



## CONCENTRATION OF MEASURE FOR GRAPHON PARTICLE SYSTEM

ERHAN BAYRAKTAR \* \*\* AND  
DONGHAN KIM,\* \*\*\* *University of Michigan*

### Abstract

We study heterogeneously interacting diffusive particle systems with mean-field-type interaction characterized by an underlying graphon and their finite particle approximations. Under suitable conditions, we obtain exponential concentration estimates over a finite time horizon for both 1- and 2-Wasserstein distances between the empirical measures of the finite particle systems and the averaged law of the graphon system.

**Keywords:** Mean-field interaction; heterogeneous interaction; networks; concentration bounds; transportation inequalities;  $L^1$ -Fourier class

2020 Mathematics Subject Classification: Primary 60J60; 60K35  
Secondary 05C80

### 1. Introduction

In this article, we study concentration of measures related to the graphon particle system and its finite particle approximations. This work is a continuation of the earlier papers [2–4]. A graphon particle system consists of uncountably many heterogeneous particles  $X_u$  for  $u \in [0, 1]$  whose interactions are characterized by a graphon. More precisely, for a fixed  $T > 0$  and  $d \in \mathbb{N}$ , we consider the following system:

$$X_u(t) = X_u(0) + \int_0^t \left[ \int_0^1 \int_{\mathbb{R}^d} \phi(X_u(s), y) G(u, v) \mu_{v,s}(dy) dv + \psi(X_u(s)) \right] ds + \sigma B_u(t),$$

$\mu_{u,t}$  is the probability distribution of  $X_u(t)$ , for every  $u \in [0, 1]$  and  $t \in [0, T]$ , (1.1)

where  $\{B_u\}_{u \in [0,1]}$  is a family of independent and identically distributed  $d$ -dimensional Brownian motions, and  $\{X_u(0)\}_{u \in [0,1]}$  is a collection of independent (but not necessarily identically distributed)  $\mathbb{R}^d$ -valued random variables with law  $\mu_u(0)$ , independent of  $\{B_u\}_{u \in [0,1]}$  for each  $u \in [0, 1]$ , defined on a filtered probability space  $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P})$ . Two functions  $\phi : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^d$  and  $\psi : \mathbb{R}^d \rightarrow \mathbb{R}^d$  represent pairwise interactions between the particles and single-particle drift, respectively. The quantity  $\sigma \in \mathbb{R}^{d \times d}$  is a constant, and  $G : [0, 1] \times [0, 1] \rightarrow [0, 1]$  is a graphon, that is, a symmetric measurable function.

Along with the graphon particle system, we introduce two finite particle systems with heterogeneous interactions, which approximate (1.1). For a fixed, arbitrary  $n \in \mathbb{N}$  and each  $i \in [n] := \{1, \dots, n\}$ , we first consider the ‘not-so-dense’ analogue of (1.1) introduced in

---

Received 2 October 2022; accepted 1 December 2023.

\* Postal address: Department of Mathematics, 530 Church Street, Ann Arbor, MI 48109.

\*\* Email address: [erhan@umich.edu](mailto:erhan@umich.edu)

\*\*\* Email address: [donghank@umich.edu](mailto:donghank@umich.edu)

© The Author(s), 2024. Published by Cambridge University Press on behalf of Applied Probability Trust. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

Section 4 of [2]:

$$X_i^n(t) = X_{\frac{i}{n}}(0) + \int_0^t \left[ \frac{1}{np(n)} \sum_{j=1}^n \xi_{ij}^n \phi(X_i^n(s), X_j^n(s)) + \psi(X_i^n(s)) \right] ds + \sigma B_{\frac{i}{n}}(t), \quad (1.2)$$

where  $\{p(n)\}_{n \in \mathbb{N}} \subset (0, 1]$  is a sequence of numbers and  $\{\xi_{ij}^n\}_{1 \leq i, j \leq n}$  are independent Bernoulli random variables satisfying

$$\xi_{ij}^n = \xi_{ji}^n, \quad \mathbb{P}(\xi_{ij}^n = 1) = p(n)G\left(\frac{i}{n}, \frac{j}{n}\right), \quad \text{for every } i, j \in [n],$$

independent of  $\{B_{i/n}, X_{i/n}(0) : i \in [n]\}$ . Here,  $p(n)$  represents the global sparsity parameter, and the strength of interaction between the particles in (1.2) is scaled by  $np(n)$ , the order of the number of neighbors, as in mean-field systems on Erdős–Rényi random graphs [7, 15, 27]; the convergence of  $p(n) \rightarrow 0$  as  $n \rightarrow \infty$  implies that the graph is sparse, but we shall consider the case  $np(n) \rightarrow \infty$ , i.e., the random graph is ‘not so dense’, meaning that the average degree of the graph diverges.

The other finite particle approximation system is given by

$$\bar{X}_i^n(t) = X_{\frac{i}{n}}(0) + \int_0^t \left[ \frac{1}{n} \sum_{j=1}^n \phi(\bar{X}_i^n(s), \bar{X}_j^n(s))G\left(\frac{i}{n}, \frac{j}{n}\right) + \psi(\bar{X}_i^n(s)) \right] ds + \sigma B_{\frac{i}{n}}(t). \quad (1.3)$$

Since this system has a nonrandom coefficient for the interaction term (but still models heterogeneous interaction via the graphon), it is easier to analyze than the other finite particle system (1.2). We note that the three systems (1.1)–(1.3) are coupled in the sense that they share initial particle locations  $X_{i/n}(0)$  and Brownian motions  $B_{i/n}$  for  $i \in [n]$ .

Law-of-large-numbers-type results on the convergence of the systems (1.2) and (1.3) to the graphon particle system (1.1) under suitable conditions are studied in [2]. Results on the exponential ergodicity of the two systems (1.1) and (1.2), as well as the uniform-in-time convergence of (1.2) to (1.1) under a certain dissipativity condition, are presented in [3]. There are numerous recent studies of graphon particle systems [4, 14] and works on associated heterogeneously interacting finite particle models [6, 13, 17, 25–27]. These studies have been undertaken because graphons have been widely applied in mean-field game theory for both the static and dynamic cases; see e.g. [1, 5, 10–12, 20, 21, 28, 30, 31, 33] and references therein.

Among these studies, our work is particularly linked to [4]. With  $W_1$  denoting the 1-Wasserstein distance, with the empirical measures of the three particle systems at time  $t \in [0, T]$  defined by

$$L_{n,t} := \frac{1}{n} \sum_{i=1}^n \delta_{(X_i^n(t))}, \quad \bar{L}_{n,t} := \frac{1}{n} \sum_{i=1}^n \delta_{(\bar{X}_i^n(t))}, \quad \tilde{L}_{n,t} := \frac{1}{n} \sum_{i=1}^n \delta_{(X_{i/n}(t))}, \quad (1.4)$$

and with the averaged law  $\tilde{\mu}_t := \int_0^1 \mu_{u,t} du$  of the graphon system (1.1), [4] computes concentration bounds of the types  $\mathbb{P}[\sup_{0 \leq t \leq T} W_1(\bar{L}_{n,t}, \tilde{\mu}_t) > \epsilon]$ ,  $\sup_{t \geq 0} \mathbb{P}[W_1(\bar{L}_{n,t}, \tilde{\mu}_t) > \epsilon]$  for  $\epsilon > 0$  under certain conditions. In particular, uniform-in-time concentration bounds of the latter type are studied in an infinite-time-horizon setting under an extra dissipativity condition on  $\psi$ . These results are established by computing certain sub-Gaussian estimates rather directly with the moment generating function of the standard normal random vector (Lemmas 3.7–3.10 of [4]).

In contrast, the present work focuses on the case of finite time horizon and deals with a more general sparsity sequence  $\{p(n)\}_{n \in \mathbb{N}} \subset (0, 1]$  for (1.2), whereas the results of [4] cover only the dense graphs, i.e.,  $p(n) \equiv 1$ . For our argument we adopt the method of [16], as follows. We first compute the bound on the probability that Lipschitz functions of the finite particles  $\bar{X}^n = (\bar{X}_1^n, \dots, \bar{X}_n^n)$  from the system (1.3) on the space  $(C([0, T] : \mathbb{R}^d))^n$  deviate from their means (Theorem 3.1). The derivation of this ‘concentration bound around their means’ relies on transportation inequalities from [18]. Combining this bound with the fact, presented in Section 3.1, that the expectations of the  $W_2$ -distances between the empirical measures in (1.4) converge to zero as the number of particles goes to infinity, we show exponential concentration bounds on the probabilities

$$\mathbb{P} \left[ \sup_{0 \leq t \leq T} W_p(\bar{L}_{n,t}, \tilde{\mu}_t) > \epsilon \right] \quad (1.5)$$

for  $p = 1$  (Theorem 3.3).

The advantage of using transportation inequalities to find ‘concentration around means’, compared to the approach of [4], is that we can even derive the same exponential bound (1.5) in the case of  $p = 2$  (Theorem 3.5), at the cost of assuming independence and (3.3) on the initial particles. More specifically, we can obtain a *dimension-free* concentration bound around means,

$$\mathbb{P} \left[ F(\bar{X}^n) - \mathbb{E}(F(\bar{X}^n)) > a \right] \leq 2 \exp(-\delta a^2),$$

for every Lipschitz function  $F$  and  $a > 0$  (see (3.4))—i.e., the right-hand side does not depend on  $n$ —from the quadratic transportation inequality (3.3) on the initial particles. This bound can be derived thanks to the remarkable result on the dimension-free  $W_2$ -tensorization of transportation inequalities from [22] (Lemma 2.7(ii)). Then, again by combining the dimension-free concentration bound around means with the results in Section 3.1, we ultimately arrive at the exponential bound (1.5) for  $p = 2$ . The authors of [16] apply this method to compute similar concentration bounds when  $\bar{X}^n$  represents the state of the so-called Nash equilibrium of a symmetric  $n$ -player stochastic differential game and  $\tilde{\mu}$  is the measure flow of the unique equilibrium of the corresponding mean-field game (we refer to Section 2 of [16] for a more detailed description of these terms).

Moreover, inspired by the argument using Bernstein’s inequality in [27], we compare the particles  $X_i^n$  and  $\bar{X}_i^n$  to improve the exponential bound in [4] for

$$\mathbb{P} \left[ \sup_{0 \leq t \leq T} W_p(L_{n,t}, \tilde{\mu}_t) > \epsilon \right]$$

when  $p = 1$ , at the cost of an assumption on the interaction function  $\phi$ , namely, that it is a member of the  $L^1$ -Fourier class. When  $p = 2$ , we also present a similar exponential bound for the system (1.2) on the dense graphs ( $p(n) \equiv 1$ ). This requires a refinement of the Bernstein-type inequality for the cut norm in [27]; see Lemma 3.1. Without such a condition on  $\phi$ , we have a similar exponential bound, but in terms of the bounded Lipschitz metric ( $d_{BL}$  metric), a weaker metric than  $W_1$ .

This paper is organized as follows. In Section 2 we introduce the notation, state the assumptions, and recall some of the relevant existing results concerning the particle systems (1.1)–(1.3), as well as other preliminary results. Section 3 provides our main results, and Section 4 gives the proofs of these results.

## 2. Preliminaries

In this section, we first introduce the notation which will be used throughout this paper. We then state several assumptions and some of the basic results on the particle systems (1.1)–(1.3) from [2, 4], and provide several well-known results regarding transportation cost inequalities without proof. Finally, we introduce Bernstein’s inequality and the concept of the  $L^1$ -Fourier class.

### 2.1. Notation

Given a metric space  $(S, d)$  and a function  $f : S \rightarrow \mathbb{R}$ , we define

$$\begin{aligned} \|f\|_\infty &:= \sup\{|f(x)| : x \in S\}, \\ \|f\|_{Lip} &:= \sup \left\{ \frac{|f(x) - f(y)|}{d(x, y)} : x, y \in S, x \neq y \right\}, \\ \|f\|_{BL} &:= \|f\|_\infty + \|f\|_{Lip}, \end{aligned}$$

and we say that  $f$  is Lipschitz (respectively, bounded Lipschitz) if  $\|f\|_{Lip} < \infty$  (respectively,  $\|f\|_{BL} < \infty$ ). In particular,  $f$  is called  $a$ -Lipschitz if  $\|f\|_{Lip} = a$ .

Denote by  $\mathcal{P}(S)$  the space of Borel probability measures on  $S$ . We shall use the standard notation  $\langle \mu, \varphi \rangle := \int_S \varphi \, d\mu$  for integrable functions  $\varphi$  and measures  $\mu$  on  $S$ . When  $(S, \|\cdot\|)$  is a normed space, we write  $\mathcal{P}^p(S, \|\cdot\|)$  for the set of  $\mu \in \mathcal{P}(S)$  satisfying  $\langle \mu, \|\cdot\|^p \rangle < \infty$  for a given  $p \in [1, \infty)$ . We denote by  $Lip(S, \|\cdot\|)$  the set of 1-Lipschitz functions, i.e.,  $f : S \rightarrow \mathbb{R}$  satisfying  $|f(x) - f(y)| \leq \|x - y\|$  for every  $x, y \in S$ .

For a separable Banach space  $(S, \|\cdot\|)$ , we endow  $\mathcal{P}^p(S, \|\cdot\|)$  with the  $p$ -Wasserstein metric

$$W_{p, (S, \|\cdot\|)}(\mu, \nu) := \inf_{\pi} \left( \int_{S \times S} \|x - y\|^p \pi(dx, dy) \right)^{1/p}, \quad p \geq 1, \quad (2.1)$$

where the infimum is taken over all probability measures  $\pi$  on  $S \times S$  with first and second marginals  $\mu$  and  $\nu$ . We also consider the product space  $S^n := S \times \cdots \times S$ , equipped with the  $\ell^p$  norm (for any  $p \geq 1$ )

$$\|x\|_{n, p} = \left( \sum_{i=1}^n \|x_i\|^p \right)^{1/p}, \quad (2.2)$$

for  $x = (x_1, \dots, x_n) \in S^n$ . When the space  $S$  or the norm  $\|\cdot\|$  is understood, we sometimes omit it from the above notation.

Denote by  $C([0, T] : S)$  the space of continuous functions from  $[0, T]$  to  $S$ , and  $\|x\|_{\star, t} := \sup_{0 \leq s \leq t} |x_s|$ , where  $|\cdot|$  is the usual Euclidean norm on  $\mathbb{R}^d$  for  $x \in C([0, T] : \mathbb{R}^d)$  and  $t \in [0, T]$ . We write  $\mathcal{L}(X)$  for the probability law of a random variable  $X$  and  $[n] := \{1, \dots, n\}$  for any  $n \in \mathbb{N}$ . We use  $K$  to denote various positive constants throughout the paper; its value may change from line to line.

For a Polish space  $(S, d)$  with Borel  $\sigma$ -field  $\mathcal{S}$ , we also consider the space of probability measures over  $(S, \mathcal{S})$  endowed with the topology of weak convergence, which is metrized by the  $BL$  metric, defined for  $\mu, \nu \in \mathcal{P}(S)$  by

$$d_{BL}(\mu, \nu) := \sup \left\{ \left| \int_S f \, d(\mu - \nu) \right| : f : S \rightarrow \mathbb{R} \text{ with } \|f\|_{BL} \leq 1 \right\}. \quad (2.3)$$

Note the dual representation of the 1-Wasserstein metric

$$W_1(\mu, \nu) := \sup \left\{ \left| \int_S f d(\mu - \nu) \right| : f : S \rightarrow \mathbb{R} \text{ with } \|f\|_{Lip} \leq 1 \right\}, \quad (2.4)$$

along with the relationship  $d_{BL} \leq W_1$ . We shall also use the following notation: for given  $\mu, \nu \in \mathcal{P}(C([0, T] : \mathbb{R}^d))$ ,

$$W_{p,t}(\mu, \nu) := \inf_{\pi} \left( \int \|x - y\|_{\star,t}^p \pi(dx, dy) \right)^{1/p}, \quad t \in [0, T], \quad p \geq 1,$$

where the infimum is taken over all probability measures  $\pi$  with marginals  $\mu$  and  $\nu$ .

Let us define three  $n \times n$  random matrices  $P^{(n)}, \bar{P}^{(n)}$ , and  $D^{(n)}$ , related to the systems (1.2), (1.3), for every  $n \in \mathbb{N}$ , with entries

$$\begin{aligned} P_{i,j}^{(n)} &:= \frac{\xi_{ij}^n}{np(n)}, & i, j \in [n], \\ \bar{P}_{i,j}^{(n)} &:= \frac{1}{n} G\left(\frac{i}{n}, \frac{j}{n}\right), & i, j \in [n], \\ D^{(n)} &:= P^{(n)} - \bar{P}^{(n)}. \end{aligned} \quad (2.5)$$

For these matrices, we define the  $\ell_\infty \rightarrow \ell_1$  norm of an  $n \times n$  matrix  $A$  by

$$\|A\|_{\infty \rightarrow 1} := \sup \left\{ \langle \mathbf{x}, A\mathbf{y} \rangle : \mathbf{x}, \mathbf{y} \in [-1, 1]^n \right\}. \quad (2.6)$$

This norm is known to be equivalent to the so-called *cut norm* (see (3.3) of [23]).

We denote the empirical measures of the approximation systems for each  $n \in \mathbb{N}$  by

$$L_n := \frac{1}{n} \sum_{i=1}^n \delta_{(X_i^n)}, \quad \bar{L}_n := \frac{1}{n} \sum_{i=1}^n \delta_{(\bar{X}_i^n)}, \quad (2.7)$$

all of which are random elements of  $\mathcal{P}(C([0, T] : \mathbb{R}^d))$ .

We conclude this subsection by recalling the relative entropy of two probability measures  $\mu, \nu$  over the same measurable space:

$$H(\mu|\nu) := \begin{cases} \int \log\left(\frac{d\mu}{d\nu}\right) d\mu & \text{if } \mu \ll \nu, \\ \infty & \text{otherwise.} \end{cases} \quad (2.8)$$

## 2.2. Existence and uniqueness of the solutions

We state the existence and uniqueness of strong solutions to the systems (1.1)–(1.3).

### Assumption 2.1

(i) *The function  $\phi$  is bounded; furthermore,  $\phi$  and  $\psi$  are Lipschitz, i.e., there exists a constant  $K > 0$  such that*

$$|\phi(x_1, y_1) - \phi(x_2, y_2)| + |\psi(x_1) - \psi(x_2)| \leq K(|x_1 - x_2| + |y_1 - y_2|)$$

holds. Moreover, the initial particles have finite second moments, i.e.,

$$\sup_{u \in [0,1]} \mathbb{E}|X_u(0)|^2 < \infty. \quad (2.9)$$

(ii) The map  $[0, 1] \ni u \mapsto \mu_u(0) = \mathcal{L}(X_u(0)) \in \mathcal{P}(\mathbb{R}^d)$  is measurable.

**Lemma 2.1.** (Existence and uniqueness of the particle systems)

- (i) Under Assumption 2.1(i), the two systems (1.2), (1.3) have unique strong solutions.
- (ii) Under Assumption 2.1(i)–(ii), the graphon system (1.1) has a unique strong solution, and the map  $[0, 1] \ni u \mapsto \mu_u \in \mathcal{P}(C([0, T] : \mathbb{R}^d))$  is measurable.

The proof of Lemma 2.1(i) is classical (see e.g. Theorem 5.2.9 of [24]). Part (ii) follows from Proposition 2.1 of [2]. As pointed out in Remark 2.2 of [2], we note that the boundedness condition on  $\phi$  in Assumption 2.1(i) can be removed throughout this paper, at the cost of a stronger condition  $\sup_{u \in [0,1]} \mathbb{E}|X_u(0)|^{2+\epsilon} < \infty$  for some  $\epsilon > 0$  than (2.9). We occasionally need an even stronger condition on the initial particles, as in the following.

**Assumption 2.2.** The initial particles  $\{X_u(0)\}_{u \in [0,1]}$  are independent, with law  $\mu_{u,0} \in \mathcal{P}(\mathbb{R}^d)$  satisfying

$$\sup_{u \in [0,1]} \int_{\mathbb{R}^d} e^{\kappa|x|^2} \mu_{u,0}(dx) < \infty, \quad \text{for some } \kappa > 0. \quad (2.10)$$

Under this stronger assumption, we have, in particular, the finite fourth moment of the solution to (1.1). The proof is standard and hence is omitted (see e.g. [29] or Proposition 2.1 of [4]).

**Lemma 2.2.** Under Assumptions 2.1, 2.2, the solution to (1.1) satisfies

$$\sup_{u \in [0,1]} \sup_{t \in [0,T]} \mathbb{E}[|X_u(t)|^4] < \infty.$$

### 2.3. Continuity of the graphon system

The following result, which states the continuity of the graphon system (1.1), is from Theorem 2.1 of [2].

**Assumption 2.3.** There exists a finite collection of subintervals  $\{I_i : i \in [N]\}$ , for some  $N \in \mathbb{N}$ , satisfying  $\cup_{i=1}^N I_i = [0, 1]$ . For each  $i, j \in [N]$ , the following hold:

- (i) The map  $I_i \ni u \mapsto \mu_u(0) \in \mathcal{P}(\mathbb{R}^d)$  is continuous with respect to the  $W_2$  metric.
- (ii) For each  $u \in I_i$ , there exists a Lebesgue-null set  $N_u \subset [0, 1]$  such that  $G(u, v)$  is continuous at  $(u, v) \in [0, 1] \times [0, 1] \setminus N_u$ .
- (iii) There exists  $K > 0$  such that

$$W_2(\mu_{u_1}(0), \mu_{u_2}(0)) \leq K|u_1 - u_2|, \quad u_1, u_2 \in [0, 1],$$

$$|G(u_1, v_1) - G(u_2, v_2)| \leq K(|u_1 - u_2| + |v_1 - v_2|), \quad (u_1, v_1), (u_2, v_2) \in I_i \times I_j.$$

**Lemma 2.3.** Suppose that Assumption 2.1 holds.

- (i) (Continuity.) Under Assumption 2.3(i)–(ii), the map  $I_i \ni u \mapsto \mu_u \in \mathcal{P}(C([0, T] : \mathbb{R}^d))$  is continuous with respect to the  $W_{2,T}$  metric for every  $i \in [N]$ .
- (ii) (Lipschitz continuity.) Under Assumption 2.3(iii), there exists  $\kappa > 0$ , which depends on  $T$ , such that  $W_{2,T}(\mu_u, \mu_v) \leq \kappa |u - v|$  whenever  $u, v \in I_i$  for some  $i \in [N]$ .

In Lemma 2.3(ii), note that we have, in particular,

$$\sup_{\|f\|_{Lip} \leq 1} \left| \int_{\mathbb{R}^d} f(x) \mu_{u,t}(dx) - \int_{\mathbb{R}^d} f(x) \mu_{v,t}(dx) \right| \leq W_{2,t}(\mu_u, \mu_v) \leq \kappa |u - v|, \quad \forall t \in [0, T].$$

## 2.4. A law of large numbers for the mean-field particle system

Besides the assumptions already introduced in this section, we will need the following assumption on the sparsity parameter for the system (1.2), as briefly mentioned in Section 1.

**Assumption 2.4.** The sequence  $\{p(n)\}_{n \in \mathbb{N}}$  in (1.2) satisfies  $np(n) \rightarrow \infty$  as  $n \rightarrow \infty$ .

We introduce the following law-of-large-numbers result for the mean-field particle system (1.2), which is Theorem 4.1 of [2]. We write  $\mu_u$  for the law of  $X_u$  in the graphon particle system (1.1) for each  $u \in [0, 1]$ , and define

$$\tilde{\mu} := \int_0^1 \mu_u du. \quad (2.11)$$

**Lemma 2.4.** Under Assumptions 2.1, 2.3, and 2.4,

$$L_n \rightarrow \tilde{\mu} \quad \text{in } \mathcal{P}(C([0, T] : \mathbb{R}^d)) \text{ in probability, as } n \rightarrow \infty.$$

Moreover, we have

$$\frac{1}{n} \sum_{i=1}^n \mathbb{E} \|X_i^n - X_{\frac{i}{n}}\|_{\star, T}^2 \rightarrow 0, \quad \text{as } n \rightarrow \infty.$$

## 2.5. Transportation inequalities

In this subsection, we present some preliminary results regarding transportation inequalities. The first result, from Theorem 9.1 of [32], illustrates the transportation inequality with the uniform norm for the laws of diffusion processes.

**Lemma 2.5.** For a fixed  $T > 0$  and  $k \in \mathbb{N}$ , suppose that  $X^x = \{X_t^x\}_{t \in [0, T]}$  is the unique strong solution of the stochastic differential equation (SDE)

$$dX_t^x = b(t, X^x) dt + \Sigma dW_t, \quad \forall t \in [0, T], \quad X_0 = x \in \mathbb{R}^k, \quad (2.12)$$

on a probability space  $C([0, T] : \mathbb{R}^k)$  supporting a  $k$ -dimensional Brownian motion  $W$ . Here,  $b : [0, T] \times C([0, T] : \mathbb{R}^k) \rightarrow \mathbb{R}^k$  satisfies, for any  $\xi, \eta \in C([0, T] : \mathbb{R}^k)$ ,

$$|b(t, \xi) - b(t, \eta)| \leq L \sup_{0 \leq s \leq t} |\xi(s) - \eta(s)| = L \|\xi - \eta\|_{\star, t}, \quad \forall t \in [0, T], \quad (2.13)$$

for some constants  $L > 0$  and  $\Sigma \in \mathbb{R}^{k \times k}$ . Let  $P^x \in \mathcal{P}(C([0, T] : \mathbb{R}^k))$  be the law of  $X^x$  for any  $x \in \mathbb{R}^k$ . Then for any  $Q \in \mathcal{P}(C([0, T] : \mathbb{R}^k))$ , there exist positive constants  $\kappa_1, \kappa_2$ , depending

only on  $T$ , such that the inequalities

$$W_{1,(C([0,T]:\mathbb{R}^k), \|\cdot\|_{k,2})}^2(P^x, Q) \leq W_{2,(C([0,T]:\mathbb{R}^k), \|\cdot\|_{k,2})}^2(P^x, Q) \leq \kappa_1 e^{\kappa_2 L^2} H(Q|P^x)$$

hold, where  $H(Q|P)$  is the relative entropy of  $Q$  with respect to  $P$ , defined in (2.8).

The following result (Theorem 5.1 of [16]) characterizes the concentration of a probability measure with a transportation cost inequality and Gaussian integrability property. The equivalence between (2.14) and (2.15) is originally from Theorem 3.1 of [8], and the equivalence between (2.14) and (2.17) is due to Theorem 2.3 of [18].

**Lemma 2.6.** *For a probability measure  $\mu \in \mathcal{P}^1(S)$  on a separable Banach space  $(S, \|\cdot\|)$ , the following statements are equivalent up to a universal change in the positive constant  $c$ :*

(i) *The transportation cost inequality*

$$W_{1,S}(\mu, \nu) \leq \sqrt{2cH(\nu|\mu)} \quad (2.14)$$

*holds for every  $\nu \in \mathcal{P}(S)$ .*

(ii) *For every 1-Lipschitz function  $f$  on  $S$  and  $\lambda \in \mathbb{R}$ ,*

$$\int_S e^{\lambda(f - \langle \mu, f \rangle)} d\mu \leq \exp\left(\frac{c\lambda^2}{2}\right) \quad (2.15)$$

*holds.*

(iii) *For every 1-Lipschitz function  $f$  on  $S$  and  $a > 0$ ,*

$$\mu(f - \langle \mu, f \rangle > a) \leq \exp\left(-\frac{a^2}{2c}\right). \quad (2.16)$$

(iv) *The probability measure  $\mu$  is sub-Gaussian, i.e.,*

$$\int_S e^{c\|x\|^2} \mu(dx) < \infty. \quad (2.17)$$

The next result is a well-known tensorization of transportation cost inequalities from Corollary 5 of [22]. The major difference between (i) and (ii) is that the inequality (2.18) is *dimension-free*, i.e., the right-hand side does not depend on  $n$ .

**Lemma 2.7.** *For each  $n \in \mathbb{N}$ , consider a set of probability measures  $\{\mu_i\}_{i \in [n]} \subset \mathcal{P}(S)$  on a separable Banach space  $(S, \|\cdot\|)$ .*

(i) *If the inequality  $W_{1,S}(\mu_i, \nu) \leq \sqrt{2cH(\nu|\mu_i)}$  holds for every  $i \in [n]$  and  $\nu \in \mathcal{P}^1(S)$ , then*

$$W_{1,(S^n, \|\cdot\|_{n,1})}(\mu_1 \otimes \cdots \otimes \mu_n, \rho) \leq \sqrt{2ncH(\rho|\mu_1 \otimes \cdots \otimes \mu_n)}$$

*holds for every  $\rho \in \mathcal{P}^1(S^n)$ .*

(ii) *If the inequality  $W_{2,S}(\mu_i, \nu) \leq \sqrt{2cH(\nu|\mu_i)}$  holds for every  $i \in [n]$  and  $\nu \in \mathcal{P}^2(S)$ , then*

$$W_{2,(S^n, \|\cdot\|_{n,2})}(\mu_1 \otimes \cdots \otimes \mu_n, \rho) \leq \sqrt{2cH(\rho|\mu_1 \otimes \cdots \otimes \mu_n)} \quad (2.18)$$

*holds for every  $\rho \in \mathcal{P}^2(S^n)$ .*

We finally mention the following result on the Wasserstein distance of the empirical measures of independent but not necessarily identically distributed random variables. This is Lemma A.1 of [4], a generalization of Theorem 1 of [19], where independent and identically distributed random variables are considered. This result will be used in proving Proposition 3.2.

**Lemma 2.8.** *Let  $\{Y_i\}_{i \in \mathbb{N}}$  be independent  $\mathbb{R}^d$ -valued random variables, and define*

$$v_n := \frac{1}{n} \sum_{i=1}^n \delta_{Y_i}, \quad \bar{v}_n := \frac{1}{n} \sum_{i=1}^n \mathcal{L}(Y_i).$$

For a fixed  $p > 0$ , assume that  $\sup_{i \in \mathbb{N}} \mathbb{E}|Y_i|^q < \infty$  holds for some  $q > p$ . Then there exists a constant  $K > 0$  depending only on  $p$ ,  $q$ , and  $d$  such that for every  $n \geq 1$ ,

$$\mathbb{E}[W_p^p(v_n, \bar{v}_n)] \leq K \alpha_{p,q}(n) \left( \int_{\mathbb{R}^d} |x|^q \bar{v}_n(dx) \right)^{p/q},$$

where

$$\alpha_{p,q}(n) := \begin{cases} n^{-1/2} + n^{-(q-p)/q} & \text{if } p > d/2 \text{ and } q \neq 2p, \\ n^{-1/2} \log(1+n) + n^{-(q-p)/q} & \text{if } p = d/2 \text{ and } q \neq 2p, \\ n^{-p/d} + n^{-(q-p)/q} & \text{if } p < d/2 \text{ and } q \neq d/(d-p). \end{cases}$$

## 2.6. Bernstein's inequality and the $L^1$ -Fourier class

When comparing the two approximation systems (1.2) and (1.3), it is essential to control the matrix  $D^{(n)}$  of (2.5). Thus, we introduce the following concentration of  $D^{(n)}$  in terms of the  $\|\cdot\|_{\infty \rightarrow 1}$  norm, which is from Lemma 2 of [27]. Its proof is a straightforward application of Bernstein's inequality (Lemma 2.10, or Bennett's inequality) with the distribution of the  $n^2$  independent entries of the matrix  $D^{(n)}$ . We will use Bernstein's inequality again in Section 3.3 to prove Lemma 3.1, an elaboration of Lemma 2.9.

**Lemma 2.9.** *For any  $0 < \eta \leq n$ , we have*

$$\mathbb{P}\left[\frac{\|D^{(n)}\|_{\infty \rightarrow 1}}{n} > \eta\right] \leq \exp\left(-\frac{\eta^2 n^2 p(n)}{2 + \frac{\eta}{3}}\right).$$

In particular, under Assumption 2.4, we have for every  $\eta > 0$

$$\frac{1}{n} \log \mathbb{P}\left[\frac{\|D^{(n)}\|_{\infty \rightarrow 1}}{n} > \eta\right] \rightarrow -\infty, \quad \text{as } n \rightarrow \infty.$$

**Lemma 2.10.** (Bernstein's inequality, Theorem 2.9 of [9].) *Let  $X_1, \dots, X_k$  be independent random variables with finite variance such that  $X_i \leq b$  for some  $b > 0$  almost surely for each  $i \in [k]$ . Let  $v = \sum_{i=1}^k \mathbb{E}[X_i^2]$ ; then we have*

$$\mathbb{P}\left[\sum_{i=1}^k (X_i - \mathbb{E}(X_i)) \geq u\right] \leq \exp\left(-\frac{u^2}{2(v + \frac{bu}{3})}\right).$$

When the interaction function  $\phi$  belongs to a special class of functions, we shall see in the proof of Theorem 3.4 that the distance  $W_1(L_n, \bar{L}_n)$  can easily be expressed in terms of

the quantity  $\|D^{(n)}\|_{\infty \rightarrow 1}$ . This observation is inspired by the work of [27]. To state it more precisely, we introduce the notion of the  $L^1$ -Fourier class of functions.

**Definition 2.1.** Identifying  $\mathbb{R}^{2d}$  with  $\mathbb{R}^d \times \mathbb{R}^d$ , we say that a function  $f: \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$  belongs to the  $L^1$ -Fourier class if there exists a finite complex measure  $m_f$  over  $\mathbb{R}^{2d}$  such that for every  $(x, y) \in \mathbb{R}^d \times \mathbb{R}^d$ ,

$$f(x, y) = \int_{\mathbb{R}^{2d}} \exp(2\pi\sqrt{-1}\langle (x, y), z \rangle) m_f(dz).$$

We recall that a finite complex measure  $m$  over  $\mathbb{R}^{2d}$  is a set function  $m: \mathcal{B}(\mathbb{R}^{2d}) \rightarrow \mathbb{C}$  of the form  $m = m_r^+ - m_r^- + \sqrt{-1}(m_i^+ - m_i^-)$ , where each of  $m_r^+, m_r^-, m_i^+, m_i^-$  is a finite,  $\sigma$ -additive (nonnegative) measure over  $\mathbb{R}^{2d}$ . We define the total mass of  $m$  by

$$\|m\|_{TM} := m_r^+(\mathbb{R}^{2d}) + m_r^-(\mathbb{R}^{2d}) + m_i^+(\mathbb{R}^{2d}) + m_i^-(\mathbb{R}^{2d}).$$

If a function  $f$  is an inverse  $L^1$ -transform of a function in  $L^1(\mathbb{R}^{2d})$ , then  $f$  belongs to the  $L^1$ -Fourier class. In particular, any Schwartz function belongs to the  $L^1$ -Fourier class. An example of such a function is the Kuramoto interaction; if  $d = 1$  and  $\phi(x - y) = K \sin(y - x)$  for some constant  $K$ , then the corresponding complex measure is equal to

$$m_\phi = \frac{K}{2\sqrt{-1}}(\delta_{(-1,1)} + \delta_{(1,-1)}).$$

The finite system (1.2) of ‘oscillators’ with the Kuramoto interaction function is studied in [13].

### 3. Main results

This section consists of three subsections. The first shows that expectations of the  $W_2$ -distances between two empirical measures on  $\mathbb{R}^d$  related to the systems (1.1)–(1.3) converge to zero as the number of particles goes to infinity. The second subsection gives exponential bounds on the probabilities that Lipschitz function values of the particles  $\bar{X}^n$  of the system (1.3) on  $(C([0, T] : \mathbb{R}^d))^n$  deviate from their means; the stronger the norm we use for the space  $(C([0, T] : \mathbb{R}^d))^n$ , the stronger the assumption we need on the initial distribution of the particles. The results in the first two subsections will be used in proving the results in the last subsection. In the latter, we derive several results on the concentration of the finite particle systems (1.2), (1.3) toward the graphon particle system (1.1), under different metrics.

#### 3.1. Concentration in mean of the $W_2$ -distance

Let us recall the law  $\mu_{u,t}$  of (1.1), the empirical measures (1.4) of the three systems, and the averaged law  $\tilde{\mu}_t := \int_0^1 \mu_{u,t} du$  for every  $t \in [0, T]$ . We give two expectations converging to zero as  $n \rightarrow \infty$  in the following. The proofs are provided in Section 4.1.

**Proposition 3.1.** *Under Assumptions 2.1 and 2.4,*

$$\mathbb{E} \left[ \sup_{0 \leq t \leq T} W_2(L_{n,t}, \bar{L}_{n,t}) \right] \rightarrow 0,$$

as  $n \rightarrow \infty$ .

**Proposition 3.2.** *Under Assumptions 2.1, 2.2, and 2.3(iii),*

$$\mathbb{E} \left[ \sup_{0 \leq t \leq T} W_2(\tilde{L}_{n,t}, \tilde{\mu}_t) \right] \rightarrow 0,$$

as  $n \rightarrow \infty$ .

By virtue of Lemma 2.4, we have

$$\begin{aligned} \mathbb{E} \left[ \sup_{0 \leq t \leq T} W_2^2(L_{n,t}, \tilde{L}_{n,t}) \right] &\leq \mathbb{E} \left[ \sup_{0 \leq t \leq T} \frac{1}{n} \sum_{i=1}^n |X_{\frac{i}{n}}(t) - X_i^n(t)|^2 \right] \\ &\leq \mathbb{E} \left[ \frac{1}{n} \sum_{i=1}^n \|X_{\frac{i}{n}} - X_i^n\|_{\star, T}^2 \right] \rightarrow 0, \quad \text{as } n \rightarrow \infty. \end{aligned}$$

Combining the last convergence with Propositions 3.1 and 3.2, we immediately have other convergences of the expectations.

**Corollary 3.1.** *Under assumptions of Propositions 3.1, 3.2,*

$$\mathbb{E} \left[ \sup_{0 \leq t \leq T} W_2(L_{n,t}, \tilde{\mu}_t) \right] \rightarrow 0, \quad \mathbb{E} \left[ \sup_{0 \leq t \leq T} W_2(\bar{L}_{n,t}, \tilde{\mu}_t) \right] \rightarrow 0,$$

as  $n \rightarrow \infty$ .

### 3.2. Concentration around the mean

We present in this subsection the concentration of a 1-Lipschitz function of the particles  $\bar{X}^n$  around its mean, under two different norms  $\ell^1$  and  $\ell^2$ . The proofs of the results rely on the transportation inequalities presented in Section 2.5, and they will be given in Section 4.2.

From Lemma 2.6, we note that the condition (2.10) of Assumption 2.2 in Theorem 3.1 below is equivalent to the condition

$$W_1(\mu_{u,0}, v) \leq \sqrt{2\kappa H(v|\mu_{u,0})}, \quad \text{for every } u \in [0, 1], \quad v \in \mathcal{P}^1(S). \quad (3.1)$$

**Theorem 3.1.** *Under Assumptions 2.1 and 2.2, there exists a constant  $\delta > 0$ , independent of  $n$ , such that for every  $F \in \text{Lip}((C([0, T] : \mathbb{R}^d))^n, \|\cdot\|_{n,1})$  and every  $a > 0$ ,*

$$\mathbb{P} \left[ F(\bar{X}^n) - \mathbb{E}(F(\bar{X}^n)) > a \right] \leq 2 \exp \left( -\frac{\delta a^2}{n} \right) \quad (3.2)$$

holds.

We have the following result analogous to Theorem 3.1 when the condition (3.1) is replaced by (3.3). For example, if the initial law for any  $u \in [0, 1]$  takes the same form  $\mu_{u,0}(dx) = e^{-U(x)}dx$  for some  $U \in C^2(\mathbb{R}^d)$  with Hessian bounded below in semidefinite order by  $cI$  for some  $c > 0$ , then  $\mu_{u,0}$  satisfies the condition (3.3) with  $\kappa = 1/c$ . In particular, if  $\mu_{u,0}$  has the standard normal distribution on  $\mathbb{R}^d$ , then (3.3) holds with  $\kappa = 1$  (however, the initial law may not necessarily be the same for every  $u \in [0, 1]$  in the statement of Theorem 3.2).

We note here that the positive constant  $\delta$  which appears in Theorems 3.1–3.6 of this section depends only on the constants  $c$  of Lemmas 2.6 and 2.7,  $\kappa$  of (3.1) and (3.3), the functions

$\phi$ ,  $\psi$ , and  $T$ , but not on  $n$ , the number of particles. We also emphasize that the concentration inequality (3.4) is *dimension-free*; the bound on the right-hand side does not depend on  $n$ . This property will play an essential role in deriving the exponential concentration of the empirical measures in terms of  $W_2$ -distance.

**Theorem 3.2.** *Suppose that the initial particles  $\{X_u(0)\}_{u \in [0,1]}$  are independent, with law  $\mu_{u,0} \in \mathcal{P}(\mathbb{R}^d)$  satisfying, for some  $\kappa > 0$ ,*

$$W_2(\mu_{u,0}, \nu) \leq \sqrt{2\kappa H(\nu|\mu_{u,0})}, \quad \text{for every } u \in [0, 1], \quad \nu \in \mathcal{P}^2(S). \quad (3.3)$$

*Under Assumption 2.1, there exists a constant  $\delta > 0$ , independent of  $n$ , such that for every  $F \in \text{Lip}\left(\left(C([0, T] : \mathbb{R}^d)\right)^n, \|\cdot\|_{n,2}\right)$  and every  $a > 0$ ,*

$$\mathbb{P}\left[F(\bar{X}^n) - \mathbb{E}(F(\bar{X}^n)) > a\right] \leq 2 \exp(-\delta a^2) \quad (3.4)$$

*holds.*

### 3.3. Concentration toward the graphon system

Recalling the notation in (1.4), we now provide the concentration, in terms of the (1- and 2-) Wasserstein distance, of the empirical measures of the finite particle systems toward the averaged measure  $\tilde{\mu}_t$  of the graphon system. Proofs will be given in Section 4.3.

First, we have the following result on the concentration of  $\bar{L}_{n,t}$  toward  $\tilde{\mu}_t$  in terms of the  $W_1$ -distance, due to Theorem 3.1.

**Theorem 3.3.** *Under Assumptions 2.1, 2.2, and 2.3(iii), there exist constants  $\delta > 0$ , which is independent of  $n$ , and  $N \in \mathbb{N}$  such that*

$$\mathbb{P}\left[\sup_{0 \leq t \leq T} W_1(\bar{L}_{n,t}, \tilde{\mu}_t) > a\right] \leq 2 \exp\left(-\frac{\delta a^2 n}{4}\right) \quad (3.5)$$

*holds for every  $a > 0$  and every  $n \geq N$ .*

**Remark 3.1.** Theorem 3.3 gives the same exponential bound as in Theorem 2.1 of [4]. The proof in [4] mainly focuses on computing certain sub-Gaussian estimates, whereas our argument relies on the concentration property (3.2) of the system (1.3). Applying the same argument, we can even deduce the exponential bound in terms of the  $W_2$  metric (in Theorem 3.5 below).

In Section 2.6, we introduced the concept of the  $L^1$ -Fourier class, along with Bernstein's inequality, to express  $W_1(L_n, \bar{L}_n)$  in terms of  $\|D^{(n)}\|_{\infty \rightarrow 1}$ . This gives rise to the following result on the concentration of the particle system (1.2) toward the graphon system.

**Theorem 3.4.** *Suppose that the components of the interaction function  $\phi$  belong to the  $L^1$ -Fourier class (Definition 2.1). Under Assumptions 2.1, 2.2, 2.3(iii), and 2.4, there exist constants  $\delta > 0$ , which is independent of  $n$ , and  $N \in \mathbb{N}$  such that*

$$\mathbb{P}\left[\sup_{0 \leq t \leq T} W_1(L_{n,t}, \tilde{\mu}_t) > a\right] \leq 3 \exp\left(-\frac{\delta a^2 n}{16}\right) \quad (3.6)$$

*holds for every  $a > 0$  and every  $n \geq N$ . For general interaction functions  $\phi$  (which do not necessarily belong to the  $L^1$ -Fourier class), we have instead*

$$\mathbb{P}\left[\sup_{0 \leq t \leq T} d_{BL}(L_{n,t}, \tilde{\mu}_t) > a\right] \leq 3 \exp\left(-\frac{\delta a^2 n}{16}\right). \quad (3.7)$$

The following result gives the concentration of  $\bar{L}_{n,t}$  toward  $\tilde{\mu}_t$  as in Theorem 3.3, but in terms of the  $W_2$  metric. Its proof is similar to that of Theorem 3.3, but Theorem 3.2 is used in place of Theorem 3.1.

**Theorem 3.5.** *Under Assumptions 2.1, 2.2, and 2.3, together with the condition (3.3), there exist constants  $\delta > 0$ , independent of  $n$ , and  $N \in \mathbb{N}$  such that*

$$\mathbb{P}\left[\sup_{0 \leq t \leq T} W_2(\bar{L}_{n,t}, \tilde{\mu}_t) > a\right] \leq 2 \exp\left(-\frac{\delta a^2 n}{4}\right) \quad (3.8)$$

holds for every  $a > 0$  and every  $n \geq N$ .

Since we have the exponential bound in (3.8) in the  $W_2$  metric, one naturally expects to obtain a bound similar to (3.6) in the  $W_2$  metric as well. In order to do this, we need to find the exponential bound for the probability  $\mathbb{P}[\sup_{0 \leq t \leq T} W_2(L_{n,t}, \bar{L}_{n,t}) > a]$ , which requires us to handle the quantity  $\|(D^{(n)})^\top D^{(n)}\|_{\infty \rightarrow 1}$ , instead of  $\|D^{(n)}\|_{\infty \rightarrow 1}$  as in the proof of Theorem 3.4. The control of this quantity is achieved in Lemma 3.1 under an extra condition on the sparsity parameter  $p(n)$ , a more restrictive condition than the one in Assumption 2.4.

**Assumption 3.1.** *The sparsity parameter sequence  $\{p(n)\}_{n \in \mathbb{N}} \subset (0, 1]$  of the system (1.2) satisfies one of the following:*

(i)  $p(n) \rightarrow 0$  and  $np(n)^2 \rightarrow \infty$  as  $n \rightarrow \infty$ , or

(ii)  $p(n) \equiv 1$  for every  $n \in \mathbb{N}$ .

Recalling the notation of (2.5), we state the following lemma, which is needed in proving Theorem 3.6. Its proof, given in Section 4.3, is similar to that of Lemma 2.9, but requires more involved applications of Bernstein's inequality.

**Lemma 3.1.** *Under Assumption 3.1, there exists  $N \in \mathbb{N}$  such that*

$$\mathbb{P}\left[\frac{\|(D^{(n)})^\top D^{(n)}\|_{\infty \rightarrow 1}}{n} > \eta\right] \leq 3n^2 \exp\left(-\frac{2\eta^2 np(n)^4}{9 + 4\eta}\right)$$

holds for every  $n \geq N$  and  $\eta > 0$ .

**Theorem 3.6.** *Suppose that the components of the interaction function  $\phi$  belong to the  $L^1$ -Fourier class. Under Assumptions 2.1, 2.2, 2.3, and 3.1(i), together with the condition (3.3), there exist constants  $K > 0$ , which is independent of  $n$ , and  $N \in \mathbb{N}$  such that*

$$\mathbb{P}\left[\sup_{0 \leq t \leq T} W_2(L_{n,t}, \tilde{\mu}_t) > a\right] \leq 4n^2 \exp\left(-\frac{a^4 np(n)^4}{72K^2 + 8a^2 K}\right) \quad (3.9)$$

holds for every  $a > 0$  and every  $n \geq N$ .

Furthermore, if Assumption 3.1(i) is replaced by Assumption 3.1(ii), we have the exponential bound in  $n$ : there exist constants  $\delta > 0$ , which is independent of  $n$ , and  $N \in \mathbb{N}$  such that

$$\mathbb{P}\left[\sup_{0 \leq t \leq T} W_2(L_{n,t}, \tilde{\mu}_t) > a\right] \leq \exp\left(-\frac{\delta a^4 n}{a^2 + \delta}\right) \quad (3.10)$$

holds for every  $a > 0$  and every  $n \geq N$ .

## 4. Proofs

In this section we provide the proofs of the results stated in Section 3.

### 4.1. Proofs of results in Section 3.1

4.1.1. *Proof of Proposition 3.1.* Let us recall the identity (4.7), along with the notation in (2.5). By Hölder's inequality there exists  $K > 0$ , depending on  $\phi$  and  $\psi$ , such that for every  $t \in [0, T]$ ,

$$\begin{aligned} \sup_{0 \leq t \leq T} |X_i^n(t) - \bar{X}_i^n(t)|^2 &\leq KT \int_0^T \left| \sum_{j=1}^n D_{i,j}^{(n)} \phi(X_i^n(s), X_j^n(s)) \right|^2 ds \\ &\quad + KT \int_0^T \left| \sum_{j=1}^n \bar{P}_{i,j}^{(n)} (\phi(X_i^n(s), X_j^n(s)) - \phi(\bar{X}_i^n(s), \bar{X}_j^n(s))) \right|^2 ds \\ &\quad + KT \int_0^T |X_i^n(s) - \bar{X}_i^n(s)|^2 ds. \end{aligned}$$

Taking the expectation of the first term, and using the independence of  $\{D_{i,j}^{(n)}\}_{j \in [n]}$  and the boundedness of  $\phi$ , we have

$$KT \mathbb{E} \int_0^T \left| \sum_{j=1}^n D_{i,j}^{(n)} \phi(X_i^n(s), X_j^n(s)) \right|^2 ds \leq KT \int_0^T \sum_{j=1}^n \mathbb{E}[(D_{i,j}^{(n)})^2] ds \leq \frac{KT^2}{np(n)}.$$

For the second term, Hölder's inequality and the Lipschitz continuity of  $\phi$  give

$$\begin{aligned} KT \mathbb{E} \int_0^T \left| \sum_{j=1}^n \bar{P}_{i,j}^{(n)} (\phi(X_i^n(s), X_j^n(s)) - \phi(\bar{X}_i^n(s), \bar{X}_j^n(s))) \right|^2 ds \\ \leq KT \mathbb{E} \int_0^T \frac{1}{n} \sum_{j=1}^n (|X_i^n(s) - \bar{X}_i^n(s)|^2 + |X_j^n(s) - \bar{X}_j^n(s)|^2) ds. \end{aligned}$$

Combining above inequalities and averaging over  $i \in [n]$ , we obtain

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n \mathbb{E} \left[ \sup_{0 \leq t \leq T} |X_i^n(t) - \bar{X}_i^n(t)|^2 \right] \\ \leq \frac{KT^2}{np(n)} + KT \int_0^T \frac{1}{n} \sum_{i=1}^n \mathbb{E} \left[ \sup_{0 \leq u \leq s} |X_i^n(u) - \bar{X}_i^n(u)|^2 \right] ds. \end{aligned}$$

Grönwall's inequality yields

$$\frac{1}{n} \sum_{i=1}^n \mathbb{E} \left[ \sup_{0 \leq t \leq T} |X_i^n(t) - \bar{X}_i^n(t)|^2 \right] \leq \frac{KT^2 \exp(KT^2)}{np(n)},$$

and thus

$$\begin{aligned} \mathbb{E} \left[ \sup_{0 \leq t \leq T} W_2^2(L_{n,t}, \bar{L}_{n,t}) \right] &\leq \mathbb{E} \left[ \sup_{0 \leq t \leq T} \frac{1}{n} \sum_{i=1}^n |X_i^n(t) - \bar{X}_i^n(t)|^2 \right] \\ &\leq \frac{1}{n} \sum_{i=1}^n \mathbb{E} \left[ \sup_{0 \leq t \leq T} |X_i^n(t) - \bar{X}_i^n(t)|^2 \right] \leq \frac{KT^2 \exp(KT^2)}{np(n)} \rightarrow 0, \text{ as } n \rightarrow \infty. \end{aligned}$$

4.1.2. *Proof of Proposition 3.2.* We divide the interval  $[0, T]$  into  $M := \lceil \frac{T}{\Delta} \rceil$  subintervals of length  $\Delta > 0$ :

$$[0, T] = [0, \Delta] \cup [\Delta, 2\Delta] \cup \dots \cup [(M-1)\Delta, T] =: \cup_{h=1}^M \Delta_h,$$

where  $\Delta_h := [(h-1)\Delta, h\Delta]$  for  $h = 1, \dots, M-1$  and  $\Delta_M = [(M-1)\Delta, T]$ . (We choose the value of  $\Delta$  later.) With the notation

$$\tilde{\mu}_{n,t} := \frac{1}{n} \sum_{i=1}^n \mu_{\frac{i}{n}, t} = \frac{1}{n} \sum_{i=1}^n \mathcal{L}(X_{\frac{i}{n}}(t)),$$

the triangle inequality gives

$$\begin{aligned} \mathbb{E} \left[ \sup_{0 \leq t \leq T} W_2(\tilde{L}_{n,t}, \tilde{\mu}_t) \right] &= \mathbb{E} \left[ \max_{h \in [M]} \sup_{t \in \Delta_h} W_2(\tilde{L}_{n,t}, \tilde{\mu}_t) \right] \\ &\leq \mathbb{E} \left[ \max_{h \in [M]} \sup_{t \in \Delta_h} W_2(\tilde{L}_{n,t}, \tilde{L}_{n,(h-1)\Delta}) \right] + \mathbb{E} \left[ \sup_{h \in [M]} W_2(\tilde{L}_{n,(h-1)\Delta}, \tilde{\mu}_{n,(h-1)\Delta}) \right] \\ &\quad + \mathbb{E} \left[ \sup_{h \in [M]} W_2(\tilde{\mu}_{n,(h-1)\Delta}, \tilde{\mu}_{(h-1)\Delta}) \right] + \mathbb{E} \left[ \sup_{h \in [M]} \sup_{t \in \Delta_h} W_2(\tilde{\mu}_{(h-1)\Delta}, \tilde{\mu}_t) \right]. \\ &=: E_1 + E_2 + E_3 + E_4. \end{aligned}$$

For the first term,  $E_1$ , we note that there exists  $K > 0$ , depending on the bounds of  $\phi$ ,  $\psi$ , and  $\sigma$ , such that

$$\left| X_{\frac{i}{n}}(s) - X_{\frac{i}{n}}(u) \right|^2 \leq K|s-u|^2 + K \left| B_{\frac{i}{n}}(s) - B_{\frac{i}{n}}(u) \right|^2$$

holds for every  $0 \leq u \leq s \leq T$ , and thus we have

$$\begin{aligned} (E_1)^2 &\leq \mathbb{E} \left[ \max_{h \in [M]} \sup_{t \in \Delta_h} W_2^2(\tilde{L}_{n,t}, \tilde{L}_{n,(h-1)\Delta}) \right] \\ &\leq \mathbb{E} \left[ \max_{h \in [M]} \sup_{t \in \Delta_h} \frac{1}{n} \sum_{i=1}^n \left| X_{\frac{i}{n}}(t) - X_{\frac{i}{n}}((h-1)\Delta) \right|^2 \right] \\ &\leq \mathbb{E} \left[ \max_{h \in [M]} \sup_{t \in \Delta_h} \frac{1}{n} \sum_{i=1}^n \left( K\Delta^2 + K \left| B_{\frac{i}{n}}(t) - B_{\frac{i}{n}}((h-1)\Delta) \right|^2 \right) \right] \\ &\leq K\Delta^2 + K \mathbb{E} \left[ \max_{h \in [M]} \sup_{t \in \Delta_h} \frac{1}{n} \sum_{i=1}^n \left| B_{\frac{i}{n}}(t) - B_{\frac{i}{n}}((h-1)\Delta) \right|^2 \right]. \end{aligned}$$

Applying Hölder's inequality twice, we find that the last expectation is bounded above by

$$\begin{aligned} & \mathbb{E} \left[ \max_{h \in [M]} \sup_{t \in \Delta_h} \left( \frac{1}{n} \sum_{i=1}^n |B_{\frac{i}{n}}(t) - B_{\frac{i}{n}}((h-1)\Delta)|^4 \right)^{\frac{1}{2}} \right] \\ & \leq \sqrt{\mathbb{E} \left[ \max_{h \in [M]} \sup_{t \in \Delta_h} \frac{1}{n} \sum_{i=1}^n |B_{\frac{i}{n}}(t) - B_{\frac{i}{n}}((h-1)\Delta)|^4 \right]} \\ & \leq \sqrt{\sum_{h \in [M]} \frac{1}{n} \sum_{i=1}^n \mathbb{E} \left[ \sup_{t \in \Delta_h} |B_{\frac{i}{n}}(t) - B_{\frac{i}{n}}((h-1)\Delta)|^4 \right]} \leq \sqrt{MC_4 \mathbb{E}[\Delta^2]} \leq \sqrt{(T+1)C_4\Delta}. \end{aligned}$$

The second-to-last inequality uses the properties of the increments of Brownian motion and the Burkholder–Davis–Gundy inequality with the positive constant  $C_4$ . Therefore, we have the bound

$$(E_1)^2 \leq K\Delta^2 + K\sqrt{(T+1)C_4\Delta}.$$

For the second expectation,  $E_2$ , a series of applications of Hölder's inequality and Lemma 2.8 give

$$\begin{aligned} E_2 & \leq \sum_{h \in [M]} \mathbb{E}[W_2(\tilde{L}_{n,(h-1)\Delta}, \tilde{\mu}_{n,(h-1)\Delta})] \leq \sum_{h \in [M]} \mathbb{E}[W_3(\tilde{L}_{n,(h-1)\Delta}, \tilde{\mu}_{n,(h-1)\Delta})] \\ & \leq \sum_{h \in [M]} \left( \mathbb{E}[W_3^3(\tilde{L}_{n,(h-1)\Delta}, \tilde{\mu}_{n,(h-1)\Delta})] \right)^{1/3} \\ & \leq K \sum_{h \in [M]} \left( \int_{\mathbb{R}^d} |x|^4 \tilde{\mu}_{n,(h-1)\Delta}(dx) \right)^{1/4} \alpha_{3,4}^{1/3}(n) \leq KM\alpha_{3,4}^{1/3}(n) \rightarrow 0, \end{aligned}$$

as  $n \rightarrow \infty$ , where the last inequality follows from Lemma 2.2.

On the other hand, by the convexity of  $W_2^2(\cdot, \cdot)$  and Lemma 2.3(ii), there exists  $K > 0$  satisfying

$$\begin{aligned} E_3^2 & \leq \mathbb{E} \left[ \sup_{h \in [M]} W_2^2(\tilde{\mu}_{n,(h-1)\Delta}, \tilde{\mu}_{(h-1)\Delta}) \right] \\ & \leq \mathbb{E} \left[ \sup_{h \in [M]} \int_0^1 W_2^2(\mu_{\lceil nu \rceil}, \mu_{u,(h-1)\Delta}) du \right] \leq \frac{K}{n^2}. \end{aligned}$$

Finally, for the last term,  $E_4$ , we note from a straightforward computation that there exists  $K > 0$  satisfying

$$\mathbb{E}|X_u(t) - X_u(s)|^2 \leq K|t-s|^2 + K\mathbb{E}|B_u(t) - B_u(s)|^2 \leq K|t-s|^2$$

for every  $u \in [0, 1]$  and  $s, t \in [0, T]$  satisfying  $|t-s| \leq 1$ . Thus, we have

$$\begin{aligned} E_4^2 & \leq \sup_{h \in [M]} \sup_{t \in \Delta_h} \int_0^1 W_2^2(\mu_{u,(h-1)\Delta}, \mu_{u,t}) du \\ & \leq \sup_{h \in [M]} \sup_{t \in \Delta_h} \int_0^1 \mathbb{E}|X_u((h-1)\Delta) - X_u(t)|^2 du \leq K\Delta. \end{aligned}$$

Let us combine all the bounds from  $E_1$  to  $E_4$ . For any given  $\epsilon > 0$ , we can choose  $\Delta$  small enough so that  $E_1 + E_4 < \epsilon/2$ . Then we can choose  $N \in \mathbb{N}$  large enough so that  $E_2 + E_3 < \epsilon/2$  for every  $n \geq N$ , which implies  $\mathbb{E}[\sup_{0 \leq t \leq T} W_2(\tilde{L}_{n,t}, \tilde{\mu}_t)] < \epsilon$  for every  $n \geq N$ .

#### 4.2. Proofs of results in Section 3.2

4.2.1. *Proof of Theorem 3.1.* Let us fix an arbitrary  $n \in \mathbb{N}$ . We shall naturally identify elements of  $(\mathbb{R}^d)^n$  with those of  $\mathbb{R}^{dn}$ , and elements of  $(C([0, T] : \mathbb{R}^d))^n$  with those of  $C([0, T] : (\mathbb{R}^d)^n)$ ; we shall specify which norm we use for each space. We can express the SDE (1.3) in the form of (2.12) with  $k = dn$ , by making the following definitions:

- (i)  $(\mathbb{R}^d)^n \ni x = (x_i)_{i \in [n]}$ , where  $x_i = X_{i/n}(0)$ ;
- (ii)  $C([0, T] : (\mathbb{R}^d)^n) \ni X^x = (\bar{X}_i^n)_{i \in [n]}$ , where  $\bar{X}_i^n = (\bar{X}_{i,k}^n)_{k \in [d]}$ ;
- (iii)  $b : [0, T] \times C([0, T] : (\mathbb{R}^d)^n) \rightarrow (\mathbb{R}^d)^n$  is such that  $b = (b_i)_{i \in [n]}$ ,  $b_i = (b_{i,k})_{k \in [d]}$ , where

$$b_{i,k}(t, X^x) = \frac{1}{n} \sum_{j=1}^n \phi_k(\bar{X}_i^n(t), \bar{X}_j^n(t)) G\left(\frac{i}{n}, \frac{j}{n}\right) + \psi_k(\bar{X}_i^n(t));$$

- (iv)  $W = (W_i)_{i \in [n]}$  is a  $(dn)$ -dimensional Brownian motion, where  $W_i \equiv B_{i/n}$ ;
- (v)  $\Sigma$  is a block-diagonal  $(dn) \times (dn)$  matrix with block diagonal entries  $\sigma$ .

In order to apply Lemma 2.5, it suffices to check the condition (2.13): for any  $X, Y \in C([0, T] : \mathbb{R}^{dn})$ , Hölder's inequality and the Lipschitz continuity of  $\phi, \psi$  indeed yield, for every  $t \in [0, T]$ ,

$$\begin{aligned} & |b(t, X) - b(t, Y)|^2 \\ &= \sum_{i=1}^n \sum_{k=1}^d \left| \frac{1}{n} \sum_{j=1}^n (\phi_k(X_i(t), X_j(t)) - \phi_k(Y_i(t), Y_j(t))) G\left(\frac{i}{n}, \frac{j}{n}\right) + \psi_k(X_i(t)) - \psi_k(Y_i(t)) \right|^2 \\ &\leq 2 \sum_{i=1}^n \sum_{k=1}^d \left[ \left| \frac{1}{n} \sum_{j=1}^n (\phi_k(X_i(t), X_j(t)) - \phi_k(Y_i(t), Y_j(t))) G\left(\frac{i}{n}, \frac{j}{n}\right) \right|^2 \right. \\ &\quad \left. + \left| \psi_k(X_i(t)) - \psi_k(Y_i(t)) \right|^2 \right] \\ &\leq 2 \sum_{i=1}^n \left[ K^2 |X_i(t) - Y_i(t)|^2 + \sum_{k=1}^d \frac{1}{n} \sum_{j=1}^n \left| \phi_k(X_i(t), X_j(t)) - \phi_k(Y_i(t), Y_j(t)) \right|^2 \right] \\ &\leq (4K^2 + 1) \sum_{i=1}^n |X_i(t) - Y_i(t)|^2 \leq (4K^2 + 1) \|X - Y\|_{\star, t}^2. \end{aligned}$$

Let  $P^x \in \mathcal{P}(C([0, T] : \mathbb{R}^{dn}))$  be the law of the solution of (1.3) in the notation of (1)–(5) above; then, from Lemma 2.5, for any  $Q \in \mathcal{P}(C([0, T] : \mathbb{R}^{dn}))$  we have

$$W_{1, (C([0, T] : \mathbb{R}^{dn}), \|\cdot\|_{dn,2})}(P^x, Q) \leq \sqrt{2c_1 H(Q|P^x)},$$

for some  $c_1 > 0$ .

For an arbitrary  $F \in Lip((C([0, T] : \mathbb{R}^d))^n, \|\cdot\|_{n,1})$ , Hölder's inequality shows that  $F$  is a  $\sqrt{n}$ -Lipschitz function on the space  $(C([0, T] : \mathbb{R}^{dn}), \|\cdot\|_{dn,2})$ ; indeed, for  $X, Y \in ((C([0, T] : \mathbb{R}^d))^n, \|\cdot\|_{n,1})$  we obtain

$$\begin{aligned} |F(X) - F(Y)| &\leq \|X - Y\|_{n,1} = \sum_{i=1}^d \|X_i - Y_i\|_{\star,T} \leq \sqrt{n \sum_{i=1}^d \|X_i - Y_i\|_{\star,T}^2} \\ &\leq \sqrt{n \sum_{i=1}^d \sum_{k=1}^d \|X_{i,k} - Y_{i,k}\|_{\star,T}^2} = \sqrt{n} \|X - Y\|_{dn,2}. \end{aligned}$$

Thus, Lemma 2.6 implies

$$P^x(F - \langle P^x, F \rangle > a) \leq \exp\left(-\frac{a^2}{2c_1 n}\right), \quad (4.1)$$

for any  $a > 0$ .

We now claim that there exists a positive constant  $c_2$ , which does not depend on  $n$ , such that the map  $x \mapsto \langle P^x, F \rangle$  is  $c_2$ -Lipschitz on  $(\mathbb{R}^d)^n$  with respect to the Euclidean  $\ell^p$  norm for any  $F \in Lip((C([0, T] : \mathbb{R}^d))^n, \|\cdot\|_{n,p})$  and for any  $p = 1, 2$ . Given any  $x, y \in (\mathbb{R}^d)^n$ , we couple  $P^x$  and  $P^y$  by solving the system (1.3) from the two initial states  $x, y$  with the same Brownian motion, and denote the coupling by  $\pi_{x,y}$ . We deduce that for  $\mathcal{L}(X) = P^x$  and  $\mathcal{L}(Y) = P^y$ ,

$$|\langle P^x, F \rangle - \langle P^y, F \rangle|^p \leq \int |F(X) - F(Y)|^p \pi_{x,y}(dX, dY) \leq \int \|X - Y\|_{n,p}^p \pi_{x,y}(dX, dY). \quad (4.2)$$

When  $p = 2$ , we use a standard argument (the trivial inequality  $(a + b)^2 \leq 2(a^2 + b^2)$ , the Lipschitz continuity from Assumption 2.1(i), and a series of applications of Hölder's inequality) to derive

$$\begin{aligned} &\sum_{i=1}^n |X_i(t) - Y_i(t)|^2 \\ &\leq 2\|x - y\|_{n,2}^2 + 2Kt \sum_{i=1}^n \int_0^t \left( |X_i(s) - Y_i(s)| + \frac{1}{n} \sum_{j=1}^n |X_j(s) - Y_j(s)| \right)^2 ds \\ &\leq 2\|x - y\|_{n,2}^2 + 8Kt \int_0^t \sum_{i=1}^n |X_i(s) - Y_i(s)|^2 ds, \quad \forall t \in [0, T]. \end{aligned}$$

Grönwall's inequality yields that the last integrand in (4.2) for  $p = 2$  is bounded by

$$\|X - Y\|_{n,2}^2 \leq c_2^2 \|x - y\|_{n,2}^2$$

for some constant  $c_2 > 0$ , which depends on  $\phi$ ,  $\psi$ , and  $T$ , but not on  $n$ . When  $p = 1$ , proving  $\|X - Y\|_{n,1} \leq c_2\|x - y\|_{n,1}$  is easier, and the claim follows.

On the other hand, we apply Lemmas 2.7, 2.6 to the assumption (3.1) to obtain, for every  $f \in Lip((\mathbb{R}^d)^n, \|\cdot\|_{n,1})$  and for any  $a > 0$ ,

$$\mu_0^n(f - \langle \mu_0^n, f \rangle > a) \leq \exp\left(-\frac{a^2}{2\kappa n}\right). \quad (4.3)$$

We conclude from (4.1), the above claim, and (4.3) that

$$\begin{aligned} \mathbb{P}\left[F(\bar{X}^n) - \mathbb{E}(F(\bar{X}^n)) > a\right] &\leq \mathbb{E}\left[\mathbb{P}\left(F(\bar{X}^n) - \langle P^x, F \rangle > \frac{a}{2} \mid \bar{X}^n(0) = x\right)\right] \\ &\quad + \mathbb{P}\left(\langle P^{\bar{X}^n(0)}, F \rangle - \mathbb{E}[\langle P^{\bar{X}^n(0)}, F \rangle] > \frac{a}{2}\right) \\ &\leq \exp\left(-\frac{a^2}{8c_1 n}\right) + \exp\left(-\frac{a^2}{8\kappa c_2^2 n}\right). \end{aligned}$$

The assertion (3.4) follows from choosing  $1/\delta = 8 \max(c_1, \kappa c_2^2)$ .

**4.2.2. Proof of Theorem 3.2.** We follow the proof of Theorem 3.1. Identifying the elements of  $(C([0, T] : \mathbb{R}^d))^n$  with those of  $C([0, T] : \mathbb{R}^{dn})$ , expressing the SDE (1.3) in the form of (2.12), and applying Lemma 2.5, we have that there exists a positive constant  $c_1 > 0$  such that

$$W_{1, (C([0, T] : \mathbb{R}^{dn}), \|\cdot\|_{dn,2})}(P^x, Q) \leq \sqrt{2c_1 H(Q|P^x)}$$

holds for any  $Q \in \mathcal{P}(C([0, T] : \mathbb{R}^{dn}))$ . Here,  $P^x$  is the law of the solution of (1.3). Moreover, Lemma 2.6 implies

$$P^x(F - \langle P^x, F \rangle > a) \leq \exp\left(-\frac{a^2}{2c_1}\right), \quad (4.4)$$

for any  $a > 0$  and every  $F \in Lip(C([0, T] : \mathbb{R}^{dn}), \|\cdot\|_{dn,2})$ . It is easy to check that every function in  $Lip(C([0, T] : \mathbb{R}^{dn}), \|\cdot\|_{dn,2})$  also belongs to  $Lip((C([0, T] : \mathbb{R}^d))^n, \|\cdot\|_{n,2})$ ; thus the inequality (4.4) also holds for every  $F \in Lip((C([0, T] : \mathbb{R}^d))^n, \|\cdot\|_{n,2})$ .

We now apply Lemmas 2.7(ii) and 2.6 to the assumption (3.3) to deduce

$$\mu_0^n(f - \langle \mu_0^n, f \rangle > a) \leq \exp\left(-\frac{a^2}{2\kappa}\right), \quad (4.5)$$

for every  $f \in Lip((\mathbb{R}^d)^n, \|\cdot\|_{n,2})$  and for any  $a > 0$ .

From (4.4), (4.5), and the claim in the proof of Theorem 3.1, we conclude that

$$\begin{aligned} \mathbb{P}\left[F(\bar{X}^n) - \mathbb{E}(F(\bar{X}^n)) > a\right] &\leq \mathbb{E}\left[\mathbb{P}\left(F(\bar{X}^n) - \langle P^x, F \rangle > \frac{a}{2} \mid \bar{X}^n(0) = x\right)\right] \\ &\quad + \mathbb{P}\left[\langle P^{\bar{X}^n(0)}, F \rangle - \mathbb{E}[\langle P^{\bar{X}^n(0)}, F \rangle] > \frac{a}{2}\right] \\ &\leq \exp\left(-\frac{a^2}{8c_1}\right) + \exp\left(-\frac{a^2}{8\kappa c_2^2}\right). \end{aligned}$$

The result (3.4) follows from choosing  $1/\delta = 8 \max(c_1, \kappa c_2^2)$ .

### 4.3. Proofs of results in Section 3.3

4.3.1. *Proof of Theorem 3.3.* First, we claim that

$$Y \mapsto \sup_{0 \leq t \leq T} W_1 \left( \frac{1}{n} \sum_{i=1}^n \delta_{Y_i(t)}, \tilde{\mu}_t \right)$$

is  $(1/n)$ -Lipschitz from  $(C([0, T] : \mathbb{R}^d))^n, \|\cdot\|_{n,1}$  to  $\mathbb{R}$ . For  $Y, Z \in (C([0, T] : \mathbb{R}^d))^n$ , applying the triangle inequality for the Wasserstein metric and taking the supremum over  $t \in [0, T]$  gives the inequality

$$\begin{aligned} & \left| \sup_{0 \leq t \leq T} W_1 \left( \frac{1}{n} \sum_{i=1}^n \delta_{Y_i(t)}, \tilde{\mu}_t \right) - \sup_{0 \leq t \leq T} W_1 \left( \frac{1}{n} \sum_{i=1}^n \delta_{Z_i(t)}, \tilde{\mu}_t \right) \right| \\ & \leq \sup_{0 \leq t \leq T} W_1 \left( \frac{1}{n} \sum_{i=1}^n \delta_{Y_i(t)}, \frac{1}{n} \sum_{i=1}^n \delta_{Z_i(t)} \right). \end{aligned}$$

From the definitions (2.1) and (2.2), the last expression is less than or equal to  $\frac{1}{n} \sum_{i=1}^n \|Y_i - Z_i\| = \frac{1}{n} \|Y - Z\|_{n,1}$ , and the claim follows.

Then, for any  $a > 0$ ,

$$\begin{aligned} \mathbb{P} \left[ \sup_{0 \leq t \leq T} W_1(\bar{L}_{n,t}, \tilde{\mu}_t) > a \right] & \leq \mathbb{P} \left[ \sup_{0 \leq t \leq T} W_1(\bar{L}_{n,t}, \tilde{\mu}_t) - \mathbb{E} \left[ \sup_{0 \leq t \leq T} W_1(\bar{L}_{n,t}, \tilde{\mu}_t) \right] > \frac{a}{2} \right] \\ & \quad + \mathbb{P} \left[ \mathbb{E} \left[ \sup_{0 \leq t \leq T} W_1(\bar{L}_{n,t}, \tilde{\mu}_t) \right] > \frac{a}{2} \right]. \end{aligned}$$

The first term is bounded by the right-hand side of (3.5) from Theorem 3.1.

Let us consider the auxiliary particle system (1.2) satisfying Assumption 2.4. Corollary 3.1 shows that the last probability vanishes for all but finitely many  $n$ , and the result follows.

4.3.2. *Proof of Theorem 3.4.* We first prove (3.6). From the triangle inequality, we obtain

$$\mathbb{P} \left[ \sup_{0 \leq t \leq T} W_1(L_{n,t}, \tilde{\mu}_t) > a \right] \leq \mathbb{P} \left[ \sup_{0 \leq t \leq T} W_1(\bar{L}_{n,t}, \tilde{\mu}_t) > \frac{a}{2} \right] + \mathbb{P} \left[ \sup_{0 \leq t \leq T} W_1(L_{n,t}, \bar{L}_{n,t}) > \frac{a}{2} \right]. \quad (4.6)$$

In what follows, we compute the bound for the last probability on the right-hand side. For fixed  $t \in [0, T]$  and  $i \in [n]$ , we use the notation (2.5) to obtain

$$\begin{aligned} X_i^n(t) - \bar{X}_i^n(t) &= \int_0^t \sum_{j=1}^n D_{i,j}^{(n)} \phi(X_i^n(s), X_j^n(s)) ds \\ &+ \int_0^t \sum_{j=1}^n \bar{P}_{i,j}^{(n)} \left( \phi(X_i^n(s), X_j^n(s)) - \phi(\bar{X}_i^n(s), \bar{X}_j^n(s)) \right) + \psi(X_i^n(s)) - \psi(\bar{X}_i^n(s)) ds. \end{aligned} \quad (4.7)$$

We define  $\Delta(t) := \frac{1}{n} \sum_{i=1}^n ||X_i^n - \bar{X}_i^n||_{\star, t}$ , then deduce from the continuity of  $X_i^n(\cdot) - \bar{X}_i^n(\cdot)$  that there exists  $t_i \in [0, t]$  for each  $i \in [n]$  satisfying

$$\Delta(t) = \frac{1}{n} \sum_{i=1}^n |X_i^n(t_i) - \bar{X}_i^n(t_i)| \leq \int_0^t \frac{1}{n} \sum_{i,j=1}^n \left| D_{i,j}^{(n)} \mathbb{1}_{[0,t_i]}(s) \phi(X_i^n(s), X_j^n(s)) \right| ds \quad (4.8)$$

$$+ \frac{1}{n} \int_0^t \sum_{i,j=1}^n \bar{P}_{i,j}^{(n)} \mathbb{1}_{[0,t_i]}(s) \left| \phi(X_i^n(s), X_j^n(s)) - \phi(\bar{X}_i^n(s), \bar{X}_j^n(s)) \right| ds \quad (4.9)$$

$$+ \frac{1}{n} \int_0^t \sum_{i=1}^n \mathbb{1}_{[0,t_i]}(s) \left| \psi(X_i^n(s)) - \psi(\bar{X}_i^n(s)) \right| ds. \quad (4.10)$$

Since each component  $\phi_k$  of  $\phi$  belongs to the  $L^1$ -Fourier class, there exists a finite complex measure  $m_{\phi_k}$  such that we can write, for every  $k \in [d]$ ,

$$\phi_k(X_i^n(s), X_j^n(s)) = \int_{\mathbb{R}^{2d}} a_i^k(z, s) b_j^k(z, s) m_{\phi_k}(dz), \quad z = (z_1, z_2), \quad (4.11)$$

for some complex functions  $a_i^k, b_j^k$  of the form

$$a_i^k(z, s) := \exp(2\pi\sqrt{-1}\langle X_i^n(s), z_1 \rangle), \quad b_j^k(z, s) := \exp(2\pi\sqrt{-1}\langle X_j^n(s), z_2 \rangle).$$

Using the representation (4.11) with the elementary inequality

$$|\phi(X_i^n(s), X_j^n(s))| \leq \sqrt{d} \max_{1 \leq k \leq d} |\phi_k(X_i^n(s), X_j^n(s))|,$$

we find that the integral of (4.8) is bounded above by

$$\begin{aligned} \max_{1 \leq k \leq d} \frac{\sqrt{d}}{n} \int_0^t \int_{\mathbb{R}^{2d}} \sum_{i,j=1}^n \left| D_{i,j}^{(n)} \mathbb{1}_{[0,t_i]}(s) a_i^k(z, s) b_j^k(z, s) \right| m_{\phi_k}(dz) ds \\ = \max_{1 \leq k \leq d} \frac{\sqrt{d}}{n} \int_0^t \int_{\mathbb{R}^{2d}} \left| \langle \mathbf{a}^k(z, s), D^{(n)} \mathbf{b}^k(z, s) \rangle \right| m_{\phi_k}(dz) ds, \end{aligned} \quad (4.12)$$

where we define the complex vectors

$$\mathbf{a}^k(z, s) := \left( \mathbb{1}_{[0,t_i]}(s) a_i^k(z, s) \right)_{i \in [n]}, \quad \mathbf{b}^k(z, s) := (b_j^k(z, s))_{j \in [n]}, \quad \text{for each } k \in [d]. \quad (4.13)$$

Since the  $\ell^\infty$  norms of these vectors are bounded by 1, decomposing them into real and complex parts gives, for each  $k \in [d]$ ,

$$\left| \langle \mathbf{a}^k(z, s), D^{(n)} \mathbf{b}^k(z, s) \rangle \right| \leq 4 \sup \left\{ \langle \mathbf{x}, D^{(n)} \mathbf{y} \rangle : \mathbf{x}, \mathbf{y} \in [-1, 1]^n \right\} = 4 ||D^{(n)}||_{\infty \rightarrow 1}.$$

Thus, the right-hand side of (4.12) is bounded above by

$$\frac{4t\sqrt{d}}{n} \left( \max_{1 \leq k \leq d} ||m_{\phi_k}||_{TM} \right) ||D^{(n)}||_{\infty \rightarrow 1}.$$

For the integrals of (4.9) and (4.10), we use the Lipschitz continuity of  $\phi$  and  $\psi$ ; thus, there exists a constant  $K > 0$  such that

$$\Delta(t) \leq K \int_0^t \Delta(s) ds + \frac{\|D^{(n)}\|_{\infty \rightarrow 1}}{n} Kt, \quad \forall t \in [0, T]. \quad (4.14)$$

Grönwall's inequality yields

$$\Delta(T) \leq \frac{K\|D^{(n)}\|_{\infty \rightarrow 1}}{n}, \quad (4.15)$$

where  $K$  is now a positive constant depending on the time horizon  $T$ . Recalling the notation  $\Delta(t)$ , we obtain

$$\sup_{0 \leq t \leq T} W_1(L_{n,t}, \bar{L}_{n,t}) \leq \Delta(T) \leq \frac{K\|D^{(n)}\|_{\infty \rightarrow 1}}{n},$$

and finally Lemma 2.9 gives the bound for the last probability of (4.6),

$$\mathbb{P}\left[\sup_{0 \leq t \leq T} W_1(L_{n,t}, \bar{L}_{n,t}) > \frac{a}{2}\right] \leq \exp\left(-\frac{a^2 n^2 p(n)}{8K^2 + 2aK/3}\right). \quad (4.16)$$

For the first probability on the right-hand side of (4.6), Theorem 3.3 yields, for every  $n \geq N$ ,

$$\mathbb{P}\left[\sup_{0 \leq t \leq T} W_1(\bar{L}_{n,t}, \tilde{\mu}_t) > \frac{a}{2}\right] \leq 2 \exp\left(-\frac{\delta a^2 n}{16}\right).$$

Thanks to Assumption 2.4, by choosing a larger value for  $N \in \mathbb{N}$  than the one in Theorem 3.3, we can ensure that

$$\exp\left(-\frac{a^2 n^2 p(n)}{8K^2 + 2aK/3}\right) \leq \exp\left(-\frac{\delta a^2 n}{16}\right)$$

for every  $n \geq N$ , and the assertion (3.6) follows.

For the result (3.7), we can approximate a general  $\phi$  by those in the  $L^1$ -Fourier class, using the approximation method in Section 5.1.3 of [27], to find the exponential bound for the probability  $\mathbb{P}[\sup_{0 \leq t \leq T} d_{BL}(L_{n,t}, \bar{L}_{n,t}) > a/2]$  (similar to (4.16)). By recalling the fact  $d_{BL} \leq W_1$  and replacing all the  $W_1$  metrics with the  $d_{BL}$  metrics in (4.6), we arrive at (3.7).

**4.3.3. Proof of Theorem 3.5.** As in the proof of Theorem 3.3, for  $Y, Z \in (C([0, T] : \mathbb{R}^d))^n$  we derive

$$\begin{aligned} \left| \sup_{0 \leq t \leq T} W_2\left(\frac{1}{n} \sum_{i=1}^n \delta_{Y_i(t)}, \tilde{\mu}_t\right) - \sup_{0 \leq t \leq T} W_2\left(\frac{1}{n} \sum_{i=1}^n \delta_{Z_i(t)}, \tilde{\mu}_t\right) \right| &\leq \sup_{0 \leq t \leq T} W_2\left(\frac{1}{n} \sum_{i=1}^n \delta_{Y_i(t)}, \frac{1}{n} \sum_{i=1}^n \delta_{Z_i(t)}\right) \\ &\leq \left(\frac{1}{n} \sum_{i=1}^n \|Y_i - Z_i\|^2\right)^{\frac{1}{2}} = \frac{1}{\sqrt{n}} \|Y - Z\|_{n,2}, \end{aligned}$$

again from (2.1) and (2.2). This verifies the  $\left(\frac{1}{\sqrt{n}}\right)$ -Lipschitz continuity of the map

$$Y \mapsto \sup_{0 \leq t \leq T} W_2\left(\frac{1}{n} \sum_{i=1}^n \delta_{Y_i(t)}, \tilde{\mu}_t\right)$$

from  $\left( (C([0, T] : \mathbb{R}^d))^n, \|\cdot\|_{n,2} \right)$  to  $\mathbb{R}$ . Then, for any  $a > 0$ ,

$$\begin{aligned} \mathbb{P} \left[ \sup_{0 \leq t \leq T} W_2(\bar{L}_{n,t}, \tilde{\mu}_t) > a \right] &\leq \mathbb{P} \left[ \sup_{0 \leq t \leq T} W_2(\bar{L}_{n,t}, \tilde{\mu}_t) - \mathbb{E} \left[ \sup_{0 \leq t \leq T} W_2(\bar{L}_{n,t}, \tilde{\mu}_t) \right] > \frac{a}{2} \right] \\ &\quad + \mathbb{P} \left[ \mathbb{E} \left[ \sup_{0 \leq t \leq T} W_2(\bar{L}_{n,t}, \tilde{\mu}_t) \right] > \frac{a}{2} \right]. \end{aligned}$$

The first term is bounded by the right-hand side of (3.4) from Theorem 3.2. The last probability vanishes for all but finitely many  $n$  from Corollary 3.1.

**4.3.4. Proof of Lemma 3.1.** We note from (2.5) that  $\{D_{i,j}^{(n)}\}_{1 \leq i,j \leq n}$  are independent zero-mean random variables, and for every  $i, j \in [n]$ ,

$$\mathbb{E}[(D_{i,j}^{(n)})^2] = \frac{p(n)G(\frac{i}{n}, \frac{j}{n})(1 - p(n)G(\frac{i}{n}, \frac{j}{n}))}{(np(n))^2}. \quad (4.17)$$

In particular, since  $p(n) \leq 1$ , we have  $0 \leq p(n)G(\frac{i}{n}, \frac{j}{n}) \leq 1$ , and thus  $\mathbb{E}[(D_{i,j}^{(n)})^2] \leq 1/(4n^2p(n)^2)$ .

Let us fix any  $n \in \mathbb{N}$ . For arbitrary  $n$ -dimensional vectors  $\mathbf{x}, \mathbf{y} \in [-1, 1]^n$ , we have

$$\begin{aligned} \langle \mathbf{x}, (D^{(n)})^\top D^{(n)} \mathbf{y} \rangle &= \sum_{i=1}^n (D^{(n)}\mathbf{x})_i (D^{(n)}\mathbf{y})_i \\ &= \sum_{i=1}^n \sum_{j=1}^n \sum_{\substack{k=1 \\ k \neq j}}^n D_{i,j}^{(n)} D_{i,k}^{(n)} x_j y_k + \sum_{i=1}^n \sum_{j=1}^n (D_{i,j}^{(n)})^2 x_j y_j \\ &\leq \sum_{i=1}^n \sum_{j=1}^n \sum_{\substack{k=1 \\ k \neq j}}^n D_{i,j}^{(n)} D_{i,k}^{(n)} x_j y_k + \sum_{i=1}^n \sum_{j=1}^n \left( (D_{i,j}^{(n)})^2 x_j y_j - \mathbb{E}[(D_{i,j}^{(n)})^2 x_j y_j] \right) \\ &\quad + \sum_{i=1}^n \sum_{j=1}^n \frac{x_j y_j}{4n^2p(n)^2}. \end{aligned}$$

Thus, for fixed arbitrary  $\eta > 0$  we have

$$\begin{aligned} \mathbb{P} \left[ \frac{\|(D^{(n)})^\top D^{(n)}\|_{\infty \rightarrow 1}}{n} > \eta \right] &\leq \mathbb{P} \left[ \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \sum_{\substack{k=1 \\ k \neq j}}^n D_{i,j}^{(n)} D_{i,k}^{(n)} x_j y_k > \frac{\eta}{3} \right] \\ &\quad + \mathbb{P} \left[ \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \left( (D_{i,j}^{(n)})^2 x_j y_j - \mathbb{E}[(D_{i,j}^{(n)})^2 x_j y_j] \right) > \frac{\eta}{3} \right] \\ &\quad + \mathbb{1}_{\{\frac{1}{4np(n)^2} > \frac{\eta}{3}\}} =: P_1 + P_2 + P_3. \end{aligned} \quad (4.18)$$

From Assumption 3.1, there exists  $N \in \mathbb{N}$  such that  $P_3$  vanishes for every  $n \geq N$ . In the following, we find the bounds for  $P_1$  and  $P_2$ . Using the distribution

$$D_{i,j}^{(n)} = \begin{cases} \frac{1-p(n)G(i/n, j/n)}{np(n)} \leq \frac{1}{np(n)} & \text{with probability } p(n)G(i/n, j/n), \\ -\frac{1}{n}G(i/n, j/n) \geq -\frac{1}{n} & \text{with probability } 1 - p(n)G(i/n, j/n), \end{cases}$$

for each  $i, j \in [n]$ , we derive for  $P_1$

$$\begin{aligned}
P_1 &\leq \sum_{i=1}^n \mathbb{P} \left[ \sum_{j=1}^n \sum_{\substack{k=1 \\ k \neq j}}^n D_{i,j}^{(n)} D_{i,k}^{(n)} x_j y_k > \frac{\eta}{3} \right] \leq \sum_{i=1}^n \sum_{j=1}^n \mathbb{P} \left[ \sum_{\substack{k=1 \\ k \neq j}}^n D_{i,j}^{(n)} D_{i,k}^{(n)} x_j y_k > \frac{\eta}{3n} \right] \\
&= \sum_{i=1}^n \sum_{j=1}^n \mathbb{E} \left[ \mathbb{P} \left( \sum_{\substack{k=1 \\ k \neq j}}^n D_{i,j}^{(n)} D_{i,k}^{(n)} x_j y_k > \frac{\eta}{3n} \middle| D_{i,j}^{(n)} \right) \right] \\
&\leq \sum_{i=1}^n \sum_{j=1}^n \left( \mathbb{P} \left[ \sum_{\substack{k=1 \\ k \neq j}}^n D_{i,k}^{(n)} x_j y_k > \frac{\eta p(n)}{3} \right] + \mathbb{P} \left[ \sum_{\substack{k=1 \\ k \neq j}}^n D_{i,k}^{(n)} x_j y_k < -\frac{\eta}{3} \right] \right).
\end{aligned}$$

The summands  $D_{i,k}^{(n)} x_j y_k$  in the last two probabilities are independent zero-mean random variables bounded above by  $1/(np(n))$ , bounded below by  $-1/n$ , and satisfying

$$\sum_{\substack{k=1 \\ k \neq j}}^n \mathbb{E}[(D_{i,k}^{(n)} x_j y_k)^2] \leq \frac{n-1}{4n^2 p(n)^2} \leq \frac{1}{4np(n)^2},$$

from (4.17). From Bernstein's inequality (Lemma 2.10), we have

$$\begin{aligned}
\mathbb{P} \left[ \sum_{\substack{k=1 \\ k \neq j}}^n D_{i,k}^{(n)} x_j y_k > \frac{\eta p(n)}{3} \right] &\leq \exp \left( -\frac{2np(n)^4 \eta^2}{9 + 4np(n)^2} \right), \\
\mathbb{P} \left[ -\sum_{\substack{k=1 \\ k \neq j}}^n D_{i,k}^{(n)} x_j y_k > \frac{\eta}{3} \right] &\leq \exp \left( -\frac{2np(n)^2 \eta^2}{9 + 4np(n)^2} \right);
\end{aligned}$$

thus

$$P_1 \leq 2n^2 \exp \left( -\frac{2np(n)^4 \eta^2}{9 + 4np(n)^2} \right). \quad (4.19)$$

We now compute the bound for  $P_2$ . We have

$$P_2 \leq \sum_{i=1}^n \mathbb{P} \left[ \sum_{j=1}^n \left( (D_{i,j}^{(n)})^2 x_j y_j - \mathbb{E}[(D_{i,j}^{(n)})^2 x_j y_j] \right) > \frac{\eta}{3} \right], \quad (4.20)$$

and the summands  $(D_{i,j}^{(n)})^2 x_j y_j - \mathbb{E}[(D_{i,j}^{(n)})^2 x_j y_j]$  in the probability are independent zero-mean random variables bounded above by  $5/(2np(n))^2$ . Moreover, we easily obtain the bound

$\mathbb{E}[(D_{i,j}^{(n)})^4] \leq 1/(np(n))^4$ , and thus the sum of the variances of the summands are

$$\begin{aligned} \sum_{j=1}^n \mathbb{E} \left[ \left( (D_{i,j}^{(n)})^2 x_{jy_j} - \mathbb{E}[(D_{i,j}^{(n)})^2 x_{jy_j}] \right)^2 \right] &\leq \sum_{j=1}^n \mathbb{E} \left[ \left( (D_{i,j}^{(n)})^2 - \mathbb{E}[(D_{i,j}^{(n)})^2] \right)^2 \right] \\ &\leq \sum_{j=1}^n \left( \mathbb{E}[(D_{i,j}^{(n)})^4] + (\mathbb{E}[(D_{i,j}^{(n)})^2])^2 \right) \leq \frac{17}{16n^3 p(n)^4}. \end{aligned}$$

Applying Bernstein's inequality (Lemma 2.10) to each probability in (4.20) yields

$$P_2 \leq n \exp \left( - \frac{8n^3 p(n)^4 \eta^2}{153 + 20np(n)^2 \eta} \right). \quad (4.21)$$

Comparing the bounds of (4.19) and (4.21), modifying the value of  $N$  if necessary, and plugging these into (4.18), we obtain the result.

**4.3.5. Proof of Theorem 3.6.** The argument is similar to the proof of Theorem 3.4. The triangle inequality gives

$$\mathbb{P} \left[ \sup_{0 \leq t \leq T} W_2(L_{n,t}, \tilde{\mu}_t) > a \right] \leq \mathbb{P} \left[ \sup_{0 \leq t \leq T} W_2(\bar{L}_{n,t}, \tilde{\mu}_t) > \frac{a}{2} \right] + \mathbb{P} \left[ \sup_{0 \leq t \leq T} W_2(L_{n,t}, \bar{L}_{n,t}) > \frac{a}{2} \right]. \quad (4.22)$$

Recalling the identity (4.7), applying Hölder's inequality several times, and using the Lipschitz property, we obtain

$$\begin{aligned} &|X_i^n(t) - \bar{X}_i^n(t)|^2 - 2t \int_0^t \left| \sum_{j=1}^n D_{i,j}^{(n)} \phi(X_i^n(s), X_j^n(s)) \right|^2 ds \\ &\leq 2K^2 t \int_0^t 2 \left( \sum_{j=1}^n \bar{P}_{i,j}^{(n)} \left( |X_i^n(s) - \bar{X}_i^n(s)| + |X_j^n(s) - \bar{X}_j^n(s)| \right)^2 + 2|X_i^n(s) - \bar{X}_i^n(s)|^2 \right) ds \\ &\leq 2K^2 t \int_0^t 6|X_i^n(s) - \bar{X}_i^n(s)|^2 + \frac{2}{n} \sum_{j=1}^n |X_j^n(s) - \bar{X}_j^n(s)|^2 ds, \quad \forall t \in [0, T]. \end{aligned}$$

For a fixed  $t \in [0, T]$ , by the continuity of  $X_i^n(\cdot) - \bar{X}_i^n(\cdot)$ , there exists  $t_i \in [0, t]$  for each  $i \in [n]$  satisfying  $\square(t) := \frac{1}{n} \sum_{i=1}^n \|X_i^n - \bar{X}_i^n\|_{\star,t}^2 = \frac{1}{n} \sum_{i=1}^n |X_i^n(t_i) - \bar{X}_i^n(t_i)|^2$ . Combining this with the last inequality, we have

$$\square(t) \leq 2t \int_0^t \frac{1}{n} \sum_{i=1}^n \left| \sum_{j=1}^n D_{i,j}^{(n)} \mathbb{1}_{[0,t_i]}(s) \phi(X_i^n(s), X_j^n(s)) \right|^2 ds + 16K^2 t \int_0^t \square(s) ds. \quad (4.23)$$

We recall the representations (4.11) and (4.13) and use Hölder's inequality to derive for the first integral on the right-hand side

$$\begin{aligned}
& \int_0^t \frac{1}{n} \sum_{i=1}^n \left| \sum_{j=1}^n D_{i,j}^{(n)} \mathbb{1}_{[0,t_i]}(s) \phi(X_i^n(s), X_j^n(s)) \right|^2 ds \\
& \leq d \int_0^t \frac{1}{n} \sum_{i=1}^n \sum_{k=1}^d \left| \sum_{j=1}^n \int_{\mathbb{R}^{2d}} |D_{i,j}^{(n)} \mathbb{1}_{[0,t_i]}(s) a_i^k(z, s) b_j^k(z, s)| m_{\phi_k}(dz) \right|^2 ds \\
& \leq d \int_0^t \frac{1}{n} \sum_{i=1}^n \sum_{k=1}^d \|m_{\phi_k}\|_{TM} \int_{\mathbb{R}^{2d}} |(\mathbf{a}^k(z, s))_i (D^{(n)} \mathbf{b}^k(z, s))_i|^2 m_{\phi_k}(dz) ds \\
& \leq d \int_0^t \sum_{k=1}^d \|m_{\phi_k}\|_{TM} \int_{\mathbb{R}^{2d}} \frac{1}{n} \sum_{i=1}^n |(D^{(n)} \mathbf{b}^k(z, s))_i|^2 m_{\phi_k}(dz) ds \\
& \leq \max_{1 \leq k \leq d} \|m_{\phi_k}\|_{TM}^2 d^2 t \frac{\|(D^{(n)})^\top D^{(n)}\|_{\infty \rightarrow 1}}{n}.
\end{aligned}$$

In the last two inequalities, we used the fact that the  $\ell^\infty$  norms of the two vectors  $\mathbf{a}^k(z, s)$  and  $\mathbf{b}^k(z, s)$  are bounded by 1. Thus, from (4.23), there exists a constant  $K > 0$  such that

$$\square(t) \leq Kt^2 \frac{\|(D^{(n)})^\top D^{(n)}\|_{\infty \rightarrow 1}}{n} + Kt \int_0^t \square(s) ds$$

holds for every  $t \in [0, T]$ , and applying Grönwall's inequality gives

$$\square(T) \leq \frac{K \|(D^{(n)})^\top D^{(n)}\|_{\infty \rightarrow 1}}{n},$$

where the constant  $K$  now depends on  $T$ .

Since we have

$$\sup_{0 \leq t \leq T} W_2(L_{n,t}, \bar{L}_{n,t}) \leq \sup_{0 \leq t \leq T} \sqrt{\frac{1}{n} \sum_{i=1}^n |X_i^n(t) - \bar{X}_i^n(t)|^2} \leq \sqrt{\square(T)},$$

Lemma 3.1 shows that there exists  $N \in \mathbb{N}$  such that the last probability in (4.22) has the bound

$$\begin{aligned}
\mathbb{P} \left[ \sup_{0 \leq t \leq T} W_2(L_{n,t}, \bar{L}_{n,t}) > \frac{a}{2} \right] & \leq \mathbb{P} \left[ \square(T) > \frac{a^2}{4} \right] \leq \mathbb{P} \left[ \frac{\|(D^{(n)})^\top D^{(n)}\|_{\infty \rightarrow 1}}{n} > \frac{a^2}{4K} \right] \\
& \leq 3n^2 \exp \left( -\frac{a^4 np(n)^4}{72K^2 + 8a^2K} \right),
\end{aligned}$$

for every  $n \geq N$ .

On the other hand, Theorem 3.5 gives the bound for the other probability in (4.22). By comparing the two bounds under Assumption 3.1(i), we obtain the assertion (3.9). The result (3.10) is now clear under Assumption 3.1(ii), if we set  $p(n) \equiv 1$  and redefine the constants  $\delta > 0$  and  $N \in \mathbb{N}$  appropriately.

## Funding information

E. Bayraktar is supported in part by the National Science Foundation under the grant DMS-2106556 and by the Susan M. Smith Professorship.

## Competing interests

There were no competing interests to declare which arose during the preparation or publication process of this article.

## References

- [1] AURELL, A., CARMONA, R. AND LAURIÈRE, M. (2022). Stochastic graphon games: II. The linear-quadratic case. *Appl. Math. Optimization* **85**, article no. 39.
- [2] BAYRAKTAR, E., CHAKRABORTY, S. AND WU, R. (2023). Graphon mean field systems. *Ann. Appl. Prob.* **33**, 3587–3619.
- [3] BAYRAKTAR, E. AND WU, R. (2022). Stationarity and uniform in time convergence for the graphon particle system. *Stoch. Process. Appl.* **150**, 532–568.
- [4] BAYRAKTAR, E. AND WU, R. (2023). Graphon particle system: uniform-in-time concentration bounds. *Stoch. Process. Appl.* **156**, 196–225.
- [5] BAYRAKTAR, E., WU, R. AND ZHANG, X. (2023). Propagation of chaos of forward–backward stochastic differential equations with graphon interactions. *Appl. Math. Optimization* **88**, article no. 25.
- [6] BET, G., COPPINI, F. AND NARDI, F. R. (2023). Weakly interacting oscillators on dense random graphs. *J. Appl. Prob.* (published online).
- [7] BHAMIDI, S., BUDHIRAJA, A. AND WU, R. (2019). Weakly interacting particle systems on inhomogeneous random graphs. *Stoch. Process. Appl.* **129**, 2174–2206.
- [8] BOBKOV, S. AND GÖTZE, F. (1999). Exponential integrability and transportation cost related to logarithmic Sobolev inequalities. *J. Funct. Anal.* **163**, 1–28.
- [9] BOUCHERON, S., LUGOSI, G. AND MASSART, P. (2013). *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford University Press.
- [10] CAINES, P. E. AND HUANG, M. (2018). Graphon mean field games and the GMFG equations. In *2018 IEEE Conference on Decision and Control (CDC)*, Institute of Electrical and Electronics Engineers, Piscataway, NJ, pp. 4129–4134.
- [11] CAINES, P. E. AND HUANG, M. (2021). Graphon mean field games and their equations. *SIAM J. Control Optimization* **59**, 4373–4399.
- [12] CARMONA, R. A., COONEY, D. B., GRAVES, C. V. AND LAURIÈRE, M. (2021). Stochastic graphon games: I. The static case. *Math. Operat. Res.* **47**, 750–778.
- [13] COPPINI, F. (2022). Long time dynamics for interacting oscillators on graphs. *Ann. Appl. Prob.* **32**, 360–391.
- [14] COPPINI, F. (2022). A note on Fokker–Planck equations and graphons. *J. Statist. Phys.* **187**, article no. 15.
- [15] DELARUE, F. (2017). Mean field games: a toy model on an Erdős–Renyi graph. *ESAIM Proc. Surveys* **60**, 1–26.
- [16] DELARUE, F., LACKER, D. AND RAMANAN, K. (2020). From the master equation to mean field game limit theory: large deviations and concentration of measure. *Ann. Prob.* **48**, 211–263.
- [17] DELATTRE, S., GIACOMIN, G. AND LUÇON, E. (2016). A note on dynamical models on random graphs and Fokker–Planck equations. *J. Statist. Phys.* **165**, 785–798.
- [18] DJELLOUT, H., GUILLIN, A. AND WU, L. (2004). Transportation cost-information inequalities and applications to random dynamical systems and diffusions. *Ann. Prob.* **32**, 2702–2732.
- [19] FOURNIER, N. AND GUILLIN, A. (2015). On the rate of convergence in Wasserstein distance of the empirical measure. *Prob. Theory Relat. Fields* **162**, 707–738.
- [20] GAO, S., CAINES, P. E. AND HUANG, M. (2021). LQG graphon mean field games: graphon invariant subspaces. In *2021 60th IEEE Conference on Decision and Control (CDC)*, Institute of Electrical and Electronics Engineers, Piscataway, NJ, pp. 5253–5260.
- [21] GAO, S., TCHUENDOM, R. F. AND CAINES, P. E. (2021). Linear quadratic graphon field games. *Commun. Inf. Systems* **21**, 341–369.
- [22] GOZLAN, N. AND LÉONARD, C. (2007). A large deviation approach to some transportation cost inequalities. *Prob. Theory Relat. Fields* **139**, 235–283.
- [23] GUÉDON, O. AND VERSHYNIN, R. (2016). Community detection in sparse networks via Grothendieck’s inequality. *Prob. Theory Relat. Fields* **165**, 1025–1049.
- [24] KARATZAS, I. AND SHREVE, S. E. (1991). *Brownian Motion and Stochastic Calculus*, 2nd edn. Springer, New York.

- [25] LACKER, D. AND SORET, A. (2021). A case study on stochastic games on large graphs in mean field and sparse regimes. *Math. Operat. Res.* **47**, 1530–1565.
- [26] LUÇON, E. (2020). Quenched asymptotics for interacting diffusions on inhomogeneous random graphs. *Stoch. Process. Appl.* **130**, 6783–6842.
- [27] OLIVEIRA, R. AND REIS, G. (2019). Interacting diffusions on random graphs with diverging average degrees: hydrodynamics and large deviations. *J. Statist. Phys.* **176**, 1057–1087.
- [28] PARISE, F. AND OZDAGLAR, A. E. (2023). Graphon games: a statistical framework for network games and interventions. *Econometrica* **91**, 191–225.
- [29] SZNITMAN, A.-S. (1991). Topics in propagation of chaos. In *École d'Été de Probabilités de Saint-Flour XIX—1989*, Springer, Berlin, Heidelberg, pp. 165–251.
- [30] TCHUENDOM, R. F., CAINES, P. E. AND HUANG, M. (2020). On the master equation for linear quadratic graphon mean field games. In *2020 59th IEEE Conference on Decision and Control (CDC)*, Institute of Electrical and Electronics Engineers, Piscataway, NJ, pp. 1026–1031.
- [31] TCHUENDOM, R. F., CAINES, P. E. AND HUANG, M. (2021). Critical nodes in graphon mean field games. In *2021 60th IEEE Conference on Decision and Control (CDC)*, Institute of Electrical and Electronics Engineers, Piscataway, NJ, pp. 166–170.
- [32] ÜSTÜNEL, A. (2012). Transportation cost inequalities for diffusions under uniform distance. In *Stochastic Analysis and Related Topics*, Springer, Berlin, Heidelberg, pp. 203–214.
- [33] VASAL, D., MISHRA, R. AND VISHWANATH, S. (2021). Sequential decomposition of graphon mean field games. In *2021 American Control Conference (ACC)*, Institute of Electrical and Electronics Engineers, Piscataway, NJ, pp. 730–736.