Sharp Propagation of Chaos for the Ensemble Langevin Sampler

Urbain Vaes

Sharp Propagation of Chaos for the Ensemble Langevin Sampler

U. Vaes*

9th April 2024

Abstract

The aim of this note is to revisit propagation of chaos for a Langevin-type interacting particle system recently proposed as a method to sample probability measures. The interacting particle system we consider coincides, in the setting of a log-quadratic target distribution, with the ensemble Kalman sampler [13], for which propagation of chaos was first proved by Ding and Li in [9]. Like these authors, we prove propagation of chaos using a synchronous coupling as a starting point, as in Sznitman's classical argument [25]. Instead of relying on a boostrapping argument, however, we use a technique based on stopping times in order to handle the presence of the empirical covariance in the coefficients of the dynamics. This approach originates from numerical analysis [17] and was recently employed to prove mean field limits for consensus-based optimization and related interacting particle systems [18, 15]. In the context of ensemble Langevin sampling, it enables proving pathwise propagation of chaos with optimal rate, whereas previous results were optimal only up to a positive ε.

1 Introduction

Context. In a wide variety of applications, ranging from Bayesian inference to statistical physics and computational biology, it is necessary to produce samples from high-dimensional probability distributions of the form

$$\mu = \frac{\mathrm{e}^{-\phi}}{Z}, \qquad Z = \int_{\mathbf{R}^d} \mathrm{e}^{-\phi}. \tag{1.1}$$

where $\phi\colon \mathbf{R}^d \to \mathbf{R}$ is a given function. In [13], the authors propose to simulate the following interacting particle system in order to generate approximate samples from μ:

$$\mathrm{d}X^j = -\mathcal{C}(\mu_t^J)\nabla\phi(X^j)\,\mathrm{d}t + \sqrt{2\,\mathcal{C}(\mu_t^J)}\,\mathrm{d}W_t^j, \qquad j \in [\![1, J]\!], \tag{1.2}$$

where $(W^j)_{j\in[\![1,J]\!]}$ are independent standard Brownian motions in $\mathbf{R}^d$, $\mu_t^J = \frac{1}{J}\sum_{j=1}^J \delta_{X_t^j}$ is the associated empirical measure, and $\mathcal{C}(\mu_t^J)$ denotes the covariance under μ_t^J; see (1.7) below for the precise definition. They also present a gradient-free approximation of (1.2) that is well-suited to those Bayesian inverse problems for which it is difficult or undesirable to calculate derivatives of the forward model [10]. Taking formally the limit $J \to \infty$ in (1.2), they conjecture that the mean field limit of the system is given by the following McKean stochastic differential equation

$$\begin{cases} \mathrm{d}\overline{X}_t = -\mathcal{C}(\overline{\rho}_t)\nabla\phi(\overline{X}_t)\,\mathrm{d}t + \sqrt{2\,\mathcal{C}(\overline{\rho}_t)}\,\mathrm{d}W_t, \\ \overline{\rho}_t = \mathrm{Law}(\overline{X}_t)\,. \end{cases} \tag{1.3}$$

For background material on propagation of chaos, we refer to [6, 7]. The mean field equation (1.3) is often simpler to analyze mathematically than the interacting particle system (1.2), and doing so is useful in order to better understand the behavior of the particle system for large J. In the setting where ϕ is quadratic, for example, it is possible to show that the law $\overline{\rho}_t$ converges exponentially, in an appropriate sense, to the target distribution [5, 13].

In [23], a correction of (1.2) is proposed to ensure that the interacting particle system possesses as invariant distribution the tensorized measure $\mu^{\otimes J}$ for every value of $J \geqslant d+1$. The properties of the resulting interacting particle system, with acronym ALDI, are analyzed in [14].

*MATHERIALS team, Inria Paris & CERMICS, École des Ponts, France (urbain.vaes@inria.fr)

Summary of previous work. In [9], Ding and Li study propagation of chaos for (1.2) in the particular situation where ϕ is quadratic. In the setting of Bayesian inverse problems [24], this situation arises when the forward model is linear, observational noise is Gaussian and the prior distribution is Gaussian. In order to show convergence, in an appropriate sense, of the interacting particle system (1.2) to the mean field limit (1.3), they employ as a pivot the following synchronously coupled system, with the same initial condition and the same Brownian motions:

$$\begin{cases} \mathrm{d}\overline{X}^j = -\mathcal{C}(\overline{\rho}_t)\nabla\phi(\overline{X}^j)\,\mathrm{d}t + \sqrt{2\,\mathcal{C}(\overline{\rho}_t)}\,\mathrm{d}W_t^j, \qquad j \in [\![1, J]\!], \\ \overline{\rho}_t = \mathrm{Law}\left(\overline{X}_t^j\right). \end{cases} \tag{1.4}$$

Using a novel bootstrapping argument, which is summarized in [7, p. 142], they prove that for every $\varepsilon > 0$, there exists C independent of J such that

$$\sup_{t\in[0,T]} \mathbf{E}\left[\left|X_t^j - \overline{X}_t^j\right|^2\right] \leqslant CJ^{-\frac{1}{2}+\varepsilon}. \tag{1.5}$$

Combining this bound with Wasserstein convergence estimates for empirical measures formed from i.i.d. samples [12], they then deduce a convergence estimate of the form

$$\forall J \in \mathbf{N}^+, \qquad \mathbf{E}\Big[W_2(\mu_T^J, \overline{\rho}_T)\Big] \leqslant CJ^{-\alpha}, \tag{1.6}$$

for an appropriate $\alpha > 0$. This part of the argument is straightforward and independent of the interacting particle system considered. In the terminology of [6], estimate (1.5) establishes pointwise infinite-dimensional Wasserstein-2 chaos, whereas estimate (1.6) implies pointwise Wasserstein-2 empirical chaos.

Contributions of this note. The aim of this note is to revisit propagation of chaos for the interacting particle system (1.2). For simplicity, we consider neither the modification proposed in [23, 14] nor gradient-free approximations, but note that the approach we present generalizes to the ALDI sampler from [14] in a straightforward manner. Our contributions are the following:

- We generalize the work of [9] by relaxing the assumption that ϕ is quadratic. The assumptions made on ϕ in our main result are similar to those in [14].
- Whereas (1.5), in the terminology of [7], may be viewed as a *pointwise* estimate for the Wasserstein-2 norm, the result we prove is a general *pathwise* estimate for the Wasserstein-p norm.
- We employ an approach based on appropriate stopping times, which is novel in the context of mean field limits and may prove useful for the analysis of other interacting particle systems. This leads to an estimate which is is optimal, in the sense that (1.5) holds with $\varepsilon = 0$.
- Finally, we prove a well-posedness result for the mean-field dynamics (1.4).

The rest of this document is organized as follows. Well-posedness results for the interacting particle system and its formal mean field limit are presented in Section 2. We then present additional auxiliary results in Section 3, and the main results in Section 4. The appendices contain the proofs of well-posedness results (Appendix A) and auxiliary results (Appendix B).

Notation. We let $\mathbf{N} := \{0, 1, 2, 3, \dots\}$ and $\mathbf{N}^+ := \{1, 2, 3, \dots\}$. For a matrix $X \in \mathbf{R}^{d\times d}$, the notation $\|X\|_{\mathrm{F}}$ refers to the Frobenius norm. The set of probability measures over $\mathbf{R}^d$ is denoted by $\mathcal{P}(\mathbf{R}^d)$, and the notation $\mathcal{P}_p(\mathbf{R}^d) \subset \mathcal{P}(\mathbf{R}^d)$ refers to the set of probability measures with finite moments up to order p. For a probability measure $\mu \in \mathcal{P}(\mathbf{R}^d)$ the following notation is used to denote the mean and covariance under μ:

$$\mathcal{M}(\mu) = \int_{\mathbf{R}^d} x\,\mu(\mathrm{d}x), \qquad \mathcal{C}(\mu) = \int_{\mathbf{R}^d} \big(x - \mathcal{M}(\mu)\big) \otimes \big(x - \mathcal{M}(\mu)\big)\,\mu(\mathrm{d}x). \tag{1.7}$$

For a probability measure $\mu \in \mathcal{P}(\mathbf{R}^d)$ and a function $f\colon \mathbf{R}^d \to \mathbf{R}$, we use the short-hand notation

$$\mu[f] = \int_E f(x)\,\mu(\mathrm{d}x).$$

By a slight abuse of notation, we sometimes write $\mu[f(x)]$ instead of $\mu[f]$ for convenience. For example, for a probability measure $\mu \in \mathcal{P}_2(\mathbf{R})$, the notation $\mu\left[x^2\right]$ refers to the second raw moment of μ. Throughout this note, the notation Ω refers to the sample space, and C refers to a constant whose exact value is irrelevant in the context and may change from occurrence to occurrence. Furthermore, in expressions such as $\mathbf{E}\left|f(X)\right|^p$, it is always assumed that the exponent is inside the expectation; otherwise we write $\left(\mathbf{E}\left[f(X)\right]\right)^p$. Finally, for convenience, we let $|x|_* = \max\{1, |x|\}$, so that $|x|_*^a \leqslant |x|_*^b$ for all $a \leqslant b$ and all $x \in \mathbf{R}$.

2 Assumption and well-posedness

Throughout this note, we denote by $\mathcal{A}(\ell)$ for $\ell \geqslant 0$ the set of negative log-densities ϕ that satisfy the following assumptions, with that value of ℓ.

Assumption H. *Without loss of generality, we assume that the function $\phi\colon \mathbf{R}^d \to \mathbf{R}$ is bounded from below by 1. We assume furthermore that $\phi \in \mathrm{C}^2(\mathbf{R}^d)$ and that there are positive constants c_ℓ, c_u and a compact set K such that the following inequalities are satisfied for all $x \in \mathbf{R}^d \setminus K$:*

$$c_\ell |x|^{\ell+2} \leqslant \phi(x) \leqslant c_u |x|^{\ell+2}, \tag{2.1a}$$

$$c_\ell |x|^{\ell+1} \leqslant |\nabla \phi(x)| \leqslant c_u |x|^{\ell+1}, \tag{2.1b}$$

$$c_\ell |x|^{\ell}\, \mathrm{I}_d \preccurlyeq \mathrm{D}^2 \phi(x) \preccurlyeq c_u\, \mathrm{I}_d\, |x|^{\ell}. \tag{2.1c}$$

Remark 1. *A few comments are in order.*

- *The upper bound in (2.1c) implies that $\nabla\phi$ is locally Lipschitz continuous: there is $L_\phi > 0$ such that*

$$\forall x, y \in \mathbf{R}^d, \qquad \left|\nabla\phi(x) - \nabla\phi(y)\right| \leqslant L_\phi \left(1 + |x|^\ell + |y|^\ell\right) |x - y|. \tag{2.2}$$

- *Assumption H implies that there are constants $\widetilde{c_\ell}, \widetilde{c_u}$ such that*

$$\forall x \in \mathbf{R}^d, \qquad \widetilde{c_\ell} |x|_*^{\ell+2} \leqslant \phi(x) \leqslant \widetilde{c_u} |x|_*^{\ell+2}, \qquad \left|\nabla\phi(x)\right| \leqslant \widetilde{c_u} |x|_*^{\ell+1}, \qquad \left\|\mathrm{D}^2 \phi(x)\right\|_{\mathrm{F}} \leqslant \widetilde{c_u} |x|_*^{\ell}. \tag{2.3}$$

  We shall henceforth assume that these inequalities are satisfied with the same constants as in Assumption H.

- *When the target distribution (1.1) is the Bayesian posterior associated with an inverse problem with a linear forward model, Gaussian noise and a Gaussian prior, the function ϕ is quadratic. In this case Assumption H is indeed satisfied with $\ell = 0$.*

Before presenting well-posedness results, we describe the mathematical setting more precisely than in Section 1. Given independent standard Brownian motions $\left(W^j\right)_{j \in \mathbf{N}^+}$ in $\mathbf{R}^d$ and independent random variables $\left(X_0^j\right)_{j \in \mathbf{N}^+}$ sampled from some $\overline{\rho}_0 \in \mathcal{P}(\mathbf{R}^d)$, we consider for each $J \in \mathbf{N}^+$ the following interacting particle system

$$X_t^j = X_0^j + \int_0^t b\left(X_s^j, \mu_s^J\right) \mathrm{d}s + \int_0^t \sigma\left(X_s^j, \mu_s^J\right) \mathrm{d}W_s^j, \qquad j \in [\![1, J]\!] \tag{2.4}$$

where the drift $b\colon \mathbf{R}^d \times \mathcal{P}(\mathbf{R}^d) \to \mathbf{R}^d$ and diffusion $\sigma\colon \mathbf{R}^d \times \mathcal{P}(\mathbf{R}^d) \to \mathbf{R}^{d \times d}$ are given by

$$b(x, \mu) := -\mathcal{C}(\mu) \nabla\phi(x), \qquad \sigma(x, \mu) := \sqrt{2\,\mathcal{C}(\mu)}. \tag{2.5}$$

Following the classical synchronous coupling approach [25], we couple to (2.4) the system of i.i.d. mean-field McKean-Vlasov diffusions

$$\forall j \in [\![1, J]\!], \qquad \begin{cases} \overline{X}_t^j = X_0^j + \displaystyle\int_0^t b\left(\overline{X}_s^j, \overline{\rho}_s\right) \mathrm{d}s + \int_0^t \sigma\left(\overline{X}_s^j, \overline{\rho}_s\right) \mathrm{d}W_s^j, \\ \overline{\rho}_t = \mathrm{Law}\left(\overline{X}_t^j\right), \end{cases} \tag{2.6}$$

driven by the same Brownian motions and with the same initial conditions as (2.4). The well-posedness of (2.4) and (2.5) follows from Proposition 1 and Proposition 2 respectively, stated hereafter and proved in Appendix A.

Proposition 1 (Well-posedness for the interacting particle system)**.** *Assume that $\phi \in \mathcal{A}(\ell)$ for some $\ell \geqslant 0$ and that $\overline{\rho}_0 \in \mathcal{P}_p(\mathbf{R}^d)$ for some $p \geqslant 2$. Then for any $J \in \mathbf{N}^+$, the system of stochastic differential equations (2.4) has a unique globally defined strong solutions that is almost surely continuous. Furthermore, for all $T > 0$ there is $\kappa = \kappa(p) > 0$ independent of J such that*

$$\mathbf{E}\left[\sup_{t\in[0,T]} \left|X_t^j\right|^p\right] \leqslant \kappa. \tag{2.7}$$

Proof. The proof of well-posedness is essentially that of [14, Proposition 4.4]. The moment estimate (2.7) generalizes those from [9]. See Appendix A for details. □

Proposition 2 (Well-posedness for the mean field dynamics)**.** *Suppose that $\phi \in \mathcal{A}(\ell)$, that $\overline{\rho}_0 \in \mathcal{P}_p(\mathbf{R}^d)$ for $p \geqslant \ell+2$, and that $\mathcal{C}(\overline{\rho}_0) \succ 0$. Fix $x_0 \sim \overline{\rho}_0$ and $T > 0$. Then, there exists a unique strong solution $\overline{X} \in C([0,T],\mathbf{R}^d)$ to (1.3) such that $\overline{X}_0 = x_0$ and $t \mapsto \mathcal{C}(\overline{\rho}_t)$ is continuous in $[0,T]$. Furthermore, the function $t \mapsto \mathcal{C}(\overline{\rho}_t)$ is differentiable and there is $\overline{\kappa} = \overline{\kappa}(p) > 0$ such that*

$$\mathbf{E}\left[\sup_{t\in[0,T]} |\overline{X}_t|^p\right] \leqslant \overline{\kappa}, \qquad \sup_{t\in[0,T]} \left\|\mathcal{C}(\overline{\rho}_t)^{-1}\right\|_{\mathrm{F}} \leqslant \overline{\kappa}, \qquad \sup_{t\in[0,T]} \left\|\frac{\mathrm{d}\,\mathcal{C}(\overline{\rho}_t)}{\mathrm{d}t}\right\|_{\mathrm{F}} \leqslant \overline{\kappa}, \tag{2.8}$$

Proof. The proof uses a classical fixed-point approach, similar to that used in [4]. See Appendix A for details. □

3 Auxiliary results

The proof of the main results presented in Section 4 relies on the moment bounds (2.6) and (2.7), as well as the following auxiliary lemmas.

Lemma 1 (Bound on the probability of large excursions)**.** *Let $(Z_j)_{j\in\mathbf{N}^+}$ be a family of i.i.d. $\mathbf{R}$-valued random variables such that $\mathbf{E}\left[|Z_1|^r\right] < \infty$ for some $r \geqslant 1$. Then for all $R > \mathbf{E}\left[|Z_1|\right]$, there exists a constant $C > 0$ such that*

$$\forall J \in \mathbf{N}^+, \qquad \mathbf{P}\left[\frac{1}{J}\sum_{j=1}^{J} Z_j \geqslant R\right] \leqslant CJ^{-\frac{r}{2}}.$$

Proof. This follows from a generalization of Chebychev's inequality [22, Exercise 3.21]. Let

$$X = \frac{1}{J}\sum_{j=1}^{J} Z_j.$$

By the classical Marcinkiewicz–Zygmund inequality, it holds that $\mathbf{E}\left|X - \mathbf{E}[X]\right|^r \leqslant C_{\mathrm{MZ}}(r) J^{-\frac{r}{2}} \mathbf{E}\left|Z_1 - \mathbf{E}[Z_1]\right|^r$. Therefore, using the Markov inequality, we deduce that

$$\mathbf{P}\left[X \geqslant R\right] \leqslant \mathbf{P}\left[\left|X - \mathbf{E}[X]\right|^r \geqslant \left(R - \mathbf{E}[X]\right)^r\right] \leqslant \mathbf{E}\left[\frac{\left|X - \mathbf{E}[X]\right|^r}{\left(R - \mathbf{E}[X]\right)^r}\right] \leqslant \frac{CJ^{-\frac{r}{2}}}{\left(R - \mathbf{E}[X]\right)^r},$$

which concludes the proof. □

Lemma 2 (Wasserstein stability estimates)**.** *For all $(\mu,\nu) \in \mathcal{P}_2\big(\mathbf{R}^d\big) \times \mathcal{P}_2\big(\mathbf{R}^d\big)$, it holds that*

$$\left\|\mathcal{C}(\mu) - \mathcal{C}(\nu)\right\|_{\mathrm{F}} \leqslant 2\Big(W_2(\mu,\delta_0) + W_2(\nu,\delta_0)\Big) W_2(\mu,\nu), \tag{3.1a}$$

$$\left\|\sqrt{\mathcal{C}(\mu)} - \sqrt{\mathcal{C}(\nu)}\right\|_{\mathrm{F}} \leqslant \sqrt{2}\, W_2(\mu,\nu). \tag{3.1b}$$

Proof. See Appendix B.1. □

Lemma 3 (Convergence of the empirical covariance for i.i.d. samples)**.** *For all $\mu \in \mathcal{P}_{2p}(\mathbf{R}^d)$, there is C depending only on p and the $2p$-th moment of μ such that for all $J \in \mathbf{N}^+$,*

$$\mathbf{E}\left\|\mathcal{C}(\overline{\mu}^J) - \mathcal{C}(\mu)\right\|_{\mathrm{F}}^p \leqslant CJ^{-\frac{p}{2}}, \qquad \overline{\mu}^J := \frac{1}{J}\sum_{j=1}^J \delta_{\overline{X}^j}, \qquad \left\{\overline{X}^j\right\}_{j\in\mathbf{N}} \overset{\text{i.i.d.}}{\sim} \mu. \tag{3.2}$$

Furthermore, for all $\mu \in \mathcal{P}_{2p}(\mathbf{R}^d)$ satisfying $\mathcal{C}(\mu) \succcurlyeq \eta\, \mathrm{I}_d \succ 0$, there is C depending only on (p, η) and the $2p$-th moment of μ such that for all $J \in \mathbf{N}^+$,

$$\mathbf{E}\left\|\sqrt{\mathcal{C}(\overline{\mu}^J)} - \sqrt{\mathcal{C}(\mu)}\right\|_{\mathrm{F}}^p \leqslant CJ^{-\frac{p}{2}}. \tag{3.3}$$

Proof. This is essentially [9, Lemma 3]. We include a short proof in Appendix B.2 for the reader's convenience. □

4 Main results

We first present, in Subsection 4.1, a sharp quantitative result in the setting where $\phi \in \mathcal{A}(0)$, in which case $\nabla\phi$ is globally Lipschitz continuous. Then, in Subsection 4.2, we extend the result to the setting where $\phi \in \mathcal{A}(\ell)$ for $\ell > 0$. While the methodology used in the proof of Theorem 3 is rather general and potentially applicable to other interacting particle systems, the approach used in the proof of Theorem 4 is more technical and tailored to the interacting particle system (2.4). Finally, in Subsection 4.3, we present a corollary of these theorems with applications in sampling.

4.1 Sharp propagation of chaos for globally Lispchitz $\nabla\phi$

Theorem 3. *Suppose that $\phi \in \mathcal{A}(0)$, and consider the systems (2.4) and (2.6) with the coefficients given in (2.5). Assume that $\overline{\rho}_0 \in \mathcal{P}_q(\mathbf{R}^d)$, for some $q \geqslant 6$, and that $\mathcal{C}(\overline{\rho}_0) \succ 0$. Then for all $p \in [2, \frac{q}{3}]$, there is $C > 0$ independent of J such that*

$$\forall J \in \mathbf{N}^+, \qquad \forall j \in [\![1, J]\!], \qquad \mathbf{E}\left[\sup_{t\in[0,T]}\left|X_t^j - \overline{X}_t^j\right|^p\right] \leqslant CJ^{-\min\left\{\frac{p}{2}, \frac{q-p}{2\sqrt{q}}, \frac{q-p}{2p}, \frac{q-p}{6}\right\}}. \tag{4.1}$$

Proof. Fix $p \in [2, \frac{q}{3}]$ and $r \in [p, \frac{q}{3}]$ to be determined later. Fix also $R \in (0, \infty)$ such that

$$\left(\frac{R}{2}\right)^r > \mathbf{E}\left[\sup_{t\in[0,T]}\left|\overline{X}_t^j\right|^r\right]. \tag{4.2}$$

By Proposition 2, the right-hand side is indeed finite. Consider the stopping times

$$\tau_J(R) = \inf\left\{t \geqslant 0 : \frac{1}{J}\sum_{j=1}^J \left|X_t^j\right|^r \geqslant R^r\right\} = \inf\Big\{t \geqslant 0 : W_r(\mu_t^J, \delta_0) \geqslant R\Big\}, \tag{4.3a}$$

$$\overline{\tau}_J(R) = \inf\left\{t \geqslant 0 : \frac{1}{J}\sum_{j=1}^J \left|\overline{X}_t^j\right|^r \geqslant R^r\right\} = \inf\Big\{t \geqslant 0 : W_r(\overline{\mu}_t^J, \delta_0) \geqslant R\Big\}, \tag{4.3b}$$

and let $\theta_J(R) := \min\Big\{\tau_J(R), \overline{\tau}_J(R)\Big\}$. Since R will be fixed until the end of the proof, we omit this dependence in the notation. By Hölder's inequality, it holds that

$$\begin{aligned}
&\frac{1}{2^{p-1}}\mathbf{E}\left[\sup_{t\in[0,T]}\left|X_t^j - \overline{X}_t^j\right|^p\right] \\
&\quad= \mathbf{E}\left[\sup_{t\in[0,T]}\left|X_t^j - \overline{X}_t^j\right|^p 1_{\{\theta_J > T\}}\right] + \mathbf{E}\left[\sup_{t\in[0,T]}\left|X_t^j - \overline{X}_t^j\right|^p 1_{\{\theta_J \leqslant T\}}\right] \\
&\quad\leqslant \mathbf{E}\left[\sup_{t\in[0,T]}\left|X_{t\wedge\theta_J}^j - \overline{X}_{t\wedge\theta_J}^j\right|^p\right] + \left(\mathbf{E}\left[\sup_{t\in[0,T]}\left|X_t^j - \overline{X}_t^j\right|^q\right]\right)^{\frac{p}{q}} \Big(\mathbf{P}[\theta_J \leqslant T]\Big)^{\frac{q-p}{q}},
\end{aligned} \tag{4.4}$$

We bound the two terms on the right-hand side separately, and then conclude the proof.

A. Bounding the first term in (4.4). Since $\|Y\|_{L^a(\Omega)} \leqslant \|Y\|_{L^b(\Omega)}$ for any random variable Y and any $a \leqslant b$, it holds that

$$\mathbf{E}\left[\sup_{t\in[0,T]}\left|X^j_{t\wedge\theta_J} - \overline{X}^j_{t\wedge\theta_J}\right|^p\right] \leqslant \left(\mathbf{E}\left[\sup_{t\in[0,T]}\left|X^j_{t\wedge\theta_J} - \overline{X}^j_{t\wedge\theta_J}\right|^r\right]\right)^{\frac{p}{r}}. \tag{4.5}$$

Bounding the right-hand side of this inequality is not simpler than bounding the left-hand side directly. However, the bound obtained with $r \geqslant p$ will be useful to bound the second term in (4.4). In order to bound the right-hand side of (4.5), we adapt Sznitman's classical argument. We have

$$\begin{aligned}\frac{1}{2^{r-1}}\left|X^j_{t\wedge\theta_J} - \overline{X}^j_{t\wedge\theta_J}\right|^r \leqslant &\left|\int_0^{t\wedge\theta_J} b\left(X^j_s,\mu^J_s\right) - b\left(\overline{X}^j_s,\overline{\rho}_s\right)\,\mathrm{d}s\right|^r \\ &+ \left|\int_0^{t\wedge\theta_J} \sigma\left(X^j_s,\mu^J_s\right) - \sigma\left(\overline{X}^j_s,\overline{\rho}_s\right)\,\mathrm{d}W^j_s\right|^r.\end{aligned}$$

Let us introduce the martingale

$$M^j_t = \int_0^t \sigma\left(X^j_s,\mu^J_s\right) - \sigma\left(\overline{X}^j_s,\overline{\rho}_s\right)\,\mathrm{d}W^j_s.$$

By Doob's optional stopping theorem [17, Theorem 3.3], see also [11, Equation 2.29, p.285], the process $(M^j_{t\wedge\theta_J})_{t\geqslant 0}$ is a martingale, with a quadratic variation process given by $(\langle M^j\rangle_{t\wedge\theta_J})_{t\geqslant 0}$, where $\langle M^j\rangle$ is the quadratic variation of M^j. Therefore, by the Burkholder–Davis–Gundy inequality, we have for all $t\in[0,T]$ that

$$\begin{aligned}\mathbf{E}\left[\sup_{s\in[0,t]}\left|X^j_{s\wedge\theta_J} - \overline{X}^j_{s\wedge\theta_J}\right|^r\right] \leqslant{}& (2T)^{r-1}\mathbf{E}\int_0^{t\wedge\theta_J}\left|b\left(X^j_s,\mu^J_s\right) - b\left(\overline{X}^j_s,\overline{\rho}_s\right)\right|^r\,\mathrm{d}s \\ &+ C_{\mathrm{BDG}}2^{r-1}T^{\frac{r}{2}-1}\mathbf{E}\int_0^{t\wedge\theta_J}\left\|\sigma\left(X^j_s,\mu^J_s\right) - \sigma\left(\overline{X}^j_s,\overline{\rho}_s\right)\right\|^r_{\mathrm{F}}\,\mathrm{d}s.\end{aligned} \tag{4.6}$$

From the triangle inequality, it holds that

$$\begin{aligned}\frac{1}{2^{r-1}}\mathbf{E}\int_0^{t\wedge\theta_J}\left|b\left(X^j_s,\mu^J_s\right) - b\left(\overline{X}^j_s,\overline{\rho}_s\right)\right|^r\,\mathrm{d}s \leqslant{}& \int_0^t \mathbf{E}\left|b\left(X^j_{s\wedge\theta_J},\mu^J_{s\wedge\theta_J}\right) - b\left(\overline{X}^j_{s\wedge\theta_J},\overline{\mu}^J_{s\wedge\theta_J}\right)\right|^r\,\mathrm{d}s \\ &+ \int_0^t \mathbf{E}\left|b\left(\overline{X}^j_s,\overline{\mu}^J_s\right) - b\left(\overline{X}^j_s,\overline{\rho}_s\right)\right|^r\,\mathrm{d}s.\end{aligned} \tag{4.7}$$

Similarly, for the diffusion term, we have

$$\begin{aligned}\frac{1}{2^{r-1}}\mathbf{E}\int_0^{t\wedge\theta_J}\left\|\sigma\left(X^j_s,\mu^J_s\right) - \sigma\left(\overline{X}^j_s,\overline{\rho}_s\right)\right\|^r_{\mathrm{F}}\,\mathrm{d}s \leqslant{}& \int_0^t \mathbf{E}\left\|\sigma\left(X^j_{s\wedge\theta_J},\mu^J_{s\wedge\theta_J}\right) - \sigma\left(\overline{X}^j_{s\wedge\theta_J},\overline{\mu}^J_{s\wedge\theta_J}\right)\right\|^r_{\mathrm{F}}\,\mathrm{d}s \\ &+ \int_0^t \mathbf{E}\left\|\sigma\left(\overline{X}^j_s,\overline{\mu}^J_s\right) - \sigma\left(\overline{X}^j_s,\overline{\rho}_s\right)\right\|^r_{\mathrm{F}}\,\mathrm{d}s.\end{aligned} \tag{4.8}$$

Next, we bound the terms on the right-hand side of (4.7) and (4.8).

A.1. Bounding the first term in (4.7). By (2.5) and the triangle inequality, it holds that

$$\begin{aligned}\mathbf{E}\left|b\left(X^j_{s\wedge\theta_J},\mu^J_{s\wedge\theta_J}\right) - b\left(\overline{X}^j_{s\wedge\theta_J},\overline{\mu}^J_{s\wedge\theta_J}\right)\right|^r \leqslant{}& 2^{r-1}\mathbf{E}\left|\left(\mathcal{C}\left(\mu^J_{s\wedge\theta_J}\right) - \mathcal{C}\left(\overline{\mu}^J_{s\wedge\theta_J}\right)\right)\nabla\phi\left(X^j_{s\wedge\theta_J}\right)\right|^r \\ &+ 2^{r-1}\mathbf{E}\left|\mathcal{C}\left(\overline{\mu}^J_{s\wedge\theta_J}\right)\left(\nabla\phi\left(X^j_{s\wedge\theta_J}\right) - \nabla\phi\left(\overline{X}^j_{s\wedge\theta_J}\right)\right)\right|^r.\end{aligned}$$

By (3.1a) in Lemma 2, we obtain

$$\left\|\mathcal{C}\left(\mu^J_{s\wedge\theta_J}\right) - \mathcal{C}\left(\overline{\mu}^J_{s\wedge\theta_J}\right)\right\|_{\mathrm{F}} \leqslant 2\Big(W_2\left(\mu^J_{s\wedge\theta_J},\delta_0\right) + W_2\left(\overline{\mu}^J_{s\wedge\theta_J},\delta_0\right)\Big)W_2\left(\mu^J_{s\wedge\theta_J},\overline{\mu}^J_{s\wedge\theta_J}\right) \leqslant 4R\,W_2\left(\mu^J_{s\wedge\theta_J},\overline{\mu}^J_{s\wedge\theta_J}\right),$$

where we used the definition of the stopping times in the second inequality. Therefore, we deduce that

$$\begin{aligned}\mathbf{E}\left|\left(\mathcal{C}\left(\mu^J_{s\wedge\theta_J}\right)-\mathcal{C}\left(\overline{\mu}^J_{s\wedge\theta_J}\right)\right)\nabla\phi\left(X^j_{s\wedge\theta_J}\right)\right|^r &\leqslant (4R)^r\mathbf{E}\left|W_2\left(\mu^J_{s\wedge\theta_J},\overline{\mu}^J_{s\wedge\theta_J}\right)\nabla\phi\left(X^j_{s\wedge\theta_J}\right)\right|^r\\ &= (4R)^r\mathbf{E}\left[W_2\left(\mu^J_{s\wedge\theta_J},\overline{\mu}^J_{s\wedge\theta_J}\right)^r\frac{1}{J}\sum_{j=1}^J\left|\nabla\phi\left(X^j_{s\wedge\theta_J}\right)\right|^r\right],\end{aligned}$$

where we used exchangeability in the second line. By the assumption (2.1b) of linear growth of $\nabla\phi$ and the definition of θ_J, this leads to

$$\begin{aligned}\mathbf{E}\left|\left(\mathcal{C}\left(\mu^J_{s\wedge\theta_J}\right)-\mathcal{C}\left(\overline{\mu}^J_{s\wedge\theta_J}\right)\right)\nabla\phi\left(X^j_{s\wedge\theta_J}\right)\right|^r &\leqslant C\mathbf{E}\left[W_2\left(\mu^J_{s\wedge\theta_J},\overline{\mu}^J_{s\wedge\theta_J}\right)^r\left(\frac{1}{J}\sum_{j=1}^J\left|X^j_{s\wedge\theta_J}\right|_*^r\right)\right]\\ &\leqslant C\mathbf{E}\left[W_2\left(\mu^J_{s\wedge\theta_J},\overline{\mu}^J_{s\wedge\theta_J}\right)^r\left(1+W_r(\mu^J_{s\wedge\theta_J},\delta_0)^r\right)\right]\\ &\leqslant C\mathbf{E}\left[W_2\left(\mu^J_{s\wedge\theta_J},\overline{\mu}^J_{s\wedge\theta_J}\right)^r\right].\end{aligned}\tag{4.9}$$

On the other hand, by definition of the stopping time θ_J, and the inequality $\|\mathcal{C}(\mu)\|_{\mathrm{F}}\leqslant W_2(\mu,\delta_0)^2$ which holds for all $\mu\in\mathcal{P}(\mathbf{R}^d)$, it holds that

$$\begin{aligned}\mathbf{E}\left|\mathcal{C}\left(\overline{\mu}^J_{s\wedge\theta_J}\right)\left(\nabla\phi\left(X^j_{s\wedge\theta_J}\right)-\nabla\phi\left(\overline{X}^j_{s\wedge\theta_J}\right)\right)\right|^r &\leqslant R^{2r}\mathbf{E}\left|\nabla\phi\left(X^j_{s\wedge\theta_J}\right)-\nabla\phi\left(\overline{X}^j_{s\wedge\theta_J}\right)\right|^r,\\ &\leqslant R^{2r}L^r_\phi\mathbf{E}\left|X^j_{s\wedge\theta_J}-\overline{X}^j_{s\wedge\theta_J}\right|^r.\end{aligned}\tag{4.10}$$

where we used the assumption of Lipschitz continuity on $\nabla\phi$ in Assumption H. In view of the inequalities (4.9) and (4.10), and of the bound

$$\mathbf{E}\left[W_r\left(\mu^J_{s\wedge\theta_J},\overline{\mu}^J_{s\wedge\theta_J}\right)^r\right]\leqslant\mathbf{E}\left[\frac{1}{J}\sum_{j=1}^J\left|X^j_{s\wedge\theta_J}-\overline{X}^j_{s\wedge\theta_J}\right|^r\right]=\mathbf{E}\left|X^j_{s\wedge\theta_J}-\overline{X}^j_{s\wedge\theta_J}\right|^r,\tag{4.11}$$

which holds by definition of the Wasserstein distance, we deduce that

$$\mathbf{E}\left|b\left(X^j_{s\wedge\theta_J},\mu^J_{s\wedge\theta_J}\right)-b\left(\overline{X}^j_{s\wedge\theta_J},\overline{\mu}^J_{s\wedge\theta_J}\right)\right|^r\leqslant C\mathbf{E}\left|X^j_{s\wedge\theta_J}-\overline{X}^j_{s\wedge\theta_J}\right|^r.$$

A.2. Bounding the second term in (4.7)**.** For this term, we have

$$\begin{aligned}\mathbf{E}\left|b\left(\overline{X}^j_s,\overline{\mu}^J_s\right)-b\left(\overline{X}^j_s,\overline{\rho}_s\right)\right|^r &= \mathbf{E}\left|\left(\mathcal{C}(\overline{\mu}^J_s)-\mathcal{C}(\overline{\rho}_s)\right)\nabla\phi(\overline{X}^j_s)\right|^r\\ &\leqslant C\mathbf{E}\left[\left\|\mathcal{C}(\overline{\mu}^J_s)-\mathcal{C}(\overline{\rho}_s)\right\|^r_{\mathrm{F}}\left(1+|\overline{X}^j_s|^r\right)\right]\\ &\leqslant C\left(\mathbf{E}\left\|\mathcal{C}(\overline{\mu}^J_s)-\mathcal{C}(\overline{\rho}_s)\right\|^{\frac{3r}{2}}_{\mathrm{F}}\right)^{\frac{2}{3}}\left(\mathbf{E}\left[1+|\overline{X}^j_s|^{3r}\right]\right)^{\frac{1}{3}},\end{aligned}$$

where we used (2.1b) in Assumption H and Hölder's inequality. Using the moment bound in Proposition 2 and then using Lemma 3, noting that $\overline{\rho}_0\in\mathcal{P}_{3r}(\mathbf{R}^d)$ by assumption, we deduce that

$$\mathbf{E}\left|b\left(\overline{X}^j_s,\overline{\mu}^J_s\right)-b\left(\overline{X}^j_s,\overline{\rho}_s\right)\right|^r\leqslant CJ^{-\frac{r}{2}}.$$

A.3. Bounding the first term in (4.8)**.** For the first diffusion term, we have that

$$\left\|\sigma\left(X^j_{s\wedge\theta_J},\mu^J_{s\wedge\theta_J}\right)-\sigma\left(\overline{X}^j_{s\wedge\theta_J},\overline{\mu}^J_{s\wedge\theta_J}\right)\right\|^r_{\mathrm{F}}=2^r\left\|\sqrt{\mathcal{C}\left(\mu^J_{s\wedge\theta_J}\right)}-\sqrt{\mathcal{C}\left(\overline{\mu}^J_{s\wedge\theta_J}\right)}\right\|^r_{\mathrm{F}}.$$

Using Lemma 2 together with the bound (4.11), we obtain

$$\mathbf{E}\left\|\sigma\left(X^j_{s\wedge\theta_J},\mu^J_{s\wedge\theta_J}\right)-\sigma\left(\overline{X}^j_{s\wedge\theta_J},\overline{\mu}^J_{s\wedge\theta_J}\right)\right\|^r_{\mathrm{F}}\leqslant 2^{\frac{r+1}{2}}\mathbf{E}\left|X^j_{s\wedge\theta_J}-\overline{X}^j_{s\wedge\theta_J}\right|^r.$$

***A.4. Bounding the second term in* (4.8).** By Proposition 2, there is $\eta > 0$ such that $\mathcal{C}(\overline{\rho}_t) \succcurlyeq \eta \, \mathrm{I}_d$ for all $t \in [0,T]$. Thus, it follows from Lemma 3 that

$$\mathbf{E}\left\|\sigma\left(\overline{X}^j_s, \overline{\mu}^J_s\right) - \sigma\left(\overline{X}^j_s, \overline{\rho}_s\right)\right\|^r_{\mathrm{F}} = 2^{\frac{r}{2}} \mathbf{E}\left\|\sqrt{\mathcal{C}\left(\overline{\mu}^J_s\right)} - \sqrt{\mathcal{C}\left(\overline{\rho}_s\right)}\right\|^r_{\mathrm{F}} \leqslant CJ^{-\frac{r}{2}}.$$

A.5. Concluding part A. Combining the bounds on all terms, we finally obtain from (4.6) that

$$\mathbf{E}\left[\sup_{s\in[0,t]}\left|X^j_{s\wedge\theta_J} - \overline{X}^j_{s\wedge\theta_J}\right|^r\right] \leqslant CJ^{-\frac{r}{2}} + C\int_0^t \mathbf{E}\left[\sup_{u\in[0,s]}\left|X^j_{u\wedge\theta_J} - \overline{X}^j_{u\wedge\theta_J}\right|^r\right] \mathrm{d}s.$$

By Grönwall's inequality, this implies that

$$\mathbf{E}\left[\sup_{t\in[0,T]}\left|X^j_{t\wedge\theta_J} - \overline{X}^j_{t\wedge\theta_J}\right|^r\right] \leqslant CJ^{-\frac{r}{2}}. \tag{4.12}$$

B. Bounding the second term in (4.4). We have by Propositions 1 and 2 that

$$\mathbf{E}\left[\sup_{t\in[0,T]}\left|X^j_t - \overline{X}^j_t\right|^q\right] \leqslant 2^{q-1}\left(\mathbf{E}\left[\sup_{t\in[0,T]}\left|X^j_t\right|^q\right] + \mathbf{E}\left[\sup_{t\in[0,T]}\left|\overline{X}^j_t\right|^q\right]\right) \leqslant 2^{q-1}\big(\kappa(q) + \overline{\kappa}(q)\big).$$

In order to complete the proof of the theorem, it remains to bound the probability $\mathbf{P}\left[\theta_J \leqslant T\right]$, which can be achieved by noticing that

$$\mathbf{P}\big[\theta_J \leqslant T\big] = \mathbf{P}\big[\tau_J \leqslant T < \overline{\tau}_J\big] + \mathbf{P}\big[\overline{\tau}_J \leqslant T\big].$$

Using the almost sure continuity of the solution to the interacting particle system, together with the triangle inequality, we bound the first probability as follows:

$$\begin{aligned}
\mathbf{P}[\tau_J \leqslant T < \overline{\tau}_J] &\leqslant \mathbf{P}\left[\sup_{t\in[0,T]} W_r(\mu^J_{t\wedge\theta_J}, \delta_0) = R\right] \\
&\leqslant \mathbf{P}\left[\sup_{t\in[0,T]} W_r(\mu^J_{t\wedge\theta_J}, \overline{\mu}^J_{t\wedge\theta_J}) + \sup_{t\in[0,T]} W_r(\overline{\mu}^J_{t\wedge\theta_J}, \delta_0) \geqslant R\right] \\
&\leqslant \mathbf{P}\left[\sup_{t\in[0,T]} W_r(\mu^J_{t\wedge\theta_J}, \overline{\mu}^J_{t\wedge\theta_J}) \geqslant \frac{R}{2}\right] + \mathbf{P}\left[\sup_{t\in[0,T]} W_r(\overline{\mu}^J_{t\wedge\theta_J}, \delta_0) \geqslant \frac{R}{2}\right],
\end{aligned}$$

where we used that $\mathbf{P}[A + B \geqslant k] \leqslant \mathbf{P}[A \geqslant k/2] + \mathbf{P}[B \geqslant k/2]$ for any two real-valued random variables A and B, because $\{A + B \geqslant k\} \subset \{A \geqslant k/2\} \cup \{B \geqslant k/2\}$. The probability $\mathbf{P}[\theta_J \leqslant T]$ can then be bounded as follows:

$$\begin{aligned}
\mathbf{P}[\theta_J \leqslant T] &= \mathbf{P}\left[\tau_J \leqslant T < \overline{\tau}_J\right] + \mathbf{P}\left[\overline{\tau}_J \leqslant T\right] \\
&\leqslant \mathbf{P}\left[\sup_{t\in[0,T]} W_r(\mu^J_{t\wedge\theta_J}, \overline{\mu}^J_{t\wedge\theta_J}) \geqslant \frac{R}{2}\right] + 2\mathbf{P}\left[\sup_{t\in[0,T]} W_r(\overline{\mu}^J_t, \delta_0) \geqslant \frac{R}{2}\right].
\end{aligned}$$

Using (4.11) and Markov's inequality for the first term, together with the inequality $\sup \sum \leqslant \sum \sup$, we then obtain

$$\mathbf{P}[\theta_J \leqslant T] \leqslant \frac{2^r}{R^r}\mathbf{E}\left[\frac{1}{J}\sum_{j=1}^J \sup_{t\in[0,T]}\left|X^j_{t\wedge\theta_J} - \overline{X}^j_{t\wedge\theta_J}\right|^r\right] + 2\mathbf{P}\left[\frac{1}{J}\sum_{j=1}^J \sup_{t\in[0,T]}\left|\overline{X}^j_t\right|^r \geqslant \frac{R^r}{2^r}\right]. \tag{4.13}$$

The first term is bounded from above by $CJ^{-\frac{r}{2}}$ by exchangeability and (4.12). In order to bound the second term, let us introduce the i.i.d. random variables

$$Z_j = \sup_{t\in[0,T]}\left|\overline{X}^j_t\right|^r, \qquad j \in [\![1, J]\!].$$

By the moment bounds in Proposition 2 and the assumption that $\overline{\rho}_0 \in \mathcal{P}_q(\mathbf{R}^d)$, the random variable Z_1 has finite moments up to order $\frac{q}{r} \geqslant 1$. Furthermore, by definition (4.2) of R, it holds that $\mathbf{E}[Z_1] < \frac{R^r}{2^r}$. Thus, it follows

from Lemma 1 that there is a constant C independent of J such that

$$\mathbf{P}\left[\frac{1}{J}\sum_{j=1}^{J}\sup_{t\in[0,T]}\left|\overline{X}_t^j\right|^r \geqslant \frac{R^r}{2^r}\right] = \mathbf{P}\left[\frac{1}{J}\sum_{j=1}^{J} Z_j \geqslant \frac{R^r}{2^r}\right] \leqslant CJ^{-\frac{q}{2r}}. \tag{4.14}$$

C. Concluding the proof. Substituting the bounds (4.12) and (4.14) in (4.4), we finally obtain that

$$\mathbf{E}\left[\sup_{t\in[0,T]}\left|X_t^j - \overline{X}_t^j\right|^p\right] \leqslant C\left(J^{-\frac{p}{2}} + J^{-\frac{r(q-p)}{2q}} + J^{-\frac{q-p}{2r}}\right).$$

This inequality is true for any $r \in [p, \frac{q}{3}]$, with a constant C depending on r. The best estimate is obtained when the exponents of the second and third terms are as close to equal as possible, that is to say when $r = \max\{p, \min\{\sqrt{q}, \frac{q}{3}\}\}$, which leads to the estimate (4.1) and concludes the proof. □

Remark 2. *A few comments are in order.*

- *If $\overline{\rho}_0$ has sufficiently many moments, then (4.1) recovers the optimal convergence rate $J^{-\frac{p}{2}}$, which is obtained in the classical setting with globally Lipschitz coefficients, see [7, Theorem 3.1].*
- *In the proof of Theorem 3, the probability $\mathbf{P}[\tau_J(R) \leqslant T]$ was bounded from above in terms of $\mathbf{P}[\overline{\tau}_J(\frac{R}{2}) \leqslant T]$ and an appropriate distance between the stopped particle systems. The probability $\mathbf{P}[\overline{\tau}_J(\frac{R}{2}) \leqslant T]$ is simple to bound directly, because the synchronously coupled mean field particles are independent and identically distributed.*
- *In order to obtain an estimate with a scaling that is optimal in J, it was crucial to first prove in (4.12) a propagation of chaos result for the stopped particle system in a metric L^r with r larger than the value p in the final L^p estimate (4.1).*
- *For $\phi \in \mathcal{A}(\ell)$ with $\ell > 0$, the proof presented above does not go through. The issue in this case is to obtain a bound similar to (4.10). It is still possible, however, to prove sharp propagation of chaos using a slightly different and more technical approach; see Theorem 4.*
- *Reasoning as in [9], we deduce from Theorem 3 that Wasserstein-p empirical chaos holds for the ensemble Langevin sampler (2.4), in the sense of (4.15) below. Indeed, an application of the triangle inequality gives*

$$\mathbf{E}\Big[W_p(\mu_t^J, \overline{\rho}_t)\Big] \leqslant \mathbf{E}\Big[W_p(\mu_t^J, \overline{\mu}_t^J)\Big] + \mathbf{E}\Big[W_p(\overline{\mu}_t^J, \overline{\rho}_t)\Big].$$

  Under the assumptions of Theorem 3, the first term on the right-hand side tends to 0 in the limit $J \to \infty$. On the other hand, by [12, Theorem 1], the second term decreases as $J^{-\alpha}$, for a constant $\alpha \in (0, \frac{1}{2}]$ depending on p, the dimension d and the number of bounded moments of $\overline{\rho}_t$, We refer to [12], see also [6, Lemma 4.2], for the explicit value of α, which is in general strictly smaller than the Monte Carlo rate $\frac{1}{2}$, even when $\overline{\rho}_t$ has infinitely many moments. Therefore, it holds that

$$\lim_{J\to\infty}\left(\sup_{t\in[0,T]}\mathbf{E}\Big[W_p(\mu_t^J, \overline{\rho}_t)\Big]\right) = 0. \tag{4.15}$$

4.2 Extension: sharp propagation of chaos for locally Lipschitz $\nabla\phi$

In order to prove sharp propagation of chaos for this case, we need the following additional auxiliary result.

Lemma 4 (Convexity inequality)**.** *If $\phi \in \mathcal{A}(\ell)$ for some $\ell \geqslant 0$, then there are $c_1 > 0$ and $c_2 > 0$ such that*

$$\forall (x,y) \in \mathbf{R}^d \times \mathbf{R}^d, \qquad \Big\langle y - x, \nabla\phi(y) - \nabla\phi(x)\Big\rangle \geqslant c_1\left(1 + |x|^\ell + |y|^\ell\right)|y-x|^2 - c_2|y-x|^2.$$

Proof. By the fundamental theorem, it holds that

$$\nabla\phi(y) - \nabla\phi(x) = \int_0^1 \mathrm{D}^2\,\phi\big(x + t(y-x)\big)\,(y-x)\,\mathrm{d}t. \tag{4.16}$$

Assumption (2.1c) implies that there is $c_2 > 0$ such that

$$\forall x \in \mathbf{R}^d, \qquad \mathrm{D}^2\,\phi(x) \succcurlyeq c_\ell |x|^\ell\,\mathrm{I}_d - c_2\,\mathrm{I}_d\,.$$

Therefore, taking the inner product with $y - x$ on both sides of (4.16), we obtain that

$$\Big\langle y - x, \nabla\phi(y) - \nabla\phi(x)\Big\rangle \geqslant c_\ell \int_0^1 |x + t(y-x)|^\ell\,|y-x|^2\,\mathrm{d}t - c_2 \int_0^1 |y-x|^2\,\mathrm{d}t.$$

Suppose without loss of generality that $|y| \geqslant |x|$. Then, using the triangle inequality, we obtain that

$$\begin{aligned}\int_0^1 |x + t(y-x)|^\ell\,\mathrm{d}t &\geqslant \int_{\frac{3}{4}}^1 |ty + (1-t)x|^\ell\,\mathrm{d}t \geqslant \int_{\frac{3}{4}}^1 \Big(t|y| - (1-t)|x|\Big)^\ell\,\mathrm{d}t \\ &\geqslant \int_{\frac{3}{4}}^1 \Big(2t-1\Big)|y|^\ell\,\mathrm{d}t \geqslant C|y|^\ell \geqslant \frac{C}{2}\left(|x|^\ell + |y|^\ell\right),\end{aligned}$$

which concludes the proof. □

In the following result, we assume for simplicity that the probability measure $\overline{\rho}_0$ has infinitely many moments, but this assumption is not required.

Theorem 4. *Suppose that $\phi \in \mathcal{A}(\ell)$ for some $\ell \geqslant 0$, and consider the systems (2.4) and (2.6) with the coefficients given in (2.5). Assume that $\overline{\rho}_0 \in \mathcal{P}_q(\mathbf{R}^d)$ for all $q \in \mathbf{N}$ and that $\mathcal{C}(\overline{\rho}_0) \succ 0$. Then for all $p > 0$, there is $C > 0$ independent of J such that it holds that*

$$\forall J \in \mathbf{N}^+, \qquad \forall j \in [\![1, J]\!], \qquad \mathbf{E}\left[\sup_{t\in[0,T]}\left|X_t^j - \overline{X}_t^j\right|^p\right] \leqslant C J^{-\frac{p}{2}}. \tag{4.17}$$

Proof. It suffices to prove the statement for $p \geqslant 4$ because, by Jensen's inequality, it holds for all $p \leqslant \mathfrak{p}$ that

$$\mathbf{E}\left[\sup_{t\in[0,T]}\left|X_t^j - \overline{X}_t^j\right|^p\right] \leqslant \left(\mathbf{E}\left[\sup_{t\in[0,T]}\left|X_t^j - \overline{X}_t^j\right|^{\mathfrak{p}}\right]\right)^{\frac{p}{\mathfrak{p}}}.$$

Fix $p \geqslant 4$, as well as $r = 2p$ and $q = 4p(\ell+1)$. Fix also $\varepsilon > 0$ to be determined later, and $R \in (0, \infty)$ such that

$$R^r > \mathbf{E}\left[\sup_{t\in[0,T]}\left|\overline{X}_t^j\right|^r\right], \tag{4.18}$$

and let $\theta_J := \min\Big\{\mathfrak{t}_J, \overline{\tau}_J\Big\}$, where the stopping times $\mathfrak{t}_J$ and $\overline{\tau}_J$ are given by

$$\mathfrak{t}_J = \inf\Big\{t \geqslant 0 : W_r(\mu_t^J, \overline{\mu}_t^J) \geqslant \varepsilon\Big\}, \tag{4.19a}$$

$$\overline{\tau}_J = \inf\Big\{t \geqslant 0 : W_{r(\ell+1)}(\overline{\mu}_t^J, \delta_0) \geqslant R\Big\}. \tag{4.19b}$$

By the triangle inequality and the definition of θ_J, it holds for all $t \in [0,T]$ that

$$W_r(\mu_{t\wedge\theta_J}^J, \delta_0) \leqslant W_r(\mu_{t\wedge\theta_J}^J, \overline{\mu}_{t\wedge\theta_J}^J) + W_r(\overline{\mu}_{t\wedge\theta_J}^J, \delta_0) \leqslant \varepsilon + R.$$

In addition, in view of the Wasserstein stability estimate (3.1a), it holds that

$$\Big\|\mathcal{C}(\mu_{t\wedge\theta_J}^J) - \mathcal{C}(\overline{\mu}_{t\wedge\theta_J}^J)\Big\|_{\mathrm{F}} \leqslant 2(2R+\varepsilon)\varepsilon. \tag{4.20}$$

As in the proof of Theorem 3, we have by Hölder's inequality that

$$\frac{1}{2^{p-1}}\mathbf{E}\left[\sup_{t\in[0,T]}\left|X_t^j-\overline{X}_t^j\right|^p\right] \leqslant \mathbf{E}\left[\sup_{t\in[0,T]}\left|X_{t\wedge\theta_J}^j-\overline{X}_{t\wedge\theta_J}^j\right|^p\right]+\left(\mathbf{E}\left[\sup_{t\in[0,T]}\left|X_t^j-\overline{X}_t^j\right|^q\right]\right)^{\frac{p}{q}}\left(\mathbf{P}[\theta_J\leqslant T]\right)^{\frac{q-p}{q}}. \tag{4.21}$$

A. Bounding the first term in (4.21). Here, the approach is slightly more technical than that in the proof of Theorem 3. Recall that, by Proposition 2, there is $\kappa>0$ such that

$$\forall t\in[0,T],\qquad \left\|\mathcal{C}(\overline{\rho}_t)\right\|_{\mathrm{F}}\vee\left\|\mathcal{C}(\overline{\rho}_t)^{-1}\right\|_{\mathrm{F}}\vee\left\|\frac{\mathrm{d}\,\mathcal{C}(\overline{\rho}_t)}{\mathrm{d}t}\right\|_{\mathrm{F}}\leqslant\kappa. \tag{4.22}$$

Using this bound, applying Itô's lemma to $f(x,\overline{x},t)=\frac{1}{2}(x-\overline{x})^{\mathsf{T}}\mathcal{C}(\overline{\rho}_t)^{-1}(x-\overline{x})$, and noting that $f\left(X_0^j,\overline{X}_0^j,0\right)=0$, we obtain that, for all $t\in[0,T]$,

$$\begin{aligned}\frac{1}{\kappa}\left|X_t^j-\overline{X}_t^j\right|^2\leqslant f\left(X_t^j,\overline{X}_t^j,t\right)&=-\int_0^t\left\langle X_s^j-\overline{X}_s^j,\nabla\phi(X_s^j)-\nabla\phi(\overline{X}_s^j)\right\rangle\mathrm{d}s\\&+\int_0^t\left\langle X_s^j-\overline{X}_s^j,\mathcal{C}(\overline{\rho}_s)^{-1}\Big(\mathcal{C}(\overline{\rho}_s)-\mathcal{C}(\mu_s^J)\Big)\nabla\phi(X_s^j)\right\rangle\mathrm{d}s\\&+\int_0^t\begin{pmatrix}\mathcal{C}(\mu_s^J)&\sqrt{\mathcal{C}(\mu_s^J)}\sqrt{\mathcal{C}(\overline{\rho}_s)}\\\sqrt{\mathcal{C}(\overline{\rho}_s)}\sqrt{\mathcal{C}(\mu_s^J)}&\mathcal{C}(\overline{\rho}_s)\end{pmatrix}:\begin{pmatrix}\mathcal{C}(\overline{\rho}_s)^{-1}&-\mathcal{C}(\overline{\rho}_s)^{-1}\\-\mathcal{C}(\overline{\rho}_s)^{-1}&\mathcal{C}(\overline{\rho}_s)^{-1}\end{pmatrix}\mathrm{d}s\\&+\int_0^t\frac{1}{2}\big(X_s^j-\overline{X}_s^j\big)^{\mathsf{T}}\mathcal{C}(\overline{\rho}_s)^{-1}\frac{\mathrm{d}\,\mathcal{C}(\overline{\rho}_s)}{\mathrm{d}s}\mathcal{C}(\overline{\rho}_s)^{-1}\big(X_s^j-\overline{X}_s^j\big)^{\mathsf{T}}\mathrm{d}s\\&+\int_0^t\Big(X_s^j-\overline{X}_s^j\Big)^{\mathsf{T}}\mathcal{C}(\overline{\rho}_s)^{-1}\left(\sqrt{2\,\mathcal{C}(\mu_s^J)}-\sqrt{2\,\mathcal{C}(\overline{\rho}_s)}\right)\mathrm{d}W_s^j.\end{aligned} \tag{4.23}$$

Let us bound the terms one by one.

- By Lemma 4 and Assumption H, the integrand in the first term satisfies

$$\begin{aligned}-\left\langle X_s^j-\overline{X}_s^j,\nabla\phi(X_s^j)-\nabla\phi(\overline{X}_s^j)\right\rangle&\leqslant-c_1\left(1+\left|X_s^j\right|^\ell+\left|\overline{X}_s^j\right|^\ell\right)\left|X_s^j-\overline{X}_s^j\right|^2+c_2\left|X_s^j-\overline{X}_s^j\right|^2\\&\leqslant-\frac{c_1}{L_\phi}\left|X_s^j-\overline{X}_s^j\right|\left|\nabla\phi(X_s^j)-\nabla\phi(\overline{X}_s^j)\right|+c_2\left|X_s^j-\overline{X}_s^j\right|^2.\end{aligned}$$

- By (4.22) and the triangle inequality, the integrand in the second term satisfies

$$\begin{aligned}&\left\langle X_s^j-\overline{X}_s^j,\mathcal{C}(\overline{\rho}_s)^{-1}\Big(\mathcal{C}(\overline{\rho}_s)-\mathcal{C}(\mu_s^J)\Big)\nabla\phi(X_s^j)\right\rangle\\&\quad\leqslant\kappa\left|X_s^j-\overline{X}_s^j\right|\left\|\mathcal{C}(\overline{\rho}_s)-\mathcal{C}(\overline{\mu}_s^J)\right\|_{\mathrm{F}}\left|\nabla\phi(X_s^j)\right|+\kappa\left|X_s^j-\overline{X}_s^j\right|\left\|\mathcal{C}(\overline{\mu}_s^J)-\mathcal{C}(\mu_s^J)\right\|_{\mathrm{F}}\left|\nabla\phi(X_s^j)-\nabla\phi(\overline{X}_s^j)\right|\\&\qquad+\kappa\left|X_s^j-\overline{X}_s^j\right|\left\|\mathcal{C}(\overline{\mu}_s^J)-\mathcal{C}(\mu_s^J)\right\|_{\mathrm{F}}\left|\nabla\phi(\overline{X}_s^j)\right|\\&\quad\leqslant\kappa\left|X_s^j-\overline{X}_s^j\right|^2+\frac{\kappa}{2}\left\|\mathcal{C}(\overline{\rho}_s)-\mathcal{C}(\overline{\mu}_s^J)\right\|_{\mathrm{F}}^2\left|\nabla\phi(X_s^j)\right|^2+\frac{\kappa}{2}\left\|\mathcal{C}(\overline{\mu}_s^J)-\mathcal{C}(\mu_s^J)\right\|_{\mathrm{F}}^2\left|\nabla\phi(\overline{X}_s^j)\right|^2\\&\qquad+\kappa\left|X_s^j-\overline{X}_s^j\right|\left\|\mathcal{C}(\overline{\mu}_s^J)-\mathcal{C}(\mu_s^J)\right\|_{\mathrm{F}}\left|\nabla\phi(X_s^j)-\nabla\phi(\overline{X}_s^j)\right|,\end{aligned}$$

  where we used Young's inequality in the second inequality.

- For the third term in (4.23), a simple calculation gives that

$$\begin{aligned}
&\begin{pmatrix} \mathcal{C}(\mu_s^J) & \sqrt{\mathcal{C}(\mu_s^J)}\sqrt{\mathcal{C}(\overline{\rho}_s)} \\ \sqrt{\mathcal{C}(\overline{\rho}_s)}\sqrt{\mathcal{C}(\mu_s^J)} & \mathcal{C}(\overline{\rho}_s) \end{pmatrix} : \begin{pmatrix} \mathcal{C}(\overline{\rho}_s)^{-1} & -\mathcal{C}(\overline{\rho}_s)^{-1} \\ -\mathcal{C}(\overline{\rho}_s)^{-1} & \mathcal{C}(\overline{\rho}_s)^{-1} \end{pmatrix} \\
&\qquad = \operatorname{tr}\left(\left(\sqrt{\mathcal{C}(\mu_s^J)} - \sqrt{\mathcal{C}(\overline{\rho}_s)} \right) \left(\sqrt{\mathcal{C}(\mu_s^J)} - \sqrt{\mathcal{C}(\overline{\rho}_s)} \right) \mathcal{C}(\overline{\rho}_s)^{-1} \right) \\
&\qquad = \left\| \sqrt{\mathcal{C}(\overline{\rho}_s)^{-1}} \left(\sqrt{\mathcal{C}(\mu_s^J)} - \sqrt{\mathcal{C}(\overline{\rho}_s)} \right) \right\|_{\mathrm{F}}^2 \leqslant \kappa \left\| \sqrt{\mathcal{C}(\mu_s^J)} - \sqrt{\mathcal{C}(\overline{\rho}_s)} \right\|_{\mathrm{F}}^2 \qquad (4.24a) \\
&\qquad \leqslant 2\kappa W_2(\mu_s^J, \overline{\mu}_s^J)^2 + 2\kappa \left\| \sqrt{\mathcal{C}(\overline{\mu}_s^J)} - \sqrt{\mathcal{C}(\overline{\rho}_s)} \right\|_{\mathrm{F}}^2 . \qquad (4.24b)
\end{aligned}$$

  where we used (4.22) in (4.24a), and then the triangle inequality and the Wasserstein stability estimate (3.1b) from Lemma 2 in (4.24b).

- For the fourth term in (4.23), we have in view of the bounds (4.22) on the mean field covariance that

$$\left(X_s^j - \overline{X}_s^j\right)^{\mathsf{T}} \mathcal{C}(\overline{\rho}_s)^{-1} \frac{\mathrm{d}\,\mathcal{C}(\overline{\rho}_s)}{\mathrm{d}s} \mathcal{C}(\overline{\rho}_s)^{-1} \left(X_s^j - \overline{X}_s^j\right)^{\mathsf{T}} \leqslant \kappa^3 \left| X_s^j - \overline{X}_s^j \right|^2 .$$

- Finally, for the last term in (4.23), let M_t denote the martingale

$$M_t^j = \int_0^t \left(X_s^j - \overline{X}_s^j \right)^{\mathsf{T}} \mathcal{C}(\overline{\rho}_t)^{-1} \left(\sqrt{2\,\mathcal{C}(\mu_s^J)} - \sqrt{2\,\mathcal{C}(\overline{\rho}_s)} \right) \mathrm{d}W_s^j .$$

  By Itô's isometry, the bounds (4.22), the triangle inequality, and the Wasserstein stability estimate (3.1b), the quadratic variation of this process is bounded from above as follows:

$$\begin{aligned}
\langle M^j \rangle_t &\leqslant C \int_0^t \left| X_s^j - \overline{X}_s^j \right|^2 \left\| \sqrt{\mathcal{C}(\mu_s^J)} - \sqrt{\mathcal{C}(\overline{\rho}_s)} \right\|_{\mathrm{F}}^2 \mathrm{d}s \\
&\leqslant C \int_0^t \left| X_s^j - \overline{X}_s^j \right|^4 + W_2\left(\mu_s^J, \overline{\mu}_s^J\right)^4 + \left\| \sqrt{\mathcal{C}(\overline{\mu}_s^J)} - \sqrt{\mathcal{C}(\overline{\rho}_s)} \right\|_{\mathrm{F}}^4 \mathrm{d}s. \qquad (4.25)
\end{aligned}$$

Combining all the bounds, we obtain

$$\begin{aligned}
\frac{1}{\kappa} \left| X_t^j - \overline{X}_t^j \right|^2 &\leqslant \int_0^t \left(\kappa \left\| \mathcal{C}(\mu_s^J) - \mathcal{C}(\overline{\mu}_s^J) \right\|_{\mathrm{F}} - \frac{c_1}{L_\phi} \right) \left| X_s^j - \overline{X}_s^j \right| \left| \nabla\phi(X_s^j) - \nabla\phi(\overline{X}_s^j) \right| \mathrm{d}s \\
&\quad + C \int_0^t \left| X_s^j - \overline{X}_s^j \right|^2 + W_2(\mu_s^J, \overline{\mu}_s^J)^2 + \left\| \mathcal{C}(\mu_s^J) - \mathcal{C}(\overline{\mu}_s^J) \right\|_{\mathrm{F}}^2 \left| \nabla\phi(\overline{X}_s^j) \right|^2 \mathrm{d}s \\
&\quad + C \int_0^t \left\| \mathcal{C}(\overline{\mu}_s^J) - \mathcal{C}(\overline{\rho}_s) \right\|_{\mathrm{F}}^2 \left| \nabla\phi(X_s^j) \right|^2 + \left\| \sqrt{\mathcal{C}(\overline{\mu}_s^J)} - \sqrt{\mathcal{C}(\overline{\rho}_s)} \right\|_{\mathrm{F}}^2 \mathrm{d}s + M_t^j \\
&=: \mathcal{I}_t^1 + \mathcal{I}_t^2 + \mathcal{I}_t^3 + M_t^j . \qquad (4.26)
\end{aligned}$$

In view of (4.20), it holds that

$$\mathcal{I}_{t\wedge\theta_J}^1 \leqslant \left(2\kappa(2R+\varepsilon)\varepsilon - \frac{c_1}{L_\phi} \right) \int_0^{t\wedge\theta_J} \left| X_s^j - \overline{X}_s^j \right| \left| \nabla\phi(X_s^j) - \nabla\phi(\overline{X}_s^j) \right| \mathrm{d}s.$$

Let ε be such that the first factor on the right-hand side is negative. Then, evaluating both sides of (4.26) at the stopping time $s \wedge \theta_J$, raising to the power $\frac{r}{2}$, and taking the supremum, we obtain

$$\begin{aligned}
\frac{1}{3^{\frac{r}{2}-1}} \sup_{s\in[0,t]} \left| X_{s\wedge\theta_J}^j - \overline{X}_{s\wedge\theta_J}^j \right|^r &\leqslant \sup_{s\in[0,t]} |\mathcal{I}_{s\wedge\theta_J}^2|^{\frac{r}{2}} + \sup_{s\in[0,t]} |\mathcal{I}_{s\wedge\theta_J}^3|^{\frac{r}{2}} + \sup_{s\in[0,t]} |M_{t\wedge\theta_J}^j|^{\frac{r}{2}} \\
&\leqslant |\mathcal{I}_{t\wedge\theta_J}^2|^{\frac{r}{2}} + |\mathcal{I}_t^3|^{\frac{r}{2}} + \sup_{s\in[0,t]} |M_{t\wedge\theta_J}^j|^{\frac{r}{2}} . \qquad (4.27)
\end{aligned}$$

As in the proof of Theorem 3, we use that

$$|\mathcal{I}^2_{t\wedge\theta_J}|^{\frac{r}{2}} \leqslant C\int_0^t \left|X^j_{s\wedge\theta_J} - \overline{X}^j_{s\wedge\theta_J}\right|^r + W_2\left(\mu^J_{s\wedge\theta_J}, \overline{\mu}^J_{s\wedge\theta_J}\right)^r + \left\|\mathcal{C}(\mu^J_{s\wedge\theta_J}) - \mathcal{C}(\overline{\mu}^J_{s\wedge\theta_J})\right\|^r_{\mathrm{F}} \left|\nabla\phi(\overline{X}^j_{s\wedge\theta_J})\right|^r \mathrm{d}s.$$

Taking the expectation in (4.27), then using (4.11), the optional stopping theorem and the Burkholder–Davis–Gundy inequality, and finally using the bound (4.25) on the quadratic variation of M^j, we deduce that

$$\begin{aligned}\mathbf{E}\left[\sup_{s\in[0,t]}\left|X^j_{s\wedge\theta_J} - \overline{X}^j_{s\wedge\theta_J}\right|^r\right] &\leqslant C\int_0^t \mathbf{E}\left|X^j_{s\wedge\theta_J} - \overline{X}^j_{s\wedge\theta_J}\right|^r + \mathbf{E}\left[\left\|\mathcal{C}(\mu^J_{s\wedge\theta_J}) - \mathcal{C}(\overline{\mu}^J_{s\wedge\theta_J})\right\|^r_{\mathrm{F}}\left|\nabla\phi(\overline{X}^j_{s\wedge\theta_J})\right|^r\right]\mathrm{d}s \\ &\quad + C\int_0^t \mathbf{E}\left[\left\|\mathcal{C}(\overline{\mu}^J_s) - \mathcal{C}(\overline{\rho}_s)\right\|^r_{\mathrm{F}}\left|\nabla\phi(X^j_s)\right|^r\right] + \mathbf{E}\left\|\sqrt{\mathcal{C}(\overline{\mu}^J_s)} - \sqrt{\mathcal{C}(\overline{\rho}_t)}\right\|^{2r}_{\mathrm{F}}\mathrm{d}s.\end{aligned}$$

By exchangeability, the Wasserstein stability estimate (3.1a) in Lemma 2, and the definition of the stopping time θ_J, it holds for some $C = C(R,\varepsilon)$ that

$$\begin{aligned}\mathbf{E}\left[\left\|\mathcal{C}(\mu^J_{s\wedge\theta_J}) - \mathcal{C}(\overline{\mu}^J_{s\wedge\theta_J})\right\|^r_{\mathrm{F}}\left|\nabla\phi(\overline{X}^j_{s\wedge\theta_J})\right|^r\right] &= \mathbf{E}\left[\left\|\mathcal{C}(\mu^J_{s\wedge\theta_J}) - \mathcal{C}(\overline{\mu}^J_{s\wedge\theta_J})\right\|^r_{\mathrm{F}}\frac{1}{J}\sum_{k=1}^J\left|\nabla\phi(\overline{X}^k_{s\wedge\theta_J})\right|^r\right] \\ &\leqslant C\mathbf{E}\left[W_2\left(\mu^J_{s\wedge\theta_J}, \overline{\mu}^J_{s\wedge\theta_J}\right)^r\frac{1}{J}\sum_{k=1}^J\left|\overline{X}^k_{s\wedge\theta_J}\right|^{r(\ell+1)}_*\right] \\ &\leqslant C\mathbf{E}\left[W_2\left(\mu^J_{s\wedge\theta_J}, \overline{\mu}^J_{s\wedge\theta_J}\right)^r\right] \leqslant C\mathbf{E}\left|X^j_{s\wedge\theta_J} - \overline{X}^j_{s\wedge\theta_J}\right|^r.\end{aligned} \tag{4.28}$$

On the other hand, by the moment bounds in Proposition 2 and by Lemma 3, it holds that

$$\begin{aligned}\mathbf{E}\left\|\sqrt{\mathcal{C}(\overline{\mu}^J_s)} - \sqrt{\mathcal{C}(\overline{\rho}_s)}\right\|^r_{\mathrm{F}} &\leqslant CJ^{-\frac{r}{2}}, \\ \mathbf{E}\left[\left\|\mathcal{C}(\overline{\mu}^J_s) - \mathcal{C}(\overline{\rho}_s)\right\|^r_{\mathrm{F}}\left|\nabla\phi(X^j_s)\right|^r\right] &\leqslant \left(\mathbf{E}\left\|\mathcal{C}(\overline{\mu}^J_s) - \mathcal{C}(\overline{\rho}_s)\right\|^{2r}_{\mathrm{F}}\right)^{\frac{1}{2}}\left(\mathbf{E}\left|\nabla\phi(X^j_s)\right|^{2r}\right)^{\frac{1}{2}} \leqslant CJ^{-\frac{r}{2}}.\end{aligned}$$

Therefore, we obtain that

$$\mathbf{E}\left[\sup_{s\in[0,t]}\left|X^j_{s\wedge\theta_J} - \overline{X}^j_{s\wedge\theta_J}\right|^r\right] \leqslant CJ^{-\frac{r}{2}} + C\int_0^t \mathbf{E}\left|X^j_{s\wedge\theta_J} - \overline{X}^j_{s\wedge\theta_J}\right|^r \mathrm{d}s.$$

By Grönwall's inequality, this implies that

$$\mathbf{E}\left[\sup_{s\in[0,t]}\left|X^j_{s\wedge\theta_J} - \overline{X}^j_{s\wedge\theta_J}\right|^r\right] \leqslant CJ^{-\frac{r}{2}}. \tag{4.29}$$

B. Bounding the second term in (4.21). We have by Propositions 1 and 2 that

$$\mathbf{E}\left[\sup_{t\in[0,T]}\left|X^j_t - \overline{X}^j_t\right|^q\right] \leqslant 2^{q-1}\left(\mathbf{E}\left[\sup_{t\in[0,T]}\left|X^j_t\right|^q\right] + \mathbf{E}\left[\sup_{t\in[0,T]}\left|\overline{X}^j_t\right|^q\right]\right) \leqslant 2^{q-1}\big(\kappa(q) + \overline{\kappa}(q)\big).$$

In order to complete the proof of the theorem, it remains to bound the probability $\mathbf{P}\left[\theta_J(R) \leqslant T\right]$. By using the same strategy as in (4.14) in the proof of Theorem 3, it is simple to show that there is C such that

$$\mathbf{P}\left[\overline{\tau}_J \leqslant T\right] \leqslant CJ^{-\frac{q}{2r(\ell+1)}}.$$

On the other hand, by the Markov inequality and (4.29), it holds that

$$\begin{aligned}\mathbf{P}\left[\mathfrak{t}_J \leqslant T \leqslant \overline{\tau}_J\right] &= \mathbf{P}\left[\sup_{t\in[0,T]} W_r\left(\mu^J_{t\wedge\theta_J}, \overline{\mu}^J_{t\wedge\theta_J}\right) \geqslant \varepsilon\right] \leqslant \mathbf{P}\left[\sup_{t\in[0,T]} \frac{1}{J}\sum_{j=1}^J \left|X^j_{t\wedge\theta_J} - \overline{X}^j_{t\wedge\theta_J}\right|^r \geqslant \varepsilon^r\right] \\ &\leqslant \mathbf{P}\left[\frac{1}{J}\sum_{j=1}^J \sup_{t\in[0,T]} \left|X^j_{t\wedge\theta_J} - \overline{X}^j_{t\wedge\theta_J}\right|^r \geqslant \varepsilon^r\right] \leqslant \frac{1}{\varepsilon^r}\mathbf{E}\left[\sup_{t\in[0,T]} \left|X^j_{t\wedge\theta_J} - \overline{X}^j_{t\wedge\theta_J}\right|^r\right] \leqslant \frac{C}{\varepsilon^r}J^{-\frac{r}{2}}.\end{aligned}$$

Given the definitions of p and q, it follows that

$$\Big(\mathbf{P}[\theta_J \leqslant T]\Big)^{\frac{q-p}{q}} \leqslant C\Big(J^{-\frac{q}{2r(\ell+1)}} + J^{-\frac{r}{2}}\Big)^{\frac{q-p}{q}} = C\Big(J^{-p} + J^{-p}\Big)^{1-\frac{1}{4(\ell+1)}} \leqslant CJ^{-\frac{p}{2}}.$$

The proof can then be concluded exactly as the proof of Theorem 3. □

4.3 Corollary: bound on the sampling error

To conclude this section, we mention the following corollary of Theorems 3 and 4, which is similar to [8, Theorem 2] and useful in the context of sampling. It states that, for the purpose of calculating the average of an observable $f\colon \mathbf{R}^d \to \mathbf{R}$ with respect to $\overline{\rho}_t$, the interacting particle system at time t is as good, in terms of convergence rate of the approximation error with respect to J, as an estimator constructed from J i.i.d. samples from $\overline{\rho}_t$.

Corollary 5 (L^p bound on the sampling error)**.** *Suppose that $\phi \in \mathcal{A}(\ell)$ for some $\ell > 0$, that $\overline{\rho}_0 \in \mathcal{P}_q(\mathbf{R}^d)$ for all $q > 0$ and that $\mathcal{C}(\overline{\rho}_0) \succ 0$. Assume additionally that f satisfies the following local Lipschitz continuity assumption:*

$$\forall x, y \in \mathbf{R}^d, \qquad \big|f(x) - f(y)\big| \leqslant L\Big(1 + |x|^s + |y|^s\Big)|x - y|. \tag{4.30}$$

Consider the systems (2.4) and (2.6) with the coefficients given in (2.5). Then for any $p \geqslant 1$, there is $C > 0$ depending on (p, L, s) and a finite number of moments of $\overline{\rho}_0$ such that

$$\left(\mathbf{E}\left|\frac{1}{J}\sum_{j=1}^J f(X^j_t) - \overline{\rho}_t[f]\right|^p\right)^{\frac{1}{p}} \leqslant CJ^{-\frac{1}{2}}.$$

Proof. By the triangle inequality, it holds that

$$\left(\mathbf{E}\left|\frac{1}{J}\sum_{j=1}^J f(X^j_t) - \overline{\rho}_t[f]\right|^p\right)^{\frac{1}{p}} \leqslant \left(\mathbf{E}\left|\frac{1}{J}\sum_{j=1}^J \Big(f\big(X^j_t\big) - f\big(\overline{X}^j_t\big)\Big)\right|^p\right)^{\frac{1}{p}} + \left(\mathbf{E}\left|\frac{1}{J}\sum_{j=1}^J f\big(\overline{X}^j_t\big) - \overline{\rho}_t[f]\right|^p\right)^{\frac{1}{p}}. \tag{4.31}$$

By Jensen's inequality, exchangeability, the local Lipschitz continuity of f, the Cauchy–Schwarz inequality, the moment bound in Proposition 2, and Theorem 3, the first term satisfies

$$\begin{aligned}\left(\mathbf{E}\left|\frac{1}{J}\sum_{j=1}^J \Big(f\big(X^j_t\big) - f\big(\overline{X}^j_t\big)\Big)\right|^p\right)^{\frac{1}{p}} &\leqslant \left(\mathbf{E}\left[\frac{1}{J}\sum_{j=1}^J \left|f\big(X^j_t\big) - f\big(\overline{X}^j_t\big)\right|^p\right]\right)^{\frac{1}{p}} = \Big(\mathbf{E}\left|f\big(X^1_t\big) - f\big(\overline{X}^1_t\big)\right|^p\Big)^{\frac{1}{p}} \\ &\leqslant L\left(3^{p-1}\mathbf{E}\left[1 + \left|X^1_t\right|^{2ps} + \left|\overline{X}^1_t\right|^{2ps}\right]\right)^{\frac{1}{2p}} \left(\mathbf{E}\left|X^1_t - \overline{X}^1_t\right|^{2p}\right)^{\frac{1}{2p}} \leqslant CJ^{-\frac{1}{2}}.\end{aligned}$$

By the Marcinkiewicz–Zygmund inequality and the moment bound in Proposition 2, the second term on the right-hand side of (4.31) also tends to 0 at the classical Monte Carlo rate $J^{-\frac{1}{2}}$, which concludes the proof. □

A Proof of well-posedness and moment bounds

We present first the proof of well-posedness for the interacting particle system in Appendix A.1, then auxiliary results in Appendix A.2, and finally the proof of well-posedness for the mean field dynamics, relying on these auxiliary results,

in Appendix A.3. In several proofs in this section, we will use that if $\phi \in \mathcal{A}(\ell)$ and $f(x) = \phi(x)^q$ for $q > 0$, then by the upper bounds in Assumption H, it holds that

$$\forall x \in \mathbf{R}^d, \qquad \left\|\mathrm{D}^2 f(x)\right\|_{\mathrm{F}} = \left\|q(q-1)\phi(x)^{q-2}\nabla\phi(x)\otimes\nabla\phi(x) + q\phi(x)^{q-1}\,\mathrm{D}^2\phi(x)\right\|_{\mathrm{F}}$$
$$\leqslant C|x|_*^{(q-2)(\ell+2)+2(\ell+1)} + C|x|_*^{(q-1)(\ell+2)+\ell} = 2C|x|_*^{q(\ell+2)-2}. \tag{A.1}$$

In addition, we repeatedly use the following inequality, which is implied by Jensen's inequality and holds for any probability distribution $\mu \in \mathbf{R}^d$, any function $h\colon \mathbf{R}^d \to \mathbf{R}^+$ and any $a, b \in \mathbf{R}^+$:

$$\int_{\mathbf{R}^d} h(x)^a\,\mu(\mathrm{d}x)\int_{\mathbf{R}^d} h(x)^b\,\mu(\mathrm{d}x) \leqslant \left(\int_{\mathbf{R}^d} h(x)^{a+b}\mu(\mathrm{d}x)\right)^{\frac{a}{a+b}}\left(\int_{\mathbf{R}^d} h(x)^{a+b}\mu(\mathrm{d}x)\right)^{\frac{b}{a+b}} = \int_{\mathbf{R}^d} h(x)^{a+b}\mu(\mathrm{d}x). \tag{A.2}$$

A.1 Proof of Proposition 1

We first prove well-posedness, and then the moment bound (2.7).

Well-posedness of the interacting particle system. This part is similar to the proof of [14, Proposition 4.4], but slightly simpler because we do not need to prove nondegeneracy of the empirical covariance. Let $\mathcal{L}$ denote the generator of the interacting particle system and let $\mathscr{X} \in \mathbf{R}^{dJ}$ be the collection $(X^1, \dots, X^J)$. Fix $j \in [\![1, J]\!]$ and $q > 0$ and let $\mathscr{V}$ denote the Lyapunov functional

$$\mathscr{V}(\mathscr{X}) = \frac{1}{J}\sum_{j=1}^J \phi(X^j)^q.$$

Let also $f(x) = \phi(x)^q$ and $\mu^J = \frac{1}{J}\sum_{j=1}^J \delta_{X^j}$. It holds that

$$\mathcal{L}\mathscr{V}(\mathscr{X}) = \frac{1}{J}\sum_{j=1}^J\Big(-\nabla\phi(X^j)^{\mathsf{T}}\,\mathcal{C}(\mu^J)\nabla_j\mathscr{V}(\mathscr{X}) + \mathcal{C}(\mu^J) : \mathrm{D}_j^2\,\mathscr{V}(\mathscr{X})\Big)$$
$$= \frac{1}{J}\sum_{j=1}^J\Big(-q\phi(X^j)^{q-1}\nabla\phi(X^j)^{\mathsf{T}}\,\mathcal{C}(\mu^J)\nabla\phi(X^j) + \mathcal{C}(\mu^J) : \mathrm{D}^2\,f(X^j)\Big) \leqslant \frac{1}{J}\sum_{j=1}^J \mathcal{C}(\mu^J) : \mathrm{D}^2\,f(X^j),$$

where ∇_j denotes the gradient with respect to X^j. Using (A.1), we deduce that

$$\mathcal{L}\mathscr{V}(\mathscr{X}) \leqslant C\left\|\mathcal{C}(\mu^J)\right\|_{\mathrm{F}}\left(\frac{1}{J}\sum_{j=1}^J|X^j|_*^{q(\ell+2)-2}\right) \leqslant C\left(\frac{1}{J}\sum_{k=1}^J|X^j|_*^2\right)\left(\frac{1}{J}\sum_{j=1}^J|X^j|_*^{q(\ell+2)-2}\right)$$
$$= C\left(\int_{\mathbf{R}^d}|x|_*^2\,\mu^J(\mathrm{d}x)\right)\left(\int_{\mathbf{R}^d}|x|_*^{q(\ell+2)-2}\,\mu^J(\mathrm{d}x)\right) \leqslant C\int_{\mathbf{R}^d}|x|_*^{q(\ell+2)}\,\mu^J(\mathrm{d}x),$$

where we used (A.2) in the last bound. Using (2.1a), we conclude that

$$\forall\mathscr{X} \in \mathbf{R}^{dJ}, \qquad \mathcal{L}\mathscr{V}(\mathscr{X}) \leqslant C\mathscr{V}(\mathscr{X}). \tag{A.3}$$

From this inequality with $q = \frac{p}{\ell+2}$, it then follows from [19, Theorem 3.5], see also [21, Theorem 2.1], that there exists a unique strong globally-defined solution to the interacting particle system if $\overline{\rho}_0 \in \mathcal{P}_p(\mathbf{R}^d)$, and that furthermore

$$\sup_{t\in[0,T]}\mathbf{E}\big[\mathscr{V}(\mathscr{X}_t)\big] < \infty. \tag{A.4}$$

Proving the moment bound (2.7). Fix $j \in [\![1, J]\!]$ and let again $f(x) = \phi(x)^q$. By Itô's formula and a reasoning similar to above, it holds that

$$f(X_t^j) \leqslant f(X_0^j) + \int_0^t \mathcal{C}(\mu_s^J) : \mathrm{D}^2\,f(X_s^j)\,\mathrm{d}s + \int_0^t \sqrt{2\,\mathcal{C}(\mu_s^J)}\nabla f(X_s^j)\,\mathrm{d}W_s^j.$$

Taking the square and the supremum, then taking the expectation and using the Burkholder–Davis–Gundy inequality, we obtain that

$$\frac{1}{3}\mathbf{E}\left[\sup_{s\in[0,t]}\left|f(X_s^j)\right|^2\right] \leqslant \mathbf{E}\left[\left|f(X_0^j)\right|^2\right] + CT\int_0^t \mathbf{E}\left|X_s^j\right|_*^{2q(\ell+2)}\,\mathrm{d}s + 2C_{\mathrm{BDG}}\int_0^t \mathbf{E}\left|\nabla f(X_s^j)^{\mathsf{T}}\,\mathcal{C}(\mu_s^J)\nabla f(X_s^j)\right|\,\mathrm{d}s. \quad \text{(A.5)}$$

Here we used that, by (A.1) and (A.2),

$$\mathbf{E}\left|\mathcal{C}(\mu_s^J) : \mathrm{D}^2 f(X_s^j)\right| \leqslant C\mathbf{E}\left|\left\|\mathcal{C}(\mu_s^J)\right\|_{\mathrm{F}}\left|X_s^j\right|_*^{q(\ell+2)-2}\right| \leqslant C\mathbf{E}\left|X_s^j\right|_*^{q(\ell+2)}.$$

Since $\left|\nabla f(x)\right| \leqslant C|x|_*^{q(\ell+2)-1}$ for all $x \in \mathbf{R}^d$ by (2.1b), we have using exchangeability that

$$\begin{aligned}\mathbf{E}\left|\nabla f(X_s^j)^{\mathsf{T}}\,\mathcal{C}(\mu_s^J)\nabla f(X_s^j)\right| &\leqslant \mathbf{E}\left[\left\|\mathcal{C}(\mu_s^J)\right\|_{\mathrm{F}}\left(\frac{1}{J}\sum_{j=1}^J\left|\nabla f(X_s^j)\right|^2\right)\right]\\ &\leqslant C\mathbf{E}\left[\int_{\mathbf{R}^d}|x|^2\mu_s^J(\mathrm{d}x)\int_{\mathbf{R}^d}|x|_*^{2q(\ell+2)-2}\,\mu_s^J(\mathrm{d}x)\right] \leqslant C\mathbf{E}|X_s^j|_*^{2q(\ell+2)}.\end{aligned}$$

where we used again inequality (A.2). Substituting this bound in (A.5) and using (2.1a), we deduce that

$$\mathbf{E}\left[\sup_{s\in[0,t]}\left|X_s^j\right|_*^{2q(\ell+2)}\right] \leqslant \mathbf{E}\left[\left|X_0^j\right|_*^{2q(\ell+2)}\right] + C\int_0^t \mathbf{E}\left|X_s^j\right|_*^{2q(\ell+2)}\,\mathrm{d}s.$$

The moment bound (2.7) then follows from (A.4), or from an application of Grönwall's inequality.

A.2 Auxiliary lemmas to establish well-posedness of the mean field dynamics

In the following, we endow the vector space $\mathcal{X} := \mathbf{R}^{d\times d}$ with the Frobenius norm, and we let $\mathcal{R}\colon \mathbf{R}^{d\times d} \to \mathbf{R}^{d\times d}$ be the map defined by $\mathcal{R}(\Gamma) = \sqrt{\Gamma\Gamma^{\mathsf{T}}}$. We prove auxiliary lemmas in this section, and postpone the proof of Proposition 2 to Appendix A.3.

Lemma 5. *Suppose that $\phi \in \mathcal{A}(\ell)$ and that $\overline{\rho}_0 \in \mathcal{P}_p(\mathbf{R}^d)$ for $p \geqslant \ell+2$. Fix $\Gamma \in C([0,T],\mathcal{X})$ and $y_0 \sim \overline{\rho}_0$. Then there is a unique strong solution $Y \in C([0,T],\mathbf{R}^d)$ to the stochastic differential equation*

$$\mathrm{d}Y_t = -\mathcal{R}(\Gamma_t)\nabla\phi(Y_t)\,\mathrm{d}t + \sqrt{2\mathcal{R}(\Gamma_t)}\,\mathrm{d}W_t, \qquad Y_0 = y_0\,, \quad \text{(A.6)}$$

In addition, the function $\mathcal{K}\colon [0,T] \to \mathcal{X}$ given by $\mathcal{K}(t) = \mathcal{C}(\rho_t)$, where $\rho_t = \mathrm{Law}(Y_t)$, belongs to $\mathrm{C}^1([0,T],\mathcal{X})$, and there is $\sigma > 0$ depending on $\|\Gamma\|_{\mathrm{C}([0,T],\mathcal{X})}$ such that

$$\|\mathcal{K}\|_{\mathrm{C}^1\left([0,T],\mathcal{X}\right)} \leqslant \sigma. \quad \text{(A.7)}$$

Proof. The existence of a unique continuous strong solution $Y \in C([0,T],\mathbf{R}^d)$ to (A.6) follows from classical theory of stochastic differential equations, for example from [19, Theorem 3.5] with the Lyapunov function $x \mapsto \phi(x)^{\frac{2}{\ell+2}}$. It remains to prove (A.7). By Itô's formula, it holds with $f(x) = \phi(x)^q$ for any $q > 0$ that

$$\mathrm{d}f(Y_t) = -q\phi(x)^{q-1}\nabla\phi(Y_t)^{\mathsf{T}}\mathcal{R}(\Gamma_t)\nabla\phi(Y_t)\,\mathrm{d}t + \mathcal{R}(\Gamma_t) : \mathrm{D}^2 f(Y_t)\,\mathrm{d}t + \nabla f(Y_t)^{\mathsf{T}}\sqrt{2\mathcal{R}(\Gamma_t)}\,\mathrm{d}W_t. \quad \text{(A.8)}$$

By (2.1b) in Assumption H, it holds for all $x \in \mathbf{R}^d$ that $\left|\nabla f(x)\right| \leqslant C|x|_*^{q(\ell+2)-1}$. Thus, writing (A.8) in integral form,

then taking the supremum and using the Burkholder–Davis–Gundy inequality, we obtain for all $t \in [0,T]$

$$\begin{aligned}\frac{1}{3}\mathbf{E}\left[\sup_{s\in[0,t]}\left|f(Y_s)\right|^2\right] &\leqslant \mathbf{E}\left|f(Y_0)\right|^2 + T\int_0^t \left\|\mathcal{R}(\Gamma_s)\right\|_{\mathrm{F}}^2 \mathbf{E}\left\|\mathrm{D}^2 f(Y_s)\right\|_{\mathrm{F}}^2 \mathrm{d}s \\ &\quad + 2C_{\mathrm{BDG}}\int_0^t \mathbf{E}\left|\nabla f(Y_s)^{\mathsf{T}}\mathcal{R}\big(\Gamma(s)\big)\nabla f(Y_s)\right| \mathrm{d}s \\ &\leqslant C\mathbf{E}|Y_0|_*^{2q(\ell+2)} + C\left(\|\Gamma\|_{C([0,T],\mathcal{X})}^2 + 1\right)\int_0^t \mathbf{E}|Y_s|_*^{2q(\ell+2)-4} + \mathbf{E}|Y_s|_*^{2q(\ell+2)-2}\,\mathrm{d}s.\end{aligned}$$

Using the lower bound in (2.1a) and rearranging, we finally obtain

$$\mathbf{E}\left[\sup_{s\in[0,t]}\left|Y\right|_*^{2q(\ell+2)}\right] \leqslant C\mathbf{E}|Y_0|_*^{2q(\ell+2)} + C\left(\|\Gamma\|_{C([0,T],\mathcal{X})}^2 + 1\right)\int_0^t \mathbf{E}\left[\sup_{u\in[0,s]}|Y_u|_*^{2q(\ell+2)}\right]\mathrm{d}s.$$

From Grönwall's inequality, it follows that

$$\mathbf{E}\left[\sup_{s\in[0,t]}\left|Y\right|_*^{2q(\ell+2)}\right] \leqslant C\mathbf{E}\left|Y_0\right|_*^{2q(\ell+2)}\,\sigma\left(\|\Gamma\|_{C([0,T],\mathcal{X})}\right), \tag{A.9}$$

for some increasing function $\sigma\colon \mathbf{R}^+ \to \mathbf{R}^+$. In particular, since $\overline{\rho}_0 \in \mathcal{P}_{\ell+2}(\mathbf{R}^d)$, we can use this inequality with $q = \frac{1}{\ell+2}$ and dominated convergence to deduce that the functions $t \mapsto \mathbf{E}[Y_t]$ and $t \mapsto \mathbf{E}[Y_t \otimes Y_t]$ are continuous. Furthermore, using (A.9) with $q = \frac{1}{2}$ we deduce that these functions are also differentiable on $[0,T]$ because, by Itô's lemma and dominated convergence,

$$\frac{1}{h}\Big(\mathbf{E}[Y_{t+h}] - \mathbf{E}[Y_t]\Big) = -\frac{1}{h}\mathbf{E}\left[\int_t^{t+h}\mathcal{R}(\Gamma_u)\nabla\phi(Y_u)\,\mathrm{d}u\right] \xrightarrow[h\to 0]{} \mathbf{E}\left[\mathcal{R}(\Gamma_t)\nabla\phi(Y_t)\right], \tag{A.10}$$

and similarly

$$\lim_{h\to 0}\frac{1}{h}\Big(\mathbf{E}[Y_{t+h}\otimes Y_{t+h}] - \mathbf{E}[Y_t\otimes Y_t]\Big) = -\mathbf{E}\Big[\big(\mathcal{R}(\Gamma_t)\nabla\phi(Y_t)\big)\otimes Y_t + Y_t\otimes\big(\mathcal{R}(\Gamma_t)\nabla\phi(Y_t)\big)\Big] + 2\mathcal{R}(\Gamma_t). \tag{A.11}$$

Another application of dominated convergence yields continuity of the derivatives. Finally, we bound the right-hand side of (A.10) as follows:

$$\Big|\mathbf{E}\left[\mathcal{R}(\Gamma_t)\nabla\phi(Y_t)\right]\Big| \leqslant \mathbf{E}\Big[\big\|\Gamma_t\big\|_{\mathrm{F}}\big|\nabla\phi(Y_t)\big|\Big] \leqslant c_u\big\|\Gamma\big\|_{C([0,T],\mathcal{X})}\mathbf{E}\big|Y_t\big|_*^{\ell+1}.$$

Employing a similar reasoning for the right-hand side of (A.11), and using (A.9), we finally obtain (A.7). □

Lemma 6. *Suppose that $\phi \in \mathcal{A}(\ell)$ for $\ell \geqslant 0$ and $\overline{\rho}_0 \in \mathcal{P}_p(\mathbf{R}^d)$ for some $p \geqslant 2$. Fix $x_0 \sim \overline{\rho}_0$ and $\xi \in [0,1]$, and suppose that $\overline{X} \in C\left([0,T],\mathbf{R}^d\right)$ is a strong solution to*

$$\begin{cases}\mathrm{d}\overline{X}_t = -\xi\,\mathcal{C}(\overline{\rho}_t)\nabla\phi(\overline{X}_t)\,\mathrm{d}t + \sqrt{2\xi\,\mathcal{C}(\overline{\rho}_t)}\,\mathrm{d}W_t, \\ \overline{\rho}_t = \mathrm{Law}(\overline{X}_t),\end{cases} \tag{A.12}$$

with initial condition $\overline{X}_0 = x_0$. Then there is $\kappa > 0$ independent of ξ such that

$$\mathbf{E}\left[\sup_{t\in[0,T]}\left|\overline{X}_t^j\right|^p\right] \leqslant \kappa, \tag{A.13}$$

Finally, if $p \geqslant \ell+2$ and $\mathcal{C}(\overline{\rho}_0) \succ 0$, then there is $\eta > 0$ such that

$$\forall t \in [0,T], \qquad \mathcal{C}(\overline{\rho}_t) \succcurlyeq \eta\,\mathrm{I}_d\,. \tag{A.14}$$

Proof. The statements (A.13) and (A.14) are obvious if $\xi = 0$. The proof for a fixed value of $\xi \in (0,1]$ is the same for all values of ξ, so for simplicity we assume from now on that $\xi = 1$.

Proof of the bound (A.13). The proof of this bound is similar to the proof of (A.9) in Lemma 5. Let $f(x) = \phi(x)^q$ for some $q > 0$. By Itô's formula, it holds that

$$\begin{aligned} \mathrm{d}f(\overline{X}_t) &= -\nabla\phi(\overline{X}_t)^\mathsf{T}\,\mathcal{C}(\overline{\rho}_t)\nabla f(\overline{X}_t)\,\mathrm{d}t + \mathcal{C}(\overline{\rho}_t) : \mathrm{D}^2 f(\overline{X}_t)\,\mathrm{d}t + \nabla f(\overline{X}_t)^\mathsf{T}\sqrt{2\,\mathcal{C}(\overline{\rho}_t)}\,\mathrm{d}W_t \\ &= -q\phi(\overline{X}_t)^{q-1}\nabla\phi(\overline{X}_t)^\mathsf{T}\,\mathcal{C}(\overline{\rho}_t)\nabla\phi(\overline{X}_t)\,\mathrm{d}t + \mathcal{C}(\overline{\rho}_t) : \mathrm{D}^2 f(\overline{X}_t)\,\mathrm{d}t + \nabla f(\overline{X}_t)^\mathsf{T}\sqrt{2\,\mathcal{C}(\overline{\rho}_t)}\,\mathrm{d}W_t. \end{aligned} \tag{A.15}$$

Writing this equation in integral form, taking the supremum and using the Burkholder–Davis–Gundy inequality, we obtain that

$$\begin{aligned} \frac{1}{3}\mathbf{E}\left[\sup_{s\in[0,t]}\left|f(\overline{X}_t)\right|^2\right] \leqslant \mathbf{E}\left[\left|f(\overline{X}_0)\right|^2\right] + T\int_0^t \left\|\mathcal{C}(\overline{\rho}_s)\right\|_{\mathrm{F}}^2 \mathbf{E}\left\|\mathrm{D}^2 f(\overline{X}_s)\right\|_{\mathrm{F}}^2 \mathrm{d}s \\ + 2C_{\mathrm{BDG}}\int_0^t \mathbf{E}\left|\nabla f(\overline{X}_s)^\mathsf{T}\,\mathcal{C}(\overline{\rho}_s)\nabla f(\overline{X}_s)\right|\,\mathrm{d}s. \end{aligned}$$

Let us bound the terms in the integrals on the right-hand side. Using the inequality $\left\|\mathcal{C}(\overline{\rho}_t)\right\|_{\mathrm{F}} \leqslant \mathbf{E}|\overline{X}_t|^2$, the bound (A.1), and then Jensen's inequality, we obtain

$$\begin{aligned} \left\|\mathcal{C}(\overline{\rho}_s)\right\|_{\mathrm{F}}^2 \mathbf{E}\left\|\mathrm{D}^2 f(\overline{X}_s)\right\|_{\mathrm{F}}^2 &\leqslant C\left(\mathbf{E}\left[\left|\overline{X}_s\right|^2\right]\right)^2 \mathbf{E}\left|\overline{X}_s\right|_*^{2q(\ell+2)-4} \\ &\leqslant C\left(\mathbf{E}\left[\left|\overline{X}_s\right|^{2q(\ell+2)}\right]\right)^{\frac{4}{2q(\ell+2)}}\left(\mathbf{E}\left[\left|\overline{X}_s\right|_*^{2q(\ell+2)}\right]\right)^{\frac{2q(\ell+2)-4}{2q(\ell+2)}} = C\mathbf{E}\left|\overline{X}_s\right|_*^{2q(\ell+2)}. \end{aligned}$$

By a similar reasoning, it holds that

$$\mathbf{E}\left|\nabla f(\overline{X}_s)^\mathsf{T}\,\mathcal{C}(\overline{\rho}_s)\nabla f(\overline{X}_s)\right| \leqslant C\mathbf{E}\left|\overline{X}_s\right|_*^{2q(\ell+2)}.$$

Therefore, using the lower bound in (2.1a), we deduce that

$$\mathbf{E}\left[\sup_{s\in[0,t]}\left|\overline{X}_s\right|_*^{2q(\ell+2)}\right] \leqslant C\mathbf{E}\left|\overline{X}_0\right|_*^{2q(\ell+2)} + C\int_0^t \mathbf{E}\left[\sup_{u\in[0,s]}\left|\overline{X}_u\right|_*^{2q(\ell+2)}\right]\mathrm{d}s$$

Letting $q = \frac{p}{2(\ell+2)}$ and using Grönwall's inequality, we obtain (A.13).

Proof of (A.14). Let $g(x) = x \otimes x$. By Itô's formula, it holds that

$$\begin{aligned} \mathrm{d}g(\overline{X}_t) = &-\big(\mathcal{C}(\overline{\rho}_t)\nabla\phi(\overline{X}_t)\big)\otimes\overline{X}_t\,\mathrm{d}t - \overline{X}_t\otimes\big(\mathcal{C}(\overline{\rho}_t)\nabla\phi(\overline{X}_t)\big)\,\mathrm{d}t + 2\,\mathcal{C}(\overline{\rho}_t)\,\mathrm{d}t \\ &+ \overline{X}_t\otimes\big(\sqrt{2\,\mathcal{C}(\overline{\rho}_t)}\,\mathrm{d}W_t\big) + \big(\sqrt{2\,\mathcal{C}(\overline{\rho}_t)}\,\mathrm{d}W_t\big)\otimes\overline{X}_t\,. \end{aligned}$$

Therefore, taking expectations, we deduce that

$$\mathbf{E}g(\overline{X}_t) - \mathbf{E}g(\overline{X}_0) = \int_0^t \mathcal{C}(\overline{\rho}_s) - \mathbf{E}\Big[\big(\mathcal{C}(\overline{\rho}_s)\nabla\phi(\overline{X}_s)\big)\otimes\overline{X}_s + \overline{X}_s\otimes\big(\mathcal{C}(\overline{\rho}_s)\nabla\phi(\overline{X}_s)\big)\Big]\,\mathrm{d}s\,.$$

On the other hand

$$\mathbf{E}[\overline{X}_t] - \mathbf{E}[\overline{X}_0] = -\int_0^t \mathcal{C}(\overline{\rho}_s)\mathbf{E}\left[\nabla\phi(\overline{X}_s)\right]\,\mathrm{d}s.$$

Combining these equations, and noting that $t \mapsto \mathcal{C}(\overline{\rho}_t)$ is differentiable by Lemma 5, we deduce that

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{C}(\overline{\rho}_t) = 2\,\mathcal{C}(\overline{\rho}_t) - \mathbf{E}\left[\Big(\mathcal{C}(\overline{\rho}_t)\nabla\phi(\overline{X}_t)\Big)\otimes\Big(\overline{X}_t - \mathcal{M}(\overline{\rho}_t)\Big)\right] - \mathbf{E}\left[\Big(\overline{X}_t - \mathcal{M}(\overline{\rho}_t)\Big)\otimes\Big(\mathcal{C}(\overline{\rho}_t)\nabla\phi(\overline{X}_t)\Big)\right].$$

Inspired by [14, Lemma A.1], we define the Lyapunov function

$$\mathcal{V}_{\mathcal{C}}(\mu) = -\log\det\big(\mathcal{C}(\mu)\big).$$

Using Jacobi's formula for the determinant, we deduce that

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{V}_{\mathcal{C}}(\overline{\rho}_t) = \mathrm{tr}\left(\mathcal{C}(\overline{\rho}_t)^{-1}\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{C}(\overline{\rho}_t)\right) = \mathrm{tr}\Big(\mathrm{I}_d - 2\mathbf{E}\Big[\nabla\phi(\overline{X}_t)\otimes\big(\overline{X}_t - \mathcal{M}(\overline{\rho}_t)\big)\Big]\Big).$$

It follows that

$$\begin{aligned}\mathcal{V}_{\mathcal{C}}(\overline{\rho}_t) - \mathcal{V}_{\mathcal{C}}(\overline{\rho}_0) &\leqslant td + 2\sqrt{d}\int_0^t \mathbf{E}\Big[\big\|\nabla\phi(\overline{X}_s)\otimes\big(\overline{X}_s - \mathcal{M}(\overline{\rho}_s)\big)\big\|_{\mathrm{F}}\Big]\\ &= td + 2\sqrt{d}\int_0^t \mathbf{E}\Big[\big|\nabla\phi(\overline{X}_s)\big|\cdot\big|\overline{X}_s - \mathcal{M}(\overline{\rho}_s)\big|\Big]\,\mathrm{d}s.\end{aligned}$$

The integrand can be bounded from (A.13) with $p = \ell + 2$, which concludes the proof since for any symmetric positive definite matrix A it holds that

$$\det(A) \leqslant \lambda_{\min}(A)\lambda_{\max}(A)^{d-1} \qquad \Rightarrow \qquad \lambda_{\min}(A) \geqslant \frac{\det A}{\|A\|_{\mathrm{F}}^{d-1}},$$

where $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$ are the minimum and maximum eigenvalues of A. □

A.3 Proof of Proposition 2

The proof of well-posedness of the mean field dynamics is based on a classical fixed point argument, applied in the vector space $\mathrm{C}\big([0,T],\mathbf{R}^{d\times d}\big)$. Similarly to the proof of [4, Theorem 3.1], we first construct a map

$$\mathcal{T}\colon C\left([0,T],\mathcal{X}\right) \to C\left([0,T],\mathcal{X}\right)$$

whose fixed points correspond to solutions of (1.3). This section follows closely the proof of [15, Theorem 2.4].

Step 1: Constructing the map $\mathcal{T}$. Consider the map

$$\begin{aligned}\mathcal{T}\colon C([0,T],\mathcal{X}) &\to C([0,T],\mathcal{X})\\ \Gamma &\mapsto \Big(t\mapsto \mathcal{C}(\rho_t)\Big),\end{aligned}$$

where $\rho_t := \mathrm{Law}(Y_t)$ and $(Y_t)_{t\in[0,T]}$ is the unique solution to (A.6) with matrix Γ. By Lemma 5, the map $\mathcal{T}$ is well-defined. Fixed points of $\mathcal{T}$ correspond to solutions of the McKean-Vlasov SDE (1.3). The existence of a fixed point follows from applying the Leray-Schauder fixed point theorem [16, Chapter 11] in the space $C\left([0,T],\mathcal{X}\right)$, once we have proved that $\mathcal{T}$ is compact and that the following set is bounded:

$$\Big\{\Gamma\in\mathrm{C}\big([0,T],\mathcal{X}\big) : \exists\xi\in[0,1] \text{ such that } \Gamma = \xi\mathcal{T}\left(\Gamma\right)\Big\}. \tag{A.16}$$

Step 2: Showing that $\mathcal{T}$ is compact. To prove that $\mathcal{T}$ is a compact operator, fix $R>0$ and consider the ball

$$B_R := \Big\{\Gamma\in C([0,T],\mathcal{X}) : \|\Gamma\|_{C([0,T],\mathcal{X})} \leqslant R\Big\}.$$

By the Arzelà–Ascoli theorem, we have a compact embedding

$$C^1\left([0,T],\mathcal{X}\right) \hookrightarrow C\left([0,T],\mathcal{X}\right).$$

Thus, it suffices to show that $\mathcal{T}(B_R)$ is bounded in $C^1\left([0,T],\mathcal{X}\right)$, which follows immediately from the assertion (A.7) in Lemma 5.

Step 3: Showing that the set (A.16) is bounded. To this end, assume that $\Gamma\in C([0,T],\mathcal{X})$ satisfies

$$\Gamma = \xi\mathcal{T}(\Gamma) \tag{A.17}$$

for some $\xi \in [0,1]$, and let (Y_t) denote the corresponding solution to (A.6). By (A.17), the stochastic process (Y_t) is also a solution to

$$\mathrm{d}Y_t = -\xi\, \mathcal{C}(\rho_t)\nabla\phi(Y_t)\,\mathrm{d}t + \sqrt{2\xi\,\mathcal{C}(\rho_t)}\,\mathrm{d}W_t, \qquad \rho_t = \mathrm{Law}(Y_t)\,.$$

By Lemma 6, it holds that $\|\mathcal{C}(\rho_t)\|$ is bounded uniformly in $[0,T]$ by a constant independent of ξ, and so the set (A.16) is indeed bounded. This establishes the existence of a fixed point of $\mathcal{T}$. Furthermore, Lemma 6 yields the first and second moment bounds in (2.8), whereas (A.7) in Lemma 5 yields the third bound in (2.8).

Step 4a: Showing uniqueness when $\phi \in \mathcal{A}(0)$. Let Γ and $\widehat{\Gamma}$ be two fixed points of $\mathcal{T}$ with corresponding solutions $Y_t, \widehat{Y}_t$ of (A.6). By definition of $\mathcal{T}$ and since $\mathcal{C}(\mu)$ is symmetric and positive semidefinite for all $\mu \in \mathcal{P}_2(\mathbf{R}^d)$, it holds that $\mathcal{R}(\Gamma_t) = \Gamma_t$ and $\mathcal{R}(\widehat{\Gamma}_t) = \widehat{\Gamma}_t$ for all $t \in [0,T]$. Let $\rho_t, \widehat{\rho}_t \in \mathcal{P}(\mathbf{R}^d)$ denote the marginal laws of Y_t and $\widehat{Y}_t$, respectively. By the Burkholder–Davis–Gundy inequality, we have for all $t \in [0,T]$

$$\frac{1}{2}\mathbf{E}\left[\sup_{s\in[0,t]}\left|Y_s - \widehat{Y}_s\right|^2\right] \leqslant T\int_0^t \mathbf{E}\left|\mathcal{C}(\rho_s)\nabla\phi(Y_s) - \mathcal{C}(\widehat{\rho}_s)\nabla\phi(\widehat{Y}_s)\right|^2 \mathrm{d}s + 2C_{\mathrm{BDG}}\int_0^t \left\|\sqrt{\mathcal{C}(\rho_s)} - \sqrt{\mathcal{C}(\widehat{\rho}_s)}\right\|_{\mathrm{F}}^2 \mathrm{d}s. \quad \text{(A.18)}$$

By Lemma 2, together with the inequality $W_p(\rho_s,\widehat{\rho}_s)^p \leqslant \mathbf{E}|Y_s - \widehat{Y}_s|^p$, which follows from the definition of the Wasserstein distance, we have that

$$\left\|\sqrt{\mathcal{C}(\rho_s)} - \sqrt{\mathcal{C}(\widehat{\rho}_s)}\right\|_{\mathrm{F}}^2 \leqslant 2W_2\big(\rho_s,\widehat{\rho}_s\big)^2 \leqslant 2\mathbf{E}|Y_s - \widehat{Y}_s|^2.$$

For the first term on the right-hand side of (A.18), we use the triangle inequality to obtain

$$\begin{aligned}
&\frac{1}{2}\mathbf{E}\left|\mathcal{C}(\rho_s)\nabla\phi(Y_s) - \mathcal{C}(\widehat{\rho}_s)\nabla\phi(\widehat{Y}_s)\right|^2 \\
&\quad\leqslant \mathbf{E}\left|\big(\mathcal{C}(\rho_s) - \mathcal{C}(\widehat{\rho}_s)\big)\nabla\phi(Y_s)\right|^2 + \mathbf{E}\left|\mathcal{C}(\widehat{\rho}_s)\big(\nabla\phi(Y_s) - \nabla\phi(\widehat{Y}_s)\big)\right|^2 \\
&\quad\leqslant C\big\|\mathcal{C}(\rho_s) - \mathcal{C}(\widehat{\rho}_s)\big\|_{\mathrm{F}}^2\mathbf{E}|Y_s|_*^2 + L_\phi\big\|\mathcal{C}(\widehat{\rho}_s)\big\|_{\mathrm{F}}^2\,\mathbf{E}\left|Y_s - \widehat{Y}_s\right|^2 \\
&\quad\leqslant C\Big(W_2(\rho_s,\delta_0) + W_2(\widehat{\rho}_s,\delta_0)\Big)^2 W_2(\rho_s,\widehat{\rho}_s)^2 + C\mathbf{E}\left|Y_s - \widehat{Y}_s\right|^2 \leqslant C\mathbf{E}\left|Y_s - \widehat{Y}_s\right|^2,
\end{aligned}$$

where we used the Lipschitz continuity (2.2), the Wasserstein stability estimate for the covariance in Lemma 2, and the moment bound established in Lemma 6. We conclude that for all $t \in [0,T]$, it holds that

$$\mathbf{E}\left[\sup_{s\in[0,t]}\left|Y_s - \widehat{Y}_s\right|^2\right] \leqslant C\int_0^t \mathbf{E}\left[\sup_{u\in[0,s]}\left|Y_u - \widehat{Y}_u\right|^2\right]\mathrm{d}s.$$

By Grönwall's lemma, it follows that the left-hand side is 0 for all $t \in [0,T]$, and so $\Gamma_t = \widehat{\Gamma}_t$ for all $t \in [0,T]$, which concludes the proof of uniqueness.

Step 4b: Showing uniqueness when $\phi \in \mathcal{A}(\ell)$ with $\ell > 0$. Recall that Y_t and $\widehat{Y}_t$ satisfy

$$\begin{aligned}
\mathrm{d}Y_t &= -\Gamma_t\nabla\phi(Y_t) + \sqrt{2\Gamma_t}\,\mathrm{d}W_t, \\
\mathrm{d}\widehat{Y}_t &= -\widehat{\Gamma}_t\nabla\phi(\widehat{Y}_t) + \sqrt{2\widehat{\Gamma}_t}\,\mathrm{d}W_t.
\end{aligned}$$

Applying Itô's formula to $f(y,\widehat{y},t) = \frac{1}{2}\big(y-\widehat{y}\big)^{\mathsf{T}}\Gamma_t^{-1}\big(y-\widehat{y}\big)$, and noting that $f(Y_0,\widehat{Y}_0,0) = 0$, we obtain that

$$\begin{aligned}
\mathbf{E}f(Y_t,\widehat{Y}_t,t) = -\mathbf{E}\int_0^t \Big\langle Y_s - \widehat{Y}_s, \nabla\phi(Y_s) - \nabla\phi(\widehat{Y}_s)\Big\rangle\,\mathrm{d}s + \mathbf{E}\int_0^t \Big\langle Y_s - \widehat{Y}_s, \Gamma_s^{-1}(\widehat{\Gamma}_s - \Gamma_s)\nabla\phi(\widehat{Y}_s)\Big\rangle\,\mathrm{d}s \\
+ \int_0^t \begin{pmatrix} \Gamma_s & \sqrt{\Gamma_s}\sqrt{\widehat{\Gamma}_s} \\ \sqrt{\widehat{\Gamma}_s}\sqrt{\Gamma_s} & \widehat{\Gamma}_s \end{pmatrix} : \begin{pmatrix} \Gamma_s^{-1} & -\Gamma_s^{-1} \\ -\Gamma_s^{-1} & \Gamma_s^{-1} \end{pmatrix} \mathrm{d}s \\
+ \mathbf{E}\int_0^t \frac{1}{2}\big(Y_s - \widehat{Y}_s\big)^{\mathsf{T}}\Gamma_s^{-1}\frac{\mathrm{d}\Gamma_s}{\mathrm{d}s}\Gamma_s^{-1}\big(Y_s - \widehat{Y}_s\big)^{\mathsf{T}}\,\mathrm{d}s.
\end{aligned}$$

Let us bound the four terms on the right-hand side.

- In view of Lemma 4, it holds that

$$-\Big\langle Y_s - \widehat{Y}_s, \nabla\phi(Y_s) - \nabla\phi(\widehat{Y}_s)\Big\rangle \leqslant c_2 \left|Y_s - \widehat{Y}_s\right|^2.$$

- For the second term, we have from the Cauchy–Schwarz inequality

$$\begin{aligned}\mathbf{E}\Big\langle Y_s - \widehat{Y}_s, \Gamma_s^{-1}(\widehat{\Gamma}_s - \Gamma_s)\nabla\phi(\widehat{Y}_s)\Big\rangle &\leqslant \left\|\Gamma_s^{-1}\right\|_{\mathrm{F}}\left\|\widehat{\Gamma}_s - \Gamma_s\right\|_{\mathrm{F}}\left(\mathbf{E}|Y_s - \widehat{Y}_s|^2\right)^{\frac{1}{2}}\left(\mathbf{E}|\nabla\phi(Y_s)|^2\right)^{\frac{1}{2}} \\ &\leqslant C\left\|\mathcal{C}(\widehat{\rho}_s) - \mathcal{C}(\rho_s)\right\|_{\mathrm{F}}\left(\mathbf{E}|Y_s - \widehat{Y}_s|^2\right)^{\frac{1}{2}} \leqslant C\mathbf{E}|Y_s - \widehat{Y}_s|^2,\end{aligned}$$

  where we used the moment bounds from Lemma 6 and the Wasserstein stability estimate from Lemma 2.

- For the third term, a simple calculation gives that

$$\begin{aligned}\begin{pmatrix}\Gamma_s & \sqrt{\Gamma_s}\sqrt{\widehat{\Gamma}_s} \\ \sqrt{\widehat{\Gamma}_s}\sqrt{\Gamma_s} & \widehat{\Gamma}_s\end{pmatrix} : \begin{pmatrix}\Gamma_s^{-1} & -\Gamma_s^{-1} \\ -\Gamma_s^{-1} & \Gamma_s^{-1}\end{pmatrix} &= \operatorname{tr}\left(\left(\sqrt{\Gamma_s} - \sqrt{\widehat{\Gamma}_s}\right)\left(\sqrt{\Gamma_s} - \sqrt{\widehat{\Gamma}_s}\right)\Gamma_s^{-1}\right) \\ &= \left\|\sqrt{\Gamma_s^{-1}}\left(\sqrt{\Gamma_s} - \sqrt{\widehat{\Gamma}_s}\right)\right\|_{\mathrm{F}}^2 \leqslant C\mathbf{E}\left|Y_t - \widehat{Y}_t\right|^2.\end{aligned}$$

  where we used the estimate (A.14) in Lemma 6 and Lemma 2 in the last inequality.

- Finally, it immediately follows from the bounds on $\Gamma_s = \mathcal{C}(\rho_s)$ and $\widehat{\Gamma}_s = \mathcal{C}(\widehat{\rho}_s)$ given Lemma 6, as well as the upper bound (A.7) on the time derivative of Γ_s, that

$$\frac{1}{2}(Y_s - \widehat{Y}_s)^{\mathsf{T}}\Gamma_s^{-1}\frac{\mathrm{d}\Gamma_s}{\mathrm{d}s}\Gamma_s^{-1}(Y_s - \widehat{Y}_s)^{\mathsf{T}} \leqslant C\left|Y_s - \widehat{Y}_s\right|^2.$$

Combining all the bounds, we deduce that

$$\mathbf{E}\left|Y_t - \widehat{Y}_t\right|^2 \leqslant C\mathbf{E}f(Y_t, \widehat{Y}_t, t) \leqslant C\int_0^t \mathbf{E}\left|Y_s - \widehat{Y}_s\right|^2 \mathrm{d}s.$$

Grönwall's inequality then gives that the left-hand side is 0 for all times, which concludes the proof.

B Proof of auxiliary results

B.1 Proof of Lemma 2

We prove the statements separately, using exactly the same approach as in [15].

Proof of (3.1a). By the triangle inequality, it holds that

$$\left\|\mathcal{C}(\mu) - \mathcal{C}(\nu)\right\|_{\mathrm{F}} \leqslant \left\|\mu[x\otimes x] - \nu[x\otimes x]\right\|_{\mathrm{F}} + \left\|\mathcal{M}(\mu)\otimes\mathcal{M}(\mu) - \mathcal{M}(\nu)\otimes\mathcal{M}(\nu)\right\|_{\mathrm{F}}. \tag{B.1}$$

Let $\pi \in \Pi(\mu,\nu)$ denote an arbitrary coupling between μ and ν. By Jensen's inequality, it holds that

$$\left\|\mu[x\otimes x] - \nu[x\otimes x]\right\|_{\mathrm{F}} \leqslant \iint_{\mathbf{R}^d\times\mathbf{R}^d} \|x\otimes x - y\otimes y\|_{\mathrm{F}}\,\pi(\mathrm{d}x\,\mathrm{d}y).$$

Recall that for all $(x,y) \in \mathbf{R}^d \times \mathbf{R}^d$, it holds that

$$\begin{aligned}\|x\otimes x - y\otimes y\|_{\mathrm{F}} &= \left\|x(x-y)^{\mathsf{T}} + (x-y)y^{\mathsf{T}}\right\|_{\mathrm{F}} \\ &\leqslant \left\|x(x-y)^{\mathsf{T}}\right\|_{\mathrm{F}} + \left\|(x-y)y^{\mathsf{T}}\right\|_{\mathrm{F}} = |x|\cdot|x-y| + |y|\cdot|x-y| = \big(|x|+|y|\big)|x-y|.\end{aligned} \tag{B.2}$$

Therefore, we deduce that

$$\left\|\mu[x\otimes x]-\nu[x\otimes x]\right\|_{\mathrm{F}}\leqslant\sqrt{\iint_{\mathbf{R}^d\times\mathbf{R}^d}\left(|x|+|y|\right)^2\pi(\mathrm{d}x\,\mathrm{d}y)}\sqrt{\iint_{\mathbf{R}^d\times\mathbf{R}^d}|x-y|^2\,\pi(\mathrm{d}x\,\mathrm{d}y)}.$$

Infimizing over all couplings $\pi\in\Pi(\mu,\nu)$, we conclude that

$$\left\|\mu[x\otimes x]-\nu[x\otimes x]\right\|_{\mathrm{F}}\leqslant\Big(W_2(\mu,\delta_0)+W_2(\nu,\delta_0)\Big)W_2(\mu,\nu).$$

Applying a similar reasoning to the second term on the right-hand side of (B.1), we deduce (3.1a).

Proof of (3.1b). It is sufficient to check the claim for measures μ and ν of the form

$$\mu^J=\frac{1}{J}\sum_{j=1}^{J}\delta_{X^j},\qquad\nu^J=\frac{1}{J}\sum_{j=1}^{J}\delta_{Y^j},\qquad J\in\mathbf{N}^+. \tag{B.3}$$

Indeed, assume that the statement holds for all such pairs of probability measures, and take $(\mu,\nu)\in\mathcal{P}_2(\mathbf{R}^d)\times\mathcal{P}_2(\mathbf{R}^d)$. By [27, Theorem 6.18], there exists a sequence $\left\{(\mu^J,\nu^J)\right\}_{J\in\mathbf{N}^+}$ in $\mathcal{P}_2(\mathbf{R}^d)\times\mathcal{P}_2(\mathbf{R}^d)$ such that $W_2(\mu^J,\mu)\to 0$ and $W_2(\nu^J,\nu)\to 0$ in the limit as $J\to\infty$. Then

$$\begin{aligned}\left\|\sqrt{\mathcal{C}(\mu)}-\sqrt{\mathcal{C}(\nu)}\right\|_{\mathrm{F}}\leqslant&\left\|\sqrt{\mathcal{C}(\mu^J)}-\sqrt{\mathcal{C}(\nu^J)}\right\|_{\mathrm{F}}\\&+\left\|\sqrt{\mathcal{C}(\mu)}-\sqrt{\mathcal{C}(\mu^J)}\right\|_{\mathrm{F}}+\left\|\sqrt{\mathcal{C}(\nu)}-\sqrt{\mathcal{C}(\nu^J)}\right\|_{\mathrm{F}}.\end{aligned}$$

The first term is bounded from above by $\sqrt{2}W_2(\mu^J,\nu^J)$ by the base case, while the other two terms converge to 0 in the limit as $J\to\infty$ by (3.1a) in Lemma 2. Taking the limit $J\to\infty$, we deduce that

$$\left\|\sqrt{\mathcal{C}(\mu)}-\sqrt{\mathcal{C}(\nu)}\right\|_{\mathrm{F}}\leqslant\sqrt{2}W_2(\mu,\nu).$$

Proof of the statement for empirical measures. By [28, p.5], the Wasserstein distance between empirical measures μ^J and ν^J of the form (B.3) is equal to

$$W_2(\mu^J,\nu^J)=\min_{\sigma\in\mathcal{S}_J}\left(\frac{1}{J}\sum_{j=1}^{J}\left|X^j-Y^{\sigma(j)}\right|^2\right)^{\frac{1}{2}},$$

where $\mathcal{S}_J$ denotes the set of permutations in $\{1,\dots,J\}$. Thus, the claim will follow if we can prove that, for any pair of probability measures $(\mu^J,\nu^J)\in\mathcal{P}_2(\mathbf{R}^d)\times\mathcal{P}_2(\mathbf{R}^d)$ of the form (B.3), it holds that

$$\left\|\sqrt{\mathcal{C}(\mu^J)}-\sqrt{\mathcal{C}(\nu^J)}\right\|_{\mathrm{F}}\leqslant\sqrt{2}\left(\frac{1}{J}\sum_{j=1}^{J}\left|X^j-Y^j\right|^p\right)^{\frac{1}{p}}. \tag{B.4}$$

We henceforth drop the superscript J in μ^J,ν^J for simplicity, and write $\mathbf{X}=\left(X^1,\dots,X^J\right)$ and $\mathbf{Y}=\left(Y^1,\dots,Y^J\right)$. The proof of (B.4) presented below follows the lines of a proof shown to me by N. J. Gerber, who proved this inequality in preliminary work with F. Hoffmann which eventually lead to the preprint [15]. First note that

$$\mathcal{C}(\mu)=M_{\mathbf{X}}M_{\mathbf{X}}^{\mathsf{T}},\qquad M_{\mathbf{X}}:=\frac{1}{\sqrt{J}}\begin{pmatrix}\left(X^1-\mathcal{M}(\mu)\right)&\dots&\left(X^J-\mathcal{M}(\mu)\right)\end{pmatrix}.$$

Proceeding in the same manner, we construct a matrix $M_{\mathbf{Y}}\in\mathbf{R}^{d\times J}$ such that $\mathcal{M}(\nu)=M_{\mathbf{Y}}M_{\mathbf{Y}}^{\mathsf{T}}$. A result by Araki and Yamagami [1], later generalized by Kittaneh [20] and Bhatia [3], states for any two matrices A and B with the same shape, it holds that

$$\left\|\sqrt{A^{\mathsf{T}}A}-\sqrt{B^{\mathsf{T}}B}\right\|_{\mathrm{F}}\leqslant\sqrt{2}\left\|A-B\right\|_{\mathrm{F}}.$$

See also [2, Theorem VII.5.7] for a textbook presentation. This result, applied with $A = M_{\mathbf{X}}$ and $B = M_{\mathbf{Y}}$, yields

$$\left\|\sqrt{\mathcal{C}(\mu)} - \sqrt{\mathcal{C}(\nu)}\right\|_{\mathrm{F}} \leqslant \sqrt{2}\left(\frac{1}{J}\sum_{j=1}^{J}\left|\left(X^j - \mathcal{M}(\mu)\right) - \left(Y^j - \mathcal{M}(\nu)\right)\right|^2\right)^{\frac{1}{2}}$$
$$= \min_{a\in\mathbf{R}^d} \sqrt{2}\left(\frac{1}{J}\sum_{j=1}^{J}\left|X^j - Y^j - a\right|^2\right)^{\frac{1}{2}} \leqslant \sqrt{2}\left(\frac{1}{J}\sum_{j=1}^{J}\left|X^j - Y^j\right|^2\right)^{\frac{1}{2}},$$

which shows (B.4) and completes the proof.

B.2 Proof of Lemma 3

Equation (3.3) follows from (3.2), and from an inequality due to van Hemmen and Ando [26], see also [2, Problem X.5.5], which in view of the assumption $\mathcal{C}(\mu) \succcurlyeq \eta\,\mathrm{I}_d \succ 0$ gives that

$$\left\|\sqrt{\mathcal{C}(\overline{\mu}^J)} - \sqrt{\mathcal{C}(\mu)}\right\|_{\mathrm{F}} \leqslant \frac{1}{\eta}\left\|\mathcal{C}(\overline{\mu}^J) - \mathcal{C}(\mu)\right\|_{\mathrm{F}}.$$

The bound (3.2) follows from usual Monte Carlo estimates. Using the triangle inequality, we have that

$$\left\|\mathcal{C}(\overline{\mu}^J) - \mathcal{C}(\mu)\right\|_{\mathrm{F}}^p = \left\|\overline{\mu}^J\left[x\otimes x\right] - \mu\left[x\otimes x\right] - \mathcal{M}(\overline{\mu}^J)\otimes\mathcal{M}(\overline{\mu}^J) + \mathcal{M}(\mu)\otimes\mathcal{M}(\mu)\right\|_{\mathrm{F}}^p$$
$$\leqslant 2^{p-1}\left\|\overline{\mu}^J\left[x\otimes x\right] - \mu\left[x\otimes x\right]\right\|_{\mathrm{F}}^p + 2^{p-1}\left\|\mathcal{M}(\overline{\mu}^J)\otimes\mathcal{M}(\overline{\mu}^J) - \mathcal{M}(\mu)\otimes\mathcal{M}(\mu)\right\|_{\mathrm{F}}^p.$$

Convergence to zero of the expectation of the first term with rate $J^{-\frac{p}{2}}$ follows from the Marcinkiewicz–Zygmund inequality. For the second term, we use (B.2) to obtain that

$$\mathbf{E}\left\|\mathcal{M}(\overline{\mu}^J)\otimes\mathcal{M}(\overline{\mu}^J) - \mathcal{M}(\mu)\otimes\mathcal{M}(\mu)\right\|_{\mathrm{F}}^p$$
$$\leqslant \mathbf{E}\left[\left(\left|\mathcal{M}(\overline{\mu}^J)\right| + \left|\mathcal{M}(\mu)\right|\right)^p\left|\mathcal{M}(\overline{\mu}^J) - \mathcal{M}(\mu)\right|^p\right]$$
$$\leqslant \left(\mathbf{E}\left[\left(\left|\mathcal{M}(\overline{\mu}^J)\right| + \left|\mathcal{M}(\mu)\right|\right)^{2p}\right]\mathbf{E}\left[\left|\mathcal{M}(\overline{\mu}^J) - \mathcal{M}(\mu)\right|^{2p}\right]\right)^{\frac{1}{2}}$$
$$\leqslant CJ^{-\frac{p}{2}}\left(\mathbf{E}\left[\left(\left|\mathcal{M}(\overline{\mu}^J)\right| + \left|\mathcal{M}(\mu)\right|\right)^{2p}\right]\right)^{\frac{1}{2}},$$

where we used again the Marcinkiewicz–Zygmund inequality. The claim follows since $\mathbf{E}\left[\left|\mathcal{M}(\overline{\mu}^J)\right|^{2p}\right] \leqslant \mathbf{E}\left[\left|\overline{X}^j\right|^{2p}\right]$ by Jensen's inequality.

Acknowledgements. The author is grateful to Zhiyan Ding, Nicolai Gerber, Franca Hoffmann, Qin Li and Julien Reygner for useful discussions. UV is partially supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 810367), and by the Agence Nationale de la Recherche under grants ANR-21-CE40-0006 (SINEQ) and ANR-23-CE40-0027 (IPSO).

References

[1] H. Araki and S. Yamagami. An inequality for Hilbert-Schmidt norm. *Comm. Math. Phys.*, **81**(1):89–96, 1981.

[2] R. Bhatia. *Matrix analysis*, volume **169** of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1997.

[3] R. Bhatia. Matrix factorizations and their perturbations. In volume **197/198**, 1994. Second Conference of the International Linear Algebra Society (Lisbon, 1992).

[4] J. A. Carrillo, Y.-P. Choi, C. Totzeck and O. Tse. An analytical framework for consensus-based global optimization method. *Math. Models Methods Appl. Sci.*, **28**(6):1037–1066, 2018.

[5] J. A. Carrillo and U. Vaes. Wasserstein stability estimates for covariance-preconditioned Fokker-Planck equations. *Nonlinearity*, **34**(4):2275–2295, 2021.

[6] L.-P. Chaintron and A. Diez. Propagation of chaos: a review of models, methods and applications. I. Models and methods. *Kinet. Relat. Models*, **15**(6):895–1015, 2022.

[7] L.-P. Chaintron and A. Diez. Propagation of chaos: a review of models, methods and applications. II. Applications. *Kinet. Relat. Models*, **15**(6):1017–1173, 2022.

[8] Z. Ding and Q. Li. Ensemble Kalman inversion: mean-field limit and convergence analysis. *Stat. Comput.*, **31**(1):Paper No. 9, 21, 2021.

[9] Z. Ding and Q. Li. Ensemble Kalman sampler: mean-field limit and convergence analysis. *SIAM J. Math. Anal.*, **53**(2):1546–1578, 2021.

[10] O. R. A. Dunbar, A. B. Duncan, A. M. Stuart and M.-T. Wolfram. Ensemble inference methods for models with noisy and expensive likelihoods. *SIAM J. Appl. Dyn. Syst.*, **21**(2):1539–1572, 2022.

[11] S. N. Ethier and T. G. Kurtz. *Markov processes*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York, 1986. Characterization and convergence.

[12] N. Fournier and A. Guillin. On the rate of convergence in Wasserstein distance of the empirical measure. *Probab. Theory Related Fields*, **162**(3-4):707–738, 2015.

[13] A. Garbuno-Inigo, F. Hoffmann, W. Li and A. M. Stuart. Interacting Langevin diffusions: gradient structure and ensemble Kalman sampler. *SIAM J. Appl. Dyn. Syst.*, **19**(1):412–441, 2020.

[14] A. Garbuno-Inigo, N. Nüsken and S. Reich. Affine invariant interacting Langevin dynamics for Bayesian inference. *SIAM J. Appl. Dyn. Syst.*, **19**(3):1633–1658, 2020.

[15] N. J. Gerber, F. Hoffmann and U. Vaes. Mean-field limits for consensus-based optimization and sampling. **2312.07373**, 2023.

[16] D. Gilbarg and N. S. Trudinger. *Elliptic Partial Differential Equations of Second Order*. Classics in Mathematics. Springer-Verlag, Berlin, 2001. Reprint of the 1998 edition.

[17] D. J. Higham, X. Mao and A. M. Stuart. Strong convergence of Euler-type methods for nonlinear stochastic differential equations. *SIAM J. Numer. Anal.*, **40**(3):1041–1063, 2002.

[18] D. Kalise, A. Sharma and M. V. Tretyakov. Consensus-based optimization via jump-diffusion stochastic differential equations. *Math. Models Methods Appl. Sci.*, **33**(2):289–339, 2023.

[19] R. Khasminskii. *Stochastic Stability of Differential Equations*, volume **66** of *Stochastic Modelling and Applied Probability*. Springer, Heidelberg, second edition, 2012. With contributions by G. N. Milstein and M. B. Nevelson.

[20] F. Kittaneh. On Lipschitz functions of normal operators. *Proc. Amer. Math. Soc.*, **94**(3):416–418, 1985.

[21] S. P. Meyn and R. L. Tweedie. Stability of Markovian processes. III. Foster-Lyapunov criteria for continuous-time processes. *Adv. in Appl. Probab.*, **25**(3):518–548, 1993.

[22] M. Mitzenmacher and E. Upfal. *Probability and computing*. Cambridge University Press, Cambridge, second edition, 2017, Randomization and probabilistic techniques in algorithms and data analysis.

[23] N. Nüsken and S. Reich. Note on interacting langevin diffusions: gradient structure and ensemble kalman sampler by garbuno-inigo, hoffmann, li and stuart. **1908.10890**, 2019.

[24] A. M. Stuart. Inverse problems: a Bayesian perspective. *Acta Numer.*, **19**:451–559, 2010.

[25] A.-S. Sznitman. Topics in propagation of chaos. In *École d'Été de Probabilités de Saint-Flour XIX—1989*. volume **1464**, Lecture Notes in Math. Springer, Berlin, 1991.

[26] J. L. van Hemmen and T. Ando. An inequality for trace ideals. *Comm. Math. Phys.*, **76**(2):143–148, 1980.

[27] C. Villani. *Optimal Transport*, volume **338** of *Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 2009.

[28] C. Villani. *Topics in Optimal Transportation*, volume **58** of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2003.