QUANTITATIVE CONCENTRATION INEQUALITIES ON SAMPLE PATH SPACE FOR MEAN FIELD INTERACTION

FRANÇOIS BOLLEY

Abstract. We consider the approximation of a mean field stochastic process by a large interacting particle system. We derive non-asymptotic large deviation bounds measuring the concentration of the empirical measure of the paths of the particles around the law of the process. The method is based on a coupling argument, strong integrability estimates on the paths in Hölder norm, and a general concentration result for the empirical measure of identically distributed independent paths.

Contents

Introduction 1
1. Statement of the results 2
2. Coupling 8
3. A preliminary result on independent variables 10
4. Integrability in Hölder norm 15
5. An example of application 18
Appendix. Metric entropy of a Hölder space 21
References 25

Introduction

This paper is devoted to the study of the particle approximation of a mean field stochastic process. In the models to be considered, the evolution is governed by a random diffusive term, an exterior force field and a mean field interaction depending on the law of the process itself. The particle approximation of such processes has been 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 question in the setting of large deviations.

This follows some works addressing this issue at the level of the time marginals, that we now summarize. For this purpose let μ_t be the law of the considered process at time t, and $(X^i_t)_{1\leq i\leq N}$ be the position of the N particles in the phase space $\mathbb{R}^d$. As regards Lipschitz observables, F. Malrieu [15] adapted concentration of measure ideas to obtain bounds like

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

Key words and phrases. Mean field limits, Particle approximation, Transportation inequalities.

Université Paris-Dauphine, Ceremade, UMR CNRS 7534, Place du Maréchal de Lattre de Tassigny, F-75775 Paris cedex 16.

E-mail address: bolley@ceremade.dauphine.fr.

here C and λ are explicit constants independent of ε and N and $[\,.\,]_1$ 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 denote 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 the law μ_t in a stronger way, namely, at the very level of the measures. For this, transposing Sanov's large deviation argument to their setting, the authors in [4] got precise 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 go one step further again by considering the law $\mu_{[0,T]}$ of the paths of the process on a given time interval $[0,T]$: this is a 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 $[0,T]$. A natural object to consider in the particle approximation is the empirical measure of the trajectories $(X_t^i)_{0\leq t\leq T}$; it 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 $(X_t^i)_{0\leq t\leq T}$.

We shall give a precise meaning to the convergence of $\hat{\mu}_{[0,T]}$ to $\mu_{[0,T]}$, and estimates which are the analogue of (2) in the path space; we shall see that they imply (2) by projection at time t, but above all they give concentration estimates at the level of the trajectories.

In the first 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

We are interested in the particle approximation of the $\mathbb{R}^d$-valued process $(X_t)_{t\geq 0}$ evolving according to the mean field stochastic differential equation

$$dX_t = \sigma\, dB_t - b(X_t)\, dt - c * \mu_t(X_t)\, dt. \tag{3}$$

Here σ is a $d \times d$ real matrix, $(B_t)_{t\geq 0}$ a standard Brownian motion on $\mathbb{R}^d$, b and c are $\mathbb{R}^d$ to $\mathbb{R}^d$ maps, $*$ stands for the convolution and μ_t is the law on $\mathbb{R}^d$ of the random variable X_t.

Two instances of such processes are particularly interesting. First of all, when $\mathbb{R}^d$ is the phase space of positions $x \in \mathbb{R}^{d'}$ and velocities $v \in \mathbb{R}^{d'}$ with $d = 2d'$, one is interested in the process $(X_t)_{t\geq 0} = ((x_t, v_t))_{t\geq 0}$ solution to the diffusive Newton's equations

$$\begin{cases} dx_t &= v_t\, dt \\ dv_t &= \sqrt{2}\, db_t - \lambda\, v_t\, dt - \nabla_x U *_x \rho_t(x_t)\, dt. \end{cases} \tag{4}$$

Here $(b_t)_{t\geq 0}$ is a Brownian motion in the velocity space $\mathbb{R}^{d'}$, $U = U(x)$ is an interaction potential in the position space and ρ_t is the law of x_t on $\mathbb{R}^{d'}$; moreover ∇_x and $*_x$ respectively stand for the gradient and convolution with respect to the position variable $x \in \mathbb{R}^{d'}$. By Itô's formula the distribution μ_t of X_t is solution to the Vlasov-Fokker-Planck equation

$$\frac{\partial \mu_t}{\partial t} + v \cdot \nabla_x \mu_t - \big(\nabla_x U *_x \rho_t\big) \cdot \nabla_v \mu_t = \Delta_v \mu_t + \lambda \nabla_v \cdot (v \mu_t), \qquad t > 0,\ x, v \in \mathbb{R}^{d'}.$$

Here $a \cdot b$ denotes the scalar product of two vectors a and b in $\mathbb{R}^{d'}$, whereas ∇_v, $\nabla_v \cdot$ and Δ_v respectively stand for the gradient, divergence and Laplace operators with respect to the velocity variable $v \in \mathbb{R}^{d'}$. This equation is used in the modelling of diffusive stellar matter.

We are also concerned with the process $(X_t)_{t\geq 0} = ((x_t, v_t))_{t\geq 0}$ solution to

$$\begin{cases} dx_t &= v_t\, dt \\ dv_t &= \sqrt{2}\, db_t - \nabla_v V(v_t)\, dt - \nabla_v W * \nu_t(v_t)\, dt \end{cases} \tag{5}$$

where V and W are respectively exterior and interaction potentials in the velocity space and ν_t is the law of v_t on $\mathbb{R}^{d'}$. By Itô's formula, the distribution μ_t of X_t is solution to

$$\frac{\partial \mu_t}{\partial t} + v \cdot \nabla_x \mu_t = \Delta_v \mu_t + \nabla_v \cdot (\mu_t (\nabla_v V + \nabla_v W *_v \mu_t)), \qquad t > 0,\ x, v \in \mathbb{R}^{d'}.$$

This equation is used in the modelling of granular media.

For position homogeneous distributions, we are brought to study the solution $(v_t)_{t\geq 0}$ to

$$dv_t = \sqrt{2}\, db_t - \nabla_v V(v_t)\, dt - \nabla_v W *_v \mu_t(v_t)\, dt \tag{6}$$

in $\mathbb{R}^d$ with $d = d'$; here μ_t is the law of v_t and is solution to the McKean-Vlasov equation

$$\frac{\partial \mu_t}{\partial t} = \Delta_v \mu_t + \nabla_v \cdot (\mu_t (\nabla_v V + \nabla_v W *_v \mu_t)), \qquad t > 0,\ v \in \mathbb{R}^{d'}. \tag{7}$$

In the general formulation (3), X_t can be seen as the position in the phase space of a particle initially chosen at random in a physical system according to its initial distribution μ_0. The influence of the system on the particle evolution is given by averaging the drift term $c(y - X_t)$, which measures the influence on X_t of the particles located at y, over the whole system according to the distribution μ_t. In this sense $(X_t)_{t\geq 0}$ is a mean field process.

The *particle approximation* of such a process consists in introducing N processes $(X_t^i)_{t\geq 0}$, with $1 \leq i \leq N$, which evolve no more according to the distribution μ_t of the physical system, but according to its discrete counterpart, namely the empirical measure

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

of the particle system $(X_t^1, \dots, X_t^N)$. In other words we let the processes $(X_t^i)_{t \geq 0}$ solve

$$dX_t^i = \sigma \, dB_t^i - b(X_t^i) dt - \frac{1}{N} \sum_{j=1}^N c(X_t^i - X_t^j) dt, \qquad 1 \leq i \leq N. \tag{8}$$

Here the $(B_t^i)_{t \geq 0}$'s are N independent standard Brownian motions on $\mathbb{R}^d$.

Under regularity and growth assumptions on σ, b and c, and if the particles are initially distributed in a chaotic way, for instance as independent and identically distributed variables, then $\hat{\mu}_t^N$ indeed converges as N tends to infinity to the distribution μ_t of X_t. The convergence of $\hat{\mu}_t^N$ is strongly linked with the phenomenon of *propagation of chaos* for the N interacting particles $(X_t^i)_{t \geq 0}$, as we shall see below more in detail, and both issues have been studied in [1], [7], [16] and [17] for instance. Then quantitative estimates on this convergence have been obtained in [15] at the level of observables, and at the very level of the law in [4].

In this work we go one step further and give precise estimates on the approximation of the law $\mu_{[0,T]}$ of the *path* $X = (X_t)_{0 \leq t \leq T}$ on a time interval $[0, T]$ by the empirical measure

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

of the N trajectories $X^i = (X_t^i)_{0 \leq t \leq T}$. The convergence of $\hat{\mu}_{[0,T]}^N$ has been proved in works mentionned above, and here we extend to this new setting the techniques developed in [4].

We shall assume that b and c are Lipschitz on $\mathbb{R}^d$. Then global existence and uniqueness, pathwise and in law, of the solutions to (3) and (8) are proven in [16] for instance for square integrable initial data; moreover the paths are continuous (in time).

To state our main theorem on the particle approximation, we first give some notation. If (S, d) is a separable and complete metric space, and p is a real number ≥ 1, the *Wasserstein distance* of order p between two Borel probability measures μ and ν on S is

$$W_p(\mu, \nu) := \inf_{X,Y} \left(\mathbb{E} \, d(X, Y)^p \right)^{1/p}$$

where X and Y are S-valued random variables with respective law μ and ν. W_p induces a metric on the set of Borel probability measures on S with moment $\int_S d(x_0, x)^p \, d\mu(x)$ finite for some (and thus any) x_0 in S; convergence in this metric is equivalent to narrow convergence plus some tightness condition on the moments (see for instance [19]).

In this work (S,d) will be the space $\mathcal{C} := \mathcal{C}([0,T],\mathbb{R}^d)$ 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 $W_{p,[0,T]}$.

Theorem 1. *Let μ_0 be a probability measure on $\mathbb{R}^d$ such that $\int_{\mathbb{R}^d} e^{a_0|x|^2}\, d\mu_0(x)$ be finite for some $a_0>0$, and let b and c be Lipschitz functions on $\mathbb{R}^d$. Given $T\geq 0$, let $\mu_{[0,T]}$ be the law of the solution to (3) on $[0,T]$ for some initial value distributed according to μ_0. Let also $(X_0^i)_{1\leq i\leq N}$ be N independent random variables with common law μ_0 and $\hat{\mu}^N_{[0,T]}$ be the empirical measure of the solutions to (8) on $[0,T]$, with respective initial value X_0^i.*

Then, for any $\alpha\in(0,1/2)$, there exist 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 $N\geq N_0\,\varepsilon^{-2}\exp\left(N_0\,\varepsilon^{-1/\alpha}\right)$.

The constants K and N_0 depend on T,b,c,α and a finite square exponential moment of μ_0.

Kantorovich-Rubinstein dual formulation of the W_1 distance on a general space (S,d) reads

$$W_1(\mu,\nu) = \sup_{[\varphi]_1\leq 1}\left\{\int_S \varphi\, d\mu - \int_S \varphi\, d\nu\right\} \tag{9}$$

where $[\varphi]_1 := \sup_{x\neq y}\dfrac{|\varphi(x)-\varphi(y)|}{d(x,y)}\cdot$ Then the bound in Theorem 1 can be written as

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

where the supremum runs over all 1-Lipschitz functions φ on $\mathcal{C}$. By projection at time t, it implies the concentration inequalities

$$\mathbb{P}\left[\sup_{[\varphi]_1\leq 1}\left\{\frac{1}{N}\sum_{i=1}^N \varphi(X_t^i) - \int_{\mathbb{R}^d} \varphi\, d\mu_t\right\} > \varepsilon\right] \leq e^{-\lambda N\varepsilon^2} \tag{10}$$

given in [4] for the time marginals, provided $N\geq N_0\,\varepsilon^{-(d+2)}$ for some N_0; here the supremum runs over all 1-Lipschitz functions φ on $\mathbb{R}^d$. But above all it gives error bounds in the approximation by $\frac{1}{N}\sum_{i=1}^N \varphi(X^i)$ of the expectation of quantities $\varphi(X)$ which depend on the whole path X. In return we impose a stronger condition on the required size of the sample.

An example is the distance $d(X, A) = \inf\{|X_t - y|; t \in [0, T], y \in A\}$ of the trajectory to a given set A in $\mathbb{R}^d$, which measures how close X_t has been to A; under the assumptions of Theorem 1, for any $\alpha \in (0, 1/2)$ there exist constants K and N_0 such that

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

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

A more involved example of such error bounds will be discussed in detail in Section 5.

As pointed out in [17], the convergence of the empirical measure $\hat{\mu}^N_{[0,T]}$ towards the distribution $\mu_{[0,T]}$ is strongly linked with the phenomenon of propagation of chaos, namely, that the interacting particles X^i tend to behave like independent variables with law $\mu_{[0,T]}$, as N goes to infinity. 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 of pairs of paths, the asymptotic independence of two paths (among N) can be estimated as

Theorem 2. *With the same notation and assumptions as in Theorem 1, for all $T \geq 0$ and $\alpha \in (0, 1/2)$ there exist 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, b, c, 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 $\mathcal{C} \times \mathcal{C}$.

In turn this leads to error bounds in the approximation of functions of the paths of two independent solutions to (3). Let for instance $(X_t)_{t \geq 0}$ and $(\tilde{X}_t)_{t \geq 0}$ be the solutions to (3) respectively driven by two independent Brownian motions B and $\tilde{B}$ and with independent initial data. Then, in average, the minimal distance $d(X, \tilde{X}) = \inf\{|X_t - \tilde{X}_t|; t \in [0, T]\}$ between X_t and $\tilde{X}_t$ on $[0, T]$ is approximated by $\displaystyle\frac{1}{N(N-1)} \sum_{i \neq j} d(X^i, X^j)$ with an error controlled by

$$\mathbb{P}\Big[\big|\mathbb{E}[d(X, \tilde{X})] - \frac{1}{N(N-1)} \sum_{i \neq j} d(X^i, X^j)\big| > \varepsilon\Big] \leq e^{-KN\varepsilon^2}$$

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

The proof of Theorem 1 is based on this phenomenon of propagation of chaos: as N goes to infinity, the interacting particles $(X^i_t)_{t \geq 0}$ tend to behave like the N independent and identically distributed processes $(Y_t^i)_{t\geq 0}$ solution to

$$\left\{ \begin{array}{rcl} dY_t^i & = & \sigma\, dB_t^i - b(Y_t^i)\, dt - c * \mu_t(Y_t^i)\, dt \\ Y_0^i & = & X_0^i \end{array} \right. \qquad 1 \leq i \leq N\,. \tag{11}$$

Here μ_t is the law of X_t, but is also the law of any Y_t^i and, for each i, $(B_t^i)_{t\geq 0}$ is the Brownian motion driving the evolution of $(X_t^i)_{t\geq 0}$. Then the paths $(Y_t^i)_{t\geq 0}$ are close to the paths $(X_t^i)_{t\geq 0}$ and Proposition 3 ensures the existence of a constant C (depending only on b, c and T) such that

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

holds almost surely, where $\hat{\nu}^N_{[0,T]} := \displaystyle\frac{1}{N}\sum_{i=1}^N \delta_{Y^i}$ and $Y^i = (Y_t^i)_{0\leq t\leq T}$; 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 paths Y^i for $1 \leq i \leq N$ are independent and distributed according to $\mu_{[0,T]}$. Then Propositions 5 and 7 ensure 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 a condition on ε and N.

The proof will be given in greater detail in Section 4. It is an adaptation of the argument given in [4, Section 2.6] of estimates (10) for time marginals. The current proof turns out to be simpler since it consists in fewer steps; in return each step of the present infinite dimensional case, where $\mu_{[0,T]}$ is a measure on the functional space $\mathcal{C}$, exhibits new difficulties with respect to the finite dimensional case of [4] where μ_t is a measure on $\mathbb{R}^d$.

The proof of Theorem 2 consists in writing this coupling argument for *pairs* of paths and comparing $\mu_{[0,T]} \otimes \mu_{[0,T]}$ and $\hat{\nu}^{N,2}_{[0,T]} := \displaystyle\frac{1}{N(N-1)}\sum_{i\neq j}\delta_{(Y^i,Y^j)}$ by means of $\displaystyle\frac{1}{N^2}\sum_{i,j}\delta_{(Y^i,Y^j)}$.

Let us finally note that it would be desirable to relax the assumptions made on the drift terms b and c: first of all to include the interesting case of the cubic potential $W(z) = |z|^3/3$ on $\mathbb{R}$ in the case of equation (5), which models the interaction among one-dimensional granular media (see [2] for instance); then to treat the fundamental cases of the Coulomb and gravitational potentials in the diffusive Newton's equations (4) (see [10] for instance).

It would also be interesting to consider the whole trajectories $(X_t)_{t\geq 0}$ on $[0,+\infty)$, and derive concentration bounds on functionals such as hitting times for instance.

Before turning to the proofs we briefly give the plan of the paper. In the coming section we reduce our concentration issue on interacting particles to the same issue for independent variables by a coupling argument. In Section 3 we prove a general concentration result for $\mathcal{C}$-valued independent variables and in Section 4 we check that it applies to our framework. An example of error bounds implied by Theorem 1 is finally discussed in Section 5.

2. Coupling

Let us recall that we want to measure the distance between the law $\mu_{[0,T]}$ of the solution $X = (X_t)_{0\leq t\leq T}$ on $[0,T]$ to (3) and the empirical measure $\hat\mu^N_{[0,T]}$ of the N solutions $X^i = (X^i_t)_{0\leq t\leq T}$ on $[0,T]$ to (8).

In the following proposition we reduce this issue to measuring the distance between $\mu_{[0,T]}$ and the empirical measure

$$\hat\nu^N_{[0,T]} := \frac{1}{N}\sum_{i=1}^N \delta_{Y^i} \tag{12}$$

of the N independent solutions $Y^i = (Y^i_t)_{0\leq t\leq T}$ to (11) on $[0,T]$:

Proposition 3. *Let us assume that there exist real numbers β, γ and Γ such that*

$$(b(x)-b(y))\cdot(x-y) \geq \beta\,|x-y|^2, \quad (c(x)-c(y))\cdot(x-y) \geq \gamma\,|x-y|^2, \quad |c(x)-c(y)| \leq \Gamma|x-y|$$

for all x and y in $\mathbb{R}^d$. Then there exists a constant C depending only on β, γ, Γ and T such that

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

almost surely in the above notation.

Proof. We first follow the lines of the proof of [4] or [15] in the case when $b = \nabla V$ and $c = \nabla W$, but in the end we want an estimate on the *paths*. 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

$$d(X^i_t - Y^i_t) = -\big(b(X^i_t) - b(Y^i_t)\big)\,dt - \big(c * \hat\mu^N_t(X^i_t) - c * \mu_t(Y^i_t)\big)\,dt.$$

In particular

$$\frac{1}{2}\frac{d}{dt}|X^i_t - Y^i_t|^2 = -\,\big(b(X^i_t) - b(Y^i_t)\big)\cdot(X^i_t - Y^i_t) - \,\big(c * \hat\mu^N_t(X^i_t) - c * \mu_t(Y^i_t)\big)\cdot(X^i_t - Y^i_t). \tag{13}$$

We decompose the last term according to

$$c * \hat\mu^N_t(X^i_t) - c * \mu_t(Y^i_t) = (c * \hat\mu^N_t - c * \mu_t)(X^i_t) + \big(c * \mu_t(X^i_t) - c * \mu_t(Y^i_t)\big).$$

Then the map $c(X^i_t - \,\cdot\,)$ is Γ-Lipschitz, so the Kantorovich-Rubinstein formulation (9) of the Wasserstein distance of order 1 ensures that

$$\Big|c * (\hat\mu^N_t - \mu_t)(X^i_t)\Big| = \left|\int_{\mathbb{R}^d} c(X^i_t - y)\, d(\hat\mu^N_t - \mu_t)(y)\right| \leq \Gamma\, w_1(\hat\mu^N_t, \mu_t)$$

where w_p is the Wasserstein distance of order $p \geq 1$ between measures on $\mathbb{R}^d$ equipped with the Euclidean distance $|\cdot|$. Then (13) and the assumptions on b and c imply

$$\frac{1}{2}\frac{d}{dt}|X^i_t - Y^i_t|^2 \leq \Gamma\, w_1(\hat\mu^N_t, \mu_t)\, |X^i_t - Y^i_t| - (\beta + \gamma)\, |X^i_t - Y^i_t|^2. \tag{14}$$

In particular, by integration,

$$|X^i_s - Y^i_s| \leq \Gamma \int_0^s w_1(\hat\mu^N_u, \mu_u)\, du - (\beta + \gamma)\int_0^s |X^i_u - Y^i_u|\, du \tag{15}$$

since initially $X_0^i = Y_0^i$. But

$$w_1(\hat\mu_u^N, \mu_u) \le W_{1,[0,u]}(\hat\mu_{[0,u]}^N, \mu_{[0,u]})$$

since $\hat\mu_u^N$ and μ_u are the respective image measures of $\hat\mu_{[0,u]}^N$ and $\mu_{[0,u]}$ by the 1-Lipschitz projection π_u from $\mathcal{C}([0,u],\mathbb{R}^d)$ into $\mathbb{R}^d$ defined by $\pi_u(f) = f(u)$. Moreover

$$W_{1,[0,u]}(\hat\mu_{[0,u]}^N, \mu_{[0,u]}) \le \frac{1}{N}\sum_{i=1}^N \sup_{0\le s\le u} |X_s^i - Y_s^i| + W_{1,[0,u]}(\hat\nu_{[0,u]}^N, \mu_{[0,u]}).$$

by triangular inequality. Hence, by averaging (15) over i, and by Gronwall's lemma,

$$\frac{1}{N}\sum_{i=1}^N \sup_{0\le s\le t} |X_s^i - Y_s^i| \le \Gamma \int_0^t e^{(\Gamma+m)(t-u)} W_{1,[0,u]}(\hat\nu_{[0,u]}^N, \mu_{[0,u]})\, du \tag{16}$$

where $m := \max\big(0, -(\beta+\gamma)\big)$. On the other hand, for $0 \le u \le t$,

$$W_{1,[0,u]}(\hat\nu_{[0,u]}^N, \mu_{[0,u]}) \le W_{1,[0,t]}(\hat\nu_{[0,t]}^N, \mu_{[0,t]})$$

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

$$W_{1,[0,t]}(\hat\mu_{[0,t]}^N, \hat\nu_{[0,t]}^N) \le \frac{1}{N}\sum_{i=1}^N \sup_{0\le s\le t} |X_s^i - Y_s^i| \le \Gamma(\Gamma+m)^{-1}(e^{(\Gamma+m)T} - 1)\, W_{1,[0,t]}(\hat\nu_{[0,t]}^N, \mu_{[0,t]})$$

by (16). This concludes the argument by triangular inequality. □

Remark 4. In the case of the granular media equation (6), and under convexity assumptions on V and W, such as $\beta > 0$, $\beta + 2\gamma > 0$, it has been proven in [6], [7], [15] that the time marginal μ_t converges, as t goes to infinity, to the stationary solution to (7); one can also prove that in expectation observables of the particle system are bounded in time.

Hence, under this kind of assumptions, one can hope for uniform in time constants in this coupling argument. This was obtained in [4] for the time marginals of the granular media equation; here, for the whole processes, if $\beta+\gamma > \Gamma$, then we can let C be $(\beta+\gamma)\,(\beta+\gamma-\Gamma)^{-1}$ in Proposition 3, independently of T. Indeed, if $\beta + \gamma > 0$, then (14) leads to

$$|X_s^i - Y_s^i| \le \Gamma \int_0^s e^{-(\beta+\gamma)(s-u)}\, w_1(\hat\mu_u^N, \mu_u)\, du$$

by integration. Consequently

$$W_{1,[0,t]}(\hat\mu_{[0,t]}^N, \hat\nu_{[0,t]}^N) \le \Gamma \sup_{0\le s\le t} \int_0^s e^{-(\beta+\gamma)(s-u)}\, du \sup_{0\le u\le t} w_1(\hat\mu_u^N, \mu_u) \le \frac{\Gamma}{\beta+\gamma} W_{1,[0,t]}\,(\hat\mu_{[0,t]}^N, \mu_{[0,t]}).$$

If moreover $\beta + \gamma > \Gamma$, then by triangular inequality

$$W_{1,[0,t]}(\hat\mu_{[0,t]}^N, \mu_{[0,t]}) \le \frac{\beta+\gamma}{\beta+\gamma-\Gamma} W_{1,[0,t]}(\hat\nu_{[0,t]}^N, \mu_{[0,t]}).$$

Let us note that, contrary to [4] 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.

3. A preliminary result on independent variables

Our main theorem on the particle approximation is based on a general concentration result for the empirical measure of $\mathcal{C}$-valued independent and identically distributed random variables. In this section we state this result in a more general formulation and for this purpose we first introduce some notation.

If μ and ν are two measures on $\mathcal{C}$, the *relative entropy* of ν with respect to μ is defined by

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

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

This notion is linked with the Wasserstein distances by the family of *transportation inequalities*: given $p \geq 1$ and $\lambda > 0$, a probability measure μ on $\mathcal{C}$ 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.

Moreover, given a Borel probability measure μ on $\mathcal{C}$ and N independent random variables $(X^i)_{1 \leq i \leq N}$ with law μ, we let $\hat{\mu}^N$ denote their empirical measure, defined as

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

Given a 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 $\mathcal{C}$ 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:

Proposition 5. *Let μ be a Borel probability measure on $\mathcal{C}$ satisfying a $T_p(\lambda)$ inequality for some $\lambda > 0$ and $p \in [1,2]$, and such that $\int_{\mathcal{C}} e^{a\|x\|_\alpha^2}\, d\mu(x)$ be finite for some $a > 0$ and $\alpha \in (0,1]$. Then, for any $\alpha' < \alpha$ and $\lambda' < \lambda$, there exists a constant N_0 such that*

$$\mathbb{P}\left[W_{p,[0,T]}(\mu, \hat{\mu}^N) > \varepsilon\right] \leq e^{-\beta_p \frac{\lambda'}{2} N \varepsilon^2} \tag{17}$$

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 $\int_{\mathcal{C}} e^{a\|x\|_\alpha^2}\, d\mu(x)$.

This proposition will be applied with $p = 1$ to the distribution $\mu_{[0,T]}$ of the process X. For $p = 1$, a $T_1(\lambda)$ inequality is equivalent to the existence of $a > 0$ such that $\int_{\mathcal{C}} e^{a\,\|x\|_\infty^2}\, d\mu(x)$ be finite (see [8]). Numerical relations between such a and λ are given in [5] and [12]. In particular this condition if fulfilled if $\int_{\mathcal{C}} e^{a\,\|x\|_\alpha^2}\, d\mu(x)$ is finite, as will be the case for $\mu_{[0,T]}$.

For $p = 1$ again, a result by S. Bobkov and F. Götze [3], based on (9), ensures that a $T_1(\lambda)$ inequality for μ is equivalent to the following precise version of the central limit theorem:

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

By comparison, the bound given in Proposition 5 implies

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

by (9), but a modification of the proof would also lead to

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

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

This result is reasonable in view of Sanov's theorem (stated in [9] for instance). Indeed, by this theorem, one can hope for an upper bound like

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

for large N, where $A := \{\nu; W_{p,[0,T]}(\nu, \mu) > \varepsilon\}$. But μ satisfies $T_p(\lambda)$, so that

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

and one indeed obtains an upper bound like (17), but only asymptotically, whereas Proposition 5 gives 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, the space $\mathcal{C}$ being unbounded, 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 bound $\mathbb{P}[\hat{\mu}^N \in A] \leq \exp(-N \inf\{H(\nu|\mu);\, \nu \in \overline{A}\}) = 1$.

Finally this result can be seen as an extension of the following similar concentration result given in [4, Theorem 2.1] in the case of measures on $\mathbb{R}^d$: if m satisfies $T_p(\lambda)$, then

$$\mathbb{P}\left[w_p(m, \hat{m}^N) > \varepsilon\right] \leq e^{-\gamma_p \frac{\lambda'}{2} N \varepsilon^2}, \quad \varepsilon > 0\,, \quad N \geq N_0 \max(\varepsilon^{-(d'+2)}, 1). \tag{18}$$

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 Proposition 5 (one can take any $a < \lambda/2$), and the bound (18) follows by (17) with the constant γ_p obtained in [4]. Note however that the required size of the sample is here larger for small ε.

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 argument by optimizing the introduced parameters. Since the argument follows the lines of the proof given in [4, Section 3.1] in the finite dimensional case, in which μ is a measure on $\mathbb{R}^d$, we shall only sketch it, stressing only on the bounds specific to our new framework. We refer to [4] for further details.

1. Truncation. Given $R > 0$, to be chosen later on, we let $\mathcal{B}_R^\alpha$ denote 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 $E := \displaystyle\int_{\mathcal{C}} e^{a\|x\|_\alpha^2}\, d\mu(x)$ and a. In this step we reduce the concentration issue 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 obtain the bound

$$\mathbb{P}\left[W_{p,[0,T]}(\mu, \hat{\mu}^N) > \varepsilon\right] \leq \mathbb{P}\left[W_{p,[0,T]}(\mu_R, \hat{\mu}_R^N) > \eta\,\varepsilon - 2\, E^{\frac{1}{p}} R\, e^{-\frac{a}{p}R^2}\right] + \exp\Big(-N\big(\theta(1-\eta)^p\varepsilon^p - E\, e^{(a_1 - a)\, R^2}\big)\Big)\,; \tag{19}$$

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_1 and p.

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

$$\mathbb{P}\left[W_{2,[0,T]}(\mu, \hat{\mu}^N) > \varepsilon\right] \leq \mathbb{P}\left[W_{2,[0,T]}(\mu_R, \hat{\mu}_R^N) > \eta\,\varepsilon - 2\,E^{\frac{1}{2}} R\, e^{-\frac{a}{2}R^2}\right] + \exp\left(-N\big(\frac{a_1}{2}(1-\eta)^2\varepsilon^2 - 2E^2\, e^{(a_1-a)\,R^2}\big)\right). \quad (20)$$

2. Sanov's argument on small balls. In view of (19) for $p < 2$ or (20) for $p = 2$, we now aim at bounding $\mathbb{P}\,[\hat{\mu}_R^N \in \mathcal{A}]$ where

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

For that purpose we let $\delta > 0$ and cover $\mathcal{A}$ with $\mathcal{N}_p(\mathcal{A}, \delta)$ balls B_i 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

$$\mathbb{P}\big[\hat{\mu}_R^N \in \mathcal{A}\big] \leq \mathbb{P}\Big[\hat{\mu}_R^N \in \bigcup_{i=1}^{\mathcal{N}_p(\mathcal{A},\delta)} B_i\Big] \leq \sum_{i=1}^{\mathcal{N}_p(\mathcal{A},\delta)} \mathbb{P}\left[\hat{\mu}_R^N \in B_i\right] \leq \sum_{i=1}^{\mathcal{N}_p(\mathcal{A},\delta)} \exp\Big(-N \inf_{\nu \in B_i} H(\nu|\mu_R)\Big). \quad (21)$$

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 (21) the existence of positive constants δ_1, η_1 and K_1 such that

$$\mathbb{P}\left[W_{p,[0,T]}(\mu_R, \hat{\mu}_R^N) > \eta\,\varepsilon - 2\,E^{\frac{1}{p}} R e^{-\frac{a}{p}R^2}\right] \leq \mathcal{N}_p(\mathcal{A}, \delta)\, \exp\left(-N\Big(\frac{\lambda_2}{2}\,\varepsilon^2 - K_1 R^2 e^{-aR^2}\Big)\right) \quad (22)$$

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

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

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

Then, since $\mathcal{A}$ is a subset of $\mathcal{P}(\mathcal{B}_R^\alpha)$, Theorem 11 in the Appendix gives the bound

$$\mathcal{N}_p(\mathcal{A}, \delta_1\,\varepsilon) \leq \exp\left(K_2 (R\,\varepsilon^{-1})^d\, 3^{K_2 (R\,\varepsilon^{-1})^{1/\alpha}}\, \ln\big(\max(1, K_2\, R\,\varepsilon^{-1})\big)\right) \quad (23)$$

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

Remark 6. 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 [4], and will influence the final condition on the size N of the sample.

3. Conclusion of the argument. We first focus on the case when $p \in [1,2)$. By estimates (19), (22) and (23), and given $\lambda_2 < \lambda$ and $a_1 < a$, we obtain the existence of positive constants K_1, K_2, K_3 and R_3 depending on $E, a, a_1, \alpha, \lambda$ and λ_2 such that

$$\mathbb{P}\left[W_{p,[0,T]}(\mu,\hat{\mu}^N) > \varepsilon\right] \leq \exp\Big(-N\big(K_3\,\theta\,\varepsilon^p - K_4 e^{(a_1-a)\,R^2}\big)\Big)$$
$$+\exp\left(K_2(R\,\varepsilon^{-1})^d\,3^{K_2(R\,\varepsilon^{-1})^{1/\alpha}}\,\ln\big(\max(1,K_2\,R\,\varepsilon^{-1})\big) - N\Big(\frac{\lambda_2}{2}\,\varepsilon^2 - K_1 R^2 e^{-\alpha R^2}\Big)\right) \quad (24)$$

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

Then let $\lambda_3 < \lambda_2$. One can prove that the second term in the right-hand side in (24) is bounded by $\exp\big(-\frac{\lambda_3}{2}\,N\,\varepsilon^2\big)$ provided

$$R^2 \geq A\max(1,\varepsilon^2,\ln(\varepsilon^{-2}))\,, \quad N\varepsilon^2 \geq B\,3^{C(R\varepsilon^{-1})^{1/\alpha}} \quad (25)$$

for 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 first term in the right-hand side in (24) is bounded by $\exp\big(-\frac{\lambda_3}{2}\,N\,\varepsilon^2\big)$ as soon as $R^2 \geq R_4 \max(1, \ln(\varepsilon^{-2}))$, for a constant R_4 depending on λ_3.

Letting $R = \varepsilon\left(\frac{1}{C\,\ln 3}\ln\frac{N\varepsilon^2}{B}\right)^{\alpha}$ if $\varepsilon \in (0,1)$ and $R = \sqrt{A}\,\varepsilon$ otherwise, and $\alpha' < \alpha$, both conditions in (25) hold true as soon as $N \geq N_0\,\varepsilon^{-2}\exp(N_0\,\varepsilon^{-1/\alpha'})$ for a constant N_0 depending on $E, a, \lambda, \lambda_3, \alpha$ and α'. Finally, given $\lambda' < \lambda_3 < \lambda$, this condition ensures that

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

for a possibly 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

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

to hold (by (20)). 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

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

for any $\varepsilon > 0$ and $N \geq N_0\,\varepsilon^{-2}\,\exp(N_0\,\varepsilon^{-1/\alpha'})$. This concludes the proof of Proposition 5 in this second and last case.

4. Integrability in Hölder norm

In Section 2 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 Proposition 5 with $p = 1$, namely, that there exist $\alpha \in (0,1]$ and $a > 0$ such that

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

Proposition 7. *Let μ_0 be a probability measure on $\mathbb{R}^d$ with a finite square exponential moment and let X_0 be with law μ_0. Given $T \geq 0$, let X be the solution on $[0,T]$ to (3) starting at X_0, where b and c are Lipschitz on $\mathbb{R}^d$. 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 $\mathbb{E}\exp(a\|X\|^2_\alpha)$ be finite.*

Assuming this result for the moment we can now conclude the *proof of Theorem 1.* Let indeed α be given in $(0,1/2)$ and $\alpha_0 \in (\alpha,1/2)$. Then, by Propositions 5 and 7, applied with $\alpha = \alpha_0$ and $\alpha' = \alpha$, there exist positive constants $\tilde{K}$ and $\tilde{N}_0$, depending on α_0, α, T and a square exponential moment of μ_0, such that

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

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 (11) and (12). Then, by Proposition 3, 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

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

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. We separately prove the existence of positive constants a_1 and a_2 such that $\mathbb{E}\exp(a_1\|X\|^2_\infty)$ and $\mathbb{E}\exp(a_2[X]^2_\alpha)$ be finite, where $[\,\cdot\,]_\alpha$ stands for the Hölder seminorm defined in Section 3.

1. We start with the expectation in uniform norm.

1.1. We first prove that $\mathbb{E}|X_t|^2$ is bounded on $[0,T]$. Indeed, by Itô's formula,

$$|X_t|^2 = |X_0|^2 + 2\int_0^t X_s \cdot \sigma\, dB_s + \int_0^t tr(\sigma\sigma^*) - 2\,X_s \cdot (b(X_s) + c * \mu_s(X_s))\, ds,$$

so that

$$\mathbb{E}|X_t|^2 = \mathbb{E}|X_0|^2 + \int_0^t tr(\sigma\sigma^*) - \mathbb{E}\big[2\,X_s \cdot (b(X_s) + c * \mu_s(X_s))\big]\, ds.$$

If B is the Lipschitz seminorm of b, then for any x in $\mathbb{R}^d$ we have

$$-2\,x\,\cdot\,b(x) \leq 2B|x|^2 + 2|x\cdot b(0)| \leq (2B+1)|x|^2 + |b(0)|^2. \tag{26}$$

Furthermore, if Γ is the Lipschitz seminorm of c, then

$$-2\,x\cdot c{*}\mu_s(x) \leq 2\int_{\mathbb{R}^d} |x|(\Gamma|x-z|+|c(0)|)d\mu_s(z) \leq (3\,\Gamma+1)|x|^2+|c(0)|^2+\Gamma\int_{\mathbb{R}^d}|z|^2\,d\mu_s(z). \tag{27}$$

But $\displaystyle\int_{\mathbb{R}^d}|z|^2\,d\mu_s(z) = \mathbb{E}\,|X_s|^2$, so collecting all terms together we obtain the bound

$$\mathbb{E}|X_t|^2 \leq \mathbb{E}|X_0|^2 + \int_0^t (C + D\,\mathbb{E}|X_s|^2)\,ds$$

where $C = tr(\sigma\sigma^*) + |b(0)|^2 + |c(0)|^2$ and $D = 2B + 4\Gamma + 2$. By Gronwall's lemma this ensures the existence of a constant m such that $\mathbb{E}|X_t|^2 \leq m$ for any $t \in [0,T]$.

1.2. Then we prove the existence of $K > 0$ such that $\mathbb{E}\exp(K|X_t|^2)$ be bounded on $[0,T]$. For this purpose we let k be a smooth function on $[0,T]$, to be chosen later on, and we let $Z_t = \exp(k(t)\,|X_t|^2)$. By Itô's formula,

$$Z_t = Z_0+M_t+\int_0^t\big[k'(s)|X_s|^2+tr(\sigma\sigma^*)\,k(s)+2k(s)^2X_s\cdot\sigma\sigma^*X_s-2k(s)X_s\cdot(b(X_s)+c{*}\mu_s(X_s))\big]\,Z_s\,ds$$

where

$$M_t = 2\int_0^t k(s)\,Z_s\,X_s\cdot\sigma\,dB_s.$$

Bounds (26) and (27), with $\displaystyle\int_{\mathbb{R}^d}|z|^2\,d\mu_s(z) = \mathbb{E}\,|X_s|^2$ bounded by m on $[0,T]$, ensure that

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

where $C(s) = \Big(tr(\sigma\sigma^*)+\Gamma\,m+|b(0)|^2+|c(0)|^2\Big)\,k(s)$ and $D(s) = k'(s)+2\|\sigma\|^2\,k(s)^2+u\,k(s)$ with $u = 2B+3\Gamma+2$. Here $\|\sigma\|$ is the matrix norm of σ associated to the Euclidean norm on $\mathbb{R}^d$. We let $k(s)$ such that $D(s) \equiv 0$, namely

$$k(s) = e^{-us}\left(k(0)^{-1} + 2\|\sigma\|^2\,u^{-1}(1-e^{-us})\right)^{-1} \tag{28}$$

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

$$Z_t \leq Z_0 + M_t + C(0)\int_0^t Z_s\,ds. \tag{29}$$

In particular

$$\mathbb{E}\,Z_t \leq \mathbb{E}\,Z_0 + C(0)\int_0^T \mathbb{E}\,Z_s\,ds. \tag{30}$$

Now μ_0 is assumed to have a finite square exponential moment, so there exists $k(0)$ such that $\mathbb{E}Z_0 = \mathbb{E}\exp(k(0)|X_0|^2)$ be finite. Then (30) ensures that $\mathbb{E}\exp(k(t)\,|X_t|^2) = \mathbb{E}Z_t$ is finite and bounded on $[0,T]$ by Gronwall's lemma. Furthermore k is decreasing so $\exp(K\,|X_t|^2)$ is bounded on $[0,T]$ with $K = k(T)$.

1.3. We now prove the existence of $a_1 > 0$ such that $\mathbb{E}\exp(a_1\|X\|_\infty^2) = \mathbb{E}\sup_{0\le t\le T}\exp(a_1|X_t|^2)$ be finite. We let again $Z_t = \exp(k(t)\,|X_t|^2)$ with k given by (28), but we shall choose a new $k(0)$ later on. Then, by (29),

$$\mathbb{E}\sup_{0\le t\le T} Z_t \le \mathbb{E}\,Z_0 + \mathbb{E}\sup_{0\le t\le T} M_t + C(0)\int_0^T \mathbb{E}\,Z_s\,ds. \tag{31}$$

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

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

Then, by Itô's formula again,

$$\begin{aligned}
\mathbb{E}|M_t|^2 &= 4\int_0^t k(s)^2\,\mathbb{E}\left[Z_s^2\,|\sigma^* X_s|^2\right] ds\\
&\le 4\|\sigma\|^2\,k(0)\int_0^t \mathbb{E}\left[k(s)\,|X_s|^2\,\exp(2\,k(s)\,|X_s|^2)\right] ds\\
&\le 4\|\sigma\|^2\,k(0)\int_0^t \mathbb{E}\,\exp(3\,k(0)\,|X_s|^2)\,ds.
\end{aligned}$$

Choosing $k(0) \le K/3$ ensures that $\sup_{0\le t\le T}\mathbb{E}|M_t|^2$, whence $\mathbb{E}\sup_{0\le t\le T} M_t$, is finite.

Since, for this $k(0)$, $\sup_{0\le t\le T}\mathbb{E}\,Z_t$ also is finite, it follows from (31) that so is $\mathbb{E}\sup_{0\le t\le T} Z_t$, which proves that $\mathbb{E}\exp(a_1\|X\|_\infty^2)$ is finite with $a_1 = k(T)$.

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

$$X_t = X_0 + \sigma\,B_t - \int_0^t \big(b(X_s) + c * \mu_s(X_s)\big)\,ds$$

so that

$$[X]_\alpha \le [\sigma\,B]_\alpha + \Big[\int_0^\cdot \big(b(X_s) + c * \mu_s(X_s)\big)\,ds\Big]_\alpha$$

almost surely; here X and $\sigma\,B$ stand as before for the maps $t \mapsto X_t$ and $t \mapsto \sigma\,B_t$ respectively on $[0,T]$, and $\int_0^\cdot \varphi(s)\,ds$ is an antiderivative of φ. Hence, by Cauchy-Schwarz' inequality,

$$\mathbb{E}\exp(a_2[X]_\alpha^2) \le \big(\mathbb{E}\exp(4\,a_2\,\|\sigma\|^2\,[B]_\alpha^2)\big)^{1/2}\Big(\mathbb{E}\exp 4\,a_2\Big[\int_0^\cdot (b(X_s) + c * \mu_s(X_s))\,ds\Big]_\alpha^2\Big)^{1/2}.$$

But, on one hand, $\mathbb{E}\exp(4\,a_2\,\|\sigma\|^2\,[B]_\alpha^2)$ is finite for a_2 small enough (see [11, Theorem 1.3.2] for instance, with $E = \mathcal{C}$ and $N(f) = [f]_\alpha$). On the other hand, by step 1 and assumption on b and c, there exists a constant A such that

$$|b(x) + c * \mu_s(x)| \leq A + (B + \Gamma)|x|$$

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

$$\Big[\int_0^\cdot b(X_s)+c*\mu_s(X_s)\,ds\Big]_\alpha \leq \sup_{0\leq s<t\leq T} \frac{1}{|t-s|^\alpha}\int_s^t A+(B+\Gamma)|X_u|\,du \leq T^{1-\alpha}\big(A+(B+\Gamma)\|X\|_\infty\big)$$

almost surely, and

$$\mathbb{E}\exp 4\,a_2\Big[\int_0^\cdot \big(b(X_s)+c*\mu_s(X_s)\big)\,ds\Big]_\alpha^2 \\ \leq \exp(8\,a_2\,T^{2-2\alpha}A^2)\;\mathbb{E}\exp\big(8\,a_2\,T^{2-2\alpha}(B+\Gamma)^2\|X\|_\infty^2\big)$$

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

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

5. An example of application

In this section we give an instance of error bound in the approximation by $\displaystyle\frac{1}{N}\sum_{i=1}^N \varphi(X^i)$ of the expectation of a quantity $\varphi(X)$ depending on the whole path of the considered process.

Let $0 < t_1 < \cdots < t_n \leq T$ and for $1 \leq j \leq n$ let B_j be the (for instance closed) ball $B(x_j, r_j)$ of center x_j and radius $r_j > 0$ in the Euclidean space $\mathbb{R}^d$. Then we can approximate the probability $\mathbb{P}\big[X_{t_j} \in B_j; 1 \leq j \leq n\big]$ for X_t to be in B_j at $t = t_j$ for each j as follows:

In the notation and assumptions of Theorem 1, assume moreover that all partial derivatives $\partial^m b$ and $\partial^m c$ of b and c are continuous and bounded on $\mathbb{R}^d$, for any multi-index $m \in \mathbb{N}^d$ with $1 \leq |m| \leq s$ where s is the smallest integer number larger than $d/2$, and that the diffusion matrix σ is (for instance) a nonzero multiple of the identity, as in the example (6). Then, for any $\alpha \in (0, 1/2)$, there exist positive constants K and N_0 such that

$$\mathbb{P}\left[\left|\frac{1}{N}\sum_{i=1}^N\prod_{j=1}^n \chi_j(X^i_{t_j}) - \mathbb{P}\big[X_{t_j} \in B_j; 1 \leq j \leq n\big]\right| > \varepsilon\right] \leq e^{-K\,N\,\varepsilon^4}$$

for all $\varepsilon \in (0, 1)$ and $N \geq N_0\,\varepsilon^{-4}\,\exp\big(N_0\,\varepsilon^{-2/\alpha}\big)$.

Here χ_j is the indicator function of the ball B_j defined by $\chi_j(x) = 1$ if $x \in B_j$ and 0 otherwise, and let us note that only the $\varepsilon \in (0, 1)$ have to be considered.

Indeed

$$
\begin{aligned}
&\Big|\frac{1}{N}\sum_{i=1}^{N}\prod_{j=1}^{n}\chi_j(X^i_{t_j}) - \mathbb{P}\big[X_{t_j}\in B_j; 1\le j\le n\big]\Big| \\
= &\ \Big|\int_{\mathcal{C}}\prod_{j=1}^{n}\chi_j(f(t_j))\,d\hat\mu^N(f) - \int_{\mathcal{C}}\prod_{j=1}^{n}\chi_j(f(t_j))\,d\mu(f)\Big| \\
\le &\ \int_{\mathcal{C}}\Big|\prod_{j=1}^{n}\chi_j(f(t_j)) - \prod_{j=1}^{n}\varphi_j(f(t_j))\Big|\,d\hat\mu^N(f) + \Big|\int_{\mathcal{C}}\prod_{j=1}^{n}\varphi_j(f(t_j))\,d(\hat\mu^N-\mu)(f)\Big| \\
&+\int_{\mathcal{C}}\Big|\prod_{j=1}^{n}\varphi_j(f(t_j)) - \prod_{j=1}^{n}\chi_j(f(t_j))\Big|\,d\mu(f). \qquad (32)
\end{aligned}
$$

Here φ_j is the $\mathbb{R}^d$ to $\mathbb{R}$ map defined by $\varphi_j(x) = \Big(1-\dfrac{d(x,B_j)}{\delta}\Big)_+$ for some $0<\delta\le\min\limits_{1\le j\le n} r_j$ to be chosen later, where $d(x,A)$ is the distance of a point x to a set A and $u_+=\max(u,0)$ for all real u. For simplicity we write $\hat\mu^N,\mu$ and W_1 instead of $\hat\mu^N_{[0,T]},\mu_{[0,T]}$ and $W_{1,[0,T]}$ respectively.

The second term in (32), which is the main term, will be bounded by the Kantorovich-Rubinstein formulation (9). Indeed, if f and g are two functions in $\mathcal{C}$, then

$$
\left|\prod_{j=1}^{n}\varphi_j(f(t_j)) - \prod_{j=1}^{n}\varphi_j(g(t_j))\right| \le \sum_{j=1}^{n}|\varphi_j(f(t_j))-\varphi_j(g(t_j))|
$$

since the maps φ_j take values in $[0,1]$ and because of the following elementary bound:

Lemma 8. *If a_j and b_j are real numbers in $[0,1]$ for $1\le j\le n$, then*

$$
\left|\prod_{j=1}^{n}a_j - \prod_{j=1}^{n}b_j\right| \le \sum_{j=1}^{n}|a_j-b_j|.
$$

Then the maps φ_j are δ^{-1}-Lipschitz, so

$$
\sum_{j=1}^{n}|\varphi_j(f(t_j))-\varphi_j(g(t_j))| \le n\delta^{-1}\|f-g\|_\infty.
$$

Consequently the map $f\mapsto\prod\limits_{j=1}^{n}\varphi_j(f(t_j))$ is $n\delta^{-1}$-Lipschitz on $\mathcal{C}$, and by (9) the second term in (32) is bounded by $n\delta^{-1}\,W_1(\hat\mu^N,\mu)$.

Then the first term in (32) is bounded by

$$
\int_{\mathcal{C}}\sum_{j=1}^{n}|\chi_j(f(t_j))-\varphi_j(f(t_j))|\,d\hat\mu^N(f)
$$

by Lemma 8 again. Now, for all $1 \leq j \leq n$ and $x \in \mathbb{R}^d$,

$$|\chi_j(x) - \varphi_j(x)| = \varphi_j(x) - \chi_j(x) \leq \gamma_j(x)$$

where $\gamma_j(x) = \left(1 - \dfrac{d(x, B_j) + d(x, \mathbb{R}^d \setminus B_j)}{\delta}\right)_+$. The map $x \mapsto d(x, B_j) + d(x, \mathbb{R}^d \setminus B_j)$ is 1-Lipschitz, so γ_j is δ^{-1}-Lipschitz, and by (9) again the first term in (32) is bounded by

$$\sum_{j=1}^n \int_{\mathcal{C}} \gamma_j(f(t_j)) d\hat{\mu}^N(f) \leq \sum_{j=1}^n \Big[\delta^{-1} W_1(\hat{\mu}^N, \mu) + \int_{\mathcal{C}} \gamma_j(f(t_j)) d\mu(f)\Big].$$

Moreover, if for any j we let μ_{t_j} be the marginal at time t_j of the distribution μ, then

$$\int_{\mathcal{C}} \gamma_j(f(t_j)) d\mu(f) = \int_{\mathbb{R}^d} \gamma_j(x)\, d\mu_{t_j}(x) \leq \mu_{t_j}[C_j]$$

where $C_j = B(x_j, r_j+\delta) \setminus B(x_j, r_j-\delta)$, since γ_j is zero outside of C_j and bounded by 1 on C_j.

In the same way the third term in (32) is bounded by

$$\sum_{j=1}^n \mu_{t_j}[D_j]$$

where $D_j = B(x_j, r_j + \delta) \setminus B(x_j, r_j)$, since $\varphi_j - \chi_j$ is zero outside of D_j and bounded by 1 on D_j.

Now, under our assumptions, we can adapt the techniques in [4, Theorem B.1] to prove that for any $t > 0$ the time marginal μ_t belongs to the Sobolev space $H^s(\mathbb{R}^d)$, with $s > d/2$, whence to $L^\infty(\mathbb{R}^d)$. Moreover there exists a constant K_1, depending only on $t_1, t_n, d, b, c, \sigma$ and a square exponential moment of μ_0, such that

$$\sup_{t_1 \leq t \leq t_n} \|\mu_t\|_{L^\infty} \leq K_1.$$

In particular the first and third terms in (32) are together bounded by

$$n\delta^{-1}\, W_1(\hat{\mu}^N, \mu) + K_1 \sum_{j=1}^n Leb[C_j] + Leb[D_j]$$

where $Leb[A]$ stands for the Lebesgue measure of a Borel set A in $\mathbb{R}^d$.

Moreover

$$Leb[C_j] = \omega_d\big((r_j + \delta)^d - (r_j - \delta)^d\big) \leq 2^d\, \omega_d\, r_j^{d-1}\delta$$

and

$$Leb[D_j] = \omega_d\big((r_j + \delta)^d - r_j^d\big) \leq (2^d - 1)\, \omega_d\, r_j^{d-1}\delta$$

for any $1 \leq j \leq n$ and $0 < \delta \leq r_j$, where ω_d is the Lebesgue measure of the Euclidean unit ball in $\mathbb{R}^d$. Hence it follows from (32) that for a constant K_2, independent of N and the r_j,

$$\Big|\frac{1}{N}\sum_{i=1}^N \prod_{j=1}^n \chi_j(X^i_{t_j}) - \mathbb{P}\big[X_{t_j} \in B_j; 1 \leq j \leq n\big]\Big| \leq 2n\delta^{-1} W_1(\hat{\mu}^N, \mu) + K_2 \sum_{j=1}^n r_j^{d-1}\delta$$

for any $0 < \delta \leq \min_{1 \leq j \leq n} r_j$ and in particular

$$\mathbb{P}\Big[\Big|\frac{1}{N}\sum_{i=1}^N \prod_{j=1}^n \chi_j(X^i_{t_j}) - \mathbb{P}\big[X_{t_j} \in B_j; 1 \leq j \leq n\big]\Big| > \varepsilon\Big] \leq \mathbb{P}\Big[W_1(\hat{\mu}^N, \mu) > \frac{\delta}{2n}(\varepsilon - K_2 R\delta)\Big]$$

for any $0 < \delta \leq r$, in the notation $r = \min_{1 \leq j \leq n} r_j$ and $R = \sum_{j=1}^n r_j^{d-1}$.

In the case when $\varepsilon \leq 2K_2Rr$ we choose $\delta = \dfrac{\varepsilon}{2K_2R} \in (0, r]$, so that, by Theorem 1, for any $\alpha \in (0, 1/2)$ there exist some constants K and N_0, independent of ε and N, such that

$$\mathbb{P}\Big[W_1(\hat{\mu}^N, \mu) > \frac{\delta}{2n}(\varepsilon - K_2R\delta)\Big] = \mathbb{P}\Big[W_1(\hat{\mu}^N, \mu) > \frac{\varepsilon^2}{8K_2Rn}\Big] \leq \exp\Big(- KN\Big(\frac{\varepsilon^2}{8K_2Rn}\Big)^2\Big)$$

for all $N \geq N_0 \left(\frac{\varepsilon^2}{8K_2Rn}\right)^{-2} \exp\Big(N_0\big(\frac{\varepsilon^2}{8K_2Rn}\big)^{-1/\alpha}\Big)$.

In the case when $\varepsilon > 2K_2Rr$ we choose $\delta = r$, so that

$$\mathbb{P}\Big[W_1(\hat{\mu}^N, \mu) > \frac{\delta}{2n}(\varepsilon - K_2R\delta)\Big] \leq \mathbb{P}\Big[W_1(\hat{\mu}^N, \mu) > \frac{\varepsilon r}{4n}\Big] \leq \exp\Big(- KN\big(\frac{\varepsilon r}{4n}\big)^2\Big)$$

for all $N \geq N_0 \left(\frac{\varepsilon r}{4n}\right)^{-2} \exp\Big(N_0\big(\frac{\varepsilon r}{4n}\big)^{-1/\alpha}\Big)$.

As a conclusion, there exist two new constants K and N_0, depending on μ_0 only through a square exponential moment, such that

$$\mathbb{P}\Big[\Big|\frac{1}{N}\sum_{i=1}^N \prod_{j=1}^n \chi_j(X^i_{t_j}) - \mathbb{P}\big[X_{t_j} \in B_j; 1 \leq j \leq n\big]\Big| > \varepsilon\Big] \leq e^{-KN\varepsilon^4}$$

for all $\varepsilon \in (0, 1)$ and $N \geq N_0\, \varepsilon^{-4} \exp(N_0 \varepsilon^{-2/\alpha})$. This concludes the argument.

Appendix. Metric entropy of a Hölder space

In this appendix we establish the bound (23) used in the covering argument in the proof of Proposition 5, which amounts to studying the metric entropy of a Hölder space and of a related space of probability measures.

In the notation introduced in Section 3, it follows from Ascoli's theorem that the closed ball $\mathcal{B}^\alpha_R := \mathcal{B}^\alpha_R([0, T], \mathbb{R}^d) = \{f \in \mathcal{C}^\alpha; \|f\|_\alpha \leq R\}$ 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 $\mathcal{B}^\alpha_R$ the compact metric space $\mathcal{B}^\alpha_R$ can be covered. We note that for $r \geq R$ the sole ball $\{f \in \mathcal{B}^\alpha_R; \|f\|_\infty \leq r\}$ covers $\mathcal{B}^\alpha_R$.

Notation: Given $r > 0$, the *covering number* $\mathcal{N}(S, r)$ of a compact metric space (S, d) is the smallest integer n such that S can be covered by n balls centered in S and of radius r in d metric. The following result gives lower and upper bounds on the covering number $\mathcal{N}(\mathcal{B}^\alpha_R, r)$ and in our case makes more precise those given for instance in [12], [14] or [18]:

Theorem 9. *Given some integer $d \geq 1$, 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 Proposition 5 and hence in Theorem 1, 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 some integers J and K larger or equal to 1, we let $\tau = \dfrac{T}{J}$ and $\eta = \dfrac{R}{K}$, and then

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

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 $\mathcal{B}_R^\alpha([0,T],\mathbb{R})$. The intervals $[y_k - \frac{\eta}{2}, y_k + \frac{\eta}{2}]$ cover $[-R,R]$, so for any integer $j \in [1,J]$ there exists an 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 assume $KT^\alpha < J^\alpha$. But the y_k take values regularly distant of η, so 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

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

In particular we note that this function f_k is Lipschitz with

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

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

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

Now if K is the smallest integer larger than $4\,\dfrac{R}{r}$ and J is such that $KT^\alpha < J^\alpha$, then

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

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

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

Then we can cover $\mathcal{B}_R^\alpha([0,T],\mathbb{R})$ 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 we have $KT^\alpha < J^\alpha$ and

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

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

Let F be a given function in $\mathcal{B}_R^\alpha([0,T],\mathbb{R}^d)$ with components $F_i \in \mathcal{B}_R^\alpha([0,T],\mathbb{R})$ for $1 \le i \le d$. 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

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

where the f_k are the functions in $\mathcal{C}([0,T],\mathbb{R})$ defined in step 1 with $\dfrac{r}{\sqrt{d}}$ instead of r.

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

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 $\dfrac{r}{2}$ 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 step 1. 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

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

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

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

Given some integer ℓ such that $-\tau^{-\alpha} + \dfrac{1}{2} \le \ell \le \tau^{-\alpha} - \dfrac{1}{2}$, 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 $\|f_{k\ell} - f_{k'\ell'}\|_\infty \ge \eta$ if $f_{k\ell} \ne f_{k'\ell'}$.

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 $f_{k\ell} \ne f_{k'\ell'}$.

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 cardinality of integers ℓ in $\big[-\tau^{-\alpha} + \dfrac{1}{2}, \tau^{-\alpha} - \dfrac{1}{2}\big]$. Hence

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

But

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

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

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

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

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

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

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

Then we have the following general result which is proven in [4] (see also [9], [13]):

Theorem 10. *Let (S, d) be a compact metric space with finite diameter D, and p and δ be real numbers with $p \geq 1$ and $0 < \delta < D$. Then the covering numbers of S and $\mathcal{P}(S)$ satisfy*

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

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

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

Theorem 11. *Let $d \geq 1$, p, T, R, δ and α be 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*

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

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

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

Acknowledgments. The author thanks Professor M. Ledoux for his kind interest in this work, in particular pointing out Reference [11], and Professors A. Guillin and C. Villani for stimulating discussions during the preparation of [4].

References

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

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

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

[4] F. Bolley, A. Guillin and C. Villani. Quantitative concentration inequalities for empirical measures on non-compact spaces. *Prob. Theor. Rel. Fields 137*, 3-4 (2007), 541–593.

[5] 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.

[6] J. A. Carrillo, R. J. McCann and C. Villani. Contractions in the 2-Wasserstein length space and thermalization of granular media. *Arch. Rat. Mech. Anal. 179* (2006), 217–263.

[7] P. Cattiaux, A. Guillin and F. Malrieu. Probabilistic approach for granular media equations in the non uniformly case. *Prob. Theor. Rel. Fields 140*, 1-2 (2008), 19–40.

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

[9] A. Dembo and O. Zeitouni. *Large deviations techniques and applications.* Springer, NewYork, 1998.
[10] J. Dolbeault. Free energy and solutions of the Vlasov-Poisson-Fokker-Planck system: external potential and confinement (large time behavior and steady states). *J. Math. Pures Appl. 9*, 78, 2 (1999), 121–157.
[11] X. Fernique. Régularité des trajectoires des fonctions aléatoires gaussiennes. Lecture Notes in Math. 480, Springer, Berlin, 1975.
[12] 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.
[13] S. R. Kulkarni and O. Zeitouni. A general classification rule for probability measures. *Ann. Stat. 23*, 4 (1995), 1393–1407.
[14] G. G. Lorentz. *Approximation of functions.* Holt, Rinehart and Winston, New York, 1966.
[15] F. Malrieu. Logarithmic Sobolev inequalities for some nonlinear PDE's. *Stoch. Proc. Appl. 95*, 1 (2001), 109–132.
[16] S. Méléard. Asymptotic behaviour of some interacting particle systems; McKean-Vlasov and Boltzmann models. Lecture Notes in Math. 1627, Springer, Berlin, 1996.
[17] A.-S. Sznitman. Topics in propagation of chaos. Lecture Notes in Math. 1464, Springer, Berlin, 1991.
[18] A. van der Vaart and J. Wellner. *Weak convergence and empirical processes.* Springer, Berlin, 1995.
[19] C. Villani. *Topics in optimal transportation.* Grad. Stud. Math. 58, A.M.S., Providence, 2003.