Bernoulli **21**(1), 2015, 505–536
DOI: 10.3150/13-BEJ577

Exponential ergodicity for Markov processes with random switching

BERTRAND CLOEZ[1] and MARTIN HAIRER[2]

[1] *Institut de Mathématiques de Toulouse, Université de Toulouse, F-31062 Toulouse, France.*
E-mail: bertrand.cloez@univ-mlv.fr
[2] *Mathematics Department, University of Warwick, Coventry, CV4 7AL, U.K.*
E-mail: M.Hairer@Warwick.ac.uk

We study a Markov process with two components: the first component evolves according to one of finitely many underlying Markovian dynamics, with a choice of dynamics that changes at the jump times of the second component. The second component is discrete and its jump rates may depend on the position of the whole process. Under regularity assumptions on the jump rates and Wasserstein contraction conditions for the underlying dynamics, we provide a concrete criterion for the convergence to equilibrium in terms of Wasserstein distance. The proof is based on a coupling argument and a weak form of the Harris theorem. In particular, we obtain exponential ergodicity in situations which do not verify any hypoellipticity assumption, but are not uniformly contracting either. We also obtain a bound in total variation distance under a suitable regularising assumption. Some examples are given to illustrate our result, including a class of piecewise deterministic Markov processes.

Keywords: ergodicity; exponential mixing; piecewise deterministic Markov process; switching; Wasserstein distance

1. Introduction

Markov processes with switching are intensively used for modelling purposes in applied subjects like biology [10,12,16], storage modelling [7], neuronal activity [17,31]. This class of Markov processes is reminiscent of the so-called iterated random functions [14] or branching processes in random environment [32] in the discrete time setting. Several recent works [1,3–5,11,13,18,19] deal with their long time behaviour (existence of an invariant probability measure, Harris recurrence, exponential ergodicity, hypoellipticity...). In particular, in [1,4], the authors provide a kind of hypoellipticity criterion with Hörmander-like bracket conditions. Under these conditions, they deduce the uniqueness and absolute continuity of the invariant measure, provided that a suitable tightness condition is satisfied. They also obtain geometric convergence in the total variation distance. Nevertheless, there are many simple processes with switching which do not verify any hypoellipticity condition. To illustrate this fact, let us consider the simple example of [5]. Let (X, I) be the Markov process on $\mathbb{R}^2 \times \{-1, 1\}$ generated by

$$Af(x,i) = -\big(x-(i,0)\big)\cdot\nabla_x f(x,i) + \big(f(x,-i) - f(x,i)\big). \tag{1.1}$$

This process is ergodic and the first marginal π of its invariant measure is supported on $\mathbb{R} \times \{0\}$. This suggests that, in general, the law of the process does not converge to its invariant

1350-7265

measure in the total variation distance. However, it was proved in [5] that it converges in a certain Wasserstein distance. Let us recall that the pth Wasserstein distance $\mathcal{W}^{(p)}$, with $p \geq 1$, on a Polish space (E, d) is defined by

$$\mathcal{W}_d^{(p)}(\mu_1, \mu_2) = \inf_{X_1, X_2} \mathbb{E}\big[d(X_1, X_2)^p\big]^{1/p},$$

for any two probability measures μ_1, μ_2 on E, where the infimum is taken over all pairs of E-valued random variables X_1, X_2 with respective laws μ_1, μ_2. When $p = 1$, we set $\mathcal{W}_d = \mathcal{W}_d^{(1)}$. The Kantorovich–Rubinstein duality ([33], Theorem 5.10) shows that one also has

$$\mathcal{W}_d(\mu_1, \mu_2) = \sup_{f \in \mathrm{Lip}_1} \left(\int_E f \,\mathrm{d}\mu_1 - \int_E f \,\mathrm{d}\mu_2 \right),$$

where $f : E \mapsto \mathbb{R}$ is in Lip_1 if and only if it is a 1-Lipschitz function, namely

$$\forall x, y \in E, \qquad \big|f(x) - f(y)\big| \leq d(x, y).$$

The total variation distance d_{TV} can be viewed as the Wasserstein distance associated to the trivial distance function, namely

$$d_{\mathrm{TV}}(\mu_1, \mu_2) = \inf_{X_1, X_2} \mathbb{P}(X_1 \neq X_2) = \frac{1}{2} \sup_{\|f\|_\infty \leq 1} \left(\int_E f \,\mathrm{d}\mu_1 - \int_E f \,\mathrm{d}\mu_2 \right),$$

where the infimum is again taken over all random variables X_1, X_2 with laws μ_1, μ_2. In the present article, we will give convergence criteria for a general class of switching Markov processes. These processes are built from the following ingredients:

- a Polish space (E, d) and a finite set F;
- a family $(Z^{(n)})_{n \in F}$ of E-valued strong Markov processes represented by their semigroups $(P^{(n)})_{n \in F}$, or equivalently by their generators $(\mathcal{L}^{(n)})_{n \in F}$ with domains $(\mathcal{D}^{(n)})_{n \in F}$;
- a family $(a(\cdot, i, j))_{i, j \in F}$ of non-negative functions on E.

We are interested by the process $(\mathbf{X}_t)_{t \geq 0} = (X_t, I_t)_{t \geq 0}$, defined on $\mathbf{E} = E \times F$, which jumps between these dynamics. Roughly speaking, X_t behaves like $Z_t^{(I_t)}$ as long as I does not jump. The process I is discrete and jumps at a rate given by a. More precisely, the dynamics of $(\mathbf{X}_t)_{t \geq 0}$ is as follows:

- Given a starting point $(x, i) \in E \times F$, we take for $Z^{(i)}$ an instance as above with initial condition $Z_0^{(i)} = x$. The initial conditions for $Z^{(j)}$ with $j \neq i$ are irrelevant.
- The discrete component I is constant and equal to i until the time $T = \min_{j \in F} T_j$, where $(T_j)_{j \geq 0}$ is a family of random variables that are conditionally independent given $Z^{(i)}$ and that verify

$$\forall j \in F, \qquad \mathbb{P}(T_j > t | \mathcal{F}_t) = \exp\left(- \int_0^t a\big(Z_s^{(i)}, i, j\big) \,\mathrm{d}s \right),$$

where $\mathcal{F}_t = \sigma\{Z_s^{(i)} | s \leq t\}$.

- For all $t \in [0, T)$, we then set $X_t = Z_t^{(i)}$ and $I_t = i$.
- At time T, there exists a unique $j \in F$ such that $T = T_j$ and we set $I_T = j$ and $X_T = X_{T-}$.
- We take (X_T, I_T) as a new starting point at time T.

Let us make a few remarks about this construction. First, this algorithm guarantees the existence of our process under the condition that there is no explosion in the switching rate. In other words, our construction is global as long as I only switches value finitely many time in any finite time interval. Assumption 1.1 below will be sufficient to guarantee this non-explosion. Also note that, in general, X and I are not Markov processes by themselves, contrary to $\mathbf{X}$. Nevertheless, we have that I is a Markov process if a does not depend on its first component. The construction given above shows that, provided that there is no explosion, the infinitesimal generator of $\mathbf{X}$ is given by

$$\mathbf{L} f(x,i) = \mathcal{L}^{(i)} f(x,i) + \sum_{j \in F} a(x,i,j)\big(f(x,j) - f(x,i)\big), \tag{1.2}$$

for any bounded function f such that $f(\cdot, i)$ belongs to $\mathcal{D}^{(i)}$ for every $i \in F$. We will denote by $(\mathbf{P}_t)_{t\geq 0}$ the semigroup of $\mathbf{X}$. To guarantee the existence of our process, we will consider the following natural assumption:

Assumption 1.1 (Regularity of the jumps rates). *The following boundedness condition is verified*:

$$\bar{a} = \sup_{x\in E} \sup_{i \in F} \sum_{j\in F} a(x,i,j) < +\infty,$$

and the following Lipschitz condition is also verified:

$$\sup_{i\in F} \sum_{j \in F} \big| a(x,i,j) - a(x,i,j)\big| \leq \kappa d(x,y),$$

for some $\kappa > 0$.

We will also assume the following hypothesis to guarantee the recurrence of I:

Assumption 1.2 (Recurrence assumption). *The matrix* $(\underline{a}(i,j))_{i,j\in F}$ *defined by*

$$\underline{a}(i,j) = \inf_{x \in E} a(x,i,j),$$

yields the transition rates of an irreducible and positive recurrent Markov chain.

With these two assumptions, we are able to get exponential stability in two situations. The first situation is one where each underlying dynamics does on average yield a contraction in some Wasserstein distance, but no regularising assumption is made. The second situation is the opposite, where we replace the contraction by a suitable regularising property.

1.1. Two criteria without hypoellipticity assumption

In this section, we assume that we have some information on the Lipschitz contraction (or expansion) of our underlying processes:

Assumption 1.3 (Lipschitz contraction). *For each $i \in F$, there exists $\rho(i) \in \mathbb{R}$ such that*

$$\forall t \geq 0, \qquad \mathcal{W}_d\big(\mu P_t^{(i)}, \nu P_t^{(i)}\big) \leq \mathrm{e}^{-\rho(i)t} \mathcal{W}_d(\mu, \nu), \tag{1.3}$$

for any two probability measures μ, ν. Furthermore there exist $x_0 \in E$ and $t_{x_0} > 0$ such that if $V_{x_0} : x \mapsto d(x, x_0)$ then

$$\sup_{t \in [0, t_{x_0}]} P_t V_{x_0}(x_0) < +\infty.$$

In the previous assumption, given a semigroup $(P_t)_{t \geq 0}$, we used the notation μP_t to denote the measure defined by

$$(\mu P_t) f = \int P_t f \, \mathrm{d}\mu,$$

if $\mu = \delta_x$, for some x, then in this work, we also use the notation $\delta_x P_t(\mathrm{d}y) = P_t(x, \mathrm{d}y)$.

To verify equation (1.3) is not much of a restriction because we do not assume that $\rho(i) > 0$. The best constant in this inequality is called the Wasserstein curvature in [26,27] and the coarse Ricci curvature in [29,30], since it is heavily related to the geometry of the underlying space as illustrated in [34], Theorem 2. If $\rho(i) > 0$, then we can deduce some properties like geometric ergodicity, a Poincaré inequality or some concentration inequalities [9,25–27,30]. A trivial bound on $\rho(i)$ is given in the special case of diffusion processes in Section 4.1.

The bound (1.3) is quite stringent since, if $\rho(i) > 0$, it implies that there is some Wasserstein contraction for *every* $t > 0$ and not just for sufficiently long times. This is essentially equivalent to the existence of a Markovian coupling between two instances X_t and Y_t of the Markov process with generator $\mathcal{L}^{(i)}$ such that $\mathbb{E} d(X_t, Y_t) \leq \mathrm{e}^{-\rho t} d(X_0, Y_0)$.

In principle, this condition could be slightly relaxed by the addition of a proportionality constant C_i, provided that one assumes that the switching rate of the process is sufficiently slow. This ensures that, most of the time, it spends a sufficiently long time in any one state for this proportionality constant not to play a large role.

One could also imagine allowing for jumps of the component in E at the switching times, and this would lead to a similar difficulty.

In the same way, the distance d appearing in Assumption 1.3 is the *same* for every i and that it does not allow for a constant prefactor in the right-hand side of (1.3). This may seem like a very strong assumption since usual convergence theorems, like Harris' theorem, do not give this kind of bound. We will see however in Section 5 an example which illustrates that there is no obvious way in general to weaken this condition. The intuitive reason why this is so is that if the process switches rapidly, then it is crucial to have some local information (small times) and not only global information (large times) on the behaviour of each underlying dynamics.

We have now presented all the assumptions required to state our main results. The first one describes the simplest situation, that is when a does not depend on its first component:

Theorem 1.4 (Wasserstein exponential ergodicity in the constant case). *Under Assumptions 1.1, 1.2 and 1.3, if $a(x, i, j)$ does not depend on x and the Markov process I has an invariant probability measure ν verifying*

$$\sum_{i \in F} \nu(i)\rho(i) > 0,$$

then there exist a probability measure $\boldsymbol{\pi}$, some constants $C, \lambda > 0$ and $q \in (0, 1]$ such that

$$\forall t \geq 0, \qquad \mathcal{W}_{\mathbf{d}}(\delta_{\mathbf{y}_0}\mathbf{P}_t, \boldsymbol{\pi}) \leq C\mathrm{e}^{-\lambda t}\left(1 + \sum_{i \in F} \int_E d(y_0, x)^q \boldsymbol{\pi}(\mathrm{d}x, i)\right),$$

for every $\mathbf{y}_0 = (y_0, j_0) \in \mathbf{E}$, where the distance $\mathbf{d}$, on $\mathbf{E}$, is defined by

$$\mathbf{d}(\mathbf{x}, \mathbf{y}) = \mathbf{1}_{i \neq j} + \mathbf{1}_{i=j}\big(1 \wedge d^q(x, y)\big), \tag{1.4}$$

for every $\mathbf{x} = (x, i)$, $\mathbf{y} = (y, j)$ belonging to $\mathbf{E}$.

This statement is not surprising: it states that if the process contracts in mean, then it converges exponentially to an invariant distribution. The conditions are rather sharp as will be illustrated in Section 5. In particular, we recover [5], Theorem 1.10, and this (slight) generalisation could be deduced from the argument given there. Using Hölder's inequality, we can also deduce convergence in the pth Wasserstein distance $\mathcal{W}_{\mathbf{d}}^{(p)}$ with $p \geq 1$ provided that $\mathbf{X}$ satisfies a moment condition.

We provided Theorem 1.4 and its proof for sake of completeness and for a better understanding of the more complicated case, where a is allowed to depend on its first argument. In this situation, our main result reads as follows.

Theorem 1.5 (Wasserstein exponential ergodicity with an on–off type criterion). *Suppose that Assumptions 1.1, 1.2, and 1.3 hold and set*

$$\begin{aligned}
&F_0 = \big\{i \in F \,|\, \rho(i) > 0\big\} \quad \textit{and} \quad F_1 = \big\{i \in F \,|\, \rho(i) \leq 0\big\}, \\
&\rho_0 = \min_{i \in F_0} \rho(i) > 0 \quad \textit{and} \quad \rho_1 = \min_{i \in F_1} \rho(i) \leq 0, \\
&a_0 = \max_{i \in F_0} \sup_{x \in E} \sum_{j \in F_1} a(x, i, j) \quad \textit{and} \quad a_1 = \min_{i \in F_1} \inf_{x \in E} \sum_{j \in F_0} a(x, i, j).
\end{aligned}$$

If

$$\rho_0 a_1 + \rho_1 a_0 > 0,$$

then there exist a probability measure $\boldsymbol{\pi}$, *some constants* $C, \lambda > 0$ *and* $q \in (0,1]$ *such that*

$$\forall t \geq 0, \qquad \mathcal{W}_{\mathbf{d}}(\delta_{\mathbf{y}_0}\mathbf{P}_t, \boldsymbol{\pi}) \leq C\mathrm{e}^{-\lambda t}\left(1 + \sum_{i \in F} \int_E d(y_0, x)^q \boldsymbol{\pi}(\mathrm{d}x, i)\right),$$

for every $\mathbf{y}_0 = (y_0, j_0) \in \mathbf{E}$, *where the distance* $\mathbf{d}$ *on* $\mathbf{E}$ *is again given by* (1.4).

With this result, we not only recover [5], Theorem 1.15, but we extend it significantly. In our case, the underlying dynamics are not necessarily deterministic and do not need to be strictly contracting in a Wasserstein distance. One drawback is that the constants λ and C are much less explicit. This theorem is a direct consequence of the more general Theorem 3.3 below. These two theorems are our main results and, contrary to Theorem 1.4, it seems that they cannot be deduced directly from the approach of [5].

1.2. Two criteria with hypoellipticity assumption

In the previous subsection, we have supposed that some of the underlying dynamics contract at sufficiently high rate in a Wasserstein distance. This is of course not a necessary condition for geometric ergodicity in general. Using some arguments of the proof of Theorem 1.4 and Theorem 1.5, we can deduce a different criterion which uses instead a Lyapunov-type argument to prove that $\mathbf{X}$ converges. We begin by stating an assumption similar to Assumption 1.3:

Assumption 1.6 (Existence of a Lyapunov function). *There exist* $K \geq 0$, *a function* $V \geq 0$, *and for every* $i \in F$ *there exists* $\lambda(i) \in \mathbb{R}$ *such that*

$$\forall t \geq 0, \forall x \in E, \qquad P_t^{(i)} V(x) \leq \mathrm{e}^{-\lambda(i)t} V(x) + K. \tag{1.5}$$

Note again that we have not supposed that $\lambda(i) > 0$. One way to prove this kind of bound is to use the classical drift condition on the generator (see (2.2) below). With this assumption we are able to prove the following result, where the definition of a "small set" will be recalled in Definition 2.10 below.

Theorem 1.7 (Exponential ergodicity in the constant case). *Suppose that Assumptions 1.1, 1.2 and 1.6 hold, that* $a(x,i,j)$ *does not depend on* x *and that* I *has an invariant probability measure* ν *verifying*

$$\sum_{i \in F} \nu(i)\lambda(i) > 0.$$

If there exists $i_0 \in F$ *and* $t_0 \geq 0$ *such that the sublevel sets* $\{x \in E | V(x) \leq K\}$ *of* V *are small for* $P_t^{(i_0)}$ *for every* $K > 0$ *and* $t \geq t_0$, *then there exist a probability measure* $\boldsymbol{\pi}$ *and two constants* $C, \lambda > 0$ *such that*

$$\forall t \geq 0, \qquad d_{\mathrm{TV}}(\delta_{\mathbf{x}}\mathbf{P}_t, \boldsymbol{\pi}) \leq C\mathrm{e}^{-\lambda t}\bigl(1 + V(x)\bigr),$$

for every $\mathbf{x} = (x, i) \in \mathbf{E}$.

We give also the result analogous to Theorem 1.5.

Theorem 1.8 (Exponential ergodicity with an on–off type criterion). *Suppose that Assumptions 1.1, 1.2, 1.3 hold and set*

$$F_0 = \{i \in F | \lambda(i) > 0\} \quad \textit{and} \quad F_1 = \{i \in F | \lambda(i) \le 0\},$$
$$\lambda_0 = \min_{i \in F_0} \lambda(i) > 0 \quad \textit{and} \quad \lambda_1 = \min_{i \in F_1} \lambda(i) \le 0,$$
$$a_0 = \max_{i \in F_0} \sup_{x \in E} \sum_{j \in F_1} a(x, i, j) \quad \textit{and} \quad a_1 = \min_{i \in F_1} \inf_{x \in E} \sum_{j \in F_0} a(x, i, j).$$

If

$$\lambda_0 a_1 + \lambda_1 a_0 > 0,$$

and there exists $i_0 \in F$ and $t_0 \ge 0$ such that all sublevel sets of V are small for $P_t^{(i_0)}$, for every $t \ge t_0$, then there exist a probability measure $\boldsymbol{\pi}$ and two constants $C, \lambda > 0$ such that

$$\forall t \ge 0, \qquad d_{\mathrm{TV}}(\delta_{\mathbf{x}} \mathbf{P}_t, \boldsymbol{\pi}) \le C \mathrm{e}^{-\lambda t}\big(1 + V(x)\big),$$

for every $\mathbf{x} = (x, i) \in \mathbf{E}$.

Note that in general it is not necessary to assume that sublevel sets of V are small for any single one of the underlying dynamics. For example, using the results of [1,4], Section 4.2 gives results analogous to the two previous theorems, in the special case of piecewise deterministic Markov processes where the only small sets for the underlying dynamics consist of single points.

The remainder of the paper is organised as follows. The proofs of our four main theorems are split over two sections: Section 2 deals with the proof of Theorem 1.4 and Theorem 1.7. In Section 3, we begin by giving a more general assumption in the non-constant case than our on–off criterion. Then, we introduce a weak form of Harris' theorem that we will use to prove Theorem 1.5. The proof of this theorem is then decomposed in such a way to verify each point of the weak Harris' theorem. Section 4.1 gives sufficient conditions to verify our main assumption in the special case of diffusion processes. The section which follows deals with the special case of switching dynamical system. We conclude with Section 5, where we give some very simple examples illustrating the sharpness of our conditions.

2. Constant jump rates

In this section, we begin by proving that under Assumptions 1.3 or 1.6, the process $\mathbf{X}$ cannot wander off to infinity, that is, its semigroup possesses a Lyapunov function. We then prove Theorems 1.4 and 1.7 using a similar argument to [5] for the first one and Harris' theorem for the second one.

2.1. Construction of a Lyapunov function

We begin by recalling the definition of a Lyapunov function

Definition 2.1 (Lyapunov function). *A Lyapunov function for a Markov semigroup $(P_t)_{t\geq 0}$ over a Polish space (X, d_X) is a function $V : X \mapsto [0, \infty]$ such that V is integrable with respect to $P_t(x, \cdot)$ for every $x \in X$ and $t > 0$ and such that there exist constants $C_V, \gamma, K_V > 0$ verifying*

$$P_t V(x) = \int_X V(y) P_t(x, \mathrm{d}y) \leq C_V \mathrm{e}^{-\gamma t} V(x) + K_V, \tag{2.1}$$

for every $x \in X$ and $t \geq 0$.

A well-known sufficient condition for finding a Lyapunov function is the following drift condition:

$$\mathcal{L}V \leq -\gamma V + C, \tag{2.2}$$

where $\mathcal{L}$ is the generator of the semigroup $(P_t)_{t\geq 0}$. The condition (2.2) implies a bound like (1.5) and is clearly stronger than (2.1). In general, our switching Markov process $\mathbf{X}$ may not verify the drift condition (2.2) but, in Lemmas 2.8 and 3.9, we give a sharp condition under which it verifies (2.1). In this section, we first prove that a Wasserstein contraction as in Assumption 1.3 implies the existence of a Lyapunov-type function as in Assumption 1.6. Then, we will prove that Assumption 1.6 implies the existence of a Lyapunov function for $\mathbf{X}$.

Lemma 2.2 (Wasserstein contraction implies the existence of a Lyapunov-type function). *Let $(P_t)_{t\geq 0}$ be the semigroup of a Markov process, on a Polish space (X, d_X), such that there exists $\lambda \in \mathbb{R}^*$ verifying*

$$\mathcal{W}_{d_X}(\delta_x P_t, \delta_y P_t) \leq \mathrm{e}^{-\lambda t} d_X(x, y), \tag{2.3}$$

for every $x, y \in X$ and $t \geq 0$. If there exist $x_0 \in X$ and $t_{x_0} > 0$ such that the function $V_{x_0} : x \mapsto d(x, x_0)$ verifies

$$\sup_{t\in[0,t_{x_0}]} P_t V_{x_0}(x_0) < +\infty, \tag{2.4}$$

then there exist $C_1, C_2 > 0$ such that

$$P_t V_{x_0}(x) \leq \mathrm{e}^{-\lambda t}\big(V_{x_0}(x) + C_1\big) + C_2, \tag{2.5}$$

for every $x \in X$ and $t \geq 0$.

Proof. Note first that the bound (2.3) is equivalent to the bound

$$\mathcal{W}_{d_X}(\mu P_t, \nu P_t) \leq \mathrm{e}^{-\lambda t} \mathcal{W}_{d_X}(\mu, \nu), \tag{2.6}$$

for *every* probability measure μ and ν, as a consequence of the bound $\mathcal{W}_{d_X}(\mu P_t, \nu P_t) \leq \int \mathcal{W}_{d_X}(\delta_x P_t, \delta_y P_t)\pi(\mathrm{d}x, \mathrm{d}y)$ which follows immediately from the definitions and is true for any measure π with marginals μ and ν.

For any $t \geq t_{x_0}$ and $n \geq 0$, it then follows from (2.6) that

$$\begin{aligned} P_t V_{x_0}(x_0) = \mathcal{W}_{d_X}(\delta_{x_0} P_t, \delta_{x_0}) &\leq \sum_{k=0}^{n-1} \mathcal{W}_{d_X}(\delta_{x_0} P_{(k+1)t/n}, \delta_{x_0} P_{kt/n}) \\ &\leq \sum_{k=0}^{n-1} \mathrm{e}^{-\lambda kt/n} \mathcal{W}_{d_X}(\delta_{x_0} P_{t/n}, \delta_{x_0}) \leq \frac{\mathrm{e}^{-\lambda t} - 1}{\mathrm{e}^{-\lambda t/n} - 1} P_{t/n} V_{x_0}(x_0). \end{aligned}$$

Taking $n = \lfloor t/t_{x_0} \rfloor + 1$, where $\lfloor \lambda \rfloor$ denotes the integer part of λ, we conclude that

$$P_t V_{x_0}(x_0) \leq \big(\mathrm{e}^{-\lambda t} + 1\big) C', \qquad C' = \sup_{u \in [t_{x_0}/2, t_{x_0}]} \frac{P_u V_{x_0}(x_0)}{|\mathrm{e}^{-\lambda u} - 1|},$$

which is finite by (2.4). Finally, for every $x \in X$ and $t \geq 0$, we have

$$\begin{aligned} P_t V_{x_0}(x) = \mathcal{W}_{d_X}(\delta_x P_t, \delta_{x_0}) &\leq \mathcal{W}_{d_X}(\delta_x P_t, \delta_{x_0} P_t) + \mathcal{W}_{d_X}(\delta_{x_0} P_t, \delta_{x_0}) \\ &\leq \mathrm{e}^{-\lambda t} V_{x_0}(x) + \big(\mathrm{e}^{-\lambda t} + 1\big) C', \end{aligned}$$

thus concluding the proof. □

Remark 2.3. The point of this lemma is to also allow for negative values of λ. When $\lambda > 0$, then it is immediate that P_t admits a unique invariant measure and exhibits geometric ergodicity.

Remark 2.4. If V_{x_0} is in the domain of the generator $\mathcal{L}$ of $(P_t)_{t \geq 0}$, then we have

$$\forall t \geq 0, \qquad P_t V_{x_0}(x_0) \leq \frac{\mathrm{e}^{-\lambda t} - 1}{\mathrm{e}^{-\lambda t/n} - 1} P_{t/n} V_{x_0}(x_0),$$

for some $n \geq 1$. Now, taking the limit $n \to +\infty$, we deduce the following bound:

$$\mathcal{W}_{d_X}(\delta_{x_0} P_t, \delta_{x_0}) \leq \frac{\mathrm{e}^{-\lambda t} - 1}{-\lambda} \mathcal{L} V(x_0).$$

Finally, for every $x \in X$, we have

$$\begin{aligned} P_t V(x) = \mathcal{W}_{d_X}(\delta_x P_t, \delta_{x_0}) &\leq \mathcal{W}_{d_X}(\delta_x P_t, \delta_{x_0} P_t) + \mathcal{W}_{d_X}(\delta_{x_0} P_t, \delta_{x_0}) \\ &\leq \mathrm{e}^{-\lambda t} V(x) + \frac{\mathrm{e}^{-\lambda t} - 1}{-\lambda} \mathcal{L} V(x_0). \end{aligned}$$

However, V_{x_0} does not belong to the domain of the generator in general, as can be seen already in the example of simple Brownian motion.

Remark 2.5 (The special case $\lambda = 0$***).*** The assumption $\lambda \neq 0$ is required for our conclusion to hold. Indeed, if $(B_t)_{t\geq 0}$ is a Brownian motion then

$$\lim_{t\to+\infty} \mathbb{E}\big[|B_t|\big] = +\infty,$$

and inequality (2.5) does not hold. Instead, it is straightforward to follow the argument of the proof to show that if $\lambda = 0$ in Lemma 2.2, then one has the bound

$$P_t V_{x_0}(x) \leq V_{x_0}(x) + Ct,$$

for some fixed constant $C > 0$, every $x \in E$, and every $t \geq 0$.

Remark 2.6. By Lemma 2.2, Assumption 1.3 implies Assumption 1.6 with $\lambda = \rho$ and $V = V_{x_0}$ as long as one has $\rho(i) \neq 0$ for every i. In general, without any assumption on ρ, it does of course imply Assumption 1.6 for any function λ with $\lambda(i) < \rho(i)$, which is sufficient for our needs.

We now show that if Assumption 1.6 holds and the mean of $(\lambda(i))_{i\in F}$ is positive, then $\mathbf{X}$ admits a Lyapunov function. As in [5], this result is obtained as a consequence of the following lemma:

Lemma 2.7. *Let* $(K_t)_{t\geq 0}$ *be a continuous-time Markov chain on a finite set* S*, and assume that it is irreducible and positive recurrent with invariant measure* ν_K*. If* $\alpha : S \to \mathbb{R}$ *is a function verifying*

$$\sum_{n\in S} \nu_K(n)\alpha(n) > 0,$$

then there exist $C, c, \eta > 0$ *and* $p \in (0, 1]$ *such that*

$$c\mathrm{e}^{-\eta t} \leq \mathbb{E}\big[\mathrm{e}^{-\int_0^t p\alpha(K_s)\,\mathrm{d}s}\big] \leq C\mathrm{e}^{-\eta t},$$

for any initial condition K_0 *and every* $t \geq 0$.

Proof. It is a consequence of Perron–Frobenius theorem and the study of eigenvalues. See [3], Proposition 4.1, and [3], Proposition 4.2, for further details. □

Now we are able to prove that $\mathbf{P}$ possesses a Lyapunov function in the case where the switching rates do not depend on the location of the process.

Lemma 2.8. *Under Assumptions 1.1, 1.2 and 1.6, if* $a(x, i, j)$ *does not depend on* x *and* I *has an invariant measure* ν *satisfying*

$$\sum_{i\in F} \lambda(i)\nu(i) > 0,$$

then there exist $C_V, K_V, \lambda_V > 0$ *and* $q \in (0, 1]$ *such that*

$$\forall t \geq 0, \forall x \in E, \qquad \mathbf{P}_t V^q(x, i) \leq C_V \mathrm{e}^{-\lambda_V t} V^q(x) + K_V.$$

In the previous lemma, we used a slight abuse of notation. Indeed, if f is a function defined on E, we also denote by f the mapping $(x, i) \mapsto f(x)$ on $\mathbf{E}$.

Proof. First, Jensen's inequality gives this weaker form of (1.5):

$$P_t^{(i)}\big(V^q\big)(x) \le \mathrm{e}^{-q\lambda(i)t} V^q(x) + K^q,$$

for every $q \in (0, 1]$. Now, for all $t \ge 0$ and $(x, i) \in \mathbf{E}$, a straightforward recurrence gives

$$\begin{aligned}\mathbf{P}_t V^q(x, i) &= \mathbb{E}\big[P_{t-T_{N_t}}^{(I_{T_{N_t}})} \circ P_{T_{N_t}-T_{N_t-1}}^{(I_{T_{N_t}-1})} \circ \cdots \circ P_{T_1-T_0}^{(I_0)}\big(V^q\big)(x)\big] \\ &\le \mathbb{E}\big[\mathrm{e}^{-\int_0^t q\lambda(I_s)\,\mathrm{d}s}\big] V^q(x) + K^q \sum_{n\ge 0} \mathbb{E}\big[\mathrm{e}^{-q\int_0^{T_n} \lambda(I_s)\,\mathrm{d}s}\big],\end{aligned}$$

where $(T_k)_{k\ge 0}$ is the sequence of jump times of I, with $T_0 = 0$, and N_t the number of jumps before t. By Lemma 2.7, there exist $C > 0$, $\eta > 0$ and $q \in (0, 1]$ such that

$$\mathbb{E}\big[\mathrm{e}^{-\int_0^t q\lambda(I_s)\,\mathrm{d}s}\big] \le C\mathrm{e}^{-\eta t}.$$

Furthermore, one can show that T_n is of order n and that

$$K_V = K^q \sum_{n\ge 0} \mathbb{E}\big[\mathrm{e}^{-q\int_0^{T_n} \lambda(I_s)\,\mathrm{d}s}\big] \lesssim K^q \sum_{n\ge 0} \mathrm{e}^{-\varepsilon n} < +\infty,$$

for some $\varepsilon > 0$. We do not detail this argument now, but we will prove it in the slightly more difficult context of non-constant rate a in Lemma 3.9. This concludes the proof. □

Remark 2.9 (On the assumption that F is finite). It is natural to extend our results to the case where F is countably infinite. Obviously, we then have to add the assumption that I is positive recurrent, but this is not enough. Indeed, if for each $i \in F$, $C_1(i)$ and $C_2(i)$ denote the constants C_1, C_2, appearing in Lemma 2.2 applied on $Z^{(i)}$, then we should furthermore assume that

$$\sup_{i\in F}\big(C_1(i) + C_2(i)\big) < +\infty,$$

for the argument to go through.

2.2. Proof of Theorem 1.4

The proof of this result is obtained by a coupling construction. We first give a description of this construction and we then turn to the proof itself. Throughout this section, we make the standing assumption that the hypotheses of Theorem 1.4 hold. In particular, I is an ergodic finite-state Markov chain.

Let $\mathbf{x} = (x, i)$ and $\mathbf{y} = (y, j)$ be two points of $\mathbf{E}$, we will build a coupling $(\mathbf{X}, \mathbf{Y})$, starting from $(\mathbf{x}, \mathbf{y})$, such that each component is an instance of the Markov process generated by $\mathbf{L}$, and such that the distance $\mathbf{d}(\mathbf{X}_t, \mathbf{Y}_t)$ decreases to 0 at exponential rate. From now on, we fix the starting points of our coupling $\mathbf{x} = (x, i)$, $\mathbf{y} = (y, j)$. The processes $(\mathbf{X}_t)_{t\geq 0} = (X_t, I_t)_{t\geq 0}$ and $(\mathbf{Y}_t)_{t\geq 0} = (Y_t, J_t)_{t\geq 0}$ are then constructed as follows:

- First, we run both processes independently until the first hitting time $T_c = \inf\{t \geq 0 | I_t = J_t\}$ of the two components I and J. In case we start with an initial condition such that $i = j$, then we simply set $T_c = 0$.
- For times $s \geq T_c$, we set $I_s = J_s$ and we couple X and Y in such a way that

$$\forall k \geq 0, \qquad \mathbb{E}\big[d(X_{S_k}, Y_{S_k}) | \mathcal{F}_{S_{k-1}}\big] \leq \mathrm{e}^{-\rho(I_{S_{k-1}})(S_k - S_{k-1})} d(X_{S_{k-1}}, Y_{S_{k-1}}),$$

  where $(T_k)_{k\geq 0}$ is the sequence of jumps times of I, $S_k = T_k \wedge t$ and $(\mathcal{F}_s)_{s\geq 0}$ is the natural filtration associated to $(\mathbf{X}, \mathbf{Y})$.

The existence of a coupling satisfying the second point is an immediate consequence of Assumption 1.3.

Proof of Theorem 1.4. Recall first that if I and J are two independent finite-state Markov chains with transition rate a as in the statement of Theorem 1.4, then there exist constants $C_c, \theta_c > 0$ such that

$$\forall t \geq 0, \qquad \mathbb{P}(T_c > t) \leq C_c \mathrm{e}^{-\theta_c t}, \tag{2.7}$$

for any two initial conditions I_0 and J_0.

If $i = j$, then by Jensen's inequality and iteration, we have similarly to before

$$\mathbb{E}\big[d(X_t, Y_t)^q\big] \leq \mathbb{E}\big[\mathrm{e}^{-q\int_0^t \rho(I_s)\,\mathrm{d}s}\big] d(x, y)^q,$$

where $q \in (0, 1]$. By Lemma 2.7, there exist $C, \eta > 0$ and $q \in (0, 1]$ such that

$$\mathbb{E}\big[d(X_t, Y_t)^q\big] \leq C\mathrm{e}^{-\eta t} d(x, y)^q.$$

Now, for general i and j, we have

$$\begin{aligned}\mathbb{E}\big[\mathbf{d}(\mathbf{X}_t, \mathbf{Y}_t)\big] \leq{}& \mathbb{E}\Big[\sqrt{\mathbf{1}_{T_c \geq t/2}\big(1 + V^q(X_t) + V^q(Y_t)\big)}\Big] \\ &+ \mathbb{E}\Big[\sqrt{\mathbf{1}_{T_c \leq t/2} d(X_t, Y_t)^q \big(1 + V^q(X_t) + V^q(Y_t)\big)}\Big],\end{aligned}$$

where $V(x) = d(x, x_0)$. Now, Cauchy–Schwarz inequality, Equation (2.7), Lemma 2.2 and Lemma 2.8 give

$$\begin{aligned}\mathbb{E}\Big[\sqrt{\mathbf{1}_{T_c \geq t/2}\big(1 + V^q(X_t) + V^q(Y_t)\big)}\Big] &\leq \mathbb{P}(T_c \geq t/2)^{1/2} \mathbb{E}\big[1 + V^q(X_t) + V^q(Y_t)\big]^{1/2} \\ &\leq C_c \mathrm{e}^{-\theta_c t/4}\big(1 + C_V \mathrm{e}^{-\lambda_V t}\big(V^q(x) + V^q(y)\big) + 2K_V\big)^{1/2}.\end{aligned}$$

In the other hand, one has the bound

$$\begin{aligned}
&\mathbb{E}\Big[\sqrt{\mathbf{1}_{T_c\le t/2}d(X_t,Y_t)^q\big(1+V^q(X_t)+V^q(Y_t)\big)}\Big]\\
&\quad\le \mathbb{E}\big[\mathbf{1}_{T_c\le t/2}d(X_t,Y_t)^q\big]^{1/2}\mathbb{E}\big[1+V^q(X_t)+V^q(Y_t)\big]^{1/2}.
\end{aligned}\tag{2.8}$$

As a consequence of Lemmas 2.2 and 2.8, we also have the bound

$$\begin{aligned}
\mathbb{E}\big[\mathbf{1}_{T_c\le t/2}d(X_t,Y_t)^q\big]^{1/2} &\le C\mathrm{e}^{-\eta t/2}\mathbb{E}\big[d(X_{T_c},Y_{T_c})^q\mathbf{1}_{T_c\le t/2}\big]^{1/2}\\
&\le C\mathrm{e}^{-\eta t/2}\mathbb{E}\big[\big(V(X_{T_c})^q+V(Y_{T_c})^q\big)\mathbf{1}_{T_c\le t/2}\big]^{1/2}\\
&\le C\mathrm{e}^{-\eta t/2}\big[C_V V^q(x_0)+C_V V^q(y_0)+2K_V\big]^{1/2}.
\end{aligned}$$

Assembling these inequalities and using again Lemma 2.8 to bound the second factor in (2.8), we find that there exist constants $C>0$ and $\lambda>0$ such that

$$\mathbb{E}\big[\mathbf{d}(\mathbf{X}_t,\mathbf{Y}_t)\big]\le C\mathrm{e}^{-\lambda t}\big(1+V(x)+V(y)\big),$$

for every $t\ge 0$ and $x,y\in E$. (Recall that x and y denote the E-components of the initial conditions.) As a consequence of this bound and the definition of the Wasserstein distance, we deduce that

$$\mathcal{W}_{\mathbf{d}}(\boldsymbol{\mu}\mathbf{P}_t,\boldsymbol{\nu}\mathbf{P}_t)\le C\mathrm{e}^{-\lambda t}\left(1+\sum_{i\in F}\int_E\big(V(x)\boldsymbol{\nu}(\mathrm{d}x,i)+V(x)\boldsymbol{\mu}(\mathrm{d}x,i)\big)\right),\tag{2.9}$$

for any two probability measures $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$. Now, mimicking the proof of [23], Corollary 4.10, we can prove the existence of an invariant measure. More precisely, fix a probability measure $\boldsymbol{\mu}$ and note that (2.9) implies that $(\boldsymbol{\mu}\mathbf{P}_n)_{n\ge 0}$ is a Cauchy sequence with respect to the distance $\mathcal{W}_{\mathbf{d}}$. We deduce that it converges to a measure $\boldsymbol{\mu}_\infty$ verifying

$$\boldsymbol{\mu}_\infty\mathbf{P}_1=\boldsymbol{\mu}_\infty.$$

It immediately follows that $\boldsymbol{\pi}=\int_0^1\boldsymbol{\mu}_\infty\mathbf{P}_u\,\mathrm{d}u$ is invariant, just like in the classical proof of the Krylov–Bogolioubov criterion. $\square$

2.3. Proof of Theorem 1.7

Before we start the proof proper, we recall a version of Harris' theorem (also called Foster, Lyapunov, Meyn-Tweedie, Doeblin in the literature) that is suitable for our needs. This theorem yields exponential convergence to stationarity for a process which does not "escape to infinity" and verifies furthermore a Doeblin-type condition. More precisely, we use the following notion of a small set:

Definition 2.10. *A set $A \subset X$ is small for the semigroup $(P_t)_{t\geq 0}$ over a Polish space (X, d_X), if there exists a time $t > 0$ and a constant $\varepsilon > 0$ such that*

$$d_{\mathrm{TV}}(\delta_x P_t, \delta_y P_t) \leq 1 - \varepsilon$$

for every $x, y \in A$.

The classical Harris theorem [23,28] then states that

Theorem 2.11 (Harris). *Let $(P_t)_{t\geq 0}$ be a Markov semigroup over a Polish space (X, d_X) such that there exists a Lyapunov function V with the additional property that the sublevel sets $\{x \in X | V(x) \leq C\}$ are small for every $C > 0$. Then $(P_t)_{t\geq 0}$ has a unique invariant measure π and*

$$d_{\mathrm{TV}}(\delta_x P_t, \pi) \leq C\mathrm{e}^{-\gamma_* t}\big(1 + V(x)\big),$$

for some positive constants C and γ_*.

Note that one does not really need that *all* sublevel sets are small and one can have a slightly stronger conclusion by using a total variation distance weighted by V, see, for example, [23], Theorem 1.3.

Proof of Theorem 1.7. By Lemma 2.8, $\mathbf{P}$ admits V as Lyapunov function so, by Harris' theorem, it only remains to show that $\{V \leq C\}$ is small for $\mathbf{P}$, for every $C > 0$. Since V is a Lyapunov function, there exists $t_*^{(1)} > 0$ and $K > K_V$ (with K_V as in Lemma 2.8) such that

$$\forall t \geq t_*^{(1)}, \qquad \mathbb{E}\big[V(X_t)\big] \leq K,$$

uniformly over all $x \in E$ such that $V(x) \leq C$. Therefore, if $\mathbf{X}$ is a process generated by $\mathbf{L}$, it follows from Markov's inequality that

$$\mathbb{P}\big(V(X_t) \leq 2K\big) \geq \tfrac{1}{2},$$

uniformly over $t \geq t_*^{(1)}$.

Let now $i_0 \in F$ be as in the statement. Since $A = \{V \leq 2K\}$ is small for $P^{(i_0)}$, we obtain some $t_0 > 0$ and $\varepsilon > 0$, such that for all $x, y \in A$ there exists a coupling $(Z_t^{i_0,x}, Z_t^{i_0,y})$ verifying

$$\mathbb{P}\big(Z_t^{i_0,x} = Z_t^{i_0,y}\big) \geq \varepsilon, \qquad t \geq t_0, \tag{2.10}$$

and $Z_t^{i_0,x}$, $Z_t^{i_0,y}$ have respective law $\delta_x P_t^{(i_0)}$, $\delta_y P_t^{(i_0)}$.

By the irreducibility of the process I, one can find $t_* > t_*^{(1)}$ and $\delta > 0$ such that $\mathbb{P}(I_s = i_0, \forall s \in [t_*, t_* + t_0]) > \delta$, uniformly over the starting distributions. Let now $(\mathbf{X}_t, \mathbf{Y}_t)$ be the following coupling:

- the Markov chains I and J are independent over $t \in [0, t_* + t_0]$;
- the processes X and Y are independent over $t \in [0, t_*]$;

- conditionally on the set

$$B = \{V(X_{t_*}) \le 2K, V(Y_{t_*}) \le 2K, I_s = J_s = i_0, \forall s \in [t_*, t_* + t_0]\},$$

  the processes X and Y are coupled in such a way to verify (2.10), over $t \in [t_*, t_* + t_0]$;
- conditionally on B^c, they are coupled independently from each other.

The Markov property gives

$$\mathbb{P}\big(V(X_{t_*}) \le 2K, I_s = i_0, \forall s \in [t_*, t_* + t_0]\big) \ge \frac{\delta}{2}, \tag{2.11}$$

and so $\mathbb{P}(B) \ge \delta^2/4$. Combining this inequality with (2.10), we conclude that $\mathbb{P}(\mathbf{X}_{t_*+t_0} = \mathbf{Y}_{t_*+t_0}) \ge \delta^2\varepsilon/4$, uniformly over all initial conditions $\mathbf{x}$ and $\mathbf{y}$ with $V(x) \le C$ and $V(y) \le C$, as required. □

3. Non-constant jump rates

In all of this section, we now assume that a depends non-trivially on its first component, so that I by itself is not a Markov process anymore. We want to use again Lemma 2.7 to show that $\mathbf{X}$ converges, but this time we cannot use it directly on I. The idea is to consider an auxiliary process which does not depend to X and which will bound $(\rho(I_t))_{t\ge0}$ or $(\lambda(I_t))_{t\ge0}$. More precisely, we will assume the following assumption.

Assumption 3.1 (Birth–death type criterion in the non constant case). *There exist $\bar{n} \in \mathbb{N}$ and a partition $(F_n)_{0\le n\le\bar{n}}$ of F such that*

$$\forall n \le \bar{n}, \forall i \in F_n, \forall j \notin F_{n-1} \cup F_n \cup F_{n+1}, \forall x \in E, \qquad a(x, i, j) = 0,$$

where we have set $F_{-1} = F_{\bar{n}+1} = \varnothing$. Let $(L_t)_{t\ge0}$ be the continuous-time Markov chain on $\{0, \dots, \bar{n}\}$ with generator

$$Gf(n) = b(n)\big(f(n+1) - f(n)\big) + d(n)\big(f(n-1) - f(n)\big), \tag{3.1}$$

for every $n \le \bar{n}$, where $d(0) = b(\bar{n}) = 0$,

$$b(n) = \inf_{x\in E} \inf_{i\in F_n} \sum_{j\in F_{n+1}} a(x, i, j) > 0,$$

for $n < \bar{n}$ and

$$d(n) = \sup_{x\in E} \sup_{i\in F_n} \sum_{j\in F_{n-1}} a(x, i, j) > 0,$$

for $n > 0$.

Remark 3.2. The process with generator G is irreducible, non-explosive and positive recurrent. We will henceforth denote its invariant measure by ν.

If Assumption 3.1 holds then, for every $i \in F$, we denote by n_i the only $n \le \bar{n}$ verifying $i \in F_n$. Let us recall that, for every $n \le \bar{n}$, the invariant measure ν is given by

$$\nu(n) = \nu(0) \prod_{k=1}^{n} \frac{b(k-1)}{d(k)} \quad \text{and} \quad \nu(0) = (1+\Xi)^{-1},$$

where

$$\Xi = \sum_{n=1}^{\bar{n}} \frac{b(0)\cdots b(n-1)}{d(1)\cdots d(n)}.$$

Now we can state two slight generalisations of Theorems 1.5 and 1.8. The first one is

Theorem 3.3 (Wasserstein exponential ergodicity). *Suppose that Assumptions 1.1, 1.2, 1.3, and 3.1 hold. If*

$$\sum_{n=0}^{\bar{n}} \nu(n)\alpha(n) > 0,$$

where $(\alpha(n))_{n\ge 0}$ is an increasing sequence verifying $\alpha(n) \le \inf_{i\in F_n} \rho(i)$, then there exist a probability measure $\boldsymbol{\pi}$ and some constants $C, \lambda, t_0 > 0$ and $q \in (0,1]$ such that

$$\forall t \ge t_0, \qquad \mathcal{W}_{\mathbf{d}}(\delta_{\mathbf{y}_0}\mathbf{P}_t, \boldsymbol{\pi}) \le C\mathrm{e}^{-\lambda t}\left(1 + \sum_{i\in F} \int_E d(y_0, x)^q \boldsymbol{\pi}(\mathrm{d}x, i)\right),$$

for every $\mathbf{y}_0 = (y_0, j_0) \in \mathbf{E}$. Here, the distance $\mathbf{d}$ on $\mathbf{E}$ was defined in (1.4).

If Assumption 3.1 holds with $\bar{n} = 0$ then all contraction parameters are positive and we recover [5], Theorem 1.15. If it holds with $\bar{n} = 1$, then we have the on–off criterion which was given in introduction. We can also state the analogous result in the setting of Theorem 1.8:

Theorem 3.4 (Exponential ergodicity). *Suppose that Assumptions 1.1, 1.2, 1.3 and 3.1 hold and there exist $i_0 \in F$ and $t_0 \ge 0$ such that the sublevel sets of V are small for $P_t^{(i_0)}$, for every $t \ge t_0$. If*

$$\sum_{n=0}^{\bar{n}} \nu(n)\alpha(n) > 0,$$

where $(\alpha(n))_{n\ge 0}$ is an increasing sequence verifying $\alpha(n) \le \inf_{i\in F_n} \lambda(i)$, then there exist a probability measure $\boldsymbol{\pi}$ and two constants $C, \lambda > 0$ such that

$$\forall t \ge 0, \qquad d_{\mathrm{TV}}(\delta_{\mathbf{x}}\mathbf{P}_t, \boldsymbol{\pi}) \le C\mathrm{e}^{-\lambda t}\big(1 + V(x)\big)$$

for every $\mathbf{x} = (x, i) \in \mathbf{E}$.

We do not give the proofs of Theorem 1.8 and Theorem 3.4, as their proofs are very similar to the proof of Theorem 1.7, combined with the argument of Lemma 3.9 below. To prove Theorem 3.3 however, we cannot use classical Harris' Theorem. Its proof follows the same idea as the proof of Theorem 1.4, but there is no direct equivalent to the meeting time. Instead, we use a weak version of Harris' Theorem which yields geometric ergodicity under the existence of a Lyapunov function and a modified "small set" condition. This theorem was previously applied to the stochastic Navier–Stokes equation [22], stochastic delay differential equations [24], and linear response theory [21]. It is an extension of the classic Harris' Theorem which allows to deal with some degenerate examples like the one given in (1.1).

3.1. Weak form of Harris' Theorem

As already mentioned earlier, there are situations in which we cannot expect convergence in total variation. The problem here is that bounded sets may not be small sets. We will therefore replace the notion of small set by the following notion of "closedness" between transition probabilities introduced in [24], which takes into account the topology of the underlying space X.

***Definition 3.5** (d**-small set**).* *Let P be a Markov operator over a Polish space X endowed with a distance $d_X : X \times X \mapsto [0, 1]$. A set $A \subset X$ is said to be d_X-small if there exists a constant ε such that*

$$\mathcal{W}_{d_X}(\delta_x P, \delta_y P) \leq 1 - \varepsilon,$$

for every $x, y \in A$.

This notion is a generalisation of the notion of small set, since small sets are d-small for the trivial distance. This definition can also be extended to situations when d is not a distance [24]. As remarked in that paper, having a Lyapunov function V with d-small sublevel sets cannot be sufficient to imply the ergodicity of a Markov semigroup. To obtain some convergence result, we further impose that d is contracting for our semigroup:

***Definition 3.6** (d**-contracting operator**).* *Let P be a Markov operator over a Polish space X endowed with a distance $d_X : X \times X \mapsto [0, 1]$. The distance d_X is said to be contracting for P if there exists $\alpha < 1$ such that the bound*

$$\mathcal{W}_{d_X}(\delta_x P, \delta_y P) \leq \alpha d_X(x, y)$$

holds for every $x, y \in X$ verifying $d(x, y) < 1$.

Note that this condition alone is not sufficient to guarantee the convergence of transition probabilities toward a unique invariant measure since we only impose a contraction when $d(x, y) < 1$. In typical situations, "most" pairs (x, y) may satisfy $d(x, y) = 1$, as would be the case for the total variation distance. However, when combined with the existence of a Lyapunov function V that has d-small sublevel sets, it gives geometrical ergodicity ([24], Theorem 4.7):

Theorem 3.7 (Weak form of Harris' Theorem). *Let $(P_t)_{t\geq 0}$ be a Markov semigroup over a Polish space X admitting a continuous Lyapunov function V. Assume furthermore that there exist $t^* > t_* > 0$ and a distance $d_X : X \times X \mapsto [0,1]$ which is contracting for P_t and such that the sublevel set $\{x \in X | V(x) \leq 4K_V\}$ is d_X-small for P_t, for every $t \in [t_*, t^*]$. Here K_V is as in Definition 2.1. Then, $(P_t)_{t\geq 0}$ has an invariant probability measure π. Furthermore, defining*

$$\delta_X(x,y) = \sqrt{d_X(x,y)\big(1 + V(x) + V(y)\big)},$$

there exist $r > 0$ and $t_0 > 0$ such that

$$\forall t \geq t_0, \qquad \mathcal{W}_{\delta_X}(\mu P_t, \nu P_t) \leq \mathrm{e}^{-rt}\mathcal{W}_{\delta_X}(\mu, \nu),$$

for all of probability measures μ, ν on X.

Remark 3.8 (On the contracting distances). The main difficulty when applying the previous theorem is to find a contracting distance. The construction of this distance represents the main part of our paper. In [21], there is a general way to build a contracting distance of a Markov operator P over a Banach space $(\mathbb{B}, \|\cdot\|)$, based on a gradient estimate for P and the existence of a super-Lyapunov function. This technique was efficient in [21,22].

3.2. Construction of a Lyapunov function

As in the constant case, we first show that if each underlying Markov process verifies a weaker form of the drift condition (2.2) then $\mathbf{X}$ possesses a Lyapunov function:

Lemma 3.9 (Construction of a Lyapunov function). *Suppose that Assumptions 1.1, 1.2, 1.6 and 3.1 hold, if*

$$\sum_{n\geq 0} \nu(n)\alpha(n) > 0,$$

where $(\alpha(n))_{n\geq 0}$ is an increasing sequence verifying $\alpha(n) \leq \inf_{i\in F_n} \lambda(i)$, then there exist $C_V, K_V, \lambda_V > 0$ and $q \in (0,1)$ such that, for all $t \geq 0$ and all $(x,i) \in \mathbf{E}$, the bound

$$\mathbf{P}_t V^q(x,i) \leq C_V \mathrm{e}^{-\lambda_V t} V^q(x) + K_V \tag{3.2}$$

holds.

Proof. Recall again that Jensen's inequality gives this weaker form of (1.5):

$$\big(P_t^{(i)} V^q\big)(x) \leq \mathrm{e}^{-q\alpha(i)t} V^q(x) + K^q,$$

for every $x \in E$ and $q \in (0,1]$. Note also that, as a consequence of the Markov property, (3.2) follows if we are able to find *some* $T > 0$ and constants $C, K > 0$ and $q \in (0,1]$ such that

$$\mathbf{P}_T V^q(x,i) \leq \tfrac{1}{2} V^q(x) + K, \tag{3.3}$$

and such that

$$\mathbf{P}_t V^q(x,i) \le C V^q(x) + K, \tag{3.4}$$

for all $t \in [0,T]$. In order to find such a time T, we will build a process which couples a copy of $\mathbf{X}$ with the birth and death process L of Assumption 3.1. We define a generator $\mathcal{G}$ on $\mathbf{E} \times \{0,\dots,\bar{n}\}$ by

$$\begin{aligned}\mathcal{G} f(x,i,l) &= \mathcal{L}^{(i)} f(x,i,l) + \sum_{j\in F} a(x,i,j)\big(f(x,j,l) - f(x,i,l)\big)\\ &\quad + b(l)\big(f(x,i,l+1) - f(x,i,l)\big) + d(l)\big(f(x,i,l-1) - f(x,i,l)\big)\end{aligned}$$

for $l \neq n_i$. For $l = n_i$ on the other hand, we set

$$\begin{aligned}\mathcal{G} f(x,i,l) &= \mathcal{L}^{(i)} f(x,i,l) + \sum_{j\in F_{l-1}} a(x,i,j)\big(f(x,j,l-1) - f(x,j,l)\big)\\ &\quad + \Big(d(l) - \sum_{j\in F_{l-1}} a(x,i,j)\Big)\big(f(x,i,l-1) - f(x,j,l)\big)\\ &\quad + \sum_{j\in F_l} a(x,i,j)\big(f(x,j,l) - f(x,i,l)\big)\\ &\quad + \frac{b(l)}{\sum_{k\in F_{l+1}} a(x,i,k)} \sum_{j\in F_{l+1}} a(x,i,j)\big(f(x,j,l+1) - f(x,j,l)\big)\\ &\quad + \frac{\sum_{k\in F_{l+1}} a(x,i,k) - b(l)}{\sum_{k\in F_{l+1}} a(x,i,k)} \sum_{j\in F_{l+1}} a(x,i,j)\big(f(x,j,l) - f(x,j,l)\big).\end{aligned}$$

In words, as long as $L \neq n_I$, L and $\mathbf{X}$ move independently from each other until the time where n_I and L agree. After that time, the coupling is designed in such a way that one always has $n_I \ge L$. If we start the process with an initial condition (x,i,l) such that $n_i \ge l$, this construction ensures in particular that, for all times, one has

$$\alpha(I_t) \ge \alpha(L_t).$$

We now denote by $\{\tau_n\}_{n\ge 1}$ the times at which the process I_t jumps and by N_t the number of such jumps before time t.

With these notations at hand, we then have

$$\begin{aligned}\mathbf{P}_t V^q(x) &= \mathbb{E}\big[P_{t-\tau_{N_t}}^{(I_{\tau_{N_t}})} V^q(X_{\tau_{N_t}})\big] \le \mathbb{E}\big[\mathrm{e}^{-\alpha(I_{\tau_{N_t}})(t-\tau_{N_t})} V^q(X_{\tau_{N_t}}) + K^q\big]\\ &\le \mathbb{E}\big[\mathrm{e}^{-\int_{\tau_{N_t}}^t \alpha(L_s)\,\mathrm{d}s} V^q(X_{\tau_{N_t}})\big] + K^q\end{aligned} \tag{3.5}$$

$$= \mathbb{E}\left[\mathrm{e}^{-\int_{\tau_{N_t}}^t \alpha(L_s)\,\mathrm{d}s} P_{\tau_{N_t}-\tau_{N_t-1}}^{(I_{\tau_{N_t-1}})} V^q(X_{\tau_{N_t-1}})\right] + K^q$$
$$\leq \cdots \leq \mathbb{E}\left[\mathrm{e}^{-\int_0^t q\alpha(L_s)\,\mathrm{d}s}\right] V^q(x) + K^q \mathbb{E}\left[\sum_{n\leq N_t} \mathrm{e}^{-q\int_{\tau_n}^t \alpha(L_s)\,\mathrm{d}s}\right].$$

Now, using Lemma 2.7, there exist $C, \eta > 0$ and $q \in (0, 1]$ such that

$$\mathbb{E}\left[\mathrm{e}^{-\int_0^t q\alpha(L_s)\,\mathrm{d}s}\right] \leq C\mathrm{e}^{-\eta t}. \tag{3.6}$$

Hence, in view of (3.3) and (3.4), it only remains to prove that, for any fixed time T, one has the bound

$$\sup_{t\leq T} \mathbb{E}\left[\sum_{n\leq N_t} \mathrm{e}^{-q\int_{\tau_n}^t \alpha(L_s)\,\mathrm{d}s}\right] < +\infty.$$

Since the function α is bounded from below and the function $t \mapsto N_t$ is increasing, this boils down to the bound $\mathbb{E}N_T < \infty$, which is a simple consequence of the fact that by Assumption 1.1 the jump rates are also bounded from above. □

3.3. The contracting distance

This section is divided in three parts. We introduce the distance $\widetilde{d}$ that we will use in Theorem 3.7, we build our coupling in such a way that $\widetilde{d}$ will be contracting for it, and we finally prove that it is indeed contracting.

3.3.1. *Definition of* $\widetilde{d}$

Here, we build a distance $\widetilde{d} : (E \times F) \times (E \times F) \to [0, 1]$ such that there exist $t_* > 0$ and $\alpha \in (0, 1)$ verifying

$$\widetilde{d}(\mathbf{x}, \mathbf{y}) < 1 \quad \Rightarrow \quad \forall t \geq t_*, \qquad \mathcal{W}_{\tilde{d}}(\delta_{\mathbf{x}} P_t, \delta_{\mathbf{y}} P_t) \leq \alpha \widetilde{d}(\mathbf{x}, \mathbf{y}). \tag{3.7}$$

where $\mathbf{x} = (x, i)$ and $\mathbf{y} = (y, j)$ belong to $E \times F$. Since we can say nothing when $i \neq j$, we will take $\widetilde{d}(\mathbf{x}, \mathbf{y})$ constant equal to 1 in this case. When $i = j$ we want to use Assumption 1.3 to prove a decay. But it is more useful to "decrease the contraction" of the underlying Markov semigroup. More precisely, by Jensen inequality, Assumption 1.3 gives

$$\mathcal{W}_{d^q}\left(\mu P_t^{(i)}, \nu P_t^{(i)}\right) \leq \mathrm{e}^{-q\rho(i)t} \mathcal{W}_{d^q}(\mu, \nu),$$

for all $t \geq 0$, $q \in (0, 1]$ and every probability measures μ, ν. Finally, we define $\widetilde{d}$ by

$$\widetilde{d}(\mathbf{x}, \mathbf{y}) = \mathbf{1}_{i\neq j} + \mathbf{1}_{i=j}\left(\delta^{-1} d^q(x, y) \wedge 1\right),$$

where $\delta > 0$ will be determined later. Now, if a realisation of the coupling $(\mathbf{X}_t, \mathbf{Y}_t)_{t\geq 0} = ((X_t, I_t), (Y_t, J_t))_{t\geq 0}$ starting from $(\mathbf{x}, \mathbf{y})$, verifies $\widetilde{d}(\mathbf{x}, \mathbf{y}) < 1$, then $I_0 = J_0 = i = j$. So, we will try to build our coupling in such a way that I and J remain equal for as long as possible. More precisely, if we set

$$T = \inf\{s \geq 0 | I_s \neq J_s\}, \tag{3.8}$$

then we will prove that there exists $K > 0$ and a choice of coupling such that

$$\mathbb{P}(T < \infty) \leq K d(x, y).$$

3.3.2. *Construction of our coupling*

Here, we fix $\mathbf{x} = (x, i)$, $\mathbf{y} = (y, j)$ in $\mathbf{E}$ and we let $t > 0$. Let $r \geq 0$ and $(N_t)_{t\geq 0}$ be a Poisson process of intensity r with $N_t = \sum_{n\geq 0} \mathbf{1}_{\{\tau_n \leq t\}}$ and $\tau_n = \sum_{k=1}^{n} E_k$ for a family $(E_k)_{k\geq 0}$ of i.i.d. exponential variables and $\tau_0 = 0$. We assume that $r \geq 2\bar{a}$, that is r is larger than the jump rates of I or J. As in the proof of Theorem 1.4, we give the construction of our coupling $(\mathbf{X}, \mathbf{Y})$ at the jump times of N. Let $n \in \{0, \dots, N_t\}$, we consider the following dynamics:

- If $I_{\tau_n} \neq J_{\tau_n}$, then X_s and Y_s evolve independently for every $s \in [\tau_n, \tau_{n+1} \wedge t)$.
- If $I_{\tau_n} = J_{\tau_n}$, then by Assumption 1.3, we can couple X and Y in such a way that

$$\mathbb{E}\big[d(X_{\tau_{n+1}\wedge t}, Y_{\tau_{n+1}\wedge t}) | \mathcal{G}_{\tau_n}\big] \leq \mathrm{e}^{-\rho(I_{\tau_n})(\tau_{n+1}\wedge t - \tau_n)} d(X_{\tau_n}, Y_{\tau_n}),$$

  where $\mathcal{G}_n = \sigma\{(\mathbf{X}_{\tau_n}, \mathbf{Y}_{\tau_n}), (\tau_k)_{k\geq 0}\}$.

At the jump times of N the situation is different since I or J may jump. We will optimise the chance that I and J jump simultaneously. For each $n \in \mathbb{N}^*$, we cut $[0, 1]$ in four parts $I_0^n, I_1^n, I_2^n, I_3^n$ in such a way that

$$\lambda(I_0^n) = \frac{1}{r} \sum_{j\in F} \big(a(X_{\tau_n-}, I_{\tau_n}, j) - a(Y_{\tau_n-}, I_{\tau_n}, j)\big)_+,$$

$$\lambda(I_1^n) = \frac{1}{r} \sum_{j\in F} \big(a(Y_{\tau_n-}, I_{\tau_n}, j) - a(X_{\tau_n-}, I_{\tau_n}, j)\big)_+,$$

$$\lambda(I_2^n) = \frac{1}{r} \sum_{j\in F} a(X_{\tau_n-}, I_{\tau_n}, j) \wedge a(Y_{\tau_n-}, I_{\tau_n}, j),$$

$$\lambda(I_3^n) = 1 - \frac{1}{r} \sum_{j\in F} a(X_{\tau_n-}, I_{\tau_n}, j) \vee \sum_{j\in F} a(Y_{\tau_n-}, I_{\tau_n}, j),$$

where λ is the Lebesgue measure and $(x)_+ = \max(x, 0)$. Let $(U_n)_{n\geq 0}$ be a sequence of i.i.d. random variables uniformly distributed on $[0, 1]$, we couple I and J at the jump times as follows:

- For $U_n \in I_0^n$, I jumps, but J does not jump.
- For $U_n \in I_1^n$, J jumps, but I does not jump.
- For $U_n \in I_2^n$, I and J both jump simultaneously to the same location.
- For $U_n \in I_3^n$, I and J both stay in place.

The second components, X and Y, do not jump. Finally, we also couple $\mathbf{X}$ and $\mathbf{Y}$ with a continuous Markov chain L which only depend to U and N and which verifies

$$\forall t \geq 0, \qquad \rho(I_t) \geq \alpha(L_t).$$

This Markov chain L is constructed as in the proof of Lemma 3.9.

Remark 3.10. This coupling is not quite Markovian since, between times τ_n and τ_{n+1}, it already uses information about the pair (X_t, Y_t) at time τ_{n+1}. However, in many situations to which our results apply there exists a Markovian coupling with generator $\mathbb{L}^{(i)}$ which yields a good coupling for each of the underlying processes. In this case, we can make our coupling Markovian with generator

$$\begin{aligned}
\mathbb{L} f(\mathbf{x}, \mathbf{y}, n) = \mathbb{L}^{(i)} f(\mathbf{x}, \mathbf{y}, n) &+ \sum_{k \in F} \big(a(x, i, k) - a(y, j, k)\big)_+ f\big((x, k), \mathbf{y}, n+1\big) \\
&+ \sum_{k \in F} \big(a(y, j, k) - a(x, i, k)\big)_+ f\big(\mathbf{x}, (y, k), n+1\big) \\
&+ \sum_{k \in F} a(x, i, k) \wedge a(y, j, k) f\big((x, k), (y, k), n+1\big) \\
&+ \Big(r - \sum_{k \in F} a(x, i, k) \vee a(y, j, k)\Big) f(\mathbf{x}, \mathbf{y}, n+1) - r f(\mathbf{x}, \mathbf{y}, n).
\end{aligned}$$

3.3.3. *The distance $\widetilde{d}$ is contracting for* $\mathbf{P}$

In this subsection, we show that the distance $\widetilde{d}$ defined above is indeed contracting for the coupling constructed in the previous subsection. This is formulated in the following result.

Lemma 3.11. *Let $(\mathbf{X}_t, \mathbf{Y}_t)_{t \geq 0}$ be the coupling of the previous section. Under the assumptions of Theorem 3.3, we can choose r and δ in such a way that*

$$\forall t \geq t_*, \qquad \mathbb{E}\big[\widetilde{d}(\mathbf{X}_t, \mathbf{Y}_t)\big] \leq \gamma \widetilde{d}(\mathbf{x}, \mathbf{y}),$$

for some $\gamma \in (0, 1)$ and $t_* > 0$, and all $\mathbf{x}, \mathbf{y} \in E \times F$ verifying $\widetilde{d}(\mathbf{x}, \mathbf{y}) < 1$.

Proof. Recall that since $\widetilde{d}(\mathbf{x}, \mathbf{y}) < 1$ one has $I_0 = J_0$ and that T, defined in (3.8), denotes the first time of separation of I and J. Using Lemma 2.7, there exist $q \in (0, 1]$ and $C, \eta > 0$ such that

$$\begin{aligned}
\mathbb{E}\big[\widetilde{d}(\mathbf{X}_t, \mathbf{Y}_t)\big] &\leq \mathbb{E}\left[\mathbf{1}_{\{T=\infty\}} \frac{1}{\delta} d^q(X_t, Y_t) + \mathbf{1}_{\{T<+\infty\}}\right] \\
&\leq \frac{1}{\delta} \mathbb{E}\big[\mathrm{e}^{-\int_0^t q\alpha(L_s)\,\mathrm{d}s}\big] \mathbb{E}\big[d^q(x, y)\big] + \mathbb{P}(T < +\infty) \\
&\leq C \mathrm{e}^{-\eta t} \widetilde{d}(\mathbf{x}, \mathbf{y}) + \mathbb{P}(T < +\infty).
\end{aligned}$$

Here, we have used the fact that

$$\begin{aligned}
\mathbb{E}\big[\mathbf{1}_{\{T=\infty\}} d^q(X_t, Y_t)\big] &\le \mathbb{E}\big[\mathbf{1}_{\{T \ge \tau_{N_t}\}} \mathrm{e}^{-q\alpha(L_{\tau_{N_t}})(t-\tau_{N_t})} d^q(X_{\tau_{N_t}}, Y_{\tau_{N_t}})\big] \\
&\le \mathbb{E}\big[\mathbf{1}_{\{T \ge \tau_{N_t}\}} \mathrm{e}^{-q\alpha(L_{\tau_{N_t}})(t-\tau_{N_t})} \mathbb{E}\big[d^q(X_{\tau_{N_t}}, Y_{\tau_{N_t}})|\mathcal{G}_n\big]\big] \\
&\le \mathbb{E}\big[\mathbf{1}_{\{T \ge \tau_{N_t-1}\}} \mathrm{e}^{-\int_{\tau_{N_t-1}}^t q\alpha(L_s)\,\mathrm{d}s} d^q(X_{\tau_{N_t-1}}, Y_{\tau_{N_t-1}})\big] \\
&\le \mathbb{E}\big[\mathrm{e}^{-\int_0^t q\alpha(L_s)\,\mathrm{d}s}\big]\mathbb{E}\big[d^q(x, y)\big].
\end{aligned}$$

It remains to obtain a bound on $\mathbb{P}(T < +\infty)$. Since I and J can only jump when N jumps, T can be finite only if it is one of the jump times of N. So, we set

$$A_n = \{T = \tau_n\} = \{T \ge \tau_n \text{ and } I_{\tau_n} \ne J_{\tau_n}\}.$$

By Assumption 1.1, we have

$$\begin{aligned}
\mathbb{P}(A_n) &= \mathbb{P}\big(\big\{U_n \in I_0^n \cup I_1^n \cup I_3^n\big\} \cap \{T \ge \tau_n\}\big) \\
&\le \mathbb{E}\left[\frac{2\mathbf{1}_{\{T\ge\tau_n\}} \sum_{j\in F} |a(X_{\tau_n-}, I_{\tau_n-}, j) - a(Y_{\tau_n-}, I_{\tau_n-}, j)|}{r}\right] \\
&\le \mathbb{E}\left[\left(\frac{2\mathbf{1}_{\{T\ge\tau_n\}} \sum_{j\in F} |a(X_{\tau_n-}, I_{\tau_n-}, j) - a(Y_{\tau_n-}, I_{\tau_n-}, j)|}{r}\right)^q\right] \\
&\le \frac{2^q\kappa^q}{r^q}\mathbb{E}\big[d(X_{\tau_n-}, Y_{\tau_n-})^q\big] \le \frac{2^q\kappa^q}{r^q}\mathbb{E}\big[\mathrm{e}^{-q\int_0^{\tau_n} \alpha(L_s)\,\mathrm{d}s}\big]d(x, y)^q.
\end{aligned}$$

Hence,

$$\mathbb{P}(T < \infty) = \sum_{n\ge1} \mathbb{P}(A_n) \le \frac{2^q\kappa^q}{r^q} d(x, y)^q \sum_{n\ge1} \mathbb{E}\big[\mathrm{e}^{-q\int_0^{\tau_n} \alpha(L_s)\,\mathrm{d}s}\big].$$

Now, similarly to the proof of Lemma 3.9, provided that r is sufficiently large, there exist $C' > 0$ and $\varepsilon > 0$ verifying

$$\sum_{n\ge1} \mathbb{E}\big[\mathrm{e}^{-q\int_0^{\tau_n} \alpha(L_s)\,\mathrm{d}s}\big] \le \sum_{n\ge1} C'\mathrm{e}^{-\varepsilon n} =: \tilde{C} < +\infty.$$

Combining these bounds, we obtain the estimate

$$\mathbb{E}\big[\widetilde{d}(\mathbf{X}_t, \mathbf{Y}_t)\big] \le \left(C\mathrm{e}^{-\eta t} + \frac{(2\kappa)^q\tilde{C}}{r^q}\delta\right)\widetilde{d}(\mathbf{x}, \mathbf{y}).$$

First making δ sufficiently small and then taking t large enough, we thus obtain the announced result. □

3.4. Bounded sets are $\widetilde{d}$-small

Here, we prove that if a set is bounded then it is $\widetilde{d}$-small.

Lemma 3.12. *Under the assumptions of Theorem 3.3, if $S \subset E \times F$ is of bounded diameter in the sense that*

$$R = \sup\big\{d(x, y) | \mathbf{x}, \mathbf{y} \in S\big\} < +\infty,$$

then there exist $t_*, t^* > 0$ such that S is $\widetilde{d}$-small for P_t, for all $t \in [t_*, t^*]$.

Proof. Let $\mathbf{x} = (x, i)$ and $\mathbf{y} = (y, j)$ be two different points of S. By the assumptions of Theorem 3.3, there exists $i_0 \in F$ such that $\rho(i_0) > 0$. Let $(\mathbf{X}_t)_{t\geq 0}$ and $(\mathbf{Y}_t)_{t\geq 0}$ be two independent processes generated by (1.2) and starting respectively from $\mathbf{x}$ and $\mathbf{y}$. Let us denote

$$\tau_{\text{in}} = \inf\{t \geq 0 | I_t = J_t = i_0\} \quad \text{and} \quad \tau_{\text{out}} = \inf\{t \geq \tau_{\text{in}} | I_t \neq i_0 \text{ or } J_t \neq i_0\}.$$

For every $b, c > 0$ such that $b > c$, we define

$$p_{c,b}(\mathbf{x}, \mathbf{y}) = \mathbb{P}(\tau_{\text{in}} < c, \tau_{\text{out}} > b).$$

By Assumptions 1.1 and 1.2, we have $p_{c,b}(\mathbf{x}, \mathbf{y}) > 0$. Using the fact that a is bounded, a coupling argument shows that $p_{c,b}$ is lower bounded by a positive quantity which only depends on i and j. We then obtain the bound

$$\begin{aligned}
\mathbb{E}\big[\widetilde{d}(\mathbf{X}_t, \mathbf{Y}_t)\big] &\leq \mathbb{E}\big[\mathbf{1}_{\{\tau_{\text{in}} < c, \tau_{\text{out}} > b\}} \widetilde{d}(\mathbf{X}_t, \mathbf{Y}_t)\big] + 1 - p_{c,b}(\mathbf{x}, \mathbf{y}) \\
&\leq 1 - p_{c,b}(\mathbf{x}, \mathbf{y})\big(1 - \delta^{-1} \mathrm{e}^{\varrho c} \mathrm{e}^{-\rho(i_0)t} d(x, y)\big) \\
&\leq 1 - p_{c,b}(\mathbf{x}, \mathbf{y})\big(1 - \delta^{-1} \mathrm{e}^{\varrho c} \mathrm{e}^{-\rho(i_0)t} R\big),
\end{aligned}$$

where ϱ is given by

$$\varrho = -\min\big\{q\alpha(k) | k \in F\big\}.$$

There exist $c > 0$ and $t_* > c$ such that $1 - \delta^{-1} \mathrm{e}^{\varrho c} \mathrm{e}^{-\rho(i_0)t_*} R > 0$. Since F is finite, we can furthermore bound $p_{c,b}$ from below by the minimum over all $i, j \in F$, and the result follows for any $b > t_*$ and $t^* \in (t_*, b)$. □

Remark 3.13. One can see from this proof that it is not necessary that the jump rates are lower bounded, as in Assumption 1.2. Indeed, we need that, for each $i, j \in F$, the jump times of I are stochastically smaller than a variable which does not depend of the dynamics of X.

3.5. Proofs of Theorem 1.5 and Theorem 3.3

Recall that Lemmata 2.2 and 3.9 yield the existence of a Lyapunov function $V = V_{x_0}$, for some $x_0 \in E$, Lemma 3.11 shows that $\widetilde{d}$ is contracting for $\mathbf{P}$, and Lemma 3.12 proves that sublevel sets of V are $\widetilde{d}$-small. So we can use Theorem 3.7 to deduce that there exist a probability measure $\boldsymbol{\pi}$ and some constants $C, \lambda, t_0 > 0$ such that, for all $t \geq t_0$,

$$\mathcal{W}_{\widetilde{\mathbf{d}}}(\boldsymbol{\mu}\mathbf{P}_t, \boldsymbol{\pi}) \leq C\mathrm{e}^{-\lambda t}\mathcal{W}_{\widetilde{\mathbf{d}}}(\boldsymbol{\mu}, \boldsymbol{\pi}),$$

for every probability measure $\boldsymbol{\mu}$ on $\mathbf{E}$. In this expression, $\widetilde{\mathbf{d}}$ is defined by

$$\widetilde{\mathbf{d}}(\mathbf{x}, \mathbf{y}) = \sqrt{\big(\mathbf{1}_{i\neq j} + \mathbf{1}_{i=j}\big(1 \wedge d^q(x, y)\big)\big)\big(1 + d^q(x, x_0) + d^q(y, x_0)\big)},$$

where $\mathbf{x} = (x, i)$, $\mathbf{y} = (y, j)$ belong to $\mathbf{E}$, x_0 is as in Assumption 1.3 and $q \in (0, 1]$. Noting that $\mathbf{d} \leq \widetilde{\mathbf{d}}$ we conclude that for $t \geq t_0$ one has

$$\mathcal{W}_{\mathbf{d}}(\delta_{\mathbf{y}_0}\mathbf{P}_t, \boldsymbol{\pi}) \leq C\mathrm{e}^{-\lambda t}\left(1 + \sum_{i\in F}\int_E d(y_0, x)^q \boldsymbol{\pi}(\mathrm{d}x, i)\right).$$

Since furthermore

$$\mathcal{W}_{\mathbf{d}}(\delta_{\mathbf{y}_0}\mathbf{P}_t, \boldsymbol{\pi}) \leq 1,$$

for all $t \leq t_0$, this ends the proof.

4. Two special cases

Here, we give some sufficient conditions allowing to verify our main assumptions in situations where the underlying processes are deterministic or diffusive. Note that we can find sufficient conditions in [9] for stochastically monotone processes, in [8] for birth–death processes and in [15] for diffusion processes.

4.1. The case of diffusion processes

Let us recall that a diffusion process on $\mathbb{R}^d$, $d \in \mathbb{N}^*$, is a process generated by

$$\forall x \in \mathbb{R}^d, \qquad \mathcal{L}f(x) = \sum_{i=1}^d b_i(x)\,\partial_i f(x) + \sum_{i,j=1}^d \big(\sigma(x)\sigma(x)^t\big)_{i,j}\,\partial_{i,j} f(x), \tag{4.1}$$

where f is a smooth enough function and b, σ are regular enough, say

$$\forall x, y \in \mathbb{R}^d, \qquad \big\|\sigma(x) - \sigma(y)\big\| + \big\|b(x) - b(y)\big\| \leq K\|x - y\| \tag{4.2}$$

for some $K > 0$. In the previous expression, $\|\cdot\|$ denotes both the Euclidean norm and the subordinate norm.

Lemma 4.1. *Let $(P_t)_{t\geq 0}$ be the Markov semigroup generated by (4.1). If σ is constant and*

$$\forall x, y \in \mathbb{R}^d, \qquad \big\langle b(x) - b(y), x - y\big\rangle \leq -\alpha\|x - y\|^2, \tag{4.3}$$

for some $\alpha \in \mathbb{R}$, *then*

$$\forall t \geq 0, \qquad \mathcal{W}_{\|\cdot\|}(\mu P_t, \nu P_t) \leq \mathrm{e}^{-\alpha t} \mathcal{W}_{\|\cdot\|}(\mu, \nu),$$

for any probability measures μ *and* ν.

Proof. This is an immediate consequence of (4.3). One can see this by using the same Brownian motion for two different solutions of the SDE starting with different initial measures. □

If σ is not constant, then one can also use [2], Proposition 6.1, which essentially requires the Lipschitz constant of σ to be sufficiently small compared to the rate of contraction α, see also [15,34]. The "small level sets" assumption of Theorem 1.7 or Theorem 1.8 is satisfied if one of the underlying diffusions verifies Hörmander's hypoellipticity assumption and satisfies furthermore a natural controllability assumption. See, for instance, [20] for an introduction on this subject.

Remark 4.2 (Exponential convergence for an infinite dimensional process). The previous result gives also the convergence for switching Fokker–Planck processes. Indeed, we can consider that each underlying Markov process $(Z_t^{(i)})_{t\geq 0}$ is deterministic, belongs to the space of smooth density functions, and verifies

$$\partial_t Z_t^{(i)}(x) = \sum_{k=1}^{d} -\partial_k \big(b_k Z_t^{(i)}\big)(x) + \sum_{k,l=1}^{d} \partial_{k,l} \big(\sigma_{k,l} Z_t^{(i)}\big)(x)$$

for all $x \in \mathbb{R}^d$, and $t \geq 0$. The previous lemma gives a contraction as in Assumption 1.3, for each underlying process, where d is the Wasserstein metric.

4.2. Case of piecewise deterministic Markov processes

Let us assume that each one of the underlying Markov processes is actually deterministic. More precisely, we consider that E is an open of $\mathbb{R}^d$, $d \in \mathbb{N}^*$ and $\mathcal{L}^{(i)} f = G^{(i)} \cdot \nabla f$, for every $i \in F$, where $(G^{(i)})_{i\in F}$ is a family of vector fields such that the ordinary differential equations $x' = G^{(i)}(x)$ have a unique and global solution for any initial condition, for every $i \in F$. Lemma 4.1 gives the assumption in order to apply Theorem 1.4 and Theorem 1.5. In general, we cannot apply Theorem 1.7 or Theorem 3.4 but [1,4] give a sufficient condition ensuring that $\mathbf{X}$ generates densities:

Assumption 4.3 (Hörmander-type bracket conditions). *Let* $\mathcal{G}_0 = \{G^{(i)} - G^{(j)}, i \neq j\}$ *and for all* $k \geq 0$,

$$\mathcal{G}_{k+1} = \big\{\big[G^{(i)}, G\big] | i \in F, G \in \mathcal{G}_k\big\},$$

where $[\cdot,\cdot]$ *designs the Lie bracket. We have* $\mathcal{G}_k(x) = \{G(x) | G \in \mathcal{G}_k\} = \mathbb{R}^d$, *for every* $x \in E$.

In this case our main result gives the following theorem.

Theorem 4.4. *Let us suppose that Assumptions 1.1, 1.2 and 4.3 hold. If one of the two following assumptions is satisfied*:

- $a(x,i,j)$ *does not depend to* x *and* I *is ergodic with an invariant measure* ν *satisfying*

$$\sum_{i\in F}\nu(i)\lambda(i) > 0;$$

- *Assumption 3.1 holds and*

$$\sum_{i\in F}\nu(i)\alpha(i) > 0,$$

  for some increasing sequence α *satisfying* $\alpha(n) \le \min_{i\in F_n}\lambda(i)$, *for all* $n \le \bar{n}$

then there exist a probability measure $\boldsymbol{\pi}$ *and three constants* $C, \lambda, t_0 > 0$ *such that*

$$\forall t \ge t_0, \qquad d_{\mathrm{TV}}(\delta_{\mathbf{x}}\mathbf{P}_t, \boldsymbol{\pi}) \le C\mathrm{e}^{-\lambda t}\big(1 + V(x)\big),$$

for every $\mathbf{x} = (x,i) \in \mathbf{E}$.

Proof. Using [4], Theorem 6.6, we see that compact sets are small for $\mathbf{X}$. Using Lemma 2.8 in the first case and Lemma 3.9 in the second case, we see that we can apply Theorem 2.11. $\square$

5. Examples

Here, we give three simple examples to illustrate our results.

5.1. The most elementary example

Let us consider the example where X belongs to $\mathbb{R}$ and verifies

$$\forall t \ge 0, \qquad \partial_t X_t = I_t X_t,$$

where $(I_t)_{t\ge 0}$ is the continuous time Markov chain, on $\{-1, 1\}$, which jumps from 1 to -1 with rate $a_1 > 0$ and from -1 to 1 with rate $a_{-1} > 0$. If $a_1 > a_{-1}$ then Theorems 1.4 and 1.5 give the exponential ergodicity of $\mathbf{X}$ in the Wasserstein distance. Here, the invariant law is

$$\delta_0 \otimes \frac{1}{a_{-1} + a_1}(a_1\delta_{-1} + a_{-1}\delta_1),$$

and there is clearly no convergence in total variation. Thus, classical Harris' theorem does not work here. Furthermore, the classical law of large number gives

$$\lim_{t\to+\infty} X_t = \begin{cases} 0 & \text{a.s., if } a_1 > a_{-1}, \\ +\infty & \text{a.s., if } a_1 < a_{-1}. \end{cases}$$

In particular, there is no convergence when $a_1 < a_{-1}$.

Remark 5.1. In our main theorems, we use a Wasserstein distance associated to a distance comparable to d^q rather than d. We choose this distance because, in general, moments of $\mathbf{X}$ can explode even though $\mathbf{X}$ converges in law. For instance, in the above example, one has $\lim_{t\to\infty} \mathbb{E}X_t = \infty$ as soon as $a_1 < 1$. See also [3] for comments on the optimal choice of the parameter q.

5.2. Wasserstein contraction of some switching dynamical systems

Let us consider a slight generalisation of the previous example; that is X belongs to $\mathbb{R}$ and verifies

$$\forall t \geq 0, \qquad \partial_t X_t = -a(I_t)X_t, \tag{5.1}$$

where $(I_t)_{t\geq 0}$ is a recurrent continuous time Markov chain on a finite state space F and a a function from F to $\mathbb{R}$. Theorem 1.4 gives the exponential–Wasserstein ergodicity under the condition that

$$\sum_{i\in F} a(i)\nu(i) > 0, \tag{5.2}$$

where ν is a invariant measure of I. This simple example satisfies a bound like in Assumption 1.3. Indeed we have the following lemma.

Lemma 5.2. *If* (5.1) *and* (5.2) *are satisfied then there is a distance* $\boldsymbol{\delta}$ *on* $\mathbf{E}$ *such that the Wasserstein curvature of the semigroup of* $\mathbf{X}$ *is positive*, *that is*, *there exists* $\lambda > 0$ *such that*

$$\forall t \geq 0, \qquad \mathcal{W}_{\boldsymbol{\delta}}(\delta_{\mathbf{x}}\mathbf{P}_t, \delta_{\mathbf{y}}\mathbf{P}_t) \leq \mathrm{e}^{-\lambda t}\boldsymbol{\delta}(\mathbf{x},\mathbf{y}),$$

for all $\mathbf{x}, \mathbf{y} \in \mathbf{E}$.

Proof. First, let us give a complement on the conclusion of Lemma 2.7. The Markov chain I satisfies its assumptions and using the results of [3], there exist a function ψ on F, $\rho > 0$ and $p \in (0, 1)$ verifying

$$\forall t \geq 0, \qquad \mathbb{E}\big[\psi(I_t)\mathrm{e}^{-\int_0^t pa(I_s)\,\mathrm{d}s}\big] = \mathrm{e}^{-\rho t}\mathbb{E}\big[\psi(I_0)\big].$$

Now let $\boldsymbol{\delta}$ be the distance, on $\mathbf{E}$, defined by

$$\forall \mathbf{x}, \mathbf{y} \in \mathbf{E}, \qquad \boldsymbol{\delta}(\mathbf{x},\mathbf{y}) = \mathbf{1}_{\{i=j\}}\psi(i)|x-y|^p + \mathbf{1}_{\{i\neq j\}}\frac{\overline{\psi}}{\underline{\psi}}\big(\psi(i)|x|^p + \psi(j)|y|^p + 1\big),$$

where

$$\overline{\psi} = \max_{k\in F}\psi(k) \quad \text{and} \quad \underline{\psi} = \min_{k\in F}\psi(k).$$

Now, using the fact that for all $t > 0$ one has

$$X_t = X_0 \mathrm{e}^{-\int_0^t a(I_s)\,\mathrm{d}s},$$

the proof is straightforward. □

5.3. Surprising blow-up under exponential ergodicity assumptions

Here we give some comments on [6], Example 1.4, which also illustrate the sharpness of our criteria. Let us consider $E = \mathbb{R}^2$, $F = \{0, 1\}$, $\mathcal{L}^{(i)} f = A_i \cdot \nabla f$ where

$$A_0 = \begin{pmatrix} -1 & 3 \\ -1/3 & -1 \end{pmatrix} \quad \text{and} \quad A_1 = \begin{pmatrix} -1 & -1/3 \\ 3 & -1 \end{pmatrix},$$

$a(x, 0, 0) = a(x, 1, 1) = 0$, and $a(x, 1, 0) = a(x, 0, 1) = a > 0$, for all $x \in \mathbb{R}^2$. In short, $\mathbf{X}$ is generated, for all $x \in \mathbb{R}^2$ and $i \in \{0, 1\}$, by

$$\mathbf{L} f(x, i) = A_i \cdot \nabla f(x, i) + a\big(f(x, 1 - i) - f(x, i)\big). \tag{5.3}$$

Since a does not depend on its first component, I is a Markov process and it converges exponentially to

$$\nu = \tfrac{1}{2}\delta_0 + \tfrac{1}{2}\delta_1.$$

For each $i \in \{0, 1\}$, we have $\partial_t Z_t^{(i)} = A_i Z_t^{(i)}$ and thus we easily prove that

$$\big\|Z_t^{(i)}\big\|_i \le \mathrm{e}^{-t}\big\|Z_0^{(i)}\big\|_i \quad \text{and} \quad \big\|Z_t^{(i)}\big\|_{1-i} \le 3\mathrm{e}^{-t}\big\|Z_0^{(i)}\big\|_{1-i}, \tag{5.4}$$

for every $t \ge 0$, where the norms $\|\cdot\|_0$ and $\|\cdot\|_1$ are defined by

$$\forall u = (u_1, u_2) \in \mathbb{R}^2, \qquad \|u\|_0 = \sqrt{(u_1/3)^2 + u_2^2} \quad \text{and} \quad \|u\|_1 = \sqrt{u_1^2 + (u_2/3)^2}.$$

Thus each flow $i \in \{0, 1\}$ contracts, with the norm $\|\cdot\|_i$, and converges geometrically, with the norm $\|\cdot\|_{1-i}$, to the same limit. Nevertheless, if a is large enough then [6], Example 1.4, shows that

$$\lim_{t\to+\infty} \|X_t\| = +\infty.$$

In particular, the conclusion of Theorem 1.4 is not satisfied. This illustrates the fact that assuming that each underlying dynamics converges geometrically is not sufficient in general to guarantee the convergence of X. Moreover, this shows that it is essential in Theorem 1.4 to measure the constants $\rho(i)$ with respect to the *same* distance for every i. Note that the Wasserstein curvature of $Z^{(i)}$, with respect to $\|\cdot\|_{1-i}$, is negative and given by $-37/3$.

5.4. Non-convergence when I is recurrent but not positive recurrent

A last example is the following: the process X verifies

$$\forall t \geq 0, \qquad \mathrm{d}X_t = -(X_t - a_{I_t})\,\mathrm{d}t,$$

where $(a_n)_{n\geq 0}$ is a bounded real sequence and I is an irreducible and recurrent continuous time Markov chain which is not positive recurrent. It is easy to see that the sequence of laws of $(X_t)_{t\geq 0}$ is tight and we can hope that there exists a probability measure π verifying

$$\lim_{t\to+\infty} \mathbb{E}\big[f(X_t)\big] = \int f\,\mathrm{d}\pi,$$

for every continuous and bounded function f and any starting distribution. But in general, this is false. To illustrate it, let us consider the case when I is the classical continuous-time random walk on $\mathbb{N}$ reflected at 0. Namely, I is generated by

$$Jf(i) = \tfrac{1}{2}f(i+1) + \tfrac{1}{2}f(i-1) - f(i),$$

if $i \neq 0$ and

$$Jf(0) = f(1) - f(0).$$

The sequence a on the other hand is defined recursively by:

$$a_{n+1} = \begin{cases} a_n & \text{if } n \notin \{2^k | k \in \mathbb{N}\}, \\ -a_n & \text{if } n \in \{2^k | k \in \mathbb{N}\}. \end{cases}$$

In this case, the central limit theorem gives that $I_t \approx \sqrt{t}$ and so, for very large times, I and a do not switch on the same time scale. As a matter of fact, the process a_{I_t} stays constant during longer and longer stretches of time. It is then possible to find two sequences of *deterministic* times $(t_n)_{n\geq 0}$ and $(s_n)_{n\geq 0}$, both converging to infinity, and such that

$$\lim_{n\to+\infty} \mathbb{E}\big[f(X_{t_n})\big] = f(0) \quad \text{and} \quad \lim_{n\to+\infty} \mathbb{E}\big[f(X_{s_n})\big] = f(1).$$

Thus this process exhibits ageing and is not exponentially stable, even though there exists $C > 0$, such that for any two starting points $\mathbf{x} = (x, i)$ and $\mathbf{y} = (y, j)$, we have

$$\forall t \geq 0, \qquad \mathcal{W}_{\mathbf{d}_0}(\delta_{\mathbf{x}}\mathbf{P}_t, \delta_{\mathbf{y}}\mathbf{P}_t) \leq \frac{C}{\sqrt{t}}|i - j|,$$

where $\mathbf{d}_0(\mathbf{x}, \mathbf{y}) = \mathbf{1}_{i=j}\|x - y\| \wedge 1 + \mathbf{1}_{i\neq j}$.

Acknowledgements

We would like to thank the referees for their careful reading of the manuscript. Support for MH's research was provided by the Royal Society through a Wolfson Research Merit Award. BC was partially supported by Ecole Doctorale MSTIC, Université Paris-Est, and by ANR MANEGE (09-BLAN-0215).

References

[1] Bakhtin, Y. and Hurth, T. (2012). Invariant densities for dynamical systems with random switching. *Nonlinearity* **25** 2937–2952. MR2979976

[2] Bardet, J.-B., Christen, A., Guillin, A., Malrieu, F. and Zitt, P.-A. (2013). Total variation estimates for the TCP process. *Electron. J. Probab.* **18** no. 10, 21. MR3035738

[3] Bardet, J.-B., Guérin, H. and Malrieu, F. (2010). Long time behavior of diffusions with Markov switching. *ALEA Lat. Am. J. Probab. Math. Stat.* **7** 151–170. MR2653702

[4] Benaïm, M., Le Borgne, S., Malrieu, F. and Zitt, P.-A. (2012). Qualitative properties of certain piecewise deterministic Markov processes. ArXiv e-prints.

[5] Benaïm, M., Le Borgne, S., Malrieu, F. and Zitt, P.-A. (2012). Quantitative ergodicity for some switched dynamical systems. *Electron. Commun. Probab.* **17** no. 56, 14. MR3005729

[6] Benaïm, M., Le Borgne, S., Malrieu, F. and Zitt, P.-A. (2014). On the stability of planar randomly switched systems. *Ann. Appl. Probab.* **24** 292–311. MR3161648

[7] Boxma, O., Kaspi, H., Kella, O. and Perry, D. (2005). On/off storage systems with state-dependent input, output, and switching rates. *Probab. Engrg. Inform. Sci.* **19** 1–14. MR2104547

[8] Chafaï, D. and Joulin, A. (2013). Intertwining and commutation relations for birth–death processes. *Bernoulli* **19** 1855–1879. MR3129037

[9] Cloez, B. (2012). Wasserstein decay of one dimensional jump-diffusions. ArXiv e-prints.

[10] Collet, P., Martínez, S., Méléard, S. and San Martín, J. (2013). Stochastic models for a chemostat and long-time behavior. *Adv. in Appl. Probab.* **45** 822–836. MR3102473

[11] Costa, O.L.V. and Dufour, F. (2008). Stability and ergodicity of piecewise deterministic Markov processes. *SIAM J. Control Optim.* **47** 1053–1077. MR2385873

[12] Crudu, A., Debussche, A., Muller, A. and Radulescu, O. (2012). Convergence of stochastic gene networks to hybrid piecewise deterministic processes. *Ann. Appl. Probab.* **22** 1822–1859. MR3025682

[13] de Saporta, B. and Yao, J.-F. (2005). Tail of a linear diffusion with Markov switching. *Ann. Appl. Probab.* **15** 992–1018. MR2114998

[14] Diaconis, P. and Freedman, D. (1999). Iterated random functions. *SIAM Rev.* **41** 45–76. MR1669737

[15] Eberle, A. (2011). Reflection coupling and Wasserstein contractivity without convexity. *C. R. Math. Acad. Sci. Paris* **349** 1101–1104. MR2843007

[16] Fontbona, J., Guérin, H. and Malrieu, F. (2012). Quantitative estimates for the long-time behavior of an ergodic variant of the telegraph process. *Adv. in Appl. Probab.* **44** 977–994. MR3052846

[17] Genadot, A. and Thieullen, M. (2012). Averaging for a fully coupled piecewise-deterministic Markov process in infinite dimensions. *Adv. in Appl. Probab.* **44** 749–773. MR3024608

[18] Goldie, C.M. and Grübel, R. (1996). Perpetuities with thin tails. *Adv. in Appl. Probab.* **28** 463–480. MR1387886

[19] Guyon, X., Iovleff, S. and Yao, J.-F. (2004). Linear diffusion with stationary switching regime. *ESAIM Probab. Stat.* **8** 25–35. MR2085603

[20] Hairer, M. (2011). On Malliavin's proof of Hörmander's theorem. *Bull. Sci. Math.* **135** 650–666. MR2838095

[21] Hairer, M. and Majda, A.J. (2010). A simple framework to justify linear response theory. *Nonlinearity* **23** 909–922. MR2602020

[22] Hairer, M. and Mattingly, J.C. (2008). Spectral gaps in Wasserstein distances and the 2D stochastic Navier-Stokes equations. *Ann. Probab.* **36** 2050–2091. MR2478676

[23] Hairer, M. and Mattingly, J.C. (2011). Yet another look at Harris' ergodic theorem for Markov chains. In *Seminar on Stochastic Analysis*, *Random Fields and Applications VI*. *Progress in Probability* **63** 109–117. Basel: Birkhäuser. MR2857021

[24] Hairer, M., Mattingly, J.C. and Scheutzow, M. (2011). Asymptotic coupling and a general form of Harris' theorem with applications to stochastic delay equations. *Probab*. *Theory Related Fields* **149** 223–259. MR2773030

[25] Hairer, M., Stuart, A. and Vollmer, S. Spectral gap for a Metropolis–Hasting algorithm in infinite dimensions. Arxiv e-print.

[26] Joulin, A. (2007). Poisson-type deviation inequalities for curved continuous-time Markov chains. *Bernoulli* **13** 782–798. MR2348750

[27] Joulin, A. (2009). A new Poisson-type deviation inequality for Markov jump processes with positive Wasserstein curvature. *Bernoulli* **15** 532–549. MR2543873

[28] Meyn, S.P. and Tweedie, R.L. (1993). Stability of Markovian processes. III. Foster-Lyapunov criteria for continuous-time processes. *Adv. in Appl*. *Probab*. **25** 518–548. MR1234295

[29] Ollivier, Y. (2009). Ricci curvature of Markov chains on metric spaces. *J*. *Funct*. *Anal*. **256** 810–864. MR2484937

[30] Ollivier, Y. (2010). A survey of Ricci curvature for metric spaces and Markov chains. In *Probabilistic Approach to Geometry*. *Adv*. *Stud*. *Pure Math*. **57** 343–381. Tokyo: Math. Soc. Japan. MR2648269

[31] Pakdaman, K., Thieullen, M. and Wainrib, G. (2012). Asymptotic expansion and central limit theorem for multiscale piecewise-deterministic Markov processes. *Stochastic Process*. *Appl*. **122** 2292–2318. MR2922629

[32] Smith, W.L. (1968). Necessary conditions for almost sure extinction of a branching process with random environment. *Ann*. *Math*. *Statist* **39** 2136–2140. MR0237006

[33] Villani, C. (2009). *Optimal Transport*, *Old and New*. *Grundlehren der Mathematischen Wissenschaften* [*Fundamental Principles of Mathematical Sciences*] **338**. Berlin: Springer. MR2459454

[34] von Renesse, M.-K. and Sturm, K.-T. (2005). Transport inequalities, gradient estimates, entropy, and Ricci curvature. *Comm*. *Pure Appl*. *Math*. **58** 923–940. MR2142879

Received March 2013 and revised September 2013