Bayesian Nonparametric Inference in McKean-Vlasov models

Richard Nickl [*]   Grigorios A. Pavliotis [†]   Kolyan Ray [†]

April 26, 2024

Abstract

We consider nonparametric statistical inference on a periodic interaction potential W from noisy discrete space-time measurements of solutions $\rho = \rho_W$ of the nonlinear McKean-Vlasov equation, describing the probability density of the mean field limit of an interacting particle system. We show how Gaussian process priors assigned to W give rise to posterior mean estimators that exhibit fast convergence rates for the implied estimated densities $\bar{\rho}$ towards ρ_W. We further show that if the initial condition ϕ is not too smooth and satisfies a standard deconvolvability condition, then one can consistently infer the potential W itself at convergence rates $N^{-\theta}$ for appropriate $\theta > 0$, where N is the number of measurements. The exponent θ can be taken to approach $1/2$ as the regularity of W increases corresponding to 'near-parametric' models.

1   Introduction

We investigate the problem of conducting statistical inference in a class of dynamical systems whose state at time $t \in [0, T]$ is described by the probability density $\rho = \rho_W(t, \cdot)$ solving the nonlinear partial differential equation (PDE)

$$\frac{\partial \rho}{\partial t} = \Delta \rho + \nabla \cdot (\rho \nabla (W * \rho)) \tag{1}$$
$$\rho(0, \cdot) = \phi,$$

known as the *McKean-Vlasov equation*, where $\Delta, \nabla \cdot, \nabla$ denote the Laplacian, divergence and gradient operators, respectively. Further ϕ is the probability distribution of the initial state and W is an interaction potential with Lipschitz gradient. In the present paper we consider this equation on the d-dimensional torus $\mathbb{T}^d$, with periodic boundary conditions. This PDE is a non-linear and non-local Fokker-Planck type equation and plays a fundamental role in a variety of scientific application areas ranging from opinion dynamics [31] and other models in the social sciences [51] to mathematical biology [50], sampling and optimization [59, 55], plasma physics [18], fluid mechanics [28], as well as statistical physics [29, 19] and kinetic theory [63, 25, 41].

The problem we shall study here is how to recover the interaction potential W from a discrete statistical measurement of the evolution of the macrosopic state ρ of the system over

[*]Department of Pure Mathematics and Mathematical Statistics, University of Cambridge, email: nickl@maths.cam.ac.uk

[†]Department of Mathematics, Imperial College London, email: g.pavliotis@imperial.ac.uk, kolyan.ray@imperial.ac.uk

time. We shall take a non-parametric approach where a possibly high- or infinite-dimensional model is postulated for the function W, arising from a Gaussian process prior Π, which then is updated via Bayes rule from 'regression type' data $(Y_i, t_i, X_i)_{i=1}^N$ of the form

$$Y_i = \rho_W(t_i, X_i) + \varepsilon_i, \quad \varepsilon_i \sim^{iid} N(0,1), \quad i = 1, \ldots, N, \tag{2}$$

where the t_i, X_i are sampled discretely from the the time-space cylinder $\mathcal{X} = [0,T] \times \mathbb{T}^d$. This approach is frequently taken in applications, e.g., [27, 32, 37], and fits into the general paradigm of Bayesian non-linear inverse problems with PDEs [62], which is amenable to an algorithmic treatment via MCMC and numerical PDE methods even in high- and infinite-dimensional models [10, 3] – see after (6) below for more details. It also allows us to take advantage of recent theoretical developments in the field, e.g., [43, 42] and especially [44], but before we do so let us shed some more light on how such measurements may arise in concrete physical situations, explaining also the interpretation of the function W as driving interactions.

A basic interacting particle model postulates the simultaneous evolution in time $t \in [0,T]$ of n particles X_t^i in a d-dimensional state space $\mathbb{T}^d$, solving the coupled system of n stochastic differential equations (SDEs)

$$dX_t^i = -\frac{1}{n} \sum_{i \neq j} \nabla W(X_t^i - X_t^j) dt + \sqrt{2} dB_t^i, \quad i = 1, \ldots, n, \tag{3}$$

where the B_t^i are independent d-dimensional Brownian motions, and where *chaotic* initial conditions are assumed, i.e., the X_0^i are all started independently, drawn randomly from the given initial distribution ϕ. This setup can be considered as well in $\mathbb{R}^d$, in particular if we add a drift $-\nabla U(X_t^i) dt$ arising from a (known) confining potential U to the SDE (3). There has been a sequence of interesting recent papers trying to address the problem of inferring W from measurements in models such as (3), see [2, 61, 15, 53, 54, 20, 22, 1, 9, 14] – with earlier references being [4, 34]. Key statistical aspects of this task, such as dealing with the multi-dimensional setting $d > 1$ and clarifying when the potential W is indeed statistically identifiable, still remain broadly open, however. The important reference [14] shows how the particle densities can be estimated empirically from the data, but as our results indicate, one can obtain faster convergence rates than in [14] in important physically relevant settings (see below for more discussion). Furthermore, the hypotheses imposed in all the preceding references to identify W are typically implicit and can generally not be verified for natural parameter spaces. Note also that recently developed proof techniques for non-parametric inference in multi-dimensional diffusion models (see [46, 24, 45, 33] and references therein) do not apply to the model (3) when the number n of particles is large as the dimension nd of the underlying state space then diverges rapidly.

To exploit the information present in measurements of such systems in the mean field limit $(n \to \infty)$, one ought to take advantage of the rich mathematical theory that describes the macroscopic behaviour of particles interacting according to (3). In particular, it is well known that under mild assumptions on the interaction potential, and with chaotic initial data, the *propagation of chaos* property holds– we refer to [63] for a classical reference, to [49, 39] for particularly clear proofs, and [28, 36, 13] for important recent developments. Specifically, it is shown in [36] that if one considers the *marginal distributions* $P_t^{k,n}$ of any fixed number of $k = o(n)$ particles, then in the mean field limit $n \to \infty$ the laws $P_t^{k,n}$ are approximated in relative entropy by a *product measure* P_t^k whose coordinate distributions have probability

2

density ρ_W solving the PDE (1). Moreover, $\rho = \rho_W$ is the density of the law $Law(X_t)$ where X_t is the solution of the *nonlinear* (in the sense of McKean) SDE

$$dX_t = -\nabla W * \rho(t, X_t)dt + \sqrt{2}dB_t, \ t \geq 0, \ \ X_0 \sim \phi. \tag{4}$$

The existence of the solution to the last equation is not difficult to prove, in particular when the state space is the torus, by a fixed point argument in path space, see, e.g., Ch.I in [63]. The intuition is that as n increases, the distribution of any fixed particle is determined by the local dynamics of a Brownian motion with drift term reacting to the local density ρ of nearby particles by convolution with the gradient of the interaction potential. Moreover, we can regard the laws of each distinct (fixed) pair chosen among $k = o(n)$ particles as *approximately independent*. One can then record the 'relative frequencies' of the proportion of these k particles spent in each of N 'bins' of a dissection of the time-space cylinder $[0, T] \times \mathbb{T}^d$. This allows one to obtain approximate ('histogram' type) measurements of the particle density ρ_W, and by the approximate independence of the k marginals we can justify the measurement model (2). We believe that the results obtained here also strongly inform the non-iid models (3), (4), possibly even by using the non-asymptotic approximations in [36] combined with results in [58] to bound the (one-sided) Le Cam distance for equivalence of statistical experiments. These questions are of independent interest and will be investigated elsewhere – in the present article we content ourselves with aggregating the information in the data to a measurement (2) of the particle densities.

Once we have adopted the paradigm of Bayesian nonlinear regression with data (2), the question of whether consistent statistical inference on W, ρ_W is feasible can be determined by studying analytical properties of the parameter to solution map $W \mapsto \rho_W$ of the underlying PDE – this was first shown in [42] and a general theory is laid out in [44]. For the present non-linear Fokker-Planck equation (1) we exploit ideas from [7, 8] that allow to realise ρ_W as a limiting fixed point of sequences of suitable linear parabolic problems, as well as a 'Girsanov type' estimate in relative entropy for solutions to linear and nonlinear Fokker-Planck equations [5, 36]; this bound permits us to replace the arguments for linear PDEs from Sec. 2.1.1 in [44]. Our key analytical results consist, first (Lemma 3), of proving Lipschitz continuity of the non-linear forward map $W \mapsto \rho_W$ on bounded subsets of $C^2(\mathbb{T}^d)$;

$$\|\rho_{W_1} - \rho_{W_2}\|_{L^2(\mathcal{X})} \lesssim \|\nabla(W_1 - W_2) * \rho_{W_2}\|_{L^2(\mathcal{X})} \lesssim \|W_1 - W_2\|_{H^{-\beta}}$$

where $\beta > 0$ is a measure of the regularity of the initial condition $\phi \in H^\beta$. Using the theory from [44] this implies that Gaussian process priors for interaction potentials W give rise to sensible non-parametric models for the relevant regression functions ρ_W. In particular, this enables us to obtain fast convergence rates for recovery of ρ_W by Bayesian plug-in estimates $\rho_{\bar{W}_N}$ arising from the posterior mean $\bar{W}_N = E^\Pi[W|(Y_i, t_i, X_i)_{i=1}^N]$. We then tackle the inverse problem and prove a partial converse to the first inequality in the last display, specifically a stability estimate

$$\|(W_1 - W_2) * \rho_{W_2}\|_{L^2(\mathcal{X})} \lesssim \|\rho_{W_1} - \rho_{W_2}\|_+$$

for a norm $\|\cdot\|_+$ that involves partial derivatives in the space *and* time variables, see (22) for details. When W_1 has a bound on the number of non-zero Fourier modes, this further implies novel injectivity (or 'inverse stabilty') estimate for the entire map $W \mapsto \rho_W$ as soon as deconvolving the 'factor' ρ_{W_2} is possible. This permits us to provide concrete verifiable hypotheses on only the initial condition ϕ under which full inference on the interaction potential becomes

3

feasible at polynomial convergence rates $N^{-\theta}, \theta > 0$, as soon as W is Sobolev smooth. The Bayesian posterior mean estimators that attain these convergence rate can be implemented by MCMC techniques, and unlike frequentist methods studied in earlier references, our approach does not require or postulate an explicit identification formula for W. Rather, the aforementioned stability estimates are sufficient to guarantee recovery, which is one of the main reason why Bayes methods are attractive in nonlinear inverse problems arising with PDEs and data assimiliation tasks, see [44] and references therein.

Our proofs exploit information in the observations available at finite (in fact, short) time horizons, and are not based on any kind of ergodicity properties of the particle system. Several of the recent papers on inference for the McKean SDE cited earlier develop inference methodologies based on observing particles over a long time interval, and rely on the ergodic properties of the dynamics in the limit as $T \to \infty$. This approach is not necessarily well-suited for the mean field dynamics. Indeed, it is well known [11, 7, 13, 12] that the McKean-Vlasov dynamics can exhibit phase transitions, i.e. multiple stationary states can exist. In particular, for the dynamics on the torus in the absence of a confining potential, the uniform measure is always a stationary state for the mean field dynamics, and clearly no information about the interaction potential can be deduced from it. For example, the spectral theoretic approach that was proposed in [53] and that is based on linearising the McKean-Vlasov operator around the invariant measure, does not work in this case, since the linearised Mckean-Vlasov operator is simply the Laplacian. But even when the McKean SDE has a unique *informative* invariant measure, long term horizon asymptotics may not be useful as the invariant measures are generally smooth and hence not easily deconvolvable.

2 Main results

2.1 Notation

We will denote by $C(\mathbb{T}^d)$ the space of (bounded) continuous functions on $\mathbb{T}^d$ while $L^2(\mathbb{T}^d)$ denotes the usual Hilbert space of square integrable functions for Lebesgue (probability) measure dx on $\mathbb{T}^d$. The spaces $H^\alpha(\mathbb{T}^d), C^\alpha(\mathbb{T}^d)$ consist of all functions that have partial (in the former case, weak) derivatives up to order $\alpha \in \mathbb{N}$ defining elements of $L^2(\mathbb{T}^d), C(\mathbb{T}^d)$, respectively, and we set $C^\infty(\mathbb{T}^d) = \cap_{\alpha>0} C^\alpha(\mathbb{T}^d)$, $C^0(\mathbb{T}^d) = C(\mathbb{T}^d)$ by convention. For regularity estimates for the PDE (1), it is also convenient to introduce the space $W^{\alpha,\infty}(\mathbb{T}^d)$ of functions in $C(\mathbb{T}^d)$ whose weak partial derivatives up to order α lie in the space $L^\infty(\mathbb{T}^d)$ of bounded functions. We have $C^\alpha(\mathbb{T}^d) \hookrightarrow W^{\alpha,\infty}(\mathbb{T}^d)$ where $\hookrightarrow$ means a norm-continuous imbedding. We can further define Sobolev spaces $H^\alpha(\mathbb{T}^d) = (H^{-\alpha}(\mathbb{T}^d))^*$ also for $\alpha < 0$ as the topological dual space. The convolution $*$ of functions and measures is defined as usual, for instance if $f, g \in C(\mathbb{T}^d)$ then $f * g(x) = \int_{\mathbb{T}^d} f(x - y)g(y)dy, x \in \mathbb{T}^d$.

We further define function spaces on the time-space cylinder $\mathcal{X} = [0, T] \times \mathbb{T}^d$, such as the Hilbert space $L^2(\mathcal{X}, \lambda)$ where $\lambda = \lambda_T$ is the uniform probability measure on $\mathcal{X}$. For B a normed space we also use standard parabolic PDE notation for the function spaces $L^p([0, T], B)$ of maps $H : \mathcal{X} \to \mathbb{R}$ whose norms $\|H(t, \cdot)\|_B$ lie in $L^p([0, T])$. The spaces $H^m([0, T], B)$ then denote those functions $H \in L^2([0, T], B)$ whose (weak) time derivatives $(\partial^j/\partial t)H$ for all $0 \leq j \leq m$ lie in $L^2([0, T], B)$, with corresponding Hilbert space norm. We also write $\mathcal{P} = \mathcal{P}(\mathbb{T}^d)$ for the set of all (Borel-) probability measures on $\mathbb{T}^d$.

2.2 Gaussian process regression in the McKean-Vlasov model

For an initial condition $\phi \in H^\beta$, consider the time marginal densities $\rho = \rho_W$ providing the unique periodic solutions to the non-linear parabolic PDE (1) on the time-space cylinder $\mathcal{X} = [0,T] \times \mathbb{T}^d$, see Theorem 4. Consider N independent and identically distributed (iid) observations $(t_i, X_i, Y_i)_{i=1}^N$ arising from the model (2), where the (t_i, X_i) are drawn iid from the uniform distribution λ on $[0,T] \times \mathbb{T}^d$, independently of the noise ε_i. We denote by $D_N = ((t_i, X_i, Y_i))_{i=1}^N$ the observation vector, with law P_W^N. As explained in the introduction, such regression-type forward models are widely used in (Bayesian) inverse problems [44] and are approximately justified in interacting particle models (3) by the propagation of chaos phenomenon.

We devise a Bayesian model for the solutions ρ_W of the PDE (1) in the non-linear regression framework (2), placing a prior on the interaction potential W, our quantity of interest, while treating the initial condition ϕ as given. Another common approach in data assimilation with such time evolution equations would be to assign a prior to the initial condition ϕ – see for instance the recent article [47] in the $2D$ Navier-Stokes model.

We employ Gaussian process priors [57] for the interaction potential – these will be supported on a separable normed linear space $(\mathcal{W}, \|\cdot\|_{\mathcal{W}})$ satisfying the imbedding

$$\mathcal{W} \hookrightarrow C^2(\mathbb{T}^d) \cap \left\{ W : \int_{\mathbb{T}^d} W(x)dx = 0 \right\}. \tag{5}$$

Note that the potential W is identifiable only up to constants as the gradient ∇W relevant in models (1),(3),(4) annihilates constants, so fixing $\int W = 0$ is natural for identifiability reasons. The separability is required to apply Gaussian process techniques from [23, 44] – we can think of $\mathcal{W}$ to equal a Sobolev space $H^{\alpha+1}(\mathbb{T}^d)$ for $\alpha > 1 + d/2$ or, to obtain sharper results, we can also take periodic Besov spaces $\mathcal{W} = B_{\infty 1}^\alpha(\mathbb{T}^d), \alpha \geq 2$ ([23], p.370). Concrete examples will be discussed below, but for now we let Π be any Borel probability measure on some space $\mathcal{W}$ satisfying (5).

The posterior distribution obtained from such a prior Π and data (2) is given by Bayes' formula via standard arguments [21, 44];

$$d\Pi(W|D_N) \propto e^{\ell_N(W)} d\Pi(W); \quad \ell_N(W) = -\frac{1}{2} \sum_{i=1}^N |Y_i - \rho_W(t_i, X_i)|^2, \quad W \in \mathcal{W}. \tag{6}$$

Posterior draws $w \sim \Pi(\cdot|D_N)$ can be approximately calculated from Markov chain Monte Carlo (MCMC) techniques, for instance by the pCN, ULA or MALA algorithm, following the now well established paradigm of Bayesian inversion and uncertainty quantification in PDE models, see [62, 10, 3, 52] and also Section 1.2.4 in [44] for an overview. Each MCMC step requires one numerical solution of the non-linear Fokker-Planck equation (1), for which a variety of methods exists, see [6, 26, 60] for recent references. We can then form ergodic averages $(1/K) \sum_{k=1}^K w_k$ of these MCMC draws w_k to approximately compute the posterior mean estimate $\bar{W} = E^\Pi[W|D_N]$ of the interaction potential W, which in turn delivers estimates also of the particle densities $\rho_{\bar{W}}$ by one further numerical solution of the Fokker-Planck equation (1) with interaction potential $\bar{W}$. For a concrete example and further discussion on applying MCMC method for instance to the Keller-Segel model for chemotaxis (which is a non-linear, non-local PDE of the form (1), [7, Sec. 6.5]), see for instance [37].

One can also attempt to minimise the negative log-posterior over $\mathcal{W}$ (amounting to a MAP estimate, or Tikhonov regulariser), but we note that the problem is non-convex due to the

nonlinearity of $W \mapsto \rho_W$, hence MCMC may be more robust to the presence of local optima in the criterion function. Computational guarantees for gradient based MCMC methods can be obtained in principle following ideas from [48] combined with some of the stability estimates from the proofs that follow. But this is beyond the scope of the present paper – we shall focus here on analysing the statistical properties of the posterior distribution $\Pi(\cdot|D_N)$ under the objective ('frequentist') assumption that an actual ground truth potential W_0 generated the interacting particle system and thus data from equation (2). Therefore the statistical analysis is under the law $P_{W_0}^N$ of the data vector D_N, and $\to^{P_{W_0}^N}$ will denote convergence in $P_{W_0}^N$-probability.

2.3 Posterior contraction rates for the particle densities

We first establish contraction rates for posterior distributions (6) arising from rescaled and possibly projected Gaussian process priors over $\mathbb{T}^d$. We start with a mean-zero Gaussian process $(V(x) : x \in \mathbb{T}^d)$ with reproducing kernel Hilbert space (RKHS) $\mathcal{H}$ whose law Π_V satisfies the following assumption. For definitions and background material on Gaussian processes and their associated RKHS, see Chapter 11 of [21] or [57].

Condition α. *Let Π_V be a centred Gaussian probability measure on $\mathcal{W}$ from (5) with RKHS $(\mathcal{H}, \| \cdot \|_{\mathcal{H}})$. Further suppose that for some $\alpha > d/2 + 1$, the continuous embedding $\mathcal{H} \hookrightarrow H^{\alpha+1}(\mathbb{T}^d)$ holds.*

To deal with various non-linearities in our regression problem, we follow ideas in [42] and rescale the 'base prior' via

$$W = \frac{V}{\sqrt{N}\delta_N}; \qquad \delta_N = N^{-\frac{\alpha+1+\beta}{2(\alpha+1)+2\beta+d}}, \tag{7}$$

to introduce extra regularisation in the posterior distribution. Here, $\alpha + 1$ models the regularity of the interaction potential W (so α models the regularity of ∇W), while $\beta > 0$ is determined by that of the initial condition $\rho_W(0, \cdot) = \phi \in H^\beta$. The extra regularisation allows us for instance to deduce space or time regularity estimates for ρ_W corresponding to interaction potentials W drawn from the posterior, see Theorem 5 below. Note that such priors are special cases of the rescaled Gaussian process priors considered in several 'direct' statistical settings [21].

We will further consider $L^2(\mathbb{T}^d)$-projections $\pi_{\mathcal{W}_N}(W)$ of the law of W onto sequences of linear subspaces $\mathcal{W}_N \subseteq \mathcal{W}$, including the case $\mathcal{W}_N = \mathcal{W}$ for all N. The final prior law of such W on $\mathcal{W}$ or $\mathcal{W}_N$ will be denoted by $\Pi = \Pi_N$. A typical example of such a projection is the truncation of an infinite dimensional Gaussian process onto its finite-dimensional counterpart, as is usually done in practice. We note that less aggressive shrinkage would be permitted in (7) but we opt for the present choice as this allows us to directly use results from [44] in the proofs.

The following theorem shows that we can solve the regression problem underlying (2) via the implied prior arising from a Gaussian process model for the interaction potential W. [We restrict to *even* integers β to facilitate the application of parabolic PDE theory in the proofs.]

Theorem 1. *Suppose $\phi \in H^\beta(\mathbb{T}^d)$ for some even integer $\beta \geq 3 + d$ and let Π_V be a Gaussian measure satisfying Condition α. Consider the rescaled Gaussian process prior from*

6

(7) *projected onto a linear subspace $\mathcal{W}_N$ of $\mathcal{W}$. Let $W_0 \in \mathcal{H}$ satisfy $\int_{\mathbb{T}^d} W_0 = 0$, and assume further that there exists a sequence $W_{0,N} \in \mathcal{H} \cap \mathcal{W}_N$ such that $\|W_{0,N}\|_{\mathcal{H}} = O(1)$ and $\|W_0 - W_{0,N}\|_{H^{-\beta}} = O(\delta_N)$ as $N \to \infty$. Then for $M, L > 0$ large enough, we have*

$$\Pi(W \in \overline{\mathcal{W}}_{M,N} : \|\rho_W - \rho_{W_0}\|_{L^2(\mathcal{X})} \leq L\delta_N | D_N) \to^{P_{W_0}^N} 1$$

as $N \to \infty$, where

$$\overline{\mathcal{W}}_{M,N} = \{W = W_1 + W_2 \in \mathcal{W}_N : \|W_1\|_{H^{-\beta}} \leq M\delta_N, \ \|W_2\|_{\mathcal{H}} \leq M, \ \|W\|_{C^2} \leq M\}. \tag{8}$$

The convergence rate $\delta_N = N^{-\frac{\alpha+1+\beta}{2(\alpha+1)+2\beta+d}}$ derived above corresponds to the forward rate of an $(\alpha + 1)$-smooth truth in an inverse problem with 'polynomial ill-posedness' β. The self-convolutional term $\nabla W * \rho_W$ appearing in models (1) and (4) suggests the ill-posedness here should be related to the smoothness of ρ_{W_0}. In the present nonlinear parabolic PDE setting, the regularity of the densities ρ_{W_0} is driven by that of the initial condition ϕ rather than that of W_0, see Theorem 5, which is reflected in the role of β in the forward rate, see (19) below. Whether this is optimal is a delicate question concerning the optimality of regularity estimates in nonlinear parabolic PDEs. Note that the convergence rate δ_N accelerates to the 'parametric' rate $1/\sqrt{N}$ as soon as either the model for W or the initial condition ϕ is very smooth, i.e., as α or $\beta \to \infty$.

To compare the convergence rates obtained in the preceding theorem to the existing literature, let us consider a general version of the interacting particle system (3) on all of $\mathbb{R}^d$:

$$dX_t^i = -\nabla U(X_t^i)dt - \frac{1}{n}\sum_{i \neq j} \nabla W(X_t^i - X_t^j)dt + \sqrt{2}dB_t^i, \quad i = 1, \ldots, n, \tag{9}$$

where U is a suitable confining potential. For (spatially) β-smooth ρ solving the corresponding non-linear Fokker-Planck equation, Della Maestra and Hoffmann [14] show that the minimax rate for pointwise estimation of ρ is the 'usual' nonparametric rate $n^{-\frac{\beta}{2\beta+d}}$. One can check that their minimax lower bound is established by considering $W = 0$, that is, in a diffusion model without interaction term, where regularity of U translates directly into that of ρ. However, when U is zero or much smoother than W, our results show that we can expect significantly faster rates $n^{-\frac{\alpha+1+\beta}{2(\alpha+1)+2\beta+d}}$ for the forward map ρ_W if the deconvolution structure is correctly taken advantage of.

This suggests that methods that are only based on the smoothness of $\rho \in H^\beta$ and not (as we do via our Lemma 3) on properties of the underlying parameter to solution map $(U, W) \mapsto \rho_{U,W}$ may not be able to fully exploit the information available in the data. This highlights a significant advantage of modelling W directly, as the Bayes method does. We also note that by placing a prior on W, and hence ρ_W, one obtains posterior draws ρ_W that lie in the range of the forward map (i.e., they solve the PDE (1)). This is key for applying our stability estimate (14) below for recovery of W. While the Bayes method does this automatically, it is not necessarily the case for other methods that first estimate ρ_W directly by some smoothing method which relaxes the PDE constraint.

In summary, our results demonstrate that the estimation rates for ρ are more complex than the regularity of ρ alone might suggest, and that they depend on a delicate interplay of the interaction term W, the initial condition ϕ, and the confining potential U.

2.3.1 Examples of Gaussian priors

We now discuss examples of Gaussian priors satisfying the conditions of Theorem 1.

Example 1 (Periodized Matérn process). Consider the Gaussian process on $\mathbb{T}^d$ given by

$$V(x) = (2\pi)^{d/2} \sum_{0 \neq k \in \mathbb{Z}^d} \frac{1}{(1 + 4\pi^2|k|^2)^{(\alpha+1)/2}} g_k e_k(x), \qquad g_k \sim^{iid} N(0,1), \quad x \in \mathbb{T}^d,$$

where the $e_k \propto e^{2\pi i k(\cdot)}$ are the $L^2(\mathbb{T}^d)$-orthonormal trigonometric polynomials over $\mathbb{T}^d$. This is the $L^2(\mathbb{T}^d)$-series expansion resulting from periodising the usual Matérn process on $\mathbb{R}^d$, see Section A.1.1 of [24] for details. We have excluded the constant function from the trigonometric basis to incorporate our identifiability condition $\int_{\mathbb{T}^d} W = 0$.

By standard arguments (Ex. 2.6.15 in [23]), the RKHS of V equals

$$\mathcal{H} := H^{\alpha+1}(\mathbb{T}^d) \cap \left\{ \int_{\mathbb{T}^d} W = 0 \right\}, \tag{10}$$

with equivalent norm $\|\cdot\|_{\mathcal{H}} \simeq \|\cdot\|_{H^{\alpha+1}}$. The above process has a version V whose sample paths are in $C^{\alpha+1-d/2-\eta}(\mathbb{T}^d)$ for any $\eta > 0$ ([21], Proposition I.4). Thus using [21], Lemma I.7, we see that V defines a Gaussian random element of $C^{r_0}(\mathbb{T}^d)$ for any $r_0 < \alpha + 1 - d/2$, and since for $\alpha > 1 + d/2$ the space $C^{r_0}(\mathbb{T}^d)$ imbeds continuously into the separable Besov space $B^r_{\infty 1}(\mathbb{T}^d) \hookrightarrow C^2(\mathbb{T}^d), 2 < r < r_0$, (cf. p.370f. in [23]) we can realise the law of V as a Gaussian Borel probability measure on

$$\mathcal{W} = B^r_{\infty 1}(\mathbb{T}^d) \cap \left\{ \int_{\mathbb{T}^d} W = 0 \right\}, \text{ any } r < \alpha + 1 - d/2,$$

so that Condition α is verified.

Remark 1. Passing through a separable space is important to apply techniques from Gaussian process theory [23], Ch.2, but the introduction of Besov spaces could be avoided by increasing the smoothness to $\alpha > 1 + d$ and taking $\mathcal{W} = H^{\alpha-d/2}$ as in Theorem B.1.3 in [44].

Now assume $W_0 \in \mathcal{H}$ defined in (10) which also lies in $\mathcal{W}$ (by what precedes, or abstractly by Corollary 2.6.17 in [23]). Applying Theorem 1 with $\mathcal{W}_N = \mathcal{W}, W_{0,N} = W_0$, then gives that for $M, L > 0$ large enough,

$$\Pi\left(W \in \mathcal{W}_N : \|\rho_W - \rho_{W_0}\|_{L^2(\mathcal{X})} \leq LN^{-\frac{\alpha+1+\beta}{2(\alpha+1)+2\beta+d}}, \|W\|_{C^2} \leq M | D_N \right) \to^{P^N_{W_0}} 1,$$

as $N \to \infty$ for this prior.

Example 2 (Truncated Fourier prior). For implementation by MCMC, and also in order to conduct inference on the interaction potential later, it is of interest to remove the higher Fourier frequencies from the Matérn prior for W just constructed. For $K_N \in \mathbb{N}, K_N \to_{N \to \infty} \infty$, consider the truncated Fourier series prior

$$\pi_{\mathcal{W}_N}(W)(x) = \frac{1}{\sqrt{N}\delta_N} \sum_{k \in \mathbb{Z}^d : 0 < |k| \leq K_N} \frac{1}{(1+|k|^2)^{(\alpha+1)/2}} g_k e_k(x), \qquad g_k \sim^{iid} N(0,1), \tag{11}$$

which equals the L^2-projection of W onto $\mathcal{W}_N = E_{K_N}$, the span of the trigonometric polynomials up to frequency K_N (excluding constants). Let $W_0 \in \mathcal{H}$, and set $W_{0,N} =$

8

$\sum_{k\in\mathbb{Z}^d:0<|k|\leq K_N}\langle W_0, e_k\rangle_{L^2}e_k \in \mathcal{H}\cap\mathcal{W}_N$ to be the Fourier projection of W_0 onto the first K_N frequencies. Then $\|W_{0,N}\|_{\mathcal{H}} \leq \|W_0\|_{H^{\alpha+1}}$ and $\|W_0 - W_{0,N}\|_{H^{-\beta}} \lesssim K_N^{-\alpha-1-\beta} = O(\delta_N)$ for $K_N \gtrsim N^{\frac{1}{2(\alpha+1)+2\beta+d}}$. Applying Theorem 1 gives that for $M, L > 0$ large enough,

$$\Pi(W \in \overline{\mathcal{W}}_{M,N} : \|\rho_W - \rho_{W_0}\|_{L^2(\mathcal{X})} \leq LN^{-\frac{\alpha+1+\beta}{2(\alpha+1)+2\beta+d}}|D_N) \to^{P_{W_0}^N} 1,$$

as $N \to \infty$, with regularisation $\overline{\mathcal{W}}_{M,N}$ set given in (8). Note the condition $K_N \gtrsim N^{\frac{1}{2(\alpha+1)+2\beta+d}}$ means the prior scaling rather than the truncation drives the prior smoothness.

Remark 2 (Symmetric potentials). In many physical applications, the interaction potential W is symmetric. This can be encoded into the prior by separating the real and imaginary parts of the trigonometric polynomials $e_k \propto e^{2\pi i k(\cdot)}$ into $\phi_{2k}(x) \propto \cos(2\pi kx)$ and $\phi_{2k+1}(x) \propto \sin(2\pi kx)$, and setting the prior coefficients of ϕ_{2k+1} to be zero. This yields prior and then also posterior draws that are symmetric about zero and all our results apply equally for Sobolev smooth *symmetric* truths W_0.

Remark 3 (Almost finite-dimensional models). One can also consider smoother Gaussian processes, for instance

$$V(x) = (2\pi)^{d/2} \sum_{0\neq k\in\mathbb{Z}^d} e^{-r|k|_1/2}g_k e_k(x), \qquad g_k \sim^{iid} N(0,1), \quad x \in \mathbb{T}^d,$$

where $|k|_1 = |k_1|+\cdots+|k_d|$ and $r > 0$. These model infinitely differentiable periodic functions with exponentially decaying Fourier coefficients. For example, as in [56], we consider the rescaled Gaussian process

$$W = \frac{V}{\log N},$$

where less rescaling is needed than above since the sample paths of V are already very regular. The RKHS of V is

$$\mathcal{H}_r = \left\{h = \sum_{0\neq k\in\mathbb{Z}^d} h_k e_k : \|h\|_{\mathcal{H}}^2 := \sum_{0\neq k\in\mathbb{Z}^d} h_k^2 e^{r|k|_1} < \infty\right\}, \tag{12}$$

which embeds continuously into any Sobolev space $H^{\alpha+1}$, $\alpha > 0$, meaning we have all the required regularity properties needed above. Our arguments can then be modified to extend to such settings by standard arguments, see Section C of [56] for details. For $W_0 \in \mathcal{H}_r$, we thus have for $M, L > 0$ large enough,

$$\Pi\big(W \in \mathcal{W}_N' : \|\rho_W - \rho_{W_0}\|_{L^2(\mathcal{X})} \leq L\delta_N', \ \|W\|_{W^{2,\infty}} \leq M|D_N) \to^{P_{W_0}^N} 1$$

as $N \to \infty$, where $\delta_N' = (\log N)^\eta/\sqrt{N}$ for some $\eta > 0$ and $\mathcal{W}_N' = \{W = W_1+W_2 : \|W_1\|_{H^{-\beta}} \leq M\delta_N', \|W_2\|_{\mathcal{H}_r} \leq M, \|W\|_{C^2} \leq M\}$. Such priors work for truths with exponentially decaying Fourier coefficients and are thus relevant when the interaction potential W_0 can be well approximated by relatively few Fourier modes. This occurs in several physical applications, for example the Kuramoto-Shinomoto-Sakaguchi ($O(2)$) model for synchronization or the (wrapped) Gaussian (attractive or repulsive) interaction potential; we refer to [7, Sec. 6] for further discussion and several examples.

2.4 Recovering the interaction potential

We now consider conditions under which we can consistently recover the interaction potential W_0 from noisy observations (2) of the McKean-Vlasov PDE (1).

We notice that the posterior contraction rate δ_N for $\|\rho_W - \rho_{W_0}\|_{L^2}$ from the previous theorem becomes close to the parametric rate $1/\sqrt{N}$ as $\beta \to \infty$, even if we keep the regularity α of the prior model for the potential W fixed. This reflects the self-convolving nature of the McKean-Vlasov dynamics: the density ρ_W can be very smooth even when W is not, as long as the initial condition is sufficiently regular in a Sobolev sense. A consequence is that the interacting particle dynamics partially mask information about W unless the initial condition is somehow atypical. Viewed this way, inferring W becomes a statistical inverse problem which compounds a deconvolution step and a nonlinear inverse problem for the Fokker-Planck equation (1). We now show that this inverse problem can be solved, provided that the initial condition ϕ is not too smooth, measured by a lower bound on the decay of its Fourier coefficients. In other words, if we can prepare the system to start in a somewhat irregular initial state, then the observed dynamics of the interacting particle system will reveal the potential W to the Bayesian posterior distribution arising from a prior that concentrates on at most a finite (but growing) number of Fourier modes.

At the heart of the main statistical result Theorem 3 is the following stability estimate for the forward map $W \mapsto \rho_W$ under such hypotheses.

Theorem 2. *Suppose $W_0 \in H^{\alpha+1}$ for $\alpha > 1 + d/2$ and $\phi \in H^\beta(\mathbb{T}^d)$ where $\beta \geq 3 + d$ is an even integer. Assume ϕ is strictly positive $\phi \geq \phi_{min} > 0$ on $\mathbb{T}^d$ and that the Fourier coefficients $\hat{\phi}$ of ϕ satisfy*

$$|\hat{\phi}(k)| \geq c|k|^{-\zeta}, \qquad k \in \mathbb{Z}^d, \tag{13}$$

for some $\zeta > \beta + d/2$. Let $W = \sum_{k \in \mathbb{Z}^d:|k| \leq K} W_k e_k \in E_K$ be band-limited by K Fourier frequencies and assume $\|W\|_{W^{2,\infty}} + \|W_0\|_{W^{2,\infty}} \leq M$ as well as $\int_{\mathbb{T}^d}(W - W_0)dx = 0$. Then there exists a constant $C = C(d, T, M, \|\phi\|_{H^\beta}, \phi_{min}) > 0$ such that for any $0 < t_0 \leq \frac{c}{2C}K^{-\zeta}$ we have

$$\|W - W_0\|_{L^2}^2 \leq C\left(K^{-2\alpha-2} + t_0^{-1}K^{2\zeta}\left(\|\rho_W - \rho_{W_0}\|_{H^1([0,t_0],H^{-1})}^2 + \|\rho_W - \rho_{W_0}\|_{L^2([0,t_0],H^1)}^2\right)\right). \tag{14}$$

The analytical intuition of this stability estimate could be phrased as saying that if we can obtain precise estimates of the space and time derivatives of ρ_{W_0}, we will be able to identify $W_0 \in \mathcal{H}$ under the hypothesis (13). So *both* the space and time dynamics of the densities ρ_W are relevant for reconstruction. We can combine this estimate with the truncated Fourier series prior in Example 2 to consistently recover the underlying potential.

Theorem 3. *Suppose $W_0 \in \mathcal{H}$ for $\alpha > 1 + d/2$ and $\phi \in H^\beta(\mathbb{T}^d)$, where $\beta \geq 3 + d$ is an even integer. Further assume that ϕ is lower bounded by $\phi_{min} > 0$ and satisfies (13). Consider the truncated Fourier series prior in Example 2 with $K_N \simeq N^{\frac{1}{2(\alpha+1)+2\beta+d}}$. Then for $L > 0$ large enough,*

$$\Pi(W \in E_{K_N} : \|W - W_0\|_{L^2} \geq LK_N^{3\zeta/2}\delta_N^{\frac{\beta-2}{\beta}} | D_N) \to^{P_{W_0}^N} 0,$$

as $N \to \infty$, where $\delta_N = N^{-\frac{\alpha+1+\beta}{2(\alpha+1)+2\beta+d}}$. In particular, the rate is $N^{-\theta}$ for

$$\theta = \frac{(\alpha + 1 + \beta)(\beta - 2)/\beta - 3\zeta/2}{2(\alpha + 1) + 2\beta + d}. \tag{15}$$

10

Moreover, if $\bar{W}_N = E^\Pi[W|D_N] \in E_{K_N}$ is the posterior mean, then we also have

$$\|\bar{W}_N - W_0\|_{L^2} = O_{P_{W_0}^N}(N^{-\theta}) \text{ as } N \to \infty.$$

Theorem 3 uses the stability estimate (14) with $t_0 \simeq K_N^{-\zeta} \to 0$ small, which shows that as long as the initial condition ϕ is sufficiently irregular, the *short-time* dynamics of the system are already enough to consistently recover W as $N \to \infty$. This is especially relevant in our setting where ergodic averages may be uninformative for W, since the uniform distribution is always a stationary state for the mean-field dynamics on the torus [7]. Whether one can further exploit long-time information under relevant physical scenarios is a subtle question given the existence of phase transitions and, consequently, of multiple stationary states for McKean-Vlasov dynamics. In any case, our result already shows that Bayesian methods can extract relevant short-time information from the data.

The condition $\zeta > \beta + d/2$ ensures that ϕ can satisfy both $\phi \in H^\beta$ and (13) simultaneously. If the lower bound (13) is too weak, i.e. the initial density is much harder to deconvolve that its Sobolev smoothness β suggests, then our result may not imply consistency. Consistency with polynomial rates is possible as soon as $\theta > 0$, which holds for ζ satisfying

$$\beta + d/2 < \zeta < \frac{2(\alpha + 1 + \beta)(\beta - 2)}{3\beta}.$$

This range is non-empty for $\alpha > 0$ large enough. For semi-parametric applications to Bernstein-von Mises theorems (e.g., as in Sec. 4.1.1 in [44]) it is important that these rates can be sufficiently fast in regular models: for instance we can always obtain $3\theta > 1$ as long as $\alpha \geq \alpha_0(\beta, d)$ for some $\alpha_0 = \alpha_0(\beta, d)$, if (13) holds for $\zeta = \beta + d/2 + \eta$ and all $\eta > 0$.

One can further check that the rate $K_N^{3\zeta/2}\delta_N^{\frac{\beta-2}{\beta}}$ holds in Theorem 3 if one takes any $K_N \gtrsim N^{\frac{1}{2(\alpha+1)+2\beta+d}}$, though taking larger K_N here deteriorates the final rate. One can also extend the theorem to $K_N \lesssim N^{\frac{1}{2(\alpha+1)+2\beta+d}}$ giving deconvolution rate

$$K_N^{-\alpha-1} + K_N^{3\zeta/2}\left(\xi^{\frac{\beta-2}{\beta}} + \xi_N^{\frac{\beta}{\beta+1}}\right),$$

where now $\xi_N = \delta_N + K_N^{-\alpha-1-\beta}$ is a possibly slower contraction rate (since the prior truncation may now drive the forward rate).

Remark 4 (Deconvolution rates for almost finite-dimensional models)**.** We can also consider the truncated version of the prior in Remark 3, truncating at level K_N as in (11). For $W_0 \in \mathcal{H}_r$ defined in (12) and $K_N \geq \frac{1}{r} \log N$, we obtain the same near-parametric contraction rate $\delta_N' = (\log N)^\eta/\sqrt{N}$ for the densities ρ_W as in Remark 3. Turning to the deconvolution, we deduce the rate

$$\|W - W_0\|_{L^2}^2 \lesssim e^{-rK_N} + K_N^{3\zeta}(\delta_N')^{\frac{2(\beta-2)}{\beta}} \lesssim (\log N)^{3\zeta + \frac{2(\beta-2)\eta}{\beta}} N^{-\frac{\beta-2}{\beta}}.$$

Thus for $W_0 \in \mathcal{H}_r$, one can also recover the true potential W_0, and thus solve the inverse problem, at the near-parametric rate $(\log N)^{\eta'}/\sqrt{N}$ as the regularity $\beta \to \infty$.

We next give a concrete example of a distribution ϕ for which the above theorem applies and consistent recovery of the potential W_0 with polynomial rates is possible.

Example 3 (Periodised symmetric multivariate Laplace distribution). Consider the symmetric d-dimensional multivariate Laplace distribution on $\mathbb{R}^d$, usually defined via its characteristic function $\frac{1}{1+|\xi|^2/2}$, $\xi \in \mathbb{R}^d$. This distribution has a density function of the form ([35], p. 235)

$$f_d(y) = \frac{2}{(2\pi)^{d/2}} \left(\frac{|y|^2}{2} \right)^{\frac{2-d}{2}} K_{(2-d)/2}(\sqrt{2}|y|), \qquad y \in \mathbb{R}^d,$$

where K_λ is the modified Bessel function of the second kind, having integral representation $K_\lambda(x) = \int_0^\infty e^{-x\cosh t} \cosh(\lambda t) dt$. Consider the periodized version of this density

$$f_d^{per}(x) = \sum_{k \in \mathbb{Z}^d} f_d(x+k), \qquad x \in (0,1]^d,$$

which defines a strictly positive density function on the torus $\mathbb{T}^d$ with Fourier series

$$\hat{f}_d^{per}(k) = F_{\mathbb{T}^d}[f_d^{per}](k) = \int_{\mathbb{T}^d} f_d^{per}(x) e^{-2\pi i k.x} dx = F_{\mathbb{R}^d}[f_d](k) = \frac{1}{1 + 2\pi^2|k|^2}, \qquad k \in \mathbb{Z}^d,$$

where the relation with the Fourier transform of the unperiodized density f_d follows from Theorem 8.35 of [17]. For any $m \geq 1$, consider the m-fold convolution $\phi_m = f_d^{per} * \cdots * f_d^{per}$ which defines a positive probability density function on $\mathbb{T}^d$ with Fourier transform

$$\hat{\phi}_m(k) = \frac{1}{(1 + 2\pi^2|k|^2)^m}.$$

Then $\phi_m \in H^\beta$ for any $\beta < 2m - d/2$ and satisfies the deconvolution condition (13) with $\zeta = 2m$.

Applying Theorem 3 with $W_0 \in \mathcal{H} \subset H^{\alpha+1} \cap W^{2,\infty}$ and $\beta = 2m - d/2 - \eta$ for $\eta > 0$ arbitrarily small, we obtain contraction rate

$$\Pi(W \in E_{K_N} : \|W - W_0\|_{L^2} \geq LN^{-\theta'}|D_N) \to^{P_{W_0}} 0$$

as $N \to \infty$, where θ' is any constant satisfying

$$\theta' < \frac{(\alpha + 1 + 2m - d/2)^{\frac{2m-d/2-2}{2m-d/2}} - 3m}{2(\alpha+1) + 4m}$$

(which can get arbitrarily close to the upper bound by taking $\eta > 0$ small enough). The right-hand side is strictly positive if and only if

$$\alpha + 1 > \frac{(2m - d/2)(m + d/2 + 2)}{(2m - d/2 - 2)},$$

in which case we can take $\theta' > 0$. Thus for any fixed m, one can consistently recover $W_0 \in \mathcal{H}$ with polynomial rates for α large enough.

3 Proofs

3.1 Proof of Theorem 1: forward contraction rate

We verify the conditions of Theorem 2.2.2 in [44], namely Condition 2.1.1 of [44] with regularization space $\mathcal{R} = W^{2,\infty}$ on which the prior concentrates (in view of (5) and the continuous

12

imbedding $C^2 \hookrightarrow W^{2,\infty}$). This consists of showing (i) boundedness of ρ_W on the domain $\mathcal{X}$ and (ii) Lipschitz continuity of the forward map $W \mapsto \rho_W : H^{-\beta}(\mathbb{T}^d) \to L^2(\mathcal{X})$, uniformly over $\| \cdot \|_{W^{2,\infty}}$-balls of radius $M > 0$. While that theorem requires that $W_0 \in \mathcal{H} \cap \mathcal{R}$, it is straightforward to modify the proof to require only $W_0 \in \mathcal{R} = W^{2,\infty}$ along with a suitably good approximation sequence $W_{0,N}$ in the RKHS $\mathcal{H}$, see (2.22) or Exercise 2.4.3 in [44].

(i) *Boundedness*: using the Sobolev embedding theorem $H^{d/2+\varepsilon}(\mathbb{T}^d) \hookrightarrow C(\mathbb{T}^d)$ for any $\varepsilon > 0$ and the $L^\infty([0,T]; H^{k+1}(\mathbb{T}^d))$ regularity estimate (27) from Theorem 5 with $k = d-1$,

$$
\begin{aligned}
\sup_{W : \|W\|_{\mathcal{R}} \leq M} \|\rho_W\|_{L^\infty(\mathcal{X})} &\leq \sup_{W : \|W\|_{W^{2,\infty}} \leq M} \sup_{0 \leq t \leq T} \|\rho_W(t)\|_{H^{d/2+\varepsilon}} \\
&\leq \sup_{W : \|W\|_{W^{2,\infty}} \leq M} \|\rho_W\|_{L^\infty([0,T]; H^d(\mathbb{T}^d))} \leq C_{M,d,T,\|\phi_0\|_{H^d}},
\end{aligned}
\tag{16}
$$

since $\phi \in H^\beta$ with $\beta \geq 3 + d$. This verifies (2.3) in Condition 2.1.1 of [44].

(ii) *Lipschitz continuity of the forward map*: we control the distance between solutions to the McKean-Vlasov equation (1) for different potentials via the relative entropy/Kullback-Leibler divergence of the marginal densities. We first require the following estimate from [36, 5], adapted on the torus.

Lemma 1. *Let $W_i \in W^{2,\infty}(\mathbb{T}^d)$, $i = 1, 2$, $\phi \in H^{3+d}(\mathbb{T}^d) \cap \mathcal{P}(\mathbb{T}^d)$, and consider the solutions ρ_i, $i = 1, 2$, of the corresponding Fokker-Planck equations*

$$
\begin{aligned}
\frac{\partial \rho_i}{\partial t} &= \Delta \rho_i + \nabla \cdot (\rho_i (\nabla W_i * \rho_i)), \\
\rho_i(0, \cdot) &= \phi.
\end{aligned}
\tag{17}
$$

Then for all $t > 0$,

$$
\begin{aligned}
H(\rho_1(t)|\rho_2(t)) &\equiv \int_{\mathbb{T}^d} \rho_1(t,x) \log \frac{\rho_1(t,x)}{\rho_2(t,x)} dx \\
&\leq \frac{1}{2} \int_0^t \int_{\mathbb{T}^d} |\nabla W_1 * \rho_1(u) - \nabla W_2 * \rho_2(u)|^2 \rho_1(u) \, dx \, du.
\end{aligned}
$$

Proof. The estimate follows from [36], Lemma 3.1, which follows from [5], Theorem 1.1. The proof can be easily adapted so that it holds for Fokker-Planck equations posed on $\mathbb{T}^d$ instead of $\mathbb{R}^d$. We now check the conditions (H1)-(H3) of [36][Lem. 3.1].

Under our assumptions on the interaction potential and on the initial condition ϕ, from Theorem 4 below it follows that the Fokker-Planck equations (17) have a unique, strictly positive solution $\rho_i(t) \in \mathcal{P}(\mathbb{T}^d) \cap C^2(\mathbb{T}^d)$ for all $t > 0$. By standard convolution inequalities, $\nabla W_2 * \rho_2$ is bounded, therefore (H1) is satisfied. This implies, in particular, that the weighted L^p spaces in the proof of [36], Lemma 3.1 can be replaced by the flat spaces $L^p(\mathbb{T}^d)$. Furthermore, since $W_i \in W^{2,\infty}(\mathbb{T}^d)$ and ρ_i is bounded on $\mathbb{T}^d$ for all $T > 0$, together with Young's inequality for convolutions, we have that $\nabla W_1 * \rho_1 \in L^2((0,T); L^p(\mathbb{T}^d))$ for $p > d+2$ (in fact, we can take $p = \infty$), $\nabla W_2 * \rho_2 \in L^2((0,T); L^2(\mathbb{T}^d))$ and $\nabla W_1 * \rho_1 \in L^\infty((0,T); L^2(\mathbb{T}^d))$, i.e. (H2)-(H3) are satisfied. $\square$

Lemma 2. *Let $W_i \in W^{2,\infty}(\mathbb{T}^d)$, $\phi \in H^{3+d}(\mathbb{T}^d) \cap \mathcal{P}(\mathbb{T}^d)$ and consider the corresponding solutions $\rho_i = \rho_{W_i}$ to (1). If $\|W_1\|_{W^{2,\infty}} \leq M$, then for all $t \in [0, T]$,*

$$
\int_0^t H(\rho_1(u)|\rho_2(u)) \, du \leq C \int_0^t \|\nabla(W_1 - W_2) * \rho_2(t)\|_{L^2(\mathbb{T}^d)}^2,
$$

where $C = C(d, T, M, \|\phi\|_{H^d})$.

Proof. Arguing as in (16) and using Theorem 5,

$$\|\rho_1\|_{L^\infty(\mathcal{X})} \leq C(d)\|\rho_1\|_{L^\infty([0,T];H^d(\mathbb{T}^d))} \leq C_0(d,T,M,\|\phi\|_{H^d}). \tag{18}$$

Using Lemma 1, Young's convolution inequality and Pinsker's inequality (e.g., Prop. 6.1.7 in [23]) we have for all $0 \leq t \leq T$:

$$
\begin{aligned}
H(\rho_1(t)|\rho_2(t)) &\leq \frac{1}{2}\|\rho_1\|_{L^\infty(\mathcal{X})} \int_0^t \int_{\mathbb{T}^d} |\nabla W_1 * \rho_1(u) - \nabla W_2 * \rho_2(u)|^2 \, dx du \\
&\leq C_0 \int_0^t \int_{\mathbb{T}^d} |\nabla W_1 * (\rho_1(u) - \rho_2(u))|^2 \, dx du \\
&\quad + C_0 \int_0^t \int_{\mathbb{T}^d} |\nabla(W_1 - W_2) * \rho_2(u)|^2 \, dx du \\
&\leq C_0 \|\nabla W_1\|^2_{L^2(\mathbb{T}^d)} \int_0^t \|\rho_1(u) - \rho_2(u)\|^2_{L^1(\mathbb{T}^d)} \, du \\
&\quad + C_0 \|\nabla(W_1 - W_2) * \rho_2\|^2_{L^2([0,t]\times\mathbb{T}^d)} \\
&\leq 2C_0 \|W_1\|^2_{H^1} \int_0^t H(\rho_1(u)|\rho_2(u)) \, du + C_0 \|\nabla(W_1 - W_2) * \rho_2\|^2_{L^2([0,t]\times\mathbb{T}^d)}.
\end{aligned}
$$

Setting $\eta(t) = \int_0^t H(\rho_1(u)|\rho_2(u)) \, du$, the above estimate can be written as

$$\dot{\eta}(t) \leq 2C_0 M^2 \eta(t) + C_0 \|\nabla(W_1 - W_2) * \rho_2\|^2_{L^2([0,t]\times\mathbb{T}^d)}.$$

The result now follows using the differential form of Gronwall's inequality [16, Sec. B.2]. $\square$

The second part of the following lemma shows that the map $W \mapsto \rho_W$ is β-smoothing, suggesting a mildly ill-posed inverse problem. We have not attempted to optimise β here but believe it close to being sharp.

Lemma 3. *Let $W_i \in W^{2,\infty}(\mathbb{T}^d)$, $\phi \in H^{3+d}(\mathbb{T}^d) \cap \mathcal{P}(\mathbb{T}^d)$ and consider the corresponding solutions ρ_{W_i} to (1). Suppose $\|W_i\|_{W^{2,\infty}} \leq M$ for $i = 1, 2$.*

(i) Then for some constant $C = C(d, T, M, \|\phi\|_{H^d})$,

$$\|\rho_{W_1} - \rho_{W_2}\|^2_{L^2(\mathcal{X},\lambda)} \leq C \int_0^T \|\nabla(W_1 - W_2) * \rho_{W_2}(t)\|^2_{L^2(\mathbb{T}^d)} dt.$$

(ii) If in addition $\phi \in H^\beta(\mathbb{T}^d)$, where $\beta \geq 3 + d$ is an even integer, then

$$\|\rho_{W_1} - \rho_{W_2}\|_{L^2(\mathcal{X},\lambda)} \leq C\|W_1 - W_2\|_{H^{-\beta}} \tag{19}$$

for some constant $C(d, \beta, T, M, \|\phi\|_{H^\beta})$.

Proof. (i) For any densities p, q, recall that the Hellinger distance $h(p, q) = \int(\sqrt{p} - \sqrt{q})^2$ satisfies $h^2(p, q) \leq H(p|q)$ (e.g. Lemma B.1 in [21]). Moreover, writing $(p - q)^2 = (\sqrt{p} + \sqrt{q})^2(\sqrt{p} - \sqrt{q})^2$ yields $\|p - q\|^2_{L^2} \leq 4\max(\|p\|_\infty, \|q\|_\infty)h^2(p, q)$. Using these facts and that $\rho_{W_i}(t)$ are probability densities for all $t \geq 0$,

$$\frac{1}{T}\int_0^T \|\rho_{W_1}(t, \cdot) - \rho_{W_2}(t, \cdot)\|^2_{L^2(\mathbb{T}^d)} dt \leq \frac{4}{T}\max_{i=1,2}\|\rho_{W_i}\|_{L^\infty(\mathcal{X})} \int_0^T H(\rho_{W_1}(t, \cdot)|\rho_{W_2}(t, \cdot)) dt.$$

14

Using (18) and Lemma 1 then gives the result.

(ii) Let $e_m(x) = e^{2\pi i m.x}$, $m \in \mathbb{Z}^d$, denote the usual trigonometric polynomials. Since $\langle f(x-\cdot), e_m \rangle_{L^2(\mathbb{T}^d)} = -e_{-m}(x)\langle f, e_{-m} \rangle_{L^2(\mathbb{T}^d)}$, Parseval's theorem yields

$$
|\partial_{x_i}(W_1 - W_2) * \rho_{W_2}(t, \cdot)(x)|^2 \leq 4\pi^2 \left(\sum_{m \in \mathbb{Z}^d} |m_i| |\langle W_1 - W_2, e_m \rangle_{L^2}| |\langle \rho_{W_2}(t, \cdot), e_m \rangle_{L^2}| \right)^2
$$
$$
\leq 4\pi^2 \sum_{m \in \mathbb{Z}^d} |m_i|^2 (1 + |m|^2)^{-1-\beta} |\langle W_1 - W_2, e_m \rangle_{L^2}|^2
$$
$$
\times \sum_{m \in \mathbb{Z}^d} (1 + |m|^2)^{1+\beta} |\langle \rho_{W_2}(t, \cdot), e_m \rangle_{L^2}|^2
$$

for any $x \in \mathbb{T}^d$. This implies that

$$
|\nabla(W_1 - W_2) * \rho_{W_2}(t, \cdot)(x)|^2 \leq 4\pi^2 \|\rho_{W_2}(t, \cdot)\|^2_{H^{1+\beta}} \|W_1 - W_2\|^2_{H^{-\beta}}.
$$

By the regularity estimate (27) in Theorem 5, $\|\rho_{W_2}\|_{L^2([0,T];H^{\beta+1}(\mathbb{T}^d))} \leq C(d, \beta, T, M, \|\phi\|_{H^\beta}) < \infty$. Integrating the second last display in space and time, and substituting this into the conclusion of (i) then implies

$$
\|\rho_{W_1} - \rho_{W_2}\|^2_{L^2(\mathcal{X})} \leq C \int_0^T \|\rho_{W_2}(t, \cdot)\|^2_{H^{\beta+1}} dt \, \|W_1 - W_2\|^2_{H^{-\beta}} \leq C' \|W_1 - W_2\|^2_{H^{-\beta}}.
$$

$\square$

The estimate given in Lemma 3(ii) is uniform over $\|\cdot\|_{\mathcal{R}} = \|\cdot\|_{W^{2,\infty}}$-balls for W, which verifies the Lipschitz estimate (2.4) in Condition 2.1.1 of [44] with $\kappa = \beta$. We may thus apply Theorem 2.2.2 of [44], which completes the proof of Theorem 1.

3.2   Proof of Theorems 2 and 3: stability estimate and deconvolution

Proof of Theorem 2. Let $\mathscr{L}_{W,n}\psi := \Delta\psi + \nabla\cdot(\psi\nabla W * \rho_{n-1})$, for smooth ψ, denote the operator defined in the linearized McKean-Vlasov equation (25). We will prove the result for the linearized solutions before passing to the limit $n \to \infty$. Further define the linear parabolic operator $\mathscr{P}_{W,n} = (\partial/\partial t - \mathscr{L}_{W,n})$, which satisfies $\mathscr{P}_{W,n}\rho_{W,n} = 0$ where $\rho_{W,n}$ is a solution to (25), see Theorem 4.

Step 1: Reduction to deconvolution with the particle density. Let ρ_W, ρ_{W_0} be two solutions of (1) for distinct W, W_0, and denote by $\rho_{W,n}, \rho_{W_0,n}$ their n-approximated linearised solutions. As the $\mathscr{P}_{W,n}$ operators are linear, we can write

$$
\mathscr{P}_{W,n}(\rho_{W,n} - \rho_{W_0,n}) = 0 - \frac{\partial\rho}{\partial t}\rho_{W_0,n} + \mathscr{L}_{W,n}\rho_{W_0,n}
$$
$$
= (\mathscr{L}_{W,n} - \mathscr{L}_{W_0,n})\rho_{W_0,n}
$$
$$
= \nabla\cdot(\rho_{W_0,n}\nabla(W * \rho_{W,n-1} - W_0 * \rho_{W_0,n-1})). \tag{20}
$$

The densities $\rho_{W_0,n}(t, \cdot)$ are strictly positive throughout $[0, T] \times \mathbb{T}^d$ in view of (26). Hence by standard theory for periodic elliptic equations, the operator $L_{0,n,t}\psi := \nabla\cdot(\rho_{W_0,n}(t)\nabla\psi)$ is invertible on $L_0^2(\mathbb{T}^d) := L^2(\mathbb{T}^d) \cap \{\int_{\mathbb{T}^d} f = 0\}$ for every t: if we select the integral zero solution

15

$u = L_{0,n,t}^{-1} h$ to the elliptic equation $L_{0,n,t} u = h$, we have a uniform in n, t elliptic regularity estimate (proved just as on p.127-128 in [44]),

$$\|L_{0,n,t}^{-1} h\|_{L^2} \leq c \|h\|_{H^{-1}}, \quad h \in L_0^2(\mathbb{T}^d),$$

with c depending only on the (universal) constant in the Poincaré inequality on $\mathbb{T}^d$ as well as the lower bound ϕ_{min} from (26). We also notice that for any t,

$$\int_{\mathbb{T}^d} \mathscr{P}_{W,n}(\rho_{W,n} - \rho_{W_0,n}) = \int_{\mathbb{T}^d} (\mathscr{L}_{W,n} - \mathscr{L}_{W_0,n}) \rho_{W_0,n} = 0$$

as both $\mathscr{L}_{W_n}$ and $\mathscr{L}_{W_0,n}$ are in divergence form. Therefore we can apply the inverse of $L_{0,n,t}$ to (20) and obtain

$$W * \rho_{W,n-1} - W_0 * \rho_{W_0,n-1} = L_{0,n,t}^{-1}[\mathscr{P}_{W,n}(\rho_{W,n} - \rho_{W_0,n})], \quad \text{on } [0, T] \times \mathbb{T}^d,$$

where we note that the left hand side integrates to zero since $\int W = \int W_0, \int \rho_{W_0,n-1} = \int \rho_{W,n-1} = 1$. Rearranging we obtain the identity

$$(W - W_0) * \rho_{W_0,n-1} = L_{0,n,t}^{-1}[\mathscr{P}_{W,n}(\rho_{W,n} - \rho_{W_0,n})] + W * (\rho_{W_0,n-1} - \rho_{W,n-1}), \quad \text{on } [0, T] \times \mathbb{T}^d.$$

We can now fix any time horizon $0 < t_0 \leq T$ and estimate

$$\int_0^{t_0} \|(W - W_0) * \rho_{W_0,n-1}(t)\|_{L^2(\mathbb{T}^d)}^2 dt$$

$$\lesssim \int_0^{t_0} \|L_{0,n,t}^{-1}[\mathscr{P}_{W,n}(\rho_{W,n} - \rho_{W_0,n})(t)]\|_{L^2(\mathbb{T}^d)}^2 dt + \int_0^{t_0} \|W * (\rho_{W_0,n-1} - \rho_{W,n-1})(t)\|_{L^2(\mathbb{T}^d)}^2 dt$$

$$\lesssim \int_0^{t_0} \|(\partial/\partial t - \mathscr{L}_{W,n})(\rho_{W,n} - \rho_{W_0,n})(t)\|_{H^{-1}(\mathbb{T}^d)}^2 dt + \int_0^{t_0} \|\rho_{W_0,n-1} - \rho_{W,n-1}(t)\|_{L^2(\mathbb{T}^d)}^2 dt$$

$$\lesssim \|\rho_{W,n} - \rho_{W_0,n}\|_{H^1([0,t_0],H^{-1})}^2 + \|\rho_{W,n} - \rho_{W_0,n}\|_{L^2([0,t_0],H^1)}^2. \tag{21}$$

We can take limits in this inequality: this is clear for the left hand side by Lemma 4 and since convolution with $W - W_0$ is uniformly Lipschitz on $L^2(\mathbb{T}^d)$. For the right hand side of (21) we notice that by Theorem 5 and Rellich's compactness theorem (p.305 in [17]), the sequences $\{\rho_{W,n} - \rho_{W_0,n} : n \in \mathbb{N}\}$ converge in $H^1([0,t_0], H^{-1})$ and $L^2([0,t_0], H^1)$ along a subsequence to their unique limit, which in view of Lemma 4 equals $\rho_W - \rho_{W_0}$. As the constants implied by $\lesssim$ in (21) are n-independent, we deduce for any $0 < t_0 \leq T$ that

$$\int_0^{t_0} \|(W - W_0) * \rho_{W_0}(t)\|_{L^2(\mathbb{T}^d)}^2 dt \leq C\left(\|\rho_W - \rho_{W_0}\|_{H^1([0,t_0],H^{-1})}^2 + \|\rho_W - \rho_{W_0}\|_{L^2([0,t_0],H^1)}^2\right) \tag{22}$$

for a constant $C = C(T, d, M, \phi_{min}, \|\phi\|_{H^\beta}) > 0$.

Step 2. Deconvolving at small times. Let $W_{0,K} = \sum_{k \in \mathbb{Z}^d : |k| \leq K} \langle W_0, e_k \rangle_{L^2} e_k$ denote the Fourier projection of $W_0 \in H^{\alpha+1}$ onto the first K frequencies. By Parseval's theorem and since $\int W = \int W_0$ we have

$$\|W - W_0\|_{L^2}^2 \leq \|W_0\|_{H^{\alpha+1}}^2 K^{-2(\alpha+1)} + \|W - W_{0,K}\|_{L^2}^2$$

$$\lesssim K^{-2\alpha-2} + \sum_{0 < |k| \leq K} |\hat{W}(k) - \hat{W}_0(k)|^2 \frac{|\hat{\rho}_{W_0,n}(t,k)|^2}{|\hat{\rho}_{W_0,n}(t,k)|^2} \tag{23}$$

$$\lesssim K^{-2\alpha-2} + \sup_{0 < |k| \leq K} \frac{1}{|\hat{\rho}_{W_0,n}(t,k)|^2} \|(W - W_0) * \rho_{W_0,n}(t, \cdot)\|_{L^2}^2$$

16

For $t > 0$, define the stability constant

$$\iota_{K,n}(t) \equiv \sup_{0 < |k| \leq K} \frac{1}{|\hat{\rho}_{W_0,n}(t,k)|^2}, \tag{24}$$

which we now upper bound. By Theorem 5 we know

$$\left\| \frac{d\rho_{W_0,n}}{dt} \right\|_{L^\infty([0,T];H^\beta)} \leq C(d,T,\beta,M,\|\phi\|_{H^\beta}),$$

and hence $t \mapsto \rho_{W_0,n}(t)$ is Lipschitz as a map from $[0,T] \to L^1(\mathbb{T}^d)$ which implies $\|\rho_{W_0,n}(t) - \phi\|_{L^1(\mathbb{T}^d)} \leq Ct$. But since L^1-norms bound the Fourier coefficients uniformly, for any $k \in \mathbb{Z}^d$,

$$|\hat{\rho}_{W_0,n}(k)| \geq |\hat{\phi}(k)| - |\hat{\rho}_{W_0,n}(k) - \hat{\phi}(k)| \geq c|k|^{-\zeta} - Ct.$$

Thus for any $0 < t \leq t_0 = \frac{c}{2C}K^{-\zeta}$, the stability constant in (24) satisfies $\iota_{K,n}(t) \leq \frac{c^2}{4}K^{2\zeta}$.

Integrating the inequality (23) in time then gives

$$\int_0^{t_0} \|W - W_0\|_{L^2}^2 dt \lesssim t_0 K^{-2\alpha-2} + K^{2\zeta} \int_0^{t_0} \|(W - W_0) * \rho_{W_0,n}(t,\cdot)\|_{L^2}^2 dt$$

and hence by (21),

$$\|W - W_0\|_{L^2}^2 \lesssim K^{-2\alpha-2} + t_0^{-1} K^{2\zeta} (\|\rho_W - \rho_{W_0}\|_{H^1([0,t_0],H^{-1})}^2 + \|\rho_W - \rho_{W_0}\|_{L^2([0,t_0],H^1)}^2),$$

completing the proof. $\qquad\square$

Proof of Theorem 3. The proof follows from combining the contraction rate for the forward map from Example 2 with the stability estimate from Theorem 2. By Example 2, the posterior concentrates on the set

$$\overline{\mathcal{W}}_N = \{ W \in E_{K_N} : \|\rho_W - \rho_{W_0}\|_{L^2(\mathcal{X})} > L\delta_N, \ \|W\|_{W^{2,\infty}} \leq M \},$$

where $\delta_N = N^{-\frac{\alpha+1+\beta}{2(\alpha+1)+2\beta+d}}$, so it suffices to prove the result for $W \in \overline{\mathcal{W}}_N$. Using the interpolation inequality (A.5) in [44], $\|u\|_{H^1(\mathbb{T}^d)} \lesssim \|u\|_{L^2(\mathbb{T}^d)}^{1-1/\gamma} \|u\|_{H^\gamma(\mathbb{T}^d)}^{1/\gamma}$ for any $\gamma > 1$. By (27) of Theorem 5, $\|\rho_W\|_{L^2([0,T];H^{\beta+1})} \leq C$ for any $W \in \overline{\mathcal{W}}_N$, and hence applying the interpolation with $\gamma = \beta + 1$ and using Hölder's inequality yields

$$\|\rho_W - \rho_{W_0}\|_{L^2([0,T];H^1)}^2 \lesssim \left(\int_0^T \|\rho_W(t) - \rho_{W_0}(t)\|_{L^2(\mathbb{T}^d)}^2 dt \right)^{\frac{\gamma-1}{\gamma}} = \|\rho_W - \rho_{W_0}\|_{L^2(\mathcal{X})}^{\frac{2\beta}{\beta+1}}.$$

Turning to the time regularity, using Proposition 2.1 in Ch. 4 in [38], we have the interpolation inequality

$$\|v\|_{H^1([0,T];L^2)} \lesssim \|v\|_{L^2([0,T];L^2)}^{(m-1)/m} \|v\|_{H^m([0,T];L^2)}^{1/m}$$

for any integer $m \geq 1$. But by (28) of Theorem 5, $\|\frac{d^\ell \rho_W}{dt^\ell}\|_{L^2([0,T];H^1)} \leq C$ for all $\ell = 0, 1, \ldots, \beta/2$ and hence we may take $m = \beta/2$ to obtain

$$\|\rho_W - \rho_{W_0}\|_{H^1([0,T];H^{-1})} \lesssim \|\rho_W - \rho_{W_0}\|_{L^2([0,T];L^2)}^{\frac{m-1}{m}} = \|\rho_W - \rho_{W_0}\|_{L^2(\mathcal{X})}^{\frac{\beta-2}{\beta}}.$$

17

Apply Theorem 2 with $t_0 \simeq K_N^{-\zeta}$ gives that for any $W \in \overline{\mathcal{W}}_N \subset E_{K_N}$,

$$\|W - W_0\|_{L^2(\mathbb{T}^d)}^2 \lesssim K_N^{-2\alpha-2} + K_N^{3\zeta}\left(\delta_N^{\frac{2(\beta-2)}{\beta}} + \delta_N^{\frac{2\beta}{\beta+1}}\right) \lesssim K_N^{-2\alpha-2} + K_N^{3\zeta}\delta_N^{\frac{2(\beta-2)}{\beta}}.$$

Substituting in $K_N \simeq N^{\frac{1}{2(\alpha+1)+2\beta+d}}$ and δ_N shows that the second term dominates for all $\alpha > 0$ and $\beta \geq 3 + d$ since $\zeta > \beta + d/2$. The convergence of the posterior mean vector now follows from arguments as in Theorem 2.3.2 in [44] and the details are left to the reader. $\square$

3.3 Regularity properties for the McKean-Vlasov PDE

In this section, we collect properties and regularity estimates for the nonlinear parabolic McKean-Vlasov PDE (1) driving the dynamics of our model, as well as its linearization. Recall that we work on the torus $\mathbb{T}^d = \mathbb{R}^d/\mathbb{Z}^d = (0,1]^d$ and that $\mathcal{P} = \mathcal{P}(\mathbb{T}^d)$ denotes the set of probability densities (with respect to Lebesgue measure) on $\mathbb{T}^d$. The basic existence and uniqueness of classical solutions to (1) follows from [7, 30]; see also [28, Thm. 2].

Theorem 4 ([7], Theorem 2.2). *If $W \in W^{2,\infty}(\mathbb{T}^d)$ and $\phi \in H^{3+d}(\mathbb{T}^d) \cap \mathcal{P}(\mathbb{T}^d)$, then there exists a classical solution $\rho : [0,T] \times \mathbb{T}^d \to [0,\infty)$ such that $\rho(t,\cdot) \in \mathcal{P}(\mathbb{T}^d) \cap C^2(\mathbb{T}^d)$ for all $t > 0$.*

Similar to [63, 8, 7], we approximate the McKean-Vlasov equation (1) by a sequence of linear parabolic PDEs. For $n = 1, 2, \dots$, consider the sequence of linear equations on $\mathbb{T}^d$

$$\frac{\partial \rho_n}{\partial t} = \Delta \rho_n + \nabla \cdot (\rho_n \nabla W * \rho_{n-1}) =: \mathscr{L}_{W,n}\rho_n, \tag{25}$$
$$\rho(0,\cdot) = \phi,$$

where $\phi \in H^\beta(\mathbb{T}^d) \cap \mathcal{P}(\mathbb{T}^d)$ and we take a smooth time-independent initialization $\rho_0 \in C^\infty((\mathbb{T}^d))\mathcal{P}(\mathbb{T}^d)$. One then shows as in [63, 8, 7]:

Lemma 4. *Let $W \in W^{2,\infty}(\mathbb{T}^d)$ and $\phi \in H^{3+d}(\mathbb{T}^d) \cap \mathcal{P}(\mathbb{T}^d)$. As $k \to \infty$ we have (if necessary along a subsequence)*

$$\sup_{0 \leq t \leq T} \|\rho_{W,k}(t) - \rho_W(t)\|_{H^1} \to 0.$$

We now establish quantitative regularity estimates which hold uniformly over potentials W of bounded norm, which was crucial in our proofs above. It builds on the results of [8, 30]. We remark that the smoothness of ρ_W is driven by the regularity of the initial condition ϕ, while the regularity of W is less crucial as it factors into the estimates only after convolution with ρ_{n-1}, which is a smoothing operation.

Theorem 5. *Let $\beta \geq 3+d$ be an integer and suppose $W \in W^{2,\infty}(\mathbb{T}^d)$ satisfies $\|W\|_{W^{2,\infty}} \leq M$. Let further $\phi \in H^\beta(\mathbb{T}^d) \cap \mathcal{P}(\mathbb{T}^d)$ be a strictly positive probability density $\inf_x \phi(x) \geq \phi_{min} > 0$ and let $\rho_0 \in C^\infty(\mathbb{T}^d) \cap \mathcal{P}(\mathbb{T}^d)$ be strictly positive. Then the sequence of PDEs (25) has a unique solution $\rho_n \in C^{1,2}([0,T] \times \mathbb{T}^d)$, which is strictly positive*

$$\inf_{x \in \mathbb{T}^d, t \in [0,T]} \rho_n(t,x) \geq \lambda(\phi_{min}, M, T) > 0, \tag{26}$$

for a constant λ that depends only on ϕ_{min}, M, T, and satisfies $\int_{\mathbb{T}^d} \rho_n(t,x)\,dx = 1$ for all $t \in [0,T]$. For $k = 0, 1, \dots, \beta - 1$, we have the following regularity estimate in space:

$$\|\rho_n\|_{L^2([0,T];H^{k+2}(\mathbb{T}^d))} + \|\rho_n\|_{L^\infty([0,T];H^{k+1}(\mathbb{T}^d))} \leq C(d,k,T,M,\|\phi\|_{H^{k+1}}). \tag{27}$$

18

Furthermore, if $\beta = 2j$ is an even integer, we have the time regularity

$$\frac{d^\ell \rho_n}{dt^\ell} \in L^2([0,T]; H^{2j-2\ell+1}(\mathbb{T}^d)) \cap L^\infty([0,T]; H^{2j-2\ell}(\mathbb{T}^d)) \quad for\ 0 \leq \ell \leq j,$$

with the estimate

$$\sum_{\ell=0}^{j} \left(\left\| \frac{d^\ell \rho_n}{dt^\ell} \right\|_{L^2([0,T]; H^{2j-2\ell+1}(\mathbb{T}^d))} + \left\| \frac{d^\ell \rho_n}{dt^\ell} \right\|_{L^\infty([0,T]; H^{2j-2\ell}(\mathbb{T}^d))} \right) \tag{28}$$
$$\leq C(d, j, T, M, \|\phi\|_{H^{2j}}).$$

Moreover, all the above estimates hold with ρ_n replaced by ρ, the solution to the McKean-Vlasov equation (1).

Proof. The existence of a unique solution ρ_n to (25) that is a probability density follows from standard linear parabolic PDE theory, e.g. [16, Ch. 7]. The basic regularity in time and space follow from Theorems 27.2-27.3 (time) and Theorem 27.5 (space) of [64], see also [40, Thm. 2, Thm. 3]. In particular, since ρ_0 is smooth, for $n = 1$ we have a linear parabolic PDE (written in non-divergence form) with smooth coefficients $\nabla W * \rho_0 \in C^\infty$ and $\Delta W * \rho_0 \in C^\infty$. Furthermore, the compatibility conditions are automatically satisfied since we have periodic boundary conditions. Therefore, from [40, Thm. 2] it follows that $\rho_1 \in L^2([0,T]; H^{\beta+1}(\mathbb{T}^d)) \cap H^{\frac{\beta+1}{2}}([0,T]; L^2(\mathbb{T}^d))$. From this we deduce that we have the same space/time regularity as ρ_1 for $\nabla W * \rho_{n-1}$ and of $\Delta W * \rho_{n-1}$, since $W \in W^{2,\infty}$. Applying again [40, Thm. 2] we conclude the existence and uniqueness of solutions $\rho_n \in L^2([0,T]; H^{\beta+1}(\mathbb{T}^d)) \cap H^{\frac{\beta+1}{2}}([0,T]; L^2(\mathbb{T}^d))$ to (25), which justifies taking any derivatives we need in what follows. It remains to show the quantitative regularity estimates, which we show hold uniformly in n and over potentials $\|W\|_{W^{2,\infty}} \leq M$.

We start with space regularity and follow ideas in [8, 30]. We prove (27) by induction on k, starting with $k = 0$. We multiply (25) by ρ_n, integrate in x, do the standard integration by parts, use Hölder and Young inequalities to deduce

$$\begin{aligned}
\frac{1}{2}\frac{d}{dt}\|\rho_n\|_{L^2}^2 + \|\nabla \rho_n\|_{L^2}^2 &= \langle \rho_n, \frac{d}{dt}\rho_n - \Delta \rho_n \rangle_{L^2} \\
&= -\langle \nabla \rho_n, \rho_n \nabla W_{n-1} \rangle_{L^2} \\
&\leq \frac{1}{2}\|\nabla \rho_n\|_{L^2}^2 + \frac{1}{2}\|\rho_n\|_{L^2}^2 \|\nabla W_{n-1}\|_{L^2}^2,
\end{aligned}$$

where we have used the notation $W_{n-1} := W * \rho_{n-1}$. Since $\|\nabla W_{n-1}\|_{L^2} \leq \|\nabla W\|_{L^2} \|\rho_{n-1}\|_{L^1} \leq M$ by Young's convolution inequality, we deduce

$$\frac{1}{2}\frac{d}{dt}\|\rho_n\|_{L^2}^2 + \frac{1}{2}\|\nabla \rho_n\|_{L^2}^2 \leq \frac{M^2}{2}\|\rho_n\|_{L^2}^2 \tag{29}$$

for all $t \in [0,T]$, where we recall $\rho_n = \rho_n(t, \cdot)$. Using the differential form of Gronwall's inequality ([16], Section B.2), we have

$$\|\rho_n(t, \cdot)\|_{L^2}^2 \leq e^{M^2 T}\|\rho(0, \cdot)\|_{L^2}^2 = e^{M^2 T}\|\phi\|_{L^2}^2 \tag{30}$$

for all $t \in [0,T]$. Furthermore, integrating (29) in time and using (30),

$$\sup_{t \in [0,T]} \|\rho_n(t, \cdot)\|_{L^2}^2 + \int_0^T \|\nabla \rho_n(t, \cdot)\|_{L^2}^2 \, dt \leq (M^2 T e^{M^2 T} + 1)\|\phi\|_{L^2}^2. \tag{31}$$

19

We proceed in similar manner with the higher regularity estimates. For the H^2-in-space estimate, we multiply (25) by $\nabla \rho_n$, integrate by parts, use Hölder's and Young's inequalities and the fact that $\|\rho_{n-1}\|_{L^1} = 1$ to deduce that

$$
\begin{aligned}
\frac{1}{2}\frac{d}{dt}\|\nabla\rho_n\|_{L^2}^2 + \|\Delta\rho_n\|_{L^2}^2 &= \int_{\mathbb{T}^d} \nabla\cdot(\rho_n\nabla W_{n-1})\,\Delta\rho_n\,dx \\
&\leq \frac{1}{2}\|\Delta\rho_n\|_{L^2}^2 + \frac{1}{2}\int_{\mathbb{T}^d}\left|\nabla\cdot(\rho_n\nabla W_{n-1})\right|^2 dx \\
&\leq \frac{1}{2}\|\Delta\rho_n\|_{L^2}^2 + \|\nabla\rho_n\cdot\nabla W_{n-1}\|_{L^2}^2 + \|\rho_n\Delta W_{n-1}\|_{L^2}^2 \\
&\leq \frac{1}{2}\|\Delta\rho_n\|_{L^2}^2 + \|\nabla W_{n-1}\|_{L^\infty}^2\|\nabla\rho_n\|_{L^2}^2 + \|\Delta W_{n-1}\|_{L^\infty}^2\|\rho_n\|_{L^2}^2.
\end{aligned}
$$

Using (29)-(30), we similarly obtain $\frac{d}{dt}\|\nabla\rho_n\|_{L^2}^2 + \|\Delta\rho_n\|_{L^2}^2 \leq C(M,T)\|\phi\|_{L^2}^2$. Integrating in time and using (31) and Poincaré's inequality on the torus,

$$
\sup_{t\in[0,T]}\|\nabla\rho_n\|_{L^2}^2 + c\int_0^T \|\rho_n\|_{H^2}^2\,dt \leq C(T,M)\|\phi\|_{H^1}^2. \tag{32}
$$

This establishes (27) for $k=0$.

For the higher regularity estimates (27) with integers $1 \leq k \leq \beta - 1$, we proceed by induction on k as in [8, 30]. Assume that (27) holds for $k-1$. We differentiate (25) by ∂_ν for some multi-index ν with $|\nu| = k$, multiply by $\Delta\partial_\nu\rho_n$, integrate by parts and use Young's inequality to obtain

$$
\begin{aligned}
\frac{1}{2}\frac{d}{dt}\|\nabla\partial_\nu\rho_n\|_{L^2}^2 + \|\Delta\partial_\nu\rho_n\|_{L^2}^2 &= \int_{\mathbb{T}^d}\left|\Delta\partial_\nu\rho_n\partial_\alpha\nabla\cdot(\rho_n\nabla W_{n-1})\right|dx \\
&\leq \frac{1}{2}\|\Delta\partial_\nu\rho_n\|_{L^2}^2 + C_d\sum_{i=1}^d\|\rho_n(\partial_{x_i}W * \rho_{n-1})\|_{H^{k+1}}^2.
\end{aligned} \tag{33}
$$

But by Lemma 5,

$$
\|\rho_n(\partial_{x_i}W * \rho_{n-1})\|_{H^{k+1}}^2 \leq C(d,k,M)\|\rho_n\|_{H^{k+1}}\|\rho_{n-1}\|_{H^k}.
$$

Substituting this estimate into (33), summing over all $\nu = |k|$, integrating over $[0,T]$ and using the induction hypothesis,

$$
\begin{aligned}
\sup_{t\in[0,T]}\|\rho_n(t)\|_{H^{k+1}}^2 + \int_0^T \|\rho_n(t)\|_{H^{k+2}}^2 dt &\leq C(d,k,M)\int_0^T\|\rho_n(t)\|_{H^{k+1}}^2\|\rho_{n-1}(t)\|_{H^k}dt + \|\phi\|_{H^{k+1}}^2 \\
&\leq C(d,k,M,T,\|\phi\|_{H^{k+1}},\|\rho_0\|_{H^{k+1}}),
\end{aligned}
$$

which establishes (27) for k as desired. Now since the ρ_n are uniformly bounded in the relevant Hilbert space norms, the Banach-Alaoglu theorem implies that they converge weakly in these Hilbert spaces to their limits along a subsequence. Combined with the inequality $\|h\|_H \leq \liminf_n \|h_n\|_H$ for any weakly convergent sequence $h_n \to h$ in a Hilbert space, we may take limits in the last inequalities, to give the following apriori estimate for the solution to the McKean-Vlasov PDE

$$
\|\rho\|_{L^2([0,T];H^{k+2}(\mathbb{T}^d))} + \|\rho\|_{L^\infty([0,T];H^{k+1}(\mathbb{T}^d))} \leq C(T,\rho_0,M). \tag{34}
$$

We now consider regularity in time. For $\beta = 2j$, we have that $\rho_n \in L^2([0,T]; H^{2j+1}(\mathbb{T}^d)) \cap H^{j+\frac{1}{2}}([0,T]; L^2(\mathbb{T}^d))$ with $\partial_t^{j-1} \rho_n \in C([0,T]; H^1(\mathbb{T}^d))$, see [40, Thm. 2]. To get the quantitative estimates, we follow the proof of [8, Theorem 4.3], proceeding by induction on the number of time-derivatives ℓ. For $\ell = 0$, this follows immediately from (27). Suppose now we have the time regularity estimate (28) for some integer $0 \le \ell < j$. We differentiate (25) ℓ-times in t to obtain, using the notation $\rho^{(\ell)} := \frac{d^\ell \rho}{dt^\ell}$,

$$
\begin{aligned}
\rho_n^{(\ell+1)} &= \Delta \rho_n^{(\ell)} + \nabla \cdot (\rho_n \nabla W * \rho_{n-1})^{(\ell)} \\
&= \Delta \rho_n^{(\ell)} + \sum_{k=0}^{\ell} \binom{\ell}{k} \nabla \cdot \left(\rho_n^{(k)} \nabla W * \rho_{n-1}^{(\ell-k)} \right).
\end{aligned}
$$

Taking the $H^{2j-2\ell-1}(\mathbb{T}^d)$-norm in space and using Lemma 5 since $2j - 2\ell \ge 2$,

$$
\begin{aligned}
\|\rho_n^{(\ell+1)}\|_{H^{2j-2\ell-1}(\mathbb{T}^d)}^2 &\le 2\|\rho_n^{(\ell)}\|_{H^{2j-2\ell+1}(\mathbb{T}^d)}^2 + C_\ell \sum_{k=0}^{\ell} \|\rho_n^{(k)} \nabla W * \rho_{n-1}^{(\ell-k)}\|_{H^{2j-2\ell}(\mathbb{T}^d)}^2 \\
&\le 2\|\rho_n^{(\ell)}\|_{H^{2j-2\ell+1}(\mathbb{T}^d)}^2 + C_{j,M} \sum_{k=0}^{\ell} \|\rho_n^{(k)}\|_{H^{2j-2\ell}(\mathbb{T}^d)}^2 \|\rho_{n-1}^{(\ell-k)}\|_{H^{2j-2\ell-1}(\mathbb{T}^d)}^2,
\end{aligned}
$$

where C depends only on j and $\|W\|_{W^{2,\infty}(\mathbb{T}^d)} \le M$. Integrating over time $t \in [0,T]$ then gives

$$
\begin{aligned}
\|\rho_n^{(\ell+1)}\|_{L^2(0,T; H^{2j-2\ell-1}(\mathbb{T}^d))}^2 &\le 2\|\rho_n^{(\ell)}\|_{L^2(0,T; H^{2j-2\ell+1}(\mathbb{T}^d))}^2 \\
&\quad + C_{j,M} \sum_{k=0}^{\ell} \|\rho_n^{(k)}\|_{L^2(0,T; H^{2j-2\ell}(\mathbb{T}^d))}^2 \|\rho_{n-1}^{(\ell-k)}\|_{L^\infty(0,T; H^{2j-2\ell-1}(\mathbb{T}^d))}^2.
\end{aligned}
$$

Since $0 \le k, \ell - k \le \ell$, we may now apply the inductive hypothesis (28) to conclude that

$$
\|\rho_n^{(\ell+1)}\|_{L^2(0,T; H^{2j-2\ell-1}(\mathbb{T}^d))}^2 \le C(d,j,T, \|W\|_{W^{2,\infty}}, \|\phi\|_{H^{2j}}, \|\rho_0\|_{H^{2j}}),
$$

which is the first part of the required estimate for $\ell + 1$. Similarly, taking the $H^{2j-2\ell-2}$-norm in space and arguing as above yields

$$
\|\rho_n^{(\ell+1)}\|_{L^\infty(0,T; H^{2j-2\ell-2}(\mathbb{T}^d))}^2 \le C(d,j,T, \|W\|_{W^{2,\infty}}, \|\phi\|_{H^{2j}}, \|\rho_0\|_{H^{2j}}),
$$

which proves the desired estimate for $\ell + 1$. This proves (28) for $\ell = 0, \ldots, j$ as required, based on the smoothness of the initial condition $\phi \in H^{2j}$. As before, we can then pass to the limit $n \to \infty$ to obtain the estimates for the mean field PDE.

To obtain the lower bound (26) to the solution to the McKean-Vlasov PDE, we follow the proof of [36][Lem. 5.1] that is based on the argument presented in [28][Thm. 2]. Fix $T' \in (0,T)$, let $\rho_t(x) = \rho(t,x)$ denote the solution to the McKean-Vlasov PDE (1), and consider the unique strong solution of the *linear* (in the sense of McKean) SDE

$$
dY_t = \nabla W * \rho_{T'-t}(Y_t)\, dt + \sqrt{2}\, dB_t, \quad Y_0 = x.
$$

Using Itô's formula, the PDE (1) and taking expectations gives for $t \in [0, T']$,

$$
\mathbb{E}^x \rho_{T'-t}(Y_t) = \rho_{T'}(x) - \mathbb{E}^x \int_0^t \rho_{T'-s}(Y_s) \Delta W * \rho_{T'-s}(Y_s)\, ds, \tag{35}
$$

21

where $\mathbb{E}^x$ denotes the expectation with respect to the law of the process Y_t starting at x. We use now the fact that $\rho_u(x) \geq 0$, $u \in [0, T]$, together with the estimate $\|\Delta W * \rho_u\|_{L^\infty} \leq \|W\|_{W^{2,\infty}} \|\rho_u\|_{L^1} \leq M$ to deduce that

$$\mathbb{E}^x \rho_{T'-t}(Y_t) \leq \rho_{T'}(x) + M \int_0^t \mathbb{E}^x \rho_{T'-s}(Y_s) \, ds.$$

From the integral form of Gronwall's inequality, it follows that for all $t \in [0, T']$,

$$\mathbb{E}^x \rho_{T'-t}(Y_t) \leq \rho_{T'}(x) \left(1 + M t e^{Mt}\right) \leq \rho_{T'}(x) \left(1 + M T e^{MT}\right).$$

Setting now $t = T'$ gives that for any $T' \in (0, T]$,

$$\rho_{T'}(x) \geq \frac{\phi_{min}}{1 + M T e^{MT}},$$

as required. The same argument also applies to the sequence of PDEs (25). We omit the details. $\square$

Lemma 5. *Let* $W \in W^{2,\infty}(\mathbb{T}^d)$, $u \in H^k$ *and* $v \in H^{k-1}$ *for integer* $k \geq 2$. *Then*

$$\|f(\nabla W * g)\|_{H^k} \leq C(k, d, \|W\|_{W^{2,\infty}}) \|f\|_{H^k} \|g\|_{H^{k-1}}. \tag{36}$$

Proof. The proof follows from the estimates in [30], pp. 65-66. $\square$

References

[1] C. Amorino, D. Belomestny, V. Pilipauskaitė, M. Podolskij, and S.-Y. Zhou. Polynomial rates via deconvolution for nonparametric estimation in mckean-vlasov sdes, 2024.

[2] C. Amorino, A. Heidari, V. Pilipauskaitė, and M. Podolskij. Parameter estimation of discretely observed interacting particle systems. *Stochastic Process. Appl.*, 163:350–386, 2023.

[3] A. Beskos, M. Girolami, S. Lan, P. E. Farrell, and A. M. Stuart. Geometric MCMC for infinite-dimensional inverse problems. *J. Comput. Phys.*, 335:327–351, 2017.

[4] J. P. N. Bishwal. Estimation in interacting diffusions: continuous and discrete sampling. *Appl. Math. (Irvine)*, 2(9):1154–1158, 2011.

[5] V. I. Bogachev, M. Röckner, and S. V. Shaposhnikov. Distances between transition probabilities of diffusions and applications to nonlinear Fokker-Planck-Kolmogorov equations. *J. Funct. Anal.*, 271(5):1262–1300, 2016.

[6] J. A. Carrillo, A. Chertock, and Y. Huang. A finite-volume method for nonlinear nonlocal equations with a gradient flow structure. *Commun. Comput. Phys.*, 17(1):233–258, 2015.

[7] 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(1):635–690, 2020.

[8] B. Chazelle, Q. Jiu, Q. Li, and C. Wang. Well-posedness of the limiting equation of a noisy consensus model in opinion dynamics. *J. Differential Equations*, 263(1):365–397, 2017.

[9] F. Comte and V. Genon-Catalot. Nonparametric adaptive estimation for interacting particle systems. *Scand. J. Stat.*, 50(4):1716–1755, 2023.

[10] S. L. Cotter, G. O. Roberts, A. M. Stuart, and D. White. MCMC methods for functions: modifying old algorithms to make them faster. *Statist. Sci.*, 28(3):424–446, 2013.

[11] D. A. Dawson. Critical dynamics and fluctuations for a mean-field model of cooperative behavior. *J. Statist. Phys.*, 31(1):29–85, 1983.

[12] 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(1):91–148, 2021.

[13] M. G. Delgadino, R. S. Gvalani, G. A. Pavliotis, and S. A. Smith. Phase Transitions, Logarithmic Sobolev Inequalities, and Uniform-in-Time Propagation of Chaos for Weakly Interacting Diffusions. *Comm. Math. Phys.*, 401(1):275–323, 2023.

[14] L. Della Maestra and M. Hoffmann. Nonparametric estimation for interacting particle systems: McKean-Vlasov models. *Probab. Theory Related Fields*, 182(1-2):551–613, 2022.

[15] L. Della Maestra and M. Hoffmann. The lan property for mckean–vlasov models in a mean-field regime. *Stochastic Processes and their Applications*, 155:109–146, 2023.

[16] L.C. Evans. *Partial Differential Equations*. AMS, Providence, Rhode Island, 1998.

[17] G.B. Folland. *Real analysis*. Pure and Applied Mathematics. John Wiley & Sons Inc., New York, 1999.

[18] N. Fournier and B. Jourdain. Stochastic particle approximation of the Keller-Segel equation and two-dimensional generalization of Bessel processes. *Ann. Appl. Probab.*, 27(5):2807–2861, 2017.

[19] T. D. Frank. *Nonlinear Fokker-Planck equations*. Springer Series in Synergetics. Springer-Verlag, Berlin, 2005.

[20] V. Genon-Catalot and C. Larédo. Parametric inference for small variance and long time horizon McKean-Vlasov diffusion models. *Electron. J. Stat.*, 15(2):5811–5854, 2021.

[21] Subhashis Ghosal and Aad van der Vaart. *Fundamentals of nonparametric Bayesian inference*, volume 44 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 2017.

[22] K. Giesecke, G. Schwenkler, and J. A. Sirignano. Inference for large financial systems. *Math. Finance*, 30(1):3–46, 2020.

[23] E. Giné and R. Nickl. *Mathematical foundations of infinite-dimensional statistical models*. Cambridge Series in Statistical and Probabilistic Mathematics, [40]. Cambridge University Press, New York, 2016.

[24] M. Giordano and K. Ray. Nonparametric Bayesian inference for reversible multidimensional diffusions. *Ann. Statist.*, 50(5):2872–2898, 2022.

[25] 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*, volume 3 of *Lect. Notes Appl. Math. Mech.*, pages 1–144. Springer, [Cham], 2016.

[26] S. N. Gomes, G. A. Pavliotis, and U. Vaes. Mean Field Limits for Interacting Diffusions with Colored Noise: Phase Transitions and Spectral Numerical Methods. *Multiscale Model. Simul.*, 18(3):1343–1370, 2020.

[27] T. Goodwin, C. Evenhuis, S. Woodcock, and M. Quiroz. Bayesian inference on the

Keller–Segel model. In *Proceedings of the 14th Biennial Engineering Mathematics and Applications Conference, EMAC-2019*, volume 61 of *ANZIAM J.*, pages C181–C196, 2020.

[28] Arnaud Guillin, Pierre Le Bris, and Pierre Monmarché. Uniform in time propagation of chaos for the 2d vortex model and other singular stochastic systems, 2023.

[29] S. Gupta, A. Campa, and S. Ruffo. *Statistical physics of synchronization*. SpringerBriefs in Complexity. Springer, Cham, 2018. With a foreword by Steven Strogatz.

[30] R. S. Gvalani. *Many-particle systems: phase transitions, distinguished limits, and gradient flows*. PhD thesis, Imperial College London, 2020.

[31] R. Hegselmann and U. Krause. Opinion dynamics and bounded confidence: models, analysis and simulation. *J. Artif. Soc. Soc. Simul.*, 5(3), 2002.

[32] K. Hellmuth, C. Klingenberg, Q. Li, and M. Tang. Multiscale convergence of the inverse problem for chemotaxis in the bayesian setting. *Computation*, 9(11), 2021.

[33] Marc Hoffmann and Kolyan Ray. Bayesian estimation in a multidimensional diffusion model with high frequency data. *arXiv preprint 2022*, 2022.

[34] R. A. Kasonga. Maximum likelihood theory for large interacting systems. *SIAM J. Appl. Math.*, 50(3):865–875, 1990.

[35] S. Kotz, T. J. Kozubowski, and K. Podgórski. *The Laplace distribution and generalizations*. Birkhäuser Boston, Inc., Boston, MA, 2001. A revisit with applications to communications, economics, engineering, and finance.

[36] D. Lacker and L. Le Flem. Sharp uniform-in-time propagation of chaos. *Probab. Theory Related Fields*, 187(1-2):443–480, 2023.

[37] S. Lee, Y. M. Psarellis, C. I. Siettos, and I. G Kevrekidis. Learning black-and gray-box chemotactic PDEs/closures from agent based Monte Carlo simulation data. *Journal of Mathematical Biology*, 87(1):15, 2023.

[38] J.-L. Lions and E. Magenes. *Non-homogeneous boundary value problems and applications. Vol. II*. Die Grundlehren der mathematischen Wissenschaften, Band 182. Springer-Verlag, New York-Heidelberg, 1972. Translated from the French by P. Kenneth.

[39] F. Malrieu. Logarithmic Sobolev inequalities for some nonlinear PDE's. *Stochastic Process. Appl.*, 95(1):109–132, 2001.

[40] A. Milani. A remark on the Sobolev regularity of classical solutions to uniformly parabolic equations. *Math. Nachr.*, 199:115–144, 1999.

[41] S. Mischler and C. Mouhot. Kac's program in kinetic theory. *Invent. Math.*, 193(1):1–147, 2013.

[42] F. Monard, R. Nickl, and G. P. Paternain. Consistent inversion of noisy non-Abelian X-ray transforms. *Comm. Pure Appl. Math.*, 74(5):1045–1099, 2021.

[43] R. Nickl. Bernstein–von Mises theorems for statistical inverse problems I: Schrödinger equation. *J. Eur. Math. Soc. (JEMS)*, 22(8):2697–2750, 2020.

[44] R. Nickl. *Bayesian non-linear statistical inverse problems*. EMS Press, Berlin, 2023.

[45] R. Nickl. Consistent inference for multi-dimensional diffusions from low frequency observations. *Annals of Statistics*, page to appear, 2024.

[46] R. Nickl and K. Ray. Nonparametric statistical inference for drift vector fields of multi-dimensional diffusions. *Ann. Statist.*, 48(3):1383–1408, 2020.

[47] R. Nickl and E. S. Titi. On posterior consistency of data assimilation with gaussian process priors: the 2d Navier-Stokes equations. *arXiv preprint 2023*, 2023.

[48] R. Nickl and S. Wang. On polynomial-time computation of high-dimensional posterior measures by Langevin-type algorithms. *J. Eur. Math. Soc. (JEMS)*, 26(3):1031–1112,

2024.

[49] K. Oelschläger. A martingale approach to the law of large numbers for weakly interacting stochastic processes. *Ann. Probab.*, 12(2):458–479, 1984.

[50] K. J. Painter, T. Hillen, and J. R. Potts. Biological modeling with nonlocal advection-diffusion equations. *Math. Models Methods Appl. Sci.*, 34(1):57–107, 2024.

[51] L. Pareschi and G. Toscani. Interacting multiagent systems: kinetic equations and monte carlo methods, 2013.

[52] G. A. Pavliotis, A. M. Stuart, and U. Vaes. Derivative-free Bayesian inversion using multiscale dynamics. *SIAM J. Appl. Dyn. Syst.*, 21(1):284–326, 2022.

[53] G. A. Pavliotis and A. Zanoni. Eigenfunction Martingale Estimators for Interacting Particle Systems and Their Mean Field Limit. *SIAM J. Appl. Dyn. Syst.*, 21(4):2338–2370, 2022.

[54] G. A. Pavliotis and A. Zanoni. A Method of Moments Estimator for Interacting Particle Systems and their Mean Field Limit. *SIAM/ASA J. Uncertain. Quantif.*, 12(2):262–288, 2024.

[55] R. Pinnau, C. Totzeck, O. Tse, and S. Martin. A consensus-based model for global optimization and its mean-field limit. *Math. Models Methods Appl. Sci.*, 27(1):183–204, 2017.

[56] A. K. Rasmussen, F. Seizilles, M. Girolami, and I. Kazlauskaite. The Bayesian approach to inverse robin problems. *arXiv preprint arXiv:2311.17542*, 2023.

[57] C. E. Rasmussen and C. K. I. Williams. *Gaussian processes for machine learning.* Adaptive Computation and Machine Learning. MIT Press, Cambridge, MA, 2006.

[58] M. Reiß. Asymptotic equivalence for nonparametric regression with multivariate and random design. *Ann. Statist.*, 36(4):1957–1982, 2008.

[59] Björn S., Simon W., and Jakob Z. Metropolis-adjusted interacting particle sampling, 2023.

[60] A. Schlichting and C. Seis. The Scharfetter-Gummel scheme for aggregation-diffusion equations. *IMA J. Numer. Anal.*, 42(3):2361–2402, 2022.

[61] L. Sharrock, N. Kantas, P. Parpas, and G. A. Pavliotis. Online parameter estimation for the McKean-Vlasov SDE. *Stoch. Proc. Appl.*, 162:481–546, 2023.

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

[63] A.-S. Sznitman. Topics in propagation of chaos. In *École d'Été de Probabilités de Saint-Flour XIX—1989*, volume 1464 of *Lecture Notes in Math.*, pages 165–251. Springer, Berlin, 1991.

[64] J. Wloka. *Partial differential equations.* Cambridge University Press, Cambridge, 1987. Translated from the German by C. B. Thomas and M. J. Thomas.