Convergence of the Fleming-Viot algorithm: uniform in time estimates in a compact soft case.

Lucas Journel, Pierre Monmarché

October 14, 2019

Abstract

We establish the convergences (with respect to the simulation time \( t \); the number of particles \( N \); the timestep \( \gamma \)) of the Fleming-Viot algorithm toward the quasi-stationary distribution of a diffusion on the \( d \)-dimensional torus, killed at a smooth rate. In these conditions, quantitative bounds are obtained that, for each parameter \( t \to \infty \), \( N \to \infty \) or \( \gamma \to 0 \) are independent from the two others.

1 Introduction

1.1 The problem

Start from the diffusion on the \( d \)-dimensional periodic flat torus \( \mathbb{T}_d \)

\[
\mathrm{d}Z_t = b(Z_t) \mathrm{d}t + \mathrm{d}B_t
\]

with \( b \in C^1(\mathbb{T}_d) \), where \( (B_t)_{t \geq 0} \) is a \( d \)-dimensional Brownian motion. Add a killing rate \( \lambda \in C(\mathbb{T}_d) \) and, given a standard exponential random variable \( E \) independent from \( (Z_t)_{t \geq 0} \), define the death time

\[
T = \inf \left\{ t \geq 0, \ E \leq \int_0^t \lambda(Z_s) \mathrm{d}s \right\}.
\]

Then a probability measure \( \nu \) on \( \mathbb{T}_d \) is said to be a quasi-stationary distribution (QSD) associated to the SDE (1) and the rate \( \lambda \) if

\[
\text{Law}(Z_0) = \nu \quad \Rightarrow \quad \forall t \geq 0, \ \text{Law}(Z_t \mid T > t) = \nu.
\]

In our case, there exists a unique QSD \( \nu_* \) (see e.g. [2, Theorem 2.1]) and, whatever the initial distribution \( \eta_0 \) of \( Z_0 \),

\[
\text{Law}(Z_t \mid T > t) \underset{t \to \infty}{\longrightarrow} \nu_*. \]

The Fleming-Viot algorithm is designed to approximate \( \nu_* \). The present work is dedicated to the proof of convergence of this algorithm. This problem has already been addressed by many authors in various contexts (see e.g. [3, 4] and references within). A first novelty of the present work is that we take into account the time-discretization of the continuous-time diffusion. That way, we establish error bounds between the theoretical target QSD and the empirical measure indeed obtained with an actual implementation of the algorithm. There are three sources of errors: first, the continuous-time SDE (1) has to be discretized with some time step \( \gamma > 0 \). Second, as will be detailed below,
a non-linearity in the theoretical algorithm has to be approximated by a system of \( N \) particles. This leads to the definition of an ergodic Markov chain whose invariant measure is close, in some sense, to the QSD. But then this Markov chain is only run for a finite simulation time \( t = m\gamma, m \in \mathbb{N} \). A third error term then comes from the fact that stationarity is not fully achieved. We will obtained quantitative error bounds in \( \gamma, N \) and \( t \).

Note that we restrict the study to a compact state space. Moreover, we only consider soft killing at some continuous rate, and no hard killing which would correspond to the case where \( T \) is the escape time from some sub-domain (see e.g. [1] [5]). Finally, as will be seen below, as far as the long-time behaviour of the process is concerned we will work in a perturbative regime, namely we will assume that the variations of \( \lambda \) are small with respect to the mixing time of the diffusion [11]. Although already interesting by itself, this restricted framework can be thought as a toy model motivated in particular by the case that arises in the parallel replica algorithm [8]. In that case, \( T \) is the escape time for (1) from a bounded metastable domain, so that the lifespan of the process is expected to be larger than its mixing time (and to depend little from the initial condition, given it is far enough from the boundary). Hence, the compact and perturbative assumptions are consistent with this objective. The restriction to smooth killing rate, however, is made to avoid additional difficulties in the hard case where, even in the metastable case, the probability to leave the domain is high (and exhibits high variations) when the process is close to its boundary. A motivation of our study is that we hope our method can be extended to the metastable hard case by combining it with some Lyapunov arguments. This study is postponed to future work.

### Notations and conventions

We respectively denote \( \mathcal{P}(F) \) and \( \mathcal{B}(F) \) the set of probability measures and of Borel sets of a Polish space \( F \). Functions on \( \mathbb{T}^d \) are sometimes identified to \([0, 1]^d\)-periodic functions, and similar non-ambiguous identifications are performed, for instance if \( x \in \mathbb{T}^d \) and \( G \) is a \( d \)-dimensional standard gaussian random variable, \( x + G \) has to be understood in \( \mathbb{T}^d \), etc. A Markov kernel \( Q \) on \( F \) is indiscriminately understood as, first, a function from \( F \) to \( \mathcal{P}(F) \), in which case we denote \( Q : x \mapsto Q(x, \cdot) \) (where \( Q(x, \cdot) \) denotes the probability \( A \in \mathcal{B}(F) \mapsto Q(x, A) \in [0, 1] \)); second, a Markov operator on bounded measurable functions on \( F \), in which case we denote \( Q : f \mapsto Qf \) (where \( Qf(x) = \int f(w)Q(x, dw) \)); third, by duality, a function on \( \mathcal{P}(F) \), in which case we denote \( Q : \mu \mapsto \mu Q \) (so that \( \mu(Qf) = (\mu Q)f \)). In particular, \( Q(x, \cdot) = \delta_xQ \) for \( x \in F \). If \( \mu \in \mathcal{P}(F) \) and \( k \in \mathbb{N}_+ \), we denote \( \mu^{\otimes k} \in \mathcal{P}(F^k) \) the law of a \( k \)-uplet of independent random variables with law \( \mu \). Similarly, if \( Q \) is a Markov kernel on \( F \), we denote \( Q^{\otimes k} \) the kernel on \( F^k \) such that \( Q^{\otimes k}(x, \cdot) = Q(x_1, \cdot) \otimes \cdots \otimes Q(x_k, \cdot) \) for all \( x = (x_1, \ldots, x_k) \in F^k \). We denote \( \mathcal{E}(1) \) the exponential law with parameter 1, \( \mathcal{U}(I) \) the uniform law on a set \( I \) and \( \mathcal{N}(m, \Sigma) \) the Gaussian law with mean \( m \) and variance matrix \( \Sigma \). We use bold letters for random variables in \( \mathbb{T}^d \mathbb{N} \) and decompose them in \( d \)-dimensional coordinates, like \( X = (X_1, \ldots, X_N) \) with \( X_i \in \mathbb{T}^d \), or \( X_1 = (X_{1,1}, \ldots, X_{N,1}) \).

#### 1.2 The algorithm and main result

Starting from the diffusion (1) killed at time \( T \) given by (2), we introduce two successive approximations. The first is time discretization. For a given time step \( \gamma > 0 \) and a sequence \((G_k)_{k \in \mathbb{N}}\) of independent random variables with law \( \mathcal{N}(0, I_d) \), we consider the Markov chain on \( \mathbb{T}^d \) given by \( \tilde{Z}_0 = Z_0 \) and

\[
\forall k \in \mathbb{N}, \quad \tilde{Z}_{k+1} = \tilde{Z}_k + \gamma b(\tilde{Z}_k) + \sqrt{\gamma} G_k \tag{3}
\]
and, given $E \sim \mathcal{E}(1)$ independent from $(G_k)_{k \in \mathbb{N}}$ and $Z_0$, 
\[
\hat{T} = \inf \left\{ t = n\gamma, \; n \in \mathbb{N}_*, \; E \leq \sum_{k=1}^{n} \lambda(\tilde{Z}_k) \right\}.
\]

From classical results for Euler schemes of diffusions, it is quite clear that, for any $A \in \mathcal{B}(\mathbb{T}^d)$ and all $t \geq 0$,
\[
\mathbb{P}\left( \tilde{Z}_{[t/\gamma]} \in A, \; \hat{T} < t \right) \xrightarrow{\gamma \to 0} \mathbb{P}(Z_t \in A, \; T < t),
\]
(see Corollary 3 for a proof) from which, for all $t \geq 0$,
\[
\text{Law}\left( \tilde{Z}_{[t/\gamma]} \mid \hat{T} < t \right) \xrightarrow{\gamma \to 0} \text{Law}(Z_t \mid T < t).
\]

Note that, from the memoryless property of the exponential law, given a sequence $(U_k)_{k \in \mathbb{N}}$ of independent variables uniformly distributed over $[0, 1]$ and independent from $(G_k)_{k \in \mathbb{N}}$ and $Z_0$, then $((Z_n)_{n \in \mathbb{N}}, \hat{T})$ has the same joint distribution as $((\tilde{Z}_n)_{n \in \mathbb{N}}, \hat{T})$ with
\[
\hat{T} = \inf \left\{ t = n\gamma, \; n \in \mathbb{N}_*, \; U_n \leq p(\tilde{Z}_n) \right\}
\]
where $p(z) = 1 - \exp(-\gamma \lambda(z))$ is the probability that, arriving at state $z$, the chain is killed.

A naive Monte Carlo sampler for the QSD would be to simulate $N$ independent copies of the chain (3) killed with probability $z \mapsto p(z)$ and to consider after a large number of iterations the distribution of the copies that have survived. However, after a long time, most copies (possibly all) would have died and the estimator would be very bad. We now introduce the Fleming-Viot algorithm that tackles this issue by resurrecting dead particles.

Denote $K : \mathbb{T}^d \rightarrow \mathcal{P}(\mathbb{T}^d)$ the Markov kernel associated with the transition (3), i.e.
\[
K f(x) = (2\pi)^{-d/2} \int_{\mathbb{R}^d} f(x + \gamma b(y) + \sqrt{\gamma} y) e^{-\frac{1}{2} |y|^2} dy.
\]

For $\mu \in \mathcal{P}(\mathbb{T}^d)$, let $Q_{\mu}$ be the Markov kernel such that, for all $x \in \mathbb{T}^d$, $Q_{\mu}(x, \cdot)$ is the law of the random variable $X$ defined as follows. Let $(X_k, U_k)_{k \in \mathbb{N}}$ be a sequence of independent random variables such that, for all $k \in \mathbb{N}$, $X_k$ and $U_k$ are independent, $U_k \sim \mathcal{U}([0, 1])$ and $X_0 \sim K(x, \cdot)$ while, for $k \geq 1$, $X_k \sim \mu K$. Let $H = \inf\{k \in \mathbb{N}, U_k \geq p(X_k)\}$, and set $X = X_H$. Since $\lambda$ is bounded, $p$ is uniformly bounded away from 1 and thus $H$ is almost surely finite, so that $Q_{\mu}$ is well-defined.

In other words, a random variable $X \sim Q_{\mu}(x, \cdot)$ may be constructed through the following algorithm (in which new means: independent from all the variables previously drawn).

1. Draw $X_0 \sim \mathcal{N}(x + \gamma b(x), \gamma I_d)$ and a new $U_0 \sim \mathcal{U}([0, 1])$.
2. If $U_0 \geq p(X_0)$, set $X = X_0$ in $\mathbb{T}^d$ (in that case, we say the particle has moved from $x$ to $X_0$ without dying).
3. If $U_0 < p(X_0)$ then set $i = 1$ and, while $X$ is not defined, do:
   a. Draw a new $X'_i$ distributed according to $\mu$, a new $X_i \sim \mathcal{N}(X'_i + \gamma b(X'_i), \gamma I_d)$ and a new $U_i \sim \mathcal{U}([0, 1])$.
   b. if $U_i \geq p(X_i)$, set $X = X_i$ in $\mathbb{T}^d$ (in that case, we say the particle has died, resurrected at $X'_i$, moved to $X_i$ and survived).
   c. If $U_i < p(X_i)$, set $i \leftarrow i + 1$ (in that case, we say the particle has died, resurrected at $X'_i$, moved to $X_i$ and died again) and go back to step (a).
From this, we define a chain \((Y_k)_{k \in \mathbb{N}}\) as follows. Set \(Y_0 = Z_0\) and suppose that \(Y_k\) has been defined for some \(k \in \mathbb{N}\). Let \(\eta_k = \text{Law}(Y_k)\), and draw a new \(Y_{k+1} \sim Q_{\eta_k}(Y_k, \cdot)\). This somewhat intricate definition is motivated by the following results (whose proof is postponed to Section 2):

**Proposition 1.** For all \(n \in \mathbb{N}\)

\[
\eta_n = \mathcal{L}(\hat{Z}_n \mid \hat{T} \geq n\gamma).
\]

In particular, as \(n \to \infty\), the law \(\eta_n\) of \(Y_n\) converges toward the QSD of \(\hat{Z}\). Unfortunately, it is impossible to sample \((Y_k)_{k \in \mathbb{N}}\) in practice since this would require to sample according to \(\eta_k\) for any \(k \in \mathbb{N}\). This is a classical problem for non-linear McKean-Vlasov diffusions. Thus, motivated by the Law of Large Numbers, we lead to a second approximation, which is to use mean-field interacting particles. For a fixed \(N \in \mathbb{N}_*\) and for \(x = (x_i)_{i \in [1, N]} \in \mathbb{T}^d\), we denote

\[
\pi(x) := \frac{1}{N} \sum_{i=1}^N \delta_{x_i} \in \mathcal{P}(\mathbb{T}^d)
\]

the associated empirical distribution. Then we define the Markov operator \(R\) on \(\mathbb{T}^{dN}\) as

\[
R(x, \cdot) = Q_{\pi(x)}(x_1, \cdot) \otimes \cdots \otimes Q_{\pi(x)}(x_N, \cdot).
\]

In other words, a random variable \(Y \sim Q(x, \cdot)\) is such that the \(Y_i\)'s are independent with \(Y_i \sim Q_{\pi(x)}(x_i, \cdot)\). In order to specify the parameters involved, we will sometimes write \(R_{N, \gamma}\) for \(R\).

Let us informally describe the transitions of a Markov chain \((X_k)_{k \in \mathbb{N}}\) associated to \(R\), which we call a Fleming-Viot system of \(N\) interacting particles: the \(i^{th}\) particle follows the transition given by \([8]\) independently from the other particles until it dies. If it dies at a step \(k \in \mathbb{N}_*\), then it is resurrected on another particle \(X_{J, k-1}\) with \(J\) uniformly distributed over \([1, N]\) (in particular and contrary to some other works, \(J = i\) is not excluded) and immediately performs a step of \([8]\); if it dies again after this unique step, it is resurrected again and performs a new step, and so on until it is not killed after a resurrection and an Euler scheme step. Then this is the new value \(X_{i, k}\) from which the particle follows again the transitions \([8]\) until its next death, etc.

Note that there is no problem of simultaneous death since at step \(k\) the particles are resurrected on positions at step \(k - 1\), which are well-defined even if all particles die at once at step \(k\).

It is easily seen that \(R\) admits a unique invariant measure toward which the law of the associated Markov chain converges exponentially fast (in the total variation sense for instance), but a naive argument yields a convergence rate that heavily depends on \(N\) (and possibly \(\gamma\)). Similarly, classical studies can be conducted for the limits \(N \to \infty\) and \(\gamma \to 0\) but again with estimates that are typically exponentially bad with respect to the total simulation time (see in particular Propositions \([7]\) and \([10]\)). In the following we will focus on a somewhat perturbative regime under which we will establish estimates for each of these limits that are uniform with respect to the other parameters. Even for the continuous process (\(\gamma = 0\)), such uniform results are new (see Corollaries \([13]\) and \([16]\)).

Recall that the \(W_1\) Wasserstein distance between \(\mu, \nu \in \mathcal{P}(\mathbb{T}^d)\) is defined by

\[
W_1(\mu, \nu) = \inf \{ \mathbb{E}(|X - Y|) : X \sim \mu, \ Y \sim \nu \}.
\]

Our main result give a quantitative bound on the error made in practice by approximating \(\nu_s\) by the empirical distribution of a Fleming-Viot particle system:
Theorem 2. There exists $c_0, \gamma_0 > 0$ that depends only on the drift $b$ and the dimension $d$ such that, if $\lambda$ is Lipschitz with a constant $L_\lambda$ such that
\[
L_\lambda e^{\gamma \|\lambda\|_{\infty}} < c_0 ,
\]
then there exists $C, \kappa > 0$ such that for all $N \in \mathbb{N}$, $\gamma \in (0, \gamma_0]$, $t \geq 0$ and $\mu_0 \in \mathcal{P}(\mathbb{T}^{dN})$, if $(X^n_k)_{k \in \mathbb{N}}$ is a Markov chain with initial distribution $\mu_0$ and transition kernel $R_{N, \gamma}$, then
\[
\mathbb{E} \left[ \mathcal{W}_1 \left( \pi(X_{\lfloor t \gamma \rfloor}), \nu_x \right) \right] \leq C \left( \sqrt{\gamma} + \alpha(N) + e^{-\kappa t} \right),
\]
where
\[
\alpha(N) = \begin{cases} 
N^{-1/2} & \text{if } d = 1, \\
N^{-1/2} \ln(1 + N) & \text{if } d = 2, \\
N^{-1/d} & \text{if } d > 2.
\end{cases}
\]

The speeds of the different convergences (exponential in the simulation time, with the square-root of the timestep and with $\alpha$ of the number of particles) are optimal since they are optimal for non-interacting diffusions (i.e. the case $\lambda = 0$), see in particular [6] for the large $N$ asymptotic.

Other intermediary results will be established in the rest of the paper that are interesting by themselves: long-time convergence at fixed $N, \gamma$ in Proposition 5, propagation of chaos (i.e. $N \to \infty$) and continuous-time limit at a fixed time (even without the condition (4)) respectively in Propositions 7 and 10. From that, results for the continuous-time process ($\gamma = 0$), the equilibria ($t = \infty$) or the non-linear process ($N = \infty$), or when two parameters among three are sent to their limits, are then simple corollaries, see Section 2.5.

Note that $\exp(-\gamma \lambda(x))$ is the probability that the chain is not killed when it arrives at state $x$. The time step $\gamma$ should be chosen in such a way that this probability is relatively large, say at least one half. In that case, $\exp(\gamma \|\lambda\|_{\infty})$ is typically close to 1. In other words, (4) is mostly a condition about $L_\lambda$ being small enough.

This perturbation condition is different from the one considered in [11], where $\|\lambda\|_{\infty}$ rather than $L_\lambda$ is supposed to be small (while our main arguments are a direct adaptation of the coupling arguments of [11]). This difference comes from the fact that, in the present study, we work with the $\mathcal{W}_1$ distance rather than the total variation one (which is a Wasserstein distance but associated to the discrete metric $d(x, y) = 1_{x \neq y}$). Indeed, in our coupling arguments, we need to control $|\lambda(x) - \lambda(y)|$ the difference between the death rates of two processes at different locations, which is bounded here by $L_\lambda |x - y|$ and in [11] by $2\|\lambda\|_{\infty} 1_{x \neq y}$. In fact our argument for the long-time convergence may easily be adapted to the total variation distance framework, following [11]. Nevertheless this would be more troublesome in the study of the limit $N \to \infty$. Then, one needs to couple $\eta_k$ (that admits a density with respect to the Lebesgue measure) with $\pi(X_k)$ (which is a sum of Dirac masses), so that the total variation distance is not adapted. This may be solved by considering $\mathcal{W}_1 \to$ total variation regularization results for (Euler schemes of) diffusions, that can be established by coupling arguments again. Nevertheless, in order to focus on the other difficulties of the problem and for the sake of clarity, we decided to stick to the $\mathcal{W}_1$ distance in all the different results of this work.

2 Proofs

Let us first establish the preliminary result stated in the introduction:

Proof of Proposition 4. For $n \in \mathbb{N}$, denote
\[
\eta_n = \text{Law}(Y_n), \quad \nu_n = \text{Law} \left( \tilde{Z}_n \mid \tilde{T} \geq n\gamma \right).
\]
Since \( \nu_0 = \eta_0 \), suppose by induction that \( \nu_n = \eta_n \) for some \( n \in \mathbb{N} \). Keeping the notations introduced of the definition of the kernel \( Q_\mu \), consider the events \( B_k = \{ U_k \geq p(X_k) \} \).

Then, for all bounded measurable \( f \),

\[
Q_\mu f(x) = \mathbb{E}(f(X)) = \mathbb{E}\left( f(X) \sum_{k \in \mathbb{N}} \mathbb{1}_{B_k \cap (\bigcap_{j=0}^{k-1} B_j^c)} \right) = \mathbb{E}(f(X_0)\mathbb{1}_{U_0 \geq p(X_0)}) + \sum_{k \geq 1} \mathbb{E}(f(X_k)\mathbb{1}_{U_k \geq p(X_k)}) \prod_{j=0}^{k-1} \mathbb{P}(B_j^c) = K[f(1-p)](x) + \sum_{k \geq 1} \mu K[f(1-p)](\mu Kp)^k Kp(x).
\]

In particular, integrating with respect to \( \mu \), we obtain

\[
\mu Q_\mu f = \mu K[f(1-p)] \sum_{k \in \mathbb{N}} (\mu Kp)^k = \frac{\mu K[f(1-p)]}{\mu K[1-p]}.
\]

Applied with \( \mu = \eta_n \), this reads

\[
\eta_{n+1} f = \mathbb{E}(f(Y_{n+1})) = \mathbb{E}(\mathbb{E}(f(Y_{n+1}) | Y_n)) = \eta_n Q_{\eta_n} f = \frac{\eta_n K[f(1-p)]}{\eta_n K[1-p]}.
\]

On the other hand,

\[
\mathbb{E}\left( f(\tilde{Z}_{n+1}) \mathbb{1}_{\tilde{T} > (n+1)\gamma} \right) = \mathbb{E}\left( f(\tilde{Z}_{n+1}) \mathbb{1}_{\tilde{T} > n\gamma} \mathbb{1}_{U_n \geq p(\tilde{Z}_{n+1})} \right) = \mathbb{E}\left( f(\tilde{Z}_{n+1}) \left( 1 - p \left( \tilde{Z}_{n+1} \right) \right) \mathbb{1}_{\tilde{T} > n\gamma} \right) = \mathbb{P}(\tilde{T} > n\gamma) \nu_n K[f(1-p)],
\]

from which

\[
\nu_{n+1} f = \frac{\mathbb{E}\left( f(\tilde{Z}_{n+1}) \mathbb{1}_{\tilde{T} > (n+1)\gamma} \right)}{\mathbb{P}(\tilde{T} > (n+1)\gamma)} = \frac{\mathbb{P}(\tilde{T} > n\gamma) \nu_n K[f(1-p)]}{\mathbb{P}(\tilde{T} > n\gamma) \nu_n K[1-p]} = \frac{\nu_n K[f(1-p)]}{\nu_n K[1-p]},
\]

which concludes.

\[\square\]

### 2.1 The basic coupling

The long-time estimates needed to prove convergence toward equilibrium and uniform in time estimates in \( N \) and \( \gamma \) are based on the fact that, as long as particles don’t die, they follow the chain \([3]\) which, like its continuous-time counterpart \([1]\), have some mixing properties. In order to quantify the latters, we start by stating \([10]\) Corollary 2.2 in a suitable way in our context. For \( \rho \) a distance on some Polish space \( F \), denote \( \mathcal{W}_\rho \) the corresponding Wasserstein distance on \( \mathcal{P}(F) \), defined by

\[
\mathcal{W}_\rho(\mu, \nu) = \inf \{ \mathbb{E}(\rho(X, Y)) : X \sim \mu, Y \sim \nu \}.
\]

If \( X \sim \mu \) and \( Y \sim \nu \), we call \((X, Y)\) a coupling of \( \mu \) and \( \nu \). If \((X, Y)\) is a coupling for which the infimum in \((5)\) is attained, we say that it is an optimal coupling. From \([12]\) Corollary 5.22, such an optimal coupling always exists.
Proposition 3. There exists $c_1, a, \gamma_0 > 0$ (that all depend only on the drift $b$ of $\Pi$ and on the dimension $d$) such that, denoting $\rho(x,y) = (1 - \exp(-a|x-y|))/a$ for $x,y \in \mathbb{T}^d$, then $\rho$ is a metric on $\mathbb{T}^d$ with
\[
\forall \gamma \in (0, \gamma_0], \forall \mu, \nu \in \mathcal{P}(\mathbb{T}^d), \quad \mathcal{W}_\rho(\mu K, \nu K) \leq (1 - c_1 \gamma)\mathcal{W}_\rho(\mu, \nu).
\]

Proof. This is [10] Corollary 2.2, except that the latter is stated in $\mathbb{R}^d$ with some contraction assumption outside some compact ball. The proof is straightforwardly adapted to the case of a diffusion with smooth drift on the torus. In particular, considering the notations of [10], in this proof we can take $\mathcal{R}$ (hence $r_1$ and $r_2$) larger than the diameter of the torus, in which case the function $f$ defined in [10] Equation (1.1)] is simply $f(r) = (1 - \exp(-ar))/a$, which concludes. 

In the rest of the paper, $\rho$ is the metric and $c_1, a, \gamma_0$ are the constants given by Proposition 3. Remark that $\rho$ is equivalent to the Euclidian metric, with
\[
\beta|x-y| \leq \rho(x,y) \leq |x-y| \quad \text{for} \quad \beta = 2(1 - e^{-a\sqrt{d}/2})/(a\sqrt{d}),
\]
where we used that the diameter of $\mathbb{T}^d$ is $\sqrt{d}/2$ and that $r \mapsto (1 - \exp(-ar))/a$ is a concave function with derivative 1 at zero. In particular, $W_1$ and $W_\rho$ are equivalent.

Now, in the Fleming-Viot algorithm, the contraction property of the chain [5] may be counterbalanced by the death/resurrection mechanism through which particles interact. Indeed, considering two systems of $N$ interacting particles, for $i \in [1,N]$ the previous result means that we can couple the $i$th particles of both systems to get closer one to the other (on average), as long as they don’t die. But then, one of the two particles can die and resurrect far from the other, or even if they die simultaneously they may resurrect far apart one from the other. That being said, first, the closer they get, the easier it is to couple them in order to die simultaneously, and second, when they die simultaneously, keeping the particles close one to the other amount to do a suitable coupling of the laws from which the particles are resurrected. This is quantified in the following proposition.

In all the rest of the paper, we suppose that $\lambda$ is $L_\lambda$-Lipschitz (but not necessarily that $\Pi$ holds).

Proposition 4. Let $\mu_1, \mu_2, \nu_1, \nu_2 \in \mathcal{P}(\mathbb{T}^d)$ and let $(X_0, Y_0)$ (resp. $(X_1, Y_1)$) be a coupling of $\nu_1 K$ and $\nu_2 K$ (resp. $\mu_1 K$ and $\mu_2 K$). Then
\[
\mathcal{W}_\rho(\nu_1 Q_{\mu_1}, \nu_2 Q_{\mu_2}) \leq h \left( \mathbb{E}(\rho(X_0, Y_0)) + \frac{q_0}{1 - q_1} \mathbb{E}(\rho(X_1, Y_1)) \right)
\]
where
\[
h = 1 - \min p + (a\beta)^{-1}\gamma L_\lambda
\]
and, considering $U \sim \mathcal{U}([0,1])$ independent from $(W_0, Z_0)$ and $(W_1, Z_1)$,
\[
q_i = \mathbb{P}(U < p(X_i) \land p(Y_i)), \quad i = 0, 1.
\]

Proof. Let $(X_k, Y_k, U_k)_{k \in \mathbb{N}}$ be a sequence of independent triplet of random variables such that, for all $k \in \mathbb{N}$, $U_k \sim \mathcal{U}([0,1])$ is independent from $(X_k, Y_k)$, which are such as defined in the proposition for $k = 0$ and 1 and, for $j > 1$, have the same distribution as $(X_1, Y_1)$. Set $H_1 = \inf\{n \in \mathbb{N}, U_n < p(X_n)\}$ and $H_2 = \inf\{n \in \mathbb{N}, U_n < p(Y_n)\}$. Then, by considering the law of $(X_k, U_k)_{k \in \mathbb{N}}$ alone, it is clear that $X_{H_1} \sim \nu_1 Q_{\mu_1}$ and, similarly, $Y_{H_2} \sim \nu_2 Q_{\mu_2}$, so that
\[
\mathcal{W}_\rho(\nu_1 Q_{\mu_1}, \nu_2 Q_{\mu_2}) \leq \mathbb{E}(\rho(X_{H_1}, Y_{H_2})).
\]
Different cases are distinguished depending on the value of \( H_1 \) and \( H_2 \). In the simplest case, none of the particles dies:

\[
\mathbb{E}(\rho(X_{H_1}, Y_{H_2}) \mathbb{1}_{H_1=H_2=0}) = \mathbb{E}(\rho(X_0, Y_0) \mathbb{1}_{U_0 > \rho(X_0) \vee \rho(Y_0)}) \\
\leq \mathbb{E}(\rho(X_0, Y_0) \mathbb{1}_{U_0 > \min p}) \\
\leq (1 - \min p) \mathbb{E}(\rho(X_0, Y_0))
\]

where we used the independence between \( U_0 \) and \((X_0, Y_0)\). In the second case, only one particle dies: using that \( \|\rho\|_{\infty} \leq 1/a \),

\[
\mathbb{E}(\rho(X_{H_1}, Y_{H_2}) \mathbb{1}_{H_1 \wedge H_2 = 0 < H_1 \vee H_2}) \leq a^{-1} \mathbb{P}(U_0 \in [p(X_0) \wedge p(Y_0), p(X_0) \vee p(Y_0)]) \\
= a^{-1} \mathbb{E}(|p(X_0) - p(Y_0)|) \\
\leq a^{-1} \gamma L \lambda \mathbb{E}(|X_0 - Y_0|) \\
\leq (a \beta)^{-1} \gamma L \lambda \mathbb{E}(\rho(X_0, Y_0))
\]

In the third case, both particles die \( k \geq 1 \) times:

\[
\mathbb{E}(\rho(X_{H_1}, Y_{H_2}) \mathbb{1}_{H_1 \wedge H_2 = k}) = \mathbb{E}\left(\left(\rho(X_k, Y_k) \mathbb{1}_{U_k > \rho(X_k) \vee \rho(Y_k)} \prod_{j=0}^{k-1} \mathbb{1}_{U_j > \rho(X_j) \wedge \rho(Y_j)}\right)\right) \\
\leq q_0 q_1^{-1} \mathbb{E}(\rho(X_k, Y_k) \mathbb{1}_{U_k > \min p}) \\
\leq q_0 q_1^{-1} (1 - \min p) \mathbb{E}(\rho(X_1, Y_1))
\]

Finally, combining the computations of the last two cases, the fourth one reads, for \( k \geq 1 \),

\[
\mathbb{E}(\rho(X_{H_1}, Y_{H_2}) \mathbb{1}_{H_1 \wedge H_2 = k < H_1 \vee H_2}) \leq a^{-1} q_0 q_1^{-1} \mathbb{P}(U_k \in [p(X_k) \wedge p(Y_k), p(X_k) \vee p(Y_k)]) \\
\leq (a \beta)^{-1} q_0 q_1^{-1} \gamma L \lambda \mathbb{E}(\rho(X_1, Y_1))
\]

Summing these four cases concludes. \(\square\)

### 2.2 Long-time convergence

For \( N \in \mathbb{N}_+ \) denote \( \rho_N \) the metric on \( \mathbb{T}^d \mathbb{N} \) given by

\[
\rho_N(x, y) = \sum_{i=1}^{N} \rho(x_i, y_i).
\]

**Proposition 5.** There exists \( c_2 > 0 \) (that depends only on the drift \( b \) of \( \Pi \) and on the dimension \( d \)) such that for all \( \gamma \in (0, \gamma_0] \) \( N \in \mathbb{N} \), and all \( \mu, \nu \in \mathcal{P}(\mathbb{T}^d \mathbb{N}) \),

\[
\mathcal{W}_{\rho_N}(\mu R_{N, \gamma}, \nu R_{N, \gamma}) \leq (1 - \gamma \kappa) \mathcal{W}_{\rho_N}(\mu, \nu).
\]

with

\[
\kappa = c_1 - c_2 L \lambda e^{\gamma \|\lambda\|_{\infty}}.
\]

This means that, with respect to the metric \( \rho_N \), \( R_{N, \gamma} \) has a Wasserstein curvature of \( \gamma \kappa \) in the sense of \( \mathcal{W} \).

**Proof.** It is in fact sufficient to prove this for \( \mu = \delta_x \) and \( \nu = \delta_y \) for any \( x, y \in \mathbb{T}^d \mathbb{N} \).

Indeed, assuming the result proven for Dirac masses, in the general case, considering
(X₀, Y₀) an optimal coupling of µ and ν and (X₁, Y₁) an optimal coupling of R(X₀, ·) and R(Y₀, ·), then X₁ ∼ µR and Y₁ ∼ νR, so that

\[ W_{ρ_N} (µR, νR) \leq \mathbb{E} (ρ_N (X₁, Y₁)) \]

\[ = \mathbb{E} (\mathbb{E} (ρ_N (X₁, Y₁) | (X₀, Y₀))) \]

\[ = \mathbb{E} (\mathbb{E} (W_{ρ_N} (δ_{X₀}R, δ_{Y₀}R) | (X₀, Y₀))) \]

\[ \leq (1 - γκ) \mathbb{E} (ρ_N (X₀, Y₀)) \]

\[ = (1 - γκ) W_{ρ_N} (µ, ν) . \]

Hence, in the following, we fix x, y ∈ TᵈN. Let (Xᵢ, Yᵢ)∈ [1, N] be independent pairs of random variables in Tᵈ where, for all i ∈ [1, N], (Xᵢ, Yᵢ) is an optimal coupling of Qₚᵢ(xᵢ, ·) and Qₚᵢ(yᵢ, ·). Then (X, Y) is a coupling of R(x, ·) and R(y, ·), so that

\[ W_{ρ_N} (δ_{X}R, δ_{Y}R) \leq \mathbb{E} (ρ_N (X, Y)) = \sum_{i=1}^{N} \mathbb{E} (ρ (Xᵢ, Yᵢ)) \]

\[ = \sum_{i=1}^{N} W_{ρ} (Qₚₐₙ(xᵢ, ·), Qₚₐₙ(yᵢ, ·)) . \]

We want to apply Proposition 4 with µ₁ = π(x), ν₁ = δₓᵢ, µ₂ = π(y) and ν₂ = δᵧᵢ. To do so, for all i ∈ [1, N], we consider (Xᵢ, Yᵢ) an optimal coupling of K(xᵢ, ·) and K(yᵢ, ·). From Proposition 3

\[ \mathbb{E} \left( ρ \left( \tilde{X}ᵢ, \tilde{Y}ᵢ \right) \right) \leq (1 - c₃γ) ρ(xᵢ, yᵢ) . \]

Moreover, if J ∼ υ([1, N]) is independent from the (Xᵢ, Yᵢ)’s, we remark that (X_J, Y_J) is a coupling of π(x)K and π(y)K. Proposition 4 applied with these couplings reads, for all i ∈ [1, N],

\[ W_{ρ} (Qₚₐₙ(xᵢ, ·), Qₚₐₙ(yᵢ, ·)) \leq h \left( \mathbb{E} \left( ρ (\tilde{X}ᵢ, \tilde{Y}ᵢ) \right) + \frac{qᵢ}{1 - qₛ} \mathbb{E} \left( ρ (\tilde{X}_J, \tilde{Y}_J) \right) \right) \]

(8)

where, if U ∼ υ([0, 1]) is independent from the previous variables,

\[ qᵢ := \mathbb{P} (U < p(\tilde{X}ᵢ) ∧ p(\tilde{Y}ᵢ)) \]

and, conditioning on the value of J,

\[ qₛ := \mathbb{P} (U < p(\tilde{X}_J) ∧ p(\tilde{Y}_J)) = \frac{1}{N} \sum_{i=1}^{N} qᵢ . \]

Summing 3 over i ∈ [1, N] and applying 7 yields

\[ W_{ρ_N} (δ_{X}R, δ_{Y}R) \leq h \left( (1 - c₃γ) \sum_{i=1}^{N} ρ(xᵢ, yᵢ) + \frac{N qₛ}{1 - qₛ} \mathbb{E} (ρ (\tilde{X}_J, \tilde{Y}_J)) \right) . \]

Applying Proposition 3 again,

\[ \mathbb{E} \left( ρ (\tilde{X}_J, \tilde{Y}_J) \right) = \frac{1}{N} \sum_{i=1}^{N} \mathbb{E} \left( ρ (\tilde{X}ᵢ, \tilde{Y}ᵢ) \right) \leq \frac{1}{N} (1 - c₃γ) \sum_{i=1}^{N} ρ(xᵢ, yᵢ) , \]

and the previous inequality becomes

\[ W_{ρ_N} (δ_{X}R, δ_{Y}R) \leq \frac{h (1 - c₃γ)}{1 - qₛ} ρ_N (x, y) . \]
Bounding $1 - q_* \geq 1 - \max p \geq \exp(-\gamma \|\lambda\|_\infty)$ and $\max p - \min p \leq \sqrt{d/2} \gamma L_\lambda$ yields
\[
\frac{h(1 - c_1 \gamma)}{1 - q_*} \leq (1 - c_1 \gamma) \frac{1 - \max p + \max p - \min p + (a\beta)^{-1} \gamma L_\lambda}{1 - \max p} \\
\leq 1 - c_1 \gamma + \gamma L_\lambda e^{\gamma \|\lambda\|_\infty} \left((a\beta)^{-1} + \sqrt{d}\right),
\]
which concludes. \hfill \qed

As a direct consequence, assuming that \((\ref{equation:optimality-condition})\) holds with $c_0 = c_1/c_2$, then Proposition \((\ref{proposition:optimality})\) gives the contraction
\[
W_{\rho_N}(\mu R^m, \nu R^m) \leq e^{-\kappa m} W_{\rho_N}(\mu, \nu),
\]
with $\kappa > 0$ that does not depends on $N$ nor on $\gamma$. Since $\mathcal{P}(\mathbb{T}^d N)$ is complete for $W_1$ (hence with $W_{\rho_N}$) the Banach fixed-point theorem implies then that $R$ admits a unique invariant measure toward which it converges at rate $\gamma \kappa$.

In the rest of the paper, $\kappa$ is given by \((\ref{equation:noise-noise-distance})\) (but is not necessarily assumed positive).

### 2.3 Propagation of chaos

Recall that $\eta_k$ is the law at time $k$ of the non-homogeneous Markov chain $(Y_k)_{k \in \mathbb{N}}$ on $\mathbb{T}^d$ introduced in Section 1.2 with transition kernels $Q_{\eta_k}$ and initial condition $\eta_0$, and that $R = R_{\gamma} N$ is the transition kernel of the Markov chain $(X_k)_{k \in \mathbb{N}}$ on $\mathbb{T}^{d N}$.

**Lemma 6.** There exist $C_1 > 0$ such that for all $N \in \mathbb{N}$, $\gamma \in (0, \gamma_0]$, $\eta \in \mathcal{P}(\mathbb{T}^d)$ and $\mu \in \mathcal{P}(\mathbb{T}^{d N})$,
\[
W_{\rho_N}(\mu R^m, \nu Q^\otimes N) \leq \gamma N C_1 \int_{\mathbb{T}^d N} W_{\rho}(\pi(x), \eta) \mu(dx).
\]

**Proof.** Similarly to the proof of Proposition \((\ref{proposition:optimality})\) we start with the case $\mu = \delta_x$ for some $x \in \mathbb{T}^{d N}$. Let $(X_i, Y_i)_{i \in [1, N]}$ be $N$ independent pairs of random variable such that for all $i \in [1, N]$, $(X_i, Y_i)$ is an optimal coupling of $Q_{\pi(x)}(x_i, \cdot)$ and $Q_{\eta}(x_i, \cdot)$. Then $(X, Y)$ is a coupling of $R_{\gamma} (x, \cdot)$ and $Q^\otimes N(x, \cdot)$, so that
\[
W_{\rho_N}(\delta_x R_N, \delta_x Q_{\eta}^\otimes N) \leq \mathbb{E}(\rho_N(X, Y)) = \sum_{i=1}^N \mathbb{E}(\rho(X_i, Y_i)) = \sum_{i=1}^N W_{\rho}(\delta_x Q_{\pi(x)}(x_i), \delta_x \eta).
\]

From Proposition \((\ref{proposition:optimality})\) (bounding $q_0 \leq \max p \leq \gamma \|\lambda\|_\infty$ and $1 - q_1 \geq 1 - \max p \geq \exp(-\gamma_0 \|\lambda\|_\infty)$)
\[
W_{\rho}(\delta_x Q_{\pi(x)}, \delta_x \eta) \leq \gamma \|\lambda\|_\infty (1 + (a\beta)^{-1} \gamma_0 L_\lambda) e^{\gamma_0 \|\lambda\|_\infty} W_{\rho}(\pi(x), \eta) =: \gamma C_1 W_{\rho}(\pi(x), \eta).
\]

Now in the general case where $\mu$ is not a Dirac mass, considering $Z_0 \sim \mu$, and $(Z_1, Z_2)$ an optimal coupling of $R(Z_0, \cdot)$ and $Q_{\eta}^\otimes (Z_0, \cdot)$ and conditioning with respect to $Z_0$,
\[
W_{\rho_N}(\mu R_N, \mu Q_{\eta}^\otimes N) \leq \mathbb{E}(\rho_N(Z_1, Z_N)) \leq \gamma N C_1 \mathbb{E}(W_{\rho}(\pi(Z_0), \eta)). \quad \hfill \qed
\]

**Proposition 7.** There exist $C_2, C_3 > 0$ such that for all $N \in \mathbb{N}$, $\gamma \in (0, \gamma_0]$, $m \in \mathbb{N}$ and $\eta_0 \in \mathcal{P}(\mathbb{T}^d)$, first,
\[
W_{\rho_N}(\eta_0^\otimes N R^m, \eta_m^\otimes N) \leq C_2 N \alpha(N) \gamma \sum_{s=1}^m (1 - \gamma s)^{s-1}, \quad \hfill (9)
\]
and second, if \((X_k)_{k \in \mathbb{N}}\) is a Markov chain with initial distribution \(\eta_0^\otimes N\) and transition kernel \(R\), then
\[
\mathbb{E}(W_{\rho}(\pi(X_m), \eta_m)) \leq C_3 \alpha(N) \left(1 + \gamma \sum_{s=1}^{m}(1 - \gamma \kappa)^{s-1}\right).
\] (10)

Remark that when \(\kappa > 0\), \(\gamma \sum_{s=1}^{m}(1 - \gamma \kappa)^{s-1} \leq 1/\kappa\) so that (9) and (10) yield uniform in time estimates. On the contrary, when \(k < 0\), the estimates are exponentially bad in \(t = m\gamma\).

Proof. We start with the proof of (9), for \(m \geq 1\) (the case \(m = 0\) being trivial). From the triangular inequality, Proposition 5 and Lemma 6,
\[
r_m := W_{\rho_N}(\eta_0^\otimes N R^m, \eta_m^\otimes N) \leq W_{\rho_N}(\eta_0^\otimes N R^m, \eta_{m-1}^\otimes N R) + W_{\rho_N}(\eta_{m-1}^\otimes N R, \eta_{m-1}^\otimes N Q_{\eta_{m-1}}) \leq (1 - \kappa \gamma) r_{m-1} + \gamma N C_1 \int_{T^d N} W_{\rho}(\pi(x), \eta_{m-1}) \eta_{m-1}^\otimes N(dx).
\]
Since \(W_{\rho} \leq W_1\), estimating the last term is a classical question, that is to bound the expected Wasserstein distance between the empirical measure of a sample of \(N\) independent and identically distributed random variables and their common law. From [6, Theorem 1] (and since on the torus the moments of probability measures are uniformly bounded), there exists some \(C' > 0\) independent from \(\eta_0\), \(m\), \(N\) and \(\gamma\) such that
\[
\int_{T^d N} W_1(\pi(x), \eta_{m-1}) \eta_{m-1}^\otimes N(dx) \leq C' \alpha(N).
\]
Since \(r_0 = 0\), a direct induction concludes the proof of (9).

To prove (10), let \((X, Y)\) be an optimal coupling of \(\eta_0^\otimes N R^m\) and \(\eta_m^\otimes N\). Considering \(J \sim \mathcal{U}([1, N])\) independent from \((X, Y)\) then, conditionally to \((X, Y)\), \((X_J, Y_J)\) is a coupling of \(\pi(X)\) and \(\pi(Y)\), so that
\[
W_{\rho}(\pi(X), \pi(Y)) \leq \mathbb{E}(\rho(X_J, Y_J) \mid (X, Y)) = \frac{1}{N} \rho_N(X, Y).
\]
Taking the expectation in
\[
W_{\rho}(\pi(X), \eta_k) \leq W_{\rho}(\pi(X), \pi(Y)) + W_{\rho}(\pi(Y), \eta_k),
\]
we conclude with (9) and [6, Theorem 1] again.

**Corollary 8.** With the notations of Proposition 7 for all \(k \in [1, N]\),
\[
W_{\rho_k}(\text{Law}(X_{1,m}, \ldots, X_{k,m}), \eta_{m}^\otimes k) \leq C_2 k \alpha(N) \gamma \sum_{s=1}^{m}(1 - \gamma \kappa)^{s-1}.
\]

Proof. Let \((X, Y)\) be an optimal coupling of \(\eta_0^\otimes N R^m\) and \(\eta_m^\otimes N\), and let \(\sigma\) be uniformly distributed over the set of permutations of \(N\) elements, independent from \((X, Y)\). Since the laws of \(X\) and \(Y\) are exchangeable, \(X_{\sigma} = (X_{\sigma(1)}, \ldots, X_{\sigma(N)})\) has the same law as \(X\), in particular \((X_{\sigma(1)}, \ldots, X_{\sigma(k)})\) has the same law as \((X_1, \ldots, X_k)\). The same goes for \(Y_{\sigma}\), and
\[
\mathbb{E}\left(\sum_{i=1}^{k} \rho(X_{\sigma(i)}, Y_{\sigma(i)})\right) = k \mathbb{E}(\rho(X_{\sigma(1)}, Y_{\sigma(1)})) = \frac{k}{N} \mathbb{E}(\rho_N(X, Y)) = \frac{k}{N} W_{\rho_N}(\eta_0^\otimes N R^m, \eta_m^\otimes N),
\]
and Proposition 5 concludes.

Corollary 8 means that, for any fixed \(k \in \mathbb{N}_0\), as \(N\) goes to infinity, the \(k\)-marginals of the system of particles converge toward the law of \(k\) independent non-linear chains, which is the so-called propagation of chaos phenomenon.
2.4 Discrete to continuous time

We start by defining \((\bar{Y}_t)_{t \geq 0}\) and \((\bar{X}_t)_{t \geq 0}\) the continuous-time analogous of the chains \((Y_k)_{k \in \mathbb{N}}\) on \(\mathbb{T}^d\) and \((X_k)_{k \in \mathbb{N}}\) on \(\mathbb{T}^{dN}\) defined in Section 1.2. We start with the non-linear process. For \(t \geq 0\), let

\[
\eta_t = \text{Law}(Z_t \mid T > t)
\]

where \(Z\) solves (11) with initial distribution \(\eta_0\) and \(T\) is given by (2). We define \((\bar{Y}_t)_{t \geq 0}\) as follows. Set \(\bar{Y}_0 = Z_0 \sim \eta_0, T_0 = 0\) and suppose that \(T_n\) and \((\bar{Y}_t)_{t \in [0,T_n]}\) have been defined for some \(n \in \mathbb{N}\). Let \((B_t)_{t \geq 0}\) be a new Brownian motion on \(\mathbb{T}^d\) and \(E \sim \mathcal{E}(1)\), independent one from the other. Let \(\bar{Y}\) be the solution of

\[
d\bar{Y}_t = b(\bar{Y}_t)dt + dB_t
\]

for \(t \geq T_n\) with \(\bar{Y}_{T_n} = \bar{Y}_{T_n}\) and let

\[
T_{n+1} = \inf \left\{ t > T_n, \ E \leq \int_{T_n}^t \lambda(\bar{Y}_s)ds \right\}.
\]

For \(t \in (T_n, T_{n+1})\), set \(\bar{Y}_t = \bar{Y}_{T_n}\). Finally, draw a new \(\bar{Y}_{T_{n+1}}\) according to \(\eta_{T_{n+1}}\). By induction \(T_n\) and \((\bar{Y}_t)_{t \in [0,T_n]}\) are then defined for all \(n \in \mathbb{N}\). Since \(\lambda\) is bounded, \(T_n\) almost surely goes to infinity when \(n \to \infty\) so that \((\bar{Y}_t)_{t \geq 0}\) is defined for all \(t \geq 0\). Similarly to Proposition 1, it can be established that \(\text{Law}(\bar{Y}_t) = \eta_t\) for all \(t \geq 0\).

Now, as in Section 1.2 from the non-linear process \((\bar{Y}_t)_{t \geq 0}\), the interacting particles \((\bar{X}_t)_{t \geq 0}\) are obtained by replacing \(\eta_t\) by the empirical distribution of the system when particles die and are resurrected.

More precisely, let \((E_{i,k}, B_{i,k}, J_{i,k})_{i \in [1,N], k \in \mathbb{N}}\) be a family of independent triplet of independent random variables where, for all \(i \in [1,N]\) and \(k \in \mathbb{N}\), \(E_{i,k} \sim \mathcal{E}(1), J_{i,k} \sim U([1,N])\) (except if \(k = 0\), in which case \(J_{i,k} = i\) almost surely) and \(B_{i,k} = (B_{i,k,t})_{t \geq 0}\) is a \(d\)-dimensional Brownian motion. From these variables, we simultaneously define by induction the process and its death times \((T_{i,k})_{i \in [1,N], k \in \mathbb{N}}\) as follows. First, set \(\bar{X}_0 = x\) and \(T_{i,0} = 0\) for all \(i \in [1,N]\). For all \(i \in [1,N]\), set \(\hat{X}_{i,0,0} = x_i\) and for \(k \geq 1\), set

\[
\hat{X}_{i,k,T_{i,k}} = \lim_{t \to T_{i,k}} \bar{X}_{J_{i,k},t}.
\]

(11)

For all \(k \in \mathbb{N}\), for \(t \geq T_{i,k}\), let \(\hat{X}_{i,k}\) solve

\[
d\hat{X}_{i,k,t} = b(\hat{X}_{i,k,t})dt + dB_{i,k,t},
\]

set

\[
T_{i,k+1} = T_{i,k} + \inf \left\{ t \geq 0, \ E_{i,k} \leq \int_0^t \lambda(\hat{X}_{i,k,s})ds \right\}
\]

and for all \(t \in [T_{i,k}, T_{i,k+1})\), set \(\bar{X}_t = \hat{X}_{i,k,t}\).

Then \(\bar{X}_t = (\bar{X}_{1,t}, \ldots, \bar{X}_{N,t})\) is well-defined for all \(t \geq 0\). Indeed, it is well defined for all \(t < S_1 := \min\{T_{i,1}, i \in [1,N]\}\) the first death time of some particle, and is equal on this interval to \((\hat{X}_{1,0,t}, \ldots, \hat{X}_{N,0,t})\), which is continuous on \([0,S_1]\). Hence, the limits involved in (11) are well defined for \(k = 1\) and all \(i \in [1,N]\) such that \(T_{i,1} = S_1\). Then the algorithm above similarly defines the process up to the second time some particles die, etc.

Remark that most of the times (11) simply reads \(\hat{X}_{i,k,T_{i,k}} = \bar{X}_{J_{i,k},T_{i,k}}\) (at its \(k\)th death time, the \(i\)th particle is resurrected at the current position of the \(J_{i,k}\)th particle). Indeed, the only case when this is not true is when the \(J_{i,k}\)th particle dies at time \(T_{i,k}\). Since the probability that two or more particles die simultaneously is zero, this almost surely only occurs if \(J_{i,k} = i\), i.e. if the particle is resurrected at its own position.
Denote \((P_t)_{t \geq 0}\) the Markov semi-group associated with \((X_t)_{t \geq 0}\), i.e. for all \(t \geq 0\), \(P_t\) is the Markov kernel given by

\[
P_t f(x) = \mathbb{E}(f(X_t) \mid X_0 = x)
\]

We sometimes write \(P_t = P_{N,t}\) to specify the number of particles.

**Lemma 9.** There exist \(C_4 > 0\) such that for all \(N \in \mathbb{N}\), \(\gamma \in (0, \gamma_0]\) and \(\mu \in \mathcal{P}(\mathbb{T}^d N)\),

\[
W_{\rho_N} \left(\mu R_{N,\gamma}, \mu P_{N,\gamma}\right) \leq NC_4 \gamma^{3/2}.
\]

**Proof.** As in the proof of Lemma 9, it is sufficient to treat the case \(\mu = \delta_x\) with a fixed \(x \in \mathbb{T}^d N\). Let \((X_t)_{t \geq 0}\) be defined as above from random variables \((E_{i,k}, B_{i,k}, J_{i,k})_{i \in [1,N], k \in \mathbb{N}}\).

In particular, \(X_\gamma \sim \delta_x P_{\gamma}\).

To define \(X_1 \sim \delta_x R\), for all \(i \in [1,N]\) and \(k \in \mathbb{N}\), consider \((\tilde{X}_{i,k} t)_{t \geq 0}\) the solution to \(\tilde{X}_{i,k,0} = x, J_{i,k}\) and

\[
d\tilde{X}_{i,k,t} = b\left(\tilde{X}_{i,k,0}\right) dt + dB_{i,k,t}.
\]

Denoting

\[
H_i = \inf \left\{ k \in \mathbb{N} \mid E_{i,k} \geq \gamma \lambda \left(\tilde{X}_{i,k,\gamma}\right) \right\},
\]

set \(X_1 := (\tilde{X}_{1,H_{1,\gamma}}, \ldots, \tilde{X}_{N,H_{N,\gamma}})\).

Then \((X_1, X_\gamma)\) is a coupling of \(R(x, \cdot)\) and \(P_{\gamma}(x, \cdot)\), so that

\[
W_{\rho_N} \left(R(x, \cdot), P_{\gamma}(x, \cdot)\right) \leq \mathbb{E}(\rho_N(X_1, X_\gamma)) = \sum_{i=1}^N \mathbb{E}(\rho(X_{i,1}, X_{i,\gamma}))
\]

We now distinguish four cases, considering the events

\[
\begin{align*}
B_{i,1} &= \{H_i = 0 \text{ and } T_{i,1} > \gamma\} \\
B_{i,2} &= \{H_i = 1 \text{ and } T_{i,1} \leq \gamma < T_{i,2} \wedge T_{i,0,1}\} \\
B_{i,3} &= \{H_i = 1 \text{ and } T_{i,1} > \gamma\} \cup \{H_i = 0 \text{ and } T_{i,1} \leq \gamma\} \\
B_{i,4} &= \{H_i \geq 2\} \cup \{T_{i,2} \leq \gamma\} \cup \{T_{i,1} \vee T_{i,0,1} \leq \gamma\},
\end{align*}
\]

that is, respectively: none of the two \(i^{th}\) particles dies; both the \(i^{th}\) particles die exactly once; one particle dies but not the other; at least two deaths are involved for one of the two particle. For all \(i \in [1, N]\), \(\Omega = \bigcup_{j=1}^4 B_{i,j}\), so that

\[
\mathbb{E}(\rho(X_{i,1}, X_{i,\gamma})) \leq \mathbb{E}(\rho(X_{i,1}, X_{i,\gamma})(\mathbb{1}_{B_{i,1}} + \mathbb{1}_{B_{i,2}} + \mathbb{1}_{B_{i,3}} + \mathbb{1}_{B_{i,4}})).
\]

Conclusion follows by gathering the four cases.

**Case 1.** It reduces to the classical case of diffusions, since

\[
\mathbb{E}(|X_{i,1} - X_{i,\gamma}| \mathbb{1}_{B_{i,1}}) = \mathbb{E}(|\tilde{X}_{i,0,\gamma} - \tilde{X}_{i,0,\gamma}| \mathbb{1}_{B_{i,1}}) \leq \mathbb{E}(|\tilde{X}_{i,0,\gamma} - \tilde{X}_{i,0,\gamma}|).
\]

Then

\[
|\tilde{X}_{i,0,t} - \tilde{X}_{i,0,t}| = \left| \int_0^t \left(b(x_i) - b\left(\tilde{X}_{i,0,s}\right)\right) ds \right| \\
\leq \|\nabla b\|_{\infty} \int_0^t \left(|\tilde{X}_{i,0,s} - \tilde{X}_{i,0,s}| + |x_i - \tilde{X}_{i,0,s}|\right) ds
\]

By the Gronwall Lemma, for all \(t \geq 0\), almost surely,

\[
\sup_{s \in [0,t]} |\tilde{X}_{i,0,t} - \tilde{X}_{i,0,t}| \leq \|\nabla b\|_{\infty} e^{ct\|\nabla b\|_{\infty}} \int_0^t |x_i - \tilde{X}_{i,0,s}| ds.
\]  \((12)\)
Since $\tilde{X}_{i,0,s}$ is a Gaussian variable with mean $x_i + sb(x_i)$ and variance $s$,
\begin{equation}
\mathbb{E}\left( |x_i - \tilde{X}_{i,0,s}| \right) \leq sb(x_i) + \mathbb{E}\left( |x_i + sb(x_i) - \tilde{X}_{i,0,s}| \right) \leq \|b\|_\infty s + \sqrt{s}. \tag{13}
\end{equation}

As a consequence, for $\gamma \leq \gamma_0$,
\begin{equation}
\mathbb{E}\left( |\tilde{X}_{i,0,\gamma} - \tilde{X}_{i,0,\gamma}| \right) \leq \|\nabla b\|_\infty e^{\gamma_0 \|\nabla b\|_\infty} \int_0^\gamma \mathbb{E}\left( |x_i - \tilde{X}_{i,0,s}| \right) ds \leq c\gamma^{3/2}. \tag{14}
\end{equation}

Case 2. We bound
\begin{equation}
\mathbb{E}\left( |X_{i,1} - \tilde{X}_{i,\gamma}| \mathbb{1}_{B_{i,2}} \right) \leq \mathbb{E}\left( \left| \tilde{X}_{i,1,\gamma} - x_{J_i,0} \right| + \left| \tilde{X}_{i,1,\gamma} - x_{J_i,0} \right| \mathbb{1}_{B_{i,2}} \right).
\end{equation}

Similarly to \[13\],
\begin{equation}
\mathbb{E}\left( \left| \tilde{X}_{i,1,\gamma} - x_{J_i,0} \mathbb{1}_{B_{i,2}} \right| \right) \leq \mathbb{E}\left( \left| \tilde{X}_{i,1,\gamma} - x_{J_i,0} \mathbb{1}_{E_{i,0} \leq \gamma \|\lambda\|_\infty} \right| \right) \leq c\gamma^{3/2},
\end{equation}
where we used the independence of $E_{i,0}$ from $J_{i,1}$ and $(\tilde{X}_{i,1,t})_{t \geq 0}$. Denote $(X'_i)_{t \geq 0}$ the solution of
\begin{equation}
dX'_t = b(X'_t) dt + \begin{cases}
\mathbb{d}B_{i,0,0,t} & \text{for } t < T_{i,0} \\
\mathbb{d}B_{i,1,t} & \text{for } t \geq T_{i,0}
\end{cases}
\end{equation}
with $X'_{i,0} = x_{J_i,0}$. Under the event $B_{i,2}$, $\tilde{X}_{i,1,\gamma} = X'_i \gamma$. Moreover, $J_{i,0}$, $B_{i,0,0}$ and $B_{i,1}$ are independent from $T_{i,0}$ and thus, by the strong Markov property, $(X'_i)_{t \geq 0}$ is independent from $T_{i,0}$ and conditionally to $J_{i,0}$ it has the same distribution as $\tilde{X}_{i,0,0,0}$ (namely it is a diffusion solving \[1\] with initial condition $x_{J_i,0}$). Hence,
\begin{equation}
\mathbb{E}\left( \left| \tilde{X}_{i,1,\gamma} - x_{J_i,0} \mathbb{1}_{B_{i,2}} \right| \right) \leq \mathbb{E}\left( \left| X'_i - y_{J_i,0} \mathbb{1}_{E_{i,0} \leq \gamma \|\lambda\|_\infty} \right| \right) \leq c'\gamma^{3/2}.
\end{equation}

Case 3. We bound
\begin{equation}
\mathbb{E}\left( \rho(X_{i,1}, \tilde{X}_{i,\gamma}) \mathbb{1}_{B_{i,3}} \right) \leq \frac{1}{d} \mathbb{P}(B_{i,3})
\end{equation}
\begin{equation}
\leq \frac{1}{d} \mathbb{P}\left( \int_0^\gamma \lambda(\tilde{X}_{i,0,s}) ds \land \left( \gamma \lambda(\tilde{X}_{i,0,\gamma}) \right) \leq E_{i,0} \leq \int_0^\gamma \lambda(\tilde{X}_{i,0,s}) ds \lor \left( \gamma \lambda(\tilde{X}_{i,0,\gamma}) \right) \right)
\end{equation}
\begin{equation}
= \frac{1}{d} \mathbb{E}\left( \left| \exp\left( -\int_0^\gamma \lambda(\tilde{X}_{i,0,s}) ds \right) - \exp\left( -\gamma \lambda(\tilde{X}_{i,0,\gamma}) \right) \right| \right)
\end{equation}
\begin{equation}
\leq \frac{1}{d} \mathbb{E}\left( \left| \int_0^\gamma \lambda(\tilde{X}_{i,0,s}) ds - \gamma \lambda(\tilde{X}_{i,0,\gamma}) \right| \right).
\end{equation}

Now,
\begin{equation}
\left| \int_0^\gamma \lambda(\tilde{X}_{i,0,s}) ds - \gamma \lambda(\tilde{X}_{i,0,\gamma}) \right| \leq L_\lambda \left( \int_0^\gamma |\tilde{X}_{i,0,s} - x_i| ds + \gamma |x_i - \tilde{X}_{i,0,\gamma}| \right).
\end{equation}

Using \[12\] together with \[13\] yields
\begin{equation}
\mathbb{E}\left( \rho(X_{i,1}, \tilde{X}_{i,\gamma}) \mathbb{1}_{B_{i,3}} \right) \leq c_3\gamma^{3/2}.
\end{equation}

Case 4. We bound
\begin{equation}
\mathbb{E}\left( \rho(X_{i,1}, \tilde{X}_{i,\gamma}) \mathbb{1}_{B_{i,4}} \right) \leq \frac{1}{d} \mathbb{P}(B_{i,4})
\end{equation}
\begin{equation}
\leq \mathbb{P}(E_{i,0} \lor E_{i,1} \leq \gamma \|\lambda\|_\infty) + \mathbb{P}(E_{i,0} \lor E_{J_i,0} \leq \gamma \|\lambda\|_\infty)
\end{equation}
\begin{equation}
\leq 2 \left( 1 - e^{-\gamma \|\lambda\|_\infty} \right)^2 \leq 2 \gamma^2 \|\lambda\|_\infty^2.
\end{equation}
Proposition 10. There exist $C_5 > 0$ such that for all $N \in \mathbb{N}$, $\gamma \in (0, \gamma_0]$ and $\eta_0 \in \mathcal{P}(\mathbb{T}^d)$,

$$W_{\rho_N} (\mu R_{N,\gamma}^m, \mu P_{N,m\gamma}) \leq \sqrt{\gamma NC_5\gamma} \sum_{s=1}^{m} (1 - \gamma \kappa)^{s-1}.$$ 

As for Proposition 7, when $\kappa > 0$, $\gamma \sum_{s=1}^{m} (1 - \gamma \kappa)^{s-1} \leq 1/\kappa$ so that [9] and [10] yield uniform in time estimates. On the contrary, when $k < 0$, the estimates are exponentially bad in $t = m\gamma$.

Proof. The proof is similar to Proposition 7. Denoting $r_m = \mu R_m$ and $\nu_m = \mu P_{m\gamma}$, from the triangular inequality, Proposition 5 and Lemma 9,

$$r_m := W_{\rho_N} (\mu_m, \nu_m) \leq W_{\rho_N} (\mu_m, \nu_{m-1} R) + W_{\rho_N} (\nu_{m-1} R, \nu_{m-1} P_s) \leq (1 - \gamma \kappa)r_{m-1} + NC_4 \gamma^{3/2},$$

and an induction concludes. \qed

2.5 Conclusion

In this section we use the notations of the previous ones, in particular $\kappa$ is given by [9] and the constants $C_2$, $C_3$ and $C_5$ are those of Propositions 7 and 10. We can now gather all these previous results.

We start with the following simple preliminary result.

Lemma 11. For all $N \in \mathbb{N}$ and $\mu, \nu \in \mathcal{P}(\mathbb{T}^d)$,

$$W_{\rho_N} (\mu \otimes N, \nu \otimes N) = NW_{\rho_N} (\mu, \nu).$$

Proof. By considering $N$ independent couplings $(X_i, Y_i)_{i \in [1, N]},$

$$W_{\rho_N} (\mu \otimes N, \nu \otimes N) \leq \mathbb{E} (\rho_N (X, Y)) = \sum_{i=1}^{N} \mathbb{E} (\rho (X_i, Y_i)) = NW_{\rho_N} (\mu, \nu).$$

Conversely, if $(X, Y)$ is an optimal coupling of $\mu \otimes N$ and $\nu \otimes N$, then

$$W_{\rho} (\mu, \nu) \leq \mathbb{E} (\rho (X_1, Y_1)) = \frac{1}{N} \mathbb{E} (\rho_N (X, Y)) = \frac{1}{N} W_{\rho_N} (\mu \otimes N, \nu \otimes N).$$ \qed

Remark that the second part of the proof also applies for $\mu, \nu \in \mathcal{P}(\mathbb{T}^{dN})$ that are exchangeable (i.e. invariant by any permutation of the $d$-dimensional coordinates), in which case, denoting, $\mu^{(1)}$ and $\nu^{(1)}$ their $d$-dimensional marginals, we get that

$$W_{\rho} (\mu^{(1)}, \nu^{(1)}) \leq \frac{1}{N} W_{\rho_N} (\mu, \nu).$$

Letting either $\gamma$ vanish or $N$ go to infinity in Proposition 5 we obtain long-time convergence for, respectively, the non-homogeneous self-interacting Markov chain $(Y_k)_{k \in \mathbb{N}}$ introduced in Section 1.2 and the continuous-time Markov chain $(\overline{X}_t)_{t \geq 0}$ defined in Section 2.4.

Corollary 12. Let $(\eta_n)_{n \in \mathbb{N}}$ be such as defined in Section 1.2 and $(\tilde{\eta}_n)_{n \in \mathbb{N}}$ be similarly defined but with a different initial distribution $\tilde{\eta}_0 \in \mathcal{P}(\mathbb{T}^d)$. For all $m \in \mathbb{N}$ and all $\gamma \in (0, \gamma_0],$

$$W_{\rho} (\eta_m, \tilde{\eta}_m) \leq (1 - \gamma \kappa)^m W_{\rho} (\eta_0, \tilde{\eta}_0).$$
Corollary 13. For all $N \in \mathbb{N}^*$, $t \geq 0$ and $\mu, \nu \in \mathcal{P}(\mathbb{T}^{dN})$,
\[
W_{\rho_N} (\mu P_{N,t}, \nu P_{N,t}) \leq e^{-\kappa t} W_{\rho_N} (\mu, \nu).
\]

Proofs of Corollaries 12 and 13. By the triangular inequality,
\[
W_{\rho} (\eta_m^{\otimes N}, \bar{\eta}_m^{\otimes N}) \leq W_{\rho} (\eta_m^{\otimes N}, \tilde{\eta}_0^{\otimes N} R^m) + W_{\rho} (\tilde{\eta}_0^{\otimes N} R^m, \bar{\eta}_m^{\otimes N}) + W_{\rho} (\tilde{\eta}_0^{\otimes N} R^m, \eta_m^{\otimes N})
\]
\[
\leq (1 - \gamma \kappa)^m W_{\rho} (\eta_0^{\otimes N}, \eta_0^{\otimes N}) + 2C_2 N a(N) \gamma \sum_{s=1}^{m} (1 - \gamma \kappa)^s - 1,
\]
where we applied Propositions 5 and 7. Using Lemma 11, dividing by $N$ and letting $N$ go to infinity concludes the proof of Corollary 12.

Similarly, Corollary 13 is a direct consequence of Propositions 5 and 10, letting $m$ go to infinity at a fixed $t$ and $N$ in
\[
W_{\rho_N} (\mu P_{N,t}, \nu P_{N,t}) \leq W_{\rho_N} (\mu P_{N,t}, \mu R_{N,t/m}^m) + W_{\rho_N} (\mu R_{N,t/m}^m, \nu R_{N,t/m}^m) + W_{\rho_N} (\nu R_{N,t/m}^m, \nu P_{N,t}).
\]

We now turn to the continuous-time limit of the non-linear chain $(Y_k)_{k \in \mathbb{N}}$.

Lemma 14. There exists $C_6 > 0$ such that for all $\eta_0 \in \mathcal{P}(\mathbb{T}^d)$ and all $\gamma \in (0, \gamma_0]$, if $(\eta_n)_{n \in \mathbb{N}}$ is such as defined in Section 2.2 and $(\eta_t)_{t \geq 0}$ is such as defined in Section 2.3 (with $\overline{\eta}_0 = \eta_0$), then
\[
W_1 (\eta_1, \overline{\eta}_\gamma) \leq C_6 \gamma^{3/2}.
\]

Proof. We could follow the proof of Lemma 9 but, using the notations of the introduction, we will rather use the fact that
\[
\eta_1 = \text{Law} \left( Z_t \mid \tilde{T} > \gamma \right), \quad \overline{\eta}_\gamma = \text{Law} \left( Z_t \mid T > \gamma \right),
\]
where the gaussian variable $G_0$ in 14 is equal to $B_\gamma / \sqrt{T}$ where $(B_t)_{t \geq 0}$ is the Brownian motion involved in 11, and $T$ and $\tilde{T}$ are defined with the same $E \sim \mathcal{E}(1)$. Recall the estimate 14 for the error from an Euler scheme to its initial diffusion. Then we bound
\[
\mathbb{E} \left( |Z_1 - Z_\gamma| \mid T > \gamma, \tilde{T} > \gamma \right) \leq \mathbb{E} \left( P \left( T > \gamma, \tilde{T} > \gamma \right) \right)^{-1} \mathbb{E} \left( |Z_1 - Z_\gamma| \right)
\]
\[
\leq \left( 1 - e^{-\gamma \kappa \lambda \| \lambda \|_{\infty}} \right)^{-1} c_\gamma^{3/2},
\]
which concludes.

Corollary 15. Keeping the notations of Lemma 14 for all $\gamma \in (0, \gamma_0]$, $m \in \mathbb{N}$ and $\eta_0 \in \mathcal{P}(\mathbb{T}^{dN})$,
\[
W_{\rho} (\eta_m, \overline{\eta}_m) \leq \sqrt{T} C_6 \gamma \sum_{s=1}^{m} (1 - \gamma \kappa)^s - 1.
\]

Proof. Denoting $r_m = W_{\rho} (\eta_m, \overline{\eta}_m)$, we bound
\[
r_m \leq W_{\rho} (\eta_m, \overline{\eta}_m - Q_{\eta_m}) + W_{\rho} (\overline{\eta}_m - Q_{\eta_m}, \overline{\eta}_m) + W_{\rho} (\overline{\eta}_m - Q_{\eta_m}, \eta_m)
\]
\[
\leq (1 - \gamma \kappa) r_{m-1} + C_6 \gamma^{3/2},
\]
where we used Lemma 14 and Corollary 9. An induction concludes.
We can now prove propagation of chaos results for the continuous-time process:

**Corollary 16.** For all $N \in \mathbb{N}$, $k \in [1, N]$ and all $t \geq 0$, if $(\bar{X}_t)_{t \geq 0}$ is a Markov process with initial distribution $\eta_0^{\otimes N}$ associated to the semigroup $(P_{N,t})_{t \geq 0}$ then, first,

$$\mathcal{W}_{p_k} \left( \text{Law}(\bar{X}_{1,t}, \ldots, \bar{X}_{k,t}, \eta_t^{\otimes k}) \right) \leq C_2k\alpha(N) \int_0^t e^{-\kappa s} ds,$$

and second,

$$\mathbb{E} \left( \mathcal{W}_p (\pi(\bar{X}_t), \eta_t) \right) \leq C_3\alpha(N) \left( 1 + \int_0^t e^{-\kappa s} ds \right).$$

**Proof.** As shown in the proof of Proposition 7 if $(X, Y)$ is an optimal coupling of $\mu$ and $\nu$, 

$$\mathbb{E} \left( \mathcal{W}_p (\pi(X), \pi(Y)) \right) \leq \frac{1}{N} \mathcal{W}_{p_N}(\mu, \nu).$$

Thus, considering a time step $\gamma = t/m$, $m \in \mathbb{N}$, we decompose

$$\mathcal{W}_p (\pi(\bar{X}_t), \eta_t) \leq \mathcal{W}_p (\pi(\bar{X}_t), \pi(X_m)) + \mathcal{W}_p (\pi(X_m), \eta_m) + \mathcal{W}_p (\eta_m, \eta_t),$$

take the expectation, apply Propositions 7 and 10 and Corollary 15 and let $m$ go to infinity. This proves the second point, and the proof of the first one is similar, with Corollary 8.

Up to now, we have sent either $N$ or $\gamma$ to their limit. When $\kappa > 0$, if we let $t = m\gamma$ go to infinity at fixed $N$ and $\gamma$, we recover results on the equilibria of the processes. Indeed, note that Corollary 12 together with the Banach fixed-point theorem imply that $n \rightarrow \eta_n$ admits a limit which is independent from $\eta_0$. Together with Proposition 11 this is the unique QSD of the Markov chain 13. Denote it $\nu_\gamma$. Similarly, Proposition 5 implies that $R_{N,\gamma}$ admits a unique invariant measure. Denote it $\mu_{\infty,N,\gamma}$ and $\mu_{\infty,N,\gamma}^{(k)}$ its first $kd$-dimensional marginal for $k \in [1, N]$ (i.e. the law of $(X_1, \ldots, X_k)$ if $X \sim \mu_{\infty,N,\gamma}$). Third, Corollary 15 implies that $(P_{N,t})_{t \geq 0}$ admits a unique invariant measure $\overline{\mu}_{\infty,N}$.

**Corollary 17.** If $\kappa > 0$, then for all $N \in \mathbb{N}$ and $\gamma \in (0, \gamma_0]$

$$\mathcal{W}_{p_N} \left( \mu_{\infty,N,\gamma}, \overline{\mu}_{\infty,N} \right) \leq \sqrt{\gamma} N^{-1} \kappa C_5,$$

**Corollary 18.** If $\kappa > 0$, then for all $N \in \mathbb{N}$, $k \in [1, N]$ and $\gamma \in (0, \gamma_0]$, first,

$$\mathcal{W}_{p_k} \left( \mu_{\infty,N,\gamma}^{(k)}, \nu_{\gamma}^{(k)} \right) \leq \kappa^{-1} C_2 k\alpha(N),$$

and second,

$$\mathbb{E}_{\mu_{\infty,N,\gamma}} \left( \mathcal{W}_p (\pi(X), \nu_{\gamma}) \right) \leq \kappa^{-1} C_3 \alpha(N).$$

**Proofs of Corollaries 17 and 18.** Considering any $\eta_0 \in \mathcal{P}(\mathbb{T}^d)$ and $m \in \mathbb{N}$,

$$\mathcal{W}_{p_N} \left( \mu_{\infty,N,\gamma}, \overline{\mu}_{\infty,N} \right) \leq \mathcal{W}_{p_N} \left( \mu_{\infty,N,\gamma}, \eta_0^{\otimes N} P^m \right) + \mathcal{W}_{p_N} \left( \eta_0^{\otimes N} P^m, \eta_0^{\otimes N} P_{\gamma m} \right) + \mathcal{W}_{p_N} \left( \eta_0^{\otimes N} P_{\gamma m}, \overline{\mu}_{\infty,N} \right).$$

Apply Proposition 5 with $\mu = \mu_{\infty,N,\gamma}$ and $\nu = \eta_0^{\otimes N}$, Corollary 15 with the same $\nu$ and with $\mu = \overline{\mu}_{\infty,N}$, and Proposition 10. Letting $m$ go to infinity concludes the proof of Corollary 17. The proof of Corollary 18 is similar (based on Proposition 7 and Corollary 8).
Next, we can send two parameters to their limit. Sending $N$ to infinity and $\gamma$ to zero, we get the long time convergence of the non-linear process $(Y_t)_{t \geq 0}$ introduced in Section 2.4 (or, equivalently, of the process $Z$ solving \eqref{eq:process} conditioned not to be dead):

\begin{corollary}
Let $(\eta_t)_{t \geq 0}$ be such as defined in Section 2.4 and $(\hat{\eta}_t)_{t \geq 0}$ be similarly defined but with a different initial distribution $\hat{\eta}_0 \in \mathcal{P}(\mathbb{T}^d)$. For all $t \geq 0$,
\[ W_\rho(\eta_t, \hat{\eta}_t) \leq e^{-\kappa t} W_\rho(\eta_0, \hat{\eta}_0). \]
\end{corollary}

\begin{proof}
Thanks to Corollary 15, let $\gamma = t/m$ vanish in Corollary 12.
\end{proof}

In particular, if $\hat{\eta}_0$ is the QSD $\nu_\star$, by definition, $\hat{\eta}_t = \nu_\star$ for all $t \geq 0$, so that Corollary 19 yields the uniqueness of the QSD and the exponential convergence of $\text{Law}(Z_t \mid T > t)$ toward $\nu_\star$ (which is a result in the spirit of \cite{benaim2015}).

Now, at a fixed $\gamma > 0$, letting $t$ and $N$ go to infinity, we obtain an error bound between the QSD $\nu_\gamma$ of the continuous process \eqref{eq:process} and the QSD $\nu_\gamma$ of the discrete scheme.

\begin{corollary}
If $\kappa > 0$, then for all $\gamma \in (0, \gamma_0]$
\[ W_\rho(\nu_\gamma, \nu_\star) \leq \sqrt{\gamma} \kappa^{-1} C_6, \]
\end{corollary}

\begin{proof}
Thanks to Corollaries 12 and 19 (applied with one of the initial condition being the equilibrium), let $m$ go to infinity in Corollary 13.
\end{proof}

Finally, letting $\gamma$ vanish and $t$ go to infinity at a fixed $N \in \mathbb{N}$, we obtain a propagation of chaos result at stationarity (as established in \cite{benaim2015} in the case of a finite state space) for the continuous time system of interacting particle $(\mathbf{X}_t)_{t \geq 0}$ introduced in Section 2.4.

\begin{corollary}
If $\kappa > 0$ and if $\mathbf{X}$ is a random variable with law $\pi_{\infty,N}$, then for all $N \in \mathbb{N}$ and $k \in [1, N]$,
\[ W_{\rho_N}(\mathcal{L}(X_1, \ldots, X_k), \nu_\star^{\otimes k}) \leq \kappa^{-1} C_2 \alpha(N), \]
and second,
\[ E(W_\rho(\pi(\mathbf{X}), \nu_\star)) \leq C_3 \alpha(N) (1 + \kappa^{-1}). \]
\end{corollary}

\begin{proof}
The proof is similar to Corollary 16, letting $t$ go to infinity in Corollary 16, thanks to Corollaries 13 and 19.
\end{proof}

Finally, we detail the proof of our main result.

\begin{proof}[Proof of Theorem 2] Let $c_0 = c_1/c_2$ so that, under the condition \eqref{eq:condition}, $\kappa > 0$. For $\eta_0 \in \mathcal{P}(\mathbb{T}^d)$, let $(\mathbf{X}, \mathbf{Y})$ be an optimal coupling of $\mu_0 R^{\lfloor t/\gamma \rfloor}$ and $\eta_0^{\otimes N} R^{\lfloor t/\gamma \rfloor}$. As in the proof of Proposition 7,
\[ E(W_\rho(\pi(\mathbf{X}), \pi(\mathbf{Y}))) \leq \frac{1}{N} W_{\rho_N}(\mu R^{\lfloor t/\gamma \rfloor}, \eta_0^{\otimes N} R^{\lfloor t/\gamma \rfloor}) \leq a e^{-\kappa(t - \gamma_0)}, \]
where we used Proposition 5 and the fact that $\rho_N(x, y) \leq Na$ for all $x, y \in \mathbb{T}^{dN}$. Then, by the triangular inequality,
\[ W_\rho(\pi(\mathbf{Y}), \nu_\star) \leq W_\rho(\pi(\mathbf{Y}), \eta_{\lfloor t/\gamma \rfloor}) + W_\rho(\eta_{\lfloor t/\gamma \rfloor}, \nu_\gamma) + W_\rho(\nu_\gamma, \nu_\star). \]
Taking the expectation, applying Proposition 7 and Corollaries 12 (applied with $\hat{\eta}_0 = \nu_\gamma$) and 20, the boundedness of $\rho$ and the equivalence of $W_\rho$ and $W_1$ concludes.
\end{proof}
Acknowledgements

Pierre Monmarché acknowledges partial support by the projects EFI ANR-17-CE40-0030 and METANOLIN of the French National Research Agency.

References

[1] M. Benaïm, N. Champagnat, and D. Villemonais. Stochastic approximation of quasi-stationary distributions for diffusion processes in a bounded domain. arXiv e-prints, page arXiv:1904.08620, Apr 2019.

[2] N. Champagnat, K. Coulibaly-Pasquier, and D. Villemonais. Criteria for exponential convergence to quasi-stationary distributions and applications to multi-dimensional diffusions. arXiv e-prints, page arXiv:1603.07909, Mar 2016.

[3] N. Champagnat and D. Villemonais. Convergence of the Fleming-Viot process toward the minimal quasi-stationary distribution. arXiv e-prints, page arXiv:1810.06849, Oct 2018.

[4] B. Cloez and M.-N. Thai. Quantitative results for the Fleming-Viot particle system and quasi-stationary distributions in discrete space. Stochastic Process. Appl., 126(3):680–702, 2016.

[5] B. Delyon, F. Cérou, A. Guyader, and M. Rousset. A Central Limit Theorem for Fleming-Viot Particle Systems with Hard Killing. To appear in Annales de l’IHP (Probability and Statistics), page arXiv:1709.06771, Sep 2017.

[6] N. Fournier and A. Guillin. On the rate of convergence in Wasserstein distance of the empirical measure. Probab. Theory Related Fields, 162(3-4):707–738, 2015.

[7] A. Joulin and Y. Ollivier. Curvature, concentration and error estimates for Markov chain Monte Carlo. Ann. Probab., 38(6):2418–2442, 2010.

[8] C. Le Bris, T. Lelièvre, M. Luskin, and D. Perez. A mathematical formalization of the parallel replica dynamics. Monte Carlo Methods Appl., 18(2):119–146, 2012.

[9] T. Lelièvre, L. Pilaud-Vivien, and J. Reygner. Central limit theorem for stationary Fleming-Viot particle systems in finite spaces. ALEA Lat. Am. J. Probab. Math. Stat., 15(2):1163–1182, 2018.

[10] M. B. Majka, A. Mijatović, and L. Szpruch. Non-asymptotic bounds for sampling algorithms without log-concavity. arXiv e-prints, page arXiv:1808.07105, Aug 2018.

[11] P. Monmarché. Elementary coupling approach for non-linear perturbation of Markov processes with mean-field jump mechanisms and related problems. ArXiv e-prints, September 2018.

[12] C. Villani. Optimal transport, volume 338 of Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]. Springer-Verlag, Berlin, 2009. Old and new.