Exponential ergodicity for Markov processes with random switching

March 27, 2013

Bertrand Cloez[1], **Martin Hairer**[2]

[1] LAMA, Université Paris-Est, France
Email: bertrand.cloez@univ-mlv.fr
[2] Mathematics Department, University of Warwick, U.K.
Email: M.Hairer@Warwick.ac.uk

Abstract

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.

Contents

1 Introduction

Markov processes with switching are intensively used for modelling purposes in applied subjects like biology [CMMS12, CDMR12, FGM12], storage modelling [BKKP05], neuronal activity [PTW12, GT11]. This class of Markov processes is reminiscent of the so-called iterated random functions [DF99] or branching processes in random environment [Smi68] in the discrete time setting. Several recent works [BH12, BGM10, BLMZ12b, BBMZ12, CD08, dSY05, GIY04, GG96] deal with their long time behaviour (existence of an invariant probability measure, Harris recurrence, exponential ergodicity, hypoellipticity...). In particular, in [BH12, BLMZ12b], 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 [BBMZ12]. Let (X, I) be the Markov process on $\mathbb{R}^2 \times \{-1, 1\}$ generated by

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

This process is ergodic and the first marginal π of its invariant measure is supported on $\mathbb{R} \times \{0\}$. Thus, in general, it does not converge in the total variation distance. However, it is proved in [BBMZ12] that it converges in a certain Wasserstein distance. Let us recall that the Wasserstein distance on a Polish space (E, d) is defined by

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

for every probability measure μ_1, μ_2 on E, where the infimum is taken over all pairs of random variables X_1, X_2 with respective laws μ_1, μ_2. The Kantorovich-Rubinstein duality [Vil09, Theorem 5.10] shows that one also has

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

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

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

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

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

where the infimum is again taken over all random variables X_1, X_2 with respective distributions 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

$$\forall j \in F,\ \mathbb{P}\left(T_j > t \mid \mathcal{F}_t\right) = \exp\left(-\int_0^t a\left(Z_s^{(i)}, i, j\right) ds\right),$$

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

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

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:

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

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 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*

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

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

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

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 [Jou07, Jou09] and the coarse Ricci curvature in [Oll09, Oll10], since it is heavily related to the geometry of the underlying space as illustrated in [vRS05, Theorem 2]. If $\rho(i) > 0$, then we can deduce some properties like geometric ergodicity, a Poincaré inequality or some concentration inequalities [Clo12, Jou07, Jou09, HSV, Oll10]. 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 $\mathbb{E}d(X_t, Y_t) \leq e^{-\rho t} d(X_0, Y_0)$.

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 now have presented all the assumptions that are necessary 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 $\boldsymbol{\pi}$ and some constants $C, \lambda, t_0 > 0$ such that

$$\forall t \geq t_0, \ \mathcal{W}_{\mathbf{d}}(\delta_{\mathbf{y}_0} \mathbf{P}_t, \boldsymbol{\pi}) \leq C e^{-\lambda t} (1 + \mathcal{W}_{d^q}(\delta_{y_0}, \boldsymbol{\pi})),$$

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

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

for every $\mathbf{x} = (x, i)$*,* $\mathbf{y} = (y, j)$ *belonging to* $\mathbf{E}$*,* x_0 *is as in Assumption 1.3 and* $q \in (0, 1]$.

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 [BBMZ12, 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 p^{th} Wasserstein distance $\mathcal{W}^{(p)}$ with $p \geq 1$ provided that $\mathbf{X}$ satisfies a moment condition. We give the previous theorem 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. That is

Theorem 1.5 (Wasserstein exponential ergodicity with an on-off type criterion)**.** *Let us suppose that Assumptions 1.1, 1.2, and 1.3 hold. We set*

$$F_0 = \{i \in F \mid \rho(i) > 0\} \;\; \text{and} \;\; F_1 = \{i \in F \mid \rho(i) \leq 0\},$$

$$\rho_0 = \min_{i \in F_0} \rho(i) > 0 \;\; \text{and} \;\; \rho_1 = \min_{i \in F_1} \rho(i) \leq 0,$$

$$a_0 = \max_{i \in F_0} \sup_{x \in E} \sum_{j \in F_1} a(x, i, j) \;\; \text{and} \;\; a_1 = \min_{i \in F_1} \inf_{x \in E} \sum_{j \in F_0} a(x, i, j).$$

If

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

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

$$\forall t \geq t_0, \; \mathcal{W}_{\mathbf{d}}\left(\delta_{\mathbf{y}_0} \mathbf{P}_t, \boldsymbol{\pi}\right) \leq C e^{-\lambda t}(1 + \mathcal{W}_{d^q}(\delta_{y_0}, \boldsymbol{\pi})),$$

for every $\mathbf{y}_0 = (y_0, j_0) \in \mathbf{E}$*, where the distance* $\mathbf{d}$*, on* $\mathbf{E}$*, is defined in* (1.4)*,* x_0 *is as in Assumption 1.3 and* $q \in (0, 1]$.

With this result, we not only recover [BBMZ12, 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 3.2 below. These two theorems are our main result and, contrary to the previous theorem, it seems that they cannot be deduced directly from the approach of [BBMZ12].

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 1.5 and Theorem 1.6, 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*

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

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

Theorem 1.7 (Exponential ergodicity in the constant case)**.** *Suppose that assumptions 1.1, 1.2 and 1.7 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 $i_0 \in F$ such that the sublevel sets of V are small for $P_t^{(i_0)}$, then there exist a probability measure $\boldsymbol{\pi}$ and two constants $C, \lambda > 0$ such that

$$d_{\mathrm{TV}}\left(\delta_{\mathbf{x}}\mathbf{P}_t, \boldsymbol{\pi}\right) \le Ce^{-\lambda t}(1+V(x)),$$

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

The definition of a small set is recalled in Definition 2.9. We give also the analogous of Theorem 1.6:

Theorem 1.8 (Exponential ergodicity with an on-off type criterion)**.** *Let us suppose that Assumptions 1.1, 1.2, 1.3 hold. We set*

$$F_0 = \{i \in F \mid \lambda(i) > 0\} \ \textit{and}\ \ F_1 = \{i \in F \mid \lambda(i) \le 0\},$$

$$\lambda_0 = \min_{i\in F_0}\lambda(i) > 0 \ \textit{and}\ \ \lambda_1 = \min_{i\in F_1}\lambda(i) \le 0,$$

$$a_0 = \max_{i\in F_0}\sup_{x\in E}\sum_{j\in F_1} a(x,i,j) \ \textit{and}\ \ a_1 = \min_{i\in F_1}\inf_{x\in E}\sum_{j\in F_0} a(x,i,j).$$

If

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

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

$$d_{\mathrm{TV}}\left(\delta_{\mathbf{x}}\mathbf{P}_t, \boldsymbol{\pi}\right) \le Ce^{-\lambda t}(1+V(x)),$$

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 [BH12, BLMZ12b], 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 1.5 and Theorem 1.8. 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 1.6. 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 1.7, the process **X** cannot wander off to infinity, i.e. its semigroup possesses a Lyapunov function. We then prove Theorems 1.5 and 1.8 using a similar argument to [BBMZ12] 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*

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

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

A well know 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 **X** may not verify the drift condition (2.2) but, in Lemmas 2.7 and 3.8, 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 1.7. Then, we will prove that Assumption 1.7 implies the existence of a Lyapunov function for **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*

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

for every $x, y \in X$ and $t \geq 0$. If there exist $x_0 \in E$ 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

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

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

Proof. For any $t \geq t_{x_0}$ and $n \geq 0$, it follows from (2.3) that

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

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

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

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

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

thus concluding the proof. □

We deduce that Assumption 1.3 implies Assumption 1.7 with $V = V_{x_0}$ and $\lambda = \rho$.

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*

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

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

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

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

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

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 $\lambda = 0$)**.** *In the previous lemma, we have supposed that $\lambda \neq 0$, and this assumption is necessary for our conclusion to hold. Indeed, if $(B_t)_{t \geq 0}$ is a Brownian motion then*

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

and inequality (2.5) *does not hold.*

We now show that if Assumption 1.7 holds and the mean of $(\lambda(i))_{i \in F}$ is positive, then $\mathbf{X}$ admits a Lyapunov function. As in [BBMZ12], this result comes from the following lemma:

Lemma 2.6. *Let (K_t) 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 $\alpha \colon S \to \mathbb{R}$ 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

$$c e^{-\eta t} \leq \mathbb{E}\left[e^{-\int_0^t p\alpha(K_s) ds} \right] \leq C e^{-\eta t}.$$

Proof. It is a consequence of Perron-Frobenius Theorem and the study of eigenvalues. See [BGM10, Proposition 4.1] and [BGM10, Proposition 4.2] for further details. □

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

Lemma 2.7. *Under Assumption 1.1, 1.2 and 1.7, 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

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

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):

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

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

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

where $(T_k)_{k \geq 0}$ is the sequence of jump times of I, with $T_0 = 0$, and N_t the number of jumps before t. By Lemma 2.6, there exist $C > 0, \eta > 0$ and $q \in (0, 1]$ such that

$$\mathbb{E}\left[e^{-\int_0^t q\lambda(I_s)ds}\right] \leq Ce^{-\eta t}.$$

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

$$K_V = K \sum_{n \geq 0} \mathbb{E}\left[e^{-q\int_0^{T_n} \lambda(I_s)ds}\right] \lesssim K \sum_{n \geq 0} e^{-\varepsilon n} < +\infty,$$

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 3.8. This concludes the proof. □

Remark 2.8 (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*

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

for the argument to go through.

2.2 Proof of Theorem 1.5

This section is split into two parts. We begin by introducing our coupling construction, and we then proceed to prove Theorem 1.5. In both parts, we make the standing assumption that the hypotheses of Theorem 1.5 hold. In particular, I is an ergodic Markov chain.

2.2.1 Our coupling

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

$$\forall t \geq 0,\ \mathbb{E}\left[\widetilde{\mathbf{d}}(\mathbf{X}_t, \mathbf{Y}_t)\right] \leq Ce^{-\alpha t}\widetilde{\mathbf{d}}(\mathbf{x}, \mathbf{y}),$$

for some $C, \alpha > 0$. Here the "distance function" $\widetilde{\mathbf{d}}$ is defined by

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

Here, we put "distance function" in quotation marks since $\widetilde{\mathbf{d}}$ does not in general satisfy the triangle inequality. Obviously, we have

$$\mathbf{d}(\mathbf{x}, \mathbf{y}) \leq \widetilde{\mathbf{d}}(\mathbf{x}, \mathbf{y}) \leq \mathbf{1}_{\mathbf{x}\neq\mathbf{y}}\sqrt{1 + d(x, x_0)^q + d(y, x_0)^q}.$$

Remark that it is well-known that if I and J are two independent processes with transition rate a then there exists $\theta_c > 0$ such that

$$\forall t \geq 0,\ \mathbb{P}\left(T_c > t\right) \leq e^{-\theta_c t},$$

where $T_c = \inf\{t \geq 0 \mid I_t = J_t\}$ is their first meeting time. From now on, we fix the starting points of our coupling $\mathbf{x} = (x, i)$, $\mathbf{y} = (y, j)$ and the time $t \geq 0$. 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 coupled as follow:

- if $i \neq j$ then we consider that $\mathbf{X}$ and $\mathbf{Y}$ evolve independently of each other until the first meeting time T_c.
- for all $s \geq T_c$, we set $I_s = J_s$ and we couple X and Y in such a way that

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

  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 $(\mathbf{X}, \mathbf{Y})$.

Note that if $i = j$ then $T_c = 0$.

Proof of Theorem 1.5. By Lemma 2.2, it suffices to show that

$$\mathbb{E}\left[\mathbf{d}(\mathbf{X}_t, \mathbf{Y}_t)\right] \leq Ce^{-\lambda t}(1 + d^q(x_0, y_0))\ .$$

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

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

where $q \in (0,1]$. By Lemma 2.6, there exist $C, \eta > 0$ and $q \in (0,1]$ such that

$$\mathbb{E}\left[d(X_t, Y_t)^q\right] \leq Ce^{-\eta t} d(x,y)^q.$$

Now, for general i and j, we have

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

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

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

In the other hand, one has the bound

$$\begin{aligned}&\mathbb{E}\left[\sqrt{\mathbf{1}_{T_c \leq t/2} d(X_t, Y_t)^q (1 + V^q(X_t) + V^q(Y_t))}\right] \\ &\qquad \leq \mathbb{E}\left[\mathbf{1}_{T_c \leq t/2} d(X_t, Y_t)^q\right]^{1/2} \mathbb{E}\left[1 + V^q(X_t) + V^q(Y_t)\right]^{1/2}. \end{aligned} \tag{2.6}$$

As a consequence of Lemmas 2.2 and 2.7, we also have the bound

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

Assembling these inequalities and using again Lemma 2.7 to bound the second factor in (2.6), the claim follows. □

2.3 Proof of Theorem 1.8

We divide again the proof in two parts. First, we recall some tools on Harris' Theorem. Second, we give the proof of Theorem 1.8.

2.3.1 Classical Harris' Theorem

Here, we recall a version of Harris' Theorem (also called Foster, Lyapunov, Meyn-Tweedie, Doeblin) 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 2.9. *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 [HM11, MT93] then states that

Theorem 2.10 (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 $\{x \in X \mid V(x) \leq C\}$ are small for every $C > 0$. Then $(P_t)_{t\geq 0}$ has a unique invariant measure π and*

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

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.

Proof of Theorem 1.8. By Lemma 2.7, **P** admits V as Lyapunov function so, by Harris' Theorem, it only remains to show that $\{V \leq C\}$ is small for **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 2.7) such that

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

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

$$\mathbb{P}\left(V(X_t) \leq 2K\right) \geq \frac{1}{2} \ ,$$

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

$$\mathbb{P}\left(Z_t^{i_0,x} = Z_t^{i_0,y}\right) \geq \varepsilon \ , \qquad t \geq t_0 \ , \tag{2.7}$$

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
$$B = \{V(X_{t_*}) \leq 2K, V(Y_{t_*}) \leq 2K, I_s = J_s = i_0, \forall s \in [t_*, t_* + t_0]\},$$
the processes X and Y are coupled in such a way to verify (2.7), over $t \in [t_*, t_* + t_0]$;
- conditionally on B^c, they are coupled independently from each other.

The Markov property gives

$$\mathbb{P}(V(X_{t_*}) \leq 2K, I_s = i_0, \forall s \in [t_*, t_* + t_0]) \geq \frac{\delta}{2} \ ,$$

and so $\mathbb{P}(B) \geq \delta^2/4$. Combining this inequality with (2.7), we conclude that $\mathbb{P}(\mathbf{X}_{t_*+t_0} = \mathbf{Y}_{t_*+t_0}) \geq \delta^2\varepsilon/4$, uniformly over all initial conditions **x** and **y** with $V(x) \leq C$ and $V(y) \leq C$, 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 2.6 to show that **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 $(\rho(I_t))_{t\geq 0}$ or $(\lambda(I_t))_{t\geq 0}$. More precisely, we will assume

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

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

Let $(L_t)_{t\geq 0}$ be the continuous time Markov chain with generator

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

for every $n \leq \bar{n}$, where

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

and

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

if $n \neq 0$ and $d(0) = 0$. This process is irreducible, non-explosive and positive recurrent with invariant measure ν.

If this assumption holds then, for every $i \in F$, we denote by n_i the only $n \leq \bar{n}$ verifying $i \in F_n$. Let us recall that ν is defined, for every $n < \bar{n}$, by

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

where

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

Now we can state two slight generalisations of Theorem 1.6 and 1.9. The first one is

Theorem 3.2 (Wasserstein exponential ergodicity)**.** *Let us 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 $(\alpha(n))_{n\geq 0}$ is an increasing sequence verifying $\alpha(n) \leq \inf_{i\in F_n} \rho(i)$, then there exist a probability measure $\boldsymbol{\pi}$ and some constants $C, \lambda, t_0 > 0$ and $q \in (0,1]$ such that

$$\forall t \geq t_0,\ \mathcal{W}_{\mathbf{d}}\left(\delta_{\mathbf{y}_0}\mathbf{P}_t, \boldsymbol{\pi}\right) \leq Ce^{-\lambda t}(1 + \mathcal{W}_{d^q}(\delta_{y_0}, \boldsymbol{\pi})),$$

for every $\mathbf{y}_0 = (y_0, j_0) \in \mathbf{E}$, where the distance $\mathbf{d}$, on $\mathbf{E}$, is defined in (1.4), x_0 is as in Assumption 1.3.

If Assumption 3.1 holds with $\bar{n} = 0$ then all contraction parameters are positive and we recover [BBMZ12, 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 1.9:

Theorem 3.3 (Exponential ergodicity)**.** *Let us suppose that Assumptions 1.1, 1.2, 1.3 and 3.1 hold and there exists $i_0 \in F$ such that the sublevel sets of V are small for $P_t^{(i_0)}$. If*

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

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

$$d_{\mathrm{TV}}\left(\delta_{\mathbf{x}}\mathbf{P}_t, \boldsymbol{\pi}\right) \leq Ce^{-\lambda t}(1 + V(x))$$

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

We do not give the proofs of Theorem 1.9 and Theorem 3.3, as their proofs are very similar to the proof of Theorem 1.8, combined with the argument of Lemma 3.8 below. To prove Theorem 3.2 however, we cannot use classical Harris' Theorem. Its proof follows the same idea as the proof of Theorem 1.5, 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 [HM08], stochastic delay differential equations [HMS11], and linear response theory [HM10]. 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 [HMS11], which takes into account the topology of the underlying space X.

Definition 3.4 (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 [HMS11]. 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 3.5 (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 [HMS11, Theorem 4.7]:

Theorem 3.6 (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 $\{x \in X \mid V(x) \leq 4K_V\}$ 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*

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

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

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

for all of probability measures μ, ν on X.

Remark 3.7 (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 [HM10], 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 [HM10, HM08].*

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 **X** possesses a Lyapunov function:

Lemma 3.8 (Construction of a Lyapunov function). *Let us suppose that Assumptions 1.1, 1.2, 1.7 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 $C_V, K_V, \lambda > 0$ and $q \in (0,1)$ such that

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

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

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

for every $x \in E$ and $q \in (0,1]$. Now, we will describe a construction of **X** which will permit to have a better control of the jump mechanism. Let $r \geq 2\bar{a}$ and $(N_t)_{t\geq 0}$ be a Poisson process of intensity r; namely

$$N_t = \sum_{n\geq 0} \mathbf{1}_{\{\tau_n \leq t\}},$$

where $\tau_n = \sum_{k=1}^{n} E_k$ and $(E_k)_{k\geq 0}$ is a family of i.i.d. exponentially distributed random variables with mean $1/r$. We set $\tau_0 = 0$. At this stage, we do not fix the value of r, but we allow ourselves the freedom to tune it at the end of the proof. We will couple $\mathbf{X} = (X, I)$ with a process L that has generator (3.1). Let us fix $n \in \mathbb{N}$, on $[\tau_n, \tau_{n+1}]$, the process $(\mathbf{X}, L)$ is built as follow:

- conditionally on $\mathbf{X}_{\tau_n}$, $(L_s)_{s\geq 0}$, $(\tau_k)_{k\geq 0}$, the process $(X_s)_{s\in[\tau_n,\tau_{n+1}]}$ moves as $(Z^{(I_{\tau_n})}_{t-\tau_n})_{t\in[\tau_n,\tau_{n+1}]}$ starting from X_{τ_n}; more precisely,

$$\mathbb{E}\left[f(X_t)\mathbf{1}_{t\in[\tau_n,\tau_{n+1}]} \mid \mathcal{G}_n\right] = P^{(I_{\tau_n})}_{t-\tau_n} f(X_{\tau_n}),$$

  where f is a continuous and bounded function and $\mathcal{G}_n = \sigma\{\mathbf{X}_{\tau_n}, (L_s)_{s\geq 0}, (\tau_k)_{k\geq 0}\}$;
- on $[\tau_n, \tau_{n+1})$, the discrete processes I and L remain constant;
- at time τ_{n+1}, we consider a Bernouilli random variable B with parameter $1/2$ independent the previous variables and we have two situations:
  - if $n_{I_{\tau_n}} \neq L_{\tau_n}$ then
    * if $B = 0$ then I does not jump but L can jump,
    * if $B = 1$ then L does not jump but I can jump;
  - if $n_{I_{\tau_n}} = L_{\tau_n}$ then
    * if $B = 0$ then neither I nor L can jump,
    * if $B = 1$ then we have the following possibilities:
      · $L_{\tau_{n+1}} = L_{\tau_n} + 1$ and $I_{\tau_{n+1}} \in F_{n_{I_{\tau_n}}+1}$,
      · $L_{\tau_{n+1}} = L_{\tau_n}$ and $I_{\tau_{n+1}} \in F_{n_{I_{\tau_n}}} \cup F_{n_{I_{\tau_n}}+1}$,
      · $L_{\tau_{n+1}} = L_{\tau_n} - 1$ and $I_{\tau_{n+1}} \in F_{n_{I_{\tau_n}}} \cup F_{n_{I_{\tau_n}}-1}$.

Here, the respective probabilities of those jumps that are admissible are chosen in such a way that **X** and L takes separately are indeed Markov processes with respective generators **L** and G.

In words, if $L \neq n_I$, L and **X** move independently from each other until the time where n_I and L agree. After that time, it is guaranteed that one always has $n_I \geq L$. We have not detailed precisely where I jumps exactly to be concise. But, if we ignore N, the couple $(\mathbf{X}, L)$ is just the Markov process generated by

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

if $l \neq n_i$ and

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

Now, for all $t \geq 0$ and $x \in E$, we have

$$
\begin{aligned}
\mathbf{P}_t V^q(x) &= \mathbb{E}\left[P^{(I_{\tau_{N_t}})}_{t-\tau_{N_t}} V^q(X_{\tau_{N_t}}) \right] \leq \mathbb{E}\left[e^{-\alpha(I_{\tau_{N_t}})(t-\tau_{N_t})} V^q(X_{\tau_{N_t}}) + K \right] \\
&= \mathbb{E}\left[e^{-\alpha(L_{\tau_{N_t}})(t-\tau_{N_t})} V^q(X_{\tau_{N_t}}) \right] + K \\
&= \mathbb{E}\left[e^{-\alpha(L_{\tau_{N_t}})(t-\tau_{N_t})} P^{(I_{\tau_{N_t}-1})}_{\tau_{N_t}-\tau_{N_t-1}} V^q(X_{\tau_{N_t-1}}) + K \right] \\
&\leq \ldots \leq \mathbb{E}\left[e^{-\int_0^t q\alpha(L_s)ds} \right] V(x) + K \sum_{n \geq 0} \mathbb{E}\left[e^{-q \int_0^{\tau_n} \alpha(L_s)ds} \right]. \qquad (3.2)
\end{aligned}
$$

Now, using Lemma 2.6, there exist $C, \eta > 0$ and $q \in (0,1]$ such that

$$
\mathbb{E}\left[e^{-\int_0^t q\alpha(L_s)ds} \right] \leq Ce^{-\eta t}, \qquad (3.3)
$$

Hence, it only remains to prove that

$$
\sum_{n \geq 0} \mathbb{E}\left[e^{-q \int_0^{\tau_n} \alpha(L_s)ds} \right] < +\infty.
$$

We cannot deduce this directly from equation (3.3) but, heuristically, if r and n are large enough then the law of large number gives that $\tau_n \approx n/r$ and thus

$$
\mathbb{E}\left[e^{-\int_0^{\tau_n} q\alpha(I_s)ds} \right] \approx \mathbb{E}\left[e^{-\int_0^{n/r} q\alpha(I_s)ds} \right] \leq Ce^{-\eta n/r},
$$

and the previous sum is finite. Now, we estimate the left hand side of the previous equation following that τ_n is lower than n/r, close to n/r or higher than n/r. Note that we have not

$$
e^{-\int_0^{\tau_n} q\alpha(I_s)ds} \leq 1,
$$

because α can be negative. Let $\epsilon > 0$ and denote by ϱ the worst case of decay:

$$
\varrho = -\min\{ q\alpha(k) \mid k \in F \}. \qquad (3.4)
$$

If τ_n is close to n/r then we have from (3.3) the bound

$$
\mathbb{E}\Big[e^{-\int_0^{\tau_n} q\alpha(I_s)ds} \mathbf{1}_{\{\tau_n \in [nr^{-1}(1-\epsilon), nr^{-1}(1+\epsilon)]\}} \Big]
$$

$$\leq e^{2\rho\varepsilon n/r}\mathbb{E}\Big[e^{-\int_0^{nr^{-1}(1-\epsilon)} q\alpha(I_s)ds}\mathbf{1}_{\{\tau_n\in[nr^{-1}(1-\epsilon),nr^{-1}(1+\epsilon)]\}}\Big]$$
$$\leq e^{2\rho\varepsilon n/r}\mathbb{E}\Big[e^{-\int_0^{nr^{-1}(1-\epsilon)} q\alpha(I_s)ds}\Big] \leq Ce^{-nr^{-1}(\eta-\epsilon(2\varrho+\eta))}.$$

Thereafter, we therefore fix $\epsilon < \eta(2\varrho+\eta)^{-1}$. Now, if τ_n is lower than n/r then, using Markov's inequality, we have

$$\begin{aligned}\mathbb{E}\left[e^{-\int_0^{\tau_n} q\alpha(I_s)ds}\mathbf{1}_{\{\tau_n<nr^{-1}(1-\epsilon)\}}\right] &\leq e^{\varrho nr^{-1}(1-\epsilon)}\mathbb{P}\left(\tau_n<nr^{-1}(1-\epsilon)\right)\\ &\leq e^{\varrho nr^{-1}(1-\epsilon)}e^{\theta nr^{-1}(1-\epsilon)}\mathbb{E}\left[e^{-\theta\tau_n}\right]\\ &\leq \exp\left(-n\left(\ln\left(1+\frac{\theta}{r}\right)-(1-\epsilon)r^{-1}(\varrho+\theta)\right)\right),\end{aligned}$$

for every $\theta \geq 0$. And finally, if τ_n is higher than n/r then, using the Cauchy-Schwarz and Markov inequalities, we have

$$\begin{aligned}\mathbb{E}\left[e^{-\int_0^{\tau_n} q\alpha(I_s)ds}\mathbf{1}_{\{\tau_n>nr^{-1}(1+\epsilon)\}}\right] &\leq \mathbb{E}\left[e^{2\varrho\tau_n}\right]^{1/2}\mathbb{P}\left(\tau_n>nr^{-1}(1+\epsilon)\right)^{1/2}\\ &\leq \mathbb{E}\left[e^{2\varrho\tau_n}\right]^{1/2}\left(e^{-\theta' nr^{-1}(1+\epsilon)}\mathbb{E}\left[e^{\theta'\tau_n}\right]\right)^{1/2}\\ &\leq \exp\left(-\frac{n}{2}\left(\ln\left(1-\frac{2\varrho}{r}\right)+\ln\left(1-\frac{\theta'}{r}\right)+\frac{\theta'(1+\epsilon)}{r}\right)\right),\end{aligned}$$

where $\theta' \geq 0$. Note that in the previous inequality, we have supposed that $r > 2\varrho$. Let $\gamma \in (0,1)$, we set $\theta = \theta' = \gamma r$. We can find a large r and a small γ verifying

$$\ln(1+\gamma) - (1-\epsilon)\gamma + (1-\epsilon)r^{-1}\varrho > 0,$$

and

$$\ln\left(1-\frac{2\varrho}{r}\right) + \ln(1-\gamma) + \gamma(1+\epsilon) > 0.$$

and thus there exist $C' > 0$ and $\varepsilon > 0$ such that

$$\sum_{n\geq 1}\mathbb{E}\left[e^{-\int_0^{\tau_n} q\alpha(I_s)ds}\right] \leq \sum_{n\geq 1} C'e^{-\varepsilon n} < +\infty\ ,$$

thus concluding the proof by combining this with (3.2) and (3.3). □

Remark 3.9. *If all Markov processes contract, then the proof simplifies considerably. Indeed, if for all $i \in F$, one has $\alpha(i) \geq \zeta > 0$, then one has*

$$\sum_{n\geq 1}\mathbb{E}\left[e^{-\int_0^{\tau_n} q\alpha(I_s)ds}\right] \leq \sum_{n\geq 1}\mathbb{E}\left[e^{-\zeta\tau_n}\right] \leq \sum_{n\geq 1}\left(\frac{r}{r+\zeta}\right)^n = \frac{r}{\zeta} < \infty\ .$$

3.3 The contracting distance

This section is divided in three parts. We introduce the distance $\widetilde{d}$ that we will use in Theorem 3.6, 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 $\widetilde{d}\colon (E\times F)\times(E\times F)\to[0,1]$ such that there exist $t_*>0$ and $\alpha\in(0,1)$ verifying

$$\widetilde{d}(\mathbf{x},\mathbf{y})<1 \quad\Rightarrow\quad \forall t\geq t_*,\quad \mathbb{E}[\mathcal{W}_{\tilde{d}}(\delta_{\mathbf{x}}P_t,\delta_{\mathbf{y}}P_t)]\leq \alpha\widetilde{d}(\mathbf{x},\mathbf{y}). \tag{3.5}$$

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

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

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 coupling $(\mathbf{X}_t,\mathbf{Y}_t)_{t\geq0}=((X_t,I_t),(Y_t,J_t))_{t\geq0}$ 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

$$T=\inf\{s\geq 0\mid I_s\neq J_s\}, \tag{3.6}$$

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

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

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 $r\geq0$ and $(N_t)_{t\geq0}$ be a Poisson process of intensity r with $N_t=\sum_{n\geq0}\mathbf{1}_{\{\tau_n\leq t\}}$ and $\tau_n=\sum_{k=1}^n E_k$ for a family $(E_k)_{k\geq0}$ of i.i.d. exponential variables as before and $\tau_0=0$. We assume that $r\geq 2\bar{a}$, i.e. that is r is bigger than the jump rates of I or J. As in the proof of Lemma 3.8 and Theorem 1.5, we give the construction of our coupling $(\mathbf{X},\mathbf{Y})$ at the jump times of N. Let $n\in\{0,..,N_t\}$, 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

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

  where $\mathcal{G}_n=\sigma\{(\mathbf{X}_{\tau_n},\mathbf{Y}_{\tau_n}),(\tau_k)_{k\geq0}\}$.

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

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

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

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

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

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

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

This Markov chain L is constructed as in section 3.8.

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 minimises the Wasserstein distance for each of the underlying processes. In this case, we can make our coupling Markovian with generator*

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

3.3.3 The distance $\widetilde{d}$ is contracting for 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 3.2, we can choose r and δ in such a way that*

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

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 (3.6), denotes the first time of separation of I and J. Using Lemma 2.6, there exist $q \in (0,1]$ and $C, \eta > 0$ such that

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

Here, we have used the fact that

$$\begin{aligned}
\mathbb{E}\left[\mathbf{1}_{\{T=\infty\}} d^q(X_t, Y_t)\right] &\leq \mathbb{E}\left[\mathbf{1}_{\{T\geq \tau_{N_t}\}} e^{-q\alpha(L_{\tau_{N_t}})(t-\tau_{N_t})} d^q(X_{\tau_{N_t}}, Y_{\tau_{N_t}})\right] \\
&\leq \mathbb{E}\left[\mathbf{1}_{\{T\geq \tau_{N_t}\}} e^{-q\alpha(L_{\tau_{N_t}})(t-\tau_{N_t})} \mathbb{E}\left[d^q(X_{\tau_{N_t}}, Y_{\tau_{N_t}}) \mid \mathcal{G}_n\right]\right] \\
&\leq \mathbb{E}\left[\mathbf{1}_{\{T\geq \tau_{N_t-1}\}} e^{-\int_{\tau_{N_t-1}}^t q\alpha(L_s)ds} d^q(X_{\tau_{N_t-1}}, Y_{\tau_{N_t-1}})\right] \\
&\leq \mathbb{E}\left[e^{-\int_0^t q\alpha(L_s)ds}\right] \mathbb{E}[d^q(x,y)]\ .
\end{aligned}$$

It remains to obtain a bound on $\mathbb{P}\left(T < +\infty\right)$. 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

$$A_n = \{T = \tau_n\} = \{T \geq \tau_n \text{ and } I_{\tau_n} \neq J_{\tau_n}\}.$$

By Assumption 1.1, we have

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

Hence

$$\mathbb{P}\left(T < \infty\right) = \sum_{n\geq 1} \mathbb{P}\left(A_n\right) \leq \frac{2^q \kappa^q}{r^q} d(x,y)^q \sum_{n\geq 1} \mathbb{E}\left[e^{-q\int_0^{\tau_n} \alpha(L_s)ds}\right].$$

Now, similarly to the proof of Lemma 3.8, there exist $C' > 0$ and $\varepsilon > 0$ verifying

$$\sum_{n\geq 1} \mathbb{E}\left[e^{-q\int_0^{\tau_n} \alpha(L_s)ds}\right] \leq \sum_{n\geq 1} C' e^{-\varepsilon n} =: \tilde{C} < +\infty.$$

Combining these bounds, we obtain the estimate

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

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 3.2, if $S \subset E \times F$ is of bounded diameter in the sense that*

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

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 Assumption 3.1, there exists $i_0 \in F$ such that $\rho(i_0) > 0$. Let $(\mathbf{X}_t)_{t\ge 0}$ and $(\mathbf{Y}_t)_{t\ge 0}$ be two independent processes generated by (1.2) and starting respectively from $\mathbf{x}$ and $\mathbf{y}$. Let us denote

$$\tau_{\text{in}} = \inf\{t \ge 0 \mid I_t = J_t = i_0\} \quad \text{and} \quad \tau_{\text{out}} = \inf\{t \ge \tau_{\text{in}} \mid I_t \ne i_0 \text{ or } J_t \ne i_0\}.$$

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

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

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

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

where ϱ was defined in (3.4). There exist $c > 0$ and $t_* > c$ such that $1-\delta^{-1}e^{\varrho c}e^{-\rho(i_0)t_*}R > 0$. 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 1.6 and Theorem 3.2

Lemma 2.2 and Lemma 3.8 give the existence of a Lyapunov function V, 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 3.6 to deduce that there exist a probability measure $\boldsymbol{\pi}$ and some constants $C, \lambda, t_0 > 0$ such that

$$\forall t \ge t_0,\ \mathcal{W}_{\widetilde{\mathbf{d}}}\left(\boldsymbol{\mu}\mathbf{P}_t, \boldsymbol{\pi}\right) \le Ce^{-\lambda t}\mathcal{W}_{\widetilde{\mathbf{d}}}\left(\boldsymbol{\mu}, \boldsymbol{\pi}\right),$$

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

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

where $\mathbf{x} = (x,i)$, $\mathbf{y} = (y,j)$ belong to $\mathbf{E}$, x_0 is as in Assumption 1.3 and $q \in (0,1]$. We conclude the proof by noting that $\mathbf{d} \leq \widetilde{\mathbf{d}}$.

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 [Clo12] for stochastically monotone processes, in [CJ10] for birth-death processes and in [Ebe11] 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

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

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

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

for some $K > 0$.

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

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

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

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

for any probability measures μ and ν.

Proof. It is usually proved considering the same Brownian motion for two different solutions of the SDE starting with different initial measures. □

Assumptions of Theorem 1.8 or Theorem 1.9 are satisfied if one of the underlying diffusions verifies Hörmander's hypoellipticity assumption. See for instance [Hai11] 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*

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

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 $\mathcal{L}^{(i)} f = G^{(i)} \cdot \nabla f$, 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 1.5 and Theorem 1.6. In general, we can not apply Theorem 1.8 or Theorem 3.3 but [BH12, BLMZ12b] 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* $k \geq 0$*,*

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

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

In this case our main theorem gives

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* $\alpha(n) \leq \min_{i\in F_n} \lambda(i)$*, for all* $n \leq \bar{n}$*.*

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

$$\forall t \geq t_0,\ d_{\mathrm{TV}}\left(\delta_{\mathbf{x}} \mathbf{P}_t, \boldsymbol{\pi}\right) \leq Ce^{-\lambda t}(1 + V(x)),$$

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

Proof. Using [BLMZ12b, Theorem 6.6], we see that compact sets are small for $\mathbf{X}$. Using Lemma 2.7 in the first case and Lemma 3.8 in the second case, we see that we can apply Theorem 2.10. □

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

$$\forall t \geq 0,\ \partial_t X_t = I_t X_t,$$

where $(I_t)_{t\geq 0}$ 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 1.5 and 1.6 give the exponential ergodicity of **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* **X** *can explode even though* **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 [BGM10] 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

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

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 1.5 gives the exponential-Wasserstein ergodicity under the condition that

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

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

Lemma 5.2. *Under assumption* (5.1)*, there is a distance* $\boldsymbol{\delta}$ *on* **E** *such that the Wasserstein curvature of the semigroup of* **X** *is positive, i.e. there exists* $\lambda > 0$ *such that*

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

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

Proof. Firstly, let us give a complement on the conclusion of Lemma 2.6. The Markov chain I satisfies its assumptions and using the results of [BGM10], there exist a function ψ on F, $\rho > 0$ and $p \in (0,1)$ verifying

$$\forall t \geq 0,\ \mathbb{E}\left[\psi(I_t)e^{-\int_0^t pa(I_s)ds}\right] = e^{-\rho t}\mathbb{E}\left[\psi(I_0)\right].$$

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

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

where

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

Now, using the fact that

$$\forall t \geq 0, X_t = X_0 e^{-\int_0^t a(I_s)ds},$$

the proof is straightforward. □

5.3 Surprising blow-up under exponential ergodicity assumptions

Here we gives some comments on [BLMZ12a, 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

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

$a(x,0,0) = a(x,1,1) = 0$, 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

$$\mathbf{L} f(x,i) = A_i \cdot \nabla f(x,i) + a\left(f(x,1-i) - f(x,i)\right). \tag{5.2}$$

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

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

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

$$\left\| Z_t^{(i)} \right\|_i \leq e^{-t} \left\| Z_0^{(i)} \right\|_i \text{ and } \left\| Z_t^{(i)} \right\|_{1-i} \leq 3e^{-t} \left\| Z_0^{(i)} \right\|_{1-i}, \tag{5.3}$$

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

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

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 [BLMZ12a, Example 1.4] shows that

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

In particular, the conclusion of Theorem 1.5 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 1.5 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 $-74/6$.

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

A last example is the following: the process X verifies

$$\forall t \geq 0,\ dX_t = -(X_t - a_{I_t})dt,$$

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

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

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

$$Jf(i) = \frac{1}{2}f(i+1) + \frac{1}{2}f(i-1) - f(i).$$

if $i \neq 0$ and

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

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

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

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

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

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

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

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

References

[BBMZ12] M. BENAÏM, S. L. BORGNE, F. MALRIEU, and P.-A. ZITT. Quantitative ergodicity for some switched dynamical systems. *Electron. Commun. Probab.* **17**, (2012), no. 56, 1–14.

[BGM10] J.-B. BARDET, H. GUÉRIN, and F. MALRIEU. Long time behavior of diffusions with Markov switching. *ALEA Lat. Am. J. Probab. Math. Stat.* **7**, (2010), 151–170.

[BH12] Y. BAKHTIN and T. HURTH. Invariant densities for dynamical systems with random switching. *Nonlinearity* **25**, no. 10, (2012), 2937.

[BKKP05] O. BOXMA, H. KASPI, O. KELLA, and D. PERRY. On/off storage systems with state-dependent input, output, and switching rates. *Probab. Engrg. Inform. Sci.* **19**, no. 1, (2005), 1–14.

[BLMZ12a] M. BENAÏM, S. LE BORGNE, F. MALRIEU, and P.-A. ZITT. On the stability of planar randomly switched systems. *ArXiv e-prints* (2012). arXiv:1204.1921.

[BLMZ12b] M. BENAÏM, S. LE BORGNE, F. MALRIEU, and P.-A. ZITT. Qualitative properties of certain piecewise deterministic Markov processes. *ArXiv e-prints* (2012). arXiv:1204.4143.

[CD08] O. L. V. COSTA and F. DUFOUR. Stability and ergodicity of piecewise deterministic Markov processes. *SIAM J. Control Optim.* **47**, no. 2, (2008), 1053–1077.

[CDMR12] A. CRUDU, A. DEBUSSCHE, A. MULLER, and O. RADULESCU. Convergence of stochastic gene networks to hybrid piecewise deterministic processes. *Annals of Applied Probability* **22**, no. 5, (2012), 1822–1859.

[CJ10] D. CHAFAI and A. JOULIN. Intertwining and commutation relations for birth-death processes. *ArXiv e-prints* (2010). arXiv:1011.2331.

[Clo12] B. CLOEZ. Wasserstein decay of one dimensional jump-diffusions. *ArXiv e-prints* (2012). arXiv:1202.1259.

[CMMS12] P. COLLET, S. MARTINEZ, S. MELEARD, and J. SAN MARTIN. Stochastic models for a chemostat and long time behavior. *ArXiv e-prints* (2012). arXiv:1206.3691.

[DF99] P. DIACONIS and D. FREEDMAN. Iterated random functions. *SIAM Rev.* **41**, no. 1, (1999), 45–76.

[dSY05] B. DE SAPORTA and J.-F. YAO. Tail of a linear diffusion with Markov switching. *Ann. Appl. Probab.* **15**, no. 1B, (2005), 992–1018.

[Ebe11] A. EBERLE. Reflection coupling and Wasserstein contractivity without convexity. *C. R. Math. Acad. Sci. Paris* **349**, no. 19-20, (2011), 1101–1104.

[FGM12] J. FONTBONA, H. GUÉRIN, and F. MALRIEU. Quantitative estimates for the long-time behavior of an ergodic variant of the telegraph process. *Adv. Appl. Probab.* **44**, no. 4, (2012), 977–994.

[GG96] C. M. GOLDIE and R. GRÜBEL. Perpetuities with thin tails. *Adv. in Appl. Probab.* **28**, no. 2, (1996), 463–480.

[GIY04] X. GUYON, S. IOVLEFF, and J.-F. YAO. Linear diffusion with stationary switching regime. *ESAIM Probab. Stat.* **8**, (2004), 25–35.

[GT11] A. GENADOT and M. THIEULLEN. Averaging for a Fully-Coupled Piecewise Deterministic Markov Process in Infinite Dimension. *ArXiv e-prints* (2011). arXiv:1109.6581.

[Hai11] M. HAIRER. On Malliavin's proof of Hörmander's theorem. *Bull. Sci. Math.* **135**, no. 6-7, (2011), 650–666.

[HM08] M. HAIRER and J. C. MATTINGLY. Spectral gaps in Wasserstein distances and the 2D stochastic Navier-Stokes equations. *Ann. Probab.* **36**, no. 6, (2008), 2050–2091.

[HM10] M. HAIRER and A. J. MAJDA. A simple framework to justify linear response theory. *Nonlinearity* **23**, no. 4, (2010), 909–922.

[HM11] M. HAIRER and J. C. MATTINGLY. Yet another look at Harris' ergodic theorem for Markov chains. In *Seminar on Stochastic Analysis, Random Fields and Applications VI*, vol. 63 of *Progr. Probab.*, 109–117. Birkhäuser/Springer Basel AG, Basel, 2011.

[HMS11] M. HAIRER, J. C. MATTINGLY, and M. SCHEUTZOW. Asymptotic coupling and a general form of Harris' theorem with applications to stochastic delay equations. *Probab. Theory Related Fields* **149**, no. 1-2, (2011), 223–259.

[HSV] M. HAIRER, A. STUART, and S. VOLLMER. Spectral gap for a metropolis-hasting algorithm in infinite dimensions. Arxiv e-print.

[Jou07] A. JOULIN. Poisson-type deviation inequalities for curved continuous-time Markov chains. *Bernoulli* **13**, no. 3, (2007), 782–798.

[Jou09] A. JOULIN. A new Poisson-type deviation inequality for Markov jump processes with positive Wasserstein curvature. *Bernoulli* **15**, no. 2, (2009), 532–549.

[MT93] S. P. MEYN and R. L. TWEEDIE. Stability of Markovian processes. III. Foster-Lyapunov criteria for continuous-time processes. *Adv. in Appl. Probab.* **25**, no. 3, (1993), 518–548.

[Oll09] Y. OLLIVIER. Ricci curvature of Markov chains on metric spaces. *J. Funct. Anal.* **256**, no. 3, (2009), 810–864.

[Oll10] Y. OLLIVIER. A survey of Ricci curvature for metric spaces and Markov chains. In *Probabilistic approach to geometry*, vol. 57 of *Adv. Stud. Pure Math.*, 343–381. Math. Soc. Japan, Tokyo, 2010.

[PTW12] K. PAKDAMAN, M. THIEULLEN, and G. WAINRIB. Asymptotic expansion and central limit theorem for multiscale piecewise-deterministic Markov processes. *Stochastic Process. Appl.* **122**, no. 6, (2012), 2292–2318.

[Smi68] W. L. SMITH. Necessary conditions for almost sure extinction of a branching process with random environment. *Ann. Math. Statist* **39**, (1968), 2136–2140.

[Vil09] C. VILLANI. *Optimal transport, old and new*, vol. 338 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 2009.

[vRS05] M.-K. VON RENESSE and K.-T. STURM. Transport inequalities, gradient estimates, entropy, and Ricci curvature. *Comm. Pure Appl. Math.* **58**, no. 7, (2005), 923–940.