

GRAPHON PARTICLE SYSTEM: UNIFORM-IN-TIME CONCENTRATION BOUNDS

ERHAN BAYRAKTAR AND RUOYU WU*

ABSTRACT. In this paper, we consider graphon particle systems with heterogeneous mean-field type interactions and the associated finite particle approximations. Under suitable growth (resp. convexity) assumptions, we obtain uniform-in-time concentration estimates, over finite (resp. infinite) time horizon, for the Wasserstein distance between the empirical measure and its limit, extending the work of Bolley–Guillin–Villani [9].

CONTENTS

1. Introduction	1
1.1. Organization	3
1.2. Notation	3
2. Model, assumptions, results	4
3. Proofs	7
3.1. Properties of the limiting system	7
3.2. Bounds in terms of empirical measures of independent variables	8
3.3. Exponential estimates	10
3.4. Proofs of Theorems 2.1 and 2.2	16
4. Acknowledgment	19
Appendix A. Proofs of results in Section 3.1	19
Appendix B. Sub-Gaussian inequality	22
Appendix C. Bounds of independent random variables	23
References	27

1. INTRODUCTION

In this work we study the uniform-in-time exponential concentration bounds related to the graphon particle system and its finite particle approximations. This is a continuation of our work [2, 4] on the large population and long time behavior for these systems. The interaction in the graphon particle system is of mean-field type and characterized by a (directed) graphon G , which is a measurable function from $[0, 1] \times [0, 1]$ to $[0, 1]$ (see e.g. [27] for the theory of graphons). More precisely, denoting by $\bar{X}_u(t)$ the state of the particle indexed by $u \in [0, 1]$

2020 *Mathematics Subject Classification.* 05C80 60J60 60K35.

Key words and phrases. graphons, graphon particle systems, mean field interaction, heterogeneous interaction, networks, long time behavior, exponential concentration bounds, transport inequalities.

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

at time $t \geq 0$,

$$\bar{X}_u(t) = \bar{X}_u(0) + \int_0^t \left(f(\bar{X}_u(s)) + \int_0^1 \int_{\mathbb{R}^d} b(\bar{X}_u(s), x) G(u, v) \bar{\mu}_{v,s}(dx) dv \right) ds + B_u(t), \quad (1.1)$$

where $\bar{\mu}_{v,s}$ is the probability distribution of the \mathbb{R}^d -valued random variable $\bar{X}_v(s)$ for each $v \in [0, 1]$ and $s \geq 0$, f and b are suitable functions, $\{B_u : u \in [0, 1]\}$ are d -dimensional standard Brownian motions, and $\{\bar{X}_u(0), B_u : u \in [0, 1]\}$ are mutually independent. Due to the general form of G , the system (1.1) consists of uncountably many heterogeneous particles \bar{X}_u , the evolution of whose probability distributions $\bar{\mu}_{u,t}$ are fully coupled. Indeed, $\{(\bar{X}_u(0), B_u) : u \in [0, 1]\}$ are not necessarily identically distributed and may not be measurable in $u \in [0, 1]$. Although the system (1.1) of nonlinear McKean–Vlasov type processes $\bar{X}_u(t)$ is well-posed [2], we will be only interested in the concentration of finite particle empirical measures around the probability laws $\bar{\mu}_{u,t}$, and hence one may alternatively treat (1.1) as an informal description and work with equivalent formulations that do not involve uncountable $\{B_u : u \in [0, 1]\}$; see Remark 2.2 on more of this point. Still, in general, $\bar{\mu}_{u,t}$ and the associated stationary measure, provided existence under suitable assumptions, may not be tractable (even in the case of linear f and b ; see e.g. [4, Example 3.1]). However, the system (1.1) arises naturally as the limit (see e.g. [2, 4]) of the associated finite particle system with heterogeneous interactions given by

$$X_i^n(t) = \bar{X}_{\frac{i}{n}}(0) + \int_0^t \left(f(X_i^n(s)) + \frac{1}{n} \sum_{j=1}^n \xi_{ij}^n b(X_i^n(s), X_j^n(s)) \right) ds + B_{\frac{i}{n}}(t) \quad (1.2)$$

for $i \in \{1, \dots, n\}$ and $t \geq 0$, where $\{\xi_{ij}^n : 1 \leq i \leq j \leq n\}$ is a collection of independent $[0, 1]$ -valued random variables sampled from the graphon G . The study of uniform-in-time exponential concentration bounds will provide useful quantitative non-asymptotic estimates of the fluctuations for the convergence of the empirical measures of (1.2) to the probability laws of (1.1).

When $G \equiv 1$ and $\xi_{ij}^n \equiv 1$, the above systems (1.1) and (1.2) reduce to the classic McKean–Vlasov processes and associated weakly interacting diffusions. The study of this *homogeneous* setup dates back to works of Boltzmann, Vlasov, McKean and others (see [25, 29, 35] and references therein). Besides large population asymptotics such as law of large numbers (LLN) on the finite time horizon, there have been an extensive collection of results on concentration estimates and uniform-in-time LLN (see e.g. [8, 9, 11, 20, 37] and references therein).

In recent ten years, there has been a growing interest in the mean-field *inhomogeneous* particle system, where $G \equiv 1$ (or block-wise constant) and the interaction between particles is governed by their own types and/or random graphs (see e.g. [1, 3, 7, 12, 18, 19, 21, 26]). The inhomogeneity arises from random interactions ξ_{ij}^n while the limiting system is still homogeneous, as opposed to uncountable heterogeneous processes in graphon particle systems (1.1) with general graphon G .

The study of graphon particle systems and associated finite particle models with mean-field *heterogeneous* interactions emerged recently ([2, 4, 6, 16, 17, 28, 33]). There is also a growing number of applications of graphons in game theory; see e.g. [5, 13–15, 23, 34, 36] for the study of graphon mean field games in static and dynamic settings. These works focus on large population convergence over finite time horizon, except for [4, 17] on uniform-in-time convergence, where the limit is homogeneous in [17] while heterogeneous in [4]. Deterministic dynamical systems on graphons over finite time horizon are also studied (see [22, 30–32] and references therein).

The analysis of the continuum limit (1.1) is motivated by physical or biological applications such as in [16–18, 28, 30–33], by numerical questions as in [9], or by game theory applications such as in [13–15, 23, 34, 36]. In this work we are interested in the approximation errors of (1.1) by (1.2) by obtaining non-asymptotic exponential concentration results for quantities like

$$\mathbb{P}\left(\sup_{0 \leq t \leq T} W_1(\nu_t^n, \hat{\mu}_t) > \varepsilon\right) \quad \text{and} \quad \sup_{t \geq 0} \mathbb{P}(W_1(\nu_t^n, \hat{\mu}_t) > \varepsilon)$$

for $\varepsilon > 0$, in Theorem 2.1 (resp. Theorem 2.2) under suitable growth (resp. convexity) condition. Here W_1 is the Wasserstein-1 distance, ν_t^n is the empirical measure of $\{X_i^n(t) : i = 1, \dots, n\}$, and $\hat{\mu}_t$ is the averaged law of $\{\bar{X}_u(t) : u \in [0, 1]\}$. As an application, we also obtain in Corollary 2.1 an exponential concentration bound for numerical reconstruction of the invariant measure of $\{\bar{X}_u(t) : u \in [0, 1]\}$. Concentration estimates are shown in [8, 9, 11] for mean-field (type) systems over infinite time horizon and in [20] for mean-field games over finite time horizon. Theorems 2.1 and 2.2 provide for the first time quantitative non-asymptotic exponential error bounds for graphon particle systems. To the best of our knowledge, there are no results on concentration bounds in either inhomogeneous or heterogeneous regime, over either finite or infinite time horizon. Our results are consistent with those in [9] when the system is mean-field (namely $G \equiv 1$ and $\xi_{ij}^n \equiv 1$), and actually improve pre-exponential estimates in the infinite time horizon (see Remark 2.4). Although only the dissipative regime is considered for the infinite time horizon, the results here suggest that it is indeed possible to obtain concentration bounds with asymptotic independence instead of the usual assumption on homogeneity or exchangeability of the limiting system, and leave the door open for future study of more general non-dissipative regimes.

The proofs of Theorems 2.1 and 2.2 start by reducing the analysis of $W_1(\nu_t^n, \hat{\mu}_t)$ to that of $W_1(\bar{\nu}_t^n, \hat{\mu}_t)$, where $\bar{\nu}_t^n$ is the empirical measure of $\{\bar{X}_{i/n}(t) : i = 1, \dots, n\}$. In the homogeneous setup [9, Proposition 6.1], this reduction is done by a classic coupling argument. However, due to the heterogeneity of the stochastic processes $\{\bar{X}_{i/n}\}$ and the presence of random interactions $\{\xi_{ij}^n\}$, such a coupling argument does not provide the desired reduction any more. Instead, we carefully bound $W_1(\nu_t^n, \hat{\mu}_t)$ by $W_1(\bar{\nu}_t^n, \hat{\mu}_t)$ in a different way in Proposition 3.1, with certain new error terms consisting of $\{\bar{X}_{i/n}, \xi_{ij}^n\}$ and requiring new treatments. Such new terms are further estimated in the exponential scale in Proposition 3.2. In particular, the key ingredients for such estimates are exponential bounds obtained in Lemmas 3.7, 3.9 and 3.10, using concentration inequalities and sub-Gaussian properties. The distance $W_1(\bar{\nu}_t^n, \hat{\mu}_t)$ is analyzed in Lemma 3.11 using exponential concentration bounds, for empirical measures of independent (but not necessarily identically distributed) random variables, established in Proposition C.1 which is a heterogeneous version of [9, Theorem 2.1] where the i.i.d. setup is analyzed.

1.1. Organization. The paper is organized as follows. In Section 2 we state the space of graphons, the standing assumptions, and well-posedness of systems (1.1) and (1.2). The main results on concentration bounds are stated in Theorems 2.1 and 2.2. Proofs of these two results are given in Section 3, with the key exponential estimates provided in Section 3.3. Finally Appendices A–C collect proofs of some auxiliary results.

We close this section by introducing some frequently used notation.

1.2. Notation. Given a Polish space \mathbb{S} , denote by $\mathcal{B}(\mathbb{S})$ the Borel σ -field. Let $\mathcal{P}(\mathbb{S})$ be the space of probability measures on \mathbb{S} endowed with the topology of weak convergence. Denote by $\mathcal{C}([0, T] : \mathbb{R}^d)$ the space of continuous functions from $[0, T]$ to \mathbb{R}^d , endowed with the topology of uniform convergence. We will use C, C_1, C_2, \dots , to denote various positive constants in

the paper and $C(m)$ to emphasize the dependence on some parameter m . The probability law of a random variable X will be denoted by $\mathcal{L}(X)$. Expectations under \mathbb{P} will be denoted by \mathbb{E} . To simplify the notation, we will usually write $\mathbb{E}[X^k]$ as $\mathbb{E}X^k$. For vectors $x, y \in \mathbb{R}^d$, denote by $|x|$ the Euclidean norm and $x \cdot y$ the inner product. Denote by W_p , $p \in \mathbb{N}$, the Wasserstein- p distance (cf. [38, Chapter 6]) on $\mathcal{P}(\mathbb{R}^d)$:

$$W_p(m_1, m_2) := \left(\inf_{\pi} \int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^p \pi(dx dy) \right)^{1/p}, \quad m_1, m_2 \in \mathcal{P}(\mathbb{R}^d),$$

where the infimum is taken over all probability measures $\pi \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$ with marginals m_1 and m_2 , that is, $\pi(\cdot \times \mathbb{R}^d) = m_1(\cdot)$ and $\pi(\mathbb{R}^d \times \cdot) = m_2(\cdot)$.

2. MODEL, ASSUMPTIONS, RESULTS

We follow the notation used in [27, Chapters 7 and 8]. Let $I := [0, 1]$. Denote by \mathcal{G} the space of all **bounded measurable** functions $G: I \times I \rightarrow \mathbb{R}$. A **(directed) graphon** G is an element of \mathcal{G} with $0 \leq G \leq 1$.

Given a graphon $G \in \mathcal{G}$ and a collection of initial distributions $\bar{\mu}(0) := (\bar{\mu}_u(0) \in \mathcal{P}(\mathbb{R}^d) : u \in I)$, recall the graphon particle system (1.1) and the finite particle system (1.2). The following assumptions will be made throughout the paper.

Standing Assumptions:

- The map $I \ni u \mapsto \mathcal{L}(\bar{\mu}_u(0)) \in \mathcal{P}(\mathbb{R}^d)$ is measurable, and $\sup_{u \in I} \int_{\mathbb{R}^d} e^{\theta_0 |x|^2} \bar{\mu}_{u,0}(dx) < \infty$ for some $\theta_0 > 0$.
- The drift functions f and b are Lipschitz with Lipschitz constants K_f and K_b , respectively, namely

$$|f(x_1) - f(x_2)| \leq K_f |x_1 - x_2|, \quad \forall x_1, x_2 \in \mathbb{R}^d,$$

$$|b(x_1, y_1) - b(x_2, y_2)| \leq K_b(|x_1 - x_2| + |y_1 - y_2|), \quad \forall x_1, x_2, y_1, y_2 \in \mathbb{R}^d.$$

- $G \in \mathcal{G}$ is a **directed graphon** and ξ_{ij}^n is the *sampled graphon*, namely
 - (i) either $\xi_{ij}^n = G(\frac{i}{n}, \frac{j}{n})$ for $i, j \in \{1, \dots, n\}$,
 - (ii) or $\xi_{ij}^n = \text{Bernoulli}(G(\frac{i}{n}, \frac{j}{n}))$ independently for $i, j \in \{1, \dots, n\}$, and independent of $\{\bar{X}_u(0), B_u : u \in I\}$.
- There exist some $K_G \in (0, \infty)$ and a finite collection of **disjoint** intervals $\{I_i : i = 1, \dots, N\}$ for some $N \in \mathbb{N}$, such that $\cup_{i=1}^N I_i = I$ and

$$W_2(\bar{\mu}_{u_1}(0), \bar{\mu}_{u_2}(0)) \leq K_G |u_1 - u_2|, \quad u_1, u_2 \in I_i, \quad i \in \{1, \dots, N\},$$

$$|G(u_1, v_1) - G(u_2, v_2)| \leq K_G(|u_1 - u_2| + |v_1 - v_2|), \quad (u_1, v_1), (u_2, v_2) \in I_i \times I_j, \quad i, j \in \{1, \dots, N\}.$$

We note that, since f is Lipschitz, the quantity

$$c_0 := \inf_{x_1, x_2 \in \mathbb{R}^d, x_1 \neq x_2} \frac{-(x_1 - x_2) \cdot (f(x_1) - f(x_2))}{|x_1 - x_2|^2}$$

is well-defined, and $|c_0| \leq K_f$. From this we have

$$(x_1 - x_2) \cdot (f(x_1) - f(x_2)) \leq -c_0 |x_1 - x_2|^2, \quad \forall x_1, x_2 \in \mathbb{R}^d. \quad (2.1)$$

Let

$$\kappa := c_0 - 2K_b. \quad (2.2)$$

For several long-time results the following dissipativity assumption will be made.

Condition 2.1. *The coefficient f is dissipative, in the sense that $\kappa > 0$.*

Remark 2.1. A common example of b and f satisfying Condition 2.1 is linear (as in the study of linear quadratic graphon mean-field games in, e.g., [23]) and mean-reverting:

$$f(x) + b(x, y) = -c_1 x + c_2 y, \quad \text{for some } c_1 > c_2 > 0.$$

In particular, the choice of $f(x) = -(c_1 + c_2)x$ and $b(x, y) = c_2(x + y)$ satisfies Condition 2.1 since $c_0 = c_1 + c_2 > 2c_2 = 2K_b$. For a more general linear mean-reverting example, see [4, Example 3.1].

The following result gives well-posedness of systems (1.1) and (1.2). The proof is standard (see e.g. the homogeneous setup [35] and the heterogeneous setup [2] for part (i), and [24, Theorems 5.2.5 and 5.2.9] for part (ii)) and hence is omitted.

Proposition 2.1. (i) There exists a unique pathwise solution to (1.1). For every $T < \infty$, the map $I \ni u \mapsto \bar{\mu}_u \in \mathcal{P}(\mathbb{C}([0, T] : \mathbb{R}^d))$ is measurable and

$$\sup_{u \in I} \sup_{t \in [0, T]} \mathbb{E} [|\bar{X}_u(t)|^4] < \infty.$$

(ii) There exists a unique pathwise solution to (1.2). Also for every $T < \infty$,

$$\max_{i=1, \dots, n} \sup_{t \in [0, T]} \mathbb{E} [|X_i^n(t)|^4] < \infty.$$

Remark 2.2. The formulation (1.1) with Brownian motions $\{B_u : u \in I\}$ is used to describe the nonlinear McKean–Vlasov type stochastic processes $\{\bar{X}_u : u \in I\}$ and emphasize their independence. However, results in this work are on measures of $\{\bar{X}_u : u \in I\}$. Therefore, one may work with the following equivalent formulation with a single Brownian motion $\{B(t) : t \geq 0\}$

$$\bar{X}_u(t) = \bar{X}_u(0) + \int_0^t \left(f(\bar{X}_u(s)) + \int_0^1 \int_{\mathbb{R}^d} b(\bar{X}_u(s), x) G(u, v) \bar{\mu}_{v,s}(dx) dv \right) ds + B(t),$$

with $\bar{\mu}_{v,s} = \mathcal{L}(\bar{X}_v(s))$ and these laws are independent for $v \in I$; see e.g. [5, 6, 28]. In fact, one may just work with the spatially extended law $\bar{\mu}_t(du dx) := \bar{\mu}_{u,t}(dx) du$ and the nonlinear Fokker–Planck equations

$$\begin{aligned} \langle \bar{\mu}_{u,t}, h \rangle &= \langle \bar{\mu}_{u,0}, h \rangle + \int_0^t \left\langle \bar{\mu}_{u,s}, \nabla h(\cdot) \cdot \left(f(\cdot) + \int_{I \times \mathbb{R}^d} b(\cdot, x) G(u, v) \bar{\mu}_s(dv dx) \right) \right\rangle ds \\ &\quad + \int_0^t \left\langle \bar{\mu}_{u,s}, \frac{1}{2} \Delta h \right\rangle ds \end{aligned}$$

for $u \in I$ and regular test functions h ; see e.g. [6, 28].

Let

$$\nu^n(t) := \frac{1}{n} \sum_{i=1}^n \delta_{X_i^n(t)}, \quad \bar{\nu}^n(t) := \frac{1}{n} \sum_{i=1}^n \delta_{\bar{X}_{\frac{i}{n}}(t)}, \quad \hat{\mu}(t) := \int_I \bar{\mu}_{u,t} du.$$

We first have the following result on the trajectory of Wasserstein distances over finite time horizon. Note that Condition 2.1 is not used.

Theorem 2.1. Fix $T \in (0, \infty)$. Then there exists some $K \in (0, \infty)$ such that for any $d' > d$, there exist $C = C_T \in (0, \infty)$ and $N_0 \in \mathbb{N}$ such that

$$\mathbb{P} \left(\sup_{0 \leq t \leq T} W_1(\nu_t^n, \hat{\mu}_t) > \varepsilon \right) \leq C(1 + \varepsilon^{-2}) \exp(-K\sqrt{n\varepsilon}) \quad (2.3)$$

for all $\varepsilon > 0$ and $n \geq N_0 \max(\varepsilon^{-(d'+2)}, 1)$. If in addition

$$\xi_{ij}^n = G\left(\frac{i}{n}, \frac{j}{n}\right) \text{ or } b(\cdot, 0) \text{ is bounded,} \quad (2.4)$$

then $\sqrt{n\varepsilon}$ in (2.3) could be improved to $n\varepsilon^2$.

Remark 2.3. The technical assumption (2.4) will be used in Lemma 3.7 to derive certain sub-Gaussian bounds. The assumption on b holds in particular if $b(x, y) = b(y)$ is x -independent or if b is bounded.

When the convexity property in Condition 2.1 is satisfied, the following uniform-in-time marginal concentration bound holds.

Theorem 2.2. Suppose Condition 2.1 holds. Then there exists some $K \in (0, \infty)$ such that for any $d' > d$, there exist $C \in (0, \infty)$ and $N_0 \in \mathbb{N}$ such that

$$\sup_{t \geq 0} \mathbb{P}(W_1(\nu_t^n, \hat{\mu}_t) > \varepsilon) \leq C \exp(-K\sqrt{n\varepsilon}) \quad (2.5)$$

for all $\varepsilon > 0$ and $n \geq N_0 \max(\varepsilon^{-(d'+2)}, 1)$. If in addition (2.4) holds, then $\sqrt{n\varepsilon}$ in (2.5) could be improved to $n\varepsilon^2$.

Remark 2.4. In the classic mean-field regime, namely when $G \equiv 1$, one has $\xi_{ij}^n \equiv 1$ so that the condition (2.4) holds and the exponential rates in (2.3) and (2.5) are $n\varepsilon^2$. From this we observe that Theorems 2.1 and 2.2 are consistent with the mean-field results in [9, Theorems 2.9 and 2.12]. In fact, the pre-exponential term C in (2.5) does not depend on ε , while the corresponding term in [9, Theorem 2.12] is $C(1 + \varepsilon^{-2})$. This improvement is due to a different argument (Proposition 3.2) that does not use time discretization as in [9, Section 7.2].

Using Theorem 2.2 we could obtain an exponential concentration bound for numerical reconstruction of the invariant measure.

Corollary 2.1. Suppose Condition 2.1 holds. Then there exist a unique probability measure $\hat{\mu}_\infty \in \mathcal{P}(\mathbb{R}^d)$ and some $\varepsilon_0, T_0, K \in (0, \infty)$ such that for any $d' > d$, there exist $C \in (0, \infty)$ and $N_0 \in \mathbb{N}$ such that

$$\sup_{t \geq T_0 \log(\varepsilon_0/\varepsilon)} \mathbb{P}(W_1(\nu_t^n, \hat{\mu}_\infty) > \varepsilon) \leq C \exp(-K\sqrt{n\varepsilon}) \quad (2.6)$$

for all $\varepsilon > 0$ and $n \geq N_0 \max(\varepsilon^{-(d'+2)}, 1)$. If in addition (2.4) holds, then $\sqrt{n\varepsilon}$ in (2.6) could be improved to $n\varepsilon^2$.

Proof. By the exponential ergodicity of $\hat{\mu}_t$ shown in [4, Theorem 3.1], there exist a unique probability measure $\hat{\mu}_\infty \in \mathcal{P}(\mathbb{R}^d)$ and some $\tilde{C} \in (0, \infty)$ such that

$$W_2(\hat{\mu}_t, \hat{\mu}_\infty) \leq \tilde{C}e^{-\kappa t/2}, \quad t \geq 0.$$

Now let $T_0 = 2/\kappa$ and $\varepsilon_0 = 2\tilde{C}$. Then for $t \geq T_0 \log(\varepsilon_0/\varepsilon)$, we have

$$W_1(\hat{\mu}_t, \hat{\mu}_\infty) \leq W_2(\hat{\mu}_t, \hat{\mu}_\infty) \leq \tilde{C}e^{-\kappa t/2} \leq \varepsilon/2$$

and hence

$$\mathbb{P}(W_1(\nu_t^n, \hat{\mu}_\infty) > \varepsilon) \leq \mathbb{P}(W_1(\nu_t^n, \hat{\mu}_t) > \varepsilon/2).$$

The result then follows from Theorem 2.2. ■

3. PROOFS

In this section we prove Theorems 2.1 and 2.2 on the concentration inequalities of $W_1(\nu_t^n, \hat{\mu}_t)$. We first state some properties of the limiting system in Section 3.1. Using these, in Section 3.2, we reduce the analysis of $W_1(\nu_t^n, \hat{\mu}_t)$ to that of $W_1(\bar{\nu}_t^n, \hat{\mu}_t)$ and other quantities only involving independent but heterogeneous processes. Such quantities are carefully analyzed in Section 3.3. In Section 3.4 we obtain the concentration bounds for $W_1(\bar{\nu}_t^n, \hat{\mu}_t)$ and complete the proofs of Theorems 2.2 and 2.1.

3.1. Properties of the limiting system. In this section we state some properties of the limiting system that will be used later. Proofs will be given in Appendix A.

Lemma 3.1. [2, Theorem 2.1] *For every $T \in (0, \infty)$, there exists some $C(T) \in (0, \infty)$ such that*

$$\sup_{t \in [0, T]} W_2(\bar{\mu}_{u_1}(t), \bar{\mu}_{u_2}(t)) \leq C(T)|u_1 - u_2|$$

whenever $u_1, u_2 \in I_i$ for some $i \in \{1, \dots, N\}$.

Lemma 3.2. [4, Proposition 3.1 and Corollary 3.1] *Suppose Condition 2.1 holds. Then*

$$\sup_{u \in I} \sup_{t \geq 0} \mathbb{E} [|\bar{X}_u(t)|^4] < \infty.$$

Moreover, there exists some $C \in (0, \infty)$ such that

$$\sup_{t \geq 0} W_2(\bar{\mu}_{u_1}(t), \bar{\mu}_{u_2}(t)) \leq C|u_1 - u_2|$$

whenever $u_1, u_2 \in I_i$ for some $i \in \{1, \dots, N\}$.

The following result on the square exponential moments of \bar{X}_u is a generalization of [9, Proposition 4.1] and [11, Proposition 4.3]. The proof is provided in Appendix for completeness. Recall θ_0 in the standing assumption and κ in (2.2).

Lemma 3.3. (i) *For every $T \in (0, \infty)$, there exists some $\theta_T \in (0, \infty)$ such that*

$$\sup_{u \in I} \mathbb{E} \left[\sup_{t \in [0, T]} e^{\theta_T |\bar{X}_u(t)|^2} \right] < \infty.$$

(ii) *Suppose Condition 2.1 holds. Then for any $\theta \in (0, (\kappa \wedge \theta_0)/4)$, we have*

$$\sup_{t \geq 0} \sup_{u \in I} \mathbb{E} [e^{\theta |\bar{X}_u(t)|^2}] < \infty.$$

The following two lemmas on the time-regularity of \bar{X}_u , $\hat{\mu}$ and $\bar{\nu}^n$ are generalizations of results in [9, Section 5] and [11, Section 4.3.3]. We provide the proofs in Appendix for completeness.

Lemma 3.4. *For every $T \in (0, \infty)$, there exist some $C(T), \theta_T \in (0, \infty)$ such that, for all $s, t, t_0, \Delta \in [0, T]$ and $u \in I$,*

$$\begin{aligned} \mathbb{E} |\bar{X}_u(t) - \bar{X}_u(s)|^4 &\leq C(T)|t - s|^2, \\ \mathbb{E} \left[\sup_{t_0 \leq s \leq t \leq t_0 + \Delta} \exp(\theta_T |\bar{X}_u(t) - \bar{X}_u(s)|^2) \right] &\leq 1 + C(T)\Delta. \end{aligned}$$

Lemma 3.5. *For every $T \in (0, \infty)$, there exist some $C(T), \theta_T \in (0, \infty)$ such that*

$$\mathbb{P} \left(\sup_{t_0 \leq s, t \leq t_0 + \Delta} W_1(\bar{\nu}_s^n, \bar{\nu}_t^n) > \varepsilon \right) \leq \exp(-n [\theta_T \varepsilon^2 - C(T) \Delta])$$

for all $t_0 \in [0, T]$, $\Delta \in [0, T]$ and $\varepsilon \in (0, \infty)$.

The following result is an immediate consequence of Lemma 3.4.

Lemma 3.6. *For every $T \in (0, \infty)$, there exists some $C(T) \in (0, \infty)$ such that*

$$W_1(\hat{\mu}_t, \hat{\mu}_s) \leq C(T) |t - s|^{1/2}, \quad \forall s, t \in [0, T].$$

3.2. Bounds in terms of empirical measures of independent variables. The following proposition is a generalization of [9, Proposition 6.1] and reduces the analysis of $W_1(\nu_t^n, \hat{\mu}_t)$ to quantities only in terms of independent processes.

Proposition 3.1. *Let*

$$\mathcal{R}_t^{n,i} := \frac{1}{n} \sum_{j=1}^n \left(\xi_{ij}^n b(\bar{X}_{\frac{i}{n}}(t), \bar{X}_{\frac{j}{n}}(t)) - \int_{\mathbb{R}^d} b(\bar{X}_{\frac{i}{n}}(t), x) G\left(\frac{i}{n}, \frac{j}{n}\right) \bar{\mu}_{\frac{j}{n},t}(dx) \right). \quad (3.1)$$

(i) *For every $T \in (0, \infty)$, there exists some $C(T) \in (0, \infty)$ such that*

$$W_1(\nu_t^n, \hat{\mu}_t) \leq W_1(\bar{\nu}_t^n, \hat{\mu}_t) + C(T) \left(\int_0^t e^{-\kappa(t-s)} \frac{1}{n} \sum_{i=1}^n \left| |\mathcal{R}_s^{n,i}| + \frac{1}{n} (1 + |\bar{X}_{\frac{i}{n}}(s)|) \right|^2 ds \right)^{1/2}$$

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

(ii) *Suppose Condition 2.1 holds. Then there exists some $C \in (0, \infty)$, independent of $t \geq 0$, such that*

$$W_1(\nu_t^n, \hat{\mu}_t) \leq W_1(\bar{\nu}_t^n, \hat{\mu}_t) + C \left(\int_0^t e^{-\kappa(t-s)} \frac{1}{n} \sum_{i=1}^n \left| |\mathcal{R}_s^{n,i}| + \frac{1}{n} (1 + |\bar{X}_{\frac{i}{n}}(s)|) \right|^2 ds \right)^{1/2}$$

for all $t \geq 0$.

Proof. We first prove part (ii). From (1.1) and (1.2) we have

$$\begin{aligned} & d \left| X_i^n(t) - \bar{X}_{\frac{i}{n}}(t) \right|^2 \\ &= 2 \left(X_i^n(t) - \bar{X}_{\frac{i}{n}}(t) \right) \cdot \left(f(X_i^n(t)) - f(\bar{X}_{\frac{i}{n}}(t)) \right) dt \\ &+ 2 \left(X_i^n(t) - \bar{X}_{\frac{i}{n}}(t) \right) \cdot \left(\frac{1}{n} \sum_{j=1}^n \xi_{ij}^n b(X_i^n(t), X_j^n(t)) - \int_I \int_{\mathbb{R}^d} b(\bar{X}_{\frac{i}{n}}(t), x) G\left(\frac{i}{n}, v\right) \bar{\mu}_{v,t}(dx) dv \right) dt. \end{aligned}$$

By (2.1),

$$\left(X_i^n(t) - \bar{X}_{\frac{i}{n}}(t) \right) \cdot \left(f(X_i^n(t)) - f(\bar{X}_{\frac{i}{n}}(t)) \right) \leq -c_0 \left| X_i^n(t) - \bar{X}_{\frac{i}{n}}(t) \right|^2.$$

By adding and subtracting terms, we have

$$\begin{aligned}
& \left(X_i^n(t) - \bar{X}_{\frac{i}{n}}(t) \right) \cdot \left(\frac{1}{n} \sum_{j=1}^n \xi_{ij}^n b(X_i^n(t), X_j^n(t)) - \int_I \int_{\mathbb{R}^d} b(\bar{X}_{\frac{i}{n}}(t), x) G\left(\frac{i}{n}, v\right) \bar{\mu}_{v,t}(dx) dv \right) \\
&= \left(X_i^n(t) - \bar{X}_{\frac{i}{n}}(t) \right) \cdot \left[\frac{1}{n} \sum_{j=1}^n \xi_{ij}^n \left(b(X_i^n(t), X_j^n(t)) - b(\bar{X}_{\frac{i}{n}}(t), \bar{X}_{\frac{j}{n}}(t)) \right) \right] \\
&\quad + \left(X_i^n(t) - \bar{X}_{\frac{i}{n}}(t) \right) \cdot \left[\frac{1}{n} \sum_{j=1}^n \left(\xi_{ij}^n b(\bar{X}_{\frac{i}{n}}(t), \bar{X}_{\frac{j}{n}}(t)) - \int_{\mathbb{R}^d} b(\bar{X}_{\frac{i}{n}}(t), x) G\left(\frac{i}{n}, \frac{j}{n}\right) \bar{\mu}_{\frac{j}{n},t}(dx) \right) \right] \\
&\quad + \left(X_i^n(t) - \bar{X}_{\frac{i}{n}}(t) \right) \cdot \left(\frac{1}{n} \sum_{j=1}^n \int_{\mathbb{R}^d} b(\bar{X}_{\frac{i}{n}}(t), x) G\left(\frac{i}{n}, \frac{j}{n}\right) \bar{\mu}_{\frac{j}{n},t}(dx) - \int_I \int_{\mathbb{R}^d} b(\bar{X}_{\frac{i}{n}}(t), x) G\left(\frac{i}{n}, v\right) \bar{\mu}_{v,t}(dx) dv \right) \\
&=: \tilde{\mathcal{R}}_t^{n,i,1} + \tilde{\mathcal{R}}_t^{n,i,2} + \tilde{\mathcal{R}}_t^{n,i,3}.
\end{aligned}$$

For $\tilde{\mathcal{R}}_t^{n,i,1}$, using the Lipschitz property of b , we have

$$\tilde{\mathcal{R}}_t^{n,i,1} \leq \frac{K_b}{n} \sum_{j=1}^n \left| X_i^n(t) - \bar{X}_{\frac{i}{n}}(t) \right| \left(\left| X_i^n(t) - \bar{X}_{\frac{i}{n}}(t) \right| + \left| X_j^n(t) - \bar{X}_{\frac{j}{n}}(t) \right| \right).$$

For $\tilde{\mathcal{R}}_t^{n,i,2}$, from (3.1) we have

$$\tilde{\mathcal{R}}_t^{n,i,2} \leq \left| X_i^n(t) - \bar{X}_{\frac{i}{n}}(t) \right| |\mathcal{R}_s^{n,i}|.$$

For $\tilde{\mathcal{R}}_t^{n,i,3}$, using Lemma 3.2, we have

$$\begin{aligned}
\tilde{\mathcal{R}}_t^{n,i,3} &= \left(X_i^n(t) - \bar{X}_{\frac{i}{n}}(t) \right) \cdot \int_I \int_{\mathbb{R}^d} b(\bar{X}_{\frac{i}{n}}(t), x) \left(G\left(\frac{i}{n}, \frac{[nv]}{n}\right) \bar{\mu}_{\frac{[nv]}{n},t}(dx) - G\left(\frac{i}{n}, v\right) \bar{\mu}_{v,t}(dx) \right) dv \\
&\leq \frac{C_1}{n} \left| X_i^n(t) - \bar{X}_{\frac{i}{n}}(t) \right| \left(1 + \left| \bar{X}_{\frac{i}{n}}(t) \right| \right).
\end{aligned}$$

Taking the average over i and using (2.2), we get

$$\begin{aligned}
\frac{d}{dt} \frac{1}{n} \sum_{i=1}^n \left| X_i^n(t) - \bar{X}_{\frac{i}{n}}(t) \right|^2 &\leq -2\kappa \frac{1}{n} \sum_{i=1}^n \left| X_i^n(t) - \bar{X}_{\frac{i}{n}}(t) \right|^2 \\
&\quad + 2 \left(\frac{1}{n} \sum_{i=1}^n \left| X_i^n(t) - \bar{X}_{\frac{i}{n}}(t) \right|^2 \right)^{1/2} \left(\frac{1}{n} \sum_{i=1}^n \left| \mathcal{R}_t^{n,i} \right| + \frac{C_1}{n} \left(1 + \left| \bar{X}_{\frac{i}{n}}(t) \right| \right) \right)^2 \Big)^{1/2}.
\end{aligned}$$

So

$$\begin{aligned}
& \frac{d}{dt} \left(\frac{1}{n} \sum_{i=1}^n \left| X_i^n(t) - \bar{X}_{\frac{i}{n}}(t) \right|^2 \right)^{1/2} \\
&= \frac{1}{2} \left(\frac{1}{n} \sum_{i=1}^n \left| X_i^n(t) - \bar{X}_{\frac{i}{n}}(t) \right|^2 \right)^{-1/2} \frac{d}{dt} \frac{1}{n} \sum_{i=1}^n \left| X_i^n(t) - \bar{X}_{\frac{i}{n}}(t) \right|^2 \\
&\leq -\kappa \left(\frac{1}{n} \sum_{i=1}^n \left| X_i^n(t) - \bar{X}_{\frac{i}{n}}(t) \right|^2 \right)^{1/2} + \left(\frac{1}{n} \sum_{i=1}^n \left| \mathcal{R}_t^{n,i} \right| + \frac{C_1}{n} \left(1 + \left| \bar{X}_{\frac{i}{n}}(t) \right| \right) \right)^2 \Big)^{1/2}.
\end{aligned}$$

From this we have

$$\begin{aligned} \left(\frac{1}{n} \sum_{i=1}^n \left| X_i^n(t) - \bar{X}_{\frac{i}{n}}(t) \right|^2 \right)^{1/2} &\leq \int_0^t e^{-\kappa(t-s)} \left(\frac{1}{n} \sum_{i=1}^n \left| \mathcal{R}_s^{n,i} + \frac{C_1}{n} (1 + |\bar{X}_{\frac{i}{n}}(s)|) \right|^2 \right)^{1/2} ds \\ &\leq C_2 \left(\int_0^t e^{-\kappa(t-s)} \frac{1}{n} \sum_{i=1}^n \left| \mathcal{R}_s^{n,i} + \frac{1}{n} (1 + |\bar{X}_{\frac{i}{n}}(s)|) \right|^2 ds \right)^{1/2}, \end{aligned} \quad (3.2)$$

where the last line follows from Holder's inequality and Condition 2.1. Noting that

$$W_1(\nu_t^n, \bar{\nu}_t^n) \leq W_2(\nu_t^n, \bar{\nu}_t^n) \leq \left(\frac{1}{n} \sum_{i=1}^n \left| X_i^n(t) - \bar{X}_{\frac{i}{n}}(t) \right|^2 \right)^{1/2}$$

and

$$W_1(\nu_t^n, \hat{\mu}_t) \leq W_1(\bar{\nu}_t^n, \hat{\mu}_t) + W_1(\nu_t^n, \bar{\nu}_t^n),$$

we have the desired result for part (ii).

The proof of part (i) is the same as that of part (ii), except that Lemma 3.2 and all the relevant constants (such as C_1) are to be replaced by Lemma 3.1 and T -dependent constants (such as $C_1(T)$). \blacksquare

3.3. Exponential estimates. Recall $\mathcal{R}_t^{n,i}$ introduced in (3.1). In this section we will provide an exponential bound for the probability

$$p(n, t, \varepsilon) := \mathbb{P} \left(\int_0^t e^{-\kappa(t-s)} \frac{1}{n} \sum_{i=1}^n \left| \mathcal{R}_s^{n,i} + \frac{1}{n} (1 + |\bar{X}_{\frac{i}{n}}(s)|) \right|^2 ds > \varepsilon \right) \quad (3.3)$$

for $t \geq 0$ and $\varepsilon \in (0, \infty)$.

Proposition 3.2. (i) *There exist some $C(T), K_T \in (0, \infty)$ and $N_T \in \mathbb{N}$ such that*

$$\sup_{t \in [0, T]} p(n, t, \varepsilon) \leq C(T) e^{-K_T \sqrt{n\varepsilon}} \quad (3.4)$$

for all $\varepsilon > 0$ and $n \geq N_T/\sqrt{\varepsilon}$. If in addition (2.4) holds, then $\sqrt{n\varepsilon}$ in (3.4) could be improved to $n\varepsilon$.

(ii) *Suppose Condition 2.1 holds. Then there exist some $C, K \in (0, \infty)$ and $N_0 \in \mathbb{N}$ such that*

$$\sup_{t \geq 0} p(n, t, \varepsilon) \leq C e^{-K \sqrt{n\varepsilon}} \quad (3.5)$$

for all $\varepsilon > 0$ and $n \geq N_0/\sqrt{\varepsilon}$. If in addition (2.4) holds, then $\sqrt{n\varepsilon}$ in (3.5) could be improved to $n\varepsilon$.

Proof. We first prove part (ii). Using the union bound, we have

$$p(n, t, \varepsilon) \leq \sum_{i=1}^n \mathbb{P} \left(\int_0^t e^{-\kappa(t-s)} \left| \mathcal{R}_s^{n,i} + \frac{1}{n} (1 + |\bar{X}_{\frac{i}{n}}(s)|) \right|^2 ds > \varepsilon \right).$$

Now we fix $i \in \mathbb{N}$. Writing $\mathcal{R}_s^{n,i} = \tilde{\mathcal{T}}_s^{n,i} + \mathcal{T}_s^{n,i}$, where

$$\tilde{\mathcal{T}}_s^{n,i} := \frac{1}{n} \sum_{j=1}^n \left(\xi_{ij}^n - G\left(\frac{i}{n}, \frac{j}{n}\right) \right) b(\bar{X}_{\frac{i}{n}}(s), 0), \quad (3.6)$$

$$\begin{aligned} \mathcal{T}_s^{n,i} := \frac{1}{n} \sum_{j=1}^n & \left[\left(\xi_{ij}^n - G\left(\frac{i}{n}, \frac{j}{n}\right) \right) \left(b(\bar{X}_{\frac{i}{n}}(s), \bar{X}_{\frac{j}{n}}(s)) - b(\bar{X}_{\frac{i}{n}}(s), 0) \right) \right. \\ & \left. + G\left(\frac{i}{n}, \frac{j}{n}\right) \left(b(\bar{X}_{\frac{i}{n}}(s), \bar{X}_{\frac{j}{n}}(s)) - \int_{\mathbb{R}^d} b(\bar{X}_{\frac{i}{n}}(s), x) \bar{\mu}_{\frac{j}{n},s}(dx) \right) \right], \end{aligned} \quad (3.7)$$

we have

$$\left| |\mathcal{R}_s^{n,i}| + \frac{1}{n} \left(1 + |\bar{X}_{\frac{i}{n}}(s)| \right) \right|^2 \leq 4 \left(|\tilde{\mathcal{T}}_s^{n,i}|^2 + |\mathcal{T}_s^{n,i}|^2 + \frac{1}{n^2} + \frac{1}{n^2} |\bar{X}_{\frac{i}{n}}(s)|^2 \right).$$

Therefore

$$\begin{aligned} & \mathbb{P} \left(\int_0^t e^{-\kappa(t-s)} \left| |\mathcal{R}_s^{n,i}| + \frac{1}{n} \left(1 + |\bar{X}_{\frac{i}{n}}(s)| \right) \right|^2 ds > \varepsilon \right) \\ & \leq \mathbb{P} \left(\int_0^t e^{-\kappa(t-s)} |\tilde{\mathcal{T}}_s^{n,i}|^2 ds > \varepsilon/16 \right) + \mathbb{P} \left(\int_0^t e^{-\kappa(t-s)} |\mathcal{T}_s^{n,i}|^2 ds > \varepsilon/16 \right) \\ & \quad + \mathbb{P} \left(\int_0^t e^{-\kappa(t-s)} \frac{1}{n^2} ds > \varepsilon/16 \right) + \mathbb{P} \left(\int_0^t e^{-\kappa(t-s)} \frac{1}{n^2} |\bar{X}_{\frac{i}{n}}(s)|^2 ds > \varepsilon/16 \right) \\ & =: p(n, t, \varepsilon, i, 1) + p(n, t, \varepsilon, i, 2) + p(n, t, \varepsilon, i, 3) + p(n, t, \varepsilon, i, 4), \end{aligned}$$

and hence

$$p(n, t, \varepsilon) \leq \sum_{i=1}^n [p(n, t, \varepsilon, i, 1) + p(n, t, \varepsilon, i, 2) + p(n, t, \varepsilon, i, 3) + p(n, t, \varepsilon, i, 4)]. \quad (3.8)$$

Next we analyze each term. For $p(n, t, \varepsilon, i, 4)$, applying Markov's inequality and Jensen's inequality, we have

$$\begin{aligned} p(n, t, \varepsilon, i, 4) & \leq e^{-\theta\varepsilon/16} \mathbb{E} \left[\exp \left(\theta \int_0^t e^{-\kappa(t-s)} \frac{1}{n^2} |\bar{X}_{\frac{i}{n}}(s)|^2 ds \right) \right] \\ & \leq e^{-\theta\varepsilon/16} \mathbb{E} \left[\int_0^t C_1 e^{-\kappa(t-s)} \exp \left(\frac{C_1 \theta}{n^2} |\bar{X}_{\frac{i}{n}}(s)|^2 \right) ds \right] \end{aligned}$$

for each $\theta > 0$. Taking $\theta = n^2(\kappa \wedge \theta_0)/8C_1$ and using Lemma 3.3(ii), we have

$$p(n, t, \varepsilon, i, 4) \leq C_2 e^{-C_3 n^2 \varepsilon}.$$

For $p(n, t, \varepsilon, i, 3)$, we have

$$p(n, t, \varepsilon, i, 3) = 0$$

whenever $n > C_4/\sqrt{\varepsilon}$. For $p(n, t, \varepsilon, i, 2)$, we will show in Lemma 3.10(ii) below that

$$p(n, t, \varepsilon, i, 2) \leq C_5 e^{-C_6 n \varepsilon}.$$

Combining these with (3.8) and Lemma 3.7(ii) below gives

$$p(n, t, \varepsilon) \leq C_7 n e^{-C_8 \sqrt{n\varepsilon}} \leq C_9 e^{-C_{10} \sqrt{n\varepsilon}},$$

namely (3.5) holds, for $n > C_{11}/\sqrt{\varepsilon}$. It also follows from Lemma 3.7(ii) that if in addition (2.4) holds, then the above $\sqrt{n\varepsilon}$ could be improved to $n\varepsilon$. This gives part (ii).

The proof of part (i) is the same as that of part (ii), except that Lemmas 3.3(ii), 3.7(ii), 3.10(ii) and all the relevant constants (such as C_1) are to be replaced by Lemmas 3.3(i), 3.7(i), 3.10(i), and T -dependent constants (such as $C_1(T)$). \blacksquare

In the following lemma we estimate the probability $p(n, t, \varepsilon, i, 1)$. Recall that

$$p(n, t, \varepsilon, i, 1) = \mathbb{P} \left(\int_0^t e^{-\kappa(t-s)} |\tilde{\mathcal{T}}_s^{n,i}|^2 ds > \varepsilon/16 \right),$$

where $\tilde{\mathcal{T}}_s^{n,i}$ was introduced in (3.6).

Lemma 3.7. (i) For every $T \in (0, \infty)$, there exist some $C(T), K_T \in (0, \infty)$ such that

$$\max_{i=1, \dots, n} \sup_{t \in [0, T]} p(n, t, \varepsilon, i, 1) \leq C(T) e^{-K_T \sqrt{n\varepsilon}}, \quad \forall n \in \mathbb{N}, \varepsilon > 0. \quad (3.9)$$

If in addition (2.4) holds, then $\sqrt{n\varepsilon}$ in (3.9) could be improved to $n\varepsilon$.

(ii) Suppose that Condition 2.1 holds. Then there exist some $C, K \in (0, \infty)$ such that

$$\max_{i=1, \dots, n} \sup_{t \geq 0} p(n, t, \varepsilon, i, 1) \leq C e^{-K \sqrt{n\varepsilon}}, \quad \forall n \in \mathbb{N}, \varepsilon > 0. \quad (3.10)$$

If in addition (2.4) holds, then $\sqrt{n\varepsilon}$ in (3.10) could be improved to $n\varepsilon$.

Proof. We first prove part (ii). Fix n, t, ε, i . Applying Markov's inequality, we have

$$\begin{aligned} p(n, t, \varepsilon, i, 1) &= \mathbb{P} \left(\int_0^t e^{-\kappa(t-s)} \left| \frac{1}{n} \sum_{j=1}^n \left(\xi_{ij}^n - G\left(\frac{i}{n}, \frac{j}{n}\right) \right) b(\bar{X}_{\frac{i}{n}}(s), 0) \right|^2 ds > \varepsilon/16 \right) \\ &= \mathbb{P} \left(\left| \frac{1}{n} \sum_{j=1}^n \left(\xi_{ij}^n - G\left(\frac{i}{n}, \frac{j}{n}\right) \right) \right| \sqrt{\int_0^t e^{-\kappa(t-s)} |b(\bar{X}_{\frac{i}{n}}(s), 0)|^2 ds} > \sqrt{\varepsilon}/4 \right) \\ &\leq e^{-\theta \sqrt{\varepsilon}/4} \mathbb{E} \left[\exp \left(\theta \left| \frac{1}{n} \sum_{j=1}^n \left(\xi_{ij}^n - G\left(\frac{i}{n}, \frac{j}{n}\right) \right) \right| \sqrt{\int_0^t e^{-\kappa(t-s)} |b(\bar{X}_{\frac{i}{n}}(s), 0)|^2 ds} \right) \right] \end{aligned} \quad (3.11)$$

for each $\theta > 0$. Since $e^{|x|} \leq e^x + e^{-x}$, we have

$$\begin{aligned} &\mathbb{E} \left[\exp \left(\theta \left| \frac{1}{n} \sum_{j=1}^n \left(\xi_{ij}^n - G\left(\frac{i}{n}, \frac{j}{n}\right) \right) \right| \sqrt{\int_0^t e^{-\kappa(t-s)} |b(\bar{X}_{\frac{i}{n}}(s), 0)|^2 ds} \right) \right] \\ &\leq \mathbb{E} \left[\exp \left(\frac{\theta}{n} \sum_{j=1}^n \left(\xi_{ij}^n - G\left(\frac{i}{n}, \frac{j}{n}\right) \right) \sqrt{\int_0^t e^{-\kappa(t-s)} |b(\bar{X}_{\frac{i}{n}}(s), 0)|^2 ds} \right) \right] \\ &\quad + \mathbb{E} \left[\exp \left(-\frac{\theta}{n} \sum_{j=1}^n \left(\xi_{ij}^n - G\left(\frac{i}{n}, \frac{j}{n}\right) \right) \sqrt{\int_0^t e^{-\kappa(t-s)} |b(\bar{X}_{\frac{i}{n}}(s), 0)|^2 ds} \right) \right]. \end{aligned}$$

Since $0 \leq \xi_{ij}^n \leq 1$ and $\mathbb{E} \left[\xi_{ij}^n - G(\frac{i}{n}, \frac{j}{n}) \mid \bar{X}_{\frac{i}{n}} \right] = 0$, we can condition on $\bar{X}_{i/n}$ and apply Hoeffding's lemma to obtain that

$$\begin{aligned}
& \mathbb{E} \left[\exp \left(\pm \frac{\theta}{n} \sum_{j=1}^n \left(\xi_{ij}^n - G(\frac{i}{n}, \frac{j}{n}) \right) \sqrt{\int_0^t e^{-\kappa(t-s)} \left| b(\bar{X}_{\frac{i}{n}}(s), 0) \right|^2 ds} \right) \right] \\
&= \mathbb{E} \left\{ \prod_{j=1}^n \mathbb{E} \left[\exp \left(\pm \frac{\theta}{n} \left(\xi_{ij}^n - G(\frac{i}{n}, \frac{j}{n}) \right) \sqrt{\int_0^t e^{-\kappa(t-s)} \left| b(\bar{X}_{\frac{i}{n}}(s), 0) \right|^2 ds} \right) \mid \bar{X}_{\frac{i}{n}} \right] \right\} \\
&\leq \mathbb{E} \left[\exp \left(\frac{\theta^2}{8n} \int_0^t e^{-\kappa(t-s)} \left| b(\bar{X}_{\frac{i}{n}}(s), 0) \right|^2 ds \right) \right] \\
&\leq \mathbb{E} \left[\int_0^t C_1 e^{-\kappa(t-s)} \exp \left(\frac{C_1 \theta^2}{8n} \left| b(\bar{X}_{\frac{i}{n}}(s), 0) \right|^2 \right) ds \right] \\
&\leq \mathbb{E} \left[\int_0^t C_1 e^{-\kappa(t-s)} \exp \left(\frac{C_2 \theta^2}{8n} (1 + |\bar{X}_{\frac{i}{n}}(s)|^2) \right) ds \right],
\end{aligned}$$

where the second inequality follows from Jensen's inequality, and the last line uses the linear growth property of b . Combining these three estimates gives

$$p(n, t, \varepsilon, i, 1) \leq 2e^{-\theta\sqrt{\varepsilon}/4} \mathbb{E} \left[\int_0^t C_1 e^{-\kappa(t-s)} \exp \left(\frac{C_2 \theta^2}{8n} (1 + |\bar{X}_{\frac{i}{n}}(s)|^2) \right) ds \right].$$

Taking $\theta = \sqrt{n(\kappa \wedge \theta_0)}/C_2$ and using Lemma 3.3(ii), we have the desired result (3.10).

Now we prove the strengthened version of (3.10) under the extra assumption that (2.4) holds, namely $\xi_{ij}^n = G(\frac{i}{n}, \frac{j}{n})$ or $b(\cdot, 0)$ is bounded. If $\xi_{ij}^n = G(\frac{i}{n}, \frac{j}{n})$, from (3.11) we have $p(n, t, \varepsilon, i, 1) = 0$ and the desired result clearly holds. If $b(\cdot, 0)$ is bounded, from (3.11) we have

$$\begin{aligned}
p(n, t, \varepsilon, i, 1) &\leq \mathbb{P} \left(\left| \frac{1}{n} \sum_{j=1}^n \left(\xi_{ij}^n - G(\frac{i}{n}, \frac{j}{n}) \right) \right|^2 > C_3 \varepsilon \right) \\
&\leq e^{-\theta C_3 \varepsilon} \mathbb{E} \left[\exp \left(\theta \left| \frac{1}{n} \sum_{j=1}^n \left(\xi_{ij}^n - G(\frac{i}{n}, \frac{j}{n}) \right) \right|^2 \right) \right]
\end{aligned}$$

for each $\theta > 0$. Letting Z be an independent standard normal random variable, we have

$$\begin{aligned}
\mathbb{E} \left[\exp \left(\theta \left| \frac{1}{n} \sum_{j=1}^n \left(\xi_{ij}^n - G(\frac{i}{n}, \frac{j}{n}) \right) \right|^2 \right) \right] &= \mathbb{E} \left[\exp \left(Z \frac{\sqrt{2\theta}}{n} \sum_{j=1}^n \left(\xi_{ij}^n - G(\frac{i}{n}, \frac{j}{n}) \right) \right) \right] \\
&= \mathbb{E} \left\{ \prod_{j=1}^n \mathbb{E} \left[\exp \left(Z \frac{\sqrt{2\theta}}{n} \left(\xi_{ij}^n - G(\frac{i}{n}, \frac{j}{n}) \right) \right) \mid Z \right] \right\} \\
&\leq \mathbb{E} \left[\exp \left(\frac{\theta Z^2}{4n} \right) \right].
\end{aligned}$$

where the last line uses Hoeffding's lemma again. It then follows from the formula of the moment generating function of Z^2 that

$$\mathbb{E} \left[\exp \left(\theta \left| \frac{1}{n} \sum_{j=1}^n \left(\xi_{ij}^n - G\left(\frac{i}{n}, \frac{j}{n}\right) \right) \right|^2 \right) \right] \leq \left(1 - \frac{\theta}{2n} \right)^{-1/2}.$$

Taking $\theta = n$ gives the strengthened version of (3.10).

The proof of part (i) is the same as that of part (ii), except that Lemma 3.3(ii) and all the relevant constants (such as C_1) are to be replaced by Lemma 3.3(i) and T -dependent constants (such as $C_1(T)$). \blacksquare

For the analysis of $p(n, t, \varepsilon, i, 2)$, we will need the following standard property of sub-Gaussian random vectors. A proof is provided in Appendix B for completeness.

Lemma 3.8. *Let Y be an \mathbb{R}^d -valued random variable with $\mathbb{E}Y = 0$ and $\mathbb{E}[e^{|Y|^2/a}] \leq 2$ for some $a \in (0, \infty)$. Then for any $\lambda \in \mathbb{R}^d$,*

$$\mathbb{E}[e^{\lambda \cdot Y}] \leq \exp \left(\frac{5a}{2} |\lambda|^2 \right).$$

Recall $\mathcal{T}_t^{n,i}$ defined in (3.7). Let

$$\begin{aligned} \mathcal{T}_t^{n,i,j} &:= \left(\xi_{ij}^n - G\left(\frac{i}{n}, \frac{j}{n}\right) \right) \left(b(\bar{X}_{\frac{i}{n}}(t), \bar{X}_{\frac{j}{n}}(t)) - b(\bar{X}_{\frac{i}{n}}(t), 0) \right) \\ &\quad + G\left(\frac{i}{n}, \frac{j}{n}\right) \left(b(\bar{X}_{\frac{i}{n}}(t), \bar{X}_{\frac{j}{n}}(t)) - \int_{\mathbb{R}^d} b(\bar{X}_{\frac{i}{n}}(t), x) \bar{\mu}_{\frac{j}{n},t}(dx) \right). \end{aligned} \quad (3.12)$$

Note that $\mathcal{T}_t^{n,i} = \frac{1}{n} \sum_{j=1}^n \mathcal{T}_t^{n,i,j}$. We will estimate $p(n, t, \varepsilon, i, 2)$ by proving the following sub-Gaussian bounds of $\mathcal{T}_t^{n,i,j}$.

Lemma 3.9. (i) *For every $T \in (0, \infty)$, there exists some $C(T) \in (0, \infty)$ such that*

$$\max_{i=1,\dots,n} \sup_{t \in [0,T]} \mathbb{E} \left[\exp \left(s \left| \frac{1}{n} \sum_{j \neq i} \mathcal{T}_t^{n,i,j} \right|^2 \right) \right] \leq \left(1 - \frac{C(T)s}{n} \right)^{-d/2}, \quad \forall s \in \left(0, \frac{n}{C(T)} \right), n \in \mathbb{N}.$$

(ii) *Suppose Condition 2.1 holds. Then there exists some $C \in (0, \infty)$ such that*

$$\max_{i=1,\dots,n} \sup_{t \geq 0} \mathbb{E} \left[\exp \left(s \left| \frac{1}{n} \sum_{j \neq i} \mathcal{T}_t^{n,i,j} \right|^2 \right) \right] \leq \left(1 - \frac{Cs}{n} \right)^{-d/2}, \quad \forall s \in \left(0, \frac{n}{C} \right), n \in \mathbb{N}.$$

Proof. We first prove part (ii). Fix t, n, i and $j \neq i$. Using the Lipschitz property of b and Lemma 3.3(ii), we have

$$|\mathcal{T}_t^{n,i,j}|^2 \leq C_1(1 + |\bar{X}_{\frac{j}{n}}(t)|^2).$$

It then follows from Lemma 3.3(ii) that

$$\mathbb{E} \left[\exp \left(|\mathcal{T}_t^{n,i,j}|^2 / C_2 \right) \mid \bar{X}_{\frac{i}{n}}(t) \right] \leq 2.$$

Using this and the fact that $\mathbb{E} \left[\mathcal{T}_t^{n,i,j} \mid \bar{X}_{\frac{i}{n}}(t) \right] = 0$, we can apply Lemma 3.8 to get

$$\mathbb{E} \left[\exp(\lambda \cdot \mathcal{T}_t^{n,i,j}) \mid \bar{X}_{\frac{i}{n}}(t) \right] \leq \exp \left(\frac{5C_2}{2} |\lambda|^2 \right)$$

for each $\lambda \in \mathbb{R}^d$. Let Z be an independent d -dimensional standard normal random vector. Using the independence of $(Z, \{\xi_{ij}^n\}, \{\bar{X}_{\frac{i}{n}}(t)\})$, we have that for any $s > 0$,

$$\begin{aligned} \mathbb{E}\left[\exp\left(s\left|\frac{1}{n}\sum_{j \neq i}\mathcal{T}_t^{n,i,j}\right|^2\right)\right] &= \mathbb{E}\left[\exp\left(Z \cdot \frac{\sqrt{2s}}{n}\sum_{j \neq i}\mathcal{T}_t^{n,i,j}\right)\right] \\ &= \mathbb{E}\left\{\prod_{j \neq i}\mathbb{E}\left[\exp\left(Z \cdot \frac{\sqrt{2s}}{n}\mathcal{T}_t^{n,i,j}\right) \mid Z, \bar{X}_{\frac{i}{n}}(t)\right]\right\} \\ &\leq \mathbb{E}\left[\exp\left(\frac{5C_2}{2}\left|\frac{\sqrt{2s}Z}{n}\right|^2(n-1)\right)\right]. \end{aligned}$$

It then follows from the formula of the moment generating function of $|Z|^2$ that

$$\mathbb{E}\left[\exp\left(s\left|\frac{1}{n}\sum_{j \neq i}\mathcal{T}_t^{n,i,j}\right|^2\right)\right] \leq \left(1 - \frac{C_3 s}{n}\right)^{-d/2},$$

for all $t \geq 0$, $n \in \mathbb{N}$, $i = 1, \dots, n$, $s < n/C_3$. This gives part (ii).

The proof of part (i) is the same as that of part (ii), except that Lemma 3.3(ii) and all the relevant constants (such as C_1) are to be replaced by Lemma 3.3(i) and T -dependent constants (such as $C_1(T)$). \blacksquare

Now we show the following estimate for $p(n, t, \varepsilon, i, 2)$. Recall that

$$p(n, t, \varepsilon, i, 2) = \mathbb{P}\left(\int_0^t e^{-\kappa(t-s)} |\mathcal{T}_s^{n,i}|^2 ds > \varepsilon/16\right),$$

where $\mathcal{T}_s^{n,i}$ was introduced in (3.7).

Lemma 3.10. (i) For every $T \in (0, \infty)$, there exist some $C(T), K_T \in (0, \infty)$ such that

$$\max_{i=1, \dots, n} \sup_{t \in [0, T]} p(n, t, \varepsilon, i, 2) \leq C(T) e^{-K_T n \varepsilon}, \quad \forall n \in \mathbb{N}, \varepsilon > 0.$$

(ii) Suppose Condition 2.1 holds. Then there exist some $C, K \in (0, \infty)$ such that

$$\max_{i=1, \dots, n} \sup_{t \geq 0} p(n, t, \varepsilon, i, 2) \leq C e^{-K n \varepsilon}, \quad \forall n \in \mathbb{N}, \varepsilon > 0.$$

Proof. We first prove part (ii). Fix n, t, ε, i . Applying Markov's inequality and Jensen's inequality, we have

$$\begin{aligned} p(n, t, \varepsilon, i, 2) &\leq e^{-\theta \varepsilon/16} \mathbb{E}\left[\exp\left(\theta \int_0^t e^{-\kappa(t-s)} |\mathcal{T}_s^{n,i}|^2 ds\right)\right] \\ &\leq e^{-\theta \varepsilon/16} \mathbb{E}\left[\int_0^t C_1 e^{-\kappa(t-s)} \exp(C_1 \theta |\mathcal{T}_s^{n,i}|^2) ds\right] \end{aligned}$$

for each $\theta > 0$. Recall $\mathcal{T}_t^{n,i,j}$ in (3.12). Using Holder's inequality and Lemma 3.9(ii), we have

$$\begin{aligned} \mathbb{E} \left[\exp \left(C_1 \theta |\mathcal{T}_t^{n,i}|^2 \right) \right] &= \mathbb{E} \left[\exp \left(C_1 \theta \left| \frac{1}{n} \sum_{j=1}^n \mathcal{T}_t^{n,i,j} \right|^2 \right) \right] \\ &\leq \mathbb{E} \left[\exp \left(2C_1 \theta \left| \frac{1}{n} \mathcal{T}_t^{n,i,i} \right|^2 + 2C_1 \theta \left| \frac{1}{n} \sum_{j \neq i}^n \mathcal{T}_t^{n,i,j} \right|^2 \right) \right] \\ &\leq \sqrt{\mathbb{E} \left[\exp \left(4C_1 \theta \left| \frac{1}{n} \mathcal{T}_t^{n,i,i} \right|^2 \right) \right]} \sqrt{\mathbb{E} \left[\exp \left(4C_1 \theta \left| \frac{1}{n} \sum_{j \neq i}^n \mathcal{T}_t^{n,i,j} \right|^2 \right) \right]} \\ &\leq \sqrt{\mathbb{E} \left[\exp \left(4C_1 \theta \left| \frac{1}{n} \mathcal{T}_t^{n,i,i} \right|^2 \right) \right]} \left(1 - \frac{C_2 \theta}{n} \right)^{-d/2}. \end{aligned}$$

Now taking $\theta = \delta n$ for small enough $\delta > 0$ and using Lemma 3.3(ii), we have

$$p(n, t, \varepsilon, i, 2) \leq C_3 e^{-C_4 n \varepsilon}.$$

This gives part (ii).

The proof of part (i) is the same as that of part (ii), except that Lemmas 3.3(ii), 3.9(ii) and all the relevant constants (such as C_1) are to be replaced by Lemmas 3.3(i), 3.9(i) and T -dependent constants (such as $C_1(T)$). \blacksquare

3.4. Proofs of Theorems 2.1 and 2.2. Finally we need the following result on the concentration estimates of the distance $W_1(\bar{\nu}_t^n, \hat{\mu}_t)$ in Proposition 3.1.

Lemma 3.11. (i) For every $T \in (0, \infty)$, there exists some $K_T \in (0, \infty)$ such that for any $d' > d$, there exists some $N_T \in \mathbb{N}$ such that

$$\sup_{0 \leq t \leq T} \mathbb{P}(W_1(\bar{\nu}_t^n, \hat{\mu}_t) > \varepsilon) \leq \exp(-K_T n \varepsilon^2),$$

for all $\varepsilon > 0$ and $n \geq N_T \max(\varepsilon^{-(d'+2)}, 1)$.

(ii) Suppose Condition 2.1 holds. Then there exists some $K \in (0, \infty)$ such that for any $d' > d$, there exists some $N_0 \in \mathbb{N}$ such that

$$\sup_{t \geq 0} \mathbb{P}(W_1(\bar{\nu}_t^n, \hat{\mu}_t) > \varepsilon) \leq \exp(-K n \varepsilon^2),$$

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

Proof. We first prove part (ii). Let $\bar{\mu}_t^n := \frac{1}{n} \sum_{i=1}^n \bar{\mu}_{\frac{i}{n}, t}$. From Lemma 3.3(ii) we know that

$$\sup_{t \geq 0} \sup_{n \in \mathbb{N}} \max_{i=1, \dots, n} \int_{\mathbb{R}^d} e^{\alpha |x|^2} \bar{\mu}_{\frac{i}{n}, t}(dx) < \infty$$

for some $\alpha > 0$. Then there is some $\lambda > 0$ such that $\bar{\mu}_t^n$ satisfies the Talagrand (transport) $T_1(\lambda)$ inequality (see (C.1) in Proposition C.1) for each $n \in \mathbb{N}$ and $t \geq 0$ (by e.g. [10, Corollary 2.4]). It then follows from Proposition C.1 that for any $d' > d$, there exists some $N_0 \in \mathbb{N}$ such that

$$\sup_{t \geq 0} \mathbb{P}(W_1(\bar{\nu}_t^n, \bar{\mu}_t^n) > \varepsilon/2) \leq \exp(-4\lambda n \varepsilon^2),$$

for all $\varepsilon > 0$ and $n \geq N_0 \max(\varepsilon^{-(d'+2)}, 1)$. From Lemma 3.2 we know that

$$\sup_{t \geq 0} W_1(\bar{\mu}_t^n, \hat{\mu}_t) \leq \varepsilon/2$$

for all $n \geq C_1/\varepsilon$. Combining these two estimates with the triangle inequality gives the desired result.

The proof of part (i) is the same as that of part (ii), except that Lemmas 3.2, 3.3(ii) and all the relevant constants (such as C_1) are to be replaced by Lemmas 3.1, 3.3(i) and T -dependent constants (such as $C_1(T)$). \blacksquare

Now we are ready to prove Theorem 2.2.

Proof of Theorem 2.2. Using Proposition 3.1(ii) and recalling (3.3), we have

$$\mathbb{P}(W_1(\nu_t^n, \hat{\mu}_t) > \varepsilon) \leq \mathbb{P}(W_1(\bar{\nu}_t^n, \hat{\mu}_t) > \varepsilon/2) + p(n, t, C_1 \varepsilon^2).$$

The result then follows on combining this with Lemma 3.11(ii) and Proposition 3.2(ii). \blacksquare

Finally we prove Theorem 2.1.

Proof of Theorem 2.1. Fix $T, \varepsilon \in (0, \infty)$ and $d' > d$. We will suppress the dependency of constants $C(T)$'s on T . From Proposition 3.1(i) we have

$$\begin{aligned} W_1(\nu_t^n, \hat{\mu}_t) &\leq W_1(\bar{\nu}_t^n, \hat{\mu}_t) + C_1 \left(\int_0^t e^{-\kappa(t-s)} \frac{1}{n} \sum_{i=1}^n \left| |\mathcal{R}_s^{n,i}| + \frac{1}{n} \left(1 + |\bar{X}_{\frac{i}{n}}(s)| \right) \right|^2 ds \right)^{1/2} \\ &\leq W_1(\bar{\nu}_t^n, \hat{\mu}_t) + C_1 e^{|\kappa|(T-t)/2} \left(\int_0^T e^{-\kappa(T-s)} \frac{1}{n} \sum_{i=1}^n \left| |\mathcal{R}_s^{n,i}| + \frac{1}{n} \left(1 + |\bar{X}_{\frac{i}{n}}(s)| \right) \right|^2 ds \right)^{1/2}. \end{aligned}$$

Therefore

$$\begin{aligned} &\mathbb{P} \left(\sup_{0 \leq t \leq T} W_1(\nu_t^n, \hat{\mu}_t) > \varepsilon \right) \\ &\leq \mathbb{P} \left(\sup_{0 \leq t \leq T} W_1(\bar{\nu}_t^n, \hat{\mu}_t) > \frac{\varepsilon}{2} \right) + \mathbb{P} \left(\int_0^T e^{-\kappa(T-s)} \frac{1}{n} \sum_{i=1}^n \left| |\mathcal{R}_s^{n,i}| + \frac{1}{n} \left(1 + |\bar{X}_{\frac{i}{n}}(s)| \right) \right|^2 ds > C_2 \varepsilon^2 \right) \\ &=: \tilde{p}(n, T, \varepsilon, 1) + \tilde{p}(n, T, \varepsilon, 2). \end{aligned} \tag{3.13}$$

We first analyze $\tilde{p}(n, T, \varepsilon, 1)$. Let $\Delta > 0$ (to be chosen later) and $M := \lceil \frac{T}{\Delta} \rceil$. Decompose the interval $[0, T]$ as

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

Then, by triangle inequality, we have

$$\begin{aligned}
\tilde{p}(n, T, \varepsilon, 1) &\leq \mathbb{P} \left(\sup_{h=0, \dots, M-1} \sup_{h\Delta \leq t \leq (h+1)\Delta} W_1(\bar{\nu}_t^n, \hat{\mu}_t) > \frac{\varepsilon}{2} \right) \\
&\leq \mathbb{P} \left(\sup_{h=0, \dots, M-1} \sup_{h\Delta \leq t \leq (h+1)\Delta} W_1(\bar{\nu}_t^n, \bar{\nu}_{h\Delta}^n) > \frac{\varepsilon}{6} \right) \\
&\quad + \mathbb{P} \left(\sup_{h=0, \dots, M-1} W_1(\bar{\nu}_{h\Delta}^n, \hat{\mu}_{h\Delta}) > \frac{\varepsilon}{6} \right) \\
&\quad + \mathbb{P} \left(\sup_{h=0, \dots, M-1} \sup_{h\Delta \leq t \leq (h+1)\Delta} W_1(\hat{\mu}_{h\Delta}, \hat{\mu}_t) > \frac{\varepsilon}{6} \right). \tag{3.14}
\end{aligned}$$

For the first term on the right hand side of (3.14), from Lemma 3.5 we have

$$\mathbb{P} \left(\sup_{h\Delta \leq t \leq (h+1)\Delta} W_1(\bar{\nu}_t^n, \bar{\nu}_{h\Delta}^n) > \frac{\varepsilon}{6} \right) \leq \exp(-n(C_3\varepsilon^2 - C_4\Delta))$$

for all $h = 0, \dots, M-1$. So

$$\mathbb{P} \left(\sup_{h=0, \dots, M-1} \sup_{h\Delta \leq t \leq (h+1)\Delta} W_1(\bar{\nu}_t^n, \bar{\nu}_{h\Delta}^n) > \frac{\varepsilon}{6} \right) \leq M \exp(-n(C_3\varepsilon^2 - C_4\Delta)).$$

Taking $\Delta \leq C_3\varepsilon^2/2C_4$, we have $M \leq C_5(T/\varepsilon^2 + 1)$ and

$$\mathbb{P} \left(\sup_{h=0, \dots, M-1} \sup_{h\Delta \leq t \leq (h+1)\Delta} W_1(\bar{\nu}_t^n, \bar{\nu}_{h\Delta}^n) > \frac{\varepsilon}{6} \right) \leq C_5 \left(\frac{T}{\varepsilon^2} + 1 \right) \exp \left(-\frac{C_3}{2} n \varepsilon^2 \right). \tag{3.15}$$

For the second term on the right hand side of (3.14), by Lemma 3.11(i), there exist some $K \in (0, \infty)$ and $N_0 \in \mathbb{N}$ such that

$$\mathbb{P}(W_1(\bar{\nu}_{h\Delta}^n, \hat{\mu}_{h\Delta}) > \varepsilon/6) \leq e^{-Kn\varepsilon^2}$$

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

$$\begin{aligned}
\mathbb{P} \left(\sup_{h=0, \dots, M-1} W_1(\bar{\nu}_{h\Delta}^n, \hat{\mu}_{h\Delta}) > \frac{\varepsilon}{6} \right) &\leq \sum_{h=0}^{M-1} \mathbb{P}(W_1(\bar{\nu}_{h\Delta}^n, \hat{\mu}_{h\Delta}) > \varepsilon/6) \\
&\leq M e^{-Kn\varepsilon^2} \leq C_5 \left(\frac{T}{\varepsilon^2} + 1 \right) \exp(-Kn\varepsilon^2). \tag{3.16}
\end{aligned}$$

For the last term on the right hand side of (3.14), taking $\Delta \leq C_6\varepsilon^2$ for some $C_6 \in (0, \infty)$ small enough, from Lemma 3.6 we have

$$h\Delta \leq t \leq (h+1)\Delta \Rightarrow W_1(\hat{\mu}_t, \hat{\mu}_{h\Delta}) \leq \varepsilon/6,$$

and hence

$$\mathbb{P} \left(\sup_{h=0, \dots, M-1} \sup_{h\Delta \leq t \leq (h+1)\Delta} W_1(\hat{\mu}_{h\Delta}, \hat{\mu}_t) > \frac{\varepsilon}{6} \right) = 0. \tag{3.17}$$

Combining (3.14)–(3.17) and taking $\Delta = \min(C_3\varepsilon^2/2C_4, C_6\varepsilon^2)$, we have

$$\tilde{p}(n, T, \varepsilon, 1) \leq 2C_5 \left(\frac{T}{\varepsilon^2} + 1 \right) \exp(-C_7n\varepsilon^2) \tag{3.18}$$

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

Next we analyze $\tilde{p}(n, T, \varepsilon, 2)$ in (3.13). Using (3.3) and Proposition 3.2(i) we have

$$\tilde{p}(n, T, \varepsilon, 2) = p(n, T, C_2 \varepsilon^2) \leq C_8 e^{-C_9 \sqrt{n} \varepsilon}, \quad (3.19)$$

for all $\varepsilon > 0$ and $n \geq C_{10}/\varepsilon$.

Combining (3.13), (3.18) and (3.19), we have the desired result

$$\mathbb{P} \left(\sup_{0 \leq t \leq T} W_1(\nu_t^n, \hat{\mu}_t) > \varepsilon \right) \leq C_{11} \left(\frac{T}{\varepsilon^2} + 1 \right) e^{-C_{12} \sqrt{n} \varepsilon} \quad (3.20)$$

for all $\varepsilon > 0$ and $n \geq \max(N_0, C_{10}) \max(\varepsilon^{-(d'+2)}, 1)$. If in addition (2.4) holds, then using Proposition 3.2(i) we can improve $\sqrt{n} \varepsilon$ in (3.19) and (3.20) to $n \varepsilon^2$. This completes the proof. \blacksquare

4. ACKNOWLEDGMENT

We would like to thank the anonymous referees for their careful reading and many valuable suggestions on the paper.

APPENDIX A. PROOFS OF RESULTS IN SECTION 3.1

Proof of Lemma 3.3. The proof is similar to that of [11, Proposition 4.3] and hence we only provide a sketch.

(i) For each $t \in [0, T]$ and $u \in I$, using Proposition 2.1(i) and the linear growth properties of f and b , we have

$$|\bar{X}_u(t)| \leq |\bar{X}_u(0)| + \int_0^t C_1(T)(1 + |\bar{X}_u(s)|) ds + |B_u(t)|.$$

It then follows from Gronwall's inequality that

$$\sup_{t \in [0, T]} |\bar{X}_u(t)| \leq C_2(T) \left(|\bar{X}_u(0)| + 1 + \sup_{t \in [0, T]} |B_u(t)| \right).$$

Thus for $\theta > 0$,

$$\begin{aligned} \mathbb{E} \left[\sup_{t \in [0, T]} e^{\theta |\bar{X}_u(t)|^2} \right] &\leq \mathbb{E} \left[e^{C_3(T)\theta |\bar{X}_u(0)|^2} e^{C_3(T)\theta} e^{C_3(T)\theta \sup_{t \in [0, T]} |B_u(t)|^2} \right] \\ &\leq \left(\mathbb{E} \left[e^{2C_3(T)\theta |\bar{X}_u(0)|^2} \right] \right)^{1/2} e^{C_3(T)\theta} \left(\mathbb{E} \left[e^{2C_3(T)\theta \sup_{t \in [0, T]} |B_u(t)|^2} \right] \right)^{1/2}. \end{aligned}$$

The proof of part (i) follows by choosing $\theta > 0$ such that $\mathbb{E} \left[e^{2C_3(T)\theta |\bar{X}_u(0)|^2} \right] < \infty$ and $\mathbb{E} \left[e^{2C_3(T)\theta \sup_{t \in [0, T]} |B_u(t)|^2} \right] < \infty$.

(ii) Fix $\theta \in (0, (\kappa \wedge \theta_0)/4)$. Note that for the function $\phi(x) := e^{\theta |x|^2}$, we have $\nabla \phi(x) = 2\theta e^{\theta |x|^2} x$ and $\Delta \phi(x) = 2\theta e^{\theta |x|^2} (d + 2\theta |x|^2)$. It then follows from Itô's formula that

$$\begin{aligned} d e^{\theta |\bar{X}_u(t)|^2} &= e^{\theta |\bar{X}_u(t)|^2} \left[2\theta \bar{X}_u(t) \cdot \left(f(\bar{X}_u(t)) dt + \int_0^1 \int_{\mathbb{R}^d} b(\bar{X}_u(t), x) G(u, v) \bar{\mu}_{v,t}(dx) dv dt \right. \right. \\ &\quad \left. \left. + dB_u(t) \right) + \theta (d + 2\theta |\bar{X}_u(t)|^2) dt \right]. \end{aligned} \quad (A.1)$$

Using (2.1) we have

$$\begin{aligned}\bar{X}_u(t) \cdot f(\bar{X}_u(t)) &= (\bar{X}_u(t) - 0) \cdot (f(\bar{X}_u(t)) - f(0)) + \bar{X}_u(t) \cdot f(0) \\ &\leq -c_0|\bar{X}_u(t)|^2 + C_1|\bar{X}_u(t)| \leq -(c_0 - \frac{1}{2}K_b)|\bar{X}_u(t)|^2 + C_2.\end{aligned}\quad (\text{A.2})$$

Using Lemma 3.2 and the Lipschitz property of b we have

$$\bar{X}_u(t) \cdot \int_0^1 \int_{\mathbb{R}^d} b(\bar{X}_u(t), x) G(u, v) \bar{\mu}_{v,t}(dx) dv \leq K_b|\bar{X}_u(t)|^2 + C_3|\bar{X}_u(t)| \leq \frac{3}{2}K_b|\bar{X}_u(t)|^2 + C_4. \quad (\text{A.3})$$

Similar to the proof of [11, Proposition 4.3(ii)], one can verify that

$$\int_0^t e^{\theta|\bar{X}_u(s)|^2} \bar{X}_u(s) \cdot dB_u(s)$$

is a square integrable martingale for all $\theta \in (0, (\kappa \wedge \theta_0)/4)$. Then

$$\mathbb{E} \int_{t_1}^{t_2} e^{\theta|\bar{X}_u(t)|^2} 2\theta \bar{X}_u(t) \cdot dB_u(t) = 0, \quad \forall t_2 \geq t_1 \geq 0.$$

Combining this and (A.1)–(A.3) gives

$$d\mathbb{E} \left[e^{\theta|\bar{X}_u(t)|^2} \right] \leq \mathbb{E} \left[e^{\theta|\bar{X}_u(t)|^2} (A + B|\bar{X}_u(t)|^2) \right] dt,$$

where

$$A = \theta C_5, \quad B = 2\theta^2 - 2\kappa\theta.$$

Since $\theta < (\kappa \wedge \theta_0)/4 < \kappa$, we have $B < 0$. Therefore

$$d\mathbb{E} \left[e^{\theta|\bar{X}_u(t)|^2} \right] \leq C_6(\theta) \mathbb{E} \left[e^{\theta|\bar{X}_u(t)|^2} (C_7(\theta) - |\bar{X}_u(t)|^2) \right] dt. \quad (\text{A.4})$$

Decomposing the expectation on the right hand side of (A.4) according to the size of $|\bar{X}_u(t)|$, we get

$$d\mathbb{E} \left[e^{\theta|\bar{X}_u(t)|^2} \right] \leq \left(C_8(\theta) - C_9(\theta) \mathbb{E} \left[e^{\theta|\bar{X}_u(t)|^2} \right] \right) dt.$$

Since $\theta < (\kappa \wedge \theta_0)/4 < \theta_0$, $\mathbb{E} \left[e^{\theta|\bar{X}_u(0)|^2} \right] < \infty$. A standard estimate shows that $\sup_{t \geq 0} \mathbb{E} \left[e^{\theta|\bar{X}_u(t)|^2} \right] < \infty$, uniformly in $u \in I$. \blacksquare

Proof of Lemma 3.4. Fix $T \in (0, \infty)$, $0 \leq s \leq t \leq T$ and $u \in I$. We will suppress the dependency of constants $C(T)$'s on T . From (1.1) we have

$$\bar{X}_u(t) - \bar{X}_u(s) = \int_s^t \left(f(\bar{X}_u(r)) + \int_0^1 \int_{\mathbb{R}^d} b(\bar{X}_u(r), x) G(u, v) \bar{\mu}_{v,r}(dx) dv \right) dr + B_u(t) - B_u(s).$$

It then follows from the linear growth properties of f and b and Proposition 2.1(i) that

$$|\bar{X}_u(t) - \bar{X}_u(s)| \leq C_1 \int_s^t (1 + |\bar{X}_u(r)|) dr + |B_u(t) - B_u(s)|.$$

Using Proposition 2.1(i) and Holder's inequality we have

$$\begin{aligned}\mathbb{E}|\bar{X}_u(t) - \bar{X}_u(s)|^4 &\leq C_2(t-s)^3 \int_s^t (1 + \mathbb{E}|\bar{X}_u(r)|^4) dr + C_2\mathbb{E}|B_u(t) - B_u(s)|^4 \\ &\leq C_3(t-s)^4 + C_3(t-s)^2.\end{aligned}$$

This gives the first assertion.

For the last assertion, fix $t_0, \Delta \in [0, T]$, $t_0 \leq s \leq t \leq t_0 + \Delta$ and $u \in I$. Since

$$\begin{aligned} & \mathbb{E} \left[\sup_{t_0 \leq s \leq t \leq t_0 + \Delta} \exp(\theta |\bar{X}_u(t) - \bar{X}_u(s)|^2) \right] \\ & \leq \mathbb{E} \left[\sup_{t_0 \leq s \leq t \leq t_0 + \Delta} \exp(2\theta |\bar{X}_u(t) - \bar{X}_u(t_0)|^2 + 2\theta |\bar{X}_u(s) - \bar{X}_u(t_0)|^2) \right] \\ & \leq \mathbb{E} \left[\sup_{t_0 \leq t \leq t_0 + \Delta} \exp(4\theta |\bar{X}_u(t) - \bar{X}_u(t_0)|^2) \right] \end{aligned}$$

for $\theta > 0$, it suffices to show that

$$\mathbb{E} \left[\sup_{t_0 \leq t \leq t_0 + \Delta} \exp(\theta_T |\bar{X}_u(t) - \bar{X}_u(t_0)|^2) \right] \leq 1 + C(T)\Delta \quad (\text{A.5})$$

for some $\theta_T, C(T) > 0$. Let $\theta: [0, T] \rightarrow \mathbb{R}$ be a non-negative continuously differentiable function (to be chosen later) and write

$$Z_u(r) := e^{\theta(r)|\bar{X}_u(r) - \bar{X}_u(t_0)|^2}.$$

Using Itô's formula we have

$$\begin{aligned} Z_u(t) - 1 &= M_u(t) + \int_{t_0}^t Z_u(r) \left[2\theta(r)(\bar{X}_u(r) - \bar{X}_u(t_0)) \cdot \left(f(\bar{X}_u(r)) + \int_0^1 \int_{\mathbb{R}^d} b(\bar{X}_u(r), x) G(u, v) \bar{\mu}_{v,r}(dx) dv \right) \right. \\ &\quad \left. + \theta(r)(d + 2\theta(r)|\bar{X}_u(r) - \bar{X}_u(t_0)|^2) + \theta'(r)|\bar{X}_u(r) - \bar{X}_u(t_0)|^2 \right] dr, \end{aligned}$$

where

$$M_u(t) := \int_{t_0}^t 2Z_u(r)\theta(r)(\bar{X}_u(r) - \bar{X}_u(t_0)) \cdot dB_u(r).$$

Using Lemma 3.3(i) we can choose some small $\zeta \in (0, \infty)$ such that

$$\sup_{u \in I} \sup_{0 \leq r \leq T} \mathbb{E} e^{\zeta |\bar{X}_u(r) - \bar{X}_u(t_0)|^2} < \infty. \quad (\text{A.6})$$

The function $\theta(r)$ will be chosen below such that $\sup_{0 \leq r \leq T} \theta(r) < \zeta/2$. With such a choice of $\theta(r)$, $M_u(r)$ is a martingale. By Young's inequality and the linear growth properties of f and b , we have that for every $\eta > 0$,

$$\begin{aligned} & 2(\bar{X}_u(r) - \bar{X}_u(t_0)) \cdot \left(f(\bar{X}_u(r)) + \int_0^1 \int_{\mathbb{R}^d} b(\bar{X}_u(r), x) G(u, v) \bar{\mu}_{v,r}(dx) dv \right) \\ & \leq \eta |\bar{X}_u(r) - \bar{X}_u(t_0)|^2 + \frac{C_4(1 + |\bar{X}_u(r)|^2)}{\eta}. \end{aligned}$$

So, by letting

$$\begin{aligned} A(r) &:= \theta(r) \left[d + \frac{C_4(1 + |\bar{X}_u(r)|^2)}{\eta} \right], \\ B(r) &:= \eta\theta(r) + 2\theta^2(r) + \theta'(r), \end{aligned}$$

we obtain

$$Z_u(t) \leq 1 + M_u(t) + \int_{t_0}^t Z_u(r) [A(r) + B(r)|\bar{X}_u(r) - \bar{X}_u(t_0)|^2] dr.$$

The remaining argument is similar to that in [9, Section 5.1] and so we only give a sketch. Choose $\theta(r)$ to be the solution of the ODE

$$\eta\theta(r) + 2\theta^2(r) + \theta'(r) = 0$$

with $\theta(0) \in (0, \zeta/2)$. It is easy to see that the solution is decreasing and strictly positive. Thus $B(r) = 0$ and $0 < \theta(T) \leq \theta(r) < \zeta/2$ for every $r \in [0, T]$. As a consequence,

$$\mathbb{E} \sup_{t_0 \leq t \leq t_0 + \Delta} Z_u(t) \leq 1 + \mathbb{E} \sup_{t_0 \leq t \leq t_0 + \Delta} M_u(t) + \int_{t_0}^{t_0 + \Delta} \mathbb{E} Z_u(r) A(r) dr.$$

Using the bound in (A.6) one can check exactly as in [9, Section 5.1] that

$$\sup_{0 \leq r \leq T} \mathbb{E} Z_u(r) A(r) < \infty \text{ and } \mathbb{E} \sup_{t_0 \leq t \leq t_0 + \Delta} M_u(t) \leq C_5 \Delta.$$

This proves (A.5), with $C(T) := C_5 + \sup_{0 \leq r \leq T} \mathbb{E} Z_u(r) A(r) < \infty$, and hence the result follows. \blacksquare

Proof of Lemma 3.5. Noting that $W_1(\bar{\nu}_s^n, \bar{\nu}_t^n) \leq \frac{1}{n} \sum_{i=1}^n |\bar{X}_{\frac{i}{n}}(t) - \bar{X}_{\frac{i}{n}}(s)|$, we have

$$\mathbb{P} \left(\sup_{t_0 \leq s, t \leq t_0 + \Delta} W_1(\bar{\nu}_s^n, \bar{\nu}_t^n) > \varepsilon \right) \leq \mathbb{P} \left(\frac{1}{n} \sum_{i=1}^n V_i^n > \varepsilon \right),$$

where $V_i^n := \sup_{t_0 \leq s, t \leq t_0 + \Delta} |\bar{X}_{\frac{i}{n}}(t) - \bar{X}_{\frac{i}{n}}(s)|$.

By Markov's inequality and the independence of \bar{X}_u , we have

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

But for any given $\theta \geq 0$, with $C(T)$ and θ_T as in Lemma 3.4, we have

$$\mathbb{E} \exp(\theta V_i^n) \leq \mathbb{E} \exp \left(\frac{\theta^2}{4\theta_T} + \theta_T (V_i^n)^2 \right) \leq (1 + C(T)\Delta) \exp \left(\frac{\theta^2}{4\theta_T} \right).$$

Therefore

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

This completes the proof. \blacksquare

APPENDIX B. SUB-GAUSSIAN INEQUALITY

Proof of Lemma 3.8. Since $\mathbb{E}Y = 0$, we have

$$\mathbb{E}[e^{\lambda \cdot Y}] = 1 + \mathbb{E} \left[\sum_{k=2}^{\infty} \frac{(\lambda \cdot Y)^k}{k!} \right] \leq 1 + \frac{|\lambda|^2}{2} \mathbb{E} \left[\sum_{k=2}^{\infty} \frac{|\lambda|^{k-2} |Y|^k}{(k-2)!} \right] = 1 + \frac{|\lambda|^2}{2} \mathbb{E} [|Y|^2 e^{|\lambda| |Y|}].$$

Since $x^2 < 2a(1 + \frac{x^2}{2a}) \leq 2ae^{x^2/2a}$ and $tx \leq \frac{at^2}{2} + \frac{x^2}{2a}$ for $x, t \in \mathbb{R}$, we have

$$\begin{aligned} \mathbb{E}[e^{\lambda \cdot Y}] &\leq 1 + \frac{|\lambda|^2}{2} \mathbb{E} \left[2a \exp \left(\frac{|Y|^2}{2a} \right) \exp \left(\frac{a|\lambda|^2}{2} + \frac{|Y|^2}{2a} \right) \right] \\ &= 1 + a|\lambda|^2 \exp \left(\frac{a|\lambda|^2}{2} \right) \mathbb{E} \left[\exp \left(\frac{|Y|^2}{a} \right) \right] \\ &\leq (1 + 2a|\lambda|^2) \exp \left(\frac{a|\lambda|^2}{2} \right) \\ &\leq \exp \left(\frac{5a}{2} |\lambda|^2 \right), \end{aligned}$$

where the last line follows on noting that $1 + 2ax \leq e^{2ax}$ for $x \in \mathbb{R}$. ■

APPENDIX C. BOUNDS OF INDEPENDENT RANDOM VARIABLES

In this section we prove the following heterogeneous version of the result with i.i.d. setup in [9, Theorem 2.1]. We will keep the use of X , μ , etc. as in [9]. These notations are used in this section only.

Proposition C.1. *Let $\{X^i : i \in \mathbb{N}\}$ be independent \mathbb{R}^d -valued random variables. Write*

$$\hat{\mu}^n := \frac{1}{n} \sum_{i=1}^n \delta_{X^i}, \quad \mu_i := \mathcal{L}(X^i), \quad \mu^n := \frac{1}{n} \sum_{i=1}^n \mu_i.$$

Suppose $E_\alpha := \sup_{i \in \mathbb{N}} \int_{\mathbb{R}^d} e^{\alpha|x|^2} \mu_i(dx) < \infty$ for some $\alpha > 0$. Let $p \in [1, 2]$. Suppose there is some $\lambda > 0$ such that μ^n satisfies the Talagrand $T_p(\lambda)$ inequality for each $n \in \mathbb{N}$, namely

$$W_p(\nu, \mu^n) \leq \sqrt{\frac{2}{\lambda} H(\nu \| \mu^n)} \quad (\text{C.1})$$

for any $\nu \in \mathcal{P}(\mathbb{R}^d)$, where $H(\cdot \| \cdot)$ is the relative entropy. Then, for any $d' > d$ and $\lambda' < \lambda$, there exists some $N_0 \in \mathbb{N}$, depending only on $\lambda', d', \alpha, E_\alpha$, such that for any $\varepsilon > 0$ and $n \geq N_0 \max(\varepsilon^{-(d'+2)}, 1)$,

$$\mathbb{P}(W_p(\hat{\mu}^n, \mu^n) > \varepsilon) \leq \exp \left(-\gamma_p \frac{\lambda'}{2} n \varepsilon^2 \right),$$

where

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

Remark C.1. *The assumption that μ^n satisfies the $T_p(\lambda)$ inequality, for some $p \in [1, 2]$ and $\lambda > 0$, implies the existence of a square-exponential moment of μ^n . However, this is different from the assumption $E_\alpha < \infty$ which is a uniform bound on the square-exponential moment of each μ_i . The latter is not redundant and will be used in the proof of Proposition C.1.*

Proof of Proposition C.1. The proof follows the three-step argument in [9, Section 3.1] and hence we only report what is different. In particular, details are provided for Step 1 and a part of Step 2 where the heterogeneity requires some additional estimates. Only a sketch is provided for the remaining part of Step 2 and Step 3.

Step 1 (Truncation) Let $R > 0$ and $B_R := \{x \in \mathbb{R}^d : |x| \leq R\}$. Truncate μ_i into a probability measure $\mu_{i,R}$ on the ball B_R , namely $\mu_{i,R}(\cdot) := \mathbf{1}_{B_R} \mu_i(\cdot) / \mu_i(B_R)$. Let $\mu_R^n := \frac{1}{n} \sum_{i=1}^n \mu_{i,R}$.

Let $\{Y^i : i \in \mathbb{N}\}$ be independent \mathbb{R}^d -valued random variables such that $\mathcal{L}(Y^i) = \mu_{i,R}$, and $\{X^i : i \in \mathbb{N}\}$ and $\{Y^i : i \in \mathbb{N}\}$ are independent. Define

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

Then

$$|X^i - X_R^i| = |X^i - Y^i| \mathbf{1}_{|X^i| > R} \leq (|X^i| + R) \mathbf{1}_{|X^i| > R} \leq 2|X^i| \mathbf{1}_{|X^i| > R} \quad (\text{C.2})$$

and hence

$$W_p^p(\mu^n, \mu_R^n) \leq \frac{1}{n} \sum_{i=1}^n \mathbb{E} |X^i - X_R^i|^p \leq \frac{1}{n} \sum_{i=1}^n 2^p \mathbb{E} (|X^i|^p \mathbf{1}_{|X^i| > R}) = \frac{1}{n} \sum_{i=1}^n 2^p \int_{\{|x| > R\}} |x|^p \mu_i(dx).$$

If $R \geq \sqrt{p/(2\alpha)}$, then the function $r \mapsto r^p/e^{\alpha r^2}$ is nonincreasing for $r \geq R$, and then

$$W_p^p(\mu^n, \mu_R^n) \leq \frac{1}{n} \sum_{i=1}^n 2^p \frac{R^p}{e^{\alpha R^2}} \int_{\{|x| > R\}} e^{\alpha |x|^2} \mu_i(dx).$$

We conclude that

$$W_p^p(\mu^n, \mu_R^n) \leq 2^p E_\alpha R^p e^{-\alpha R^2} \quad \text{for all } R \geq \sqrt{p/(2\alpha)}. \quad (\text{C.3})$$

On the other hand, the empirical measures

$$\hat{\mu}^n = \frac{1}{n} \sum_{i=1}^n \delta_{X^i}, \quad \hat{\mu}_R^n := \frac{1}{n} \sum_{i=1}^n \delta_{X_R^i}$$

satisfy

$$W_p^p(\hat{\mu}_R^n, \hat{\mu}^n) \leq \frac{1}{n} \sum_{i=1}^n |X_R^i - X^i|^p \leq \frac{1}{n} \sum_{i=1}^n Z^i$$

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

$$\mathbb{P}[W_p(\hat{\mu}_R^n, \hat{\mu}^n) > \varepsilon] \leq \mathbb{P}\left[\frac{1}{n} \sum_{i=1}^n Z^i > \varepsilon^p\right] \leq \exp\left(-n\theta\varepsilon^p + \sum_{i=1}^n \log \mathbb{E} \exp(\theta Z^i)\right). \quad (\text{C.4})$$

In the case when $p \in [1, 2]$, for any $\alpha_1 \in (0, \alpha)$, there exists some constant $R_1 = R_1(\alpha_1, p)$ such that $2^p \theta r^p \leq \alpha_1 r^2$ for all $\theta > 0$ and $r \geq R_1 \theta^{\frac{1}{2-p}}$. Hence

$$\mathbb{E} \exp(\theta Z^i) \leq \mathbb{E} \exp(\alpha_1 |X^i|^2 \mathbf{1}_{|X^i| > R}) \leq 1 + \mathbb{E} \left(e^{\alpha_1 |X^i|^2} \mathbf{1}_{|X^i| > R} \right) \leq 1 + e^{(\alpha_1 - \alpha) R^2} E_\alpha$$

for $R \geq R_1 \theta^{\frac{1}{2-p}}$. From this and (C.4) we have

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

From this, (C.3) and the triangle inequality for W_p , we have

$$\begin{aligned} & \mathbb{P}[W_p(\mu^n, \hat{\mu}^n) > \varepsilon] \\ & \leq \mathbb{P}[W_p(\mu^n, \mu_R^n) + W_p(\mu_R^n, \hat{\mu}_R^n) + W_p(\hat{\mu}_R^n, \hat{\mu}^n) > \varepsilon] \\ & \leq \mathbb{P}\left[W_p(\mu_R^n, \hat{\mu}_R^n) > \eta\varepsilon - 2E_\alpha^{1/p} R e^{-\frac{\alpha}{p} R^2}\right] + \mathbb{P}[W_p(\hat{\mu}_R^n, \hat{\mu}^n) > (1 - \eta)\varepsilon] \\ & \leq \mathbb{P}\left[W_p(\mu_R^n, \hat{\mu}_R^n) > \eta\varepsilon - 2E_\alpha^{1/p} R e^{-\frac{\alpha}{p} R^2}\right] + \exp\left(-n \left[\theta (1 - \eta)^p \varepsilon^p - E_\alpha e^{(\alpha_1 - \alpha) R^2} \right]\right). \end{aligned} \quad (\text{C.5})$$

This estimate was established for any given $p \in [1, 2)$, $\eta \in (0, 1)$, $\varepsilon, \theta > 0$, $\alpha_1 \in (0, \alpha)$ and $R \geq \sqrt{p/(2\alpha)} \vee R_1 \theta^{\frac{1}{2-p}}$, where R_1 is a constant depending only on α_1 and p .

In the case when $p = 2$, we let $Z^i := |X^i - Y^i|^2 \mathbf{1}_{|X^i| > R}$, $i = 1, \dots, n$. Starting from (C.4) again, we choose $\alpha_1 \in (0, \alpha)$ and then $\theta := \alpha_1/2$. By definition of Z^i and $\mu_{i,R}$,

$$\begin{aligned} \mathbb{E} \left[\exp \left(\frac{\alpha_1}{2} Z^i \right) \right] &= \int_{\mathbb{R}^{2d}} \exp \left(\frac{\alpha_1}{2} |y - x|^2 \mathbf{1}_{|x| \geq R} \right) \mu_i(dx) \mu_{i,R}(dy) \\ &= \mu_i(B_R) + \frac{1}{\mu_i(B_R)} \int_{|y| \leq R} \int_{|x| \geq R} \exp \left(\frac{\alpha_1}{2} |y - x|^2 \right) \mu_i(dx) \mu_i(dy) \\ &\leq 1 + \left(1 - E_\alpha e^{-\alpha R^2} \right)^{-1} \int_{|y| \leq R} e^{\alpha_1 |y|^2} \mu_i(dy) \int_{|x| \geq R} e^{\alpha_1 |x|^2} \mu_i(dx) \\ &\leq 1 + 2E_\alpha^2 e^{(\alpha_1 - \alpha)R^2} \end{aligned}$$

for $R \geq R_2$, where $R_2 = R_2(\alpha, E_\alpha)$ is some constant. From this and (C.4) we have

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

So a similar argument as in (C.5) gives

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

This estimate was established for any given $\eta \in (0, 1)$, $\varepsilon > 0$, $\alpha_1 \in (0, \alpha)$ and $R \geq \sqrt{p/(2\alpha)} \vee R_2$, where R_2 is a constant depending only on α and E_α .

So, apart from some error terms, it will be sufficient to establish the result for the “truncated” law μ_R^n , whose support lies in the compact set B_R .

Next we prove that μ_R^n satisfies some modified T_p inequality. Let ν be a probability measure on B_R , absolutely continuous with respect to μ (and hence with respect to μ_R^n). Then, when $R \geq R_3$ for some constant $R_3 = R_3(E_\alpha)$, we can write

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

Since μ^n satisfies the $T_p(\lambda)$ inequality,

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

by triangle inequality. Combining these two displays, we obtain

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

From this, (C.3) and elementary inequality

$$\forall a \in (0, 1), \exists C_a > 0 \text{ such that } \forall x, y \in \mathbb{R}, (x - y)^2 \geq (1 - a)x^2 - C_a y^2,$$

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

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

Step 2 (Covering by small balls) In this second step we derive quantitative estimates on $\hat{\mu}_R^n$. Let ϕ be a bounded continuous function on B_R , and let U be a Borel set in $\mathcal{P}(B_R)$. By Chebyshev's exponential inequality, the independence of the variables X_R^i , and the concavity of $\log(\cdot)$,

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

As ϕ is arbitrary, we can show that [9, Equation (30)] holds if U is convex and compact, namely

$$\mathbb{P}(\hat{\mu}_R^n \in U) \leq \exp \left(-n \inf_{\nu \in U} H(\nu \|\mu_R^n) \right).$$

Following the argument in [9, Pages 557–559], we can show that equations (36) and (37) therein hold. More precisely, given $p \in [1, 2)$, $\lambda_2 < \lambda$ and $\alpha_1 < \alpha$, there exist some constants $K_1, K_2, K_3, R_4 \in (0, \infty)$ such that for all $\varepsilon, \zeta > 0$ and $R \geq R_4 \max(1, \zeta^{\frac{1}{2-p}})$,

$$\begin{aligned} \mathbb{P}(W_p(\hat{\mu}^n, \mu^n) > \varepsilon) &\leq \left(K_2 \frac{R}{\varepsilon} \vee 1 \right)^{K_2 \left(\frac{R}{\varepsilon} \right)^d} \exp \left(-n \left[\frac{\lambda_2}{2} \varepsilon^2 - K_1 R^2 e^{-\alpha R^2} \right] \right) \\ &\quad + \exp \left(-n \left[K_3 \zeta \varepsilon^p - K_4 e^{(\alpha_1 - \alpha) R^2} \right] \right) \end{aligned} \tag{C.6}$$

for some constant $K_4 = K_4(\theta, \alpha_1)$; while in the case $p = 2$,

$$\begin{aligned} \mathbb{P}(W_2(\hat{\mu}^n, \mu^n) > \varepsilon) &\leq \left(K_2 \frac{R}{\varepsilon} \vee 1 \right)^{K_2 \left(\frac{R}{\varepsilon} \right)^d} \exp \left(-n \left[\frac{\lambda_2}{2} \eta^2 \varepsilon^2 - K_1 R^2 e^{-\alpha R^2} \right] \right) \\ &\quad + \exp \left(-n \left[\frac{\alpha_1}{2} (1 - \eta)^2 \varepsilon^2 - K_4 e^{(\alpha_1 - \alpha) R^2} \right] \right) \end{aligned} \tag{C.7}$$

for any $\eta \in (0, 1)$ and $R \geq R_4$.

Step 3 (Choice of the parameters) The argument is the same as that in [9, Pages 559–561]. Therefore we only provide a sketch here. First consider the case $p \in [1, 2)$. Let $\lambda' < \lambda_2$, $\alpha' < \alpha$ and $d_1 > d$. Using (C.6) one can show that [9, (38)] holds, that is,

$$\mathbb{P}(W_p(\hat{\mu}^n, \mu^n) > \varepsilon) \leq \exp \left(-\frac{\lambda'}{2} n \varepsilon^2 \right) + \exp(-\alpha' n \varepsilon^2) \tag{C.8}$$

as soon as

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

for some constants R_2 and K_5 depending on μ^n only through λ , α and E_α . By choosing parameters R, n, d' carefully as in [9, (39)], one can find some $N_0 = N_0(\lambda', d', \alpha, E_\alpha)$ such that for all $\varepsilon > 0$, (C.8) holds as soon as $n \geq N_0 \max(\varepsilon^{-(d'+2)}, 1)$. Choosing $\alpha < \lambda/2$ gives the desired upper bound for the right hand side of (C.8).

Finally consider the case $p = 2$. Given $\lambda_3 < \lambda_2$ and $\alpha_2 < \alpha_1$, using (C.7) one can show that similar estimates as in (C.8) holds, that is,

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

Taking $\alpha_2 = \lambda_3/2$ and $\eta = \sqrt{2} - 1$ gives the desired result. ■

REFERENCES

- [1] J. Barré, P. Dobson, M. Ottobre, and E. Zatorska, *Fast non-mean-field networks: Uniform in time averaging*, SIAM Journal on Mathematical Analysis **53** (2021), no. 1, 937–972.
- [2] E. Bayraktar, S. Chakraborty, and R. Wu, *Graphon mean field systems*, The Annals of Applied Probability, accepted (2022).
- [3] E. Bayraktar and R. Wu, *Mean field interaction on random graphs with dynamically changing multi-color edges*, Stochastic Processes and their Applications **141** (2021), 197–244.
- [4] E. Bayraktar and R. Wu, *Stationarity and uniform in time convergence for the graphon particle system*, Stochastic Processes and their Applications **150** (2022), 532–568.
- [5] E. Bayraktar, R. Wu, and X. Zhang, *Propagation of chaos of forward-backward stochastic differential equations with graphon interactions*, arXiv preprint arXiv:2202.08163 (2022).
- [6] G. Bet, F. Coppini, and F. R. Nardi, *Weakly interacting oscillators on dense random graphs*, arXiv preprint arXiv:2006.07670 (2020).
- [7] S. Bhamidi, A. Budhiraja, and R. Wu, *Weakly interacting particle systems on inhomogeneous random graphs*, Stochastic Processes and their Applications **129** (2019), no. 6, 2174–2206.
- [8] F. Bolley, A. Guillin, and F. Malrieu, *Trend to equilibrium and particle approximation for a weakly self-consistent vlasov-fokker-planck equation*, ESAIM: Mathematical Modelling and Numerical Analysis **44** (2010), no. 5, 867–884.
- [9] F. Bolley, A. Guillin, and C. Villani, *Quantitative concentration inequalities for empirical measures on non-compact spaces*, Probability Theory and Related Fields **137** (2007), no. 3-4, 541–593.
- [10] F. Bolley and C. Villani, *Weighted csiszár-kullback-pinsker inequalities and applications to transportation inequalities*, Annales de la faculté des sciences de toulouse: Mathématiques, 2005, pp. 331–352.
- [11] A. Budhiraja and W.-T. L. Fan, *Uniform in time interacting particle approximations for nonlinear equations of Patlak-Keller-Segel type*, Electron. J. Probab. **22** (2017), 37 pp.
- [12] A. Budhiraja, D. Mukherjee, and R. Wu, *Supermarket model on graphs*, Ann. Appl. Probab. **29** (201906), no. 3, 1740–1777.
- [13] P. E. Caines and M. Huang, *Graphon mean field games and the GMFG equations*, 2018 IEEE Conference on Decision and Control (CDC), 2018, pp. 4129–4134.
- [14] P. E. Caines and M. Huang, *Graphon mean field games and their equations*, SIAM Journal on Control and Optimization **59** (2021), no. 6, 4373–4399.
- [15] R. Carmona, D. B. Cooney, C. V. Graves, and M. Lauriere, *Stochastic graphon games: I. the static case*, Mathematics of Operations Research **47** (2022), no. 1, 750–778.
- [16] F. Coppini, *A Note on Fokker–Planck Equations and Graphons*, Journal of Statistical Physics **187** (2022), no. 2, 1–12.
- [17] F. Coppini, *Long time dynamics for interacting oscillators on graphs*, The Annals of Applied Probability **32** (2022), no. 1, 360–391.
- [18] F. Coppini, H. Dietert, and G. Giacomini, *A law of large numbers and large deviations for interacting diffusions on Erdős–Rényi graphs*, Stochastics and Dynamics **0** (2019), no. 0, 2050010, available at <https://doi.org/10.1142/S0219493720500100>.

- [19] F. Delarue, *Mean field games: A toy model on an Erdős-Renyi graph.*, ESAIM: Proceedings and Surveys **60** (2017), 1–26.
- [20] F. Delarue, D. Lacker, and K. Ramanan, *From the master equation to mean field game limit theory: Large deviations and concentration of measure*, The Annals of Probability **48** (2020), no. 1, 211–263.
- [21] S. Delattre, G. Giacomini, and E. Luçon, *A note on dynamical models on random graphs and Fokker–Planck equations*, Journal of Statistical Physics **165** (2016), no. 4, 785–798.
- [22] P. Dupuis and G. S. Medvedev, *The large deviation principle for interacting dynamical systems on random graphs*, Communications in Mathematical Physics **390** (2022), no. 2, 545–575.
- [23] S. Gao, R. F. Tchuendom, and P. E. Caines, *Linear quadratic graphon field games*, arXiv preprint arXiv:2006.03964 (2020).
- [24] I. Karatzas and S. E. Shreve, *Brownian Motion and Stochastic Calculus*, Graduate Texts in Mathematics, vol. 113, Springer New York, 1991.
- [25] V. N. Kolokoltsov, *Nonlinear Markov Processes and Kinetic Equations*, Cambridge Tracts in Mathematics, vol. 182, Cambridge University Press, 2010.
- [26] D. Lacker and A. Soret, *A case study on stochastic games on large graphs in mean field and sparse regimes*, Mathematics of Operations Research **47** (2022), no. 2, 1530–1565.
- [27] L. Lovász, *Large networks and graph limits*, Vol. 60, American Mathematical Soc., 2012.
- [28] E. Luçon, *Quenched asymptotics for interacting diffusions on inhomogeneous random graphs*, Stochastic Processes and their Applications (2020).
- [29] H. P. McKean, *Propagation of chaos for a class of non-linear parabolic equations*, Stochastic differential equations (Lecture Series in Differential Equations, Session 7, Catholic University, 1967), 1967, pp. 41–57.
- [30] G. S. Medvedev, *The nonlinear heat equation on dense graphs and graph limits*, SIAM Journal on Mathematical Analysis **46** (2014), no. 4, 2743–2766.
- [31] G. S. Medvedev, *The nonlinear heat equation on w-random graphs*, Archive for Rational Mechanics and Analysis **212** (2014), no. 3, 781–803.
- [32] G. S. Medvedev, *The continuum limit of the Kuramoto model on sparse random graphs*, arXiv preprint arXiv:1802.03787 (2018).
- [33] R. Oliveira and G. Reis, *Interacting diffusions on random graphs with diverging average degrees: Hydrodynamics and large deviations*, Journal of Statistical Physics **176** (2019), 1057–1087.
- [34] F. Parise and A. E. Ozdaglar, *Graphon games: A statistical framework for network games and interventions*, Available at SSRN: <https://ssrn.com/abstract=3437293> (2019).
- [35] A.-S. Sznitman, *Topics in propagation of chaos*, Ecole d’Été de Probabilités de Saint-Flour XIX—1989, 1991, pp. 165–251.
- [36] D. Vasal, R. Mishra, and S. Vishwanath, *Sequential decomposition of graphon mean field games*, 2021 American Control Conference (ACC), 2021, pp. 730–736.
- [37] A. Y. Veretennikov, *On ergodic measures for mckean-vlasov stochastic equations*, Monte carlo and quasi-monte carlo methods 2004, 2006, pp. 471–486.
- [38] C. Villani, *Optimal transport: old and new*, Vol. 338, Springer Science & Business Media, 2008.

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF MICHIGAN, 530 CHURCH STREET, ANN ARBOR, MI 48109

DEPARTMENT OF MATHEMATICS, IOWA STATE UNIVERSITY, 411 MORRILL ROAD, AMES, IA 50011
 Email address: erhan@umich.edu, ruoyu@iastate.edu