# SECOND ORDER ENSEMBLE LANGEVIN METHOD FOR SAMPLING AND INVERSE PROBLEMS[*]

ZIMING LIU[†], ANDREW STUART[‡], AND YIXUAN WANG[§]

**Abstract.** We propose a sampling method based on an ensemble approximation of second order Langevin dynamics. The log target density is appended with a quadratic term in an auxiliary momentum variable and damped-driven Hamiltonian dynamics introduced; the resulting stochastic differential equation is invariant to the Gibbs measure, with marginal on the position coordinates given by the target. A preconditioner based on covariance under the law of position coordinates under the dynamics does not change this invariance property, and is introduced to accelerate convergence to the Gibbs measure. The resulting mean-field dynamics may be approximated by an ensemble method; this results in a gradient-free and affine-invariant stochastic dynamical system with desirable provably uniform convergence properties across the class of all Gaussian targets. Numerical results demonstrate the potential of the method as the basis for a numerical sampler in Bayesian inverse problems, beyond the Gaussian setting.

**Keywords.** Bayesian inverse problems; sampling; ensemble method; second order Langevin equation; Hybrid Monte Carlo; mean-field models; nonlinear Fokker-Planck equation.

**AMS subject classifications.** 62F15; 60H10; 65C30; 65C35.

## 1. Introduction

**1.1. Set-up.** Consider sampling the density

$$\pi(q) = \frac{1}{Z_q} \exp(-\Phi(q)),$$

where $\Phi : \mathbb{R}^N \to \mathbb{R}$ is termed the *potential* function and $Z_q$ the *normalization constant*. A broad family of problems can be cast into this formulation; the Bayesian approach to inverse problems provides a particular focus for our work [34]. The point of departure for the algorithms considered in this paper is the following mean-field Langevin equation:

$$\frac{\mathrm{d}q}{\mathrm{d}t} = -\mathcal{C}(\rho)D\Phi(q) + \sqrt{2\mathcal{C}(\rho)}\frac{\mathrm{d}W}{\mathrm{d}t}, \tag{1.1}$$

where $D$ denotes the gradient operator, $W$ is an $N$-dimensional standard Brownian motion, $\rho$ is the density associated to the law of $q$, and $\mathcal{C}(\rho)$ is the covariance under this density. This constitutes a mean-field generalization [23] of the standard Langevin equation [44]. Applying a particle approximation to the mean-field model results in an interacting particle system, and coupled Langevin dynamics [23, 24, 37, 42]. The benefit of preconditioning using the covariance is that it leads to mixing rates independent of the problem, provably for quadratic $\Phi$ and empirically beyond this setting [23], because of the affine invariance [26] of the resulting algorithms [24].

In order to further accelerate mixing and achieve sampling efficiency, we introduce an auxiliary variable $p \in \mathbb{R}^N$ and consider the Hamiltonian

$$\mathcal{H}(z) = \frac{1}{2}\langle p, \mathcal{M}^{-1} p \rangle + \Phi(q), \tag{1.2}$$

where the new state variable $z := (q^\top, p^\top)^\top \in \mathbb{R}^{2N}$. Define a measure via its density on $\mathbb{R}^{2N}$ by

$$\Pi(z) = \frac{1}{Z_{q,p}} \exp(-\mathcal{H}(z)), \tag{1.3}$$

where $Z_{q,p}$ is the normalization constant. The marginal distribution of $\Pi$ in the $q$ variable gives the desired distribution $\pi$, i.e. $\int \Pi(z) \mathrm{d}p = \pi(q)$. We now aim at sampling the joint distribution. To this end, consider the following underdamped Langevin dynamics in $\mathbb{R}^{2N}$:

$$\frac{\mathrm{d}z}{\mathrm{d}t} = \mathcal{J} D\mathcal{H}(z) - \mathcal{K} D\mathcal{H}(z) + \sqrt{2\mathcal{K}} \frac{\mathrm{d}W_0}{\mathrm{d}t}, \tag{1.4}$$

with the choices

$$\mathcal{J} = \begin{pmatrix} 0 & \mathcal{C} \\ -\mathcal{C} & 0 \end{pmatrix}, \quad \mathcal{K} = \begin{pmatrix} \mathcal{K}_1 & 0 \\ 0 & \mathcal{K}_2 \end{pmatrix}. \tag{1.5}$$

Here $W_0$ is a standard Brownian motion in $\mathbb{R}^{2N}$ with components $W', W \in \mathbb{R}^N$. Then we have the following, proved in Subsection A.1:

PROPOSITION 1.1.    *Assume that $\mathcal{K}_1$, $\mathcal{K}_2$ are symmetric and non-negative definite, and that $\mathcal{C}$ is symmetric positive definite. Assume further that $\mathcal{C}$, $\mathcal{K}$ and $\mathcal{M}$ depend on the law of $z$ under the dynamics defined by (1.4) and (1.5), but are are independent of $z$: all derivatives with respect to $z$ are zero. Then the Gibbs measure $\Pi(z)$ is invariant under the dynamics defined by (1.4), (1.5).*

In practice, to simulate from such a mean-field model, it will be necessary to consider a particle approximation of the form

$$\frac{\mathrm{d}z^{(i)}}{\mathrm{d}t} = J(Z) D_z H(z^{(i)}; Z) - K(Z) D_z H(z^{(i)}; Z) + \sqrt{2K(Z)} \frac{\mathrm{d}W_0^{(i)}}{\mathrm{d}t}, \tag{1.6}$$

for the set of $I$ particles $Z = \{z^{(i)}\}_{i=1}^I$, and where $M(Z), K(Z), J(Z)$ are appropriate empirical approximations of $\mathcal{M}(\rho), \mathcal{K}(\rho), \mathcal{J}(\rho)$ based on replacing $\rho$ by $\rho^I$ where

$$\rho^I = \frac{1}{I} \sum_{i=1}^I \delta_{z^{(i)}},$$

and the Hamiltonian is given by

$$H(z; Z) = \frac{1}{2}\langle p, M(Z)^{-1} p \rangle + \Phi(q). \tag{1.7}$$

Thus

$$D_z H(z; Z) = \left( D\Phi(q)^\top, (M(Z)^{-1} p)^\top \right)^\top.$$

Note that $H$ is the appropriate finite particle approximation of $\mathcal{H}$, given the particle approximation $M$ of $\mathcal{M}$. The dependence of $\mathcal{H}$ on the law of $z$ has been replaced by dependence on the collection of particles $Z$.[1]

REMARK 1.1.    Unlike (1.4), the Equation (1.6) is no longer a damped-driven Hamiltonian system; this is because of the dependence of the Hamiltonian on the particle positions $Z$, through the mass matrix. Furthermore, its marginal on any coordinate $q^{(i)}$ does not necessarily preserve the desired target measure under the dynamics. However we expect it to do so *approximately* when $I$ is large. This justifies the use of algorithms based on (1.6).

In this paper we will concentrate on a specific choice of mean-field operators within the above general construction, which we now describe. Let $\mathcal{C}_q(\rho)$ denote the $q$-marginal in the covariance under the law of (1.4). We make the choices $\mathcal{K}_1 = 0$, $\mathcal{C} = \mathcal{M} = \mathcal{C}_q(\rho)$, and $\mathcal{K}_2 = \gamma \mathcal{C}_q(\rho)$, for a scalar damping parameter $\gamma > 0$. Then the underdamped Langevin dynamics yields

$$
\begin{aligned}
\frac{dq}{dt} &= p, \\
\frac{dp}{dt} &= -\mathcal{C}_q(\rho) D\Phi(q) - \gamma p + \sqrt{2\gamma \mathcal{C}_q(\rho)} \frac{dW}{dt}.
\end{aligned}
\tag{1.8}
$$

To implement a particle approximation of the mean-field dynamics (1.8) we introduce particles in the form $z^{(i)}(t) = \big((q^{(i)}(t))^\top, (p^{(i)}(t))^\top\big)$ and we use the ensemble covariance and mean approximations

$$
\mathcal{C}_q(\rho) \approx C_q(Z) := \frac{1}{I} \sum_{i=1}^{I} \Big(q^{(i)} - \bar{q}\Big) \otimes \Big(q^{(i)} - \bar{q}\Big),
\tag{1.9a}
$$

$$
\bar{q} := \frac{1}{I} \sum_{i=1}^{I} q^{(i)}.
\tag{1.9b}
$$

In order to obtain affine invariance, we take the generalized square root of the ensemble covariance $C_q(Z)$, similarly to [24]. We introduce the $N \times I$ matrix

$$
Q := \Big(q^{(1)} - \bar{q}, q^{(2)} - \bar{q}, \cdots, q^{(I)} - \bar{q}\Big),
$$

which allows us to define the empirical covariance and generalized (nonsymmetric) square root via

$$
C_q(Z) = \frac{1}{I} Q Q^T, \quad \sqrt{C_q(Z)} := \frac{1}{\sqrt{I}} Q.
$$

Now with $I$ independent standard Brownian motions $\{W^{(i)}\}_{i=1}^{I} \in \mathbb{R}^I$, a natural particle approximation of (1.8) is, for $i = 1, \ldots, I$,

$$
\begin{aligned}
\frac{dq^{(i)}}{dt} &= p^{(i)}, \\
\frac{dp^{(i)}}{dt} &= -C_q(Z) D\Phi(q^{(i)}) - \gamma p^{(i)} + \sqrt{2\gamma C_q(Z)} \frac{dW^{(i)}}{dt}.
\end{aligned}
\tag{1.10}
$$

---

[1]The reader is asked to note that collection of particles $Z$ is different from normalization constant $Z_{q,p}$ appearing in (1.3).

In subsequent sections we will employ an ensemble approximation of $C_q(Z)D\Phi(\cdot)$, as in [23], thereby avoiding the need to compute adjoints of the forward model; we note also that in the linear case this approximation is exact. We will show that the resulting interacting particle system has the potential to provide accurate derivative-free inference for certain classes of inverse problems.

In the remainder of this section we provide a literature review, we highlight our contributions, and we outline the structure of the paper.

**1.2. Literature review.** The overdamped Langevin equation is the canonical SDE that is invariant with respect to a given target density. Many sampling algorithms are built upon this idea, and in particular, it is shown to govern a large class of Monte Carlo Markov Chain (MCMC) methods; see [40, 43, 47]. To enhance mixing and accelerate convergence, a second-order method, Hybrid Monte Carlo (HMC, also referred to as Hamiltonian Monte Carlo) [19, 41] has been proposed, leading to underdamped Langevin dynamics. There have been many attempts to justify the empirically observed fast convergence speed of second-order methods in comparison with first-order methods [4]. Recently a quantitative convergence rate is established in [10], showing that the underdamped Langevin dynamics converges faster than the overdamped Langevin dynamics when the log of the target $\pi$ has a small Poincaré constant; see also [21].

The idea of introducing preconditioners within the context of interacting particle systems used for sampling is developed in [37]. Preconditioning via ensemble covariance is shown to boost convergence and numerical performance [23]. Other choices of preconditioners in sampling will lead to different forms of SDEs and associated Fokker-Planck equations with different structures, which can result in different sampling methods effective in different scenarios; see for example [38]. Affine invariance is introduced in [26] where it is argued that this property leads to desirable convergence properties for interacting particle systems used for sampling: methods that satisfy affine invariance are invariant under an affine change of coordinates, and are thus uniformly effective for problems that can be rendered well-conditioned under an affine transformation; in particular for the sampling of the class of all Gaussian measures. An affine invariant version of the mean-field underdamped Langevin dynamics of [23] is proposed in [24, 42].

Kalman methods have shown wide success in state estimation problems since their introduction by Evensen; see [22] for an overview of the field, and the papers [31, 45] for discussion of their use in inverse problems. Using the ensemble covariance as a preconditioner leads to affine invariant [24] and gradient-free [23] approximations of Langevin dynamics; this is desirable in practical computations in view of the intractability of derivatives in many large-scale models arising in science and engineering [15, 28, 29, 35]. See [2, 49] for analysis of these methods and, in the context of continuous data-assimilation, see [3, 17, 51]. There are other derivative-free methods that can be derived from the mean-field perspective, and in particular consensus-based methods show promise for optimization [12] and have recently been extended to sampling in [11].

Recent work has established the convergence of ensemble preconditioning methods to mean field limits; see for example [18]. For other works on rigorous derivation of mean field limits of interacting particle systems, see [13, 32, 50]. For underpinning theory of Hamiltonian-based sampling, see [5, 7, 8, 39].

**1.3. Our contributions.** The following contributions are made in this paper:

(1) We introduce an underdamped second order mean field Langevin dynamics, with a covariance-based preconditioner.

(2) In the case of Bayesian inverse problems defined by a linear forward map, we show

that that this mean field model preserves Gaussian distributions under time evolution and, if initialized at a Gaussian, converges to the desired target at a rate independent of the linear map.

(3) We introduce finite particle approximations of the mean field model, resulting in an affine invariant method.

(4) For Bayesian inverse problems, we introduce a gradient-free approximation of the algorithm, based on ensemble Kalman methodology.

(5) In the context of Bayesian inverse problems we provide numerical examples to demonstrate that the algorithm resulting from the previous considerations has desirable sampling properties.

In Section 2, we introduce the inverse problems context that motivates us. In Section 3 we discuss the equilibrium distribution of the mean field model. Section 4 introduces the ensemble Kalman approximation of the finite particle system; and in that section we also demonstrate affine invariance of the resulting method. Section 5 presents analysis of the finite particle system in the case of linear inverse problems, where the ensemble Kalman approximation is exact; we demonstrate that the relaxation time to equilibrium is independent of the specific linear inverse problem considered, a consequence of affine invariance. In Section 6 we provide numerical results which demonstrate the efficiency and potential value of our method, and in Section 7 we draw conclusions. Proofs of the propositions are given in the Appendix.

## 2. Inverse problem

Consider the Bayesian inverse problem of finding $q$ from an observation $y$ determined by the forward model

$$y = \mathcal{G}(q) + \eta.$$

Here $\mathcal{G} : \mathbb{R}^N \to \mathbb{R}^J$ is a (in general) nonlinear forward map. We assume a prior zero-mean Gaussian $\pi_0 = \mathsf{N}(0, \Gamma_0)$ on unknown $q$ and assume that the random variable $\eta \sim \mathsf{N}(0, \Gamma)$, representing measurement error, is independent of the prior on $q$. We also assume that $\Gamma, \Gamma_0$ are positive definite. Then by Bayes rule, the posterior density that we aim to sample is given by[2]

$$\pi(q) \propto \exp\left(-\frac{1}{2}\|y - \mathcal{G}(q)\|_\Gamma^2\right) \pi_0(q) \propto \exp(-\Phi(q)),$$

where potential function $\Phi(q)$ has the following form:

$$\Phi(q) = \frac{1}{2}\|y - \mathcal{G}(q)\|_\Gamma^2 + \frac{1}{2}\|q\|_{\Gamma_0}^2. \tag{2.1}$$

In the linear case when $\mathcal{G}(q) = Aq$, $\Phi(q)$ is quadratic and the gradient $D\Phi(q)$ can be written as a linear function:

$$\Phi(q) = \frac{1}{2}\|y - Aq\|_\Gamma^2 + \frac{1}{2}\|q\|_{\Gamma_0}^2, \tag{2.2a}$$

$$D\Phi(q) = B^{-1}q - c, \tag{2.2b}$$

$$B = (A^T\Gamma^{-1}A + \Gamma_0^{-1})^{-1}, \quad c = A^T\Gamma^{-1}y. \tag{2.2c}$$

In this linear setting, the posterior distribution $\pi(q)$ is the Gaussian $\mathsf{N}(Bc, B)$.

---

[2]In what follows $\|\cdot\|_C = \|C^{-\frac{1}{2}}\cdot\|$, with analogous notation for the inducing inner-product, for any positive definite covariance $C$ and for $\|\cdot\|$ the Euclidean norm.

### 3. Equilibrium distributions for the mean field Fokker-Planck equation

The mean-field underdamped Langevin Equation (1.4) has an associated nonlinear and nonlocal Fokker-Planck equation giving the evolution of the law of particles $z(t)$, denoted $\rho(z,t)$. The equation for this law is (see proof in Subsection A.1.)

$$\partial_t \rho = \nabla \cdot \big( (\mathcal{K} - J)(\rho \nabla \mathcal{H} + \nabla \rho) \big). \tag{3.1}$$

By Proposition 1.1 this Fokker-Planck equation has $\Pi(z)$ as its equilibrium; this follows as for standard linear Fokker-Planck equations [44] since the dependence of the sample paths on $\rho$ involves only the mean and covariance; see the proof in Subsection A.1 and see also [20, 27, 33, 44, 46] for discussions on how to derive Fokker-Planck equations with a prescribed stationary distribution. In the specific case (1.8), recall that $\mathcal{J}$ and $\mathcal{K}$ are given by

$$\mathcal{J} = \begin{pmatrix} 0 & \mathcal{C}_q(\rho) \\ -\mathcal{C}_q(\rho) & 0 \end{pmatrix}, \quad \mathcal{K} = \begin{pmatrix} 0 & 0 \\ 0 & \gamma \mathcal{C}_q(\rho) \end{pmatrix}. \tag{3.2}$$

These choices satisfy the assumption of Proposition 1.1. We approximate (1.8) by the interacting particle system (1.10). In this context, we note Remark 1.1 to motivate computational methods based on integrating (1.10).

### 4. Ensemble Kalman Approximation

**4.1. Derivatives via differences.** We now make the ensemble Kalman approximation to approximate the gradient term by differences, as in [23]:

$$D\mathcal{G}\left(q^{(i)}\right)\left(q^{(k)} - \bar{q}\right) \approx \left(\mathcal{G}\left(q^{(k)}\right) - \bar{\mathcal{G}}\right),$$

where $\bar{\mathcal{G}} := \frac{1}{I} \sum_{k=1}^{I} \mathcal{G}\left(q^{(k)}\right)$. Invoking this approximation within (1.10), using the specific form (2.1) of $\Phi$, yields the following system of interacting particles in $\mathbb{R}^N$, for $i = 1, \ldots, I$:

$$\dot{q}^{(i)} = p^{(i)},$$

$$\dot{p}^{(i)} = -C_q(Z)\Gamma_0^{-1} q^{(i)} - \frac{1}{I} \sum_{k=1}^{I} \langle \mathcal{G}(q^{(k)}) - \bar{\mathcal{G}}, \mathcal{G}(q^{(i)}) - y \rangle_\Gamma q^{(k)} - \gamma p^{(i)} + \sqrt{2\gamma C_q(Z)} \dot{W}^{(i)}. \tag{4.1}$$

We will use this system as the basis of all our numerical experiments.

**4.2. Affine invariance.** In this subsection, we show the affine invariance property [24, 26, 37] for the Fokker-Planck equations in the mean-field regime (3.1), for the particle equation in the mean-field regime (1.8), for the ensemble approximation (1.10), and the gradient-free approximation (4.1). For simplicity of presentation, we only state the results in the case of ensemble approximation, and the mean-field case is a straightforward analogy upon dropping all of the particle superscripts.

DEFINITION 4.1 (Affine invariance for particle formulation). *We say a particle formulation is affine invariant, if under all affine transformations of the form*

$$q^{(i)} = Av^{(i)} + b, \quad p^{(i)} = Au^{(i)}, \tag{4.2}$$

*the equations on the transformed particle systems are given by the same equations with $q^{(i)}$, $p^{(i)}$ replaced by $v^{(i)}$, $u^{(i)}$ respectively, and with potential $\Phi$ replaced by $\tilde{\Phi}$ via*

$$\tilde{\Phi}(v^{(i)}) = \Phi(q^{(i)}) = \Phi(Av^{(i)} + b).$$

*Here A is any invertible matrix and b is a vector.*

DEFINITION 4.2 (Affine invariance for Fokker-Planck equation). *We say a Fokker-Planck equation is affine invariant, if under all affine transformations of the form*

$$q^{(i)} = Av^{(i)} + b, \quad p^{(i)} = Au^{(i)},$$

*the equations on the pushforward PDF $\tilde{\rho}^I$ are given by the same equation on $\rho^I$ with $q^{(i)}$, $p^{(i)}$ replaced by $v^{(i)}$, $u^{(i)}$ respectively, and with Hamiltonian $H$ replaced by $\tilde{H}$ via*

$$\tilde{H}(v^{(i)}, u^{(i)}) = H(q^{(i)}, p^{(i)}) = H(Av^{(i)} + b, Au^{(i)}).$$

*Here A is any invertible matrix and b is a vector.*

The key dynamical systems introduced in this paper are affine invariant:

PROPOSITION 4.1. *The particle formulations* (1.8)*,* (1.10) *and* (4.1) *are affine invariant. The Fokker-Planck Equation* (3.1) *is also affine invariant.*

We defer the proof to Subsection A.2. The significance of affine invariance is that it implies that the rate of convergence is preserved under affine transformations. The proposed methodology is thus uniformly effective for problems that become well-conditioned under an affine transformation. Proposition 5.1, which follows in the next section, illustrates this property in the setting of linear forward map $\mathcal{G}(\cdot)$.

REMARK 4.1. The affine invariance of the methodology introduced in [37] involves a definition different from that in Definition 4.1. In particular (4.2) is replaced by

$$q^{(i)} = Av^{(i)} + b, \quad p^{(i)} = u^{(i)}. \tag{4.3}$$

## 5. Mean field model for linear inverse problems

We consider the mean field SDE (1.8) in the linear inverse problem setting of Section 2 with $\mathcal{G}(q) = Aq$; thus (2.2) holds. We note that $B$ in (2.2) is both well-defined and symmetric positive definite since $\Gamma_0, \Gamma$ are assumed to be symmetric positive definite. The two Propositions 5.1, 5.2 demonstrate problem-independent rates of convergence, across the set of all linear Gaussian inverse problems; this is a consequence of affine invariance which in turn is a consequence of our choice of preconditioned mean field system.

In the setting of the linear inverse problem, the mean field model (1.8) reduces to

$$\begin{aligned} \dot{q} &= p, \\ \dot{p} &= -\mathcal{C}_q(\rho)(B^{-1}q - c) - \gamma p + \sqrt{2\gamma \mathcal{C}_q(\rho)}\dot{W}. \end{aligned} \tag{5.1}$$

We prove the following result about this system in Subsection A.3:

PROPOSITION 5.1. *Write the mean $m(\rho)$ and the covariance $\mathcal{C}(\rho)$ of the law $\rho(z)$ of particles in Equation* (5.1) *in the block form*

$$m(\rho) = \begin{pmatrix} m_q(\rho) \\ m_p(\rho) \end{pmatrix}, \quad \mathcal{C}(\rho) = \begin{pmatrix} \mathcal{C}_q(\rho) & \mathcal{C}_{q,p}(\rho) \\ \mathcal{C}_{q,p}^T(\rho) & \mathcal{C}_p(\rho) \end{pmatrix}.$$

*The evolution of the mean and covariance is prescribed by the following system of ODEs:*

$$\dot{m}_q = m_p,$$
$$\dot{m}_p = -\mathcal{C}_q(B^{-1}m_q - c) - \gamma m_p,$$
$$\dot{\mathcal{C}}_q = \mathcal{C}_{q,p} + \mathcal{C}_{q,p}^T, \tag{5.2}$$
$$\dot{\mathcal{C}}_p = -\mathcal{C}_q B^{-1}\mathcal{C}_{q,p} - (\mathcal{C}_q B^{-1}\mathcal{C}_{q,p})^T - 2\gamma\mathcal{C}_p + 2\gamma\mathcal{C}_q,$$
$$\dot{\mathcal{C}}_{q,p} = -\gamma\mathcal{C}_{q,p} - \mathcal{C}_q B^{-1}\mathcal{C}_q + \mathcal{C}_p.$$

*The unique steady solution with positive definite covariance is the Gibbs measure $m_q = Bc, m_p = 0, C_{q,p} = 0, C_q = C_p = B$; the marginal on $q$ gives the solution of the linear Gaussian Bayesian inverse problem. All other steady state solutions have degenerate covariance, are unstable, and take the form $m_q = B(c + m), m_p = 0, C_{q,p} = 0, C_q = C_p = B^{1/2}XB^{1/2}$ for a projection matrix $X$ and $m$ in the nullspace of $C_q$. The equilibrium point with positive definite covariance is hyperbolic and linearly stable and hence its basin of attraction is an open set, containing the equilibrium point itself, in the set of all mean vectors and positive definite covariances. Furthermore, the mean and covariance converge to this equilibrium, from all points in its basin of attraction, with a speed independent of $B$ and $c$.*

REMARK 5.1.    Proposition 5.1 demonstrates that the convergence speed to the non-degenerate equilibrium point is independent of the specific linear inverse problem to be solved; the rate does, however, depend on $\gamma$. Analysis of the linear stability problem shows that $\gamma \approx 1.83$ gives the best local convergence rate; see Remark A.1. In the case where mean field preconditioning is not used, the optimal choice of $\gamma$ for underdamped Langevin dynamics depends on the linear inverse problem being solved, and can be identified explicitly in the scalar setting [44]; for analysis in the non-Gaussian setting see [14]. Motivated by analogies with the work in [14, 44] we expect the optimal choice of $\gamma$ to be problem dependent in the nonlinear case, but motivated by our analysis in the linear setting we expect to see a good choice which is not too small or too large and can be identified by straightforward optimization.

Now we show that for Gaussian initial data, the solution remains Gaussian and is thus determined by the evolution of the mean and covariance via the ODE (5.2). We prove this by investigating the mean field Fokker-Planck equation in the linear case. The evolution in time of the law $\rho(z)$ of Equation (5.1) is governed by the equation

$$\partial_t \rho = \nabla \cdot \left( \begin{pmatrix} -p \\ \mathcal{C}_q(\rho)(B^{-1}q - c) + \gamma p \end{pmatrix} \rho + \begin{pmatrix} 0 & -\mathcal{C}_q(\rho) \\ \mathcal{C}_q(\rho) & \gamma\mathcal{C}_q(\rho) \end{pmatrix} \nabla \rho \right). \tag{5.3}$$

We prove the following result in Subsection A.4; by virtue of Proposition 5.1 it establishes that the eigenvalues that determine the local stability of the posterior Gaussian density are the same across all linear Gaussian inverse problems:

PROPOSITION 5.2.    *Let $m(t)$, $C(t)$ solve the ODE (5.2) with initial conditions $m_0$ and $C_0$. Assume that $\rho_0$ is a Gaussian distribution, so that*

$$\rho_0(z) := \frac{1}{(2\pi)^N} (\det C_0)^{-1/2} \exp\left( -\frac{1}{2} \|z - m_0\|_{C_0}^2 \right)$$

*with mean $m_0$ and covariance $C_0$. Then the Gaussian profile*

$$\rho(t, z) := \frac{1}{(2\pi)^N} (\det C(t))^{-1/2} \exp\left( -\frac{1}{2} \|z - m(t)\|_{C(t)}^2 \right)$$

*solves the Fokker-Planck Equation (5.3) with initial condition $\rho(0, z) = \rho_0(z)$.*

## 6. Numerical results

We introduce, and study, a numerical method for sampling the Bayesian inverse problem of Section 2; the method is based on numerical time stepping of the interacting particle system (4.1). In this section, we demonstrate that the proposed sampler, which we refer to as *ensemble Kalman hybrid Monte Carlo* (EKHMC) can effectively approximate posterior distributions for two widely studied inverse problem test cases. We compare EKHMC with its first-order version EKS [23] and a gold standard MCMC [9]. EKHMC inherits two major advantages of EKS: (1) exact gradients are not required (i.e., derivative-free); (2) the ensemble can faithfully approximate the spread of the posterior distribution, rather than collapse to a single point as happens with the basic EKI algorithm [49]. Furthermore, we show empirically that EKHMC can obtain samples of similar quality to EKS, and has faster convergence than EKS. We detail our numerical time-stepping scheme in the first subsection, before studying two examples (one low dimensional, one a PDE inverse problem for a field) in the subsequent subsections.

**6.1. Time integration schemes.**    We employ a splitting method to integrate the stochastic dynamical system given by Equation (4.1): the first capturing the finite particle approximation of the Hamiltonian evolution, and the second capturing an OU process in momentum space. The Hamiltonian evolution follows the equation:

$$
\begin{aligned}
\dot{q}^{(i)} &= p^{(i)}, \\
\dot{p}^{(i)} &= F_H := -C_q(Z)\Gamma_0^{-1}q^{(i)} - \frac{1}{I}\sum_{k=1}^{I}\langle \mathcal{G}(q^{(k)}) - \bar{\mathcal{G}}, \mathcal{G}(q^{(i)}) - y\rangle_\Gamma q^{(k)}.
\end{aligned}
\tag{6.1}
$$

The OU process follows the equation:

$$
\begin{aligned}
\dot{q}^{(i)} &= 0, \\
\dot{p}^{(i)} &= -\gamma p^{(i)} + \sqrt{2\gamma C_q(Z)}\dot{W}^{(i)}.
\end{aligned}
\tag{6.2}
$$

We implement a symplectic Euler integrator [48] for the part of the particle system arising from approximation of the Hamiltonian contribution to the damped-driven mean-field equations. That is, we take a half step $\epsilon/2$ of momentum updates, then a full step $\epsilon$ of position updates, and finally a half step $\epsilon/2$ of momentum updates. With the ensemble approximation the system is only approximately Hamiltonian; the splitting used in symplectic Euler is still well-defined, however. And we also expect it to perform well in the large particle limit because the mean-field limit is itself Hamiltonian. Let $Z_j$ be the collection of all position and momentum particles at time $j$: $\{q_j^{(i)}, p_j^{(i)}\}_{i=1}^I$. Starting from time $j$ this symplectic Euler integration gives map $Z_j \mapsto \hat{Z}_j$. We set $q_{j+1}^{(i)}$ to be the $i^{\text{th}}$ position coordinates of $\hat{Z}_j$ and then update the momentum coordinates using the OU process which provides the damped-driven component of the mean-field limiting process. The damping coefficient $\gamma > 0$ is treated as a hyperparameter of EKHMC. Vector $Z$ is set at the value given by output of the preceding symplectic Euler integrator, denoted by $\hat{Z}_j$. The update of the $i^{\text{th}}$ momentum coordinate, given by solving the OU process exactly in law, is then

$$
\begin{aligned}
\tilde{p}_j^{(i)} &= \exp(-\gamma\epsilon)\hat{p}_j^{(i)}, \\
p_{j+1}^{(i)} &= \tilde{p}_j^{(i)} + \eta, \quad \eta \sim \mathsf{N}\Big(0, \big(1 - \exp(-2\gamma\epsilon)\big)C_q(\hat{Z}_j)\Big),
\end{aligned}
\tag{6.3}
$$

where $\hat{p}_j^{(i)}$ are the momentum coordinates from $\hat{Z}_j$. Within the OU process the damping coefficient $\gamma > 0$ is treated as a hyperparameter of EKHMC.

It is possible to consider use of a Metropolis-Hastings (Metropolization) step to correct the dynamics; however because the underlying continuous time system is not (for finite number of particles) invariant with respect to the target, doing so would be very complicated and so we do not pursue this. Aside from invariance, Metropolization also imparts stability of the integrator and we may address this in different ways. Indeed, similarly to [23], and as there with the goal of improving stability and convergence speed towards posterior distribution, we implement an adaptive step size, i.e., the true step size is rescaled by the magnitude of the "force field" $F_H$ (defined in Equation (6.1)):

$$\tilde{\epsilon} = \frac{\epsilon}{a|F_H| + 1}. \tag{6.4}$$

**6.2. Low dimensional parameter space.** We follow the example presented in Section 4.3 of the paper [23]. We start by defining the forward map which is given by the one-dimensional elliptic boundary value problem

$$-\frac{d}{dx}\Big(\exp(u_1)\frac{d}{dx}p(x)\Big) = 1, \quad x \in (0,1), \tag{6.5a}$$

$$p(0) = 0, \quad p(1) = u_2. \tag{6.5b}$$

The solution is given explicitly by

$$p(x) = u_2 x + \exp(-u_1)\Big(-\frac{x^2}{2} + \frac{x}{2}\Big). \tag{6.6}$$

The forward model operator $\mathcal{G}$ is then defined by

$$\mathcal{G}(u) = \begin{pmatrix} p(x_1) \\ p(x_2) \end{pmatrix}. \tag{6.7}$$

Here $u = \big(u_1, u_2\big)^T$ is a constant vector that we want to find and we assume that we are given noise measurements $y$ of $p(\cdot)$ at locations $x_1 = 0.25$ and $x_2 = 0.75$. Parameters are chosen according to [23], but we summarize them here for completeness:

- noise $\eta \sim \mathsf{N}(0,\Gamma)$, $\Gamma = 0.1^2 I_2$;
- prior $\pi_0 = \mathsf{N}(0,\Gamma_0)$, $\Gamma_0 = \sigma^2 I_2$, $\sigma = 10$;
- measurement $y = (27.5, 79.7)^T$;
- number of particles $I = 10^3$;
- initialization: $(u_1, u_2) \sim \mathsf{N}(-3.5, 0.1^2) \times \mathsf{U}(70, 110)$.

Here $\mathsf{U}$ is the uniform distribution.

We choose $a = 0.01$, $\epsilon = 0.2$ in both EKS and EKHMC. We find choosing $\gamma = 1$ causes overshooting (the trajectory exhibits oscillatory rather than monotonic convergence), a phenomenon which can be ameliorated by increasing to $\gamma = 100$; indeed this latter value appears, empirically, to be close to optimal in terms of convergence speed. We conjecture this difference from the linear case, where the optimal value is $\gamma = 1.83$ (see Remark 5.1), is due to the non-Gaussianity of the desired target: the particles have accumulated considerable momentum when entering the linear convergence regime, and so extra damping is required to counteract this and avoid overshooting. Despite the desirable problem independence of the optimal $\gamma$ in the linear case, we expect case-by-case optimization to be needed for nonlinear problems. We also empirically study how
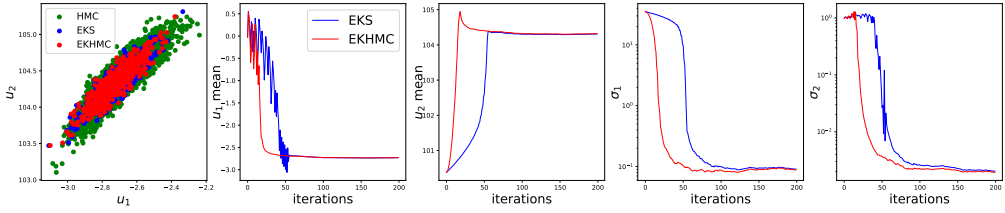
FIG. 6.1. *The low dimensional parameter space example. From left to right: samples; mean $u_1$; mean $u_2$; the first singular value $\sigma_1$; the second singular value $\sigma_2$.*
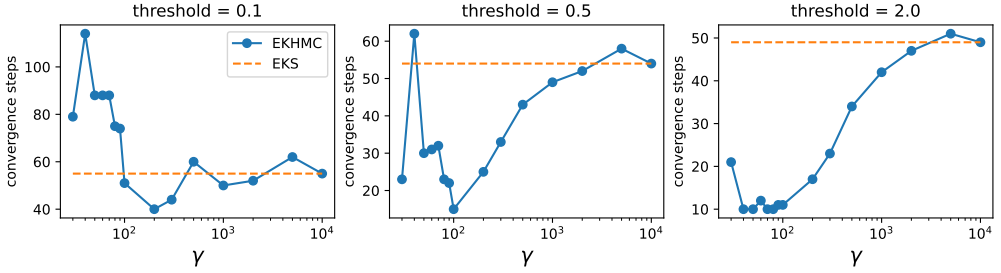


FIG. 6.2. *Convergence time (of $u_2$ mean) as a function of damping coefficient $\gamma$. Left, Middle, Right: thresholds are 0.1, 0.5, 2.0, respectively. The takeaway from these plots is: large damping converges faster eventually (small threshold), while small damping converges faster initially (large threshold).*

$\gamma$ affects the convergence speed for this 2D problem: we sweep $\gamma \in [30, 10^4]$, compute the number of steps needed to be close to the convergent point by a threshold, shown in Figure 6.2. When the threshold is small, large $\gamma$ converges faster; when the threshold is large, small $\gamma$ converges faster. The takeaway is that: large $\gamma$ converges faster eventually (small threshold), while small $\gamma$ converges faster initially (large threshold).

We evolve the ensemble of particles for 200 iterations, and record their positions in the last iteration as the approximation of the posterior distribution, shown in Figure 6.1. The EKHMC samples are quite similar to EKS samples, both of which are reasonable approximations of the samples obtained from the gold standard HMC [19] simulations – but both approximations of the gold standard miss the full spread of the true distribution, because of the ensemble Kalman approximation they invoke. We also compute four ensemble quantities to characterize the evolution of EKHMC and EKS: $u_1$ mean, $u_2$ mean, two eigenvalues $\sigma_1$ and $\sigma_2$ of the covariance matrix $C_q(Z)$. EKHMC has faster convergence towards equilibrium than EKS, benefiting from the second-order dynamics.

**6.3. Darcy flow.** This example follows the problem setting detailed in [23]. We summarize the essential problem specification here for completeness. The forward problem of porous medium flow, defined by permeability $a(\cdot)$ and source term $f(\cdot)$, is to find the pressure field $p(\cdot)$ where $p(\cdot)$ is solution of the following elliptic PDE:

$$-\nabla \cdot \Big( a(x)\nabla p(x) \Big) = f(x), \quad x \in D = [0,1]^2,$$
$$p(x) = 0, \qquad\qquad\qquad x \in \partial D.$$
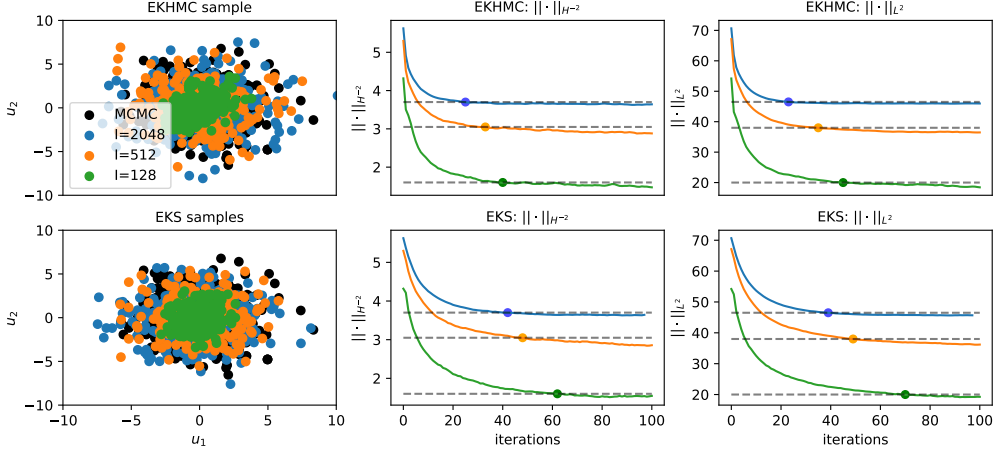$$(6.8)$$

FIG. 6.3. *Darcy flow. Left: samples obtained from EKHMC (top) and EKS (bottom), compared with MCMC. Middle: Evolution of $||u||_{H^{-2}}$ for EKHMC and EKS for different $I = 128, 512, 2048$. Right: The same as the middle, but $||u||_{L^2}$ instead of $||u||_{H^{-2}}$. EKHMC converges faster than EKS.*

We assume that the permeability $a(x) = a(x; u)$ depends on some unknown parameters $u \in \mathbb{R}^d$. The resulting inverse problem is, given (noisy) pointwise measurements of $p(x)$ on a grid, to infer $a(x; u)$ or $u$. We model $a(x; u)$ as a log-Gaussian field. The Gaussian underlying this log-Gaussian model has zero and has precision operator defined as $\mathcal{C}^{-1} = (-\triangle + \tau^2 \mathcal{I})^\alpha$; here $\triangle$ is equipped with Neumann boundary conditions on the spatial-mean zero functions. We set $\tau = 3$ and $\alpha = 2$ in the experiments. Such parametrization yields as Karhunen-Loeve (KL) expansion

$$\log a(x; u) = \sum_{l \in K} u_l \sqrt{\lambda_l} \varphi_l(x), \tag{6.9}$$

where $\varphi_l(x) = \cos(\pi \langle l, x \rangle)$, $\lambda_l = (\pi^2 |\ell|^2 + \tau^2)^{-\alpha}$, $K \equiv \mathbb{Z}^2$. Draws from this random field are Hölder with any exponent less than one. In practice, we truncate $K$ to have dimension $d = 2^8$. We generate a truth random field by sampling $u \sim \mathsf{N}(0, I_d)$. We create data $y$ with $\eta \sim \mathsf{N}(0, 0.1^2 I_K)$. We choose the prior covariance $\Gamma_0 = 10^2 I_d$.

To characterize the evolution of EKHMC and compare it with EKS, we compute two metrics:

$$d_{H^{-2}}(\cdot) = \sqrt{\frac{1}{I} \sum_{j=1}^{I} ||u^{(j)}(t) - \cdot||_{H^{-2}}^2}, \quad d_{L^2}(\cdot) = \sqrt{\frac{1}{I} \sum_{j=1}^{I} ||u^{(j)}(t) - \cdot||_{L^2}^2}. \tag{6.10}$$

For these metrics, we use norms

$$||u||_{H^{-2}} = \sqrt{\sum_{l \in K_d} |u_l|^2 \lambda_l}, \quad ||u||_{L^2} = \sqrt{\sum_{l \in K_d} |u_l|^2}, \tag{6.11}$$

where the first is defined in the negative Sobolev space $H^{-2}$, whilst the second in the $L^2$ space.

We set $a = 0.01$ and $\epsilon = 1.0$ for both EKHMC and EKS. We set $\gamma = 1$ in EKHMC; unlike the previous example, this choice does not lead to over-shooting. We simulate the particles for 100 iterations, and use their positions in the last iteration as the

approximation to the posterior distribution. We compare the samples obtained from MCMC (we use the pCN variant on RWMH [16]), EKS ($I = 128, 512, 2048$) and EKHMC ($I = 128, 512, 2048$) in Figure 6.3. The evolution of $||u||_{H^{-2}}$ and $||u||_{L^2}$ are plotted to show that convergence is achieved, and that EKHMC converges faster than EKS.

## 7. Conclusions

Gradient-free methods for inverse problems are increasingly important in large-scale multi-physics science and engineering problems; see [30] and the references therein. In this paper we have provided an initial study of gradient-free methods which leverage the potential power of Hamiltonian based sampling. Analysis of the resulting method is hard, and our initial theoretical results leave many avenues open; in particular convergence to equilibrium is not established for the underlying nonlinear, nonlocal Fokker-Planck equation arising from the mean field model, and optimization of convergence rates over $\gamma$ is not understood, except in the setting of linear Gaussian inverse problems. The Fokker-Planck equation associated with standard underdamped Langevin dynamics has been studied in the context of hypocoercity – see [1, 52] – and generalization of these methods will be of potential interest. Preconditioned HMC is also proven to converge in [6]; generalization to ensemble approximations would be of interest. On the computational side there are other approaches to avoiding gradient computations, yet leveraging Hamiltonian structure, which need to be evaluated and compared with what is proposed here [36].

**Appendix.** We present the proofs of various results from the paper here.

### A.1. Proof of Proposition 1.1.

*Proof.* The determination of the most general form of diffusion process which is invariant with respect to a given measure has a long history; see [20][Theorem 1] for a statement and the historical context. Note, however, that this theory is not developed in the mean-field setting and concerns only the linear Fokker-Planck equation. However, the dependence of the matrices $\mathcal{K}, \mathcal{J}$ and $\mathcal{M}$ on $\rho$ readily allows use of the approach taken in the linear case. We find it expedient to use the exposition of this topic in [40]. The same derivation as used to obtain Equation (5) from [40] shows[3] that the density $\rho$ associated with the dynamics (1.4), (1.5) satisfies Equation (3.1); this follows from the (resp. skew-) symmetry properties of (resp. $\mathcal{J}$) $\mathcal{K}$ and the fact that $\mathcal{J}$, $\mathcal{K}$ and $\mathcal{M}$ are assumed independent of $z$, despite their dependence on $\rho$. It is also manifestly the case that the Gibbs measure is invariant for (3.1) since the mass term $\mathcal{M}$ appearing in $\mathcal{H}$ is independent of $z$ so that

$$\rho \nabla \mathcal{H} + \nabla \rho = 0$$

when $\rho$ is the Gibbs measure and (3.1) shows that then $\partial_t \rho = 0$. □

### A.2. Proof of Proposition 4.1.

*Proof.* In fact, consider the affine transformation

$$q^{(i)} = Av^{(i)} + b, \quad p^{(i)} = Au^{(i)},$$

---

[3]We use the notational convention concerning divergence of matrix fields that is standard in continuum mechanics [25]; this differs from the notational convention adopted in [40].

along with

$$\tilde{\Phi}(v^{(i)}) = \Phi(q^{(i)}) = \Phi(Av^{(i)}+b), \quad \tilde{H}(v^{(i)},u^{(i)}) = H(q^{(i)},p^{(i)}) = H(Av^{(i)}+b, Au^{(i)}).$$

Then we have that the gradient term scales like $\nabla_{v^{(i)}}\tilde{\Phi}(v^{(i)}) = A^T\nabla_{q^{(i)}}\Phi(q^{(i)})$ and the covariance preconditioner scales like $C_v = A^{-1}C_q A^{-1^T}$. The generalized square root scales like $\sqrt{C_v} = A^{-1}\sqrt{C_q}$. Therefore we can check that affine invariance holds for the particle systems (1.10) and (1.8). Affine invariance for the gradient free approximation (4.1) is more easily seen to be true. Finally for the Fokker-Planck Equation (3.1), we can check similarly via the scaling of the terms. See a similar argument in the proof of Lemma 4.7 in [24]. □

### A.3. Proof of Proposition 5.1.

*Proof.* We can take expectations in (5.1) to obtain the first two ODEs in (5.2). Now define

$$\hat{z} = z - m(\rho) \triangleq \begin{pmatrix} \hat{q} \\ \hat{p} \end{pmatrix},$$

to obtain the following evolution for $\hat{z}$:

$$\dot{\hat{q}} = \hat{p},$$

$$\dot{\hat{p}} = -\mathcal{C}_q(\rho)B^{-1}\hat{q} - \gamma\hat{p} + \sqrt{2\gamma\mathcal{C}_q(\rho)}\dot{W}.$$

Note that $\mathcal{C}_q = \mathbb{E}[\hat{q}\otimes\hat{q}]$, $\mathcal{C}_p = \mathbb{E}[\hat{p}\otimes\hat{p}]$, and $\mathcal{C}_{q,p} = \mathbb{E}[\hat{q}\otimes\hat{p}]$. We can use Ito's formula and the above evolution to derive the last three ODEs in (5.2). The form of the steady state solutions is immediate from setting the right-hand side of (5.2) to zero.

Now, we will establish the independence of the essential dynamics of the mean and covariance on the specific choice of $B, c$. To this end, denote $x_1 = B^{1/2}(B^{-1}m_q - c)$, $x_2 = B^{-1/2}m_p$, $X = B^{-1/2}\mathcal{C}_q B^{-1/2}$, $Y = B^{-1/2}\mathcal{C}_p B^{-1/2}$, $Z = B^{-1/2}\mathcal{C}_{q,p}B^{-1/2}$. Applying this change of variables we obtain, from (5.2), the system

$$\begin{aligned}
\dot{x_1} &= x_2, \\
\dot{x_2} &= -Xx_1 - \gamma x_2, \\
\dot{X} &= Z + Z^T, \\
\dot{Y} &= -XZ - Z^T X - 2\gamma Y + 2\gamma X, \\
\dot{Z} &= -\gamma Z - X^2 + Y.
\end{aligned} \tag{A.1}$$

We notice that the steady state solutions take the form $(x_1,x_2,X,Y,Z) = (w,0,X,X,0)$ for a projection matrix $X^2 = X$ and $w$ such that $Xw = 0$. The steady state with $X = I_N$ and $w = 0$ corresponds to the desired posterior mean. The transformed equation is indeed independent of $B$ and $c$. Therefore the speed of convergence $(x_1,x_2,X,Y,Z) \to (0,0,I_N,I_N,0)$, within its basin of attraction, is indeed independent of $B$ and $c$. The same is true in the original variables.

To complete the proof of stability it suffices to establish the local exponential convergence to the steady solution $(x_1,x_2,X,Y,Z) = (0,0,I_N,I_N,0)$. As a first step, we show that the following system converges to $(X,Y,Z) = (I_N,I_N,0)$:

$$\begin{aligned}
\dot{X} &= Z + Z^T, \\
\dot{Y} &= -XZ - Z^T X - 2\gamma Y + 2\gamma X, \\
\dot{Z} &= -\gamma Z - X^2 + Y.
\end{aligned} \tag{A.2}$$

Linearization around $(X,Y,Z)=(I_N,I_N,0)$ in variables of entries of $(X,Y,Z+Z^T)$ gives the matrix

$$\begin{pmatrix} 0 & 0 & 2I_{N^2} \\ 2\gamma I_{N^2} & -2\gamma I_{N^2} & -2I_{N^2} \\ -2I_{N^2} & I_{N^2} & -\gamma I_{N^2} \end{pmatrix}$$

whose eigenvalues satisfy $x^3+3\gamma x^2+(2\gamma^2+6)x+4\gamma=0$ which all have strictly negative real part for $\gamma>0$, since we can show the unique real eigenvalue is in the interval $(-\gamma,0)$. Therefore we conclude the local exponential convergence of covariances $X,Y,Z$ in a neighbourhood of $(X,Y,Z)=(I_N,I_N,0)$. The exponential convergence of means $(x_1,x_2)\to(0,0)$ then follows linearizing the first two linear ODEs in the system (A.1) at $X=I_N$.

Similarly, for any other steady state $(X,Y,Z)=(X,X,0)$ with a projection matrix $X\neq I_N$, we will show that there is an unstable direction in (A.2). In fact, since $X$ is symmetric with only eigenvalues 1 and 0, there exists a nonzero vector $v$ such that $Xv=0$. Linearization around $(X,Y,Z)=(X,X,0)$ in the direction $(avv^T,bvv^T,cvv^T)$ gives the following $3\times 3$ matrix for scalars $a,b,c$:

$$\begin{pmatrix} 0 & 0 & 2 \\ 2\gamma & -2\gamma & 0 \\ 0 & 1 & -\gamma \end{pmatrix}$$

whose eigenvalues satisfy $x^3+3\gamma x^2+2\gamma^2 x-4\gamma=0$. For all $\gamma>0$ it may be shown that there exists a real eigenvalue in the interval $(0,1)$, since $f(0)f(1)<0$; thus there is an unstable direction determined by $(a,b,c)$, and therefore in the original formulation.     □

REMARK A.1.    We can further investigate the spectral gap around $(X,Y,Z)=(I_N,I_N,0)$, which is the absolute value in the real root of equation $x^3+3\gamma x^2+(2\gamma^2+6)x+4\gamma=0$. To be precise, we can show that for $x_0=-\sqrt{12-\sqrt{128}}$ and $\gamma_0=-(4+3x_0^2)/(4x_0)\approx 1.83$, the spectral gap is maximized as $-x_0$ and we expect fastest convergence for our method in the linear setting.

To establish this claim we proceed as follows. By the intermediate value theorem, in order to show there always exists a root in $[x_0,0)$ and the spectral gap is at most $-x_0$, we only need to show that

$$x_0^3+3\gamma x_0^2+(2\gamma^2+6)x_0+4\gamma\leq 0.$$

The claim is true because $x_0<0$ and $2x_0(x_0^3+6x_0)=(4+3x_0^2)^2/4$. By the basic inequality $a^2+b^2\geq 2ab$, we have

$$x_0^3+3\gamma x_0^2+(2\gamma^2+6)x_0+4\gamma=2x_0\gamma^2+(4+3x_0^2)\gamma+x_0^3+6x_0\leq 0.$$

The maximal spectral gap is attained if and only if the equality in $a^2+b^2\geq 2ab$ holds, i.e. when $\gamma=\gamma_0$.

We also plot the spectral gap as a function of $\gamma$ to help visualize the dependence of the rate of convergence on damping for the linear problem; see Figure A.1. The clear message from this figure is that there is a natural optimization problem for $\gamma$ in this linear Gaussian setting; this is used to motivate searches for optimal parameters in the non-Gaussian setting.
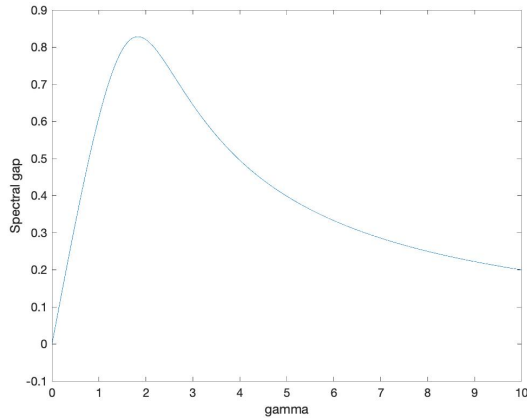
FIG. A.1. *Spectral gap as a function of $\gamma$*

### A.4. Proof of Proposition 5.2.

*Proof.* Note that since $m(\rho), \mathcal{C}(\rho)$ are independent of the particle $z$, we have

$$\nabla \rho = -\mathcal{C}(\rho)^{-1}(z - m(\rho))\rho,$$

$$D^2 \rho = \left(-\mathcal{C}(\rho)^{-1} + \left(\mathcal{C}(\rho)^{-1}(z - m(\rho))\right) \otimes \left(\mathcal{C}(\rho)^{-1}(z - m(\rho))\right)\right)\rho.$$

Using this we can substitute the Gaussian profile into Equation (5.3). We compute the first term on the right hand to obtain

$$\left(-\mathcal{C}(\rho)^{-1}(z - m(\rho)) \cdot \begin{pmatrix} -p \\ \mathcal{C}_q(\rho)(B^{-1}q - c) + \gamma p \end{pmatrix} + \gamma N\right)\rho \triangleq I_1 \rho.$$

The second term on the right-hand side can be written as:

$$\begin{pmatrix} 0 & -\mathcal{C}_q(\rho) \\ \mathcal{C}_q(\rho) & \gamma \mathcal{C}_q(\rho) \end{pmatrix} : \left(-\mathcal{C}(\rho)^{-1} + \left(\mathcal{C}(\rho)^{-1}(z - m(\rho))\right) \otimes \left(\mathcal{C}(\rho)^{-1}(z - m(\rho))\right)\right)\rho \triangleq (I_2 + I_3)\rho.$$

For the left-hand side of (5.3), note that

$$\frac{\mathrm{d}}{\mathrm{d}t} \det \mathcal{C}(\rho) = \mathrm{Tr}\left[\det \mathcal{C}(\rho)\mathcal{C}(\rho)^{-1}\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{C}(\rho)\right], \quad \frac{\mathrm{d}}{\mathrm{d}t}\mathcal{C}(\rho)^{-1} = -\mathcal{C}(\rho)^{-1}\left(\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{C}(\rho)\right)\mathcal{C}(\rho)^{-1}.$$

Using the evolution of the mean (5.2) we obtain

$$\frac{\mathrm{d}}{\mathrm{d}t}\|z - m(\rho)\|^2_{\mathcal{C}(\rho)} = 2\left\langle \frac{\mathrm{d}}{\mathrm{d}t}(z - m(\rho)), \mathcal{C}(\rho)^{-1}(z - m(\rho)) \right\rangle$$

$$+ \left\langle (z - m(\rho)), \frac{\mathrm{d}}{\mathrm{d}t}\left(\mathcal{C}(\rho)^{-1}\right)(z - m(\rho)) \right\rangle$$

$$= 2\left\langle \begin{pmatrix} -m_p(\rho) \\ \mathcal{C}_q(\rho)(B^{-1}m_q(\rho) - c) + \gamma m_p(\rho) \end{pmatrix}, \mathcal{C}(\rho)^{-1}(z - m(\rho)) \right\rangle$$

$$- \left\langle \mathcal{C}(\rho)^{-1}(z - m(\rho)), \frac{\mathrm{d}}{\mathrm{d}t}(\mathcal{C}(\rho))\mathcal{C}(\rho)^{-1}(z - m(\rho)) \right\rangle \triangleq 2(I_4 - I_5).$$

Therefore we can compute the left-hand side as:

$$\partial_t \rho = \left[ -\frac{1}{2} (\det \mathcal{C}(\rho))^{-1} \left( \frac{\mathrm{d}}{\mathrm{d}t} \det \mathcal{C}(\rho) \right) - \frac{1}{2} \frac{\mathrm{d}}{\mathrm{d}t} \| z - m(\rho) \|^2_{\mathcal{C}(\rho)} \right] \rho$$

$$= \left[ -\frac{1}{2} \mathrm{Tr} \left[ \mathcal{C}(\rho)^{-1} \frac{\mathrm{d}}{\mathrm{d}t} \mathcal{C}(\rho) \right] - I_4 + I_5 \right] \rho \triangleq (I_6 - I_4 + I_5) \rho.$$

In order to show that the Gaussian profile is indeed a solution to (5.3), we only need to show that $I_1 + I_2 + I_3 = -I_4 + I_5 + I_6$. Note that

$$I_1 + I_4 = \gamma N + \mathcal{C}(\rho)^{-1} (z - m(\rho)) \cdot \begin{pmatrix} p - m_p(\rho) \\ -\mathcal{C}_q(\rho) B^{-1} (q - m_q(\rho)) - \gamma (p - m_p(\rho)) \end{pmatrix}$$

$$= \gamma N + \mathcal{C}(\rho)^{-1} (z - m(\rho)) \cdot \begin{pmatrix} 0 & I_N \\ -\mathcal{C}_q(\rho) B^{-1} & -\gamma I_N \end{pmatrix} (z - m(\rho)).$$

Defining $\hat{z} = z - m(\rho)$, we collect the terms in the to-be-proven identity $I_1 + I_2 + I_3 = -I_4 + I_5 + I_6$ as:

$$\hat{z}^T \left[ \mathcal{C}(\rho)^{-1} \begin{pmatrix} 0 & I_N \\ -\mathcal{C}_q(\rho) B^{-1} & -\gamma I_N \end{pmatrix} + \mathcal{C}(\rho)^{-1} \left( \begin{pmatrix} 0 & -\mathcal{C}_q(\rho) \\ \mathcal{C}_q(\rho) & \gamma \mathcal{C}_q(\rho) \end{pmatrix} - \frac{1}{2} \frac{\mathrm{d}}{\mathrm{d}t} (\mathcal{C}(\rho)) \right) \mathcal{C}(\rho)^{-1} \right] \hat{z}$$

$$+ \gamma N - \mathrm{Tr} \left[ \begin{pmatrix} 0 & -\mathcal{C}_q(\rho) \\ \mathcal{C}_q(\rho) & \gamma \mathcal{C}_q(\rho) \end{pmatrix} \mathcal{C}(\rho)^{-1} \right] + \frac{1}{2} \mathrm{Tr} \left[ \frac{\mathrm{d}}{\mathrm{d}t} \mathcal{C}(\rho) \mathcal{C}(\rho)^{-1} \right] \triangleq \hat{z}^T M_1 \hat{z} + M_2 = 0.$$

We only need to show that $M_1 + M_1^T = M_2 = 0$.

In fact, we compute $\mathcal{C}(\rho)(M_1 + M_1^T)\mathcal{C}(\rho)$ to be:

$$\begin{pmatrix} 0 & I_N \\ -\mathcal{C}_q(\rho) B^{-1} & -\gamma I_N \end{pmatrix} \mathcal{C}(\rho) + \mathcal{C}(\rho) \begin{pmatrix} 0 & -(B^{-1})^T \mathcal{C}_q(\rho) \\ I_N & -\gamma I_N \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & 2\gamma \mathcal{C}_q(\rho) \end{pmatrix} - \frac{\mathrm{d}}{\mathrm{d}t} (\mathcal{C}(\rho)),$$

which equals zero upon blockwise computation via (5.2). Thus $M_1 + M_1^T = 0$. Moreover, note that $M_2 = -\mathrm{Tr}[\mathcal{C}(\rho) M_1]$: $M_1 + M_1^T = 0$ implies $\mathcal{C}(\rho)(M_1 + M_1^T) = 0$ and taking the trace on both sides leads to the conclusion that $M_2 = 0$. Therefore the Gaussian profile indeed solves the Fokker-Planck Equation (5.3). □

## REFERENCES

[1] D. Bakry and M. Émery, *Diffusions hypercontractives*, in J. Azéma, and M. Yor (eds.), Seminaire de Probabilités XIX 1983/84, Springer, 177–206, 1985. 7

[2] K. Bergemann and S. Reich, *A mollified ensemble Kalman filter*, Quart. J. Roy. Meteor. Soc., 136(651):1636–1643, 2010. 1.2

[3] K. Bergemann and S. Reich, *An ensemble Kalman-Bucy filter for continuous data assimilation*, Meteorol. Z., 21(3):213–219, 2012. 1.2

[4] M. Betancourt, *A conceptual introduction to Hamiltonian Monte Carlo*, arXiv preprint, arXiv:1701.02434, 2017. 1.2

[5] M. Betancourt, S. Byrne, S. Livingstone, and M. Girolami, *The geometric foundations of Hamiltonian Monte Carlo*, Bernoulli, 23(4A):2257–2298, 2017. 1.2

[6] N. Bou-Rabee and A. Eberle, *Two-scale coupling for preconditioned Hamiltonian Monte Carlo in infinite dimensions*, Stoch. Partial Differ. Equ.: Anal. Comput., 9(1):207–242, 2021. 7

[7] N. Bou-Rabee and J.M. Sanz-Serna, *Randomized Hamiltonian Monte Carlo*, Ann. Appl. Probab., 27(4):2159–2194, 2017. 1.2

[8] N. Bou-Rabee and J.M. Sanz-Serna, *Geometric integrators and the Hamiltonian Monte Carlo method*, Acta Numer., 27:113–206, 2018. 1.2

[9] S. Brooks, A. Gelman, G. Jones, and X.-L. Meng, *Handbook of Markov Chain Monte Carlo*, CRC Press, 2011. 6

[10] Y. Cao, J. Lu, and L. Wang, *On explicit $L^2$-convergence rate estimate for underdamped Langevin dynamics*, Arch. Ration. Mech. Anal., 247:90, 2023. 1.2

[11] J.A. Carrillo, F. Hoffmann, A.M. Stuart, and U. Vaes, *Consensus-based sampling*, Stud. Appl. Math., 148(3):1069–1140, 2022. 1.2

[12] J.A. Carrillo, Y.-P. Choi, C. Totzeck, and O. Tse, *An analytical framework for consensus-based global optimization method*, Math. Models Meth. Appl. Sci., 28(06):1037–1066, 2018. 1.2

[13] J.A. Carrillo, M. Fornasier, G. Toscani, and F. Vecil, *Particle, kinetic, and hydrodynamic models of swarming*, in G. Naldi, L. Pareschi, and G. Toscani (eds.), Mathematical Modeling of Collective Behavior in Socio-economic and Life Sciences, Springer, 297–336, 2010. 1.2

[14] M. Chak, N. Kantas, T. Lelièvre, and G.A. Pavliotis, *Optimal friction matrix for underdamped Langevin sampling*, ESAIM, 57(6):3335–3371,2023 5.1

[15] E. Cleary, A. Garbuno-Inigo, S. Lan, T. Schneider, and A.M. Stuart, *Calibrate, emulate, sample*, J. Comput. Phys., 424:109716, 2021. 1.2

[16] S.L. Cotter, G.O. Roberts, A.M. Stuart, and D. White, *MCMC methods for functions: modifying old algorithms to make them faster*, Stat. Sci., 28(3):424–446, 2013. 6.3

[17] P. Del Moral, A. Kurtzmann, and J. Tugaut, *On the stability and the uniform propagation of chaos of a class of extended ensemble Kalman–Bucy filters*, SIAM J. Control Optim., 55(1):119–155, 2017. 1.2

[18] Z. Ding and Q. Li, *Ensemble Kalman sampler: Mean-field limit and convergence analysis*, SIAM J. Math. Anal., 53(2):1546–1578, 2021. 1.2

[19] S. Duane, A.D. Kennedy, B.J. Pendleton, and D. Roweth, *Hybrid Monte Carlo*, Phys. Lett. B, 195(2):216–222, 1987. 1.2, 6.2

[20] A.B. Duncan, N. Nüsken, and G.A. Pavliotis, *Using perturbed underdamped Langevin dynamics to efficiently sample from probability distributions*, J. Stat. Phys., 169(6):1098–1131, 2017. 3, A.1

[21] A. Eberle, A. Guillin, and R. Zimmer, *Couplings and quantitative contraction rates for Langevin dynamics*, Ann. Probab., 47(4):1982–2010, 2019. 1.2

[22] G. Evensen, *Data Assimilation: The Ensemble Kalman Filter*, Springer, 2, 2009. 1.2

[23] A. Garbuno-Inigo, F. Hoffmann, W. Li, and A.M. Stuart, *Interacting Langevin diffusions: Gradient structure and ensemble Kalman sampler*, SIAM J. Appl. Dyn. Syst., 19(1):412–441, 2020. 1.1, 1.1, 1.2, 4.1, 6, 6.1, 6.2, 6.2, 6.3

[24] A. Garbuno-Inigo, N. Nüsken, and S. Reich, *Affine invariant interacting Langevin dynamics for Bayesian inference*, SIAM J. Appl. Dyn. Syst., 19(3):1633–1658, 2020. 1.1, 1.1, 1.2, 4.2, A.2

[25] O. Gonzalez and A.M. Stuart, *A First Course in Continuum Mechanics*, Cambridge University Press, 42, 2008. 3

[26] J. Goodman and J. Weare, *Ensemble samplers with affine invariance*, Commun. Appl. Math. Comput. Sci., 5(1):65–80, 2010. 1.1, 1.2, 4.2

[27] R. Graham, *Covariant formulation of non-equilibrium statistical thermodynamics*, Z. Phys. B Condens. Matter, 26(4):397–405, 1977. 3

[28] E. Haber, F. Lucka, and L. Ruthotto, *Never look back - a modified ENKF method and its application to the training of neural networks without back propagation*, arXiv preprint, arXiv:1805.08034, 2018. 1.2

[29] D.Z. Huang, T. Schneider, and A.M. Stuart, *Unscented Kalman inversion*, arXiv preprint, arXiv:2102.01580, 2021. 1.2

[30] D.Z. Huang, J. Huang, S. Reich, and A.M. Stuart, *Efficient derivative-free Bayesian inference for large-scale inverse problems*, Inverse Probl., 38(12):125006, 2022. 7

[31] M.A. Iglesias, K.J.H. Law, and A.M. Stuart, *Ensemble Kalman methods for inverse problems*, Inverse Probl., 29(4):045001, 2013. 1.2

[32] P.-E. Jabin and Z. Wang, *Mean field limit for stochastic particle systems*, in N. Bellomo, P. Degond, and E. Tadmor (eds.), Active Particles, Springer, 1:379–402, 2017. 1.2

[33] D.-Q. Jiang and D. Jiang, *Mathematical Theory of Nonequilibrium Steady States: On the Frontier of Probability and Dynamical Systems*, Springer Science & Business Media, 2004. 3

[34] J. Kaipio and E. Somersalo, *Statistical and Computational Inverse Problems*, Springer Science & Business Media, 160, 2006. 1.1

[35] N.B. Kovachki and A.M. Stuart, *Ensemble Kalman inversion: A derivative-free technique for machine learning tasks*, Inverse Probl., 35(9):095005, 2019. 1.2

[36] S. Lan, T. Bui-Thanh, M. Christie, and M. Girolami, *Emulation of higher-order tensors in manifold Monte Carlo methods for Bayesian inverse problems*, J. Comput. Phys., 308:81–101, 2016. 7

[37] B. Leimkuhler, C. Matthews, and J. Weare, *Ensemble preconditioning for Markov chain Monte Carlo simulation*, Stat. Comput., 28(2):277–290, 2018. 1.1, 1.2, 4.2, 4.1

[38] M. Lindsey, J. Weare, and A. Zhang, *Ensemble Markov chain Monte Carlo with teleporting*

*walkers*, SIAM-ASA J. Uncertain. Quantif., 10(3):860–885, 2022. 1.2

[39] S. Livingstone, M. Betancourt, S. Byrne, and M. Girolami, *On the geometric ergodicity of Hamiltonian Monte Carlo*, Bernoulli, 25(4A):3109–3138, 2019. 1.2

[40] Y.A. Ma, T. Chen, and E. Fox, *A complete recipe for stochastic gradient MCMC*, Adv. Neur. Inf. Process. Syst., 2917–2925, 2015. 1.2, A.1, 3

[41] R.M. Neal, *An improved acceptance procedure for the hybrid Monte Carlo algorithm*, J. Comput. Phys., 111(1):194–203, 1994. 1.2

[42] N. Nüsken and S. Reich, *Note on interacting Langevin diffusions: Gradient structure and ensemble Kalman sampler by Garbuno-Inigo, Hoffmann, Li and Stuart*, arXiv preprint, arXiv:1908.10890, 2019. 1.1, 1.2

[43] M. Ottobre and G.A. Pavliotis, *Asymptotic analysis for the generalized Langevin equation*, Nonlinearity, 24(5):1629, 2011. 1.2

[44] G.A. Pavliotis, *Stochastic Processes and Applications: Diffusion Processes, the Fokker-Planck and Langevin Equations*, Springer, 60, 2014. 1.1, 3, 5.1

[45] S. Reich, *A dynamical systems framework for intermittent data assimilation*, BIT Numer. Math., 51(1):235–249, 2011. 1.2

[46] H. Risken, *The Fokker-Planck Equation*, Springer, 63–95, 1996. 3

[47] G.O. Roberts and J.S. Rosenthal, *Optimal scaling for various Metropolis-Hastings algorithms*, Stat. Sci., 16(4):351–367, 2001. 1.2

[48] J.-M. Sanz-Serna and M.-P. Calvo, *Numerical Hamiltonian Problems*, Courier Dover Publications, 2018. 6.1

[49] C. Schillings and A.M. Stuart, *Analysis of the ensemble Kalman filter for inverse problems*, SIAM J. Numer. Anal., 55(3):1264–1290, 2017. 1.2, 6

[50] A.-S. Sznitman, *Topics in propagation of chaos*, in C. Bahadoran, H. Djellout, and A. Guillin (eds.), *Ecole d'été de probabilités de Saint-Flour XIX–1989*, Springer, 165–251, 1991. 1.2

[51] A. Taghvaei, J. De Wiljes, P.G. Mehta, and S. Reich, *Kalman filter and its modern extensions for the continuous-time nonlinear filtering problem*, J. Dyn. Syst. Meas. Control, 140(3):030904, 2018. 1.2

[52] C. Villani, *Hypocoercivity*, Amer. Math. Soc., 2009. 7