Eigenfunction martingale estimators for interacting particle systems and their mean field limit

Grigorios A. Pavliotis * and Andrea Zanoni †

Abstract. We study the problem of parameter estimation for large exchangeable interacting particle systems when a sample of discrete observations from a single particle is known. We propose a novel method based on martingale estimating functions constructed by employing the eigenvalues and eigenfunctions of the generator of the mean field limit, linearized around the (unique) invariant measure of the mean field dynamics. We then prove that our estimator is asymptotically unbiased and asymptotically normal when the number of observations and the number of particles tend to infinity, and we provide a rate of convergence towards the exact value of the parameters. Finally, we present several numerical experiments which show the accuracy of our estimator and corroborate our theoretical findings, even in the case the mean field dynamics exhibit more than one steady states.

Key words. Interacting particle systems, exchangeability, mean field limit, inference, Fokker–Planck operator, eigenvalue problem, martingale estimators.

AMS subject classifications. 35Q70, 35Q83, 60J60, 62M15, 65C30.

1. Introduction.

Interacting particle systems and, more generally interacting multiagent models, appear frequently in the natural and social sciences. In addition to the well known applications, e.g., plasma physics [22] and stellar dynamics [7], new applications include, e.g., the modeling of chemotaxis [40], pedestrian dynamics [30, 24], crowd dynamics [32], urban modeling [14], models for opinion formation [18, 21], collective behavior [11], and models for systemic risk [20]. In many of these applications, the phenomenological models involve unknown parameters that need to be estimated from data. This is particularly the case for multiagent models used in the social sciences and in economics, where no physics-informed choices of parameters are available. Learning parameters or even models, in a nonparametric setting, from data is becoming an increasingly important aspect of the overall mathematical modeling strategy. This is particularly the case in view of the huge quantity of available data in different areas, which allows the development of accurate data-driven techniques for learning parameters from data.

In this paper we study the problem of inference for systems of (weakly) interacting diffusions for which the mean field limit exists and is described by a nonlinear diffusion process of McKean type, obtained in the limit as the number of interacting processes N goes to infinity. When the number of interacting stochastic differential equations (SDEs) is large, the inference problem can become computationally intractable and it is often useful to study the problem of parameter estimation for the limiting mean field SDE. This is related, but distinct, from the problem of inference for multiscale diffusions [37, 35, 1, 2, 17] where the objective is to learn the parameters in the homogenized (limiting) SDE from observations of the full dynamics. Our goal is to show how the inference methodology using eigenfunction martingale estimating

*Department of Mathematics, Imperial College London, London SW7 2AZ, UK, g.pavliotis@imperial.ac.uk

†Institute of Mathematics, École Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland, andrea.zanoni@epfl.ch

functions that was applied in [2] to multiscale diffusions can be modified so that it can also be applied to interacting diffusions with a well defined mean field limit. It is useful to keep in mind the analogy between the homogenization and mean field limits, in the context of parameter estimation.

Inference for large interacting systems has attracted considerable attention, starting from the work of Kasonga [26], in which the maximum likelihood estimator (MLE) was considered. In particular, it was proved that the MLE for estimating parameters in the drift, when the drift is linearly dependent on the parameters, given continuous time observations of *all* the particles of the N-particle system, is consistent and asymptotically normal in the limit as $N \to \infty$. In this setting, it is possible to test whether the particles are interacting or not, at least in the linear case, i.e., for a system of interacting Ornstein–Uhlenbeck processes. Consistency and asymptotic normality of the sieve estimator and an approximate MLE estimator, i.e., when discrete observations of all the particles are given, was studied in [8] in the same framework of linear dependence on the parameters for the drift and known diffusion coefficient. Moreover, MLE inference of the mean field Ornstein–Uhlenbeck SDE was also considered. Properties of the MLE for the McKean SDE, when a continuous path of the SDE is observed, were studied in [43]. Consistency of the MLE was proved and an application to a model for ionic diffusion was presented. The MLE estimator for the McKean SDE was also considered in [29] and numerical experiments for the mean field Ornstein–Uhlenbeck process were presented. The combined large particle and long time asymptotics, $N \to \infty$ and $T \to \infty$, of the MLE for the case of a quadratic interaction, i.e., for interacting Ornstein–Uhlenbeck processes, was studied in [10]. Unlike the previous works mentioned in this literature review, the case where only a single particle trajectory is observed was considered in this paper. It was shown that the parameters in the drift can be estimated with optimal rate of convergence simultaneously in mean-field limit and in long-time dynamics. Offline and online inference for the McKean SDE was studied in [39]. Consistency and asymptotic normality of the offline MLE for the interacting particle system in the limit as the number of particles $N \to \infty$ was shown. In addition, an online parameter estimator for the mean field SDE was proposed, which evolves according to a continuous-time stochastic gradient descent algorithm on the asymptotic log-likelihood of the interacting particle system.

In this paper we consider systems of exchangeable weakly interacting diffusions for which uniform propagation of chaos results are known [33, 4, 5, 31, 12] and for which the mean field SDE has a unique invariant measure. We assume that we are given a sample of discrete-time observations of a single particle. Due to exchangeability, this amount of information should be sufficient to infer parameters in the mean field SDE, in the joint asymptotic limit as the number of observations and the number of particles go to infinity. Our approach consists of constructing martingale estimating functions [6, 27] based on the eigenvalues and the eigenfunctions of the generator of the mean field dynamics. Then, our eigenfunction estimator is the zero of the estimating function. The martingale estimator based on the eigenfunctions of the generator was used to study the inference problem for multiscale diffusions in [2]. Unlike the finite dimensional case, the mean field SDE is a measure-valued process and the generator is a nonlinear operator, dependent on the law of the process. A direct application of the martingale eigenfunction estimator would require the solution of a nonlinear eigenvalue problem that can be computationally demanding and that would also lead to eigenfunctions depending on time via their dependence on the law of the process. We circumvent this difficulty by linearizing the generator around the (unique) invariant measure of the mean field dynamics. In particular, we replace the density of the law with the density of the invariant measure of the process. This leads to a standard Sturm–Liouville type of eigenvalue problem that we can analyze and also solve numerically at a low computational cost. In this paper we consider the framework where the invariant measure of the mean field SDE is unique. We remark, however, that our numerical experiments show that our methodology applies to McKean SDEs that exhibit phase transitions, i.e., that have multiple stationary measures, as long as we are below the transition point, or the form of the invariant measure is known up to a finite set of parameters, e.g., moments.

When the mean field dynamics has a unique invariant measure, we first show the existence of the estimator with high probability when the number of available data and particles is large enough, and then analyze its consistency proving the asymptotic convergence towards the true value of the unknown parameter and providing a rate. Moreover, we prove that the estimator is asymptotically normal. We also note that the relationship between the number of observations and particles plays an important role in the study of the asymptotic properties of the estimator, in particular the latter must be sufficiently greater than the former in order for the previous results to hold. We then present a series of numerical experiments which confirm our theoretical results and we show the advantages of our method with respect to the MLE. In particular, in contrast with our estimator, the MLE is biased when we have sparse observations, i.e., when the sampling rate Δ is far from the asymptotic limit $\Delta \to 0$.

Main contributions. The main contributions of our work are summarized below.

- We propose a new methodology for estimating parameters in the drift of large interacting particle systems when a sequence of discrete observations of a single particle is given. Our proposed estimator is based on the eigenvalues and eigenfunctions of the generator of the mean field SDE, linearized around the steady state.
- We show theoretically that our estimator is asymptotically unbiased and asymptotically normal in the limit as the number of observations and the number of particles go to infinity and we compute the rate of convergence.
- We demonstrate numerically that our proposed estimator is reliable and robust with respect to the sampling rate.

Outline. The rest of the paper is organized as follows. In Section 2 we introduce the framework of the problem under investigation and we present the main theoretical results, and in Section 3 we show several numerical experiments illustrating the potentiality of our approach. Finally, Section 4 is devoted to the proofs of the main theorems.

2. Problem setting.

In this work we consider a system of interacting particles in one dimension moving in a confining potential over the time interval $[0,T]$ whose interaction is governed by an interaction potential

(2.1)
$$\mathrm{d}X_t^{(n)} = -V'(X_t^{(n)};\alpha)\,\mathrm{d}t - \frac{1}{N}\sum_{i=1}^N W'(X_t^{(n)} - X_t^{(i)};\kappa)\,\mathrm{d}t + \sqrt{2\sigma}\,\mathrm{d}B_t^{(n)}, \qquad n = 1,\dots,N,$$
$$X_0^{(n)} \sim \nu, \qquad n = 1,\dots,N,$$

where N is the number of particles, $\{B_t^{(n)}\}_{n=1}^N$ are standard independent one dimensional Brownian motions, $V(\cdot;\alpha)$ and $W(\cdot;\kappa)$ are the confining and interaction potentials, respectively, which depend on some parameters $\alpha \in \mathbb{R}^{p_1}, \kappa \in \mathbb{R}^{p_2}$, and $\sigma > 0$ is the diffusion coefficient. We assume chaotic initial conditions, i.e., that the particles are initially distributed according to the same measure ν.

Remark 2.1. We consider the case when the particles move in one dimension for the clarity of exposition. In fact, the proposed method and our rigorous results can be easily generalized to the case of N interacting particles moving in dimension $d > 1$. However in higher dimensions the problem becomes more complex and expensive from a computational point of view.

We place ourselves in the same framework of [31], which is summarized in the following assumption.

Assumption 2.2. The confining and interaction potentials V and W, respectively, satisfy:
- $V(\cdot;\alpha) \in \mathcal{C}^2(\mathbb{R})$ is uniformly convex and polynomially bounded along with its derivatives uniformly in α;
- $W(\cdot;\kappa) \in \mathcal{C}^2(\mathbb{R})$ is even, convex and polynomially bounded along with its derivatives uniformly in κ.

It is well-known (see, e.g., [36, Chapter 4]) that under Assumption 2.2 the dynamics described by the system (2.1) is geometrically ergodic with unique invariant measure given by the Gibbs measure $\mu_\theta^N(\mathrm{d}\mathbf{x}) = \rho^N(\mathbf{x};\theta)\,\mathrm{d}\mathbf{x}$, where

$$\rho^N(\mathbf{x};\theta) = \frac{1}{Z^N}\exp\left\{-\frac{1}{\sigma}E^N(\mathbf{x};\theta)\right\}, \qquad Z^N = \int_{\mathbb{R}^N}\exp\left\{-\frac{1}{\sigma}E^N(\mathbf{x};\theta)\right\}\mathrm{d}\mathbf{x},$$

and $E^N(\cdot;\theta)$ is defined by

$$E^N(\mathbf{x};\theta) := \sum_{n=1}^N V(x_n;\alpha) + \frac{1}{2N}\sum_{n=1}^N\sum_{i=1}^N W(x_n - x_i;\kappa).$$

for $\theta = \begin{pmatrix}\alpha^\top & \kappa^\top\end{pmatrix}^\top \in \Theta \subseteq \mathbb{R}^p$ with $p = p_1 + p_2$ and Θ the set of admissible parameters. The main goal of this paper is the estimation of the unknown parameter $\theta \in \Theta$, given discrete observations of the path of one single particle. We are interested in applications involving large interacting particle systems, i.e., when $N \gg 1$, hence studying the whole system is not practical and can be computationally unfeasible. Therefore, our approach consists of considering the mean field limit which has already been thoroughly studied (see, e.g., [11, 19]). Letting the number of particles N go to infinity we obtain the nonlinear, in the sense of McKean, SDE

$$\begin{aligned}\mathrm{d}X_t &= -V'(X_t;\alpha)\,\mathrm{d}t - (W'(\cdot;\kappa) * u(\cdot,t))(X_t)\,\mathrm{d}t + \sqrt{2\sigma}\,\mathrm{d}B_t,\\ X_0 &\sim \nu,\end{aligned} \tag{2.2}$$

where $u(\cdot,t)$ is the density with respect to the Lebesgue measure of the law of X_t and the nonlinearity means that the drift of the SDE (2.2) depends on the law of the process. The density u is the solution of the nonlinear Fokker–Planck (McKean–Vlasov) equation

$$\frac{\partial u}{\partial t}(x,t) = \frac{\partial}{\partial x}\left(V'(x;\alpha)u(x,t) + (W'(\cdot;\kappa) * u(\cdot,t))(x,t)u(x,t) + \sigma \frac{\partial u}{\partial x}(x,t)\right),$$

with initial condition $u(x,0)\,\mathrm{d}x = \nu(\,\mathrm{d}x)$. It is well known that, in contrast to the finite dimensional dynamics, the mean field limit (2.2) can have, in the non-convex case more than one invariant measures $\mu_\theta(\,\mathrm{d}x) = \rho(x;\theta)\,\mathrm{d}x$ [11, 9]. The density of the stationary state(s) satisfies the stationary Fokker–Planck equation

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(V'(x;\alpha)\rho(x;\theta) + (W'(\cdot;\kappa) * \rho(\cdot;\theta))(x)\rho(x;\theta) + \rho'(x;\theta)\right) = 0,$$

where the second variable θ emphasizes the fact that ρ depends on the parameters α and κ of the potentials V and W, respectively. However, under Assumption 2.2 it has been proven in [31] that there exists a unique invariant measure which is the solution of

$$\rho(x;\theta) = \frac{1}{Z}\exp\left\{-\frac{1}{\sigma}\left(V(x;\alpha) + (W(\cdot;\kappa) * \rho(\cdot;\theta))(x)\right)\right\}, \tag{2.3}$$

where Z is the normalization constant

$$Z = \int_{\mathbb{R}} \exp\left\{-\frac{1}{\sigma}\left(V(x;\alpha) + (W(\cdot;\kappa) * \rho(\cdot;\theta))(x)\right)\right\}\,\mathrm{d}x.$$

Example 2.3. A particular choice for the interaction potential is the Curie–Weiss quadratic interaction [11]. We take $\kappa > 0$ and consider the confining potential

$$W(x;\kappa) = \frac{\kappa}{2}x^2.$$

The interacting particles system (2.1) becomes, for all $n = 1,\dots,N$

$$\mathrm{d}X_t^{(n)} = -V'(X_t^{(n)};\alpha)\,\mathrm{d}t - \kappa\left(X_t^{(n)} - \bar{X}_t^N\right)\,\mathrm{d}t + \sqrt{2\sigma}\,\mathrm{d}B_t^{(n)},$$

where $\bar{X}_t^N$ denotes the empirical mean

$$\bar{X}_t^N = \frac{1}{N}\sum_{i=1}^N X_t^{(i)}.$$

This interaction term creates a tendency for the particles to relax toward the center of gravity of the ensemble and the parameter κ measures the strength of the interaction between the agents, hence this model provides a simple example of cooperative interaction.

The mean field limit (2.2) then becomes

$$\mathrm{d}X_t = -V'(X_t;\alpha)\,\mathrm{d}t - \kappa\left(X_t - m_t\right)\,\mathrm{d}t + \sqrt{2\sigma}\,\mathrm{d}B_t,$$

where m_t denotes the expectation of X_t, $m_t = \mathbb{E}[X_t]$, and its unique (when the confining potential V is convex) invariant measure $\mu_\theta(\,\mathrm{d}x) = \rho(x;\theta)\,\mathrm{d}x$ is given by

$$\rho(x;\theta) = \frac{1}{Z} \exp\left\{ -\frac{1}{\sigma} \left(V(x;\alpha) + \kappa \left(\frac{1}{2}x^2 - mx \right) \right) \right\}, \tag{2.4}$$

with the constraint for the expectation with respect to the invariant measure

$$m = \int_{\mathbb{R}} x\rho(x;\theta)\,\mathrm{d}x, \tag{2.5}$$

and where

$$Z = \int_{\mathbb{R}} \exp\left\{ -\frac{1}{\sigma} \left(V(x;\alpha) + \kappa \left(\frac{1}{2}x^2 - mx \right) \right) \right\} \mathrm{d}x.$$

Equation (2.5) is the self-consistency equation [11, 15, 23] that enables us to calculate the invariant measure and, then, the stationary state(s). In the case where the confining potential is quadratic, we have a system linear SDEs and the mean field limit reduces to the mean field Ornstein-Uhlenbeck SDE. In this case the first moment vanishes, $m = 0$, and the invariant measure is unique (this is the case, of course, of arbitrary strictly convex confining potentials). The inference problem for the linear interacting particle system and for the corresponding mean field limit is easier than that of the general case. We emphasize that, unlike this present work, most earlier papers, e.g., [26, 8], focus on this linear case, i.e., on systems of weakly interacting linear stochastic differential equations. The estimator proposed and studied in this paper can be applied to arbitrary non-quadratic interaction and confining potentials.

2.1. Parameter estimation problem. We now present our method for the estimation of the unknown parameter $\theta = (\alpha, \kappa) \in \Theta \subseteq \mathbb{R}^p$, given discrete observation of a single particle of the system (2.1). Consider $M + 1$ uniformly distributed observation times $0 = t_0 < t_1 < \cdots < t_M = T$, let $\Delta = t_m - t_{m-1}$ be the sampling rate and let $(X_t^{(n)})_{t \in [0,T]}$ be a realization of the n-th particle of the solution of the system (2.1) for some $n = 1, \dots, N$. We then aim to estimate the unknown parameter θ given a sample $\{\widetilde X_m^{(n)}\}_{m=0}^M$ of the realization where $\widetilde X_m^{(n)} = X_{t_m}^{(n)}$ and $t_m = \Delta m$. We want to construct martingale estimating functions based on the eigenfunctions and the eigenvalues of the generator of the dynamics, a technique which was initially proposed in [27] for single-scale SDEs and then successfully applied to multiscale SDEs in [2]. In principle, the methodology developed in [27] can be applied to the $N-$particle system. However, this would require solving the eigenvalue problem for the generator of an $N-$dimensional diffusion process, which is computationally expensive. Moreover, our fundamental assumption is that are observing a single particle and thus we do not have a complete knowledge of the system. Therefore, we construct the martingale estimating functions employing the generator of the mean field dynamics, which is a good approximation of the path of a single particle when the number N of particles is large [41]. Let $\mathcal{L}_t$ be the generator of the mean field limit SDE (2.2)

$$\mathcal{L}_t = -\left(V'(\cdot;\alpha) + (W'(\cdot;\kappa) * u(\cdot,t)) \right) \frac{\mathrm{d}}{\mathrm{d}x} + \sigma \frac{\mathrm{d}^2}{\mathrm{d}x^2},$$

and let $\mathcal{L}$ be the generator obtained replacing the density $u(\cdot, t)$ with the density $\rho(\cdot; \theta)$, i.e., linearizing the generator around the invariant measure μ_θ

$$\mathcal{L} = -\left(V'(\cdot; \alpha) + (W'(\cdot; \kappa) * \rho(\cdot; \theta))\right) \frac{\mathrm{d}}{\mathrm{d}x} + \sigma \frac{\mathrm{d}^2}{\mathrm{d}x^2}.$$

We then consider the eigenvalue problem $-\mathcal{L}\phi(\cdot; \theta) = \lambda(\theta)\phi(\cdot; \theta)$, which reads

$$\sigma\phi''(x; \theta) - \left(V'(x; \alpha) + (W'(\cdot; \kappa) * \rho(\cdot; \theta))(x)\right) \phi'(x; \theta) + \lambda(\theta)\phi(x; \theta) = 0, \tag{2.6}$$

and from the well-known spectral theory of diffusion processes (see, e.g., [25]) we deduce the existence of a countable set of eigenvalues $0 = \lambda_0(\theta) < \lambda_1(\theta) < \cdots < \lambda_j(\theta) \uparrow \infty$ whose corresponding eigenfunctions $\{\phi_j(\cdot; \theta)\}_{j=0}^\infty$ form an orthonormal basis of the weighted space $L^2(\rho(\cdot; \theta))$. In fact, even if the SDE (2.2) is nonlinear, when $X_0 \sim \rho(\cdot; \theta)$ then the solution X_t behaves like a classic diffusion process with drift function $-V'(\cdot; \alpha) - W'(\cdot; \kappa) * \rho(\cdot; \theta)$, hence the spectral theory for diffusion processes still holds. We also state here the variational formulation of the eigenvalue problem, which will be employed to implement numerically the proposed methodology. Let φ be a test function and multiply equation (2.6) by $\varphi\rho(\cdot; \theta)$, where the density $\rho(\cdot; \theta)$ of the invariant measure μ_θ is defined in (2.3). Then, integrating over $\mathbb{R}$ and by parts we obtain

$$\sigma \int_{\mathbb{R}} \phi'(x; \theta)\varphi'(x)\rho(x; \theta) \,\mathrm{d}x = \lambda(\theta) \int_{\mathbb{R}} \phi(x; \theta)\varphi(x)\rho(x; \theta) \,\mathrm{d}x.$$

We are now ready to present how to employ the eigenvalue problem in the construction of the martingale estimation function and afterwords in the definition of our estimator. Let J be a positive integer and let $\psi_j(\cdot; \theta) \colon \mathbb{R} \to \mathbb{R}^p$ for $j = 1, \dots, J$ be arbitrary functions dependent on the parameter θ which satisfy Assumption 2.5 below, and define the martingale estimating function $G_{M,N}^J \colon \Theta \to \mathbb{R}^p$ as

$$G_{M,N}^J(\theta) \coloneqq \frac{1}{M} \sum_{m=0}^{M-1} \sum_{j=1}^{J} g_j(\widetilde{X}_m^{(n)}, \widetilde{X}_{m+1}^{(n)}; \theta),$$

where

$$g_j(x, y; \theta) \coloneqq \psi_j(x; \theta) \left(\phi_j(y; \theta) - e^{-\lambda_j(\theta)\Delta}\phi_j(x; \theta)\right), \tag{2.7}$$

and $\{\widetilde{X}_m^{(n)}\}_{m=0}^M$ is the set of observations of the n-th particle from the system with N particles. The estimator we propose is then given by the solution $\widehat{\theta}_{M,N}^J$ of the p-dimensional nonlinear system

$$G_{M,N}^J(\theta) = \mathbf{0}, \tag{2.8}$$

where $\mathbf{0} \in \mathbb{R}^p$ denotes the vector with all components equal to zero. The main steps needed to obtain the estimator $\widehat{\theta}_{M,N}^J$ are summarized in Algorithm 2.1. For further details about the implementation and for discussions about the choice of the arbitrary functions $\{\psi_j(\cdot; \theta)\}_{j=1}^J$ we refer to Appendix B and Remark 2.6 in [2].

Remark 2.4. The main limitation of our approach is that the knowledge of the invariant measure is required in order to construct the martingale estimating function (step 1 in Algorithm 2.1). However, it is often the case that the invariant measure is known up to a set of parameters, such as moments, i.e., only the functional form of the invariant measure is known. These parameters (moments) are obtained by solving appropriate self-consistency equations [15, Section 2.3]. When such a situation arises, it is possible to first learn these parameters using the available data, e.g., estimate the moments that appear in the invariant measure by employing the law of large numbers. Then, we are in the setting where our technique applies and we can proceed in the same way, as shown in the numerical experiments in Sections 3.5 and 3.6. In summary, it is sufficient to replace step 1 in Algorithm 2.1 with "estimate the moments in the invariant measure $\rho(\cdot;\theta)$".

We finally introduce a technical hypothesis which will be needed for the proofs of our main results.

Assumption 2.5. Let $\Theta \subseteq \mathbb{R}^p$ be a compact set. Then the following hold for all $\theta \in \Theta$ and for all $j = 1, \dots, J$:

1. $\psi_j(x;\theta)$ is continuously differentiable with respect to θ for all $x \in \mathbb{R}$;
2. all components of $\psi_j(\cdot;\theta)$, $\psi_j'(\cdot;\theta)$, $\dot\psi_j(\cdot;\theta)$, $\dot\psi_j'(\cdot;\theta)$ are polynomially bounded uniformly in θ;
3. the potentials V and W are such that $\phi_j(\cdot;\theta)$, $\phi_j'(\cdot;\theta)$ and all components of $\dot\phi_j(\cdot;\theta)$, $\dot\phi_j'(\cdot;\theta)$ are polynomially bounded uniformly in θ;

where the dot denotes either the Jacobian matrix or the gradient with respect to θ.

Remark 2.6. Assumption 2.5(i) together with [38, Sections 2 and 6] gives the continuous differentiability of the vector-valued function $G_{M,N}^J(\theta)$ with respect to the unknown parameter θ.

Remark 2.7. In this paper we always assume that the diffusion coefficient σ in (2.1) is known. We remark that this is not an essential limitation of our methodology; in fact, if the diffusion coefficient is also unknown, we can consider the parameter set to be estimated to be $\widetilde\theta = (\theta,\sigma) = (\alpha,\kappa,\sigma) \in \mathbb{R}^{p+1}$ and repeat the same procedure. The estimator is then obtained as the solution of the nonlinear system of dimension $p+1$ corresponding to (2.8). A numerical experiment illustrating this procedure is given in Section 3.3. Moreover, our main theoretical results remain valid and the proofs do not need any major changes. Alternatively, it is possible to first estimate the diffusion coefficient using the quadratic variation and then proceed with the methodology proposed in this paper.

Example 2.8. Let us consider the Curie–Weiss quadratic interaction introduced in Example 2.3 as well as a quadratic–Ornstein–Uhlenbeck–confining potential $V(x;\alpha) = \frac{1}{2}x^2$. In this case the only unknown parameter is κ and the eigenvalue problem (2.6) reads

$$\sigma\phi''(x;\theta) - (1+\kappa)x\phi'(x;\theta) + \lambda(\theta)\phi(x;\theta) = 0, \tag{2.9}$$

so that the eigenvalue and eigenfunctions can be computed analytically [2, Section 3.1]. In particular, the first eigenvalue and eigenfunction are given by $\lambda_1(\theta) = 1+\kappa$ and $\phi_1(x;\theta) = x$,

Algorithm 2.1 Estimation of $\theta \in \Theta$

Input: Observations $\{\widetilde{X}_m^{(n)}\}_{m=0}^M$.
Distance between two consecutive observations Δ.
Number of eigenvalues and eigenfunctions J.
Functions $\{\psi_j(x;\theta)\}_{j=1}^J$.
Confining potential V and interaction potential W.
Diffusion coefficient σ.

Output: Estimation $\widehat{\theta}_{M,N}^J$ of θ.

1: Find the invariant measure $\rho(\cdot;\theta)$.
2: Consider the equation
$\sigma\phi''(x;\theta) - (V'(x;\alpha) + (W'(\cdot;\kappa) * \rho(\cdot;\theta))(x))\,\phi'(x;\theta) + \lambda(\theta)\phi(x;\theta) = 0.$
3: Compute the first J eigenvalues $\{\lambda_j(\theta)\}_{j=1}^J$ and eigenfunctions $\{\phi_j(\cdot;\theta)\}_{j=1}^J$.
4: Construct the function $g_j(x,y;\theta) = \psi_j(x;\theta)\left(\phi_j(y;\theta) - e^{-\lambda_j(\theta)\Delta}\phi_j(x;\theta)\right)$.
5: Construct the score function $G_{M,N}^J(\theta) = \frac{1}{M}\sum_{m=0}^{M-1}\sum_{j=1}^J g_j(\widetilde{X}_m^{(n)}, \widetilde{X}_{m+1}^{(n)};\theta)$.
6: Let $\widehat{\theta}_{M,N}^J$ be the solution of the nonlinear system $G_{M,N}^J(\theta) = \mathbf{0}$.

respectively. Therefore, letting $\psi_1(x;\theta) = x$ we have an explicit expression for our estimator

$$\widehat{\theta}_{M,N}^1 = -1 - \frac{1}{\Delta}\log\left(\frac{\sum_{m=0}^{M-1}\widetilde{X}_m^{(n)}\widetilde{X}_{m+1}^{(n)}}{\sum_{m=0}^{M-1}(\widetilde{X}_m^{(n)})^2}\right). \tag{2.10}$$

For additional details regarding the eigenvalue problem (2.9) we refer to [2, Section 3.1]. We also remark that when the drift coefficient of the Ornstein–Uhlenbeck process is unknown, i.e., if we consider the confining potential $V(x;\alpha) = \frac{\alpha}{2}x^2$, then the eigenvalue problem reads

$$\sigma\phi''(x;\theta) - (\alpha+\kappa)x\phi'(x;\theta) + \lambda(\theta)\phi(x;\theta) = 0,$$

which only depends on the sum $\alpha + \kappa$ and not on the single parameters alone. Therefore, in this case it is not possible to estimate the unknown coefficients α and κ, but we can only estimate their sum. This is in contrast with the set up in [26], where *all* the particles are observed in continuous time. When this amount of information is available, it is possible to check whether or not the particles are interacting, i.e., to check whether $\kappa = 0$ or not (see [26, Section 4]).

2.2. Main results. In this section we present the main theoretical results of this work. In particular, we prove that our estimator $\widehat{\theta}_{M,N}^J$ is asymptotically unbiased (consistent) and asymptotically normal as the number of observations M and particles N go to infinity and we compute the rate of convergence towards the true value of the parameter, which we denote by θ_0. Part of the proof of the consistency of the estimator, which will be presented in detail in Section 4, is inspired by our previous work [2, Section 5]. In this paper we studied the asymptotic properties of a similar estimator for multiscale SDEs letting the number of observations go to infinity and the multiscale parameter vanish. The proofs or our results in the present work also requires us to perform a rigorous asymptotic analysis with respect to two parameters, the number of observations and the number of particles.

We first define the Jacobian matrix of the function g_j introduced in (2.7) with respect to the parameter θ, with $\otimes$ denoting the outer product in $\mathbb{R}^p$,

$$\begin{aligned} h_j(x,y;\theta) &\coloneqq \dot g_j(x,y;\theta) \\ &= \dot\psi_j(x;\theta)\left(\phi_j(y;\theta) - e^{-\lambda_j(\theta)\Delta}\phi_j(x;\theta)\right) \\ &\quad + \psi_j(x;\theta)\otimes\left(\dot\phi_j(y;\theta) - e^{-\lambda_j(\theta)\Delta}\left(\dot\phi_j(x;\theta) - \Delta\dot\lambda_j(\theta)\phi_j(x,\theta)\right)\right), \end{aligned}$$

as well as the following quantity

$$\ell_{j,k}(x,y;\theta) \coloneqq (\psi_j(x;\theta)\otimes\psi_k(x;\theta))\left(\phi_j(y;\theta)\phi_k(y;\theta) - e^{-(\lambda_j(\theta)+\lambda_k(\theta))\Delta}\phi_j(x;\theta)\phi_k(x;\theta)\right).$$

We remark that whenever we write $\mathbb{E}^{\mu_\theta}$ we mean that $X_0 \sim \mu_\theta$ and similarly for the other probability measures.

We now present our main results. In Theorem 2.9 we prove that our estimator is consistent.

Theorem 2.9. *Let J be a positive integer and let $\{\widetilde X_m^{(n)}\}_{m=1}^M$ be a set of observations obtained by system (2.1) with true parameter θ_0. Under Assumptions 2.2 and 2.5 and if*

$$\det\left(\sum_{j=1}^J \mathbb{E}^{\mu_{\theta_0}}\left[h_j(X_0,X_\Delta;\theta_0)\right]\right) \neq 0, \tag{2.11}$$

there exists $N_0 > 0$ such that for all $N > N_0$ an estimator $\widehat\theta^J_{M,N}$, which solves the system $G^J_{M,N}(\theta) = 0$, exists with probability tending to one as M goes to infinity. Moreover, the estimator $\widehat\theta^J_{M,N}$ is asymptotically unbiased, i.e.,

$$\lim_{N\to\infty}\lim_{M\to\infty}\widehat\theta^J_{M,N} = \theta_0, \qquad \text{in probability}, \tag{2.12}$$

$$\lim_{M\to\infty}\lim_{N\to\infty}\widehat\theta^J_{M,N} = \theta_0, \qquad \text{in probability}, \tag{2.13}$$

and if $M = o(N)$

$$\lim_{M,N\to\infty}\widehat\theta^J_{M,N} = \theta_0, \qquad \text{in probability.} \tag{2.14}$$

Then, in Theorem 2.10 we provide a rate of convergence for our estimator.

Theorem 2.10. *Let the assumptions of Theorem 2.9 hold, and let us introduce the notation*

$$\Xi^J_{M,N} \coloneqq \left(\frac{1}{\sqrt M} + \frac{1}{\sqrt N}\right)^{-1}\left\|\widehat\theta^J_{M,N} - \theta_0\right\|.$$

Then, for all $\varepsilon > 0$ there exists $K_\varepsilon > 0$ such that

$$\lim_{N\to\infty}\lim_{M\to\infty} \mathbb{P}\left(\Xi^J_{M,N} > K_\varepsilon\right) < \varepsilon, \tag{2.15}$$

$$\lim_{M\to\infty}\lim_{N\to\infty} \mathbb{P}\left(\Xi^J_{M,N} > K_\varepsilon\right) < \varepsilon, \tag{2.16}$$

and if $M = o(\sqrt{N})$

$$\lim_{M,N\to\infty} \mathbb{P}\left(\Xi^J_{M,N} > K_\varepsilon\right) < \varepsilon. \tag{2.17}$$

Finally, in Theorem 2.11 we show that our estimator is asymptotically normal.

Theorem 2.11. *Let the assumptions of Theorem 2.9 hold with $M = o(\sqrt{N})$. Then, the estimator $\widehat{\theta}^J_{M,N}$ is asymptotically normal, i.e.,*

$$\lim_{M,N\to\infty} \sqrt{M}\left(\widehat{\theta}^J_{M,N} - \theta_0\right) = \Lambda^J \sim \mathcal{N}(\mathbf{0}, \Gamma^J_0), \qquad \text{in distribution,}$$

where

$$\begin{aligned}\Gamma^J_0 = &\left(\sum_{j=1}^J \mathbb{E}^\mu\left[h_j(X_0, X_\Delta; \theta_0)\right]\right)^{-1} \left(\sum_{j=1}^J\sum_{k=1}^J \mathbb{E}^\mu\left[\ell_{j,k}(X_0, X_\Delta; \theta_0)\right]\right) \\ &\times \left(\sum_{j=1}^J \mathbb{E}^\mu\left[h_j(X_0, X_\Delta; \theta_0)\right]\right)^{-\top}.\end{aligned} \tag{2.18}$$

Remark 2.12. We note that the technical assumption (2.11) is not a serious limitation of the validity of the theorem; in fact, it is a nondegeneracy hypothesis which holds true in all nonpathological cases and is equivalent to [27, Condition 4.2(a)] and [2, Assumption 3.1].

Remark 2.13. For the proof of the main results, we need to assume that, roughly speaking, the number of particles goes to infinity faster than the number of observations. It is not clear whether this assumption is strictly necessary. We expect that noncommutativity issues between the different distinguished limits may arise in the case where the mean field dynamics exhibits phase transitions, i.e., when the stationary state is not unique, see [13]. We will study the consequences of this noncommutativity due to phase transitions to the performance of our estimator and, more generally, to the inference problem in future work.

3. Numerical experiments. In this section we present a series of numerical experiments to validate our theoretical results and demonstrate the effectiveness of our estimator in estimate unknown drift parameters of interacting particle systems. In order to generate synthetic data we employ the Euler–Maruyama method with a time step $h = 0.01$ to solve numerically system (2.1) and obtain $(X^{(n)}_t)_{t\in[0,T]}$ for all $n = 1, \dots, N$. Notice that in order to preserve the exchangeability property of the system it is important to set the same initial condition for all the particles, hence we take $X^{(n)}_0 = 0$ for all $n = 1, \dots, N$. We then randomly choose a value $n^* \in \{1, \dots, N\}$ and we assume to know a sample $\{X^{(n^*)}_m\}_{m=0}^M$ of observations obtained

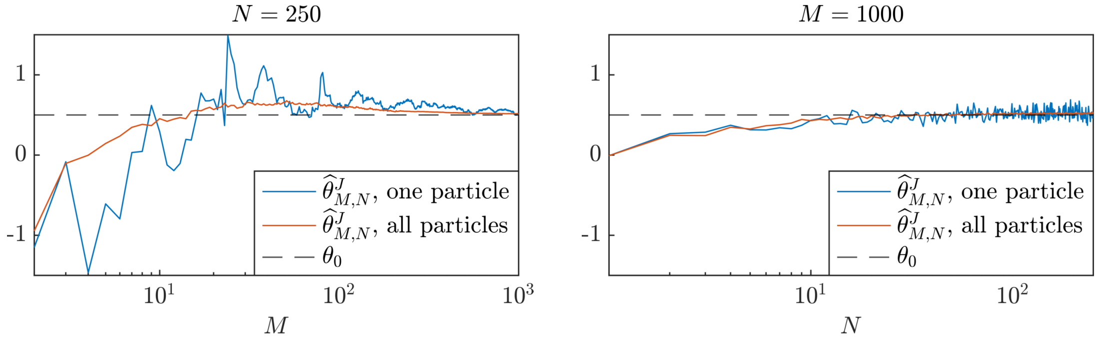


Figure 1. *Sensitivity analysis for the Ornstein–Uhlenbeck potential with respect to the number M of observations and N of particles, for the estimator $\widehat{\theta}^J_{M,N}$ with $J = 1$.*

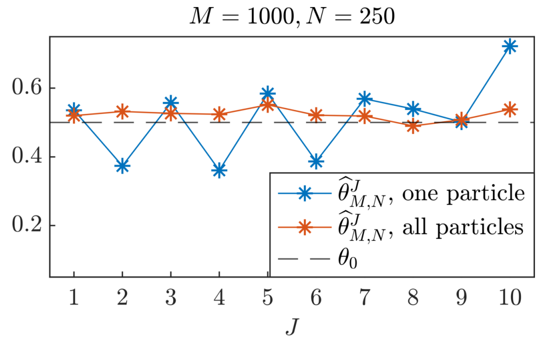


Figure 2. *Sensitivity analysis for the Ornstein–Uhlenbeck potential with respect to the number J of eigenvalues and eigenfunctions, for the estimator $\widehat{\theta}^J_{M,N}$.*

from the n^*-th particle with sampling rate Δ. We remark that the parameters h and Δ are not related to each other, in fact the former is only used to generate the data, while the latter is the actual distance between two consecutive observations. We repeat the same procedure for $L = 5$ different realizations of the Brownian motions and then we compute the average of the values obtained employing our estimator $\widehat{\theta}^J_{M,N}$. In the following, we first perform a sensitivity analysis with respect to the number of observations M, particles N and eigenvalues and eigenfunctions employed in the estimation J, then we confirm our theoretical results given in Theorems 2.9 to 2.11 and finally we test our technique with more challenging academic examples which do not exactly fit into the theory.

3.1. Sensitivity analysis and rate of convergence. We consider the setting of Example 2.8 choosing $\sigma = 1$, i.e., the interacting particles system reads

$$\mathrm{d}X_t^{(n)} = -X_t^{(n)}\,\mathrm{d}t - \kappa\left(X_t^{(n)} - \bar{X}_t^N\right)\mathrm{d}t + \sqrt{2}\,\mathrm{d}B_t^{(n)}, \qquad n = 1, \dots, N, \tag{3.1}$$

and we aim to estimate the interaction parameter κ, so we write $\theta = \kappa$. We set $\kappa = 0.5$ and the number of eigenvalues and eigenfunctions $J = 1$ with $\psi_1(x;\theta) = x$, so that we can employ the analytical expression of our estimator given in (2.10). In Figure 1 we perform a sensitivity analysis for the estimator $\widehat{\theta}^1_{M,N}$ fixing $\Delta = 1$, varying the number M of observations

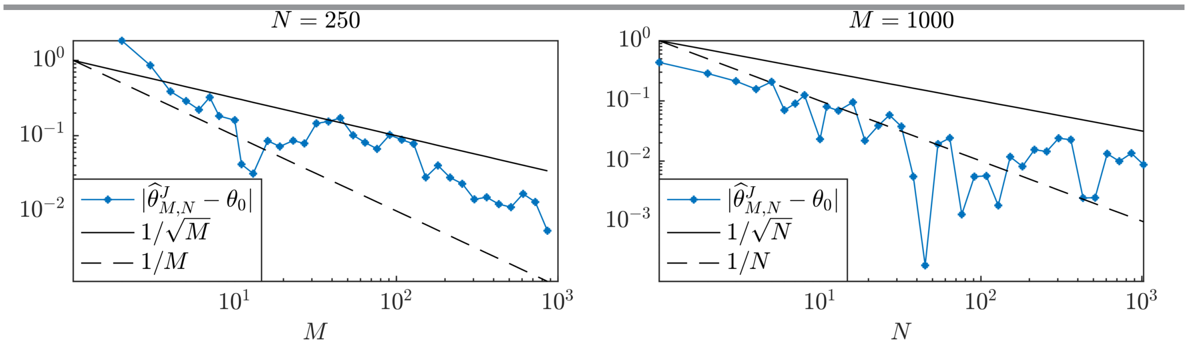


Figure 3. *Rates of convergence for the Ornstein–Uhlenbeck potential with respect to the number M of observations and N of particles, for the estimator $\widehat{\theta}^J_{M,N}$ with $J = 1$.*

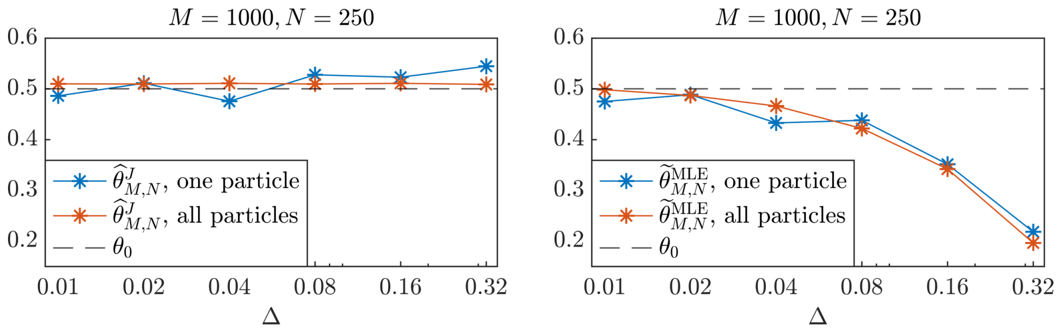


Figure 4. *Comparison between the estimator $\widehat{\theta}^J_{M,N}$ with $J = 1$ (left) and the maximum likelihood estimator $\widetilde{\theta}^{\mathrm{MLE}}_{M,N}$ (right) varying the distance Δ between two consecutive observations for the Ornstein–Uhlenbeck potential.*

and N of particles and choosing as other parameter respectively $N = 250$ and $M = 1000$, for which convergence has been reached. The blue line is the estimation given by one single particle while the red line is obtained by averaging the estimations computed employing all the different particles. We notice that convergence is reached when both N and M are large enough and, as expected, the estimation computed by averaging over all the particles stabilizes faster. Moreover, in Figure 2 we fix $M = 1000$ and $N = 250$ and we compare the results for different numbers J of eigenvalues and eigenfunctions employed in the construction of the estimating function. We observe that increasing the value of J does not significantly improves the results, hence it seems preferable to always choose $J = 1$ in order to reduce the computational cost. Finally, in Figure 3 we verify that the rates of convergence of the estimator $\widehat{\theta}^1_{M,N}$ towards the exact value θ_0 with respect to the number of observations M and particles N are consistent with the theoretical results given in Theorem 2.10. In particular, we observe that approximately it holds

$$\left|\widehat{\theta}^1_{M,N} - \theta_0\right| \simeq \mathcal{O}\left(\frac{1}{\sqrt{M}} + \frac{1}{\sqrt{N}}\right).$$

3.2. Comparison with the maximum likelihood estimator. We keep the same setting of Section 3.1 and we compare the results of our estimator with a maximum likelihood estimator.

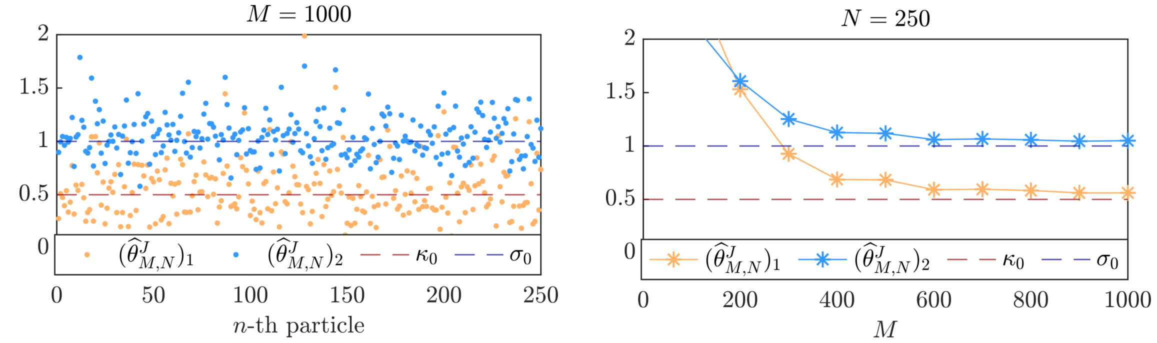


Figure 5. *Simultaneous inference of the interaction and diffusion coefficients for the Ornstein–Uhlenbeck potential. Left: estimation $\widehat\theta^J_{M,N}$ obtained from each particle with $J = 2$. Right: average of the estimations varying the number of observations.*

In particular, in [26] the MLE for the interacting particles system with continuous observations is rigorously derived. Since for large values of N all the particles are approximately independent and identically distributed and we are assuming to observe only one particle, we replace the sample mean with the expectation with respect to the invariant measure, i.e., $\bar X_t^N = 0$, and we ignore the sum over all the particles. We then discretize the integrals in the formulation obtaining a modified MLE

$$\widetilde\theta^{\mathrm{MLE}}_{M,N} = -1 - \frac{\sum_{m=0}^{M-1} \widetilde X_m^{(n)} (\widetilde X_{m+1}^{(n)} - \widetilde X_m^{(n)})}{\Delta \sum_{m=0}^{M-1} (\widetilde X_m^{(n)})^2}. \tag{3.2}$$

In Figure 4 we repeat the estimation for different values of $\Delta = 0.01 \cdot 2^i$, for $i = 0, \dots, 5$, and we observe that, differently from our estimator, the MLE is unbiased only for small values of the sampling rate Δ, hence when the discrete observations approximate well the continuous trajectory. Notice also that, as highlighted by the numerical experiments, our estimator $\widehat\theta^1_{M,N}$ and the MLE $\widetilde\theta^{\mathrm{MLE}}_{M,N}$ defined respectively in (2.10) and (3.2) coincide in the limit of vanishing Δ. In fact, we can rewrite equation (2.10) as

$$\widehat\theta^1_{M,N} = -1 - \frac{1}{\Delta} \log \left(1 + \frac{\sum_{m=0}^{M-1} \widetilde X_m^{(n)} (\widetilde X_{m+1}^{(n)} - \widetilde X_m^{(n)})}{\sum_{m=0}^{M-1} (\widetilde X_m^{(n)})^2} \right),$$

observe that the fraction in the argument of the logarithm is $\mathcal{O}(\Delta)$ and employ the asymptotic expansion $\log(1 + x) \sim x$ for $x = o(1)$.

3.3. Diffusion coefficient. We still consider the setting of Example 2.8, but, differently from Section 3.1, we now assume the diffusion coefficient to be unknown and we aim to simultaneously retrieve the correct values of the interaction parameter and the diffusion coefficient, which are given by $\kappa = 0.5$ and $\sigma = 1$, respectively. We write $\theta = \begin{pmatrix} \kappa & \sigma \end{pmatrix}^\top$ and we set the number of particles $N = 250$ and the number of observations $M = 1000$ with sampling rate $\Delta = 1$. In order to construct the estimating functions we then employ $J = 2$ eigenvalues and eigenfunctions with functions $\psi_1(x; \theta) = \psi_2(x; \theta) = \begin{pmatrix} x^2 & x \end{pmatrix}^\top$. We remark that in the particular case of the Ornstein–Uhlenbeck process it is possible to express the eigenvalues and eigenfunctions analytically and the first two are given by

$$\begin{aligned} \lambda_1 &= 1+\kappa, & \phi_1(x;\theta) &= x, \\ \lambda_2 &= 2(1+\kappa), & \phi_2(x;\theta) &= x^2 - \frac{\sigma}{1+\kappa}. \end{aligned}$$

Note that the first eigenvalue and eigenfunction do not depend on the diffusion coefficient σ and therefore they alone do not provide enough information, hence it is important to choose at least $J = 2$. In Figure 5 we show the numerical results. On the left and we plot the estimation computed employing one single particle for all the N particles and we observe that the estimators are concentrated around the exact values. On the other hand, on the right, we average all the estimations previously computed and we pot the results varying the number of observations M. We notice that the estimations stabilize fast near the correct coefficients.

3.4. Central limit theorem. We keep the same setting of Section 3.1 and we validate numerically the central limit theorem which we proved theoretically in Theorem 2.11. In this particular case, the asymptotic variance Γ_0^J can be computed analytically. In fact, the mean field limit of (3.1) at stationarity is

$$\mathrm{d}X_t = -(1+\kappa)X_t\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}B_t^{(n)},$$

and its solution $(X_t)_{t\in[0,T]}$ is a Gaussian process, i.e., $X \sim \mathcal{GP}(m(t), \mathcal{C}(t,s))$, where $m(t) = 0$ and

$$\mathcal{C}(t,s) = \frac{1}{1+\kappa} e^{-(1+\kappa)|t-s|}.$$

Moreover, we have

$$h_1(x,y;\theta) = \Delta e^{-(1+\kappa)\Delta} x^2 \qquad \text{and} \qquad \ell_{1,1}(x,y;\theta) = x^2\left(y^2 - e^{-2(1+\kappa)\Delta}x^2\right),$$

and therefore we obtain

$$\Gamma_0^J = \frac{e^{2(1+\kappa)\Delta} - 1}{\Delta^2}.$$

We then fix the number of particles $N = 1500$, the number of observations $M = 1000$ and the sampling rate $\Delta = 1$. In Figure 6 we plot the quantity $\sqrt{M}(\widehat{\theta}_{M,N}^J - \theta_0)$ for any particle $n = 1, \dots, N$ and for $L = 500$ realizations of the Brownian motion and we observe that it is approximately distributed as $\mathcal{N}(0, \Gamma_0^J)$ accordingly to the theoretical result.

3.5. Bistable potential. We consider the setting of Example 2.3 and we analyse the bistable potential, i.e., we let the confining potential $V(\cdot;\alpha)$ be

$$V(x;\alpha) = \alpha \cdot \begin{pmatrix} \frac{x^4}{4} & -\frac{x^2}{2} \end{pmatrix}^\top,$$

with $\alpha = \begin{pmatrix} 1 & 2 \end{pmatrix}^\top$, which is the parameter that we aim to estimate, so we write $\theta = \alpha$. Moreover, we set the interaction term $\kappa = 0.5$ and the number of observations $M = 2000$

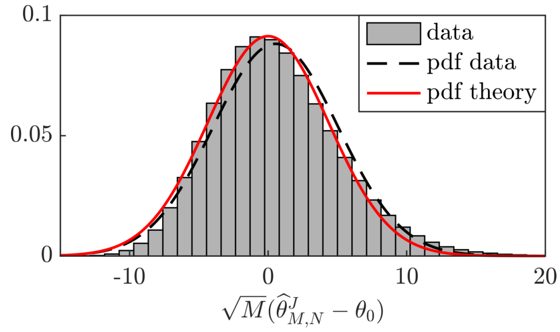


Figure 6. *Central limit theorems for the Ornstein–Uhlenbeck potential, for the estimator* $\widehat{\theta}^J_{M,N}$ *with* $J = 1$*.*

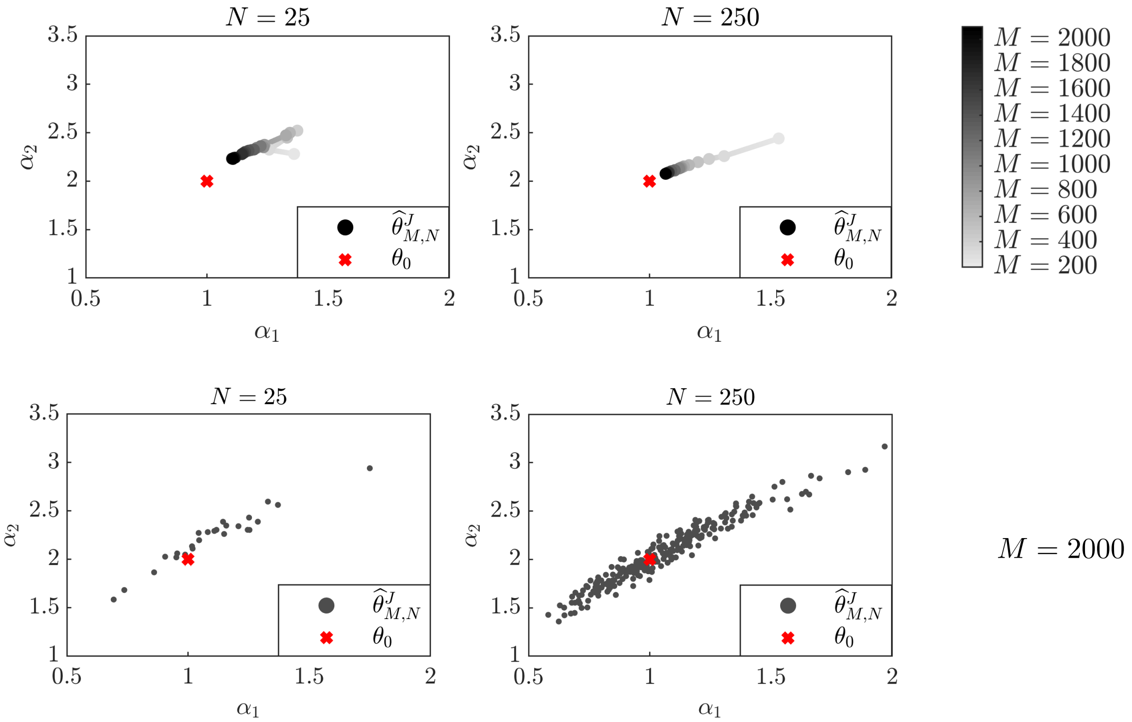


Figure 7. *Inference of the two-dimensional drift coefficient of the bistable potential below the phase transition. Top: average of the estimations* $\widehat{\theta}^J_{M,N}$ *with* $J = 1$ *varying the number of observations. Bottom: scatter plot of the estimations obtained from each particle.*

with sampling rate $\Delta = 0.5$. Finally, to construct the estimating functions we use $J = 1$ eigenfunctions and eigenvalues and we employ the function $\psi_1(x;\theta) = \begin{pmatrix} x & x^3 \end{pmatrix}^\top$. We remark that this example does not fit in Assumption 2.2, but if the diffusion coefficient σ is chosen sufficiently large, then we are below the phase transition and the mean field limit admits a unique invariant measure [11], so the theory applies. However, when the diffusion coefficient σ is below the critical noise strength, then a continuous phase transition occurs and two stationary states exist [23]. In particular, the transition point occurs at $\sigma \simeq 0.6$ with these

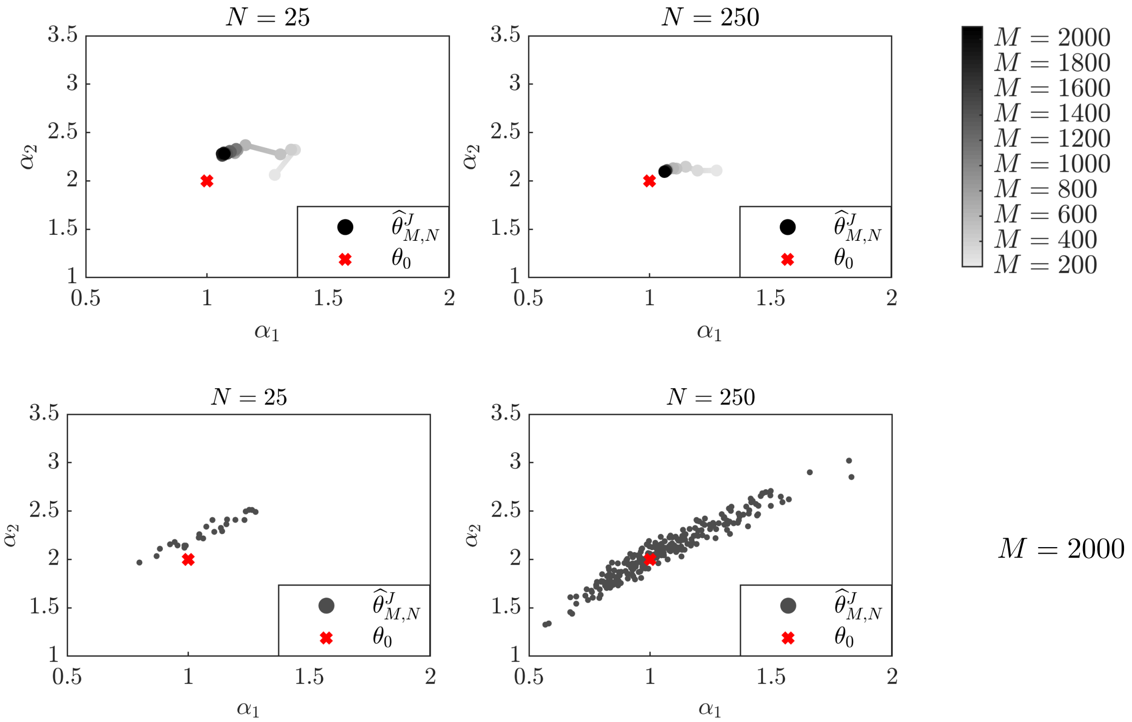


Figure 8. *Inference of the two-dimensional drift coefficient of the bistable potential above the phase transition. Top: average of the estimations $\widehat{\theta}^J_{M,N}$ with $J = 1$ varying the number of observations. Bottom: scatter plot of the estimations obtained from each particle.*

data. We therefore perform two numerical experiments, one below and one above the phase transition, setting $\sigma = 0.75$ and $\sigma = 0.5$. In the former we have a unique invariant measure, so we can follow the usual approach, while in the latter we do not know in which state the data are converging. Nevertheless, the invariant distribution is known up to the first moment by equation (2.4), so we first estimate the expectation using the law of large numbers with the available observations and then repeat the same procedure as in the previous case. In Figures 7 and 8 we plot the results of these two experiments. On the top of the figures we plot the evolution of our estimator varying the number of observations M for two different values of the number of particles, in particular $N = 25$ and $N = 250$. We observe that the estimator approaches the correct drift coefficient α as the number of observations M increases and, as expected, the final approximation is better when the number of particles is sufficiently big. Moreover, on the bottom of the same figures we show the scatter plot of the estimations obtained from each particle with $M = 2000$ observations and we can see that they are concentrated around the exact drift coefficient α. We finally remark that we do not notice significant differences between the two cases, yielding that the initial estimation of the first moment of the invariant measure does not affect the final results and thus that our methodology can be employed even when multiple stationary states exist.

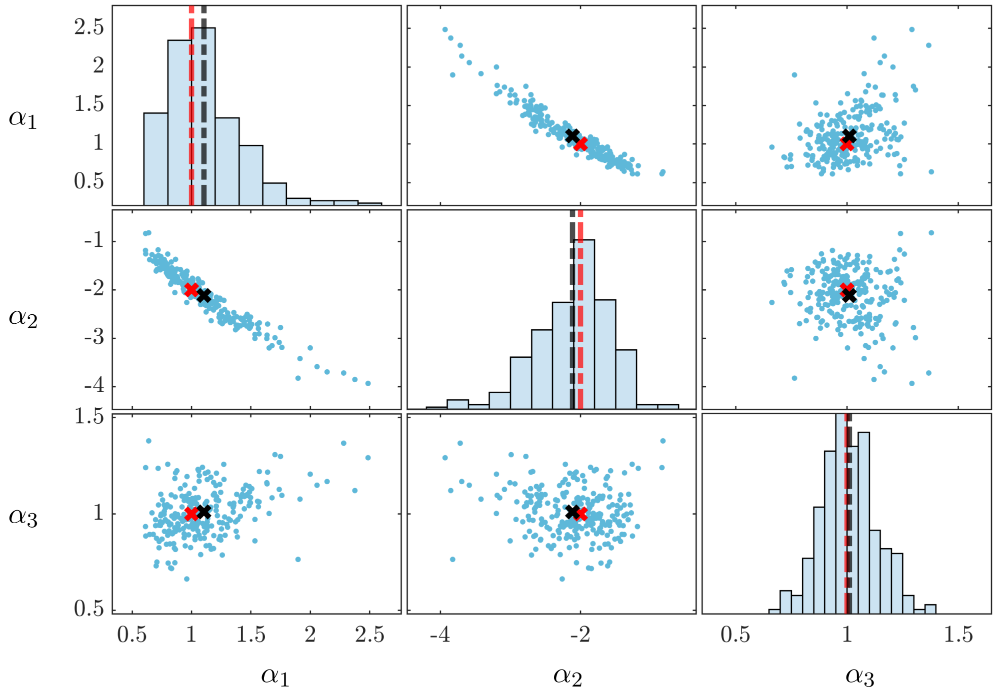


Figure 9. *Inference of the three-dimensional drift coefficient of a nonsymmteric potential for the estimator* $\widehat{\theta}^J_{M,N}$ *with* $J = 1$*. Diagonal: histogram of the estimations of each component obtained from all particles. Off-diagonal: scatter plot of the estimations obtained from all particles for two components at a time. Black and red stars/lines represent the average of the estimations and the exact value, respectively.*

3.6. Nonsymmetric confining potential. We still consider the same setting of Example 2.3 and we now study the case of a nonsymmetric potential. In particular, we let the confining potential $V(\cdot;\alpha)$ be

$$V(x;\alpha) = \alpha \cdot \begin{pmatrix} \frac{x^4}{4} & \frac{x^2}{2} & x \end{pmatrix}^\top,$$

with $\alpha = \begin{pmatrix} 1 & -2 & 1 \end{pmatrix}^\top$, which is the unknown parameter that we want to infer, hence we set $\theta = \alpha$. Notice that the confining potential is given by the sum of the bistable potential and a linear term which breaks the symmetry. This type of potentials of the form $V(x) = \sum_{\nu=1}^{\mathscr{N}} a_{2\nu} s^{2\nu} + a_1 s$, where $\mathscr{N} \geq 2$, $a_1, a_2 \in \mathbb{R}$, $a_4, \dots, a_{2(\mathscr{N}-1)} \geq 0$ and $a_{2\mathscr{N}} > 0$, which is used in the study of metastability and phase transitions and may have arbitrarily deep double wells, has been analyzed in [44, 42]. Similarly to the experiment in Section 3.5, this example does not satisfy Assumption 2.2 and more stationary states can exist. In particular, in [42] it has been proved the existence of an invariant measure around each critical point of the potential. We therefore adopt the same strategy as in the second part of Section 3.5 and, since the invariant measure is known up to the first moment by equation (2.4), we first approximate the expectation using the sample mean of the available observations, and then proceed with the following steps of the algorithm. We further set the interaction term $\kappa = 0.5$, the diffusion coefficient $\sigma = 1.5$, the number of particles $N = 250$ and the number of observations $M = 2000$

with sampling rate $\Delta = 0.5$. Moreover, to construct the estimating functions we use $J = 1$ eigenfunctions and eigenvalues and we employ the function $\psi_1(x;\theta) = \begin{pmatrix} x & x^2 & x^3 \end{pmatrix}^\top$. In Figure 9 we plot the results of the inference procedure considering two components of the three-dimensional drift coefficient at a time and the single components alone. We observe that the majority of the estimations obtained from all particles are concentrated around the exact values and that their average provides a reliable approximation of the true unknown. A peculiarity of this numerical experiment is the relationship between the first and second components of the estimated drift coefficient, in fact one increases when the other decreases and vice-versa, meaning that the two approximations appear to be correlated.

4. Proof of the main results.

In this section we present the proof of Theorems 2.9 to 2.11, which are the main results of this work. We first recall that due to [16, Lemma 2.3.1] the solution of the interacting particle system $X_t^{(n)}$ and of its mean field limit X_t have bounded moments of any order, in particular there exists a constant $C > 0$ independent of N such that for all $t \in [0,T]$, $n = 1,\dots,N$ and $q \geq 1$

$$\mathbb{E}\left[\left|X_t^{(n)}\right|^q\right]^{1/q} \leq C \qquad \text{and} \qquad \mathbb{E}\left[|X_t|^q\right]^{1/q} \leq C. \tag{4.1}$$

Moreover, in [31, Theorem 3.3] it is shown that each particle converges to the solution of the mean field limit with the same Brownian motion in L^2, i.e, that

$$\sup_{t\in[0,T]} \mathbb{E}\left[\left|X_t^{(n)} - X_t\right|^2\right]^{1/2} \leq \frac{C}{\sqrt{N}}, \tag{4.2}$$

where the constant C is also independent of the final time T. We also state here a formula which has been proved in [27] and will be crucial in the last part of the proof

$$\mathbb{E}^{\mu_{\theta_0}}[\phi_j(X_\Delta;\theta_0) \mid X_0 = x] = e^{-\lambda_j(\theta_0)\Delta}\phi_j(x;\theta_0), \qquad \text{for all } j = 1,\dots,J, \tag{4.3}$$

where θ_0 is the true parameter which generates the path $(X_t)_{t\in[0,T]}$ and $\mathbb{E}^{\mu_{\theta_0}}$ denotes the fact that $X_0 \sim \mu_{\theta_0}$. Before entering the main part of the proof, we introduce some notation and technical results which will be used later. We finally remark that all the constants will be denoted by C and their value can change from line to line.

4.1. Limits of the estimating function and its derivative.

Let us first define the following vector-valued functions $\mathbb{G}_M^J(\theta), \mathcal{G}_N^J(\theta), \mathscr{G}^J(\theta) \colon \mathbb{R}^p \to \mathbb{R}^p$ and matrix-valued functions $\mathbb{H}_M^J(\theta), \mathcal{H}_N^J(\theta), \mathscr{H}^J(\theta) \colon \mathbb{R}^p \to \mathbb{R}^{p\times p}$

$$\begin{aligned}
&\mathbb{G}_M^J(\theta) \coloneqq \frac{1}{M}\sum_{m=0}^{M-1}\sum_{j=1}^J g_j(\widetilde{X}_m, \widetilde{X}_{m+1};\theta), && \mathbb{H}_M^J(\theta) \coloneqq \frac{1}{M}\sum_{m=0}^{M-1}\sum_{j=1}^J h_j(\widetilde{X}_m, \widetilde{X}_{m+1};\theta), \\
&\mathcal{G}_N^J(\theta) \coloneqq \sum_{j=1}^J \mathbb{E}^{\mu_N}\left[g_j(X_0^{(n)}, X_\Delta^{(n)};\theta)\right], && \mathcal{H}_N^J(\theta) \coloneqq \sum_{j=1}^J \mathbb{E}^{\mu_N}\left[h_j(X_0^{(n)}, X_\Delta^{(n)};\theta)\right], \\
&\mathscr{G}^J(\theta) \coloneqq \sum_{j=1}^J \mathbb{E}^{\mu}\left[g_j(X_0, X_\Delta;\theta)\right], && \mathscr{H}^J(\theta) \coloneqq \sum_{j=1}^J \mathbb{E}^{\mu}\left[h_j(X_0, X_\Delta;\theta)\right].
\end{aligned} \tag{4.4}$$

The following lemma then shows that these quantities are bounded in a suitable norm and thus well defined.

Lemma 4.1. *Under Assumptions 2.2 and 2.5 there exists a constant $C > 0$ independent of M, N such that for all $q \geq 1$*

$$(i)\ \mathbb{E}\left[\left\|G^J_{M,N}(\theta)\right\|^q\right] \leq C, \qquad (ii)\ \mathbb{E}\left[\left\|\mathbb{G}^J_M(\theta)\right\|^q\right] \leq C,$$
$$(iii)\ \left\|\mathcal{G}^J_N(\theta)\right\| \leq C, \qquad (iv)\ \left\|\mathscr{G}^J(\theta)\right\| \leq C.$$

Proof. Since the argument is similar for the four cases, we only write the details of (i). Using the triangular inequality we have

$$\mathbb{E}\left[\left\|G^J_{M,N}(\theta)\right\|^q\right] \leq \frac{2^{q-1}}{M}\sum_{m=0}^{M-1}\sum_{j=1}^{J}\mathbb{E}\left[\left\|\psi_j(\widetilde X^{(n)}_m;\theta)\right\|^q\left(\left|\phi_j(\widetilde X^{(n)}_{m+1};\theta)\right|^q + \left|\phi_j(\widetilde X^{(n)}_m;\theta)\right|^q\right)\right],$$

and due to the Cauchy–Schwarz inequality we obtain

$$\begin{aligned}\mathbb{E}\left[\left\|G^J_{M,N}(\theta)\right\|\right] \leq \frac{2^{q-1}}{M}\sum_{m=0}^{M-1}\sum_{j=1}^{J}\mathbb{E}\left[\left\|\psi_j(\widetilde X^{(n)}_m;\theta)\right\|^{2q}\right]^{1/2}\mathbb{E}\left[\left|\phi_j(\widetilde X^{(n)}_{m+1};\theta)\right|^{2q}\right]^{1/2}\\ + \frac{2^{q-1}}{M}\sum_{m=0}^{M-1}\sum_{j=1}^{J}\mathbb{E}\left[\left\|\psi_j(\widetilde X^{(n)}_m;\theta)\right\|^{2q}\right]^{1/2}\mathbb{E}\left[\left|\phi_j(\widetilde X^{(n)}_{m};\theta)\right|^{2q}\right]^{1/2}.\end{aligned}$$

Finally, bound (4.1) together with the fact that ψ_j and ϕ_j are polynomially bounded for all $j = 1, \dots, J$ by Assumption 2.5 gives the desired result. ∎

In the next proposition we study the behaviour of the estimating function $G^J_{M,N}$ as the number of observations M and particles N go to infinity.

Proposition 4.2. *Under Assumptions 2.2 and 2.5 it holds for all $1 \leq q < 2$*

$$(i)\ \lim_{N\to\infty} G^J_{M,N}(\theta) = \mathbb{G}^J_M(\theta), \quad \text{in } L^q, \qquad (ii)\ \lim_{M\to\infty}\mathbb{G}^J_M(\theta) = \mathscr{G}^J(\theta), \quad \text{in } L^2,$$
$$(iii)\ \lim_{M\to\infty} G^J_{M,N}(\theta) = \mathcal{G}^J_N(\theta), \quad \text{in } L^2, \qquad (iv)\ \lim_{N\to\infty}\mathcal{G}^J_N(\theta) = \mathscr{G}^J(\theta).$$

Moreover, there exists a constant $C > 0$ independent of M, N and θ such that

$$(i)'\ \mathbb{E}\left[\left\|G^J_{M,N}(\theta) - \mathbb{G}^J_M(\theta)\right\|^q\right]^{1/q} \leq \frac{C}{\sqrt N}, \qquad (iv)'\ \left\|\mathcal{G}^J_N(\theta) - \mathscr{G}^J(\theta)\right\| \leq \frac{C}{\sqrt N}.$$

Proof. Results (ii) and (iii) are direct consequences of [6, Lemma 3.1] and of the ergodicity of the processes $(X^{(n)}_t)_{t\in[0,T]}$ and $(X_t)_{t\in[0,T]}$ given by [23, Section 1] and [31, Theorem 3.16], respectively. Let us now consider cases (i) and $(i)'$. Using the triangle inequality we have

$$\mathbb{E}\left[\left\|G^J_{M,N}(\theta) - \mathbb{G}^J_M(\theta)\right\|^q\right] \leq \frac{4^{q-1}}{M}\sum_{m=0}^{M-1}\sum_{j=1}^{J}\left(Q^{(1)}_{m,j} + Q^{(2)}_{m,j} + Q^{(3)}_{m,j} + Q^{(4)}_{m,j}\right),$$

where

$$\begin{aligned}
Q_{m,j}^{(1)} &\coloneqq \mathbb{E}\left[\left\|\psi_j(\widetilde{X}_m^{(n)};\theta)\right\|^q \left|\phi_j(\widetilde{X}_{m+1}^{(n)};\theta) - \phi_j(\widetilde{X}_{m+1};\theta)\right|^q\right], \\
Q_{m,j}^{(2)} &\coloneqq \mathbb{E}\left[\left\|\psi_j(\widetilde{X}_m^{(n)};\theta)\right\|^q \left|\phi_j(\widetilde{X}_m^{(n)};\theta) - \phi_j(\widetilde{X}_m;\theta)\right|^q\right], \\
Q_{m,j}^{(3)} &\coloneqq \mathbb{E}\left[\left\|\psi_j(\widetilde{X}_m^{(n)};\theta) - \psi_j(\widetilde{X}_m;\theta)\right\|^q \left|\phi_j(\widetilde{X}_{m+1};\theta)\right|^q\right], \\
Q_{m,j}^{(4)} &\coloneqq \mathbb{E}\left[\left\|\psi_j(\widetilde{X}_m^{(n)};\theta) - \psi_j(\widetilde{X}_m;\theta)\right\|^q \left|\phi_j(\widetilde{X}_m;\theta)\right|^q\right],
\end{aligned}$$

and applying the mean value theorem we obtain

$$\begin{aligned}
Q_{m,j}^{(1)} &\le \mathbb{E}\left[\left\|\psi_j(\widetilde{X}_m^{(n)};\theta)\right\|^q \left|\int_0^1 \phi_j'(\widetilde{X}_{m+1} + s(\widetilde{X}_{m+1}^{(n)} - \widetilde{X}_{m+1});\theta)\,\mathrm{d}s\right|^q \left|\widetilde{X}_{m+1}^{(n)} - \widetilde{X}_{m+1}\right|^q\right], \\
Q_{m,j}^{(2)} &\le \mathbb{E}\left[\left\|\psi_j(\widetilde{X}_m^{(n)};\theta)\right\|^q \left|\int_0^1 \phi_j'(\widetilde{X}_m + s(\widetilde{X}_m^{(n)} - \widetilde{X}_m);\theta)\,\mathrm{d}s\right|^q \left|\widetilde{X}_m^{(n)} - \widetilde{X}_m\right|^q\right], \\
Q_{m,j}^{(3)} &\le \mathbb{E}\left[\left\|\int_0^1 \psi_j'(\widetilde{X}_m + s(\widetilde{X}_m^{(n)} - \widetilde{X}_m);\theta)\,\mathrm{d}s\right\|^q \left|\widetilde{X}_m^{(n)} - \widetilde{X}_m\right|^q \left|\phi_j(\widetilde{X}_{m+1};\theta)\right|^q\right], \\
Q_{m,j}^{(4)} &\le \mathbb{E}\left[\left\|\int_0^1 \psi_j'(\widetilde{X}_m + s(\widetilde{X}_m^{(n)} - \widetilde{X}_m);\theta)\,\mathrm{d}s\right\|^q \left|\widetilde{X}_m^{(n)} - \widetilde{X}_m\right|^q \left|\phi_j(\widetilde{X}_m;\theta)\right|^q\right].
\end{aligned}$$

Then, employing the Hölder's inequality with exponents $4/(2-q), 4/(2-q), 2/q$ and since $\phi_j, \phi_j', \psi_j, \psi_j'$ are polynomially bounded by Assumption 2.5 and $\widetilde{X}_m^{(n)}, \widetilde{X}_m$ have bounded moments of any order by (4.1) we deduce

$$\mathbb{E}\left[\left\|G_{M,N}^J(\theta) - \mathbb{G}_M^J(\theta)\right\|^q\right] \le \frac{C}{M}\sum_{m=0}^{M-1}\sum_{j=1}^J \left(\mathbb{E}\left[(\widetilde{X}_m^{(n)} - \widetilde{X}_m)^2\right]^{\frac{q}{2}} + \mathbb{E}\left[(\widetilde{X}_{m+1}^{(n)} - \widetilde{X}_{m+1})^2\right]^{\frac{q}{2}}\right),$$

which due to (4.2) proves $(i)'$, which directly implies (i). Finally, the proofs of results (iv) and $(iv)'$ are similar to cases (i) and $(i)'$, respectively, and are omitted here. ■

Corollary 4.3. *Under Assumptions 2.2 and 2.5 it holds for all* $1 \le q < 2$

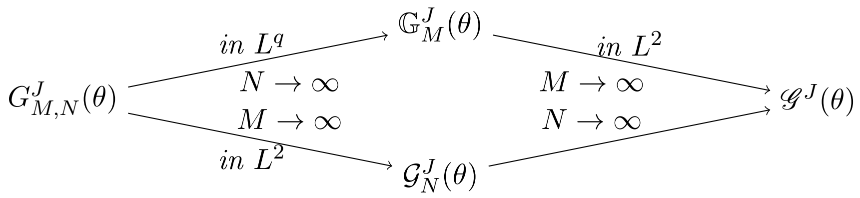
$$\lim_{M,N\to\infty} G_{M,N}^J(\theta) = \mathscr{G}^J(\theta), \quad \text{in } L^q.$$

Proof. Employing the triangular inequality we have

$$\mathbb{E}\left[\left\|G_{M,N}^J(\theta) - \mathscr{G}^J(\theta)\right\|^q\right] \le 2^{q-1}\left(\mathbb{E}\left[\left\|G_{M,N}^J(\theta) - \mathbb{G}_M^J(\theta)\right\|^q\right] + \mathbb{E}\left[\left\|\mathbb{G}_M^J(\theta) - \mathscr{G}^J(\theta)\right\|^q\right]\right),$$

where the right-hand side vanishes by $(i)'$ and (ii) in Proposition 4.2, yielding the desires result. ■

The limits considered in Proposition 4.2 are summarized schematically in the following graph



where $q \in [1, 2)$.

Remark 4.4. Notice that all the results in this section holds true also for the derivatives $\mathbb{H}^J_M(\theta)$, $\mathcal{H}^J_N(\theta)$, $\mathscr{H}^J(\theta)$ with respect to the parameter θ defined in (4.4). Since the arguments are analogous we omit the details here.

4.2. Zeros of the limits of the estimating function. The goal of this section is to show that the limits of the estimating functions previously defined admit zeros and to study their asymptotic limit. We already know by (4.3) that $\mathscr{G}^J(\theta_0) = 0$, where θ_0 is the true parameter. Then, in the following lemma we consider the zero of the function $\mathcal{G}^J_N(\theta)$ and its limit as $N \to \infty$.

Lemma 4.5. *Under Assumptions 2.2 and 2.5 and if* $\det(\mathscr{H}^J(\theta_0)) \neq 0$ *there exists* $N_0 > 0$ *such that for all* $N > N_0$ *there exists* $\vartheta^J_N \in \Theta$ *which solves the system* $\mathcal{G}^J_N(\theta) = 0$ *and satisfies* $\det(\mathcal{H}^J_N(\vartheta^J_N)) \neq 0$*. Moreover, there exists a constant* $C > 0$ *independent of* N *such that*

$$\left\| \vartheta^J_N - \theta_0 \right\| \leq \frac{C}{\sqrt{N}}. \tag{4.5}$$

Proof. We first remark that by (4.3) we have $\mathscr{G}^J(\theta_0) = 0$ and, without loss of generality, we can assume that $\det(\mathscr{H}^J(\theta_0)) > 0$. Let $\delta > 0$ sufficiently small, by point $(iv)'$ in Proposition 4.2 and Remark 4.4 we know that $\mathcal{H}^J_N(\theta)$ converges to $\mathscr{H}^J(\theta)$ uniformly in θ and therefore there exist $N_1 > 0$ and $\varepsilon > 0$ such that for all $N > N_1$ and for all $\theta \in B_\varepsilon(\theta_0)$

$$0 < \det(\mathscr{H}^J(\theta_0)) - \delta \leq \det(\mathcal{H}^J_N(\theta)) \leq \det(\mathscr{H}^J(\theta_0)) + \delta, \tag{4.6}$$

$$0 < \left\| \mathscr{H}^J(\theta_0)^{-1} \right\| - \delta \leq \left\| \mathcal{H}^J_N(\theta)^{-1} \right\| \leq \left\| \mathscr{H}^J(\theta_0)^{-1} \right\| + \delta. \tag{4.7}$$

Hence, due to equation (4.6) and applying the inverse function theorem we deduce the existence of $\eta > 0$ such that

$$B_\eta(\mathcal{G}^J_N(\theta_0)) \subseteq \mathcal{G}^J_N(B_\varepsilon(\theta_0)).$$

Notice that the radius $\eta > 0$ can be chosen independently of $N > N_1$. In fact, by the proof of [34, Theorem 2.3] and [28, Lemma 1.3] we observe that η is dependent on the radius ε of the ball $B_\varepsilon(\theta_0)$ and the quantity $\left\| \mathcal{H}^J_N(\theta_0)^{-1} \right\|$, which can be bounded independently of $N > N_1$ due to estimate (4.7). Moreover, since

$$\lim_{N \to \infty} \mathcal{G}^J_N(\theta_0) = \mathscr{G}^J(\theta_0) = 0,$$

then there exists $N_2 > 0$ such that for all $N > N_2$ we have $0 \in B_\eta(\mathcal{G}^J_N(\theta_0))$. Therefore, setting $N_0 = \max\{N_1, N_2\}$ for all $N > N_0$ there exists $\vartheta^J_N \in B_\varepsilon(\theta_0)$ such that $\mathcal{G}^J_N(\vartheta^J_N) = 0$, which proves the existence. Furthermore, equation (4.6) gives $\det(\mathcal{H}_N^J(\vartheta_N^J)) \neq 0$. It now remains to show estimate (4.5). Since the set $\overline{B_\varepsilon(\theta_0)}$ is compact, there exist $\widetilde{\vartheta}^J \in \overline{B_\varepsilon(\theta_0)}$ and a subsequence $\vartheta_{N_k}^J$ such that

$$\lim_{k\to\infty} \vartheta_{N_k}^J = \widetilde{\vartheta}^J.$$

By point $(iv)'$ in Proposition 4.2 the function $\mathcal{G}_N^J(\theta)$ converges to $\mathscr{G}^J(\theta)$ uniformly in θ, thus we have

$$0 = \lim_{k\to\infty} \mathcal{G}_{N_k}^J(\vartheta_{N_k}^J) = \lim_{k\to\infty} \left[\mathcal{G}_{N_k}^J(\vartheta_{N_k}^J) - \mathscr{G}^J(\vartheta_{N_k}^J) + \mathscr{G}^J(\vartheta_{N_k}^J)\right] = \mathscr{G}^J(\widetilde{\vartheta}^J),$$

which yields $\widetilde{\vartheta}^J = \theta_0$. This is guaranteed by the fact that ε can be previously chosen sufficiently small such that θ_0 is the only zero of the function $\mathscr{G}^J(\theta)$ in $B_\varepsilon(\theta_0)$. Since θ_0 is the unique limit point for the subsequence $\vartheta_{N_k}^J$, it follows that the whole sequence converges. Then, applying the mean value theorem we obtain

$$\mathscr{G}^J(\vartheta_N^J) - \mathcal{G}_N^J(\vartheta_N^J) = \mathscr{G}^J(\vartheta_N^J) - \mathscr{G}^J(\theta_0) = \left(\int_0^1 \mathscr{H}^J(\theta_0 + t(\vartheta_N^J - \theta_0))\,\mathrm{d}t\right)(\vartheta_N^J - \theta_0),$$

which implies

$$\left\|\vartheta_N^J - \theta_0\right\| \leq \left\|\left(\int_0^1 \mathscr{H}^J(\theta_0 + t(\vartheta_N^J - \theta_0))\,\mathrm{d}t\right)^{-1}\right\| \left\|\mathscr{G}^J(\vartheta_N^J) - \mathcal{G}_N^J(\vartheta_N^J)\right\|.$$

Since ϑ_N^J converges to θ_0 as N goes to infinity, then

$$\lim_{N\to\infty} \left\|\left(\int_0^1 \mathscr{H}^J(\theta_0 + t(\vartheta_N^J - \theta_0))\,\mathrm{d}t\right)^{-1}\right\| = \left\|\mathscr{H}^J(\theta_0)^{-1}\right\|,$$

where the right-hand side is well defined because $\det(\mathscr{H}^J(\theta_0)) \neq 0$. Therefore, if N is sufficiently big there exists a constant $C > 0$ independent of N such that

$$\left\|\left(\int_0^1 \mathscr{H}^J(\theta_0 + t(\vartheta_N^J - \theta_0))\,\mathrm{d}t\right)^{-1}\right\| \leq C,$$

which together with point $(iv)'$ in Proposition 4.2 yields estimate (4.5) and concludes the proof. ∎

In the next lemma we study the zero of the random function $\mathbb{G}_M^J(\theta)$ and its limit as $M \to \infty$. This result is almost the same as [27, Theorem 4.3].

Lemma 4.6. *Let the assumptions of Lemma 4.5 hold. Then, an estimator $\widehat{\vartheta}_M^J$, which solves the equation $\mathbb{G}_M^J(\theta) = 0$ and is such that $\det(\mathbb{H}_M^J(\widehat{\vartheta}_M^J)) \neq 0$, exists with a probability tending to one as $M \to \infty$. Moreover,*

$$\lim_{M\to\infty} \widehat{\vartheta}_M^J = \theta_0, \qquad \textit{in probability,}$$

and

$$\lim_{M\to\infty} \sqrt{M} \left(\widehat{\vartheta}_M^J - \theta_0 \right) = \Lambda^J \sim \mathcal{N}(\mathbf{0}, \Gamma_0^J), \qquad \textit{in distribution,}$$

where Γ_0^J *is defined in* (2.18).

Proof. The existence of the estimator $\widehat{\vartheta}_M^J$ which solves the equation $\mathbb{G}_M^J(\theta) = 0$ with a probability tending to one as $M \to \infty$ and its asymptotic unbiasedness and normality is given by [27, Theorem 4.3], whose prove can be found in [6, Theorem 3.2] and is based on [3, Theorem A.1]. Moreover, by the last line of the proof of [6, Theorem 3.2] or by (A.5) in [27, Theorem 4.3] we have

$$\lim_{M\to\infty} \mathbb{H}_M^J(\widehat{\vartheta}_M^J) = \mathscr{H}^J(\theta_0), \qquad \text{in probability,} \tag{4.8}$$

where $\det(\mathscr{H}^J(\theta_0)) \neq 0$ by assumption. Hence, there exists $\delta > 0$ such that if

$$\left\| \mathbb{H}_M^J(\widehat{\vartheta}_M^J) - \mathscr{H}^J(\theta_0) \right\| \leq \delta,$$

then $\det(\mathbb{H}_M^J(\widehat{\vartheta}_M^J))) \neq 0$. Moreover, for M large enough it holds

$$\mathbb{P}\left(\left\| \mathbb{H}_M^J(\widehat{\vartheta}_M^J) - \mathscr{H}^J(\theta_0) \right\| \leq \delta \right) \geq 1 - \varepsilon_M,$$

where $\varepsilon_M \to 0$ as $M \to \infty$. Let us now define the events

$$A_M := \left\{ \exists\, \widehat{\vartheta}_M^J \colon \mathbb{G}_M^J(\widehat{\vartheta}_M^J) \right\} \qquad \text{and} \qquad B_M := \left\{ \left\| \mathbb{H}_M^J(\widehat{\vartheta}_M^J) - \mathscr{H}^J(\theta_0) \right\| \leq \delta \right\},$$

and notice that by the first part of the proof we have $\mathbb{P}(A_M) = p_M$ where $p_M \to 1$ as $M \to \infty$. Then, using basic properties of probability measures we obtain

$$\mathbb{P}\left(A_M \cap \{\det(\mathbb{H}_M^J(\widehat{\vartheta}_M^J)) \neq 0\} \right) \geq \mathbb{P}\left(A_M \cap B_M\right) \geq \mathbb{P}(A_M) + \mathbb{P}(B_M) - 1 \geq p_M - \varepsilon_M,$$

where the last term tends to one as $M \to \infty$, and which gives the desired result. ∎

We now consider the zero of the actual estimating function $G_{M,N}^J(\theta)$ and we first analyze its limit as $M \to \infty$.

Lemma 4.7. *Let the assumptions of Theorem 2.9 hold. Then, there exists $N_0 > 0$ such that for all $N > N_0$ an estimator $\widehat{\theta}_{M,N}^J$, which solves the system $G_{M,N}^J(\theta) = 0$, exists with a probability tending to one as M goes to infinity. Moreover, there exist ϑ_N^J solving $\mathcal{G}_N^J(\theta) = 0$ such that*

$$\lim_{M\to\infty} \widehat{\theta}_{M,N}^J = \vartheta_N^J, \qquad \textit{in probability,}$$

and

$$\lim_{M\to\infty} \sqrt{M} \left(\widehat{\theta}_{M,N}^J - \vartheta_N^J \right) = \Lambda_N^J \sim \mathcal{N}(\mathbf{0}, \Gamma_N^J), \qquad \textit{in distribution,}$$

where Γ_N^J is a positive definite covariance matrix such that $\lim_{N\to\infty}\Gamma_N^J = \Gamma_0^J$ where Γ_0^J is defined in (2.18)*.*

Proof. First, by Lemma 4.5 there exists $N_0 > 0$ such that for all $N > N_0$ there exists ϑ_N^J such that

$$\mathcal{G}_N^J(\vartheta_N^J) = 0 \qquad \text{and} \qquad \det(\mathcal{H}_N^J(\vartheta_N^J)) \neq 0.$$

Then, the results are equivalent to Lemma 4.6 and therefore the argument follows the same steps of its proof, which is given in detail in [6, Theorem 3.2] and is based on [3, Theorem A.1]. Finally, the convergence of the covariance matrix Γ_N^J is implied by (4.2). ∎

We then study the limit of the zero of $G_{M,N}^J(\theta)$ as $N \to \infty$.

Lemma 4.8. *Let the assumptions of Lemma 4.7 hold and let $M \ll N$. Then, the estimator $\widehat{\theta}_{M,N}^J$ satisfies for some $\widehat{\vartheta}_M^J$ solving $\mathbb{G}_M^J(\theta) = 0$ and for a constant $C > 0$ independent of M and N*

$$\mathbb{E}\left[\left\|\widehat{\theta}_{M,N}^J - \widehat{\vartheta}_M^J\right\|\right] \leq C\sqrt{\frac{M}{N}}.$$

Proof. The existence of the estimators $\widehat{\vartheta}_M^J$, such that $\mathbb{G}_M^J(\widehat{\vartheta}_M^J) = 0$ and $\det(\mathbb{H}_M^J(\widehat{\vartheta}_M^J)) \neq 0$, and $\widehat{\theta}_{M,N}^J$, such that $G_{M,N}^J(\widehat{\theta}_{M,N}^J) = 0$, with a probability tending to one as M goes to infinity is guaranteed by Lemmas 4.6 and 4.7, respectively. Then, all the following events are considered as conditioned on the existence of $\widehat{\vartheta}_M^J$ and $\widehat{\theta}_{M,N}^J$ and the fact that $\det(\mathbb{H}_M^J(\widehat{\vartheta}_M^J)) \neq 0$. Let us now define the function $f\colon \mathbb{R}^p \times \mathbb{R}^{M+1} \to \mathbb{R}^p$ as

$$f(\theta, x) = \frac{1}{M}\sum_{m=0}^{M-1}\sum_{j=1}^{J} g_j(x_m, x_{m+1}; \theta),$$

where x_m denotes the m-th component of the vector $x \in \mathbb{R}^{m+1}$, and the vectors $\mathbb{X}^{(n)}$ and $\mathbb{X}$ whose m-th components for $m = 0, \dots, M$ are given by

$$\mathbb{X}_m^{(n)} = \widetilde{X}_m^{(n)} \qquad \text{and} \qquad \mathbb{X}_m = \widetilde{X}_m,$$

where $\{\widetilde{X}_m^{(n)}\}_{m=0}^M$ is the set of observations and $\{\widetilde{X}_m\}_{m=0}^M$ are the corresponding realizations of the mean field limit. Notice that $f \in C^1(\Theta \times \mathbb{R}^{M+1})$ due to Assumption 2.5 and Remark 2.6 and by definition we have

$$f(\widehat{\vartheta}_M^J, \mathbb{X}) = 0 \qquad \text{and} \qquad \det\left(\frac{\partial f}{\partial \theta}(\widehat{\vartheta}_M^J, \mathbb{X})\right) \neq 0.$$

Therefore, applying the implicit function theorem there exist $\varepsilon, \delta > 0$ and a continuously differentiable function $F\colon B_\varepsilon(\mathbb{X}) \to B_\delta(\widehat{\vartheta}_M^J)$ such that $f(F(x), x) = 0$ for all $x \in B_\varepsilon(\mathbb{X})$. Hence, if $\mathbb{X}^{(n)}$ is close enough to $\mathbb{X}$ then there must be one $\widehat{\theta}_{M,N}^J \in B_\delta(\widehat{\vartheta}_M^J)$ such that $F(\mathbb{X}^{(n)}) = \widehat{\theta}_{M,N}^J$.

Then, employing Jensen's inequality and by estimate (4.2) we have

$$\mathbb{E}\left[\left\|\mathbb{X}^{(n)}-\mathbb{X}\right\|\right]=\mathbb{E}\left[\left(\sum_{m=0}^{M}\left|\widetilde{X}_m^{(n)}-\widetilde{X}_m\right|^2\right)^{1/2}\right]\leq\left(\sum_{m=0}^{M}\mathbb{E}\left[\left|\widetilde{X}_m^{(n)}-\widetilde{X}_m\right|^2\right]\right)^{1/2}\leq C\sqrt{\frac{M}{N}},$$

where the constant C is independent of M and N. Therefore, letting $\varepsilon>0$ and applying Markov's inequality we obtain

$$\mathbb{P}\left(\left\|\mathbb{X}^{(n)}-\mathbb{X}\right\|\geq\varepsilon\right)\leq\frac{1}{\varepsilon}\mathbb{E}\left[\left\|\mathbb{X}^{(n)}-\mathbb{X}\right\|\right]\leq\frac{C}{\varepsilon}\sqrt{\frac{M}{N}}. \tag{4.9}$$

Defining the event $A=\{\left\|\mathbb{X}^{(n)}-\mathbb{X}\right\|<\varepsilon\}$ and using the law of total expectation conditioning on A we deduce

$$\mathbb{E}\left[\left\|\widehat{\theta}_{M,N}^J-\widehat{\vartheta}_M^J\right\|\right]=\mathbb{E}\left[\left\|\widehat{\theta}_{M,N}^J-\widehat{\vartheta}_M^J\right\||A\right]\mathbb{P}(A)+\mathbb{E}\left[\left\|\widehat{\theta}_{M,N}^J-\widehat{\vartheta}_M^J\right\||A^{\mathsf{C}}\right]\mathbb{P}(A^{\mathsf{C}}),$$

which since $\widehat{\theta}_{M,N}^J,\widehat{\vartheta}_M^J\in\Theta$, a compact set, and due to estimate (4.9) implies

$$\mathbb{E}\left[\left\|\widehat{\theta}_{M,N}^J-\widehat{\vartheta}_M^J\right\|\right]\leq\mathbb{E}\left[\left\|\widehat{\theta}_{M,N}^J-\widehat{\vartheta}_M^J\right\||A\right]+C\sqrt{\frac{M}{N}}. \tag{4.10}$$

It now remains to study the first term in the right-hand side. Applying the mean value theorem we obtain

$$\begin{aligned}\mathbb{G}_M^J(\widehat{\theta}_{M,N}^J)-G_{M,N}^J(\widehat{\theta}_{M,N}^J)&=\mathbb{G}_M^J(\widehat{\theta}_{M,N}^J)-\mathbb{G}_M^J(\widehat{\vartheta}_M^J)\\&=\left(\int_0^1\mathbb{H}_M^J(\widehat{\vartheta}_M^J+t(\widehat{\theta}_{M,N}^J-\widehat{\vartheta}_M^J))\,\mathrm{d}t\right)(\widehat{\theta}_{M,N}^J-\widehat{\vartheta}_M^J),\end{aligned}$$

which implies

$$\left\|\widehat{\theta}_{M,N}^J-\widehat{\vartheta}_M^J\right\|\leq\left\|\left(\int_0^1\mathbb{H}_M^J(\widehat{\vartheta}_M^J+t(\widehat{\theta}_{M,N}^J-\widehat{\vartheta}_M^J))\,\mathrm{d}t\right)^{-1}\right\|\left\|\mathbb{G}_M^J(\widehat{\theta}_{M,N}^J)-G_{M,N}^J(\widehat{\theta}_{M,N}^J)\right\|.$$

Using Hölder's inequality with exponents $q\in(1,2)$ and its conjugate q' such that $1/q+1/q'=1$ we have

$$\mathbb{E}\left[\left\|\widehat{\theta}_{M,N}^J-\widehat{\vartheta}_M^J\right\||A\right]\leq Q\,\mathbb{E}\left[\left\|\mathbb{G}_M^J(\widehat{\theta}_{M,N}^J)-G_{M,N}^J(\widehat{\theta}_{M,N}^J)\right\|^q|A\right]^{1/q}, \tag{4.11}$$

where

$$Q=\mathbb{E}\left[\left\|\left(\int_0^1\mathbb{H}_M^J(\widehat{\vartheta}_M^J+t(\widehat{\theta}_{M,N}^J-\widehat{\vartheta}_M^J))\,\mathrm{d}t\right)^{-1}\right\|^{q'}|A\right]^{1/q'}.$$

Employing the inequality $\mathbb{E}[Y|A] \le \mathbb{E}[Y]/\mathbb{P}(A)$, which holds for any positive random variable Y, point $(i)'$ in Proposition 4.2 and estimate (4.9), the second term in the right-hand side can be bounded by

$$\mathbb{E}\left[\left\|\mathbb{G}_M^J(\widehat{\theta}_{M,N}^J) - G_{M,N}^J(\widehat{\theta}_{M,N}^J)\right\|^q |A\right]^{1/q} \le \frac{C}{\sqrt{N}}\left(\frac{1}{1-C\sqrt{\frac{M}{N}}}\right)^{1/q} \le \frac{C}{\sqrt{N}}, \tag{4.12}$$

where the last inequality is justified by the fact that $M \ll N$ and by changing the value of the constant C. We now have to bound the first term Q in the right-hand side of equation (4.11). Employing the inequality $\left\|M^{-1}\right\| \le \|M\|^{p-1} / |\det(M)|$, which holds for any square nonsingular matrix $M \in \mathbb{R}^{p\times p}$, we have

$$Q \le \mathbb{E}\left[\frac{\left\|\int_0^1 \mathbb{H}_M^J(\widehat{\vartheta}_M^J + t(\widehat{\theta}_{M,N}^J - \widehat{\vartheta}_M^J))\,\mathrm{d}t\right\|^{q'(p-1)}}{\left|\det\left(\int_0^1 \mathbb{H}_M^J(\widehat{\vartheta}_M^J + t(\widehat{\theta}_{M,N}^J - \widehat{\vartheta}_M^J))\,\mathrm{d}t\right)\right|^{q'}}|A\right].$$

Since we are conditioning on the event A, by the first part of the proof, we know that $\left\|\widehat{\theta}_{M,N}^J - \widehat{\vartheta}_M^J\right\| \le \delta$ and, by taking ε sufficiently small, we can always find δ small enough, but still finite, such that the absolute value of the determinant in the denominator is lower bounded by a constant independent of M and N because $\det(\mathbb{H}_M^J(\widehat{\vartheta}_M^J)) \neq 0$ and by (4.8) it converges in probability to $\det(\mathscr{H}^J(\theta_0))$, which is invertible. Hence, applying Jensen's inequality we obtain

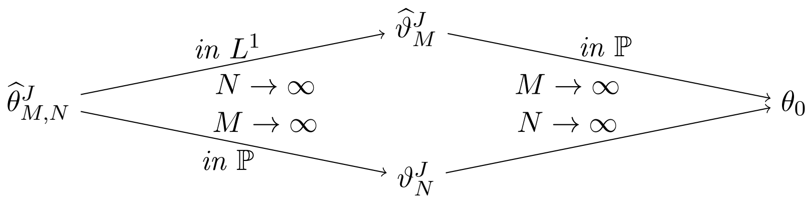
$$\begin{aligned} Q &\le C\,\mathbb{E}\left[\left\|\int_0^1 \mathbb{H}_M^J(\widehat{\vartheta}_M^J + t(\widehat{\theta}_{M,N}^J - \widehat{\vartheta}_M^J))\,\mathrm{d}t\right\|^{q'(p-1)} |A\right] \\ &\le C\,\mathbb{E}\left[\int_0^1 \left\|\mathbb{H}_M^J(\widehat{\vartheta}_M^J + t(\widehat{\theta}_{M,N}^J - \widehat{\vartheta}_M^J))\right\|^{q'(p-1)} \mathrm{d}t|A\right], \end{aligned}$$

which due to Lemma 4.1, Remark 4.4, the property $\mathbb{E}[Y|A] \le \mathbb{E}[Y]/\mathbb{P}(A)$, which holds for any positive random variable Y, and estimate (4.9) yields

$$Q \le \frac{C}{\mathbb{P}(A)}\int_0^1 \mathbb{E}\left[\left\|\mathbb{H}_M^J(\widehat{\vartheta}_M^J + t(\widehat{\theta}_{M,N}^J - \widehat{\vartheta}_M^J))\right\|^{q'(p-1)}\right] \mathrm{d}t \le C,$$

which together with equations (4.10), (4.11) and (4.12) gives the desired result. ■

The results of this section are summarized in the following graph



where $\mathbb{P}$ stands for convergence in probability.

Remark 4.9. All the previous results only prove the existence of such estimators with high probability and do not guarantee their uniqueness. However, as we will see in the next section, any of these estimators converge to the exact value of the unknown.

4.3. Proof of the main theorems. In this section we finally present the proofs of the main results of this work, i.e., Theorems 2.9 to 2.11.

Proof of Theorem 2.9. First, by Lemma 4.7 we deduce the existence of $N_0 > 0$ such that for all $N > N_0$ the estimator $\widehat{\theta}^J_{M,N}$ exists with a probability tending to one as M goes to infinity. Then, we prove separately equations (2.12), (2.13) and (2.14).
Proof of (2.12)**.** By Lemmas 4.5 and 4.7 we have

$$\lim_{N\to\infty}\lim_{M\to\infty}\widehat{\theta}^J_{M,N} = \lim_{N\to\infty}\vartheta^J_N = \theta_0, \qquad \text{in probability,}$$

which proves (2.12).
Proof of (2.13)**.** By Lemma 4.8 the estimator $\widehat{\theta}^J_{M,N}$ converges to $\widehat{\vartheta}^J_M$ in L^1 as N goes to infinity and hence in probability. Therefore, applying Lemma 4.6 we obtain

$$\lim_{M\to\infty}\lim_{N\to\infty}\widehat{\theta}^J_{M,N} = \lim_{M\to\infty}\widehat{\vartheta}^J_M = \theta_0, \qquad \text{in probability,}$$

which shows (2.13).
Proof of (2.14)**.** We introduce the following decomposition

$$\widehat{\theta}^J_{M,N} - \theta_0 = (\widehat{\theta}^J_{M,N} - \widehat{\vartheta}^J_M) + (\widehat{\vartheta}^J_M - \theta_0) =: Q_1 + Q_2,$$

where $\widehat{\vartheta}^J_M$ is defined in Lemma 4.6 and due to Lemma 4.8 the first quantity satisfies

$$\mathbb{E}\left[|Q_1|\right] \le C\sqrt{\frac{M}{N}}, \tag{4.13}$$

with the constant C independent of M and N. Therefore, since $M = o(N)$, estimate (4.13) together with Lemma 4.6 and the fact that convergence in L^1 implies convergence in probability gives the desired result (2.14) and ends the proof. ■

Proof of Theorem 2.10. The existence of the estimator $\widehat{\theta}^J_{M,N}$ is given by Theorem 2.9. Then, we prove separately equations (2.15), (2.16) and (2.17).
Proof of (2.15)**.** Let ϑ_N be defined in Lemma 4.5. Using basic properties of probability measures we have

$$\begin{aligned}\mathbb{P}\left(\Xi^J_{M,N} > K_\varepsilon\right) &= \mathbb{P}\left(\left\|\widehat{\theta}^J_{M,N} - \theta_0\right\| > \left(\frac{1}{\sqrt{M}} + \frac{1}{\sqrt{N}}\right)K_\varepsilon\right) \\ &\le \mathbb{P}\left(\left\|\widehat{\theta}^J_{M,N} - \vartheta_N\right\| + \|\vartheta_N - \theta_0\| > \left(\frac{1}{\sqrt{M}} + \frac{1}{\sqrt{N}}\right)K_\varepsilon\right),\end{aligned} \tag{4.14}$$

which implies

$$\begin{aligned}\mathbb{P}\left(\Xi^J_{M,N} > K_\varepsilon\right) &\leq \mathbb{P}\left(\left\|\widehat{\theta}^J_{M,N} - \vartheta_N\right\| > \left(\frac{1}{\sqrt{M}} + \frac{1}{\sqrt{N}}\right)\frac{K_\varepsilon}{2}\right) \\ &\quad + \mathbb{P}\left(\|\vartheta_N - \theta_0\| > \left(\frac{1}{\sqrt{M}} + \frac{1}{\sqrt{N}}\right)\frac{K_\varepsilon}{2}\right) \\ &\leq \mathbb{P}\left(\sqrt{M}\left\|\widehat{\theta}^J_{M,N} - \vartheta_N\right\| > \frac{K_\varepsilon}{2}\right) + \mathbb{P}\left(\|\vartheta_N - \theta_0\| > \frac{K_\varepsilon}{2\sqrt{N}}\right),\end{aligned}$$

and we now study the two terms in the right-hand side separately. First, letting M and N go to infinity by Lemma 4.7 we obtain

$$\lim_{N\to\infty}\lim_{M\to\infty}\mathbb{P}\left(\sqrt{M}\left\|\widehat{\theta}^J_{M,N} - \vartheta_N\right\| > \frac{K_\varepsilon}{2}\right) = \mathbb{P}\left(\left\|\Lambda^J\right\| > \frac{K_\varepsilon}{2}\right),$$

where the right-hand side can be made arbitrarily small by taking $K_\varepsilon > 0$ sufficiently big. Moreover, we have

$$\mathbb{P}\left(\|\vartheta_N - \theta_0\| > \frac{K_\varepsilon}{2\sqrt{N}}\right) = \mathbb{1}_{\left\{\|\vartheta_N - \theta_0\| > \frac{K_\varepsilon}{2\sqrt{N}}\right\}},$$

where the right-hand side is identically equal to zero if we set $K_\varepsilon > 2C$, where the constant C is given by Lemma 4.5. Hence, for all $\varepsilon > 0$ we can take $K_\varepsilon > 0$ sufficiently big such that

$$\lim_{N\to\infty}\lim_{M\to\infty}\mathbb{P}\left(\Xi^J_{M,N} > K_\varepsilon\right) < \varepsilon,$$

which proves (2.15).

Proof of (2.16). Let $\widehat{\vartheta}_M$ be defined in Lemma 4.6. Repeating a procedure similar to (4.14) and applying Markov's inequality we get

$$\begin{aligned}\mathbb{P}\left(\Xi^J_{M,N} > K_\varepsilon\right) &\leq \mathbb{P}\left(\left\|\widehat{\theta}^J_{M,N} - \widehat{\vartheta}_M\right\| > \left(\frac{1}{\sqrt{M}} + \frac{1}{\sqrt{N}}\right)\frac{K_\varepsilon}{2}\right) + \mathbb{P}\left(\sqrt{M}\left\|\widehat{\vartheta}_M - \theta_0\right\| > \frac{K_\varepsilon}{2}\right) \\ &\leq \frac{2\sqrt{MN}}{K_\varepsilon(\sqrt{M} + \sqrt{N})}\mathbb{E}\left[\left\|\widehat{\theta}^J_{M,N} - \widehat{\vartheta}_M\right\|\right] + \mathbb{P}\left(\sqrt{M}\left\|\widehat{\vartheta}_M - \theta_0\right\| > \frac{K_\varepsilon}{2}\right),\end{aligned}$$

and we now study the two terms in the right-hand side separately. First, by Lemma 4.6 we have

$$\lim_{M\to\infty}\mathbb{P}\left(\sqrt{M}\left\|\widehat{\vartheta}_M - \theta_0\right\| > \frac{K_\varepsilon}{2}\right) = \mathbb{P}\left(\left\|\Lambda^J\right\| > \frac{K_\varepsilon}{2}\right),$$

where the right-hand side can be made arbitrarily small by taking $K_\varepsilon > 0$ sufficiently big. Moreover, by Lemma 4.8 we have

$$\frac{2\sqrt{MN}}{K_\varepsilon(\sqrt{M} + \sqrt{N})}\mathbb{E}\left[\left\|\widehat{\theta}^J_{M,N} - \widehat{\vartheta}_M\right\|\right] \leq \frac{2CM}{K_\varepsilon(\sqrt{M} + \sqrt{N})}, \tag{4.15}$$

where the constant C is independent of M and N. Hence, for all $\varepsilon > 0$ we can take $K_\varepsilon > 0$ sufficiently big such that

$$\lim_{M\to\infty} \lim_{N\to\infty} \mathbb{P}\left(\Xi^J_{M,N} > K_\varepsilon\right) < \varepsilon,$$

which shows (2.16).

Proof of (2.17)**.** Equation (2.17) is obtained following verbatim the proof of (2.16) in the previous step and using the fact that $M = o(\sqrt{N})$ to show that the right-hand side in equation (4.15) vanishes. ■

Proof of Theorem 2.11. The existence of the estimator $\widehat{\theta}^J_{M,N}$ is given by Theorem 2.9. Then, let us introduce the following decomposition

$$\sqrt{M}\left(\widehat{\theta}^J_{M,N} - \theta_0\right) = \sqrt{M}\left(\widehat{\theta}^J_{M,N} - \widehat{\vartheta}^J_M\right) + \sqrt{M}\left(\widehat{\vartheta}^J_M - \theta_0\right),$$

where $\widehat{\vartheta}^J_M$ is defined in Lemma 4.6. We now study the two terms in the right-hand side separately. By Lemma 4.8 we have

$$\sqrt{M}\,\mathbb{E}\left[\left\|\widehat{\theta}^J_{M,N} - \widehat{\vartheta}^J_M\right\|\right] \le C\frac{M}{\sqrt{N}},$$

where the constant C is independent of M and N, hence since $M = o(\sqrt{N})$ by hypothesis we obtain

$$\lim_{M,N\to\infty} \sqrt{M}\left(\widehat{\theta}^J_{M,N} - \widehat{\vartheta}^J_M\right) = 0, \qquad \text{in probability.} \tag{4.16}$$

Moreover, by Lemma 4.6 we know that

$$\lim_{M\to\infty} \sqrt{M}\left(\widehat{\vartheta}^J_M - \theta_0\right) = \Lambda^J \sim \mathcal{N}(0, \Gamma^J_0), \qquad \text{in distribution,} \tag{4.17}$$

where the covariance matrix Γ^J_0 is defined in (2.18). Finally, limits (4.16) and (4.17) together with Slutsky's theorem imply the desired result. ■

5. Conclusion.

In this work we considered inference problems for large systems of exchangeable interacting particles. When the number of particles is large, then the path of a single particle is well approximated by its mean field limit. The limiting mean field SDE is on the one hand more complex because it is a nonlinear SDE (in the sense of McKean), but on the other hand more tractable from a computational viewpoint as it reduces an N-dimensional SDE to a one dimensional one. Our aim was to infer unknown parameters of the dynamics, in particular of the confining and interaction potentials, from a set of discrete observations of a single particle. We propose a novel estimator which is obtained by computing the zero of a martingale estimating function based on the eigenvalues and the eigenfunctions of the generator of the mean field limit, linearized around the (unique) invariant measure of the mean field dynamics. We showed both theoretically and numerically the asymptotic unbiasedness and normality of our estimator in the limit of infinite data and particles, providing also a rate of convergence towards the true value of the unknown parameter. In particular, we observed that these properties hold true if the number of particles is much larger than the number of observations. Even though our theoretical results require uniqueness of the steady state for the mean field dynamics, our numerical experiments suggest that our method works well even when phase transitions are present, i.e., when there are more than one stationary states. Moreover, we compared our estimator with the maximum likelihood estimator, demonstrating that our approach is more robust with respect to small values of the sampling rate. We believe, therefore, that the inference methodology proposed and analyzed in this paper can be very efficient when learning parameters in mean field SDE models from data.

The work presented in this paper can be extended in several interesting directions. First, the main limitation of our methodology is the fact that in order to construct the martingale estimating function we have to know the functional form of the invariant measure of the mean field SDE, possibly parameterized in terms of a finite number of moments. There are many interesting examples of mean field PDEs where the self-consistency equation cannot be solved analytically or, at least, its solution depends on the unknown parameters in the model. Therefore, it would be interesting to lift this assumption by first learning the invariant measure from data and then applying our martingale eigenfunction estimator approach. This leads naturally to our second objective, namely the extension of our methodology to a nonparametric setting, i.e., when the functional form of the confining and interaction potentials are unknown. Thirdly, we want to obtain more detailed information on the computational complexity of the proposed algorithm, in particular when more eigenfunctions are needed for our martingale estimator and when we are in higher dimensions in space. We will return to these problems in future work.

Acknowledgements. The work of GAP was partially funded by the EPSRC, grant number EP/P031587/1, and by JPMorgan Chase & Co. AZ is partially supported by the Swiss National Science Foundation, under grant No. 200020_172710. AZ is grateful to Massimo Sorella for fruitful discussions.

REFERENCES

[1] A. Abdulle, G. Garegnani, G. Pavliotis, A. Stuart, and A. Zanoni, *Drift estimation of multiscale diffusions based on filtered data*, Found. Comput. Math., (2021), pp. 1–52.

[2] A. Abdulle, G. A. Pavliotis, and A. Zanoni, *Eigenfunction martingale estimating functions and filtered data for drift estimation of discretely observed multiscale diffusions*. Preprint arXiv:2104.10587, 2021.

[3] O. E. Barndorff-Nielsen and M. Sørensen, *A review of some aspects of asymptotic likelihood theory for stochastic processes*, International Statistical Review / Revue Internationale de Statistique, 62 (1994), pp. 133–165.

[4] S. Benachour, B. Roynette, D. Talay, and P. Vallois, *Nonlinear self-stabilizing processes. I. Existence, invariant probability, propagation of chaos*, Stochastic Process. Appl., 75 (1998), pp. 173–201, https://doi.org/10.1016/S0304-4149(98)00018-0.

[5] S. Benachour, B. Roynette, and P. Vallois, *Nonlinear self-stabilizing processes. II. Convergence to invariant probability*, Stochastic Process. Appl., 75 (1998), pp. 203–224, https://doi.org/10.1016/S0304-4149(98)00019-2.

[6] B. Bibby and M. Sø rensen, *Martingale estimation functions for discretely observed diffusion processes*, Bernoulli, 1 (1995), pp. 17–39, https://doi.org/10.2307/3318679.

[7] J. Binney and S. Tremaine, *Galactic Dynamics*, Princeton University Press, Princeton, second ed., 2008.
[8] J. P. N. Bishwal, *Estimation in interacting diffusions: continuous and discrete sampling*, Appl. Math. (Irvine), 2 (2011), pp. 1154–1158, https://doi.org/10.4236/am.2011.29160.
[9] J. A. Carrillo, R. S. Gvalani, G. A. Pavliotis, and A. Schlichting, *Long-time behaviour and phase transitions for the Mckean-Vlasov equation on the torus*, Arch. Ration. Mech. Anal., 235 (2020), pp. 635–690, https://doi.org/10.1007/s00205-019-01430-4.
[10] X. Chen, *Maximum likelihood estimation of potential energy in interacting particle systems from single-trajectory data*, Electron. Commun. Probab., 26 (2021), pp. Paper No. 45, 13, https://doi.org/10.1214/21-ecp416.
[11] D. A. Dawson, *Critical dynamics and fluctuations for a mean-field model of cooperative behavior*, J. Statist. Phys., 31 (1983), pp. 29–85, https://doi.org/10.1007/BF01010922.
[12] F. Delarue and A. Tse, *Uniform in time weak propagation of chaos on the torus*, 2021, https://arxiv.org/abs/2104.14973.
[13] M. G. Delgadino, R. S. Gvalani, and G. A. Pavliotis, *On the Diffusive-Mean Field Limit for Weakly Interacting Diffusions Exhibiting Phase Transitions*, Arch. Ration. Mech. Anal., 241 (2021), pp. 91–148, https://doi.org/10.1007/s00205-021-01648-1.
[14] L. Ellam, M. Girolami, G. A. Pavliotis, and A. Wilson, *Stochastic modelling of urban structure*, Proc. A., 474 (2018), pp. 20170700, 20, https://doi.org/10.1098/rspa.2017.0700.
[15] T. D. Frank, *Nonlinear Fokker-Planck equations*, Springer Series in Synergetics, Springer-Verlag, Berlin, 2005. Fundamentals and applications.
[16] A. Ganz Bustos, *Approximations des distributions d'équilibre de certains systèmes stochastiques avec interactions McKean-Vlasov*, PhD thesis, Nice, 2008.
[17] G. Garegnani and A. Zanoni, *Robust estimation of effective diffusions from multiscale data.* Preprint arXiv:2109.03132, 2021.
[18] J. Garnier, G. Papanicolaou, and T.-W. Yang, *Consensus convergence with stochastic effects*, Vietnam J. Math., 45 (2017), pp. 51–75, https://doi.org/10.1007/s10013-016-0190-2.
[19] J. Gärtner, *On the McKean-Vlasov limit for interacting diffusions*, Math. Nachr., 137 (1988), pp. 197–248, https://doi.org/10.1002/mana.19881370116.
[20] K. Giesecke, G. Schwenkler, and J. A. Sirignano, *Inference for large financial systems*, Math. Finance, 30 (2020), pp. 3–46, https://doi.org/10.1111/mafi.12222.
[21] B. D. Goddard, B. Gooding, H. Short, and G. A. Pavliotis, *Noisy bounded confidence models for opinion dynamics: the effect of boundary conditions on phase transitions*, IMA Journal of Applied Mathematics, (2021), https://doi.org/10.1093/imamat/hxab044, https://arxiv.org/abs/https://academic.oup.com/imamat/advance-article-pdf/doi/10.1093/imamat/hxab044/41147076/hxab044.pdf.
[22] F. Golse, *On the dynamics of large particle systems in the mean field limit*, in Macroscopic and large scale phenomena: coarse graining, mean field limits and ergodicity, vol. 3 of Lect. Notes Appl. Math. Mech., Springer, [Cham], 2016, pp. 1–144, https://doi.org/10.1007/978-3-319-26883-5_1.
[23] S. N. Gomes and G. A. Pavliotis, *Mean field limits for interacting diffusions in a two-scale potential*, J. Nonlinear Sci., 28 (2018), pp. 905–941, https://doi.org/10.1007/s00332-017-9433-y.
[24] S. N. Gomes, A. M. Stuart, and M.-T. Wolfram, *Parameter estimation for macroscopic pedestrian dynamics models from microscopic data*, SIAM J. Appl. Math., 79 (2019), pp. 1475–1500, https://doi.org/10.1137/18M1215980.
[25] L. P. Hansen, J. A. Scheinkman, and N. Touzi, *Spectral methods for identifying scalar diffusions*, J. Econometrics, 86 (1998), pp. 1–32, https://doi.org/10.1016/S0304-4076(97)00107-3.
[26] R. A. Kasonga, *Maximum likelihood theory for large interacting systems*, SIAM J. Appl. Math., 50 (1990), pp. 865–875, https://doi.org/10.1137/0150050.
[27] M. Kessler and M. Sø rensen, *Estimating equations based on eigenfunctions for a discretely observed diffusion process*, Bernoulli, 5 (1999), pp. 299–314, https://doi.org/10.2307/3318437.
[28] S. Lang, *Real and functional analysis*, vol. 142 of Graduate Texts in Mathematics, Springer-Verlag, New York, third ed., 1993, https://doi.org/10.1007/978-1-4612-0897-6.
[29] M. Liu and H. Qiao, *Parameter estimation of path-dependent McKean-Vlasov stochastic differential equations*, 2020, https://arxiv.org/abs/2004.09580.

[30] N. K. Mahato, A. Klar, and S. Tiwari, *Particle methods for multi-group pedestrian flow*, Appl. Math. Model., 53 (2018), pp. 447–461, https://doi.org/10.1016/j.apm.2017.08.024.

[31] F. Malrieu, *Logarithmic Sobolev inequalities for some nonlinear PDE's*, Stochastic Process. Appl., 95 (2001), pp. 109–132, https://doi.org/10.1016/S0304-4149(01)00095-3.

[32] B. Maury and S. Faure, *Crowds in equations*, Advanced Textbooks in Mathematics, World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ, 2019. An introduction to the microscopic modeling of crowds, With a foreword by Laure Saint-Raymond.

[33] K. Oelschläger, *A martingale approach to the law of large numbers for weakly interacting stochastic processes*, Ann. Probab., 12 (1984), pp. 458–479.

[34] C. D. Pagani and S. Salsa, *Analisi matematica - Volume 1*, Masson, 1995.

[35] A. Papavasiliou, G. A. Pavliotis, and A. M. Stuart, *Maximum likelihood drift estimation for multiscale diffusions*, Stochastic Process. Appl., 119 (2009), pp. 3173–3210, https://doi.org/10.1016/j.spa.2009.05.003.

[36] G. A. Pavliotis, *Stochastic processes and applications*, vol. 60 of Texts in Applied Mathematics, Springer, New York, 2014, https://doi.org/10.1007/978-1-4939-1323-7. Diffusion processes, the Fokker-Planck and Langevin equations.

[37] G. A. Pavliotis and A. M. Stuart, *Parameter estimation for multiscale diffusions*, J. Stat. Phys., 127 (2007), pp. 741–781.

[38] T. B. Scheffler, *Analyticity of the eigenvalues and eigenfunctions of an ordinary differential operator with respect to a parameter*, Proc. Roy. Soc. London Ser. A, 336 (1974), pp. 475–486, https://doi.org/10.1098/rspa.1974.0030.

[39] L. Sharrock, N. Kantas, P. Parpas, and G. A. Pavliotis, *Parameter estimation for the McKean-Vlasov stochastic differential equation*, 2021, https://arxiv.org/abs/2106.13751.

[40] T. Suzuki, *Free energy and self-interacting particles*, vol. 62 of Progress in Nonlinear Differential Equations and their Applications, Birkhäuser Boston, Inc., Boston, MA, 2005, https://doi.org/10.1007/0-8176-4436-9.

[41] A.-S. Sznitman, *Topics in propagation of chaos*, in École d'Été de Probabilités de Saint-Flour XIX—1989, vol. 1464 of Lecture Notes in Math., Springer, Berlin, 1991, pp. 165–251, https://doi.org/10.1007/BFb0085169.

[42] J. Tugaut, *Phase transitions of McKean-Vlasov processes in double-wells landscape*, Stochastics, 86 (2014), pp. 257–284, https://doi.org/10.1080/17442508.2013.775287.

[43] J. Wen, X. Wang, S. Mao, and X. Xiao, *Maximum likelihood estimation of McKean-Vlasov stochastic differential equation and its application*, Appl. Math. Comput., 274 (2016), pp. 237–246, https://doi.org/10.1016/j.amc.2015.11.019.

[44] N. Yoshida, *Phase transition from the viewpoint of relaxation phenomena*, Rev. Math. Phys., 15 (2003), pp. 765–788, https://doi.org/10.1142/S0129055X03001746.