M}_d(\mathbb{R})$  and  $\Phi_\sigma(\cdot)x_0$  in  $\mathbb{R}^d$ . When clear from the context, we still identify such processes with the triple  $(\mathbf{v}, \mu, P)$ . The matrices  $A_1, \dots, A_N$  are called the *modes* of the Markov process and, for distinct  $i, j \in \llbracket 1, N \rrbracket$ ,  $\lambda(i, j) = \mu p_{ij}$  is the *jump rate from i to j*. Notice that, although  $x(\cdot) = \Phi_\sigma(\cdot)x_0$  is not a Markov process by itself, this is the case for  $(x(\cdot), \sigma(\cdot))$ , which is a PDMP. However, with a slight abuse of language, we will sometimes refer to  $x(\cdot)$  as a Markov process for  $A$ .

We say that a stochastic matrix  $P$  is *strongly connected* if it is not similar via a permutation to a nontrivial block upper triangular matrix, i.e., if its associated directed graph

is strongly connected. (Such a matrix is usually called *irreducible*, but we already use the latter term in its linear algebraic meaning.) More generally, every stochastic matrix  $P \in \mathcal{M}_N(\mathbb{R})$  admits, up to a permutation in the set of indices  $\llbracket 1, N \rrbracket$ , the decomposition into strongly connected blocks (see, e.g., [22]) given by

$$P = \begin{pmatrix} P_1 & 0 & \cdots & \cdots & 0 \\ 0 & \ddots & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & P_R & 0 \\ * & \cdots & \cdots & * & Q \end{pmatrix}, \quad (2.4)$$

where  $\rho(Q) < 1$  and, for  $i \in \llbracket 1, R \rrbracket$ ,  $P_i \in \mathcal{M}_{n_i}(\mathbb{R})$  is a stochastic and strongly connected matrix for some positive integers  $R, n_1, \dots, n_R$ . For every  $i \in \llbracket 1, R \rrbracket$ , we define the recurrence class  $\mathcal{J}(i)$  for  $P$  by

$$\mathcal{J}(i) = \llbracket n_1 + \cdots + n_{i-1} + 1, n_1 + \cdots + n_i \rrbracket, \quad (2.5)$$

and the set of transient states by  $\mathcal{T} = \llbracket n_1 + \cdots + n_R + 1, N \rrbracket$  (possibly empty). Given a probability vector  $\mathbf{v} \in \mathbb{R}^N$ , a recurrence class  $\mathcal{J}(i)$  is said to be *accessible from*  $\mathbf{v}$  if  $\mathbb{P}_{\mathbf{v}}(\sigma(\tau) \in \mathcal{J}(i)) > 0$ , where  $\tau = \inf\{t \geq 0 \mid \sigma(t) \notin \mathcal{T}\}$ .

A probability vector  $\mathbf{v} \in \mathbb{R}^N$  is said to be *invariant for*  $P$  if  $\mathbf{v}P = \mathbf{v}$ . When  $P$  is strongly connected, such an invariant law exists and is unique. More generally, considering a decomposition in strongly connected blocks (2.4), we let  $\mathbf{v}^{[i]}$  be the unique invariant probability vector for  $P_i$ , which is canonically extended (adding zero components) to a vector in  $\mathbb{R}^N$  still denoted by  $\mathbf{v}^{[i]}$ . Then every invariant probability vector  $\mathbf{v} \in \mathbb{R}^N$  of  $P$  can be uniquely decomposed as

$$\mathbf{v} = \sum_{i=1}^R \alpha_i \mathbf{v}^{[i]}, \quad (2.6)$$

where  $\alpha_1, \dots, \alpha_R \in [0, 1]$  and  $\sum_{i=1}^R \alpha_i = 1$ . This means that the corresponding Markov process  $\sigma$  starts in the recurrence class  $\mathcal{J}(i)$  with probability  $\alpha_i$ , for  $i \in \llbracket 1, R \rrbracket$ , and remains there for all positive times.

### 2.3 Probabilistic Lyapunov exponent

Let  $(\mathbf{v}, \mu, P)$  define a Markov process in  $\Sigma$  and  $A \in \mathcal{M}_d(\mathbb{R})^N$ . The *probabilistic Lyapunov exponent* for  $(\mathbf{v}, \mu, P, A)$  is the quantity defined by

$$\lambda_p(\mathbf{v}, \mu, P, A) = \limsup_{t \rightarrow +\infty} \frac{1}{t} \mathbb{E}[\log \|\Phi_{\sigma}(t)\|], \quad (2.7)$$

where  $\mathbb{E}$  denotes the expectation with respect to  $\sigma$  distributed according to  $(\mathbf{v}, \mu, P)$ . As for  $\lambda_d(A)$ , the above expression is independent on the choice of the specific induced norm  $\|\cdot\|$  in  $\mathcal{M}_d(\mathbb{R})$ .

Given  $A \in \mathcal{M}_d(\mathbb{R})^N$ , we denote by  $\lambda_p^{\sup}(A)$  the supremum of  $\lambda_p(\mathbf{v}, \mu, P, A)$  over all parameters  $(\mathbf{v}, \mu, P)$ . For every Markov process  $(\mathbf{v}, \mu, P)$ , we have

$$\lambda_p(\mathbf{v}, \mu, P, A) \leq \lambda_p^{\sup}(A) \leq \lambda_d(A), \quad (2.8)$$



since, for fixed  $t > 0$ ,

$$\mathbb{E}[\log \|\Phi_\sigma(t)\|] \leq \sup_{\sigma' \in \Sigma} \log \|\Phi_{\sigma'}(t)\|.$$

Provided that  $\nu$  is invariant for  $P$ , by classical subadditivity arguments on the function  $t \mapsto \mathbb{E}[\log \|\Phi_\sigma(t)\|]$ , one has

$$\lambda_p(\nu, \mu, P, A) = \lim_{t \rightarrow +\infty} \frac{1}{t} \mathbb{E}[\log \|\Phi_\sigma(t)\|] = \inf_{t > 0} \frac{1}{t} \mathbb{E}[\log \|\Phi_\sigma(t)\|]. \quad (2.9)$$

In fact, the initial condition  $\nu$  is not very important, due to the Markov property and the ergodic behaviour within each recurrent class. More precisely, we can state the following.

**Proposition 2.3.** *Consider the decomposition (2.4) and let  $\tau = \inf\{t \geq 0 \mid \sigma(t) \notin \mathcal{J}\}$  be the first time at which  $\sigma$  reaches a recurrent class. Then, for each probability vector  $\nu \in \mathbb{R}^N$ ,*

$$\lambda_p(\nu, \mu, P, A) = \sum_{i=1}^R \mathbb{P}_\nu(\sigma(\tau) \in \mathcal{J}(i)) \lambda_p(\nu^{[i]}, \mu, P, A).$$

*Proof.* Let  $K = \max_{i \in \llbracket 1, N \rrbracket} \|A_i\|$ . Then, for all  $\sigma \in \Sigma$  and all  $t, s \geq 0$ ,

$$e^{-Kt} \|\Phi_\sigma(s)\| \leq \|\Phi_\sigma(t+s)\| \leq e^{Kt} \|\Phi_\sigma(s)\|. \quad (2.10)$$

Fix a probability vector  $\nu \in \mathbb{R}^N$  and let  $\alpha_i = \mathbb{P}_\nu(\sigma(\tau) \in \mathcal{J}(i))$  for  $i \in \llbracket 1, R \rrbracket$  and  $\tilde{\nu} = \sum_{i=1}^R \alpha_i \nu^{[i]}$ . By standard arguments on Markov chains, denoting  $\nu(t) = \nu e^{t\mu(P-I)}$  the law at time  $t$  of a chain associated with  $P$  with initial condition  $\nu$ , then  $\nu(t)$  converges to  $\tilde{\nu}$  as  $t \rightarrow +\infty$ . For an arbitrary  $\varepsilon > 0$ , let  $t_0$  be such that the total variation norm of  $\nu(t_0) - \tilde{\nu}$  is less than  $\varepsilon$ . It means that there exist random variables  $\sigma_0, \sigma'_0$  on  $\llbracket 1, N \rrbracket$  respectively distributed according to  $\nu(t_0)$  and  $\tilde{\nu}$  such that  $\mathbb{P}(\sigma_0 \neq \sigma'_0) \leq \varepsilon/2$ . Considering two chains  $\sigma$  and  $\sigma'$  with respective initial conditions  $\sigma_0$  and  $\sigma'_0$  and such that, conditionally to  $\{\sigma_0 = \sigma'_0\}$ ,  $\{\sigma(t) = \sigma'(t) \forall t \geq 0\}$ , we get that, for all  $t \geq t_0$ ,

$$|\mathbb{E}_{\nu(t_0)}(\log \|\Phi_\sigma(t-t_0)\|) - \mathbb{E}_{\tilde{\nu}}(\log \|\Phi_\sigma(t-t_0)\|)| \leq \varepsilon K(t-t_0).$$

From (2.10),

$$|\log \|\Phi_\sigma(t)\| - \log \|\Phi_{\sigma(t_0+\cdot)}(t-t_0)\| \leq Kt_0.$$

Thus, using the Markov property, for all  $t \geq t_0$ ,

$$|\mathbb{E}_\nu(\log \|\Phi_\sigma(t)\|) - \mathbb{E}_{\nu(t_0)}(\log \|\Phi_\sigma(t-t_0)\|)| \leq Kt_0.$$

Combining these two bounds, dividing by  $t$ , taking the lim sup as  $t \rightarrow +\infty$ , and using that  $\varepsilon$  is arbitrary, we get that

$$\lambda_p(\nu, \mu, P, A) = \lambda_p(\tilde{\nu}, \mu, P, A).$$

Besides, conditioning with respect to the recurrence class of the initial condition, we immediately get that

$$\mathbb{E}_{\tilde{\nu}}(\log \|\Phi_\sigma(t)\|) = \sum_{i=1}^R \alpha_i \mathbb{E}_{\nu^{[i]}}(\log \|\Phi_\sigma(t)\|)$$

for all  $t \geq 0$ . Dividing by  $t$  and letting  $t \rightarrow +\infty$  yield the conclusion.  $\square$

From Proposition 2.3, we deduce that  $\lambda_p(\mathbf{v}, \mu, P, A) \leq \max_{i \in \llbracket 1, R \rrbracket} \lambda_p(\mathbf{v}^{[i]}, \mu, P, A)$ . When  $\sigma(0)$  is distributed according to  $\mathbf{v}^{[i]}$  for some  $i \in \llbracket 1, R \rrbracket$ ,  $\sigma$  stays for all times in the class  $\mathcal{J}(i)$ , where  $P_i$  is strongly connected. Since we are interested in maximal values of the Lyapunov exponent, we are going to use several times in what follows Proposition 2.3 to reduce to the strongly connected case.

### 3 Statements of the main results

Our main result concerning equality between  $\lambda_p(\mathbf{v}, \mu, P, A)$  and  $\lambda_d(A)$  for a given Markov process  $(\mathbf{v}, \mu, P)$  and a given  $A \in \mathcal{M}_d(\mathbb{R})^N$  is the following.

**Theorem 3.1.** *Let  $P \in \mathcal{M}_N(\mathbb{R})$  be a stochastic matrix,  $\mathbf{v}$  be a probability vector of  $\mathbb{R}^N$ ,  $\mu > 0$ , and  $A = (A_1, \dots, A_N) \in \mathcal{M}_d(\mathbb{R})^N$ . Then the following statements are equivalent:*

(a)  $\lambda_d(A) = \lambda_p(\mathbf{v}, \mu, P, A)$ .

(b) *For every recurrent class  $\mathcal{J}$  of  $P$  which is accessible from  $\mathbf{v}$ , every  $k \in \mathbb{N}$ ,  $i_1, \dots, i_k \in \mathcal{J}$ , and  $t_1, \dots, t_k \geq 0$ , it holds*

$$\rho(e^{A_{i_k} t_k} \dots e^{A_{i_1} t_1}) = e^{\lambda_d(A)(t_1 + \dots + t_k)}.$$

The proof of Theorem 3.1 is the main goal of Section 4.

The sequel of the paper is motivated by the problem of characterizing equality between  $\lambda_d(A)$  and  $\lambda_p^{\text{sup}}(A)$ . A first step in that direction is to understand the behavior of sequences  $(\mathbf{v}_n, \mu_n, P_n)_{n \in \mathbb{N}}$  of Markov processes for a given  $A = (A_1, \dots, A_N)$ . For that purpose, we introduce the following definition.

**Definition 3.2.** *A convexified Markov process  $(x, \sigma)$  for  $A$  is a continuous-time Markov process with modes  $B_1, \dots, B_k$ , where  $k \in \llbracket 1, N \rrbracket$ ,  $B_j \in \text{co}\{A_\ell \mid \ell \in I_j\}$  for  $j = 1, \dots, k$ , and  $I_1, \dots, I_k$  are pairwise disjoint nonempty subsets of  $\llbracket 1, N \rrbracket$ .*

We also define the quantity

$$\lambda_p^{\text{conv}}(A) = \sup_{(\mathbf{v}, \mu, P, B)} \lambda_p(\mathbf{v}, \mu, P, B), \quad (3.1)$$

where the supremum is taken among all convexified Markov processes  $(\mathbf{v}, \mu, P, B)$  for  $A$ .

Note that the quantity  $\lambda_p^{\text{conv}}(A)$  introduced above satisfies

$$\lambda_p^{\text{sup}}(A) \leq \lambda_p^{\text{conv}}(A) \leq \lambda_d(A), \quad (3.2)$$

where the last inequality follows from the fact that, for every convexified Markov processes  $(\mathbf{v}, \mu, P, B)$  for  $A$ , we have by (2.8) that  $\lambda_p(\mathbf{v}, \mu, P, B) \leq \lambda_d(B)$  and, in addition,  $\lambda_d(B) \leq \lambda_d(A)$ , the latter inequality being a consequence of the fact that  $\lambda_d(\widehat{A}) = \lambda_d(\text{co}(\widehat{A}))$  for every  $\widehat{A} \in \mathcal{M}_d(\mathbb{R})^N$  (see, e.g., [23]).

Our main result concerning convexified Markov processes is that they compactify the space of Markov processes. More precisely, we prove the following theorem.

**Theorem 3.3.** Consider a sequence  $(x_n, \sigma_n)$  of Markov processes for  $A$  with parameters  $(\nu_n, \mu_n, P_n)_{n \in \mathbb{N}}$ . Up to extracting a subsequence, there exists a convexified Markov process  $(x, \sigma)$  for  $A$  such that, for all  $T, \delta > 0$ ,

$$\mathbb{P} \left( \sup_{t \in [0, T]} |x(t) - x_n(t)| > \delta \right) \xrightarrow{n \rightarrow +\infty} 0.$$

The proof of Theorem 3.3 can be found in Section 5.

Thanks to the compactification result from Theorem 3.3, we are able to provide the following result on the relations between  $\lambda_p^{\text{sup}}(A)$ ,  $\lambda_p^{\text{conv}}(A)$ , and  $\lambda_d(A)$ . Recall that, given a square matrix  $M$ ,  $\lambda(M)$  denotes its spectral abscissa.

**Theorem 3.4.** Let  $A = (A_1, \dots, A_N) \in \mathcal{M}_d(\mathbb{R})^N$ .

(a) The equality  $\lambda_d(A) = \lambda_p^{\text{conv}}(A)$  holds true if and only if there exists  $M \in \text{co}(A)$  such that  $\lambda(M) = \lambda_d(A)$ .

(b) If  $\lambda_d(A) = \lambda_p^{\text{sup}}(A)$  then there exists  $M \in \text{co}(A)$  such that  $\lambda(M) = \lambda_d(A)$ .

(c) Assume that there exists  $M \in \text{co}(A)$  such that  $\lambda(M) = \lambda_d(A)$  and that, for every  $\varepsilon > 0$ , there exist  $M^\varepsilon \in \text{co}(A)$  and a sequence  $(\nu_n, \mu_n, P_n)_{n \in \mathbb{N}}$  of Markov processes for  $A$  such that  $\|M - M^\varepsilon\| < \varepsilon$  and

$$\lambda(M^\varepsilon) = \lim_{n \rightarrow \infty} \lambda_p(\nu_n, \mu_n, P_n, A). \quad (3.3)$$

Then  $\lambda_d(A) = \lambda_p^{\text{sup}}(A)$ .

We conjecture that the converse of Theorem 3.4 (b) is true. However, we are only able to prove it in low dimension, as stated in the following result.

**Proposition 3.5.** Let  $d \leq 3$  and  $A \in \mathcal{M}_d(\mathbb{R})^N$ . Then  $\lambda_d(A) = \lambda_p^{\text{sup}}(A)$  if and only if there exists  $M \in \text{co}(A)$  such that  $\lambda(M) = \lambda_d(A)$ .

The proofs of Theorem 3.4 and Proposition 3.5 are provided in Section 6.

## 4 Characterization of equality between $\lambda_d(A)$ and $\lambda_p(\nu, \mu, P, A)$

The goal of this section is to prove Theorem 3.1. We start with the particular situation in which  $P$  is strongly connected and  $A$  is irreducible.

**Proposition 4.1.** Let  $P \in \mathcal{M}_N(\mathbb{R})$  be a stochastic strongly connected matrix,  $\nu$  a probability vector of  $\mathbb{R}^N$ , and  $\mu > 0$ . Let  $A = (A_1, \dots, A_N) \in \mathcal{M}_d(\mathbb{R})^N$  be irreducible and  $\|\cdot\|_e$  be an extremal norm for  $A$ . Then the following statements are equivalent:

(a)  $\lambda_d(A) = \lambda_p(\nu, \mu, P, A)$ .

(b) For every  $k \in \mathbb{N}$ ,  $i_1, \dots, i_k \in \llbracket 1, N \rrbracket$ , and  $t_1, \dots, t_k \geq 0$ , one has

$$\left\| e^{A_{i_k} t_k} \dots e^{A_{i_1} t_1} \right\|_e = e^{\lambda_d(A)(t_1 + \dots + t_k)}.$$

(c) For every  $k \in \mathbb{N}$ ,  $i_1, \dots, i_k \in \llbracket 1, N \rrbracket$ , and  $t_1, \dots, t_k \geq 0$ , one has

$$\rho(e^{A_{i_k} t_k} \dots e^{A_{i_1} t_1}) = e^{\lambda_d(A)(t_1 + \dots + t_k)}.$$

(d) Up to a common change of basis, the matrices  $A_i - \lambda_d(A)\text{Id}$ ,  $i \in \llbracket 1, N \rrbracket$ , are skew-symmetric.

*Proof.* Note that, in terms of the flow  $\Phi_\sigma$ , items (b) and (c) can be equivalently stated by saying that the quantities  $\frac{1}{t} \log \|\Phi_\sigma(t)\|_e$  and  $\frac{1}{t} \log \rho(\Phi_\sigma(t))$ , respectively, are independent of  $t > 0$  and  $\sigma \in \Sigma$  and are equal to  $\lambda_d(A)$ .

We will first show that (a), (b), and (c) are equivalent. The fact that (b) implies (c) follows from Gelfand's formula for the spectral radius. To show that (c) implies (a), notice that, for every induced norm  $\|\cdot\|$  in  $\mathcal{M}_d(\mathbb{R})$ , it follows from (c) that

$$\lambda_d(A) = \frac{1}{t} \log \rho(\Phi_\sigma(t)) \leq \frac{1}{t} \log \|\Phi_\sigma(t)\|$$

for every  $\sigma \in \Sigma$  and  $t > 0$ . Hence, by first taking the expectation with respect to  $\sigma$  and then the limsup as  $t$  tends to infinity, we obtain that  $\lambda_d(A) \leq \lambda_p(\nu, \mu, P, A)$ , yielding (a) thanks to (2.8).

Let us now prove that (a) implies (b). Without loss of generality, by Proposition 2.3, we assume that  $\nu$  is the unique invariant probability vector for  $P$ . By definition of extremal norm, for every  $k \in \mathbb{N}$ ,  $i_1, \dots, i_k \in \llbracket 1, N \rrbracket$ , and  $t_1, \dots, t_k \geq 0$ , one has

$$\left\| e^{A_{i_k} t_k} \dots e^{A_{i_1} t_1} \right\|_e \leq e^{\lambda_d(A)(t_1 + \dots + t_k)},$$

which can be equivalently rewritten as

$$\frac{1}{t} \log \|\Phi_\sigma(t)\|_e \leq \lambda_d(A), \quad (4.1)$$

for every  $t > 0$  and  $\sigma \in \Sigma$ .

Arguing by contradiction, there exist  $k \in \mathbb{N}$ ,  $i_1, \dots, i_k \in \llbracket 1, N \rrbracket$ , and  $t_1, \dots, t_k \geq 0$  such that

$$\left\| e^{A_{i_k} t_k} \dots e^{A_{i_1} t_1} \right\|_e < e^{\lambda_d(A)(t_1 + \dots + t_k)}. \quad (4.2)$$

We claim that, with no loss of generality,  $p_{i_1 i_2} \dots p_{i_{k-1} i_k} > 0$ . Indeed, if it were not the case, then  $p_{i_\ell i_{\ell+1}} = 0$  for some  $\ell \in \llbracket 1, k-1 \rrbracket$ . Since  $P$  is strongly connected, there exist  $r \in \mathbb{N}$  and  $j_1, \dots, j_r \in \llbracket 1, N \rrbracket$  such that  $j_1 = i_\ell$ ,  $j_r = i_{\ell+1}$ , and  $p_{j_1 j_2} \dots p_{j_{r-1} j_r} > 0$ . Letting  $s_1 = t_\ell$ ,  $s_r = t_{\ell+1}$ , and  $s_2 = \dots = s_{r-1} = 0$ , we may then replace  $e^{A_{i_{\ell+1}} t_{\ell+1}} e^{A_{i_\ell} t_\ell}$  by  $e^{A_{j_r} s_r} \dots e^{A_{j_1} s_1}$  in (4.2). Repeating the previous construction for every  $\ell$  such that  $p_{i_\ell i_{\ell+1}} = 0$ , the claim is proved. Note also that, by continuity, (4.2) holds for an open subset of times  $t_1, \dots, t_k$  in  $(0, +\infty)^k$ . We have thus proved that there exists  $t > 0$  and a set of positive probability of signals  $\sigma \in \Sigma$  such that

$$\frac{1}{t} \log \|\Phi_\sigma(t)\|_e < \lambda_d(A).$$

Combining with (4.1), we deduce from (2.9) that  $\lambda_p(\nu, \mu, P, A) < \lambda_d(A)$ .

To prove that (d) implies (c), notice that, if  $A_i - \lambda_d(A)\text{Id}$  is skew-symmetric for every  $i \in \llbracket 1, N \rrbracket$ , then, for every  $k \in \mathbb{N}$ ,  $i_1, \dots, i_k \in \llbracket 1, N \rrbracket$ , and  $t_1, \dots, t_k \geq 0$ , the matrix

$$e^{(A_{i_k} - \lambda_d(A)\text{Id})t_k} \dots e^{(A_{i_1} - \lambda_d(A)\text{Id})t_1} = e^{A_{i_k}t_k} \dots e^{A_{i_1}t_1} e^{-\lambda_d(A)(t_1 + \dots + t_k)}$$

is the product of orthogonal matrices, hence it is itself orthogonal and its spectral radius is equal to 1. The conclusion follows.

Finally, to prove that (c) implies (d), notice that the semigroup

$$\{e^{(A_{i_k} - \lambda_d(A)\text{Id})t_k} \dots e^{(A_{i_1} - \lambda_d(A)\text{Id})t_1} \mid k \in \mathbb{N}, i_1, \dots, i_k \in \llbracket 1, N \rrbracket, t_1, \dots, t_k \geq 0\}$$

is irreducible since  $A$  is irreducible. Moreover, (c) is equivalent to saying that the above semigroup has constant spectral radius. Then, using [21, Theorem 2], we deduce that, up to a common change of basis,  $e^{(A_i - \lambda_d(A)\text{Id})t}$  is orthogonal for every  $i \in \llbracket 1, N \rrbracket$  and  $t \geq 0$ , yielding the conclusion.  $\square$

**Remark 4.2.** The equivalences between (a), (b), and (c) only rely on the extremality of the norm  $\|\cdot\|_e$  for  $A$ , and hence hold under the weaker assumption that  $A$  is nondefective instead of irreducible (cf. Remark 2.2). Notice also that the proof that (c) implies (a) requires neither the irreducibility of  $A$  nor the strong connectedness of  $P$ .

In the next result, we extend Proposition 4.1 to the more general case where  $A$  is not necessarily irreducible, but  $P$  is still assumed to be strongly connected.

**Proposition 4.3.** *Let  $P \in \mathcal{M}_N(\mathbb{R})$  be a stochastic strongly connected matrix,  $\nu$  be a probability vector of  $\mathbb{R}^N$ ,  $\mu > 0$ , and  $A = (A_1, \dots, A_N) \in \mathcal{M}_d(\mathbb{R})^N$ . Then the following statements are equivalent:*

(a)  $\lambda_d(A) = \lambda_p(\nu, \mu, P, A)$ .

(b) For every  $k \in \mathbb{N}$ ,  $i_1, \dots, i_k \in \llbracket 1, N \rrbracket$ , and  $t_1, \dots, t_k \geq 0$ , it holds

$$\rho(e^{A_{i_k}t_k} \dots e^{A_{i_1}t_1}) = e^{\lambda_d(A)(t_1 + \dots + t_k)}.$$

*Proof.* As in the proof of Proposition 4.1, thanks to Proposition 2.3, we can suppose that  $\nu$  is the unique invariant measure of  $P$ . Due to Remark 4.2, we are only left to show that (a) implies (b). As in [7, Lemma 3.5], a key ingredient of the argument is a simultaneous block decomposition of the matrices  $A_1, \dots, A_N$ . The idea is the following: if  $A_1, \dots, A_N$  admit a common proper subspace  $V$  of dimension  $d'$ , then, up to a linear change of coordinates corresponding to a basis of  $\mathbb{R}^d$  consisting of a basis of  $V$  in its first  $d'$  elements, each matrix  $A_j$  can be written as  $\begin{pmatrix} B_j & C_j \\ 0 & D_j \end{pmatrix}$  for some matrices  $B_j, C_j, D_j$ ,  $j \in \llbracket 1, N \rrbracket$ , with  $B_j \in \mathcal{M}_{d'}(\mathbb{R})$ . By an immediate inductive argument, up to a linear change of coordinates,  $A_1, \dots, A_N$  can be presented in block-triangular form as

$$A_j = \begin{pmatrix} A_j^{(1)} & * & * & \dots & * \\ 0 & A_j^{(2)} & * & \dots & * \\ 0 & 0 & A_j^{(3)} & \ddots & * \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & A_j^{(s)} \end{pmatrix}, \quad j \in \llbracket 1, N \rrbracket, \quad (4.3)$$

for some appropriate integer  $S$ , with  $A^{(s)} = (A_1^{(s)}, \dots, A_N^{(s)})$  irreducible for every  $s \in \llbracket 1, S \rrbracket$ . Both deterministic and probabilistic Lyapunov exponents are obtained as maxima of the corresponding Lyapunov exponents over the diagonal blocks (see, e.g., [6, Proposition 2] for the deterministic case and [15] for the probabilistic one). Notice also that, for every  $k \in \mathbb{N}$ ,  $i_1, \dots, i_k \in \llbracket 1, N \rrbracket$ , and  $t_1, \dots, t_k \geq 0$ , it holds

$$\lambda_d(A) \geq \frac{\log \rho \left( e^{A_{i_k} t_k} \dots e^{A_{i_1} t_1} \right)}{t_1 + \dots + t_k} = \max_{s \in \llbracket 1, S \rrbracket} \frac{\log \rho \left( e^{A_{i_k}^{(s)} t_k} \dots e^{A_{i_1}^{(s)} t_1} \right)}{t_1 + \dots + t_k}, \quad (4.4)$$

where the inequality comes from (2.3) and the equality follows 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.

Let  $\bar{s} \in \llbracket 1, S \rrbracket$  be the index such that  $\lambda_p(\nu, \mu, P, A) = \lambda_p(\nu, \mu, P, A^{(\bar{s})})$  and notice that, thanks to (a) and (2.8),  $\lambda_d(A) = \lambda_d(A^{(\bar{s})})$ . By Proposition 4.1 and (4.4), we deduce that, for every  $k \in \mathbb{N}$ ,  $i_1, \dots, i_k \in \llbracket 1, N \rrbracket$ , and  $t_1, \dots, t_k \geq 0$ , it holds

$$\lambda_d(A) = \lambda_d(A^{(\bar{s})}) = \frac{\log \rho \left( e^{A_{i_k}^{(\bar{s})} t_k} \dots e^{A_{i_1}^{(\bar{s})} t_1} \right)}{t_1 + \dots + t_k} \leq \frac{\log \rho \left( e^{A_{i_k} t_k} \dots e^{A_{i_1} t_1} \right)}{t_1 + \dots + t_k} \leq \lambda_d(A),$$

yielding (b). □

**Remark 4.4.** As a byproduct of the block-decomposition argument in the above proof, we get another statement equivalent to (a) and (b): there exist a linear change of variables and  $p \in \llbracket 1, d \rrbracket$  such that, for every  $i \in \llbracket 1, N \rrbracket$ ,

$$A_i - \lambda_d(A) \text{Id} = \begin{pmatrix} * & * \\ 0 & B_i \end{pmatrix}$$

with  $B_i$  a  $p \times p$  skew-symmetric matrix.

Moreover, any of the previous statements implies that

$$\lambda(M) = \lambda_d(A), \quad \forall M \in \text{co}(A).$$

Indeed, write  $M = \beta_1 A_1 + \dots + \beta_N A_N$  with  $\beta_1, \dots, \beta_N \in [0, 1]$  and  $\beta_1 + \dots + \beta_N = 1$ . Take  $k = N$  and  $i_j = j$  and  $t_j = t \beta_j$  for  $j \in \llbracket 1, N \rrbracket$  in (b). The conclusion follows by letting  $t \rightarrow 0$ .

Finally, we can turn to the proof of Theorem 3.1.

*Proof of Theorem 3.1.* According to Proposition 2.3, (a) is equivalent to the fact that for all  $i \in \llbracket 1, R \rrbracket$  such that  $\mathcal{J}(i)$  is accessible from  $\nu$ ,  $\lambda_d(A) = \lambda_p(\nu^{[i]}, \mu, P, A)$ . Replacing  $P$  by  $P_i$  and  $A$  by  $(A_j)_{j \in \mathcal{J}(i)}$ , we are in a strongly connected case and Proposition 4.3 concludes. □

## 5 Compactification of the space of Markov processes

The aim of this section is to prove Theorem 3.3, that is, that any sequence  $(\nu_n, \mu_n, P_n)_{n \in \mathbb{N}}$  of Markov processes for  $A = (A_1, \dots, A_N)$  admits a subsequence converging in law to a convexified Markov process for  $A$ .

Notice that an important difference from more classical averaging results (such as [4, Section 2.5] or [5, Corollary 2.15]) is that  $P_n$  is not fixed and, in particular, the jump rates  $\mu_n(P_n)_{i,j}$  may have different asymptotic behaviours as  $n$  goes to infinity depending on  $i, j$ , which is why, when  $\lim_{n \rightarrow \infty} \mu_n = +\infty$ , the limit process is not necessarily a deterministic ODE  $\dot{x} = Mx$  for some  $M \in \text{co}(A)$ .

### 5.1 Convexified Markov processes as limits of Markov processes

First, let us show that any convexified Markov process for  $A$  can be obtained as the limit of a sequence of Markov processes for  $A$ . For the sake of clarity, let us stress that we only consider here the convergence of the continuous component  $x(\cdot)$ .

**Proposition 5.1.** *Let  $(x, \sigma)$  be a convexified Markov process for  $A$ . Then there exists a sequence  $(x_n, \sigma_n)_{n \in \mathbb{N}}$  of Markov processes for  $A$  such that, for all  $T, \delta > 0$ ,*

$$\mathbb{P} \left( \sup_{t \in [0, T]} |x(t) - x_n(t)| > \delta \right) \xrightarrow{n \rightarrow +\infty} 0.$$

*Proof.* Denote by  $B_1, \dots, B_k$  the modes of  $(x, \sigma)$  with  $B_r = \sum_{j \in I_r} \pi_r(j) A_j$ , where  $I_1, \dots, I_k$  are pairwise disjoint nonempty subsets of  $\llbracket 1, N \rrbracket$  and, for all  $r \in \llbracket 1, k \rrbracket$ ,  $\pi_r$  is a probability measure on  $I_r$ . In particular,  $(\sigma(t))_{t \geq 0}$  is a continuous-time Markov chain on  $\llbracket 1, k \rrbracket$ . Denote by  $\mu$  its jump rate,  $P$  its transition matrix, and  $\nu$  its initial probability law. Let  $T_0 = 0$  and  $(T_m)_{m \geq 0}$  be the jump times of  $\sigma$ , so that  $(T_{m+1} - T_m)_{m \in \mathbb{N}}$  is an i.i.d. sequence of random variables distributed according to the exponential law with parameter  $\mu > 0$ . We will construct for all  $n \in \mathbb{N}$  a Markov chain  $(\sigma_n(t))_{t \geq 0}$  on  $\llbracket 1, N \rrbracket$  such that  $\sigma_n(t) \in I_{\sigma(t)}$  for all  $t \geq 0$  and which is moreover re-sampled at rate  $n$  according to  $\pi_{\sigma(T_m)}$  between consecutive slow jump times  $T_m$  and  $T_{m+1}$ .

More precisely, let  $(N_t)_{t \geq 0}$  be a standard Poisson process with intensity 1 and  $\mathcal{U} = (\mathcal{U}_{p,r})_{p \in \mathbb{N}, r \in \llbracket 1, k \rrbracket}$  be a family of independent random variables such that, for all  $p \in \mathbb{N}$  and  $r \in \llbracket 1, k \rrbracket$ ,  $\mathcal{U}_{p,r}$  takes values in  $I_r$  and is distributed according to  $\pi_r$ , with moreover  $(N_t)_{t \geq 0}$ ,  $\mathcal{U}$ , and  $\sigma$  independent. For all  $n \in \mathbb{N}$  and  $t \geq 0$ , set  $M_t^n = N_{nt} + \sum_{m \in \mathbb{N}} \mathbb{1}_{t \geq T_m}$ , so that  $(M_t^n)_{t \geq 0}$  is a Poisson process with intensity  $\mu + n$  such that that all jumps of  $\sigma$  are jumps of  $M^n$ . For all  $n \in \mathbb{N}$  and  $t \geq 0$ , set  $\sigma_n(t) = \mathcal{U}_{M_t^n, \sigma(t)}$ .

Then, for all  $n \in \mathbb{N}$ ,  $(\sigma_n(t))_{t \geq 0}$  is a Markov chain on  $\llbracket 1, N \rrbracket$  with initial condition  $\nu_n$  and jump rates  $\lambda_n(i, j)$  for  $i, j \in \llbracket 1, N \rrbracket$  given as follows:

- for all  $i \in \llbracket 1, N \rrbracket$ ,  $\nu_n(i) = \nu(r) \pi_r(i)$  if  $i \in I_r$  with  $r \in \llbracket 1, k \rrbracket$ ;
- for all  $r \in \llbracket 1, k \rrbracket$  and all  $i, j \in I_r$ ,  $\lambda_n(i, j) = n \pi_r(j)$ ;
- for all distinct  $r, s \in \llbracket 1, k \rrbracket$  and all  $i \in I_r, j \in I_s$ ,  $\lambda_n(i, j) = \mu \pi_s(j) P(r, s)$ ;
- for all  $i \notin I := \bigcup_{r=1}^k I_r$  and all  $j \in \llbracket 1, N \rrbracket$ ,  $\lambda_n(i, j) = \lambda_n(j, i) = 0$ .

Remark that, for all  $n \in \mathbb{N}$  and  $t \geq 0$ ,  $\sigma_n(t) \in I_{\sigma(t)}$  (and in particular  $\sigma_n(t) \in I$ ), so that  $\sigma$  is completely determined by  $\sigma_n$ .

The proof is then similar to [5, Lemma 2.14]. From [13, Chapter 2, Theorem 1.3], it is sufficient to prove that for all  $t_0, T > 0$ ,  $\int_{t_0}^{t_0+T} A_{\sigma_n(s)} ds$  converges in probability as  $n \rightarrow \infty$  towards  $\int_{t_0}^{t_0+T} B_{\sigma(s)} ds$ , uniformly with respect to  $t_0$ . As in the proof of [5, Lemma 2.14], it is thus sufficient to prove that, for all  $r \in \llbracket 1, k \rrbracket$  and all  $j \in I_r$ ,  $\int_{t_0}^{t_0+T} \mathbb{1}_{\sigma_n(s)=j} ds$  converges in probability as  $n \rightarrow \infty$  towards  $\pi_r(j) \int_{t_0}^{t_0+T} \mathbb{1}_{\sigma(s)=r} ds$ , uniformly with respect to  $t_0$ . By the Markov property, for all  $\delta > 0$ ,

$$\begin{aligned} \mathbb{P} \left( \left| \int_{t_0}^{t_0+T} (\mathbb{1}_{\sigma_n(s)=j} - \pi_r(j) \mathbb{1}_{\sigma(s)=r}) ds \right| > \delta \right) \\ \leq \sup_{i \in I} \mathbb{P}_i \left( \left| \int_0^T (\mathbb{1}_{\sigma_n(s)=j} - \pi_r(j) \mathbb{1}_{\sigma(s)=r}) ds \right| > \delta \right), \end{aligned}$$

where we recall that the subscript  $i$  denotes the conditioning  $\sigma_n(0) = i$ .

Let  $R$  be a positive integer. For all  $b \in \llbracket 0, R-1 \rrbracket$  and all  $t \geq bT/R$ , denote

$$\tilde{\sigma}_n^b(t) = \mathcal{U}_{N_{nt} - N_{nbT/R} + M_{bT/R}^n, \sigma(bT/R)}.$$

In other words,  $\tilde{\sigma}_n^b(t)$  is initialized at time  $bT/R$  with  $\tilde{\sigma}_n^b(bT/R) = \sigma_n(bT/R)$  and then is re-sampled on  $I_{\sigma(bT/R)}$  at each jump of  $N_{nt}$ . In particular, up to  $\inf\{T_m \mid m \in \mathbb{N}, T_m > bT/R\}$ , the first slow jump time after time  $bT/R$ , we have  $\sigma_n = \tilde{\sigma}_n^b$ . In particular,  $\sigma_n$  and  $\tilde{\sigma}_n^b$  coincide on the interval  $[bT/R, (b+1)T/R]$  if the latter does not contain any slow jump time  $T_m$ . Hence, using the identity  $\pi_r(j) \mathbb{1}_{r'=r} = \pi_{r'}(j)$ , we deduce the bound

$$\begin{aligned} \left| \int_0^T (\mathbb{1}_{\sigma_n(s)=j} - \pi_r(j) \mathbb{1}_{\sigma(s)=r}) ds \right| \\ \leq \sum_{b=0}^{R-1} \left| \int_{bT/R}^{(b+1)T/R} (\mathbb{1}_{\sigma_n(s)=j} - \pi_r(j) \mathbb{1}_{\sigma(s)=r}) ds \right| \\ \leq \frac{T}{R} \sum_{m \in \mathbb{N}} \mathbb{1}_{T_m \leq T} + \sum_{b=0}^{R-1} \left| \int_{bT/R}^{(b+1)T/R} (\mathbb{1}_{\tilde{\sigma}_n^b(s)=j} - \pi_{\sigma(bT/R)}(j)) ds \right|. \end{aligned}$$

Consider the events

$$\mathcal{A}_b = \left\{ \left| \int_{bT/R}^{(b+1)T/R} (\mathbb{1}_{\tilde{\sigma}_n^b(s)=j} - \pi_{\sigma(bT/R)}(j)) ds \right| > \frac{\delta}{2R} \right\}$$

for all  $b \in \llbracket 0, R-1 \rrbracket$  and

$$\mathcal{A}_R = \left\{ \frac{T}{R} \sum_{m \in \mathbb{N}} \mathbb{1}_{T_m \leq T} > \frac{\delta}{2} \right\}.$$

Then

$$\left\{ \left| \int_0^T (\mathbb{1}_{\sigma_n(s)=j} - \pi_r(j) \mathbb{1}_{\sigma(s)=r}) ds \right| > \delta \right\} \subset \bigcup_{b=0}^R \mathcal{A}_b$$



and thus

$$\mathbb{P}_i \left( \left| \int_0^T (\mathbb{1}_{\sigma_n(s)=j} - \pi_r(j) \mathbb{1}_{\sigma(s)=r}) ds \right| > \delta \right) \leq \sum_{b=0}^R \mathbb{P}_i(\mathcal{A}_b).$$

Since the number of jumps occurring before time  $T$  follows a Poisson distribution with intensity  $T\mu$ ,

$$\mathbb{P}_i(\mathcal{A}_R) \leq \frac{2T\mathbb{E}(\sum_{m \in \mathbb{N}} \mathbb{1}_{T_m \leq T})}{\delta R} \leq \frac{2T^2\mu}{\delta R}.$$

Moreover, conditioning on the value  $\sigma_n(bT/R)$ , we get that for all  $i \in I$  and  $b \in \llbracket 0, R-1 \rrbracket$ ,

$$\begin{aligned} \mathbb{P}_i(\mathcal{A}_b) &\leq \sup_{u \in I} \mathbb{P}_u \left( \left| \int_0^{T/R} (\mathbb{1}_{\tilde{\sigma}_n^0(s)=j} - \pi_{\sigma(0)}(j)) ds \right| > \frac{\delta}{2R} \right) \\ &\leq \frac{4R^2}{\delta^2} \sup_{u \in I} \mathbb{E}_u \left( \left| \int_0^{T/R} (\mathbb{1}_{\tilde{\sigma}_n^0(s)=j} - \pi_{\sigma(0)}(j)) ds \right|^2 \right). \end{aligned}$$

It only remains to prove that for all  $u \in I$  the expectation vanishes as  $n \rightarrow +\infty$ . Indeed, in that case, we obtain that for all  $T, \delta, R > 0$ ,

$$\limsup_{n \rightarrow +\infty} \sup_{i \in I} \mathbb{P}_i \left( \left| \int_0^T (\mathbb{1}_{\sigma_n(s)=j} - \pi_r(j) \mathbb{1}_{\sigma(s)=r}) ds \right| > \delta \right) \leq \frac{2T^2\mu}{\delta R},$$

yielding the conclusion, since the left-hand side does not depend on  $R$ , which can thus be taken arbitrarily large.

Let us fix  $r \in \llbracket 1, k \rrbracket$  and  $u \in I_r$ , and work conditionally to  $\{\sigma_n(0) = u\}$ . Under this event,  $\tilde{\sigma}_n^0$  is simply a Markov chain starting at  $u$  and re-sampled according to  $\pi_r$  at rate  $n$ . We are back to a problem similar to [5, Lemma 2.14], and follow the same proof. In particular, for all  $j \in I_r$ ,

$$\mathbb{P}_u(\tilde{\sigma}_n^0(t) = j) = e^{-nt} \mathbb{1}_{j=u} + (1 - e^{-nt}) \pi_r(j),$$

and thus

$$\left| \mathbb{E}_u \left( \int_0^{T/R} (\mathbb{1}_{\tilde{\sigma}_n^0(s)=j} - \pi_r(j)) ds \right) \right| \leq \int_0^{T/R} e^{-ns} ds \leq \frac{1}{n}.$$

Similarly

$$\begin{aligned} \mathbb{E}_u \left( \left( \int_0^{T/R} \mathbb{1}_{\tilde{\sigma}_n^0(s)=j} ds \right)^2 \right) &= \int_0^{T/R} \int_0^{T/R} \mathbb{P}_u(\tilde{\sigma}_n^0(s) = j, \tilde{\sigma}_n^0(t) = j) ds dt \\ &= 2 \int_0^{T/R} \int_0^t \mathbb{P}_u(\tilde{\sigma}_n^0(s) = j, \tilde{\sigma}_n^0(t) = j) ds dt. \end{aligned}$$

For all  $s < t$ ,

$$\begin{aligned} &|\mathbb{P}_u(\tilde{\sigma}_n^0(s) = j, \tilde{\sigma}_n^0(t) = j) - \pi_r(j)^2| \\ &= \left| (e^{-ns} \mathbb{1}_{j=u} + (1 - e^{-ns}) \pi_r(j)) (e^{-n(t-s)} + (1 - e^{-n(t-s)}) \pi_r(j)) - \pi_r(j)^2 \right| \\ &\leq e^{-nt} \mathbb{1}_{j=u} + e^{-ns} (1 - e^{-n(t-s)}) \mathbb{1}_{j=u} \pi_r(j) + (1 - e^{-ns}) e^{-n(t-s)} \pi_r(j) \\ &\quad + \left| (1 - e^{-ns})(1 - e^{-n(t-s)}) - 1 \right| \pi_r(j)^2 \\ &\leq e^{-nt} + 2e^{-ns} + 2e^{-n(t-s)}, \end{aligned}$$

so that

$$\left| \mathbb{E}_u \left( \left( \int_0^{T/R} \mathbb{1}_{\bar{\sigma}_n^0(s)=j} ds \right)^2 \right) - T^2 \pi_r(j)^2 / R^2 \right| \xrightarrow{n \rightarrow +\infty} 0.$$

This concludes since, denoting  $Z_n = \int_0^{T/R} \mathbb{1}_{\bar{\sigma}_n^0(s)=j} ds$ , we have then obtained that

$$\begin{aligned} \mathbb{E}_u \left( |Z_n - T \pi_r(j) / R|^2 \right) &= \mathbb{E}_u(Z_n^2) - 2T \pi_r(j) \mathbb{E}_u(Z_n) / R + T^2 \pi_r(j)^2 / R^2 \\ &\xrightarrow{n \rightarrow +\infty} T^2 \pi_r(j)^2 / R^2 - 2T^2 \pi_r(j)^2 / R^2 + T^2 \pi_r(j)^2 / R^2 = 0. \quad \square \end{aligned}$$

## 5.2 Convexified Markov processes compactify the space of Markov processes

We now want to prove a converse of Proposition 5.1, namely that, from any sequence  $(x_n)_{n \in \mathbb{N}}$  of Markov processes for  $A$ , we can extract a subsequence that converges to a convexified process, cf. Theorem 3.3 below. We start by treating separately the simple case where the maximal jump rate  $\mu_n$  is bounded.

**Proposition 5.2.** *Consider a sequence of parameters  $(v_n, \mu_n, P_n)_{n \in \mathbb{N}}$  of Markov processes for  $A$ . Suppose that  $(\mu_n)_{n \in \mathbb{N}}$  is bounded. Then, up to extracting a subsequence, there exist parameters  $(v, \mu, P)$  of a Markov process for  $A$  such that the following holds: There exist Markov processes  $(x_n, \sigma_n)_{n \in \mathbb{N}}$  and  $(x, \sigma)$  associated respectively with  $(v_n, \mu_n, P_n)_{n \in \mathbb{N}}$  and  $(v, \mu, P)$  such that, for all  $T > 0$  and  $\delta > 0$ ,*

$$\mathbb{P} \left( \sup_{t \in [0, T]} |x(t) - x_n(t)| > \delta \right) \xrightarrow{n \rightarrow +\infty} 0.$$

*Proof.* Up to an extraction we can suppose that  $v_n$ ,  $\mu_n$ , and  $P_n$  have limits as  $n \rightarrow +\infty$  (in fact it would have been sufficient to assume that  $\liminf_{n \rightarrow +\infty} \mu_n < +\infty$ ), that we denote by  $v$ ,  $\mu$ , and  $P$ , respectively.

We are going to prove that for all  $T, \varepsilon > 0$ , there exists  $n_0 \in \mathbb{N}$  such that, for all  $n \geq n_0$ , there exist Markov processes  $(x_n, \sigma_n)$  and  $(x, \sigma)$  associated respectively with  $(v_n, \mu_n, P_n)$  and  $(v, \mu, P)$  such that

$$\mathbb{P}(x(t) = x_n(t), \forall t \in [0, T]) \geq 1 - \varepsilon.$$

Remark that, in this statement,  $(x, \sigma)$  may depend on  $n$ . Nevertheless, this yields the convergence of the distribution of  $(x_n(t))_{t \geq 0}$  to the distribution of  $(x(t))_{t \geq 0}$  on all compact time intervals. The result then follows from Skorokhod's representation theorem.

The proof relies on a synchronous coupling of the Markov chains, namely, for each  $n \in \mathbb{N}$ , we can define simultaneously two Markov processes  $(z_n(t))_{t \geq 0} := (x_n(t), \sigma_n(t))_{t \geq 0}$  and  $(z(t))_{t \geq 0} := (x(t), \sigma(t))_{t \geq 0}$  associated respectively with  $(v_n, \mu_n, P_n)$  and  $(v, \mu, P)$  on the same probability space in such a way that they have the same initial value with maximal probability (i.e.,  $\mathbb{P}(z_n(0) \neq z(0)) = |v - v_n|_1 / 2$ ) and that, as long as they stay at the same position, they jump as much as possible at the same times and to the same locations. The

precise construction of this coupling is given in [12, Section 6], to which we refer for details. The Markov generator  $\mathcal{L}$  on  $\mathbb{R}^d \times \llbracket 1, N \rrbracket$  associated with  $(z(t))_{t \geq 0}$  is given by

$$\mathcal{L}f(x, \sigma) = (A_\sigma x) \cdot \nabla_x f(x, \sigma) + \mu \sum_{j=1}^N (P)_{\sigma,j} (f(x, j) - f(x, \sigma)),$$

and similarly for the generator  $\mathcal{L}_n$  of  $z_n$ . Then

$$\varepsilon_n := \sup_{\|f\|_\infty \leq 1} \|\mathcal{L}f - \mathcal{L}_n f\|_\infty \leq 2|\mu_n - \mu| + 2\mu \|P_n - P\|_1 \xrightarrow{n \rightarrow +\infty} 0.$$

From [12, Theorem 11],

$$\mathbb{P}(\exists t \in [0, T], x(t) \neq x_n(t)) \leq \mathbb{P}(z_n(0) \neq z(0)) + 1 - e^{-\varepsilon_n T} \xrightarrow{n \rightarrow +\infty} 0. \quad \square$$

For the rest of this section we consider a given sequence  $(v_n, \mu_n, P_n)_{n \in \mathbb{N}}$  of Markov processes for  $A$  such that, for at least a pair  $(i, j)$  of distinct elements of  $\llbracket 1, N \rrbracket$ , the jump rate  $\lambda_n(i, j) = \mu_n (P_n)_{i,j}$  is unbounded. We will repeatedly consider successive extractions of this sequence and keep writing them  $(v_n, \mu_n, P_n)_{n \in \mathbb{N}}$ . First, up to extracting a subsequence, we suppose that  $\lambda_n(i, j) \rightarrow +\infty$  for some distinct  $i, j$ .

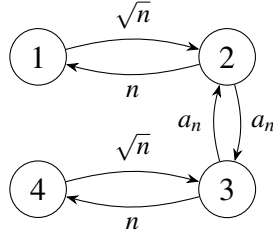


Figure 5.1: A chain on  $\llbracket 1, 4 \rrbracket$  whose jump rates all go to  $+\infty$ .

As in the proof of Proposition 5.1, we would like to separate in the chain fast transitions that happen in arbitrarily small time as  $n \rightarrow +\infty$  and slow transitions. It is not sufficient to consider the pairs  $(i, j)$  such that  $\lambda_n(i, j) \rightarrow +\infty$  or such that  $\lambda_n(i, j)$  is of the order of  $\mu_n$ . Indeed, consider the example given in Figure 5.1, where the values over the arrows denote the jump rates and we assume that  $a_n$  goes to infinity as  $n \rightarrow +\infty$  with the assumption that  $\lim_{n \rightarrow +\infty} \frac{a_n}{n} = 0$ . For instance, starting from the state 2, the Markov chain will go to 1 with high probability (for large  $n$ ). Then, each time it will go back to state 2, it will have a probability  $a_n/(n + a_n)$  to go to state 3, from which it will go very fast to state 4 with high probability. Since the time taken by transitions from 2 to 1, of order  $1/n$ , is negligible with respect to the time taken by transitions from 1 to 2 which is of order  $1/\sqrt{n}$ , and since the number of transitions from 1 to 2 before the chain reaches 3 follows a geometric law with parameter  $a_n/(n + a_n)$ , the typical time to see a transition between 1 and 4 is of order  $\sqrt{n}/a_n$ . If  $a_n = n^{1/3}$ , for large  $n$ , it is unlikely to see such a transition before a given time  $T$  (independent of  $n$ ), so that the corresponding Markov process  $x_n$  is expected to converge to the deterministic solution of  $\dot{x} = A_1 x$ . If  $a_n = n^{2/3}$ , transitions between 1 and 4 get arbitrarily fast for large  $n$  and a fast averaging phenomenon leads to  $\dot{x} = [(A_1 + A_4)/2]x$  (the time spent in 2 and 3 being negligible). If  $a_n = \sqrt{n}$ , transitions

between 1 and 4 occur at a rate of order 1, so the limit process is  $\dot{x} = A_\sigma x$ , where  $\sigma$  is an irreducible Markov chain on  $\{1, 4\}$ .

Trying to adapt this analysis to a general chain leads to a recursive construction of several timescales at which different transitions occur. Such a rigorous construction is precisely the topic of the work [19] by Landim and Xu (itself based on [3] which deals with reversible Markov chains), upon which we will rely. Nevertheless, stated as they are, the results of [19] do not fully match our needs. For this reason, we slightly reformulate them below.

Let us check that the assumptions of [19] are satisfied (at least up to extracting a subsequence). The first one is that the chain is strongly connected for all  $n \in \mathbb{N}$ . This does not necessarily hold in our case, but we will be able to reduce the problem to this case by a standard argument, see the proof of Theorem 3.3. For this reason, we can suppose that  $P_n$  is strongly connected for all  $n \in \mathbb{N}$ .

Up to extracting a subsequence, we can suppose that

$$\forall (i, j) \in \llbracket 1, N \rrbracket^2, \quad \begin{cases} \text{either} & \lambda_n(i, j) = 0 \ \forall n \in \mathbb{N}, \\ \text{or} & \lambda_n(i, j) > 0 \ \forall n \in \mathbb{N}, \end{cases} \quad (5.1)$$

and we denote  $\mathbb{B} = \{(i, j) \in \llbracket 1, N \rrbracket^2 \mid \lambda_n(i, j) > 0 \ \forall n \in \mathbb{N}\}$ . Notice that, the chains being strongly connected, necessarily  $\mathbb{B} \neq \emptyset$ .

**Definition 5.3.** For  $r \geq 2$ , a family  $\{(a_n^i)_{n \in \mathbb{N}}\}_{i \in \llbracket 1, r \rrbracket}$  of positive sequences is said to be *ordered* if  $\arctan(a_n^i/a_n^j)$  converges as  $n \rightarrow +\infty$  for all  $i, j \in \llbracket 1, r \rrbracket$ .

For a pair of ordered positive sequences we write  $a_n \ll b_n$  (resp.,  $\simeq, \gg$ ) if  $a_n/b_n \rightarrow 0$  (resp.,  $1, +\infty$ ) as  $n \rightarrow +\infty$ .

Two positive sequences form an ordered pair up to extracting a subsequence, and the same is true for a finite family of sequences. As a consequence, denoting by  $\mathfrak{A}_m$ , for all  $m \in \mathbb{N}$ , the set of functions  $k : \mathbb{B} \rightarrow \mathbb{N}$  such that  $\sum_{(i,j) \in \mathbb{B}} k(i, j) = m$ , we see that, up to extracting a subsequence, by a diagonal argument,

$$\forall m \in \mathbb{N}, \quad \left\{ \left( \prod_{(i,j) \in \mathbb{B}} \lambda_n(i, j)^{k(i,j)} \right)_{n \in \mathbb{N}} \right\}_{k \in \mathfrak{A}_m} \quad \text{is ordered,} \quad (5.2)$$

which is [19, Assumption 2.6].

Let us now describe the consequences of this, established in [19]. The following result is an adaptation from Theorems 2.1, 2.7, and 2.12 of [19]. As such an adaptation requires the introduction of several definitions and notations, it is postponed to Appendix A.

**Theorem 5.4.** Consider for all  $n \in \mathbb{N}$  a strongly connected Markov chain  $(\sigma_n(t))_{t \geq 0}$  on  $\llbracket 1, N \rrbracket$  with jump rates  $(\lambda_n(i, j))_{i, j \in \llbracket 1, N \rrbracket}$ . Under conditions (5.1) and (5.2), there exist  $\mathfrak{p} \geq 1$ , a decreasing sequence  $n_1, \dots, n_{\mathfrak{p}+1}$  in  $\llbracket 1, N \rrbracket$ , a family of  $\mathfrak{p} + 1$  partitions  $\{\mathcal{E}_1^j, \dots, \mathcal{E}_{n_j}^j, \Delta^j\}$  of  $\llbracket 1, N \rrbracket$ ,  $j \in \llbracket 1, \mathfrak{p} + 1 \rrbracket$ , and  $\mathfrak{p}$  positive sequences  $\theta^j = (\theta_n^j)_{n \in \mathbb{N}}$ ,  $j \in \llbracket 1, \mathfrak{p} + 1 \rrbracket$ , with the following properties:

(a) The timescales  $\theta^j$  are increasing with respect to  $j$ , in the sense that for all  $j \in \llbracket 1, \mathfrak{p} - 1 \rrbracket$ ,  $\theta_n^j \ll \theta_n^{j+1}$ . Moreover, the fastest timescale  $\theta^1$  is obtained by taking  $1/\theta_n^1 = \sum_{i, j=1}^N \lambda_n(i, j)$  for all  $n \in \mathbb{N}$ .

(b) Denote, for all  $i \in \llbracket 1, N \rrbracket$  and  $j \in \llbracket 1, \mathfrak{p} + 1 \rrbracket$ ,

$$\Psi^j(i) = \sum_{x=1}^{n_j} x \mathbb{1}_{i \in \mathcal{E}_x^j},$$

which we call the coarse-grained variable at level  $j$ . For all  $j \in \llbracket 1, \mathfrak{p} \rrbracket$ , the transitions of  $\Psi^j(\sigma_n)$  occur at the timescale  $\theta^j$ , and are approximately Markovian, in the sense that there exists a Markov chain  $(X^j(t))_{t \geq 0}$  on  $\llbracket 1, n_j \rrbracket$  that has at least one nonzero jump rate and such that, for all  $x, y \in \llbracket 1, n_j \rrbracket$ ,  $i \in \mathcal{E}_x^j$ , and  $t > 0$ ,

$$\mathbb{P}_i(\Psi^j(\sigma_n(t\theta_n^j)) = y) \xrightarrow{n \rightarrow +\infty} \mathbb{P}_x(X^j(t) = y) \quad (5.3)$$

and, for all  $\delta > 0$ ,

$$\mathbb{P}_i\left(\left|\int_0^t \left(\mathbb{1}_{\Psi^j(\sigma_n(s\theta_n^j))=y} - \mathbb{1}_{X^j(s)=y}\right) ds\right| > \delta\right) \xrightarrow{n \rightarrow +\infty} 0. \quad (5.4)$$

More generally, (5.3) and (5.4) still hold if  $\theta_n^j$  is replaced by  $\tilde{\theta}_n^j$  with  $\tilde{\theta}_n^j \simeq \theta_n^j$ .

(c) There are no transitions between two successive timescales  $\theta^j$  and  $\theta^{j+1}$  or at a larger timescale than  $\theta^{\mathfrak{p}}$ , in the sense that, for all  $j \in \llbracket 1, \mathfrak{p} \rrbracket$ , for all positive sequence  $(\alpha_n)_{n \in \mathbb{N}}$  with  $\theta_n^j \ll \alpha_n \ll \theta_n^{j+1}$  (where, for  $j = \mathfrak{p}$ , we set  $\theta_n^{\mathfrak{p}+1} = +\infty$  for all  $n \in \mathbb{N}$ ), for all  $x, y \in \llbracket 1, n_{j+1} \rrbracket$ ,  $i \in \mathcal{E}_x^{j+1}$ , and  $t > 0$ ,

$$\mathbb{P}_i(\Psi^{j+1}(\sigma_n(t\alpha_n)) = y) \xrightarrow{n \rightarrow +\infty} \mathbb{1}_{x=y}$$

and for all  $\delta > 0$ ,

$$\mathbb{P}_i\left(\left|\int_0^t \mathbb{1}_{\Psi^{j+1}(\sigma_n(s\alpha_n))=y} ds - t \mathbb{1}_{x=y}\right| > \delta\right) \xrightarrow{n \rightarrow +\infty} 0.$$

(d) The time spent in  $\Delta^j$  at a scale larger than  $\theta^{j-1}$  is negligible, in the sense that for all  $j \in \llbracket 2, \mathfrak{p} \rrbracket$ , all positive sequence  $(\alpha_n)_{n \in \mathbb{N}}$  with  $\theta_n^{j-1} \ll \alpha_n$  and  $t \geq 0$ ,

$$\max_{i \in \llbracket 1, N \rrbracket} \mathbb{E}_i\left(\int_0^t \mathbb{1}_{\sigma_n(s\alpha_n) \in \Delta^j} ds\right) \xrightarrow{n \rightarrow +\infty} 0.$$

(e) For  $j = 1$ , we have  $n_1 = N$ ,  $\mathcal{E}_i^1 = \{i\}$  for every  $i \in \llbracket 1, N \rrbracket$  (i.e., all the points are separated), and  $\Delta^1 = \emptyset$ . Then, the partitions get coarser and are given by the recurrence classes of the limit chains. More precisely, for all  $j \in \llbracket 1, \mathfrak{p} \rrbracket$ , the limit chain  $X^j$  admits  $n_{j+1}$  recurrence classes  $\mathcal{C}_1^j, \dots, \mathcal{C}_{n_{j+1}}^j$ , and for all  $x \in \llbracket 1, n_{j+1} \rrbracket$ ,  $\mathcal{E}_x^{j+1} = \bigcup_{y \in \mathcal{C}_x^j} \mathcal{E}_y^j$ . Similarly, the set  $\Delta^j$  is increasing, and is given by the transient points of  $X^j$ . More precisely, denoting  $\mathcal{T}_j$  the set of transient points of  $X^j$ , then  $\Delta^{j+1} = \Delta^j \cup \left(\bigcup_{y \in \mathcal{T}_j} \mathcal{E}_y^j\right)$ . In particular, for all  $j \in \llbracket 1, \mathfrak{p} + 1 \rrbracket$  and all  $x \in \llbracket 1, n_j \rrbracket$ ,  $\mathcal{E}_x^j \neq \emptyset$  (while possibly  $\Delta^j = \emptyset$ ). The last partition is trivial, in the sense that  $n_{\mathfrak{p}+1} = 1$ .

(f) For all  $j \in \llbracket 1, \mathfrak{p} + 1 \rrbracket$  and all  $x \in \llbracket 1, \mathfrak{n}_j \rrbracket$ , consider the escape time  $\tau_x^j = \inf\{t > 0 \mid \sigma_n(t) \notin \mathcal{E}_x^j \cup \Delta^j\}$ . For all  $j \in \llbracket 1, \mathfrak{p} \rrbracket$ ,  $x \in \llbracket 1, \mathfrak{n}_j \rrbracket$ , and all initial conditions  $i \in \mathcal{E}_x^j$ ,  $\tau_x^j / \theta_n^j$  converges in law towards an exponential distribution with some parameter  $r \geq 0$  (where  $r = 0$  means that  $\tau_x^j / \theta_n^j \rightarrow +\infty$  in probability). For  $j = \mathfrak{p} + 1$ , for all  $x \in \llbracket 1, \mathfrak{n}_{\mathfrak{p}+1} \rrbracket$  and all initial conditions  $i \in \mathcal{E}_x^{\mathfrak{p}+1}$ , almost surely  $\tau_x^{\mathfrak{p}+1} = +\infty$ .

**Remark 5.5.** To help clarify the notations introduced in Theorem 5.4, its construction is illustrated in Figure 5.2.

Figure 5.2 (a) illustrates the first step of the construction, representing in blue the states of  $\sigma_n$ , i.e., the elements of  $\llbracket 1, N \rrbracket$  identified with  $\mathcal{E}_1^1, \dots, \mathcal{E}_N^1$ . We only represent the fastest transitions of  $\sigma_n$  (black arrows), i.e., the transition whose rates are of the same order as the total jump rate of  $\sigma_n$ , which we call  $1/\theta_n^1$ . This defines recurrence classes (in dashed lines) and transient states. We call  $\Delta^2$  the set of transient states, and  $\mathcal{E}_1^2, \mathcal{E}_2^2, \mathcal{E}_3^2, \mathcal{E}_4^2$  the recurrence classes. After rescaling the time by  $\theta_n^1$ ,  $\sigma_n$  converges to some Markov chain  $X^1$  on  $\llbracket 1, N \rrbracket$ .

Figure 5.2 (b) represents the second step of the construction. At a timescale larger than  $\theta_n^1$ , the time spent in  $\Delta^2$ , the transient states of  $X^1$ , is negligible and  $\sigma_n$  is averaged within the recurrence classes, so that we can approximately consider that the “macroscopic states” are the recurrence classes  $\mathcal{E}_1^2, \mathcal{E}_2^2, \mathcal{E}_3^2, \mathcal{E}_4^2$  (corresponding to the coarse-grained variable  $\Psi^2(\sigma_n)$ ). We can now consider the first timescale  $\theta_n^2$  at which transitions between these macroscopic states (the black arrows in Figure 5.2 (b)) occur. Remark that, during any such a transition,  $\sigma_n$  may have to cross  $\Delta^2$ . These transitions define a new Markov chain over the macroscopic states. More precisely, after rescaling the time by  $\theta_n^2$ ,  $\Psi^2(\sigma_n)$  converges to some Markov chain  $X^2$  on  $\{1, 2, 3, 4\}$ . In this example, the recurrence classes of  $X^2$  are  $\mathcal{C}_1^2 = \{1, 2\}$  and  $\mathcal{C}_2^2 = \{4\}$ , corresponding to the sets  $\mathcal{E}_1^3 = \mathcal{E}_1^2 \cup \mathcal{E}_2^2$  and  $\mathcal{E}_2^3 = \mathcal{E}_4^2$  (in dashed lines). The last macroscopic state,  $\mathcal{E}_3^2$ , is transient for  $X^2$ , so  $\Delta^3 = \Delta^2 \cup \mathcal{E}_3^2$  is the set of states that are negligible for any timescale larger than  $\theta_n^2$ .

The third step of the construction is represented in Figure 5.2 (c). At a timescale larger than  $\theta_n^2$ , the time spent in  $\Delta^3$  is negligible and the chain  $\sigma_n$  is averaged within either  $\mathcal{E}_1^3$  or  $\mathcal{E}_2^3$ . So, at this scale, there are two macroscopic states. The next timescale is given by the transitions between them (afterwards, only one class remains and the construction stops, in other words here  $\mathfrak{p} = 3$ ). Besides, we are not interested in these transitions if they occur at a timescale larger than  $\mathcal{O}(1)$  since, in this case, they are not seen in the limit convexified process.

From now on, we suppose that the sequence  $(v_n, \mu_n, P_n)_{n \in \mathbb{N}}$  is such that  $\mu_n \rightarrow +\infty$ , that  $P_n$  is strongly connected for all  $n \in \mathbb{N}$ , and that (5.1) and (5.2) hold, so that Theorem 5.4 holds. Up to extracting a subsequence, we assume that for all  $j \in \llbracket 1, \mathfrak{p} \rrbracket$ , the sequence  $\theta^j$  is monotone (and in particular admits a limit in  $[0, +\infty]$ ). Since we assumed that  $\mu_n \rightarrow +\infty$ , necessarily  $\theta_n^1 \rightarrow 0$ . Let  $h = \max\{j \in \llbracket 1, \mathfrak{p} \rrbracket \mid \theta_n^j \rightarrow 0\}$ , so that  $\theta^h$  is the slowest of all the fast scales of the chain. We now have to distinguish whether there are slow transitions (i.e., occurring at a time of order 1 with respect to  $n$ ) or not.

- **Case 1.** If  $h = \mathfrak{p}$  then by Theorem 5.4 (c), for any sequence  $\alpha_n \simeq 1 \gg \theta_n^{\mathfrak{p}}$ , for all  $t > 0$ ,  $\Psi^{h+1}(\sigma_n(t\alpha_n))$  converges in law to the value at time  $t$  of the constant Markov chain on  $\{1\}$  (since, by Theorem 5.4(e),  $\mathfrak{n}_{\mathfrak{p}+1} = 1$ ).

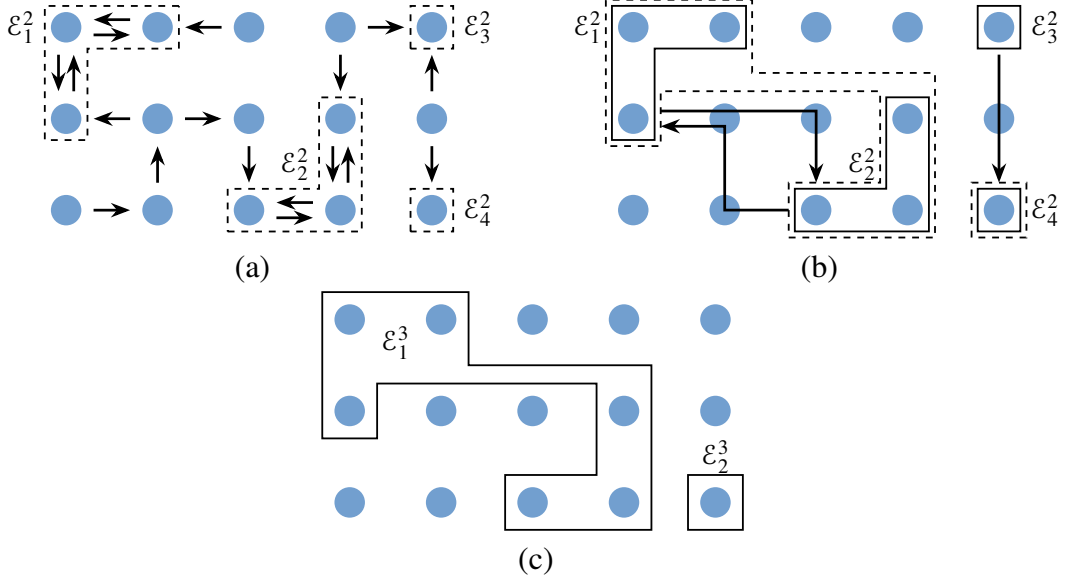


Figure 5.2: Construction of Theorem 5.4, described in Remark 5.5.

- **Case 2.** If  $h < p$  and  $\theta_n^{h+1}$  does not converge to  $+\infty$  as  $n \rightarrow +\infty$ , then it converges to some  $\theta_* > 0$ . In particular, for all  $\alpha_n \simeq 1$  and  $t > 0$ ,  $\Psi^{h+1}(\sigma_n(t\alpha_n))$  converges in law towards  $X^{h+1}(t/\theta_*)$ .
- **Case 3.** If  $h < p$  and  $\theta_n^{h+1} \rightarrow +\infty$  as  $n \rightarrow +\infty$  then, as in the first case, by Theorem 5.4 (c), for any sequence  $(\alpha_n)_{n \in \mathbb{N}}$  with  $\theta_n^h \ll \alpha_n \simeq 1 \ll \theta_n^{h+1}$ , for all  $t > 0$ ,  $\Psi^{h+1}(\sigma_n(t\alpha_n))$  converges in law to the value at time  $t$  of the constant Markov chain on  $\llbracket 1, n_{h+1} \rrbracket$ .

Summarizing the above three cases, there is a Markov chain on  $\llbracket 1, n_{h+1} \rrbracket$ , that we denote  $(\tilde{X}^{h+1}(t))_{t \geq 0}$ , such that, for all sequence  $\alpha_n \simeq 1$ ,  $t > 0$ ,  $x, y \in \llbracket 1, n_{h+1} \rrbracket$  and  $i \in \mathcal{E}_x^{h+1}$ ,

$$\mathbb{P}_i\left(\Psi^{h+1}(\sigma_n(t\alpha_n)) = y\right) \xrightarrow{n \rightarrow +\infty} \mathbb{P}_x\left(\tilde{X}^{h+1}(t) = y\right).$$

Reasoning similarly, we can ensure, in addition, that for all  $\delta > 0$ ,

$$\mathbb{P}_i\left(\left|\int_0^t \left(\mathbb{1}_{\Psi^{h+1}(\sigma_n(s\alpha_n))=y} - \mathbb{1}_{\tilde{X}^{h+1}(s)=y}\right) ds\right| > \delta\right) \xrightarrow{n \rightarrow +\infty} 0. \quad (5.5)$$

Moreover, from Theorem 5.4 (d), for all sequences  $\alpha_n \simeq 1$  and  $t \geq 0$ ,

$$\max_{i \in \llbracket 1, M \rrbracket} \mathbb{E}_i\left(\int_0^t \mathbb{1}_{\sigma_n(s\alpha_n) \in \Delta^{h+1}} ds\right) \xrightarrow{n \rightarrow +\infty} 0. \quad (5.6)$$

By analogy with the proof of Proposition 5.1, the transitions of  $\tilde{X}^{h+1}$  will play the role of the slow transition (at rate  $\mu$  in Proposition 5.1). It remains to identify a Markov chain that plays the same role as  $\tilde{\sigma}_n^b$  in the proof of Proposition 5.1, namely a Markov chain that is equal to  $\sigma_n$  up to the first slow transition and that mixes fast within the sets  $\mathcal{E}_x^{h+1}$ ,  $x \in \llbracket 1, n_{h+1} \rrbracket$ . We could consider the trace process of  $\sigma_n$  on  $\mathcal{E}_x^{h+1}$  (see Section A for the definition), but for simplicity (in order to avoid the question of time change) we will consider another process.

For all  $z \in \llbracket 1, n_{h+1} \rrbracket$ , fix some  $i_z \in \mathcal{E}_z^{h+1}$  (for instance,  $i_z = \min \mathcal{E}_z^{h+1}$ ). We denote by  $\tilde{\sigma}_n^z$  the Markov chain on  $\bar{\mathcal{E}}_z^{h+1} := \mathcal{E}_z^{h+1} \cup \Delta^{h+1}$  with jump rates  $\tilde{\lambda}_n$  defined as follows:

- $\tilde{\lambda}_n(i, j) = \lambda_n(i, j)$  for all  $i, j \in \bar{\mathcal{E}}_z^{h+1}$  with  $j \neq i_z$ ;
- $\tilde{\lambda}_n(i, i_z) = \lambda_n(i, i_z) + \sum_{k \notin \bar{\mathcal{E}}_z^{h+1}} \lambda_n(i, k)$  for all  $i \in \bar{\mathcal{E}}_z^{h+1}$ .

In other words, starting at a point of  $\bar{\mathcal{E}}_z^{h+1}$ ,  $\tilde{\sigma}_n^z$  has the same transitions as  $\sigma_n$  except that all transitions that would leave  $\bar{\mathcal{E}}_z^{h+1}$  are replaced by a transition to  $i_z$ . By Theorem 5.4 (f), from an initial condition in  $\mathcal{E}_z^{h+1}$ , the first transition of  $\sigma_n$  out of  $\bar{\mathcal{E}}_z^{h+1}$  occurs at a rate at most of order 1 (in  $n$  as  $n \rightarrow \infty$ ). On the other hand, let us prove that the mixing time of  $\tilde{\sigma}_n^z$  is of order  $\theta_n^h$  (in the sense of Equation (5.7) below). Denote by  $(P_t^{n,z})_{t \geq 0}$  the semigroup associated with  $\tilde{\sigma}_n^z$ , i.e.,  $P_t^{n,z}(i, j) = \mathbb{P}(\tilde{\sigma}_n^z(t) = j | \tilde{\sigma}_n^z(0) = i)$  for all  $i, j \in \bar{\mathcal{E}}_z^{h+1}$ .

**Proposition 5.6.** *There exist  $\rho, C > 0$  such that for all  $z \in \llbracket 1, n_{h+1} \rrbracket$  and all  $n \in \mathbb{N}$ ,  $(P_t^{n,z})_{t \geq 0}$  admits a unique invariant probability measure  $\pi_n^z$  on  $\bar{\mathcal{E}}_z^{h+1}$  and, for all  $t \geq 0$  and all  $i \in \bar{\mathcal{E}}_z^{h+1}$ ,*

$$|P_t^{n,z}(i, \cdot) - \pi_n^z| \leq C e^{-\rho t / \theta_n^h}. \quad (5.7)$$

*Proof.* The statement does not depend on the norm used in the left hand side of (5.7). In the following, we consider the total variation distance  $|x| = \sum_{i=1}^N |x_i|$ .

Since there is a finite number of values  $z \in \llbracket 1, n_{h+1} \rrbracket$ , it is sufficient to prove the result for each one, so let  $z$  be fixed.

Assuming that  $P_{4\theta_n^h}^{n,z}$  satisfies the Doeblin condition

$$\liminf_{n \rightarrow +\infty} \min_{i \in \bar{\mathcal{E}}_z^{h+1}} P_{4\theta_n^h}^{n,z}(i, i_z) > 0, \quad (5.8)$$

we deduce the conclusion of the proposition by the following classical argument. By (5.8),  $P^{n,z}$  has a unique recurrence class, since all points can go to  $i_z$ , and thus a unique invariant measure  $\pi_n^z$  (supported by this recurrence class). Let  $\alpha \in (0, 1)$  be such that, for  $n$  large enough,

$$\min_{i \in \bar{\mathcal{E}}_z^{h+1}} P_{4\theta_n^h}^{n,z}(i, i_z) \geq \alpha.$$

For the total variation, we have (see e.g. [11, Theorem S7] with  $V = 1$ )

$$\sup_{i \in \bar{\mathcal{E}}_z^{h+1}} |P_t^{n,z}(i, \cdot) - \pi_n^z| \leq 2(1 - \alpha)^{\lfloor t / (4\theta_n^h) \rfloor} \leq \frac{2}{1 - \alpha} e^{t \ln(1 - \alpha) / (4\theta_n^h)},$$

which concludes the proof of (5.7).

We are left to prove that (5.8) holds true and for that purpose we introduce some useful notations. In the sequel we say that a sequence  $(a_n)_{n \in \mathbb{N}}$  is nonvanishing if  $\liminf_{n \rightarrow +\infty} a_n > 0$ . Let  $x_1, \dots, x_{h+1}$  be such that

$$i_z \in \mathcal{E}_{x_1}^1 \subset \dots \subset \mathcal{E}_{x_{h+1}}^{h+1}.$$



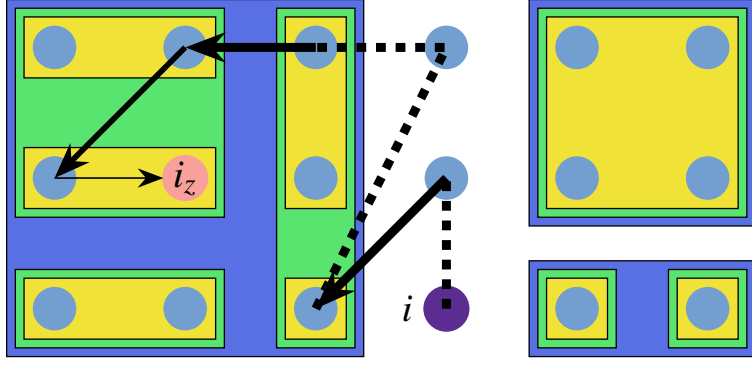


Figure 5.3: At the timescale  $\theta^1$ , the state  $i_z$ , which is also  $\mathcal{E}_{i_z}^1$ , is averaged in some recurrence class  $\mathcal{E}_{x_2}^2$  (in yellow). Then, at the slower timescale  $\theta^2$ , this class is itself averaged with other yellow classes, in a coarser  $\mathcal{E}_{x_3}^3$  (in green) which is itself averaged at the slower timescale  $\theta^3$  with some other green classes to give a coarser  $\mathcal{E}_{x_4}^4$  (in blue). Here,  $h = 3$ , i.e., there is no more fast averaging (there may be slow transitions at order 1 between several blue classes, but we do not consider them here). Now, starting from some point  $i$  (in  $\Delta^4$  here), after a time  $\theta^3$ , the chain  $\sigma_n$  has a positive (nonvanishing) probability to be in one of the blue classes (and thus  $\tilde{\sigma}_n^z$  to be in the blue class of  $i_z$ , since transitions to other blue classes are replaced by a transition to  $i_z$ ; and after reaching this blue class the transitions of  $\sigma_n$  and  $\tilde{\sigma}_n^z$  are similar in the timescale  $\theta^3$  since a transition to a different blue class is unlikely). Then, starting in the blue class of  $i_z$ , i.e., in some green class within  $\mathcal{E}_{x_4}^4$ , the chain has a positive probability to be in  $\mathcal{E}_{x_3}^3$ , the green class of  $i_z$ , in a time  $\theta^3$  (or equivalently  $\theta^3 - \theta^2$ ) as this is the timescale of motion between green classes and all green classes within a given blue class are linked at this timescale (by definition of the blue classes as recurrence classes). Similarly, from there, the chain has a positive probability to be in  $\mathcal{E}_{x_2}^2$ , the yellow class of  $i_z$ , in a time  $\theta^2$  (or  $\theta^2 - \theta^1$ ). Finally, starting from any point in  $\mathcal{E}_{x_2}^2$ , it has a positive probability to be at  $i_z$  in a time  $\theta^1$ . The arrows represent the transitions (the timescale being represented by the thickness of the arrow). The dotted lines are here to remind that a “macroscopic transition” happening between two classes (or from the point  $i \in \Delta^4$  to some point of the blue class  $\mathcal{E}_{x_4}^4$ ) may be constituted of many (possibly fastest) “microscopic” transitions (that may cross  $\Delta^4$ , which explains why  $\tilde{\sigma}_n^z$  is defined on  $\bar{\mathcal{E}}_{x_4}^4$  and not on  $\mathcal{E}_{x_4}^4$ ).

In other words,  $x_1 = i_z$ ,  $x_{h+1} = z$ , and for all  $j \in \llbracket 1, h+1 \rrbracket$  and  $n \in \mathbb{N}$ ,  $x_j = \Psi^j(i_z)$ . Denote by  $(P_t^n)_{t \geq 0}$  the semigroup associated with  $\sigma_n$ . Also, set  $\theta_n^0 = 0$  for all  $n \in \mathbb{N}$ . The strategy of the proof of (5.8) consists in finding a path from any  $i \in \bar{\mathcal{E}}_z^{h+1}$  to  $i_z$  that has a nonvanishing probability. An informal description of such a path is given in Figure 5.3.

We decompose the argument in several steps.

**Step 1.** Let us prove that, from any initial condition in  $\bar{\mathcal{E}}_z^{h+1}$ , the probability that  $\tilde{\sigma}_n^z$  hits  $\mathcal{E}_z^{h+1}$  before the time  $\theta_n^h$  is nonvanishing. More precisely denote

$$\tau_* = \inf\{t \geq 0 \mid \sigma_n(t) \notin \Delta^{h+1}\} = \inf\{t \geq 0 \mid \tilde{\sigma}_n^z(t) \in \mathcal{E}_z^{h+1}\},$$

where  $\sigma_n$  and  $\tilde{\sigma}_n^z$  are synchronously coupled (as in the proof of Proposition 5.2, see [12, Section 6] for details) with the same initial condition. It is clear that, for all initial condi-

tions  $i \in \Delta^{h+1}$ ,

$$\liminf_{n \rightarrow +\infty} \mathbb{P}_i(\tau_* \leq \theta_n^h) > 0. \quad (5.9)$$

Indeed, the time spent in  $\Delta^h$  is negligible at the timescale  $\theta_n^h$  (Theorem 5.4 (d)), so the probability for  $\sigma_n$  to leave  $\Delta^h$  before time  $\theta_n^h/2$  goes to 1 as  $n \rightarrow +\infty$ . Then, starting from a point  $i' \in \Delta^{h+1} \setminus \Delta^h$  (which corresponds to the transient points of  $X^h$ ), we can use Theorem 5.4 (b) to see that

$$\mathbb{P}_{i'}(\tau_* \leq \theta_n^h/2) \geq \mathbb{P}_{i'}(\Psi^h(\sigma_n(\theta_n^h/2)) \notin \mathcal{T}_h \cup \{0\}) \xrightarrow{n \rightarrow +\infty} \mathbb{P}_{\Psi^h(i')} (X^h(1/2) \notin \mathcal{T}_h) > 0$$

(where we recall that, by definition,  $\Psi^h(\sigma) = 0$  means that  $\sigma \in \Delta^h$ ). The strong Markov property concludes the proof of (5.9).

**Step 2.** Starting from an initial condition  $i \in \mathcal{E}_z^{h+1}$ , we consider again the synchronous coupling of  $\sigma_n$  and  $\tilde{\sigma}_n^z$ . The two chains are then equal up to the first time at which  $\sigma_n$  hits  $\mathcal{E}_y^{h+1}$  for some  $y \neq z$ . According to Theorem 5.4 (f), this occurs at a time of order at least 1. In other words, for any  $M > 0$ , the probability that this happens before time  $M\theta_n^h$  goes to 0 as  $n \rightarrow +\infty$ . Hence, for all  $M > 0$ ,

$$\sup_{s \in [0, M\theta_n^h]} |P_s^n(i, u) - P_s^{n,z}(i, u)| \leq \mathbb{P}_i(\exists s \in [0, M\theta_n^h], \sigma_n(s) \neq \tilde{\sigma}_n^z(s)) \xrightarrow{n \rightarrow +\infty} 0.$$

For this reason, we can focus on  $\sigma_n$  rather than  $\tilde{\sigma}_n^z$ .

**Step 3.** Let us prove that, for all  $j \in \llbracket 1, h \rrbracket$ ,

$$\liminf_{n \rightarrow +\infty} \min_{i \in \mathcal{E}_{x_{j+1}}^{j+1}} \sum_{u \in \mathcal{E}_{x_j}^j} P_{\theta_n^j - \theta_n^{j-1}}^n(i, u) > 0. \quad (5.10)$$

Indeed, for all  $s > 0$ ,  $n \in \mathbb{N}$ ,

$$\sum_{u \in \mathcal{E}_{x_j}^j} P_s^n(i, u) = \mathbb{P}_i(\Psi^j(\sigma_n(s)) = x_j).$$

Since  $\theta_n^j - \theta_n^{j-1} \simeq \theta_n^j$ , Theorem 5.4 (b) implies that for all  $y \in \llbracket 1, n_j \rrbracket$  and  $i \in \mathcal{E}_y^j$ ,

$$\sum_{u \in \mathcal{E}_{x_j}^j} P_{\theta_n^j - \theta_n^{j-1}}^n(i, u) \xrightarrow{n \rightarrow +\infty} \mathbb{P}_y(X^j(1) = x_j).$$

Recall from Theorem 5.4 (e) that  $\mathcal{E}_{x_{j+1}}^{j+1} = \bigcup_{y \in \mathcal{C}_{x_{j+1}}^j} \mathcal{E}_y^j$ , where  $\mathcal{C}_{x_{j+1}}^j$  is the  $x_{j+1}$ -th recurrence class of the limit chain  $X^j$ . In particular,  $x_j \in \mathcal{C}_{x_{j+1}}^j$  and, by definition of a recurrence class, for all  $x, y \in \mathcal{C}_{x_{j+1}}^j$ ,  $\mathbb{P}_x(X^j(1) = y) > 0$ . This concludes the proof of (5.10).

**Step 4.** We now deduce from (5.10) that, for all  $j \in \llbracket 1, h \rrbracket$ ,

$$\liminf_{n \rightarrow +\infty} \min_{i \in \mathcal{E}_{x_{j+1}}^{j+1}} P_{\theta_n^j}^n(i, i_z) > 0. \quad (5.11)$$

Indeed, reasoning by induction, the case  $j = 1$  is just (5.10) with  $j = 1$  since  $\mathcal{E}_{x_1}^1 = \{i_z\}$  and  $\theta_n^0 = 0$ . Then, using the Markov property, for  $i \in \mathcal{E}_{x_{j+1}}^{j+1}$ ,

$$P_{\theta_n^j}^n(i, i_z) \geq \sum_{u \in \mathcal{E}_{x_j}^j} P_{\theta_n^j - \theta_n^{j-1}}^n(i, u) P_{\theta_n^{j-1}}^n(u, i_z),$$

and the conclusion follows from (5.10) and, by induction, (5.11) for  $j - 1$ .

**Step 5.** Let us prove by induction on  $j$  that, for all  $j \in \llbracket 1, h \rrbracket$ ,

$$\forall M > m > 1, \quad \liminf_{n \rightarrow +\infty} \inf_{s \in [m, M]} P_{s\theta_n^j}^n(i_z, i_z) > 0. \quad (5.12)$$

For  $j = 1$ , this follows from the fact that  $1/\theta_n^1$  is of the same order as the total jump rate  $\mu_n$  of the chain, so that for all  $M > 0$  there is a nonvanishing probability that no jump occurs in  $[0, M\theta_n^1]$ , and then

$$\inf_{s \in [0, M]} P_{s\theta_n^1}^n(i_z, i_z) \geq \mathbb{P}_{i_z}(\sigma_n(t) = i_z \forall t \in [0, M\theta_n^1]).$$

Now, suppose that (5.12) is true for  $j - 1$  for some  $j \in \llbracket 2, h \rrbracket$ . Fix  $M > m > 1$ . For  $M' > 0$  to be chosen later on, consider the events

$$E_n = \left\{ \sigma_n(t) \in \mathcal{E}_{x_j}^j \cup \Delta^j \forall t \in [0, M\theta_n^j] \right\}, \quad F_n = \left\{ \tau' \leq M'\theta_n^{j-1} \right\},$$

where  $\tau' = \inf\{t \geq 0 \mid \sigma_n(t) \notin \Delta^j\}$ . From Theorem 5.4 (f),

$$\varepsilon := \liminf_{n \rightarrow +\infty} \mathbb{P}_{i_z}(E_n) > 0.$$

Besides, from Theorem 5.4 (b) and (e), for all  $i \in \Delta^j \setminus \Delta^{j-1}$ ,

$$\begin{aligned} \liminf_{n \rightarrow +\infty} \mathbb{P}_i(\tau' \leq M'\theta_n^{j-1}) &\geq \liminf_{n \rightarrow +\infty} \mathbb{P}_i(\sigma_n(M'\theta_n^{j-1}) \notin \Delta^j) \\ &\geq \min_{y \in \mathcal{J}^j} \mathbb{P}_y(X^j(M') \notin \mathcal{J}^j) \xrightarrow{M' \rightarrow +\infty} 1, \end{aligned}$$

and from Theorem 5.4 (d), for all  $i \in \Delta^{j-1}$ ,

$$\mathbb{P}_i(\sigma_n(t) \in \Delta^{j-1} \forall t \in [0, \theta_n^{j-1}]) \xrightarrow{n \rightarrow +\infty} 0,$$

so that, combining these two facts thanks to the Markov property and denoting  $q_n = \min_{i \in \llbracket 1, N \rrbracket} \mathbb{P}_i(\tau' \leq M'\theta_n^{j-1})$ , we get

$$\liminf_{n \rightarrow +\infty} q_n \xrightarrow{M' \rightarrow +\infty} 1.$$

From now on, we choose  $M'$  large enough so that  $q_n \geq 1 - \varepsilon/2$  for all  $n \geq n_0$  for some sufficiently large  $n_0$ . We also suppose that  $n_0$  is large enough so that  $(m - 1)\theta_n^j - (2 + M')\theta_n^{j-1} > 0$  for all  $n \geq n_0$ , and implicitly up to the end of Step 4 we always assume  $n \geq n_0$ .

For  $s \in [m, M]$  and all  $n \geq n_0$  consider the time interval  $I_n(s) = [(s-1)\theta_n^j - (2+M')\theta_n^{j-1}, (s-1)\theta_n^j - 2\theta_n^{j-1}]$  and the event

$$F_n(s) = \{\exists u \in I_n(s) \mid \sigma_n(u) \notin \Delta^j\}.$$

By the Markov property,  $\mathbb{P}_i(F_n(s)) \geq q_n$  for all  $i \in \llbracket 1, N \rrbracket$  and  $s \in [m, M]$ . Moreover, the event  $F_n(s) \cap E_n$ , which has a probability at least  $\varepsilon/3$  for  $n$  large enough, implies that there exists  $u \in I_n(s)$  such that  $\sigma_n(u) \in \mathcal{E}_{x_j}^j$ . Conditioning with respect to  $u$  and  $\sigma_n(u)$  for such a time  $u$ , we get by the Markov property that, for all  $s \in [m, M]$ ,

$$P_{s\theta_n^j}^n(i_z, i_z) \geq \mathbb{P}_{i_z}(F_n(s) \cap E_n) \min_{i \in \mathcal{E}_{x_j}^j} P_{\theta_n^j}^n(i, i_z) \inf_{s' \in [2, 2+M']} P_{s'\theta_n^{j-1}}^n(i_z, i_z).$$

Thanks to the induction hypothesis, (5.11), and the bound on  $q_n$ , we see that the three factors of the right-hand side are bounded from below by positive quantities independent of  $s \in [m, M]$ , which concludes the proof of (5.12) for all  $j \in \llbracket 1, h \rrbracket$ .

**Step 6.** We now prove the Doeblin condition (5.8). From Step 2, (5.11) and (5.12) are still true if  $P^n$  is replaced by  $P^{n,z}$ . According to Step 1, from any initial condition  $i \in \llbracket 1, N \rrbracket$ ,  $\tilde{\sigma}_n^z$  has a nonvanishing probability to hit  $\mathcal{E}_z^{h+1}$  before time  $\theta_n^h$ . Using the Markov property and (5.11) (with  $P^n$  replaced by  $P^{n,z}$ ), conditioning with respect to this hitting time  $u$ , the chain has a nonvanishing probability (independent of  $u$ ) to be at  $i_z$  at time  $u + \theta_n^h$ . Then, from (5.12) (with  $P^n$  replaced by  $P^{n,z}$ ), it has a nonvanishing probability (still independent of  $u$ ) to be at  $i_z$  at time  $4\theta_n^z$ , which concludes.  $\square$

Before stating the main result of the section, let us prove an additional technical property.

**Proposition 5.7.** *For all  $T > 0$ , all  $z \in \llbracket 1, n_{h+1} \rrbracket$ , and all  $j \in \mathcal{E}_z^{h+1}$ ,*

$$\max_{i \in \llbracket 1, M \rrbracket} \mathbb{E}_i \left( \left| \int_0^T \left( \mathbb{1}_{\sigma_n(s)=j} - \pi_n^z(j) \mathbb{1}_{\sigma_n(s) \in \mathcal{E}_z^{h+1}} \right) ds \right| \right) \xrightarrow{n \rightarrow +\infty} 0.$$

*Proof.* Fix  $T > 0$ ,  $z \in \llbracket 1, n_{h+1} \rrbracket$ , and  $j \in \mathcal{E}_z^{h+1}$ . Let  $R \in \mathbb{N}$ , and notice that

$$\left| \int_0^T \left( \mathbb{1}_{\sigma_n(s)=j} - \pi_n^z(j) \mathbb{1}_{\sigma_n(s) \in \mathcal{E}_z^{h+1}} \right) ds \right| \leq \sum_{b=0}^{R-1} \left| \int_{bT/R}^{(b+1)T/R} \left( \mathbb{1}_{\sigma_n(s)=j} - \pi_n^z(j) \mathbb{1}_{\sigma_n(s) \in \mathcal{E}_z^{h+1}} \right) ds \right|.$$

For all  $b \in \llbracket 0, R-1 \rrbracket$ , consider the stopping time  $\tau_b = \inf\{t \geq bT/R \mid \sigma_n(t) \notin \Delta^{h+1}\}$  and  $Z_b = \Psi^{h+1}(\sigma_n(\tau_b))$ . Let  $\tau_b^+ = \inf\{t \geq \tau_b \mid \sigma_n(t) \in \bigcup_{y \neq Z_b} \mathcal{E}_y^{h+1}\}$ . According to Theorem 5.4 (f), for all  $b \in \llbracket 0, R-1 \rrbracket$ ,  $\tau_b^+ - \tau_b$  converges as  $n \rightarrow +\infty$  towards an exponential distribution with some finite rate (possibly 0). Thus there exist  $K > 0$  such that

$$\max_{i \in \llbracket 1, M \rrbracket} \mathbb{P}_i(\tau_b^+ - \tau_b \leq T/R) \leq \frac{K}{R}$$

for all  $n$  large enough and all  $R \in \mathbb{N}$ .

For each  $b \in \llbracket 0, R-1 \rrbracket$ , consider a Markov chain  $\tilde{\sigma}_n^{Z_b}$  and its associated semigroup  $P^{n, Z_b}$ , initialized at time  $\tau_b$  by  $\tilde{\sigma}_n^{Z_b}(\tau_b) = \sigma_n(\tau_b)$  and such that  $\tilde{\sigma}_n^{Z_b}(s) = \sigma_n(s)$  for all  $s \in [\tau_b, \tau_b^+)$ . We have

$$\begin{aligned} & \left| \int_{bT/R}^{(b+1)T/R} \left( \mathbb{1}_{\sigma_n(s)=j} - \pi_n^z(j) \mathbb{1}_{\sigma_n(s) \in \mathcal{E}_z^{h+1}} \right) ds \right| \leq \frac{T}{R} \mathbb{1}_{\tau_b \geq (b+1)T/R} \\ & + \mathbb{1}_{\tau_b < (b+1)T/R} \left( \tau_b - \frac{bT}{R} + \left| \int_{\tau_b}^{(b+1)T/R} \left( \mathbb{1}_{\sigma_n(s)=j} - \pi_n^z(j) \mathbb{1}_{\sigma_n(s) \in \mathcal{E}_z^{h+1}} \right) ds \right| \right) \\ & \leq \frac{T}{R} \left( \mathbb{1}_{\tau_b \geq (b+1)T/R} + \mathbb{1}_{\tau_b^+ - \tau_b \leq T/R} \right) \\ & + \mathbb{1}_{\tau_b < (b+1)T/R} \left( \tau_b - \frac{bT}{R} + \left| \int_{\tau_b}^{(b+1)T/R} \left( \mathbb{1}_{\tilde{\sigma}_n^{Z_b}(s)=j} - \pi_n^z(j) \mathbb{1}_{\tilde{\sigma}_n^{Z_b}(s) \in \mathcal{E}_z^{h+1}} \right) ds \right| \right) \\ & \leq \frac{T}{R} \left( \mathbb{1}_{\tau_b \geq (b+1)T/R} + \mathbb{1}_{\tau_b^+ - \tau_b \leq T/R} \right) \\ & + 2 \mathbb{1}_{\tau_b < (b+1)T/R} \left( \tau_b - \frac{bT}{R} \right) + \left| \int_{\tau_b}^{\tau_b + T/R} \left( \mathbb{1}_{\tilde{\sigma}_n^{Z_b}(s)=j} - \pi_n^z(j) \mathbb{1}_{\tilde{\sigma}_n^{Z_b}(s) \in \mathcal{E}_z^{h+1}} \right) ds \right|. \end{aligned}$$

We bound separately the expectations of these terms. First, for all  $i \in \llbracket 1, N \rrbracket$ ,

$$\begin{aligned} \mathbb{E}_i \left( \mathbb{1}_{\tau_b < (b+1)T/R} \left( \tau_b - \frac{bT}{R} \right) \right) & \leq \mathbb{E}_i \left( \int_{bT/R}^{(b+1)T/R} \mathbb{1}_{\sigma_n(s) \in \Delta^{h+1}} ds \right) \\ & \leq \sup_{u \in \llbracket 1, N \rrbracket} \mathbb{E}_u \left( \int_0^{T/R} \mathbb{1}_{\sigma_n(s) \in \Delta^{h+1}} ds \right), \end{aligned}$$

which vanishes as  $n \rightarrow +\infty$  (this is (5.6)). Similarly,

$$\begin{aligned} \mathbb{P}_i(\tau_b \geq (b+1)T/R) & = \mathbb{P}_i \left( \int_{bT/R}^{(b+1)T/R} \mathbb{1}_{\sigma_n(s) \in \Delta^{h+1}} ds = \frac{T}{R} \right) \\ & \leq \frac{R}{T} \sup_{u \in \llbracket 1, N \rrbracket} \mathbb{E}_u \left( \int_0^{T/R} \mathbb{1}_{\sigma_n(s) \in \Delta^{h+1}} ds \right). \end{aligned}$$

Also, remark that if  $Z_b \neq z$  then for all  $s \geq \tau_b$

$$\mathbb{1}_{\tilde{\sigma}_n^{Z_b}(s)=j} = 0 = \pi_n^z(j) \mathbb{1}_{\tilde{\sigma}_n^{Z_b}(s) \in \mathcal{E}_z^{h+1}}.$$

Up to now we have obtained that, for all  $i \in \llbracket 1, N \rrbracket$ , for  $n$  large enough,

$$\begin{aligned} & \mathbb{E}_i \left( \left| \int_0^T \left( \mathbb{1}_{\sigma_n(s)=j} - \pi_n^z(j) \mathbb{1}_{\sigma_n(s) \in \mathcal{E}_z^{h+1}} \right) ds \right| \right) \\ & \leq \frac{KT}{R} + 3R \sup_{u \in \llbracket 1, N \rrbracket} \mathbb{E}_u \left( \int_0^{T/R} \mathbb{1}_{\sigma_n(s) \in \Delta^{h+1}} ds \right) \\ & \quad + R \sup_{u \in \bar{\mathcal{E}}_z^{h+1}} \mathbb{E}_u \left( \left| \int_0^{T/R} \left( \mathbb{1}_{\tilde{\sigma}_n^z(s)=j} - \pi_n^z(j) \mathbb{1}_{\tilde{\sigma}_n^z(s) \in \mathcal{E}_z^{h+1}} \right) ds \right| \right). \end{aligned}$$

Since the left-hand side does not depend on  $R$ , it only remains to prove that, for any fixed  $R$ , the last term converges to zero as  $n \rightarrow +\infty$ . Using Proposition 5.6, the proof is exactly the same as that given for Proposition 5.1.  $\square$

We are now ready to prove the main result of this section, Theorem 3.3.

*Proof of Theorem 3.3.* Let us first show that it is sufficient to consider the case where  $P_n$  is strongly connected for all  $n \in \mathbb{N}$ . Indeed, instead of a chain  $\sigma_n$  associated with some  $(\mathbf{v}_n, \boldsymbol{\mu}_n, P_n)_{n \in \mathbb{N}}$ , we can rather consider a chain  $\tilde{\sigma}_n$  associated with  $(\mathbf{v}_n, \boldsymbol{\mu}_n + e^{-n}, (\boldsymbol{\mu}_n P_n + e^{-n} Q) / (\boldsymbol{\mu}_n + e^{-n}))_{n \in \mathbb{N}}$  where  $Q$  is the matrix whose coefficients are all  $1/N$ . In other words,  $\tilde{\sigma}_n$  behaves like  $\sigma_n$  except that, at rate  $e^{-n}$ , it jumps to a position distributed uniformly over  $\llbracket 1, N \rrbracket$ . In particular, as in the proof of Proposition 5.2, considering the synchronous coupling of  $\sigma_n$  and  $\tilde{\sigma}_n$ , we get that, for all  $T > 0$ ,

$$\mathbb{P}(\sigma_n(t) = \tilde{\sigma}_n(t) \forall t \in [0, T]) \geq e^{-Te^{-n}}.$$

Hence, if we prove that, up to a subsequence, the Markov process associated with  $\tilde{\sigma}_n$  converges in the Skorokod topology to some convexified process, then the same convergence holds for the Markov process associated with  $\sigma_n$ . For this reason, from now on we suppose that  $P_n$  is strongly connected.

The case where  $(\boldsymbol{\mu}_n)_{n \in \mathbb{N}}$  is bounded has already been treated in Proposition 5.2, so we suppose that  $\lambda_n(i, j) \rightarrow +\infty$  for some distinct  $i, j \in \llbracket 1, N \rrbracket$  and we keep all the definitions and notations of the rest of the section. Up to extracting a subsequence, we assume that, for all  $z \in \llbracket 1, n_{h+1} \rrbracket$ ,  $\pi_n^z$  converges as  $n \rightarrow +\infty$  to some law  $\pi^z$ . Let us prove that for all  $T, \delta > 0, z \in \llbracket 1, n_{h+1} \rrbracket$ , and  $j \in \mathcal{E}_z^{h+1}$ ,

$$\sup_{i \in \llbracket 1, N \rrbracket} \mathbb{P}_i \left( \left| \int_0^T \left( \mathbb{1}_{\sigma_n(s)=j} - \pi^z(j) \mathbb{1}_{\tilde{X}^{h+1}(s)=z} \right) ds \right| > \delta \right) \xrightarrow{n \rightarrow +\infty} 0, \quad (5.13)$$

from which the conclusion follows exactly as in the proof of [5, Lemma 2.14] or of Proposition 5.1. First, by (5.5),

$$\sup_{i \in \llbracket 1, N \rrbracket} \mathbb{P}_i \left( \left| \pi^z(j) \int_0^T \left( \mathbb{1}_{\Psi^{h+1}(\sigma_n(s))=z} - \mathbb{1}_{\tilde{X}^{h+1}(s)=z} \right) ds \right| > \delta/2 \right) \xrightarrow{n \rightarrow +\infty} 0.$$

Second,

$$\sup_{i \in \llbracket 1, N \rrbracket} \mathbb{E}_i \left( \left| \pi^z(j) - \pi_n^z(j) \right| \int_0^T \mathbb{1}_{\Psi^{h+1}(\sigma_n(s))=z} ds \right) \leq T |\pi_n^z - \pi^z|_\infty \xrightarrow{n \rightarrow +\infty} 0.$$

Third, Proposition 5.7 reads

$$\sup_{i \in \llbracket 1, N \rrbracket} \mathbb{E}_i \left( \left| \int_0^T \left( \mathbb{1}_{\sigma_n(s)=j} - \pi_n^z(j) \mathbb{1}_{\Psi^{h+1}(\sigma_n(s))=z} \right) ds \right| \right) \xrightarrow{n \rightarrow +\infty} 0.$$

The combination of these three facts yields (5.13).  $\square$

We now extend Theorem 3.3 to the case where the processes  $(x_n, \sigma_n)$  are themselves already convexified. The parameters of a convexified process are  $(\mathbf{v}, \boldsymbol{\mu}, P, B)$  where  $B = (B_1, \dots, B_k)$  are the modes and  $(\mathbf{v}, \boldsymbol{\mu}, P)$  are the parameters of the Markov chain on  $\llbracket 1, k \rrbracket$ .

**Corollary 5.8.** *Consider a sequence  $(x_n, \sigma_n)$  of convexified Markov processes for  $A$  with parameters  $(\mathbf{v}_n, \boldsymbol{\mu}_n, P_n, B_n)_{n \in \mathbb{N}}$ . Up to extracting a subsequence, there exists a convexified Markov process  $(x, \sigma)$  for  $A$  such that, for all  $T, \delta > 0$ ,*

$$\mathbb{P} \left( \sup_{t \in [0, T]} |x(t) - x_n(t)| > \delta \right) \xrightarrow{n \rightarrow +\infty} 0.$$

*Proof.* For each  $n \in \mathbb{N}$ , the set of matrices in  $B_n$  is related to a partition of the matrices in  $A$ , and thus, up to extracting a subsequence, we assume that this partition is the same for all  $n \in \mathbb{N}$ . In other words,  $B_n = (B_{n,1}, \dots, B_{n,N'})$  where  $N' \in \llbracket 1, N \rrbracket$  is independent of  $n$ , and for all  $j \in \llbracket 1, N' \rrbracket$ ,  $B_{n,j} = \sum_{i \in I_j} \pi_{n,i} A_i$ , where the sets  $I_1, \dots, I_{N'}$  are disjoint, nonempty, and independent of  $n$ , and  $(\pi_{n,i})_{i \in I_j}$  is a probability vector. In particular, for all  $n \in \mathbb{N}$ ,  $v_n$  and  $P_n$  are respectively of dimension  $N'$  and  $N' \times N'$ . By compactness, up to extracting a subsequence, we assume that the coefficients  $\pi_{n,i}$  converge as  $n \rightarrow \infty$ , in other words  $B_n \rightarrow B = (B_1, \dots, B_{N'})$ .

For  $n \in \mathbb{N}$ , let  $y_n$  be the solution of  $\dot{y}_n(t) = B_{\sigma_n(t)} y_n(t)$  for  $t \geq 0$ . In particular,  $(y_n, \sigma_n)$  is a Markov process for  $B$  with parameters  $(v_n, \mu_n, P_n, B)$ . Applying Theorem 3.3, we get that there exists a convexified Markov process  $(x, \sigma)$  for  $B$  (which is in particular a convexified process for  $A$ ) such that, for all  $T, \delta > 0$ ,

$$\mathbb{P} \left( \sup_{t \in [0, T]} |x(t) - y_n(t)| > \delta \right) \xrightarrow{n \rightarrow +\infty} 0.$$

On the other hand, since  $y_n$  and  $x_n$  share the same Markov chain  $\sigma_n$ , we immediately get that, almost surely, for all  $t \geq 0$ ,

$$\begin{aligned} |x_n(t) - y_n(t)| &\leq \int_0^t (\|B_{\sigma_n(s)}\| |x_n(s) - y_n(s)| + \|B_{n, \sigma_n(s)} - B_{\sigma_n(s)}\| |x_n(s)|) ds \\ &\leq K \int_0^t |x_n(s) - y_n(s)| ds + t e^{Kt} |x_0| \max_{j \in \llbracket 1, N' \rrbracket} \|B_j - B_{n,j}\| \end{aligned}$$

where  $K = \max_{i \in \llbracket 1, N \rrbracket} \|A_i\|$  and thus, almost surely,

$$\sup_{t \in [0, T]} |x_n(t) - y_n(t)| \xrightarrow{n \rightarrow \infty} 0,$$

yielding the conclusion.  $\square$

Up to now, in this section, the initial condition  $x_0 \in \mathbb{R}^d$  has been fixed and is common to all processes  $x$  and  $x_n$ . Since we are interested in the study of Lyapunov exponents, it is useful to reformulate Corollary 5.8 in terms of the flow  $\Phi_\sigma$ . This is done by applying it to the Markov process  $(\Phi_{\sigma_n}, \sigma_n)$ , which is a PDMP corresponding to the flow in the space  $\mathcal{M}_d(\mathbb{R})$  of  $\dot{\Phi}_{\sigma_n}(t) = A_{\sigma_n(t)} \Phi_{\sigma_n}(t)$  with initial condition  $\Phi_{\sigma_n}(0) = \text{Id}$ . As a straightforward consequence, using the fact that the flow is bounded on compact time intervals uniformly with respect to  $\sigma$ , we obtain the following.

**Corollary 5.9.** *Let  $(v_n, \mu_n, P_n, B_n)_{n \in \mathbb{N}}$  be a sequence of convexified Markov processes for  $A$ . Then there exists a convexified Markov process  $(v_*, \mu_*, P_*, B)$  for  $A$  such that, up to extracting a subsequence, we have, for every  $T > 0$ ,*

$$\mathbb{E}[\log \|\Phi_{\sigma_n}(T)\|] \xrightarrow{n \rightarrow +\infty} \mathbb{E}[\log \|\Phi_\sigma(T)\|], \quad (5.14)$$

where  $\sigma_n$  and  $\sigma$  are obtained from the convexified Markov processes  $(v_n, \mu_n, P_n, B_n)$  and  $(v_*, \mu_*, P_*, B)$ , respectively.

## 6 On the equality between $\lambda_d(A)$ and $\lambda_p^{\text{sup}}(A)$

One of the difficulties in analysis of  $\lambda_p^{\text{sup}}(A)$  is that it may fail to be reached by some  $\lambda_p(\nu, \mu, P, A)$ , as detailed next.

**Example 6.1.** Consider the case  $N = d = 2$  with

$$A_1 = \begin{pmatrix} 0 & -1 \\ 1 & -1 \end{pmatrix}, \quad A_2 = \begin{pmatrix} 0 & 1 \\ -1 & -1 \end{pmatrix}, \quad M = \frac{1}{2}A_1 + \frac{1}{2}A_2 = \begin{pmatrix} 0 & 0 \\ 0 & -1 \end{pmatrix}.$$

Then the spectral abscissa of  $M$  is equal to 0. From Proposition 6.4 stated below (since it is easy to check that  $M$  satisfies condition (C) in Definition 6.3 below), for  $\nu = (\frac{1}{2}, \frac{1}{2})$  and  $P = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}$ , we have  $\lim_{\mu \rightarrow +\infty} \lambda_p(\nu, \mu, P, A) = 0$ . Hence, on the one hand,  $\lambda_p^{\text{sup}}(A) \geq 0$ .

On the other hand, it is easily seen that the standard Euclidean norm is strictly decreasing along both flows  $\dot{x} = A_i x$ ,  $i = 1, 2$ , so that  $\log \|\Phi_\sigma(t)\| < 0$  for any piecewise constant  $\sigma$ . From Proposition 2.3 and (2.9),  $\lambda_p(\nu, \mu, P, A) < 0$  for every choice of the Markov process  $(\nu, \mu, P)$ .

Thus,  $\lambda_p^{\text{sup}}(A)$  is not reached.

The equality cases between  $\lambda_d(A)$  and  $\lambda_p^{\text{sup}}(A)$  will be addressed by studying the equality cases in (3.2). A helpful property is that, contrarily to what happens for  $\lambda_p^{\text{sup}}(A)$ , the supremum in the definition of  $\lambda_p^{\text{conv}}(A)$  is always reached, as shown next.

**Proposition 6.2.** *There exists a convexified Markov process  $(\nu, \mu, P, B)$  for  $A$  such that  $\lambda_p^{\text{conv}}(A) = \lambda_p(\nu, \mu, P, B)$ .*

*Proof.* Let  $(\nu_n, \mu_n, P_n, B_n)_{n \in \mathbb{N}}$  be a maximizing sequence for  $\lambda_p^{\text{conv}}(A)$ , i.e.,  $\lambda_p(\nu_n, \mu_n, P_n, B_n) \rightarrow \lambda_p^{\text{conv}}(A)$  as  $n \rightarrow \infty$ . By Proposition 2.3, with no loss of generality, we may assume that  $\nu_n$  is invariant under  $P_n$  for every  $n \in \mathbb{N}$ .

By Corollary 5.9, there exists a convexified Markov process  $(\nu, \mu, P, B)$  for  $A$  such that, for every  $T > 0$ , (5.14) holds. Then, using (2.7) and (2.9),

$$\begin{aligned} \lambda_p(\nu, \mu, P, B) &= \limsup_{t \rightarrow +\infty} \frac{1}{t} \mathbb{E}[\log \|\Phi_\sigma(t)\|] = \limsup_{t \rightarrow +\infty} \lim_{n \rightarrow +\infty} \frac{1}{t} \mathbb{E}[\log \|\Phi_{\sigma_n}(t)\|] \\ &\geq \lim_{n \rightarrow +\infty} \lambda_p(\nu_n, \mu_n, P_n, A) = \lambda_p^{\text{conv}}(A), \end{aligned}$$

hence the conclusion.  $\square$

We are now ready to prove Theorem 3.4.

*Proof of Theorem 3.4.* Let us prove (a). Assume first that  $\lambda_d(A) = \lambda_p^{\text{conv}}(A)$  and consider a convexified Markov process  $(\nu, \mu, P, B)$  at which  $\lambda_p^{\text{conv}}(A)$  is attained, whose existence is established in Proposition 6.2. Hence

$$\lambda_d(B) \leq \lambda_d(A) = \lambda_p^{\text{conv}}(A) = \lambda_p(\nu, \mu, P, B) \leq \lambda_d(B),$$

implying that  $\lambda_p(\nu, \mu, P, B) = \lambda_d(B)$ . Let  $i$  be an index belonging to a recurrent class of  $P$  and accessible from  $\nu$ . Then, by Theorem 3.1, we deduce that  $\rho(e^{Bt}) = e^{\lambda_d(B)t}$  for



every  $t \geq 0$ , which yields  $\lambda(B_i) = \lambda_d(B)$ , and the conclusion follows since  $B_i \in \text{co}(A)$  and  $\lambda_d(B) = \lambda_d(A)$ .

To prove the converse implication, let  $M \in \text{co}(A)$  be such that  $\lambda(M) = \lambda_d(A)$ . Then the process  $\dot{x} = Mx$  is a convexified Markov process for  $A$ , with parameters  $(\nu, \mu, P, B)$  given by  $B = (M)$  and  $(\nu, \mu, P)$  determining the (constant) Markov chain on a single state. Hence

$$\lambda_d(A) = \lambda(M) = \lambda_p(\nu, \mu, P, B) \leq \lambda_p^{\text{conv}}(A) \leq \lambda_d(A),$$

yielding the conclusion.

The proof of (b) follows immediately from (a) and (3.2).

To prove (c), notice that, by the continuity of the spectral abscissa function, for every  $\eta > 0$ , there exists  $\varepsilon > 0$  such that  $\lambda(M^\varepsilon) > \lambda(M) - \eta = \lambda_d(A) - \eta$ . Let  $(\nu_n, \mu_n, P_n)_{n \in \mathbb{N}}$  be a sequence of Markov processes for  $A$  such that (3.3) holds true. Then, for  $n$  large enough

$$\lambda_p(\nu_n, \mu_n, P_n, A) > \lambda_d(A) - \eta.$$

Hence  $\lambda_d(A) < \lambda_p^{\text{sup}}(A) + \eta$  and, since  $\eta > 0$  is arbitrary, we deduce that  $\lambda_d(A) \leq \lambda_p^{\text{sup}}(A)$ , yielding the conclusion thanks to (2.8).  $\square$

Theorem 3.4 (c) raises the question of verifying the condition stated in (3.3). If  $M = A_i$  for some  $i \in \llbracket 1, N \rrbracket$ , (3.3) trivially holds by taking  $M^\varepsilon = A_i$  and the Markov chain constantly equal to  $i$ . Otherwise, one can expect  $M$  to be attained through arbitrarily fast switching. This leads to the question of the convergence of the Lyapunov exponent in this regime, which is the subject of the sequel of this section.

Let us introduce an explicit algebraic condition under which we are able to prove the convergence of Lyapunov exponents stated in (3.3).

**Definition 6.3** (Condition (C)). Let  $M$  be in  $\mathcal{M}_d(\mathbb{R})$  and denote by  $\xi_1 > \xi_2 > \dots > \xi_k$  the distinct real parts of the eigenvalues of  $M$  (with  $k \leq d$ ). For  $j \in \llbracket 1, k \rrbracket$ , denote by  $E_j$  the space spanned by the generalized eigenvectors of  $M$  corresponding to eigenvalues with real part  $\xi_j$  and by  $n_j$  the dimension of  $E_j$ . Let  $F_M$  be the vector field on the  $(d-1)$ -dimensional unit sphere  $S^{d-1}$  obtained by projecting  $x \mapsto Mx$ , i.e.,  $F_M(x) = Mx - (x \cdot Mx)x$ .

Given  $A = (A_1, \dots, A_N) \in \mathcal{M}_d(\mathbb{R})^N$ , we say that  $M$  satisfies *condition (C) for A* if, for every  $j \in \llbracket 2, k \rrbracket$ , there exists  $i \in \llbracket 1, N \rrbracket$  such that, for every  $\theta \in E_j \cap S^{d-1}$ , one has  $F_{A_i}(\theta) \notin \bigoplus_{r \geq j} E_r$ .

Condition (C) can be used to obtain the following result on the convergence of Lyapunov exponents.

**Proposition 6.4.** Let  $A = (A_1, \dots, A_N) \in \mathcal{M}_d(\mathbb{R})^N$ ,  $\pi = (\pi_1, \dots, \pi_N) \in (0, 1]^N$  be a probability vector, and define  $M = \sum_{j=1}^N \pi_j A_j$ . Let  $P \in \mathcal{M}_N(\mathbb{R})$  be the matrix with all rows equal to  $\pi$  and  $(\mu_n)_{n \in \mathbb{N}}$  be a sequence tending to  $+\infty$ . If  $M$  satisfies condition (C) for  $A$ , then the sequence  $(\pi, \mu_n, P)_{n \in \mathbb{N}}$  of Markov processes for  $A$  satisfies

$$\lim_{n \rightarrow \infty} \lambda_p(\pi, \mu_n, P, A) = \lambda(M).$$

From the results in [4, Section 2.5], the probabilistic Lyapunov exponent can be expressed in terms of the invariant measures of the Markov process  $(\theta(t), \sigma(t))_{t \geq 0} = (x(t)/\|x(t)\|, \sigma(t))_{t \geq 0}$  on  $S^{d-1} \times \llbracket 1, N \rrbracket$ . As a consequence, Proposition 6.4 follows from a simple combination of the results of [4] and of the following proposition, whose proof is given in Appendix B.

**Proposition 6.5.** *Let  $A = (A_1, \dots, A_N) \in \mathcal{M}_d(\mathbb{R})^N$ ,  $\pi = (\pi_1, \dots, \pi_N) \in (0, 1]^N$  be a probability vector, and define  $M = \sum_{j=1}^N \pi_j A_j$ . Let  $P \in \mathcal{M}_N(\mathbb{R})$  be the matrix with all rows equal to  $\pi$ ,  $\mu > 0$ , and consider the Markov process  $(\theta(t), \sigma(t))_{t \geq 0}$  on  $S^{d-1} \times \llbracket 1, N \rrbracket$  obtained from  $(\pi, \mu, P)$ .*

*If  $M$  satisfies condition (C) for  $A$ , then, for every  $\varepsilon \in (0, 1]$  and every neighborhood  $K$  of  $E_1$  in  $S^{d-1}$ , there exists  $\mu_0 > 0$  such that, for every  $\mu \geq \mu_0$  and every invariant measure  $\rho$  of the process  $(\theta(t), \sigma(t))_{t \geq 0}$ , one has  $\rho(K \times \llbracket 1, N \rrbracket) \geq 1 - \varepsilon$ .*

Notice that a similar result has recently been established in [10]. More precisely, our result is similar in spirit to [10, Theorem 1.1] in the case where the diffusion coefficient  $\delta$  is equal to 0 and the limit cycle considered in [10] is reduced to a single point, and our condition (C) corresponds to the second point of [10, Assumption 1.2]. The proof of [10] is based on estimates on the exit times of the trajectory from neighborhoods of unstable equilibria, while our argument is based on the construction of a suitable Lyapunov function (similarly to [4], where the result is established in a particular case in dimension 2).

By combining Proposition 6.4 with Theorem 3.4, one deduces the following statement.

**Corollary 6.6.** *Let  $A \in \mathcal{M}_d(\mathbb{R})^N$ . Assume that condition (C) holds for a dense subset of  $\text{co}(A)$ . Then  $\lambda_d(A) = \lambda_p^{\text{sup}}(A)$  if and only if there exists  $M \in \text{co}(A)$  such that  $\lambda(M) = \lambda_d(A)$ .*

From what precedes one would prove a full converse of Theorem 3.4 (b) if for every  $A \in \mathcal{M}_d(\mathbb{R})^N$  condition (C) held for a dense subset of  $\text{co}(A)$ . Our next result, whose proof is provided in Appendix B, states that this is the case for dimensions up to 3.

**Lemma 6.7.** *Let  $d \leq 3$  and  $A \in \mathcal{M}_d(\mathbb{R})^N$  be irreducible. Then condition (C) holds for a dense subset of  $\text{co}(A)$ .*

Thanks to Lemma 6.7, we are now in position to prove Proposition 3.5.

*Proof of Proposition 3.5.* By Theorem 3.4 (b), it remains to prove that  $\lambda_d(A) = \lambda_p^{\text{sup}}(A)$  if there exists  $M \in \text{co}(A)$  such that  $\lambda(M) = \lambda_d(A)$ . Hence, we assume that such a  $M$  exists. We next use the block decomposition of matrices of  $A$  used in the proof of Proposition 4.3 (and, in particular, the notations of (4.3), extended to the matrices in  $\text{co}(A)$ ) to deduce that there exists  $j \in \llbracket 1, S \rrbracket$  such that  $\lambda(M^{(j)}) = \lambda(M)$ . Note that  $\lambda_d(A^{(j)}) \leq \lambda_d(A) = \lambda(M) = \lambda(M^{(j)}) \leq \lambda_d(A^{(j)})$ , implying that all these inequalities are in fact equalities. Thus, by applying Lemma 6.7 and Corollary 6.6 to  $M^{(j)}$  and  $A^{(j)}$ , we deduce that  $\lambda_d(A^{(j)}) = \lambda_p^{\text{sup}}(A^{(j)})$ . Hence,

$$\lambda_d(A) = \lambda_d(A^{(j)}) = \lambda_p^{\text{sup}}(A^{(j)}) \leq \lambda_p^{\text{sup}}(A) \leq \lambda_d(A),$$

yielding  $\lambda_d(A) = \lambda_p^{\text{sup}}(A)$ . □

## Acknowledgements

The authors thank Edouard Strickler for fruitful discussions. G. Mazanti was partially supported by ANR PIA funding number ANR-20-IDEES-0002. P. Monmarché acknowledges financial support by the French ANR grant SWIDIMS (ANR-20-CE40-0022).

## A Decomposition of a Markov chain

This section is devoted to the proof of Theorem 5.4. It is based on the works of Landim and collaborators [2, 17–19], to which we refer for more details and discussions. Even though the contents of Theorem 5.4 are similar to those of Theorems 2.1, 2.7, and 2.12 of [19], we decided to state them in a modified way, more convenient in our context. Let us briefly highlight the main differences and explain how they are handled in the sequel.

- In [19], the convergence after rescaling in time is given in term of the so-called soft topology, introduced in [17], while in Theorem 5.4 (b) and (c) we state a convergence for the time marginals. The way to get this time marginals convergence instead of the soft one is established in [18, Proposition 2.1].
- The case of a timescale which is strictly between  $\theta^j$  and  $\theta^{j+1}$  is not considered in [19]. In particular, in [19, Condition H1], the total rate of the limit Markov chain is supposed to be nonzero. Nevertheless, this assumption is not used to prove the convergence of the rescaled chain, it is just used, after the convergence is established, to say that the limit chain is not constant, so that in particular the next partition will be strictly coarser than the previous one. Replacing  $\theta^{j+1}$  by some  $\alpha$  with  $\theta^j \ll \alpha \ll \theta^{j+1}$ , the conditions [19, H1–H3] are still satisfied, except that in H1 the total rate is zero.
- The fact that (5.3) and (5.4) still hold with  $\tilde{\theta}_n^j \simeq \theta_n^j$  is immediately obtained from the fact  $\tilde{\theta}^j$  satisfies the same conditions [19, H1–H3] as  $\theta^j$ .

Throughout this section, we consider for each  $n \in \mathbb{N}$  a strongly connected Markov chain  $(\sigma_n(t))_{t \geq 0}$  on  $\llbracket 1, N \rrbracket$  for some fixed  $N \in \mathbb{N}$  with jump rates  $(\lambda_n(i, j))_{i, j \in \llbracket 1, N \rrbracket}$ . We suppose that the conditions (5.1) and (5.2) are fulfilled. We denote by  $\pi_n$  the unique invariant probability measure of  $\sigma_n$ .

For a nonempty  $J \subset \llbracket 1, N \rrbracket$ , the *trace process on  $J$  associated with  $\sigma_n$*  is the process  $(\sigma_n^J(t))_{t \geq 0}$  obtained from  $\sigma_n$  by removing all the time spent outside of  $J$ . More precisely, denoting

$$T_n^J(t) = \int_0^t \mathbb{1}_{\sigma_n(s) \in J} ds \quad \text{and} \quad S_n^J(t) = \sup\{s \geq 0 \mid T_n^J(s) \leq t\},$$

we define  $\sigma_n^J(t) = \sigma_n(S_n^J(t))$ . Remark that, the chain being strongly connected,  $T_n^J(t) \rightarrow +\infty$  almost surely as  $t \rightarrow +\infty$  and  $S_n^J(t)$  is finite for all  $t \geq 0$ .

For  $J \subset \llbracket 1, N \rrbracket$ , denote by  $H_J$  and  $H_J^+$  the hitting time of  $J$  and the time of the first return to  $J$ , i.e.,

$$H_J = \inf\{t \geq 0 \mid \sigma_n(t) \in J\}, \quad H_J^+ = \inf\{t \geq \tau_1 \mid \sigma_n(t) \in J\},$$

where  $\tau_1 = \inf\{t \geq 0 \mid \sigma_n(t) \neq \sigma_n(0)\}$  is the first jump time of the chain. For  $i, j \in \llbracket 1, N \rrbracket$ , denote  $\lambda_n(i) = \sum_{k \neq i} \lambda_n(i, k)$  the holding rate of the chain at  $i$ , and  $p_n(i, j) = \lambda_n(i, j) / \lambda_n(i)$  the transition probabilities from  $i$ . For two disjoint subsets  $J_1$  and  $J_2$  of  $\llbracket 1, N \rrbracket$ , the capacity between  $J_1$  and  $J_2$  is defined by

$$\text{cap}_n(J_1, J_2) = \sum_{i \in J_1} \pi_n(i) \lambda_n(i) \mathbb{P}_i(H_{J_2} < H_{J_1}^+).$$

Let  $\mathcal{E}_1, \dots, \mathcal{E}_n, \Delta$  be a partition of  $\llbracket 1, N \rrbracket$  for some  $n \geq 1$  with  $\mathcal{E}_x \neq \emptyset$  for all  $x \in \llbracket 1, n \rrbracket$  and let  $\mathcal{E} = \bigcup_{x=1}^n \mathcal{E}_x$ . It is proven in [2, Proposition 6.1] that the trace process  $\sigma_n^\mathcal{E}$  is a Markov chain on  $\mathcal{E}$  with rates

$$R_n(i, j) = \lambda_n(i) \mathbb{P}_i(H_{\{j\}} = H_\mathcal{E}^+)$$

for  $i \neq j$ . For  $x, y \in \llbracket 1, n \rrbracket$ ,  $x \neq y$ , denote by

$$r_n^\mathcal{E}(x, y) = \frac{1}{\pi_n(\mathcal{E}_x)} \sum_{i \in \mathcal{E}_x} \pi_n(i) \sum_{j \in \mathcal{E}_y} R_n(i, j)$$

the mean (at equilibrium) rate at which the trace process jumps from  $\mathcal{E}_x$  to  $\mathcal{E}_y$ . Consider the coarse-grained variable

$$\Psi^\mathcal{E}(i) = \sum_{x=1}^n x \mathbb{1}_{i \in \mathcal{E}_x}.$$

Finally, let  $(\alpha_n)_{n \in \mathbb{N}}$ ,  $(\beta_n)_{n \in \mathbb{N}}$  be two positive sequences. We consider the following conditions:

**H1.** For all  $x, y \in \llbracket 1, n \rrbracket$  with  $x \neq y$ , there exists  $r^\mathcal{E}(x, y) \geq 0$  so that

$$\beta_n r_n^\mathcal{E}(x, y) \xrightarrow{n \rightarrow +\infty} r^\mathcal{E}(x, y).$$

**H2.** For all  $x \in \llbracket 1, n \rrbracket$  such that  $\mathcal{E}_x$  is not a singleton and all  $i, j \in \mathcal{E}_x$  with  $i \neq j$ ,

$$\liminf_{n \rightarrow +\infty} \alpha_n \frac{\text{cap}(\{i\}, \{j\})}{\pi_n(\mathcal{E}_x)} > 0.$$

Intuitively, **H2** means that mixing within a class  $\mathcal{E}_x$  occurs in a time smaller than  $\alpha_n$ , and **H1** that transitions between different classes  $\mathcal{E}_x$  and  $\mathcal{E}_y$  occur at a time of order at least  $\beta_n$ .

**Proposition A.1.** *Consider a partition  $\mathcal{E}_1, \dots, \mathcal{E}_n, \Delta$  of  $\llbracket 1, N \rrbracket$  and two positive sequences  $(\alpha_n)_{n \in \mathbb{N}}$ ,  $(\beta_n)_{n \in \mathbb{N}}$  such that  $\alpha_n \ll \beta_n$  and **H1–H2** hold. Then, for  $x \in \llbracket 1, n \rrbracket$  and for an initial condition  $i \in \mathcal{E}_x$ ,  $(\Psi^\mathcal{E}(\sigma_n^\mathcal{E}(t\beta_n)))_{t \geq 0}$  converges in law in the Skorokhod topology toward the Markov chain  $(X(t))_{t \geq 0}$  on  $\llbracket 1, n \rrbracket$  with initial condition  $x$  and jump rates  $(r^\mathcal{E}(y, z))_{y, z \in \llbracket 1, n \rrbracket}$  as  $n \rightarrow +\infty$ .*

*Proof.* This is [19, Proposition 6.1]. Remark that in [19] it is also assumed that the limit chain is not constant, i.e.,  $\sum_{x, y \in \llbracket 1, n \rrbracket} r^\mathcal{E}(x, y) \neq 0$ . Nevertheless this is not used in the proof of [19, Proposition 6.1], which relies on results from [2] in which this additional assumption is not made.  $\square$

Remark that in the particular case  $n = 1$  the result is trivial as condition **H1** is empty and  $\Psi^\mathcal{E}(\sigma_n^\mathcal{E}(t)) = 1$  for all  $n \in \mathbb{N}$  and  $t \geq 0$ .

Now, consider the following additional conditions on the partition  $\mathcal{E}_1, \dots, \mathcal{E}_n, \Delta$  and the timescale  $\beta$ :

**H3.** For all  $i \in \llbracket 1, N \rrbracket$  and  $t > 0$ ,

$$\mathbb{E}_i \left( \int_0^t \mathbb{1}_{\sigma_n(s\beta_n) \in \Delta} ds \right) \xrightarrow{n \rightarrow +\infty} 0.$$

**H4.** For all  $x \in \llbracket 1, n \rrbracket$  and  $i \in \mathcal{E}_x$ ,

$$\lim_{\delta \rightarrow 0} \limsup_{n \rightarrow +\infty} \sup_{s \in [2\delta, 3\delta]} \mathbb{P}_i(\sigma_n(s\beta_n) \in \Delta) = 0.$$

**Proposition A.2.** Consider a partition  $\mathcal{E}_1, \dots, \mathcal{E}_n, \Delta$  of  $\llbracket 1, N \rrbracket$  and two positive sequences  $(\alpha_n)_{n \in \mathbb{N}}, (\beta_n)_{n \in \mathbb{N}}$  such that  $\alpha_n \ll \beta_n$  and **H1–H3** hold. Fix  $x \in \llbracket 1, n \rrbracket$  and an initial condition  $i \in \mathcal{E}_x$ . Let  $X$  be as in Proposition A.1. Then

(a) For all  $t \geq 0$  and  $y \in \llbracket 1, n \rrbracket$ ,

$$\int_0^t \mathbb{1}_{\Psi^\varepsilon(\sigma_n(s\beta_n))=y} ds \xrightarrow{n \rightarrow +\infty} \int_0^t \mathbb{1}_{X(s)=y} ds \quad \text{in law.}$$

(b) Let  $\tau_n = \inf\{t \geq 0 \mid \sigma_n(t) \notin \mathcal{E}_x \cup \Delta\}$ . Then  $\tau_n/\beta_n$  converges in law as  $n \rightarrow +\infty$  to an exponential variable with parameter  $r^\varepsilon(x) := \sum_{y \in \llbracket 1, n \rrbracket \setminus \{x\}} r^\varepsilon(x, y)$  (to be understood as  $\tau_n/\beta_n \rightarrow +\infty$  in probability in the case where  $r^\varepsilon(x) = 0$ ).

(c) If, moreover, **H4** holds, then for all  $t_1, \dots, t_k \geq 0$ ,  $(\Psi^\varepsilon(\sigma_n(t_j\beta_n)))_{j \in \llbracket 1, k \rrbracket}$  converges in law toward  $(X(t_j))_{j \in \llbracket 1, k \rrbracket}$  as  $n \rightarrow +\infty$ .

*Proof.* For the first point, we know by Proposition A.1 that, for all  $t \geq 0$ ,  $\int_0^t \mathbb{1}_{\Psi^\varepsilon(\sigma_n^\varepsilon(s\beta_n))=y} ds$  converges in law toward  $\int_0^t \mathbb{1}_{X(s)=y} ds$  as  $n \rightarrow +\infty$ . Moreover,

$$\mathbb{E} \left( \int_0^t \left| \mathbb{1}_{\Psi^\varepsilon(\sigma_n(s\beta_n))=y} - \mathbb{1}_{\Psi^\varepsilon(\sigma_n^\varepsilon(s\beta_n))=y} \right| ds \right) \leq 2 \mathbb{E} \left( \int_0^t \mathbb{1}_{\sigma_n(s\beta_n) \in \Delta} ds \right),$$

which vanishes as  $n \rightarrow +\infty$  according to **H3**.

The proof of the second point is similar. Indeed, denoting  $\tau_n^\varepsilon = \inf\{t \geq 0 \mid \sigma_n^\varepsilon(t) \notin \mathcal{E}_x\}$ , the convergence in law of  $\Psi^\varepsilon(\sigma_n^\varepsilon(\cdot\beta_n))$  toward  $X$  in the Skorokhod topology implies the convergence in law of  $\tau_n^\varepsilon/\beta_n$  toward the first jump time of  $X$ . Moreover, necessarily,  $\tau_n^\varepsilon \leq \tau_n$  (since some time is removed in the trace process), and for all  $M, \varepsilon > 0$ ,

$$\begin{aligned} \mathbb{P}_x \left( \tau_n^\varepsilon / \beta_n \geq M \right) &\leq \mathbb{P}_x \left( \tau_n / \beta_n \geq M \right) \\ &\leq \mathbb{P}_x \left( \tau_n^\varepsilon / \beta_n \geq M - \varepsilon \right) + \mathbb{P}_x \left( \sup_{s \in [0, M\beta_n]} |s - T_n^\varepsilon(s)| \geq \varepsilon \beta_n \right) \\ &\leq \mathbb{P}_x \left( \tau_n^\varepsilon / \beta_n \geq M - \varepsilon \right) + \mathbb{P}_x \left( \int_0^{M\beta_n} \mathbb{1}_{\sigma_n(s) \in \Delta} ds \geq \varepsilon \beta_n \right). \end{aligned}$$

The last term vanishes as  $n \rightarrow +\infty$  thanks to **H3**. Noticing that  $\varepsilon$  is arbitrary, we conclude by using the convergence of  $\tau_n^\varepsilon/\beta_n$ .

Finally, the convergence of the time marginals of  $\Psi^\varepsilon(\sigma_n^\varepsilon(\cdot\theta_n))$  is established in [18, Proposition 2.1] under **H3, H4**, and the convergence in the Skorokhod topology for the trace process proven in Proposition A.1.  $\square$

*Proof of Theorem 5.4.* Let us recall the construction of [19] (with our slightly different notations). The first timescale  $\theta^1$  is defined by  $\theta_n^1 = \left( \sum_{i,j=1}^N \lambda_n(i, j) \right)^{-1}$ , and the first

partition is simply  $\mathcal{E}_x^1 = \{x\}$  for  $x \in \llbracket 1, N \rrbracket$ , hence  $n_1 = N$ , and  $\Delta^1 = \emptyset$ . As seen in [19, Section 3], condition (5.2) ensures that  $\lambda_n(i, j)\theta_n^1$  admits a nonnegative limit  $r_1(i, j)$  for all  $i, j \in \llbracket 1, N \rrbracket$ , with at least one pair  $(i, j)$  such that  $r_1(i, j) \neq 0$ . Notice that there is no transition at a timescale smaller than  $\theta^1$  since, for all  $t \geq 0$ , the probability that there has been a jump before time  $t$  is less than  $e^{-t/\theta^1}$ . Let  $X^1$  be the Markov chain in  $\llbracket 1, N \rrbracket$  with transition rates  $r_1(i, j)$ . Since the jump rates of  $\sigma_n(\cdot\theta_n^1)$  converge toward those of  $X^1$ , as in the proof of Proposition 5.2 we can consider a synchronous coupling for which

$$\mathbb{P}(X^1(t) = \sigma_n(t\theta_n^1) \forall t \in [0, T]) \xrightarrow[n \rightarrow +\infty]{} 1 \quad (\text{A.1})$$

for all  $T > 0$ . This implies the convergence in law in the Skorokhod space, hence point (b) for  $j = 1$ .

Denote by  $\mathcal{C}_1^1, \dots, \mathcal{C}_{n_2}^1$  the recurrence classes of  $X^1$  and, for all  $x \in \llbracket 1, n_2 \rrbracket$ , set  $\mathcal{E}_x^2 = \mathcal{C}_x^1$ . Denote by  $\mathcal{T}_1$  the set of transient points of  $X^1$  and set  $\Delta^2 = \mathcal{T}_1$ . Since  $X^1$  is not a constant chain, necessarily  $n_2 < n_1$ .

Then, the timescales and the partitions are defined by induction. Suppose that  $\theta^{j-1}$ ,  $n_j$ ,  $\mathcal{E}_1^j, \dots, \mathcal{E}_{n_j}^j$  (with  $\mathcal{E}_x^j \neq \emptyset$  for all  $x \in \llbracket 1, n_j \rrbracket$ ), and  $\Delta^j$  have been defined for some  $j \geq 2$ . If  $n_j = 1$ , we stop the construction (and set  $\mathfrak{p} = j - 1$ ). Otherwise, denoting  $\mathcal{E}_{\neq x} = \bigcup_{y \neq x} \mathcal{E}_y$  for  $x \in \llbracket 1, n_j \rrbracket$ , we set

$$\theta_n^j = \left( \sum_{x=1}^{n_j} \frac{\text{cap}(\mathcal{E}_x, \mathcal{E}_{\neq x})}{\mu_n(\mathcal{E}_x)} \right)^{-1},$$

which is well defined since the chain is strongly connected, so that the capacities are nonzero. By [19, Theorems 2.7 and 2.12], for all  $j \in \llbracket 2, \mathfrak{p} \rrbracket$ , the partition  $\mathcal{E}_1^j, \dots, \mathcal{E}_{n_j}^j, \Delta^j$  and the timescales  $\alpha_n = \theta_n^{j-1}$  and  $\beta_n = \theta_n^j$  satisfy **H1–H3** (with, in **H1**,  $r^\mathcal{E}(x, y) \neq 0$  for at least one pair  $(x, y)$ ) and  $\theta^j \ll \theta^{j+1}$ . Let  $X^j$  be the Markov chain on  $\llbracket 1, n_j \rrbracket$  given by Proposition A.1 for this partition and these timescales. Denote by  $\mathcal{C}_1^j, \dots, \mathcal{C}_{n_{j+1}}^j$  its recurrence classes and  $\mathcal{T}_j$  its transient points. For all  $x \in \llbracket 1, n_{j+1} \rrbracket$ , set  $\mathcal{E}_x^{j+1} = \bigcup_{y \in \mathcal{C}_x^j} \mathcal{E}_y^j$  and  $\Delta^{j+1} = \Delta^j \cup (\bigcup_{y \in \mathcal{T}_j} \mathcal{E}_y^j)$ . The chain  $X^j$  being nonconstant,  $n_{j+1} < n_j$ . This shows that this inductive definition of timescales and partitions ends in a finite number of steps.

Points (a) and (e) of Theorem 5.4 are satisfied by construction. Point (f) is a consequence of Proposition A.2.

Let us prove that **H4** holds for this choice of partitions and timescales. More precisely, let us prove by induction on  $j \in \llbracket 1, \mathfrak{p} \rrbracket$  that

$$\forall \delta > 0, \forall i \in \llbracket 1, N \rrbracket, \quad \limsup_{n \rightarrow +\infty} \sup_{s \geq \delta} \mathbb{P}_i(\sigma_n(s\theta_n^j) \in \Delta^j) = 0, \quad (\text{A.2})$$

which is stronger than **H4**. For  $j = 1$ ,  $\Delta^1 = \emptyset$ , so there is nothing to prove. For  $j = 2$ , let  $\varepsilon > 0$  and  $M > 0$  be such that

$$\sup_{t \geq M} \max_{i \in \llbracket 1, N \rrbracket} \mathbb{P}_i(X^1(t) \in \mathcal{T}_1) \leq \varepsilon.$$

Thanks to (A.1), we can consider  $n_0 \in \mathbb{N}$  such that for all  $n \geq n_0$

$$\max_{i \in \llbracket 1, N \rrbracket} \mathbb{P}_i(\sigma_n(M\theta_n^1) \in \Delta^2) \leq 2\varepsilon.$$

For  $\delta > 0$ , for all  $n \geq n_0$  large enough so that  $\delta \theta_n^2 > M \theta_n^1$ , by the Markov property, for all  $s \geq \delta$  and  $i \in \llbracket 1, N \rrbracket$ ,

$$\mathbb{P}_i(\sigma_n(s\theta_n^j) \in \Delta^2) = \sum_{k=1}^N \mathbb{P}_i(\sigma_n(s\theta_n^j - M\theta_n^1) = k) \mathbb{P}_k(\sigma_n(M\theta_n^1) \in \Delta^2) \leq 2\varepsilon.$$

Hence

$$\limsup_{n \rightarrow +\infty} \sup_{s \geq \delta} \mathbb{P}_i(\sigma_n(s\theta_n^2) \in \Delta^2) \leq 2\varepsilon$$

with an arbitrary  $\varepsilon > 0$ , which concludes the proof of (A.2) for  $j = 2$ .

Now, suppose by induction that (A.2) holds for all  $k \in \llbracket 1, j \rrbracket$  for some  $j \in \llbracket 1, p-1 \rrbracket$ . This implies **H4** with  $\beta_n = \theta_n^j$  and the partition  $\mathcal{E}_1^j, \dots, \mathcal{E}_{n_j}^j, \Delta^j$ . By Proposition A.2, we get the convergence of the time marginals of  $\Psi^j(\sigma_n(\cdot \theta_n^j))$  toward those of  $X^j$ . In particular, as in the previous case, for any  $\varepsilon > 0$  we can choose  $M, n_0 > 0$  large enough so that for all  $n \geq n_0$ ,

$$\max_{i \in \llbracket 1, N \rrbracket} \mathbb{P}_i(\sigma_n(M\theta_n^j) \in \Delta^{j+1} \setminus \Delta^j) \leq 2\varepsilon.$$

Since, by induction, we also know that  $\mathbb{P}_i(\sigma_n(M\theta_n^j) \in \Delta^j)$  vanishes as  $n \rightarrow +\infty$ , we get that

$$\max_{i \in \llbracket 1, N \rrbracket} \mathbb{P}_i(\sigma_n(M\theta_n^j) \in \Delta^{j+1}) \leq 3\varepsilon$$

for  $n$  large enough. The conclusion follows again from the Markov property, as in the case  $j = 2$ .

We have thus established that **H4** holds with  $\beta_n = \theta_n^j$  and the partition  $\mathcal{E}_1^j, \dots, \mathcal{E}_{n_j}^j$  for all  $j \in \llbracket 1, p \rrbracket$ . All remaining points of Theorem 5.4 then follow from Proposition A.2 and the fact that, when applying Proposition A.2, a sequence  $\alpha$  with  $\theta_n^{j-1} \ll \alpha_n \ll \theta_n^j$ , can replace  $\theta^{j-1}$  in condition **H2** or, alternatively, can replace  $\theta^j$  in the conditions **H1**, **H3**, and **H4** (in which case, in **H1**, the limit rates  $r^\varepsilon$  are all zero, so that the limit chain is constant). Indeed, the proof that **H3** holds in this case is given in [19, Lemma 7.2], which only requires that **H2** holds for  $\theta^{j-1}$  and that  $\alpha \gg \theta^{j-1}$ . Similarly, our proof of **H4** is unchanged if, at some step of the inductive construction of the timescales and the partitions, we replace  $\theta^j$  by some  $\alpha$  with  $\theta^{j-1} \ll \alpha \ll \theta^j$  (the only difference would be that the limit chain is constant, and thus we could not conclude that  $n_j < n_{j-1}$  to ensure that the construction ends in a finite number of steps).  $\square$

## B Technical results

### B.1 Proof of Proposition 6.5

The proof of Proposition 6.5 relies on the following linear algebra result.

**Lemma B.1.** *Let  $M$  be in  $\mathcal{M}_d(\mathbb{R})$  and denote by  $\xi_1 > \xi_2 > \dots > \xi_k$  the distinct real parts of the eigenvalues of  $M$ . For  $j \in \llbracket 1, k \rrbracket$ , denote by  $E_j$  the space spanned by the generalized eigenvectors of  $M$  corresponding to eigenvalues with real part  $\xi_j$ . Then, up to a linear change of coordinates, there exists a function  $h \in C^\infty(S^{d-1}, (0, +\infty))$  such that*

- (a)  $\nabla h(\theta) = 0$  for  $\theta \in S^{d-1} \cap (\bigcup_{i=1}^k E_i)$ ;

(b)  $h$  attains its minimum at  $\theta$  if and only if  $\theta \in E_1 \cap S^{d-1}$  and its maximum at  $\theta$  if and only if  $\theta \in E_k \cap S^{d-1}$ ;

(c)  $\nabla h(\theta) \cdot F_M(\theta) < 0$  for every  $\theta \in S^{d-1} \setminus \bigcup_{i=1}^k E_i$ , where we recall that  $F_M(\theta) = M\theta - (\theta \cdot M\theta)\theta$  for  $\theta \in S^{d-1}$ ;

(d)  $v \cdot \nabla^2 h(\theta)v \leq -|\text{pr}_{i-1}(v)|^2$  for  $i = 1, \dots, k$ ,  $\theta \in S^{d-1} \cap E_i$ , and  $v \in T_\theta S^{d-1}$ , where  $\nabla^2 h$  denotes the Hessian of  $h$  and, for  $i = 2, \dots, k$ ,  $\text{pr}_{i-1}$  denotes the projection onto  $E_1 \oplus \dots \oplus E_{i-1}$  along  $E_i \oplus \dots \oplus E_k$  and  $\text{pr}_0 = 0$ .

*Proof.* Denote by  $n_j$  the dimension of  $E_j$  for  $j \in \llbracket 1, k \rrbracket$ . Let us write a vector  $x \in \mathbb{R}^d$  as  $x = (x_1, \dots, x_k)$  with  $x_j \in \mathbb{R}^{n_j}$  identified up to a linear system of coordinates in  $E_j$  to be fixed later.

By construction,  $E_j = \{x \mid x_i = 0 \text{ for all } i \neq j\}$  and

$$M = \text{diag}(M_1, \dots, M_k),$$

where each  $M_j$  is a  $n_j \times n_j$  square matrix whose eigenvalues have all real part equal to  $\xi_j$ . Notice that, up to applying a further change of coordinates in each space  $E_j$ , we can assume that  $M_j^T + M_j$  is arbitrarily close to the  $n_j \times n_j$  identity matrix multiplied by  $2\xi_j$ . Hence, given  $\varepsilon > 0$ , for every  $x_j \in \mathbb{R}^{n_j}$  we can assume that

$$x_j^T M_j x_j = (\xi_j + \varepsilon_{j,x_j})|x_j|^2, \quad \varepsilon_{j,x_j} \in (-\varepsilon, \varepsilon). \quad (\text{B.1})$$

Set  $\phi(t) = t + t^2$  for  $t \in [0, 1]$  and let

$$h = h_1 + \dots + h_{k-1}$$

with

$$h_j(\theta) = \frac{1}{2}(|\theta_1|^2 + \dots + |\theta_j|^2 + \phi(|\theta_{j+1}|^2 + \dots + |\theta_k|^2)).$$

Notice that  $[0, 1] \ni t \mapsto 1 - t + \phi(t) = 1 + t^2$  takes its minimal value at  $t = 0$  and its maximal value at  $t = 1$ . Hence,  $h_j$  attains its minimum (respectively, maximum) at  $\theta \in S^{d-1}$  if and only if  $|\theta_{j+1}| = \dots = |\theta_k| = 0$  (respectively,  $|\theta_1| = \dots = |\theta_j| = 0$ ). In particular, (a) and (b) are satisfied.

Let us now turn to the proof of (c). For  $\theta \in S^{d-1}$  and  $j \in \llbracket 1, k-1 \rrbracket$ , denote  $t_j = |\theta_{j+1}|^2 + \dots + |\theta_k|^2$ . The derivative of  $h_j$  at  $\theta$  in the direction  $F_M(\theta)$  is equal to

$$\begin{aligned} \nabla h_j(\theta) \cdot (M\theta - (\theta \cdot M\theta)\theta) &= \sum_{\ell=1}^j (\theta_\ell^T M_\ell \theta_\ell - |\theta_\ell|^2 \sum_{i=1}^k \theta_i^T M_i \theta_i) \\ &\quad + \phi'(t_j) \sum_{\ell=j+1}^k (\theta_\ell^T M_\ell \theta_\ell - |\theta_\ell|^2 \sum_{i=1}^k \theta_i^T M_i \theta_i). \end{aligned}$$

For every  $\ell \in \llbracket 1, N \rrbracket$ , let us rewrite the term  $\theta_\ell^T M_\ell \theta_\ell - |\theta_\ell|^2 \sum_{i=1}^k \theta_i^T M_i \theta_i$  using (B.1) as



follows:

$$\begin{aligned}
\boldsymbol{\theta}_\ell^T M_\ell \boldsymbol{\theta}_\ell - |\boldsymbol{\theta}_\ell|^2 \sum_{i=1}^k \boldsymbol{\theta}_i^T M_i \boldsymbol{\theta}_i &= (\xi_\ell + \varepsilon_{\ell, \boldsymbol{\theta}_\ell}) |\boldsymbol{\theta}_\ell|^2 - |\boldsymbol{\theta}_\ell|^2 \sum_{i=1}^k (\xi_i + \varepsilon_{i, \boldsymbol{\theta}_i}) |\boldsymbol{\theta}_i|^2 \\
&= (\xi_\ell + \varepsilon_{\ell, \boldsymbol{\theta}_\ell}) |\boldsymbol{\theta}_\ell|^2 - |\boldsymbol{\theta}_\ell|^2 (\xi_\ell + \varepsilon_{\ell, \boldsymbol{\theta}_\ell} \\
&\quad + \sum_{i \neq \ell} (\xi_i - \xi_\ell + \varepsilon_{i, \boldsymbol{\theta}_i} - \varepsilon_{\ell, \boldsymbol{\theta}_\ell}) |\boldsymbol{\theta}_i|^2) \\
&= -|\boldsymbol{\theta}_\ell|^2 \sum_{i \neq \ell} (\xi_i - \xi_\ell + \varepsilon_{i, \boldsymbol{\theta}_i} - \varepsilon_{\ell, \boldsymbol{\theta}_\ell}) |\boldsymbol{\theta}_i|^2,
\end{aligned}$$

where the middle equality uses  $\sum_{i=1}^k |\boldsymbol{\theta}_i|^2 = 1$ . Hence,

$$\begin{aligned}
\nabla h_j(\boldsymbol{\theta}) \cdot (M\boldsymbol{\theta} - (\boldsymbol{\theta} \cdot M\boldsymbol{\theta})\boldsymbol{\theta}) &= - \sum_{\ell=1}^j \sum_{i \neq \ell} (\xi_i - \xi_\ell + \varepsilon_{i, \boldsymbol{\theta}_i} - \varepsilon_{\ell, \boldsymbol{\theta}_\ell}) |\boldsymbol{\theta}_i|^2 |\boldsymbol{\theta}_\ell|^2 \\
&\quad - \phi'(t_j) \sum_{\ell=j+1}^k \sum_{i \neq \ell} (\xi_i - \xi_\ell + \varepsilon_{i, \boldsymbol{\theta}_i} - \varepsilon_{\ell, \boldsymbol{\theta}_\ell}) |\boldsymbol{\theta}_i|^2 |\boldsymbol{\theta}_\ell|^2 \\
&= \sum_{\ell=1}^j \sum_{i=j+1}^k (1 - \phi'(t_j)) (\xi_\ell - \xi_i + \varepsilon_{\ell, \boldsymbol{\theta}_\ell} - \varepsilon_{i, \boldsymbol{\theta}_i}) |\boldsymbol{\theta}_i|^2 |\boldsymbol{\theta}_\ell|^2.
\end{aligned}$$

We have  $1 - \phi'(t_j) = -2t_j \leq 0$  for every  $t_j \in [0, 1]$  with equality holding only for  $t_j = 0$ . Moreover,  $\xi_\ell - \xi_i + \varepsilon_{\ell, \boldsymbol{\theta}_\ell} - \varepsilon_{i, \boldsymbol{\theta}_i} > 0$  for  $\varepsilon$  small enough, since  $\ell < i$ . Hence  $\nabla h_j(\boldsymbol{\theta}) \cdot (M\boldsymbol{\theta} - (\boldsymbol{\theta} \cdot M\boldsymbol{\theta})\boldsymbol{\theta}) < 0$  if  $t_j > 0$  and if there exist  $\ell \in \llbracket 1, j \rrbracket$  and  $i \in \llbracket j+1, k \rrbracket$  such that  $|\boldsymbol{\theta}_i| |\boldsymbol{\theta}_\ell| \neq 0$ , that is, if  $\boldsymbol{\theta} \notin V_j \cup W_j$ , where

$$V_j = E_1 \oplus \cdots \oplus E_j, \quad W_j = E_{j+1} \oplus \cdots \oplus E_k.$$

This proves (c), since if  $\boldsymbol{\theta} \in S^{d-1} \setminus \bigcup_{i=1}^k E_i$ , then there exists  $j$  such that  $\boldsymbol{\theta} \notin V_j \cup W_j$ .

We are left to prove (d). If  $\boldsymbol{\theta}$  is in  $S^{d-1} \cap E_i$ , then it is in  $V_j$  for  $j \geq i$  and in  $W_j$  for  $j < i$ . The proof works by computing  $\nabla^2 h_j$  both on  $S^{d-1} \cap V_j$  and  $S^{d-1} \cap W_j$  for  $j \in \llbracket 1, k \rrbracket$ .

In order to compute  $\nabla^2 h_j$ , we first extend  $h_j$  to a function  $H_j : x \mapsto \frac{1}{2}(|x_1|^2 + \cdots + |x_j|^2 + \phi(|x_{j+1}|^2 + \cdots + |x_k|^2))$  on  $\mathbb{R}^d$ . Then

$$\nabla h_j(\boldsymbol{\theta}) = \nabla H_j(\boldsymbol{\theta}) - (\boldsymbol{\theta} \cdot \nabla H_j(\boldsymbol{\theta}))\boldsymbol{\theta}$$

and

$$\nabla^2 h_j(\boldsymbol{\theta})v = \nabla^2 H_j(\boldsymbol{\theta})v - (\boldsymbol{\theta} \cdot \nabla^2 H_j(\boldsymbol{\theta})v)\boldsymbol{\theta} - (v \cdot \nabla H_j(\boldsymbol{\theta}))\boldsymbol{\theta} - (\boldsymbol{\theta} \cdot \nabla H_j(\boldsymbol{\theta}))v,$$

for  $v$  in  $T_\theta S^{d-1}$ , where the latter is identified with a linear subspace of  $\mathbb{R}^d$ .

A direct computation (using that  $\phi'(0) = 1$  and  $\phi'(1) = 3$ ) shows that if  $\boldsymbol{\theta}$  is in  $S^{d-1} \cap V_j$  then  $\nabla^2 h_j(\boldsymbol{\theta}) = 0$ , while if  $\boldsymbol{\theta}$  is in  $S^{d-1} \cap W_j$  then  $\nabla^2 h_j(\boldsymbol{\theta})v = -2\text{pr}_j(v)$ . Hence, for  $\boldsymbol{\theta} \in S^{d-1} \cap E_i$  and  $v \in T_\theta S^{d-1}$ ,

$$v \cdot \nabla^2 h(\boldsymbol{\theta})v = -2 \sum_{j=1}^{i-1} |\text{pr}_j(v)|^2 \leq -2 |\text{pr}_{i-1}(v)|^2. \quad \square$$

We can now proceed to the proof of Proposition 6.5.

*Proof of Proposition 6.5.* Fix  $\varepsilon, K$  as in the statement of the proposition and let  $h$  be given by Lemma B.1. To simplify the notations in this proof, for  $i \in \llbracket 1, N \rrbracket$ , we denote  $F_{A_i}$  simply by  $F_i$ .

Notice that, for  $j \in \llbracket 1, k \rrbracket$  and  $\theta \in S^{d-1} \cap E_j$ , we have  $\nabla^2 h(\theta) F_M(\theta) = 0$ . Indeed, if  $\theta$  is a generalized eigenvector of  $M$  in  $E_j$  then  $F_M(\theta) \in E_j$  and the conclusion follows from Item (a) in Lemma B.1. As a consequence, for  $j \in \llbracket 1, k \rrbracket$ ,  $\theta \in S^{d-1} \cap E_j$ , and  $i \in \llbracket 1, N \rrbracket$ ,

$$F_i(\theta) \cdot \nabla((F_i - F_M) \cdot \nabla h)(\theta) = F_i(\theta) \cdot \nabla^2 h(\theta) F_i(\theta),$$

where we again used Item (a) in Lemma B.1.

Let  $\delta = \varepsilon \min_{i \in \llbracket 1, N \rrbracket} \pi_i / 8$ . Thanks to Lemma B.1 and Condition (C), up to multiplying  $h$  by a positive constant, there exist a map  $s : \llbracket 2, k \rrbracket \rightarrow \llbracket 1, N \rrbracket$  and  $K_1, \dots, K_k$  disjoint neighborhoods in  $S^{d-1}$  of, respectively,  $S^{d-1} \cap E_1, \dots, S^{d-1} \cap E_k$  such that  $K_1 \subset K$  and

$$\begin{aligned} \max_{j \in \llbracket 2, k \rrbracket} \max_{\theta \in K_j} F_{s(j)}(\theta) \cdot \nabla((F_{s(j)} - F_M) \cdot \nabla h)(\theta) &\leq -1, \\ \max_{j \in \llbracket 1, k \rrbracket} \max_{\theta \in K_j} \max_{i \in \llbracket 1, N \rrbracket} F_i(\theta) \cdot \nabla((F_i - F_M) \cdot \nabla h)(\theta) &\leq \delta. \end{aligned}$$

For all  $j \in \llbracket 2, k \rrbracket$  we consider  $\psi_j \in \mathcal{C}^2(S^{d-1})$  such that  $\psi_j(\theta) = 1$  for all  $\theta \in S^{d-1} \cap K_j$  and 0 for all  $\theta \in S^{d-1} \cap K_\ell$  with  $\ell \neq j$ .

The generator of  $(\theta(t), \sigma(t))_{t \geq 0}$  being given by

$$Lg(\theta, \sigma) = F_\sigma(\theta) \cdot \nabla_\theta g(\theta, \sigma) + \mu \left( \sum_{i=1}^N \pi_i g(\theta, i) - g(\theta, \sigma) \right),$$

we consider the Lyapunov function

$$f(\theta, \sigma) = h(\theta) + \frac{1}{\mu} (F_\sigma(\theta) - F_M(\theta)) \cdot \nabla h(\theta) - \frac{1}{4\mu^2} \sum_{j=2}^k \psi_j(\theta) \mathbb{1}_{s(j)}(\sigma)$$

to get

$$\begin{aligned} Lf(\theta, \sigma) &= F_M(\theta) \cdot \nabla h(\theta) + \frac{1}{\mu} F_\sigma(\theta) \cdot \nabla((F_\sigma - F_M) \cdot \nabla h)(\theta) \\ &\quad - \frac{1}{4\mu^2} \sum_{j=2}^k [F_\sigma(\theta) \cdot \nabla \psi_j(\theta) \mathbb{1}_{s(j)}(\sigma) + \mu \psi_j(\theta) (\pi_{s(j)} - \mathbb{1}_{s(j)}(\sigma))]. \end{aligned}$$

We distinguish four cases. First, for all  $\theta \in S^{d-1} \setminus (\cup_{j=1}^k K_j)$  and  $\sigma \in \llbracket 1, N \rrbracket$ , we simply bound

$$Lf(\theta, \sigma) \leq -\alpha + C \left( \frac{1}{\mu} + \frac{1}{\mu^2} \right),$$

where

$$\alpha = - \sup_{\theta \in S^{d-1} \setminus (\cup_{j=1}^k K_j)} F_M(\theta) \cdot \nabla h(\theta) > 0$$

and  $C > 0$  is some constant independent of  $\mu$ . Second, for all  $\theta \in K_1$  and  $\sigma \in \llbracket 1, N \rrbracket$ ,

$$Lf(\theta, \sigma) \leq \frac{1}{\mu} F_\sigma(\theta) \cdot \nabla((F_\sigma - F_M) \cdot \nabla h)(\theta) \leq \frac{\delta}{\mu}.$$

Third, for all  $j \in \llbracket 2, k \rrbracket$  and  $\theta \in K_j$ ,

$$Lf(\theta, s(j)) \leq \frac{1}{\mu} F_{s(j)}(\theta) \cdot \nabla((F_{s(j)} - F_M) \cdot \nabla h)(\theta) - \frac{1}{4\mu}(\pi_i - 1) \leq -\frac{1}{4\mu}.$$

Fourth, for all  $j \in \llbracket 2, k \rrbracket$ ,  $\theta \in K_j$ , and  $\sigma \neq s(j)$ ,

$$Lf(\theta, \sigma) \leq \frac{1}{\mu} F_\sigma(\theta) \cdot \nabla((F_\sigma - F_M) \cdot \nabla h)(\theta) - \frac{1}{4\mu} \pi_{s(j)} \leq \frac{4\delta - \pi_{s(j)}}{4\mu} \leq -\frac{\min_{i \in \llbracket 1, N \rrbracket} \pi_i}{8\mu}.$$

Gathering these four cases, for  $\mu$  large enough we get

$$Lf(\theta, \sigma) \leq \frac{\min_{i \in \llbracket 1, N \rrbracket} \pi_i}{8\mu} (\varepsilon \mathbb{1}_{\theta \in K_1} - \mathbb{1}_{\theta \notin K_1}).$$

Any invariant measure  $\rho$  for  $L$  thus satisfies

$$0 = \rho Lf \leq \frac{\min_{i \in \llbracket 1, N \rrbracket} \pi_i}{8\mu} (\varepsilon \rho(K_1 \times \llbracket 1, N \rrbracket) - 1 + \rho(K_1 \times \llbracket 1, N \rrbracket)),$$

hence  $\rho(K_1 \times \llbracket 1, N \rrbracket) \geq 1/(1 + \varepsilon) \geq 1 - \varepsilon$ . This concludes since  $K_1 \subset K$ .  $\square$

## B.2 Proof of Lemma 6.7

Assume by contradiction that there exist an open nonempty subset  $B$  of  $\text{co}(A)$  and two positive integers  $j \leq k$  in  $\{2, 3\}$  so that for every matrix  $M \in B$ , using the notation of Definition 6.3, for every  $i \in \llbracket 1, N \rrbracket$  there exists  $\theta \in E_j \cap S^{d-1}$  with  $F_{A_i}(\theta) \in \bigoplus_{r \geq j} E_r$ . Denote by  $n_l$  the dimension of the generalized eigenspace  $E_l$  for  $l \in \llbracket 1, k \rrbracket$ . By eventually shrinking  $B$ , the integers  $n_l$  do not depend on  $M$  in  $B$ . To further simplify the discussion, we will assume that  $k = d$ , i.e., the  $n_l$  are all equal to one, since, otherwise, up to replacing the eigenvectors by projectors on the spaces  $E_l$  along the direct sum of the other  $E_{l'}$ , the subsequent computations carry over.

Let us parameterize  $B$  by an open neighborhood  $S$  of  $0$  in  $\mathbb{R}^m$ , where  $m$  denotes the dimension of  $\text{co}(A)$ . Note that one can choose the assignment  $s \mapsto M(s)$  affine in  $s$ . It is standard that the assignments given by  $s \mapsto (\lambda_l(M(s)))_{1 \leq l \leq d}$  and  $s \mapsto (v_l(M(s)))_{1 \leq l \leq d}$  for the spectrum of  $M(s)$  and a basis of its unit length eigenvectors define smooth functions on  $S$ . In the sequel we simply write  $\lambda_l(s)$  and  $v_l(s)$  and, to highlight the fact that the spaces  $E_l$  depend on  $M(s)$ , we write them as  $E_l(s)$ . (Note that if  $n_l > 1$  for some  $l \leq k$ , then the map assigning to every  $s \in S$  the projector on  $E_l(s)$  is again smooth.)

We use  $D(s)$  and  $V(s)$  to denote, respectively, the diagonal matrix made of the eigenvalues of  $M(s)$  and the matrix in  $\mathcal{M}_d(\mathbb{R})$  with columns  $v_l(s)$  and set  $W(s) = (V(s)^T)^{-1}$ , whose columns we denote by  $w_l(s)$ ,  $l \in \llbracket 1, d \rrbracket$ . We summarize these notations with the relations

$$M(s)V(s) = V(s)D(s), \quad W^T(s)V(s) = \text{Id}, \quad \text{for all } s \in S. \quad (\text{B.2})$$

Notice, moreover, that

$$v_l^T(s)v_l'(s) = 0, \quad \text{for all } s \in S \text{ and } l \in \llbracket 1, d \rrbracket, \quad (\text{B.3})$$

where by  $'$  we denote the differentiation with respect to  $s \in S$ . (One can either interpret such a differentiation in tensorial sense in the computations below, or simply consider a directional derivative in the space  $\mathbb{R}^m$  along an arbitrary direction.) Note also that  $M^T(s)W(s) = W(s)D(s)$  for  $s \in S$ , i.e., the vectors  $w_l(s)$  are eigenvectors of  $M^T$ .

We only treat the case where  $j = 2$  since the remaining case for  $j = k = d = 3$  is even simpler. Then one can choose  $\theta = v_2(s)$  in  $E_2(s) \cap S^{d-1}$  and the above assumption on  $F_{A_i}$  reads

$$F_{A_i}(v_2(s)) \in \bigoplus_{r \geq 2} E_r(s), \quad \text{for all } s \in S \text{ and } i \in \llbracket 1, N \rrbracket.$$

In turn the above equation reduces to

$$w_1^T(s)A_i v_2(s) = 0, \quad \text{for all } s \in S \text{ and } i \in \llbracket 1, N \rrbracket. \quad (\text{B.4})$$

In the sequel, for simplicity, we drop the variable  $s$  from the notations. Set  $R = V^{-1}V'$  and note that

$$v_l' = \sum_{q=1}^d r_{ql} v_q \quad \text{for } l \in \llbracket 1, d \rrbracket, \quad (\text{B.5})$$

where  $(r_{lq})_{lq} = R$ . Differentiating the first equation in (B.2) with respect to  $s$ , replacing  $V'$  by  $VR$ , left multiplying by  $W^T$ , and using that  $W^T M V = D$  yields

$$W^T M' V = [R, D] + D'. \quad (\text{B.6})$$

Focusing on the coefficient  $(1, 2)$  in the above equation and taking into account (B.4) and the fact that the eigenvalues are distinct, one deduces that  $r_{12} \equiv 0$ .

When  $d = 2$ , using (B.3) and (B.5) for  $l = 2$  and the above, we deduce that  $v_2$  is a constant vector, and hence the line supported by  $v_2$  is invariant by every matrix of  $A$ , contradicting the irreducibility assumption.

If  $d = 3$ , we further differentiate (B.6) to deduce that

$$[W^T M' V, R] = [R', D] + [R, D'] + D'',$$

where we have used the fact that  $M'' \equiv 0$  and the relation  $W' = -WR^T$ . Plugging again (B.6), we obtain that

$$[R, [D, R]] = [R', D] + 2[R, D'] + D''.$$

Again, considering only the coefficient  $(1, 2)$  in the above equation, we deduce that  $r_{13}r_{32} \equiv 0$ . If  $r_{13}$  is not identically equal to zero, then there exists an open subset of  $S$  where  $r_{32} \equiv 0$ . Using (B.5) for  $l = 2$  and the above, we deduce that  $v_2$  is a constant vector and we have a contradiction as previously. Assume now that  $r_{13} \equiv 0$ . Using again that  $W' = -WR^T$ , we have that  $w_1' = -r_{11}w_1$ . We deduce that the line spanned by  $w_1(0)$  is invariant by the matrices  $A_1^T, \dots, A_N^T$ , contradicting again the irreducibility of  $A$ . That concludes the proof of the lemma.  $\square$

## References

- [1] M. Balde, U. Boscain, and P. Mason. A note on stability conditions for planar switched systems. *Internat. J. Control*, 82(10):1882–1888, 2009.
- [2] J. Beltrán and C. Landim. Tunneling and metastability of continuous time Markov chains. *J. Stat Phys*, 140:1065–1114, 2010.
- [3] J. Beltrán and C. Landim. Metastability of reversible finite state Markov processes. *Stochastic Process. Appl.*, 121(8):1633–1677, 2011.
- [4] M. Benaïm, S. Le Borgne, F. Malrieu, and P.-A. Zitt. On the stability of planar randomly switched systems. *Ann. Appl. Probab.*, 24(1):292–311, 2014.
- [5] M. Benaïm and E. Strickler. Random switching between vector fields having a common zero. *Ann. Appl. Probab.*, 29(1):326–375, 02 2019.
- [6] Y. Chitour, P. Mason, and M. Sigalotti. On the marginal instability of linear switched systems. *Systems Control Lett.*, 61(6):747–757, 2012.
- [7] Y. Chitour, G. Mazanti, and M. Sigalotti. On the gap between deterministic and probabilistic joint spectral radii for discrete-time linear systems. *Linear Algebra Appl.*, 613:24–45, 2021.
- [8] I. Daubechies and J. C. Lagarias. Sets of matrices all infinite products of which converge. *Linear Algebra Appl.*, 161:227–263, 1992.
- [9] M. H. A. Davis. Piecewise-deterministic Markov processes: a general class of non-diffusion stochastic models. *J. Roy. Statist. Soc. Ser. B*, 46(3):353–388, 1984.
- [10] N. H. Du, A. Hening, D. H. Nguyen, and G. Yin. Dynamical systems under random perturbations with fast switching and slow diffusion: Hyperbolic equilibria and stable limit cycles. *J. Differential Equations*, 293:313–358, 2021.
- [11] A. Durmus, A. Guillin, and P. Monmarché. Geometric ergodicity of the bouncy particle sampler. *Ann. Appl. Probab.*, 30(5):2069–2098, 10 2020.
- [12] A. Durmus, A. Guillin, and P. Monmarché. Piecewise deterministic Markov processes and their invariant measures. *Ann. Inst. Henri Poincaré Probab. Stat.*, 57(3):1442–1475, 2021.
- [13] M. I. Freidlin and A. D. Wentzell. *Random perturbations of dynamical systems*, volume 260 of *Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer, Heidelberg, third edition, 2012.
- [14] H. Furstenberg and H. Kesten. Products of random matrices. *Ann. Math. Statist.*, 31:457–469, 1960.
- [15] L. Gerencsér, G. Michaletzky, and Z. Orlovits. Stability of block-triangular stationary random matrices. *Systems Control Lett.*, 57(8):620–625, 2008.

- [16] R. Jungers. *The joint spectral radius. Theory and applications*, volume 385 of *Lecture Notes in Control and Information Sciences*. Springer-Verlag, Berlin, 2009.
- [17] C. Landim. A topology for limits of Markov chains. *Stochastic Process. Appl.*, 125(3):1058–1088, 2015.
- [18] C. Landim, M. Loulakis, and M. Mourragui. Metastable Markov chains: from the convergence of the trace to the convergence of the finite-dimensional distributions. *Electron. J. Probab.*, 23:Paper No. 95, 34 pp., 2018.
- [19] C. Landim and T. Xu. Metastability of finite state Markov chains: a recursive procedure to identify slow variables for model reduction. *ALEA Lat. Am. J. Probab. Math. Stat.*, 13(2):725–751, 2016.
- [20] D. Liberzon. *Switching in Systems and Control*. Birkhäuser Boston, 2003.
- [21] V. Y. Protasov and A. S. Voynov. Matrix semigroups with constant spectral radius. *Linear Algebra Appl.*, 513:376–408, 2017.
- [22] 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].
- [23] 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.
- [24] F. Wirth. The generalized spectral radius and extremal norms. *Linear Algebra Appl.*, 342:17–40, 2002.