Electron. Commun. Probab. **26** (2021), article no. 45, 1–13.
https://doi.org/10.1214/21-ECP416
ISSN: 1083-589X

ELECTRONIC COMMUNICATIONS in PROBABILITY

Maximum likelihood estimation of potential energy in interacting particle systems from single-trajectory data*

Xiaohui Chen†

Abstract

This paper concerns the parameter estimation problem for the quadratic potential energy in interacting particle systems from continuous-time and single-trajectory data. Even though such dynamical systems are high-dimensional, we show that the vanilla maximum likelihood estimator (without regularization) is able to estimate the interaction potential parameter with optimal rate of convergence simultaneously in mean-field limit *and* in long-time dynamics. This to some extend avoids the curse-of-dimensionality for estimating large dynamical systems under symmetry of the particle interaction.

Keywords: interacting particle systems; maximum likelihood estimation; mean-field regime; stochastic Vlasov equation; symmetry.
MSC2020 subject classifications: 60H15; 62M05.
Submitted to ECP on July 21, 2020, final version accepted on June 30, 2021.

1 Introduction

Dynamical systems of interacting particles have a wide range of applications on modeling collective behaviors in physics [5], biology [14, 19], social science [15], and more recently in machine learning as useful tools to understand the stochastic gradient descent (SGD) dynamics on neural networks [13]. Due to the large number of particles, such dynamical systems are high-dimensional even for single-trajectory data from each particle, and statistical learning problems for lower-dimensional interaction functionals from data are usually challenging [1, 10, 11]. In this paper, we propose a likelihood based inference to estimate the parameters of the interaction function induced by a potential energy in an N interacting particle system based on their observed (continuous-time) trajectories, and establish its statistical guarantees.

1.1 Interacting N-particle systems

In statistical mechanics, microscopic behaviors of a large number of random particles are related to explain macroscopic physical quantities (such as temperature distributions). Specifically, a system of N interacting particles $(X_t^{N,1}, \dots, X_t^{N,N})$ can be described by stochastic differential equations (SDEs) of the form:

$$\mathrm{d}X_t^{N,i} = \frac{1}{N}\sum_{j=1}^{N} b(X_t^{N,j} - X_t^{N,i})\mathrm{d}t + \sigma \mathrm{d}W_t^i, \tag{1.1}$$

*Research was supported in part by NSF CAREER Award DMS-1752614 and a Simons Fellowship.
†University of Illinois at Urbana-Champaign, United States of America. E-mail: xhchen@illinois.edu

where $b : \mathbb{R}^d \to \mathbb{R}^d$ is a vector field representing the pairwise interaction between the particles, $(W_t^1)_{t\geqslant 0}, \dots, (W_t^N)_{t\geqslant 0}$ are N independent copies of the standard Brownian motion in $\mathbb{R}^d$ such that $W_0^i = 0$, and $\sigma \in \mathbb{R}_+$ is a diffusion parameter which is assumed to be constant and known. The scaling in (1.1) puts us in the mean-field regime, where the pairwise interaction effect is weak and decays on the order of $1/N$ as the number of particles $N \to \infty$. Thus the total interaction effect remains $O(1)$.

In this paper, we consider the estimation problem of the interaction function b parametrized by a linear approximation $b(x) = \Theta x$ for some unknown $d \times d$ (symmetric) positive-definite matrix $\Theta \succ 0$:

$$\mathrm{d}X_t^{N,i} = \Theta(\overline{X}_t^N - X_t^{N,i})\mathrm{d}t + \sigma \mathrm{d}W_t^i, \tag{1.2}$$

and $\overline{X}_t^N = N^{-1}\sum_{j=1}^N X_t^{N,j}$. Interaction in stochastic system (1.2) relates to the *Hookean* behavior for capturing the *linear elasticity* where the interaction force b scales linearly with deformation distance (due to compression and stretch) in the direction from $X_t^{N,j}$ to $X_t^{N,i}$. This type of interaction is extensively used to study the large-scale and long-time dynamics of protein folding as an elastic mass-and-spring network of small Cα atoms [7, 6].

It is a classical result [18] that the N-particle SDE system (1.1) in the mean-field limit admits a unique strong solution $(X_t^{N,i})_{t\geqslant 0}$ if b is (globally) Lipschitz, which is the case for the stochastic system (1.2). Based on the observed continuous-time and single-trajectory data of the particle movement $(X_t^{N,1})_{t\geqslant 0}, \dots, (X_t^{N,N})_{t\geqslant 0}$ in the interacting particle system (1.2), our main focus is to estimate the interaction parameter Θ in the potential energy.

Note that the first-order dynamical system (1.2) evolves as stochastic gradient flows in $\mathbb{R}^d$:

$$\frac{\mathrm{d}X_t^{N,i}}{\mathrm{d}t} = \frac{1}{N}\sum_{j=1}^N \nabla V(X_t^{N,j} - X_t^{N,i}) + \sigma\xi_t^i, \tag{1.3}$$

which corresponds to a quadratic potential energy $V(x) = \frac{1}{2}x^T\Theta x$ and ξ_t^i are i.i.d. standard Gaussian random vectors in $\mathbb{R}^d$. The left-hand side of (1.3) is the observed velocity vector of particle i at time t and the right-hand side of (1.3) is a linear function of all particle trajectories at time t corrupted by independent additive Gaussian noise. Thus, there are Nd SDEs with observed N trajectory data in dimension d to solve in (1.3) and estimation problem for Θ can be recast as a high-dimensional linear regression problem in an augmented space $\mathbb{R}^{Nd}$ (cf. the equivalent form (2.1) in Section 2). Nevertheless, the trajectory data are *temporally dependent* samples since the particles are interacting and dynamic, so that theoretical guarantees on estimating structured coefficients in high-dimensional linear models with i.i.d. samples are no long applicable in our context [3]. Moreover, due to the *symmetry* of the particles in law, the regression coefficients have very special replicated block diagonal structure in $\mathbb{R}^{Nd}$, which suggests that regularization techniques may not necessarily needed in our problem. Indeed, we show that a direct likelihood-ratio method suffices to estimate Θ with the optimal rate of convergence in this work.

1.2 Stochastic Vlasov equation: decoupled mean-field limit

Let $\rho_t^N = N^{-1}\sum_{j=1}^N \delta_{X_t^{N,j}}$ be the empirical measure of the N particles at time t. Then we can alternatively write (1.1) as

$$\mathrm{d}X_t^{N,i} = \Big(\int_{\mathbb{R}^d} b(y - X_t^{N,i})\rho_t^N(\mathrm{d}y)\Big)\mathrm{d}t + \sigma\mathrm{d}W_t^i, \tag{1.4}$$

where the drift coefficient vector depends on the individual state $X_t^{N,i}$ and the distribution ρ_t^N (due to interaction). As $N \to \infty$ (i.e., in the mean-field limit), the interaction contributed by any pair of particles in the N-particle system (1.1) vanishes and all particles are (asymptotically) i.i.d. since they have the same drift and diffusion coefficients driven by independent standard Brownian motion in $\mathbb{R}^d$. By the law of large numbers, we see that for any fixed t, $\rho_t^N \to \rho_t$ as $N \to \infty$, and the dynamic system (1.4) becomes

$$\mathrm{d}Y_t^i = \Big(\int_{\mathbb{R}^d} b(y - Y_t^i)\rho_t(\mathrm{d}y) \Big)\mathrm{d}t + \sigma \mathrm{d}W_t^i, \tag{1.5}$$

where $(\rho_t)_{t\geqslant 0}$ is a non-random measure flow. Since the particles are symmetric in distribution, ρ_t is actually the limiting law of each particle $X_t^i, i = 1, \ldots, N$. This defines a system of independent *stochastic Vlasov equations* (1.5) with $\rho_t = \mathcal{L}(Y_t^i)$, which is a class of Markov processes with nonlinear dynamics [12].

For quadratic potential $V(x) = \frac{1}{2}x^T\Theta x$, the stochastic system (1.2) depends on the empirical distribution ρ_t^N through the empirical mean $\overline{X}_t^N$ and it can be decoupled and approximated by a system of independent mean-reverting processes. Averaging the N SDEs in (1.2), we see that once again by the law of number numbers,

$$\overline{X}_t^N = \overline{W}_t^N := \frac{1}{N}\sum_{i=1}^N W_t^t \to 0, \quad \text{as } N \to \infty,$$

which means that the interaction effect $\overline{X}_t^N$ becomes deterministic and it is nicely *decoupled* in the mean-field limit. Thus we expect that the i-th particle process $(X_t^{N,i})_{t\geqslant 0}$ can be (independently) approximated by a limiting process $(Y_t^i)_{t\geqslant 0}$ given by

$$\mathrm{d}Y_t^i = -\Theta Y_t^i \mathrm{d}t + \sigma \mathrm{d}W_t^i. \tag{1.6}$$

Note that the processes $(Y_t^i)_{t\geqslant 0}$ (1.6) are *independent* copies of the Ornstein-Uhlenbeck (OU) processes, which is a linear dynamic system of N independent particles.

1.3 Existing literature

Learnability (i.e., identifiability) of interaction functions in interacting particle systems under the coercivity condition were studied in [11, 10]. In the noiseless setting, estimation of the interaction kernel, a scalar-valued function of pairwise distance between particles in the system, was first studied in [1] for single-trajectory data in the mean-field limit, where the rate of convergence is no faster than $N^{-1/d}$. To alleviate the curse-of-dimensionality, sparsity-promoting techniques were considered for some structured high-dimensional dynamical systems [2, 17]. [11] showed that a least-squares estimator achieves the optimal rate of convergence (in the number of observed trajectories for each particle) for estimating the interaction kernel based on multiple-trajectory data sampled from a deterministic system with random initialization. Estimation of the diffusion parameter for interacting particle systems from noisy trajectory data was studied in [8]. Consistency of parameter estimation of the general McKean-Vlasov equation by the maximum likelihood estimation is studied in [21]. To the best of our knowledge, there is no existing work, regularized or not, establishes the optimal rate of convergence for interaction parameter estimation simultaneously in the large N (mean-field limit) and large t (long-time dynamics) regime. This work fills this gap for the linear elasticity interacting particle systems.

1.4 Notation

For two generic vectors $a, b \in \mathbb{R}^d$, we use $a \cdot b = \sum_{j=1}^d a_j b_j$ to denote the inner product of a and b. We use $\|a\| = (a \cdot a)^{1/2}$ to denote its Euclidean norm. For a generic

matrix M, we use $\|M\|$ to denote its spectral norm. Denote the set of $d \times d$ (symmetric) positive-definite matrices by $\mathbb{S}_+^{d\times d}$. We use C and c to denote positive universal constants whose values may vary from place to place.

2 Maximum likelihood estimation

We estimate $\Theta \succ 0$ by a likelihood based method. In view of propagation of chaos [18], we assume that the initializations $X_0^{N,i}$ are some mean-zero independent random vectors in $\mathbb{R}^d$. Let $\mathbf{W}_t = (W_t^1, \dots, W_t^N) \in \mathbb{R}^{Nd}$ be the N stacked standard d-dimensional Brownian motion and $\mathbf{X}_t^N = (X_t^{N,1}, \dots, X_t^{N,N}) \in \mathbb{R}^{Nd}$ be the stacked observation process. Then system (1.2) can be rewritten as a higher-dimensional mean-reverting process in the augmented space $\mathbb{R}^{Nd}$:

$$\mathrm{d}\mathbf{X}_t^N = -\mathbf{\Theta} H \mathbf{X}_t^N \mathrm{d}t + \Sigma \mathrm{d}\mathbf{W}_t, \tag{2.1}$$

where $\mathbf{\Theta} = \mathrm{diag}(\Theta, \dots, \Theta)$ is an $(Nd) \times (Nd)$ block diagonal matrix, $\Sigma = \mathrm{diag}(\sigma^2, \dots, \sigma^2)$ is an $(Nd) \times (Nd)$ diagonal matrix, and

$$H = \frac{1}{N} \begin{pmatrix} (N-1)I_d & -I_d & \dots & -I_d \\ -I_d & (N-1)I_d & \dots & -I_d \\ \vdots & \vdots & \ddots & \vdots \\ -I_d & -I_d & \dots & (N-1)I_d \end{pmatrix}$$

is an $(Nd) \times (Nd)$ *interaction matrix*. Note that H is a projection matrix $H^2 = H$, which implies that the interaction effect is homogeneous.

Let P be the law of the standard (Nd)-dimensional Brownian motion $(\mathbf{W}_t)_{t\geqslant 0}$ and Q be the law of the augmented observation process $(\mathbf{X}_t^N)_{t\geqslant 0}$. By the multivariate Girsanov theorem for changing measures (cf. Theorem 1.12 in [9]), the likelihood ratio of $(\mathbf{X}_t^N)_{t\geqslant 0}$ in (1.2) and $(\mathbf{W}_t)_{t\geqslant 0}$ is given by the Radon-Nikodym derivative $\frac{\mathrm{d}Q}{\mathrm{d}P}(\mathbf{X}_0^{N,t}, \rho_0^t) =: e^{\overline{\ell}_t^N(A)}$, where

$$\overline{\ell}_t^N(A) = \sum_{i=1}^N \Big[-\frac{1}{2} \int_0^t \|A(\overline{X}_s^N - X_s^{N,i})\|^2 \,\mathrm{d}s + \int_0^t A(\overline{X}_s^N - X_s^{N,i}) \cdot \mathrm{d}X_s^{N,i} \Big], \tag{2.2}$$

$A \in \mathbb{S}_+^{d\times d}$, $\mathbf{X}_0^{N,t} = (\mathbf{X}_s^N)_{s\in[0,t]}$, and $\rho_0^t = (\rho_s)_{s\in[0,t]}$. Then the maximum likelihood estimator (MLE) for Θ is defined as

$$\hat{\Theta}_t^N = \mathrm{argmax}_{A\in\mathbb{S}_+^{d\times d}} \overline{\ell}_t^N(A). \tag{2.3}$$

Remark 2.1 (Computing the MLE). The MLE $\hat{\Theta}_t^N$ in (2.3) is a constrained optimization problem on the smooth manifold $\mathbb{S}_+^{d\times d}$. Since the objective function $\overline{\ell}_t^N(A)$ is quadratic in A, we may easily maximize $\overline{\ell}_t^N(A)$ over all possible $d \times d$ matrices in closed form that may not necessarily be a symmetric positive-definite matrix (cf. equation (3.4) below), and then project the unconstrained maximizer $\tilde{\Theta}_t^N = \mathrm{argmax}_{A\in\mathbb{R}^{d\times d}} \overline{\ell}_t^N(A)$ into $\mathbb{S}_+^{d\times d}$ by a procedure called *positive-definitization* that has been used to estimate high-dimensional covariance matrix (cf. Section 2.2 in [4]). Let $\breve{\Theta}_t^N = \sum_{j=1}^d \breve{\lambda}_j u_j u_j^T$ be its eigen-decomposition of the symmetrized matrix $\breve{\Theta}_t^N = (\tilde{\Theta}_t^N + (\tilde{\Theta}_t^N)^T)/2$.

Lemma 2.2 (Positive-definitization preserves the rate). If $\|\tilde{\Theta}_t^N - \Theta\| \leqslant r_n$, then $\|\breve{\Theta}_t^N - \Theta\| \leqslant 3r_n$, where $\breve{\Theta}_t^N = \sum_{j=1}^d \max(\breve{\lambda}_j, r_n) u_j u_j^T$.

Proof of Lemma 2.2. First, triangle inequality yields $\|\breve{\Theta}_t^N - \Theta\| \leqslant \|\hat{\Theta}_t^N - \Theta\| \leqslant r_n$. Note

that

$$\|\check{\Theta}_t^N - \Theta\| \leqslant \|\check{\Theta}_t^N - \breve{\Theta}_t^N\| + \|\breve{\Theta}_t^N - \Theta\| \leqslant \Big| \sum_{j=1}^{d} \big[(\max(\breve{\lambda}_j, r_n) - \breve{\lambda}_j)\big] u_j u_j^T \Big| + r_n$$
$$= \max_{j\in[d]} \big| \max(\breve{\lambda}_j, r_n) - \breve{\lambda}_j \big| + r_n.$$

If $\breve{\lambda}_j \leqslant 0$, then $|r_n - \breve{\lambda}_j| \leqslant r_n + |\breve{\lambda}_j| \leqslant r_n + |\breve{\lambda}_j - \lambda_j| \leqslant r_n + \|\breve{\Theta}_t^N - \Theta\| \leqslant 2r_n$, where λ_j is the j-th positive eigenvalue of Θ. If $\breve{\lambda}_j > 0$, then $|\max(\breve{\lambda}_j, r_n) - \breve{\lambda}_j| \leqslant r_n$. Thus, we have $\|\check{\Theta}_t^N - \Theta\| \leqslant 3r_n$. ■

Lemma 2.2 states that the estimator $\check{\Theta}_t^N = \sum_{j=1}^d \max(\breve{\lambda}_j, r_n) u_j u_j^T$ projects back $\tilde{\Theta}_t^N$ into $\mathbb{S}^{d\times d}$ with the same rate of convergence as $\tilde{\Theta}_t^N$. In view of the equivalent theoretical guarantee, we shall in practice compute the unconstrained version of the MLE $\tilde{\Theta}_t^N$ as our working definition of the MLE $\hat{\Theta}_t^N$. In Section 3, we derive the rate of convergence for $\tilde{\Theta}_t^N$ and we shall use $\hat{\Theta}_t^N$ to mean $\tilde{\Theta}_t^N$.

Next we explain the intuition why the MLE (2.3) works. Note that we can write the log-likelihood as

$$\bar{\ell}_t^N(A) = N \int_0^t \operatorname{tr}\Big[M_s\Big(-\frac{1}{2}AA^T + A\Theta\Big)\Big]\,\mathrm{d}s + \sigma \sum_{i=1}^N \int_0^t A(\overline{X}_s^N - X_s^{N,i}) \cdot \mathrm{d}W_s^i, \tag{2.4}$$

where $M_s = N^{-1}\sum_{i=1}^N (\overline{X}_s^N - X_s^{N,i})(\overline{X}_s^N - X_s^{N,i})^T$ is the instantaneous mean-field covariance matrix at time point s.

As discussed earlier in Section 1, since the interaction among the N particles in the mean-field regime is weak, we expect that those particles can be decoupled by their independent analogs. Let $(Y_t^1)_{t\geqslant 0}, \ldots, (Y_t^N)_{t\geqslant 0}$ be *independent* copies of the OU processes defined in (1.6). Effectively we can view $(Y_t^1)_{t\geqslant 0}, \ldots, (Y_t^N)_{t\geqslant 0}$ as a *decoupled system* of the N-particle system $(X_t^{N,1})_{t\geqslant 0}, \ldots, (X_t^{N,N})_{t\geqslant 0}$. Based on the decoupled processes, we can approximate $\bar{\ell}_t^N(A)$ by

$$\bar{\tilde{\ell}}_t^N(A) := N \int_0^t \operatorname{tr}\Big[\tilde{M}_s\Big(-\frac{1}{2}AA^T + A\Theta\Big)\Big]\,\mathrm{d}s - \sigma \sum_{i=1}^N \int_0^t AY_s^i \cdot \mathrm{d}W_s^i,$$

where $\tilde{M}_s = N^{-1}\sum_{i=1}^N Y_s^i Y_s^{i\,T}$. Decompose

$$\bar{\ell}_t^N(A) = \underbrace{(\bar{\ell}_t^N(A) - \bar{\tilde{\ell}}_t^N(A))}_{\text{decoupling error}} + \underbrace{(\bar{\tilde{\ell}}_t^N(A) - \mathbb{E}[\bar{\tilde{\ell}}_t^N(A)])}_{\text{OU fluctuation error}} + \underbrace{\mathbb{E}[\bar{\tilde{\ell}}_t^N(A)]}_{\text{signal}}.$$

Concentration bounds developed in Section 4 allow us to control the decoupling and OU fluctuation errors around zero. Thus information useful for the estimation purpose comes from the signal part

$$\mathbb{E}[\bar{\tilde{\ell}}_t^N(A)] = \operatorname{tr}\Big[\Big(\int_0^t \mathbb{E}[\tilde{M}_s]\mathrm{d}s\Big)\Big(-\frac{1}{2}AA^T + A\Theta\Big)\Big].$$

Since the matrix $\int_0^t \mathbb{E}[\tilde{M}_s]\mathrm{d}s$ is positive-definite, we see that the maximizer of $\mathbb{E}[\bar{\tilde{\ell}}_t^N(A)]$ is $A^* = \Theta$. This means that on the population level, the MLE equals to the true parameter. Combining this with the decoupling and OU fluctuation errors, we can obtain the rate of convergence for the MLE $\hat{\Theta}_t^N$ in (2.3).

3 Rate of convergence

In this section, we derive the rate of convergence for estimating Θ by the MLE $\hat{\Theta}_t^N$ in (2.3) from the continuous-time and single-trajectory data for each particle. Below is the main result of this paper.

Theorem 3.1. Let $(X_t^{N,1})_{t\geqslant 0},\dots,(X_t^{N,N})_{t\geqslant 0}$ be the d-dimensional N-particle system defined in (1.2) with i.i.d. initialization $X_0^{N,i}\sim N(0,D)$ for $i=1,\dots,N$. Let $\theta_1\geqslant\ldots\geqslant\theta_d>0$ and $\tau_1\geqslant\ldots\geqslant\tau_d\geqslant 0$ be the ordered eigenvalues of Θ and D, respectively. If

$$t\geqslant C\kappa\left(\frac{1}{\theta_d}+\frac{\tau_1}{\sigma^2}\right)\quad\text{and}\quad \kappa d\sqrt{\frac{\log(d/\varepsilon)}{N}}\leqslant c \tag{3.1}$$

for some universal constants C and c where $\kappa=\theta_1/\theta_d$ is the condition number of Θ, then we have with probability at least $1-14\varepsilon$,

$$\big\|\hat{\Theta}_t^N-\Theta\big\|\leqslant Cd\log\left(\frac{d}{\varepsilon}\right)\sqrt{\Big(\frac{\tau_1}{\sigma^2 t}+1\Big)\frac{\kappa\theta_1}{Nt}}. \tag{3.2}$$

Theorem 3.1 is non-asymptotic and has several appealing features.

First, the Gaussian initialization is not essential and it can be relaxed to any i.i.d. initialization with sub-Gaussian distributions with the ψ_2 norm controlled by τ_1. The sub-Gaussian tail is necessary to obtain the exponential concentration rate in N, and the effect of initialization in the approximating OU processes decays exponentially fast in t (cf. (4.2) in Section 4.1). Error bound in (3.2) also reflects the diminishing effect of initialization $\tau_1/(\sigma^2 t)\to 0$ as $t\to\infty$.

Second, the conditions in (3.1) are mild. We do not assume that the N-particle processes $(X_t^{N,1})_{t\geqslant 0},\dots,(X_t^{N,N})_{t\geqslant 0}$ starts from the stationary distribution (which is a restrictive assumption), and the (continuous) time complexity of the particle trajectories is sharp in the following sense. Suppose $\Theta=\mathrm{diag}(\theta,\dots,\theta)$ is an isotropic interaction matrix (i.e., $\kappa=1$) and $X_0^{N,1}=\cdots=X_0^{N,N}=0$ (i.e., $\tau_1=0$). Observe that the decoupled N copies of the OU processes to approximate the dynamics of the interacting N-particle system have the equilibrium distribution as $N(0,\sigma^2(2\theta)^{-1}I_d)$, and in view of (4.2), it takes at least $\Omega(\theta^{-1})$ time for $(X_t^{N,1})_{t\geqslant 0},\dots,(X_t^{N,N})_{t\geqslant 0}$ mixing to the steady states (modulo small decoupling errors in t). In particular, if θ is closer to zero, then the log-likelihood ratio becomes flatter and thus larger t is necessary to see the information from samples of the stationary distribution. On the other hand, if the processes start from the stationary distribution, then this trajectory time lower bound in t is not needed to obtain (3.2), provided that $d\sqrt{N^{-1}\log(d/\varepsilon)}\leqslant c$ which is the case whenever ϵ is not too small, e.g., $\varepsilon\geqslant d\exp(-c^2N/d^2)$.

Third, it is known that the MLE $\tilde{\theta}_t$ of a one-dimensional OU process $\mathrm{d}Y_t=-\theta Y_t\mathrm{d}t+\sigma\mathrm{d}W_t$ for the parameter $\theta>0$ has the exact rate of convergence in the sense that $\sqrt{t}(\tilde{\theta}_t-\theta)$ converges in distribution to $N(0,2\theta)$ as $t\to\infty$ (cf. Example 1.35 in [9]). Contrast this with our result, we see the rate of convergence in (3.2) is rate-optimal in both t and N, as well as θ_1. Specifically, for fixed d, we can obtain from Theorem 3.1 the large N (mean-field limit) and large t (long-time dynamics) asymptotics as:

$$\|\sqrt{t}(\hat{\Theta}_t^N-\Theta)\|=O_P\left(\theta_1/\sqrt{N}\right), \tag{3.3}$$

provided that the N particle processes start from a chaotic distribution, which agrees with the exact asymptotics of the one-dimensional OU process (or more generally, d-dimensional OU processes with isotropic Θ). In view of the architecture of approximating the N-particle system by N independent OU processes, the rate we derived in Theorem 3.1 is rate-optimal in t and N, modulo small decoupling errors. We shall highlight

that long-time dynamic behavior in (3.3) as $t \to \infty$ cannot be obtained from the classical theory of the propagation of chaos (cf. [18]), where Gronwall's lemma (cf. Appendix 1 in [16]) is typically used to control the decoupling error between the interacting N-particle processes and their independent analogs. In such case, the decoupling error is exponentially increasing in t, and thus it cannot be used to yield the rate $t^{-1/2}$. Our argument is tailored to the quadratic structure of the interacting potential $V(x) = \frac{1}{2}x^T\Theta x$, which allows for a far more efficient decoupling strategy (cf. Lemma 4.2).

Proof of Theorem 3.1. Since the objective function $\bar{\ell}_t^N(A)$ in (2.4) is quadratic in A, the first-order optimality condition for the unconstrained optimization problem implies that the MLE of Θ satisfies

$$\hat{\Theta}_t^N = \Theta + \Big(\int_0^t M_s \mathrm{d}s\Big)^{-1}\Big(\int_0^t \frac{1}{N}\sum_{i=1}^N \mathrm{d}W_s^i \otimes (\overline{X}_s^N - X_s^{N,i})\Big)\,\sigma, \tag{3.4}$$

where we recall $M_s = N^{-1}\sum_{i=1}^N (\overline{X}_s^N - X_s^{N,i})(\overline{X}_s^N - X_s^{N,i})^T$ and $a \otimes b$ denotes the tensor product of two vectors a and b, i.e., $(a \otimes b)_{jk} = a_j b_k$. Let $(Y_t^i)_{t\geqslant 0}, i \in [N]$ be independent copies of the OU process driven by the same Brownian motion $(W_t^i)_{t\geqslant 0}$ in $(X_t^{N,i})_{t\geqslant 0}$, namely,

$$\mathrm{d}Y_t^i = -\Theta Y_t^i \mathrm{d}t + \sigma \mathrm{d}W_t^i \quad \text{with} \quad Y_0^i = X_0^{N,i}.$$

First, by triangle inequality,

$$\begin{aligned}\Big\|\frac{1}{N}\sum_{i=1}^N \int_0^t \mathrm{d}W_s^i \otimes (\overline{X}_s^N - X_s^{N,i})\Big\| \leqslant &\Big\|\frac{1}{N}\sum_{i=1}^N \int_0^t \mathrm{d}W_s^i \otimes (\overline{X}_s^N - X_s^{N,i} + Y_s^i)\Big\| \\ &+ \Big\|\frac{1}{N}\sum_{i=1}^N \int_0^t \mathrm{d}W_s^i \otimes Y_s^i\Big\|\end{aligned}$$

Let $g(\varepsilon, N, d) = 1 + \sqrt{N^{-1}\log(d/\varepsilon)}$. Clearly, $g(\varepsilon, N, d) \leqslant g(\varepsilon, 1, d)$. By (4.4) in Lemma 4.1 and (4.7) in Lemma 4.2, there exists a universal constant $C > 0$ such that for any $\varepsilon \in (0,1)$,

$$\begin{aligned}\Big\|\frac{1}{N}\sum_{i=1}^N \int_0^t \mathrm{d}W_s^i \otimes (\overline{X}_s^N - X_s^{N,i})\Big\| &\leqslant Cdg(\varepsilon,1,1)\sqrt{\frac{(\tau_1 + \sigma^2 t)\log(d/\varepsilon)}{N\theta_d}} \\ &\leqslant Cd\log(d/\varepsilon)\sqrt{\frac{\tau_1 + \sigma^2 t}{N\theta_d}}\end{aligned} \tag{3.5}$$

holds with probability at least $1 - 8\varepsilon$. Next, since

$$\Big\|\Big(\int_0^t M_s \mathrm{d}s\Big)^{-1}\Big\| = \frac{1}{\lambda_{\min}\Big(\int_0^t M_s \mathrm{d}s\Big)} = \frac{1}{\lambda_d\Big(\int_0^t M_s \mathrm{d}s\Big)},$$

where $\lambda_j(M)$ is the j-th eigenvalue of a symmetric positive semidefinite matrix M and $\lambda_{\min}(M)$ is the smallest eigenvalue of M, we shall derive a lower bound for $\lambda_{\min}\Big(\int_0^t M_s \mathrm{d}s\Big)$. Note that

$$\begin{aligned}\int_0^t M_s \mathrm{d}s = &\int_0^t \mathbb{E}[Y_s^i Y_s^{iT}]\mathrm{d}s + \frac{1}{N}\sum_{i=1}^N \int_0^t (Y_s^i Y_s^{iT} - \mathbb{E}\, Y_s^i Y_s^{iT})\mathrm{d}s \\ &+ \frac{1}{N}\sum_{i=1}^N \int_0^t [(\overline{X}_s^N - X_s^{N,i})(\overline{X}_s^N - X_s^{N,i})^T - Y_s^i Y_s^{iT}]\mathrm{d}s.\end{aligned}$$

By Weyl's inequality, we have for all $j \in [d]$,

$$\Big|\lambda_j\Big(\int_0^t M_s \mathrm{d}s\Big) - \lambda_j\Big(\int_0^t \mathbb{E}[Y_s^i Y_s^{iT}]\mathrm{d}s\Big)\Big| \leqslant \Big\|\frac{1}{N}\sum_{i=1}^N \int_0^t (Y_s^i Y_s^{iT} - \mathbb{E}\, Y_s^i Y_s^{iT})\mathrm{d}s\Big\| + \Big\|\frac{1}{N}\sum_{i=1}^N \int_0^t [(\overline{X}_s^N - X_s^{N,i})(\overline{X}_s^N - X_s^{N,i})^T - Y_s^i Y_s^{iT}]\mathrm{d}s\Big\|.$$

In addition, since $\mathbb{E}[Y_s^i Y_s^{iT}] = \mathrm{Cov}(Y_s^i) = e^{-\Theta t} D e^{-\Theta t} + \sigma^2 (2\Theta)^{-1}(I_d - e^{-2\Theta t})$, by a second application of Weyl's inequality and Jensen's inequality, we have for all $j \in [d]$,

$$\Big|\lambda_j\Big(\int_0^t \mathbb{E}[Y_s^i Y_s^{iT}]\mathrm{d}s\Big) - \lambda_j\big(t\sigma^2(2\Theta)^{-1}\big)\Big| \leqslant \Big\|\int_0^t e^{-\Theta s} D e^{-\Theta s}\mathrm{d}s - \int_0^t \sigma^2 (2\Theta)^{-1} e^{-2\Theta s}\mathrm{d}s\Big\| \leqslant \int_0^t \big[\tau_1 e^{-2s\theta_d} + \sigma^2(2\theta_d)^{-1} e^{-2s\theta_d}\big]\mathrm{d}s \leqslant \frac{\tau_1}{2\theta_d} + \frac{\sigma^2}{4\theta_d^2}.$$

Combining the last two inequalities and using (4.3) in Lemma 4.1 and (4.8) in Lemma 4.2, we deduce that there is a universal constant c such that with probability at least $1 - 6\varepsilon$,

$$\lambda_{\min}\Big(\int_0^t M_s \mathrm{d}s\Big) \geqslant \frac{\sigma^2 t}{2\theta_1} - \frac{\tau_1}{2\theta_d} - \frac{\sigma^2}{4\theta_d^2} - Cd\frac{\tau_1 + \sigma^2 t}{2\theta_d}\sqrt{\frac{\log(d/\varepsilon)}{N}} \geqslant c\frac{\sigma^2 t}{\theta_1}, \tag{3.6}$$

where the last inequality follows from the conditions in (3.1). Now, putting together (3.4), (3.5) and (3.6), we conclude that there exists a universal constant C such that (3.2) holds with probability at least $1 - 14\varepsilon$. ■

4 Technical lemmas

This section provides key technical results for bounding the fluctuation of the OU process and the decoupling error of the N-particle system by the associated N independent d-dimensional Ornstein-Uhlenbeck processes.

4.1 Concentration inequalities for Ornstein-Uhlenbeck processes

The d-dimensional Ornstein-Uhlenbeck (OU) process $(Y_t)_{t\geqslant 0}$ with an isotropic diffusion parameter is a mean-reverting stochastic process defined by the following stochastic differential equation:

$$\mathrm{d}Y_t = -\Theta Y_t \mathrm{d}t + \sigma \mathrm{d}W_t, \tag{4.1}$$

where $(W_t)_{t\geqslant 0}$ is the standard Brownian motion in $\mathbb{R}^d$ such that $W_0 = 0$, $\Theta \in \mathbb{S}_+^{d\times d}$ is the mean-reverting drift parameter, and $\sigma > 0$ is the scalar diffusion parameter. The solution to (4.1) is given by

$$Y_t = e^{-\Theta t} Y_0 + \sigma \int_0^t e^{\Theta(s-t)}\, \mathrm{d}W_s, \tag{4.2}$$

where $e^{\Theta} = \sum_{k=0}^{\infty} \Theta^k/k!$ is the matrix exponential function. From (4.2), we see that $Y_t|Y_0 \sim N(e^{-\Theta t}Y_0, \sigma^2 \int_0^t e^{2\Theta(s-t)}\mathrm{d}s)$. Suppose the initialization Y_0 has mean zero and covariance matrix D, and Y_0 is independent of $(W_t)_{t\geqslant 0}$. Then we have $\mathbb{E}[Y_t] = 0$ and $\Sigma_t := \mathrm{Cov}(Y_t) = e^{-\Theta t} D e^{-\Theta t} + \sigma^2(2\Theta)^{-1}(I_d - e^{-2\Theta t})$. As $t \to \infty$, the stationary distribution Y_t is $N(0, \sigma^2(2\Theta)^{-1})$.

Denote $g(\varepsilon, N, d) = 1 + \sqrt{N^{-1}\log(d/\varepsilon)}$. Let $\theta_1 \geqslant \ldots \geqslant \theta_d > 0$ and $\tau_1 \geqslant \ldots \geqslant \tau_d \geqslant 0$ be the ordered eigenvalues of Θ and D, respectively.

Lemma 4.1 (Concentration inequalities for the OU process). Let $\mathrm{d}Y_t = -\Theta Y_t \mathrm{d}t + \sigma \mathrm{d}W_t$ be the d-dimensional OU process defined in (4.1) with $Y_0 \sim N(0, D)$ being independent of $(W_t)_{t\geqslant 0}$. Suppose that $(Y_t^1)_{t\geqslant 0}, \dots, (Y_t^N)_{t\geqslant 0}$ are independent copies of $(Y_t)_{t\geqslant 0}$. Then there exists a universal constant $C > 0$ such that for any $\varepsilon \in (0,1)$,

$$\mathbb{P}\Big(\Big\|\frac{1}{N}\sum_{i=1}^N \int_0^t (Y_s^i Y_s^{i\,T} - \mathbb{E}\, Y_s^i Y_s^{i\,T})\mathrm{d}s\Big\| \geqslant Cd\frac{\tau_1+\sigma^2 t}{\theta_d}\Big(\frac{\log(d/\varepsilon)}{N} + \sqrt{\frac{\log(d/\varepsilon)}{N}}\Big)\Big) \leqslant 2\varepsilon, \tag{4.3}$$

$$\mathbb{P}\Big(\Big\|\frac{1}{N}\sum_{i=1}^N \int_0^t \mathrm{d}W_s^i \otimes Y_s^i\Big\| \geqslant Cdg(\varepsilon,N,1)\sqrt{\frac{(\tau_1+\sigma^2 t)\log(d/\varepsilon)}{N\theta_d}}\Big) \leqslant 4\varepsilon. \tag{4.4}$$

Proof of Lemma 4.1. Denote $U_s^i = Y_s^i Y_s^{i\,T} - \mathbb{E}\, Y_s^i Y_s^{i\,T}$ and $U_s^i = (U_{s,jk}^i)_{j,k\in[d]}$. Let $\lambda > 0$ and fix a $j,k \in [d]$. By the exponential Markov inequality and the independence of the processes $(Y_t^i)_{t\geqslant 0}$, we have for all $x > 0$,

$$\mathbb{P}\Big(\frac{1}{N}\sum_{i=1}^N \int_0^t U_{s,jk}^i \mathrm{d}s \geqslant x\Big) \leqslant \exp(-\lambda x)\prod_{i=1}^N \mathbb{E}\Big[\exp\Big(\frac{\lambda}{N}\int_0^t U_{s,jk}^i \mathrm{d}s\Big)\Big]. \tag{4.5}$$

Since $Y_0^i \sim N(0,D)$, we have $Y_t \sim N(0, e^{-\Theta t}De^{-\Theta t} + \sigma^2(2\Theta)^{-1}(I_d - e^{-2\Theta t}))$ for $t \geqslant 0$. Thus, for all $0 \leqslant s \leqslant t$,

$$\mathrm{Var}(Y_{sj}^i) \leqslant \|\mathrm{Cov}(Y_s^i)\| \leqslant \|e^{-\Theta s}\|^2\,\|D\| + \sigma^2\,\|(2\Theta)^{-1}\|\,\|I_d - e^{-2\Theta s}\| \leqslant \tau_1 e^{-2s\theta_d} + \sigma^2(2\theta_d)^{-1},$$

and $\|Y_{sj}^i\|_{\psi_2}^2 \leqslant C\,\mathrm{Var}(Y_{sj}^i) \leqslant C\,(\tau_1 e^{-2s\theta_d} + \sigma^2(2\theta_d)^{-1})$ for some universal constant C. By Lemma 2.7.7 in [20], $\|Y_{sj}^i Y_{sk}^i\|_{\psi_1} \leqslant \|Y_{sj}^i\|_{\psi_2}\|Y_{sk}^i\|_{\psi_2} \leqslant C\,(\tau_1 e^{-2s\theta_d} + \sigma^2(2\theta_d)^{-1})$. This together with Jensen's inequality imply that

$$\Big\|\int_0^t U_{s,jk}^i \mathrm{d}s\Big\|_{\psi_1} \leqslant \int_0^t \big\|U_{s,jk}^i\big\|_{\psi_1}\mathrm{d}s \leqslant C\,\theta_d^{-1}(\tau_1+\sigma^2 t).$$

Using the moment-generating function property of sub-exponential random variables (cf. Proposition 2.7.1 in [20]), we have

$$\Big|\frac{\lambda}{N}\Big| \leqslant \frac{c}{\max_i \|\int_0^t U_{s,jk}^i \mathrm{d}s\|_{\psi_1}} \implies \mathbb{E}\Big[\exp\Big(\frac{\lambda}{N}\int_0^t U_{s,jk}^i \mathrm{d}s\Big)\Big] \leqslant \exp\Big(C\frac{\lambda^2}{N^2}\Big\|\int_0^t U_{s,jk}^i \mathrm{d}s\Big\|_{\psi_1}^2\Big).$$

Thus, for $0 < \lambda < cN\theta_d/(\tau_1+\sigma^2 t)$, we can bound

$$\mathbb{E}\Big[\exp\Big(\frac{\lambda}{N}\int_0^t U_{s,jk}^i \mathrm{d}s\Big)\Big] \leqslant \exp\Big(C\frac{\lambda^2}{N^2}\sum_{i=1}^N\Big\|\int_0^t U_{s,jk}^i \mathrm{d}s\Big\|_{\psi_1}^2\Big) \leqslant \exp\Big(C\frac{\lambda^2(\tau_1+\sigma^2 t)^2}{N\theta_d^2}\Big).$$

Combining the last inequality with (4.5) and optimizing the bound over λ, we have

$$\mathbb{P}\Big(\frac{1}{N}\sum_{i=1}^N \int_0^t U_{s,jk}^i \mathrm{d}s \geqslant x\Big) \leqslant \exp\Big\{-\min\Big[\frac{N\theta_d^2 x^2}{4C(\tau_1+\sigma^2 t)^2},\ \frac{cN\theta_d x}{2(\tau_1+\sigma^2 t)}\Big]\Big\}.$$

Equivalently, we can write the above inequality as: for all $\varepsilon \in (0,1)$,

$$\mathbb{P}\Big(\frac{1}{N}\sum_{i=1}^N \int_0^t U_{s,jk}^i \mathrm{d}s \geqslant C\frac{\tau_1+\sigma^2 t}{\theta_d}\Big(\frac{\log(1/\varepsilon)}{N} + \sqrt{\frac{\log(1/\varepsilon)}{N}}\Big)\Big) \leqslant \varepsilon. \tag{4.6}$$

Applying the same argument to $-U_{s,jk}^i$, together with the union bound over $j,k \in [d]$ and using the fact that $\|M\| \leqslant d\max_{j,k\in[d]}|M_{jk}|$ for any $d\times d$ matrix M, we obtain (4.3).

Next we prove (4.4). Denote $Z_t^i = \int_0^t \mathrm{d}W_s^i \otimes Y_s^i$. Clearly $\mathbb{E}[Z_t^i] = 0_{d\times d}$. Moreover, $(Z_t^i)_{t\geqslant 0}, \dots, (Z_t^N)_{t\geqslant 0}$ are independent continuous local martingales vanishing at zero with quadratic variation $[Z^i]_t := ([Z_{jk}^i]_t)_{j,k\in[d]}$ is given by $[Z_{jk}^i]_t = \int_0^t |Y_{sk}^i|^2 \mathrm{d}s$. Moreover, the process $\overline{Z}_s^N = \frac{1}{N}\sum_{i=1}^N Z_s^i$ is also a local martingale vanishing at zero with quadratic variation $[\overline{Z}_{jk}^N]_t = \frac{1}{N^2}\sum_{i=1}^N \int_0^t |Y_{sk}^i|^2 \mathrm{d}s$. Note that $\mathbb{E}[Z_{jk}^i]_t \leqslant \int_0^t \|\Sigma_s\| \mathrm{d}s \leqslant (2\theta_d)^{-1}(\tau_1 + \sigma^2 t)$ and $\mathbb{E}[\overline{Z}_{jk}^N]_t \leqslant (2N\theta_d)^{-1}(\tau_1 + \sigma^2 t)$. Applying (4.6) on the diagonal entries $(U_{s,kk}^i)_{k\in[d]}$, we have $\mathbb{P}([\overline{Z}_{jk}^N]_t \geqslant Ctg(\varepsilon, N, 1)^2(\tau_1 + \sigma^2 t)(N\theta_d)^{-1}) \leqslant \varepsilon$. By Bernstein's inequality for continuous local martingales (cf. Exercise (3.16) on page 145 in [16]), we get

$$\begin{aligned}
&\mathbb{P}\Big(\sup_{0\leqslant s\leqslant t} \frac{1}{N}\sum_{i=1}^N Z_{s,jk}^i \geqslant Cxg(\varepsilon, N, 1)\sqrt{\frac{\tau_1 + \sigma^2 t}{N\theta_d}} \Big) \\
\leqslant &\mathbb{P}\Big(\sup_{0\leqslant s\leqslant t} \overline{Z}_{s,jk}^N \geqslant Cxg(\varepsilon, N, 1)\sqrt{\frac{\tau_1 + \sigma^2 t}{N\theta_d}},\ [\overline{Z}^N]_t \leqslant Cg(\varepsilon, N, 1)^2 \frac{\tau_1 + \sigma^2 t}{N\theta_d} \Big) + \varepsilon \\
\leqslant &\exp\Big(-\frac{Cx^2}{2} \Big) + \varepsilon.
\end{aligned}$$

Choosing $x = \sqrt{2\log(d^2/\varepsilon)}$ and applying the union bound, we have with probability at least $1 - 2\varepsilon$,

$$\max_{j,k\in[d]} \sup_{0\leqslant s\leqslant t} \frac{1}{N}\sum_{i=1}^N Z_{s,jk}^i \leqslant Cg(\varepsilon, N, 1)\sqrt{\frac{(\tau_1 + \sigma^2 t)\log(d/\varepsilon)}{N\theta_d}}$$

Applying the same argument to $-Z_s^i$ and using $\|M\| \leqslant d\max_{j,k\in[d]} |M_{jk}|$ for any $d \times d$ matrix M, we obtain (4.4). ■

4.2 Decoupling error bounds for N-particle systems

Lemma 4.2 (Decoupling error bounds for the N-particle system). Let $(X_t^{N,1})_{t\geqslant 0}, \dots, (X_t^{N,N})_{t\geqslant 0}$ be the d-dimensional N-particle system defined in (1.2) with i.i.d. initialization $X_0^{N,i} \sim N(0, D)$ for $i = 1, \dots, N$. Then there exists a universal constant $C > 0$ such that for any $\varepsilon \in (0, 1)$,

$$\mathbb{P}\Big(\Big\| \frac{1}{N}\sum_{i=1}^N \int_0^t \mathrm{d}W_s^i \otimes (\overline{X}_s^N - X_s^{N,i} + Y_s^i) \Big\| \geqslant Cdg(\varepsilon, 1, 1)\sqrt{\frac{(\tau_1 + \sigma^2 t)\log(d/\varepsilon)}{N\theta_d}} \Big) \leqslant 4\varepsilon, \tag{4.7}$$

$$\begin{aligned}
&\mathbb{P}\Big(\Big\| \frac{1}{N}\sum_{i=1}^N \int_0^t [(\overline{X}_s^N - X_s^{N,i})(\overline{X}_s^N - X_s^{N,i})^T - Y_s^i {Y_s^i}^T] \mathrm{d}s \Big\| \\
&\qquad\qquad \leqslant Cd\frac{g(\varepsilon, 1, d)}{\sqrt{N}} \Big[\frac{g(\varepsilon, 1, d)}{\sqrt{N}} + g(\varepsilon, N, d) \Big] \frac{\tau_1 + \sigma^2 t}{2\theta_d} \Big) \leqslant 4\varepsilon.
\end{aligned} \tag{4.8}$$

Proof of Lemma 4.2. Recall definitions of the N-particle system and the associated approximating N independent OU processes: for $i \in [N]$,

$$\mathrm{d}X_t^{N,i} = \Theta(\overline{X}_t^N - X_t^{N,i})\mathrm{d}t + \sigma \mathrm{d}W_t^i \quad \text{and} \quad \mathrm{d}Y_t^i = -\Theta Y_t^i \mathrm{d}t + \sigma \mathrm{d}W_t^i,$$

where $(X_t^{N,i})$ and (Y_t^i) are driven by the same Brownian motion (W_t^i) and $Y_0^i = X_0^{N,i}$. Averaging the processes $(X_t^{N,i})$ over $i \in [N]$, we get $\mathrm{d}\overline{X}_t^N = \sigma \mathrm{d}\overline{W}_t^N := \sigma N^{-1}\sum_{i=1}^N \mathrm{d}W_t^i$, i.e., the mean process $(\sqrt{N}\overline{X}_t^N)$ of the N particles has the same law as a standard Brownian motion rescaled by σ. Combining the last three expressions, we obtain that

$$\mathrm{d}(\overline{X}_t^N - X_t^{N,i} + Y_t^i) = \sigma \mathrm{d}\overline{W}_t^N - \Theta(\overline{X}_t^N - X_t^{N,i} + Y_t^i),$$

which implies that the difference process $\Delta_t^{N,i} = \overline{X}_t^N - X_t^{N,i} + Y_t^i$ between the N-particle system and the decoupled OU processes is an OU process with respect to $(\overline{W}_t^N)$, i.e., we have for all $i \in [N]$,

$$\mathrm{d}\Delta_t^{N,i} = -\Theta\Delta_t^{N,i}\mathrm{d}t + \sigma\mathrm{d}\overline{W}_t^N \quad \text{with} \quad \Delta_0^{N,i} = \overline{Y}_0^N \sim N(0, N^{-1}D). \tag{4.9}$$

Equation (4.9) means that the processes $(\Delta_t^{N,1}), \dots, (\Delta_t^{N,N})$ are the same mean-reverting OU processes, all driven by $(\overline{W}_t^N)$. Thus $\overline{\Delta}_t^N = \Delta_t^{N,i}$ for all $i \in [N]$, and $N^{-1}\sum_{i=1}^N \int_0^t \mathrm{d}W_s^i \otimes \Delta_s^{N,i} = \int_0^t \mathrm{d}\overline{W}_s \otimes \overline{\Delta}_t^N$. Now applying (4.4) in Lemma 4.1 to the averaged difference process $(\overline{\Delta}_t^N)$ with diffusion parameter $N^{-1/2}\sigma$ and initialization variance $N^{-1}D$, we obtain (4.7).

Next we prove (4.8). By the triangle inequality and the Cauchy-Schwarz inequality,

$$\begin{aligned}
&\Big\|\frac{1}{N}\sum_{i=1}^N \int_0^t [(\overline{X}_s^N - X_s^{N,i})(\overline{X}_s^N - X_s^{N,i})^T - Y_s^i Y_s^{i\,T}]\mathrm{d}s\Big\| \\
\leqslant &\Big\|\frac{1}{N}\sum_{i=1}^N \int_0^t \Delta_s^{N,i}\Delta_s^{N,i\,T}\mathrm{d}s\Big\| + \Big\|\frac{1}{N}\sum_{i=1}^N \int_0^t [\Delta_s^{N,i} Y_s^{i\,T} + Y_s^i \Delta_s^{N,i\,T}]\mathrm{d}s\Big\| \\
\leqslant &\Big\|\int_0^t \overline{\Delta}_s^N \overline{\Delta}_s^{N\,T}\mathrm{d}s\Big\| + 2\Big(\int_0^t \|\overline{\Delta}_s^N\|^2\mathrm{d}s\Big)^{\frac{1}{2}}\Big(\int_0^t \frac{1}{N}\sum_{i=1}^N \|Y_s^i\|^2\mathrm{d}s\Big)^{\frac{1}{2}}.
\end{aligned}$$

Since $(\overline{\Delta}_t^N)$ is an OU process, we have $\|\int_0^t \mathbb{E}[\overline{\Delta}_s^N \overline{\Delta}_s^{N\,T}]\mathrm{d}s\| \leqslant N^{-1}\int_0^t \|\Sigma_s\|\mathrm{d}s \leqslant (\tau_1 + \sigma^2 t)(2N\theta_d)^{-1}$, and by (4.3),

$$\mathbb{P}\Big(\Big\|\int_0^t [\overline{\Delta}_s^N \overline{\Delta}_s^{N\,T} - \mathbb{E}\,\overline{\Delta}_s^N \overline{\Delta}_s^{N\,T}]\mathrm{d}s\Big\| \geqslant Cd\frac{\tau_1 + \sigma^2 t}{N\theta_d}(\log(d/\varepsilon) + \sqrt{\log(d/\varepsilon)})\Big) \leqslant 2\varepsilon.$$

So we have with probability at least $1 - 2\varepsilon$,

$$\Big\|\int_0^t \overline{\Delta}_s^N \overline{\Delta}_s^{N\,T}\mathrm{d}s\Big\| \leqslant Cdg(\varepsilon, 1, d)^2\frac{\tau_1 + \sigma^2 t}{N\theta_d}.$$

Next, applying (4.6) to the process $(\overline{\Delta}_t^N)$ and the union bound, we have

$$\mathbb{P}\Big(\max_{j\in[d]}\int_0^t (|\overline{\Delta}_{sj}^N|^2 - \mathbb{E}\,|\overline{\Delta}_{sj}^N|^2)\mathrm{d}s \geqslant C\frac{\tau_1 + \sigma^2 t}{N\theta_d}(\log(d/\varepsilon) + \sqrt{\log(d/\varepsilon)})\Big) \leqslant \varepsilon.$$

This gives

$$\int_0^t \|\overline{\Delta}_s^N\|^2\mathrm{d}s \leqslant d\max_{j\in[d]}\int_0^t |\overline{\Delta}_{sj}^N|^2\mathrm{d}s \leqslant Cdg(\varepsilon, 1, d)^2\frac{\tau_1 + \sigma^2 t}{N\theta_d},$$

where the last inequality holds with probability at least $1 - \varepsilon$. Similar argument yields

$$\mathbb{P}\Big(\frac{1}{N}\sum_{i=1}^N \int_0^t \|Y_s^i\|^2\mathrm{d}s \geqslant Cdg(\varepsilon, N, d)^2\frac{\tau_1 + \sigma^2 t}{\theta_d}\Big) \leqslant \varepsilon.$$

Then (4.8) follows from putting all pieces together. ■

References

[1] Mattia Bongini, Massimo Fornasier, Markus Hansen, and Mauro Maggioni, *Inferring interaction rules from observations of evolutive systems i: The variational approach*, Mathematical Models and Methods in Applied Sciences **27** (2017), no. 5, 909–951. MR-3636616

[2] Steven L. Brunton, Joshua L. Proctor, and J. Nathan Kutz, *Discovering governing equations from data by sparse identification of nonlinear dynamical systems*, Proceedings of the National Academy of Sciences **113** (2016), no. 15, 3932–3937. MR-3494081

[3] Peter Bühlmann and Sara van de Geer, *Statistics for high-dimensional data*, Springer Series in Statistics, Springer, Heidelberg, 2011, Methods, theory and applications. MR-2807761

[4] Xiaohui Chen, Mengyu Xu, and Wei Biao Wu, *Covariance and precision matrix estimation for high-dimensional time series*, The Annals of Statistics **41** (2013), no. 6, 2994–3021. MR-3161455

[5] M. R. D'Orsogna, Y. L. Chuang, A. L. Bertozzi, and L. S. Chayes, *Self-propelled particles with soft-core interactions: Patterns, stability, and collapse*, Phys. Rev. Lett. **96** (2006), 104302. MR-2369988

[6] Eran Eyal, Lee-Wei Yang, and Ivet Bahar, *Anisotropic network model: systematic evaluation and a new web interface*, Bioinformatics **22** (2006), no. 21, 2619–2627.

[7] Turkan Haliloglu, Ivet Bahar, and Burak Erman, *Gaussian Dynamics of Folded Proteins*, Phys. Rev. Lett. **79** (1997), no. 16, 3090–3093.

[8] Hui Huang, Jian-Guo Liu, and Jianfeng Lu, *Learning interacting particle systems: Diffusion parameter estimation for aggregation equations*, Mathematical Models and Methods in Applied Sciences **29** (2019), no. 1, 1–29. MR-3904675

[9] Yury A. Kutoyants, *Statistical inference for ergodie diffusion processes*, Springer-Verlag London Ltd., 2004. MR-2144185

[10] Zhongyang Li, Fei Lu, Mauro Maggioni, Sui Tang, and Cheng Zhang, *On the identifiability of interaction functions in systems of interacting particles*, arXiv:1912.11965 (2019). MR-4175473

[11] Fei Lu, Ming Zhong, Sui Tang, and Mauro Maggioni, *Nonparametric inference of interaction laws in systems of agents from trajectory data*, Proceedings of the National Academy of Sciences **116** (2019), no. 29, 14424–14433. MR-3984488

[12] H. P. McKean, *A class of markov processes associated with nonlinear parabolic equations*, Proceedings of the National Academy of Sciences **56** (1966), no. 6, 1907–1911. MR-0221595

[13] Song Mei, Andrea Montanari, and Phan-Minh Nguyen, *A mean field view of the landscape of two-layer neural networks*, Proceedings of the National Academy of Sciences **115** (2018), no. 33, E7665–E7671. MR-3845070

[14] Alexander Mogilner and Leah Edelstein-Keshet, *A non-local model for a swarm*, Journal of Mathematical Biology **38** (1999), no. 6, 534–570. MR-1698215

[15] Sebastien Motsch and Eitan Tadmor, *Heterophilious dynamics enhances consensus*, SIAM Review **56** (2014), no. 4, 577–621. MR-3274797

[16] Daniel Revuz and Marc Yor, *Continuous martingales and brownian motion*, Springer-Verlag, 1991. MR-1083357

[17] Hayden Schaeffer, Giang Tran, and Rachel Ward, *Extracting sparse high-dimensional dynamics from limited data*, SIAM Journal on Applied Mathematics **78** (2018), no. 6, 3279–3295. MR-3885764

[18] Alain-Sol Sznitman, *Topics in propagation of chaos*, Ecole d'Eté de Probabilités de Saint-Flour XIX — 1989 (Berlin, Heidelberg) (Paul-Louis Hennequin, ed.), Springer Berlin Heidelberg, 1991, pp. 165–251. MR-1108185

[19] Chad M. Topaz, Andrea L. Bertozzi, and Mark A. Lewis, *A nonlocal continuum model for biological aggregation*, Bulletin of Mathematical Biology **68** (2006), no. 7, 1601. MR-2257718

[20] Roman Vershynin, *High-Dimensional Probability: An Introduction with Applications in Data Science*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, 2018. MR-3837109

[21] Jianghui Wen, Xiangjun Wang, Shuhua Mao, and Xinping Xiao, *Maximum likelihood estimation of mckean-vlasov stochastic differential equation and its application*, Applied Mathematics and Computation **274** (2016), 237–246. MR-3433132

Acknowledgments. Part of this research was carried out in the Institute for Data, System, and Society (IDSS) at Massachusetts Institute of Technology. The author would like to thank Philippe Rigollet (MIT) and Yun Yang (UIUC) for helpful comments.

Electronic Journal of Probability
Electronic Communications in Probability

Advantages of publishing in EJP-ECP

- Very high standards
- Free for authors, free for readers
- Quick publication (no backlog)
- Secure publication (LOCKSS[1])
- Easy interface (EJMS[2])

Economical model of EJP-ECP

- Non profit, sponsored by IMS[3], BS[4] , ProjectEuclid[5]
- Purely electronic

Help keep the journal free and vigorous

- Donate to the IMS open access fund[6] (click here to donate!)
- Submit your best articles to EJP-ECP
- Choose EJP-ECP over for-profit journals

[1]LOCKSS: Lots of Copies Keep Stuff Safe `http://www.lockss.org/`
[2]EJMS: Electronic Journal Management System `http://www.vtex.lt/en/ejms.html`
[3]IMS: Institute of Mathematical Statistics `http://www.imstat.org/`
[4]BS: Bernoulli Society `http://www.bernoulli-society.org/`
[5]Project Euclid: `https://projecteuclid.org/`
[6]IMS Open Access Fund: `http://www.imstat.org/publications/open.htm`