Elementary coupling approach for non-linear perturbation of Markov processes with mean-field jump mechanisms and related problems.

Pierre Monmarché

Sorbonne Université

September 28, 2018

Abstract

Mean-field integro-differential equations are studied in an abstract framework, through couplings of the corresponding stochastic processes. In the perturbative regime, the equation is proven to admit a unique equilibrium, toward which the process converges exponentially fast. Similarly, in this case, the associated particle system is proven to converge toward its equilibrium at a rate independent from the number of particles.

Keywords: mean-field interaction, jump processes, PDMP, interacting particle system, coupling, integro-differential equation.

MSC : 60J75

1 Introduction

The initial motivation of this work is the study of the long-time behaviour of mean-field semi-linear integro-differential equations of the form

$$\partial_t m_t(x) = L'm_t(x) + Q'_{m_t}(\lambda_{m_t}(x))(x) - \lambda_{m_t}(x)m_t(x), \quad (1)$$

where $m_t(x)$ is the density of particles at time $t$ at point $x$, $L$ is the generator of a Markov process, $A'$ denotes the dual on measures of an operator $A$ on functions, and for all probability distribution $\nu$, $\lambda_{\nu} > 0$ is a jump rate and $Q_{\nu}$ is a jump kernel. In other words, the stochastic process whose law $m_t$ solves (1) follows the dynamics of $L$ and, at a rate $\lambda_{m_t}$, jumps to a new position drawn according to the kernel $Q_{m_t}$.

In particular, this encompasses the case of mean-field continuous-time Markov chains or of mean field Piecewise Deterministic Markov Processes (PDMP). Mixed jump/diffusion processes are also concerned, but only when the non-linearity lies in the jump mechanism. For instance, the kinetic neuron model studied in [44], for which the mean-field interaction lies in the drift, doesn’t enter this scope.

Equation (1) can model various phenomena like chemical reactions [45], neuron activity [42, 9], economical games [18, 19], communication networks [6]. In the field of...
stochastic algorithms, where PDMP samplers have recently gained interest [40, 51, 4], mean-field equations describe the limiting behaviour, as the number of particles go to infinity, of algorithms like parallel tempering [46], adaptive biasing force methods [32], selection/mutation techniques like Kalman-Bucy filter [16, 17], Fleming-Viot algorithms [28, 14], adaptive multi-level splitting [12] and so on.

We are interested in the long-time behaviour of the solutions of (1) in the perturbative regime, namely when the effects of the interaction is small with respect to the mixing effect of $L$. The general idea is simple and based on coupling methods. A classical way to obtain some contraction in (weighted) total variation distance for Markov processes is to define simultaneously two processes starting at different positions in such a way that they have merged at some positive time with some positive probability. Once they have met, in the classical setting, then they remain equal forever. However, in the non-linear case, even if we manage to merge two processes that started at different positions, they still don’t have exactly the same law, hence the same dynamics, and thus they may have to split after some time. Nevertheless, as they stay together, their laws get closer one to the other, which makes a splitting less and less likely. Provided the non-linearity is small enough, the balance between this two merge/split mechanisms is on the side of contraction.

This strategy of coupling two processes that don’t have the same dynamics in order to keep them as close as possible is classical in the study of mean-field diffusions like the McKean-Vlasov equation [35, 47]. In that case, a parallel coupling can typically be used, in other words the processes are defined as solutions of different SDE but driven by the same Brownian motion. Even if they start at the same position, both processes then drift away one from the other, but the evolution of the expectation of their squared distance is bounded by the difference in their drifts which, for mean-field processes, is typically controlled by the Wasserstein-2 distance between their law, which is itself controlled by the expectation of their squared distance. The expansion of the latter is thus controlled through the Gronwall Lemma.

When the interaction lies in the jump mechanism rather than in the deterministic drift one, although jump processes may lack the regularization properties of diffusions, the picture is somehow a bit simpler. Indeed, the analogous of the parallel coupling in that case means that, as much as possible, the two processes are prescribed to jump at the same times and to the same locations, and in particular, starting from the same position, they will stay equal for a (random) positive time. As a consequence, total variation estimates are obtained, which are then compatible with classical contraction estimates obtained via Foster-Lyapunov techniques. This synchronous coupling is used in [41] to study self-interacting PDMP and in [22] to control the difference between a smooth approximation of a PDMP and the latter.

This coupling method works both for the non-linear process and for its associated interacting particle system. Actually, the general idea is not restricted to mean-field interactions, and can be applied for instance to self-interacting processes with memory (see e.g. the discussion in Section 4.7), sampling processes when the target measure is close to be the product of its marginal (see Section 4.3), chain of oscillators, diffusions with colored noise, killed or branching processes, SPDEs etc. It works for discrete time interacting chains. Mattingly in [36, 37, 1] and Cloez and Thai in [14] use the same kind of arguments. For diffusion processes, rather than the simple parallel coupling, the analogous of our merge/split coupling should be the sticky coupling of Eberle and his co-authors [20, 24, 25]. In a word, it is a very natural and general strategy, and we expect the arguments developed in the particular framework of this paper to be of interest in many other settings.

The paper is organized as follows. The definitions and main results are presented in Section 2.2 for the non-linear problem and in Section 2.2 for the interacting particle
2 Definitions and main results

2.1 The non-linear process

Let $E$ be a Polish space, $(\Omega,\mathcal{F},\mathbb{P})$ be a probability space on which all the random variables in this work will be implicitly defined, and let $L$ be the infinitesimal generator of a conservative Feller semi-group $(P_t)_{t \geq 0}$ on $E$. Denote $\mathcal{P}(E)$ the set of probability measures on $E$ and $\mathcal{D}(E)$ the set of càdlàg functions for $\mathbb{R}_+$ to $E$, endowed with the Skorokhod topology. From [31, Chapter 17], for all initial condition $x \in E$, there exists an associated strong Markov process $(Z_t^x)_{t \geq 0}$ on $\mathcal{D}(E)$ such that $P_tf(x) = E(f(Z_t^x))$ for all $t \geq 0$ and $f \in C_0(E)$. In order to avoid regularity or well-posedness considerations, we will rather work at the level of the stochastic process than of the generator.

For all $\nu \in \mathcal{P}(E)$, let $\lambda_\nu$ be a measurable function from $E$ to $\mathbb{R}_+$, and let $Q_\nu$ be a Markov kernel on $E$, namely a measurable function from $E$ to $\mathcal{P}(E)$. Throughout all this work, if $Q : E \to \mathcal{P}(E)$ is a Markov kernel, we also denote $Q$ the Markov operator defined by $Qf(x) = (Q(x))(f)$ for all $x \in E$ and all bounded measurable functions $f$. We call $\lambda_\nu$ the non-linear jump rate, and $Q_\nu$ the non-linear jump kernel.

Let $G_\nu : E \times [0, 1] \to E$ be a representation of $Q_\nu$, namely be such that, if $U$ is a uniformly distributed random variable (r.v.) on $[0, 1]$ then, for all $x \in E$, $G_\nu(x, U)$ is distributed according to $Q_\nu(x)$. From [3, Corollary 7.16.1] such a representation always exists. Similarly, let $H : (x, u) \in E \times [0, 1] \mapsto (H_t(x, u))_{t \geq 0} \in \mathcal{D}(E)$ be a representation of the kernel $x \mapsto \text{Law}((Z_t^x)_{t \geq 0})$.

Assumption 1. There exists $\lambda_*$ such that for all $\nu \in \mathcal{P}(E)$ and $x \in E$, $\lambda_\nu(x) \leq \lambda_*$. 

This assumption that the jump rate is bounded uniformly, both in $x \in E$ and $\nu \in \mathcal{P}(E)$, is made for simplicity. Although it already holds in many interesting applications, it may sometimes be too restrictive, but in many such cases it can be circumvented via proper a priori bounds (see e.g. Section 4.6), specific to the problem at hand.

Let $\mathcal{M}(\mathbb{R}_+, \mathcal{P}(E))$ be the set of measurable function from $\mathbb{R}_+$ to $\mathcal{P}(E)$. Under Assumption 1, for $x \in E$ and $\mu : t \mapsto \mu_t$ in $\mathcal{M}(\mathbb{R}_+, \mathcal{P}(E))$, we want to define an inhomogeneous Markov process $(X_t^{\mu,x})_{t \geq 0}$ starting from $x$ which, loosely speaking, follows the Markov dynamics of the semi-group $(P_t)_{t \geq 0}$ between some random jump times drowned at rate $t \mapsto \lambda_\mu(X_t^{\mu,x})$ at which it jumps to new values drowned according to the distribution $Q_\mu(X_t^{\mu,x})$. More precisely, consider an i.i.d. sequence $(S_k, U_k, V_k, W_k)_{k \in \mathbb{N}}$ where, for all $k \in \mathbb{N}$, $S_k$, $U_k$, $V_k$ and $W_k$ are independent one from the other, $S_k$ follows an exponential law with parameter 1 and $U_k$, $V_k$ and $W_k$ a uniform distribution over $[0, 1]$. Set $T_0 = 0$, $X_0^{\mu,x} = x$ and suppose that $T_n \geq 0$ and $(X_t^{\mu,x})_{t \in [0,T_n]}$ have been defined for some $n \in \mathbb{N}$ and are independent from $(S_k, U_k, V_k, W_k)_{k \geq n}$. Set $T_{n+1} = T_n + S_n/\lambda_*$. Set $\mathcal{M}(\mathbb{R}_+^{n+1}, \mathcal{P}(E))$.

- For all $t \in (T_n, T_{n+1})$, set $X_t^{\mu,x} = H_{t-T_n}(X_{T_n}^{\mu,x}, U_n)$.
- If $V_n \leq \lambda_{T_{n+1}^{-1}}(H_{E_n/\lambda_*}(X_{T_n}^{\mu,x}, U_n))/\lambda_*$ then set
  \[ X_{T_{n+1}}^{\mu,x} = G_{T_{n+1}}(H_{S_{n}/\lambda_*}(X_{T_n}^{\mu,x}, U_n), W_n). \]
- Else, set $X_{T_{n+1}}^{\mu,x} = H_{S_{n}/\lambda_*}(X_{T_n}^{\mu,x}, U_n)$.
By induction, \((X_t^{\mu,x})_{t \in [0,T_n]}\) is thus defined for all \(n \in \mathbb{N}\) and, since \(T_n\) is the \(n^{th}\) jump time of a Poisson process, it goes almost surely to infinity as \(n\) goes to infinity, so that \(X_t^{\mu,x}\) is almost surely defined for all \(t \geq 0\). Denote by \((R_t^{\mu})_{t \geq s \geq 0}\) the inhomogeneous Markov semi-group defined by

\[
R_{s,t}^{\mu} f(x) = \mathbb{E} f(X_{t-s}^{\sigma_s \mu,x})
\]

for all bounded measurable functions \(f, t \geq s \geq 0\) and \(x \in E\), where \(\sigma_s\) is the shift of time \(s\) defined by \((\sigma_s \mu)_t = \mu_{s+t}\) for all \((\mu_t)_{t \geq 0}\) and all \(t \geq 0\). Note that, at least formally for suitable functions \(f_0, f_t = R_{0,t}^{\mu} f_0\) is a solution of

\[
\partial_t f_t = L f_t := L f_t + \lambda_{\mu_t}(Q_{\mu_t} f_t - f_t) .
\]  

**Definition 1.** We say that \(m \in \mathcal{M}(\mathbb{R}_+, \mathcal{P}(E))\) is a solution of \((1)\) with initial distribution \(m_0\) if \(m_0 R_t^{\mu} = m_t\) for all \(t \geq 0\).

Note that Definition 1 gives solutions in a quite weak sense, but in many cases additional informations on \(L\) allows to precise the meaning of \((2)\) and thus obtain, by uniqueness of the solution, that a solution in the sense of Definition 1 is in fact a strong solution of \((1)\).

For \(\nu_1, \nu_2 \in \mathcal{P}(E)\), we denote their total variation distance

\[
\| \nu_1 - \nu_2 \|_{TV} = \sup_{\| f \|_\infty \leq 1} |(\nu_1 - \nu_2)(f)| .
\]

**Assumption 2.** There exists \(\theta > 0\) such that for all \(\nu_1, \nu_2 \in \mathcal{P}(E), \)

\[
\sup_{\| f \|_\infty \leq 1} \| \lambda_{\nu_1} (Q f - f) \|_{\infty} \leq \theta \| \nu_1 - \nu_2 \|_{TV} .
\]

**Proposition 2.** Under Assumptions 1 and 2, for all \(m_0 \in \mathcal{P}(E)\), there exists a unique solution of \((1)\). Moreover, if \(t \mapsto m_t, h_t\) are solutions with respective initial conditions \(m_0, h_0\), then for all \(t \geq 0\)

\[
\| m_t - h_t \|_{TV} \leq e^{\theta t} \| m_0 - h_0 \|_{TV} .
\]

This will be proven in Section 3.2.2 through a fixed point procedure. As far as the long time behaviour of the process is concerned, let us first state a simple result:

**Assumption 3.** There exist \(t_0, \alpha \geq 0\) such that for all \(x, y \in E, \)

\[
\| \delta_x P_{t_0} - \delta_y P_{t_0} \|_{TV} \leq 2(1-\alpha) .
\]

**Theorem 3.** Under Assumptions 1, 2 and 3, if \(t \mapsto m_t, h_t\) are solutions of \((1)\) with respective initial distributions \(m_0, h_0 \in \mathcal{P}(E)\) then, for all \(t \geq 0, \)

\[
\| m_t - h_t \|_{TV} \leq e^{\theta t_0} (e^{\theta t_0} - \alpha e^{-\lambda_t t_0})^{t/t_0} \| m_0 - h_0 \|_{TV} .
\]

An equivalent result is proved in [9], for a particular model, via the Duhamel formula. We will rather establish it (in Section 3.2.2) with a similar but probabilistic point of view. This proof will introduce, in a simpler context, most of the ideas then used in the more general proof of Theorem 3 below.

The uniform Doeblin condition of Assumption 3 is quite demanding, in particular if \(E\) is not compact. The classical Foster-Lyapunov approach [30, 39] for Markov processes usually gathers a local Doeblin condition on compact sets and the existence of a Lyapunov function that tends to decrease on average along a trajectory, away from some compact. The following counterpart of this strategy for non-linear perturbations seems to be new.

For a measurable function \(V\) from \(E\) to \([1, \infty)\), denote \(P_V(E) := \{ \nu \in \mathcal{P}(E), \nu(V) < \infty \}\) and for \(\mu, \nu \in P_V(E), \| \mu - \nu \|_V \equiv \| \mu - \nu \|_V\) the so-called \(V\)-norm of \(\mu - \nu\). In particular, if \(V = 1\), then \(\| \cdot \|_V = \| \cdot \|_{TV}\).
Theorem 4. Under Assumptions 1 and 4, for all $t \in \mathbb{R}_+$, $x \in E$, $t \geq 0$, and almost all $u \in [0,1]$, \begin{equation}
abla_{\mu_0} V(x) \leq e^{-\rho t} V(x) + \rho \int_0^t e^{\rho(s-t)} (\eta \mu_a(V) + M) ds \tag{3}
abla \theta \nabla P_0 V(x) \leq e^{\rho t} V(x) \tag{4}
abla \theta \nabla (G_{\mu_0}(x,u)) \leq \gamma_a V(x) \tag{5}
abla \theta \end{equation}

- For all $\nu_1, \nu_2 \in \mathcal{P}_V(E)$, \begin{equation}
abla \theta \sup_{\|f\|_1} \|\lambda_{\nu_1} (Q_{\nu_1} f - f) - \lambda_{\nu_2} (Q_{\nu_2} f - f)\|_\infty \leq \theta \|\nu_1 - \nu_2\|_V \tag{6}
abla \theta \end{equation}

- For all $x, y \in E$ with $V(x) + V(y) \leq 8(1 + M/(1 - \eta))$, \begin{equation}
abla \theta \|\delta_x P_0 - \delta_y P_0\|_{TV} \leq 2(1 - \alpha) \tag{7}
abla \theta \end{equation}

Note that (3) and (4) can usually be established if, for all $\nu \in \mathcal{P}_V$, \begin{equation}
abla \theta LV + \lambda_\nu (Q_\nu V - V) \leq -\rho (V - \eta \nu(V) - M) , \quad LV \leq \rho_a V , \nabla \theta \end{equation}
provided additional regularity and/or truncation arguments, see e.g. [39] or Section 4.

The Doeblin condition (7) is always satisfied with $\alpha = 0$.

Theorem 4. Under Assumptions 1 and 4 for all $m_0 \in \mathcal{P}_V$, there exists a unique solution $t \mapsto m_t$ of (1) with initial condition $m_0$. Moreover, for all $t \geq 0$, \begin{equation}
abla \theta \nabla m_t(V) \leq e^{-\rho(1-\eta)t} m_0(V) + (1 - e^{-\rho(1-\eta)t}) M \frac{1}{1 - \eta} . \tag{8}
abla \theta \end{equation}

Theorem 5. Under Assumptions 1 and 4 if $\alpha > 0$, denote \begin{equation}
abla \theta \beta = \frac{2(1 - e^{-\rho t_0}) e^{\lambda_1 t_0}}{\alpha} \left( \frac{M}{1 - \eta} - 1 \right) \nabla \theta \nabla \nabla C_* = \frac{2(1 + \beta) \gamma_*(\frac{M}{1 - \eta} + 1)}{\theta (\rho + \theta \gamma_*) \left(\frac{M}{1 - \eta} + 1\right)} e^{(\rho + \theta \gamma_*) (\frac{M}{1 - \eta} + 1) t_0 - 1} \nabla \theta \nabla \kappa = \frac{1 + e^{-\rho t_0}}{2} \left(1 - \frac{1}{2} \alpha e^{-\lambda_1 t_0}\right) \nabla \theta \nabla \tilde{\kappa} = e^{-\rho(1-\eta)t_0} \left(\kappa + C_*\right). \nabla \theta \end{equation}

Suppose that $\tilde{\kappa} < 1$. Then Equation (1) admits an equilibrium $\mu_\infty \in \mathcal{P}_V(E)$, i.e. a constant solution $t \mapsto m_t = \mu_\infty$ for all $t \geq 0$, which is unique in $\mathcal{P}_V(E)$. Moreover, $\mu_\infty(V) \leq M/(1 - \eta)$, and for all $m_0 \in \mathcal{P}_V$ and all $t \geq 0$, \begin{equation}
abla \theta \nabla \|m_t - \mu_\infty\|_V \leq 2\tilde{\kappa}^{t/t_0-1} \left(\beta + 1 + \frac{M}{1 - \eta}\right) m_0(V) . \tag{9}
abla \theta \end{equation}

Remark that, for fixed values of the other parameters of Assumptions 1 and 4, $\tilde{\kappa} < 1$ if $\theta$ is small enough.
2.2 Weakly interacting Markov particles

Let $N \in \mathbb{N}$ and, for $i \in [1, N]$, let $E_i$ be a Polish space and $(P^i_t)_{t \geq 0}$ be a conservative Feller semi-group on $E_i$. Let $E = \prod_{i=1}^N E_i$ and $(P_t)_{t \geq 0}$ be the semi-group on $E$ defined by

$$P_t\left(\prod_{i=1}^N f_i\right)(x) = \prod_{i=1}^N P^i_t f_i(x_i)$$

for all $x \in E$, $t \geq 0$ and all measurable bounded functions $f_i$ on $E_i$, $i \in [1, N]$. For $i \in [1, N]$ let $H^i_t$ be a representation of $x \in E_i \mapsto \text{Law}(Z^i_{t+})$ on $\mathcal{D}(E_i)$ where, for all $z \in E_i$, $(Z^i_{t+})_{t \geq 0}$ is a Markov process associated to $(P^i_t)_{t \geq 0}$ with initial value $z$. In particular, if $(U_1, \ldots, U_N)$ are i.i.d. r.v. uniformly distributed over $[0, 1]$ and $x \in E$ then $(H^i_t(x_i, U_i))_{i \in [1, N], t \geq 0}$ is a Markov process on $E$ associated to $(P_t)_{t \geq 0}$ with initial value $x$. For all $i \in [1, N]$ let $\lambda_i$ be a jump rate on $E$, i.e. be a measurable function from $E$ to $\mathbb{R}_+$, and $Q_i$ be a measurable function from $E$ to $\mathcal{P}(E_i)$ with representation $G_i$. We extend $Q_i$ as a Markov kernel $Q'_i$ on $E$ defined for all $x \in E$ by

$$Q'_i(x) = \delta_{x_1} \otimes \cdots \otimes \delta_{x_{i-1}} \otimes Q_i(x) \otimes \delta_{x_{i+1}} \otimes \cdots \otimes \delta_{x_N}.$$

**Assumption 5.** There exists $\lambda_\ast > 0$ such that, for all $x \in E$ and all $i \in [1, N]$, $\lambda_i(x) \leq \lambda_\ast$.

Under Assumption 5 let $Q$ be the Markov kernel on $E$ defined by

$$Q(x) = \frac{1}{N} \sum_{i=1}^N \frac{\lambda_i(x)}{\lambda_\ast} Q'_i(x) + \left(1 - \frac{\lambda_i(x)}{\lambda_\ast}\right) \delta_x$$

for all $x \in E$, and let $G$ be a representation of $Q$. To jump from a position $x$ to a r.v. drawn according to $Q(x)$ is equivalent to chose a coordinate $i$ uniformly in $[1, N]$ and, with probability $\lambda_i(x)/\lambda_\ast$, make it jump to a new position drawn according to $Q_i(x)$, else leave it at its current position, and in either case leave all the other coordinates unchanged.

Let $x \in E$. We define a Markov process $(X^x_t)_{t \geq 0} = (X^x_{i,t})_{i \in [1, N], t \geq 0}$ on $E$ as follows. Consider an i.i.d. sequence $(S_k, (U_{i,k},)_{i \in [1, N]}, V_k)_{k \in \mathbb{N}}$ where, for all $k \in \mathbb{N}$, $S_k$, $V_k$ and $(U_{i,k})_{i \in [1, N]}$ are independent one from the other, $S_k$ follows an exponential law with parameter 1, $V_k$ a uniform distribution over $[0, 1]$ and $(U_{i,k})_{i \in [1, N]}$ are i.i.d. r.v. uniformly distributed over $[0, 1]$. Set $T_0 = 0$, $X^x_0 = x$ and, for all $n \in \mathbb{N}$, $T_{n+1} = T_n + \lambda_\ast^{-1} S_n$. Suppose that $(X^x_t)_{t \in [0, T_n]}$ have been defined for some $n \in \mathbb{N}$ and is independent from $(S_k, (U_{i,k})_{i \in [1, N]}, V_k)_{k \geq n}$. For all $t \in (T_n, T_{n+1})$ and all $i \in [1, N]$, set $X^x_{i,t} = H^i_{T_n - T_{i+1}}(X^x_{i,T_n}, U_{i,n})$. Finally, set

$$X^x_{n+1} = G\left(\left(H^i_{T_{n+1} - T_n}(X^x_{i,T_n}, U_{i,n})\right)_{i \in [1, N]}, V_n\right).$$

By induction, $(X^x_t)_{t \in [0, T_n]}$ is thus defined for all $n \in \mathbb{N}$ and, since $T_n$ is the $n^{\text{th}}$ jump time of a Poisson process, it goes almost surely to $+\infty$ as $n$ goes to infinity, so that $X_t$ is almost surely defined for all $t \geq 0$. Let $(R_t)_{t \geq 0}$ be the associated Markov semi-group, i.e.

$$R_t f(x) = \mathbb{E}(f(X^x_t))$$

for all bounded measurable functions $f$ on $E$.

Consider on $E$ the $l_0$ distance defined by

$$d_1(x, y) = 2 \mathbb{1}\{i \in [1, N], : x_i \neq y_i\},$$

for all $x, y \in E$, where $\mathbb{1}A$ denote the cardinality of a finite set $A$. Remark that the topology induced by $d_1$ is equivalent to the trivial one, with more precisely

$$2N \mathbb{1}_{x \neq y} \leq d_1(x, y) \leq 2N \mathbb{1}_{x \neq y}.$$
Assumption 6. There exists $\theta, t_0, \alpha \geq 0$ such that, for all $i \in [1, N]$, 

- for all $x, y \in E$ with $y_i = x_i$,
  $\sup_{\|f\|_\infty \leq 1} \|\lambda_i(x) (Q_i f(x) - f(x_i)) - \lambda_i(y) (Q_i f(y) - f(y_i))\|_\infty \leq \theta \bar{d}_1(x, y) / N. \quad (10)$

- for all $x, y \in E_i$,
  $\|\delta_x P_{t_0}^i - \delta_y P_{t_0}^i\|_{TV} \leq 2(1 - \alpha). \quad (11)$

Here is the analogous of Theorem 3 for interacting particles.

Theorem 6. Under Assumptions 5 and 6 for all $m_0, h_0 \in \mathcal{P}(E)$ and $t \geq 0$,

$\|m_0 R_t - h_0 R_t\|_{TV} \leq N e^{\theta t_0} \left(e^{\theta t_0} - \alpha e^{-\lambda_* t_0}\right)^{\lfloor t/t_0 \rfloor} \|m_0 - h_0\|_{TV}. \quad (12)$

Moreover, denoting $m'_t$ and $h'_t$ the respective laws of $X_{i,t}$ and $Y_{i,t}$ where $X_i \sim m_0 R_t$, $Y_i \sim h_0 R_t$ and $I$ is uniformly distributed over $[1, N]$ and independent from $X_t$ and $Y_t$, then

$\|m'_t - h'_t\|_{TV} \leq e^{\theta t_0} \left(e^{\theta t_0} - \alpha e^{-\lambda_* t_0}\right)^{\lfloor t/t_0 \rfloor} \|m'_0 - h'_0\|_{TV}. \quad (13)$

Note that, using a naive approach (trying to merge at time $t_0$ two whole systems of particles), it is easy to prove that the condition (11) together with Assumption 5 imply the uniform Doeblin condition

$\frac{1}{2} \|\delta_x R_{t_0} - \delta_y R_{t_0}\|_{TV} \leq 1 - \left(e^{-\lambda_* t_0 \alpha}\right)^N,$

but then the contraction rate is geometrically poor with respect to $N$. This is why the Foster-Lyapunov approach is sometimes thought to scale badly with dimension.

Let us show how Theorems 3 may apply in the case of a mean-field particle system associated to a non-linear process such as defined in Section 2.1. Suppose that $E_i = E_1$, $P_i = P_1$ and

$\lambda_i(x) = \lambda_{\mu_N(x)}(x_i), \quad Q_i(x) = Q_{\mu_N(x)}(x_i)$

for all $i \in [1, N]$ and $x \in E = E_1^N$, where $\lambda_v$ and $Q_v$ are given in Section 2.1 and

$\mu_N(x) = \frac{1}{N} \sum_{i=1}^N \delta_{x_i}$

is the empirical distribution of $x$. Remark that, for $x, y \in E$,

$\|\mu_N(x) - \mu_N(y)\|_{TV} \leq \bar{d}_1(x, y) / N. \quad (14)$

Indeed, $(x_I, y_I)$ with a r.v. $I$ uniformly distributed over $[1, N]$ is a coupling of $\mu_N(x)$ and $\mu_N(y)$ with $F(x_I = y_I) = \bar{d}_1(x, y) / N$. Hence, Assumptions 1, 2, and 3 imply Assumptions 5 and 6 with the same $\lambda_*, \theta, t_0, \alpha$.

If the initial positions $(X_{i,0})_{i \in [1, N]}$ are i.i.d. r.v. with a given law $m_0 \in \mathcal{P}(E_1)$, then the particles are exchangeables, in the sense that for all $t \geq 0$, $(X_{\sigma(i),t})_{i \in [1, N]}$ and $(X_{i,t})_{i \in [1, N]}$ have the same distribution for all permutation $\sigma$ of $[1, N]$. In particular, if $I$ is uniformly distributed over $[1, N]$ and independent from $(X_I)_{t \geq 0}$ then $X_{I,t}$ and $X_{1,t}$ have the same law. This means that, under Assumptions 1, 2, and 3 then

$\|m_t - h_t\|_{TV} \leq e^{\lambda_* \theta t_0} \left(e^{\lambda_* \theta t_0} - \alpha e^{-\lambda_* t_0}\right)^{\lfloor t/t_0 \rfloor} \|m_0 - h_0\|_{TV} \quad (15)$
holds with either $m_t$ and $h_t$ solutions of the non-linear equation \((\text{I})\) or the law of the first particle of the associated system.

We go back to the general case. In order to relax the strong uniform Doeblin condition \((\text{I})\) under the assumption that the process admits a suitable Lyapunov function, we now state an analogous of Theorem 5 for the particle system.

**Assumption 7.** There exist $\theta \geq 0$, $\rho, t_0 > 0$, $\rho_\ast \in \mathbb{R}$, $\alpha \in (0, 1]$, $\eta \in [1/2, 1)$, $M, \gamma_\ast \geq 1$ and, for all $i \in [1, N]$, a measurable $V_i : E_i \to [1, \infty)$, such that, denoting $V(x) = \sum_{i=1}^N V_i(x_i)$, the following holds for all $i \in [1, N]$:

- For all $x \in E$, $t \geq 0$, and almost all $u \in [0, 1]$,
  \[
  R_t V_i(x) \leq e^{-\rho t} V_i(x) + \rho \int_0^t e^{\rho(s-t)} \left( \frac{\eta}{N} R_s V(x) + M \right) ds \tag{12}
  \]
  \[
  P^t_i V_i(x_i) \leq e^{\rho t} V_i(x_i) \tag{13}
  \]
  \[
  V_i(G_i(x, u)) \leq \gamma_i V_i(x_i). \tag{14}
  \]

- For all $x, y \in E$ such that $x_i = y_i$,
  \[
  \sup_{\|f\|_\infty \leq 1} |\lambda_i(x)(Q_i f(x) - f(x_i)) - \lambda_i(y)(Q_i f(y) - f(y_i))|_\infty \leq \frac{\theta \bar{d}_1(x, y)}{N}. \tag{15}
  \]

- For all $x_i, y_i \in E_i$ with $V_i(x_i) + V_i(y_i) \leq 16(1 + \eta)M/(1 - \eta)^2$,
  \[
  \|\delta_{x_i} P^t_{i0} - \delta_{y_i} P^t_{i0}\|_{TV} \leq 2(1 - \alpha). \tag{16}
  \]

Under Assumption 7, denote $\bar{d}_\nu$ and $\bar{\rho}$ the distances, respectively on $E$ and $\mathcal{P}_\nu(E)$, defined for all $x, y \in E$ by

\[
\bar{d}_\nu(x, y) = \sum_{i=1}^N \mathbb{I}_{x_i \neq y_i} \left( V_i(x_i) + \frac{V(x) + V(y)}{N} \right),
\]

and for all $\mu, \nu \in \mathcal{P}_\nu(E)$ by

\[
\bar{\rho}(\mu, \nu) = \inf \left\{ E(\bar{d}_\nu(X, Y)) \mid X \sim \mu, Y \sim \nu \right\}.
\]

(See Section 3.1 for a more detailed definition of couplings.) Remark that $\|\mu - \nu\|_{TV} \leq \bar{\rho}(\mu, \nu)$ and that $\|\mu - \nu\|_\nu \leq N \bar{\rho}(\mu, \nu)$. Note also that, in the case of a particle system associated to a mean-field equation, i.e. if the particles are exchangeables and $V_i = V_j$ for all $i, j \in [1, N]$, denoting $\mu'$ and $\nu'$ the laws of $X_1$ and $Y_1$ when $X \sim \mu$ and $Y \sim \nu$ (hence of $X_i$ and $Y_i$ with an independent $I$ uniform over $[1, N]$) then $\|\mu' - \nu'\|_{V_i} \leq \bar{\rho}(\mu, \nu)/N$.

**Theorem 7.** Under Assumption 6 and 7 denote

\[
\beta = \frac{4(1 + \eta)Me^{\lambda_i t_0}}{\alpha(1 - \eta)^2} \left( 1 - e^{-\frac{1}{2}\rho(1-\eta) t_0} \right),
\]

\[
\kappa = \frac{1 + e^{-\frac{1}{2}\rho(1-\eta) t_0}}{2} \sqrt{\left( 1 - \frac{1}{2} \alpha e^{-\lambda_i t_0} \right)},
\]

\[
\bar{\kappa} = \kappa + \left( e^{\beta t_0} - 1 \right) (\gamma_\ast + 1) e^{\rho \alpha - \lambda_i (\gamma_\ast - 1) t_0} (1 + \eta).
\]

Suppose that $\bar{\kappa} < 1$. Then, $(R_{t\epsilon})_{t \geq 0}$ admits a unique invariant measure $\mu_\infty$ and for all $m_0 \in \mathcal{P}_\nu(E)$ and all $t \geq 0$,

\[
\bar{\rho}(m_0 R_t, \mu_\infty) \leq \bar{\kappa}^{[t/t_0]} \left( 2N \left( \beta + M \frac{1 + \eta}{(1 - \eta)^2} \right) + \frac{1 + \eta}{1 - \eta} \nu(V) \right).
\]
For all $X,Y$ then we also say that $(\mu_1 \sim \mu_2)$ is called a coupling of $X,Y$.

Remark that it induces the trivial topology on $V$. Let us first give an alternative representation of the $V$ function.

3.1 General considerations on couplings

Let us now give a coupling interpretation of $\langle \rangle$. In other words, in this mean-field case, the assumptions of Theorem 5. Indeed, the right-hand side of (15) involves $d_V$, which would be the analogous of (6). In other words, in this mean-field case, the assumptions of Theorem 7 are equivalent to the assumptions of Theorem 5 with the addition of Assumption 2. The interesting question of weakening the condition (15) in Theorem 7 raises non-trivial considerations about the non-independence of $V_i(X_{i,t})$ for different $i \in [1,N]$, and is postponed to a future work.

3 Proofs

3.1 General considerations on couplings

Let us first give an alternative representation of the $V$-norm, for a given measurable function $V$ from $E$ to $[1, \infty]$. Consider the distance $d_V$ on $E$ defined for all $x,y \in E$ by

$$d_V(x,y) = 1_{x \neq y} (V(x) + V(y)).$$

Remark that it induces the trivial topology on $E$, since $d_V(x_n,x) \geq 21_{x \neq y}$ for all $x,y \in E$. Hairer and Mattingly proved in [30] that

$$\|\mu_1 - \mu_2\|_V = \sup_{\phi : \|\phi\|_{V,1} \leq 1} (\mu_1(\phi) - \mu_2(\phi)) = \sup_{\phi : \|\phi\|_{V,2} \leq 1} (\mu_1(\phi) - \mu_2(\phi))$$

(17)

where, for all measurable $\phi$ from $E$ to $\mathbb{R}$,

$$\|\phi\|_{V,1} = \sup_{x \in E} \frac{\phi(x)}{V(x)}, \quad \|\phi\|_{V,2} = \sup_{x \neq y} \frac{|\phi(x) - \phi(y)|}{d_V(x,y)}.$$

Let us now give a coupling interpretation of $\|\|_V$. For $\mu_1, \mu_2 \in \mathcal{P}(E)$, we denote $\xi(\mu_1, \mu_2)$ the set of transference plane between $\mu_1$ and $\mu_2$, i.e. the set of probability measures $\nu$ on $E$ such that $(X,Y) \sim \nu$ implies that $X \sim \mu_1$ and $Y \sim \mu_2$. Such a random variable $(X,Y)$ is called a coupling of $\mu_1$ and $\mu_2$. Alternatively, in that case, if $Z_i \sim \mu_i$ for $i = 1,2$, then we also say that $(X,Y)$ is a coupling of $Z_1$ and $Z_2$.

Lemma 8. For all $\mu_1$ and $\mu_2$,

$$\|\mu_1 - \mu_2\|_V = \inf \{ \mathbb{E}(d_V(X,Y)) : (X,Y) \sim \nu, \nu \in \xi(\mu_1, \mu_2) \},$$

and the infimum is attained for some $\nu_{opt} \in \xi(\mu_1, \mu_2)$ that does not depend on $V$.

Proof. Note that, according to [17], $\|\|_V$ is the Wasserstein-1 distance on $\mathcal{P}(E)$ associated to the distance $d_V$ on $E$, so that the first statement of the lemma stems from the general result of duality for Wasserstein distances. Nevertheless, in the general case, the coupling would depend on the metric. Let us then give an elementary proof in the present specific case, straightforwardly adapted from the classical case of the total variation distance. First, note that $|\mu_1 - \mu_2| = \mu_1 + \mu_2 - 2\mu_1 \wedge \mu_2$. If $\nu \in \xi(\mu_1, \mu_2)$ and $(X,Y) \sim \nu$, then for all event $A$ of $E$,

$$\mathbb{P}(X \in A, X = Y) \leq \mu_1(A) \wedge \mu_2(A).$$

In other word, the measure $\hat{\nu}$ on $E$ defined by $\hat{\nu}(dx) = \int_E 1_{x \neq y} \nu(dx,dy)$ satisfies $\hat{\nu} \leq \mu_1 \wedge \mu_2$. Therefore, writing $1_{x \neq y} = 1 - 1_{x = y}$,

$$\mathbb{E}(d_V(X,Y)) = \mu_1(V) + \mu_2(V) - \int_{E^2} (V(x) + V(y)) 1_{x = y} \nu(dx,dy)$$

$$= \mu_1(V) + \mu_2(V) - 2 \int_E V(x) \hat{\nu}(dx)$$

$$\geq \mu_1(V) + \mu_2(V) - 2\mu_1 \wedge \mu_2(V) = \|\mu_1 - \mu_2\|_V.$$
It now remains to construct an optimal coupling, i.e. a measure $\nu \in \xi(\mu_1, \mu_2)$ such that equality holds. Let $p = \mu_1 \wedge \mu_2(E)$. If $p = 1$ then $\mu_1 = \mu_2$ and the optimal coupling is to draw $X$ according to $\mu_1$ and then set $Y = X$. If $p = 0$ then $\mu_1$ and $\mu_2$ are singular one to the other, and all couplings are optimal since

$$\mathbb{E}(d_\nu(X, Y)) \leq \mathbb{E}(\nu(X) + \nu(Y)) = \|\mu_1 - \mu_2\|_\nu.$$ 

If $p \in (0, 1)$, we consider the probability measures

$$\nu_0 = \frac{\mu_1 \wedge \mu_2}{p}, \quad \nu_1 = \frac{\mu_1 - \mu_1 \wedge \mu_2}{1 - p}, \quad \nu_2 = \frac{\mu_2 - \mu_1 \wedge \mu_2}{1 - p}.$$ 

Remark that $\nu_1$ and $\nu_2$ are singular one to the other, and that $\mu_1 = p\nu_0 + (1 - p)\nu_i$ for $i = 1, 2$. Let $E_1, E_2, E_3$ and $B$ be independent random variables with $E_i \sim \nu_1$ for $i = 1, 2, 3$ and $B \sim B(p)$ the Bernoulli law with parameter $p$. Set $X = Y = E_0$ if $B = 1$ and $X = E_1, Y = E_2$ if $B = 0$. Then $X \sim \mu_1$ and $Y \sim \mu_2$, and $\{X = Y\} = \{B = 1\}$. Conditioning on the values of $B$,

$$\mathbb{E}(d_\nu(X, Y)) = (1 - p)\mathbb{E}(d_\nu(X, Y) \mid B = 0) = (1 - p)\mathbb{E}(\nu(E_1) + \nu(E_2)) = \int \nu(x) \left(\mu_1 + \mu_2 - 2\mu_1 \wedge \mu_2\right)(dx) = \|\mu_1 - \mu_2\|_\nu,$$

which concludes.

This representation of $\| \cdot \|_\nu$ is convenient since, as soon are we able to construct a coupling of two probability measures, we get an upper bound on their distance.

**Lemma 9.** Let $(P_t)_{t \geq 0}$ be a conservative Feller semi-group on $E$, $t_0 \geq 0$, $x, y \in E$ and $\nu$ be a measurable function from $E$ to $[1, \infty)$. Then there exists a random variable $(Z^x_t, Z^y_t)_{t \geq 0}$ on $\mathcal{D}(E^2)$ such that, for $z = x, y$, $(Z^z_t)_{t \geq 0}$ is a Markov process associated to $(P_t)_{t \geq 0}$ with $Z^z_0 = z$,

$$\mathbb{E}(d_\nu(Z^x_{t_0}, Z^y_{t_0})) = \|\delta_x P_{t_0} - \delta_y P_{t_0}\|_\nu, \quad (18)$$

and moreover,

$$\forall s \geq t \geq 0, \quad Z^x_t = Z^y_t \Rightarrow Z^z_s = Z^z_t. \quad (19)$$

We call such a process an optimal coupling of $(\delta_x P_t)_{t \geq 0}$ and of $(\delta_y P_t)_{t \geq 0}$ at time $t_0$.

**Proof.** Let $(X^x, X^y)$ be an optimal coupling of $\delta_x P_t$ and $\delta_y P_t$ such as constructed in Lemma 8. For $z \in E$, let $(V^z_t)_{t \geq 0}$ be a Markov process associated with $(P_t)_{t \geq 0}$ with $V^z_0 = z$, and let $W_t = V^z_{t-t_0}$ for $t \in [0, t_0]$. Then by usual conditioning arguments, $(W_t)_{t \in [0, t_0]}$ is an inhomogeneous Markov process with transitions

$$\mathbb{P}(W_t \in A \mid W_s \in B) = \mathbb{P}(V_{t-s} \in B \mid V_{t-s} \in A) \frac{\mathbb{P}(V_{t-s} \in A)}{\mathbb{P}(V_{t-s} \in B)}$$

for all $0 \leq s \leq t \leq t_0$ and all events $A, B$ of $E$. Let $(R^x_{s,t})_{0 \leq s < t \leq t_0}$ be the inhomogeneous semi-group associated to $(W_t)_{t \in [0, t_0]}$. For $z = x, y$, let $(U^z_t)_{t \in [0, t_0]}$ be a Markov process associated to $(R^z_{s,t})_{0 \leq s < t \leq t_0}$ with $U^z_0 = X^z$. For all $t \in [0, t_0]$ and $z = x, y$, set $Z^z_t = U^z_{t-t_0}$. In other words, we have defined $(Z^z_t)_{t \in [0, t_0]}$ to be a Markov bridge from $z$ to $X^z$. Then, set $Z^z_t = H_{t-t_0}(X^z, U^z)$ for all $t \geq t_0$ and $z = x, y$, where $U^z$ is uniformly distributed on $[0, 1]$ and independent from $(Z^z_t, Z^y_t)_{t \in [0, t_0]}$, and $H$ is a representation of $z \mapsto (V^z_t)_{t \geq 0}$. By the Markov property, $(Z^z_t)_{t \geq 0}$ is then a Markov process with initial condition $Z^z_0 = z$ and such that $Z^z_{t-t} = X^z$.

Define a second coupling $(\tilde{Z}^x_t, \tilde{Z}^y_t)_{t \geq 0}$ as follows. Let $\tau = \inf\{t \geq 0 : Z^x_t = Z^y_t\}$. For $z = x, y$, set $\tilde{Z}^z_t = Z^z_t$ for all $t < \tau$ and $\tilde{Z}^z_\tau = Z^z_\tau$ for $t \geq \tau$. By the strong Markov property, $(\tilde{Z}^z_t)_{t \geq 0}$ is a Markov process associated to $(P_t)_{t \geq 0}$ with $\tilde{Z}^z_0 = z$, and the two other conditions are satisfied. \qed
Exempls of couplings such that [18] holds and [19] does not are easily constructed for discrete-time Markov chains. Here is an exemple with continuous time. Consider the process $(X_t)_{t \geq 0}$ on $[-1, 1]$ with generator

$$Lf(x) = sign(x)f'(x) + \frac{f(1) + f(-1)}{2} - f(0).$$

In other words, starting from $x \in [-1, 1] \setminus \{0\}$, then $X_t = x - sign(x)t$ for $t \in [0, |x|]$, after which $X_t = 0$ for all $t \in [|x|, |x| + T]$ where $T$ is an exponential r.v. with parameter 1 and $X_{|x|+T} = R$ where $R$ is a Rademacher r.v. with parameter 1/2. It is clear that if $|x| = |y|$ then $\delta_x P_t = \delta_y P_t$ for all $t \geq |x|$, since $\delta_x P_{|x|} = 0$. Now consider for some $t_0 > 2$ the processes $(X_t, Y_t)_{t \geq 0}$ that jump at the same time $|x| + T$, the first one to a Rademacher r.v. $R$, the second one to $R' = R(1 - 21_{|x|+T<t_0-1})$, which is indeed a Rademacher r.v. independent from $T$. In other words, as long as the jump occurs early enough for the processes to have the time to come back to 0 before time $t_0$, then we send them to opposite points, else to the same one. Obviously [18] holds and [19] does not.

### 3.2 Study of the non-linear process

We use in this section the notations of Section 2.1. In particular we consider a Markov semi-group $(P_t)_{t \geq 0}$ and non-linear jump rate and kernel $\nu \mapsto \lambda_\nu, Q_\nu$.

#### 3.2.1 Coupling with different inhomogeneous jump mechanisms

Consider $\mu^1, \mu^2 \in \mathcal{M}(\mathbb{R}_+, \mathcal{P}(E))$. For all $x, y \in E$, $t \geq 0$ and $i = 1, 2$, denote

$$\tilde{Q}^i_t(x) = \frac{\lambda_{\mu^i}(x)}{\lambda_x} Q^i_t(x) + \left(1 - \frac{\lambda_{\mu^i}(x)}{\lambda_x}\right) \delta_x,$$

and $p_t(x, y) = (Q^i_t(x) \wedge \tilde{Q}^i_t(y))(E)$. If $p_t(x, y) = 0$, set

$$Q^i_t(x, y) = \delta_x, \quad Q^{i, \neq}_t(x, y) = \tilde{Q}^i_t(x).$$

If $p_t(x, y) \in (0, 1)$, set

$$Q^i_t(x, y) = \frac{\tilde{Q}^i_t(x) \wedge \tilde{Q}^2_t(y)}{p_t(x, y)}, \quad Q^{i, \neq}_t(x, y) = \frac{\tilde{Q}^i_t(x) - \tilde{Q}^1_t(x) \wedge \tilde{Q}^2_t(y)}{1 - p_t(x, y)}.$$

Finally, if $p_t(x, y) = 1$, set

$$Q^i_t(x, y) = \tilde{Q}^i_t(x), \quad Q^{i, \neq}_t(x, y) = \delta_x.$$

Then $Q^i_t$ and $Q^{i, \neq}_t$ for $i = 1, 2$ are Markov kernels from $E^2$ to $\mathcal{P}(E)$ such that, for all $x, y \in E$ and $i = 1, 2$,

$$\tilde{Q}^i_t(x) = p_t(x, y) Q^i_t(x, y) + (1 - p_t(x, y)) Q^{i, \neq}_t(x, y),$$

and, as shown in the proof of Lemma 8, $\|\tilde{Q}^i_t(x) - \tilde{Q}^2_t(y)\|_{TV} = 2(1 - p_t(x, y))$, so that the condition 6 of Assumption 4 reads

$$\forall x \in E, \forall t \geq 0, \quad 1 - p_t(x, y) \leq \frac{\theta}{2\lambda_x} \|\mu^1_t - \mu^2_t\|_\nu,$$  \hspace{1cm} (20)$$

and similarly with $\nu = 1$ for Assumption 2. For all $t \geq 0$ and $i = 1, 2$, let $G^i_t$ (resp. $G^{i, \neq}_t$) be a representation of $Q^i_t$ (resp. $Q^{i, \neq}_t$), in the sense that for all $x, y \in E$, if $U$ is a r.v. uniformly distributed over $[0, 1]$, then $G^i_t(x, y, U) \sim Q^i_t(x, y)$. 

11
We define a Markov process \((X_t, Y_t)_{t \geq 0}\) on \(E^2\) as follows. Let \(m_0, h_0 \in \mathcal{P}(E)\) and let \((X_0, Y_0)\) be an optimal coupling of \(m_0\) and \(h_0\) such as constructed in the proof of Lemma 8. For a given \(t_0\), let \((Z_t, \tilde{Z}_t)_{t \geq 0}\) be an optimal coupling of \((\delta_{X_0}, P_t)_{t \geq 0}\) and \((\delta_{Y_0}, P_t)_{t \geq 0}\) in the sense of Lemma 9. Consider an i.i.d. sequence \((S_k, U_k, V_k, W_k)_{k \in \mathbb{N}}\), independent from \((Z_t, \tilde{Z}_t)_{t \geq 0}\) where, for all \(k \in \mathbb{N}\), \(S_k, U_k, V_k\) and \(W_k\) are independent one from the other, \(S_k\) follows a standard exponential law with parameter 1 and \(U_k, V_k\) and \(W_k\) a uniform distribution over \([0, 1]\). Set \(T_0 = 0, T_{n+1} = T_n + S_n/\lambda_*\) for all \(n \in \mathbb{N}\) and suppose that \((X_t, Y_t)_{t \in [0, T_n]}\) have been defined for some \(n \in \mathbb{N}\) and are independent from \((S_k, (U_{i,k})_{i \in [1, N]}, V_k, W_k)_{k \geq n}\).

- Set \((X_t, Y_t) = (Z_t, \tilde{Z}_t)\) for all \(t \in [0, T_1)\) and \((\tilde{X}_{T_1}, \tilde{Y}_{T_1}) = (Z_{T_1}, \tilde{Z}_{T_1})\). If \(n \geq 1\), for all \(t \in (T_n, T_{n+1})\), set
  \[
  (X_t, Y_t) = (H_{t-T_n}(X_{T_n}, U_n), H_{T_n}(Y_{T_n}, U_n))
  \]
  and
  \[
  (\tilde{X}_{T_{n+1}}, \tilde{Y}_{T_{n+1}}) = (H_{T_{n+1}-T_n}(X_{T_n}, U_n), H_{T_{n+1}+T_n}(Y_{T_n}, U_n))
  \]

- If \(V_n \leq p T_n\left(\tilde{X}_{T_{n+1}}, \tilde{Y}_{T_{n+1}}\right)\), then set
  \[
  X_{T_{n+1}} = Y_{T_{n+1}} = G_{T_n}^1(\tilde{X}_{T_{n+1}}, \tilde{Y}_{T_{n+1}}, W_n),
  \]
  and else set
  \[
  X_{T_{n+1}} = G_{T_n}^1(\tilde{X}_{T_{n+1}}, \tilde{Y}_{T_{n+1}}, W_n), \quad Y_{T_{n+1}} = G_{T_n}^2(\tilde{Y}_{T_{n+1}}, \tilde{X}_{T_{n+1}}, W_n).
  \]

**Proposition 10.** For all \(t \geq 0\), \(X_t \sim m_0 R_{0,t}^{\mu_1}\) and \(Y_t \sim h_0 R_{0,t}^{\mu_2}\).

**Proof.** By symmetry, we only consider the case of \((X_t)_{t \geq 0}\). Note that in the definition of \((X_t)_{t \geq 0}\), as in the definition of \((X_t^{\mu_1,x})_{t \geq 0}\) in Section 2.1, \((T_n)_{n \geq 0}\) is a Poisson process with intensity \(\lambda_*\), moreover, for all \(n \geq 0\), conditionally to \(X_{T_n}\), \((X_t)_{t \in [T_n, T_{n+1})}\) is a Markov process associated to \((P_t)_{t \geq 0}\), i.e. has the same distribution as \((Z^{X_{T_n}}_t)_{t \in [0, T_{n+1} - T_n]}\), and the same goes for \((X_t^{\mu_1,x})_{t \in [T_n, T_{n+1})}\). As a consequence, it only remains to check that the Markov chain \((X_{T_n})_{n \in \mathbb{N}}\) has the same distribution as the Markov chain \((X_t^{\mu_1,x})_{n \in \mathbb{N}}\), which is equivalent to say that they have the same transition kernel. Yet, for any bounded measurable \(f\) on \(E\),

\[
\mathbb{E}\left(f(X_{T_1}) \mid T_1 = t, \tilde{X}_{T_1} = \tilde{x}, \tilde{Y}_{T_1} = \tilde{y}\right) = \mathbb{E}\left[1_{V_0 \in p_t(\tilde{x}, \tilde{y})} f(G_{t}^{\mu_1}(\tilde{x}, \tilde{y}, W_0)) \right]
+ 1_{V_0 > p_t(\tilde{x}, \tilde{y})} f(G_{t}^{1,\neq}(\tilde{x}, \tilde{y}, W_n))
= p_t(\tilde{x}, \tilde{y}) Q_t^{\mu}(\tilde{x}, \tilde{y}) + (1 - p_t(\tilde{x}, \tilde{y})) Q_t^{1,\neq}(\tilde{x}, \tilde{y})
= Q_t f(\tilde{x}).
\]

Hence,

\[
\mathbb{E}(f(X_{T_1})) = \int_{0}^{+\infty} P_t^{1} Q_t^{1} f(x) \lambda_* e^{-\lambda_* t} dt
= \int_{0}^{+\infty} P_t^{1} \left(\lambda_1^{\mu_1} Q_{\mu_1} f + (\lambda_* - \lambda_1^{\mu_1}) \right) e^{-\lambda_* t} dt.
\]
Similarly,
\[
\mathbb{E} \left( f(X_{T_1}^{\mu, x}) \mid T_1 = t, H_{E_0/\lambda} (x, U_0) = \tilde{x} \right) = \mathbb{E} \left[ \mathbb{I}_{0 < \lambda_{\mu} (\tilde{x})} \lambda^r f(G_{\mu} (\tilde{x}, W_0)) + \mathbb{I}_{0 > \lambda_{\mu} (\tilde{x})} f(\tilde{x}) \right]
\]
\[
= \frac{\lambda_{\mu} (\tilde{x})}{\lambda_s} Q_{\mu} f(x) + \left( 1 - \frac{\lambda_{\mu} (\tilde{x})}{\lambda_s} \right) f(\tilde{x})
\]
and again
\[
\mathbb{E} \left( f(X_{T_1}^{\mu, x}) \right) = \int_0^{+\infty} P_t \left( \lambda_{\mu} Q_{\mu} f + \left( \lambda_s - \lambda_{\mu} \right) f \right) e^{-\lambda_s t} dt,
\]
which concludes. \hfill \Box

### 3.2.2 The total variation case

**Proof of Proposition 2**. For \((\mu_1, \mu_2)_{t \geq 0}, m_0, h_0 \in \mathcal{P}(E)\), consider the associated process \((X_t, Y_t)_{t \geq 0}\) defined in section 3.2.1. From Proposition 10, for all \(t \geq 0\),
\[
\frac{1}{2} \| m_0 R_{0,t}^{\mu_1} - h_0 R_{0,t}^{\mu_2} \|_{TV} \leq \mathbb{P} \left( X_t \neq Y_t \right)
\]
\[
\leq \mathbb{P} \left( X_0 = Y_0 \right) + \mathbb{P} \left( \tau_{\text{split}} < t \mid X_0 = Y_0 \right)
\]
where \(\tau_{\text{split}} = \inf \{ T_n : n \in \mathbb{N}, V_n \geq \lambda_T (\tilde{X}_{T_n+1}, \tilde{Y}_{T_n+1}) \} \). Note that, by construction, \(X_0 = Y_0\) implies that \(X_t = Y_t\) for all \(t < \tau_{\text{split}}\) so that, conditionally to \(X_0 = Y_0\),
\[
\tau_{\text{split}} \geq \inf \{ T_n : n \in \mathbb{N}, 1 - V_n < \theta \| \mu_{T_n} - \mu_{T_n}^2 \|_{TV} / (2 \lambda_s) \} := \tilde{\tau}_{\text{split}},
\]
where we used (20). Note that \(\tilde{\tau}_{\text{split}}\) is independent from \((X_0, Y_0)\), so that
\[
\mathbb{P} \left( \tau_{\text{split}} < t \mid X_0 = Y_0 \right) \leq \mathbb{P} \left( \tilde{\tau}_{\text{split}} < t \right)
\]
\[
= 1 - \exp \left( - \frac{\theta}{2} \int_0^t \| \mu_s^{1} - \mu_s^2 \|_{TV} ds \right).
\]
Since \((X_0, Y_0)\) is an optimal coupling of \(m_0\) and \(h_0\), using that \(1 - \exp(-a) \leq a\) for all \(a \in \mathbb{R}\), we have thus obtained
\[
\| m_0 R_{0,t}^{\mu_1} - h_0 R_{0,t}^{\mu_2} \|_{TV} \leq \| m_0 - h_0 \|_{TV} + \theta \int_0^t \| \mu_s^{1} - \mu_s^2 \|_{TV} ds. \tag{21}
\]
Now, remark that, for fixed \(m_0 \in \mathcal{P}(E), t_1 \geq 0\), \((m_0 R_{0,t}^{\mu_1})_{t \in [0,t_1]}\) only depends on \((\mu)_t \in [0,t_1]\). Set \(t_1 = 1/(2\theta)\) and consider \(\Psi : (\mu)_t \in [0,t_1] \rightarrow (m_0 R_{0,t}^{\mu_1})_{t \in [0,t_1]}\). Then
\[
\sup_{s \in [0,t_1]} \| m_0 R_{0,s}^{\mu_1} - m_0 R_{0,s}^{\mu_2} \|_{TV} \leq \frac{1}{2} \sup_{s \in [0,t_1]} \| \mu_s^{1} - \mu_s^2 \|_{TV},
\]
In other words \(\Psi\) is a contraction of \(L^\infty ([0,t_1], \mathcal{P}(E))\), which is complete, hence admits a unique fixed point, which is a solution of (1) in the sense of Definition 1 on \([0, t_1]\). Then there exists a unique solution on \([nt_1, (n+1)t_1]\) for all \(n \in \mathbb{N}\), thus on \(\mathbb{R}_+\).

Considering two such solutions with respective initial distributions \(m_0\) and \(h_0\), (21) reads
\[
\| m_t - h_t \|_{TV} \leq \| m_0 - h_0 \|_{TV} + \theta \int_0^t \| m_s - h_s \|_{TV} ds,
\]
and Gronwall’s Lemma concludes. \hfill \Box
Proof of Theorem 3. Let $t \mapsto m_t, h_t$ be two solutions of (1). Consider again the process $(X_t, Y_t)$ defined in Section 3.2.1 with $\mu^1 = m$ and $\mu^2 = h$. In particular,

$$ \frac{1}{2} \|m_{t_0} - h_{t_0}\|_{TV} \leq P(X_{t_0} \neq Y_{t_0}) \leq P(X_{t_0} \neq Y_{t_0}, X_0 \neq Y_0) + P(X_{t_0} \neq Y_{t_0}, X_0 = Y_0) .$$

In the proof of Proposition 2, we established that

$$ 2P(X_{t_0} \neq Y_{t_0} | X_0 = Y_0) \leq \theta \int_0^t \|m_s - h_s\|_{TV} ds \leq \theta \int_0^t e^{\gamma_s} \|m_0 - h_0\|_{TV} ds = (e^{\theta t_0} - 1) \|m_0 - h_0\|_{TV} .$$

By construction, for all $t < T_1$, $(X_t, Y_t) = (Z_t, \tilde{Z}_t)$, so that

$$ P(X_{t_0} = Y_{t_0} | X_0 \neq Y_0) \geq P\left(Z_{t_0} = \tilde{Z}_{t_0}, T_1 < t_0 | X_0 \neq Y_0\right) = P\left(Z_{t_0} = \tilde{Z}_{t_0} | Z_0 \neq \tilde{Z}_0\right) P(T_1 < t_0) \geq \alpha e^{-\lambda_t t_0} .$$

Since $P(X_0 \neq Y_0) = \|m_0 - h_0\|_{TV}/2$, we have obtained

$$ \|m_{t_0} - h_{t_0}\|_{TV} \leq \left(e^{\theta t_0} - \alpha e^{-\lambda_t t_0}\right) \|m_0 - h_0\|_{TV} .$$

Now, for any $t \geq 0$, using again Proposition 2, we get

$$ \|m_t - h_t\|_{TV} \leq e^{\theta(t - \tau_{split})} \|m_{[t/\tau_{split}]t_0} - h_{[t/\tau_{split}]t_0}\|_{TV} \leq e^{\theta t_0} \left(e^{\theta t_0} - \alpha e^{-\lambda_t t_0}\right) \|m_0 - h_0\|_{TV} .$$

\[ \square \]

3.2.3 The general $V$ norm case

Lemma 11. Under Assumptions 1 and 4, for all $\mu^1, \mu^2 \in \mathcal{M}(\mathbb{R}_+, \mathcal{P}(E))$, $x \in E$ and $t \geq 0$,

$$ \|\delta_x R^{\mu^1}_{0,t} - \delta_x R^{\mu^2}_{0,t}\|_{\mathcal{V}} \leq \theta \gamma_s \mathcal{V}(x) e^{(\theta \gamma_s + \lambda_t (\gamma_s - 1))t} \int_0^t \|\mu^1_u - \mu^2_u\|_{\mathcal{V}} du .$$

Proof. Consider the process $(X_t, Y_t)_{t \geq 0}$ defined in Section 3.2.1 with $X_0 = Y_0 = x$, and $t_1 \geq 0$. As in the proof of Proposition 2 let $\tau_{split} = \inf\{T_n : n \in \mathbb{N}, V_n \geq pr_n(X_{T_{n+1}}, Y_{T_{n+1}})\}$. Since $X_t = Y_t$ for all $t < \tau_{split}$,

$$ \tau_{split} \geq \inf\{T_n : n \in \mathbb{N}, 1 - V_n \leq \theta \|\mu^1 - \mu^2\|_{\mathcal{V}} / (2\lambda_t)\} := \tilde{\tau}_{split} ,$$

and we bound

$$ \mathbb{E}(d\mathcal{V}(X_{t_1}, Y_{t_1})) \leq \mathbb{E}(\mathbb{1}_{\tilde{\tau}_{split} \leq t_1}(\mathcal{V}(X_{t_1}) + \mathcal{V}(Y_{t_1}))) .$$

Set $\Gamma = (S_k, V_k)_{k \in \mathbb{N}}$, and remark that $\tilde{\tau}_{split}$ and $K = \{n \in \mathbb{N} : T_n \leq t_1\}$ are deterministic functions of $\Gamma$, while $(Z_t, \tilde{Z}_t)_{t \geq 0}$ and $(U_k, W_k)_{k \in \mathbb{N}}$ are independent from $\Gamma$. Using (4) and (5), for all $n \in \mathbb{N}$,

$$ \mathbb{E}(\mathcal{V}(X_{T_n}) | \Gamma) \leq \gamma_s \mathbb{E}\left(\mathcal{V}(X_{T_n}) | \Gamma\right) = \gamma_s \mathbb{E}(F_{T_n - T_{n-1}}^{\mu^1} \mathcal{V}(X_{T_{n-1}}) | \Gamma) \leq \gamma_s e^{\gamma_s (T_n - T_{n-1})} \mathbb{E}(\mathcal{V}(X_{T_{n-1}}) | \Gamma) .$$

14
Similarly, and then by direct induction,
\[
E (V(X_{t_1}) \mid \Gamma) \leq e^{\rho t_1 - T_K} E (V(X_{T_K}) \mid \Gamma) \leq \gamma_s^K e^{\rho t_1} V(x).
\]
The case of \( Y_{t_1} \) being identical, with \( Y_0 = X_0 = x \), we have obtained
\[
E (d_Y (X_{t_1}, Y_{t_1})) \leq 2 e^{\rho t_1} V(x) E (1_{\tau_{\text{split}} \leq t_1} \gamma_s^K).
\]
Besides, \((V_k)_{k \in \mathbb{N}} \) is independent from \( K \), and conditionally to \( K \), \( \{T_i\}_{i \in [1,K]} \) are distributed like \( K \) independent uniform r.v. over \([0,t_1] \) and independent from \((V_k)_{k \in \mathbb{N}} \), so that
\[
\mathbb{P} (\tau_{\text{split}} > t_1 \mid K) \geq \mathbb{P} \left( 1 - V_n \geq \theta \| \mu_{T_{n+1}}^1 - \mu_{T_{n+1}}^2 \|_V / (2 \lambda_s), \quad \forall n \in \{0,K-1\} \mid K \right) = \left( 1 - \frac{1}{t_1} \int_0^{t_1} 1 \wedge \left( \frac{\theta}{2 \lambda_s} \| \mu_u^1 - \mu_u^2 \|_V \right) du \right)^K := s^K.
\]
Since \( K \) follows a Poisson law with parameter \( \lambda_s t_1 \),
\[
E (d_Y (X_{t_1}, Y_{t_1})) \leq 2 e^{\rho t_1} V(x) E (1 - s^K) \gamma_s^K = 2 e^{\rho t_1} V(x) \left( e^{\lambda_s t_1 (\gamma_s - 1)} - e^{\lambda_s t_1 (\gamma_s - 1)} \right) \leq 2 \lambda_s t_1 \gamma_s (1 - s) V(x) e^{(\rho + \lambda_s \gamma_s - 1) t_1},
\]
which concludes.

\[\square\]

**Proof of Theorem 4.** Let \( t \mapsto \mu_t^1, \mu_t^2 \) be two measurable functions from \( \mathbb{R}_+ \) to \( \mathcal{P}_V \), and \( m_0 \in \mathcal{P}_V \). From Lemma 11,
\[
\| m_0 R_{0,t}^{+} - m_0 R_{0,t}^{-} \|_V = \int \| \delta_x R_{0,t}^{+} - \delta_x R_{0,t}^{-} \|_V m_0 (dx) \leq \theta \gamma_s m_0 (V) e^{(\rho_s + \lambda_s (\gamma_s - 1)) t} \int_0^t \| \mu_u^1 - \mu_u^2 \|_V du.
\]
Moreover, if \( \mu_t^1 (V) \leq m_0 (V) \lor (M/(1 - \eta)) \) for all \( t \geq 0 \), then (3) implies that
\[
m_0 R_{0,t} (V) \leq e^{-\rho t} m_0 (V) + \left( e^{-\rho t} \right) \left( \eta \left( m_0 (V) \lor \frac{M}{1 - \eta} \right) + M \right) \leq m_0 (V) \lor \frac{M}{1 - \eta}.
\]
Let \( t_1 \) be small enough so that
\[
t_1 \theta \gamma_s \left( m_0 (V) \lor \frac{M}{1 - \eta} \right) e^{(\rho_s + \lambda_s (\gamma_s - 1)) t_1} \leq \frac{1}{2}.
\]
Let \( \mathcal{A} \) be the set of measurable functions \( t \mapsto \mu_t \) from \([0,t_1]\) to \( \mathcal{P}_V \) such that \( \mu_t (V) \leq m_0 (V) \lor (M/(1 - \eta)) \) for all \( t \in [0,t_1] \). Remark that \( \mathcal{A} \) is not empty since it contains the function with constant value \( m_0 \). As a closed subset of \( L^\infty ([0,t_1], \mathcal{P}_V) \), \( \mathcal{A} \) is complete with the distance
\[
dist_{\mathcal{A}} (\mu^1, \mu^2) = \sup_{s \in [0,t_1]} \| \mu^1 - \mu^2 \|_V.
\]
The map \( \Psi : (\mu_t)_{t \in [0,t_1]} \rightarrow (m_0 R_{0,t} (V))_{t \in [0,t_1]} \) being a contraction of \( \mathcal{A} \), it admits a unique fixed point in \( \mathcal{A} \), which is then a solution of (1) over the time interval \([0,t_1]\). In particular, \( m_{t_1} \leq m_0 (V) \lor (M/(1 - \eta)) \), and the same arguments (with the same time \( t_1 \)) yield a solution on the time interval \([t_1,2t_1] \) and then by induction on \( \mathbb{R}_+ \).

If \( m \) is such a solution, then (3) integrated with respect to \( m_0 \) and applied with \( \mu_s = m_s \) for all \( s \geq 0 \) yields
\[
e^{\rho t} m_t (V) \leq m_0 (V) + \eta \rho \int_0^t e^{\rho s} m_s (V) ds + (e^{\rho t} - 1) M := j_t
\]
As in the proof of Theorem 4, the Gronwall’s Lemma yields, for all \( t \to m_t \) solution of (1) with \( m_0(V) \leq M/(1-\eta) + 1 \), all \( t \geq 0 \) and all \( x \in E \),

\[
R_{0,t}^m V(x) \leq e^{-\rho t} V(x) + (1 - e^{-\rho t}) \left( \frac{M}{1-\eta} + 1 \right).
\]

(22)

Lemma 12. Under Assumptions 7 and 12, if \( t \to m_t, h_t \) are solutions of (1) with \( m_0(V) \vee m_0(V) \leq M/(1-\eta) + 1 \), then for all \( t \geq 0 \),

\[
\|m_t - h_t\|_V \leq e^{(\rho + \theta_\gamma t)} \left( \frac{M}{1-\eta} + 1 \right) \|m_0 - h_0\|_V.
\]

Proof. Consider \((X_0, Y_0)\) an optimal coupling of \( m_0 \) and \( h_0 \) such as given by Lemma 8 and let \((X_t, Y_t)\) be an optimal coupling of \( \delta_{X_0} R_{0,t}^m \) and \( \delta_{Y_0} R_{0,t}^h \), so that

\[
\|m_t - h_t\|_V \leq E(\delta_{Y_0}(X_t, Y_t)) = E((1_{X_0=Y_0} + 1_{X_0\neq Y_0}) \delta_{Y_0}(X_t, Y_t))
\]

Since \( E((1_{X_0=Y_0} V(X_0)) \leq m_0(V) \vee h_0(V) \), from Lemma 11,

\[
E((1_{X_0=Y_0} V(X_t) + V(Y_t))) \leq \theta_\gamma s \left( \frac{M}{1-\eta} + 1 \right) e^{(\rho + \lambda_\gamma(t-1))t} \int_0^t \|m_u - h_u\|_V du
\]

\[
:= a_1(t) \int_0^t \|m_u - h_u\|_V du.
\]

From (22),

\[
E((1_{X_0\neq Y_0} V(X_t) + V(Y_t))) \leq e^{-\rho t} E((1_{X_0\neq Y_0} V(X_0) + V(Y_0))
\]

\[
+ 2(1 - e^{-\rho t}) \left( \frac{M}{1-\eta} + 1 \right) \mathbb{P}(X_0 \neq Y_0)
\]

\[
\leq \left[ e^{-\rho t} + (1 - e^{-\rho t}) \left( \frac{M}{1-\eta} + 1 \right) \right] \|m_0 - h_0\|_V
\]

\[
:= a_2(t) \|m_0 - h_0\|_V.
\]

Thus

\[
\|m_t - h_t\|_V \leq a_2(t) \|m_0 - h_0\|_V + a_1(t) \int_0^t \|m_u - h_u\|_V du.
\]

As in the proof of Theorem 4, the Gronwall’s Lemma yields, for all \( t \geq 0 \),

\[
\|m_t - h_t\|_V \leq \left( a_2(t) + a_1(t) \int_0^t a_2(s) e^{\int_s^t a_1(u) du} ds \right) \|m_0 - h_0\|_V.
\]

For \( t \geq 0 \) and \( n \geq 0 \), applying \( n \) times this result with the time \( t/n \) we get

\[
\|m_t - h_t\|_V \leq \left( a_2(t/n) + a_1(t/n) \int_0^{t/n} a_2(s) e^{\int_0^{t/n} a_1(u) du} ds \right)^n \|m_0 - h_0\|_V
\]

where we used that, from (8), \( m_{kt/n}(V) \leq M/(1-\eta) + 1 \) for all \( k \in \mathbb{N} \). Letting \( n \) go to infinity,

\[
\|m_t - h_t\|_V \leq e^{(a_2(0)+a_1(0)a_2(0))t} \|m_0 - h_0\|_V,
\]

which concludes.

\[
\square
\]
In the following, as in [30], we consider for a scale parameter $\beta > 0$ the distance on $\mathcal{P}_\nu(E)$ given by $\rho_\beta(\mu_1, \mu_2) = \|\mu_1 - \mu_2\|_{\beta+\nu}$. In other words, denoting $d_\beta(x, y) = \mathbb{1}_{x \neq y}(2\beta + \nu(x) + \nu(y))$ and applying Lemma 8 means that

$$\rho_\beta(\mu_1, \mu_2) = \inf \{ \mathbb{E}(d_\beta(X, Y)) : X \sim \mu_1, Y \sim \mu_2 \} .$$

Lemma 13. Under Assumptions [4] and [7], if $\alpha > 0$ and if $t \mapsto m_t, h_t$ are solutions of (1) with $h_0(\nu) \lor m_0(\nu) \leq M/(1 - \eta) + 1$, then for all $x, y \in E$ such that $x \neq y$,

$$\rho_\beta \left( \delta_x R_{0,t_0}^m, \delta_y R_{0,t_0}^h \right) \leq \kappa d_\beta(x, y)$$

with

$$\beta = \frac{2 \left( 1 - e^{-\rho t_0} \right) e^{\lambda t_0}}{\alpha} \left( \frac{M}{1 - \eta} + 1 \right), \quad \kappa = \frac{1 + e^{-\rho t_0}}{2} \lor \left( 1 - \frac{1}{2} \alpha e^{-\lambda t_0} \right)$$

(23)

Proof. This is essentially the proof of [30] Theorem 3.1. Consider the process $(X_t, Y_t)_{t \geq 0}$ defined in Section 3.2.1 with $X_0 = x, Y_0 = y$, and bound

$$\rho_\beta \left( \delta_x R_{0,t_0}^m, \delta_y R_{0,t_0}^h \right) \leq \mathbb{E}(d_\beta(X_{t_0}, Y_{t_0})) \leq 2 \beta \mathbb{P}(X_{t_0} \neq Y_{t_0}) + R_{0,t_0}^m \nu(x) + R_{0,t_0}^h \nu(y).$$

Then (22) yields

$$R_{0,t_0}^m \nu(x) + R_{0,t_0}^h \nu(y) \leq e^{-\rho t_0} \nu(x) + \nu(y) + 2 \left( 1 - e^{-\rho t_0} \right) \left( \frac{M}{1 - \eta} + 1 \right).$$

Two cases are distinguished. First, if $\nu(x) + \nu(y) \leq 8(1 + M/(1 - \eta))$, similarly to the proof of Theorem 3 from [7],

$$\mathbb{P}(X_{t_0} = Y_{t_0}) \geq \mathbb{P}(Z_{t_0} = \tilde{Z}_{t_0}, T_1 > t_0) \geq \alpha e^{-\lambda t_0},$$

so that

$$\mathbb{E}(d_\beta(X_{t_0}, Y_{t_0})) \leq 2 \beta \left( 1 - \alpha e^{-\lambda t_0} \right) + e^{-\rho t_0} (\nu(x) + \nu(y)) + 2 \left( 1 - e^{-\rho t_0} \right) \left( \frac{M}{1 - \eta} + 1 \right)$$

$$= 2 \beta \left( 1 - \frac{1}{2} \alpha e^{-\lambda t_0} \right) + e^{-\rho t_0} (\nu(x) + \nu(y)),$$

where we used the definition (23) of $\beta$. Second, if $\nu(x) + \nu(y) \geq 8(1 + M/(1 - \eta))$, then

$$\mathbb{E}(d_\beta(X_{t_0}, Y_{t_0})) \leq 2 \beta + e^{-\rho t_0} (\nu(x) + \nu(y)) + 2 \left( 1 - e^{-\rho t_0} \right) \left( \frac{M}{1 - \eta} + 1 \right)$$

$$\leq 2 \beta + \frac{1 + e^{-\rho t_0}}{2} (\nu(x) + \nu(y)) - 2 \left( 1 - e^{-\rho t_0} \right) \left( \frac{M}{1 - \eta} + 1 \right)$$

$$= 2 \beta \left( 1 - \frac{1}{2} \alpha e^{-\lambda t_0} \right) + \frac{1 + e^{-\rho t_0}}{2} (\nu(x) + \nu(y)),$$

where we used again the definitions (23) of $\beta$.

Proof of Theorem 2. Applying (3) with $x = x_0$ and $\mu_s = \delta_{x_0}$ for all $s \geq 0$ where $x_0 \in E$ is such that $\nu(x_0)$ is arbitrarily close to $\inf \nu$, we obtain that $\inf \nu \leq M(1 - \eta)$. In particular, $\mathcal{P}_{\nu'}(E) := \{ \nu \in \mathcal{P}_\nu, \nu(\nu) \leq M(1 - \eta) + 1 \}$ is not empty since it contains $\delta_{x_0}$.
for all $x'_0 \in E$ with $\mathcal{V}(x'_0)$ sufficiently close to $\inf \mathcal{V}$. From (22), if $t \mapsto m_t$ is a solution of \[ \] with $m_0 \in \mathcal{P}_{\mathcal{V},1}(E)$ and then $m_t \in \mathcal{P}_{\mathcal{V},1}(E)$ for all $t \geq 0$.

Let $m_0, h_0 \in \mathcal{P}_{\mathcal{V},1}(E)$ and $(X_0, Y_0)$ be an optimal coupling of $m_0$ and $h_0$ given by Lemma 8, i.e. be such that, for any $\beta > 0$,

$$
\rho_\beta(m_0, h_0) = \mathbb{E}(d_\beta(X_0, Y_0)).
$$

Conditioning on the initial value, let $(X_{t_0}, Y_{t_0})$ be an optimal coupling of $\delta_{X_0}R_{0,t_0}^m$ and $\delta_{Y_0}R_{0,t_0}^h$, so that

$$
\mathbb{E}(d_\beta(X_{t_0}, Y_{t_0})) = \mathbb{E}\left(\rho_\beta\left(\delta_{X_0}R_{0,t_0}^m, \delta_{Y_0}R_{0,t_0}^h\right)\right) = \mathbb{E}\left(\mathbb{E}(1_{X_0 = Y_0} + 1_{X_0 \neq Y_0}) \rho_\beta\left(\delta_{X_0}R_{0,t_0}^m, \delta_{Y_0}R_{0,t_0}^h\right)\right).
$$

Considering $\kappa$ and $\beta$ such as defined in Lemma 13 and using Lemma 11 and the fact that $d_\beta \leq (1 + \beta)d_\nu$,

$$
\mathbb{E}(d_\beta(X_t, Y_t)) \leq \kappa \mathbb{E}(d_\beta(X_0, Y_0)) + 2(1 + \beta)\theta\gamma s(M(1 - \eta) + 1) e^{(\nu_\beta + \lambda_\nu(\gamma - 1))t_0} \int_0^{t_0} \|m_u - h_u\|d\nu.
$$

Together with Lemma 12 and the fact that $\|\cdot\|_\nu \leq \rho_\beta$ (since $d_\nu \leq d_\beta$), this means

$$
\rho_\beta(m_{t_0}, h_{t_0}) \leq \kappa \rho_\beta(m_0, h_0) \left[\kappa + \theta e^{(\nu_\beta + \lambda_\nu(\gamma - 1))t_0} \left(e^{(\nu + \theta\gamma s)}(\frac{M}{1 - \eta} + 1) - 1\right)\right]
$$

By assumption, $\kappa < 1$, so that $\Psi_{t_0} : \mathcal{P}_{\mathcal{V},1}(E) \rightarrow \mathcal{P}_{\mathcal{V},1}(E)$ that maps $m_0$ to $m_{t_0}$ where $t \mapsto m_t$ is a solution of \[ \] is a contraction. If a sequence $(\mu_n)_{n \in \mathbb{N}}$ in $\mathcal{P}$ is such that $\rho_\beta(\mu_n, \nu) \rightarrow 0$ for some $\nu \in \mathcal{P}$ as $n \rightarrow \infty$, then $\mu_n(\nu) \rightarrow \nu(\mathcal{V})$. Hence, $\mathcal{P}_{\mathcal{V},1}(E)$ is a closed subset of $\mathcal{P}(E)$ endowed with the metric $\rho_\beta$, which is complete. As a consequence, $\Psi_{t_0}$ admits a unique fixed point $\mu_\infty$ in $\mathcal{P}_{\mathcal{V},1}(E)$, which for now may depend on $t_0$.

From Lemma 8, for all $n \in \mathbb{N}$ and $m_0 \in \mathcal{P}_{\mathcal{V}}(E)$,

$$
m_{nt_0}(\mathcal{V}) \leq e^{-\nu(1 - \eta)nt_0}m_0(\mathcal{V}) + \left(1 - e^{-\nu(1 - \eta)nt_0}\right)\frac{M}{1 - \eta},
$$

which, applied to $m_0 = \mu_\infty$ and letting $n$ go to infinity implies that $\mathcal{V}(\mu_\infty) \leq M/(1 - \eta)$. More generally, Lemma 8 means that for all $m_0 \in \mathcal{P}_{\mathcal{V}}(E)$ and all $t \geq s_0 := \ln(m_0(\mathcal{V}))/\nu(1 - \eta)$, $m_t \in \mathcal{P}_{\mathcal{V},1}(E)$. For all $\nu \in \mathcal{P}_{\mathcal{V},1}(E)$,

$$
\rho_\beta(\mu_\infty, \nu) \leq 2\beta + 2\frac{M}{1 - \eta} + 2.
$$

Combining this with (24), we obtain that, for all $m_0 \in \mathcal{P}_{\mathcal{V}}(E)$,

$$
\rho_\beta(\mu_\infty, m_t) \leq \tilde{\kappa}^{(t - s_0)/t_0} \rho_\beta(\mu_\infty, m_{t_0 - s_0}/t_0) \leq 2\tilde{\kappa}^{(t - s_0)/t_0 - 1}\left(\beta + 1 + \frac{M}{1 - \eta}\right).
$$

Since $\tilde{\kappa} \geq e^{-\nu(1 - \eta)t_0}$, $\tilde{\kappa}^{-s_0/t_0} \leq m_0(\mathcal{V})$. We have then obtained, for all $t \geq 0$ and all $m_0 \in \mathcal{P}_{\mathcal{V}}$,

$$
\rho_\beta(\mu_\infty, m_t) \leq 2\tilde{\kappa}^{t/t_0 - 1}\left(\beta + 1 + \frac{M}{1 - \eta}\right)m_0(\mathcal{V}).
$$

In particular, if $m_0 = \mu_\infty$, since in that case $m_s = h_{s + nt_0}$ for all $s \geq 0$ and $n \in \mathbb{N}$, letting $n$ go to infinity, we get that $m_s = \mu_\infty$ for all $s \geq 0$, in other words $\mu_\infty$ is an equilibrium of \[ \] Finally, $\rho_\beta \geq \|\cdot\|_\nu$, which concludes. \qed
3.3 Study of the particle system

We use in this section the notations of Section 2.2. In particular for $i \in [1, N]$ we consider a Markov semi-group $(P^i_t)_{t \geq 0}$ and jump rate and kernel $\lambda_i, Q_i$.

3.3.1 Coupling for interacting particles

For all $x, y \in E$ and $i \in [1, N]$, denote

$$\tilde{Q}_i(x) = \frac{\lambda_i(x)}{\lambda_s} Q_i(x) + \left(1 - \frac{\lambda_i(x)}{\lambda_s}\right) \delta_{x_i},$$

and $p_i(x, y) = (\tilde{Q}_i(x) \wedge \tilde{Q}_i(y))(E_t)$. If $p_i(x, y) = 0$, set

$$Q^\sim_i(x, y) = \delta_{x_i}, \quad Q^\wedge_i(x, y) = \tilde{Q}_i(x).$$

If $p_i(x, y) \in (0, 1)$, set

$$Q^\sim_i(x, y) = \frac{\tilde{Q}_i(x) \wedge \tilde{Q}_i(y)}{p_i(x, y)}, \quad Q^\wedge_i(x, y) = \frac{\tilde{Q}_i(x) - \tilde{Q}_i(x) \wedge \tilde{Q}_i(y)}{1 - p_i(x, y)}.$$

Finally, if $p_i(x, y) = 1$, set

$$Q^\sim_i(x, y) = \tilde{Q}_i(x), \quad Q^\wedge_i(x, y) = \delta_{x_i}.$$

Then $Q^\sim_i$ and $Q^\wedge_i$ are Markov kernels from $E^2$ to $\mathcal{P}(E_t)$ such that, for all $x, y \in E$,

$$\tilde{Q}_i(x) = p_i(x, y) Q^\sim_i(x, y) + (1 - p_i(x, y)) Q^\wedge_i(x, y),$$

and, as shown in the proof of Lemma 8, $||\tilde{Q}_i(x) - \tilde{Q}_i(y)||_{TV} = 2(1 - p_i(x, y))$, so that (10) reads

$$\forall x, y \in E \text{ such that } x_i = y_i, \quad 1 - p_i(x, y) \leq \frac{\theta d_i(x, y)}{2N\lambda_s},$$

and similarly for (15). For all $i \in [1, N]$, let $G^\sim_i : E^2 \times [0, 1] \to E_t$ (resp. $G^\wedge_i$) be a representation of $Q^\sim_i$ (resp. $Q^\wedge_i$), in the sense that for all $x, y \in E$, if $U$ is a r.v. uniformly distributed over $[0, 1]$, then $G^\sim_i(x, y; U) \sim Q^\sim_i(x, y)$.

Let $x, y \in E$. We define a Markov process $(X_i, Y_i)_{t \geq 0} = (X_{i, t}, Y_{i, t})_{i \in [1, N], t \geq 0}$ on $E^2$ and an auxiliary Markov process $(J_{i, t})_{t \geq 0}$ on $[1, N]$ as follows. For a given $t_0$ and all $i \in [1, N]$ let $(Z_{i, t}, \bar{Z}_{i, t})_{t \geq 0}$ be an optimal coupling of $(\delta_i P^i_{1,t})_{t \geq 0}$ and $(\delta_j P^i_{1,t})_{t \geq 0}$ in the sense of Lemma 9 independent one from the other for $j \neq i$. Consider an i.i.d. sequence $(S_k, (U_{i,k})_{i \in [1, N]}, V_k, W_k, I_k)_{k \in \mathbb{N}}$, independent from $(X_0, Y_0)$ and $(Z_{i, t}, \bar{Z}_{i, t})_{t \geq 0}, i \in [1, N]$ and where, for all $k \in \mathbb{N}$, $S_k, V_k, W_k, I_k$ and $(U_{i,k})_{i \in [1, N]}$ are independent one from the other, $S_k$ follows a standard exponential law with parameter 1, $V_k$ and $W_k$ (resp. $I_k$) a uniform distribution over $[0, 1]$ (resp. over $[1, N]$), and $(U_{i,k})_{i \in [1, N]}$ are i.i.d. r.v. uniformly distributed over $[0, 1]$. Set $T_0 = 0, X_0 = x, Y_0 = y$ and $J_0 = d_1(x, y)/2$ and suppose that $T_n \geq 0$ and $(X_i, Y_i, J_i)_{i \in [0, T_n]}$ have been defined for some $n \in \mathbb{N}$ and are independent from $(S_k, (U_{i,k})_{i \in [1, N]}, V_k, W_k, I_k)_{k \geq n}$. Set $T_{n+1} = T_n + (N\lambda_s)^{-1} S_n$.

- For all $i \in [1, N]$ and all $t \in (T_n, T_{n+1})$, set $J_t = J_{T_n}$,

$$(X_{i,t}, Y_{i,t}) = \begin{cases} (Z_{i,t}, \bar{Z}_{i,t}) & \text{if } (X_{i,T_n}, Y_{i,T_n}) = (Z_{i,T_n}, \bar{Z}_{i,T_n}) \\ (H^i_{t-T_n}(X_{i,T_n}, U_{i,n}), H^i_{t-T_n}(Y_{i,T_n}, U_{i,n})) & \text{else}, \end{cases}$$

and

$$(\bar{X}_{i,T_{n+1}}, \bar{Y}_{i,T_{n+1}}) = \begin{cases} (Z_{i,T_n+1}, \bar{Z}_{i,T_n+1}) & \text{if } (X_{i,T_n}, Y_{i,T_n}) = (Z_{i,T_n}, \bar{Z}_{i,T_n}) \\ (H^i_{T_{n+1}-T_n}(X_{i,T_n}, U_{i,n}), H^i_{T_{n+1}-T_n}(Y_{i,T_n}, U_{i,n})) & \text{else}. \end{cases}$$
• If \( V_n \leq p_{I_n} \left( \tilde{X}_{T_{n+1}}, \tilde{Y}_{T_{n+1}} \right) \) then set

\[
X_{T_{n+1}} = Y_{T_{n+1}} = G_{I_n}^T \left( \tilde{X}_{T_{n+1}}, \tilde{Y}_{T_{n+1}}, W_n \right),
\]

and else set

\[
X_{T_{n+1}} = G_{I_n}^T \left( \tilde{X}_{T_{n+1}}, \tilde{Y}_{T_{n+1}}, W_n \right), \quad Y_{T_{n+1}} = G_{I_n}^T \left( \tilde{Y}_{T_{n+1}}, \tilde{X}_{T_{n+1}}, W_n \right).
\]

• If \( V_n \geq 1 - \theta J_{T_n} / (N \lambda_n) \) and \( x_{I_n} = y_{I_n} \) then set \( J_{T_{n+1}} = J_{T_n} + 1 \), else set \( J_{T_{n+1}} = J_{T_n} \). Then \((X_t, Y_t, J_t)\) is defined for all \( t \geq 0 \).

**Proposition 14.** For all \( t \geq 0 \), \( X_t \sim \delta_{x_t} \), \( Y_t \sim \delta_{y_t} \) and \( 2J_t \geq \tilde{d}_1(X_t, Y_t) \).

**Proof.** The proof of the first statements is similar to the proof of Proposition 10 hence omitted. For the last one, we note that

\[
J_t \geq \tilde{J}_t := \frac{1}{2} \tilde{d}_1(x, y) + \sharp \{ i \in [1, N] : x_i = y_i, X_{i,t} \neq Y_{i,t} \}.
\]

Indeed, it is clear by definition that \( 2\tilde{J}_t \geq \tilde{d}_1(X_t, Y_t) \) for all \( t \geq 0 \). Since \( J_0 = \tilde{J}_0 \), suppose that \( J_{T_n} \geq J_{T_n} \) for some \( n \in \mathbb{N} \). Then \( J_t = J_{T_n} \geq \tilde{J}_{T_n} = \tilde{J}_t \) for all \( t \in [T_n, T_{n+1}) \). From

\[
\tilde{d}_1 \left( \tilde{X}_{T_{n+1}}, \tilde{Y}_{T_{n+1}} \right) \leq \tilde{d}_1 \left( X_{T_n}, Y_{T_n} \right) \leq 2\tilde{J}_T,
\]

we get that

\[
\tilde{J}_{T_{n+1}} = \tilde{J}_{T_n} + \sharp \{ i \in [1, N] : x_i = y_i \} \leq \tilde{J}_{T_n} + \sharp \{ i \in [1, N] : x_i = y_i, X_{i,t} \neq Y_{i,t} \} \leq J_t - \sharp A_t.
\]

which concludes. \( \square \)

### 3.3.2 The total variation case

**Proof of Theorem 6** Keep all the notations of Section 3.3.1 and, for \( t \leq 0 \), let

\[
\mathcal{A}_t = \{ i \in \mathbb{N} : Z_{i,t} = \tilde{Z}_{i,t}, x_i \neq y_i \text{ and } I_n \neq i \forall n \in \mathbb{N} \text{ such that } T_n < t \}.
\]

Remark that, if \( i \in \mathcal{A}_t \), then \( X_{i,t} = Y_{i,t} \), so that

\[
\frac{1}{2} \tilde{d}_1(X_t, Y_t) = \frac{1}{2} \tilde{d}_1(x, y) + \sharp \{ i \in [1, N] : x_i = y_i, X_{i,t} \neq Y_{i,t} \} - \sharp \{ i \in [1, N] : x_i = y_i, X_{i,t} = Y_{i,t} \} \leq J_t - \sharp \mathcal{A}_t.
\]

For all \( i \in \mathbb{N} \), \( \mathcal{T}_i = \{ T_n > 0 : I_n = i, n \in \mathbb{N} \} \) are the jump times of a Poisson process of intensity \( \lambda_i \), independent from \( \mathcal{T}_j \) for \( j \neq i \) and from \((Z_{i,t}, \tilde{Z}_{i,t})_{t \geq 0}\). As a consequence, for all \( i \in [1, N] \) with \( x_i \neq y_i \), \( 1_{i \in \mathcal{A}_t} \) is a Bernoulli r.v. with parameter \( \exp(-\lambda_i t) \mathbb{P}(Z_{i,t} = \tilde{Z}_{i,t}) \), independent from \( 1_{j \in \mathcal{A}_i} \) if \( j \neq i \) and from \( J_t \). In particular,

\[
\mathbb{E}(\sharp \mathcal{A}_0) \geq \frac{1}{2} \tilde{d}_1(x, y) \exp(-\lambda_i t_0) \alpha.
\]

On the other hand, the generator of the Markov process \((J_t)_{t \geq 0}\) is

\[
K f(s) = \theta s \left( 1 - \frac{\tilde{d}_1(x, y)}{2N} \right) (f(s + 1) - f(s)).
\]
Applied to $f(s) = s$, this yields
\[
\partial_t \mathbb{E}(J_t) = \theta \left( 1 - \frac{d_1(x,y)}{2N} \right) \mathbb{E}(J_t) \leq \theta \mathbb{E}(J_t),
\]
and thus
\[
\mathbb{E}(\tilde{d}_1(X_{t_0},Y_{t_0})) \leq \tilde{d}_1(x,y)e^{\theta t_0} - \alpha e^{-\lambda_s t_0}\tilde{d}_1(x,y).
\]
On the other hand, for all $t \in [0,t_0]$,\[
\mathbb{E}(\tilde{d}_1(X_t,Y_t)) \leq 2\mathbb{E}(J_t) \leq \tilde{d}_1(x,y)e^{\theta t_0}.
\]
Considering on $\mathcal{P}(E)$ the distance $\rho$ defined by
\[
\rho(\mu_1,\mu_2) = \inf \left\{ \mathbb{E}(\tilde{d}_1(X,Y)) : X \sim \mu_1, Y \sim \mu_2 \right\},
\]
we have thus obtained that, for all $t \in [0,t_0]$ and $x, y \in E$,
\[
\rho(\delta_x R_t, \delta_y R_t) \leq \left( e^{t_0} - 1 - t_0 \alpha e^{-\lambda_s t_0} \right) \tilde{d}_1(x,y).
\]
Finally, for all $t \geq 0$ and $x, y \in E$,
\[
\|\delta_x R_t - \delta_y R_t\|_{TV} \leq \rho(\delta_x R_t, \delta_y R_t)
\leq \tilde{d}_1(x,y)e^{t_0} \left( e^{t_0} - \alpha e^{-\lambda_s t_0} \right)^{\lfloor t/t_0 \rfloor}
\leq 2N \|_{x \neq y} e^{t_0} \left( e^{t_0} - \alpha e^{-\lambda_s t_0} \right)^{\lfloor t/t_0 \rfloor},
\]
which, integrated with respect to any initial distribution, concludes the proof of the first statement of Theorem 8.

Now, let $m_0'$ and $h_0'$ be the respective distributions of $X_{I,t}$ and $Y_{I,t}$, where $I$ is independent from $(X_t,Y_t)_{t \geq 0}$. Then
\[
\|m_0' - h_0'\|_{TV} \leq 2\mathbb{P}(X_{I,t} \neq Y_{I,t}) = \frac{1}{N} \mathbb{E}(\tilde{d}_1(X_t,Y_t)).
\]
Considering an optimal coupling $(X_0,Y_0)$ of $m_0$ and $h_0$ and, conditionally to $(X_0,Y_0)$, an optimal coupling $(X_t,Y_t)$ of $\delta_{X_0} R_t$ and $\delta_{Y_0} R_t$,\[
\mathbb{E}(\tilde{d}_1(X_t,Y_t)) \leq \mathbb{E}(\tilde{d}_1(X_0,Y_0)) e^{t_0} \left( e^{t_0} - \alpha e^{-\lambda_s t_0} \right)^{\lfloor t/t_0 \rfloor}
= 2N \mathbb{P}(X_{I,0} \neq Y_{I,0}) e^{t_0} \left( e^{t_0} - \alpha e^{-\lambda_s t_0} \right)^{\lfloor t/t_0 \rfloor}
= N \|m_0' - h_0'\|_{TV} e^{t_0} \left( e^{t_0} - \alpha e^{-\lambda_s t_0} \right)^{\lfloor t/t_0 \rfloor},
\]
which concludes.\hfill \Box

### 3.3.3 The general $\mathcal{V}$ case

In all this section, Assumptions 5 and 7 are enforced. For $i \in \{1,N\}$ and $x \in E$, set
\[
\mathcal{W}_i(x) = \mathcal{V}_i(x_i) + \frac{2\eta}{N(1-\eta)} \mathcal{V}(x).
\]

**Lemma 15.** For all $t \geq 0$ and $x \in E$, and $i \in \{1,N\}$,
\[
R_t \mathcal{V}(x) \leq e^{-\rho(1-\eta)t} \mathcal{V}(x) + \left( 1 - e^{-\rho(1-\eta)t} \right) \frac{NM}{1-\eta},
\]
\[
R_t \mathcal{W}_i(x) \leq e^{-\frac{1}{2}\rho(1-\eta)t} \mathcal{W}_i(x) + \frac{2(1+\eta)M}{(1-\eta)^2} \left( 1 - e^{-\frac{1}{2}\rho(1-\eta)t} \right).
\]

21
Proof. Summing (12) over \( i \in \llbracket 1, N \rrbracket \) reads

\[
R_t \mathcal{V}(x) \leq e^{-\rho t} \mathcal{V}(x) + \rho \int_0^t e^{\rho(s-t)} (\eta R_s \mathcal{V}(x + NM) \, ds
\]

for all \( t \geq 0 \) and \( x \in E \). The Gronwall’s Lemma then yields (26) which, reintegrated in (12), gives

\[
R_t \mathcal{W}_i(x) \leq e^{-\rho t} \mathcal{W}_i(x_i) + e^{-\rho(1-\eta)t} \frac{2 \eta \mathcal{V}(x)}{N(1-\eta)} + \left(1 - e^{-\rho(1-\eta)t}\right) \frac{2 \eta M}{(1-\eta)^2} \\
+ \rho \int_0^t e^{\rho(s-t)} \left( \eta e^{-\rho(1-\eta)s} \left( \frac{\mathcal{V}(x)}{N} - \frac{M}{1-\eta} \right) + \frac{M}{1-\eta} \right) \, ds
\]

which is (27).

Letting \( n \rightarrow \infty \),

\[
R_t \mathcal{W}_i(x) \leq a(t) R_s \mathcal{W}_i(x) + b(t)
\]

with

\[
a(t) = e^{-\rho(1-\eta)t} + \frac{1-\eta}{2\eta} \left( e^{-\rho(1-\eta)t} - e^{-\rho t} \right)
\]

\[
b(t) = \frac{(1+\eta)M}{(1-\eta)^2} \left( 1 - e^{-\rho(1-\eta)t} \right)
\]

Applying \( R_s \) for some \( s \geq 0 \) to this inequality and using the semi-group property, we have thus obtained for all \( s, t \geq 0 \)

\[
R_{t+s} \mathcal{W}_i(x) \leq a(t) R_s \mathcal{W}_i(x) + b(t).
\]

Hence, for all \( n \in \mathbb{N} \) and \( t \geq 0 \),

\[
R_t \mathcal{W}_i(x) \leq a(t/n) R_{t/n} \mathcal{W}_i(x) + b(t/n) \leq \ldots \leq (a(t/n))^n \mathcal{W}_i(x) + b(t/n) \frac{1-(a(t/n))^n}{1-a(t/n)}.
\]

Letting \( n \) go to infinity yields

\[
R_t \mathcal{W}_i(x) \leq e^{\alpha(t) \mathcal{W}_i(x)} + b(0) \frac{\alpha(0)}{\alpha(0)} \left( 1 - e^{\alpha(0)t} \right),
\]

which is (27).

Similarly to (30) or Section 3.2.3, we now consider some parameter \( \beta > 0 \) and, for all \( x, y \in E \) and \( i \in \llbracket 1, N \rrbracket \),

\[
d_{i,\beta}(x, y) = 1_{x_i \neq y_i} (2\beta + W_i(x) + W_i(y)).
\]

Lemma 16. For all \( i \in \llbracket 1, N \rrbracket \) and all \( x, y \in E \) with \( x_i \neq y_i \), let \( (X_t, Y_t)_{t \geq 0} \) be the coupling of \( (\delta_x R_t)_{t \geq 0} \) and \( (\delta_y R_t)_{t \geq 0} \) defined in Section 3.3.1. Then

\[
\mathsf{E} (d_{i,\beta}(X_{t_0}, Y_{t_0})) \leq \kappa d_{i,\beta}(x, y)
\]

with

\[
\beta = \frac{4(1+\eta)Me^{\lambda_{t_0}}}{\alpha(1-\eta)^2} \left( 1 - e^{-\frac{1}{2} \rho(1-\eta)t_0} \right)
\]

\[
\kappa = \frac{1 + e^{-\frac{1}{2} \rho(1-\eta)t_0}}{2} \sqrt{1 - \frac{1}{2} \alpha e^{-\lambda t_0}}.
\]
Proof. This is again essentially the proof of [30, Theorem 3.1]. We bound
\[ \mathbb{E}(d_{i, \beta}(X_{i, t}, Y_{i, t})) \leq 2\beta \mathbb{P}(X_{i, t_0} \neq Y_{i, t_0}) + R_{t_0} W_i(x) + R_{t_0} W_i(y). \]
From [27],
\[ R_{t_0} W_i(x) + R_{t_0} W_i(y) \leq e^{-\frac{1}{2} \rho(1-\eta) t_0} (W_i(x) + W_i(y)) + \frac{4(1 + \eta) M}{(1 - \eta)^2} \left( 1 - e^{-\frac{1}{2} \rho(1-\eta) t_0} \right). \]
Two cases are distinguished. First, if \( V_i(x_i) + V_i(y_i) \leq 16(1 + \eta) M / (1 - \eta)^2 \), as in the previous section, from [16],
\[ \mathbb{P}(X_{i, t_0} = Y_{i, t_0}) \geq \mathbb{P}(Z_{i, t_0} = Z_{i, t_0} \text{ and } I_n \neq i \forall n \text{ such that } T_n < t) \geq \alpha e^{-\lambda_i t_0}, \]
so that
\[ \mathbb{E}(d_{i, \beta}(X_{i, t_0}, Y_{i, t_0})) \leq 2\beta \left( 1 - \alpha e^{-\lambda_i t_0} \right) + e^{-\frac{1}{2} \rho(1-\eta) t_0} (W_i(x) + W_i(y)) + \frac{4(1 + \eta) M}{(1 - \eta)^2} \left( 1 - e^{-\frac{1}{2} \rho(1-\eta) t_0} \right) \]
\[ \leq 2\beta \left( 1 - \frac{1}{2} \alpha e^{-\lambda_i t_0} \right) + e^{-\frac{1}{2} \rho(1-\eta) t_0} (W_i(x) + W_i(y)) \]
where we used the definition (28) of \( \beta \). Second, if \( V_i(x_i) + V_i(y_i) \geq 16(1 + \eta) M / (1 - \eta)^2 \), since \( W_i \geq V_i \), then
\[ \mathbb{E}(d_{i, \beta}(X_{i, t_0}, Y_{i, t_0})) \leq 2\beta + e^{-\frac{1}{2} \rho(1-\eta) t_0} (W_i(x) + W_i(y)) + \frac{4(1 + \eta) M}{(1 - \eta)^2} \left( 1 - e^{-\frac{1}{2} \rho(1-\eta) t_0} \right) \]
\[ \leq 2\beta + \frac{1 + e^{-\frac{1}{2} \rho(1-\eta) t_0}}{2} (W_i(x) + W_i(y)) - \frac{4(1 + \eta) M}{(1 - \eta)^2} \left( 1 - e^{-\frac{1}{2} \rho(1-\eta) t_0} \right) \]
\[ = 2\beta \left( 1 - \frac{1}{2} \alpha e^{-\lambda_i t_0} \right) + \frac{1 + e^{-\frac{1}{2} \rho(1-\eta) t_0}}{2} (W_i(x) + W_i(y)) \]
where we used again the definition (28) of \( \beta \).

Lemma 17. For all \( i \in [1, N] \), \( \beta > 0 \), \( t \geq 0 \) and \( x, y \in E \) with \( x_i = y_i \), if \( (X_t, Y_t)_{t \geq 0} \) is the coupling of \( (\delta_t R_t)_{t \geq 0} \) and \( (\delta_y R_t)_{t \geq 0} \) defined in Section 3.3.1, then
\[ \mathbb{E}(d_{i, \beta}(X_t, Y_t)) \leq \frac{\bar{d}_i(x, y)}{2N} \left( e^{\beta t} - 1 \right) (2 \beta + c_s(t) (W_i(x) + W_i(y))), \]
with \( c_s(t) = (\gamma_s + 1) \exp((\rho_s + \lambda_s (\gamma_s - 1)) t) \).

Proof. Let \( t_1 \geq 0 \) and \( i \in [1, N] \) be such that \( x_i = y_i \). For all \( j \in [1, N] \), let \( \tau_j = \inf\{ T_n \geq 0 : n \in N, I_n = j, V_n \geq 1 - \theta J_{t_n} / N \} \). By construction and Lemma [14], \( X_{i, t} = Y_{i, t} \) for all \( t \leq \tau_i \), so that
\[ \mathbb{E}(d_{i, \beta}(X_{i, t}, Y_{i, t})) \leq \mathbb{E}(1_{\tau_i \leq t_1} (2 \beta + W_i(X_{i, t}) + W_i(Y_{i, t}))). \]
Let \( \Gamma = (S_k, V_k, I_k)_{k \in \mathbb{N}}, K = \sharp\{ n \in \mathbb{N} : T_n \leq t_1 \} \) and for all \( j \in [1, N] \), \( A_j = \{ n \in \mathbb{N} : T_n \leq t_1, I_n = j \} \) and \( K_j = \sharp A_j \). Note that \( K \) and \( \{ \tau_j, K_j \}_{j \in [1, N]} \) are deterministic.
functions of $\Gamma$, and in particular are independent from $(U_{i,k})_{i \in [1,N], k \in \mathbb{N}}$. Hence, for all $n \in \mathbb{N}$ and $j \in [1,N]$, using (13) and (14),

$$
\mathbb{E}(\mathcal{V}_j(X_j,t_n) \mid \Gamma) \leq (\mathbb{I}_{I_n \neq j} + \gamma_n \mathbb{I}_{I_n = j}) \mathbb{E}(\mathcal{V}_j(X_j,t_n) \mid \Gamma)
= (\mathbb{I}_{I_n \neq j} + \gamma_n \mathbb{I}_{I_n = j}) \mathbb{E}(P_{t_n-T_n-1}^{j} \mathcal{V}_j(X_j,t_{n-1}) \mid \Gamma)
\leq (\mathbb{I}_{I_n \neq j} + \gamma_n \mathbb{I}_{I_n = j}) e^{\rho_\ast(T_n-T_{n-1})} \mathbb{E}(\mathcal{V}_j(X_j,t_{n-1}) \mid \Gamma).
$$

Similarly, and then by direct induction,

$$
\mathbb{E}(\mathcal{V}_j(X_j,t_1) \mid \Gamma) \leq e^{\rho_\ast(t_1-T_K)} \mathbb{E}(\mathcal{V}_j(X_j,t_K) \mid \Gamma)
\leq e^{\rho_\ast t_1 \gamma_{K_j}} \mathcal{V}_j(x_j).
$$

We have thus obtained for all $j \in [1,N]$

$$
\mathbb{E}(1_{X_{j,t_1} \neq Y_{j,t_1}} \mathcal{V}_j(X_j,t_1)) \leq e^{\rho_\ast t_1} \mathcal{V}_j(x_j) e^{\gamma_{\ast} \gamma_j}.
$$

For $j \in [1,N]$, set $\Gamma_j = \{(S_k, V_k, I_k) : I_k = j, k \in \mathbb{N}\}$. Remark that $(J_{t})_{t \geq 0}$ is a deterministic function of $\{\Gamma_j : j \in [1,N], x_j = y_j\}$ and that, by Poisson thinning, $\Gamma_j$ (hence $K_j$) is independent from $\Gamma_k$ for all $k \neq j$. In particular, if $j \in [1,N]$ is such that $x_j \neq y_j$,

$$
\mathbb{E}(1_{\tau_1 \leq t_1} \gamma_{\ast}^{K_j}) = \mathbb{P}(\tau_1 \leq t_1) \mathbb{E}(\gamma_{\ast}^{K_j}) = \mathbb{P}(\tau_1 \leq t_1) e^{\lambda_\ast(\gamma_{\ast}-1)t_1}.
$$

For $j \in [1,N]$ with $x_j \neq y_j$ we define the process $(J_{j,t})_{t \geq 0}$ as follows. Set $J_{j,0} = 0$ and, for all $n \in \mathbb{N}$ and $t \in (T_n, T_{n+1})$, set $J_{j,t} = J_{j,T_n}$ and

$$
J_{j,T_{n+1}} = J_{j,T_n} + 1_{I_n \neq j} 1_{\tau_n \geq 1-\theta(J_{j,T_n}/(N\lambda_\ast))}.
$$

In other words, $(J_{j,t})_{t \geq 0}$ is like $(J_t)_{t \geq 0}$ but ignoring the jumps of the $j^{th}$ particle. By thinning of Poisson processes, $(J_{j,t})_{t \geq 0}$ is independent from $\Gamma_j$. Moreover $J_{j,t} \leq J_t$ for all $t$ and, conditionnally to $\tau_j > t$, $J_{j,t} = J_t$. As a consequence, if $i \neq j$, we bound

$$
\mathbb{P}(\tau_i \leq t_1, \tau_j > t_1 \mid K_j) = \mathbb{P}(\tau_j > t_1, \exists n \in \mathcal{A}_i \text{ s.t. } V_n \geq 1 - \theta J_{j,T_n}/(N\lambda_\ast) \mid K_j)
\leq \mathbb{P}(\exists n \in \mathcal{A}_i \text{ s.t. } V_n \geq 1 - \theta J_{j,T_n}/(N\lambda_\ast) \mid K_j)
= \mathbb{P}(\exists n \in \mathcal{A}_i \text{ s.t. } V_n \geq 1 - \theta J_{j,T_n}/(N\lambda_\ast))
\leq \mathbb{P}(\tau_i \leq t_1),
$$

and thus, for all $j \neq i$ with $x_j = y_j$,

$$
\mathbb{E}(\mathbb{I}_{\tau_i \leq t_1} \gamma_{\ast}^{K_j}) \leq \mathbb{E}(\mathbb{I}_{\tau_j \leq t_1} \gamma_{\ast}^{K_j}) + \mathbb{E}(\mathbb{I}_{\tau_i \leq t_1} \mathbb{I}_{\tau_j > t_1} \gamma_{\ast}^{K_j})
\leq \mathbb{E}(\mathbb{I}_{\tau_j \leq t_1} \gamma_{\ast}^{K_j}) + \mathbb{P}(\tau_i \leq t_1) \mathbb{E}(\gamma_{\ast}^{K_j})
= \mathbb{E}(\mathbb{I}_{\tau_j \leq t_1} \gamma_{\ast}^{K_j}) + \mathbb{P}(\tau_i \leq t_1) e^{\lambda_\ast(\gamma_{\ast}-1)t_1}.
$$

Integrating (32) and (33) in (31) and then in (30) by definition of $\mathcal{W}_i$, we have obtained so far that

$$
\mathbb{E}(d_{i,\beta}(X_{t_1}, Y_{t_1})) \leq \mathbb{P}(\tau_i \leq t_1) \left(2\beta + \frac{2\eta_\rho(\rho_\ast + \lambda_\ast(\gamma_{\ast}-1))t_1}{N(1-\eta)}(\mathcal{V}(x) + \mathcal{V}(y)) \right)
+ 2e^{\rho_\ast t_1} \mathcal{V}_i(x_i) \mathbb{E}(\mathbb{I}_{\tau_i \leq t_1} \gamma_{\ast}^{K_j}) + \frac{4\eta_\rho t_1}{N(1-\eta)} \sum_{j=1}^{N} \mathbb{I}_{x_j = y_j} \mathcal{V}_j(x_j) \mathbb{E}(\mathbb{I}_{\tau_j \leq t_1} \gamma_{\ast}^{K_j}).
$$

(34)
Using that \((J_i, t) \geq 0\), \(K_i\) and \((V_n)_{n \in A_i}\) are independent r.v. and that, conditionally to \(K_i\), \(\{T_n : n \in A_i\}\) are distributed like \(K_i\) independent r.v. uniformly distributed over \([0, t_1]\),

\[
\mathbb{P}(\tau_i > t_1 | K_i) = \mathbb{P}\left( V_n \leq 1 - \theta J_{i, T_n} / (N \lambda_s), \forall n \in A_i | K_i \right) \\
= \left( 1 - \frac{1}{t_1} \int_0^{t_1} 1 \wedge \mathbb{E}(\theta J_{i, w} / (N \lambda_s)) \, dw \right) K_i \\
\geq \left( 1 - \frac{1}{t_1} \int_0^{t_1} 1 \wedge \mathbb{E}(\theta J_{w} / (N \lambda_s)) \, dw \right) K_i := s^{K_i}.
\]

With the same computations as in the proof of Lemma 16 this leads to

\[
\mathbb{E}\left( \mathbf{1}_{\tau_i \leq t_1} \gamma_s^{K_i} \right) \leq \lambda_s t_1 \gamma_s (1 - s)e^{\lambda_s (\gamma_s - 1)t_1} \\
\leq \gamma_s e^{\lambda_s (\gamma_s - 1)t_1} \frac{\bar{d}_1(x, y)}{2N} \int_0^{t_1} \theta e^{\theta u} \, du,
\]

where we used that, from (25), \(\mathbb{E}(J_t) \leq \exp(\theta t) \bar{d}_1(x, y)/2\) for all \(t \geq 0\). The case of \(j \neq i\) with \(x_j = y_j\) is identical and, similarly (i.e. applying this bound with \(\gamma_s = 1\)),

\[
\mathbb{P}(\tau_i \leq t_1) \leq \frac{\bar{d}_1(x, y)}{2N} \left( e^{\theta t_1} - 1 \right).
\]

Together with (54), this yields

\[
\mathbb{E}(d_{i, \beta}(X_{t_1}, Y_{t_1})) \leq \frac{\bar{d}_1(x, y)}{2N} \left( e^{\theta t_1} - 1 \right) \left( 2\beta + (\gamma_s + 1) e^{(\rho_s + \lambda_s (\gamma_s - 1)t_1)} (W_i(x) + W_i(y)) \right).
\]

**Proof of Theorem 7**. Let \(\beta\) and \(\kappa\) be given by Lemma 16 and consider on \(\mathcal{P}(E)\) the distance \(\bar{\rho}_\beta\) given for all \(\mu, \nu \in \mathcal{P}(E)\) by

\[
\bar{\rho}_\beta(\mu, \nu) = \inf \left\{ \mathbb{E}\left( \sum_{i=1}^N d_{i, \beta}(X_i, Y_i) \right) : X \sim \mu, Y \sim \nu \right\}.
\]

Lemmas 16 and 17 means that for all \(x, y \in E\),

\[
\bar{\rho}_\beta(\delta_x R_{t_0}, \delta_y R_{t_0}) \leq \kappa \bar{\rho}_\beta(\delta_x, \delta_y) + \frac{\bar{d}_1(x, y)}{2} \left( e^{\theta t} - 1 \right) \left( 2\beta + \frac{c_s(t)}{N} \sum_{i=1}^N (W_i(x) + W_i(y)) \right)
\]

\[
\leq \kappa \bar{\rho}_\beta(\delta_x, \delta_y) + \frac{\bar{d}_1(x, y)}{2} \left( e^{\theta t} - 1 \right) \frac{c_s(t)}{N} \left( 2\beta + \frac{1 + \eta}{N(1 - \eta)} (\mathcal{V}(x) + \mathcal{V}(y)) \right).
\]

Since \(\eta \geq 1/2\) by assumption,

\[
\frac{\bar{d}_1(x, y)}{2N(1 - \eta)} (\mathcal{V}(x) + \mathcal{V}(y)) = \sum_{i=1}^N \mathbf{1}_{x_i \neq y_i} \frac{\mathcal{V}(x) + \mathcal{V}(y)}{2N(1 - \eta)}
\]

\[
\leq 2 \sum_{i=1}^N \mathbf{1}_{x_i \neq y_i} (W_i(x) + W_i(y)),
\]

and we get that

\[
\bar{\rho}_\beta(\delta_x R_{t_0}, \delta_y R_{t_0}) \leq \left( \kappa + (e^{\theta t} - 1) c_s(t_0)(1 + \eta) \right) \bar{\rho}_\beta(\delta_x, \delta_y) = \tilde{\kappa} \bar{\rho}_\beta(\delta_x, \delta_y).
\]
Then for all initial distributions \( \mu, \nu \in \mathcal{P}(E) \), conditioning with respect to the initial values
\[
\rho_\beta(\mu R_{t_0}, \nu R_{t_0}) = \inf \{ \mathbb{E}(\tilde{\rho}_\beta(\delta_{X_0}, \delta_{Y_0}) : X_0 \sim \mu, Y_0 \sim \nu) \} \leq \bar{\kappa}_\beta(\mu, \nu). 
\]
As in the proof of Theorem \[5\] the assumption that \( \bar{\kappa} < 1 \) implies that \( R_{t_0} \) admits a unique fixed point \( \mu_\infty \in \mathcal{P}(E) \). Moreover, by the semi-group property, for all \( s \geq 0 \),
\[
\mu_\infty R_s R_t = \mu_\infty R_{s+t} = \mu_\infty R_{s+t} = \mu_\infty R_s,
\]
so that \( \mu_\infty R_s \) is a fixed point of \( R_{t_0} \) and thus by uniqueness it is equal to \( \mu_\infty \). Integrating (26) with respect to \( \mu_\infty \) and letting \( t \) go to infinity yields \( \mu_\infty(V) \leq N\gamma/(1-\eta) \), and thus for all \( \nu \in \mathcal{P}_V(E) \),
\[
\bar{\rho}_\beta(\nu, \mu_\infty) \leq 2\beta N + \frac{1}{1-\eta} (\nu(V) + \mu_\infty(V)) \leq \left( 2\beta + \frac{1}{1-\eta} \right) N + \frac{1}{1-\eta} \nu(V).
\]
Finally, for all \( \nu \in \mathcal{P}_V(E) \) and \( t \geq 0 \)
\[
\bar{\rho}_\beta(\nu R_t, \mu_\infty) \leq \bar{\kappa}^{[t/t_0]} \bar{\rho}_\beta(\nu R_{t-t_0} R_{t_0}, \mu_\infty) \leq \bar{\kappa}^{[t/t_0]} \left( 2\beta + \frac{1}{1-\eta} \right) N + \frac{1}{1-\eta} \nu R_{t-t_0} (V) \leq \bar{\kappa}^{[t/t_0]} \left( 2N \left( \beta + \frac{1}{1-\eta} \right) + \frac{1}{1-\eta} \nu(V) \right),
\]
where we used (26). The conclusion then follows from the fact that \( \bar{\rho}_\beta \geq \bar{d}_V \) for all \( \beta \geq 0 \).

\[\Box\]

4 Exemples

This section provides illustrations of our main results. For the sake of clarity and since the approach would be the same in more general or sophisticated applications, the models are chosen to be simple in order to highlight the core arguments.

As has already been mentioned, the neuron network model of [9] provides a first example where Theorem [3] applies.

4.1 Mean-field run-\&-tumble process

Run-\&-tumble processes model, among other things, the motion of some bacteria [16][20] (see also [27][10][29] for more recent details and references). The rate at which a bacterium tumbles depends on the concentration of given chemo-attractants in the medium. For a large population of bacterium, mean-field interaction is a natural extension of these dynamics.

We consider the non-linear integro-differential equation on \( E = \mathbb{R} \times \{-1, 1\} \) given by
\[
\partial_t m_t(x,y) + y \partial_x m_t(x,y) = \lambda_\nu(x,-y)m_t(x,-y) - \lambda_\nu(x,y)m_t(x,y),
\]
with \( \lambda_\nu(x) = r(y(x - \theta x_\nu)) \) where \( x_\nu = \int_E x \nu(dx,dy) \) is the barycentre of \( \nu \), \( \theta \in (0,1) \) and \( r : \mathbb{R} \to \mathbb{R}_+ \) is a Lipschitz function such that
\[
\lim_{s \to -\infty} r(s) := a < b := \lim_{s \to -\infty} r(s).
\]
The linear case $\theta = 0$ corresponds to the run-\&-tumble process attracted to a neighborhood of origin studied in \cite{29} (except that we don’t assume that $r$ is non-decreasing). Indeed, when $|x|$ is large, the jump rate $r(xy)$ is larger when $xy > 0$, namely when the process is drifting away from a given compact, than when $xy < 0$, i.e. when the process is going toward the compact. Similarly, the case $\theta = 1$ would correspond to a process attracted toward the barycentre of its law. For $\theta \in (0, 1)$, the process is attracted to an average of the origin and of the barycentre of its law. Note that, if $s \mapsto r(s) - r(0)$ were anti-symmetric, then the linear process would admit a symmetric equilibrium, which would then be an equilibrium for the non-linear equation since its mean is zero. In the following, this symmetry is not assumed.

A similar multi-dimensional process could be considered with the same assumptions as in \cite{27} to ensure the existence of a Lyapunov function. The arguments below could then be straightforwardly adapted.

**Proposition 18.** Assume \cite{36} holds for some $0 < a < b$ and $c := \inf_{x \in \mathbb{R}} r(s) > 0$. Then there exist $C, \theta > 0$ and $\kappa \in (0, 1)$ such that, if $\theta \leq \theta_\ast$, then \cite{35} admits a unique equilibrium $\mu \in \mathcal{P}(E)$ such that for all solution $t \mapsto m_t$ of \cite{35} and all $t \geq 0$,

$$||m_t - \mu||_W \leq C\kappa^t m_0(W),$$

with $W(x, y) = \exp((b - a)|x|/4)$.

**Proof.** Let us establish that Assumption \cite{4} holds for \cite{35}. To recover the notations of Section \ref{sec:2.1}, consider the generator $L$ on $E$ given by

$$Lf(x, y) = y\partial_x f(x, y) + c \left(f(x, -y) - f(x, y)\right),$$

the jump rate $\lambda_{\nu}(x, y) = r(y(x - \theta x_{\nu})) - c$ and the kernel $Q_{\nu}(x, y) = Q(x, y) = \delta(x, -y)$. In particular, for $z = (x, y) \in E$,

$$||\lambda_{\nu}(z)\left(Q(z) - \delta_z\right) - \lambda_{\mu}(z)\left(Q(z) - \delta_z\right)||_{TV} = ||\lambda_{\nu}(z) - \lambda_{\mu}(z)||_{TV} \leq \theta ||r'||_\infty |x_{\nu} - x_{\mu}|,$$

and

$$|x_{\nu} - x_{\mu}| \leq \int |x||\mu - \nu| \leq k^{-1}||\mu - \nu||_V$$

for all $V$ and $k > 0$ such that $V(x, y) \geq k|x|$ for all $(x, y) \in E$.

The Markov process with generator $L$ is the so-called integrated telegraph process and it is clear, either by simple controllability argument as in \cite{10}, more precise coupling estimates as \cite{29} or just explicit computations of the density transition, that for any compact set $K$ of $\mathbb{R}$ there exist $t_0, \alpha > 0$ such that for all $x, x' \in K$ and $y, y' \in \{-1, 1\}$, the Doeblin condition

$$||\delta_{(x, y)}P_{t_0} - \delta_{(x', y')}P_{t_0}||_{TV} \leq 2(1 - \alpha)$$

holds. Note that $t_0$ and $\alpha$ only depends on $K$ and $c$.

It only remains to construct a Lyapunov function $V$ that satisfies \cite{3}, \cite{4} and \cite{5} and such that $V(x, y) \geq k|x|$ for all $(x, y) \in E$ for some $k > 0$. Given a measurable function $t \mapsto \mu_t$ from $\mathbb{R}_+$ to $\mathcal{P}(E)$, recall that we denote $(P_{t, \nu})_{t \geq 0}$ the semi-group associated to the inhomogeneous process defined in Section \ref{sec:2.1} with the generator

$$L_{\nu}f(x, y) = y\partial_x f(x, y) + \lambda_{\mu}(x, y) \left(f(x, -y) - f(x, y)\right).$$

Let $\varphi : \mathbb{R} \to [0, 1]$ be defined by $\varphi(s) = 1$ for $s \geq 1$, $\varphi(s) = 0$ for $s \leq -1$ and $\varphi(s) = (1 + \sin(\pi s/2))/2$ for $s \in [-1, 1]$, and let $h : \mathbb{R} \to \mathbb{R}_+$ be defined by $h(s) = |s|$ for $s \geq 1$ and $h(s) = (s^2 + 1)/2$ for $s \in [-1, 1]$. Set

$$V(x, y) = e^\frac{b - a}{4}h(x) \left(\frac{5a + 3b}{2(b - a)} + \varphi(y|x)\right).$$

(39)
As a smooth function in \(x\), it belongs to the domain of the generalized generator of \(L_t\) for all \(t \geq 0\), see [15, 21]. For \((x, y) \in E\) with \(|x| \geq 1\),

\[
L_t \mathcal{V}(x, y) = e^{\frac{b-a|x|}{4}} \left[ \mathrm{sign}(xy) \frac{b-a}{4} \left( \frac{5a+3b}{2(b-a)} + 1_{xy>0} \right) + r(y(x-\theta x_{\mu_1})(1_{xy<0} - 1_{xy>0}) \right].
\]

Take \(R_0 \geq 1\) large enough so that \(r(z) > (7b+a)/8\) and \(r(-z) < (b+7a)/8\) for all \(z \geq R_0\). Then, for all \(t \geq 0\) and \((x, y) \in E\) with \(|x| \geq R_0 + \theta|x_{\mu_1}|\), if \(xy > 0\) then

\[
r(y(x-\theta x_{\mu_1})) > (7b+a)/8,
\]

so that

\[
L_t \mathcal{V}(x, y) \leq e^{\frac{b-a|x|}{4}} \left[ \frac{5a+3b}{8} + \frac{b-a}{4} - \frac{7b+a}{8} \right] \leq -\frac{b-a}{4} e^{\frac{b-a|x|}{4}}
\]

while if \(xy < 0\) then

\[
L_t \mathcal{V}(x, y) \leq e^{\frac{b-a|x|}{4}} \left[ -\frac{5a+3b}{8} + \frac{b+7a}{8} \right] \leq -\frac{b-a}{4} e^{\frac{b-a|x|}{4}}.
\]

Hence, for all \(x, y \in E\) with \(|x| \geq R_0 + \theta|x_{\mu_1}|\) and all \(t \geq 0\),

\[
L_t \mathcal{V}(x, y) \leq -\rho \mathcal{V}(x, y)
\]

with \(\rho = (b-a)^2/[10(b+a)]\). Besides, for all \(x, y \in E\) with \(|x| \leq R_0 + \theta|x_{\mu_1}|\), a rough bound is

\[
L_t \mathcal{V}(x, y) + \rho \mathcal{V}(x, y) \leq e^{\frac{b-a}{4} h(x)} (a + 2b + \pi/4)
\]

\[
\leq e^{\frac{b-a}{4} (1+R_0+\theta|x_{\mu_1}|)} (a + 2b + \pi/4)
\]

\[
\leq e^{\frac{b-a}{4} (1+R_0)} \left( 1 + \theta e^{\frac{b-a}{4} |x_{\mu_1}|} \right) (a + 2b + \pi/4),
\]

where we used that \(\theta \in [0, 1]\). Moreover, by the Jensen Inequality,

\[
e^{\frac{b-a}{4} |x_{\mu_1}|} \leq \int_E e^{\frac{b-a}{4} |x|} \mu_t(dx, dy) \leq \frac{2}{3} \mu_t(\mathcal{V}).
\]

We have obtained that, for all \((x, y) \in E\) and all \(t \geq 0\),

\[
L_t \mathcal{V}(x, y) \leq -\rho (\mathcal{V}(x, y) - \eta \mu_t(\mathcal{V}) - M).
\]

with

\[
M = \rho^{-1}(a + 2b + \pi/4) e^{\frac{b-a}{4} (1+R_0)}, \quad \eta = 2\theta M/3.
\]

For all \((x, y) \in E\), let \((X_t, Y_t)_{t \geq 0}\) be a process associated to the semi-group \((R^a_{s,t})_{t \geq s \geq 0}\) and initial conditions \((x, y)\) and let \(N_t\) be its number of jumps in \([0, t]\), which is stochastically less than a r.v. with Poisson law with parameter \(||r||_{\infty} t\). For \(n \in \mathbb{N}\), let \(\tau_n = \inf\{t, N_t \leq n, |X_t| \geq n\}\), which almost surely goes to infinity as \(n\) goes to infinity (note that \(|X_t| \leq |x| + t\) for all \(t \geq 0\)). For all \(n \in \mathbb{N}\), the Dynkin formula yields

\[
E \left( e^{\rho(t \wedge \tau_n)} \mathcal{V}(X_{t \wedge \tau_n}) \right) = \mathcal{V}(x) + E \left( \int_0^{t \wedge \tau_n} e^{\rho s} (L_s \mathcal{V}(X_s) + \rho \mathcal{V}(X_s)) ds \right)
\]

\[
\leq \mathcal{V}(x) + \rho \int_0^t e^{\rho s} (\eta \mu_s(\mathcal{V}) + M) ds.
\]

Letting \(n\) go to infinity we obtain [3]. If \(\theta \leq \theta_* \leq 1/M, \eta \leq 2/3\).
The two other Lyapunov conditions of Assumption \[4\] are readily checked. Indeed, similar computations yield \(LV(x,y) \leq (b + \pi/2)V(x,y)\) and then
\[
P_t V(x,y) \leq e^{(b+\pi/2)t}V(x,y)
\]
and, since \(G_\nu((x,y),u) = (x,-y)\) for all \(\nu \in \mathcal{P}(E)\), \((x,y) \in E\) and almost all \(u \in [0,1],\)
\[
\frac{V(G_\nu((x,y),u))}{V(x,y)} \leq \frac{5\pi + 3b_\ast}{2(b-a)} + \varphi(|x|) \leq 2.
\]
Condition \(6\) is deduced from \(38\). Finally, \(7\) ensues from \(37\) since
\[
|x| \leq \frac{4}{b-a} e^{\frac{b-a}{2}|x|} \leq \frac{8}{3(b-a)} V(x,y)
\]
for all \((x,y) \in E\). As a conclusion, if \(\theta \leq \theta_\ast < 1/M\), Assumption \[4\] holds, and Theorems \[4\] and \[5\] state that \(35\) admits a solution for all initial condition \(m_0 \in \mathcal{P}_V\) and that, provided \(\theta_\ast\) is sufficiently small (depending on \(r\) or more precisely on \(a, b, c\) and \(\|r^\prime\|_\infty\)), then \(35\) admits a unique equilibrium toward which all solutions \(t \mapsto m_t\) with \(m_0 \in \mathcal{P}_V(E)\) converges geometrically. The equivalence of \(V\) and \(W\), hence of \(\|\cdot\|_V\) and \(\|\cdot\|_W\), concludes.

\[\Box\]

### 4.2 MCMC for granular media equilibrium

Let \(E_1 = (\mathbb{R}/\mathbb{Z})^d\) for some \(d \in \mathbb{N}_\ast\) and let \(U : E_1 \to \mathbb{R}\) and \(W : E_1^2 \to \mathbb{R}\) be \(C^1\) functions respectively called the exterior and interaction potential. We want to sample according to the probability law \(\mu_V\) on \(E_1\) with density proportional to \(\exp(-\beta V)\), where \(\beta > 0\) is the so-called inverse temperature and \(V\) solves
\[
V(x) = U(x) + \frac{\int_{E_1} W(x,y) e^{-\beta V(y)} \, dz}{\int_{E_1} e^{-\beta V(z)} \, dz}. \tag{40}
\]

In fact, this problem doesn’t necessarily have a unique solution. More precisely, \(V\) solves \(40\) iff \(\mu_V\) is an equilibrium of the McKean-Vlasov (or granular media) equation \(11\ 38\ 39\), which admits a unique such equilibrium at large temperature \(30\ 49\), i.e. for \(\beta\) smaller than some threshold \(\beta_0 > 0\). We will only consider this large temperature regime.

Sampling according to \(\mu_V\) can be achieved through interacting particles MCMC. To that purpose, we consider three classical Markov chains: the Metropolis-Hastings (MH) chain with Gaussian proposal, the Unadjusted Langevin Algorithm (ULA) and the Metropolis Adjusted Langevin Algorithm (MALA), which we now define. Let \(N \in \mathbb{N}_\ast\) and, for \(x \in E = E_1^N\), write
\[
U_N(x) = \sum_{i=1}^N \left( U(x_i) + \frac{1}{2N} \sum_{j=1}^N \mathbb{1}_{j \neq i} W(x_i, x_j) \right).
\]

For \(i \in [1, N]\), we denote by \(e_i\) the \((Nd) \times d\) matrix whose \(d \times d\) blocks are all zero except the \(i^{th}\) one which is the \(d\)-dimensional identity matrix. In other words, \(e_i\) such that, if \(x \in E\) and \(y \in \mathbb{R}^d\), then \(x + e_i y = (x_1, \ldots, x_{i-1}, x_i + y, x_{i+1}, \ldots, x_N)\), where \(x_i + y\) is understood as a sum in \((\mathbb{R}/\mathbb{Z})^d\). For \(i \in [1, N]\), \(x \in E^N\) and \(y \in \mathbb{R}^d\), set
\[
p_i(x, y) = 1 \land \exp(\beta (U_N(x) - U_N(x + e_i (y - x_i))))
\]
\[
\tilde{p}_i(x, y) = 1 \land \frac{\exp(\beta U_N(x) + \frac{1}{4\tau^2}|y - \tau \beta \nabla x, U_N(x_i)|^2)}{\exp(\beta U_N(x + e_i y) + \frac{1}{4\tau^2}|y - \tau \beta \nabla x, U_N(x + e_i y)|^2)}
\]

29
for some \( \tau > 0 \) and let \( q : E_1^2 \to \mathbb{R}_+ \) be a symmetric Markov density kernel, i.e. be such that for all \( x, q(x, \cdot) \) is the density of a probability measure on \( E_1 \) and \( q(x, y) = q(y, x) \) for all \( x, y \in E_1 \). For instance, the image on the periodic torus of \( q(x, y) \propto \exp(-|x - y|^2/(2\sigma^2)) \) for some \( \sigma > 0 \), or \( q(x, y) \propto 1 \). Define on \( E \) the Markov kernels \( Q_{\text{MH}}, Q_{\text{ULA}} \) and \( Q_{\text{MALA}} \) by

\[
Q_{\text{MH}} f(x) = \frac{1}{N} \sum_{i=1}^{N} \int_{E_1} (p_i(x, y)f(x + e_i(y - x_i)) + (1 - p_i(x, y))f(x)) q(x_i, y) \, dy
\]

\[
Q_{\text{ULA}} f(x) = \frac{1}{N} \sum_{i=1}^{N} \int_{\mathbb{R}^d} f(x + e_i y) e^{-\frac{1}{\tau} |y - \tau \beta \nabla_x U_N(x_i)|^2} \sqrt{8\pi \tau} \, dy
\]

\[
Q_{\text{MALA}} f(x) = \frac{1}{N} \sum_{i=1}^{N} \int_{\mathbb{R}^d} (\tilde{p}_i(x, y)f(x + e_i y) + (1 - \tilde{p}_i(x, y))f(x)) e^{-\frac{1}{\tau} |y - \tau \beta \nabla_x U_N(x_i)|^2} \sqrt{8\pi \tau} \, dy
\]

Let \( x \in E \), and consider independent r.v. \( Y, Z, U \) and \( I \) where \( Y \) follows the standard (mean zero, variance one) Gaussian distribution, and \( Z, U \) and \( I \) the uniform one respectively on \( E_1, [0, 1] \) and \( [1, N] \). Then

\[
x + e_i (-\tau \beta \nabla_x U_N(x) + \sqrt{2\tau} Y) \sim Q_{\text{ULA}}(x)
\]

\[
x + \mathbb{1}_{\{p_t(x, -\tau \beta \nabla_x U_N(x) + \sqrt{2\tau} Y) < U \}} e^I (-\tau \beta \nabla_x U_N(x) + \sqrt{2\tau} Y) \sim Q_{\text{MALA}}(x)
\]

\[
x + \mathbb{1}_{\{p_t(x, \sigma Y) < U \}} e^I \sigma Y \sim Q_{\text{MH}}(x)
\]

if \( q(x, y) \propto \exp(-|x - y|^2/(2\sigma^2)) \) while, if \( q(x, y) \propto 1 \),

\[
x + \mathbb{1}_{\{p_t(x, Z - x) < U \}} e^I (Z - x) \sim Q_{\text{MH}}(x)
\]

More discussions, motivations and comparison of these processes can be found in [8, 7, 23] and references within. As far as the present work is concerned, these three cases are similar, so that we focus in the following on the MH case alone, with \( q(x, y) \propto 1 \). For a given jump rate \( \tilde{\lambda} > 0 \), consider the continuous-time Markov chain on \( E \) with generator

\[
L_{\text{MH}} f(x) = \tilde{\lambda} N (Q_{\text{MH}} f(x) - f(x))
\]

Let \( (R_t)_{t \geq 0} \) be the associated semi-group. The associated process is constructed as follows. Let \( (S_k, I_k, U_k, Z_k)_{k \geq 0} \) be an i.i.d. sequence where, for all \( k \in \mathbb{N} \), \( S_k \) is a standard exponential r.v., and \( Z_k, U_k \) and \( I_k \) are uniformly distributed respectively on \( E_1, [0, 1] \) and \( [1, N] \). Let \( x \in E \), set \( X_{i,0} = x_i \) for all \( i \in [1, N] \), \( T_0 = 0 \) and \( T_{n+1} = T_n + ([N \tilde{\lambda}]^{-1} S_n) \) for all \( n \in \mathbb{N} \). Suppose that \( (X_i)_{i \in [0, T_n]} \) has been defined for some \( n \in \mathbb{N} \) and is independent from \( (S_k, I_k, U_k, Y_k)_{k \geq n} \). Set \( X_t = X_{T_n} \) for all \( t \in (T_n, T_{n+1}) \). For all \( j \neq i_n \), set \( X_{i, T_{n+1}} = X_{j, T_n} \). If \( U_n < p_{1,1}(X_{T_n}, Z_n - X_{i,n}) \), set \( X_{i,n, T_{n+1}} = Z_n \) and else, set \( X_{i,n, T_{n+1}} = X_{i,n, T_n} \).

**Proposition 19.** Suppose that \( \beta = \tilde{\beta}/((\text{osc}(U) + \text{osc}(W))) \) with \( \tilde{\beta} \) such that

\[
\rho := e^{-\tilde{\beta}} - \frac{4\text{osc}(W)}{\text{osc}(U) + \text{osc}(W)} \tilde{\beta} (e^{\tilde{\beta}} - 1) > 0.
\]

Then \([40]\) admits a unique solution \( V \) and, moreover, for all \( m_0, h_0 \in \mathcal{P}(E) \) and all \( t \geq 0 \),

\[
\|m_0 R_t - h_0 R_t\|_{TV} \leq N e^{-\tilde{\lambda} t\rho}\|m_0 - h_0\|_{TV}.
\]

**Proof.** To recover the notations of Section [2.2], denote \( p_* = \exp(-\beta(\text{osc}(U) + \text{osc}(W))) \) and for all \( i \in [1, N] \) consider the generator \( L_i \) on \( E_i = E_1 \) defined by

\[
L_i f(z) = \tilde{\lambda} p_*(\int_{E_1} f(y) dy - f(z))
\]

30
for all \( z \in E_1 \) and bounded measurable function \( f \) on \( E_1 \). Let \((P_t^i)_{t \geq 0}\) be the associated Markov semi-group, which is simply

\[
P_t^i f(z) = e^{-\bar{\lambda}p_t^i f(z)} + \left(1 - e^{-\bar{\lambda}p_t^i}\right) \int_{E_1} f(y) dy. \tag{41}
\]

For all \( i \in [1,N] \) and \( x \in E \), set \( \lambda_i(x) = \bar{\lambda}(1 - p_x) \) and, for a measurable bounded \( f \) on \( E_i \),

\[
Q_i f(x) = \int_{E_i} \left( \frac{p_i(x,y) - p_x f(y)}{1 - p_x} f(x) + \frac{1 - p_i(x,y)}{1 - p_x} f(x_i) \right) dy.
\]

Remark that \( p_x \leq p_i(x,y) \) for all \( i \in [1,N] \), \( x \in E \) and \( y \in E_1 \), so that \( Q_i \) is indeed a Markov kernel. Then the MH process with generator \( \bar{L}_{MH} \) is the system of particles such as defined in Section 2.2 associated to \((P_t^i)_{t \geq 0}, \lambda_i, Q_i)_{i \in [1,N]}\). Note that for all \( i \in [1,N] \) and \( x, z \in E \),

\[
\left\| \lambda_i(x) (Q_i(x) - \delta_x) - \lambda_i(z) (Q_i(z) - \delta_z) \right\|_{TV} = \bar{\lambda}(1 - p_x) \left\| Q_i(x) - Q_i(z) \right\|_{TV}.
\]

If \( x, z \in E \) are such that \( x_i = z_i \), then an optimal coupling of \( Q_i(x) \) and \( Q_i(z) \) is constructed as follows. Let \( Y \) and \( U \) be independent r.v. uniformly distributed over, respectively, \( E_1 \) and \([0,1]\). Set

\[
X_i = x_i + 1_{(1-p_x)U<p_i(x,Y)-p_x}(Y-x_i), \quad Z_i := z_i + 1_{(1-p_x)U<p_i(z,Y)-p_x}(Y-z_i).
\]

Then \( X_i \sim Q_i(x) \), \( Z_i \sim Q_i(z) \) and

\[
\mathbb{P}(X_i \neq Z_i) = \mathbb{P}[p_i(x,Y) \wedge p_i(z,Y) - p_x < (1 - p_x)U < p_i(x,Y) \vee p_i(z,Y) - p_x] = (1 - p_x)^{-1} \int_{E_i} |p_i(x,y) - p_i(z,y)| dy.
\]

Now, for all \( a, b \in \mathbb{R}, \) \( 1 \wedge e^a - 1 \wedge e^b \leq |e^a - e^b| \leq e^{b\lambda_1} |b - a| \), so that if \( x_i = z_i \) then

\[
|p_i(x,y) - p_i(z,y)| \leq \beta e^{\beta \lambda_{osc}(U) + osc(W)}|U_N(x) - U_N(x + e_i y) - U_N(z) + U_N(z + e_i y)| = \frac{\beta e^{\beta \lambda}}{N} \sum_{j \neq i} \left| W(x_i, x_j) - W(x_i + y, x_j) - W(z_i, z_j) + W(z_i + y, z_j) \right|
\]

\[
\leq 2osc(W)\beta e^{\beta \lambda} \frac{d_1(x,y)}{N}.
\]

We have thus obtained, for all \( i \in [1,N] \) and \( x, z \in E \) with \( x_i = z_i \),

\[
\left\| \lambda_i(x) (Q_i(x) - \delta_x) - \lambda_i(z) (Q_i(z) - \delta_z) \right\|_{TV} \leq \frac{\theta d_1(x,y)}{N}.
\]

which is \((10)\), with \( \theta = 4\bar{\lambda}(1 - p_x) osc(W)\beta \exp(\bar{\beta}) \). On the other hand, the Doeblin condition \((11)\) is clear, since \((41)\) immediatly yields that for all \( t \geq 0 \) and \( x, y \in E_1 \) and all \( i \in [1,N] \),

\[
\left\| \delta_x P_t^i - \delta_y P_t^i \right\|_{TV} = 21_{x \neq y} e^{-\lambda p_t^i t}.
\]

Hence, Assumption \((6)\) holds and denoting \((R_t)_{t \geq 0}\) the semi-group on \( E \) associated to the particle system \((X_i, t)_{i \in [1,N], t \geq 0}\). Theorem \((7)\) means that for all \( t_0 > 0, n \in \mathbb{N} \) and \( h_0, m_0 \in \mathcal{P}(E) \),

\[
\left\| m_0 R_{nt_0} - h_0 R_{nt_0} \right\|_{TV} \leq N e^{\theta t_0} \left( e^{\theta t_0} - \left(1 - e^{-\lambda p_{t_0}}\right) e^{-\lambda(1 - p_x) t_0} \right)^n \left\| m_0 - h_0 \right\|_{TV},
\]

31
which, applied for \( t_0 = t/n \) for a fixed \( t \geq 0 \) and letting \( n \) go to infinity, reads
\[
\|m_0 R_t - h_0 R_t\|_{TV} \leq N \epsilon^{(\theta - \lambda \rho_\ast)t} \|m_0 - h_0\|_{TV}.
\]
Similarly, provided \( m_0 = (m'_0)^{\otimes N} \) for some \( m'_0 \in \mathcal{P}(E_1) \) and similarly for \( h_0 \), following the remark made after Theorem 6,
\[
\|m'_t - h'_t\|_{TV} \leq e^{(\theta - \lambda \rho_\ast)t} \|m'_0 - h'_0\|_{TV}
\]
holds for all \( t \geq 0 \) if \( m'_t \) is either the law of \( X_{t,i} \) or the solution of the non-linear mean-field limit of the particle system (applying Theorem 3). In particular, remark that \( V \) solves (40) if the probability density proportional to \( \exp(-\beta V) \) is an equilibrium of this non-linear equation. Yet, if \( \rho = p_\ast - \theta/\lambda > 0 \) then the contraction of the total variation norm implies that the latter admits a unique equilibrium, which concludes.

4.3 The Zig-Zag process with a close to tensor target

Let \( \pi \in \mathcal{P}(\mathbb{R}^N) \) be a probability law with a density proportional to \( \exp(-U) \), where \( U \in C^1(\mathbb{R}^N) \). Consider the so-called Zig-Zag process \([4]\) on \( E = \mathbb{R}^N \times \{-1, 1\}^N \) with generator
\[
L_{ZZ} f(x, y) = y \cdot \nabla_x f(x, y) + \sum_{i=1}^N (y_i \partial_{x_i} U(x))_+ (f(x, y_i - y) - f(x, y))
\]
where, for \( y \in \{-1, 1\}^N \) and \( i \in [1, N] \), \( y_i = (y_1, \ldots, y_{i-1}, y_i, y_{i+1}, \ldots, y_N) \). The Zig-Zag process admits \( \pi(dx) \otimes ((\delta_{-1} + \delta_1)/2)^{\otimes N} \) as an invariant measure and it is ergodic under general conditions on \( U \) (see [5]), so that for all reasonable (e.g. bounded) functions \( f \) on \( \mathbb{R}^N \),
\[
\frac{1}{t} \int_0^t f(X_s) ds \overset{t \to \infty}{\longrightarrow} \pi(f).
\]
This makes it suitable for MCMC purpose. If the target measure is of tensor form, i.e. if \( U(x) = U_1(x_1) + \cdots + U_N(x_N) \) for some one-dimensional functions \( U_i \), then the \( N \) coordinates of the process are independent one-dimensional Zig-Zag processes, so that the convergence to equilibrium of each of these coordinates is independent of \( N \). Let us prove that this property is stable under the addition of correlations between the coordinates, provided they are small.

Proposition 20. Suppose that there exist \( U_1, \ldots, U_N \in C^1(\mathbb{R}) \), \( W \in C^1(\mathbb{R}^N) \), \( \rho > 0 \) and \( R \geq 1 \) such that for all \( i \in [1, N] \) and \( x \in \mathbb{R} \), if \( |x| \geq R \) then \( xU'_i(x) \geq \rho|x| \). Let \( (R_t)_{t \geq 0} \) be the semi-group associated to the generator \( L_{ZZ} \) with \( U(x) = W(x) + \sum_{i=1}^N U_i(x_i) \). Then there exist \( \theta_\ast, C' > 0 \) and \( \kappa \in (0, 1) \) that depend only on \( R, \rho \) and \( C := \sup\{|U'_i(x)|, i \in [1, N], |x| < R\} \) such that, if \( \sup\{\|\partial_{x_i} W\|_\infty, i \in [1, N]\} < \theta_\ast \), then for all \( \mu \in \mathcal{P}(\mathbb{R}^N) \),
\[
\|\mu R_t - \pi\|_{TV} \leq C' e^{\kappa t} \mu(V).
\]

Proof. Let \( h, \varphi \in C^1(\mathbb{R}) \) be as defined in the proof of Proposition 18. For \( i \in [1, N] \) and \( (z, w) \in E = \mathbb{R} \times \{-1, 1\} \), set \( V_i(z, w) = \exp(\rho h(z)/4) (1 + \varphi(zw)) \). We extend \( V_i \) as a function on \( E^N \) by \( V_i(x, y) = V_i(x_i, y_i) \) for \( (x, y) \in E^N \) and write \( V = \sum_{i=1}^N V_i \). Denote \( \theta = \sup_{i \in [1, N]} \|\partial_{x_i} W\|_\infty \) and suppose that \( \theta \leq \rho/8 \). If \( (x, y) \in E^N \) is such that \( |x_i| \geq R, \)
\[
L_{ZZ} V_i(x, y) = \left[ -\frac{1}{4} \rho (1 + \mathbbm{1}_{x_i > 0}) \text{sign}(x_i) \right] - \frac{1}{4} \rho + \|\partial_{x_i} W\|_\infty \right] e^{\frac{\rho h(x)}{4}} \leq -\frac{\rho}{16} V_i(x, y).
\]
On the other hand if \(|x_i| \leq R\),
\[
L_{\text{ZZ}}V_i(x,y) + \frac{\rho}{16}V_i(x,y) \leq 2e^{\rho R^2/2} \left( \frac{\rho}{16} + \frac{\rho}{2} + \frac{\pi}{4} + C + \|\partial_x W\|_\infty \right) \leq \frac{\rho}{16} C'
\]
with \(C' = 32\exp(\rho R/2)(\rho + 1 + C)/\rho\). Integrating in time similarly as in the proof of Proposition 18 we get that
\[
R_t V_i(x,y) \leq e^{-\frac{\rho}{16}t} V_i(x_i, y_i) + \left(1 - e^{-\frac{\rho}{16}t}\right) C'.
\]
for all \(t \geq 0\), \((x, y) \in E^N \text{ and } i \in [1, N]\). In particular, summing over \(i \in [1, N]\),
\[
R_t \mathcal{V}(x,y) \leq e^{-\frac{\rho}{16}t} \mathcal{V}(x,y) + N \left(1 - e^{-\frac{\rho}{16}t}\right) C'.
\]

To check the other conditions of Assumption 5 and 7, we consider the semi-group \((P_t^i)_{t \geq 0}\) on \(R\) with generator \(L_i\) given by
\[
L_i f(z,w) = w \partial_z f(z,w) + (wU_i'(z) - \theta)_+ (f(z,w) - f(z,w)) ,
\]
and the jump rates and kernels
\[
\lambda_i(x,y) = (y_i \partial_x U(x))_+ - (y_i U_i'(x_i) - \theta)_+ \\
Q_i(x,y) = \delta_{(x,y_i)}.
\]

Then the Zig-Zag process corresponds to the process defined in Section 2.2 associated to the semi-groups \((P_t^i)_{t \geq 0, i \in [1, N]}\) and the jump mechanisms \((\lambda_i, Q_i)_{i \in [1, N]}\). Note that, if \(G_i\) is a representation of \(Q_i\), then for all \(x \in E^N\) and almost all \(u \in [0, 1]\), \(G_i((x,y), u) = (x,y_i)\) so that \(V_i(G_i((x,y), u)) \leq 2V_i(x_i, y_i)\). To check (13), through computations similar to the previous ones, it is clear that
\[
L_i V_i(z,w) \leq -\frac{\rho}{16} V_i(z,w) + \tilde{C}
\]
for some \(\tilde{C}\) that does not depend on \(\theta\). Finally, the Doeblin condition (16) is a consequence of the ergodicity of the one-dimensional Zig-Zag process as established in 29, 5. Hence, Theorem 7 holds, which concludes.

### 4.4 Hybrid drift/bounce kinetic samplers

Let \(E_1 = \mathbb{R}^d\) for some \(d \in \mathbb{N}\), \(U \in C^1(E_1)\) be an exterior potential, \(W \in C^1(E_1^2)\) an interaction potential and \(\beta > 0\) be an inverse temperature. Similarly as in Section 4.2 we want to compute expectations with respect to the granular media equilibrium, i.e. the probability measure with density proportional to \(\exp(-\beta V)\) where \(V\) solves (40). Again, this can be approximated with a system of \(N\) interacting particles. Denoting, for \(x \in E := E_1^N\),
\[
U_N(x) = \sum_{i=1}^N U(x_i), \quad W_N(x) = \frac{1}{2N} \sum_{j=1}^N W(x_i, x_j),
\]
consider the Markov process on \(E^2\) with generator \(L_N\) defined by
\[
L_N f(x,v) = v \cdot \nabla_x f(x,v) - \beta \nabla U(x) \cdot \nabla_v f(x,v) \\
+ \beta \sum_{i=1}^N (v_i \cdot \nabla_x W_N(x))_+ (f(x,R_N;_i(x,v)) - f(x,v)) + D f(x,v)
\]
with

\[ R_{N,i}(x,v) = \left( v_1, v_2, \ldots, v_{i-1}, v_i - 2 \frac{v_i \cdot \nabla_x W_N(x) \nabla_x W_N(x), v_{i+1}, \ldots, v_N} {\| \nabla_x W_N(x) \|^2} \right), \]

and \( D \) is some dissipativity operator on the velocities, ergodic with respect to the standard Gaussian measure \( \gamma_d \) on \( E = \mathbb{R}^d \), for instance \( D = hD_1, \ i = 1,3 \) with \( h > 0 \) and

\[ D_1 f(x,v) = -v \cdot \nabla_x f(x,v) + \Delta_v f(x,v) \]
\[ D_2 f(x,v) = \int \left( f \left( x, \sqrt{\rho} v + \sqrt{1 - \rho w} \right) - f(x,v) \right) \gamma_d(dw) \]
\[ D_3 f(x,v) = \sum_{i=1}^{N} \int \left( f \left( x, \left( v_1, \ldots, v_{i-1}, \sqrt{\rho} v_i + \sqrt{1 - \rho w}, v_{i+1}, \ldots, v_N \right) \right) - f(x,v) \right) \gamma_d(dw) \]

for some \( p \in [0,1) \). Then the probability measure \( \mu_N \propto \exp(-\beta(U_N + W_N)) \) on \( E \) is invariant for \( L_N \).

The motivation to use such an hybrid process to sample \( \mu_N \) is the following. The computations of \( \nabla U_N \) and \( \nabla_x W_N \) have a respective numerical cost of \( O(N) \) and \( O(N^2) \). Suppose that \( W \) is Lipschitz with a known bound \( \|\nabla W\|_\infty \leq \eta \). Then, by thinning method \cite{33,34}, it is rather simple to sample jump times with rate \( (v_i \cdot \nabla_x W_N(x))_+ \), by proposing jumps at rate \( |v_i|\eta \) and then accepting them with some probability. In that case, we only have to compute \( \nabla_x W \) at each of these proposed jump times, instead of computing it at all times \( n\delta t, n \in \mathbb{N} \), where \( \delta t \) is the timestep used for the discretization of the trajectory. If, on the other hand, no efficient bounds on \( \nabla U \) are available, then it makes more sense to deal with \( U \) with a drift operator (with a discretization scheme with timestep \( \delta t \)) rather than with a jump one. This argument is in fact not restricted to mean-field processes: each time the forces can be decomposed as a bounded but expensive part and a cheap but singular one (for instance, long-range interactions versus short-range interactions in molecular dynamics) it is reasonable to treat the bounded part with jumps and the singular one with drift. Contrary to multiple timestep methods like RESPA \cite{48}, there is no contribution of the bounded forces to the systematic bias on the invariant measure.

Now, to study the long-time behaviour of the particle system with generator \( L_N \), we can decompose \( L_N = L'_N + L''_N \) with

\[ L'_N f(x,v) = v \cdot \nabla_x f(x,v) - \beta \nabla U(x) \cdot \nabla_v f(x,v) + Df(x,v) \]
\[ L''_N f(x,v) = \beta \sum_{i=1}^{N} \left( v_i \cdot \nabla_x W_N(x) \right)_+ (f(x,R_{N,i}(x,v)) - f(x,v)) \]

If \( D = D_1 \) (resp. \( D_3 \)) for instance, then \( L'_N \) is the generator of \( N \) independent kinetic Langevin (resp. BGK-like) processes. Remark that, in the case of \( D = D_2 \), then \( L'_N \) is not the generator of \( N \) independent processes because the partial refreshment of the velocities occurs at the same time for all the particles. As a consequence, this example doesn’t exactly enter the framework of Section 2.2. Nevertheless, it is rather clear that, in this case, a coupling in the spirit of \cite{21} will still give for each coordinate, independently from the others, a probability to merge in some time \( t_0 \) independent from \( N \). This is what is really required in the proof of Theorem 7 (more precisely, of Lemma 16). In fact, the partial refreshment of the velocities could also be done, for all particles at once, at a given deterministic time period, as in Hamiltonian Monte-Carlo.

In any cases, as in the previous section, establishing the existence of a Lyapunov function \( \mathcal{V}(x) = \sum_{i=1}^{N} \mathcal{V}_i(x_i) \) and a local Doeblin condition for \( L'_N \) is possible under some assumptions on \( U \) and, assuming that \( \eta \) is small enough, one can then prove through Theorem 7 that the particle system converges towards its equilibrium at a rate independent from \( N \).
4.5 Selection/Mutation algorithms

Let \((P_t)_{t \geq 0}\) be a Markov semi-group on a Polish space \(E\) that satisfies Assumption 3. Consider \(\lambda_* > 0, N \in \mathbb{N}\), and a function \(p : E^2 \to [0, 1]\). For all \(i \in [1, N]\) and \(x \in E^N\), set \(\lambda_i(x) = \lambda_*\)

\[
Q_i f(x) = \frac{1}{N} \sum_{j=1}^{N} p(x_i, x_j) f(x_{j \rightarrow i}) + (1 - p(x_i, x_j)) f(x)
\]

where \(z = x_{j \rightarrow i}\) is defined by \(z_k = x_k\) for all \(k \in [1, N] \setminus \{i\}\) and \(z_i = x_j\). In other words, if \(x = (x_1, \ldots, x_N)\) is the position of \(N\) particles, a r.v. with law \(Q_i(x)\) is drawn as follows: draw \(J\) uniformly over \([1, N]\) and, with probability \(p(x_i, x_j)\), kill the \(i\)th particle and replace it by a copy of the \(J^{th}\) one. This kind of dynamics is used in a variety of algorithms, see [16, 17, 12, 28, 14] and references within. Then Assumption 6 clearly holds since, if \(x, Z \in E^N\) are such that \(x_i = z_i\),

\[
||Q_i(x) - Q_i(z)||_{TV} \leq \frac{1}{N} \sum_{j=1}^{N} |p(x_i, x_j) - p(x_i, z_j)| \leq \frac{d_1(x, z)}{N}.
\]

Hence, Theorem 6 holds.

4.6 The mean-field TCP process

We consider the non-linear integro-differential equation on \(\mathbb{R}_+\) given by

\[
\partial_t m_t(x) + \partial_x m_t(x, y) = \lambda_{m_t}(2x)m_t(2x) - \lambda_{m_t}(x)m_t(x),
\]

with, for \(\nu \in \mathcal{P}(\mathbb{R}_+)\) and \(x \in \mathbb{R}_+\),

\[
\lambda_{\nu}(x) = 1 + g_1(x) + \int_0^\infty g_2(x + y)\nu(dy),
\]

where \(g_1\) and \(g_2\) are both positive, non-decreasing functions on \(\mathbb{R}_+\) that goes to infinity at infinity. For references on this model, called the TCP process, see [13, 2] and references within. The choice for this particular expression of \(\lambda_{\nu}\) is not motivated by any modelling consideration; it is only meant to provide a very basic, yet interesting, exemple where Assumption 1 is not satisfied, since the non-linear jump rate is unbounded. In particular, it should be checked that, given a measurable function \(t \mapsto \mu_t\) from \(\mathbb{R}_+\) to \(\mathcal{P}(\mathbb{R}_+)\), then the associated process \((X^{p, x}_t)_{t \geq 0}\), such as defined in Section 2.1, is well-defined for all time, namely that the probability that an infinite number of jumps occurs in a finite time is zero.

This process is defined as follows. Suppose that \(t \mapsto \mu_t\) is such that for all \(x \geq 0\),

\[
g_\nu(x) := \sup_{t \geq 0} \int_0^\infty g_2(x + y)\mu_t(dy) < \infty.
\]

Let \((S_k)_{k \geq 0}\) be an i.i.d. sequence of standard exponential r.v., \(X_0 = x\) and \(T_0 = 0\). Suppose that \(T_n \geq 0\) and \((X_t)_{t \in [0, T_n]}\) have been defined for some \(n \in \mathbb{N}\). Set

\[
T_{n+1} = \inf \left\{ t > T_n, S_n < \int_{T_n}^t \lambda_{\mu_s}(X_{T_n} + (s - T_n))ds \right\}, \quad X_{T_{n+1}} = \frac{X_{T_n} + T_{n+1} - T_n}{2}
\]

and, for \(t \in (T_n, T_{n+1})\), \(X_t = X_{T_n} + t - T_n\). The process is then constructed by induction up to time \(T_n\) for all \(n \in \mathbb{N}\). Moreover, by construction, \(X_{t \wedge T_n} \leq X_0 + t\) for all \(n \in \mathbb{N}\) and
all $t \geq 0$, and since $g_1$ and $g_2$ are non-decreasing, $\lambda_{m_i}(X_s) \leq 1 + g_1(X_0 + t) + g_2(X_0 + t)$ for all $s \leq t \wedge T_n$. Hence, for all $n$ such that $T_n \leq t$,

$$T_n \geq \frac{1}{1 + g_1(X_0 + t) + g_2(X_0 + t)} \sum_{k=0}^{n-1} S_k.$$ 

In particular, for all $t \geq 0$, there are almost surely a finite number of jumps before time $t$. In other words, $T_n$ almost surely goes to infinity as $n$ goes to infinity, so that $X_t$ is almost surely defined for all $t \geq 0$. Let $(R^\mu_{s,t})_{t \geq s \geq 0}$ be the associated inhomogeneous Markov semi-group, and $L_i$ be its generator, given by

$$L_tf(x) = f'(x) + \lambda_{m_i}(x) (f(x/2) - f(x)).$$

In the following we suppose that $g_2(x) \leq K \exp(\rho x)$ for some $K, \rho > 0$, and that $\sup_{t \geq 0} \mu_t(V) < \infty$, where $V(x) = \exp(\rho x)$. In particular, (43) holds, so that the associated process and semi-group are well defined.

Then, for $x \geq R := \inf \{x \geq 0, g_1(x) \geq 2\rho\}$,

$$L_tV(x) = \rho V(x) + \left(1 + g_1(x) + \int_0^t g_2(x + y)\mu_t(dy)\right) (V\left(x^2\right) - V(x)) \leq (1 + 2\rho) \sqrt{V(x)} - (1 + \rho) V(x) \leq -\frac{1}{2} V(x) + \frac{1}{2} + \rho.$$

For $x \in [0, R]$, $L_tV(x) \leq \rho V(x) \leq \rho \exp(\rho R)$ and thus, for all $x \in \mathbb{R}_+$,

$$L_tV(x) \leq -\frac{1}{2} (V(x) - \tilde{C}),$$

with $\tilde{C} = (1 + 2\rho)(1 + e^{\rho R})$. As in the proof of Proposition 18, this yields

$$R^\mu_{0,t}V(x) \leq e^{-t/2} V(x) + \left(1 - e^{-t/2}\right) \tilde{C}. \quad (44)$$

In particular, for any $m_0 \in \mathcal{P}(\mathbb{R}_+)$, $t \geq 0$ and $x \geq 0$,

$$m_0 R^\mu_{0,t} (g_2(x + \cdot)) \leq Ke^{\rho x} m_0 R^\mu_{0,t}(V) \leq Ke^{\rho x} \left(\tilde{C} \vee m_0(V)\right).$$

Note that, simply by changing $\rho$ to $2\rho$, the same computations shows that there also exist $\tilde{C}$ such that

$$R^\mu_{0,t}(V^2)(x) \leq e^{-t/2} V^2(x) + \left(1 - e^{-t/2}\right) \tilde{C}. \quad (45)$$

Besides, for all $\nu_1, \nu_2 \in \mathcal{P}(\mathbb{R}_+)$, denoting $(Y, \tilde{Y})$ an optimal coupling of $\nu_1$ and $\nu_2$ such as given by Lemma 8,

$$|\lambda_{\nu_1}(x) - \lambda_{\nu_2}(x)| = |\mathbb{E}\left(g_2(x + Y) - g_2(x + \tilde{Y})\right)| \leq \mathbb{E}\left(1_{Y \neq \tilde{Y}}\left(g_2(x + Y) + g_2(x + \tilde{Y})\right)\right) \leq \frac{e^{\rho x}}{K} \Vert \nu_1 - \nu_2 \Vert_V.$$

Let $m_0 \in \mathcal{P}(\mathbb{R}_+)$ and $t \mapsto \mu_i^1, \mu_i^2$ be such that $\mu_i^1(V) \leq m_0(V) \vee \tilde{C}$ for $i = 1, 2$ and $t \geq 0$. Consider the synchronous coupling $(X_t, \tilde{X}_t)$ of $m_0 R^\mu_{0,t}$ and $m_0 R^\mu_{0,t}$ with $X_0 = \tilde{X}_0 \sim m_0$, such as defined in Section 3.2. Then

$$\|m_0 R^\mu_{0,t} - m_0 R^\mu_{0,t}\|_V \leq \mathbb{E}\left(1_{X_t \neq \tilde{X}_t}\left(V(X_t) + V(\tilde{X}_t)\right)\right) \leq 2\mathbb{E}\left(e^{\rho(X_0 + t)} I\left(X_t \neq \tilde{X}_t \mid X_0\right)\right).$$

36
Denoting $\tau_{\text{split}} = \inf\{t \geq 0, \ X_t \neq \bar{X}_t\}$, remark that, from $X_t \leq X_0 + t$, almost surely, for all $s \in [0, \tau_{\text{split}})$,

$$|\lambda_{\mu_1}^*(X_s) - \lambda_{\mu_2}^*(X_s)| \leq \frac{e^{\rho(X_0 + s)}}{K} \|\mu_1^s - \mu_2^s\|_V, \quad (46)$$

and then

$$\mathbb{P}\left(X_t \neq \bar{X}_t \mid X_0\right) = \mathbb{P}(\tau_{\text{split}} \leq t \mid X_0) \leq 1 - \exp\left(-\frac{e^{\rho(X_0 + t)}}{K} \int_0^t \|\mu_1^s - \mu_2^s\|_V ds\right) \leq \frac{e^{\rho(X_0 + t)}}{K} \int_0^t \|\mu_1^s - \mu_2^s\|_V ds,$$

so that, at the end,

$$\|m_0 R_{0,t}^{\mu_1} - m_0 R_{0,t}^{\mu_2}\|_V \leq \mathbb{E}\left(1_{X_t \neq \bar{X}_t} \left(V(X_t) + V(\bar{X}_t)\right)\right) \leq \frac{2}{K} e^{\rho (m_0 V^2)} \int_0^t \|\mu_1^s - \mu_2^s\|_V ds. \quad (47)$$

Similarly as in the proof of Lemma 4, taking $t_1 > 0$ small enough so that $2t_1 \exp(\rho t_1) |m_0(V^2)\vee C|/K \leq 1/2$, the map $(\mu_t)_{t \in [0,t_1]} \rightarrow (m_0 R_{0,t}^{\mu_1})_{t \in [0,t_1]}$ admits a unique fixed point, which we call by definition a solution of (42). Moreover, from (45), $m_{t_1}(V^2) \leq m_0(V^2) \vee \bar{C}$, so that the same argument (with the same $t_1$) yields a solution on $[t_1, 2t_1]$, and then on all $\mathbb{R}_+$. From (44), such a solution $t \mapsto m_t$ satisfies

$$m_t(V) \leq e^{-t/2} m_0(V) + \left(1 - e^{-t/2}\right) \bar{C} \Rightarrow \lambda_{m_t}(x) \leq 1 + g_t(x) + Ke^{\rho x} \left(m_0(V) + \bar{C}\right), \quad (48)$$

for all $t, x \geq 0$. This bound is now uniform in time, although not uniform in $x$.

For $t$ large enough, from (45), $m_t(V^2) \leq \bar{C} + 1$, so we can now restrict the study to initial distributions that satisfies $m_0(V^2) \leq \bar{C} + 1$. Then, with the same argument used to establish (47), there exists $C'' > 0$ such that for all $m_0$, $h_0$ with $m_0(V^2)$, $h(V^2) \leq \bar{C} + 1$ and all $t_0 \geq 0$,

$$\|m_t - h_t\|_V \leq C'' \int_0^{t_0} \|m_s - h_s\|_V ds,$$

which is the “splitting estimates” part of the proof of Theorem 3. Since the Lyapunov contraction (44) has already been established, what remains to obtain the “merging estimates” part, i.e. the equivalent of Lemma 13 is just a time $t_0$ and a probability $\alpha > 0$ to merge two processes in a time $t_0$, given that they started in some compact set. For the linear process (i.e. with $g_0 = 0$), this has been done in [13, 2]. As in Section 3.2.1 to couple two non-linear processes, we simply couple two linear processes (i.e. with jump rate $1 + g_1$ and we hope that no non-linear jump occurs in the time interval $[0, t_0]$. In the present case, however, we should be cautious, since the non-linear jump rate is not uniformly bounded in $x$. Nevertheless, if the starting point are taken in $[0, R_0]$ for some $R_0 > 0$, then during the time interval $[0, t_0]$ the processes remain in $[0, R_0 + t_0]$, on which the non-linear jump rate is bounded by (48). Thus, conditionally to the fact that they started in $[0, R_0]$ the probability that two coupled non-linear processes with initial law $m_0, h_0$ that satisfy $m_0(V^2), h(V^2) \leq \bar{C} + 1$ have merged at time $t_0$ is uniformly bounded away from 0. Then the strategy of the proof of Theorem 5 can be adapted to get, provided that $K$ is small enough, that (42) admits a unique equilibrium, toward which all solutions with $m_0 \in \mathcal{P}_V$ converges exponentially fast.
4.7 Markov processes with delay

We claimed in the introduction that the coupling arguments used in\[3\] to deal with non-linear equation can also be used to deal with self-interacting processes. A general study of processes $$(X_t)_{t \geq 0}$$ interacting with a weighted empirical distribution

$$\nu_t = \left( \int_0^t w(t-s)ds \right)^{-1} \int_0^t \delta_{X_s} w(t-s)ds$$

for all $$t \geq 0$$ such that $$\int_0^t w(t-s)ds \neq 0$$ exceeds the scope of the present paper, and is thus postponed. Nevertheless, as a proof of principle, consider the simple case where $$w$$ is a Dirac mass at some given time $$t_0$$. In other words, $$X_t$$ follows some Markovian dynamics and jumps at a rate and to a position that depends on $$X_{t-t_0}$$. In particular, $$Y_t = (X_s)_{s \in [t-t_0, t]}$$ is a Markov process. Nevertheless it is not necessary to consider the somewhat complicated process $$(Y_t)_{t \geq 0}$$ to obtain a speed of convergence toward equilibrium for the law of $$X_t$$. If there is some probability to couple two non-delayed processes starting at different positions and if the delayed jump rate is bounded, then there is some probability to merge two delayed processes at some positive time $$t_1$$ and then there is some probability that no delayed jump occurs in a time interval of length $$t_0$$. If that happens, then after time $$t_0 + t_1$$ the two processes share the same position and the same memory, and they will stay equal forever. An exponential convergence toward some equilibrium is obtained for the law of $$X_t$$ (which is not the solution of an integro-differential equation).

Acknowledgements

Pierre Monmarché acknowledges support from the French ANR-17-CE40-0030 - EFI - Entropy, flows, inequalities.

References


