

QUANTITATIVE CONCENTRATION INEQUALITIES FOR EMPIRICAL MEASURES ON NON-COMPACT SPACES

FRANÇOIS BOLLEY, ARNAUD GUILLIN, AND CÉDRIC VILLANI

ABSTRACT. We establish some quantitative concentration estimates for the empirical measure of many independent variables, in transportation distances. As an application, we provide some error bounds for particle simulations in a model mean field problem. The tools include coupling arguments, as well as regularity and moments estimates for solutions of certain diffusive partial differential equations.

CONTENTS

Introduction	1
1. Tools and main results	4
2. The case of independent variables	13
3. PDE estimates	25
4. The limit empirical measure	30
5. Coupling	36
6. Conclusion	39
Appendix A. Metric entropy of a probability space	42
Appendix B. Regularity estimates on the limit PDE	46
References	50

INTRODUCTION

Large stochastic particle systems constitute a popular way to perform numerical simulations in many contexts, either because they are used in some physical model (as in e.g. stellar or granular media) or as an approximation of a continuous model (as in e.g. vortex simulation for Euler equation, see [21, Chapter 5] for instance). For such systems one may wish to establish concentration estimates showing that the behavior of the system is sharply stabilized as the number N of particles goes to infinity. It is natural to search for these estimates in the setting of large (or moderate) deviations, since one wishes to make sure that the numerical method has a very small probability to give wrong results. From a physical perspective, concentration estimates may be useful to establish the validity of a continuous approximation such as a mean-field limit.

Key words and phrases. Transport inequalities, Sanov Theorem.

When one is interested in the asymptotic behavior of just one, or a few observables (such as the mean position...), there are efficient methods, based for instance on concentration of measure theory. As a good example, Malrieu [19] recently applied tools from the fields of Logarithmic Sobolev inequalities, optimal transportation and concentration of measure, to prove very neat bounds like

$$\sup_{\|\varphi\|_{\text{Lip}} \leq 1} \mathbb{P} \left[\left| \frac{1}{N} \sum_{i=1}^N \varphi(X_t^i) - \int \varphi d\mu_t \right| > \frac{C}{\sqrt{N}} + \varepsilon \right] \leq 2e^{-\lambda N \varepsilon^2}. \quad (0.1)$$

Here $(X_t^i)_{1 \leq i \leq N}$ stand for the positions of particles (in phase space) at time t , ε is a given error, \mathbb{P} stands for the probability, μ_t is a probability measure governing the limit behavior of the system, C and λ are positive constants depending on the particular system he is considering (a simple instance of McKean-Vlasov model used in particular in the modelling of granular media). Moreover,

$$\|\varphi\|_{\text{Lip}} := \sup_{x \neq y} \frac{|\varphi(x) - \varphi(y)|}{d(x, y)},$$

where d is the distance in phase space (say the Euclidean norm $|\cdot|$ in \mathbb{R}^d).

This approach can lead to nice bounds, but has the drawback to be limited to a finite number of observables. Of course, one may apply (0.1) to many functions φ , and obtain something like

$$\mathbb{P} \left[\sum_{k=1}^{\infty} \frac{1}{k^2} \left| \frac{1}{N} \sum_{i=1}^N \varphi_k(X_t^i) - \int \varphi_k d\mu_t \right| > \frac{C}{\sqrt{N}} + \varepsilon \right] \leq Ce^{-N\lambda\varepsilon^2}, \quad (0.2)$$

where $(\varphi_k)_{k \in \mathbb{N}}$ is an arbitrarily chosen dense family in the set of all 1-Lipschitz functions converging to 0 at infinity. If we denote by δ_x the Dirac mass at point x , and by

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

the empirical measure associated with the system (this is a random probability measure), then estimate (0.2) can be interpreted as a bound on how close $\widehat{\mu}_t^N$ is to μ_t . Indeed,

$$d(\mu, \nu) := \sum_{k=1}^{\infty} \frac{1}{k^2} \left| \int \varphi_k d(\mu - \nu) \right| \quad (0.3)$$

defines a distance on probability measures, associated with a topology which is at least as strong as the weak convergence of measures (convergence against bounded continuous test functions). However, this point of view is deceiving: for practical purposes, the distance d can hardly be estimated, and in any case (0.2) does not contain more information than (0.1): it is only useful if one considers a finite number of observables.

Sanov's large deviation principle [12, Theorem 6.2.10] provides a more satisfactory tool to estimate the distance between the empirical measure and its limit. Roughly speaking,

it implies, for independent variables X_t^i , an estimate of the form

$$\mathbb{P} [\text{dist}(\widehat{\mu}_t^N, \mu) \geq \varepsilon] \simeq e^{-N\alpha(\varepsilon)} \quad \text{as } N \rightarrow \infty,$$

where

$$\alpha(\varepsilon) := \inf \left\{ H(\nu|\mu); \text{dist}(\nu, \mu) \geq \varepsilon \right\} \quad (0.4)$$

and H is the relative H functional:

$$H(\nu|\mu) = \int \frac{d\nu}{d\mu} \log \frac{d\nu}{d\mu} d\mu$$

(to be interpreted as $+\infty$ if ν is not absolutely continuous with respect to μ). Since H behaves in many ways like a square distance, one can hope that $\alpha(\varepsilon) \geq \text{const.} \varepsilon^2$. Here “dist” may be any distance which is continuous with respect to the weak topology, a condition which might cause trouble on a non-compact phase space.

Yet Sanov’s theorem is not the final answer either: it is actually asymptotic, and only implies a bound like

$$\limsup \frac{1}{N} \log \mathbb{P} [\text{dist}(\widehat{\mu}_t^N, \mu) \geq \varepsilon] \leq -\alpha(\varepsilon),$$

which, unlike (0.1), does not contain any explicit estimate for a given N . Fortunately, there are known techniques to obtain quantitative upper bounds for such theorems, see in particular [12, Exercise 4.5.5]. Since these techniques are devised for compact phase spaces, a further truncation will be necessary to treat more general situations.

In this paper, we shall show how to combine these ideas with recent results about measure concentration and transportation distances, in order to derive in a systematic way estimates that are explicit, deal with the empirical measure as a whole, apply to non-compact phase spaces, and can be used to study some particle systems arising in practical problems. Typical estimates will be of the form

$$\mathbb{P} \left[\sup_{\|\varphi\|_{\text{Lip}} \leq 1} \left(\frac{1}{N} \sum_{i=1}^N \varphi(X_t^i) - \int \varphi d\mu_t \right) > \varepsilon \right] \leq C e^{-\lambda N \varepsilon^2}. \quad (0.5)$$

As a price to pay, the constant C in the right-hand side will be much larger than the one in (0.1).

Here is a possible application of (0.5) in a numerical perspective. Suppose your system has a limit invariant measure $\mu_\infty = \lim \mu_t$ as $t \rightarrow \infty$, and you wish to numerically plot its density f_∞ . For that, you run your particle simulation for a long time $t = T$, and plot, say,

$$\widetilde{f}_t(x) := \frac{1}{N} \sum_{i=1}^N \zeta_\alpha(x - X_t^i), \quad (0.6)$$

where $\zeta_\alpha = \alpha^{-d} \zeta(x/\alpha)$ is a smooth approximation of a Dirac mass as $\alpha \rightarrow 0$ (as usual, ζ is a nonnegative smooth radial function on \mathbb{R}^d with compact support and unit integral). With the help of estimates such as (0.5), it is often possible to compute bounds on, say,

$$\mathbb{P} \left[\|\widetilde{f}_T - f_\infty\|_{L^\infty} > \varepsilon \right]$$

in terms of N , ε , T and α . In this way one can “guarantee” that all details of the invariant measure are captured by the stochastic system. While this problem is too general to be treated abstractly, we shall show on some concrete model examples how to derive such bounds for the same kind of systems that was considered by Malrieu.

In the next section, we shall explain about our main tools and results; the rest of the paper will be devoted to the proofs. Some auxiliary estimates of general interest are postponed in Appendix.

1. TOOLS AND MAIN RESULTS

1.1. Wasserstein distances. To measure distances between probability measures, we shall use transportation distances, also called **Wasserstein distances**. They can be defined in an abstract Polish space X as follows: given p in $[1, +\infty)$, d a lower semi-continuous distance on X , and μ and ν two Borel probability measures on X , the Wasserstein distance of order p between μ and ν is

$$W_p(\mu, \nu) := \inf_{\pi \in \Pi(\mu, \nu)} \left(\iint d(x, y)^p d\pi(x, y) \right)^{1/p}$$

where π runs over the set $\Pi(\mu, \nu)$ of all joint probability measures on the product space $X \times X$ with marginals μ and ν ; it is easy to check [29, Theorem 7.3] that W_p is a distance on the set $P_p(X)$ of Borel probability measures μ on X such that $\int d(x_0, x)^p d\mu(x) < +\infty$.

For this choice of distance, in view of Sanov’s theorem, a very natural class of inequalities is the family of so-called transportation inequalities, or **Talagrand inequalities** (see [17] for instance): by definition, given $p \geq 1$ and $\lambda > 0$, a probability measure μ on X satisfies $T_p(\lambda)$ if the inequality

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

holds for any probability measure ν . We shall say that μ satisfies a T_p inequality if it satisfies $T_p(\lambda)$ for some $\lambda > 0$. By Jensen’s inequality, these inequalities become stronger as p becomes larger; so the weakest of all is T_1 . Some variants introduced in [8] will also be considered.

Of course T_p is not a very explicit condition, and a priori it is not clear how to check that a given probability measure satisfies it. It has been proven [7, 14, 8] that T_1 is *equivalent* to the existence of a square-exponential moment: in other words, a reference measure μ satisfies T_1 if and only if there is $\alpha > 0$ such that

$$\int e^{\alpha d(x, y)^2} d\mu(x) < +\infty$$

for some (and thus any) $y \in X$. If that condition is satisfied, then one can find explicitly some λ such that $T_1(\lambda)$ holds true: see for instance [8].

This criterion makes T_1 a rather convenient inequality to use. Another popular inequality is T_2 , which appears naturally in many situations where a lot of structure is available, and which has good tensorization properties in many dimensions. Up to now, T_2 inequalities have not been so well characterized: it is known that they are implied by a Logarithmic

Sobolev inequality [23, 6, 30], and that they imply a Poincaré, or spectral gap, inequality [23, 6]. See [11] for an attempt to a criterion for T_2 . In any case, contrary to the case $p = 1$, there is no hope to obtain T_2 inequalities from just integrability or decay estimates.

In this paper, we shall mainly focus on the case $p = 1$, which is much more flexible.

1.2. Metric entropy. When X is a compact space, the minimum number $m(X, r)$ of balls of radius r needed to cover X is called the **metric entropy** of X . This quantity plays an important role in quantitative variants of Sanov's Theorem [12, Exercise 4.5.5]. In the present paper, to fix ideas we shall always be working in the particular Euclidean space \mathbb{R}^d , which of course is not compact; and we shall reduce to the compact case by truncating everything to balls of finite radius R . This particular choice will influence the results through the function $m(\mathcal{P}_p(B_R), r)$, where B_R is the ball of radius R centered at some point, say the origin, and $\mathcal{P}_p(B_R)$ is the space of probability measures on B_R , metrized by W_p .

1.3. Sanov-type theorems. The core of our estimates is based on variants of Sanov's Theorem, all dealing with *independent* random variables. Let μ be a given probability measure on \mathbb{R}^d , and let $(X^i)_{i=1, \dots, N}$ be a sample of independent variables, all distributed according to μ ; let also

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

be the associated empirical measure. In our first main result we assume a T_p inequality for the measure μ , and deduce from that an upper bound in W_p distance:

Theorem 1.1. *Let $p \in [1, 2]$ and let μ be a probability measure on \mathbb{R}^d satisfying a $T_p(\lambda)$ inequality. Then, for any $d' > d$ and $\lambda' < \lambda$, there exists some constant N_0 , depending on λ', d' and some square-exponential moment of μ , such that for any $\varepsilon > 0$ and $N \geq N_0 \max(\varepsilon^{-(d'+2)}, 1)$,*

$$\mathbb{P} [W_p(\mu, \widehat{\mu}^N) > \varepsilon] \leq e^{-\gamma_p \frac{\lambda'}{2} N \varepsilon^2}, \quad (1.1)$$

where

$$\gamma_p = \begin{cases} 1 & \text{if } 1 \leq p < 2 \\ 3 - 2\sqrt{2} & \text{if } p = 2 \end{cases}.$$

Compared to Sanov's Theorem, this result is more restrictive in the sense that it requires some extra assumptions on the reference measure μ , but under these hypotheses we are able to replace a result which was only asymptotic by a pointwise upper bound on the error probability, together with a lower bound on the required size of the sample.

In view of the Kantorovich-Rubinstein duality formula

$$W_1(\mu, \nu) = \sup \left\{ \int f d(\mu - \nu); \|f\|_{\text{Lip}} \leq 1 \right\}, \quad (1.2)$$

Theorem 1.1 implies concentration inequalities such as

$$\mathbb{P} \left[\sup_{f; \|f\|_{\text{Lip}} \leq 1} \left(\frac{1}{N} \sum_{k=1}^N f(X_k) - \int f d\mu \right) > \varepsilon \right] \leq e^{-\frac{\lambda'}{2} N \varepsilon^2}$$

for $\lambda' < \lambda$, and N sufficiently large, under the assumption that μ satisfies a T_1 inequality, or equivalently admits a finite square-exponential moment. Those types of inequalities are of interest in non-parametric statistics and choice models [22].

Remark 1.2. The sole inequality $T_1(\lambda)$ implies that for all 1-Lipschitz function f ,

$$\mathbb{P} \left[\frac{1}{N} \sum_{k=1}^N f(X_k) - \int f d\mu > \varepsilon \right] \leq e^{-\frac{\lambda}{2} N \varepsilon^2},$$

and it is easy to see that the coefficient λ in this inequality is the best possible. While the quantity controlled in Theorem 1.1 is much stronger, the estimate is weakened only in that λ is replaced by some $\lambda' > \lambda$ (arbitrarily close to λ) and that N has to be large enough. In fact, a variant of the proof below would yield estimates such as

$$\mathbb{P} [W_p(\mu, \hat{\mu}^N) > \varepsilon] \leq C(\varepsilon) e^{-\gamma \frac{\lambda'}{2} N \varepsilon^2},$$

where now there is no restriction on N , but $C(\varepsilon)$ is a larger constant, explicitly computable from the proof.

Remark 1.3. As pointed out to us by M. Ledoux, there is another way to concentration estimates on the empirical measure when $d = p = 1$. Indeed, in this specific case,

$$W_1(\hat{\mu}^N, \mu) = \left\| \frac{1}{N} \sum_{i=1}^N H(\cdot - X_i) - F \right\|_{L^1(\mathbb{R})}$$

where $H = \mathbf{1}_{[0, +\infty)}$ stands for the Heaviside function on \mathbb{R} and F denotes the repartition function of μ , so that

$$\mathbb{P} [W_1(\hat{\mu}^N, \mu) \geq \varepsilon] = \mathbb{P} \left[\left\| \frac{1}{N} \sum_{i=1}^N F_i \right\|_{L^1} > \varepsilon \right]$$

where

$$F_i := H(\cdot - X_i) - F \quad (1 \leq i \leq N)$$

are centered $L^1(\mathbb{R})$ -valued independent identically distributed random variables. But, according to [1, Exercise 3.8.14], a centered $L^1(\mathbb{R})$ -valued random variable Y satisfies a Central Limit Theorem if and only if

$$\int_{\mathbb{R}} (\mathbb{E}[Y^2(t)])^{1/2} dt < +\infty,$$

a condition which for the random variables F_i 's can be written

$$\int_{\mathbb{R}} \sqrt{F(t)(1 - F(t))} dt < +\infty. \tag{1.3}$$

Condition (1.3) in turn holds true as soon as (for instance) $\int_{\mathbb{R}} |x|^{2+\delta} d\mu(x)$ is finite for some positive δ . Then we may apply a quantitative version of the Central Limit Theorem for random variables in the Banach space $L^1(\mathbb{R})$. See [16] and [18] for related works.

Remark 1.4. Theorem 1.1 applies if N is at least as large as ε^{-r} for some $r > d + 2$; we do not know whether $d + 2$ here is optimal.

For the applications that we shall treat, in which the tails of the probability distributions will be decaying very fast, Theorem 1.1 will be sufficient. However, it is worthwhile pointing out that the technique works under much broader assumptions: weaker estimates can be proven for probability measures that do not decay fast enough to admit finite square-exponential moments. Here below are some such results using only polynomial moment estimates:

Theorem 1.5. *Let $q \geq 1$ and let μ be a probability measure on \mathbb{R}^d such that*

$$\int_{\mathbb{R}^d} |x|^q d\mu(x) < +\infty.$$

Then

(i) *For any $p \in [1, q/2)$, $\delta \in (0, q/p - 2)$ and $d' > d$, there exists a constant N_0 such that*

$$\mathbb{P} [W_p(\mu, \widehat{\mu}^N) > \varepsilon] \leq \varepsilon^{-q} N^{-\frac{q}{2p} + \frac{\delta}{2}}$$

for any $\varepsilon > 0$ and $N \geq N_0 \max(\varepsilon^{-q\frac{2p+d'}{q-p}}, \varepsilon^{d'-d})$;

(ii) *For any $p \in [q/2, q)$, $\delta \in (0, q/p - 1)$ and $d' > d$ there exists a constant N_0 such that*

$$\mathbb{P} [W_p(\mu, \widehat{\mu}^N) > \varepsilon] \leq \varepsilon^{-q} N^{1 - \frac{q}{p} + \delta}$$

for any $\varepsilon > 0$ and $N \geq N_0 \max(\varepsilon^{-q\frac{2p+d'}{q-p}}, \varepsilon^{d'-d})$.

Here are also some variants under alternative ‘‘regularity’’ assumptions:

Theorem 1.6. (i) *Let $p \geq 1$; assume that $\mathcal{E}_\alpha := \int e^{\alpha|x|} d\mu$ is finite for some $\alpha > 0$. Then, for all $d' > d$, there exist some constants K and N_0 , depending only on d , α and \mathcal{E}_α , such that*

$$\mathbb{P} [W_p(\mu, \widehat{\mu}^N) > \varepsilon] \leq e^{-K N^{1/p} \min(\varepsilon, \varepsilon^2)}$$

for any $\varepsilon > 0$ and $N \geq N_0 \max(\varepsilon^{-(2p+d')}, 1)$.

(ii) *Suppose that μ satisfies T_1 and a Poincaré inequality, then for all $a < 2$ there exists some constants K and N_0 such that*

$$\mathbb{P} [W_2(\mu, \widehat{\mu}^N) > \varepsilon] \leq e^{-K N \min(\varepsilon^2, \varepsilon^a)} \tag{1.4}$$

for any $\varepsilon > 0$ and $N \geq N_0 \max(\varepsilon^{-(4+d')}, 1)$.

- (iii) Let $p > 2$ and let μ be a probability measure on \mathbb{R}^d satisfying $T_p(\lambda)$. Then for all $\lambda' < \lambda$ and $d' > d$ there exists some constant N_0 , depending on μ only through λ and some square-exponential moment, such that

$$\mathbb{P} [W_p(\mu, \widehat{\mu}^N) > \varepsilon] \leq \min \left(e^{-\frac{\lambda'}{2} N \varepsilon^2} + e^{-(N \varepsilon^{d'+2})^{2/d'}}, 2 e^{-\frac{\lambda'}{4} N^{2/p} \varepsilon^2} \right) \quad (1.5)$$

for any $\varepsilon > 0$ and $N \geq N_0 \max(\varepsilon^{-(d'+2)}, 1)$.

1.4. Interacting systems of particles. We now consider a system of N interacting particles whose time-evolution is governed by the system of coupled stochastic differential equations

$$dX_t^i = \sqrt{2} dB_t^i - \nabla V(X_t^i) dt - \frac{1}{N} \sum_{j=1}^N \nabla W(X_t^i - X_t^j) dt, \quad i = 1, \dots, N. \quad (1.6)$$

Here X_t^i is the position at time t of particle number i , the B^i 's are N independent Brownian motions, and V and W are smooth potentials, sufficiently nice that (1.6) can be solved globally in time. We shall always assume that W (which can be interpreted as an interaction potential) is a symmetric function, that is $W(-z) = W(z)$ for all $z \in \mathbb{R}^d$.

Equation (1.6) is a particularly simple instance of coupled system; in the case when V is quadratic and W has cubic growth, it was used as a simple mean-field kinetic model for granular media (see e.g. [19]). While many of our results could be extended to more general systems, that particular one will be quite enough for our exposition.

To this system of particles is naturally associated the empirical measure, defined for each time $t \geq 0$ by

$$\widehat{\mu}_t^N := \sum_{i=1}^N \delta_{X_t^i}. \quad (1.7)$$

Under suitable assumptions on the potentials V and W , it is a classical result that, if the initial positions of the particle system are distributed chaotically (for instance, if they are identically distributed, independent random variables), then the empirical measure $\widehat{\mu}_t^N$ converges as $N \rightarrow \infty$ to a solution of the nonlinear partial differential equation

$$\frac{\partial \mu_t}{\partial t} = \Delta \mu_t + \nabla \cdot \left(\mu_t \nabla (V + W * \mu_t) \right), \quad (1.8)$$

where $\nabla \cdot$ stands for the divergence operator. Equation 1.8 is a simple instance of McKean-Vlasov equation. This convergence result is part of the by now well-developed theory of propagation of chaos, and was studied by Sznitman for pedagogical reasons [27], in the case of potentials that grow at most quadratically at infinity. Later, Benachour, Roynette, Talay and Vallois [2, 3] considered the case where the interaction potential grows faster than quadratically. As far as the limit equation (1.8) is concerned, a discussion of its use in the modelling of granular media in kinetic theory was performed by Benedetto, Caglioti, Carrillo and Pulvirenti [4, 5], while the asymptotic behavior in large time was studied by Carrillo, McCann and Villani [9, 10] with the help of Wasserstein distances and entropy inequality methods. Then Malrieu [19] presented a detailed study of both limits $t \rightarrow \infty$

and $N \rightarrow \infty$ by probabilistic methods, and established estimates of the type of (0.1) under adequate convexity assumptions on V and W (see also [29, Problem 15]).

As announced before, we shall now give some estimates on the convergence at the level of the law itself. To fix ideas, we assume that V and W have locally bounded Hessian matrices satisfying

$$\begin{cases} \text{(i)} & D^2V(x) \geq \beta I, \quad \gamma I \leq D^2W(x) \leq \gamma' I, \quad \forall x \in \mathbb{R}^d, \\ \text{(ii)} & |\nabla V(x)| = O(e^{a|x|^2}) \quad \text{for any } a > 0. \end{cases} \quad (1.9)$$

Under these assumptions, we shall derive the following bounds.

Theorem 1.7. *Let μ_0 be a probability measure on \mathbb{R}^d , admitting a finite square-exponential moment:*

$$\exists \alpha_0 > 0; \quad M_{\alpha_0} := \int e^{\alpha_0|x|^2} d\mu_0(x) < +\infty.$$

Let $(X_0^i)_{1 \leq i \leq N}$ be N independent random variables with common law μ_0 . Let (X_t^i) be the solution of (1.6) with initial value (X_0^1, \dots, X_0^N) , where V and W are assumed to satisfy (1.9); and let μ_t be the solution of (1.8) with initial value μ_0 . Let also $\widehat{\mu}_t^N$ be the empirical measure associated with the $(X_t^i)_{1 \leq i \leq N}$. Then, for all $T \geq 0$, there exists some constant $K = K(T)$ such that, for any $d' > d$, there exists some constants N_0 and C such that for all $\varepsilon > 0$

$$N \geq N_0 \max(\varepsilon^{-(d'+2)}, 1) \implies \mathbb{P} \left[\sup_{0 \leq t \leq T} W_1(\widehat{\mu}_t^N, \mu_t) > \varepsilon \right] \leq C(1 + T\varepsilon^{-2}) \exp(-KN\varepsilon^2).$$

Note that in the above theorem we have proven not only that for all t , the empirical measure is close to the limit measure, but also that the probability of observing any significant deviation during a whole time period $[0, T]$ is small.

The fact that $\widehat{\mu}_t^N$ is very close to the deterministic measure μ_t implies the propagation of chaos: two particles drawn from the system behave independently of each other as $N \rightarrow \infty$ (see Sznitman [27] for more details). But we can also directly study correlations between particles and find more precise estimates: for that purpose it is convenient to consider the empirical measure on *pairs* of particles, defined as

$$\widehat{\mu}_t^{N,2} := \frac{1}{N(N-1)} \sum_{i \neq j} \delta_{(X_t^i, X_t^j)}.$$

By a simple adaptation of the computations appearing in the proof of Theorem 1.7, one can prove

Theorem 1.8. *With the same notation and assumptions as in Theorem 1.7, for all $T \geq 0$ and $d' > d$, there exists some constants $K > 0$ and N_0 such that for all $\varepsilon > 0$*

$$N \geq N_0 \max(\varepsilon^{-(d'+2)}, 1) \implies \mathbb{P} \left[\sup_{0 \leq t \leq T} W_1(\widehat{\mu}_t^{N,2}, \mu_t \otimes \mu_t) > \varepsilon \right] \leq \exp(-KN\varepsilon^2).$$

(Here W_1 stands for the Wasserstein distance or order 1 on $P_1(\mathbb{R}^d \times \mathbb{R}^d)$.) Of course, one may similarly consider the problem of drawing k particles with $k \geq 2$.

Theorems 1.7 and 1.8 use Theorem 1.1 as a crucial ingredient, which is why a strong integrability assumption is imposed on μ_0 . Note however that, under stronger assumptions on the behaviour at infinity of V or W , as the existence of some $\beta \in \mathbb{R}$, $B, \varepsilon > 0$ such as

$$D^2V(x) \geq (B|x|^\varepsilon + \beta)I, \quad , \forall x \in \mathbb{R}^d,$$

it can be proven that any square exponential moment for μ_t becomes instantaneously finite for $t > 0$. Note also that, by using Theorem 1.5, one can obtain weaker but still relevant results of concentration of the empirical measure under just polynomial moment assumptions on μ_0 , provided that ∇V does not grow too fast at infinity. To limit the size of this paper, we shall not go further into such considerations.

1.5. Uniform in time estimates. In the “uniformly convex case” when $\beta > 0, \beta + 2\gamma > 0$, it can be proven [9, 10, 19] that μ_t converges exponentially fast, as $t \rightarrow \infty$, to some equilibrium measure μ_∞ . In that case, it is natural to expect that the empirical measure is a good approximation of μ_∞ as $N \rightarrow \infty$ and $t \rightarrow \infty$, uniformly in time. This is what we shall indeed prove:

Theorem 1.9. *With the same notation and assumptions as in Theorem 1.7, suppose that $\beta > 0, \beta + 2\gamma > 0$. Then there exists some constant $K > 0$ such that for any $d' > d$, there exists some constants C and N_0 such that for all $\varepsilon > 0$*

$$N \geq N_0 \max(\varepsilon^{-(d'+2)}, 1) \implies \sup_{t \geq 0} \mathbb{P} [W_1(\hat{\mu}_t^N, \mu_t) > \varepsilon] \leq C(1 + \varepsilon^{-2}) \exp(-K N \varepsilon^2)$$

As a consequence, there are constants T_0, ε_0 (depending on the initial datum) and $K' = K/4$ such that, under the same conditions on N and ε ,

$$\sup_{t \geq T_0 \log(\varepsilon_0/\varepsilon)} \mathbb{P} [W_1(\hat{\mu}_t^N, \mu_\infty) > \varepsilon] \leq C(1 + \varepsilon^{-2}) \exp(-K' N \varepsilon^2).$$

Remark 1.10. In view of the results in [9], it is natural to expect that a similar conclusion holds true when $V = 0$ and W is convex enough. Propositions 3.1 and 3.8 below extend to that case, but it seems trickier to adapt the proof of Proposition 3.8.

We conclude with an application to the numerical reconstruction of the invariant measure.

Theorem 1.11. *With the same notation and assumptions as in Theorem 1.9, consider the mollified empirical measure (0.6). Then one can choose $\alpha = O(\varepsilon)$ in such a way that*

$$\begin{aligned} N \geq N_0 \max(\varepsilon^{-(d'+2)}, 1) \implies \sup_{t \geq T_0 \log(\varepsilon_0/\varepsilon)} \mathbb{P} \left[\|\tilde{f}_t - f_\infty\|_{L^\infty} > \varepsilon \right] \\ \leq C(1 + \varepsilon^{-(2d+4)}) \exp(-K' N \varepsilon^{2d+4}). \end{aligned}$$

These results are effective: all the constants therein can be estimated explicitly in terms of the data.

1.6. Strategy and plan. The strategy is rather systematic. First, we shall establish Sanov-type bounds for independent variables in \mathbb{R}^d (not depending on time), resulting in concentration results such as Theorems 1.1 to 1.6. This will be achieved along the ideas in [12, Exercices 4.5.5 and 6.2.19] (see also [25, Section 5]), by first truncating to a compact ball, and then covering the set of probability measures on this ball by a finite number of small balls (in the space of probability measures); the most tricky part will actually lie in the optimization of parameters.

With such results in hand, we will start the study of the particle system by introducing the nonlinear partial differential equation (1.8). For this equation, the Cauchy problem can be solved in a satisfactory way, in particular existence and uniqueness of a solution, which for $t > 0$ is reasonably smooth, can be shown under various assumptions on V and W (see e.g. [9, 10]). Other regularity estimates such as the decay at infinity, or the smoothness in time, can be established; also the convergence to equilibrium in large time can sometimes be proven.

Next, following the presentation by Sznitman [27], we introduce a family of independent processes $(Y_t^i)_{1 \leq i \leq N}$, governed by the stochastic differential equation

$$\begin{cases} dY_t^i &= \sqrt{2} dB_t^i - \nabla V(Y_t^i) dt - \nabla W * \mu_t(Y_t^i) dt, \\ Y_0^i &= X_0^i. \end{cases} \quad (1.10)$$

As a consequence of Itô's formula, the law ν_t of each Y_t^i is a solution of the linear partial differential equation

$$\frac{\partial \nu_t}{\partial t} = \Delta \nu_t + \nabla \cdot \left(\nabla (V + W * \mu_t) \nu_t \right), \quad \nu_0 = \mu_0.$$

But this linear equation is also solved by μ_t , and a uniqueness theorem implies that actually $\nu_t = \mu_t$, for all $t \geq 0$. See [2, 3] for related questions on the stochastic differential equation (1.10).

For each given t , the independence of the variables Y_t^i and the good decay of μ_t will imply a strong concentration of the empirical measure

$$\widehat{\nu}_t^N := \frac{1}{N} \sum_{i=1}^N \delta_{Y_t^i}.$$

To go further, we shall establish a more precise information, such as a control on

$$\mathbb{P} \left[\sup_{0 \leq t \leq T} W_1(\widehat{\nu}_t^N, \mu_t) > \varepsilon \right].$$

Such bounds will be obtained by combining the estimate of concentration at fixed time t with some estimates of regularity of $\widehat{\nu}_t^N$ (and μ_t) in t , obtained via basic tools of stochastic differential calculus (in particular Doob's inequality).

Finally, we can show by a Gronwall-type argument that the control of the distance of $\widehat{\mu}_t^N$ to μ_t reduces to the control of the distance of $\widehat{\nu}_t^N$ to μ_t : for instance,

$$\mathbb{P} \left[\sup_{0 \leq t \leq T} W_1(\widehat{\mu}_t^N, \mu_t) > \varepsilon \right] \leq \mathbb{P} \left[\sup_{0 \leq t \leq T} W_1(\widehat{\nu}_t^N, \mu_t) > C\varepsilon \right] \quad (1.11)$$

for some constant C . We shall also show how a variant of this computation provides estimates of the type of those in Theorem 1.9, and how to get data reconstruction estimates as in Theorem 1.11.

1.7. Remarks and further developments. The results in this paper confirm what seems to be a rather general rule about Wasserstein distances: results in distance W_1 are very robust and can be used in rather hard problems, with no particular structure; on the contrary, results in distance W_2 are stronger, but usually require much more structure and/or assumptions. For instance, in the study of the equation (1.8), the distance W_2 works beautifully, and this might be explained by the fact that (1.8) has the structure of a gradient flow with respect to the W_2 distance [9, 10]. In the problem considered by Malrieu [19], W_2 is also well-adapted, but leads him to impose strong assumptions on the initial datum μ_0 , such as the existence of a Logarithmic Sobolev inequality for μ_0 , considered as a reference measure. As a general rule, in a context of geometric inequalities with more or less subtle isoperimetric content, related to Brenier's transportation mapping theorem, W_2 is also the most natural distance to use [29]. On the contrary, here we are considering quite a rough problem (concentration for the law of a random probability measure, driven by a stochastic differential equation with coupling) and we wish to impose only natural integrability conditions; then the distance W_1 is much more convenient.

Further developments could be considered. For instance, one may desire to prove some deviation inequalities for dependent sequences, say Markov chains, as both Sanov's theorem and transportation inequality can be established under appropriate ergodicity and integrability conditions.

Considering again the problem of the particle system, in a numerical context, one may wish to take into account the numerical errors associated with the time-discretization of the dynamics (say an implicit Euler scheme). For concentration estimates in one observable, a beautiful study of these issues was performed by Malrieu [20]. For concentration estimates on the whole empirical measure, to our knowledge the study remains to be done. Also errors due to the boundedness of the phase space actually used in the simulation might be taken into account, etc.

At a more technical level, it would be desirable to relax the assumption of boundedness of D^2W in Theorem 1.7, so as to allow for instance the interesting case of cubic interaction. This is much more technical and will be considered in a separate work.

Another issue of interest would be to consider concentration of the empirical measure *on path space*, i.e.

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

where T is a fixed time length. Here $\widehat{\mu}_{[0,T]}^N$ is a random measure on $C([0, T]; \mathbb{R}^d)$ and we would like to show that it is close to the law of the trajectories of the nonlinear stochastic differential equation

$$dY_t = \sqrt{2} dB_t - \nabla V(Y_t) dt - (\nabla W * \mu_t)(Y_t) dt, \quad (1.12)$$

where the initial datum Y_0 is drawn randomly according to μ_0 . This will imply a quantitative information on the whole trajectory of a given particle in the system.

When one wishes to adapt the general method to this question, a problem immediately occurs: not only is $C([0, T]; \mathbb{R}^d)$ not compact, but also balls with finite radius in this space are not compact either (of course, this is true even if the phase space of particles is compact). One may remedy to this problem by embedding $C([0, T]; B_R)$ into a space such as $L^2([0, T]; B_R)$, equipped with the weak topology; but we do not know of any “natural” metric on that space. There is (at least) another way out: we know from classical stochastic processes theory that integral trajectories of differential equations driven by white noise are typically Hölder- α for any $\alpha < 1/2$. This suggests a natural strategy: choose any fixed $\alpha \in (0, 1/2)$ and work in the space $\mathcal{H}^\alpha([0, T]; \mathbb{R}^d)$, equipped with the norm

$$\|w\|_{\mathcal{H}^\alpha} := \sup_{0 \leq t \leq T} |w(t)| + \sup_{s \neq t} \frac{|w(t) - w(s)|}{|t - s|^\alpha}.$$

For any $R > 0$, the ball of radius R and center 0 (the zero function) in \mathcal{H}^α is compact, and one may estimate its metric entropy. Then one can hope to perform all estimates by using the norm \mathcal{H}^α ; for instance, establish a bound on, say, a square-exponential moment on the law of Y_t :

$$\mathbb{E} \exp \left(\beta \| (Y_t)_{0 \leq t \leq T} \|_{\mathcal{H}^\alpha}^2 \right) < +\infty.$$

Again, to avoid expanding the size of the present paper too much, these issues will be addressed separately.

2. THE CASE OF INDEPENDENT VARIABLES

In this section we consider the case where we are given N *independent* variables $X^i \in \mathbb{R}^d$, distributed according to a certain law μ . There is no time dependence at this stage. We shall first examine the case when the law μ has very fast decay (Theorem 1.1), then variants in which it decays in a slower way (Theorem 1.5 and 1.6).

2.1. Proof of Theorem 1.1. The proof splits into three steps: (1) Truncation to a compact ball B_R of radius R , (2) covering of $\mathcal{P}(B_R)$ by small balls of radius r and Sanov’s argument, and (3) optimization of the parameters.

Step 1: Truncation. Let $R > 0$, to be chosen later on, and let B_R stand for the ball of radius R and center 0 (say) in \mathbb{R}^d . Let $\mathbf{1}_{B_R}$ stand for the indicator function of B_R . We truncate μ into a probability measure μ_R on the ball B_R :

$$\mu_R = \frac{\mathbf{1}_{B_R} \mu}{\mu[B_R]}.$$

We wish to bound the quantity $\mathbb{P} [W_p(\widehat{\mu}^N, \mu) > \varepsilon]$ in terms of μ_R and the associated empirical measure. For this purpose, consider independent variables $(X^k)_{1 \leq k \leq N}$ drawn according to μ , and $(Y^k)_{1 \leq k \leq N}$ drawn according to μ_R , independent of each other; then define

$$X_R^k := \begin{cases} X^k & \text{if } |X^k| \leq R \\ Y^k & \text{if } |X^k| > R. \end{cases}$$

Since X^1 and X_R^1 are distributed according to μ and μ_R respectively, we have, by definition of Wasserstein distance,

$$\begin{aligned} W_p^p(\mu, \mu_R) &\leq \mathbb{E}|X^1 - X_R^1|^p = \mathbb{E}\left(|X^1 - Y^1|^p \mathbf{1}_{|X^1| > R}\right) \leq 2^p \mathbb{E}\left(|X^1|^p \mathbf{1}_{|X^1| > R}\right) \\ &= 2^p \int_{\{|x| > R\}} |x|^p d\mu(x). \end{aligned}$$

But μ satisfies a $T_p(\lambda)$ inequality for some $p \geq 1$, hence a fortiori a $T_1(\lambda)$ inequality, so

$$E_\alpha := \int_{\mathbb{R}^d} e^{\alpha|x|^2} d\mu(x) < +\infty$$

for some $\alpha > 0$ (any $\alpha < \lambda/2$ would do). If R is large enough (say, $R \geq \sqrt{p/(2\alpha)}$), then the function $r \mapsto \frac{r^p}{e^{\alpha r^2}}$ is nonincreasing for $r \geq R$, and then

$$W_p^p(\mu, \mu_R) \leq 2^p \left(\frac{R^p}{e^{\alpha R^2}} \right) \int_{\{|x| > R\}} e^{\alpha|x|^2} d\mu(x).$$

We conclude that

$$W_p^p(\mu, \mu_R) \leq 2^p E_\alpha R^p e^{-\alpha R^2} \quad (\alpha < \lambda/2, R \geq \sqrt{p/2\alpha}). \quad (2.1)$$

On the other hand, the empirical measures

$$\widehat{\mu}^N := \frac{1}{N} \sum_{k=1}^N \delta_{X^k}, \quad \widehat{\mu}_R^N := \frac{1}{N} \sum_{k=1}^N \delta_{X_R^k}$$

satisfy

$$W_p^p(\widehat{\mu}_R^N, \widehat{\mu}^N) \leq \frac{1}{N} \sum_{k=1}^N |X_R^k - X^k|^p \leq \frac{1}{N} \sum_{k=1}^N Z^k,$$

where $Z^k := 2^p |X^k|^p \mathbf{1}_{|X^k| > R}$ ($k = 1, \dots, N$). Then, for any $p \in [1, 2]$, we can introduce parameters ε and $\theta > 0$, and use Chebyshev's exponential inequality and the independence of the variables Z^k to obtain

$$\begin{aligned} \mathbb{P} [W_p^p(\widehat{\mu}_R^N, \widehat{\mu}^N) > \varepsilon] &\leq \mathbb{P} \left[\frac{1}{N} \sum_{k=1}^N Z^k > \varepsilon^p \right] \\ &= \mathbb{P} \left[\exp \sum_{k=1}^N \theta (Z^k - \varepsilon^p) > 1 \right] \\ &\leq \mathbb{E} \left(\exp \sum_{k=1}^N \theta (Z^k - \varepsilon^p) \right) \\ &= \exp(-N [\theta \varepsilon^p - \log \mathbb{E} \exp(\theta Z_1)]). \end{aligned} \quad (2.2)$$

In the case when $\underline{p < 2}$, for any $\alpha_1 < \alpha < \frac{\lambda}{2}$, there exists some constant $R_0 = R_0(\alpha_1, p)$ such that

$$2^p \theta r^p \leq \alpha_1 r^2 + C,$$

for all $\theta > 0$ and $r \geq R_0 \theta^{\frac{1}{2-p}}$, whence

$$\mathbb{E} \exp(\theta Z_1) \leq \mathbb{E} \exp(\alpha_1 |X_1|^2 \mathbf{1}_{|X^k| > R}) \leq 1 + E_\alpha e^{(\alpha_1 - \alpha) R^2}.$$

As a consequence,

$$\mathbb{P} [W_p(\widehat{\mu}_R^N, \widehat{\mu}^N) > \varepsilon] \leq \exp \left(-N \left[\theta \varepsilon^p - E_\alpha e^{(\alpha_1 - \alpha) R^2} \right] \right). \quad (2.3)$$

From (2.1), (2.3) and the triangular inequality for W_p ,

$$\begin{aligned} \mathbb{P} [W_p(\mu, \widehat{\mu}^N) > \varepsilon] &\leq \mathbb{P} [W_p(\mu, \mu_R) + W_p(\mu_R, \widehat{\mu}_R^N) + W_p(\widehat{\mu}_R^N, \widehat{\mu}^N) > \varepsilon] \\ &\leq \mathbb{P} [W_p(\mu_R, \widehat{\mu}_R^N) > \eta \varepsilon - 2E_\alpha^{1/p} R e^{-\frac{\alpha}{p} R^2}] + \mathbb{P} [W_p(\widehat{\mu}_R^N, \widehat{\mu}^N) > (1 - \eta) \varepsilon] \\ &\leq \mathbb{P} [W_p(\mu_R, \widehat{\mu}_R^N) > \eta \varepsilon - 2E_\alpha^{1/p} R e^{-\frac{\alpha}{p} R^2}] \\ &\quad + \exp \left(-N \left(\theta (1 - \eta)^p \varepsilon^p - E_\alpha e^{(\alpha_1 - \alpha) R^2} \right) \right). \end{aligned} \quad (2.4)$$

This estimate was established for any given $p \in [1, 2)$, $\eta \in (0, 1)$, $\varepsilon, \theta > 0$, $\alpha_1 < \alpha < \frac{\lambda}{2}$ and $R \geq \max \left(\sqrt{p/2\alpha}, R_0 \theta^{\frac{1}{2-p}} \right)$, where R_0 is a constant depending only on α_1 and p .

In the case when $\underline{p = 2}$, we let $Z^k := |Y_k - X_k|^2 \mathbf{1}_{|X_k| > R}$ ($k = 1, \dots, N$), and starting from inequality (2.2) again, we choose $\alpha_1 < \alpha$ and then $\theta := \alpha_1/2$: by definition of Z_1 and μ_R ,

$$\begin{aligned} \mathbb{E} \left(\exp \left(\frac{\alpha_1}{2} Z_1 \right) \right) &= \int_{\mathbb{R}^{2d}} \exp \left(\frac{\alpha_1}{2} |y - x|^2 \mathbf{1}_{|x| \geq R} \right) d\mu(x) d\mu_R(y) \\ &= \mu[B_R] + \frac{1}{\mu[B_R]} \int_{|y| \leq R} \int_{|x| \geq R} \exp \left(\frac{\alpha_1}{2} |y - x|^2 \right) d\mu(x) d\mu(y) \\ &\leq 1 + (1 - E_\alpha e^{-\alpha R^2})^{-1} \int_{|y| \leq R} e^{\alpha_1 |y|^2} d\mu(y) \int_{|x| \geq R} e^{\alpha_1 |x|^2} d\mu(x) \\ &\leq 1 + 2 E_\alpha^2 e^{(\alpha_1 - \alpha) R^2} \end{aligned}$$

for R large enough, from which

$$\mathbb{P} [W_2(\widehat{\mu}_R^N, \widehat{\mu}^N) > \varepsilon] \leq \exp \left(-N \left[\frac{\alpha_1}{2} \varepsilon^2 - 2 E_\alpha^2 e^{(\alpha_1 - \alpha) R^2} \right] \right). \quad (2.5)$$

To sum up, in the case $p = 2$ equation (2.4) writes

$$\begin{aligned} \mathbb{P} [W_2(\mu, \widehat{\mu}^N) > \varepsilon] &\leq \mathbb{P} [W_2(\mu_R, \widehat{\mu}_R^N) > \eta \varepsilon - 2E_\alpha^{1/2} R e^{-\frac{\alpha}{2} R^2}] \\ &\quad + \exp \left(-N \left(\frac{\alpha_1}{2} (1 - \eta)^2 \varepsilon^2 - 2 E_\alpha^2 e^{(\alpha_1 - \alpha) R^2} \right) \right). \end{aligned} \quad (2.6)$$

So, apart from some error terms, for all $p \in [1, 2]$ we have reduced the initial problem to establishing the result only for the probability law μ_R , whose support lies in the compact set B_R .

We end up this truncation procedure by proving that μ_R satisfies some modified T_p inequality. Let indeed ν be a probability measure on B_R , absolutely continuous with respect to μ (and hence with respect to μ_R); then, when R is larger than some constant depending only on E_α , we can write

$$\begin{aligned} H(\nu|\mu_R) - H(\nu|\mu) &= \int_{B_R} \log \frac{d\nu}{d\mu_R} d\nu - \int_{B_R} \log \frac{d\nu}{d\mu} d\nu = \log \mu[B_R] \\ &\geq \log \left(1 - E_\alpha e^{-\alpha R^2} \right) \\ &\geq -2E_\alpha e^{-\alpha R^2}. \end{aligned} \quad (2.7)$$

But μ satisfies a $T_p(\lambda)$ inequality, so

$$H(\nu|\mu) \geq \frac{\lambda}{2} W_p^2(\mu, \nu) \geq \frac{\lambda}{2} \left(W_p(\mu_R, \nu) - W_p(\mu_R, \mu) \right)^2$$

by triangular inequality. Combining this with (2.7), we obtain

$$H(\nu|\mu_R) \geq \frac{\lambda}{2} \left(W_p(\mu_R, \nu) - W_p(\mu_R, \mu) \right)^2 - 2E_\alpha e^{-\alpha R^2}$$

From this, inequality (2.1) and the elementary inequality

$$\forall a \in (0, 1) \quad \exists C_a > 0; \quad \forall x, y \in \mathbb{R}, \quad (x - y)^2 \geq (1 - a)x^2 - C_a y^2, \quad (2.8)$$

we deduce that for any $\lambda_1 < \lambda$ there exists some constant K such that

$$H(\nu|\mu_R) \geq \frac{\lambda_1}{2} W_p(\mu_R, \nu)^2 - KR^2 e^{-\alpha R^2}. \quad (2.9)$$

Step 2: Covering by small balls. In this second step we derive quantitative estimates on $\widehat{\mu}_R^N$. Let ϕ be a bounded continuous function on B_R , and let \mathcal{B} be a Borel set in $\mathcal{P}(B_R)$ (equipped with the weak topology of convergence against bounded continuous test functions). By Chebyshev's exponential inequality and the independence of the variables X_R^k ,

$$\begin{aligned} \mathbb{P}[\widehat{\mu}_R^N \in \mathcal{B}] &\leq \exp \left(-N \inf_{\nu \in \mathcal{B}} \int_{B_R} \phi d\nu \right) \mathbb{E} \left(e^{N \int_{B_R} \phi d\widehat{\mu}_R^N} \right) \\ &= \exp \left(-N \inf_{\nu \in \mathcal{B}} \left[\int_{B_R} \phi d\nu - \frac{1}{N} \log \mathbb{E} \left(e^{N \int_{B_R} \phi d\widehat{\mu}_R^N} \right) \right] \right) \\ &= \exp \left(-N \inf_{\nu \in \mathcal{B}} \left[\int_{B_R} \phi d\nu - \frac{1}{N} \log \mathbb{E} \left(e^{\sum_{k=1}^N \phi(X_R^k)} \right) \right] \right) \\ &= \exp \left(-N \inf_{\nu \in \mathcal{B}} \left[\int_{B_R} \phi d\nu - \log \int_{B_R} e^\phi d\mu_R \right] \right). \end{aligned}$$

As ϕ is arbitrary, we can pass to the supremum and find

$$\mathbb{P}[\widehat{\mu}_R^N \in \mathcal{B}] \leq \exp \left(-N \sup_{\phi \in C_b(B_R)} \inf_{\nu \in \mathcal{B}} \left[\int_{B_R} \phi d\nu - \log \int_{B_R} e^\phi d\mu_R \right] \right).$$

Now we note that the quantity $\int \phi d\nu - \log \int e^\phi d\mu_R$ is concave in ϕ and linear continuous in ν ; if we further assume that \mathcal{B} is convex and compact, then (for instance) Sion's min-max theorem [26, Theorem 4.2'] ensures that

$$\sup_{\phi \in C_b(B_R)} \inf_{\nu \in \mathcal{B}} \left[\int_{B_R} \phi d\nu - \log \int e^\phi d\mu_R \right] = \inf_{\nu \in \mathcal{B}} \sup_{\phi \in C_b(B_R)} \left[\int_{B_R} \phi d\nu - \log \int e^\phi d\mu_R \right].$$

By the dual formulation of the H functional [12, Lemma 6.2.13], we conclude that

$$\mathbb{P}[\widehat{\mu}_R^N \in \mathcal{B}] \leq \exp \left(-N \inf_{\nu \in \mathcal{B}} H(\nu | \mu_R) \right). \quad (2.10)$$

Now, let $\delta > 0$ and let \mathcal{A} be a measurable subset of $\mathcal{P}(B_R)$. We cover the latter with $\mathcal{N}^{\mathcal{A}}$ balls $(B_i)_{1 \leq i \leq \mathcal{N}^{\mathcal{A}}}$ with radius $\delta/2$ in W_p metric. Each of these balls is convex and compact, and it is included in the δ -thickening of \mathcal{A} in W_p metric, defined as

$$\mathcal{A}_\delta := \left\{ \nu \in \mathcal{P}(B_R); \exists \nu_a \in \mathcal{A}, W_p(\nu, \nu_a) \leq \delta \right\}.$$

So, by (2.10) we get

$$\begin{aligned} \mathbb{P}[\widehat{\mu}_R^N \in \mathcal{A}] &\leq \mathbb{P} \left[\widehat{\mu}_R^N \in \bigcup_{i=1}^{\mathcal{N}^{\mathcal{A}}} B_i \right] \\ &\leq \sum_{i=1}^{\mathcal{N}^{\mathcal{A}}} \mathbb{P}(\widehat{\mu}_R^N \in B_i) \\ &\leq \sum_{i=1}^{\mathcal{N}^{\mathcal{A}}} \exp \left(-N \inf_{\nu \in B_i} H(\nu | \mu_R) \right) \\ &\leq \mathcal{N}^{\mathcal{A}} \exp \left(-N \inf_{\nu \in \mathcal{A}_\delta} H(\nu | \mu_R) \right). \end{aligned} \quad (2.11)$$

We now apply this estimate with

$$\mathcal{A} := \left\{ \nu \in \mathcal{P}(B_R); W_p(\nu, \mu_R) \geq \eta\varepsilon - 2E_\alpha^{1/p} R e^{-\frac{\alpha}{p} R^2} \right\}.$$

From (2.9) we have, for any $\nu \in \mathcal{A}_\delta$,

$$H(\nu | \mu_R) \geq \frac{\lambda_1}{2} W_p(\nu, \mu_R)^2 - KR^2 e^{-\alpha R^2} \geq \frac{\lambda_1}{2} \mu^2 - KR^2 e^{-\alpha R^2},$$

where

$$\mu := \max \left(\eta\varepsilon - 2E_\alpha^{1/p} R e^{-\frac{\alpha}{p} R^2} - \delta, 0 \right).$$

Combining this with (2.11), we conclude that

$$\mathbb{P} \left[W_p(\mu_R, \widehat{\mu}_R^N) \geq \eta\varepsilon - 2E_\alpha^{1/p} R e^{-\frac{\alpha}{p}R^2} \right] \leq \mathcal{N}^{\mathcal{A}} \exp \left(-N \left[\frac{\lambda_1}{2} \mu^2 - KR^2 e^{-\alpha R^2} \right] \right). \quad (2.12)$$

Now, given any $\lambda_2 < \lambda_1$, it follows from (2.8) that there exist δ_1, η_1 and K_1 , depending on $\alpha, \lambda_1, \lambda_2$, such that

$$\frac{\lambda_1}{2} \mu^2 - KR^2 e^{-\alpha R^2} \geq \frac{\lambda_2}{2} \varepsilon^2 - K_1 R^2 e^{-\alpha R^2} \quad (2.13)$$

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

Though this inequality holds independently of p , we shall use it only in the case when $p < 2$. In the case $p = 2$, on the other hand, we note that for any $\eta \in (0, 1)$,

$$\frac{\lambda_1}{2} \mu^2 - KR^2 e^{-\alpha R^2} \geq \frac{\lambda_2}{2} \eta^2 \varepsilon^2 - K_1 R^2 e^{-\alpha R^2} \quad (2.14)$$

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

Finally, we bound $\mathcal{N}^{\mathcal{A}}$ by means of Theorem A.1 in Appendix A: there exists some constant C (only depending on d) such that for all $R > 0$ and $\delta > 0$ the set $\mathcal{P}(B_R)$ can be covered by

$$\left(C \frac{R}{\delta} \vee 1 \right)^{\left(C \frac{R}{\delta} \right)^d}$$

balls of radius δ in W_p metric, where $a \vee b$ stands for $\max(a, b)$. In particular, given $\delta = \delta_1 \varepsilon$, we can choose

$$\mathcal{N}^{\mathcal{A}} \leq \left(K_2 \frac{R}{\varepsilon} \vee 1 \right)^{\left(K_2 \frac{R}{\varepsilon} \right)^d} \quad (2.15)$$

balls of radius δ , for some constant K_2 depending on λ_1 and λ_2 (via δ_1) but neither on ε nor on R . (The purpose of the 1 in $(K_2 R/\varepsilon \vee 1)$ is to make sure that the estimate is also valid when $\varepsilon > R$.)

Combining (2.4), (2.12), (2.13) and (2.15), we find that, given $p \in [1, 2)$, $\lambda_2 < \lambda$ and $\alpha_1 < \alpha < \frac{\lambda}{2}$, there exist some constants K_1, K_2, K_3 and R_1 such that for all $\varepsilon, \zeta > 0$ and $R \geq R_1 \max(1, \zeta^{\frac{1}{2-p}})$,

$$\begin{aligned} \mathbb{P} [W_p(\mu, \widehat{\mu}^N) > \varepsilon] &\leq \left(K_2 \frac{R}{\varepsilon} \vee 1 \right)^{K_2 \left(\frac{R}{\varepsilon} \right)^d} \exp \left(-N \left[\frac{\lambda_2 \varepsilon^2}{2} - K_1 R^2 e^{-\alpha R^2} \right] \right) \\ &\quad + \exp \left(-N \left(K_3 \zeta \varepsilon^p - K_4 e^{(\alpha_1 - \alpha) R^2} \right) \right) \end{aligned} \quad (2.16)$$

for some constant $K_4 = K_4(\theta, \alpha_1)$. In the case when $\underline{p = 2}$, we obtain similarly

$$\begin{aligned} \mathbb{P} [W_2(\mu, \widehat{\mu}^N) > \varepsilon] &\leq \left(K_2 \frac{R}{\varepsilon} \vee 1 \right)^{K_2 \left(\frac{R}{\varepsilon} \right)^d} \exp \left(-N \left[\frac{\lambda_2}{2} \eta^2 \varepsilon^2 - K_1 R^2 e^{-\alpha R^2} \right] \right) \\ &\quad + \exp \left(-N \left(\frac{\alpha_1}{2} (1 - \eta)^2 \varepsilon^2 - K_4 e^{(\alpha_1 - \alpha) R^2} \right) \right) \end{aligned} \quad (2.17)$$

for any $\eta \in (0, 1)$ and $R \geq R_1$.

These estimates are not really appealing (!), but they are rather precise and general. In the rest of the section we shall show that an adequate choice of R leads to a simplified expression.

Step 3: Choice of the parameters.

We first consider the case when $\underline{p \in [1, 2)}$. Let $\lambda' < \lambda_2$, $\alpha' < \alpha$ and $d_1 > d$. We claim that

$$\mathbb{P} [W_p(\mu, \widehat{\mu}^N) > \varepsilon] \leq \exp \left(-\frac{\lambda'}{2} N \varepsilon^2 \right) + \exp(-\alpha' N \varepsilon^2)$$

as soon as

$$R^2 \geq R_2 \max \left(1, \varepsilon^2, \log \left(\frac{1}{\varepsilon^2} \right) \right), \quad N \varepsilon^{d_1+2} \geq K_5 R^{d_1} \quad (2.18)$$

for some constants R_2 and K_5 depending on μ only through λ, α and E_α .

Indeed, on one hand

$$K_2 \left(\frac{R}{\varepsilon} \right)^d \log \left(K_2 \frac{R}{\varepsilon} \right) \leq K_6 \left(\frac{R}{\varepsilon} \right)^{d_1}$$

for some constant K_6 , on the other hand

$$K_1 R^2 e^{-\alpha R^2} \leq e^{-\alpha_1 R^2}$$

for R large enough, and then

$$K_6 \left(\frac{R}{\varepsilon} \right)^{d_1} - N \left[\frac{\lambda_2 \varepsilon^2}{2} - e^{-\alpha_1 R^2} \right] \leq -N \frac{\lambda' \varepsilon^2}{2}$$

for $R^2 / \log(\frac{1}{\varepsilon^2})$ and $N \varepsilon^{d_1+2} / R^{d_1}$ large enough; this is enough to bound the first term in the right-hand side of (2.16) if moreover R/ε is large enough.

Moreover, letting $\alpha_2 \in (\alpha', \alpha_1)$, we can choose ζ in such a way that $K_3 \zeta = \varepsilon^{2-p}$, so that

$$\exp \left(-N \left(K_3 \zeta \varepsilon^p - K_4 e^{(\alpha_1 - \alpha) R^2} \right) \right) = \exp \left(-N \left(\alpha_2 \varepsilon^2 - K_4 e^{(\alpha_1 - \alpha) R^2} \right) \right),$$

which in the end can be bounded by

$$\exp(-N \alpha' \varepsilon^2)$$

if R and $R^2 / \log(\frac{1}{\varepsilon^2})$ are large enough. With this one can get a bound on the right-hand side of (2.16).

Now let us check that conditions (2.18) can indeed be fulfilled. Clearly, the first condition holds true for all $\varepsilon \in (0, 1)$ and $R^2 \geq R_3 \log(\frac{K_6}{\varepsilon^2})$, where R_3 and K_6 are positive constants. Then, we can choose

$$R := \left(\frac{N}{K_5} \varepsilon^{d_1+2} \right)^{1/d_1}$$

so that the second condition holds as an equality. This choice is admissible as soon as

$$\left(\frac{N}{K_5} \varepsilon^{d_1+2} \right)^{2/d_1} \geq R_3 \log \left(\frac{K_5}{\varepsilon^2} \right)$$

and this, in turn, holds true as soon as

$$N \geq K_7 \varepsilon^{-(d'+2)}, \quad (2.19)$$

where d' is such that $d' > d$, and K_7 is large enough.

If $\varepsilon \geq 1$, then we can choose $R^2 = R_2 \varepsilon^2$, i.e. $R = \sqrt{R_2} \varepsilon$, and then the second inequality in (2.18) will be true as soon as N is large enough.

To sum up: Given $d' > d$, $\lambda' < \lambda$ and $\alpha' < \alpha$, there exists some constant N_0 , depending on d' and depending on μ only through λ, α and E_α , such that for all $\varepsilon > 0$,

$$\mathbb{P} [W_p(\mu, \widehat{\mu}^N) > \varepsilon] \leq \exp \left(-\frac{\lambda'}{2} N \varepsilon^2 \right) + \exp (-\alpha' N \varepsilon^2)$$

as soon as $N \geq N_0 \max(\varepsilon^{-(d'+2)}, 1)$. Then we note that, given $K < \min \left(\frac{\lambda'}{2}, \alpha' \right)$, the inequality

$$\exp \left(-\frac{\lambda'}{2} N \varepsilon^2 \right) + \exp (-\alpha' N \varepsilon^2) \leq \exp (-K N \varepsilon^2)$$

holds if condition (2.19) is satisfied for some K_7 large enough. To conclude the proof of Theorem 1.1 in the case when $p \in [1, 2)$, it is sufficient to choose $\lambda' < \lambda$, $\alpha < \lambda/2$.

Now, in the case when $\underline{p} = 2$, given $\lambda_3 < \lambda_2$ and $\alpha_2 < \alpha_1$, conditions (2.18) imply

$$\mathbb{P} [W_2(\mu, \widehat{\mu}^N) > \varepsilon] \leq \exp \left(-\frac{\lambda_3}{2} \eta^2 N \varepsilon^2 \right) + \exp \left(-\frac{\alpha_2}{2} (1 - \eta)^2 N \varepsilon^2 \right).$$

Then we let $\alpha_2 := \frac{\lambda_3}{2}$ and $\eta := \sqrt{2} - 1$, so that

$$\frac{\lambda_3}{2} \eta^2 = \frac{\alpha_2}{2} (1 - \eta)^2.$$

Then

$$\mathbb{P} [W_2(\mu, \widehat{\mu}^N) > \varepsilon] \leq 2 \exp \left(-(3 - 2\sqrt{2}) \frac{\lambda_3}{2} N \varepsilon^2 \right);$$

for $\lambda' < \lambda$, the above quantity is bounded by

$$\exp \left(-(3 - 2\sqrt{2}) \frac{\lambda'}{2} N \varepsilon^2 \right)$$

as soon as (2.19) is enforced with K_7 large enough. This concludes the argument.

2.2. Proof of Theorem 1.5. It is very similar to the proof of Theorem 1.1, so we shall only explain where the differences lie. Obviously, the main difficulty will consist in the control of tails.

We first let $p \in [1, q)$, $\alpha \in [1, \frac{q}{p})$ and $R > 0$, and introduce

$$M_q := \int_{\mathbb{R}^d} |x|^q d\mu(x).$$

Then (2.1) may be replaced by

$$W_p^p(\mu, \mu_R) \leq 2^p M_q R^{p-q}, \quad (2.20)$$

and (2.2) by

$$\mathbb{P} [W_p(\hat{\mu}_R^N, \hat{\mu}^N) > \varepsilon] \leq C N^{\bar{\alpha}-\alpha} \frac{R^{\alpha p-q}}{(\varepsilon^p - C R^{p-q})^\alpha} \quad (2.21)$$

for some constant C depending on α and M_q .

Let us establish for instance (2.21). Introduce

$$Z_k = |Y_k - X_k|^p \mathbf{1}_{|X_k|>R} \quad (1 \leq k \leq N).$$

By Chebychev's inequality,

$$\begin{aligned} \mathbb{P} [W_p(\hat{\mu}_R^N, \hat{\mu}^N) > \varepsilon] &\leq \mathbb{P} \left[\frac{1}{N} \sum_{k=1}^N Z_k > \varepsilon^p \right] = \mathbb{P} \left[\frac{1}{N} \sum_{k=1}^N (Z_k - \mathbb{E} Z_k) > \varepsilon^p - \mathbb{E} Z_1 \right] \\ &\leq \frac{\mathbb{E} \left| \sum_{k=1}^N (Z_k - \mathbb{E} Z_k) \right|^\alpha}{(N (\varepsilon^p - \mathbb{E} Z_1))^\alpha} \end{aligned}$$

provided that $\varepsilon^p > \mathbb{E} Z_1$. But, since the random variables $(Z_k - \mathbb{E} Z_k)_k$ are independent and identically distributed, with zero mean, there exists some constant C depending on α such that

$$\mathbb{E} \left| \sum_{k=1}^N (Z_k - \mathbb{E} Z_k) \right|^\alpha \leq C N^{\bar{\alpha}} \mathbb{E} |Z_1 - \mathbb{E} Z_1|^\alpha$$

where $\bar{\alpha} := \max(\alpha/2, 1)$. This inequality is a consequence of Rosenthal's inequality in the case when $\alpha \geq 2$, but also holds true if $\alpha \in [1, 2)$ (see for instance [24, pp. 62 and 82]). Then, on one hand,

$$\mathbb{E} Z_1 = \mathbb{E} |Y_1 - X_1|^p \mathbf{1}_{|X_1|>R} \leq 2^p M_q R^{p-q},$$

while on the other hand,

$$\begin{aligned} \mathbb{E} |Z_1 - \mathbb{E} Z_1|^\alpha &= \mathbb{E} \left| |Y_1 - X_1|^p \mathbf{1}_{|X_1|>R} - \mathbb{E} |Y_1 - X_1|^p \mathbf{1}_{|X_1|>R} \right|^\alpha \\ &\leq C \mathbb{E} |Y_1 - X_1|^{\alpha p} \mathbf{1}_{|X_1|>R} \leq C M_q R^{\alpha p-q} \end{aligned}$$

with C standing for various constants. Collecting these two estimates, we conclude to the validity of (2.21) for $R^{q-p} \varepsilon^p$ large enough.

Then (2.20) and (2.21) together ensure that

$$\mathbb{P}[W_p(\mu, \widehat{\mu}^N) > \varepsilon] \leq \mathbb{P}[W_p(\mu_R, \widehat{\mu}_R^N) > \eta \varepsilon - 2M_q^{1/p} R^{1-q/p}] + C N^{\bar{\alpha}-\alpha} \frac{R^{\alpha p-q}}{((1-\eta)^p \varepsilon^p - C R^{p-q})^\alpha} \quad (2.22)$$

for any $\varepsilon \in (0, 1)$, $\eta > 0$ and $R^{q-p} \varepsilon^p (1-\eta)^p$ large enough.

Since μ_R is supported in B_R , the Csiszár-Kullback-Pinsker inequality and Kantorovich-Rubinstein formulation of the W_1 distance together ensure that it satisfies a $T_1(R^{-2})$ inequality (see e.g. [8, Particular Case 5] with $p = 1$). This estimate also extends to any W_p distance, not as a penalized T_p inequality as in (2.9), but rather as

$$W_p^{2p}(\nu, \mu_R) \leq 2^{2p-1} R^{2p} H(\nu|\mu_R) \quad (2.23)$$

(see again [8, Particular Case 5]).

From (2.22) and (2.23) we deduce (as in (2.17)) that

$$\mathbb{P}[W_p(\widehat{\mu}^N, \mu) > \varepsilon] \leq \left(K_1 \frac{R}{\delta}\right)^{K_1 \left(\frac{R}{\delta}\right)^d} \exp\left(-\frac{N \mu^{2p}}{2^{2p-1} R^{2p}}\right) + C N^{\bar{\alpha}-\alpha} \frac{R^{\alpha p-q}}{((1-\eta)^p \varepsilon^p - C R^{p-q})^\alpha} \quad (2.24)$$

for any δ , where now

$$\mu := (\eta \varepsilon - 2M_q^{1/p} R^{1-q/p} - \delta)^+.$$

Letting $\eta_1 < \eta$ and $d' > d$, and choosing $\delta = \delta_0 \varepsilon$, we deduce

$$\mathbb{P}[(W_p(\widehat{\mu}^N, \mu) > \varepsilon] \leq \exp\left(\left(\frac{R}{\varepsilon}\right)^{d'} - \frac{\eta_1^{2p}}{2^{2p-1}} \frac{N \varepsilon^{2p}}{R^{2p}} + \frac{K_1}{2^{2p-1}} \frac{N}{R^{2q}}\right) + C N^{\bar{\alpha}-\alpha} \frac{R^{\alpha p-q}}{((1-\eta_1)^p \varepsilon^p - C R^{p-q})^\alpha}$$

for $R^{q-p} \varepsilon^p (1-\eta_1)^p$ large enough, and then

$$\mathbb{P}[W_p(\widehat{\mu}^N, \mu) > \varepsilon] \leq \exp\left(-\frac{\eta_2^{2p}}{2^{2p-1}} \frac{N \varepsilon^{2p}}{R^{2p}}\right) + C N^{\bar{\alpha}-\alpha} \frac{R^{\alpha p-q}}{(1-\eta_2)^{\alpha p} \varepsilon^{\alpha p}} \quad (2.25)$$

for $\eta_2 < \eta_1$, provided that the conditions

$$R \geq R_1 \varepsilon^{-\frac{p}{q-p}}, \quad N \geq K_2 \left(\frac{R}{\varepsilon}\right)^{2p+d'} \quad (2.26)$$

hold for some R_1 and K_2 .

Given any choice of R as a product of powers of N and ε , the first term in the right-hand side of (2.25) will always be smaller than the second one, if N goes to infinity while ε is kept fixed; thus we can choose R minimizing the second term under the above conditions. Then the second condition in (2.26) will be fulfilled as an equality:

$$R = K_3 \varepsilon N^{\frac{1}{2p+d'}}.$$

As for the first condition in (2.26), it can be rewritten as

$$N \geq N_0 \varepsilon^{-q \frac{2p+d'}{q-p}},$$

and then, by (2.25),

$$\mathbb{P} [W_p(\widehat{\mu}^N, \mu) > \varepsilon] \leq \exp \left(-K_5 N^{\frac{d'}{2p+d'}} \right) + K_6 \varepsilon^{-q} N^{\bar{\alpha}-\alpha+\frac{\alpha p-q}{2p+d'}}.$$

Hence

$$\mathbb{P} [W_p(\widehat{\mu}^N, \mu) > \varepsilon] \leq \varepsilon^{-q} N^{\bar{\alpha}-\alpha} \quad (2.27)$$

for all $\varepsilon \in (0, 1)$ and N larger than some constant and, given $d' > d$, for all $\varepsilon \geq 1$ and $N \geq M\varepsilon^{d'-d}$ where M is large enough.

In the first case when $p \geq q/2$, any admissible α belongs to $[1, q/p) \subset [1, 2]$, so $\bar{\alpha} = 1$. If $\delta \in (0, q/p - 1)$, we get from (2.27), with $\alpha = q/p - \delta$, that

$$\mathbb{P} [W_p(\widehat{\mu}^N, \mu) > \varepsilon] \leq \varepsilon^{-q} N^{1-q/p+\delta}$$

for all $\varepsilon > 0$ and

$$N \geq N_0 \max(\varepsilon^{-q\frac{2p+d'}{q-p}}, \varepsilon^{d'-d}).$$

In the second case when $p < q/2$, we only consider admissible α 's in $[2, q/p) \subset [1, q/p)$, so that $\bar{\alpha} - \alpha = -\alpha/2$. Choosing $\delta \in (0, q/p - 2)$, we get from (2.27)

$$\mathbb{P} [W_p(\widehat{\mu}^N, \mu) > \varepsilon] \leq \varepsilon^{-q} N^{-q/2p+\delta/2}$$

under the same conditions on N as before. This concludes the argument.

2.3. Proof of Theorem 1.6. It is again based on the same principles as the proofs of Theorems 1.1 and 1.5, with the help of functional inequalities investigated in [8] and [11]. We skip the argument, which the reader can easily reconstruct by following the same lines as above.

2.4. Data reconstruction estimates. Finally, we show how the above concentration estimates imply data reconstruction estimates. This is a rather general estimate, which is treated here along the lines of [25, Section 5] and [29, Problem 10].

Proposition 2.1. *Let μ be a probability measure on \mathbb{R}^d , with density f with respect to Lebesgue measure. Let X_1, \dots, X_N be random points in \mathbb{R}^d , and let ζ be a Lipschitz, non-negative kernel with unit integral. Define the random measure $\widehat{\mu}$ and the random function $\widehat{f}_{\zeta, \alpha}$ by*

$$\widehat{\mu} := \frac{1}{N} \sum_{i=1}^N \delta_{X_i}, \quad \widehat{f}_{\zeta, \alpha}(x) := \frac{1}{N} \sum_{i=1}^N \zeta_{\alpha}(x - X_i), \quad \zeta_{\alpha}(x) = \frac{1}{\alpha^d} \zeta\left(\frac{x}{\alpha}\right).$$

Then,

$$\sup_{x \in \mathbb{R}^d} |\widehat{f}_{\zeta, \alpha}(x) - f(x)| \leq \frac{\|\zeta\|_{\text{Lip}}}{\alpha^{d+1}} W_1(\widehat{\mu}, \mu) + \delta(\alpha), \quad (2.28)$$

where δ stands for the modulus of continuity of f , defined as

$$\delta(\varepsilon) := \sup_{|x-y| \leq \varepsilon} |f(x) - f(y)|.$$

As a consequence, if f is Lipschitz, then there exist some constants $a, K > 0$, only depending on d , $\|f\|_{\text{Lip}}$ and $\|\zeta\|_{\text{Lip}}$, such that

$$\mathbb{P}\left[\|\widehat{f}_{\zeta, a\varepsilon} - f\|_{L^\infty} > \varepsilon\right] \leq \mathbb{P}\left[W_1(\widehat{\mu}, \mu) > K\varepsilon^{d+2}\right] \quad (2.29)$$

for all $\varepsilon > 0$.

Proof. First,

$$\begin{aligned} |\mu * \zeta_\alpha(x) - f(x)| &= \left| \int_{\mathbb{R}^d} \zeta_\alpha(x-y) (f(y) - f(x)) dy \right| \\ &\leq \int_{\mathbb{R}^d} \zeta_\alpha(x-y) |f(y) - f(x)| dy. \end{aligned}$$

Since $\zeta_\alpha(x-y)$ is supported in $\{|x-y| \leq \alpha\}$, and ζ_α is a probability density, we deduce

$$|\mu * \zeta_\alpha(x) - f(x)| \leq \delta(\alpha). \quad (2.30)$$

Now, if x is some point in \mathbb{R}^d , then, thanks to the Kantorovich-Rubinstein dual formulation (1.2),

$$\begin{aligned} \left| \widehat{f}_{\zeta, \alpha} - \mu * \zeta_\alpha \right|(x) &= \left| \int_{\mathbb{R}^d} \zeta_\alpha(x-y) d[\widehat{\mu} - \mu](y) \right| \\ &\leq \|\zeta_\alpha(x - \cdot)\|_{\text{Lip}} W_1(\widehat{\mu}, \mu) \\ &= \frac{\|\zeta\|_{\text{Lip}}}{\alpha^{d+1}} W_1(\widehat{\mu}, \mu). \end{aligned}$$

To conclude the proof of (2.28), it suffices to combine this bound with (2.30).

Now, let $L := \max(\|f\|_{\text{Lip}}, \|\zeta\|_{\text{Lip}})$, and $\alpha := \varepsilon/(2L)$. The bound (2.28) turns into

$$\|\widehat{f}_{\zeta, \alpha} - f\|_{L^\infty} \leq L \left(\frac{W_1(\widehat{\mu}, \mu)}{\alpha^{d+1}} + \alpha \right) \leq \left(\frac{(2L)^{d+1} L}{\varepsilon^{d+1}} \right) W_1(\widehat{\mu}, \mu) + \frac{\varepsilon}{2}.$$

In particular,

$$\mathbb{P}\left[\|\widehat{f}_{\zeta, \alpha} - f\|_{L^\infty} > \varepsilon\right] \leq \mathbb{P}\left[W_1(\widehat{\mu}, \mu) > \frac{\varepsilon^{d+2}}{(2L)^{d+2}}\right],$$

which is estimate (2.29). □

Remark 2.2. Estimate (2.29), combined with Theorem 1.1 or Theorem 1.5, yields simple quantitative (non-asymptotic) deviation inequalities for empirical distribution functions in supremum norm. We refer to Gao [15] for a recent study of deviation inequalities for empirical distribution functions, both in moderate and large deviations regimes.

3. PDE ESTIMATES

Now we start the study of our model system for interacting particles. The first step towards our proof of Theorem 1.7 consists in deriving suitable a priori estimates on the solution to the nonlinear limit partial differential equation (1.8). In this section, we recall some estimates which have already been established by various authors, and derive some new ones. All estimates will be effective.

3.1. Notation. In the sequel, μ_0 is a probability measure, taken as an initial datum for equation (1.8), and various regularity assumptions will later be made on μ_0 . Assumptions (1.9) will always be made on V and W , even if they are not recalled explicitly; we shall only mention additional regularity assumptions, when used in our estimates. Moreover, we shall write

$$\Gamma := \max(|\gamma|, |\gamma'|). \quad (3.1)$$

The notation μ_t will always stand for the solution (unique under our assumptions) of (1.8).

We also write

$$e(t) := \int_{\mathbb{R}^d} |x|^2 d\mu_t(x)$$

for the (kinetic) energy associated with μ_t , and

$$M_\alpha(t) := \int_{\mathbb{R}^d} e^{\alpha|x|^2} d\mu_t(x)$$

for the square exponential moment of order α .

The scalar product between two vectors $v, w \in \mathbb{R}^d$ will be denoted by $v \cdot w$. The symbols C and K will often be used to denote various positive constants; in general what will matter is an upper bound on constants denoted C , and a lower bound on constants denoted K . The space C^k is the space of k times differentiable continuous functions.

3.2. Decay at infinity. In this subsection, we prove the propagation of strong decay estimates at infinity:

Proposition 3.1. *With the conventions of Subsection 3.1, let $\bar{\eta}$ be $-\gamma$ if $\gamma < 0$, and an arbitrary negative number otherwise. Let*

$$a := 2(\beta + \bar{\eta}), \quad \bar{G} := 2d + \frac{|\nabla V(0)|^2}{2|\bar{\eta}|}.$$

Then

$$(i) \quad e(t) \leq e^{-at} \left[e(0) + \bar{G} \frac{e^{at} - 1}{a} \right];$$

(ii) For any $\alpha_0 > 0$ there is a continuous positive function $\alpha(t)$ such that $\alpha(0) = \alpha_0$ and

$$M_{\alpha_0}(0) < +\infty \implies M_{\alpha(t)}(t) < +\infty. \quad (3.2)$$

(iii) Moreover, in the “uniformly convex case” when $\beta > 0$ and $\beta + \gamma > 0$, then there is $\alpha > 0$ such that

$$\sup_{t \geq 0} e(t) < +\infty, \quad \sup_{t \geq 0} M_\alpha(t) < +\infty.$$

Corollary 3.2. *If μ_0 admits a finite square exponential moment, then μ_t satisfies $T_1(\lambda_t)$, for some function $\lambda_t > 0$, bounded below on any interval $[0, T]$ ($T < \infty$).*

Proof. We start with (i). For simplicity we shall pretend that μ_t is a smoothly differentiable function of t , with rapid decay, so that all computations based on integrating equation (1.8) against $|x|^2$ are justified. These assumptions are not a priori satisfied, but the resulting bounds can easily be rigorously justified with standard but tedious approximation arguments. With that in mind, we compute

$$e'(t) = 2d - 2 \int_{\mathbb{R}^d} (x \cdot \nabla V(x) + x \cdot \nabla W * \mu_t(x)) d\mu_t(x)$$

with

$$-2 \int_{\mathbb{R}^d} x \cdot \nabla V(x) d\mu_t(x) \leq -2\beta \int_{\mathbb{R}^d} |x|^2 d\mu_t(x) - 2\nabla V(0) \cdot \int_{\mathbb{R}^d} x d\mu_t(x).$$

Since ∇W is an odd function, we have

$$\begin{aligned} -2 \int_{\mathbb{R}^d} x \cdot \nabla W * \mu_t(x) d\mu_t(x) &= -2 \iint x \cdot \nabla W(x - y) d\mu_t(y) d\mu_t(x) \\ &= - \iint (x - y) \cdot \nabla W(x - y) d\mu_t(y) d\mu_t(x) \\ &\leq -\gamma \iint |x - y|^2 d\mu_t(y) d\mu_t(x) \\ &= -2\gamma \left[\int |x|^2 d\mu_t(x) - \left| \int x d\mu_t(x) \right|^2 \right]. \end{aligned}$$

If $\gamma < 0$, then

$$\begin{aligned} e'(t) &\leq 2d - 2(\gamma + \beta)e(t) + 2\gamma \left| \int x d\mu_t(x) \right|^2 + \frac{|\nabla V(0)|^2}{2|\gamma|} \\ &\leq 2d - 2(\gamma + \beta)e(t) + \frac{|\nabla V(0)|^2}{2|\gamma|}, \end{aligned}$$

and if $\gamma \geq 0$, then for any $\bar{\eta} < 0$

$$\begin{aligned} e'(t) &\leq 2d - 2(\bar{\eta} + \beta)e(t) - 2\gamma \left(\int |x|^2 d\mu_t(x) - \left| \int x d\mu_t(x) \right|^2 \right) + \frac{|\nabla V(0)|^2}{2|\bar{\eta}|} \\ &\leq 2d - 2(\bar{\eta} + \beta)e(t) + \frac{|\nabla V(0)|^2}{2|\bar{\eta}|}. \end{aligned}$$

This leads to

$$e'(t) \leq \bar{G} - a e(t),$$

and the conclusion follows easily by Gronwall's lemma.

We now turn to (ii). Let α be some arbitrary nonnegative C^1 function on \mathbb{R}_+ . By using the equation (1.8), we compute

$$\frac{d}{dt} \int e^{\alpha(t)|x|^2} d\mu_t(x) = \int [2d\alpha + 4\alpha^2|x|^2 - 2\alpha x \cdot \nabla V(x) - 2\alpha x \cdot \nabla W * \mu_t(x) + \alpha'(t)|x|^2] e^{\alpha|x|^2} d\mu_t(x).$$

Since $D^2V(x) \geq \beta I$ for all $x \in \mathbb{R}^d$, we can write

$$-x \cdot \nabla V(x) \leq -x \cdot \nabla V(0) - \beta|x|^2 \leq -\beta|x|^2 + |\nabla V(0)||x| \leq (\delta - \beta)|x|^2 + \frac{C}{4\delta} \quad (3.3)$$

for any $\delta > 0$ and $x \in \mathbb{R}^d$.

Next, our assumptions on W imply $\nabla W(0) = 0$, and $\gamma I \leq D^2W(x) \leq \gamma' I$, so

$$x \cdot \nabla W(x) \geq \gamma|x|^2 \quad \text{and} \quad |x \cdot D^2W(z)y| \leq \Gamma|x||y|$$

for all $x, y, z \in \mathbb{R}^d$, with Γ defined by (3.1). Hence, by Taylor's formula,

$$\begin{aligned} -x \cdot \nabla W * \mu_t(x) &= - \int_{\mathbb{R}^d} x \cdot \nabla W(x-y) d\mu_t(y) \\ &= -x \cdot \nabla W(x) + \int_{\mathbb{R}^d} \int_0^1 x \cdot D^2W(x-sy)y d\mu_t(y) ds \\ &\leq -\gamma|x|^2 + \Gamma|x| \int_{\mathbb{R}^d} |y| d\mu_t(y) \\ &\leq (-\gamma + \Gamma\eta)|x|^2 + \frac{\Gamma}{4\eta} e(t), \end{aligned} \quad (3.4)$$

where η is any positive number.

From (3.3) and (3.4) we obtain

$$\frac{d}{dt} \left(M_{\alpha(t)}(t) \right) \leq \int_{\mathbb{R}^d} [A(t) + B(t)|x|^2] e^{\alpha(t)|x|^2} d\mu_t(x) \quad (3.5)$$

where

$$A(t) = C\alpha(t)(1 + e(t)), \quad B(t) = \alpha'(t) + 4\alpha(t)^2 + b\alpha(t),$$

and C is a finite constant, while $b = -2(\gamma + \beta - \delta - \Gamma\eta)$.

We now choose $\alpha(t)$ in such a way that $B(t) \equiv 0$, i.e.

$$\alpha'(t) + 4\alpha^2(t) + b\alpha(t) = 0, \quad \alpha(0) = \alpha_0.$$

This integrates to

$$\alpha(t) = e^{-bt} \left(\frac{1}{\alpha_0} + 4 \frac{1 - e^{-bt}}{b} \right)^{-1} \quad \left(= \left(\frac{1}{\alpha_0} + 4t \right)^{-1} \text{ if } b = 0 \right).$$

Obviously α is a continuous positive function, and our estimates imply

$$\frac{d}{dt} \left(M_{\alpha(t)}(t) \right) \leq A(t)M_{\alpha(t)}(t).$$

We conclude by using Gronwall's lemma that

$$M_{\alpha(t)}(t) \leq \exp\left(\int_0^t A(s) ds\right) M_{\alpha_0}(0).$$

Next, the estimate (iii) for $e(t)$ is an easy consequence of our explicit estimates when $\beta > 0, \beta + \gamma > 0$ (in the case when $\gamma \geq 0$ and $\beta > 0$, we choose $\bar{\eta} \in (0, \beta)$).

As for the estimate about $M_{\alpha}(t)$, it will result from a slightly more precise computation. From (3.5), we have

$$\frac{d}{dt} \int_{\mathbb{R}^d} e^{\alpha|x|^2} d\mu_t(x) \leq \int_{\mathbb{R}^d} [A(t) + B|x|^2] e^{\alpha|x|^2} d\mu_t(x) \quad (3.6)$$

where A is bounded on \mathbb{R}_+ by some constant a , and

$$B = 2\alpha [2\alpha - (\beta + \gamma - \delta - \Gamma\eta)].$$

Since $\beta + \gamma > 0$, for any fixed α in $\left(0, \frac{\beta + \gamma}{2}\right)$ we can choose $\delta, \eta > 0$ such that $B < 0$.

Letting $R^2 = -a/B$ and $G = -B > 0$, equation (3.6) becomes

$$\frac{d}{dt} \int_{\mathbb{R}^d} e^{\alpha|x|^2} d\mu_t(x) \leq G \int_{\mathbb{R}^d} (R^2 - |x|^2) e^{\alpha|x|^2} d\mu_t(x). \quad (3.7)$$

Let $p > 1$. The formula

$$\begin{aligned} \int_{|x|>pR} (R^2 - |x|^2) e^{\alpha|x|^2} d\mu_t(x) &\leq R^2(1 - p^2) \int_{|x|>pR} e^{\alpha|x|^2} d\mu_t(x) \\ &= R^2(1 - p^2) \left[\int_{\mathbb{R}^d} e^{\alpha|x|^2} d\mu_t(x) - \int_{|x|\leq pR} e^{\alpha|x|^2} d\mu_t(x) \right] \end{aligned}$$

leads to

$$\int_{\mathbb{R}^d} (R^2 - |x|^2) e^{\alpha|x|^2} d\mu_t(x) \leq \int_{|x|\leq pR} (R^2 p^2 - |x|^2) e^{\alpha|x|^2} d\mu_t(x) + R^2(1 - p^2) M_{\alpha}.$$

by decomposing the integral on the sets $\{|x| \leq pR\}$ and $\{|x| > pR\}$. From (3.7) we deduce

$$(M_{\alpha})'(t) + \omega_1 M_{\alpha}(t) \leq \omega_2$$

where ω_1 and ω_2 are positive constants. It follows that $M_{\alpha}(t)$ remains bounded on \mathbb{R}_+ if $M_{\alpha}(0) < +\infty$, and this concludes the argument. \square

3.3. Time-regularity. Now we study the time-regularity of μ_t .

Proposition 3.3. *With the conventions of Subsection 3.1, for any $T < +\infty$ there exists a constant $C(T)$ such that*

$$\forall s, t \in [0, T], \quad W_1(\mu_t, \mu_s) \leq C(T) |t - s|^{1/2}. \quad (3.8)$$

Remark 3.4. The exponent $1/2$ is natural in small time if no regularity assumption is made on μ_0 ; it can be improved if t, s are assumed to be bounded below by some $t_0 > 0$. Also, in view of the results of convergence to equilibrium recalled later on, the constant $C(T)$ might be chosen independent of T if $\beta > 0, \beta + 2\gamma > 0$.

Remark 3.5. A stochastic proof of (3.8) is possible, via the study of continuity estimates for Y_t , which in any case will be useful later on. But here we prefer to present an analytical proof, to stress the fact that estimates in this section are purely analytical statements.

Proof. Let L be the linear operator $-\Delta - \nabla \cdot (\nabla V + \nabla(W * \mu_t))$, and let e^{-tL} be the associated semigroup: from our assumptions and estimates it follows that it is well-defined, at least for initial data which admit a finite square exponential moment. Of course $\mu_t = e^{-tL}\mu_0$. It follows that

$$\begin{aligned} W_2(\mu_s, \mu_t) &= W_2(\mu_s, e^{-(t-s)L}\mu_s) = W_2\left(\int_{\mathbb{R}^d} \delta_y d\mu_s(y), \int_{\mathbb{R}^d} e^{-(t-s)L}\delta_y d\mu_s(y)\right) \\ &\leq \int_{\mathbb{R}^d} W_2(\delta_y, e^{-(t-s)L}\delta_y) d\mu_s(y). \end{aligned}$$

Our goal is to bound this by $O(\sqrt{t-s})$. In view of Proposition 3.1, it is sufficient to prove that for all $a > 0$,

$$W_2^2(\delta_y, e^{-(t-s)L}\delta_y) = O(t-s) O(e^{a|y|^2}).$$

This estimate is rather easy, since the left-hand side is just the variance of the solution of a linear diffusion equation, starting with a Dirac mass at y as initial datum. Without loss of generality, we assume $s = 0$, and write $\tilde{\mu}_t := e^{-tL}\delta_y$. For simplicity we write the computations in a sketchy way, but they are not hard to justify.

Since the initial datum is δ_y , its square exponential moment \tilde{M}_α of order α is $e^{\alpha|y|^2}$. With an argument similar to the proof of Proposition 3.1(ii), one can show that

$$0 \leq t \leq T \implies \int e^{\alpha|x|^2} d\tilde{\mu}_t(x) \leq C(T)(1 + \tilde{M}_\alpha) \leq C(T) e^{\alpha|y|^2}.$$

Now, since $|\nabla V|(x) = O(e^{a|x|^2})$, $a < \alpha$, $|\nabla W * \mu_t|$ grows at most polynomially, and $\tilde{\mu}_t$ admits a square exponential moment of order α , we easily obtain

$$\frac{d}{dt} \int x d\tilde{\mu}_t = - \int \nabla(V + W * \mu_t) d\tilde{\mu}_t = \int O(e^{a|x|^2}) d\tilde{\mu}_t = O(e^{\alpha|y|^2});$$

$$\frac{d}{dt} \int \frac{|x|^2}{2} d\tilde{\mu}_t = d - \int x \cdot \nabla(V + W * \mu_t) d\tilde{\mu}_t = O(e^{\alpha|y|^2}).$$

From these estimates we deduce that the time-derivative of the variance $V(\tilde{\mu}_t) := \int |x|^2 d\tilde{\mu}_t - (\int x d\tilde{\mu}_t)^2$ is bounded by $O(e^{b|y|^2})$ for any $b > 0$. Since $\tilde{\mu}_0$ has zero variance, it follows that the variance of $\tilde{\mu}_t$ is $O(te^{b|y|^2})$, which was our goal. \square

3.4. Regularity in phase space. Regularity estimates will be useful for Theorem 1.11. Equation (1.8) is a (weakly nonlinear) parabolic equation, for which regularization effects can be studied by standard tools. Some limits to the strength of the regularization are imposed by the regularity of V . So as not to be bothered by these nonessential considerations, we shall assume strong regularity conditions on V here. Then in Appendix B we shall prove the following estimates:

Proposition 3.6. *With the conventions of Subsection 3.1, assume in addition that V has all its derivatives growing at most polynomially at infinity. Then, for each $k \geq 0$ and for all $t_0 > 0$, $T > t_0$ there is a finite constant $C(t_0, T)$, only depending on t_0, T, k and a square exponential moment of the initial measure μ_0 , such that the density f_t of μ_t is of class \mathcal{C}^k , with*

$$\sup_{t_0 \leq t \leq T} \|f_t\|_{\mathcal{C}^k} \leq C(t_0, T).$$

If moreover $\beta > 0$, $\beta + \gamma > 0$, then $C(t_0, T)$ can be chosen to be independent of T for any fixed t_0 .

Remark 3.7. For regular initial data and under some adequate assumptions on V and W , some regularity estimates on f_t/f_∞ , where f_∞ is the limit density in large time, are established in [10, Lemma 6.7]. These estimates allow a much more precise uniform decay, but are limited to just one derivative. Here there will be no need for them.

3.5. Asymptotic behavior. In the “uniformly convex” case when $\beta + \gamma > 0$, the measure μ_t converges to a definite limit μ_∞ as $t \rightarrow \infty$. This was investigated in [19, 9, 10]. The following statement is a simple variant of [9, Theorems 2.1 and 5.1].

Proposition 3.8. *With the conventions of Subsection 3.1, assuming that $\beta > 0$, $\beta + 2\gamma > 0$, there exists a probability measure μ_∞ such that*

$$W_2(\mu_t, \mu_\infty) \leq C e^{-\lambda t}, \quad \lambda > 0.$$

Here the constants C and λ only depend on the initial datum μ_0 .

4. THE LIMIT EMPIRICAL MEASURE

Consider the random time-dependent measure

$$\widehat{\nu}_t^N := \frac{1}{N} \sum_{i=1}^N \delta_{Y_t^i}, \quad (4.1)$$

where $(Y_t^i)_{t \geq 0}$, $1 \leq i \leq N$, are N independent processes solving the same stochastic differential equation

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

and such that the law of Y_0^i is μ_0 . As we already mentioned, for each t and i , Y_t^i is distributed according to the law μ_t . We call $\widehat{\nu}_t^N$ the “limit empirical measure” because it is expected to be a rather accurate description, in some well-chosen sense, of the empirical measure $\widehat{\mu}_t^N$ as $N \rightarrow \infty$.

Our estimates on μ_t , and the fact that $\widehat{\nu}_t^N$ is the empirical measure for *independent* processes, are sufficient to imply good properties of concentration of $\widehat{\nu}_t^N$ around its mean μ_t , as $N \rightarrow \infty$, for each t . But later on we shall use some estimates about the time-dependent measure (even to obtain a result of concentration for $\widehat{\mu}_t^N$ with fixed t). To get such results, we shall study the time-regularity of $\widehat{\nu}_t^N$. Our final goal in this section is the following

Proposition 4.1. *With the conventions of Subsection 3.1, for any $T \geq 0$ there are constants $C = C(T)$ and $a = a(T) > 0$ such that the limit empirical measure (4.1) satisfies*

$$\forall \Delta \in [0, T], \forall \varepsilon > 0, \quad \mathbb{P} \left[\sup_{t_0 \leq s, t \leq t_0 + \Delta} W_1(\widehat{\nu}_s^N, \widehat{\nu}_t^N) > \varepsilon \right] \leq \exp(-N(a\varepsilon^2 - C\Delta)).$$

To prove Proposition 4.1, we shall use a bit of classical stochastic calculus tools.

4.1. SDE estimates. In this subsection we establish the following estimates of time regularity for the stochastic process Y_t : For all $T > 0$, there exist positive constants a and C such that, for all $s, t, t_0, \Delta \in [0, T]$,

- (i) $\mathbb{E} |Y_t - Y_s|^2 \leq C|t - s|$
- (ii) $\mathbb{E} |Y_t - Y_s|^4 \leq C|t - s|^2$
- (iii) $\mathbb{E} \left[\sup_{t_0 \leq s \leq t \leq t_0 + \Delta} \exp(a|Y_t - Y_s|^2) \right] \leq 1 + C\Delta.$

Proof. We start with (i). We use Itô's formula to write a stochastic equation on the process $(|Y_t - Y_s|^2)_{t \geq s}$:

$$|Y_t - Y_s|^2 = M_{s,t} + 2d(t - s) - 2 \int_s^t (\nabla V(Y_u) + \nabla W * \mu_u(Y_u)) \cdot (Y_u - Y_s) du,$$

where $M_{s,t}$, viewed as a process depending on t , is a martingale with zero expectation. Hence

$$\mathbb{E} |Y_t - Y_s|^2 = 2d(t - s) - 2 \int_s^t \mathbb{E} (\nabla V(Y_u) + \nabla W * \mu_u(Y_u)) \cdot (Y_u - Y_s) du. \quad (4.2)$$

On one hand

$$\mathbb{E} \left| (\nabla V(Y_u) + \nabla W * \mu_u(Y_u)) \cdot (Y_u - Y_s) \right|^2 \leq 4 \left(\mathbb{E} |\nabla V(Y_u)|^2 + \mathbb{E} |\nabla W * \mu_u(Y_u)|^2 \right) \left(\mathbb{E} |Y_u|^2 + \mathbb{E} |Y_s|^2 \right). \quad (4.3)$$

On the other hand, by Proposition 3.1, μ_u has a finite square exponential moment, uniformly bounded for $u \in [0, T]$. More precisely, there exist $\alpha > 0$ and $M < +\infty$ such that $\int e^{\alpha|x|^2} d\mu_u(x) \leq M$ for all $u \leq T$. Since by assumption $|\nabla W(z)| \leq L|z|$ and

$|\nabla V(x)| = O(e^{\alpha|x|^2})$, we deduce

$$\sup_{s \leq u \leq T} \left(\mathbb{E} (\nabla V(Y_u) + \nabla W * \mu_u(Y_u)) \cdot (Y_u - Y_s) \right) < +\infty.$$

In view of (4.2), it follows that there exists a constant $C = C(T)$ such that

$$\mathbb{E} |Y_t - Y_s|^2 \leq (2d + C)(t - s).$$

This concludes the proof of (i).

To establish (ii), we perform a very similar computation. For given s , let $Z_{s,t} := (|Y_t - Y_s|^4)_{t \geq s}$. Another application of Itô's formula yields

$$\begin{aligned} \mathbb{E} Z_{s,t} &= 4(2 + d) \int_s^t \mathbb{E} |Y_u - Y_s|^2 du \\ &\quad - 4 \int_s^t \mathbb{E} |Y_u - Y_s|^2 (Y_u - Y_s) \cdot (\nabla V(Y_u) + \nabla W * \mu_u(Y_u)) du. \end{aligned}$$

On one hand, from (i),

$$\int_s^t \mathbb{E} |Y_u - Y_s|^2 du \leq 2C \int_s^t (u - s) ds = C(t - s)^2.$$

On the other hand

$$\begin{aligned} &\int_s^t \mathbb{E} |Y_u - Y_s|^2 (Y_u - Y_s) \cdot (\nabla V(Y_u) + \nabla W * \mu_u(Y_u)) du \\ &\leq \left(\int_s^t \mathbb{E} Z_{s,u} du \right)^{3/4} \left(\int_s^t \mathbb{E} |\nabla V(Y_u) + \nabla W * \mu_u(Y_u)|^4 du \right)^{1/4} \end{aligned} \quad (4.4)$$

by Hölder's inequality. But again, since the measures μ_t admit a bounded square exponential moment, $\mathbb{E} |\nabla V(Y_u) + \nabla W * \mu_u(Y_u)|^4$ is bounded on $[0, T]$. We conclude that

$$\mathbb{E} Z_{s,t} \leq C \left((t - s)^2 + (t - s)^{1/4} \left(\int_s^t \mathbb{E} Z_{s,u} du \right)^{3/4} \right). \quad (4.5)$$

Then, with C standing again for various constants which are independent of s and t ,

$$\mathbb{E} Z_{s,u} \leq C(\mathbb{E} |Y_u|^4 + \mathbb{E} |Y_s|^4) \leq 2C \sup_{0 \leq t \leq T} \int |x|^4 d\mu_u(x) \leq C;$$

so, from (4.5),

$$\mathbb{E} Z_{s,t} \leq C((t - s)^2 + (t - s)^{1/4}(t - s)^{3/4}) \leq C(t - s),$$

and by (4.5) again we successively obtain

$$\mathbb{E} Z_{s,t} \leq C(t - s)^{7/4},$$

and finally

$$\mathbb{E} Z_{s,t} \leq C(t - s)^2.$$

This concludes the proof of (ii).

We finally turn to the proof of (iii). Without real loss of generality, we set $t_0 = 0$. We shall proceed as in the proof of Proposition 3.1, and prove the existence of some constant C and some continuous positive function a on \mathbb{R}_+ such that

$$\mathbb{E} \left(\sup_{0 \leq s \leq t \leq \Delta \leq T} \exp(a(t)|Y_t - Y_s|^2) \right) \leq 1 + C \Delta. \quad (4.6)$$

Let $a(t)$ be a smooth function, and

$$Z_{s,t} := e^{a(t)|Y_t - Y_s|^2}.$$

By Itô's formula,

$$Z_{s,t} = 1 + M_{s,t} + \int_s^t \left[2a(u) \left(d + 2a|Y_u - Y_s|^2 - (\nabla V + \nabla W * \mu_u)(Y_u) \cdot (Y_u - Y_s) \right) + a'(u)|Y_u - Y_s|^2 \right] Z_{s,u} du$$

where

$$M_{s,t} := \int_s^t a(u) (Y_u - Y_s) Z_u dB_u.$$

For each s , $M_{s,t}$, viewed as a stochastic process in t , is a martingale.

By Young's inequality, for any $b > 0$,

$$-2(\nabla V + \nabla W * \mu_u)(Y_u) \cdot (Y_u - Y_s) \leq b|Y_u - Y_s|^2 + \frac{1}{b} |\nabla V + \nabla W * \mu_u|^2(Y_u).$$

So, by letting

$$A_u := a(u) \left[2d + \frac{1}{b} |\nabla V(Y_u) + \nabla W * \mu_u(Y_u)|^2 \right]$$

and

$$B(u) := a'(u) + 4a^2(u) + ba(u)$$

we obtain

$$Z_{s,t} \leq 1 + M_{s,t} + \int_s^t [A_u + B(u)|Y_u - Y_s|^2] Z_{s,u} du.$$

We choose a in such a way that the function B is identically zero, that is

$$a(u) = e^{-bu} \left(\frac{1}{a(0)} + 4 \frac{1 - e^{-bu}}{b} \right)^{-1},$$

where $a(0)$ is to be fixed later. Then

$$Z_{s,t} \leq 1 + M_{s,t} + \int_s^t A_u Z_{s,u} du$$

from which it is clear that

$$\mathbb{E} \sup_{s \leq t \leq \Delta} Z_{s,t} \leq 1 + \mathbb{E} \sup_{s \leq t \leq \Delta} M_{s,t} + \int_s^\Delta \mathbb{E} A_u Z_{s,u} du. \quad (4.7)$$

By Cauchy-Schwarz and Doob's inequalities,

$$\left(\mathbb{E} \sup_{s \leq t \leq \Delta} M_{s,t} \right)^2 \leq \mathbb{E} \left| \sup_{s \leq t \leq \Delta} M_{s,t} \right|^2 \leq 2 \sup_{s \leq t \leq \Delta} \mathbb{E} |M_{s,t}|^2. \quad (4.8)$$

Also, by Itô's formula and the Cauchy-Schwarz inequality again,

$$\begin{aligned} \mathbb{E} |M_{s,t}|^2 &= \int_s^t a(u)^2 \mathbb{E} |Y_u - Y_s|^2 Z_{s,u}^2 du \\ &\leq \frac{1}{2} \int_s^t a(u)^2 (\mathbb{E} |Y_u - Y_s|^4)^{1/2} (\mathbb{E} Z_{s,u}^4)^{1/2} du. \end{aligned} \quad (4.9)$$

In view of (ii), there exists a constant C such that

$$\mathbb{E} |Y_u - Y_s|^4 \leq C (u - s)^2. \quad (4.10)$$

Furthermore,

$$\mathbb{E} Z_{s,u}^4 = \mathbb{E} \exp 4 a(u) |Y_u - Y_s|^2 \leq (\mathbb{E} \exp 16 a(u) |Y_u|^2)^{1/2} (\mathbb{E} \exp 16 a(u) |Y_s|^2)^{1/2}. \quad (4.11)$$

Recall from Proposition 3.1 that there exist constants M and $\alpha > 0$ such that

$$\sup_{s \leq u \leq \Delta} \int e^{\alpha |y|^2} d\mu_u(y) \leq M.$$

If we choose $a(0) \leq \alpha/16$, the decreasing property of a will ensure that $a(u) \leq \alpha/16$ for all $u \in [0, \Delta]$, and

$$\mathbb{E} \exp 16 a(u) |Y_u|^2 \left(= \int e^{16 a(u) |y|^2} d\mu_u(y) \right) \leq M.$$

Then, from (4.11),

$$\sup_{s \leq u \leq \Delta} \mathbb{E} Z_{s,u}^4 \leq M.$$

Now, from (4.9) and (4.10) we deduce

$$\sup_{s \leq t \leq \Delta} \mathbb{E} |M_{s,t}|^2 \leq C (t - s)^2.$$

Combining this with (4.8), we conclude that

$$\mathbb{E} \sup_{s \leq t \leq \Delta} M_{s,t} \leq C \Delta.$$

In the same way, we can prove that $\mathbb{E}(A_t Z_{s,t})$ is bounded for $t \in [s, \Delta]$ by bounding $\mathbb{E} Z_{s,t}^2$ and $\mathbb{E} A_t^2$. This concludes the proof of (4.6), and therefore of (iii) above. \square

4.2. Time-regularity of the limit empirical measure. We are now ready to prove Proposition 4.1.

On one hand

$$W_1(\widehat{\nu}_s^N, \widehat{\nu}_t^N) \leq \frac{1}{N} \sum_{i=1}^N |Y_t^i - Y_s^i|,$$

so

$$\mathbb{P} \left[\sup_{0 \leq s \leq t \leq \Delta} W_1(\widehat{\nu}_s^N, \widehat{\nu}_t^N) > \varepsilon \right] \leq \mathbb{P} \left[\frac{1}{N} \sum_{i=1}^N V^i > \varepsilon \right] \quad (4.12)$$

where

$$V^i := \sup_{0 \leq s \leq t \leq \Delta} |Y_t^i - Y_s^i|.$$

By Chebyshev's exponential inequality and the independence of the $(Y_t^i - Y_s^i)$,

$$\mathbb{P} \left[\frac{1}{N} \sum_{i=1}^N V^i > \varepsilon \right] \leq \exp \left(-N \sup_{\zeta \geq 0} [\varepsilon \zeta - \log \mathbb{E} \exp(\zeta V^1)] \right).$$

But, for any given ζ and $\omega \geq 0$,

$$\mathbb{E} \exp(\zeta V^1) \leq \mathbb{E} \exp \left(\zeta \left(\frac{\omega^2 + (V^1)^2}{2\omega} \right) \right) \leq \exp \frac{\zeta \omega}{2} \mathbb{E} \exp \frac{\zeta}{2\omega} (V^1)^2.$$

Let $\omega = \frac{\zeta}{2a}$, so that $\frac{\zeta}{2\omega} = a$. Then, from estimate (iii) in Subsection 4.1,

$$\mathbb{E} \exp \frac{\zeta}{2\omega} (V^1)^2 \leq 1 + C \Delta,$$

uniformly in s and Δ . Hence, for any $\zeta > 0$,

$$\mathbb{E} \exp(\zeta V^1) \leq \mathbb{E} \exp \frac{\zeta^2}{4a} (1 + C \Delta).$$

Consequently,

$$\begin{aligned} \mathbb{P} \left[\frac{1}{N} \sum_{i=1}^N V^i > \varepsilon \right] &\leq \exp \left(-N \sup_{\zeta \geq 0} \left[\varepsilon \zeta - \frac{\zeta^2}{4} - \log(1 + C \Delta) \right] \right) \\ &= \exp \left(-N [a \varepsilon^2 - \log(1 + C \Delta)] \right) \\ &\leq \exp \left(-N [a \varepsilon^2 - C \Delta] \right). \end{aligned}$$

The proof of Proposition 4.1 follows by (4.12).

5. COUPLING

We now (as is classical) reduce the proof of convergence for $\widehat{\mu}_t^N$ to a proof of convergence for the empirical measure $\widehat{\nu}_t^N$ constructed on the auxiliary independent system (Y_t^i) . The final goal of this section is the following estimate.

Proposition 5.1. *With the conventions of Subsection 3.1,*

$$W_1(\widehat{\mu}_t^N, \mu_t) \leq \Gamma \int_0^t e^{-\alpha(t-s)} W_1(\widehat{\nu}_s^N, \mu_s) ds + W_1(\widehat{\nu}_t^N, \mu_t),$$

where Γ is defined by (3.1), and $\alpha := \beta + 2 \min(\gamma, 0)$.

Proof. For the sake of simplicity we give a slightly sketchy proof. We couple the stochastic systems (X_t^i) and (Y_t^i) by assuming that (i) $X_0^i = Y_0^i$ and (ii) both systems are driven by the *same* Brownian processes B_t^i . In particular, for each $i \in \{1, \dots, N\}$, the process $X_t^i - Y_t^i$ satisfies the equation

$$d(X_t^i - Y_t^i) = -(\nabla V(X_t^i) - \nabla V(Y_t^i)) dt - \left(\nabla W * \widehat{\mu}_t^N(X_t^i) - \nabla W * \mu_t(Y_t^i) \right) dt. \quad (5.1)$$

From (5.1) we deduce

$$\begin{aligned} \frac{1}{2} \frac{d}{dt} |X_t^i - Y_t^i|^2 &= -(\nabla V(X_t^i) - \nabla V(Y_t^i)) \cdot (X_t^i - Y_t^i) \\ &\quad - \left(\nabla W * \widehat{\mu}_t^N(X_t^i) - \nabla W * \mu_t(Y_t^i) \right) \cdot (X_t^i - Y_t^i). \end{aligned} \quad (5.2)$$

Our convexity assumption on V implies

$$-(\nabla V(X_t^i) - \nabla V(Y_t^i)) \cdot (X_t^i - Y_t^i) \leq -\beta |X_t^i - Y_t^i|^2;$$

so the main issue consists in the treatment of the quantity $\nabla W * \widehat{\mu}_t^N(X_t^i) - \nabla W * \mu_t(Y_t^i)$ appearing in the right-hand side of (5.2). There are (at least) two options here. The first one consists in writing

$$\begin{aligned} \nabla W * \widehat{\mu}_t^N(X_t^i) - \nabla W * \mu_t(Y_t^i) &= \\ &= (\nabla W * \widehat{\mu}_t^N - \nabla W * \mu_t)(X_t^i) + (\nabla W * \mu_t(X_t^i) - \nabla W * \mu_t(Y_t^i)); \end{aligned} \quad (5.3)$$

while the second one consists in forcing the introduction of $\widehat{\nu}_t^N$ as follows:

$$\begin{aligned} \nabla W * \widehat{\mu}_t^N(X_t^i) - \nabla W * \mu_t(Y_t^i) &= \\ &= \frac{1}{N} \sum_{j=1}^N [\nabla W(X_t^i - X_t^j) - \nabla W(Y_t^i - Y_t^j)] - (\nabla W * \widehat{\nu}_t^N - \nabla W * \mu_t)(Y_t^i). \end{aligned} \quad (5.4)$$

Both options are interesting and lead to slightly different computations. Since both lines of computations might be useful in other contexts, we shall sketch them one after the other. The second option leads to better bounds, but at the price of more complications (in particular, we shall need to sum over the index i at an early stage).

First option: We start as in (5.3). In view of our assumption on D^2W , the Lipschitz norm of $\nabla W(X_t^i - \cdot)$ is bounded by Γ . Therefore, by the Kantorovich-Rubinstein dual formulation (1.2),

$$\left| \nabla W * (\widehat{\mu}_t^N - \mu_t)(X_t^i) \right| = \left| \int_{\mathbb{R}^d} \nabla W(X_t^i - y) d(\widehat{\mu}_t^N - \mu_t)(y) \right| \leq \Gamma W_1(\widehat{\mu}_t^N, \mu_t),$$

and then our assumptions on V and W imply

$$\frac{1}{2} \frac{d}{dt} |X_t^i - Y_t^i|^2 \leq -(\gamma + \beta) |X_t^i - Y_t^i|^2 + \Gamma W_1(\widehat{\mu}_t^N, \mu_t) |X_t^i - Y_t^i|.$$

In other words, $|X_t^i - Y_t^i|$ satisfies the differential inequality

$$\frac{d}{dt} |X_t^i - Y_t^i| + (\beta + \gamma) |X_t^i - Y_t^i| \leq \Gamma W_1(\widehat{\mu}_t^N, \mu_t)$$

(X_t^i and Y_t^i separately are not Lipschitz functions of t , but their difference is). Hence, by Gronwall's lemma,

$$|X_t^i - Y_t^i| \leq \Gamma \int_0^t e^{-(\beta+\gamma)(t-s)} W_1(\widehat{\mu}_s^N, \mu_s) ds.$$

Now we sum over i ; by convexity of the distance W_1 and triangular inequality, we obtain

$$\begin{aligned} W_1(\widehat{\mu}_t^N, \widehat{\nu}_t^N) &\leq \frac{1}{N} \sum_{i=1}^N |X_t^i - Y_t^i| \leq \Gamma \int_0^t e^{-(\beta+\gamma)(t-s)} W_1(\widehat{\mu}_s^N, \mu_s) ds \\ &\leq \Gamma \int_0^t e^{-(\beta+\gamma)(t-s)} [W_1(\widehat{\mu}_s^N, \widehat{\nu}_s^N) + W_1(\widehat{\nu}_s^N, \mu_s)] ds. \end{aligned}$$

By using Gronwall's lemma again, we deduce

$$W_1(\widehat{\mu}_t^N, \widehat{\nu}_t^N) \leq \Gamma \int_0^t e^{-(\beta+\gamma-\Gamma)(t-s)} W_1(\widehat{\nu}_s^N, \mu_s) ds.$$

By applying the triangular inequality for W_1 , we conclude to the validity of Proposition (5.1), only with α replaced by the (a priori smaller) quantity $\beta + \gamma - \Gamma$.

Second option: Now we start with (5.4). This time we sum over i right from the beginning:

$$\frac{1}{2} \frac{d}{dt} \sum_{i=1}^N |X_t^i - Y_t^i|^2 = - \sum_{i=1}^N (\nabla V(X_t^i) - \nabla V(Y_t^i)) \cdot (X_t^i - Y_t^i) - \frac{1}{N} \sum_{i,j=1}^N (A_t^{ij} + B_t^{ij})$$

where

$$A_t^{ij} = (\nabla W(X_t^i - X_t^j) - \nabla W(Y_t^i - Y_t^j)) \cdot (X_t^i - Y_t^i)$$

and

$$B_t^{ij} = (W(Y_t^i - Y_t^j) - \nabla W * \mu_t(Y_t^i)) \cdot (X_t^i - Y_t^i).$$

Since ∇W is an odd function and $D^2W(x) \geq \gamma I$ for all $x \in \mathbb{R}^d$, we have

$$\begin{aligned} A_t^{ij} + A_t^{ji} &= (\nabla W(X_t^i - X_t^j) - \nabla W(Y_t^i - Y_t^j)) \cdot ((X_t^i - X_t^j) - (Y_t^i - Y_t^j)) \\ &\geq \gamma |(X_t^i - X_t^j) - (Y_t^i - Y_t^j)|^2, \end{aligned}$$

whence

$$-\sum_{i,j=1}^N A_t^{ij} \leq -\frac{\gamma}{2} \sum_{i,j=1}^N |(X_t^i - X_t^j) - (Y_t^i - Y_t^j)|^2 \leq -2N\gamma^- \sum_{i=1}^N |X_t^i - Y_t^i|^2$$

where $\gamma^- = \min(\gamma, 0)$.

Then

$$-\sum_{j=1}^N B_t^{ij} = -(X_t^i - Y_t^i) \cdot (\nabla W * \widehat{\nu}_t^N(Y_t^i) - \nabla W * \mu_t(Y_t^i)).$$

Our assumption on D^2W implies that the Lipschitz norm of $\nabla W(Y_t^i - \cdot)$ is bounded by Γ ; so, by the Kantorovich-Rubinstein dual formulation (1.2),

$$\left| \nabla W * (\widehat{\nu}_t^N - \mu_t)(Y_t^i) \right| = \left| \int_{\mathbb{R}^d} \nabla W(Y_t^i - y) d(\widehat{\nu}_t^N - \mu_t)(y) \right| \leq \Gamma W_1(\widehat{\nu}_t^N, \mu_t).$$

Collecting all terms we finally obtain

$$\frac{1}{2} \frac{d}{dt} \sum_{i=1}^N |X_t^i - Y_t^i|^2 \leq -(\beta + 2\gamma^-) \sum_{i=1}^N |X_t^i - Y_t^i|^2 + \Gamma \sum_{i=1}^N |X_t^i - Y_t^i| W_1(\widehat{\nu}_t^N, \mu_t).$$

Then, since $\sum_{i=1}^N |X_t^i - Y_t^i| \leq \left(N \sum_{i=1}^N |X_t^i - Y_t^i|^2 \right)^{1/2}$, the function $y(t) := \left(\frac{1}{N} \sum_{i=1}^N |X_t^i - Y_t^i|^2 \right)^{1/2}$ satisfies the differential inequality

$$y'(t) + (\beta + 2\gamma^-)y(t) \leq \Gamma W_1(\widehat{\nu}_t^N, \mu_t),$$

so that

$$\left(\frac{1}{N} \sum_{i=1}^N |X_t^i - Y_t^i|^2 \right)^{1/2} \leq \Gamma \int_0^t e^{-(\beta+2\gamma^-)(t-s)} W_1(\widehat{\nu}_s^N, \mu_s) ds.$$

The conclusion follows by triangular inequality again since

$$W_1(\widehat{\mu}_t^N, \widehat{\nu}_t^N) \leq W_2(\widehat{\mu}_t^N, \widehat{\nu}_t^N) \leq \left(\frac{1}{N} \sum_{i=1}^N |X_t^i - Y_t^i|^2 \right)^{1/2}.$$

□

Remark 5.2. Not only does the “second option” in the proof lead to better bounds, it also provides an estimate of the distance between $\widehat{\mu}$ and $\widehat{\nu}$ in the W_2 distance, which is stronger than the W_1 distance. However, we do not take any advantage of this refinement.

6. CONCLUSION

In this section, we paste together all the estimates established in the previous sections, so as to prove Theorems 1.7 to 1.11.

6.1. Concentration estimates. We start with the proof of Theorem 1.7. By C we shall denote various constants depending on T , on our assumptions on V and W , and also on $\int e^{\alpha|x|^2} d\mu_0(x)$, for some $\alpha > 0$.

From Proposition 5.1,

$$\sup_{0 \leq t \leq T} W_1(\widehat{\mu}_t^N, \mu_t) \leq (\Gamma e^{|\alpha|T} + 1) \sup_{0 \leq t \leq T} W_1(\widehat{\nu}_t^N, \mu_t) ds.$$

In particular, there is a constant C such that

$$\mathbb{P} \left[\sup_{0 \leq t \leq T} W_1(\widehat{\mu}_t^N, \mu_t) > \varepsilon \right] \leq \mathbb{P} \left[\sup_{0 \leq t \leq T} W_1(\widehat{\nu}_t^N, \mu_t) > \widetilde{\varepsilon} \right], \quad \widetilde{\varepsilon} = \frac{\varepsilon}{C}. \quad (6.1)$$

From Corollary 3.2 and Theorem 1.1 we know that

$$\sup_{0 \leq t \leq T} \mathbb{P} [W_1(\widehat{\nu}_t^N, \mu_t) > \widetilde{\varepsilon}] \leq e^{-KN\widetilde{\varepsilon}^2}$$

for all $t \in [0, T]$, $N \geq N_0 \max(\widetilde{\varepsilon}^{-(d'+2)}, 1)$ ($d' > d$). The issue now is to “exchange” sup and \mathbb{P} in this estimate. As we shall see, this is authorized by the continuity estimates on $\widehat{\nu}_t^N$ and μ_t .

Let $\Delta > 0$ (to be fixed later on), and let M be the integer part of $T/\Delta + 1$. We decompose the interval $[0, T]$ as

$$[0, T] = [0, \Delta] \cup [\Delta, 2\Delta] \cup \dots \cup [(M-1)\Delta, T] \subset \bigcup_{h=0}^{M-1} [h\Delta, (h+1)\Delta].$$

Proposition 3.3 guarantees that, if $\Delta \leq a\widetilde{\varepsilon}^2$ for some a small enough, then

$$h\Delta \leq t \leq (h+1)\Delta \implies W_1(\mu_t, \mu_{h\Delta}) \leq \frac{\widetilde{\varepsilon}}{2}. \quad (6.2)$$

Then, by triangular inequality and (6.2),

$$\begin{aligned} & \mathbb{P} \left[\sup_{0 \leq t \leq T} W_1(\widehat{\nu}_t^N, \mu_t) > \widetilde{\varepsilon} \right] \\ & \leq \mathbb{P} \left[\sup_{h=0, \dots, M-1} \sup_{h\Delta \leq t \leq (h+1)\Delta} W_1(\widehat{\nu}_t^N, \mu_t) > \widetilde{\varepsilon} \right] \\ & \leq \mathbb{P} \left[\sup_{h=0, \dots, M-1} \sup_{h\Delta \leq t \leq (h+1)\Delta} W_1(\widehat{\nu}_t^N, \nu_{h\Delta}^N) + \sup_{h=0, \dots, M-1} W_1(\widehat{\nu}_{h\Delta}^N, \mu_{h\Delta}) \right. \\ & \quad \left. + \sup_{h=0, \dots, M-1} \sup_{h\Delta \leq t \leq (h+1)\Delta} W_1(\mu_{h\Delta}, \mu_t) > \widetilde{\varepsilon} \right] \end{aligned}$$

$$\leq \mathbb{P} \left[\sup_{h=0, \dots, M-1} \sup_{h\Delta \leq t \leq (h+1)\Delta} W_1(\widehat{\nu}_t^N, \widehat{\nu}_{h\Delta}^N) + \sup_{h=0, \dots, M-1} W_1(\widehat{\nu}_{h\Delta}^N, \mu_{h\Delta}) > \frac{\widetilde{\varepsilon}}{2} \right],$$

which can be bounded by

$$\mathbb{P} \left[\sup_{h=0, \dots, M-1} \sup_{h\Delta \leq t \leq (h+1)\Delta} W_1(\widehat{\nu}_t^N, \widehat{\nu}_{h\Delta}^N) > \frac{\widetilde{\varepsilon}}{4} \right] + \mathbb{P} \left[\sup_{h=0, \dots, M-1} W_1(\widehat{\nu}_{h\Delta}^N, \mu_{h\Delta}) > \frac{\widetilde{\varepsilon}}{4} \right].$$

By Corollary 3.2 and Theorem 1.1, there exist some constants C and N_0 such that

$$\mathbb{P} \left[W_1(\widehat{\nu}_{h\Delta}^N, \mu_{h\Delta}) \geq \frac{\widetilde{\varepsilon}}{4} \right] \leq \exp(-C N \widetilde{\varepsilon}^2)$$

for all $h = 0, \dots, M-1$, and $N \geq N_0 \max(\widetilde{\varepsilon}^{-(d'+2)}, 1)$. Hence

$$\mathbb{P} \left[\sup_{h=0, \dots, M-1} W_1(\widehat{\nu}_{h\Delta}^N, \mu_{h\Delta}) > \frac{\widetilde{\varepsilon}}{4} \right] \leq \sum_{h=0}^{M-1} \mathbb{P} \left[W_1(\widehat{\nu}_{h\Delta}^N, \mu_{h\Delta}) > \frac{\widetilde{\varepsilon}}{4} \right] \leq M \exp(-C N \widetilde{\varepsilon}^2). \quad (6.3)$$

On the other hand, from Proposition 4.1 we deduce

$$\mathbb{P} \left[\sup_{h\Delta \leq t \leq (h+1)\Delta} W_1(\widehat{\nu}_t^N, \widehat{\nu}_{h\Delta}^N) > \frac{\widetilde{\varepsilon}}{4} \right] \leq \exp\left(-N\left(\frac{a}{4}\widetilde{\varepsilon}^2 - C\Delta\right)\right)$$

for all $h = 0, \dots, M-1$ and $\widetilde{\varepsilon} > 0$, so

$$\mathbb{P} \left[\sup_{h=0, \dots, M-1} \sup_{h\Delta \leq t \leq (h+1)\Delta} W_1(\widehat{\nu}_t^N, \widehat{\nu}_{h\Delta}^N) > \frac{\widetilde{\varepsilon}}{4} \right] \leq M \exp\left(-N\left(\frac{a}{4}\widetilde{\varepsilon}^2 - C\Delta\right)\right). \quad (6.4)$$

We can assume that $\Delta \leq \frac{a}{8C}\widetilde{\varepsilon}^2$, and $M \leq CT/\widetilde{\varepsilon}^2 + 1$; then we can bound the right-hand side of (6.4) by

$$M \exp\left(-\frac{a}{8}N\widetilde{\varepsilon}^2\right) \leq C \left(1 + \frac{T}{\widetilde{\varepsilon}^2}\right) \exp\left(-\frac{a}{8}N\widetilde{\varepsilon}^2\right) \quad (6.5)$$

From (6.3) and (6.5) we deduce that, for Δ small enough (depending on $\varepsilon!$),

$$\mathbb{P} \left[\sup_{0 \leq t \leq T} W_1(\widehat{\nu}_t^N, \mu_t) > \widetilde{\varepsilon} \right] \leq 2C \left(1 + \frac{T}{\widetilde{\varepsilon}^2}\right) \exp(-KN\widetilde{\varepsilon}^2) \quad (6.6)$$

for $N \geq N_0 \max(\widetilde{\varepsilon}^{-(d'+2)}, 1)$. So we deduce from (6.6) that

$$\mathbb{P} \left[\sup_{0 \leq t \leq T} W_1(\widehat{\nu}_t^N, \mu_t) > \widetilde{\varepsilon} \right] \leq \exp\left(\log\left(C\left(\frac{T}{\widetilde{\varepsilon}^2} + 1\right)\right) - KN\widetilde{\varepsilon}^2\right),$$

where again C, K stand for various positive constants, and $N \geq \max(N_0 \varepsilon^{-(d'+2)}, 1)$. This concludes the proof of Theorem 1.7.

6.2. Uniform in time estimates. Now, we shall focus on the case when $\beta > 0, \beta + 2\gamma > 0$ is positive, and derive Theorem 1.9 by a slightly refined estimate.

Let us start again from the bound

$$W_1(\widehat{\mu}_t^N, \mu_t) \leq \Gamma \int_0^t e^{-\alpha(t-s)} W_1(\widehat{\nu}_s^N, \mu_s) ds + W_1(\widehat{\nu}_t^N, \mu_t)$$

where $\alpha := \beta + 2 \min(\gamma, 0)$ is positive. Let $\Delta > 0$ (to be fixed later on), and k be the integer part of t/Δ . If $W_1(\widehat{\mu}_t^N, \mu_t)$ is larger than ε , then

$$\left\{ \begin{array}{l} \text{either } W_1(\widehat{\nu}_t^N, \mu_t) \geq \frac{\varepsilon}{2} \\ \text{or } \exists j \in \{0, \dots, k\}; \quad \int_{j\Delta}^{(j+1)\Delta} e^{-\alpha(t-s)} W_1(\widehat{\nu}_s^N, \mu_s) ds \geq \frac{\varepsilon}{2^{k+2-j}\Gamma}. \end{array} \right.$$

Indeed, $(\varepsilon/2) + \sum_{j \leq k} (\varepsilon/2^{k+2-j}) \leq \varepsilon$. As a consequence,

$$\left\{ \begin{array}{l} \text{either } W_1(\widehat{\nu}_t^N, \mu_t) > \frac{\varepsilon}{2} \\ \text{or } \exists j \in \{0, \dots, k\}; \quad \sup_{j\Delta \leq s \leq (j+1)\Delta} W_1(\widehat{\nu}_s^N, \mu_s) > \frac{\varepsilon \alpha e^{\alpha[t-(j+1)\Delta]}}{2^{k+2-j}\Gamma}. \end{array} \right.$$

Since, for $t \in [j\Delta, (j+1)\Delta]$,

$$\frac{e^{\alpha[t-(j+1)\Delta]}}{2^{k+2-j}} \geq \frac{e^{\alpha(k-j-1)\Delta}}{2^{k-j+2}} = \left(\frac{1}{4e^{\alpha\Delta}} \right) \left(\frac{e^{\alpha\Delta}}{2} \right)^{k-j},$$

we conclude to the existence of a constant C such that

$$\begin{aligned} \mathbb{P} [W_1(\widehat{\mu}_t^N, \mu_t) > \varepsilon] &\leq \mathbb{P} \left[W_1(\widehat{\nu}_t^N, \mu_t) > \frac{\varepsilon}{2} \right] \\ &\quad + \sum_{j=0}^k \mathbb{P} \left[\sup_{j\Delta \leq s \leq (j+1)\Delta} W_1(\widehat{\nu}_s^N, \mu_s) > C\varepsilon \left(\frac{e^{\alpha\Delta}}{2} \right)^{k-j} \right]. \end{aligned} \quad (6.7)$$

We already know that the first term in the right-hand side in (6.7) is bounded by $e^{-\lambda N \varepsilon^2}$ for some constant $\lambda > 0$, and so we focus on the other terms.

In the proof of Theorem 1.7, we have established that there are constant C and λ , depending on Δ and on bounds on square exponential moments for μ_0 , such that

$$\mathbb{P} \left[\sup_{0 \leq s \leq \Delta} W_1(\widehat{\nu}_s^N, \mu_s) > \delta \right] \leq C \left(1 + \frac{\Delta}{\delta^2} \right) e^{-\lambda N \delta^2}. \quad (6.8)$$

Proposition 3.1 guarantees that these square exponential bounds also hold true for μ_t , uniformly in t . Thus we can apply (6.8) with $\mu_{j\Delta}$ taken as initial datum, and get

$$\mathbb{P} \left[\sup_{j\Delta \leq s \leq (j+1)\Delta} W_1(\widehat{\nu}_s^N, \mu_s) > \delta \right] \leq C e^{-\lambda N \delta^2}, \quad (6.9)$$

as soon as $N \geq N_0 \max(\delta^{-(d'+2)}, 1)$.

We now use (6.9) to bound the sum appearing in the right-hand side of (6.7). Choose Δ large enough that

$$\theta := \frac{e^{\alpha\Delta}}{2} > 1.$$

Applying (6.9) with δ replaced by $C\theta^{j-k}\varepsilon$, we can bound the sum in the right-hand side of (6.7) by

$$C \sum_{j=0}^k \exp(-K\theta^{2(k-j)}N\varepsilon^2)$$

for $N \geq N_0 \max(\varepsilon^{-(d'+2)}, 1)$, where C , K and N_0 are again positive constants. Since again θ is larger than 1, there is a constant $a > 0$ such that $\theta^{2(k-j)} \geq a(k-j)$, so the sum above is bounded by

$$C \left(e^{-KN\varepsilon^2} + \sum_{\ell=1}^{\infty} e^{-K\ell N\varepsilon^2} \right) \leq C \left(e^{-KN\varepsilon^2} + \frac{e^{-KN\varepsilon^2}}{1 - e^{-KN\varepsilon^2}} \right).$$

If N_0 is large enough, our assumption $N \geq N_0 \max(\varepsilon^{-(d'+2)}, 1)$ implies that $e^{-KN\varepsilon^2}$ is always less than $1/2$, so that the above sum can be bounded by just $Ce^{-KN\varepsilon^2}$. This concludes the proof of the first point of Theorem 1.9.

The second point is proved by writing

$$\begin{aligned} W_1(\widehat{\mu}_t^N, \mu_\infty) &\leq W_1(\widehat{\mu}_t^N, \mu_t) + W_1(\mu_t, \mu_\infty) \\ &\leq W_1(\widehat{\mu}_t^N, \mu_t) + Ce^{-\lambda t} \end{aligned}$$

successively by the triangular inequality for Wasserstein distance and use of Proposition 3.8. Then the result follows from the uniform estimate obtained above.

6.3. Data reconstruction. We finally consider Theorem 1.11. Proposition 3.6 ensures that, as $t \rightarrow \infty$, f_t is uniformly bounded in C^k , where k is arbitrarily large. Since f_t converges to f_∞ as $t \rightarrow \infty$, we deduce that f_∞ is Lipschitz. Then Theorem 1.9 and Proposition 2.1 together imply Theorem 1.11.

APPENDIX A. METRIC ENTROPY OF A PROBABILITY SPACE

We now prove the covering result used in Section 2.1, as a particular case of a more general estimate. Let E be a Polish space, we look for an upper bound on the number $\mathcal{N}_p(E, \delta) := m(\mathcal{P}(E), \delta)$ of balls of radius δ in Wasserstein distance W_p needed to cover the space $\mathcal{P}(E)$ of probability measures on E . We use the same strategy as in [12, Exercise 6.2.19], where the Lévy distance is used instead of the Wasserstein distance.

Theorem A.1. *Let (E, d) be a Polish space with finite diameter D . For any $r > 0$, define $N(E, r)$ as the minimal number of balls needed to cover E by balls of radius r . Then, for*

all $p \geq 1$ and $\delta \in (0, D)$, the space $\mathcal{P}(E)$ can be covered by $\mathcal{N}_p(E, \delta)$ balls of radius δ in W_p distance, with

$$\mathcal{N}_p(E, \delta) \leq \left(\frac{8eD}{\delta} \right)^{pN(E, \frac{\delta}{2})}. \quad (\text{A.1})$$

Remark A.2. The W_p distance between any two probability measures on E is at most D , so, for all $\delta \geq D$, we have the trivial estimate $\mathcal{N}_p(E, \delta) = 1$.

Proof. Let $r > 0$, and let $\{x_j\}_{1 \leq j \leq N(E, r)}$ be such that E is covered by the balls $B(x_j, r)$ with centers $x_j \in E$ and radius r . For simplicity we shall write $N = N(E, r)$.

In a **first step** we prove that for any $\mu \in \mathcal{P}(E)$ there exist nonnegative real numbers $(\beta_j)_{1 \leq j \leq N}$, with $\sum_{j=1}^N \beta_j = 1$, such that

$$W_p(\mu, \tilde{\mu}) \leq r, \quad \tilde{\mu} := \sum_{j=1}^N \beta_j \delta_{x_j}.$$

For this we first replace the balls $B(x_j, r)$'s by the sets \tilde{B}_j 's defined by

$$\forall j, \quad \tilde{B}_j = B(x_j, r) \setminus \bigcup_{k \leq j-1} B(x_k, r),$$

so that E is partitioned into the \tilde{B}_j 's. Next define

$$\beta_j = \mu[\tilde{B}_j].$$

It is easy to check that the required properties are fulfilled. Indeed, we may transport μ onto $\tilde{\mu} = \sum_{j=1}^N \beta_j \delta_{x_j}$ by sending all x 's in \tilde{B}_j onto x_j , for each $j = 1, \dots, N$: the cost of this transport is bounded by $\sum_{j=1}^N r^p \mu(\tilde{B}_j) = r^p$.

In the **second step** we introduce an integer K (whose value will be made more precise later on), and consider the set

$$\mathcal{C}_K := \left\{ \sum_{j=1}^N \alpha_j \delta_{x_j}; \quad (\alpha_j)_{1 \leq j \leq N} \in \mathcal{A}_K \right\} \subset \mathcal{P}(E),$$

where \mathcal{A}_K is the set of all N -tuples $(\alpha_j)_{1 \leq j \leq N}$, such that each α_j is of the form k_j/K , $k_j \in \mathbb{N}$, and $\sum_{j=1}^N \alpha_j = 1$.

Given a probability measure $\tilde{\mu} = \sum_{i=1}^N \beta_i \delta_{x_i}$ (where $(\beta_i)_i$ does not necessarily belong to \mathcal{A}_K), there exists μ' in \mathcal{C}_K such that

$$W_p(\mu', \tilde{\mu}) \leq D \left(\frac{N}{K} \right)^{1/p}. \quad (\text{A.2})$$

To prove (A.2), we define n_j as the integer part $[K\beta_j]$ of $K\beta_j$ and J as the first integer such that

$$\sum_{j=1}^J (n_j + 1) + \sum_{j=J+1}^N n_j = K.$$

Since $\sum_{j=1}^N \beta_j = 1$, it is clear that $J \leq N$. Then we define a measure $\mu' \in \mathcal{C}_K$ by $\mu' =$

$\sum_{j=1}^N \alpha_j \delta_{x_j}$, where

$$\alpha_j = \begin{cases} \frac{n_j+1}{K} & \text{for } j = 1, \dots, J \\ \frac{n_j}{K} & \text{for } j = J+1, \dots, N. \end{cases}$$

Let us bound the distance between μ and μ' . For that we gradually define a transport plan between $\tilde{\mu}$ and μ' in the following way: first of all, at each point x_i , the mass n_i/K stays in place. Then, the remaining masses $\beta_i - n_i/K$ are redistributed as follows: all the remaining mass at x_1, \dots, x_ℓ is brought to x_1 , together with possibly a bit of mass at $x_{\ell+1}$, until a total mass $1/K$ has been added at location x_1 (for ℓ large enough). If $J \geq 2$, then we again bring mass from $x_{\ell+1}, \dots$, until another mass $1/K$ has been added at x_2 . We carry on until all the mass at x_J has been used, thus building a transport plan $(\pi_{ij})_{1 \leq i, j \leq N}$ which sends $\tilde{\mu}$ onto μ' , in such a way that $\pi_{ii} \geq \frac{n_i}{K}$ for all i . Hence,

$$\sum_{j \neq i} \pi_{ij} \leq \beta_i - \pi_{ii} = \beta_i - \frac{n_i}{K} \leq \frac{1}{K},$$

and this plan yields an upper bound on the Wasserstein distance:

$$W_p^p(\tilde{\mu}, \mu') \leq \sum_{i,j=1}^N d(x_i, x_j)^p \pi_{ij} = \sum_{i=1}^N \sum_{j \neq i} d(x_i, x_j)^p \pi_{ij} \leq N \frac{D^p}{K}.$$

To summarize the first two steps: for any μ in $\mathcal{P}(E)$ there exists $\mu' \in \mathcal{C}_K$ such that

$$W_p(\mu, \mu') \leq r + D \left(\frac{N}{K} \right)^{1/p}.$$

In other words, the family $\left(B(\mu', r + D(N/K)^{1/p}) \right)_{\mu' \in \mathcal{C}_K}$ covers $\mathcal{P}(E)$.

In the **third step** we choose some suitable K and r for a given δ .

We first choose K in such a way that r and $D(N/K)^{1/p}$ have the same order of magnitude, for instance

$$K = \left\lceil N \left(\frac{D}{r} \right)^p \right\rceil + 1.$$

Then

$$r + D(N/K)^{1/p} \leq 2r,$$

and the balls $B(\mu', r + D(N/K)^{1/p})$ have radius at most δ if

$$r = \frac{\delta}{2}.$$

Now K and r are fixed, $N = N(E, \delta/2)$, and we just have to estimate the cardinality $\#\mathcal{C}_K$ of \mathcal{C}_K . For this we first note that

$$\#\mathcal{C}_K = \frac{(K + N - 1)!}{(K - 1)!N!} = \frac{(K + N - 1) \dots K}{N!}$$

Without loss of generality, we have assumed $\delta < D$, so $K > N$. Then $K < \dots < K + N - 1 < 2K$, and hence

$$\#\mathcal{C}_K \leq \frac{(2K)^N}{N!} \leq \left(\frac{2Ke}{N} \right)^N.$$

Since $N \geq 1$ and $2D \geq \delta$, we can write

$$K \leq N \left(\frac{2D}{\delta} \right)^p + 1 \leq 2N \left(\frac{2D}{\delta} \right)^p,$$

and we deduce

$$\#\mathcal{C}_K \leq \left(C \frac{D}{\delta} \right)^{pN(E, \frac{\delta}{2})}$$

with $C = 2(4e)^{1/p} \leq 8e$.

Consequently, we have covered $\mathcal{P}(E)$ by the $\left(C \frac{D}{\delta} \right)^{pN(E, \frac{\delta}{2})}$ balls $(B(\mu', \delta))_{\mu' \in \mathcal{C}_K}$ with radius δ . This concludes the argument. □

In the particular case when E is the Euclidean ball B_R of radius R in \mathbb{R}^d , we have

$$N(B_R, r) \leq k \left(\frac{R}{r} \right)^d \tag{A.3}$$

for some constant k . To see this, one may for instance consider the balls with center in the lattice $\frac{r}{\sqrt{d}}\mathbb{Z}^d$ in \mathbb{R}^d . Then Theorem A.1 yields the bound

$$\mathcal{N}_p(B_R, \delta) \leq \left(C \frac{R}{\delta}\right)^{pk\left(\frac{R}{\delta}\right)^d},$$

which is used in the present paper.

APPENDIX B. REGULARITY ESTIMATES ON THE LIMIT PDE

In this appendix we study solutions to the limit equation

$$\partial_t \mu = \Delta \mu + \nabla \cdot (\mu \nabla (V + W * \mu)), \quad t \geq 0, \quad x \in \mathbb{R}^d \quad (\text{B.1})$$

and establish the regularity results stated in Proposition 3.6. Following the method in [13], we shall measure the regularity in terms of L^2 -Sobolev spaces

$$H^s(\mathbb{R}^d) = \left\{ u \in L^2(\mathbb{R}^d); \partial^\alpha u \in L^2(\mathbb{R}^d), \alpha \in \mathbb{N}^d, |\alpha| \leq s \right\} \quad (s \in \mathbb{N}).$$

Our main result is as follows.

Theorem B.1. *Let V and W such that all their partial derivatives $\partial^\alpha V$ and $\partial^\alpha W$ are continuous and grow at most polynomially at infinity, for any multi-index $\alpha \in \mathbb{N}^d$ with $|\alpha| \leq s + 1$. Let $a, E > 0$ and let μ_0 be a probability density such that*

$$\int_{\mathbb{R}^d} e^{a|x|^2} d\mu_0(x) \leq E.$$

Then, there exists a continuous function $f : (0, +\infty) \rightarrow (0, +\infty)$, only depending on d, s, V, W, a and E , such that any classical solution $\mu = \mu(t, x)$ to (B.1), starting from μ_0 , satisfies

$$\|\mu(t, \cdot)\|_{H^s(\mathbb{R}^d)} \leq f(t).$$

Proof. For the sake of simplicity we only give a formal proof, which can be turned rigorous by means of regularization arguments.

Let then $\mu = (\mu(t, \cdot))_{t \geq 0}$ be a solution of

$$\partial_t \mu = \Delta \mu + \nabla \cdot (\mu \nabla (V + W * \mu)), \quad t \geq 0, \quad x \in \mathbb{R}^d;$$

we rewrite the equation as

$$\partial_t \mu = \sum_{i=1}^d \partial_{ii} \mu + \partial_i [\mu \partial_i \phi],$$

where $\partial_i = \partial^{e_i}$ if e_i is the i -th vector of the canonical base of \mathbb{R}^d , and

$$\phi(t, x) = V(x) + W * \mu(t, x).$$

Let $\alpha \in \mathbb{N}^d$ be given. By integration by parts and Cauchy-Schwarz inequality,

$$\begin{aligned}
\frac{1}{2} \frac{d}{dt} \int_{\mathbb{R}^d} |\partial^\alpha \mu|^2 &= \int_{\mathbb{R}^d} \partial^\alpha \mu \partial_t (\partial^\alpha \mu) = \int_{\mathbb{R}^d} \partial^\alpha \mu \partial^\alpha (\partial_t \mu) \\
&= \sum_{i=1}^d \int_{\mathbb{R}^d} \partial^\alpha \mu \partial^\alpha (\partial_{ii} \mu + \partial_i [\mu \partial_i \phi]) \\
&= - \sum_{i=1}^d \int_{\mathbb{R}^d} |\partial^{\alpha+e_i} \mu|^2 + \int \partial^{\alpha+e_i} \mu \partial^\alpha [\mu \partial_i \phi] \\
&\leq - \sum_{i=1}^d \int_{\mathbb{R}^d} |\partial^{\alpha+e_i} \mu|^2 + \sum_i^d \left[\int_{\mathbb{R}^d} |\partial^{\alpha+e_i} \mu|^2 \right]^{1/2} \left[\sum_{\beta \leq \alpha} C_{\alpha, \beta} \int_{\mathbb{R}^d} |\partial^{\alpha-\beta+e_i} \phi \partial^\beta \mu|^2 \right]^{1/2} \\
&\leq -\frac{1}{2} \sum_{i=1}^d \int_{\mathbb{R}^d} |\partial^{\alpha+e_i} \mu|^2 + \sum_{\beta \leq \alpha} C_{\alpha, \beta} \int_{\mathbb{R}^d} |\partial^{\alpha-\beta+e_i} \phi \partial^\beta \mu|^2.
\end{aligned}$$

By summing over $\alpha \in \mathbb{N}^d$ with $|\alpha| \leq s$, we find

$$\frac{d}{dt} \sum_{|\alpha| \leq s} \int_{\mathbb{R}^d} |\partial^\alpha \mu|^2 \leq - \sum_{|\alpha| \leq s} \sum_{i=1}^d \int_{\mathbb{R}^d} |\partial^{\alpha+e_i} \mu|^2 + \sum_{|\alpha| \leq s} \sum_{\beta \leq \alpha} C_{\alpha, \beta} \int_{\mathbb{R}^d} |\partial^{\alpha-\beta+e_i} \phi \partial^\beta \mu|^2.$$

Given $T > 0$, by Proposition 3.1 there exist constants \widehat{a} and \widehat{E} , depending only on d , a , E and T , such that

$$\int e^{\widehat{a}|x|^2} d\mu(t, x) \leq \widehat{E} \tag{B.2}$$

for all $t \in [0, T]$. In particular, it follows from our assumptions on the derivatives of V and W that all $|\partial^{\alpha-\beta+e_i} \phi|^2$ terms are bounded by some polynomial in $|x|$, uniformly in $t \in [0, T]$.

Let $\langle x \rangle := \sqrt{1 + |x|^2}$. For $k, s \geq 0$, we introduce the weighted norms

$$\|u\|_{H_k^s} := \left(\sum_{|\alpha| \leq s} \int_{\mathbb{R}^d} \langle x \rangle^k |\partial^\alpha u(x)|^2 dx \right)^{1/2}$$

and

$$\|u\|_{L_k^1} := \int_{\mathbb{R}^d} \langle x \rangle^k |u(x)| dx.$$

Then for any $s \in \mathbb{N}$ and $T \geq 0$ there exist k and $C \geq 0$ such that

$$0 \leq t \leq T \implies \frac{d}{dt} \|\mu\|_{H^s}^2 \leq -\|\mu\|_{H^{s+1}}^2 + C \|\mu\|_{H_k^s}^2. \tag{B.3}$$

We shall prove later on the following interpolation lemma:

Lemma B.2. *Given $d \geq 1$, $s \in \mathbb{N}$ and $k \geq 0$, there exist nonnegative constants $C(d, s, k)$ and $h(d, s, k)$, and $\theta(d, s) \in (0, 1)$ such that for all $u \in L^\infty_1(\mathbb{R}^d) \cap H^{s+1}(\mathbb{R}^d)$,*

$$\|u\|_{H^s_k} \leq C(d, s, k) \|u\|_{L^1_{h(d,s,k)}}^{1-\theta(d,s)} \|u\|_{H^{s+1}}^{\theta(d,s)}.$$

Then, again from (B.2), all $\|\mu\|_{L^1_{h(d,s,k)}}(t)$ norms are bounded on $[0, T]$, so from (B.3) and Lemma B.2 there exists some constants C such that

$$\frac{d}{dt} \|\mu\|_{H^s}^2 \leq -\|\mu\|_{H^{s+1}}^2 + C \|\mu\|_{H^{s+1}}^{2\theta} \leq -\frac{1}{2} \|\mu\|_{H^{s+1}}^2 + C \leq -C \|\mu\|_{H^s}^{2/\theta} + C.$$

In other words $A(t) = \|\mu\|_{H^s}^2(t)$ satisfies on $[0, T]$ the differential inequality

$$A'(t) + c A(t)^p \leq C \tag{B.4}$$

for some constants $c, C \geq 0$ and $p = 1/\theta > 1$ depending only on d, a, E, s and T .

Let us distinguish two cases. If $A(0) \leq 1$, then we only use the inequality $A'(t) \leq C$ to make sure that

$$A(t) \leq A(0) + Ct \leq 1 + CT$$

for any $t \in [0, T]$.

If on the other hand $A(0) \geq 1$, we deduce from (B.4) that

$$A'(t) + c A(t)^p \leq C A(t),$$

as long as $A(t) \geq 1$, so that $D(t) := A(t)^{1-p}$ satisfies the inequality

$$D'(t) + (p-1) C D(t) \geq (p-1) c$$

which integrates to

$$D(t) \geq D(0) e^{(1-p)Ct} + \frac{c}{C} (1 - e^{(p-1)Ct}) \geq \frac{c}{C} (1 - e^{(p-1)t}).$$

As a consequence, as long as $A(t) \geq 1$, we have

$$A(t) \leq (c/C)^{1/1-p} (1 - e^{(p-1)t})^{1/(1-p)}.$$

In the end, we have obtained an a priori bound on $A(t) = \int |\partial^\alpha \mu|^2(t)$ for $t \in (0, T]$, depending only on d, s, a, E and T , but not on the initial value $A(0)$. Then the proof can be concluded by an approximation argument. \square

Proof of Lemma B.2. We proceed by induction on s .

In the **first step** we prove the result for $s = 0$. Given $d \geq 1$ and $a \in (0, 1]$, we write

$$\int_{\mathbb{R}^d} \langle x \rangle^k |u(x)|^2 dx = \int_{\mathbb{R}^d} \langle x \rangle^k |u(x)|^a |u(x)|^{2-a} dx,$$

so, by Hölder's inequality,

$$\|u\|_{L^2_k}^2 \leq \|u\|_{L^1_{\frac{k}{a}}}^a \|u\|_{L^{\frac{2-a}{1-a}}}^{2-a}$$

(with $\frac{2-a}{1-a} = \infty$ if $a = 1$). Then by Sobolev embedding,

$$\|u\|_{L_k^2}^2 \leq C(d, a) \|u\|_{L_{\frac{k}{a}}^1}^a \|u\|_{H^1}^{2-a},$$

where $a = 1$ if $d = 1$, a is arbitrary in $(0, 1)$ if $d = 2$, and $a = \frac{4}{d+2}$ if $d \geq 3$, that is,

$$\|u\|_{L_k^2} \leq C(d) \|u\|_{L_{\frac{k}{a}}^1}^{1-\theta(d)} \|u\|_{H^1}^{\theta(d)}$$

where $\theta(1) = \frac{1}{2}$, any $\theta(2) \in \left(\frac{1}{2}, 1\right)$ for $d = 2$, and $\theta(d) = \frac{d}{d+2}$ for $d \geq 3$.

In the **second step** we let $s \geq 1$ and assume by induction that there exist some constants $C(d, s-1, k), h(d, s-1, k) \geq 0$ and $\theta(d, s-1) \in (0, 1)$ such that for all $u \in L_\infty^1(\mathbb{R}^d) \cap H^s(\mathbb{R}^d)$:

$$\|u\|_{H_k^{s-1}} \leq C(d, s-1, k) \|u\|_{L_{h(d, s-1, k)}^1}^{1-\theta(d, s-1)} \|u\|_{H^s}^{\theta(d, s-1)}.$$

Let then $u \in L_\infty^1(\mathbb{R}^d) \cap H^{s+1}(\mathbb{R}^d)$.

Given $\alpha \in \mathbb{N}^d$ with $|\alpha| = j$ and $1 \leq j \leq s$, we split α into $\alpha = \alpha_1 + \alpha_2$ with $|\alpha_2| = 1$, and integrate by parts:

$$\begin{aligned} \|\partial^\alpha u\|_{L_k^2}^2 &\leq k \|\partial^{\alpha_1} u\|_{L_{2k-2}^2} \|\partial^\alpha u\|_{L^2} + \|\partial^{\alpha_1} u\|_{L_{2k}^2} \|\partial^{\alpha+\alpha_2} u\|_{L^2} \\ &\leq (k+1) \|\partial^{\alpha_1} u\|_{L_{2k}^2} \sup_{|\alpha| \leq j+1} \|\partial^\alpha u\|_{L^2}, \end{aligned}$$

whence

$$\begin{aligned} \sup_{|\alpha|=j} \|\partial^\alpha u\|_{L_k^2}^2 &\leq (k+1) \sup_{|\alpha|=j-1} \|\partial^\alpha u\|_{L_{2k}^2} \sup_{|\alpha| \leq j+1} \|\partial^\alpha u\|_{L^2} \\ &\leq (k+1) \sup_{|\alpha| \leq s-1} \|\partial^\alpha u\|_{L_{2k}^2} \sup_{|\alpha| \leq s+1} \|\partial^\alpha u\|_{L^2}. \end{aligned}$$

Since this holds for any $1 \leq j \leq s$ we obtain

$$\sup_{1 \leq |\alpha| \leq s} \|\partial^\alpha u\|_{L_k^2}^2 \leq (k+1) \sup_{|\alpha| \leq s-1} \|\partial^\alpha u\|_{L_{2k}^2} \sup_{|\alpha| \leq s+1} \|\partial^\alpha u\|_{L^2}.$$

Moreover

$$\|u\|_{L_k^2}^2 \leq \|u\|_{L_{2k}^2} \|u\|_{L^2} \leq \sup_{|\alpha| \leq s-1} \|\partial^\alpha u\|_{L_{2k}^2} \sup_{|\alpha| \leq s+1} \|\partial^\alpha u\|_{L^2},$$

so that finally

$$\|u\|_{H_k^s}^2 \leq (k+1) \|u\|_{H_{2k}^{s-1}} \|u\|_{H^{s+1}}.$$

Then, by induction hypothesis,

$$\|u\|_{H_k^s}^2 \leq (k+1) C(d, s-1, 2k) \|u\|_{L_{h(d, s-1, 2k)}^1}^{1-\theta(d, s-1)} \|u\|_{H^s}^{\theta(d, s-1)} \|u\|_{H^{s+1}},$$

whence

$$\|u\|_{H_k^s} \leq C(d, k, s) \|u\|_{L_{h(d, s, k)}^1}^{1-\theta(d, s)} \|u\|_{H^{s+1}}^{\theta(d, s)}$$

where $\theta(d, s) = \frac{1}{2 - \theta(d, s - 1)} \in (0, 1)$ and $h(d, s, k) = h(d, s - 1, 2k) \geq 0$. This concludes the argument. □

Acknowledgments: The authors thank M. Ledoux for his relevant comments and his interest during the preparation of this work, as well as providing Reference [16].

REFERENCES

- [1] ARAUJO, A. AND GINÉ, E. The central limit theorem for real and Banach valued random variables *John Wiley & Sons, New York*, (1980).
- [2] BENACHOUR, S., ROYNETTE, B., TALAY, D., AND VALLOIS, P. Nonlinear self-stabilizing processes. I. Existence, invariant probability, propagation of chaos. *Stochastic Process. Appl.* **75**, 2 (1998), 173–201.
- [3] BENACHOUR, S., ROYNETTE, B., AND VALLOIS, P. Nonlinear self-stabilizing processes. II. Convergence to invariant probability. *Stochastic Process. Appl.* **75**, 2 (1998), 203–224.
- [4] BENEDETTO, D., CAGLIOTI, E., CARRILLO, J. A., AND PULVIRENTI, M. A non-Maxwellian steady distribution for one-dimensional granular media. *J. Statist. Phys.* **91**, 5-6 (1998), 979–990.
- [5] BENEDETTO, D., CAGLIOTI, E., AND PULVIRENTI, M. A kinetic equation for granular media. *RAIRO Modél. Math. Anal. Numér.* **31**, 5 (1997), 615–641.
- [6] BOBKOV, S., GENTIL, I., AND LEDOUX, M. Hypercontractivity of Hamilton-Jacobi equations. *J. Math. Pures Appl.* **80**, 7 (2001), 669–696.
- [7] BOBKOV, S., AND GÖTZE, F. Exponential integrability and transportation cost related to logarithmic Sobolev inequalities. *J. Funct. Anal.* **163** (1999), 1–28.
- [8] BOLLEY, F., AND VILLANI, C. Weighted Csiszár-Kullback-Pinsker inequalities and applications to transportation inequalities. To appear in *Ann. Fac. Sci. Toulouse*. Available online via <http://www.umpa.ens-lyon.fr/~fbolley>, 2004.
- [9] CARRILLO, J. A., MCCANN, R. J., AND VILLANI, C. Kinetic equilibration rates for granular media and related equations: entropy dissipation and mass transportation estimates. *Rev. Mat. Iberoamericana* **19**, 3 (2003), 971–1018.
- [10] CARRILLO, J. A., MCCANN, R. J., AND VILLANI, C. Contractions in the 2-Wasserstein length space and thermalization of granular media. Preprint, 2004.
- [11] CATTIAUX, P., AND GUILLIN, A. Talagrand’s like quadratic transportation cost inequalities. Available online via <http://www.ceremade.dauphine.fr/~guillin/index3.html>. Preprint, 2004.
- [12] DEMBO, A., AND ZEITOUNI, O. *Large Deviations Techniques And Applications*, second ed. Springer Verlag, New York, 1998.
- [13] DESVILLETES, L., AND VILLANI, C. On the spatially homogeneous Landau equation for hard potentials. I. Existence, uniqueness and smoothness. *Comm. Partial Differential Equations* **25**, 1-2 (2000), 179–259.
- [14] DJELLOUT, H., GUILLIN, A., AND WU, L. Transportation cost-information inequalities and applications to random dynamical systems and diffusions. *Ann. Probab.* **32**, 3B (2004), 2702–2732.
- [15] GAO, F. Moderate deviations and large deviations for kernel density estimators. *J. Theor. Prob.*, **16** (2003), 401–418.
- [16] GINÉ, E. AND ZINN, J. Empirical processes indexed by Lipschitz functions *Ann. Probab.* **14**, 4 (1986), 1329–1338.
- [17] LEDOUX, M. *The concentration of measure phenomenon*, vol. 89 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, 2001.
- [18] LEDOUX, M. AND TALAGRAND, M., *Probability in Banach spaces*. Springer-Verlag, Berlin, 1991.

- [19] MALRIEU, F. Logarithmic Sobolev inequalities for some nonlinear PDE's. *Stochastic Process. Appl.* 95, 1 (2001), 109–132.
- [20] MALRIEU, F. Convergence to equilibrium for granular media equations and their Euler schemes. *Ann. Appl. Probab.* 13, 2 (2003), 540–560.
- [21] MARCHIORO, C., AND PULVIRENTI, M. *Mathematical theory of incompressible nonviscous fluids*. Springer-Verlag, New York, 1994.
- [22] MASSART, P. Saint-Flour Lecture Notes. Available at <http://www.math.u-psud.fr/~massart>, 2003.
- [23] OTTO, F., AND VILLANI, C. Generalization of an inequality by Talagrand, and links with the logarithmic Sobolev inequality. *J. Funct. Anal.* 173 (2000), 361–400.
- [24] PETROV, V. V. *Limit theorems of probability theory*. The Clarendon Press Oxford University Press, New York, 1995.
- [25] SCHOCHET, S. The point-vortex method for periodic weak solutions of the 2-D Euler equations. *Comm. Pure Appl. Math.* 49, 9 (1996), 911–965.
- [26] SION, M. On general minimax theorems. *Pac. J. Math.* 8 (1958), 171–176.
- [27] SZNITMAN, A.-S. Topics in propagation of chaos. In *École d'Été de Probabilités de Saint-Flour XIX—1989*, vol. 1464 of *Lecture Notes in Math*. Springer, Berlin, 1991.
- [28] TALAGRAND, M. Transportation cost for Gaussian and other product measures. *Geom. Funct. Anal.* 6 (1996), 587–600.
- [29] VILLANI, C. *Topics in optimal transportation*. Grad. Stud. Math. (58), American Mathematical Society, Providence, 2003.
- [30] WANG, F.-Y. Probability distance inequalities on Riemannian manifolds and path spaces. *J. Funct. Anal.* 206, 1 (2004), 167–190.

ENS LYON, UMPA, 46 ALLÉE D'ITALIE, F-69364 LYON CEDEX 07
E-mail address: fbolley@umpa.ens-lyon.fr

CEREMADE, UNIVERSITÉ PARIS DAUPHINE
E-mail address: guillin@ceremade.dauphine.fr

ENS LYON, UMPA, 46 ALLÉE D'ITALIE, F-69364 LYON CEDEX 07
E-mail address: cvillani@umpa.ens-lyon.fr