Fluctuations of the Empirical Measure of Freezing Markov Chains

Florian Bouguet, Bertrand Cloez

Fluctuations of the Empirical Measure of Freezing Markov Chains

Florian Bouguet Bertrand Cloez

Inria Nancy – Grand Est, BIGS, IECL
MISTEA, INRA, Montpellier SupAgro, Univ. Montpellier

Abstract: In this work, we consider a finite-state inhomogeneous-time Markov chain whose probabilities of transition from one state to another tend to decrease over time. This can be seen as a cooling of the dynamics of an underlying Markov chain. We are interested in the long time behavior of the empirical measure of this freezing Markov chain. Some recent papers provide almost sure convergence and convergence in distribution in the case of the freezing speed $1/n^\theta$, with different limits depending on $\theta < 1, \theta = 1$ or $\theta > 1$. Using stochastic approximation techniques, we generalize these results for any freezing speed, and we obtain a better characterization of the limit distribution as well as rates of convergence as well as functional convergence.

Contents

Keywords: Markov chain; Long-time behavior; Piecewise-deterministic Markov process; Ornstein-Uhlenbeck process; Asymptotic pseudotrajectory
MSC 2010: 60J10; 60J25; 60F05

1 Introduction

Let $(i_n)_{n\geq 1}$ be an inhomogeneous-time Markov chain with state space $\{1,\dots,D\}$ with the following transitions when $i \neq j$:

$$\mathbb{P}\left(i_{n+1}=j|i_n=i\right)=q_n(i,j), \quad q_n(i,j)=p_n(q(i,j)+r_n(i,j)),$$

where $(p_n)_{n\geq 1}$ is a decreasing sequence converging toward some $p \in [0,1]$, the remainders $r_n(i,j)$ tend to 0 (fast enough) and q is the discrete generator of some $\{1,\dots,D\}$-valued ergodic Markov chain. This model is related to the simulated annealing algorithm, and the sequence $(p_n)_{n\geq 1}$ can be interpreted as the cooling scheme of an underlying Markov chain generated by q. If $p<1$, since $\lim_{n\to+\infty} q_n(i,j)=pq(i,j)$, the probability of $(i_n)_{n\geq 1}$ to move decreases over time, from which the appellation *freezing Markov chain*.

The behavior of $(i_n)_{n\geq 1}$ is simple enough to understand, and depends on the summability of the sequence $(p_n)_{n\geq 1}$. The chain $(i_n)_{n\geq 1}$ shall converge in distribution to the unique invariant probability $\nu^\top$ associated to q if $\sum_{n=1}^\infty p_n=+\infty$ (see Theorem 2.4 below). On the other hand, if $\sum_{n=1}^\infty p_n<+\infty$, the Markov chain shall freeze along the way, as a consequence of the Borel-Cantelli Lemma. Then, we shall assume that $\sum_{n=1}^\infty p_n=+\infty$, so that we can investigate the convergence of the empirical distribution $x_n^\top=\frac{1}{n}\sum_{k=1}^n \delta_{i_k}$.

The problem of the convergence of this empirical measure can be traced back to the thesis of Dobrušin [Dob53], and several questions are still open, as pointed out in the recent article [EV16]. Some results can be obtained from the general theory developed in [SV05, Pel12], and [DS07, EV16] study the present model. In particular, convergence results are only obtained for a freezing rate of the form $p_n=a/n^\theta$ (and $r_n(i,j)=0$). More precisely,

- if $\theta<1$ then $(x_n)_{n\geq 1}$ converges to ν in probability; see [DS07, Theorem 1.2].
- if $\theta<1/2$, then $(x_n)_{n\geq 1}$ converges to ν a.s. This can be extended to $1/2\leq\theta<1$ when the state space contains only two points; see [DS07, Theorem 1.2] and [EV16, Corollary 2].
- if $\theta<1$ and $D=2$, then, up to an appropriate scaling, the empirical measure $(x_n)_{n\geq 1}$ converges in distribution to a Gaussian distribution; see [EV16, Theorem 2].
- if $\theta=1$ then $(x_n)_{n\geq 1}$ converges in distribution, and the moments of the limit probability are explicit. If q corresponds to the complete graph (see Section 4) then this limit probability is the Dirichlet distribution. When $D=2$, this covers classical distribution such as Beta, uniform, Arcsine or Wigner distributions; see [DS07, Theorems 1.3 and 1.4] and [EV16, Theorem 1].
- when $D=2$, some convergence results are established for $(x_n)_{n\geq 1}$ for general sequences $(p_n)_{n\geq 1}$, under technical conditions that we find hard to check in practice; see [EV16, Theorem 3].

We shall refer to the case $\theta<1$ as *standard*, since it is related to classic laws of large numbers and central limit theorems. This case was called *subcritical* in [EV16], in comparison with the *critical* case $\theta=1$. Since we can slightly generalize this critical case here, the term *non-standard* will be preferred from now on. In the present article, we generalize the aforementioned results by proving that, in the standard case, if $\sum_{n=1}^\infty (p_n n^2)^{-1}<+\infty$ then $(x_n)_{n\geq 1}$ converges to ν a.s., and we also give weaker conditions for convergence in probability; this is the purpose of Theorem 2.11. Under slightly stronger assumptions and up to a rescaling, we obtain convergence of $(x_n)_{n\geq 1}$ to a Gaussian distribution with explicit variance in Theorem 2.12. Finally, if $p_n\sim a/n$, then $(x_n)_{n\geq 1}$ converges in distribution exponentially fast to a limit probability (see Theorem 2.9). This distribution is characterized as the stationary measure of a piecewise-deterministic Markov process (PDMP), possesses a density with respect to the Lebesgue measure and satisfies a system of transport equations; see Propositions 3.1 and 3.4. Furthermore, Corollary 3.9 links the standard and non-standard setting by providing a convergence of the rescaled stationary measure of the PDMP to a Gaussian distribution as the switching accelerates. We also investigate the complete

graph dynamics in Section 4 and are able to derive explicit results in Propositions 4.1 and 4.2. Most of our convergence results are also provided with quantitative speeds and functional convergences.

In contrast with the Pólya Urns model (see for instance [Gou97]), all these results of convergences in distribution are not almost sure. However, note that, by letting $p_n = 1$ for all $n \geq 1$, we can recover classical limit theorems for homogeneous-time Markov chains (see [Jon04]). Furthermore, the remainder term $r_n(i,j)$ enables us to deal with different freezing schemes (see Remark 2.3). In particular, the proofs in [DS07] and [EV16] are mainly based on the method of moments, which is why more stringent assumptions are considered there. Our approach is completely different, and is based on the theory of asymptotic pseudotrajectories detailed in [Ben99] and revisited in [BBC16].

Briefly, a sequence is an asymptotic pseudotrajectory of a flow if, for any given time window, the sequence and the flow starting from the same point evolve close to each other (see for instance [BH96, Ben99]). This definition can be formalized for dynamical systems and be extended to discrete sequences of probabilities and continuous Markov semi-groups. This theory allows us to derive the behavior of the sequence of empirical measures $(x_n)_{n\geq 1}$ from the one of auxiliary continuous-time Markov processes. The interested reader may find illustrations of this phenomenon in [BBC16, Figures 3.1, 3.2 and 3.3], see also Figure 5.1. In the present paper, depending on whether we work in a standard or non-standard setting, these processes are either a diffusive process or a switching PDMP. The careful study of these limit processes is of interest *per se*, and is done in Section 3. More precisely, Gaussian distributions appear naturally since we deal with an Ornstein-Uhlenbeck process generated by

$$\mathcal{L}_{\mathrm{O}} f(y) = -y \cdot \nabla f(y) + \nabla f(y)^{\top} \Sigma^{(p,\Upsilon)} \nabla f(y), \tag{1.1}$$

where $\Sigma^{(p,\Upsilon)}$ is a $D \times D$ real-valued matrix such that

$$\Sigma^{(p,\Upsilon)}_{k,l} = \frac{1}{1+\Upsilon} \sum_{i=1}^{D} \nu_i \left[\sum_{j=1}^{D} q(i,j)(h_{l,j} - h_{l,i})(h_{k,j} - h_{k,i}) \right) - p \left(\nu_k - \mathbb{1}_{i=k}\right) \left(\nu_l - \mathbb{1}_{i=l}\right) \Bigg], \tag{1.2}$$

with p and h respectively defined in Assumption 2.1, and in (2.6). On the contrary, we shall see that, in a non-standard framework, the empirical measure is linked to a PDMP, called *exponential zig-zag process*, generated by

$$\mathcal{L}_{\mathrm{Z}} f(x,i) = (e_i - x) \cdot \nabla_x f(x,i) + \sum_{j \neq i} a q(i,j)[f(x,j) - f(x,i)]. \tag{1.3}$$

These Markov processes shall be defined and studied more rigorously in Section 3. In particular, besides some classic long-time properties (regularity, invariant measure, rate of convergence...), we prove in Theorem 3.3 the convergence of the exponential zig-zag process to the Ornstein-Uhlenbeck process when the frequency of jumps accelerates, i.e. when $a \to +\infty$.

The rest of this paper is organized as follows. In Section 2, we specifiy the notation and assumptions mentioned earlier, that will be used in the whole paper. We also state convergence results for $(x_n)_{n\geq 1}$, which are Theorems 2.9, 2.11 and 2.12. We study the long-time behavior of the two auxiliary Markov processes in Section 3 and investigate the case of the complete graph in Section 4, for which it is possible to get explicit formulas. The paper is then concluded with the proofs of the main theorems in Section 5.

2 Freezing Markov chains

2.1 Notation

We shall use the following notation throughout the paper:

- If d is a positive integer, a multi-index is a d-tuple $N = (N_1, \dots, N_d) \in (\{0, 1, \dots\} \cup \{+\infty\})^d$; the set of multi-indices is endowed with the order $N \leq \widetilde{N}$ if, for all $1 \leq i \leq d, N_i \leq \widetilde{N}_i$. We define $|N| = \sum_{i=1}^d N_i$ and and we identify an integer N with the multi-index $(N, \dots, N)$. Likewise, for any $x \in \mathbb{R}^d$, let $|x| = \sum_{i=1}^d |x_i|$.
- For some multi-index N and an open set $U \subseteq \mathbb{R}^d, \mathscr{C}^N(U)$ is the set of functions $f : U \to \mathbb{R}$ which are N_i times continuously differentiable in the direction i. For any $f \in \mathscr{C}^N(U)$, we define
$$f^{(N)} = \partial_1^{N_1} \dots \partial_d^{N_d} f, \quad \|f^{(N)}\|_\infty = \sup_{x \in \mathbb{R}^d} |f^{(N)}(x)|.$$
When there is no ambiguity, we write $\mathscr{C}^N$ instead of $\mathscr{C}^N(U)$, and denote by $\mathscr{C}_b^N$ and $\mathscr{C}_c^N$ the respective sets of bounded $\mathscr{C}^N$ functions and of compactly supported $\mathscr{C}^N$ functions.
- Let $\triangle$ be the simplex of $\mathbb{R}^D$ defined by
$$\triangle = \left\{ (x_1, \dots, x_D) \in \mathbb{R}^D : x_i \in [0, 1], \sum_{i=1}^D x_i = 1 \right\},$$
and $E = \triangle \times \{1, \dots, D\}$.
- We denote by $\mathscr{L}(X)$ the probability distribution of a random vector X, and we identify the measures over $\{1, \dots, D\}$ with the $1 \times D$ real-valued matrices. Let $\mathbb{L}$ be the Lebesgue measure over $\mathbb{R}^D$.
- If μ, ν are probability measures and f is a function, we write $\mu(f) = \int f(x)\mu(dx)$. For a class of functions $\mathscr{F}$, we define
$$d_{\mathscr{F}}(\mu, \nu) = \sup_{f \in \mathscr{F}} |\mu(f) - \nu(f)|.$$
Note that, for every class of functions $\mathscr{F}$ considered in this paper, convergence in $d_{\mathscr{F}}$ implies (and is often equivalent to) convergence in distribution (see [BBC16, Lemma 5.1]). In particular, let
$$W(\mu, \nu) = \sup_{|f(x)-f(y)| \leq |x-y|} |\mu(f) - \nu(f)|, \quad d_{\mathrm{TV}}(\mu, \nu) = \sup_{\|f\|_\infty \leq 1} |\mu(f) - \nu(f)|$$
be respectively the Wasserstein distance and the total variation distance.
- For $\theta \in (0, +\infty)^D$, let $\mathscr{D}(\theta)$ be Dirichlet distribution over $\mathbb{R}^D$, i.e. the probability distribution with probability density function
$$x \mapsto \frac{\Gamma\left(\sum_{k=1}^D \theta_k\right)}{\prod_{k=1}^D \Gamma(\theta_k)} \prod_{k=1}^D x_k^{\theta_k - 1} \mathbb{1}_{\{x \in \triangle\}}.$$
For $\theta_1, \theta_2 > 0$, let $\beta(\theta_1, \theta_2)$ be the Beta distribution over $\mathbb{R}$, i.e. the probability distribution with probability density function
$$x \mapsto \frac{\Gamma(\theta_1 + \theta_2)}{\Gamma(\theta_1)\Gamma(\theta_2)} x^{\theta_1} (1 - x)^{\theta_2} \mathbb{1}_{0 < x < 1}.$$
- Let $x \wedge y := \min(x, y)$ and $x \vee y := \max(x, y)$ for any $x, y \in \mathbb{R}$.
- We write, for $n \geq 1, u_n = O(v_n)$ if there exists some bounded sequence $(h_n)_{n \geq 1}$ such that $u_n = h_n v_n$. Moreover, if $\lim_{n \to +\infty} h_n = 0$, then we write $u_n = o(v_n)$.

2.2 Assumptions and main results

Let D be a positive integer and $(i_n)_{n\geq 1}$ be a $\{1,\dots,D\}$-valued inhomogeneous-time Markov chain such that, $\forall i \neq j$,

$$\mathbb{P}\left(i_{n+1}=j | i_n = i\right) = q_n(i,j).$$

The following assumption, which will be in force in the rest of the paper, describes the behavior of the transitions q_n as time goes by.

Assumption 2.1 (Freezing speed)**.** *Assume that that the matrix* $\mathrm{Id}+q$ *is irreducible and, for* $n\geq 1$ *and* $i\neq j$,

$$q_n(i,j) = p_n\left(q(i,j)+r_n(i,j)\right), \tag{2.1}$$

where (p_n) *is a sequence decreasing to* $p\in[0,1]$ *such that* $\sum_{n=1}^{\infty} p_n = +\infty$, *and* $\lim_{n\to+\infty} r_n(i,j)=0$. *For* $i\neq j$, *assume* $q(i,j)\geq 0, q_n(i,j)\geq 0$ *and*

$$q(i,i) = -\sum_{j\neq i} q(i,j), \quad q_n(i,i) = -\sum_{j\neq i} q_n(i,j).$$

Note that we do not require $(p_n)_{n\geq 1}$ to converge to 0. Of course, if $p>0$, then the series $\sum_n p_n$ trivially diverges; as pointed out in the introduction, if this series converge then the problem is trivial. In fact, if $p_n=1$ and $r_n(i,j)=0$ for any integers i,j,n, then the freezing Markov chain $(i_n)_{n\geq 1}$ is a classic Markov chain. When $p=0$, the dynamics of Assumption 2.1 corresponds to the lazier and lazier random walk introduced in [BBC16].

Remark 2.2 (Irreducibility or indecomposability)**.** The irreducibility of the transition matrix $\mathrm{Id}+q$ associated to q is a classic hypothesis when it comes to Markov chains, since otherwise we can split their state space into different recurrent classes. However, the result of the present article can be extended to *indecomposable*[1] Markov chains, which is a weaker concept. For instance, the transition matrix

$$\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 1/3 & 1/3 & 1/3 \end{bmatrix}$$

is indecomposable but not irreducible. Namely, $\mathrm{Id}+q$ is irreducible if it cannot be written as

$$P^{\top}(\mathrm{Id}+q)P = \begin{bmatrix} A & 0 \\ B & A' \end{bmatrix},$$

where A, A' are square matrices and P is a permutation matrix. We could allow such a decomposition, as long as B has a nonzero entry.

In any case, $\mathrm{Id}+q$ possesses a unique absorbing class of states on which it is irreducible. Using Perron-Frobenius Theorem (see [Gan59, Theorem 2p.53]), the matrix $\mathrm{Id}+q$ possesses a unique invariant measure $\nu^{\top}$, and the associated chain converges toward it under aperiodicity assumptions (see also Remark 3.3). Note that aperiodicity hypotheses are not relevant for the freezing Markov chain whenever $p<1$, since the freezing scheme automatically provides aperiodicity to the Markov chain. ♢

Under Assumption 2.1, $\mathrm{Id}+q$ possesses a unique invariant distribution $\nu^{\top}$, which writes $\nu^{\top}q=0$; let $\nu\in\triangle$ be its associated vector.

Remark 2.3 (Interpretation of the term $r_n(i,j)$)**.** The remainder $r_n(i,j)$ in (2.1) can either model small perturbations of the main freezing speed $p_n q(i,j)$, or a multiscale freezing scheme with p_n being the slowest freezing speed. For instance, the case

$$q_n = \begin{bmatrix} -n^{-\theta} & n^{-\theta} \\ n^{-(\theta+\widetilde{\theta})} & -n^{-(\theta+\widetilde{\theta})} \end{bmatrix}, \quad \theta,\widetilde{\theta}>0$$

[1] The algebric term *indecomposable* also exists for matrices, and is sometimes mistaken for irreducibility. Throughout this paper, a Markov chain (or its associated transistion matrix) is said *indecomposable* if it admits a unique recurrent class.

is covered by Assumption 2.1, with

$$q = \begin{bmatrix} -1 & 1 \\ 0 & 0 \end{bmatrix}, \quad p_n = n^{-\theta}.$$

♢

The following result characterizes the long-time behavior of the inhomogeneous Markov chain $(i_n)_{n\geq 1}$.

Theorem 2.4 (Convergence of the freezing Markov chain)**.** *Under Assumption 2.1, if either $p < 1$, or $p = 1$ and $\mathrm{Id} + q$ is aperiodic,*

$$\lim_{n\to+\infty} i_n = \nu^\top \text{ in distribution.}$$

Now, let us define $(e_1, \dots, e_D)$ the natural basis of $\mathbb{R}^D$ and introduce two different scaling rates

$$\gamma_n = \frac{1}{n}, \quad \alpha_n = \sqrt{\frac{p_n}{\gamma_n}}, \tag{2.2}$$

and the associated rescaled vectors

$$x_n = \gamma_n \sum_{k=1}^{n} e_{i_k}, \quad y_n = \alpha_n(x_n - \nu). \tag{2.3}$$

It is clear that (2.3) writes

$$x_{n+1} = \frac{\gamma_{n+1}}{\gamma_n} x_n + \gamma_{n+1} e_{i_{n+1}}, \tag{2.4}$$

that the vector x_n belongs to the simplex $\triangle$ and that $(x_n, i_n) \in E = \triangle \times \{1, \dots, D\}$. We highlight the fact that, in general, the sequence $(x_n)_{n\geq 1}$ is not a Markov chain by itself, but $(x_n, i_n)_{n\geq 1}$ is.

Remark 2.5 (Interpretation of $\triangle$)**.** The transpose $x \mapsto x^\top$ is a natural bijection between $\triangle$ and the set of probability measures over $\{1, \dots, D\}$. Then, the sequence $(x_n^\top)_{n\geq 1}$ can be viewed as the sequence of empirical measures of the Markov chain $(i_n)_{n\geq 1}$. From that viewpoint, we highlight the fact that the L^1 norm over $\triangle$ can be interpreted (up to a multiplicative constant) as the total variation distance: indeed, for any $x, \widetilde{x} \in \triangle$,

$$|x - \widetilde{x}| = \frac{1}{2} d_{\mathrm{TV}}\left(x^\top, \widetilde{x}^\top\right) = \frac{1}{2} d_{\mathrm{TV}}\left(\sum_{i=1}^{D} x_i \delta_i, \sum_{i=1}^{D} \widetilde{x}_i \delta_i\right).$$

♢

Remark 2.6 (Weighted means)**.** Note that one could consider weighted means of the form

$$x_n = \frac{1}{\sum_{k=1}^{n} \omega_k} \sum_{k=1}^{n} \omega_k e_{i_k},$$

for any sequence of positive weights $(\omega_n)_{n\geq 1}$, as in [BC15, Remark 1.1] or [BBC16, Section 3.1]. Then, we define $\gamma_n = \sum_{k=1}^{n} \omega_k$, and Theorem 2.11 below still holds with the bound

$$|x_n - \nu| \leq C \exp\left(-v \sum_{k=1}^{n} \gamma_k\right).$$

♢

Following [Ben99, BBC16], and given sequences $(\gamma_n)_{n\geq 1}, (\epsilon_n)_{n\geq 1}$, we define the following parameter which rules the speed of convergence in the context of standard fluctuations:

$$\lambda(\gamma, \epsilon) = -\limsup_{n\to+\infty} \frac{\log(\gamma_n \vee \epsilon_n)}{\sum_{k=1}^{n} \gamma_k}. \tag{2.5}$$

Finally, we need to introduce a fundamental tool in the study of the standard fluctuations: the matrix h, which is solution of the multidimensional Poisson equation

$$\sum_{j\neq i} q(i,j)(h_j - h_i) = \nu - e_i, \quad \text{or equivalently } \sum_{j\neq i} q(i,j)(h_{k,j} - h_{k,i}) = \nu_k - \mathbb{1}_{i=k} \tag{2.6}$$

for all $1 \leq i,k \leq D$, where we denoted by h_i the i-th column vector of the matrix h. This solution is classically defined by

$$h_i = -\int_0^{+\infty} \left(e_i^\top \, \mathrm{e}^{t(\mathrm{Id}\,+q)} - \nu^\top \right) dt.$$

With the help of Perron-Frobenius Theorem (see [Gan59, Theorem 2p.53]), it is easy to see that h is well-defined.

Throughout the paper, we shall treat two different cases, which entail different limit behaviors for the fluctuations of $(x_n)_{n\geq 1}$ or $(y_n)_{n\geq 1}$. Each of these cases corresponds to one of the two following assumptions.

Assumption 2.7 (Non-standard behavior)**.** *Assume that*

$$p_n \underset{n\to+\infty}{\sim} \frac{a}{n}.$$

Note that, under Assumption 2.7, the sequences $(\gamma_n)_{n\geq 1}$ and $(p_n)_{n\geq 1}$ are equivalent up to a multiplicative constant and the scaling $(\alpha_n)_{n\geq 1}$ is trivial, hence we are not interested in the behavior of $(y_n)_{n\geq 1}$.

Assumption 2.8 (Standard behavior)**.** *i) Assume that*

$$\limsup_{n\to+\infty} \frac{\gamma_n}{p_n} = 0.$$

ii) Assume that

$$\frac{p_{n+1}}{p_n} = 1 + \frac{\Upsilon}{n} + o\left(\frac{1}{n}\right), \quad \lim_{n\to+\infty} \frac{R_n}{\sqrt{p_n\gamma_n}} = 0,$$

with $R_n = \sup_i \sum_{j\neq i} |r_n(i,j)|$.

Now, we have all the tools needed to study the behavior of the empirical measure $(x_n)_{n\geq 1}$.

Theorem 2.9 (Non-standard fluctuations)**.** *Under Assumptions 2.1 and 2.7,*

$$\lim_{n\to+\infty} (x_n, i_n) = \pi \text{ in distribution,}$$

where π is characterized in Propositions 3.1 and 3.4.

Moreover, if there exist positive constants $A \geq 1, \theta \leq 1$ such that

$$\max_{j\neq i}(|r_n(i,j)|) \leq \frac{A}{n^\theta},$$

then, denoting by ρ the spectral gap of $\mathrm{Id}\,+q$*, for any*

$$u < \frac{\theta}{A + \theta\left(1 + \frac{1}{a\rho}\right)},$$

there exist a class of functions $\mathscr{F}$ defined in (5.4) and a positive constant C such that

$$d_{\mathscr{F}}\left(\mathscr{L}(x_n, i_n), \pi\right) \leq C n^{-v}.$$

It should be noted that our approach for the study of the long-time behavior of $(x_n, i_n)_{n\geq 0}$ also provides functional convergence for some interpolated process $(X_t, I_t)_{t\geq 0}$ defined in (5.3) (see Lemma 5.1, from which Theorem 2.9 is a straightforward consequence). Moreover, note that the speed of convergence provided by Theorem 2.9 writes, for any function $f : \triangle \times \{1, \dots, D\} \to \mathbb{R}$, two times differentiable in the first variable, there exists a constant C_f such that

$$|\mathbb{E}[f(x_n, i_n)] - \pi(f)| \leq C_f n^{-v}.$$

Remark 2.10 (Is it possible to generalize Assumption 2.7?)**.** This remarks leans heavily on the proof of Theorem 2.9 and may be omitted at first reading. It is interesting to wonder whether it is possible to obtain non-standard fluctuations for a more general freezing speed $(p_n)_{n\geq 1}$. To that end, let us try to mimic the computations of the proof of Lemma 5.1 with $(\widetilde{x}_n, i_n)_{n\geq 1}$ with

$$\widetilde{x}_n = \widetilde{\gamma}_n \sum_{k=1}^{n} e_{i_k},$$

for any vanishing sequences $(\gamma_n)_{n\geq 1}$ and $(\widetilde{\gamma}_n)_{n\geq 1}$. Our method being based on asymptotic pseudotrajectories, the limit of the rescaled process of $(x_n, i_n)_{n\geq 1}$ belongs to a certain class of PDMPs which can be attained if, and only if,

$$\lim_{n\to+\infty} \frac{p_n}{\gamma_n} = C_1, \quad \lim_{n\to+\infty} \frac{\widetilde{\gamma}_n}{\gamma_n} = C_2, \quad \lim_{n\to+\infty} \frac{1}{\gamma_n}\left(\frac{\widetilde{\gamma}_{n+1}}{\widetilde{\gamma}_n} - 1\right) = -C_3, \tag{2.7}$$

with $C_1, C_2, C_3 > 0$. Without loss of generality, one can choose $\gamma_n = \widetilde{\gamma}_n$ and $C_2 = C_3 = 1$. Then, the third term of (2.7) entails $\gamma_n = (n + o(1))^{-1}$ as $n \to +\infty$, which in turn implies $p_n = C_1 n^{-1} + o(n^{-1})$ when injected in the first term of (2.7).

Also, note that assuming $A < 1$ or $\theta > 1$ in Theorem 2.9 would not provide better speeds of convergence, since one would obtain a speed of the form

$$u < \frac{\theta \wedge 1}{A \vee 1 + \theta \wedge 1 \left(1 + \frac{1}{a\rho}\right)}.$$

♢

Theorem 2.11 (Standard convergence of the empirical measure)**.** *Under Assumptions 2.1 and 2.8.i),*

$$\lim_{n\to+\infty} x_n = \nu \text{ in probability,}$$

or equivalently in L^1*.*

Moreover, if $\sum_{n=1}^{\infty} \gamma_n^2 p_n^{-1} < +\infty$*, then* $\lim_{n\to+\infty} x_n = \nu$ *a.s.*

Moreover, if $\ell = \lambda(\gamma, \gamma/p) \wedge \lambda(\gamma, R) > 0$*, then, for any* $v < \ell$ *there exists a (random) constant* $C > 0$ *such that*

$$|x_n - \nu| \leq \frac{C}{n^v} \text{ a.s.}$$

Theorem 2.12 (Standard fluctuations)**.** *Under Assumptions 2.1 and 2.8,* $(y_n)_{n\geq 1}$ *converges in distribution to the Gaussian distribution* $\mathcal{N}\left(0, \Sigma^{(p,\Upsilon)}\right)$

The precise proofs of the main results are deferred to Section 5. As pointed out in the introduction, our proofs of Theorems 2.9 and 2.12 rely on comparing $(x_n)_{n\geq 1}$ and $(y_n)_{n\geq 1}$ with auxiliary continuous-time Markov processes, using the theory of asymptotic pseudotrajectories and the SDE method. Then, these discrete Markov chains will inherit some properties of the Markov processes that we shall prove in Section 3. In particular, the results we use provide functional convergence of the rescaled interpolating processes to the auxiliary Markov processes (see [BBC16, Theorem 2.12] and [Duf96, Théorème 4.II.4]).

Remark 2.13 (Examples of freezing rates)**.** For the sake of simplicity, consider $r_n(i,j) = 0$ for all i, j, n. Assumption 2.8 covers sequences $(p_n)_{n\geq 1}$ of the form $p_n = n^{-\theta}$ for any $0 < \theta < 1$, since $\gamma_n^2 p_n^{-1} = n^{\theta-2}$. In this case, $\ell = \lambda(n^{-1}, n^{\theta-1}) = 1 - \theta > 0$.

But we can also consider more exotic freezing rates, for instance $p_n = \log(n)^{\zeta} n^{-1}$, for some $\zeta \geq 1$. Then, $\gamma_n^2 p_n^{-1} = n^{-1}\log(n)^{-\zeta}$. If $\zeta > 1$, then the series converges and $\ell = 1$. Our results do not provide almost sure convergence in the case $\zeta = 1$, however, but only convergence in probability.

It should be noted that assuming that $(p_n)_{n\geq 1}$ is decreasing, $\lim_{n\to+\infty} p_n = 0$ and $\sum_{n=1}^{\infty} p_n = +\infty$ do not imply in general that $p_{n+1} \sim p_n$. A slight modification of the proof shows that, if p_{n+1} is not equivalent to p_n, we have to assume the existence of a sequence $(\beta_n)_{n\geq 1}$ such that

$$\lim_{n\to+\infty} \frac{p_n}{\gamma_n^2\beta_n^2}\left(\frac{\beta_n}{\beta_{n-1}}(1-\gamma_n) - 1\right) = -1, \quad \sum_{n=1}^{\infty} \frac{\gamma_n^2\beta_n^2}{p_n} = +\infty, \quad \lim_{n\to+\infty} \beta_n\gamma_n = -1$$

and such that the sequences $(\gamma_n^2\beta_n^2 p_n^{-1})_{n\geq 1}$ and $(\beta_n\gamma_n)_{n\geq 1}$ are decreasing; then the conclusion of Theorem 2.12 holds. ♢

3 The auxiliary Markov processes

In this section, we study the ergodicity of the processes arising as limits of the freezing Markov from Section 2. We also study their invariant measure, and provide explicit formulas when it is possible.

3.1 The exponential zig-zag process

In this section, we investigate the asymptotic properties of the exponential zig-zag process, which arise from the non-standard scaling of the Markov chain $(i_n)_{n\geq 1}$. To this end, let $(\mathbf{X}_t, \mathbf{I}_t)_{t\geq 0}$ be the strong solution of the following SDE (see [IW89]), with values in E:

$$(\mathbf{X}_t, \mathbf{I}_t) = (\mathbf{X}_0, \mathbf{I}_0) + \int_0^t \left(A(\mathbf{X}_{s^-}, \mathbf{I}_{s^-}) + e_{\mathbf{I}_{s^-}}\right) ds + \sum_{j=1}^{D} \int_0^t B_{\mathbf{I}_{s^-},j}(\mathbf{X}_{s^-}, \mathbf{I}_{s^-}) N_{\mathbf{I}_{s^-},j}(ds), \tag{3.1}$$

where the $N_{i,j}$ are independent Poisson processes of intensity $aq(i,j)\mathbb{1}_{\{i\neq j\}}$ and

$$A = \begin{bmatrix} -1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & -1 & 0 \\ 0 & \cdots & 0 & 0 \end{bmatrix}, \qquad B_{i,j} \begin{bmatrix} & & & 0 \\ & (0) & & \vdots \\ & & & 0 \\ 0 & \cdots & 0 & i-j \end{bmatrix}. \tag{3.2}$$

Thus, the infinitesimal generator of this process is $\mathcal{L}_{\mathrm{Z}}$ defined in (1.3) (see e.g. [EK86, Dav93, Kol11]). Actually, the exponential zig-zag process is a PDMP; the interested reader can consult [Dav93, BLBMZ15] for a detailed construction of the process $(\mathbf{X}, \mathbf{I})$. Let us describe briefly its dynamics: setting $\mathbf{I}_0 = i$, the process possesses a continuous component $\mathbf{X}$ which is exponentially attracted to the vector e_i. The discrete component $\mathbf{I}_t$ is piecewise-constant, and jumps from i to j following the epochs of the processes $N_{i,j}$, which in turn leads the continuous component to be attracted to e_j (see Figure 3.1 for sample paths of the exponential zig-zag process, and Figure 4.2 for a typical path in the framework of Section 4.2).

The following result might be seen as a direct consequence of [BLBMZ12, Theorem 1.10] or [CH15, Theorem 1.4], although these articles do not provide explicit rates of convergence, which are useful for instance in the proof of Corollary 3.9.

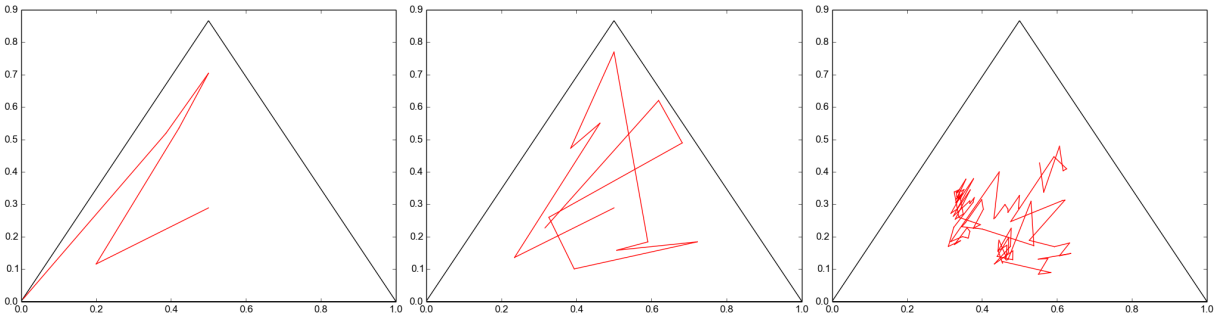


Figure 3.1: Sample paths on $[0,5]$ of the exponential zig-zag process for $X_0 = (1/3, 1/3, 1/3), q(i,j) = 1$ and $a = 0.5, a = 2, a = 20$ (from left to right).

Proposition 3.1 (Ergodicity)**.** *The exponential zig-zag process $(\mathbf{X}_t, \mathbf{I}_t)_{t\geq 0}$ admits a unique stationary distribution π. If ρ is the spectral gap of q, then for any for any $v < a\rho(1+a\rho)^{-1}$, there exists a constant $C > 0$ such that*

$$\forall t \geq 0, \quad W\left((\mathbf{X}_t, \mathbf{I}_t), \pi\right) \leq C\,\mathrm{e}^{-vt}\,.$$

Moreover, if $\mathscr{L}(\mathbf{I}_0) = \nu^\top$, then

$$\forall t \geq 0, \quad W\left((\mathbf{X}_t, \mathbf{I}_t), \pi\right) \leq W\left((\mathbf{X}_0, \mathbf{I}_0), \pi\right)\mathrm{e}^{-t}\,.$$

Note that the speed of convergence provided in Proposition 3.1 can be improved when $D = 2$, since we are able to use more refined couplings (see Proposition 4.5).

Proof of Proposition 3.1: The pattern of this proof follows [BLBMZ12]. Let $(\mathbf{X}_t, \mathbf{I}_t, \widetilde{\mathbf{X}}_t, \widetilde{\mathbf{I}}_t)_{t\geq 0}$ be the coupling for which the discrete components $\mathbf{I}$ and $\widetilde{\mathbf{I}}$ are equal forever once they are equal once. Let $t > 0$ and $\alpha \in (0,1)$. Firstly, note that, if $\mathbf{I}_{\alpha_t} = \widetilde{\mathbf{I}}_{\alpha_t}$, then the processes always have common jumps and

$$|\mathbf{X}_t - \widetilde{\mathbf{X}}_t| = |\mathbf{X}_{\alpha_t} - \widetilde{\mathbf{X}}_{\alpha_t}|\,\mathrm{e}^{-t} \leq 2\,\mathrm{e}^{-t}\,. \tag{3.3}$$

From the Perron-Frobenius theorem (see [Gan59, SC97]), for any $\varepsilon > 0$, there exists $\widetilde{C} > 0$ such that

$$d_{\mathrm{TV}}(\mathbf{I}_t, \widetilde{\mathbf{I}}_t) \leq \widetilde{C}\,\mathrm{e}^{-(a\rho-\varepsilon)t}\,.$$

Then there exists a coupling of the random variables $\mathbf{I}_{\alpha t}$ and $\widetilde{\mathbf{I}}_{\alpha t}$ such that

$$\mathbb{P}(\mathbf{I}_{\alpha t} \neq \widetilde{\mathbf{I}}_{\alpha t}) \leq \widetilde{C}\,\mathrm{e}^{-(a\rho-\varepsilon)\alpha t}\,. \tag{3.4}$$

Now, combining (3.3) and (3.4),

$$\begin{aligned}
\mathbb{E}\left[|(\mathbf{X}_t, \mathbf{I}_t) - (\widetilde{\mathbf{X}}_t, \widetilde{\mathbf{I}}_t)|\right] &\leq \mathbb{E}\left[|(\mathbf{X}_t, \mathbf{I}_t) - (\widetilde{\mathbf{X}}_t, \widetilde{\mathbf{I}}_t)|\middle|\, \mathbf{I}_{\alpha t} \neq \widetilde{\mathbf{I}}_{\alpha t}\right]\mathbb{P}(\mathbf{I}_{\alpha t} \neq \widetilde{\mathbf{I}}_{\alpha t}) \\
&\quad + \mathbb{E}\left[|(\mathbf{X}_t, \mathbf{I}_t) - (\widetilde{\mathbf{X}}_t, \widetilde{\mathbf{I}}_t)|\middle|\, \mathbf{I}_{\alpha t} \neq \widetilde{\mathbf{I}}_{\alpha t}\right]\mathbb{P}(\mathbf{I}_{\alpha t} \neq \widetilde{\mathbf{I}}_{\alpha t}) \\
&\leq 2\mathbb{P}(\mathbf{I}_{\alpha t} \neq \widetilde{\mathbf{I}}_{\alpha t}) + 2\,\mathrm{e}^{-(1-\alpha)t} \\
&\leq 2\widetilde{C}\,\mathrm{e}^{-(a\rho-\varepsilon)\alpha t} + 2\,\mathrm{e}^{-(1-\alpha)t}\,.
\end{aligned}$$

One can optimize this speed of convergence by taking $\alpha = (1 + a\rho - \varepsilon)^{-1}$, and get

$$W\left((\mathbf{X}_t, \mathbf{I}_t), (\widetilde{\mathbf{X}}_t, \widetilde{\mathbf{I}}_t)\right) \leq C\,\mathrm{e}^{-vt} \tag{3.5}$$

with $C = 2\widetilde{C} + 2$ and $v = (a\rho - \varepsilon)(1 + a\rho - \varepsilon)^{-1}$. Then, $(\mathscr{L}((\mathbf{X}_t, \mathbf{I}_t))$ is a Cauchy sequence and converges to a (stationary) distribution π. Letting $\mathscr{L}(\widetilde{\mathbf{X}}_0, \widetilde{\mathbf{I}}_0) = \pi$ in (3.5), achieves the proof in the general case.

Now, if $\mathscr{L}(\mathbf{I}_0) = \nu^\top$, then $\mathscr{L}(\mathbf{I}_0) = \mathscr{L}(\widetilde{\mathbf{I}}_0)$; we can let $\mathbf{I}_0 = \widetilde{\mathbf{I}}_0$, and then it suffices to use (3.3) with $\alpha = 0$. □

If Assumption 2.1 is in force, there exists a unique invariant measure π, which satisfies

$$\int_E \mathcal{L}_Z f(x,i)\pi(dx,di) = 0,$$

for any function f smooth enough. Now, let us establish the absolute continuity of this invariant distribution with respect to the Lebesgue measure $\mathbb{L}$.

Lemma 3.2 (Absolute continuity of the exponential zig-zag process)**.** *Let $K \subset \mathring{\triangle}$ be a compact set. There exist constants $t_0, c_0 > 0$ and a neighborhood V of K such that, for any $(x,i) \in E$ and for all $t \geq t_0$,*

$$\mathbb{P}(\mathbf{X}_t \in \cdot, \mathbf{I}_t = j | \mathbf{X}_0 = x, \mathbf{I}_0 = i) \geq c_0 \mathbb{L}(\cdot \cap V(y)). \tag{3.6}$$

Remark 3.3 (When $\mathrm{Id} + q$ is only indecomposable)**.** This remark echoes Remark 2.2 and describes the behavior of the Markov chain $(x_n, i_n)_{n\geq 1}$ when $\mathrm{Id} + q$ is reducible but indecomposable. In that case, Proposition 3.1 holds as well. However, $\mathrm{Id} + q$ possesses a unique recurrent class which is strictly contained in $\{1, \dots, D\}$, the vector ν possesses at least one zero and belongs to the frontier of the simplex $\triangle$, and $\pi(\mathring{\triangle}) = 0$. It is then impossible to obtain an equivalent to Proposition 3.1 with a convergence in total variation; when $\mathrm{Id} + q$ is irreducible, this is possible using techniques inspired from [BMP+15, Proposition 2.5].

If $\mathrm{Id} + q$ is indecomposable, one can obtain equivalents of Lemma 3.2 and Proposition 3.4 below by replacing the Lebesgue measure $\mathbb{L}$ on $\mathbb{R}^D$ by the Lebesgue measure on the linear subspace spanned by the recurrent class of $\mathrm{Id} + q$. ♢

Proof of Lemma 3.2: The proof is mainly based on Hörmander-type conditions for switching dynamical systems obtained in [BH12, BLBMZ15]. Using the notation of [BLBMZ15], let $F^i : x \mapsto e_i - x$ and then, if $D \geq 3$,

$$\forall x \in \triangle, \quad \mathcal{G}_0(x) = \mathrm{Vect}\{F^i(x) - F^j(x) : i \neq j\} = \mathrm{Vect}\{e_i - e_j : i \neq j\} = \mathbb{R}^D,$$

where Vect A denotes the vector space spanned by $A \subseteq \mathbb{R}^D$. If $D = 2$, then $\mathcal{G}_1(x) = \mathbb{R}^2$. As a consequence, the strong bracket condition of [BLBMZ15, Definition 4.3] is satisfied. In particular, using [BLBMZ15, Theorems 4.2 and 4.4], we have that, for every $x \in \triangle$, $y \in \mathring{\triangle}$, there exist $t_0(x), c_0(x) > 0$ and open sets $U_0(x), V(x,y)$, such that for all $x_0 \in U_0(x), i, j \in \{1, \dots, D\}, A \subseteq \triangle$ and $t > t_0(x)$,

$$\mathbb{P}(\mathbf{X}_t \in A, \mathbf{I}_t = j | \mathbf{X}_0 = x_0, \mathbf{I}_0 = i) \geq c_0(x) \mathbb{L}(A \cap V(x,y)).$$

Now, $\triangle = \cup_{x\in\triangle} U_0(x)$ and is compact, so there exist $x_1, ..., x_n$ such that $\triangle = \cup_{k=1}^n U_0(x_k)$. In particular, setting $V(y) = \cup_{k=1}^n V(x_k, y)$, $c_0 = \min_{1\leq k\leq n} c_0(x_k)$, $t_0 = \max_{1\leq k\leq n} t_0(x_k)$, we have, for all $x_0 \in \triangle, i, j \in \{1, \dots, D\}, A \subseteq \triangle$ and $t > t_0$,

$$\mathbb{P}(\mathbf{X}_t \in A, \mathbf{I}_t = j | \mathbf{X}_0 = x_0, \mathbf{I}_0 = i) \geq c_0 \mathbb{L}(A \cap V(y)),$$

Once again, K is compact so we can extract a finite family from the open sets $(V(y))_{y\in K}$. Using the Markov property, this holds for every $t \geq t_0$, which entails (3.6). □

Proposition 3.4 (System of transport equations for π)**.** *The distribution π introduced in Proposition 3.1 admits the following decomposition:*

$$\pi = \sum_{i=1}^{D} \nu_i \pi_i \otimes \delta_i, \quad \pi_i(dx) = \varphi(x,i)dx. \tag{3.7}$$

where the function φ satisfies, for any $(x,i) \in E$,

$$(D-1)\varphi(x,i) + \sum_{k=1}^{D} x_k \partial_k \varphi(x,i) - \partial_i \varphi(x,i) + \sum_{j=1}^{D} \frac{\nu_j}{\nu_i} aq(j,i)\varphi(x,j) = 0. \tag{3.8}$$

Once we will have proved that π admits the decomposition (3.7), the next step is the characterization of φ. Indeed, since it satisfies

$$\sum_{i=1}^{D} \nu_i \int_{\triangle} \mathcal{L}_{\mathrm{Z}} f(x,i)\varphi(x,i)dx = 0, \tag{3.9}$$

for every smooth enough function f, all we have to do is compute the adjoint operator of $\mathcal{L}_{\mathrm{Z}}$. For general switching model, it would not possible to characterize φ as a solution of a simple system of PDEs like (3.8). However, the present form of the flow enables us to derive a simple expression for the adjoint operator of $\mathcal{L}_{\mathrm{Z}}$. Before turning to the proof of Proposition 3.4, let us present the following formula of integration by parts over the simplex $\triangle$.

Lemma 3.5 (Integration by parts over $\triangle$)**.** *For all $f, g \in \mathscr{C}_c^1\left(\mathring{\triangle}\right)$, and $k, l \in \{1, \dots, D\}$, we have*

$$\int_{\triangle} g(x)(\partial_k - \partial_l) f(x) dx = -\int_{\triangle} (\partial_k - \partial_l) g(x) f(x) dx.$$

Proof of Lemma 3.5: Fix $l = 1$ and let $\triangle_1 = \left\{x_2, \dots, x_D \in [0,1] : \sum_{i=2}^{D} x_i \leq 1\right\}$. Then,

$$\begin{aligned}
&\int_{\triangle} g(x)\partial_k f(x) dx_1 \dots dx_D = \int_{\triangle_1} g\left(1 - \sum_{i=2}^{D} x_i, x_2, \dots\right) \partial_k f\left(1 - \sum_{i=2}^{D} x_i, x_2, \dots\right) dx \\
&= \int_{\triangle_1} \left[\partial_k \left(g\left(1 - \sum_{i=2}^{D} x_i, x_2, \dots\right) f\left(1 - \sum_{i=2}^{D} x_i, x_2, \dots\right)\right)\right. \\
&+ \partial_1 g\left(1 - \sum_{i=2}^{D} x_i, x_2, \dots\right) f\left(1 - \sum_{i=2}^{D} x_i, x_2, \dots\right) + g\left(1 - \sum_{i=2}^{D} x_i, x_2, \dots\right) \partial_1 f\left(1 - \sum_{i=2}^{D} x_i, x_2, \dots\right) \\
&\left. - \partial_k g\left(1 - \sum_{i=2}^{D} x_i, x_2, \dots\right) f\left(1 - \sum_{i=2}^{D} x_i, x_2, \dots\right)\right] dx_1 \dots dx_D.
\end{aligned}$$

Now, as $g(0, x_2, \dots) = f(0, x_2, \dots) = 0$ and $\partial_k 1 = 0$, use a (classic) multidimensional integration by parts to establish that

$$\int_{\triangle_1} \partial_k \left(g\left(1 - \sum_{i=2}^{D} x_i, x_2, \dots\right) f\left(1 - \sum_{i=2}^{D} x_i, x_2, \dots\right)\right) dx_1 \dots dx_D = 0,$$

which entails Lemma 3.5. □

Proof of Proposition 3.4: Integrating (3.6) with respect to the unique invariant measure π, we obtain that π admits an absolutely continuous part (note that uniqueness comes from Proposition 3.1). Since π cannot have an absolutely continuous part and a singular one (see [BH12, Theorem 6]), π admits a density with respect to the Lebesgue measure, which entails (3.7).

Now, let us characterize the function φ. We have

$$\begin{aligned}
&\sum_{i,k=1}^{D} \nu_i \int_{\triangle} (-x_k + \mathbf{1}_{i=k})\varphi(x,i)\partial_k f(x,i) dx \\
&\quad = \sum_{i=1}^{D} \left(-\sum_{k=1}^{D} \nu_i \int_{\triangle} x_k \varphi(x,i)\partial_k f(x,i) dx + \nu_i \int_{\triangle} \varphi(x,i)\partial_i f(x,i) dx\right)
\end{aligned}$$

and, using Lemma 3.5, for any $1 \leq i \leq D$,

$$
\begin{aligned}
&-\sum_{k=1}^{D} \int_{\triangle} x_k \varphi(x,i) \partial_k f(x,i) dx + \int_{\triangle} \varphi(x,i) \partial_i f(x,i) dx \\
&= -\sum_{k \neq i} \int_{\triangle} x_k \varphi(x,i) \partial_k f(x,i) dx - \int_{\triangle} x_i \varphi(x,i) \partial_i f(x,i) dx + \int_{\triangle} \varphi(x,i) \partial_i f(x,i) dx \\
&= \sum_{k \neq i} \left[\int_{\triangle} \partial_k \left(x_k \varphi(x,i) \right] f(x,i) dx - \int_{\triangle} x_k \varphi(x,i) \partial_i f(x,i) dx - \int_{\triangle} \partial_i \left(x_k \varphi(x,i) \right) f(x,i) dx \right) \\
&\quad - \int_{\triangle} x_i \varphi(x,i) \partial_i f(x,i) dx + \int_{\triangle} \varphi(x,i) \partial_i f(x,i) dx \\
&= \sum_{k \neq i} \left[\int_{\triangle} \partial_k \left(x_k \varphi(x,i) \right) f(x,i) dx - \int_{\triangle} \partial_i \left(x_k \varphi(x,i) \right) f(x,i) dx \right] \\
&= \sum_{k \neq i} \int_{\triangle} \partial_k \left(x_k \varphi(x,i) \right) f(x,i) dx - (1 - x_i) \int_{\triangle} \partial_i \varphi(x,i) f(x,i) dx.
\end{aligned}
$$

Hence, (3.9) writes

$$
\sum_{i=1}^{D} \int_{\triangle} \nu_i f(x,i) \left[\sum_{k \neq i} \partial_k \left(x_k \varphi(x,i) \right) - (1 - x_i) \partial_i \varphi(x,i) + \sum_{j=1}^{D} \frac{\nu_j}{\nu_i} q(j,i) \varphi(x,j) - \sum_{j=1}^{D} q(i,j) \varphi(x,i) \right] dx = 0.
$$

It follows that φ is the solution of (3.8). □

3.2 The Ornstein-Uhlenbeck process

In this short section, we recall a classic property of multidimensional Ornstein-Uhlenbeck processes, which is useful to characterize the behavior of $(y_n)_{n \geq 1}$ in a standard setting. Thus, we define $(\mathbf{Y}_t)_{t \geq 0}$ as the strong solution of the following SDE, with values in $\mathbb{R}^D$:

$$
\mathbf{Y}_t = \mathbf{Y}_0 - \int_0^t \mathbf{Y}_{s^-} ds + \sqrt{2} \int_0^t (\Sigma^{(p,\Upsilon)})^{1/2} dW_s, \tag{3.10}
$$

where W is a standard D-dimensional Brownian motion and $(\Sigma^{(p,\Upsilon)})^{1/2}$ is the square root of the positive-definite symmetric matrix $\Sigma^{(p,\Upsilon)}$, i.e. $(\Sigma^{(p,\Upsilon)})^{1/2} ((\Sigma^{(p,\Upsilon)})^{1/2})^{\top} = \Sigma^{(p,\Upsilon)}$. The process $\mathbf{Y}$ is a classic Ornstein-Uhlenbeck process with infinitesimal generator $\mathcal{L}_{\mathrm{O}}$ defined in (1.1). Such processes have already been thoroughly studied, so we present only the following proposition, which quantifies the speed of convergence of $\mathbf{Y}$ to its equilibrium.

Proposition 3.6 (Ergodicity of the Ornstein-Uhlenbeck process)**.** *The Markov process $(\mathbf{Y}_t)_{t \geq 0}$ generated by $\mathcal{L}_O$ in (1.1), with values in $\mathbb{R}$, admits a unique stationary distribution $\mathcal{N}\left(0, \Sigma^{(p,\Upsilon)}\right)$.*

Moreover,

$$
W\left(\mathbf{Y}_t, \mathcal{N}\left(0, \Sigma^{(p,\Upsilon)}\right)\right) = W(\mathbf{Y}_0, \pi)\, \mathrm{e}^{-t}.
$$

Proof of Proposition 3.6: First, since

$$
\mathcal{N}\left(0, \Sigma^{(p,\Upsilon)}\right)(dx) = C \exp\left(\frac{x^{\top} (\Sigma^{(p,\Upsilon)})^{-1} x}{2} \right) dx,
$$

a straightforward integration by parts shows that, for any $f \in \mathscr{C}_c^2$, $\mathcal{N}\left(0, \Sigma^{(p,\Upsilon)}\right)(\mathcal{L}_{\mathrm{O}} f) = 0$ so that $\mathcal{N}\left(0, \Sigma^{(p,\Upsilon)}\right)$ is an invariant measure for the Ornstein-Uhlenbeck process $\mathbf{Y}$.

It is well-known and easy to check that $(\mathbf{Y}_t)_{t\geq 0}$ writes

$$\mathbf{Y}_t = \mathbf{Y}_0\,\mathrm{e}^{-t} + \sqrt{2}(\Sigma^{(p,\Upsilon)})^{1/2}\int_0^t \mathrm{e}^{-(t-s)}\,dW_s,$$

where W is a standard Brownian motion. Consequently, if we consider $\widetilde{\mathbf{Y}}$ another Ornstein-Uhlenbeck process generated by $\mathcal{L}_{\mathrm{O}}$ and driven by the (same) Brownian motion W,

$$\mathbb{E}\left[|\mathbf{Y}_t - \widetilde{\mathbf{Y}}_t|\right] = \mathbb{E}\left[|\mathbf{Y}_0 - \widetilde{\mathbf{Y}}_0|\right]\mathrm{e}^{-t}. \tag{3.11}$$

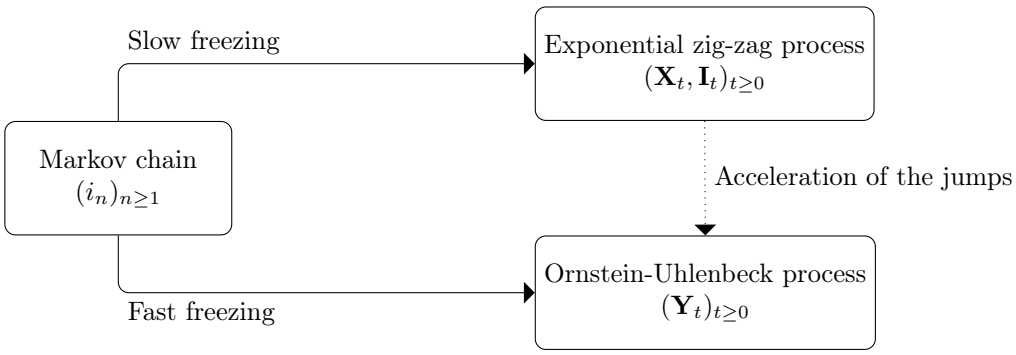
Taking the infimum over all the couplings gives a contraction in Wasserstein distance. Now, if $\mathscr{L}(\widetilde{\mathbf{Y}}_0) = \mathscr{N}\left(0,\Sigma^{(p,\Upsilon)}\right)$ and $(\mathbf{Y}_0,\widetilde{\mathbf{Y}}_0)$ is the optimal coupling between $\mathscr{L}(\mathbf{Y}_0)$ and $\mathscr{N}\left(0,\Sigma^{(p,\Upsilon)}\right)$ with respect to W, then (3.11) writes

$$W\left(\mathbf{Y}_t,\mathscr{N}\left(0,\Sigma^{(p,\Upsilon)}\right)\right) = W\left(\mathbf{Y}_0,\mathscr{N}\left(0,\Sigma^{(p,\Upsilon)}\right)\right)\mathrm{e}^{-t},$$

which entails the uniqueness of the invariant probability distribution as well as the exponential ergodicity of the process. □

3.3 Acceleration of the jumps

The current section links the Sections 3.1 and 3.2 in the following sense:



Indeed, we prove in Theorem 3.7 the convergence of the (rescaled) exponential zig-zag process to a diffusive process as the jump rates go to infinity. Such results are fairly standard and are already known in the cases of (linear) zig-zag processes (see [FGM12, BD16]) or of particle transport processes (see [CK06]). Heuristically, since there are more frequent jumps, the process tends to concentrate around its mean ν, and the effect of the discrete component fades away. This phenomenon can be seen on Figure 3.1. We shall end this section with Corollary 3.9, which provides the convergence of the stationary distribution of the exponential zig-zag process toward a Gaussian distribution.

To this end, let $(a_n)_{n\geq 1}$ be a sequence of positive numbers such that $a_n \to +\infty$ as $n \to +\infty$ and, for any integer n, let $(\mathbf{X}_t^{(n)},\mathbf{I}_t^{(n)})_{t\geq 0}$ be a Markov process with values in E generated by

$$\mathcal{L}^{(n)}f(x,i) = (e_i - x)\cdot\nabla_x f(x,i) + a_n\sum_{j\neq i} q(i,j)[f(x,j) - f(x,i)].$$

We define $\mathbf{Y}_t^{(n)} = \sqrt{a_n}(\mathbf{X}^{(n)} - \nu)$ and denote by $\mathbf{Y}_t^{(n)}(k)$ and $\mathbf{X}_t^{(n)}(k)$ the respective k^{th} component of $\mathbf{Y}_t^{(n)}$ and $\mathbf{X}_t^{(n)}$.

Theorem 3.7 (Convergence of the processes)**.** *If $(\mathbf{Y}_0^{(n)})_{n\geq 1}$ converges in distribution to a probability distribution μ, then the sequence of processes $(\mathbf{Y}^{(n)})_{n\geq 1}$ converges in distribution to the diffusive Markov process generated by*

$$\mathcal{L}_O f(y) = -y\cdot\nabla f(y) + \nabla f(y)^\top \Sigma^{(0,1)} \nabla f(y)$$

with initial condition μ.

Proof of Theorem 3.7: We shall use a diffusion approximation and follow the proof of [FGM12, Proposition 1.1]. For now, we drop the superscript (n), and let, for any $1\leq k,l\leq D$,

$$\varphi_k(x,i) = \sqrt{a_n}(x_k-\nu_k) + \frac{1}{\sqrt{a_n}} h_{k,i}, \quad \psi_{k,l}(x,i) = \varphi_k(x,i)\varphi_l(x,i).$$

Then,

$$\begin{aligned}
\mathcal{L}\varphi_k(x,i) &= \sqrt{a_n}(\nu_k - x_k),\\
\mathcal{L}\psi_{k,l}(x,i) &= \sqrt{a_n}\left((\mathbb{1}_{i=k}-x_k)\varphi_l(x,i) + (\mathbb{1}_{i=l}-x_l)\varphi_k(x,i)\right)\\
&\quad + a_n\left((x_k-\nu_k)(\nu_l-\mathbb{1}_{i=l}) + (x_l-\nu_l)(\nu_k-\mathbb{1}_{i=k})\right) + \sum_{j\neq i} q(i,j)\left(h_{k,j}h_{l,j} - h_{k,i}h_{l,i}\right).
\end{aligned}$$

Then, by Dynkin's formula, for fixed n, the processes $(M_t(k))_{t\geq 0}$ and $(N_t(k,l))_{t\geq 0}$ are local martingales with respect to the filtration generated by $(\mathbf{X}^{(n)},\mathbf{I}^{(n)})$, where

$$\begin{aligned}
M_t(k) &= \mathbf{Y}_t(k) - \sqrt{a_n}\int_0^t (\nu_k - \mathbf{X}_s(k))ds + \frac{1}{\sqrt{a_n}} h_{k,\mathbf{I}_t},\\
N_t(k,l) &= \mathbf{Y}_t(k)\mathbf{Y}_t(l) + \mathbf{Y}_t(k)h_{l,\mathbf{I}_t} + \mathbf{Y}_t(l)h_{k,\mathbf{I}_t} + \frac{1}{a_n}h_{k,\mathbf{I}_t}h_{l,\mathbf{I}_t}\\
&\quad - \int_0^t \big[-2\mathbf{Y}_s(l)\mathbf{Y}_s(k) + h_{k,\mathbf{I}_s}(\mathbb{1}_{\{\mathbf{I}_s=l\}} - \mathbf{X}_s(l)) + h_{l,\mathbf{I}_s}(\mathbb{1}_{\{\mathbf{I}_s=k\}} - \mathbf{X}_s(k))\\
&\quad + \sum_{j\neq \mathbf{I}_s} q(\mathbf{I}_s,j)\left(h_{k,j}h_{l,j} - h_{k,\mathbf{I}_s}h_{l,\mathbf{I}_s}\right)\big] ds.
\end{aligned}$$

Remark that, for any $1\leq i\leq D$, if $\sigma_{k,l}(i) = \sum_{j=1}^D q(i,j)(h_{l,j}-h_{l,i})(h_{k,j}-h_{k,i})$),

$$\begin{aligned}
\sum_{j=1}^D q(i,j)\left(h_{k,j}h_{l,j} - h_{k,i}h_{l,i}\right) &= \sum_{j=1}^D q(i,j)\left(h_{l,j}-h_{l,i}\right)\left(h_{k,j}-h_{k,i}\right) + h_{l,i}\left(\nu_k - \mathbb{1}_{i=k}\right) + h_{k,i}\left(\nu_l - \mathbb{1}_{i=l}\right)\\
&= \sigma_{k,l}(i) + h_{l,i}\left(\nu_k - \mathbb{1}_{i=k}\right) + h_{k,i}\left(\nu_l - \mathbb{1}_{i=l}\right).
\end{aligned}$$

Then, denoting by $\mathbf{Z}_s(k) = \int_0^t \mathbf{Y}_s(k)ds$,

$$\begin{aligned}
N_t(k,l) &= \mathbf{Y}_t(k)\mathbf{Y}_t(l) + 2\int_0^t \mathbf{Y}_s(k)\mathbf{Y}_s(l)ds - \int_0^t \sigma_{k,l}(\mathbf{I}_s)ds + \frac{1}{a_n}h_{k,\mathbf{I}_t}h_{l,\mathbf{I}_t}\\
&\quad + \frac{1}{\sqrt{a_n}}h_{k,\mathbf{I}_t}\left(\mathbf{Y}_t(l) + \mathbf{Z}_t(l)\right) + \frac{1}{\sqrt{a_n}}h_{l,\mathbf{I}_t}\left(\mathbf{Y}_t(k) + \mathbf{Z}_t(k)\right),
\end{aligned}$$

and

$$\begin{aligned}
M_t(k)M_t(l) &= N_t(k,l) + \mathbf{Y}_t(k)\mathbf{Z}_t(l) + \mathbf{Y}_t(l)\mathbf{Z}_t(k) + \mathbf{Z}_t(k)\mathbf{Z}_t(l) - 2\int_0^t \mathbf{Y}_s(k)\mathbf{Y}_s(l)ds + \int_0^t \sigma_{k,l}(\mathbf{I}_s)ds\\
&\quad + \frac{1}{\sqrt{a_n}}\left(h_{k,\mathbf{I}_t}\mathbf{Z}_t(l) + h_{l,\mathbf{I}_t}\mathbf{Z}_t(k) - \int_0^t h_{k,\mathbf{I}_s}\mathbf{Y}_s(l)ds - \int_0^t h_{l,\mathbf{I}_s}\mathbf{Y}_s(k)ds\right).
\end{aligned}$$

By integration by parts,

$$\mathbf{Y}_t(k)\mathbf{Z}_t(l) = \int_0^t \mathbf{Z}_s(l)dM_s(k) - \int_0^t \mathbf{Z}_s(l)\mathbf{Y}_s(k)ds + \int_0^t \mathbf{Y}_s(k)\mathbf{Y}_s(l)ds + \frac{1}{\sqrt{a_n}}\left(\int_0^t h_{k,\mathbf{I}_s}\mathbf{Y}_s(l)ds - h_{k,\mathbf{I}_t}\mathbf{Z}_s(l)\right),$$

hence

$$M_t(k)M_t(l) = N_t(k,l) + \int_0^t \mathbf{Z}_s(k)dM_s(l) + \int_0^t \mathbf{Z}_s(l)dM_s(k) + \int_0^t \sigma_{k,l}(\mathbf{I}_s)ds.$$

Finally, for any $1 \leq k,l \leq D$, the processes $M^{(n)}(k) - B^{(n)}(k)$ and $M^{(n)}(k)M^{(n)}(l) - A^{(n)}(k,l)$ are local martingales, with

$$A_t^{(n)}(k,l) = \int_0^t \sigma_{k,l}(\mathbf{I}_s^{(n)})ds, \quad B_t^{(n)}(k) = -\int_0^t \mathbf{Y}_s^{(n)}ds + \frac{1}{\sqrt{a_n}}h_{k,\mathbf{I}_t^{(n)}}.$$

Note that $\mathbf{I}^{(n)}$ is a Markov process on its own, generated by

$$\mathcal{L}_{\mathrm{I}}^{(n)}f(i) = a_n\sum_{j\neq i} q(i,j)[f(j) - f(i)].$$

In other words, for any $t > 0$, we can write $\mathbf{I}_t^{(n)} = \mathbf{I}_{a_n t}$ a.s., for some pure-jump Markov process $(\mathbf{I}_t)_{t\geq 0}$ generated by

$$\mathcal{L}_{\mathrm{I}}f(i) = \sum_{j\neq i} q(i,j)[f(j) - f(i)].$$

Using the ergodicity of $(\mathbf{I}_t)_{t\geq 0}$ together with $\lim_{n\to+\infty} a_n = +\infty$, we have

$$\lim_{n\to+\infty} A_t^{(n)}(k,l) = \lim_{n\to+\infty}\int_0^t \sigma_{k,l}(\mathbf{I}_{a_n s})ds = \lim_{n\to+\infty}\frac{1}{a_n}\int_0^{a_n t}\sigma_{k,l}(\mathbf{I}_u)du = t\sum_{i=1}^D \nu_i\sigma_{k,l}(i) = t\nu(\sigma_{k,l}).$$

Thus, the processes $\mathbf{Y}^{(n)}(k), B^{(n)}(k), A^{(n)}(k,l)$ satisfy the assumptions of [EK86, Chapter 7, Theorem 4.1], which entails Theorem 3.7. □

Remark 3.8 (Heuristics for a direct Taylor expansion of the generator)**.** As for many limit theorems for Markov processes, one would like to predict the convergence of the exponential zig-zag process to the Ornstein-Uhlenbeck diffusion from a Taylor expansion of the generator. Let us describe here a quick heuristic argument based on [CK06], which justifies the particular choice of functions φ_k in the proof of Theorem 3.7. For the sake of simplicitylet us work in the setting of Section 4.2, that is the generator of $(\mathbf{X}_t, \mathbf{I}_t)_{t\geq 0}$ is of the form

$$\mathcal{L}_{\mathrm{Z}}f(x,i) = g_i(x)\partial_x f(x,i) + a\theta_{3-i}[f(x,3-i) - f(x,i)]$$

where $g_i : x \mapsto (\mathbb{1}_{\{i=1\}} - x)$. For some smooth function $f : \mathbb{R}^D \to \mathbb{R}$, we have $\mathcal{L}_{\mathrm{Z}}f(x,i) = g_i(x)\cdot\nabla_x f(x)$ which cannot be rescaled to converge to some diffusive operator. We need an approximation f_a of f in a sense that $\lim_{a\to+\infty} f_a = f$ and $\mathcal{L}_{\mathrm{Z}}f_a$ has the form of a second order operator. Then, let

$$f_a : (x,i) \mapsto f(x) + a^{-1}h(x,i)\cdot\nabla_x f(x)$$

where $h(x,i)$ is the solution of the multidimensional Poisson equation associated to the transitions of the flows

$$\sum_{j\neq i} q(i,j)[h(x,j) - h(x,i)] = \theta_{3-i}[h(x,3-i) - h(x,i)] = \sum_j \nu_j g_j(x) - g_i(x) = \nu_1 - \mathbb{1}_{\{i=1\}}.$$

Then,

$$\mathcal{L}_{\mathrm{Z}} f_a(x,i) = \frac{1}{a} g_i(x) \cdot \nabla_x (h \cdot \nabla_x f)(x,i) + \nabla_x f(x) \sum_j \nu_j g_j(x).$$

Here, $\sum_j \nu_j g_j(x) - g_i(x) = \nu_1 - \mathbb{1}_{\{i=1\}}$ does not depend on x, neither does the function h, which is thus defined by (2.6). Furthermore, $h(x,i) = (\theta_1+\theta_2)^{-1}\mathbb{1}_{i=1}$. Moreover $\lim_{a\to+\infty} g_i(\nu + y/\sqrt{a}) = e_i - \nu$, so $\lim_{a\to+\infty} \mathcal{L}_{\mathrm{Z}} f_a(x,i) = \mathcal{L}_{\mathrm{O}} f(x)$ up to renormalization. ♢

From Proposition 3.1, for any fixed $n \geq 1$, the process $(\mathbf{X}_t^{(n)}, \mathbf{I}_t^{(n)})_{t\geq 0}$ admits and converges to a unique invariant distribution $\pi^{(n)}$, characterized in (3.7) as

$$\pi^{(n)} = \sum_{i=1}^{D} \nu_i \pi_i^{(n)} \otimes \delta_i, \quad \pi_i^{(n)}(dx) = \varphi^{(n)}(x,i)dx.$$

Let $\bar{\pi}^{(n)}$ be the first margin of the invariant measure of the Markov process $(\mathbf{Y}_t^{(n)}, \mathbf{I}_t^{(n)})_{t\geq 0}$, i.e. the probability distribution over $\mathbb{R}^D$ defined by

$$\bar{\pi}^{(n)}(dy) = \sum_{i=1}^{D} \frac{\nu_i}{\sqrt{a_n}} \varphi^{(n)}\left(\frac{y}{\sqrt{a_n}} + \nu, i\right) dy.$$

Corollary 3.9 (Convergence of the stationary distributions)**.** *The sequence of probability measures* $(\bar{\pi}^{(n)})_{n\geq 1}$ *converges to* $\mathcal{N}\left(0, \Sigma^{(0,1)}\right)$.

Proof of Corollary 3.9: Let $n \geq 1, t \geq 0$ and

$$\mathscr{F} = \left\{ f \in \mathscr{C}_c^2(\mathbb{R}^D) \ : \ \|f\|_\infty \leq 1, |f(x) - f(y)| \leq |x-y| \right\}.$$

Up to a constant, $d_{\mathscr{F}}$ is the Fortet-Mourier distance and metrizes the weak convergence. Fix $t \geq 0$ and let $\mathbf{X}_0^{(n)} = \nu$ and $\mathscr{L}(\mathbf{I}_0^{(n)}) = \nu^\top$. From Theorem 3.7,

$$\lim_{n\to+\infty} d_{\mathscr{F}}\left(\mathbf{Y}_t^{(n)}, \mathbf{Y}_t\right) = 0,$$

where $\mathbf{Y}$ is an Ornstein-Uhlenbeck process with generator $\mathcal{L}_{\mathrm{O}}$ and initial condition 0. Using the definition of $d_{\mathscr{F}}$ and Proposition 3.1,

$$d_{\mathscr{F}}\left(\mathbf{Y}_t^{(n)}, \bar{\pi}^{(n)}\right) \leq W\left((\mathbf{Y}_t^{(n)}, \mathbf{I}_t^{(n)}), \pi^{(n)}\right) \leq W\left(\delta_0 \otimes \nu, \pi^{(n)}\right) e^{-t} = W\left(\delta_0, \bar{\pi}^{(n)}\right) e^{-t}.$$

Let us check that the term $W\left(\delta_0, \bar{\pi}^{(n)}\right)$ is uniformly bounded. To that end, let

$$f^{(n)}(x,i) = x_k^2 + \frac{2}{a_n} h_{k,i} x_k + 2\nu_k x_k,$$

so that

$$\mathcal{L}^{(n)} f^{(n)}(x,i) = -2x_k^2 + \frac{2}{a_n}\left(\mathbb{1}_{\{i=k\}} - x_k\right) h_{k,i} + 2\nu_k \mathbb{1}_{\{i=k\}}.$$

Since $\pi^{(n)}\left(\mathcal{L}^{(n)} f^{(n)}\right) = 0$,

$$\int_E x_k^2 \pi^{(n)}(dx,di) - \nu_k^2 = \frac{1}{a_n}\left(h_{k,k}\nu_k - \int_E x_k h_{k,i} \pi^{(n)}(dx,di)\right).$$

Hence, with $C = \sum_{k=1}^D h_{k,k}\nu_k - \min_{i,j} h_{i,j}$, and since $\int_E x_k \pi^{(n)}(dx,di) = \nu_k$,

$$\begin{aligned}
\int_E \|x-\nu\|_2^2 \pi^{(n)}(dx,di) &= \sum_{k=1}^D \int_E (x_k - \nu_k)^2 \pi^{(n)}(dx,di) = \sum_{k=1}^D \int_E \left(x_k^2 - 2\nu_k x_k + \nu_k^2\right) \pi^{(n)}(dx,di) \\
&= \sum_{k=1}^D \int_E x_k^2 \pi^{(n)}(dx,di) - \nu_k^2 = \frac{1}{a_n}\left(\sum_{k=1}^D h_{k,k}\nu_k - \int_E x_k h_{k,i} \pi^{(n)}(dx,di)\right) \\
&\leq \frac{1}{a_n}\left(\sum_{k=1}^D h_{k,k}\nu_k - \min_{i,j} h_{i,j}\right) \leq \frac{C}{a_n}.
\end{aligned}$$

By Hölder's inequality,

$$W\left(\delta_0, \bar{\pi}^{(n)}\right) = \int_{\mathbb{R}} |y| \bar{\pi}^{(n)}(dy) = \int_E \sqrt{a_n}|x-\nu| \pi^{(n)}(dx, di) \leq \sqrt{C}.$$

Consequently to Proposition 3.6,

$$\begin{aligned} d_{\mathscr{F}}\left(\bar{\pi}^{(n)}, \mathscr{N}\left(0, \Sigma^{(0,1)}\right)\right) &\leq d_{\mathscr{F}}\left(\bar{\pi}^{(n)}, \mathbf{Y}_t^{(n)}\right) + d_{\mathscr{F}}\left(\mathbf{Y}_t^{(n)}, \mathbf{Y}_t\right) + d_{\mathscr{F}}\left(\mathbf{Y}_t, \mathscr{N}\left(0, \Sigma^{(0,1)}\right)\right) \\ &\leq 2\sqrt{C}e^{-t} + d_{\mathscr{F}}\left(\mathbf{Y}_t^{(n)}, \mathbf{Y}_t\right). \end{aligned}$$

Then,

$$\limsup_{t\to+\infty} d_{\mathscr{F}}\left(\bar{\pi}^{(n)}, \mathscr{N}\left(0, \Sigma^{(0,1)}\right)\right) \leq 2\sqrt{C}e^{-t},$$

which goes to 0 as $t \to +\infty$. □

4 Complete graph

In this section, we consider a particular case of freezing Markov chain, where all the states are connected, and the jump rate to a state does not depend on the position of the chain. This example of Markov chain has already been studied in the literature, for instance in [DS07]. Section 4.1 deals with the general D-dimensional case, for which most of the results of Section 3 can be written explicitly, notably the invariant measure of the exponential zig-zag process, which is a mixture of Dirichlet distributions (see Figure 4.1). Section 4.2 studies more deeply the case $D = 2$, where we can refine the speed of convergence provided in Proposition 3.1.

4.1 General case

Throughout this section, following [DS07], we assume that there exists a positive vector $\theta \in (0, +\infty)^D$ such that, for any $1 \leq i, j \leq D$,

$$q(i,j) = \theta_j - |\theta|\mathbb{1}_{i=j}, \quad |\theta| = \sum_{j=1}^{D} \theta_j, \tag{4.1}$$

and we will recover [DS07, Theorem 1.4]. If $D = 2$, let us highlight that an irreducible matrix $\mathrm{Id} + q$ automatically satisfies (4.1) (if $\mathrm{Id} + q$ is indecomposable then this is true as soon as $q(1,2)q(2,1) \neq 0$).

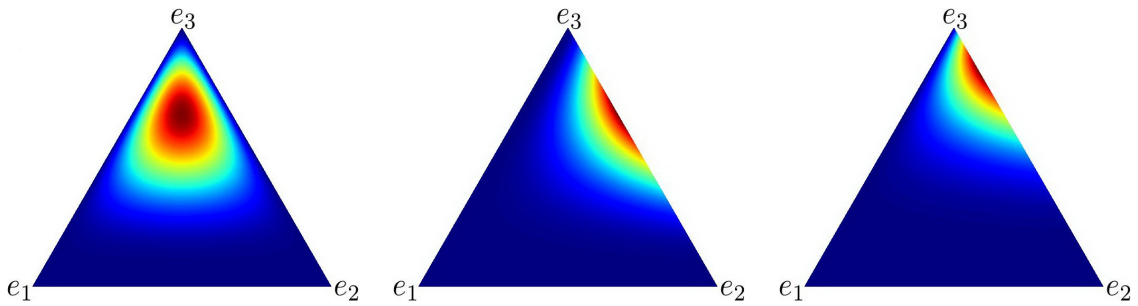


Figure 4.1: Probability density functions of $\pi_1 = \mathscr{D}(2,2,5), \pi_2 = \mathscr{D}(1,3,5), \pi_3 = \mathscr{D}(1,2,6)$, for $\theta_1 = 1, \theta_2 = 2, \theta_3 = 5$.

Proposition 4.1 (Limit distribution for the complete graph in the non-standard setting)**.** *Under Assumptions 2.1 and 2.7, and if q satisfies (4.1), then $\nu_i = \theta_i|\theta|^{-1}$ and*

$$\lim_{n\to+\infty}(x_n, i_n) = \sum_{i=1}^{D} \nu_i \mathscr{D}(a\theta + e_i) \otimes \delta_i \text{ in distribution.}$$

In particular,

$$\lim_{n\to+\infty} x_n = \mathscr{D}(a\theta) \text{ in distribution,} \qquad \lim_{n\to+\infty} i_n = \nu^{\top} \text{ in distribution.}$$

Proof of Proposition 4.1: If q satisfies (4.1), it is straightforward that its invariant distribution $\nu^{\top}$ is given by $\nu_i = \theta_i|\theta|^{-1}$ for any $1 \leq i \leq D$. The convergence of $(i_n)_{n\geq 1}$ to $\nu^{\top}$ and of $(x_n, i_n)_{n\geq 1}$ to some distribution π are direct corollaries of Theorems 2.4 and 2.9. Moreover, Proposition 3.4 holds, hence π satisfies (3.7) and it is clear that

$$\varphi(x,i) = \frac{\Gamma(|\theta|+1)}{\Gamma(\theta_i+1)\prod_{j\neq i}\Gamma(\theta_j)} x_i^{\theta_i} \prod_{j\neq i} x_j^{a\theta_j - 1} = \nu_i \frac{\Gamma(|\theta|)}{\prod_{j=1}^{D}\Gamma(\theta_j)} x_i^{\theta_i} \prod_{j\neq i} x_j^{a\theta_j - 1}$$

is the unique (up to a multiplicative constant) solution of (3.8), which entails that

$$\pi = \sum_{i=1}^{D} \nu_i \mathscr{D}(a\theta + e_i) \otimes \delta_i.$$

Finally, if $\mathscr{L}(X, I) = \pi$, it is clear that $\mathscr{L}(I) = \nu^{\top}$ and that

$$\mathscr{L}(X)(dx) = \sum_{i=1}^{D} \nu_i \varphi(x,i)dx = \frac{\Gamma(|\theta|)}{\prod_{j=1}^{D}\Gamma(\theta_j)} \prod_{j\neq i} x_j^{a\theta_j - 1} dx = \mathscr{D}(\theta)(dx).$$

□

In the framework of (4.1), it is also possible to obtain explicitly the solution of the Poisson equation related to q as well as the covariance matrix of the limit distribution in the standard setting. This is the purpose of the following result, whose proof is straightforward using Theorem 2.12 together with the expressions (1.2) and (2.6).

Proposition 4.2 (Limit distribution for the complete graph in the standard setting)**.** *Under Assumptions 2.1 and 2.8, and if q satisfies (4.1), then $\nu = |\theta|^{-1}\theta$ and $h_i = |\theta|^{-1}e_i$ and*

$$\lim_{n\to+\infty} y_n = \mathscr{N}\left(0, \Sigma^{(p,\Upsilon)}\right) \text{ in distribution,} \quad \text{with } \Sigma_{k,l}^{(p,\Upsilon)} = \begin{cases} -\frac{2-p}{1+\Upsilon}\nu_k\nu_l & \text{if } k \neq l \\ \frac{2-p}{1+\Upsilon}\nu_k(1-\nu_k) & \text{if } k = l \end{cases}.$$

Finally, let us emphasize the fact that Corollary 3.9 provides an interesting convergence of rescaled Dirichlet distributions, when considered in the particular case of the complete graph.

Corollary 4.3 (Convergence of the rescaled Dirichlet distribution to a Gaussian law)**.** *For any vector $\theta \in (0,+\infty)^D$, if $(X_n)_{n\geq 1}$ is a sequence of independent random variables such that $\mathscr{L}(X_n) = \mathscr{D}(a_n\theta)$, then*

$$\lim_{n\to+\infty} \sqrt{a_n}\,(X_n - \nu) = \mathscr{N}\left(0, \mathrm{diag}(\nu) - \nu\nu^{\top}\right) \text{ in distribution.}$$

4.2 The turnover algorithm

In this subsection, we consider the turnover algorithm introduced in [EV16]. This algorithm studies empirical frequency of *heads* when a coin is turned over with a certain probability, instead of being tossed

as usual. The authors provide various convergences in distribution for this proportion, depending on the asymptotic behavior of the turnover probability, which corresponds to $(p_n)_{n\geq 1}$ in the present paper. However, this turnover algorithm can be seen as a particular case of freezing Markov chain, and can then be written as the stochastic algorithm defined in (2.4), in the special case $D = 2$. Since $x_n(1) = 1-x_n(2)$, there is only one relevant variable in this section, which belongs to $[0, 1]$:

$$x_n = x_n(1) = \gamma_n \sum_{k=1}^{n} \mathbb{1}_{\{i_k=1\}}. \tag{4.2}$$

Note that we are in the framework of Section 4.1, with $\theta_1 = q(2, 1)$ and $\theta_2 = q(1, 2)$, and that Propositions 4.1 and 4.2 hold. In particular, we have $\nu_i = \theta_i(\theta_1 + \theta_2)^{-1}$. Then, for any $y \in \mathbb{R}$ and $(x, i) \in [0, 1] \times \{1, 2\}$, the infinitesimal generators defined in (1.1) and (1.3) write

$$\mathcal{L}_{\mathrm{O}} f(y) = -yf'(y) + \frac{2-p}{1+\Upsilon}\nu_1(1-\nu_1)f''(y) \tag{4.3}$$

and

$$\mathcal{L}_{\mathrm{Z}} f(x, i) = (\mathbb{1}_{\{i=1\}} - x)\partial_x f(x, i) + a\theta_{3-i}[f(x, 3-i) - f(x, i)]. \tag{4.4}$$

Remark 4.4 (Comparison with [EV16])**.** In the present paper, we recover [EV16, Theorems 1 and 2] as direct consequences of Theorems 2.9 and 2.12. The aforementioned results are extended by allowing $q(1, 2) \neq q(2, 1)$, but mostly by obtaining results for general sequences $(p_n)_{n\geq 1}$ while [EV16] deals only with $p_n = an^{-\theta}$ for positive constants a and θ. It should be noted that, in order to perfectly mimic the algorithm of the aforementioned article, one should consider the chain $x^\star_n = \gamma_n \sum_{k=1}^{n} \left(\mathbb{1}_{\{i_k=1\}} - \mathbb{1}_{\{i_k=2\}}\right)$, which evolves in $[-1, 1]$. The behavior of this sequence being completely similar to the one we are studying, we chose to work with (4.2) for the sake of consistence.

However, the reader should notice that the invariant measure of the process generated by (4.3) is a Gaussian distribution with variance $\Sigma^{(p,\Upsilon)}_{1,1}$. In the particular case where $p = 0$ and $\theta_1 = \theta_2$, this variance writes

$$\Sigma^{(0,\Upsilon)}_{1,1} = \frac{1}{2(1+\Upsilon)},$$

which is, at first glance, different from the variance provided in [EV16], which is (under our notation)

$$\sigma^2 = \frac{1}{a^2(1+\Upsilon)}.$$

The factor a^2 comes from the fact that [EV16] studies the behavior of $a^{-1}y_n$. The factor 2 comes from the choice of normalization mentioned earlier, since $x_n \in [0, 1]$ and $x^\star_n \in [-1, 1]$. ♢

Whenever $D = 2$, it is easier to visualize the dynamics of $(\mathbf{X}, \mathbf{I})$ (see Figure 4.2), and we can improve the results of Proposition 3.1 concerning the speed of convergence of the exponential zig-zag process to its stationary measure π.

Proposition 4.5 (Ergodicity when $D = 2$)**.** *The Markov process $(\mathbf{X}_t, \mathbf{I}_t)_{t\geq 0}$ generated by $\mathcal{L}_Z$ in (4.4), with values in $[0, 1] \times \{1, 2\}$, admits a unique stationary distribution*

$$\pi = \frac{\theta_1}{\theta_1+\theta_2}\beta(a\theta_1 + 1, a\theta_2) \otimes \delta_1 + \frac{\theta_2}{\theta_1+\theta_2}\beta(a\theta_1, a\theta_2 + 1) \otimes \delta_2.$$

Moreover, let $v = a(\theta_1 \vee \theta_2)$, then

$$W((\mathbf{X}_t, \mathbf{I}_t), \pi) \leq \begin{cases} \left(2 + \frac{2v}{|1-v|}\right) \mathrm{e}^{-(1\wedge v)t} & \text{if } v \neq 1 \\ (2+t)\,\mathrm{e}^{-t} & \text{if } v = 1 \\ W((\mathbf{X}_0, \mathbf{I}_0), \pi)\,\mathrm{e}^{-t} & \text{if } \mathscr{L}(\mathbf{I}_0) = \frac{\theta_1}{\theta_1+\theta_2}\delta_1 + \frac{\theta_2}{\theta_1+\theta_2}\delta_2 \end{cases}.$$

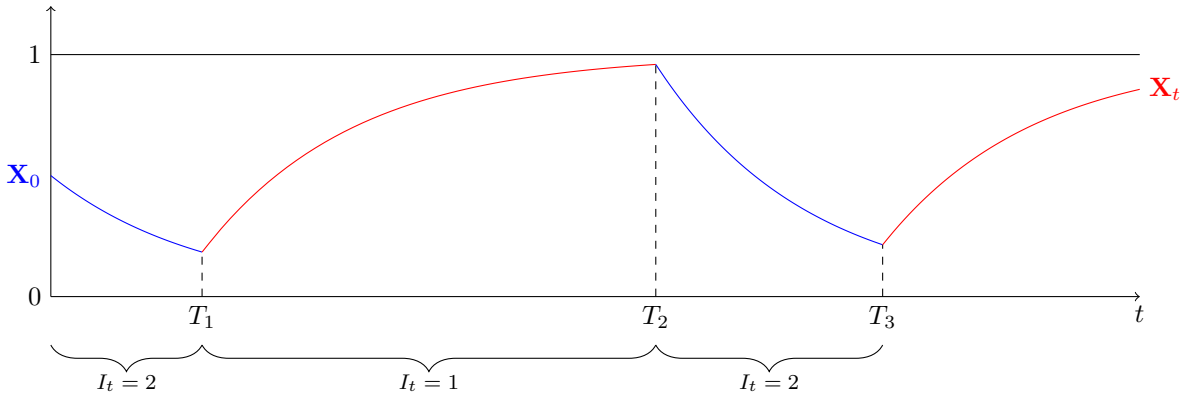


Figure 4.2: Typical path of the exponential zig-zag process when $D = 2$.

Since the inter-jump times of the exponential zig-zag process are spread-out, it is also possible to show convergence in total variation with a method similar to [BMP+15, Proposition 2.5]. Note that, following Proposition 4.1, the limit distribution of $(X_t)_{t\geq 0}$ is the first margin of π, namely $\beta(a\theta_1, a\theta_2)$.

Proof of Proposition 4.5: Without loss of generality, let us assume that $\theta_1 \geq \theta_2$, that is $v = a\theta_1$. Using Proposition 4.1, it is clear that π is the limit distribution of $(\mathbf{X}, \mathbf{I})$. Let us turn to the quantification of the ergodicity of the process. Since the flow is exponentially contracting at rate 1, one can expect the Wasserstein distance of the spatial component $\mathbf{X}$ to decrease exponentially. The only issue is to bring $\mathbf{I}$ to its stationary measure first. So, consider the Markov coupling $\Big((\mathbf{X}, \mathbf{I}), (\widetilde{\mathbf{X}}, \widetilde{\mathbf{I}})\Big)$ of $\mathcal{L}_Z$ on $E \times E$, which evolves independently if $\mathbf{I} \neq \widetilde{\mathbf{I}}$, and else follows the same flow with common jumps. We set $T_0 = 0$ and denote by T_n the epoch of its n^{th} jump. If $\mathbf{I}_0 \neq \widetilde{\mathbf{I}}_0$, the first jump is not common a.s., but in any case, since $D = 2$, $\mathbf{I}_{T_1} = \widetilde{\mathbf{I}}_{T_1}$ a.s. and $\mathscr{L}(T_1) = \mathscr{E}(v)$. Consequently,

$$\begin{aligned}
\mathbb{E}\left[|(\mathbf{X}_t, \mathbf{I}_t) - (\widetilde{\mathbf{X}}_t, \widetilde{\mathbf{I}}_t)|\right] &= \mathbb{E}\left[|\mathbf{X}_t - \widetilde{\mathbf{X}}_t|\right] + \mathbb{P}(\mathbf{I}_t \neq \widetilde{\mathbf{I}}_t) \\
&\leq \int_0^t \mathbb{E}\left[|\mathbf{X}_t - \widetilde{\mathbf{X}}_t| \,\middle|\, T_1 = s\right] v\,\mathrm{e}^{-vs}\,ds + \left(\mathbb{E}\left[|\mathbf{X}_t - \widetilde{\mathbf{X}}_t| \,\middle|\, T_1 > t\right] + 1\right)\mathbb{P}(T_1 > t) \\
&\leq 2\,\mathrm{e}^{-vt} + \int_0^t \mathbb{E}\left[|\mathbf{X}_s - \widetilde{\mathbf{X}}_s|\right]\mathrm{e}^{-(t-s)}\,v\,\mathrm{e}^{-vs}\,ds \\
&\leq 2\,\mathrm{e}^{-vt} + v\,\mathrm{e}^{-t}\int_0^t \mathrm{e}^{(1-v)s}\,ds \\
&\leq \left[\left(2 + \frac{v}{1-v}\right)\mathrm{e}^{-vt} - \frac{v}{1-v}\,\mathrm{e}^{-t}\right]\mathbb{1}_{\{v\neq 1\}} + (2+vt)\,\mathrm{e}^{-vt}\,\mathbb{1}_{\{v=1\}} \\
&\leq \left(2 + \frac{2v}{|1-v|}\right)\mathrm{e}^{-(1\wedge v)t}\,\mathbb{1}_{\{v\neq 1\}} + (2+t)\,\mathrm{e}^{-t}\,\mathbb{1}_{\{\lambda=1\}}.
\end{aligned}$$

Note that if $\mathscr{L}(\mathbf{I}_0) = \mathscr{L}(\widetilde{\mathbf{I}}_0)$, let $\mathbf{I}_0 = \widetilde{\mathbf{I}}_0$, so that the coupling $\Big((\mathbf{X}, \mathbf{I}), (\widetilde{\mathbf{X}}, \widetilde{\mathbf{I}})\Big)$ always has common jumps and

$$|\mathbf{X}_t - \widetilde{\mathbf{X}}_t| = |\mathbf{X}_0 - \widetilde{\mathbf{X}}_0|\,\mathrm{e}^{-t} \text{ a.s.}$$

Letting $(\mathbf{X}_0, \widetilde{\mathbf{X}}_0)$ be the optimal Wasserstein coupling entails Wasserstein contraction. The results above hold for any initial conditions $(\widetilde{\mathbf{X}}_0, \widetilde{\mathbf{I}}_0)$. Then, let $\mathscr{L}(\widetilde{\mathbf{X}}_0, \widetilde{\mathbf{I}}_0) = \pi$ to achieve the proof; in particular, $\mathscr{L}(\widetilde{\mathbf{I}}_0) = \nu^\top = \frac{\theta_1}{\theta_1+\theta_2}\delta_1 + \frac{\theta_2}{\theta_1+\theta_2}\delta_2$. □

5 Proofs

In this section, we provide the proofs of the main results of this paper that were stated throughout Section 2.

Proof of Theorem 2.4: Under Assumption 2.1, let us first assume that $p > 0$. The matrix $(\mathrm{Id} + q)$ is irreducible, and so is $(\mathrm{Id} + pq)$. Moreover, $\nu^\top$ is also the invariant measure of pq, and Perron-Frobenius Theorem entails that there exist $C > 0$ and $\rho \in (0,1)$ such that for every $n \geq 1$ and $i \in \{1, \dots, D\}$,

$$d_{\mathrm{TV}}\left(\delta_i(\mathrm{Id} + pq)^n, \nu^\top\right) \leq C\rho^n.$$

Now, let us prove that $(i_n)_{n\geq 1}$ is an asymptotic pseudotrajectory of the dynamical system induced by $\mathrm{Id} + pq$. The limit set of such a system being contained in every global attractor (see [Ben99, Theorems 6.9 and 6.10]), we have

$$\begin{aligned} d_{\mathrm{TV}}\left(\delta_{i_n}(\mathrm{Id} + pq), i_{n+1}\right) &= d_{\mathrm{TV}}\left(\delta_{i_n}(\mathrm{Id} + pq), \delta_{i_n}(\mathrm{Id} + q_n)\right) \\ &\leq |p_n - p| + \sum_{j \neq i_n} |r_n(i_n, j)| \leq |p_n - p| + \sum_{i,j=1}^{D} |r_n(i,j)|, \end{aligned} \tag{5.1}$$

and the right-hand side of (5.1) converges to 0, which ends the proof.

The case $p = 0$ is a mere application of [BBC16, Proposition 3.13]. □

5.1 Asymptotic pseudotrajectories in the non-standard setting

In this section, we prove Theorem 2.9 using results from [BBC16], based on the theory of asymptotic pseudotrajectories for inhomogeneous-time Markov chains. Indeed, with the convention $\sum_{k=1}^{0} = 0$, let

$$\tau_n = \sum_{k=1}^{n} \gamma_k, \quad m(t) = \sup\{k \geq 0 : \tau_k \leq t\}, \tag{5.2}$$

and define the piecewise-constant processes

$$X_t = \sum_{n=1}^{\infty} x_n \mathbb{1}_{\tau_n \leq t < \tau_{n+1}}, \quad I_t = \sum_{n=1}^{\infty} i_n \mathbb{1}_{\tau_n \leq t < \tau_{n+1}}. \tag{5.3}$$

We shall show that, as $t \to +\infty$, the process $(X_t, I_t)_{t\geq 0}$ converges in a way (see Figure 5.1) to the exponential zig-zag process $(\mathbf{X}_t, \mathbf{I}_t)_{t\geq 0}$ solution of (3.1), that we already studied in Section 3.1. To that end, let $(P_t)_{t\geq 0}$ be the Markov semigroup of $(\mathbf{X}, \mathbf{I})$, $N_1 = (2, \dots, 2, 0)$ and

$$\mathscr{F} = \left\{ f \in \mathcal{D}(\mathcal{L}_{\mathrm{Z}}) \cap \mathscr{C}_b^{N_1} : \mathcal{L}_{\mathrm{Z}} f \in \mathcal{D}(\mathcal{L}_{\mathrm{Z}}), \|\mathcal{L}_{\mathrm{Z}} f\|_\infty + \|\mathcal{L}_{\mathrm{Z}}\mathcal{L}_{\mathrm{Z}} f\|_\infty + \sum_{j=0}^{N_1} \|f^{(j)}\|_\infty \leq 1 \right\}. \tag{5.4}$$

Note that convergence with respect to $d_{\mathscr{F}}$ implies convergence in distribution (see [BBC16, Lemma 5.1]).

Lemma 5.1 (Asymptotic pseudotrajectory for non-standard fluctuations)**.** *Under the assumptions of Theorem 2.9, the sequence of probability distributions $(\mu_t)_{t\geq 0}$ is an asymptotic pseudotrajectory of $(P_t f)_{t\geq 0}$ with respect to $d_{\mathscr{F}}$.*

Moreover, if there exist positive constants $A \geq 1, \theta \leq 1$ such that

$$\max_{j \neq i}(|r_n(i,j)|) \leq \frac{A}{n^\theta},$$

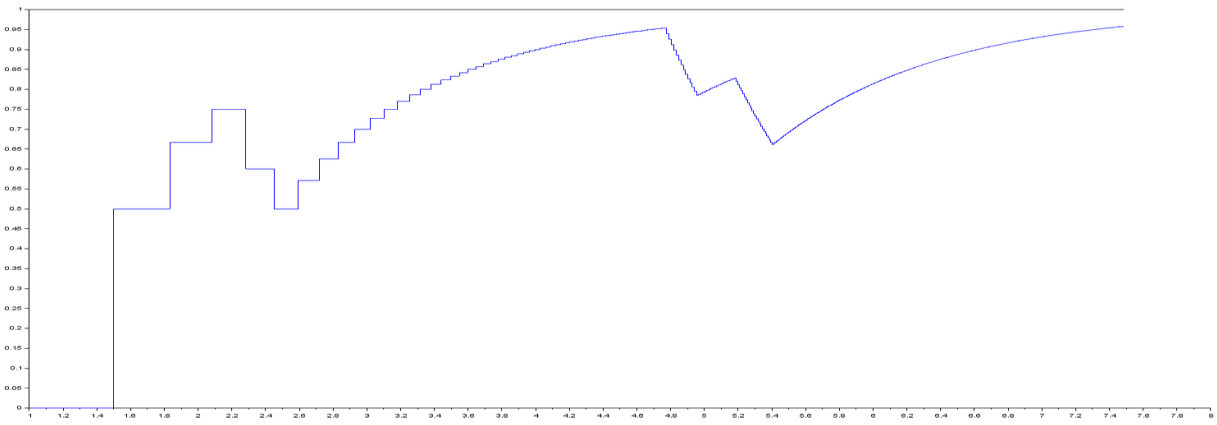


Figure 5.1: Sample path of the process $(X_t)_{t\geq 0}$ in the setting of Section 4.2 for $a = 1$, $q(1,2) = \frac{1}{3}$ and $q(2,1) = \frac{2}{3}$.

then, for any $v < a\rho(1+a\rho)^{-1}$*, there exists a positive constant* C *such that*

$$d_{\mathscr{F}\cap\mathscr{G}}(\mathscr{L}(X_t, I_t), \pi) \leq C\,\mathrm{e}^{-vt}\,. \tag{5.5}$$

Moreover, the sequence of processes $((X_{n+t}, I_{n+t})_{t\geq 0})_{n\geq 1}$ *converges in distribution, as* $t \to +\infty$*, toward* $(\mathbf{X}^\pi_t, \mathbf{I}^\pi_t)_{t\geq 0}$ *in the Skorokhod space, where* $(\mathbf{X}^\pi, \mathbf{I}^\pi)$ *is a process generated by* $\mathcal{L}_Z$ *with initial condition* π*.*

The proof of Lemma 5.1 consists in checking [BBC16, Assumptions 2.1, 2.2, 2.3 and 2.7.ii)] and relies on three ingredients:

- Convergence of a kind of discrete infinitesimal generator $\mathcal{L}_n$, which characterizes the dynamics of (X, I), to $\mathcal{L}_{\mathrm{Z}}$ defined in (1.3).
- Smoothness of the limit semigroup $(P_t)_{t\geq 0}$ and control of its derivatives with respect to the initial condition of the process.
- Uniform boundedness of the moments of $(x_n, i_n)_{n\geq 1}$ up to some order, which is trivially satisfied here since E is compact.

Proof of Lemma 5.1: In what follows, the notation O (as $n \to +\infty$) is uniform over x, i, f. We define $\mathcal{L}_n f(x,i) = \gamma_{n+1}^{-1}\mathbb{E}[f(x_{n+1}, i_{n+1})|x_n = x, i_n = i]$, and we study the convergence of $\mathcal{L}_n$ to $\mathcal{L}_{\mathrm{Z}}$ in the sense of [BBC16]. Let $(x,i) \in E$ and $\chi_{N_1}(x,i) = \prod_{k=1}^{D} x_k^2$. We recall that $q_n(i,i) = 1 + O(p_n)$ as $n \to +\infty$. With

$$\epsilon_n = \frac{1}{n} \vee \max_{j\neq i}(|r_n(i,j)|),$$

we have

$$\begin{aligned}\mathcal{L}_n f(x,i) &= \sum_{j\neq i} \frac{q_{n+1}(i,j)}{\gamma_{n+1}} \left[f\left(\frac{\gamma_{n+1}}{\gamma_n} x + \gamma_{n+1} e_j, j\right) - f(x,i)\right] \\ &\quad + \frac{1-\sum_{j\neq i} q_{n+1}(i,j)}{\gamma_{n+1}} \left[f\left(\frac{\gamma_{n+1}}{\gamma_n} x + \gamma_{n+1} e_i, i\right) - f(x,i)\right] \\ &= \sum_{j\neq i} \frac{p_{n+1}}{\gamma_{n+1}} (q(i,j) + r_{n+1}(i,j)) \left[f(x,j) - f(x,i) + \chi_{(1,\dots,1,0)}(x,i) \|f^{(1,\dots,1,0)}\|_\infty O\left(\gamma_{n+1}\right)\right] \\ &\quad + \frac{1+O(p_{n+1})}{\gamma_{n+1}} \left[\left(\left(\frac{\gamma_{n+1}}{\gamma_n} - 1 \right) + \gamma_{n+1} e_i \right) \cdot \nabla_x f(x) + \chi_{N_1}(x,i) \|f^{(N_1)}\|_\infty O\left(\gamma_{n+1}^2\right)\right] \\ &= \mathcal{L}_Z f(x,i) + \chi_{N_1}(x,i) \|f^{(N_1)}\|_\infty O\left(\epsilon_{n+1}\right).\end{aligned}$$

We turn to the study of the regularity of the limit semigroup, following [Kun84]. Let $t>0$ and note that $\|P_t f\|_\infty \leq \|f\|_\infty$. Moreover, the process $(\mathbf{X},\mathbf{I})$ is solution of the following SDE (we emphasize below the dependence on the initial condition):

$$(\mathbf{X}_t^{x,i}, \mathbf{I}_t^{x,i}) = (x,i) + \int_0^t \left(A(\mathbf{X}_{s^-}^{x,i}, \mathbf{I}_{s^-}^{x,i}) + e_{\mathbf{I}_{s^-}^{x,i}} \right) ds + \sum_{j=1}^D \int_0^t B_{\mathbf{I}_{s^-}^{x,i},j}(\mathbf{X}_{s^-}^{x,i}, \mathbf{I}_{s^-}^{x,i}) N_{\mathbf{I}_{s^-}^{x,i},j}(ds), \tag{5.6}$$

where $N_{i,j}$ is a Poisson process of intensity $aq(i,j)\mathbb{1}_{\{i\neq j\}}$ and the matrices A and $B_{i,j}$ are defined in (3.2). Then, if we denote by $\eta_t^{x,i,k,h} = h^{-1}\left[(\mathbf{X}_t^{x+he_k,i}, \mathbf{I}_t^{x+he_k,i}) - (\mathbf{X}_t^{x,i}, \mathbf{I}_t^{x,i})\right]$, we recover from (5.6) that the process $\eta^{x,i,k,h}$ satisfies the ODE

$$\eta_t^{x,i,k,h} = (e_k, 0) + \int_0^t A\eta_{s^-}^{x,i,k,h} ds,$$

so that $\eta_t^{x,i,k,h} = (\mathrm{e}^{-t} e_k, 0)$. Thus, $\eta^{x,i,k,h}$ admits a continuous modification (notably at $h=0$) and $\partial_k(\mathbf{X}^{x,i}, \mathbf{I}^{x,i}) = (\mathrm{e}^{-t} e_k, 0)$ is continuous. Using similar arguments, $\partial_k\partial_l(\mathbf{X}^{x,i}, \mathbf{I}^{x,i}) = 0$. Gathering those expressions, and since f^N is bounded for every multi-index $N \leq N_1$, it is clear that $P_t f \in \mathscr{C}^{N_1}$, with, for any $j,k \leq D$,

$$\begin{aligned}\partial_j (P_t f)(x,i) &= \mathbb{E}\left[\partial_j f(\mathbf{X}_t^{x,i}, \mathbf{I}_t^{x,i})\right] \mathrm{e}^{-t}, \\ \partial_j\partial_k (P_t f)(x,i) &= \mathbb{E}\left[\partial_j\partial_k f(\mathbf{X}_t^{x,i}, \mathbf{I}_t^{x,i})\right] \mathrm{e}^{-2t}.\end{aligned}$$

Hence, for any $j \leq N_1$, $\|(P_t f)^{(j)}\|_\infty \leq \|f^{(j)}\|_\infty$. Finally, for any $n \geq 1, |x_n| \leq 1$, so that

$$\sup_{n\geq 1} \chi_{N_1}(x_n, i_n) = \sup_{n\geq 1} \sum_{k=0}^2 |x_n|^k \leq 3.$$

Hence, we can apply [BBC16, Theorems 2.6 and 2.8.ii)] with $N_1 = N_2 = d_1 = d = (2,\dots,2,0), C_T = 1, M_2 = 3$, to obtain the existence and the announced properties of π as well as

$$\lim_{n\to+\infty} d_{\mathscr{F}}\left(\mathscr{L}(x_n,i_n), \pi\right) = 0.$$

Moreover, following [BBC16, Remark 2.5],

$$\lambda(\gamma,\epsilon) = -\limsup_{n\to+\infty} \frac{\log(\gamma \vee \epsilon_n)}{\sum_{k=1}^n \gamma_k} \leq \frac{\theta}{A}.$$

Finally, using Proposition 3.1 together with [BBC16, Theorem 2.8.ii)] entails (5.5). Recall the compactness of E, then we can apply [BBC16, Theorem 2.12] and achieve the proof. □

5.2 ODE and SDE methods in the standard setting

In the present section, we successively provide proofs for Theorems 2.11 and 2.12. We shall prove the former with a method involving an asymptotic pseudotrajectory for some interpolated process, similarly to Section 5.1 and [BC15]. On the contrary, the fluctuations obtained for $(x_n)_{n\geq 1}$ in Theorem 2.12 are obtained through a more classic result for stochastic algorithms, namely the SDE method developed in [Duf96] (see also [KY03]).

Proof of Theorem 2.11: In the following, we mimic the proof of [BC15, Lemma 2.4] (see also [MP87, Ben97]). Indeed, for any $n \geq 1$, (2.4) writes

$$x_{n+1} = x_n + \gamma_{n+1}(\nu - x_n) + \gamma_{n+1}(e_{i_{n+1}} - \nu).$$

Let the sequence $(\tau_n)_{n\geq 0}$ and the function m be as in (5.2), and define the interpolated process

$$\widehat{X}_{\tau_n+s}0 = x_n + s\frac{x_{n+1} - x_n}{\tau_{n+1} - \tau_n},$$

for all $s \in [0, \gamma_{n+1})$ and $n \geq 0$. We will show that $\widehat{X}$ is an asymptotic pseudotrajectory (and a $\ell-$pseudotrajectory) for the flow $\Phi(x,t) = \nu + e^{-t}(x-\nu)$ associated to the ODE $\partial_t\Phi(x,t) = \nu - \Phi(x,t)$. From [Ben99, Proposition 4.1] it suffices to show that, for all $T > 0$,

$$\lim_{t\to+\infty} \Delta(t,T) = 0, \quad \text{with } \Delta(t,T) = \sup_{0\leq h\leq T} \left| \sum_{k=m(t)}^{m(t+h)} \gamma_k(e_{i_{k+1}} - \nu) \right|, \tag{5.7}$$

and

$$\lim_{t\to+\infty} \frac{\log(\Delta(t,T))}{t} \leq -\ell. \tag{5.8}$$

Consider h defined in (2.6). Then,

$$\begin{aligned}
\gamma_{n+1}\left(e_{i_{n+1}} - \nu\right) &= \gamma_{n+1}\sum_{j=1}^{D} q(i_{n+1},j)\left(h_{i_{n+1}} - h_j\right) \\
&= \frac{\gamma_{n+1}}{p_n}\sum_{j=1}^{D}\left(p_n q(i_n,j) - q_n(i_n,j)\right)h_{i_{n+1}} + \frac{\gamma_{n+1}}{p_n}\left(h_{i_{n+1}} - \mathbb{E}\left[h_{i_{n+1}}|i_n\right]\right) \\
&\quad + \frac{\gamma_{n+1}}{p_n}\sum_{j=1}^{D}\left(q_n(i_n,j) - p_n q(i_n,j)\right)h_j \\
&\quad + \left(\gamma_{n+1}\sum_{j=1}^{D} q(i_n,j)h_j - \gamma_n\sum_{j=1}^{D} q(i_n,j)h_j\right) \\
&\quad + \left(\gamma_n\sum_{j=1}^{D} q(i_n,j)h_j - \gamma_{n+1}\sum_{j=1}^{D} q(i_{n+1},j)h_j\right).
\end{aligned} \tag{5.9}$$

We shall bound each term of the sum (5.9) separately. We easily have

$$\left|\gamma_{n+1}\sum_{j=1}^{D}\left(q_n(i_n,j) - p_n q(i_n,j)\right)\frac{h_j}{p_n}\right| = \left|\gamma_{n+1}\sum_{j=1}^{D} r_n(i_n,j)h_j\right| \leq \|h\|_1 R_n\gamma_{n+1}$$

and

$$\left|\gamma_{n+1}\sum_{j=1}^{D}\left(p_n q(i_n,j) - q_n(i_n,j)\right)h_{i_{n+1}}\right| = \left|\gamma_{n+1}h_{i_{n+1}}\sum_{j=1}^{D} r_n(i_{n+1},j)\right| \leq \|h\|_1 R_n\gamma_{n+1},$$

where $\|h\|_1 = \sup_j \sum_i |h_{i,j}|$ and $R_n = \sup_i \sum_j |r_n(i,j)|$. Also, for some constant $C > 0$,

$$\left| \gamma_{n+1} \sum_{j=1}^{D} q(i_n, j) h_j - \gamma_n \sum_{j=1}^{D} q(i_n, j) h_j \right| \leq C(\gamma_n - \gamma_{n+1}).$$

Note that $(\gamma_n \sum_{j=1}^{D} q(i_n, j) h_j - \gamma_{n+1} \sum_{j=1}^{D} q(i_{n+1}, j) h_j)$ is the main term of a telescoping series. It remains to bound the norm of the sum of $\gamma_{n+1} p_n^{-1} \left(h_{i_{n+1}} - \mathbb{E}[h_{i_{n+1}} | i_{n+1}] \right)$. For all $m, n \geq 1$ and $l = 1, ..., D$, set

$$M_{m,n}(l) = \sum_{k=m}^{n} \frac{\gamma_{k+1}}{p_k} \left(h_{l,i_{k+1}} - \mathbb{E}\left[h_{l,i_{k+1}} | i_{k+1} \right] \right).$$

The sequence $(M_{m,n}(l))_{m \geq n}$ is a martingale and

$$\mathbb{E}\left[M_{m,n}(l) M_{m,n}(c) \right] = \sum_{k=m}^{n} \frac{\gamma_{k+1}^2}{p_k^2} \mathbb{E}\left[\sum_{j=1}^{D} q_k(i_k, j) (h_{l,j} - \mathbb{E}[h_{l,i_{k+1}} | i_k]) (h_{c,j} - \mathbb{E}[h_{c,i_{k+1}} | i_k]) \right].$$

Moreover, as

$$\begin{aligned}
&\mathbb{E}\left[\sum_{j=1}^{D} q_k(i_k, j) (h_{l,j} - \mathbb{E}[h_{l,i_{k+1}} | i_k]) (h_{c,j} - \mathbb{E}[h_{c,i_{k+1}} | i_k]) \right] \\
=&\mathbb{E}\left[\sum_{j=1}^{D} q_k(i_k, j) (h_{l,j} h_{c,j} - \mathbb{E}[h_{l,i_{k+1}} | i_k] \mathbb{E}[h_{c,i_{k+1}} | i_k]) \right] \\
=&p_k \mathbb{E}\left[\sum_{j \neq i_k} q(i_k, j) h_{l,j} h_{c,j} \right] + \mathbb{E}\left[\left(1 - p_k \sum_{j \neq i_k} q(i_k, j) \right) h_{l,i_k} h_{c,i_k} \right] \\
&- \mathbb{E}\left[\left(\sum_{j=1}^{D} q_k(i_k, j) h_{l,j} \right) \left(\sum_{j=1}^{D} q_k(i_k, j) h_{c,j} \right) \right] + o(p_k) \\
=&p_k \mathbb{E}\left[\sum_{j \neq i_k} q(i_k, j) (h_{l,j} - h_{l,i_k}) (h_{c,j} - h_{c,i_k}) \right] \\
&+ \mathbb{E}\left[p_k^2 \left(\sum_{j \neq i_k} q(i_k, j) (h_{l,i_k} - h_{l,j}) \right) \left(\sum_{j \neq i_k} q(i_k, j) (h_{c,j} - h_{c,i_k}) \right) \right] + o(p_k),
\end{aligned}$$

by Theorem 2.11, we obtain

$$\begin{aligned}
&\mathbb{E}\left[\sum_{j=1}^{D} q_k(i_k, j) (h_{l,j} - \mathbb{E}[h_{l,i_{k+1}} | i_k]) (h_{c,j} - \mathbb{E}[h_{c,i_{k+1}} | i_k]) \right] \\
&= p_k \sum_{i=1}^{D} \nu_i \left[\sum_{j=1}^{D} q(i, j) (h_{l,j} - h_{l,i}) (h_{c,j} - h_{c,i)}) \right] \\
&\quad - p_k^2 \sum_{i=1}^{D} \nu_i \left(\nu_l - \mathbb{1}_{i=l} \right) \left(\nu_c - \mathbb{1}_{i=c} \right) + o(p_k).
\end{aligned} \tag{5.10}$$

As a consequence of (5.10), there exists some constant $C > 0$ such that

$$\sum_{l=1}^{D} \mathbb{E}\left[M_{m,n}(l)^2 \right] \leq C \sum_{k=m}^{n} \frac{\gamma_{k+1}^2}{p_k}.$$

By Doob's inequality and Assumption 2.8, it follows that, for every $k \geq 0$,

$$\mathbb{E}\left[\sup_{0\leq h\leq T}|M_{m(kT),m(kT+h)}|\right] \leq 2\sum_{l=1}^{D}\mathbb{E}\left[|M_{m(kT),m((k+1)T)}(l)|^2\right] \leq 2C\sum_{j=m(kT)}^{m((k+1)T)}\gamma_{j+1}\frac{\gamma_{j+1}}{p_j}$$
$$\leq 2CT\sup_{j\geq m(kT)}\frac{\gamma_{j+1}}{p_j},$$

which implies that $\lim_{k\to+\infty}\sup_{0\leq h\leq T}|M_{m(kT),m(kT+h)}| = 0$ and then $\lim_{k\to+\infty}\Delta(kT,T) = 0$ in probability. By the triangle inequality and [Ben99, Proposition 4.1], (5.7) holds.

Under the assumption that $\sum_{n=1}^{\infty}\gamma_{n+1}^2 p_n^{-1} < +\infty$,

$$\mathbb{E}\left[\sum_{k\geq 0}\sup_{0\leq h\leq T}|M_{m(kT),m(kT+h)}|\right] \leq \sum_{k\geq 0}2\sum_{l=1}^{D}\mathbb{E}\left[|M_{m(kT),m((k+1)T)}(l)|^2\right] \leq 2C\sum_{k\geq m(T)}\frac{\gamma_{k+1}^2}{p_k} < +\infty,$$

which implies $\lim_{k\to+\infty}\sup_{0\leq h\leq T}|M_{m(kT),m(kT+h)}| = 0$ a.s. Then, $\lim_{k\to+\infty}\Delta(kT,T) = 0$ a.s. and $\lim_{t\to+\infty}\Delta(t,T) = 0$ since

$$\Delta(t,T) \leq 2\Delta(\lfloor t/T\rfloor T,T) + \Delta((\lfloor t/T\rfloor+1)T,T).$$

In order to obtain a ℓ-pseudotrajectory, use Markov's and Doob's inequalities so that

$$\mathbb{P}\left(\sup_{0\leq h\leq T}|M_{m(kT),m(kT+h)}| \geq e^{-kT\alpha}\right) \leq e^{kT\alpha}2C\sum_{j=m(kT)}^{m((k+1)T)}\gamma_{j+1}\frac{\gamma_{j+1}}{p_j} \leq 2C(T+1)e^{kT\alpha}\sup_{j\geq m(kT)}\frac{\gamma_{j+1}}{p_j}.$$

Now, for all $\varepsilon > 0$ and k large enough,

$$\sup_{j\geq m(kT)}\frac{\gamma_{j+1}}{p_j} \leq \exp\left(-(\lambda(\gamma,\gamma/p)) - \varepsilon)\tau_{m(kT)}\right) \leq \exp\left(-(\lambda(\gamma,\gamma/p)) - \varepsilon)kT\right),$$

where $\lambda(\gamma,\gamma/p)$ is defined in (2.5). Hence,

$$\mathbb{P}\left(\sup_{0\leq h\leq T}|M_{m(kT),m(kT+h)}| \geq e^{-kT\alpha}\right) \leq 2C(T+1)\exp\left(kT(\alpha - \lambda(\gamma,\gamma/p) + \varepsilon)\right),$$

and by the Borel-Cantelli lemma, we have

$$\limsup_{k\to+\infty}\frac{1}{k}\sup_{0\leq h\leq T}|M_{m(kT),m(kT+h)}| \leq -\lambda(\gamma,\gamma/p) \text{ a.s.}$$

Then, bounding all the other terms of (5.9), we find

$$\lim_{t\to+\infty}\frac{\log(\Delta(t,T)}{t} \leq -\ell$$

with

$$\ell = \min\Big(\lambda(\gamma,\gamma/p),\lambda(\gamma,\gamma),\lambda(\gamma,R)\Big) = \lambda(\gamma,\gamma/p)\wedge\lambda(\gamma,R).$$

Since the flow Φ converges to ν exponentially fast at rate 1, use [Ben99, Theorem 6.9 and Lemma 8.7] to achieve the proof. □

Proof of Theorem 2.12: We have

$$y_{n+1} = y_n + y_n\left(\frac{\alpha_{n+1}}{\alpha_n}(1-\gamma_{n+1}) - 1\right) + \gamma_{n+1}\alpha_{n+1}(e_{i_{n+1}} - \nu).$$

Recall (5.9), so that

$$\gamma_{n+1}(e_{i_{n+1}} - \nu) = \frac{\gamma_{n+1}}{p_n}(h_{i_{n+1}} - \mathbb{E}[h_{i_{n+1}}|i_n]) + b_n,$$

with a remainder term b_n converging to 0. Now, we want to use [Duf96, Théorème 4.II.4]. In our setting, its notation reads

$$y_{n+1} = y_n + \widehat{\gamma}_n \left(h(y_n) + \widehat{r}_{n+1}\right) + \sqrt{\widehat{\gamma}_n}\epsilon_{n+1},$$

with

$$\epsilon_{n+1} = \left(\frac{1+\Upsilon}{2}\right)^{-1/2} \frac{1}{\sqrt{p_n}}(h_{i_{n+1}} - \mathbb{E}[h_{i_{n+1}}|i_n]), \quad \widehat{\gamma}_{n+1} = \left(\frac{1+\Upsilon}{2}\right)\frac{\gamma_{n+1}^2\alpha_{n+1}^2}{p_n}, \quad h : z \mapsto -z,$$

and

$$\begin{aligned}\widehat{r}_{n+1} = y_n \frac{1}{\widehat{\gamma}_{n+1}}&\left(\frac{\alpha_{n+1}}{\alpha_n}(1-\gamma_{n+1}) - 1 + \widehat{\gamma}_{n+1}\left(\frac{1+\Upsilon}{2}\right)\right)\\ &+ \frac{\alpha_{n+1}}{\widehat{\gamma}_{n+1}}\frac{\gamma_{n+1}}{p_n}\sum_{j=1}^{D} r_n(i_n,j)(h_{i_{n+1}} - h_j)\\ &+ \frac{\alpha_{n+1}}{\widehat{\gamma}_{n+1}}\gamma_{n+1}\left(\sum_{j=1}^{D} q(i_n,j)h_j - \sum_{j=1}^{D} q(i_{n+1},j)h_j\right).\end{aligned}$$

Then, by (5.10) and similar computations,

$$\mathbb{E}[\epsilon_n|\mathcal{F}_n] = 0 \quad \mathbb{E}[\epsilon_n\epsilon_n^\top|\mathcal{F}_n] = \Sigma^{(p,\Upsilon)} + o(p_n), \quad \sup_{n\geq 1}\mathbb{E}\left[|\epsilon_n|^q\right] < +\infty,\ q > 2,$$

where $\Sigma^{(p,\Upsilon)}$ is defined in (1.2). Classically, we should prove that $\lim_{n\to+\infty}\|\widehat{r}_n\| = 0$, in order to work in the framework of [Duf96, Hypothèse H4-4], which is quite difficult. Nevertheless, rather than checking that $\lim_{n\to+\infty}\|\widehat{r}_n\| = 0$ it is sufficient[2] to prove that

$$\widehat{r}_n = \widehat{r}_n^{(1)} + \widehat{r}_n^{2}, \quad \lim_{n\to+\infty}\widehat{r}_n^{(1)} = 0, \quad \lim_{n\to+\infty}\mathbb{E}\left[\sup_{0\leq s\leq T}\left|\sum_{n=m(t)}^{m(t+s)}\widehat{\gamma}_{n+1}\widehat{r}_{n+1}^{(2)}\right|\right] = 0, \tag{5.11}$$

for any $T > 0$, where $m(t)$ is defined in (5.2). Then, let

$$\begin{aligned}\widehat{r}_{n+1}^{(1)} &= y_n \frac{1}{\widehat{\gamma}_{n+1}}\left(\frac{\alpha_{n+1}}{\alpha_n}(1-\gamma_{n+1}) - 1 + \widehat{\gamma}_{n+1}\left(\frac{1+\Upsilon}{2}\right)\right) + \frac{\alpha_{n+1}}{\widehat{\gamma}_{n+1}}\frac{\gamma_{n+1}}{p_n}\sum_{j=1}^{D} r_n(i_n,j)(h_{i_{n+1}} - h_j),\\ \widehat{r}_{n+1}^{(2)} &= \frac{\alpha_{n+1}}{\widehat{\gamma}_{n+1}}\gamma_{n+1}\left(\sum_{j=1}^{D} q(i_n,j)h_j - \sum_{j=1}^{D} q(i_{n+1},j)h_j\right).\end{aligned}$$

The sequence $(\widehat{r}_n^{(1)})_{n\geq 1}$ goes to 0 a.s. and in L^1 straightforwardly under our assumptions. Furthermore

$$\begin{aligned}\widehat{\gamma}_{n+1}\widehat{r}_{n+1}^{(2)} &= \alpha_{n+1}\gamma_{n+1}\sum_{j=1}^{D} q(i_n,j)h_j - \alpha_{n+2}\gamma_{n+2}\sum_{j=1}^{D} q(i_{n+1},j)h_j\\ &\quad + (\alpha_{n+2}\gamma_{n+2} - \alpha_{n+1}\gamma_{n+1})\sum_{j=1}^{D} q(i_{n+1},j)h_j.\end{aligned} \tag{5.12}$$

[2]This assertion can be easily checked at the end of [Duf96, p.156], whose proof is based on usual arguments on diffusion approximation, such as [EK86]. The decomposition (5.11) is often assumed in more recent generalizations, see for instance [For15]. Note that we cannot use directly [For15], which besides does not provide functional convergence.

The first line of (5.12) is a telescoping series and is bounded by $\alpha_n\gamma_{n+1}$ which goes to 0. The second line of (5.12) is bounded by,

$$C\sum_{n=m(t)}^{m(t+T)}\left|\alpha_{n+2}\gamma_{n+1}-\alpha_{n+1}\gamma_n\right|, \tag{5.13}$$

for some $C>0$. Since (5.12) is a telescoping series as well, and goes to 0, we established the announced decomposition (5.11). As a conclusion, the diffusive limit $(\mathbf{Y}_t)_{t\geq 0}$ is the solution of (3.10), which trivially admits $V:z\mapsto z$ as a Lyapunov function, as required in [Duf96, Hypothèse H4-3]. The only use of an assumption on the eigenelements of $\Sigma^{(p,\Upsilon)}$ would be to guaranty the existence, uniqueness of and convergence to an invariant distribution for $\mathbf{Y}$, which was already proved in Proposition 3.6. □

Acknowledgements: Both authors acknowledge financial support from the ANR PIECE (ANR-12-JS01-0006-01) and the Chaire Modélisation Mathématique et Biodiversité.

References

[BBC16] M. Benaïm, F. Bouguet, and B. Cloez. Ergodicity of inhomogeneous Markov chains through asymptotic pseudotrajectories. *ArXiv e-prints*, January 2016. 3, 4, 5, 6, 8, 22, 23, 24

[BC15] M. Benaïm and B. Cloez. A stochastic approximation approach to quasi-stationary distributions on finite spaces. *Electron. Commun. Probab.*, 20:no. 37, 14, 2015. 6, 25

[BD16] J. Bierkens and A. Duncan. Limit theorems for the Zig-Zag process. *ArXiv e-prints*, July 2016. 14

[Ben97] M. Benaïm. Vertex-reinforced random walks and a conjecture of Pemantle. *Ann. Probab.*, 25(1):361–392, 1997. 25

[Ben99] M. Benaïm. Dynamics of stochastic approximation algorithms. In *Séminaire de Probabilités, XXXIII*, volume 1709 of *Lecture Notes in Math.*, pages 1–68. Springer, Berlin, 1999. 3, 6, 22, 25, 27

[BH96] M. Benaïm and M. W. Hirsch. Asymptotic pseudotrajectories and chain recurrent flows, with applications. *J. Dynam. Differential Equations*, 8(1):141–176, 1996. 3

[BH12] Y. Bakhtin and T. Hurth. Invariant densities for dynamical systems with random switching. *Nonlinearity*, 25(10):2937–2952, 2012. 11, 12

[BLBMZ12] M. Benaïm, S. Le Borgne, F. Malrieu, and P.-A. Zitt. Quantitative ergodicity for some switched dynamical systems. *Electron. Commun. Probab.*, 17:no. 56, 14, 2012. 9, 10

[BLBMZ15] M. Benaïm, S. Le Borgne, F. Malrieu, and P.-A. Zitt. Qualitative properties of certain piecewise deterministic Markov processes. *Ann. Inst. Henri Poincaré Probab. Stat.*, 51(3):1040–1075, 2015. 9, 11

[BMP+15] F. Bouguet, F. Malrieu, F. Panloup, C. Poquet, and J. Reygner. Long time behavior of Markov processes and beyond. *ESAIM Proc. Surv.*, 51:193–211, 2015. 11, 21

[CH15] B. Cloez and M. Hairer. Exponential ergodicity for Markov processes with random switching. *Bernoulli*, 21(1):505–536, 2015. 9

[CK06] C. Costantini and T. G. Kurtz. Diffusion approximation for transport processes with general reflection boundary conditions. *Math. Models Methods Appl. Sci.*, 16(5):717–762, 2006. 14, 16

[Dav93] M. H. A. Davis. *Markov Models & Optimization*, volume 49. CRC Press, 1993. 9

[Dob53] R. L. Dobrušin. Limit theorems for a Markov chain of two states. *Izvestiya Akad. Nauk SSSR. Ser. Mat.*, 17:291–330, 1953. 2

[DS07] Z. Dietz and S. Sethuraman. Occupation laws for some time-nonhomogeneous Markov chains. *Electron. J. Probab.*, 12:no. 23, 661–683, 2007. 2, 3, 18

[Duf96] M. Duflo. *Algorithmes stochastiques*, volume 23 of *Mathématiques & Applications (Berlin) [Mathematics & Applications]*. Springer-Verlag, Berlin, 1996. 8, 25, 28, 29

[EK86] S. N. Ethier and T. G. Kurtz. *Markov processes*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York, 1986. Characterization and convergence. 9, 16, 28

[EV16] J. Engländer and S. Volkov. Turning a coin over instead of tossing it. *ArXiv e-prints*, June 2016. 2, 3, 19, 20

[FGM12] J. Fontbona, H. Guérin, and F. Malrieu. Quantitative estimates for the long-time behavior of an ergodic variant of the telegraph process. *Adv. in Appl. Probab.*, 44(4):977–994, 2012. 14, 15

[For15] G. Fort. Central limit theorems for stochastic approximation with controlled Markov chain dynamics. *ESAIM Probab. Stat.*, 19:60–80, 2015. 28

[Gan59] F. R. Gantmacher. *The Theory of Matrices*, volume 2. Chelsea, New York, 1959. 5, 7, 10

[Gou97] R. Gouet. Strong convergence of proportions in a multicolor Pólya urn. *J. Appl. Probab.*, 34(2):426–435, 1997. 3

[IW89] N. Ikeda and S. Watanabe. *Stochastic differential equations and diffusion processes*, volume 24 of *North-Holland Mathematical Library*. North-Holland Publishing Co., Amsterdam; Kodansha, Ltd., Tokyo, second edition, 1989. 9

[Jon04] G. L. Jones. On the Markov chain central limit theorem. *Probab. Surv.*, 1:299–320, 2004. 3

[Kol11] V. N. Kolokoltsov. *Markov processes, semigroups and generators*, volume 38 of *De Gruyter Studies in Mathematics*. Walter de Gruyter & Co., Berlin, 2011. 9

[Kun84] H. Kunita. Stochastic differential equations and stochastic flows of diffeomorphisms. In *École d'été de probabilités de Saint-Flour, XII—1982*, volume 1097 of *Lecture Notes in Math.*, pages 143–303. Springer, Berlin, 1984. 24

[KY03] H. Kushner and G. Yin. *Stochastic approximation and recursive algorithms and applications*, volume 35. Springer, 2003. 25

[MP87] M. Métivier and P. Priouret. Théorèmes de convergence presque sure pour une classe d'algorithmes stochastiques à pas décroissant. *Probab. Theory Related Fields*, 74(3):403–428, 1987. 25

[Pel12] M. Peligrad. Central limit theorem for triangular arrays of non-homogeneous Markov chains. *Probab. Theory Related Fields*, 154(3-4):409–428, 2012. 2

[SC97] L. Saloff-Coste. Lectures on finite Markov chains. In *Lectures on probability theory and statistics (Saint-Flour, 1996)*, volume 1665 of *Lecture Notes in Math.*, pages 301–413. Springer, Berlin, 1997. 10

[SV05] S. Sethuraman and S. R. S. Varadhan. A martingale proof of Dobrushin's theorem for non-homogeneous Markov chains. *Electron. J. Probab.*, 10:no. 36, 1221–1235, 2005. 2