Quantitative concentration inequalities on sample path space for mean field interaction

François Bolley

QUANTITATIVE CONCENTRATION INEQUALITIES ON SAMPLE PATH SPACE FOR MEAN FIELD INTERACTION

FRANÇOIS BOLLEY

ABSTRACT. We consider a system of particles experiencing diffusion and mean field interaction, and study its behaviour when the number of particles goes to infinity. We derive non-asymptotic large deviation bounds measuring the concentration of the empirical measure of the paths of the particles around its limit. The method is based on a coupling argument, strong integrability estimates on the paths in Hölder norm, and some general concentration result for the empirical measure of identically distributed independent paths.

CONTENTS

INTRODUCTION

This paper is devoted to the study of the behaviour of some large stochastic particle system. In the models to be considered, the evolution of each particle is governed by a random diffusive term, an exterior force field and a mean field interaction with the other particles. For such models the limit behaviour has been clearly identified and studied in terms of law of large numbers, central limit theorem and large deviations. Here we shall give new quantitative estimates on the convergence in the setting of large deviations.

This follows some works addressing this issue at the level of observables or at the level of the whole system at a given time, that we now summarize. For that purpose let $(X_t^i)_{1\leq i\leq N}$ be the position at time t of the N particles in the phase space $\mathbb{R}^d$ and let μ_t be some probability measure describing the limit behaviour of the system. At the level of Lipschitz observables, F. Malrieu [16] adapted ideas of concentration of measure to obtain bounds like

Key words and phrases. Mean field limits, Sanov theorem, Transportation inequalities.

$$\sup_{[\varphi]_1 \leq 1} \mathbb{P}\left[\Big| \frac{1}{N} \sum_{i=1}^N \varphi(X_t^i) - \int_{\mathbb{R}^d} \varphi \, d\mu_t \Big| > \frac{C}{\sqrt{N}} + \varepsilon \right] \leq 2\, e^{-\lambda N \varepsilon^2}, \qquad N \geq 1, \tag{1}$$

where C and λ are some constants independent of ε and N, and $[\,\cdot\,]_1$ is the Lipschitz seminorm defined by

$$[\varphi]_1 := \sup_{x \neq y} \frac{|\varphi(x) - \varphi(y)|}{|x-y|}\cdot$$

In other words, letting δ_x stand for the Dirac mass at a point $x \in \mathbb{R}^d$, the empirical measure

$$\hat{\mu}_t^N := \frac{1}{N} \sum_{i=1}^N \delta_{X_t^i}$$

of the system, which generates the observables at time t, satisfies the deviation inequality

$$\sup_{[\varphi]_1 \leq 1} \mathbb{P}\left[\Big| \int_{\mathbb{R}^d} \varphi \, d\hat{\mu}_t^N - \int_{\mathbb{R}^d} \varphi \, d\mu_t \Big| > \frac{C}{\sqrt{N}} + \varepsilon \right] \leq 2\, e^{-\lambda N \varepsilon^2}, \qquad N \geq 1.$$

Now one can measure how this empirical measure $\hat{\mu}_t^N$ is close to its limit μ_t in a stronger sense, namely, at the very level of the measures. For this, adapting Sanov's large deviation argument, the authors in [6] got quantitative and non-asymptotic bounds on the deviation of $\hat{\mu}_t^N$ around μ_t for some distance which induces a topology stronger that the narrow topology. By comparison with (1), these bounds can be written as

$$\mathbb{P}\left[\sup_{[\varphi]_1 \leq 1} \Big| \frac{1}{N} \sum_{i=1}^N \varphi(X_t^i) - \int_{\mathbb{R}^d} \varphi \, d\mu_t \Big| > \varepsilon \right] \leq C(\varepsilon)\, e^{-\lambda N \varepsilon^2}, \qquad N \geq 1. \tag{2}$$

In this work we want to go one step further by considering the **trajectories** of the particles. A natural object to consider is the empirical measure of the trajectories $(X_t^i)_{0 \leq t \leq T}$ on some given time interval $[0, T]$, which is defined as

$$\hat{\mu}_{[0,T]}^N := \frac{1}{N} \sum_{i=1}^N \delta_{(X_t^i)_{0 \leq t \leq T}}$$

where $\delta_{(X_t^i)_{0 \leq t \leq T}}$ is the Dirac mass on the path $(X_t^i)_{0 \leq t \leq T}$. This is a random probability measure, no longer on the phase space $\mathbb{R}^d$, but now on the path space, which in our model is the space of $\mathbb{R}^d$-valued continuous functions on $[0, T]$.

Its limit behaviour is given as follows: the limit μ_t of the empirical measure $\hat{\mu}_t^N$ at time t can be seen as the law of the solution at time t to a stochastic differential equation; then the law of the whole process on $[0, T]$ so defined will be the limit of $\hat{\mu}_{[0,T]}^N$. We shall give a precise meaning to this convergence, and some estimates which are the analogue of (2) in the path space; in particular we shall see that they imply (2) by projection at time t.

In the next section we state our main results and give an insight of the proofs, which will be given in more detail in the following sections.

1. Statement of the results

1.1. **Some notation and definitions.** One of the key points in this work is to measure the discrepancy between probability measures: this will be done by means of **Wasserstein distances**, which have revealed convenient in this type of issues and are defined as follows. Let (X, d) be a separable and complete metric space, and p be a real number ≥ 1; the Wasserstein distance of order p between two Borel probability measures μ and ν on X is

$$W_p(\mu, \nu) := \inf_\pi \left(\iint_{X \times X} d(x,y)^p \, d\pi(x,y) \right)^{1/p}$$

where π runs over the set of all joint measures on $X \times X$ with marginals μ and ν. W_p induces a metric on the set of Borel probability measures on X with finite moment $\int_X d(x_0, x)^p \, d\mu(x)$ for some (and thus any) x_0 in X; convergence in this metric is equivalent to narrow convergence (against bounded continuous functions) plus some tightness condition on the moments (see for instance [5], [21] for further details on these distances).

At some point the space (X, d) will be $\mathbb{R}^d$ equipped with the Euclidean distance $|\cdot|$, and in this context W_p will be denoted $W_{p,\tau}$. But (X, d) will mainly be the space $\mathcal{C}([0,T], \mathbb{R}^d)$, also denoted $\mathcal{C}$ if no confusion is possible, of $\mathbb{R}^d$-valued continuous functions on $[0,T]$, equipped with the uniform norm

$$\|f\|_\infty := \sup_{0 \leq t \leq T} |f(t)|;$$

for this space W_p will be denoted $W_{p,[0,T]}$. The Wasserstein distances considered in these two situations are linked in the following way: if, for $0 \leq t \leq T$, π_t is the projection from $\mathcal{C}$ into $\mathbb{R}^d$ defined by $\pi_t(f) = f(t)$, then for any Borel probability measures μ and ν on $\mathcal{C}$, and any $p \geq 1$, the relation

$$W_{p,\tau}(\pi_t \sharp \mu, \pi_t \sharp \nu) \leq W_{p,[0,T]}(\mu, \nu), \qquad 0 \leq t \leq T \tag{3}$$

holds, where $\pi_t \sharp \mu$ is the image measure of μ by π_t.

The distance between two probability measures μ and ν on X can also be expressed in terms of the **relative entropy** of ν with respect to μ (for instance), defined by

$$H(\nu|\mu) = \int_X \frac{d\nu}{d\mu} \ln \frac{d\nu}{d\mu} \, d\mu$$

if ν is absolutely continuous with respect to μ, and $H(\nu|\mu) = +\infty$ otherwise.

Both notions are linked by the family of *transportation* or *Talagrand inequalities*: given $p \geq 1$ and $\lambda > 0$, we say that a probability measure μ on X satisfies the inequality $T_p(\lambda)$ if

$$W_p(\mu, \nu) \leq \sqrt{\frac{2}{\lambda} H(\nu|\mu)}$$

holds true for any measure ν, and μ satisfies T_p if it satisfies $T_p(\lambda)$ for some $\lambda > 0$. By Jensen's inequality, the weakest of all is T_1, which is also the only one for which a simple characterization is known: a measure μ satisfies $T_1(\lambda)$ for some $\lambda > 0$ if and only if it admits a square-exponential moment, in the sense that there exist $a > 0$ and x_0 in X such that $\displaystyle\int_X e^{ad(x_0,x)^2}\, d\mu(x)$ be finite. Numerical relations between such a and λ are given in [7, 11].

1.2. **A general concentration inequality for empirical measures.** The proof of our main theorem on the particle system is based on some general concentration result for the empirical measure of $\mathcal{C}$-valued independent and identically distributed random variables. We state this result separately.

For this purpose, given some Borel probability measure μ on $\mathcal{C}$ and N independent random variables $(X^i)_{1 \leq i \leq N}$ with law μ, we let

$$\hat{\mu}^N := \frac{1}{N} \sum_{i=1}^N \delta_{X^i}$$

denote their empirical measure.

Given some real number $\alpha \in (0,1]$, we let $\mathcal{C}^\alpha := \mathcal{C}^\alpha([0,T], \mathbb{R}^d)$ be the space of functions in $\mathcal{C} := \mathcal{C}([0,T], \mathbb{R}^d)$ which moreover are Hölder of order α, equipped with the Hölder norm

$$\|f\|_\alpha := \sup\left(\|f\|_\infty, [f]_\alpha\right)$$

where

$$[f]_\alpha := \sup_{0 \leq t,s \leq T} \frac{|f(t) - f(s)|}{|t-s|^\alpha}.$$

$\mathcal{C}^\alpha$ is a Borel set of the space $\mathcal{C}$ equipped with the topology induced by the uniform norm, and for Borel measures on $\mathcal{C}$, concentrated on $\mathcal{C}^\alpha$, we have in the above notation:

Theorem 1. *Let $p \in [1,2]$ and let μ be a Borel probability measure on $\mathcal{C}$ satisfying a $T_p(\lambda)$ inequality for some $\lambda > 0$, and such that $\displaystyle\int_{\mathcal{C}} e^{a\|x\|_\alpha^2}\, d\mu(x)$ be finite for some $a > 0$ and $\alpha \in (0,1]$. Then, for any $\alpha' < \alpha$ and $\lambda' < \lambda$, there exists some constant N_0 such that*

$$\mathbb{P}\left[W_{p,[0,T]}(\mu, \hat{\mu}^N) > \varepsilon\right] \leq e^{-\beta_p \frac{\lambda'}{2} N \varepsilon^2} \tag{4}$$

for any $\varepsilon > 0$ and $N \geq N_0\, \varepsilon^{-2} \exp\left(N_0\, \varepsilon^{-1/\alpha'}\right)$, where

$$\beta_p = \begin{cases} 1 & \text{if } \; 1 \leq p < 2 \\ (1 + \sqrt{\lambda/a})^{-2} & \text{if } \quad p = 2. \end{cases}$$

Here the constant N_0 depends on μ only through λ, a, α and $\displaystyle\int_{\mathcal{C}} e^{a\|x\|_\alpha^2}\, d\mu(x)$.

Let us make a few remarks on this result.

First of all, for another formulation of the obtained bound in the case when $p = 1$, we recall Kantorovich-Rubinstein dual expression of the W_1 distance on a general space (X,d):

$$W_1(\mu, \nu) = \sup_{[\varphi]_1 \leq 1} \left\{ \int_X \varphi\, d\mu - \int_X \varphi\, d\nu \right\} \tag{5}$$

where $[\varphi]_1 := \sup_{x \neq y} \frac{|\varphi(x) - \varphi(y)|}{d(x,y)}\cdot$ Then a result by S. Bobkov and F. Götze [4] ensures that a $T_1(\lambda)$ inequality for μ is equivalent to the concentration inequality

$$\sup_{[\varphi]_1 \leq 1} \mathbb{P}\left[\frac{1}{N}\sum_{i=1}^N \varphi(X^i) - \int_X \varphi\, d\mu > \varepsilon\right] \leq e^{-\frac{\lambda}{2}N\varepsilon^2}, \qquad N \geq 1\,.$$

By comparison, the bound given by Theorem 1 implies

$$\mathbb{P}\left[\sup_{[\varphi]_1 \leq 1}\Big(\frac{1}{N}\sum_{i=1}^N \varphi(X^i) - \int_{\mathcal{C}} \varphi\, d\mu\Big) > \varepsilon\right] \leq e^{-\frac{\lambda'}{2}N\varepsilon^2}, \quad \lambda' < \lambda\,, \quad N \text{ large enough}$$

by (5), but a modification of the proof would also lead to

$$\mathbb{P}\left[\sup_{[\varphi]_1 \leq 1}\Big(\frac{1}{N}\sum_{i=1}^N \varphi(X^i) - \int_{\mathcal{C}} \varphi\, d\mu\Big) > \varepsilon\right] \leq C(\varepsilon)\, e^{-\frac{\lambda'}{2}N\varepsilon^2}, \quad \lambda' < \lambda\,, \quad N \geq 1, \tag{6}$$

for some computable large constant $C(\varepsilon)$. In other words we control a much stronger quantity, up to some loss on the constant in the right-hand side, or some condition on the size of the sample.

This result seems reasonable in view of Sanov's theorem (stated in [10] for instance). By applying this theorem to $A := \{\nu; W_{p,[0,T]}(\nu,\mu) \geq \varepsilon\}$, for some given $\varepsilon > 0$, one can hope for an upper bound like

$$\mathbb{P}\left[W_{p,[0,T]}(\mu, \hat{\mu}^N) \geq \varepsilon\right] \leq \exp\Big(-N \inf\big\{H(\nu|\mu); \nu \in A\big\}\Big)$$

for large N. With this bound in hand, since

$$\inf\big\{H(\nu|\mu); \nu \in A\big\} \geq \frac{\lambda}{2}\,\varepsilon^2$$

as μ satisfies $T_p(\lambda)$, one indeed obtains an upper bound like (4), but only in an asymptotic way, whereas Theorem 1 moreover gives an estimate on a sufficient size of the sample for the deviation bound to hold. Sanov's theorem does not actually give such an upper bound here; indeed, on an unbounded space such as $\mathcal{C}$, the closure $\overline{A}$ of A (for the narrow topology) contains μ itself: in particular $\inf\{H(\nu|\mu);\, \nu \in \overline{A}\} = 0$ and Sanov's theorem only gives the trivial upper bound $\mathbb{P}[A] \leq \exp\big(-N\inf\{H(\nu|\mu); \nu \in \overline{A}\}\big) = 1$.

To get a more relevant upper bound we impose some extra integrability assumption that may at first sight seem strong and odd. The reason is that, proceeding as in [6], we first reduce the issue to a compact set of $\mathcal{C}$ that has almost full μ measure: a large ball of $\mathcal{C}^\alpha$ will do by Ascoli's theorem and the integrability assumption on μ. Then on this compact set one can get precise upper bounds by using some techniques based on a covering argument and developed in [14] (see also [10, Exercises 4.4.5 and 6.2.19] and [13]). And actually the assumption is satisfied by the Wiener measure on $\mathcal{C}$ (recall that the Brownian motion paths are almost surely Hölder of order α for any $\alpha < 1/2$) and by extension by the law of the process to be considered.

This integrability assumption again implies the existence of a square-exponential moment for μ on $\mathcal{C}$ (for the uniform norm). Since this is equivalent to some T_1 inequality for μ, the $T_p(\lambda)$ assumption is redondant when $p = 1$ if one does not care of the involved constants.

Finally this result can be seen as an extension of the following similar concentration result given in [6, Theorem 1.1] in the case of measures on $\mathbb{R}^d$: if m satisfies $T_p(\lambda)$, then

$$\mathbb{P}\left[W_{p,\tau}(m, \hat{m}^N) > \varepsilon\right] \leq e^{-\gamma_p \frac{\lambda'}{2} N \varepsilon^2}, \quad \varepsilon > 0\,, \quad N \geq N_0 \max(\varepsilon^{-(d'+2)}, 1). \tag{7}$$

Let indeed m be such a measure on $\mathbb{R}^d$. Then the law μ of a constant process on $[0, T]$ initially distributed according to m satisfies the assumptions of Theorem 1 (one can take any $a < \lambda/2$), and the bound (7) follows by (3) with the constant γ_p obtained in [6]. Note however that the required size of the sample is here much larger for small ε.

Theorem 1 will be proved in Section 2.

1.3. **Interacting particle systems.** We now turn to the study of a system of N stochastic interacting particles which positions X_t^i in the phase space $\mathbb{R}^d$ $(1 \leq i \leq N)$ evolve according to the system of coupled stochastic differential equations

$$dX_t^i = \sqrt{2}\, dB_t^i - \nabla V(X_t^i)dt - \frac{1}{N}\sum_{j=1}^N \nabla W(X_t^i - X_t^j)dt, \qquad 1 \leq i \leq N. \tag{8}$$

Here the B^i's are N standard independent Brownian motions on $\mathbb{R}^d$, V and W are exterior and interaction potentials.

The state of the system at some given time t is given by the observables $\displaystyle\frac{1}{N}\sum_{i=1}^N \varphi(X_t^i)$, and thus can be described by the random probability measure

$$\hat{\mu}_t^N := \frac{1}{N}\sum_{i=1}^N \delta_{X_t^i}$$

on the phase space $\mathbb{R}^d$, called the empirical measure of the system. Under some regularity and growth assumptions on the potentials V and W, and if the particles are initially distributed in a chaotic way, for instance as independent and identically distributed variables, then $\hat{\mu}_t^N$ converges to a solution at time t to the partial differential equation

$$\frac{\partial \mu_t}{\partial t} = \Delta\mu_t + \nabla\cdot\big(\mu_t \nabla(V + W * \mu_t)\big) \tag{9}$$

as the number N of particles goes to infinity. This nonlinear diffusive equation, in which $\Delta, \nabla\cdot$ and ∇ respectively denote the Laplace, divergence and gradient operators in $\mathbb{R}^d$, is a McKean-Vlasov equation and has been used in [3] in the modelling of one-dimensional granular media. The convergence of $\hat{\mu}_t^N$ is strongly linked with the phenomenon of propagation of chaos for the interacting particles, and both issues have been studied by H. Tanaka [19], A.-S. Sznitman [18], S. Méléard [17] or S. Benachour, B. Roynette, D. Talay and P. Vallois [1, 2] for instance. Then quantitative estimates on this convergence have been obtained by F. Malrieu [16] at the level of observables, and later at the very level of the law in [6].

In this work we go one step further and study the limit behaviour of the empirical measure

$$\hat{\mu}^N_{[0,T]} = \frac{1}{N} \sum_{i=1}^{N} \delta_{X^i}.$$

of the trajectories $X^i = (X^i_t)_{0 \leq t \leq T}$ of the particles on some time interval $[0, T]$.

For that purpose, let $Y = (Y_t)_{0 \leq t \leq T}$ be a solution to the stochastic differential equation

$$dY_t = \sqrt{2}\, dB_t - \nabla V(Y_t)\, dt - \nabla W * \nu_t(Y_t)\, dt \tag{10}$$

starting at Y_0 distributed according to the initial condition μ_0 in (9), where ν_t is the law of Y_t at time t. Then, by Itô's formula, ν_t also is a solution to equation (9) with initial datum μ_0, and a uniqueness result ensures that actually $\nu_t = \mu_t$. In other words the limit behaviour μ_t of $\hat{\mu}^N_t$ is the time-marginal of the law $\mu_{[0,T]}$ of the process Y, in the sense that it is the image measure of $\mu_{[0,T]}$ by the canonical projection π_t defined on the path space $\mathcal{C}$ by $\pi_t(f) = f(t)$. Since $\hat{\mu}^N_t$ is the time-marginal of the considered empirical measure $\hat{\mu}^N_{[0,T]}$, it is natural to hope that $\hat{\mu}^N_{[0,T]}$ converge (in some sense) to $\mu_{[0,T]}$. This convergence has indeed been proved in the first works mentionned above, and here we want to extend to this new setting the techniques developed in [16] and more particularly in [6].

We shall assume that the potentials V and W are twice differentiable on $\mathbb{R}^d$, with bounded hessian matrices in the sense that there exist real numbers β, β', γ and γ' such that

$$\beta\, I \leq D^2 V(x) \leq \beta'\, I, \qquad \gamma\, I \leq D^2 W(x) \leq \gamma'\, I, \qquad x \in \mathbb{R}^d. \tag{11}$$

In other words the force fields ∇V and ∇W are assumed to be Lipschitz on the whole $\mathbb{R}^d$.

Under these assumptions, global existence and uniqueness, pathwise and in law, of the solutions to (8) and (10) are proven in [17] for instance for square integrable initial data; moreover the paths are continuous (in time). We shall also assume that the potential W, which gives rise to an interaction term, is symmetric in the sense that $W(-z) = W(z)$ for all $z \in \mathbb{R}^d$. Then we shall prove

Theorem 2. *Let μ_0 be a probability measure on $\mathbb{R}^d$, admitting a finite square-exponential moment in the sense that there exists $a_0 > 0$ such that $\displaystyle\int_{\mathbb{R}^d} e^{a_0 |x|^2}\, d\mu_0(x)$ be finite. Let $(X^i_0)_{1 \leq i \leq N}$ be N independent random variables with common law μ_0. Given $T \geq 0$, let $(X^i)_i$ be the solution of (8) on $[0, T]$ with initial value $(X^i_0)_i$, where V and W are assumed to satisfy (11); let also $\hat{\mu}^N_{[0,T]}$ be the empirical measure associated with the N paths X^i. Let finally $\mu_{[0,T]}$ be the law of the process solution of (10) for some initial value distributed according to μ_0.*

Then, for any $\alpha \in (0, 1/2)$, there exist some positive constants K and N_0 such that

$$\mathbb{P}\left[W_{1,[0,T]}(\mu_{[0,T]}, \hat{\mu}^N_{[0,T]}) > \varepsilon\right] \leq e^{-K\, N\, \varepsilon^2}$$

for all $\varepsilon > 0$ and $N \geq N_0\, \varepsilon^{-2} \exp\left(N_0\, \varepsilon^{-1/\alpha}\right)$.

Here the constants K and N_0 depend on T, V, W, α and $\displaystyle\int_{\mathbb{R}^d} e^{a_0|x|^2}\, d\mu_0(x)$.

By Kantorovich-Rubinstein formulation again, this bound can be written as

$$\mathbb{P}\Big[\sup_{[\varphi]_1\leq 1}\Big(\frac{1}{N}\sum_{i=1}^N \varphi(X^i) - \int_{\mathcal{C}} \varphi(x)\, d\mu_{[0,T]}(x)\Big) > \varepsilon\Big] \leq e^{-K\,N\,\varepsilon^2}. \tag{12}$$

By projection at time t, it implies concentration inequalities for the time-marginals of the empirical measures similar to inequalities (1) and even (2). But above all it gives concentration estimates at the level of the paths. In return we impose some stronger condition on the required size of the sample (however, by (6), one can also get less precise estimates valid for any number N of particles).

Assume for instance that one is interested in the behaviour of a point Y_t evolving according to (10). Then, from (12), one can derive error bounds in the approximation by $\displaystyle\frac{1}{N}\sum_{i=1}^N \varphi(X^i)$ of the expectation of quantities $\varphi(Y)$ which depend on the whole path, such as the distance $d(Y,A) = \inf\big\{|Y_t - y|; t\in[0,T], y\in A\big\}$ of the trajectory to a given set A in $\mathbb{R}^d$, which measures how close Y_t has been to A, or the maximal distance $\sup\big\{|Y_t - x|; t\in[0,T]\big\}$ to a given point x in the phase space $\mathbb{R}^d$: for instance, under the assumptions of Theorem 2, for any $T\geq 0$ and $\alpha\in(0,1/2)$ there exist some positive constants K and N_0 such that

$$\mathbb{P}\Big[\Big|\mathbb{E}\left[d(Y,A)\right] - \frac{1}{N}\sum_{i=1}^N d(X^i, A)\Big| > \varepsilon\Big] \leq e^{-K\,N\,\varepsilon^2}$$

for any Borel set A in $\mathbb{R}^d$, $\varepsilon>0$ and $N\geq N_0\,\varepsilon^{-2}\,\exp\left(N_0\,\varepsilon^{-1/\alpha}\right)$.

Theorem 2 will be proven in detail in Sections 3 and 4 along the following lines. Following [18] we proceed by coupling, by introducing a family of N identically distributed processes $Y^i = (Y^i_t)_{0\leq t\leq T}$ solution to the (nonlinear) stochastic differential equations

$$\begin{cases} dY^i_t &= \sqrt{2}\, dB^i_t - \nabla V(Y^i_t)\, dt - \nabla W * \mu_t(Y^i_t)\, dt \\ Y^i_0 &= X^i_0 \end{cases} \qquad 1\leq i\leq N\,;$$

here μ_t is the solution at time t to (9), but is also the law on $\mathbb{R}^d$ of any Y^i_t by Itô's formula, and, for each i, $B^i = (B^i_t)_{0\leq t\leq T}$ is the Brownian motion driving the evolution of X^i. In particular the paths Y^i are close to the paths X^i and one can prove that there exists some constant C (depending only on T) such that

$$W_{1,[0,T]}(\mu_{[0,T]}, \hat\mu^N_{[0,T]}) \leq C\, W_{1,[0,T]}(\mu_{[0,T]}, \hat\nu^N_{[0,T]})$$

hold almost surely, where $\displaystyle\hat\nu^N_{[0,T]} := \frac{1}{N}\sum_{i=1}^N \delta_{Y^i}$; hence controlling the distance between $\mu_{[0,T]}$ and $\hat\mu^N_{[0,T]}$ reduces to the same issue with $\mu_{[0,T]}$ and $\hat\nu^N_{[0,T]}$.

But, by definition, the N processes Y^i for $1 \leq i \leq N$ are independent and distributed according to $\mu_{[0,T]}$. Then Theorem 1 ensures good concentration estimates for the empirical measure $\hat{\nu}^N_{[0,T]}$ around the common law $\mu_{[0,T]}$. In the end we obtain the bound

$$\mathbb{P}\left[W_{1,[0,T]}(\mu_{[0,T]}, \hat{\mu}^N_{[0,T]}) > \varepsilon\right] \leq \mathbb{P}\left[W_{1,[0,T]}(\mu_{[0,T]}, \hat{\nu}^N_{[0,T]}) > \frac{\varepsilon}{C}\right] \leq e^{-K\,N\,\varepsilon^2}$$

under some condition on ε and N.

The proof is actually an adaptation of the argument given in [6, Section 1.6] of estimates (2) for time-marginals. The current proof turns out to be simpler in the sense that it consists in fewer steps; in return each of these steps is somehow more delicate: for instance, as we shall see in the following sections, the proof of Theorem 2 requires the computation of the metric entropy of some space of Hölder-continuous functions, and checking that the law of Y fulfills the assumptions of this theorem needs some strong integrability in Hölder norm on solutions to stochastic differential equations.

An adaptation of this proof leads to quantitative estimates on the phenomenon of **propagation of chaos**. For instance, letting

$$\hat{\mu}^{N,2}_{[0,T]} := \frac{1}{N(N-1)} \sum_{i \neq j} \delta_{(X^i,\, X^j)}$$

be the empirical measure on pairs of paths, the asymptotic independence of two paths (among N) can be estimated as in

Theorem 3. *With the same notation and assumptions as in Theorem 2, for all* $T \geq 0$ *and* $\alpha \in (0, 1/2)$ *there exist some positive constants* K *and* N_0 *such that*

$$\mathbb{P}\left[W_{1,[0,T]}(\mu_{[0,T]} \otimes \mu_{[0,T]}, \hat{\mu}^{N,2}_{[0,T]}) > \varepsilon\right] \leq e^{-K\,N\,\varepsilon^2}$$

for all $\varepsilon > 0$ *and* $N \geq N_0\,\varepsilon^{-2}\,\exp(N_0\,\varepsilon^{-1/\alpha})$.

Here the constants K and N_0 depend on T, V, W, α and a finite square-exponential moment of μ_0, and $W_{1,[0,T]}$ stands for the Wasserstein distance of order 1 on the product space $\mathcal{C} \times \mathcal{C}$. The proof consists in writing the coupling argument for pairs of paths and comparing $\mu_{[0,T]} \otimes \mu_{[0,T]}$ and $\hat{\nu}^{N,2}_{[0,T]} := \frac{1}{N(N-1)} \sum_{i \neq j} \delta_{(Y^i, Y^j)}$ by means of $\frac{1}{N^2} \sum_{i,j} \delta_{(Y^i, Y^j)}$.

Let us finally note that it would be desirable to relax the assumptions made on the potentials V and W, in particular so as to include the interesting case of the cubic potential $W(z) = |z|^3/3$ on $\mathbb{R}$, which models the interaction among one-dimensional granular media (see [3]). It could also be interesting to consider the whole trajectories $(X_t)_{t \geq 0}$, and derive concentration bounds on functionals such as hitting times for instance.

Before turning to the proofs we briefly recall the **plan of the paper**. In the coming section we prove Theorem 1 for general $\mathcal{C}$-valued independent variables. The study of the particle system is addressed in the following two sections: in Section 3 we reduce our concentration issue on interacting particles to the same issue for independent variables by a coupling argument, whereas in Section 4 we check that we can apply our general concentration result to these independent variables; with this in hand we can prove Theorem 2. An appendix is devoted to a general metric entropy estimate in a space of Hölder-continuous functions, which enters the proof of Theorem 1.

2. A preliminary result on independent variables

The aim of this section is to prove Theorem 1 for N independent and identically distributed random variables valued in $\mathcal{C}$. We have seen how this result, applied to the artificial processes $(Y_t^i)_{0\leq t\leq T}$, enters the study of our interacting particle system.

The proof goes in three steps: truncation to a ball $\mathcal{B}_R^\alpha$ of $\mathcal{C}^\alpha$, compact for the topology induced by the uniform norm; covering of $\mathcal{B}_R^\alpha$ and then of $\mathcal{P}(\mathcal{B}_R^\alpha)$ by small balls on which one develops Sanov's argument; conclusion of the proof by optimizing the introduced parameters. Since the argument follows the lines of the proof given in [6, Section 2.1] in the finite dimensional case, in which μ is a measure on $\mathbb{R}^d$, we shall only sketch it, stressing only the bounds specific to our new framework. We refer to [6] for further details.

Step 1. Truncation. Given $R > 0$, to be chosen later on, we denote $\mathcal{B}_R^\alpha$ the ball $\{f \in \mathcal{C}^\alpha; \|f\|_\alpha \leq R\}$ of center 0 and radius R in $\mathcal{C}^\alpha$. This set $\mathcal{B}_R^\alpha$ is a compact subset of $\mathcal{C}$ for the topology induced by the uniform norm $\|\cdot\|_\infty$: indeed it is relatively compact in $\mathcal{C}$ by Ascoli's theorem, and closed since if f in $\mathcal{C}$ is the uniform limit of a sequence $(f_n)_n$ in $\mathcal{C}^\alpha$, then $\|f\|_\alpha \leq \liminf_n \|f_n\|_\alpha$, and in particular f belongs to $\mathcal{B}_R^\alpha$ if so do the f_n.

Letting $\mathbf{1}_{\mathcal{B}_R^\alpha}$ be the indicator function of $\mathcal{B}_R^\alpha$, we truncate μ into a probability measure μ_R on the ball $\mathcal{B}_R^\alpha$, defined as

$$\mu_R := \frac{\mathbf{1}_{\mathcal{B}_R^\alpha}\,\mu}{\mu[\mathcal{B}_R^\alpha]}.$$

Note that $\mu[\mathcal{B}_R^\alpha]$ is positive for R larger than some R_0 depending only on a and $E_a := \int_{\mathcal{C}} e^{a\|x\|_\alpha^2}\,d\mu(x)$. In this step we reduce the concentration problem for $\mathcal{C}$ to the same issue for the compact ball $\mathcal{B}_R^\alpha$, by bounding the quantity $\mathbb{P}\big[W_p(\mu,\hat{\mu}^N) > \varepsilon\big]$ in terms of μ_R and an associated empirical measure $\hat{\mu}_R^N := \dfrac{1}{N}\displaystyle\sum_{k=1}^N \delta_{X_R^k}$ where the X_R^k are independent variables with law μ_R.

Bounding by above the $\|\cdot\|_\infty$ norms by $\|\cdot\|_\alpha$ norms when necessary, we proceed exactly as in [6, proof of Theorem 1.1] to obtain the bound

$$\mathbb{P}\big[W_{p,[0,T]}(\mu,\hat{\mu}^N) > \varepsilon\big] \leq \mathbb{P}\Big[W_{p,[0,T]}(\mu_R,\hat{\mu}_R^N) > \eta\,\varepsilon - 2\,E_a^{1/p} R\, e^{-\frac{a}{p}R^2}\Big] + \exp\Big(-N\big(\theta(1-\eta)^p\varepsilon^p - E_a\, e^{(a_1-a)\,R^2}\big)\Big); \quad (13)$$

here p is any real number in $[1,2)$, η in $(0,1)$, $\varepsilon, \theta > 0$, $a_1 < a$ and R is constrained to be larger than $R_2 \max(1, \theta^{\frac{1}{2-p}})$ for some constant R_2 depending only on E_a, a, a_1 and p.

In the case when $p = 2$, we obtain

$$\mathbb{P}\big[W_{2,[0,T]}(\mu, \hat{\mu}^N) > \varepsilon\big] \leq \mathbb{P}\left[W_{2,[0,T]}(\mu_R, \hat{\mu}_R^N) > \eta\, \varepsilon - 2\, E_a^{1/2} R\, e^{-\frac{a}{2}R^2}\right] + \exp\left(-N\Big(\frac{a_1}{2}(1-\eta)^2\varepsilon^2 - 2E_a^2\, e^{(a_1 - a)\, R^2}\Big)\right). \quad (14)$$

Step 2. Sanov's argument on small balls. In view of (13) for $p < 2$ or (14) for $p = 2$, we now aim at bounding $\mathbb{P}\,[\hat{\mu}_R^N \in \mathcal{A}]$ where

$$\mathcal{A} := \Big\{\nu \in \mathcal{P}(\mathcal{B}_R^\alpha); \quad W_{p,[0,T]}(\nu, \mu_R) \geq \eta\, \varepsilon - 2\, E_a^{1/p} R\, e^{-\frac{a}{p}R^2}\Big\}.$$

For that purpose, reasoning as in [6], we let $\delta > 0$ and cover $\mathcal{A}$ with $\mathcal{N}(\mathcal{A}, \delta)$ balls $(B_i)_{1 \leq i \leq \mathcal{N}(\mathcal{A},\delta)}$ with radius $\delta/2$ in $W_{p,[0,T]}$ distance. Then one can develop Sanov's argument on each of these compact and convex balls, and obtain the bound

$$\mathbb{P}[\hat{\mu}_R^N \in \mathcal{A}] \leq \mathbb{P}\Big[\hat{\mu}_R^N \in \bigcup_{i=1}^{\mathcal{N}(\mathcal{A},\delta)} B_i\Big] \leq \sum_{i=1}^{\mathcal{N}(\mathcal{A},\delta)} \mathbb{P}\left[\hat{\mu}_R^N \in B_i\right] \leq \sum_{i=1}^{\mathcal{N}(\mathcal{A},\delta)} \exp\Big(-N \inf_{\nu \in B_i} H(\nu|\mu_R)\Big). \quad (15)$$

Then, from the $T_p(\lambda)$ inequality for μ, one establishes an approximate $T_p(\lambda)$ inequality for μ_R: namely, for any $\lambda_1 < \lambda$ there exists K_1 such that

$$H(\nu, \mu_R) \geq \frac{\lambda_1}{2}\, W_{p,[0,T]}(\nu, \mu_R)^2 - K_1\, R^2\, e^{-a\, R^2}$$

for any measure ν on $\mathcal{B}_R^\alpha$. With this inequality in hand, given $1 \leq p < 2$ and $\lambda_2 < \lambda_1 < \lambda$, one deduces from (15) the existence of some positive constants δ_1, η_1 and K_1 such that

$$\mathbb{P}\left[W_{p,[0,T]}(\mu_R, \hat{\mu}_R^N) > \eta\, \varepsilon - 2\, E_a^{1/p} R e^{-\frac{a}{p}R^2}\right] \leq \mathcal{N}(\mathcal{A}, \delta)\, \exp\left(-N\Big(\frac{\lambda_2}{2}\, \varepsilon^2 - K_1 R^2 e^{-aR^2}\Big)\right) \quad (16)$$

where we have chosen $\delta := \delta_1 \varepsilon$ and $\eta := \eta_1$.

In the case when $p = 2$, we do not choose η at this stage, and simply obtain

$$\mathbb{P}\left[W_{2,[0,T]}(\mu_R, \hat{\mu}_R^N) > \eta\, \varepsilon - 2\, E_a^{1/2} R\, e^{-\frac{a}{2}R^2}\right] \leq \mathcal{N}(\mathcal{A}, \delta)\, \exp\left(-N\Big(\frac{\lambda_2}{2}\, \eta^2 \varepsilon^2 - K_1 R^2 e^{-aR^2}\Big)\right)$$

where $\delta := \delta_1 \varepsilon$.

Then, since $\mathcal{A}$ is a subset of $\mathcal{P}(\mathcal{B}_R^\alpha)$, Theorem 10 in the Appendix enables to bound $\mathcal{N}(\mathcal{A}, \delta)$ with $\delta = \delta_1\, \varepsilon$ by

$$\exp\Big(K_2 (R\, \varepsilon^{-1})^d\, 3^{K_2 (R\, \varepsilon^{-1})^{1/\alpha}}\, \ln\big(\max(1, K_2\, R\, \varepsilon^{-1})\big)\Big) \quad (17)$$

for some constant K_2 depending neither on ε nor on R.

Remark 4. The order of magnitude of this covering number in an infinite-dimensional setting constitutes a main change by comparison with the finite-dimensional setting of [6], and will influence the final condition on the size N of the sample.

Step 3. Conclusion of the argument. We first focus on the case when $p \in [1,2)$. Collecting estimates (13), (16) and (17), we obtain, given $\lambda_2 < \lambda$ and $a_1 < a$, the existence of positive constants K_1, K_2, K_3 and R_3 depending on $E_a, a, a_1, \alpha, \lambda$ and λ_2 such that

$$\begin{aligned}\mathbb{P}\left[W_{p,[0,T]}(\mu, \hat{\mu}^N) > \varepsilon\right] \\ \leq \exp\left(K_2 (R\,\varepsilon^{-1})^d\, 3^{K_2 (R\,\varepsilon^{-1})^{1/\alpha}}\, \ln\left(\max(1, K_2\, R\,\varepsilon^{-1})\right) - N\Big(\frac{\lambda_2}{2}\,\varepsilon^2 - K_1 R^2 e^{-\alpha R^2}\Big)\right) \\ + \exp\left(-N\big(K_3\,\theta\,\varepsilon^p - K_4 e^{(a_1 - a)\,R^2}\big)\right) \quad (18)\end{aligned}$$

for all $\varepsilon, \theta > 0$ and $R \geq R_3 \max(1, \theta^{\frac{1}{2-p}})$, and for some constant $K_4 = K_4(\theta, a_1)$.

Then let $\lambda_3 < \lambda_2$. One can prove that the first term in the right-hand side in (18) is bounded by $\exp\Big(-\dfrac{\lambda_3}{2}\,N\,\varepsilon^2\Big)$ provided

$$R^2 \geq A \max(1, \varepsilon^2, \ln(\varepsilon^{-2}))\,, \quad N\varepsilon^2 \geq B\, 3^{C(R\varepsilon^{-1})^{1/\alpha}} \quad (19)$$

for some positive constants A, B and C depending also on λ_3. Moreover, for $\theta = \dfrac{\varepsilon^{2-p}\,\lambda_3}{2\,K_3}$, also the second term in the right-hand side in (18) is bounded by $\exp\Big(-\dfrac{\lambda_3}{2}\,N\,\varepsilon^2\Big)$ as soon as $R^2 \geq R_4 \max(1, \ln(\varepsilon^{-2}))$, for some constant R_4 depending on λ_3.

Letting $R = \varepsilon\Big(\dfrac{1}{C\,\ln 3}\ln\dfrac{N\varepsilon^2}{B}\Big)^\alpha$ if $\varepsilon \in (0,1)$ and $R = \sqrt{A}\,\varepsilon$ otherwise, and $\alpha' < \alpha$, both conditions in (19) hold true as soon as $N \geq N_0\,\varepsilon^{-2}\exp(N_0\,\varepsilon^{-1/\alpha'})$ for some constant N_0 depending on $E_a, a, \lambda, \lambda_3, \alpha$ and α'. Finally, given $\lambda' < \lambda_3 < \lambda$, this condition ensures that

$$\mathbb{P}\left[W_{p,[0,T]}(\mu, \hat{\mu}^N) > \varepsilon\right] \leq 2\,\exp\Big(-\frac{\lambda_3}{2}\,N\,\varepsilon^2\Big) \leq \exp\Big(-\frac{\lambda'}{2}\,N\,\varepsilon^2\Big),$$

possibly for some larger N_0. This concludes the argument in the case when $p \in [1,2)$.

In the case when $p = 2$, given $0 < \eta < 1$, $\lambda_3 < \lambda_2$ and $a_2 < a_1$, the same condition on N and ε (for some N_0) is sufficient for the bound

$$\mathbb{P}\left[W_{2,[0,T]}(\mu, \hat{\mu}^N) > \varepsilon\right] \leq \exp\Big(-\frac{\lambda_3}{2}\,\eta^2\,N\,\varepsilon^2\Big) + \exp\Big(-\frac{a_2}{2}(1-\eta)^2\,N\,\varepsilon^2\Big)$$

to hold (by (14)). One optimizes this bound by letting

$$a_2 = a\,\frac{\lambda_3}{\lambda}\,(\in [0,a)) \quad \text{and} \quad \eta = \frac{\sqrt{a_2}}{\sqrt{a_2} + \sqrt{\lambda_3}}.$$

Given $\lambda' < \lambda_3 < \lambda$, this ensures the existence of N_0 such that

$$\mathbb{P}\big[W_{2,[0,T]}(\mu, \hat{\mu}^N) > \varepsilon\big] \leq 2\,\exp\Big(-\frac{\lambda_3}{2}\frac{a}{(\sqrt{a} + \sqrt{\lambda})^2}\,N\,\varepsilon^2\Big) \leq \exp\Big(-\frac{\lambda'}{2}\frac{1}{(1 + \sqrt{\lambda/a})^2}\,N\,\varepsilon^2\Big)$$

for any $\varepsilon > 0$ and $N \geq N_0\,\varepsilon^{-2}\exp(N_0\,\varepsilon^{-1/\alpha'})$. This concludes the proof of Theorem 1 in this second and last case.

3. Coupling argument

Here begins the proof of Theorem 2 on the behaviour of our large interacting particle system.

We recall that we are given N independent variables X_0^i in $\mathbb{R}^d$, with common law μ_0, and N independent Brownian motions $B^i = (B_t^i)_{0\le t\le T}$ in $\mathbb{R}^d$, and we consider the solutions $X^i = (X_t^i)_{0\le t\le T}$ to the coupled stochastic differential equations

$$dX_t^i = \sqrt{2}\, dB_t^i - \nabla V(X_t^i)\, dt - \frac{1}{N}\sum_{j=1}^N \nabla W(X_t^i - X_t^j)\, dt, \qquad 1 \le i \le N.$$

We also let $\mu_{[0,T]}$ be the law of the process $Y = (Y_t)_{0\le t\le T}$ defined by

$$dY_t = \sqrt{2}\, dB_t - \nabla V(Y_t)\, dt - \nabla W * \mu_t(Y_t)\, dt$$

and starting at some Y_0 drawn according to μ_0; here $B = (B_t)_{0\le t\le T}$ also is a Brownian motion and μ_t is the law of Y_t, that is, the time-marginal of $\mu_{[0,T]}$ at time t.

We want to compare this law $\mu_{[0,T]}$ and the empirical measure of the paths

$$\hat{\mu}^N_{[0,T]} = \frac{1}{N}\sum_{i=1}^N \delta_{X^i}.$$

For this purpose we introduce N independent processes $Y^i = (Y_t^i)_{0\le t\le T}$ defined by

$$dY_t^i = \sqrt{2}\, dB_t^i - \nabla V(Y_t^i)\, dt - \nabla W * \mu_t(Y_t^i)\, dt, \qquad 1 \le i \le N \tag{20}$$

for the same Brownian motions B^i, and such that $Y_0^i = X_0^i$ initially. We let

$$\hat{\nu}^N_{[0,T]} := \frac{1}{N}\sum_{i=1}^N \delta_{Y^i} \tag{21}$$

and in this section we reduce the issue to measuring the distance between $\mu_{[0,T]}$ and $\hat{\nu}^N_{[0,T]}$.

Proposition 5. *In the above notation and under the assumptions*

$$\beta\, I \le D^2 V(x)\,, \quad \gamma\, I \le D^2 W(x) \le \gamma'\, I\,, \qquad x \in \mathbb{R}^d$$

on V and W, where β, γ and γ' are real numbers, for any $T \ge 0$ there exists some constant C depending only on β, γ, γ' and T such that

$$W_{1,[0,T]}(\mu_{[0,T]}, \hat{\mu}^N_{[0,T]}) \le C\, W_{1,[0,T]}(\mu_{[0,T]}, \hat{\nu}^N_{[0,T]})$$

almost surely.

Proof. We first follow the lines of the proof of [6, Proposition 5.1], but in the end we want an estimate on the *trajectories*. Since for each i both processes X^i and Y^i are driven by the same Brownian motion B^i, the process $X^i - Y^i$ satisfies

$$d(X_t^i - Y_t^i) = -\big(\nabla V(X_t^i) - \nabla V(Y_t^i)\big)\, dt - \big(\nabla W * \hat{\mu}_t^N(X_t^i) - \nabla W * \mu_t(Y_t^i)\big)\, dt.$$

In particular, letting $u \cdot v$ denote the scalar product of two vectors u and v in $\mathbb{R}^d$,

$$\frac{1}{2}\frac{d}{dt}|X_t^i - Y_t^i|^2 = -\big(\nabla V(X_t^i) - \nabla V(Y_t^i)\big)\cdot(X_t^i - Y_t^i) - \big(\nabla W * \hat{\mu}_t^N(X_t^i) - \nabla W * \mu_t(Y_t^i)\big)\cdot(X_t^i - Y_t^i). \quad (22)$$

We decompose the last term according to

$$\nabla W * \hat{\mu}_t^N(X_t^i) - \nabla W * \mu_t(Y_t^i) = \big(\nabla W * \hat{\mu}_t^N - \nabla W * \mu_t\big)(X_t^i) + \big(\nabla W * \mu_t(X_t^i) - \nabla W * \mu_t(Y_t^i)\big).$$

By our assumption on $D^2 W$, the map $\nabla W(X_t^i - \cdot\,)$ is Γ-Lipschitz with $\Gamma := \max(|\gamma|, |\gamma'|)$. Consequently, by the Kantorovich-Rubinstein dual formulation (5) of $W_{1,\tau}$,

$$\Big|\nabla W * (\hat{\mu}_t^N - \mu_t)(X_t^i)\Big| = \Big|\int_{\mathbb{R}^d} \nabla W(X_t^i - y)\, d(\hat{\mu}_t^N - \mu_t)(y)\Big| \leq \Gamma\, W_{1,\tau}(\hat{\mu}_t^N, \mu_t).$$

Then (22) and our convexity assumptions on V and W imply

$$\frac{1}{2}\frac{d}{dt}|X_t^i - Y_t^i|^2 \leq -(\beta+\gamma)\,|X_t^i - Y_t^i|^2 + \Gamma\, W_{1,\tau}(\hat{\mu}_t^N, \mu_t)\,|X_t^i - Y_t^i|.$$

In particular, by Gronwall's lemma,

$$|X_t^i - Y_t^i| \leq \Gamma \int_0^t e^{-(\beta+\gamma)(t-u)}\, W_{1,\tau}(\hat{\mu}_u^N, \mu_u)\, du$$

since initially $X_0^i = Y_0^i$. Consequently, by convexity of the $W_{1,[0,t]}$ distance,

$$\begin{aligned} W_{1,[0,t]}(\hat{\mu}_{[0,t]}^N, \hat{\nu}_{[0,t]}^N) &\leq \frac{1}{N}\sum_{i=1}^N \sup_{0\leq s\leq t} |X_s^i - Y_s^i| \\ &\leq \frac{1}{N}\sum_{i=1}^N \sup_{0\leq s\leq t} \Gamma \int_0^s e^{-(\beta+\gamma)(s-u)} W_{1,\tau}(\hat{\mu}_u^N, \mu_u)\, du \qquad (23) \\ &\leq \Gamma e^{|\beta+\gamma| T} \int_0^t W_{1,\tau}(\hat{\mu}_u^N, \mu_u)\, du \end{aligned}$$

for all $0 \leq t \leq T$. But

$$W_{1,\tau}(\hat{\mu}_u^N, \mu_u) \leq W_{1,[0,u]}(\hat{\mu}_{[0,u]}^N, \mu_{[0,u]}) \leq W_{1,[0,u]}(\hat{\mu}_{[0,u]}^N, \hat{\nu}_{[0,u]}^N) + W_{1,[0,u]}(\hat{\nu}_{[0,u]}^N, \mu_{[0,u]})$$

by the projection relation (3) and triangular inequality for $W_{1,[0,u]}$, so

$$W_{1,[0,t]}(\hat{\mu}_{[0,t]}^N, \hat{\nu}_{[0,t]}^N) \leq \Gamma e^{|\beta+\gamma|T} \int_0^t \exp\Big(\Gamma e^{|\beta+\gamma|T}(t-u)\Big)\, W_{1,[0,u]}(\hat{\nu}_{[0,u]}^N, \mu_{[0,u]})\, du \qquad (24)$$

for all $0 \leq t \leq T$ by Gronwall's lemma again. Then, given $0 \leq u \leq t$,

$$W_{1,[0,u]}(\hat{\nu}_{[0,u]}^N, \mu_{[0,u]}) \leq W_{1,[0,t]}(\hat{\nu}_{[0,t]}^N, \mu_{[0,t]})$$

since $\hat{\nu}_{[0,u]}^N$ and $\mu_{[0,u]}$ are the respective image measures of $\hat{\nu}_{[0,t]}^N$ and $\mu_{[0,t]}$ by the 1-Lipschitz map defined from $\mathcal{C}([0,t], \mathbb{R}^d)$ into $\mathcal{C}([0,u], \mathbb{R}^d)$ as the restriction to $[0,u]$. Hence

$$W_{1,[0,t]}(\hat{\mu}_{[0,t]}^N, \hat{\nu}_{[0,t]}^N) \leq C\, W_{1,[0,t]}(\hat{\nu}_{[0,t]}^N, \mu_{[0,t]})$$

by (24), for some constant C depending only on T, β, γ and γ'. This concludes the argument by triangular inequality. □

Remark 6. If moreover $\beta+\gamma > \Gamma$, where again $\Gamma := \max(|\gamma|, |\gamma'|)$ then we can let C be $(\beta+\gamma)\,(\beta+\gamma-\Gamma)^{-1}$ in Proposition 5, independently of T. Indeed, if $\beta+\gamma>0$, then (23) leads to

$$W_{1,[0,t]}(\hat\mu^N_{[0,t]}, \hat\nu^N_{[0,t]}) \leq \Gamma \sup_{0\leq s\leq t} \int_0^t e^{-(\beta+\gamma)(s-u)}\, du \sup_{0\leq u\leq t} W_{1,\tau}(\hat\mu^N_u, \mu_u) \leq \frac{\Gamma}{\beta+\gamma}\, W_{1,[0,t]}(\hat\mu^N_{[0,t]}, \mu_{[0,t]})$$

and by triangular inequality

$$W_{1,[0,t]}(\hat\mu^N_{[0,t]}, \mu_{[0,t]}) \leq \frac{\beta+\gamma}{\beta+\gamma-\Gamma}\, W_{1,[0,t]}(\hat\nu^N_{[0,t]}, \mu_{[0,t]})$$

provided $\beta+\gamma>\Gamma$.

This is reminiscent of the fact that, under some convexity assumptions on V and W, such as $\beta>0$, $\beta+2\,\gamma>0$, it has been proven in [8, 9, 16] that the time-marginal μ_t of the measure $\mu_{[0,t]}$ converges, as t goes to infinity, to the stationary solution to the limit equation (9). One can also prove in this context that (in expectation) observables of the particle system are bounded in time.

Hence, under this kind of assumptions, one could hope for some uniform in time constants in this coupling argument: that was obtained in [6, Proposition 5.1] for the time-marginals, and here for the whole processes. However, contrary to [6] where this property was used to approach the stationary solution by coupling together estimates of concentration of the empirical measure (as N goes to infinity) with estimates of convergence to equilibrium (as t goes to infinity), in this work we are concerned with finite time intervals only, and shall not use this specific property in the sequel.

4. Integrability in Hölder norm

In the previous section we have reduced the issue of measuring the distance between $\mu_{[0,T]}$ and $\hat\mu^N_{[0,T]}$ to measuring the distance between $\mu_{[0,T]}$ and the empirical measure $\hat\nu^N_{[0,T]}$ of N independent random variables drawn according to $\mu_{[0,T]}$.

We now solve the latter issue by proving that the measure $\mu_{[0,T]}$ fulfills the hypotheses of Theorem 1 with $p=1$, namely, that there exist $\alpha\in(0,1]$ and $a>0$ such that

$$\int_{\mathcal{C}} e^{a\|x\|^2_\alpha}\, d\mu_{[0,T]}(x) := \mathbb{E}\exp(a\|Y\|^2_\alpha) < +\infty.$$

Here again $\mathcal{C}$ stands for $\mathcal{C}([0,T],\mathbb{R}^d)$, $\|f\|_\alpha$ for the Hölder norm of a function f on $[0,T]$, and $Y=(Y_t)_{0\leq t\leq T}$ is the solution to the stochastic differential equation

$$dY_t = \sqrt{2}\, dB_t - \nabla V(Y_t)\, dt - \nabla W * \mu_t(Y_t)\, dt \tag{25}$$

starting at Y_0 drawn according to μ_0, where μ_t is the law of Y_t.

Proposition 7. *Let μ_0 be a probability measure on $\mathbb{R}^d$ admitting a finite square-exponential moment and let Y_0 be drawn according to μ_0. Given $T \geq 0$, V and W satisfying hypotheses (11), let Y be the solution to (25) starting at Y_0. Then, for any $\alpha \in (0, 1/2)$, there exists $a > 0$, depending on μ_0 only through a finite square-exponential moment, such that $\mathbb{E} \exp(a \|Y\|^2_\alpha)$ be finite.*

Assuming this result for the moment we can now conclude the *proof of Theorem 2.* Let indeed α be given in $(0, 1/2)$, and $\alpha_0 \in (\alpha, 1/2)$. Then, by Proposition 7 and Theorem 1, applied with $\alpha = \alpha_0$ and $\alpha' = \alpha$, there exist some constants $\tilde{K}$ and $\tilde{N}_0$, depending on α_0, α, T and a square-exponential moment of μ_0, such that

$$\mathbb{P}\left[W_{1,[0,T]}(\mu_{[0,T]}, \hat{\nu}^N_{[0,T]}) > \tilde{\varepsilon}\right] \leq e^{-\tilde{K} N \tilde{\varepsilon}^2}$$

for any $\tilde{\varepsilon} > 0$ and $N \geq \tilde{N}_0\, \tilde{\varepsilon}^{-2} \exp(\tilde{N}_0\, \tilde{\varepsilon}^{-1/\alpha})$, where $\hat{\nu}^N_{[0,T]}$ is defined by (20) and (21). Then, by Proposition 5, there exist some constants C, depending only on T, and then K and N_0, depending on α_0, α, T and a finite square-exponential moment of μ_0, such that

$$\mathbb{P}\left[W_{1,[0,T]}(\mu_{[0,T]}, \hat{\mu}^N_{[0,T]}) > \varepsilon\right] \leq \mathbb{P}\left[W_{1,[0,T]}(\mu_{[0,T]}, \hat{\nu}^N_{[0,T]}) > \varepsilon/C\right] \leq e^{-K N \varepsilon^2}$$

for any $\varepsilon > 0$ and $N \geq N_0\, \varepsilon^{-2} \exp(N_0\, \varepsilon^{-1/\alpha})$. This concludes the argument. □

Proof of Proposition 7. It is necessary and sufficient to prove that there exist positive constants a_1 and a_2 such that $\mathbb{E} \exp(a_1 \|Y\|^2_\infty)$ and $\mathbb{E} \exp(a_2 [Y]^2_\alpha)$ be finite, where $[\,\cdot\,]_\alpha$ stands for the Hölder seminorm defined in Section 1.2.

Step 1. We start with the expectation in uniform norm. For this we first note that, according to [6, Proposition 3.1], there exist some positive constants M and $\overline{a}$, depending on μ_0 only through a finite square-exponential moment, such that $\sup_{0 \leq t \leq T} \mathbb{E}\, |Y_t|^2$ and $\sup_{0 \leq t \leq T} \mathbb{E} \exp(\overline{a} |Y_t|^2)$ be finite and bounded by M.

Then we let b be some smooth function on $[0, T]$, to be chosen later on, and we let $Z_t = \exp(b(t)\, |Y_t|^2)$. We want to prove that $\mathbb{E} \sup_{0 \leq t \leq T} Z_t$ is finite for some positive function b. By Itô's formula,

$$Z_t = Z_0 + M_t + \int_0^t \Big[b'(s)|Y_s|^2 + 2\, d\, b(s) + 4\, b(s)^2\, |Y_s|^2 - 2\, b(s)\, Y_s \cdot (\nabla V(Y_s) + \nabla W * \mu_s(Y_s)) \Big]\, Z_s\, ds$$

where $(M_t)_{0 \leq t \leq T}$ is the martingale defined as

$$M_t = 2\sqrt{2} \int_0^t b(s)\, Z_s\, Y_s \,\cdot\, dB_s.$$

But $D^2 V(x) \geq \beta\, I$ for all $x \in \mathbb{R}^d$, so for any $\delta > 0$ and $y \in \mathbb{R}^d$ we have

$$-y \,\cdot\, \nabla V(y) \leq (\delta - \beta)|y|^2 + \frac{|\nabla V(0)|^2}{4\delta}.$$

Furthermore ∇W is Γ-Lipschitz and $\nabla W(0)=0$, so

$$-2\,y\,\cdot\,\nabla W*\mu_s(y)\leq 2\,\Gamma\int_{\mathbb{R}^d}|y|\,|y-z|\,d\mu_s(z)\leq 3\,\Gamma\,|y|^2+\Gamma\int_{\mathbb{R}^d}|z|^2\,d\mu_s(z).$$

But $\displaystyle\int_{\mathbb{R}^d}|z|^2\,d\mu_s(z)=\mathbb{E}\,|Y_s|^2\leq M$ on $[0,T]$, so collecting all terms together, we obtain

$$Z_t\leq Z_0+M_t+\int_0^t\big[C(s)+D(s)\,|Y_s|^2\big]\,Z_s\,ds$$

where $C(s)=\Big(2\,d+\Gamma\,M+\dfrac{|\nabla V(0)|^2}{2\delta}\Big)\,b(s)$ and $D(s)=b'(s)+4\,b(s)^2+\big(2\,(\delta-\beta)+3\,\Gamma\big)\,b(s)$. Given $\delta>0$ such that $c:=2\,(\delta-\beta)+3\,\Gamma$ be positive, we let $b(s)$ such that $D(s)\equiv 0$, namely

$$b(s)=e^{-cs}\,\big(b(0)^{-1}+4\,c^{-1}(1-e^{-cs})\big)^{-1}$$

for some $b(0)$ to be chosen later on. In particular b is a nonincreasing continuous positive function on $[0,+\infty)$, and, for this function b, Z_t almost surely satisfies the inequality

$$Z_t\leq Z_0+M_t+C(0)\int_0^t Z_s\,ds.$$

In particular

$$\mathbb{E}\sup_{0\leq t\leq T}Z_t\leq\mathbb{E}\,Z_0+\mathbb{E}\sup_{0\leq t\leq T}M_t+C(0)\int_0^T\mathbb{E}\,Z_s\,ds. \tag{26}$$

But, by Cauchy-Schwarz' and Doob's inequalities,

$$\Big(\mathbb{E}\sup_{0\leq t\leq T}M_t\Big)^2\leq\mathbb{E}\sup_{0\leq t\leq T}|M_t|^2\leq 2\sup_{0\leq t\leq T}\mathbb{E}\,|M_t|^2.$$

Then, by Itô's formula again,

$$\begin{aligned}\mathbb{E}|M_t|^2=8\int_0^t b(s)^2\,\mathbb{E}\big[Z_s^2\,|Y_s|^2\big]\,ds\quad&\leq\quad 8\,b(0)\int_0^t\mathbb{E}\Big[b(s)\,|Y_s|^2\,\exp(2\,b(s)\,|Y_s|^2)\Big]\,ds\\&\leq\quad 8\,b(0)\int_0^t\mathbb{E}\,\exp(3\,b(0)\,|Y_s|^2)\,ds.\end{aligned}$$

Choosing $b(0)\leq\overline{a}/3$, this ensures that $\displaystyle\sup_{0\leq t\leq T}\mathbb{E}|M_t|^2$, whence $\displaystyle\mathbb{E}\sup_{0\leq t\leq T}M_t$, is finite.

Since, for this $b(0)$, $\displaystyle\sup_{0\leq t\leq T}\mathbb{E}\,Z_t$ also is finite, it follows from (26) that so is $\displaystyle\mathbb{E}\sup_{0\leq t\leq T}Z_t$, which concludes the argument for the expectation in uniform norm with $a_1=b(T)$.

Step 2. We now turn to the expectation in Hölder seminorm. Writing the solution as

$$Y_t=Y_0+B_t-\int_0^t\big(\nabla V(Y_s)+\nabla W*\mu_s(Y_s)\big)\,ds$$

we obtain

$$[Y]_\alpha\leq[B]_\alpha+\Big[\int_0^\cdot\big(\nabla V(Y_s)+\nabla W*\mu_s(Y_s)\big)\,ds\Big]_\alpha$$

almost surely; here Y and B stand as before for the map $t \mapsto Y_t$ and $t \mapsto B_t$ respectively, and $\displaystyle\int_0^{\cdot} \varphi(s)\,ds$ is an antiderivative of φ. Hence, by Cauchy-Schwarz' inequality,

$$\mathbb{E}\exp(a_2[Y]_\alpha^2) \leq \big(\mathbb{E}\exp(4\,a_2[B]_\alpha^2)\big)^{1/2}\Big(\mathbb{E}\exp\Big(4\,a_2\Big[\int_0^{\cdot}(\nabla V(Y_s)+\nabla W*\mu_s(Y_s))\,ds\Big]_\alpha^2\Big)\Big)^{1/2}.$$

But, on one hand, $\mathbb{E}\exp\big(4\,a_2[B]_\alpha^2\big)$ is finite for a_2 small enough (see [12, Theorem 1.3.2] for instance, with $E = \mathcal{C}$ and $N(f) = [f]_\alpha$). On the other hand, by assumption (11), ∇V and ∇W are respectively B- and Γ-Lipschitz with $\mathrm{B} := \max(|\beta|, |\beta'|)$ and $\Gamma := \max(|\gamma|, |\gamma'|)$, so there exists some constant A such that

$$\big|\nabla V(y) + \nabla W * \mu_s(y)\big| \leq \mathrm{A} + (\mathrm{B}+\Gamma)|y|$$

for all $y \in \mathbb{R}^d$ and $s \in [0,T]$. In particular

$$\begin{aligned}\Big[\int_0^{\cdot}\big(\nabla V(Y_s)+\nabla W*\mu_s(Y_s)\big)\,ds\Big]_\alpha &\leq \sup_{0\leq s,t\leq T}\frac{1}{|t-s|^\alpha}\int_s^t\big(\mathrm{A}+(\mathrm{B}+\Gamma)|Y_u|\big)\,du\\ &\leq T^{1-\alpha}\big(\mathrm{A}+(\mathrm{B}+\Gamma)\|Y\|_\infty\big)\end{aligned}$$

almost surely, and

$$\begin{aligned}&\mathbb{E}\exp\Big(4\,a_2\Big[\int_0^{\cdot}\big(\nabla V(Y_s)+\nabla W*\mu_s(Y_s)\big)\,ds\Big]_\alpha^2\Big)\\ &\qquad\leq \exp\big(8\,a_2\,T^{2-2\alpha}\mathrm{A}^2\big)\,\mathbb{E}\exp\Big(8\,a_2\,T^{2-2\alpha}(\mathrm{B}+\Gamma)^2\|Y\|_\infty^2\Big)\end{aligned}$$

which by step 1 is finite as soon as $8\,a_2\,T^{2-2\alpha}(\mathrm{B}+\Gamma)^2 \leq a_1$.

On the whole, $\mathbb{E}\exp(a_2[Y]_\alpha^2)$ is indeed finite for a_2 small enough, depending on μ_0 only through a finite square-exponential moment, which concludes the argument. □

Appendix. Metric entropy of a Hölder space

The aim of this appendix is to establish the bound (17) used in the covering argument in the proof of Theorem 1, which amounts to studying the metric entropy of a Hölder space and of some related space of probability measures.

In the notation introduced in Sections 1.1 and 1.2, it follows from Ascoli's theorem that the closed ball $\mathcal{B}_R^\alpha := \mathcal{B}_R^\alpha([0,T],\mathbb{R}^d) = \{f \in \mathcal{C}^\alpha; \|f\|_\alpha \leq R\}$ of center 0 and radius R in $\mathcal{C}^\alpha$ is a compact metric space for the metric defined by the uniform norm. Here we estimate by how many balls of given radius $r < R$ and centered in $\mathcal{B}_R^\alpha$ the compact metric space $\mathcal{B}_R^\alpha$ can be covered. We note that for $r \geq R$ the sole ball $\{f \in \mathcal{B}_R^\alpha; \|f\|_\infty \leq r\}$ covers $\mathcal{B}_R^\alpha$.

Notation: Given $r > 0$, the *covering number* $\mathcal{N}(E, r)$ of a compact metric space (E, d) is defined as the infimum of the integers n such that E can be covered by n balls centered in E and of radius r in d metric. Then we have the following result which gives some lower and upper bounds on the covering number $\mathcal{N}(\mathcal{B}_R^\alpha, r)$ and in our case makes more precise the bounds given for instance in [15] or [20]:

Theorem 8. *Given some integer number $d \geq 1$, some positive numbers T, R, r and α with $r < R$ and $\alpha \leq 1$, the covering number $\mathcal{N}(\mathcal{B}_R^\alpha, r)$ of $\mathcal{B}_R^\alpha$, equipped with the uniform norm, satisfies*

$$\mathcal{N}(\mathcal{B}_R^\alpha, r) \leq \Big(10\,\sqrt{d}\,\frac{R}{r}\Big)^d\, 3^{5^{\frac{1}{\alpha}}\, d^{1+\frac{1}{2\alpha}}\, T\,(\frac{R}{r})^{\frac{1}{\alpha}}}.$$

If moreover, for instance, $r \leq \dfrac{T^\alpha}{4\,T^\alpha + 4} R$, then

$$\mathcal{N}(\mathcal{B}_R^\alpha, r) \;\geq\; \Big(\frac{\sqrt{d}}{4}\,\frac{R}{r}\Big)^d\, 2^{2^{-\frac{1}{\alpha}}\, d^{1+\frac{1}{2\alpha}}\, T\,(\frac{R}{r})^{\frac{1}{\alpha}}}.$$

The lower bound ensures that the upper bound, from which depends the condition on the size of the sample in Theorems 1 and hence 2, has the good order of growth in R/r.

Proof. **1.** We start by establishing the **upper bound**.

1.1. We first consider the case when $d = 1$.

Given J and K some integers larger or equal to 1, we let $\tau = \dfrac{T}{J}$ and $\eta = \dfrac{R}{K}$, and then

$$\begin{array}{rclll} t_j &=& (j - \frac{1}{2})\,\tau\,, & j \in \mathbb{N}\,, & 1 \;\leq j \leq J, \\ y_k &=& (k - \frac{1}{2})\,\eta\,, & k \in \mathbb{N}\,, & -K+1 \;\leq k \leq K. \end{array}$$

Then we cover the rectangle $[0, T] \times [-R, +R]$ in $\mathbb{R}_t \times \mathbb{R}_y$, which contains the graph of all functions in $\mathcal{B}_R^\alpha([0, T], \mathbb{R})$, by a lattice with step τ in $t-$axis and η in $y-$axis.

Then let f be a given function in $\mathcal{B}_R^\alpha([0, T], \mathbb{R})$. Since the intervals $[y_k - \dfrac{\eta}{2}\,, y_k + \dfrac{\eta}{2}]$ cover the interval $[-R, +R]$, for every integer $j \in [1, J]$ there exists some integer $k(j) \in [-K+1, +K]$ such that

$$|f(t_j) - y_{k(j)}| \leq \frac{\eta}{2}.$$

In particular

$$|y_{k(j+1)} - y_{k(j)}| \leq \frac{\eta}{2} + |f(t_{j+1}) - f(t_j)| + \frac{\eta}{2} \leq \eta + R\,|t_{j+1} - t_j|^\alpha \leq \eta + R\,\tau^\alpha < 2\eta$$

if we suppose $KT^\alpha < J^\alpha$. But since the y_k take values which are regularly distant of η, it follows that more precisely

$$|y_{k(j+1)} - y_{k(j)}| \leq \eta.$$

From this map $k : [1, J] \cap \mathbb{N} \to [-K+1, K] \cap \mathbb{N}$, we define the function $f_k : [0, T] \to [-R, +R]$ affine on each interval of the subdivision $(0, t_1, \cdots, t_J, T)$ and such that

$$\begin{array}{rcl} f_k(0) &=& f_k(t_1), \\ f_k(t_j) &=& y_{k(j)}, \quad 1 \leq j \leq J \\ f_k(T) &=& f_k(t_J). \end{array}$$

In particular we note that this function f_k is Lipschitz with

$$\sup_{0\leq t,s\leq T} \frac{|f_k(t)-f_k(s)|}{|t-s|} = \sup_{1\leq k\leq K} \frac{|y_{k(j+1)}-y_{k(j)}|}{|t_{j+1}-t_j|} \leq \frac{\eta}{\tau}$$

but that it does not necessarily belong to $\mathcal{B}^\alpha_R([0,T],\mathbb{R})$.

The number of such functions f_k is bounded by the number of J-uples $(y_{k(j)})_{1\leq j\leq J}$ such that $|y_{k(j+1)}-y_{k(j)}|\leq \eta$ for $1\leq j\leq J-1$, that is the number of J-uples $(k(j))_{1\leq j\leq J}$ such that $|k(j+1)-k(j)|\leq 1$ for $1\leq j\leq J-1$. Such J-uples are obtained by choosing $k(1)$ among $2K$ values, then $k(2)$ among 3 values for $-K+2\leq k(1)\leq +K-1$ or 2 values for $k(1)=-K+1$ and $+K$, and so on. Hence there exist at most $2\,K\,3^{J-1}$ such functions f_k.

If we now let K be the smallest integer larger or equal to $4\,\dfrac{R}{r}$ and J such that $KT^\alpha < J^\alpha$, then

$$\|f-f_k\|_\infty \leq \frac{r}{2}\cdot$$

Indeed, given t in $[0,T]$, there exists some integer number j in $[1,J]$ such that t belongs to $[t_j-\dfrac{\tau}{2}\,, t_j+\dfrac{\tau}{2}]$, so that

$$\begin{aligned}|f(t)-f_k(t)| &\leq |f(t)-f(t_j)|+|f(t_j)-f_k(t_j)|+|f_k(t_j)-f_k(t)|\\ &\leq R\,|t-t_j|^\alpha+|f(t_j)-y_{j(k)}|+\frac{\eta}{\tau}\,|t_j-t| \leq R\Big(\frac{\tau}{2}\Big)^\alpha+\frac{\eta}{2}+\frac{\eta}{\tau}\frac{\tau}{2}\leq 2\eta\leq\frac{r}{2}\cdot\end{aligned}$$

Hence we can cover $\mathcal{B}^\alpha_R([0,T],\mathbb{R})$ by less than $2\,K\,3^{J-1}$ balls of radius $\dfrac{r}{2}$ of the metric space $\mathcal{C}([0,T],\mathbb{R})$ equipped with the uniform norm, and if we let J and K be the smallest integers larger or equal to $5^{\frac{1}{\alpha}}\,T\,(\dfrac{R}{r})^{\frac{1}{\alpha}}$ and $4\,\dfrac{R}{r}$ respectively, then $KT^\alpha < J^\alpha$ holds true and

$$2\,K\,3^{J-1}\leq 10\,\frac{R}{r}\,3^{5^{\frac{1}{\alpha}}\,T\,(\frac{R}{r})^{\frac{1}{\alpha}}}.$$

1.2. From this we now deduce the upper bound in the general case $d\geq 1$.

Let F be a given function in $\mathcal{B}^\alpha_R([0,T],\mathbb{R}^d)$ with components $F_i\in\mathcal{B}^\alpha_R([0,T],\mathbb{R})$ for $1\leq i\leq d$. Let now J and K be the smallest integers larger or equal to $5^{\frac{1}{\alpha}}\,T\,(\sqrt{d}\,\dfrac{R}{r})^{\frac{1}{\alpha}}$ and $4\,\sqrt{d}\,\dfrac{R}{r}$ respectively. With each i, we associate an integer k_i in $[1, 2\,K\,3^{J-1}]$ such that

$$\|F_i-f_{k_i}\|_\infty\leq\frac{r}{2\,\sqrt{d}}$$

where the f_k are the functions in $\mathcal{C}([0,T],\mathbb{R})$ defined in the first step (relatively to $\dfrac{r}{\sqrt{d}}$ instead of r).

Then the function $F_{k_1,\cdots,k_d}$ with components f_{k_i} for $1\leq i\leq d$ belongs to $\mathcal{C}([0,T],\mathbb{R}^d)$ and satisfies $\|F-F_{k_1,\cdots,k_d}\|_\infty\leq\dfrac{r}{2}\cdot$ Moreover there are at most $(2\,K\,3^{J-1})^d$ such functions $F_{k_1,\cdots,k_d}$.

Consequently we can cover $\mathcal{B}_R^\alpha([0,R],\mathbb{R}^d)$ by less than $(2\,K\,3^{J-1})^d$ balls of radius $\frac{r}{2}$ of the metric space $\mathcal{C}([0,T],\mathbb{R}^d)$ equipped with the uniform norm, whence by less than $\big(2\,K\,3^{J-1}\big)^d$ balls of radius r of the metric space $\mathcal{B}_R^\alpha([0,T],\mathbb{R}^d)$ equipped with the uniform norm.

This concludes the proof of the upper bound of the covering number $\mathcal{N}(\mathcal{B}_R^\alpha([0,T],\mathbb{R}^d),r)$.

2. We now turn to the **lower bound**.

2.1. We first consider the case $d=1$.

We can give different types of lower bounds by considering special functions of the type f_k defined in the first step. Here, for instance, we give the detail for one of them.

Given some non-zero integer J, we let $\tau=\dfrac{T}{J}$ and $\eta=\tau^\alpha\,R$, and then

$$\begin{array}{rclll} t_j &=& (j-\frac{1}{2})\,\tau\,, & j\in\mathbb{N}\,, & 1\ \leq j\leq J,\\ y_k &=& (k-\frac{1}{2})\,\eta\,, & k\in\mathbb{N}\,, & -\tau^{-\alpha}+\frac{1}{2}\ \leq k\leq \tau^{-\alpha}+\frac{1}{2}. \end{array}$$

From a map $k:[1,J]\cap\mathbb{N}\to[0,1]\cap\mathbb{N}$, we define as above the function $f_k:[0,T]\to[y_0,y_1]$ affine on every interval of the subdivision $(0,t_1,\cdots,t_J,T)$ and such that

$$\begin{array}{rcll} f_k(0) &=& f_k(t_1) & \\ f_k(t_j) &=& y_{k(j)}, & 1\leq j\leq J\\ f_k(T) &=& f_k(t_J). & \end{array}$$

Given some integer ℓ such that $-\tau^{-\alpha}+\dfrac{1}{2}\ \leq\ \ell\ \leq\ \tau^{-\alpha}-\dfrac{1}{2}$, we define the function $f_{k\ell}:[0,T]\to[y_\ell,y_{\ell+1}]$ such that

$$f_{k\ell}(t)=f_k(t)+\ell\,\eta.$$

Then $f_{k\ell}$ belongs to $\mathcal{B}_R^\alpha([0,T],\mathbb{R})$ and $\|f_{k\ell}-f_{k'\ell'}\|_\infty\geq\eta$ if $f_{k\ell}\neq f_{k'\ell'}$.

If for instance $r<\inf(R,2^{-1}T^\alpha R)$ and $J+1$ is the smallest integer larger or equal to $2^{-\frac{1}{\alpha}}T\,(\dfrac{R}{r})^{\frac{1}{\alpha}}$, then $\|f_{k\ell}-f_{k'\ell'}\|_\infty>2r$ if $f_{k\ell}\neq f_{k'\ell'}$.

Thus we have found $L\,2^J$ elements in $\mathcal{B}_R^\alpha([0,T],\mathbb{R})$ mutually distant of at least $2r$ in uniform norm, where L is the number of integers ℓ between $-\tau^{-\alpha}+\dfrac{1}{2}$ and $\tau^{-\alpha}-\dfrac{1}{2}\cdot$ Hence

$$\mathcal{N}(\mathcal{B}_R^\alpha([0,T],\mathbb{R}),r)\ \geq\ L\,2^J.$$

But

$$L\ >\ 2\big((\tau^{-\alpha}-\frac{1}{2})-1\big)+1\ =\ 2\tau^{-\alpha}-2\ \geq\ \big((\frac{R}{r})^{\frac{1}{\alpha}}-\frac{2^{\frac{1}{\alpha}}}{T}\big)^\alpha-2\ \geq\ \frac{R}{r}-\frac{2}{T^\alpha}-2.$$

If moreover, for instance, $r\leq\dfrac{T^\alpha}{4\,T^\alpha+4}R$, then $L\geq\dfrac{R}{2r}$ and

$$\mathcal{N}(\mathcal{B}_R^\alpha([0,T],\mathbb{R}),r)\ \geq\ \frac{1}{4}\,\frac{R}{r}\,2^{2^{-\frac{1}{\alpha}}\,T\,(\frac{R}{r})^{\frac{1}{\alpha}}}.$$

2.2. From this we now deduce the lower bound in the general case $d \geq 1$.

The $L^d\, 2^{dJ}$ functions $F_{k_1\ell_1,\cdots,k_d\ell_d}$ with components $f_{k_j\ell_j}$ for $j = 1, \cdots, d$ where $f_{k_j\ell_j}$ have been defined in the step 1, belong to $\mathcal{B}^\alpha_R([0,T], \mathbb{R}^d)$ and are mutually distant of at least $2\sqrt{d}\, r$.

This concludes the argument for the lower bound of the number $\mathcal{N}(\mathcal{B}^\alpha_R([0,T], \mathbb{R}^d), r)$. $\square$

We now turn to the covering number of the corresponding space of probability measures: given a Polish metric space (E, d), $p \geq 1$ and $\delta > 0$, we denote $\mathcal{N}_p(\mathcal{P}(E), \delta)$ the covering number of $\mathcal{P}(E)$ for the W_p distance.

Then we have the following general result which is proven in [6] (see also [10], [14]):

Theorem 9. *Let (E, d) be a Polish metric space with finite diameter D, and let p and δ be some real numbers with $p \geq 1$ and $0 < \delta < D$. Then the covering number $\mathcal{N}_p(\mathcal{P}(E), \delta)$ of $\mathcal{P}(E)$ satisfies*

$$\mathcal{N}_p(\mathcal{P}(E), \delta) \leq \big(8\, e\, \frac{D}{\delta}\big)^{p\,\mathcal{N}(E, \frac{\delta}{2})}.$$

Note that if $\delta \geq D$ we simply have $\mathcal{N}_p(\mathcal{P}(E), \delta) = 1$ since the Wasserstein distance between any two probability measures on E is at most D.

Since $\mathcal{B}^\alpha_R$ equipped with the metric defined by the uniform norm is a Polish metric space with finite diameter $2R$, we deduce the following result:

Theorem 10. *Let $d \geq 1$, p, T, R, δ and α be some positive numbers with $p \geq 1$, $\delta < 2R$ and $\alpha \leq 1$. Let also $\mathcal{B}^\alpha_R = \{f \in \mathcal{C}^\alpha; \|f\|_\alpha \leq R\}$ be equipped with the uniform norm. Then the space $\mathcal{P}(\mathcal{B}^\alpha_R)$ of probability measures on $\mathcal{B}^\alpha_R$ can be covered by $\mathcal{N}_p(\mathcal{P}(\mathcal{B}^\alpha_R), \delta)$ balls of radius δ in Wasserstein distance W_p, with*

$$\mathcal{N}_p(\mathcal{P}(\mathcal{B}^\alpha_R), \delta) \leq \big(16\, e\, R\delta^{-1}\big)^{p\,(20\sqrt{d}\,R\,\delta^{-1})^d\, 3^{10^{\frac{1}{\alpha}}\, d^{1+\frac{1}{2\alpha}}\, T\,(R\delta^{-1})^{\frac{1}{\alpha}}}}.$$

For $\delta \geq 2R$, we have

$$\mathcal{N}_p(\mathcal{P}(\mathcal{B}^\alpha_R), \delta) = 1.$$

Acknowledgments. The author thanks Professors M. Ledoux for his kind interest in this work, in particular pointing out Reference [12], and A. Guillin and C. Villani for stimulating discussions during the preparation of [6].

References

[1] S. Benachour, B. Roynette, D. Talay and P. Vallois. Nonlinear self-stabilizing processes. I. Existence, invariant probability, propagation of chaos. *Stoch. Process. Appl. 75*, 2 (1998), 173–201.

[2] S. Benachour, B. Roynette and P. Vallois. Nonlinear self-stabilizing processes. II. Convergence to invariant probability. *Stoch. Process. Appl. 75*, 2 (1998), 203–224.

[3] D. Benedetto, E. Caglioti, J. A. Carrillo and M. Pulvirenti. A non-Maxwellian steady distribution for one-dimensional granular media. *J. Stat. Phys. 91*, 5-6 (1998), 979–990.

[4] S. Bobkov and F. Götze. Exponential integrability and transportation cost related to logarithmic Sobolev inequalities. *J. Funct. Anal. 163* (1999), 1–28.

[5] F. Bolley. Separability and Completeness for the Wasserstein distance. Report available online via `http://www.lsp.ups-tlse.fr/Fp/Bolley`, 2005.

[6] F. Bolley, A. Guillin and C. Villani. Quantitative concentration inequalities for empirical measures on non-compact spaces. To appear in *Prob. Theo. Rel. Fields*. Preprint available online via `http://www.lsp.ups-tlse.fr/Fp/Bolley`, 2005.

[7] F. Bolley and C. Villani. Weighted Csiszár-Kullback-Pinsker inequalities and applications to transportation inequalities. *Ann. Fac. Sci. Toulouse 6,* 14, 3 (2005), 331–352.

[8] J. A. Carrillo, R. J. McCann and C. Villani. Kinetic equilibration rates for granular media and related equations: entropy dissipation and mass transportation estimates. *Rev. Mat. Iberoamericana 19*, 3 (2003), 971–1018.

[9] J. A. Carrillo, R. J. McCann and C. Villani. Contractions in the 2-Wasserstein length space and thermalization of granular media. To appear in *Arch. Rat. Mech. Anal.* (2005).

[10] A. Dembo and O. Zeitouni. *Large deviation techniques and applications.* Springer, New York, 1998.

[11] H. Djellout, A. Guillin and L. Wu. Transportation cost-information inequalities and applications to random dynamical systems and diffusions. *Ann. Prob. 32*, 3B (2004), 2702–2732.

[12] X. Fernique. Régularité des trajectoires des fonctions aléatoires gaussiennes. In *Ecole d'Eté de Probabilités de Saint-Flour 1974.* Lecture Notes in Math. 480, Springer, Berlin, 1975.

[13] N. Gozlan. Principe conditionnel de Gibbs pour des contraintes fines approchées et inégalités de transport. Thèse de doctorat de l'Université de Paris 10 Nanterre, 2005.

[14] S. R. Kulkarni and O. Zeitouni. A general classification rule for probability measures. *Ann. Stat. 23*, 4 (1995), 1393–1407.

[15] G. G. Lorentz. *Approximation of functions.* Holt, Rinehart and Winston, New York, 1966.

[16] F. Malrieu. Logarithmic Sobolev inequalities for some nonlinear PDE's. *Stoch. Process. Appl. 95*, 1 (2001), 109–132.

[17] S. Méléard. Asymptotic behaviour of some interacting particle systems; McKean-Vlasov and Boltzmann models. In *Probabilistic models for nonlinear partial differential equations, Montecatini Terme, 1995.* Lecture Notes in Math. 1627, Springer, Berlin, 1996.

[18] A.-S. Sznitman. Topics in propagation of chaos. In *Ecole d'Eté de Probabilités de Saint-Flour 1989*, Lecture Notes in Math. 1464, Springer, Berlin, 1991.

[19] H. Tanaka. Limit theorems for certain diffusion processes with interaction. In *Stochastic analysis, Kataka/Tokyo, 1982.* North Holland, Amsterdam, 1984.

[20] A. van der Vaart and J. Wellner. *Weak convergence and empirical processes.* Springer, Berlin, 1995.

[21] C. Villani. *Topics in optimal transportation.* Grad. Stud. Math. 58, AMS, Providence, 2003.

ENS Lyon, UMPA (UMR 5669), 46 allée d'Italie, F-69364 Lyon Cedex 07

Current address: Institut de mathématiques - LSP (UMR C5583), Université Paul-Sabatier, Route de Narbonne, F-31062 Toulouse cedex 4

E-mail address: `bolley@cict.fr`