LONG-TIME BEHAVIORS OF MEAN-FIELD INTERACTING PARTICLE SYSTEMS RELATED TO MCKEAN-VLASOV EQUATIONS

WEI LIU, LIMING WU, AND CHAOEN ZHANG

ABSTRACT. In this paper, we investigate the concentration inequalities, the exponential convergence in the Wasserstein metric W_1, and uniform-in-time propagation of chaos for the mean-field weakly interacting particle system related to McKean-Vlasov equation. By means of the known approximate componentwise reflection coupling and with the help of some new cost function, we obtain explicit estimates for those three problems, avoiding the technical conditions in the known results. Our results apply when the confinement potential V has many wells, the interaction potential W has bounded second mixed derivative $\nabla^2_{xy}W$ which should be not big so that there is no phase transition. Two examples are provided to illustrate the sharpness and usefulness in mathematical physics of our results.

AMS 2010 Subject classifications. 82C22, 60J60, 60E15, 65C35, 65C05.

Key words and Phrases. McKean-Vlasov equation, interacting particle system, propagation of chaos, exponential convergence, concentration inequality.

1. INTRODUCTION

In this paper, we consider the following nonlinear McKean-Vlasov equation with initial condition u_0

$$\partial_t u_t = \nabla \cdot [\nabla u_t + u_t \nabla V + u_t(\nabla_x W \circledast u_t)], \tag{1.1}$$

where the unknown u_t is a time dependent probability density on $\mathbb{R}^d$ ($d \geq 1$), $V : \mathbb{R}^d \to \mathbb{R}$ is a confinement potential and $W : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is an interaction potential. Here ∇ and $\nabla \cdot$ (applied to a vector field) denote the gradient operator and the divergence operator respectively, while $\nabla_x W$ stands for the gradient of W with respect to (w.r.t. in short) the first variable, and

$$\nabla_x W \circledast u_t(x) := \int_{\mathbb{R}^d} \nabla_x W(x, y) u_t(y) dy.$$

When $W(x, y) = W_0(x - y)$ for some even potential W_0 (as in granular media), $\nabla_x W \circledast u = \nabla W_0 * u$ (the usual convolution).

The probabilistic equivalent version of (1.1) is the following self-interacting stochastic differential equation (SDE in short):

$$\begin{cases} dX_t = \sqrt{2}dB_t - \nabla V(X_t)dt - \nabla_x W \circledast \mu_t(X_t)dt, \\ X_0 \overset{law}{=} u_0(x)dx, \end{cases} \tag{1.2}$$

where μ_t is the law of X_t. The density u_t of the law μ_t of X_t with respect to (w.r.t. in short) the Lebesgue measure dx at time t is the solution of the McKean-Vlasov equation (1.1) and *vice versa*. The existence and uniqueness of the solution of the SDE (1.2) and the McKean-Vlasov equation (1.1) have been extensively studied. The reader is referred to [30, 17, 29, 34] and recent works [33, 7, 19] as well as the references therein. For the convergence to equilibrium of solution μ_t as $t \to +\infty$, it is worth mentioning that Carrillo, McCann and Villani [8] obtained the explicit exponential convergence in entropy under various kinds of convexity conditions on the potentials V and W_0, via their enlightening idea of interpreting the McKean-Vlasov equation as the gradient descent flow of the free energy on the space of probability measures equipped with the L^2-Wasserstein metric. Eberle et al. [15] got the quantitative bounds on the exponential convergence in some appropriate transport cost to equilibrium for McKean-Vlasov equations by using Lyapunov condition and reflection coupling. Eberle [14] showed the exponential contractivity for diffusion semigroups w.r.t. Kantorovich distance by using componentwise reflection coupling methods and choosing appropriate distance functions. The reader is referred also to Luo and Wang [25] for the exponential convergence of diffusion semigroups w.r.t. the L^p-Wasserstein distance for all $p \geq 1$.

The McKean-Vlasov equation (1.1) or (1.2) is the idealization of the following interacting particle system of mean-field type when the number N of particles goes to infinity:

$$\begin{cases} dX_t^{i,N} = \sqrt{2}dB_t^i - \nabla V(X_t^{i,N})dt - \frac{1}{N-1}\sum_{j:j\neq i,1\leq j\leq N} \nabla_x W(X_t^{i,N}, X_t^{j,N})dt, \\ X_0^{i,N} = X_0^i, \ i = 1, \cdots, N, \end{cases} \tag{1.3}$$

where the initial values $X_0^1, \cdots, X_0^N$ are i.i.d. random variables with common law $\mu_0(dx) = u_0(x)dx$, and $B_t^1 \cdots, B_t^N$ are N independent Brownian motions taking values in $\mathbb{R}^d$, independent of $X_0^i, 1 \leq i \leq N$. In fact this is the goal of the studies of the so-called propagation of chaos: when the number N of particles goes to infinity, the empirical measures $\frac{1}{N}\sum_{i=1}^N \delta_{X_t^{i,N}}$ of the particle system (1.3) (or the law of a single particle) converge weakly to the solution μ_t of the self-interacting diffusion (1.2). This corresponds to the law of large number in probability, see the monograph [34] of Sznitman for propagation of chaos.

The propagation of chaos for the mean-field interacting particle systems has been widely studied during the last forty years. The early studies were concentrated on the propagation of chaos in bounded time intervals, see [26, 34, 29] and the references therein. The study on the propagation of chaos in the whole time interval $\mathbb{R}^+$ is much more difficult and recent. When the confinement potential V is strictly convex and the interaction potential $W(x,y) = W_0(x-y)$ with W_0 strictly convex, Malrieu [27] showed the uniform in time propagation of chaos by applying the logarithmic Sobolev

inequality. In the case that there is no confinement (i.e. $V \equiv 0$) and the interaction potential W_0 is strictly convex, Benachour et al. [1, 2] proved propagation of chaos (but not uniform in time) and polynomial convergence to equilibrium; Malrieu [28] obtained the uniform in time propagation of chaos and exponential convergence to equilibrium for the particle system viewed from the center, by using functional inequalities. When W_0 is degenerately convex and $V = 0$, Cattiaux et al. [9] showed the uniform in time propagation of chaos and exponential convergence to equilibrium by using synchronous coupling.

An important and actively studied subject refining the propagation of chaos is the concentration inequalities which are crucial for stochastic numerical computation of solution μ_t or the equilibrium μ_∞ of the McKean-Vlasov equation. The concentration inequalities describe quantitatively why the McKean-Vlasov equation is the idealization of the particle system (1.3). For the previous works on the concentration inequalities in the convex case we refer the reader to Malriau [28], Bolley-Guillin-Villani [9] and Bolley [4] and the references therein. See Section 2 for more details. The reader is referred to the two monographies of Ledoux [23, 24] for pedagogical and enlightening treatment of concentration inequalities.

We now go to the case where V and W_0 are no longer convex. Without the convexity of V and W_0, recently Durmus, Eberle, Guillin and Zimmer [13] use the componentwise reflection coupling introduced in [14] to prove the exponential convergence in some Wasserstein metric and uniform in time propagation of chaos for weakly interacting mean-field particle system. For more results about propagation of chaos, we refer the reader to [12, 20, 21, 22, 31, 32] and the references therein. In the actual non-convex case, *phase transition* can occur if the interaction is strong, and finding explicit estimates of the weakness of the interaction for the exponential convergence of the McKean-Vlasov equation is an important question in mathematical physics.

The main purpose of this paper is to investigate the concentration inequalities, the exponential convergence in L^1-Wasserstein metric W_1 (refining the previous results in [14, 13]), and as by-product the uniform-in-time propagation of chaos of the mean-field weakly interacting particle system. Although we use the same approximate componentwise reflection coupling ([14, 13]), our next approach will be quite different from theirs:

(1) our starting point is some explicit gradient estimate of the Poisson equation $-\mathcal{L}^{(N)}G = g$ where $\mathcal{L}^{(N)}$ is the generator of (1.3), which are crucial for the concentration inequalities of the interacting particle system;

(2) we will choose a different metric from the one used in [14, 13], which allows us to obtain some explicit and almost sharp estimate of the exponential rate in the convergence of the interacting particles system to its equilibrium in the W_1-metric, uniform in the number N of the particles.

(3) As a by-product, we obtain some explicit estimate on the propagation of chaos, uniform in time.

The paper is organized as follows. In the next section, we will present our framework and main results. The proofs of the results about gradient estimate of the Poisson equation (in heat diffusion) and on the exponential convergence in Wasserstein metric

are provided in Section 3 and Section 4. The proofs of concentration inequalities are given in the last section.

2. Main resutls

2.1. **Framework: notations and conditions.**

2.1.1. *Conditions on the dissipativity rate of a single particle.* First we introduce a dissipative rate $b_0(r)$ of the drift of one single particle in (1.3) at distance $r > 0$. The function $b_0(r)$ is a continuous function on $(0, +\infty)$, such that

$$\langle x - y, -[\nabla V(x) - \nabla V(y)] - [\nabla_x W(x, z) - \nabla_x W(y, z)] \rangle \leq b_0(r)|x - y| \qquad (2.1)$$

for any $x, y, z \in \mathbb{R}^d$ with $|x - y| = r$. Throughout this paper we assume that $b_0(r)$ satisfies

$$\limsup_{r \to +\infty} \frac{b_0(r)}{r} < 0, \qquad (2.2)$$

i.e. the drift of one particle is dissipative at infinity.

We also assume that

$$\lim_{r \to 0+} b_0^+(r) = 0. \qquad (2.3)$$

Next we introduce an important reference function h which is different from the one used in [14, 13]. For any function $f \in C^2(0, +\infty)$ and $r > 0$, let $\mathcal{L}_{ref}$ be the generator defined by

$$\mathcal{L}_{ref} f(r) := 4f''(r) + b_0(r)f'(r). \qquad (2.4)$$

Let $h : \mathbb{R}^+ \to \mathbb{R}^+$ be the function determined by: $h(0) = 0$ and

$$h'(r) = \frac{1}{4} \exp\left(-\frac{1}{4} \int_0^r b_0(s)ds\right) \int_r^{+\infty} s \cdot \exp\left(\frac{1}{4} \int_0^s b_0(u)du\right) ds. \qquad (2.5)$$

As h' is C^1-smooth by the continuity of b_0, h is a well defined C^2 function, and it is a solution (the smallest in fact) of the one-dimensional Poisson equation

$$\mathcal{L}_{ref} h(r) = 4h''(r) + b_0(r)h'(r) = -r, \ r > 0 \qquad (2.6)$$

with $h(0) = 0$. This function was used by the second named author [35] for functional and isoperimetric inequalities on Riemmanian manifolds.

2.1.2. *Kantorovich-Wasserstein W_1-metric.* For the configuration space $(\mathbb{R}^d)^N$, instead of the usual Euclidean metric, we will use the l^1-metric (generalized Hamming metric)

$$d_{l^1}(x, y) = \sum_{i=1}^N |x^i - y^i|, \ \ x = (x^1, \cdots, x^N), \ y = (y^1, \cdots, y^N) \in (\mathbb{R}^d)^N.$$

We consider the Kantorovich-Wasserstein distance w.r.t. d_{l^1} metric on $(\mathbb{R}^d)^N$, i.e., for any two probability measures μ and ν on $(\mathbb{R}^d)^N$,

$$W_{1,d_{l^1}}(\mu, \nu) = \inf_{P \in \Pi(\mu,\nu)} \iint_{(\mathbb{R}^d)^N \times (\mathbb{R}^d)^N} d_{l^1}(x, y)P(dx, dy)$$

where $\Pi(\mu, \nu)$ is the set of all couplings of μ, ν, i.e. the set of all probability measures on $(\mathbb{R}^d)^N \times (\mathbb{R}^d)^N$ whose marginal distributions of x and y are respectively μ and ν.

Notice that for a C^1-function g on $(\mathbb{R}^d)^N$, its Lipschitzian norm $\|g\|_{Lip(d_{l^1})}$ w.r.t. d_{l^1} coincides with $\max_{1 \leq i \leq N} \|\nabla_i g\|_\infty$ where ∇_i is the gradient w.r.t. x_i. By Kantorovich-Rubinstein duality relation,

$$W_{d_{l^1}}(\mu, \nu) = \sup_{g \in C_b^1((\mathbb{R}^d)^N) : \max_{1 \leq i \leq N} \|\nabla_i g\|_\infty \leq 1} \left(\int g d\mu - \int g d\nu \right)$$

When $N = 1$, we write simply W_1 for $W_{1,d_{l^1}}$. We write quite often $\mu(u) := \int u d\mu$.

We notice that *for two probability measures μ, ν on $(\mathbb{R}^d)^N$,*

$$\sum_{i=1}^{N} W_1(\mu^i, \nu^i) \leq W_{d_{l^1}}(\mu, \nu) \tag{2.7}$$

and the equality holds when $\mu = \otimes_{i=1}^N \mu^i, \nu = \otimes_{i=1}^N \nu^i$ are product measures, where μ^i (resp. ν^i) is the marginal distribution of x_i of μ (resp. ν). In fact if $X = (X^1, \cdots, X^N), Y = (Y^1, \cdots, Y^N)$ are two random vectors such that the law of (X, Y) is an optimal coupling of (μ, ν) in $W_{1,d_{l^1}}$, then for each i, the law of (X^i, Y^i) is a coupling of (μ^i, ν^i), so

$$W_{d_{l^1}}(\mu, \nu) = \mathbb{E} d_{l^1}(X, Y) = \sum_{i=1}^{N} \mathbb{E}|X^i - Y^i| \geq \sum_{i=1}^{N} W_1(\mu^i, \nu^i).$$

When μ, ν are product measures, let (X^i, Y^i) (or its joint law) be an optimal coupling of (μ^i, ν^i) for $W_1(\mu^i, \nu^i)$ so that $(X^1, Y^1), \cdots, (X^N, Y^N)$ are independent. Then $(X = (X^i)_{1 \leq i \leq N}, Y = (Y^i)_{1 \leq i \leq N})$ is a coupling of (μ, ν), so we get

$$\sum_{i=1}^{N} W_1(\mu^i, \nu^i) = \sum_{i=1}^{N} \mathbb{E}|X^i - Y^i| = \mathbb{E} d_{l^1}(X, Y) \geq W_{d_{l^1}}(\mu, \nu)$$

i.e. the equality in (2.7) holds in the prodcut measures case. (This is well known.)

2.2. **An explicit gradient estimate of the Poisson equation.** Let $\{P_t^{(N)}\}_{t \geq 0}$ be the transition semigroup of the mean-field interacting particle system (1.3), whose generator is given by

$$\mathcal{L}^{(N)} f(x^1, \cdots, x^N) = \sum_{i=1}^{N} \left(\Delta_i f - \nabla V(x^i) \cdot \nabla_i f - \frac{1}{N-1} \sum_{j \neq i} \nabla_x W(x^i, x^j) \cdot \nabla_i f \right).$$

Its unique invariant probability measure is the mean-field Gibbs measure, given by

$$\mu^{(N)}(dx^1, \cdots, dx^N) = \frac{1}{C_N} \exp \left(-\sum_{i=1}^{N} V(x^i) - \frac{1}{N-1} \sum_{1 \leq i < j \leq N} W(x^i, x^j) \right) dx^1 \cdots dx^N,$$

where C_N is the normalization constant.

We introduce the following key assumption on the interaction potential:

$$(\mathbf{H}) : \ \|\nabla_{xy}^2 W\|_\infty \cdot \|h'\|_\infty < 1$$

where h is given by (2.5), $\|h'\|_\infty := \sup_{r\geq 0} h'(r)$, and $\nabla^2_{xy} W = (\frac{\partial^2}{\partial x_i \partial y_j} W)_{1\leq i,j\leq d}$,

$$\|\nabla^2_{xy} W\|_\infty := \sup_{x,y\in\mathbb{R}^d} \sup_{z\in\mathbb{R}^d, |z|=1} |\nabla^2_{xy} W(x,y)z|.$$

Notice that when the dissipativity at infinity condition (2.2) is satisfied, $b_0(r)$ can be bounded from above by $-c_1 r + c_2$ (with $c_1, c_2 > 0$), so $\|h'\|_\infty := \sup_{r\geq 0} h'(r) < +\infty$.

The assumption (**H**) is a (non-trivial) translation of Dobrushin-Zegarlinski's uniqueness condition in the framework of mean field, and it implies that the mean field has no phase transition (established by Guillin and us in [18]).

Notice that under the assumption (**H**) and (2.2), both the equations (1.2) and (1.3) have unique strong solutions. On the space of continuous paths $C([0,T], (\mathbb{R}^d)^N)$ where $T \in (0, +\infty]$, we consider the L^1-metric

$$d_{L^1[0,T]}(\gamma_1, \gamma_2) := \int_0^T d_{l^1}(\gamma_1(t), \gamma_2(t))dt. \tag{2.8}$$

Given the starting point $x \in (\mathbb{R}^d)^N$, let $\mathbb{P}_x$ be the law of $X^{(N)} = (X_t^{(N)})_{t\geq 0}$ with $X_0^{(N)} = x$.

Theorem 2.1. *Assume (2.2), (2.3) and (**H**). For any $x_0 = (x_0^1, \cdots, x_0^N) \in (\mathbb{R}^d)^N$ and $y_0 = (y_0^1, \cdots, y_0^N) \in (\mathbb{R}^d)^N$, we have*

$$\int_0^{+\infty} W_{d_{l^1}}(P_t^{(N)}(x_0, \cdot), P_t^{(N)}(y_0, \cdot))dt \leq W_{1,d_{L^1[0,\infty]}}(\mathbb{P}_{x_0}, \mathbb{P}_{y_0})$$

$$\leq \frac{1}{1 - \|\nabla^2_{xy} W\|_\infty \|h'\|_\infty} \sum_{i=1}^N h(|x_0^i - y_0^i|). \tag{2.9}$$

In particular for any $g \in C_b^1((\mathbb{R}^d)^N)$ with $\mu^{(N)}(g) = 0$, the solution G of the Poisson equation $-\mathcal{L}^{(N)} G = g$ with $\mu^{(N)}(G) = 0$ satisfies

$$\|\nabla_i G\|_\infty \leq c_{Lip} \cdot \max_{1\leq j\leq N} \|\nabla_j g\|_\infty, 1 \leq i \leq N, \tag{2.10}$$

where

$$c_{Lip} := \frac{h'(0)}{1 - \|\nabla^2_{xy} W\|_\infty \|h'\|_\infty} \tag{2.11}$$

and

$$h'(0) = \frac{1}{4} \int_0^{+\infty} s \cdot \exp\left(\frac{1}{4} \int_0^s b_0(u)du\right) ds.$$

Its proof will be given in the next section. We present two applications of the theorem above, all of physical meaning. The first is an explicit integral (over time) estimate of the nonlinear McKean-Vlasov equation (1.1).

Corollary 2.2. *Under the same assumptions as in Theorem 2.1, for any two solutions μ_t, ν_t of the self-interacting diffusion (1.2) with the initial distributions μ_0, ν_0 with finite second moment respectively, we have*

$$\int_0^\infty W_1(\mu_t, \nu_t)dt \leq \frac{\|h'\|_\infty}{1 - \|\nabla^2_{xy} W\|_\infty \|h'\|_\infty} W_1(\mu_0, \nu_0). \tag{2.12}$$

In particular for the unique equilibrium μ_∞ of the McKean-Vlasov equation (see [18]),

$$\int_0^\infty W_1(\mu_t, \mu_\infty)dt \leq \frac{\|h'\|_\infty}{1 - \|\nabla_{xy}^2 W\|_\infty \|h'\|_\infty} W_1(\mu_0, \mu_\infty).$$

Proof. By (2.9) in Theorem 2.1 and the fact that

$$h(r) \leq h(0) + \|h'\|_\infty \cdot r = \|h'\|_\infty \cdot r, \ \forall r \geq 0$$

we have

$$\int_0^\infty W_{1,d_{l1}}(\mu_0^{\otimes N} P_t^{(N)}, \nu_0^{\otimes N} P_t^{(N)})dt \leq \frac{\|h'\|_\infty}{1 - \|\nabla_{xy}^2 W\|_\infty \|h'\|_\infty} W_{1,d_{l1}}(\mu_0^{\otimes N}, \nu_0^{\otimes N}). \quad (2.13)$$

Notice that $\mu_t^{(N)} := \mu_0^{\otimes N} P_t^{(N)}$ and $\nu_t^{(N)} := \nu_0^{\otimes N} P_t^{(N)}$ are symmetric probability measures on $(\mathbb{R}^d)^N$ and their marginal distributions $\mu_t^{(i,N)}$, $\nu_t^{(i,N)}$ of x_i converge weakly to μ_t, ν_t (respectively) by the finite time propagation of chaos. By using (2.7) we have

$$NW_1(\mu_t^{(1,N)}, \nu_t^{(1,N)}) = \sum_{i=1}^N W_1(\mu_t^{(i,N)}, \nu_t^{(i,N)}) \leq W_{1,d_{l1}}(\mu_t^{(N)}, \nu_t^{(N)})$$

and then by the lower semi-continuity of W_1 in the weak convergence topology,

$$W_1(\mu_t, \nu_t) \leq \liminf_{N \to +\infty} W_1(\mu_t^{(1,N)}, \nu_t^{(1,N)}) \leq \liminf_{N \to +\infty} \frac{1}{N} W_{1,d_{l1}}(\mu_t^{(N)}, \nu_t^{(N)}). \quad (2.14)$$

Combining (2.13) and (2.14) together, we obtain by Fatou's lemma,

$$\begin{aligned}
\int_0^\infty W_1(\mu_t, \nu_t)dt &\leq \liminf_{N \to +\infty} \frac{1}{N} \int_0^\infty W_{1,d_{l1}}(\mu_t^{(N)}, \nu_t^{(N)})dt \\
&\leq \frac{\|h'\|_\infty}{1 - \|\nabla_{xy}^2 W\|_\infty \|h'\|_\infty} \liminf_{N \to +\infty} \frac{1}{N} W_{1,d_{l1}}(\mu_0^{\otimes N}, \nu_0^{\otimes N}) \\
&= \frac{\|h'\|_\infty}{1 - \|\nabla_{xy}^2 W\|_\infty \|h'\|_\infty} W_1(\mu_0, \nu_0)
\end{aligned}$$

where the last equality follows by (2.7). That completes the proof. $\square$

Remark 2.3. Before the presentation of application of Theorem 2.1 to concentration inequalities in the next paragraph, we speak quickly the consequence of the gradient estimate to the spectral gap λ_1 of the particle system (1.3) in $L^2(\mu^{(N)})$ (i.e. the best constant $\lambda_1 \geq 0$ such that the Poincaré inequality below holds

$$\lambda_1 \text{Var}_{\mu^{(N)}}(F) \leq \int_{(\mathbb{R}^d)^N} \sum_{i=1}^N |\nabla_i F|^2 d\mu^{(N)}$$

for all bounded and C^2-smooth functions F on $(\mathbb{R}^d)^N$). In fact the Poisson operator $(-\mathcal{L}^{(N)})^{-1}$, defined on some dense domain (the whole space in fact) of $L_0^2(\mu^{(N)}) = \{F \in L^2(\mu^{(N)}); \ \mu^{(N)}(F) = 0\}$ is self-adjoint. Then as in [35], we have

$$\frac{1}{\lambda_1} = \|(-\mathcal{L}^{(N)})^{-1}\|_{L_0^2(\mu^{(N)})} \leq \|(-\mathcal{L}^{(N)})^{-1}\|_{Lip(d_{l1})} \leq c_{Lip}.$$

Thus

$$\lambda_1 \geq \frac{1}{c_{Lip}} = \frac{1 - \|\nabla_{xy}^2 W\|_\infty \|h'\|_\infty}{h'(0)}. \tag{2.15}$$

This estimate is more explicit (and better in the Example 2.15 of the double-well V) than the one in [13]. The reader is referred to our joint work with A. Guillin [18] on the spectral gap and the log-Sobolev inequality of Gibbs measure (without using coupling).

2.3. **Concentration inequality for the time average of the U-statistics.** We present now another application of Theorem 2.1 in the concentration inequality about the Gaussian concentration of the *U*-statistics.

For any $1 \leq m \leq N$, let $f_m : (\mathbb{R}^d)^m \to \mathbb{R}$ be a measurable and symmetric function. The *U*-statistic of order m with kernel f_m is defined by

$$U_N(f_m)(x^1, \cdots, x^N) = \frac{1}{|I_N^m|} \sum_{(i_1, \cdots, i_m) \in I_N^m} f_m(x^{i_1}, \cdots, x^{i_m}), \ \forall (x^1, \cdots, x^N) \in (\mathbb{R}^d)^N,$$

$$\tag{2.16}$$

where

$$I_N^m := \{(i_1, \cdots, i_m) \in \mathbb{N}^k | i_1, \cdots, i_m \text{ are different }, 1 \leq i_1, \cdots, i_m \leq N\} \tag{2.17}$$

and $|I_N^m|$ denotes the number of elements in I_N^m (equal to $N!/(N-m)!$).

Next we introduce the following *Gaussian integrability* assumption of the initial distribution μ_0:

$$\int_{\mathbb{R}^d} e^{\lambda_0 |x|^2} \mu_0(dx) < +\infty, \text{ for some } \lambda_0 > 0 \tag{2.18}$$

which is equivalent to say that there is some Gaussian concentration constant $c_G(\mu_0) > 0$ such that

$$\int_{\mathbb{R}} e^{f(x) - \mu_0(f)} d\mu_0(x) \leq \exp\left(\frac{c_G(\mu_0)}{2} \|f\|_{Lip}^2\right) \tag{2.19}$$

for all Lipschitzian functions f on $\mathbb{R}^d$ (w.r.t. the usual Euclidean distance).

Remark 2.4. The equivalence between the *Gaussian integrability* (2.18) and the *Gaussian concentration inequality* (2.19) was established by H. Djellout, A. Guillin and the second named author [10], and (2.19) is the famous characterization of Bobkov-Götze [3] of the transport-entropy inequality. By the tensorization of the transport-entropy inequality for product measure, (2.19) implies that for any $N \geq 1$,

$$\int_{(\mathbb{R}^d)^N} e^{g(x) - \mu_0^{\otimes N}(g)} d\mu_0^{\otimes N}(x) \leq \exp\left(\frac{N}{2} c_G(\mu_0) \|g\|_{Lip(d_{l^1})}^2\right) \tag{2.20}$$

for all Lipschitzian functions g on $(\mathbb{R}^d)^N$.

Theorem 2.5. *Assume the conditions in Theorem 2.1 and the Gaussian integrability* (2.18) *of the initial distribution μ_0. Let $f_m \in C^2((\mathbb{R}^d)^m, \mathbb{R})$ be symmetric and 1-Lipschitz*

w.r.t. the d_{l^1}-metric on $(\mathbb{R}^d)^m$, i.e. $\max_i \|\nabla_i f\|_\infty \leq 1$. Then for any $\lambda, T > 0$, we have

$$\mathbb{E} \exp \left(\frac{\lambda}{T} \left[\int_0^T U_N(f_m)(X_t^{1,N}, \cdots, X_t^{N,N}) dt - \int_0^T \mathbb{E} f_m(X_t^{1,N}, \cdots, X_t^{m,N}) dt \right] \right)$$
$$\leq \exp \left(\frac{m^2 \lambda^2 c_{Lip}^2}{2NT} \left(1 + \frac{c_G(\mu_0)}{T} \right) \right), \tag{2.21}$$

where c_{Lip} is the same as given in (2.11). In particular we have for any $\delta > 0$

$$\mathbb{P} \left\{ \frac{1}{T} \int_0^T U_N(f_m)(X_t^{1,N}, \cdots, X_t^{N,N}) dt - \frac{1}{T} \int_0^T \mathbb{E} f_m(X_t^{1,N}, \cdots, X_t^{m,N}) dt > \delta \right\}$$
$$\leq \exp \left(-\frac{(1 - \|\nabla_{xy}^2 W\|_\infty \|h'\|_\infty)^2}{2m^2 (h'(0))^2 (1 + c_G(\mu_0)/T)} NT\delta^2 \right). \tag{2.22}$$

The explicit concentration inequality (2.22) is sharp when V is quadratic and $W(x, y) = Kxy$ (the Gaussian case), shown in Example 2.14. Its proof will given in §5 after a more general result, Proposition 5.1.

2.4. Exponential convergence of the particle system in the $W_{1,d_{l^1}}$-metric.

Theorem 2.6. *Assume (2.2) and (**H**). Suppose that there exists a constant $M \in \mathbb{R}$ such that*

$$b_0(r) \leq rM, \forall r > 0 \tag{2.23}$$

(this condition is stronger than (2.3)), then for any $\varepsilon > 0$ such that

$$K_\varepsilon := \frac{1 - \|\nabla_{xy}^2 W\|_\infty \|h'\|_\infty - \varepsilon(M + \|\nabla_{xy}^2 W\|_\infty)}{\|h'\|_\infty + \varepsilon} > 0, \tag{2.24}$$

we have for any $x_0, y_0 \in (\mathbb{R}^d)^N$

$$W_{d_{l^1}}(P_t^{(N)}(x_0, \cdot), P_t^{(N)}(y_0, \cdot)) \leq A_\varepsilon e^{-K_\varepsilon t} d_{l^1}(x_0, y_0), \ \forall t \geq 0, \tag{2.25}$$

where

$$A_\varepsilon = \sup_{r>0} \frac{r}{h(r) + \varepsilon r} \cdot \sup_{r>0} \frac{h(r) + \varepsilon r}{r}. \tag{2.26}$$

Its proof is given in Section 3.

Remark 2.7. An easy estimate of A_ε is $A_\varepsilon \leq \frac{\sup_{r \geq 0} h'(r) + \varepsilon}{\inf_{r \geq 0} h'(r) + \varepsilon}$ (since $h(0) = 0$). Note that when $M + \|\nabla_{xy}^2 W\|_\infty > 0$, the exponential rate K_ε increases (then better and better) as ε decreases to 0, but A_ε may explode once if $\inf_{r \geq 0} h'(r) = 0$ (that is the case if $V(x) = x^4 - x^2$ for example).

Remark 2.8. Notice that (2.23) is equivalent to say that

$$\nabla^2 V(x) + \nabla_{xx}^2 W(x, y) \geq -MI, \ x, y \in \mathbb{R}^d$$

i.e. the Bakry-Emery's curvature for one particle's motion is bounded from below by the constant $-M$. When $\kappa := -M - \|\nabla^2_{xy}W\|_\infty > 0$, we see that the Hessian of the Hamiltonian

$$H(x^1, \cdots, x^N) = \sum_{i=1}^{N} V(x^i) + \frac{1}{N-1} \sum_{1 \le i < j \le N} W(x^i, x^j)$$

is bounded from below by κI (this estimate of the lower bound of the Bakry-Emery curvature is sharp if $\nabla^2_{xy}W$ is constant and definitely nonnegative). Notice that when $M < 0$, we can take $b_0(r) = Mr$, so $h'(r) = -1/M$. Then $\kappa > 0$ if and only if **(H)** is satisfied. The advantage of our condition **(H)** (w.r.t. the positive curvature condition) is: it does not depend on the curvature but on the dissipativity at infinity, it holds even if V has many wells (non-convex) once if the interaction is weak enough so that there is no phase transition.

If $\kappa > 0$, we have by Bakry-Emery's curvature characterization

$$W_1(P_t^{(N)}(x, \cdot), P_t^{(N)}(y, \cdot)) \le e^{-\kappa t}|x - y|$$

in the Euclidean metric on $(\mathbb{R}^d)^N$. On the other hand as above $b_0(r) = Mr, h(r) = -r/M$, we see that $K_\varepsilon \to -M - \|\nabla^2_{xy}W\|_\infty = \kappa$ as $\varepsilon \to +\infty$, and $A_\varepsilon \equiv 1$, so (2.25) yields

$$W_{1,d_{l^1}}(P_t^{(N)}(x, \cdot), P_t^{(N)}(y, \cdot)) \le e^{-\kappa t}d_{l^1}(x, y), \tag{2.27}$$

a curious but not at all surprising phenomenon (it can be obtained by the synchronous coupling as indicated by a referee).

Theorem 2.6 above will give us an explicit exponential convergence in W_1 of the nonlinear McKean-Vlasov equation (1.1). For the exponential convergence in entropy of the nonlinear McKean-Vlasov equation (1.1) under the condition **(H)**, see Guillin and the authors [18].

Corollary 2.9. *Under the same assumptions as in Theorem 2.6, for any $\varepsilon > 0$ so that $K_\varepsilon > 0$ (i.e. (2.24)), we have for the solutions μ_t, ν_t of the self-interacting diffusion (1.2) with the initial distributions μ_0, ν_0 which have finite second moments respectively,*

$$W_1(\mu_t, \nu_t) \le A_\varepsilon e^{-K_\varepsilon t}W_1(\mu_0, \nu_0), \ \forall t \ge 0, \tag{2.28}$$

where K_ε and A_ε are given by (2.24) and (2.26) respectively.

Proof. The proof of this corollary is similar to that of Corollary 2.2, and we utilize the same notations as in the Corollary 2.2. First by Theorem 2.6, we have for any $t \ge 0$

$$W_{1,d_{l^1}}(\mu_0^{\otimes N}P_t^{(N)}, \nu_0^{\otimes N}P_t^{(N)}) \le A_\varepsilon e^{-K_\varepsilon t}W_{1,d_{l^1}}(\mu_0^{\otimes N}, \nu_0^{\otimes N}).$$

Dividing the inequality above by N, we obtain by (2.14) and propagation of chaos,

$$W_1(\mu_t, \nu_t) \le A_\varepsilon e^{-K_\varepsilon t}\liminf_{N \to +\infty} \frac{1}{N}W_{1,d_{l^1}}(\mu_0^{\otimes N}, \nu_0^{\otimes N})$$

$$= A_\varepsilon e^{-K_\varepsilon t}W_1(\mu_0, \nu_0)$$

the desired result. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

2.5. **Concentration inequality for the empirical mean uniform in time.** We go into more details in this subject, spoken already in the Introduction. Under the conditions that V is uniformly convex and W is convex, Malrieu [27] established logarithmic Sobolev inequality and then used its connection with optimal transport and concentration of measure to get the following non-asymptotic bounds on the deviation of the empirical mean of an observable f from $\mu_t(f)$,

$$\sup_{\|f\|_{Lip}\leq 1} \mathbb{P}\left\{|\frac{1}{N}\sum_{i=1}^{N} f(X_t^{i,N}) - \mu_t(f)| > \frac{A}{\sqrt{N}} + \delta\right\} \leq 2e^{-\lambda N\delta^2}, \ t > 0, \ \delta \geq 0 \quad (2.29)$$

where A and λ are positive constants depending on the particle system.

As pointed out in [6], this approach can lead to nice bounds but it is limited to a finite number of observables. Bolley-Guillin-Villani [6, Theorem 2.9] obtained for any $t > 0$ fixed and $\delta > 0$

$$\mathbb{P}\left\{\sup_{\|f\|_{Lip}\leq 1} |\frac{1}{N}\sum_{i=1}^{N} f(X_t^{i,N}) - \mu_t(f)| > \delta\right\} \leq C(1 + t\delta^{-2})e^{-K(t)N\delta^2}, \quad (2.30)$$

for all N big enough (quantifiable), where $K(t)$ depending on t is some explicitly computable constant. Furthermore, Bolley [4] got quantitative concentration inequalities on the sample path space with uniform norm, on a given time interval $[0, T]$, which implies (2.30) by projection at time $t \in [0, T]$.

Theorem 2.10. *Assume the conditions in Theorem 2.6 and the Gaussian integrability (2.18) of the initial distribution μ_0. Then for any Lipschitzian observable $f : \mathbb{R}^d \to \mathbb{R}$ with $\|f\|_{Lip} = 1$, $N \geq 2$, and for any $\delta > 0$,*

$$\mathbb{P}_{\mu_0^{\otimes N}}\left\{\frac{1}{N}\sum_{i=1}^{N} f(X_T^{i,N}) - \mathbb{E}_{\mu_0^{\otimes N}} f(X_T^{1,N}) > \delta\right\} \leq \exp\left(-\frac{NK_\varepsilon\delta^2}{A_\varepsilon^2\left[1 + 2c_G(\mu_0)K_\varepsilon e^{-2K_\varepsilon T}\right]}\right)$$
$$(2.31)$$

where $\varepsilon > 0$ is any small constant such that $K_\varepsilon > 0$ (see (2.24)), A_ε is the constant given by (2.26).

As the absolute value of the bias $|\mathbb{E}_{\mu_0^{\otimes N}} f(X_T^{1,N}) - \mu_T(f)| \leq A/\sqrt{N}$ by Remark 4.4, our result above generalizes Malrieu's result (2.29) to the case that V may have many wells. It will be proved in Section 5.

2.6. **Propagation of chaos in large time.** We have the following uniform in time propagation of chaos.

Theorem 2.11. *Assume (2.2), (2.23) and (**H**). Suppose that there exist some positive constants c_1, c_2, c_3 such that*

$$\langle x, \nabla V(x)\rangle \geq c_1|x|^2 - c_2, \forall x \in \mathbb{R}^d \quad (2.32)$$

and

$$\langle z, \nabla_{xx}^2 W(x,y)z\rangle \geq -c_3|z|^2, \forall x, y, z \in \mathbb{R}^d. \quad (2.33)$$

Assume

$$c_1 - c_3 - \|\nabla_{xy}^2 W\|_\infty > 0. \quad (2.34)$$

Then for any $\varepsilon > 0$ such that $K_\varepsilon > 0$, and $\tilde{\varepsilon} \in (0, c_1 - c_3 - \|\nabla_{xy}^2 W\|_\infty)$, the following estimates of propagation of chaos hold for the mean-field interacting particle system (1.3) with any initial probability measure μ_0 having finite second moment:

(a) **(path-type propagation of chaos)** *for any $T > 0$, $1 \leq k \leq N$, denote $\mathbb{P}_\nu(\cdot) = \int_{(\mathbb{R}^d)^N} \mathbb{P}_x(\cdot) d\nu(x)$ the law of $(X_t^{(N)})_{t \geq 0}$ with the initial distribution ν, $\mathbb{P}_\nu^{[1,k],N}|_{[0,T]}$ the joint law of paths of the k particles $((X_t^{i,N})_{t \in [0,T]}, 1 \leq i \leq k)$ in time interval $[0,T]$, and $\mathbb{Q}_{\mu_0}$ the law of the self-interacting diffusion $(X_t)_{t \geq 0}$ with the initial distribution μ_0. We have*

$$\frac{1}{kT} W_{1, d_{L^1[0,T]}}(\mathbb{P}_{\mu_0^{\otimes N}}^{[1,k],N}|_{[0,T]}, \mathbb{Q}_{\mu_0}^{\otimes k}|_{[0,T]}) \leq \frac{\max\{m_2(\mu_0), \hat{c}(\varepsilon)\}}{\sqrt{N-1}} \frac{\|\nabla_{xy}^2 W\|_\infty \|h'\|_\infty}{1 - \|\nabla_{xy}^2 W\|_\infty \|h'\|_\infty} \quad (2.35)$$

where

$$m_2(\mu_0) = \left(\int_{\mathbb{R}^d} |x|^2 d\mu_0(x) \right)^{\frac{1}{2}}, \quad \hat{c}(\varepsilon) = \left(\frac{d + c_2 + \frac{1}{4\tilde{\varepsilon}} |\nabla_x W(0,0)|^2}{c_1 - c_3 - \|\nabla_{xy}^2 W\|_\infty - \tilde{\varepsilon}} \right)^{\frac{1}{2}}. \quad (2.36)$$

(b) **(Uniform in time propagation of chaos)** *Assume moreover for all time $t > 0$ and any $1 \leq k \leq N$:*

$$W_{1, d_{l^1}}(\mu_t^{[1,k],N}, \mu_t^{\otimes k}) \leq \frac{k}{\sqrt{N-1}} \frac{A_\varepsilon}{K_\varepsilon} \|\nabla_{xy}^2 W\|_\infty \max\{m_2(\mu_0), \hat{c}(\varepsilon)\} \quad (2.37)$$

where $\mu_t = u_t dx$ is the solution of the McKean-Vlasov equation (1.1), and $\mu_t^{[1,k],N}$ is the joint law of the k particles $(X_t^{i,N}, 1 \leq i \leq k)$ in the mean-field system (1.3) of interacting particles $(X_t^{i,N})_{1 \leq i \leq N}$ with $X_0^{i,N}, 1 \leq i \leq N$ i.i.d. of law μ_0 (independent of $(B_t^{i,N})_{1 \leq i \leq N, t \geq 0}$), and the constants K_ε, A_ε, $m_2(\mu_0)$ and $\hat{c}(\varepsilon)$ are given in (2.24), (2.26) and (2.36) respectively.

Remark 2.12. The time-uniform propagation of chaos is much more difficult than the bounded time propagation of chaos, accomplished in the 80-90's of the last century. The physical reason is that the time-uniform propagation of chaos fails in the regime of phase transition. That is why we impose the condition **(H)**, which excludes the phase transition.

The reader is referred to [5, 9, 12, 13, 22] and the references therein for recent studies and progresses on this subject. The main new point here is that our estimate (2.37) is explicit and relatively neat.

Remark 2.13. All the results presented in this paper can be extended to more general case:

$$dX_t = \sqrt{2} dB_t + b(X_t, \mu_t) dt$$

where μ_t is the law of X_t, if b satisfies some dissipative condition in x (uniformly in μ) and a Lipchitz condition in μ with sufficiently small Lipchitz constant. For the sake of clarity, we deal only with the case of $b(X_t, \mu_t) = -\nabla V(X_t) - \nabla_x W \circledast \mu_t(X_t)$ in this paper.

2.7. **Examples.** We first present the Gaussian model for which the constants in Theorem 2.1 and Theorem 2.6 become exact, showing their sharpness.

Example 2.14. (Gaussian model) *Let* $d = 1$, *and*

$$V(x) = \beta \frac{x^2}{2}, \ W(x, y) = -\beta K x y$$

where $\beta > 0$ *is the inverse temperature,* $K \geq 0$.

For this model, by some simple calculations we have

$$b_0(r) = -\beta r, \ \forall r > 0.$$

and

$$h'(r) \equiv \beta^{-1}, \ \forall r \geq 0.$$

It is obvious that conditions (2.2) and (2.3) hold, and the assumption (**H**) holds once if

$$K < 1. \tag{2.38}$$

But this condition is equivalent to say that the matrix $A = (a_{ij})_{1 \leq i,j \leq N}$ is positively definite, where

$$a_{ii} = \beta, a_{ij} = \frac{-\beta K}{N - 1}, i \neq j.$$

A must be the inverse of the covariance matrix of the Gaussian measure $\mu^{(N)}$. In other words (**H**) is equivalent to well defining the equilibrium probability measure $\mu^{(N)}$.

Note that $\|\nabla_{xy}^2 W\|_\infty = \beta K$, so we have $c_{Lip} = \frac{1}{\beta(1-K)}$ under (2.38). Moreover (2.23) is satisfied with $M = -\beta$.

• *Sharpness of Theorem 2.1.* The gradient estimate (2.10) in Theorem 2.1 tells us: if $-\mathcal{L}^{(N)} G = g$, then

$$\|\nabla_i G\|_\infty \leq \frac{1}{\beta(1 - K)} \max_i \|\nabla_i g\|_\infty.$$

Let us show that it becomes equality for $g(x^1, \cdots, x^N) = \sum_{i=1}^N x^i$. In fact

$$\mathcal{L}^{(N)} g(x^1, \cdots, x^N) = -\sum_i \beta x^i + \sum_i \frac{1}{N - 1} \sum_{j \neq i} \beta K x^j = -\beta(1 - K) g.$$

In other words $G = \frac{1}{\beta(1-K)} g$ for which the gradient estimate above becomes equality. As the gradient estimate (2.10) comes from (2.9), the process level $W_{1,d_{L^1}}$ estimate (2.9) is sharp too.

• *Sharpness of Theorem 2.6.* As $\varepsilon \to +\infty$ in (2.24), we have by Theorem 2.6

$$W_{1,d_{l^1}}(P_t^{(N)}(x_0, \cdot), P_t^{(N)}(y_0, \cdot)) \leq e^{-\beta(1-K)t} d_{l^1}(x_0, y_0).$$

This is equivalent to say that

$$\max \|\nabla_i P_t^{(N)} g\|_\infty \leq e^{-\beta(1-K)t} \max \|\nabla_i g\|_\infty.$$

But it becomes equality for $g = \sum_{i=1}^N x^i$: in fact as $\mathcal{L}^{(N)} g = -\beta(1 - K)g$,

$$P_t^{(N)} g = e^{-\beta(1-K)t} g.$$

Hence the exponential convergence result (2.25) in Theorem 2.6 is sharp.

Of course for this Gaussian model all results in Theorems 2.1 and 2.6 can be derived easily by using the synchronous coupling, or from the commutativity relation

$$\nabla P_t^{(N)} g = e^{-At} P_t^{(N)} \nabla g$$

which is one of the origins of the Bakry-Emery curvature.

Next we consider a typical physical model to illustrate our results.

Example 2.15. (Double-Well confinement potential and quadratic interaction in granular media) *Let $d \geq 1$,*

$$V(x) = \beta(|x|^4/4 - |x|^2/2), \ \ W(x,y) = \beta K |x-y|^2$$

where $\beta > 0$ is the inverse temperature, $K \in \mathbb{R}$. This model has the double-well confinement potential and quadratic interaction potential $W_0(z) = K|z|^2$. Here and hereafter $|x|$ denotes the Euclidean norm of x, $\langle \cdot, \cdot \rangle$ the Euclidean inner product.

First of all, for this model we have

$$b_0(r) = \beta r(1 - 2K - r^2/4), \ \ \forall r > 0 \tag{2.39}$$

and so conditions (2.2) and (2.3) are satisfied. Indeed, set $r = |x-y|$, $\sigma = \frac{x-y}{|x-y|}$ for any $x \neq y$, then

$$\begin{aligned}
\left\langle \frac{x-y}{|x-y|}, x|x|^2 - y|y|^2 \right\rangle &= \langle \sigma, (y+r\sigma)|y+r\sigma|^2 - y|y|^2 \rangle \\
&= 2r\langle \sigma, y \rangle^2 + 3r^2\langle \sigma, y \rangle + r|y|^2 + r^3 \\
&\geq 3r\langle \sigma, y \rangle^2 + 3r^2\langle \sigma, y \rangle + r^3 \\
&\geq \frac{r^3}{4}
\end{aligned}$$

with equality if and only if $x = \frac{r}{2}\sigma$, $y = -\frac{r}{2}\sigma$. Therefore we obtain

$$\begin{aligned}
&\left\langle \frac{x-y}{|x-y|}, \nabla V(x) - \nabla V(y) + \nabla_x W(x,z) - \nabla_x W(y,z) \right\rangle \\
&= \left\langle \frac{x-y}{|x-y|}, \beta(x|x|^2 - x) - \beta(y|y|^2 - y) + 2\beta K(x-z) - 2\beta K(y-z) \right\rangle \\
&\geq \beta r \left(\frac{r^2}{4} - 1 + 2K \right).
\end{aligned}$$

Then the best $b_0(r)$ is given by (2.39).

Next we estimate $\|h'\|_\infty$. By (2.5) and some calculations, we have for any $r \geq 0$

$$\begin{aligned}
h'(r) &= \frac{1}{4} \exp\left(\frac{\beta}{64}(r^4 - 8(1-2K)r^2) \right) \int_r^{+\infty} s \cdot \exp\left(\frac{\beta}{64}(8(1-2K)s^2 - s^4) \right) ds \\
&= \frac{1}{4} \exp\left(\frac{\beta}{16}\left(\frac{r^2}{2} - 2(1-2K) \right)^2 \right) \int_{r^2/2}^{+\infty} \exp\left(-\frac{\beta}{16}(u - 2(1-2K))^2 \right) du.
\end{aligned}$$

When $K > \frac{1}{2}$, we have $\frac{r^2}{2} - 2(1 - 2K) > 0$ and so

$$
\begin{aligned}
h'(r) &\leq \frac{1}{4} \exp\left(\frac{\beta}{16}\left(\frac{r^2}{2} - 2(1 - 2K) \right)^2 \right) \cdot \sqrt{\frac{16\pi}{\beta}} \exp\left(-\frac{\beta}{16}(u - 2(1 - 2K))^2 \right) \\
&= \frac{\sqrt{\pi}}{\sqrt{\beta}}.
\end{aligned}
\tag{2.40}
$$

When $K \leq \frac{1}{2}$ and $\frac{r^2}{2} > 2(1 - 2K)$, the above bound (2.40) holds as well.
When $K \leq \frac{1}{2}$ and $\frac{r^2}{2} \leq 2(1 - 2K)$, we have by (2.6)

$$
4h''(r) = -r - \beta r(1 - 2K - r^2/4)h'(r) \leq 0,
$$

and hence

$$
\begin{aligned}
h'(r) \leq h'(0) &= \frac{1}{4} \exp\left(\frac{\beta(1 - 2K)^2}{4} \right) \int_0^{+\infty} \exp\left(-\frac{\beta}{16}(u - 2(1 - 2K))^2 \right) du \\
&< \frac{\sqrt{\pi}}{\sqrt{\beta}} \exp\left(\frac{\beta(1 - 2K)^2}{4} \right).
\end{aligned}
\tag{2.41}
$$

Combining (2.40) and (2.41), we obtain

$$
\|h'\|_\infty < \alpha := \begin{cases} \frac{\sqrt{\pi}}{\sqrt{\beta}} \exp\left(\frac{\beta(1-2K)^2}{4} \right), & \text{if } K \leq \frac{1}{2}, \\ \frac{\sqrt{\pi}}{\sqrt{\beta}}, & \text{if } K > \frac{1}{2}. \end{cases}
\tag{2.42}
$$

Since $\|\nabla_{xy}^2 W\|_\infty = 2|K|\beta$, assumption (**H**) holds once if

$$
2\beta|K|\alpha < 1.
\tag{2.43}
$$

and then the conclusion of Theorem 2.1 holds under (2.43), and then all its consequences in §2.2 and 2.3. For instance

$$
c_{Lip} \leq \frac{\sqrt{\pi}}{\sqrt{\beta}} \frac{\exp\left(\frac{\beta(1-2K)^2}{4} \right)}{1 - 2|K|\beta\alpha}, \quad \lambda_1 \geq \frac{1}{c_{Lip}} \geq \frac{\sqrt{\beta}(1 - 2|K|\beta\alpha)}{\sqrt{\pi} \exp\left(\frac{\beta(1-2K)^2}{4} \right)}.
\tag{2.44}
$$

Furthermore, note that (2.23) holds with $M = \beta(1 - 2K)$, and then

$$
M + \|\nabla_{xy}^2 W\|_\infty = \beta(1 + 4K^-) > 0.
$$

Thus the conclusion of Theorem 2.6 holds with

$$
K_\varepsilon = \frac{1 - \alpha - \varepsilon\beta(1 + 4K^-)}{\alpha + \varepsilon}
$$

where $\varepsilon > 0$ is small such that $1 - \alpha - \varepsilon\beta(1 + 4K^-) > 0$, and by Remark 2.7,

$$
A_\varepsilon \leq \frac{\sup_{r \geq 0} h'(r) + \varepsilon}{\inf_{r \geq 0} h'(r) + \varepsilon} \leq \frac{\alpha + \varepsilon}{\varepsilon}.
$$

For the result of propagation of chaos in Theorem 2.11, we can take $c_3 = 0$ when $K \geq 0$, and $c_3 = -2K\beta$ when $K < 0$. To ensure that conditions (2.32) and (2.34) are

satisfied, one can take $c_1 = 2|K|\beta + \varepsilon'$, $c_2 = \frac{\beta}{4}(1 + 2|K| + \frac{\varepsilon'}{\beta})^2$ in the case of $K > 0$ and $c_1 = -4K\beta + \varepsilon'$, $c_2 = \frac{\beta}{4}(1 - 4K + \frac{\varepsilon'}{\beta})^2$ in the case of $K < 0$, for any $\varepsilon' > 0$.

3. Proof of Theorem 2.1 and Theorem 2.6

3.1. **Coupling.** We first introduce the approximate componentwise reflection coupling by following [13] and [14]. Given $\delta > 0$, let $\lambda_\delta, \pi_\delta : \mathbb{R}^+ \to [0, 1]$ be two Lipschitz continuous functions such that

$$\lambda_\delta(r)^2 + \pi_\delta(r)^2 = 1, \ \forall r \in \mathbb{R}^+ \tag{3.1}$$

and

$$\lambda_\delta(r) = \begin{cases} 1, & \text{if } r \geq \delta, \\ 0, & \text{if } r \leq \delta/2. \end{cases} \tag{3.2}$$

Then a coupling of two solutions of the mean-field interacting particle system (1.3) with initial values $x_0, y_0 \in (\mathbb{R}^d)^N$ is given by a strong solution of the system

$$
\begin{aligned}
dX_t^{i,N} &= \sqrt{2}[\lambda_\delta(|Z_t^i|)dB_t^{1,i} + \pi_\delta(|Z_t^i|)dB_t^{2,i}] - \nabla V(X_t^{i,N})dt \\
&\qquad\qquad - \frac{1}{N-1}\sum_{j:j\neq i, 1\leq j\leq N} \nabla_x W(X_t^{i,N}, X_t^{j,N})dt, \\
dY_t^{i,N} &= \sqrt{2}[\lambda_\delta(|Z_t^i|)R_t^i dB_t^{1,i} + \pi_\delta(|Z_t^i|)dB_t^{2,i}] - \nabla V(Y_t^{i,N})dt \\
&\qquad\qquad - \frac{1}{N-1}\sum_{j:j\neq i, 1\leq j\leq N} \nabla_x W(Y_t^{i,N}, Y_t^{j,N})dt,
\end{aligned}
\tag{3.3}
$$

$1 \leq i \leq N$. Here $Z_t^i := X_t^{i,N} - Y_t^{i,N}$ and $R_t^i := I_d - 2e_t^i(e_t^i)^T$, where I_d is the d-dimensional unit matrix and $e_t^i(e_t^i)^T$ is the orthogonal projection onto the unit vector $e_t^i := Z_t^i/|Z_t^i|$ if $|Z_t^i| \neq 0$, $B_t^1 = (B_t^{1,i})_{1\leq i\leq N}$ and $B_t^2 = (B_t^{2,i})_{1\leq i\leq N}$ are two given independent standard Brownian motions taking values in $(\mathbb{R}^d)^N$.

We will denote $X_t^{(N)} = (X_t^{1,N}, \cdots, X_t^{N,N})$, $Y_t^{(N)} = (Y_t^{1,N}, \cdots, Y_t^{N,N})$ and $Z_t^{(N)} := X_t^{(N)} - Y_t^{(N)}$. To see that $(X_t^{(N)}, Y_t^{(N)})$ is a coupling process, it is enough to notice that

$$
\begin{aligned}
\hat{B}_t^i &:= \int_0^t \lambda_\delta(|Z_s^i|)dB_s^{1,i} + \int_0^t \pi_\delta(|Z_s^i|)dB_s^{2,i} \\
\check{B}_t^i &:= \int_0^t \lambda_\delta(|Z_s^i|)R_t^i dB_s^{1,i} + \int_0^t \pi_\delta(|Z_s^i|)dB_s^{2,i}, 1 \leq i \leq N,
\end{aligned}
\tag{3.4}
$$

are standard Brownian motions on $(\mathbb{R}^d)^N$. That follows by Lévy's characterization of Brownian motions, because their brackets in matrix form are

$$
\begin{aligned}
d\langle \hat{B}_t^i, \hat{B}_t^j \rangle_t &= \lambda_\delta^2(|Z_s^i|)d\langle B^{1,i}, B^{1,j} \rangle_t + \pi_\delta^2(|Z_s^i|)d\langle B^{2,i}, B^{2,j} \rangle_t \\
&\quad + \lambda_\delta(|Z_s^i|)\pi_\delta(|Z_s^i|)(d\langle B^{1,i}, B^{2,j} \rangle_t + d\langle B^{1,j}, B^{2,i} \rangle_t) \\
&= \delta_{ij} I_d dt.
\end{aligned}
$$

The same for $\check{B}_t^i$. (That is already well detailed in [13] or [14]).

Remark 3.1. (1) The coupling (3.3) behaves as a reflection coupling when the distance between the two particles $X_t^{i,N}$ and $Y_t^{i,N}$ are larger than δ. When the particles are very close (with distance less than $\frac{1}{2}\delta$), they are driven by the same Brownian motion, i.e., it is a synchronous coupling. And when the distance is between $\frac{1}{2}\delta$ and δ, it is a mixture of reflection coupling and synchronous coupling. The aim is to make λ_δ and π_δ globally Lipschitz continuous, so that the coupling SDE has a unique strong solution, given the independent Brownian motions $B_t^{1,i}, B_t^{2,i}, 1 \le i \le N$.

(2) If one adopts the componentwise reflection coupling (i.e. the limit coupling when $\delta \to 0$), since $X^{i,N}, Y^{i,N}$ will separate after the time that they meet (i.e. $X_t^{i,N} = Y_t^{i,N}$), the local times will appear when Itô's formula is applied for $|X_t^{i,N} - Y_t^{i,N}|$. This makes the control of $\sum_{i=1}^N |X_t^{i,N} - Y_t^{i,N}|$ difficult to deal with. That is the reason why A. Eberle [14] introduced the synchronous coupling when $|X_t^{i,N} - Y_t^{i,N}|$ is small (we guess).

3.2. **Proof of Theorem 2.1.**

Proof of Theorem 2.1. **1). Proof of (2.9).** The first inequality in (2.9) is trivial, and next we prove the second inequality. By doing subtraction of the equations in (3.3), we have

$$dZ_t^i = 2\sqrt{2}\lambda_\delta(|Z_t^i|)e_t^i d\tilde{B}_t^i - [\nabla V(X_t^{i,N}) - \nabla V(Y_t^{i,N})]dt$$
$$- \frac{1}{N-1} \sum_{j:j\ne i, 1\le j\le N} [\nabla_x W(X_t^{i,N}, X_t^{j,N}) - \nabla_x W(Y_t^{i,N}, Y_t^{j,N})]dt, \quad (3.5)$$
$$Z_0^i = x_0^i - y_0^i,$$

where the processes $\tilde{B}_t^i = \int_0^t (e_s^i)^T dB_s^{1,i}, 1 \le i \le N$, are one-dimensional standard Brownian motions such that $\langle \tilde{B}^i, \tilde{B}^j \rangle_t = 0$ for $i \ne j$.

Let $r_t^i = |Z_t^i|, 1 \le i \le N$. Though $|\cdot|$ is not C^2 at 0, but since Z_t^i is of bounded variation and continuous when it is close to 0 (because of the synchronous coupling when X_t^i, Y_t^i are close), we can apply Itô's formula (see [13, Lemma 7]) to obtain

$$dr_t^i = 1_{\{r_t^i\ne 0\}} 2\sqrt{2}\lambda_\delta(r_t^i) d\tilde{B}_t^i - 1_{\{r_t^i\ne 0\}}\langle e_t^i, \nabla V(X_t^{i,N}) - \nabla V(Y_t^{i,N})\rangle dt$$
$$- 1_{\{r_t^i\ne 0\}}\langle e_t^i, \frac{1}{N-1} \sum_{j:j\ne i, 1\le j\le N} [\nabla_x W(X_t^{i,N}, X_t^{j,N}) - \nabla_x W(Y_t^{i,N}, Y_t^{j,N})]\rangle dt$$
$$+ 1_{\{r_t^i\ne 0\}} \sum_{k,l=1}^d [1_{\{k=l\}}(r_t^i)^{-1} - (X_t^{i,N,k} - Y_t^{i,N,k})(X_t^{i,N,l} - Y_t^{i,N,l})(r_t^i)^{-3}]\lambda_\delta(r_t^i)^2(I_d - R_t^i)_{kl}^2 dt,$$
$$(3.6)$$

where $X_t^{i,N,k}$ and $Y_t^{i,N,k}$ denote the k-th coordinate of $X_t^{i,N}$ and $Y_t^{i,N}$ respectively, $1 \le k \le d$. Notice that the last term in the right hand side of the above equation equals to

0 by an easy calculation. Hence we get

$$dr_t^i = 1_{\{r_t^i \neq 0\}} 2\sqrt{2}\lambda_\delta(r_t^i) d\tilde{B}_t^i - 1_{\{r_t^i \neq 0\}} \langle e_t^i, \frac{1}{N-1} \sum_{j:j \neq i, 1 \leq j \leq N} [\nabla_x W(X_t^{i,N}, X_t^{j,N}) - \nabla_x W(X_t^{i,N}, Y_t^{j,N})]\rangle dt$$

$$- 1_{\{r_t^i \neq 0\}} \langle e_t^i, \nabla V(X_t^{i,N}) - \nabla V(Y_t^{i,N}) + \frac{1}{N-1} \sum_{j:j \neq i, 1 \leq j \leq N} [\nabla_x W(X_t^{i,N}, Y_t^{j,N}) - \nabla_x W(Y_t^{i,N}, Y_t^{j,N})]\rangle dt$$

$$\leq 1_{\{r_t^i \neq 0\}} 2\sqrt{2}\lambda_\delta(r_t^i) d\tilde{B}_t^i + \frac{1}{N-1} \|\nabla_{xy}^2 W\|_\infty \sum_{j:j \neq i, 1 \leq j \leq N} r_t^j dt + 1_{\{r_t^i \neq 0\}} b_0(r_t^i) dt,$$

$$(3.7)$$

where we use the definition (2.1) of b_0 in the last inequality. Here $d\xi_t \leq d\eta_t$ means that $\eta_t - \xi_t$ is a non-decreasing process.

Let L_{λ_δ} be the generator defined by for any function $f \in C^2(0, +\infty)$ and $r > 0$,

$$L_{\lambda_\delta} f(r) := 4\lambda_\delta^2(r) f''(r) + b_0(r) f'(r). \tag{3.8}$$

Note that L_{λ_δ} equals $\mathcal{L}_{ref}$ when $\lambda_\delta \equiv 1$.

Applying Itô's formula to the function $h(r_t^i)$ and using (3.7) and the fact that $h'(r) > 0$, we get for any $t > 0$ and $i = 1, \cdots, N$,

$$dh(r_t^i) \leq 2\sqrt{2}\lambda_\delta(r_t^i) h'(r_t^i) d\tilde{B}_t^i + h'(r_t^i) b_0(r_t^i) dt + 4h''\lambda_\delta(r_t^i)^2 dt$$

$$+ \frac{1}{N-1} \|\nabla_{xy}^2 W\|_\infty h'(r_t^i) \sum_{j:j \neq i, 1 \leq j \leq N} r_t^j dt$$

$$= 2\sqrt{2}\lambda_\delta(r_t^i) h'(r_t^i) d\tilde{B}_t^i + L_{\lambda_\delta} h(r_t^i) dt + \frac{1}{N-1} \|\nabla_{xy}^2 W\|_\infty h'(r_t^i) \sum_{j:j \neq i, 1 \leq j \leq N} r_t^j dt.$$

$$(3.9)$$

Notice that by the definition of L_{λ_δ} and the Poisson equation (2.6),

$$L_{\lambda_\delta} h(r) = \mathcal{L}_{ref} h(r) + 4(\lambda_\delta^2 - 1) h''(r) = -r + (1 - \lambda_\delta^2)(r + b_0(r) h'(r)). \tag{3.10}$$

Then

$$-\sum_{i=1}^N \left(L_{\lambda_\delta} h(r_t^i) + \frac{1}{N-1} \|\nabla_{xy}^2 W\|_\infty h'(r_t^i) \sum_{j:j \neq i, 1 \leq j \leq N} r_t^j \right)$$

$$\geq (1 - \|\nabla_{xy}^2 W\|_\infty \|h'\|_\infty) \sum_{i=1}^N r_t^i - \sum_{i=1}^N (1 - \lambda_\delta(r_t^i)^2)(r_t^i + b_0(r_t^i) h'(r_t^i))$$

which is bounded from below by $-N(\delta + \sup_{r \in (0,\delta)} b_0^+(r) \|h'\|_\infty)$ according to the conditions **(H)** and (2.3). By integrating from 0 to T and taking expectation in the previous inequality (3.9) for $dh(r_t^i)$ and using Fatou's lemma, we have for any $T > 0$,

$$\mathbb{E} \int_0^T \left\{ (1 - \|\nabla_{xy}^2 W\|_\infty \|h'\|_\infty) \sum_{i=1}^N r_t^i - \sum_{i=1}^N (1 - \lambda_\delta(r_t^i)^2)(r_t^i + b_0(r_t^i)h'(r_t^i)) \right\} dt \tag{3.11}$$

$$\leq \sum_{i=1}^N h(|x_0^i - y_0^i|).$$

Letting $\mathbb{P}_x|_{[0,T]}$ be the law of $(X_t^{(N)})_{t \in [0,T]}$, we obtain by assumption (**H**) and (3.11)

$$W_{1,d_{L^1[0,T]}}(\mathbb{P}_{x_0}|_{[0,T]}, \mathbb{P}_{y_0}|_{[0,T]}) \leq \mathbb{E} \int_0^T d_{l^1}(X_t^{(N)}, Y_t^{(N)})dt = \mathbb{E} \int_0^T \sum_{i=1}^N r_t^i dt$$

$$\leq \frac{1}{1 - \|\nabla_{xy}^2 W\|_\infty \|h'\|_\infty} \left\{ \sum_{i=1}^N h(|x_0^i - y_0^i|) + \sum_{i=1}^N \mathbb{E} \int_0^T (1 - \lambda_\delta(r_t^i)^2)(r_t^i + b_0^+(r_t^i)h'(r_t^i))dt \right\}. \tag{3.12}$$

By the definition of λ_δ and the assumption $\lim_{r \to 0} b_0^+(r) = 0$, the second term in the right hand side of the inequality above converges to 0, a.s., as $\delta \downarrow 0$ by dominated convergence, because $b_0^+(r_t^i)h'(r_t^i) \leq \|h'\|_\infty \sup_{r>0} b_0^+(r) < +\infty$ by condition (2.2). Hence

$$W_{1,d_{L^1[0,T]}}(\mathbb{P}_{x_0}|_{[0,T]}, \mathbb{P}_{y_0}|_{[0,T]}) \leq \frac{1}{1 - \|\nabla_{xy}^2 W\|_\infty \|h'\|_\infty} \sum_{i=1}^N h(|x_0^i - y_0^i|). \tag{3.13}$$

Let Q_n be an optimization coupling of $(\mathbb{P}_{x_0}|_{[0,n]}, \mathbb{P}_{y_0}|_{[0,n]})$ for $W_{1,d_{L^1[0,n]}}(\mathbb{P}_{x_0}|_{[0,n]}, \mathbb{P}_{y_0}|_{[0,n]})$. Then $\{Q_n|_{[0,T]}; n \geq T\}$ is tight for any finite time T (because their marginal distributions are respectively $\mathbb{P}_{x_0}|_{[0,T]}$ and $\mathbb{P}_{y_0}|_{[0,T]}$), hence one can find a probability measure Q on $C(\mathbb{R}^+, (\mathbb{R}^d)^N)^2$ such that $Q_n|_{[0,T]} \to Q|_{[0,T]}$ weakly for all $T > 0$. Thus

$$W_{1,d_{L^1[0,\infty]}}(\mathbb{P}_{x_0}, \mathbb{P}_{y_0}) \leq \mathbb{E}^Q \int_0^\infty d_{l^1}(\gamma_1(t), \gamma_2(t))dt$$

$$= \lim_{T \to +\infty} \mathbb{E}^Q \int_0^T d_{l^1}(\gamma_1(t), \gamma_2(t))dt$$

$$\leq \lim_{T \to \infty} \lim_{T \leq n \to +\infty} \mathbb{E}^{Q_n} \int_0^T d_{l^1}(\gamma_1(t), \gamma_2(t))dt$$

$$\leq \lim_{n \to \infty} W_{1,d_{L^1[0,n]}}(\mathbb{P}_{x_0}|_{[0,n]}, \mathbb{P}_{y_0}|_{[0,n]}).$$

The converse inequality is evident. Therefore we have

$$W_{1,d_{L^1}}(\mathbb{P}_{x_0}, \mathbb{P}_{y_0}) = \lim_{n \to \infty} W_{1,d_{L^1[0,n]}}(\mathbb{P}_{x_0}|_{[0,n]}, \mathbb{P}_{y_0}|_{[0,n]}).$$

From this and (3.13) we obtain (2.9).

2). Proof of (2.10). Note that for any Lipschitzian function g w.r.t the d_{l^1}-metric on $(\mathbb{R}^d)^N$, g is $\mu^{(N)}$-integrable because $\int \sum_{i=1}^N |x^i|d\mu^{(N)}(x) < +\infty$. So we can assume $\mu^{(N)}(g) = 0$ without loss of generality. In that case as $\mu^{(N)}(P_t^{(N)}g) = \mu^{(N)}(g) = 0$, we

have

$$\int_0^{+\infty} |P_t^{(N)}g(x)|dt = \int_0^{+\infty} |P_t^{(N)}g(x) - \int P_t^{(N)}g(y)d\mu^{(N)}(y)|dt$$

$$\leq \|g\|_{Lip(d_{l1})} \int_{(\mathbb{R}^d)^N} \int_0^{+\infty} W_{d_{l1}}(P_t^{(N)}(x,\cdot), P_t^{(N)}(y,\cdot))dt d\mu^{(N)}(y)$$

$$\leq \frac{\|g\|_{Lip(d_{l1})}}{1 - \|\nabla_{xy}^2 W\|_\infty \|h'\|_\infty} \int_{(\mathbb{R}^d)^N} \sum_{i=1}^N h(|x^i - y^i|)d\mu^{(N)}(y)$$

$$< +\infty,$$

then the unique solution of the Poisson equation $-\mathcal{L}^{(N)}G = g$ with $\mu^{(N)}(G) = 0$ is given by $G(x) = \int_0^{+\infty} P_t^{(N)}g(x)dt, \forall x \in (\mathbb{R}^d)^N$.

For each $1 \leq i \leq N$, letting $\tilde{x}^i \neq x^i$ and $\tilde{x} \in (\mathbb{R}^d)^N$ so that $(\tilde{x})^j = x^j$ for $j \neq i$ and $(\tilde{x})^i = \tilde{x}^i$, we have

$$|\nabla_i G(x)| \leq \limsup_{\tilde{x}^i \to x^i} \frac{|G(x) - G(\tilde{x})|}{|x^i - \tilde{x}^i|}$$

$$\leq \limsup_{\tilde{x}^i \to x^i} \frac{1}{|x^i - \tilde{x}^i|} \int_0^{+\infty} |P_t^{(N)}g(x) - P_t^{(N)}g(\tilde{x})|dt$$

$$\leq \limsup_{\tilde{x}^i \to x^i} \frac{1}{|x^i - \tilde{x}^i|} \|g\|_{Lip(d_{l1})} \int_0^{+\infty} W_{d_{l1}}(P_t^{(N)}(x,\cdot), P_t^{(N)}(\tilde{x},\cdot))dt \qquad (3.14)$$

$$\leq \frac{1}{1 - \|\nabla_{xy}^2 W\|_\infty \|h'\|_\infty} \|g\|_{Lip(d_{l1})} \lim_{\tilde{x}^i \to x^i} \frac{h(|x^i - \tilde{x}^i|)}{|x^i - y^i|}$$

$$= \frac{h'(0)}{1 - \|\nabla_{xy}^2 W\|_\infty \|h'\|_\infty} \|g\|_{Lip(d_{l1})},$$

where the fourth inequality follows from (3.13). $\qquad \square$

3.3. **Proof of Theorem 2.6.**

Proof of Theorem 2.6. Here we also adopt the coupling (3.3) . Let h be defined as in (2.5). Define for any $\varepsilon > 0$,

$$h_\varepsilon(r) := h(r) + \varepsilon r, \forall r \geq 0, \qquad (3.15)$$

and

$$H_t^\varepsilon := e^{K_\varepsilon t} \sum_{i=1}^N h_\varepsilon(r_t^i),$$

where $r_t^i = |X_t^{i,N} - Y_t^{i,N}|, 1 \le i \le N$, as in the proof of Theorem 2.1. By using Ito's formula and (3.7), we get for any $t \ge 0$,

$$dH_t^\varepsilon \le 2\sqrt{2}e^{K_\varepsilon t}\sum_{i=1}^N \lambda_\delta(r_t^i)d\tilde{B}_t^i + K_\varepsilon H_t^\varepsilon dt + e^{K_\varepsilon t}\sum_{i=1}^N(L_{\lambda_\delta}h(r_t^i) + \varepsilon b_0(r_t^i))dt$$

$$+ e^{K_\varepsilon t}\sum_{i=1}^N(h'(r_t^i) + \varepsilon)\sum_{j:j\ne i}\frac{1}{N-1}\|\nabla_{xy}^2 W\|_\infty r_t^j dt \qquad (3.16)$$

$$= 2\sqrt{2}e^{K_\varepsilon t}\sum_{i=1}^N \lambda_\delta(r_t^i)d\tilde{B}_t^i + D_t^\varepsilon dt$$

where

$$D_t^\varepsilon := K_\varepsilon H_t^\varepsilon + e^{K_\varepsilon t}\sum_{i=1}^N(L_{\lambda_\delta}h(r_t^i) + \varepsilon b_0(r_t^i)) + \sum_{i\ne j, 1\le i,j\le N}(h'(r_t^i) + \varepsilon)\frac{1}{N-1}e^{K_\varepsilon t}\|\nabla_{xy}^2 W\|_\infty r_t^j.$$
$$(3.17)$$

Calculating as in the proof of Theorem 2.1, we have

$$D_t^\varepsilon \le e^{K_\varepsilon t}\sum_{i=1}^N[1 - \lambda_\delta(r_t^i)^2][r_t^i + b_0(r_t^i)h'(r_t^i)]$$

$$+ e^{K_\varepsilon t}\sum_{i=1}^N\{K_\varepsilon h_\varepsilon(r_t^i) - [1 - (\|h'\|_\infty + \varepsilon)\|\nabla_{xy}^2 W\|_\infty]r_t^i + \varepsilon b_0(r_t^i)\}$$

$$\le e^{K_\varepsilon t}\sum_{i=1}^N[1 - \lambda_\delta(r_t^i)^2][r_t^i + b_0(r_t^i)h'(r_t^i)]$$

$$+ e^{K_\varepsilon t}\sum_{i=1}^N\{K_\varepsilon(\|h'\|_\infty + \varepsilon) + \varepsilon M - [1 - (\|h'\|_\infty + \varepsilon)\|\nabla_{xy}^2 W\|_\infty]\}r_t^i,$$
$$(3.18)$$

where we use the assumption $b_0(r) \le Mr, \forall r > 0$.

By taking

$$K_\varepsilon = \frac{1 - \|\nabla_{xy}^2 W\|_\infty\|h'\|_\infty - \varepsilon(M + \|\nabla_{xy}^2 W\|_\infty)}{\|h'\|_\infty + \varepsilon}, \qquad (3.19)$$

the second term in the right hand side of the inequality above vanishes. Then by taking expectation in (3.16) and using (3.18), we get for any $t \ge 0$,

$$\mathbb{E}e^{K_\varepsilon t}\sum_{i=1}^N h_\varepsilon(r_t^i) \le \sum_{i=1}^N h_\varepsilon(|x_0^i - y_0^i|) + \mathbb{E}\int_0^t e^{K_\varepsilon s}[1 - \lambda_\delta(r_t^i)^2][r_t^i + b_0^+(r_t^i)h'(r_t^i)]ds.$$
$$(3.20)$$

Note that the second term in the right hand side of the above inequality converges to 0 as $\delta \downarrow 0$ by dominated convergence (see the reason presented just after (3.12)).

Therefore we obtain

$$W_{1,d_{l1}}(P_t^{(N)}(x_0,\cdot),P_t^{(N)}(y_0,\cdot)) \leq \lim_{\delta\to 0}\mathbb{E}\sum_{i=1}^N r_t^i$$

$$\leq \sup_{r>0}\frac{r}{h(r)+\varepsilon r}\lim_{\delta\to 0}\mathbb{E}\sum_{i=1}^N h_\varepsilon(r_t^i)$$

$$\leq \sup_{r>0}\frac{r}{h(r)+\varepsilon r}e^{-K_\varepsilon t}\sum_{i=1}^N h_\varepsilon(|x_0^i-y_0^i|)$$

$$\leq \sup_{r>0}\frac{r}{h(r)+\varepsilon r}\cdot\sup_{r>0}\frac{h(r)+\varepsilon r}{r}e^{-K_\varepsilon t}\sum_{i=1}^N |x_0^i-y_0^i|$$

where the third inequality above follows by (3.20). That is the desired result (2.25). □

4. Uniform-in-time propagation of chaos: proof of Theorem 2.11

We begin with a uniform in time control of the second moment, which is more or less known, see e.g. Cattiaux $et\ al.$ [9].

Lemma 4.1. *Suppose that there exist some positive constants c_1,c_2,c_3 such that*

$$\langle x,\nabla V(x)\rangle \geq c_1|x|^2 - c_2, \forall x\in\mathbb{R}^d \tag{4.1}$$

and

$$\langle z,\nabla_{xx}^2 W(x,y)z\rangle \geq -c_3|z|^2, \forall x,y,z\in\mathbb{R}^d. \tag{4.2}$$

Assume

$$c_1 - c_3 - \|\nabla_{xy}^2 W\|_\infty > 0. \tag{4.3}$$

Let X_t be a solution of (1.2) with $\mathbb{E}|X_0|^2 < \infty$, then for any $\varepsilon\in(0,c_1-c_3-\|\nabla_{xy}^2 W\|_\infty)$,

$$\sup_{t\geq 0}\mathbb{E}(|X_t|^2)^{\frac{1}{2}} \leq \max\{m_2(\mu_0),\hat{c}(\varepsilon)\}, \tag{4.4}$$

where $m_2(\mu_0)$ and $\hat{c}(\varepsilon)$ are given in (2.36).

Proof. By Itô's formula, we have

$$d|X_t|^2 = -2\langle X_t,\nabla V(X_t)\rangle dt - 2\langle X_t,\nabla_x W \circledast \mu_t(X_t)\rangle dt + 2d\cdot dt + 2\sqrt{2}\langle X_t,dB_t\rangle$$

Notice that for any $x\in\mathbb{R}^d$, we have

$$\langle x,\nabla_x W \circledast \mu_t(x) - \nabla_x W \circledast \mu_t(0)\rangle = \langle x,\int_0^1 \frac{d}{ds}\nabla_x W \circledast \mu_t(sx)ds\rangle$$

$$= \langle x,\int_0^1 \frac{d}{ds}\int_{\mathbb{R}^d}\nabla_x W(sx,y)\mu_t(dy)ds\rangle \tag{4.5}$$

$$= \int_0^1\int_{\mathbb{R}^d}\langle x,\nabla_{xx}^2 W(sx,y)x\rangle\mu_t(dy)ds$$

$$\geq -c_3|x|^2,$$

where the last inequality follows from (4.2).

On the other hand,

$$\begin{aligned}
|\nabla_x W \circledast \mu_t(0)\rangle| &\leq |\nabla_x W(0,0)| + \int_{\mathbb{R}^d} |\nabla_x W(0,y) - \nabla_x W(0,0)|\mu_t(dy) \\
&\leq |\nabla_x W(0,0)| + \|\nabla_{xy}^2 W\|_\infty \mathbb{E}|X_t|.
\end{aligned} \tag{4.6}$$

Therefore we have

$$\begin{aligned}
d|X_t|^2 &\leq 2\left(c_3|X_t|^2 + \|\nabla_{xy}^2 W\|_\infty |X_t|\mathbb{E}|X_t| + |\nabla_x W(0,0)||X_t|\right) dt \\
&\quad + 2(-c_1|X_t|^2 + c_2 + d)dt + 2\sqrt{2}\langle X_t, dB_t\rangle \\
&\leq -2(c_1 - c_3 - \varepsilon)|X_t|^2 dt + 2\|\nabla_{xy}^2 W\|_\infty |X_t|\mathbb{E}|X_t|dt \\
&\quad + 2(d + c_2 + \frac{1}{4\varepsilon}|\nabla_x W(0,0)|^2)dt + 2\sqrt{2}\langle X_t, dB_t\rangle
\end{aligned}$$

where $0 < \varepsilon < c_1 - c_3 - \|\nabla_{xy}^2 W\|_\infty$. By the previous stochastic differential inequality,

$$|X_t|^2 + \int_0^t [2(c_1 - c_3 - \varepsilon)|X_s|^2 - 2\|\nabla_{xy}^2 W\|_\infty |X_s|\mathbb{E}|X_s|]ds - 2t(d + c_2 + \frac{1}{4\varepsilon}|\nabla_x W(0,0)|^2)$$

is a local supermartingale, then a supermartingale by Fatou's lemma. Then for any $T > 0$, we have

$$\begin{aligned}
\mathbb{E}|X_0|^2 &\geq \mathbb{E}|X_T|^2 + 2(c_1 - c_3 - \varepsilon)\int_0^T \mathbb{E}|X_s|^2 ds - 2\|\nabla_{xy}^2 W\|_\infty \int_0^T (\mathbb{E}|X_s|)^2 ds \\
&\quad - 2T(d + c_2 + \frac{1}{4\varepsilon}|\nabla_x W(0,0)|^2) \\
&\geq 2(c_1 - c_3 - \varepsilon - \|\nabla_{xy}^2 W\|_\infty)\int_0^T \mathbb{E}|X_s|^2 ds - 2T(d + c_2 + \frac{1}{4\varepsilon}|\nabla_x W(0,0)|^2),
\end{aligned}$$

which implies $\mathbb{E}\int_0^T |X_s|^2 ds < +\infty$. In other words $\int_0^t 2\sqrt{2}\langle X_s, dB_s\rangle$ is a L^2-martingale. By taking expectation in (4.1) we obtain by (4.5) and (4.6),

$$\begin{aligned}
\frac{d}{dt}\mathbb{E}|X_t|^2 &\leq -2c_1\mathbb{E}|X_t|^2 + 2[c_3\mathbb{E}|X_t|^2 + \|\nabla_{xy}^2 W\|_\infty (\mathbb{E}|X_t|)^2 + |\nabla_x W(0,0)|\mathbb{E}|X_t|] + 2(d + c_2) \\
&\leq -2(c_1 - c_3 - \|\nabla_{xy}^2 W\|_\infty - \varepsilon)\mathbb{E}|X_t|^2 + 2(d + c_2 + \frac{1}{4\varepsilon}|\nabla_x W(0,0)|^2)
\end{aligned} \tag{4.7}$$

where $0 < \varepsilon < c_1 - c_3 - \|\nabla_{xy}^2 W\|_\infty$. By Gronwall's lemma we get for any $t \geq 0$

$$\begin{aligned}
\mathbb{E}|X_t|^2 &\leq e^{-2(c_1 - c_3 - \|\nabla_{xy}^2 W\|_\infty - \varepsilon)t}\left(\mathbb{E}|X_0|^2 - \frac{d + c_2 + \frac{1}{4\varepsilon}|\nabla_x W(0,0)|^2}{c_1 - c_3 - \|\nabla_{xy}^2 W\|_\infty - \varepsilon}\right) \\
&\quad + \frac{d + c_2 + \frac{1}{4\varepsilon}|\nabla_x W(0,0)|^2}{c_1 - c_3 - \|\nabla_{xy}^2 W\|_\infty - \varepsilon} \\
&\leq \max\left\{\mathbb{E}|X_0|^2, \frac{d + c_2 + \frac{1}{4\varepsilon}|\nabla_x W(0,0)|^2}{c_1 - c_3 - \|\nabla_{xy}^2 W\|_\infty - \varepsilon}\right\}
\end{aligned}$$

the desired result. □

Following the proof above we have the much stronger uniform Gaussian integrability for X_t, which should be of independent interest.

Lemma 4.2. *Assume (4.1), (4.2) and (4.3). Let X_t be a solution of (1.2) with*

$$\mathbb{E}\exp\left(\lambda_0|X_0|^2\right) < \infty, \quad \text{for some } \lambda_0 > 0.$$

If

$$0 < \lambda \le \min\{\lambda_0; \frac{1}{2}(c_1 - c_3 - \|\nabla_{xy}^2 W\|_\infty - \varepsilon)\}$$

for some $\varepsilon > 0$, then

$$\sup_{t\ge 0}\mathbb{E}\exp(\lambda|X_t|^2) < +\infty.$$

Proof. By Itô's formula, we have by the estimates leading to (4.7) in the proof of Lemma 4.1,

$$d\exp(\lambda|X_t|^2)$$
$$=\lambda\exp(\lambda|X_t|^2)\left([2d - 2\langle X_t, \nabla V(X_t) + \nabla_x W \circledast \mu_t(X_t)\rangle]dt + 2\sqrt{2}\langle X_t, dB_t\rangle\right)$$
$$\quad + 4\lambda^2|X_t|^2\exp(\lambda|X_t|^2)dt$$
$$\le\lambda\exp(\lambda|X_t|^2)\left[-2(c_1 - c_3 - \|\nabla_{xy}^2 W\|_\infty - \varepsilon - 2\lambda)|X_t|^2 + 2(d + c_2 + \frac{1}{4\varepsilon}|\nabla_x W(0,0)|)\right]dt$$
$$\quad + \lambda\exp(\lambda|X_t|^2)2\sqrt{2}\langle X_t, dB_t\rangle$$

where $\varepsilon > 0, \lambda > 0$ verify $c_1 - c_3 - \|\nabla_{xy}^2 W\|_\infty - \varepsilon - 2\lambda > 0$. Taking $L > 0$ large sufficient so that

$$c_5 := 2(c_1 - c_3 - \|\nabla_{xy}^2 W\|_\infty - \varepsilon - 2\lambda)L^2 - 2(d + c_2 + \frac{1}{4\varepsilon}|\nabla_x W(0,0)|) > 0,$$

and noting that

$$-ax^2 + b \le -(aL^2 - b) + aL^2 1_{|x|\le L}, \ \forall a > 0, \forall x \in \mathbb{R},$$

we obtain by following the same argument as in Lemma 4.1

$$\frac{d}{dt}\mathbb{E}\exp(\lambda|X_t|^2) \le -\lambda c_5\mathbb{E}\exp(\lambda|X_t|^2) + 2(c_1 - c_3 - \|\nabla_{xy}^2 W\|_\infty - \varepsilon - 2\lambda)L^2\lambda e^{\lambda L^2}.$$

Therefore by Gronwall's lemma

$$\sup_{t\ge 0}\mathbb{E}\exp(\lambda|X_t|^2) < +\infty.$$

$\square$

Next we present the proof of Theorem 2.11, which is quite close to those of Theorems 2.1 and 2.6 .

Proof of Theorem 2.11. Let λ_δ and π_δ be defined as in Section 3.1. Consider the following coupling between the independent copies $\bar{X}_t^i, 1 \le i \le N$ of the nonlinear diffusion process (1.2) and the mean-field interacting particle system (1.3):

$$
\begin{aligned}
d\bar{X}_t^i &= \sqrt{2}[\lambda_\delta(|Z_t^i|)dB_t^{1,i} + \pi_\delta(|Z_t^i|)dB_t^{2,i}] - \nabla V(\bar{X}_t^i)dt - \nabla_x W \circledast \mu_t(\bar{X}_t^i)dt, \\
dX_t^{i,N} &= \sqrt{2}[\lambda_\delta(|Z_t^i|)R_t^i dB_t^{1,i} + \pi_\delta(|Z_t^i|)dB_t^{2,i}] - \nabla V(X_t^{i,N})dt \\
&\qquad\qquad\qquad\qquad\qquad - \frac{1}{N-1}\sum_{j:j\neq i,1\le j\le N}\nabla_x W(X_t^{i,N}, X_t^{j,N})dt.
\end{aligned}
\tag{4.8}
$$

Here $Z_t^i := \bar{X}_t^i - X_t^{i,N}$ and $R_t^i := I_d - 2e_t^i e_t^{i,T}$, where I_d is the d-dimensional unit matrix and $e_t^i e_t^{i,T}$ is the orthogonal projection onto the unit vector $e_t^i := Z_t^i/|Z_t^i|$ if $|Z_t^i| \neq 0$. $B_t^{1,i}$ and $B_t^{2,i}, 1 \le i \le N$, are independent standard Brownian motions in $\mathbb{R}^d$. We assume that $\bar{X}_t^i$ and $X_t^{i,N}$, $1 \le i \le N$ have the same starting points $X_0^i, 1 \le i \le N$, i.i.d. of law μ_0. The independence of $\bar{X}_t^i, 1 \le i \le N$ comes from the fact that the Brownian motions $\{\int_0^t \lambda_\delta(|Z_s^i|)dB_s^{1,i} + \int_0^t \pi_\delta(|Z_s^i|)dB_s^{2,i}, \ 1 \le i \le N\}$ are independent because their inter-brackets are zero.

By doing subtraction of the equations in (4.8), we have

$$
\begin{aligned}
dZ_t^i = 2\sqrt{2}\lambda_\delta(|Z_t^i|)e_t^i d\tilde{B}_t^i &- [\nabla V(\bar{X}_t^i) - \nabla V(X_t^{i,N})]dt - \nabla_x W \circledast \mu_t(\bar{X}_t^i)dt \\
&+ \frac{1}{N-1}\sum_{j:j\neq i,1\le j\le N}\nabla_x W(X_t^{i,N}, X_t^{j,N})dt,
\end{aligned}
$$

where the processes $\tilde{B}_t^i = \int_0^t (e_s^i)^T dB_s^{1,i}, 1 \le i \le N$, are one-dimensional standard Brownian motions such that $\langle \tilde{B}^i, \tilde{B}^j \rangle_t = 0$ for $i \neq j$.

Let $r_t^i = |Z_t^i|, 1 \le i \le N$. By applying Itô's formula, we have

$$
\begin{aligned}
dr_t^i = {}& 1_{\{r_t^i\neq 0\}}2\sqrt{2}\lambda_\delta(r_t^i)d\tilde{B}_t^i - 1_{\{r_t^i\neq 0\}}\langle e_t^i, \nabla V(\bar{X}_t^i) - \nabla V(X_t^{i,N})\rangle dt \\
& - 1_{\{r_t^i\neq 0\}}\langle e_t^i, \nabla_x W \circledast \mu_t(\bar{X}_t^i) - \frac{1}{N-1}\sum_{j:j\neq i,1\le j\le N}\nabla_x W(X_t^{i,N}, X_t^{j,N})\rangle dt \\
= {}& 1_{\{r_t^i\neq 0\}}2\sqrt{2}\lambda_\delta(r_t^i)d\tilde{B}_t^i \\
& - 1_{\{r_t^i\neq 0\}}\langle e_t^i, \nabla V(\bar{X}_t^i) - \nabla V(X_t^{i,N})\rangle dt \\
& - 1_{\{r_t^i\neq 0\}}\langle e_t^i, \nabla_x W \circledast \mu_t(\bar{X}_t^i) - \frac{1}{N-1}\sum_{j:j\neq i,1\le j\le N}\nabla_x W(\bar{X}_t^i, \bar{X}_t^j)\rangle dt \\
& - 1_{\{r_t^i\neq 0\}}\langle e_t^i, \frac{1}{N-1}\sum_{j:j\neq i,1\le j\le N}[\nabla_x W(\bar{X}_t^i, \bar{X}_t^j) - \nabla_x W(\bar{X}_t^i, X_t^{j,N})]\rangle dt \\
& - 1_{\{r_t^i\neq 0\}}\langle e_t^i, \frac{1}{N-1}\sum_{j:j\neq i,1\le j\le N}[\nabla_x W(\bar{X}_t^i, X_t^{j,N}) - \nabla_x W(X_t^{i,N}, X_t^{j,N})]\rangle dt.
\end{aligned}
$$

Remark that the sum of the first and the fourth drift terms above is $\le b_0(r_t^i)dt$, the third drift term above is $\le \frac{1}{N-1}\|\nabla_{xy}^2 W\|_\infty \sum_{j:j\neq i,1\le j\le N} r_t^j dt$, and the second drift term

is bounded by $I_t^i dt$, where

$$I_t^i := |\nabla_x W \circledast \mu_t(\bar{X}_t^i) - \frac{1}{N-1} \sum_{j:j\neq i, 1\leq j\leq N} \nabla_x W(\bar{X}_t^i, \bar{X}_t^j)|. \tag{4.9}$$

Therefore we obtain

$$dr_t^i \leq 2\sqrt{2}\lambda_\delta(r_t^i)d\tilde{B}_t^i + b_0(r_t^i)dt + \frac{1}{N-1}\|\nabla_{xy}^2 W\|_\infty \sum_{j:j\neq i, 1\leq j\leq N} r_t^j dt + I_t^i dt. \tag{4.10}$$

Recall that for any $\varepsilon \geq 0$, $h_\varepsilon(r) = h(r) + \varepsilon r, \forall r \geq 0$. By using (4.10) and Itô's formula again, we get

$$
\begin{aligned}
dh_\varepsilon(r_t^i) &\leq 2\sqrt{2}\lambda_\delta(r_t^i)h_\varepsilon'(r_t^i)d\tilde{B}_t^i + 4\lambda_\delta^2(r_t^i)h_\varepsilon''(r_t^i)dt + b_0(r_t^i)h_\varepsilon'(r_t^i)dt \\
&\quad + \frac{1}{N-1}\|\nabla_{xy}^2 W\|_\infty h_\varepsilon'(r_t^i) \sum_{j:j\neq i, 1\leq j\leq N} r_t^j dt + h_\varepsilon'(r_t^i)I_t^i dt \\
&= 2\sqrt{2}\lambda_\delta(r_t^i)h_\varepsilon'(r_t^i)d\tilde{B}_t^i + [4\lambda_\delta^2(r_t^i)h''(r_t^i) + b_0(r_t^i)h'(r_t^i)]dt + \varepsilon b_0(r_t^i)dt \\
&\quad + \frac{1}{N-1}\|\nabla_{xy}^2 W\|_\infty(h'(r_t^i) + \varepsilon) \sum_{j:j\neq i, 1\leq j\leq N} r_t^j dt + (h'(r_t^i) + \varepsilon)I_t^i dt \\
&\leq 2\sqrt{2}\lambda_\delta(r_t^i)h_\varepsilon'(r_t^i)d\tilde{B}_t^i + [1 - \lambda_\delta^2(r_t^i)][r_t^i + b_0(r_t^i)h'(r_t^i)]dt - (1 - \varepsilon M)r_t^i dt \\
&\quad + \frac{1}{N-1}\|\nabla_{xy}^2 W\|_\infty(\|h'\|_\infty + \varepsilon) \sum_{j:j\neq i, 1\leq j\leq N} r_t^j dt + (\|h'\|_\infty + \varepsilon)I_t^i dt,
\end{aligned}
\tag{4.11}
$$

where the last inequality follows from (3.8), (3.10) and (2.23).

Taking expectation in the inequality above and using the fact that $r_t^i, 1 \leq i \leq N$ have the same law, and setting

$$c_\varepsilon := 1 - \|\nabla_{xy}^2 W\|_\infty\|h'\|_\infty - \varepsilon(M + \|\nabla_{xy}^2 W\|_\infty),$$

we have

$$d\mathbb{E}h_\varepsilon(r_t^1) \leq \mathbb{E}[1 - \lambda_\delta(r_t^1)^2][r_t^1 + b_0^+(r_t^1)h'(r_t^1)]dt + (\|h'\|_\infty + \varepsilon)\mathbb{E}I_t^1 dt - c_\varepsilon\mathbb{E}r_t^1 dt \tag{4.12}$$

Proof of part (a). Choose $\varepsilon = 0$, $c_0 = 1 - \|\nabla_{xy}^2 W\|_\infty\|h'\|_\infty$. For any $1 \leq k \leq N$, by (4.12) we have

$$
\begin{aligned}
\frac{1}{k}W_{1,d_{L^1[0,T]}}(\mathbb{P}_{\mu_0^{\otimes N}}^{[1,k],N}|_{[0,T]}, \mathbb{Q}_{\mu_0}^{\otimes k}|_{[0,T]}) &\leq \frac{1}{k}\mathbb{E}\int_0^T \sum_{i=1}^k r_t^i dt = \int_0^T \mathbb{E}r_t^1 dt \\
&\leq \frac{1}{c_0}\|h'\|_\infty\int_0^T \mathbb{E}I_t^1 dt dt + \frac{1}{c_0}\mathbb{E}\int_0^T [1 - \lambda_\delta(r_t^1)^2][r_t^1 + b_0^+(r_t^1)h'(r_t^1)]dt.
\end{aligned}
$$

Letting $\delta \to 0+$, the last term tends to zero. Hence

$$\frac{1}{k}W_{1,d_{L^1[0,T]}}(\mathbb{P}_{\mu_0^{\otimes N}}^{[1,k],N}|_{[0,T]}, \mathbb{Q}_{\mu_0}^{\otimes k}|_{[0,T]}) \leq \frac{\|h'\|_\infty}{c_0}\int_0^T \mathbb{E}I_t^1 dt. \tag{4.13}$$

Next we estimate $\mathbb{E}I_t^1$, which is the only new point w.r.t. the proofs in Theorems 2.1 and 2.6. Note that $\bar{X}_t^j, 2 \leq j \leq N$ are independent copies of $\bar{X}_t^1$, and

$$\mathbb{E}[\nabla_x W(\bar{X}_t^1, \bar{X}_t^j)|\bar{X}_t^1] = \nabla_x W \circledast \mu_t(\bar{X}_t^1).$$

Thus by using Cauchy-Schwartz inequality, we get

$$
\begin{aligned}
\mathbb{E}I_t^1 &\leq \left(\mathbb{E} \left\{ \mathbb{E} \left[|\nabla_x W \circledast \mu_t(\bar{X}_t^1) - \frac{1}{N-1} \sum_{2 \leq j \leq N} \nabla_x W(\bar{X}_t^1, \bar{X}_t^j)|^2 |\bar{X}_t^1] \right] \right\} \right)^{\frac{1}{2}} \\
&= \left(\mathbb{E} \frac{1}{N-1} \int |\nabla_x W(\bar{X}_t^1, y) - \nabla_x W * \mu_t(\bar{X}_t^1)|^2 d\mu_t(y) \right)^{\frac{1}{2}} \\
&\leq \frac{1}{\sqrt{N-1}} \|\nabla_{xy}^2 W\|_\infty \left(\int_{x \in \mathbb{R}^d} |x - \mu_t(\bar{X}_t^1)|^2 \mu_t(dx) \right)^{\frac{1}{2}} \\
&\leq \frac{1}{\sqrt{N-1}} \|\nabla_{xy}^2 W\|_\infty \sup_{t \geq 0} (\mathbb{E}|X_t|^2)^{\frac{1}{2}}.
\end{aligned}
\tag{4.14}
$$

Plugging (4.14) into (4.13), we get

$$\frac{1}{k} W_{1,d_{L^1[0,T]}}(\mathbb{P}_{\mu_0^{\otimes N}}^{[1,k],N}|_{[0,T]}, \mathbb{Q}_{\mu_0}^{\otimes k}|_{[0,T]}) \leq \frac{T}{\sqrt{N-1}} \frac{\|\nabla_{xy}^2 W\|_\infty \|h'\|_\infty}{1 - \|\nabla_{xy}^2 W\|_\infty \|h'\|_\infty} \sup_{t \geq 0} (\mathbb{E}|X_t|^2)^{\frac{1}{2}}.$$

Then by using Lemma 4.1, we obtain the desired result (2.35).

Proof of part (b). For any $\varepsilon > 0$, by (4.12) we have

$$d\mathbb{E}h_\varepsilon(r_t^1) \leq \mathbb{E}[1 - \lambda_\delta(r_t^1)^2][r_t^1 + b_0^+(r_t^1)h'(r_t^1)]dt + (\|h'\|_\infty + \varepsilon)\mathbb{E}I_t^1 dt - c_\varepsilon \cdot \inf_{r>0} \frac{r}{h(r) + \varepsilon r} \mathbb{E}h_\varepsilon(r_t^1)dt \tag{4.15}$$

Plugging (4.14) into (4.15), we obtain by Gronwall's inequality that for any $\varepsilon > 0$ so that $\beta = c_\varepsilon \cdot \inf_{r>0} \frac{r}{h(r) + \varepsilon r} > 0$ (i.e. $K_\varepsilon > 0$),

$$
\begin{aligned}
\mathbb{E}h_\varepsilon(|\bar{X}_t^1 - X_t^{1,N}|) &\leq \int_0^t e^{-\beta(t-s)} \frac{1}{\sqrt{N-1}} (\|h'\|_\infty + \varepsilon) \mathbb{E}I_t^1 ds \\
&\quad + \int_0^t e^{-\beta(t-s)} \mathbb{E}[1 - \lambda_\delta(r_s^1)^2][r_s^1 + b_0^+(r_s^1)h'(r_s^1)]ds. \\
&\leq \int_0^t e^{-\beta(t-s)} \frac{1}{\sqrt{N-1}} (\|h'\|_\infty + \varepsilon) \|\nabla_{xy}^2 W\|_\infty \sup_{t \geq 0} (\mathbb{E}|X_t|^2)^{\frac{1}{2}} ds \\
&\quad + \int_0^t e^{-\beta(t-s)} \mathbb{E}[1 - \lambda_\delta(r_s^1)^2][r_s^1 + b_0^+(r_s^1)h'(r_s^1)]ds.
\end{aligned}
\tag{4.16}
$$

By letting $\delta \to 0+$, the last term tends to zero. We obtain thus

$$
\begin{aligned}
\mathbb{E}|\bar{X}_t^1 - X_t^{1,N}|) &\leq \sup_{r>0} \frac{r}{h_\varepsilon(r)} \cdot \mathbb{E}h_\varepsilon(|\bar{X}_t^1 - X_t^{1,N}|) \\
&\leq \sup_{r>0} \frac{r}{h_\varepsilon(r)} \cdot \frac{1}{\beta\sqrt{N-1}} (\|h'\|_\infty + \varepsilon) \|\nabla_{xy}^2 W\|_\infty \sup_{t \geq 0} (\mathbb{E}|X_t|^2)^{\frac{1}{2}}.
\end{aligned}
$$

As the joint law of $(\bar{X}_t^i, 1 \le i \le k)$ is $\mu_t^{\otimes k}$, we get for any $1 \le k \le N$,

$$W_{1,d_{l_1}}(\mu_t^{\otimes k}, \mu_t^{[1,k],N}) \le \limsup_{\delta \to 0} \mathbb{E} \sum_{i=1}^k |\bar{X}_t^i - X_t^{i,N}| = k \cdot \limsup_{\delta \to 0} \mathbb{E}|\bar{X}_t^1 - X_t^{1,N}|$$

$$\le k \cdot \sup_{r>0} \frac{r}{h(r) + \varepsilon r} \frac{1}{\beta \sqrt{N-1}} (\|h'\|_\infty + \varepsilon) \|\nabla_{xy}^2 W\|_\infty \sup_{t \ge 0} (\mathbb{E}|X_t|^2)^{\frac{1}{2}}$$

$$= \frac{k}{\sqrt{N-1}} \frac{A_\varepsilon}{K_\varepsilon} \|\nabla_{xy}^2 W\|_\infty \sup_{t \ge 0} (\mathbb{E}|X_t|^2)^{\frac{1}{2}},$$

(4.17)

which completes the proof by using Lemma 4.1. $\qquad \square$

The proof above yields

Proposition 4.3. *Under the conditions of Theorem 2.11, we have*

$$\mathbb{E}W_1 \left(\frac{1}{N} \sum_{i=1}^N \delta_{X_t^{i,N}}, \frac{1}{N} \sum_{i=1}^N \delta_{\bar{X}_t^i} \right) \le \frac{1}{\sqrt{N-1}} \frac{A_\varepsilon}{K_\varepsilon} \|\nabla_{xy}^2 W\|_\infty \sup_{t \ge 0} (\mathbb{E}|X_t|^2)^{\frac{1}{2}} \qquad (4.18)$$

where $(\bar{X}_t^i)_{t \ge 0}, i \ge 1$ are independent copies of the solution $(X_t)_{t \ge 0}$ of the McKean-Vlasov equation (1.2), and $X_t^{i,N}, 1 \le i \le N$ are defined as in (1.3).

Proof. Notice that

$$\mathbb{E}W_1 \left(\frac{1}{N} \sum_{i=1}^N \delta_{X_t^{i,N}}, \frac{1}{N} \sum_{i=1}^N \delta_{\bar{X}_t^i} \right) \le \mathbb{E}[\frac{1}{N} \sum_{i=1}^N |X_t^{i,N} - \bar{X}_t^i|] = \mathbb{E}\frac{1}{N} \sum_{i=1}^N r_t^i,$$

where $r_t^i, 1 \le i \le N$ are the same as defined in the proof of Theorem 2.11. And by (4.17), we have

$$\limsup_{\delta \to 0} \mathbb{E}\frac{1}{N} \sum_{i=1}^n r_t^i = \limsup_{\delta \to 0} \mathbb{E}r_t^1 \le \frac{1}{\sqrt{N-1}} \frac{A_\varepsilon}{K_\varepsilon} \|\nabla_{xy}^2 W\|_\infty \sup_{t \ge 0} (\mathbb{E}|X_t|^2)^{\frac{1}{2}}.$$

Therefore we obtain (4.18). $\qquad \square$

Remark 4.4. A consequence of Proposition 4.3 is on the bias of $\frac{1}{N} \sum_{i=1}^N f(X_t^{i,N})$ from $\mu_t(f)$: if f is Lipschitzian on $\mathbb{R}^d$,

$$bias_t(f) := |\mathbb{E}\frac{1}{N} \sum_{i=1}^N f(X_t^{i,N}) - \mu_t(f)| = |\mathbb{E}\frac{1}{N} \sum_{i=1}^N f(X_t^{i,N}) - \mathbb{E}\frac{1}{N} \sum_{i=1}^N f(\bar{X}_t^i)|$$

$$\le \|f\|_{Lip} \mathbb{E}W_1 \left(\frac{1}{N} \sum_{i=1}^N \delta_{X_t^{i,N}}, \frac{1}{N} \sum_{i=1}^N \delta_{\bar{X}_t^i} \right)$$

$$\le \frac{\|f\|_{Lip}}{\sqrt{N-1}} \frac{A_\varepsilon}{K_\varepsilon} \|\nabla_{xy}^2 W\|_\infty \sup_{t \ge 0} (\mathbb{E}|X_t|^2)^{\frac{1}{2}}.$$

It is expected that the bias is of order $O(1/N)$, which remains an open question.

5. Quantitative Concentration inequalities: proofs of Theorems 2.5 and 2.10

This section is devoted to the concentration inequalities of the mean-field interaction particle system (1.3) for more general observable (than those in Theorems 2.5 and 2.10). This kind of concentration estimate are useful to numerical simulations and mean-field limit.

5.1. **Concentration for time average.**

Proposition 5.1. *Assume* **(H)**, *(2.2) and (2.3). Given any* $T \in (0, +\infty]$, *let* F *be any* $d_{L^1[0,T]}$-*Lipschitzian continuous function on* $C([0,T], (\mathbb{R}^d)^N)$, *given by*

$$F(X_{[0,T]}^{(N)}) := G\left(\int_0^T g_1(X_t^{(N)})dt, \cdots, \int_0^T g_n(X_t^{(N)})dt\right),$$

where $G \in C^2(\mathbb{R}^n)$, $g_i \in C^2((\mathbb{R}^d)^N, \mathbb{R})$, $1 \leq i \leq n$. *Then for any convex function* φ *on* $\mathbb{R}$ *and any starting point* $X_0^{(N)} = x \in (\mathbb{R}^d)^N$, *we have*

$$\mathbb{E}_x \varphi\left(F(X_{[0,T]}^{(N)}) - \mathbb{E}_x F(X_{[0,T]}^{(N)})\right) \leq \mathbb{E}\varphi\left(\sqrt{NT}\|F\|_{Lip(d_{L^1[0,T]})} c_{Lip}\xi\right), \qquad (5.1)$$

where ξ *is some standard real Gaussian random variable of law* $\mathcal{N}(0,1)$, *and*

$$c_{Lip} = \frac{h'(0)}{1 - \|\nabla_{xy}^2 W\|_\infty \|h'\|_\infty}.$$

Proof. Let $\{\mathcal{F}_t\}_{t\geq 0}$ be the filtration generated by the process $(X_t^{(N)})_{t\geq 0}$ and

$$M_t = \mathbb{E}(F(X_{[0,T]}^{(N)})|\mathcal{F}_t), \ 0 \leq t \leq T.$$

Then by the martingale representation theorem, we have

$$F(X_{[0,T]}^{(N)}) - \mathbb{E}F(X_{[0,T]}^{(N)}) = M_T - M_0 = \sum_{i=1}^N \int_0^T \beta_t^i dB_t^i, \qquad (5.2)$$

where $\beta_t^i, 1 \leq i \leq N$ are adapted processes w.r.t. $\mathcal{F}_t$, and $B_t^i, 1 \leq i \leq N$ are N independent standard Brownian motions on $\mathbb{R}^d$.

Let $A_t^k = \int_0^t g_k(X_s^{(N)})ds, \ 1 \leq k \leq n$, and $A_t = (A_t^1, \cdots, A_t^n)$. Note that

$$M_t = \phi(A_t, X_t^{(N)})$$

where

$$\phi(a, x) := \mathbb{E}\left(G\left(a_1 + \int_t^T g_1(X_s^{(N)})ds, \cdots, a_n + \int_t^T g_N(X_s^{(N)})ds\right)|X_t^{(N)} = x\right),$$

for $a = (a_1, \cdots, a_n) \in \mathbb{R}^n, x \in (\mathbb{R}^d)^N$. Since φ is C^2 (for V, W are C^2), we can apply Itô's formula to obtain that

$$\beta_t^i = \partial_{x_i}\varphi(A_t, X_t^{(N)}).$$

For any $x = (x^1, \cdots, x^i, \cdots, x^N) \in (\mathbb{R}^d)^N$, denote $y = (x^1, \cdots, y^i, \cdots, x^N)$ which only differs from x at the i-th coordinate. Let $(X_t^{(N)})_{t\geq 0}, (Y_t^{(N)})_{t\geq 0}$ be an optimal coupling of $\mathbb{P}_x, \mathbb{P}_y$ for $W_{1,d_{L^1[0,T]}}(\mathbb{P}_x, \mathbb{P}_y)$ (this optimal coupling exists because $d_{L^1[0,T]}$ is

lower semi-continuous from $(C(\mathbb{R}^+, (\mathbb{R}^d)^N))^2$ to $[0, +\infty])$. Then for any $0 \le t \le T$ and $i = 1, \cdots, N$, we have

$$
\begin{aligned}
|\partial_{x_i}\phi(a, x)| &\le \limsup_{y^i \to x^i} \frac{|\phi(a, x) - \phi(a, y)|}{|x^i - y^i|}| \\
&= \limsup_{y^i \to x^i} \frac{1}{|x^i - y^i|} |\mathbb{E}[G(a + \int_0^{T-t} g(X_s^{(N)}ds)] - \mathbb{E}[G(a + \int_0^{T-t} g(Y_s^{(N)}ds)]| \\
&\le \limsup_{y^i \to x^i} \frac{\|F\|_{Lip(d_{L^1[0,T]})}}{|x^i - y^i|} \mathbb{E}\int_0^\infty d_{l^1}(X_s^{(N)}, Y_s^{(N)})ds \\
&= \|F\|_{Lip(d_{L^1[0,T]})} \limsup_{y^i \to x^i} \frac{W_{1,d_{L^1}}(\mathbb{P}_x, \mathbb{P}_y)}{|x^i - y^i|} \\
&\le \|F\|_{Lip(d_{L^1[0,T]})} \cdot c_{Lip}
\end{aligned}
\tag{5.3}
$$

where the last inequality follows by Theorem 2.1.

Since $\sum_{i=1}^N \int_0^T \beta_t^i dB_t^i = \xi_{\tau_T}$ where (ξ_t) is a real valued Brownian motion w.r.t. some new filtration $(\tilde{\mathcal{F}}_t)$ and $\tau_T = \int_0^T \sum_{i=1}^N |\beta_t^i|^2 dt \le \|F\|_{Lip(d_{L^1[0,T]})}^2 c_{Lip}^2 NT =: CNT$ is a stopping time w.r.t. $(\tilde{\mathcal{F}}_t)$, we obtain

$$
\begin{aligned}
\mathbb{E}_x\varphi\left(F(X_{[0,T]}^{(N)}) - \mathbb{E}F(X_{[0,T]}^{(N)})\right) = \mathbb{E}\varphi\left(\sum_{i=1}^N \int_0^T \beta_t^i dB_t^i\right) &= \mathbb{E}\varphi(\xi_{\tau_T}) \\
&= \mathbb{E}\varphi\left(\mathbb{E}(\xi_{CNT}|\tilde{\mathcal{F}}_{\tau_T})\right) \\
&\le \mathbb{E}\varphi\left(\xi_{CNT}\right) \qquad \text{(by Jensen's inequality)} \\
&= \mathbb{E}\varphi\left(\sqrt{NT}\|F\|_{Lip(d_{L^1[0,T]})} c_{Lip}\xi_1\right)
\end{aligned}
\tag{5.4}
$$

the desired result. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Next we give the proof of Theorem 2.5.

Proof of Theorem 2.5. For any given $\lambda, T > 0$, let

$$
F(X_{[0,T]}^{(N)}) = \frac{1}{T}\int_0^T U_N(f_m)(X_t^{(N)})dt.
$$

Since f_m is 1-Lipschitzian w.r.t. the d_{l^1}-metric on $(\mathbb{R}^d)^m$, by an easy calculation we have

$$
\|F\|_{Lip_{(d_{L^1[0,T]})}} \le \frac{m}{NT}.
$$

Let $g(x) = \mathbb{E}_x F$, $\forall x \in (\mathbb{R}^d)^N$. For any fixed initial value $x \in (\mathbb{R}^d)^N$, by applying Proposition 5.1 with $\varphi(z) = e^{\lambda z}$, we get

$$
\begin{aligned}
&\mathbb{E}_x \exp\left(\lambda\left[\frac{1}{T}\int_0^T U_N(f_m)(X_t^{(N)})dt - g(x)\right]\right) \\
&\le \mathbb{E}\exp\left(\frac{m\lambda}{\sqrt{NT}}c_{Lip}\xi_1\right) = \exp\left(\frac{m^2\lambda^2 c_{Lip}^2}{2NT}\right).
\end{aligned}
\tag{5.5}
$$

By the proof of Proposition 5.1,

$$\|g\|_{Lip(d_{l^1})} \le c_{Lip}\|F\|_{d_{L^1[0,T]}} \le \frac{mc_{Lip}}{NT}.$$

By the condition (2.19) and its consequence (2.20), the product measure $\mu_0^{\otimes N}$ satisfies

$$\int_{(\mathbb{R}^d)^N} e^{\lambda(g-\mu_0^{\otimes N}(g))} d\mu_0^{\otimes N} \le \exp\left(\frac{1}{2}Nc_G(\mu_0)\lambda^2\|g\|_{Lip(d_{l^1})}^2\right)$$

$$\le \exp\left(\frac{1}{2NT^2}c_G(\mu_0)\lambda^2 m^2 c_{Lip}^2\right). \tag{5.6}$$

Hence for the i.i.d. initial values $X_0^{1,N}, \cdots, X_0^{N,N}$ with the common law μ_0, noting that

$$\mathbb{E}\frac{1}{T}\int_0^T U_N(f_m)(X_t^{(N)})dt = \mu_0^{\otimes N}(g)$$

we have

$$\mathbb{E}\exp\left(\lambda\left[\frac{1}{T}\int_0^T U_N(f_m)(X_t^{(N)})dt - \mathbb{E}\frac{1}{T}\int_0^T U_N(f_m)(X_t^{(N)})dt\right]\right)$$

$$= \int_{(\mathbb{R}^d)^N} \mathbb{E}_x\left[\exp\left(\lambda\left[\frac{1}{T}\int_0^T U_N(f_m)(X_t^{(N)})dt - g(x)\right]\right)\right] e^{\lambda(g(x)-\mu_0^{\otimes N}(g))} d\mu_0^{\otimes N}(x)$$

$$\le \exp\left(\frac{m^2\lambda^2 c_{Lip}^2}{2NT}\right)\int_{(\mathbb{R}^d)^N} e^{\lambda(g(x)-\mu_0^{\otimes N}(g))} d\mu_0^{\otimes N}(x) \tag{5.7}$$

$$\le \exp\left(\frac{m^2\lambda^2 c_{Lip}^2}{2NT}\left(1 + \frac{c_G(\mu_0)}{T}\right)\right)$$

where the second inequality follows from (5.5), and the last inequality is a consequence of (5.6). This gives us (2.21). Finally (2.22) follows from (2.21), by the standard procedure of Chebyshev's inequality and optimization over $\lambda > 0$. $\qquad\square$

Remark 5.2. The time-particle average $\frac{1}{NT}\int_0^T f(X_t^{i,N})dt$ is used to approximate $\mu_\infty(f)$ where μ_∞ is the unique equilibrium state of the McKean-Vlasov equation (proved in [18] under **(H)**). For applying Theorem 2.5, it remains to bound the bias

$$|\mathbb{E}_{\mu_0^{\otimes N}}\frac{1}{NT}\int_0^T\sum_{i=1}^N f(X_t^{i,N})dt - \mu_\infty(f)|$$

$$\le |\mathbb{E}_{\mu_0^{\otimes N}}\frac{1}{T}\int_0^T [f(X_t^{1,N}) - \mu_t(f)]dt| + \frac{1}{T}\int_0^T |\mu_t(f) - \mu_\infty(f)|dt$$

$$\le \frac{1}{T}\int_0^T |\mu_t^{1,N}(f) - \mu_t(f)|dt + \|f\|_{Lip}\frac{1}{T}\int_0^T W_1(\mu_t, \mu_\infty)dt$$

$$\le \|f\|_{Lip}\left(\sup_{t\ge 0} W_1(\mu_t^{1,N}, \mu_t) + \frac{1}{T}\int_0^T W_1(\mu_t, \mu_\infty)dt\right)$$

$$\le \|f\|_{Lip}\left(\frac{A}{\sqrt{N}} + \frac{B}{T}\right)$$

where in the last inequality, the first term comes from the uniform in time propagation of chaos (2.37) in Theorem 2.11, and the second follows by (2.12) in Corollary 2.2. We believe that the bias should be of order $O(1/N + 1/T)$, but we do not know how to prove it.

5.2. **Uniform in time concentration inequality.**

Proposition 5.3. *Assume* **(H)***, (2.2) and (2.23). Let* $X_t^{(N)} = (X_t^{1,N}, \cdots, X_t^{N,N}), \forall t \geq 0$, *then for any Lipschitzian function F on $(\mathbb{R}^d)^N$, we have for any lower bounded convex function φ on $\mathbb{R}$,*

$$\mathbb{E}_x \varphi \left(F(X_T^{(N)}) - \mathbb{E}_x F(X_T^{(N)}) \right) \leq \mathbb{E} \varphi \left(\alpha A_\varepsilon \sqrt{\frac{N}{2K_\varepsilon}} \xi \right), \ \forall x \in (\mathbb{R}^d)^N, \ \forall T > 0 \quad (5.8)$$

where ξ is some standard real Gaussian random variable of law $\mathcal{N}(0,1)$, $\alpha := \|F\|_{Lip(d_{l^1})} = \max_{1 \leq i \leq N} \|\nabla_i F\|_\infty$, A_ε and K_ε are given in Theorem 2.6.

In particular for any initial distribution μ_0 satisfying the Gaussian integrability assumption on $\mathbb{R}^d$, we have for any $\delta, T > 0$

$$\mathbb{P}_{\mu_0^{\otimes N}} \left\{ F(X_T^{(N)}) - \mathbb{E}_{\mu_0^{\otimes N}} F(X_T^{(N)}) > \delta \right\} \leq \exp \left(-\frac{K_\varepsilon \delta^2}{N\alpha^2 A_\varepsilon^2 \left[1 + 2c_G(\mu_0) K_\varepsilon e^{-2K_\varepsilon T} \right]} \right). \quad (5.9)$$

Proof. Without loss of generality we may assume that $\alpha = \max_{1 \leq i \leq N} \|\nabla_i F\|_\infty = 1$.

By approximation we may assume that F is C^2-smooth with bounded derivatives of the first and the second order. For any initial position $x \in (\mathbb{R}^d)^N$, let $M_t = \mathbb{E}_x(F(X_T^{(N)})|\mathcal{F}_t)$, $0 \leq t \leq T$. Then by applying Itô's formula to $u(t,x) = P_{T-t}F(x)$, we have

$$F(X_T^{(N)}) - \mathbb{E}_x F(X_T^{(N)}) = M_T - M_0 = \sum_{i=1}^N \int_0^T \nabla_i P_{T-t} F(X_t^{(N)}) dB_t^i, \quad (5.10)$$

Note that by Theorem 2.6, for any $\varepsilon > 0$ such that $K_\varepsilon > 0$, we have

$$W_{d_{l^1}}(P_t^{(N)}(x, \cdot), P_t^{(N)}(y, \cdot)) \leq A_\varepsilon e^{-K_\varepsilon t} d_{l^1}(x, y), \forall x, y \in (\mathbb{R}^d)^N, \quad (5.11)$$

which implies that

$$|\nabla_i P_{T-t} F| \leq A_\varepsilon e^{-K_\varepsilon(T-t)}, \ 1 \leq i \leq N, \quad (5.12)$$

where A_ε and K_ε are the same as given in Theorem 2.6.

Since $M_t = \xi_{\tau_t}$ where (ξ_t) is a real valued Brownian motion w.r.t. some new filtration $(\tilde{\mathcal{F}}_t)$ and

$$\tau_t = \langle M \rangle_t = \int_0^t \sum_{i=1}^N |\nabla_i P_{t-s} F(X_s^{(N)})|^2 ds \leq \frac{A_\varepsilon^2}{2K_\varepsilon} N =: CN$$

is a stopping time w.r.t. $(\tilde{\mathcal{F}}_t)$, we obtain

$$
\begin{aligned}
\mathbb{E}_x \varphi \left(F(X_T^{(N)}) - \mathbb{E}_x F(X_T^{(N)}) \right) &= \mathbb{E}\varphi \left(M_T - M_0 \right) = \mathbb{E}\varphi(\xi_{\tau_T}) \\
&= \mathbb{E}\varphi \left(\mathbb{E}(\xi_{CN} | \tilde{\mathcal{F}}_{\tau_T}) \right) \\
&\leq \mathbb{E}\varphi \left(\xi_{CN} \right) \qquad \text{(by Jensen's inequality)} \\
&= \mathbb{E}\varphi \left(A_\varepsilon \sqrt{\frac{N}{2K_\varepsilon}} \xi_1 \right)
\end{aligned}
\tag{5.13}
$$

the desired result (5.8).

Letting $g(x) := \mathbb{E}_x F(X_T^{(N)}), \forall x \in (\mathbb{R}^d)^N$. By (5.12) we have

$$
\|g\|_{Lip(d_{l1})} = \max_{1 \leq i \leq N} \|\nabla_i g\|_\infty \leq A_\varepsilon e^{-K_\varepsilon T}.
\tag{5.14}
$$

Applying (5.8) to $\varphi(z) = e^{\lambda z}$ $(\lambda \in \mathbb{R})$, we get

$$
\begin{aligned}
&\mathbb{E}_{\mu_0^{\otimes N}} \exp \left(\lambda [F(X_T^{(N)}) - \mathbb{E}_{\mu_0^{\otimes N}} F(X_T^{(N)})]) \right) \\
&= \int_{(\mathbb{R}^d)^N} \mathbb{E}_x \exp \left(\lambda [F(X_T^{(N)}) - \mathbb{E}_x F(X_T^{(N)})] \right) \cdot \exp \left(\lambda [g(x) - \mu_0^{\otimes N}(g)] \right) d\mu_0^{\otimes N}(x) \\
&\leq \int_{(\mathbb{R}^d)^N} \mathbb{E} \exp \left(\lambda A_\varepsilon \sqrt{\frac{N}{2K_\varepsilon}} \xi_1 \right) \cdot \exp \left(\lambda [g(x) - \mu_0^{\otimes N}(g)] \right) d\mu_0^{\otimes N}(x) \\
&\leq \exp \left(\frac{N A_\varepsilon^2 \lambda^2}{4 K_\varepsilon} \right) \exp \left(\frac{\lambda^2}{2} N c_G(\mu_0) \|g\|_{Lip(d_{l1})}^2 \right) \\
&\leq \exp \left(\frac{N \lambda^2 A_\varepsilon^2}{2} \left[\frac{1}{2K_\varepsilon} + c_G(\mu_0) e^{-2K_\varepsilon T} \right] \right)
\end{aligned}
$$

where the third and the last inequality follows from the Gaussian concentration condition on the initial distribution μ_0 (see (2.20) in Remark 2.4) and (5.14) respectively.

Finally the concentration inequality (5.9) is derived from the above inequality by the standard procedure of Chebyshev's inequality and optimization over λ. $\qquad\square$

Proof of Theorem 2.10. Let $F(x) = \frac{1}{N} \sum_{i=1}^N f(x^i)$ for $x = (x^1, \cdots, x^N) \in (\mathbb{R}^d)^N$. We have

$$
\alpha = \|F\|_{Lip(d_{l1})} = \frac{1}{N} \|f\|_{Lip} = \frac{1}{N}.
$$

Then the desired concentration inequality (2.31) follows by (5.9). $\qquad\square$

References

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

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

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

[4] F. Bolley, Quantitative concentration inequalities on sample path space for mean field interaction. *ESAIM Probab. Stat.* 14 (2010), 192-209.

[5] F. Bolley, A. Guillin and F. Malrieu, Trend to equilibrium and particle approximation for a weakly selfconsistent Vlasov-Fokker-Planck equation. *ESAIM: Mathematical Modelling and Numerical Analysis* 44 (2010), 867-884.

[6] F. Bolley, A. Guillin and C. Villani, Quantitative concentration inequalities for empirical measures on non-compcact spaces. *Probab. Theory Relat. Fields* 137 (2007), 541-593.

[7] R. Buckdahn, J. Li, S. Peng and C. Rainer, Mean-field stochastic differential equations and associated PDEs. *Ann. Probab.* 45 (2017) 824-878.

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

[9] P. Cattiaux, A. Guillin, and F. Malrieu. Probabilistic approach for granular media equations in the non-uniformly convex case. *Probab. Theory Relat. Fields* 140 (2008) 19-40.

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

[11] H. Djellout, L. Wu, Lipschitzian norm estimate of one-dimensional Poisson equations. *Ann. Inst. Henri Poincaré (B) Prob. and Stat.*, 47 (2) (2011), 450-465.

[12] P. Del Moral, J. Tugaut, Uniform propagation of chaos and creation of chaos for a class of nonlinear diffusions. To appear in Stochastic Analysis and Applications.

[13] A. Durmus, A. Eberle, A. Guillin and R. Zimmer, An elementary approach to uniform in Time propagation of chaos, to appear in Proceedings of the American Mathematical Society.

[14] A. Eberle, Reflection couplings and contraction rates for diffusions. *Probab. Theory Relat. Fields* 166 (2016), 851-886.

[15] A. Eberle, A. Guillin, and R. Zimmer, Quantitative Harris type theorems for diffusions and McKean-Vlasov processes. Published electronically in Trans. Am. Math. Soc., on September 28, 2018.

[16] N. Fournier and A. Guillin, On the rate of convergence in Wasserstein distance of the empirical measure. *Probab. Theory Relat. Fields* 162 (2015), 707-738.

[17] J. Gärtner, On the McKean-Vlasov limit for interacting diffusions. *Mathematische Nachrichten* 137 (1) (1988), 197-248.

[18] A. Guillin, W. Liu, L. Wu and C. Zhang, Poincaré and logarithmic Sobolev inequalities for particles in mean field interactions. https://arxiv.org/abs/1909.07051, 2019.

[19] W. Hammersley, D. Siska, and L. Szpruch, McKean-Vlasov SDEs under measure dependent Lyapunov conditions. https://arxiv.org/abs/1802.03974, 2018.

[20] M. Hauray and S. Mischler, On Kac's chaos and related problems. *J. Funct. Anal.* 266 (10) (2014), 6055-6157.

[21] P. E. Jabin and Z. Wang, Quantitative estimates of propagation of chaos for stochastic systems with $W^{-1,\infty}$ kernels. *Invent. Math.* 214 (2018), 523-591.

[22] D. Lacker, On a strong form of propagation of chaos for McKean-Vlasov equations. *Electron. Commun. Probab. 23 (45) (2018), 1-11.*

[23] M. Ledoux. Concentration of measure and logarithmic Sobolev inequalities. Séminaire de Probabilités XXXVI. Lecture Notes in Math. 1709, 120-216. Springer-Verlag, Berlin, 1999.

[24] M. Ledoux. The concentration of measure phenomenon. Amer. Math. Soc., Providence, RI 2001.

[25] D. Luo and J. Wang, Exponential convergence in L^p-Wasserstein distance for diffusion processes without uniformly dissipative drift, *Math. Nachr.* 289 (2016), 1909-1926.

[26] M. Kac, Foundations of kinetic theory. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1954-1955, vol. III, pages 171-197.* University of California Press, Berkeley and Los Angeles, 1956.

[27] F. Malrieu, Logarithmic sobolev inequalities for some nonlinear PDE's. *Stoch. Proc. Appl.* 95 (2001), 109-132.

[28] F. Malrieu, Convergence to equilibrium for granular media equations and their Euler schemes. *Ann. Appl. Probab.* 13 (2) (2003), 540-560.

[29] S. Méléard, Asymptotic behaviour of some interacting particle systems; McKean-Vlasov and Boltzmann models. In D. Talay and L. Tubaro, editors, Probabilistic Models for Nonlinear Partial Differential Equations, Montecatini Terme, 1995, Lecture Notes in Mathematics, vol. 1627, pp. 42-95, Springer-Verlag, 1996.

[30] H.P. McKean, A class of Markov processes associated with nonlinear parabolic equations. *Proceedings of the National Academy of Sciences* 56 (6) (1966), 1907-1911.

[31] S. Mischler and C. Mouhot, Kac's program in kinetic theory. *Invent. Math.* 193 (1) (2013), 1- 147.

[32] S. Mischler, C. Mouhot, and B. Wennberg, A new approach to quantitative propagation of chaos for drift, diffusion and jump processes. *Probab. Theory Relat. Fields* 161 (1) (2015), 1-59.

[33] Y.S. Mishura and A.Y. Veretennikov, Existence and uniqueness theorems for solutions of McKean–Vlasov stochastic equations. https://arxiv.org/abs/1603.02212, 2016.

[34] A.S. Sznitman, Topics in propagation of chaos. École d'Été de Probabilités de Saint-Flour XIX, Lecture Notes in Math., vol. 1464, pp. 165-251, 1991.

[35] L. Wu, Gradient estimates of Poisson equations on Riemannian manifolds and applications. *J. Funct. Anal.* 257 (2009), 4015-4033.

Wei Liu, School of Mathematics and Statistics, Wuhan University, Wuhan, Hubei 430072, People's Republic of China; Hubei Key Laboratory of Computational Science, Wuhan University, Wuhan, Hubei 430072, People's Republic of China.
Email address: `wliu.math@whu.edu.cn`

Liming Wu. Laboratoire de Mathématiques Blaise Pascal, CNRS-UMR 6620, Université Clermont-Auvergne (UCA), 63000 Clermont-Ferrand, France; Institute for Advanced Study in Mathematics, Harbin Institute of Technology, NO.92 West Da Zhi St. Harbin, People's Republic of China.
Email address: `Li-Ming.Wu@uca.fr`

Chaoen Zhang. Institute for Advanced Study in Mathematics, Harbin Institute of Technology, NO.92 West Da Zhi St. Harbin, People's Republic of China.
Email address: `chaoenzhang@hit.edu.cn`