

Nonparametric Uncertainty Quantification for Stochastic Gradient Flows[§]

Tyrus Berry[†] and John Harlim^{†‡¶}

Abstract. This paper presents a nonparametric statistical modeling method for quantifying uncertainty in stochastic gradient systems with isotropic diffusion. The central idea is to apply the diffusion maps algorithm on a training data set to produce a stochastic matrix whose generator is a discrete approximation to the backward Kolmogorov operator of the underlying dynamics. The eigenvectors of this stochastic matrix, which we will refer to as the diffusion coordinates, are discrete approximations to the eigenfunctions of the Kolmogorov operator and form an orthonormal basis for functions defined on the data set. Using this basis, we consider the projection of three uncertainty quantification (UQ) problems (prediction, filtering, and response) into the diffusion coordinates. In these coordinates, the nonlinear prediction and response problems reduce to solving systems of infinite-dimensional linear ordinary differential equations. Similarly, the continuous-time nonlinear filtering problem reduces to solving a system of infinite-dimensional linear stochastic differential equations. Solving the UQ problems then reduces to solving the corresponding truncated linear systems in finitely many diffusion coordinates. By solving these systems we give a model-free algorithm for UQ on gradient flow systems with isotropic diffusion. We numerically verify these algorithms on a 1-dimensional linear gradient flow system where the analytic solutions of the UQ problems are known. We also apply the algorithm to a chaotically forced nonlinear gradient flow system which is known to be well approximated as a stochastically forced gradient flow.

Key words. nonparametric UQ, nonlinear filtering, nonlinear response, diffusion maps, statistical prediction, gradient flows

AMS subject classifications. 58J65, 82C31, 93E11, 60G25

1. Introduction. An important emerging scientific discipline is to quantify the evolution of low-order statistics of dynamical systems in the presence of uncertainties due to errors in modeling, numerical approximations, parameters, initial and boundary conditions [27]. While many uncertainty quantification (UQ) methods were proposed, they rely on some knowledge about the underlying dynamics, at least at the coarse grained levels. One class of popular UQ methods are low-order truncation modeling based on projection of the dynamics onto some basis [27, 16], such as the principle component analysis [12, 14], the truncated polynomial chaos basis [21, 16], the dynamically orthogonal basis [25], the Nonlinear Laplacian Spectral Analysis (NLSA) basis [9], and recently, the “ROMQG” which stands for Reduced Order Modified Quasilinear Gaussian” method [26] which was carefully designed for predicting statistics of turbulent systems.

This paper considers UQ problems in which one has no explicit parametric form for the underlying dynamics. Given only a time series of the underlying dynamics, our goal is to devise UQ methods based on nonparametric modeling, applying ideas from diffusion maps [3, 6, 4]. While the classical Cauchy problem for solving PDE’s is to find the semigroup of the

[†]Department of Mathematics, the Pennsylvania State University, University Park, PA

[‡]Department of Meteorology, the Pennsylvania State University, University Park, PA

[§]This work is partially supported by the Office of Naval Research Grants MURI N00014-12-1-0912

[¶]The research of JH is also partially supported by the Office of Naval Research Grants N00014-13-1-0797 and the National Science Foundation DMS-1317919.

given generators, the diffusion maps technique can be interpreted as the inverse of Cauchy problem. That is, given solutions of a PDE, construct a stochastic matrix, whose generator is a discrete approximation to the backward Kolmogorov operator of the underlying dynamics. Numerically, the stochastic matrix is constructed by evaluating an appropriate kernel function on all pairs of data points and carefully re-normalizing the resulting matrix to account for sampling bias in the data set and to insure the Markov property. In this paper, we apply the recently developed variable bandwidth diffusion kernel [4], which generalizes the original result of diffusion maps introduced in [6] to non-compact manifolds. Our choice is motivated by the fact that this variable bandwidth kernel is more robust and it significantly improves the operator estimation in areas of sparse sampling, e.g. on the tail of a distribution [4]. The UQ methods discussed in this paper are only applicable for a class of nonlinear dynamical systems which can be described by stochastically forced gradient flows, since this is the class of problems for which one can learn the (backward) Kolmogorov operators from diffusion maps algorithm [6, 4]. In the conclusion we discuss the possibility of applying these idea to more general systems.

We will consider three UQ problems, namely the problems of forecasting, filtering [19], and response [17]. While most stochastic projection methods chose the projection basis coordinate [27, 16] depending on the nature of the problems (such as the geometry of the state space or the distribution of the parameters), here we consider solving these UQ problems in the diffusion coordinates. For diffusion processes on nonlinear manifolds, this data-driven basis is the most natural choice since these eigenfunctions implicitly describe both the dynamics and geometry of the underlying dynamical systems. For example, in this coordinate basis, the forecasting problem reduces to solving a diagonal infinite-dimensional systems of linear ODEs. For continuous-time nonlinear filtering problems, the solutions are characterized by the unnormalized conditional distribution which solves the Zakai equation [2]. In diffusion coordinates, we shall see that this nonlinear filtering problem reduces to solving a system of infinite-dimensional multiplicatively forced linear stochastic differential equations. Finally, we consider the response problem studied in [17], which applied the fluctuation-dissipation theorem to estimate the statistical linear response of perturbed systems given the data set of the unperturbed system and the functional form of the external perturbation. Finding the response in our framework requires the diffusion coordinates of the perturbed system. By assuming the perturbation is given by a known change in the potential function, we show that this perturbed basis can be constructed using an appropriate kernel, evaluated on unperturbed data. In this basis, the corresponding nonlinear response problem reduces to solving an infinite dimensional linear system of ODEs. In each case described above, the nonparametric UQ algorithms consist of solving the truncated systems of ODEs and SDEs with finitely many diffusion coordinates.

This paper will be organized as follows: In Section 2, we briefly review the diffusion maps and several computational manipulations for obtaining the appropriate nonparametric probabilistic models by projecting the operators to a basis of the data set itself. In Section 3, we formulate the two UQ problems (prediction and filtering) in the diffusion coordinates. In Section 4, we will formula the third UQ problem (nonlinear response) in the diffusion coordinates. In Section 5, we verify our numerical methods on linear and nonlinear examples. We conclude the paper with a short summary and discussion in Section 6. We accompany this

paper with several movies in the electronic supplementary materials, depicting the evolution of the estimated time-dependent distributions.

2. Diffusion Kernels as Nonparametric Models. Consider a state variable $x \in \mathcal{M}$ evolving on a d -dimensional manifold $\mathcal{M} \subset \mathbb{R}^n$ according to the stochastic gradient flow,

$$(2.1) \quad dx = -\nabla U(x)dt + \sqrt{2D}dW_t,$$

where $U(x)$ denotes the potential of the vector field $x \in \mathcal{M}$ and D is a positive scalar that characterizes the amplitude of a d -dimensional white noise dW_t on the manifold \mathcal{M} . Given data $\{x_i\}_{i=1}^N$ sampled independently from the invariant measure of (2.1), $p_{\text{eq}}(x) \propto \exp(-U/D)$, our goal is to approximate the generator,

$$(2.2) \quad \mathcal{L} = D\Delta - \nabla U \cdot \nabla,$$

with a stochastic matrix constructed from the data set. This Markov matrix is a nonparametric model in the sense that the structure of the manifold \mathcal{M} and the form of the potential function are not assumed to be known. In (2.2), Δ is the Laplacian (with negative eigenvalues), and ∇ is the gradient operator on \mathcal{M} with respect to the Riemannian metric inherited from the ambient space \mathbb{R}^n .

The diffusion maps algorithm [6] generates a stochastic matrix by evaluating a fixed bandwidth isotropic kernel function on all pairs of data points and then renormalizing the matrix and extracting the generator. When the data set lies on a compact manifold \mathcal{M} , the resulting matrix converges to $D^{-1}\mathcal{L}$ in the limit of large data for any smooth potential U . Recently, this result was extended by both authors to non-compact manifolds by using a variable bandwidth kernel [4]. They showed that the variable bandwidth kernels produce significantly improved estimation in areas of sparse sampling (e.g. such as rare events which occur on the tail of a distribution) and are less dependent on the choice of bandwidth. As mentioned in the introduction, we will consider the variable bandwidth kernels [4] as the key ingredient in constructing the nonparametric models in this paper.

In Section 2.1 below, we briefly review the methodology in [4] for approximating the generator $D^{-1}\mathcal{L}$. Since our goal is to approximate \mathcal{L} , then we need to determine D from the data. In Section 2.2 will provide one method to determine D using the correlation time of the data and we also mention several other methods to determine D .

2.1. Approximating the Generator of Stochastic Gradient Flows. The key to our approach is the intuition that continuous notions such as functions and operators have discrete representation in the basis of the data set itself. Given a data set $\{x_i\}_{i=1}^N \subset \mathcal{M} \subset \mathbb{R}^n$, sampled independently from the invariant measure $p_{\text{eq}}(x)$, a function f is represented by a vector $\vec{f} = (f(x_1), f(x_2), \dots, f(x_N))^{\top}$. Similarly, an integral operator $G_{\epsilon}f(x) \equiv \int_{\mathcal{M}} K(x, y)f(y)p_{\text{eq}}(y)dV(y)$, where $dV(y)$ denotes the volume form on \mathcal{M} , is represented by a matrix-vector multiplication between the $N \times N$ matrix, $K_{ij} = K(x_i, x_j)$ and the N -dimensional vector, \vec{f} . With these definitions, the matrix product $K\vec{f}$ yields a vector of length N with i -th component,

$$\frac{1}{N} (K\vec{f})_i = \frac{1}{N} \sum_{j=1}^N K_{ij} \vec{f}_j = \frac{1}{N} \sum_{j=1}^N K(x_i, x_j) f(x_j) \xrightarrow{N \rightarrow \infty} G_{\epsilon}f(x_i),$$

where the limit follows from interpreting the summation as a Monte-Carlo integral. Thus, the matrix K takes functions defined on $\{x_i\}$ to functions defined on $\{x_i\}$, so in this sense we think of K as an operator written in the basis of delta functions $\{\delta_{x_i}\}$ on the data set. Notice the crucial fact that the data is not uniformly distributed on \mathcal{M} , so the operator is biased by the sampling measure $p_{\text{eq}}(y)$. This same bias applies to inner products, if f, g are functions on \mathcal{M} and \vec{f}, \vec{g} are their representations as vectors evaluated at $\{x_i\}$, then the dot product has the following interpretation,

$$\frac{1}{N} \vec{f} \cdot \vec{g} = \frac{1}{N} \sum_{i=1}^N f(x_i)g(x_i) \xrightarrow{N \rightarrow \infty} \int_{\mathcal{M}} f(y)g(y)p_{\text{eq}}(y)dV(y) \equiv \langle f, g \rangle_{L^2(\mathcal{M}, p_{\text{eq}})}.$$

The previous formula shows that inner products weighted by the sampling measure p_{eq} will be easy to compute in this framework.

With the above intuition in mind, we construct a matrix L_ϵ that will converge to the generator $D^{-1}\mathcal{L}$, where \mathcal{L} is defined in (2.2), in the limit as $N \rightarrow \infty$ in the sense that,

$$\lim_{N \rightarrow \infty} \sum_{j=1}^N (L_\epsilon)_{ij} \vec{f}_j = D^{-1}\mathcal{L}f(x_i) + \mathcal{O}(\epsilon).$$

The theory developed in [4] shows that such an L_ϵ can be constructed with a variable bandwidth kernel,

$$(2.3) \quad K_\epsilon(x, y) = \exp \left\{ \frac{-\|x - y\|^2}{4\epsilon\rho(x)\rho(y)} \right\},$$

where the bandwidth function, ρ , is chosen to be inversely proportional to the sampling density, $\rho \approx p_{\text{eq}}^{-1/2}$. This choice enables us to control the error bounds in the area of sparse sample (see [4] for the detailed error estimates). Using this kernel requires us to first estimate the sampling density p_{eq} in order to define the bandwidth function ρ . There are many methods of estimating p_{eq} , such as kernel density estimation methods [23, 22, 28, 24]. In this paper, we estimate p_{eq} using a variable bandwidth kernel method based on the distance to the nearest neighbors (see Section 4 of [4] for details).

Given the kernel in (2.3), we apply the following normalization (first introduced in [6]) to remove the sampling bias,

$$K_{\epsilon,\alpha}(x_i, x_j) = \frac{K_\epsilon(x_i, x_j)}{q_\epsilon(x_i)^\alpha q_\epsilon(x_j)^\alpha}, \quad \text{where} \quad q_\epsilon(x_i) = \sum_{j=1}^N \frac{K_\epsilon(x_i, x_j)}{\rho(x_i)^d}.$$

Note that q_ϵ is a kernel density estimate of the sampling distribution p_{eq} based on the kernel K_ϵ and this is the estimate which will be used in all our numerical examples below. Throughout this paper we will use the normalization with $\alpha = -d/4$ as suggested by the theory of [4], where d is the intrinsic dimension of the manifold \mathcal{M} . We note that d can be estimated from data, see for example [11, 15]. The next step is to construct a Markov matrix from $K_{\epsilon,\alpha}$ by

defining,

$$\hat{K}_{\epsilon,\alpha}(x_i, x_j) = \frac{K_{\epsilon,\alpha}(x_i, x_j)}{q_{\epsilon,\alpha}(x_i)}, \quad \text{where } q_{\epsilon,\alpha}(x_i) = \sum_{j=1}^N K_{\epsilon,\alpha}(x_i, x_j).$$

Finally, the discrete approximation to the continuous generator $D^{-1}\mathcal{L}$ is given by,

$$(2.4) \quad L_\epsilon(x_i, x_j) = \frac{\hat{K}_{\epsilon,\alpha}(x_i, x_j) - \delta_{ij}}{\epsilon \rho(x_i)^2}.$$

By computing the eigenvectors $\vec{\varphi}_i$ and eigenvalues λ_i of L_ϵ , for $i = 0, 1, \dots$, we are approximating the eigenfunctions and eigenvalues of $D^{-1}\mathcal{L}$, respectively. Since D is a scalar constant, obviously, the eigenvectors $\vec{\varphi}_i$ and eigenvalues $D\lambda_i$ are discrete approximations of the eigenfunctions and eigenvalues of the continuous generator \mathcal{L} , respectively. We should note that this trivial argument does not hold when D is a general non-diagonal matrix describing anisotropic diffusion on \mathcal{M} with respect to the Riemannian metric inherited from the ambient space \mathbb{R}^n . In that case, we suspect that one must use a different kernel function, such as the Local Kernels defined in [5] which we discuss in the conclusion.

In addition to the generator \mathcal{L} , we can also approximate the solution semi-group $e^{\epsilon\mathcal{L}}$ by the matrix,

$$F_\epsilon \equiv I_{N \times N} + \epsilon D L_\epsilon,$$

where F_ϵ has the same eigenvectors as L_ϵ but with eigenvalues $\xi_i = 1 + \epsilon D \lambda_i$ where λ_i are the eigenvalues of L_ϵ . For arbitrary t we can approximate the eigenvalues of the solution semi-group $e^{t\mathcal{L}}$ by $\xi_i^{t/\epsilon}$ which converges to $e^{D\lambda_i t}$ as $\epsilon \rightarrow 0$.

In order to insure that the eigenvectors $\vec{\varphi}_i$ (which approximate the eigenfunctions φ_i) form an orthonormal basis with respect to the inner product $\langle \cdot, \cdot \rangle_{L^2(\mathcal{M}, p_{\text{eq}})}$, we note that,

$$\delta_{ij} = \langle \varphi_i, \varphi_j \rangle_{L^2(\mathcal{M}, p_{\text{eq}})} = \int_{\mathcal{M}} \varphi_i(x) \varphi_j(x) p_{\text{eq}}(x) dV(x) = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{l=1}^N (\vec{\varphi}_i)_l (\vec{\varphi}_j)_l = \lim_{N \rightarrow \infty} \frac{\vec{\varphi}_i^\top \vec{\varphi}_j}{N}.$$

Since $\{\vec{\varphi}_i\}_{i=0}^\infty$ are eigenvectors, they are already orthogonal, thus we only need to renormalize $\vec{\varphi}_i$ so that $\vec{\varphi}_i^\top \vec{\varphi}_i = N$. To do this we simply replace $\vec{\varphi}_i$ with $\frac{\sqrt{N}}{\|\vec{\varphi}_i\|} \vec{\varphi}_i$. To simplify the notation below, we define $\langle \cdot, \cdot \rangle_{p_{\text{eq}}} \equiv \langle \cdot, \cdot \rangle_{L^2(\mathcal{M}, p_{\text{eq}})}$.

2.2. Determining the Diffusion Coefficient. The algorithm described in Section 2.1 approximates the generator $D^{-1}\mathcal{L}$, where \mathcal{L} is defined in (2.2). Since our aim is to obtain the generator \mathcal{L} , then we must approximate the scalar diffusion coefficient D . Several non-parametric methods for estimating D had been proposed. For example, [7] considered an optimization problem based on a variational formulation which matches information from the discrete approximation of a conditional Markov chain to the generators of the forward and backward Kolmogorov operators. Another method employed in [20] used one-lag correlation data to approximate the diffusion coefficient.

Here, we present a method to determine D by matching the correlation time, generalizing the Mean Stochastic Model (MSM) [18, 19]. Intuitively, we used the invariant measure to

implicitly determine the potential function (through the kernel based generator) and we now use the correlation time to fit the stochastic forcing. The correlation time is defined by,

$$T_c = \int_0^\infty C(\tau)C(0)^{-1}d\tau,$$

where the correlation function, $C(\tau)$, can be determined from the data by the inverse Fourier transform of the power spectrum, which follows from the Wiener-Kinchin formula, $C(\tau) = \mathcal{F}^{-1}(\|\mathcal{F}(\|x(t)\|)\|^2)$. Note that $C(0)$ represents the variance of the data set.

Once T_c is estimated from the data, we need to find the value of D which makes (2.1) have correlation time T_c . In order to do this, we will show that T_c can be estimated from the eigenvalues and eigenfunctions of \mathcal{L} . Let $I(x) = x$ be the identity map and let $e^{\tau\mathcal{L}}$ be the semigroup solution for the backward equation. Following [17] we have,

$$C(\tau) = \langle x(t + \tau)^\top x(t) \rangle = \int_{\mathcal{M}} (e^{\tau\mathcal{L}}(I)(x))^\top I(x) p_{\text{eq}}(x) dV(x).$$

Writing $I = \sum_i \langle I, \varphi_i \rangle_{p_{\text{eq}}} \varphi_i$ in the eigenbasis of \mathcal{L} , we have,

$$(2.5) \quad C(\tau) = \int_{\mathcal{M}} \sum_i e^{D\lambda_i \tau} \langle I, \varphi_i \rangle_{p_{\text{eq}}}^\top \varphi_i(x) I(x) p_{\text{eq}}(x) dV(x) = \sum_i e^{D\lambda_i \tau} \langle I, \varphi_i \rangle_{p_{\text{eq}}}^\top \langle I, \varphi_i \rangle_{p_{\text{eq}}}.$$

Note that since $\lambda_0 = 0$ and $\varphi_0 = 1$, $\langle I, \varphi_0 \rangle_{p_{\text{eq}}} = \int_{\mathcal{M}} x p_{\text{eq}}(x) dV(x) = \mathbb{E}_{p_{\text{eq}}}[x]$ so we will assume that the data is centered (otherwise the integral of the first term diverges). Noting that $\lambda_i < 0$ for $i > 0$, we can compute the correlation time analytically as,

$$T_c = \int_0^T C(\tau) C(0)^{-1} d\tau = -\frac{\sum_{i \geq 1} (D\lambda_i)^{-1} \langle I, \varphi_i \rangle_{p_{\text{eq}}}^\top \langle I, \varphi_i \rangle_{p_{\text{eq}}}}{\sum_{j \geq 1} \langle I, \varphi_j \rangle_{p_{\text{eq}}}^\top \langle I, \varphi_j \rangle_{p_{\text{eq}}}},$$

and therefore,

$$(2.6) \quad D = -\frac{1}{T_c} \frac{\sum_{i \geq 1} \lambda_i^{-1} \langle I, \varphi_i \rangle_{p_{\text{eq}}}^\top \langle I, \varphi_i \rangle_{p_{\text{eq}}}}{\sum_{j \geq 1} \langle I, \varphi_j \rangle_{p_{\text{eq}}}^\top \langle I, \varphi_j \rangle_{p_{\text{eq}}}}.$$

This gives us an equation for the diffusion coefficient D which matches any given correlation time T_c . Using the empirical correlation time allows us to find the diffusion coefficient of the system defined by the data.

To numerically estimate (2.6), we first approximate,

$$\langle I, \varphi_i \rangle_{p_{\text{eq}}} \approx \frac{1}{N} \sum_{l=1}^N x_l \varphi_i(x_l) = \frac{1}{N} X^\top \vec{\varphi}_i,$$

where X is an $N \times n$ matrix with l -th row given by the n -dimensional data point, x_l . We then approximate (2.6) as a finite sum of M modes,

$$D \approx -\frac{1}{T_c} \frac{\sum_{i=1}^M \lambda_i^{-1} \|X^\top \vec{\varphi}_i\|}{\sum_{i=1}^M \|X^\top \vec{\varphi}_i\|},$$

where $\|X^\top \vec{\varphi}_i\|$ is the ℓ_2 -norm of the n -dimensional vector $X^\top \vec{\varphi}_i$.

For one-dimensional linear problems, (2.6) reduces to the mean stochastic model (MSM) introduced in [18, 19]. To see this, let the potential function be $U(x) = -\alpha x^2/2 \in \mathbb{R}$, so that the gradient flow system in (2.1) is a one-dimensional Ornstein-Uhlenbeck (OU) process and the eigenfunctions of \mathcal{L} are the Hermite polynomials $\varphi_i(x) = H_i(x)$. All the inner products $\langle I, \varphi_i \rangle_{p_{\text{eq}}} = \langle x, H_i(x) \rangle_{p_{\text{eq}}} = \langle H_1(x), H_i(x) \rangle_{p_{\text{eq}}}$ are zero except for $i = 1$, in which case the inner product is 1 and the associated eigenvalue is $\lambda_1 = \alpha/D$ so that (2.6) becomes $\alpha = -1/T_c$ which is the MSM formula. For nonlinear problems, the correlation function (2.5) for the gradient flow system is more accurate than that of the MSM fit as we will show in Section 5.

3. Forecasting and Filtering in Diffusion Coordinates. In this section we formulate two UQ problems, namely forecasting and filtering, in diffusion coordinates. These two problems involve pushing a probability measure forward in time, and therefore they involve the Fokker-Planck operator \mathcal{L}^* . By projecting these infinite dimensional systems onto the eigenfunctions of the Fokker-Planck operator, we find infinite dimensional linear systems which govern the evolution on the projected coordinates. We can then approximate the solution by truncating the linear system to describe finitely many coordinates.

For gradient flows, it is easy to determine the eigenvalues and eigenfunctions of the Fokker-Planck operator, \mathcal{L}^* , from the eigenvalues and eigenfunctions of the generator \mathcal{L} . To see this, note that $p_{\text{eq}} \propto e^{-U/D}$, and,

$$(3.1) \quad \begin{aligned} \frac{1}{p_{\text{eq}}} \mathcal{L}^*(f p_{\text{eq}}) &= \frac{1}{p} \operatorname{div}(D \nabla(f p_{\text{eq}}) + f p_{\text{eq}} \nabla U) = \frac{1}{p_{\text{eq}}} \operatorname{div}(D p_{\text{eq}} \nabla f + D f \nabla p_{\text{eq}} + f p_{\text{eq}} \nabla U) \\ &= \frac{1}{p_{\text{eq}}} \operatorname{div}(D p_{\text{eq}} \nabla f) = D \Delta f + D \nabla f \cdot \frac{\nabla p_{\text{eq}}}{p_{\text{eq}}} = D \Delta f - \nabla f \cdot \nabla U = \mathcal{L} f, \end{aligned}$$

since $D \nabla p_{\text{eq}} = -p_{\text{eq}} \nabla U$. Also, it follows from (3.1) that $\mathcal{L}^* f = p_{\text{eq}} \mathcal{L}(f/p_{\text{eq}})$. The two operators have the same eigenvalues and it is easy to show that the semi-group solutions are related by $e^{t\mathcal{L}^*} f = p_{\text{eq}} e^{t\mathcal{L}}(f/p_{\text{eq}})$ and the eigenfunctions ψ_i of \mathcal{L}^* and $e^{t\mathcal{L}^*}$ are given by $\psi_i = p_{\text{eq}} \varphi_i$ where φ_i are the eigenfunctions of \mathcal{L} and $e^{t\mathcal{L}}$. From the orthonormality of φ_i , it is easy to deduce that,

$$(3.2) \quad \langle \psi_i, \psi_j \rangle_{1/p_{\text{eq}}} = \langle \psi_i, \varphi_j \rangle = \langle \varphi_i, \varphi_j \rangle_{p_{\text{eq}}} = \delta_{ij}.$$

Of course, since these eigenfunctions are all represented by N -dimensional eigenvectors (which represent evaluation on the data set $\{x_j\}_{j=1}^N$) the actual computation of any inner product will always be realized by rewriting it with respect to p_{eq} , following the rule in (3.2).

3.1. Nonlinear forecasting. The *forecasting* problem is, given an arbitrary initial distribution $p_0(x)$ for a state variable x , find the density $p(x, t)$ at any future time t . This will be the most straightforward problem to solve in diffusion coordinates since $p(x, t)$ solves the Fokker-Planck equation,

$$(3.3) \quad \frac{\partial p}{\partial t} = \mathcal{L}^* p, \quad p(x, 0) = p_0(x).$$

Let ψ_i be the eigenfunctions of \mathcal{L}^* with eigenvalues $D\lambda_i$, and write p in this basis as, $p(x, t) = \sum_i c_i(t)\psi_i(x)$ where $c_i = \langle p, \psi_i \rangle_{1/p_{\text{eq}}}$. We define the vector $\vec{c} = \vec{c}(t)$ of eigencoordinates and the diagonal matrix $\Lambda_{ii} = \lambda_i$, and writing (3.3) in these coordinates we have,

$$(3.4) \quad \frac{d\vec{c}}{dt} = D\Lambda\vec{c}, \quad \vec{c}(0) = \langle p_0, \psi_i \rangle_{1/p_{\text{eq}}}.$$

In order to solve the linear system in (3.4), we first need to find the initial conditions $c_i(0)$ by projecting the initial density $p(x, 0) = p_0(x)$ onto ψ_i . Since densities $p_0(x)$ and $p_{\text{eq}}(x)$ are given on the training data set x_i , we define the discrete densities $(\vec{p}_0)_l = p_0(x_l)$ and $(\vec{p}_{\text{eq}})_l = p_{\text{eq}}(x_l)$. Following the inner product rule in (3.2), we write the projection of the initial condition as,

$$(3.5) \quad c_i(0) = \langle p_0, \psi_i \rangle_{1/p_{\text{eq}}} = \langle p_0/p_{\text{eq}}, \varphi_i \rangle_{p_{\text{eq}}} \approx \frac{1}{N} \sum_{j=1}^N \frac{p_0(x_j)}{p_{\text{eq}}(x_j)} \varphi_i(x_j),$$

since this converges to the true value as $N \rightarrow \infty$. With this initial condition, each component in (3.4) can be solved analytically as $c_i(t) = e^{tD\lambda_i} c_i(0)$. Numerically, we will approximate the solutions with finitely many modes, $i = 0, 1, \dots, M$. In Section 3.3 we show how to use these solutions to reconstruct the density $p(x, t)$ or to estimate a quantity of interest at any time t .

3.2. Nonlinear filtering. The *filtering* problem is, given a time series of noisy observations z of the state variable x , find the posterior density $P(x, t|z(s), s \leq t)$ given all the observations $z(s)$ up to time t . For an observation function $h : \mathcal{M} \rightarrow \mathbb{R}^m$, we consider continuous-time observations of the form,

$$(3.6) \quad dz = h(x) dt + \sqrt{R} dW_t,$$

where dW_t are i.i.d Gaussian noise in \mathbb{R}^m . With these observations, the evolution of the unnormalized posterior density, $p(x, t)$, defined as follows,

$$P(x, t|z(s), s \leq t) = \frac{p(x, t)}{\int_{\mathcal{M}} p(x, t) dV(x)},$$

is given by the Zakai equation [2],

$$(3.7) \quad dp = \mathcal{L}^* p dt + ph^\top R^{-1} dz.$$

Writing $p(x, t) = \sum_i c_i(t)\psi_i(x)$ the evolution becomes,

$$\sum_i dc_i \psi_i = \sum_i D\lambda_i c_i \psi_i dt + \sum_i c_i \psi_i h^\top R^{-1} dz,$$

and projecting both sides of this equation onto ψ_j we find,

$$(3.8) \quad dc_j = D\lambda_j c_j dt + \sum_i c_i \langle \psi_i h^\top, \psi_j \rangle_{1/p_{\text{eq}}} R^{-1} dz,$$

which is an infinite-dimensional system of stochastic differential equations with multiplicative stochastic forcing.

To numerically estimate the solution of (3.8), we truncate this infinite-dimensional system of SDEs by solving only a system of $(M + 1)$ -dimensional SDEs for $\vec{c} = (c_0, c_1, \dots, c_M)^\top$. This strategy is simply a Galerkin approximation for the Zakai equation for which we choose the diffusion coordinates as a basis rather than the Gaussian series as proposed in [1] or the Hermite polynomial basis as proposed in [8]. We should note that for scalar OU processes, the diffusion coordinates are exactly the Hermite polynomials since these polynomials are the eigenfunctions of the backward Kolmogorov operator for the OU process.

Defining $H_{ji} = \langle \psi_i h^\top, \varphi_j \rangle$ as a $1 \times m$ dimensional vector for each pair (i, j) ; where the k -th component is given by $H_{ji}^k = \langle \psi_i h_k, \varphi_j \rangle$, we can write the truncated $(M+1)$ -dimensional system of SDEs in compact form as follows,

$$(3.9) \quad d\vec{c} = D\Lambda\vec{c}dt + (H\vec{c})R^{-1}dz.$$

To solve this system of SDEs, we first project the initial condition $p_0(x)$ to obtain $c_i(0)$ as in (3.5) for $i = 0, 1, \dots, M$. We then numerically solve the system of SDEs in (3.9) using the splitting-up method (see for example [8]). Explicitly, given $\vec{c}(t_{i-1})$ we compute $\vec{c}(t_i)$ by first using the solution to the deterministic part,

$$\vec{c}_0(t_i) = \exp(D\Lambda\Delta t)\vec{c}(t_{i-1}),$$

and then the solution to the stochastic part,

$$\vec{c}(t_i) = \exp\left(HR^{-1}dz(t_i) - \frac{1}{2}H^2R^{-1}\vec{1}\Delta t\right)\vec{c}_0(t_i).$$

Here, the exponent term consists of an $M \times M$ matrix given by

$$HR^{-1}dz(t_i) - \frac{1}{2}H^2R^{-1}\vec{1}\Delta t = \sum_k H^k(R^{-1}dz(t_i))_k - \frac{1}{2}(H^k)^2(R^{-1}\vec{1})_k\Delta t.$$

Notice that unlike the forecasting problem which had an analytic solution for any time t , this procedure must be iterated for each observation $dz(t_i)$.

3.3. Quantity of Interest. In the uncertainty quantification problems described above, we reduce the PDE and SPDE problems which describe the evolution of density $p(x, t)$ into systems of ODEs in (3.4) and SDEs in (3.9), respectively, for the coefficients $c_i(t)$. Using the coefficients $c_i(t)$ we can approximate the density function as follows,

$$(3.10) \quad p(x, t) \approx p_M(x, t) = \sum_{i=0}^M c_i(t)\psi_i(x) = \sum_{i=0}^M c_i(t)\varphi_i(x)p_{\text{eq}}(x),$$

where the coefficients $\{c_i(t)\}_{i=0}^M$ are from the solution of the $(M + 1)$ -dimensional systems of ODEs or SDEs.

One complication in some of the algorithms above is that the evolution of the coefficients may not preserve the normalization of the corresponding density. In fact, for the nonlinear filtering problems in (3.7), the solutions are unnormalized densities. Fortunately we can use

the eigenfunctions φ_i to estimate the normalization factor of an unnormalized density from its diffusion coordinates. Assume that p_M in (3.10) is an unnormalized density with diffusion coordinates c_i obtained from the solution of the $(M + 1)$ -dimensional systems of ODEs or SDEs. We can then estimate the normalization constant as,

$$Z = \int_{\mathcal{M}} p_M(x, t) dV(x) = \sum_{i=0}^M c_i \int_{\mathcal{M}} \varphi_i(x) p_{\text{eq}}(x) dV(x) = \sum_{i=0}^M c_i \langle 1, \varphi_i \rangle_{p_{\text{eq}}}.$$

where we estimate inner product, $\langle 1, \varphi_i \rangle_{p_{\text{eq}}} \approx \frac{1}{N} \sum_{j=1}^N \varphi_i(x_j)$. Subsequently, we normalize our density by replacing p_M with p_M/Z . In the remainder of this paper, we will refer to p_M as the normalized density without ambiguity. We also note that by reconstructing the density from the coefficients we are implicitly projecting the density into the subspace of Hilbert space spanned by the first $M + 1$ eigenfunctions. This has a smoothing effect which improves the clarity of figures and we always show reconstructed densities in the figures and videos.

In many cases, we are not interested in the entire density p ; instead we are often interested in the expectation of an observable $A : \mathcal{M} \rightarrow \mathbb{R}$ with respect to p , namely $\mathbb{E}_{p(x,t)}[A(x)]$. A common example of functionals of interest are the moments $A(x) = x^m$ for $m \in \mathbb{N}$. Although A is a functional on \mathcal{M} , A can be given as a functional on the ambient space \mathbb{R}^n because we only evaluate A on the data set, which lies on the manifold \mathcal{M} . Using the formula (3.10) for the density estimate at time t , we can approximate the expectation of the functional A at time t as,

$$(3.11) \quad \mathbb{E}_{p(x,t)}[A(x)] \approx \langle A, p_M \rangle = \sum_{i=0}^M c_i(t) \langle A, \varphi_i \rangle_{p_{\text{eq}}} = \sum_{i=0}^M c_i(t) a_i = \vec{c}(t)^{\top} \vec{a}.$$

The formula (3.11) shows that we can find the expectation of a functional A by writing A in diffusion coordinates as $a_i = \langle A, \varphi_i \rangle_{p_{\text{eq}}}$. In these coordinates, the expectation of A is simply the inner product of the diffusion coefficients of A and $p(x, t)$. Finally, if the quantity of interest is given by a functional A evaluated on the training data set as $(\vec{A})_l = A(x_l)$, we approximate the inner product $a_i = \langle A, \varphi_i \rangle_{p_{\text{eq}}} \approx \frac{1}{N} \vec{A}^{\top} \vec{\varphi}_i$.

4. Nonlinear Response in Diffusion Coordinates. The response problem considered in [17] is to determine the change in the expectation of a functional $A(x)$ as function of time, after perturbing the dynamics of a system at equilibrium. Letting $p^{\delta}(x, t)$ be the density evolved according to the perturbed system, we define the response of a functional A as the expected difference,

$$(4.1) \quad \delta \mathbb{E}[A(x)](t) = \mathbb{E}_{p^{\delta}}[A(x)](t) - \mathbb{E}_{p_{\text{eq}}}[A(x)], \quad t \geq 0,$$

We assume the unperturbed system has the form,

$$(4.2) \quad dx = F(x) dt + \sqrt{2D} dW.$$

with equilibrium density p_{eq} and Fokker-Planck operator \mathcal{L}^* , so that $\mathcal{L}^* p_{\text{eq}} = 0$. We assume the perturbed system has the form,

$$(4.3) \quad dx = (F(x) + \delta F(x, t)) dt + \sqrt{2D} dW,$$

with associated Fokker-Planck operator $\tilde{\mathcal{L}}^*$ so that p^δ solves the Fokker-Planck equation,

$$(4.4) \quad \frac{\partial p^\delta}{\partial t} = \tilde{\mathcal{L}}^* p^\delta = \mathcal{L}^* p^\delta + \delta \mathcal{L}^* p^\delta,$$

where $\delta \mathcal{L}^* p = -\operatorname{div}(\delta F p)$ denotes the Liouville operator corresponding to the external forcing with functional form δF .

This reveals that the response problem is closely related to the forecasting problem. Given the Fokker-Planck operator $\tilde{\mathcal{L}}^*$ for the perturbed system, one must solve (4.4) with initial condition chosen to be the equilibrium distribution of the unperturbed system, $p^\delta(x, 0) = p_{\text{eq}}(x)$. Since computational cost becomes an issue, especially for higher dimensional problems, one typically approximates p^δ by solving ensemble solutions of (4.3) with initial ensemble sampled from p_{eq} , and in fact we will use this approach as a benchmark in Section 5.2.3. However, as in [17], we are interested in cases where the Fokker-Planck operators \mathcal{L}^* and $\tilde{\mathcal{L}}^*$ and also the functional form of F are all unknown. We will assume that we only have access to data from the unperturbed system and the functional form of the external perturbation δF . Of course, since the techniques of Section 2 are restricted to gradient flow systems, we will assume that both the unperturbed and perturbed systems are well approximated by stochastically forced gradient flows.

For gradient flow systems, we have $F(x) = -\nabla U(x)$ and $\delta F(x) = -\nabla \delta U(x)$, where the perturbation is a time-independent potential. Given a data set from $\{x_i\}_{i=1}^N$ sampled from p_{eq} and the functional form of δU , it is tempting to proceed as in Section 3 by writing $p^\delta(x, t) = \sum_i c_i(t) \psi_i(x)$ and projecting (4.4) to eigenfunctions of \mathcal{L}^* . The main issue with this projection is that it is difficult to computationally attain the external perturbed forcing term,

$$\langle \delta \mathcal{L}^* \psi_i, \psi_j \rangle_{1/p_{\text{eq}}} = -\langle \operatorname{div}(\delta F \psi_i), \psi_j \rangle_{1/p_{\text{eq}}} = \langle \delta F, \psi_i \nabla \varphi_j \rangle = \langle \delta U, \operatorname{div}(\varphi_i \nabla \varphi_j) \rangle_{p_{\text{eq}}}.$$

Even if δF or δU are known, we have no provably convergent technique to compute the gradient or divergence operators on the unknown Riemannian manifold. In fact, the technique discussed in Section 2 does not provide a framework to represent vector fields on manifolds.

As a remedy, we propose to project (4.4) onto the eigenfunctions of $\tilde{\mathcal{L}}^*$ which can be accessed from the functional form of δU and the data set $\{x_i\}$. This is an application of a result in [4], which showed that one can approximate the generator of any gradient flow system with a (partially) known potential even if the data is sampled from a different gradient flow system. In particular, it was shown (see eq.(17) in [4]) that for general choice of bandwidth function ρ in (2.3) and any $f \in L^2(\mathcal{M}, p_{\text{eq}}) \cap \mathcal{C}^3(\mathcal{M})$, the matrix L_ϵ constructed in (2.4) converges to a weighted Laplacian in the following sense,

$$(4.5) \quad L_\epsilon f = \Delta f + 2 \left(1 + \frac{d}{4}\right) \nabla f \cdot \frac{\nabla p_{\text{eq}}}{p_{\text{eq}}} + (d+2) \nabla f \cdot \frac{\nabla \rho}{\rho} + \mathcal{O}(\epsilon).$$

If we choose the following bandwidth function,

$$(4.6) \quad \rho = p_{\text{eq}}^{-1/2} e^{-\frac{\delta U}{D(d+2)}},$$

where p_{eq} is estimated from the data set as before, substituting this into (4.5), we obtain,

$$\begin{aligned}
 L_\epsilon f &= \Delta f + \left(2 + \frac{d}{2}\right) \nabla f \cdot \frac{\nabla p_{\text{eq}}}{p_{\text{eq}}} - \nabla f \cdot \frac{\nabla \delta U}{D} - \left(\frac{d}{2} + 1\right) \nabla f \cdot \frac{\nabla p_{\text{eq}}}{p_{\text{eq}}} + \mathcal{O}(\epsilon) \\
 &= \Delta f + \nabla f \cdot \frac{\nabla p_{\text{eq}}}{p_{\text{eq}}} - \nabla f \cdot \frac{\nabla \delta U}{D} + \mathcal{O}(\epsilon) \\
 &= \Delta f - \nabla f \cdot \frac{(\nabla U + \nabla \delta U)}{D} + \mathcal{O}(\epsilon) \\
 (4.7) \quad &= D^{-1} \tilde{\mathcal{L}} f + \mathcal{O}(\epsilon).
 \end{aligned}$$

With this result, we construct a stochastic matrix L_ϵ following exactly the procedure described in Section 2.1, except with the bandwidth function in (4.6).

Let $\tilde{\lambda}_i$ and $\tilde{\varphi}_i$ be the eigenvalues and eigenfunctions of L_ϵ , which is a discrete approximation to the continuous operator $D^{-1} \tilde{\mathcal{L}}$ as shown in (4.7). Following the same argument as in Section 3, the corresponding Fokker-Planck operator, $\tilde{\mathcal{L}}^*$, has eigenvalues $D \tilde{\lambda}_i$ with eigenfunctions, $\tilde{\psi}_i = \tilde{\varphi}_i \tilde{p}_{\text{eq}}$, where $\tilde{p}_{\text{eq}} \propto p_{\text{eq}} e^{-\delta U/D}$ is the equilibrium measure of the perturbed system. Now letting $p^\delta(x, t) = \sum_i \tilde{c}_i(t) \tilde{\psi}_i(x)$ and projecting (4.4) with initial condition, $p^\delta(x, 0) = p_{\text{eq}}(x)$, on the perturbed coordinate basis, the evolution of the perturbed density becomes a linear forecasting problem,

$$(4.8) \quad \frac{d\tilde{c}_i}{dt} = D \tilde{\lambda}_i \tilde{c}_i, \quad \tilde{c}_i(0) = \langle p_{\text{eq}}, \tilde{\psi}_i \rangle_{1/\tilde{p}_{\text{eq}}}.$$

Therefore, the response in (4.1) formula can be approximated by,

$$(4.9) \quad \delta \mathbb{E}[A](t) \approx \sum_{i=1}^M (\tilde{c}_i(t) - \tilde{c}_i(0)) \langle A, \tilde{\varphi}_i \rangle_{\tilde{p}_{\text{eq}}}, \quad t \geq 0$$

utilizing (3.11) on the perturbed diffusion coordinates. Notice that the zeroth mode is excluded in the summation in (4.9) since $\tilde{\lambda}_0 = 0$ and therefore, $\tilde{c}_0(t) = \tilde{c}_0(0)$, for any t .

A complication in numerically implementing the above technique is that the eigenfunctions $\tilde{\varphi}_i$ are orthonormal with respect to the inner product $\langle \cdot, \cdot \rangle_{\tilde{p}_{\text{eq}}}$ weighted by the perturbed equilibrium density, \tilde{p}_{eq} , but we can only estimate inner products weighted with respect to the unperturbed equilibrium density, p_{eq} . This requires us to rewrite all of the inner products with respect to p_{eq} . In particular, we can rewrite the initial coefficients in (4.8) as,

$$\tilde{c}_i(0) = \langle p_{\text{eq}}, \tilde{\psi}_i \rangle_{1/\tilde{p}_{\text{eq}}} = \langle p_{\text{eq}}, \tilde{\varphi}_i \rangle = \langle 1, \tilde{\varphi}_i \rangle_{p_{\text{eq}}} \approx \frac{1}{N} \sum_{j=1}^N \tilde{\varphi}_i(x_j).$$

With these initial conditions, we have explicit solutions for the linear problem in (4.8) given by $\tilde{c}_i(t) = e^{D \tilde{\lambda}_i t} \tilde{c}_i(0)$. To compute the response formula in (4.9), we evaluate the inner product terms in (4.9) as follows,

$$\langle A, \tilde{\varphi}_i \rangle_{\tilde{p}_{\text{eq}}} = \langle A, \frac{\tilde{p}_{\text{eq}}}{p_{\text{eq}}} \tilde{\varphi}_i \rangle_{p_{\text{eq}}} \approx \frac{1}{N} \sum_{j=1}^N A(x_j) \frac{\tilde{p}_{\text{eq}}(x_j)}{p_{\text{eq}}(x_j)} \tilde{\varphi}_i(x_j) = \frac{1}{NZ} \sum_{j=1}^N A(x_j) e^{-\delta U(x_j)/D} \tilde{\varphi}_i(x_j),$$

using the fact that the equilibrium density of the perturbed system is given by, $\tilde{p}_{\text{eq}} = \frac{1}{Z} p_{\text{eq}} e^{-\delta U/D}$, where Z is the normalization factor for the density \tilde{p}_{eq} obtained as discussed in Section 3.3.

5. Numerical Results. For a state variable x evolving according to (2.1), we have shown how to estimate the density function $p(x, t)$, which solves either the forecasting, filtering, or response problem. We have also shown how to find the expectation of any quantity of interest, $A(x)$, with respect to this measure. In this section, we first validate the numerical algorithms developed in Sections 3 and 4 on a simple linear gradient flow system. We then test our algorithms on a chaotically forced gradient flow system which is known to be well approximated as a stochastically forced gradient flow. In each example we will use only $M = 50$ eigenvectors in the non-parametric model.

5.1. Linear Example: Ornstein-Uhlenbeck processes. We first validate our algorithms on the Ornstein-Uhlenbeck processes on $\mathcal{M} = \mathbb{R}$, given by,

$$(5.1) \quad dx = -x dt + \sqrt{2} dW_t,$$

where dW_t denotes standard Gaussian white noise. Our technique assumes that we do not know the model in (5.1), but instead we are only given sample solutions of (5.1) as a training data set. In our numerical experiment, we generate a training data set consisting of $N = 10000$ data points $x_l = x(l\Delta t)$ by numerically integrating (5.1) with the Euler-Maruyama method and sub-sample at every $\Delta t = 0.2$ time units. We intentionally sub-sample with a large Δt so that the data set is approximately independently sampled from the invariant measure. We numerically validate our methods on this simple system (5.1) since it is possible to find an analytic expression for the evolution of the moments (see Appendix A) for the forecasting and response problems. For the filtering problem, we compare the results to the Kalman-Bucy solutions which are optimal in a least square sense [13].

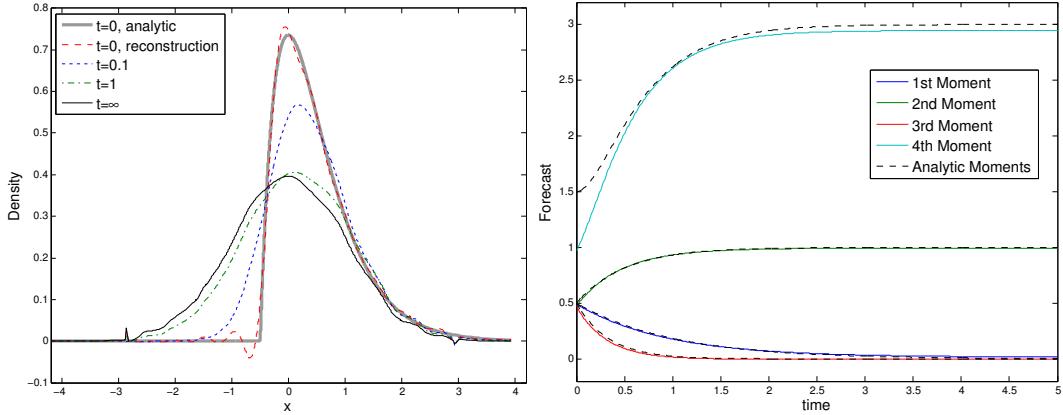


Figure 1. Forecasting the evolution of a Gamma distributed initial condition in the Ornstein-Uhlenbeck system. Left: Evolution of the density. Right: Evolution of the moments.

5.1.1. Forecasting. To validate our forecasting algorithm, we take the initial condition to be a Gamma distribution with density $p_0(x) = 4(x + 1/2)e^{-2(x+1/2)}$ on $[-1/2, \infty)$. For this

distribution the first four moments are, $\bar{x} = \mathbb{E}_{p_0}[x] = 1/2$, $\mathbb{E}_{p_0}[(x - \bar{x})^2] = 1/2$, $\mathbb{E}_{p_0}[(x - \bar{x})^3] = 1/2$, and $\mathbb{E}_{p_0}[(x - \bar{x})^4] = 3/2$. We assume that the initial distribution is known, and we simply evaluate p_0 on the training data set to find $(\vec{p}_0)_l = p_0(x_l)$. The results of the forecasting algorithm are shown in Figure 1 where we reconstruct the forecast density at various times as it approaches the invariant measure. Notice the oscillation in the reconstructed p_0 near the non-smooth part is a Gibbs phenomenon-like behavior. We also show the evolution of the first four centered moments which are evaluated as linear combinations of the uncentered moments $A(x) = x^r$, $r = 1, \dots, 4$. Notice that the initial kurtosis of the Gamma distribution is difficult to estimate since the data on the tail is very sparse (there are only 3 sample points for $x > 3$ in the training data set). The Monte-Carlo integral converges slowly since the Gamma distribution decays slower than the Gaussian distribution.

5.1.2. Filtering. To verify our filtering algorithm, we consider the observation (3.6) with $h(x) = x$ so that,

$$dz = x dt + \sqrt{R} dW_t.$$

The filtering problem for (5.1) with this observation has optimal solutions given by the Kalman-Bucy equations [13], which we approximate for a finite observation time step Δt by the discrete time Kalman filter. Since discrete time observations are typically given in the form,

$$\mathcal{Z}(t_i) = x(t_i) + \sqrt{R_o} \omega(t_i)$$

where $\omega(t_i)$ are independent random samples from $\mathcal{N}(0, 1)$, we will approximate dz for a finite time step $\Delta t = 0.01$ as,

$$dz(t_i) \approx x(t_i)\Delta t + \sqrt{\Delta t R_o} dW_{t_i} = \mathcal{Z}(t_i)\Delta t.$$

So in the numerical algorithm of Section 3.2 we use $dz(t_i) = \mathcal{Z}(t_i)\Delta t$ and we set $R = \Delta t R_o$.

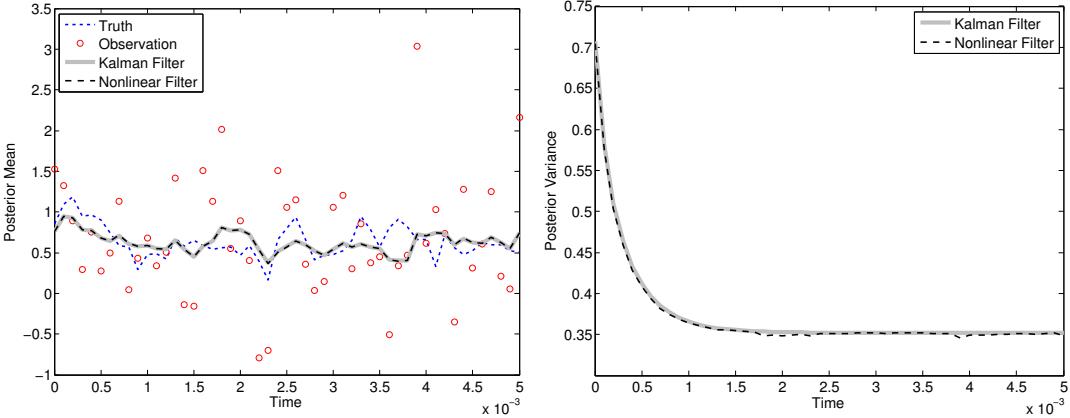


Figure 2. Nonlinear filter mean estimates as functions of time, compared to the truth, Kalman filter solutions, and observations (left). The posterior covariance estimates as functions of time (right). The nonlinear filter recovers the Kalman filter solutions in this linear example. The posterior evolution is shown in `fig2.mov`.

In our numerical simulation, we take $R_o = 1$ and assume an initial distribution $p_0(x) = p_{\text{eq}}(x)$, which is the Gaussian invariant measure of (5.1). The results are shown in Figure 2,

where we show that we recover the same statistical solutions (mean and covariance) as the discrete Kalman filter. We also include a video, `fig2.mov`, comparing the evolution of the reconstructed posterior density with the Kalman filter posterior.

5.1.3. Response. To validate the response algorithm of Section 4, we perturb the Ornstein-Uhlenbeck system in (5.1), which has potential $U(x) = x^2/2$, with a potential $\delta U(x) = \frac{a}{2}(x - b/a)^2$. The vector field for the potential $U + \delta U$ becomes $-\nabla(U + \delta U) = -x - ax + b$ so that the perturbed system is again an Ornstein-Uhlenbeck system with different mean and damping parameters. This fact allows us to easily compute the analytic response (see Appendix A) which we compare to the numerical estimate using the above algorithm in Figure 3. In this example we choose $a = -0.1$ and $b = 0.03$, since the damping is decreased by this perturbation the variance increases in response while the effect of b is a shift in the mean. Notice the accuracy of the response estimates for the mean and variance at all times.

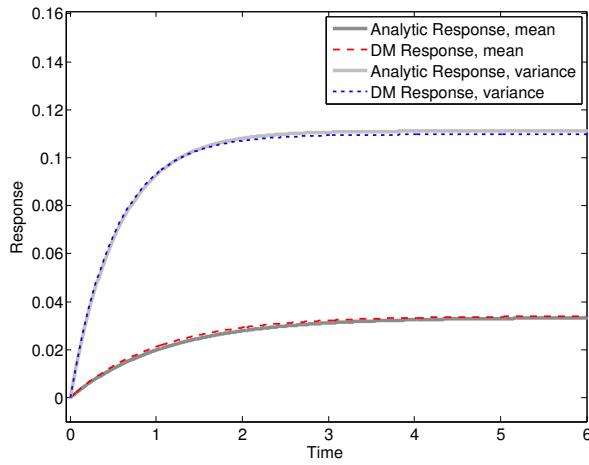


Figure 3. Nonlinear response of the first two moments of the Ornstein-Uhlenbeck system to a perturbation of the potential function. The analytic response is derived in Appendix A and the DM response is our nonparametric estimate.

5.2. Nonlinear Example: Chaotically driven double well potential. In this section we test our nonparametric UQ methods on the following nonlinear system,

$$(5.2) \quad \begin{aligned} \dot{x} &= x - x^3 + \frac{\gamma}{\epsilon} y_2, \\ \dot{y}_1 &= \frac{10}{\epsilon^2} (y_2 - y_1), \\ \dot{y}_2 &= \frac{1}{\epsilon^2} (28y_1 - y_2 - y_1 y_3), \\ \dot{y}_3 &= \frac{1}{\epsilon^2} (y_1 y_2 - \frac{8}{3} y_3). \end{aligned}$$

which was discussed in [10] as an example of a stochastic homogenization problem and used in [20] as a test model for a reduced filtering problem. The fast variables (y_1, y_2, y_3) solve a chaotic Lorenz-63 system, and x is a scalar ODE driven by this chaotic oscillator with

characteristic time ϵ^2 . In [10, 20] it was shown that for γ, ϵ sufficiently small the dynamics of the variable x are well approximated by the reduced stochastic model.

$$(5.3) \quad dX = X(1 - X^2) dt + \sigma dW_t,$$

where σ is a diffusion constant that can be numerically estimated by a linear regression based algorithm or by matching the correlation time of γy_2 . This reduced model is a stochastically forced gradient flow system, where the potential function $U(X) = X^2/2 - X^4/4$ is a standard double well potential. While this suggests that our technique should be successful, we emphasize that our method makes no use of either (5.2) or (5.3), instead it relies on a training data set to represent the reduced model for the variable x , with no assumed parametric form. Following [20], we set $\gamma = 4/90$ and $\epsilon = \sqrt{0.1}$. We note that [20] suggested $\sigma^2 = 0.113$ for the homogenization and $\sigma^2 = 0.126$ for the filtering problem.

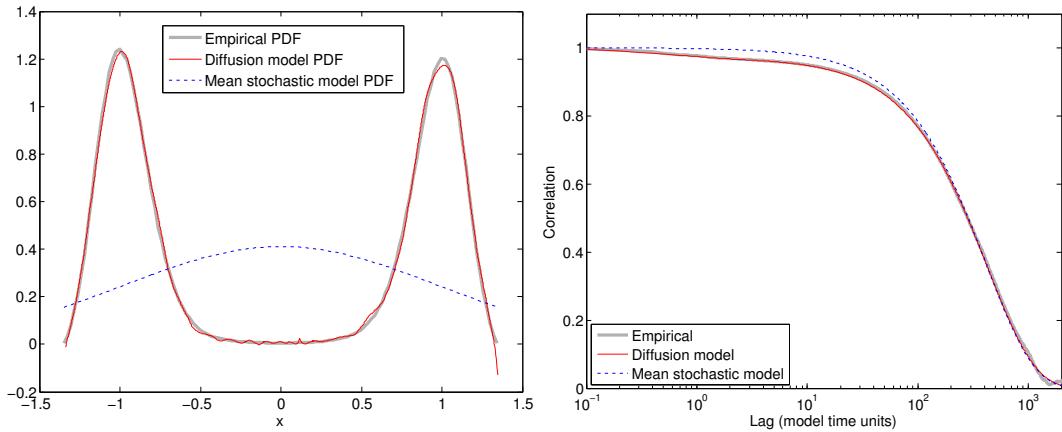


Figure 4. Left: Invariant measure estimated by the kernel. Right: Tuning the stochastic forcing using the correlation time.

In order to apply our method to this problem, we generated a training data set containing 200000 data points $\{x_j = x(t_j)\}$ with time spacing $\Delta t = t_{j+1} - t_j = .1$ by solving (5.2) using a fourth order Runge-Kutta scheme with integration time step $\delta t = 0.002$. We applied the variable bandwidth diffusion kernel as described in Section 2.1 to the data set $\{x_{10j}\}_{j=1}^{20000}$. We then estimate the diffusion constant $D = \sigma^2/2 \approx 0.0535$ as in Section 2.2, using the entire data set $\{x_j\}$ to find the correlation time and the empirical correlation function. In Figure 4 we compare the invariant measure and correlation function of the nonparametric diffusion model to the empirical estimates of these statistics taken from training data set. We also show the statistics of the mean stochastic model (MSM) of [18, 19], which is a linear (and Gaussian) model constructed by matching the correlation time and the equilibrium density variance. From Figure 4, one can see that the density and the autocorrelation function from the diffusion map fitting are very accurate.

5.2.1. Forecasting. We first test our forecasting algorithm for an initial condition given by a normal distribution centered at 0.5 with variance 0.01. Unlike the Ornstein-Uhlenbeck system in Section 5.1, we do not have an analytical expressions for the evolution the moments.

For diagnostic purposes, we will compare our forecast to a Monte-Carlo simulation. We initialize an ensemble of 100000 points $(x, y_1, y_2, y_3)^\top \in \mathbb{R}^4$ with x chosen randomly according to the initial density and triples $(y_1, y_2, y_3)^\top$ randomly selected from the training data set (so that they lie on the attractor of the chaotic system). This ensemble is then evolved according to the system (5.2) with the same Runge-Kutta integrator used to generate the training data. For any time t we generate the Monte-Carlo density using a histogram of the ensemble at time t , and we call the moments of the ensemble the Monte-Carlo moments. These Monte-Carlo estimates make use of the full true model and are computationally more intensive than our model free method, and we will consider them the ‘truth’ for the purposes of comparison.

Next, we estimate the evolution of the density and the first four centered moments using the technique of Section 3. For this initial condition, the density moves to the right and centers on the right well of the potential on the first 10 model time units. Over the next 1000 model time units, small amounts of density slowly migrate across the barrier between the two potential wells until the system is at equilibrium. In Figure 5 we compare the moment estimates of the nonparametric method to the Monte-Carlo moments. We also show snapshots of the evolution of the density as reconstructed from the diffusion coordinates at times $t = 0, 10, \infty$. The evolution of the reconstructed density is compared to the Monte-Carlo simulation in `fig5a.mov` up to time $t = 10$ and `fig5b.mov` up to time $t = 2000$.

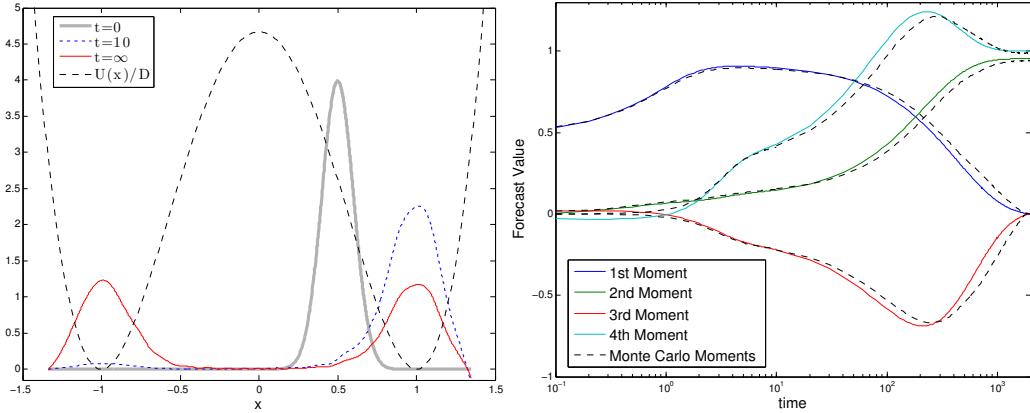


Figure 5. Left: Evolution of the density is summarized in snapshots for $t = 0, 10, \infty$ and the potential function is shown. Right: Non-parametric estimates for the evolution of the moments is compared to the Monte-Carlo simulation. The reconstructed density evolution is shown in `fig5a.mov` and `fig5b.mov`.

5.2.2. Filtering. We now test our nonlinear filtering algorithm on the chaotically driven double well potential using the observation (3.6) with $h(x) = x$ for several values of observation noise variance, R , ranging from 10^{-3} to 1. We choose a time discretization with $\Delta t = 1$ between observations, which is relatively long compared to $\Delta t = 0.1$ considered in [20]. For each noise level R , we generate 10000 discrete observations using the full system (5.2), integrated as in Section 5.2.1 and observed with the discretization of (3.6). We then apply the filtering algorithm from Section 3.2, which we refer to as the nonlinear filter in Figure 6 where we show the mean estimate as the quantity of interest. For comparison we also use the standard ensemble Kalman filter (EnKF) for the reduced model (5.3) using $\sigma^2 = .126$ as

suggested in [20, 10] with ensemble size 50 without additional variance inflation.

In Figure 6 we see that the nonlinear filter tracks the transition between the potential wells more accurately compared to EnKF for large observation noise, $R = 0.5$ (see left panel). In the large noise regime, the posterior estimate is very far from Gaussian. Conversely, for very small observation noise and also for very short discretization time, the performance of nonlinear filter and the EnKF are similar, their mean-squared errors are close to the observation noise variance (see the right panel of Figure 6). In the small noise regime, the posterior density is close to Gaussian since it is very close to the observation likelihood. The evolution of the reconstructed non-Gaussian posterior density is compared to the Gaussian posterior density of the EnKF in `fig6.mov`.

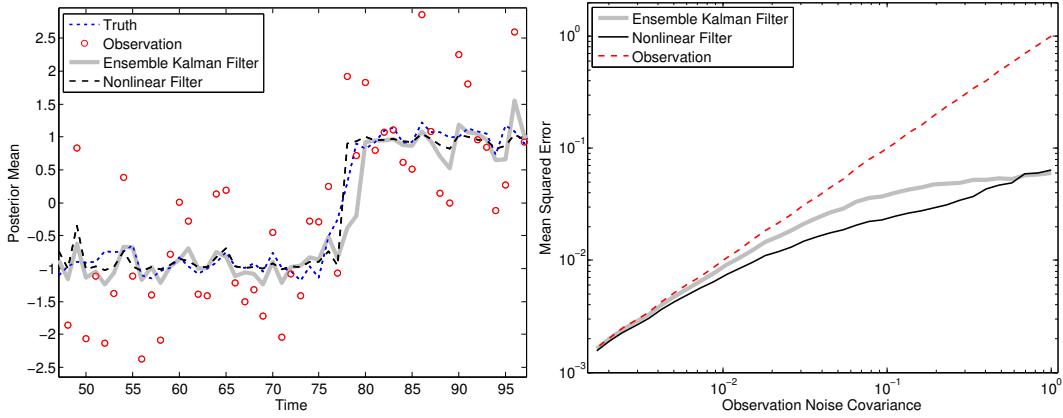


Figure 6. Left: Evolution of the posterior mean during a transition from the left potential well to the right potential well. Right: Mean squared error between filter estimates and the true state as a function of the observation noise covariance. The evolution of the posterior densities is shown in `fig6.mov`.

Next, we consider a nonlinear observation with $h(x) = |x|$, which is a pathological observation since it is impossible to determine from this observation which potential well the truth lies in. As shown in Figure 7, the nonlinear filter respects this symmetry with the posterior mean always very close to zero. In contrast, the EnKF outperforms the nonlinear filter when it happens to be on the right side of the well, however it frequently does not track the transitions leading to a higher overall mean squared error. The evolution of the posterior densities are shown in `fig7a.mov` which shows the posterior density of the nonlinear filter to be a symmetric bi-modal distribution. We also consider another nonlinear observation with $h(x) = (x - 0.05)^2$, which is nearly pathological, however, since the symmetry is not perfect, repeated observations allow the filter to determine which well the true state is in. As shown in Figure 7 and in `fig7b.mov`, the posterior of the nonlinear filter becomes increasingly bi-modal when observations occur near 0.05. However, with repeated observations which are far from 0.05 the posterior density accumulates information on the true location and the incorrect mode is progressively damped away.

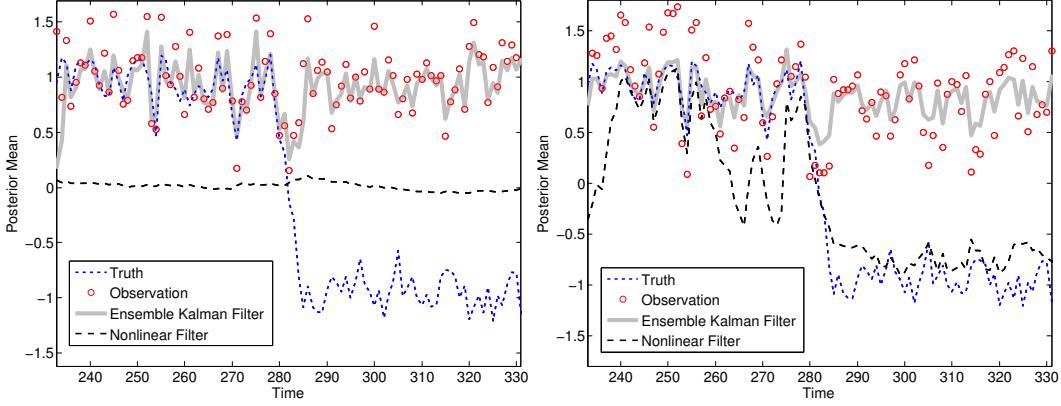


Figure 7. Left: Evolution of the posterior mean for $h(x) = |x|$. Right: Evolution of the posterior mean for $h(x) = (x - 0.05)^2$. Both filtering experiments assimilate observations at every $\Delta t = 1$ model time unit and noise variance $R = 0.05$. The evolution of the posterior densities are shown in `fig7a.mov` and `fig7b.mov` respectively.

5.2.3. Response. In this section we apply the method of Section 4 to quantify the response of the system (5.2) to perturbing the potential $U(x) = x^2/2 - x^4/4$ with

$$\delta U(x) = -\exp(-100(x - 0.5)^2)/10.$$

This introduces a third potential well centered at $x = 0.5$ as shown in Figure 8 and changes the vector field for the variable x in (5.2) which becomes,

$$(5.4) \quad \dot{x} = x - x^3 - 20(x - 0.5) \exp(-100(x - 0.5)^2) + (\gamma/\epsilon)y_2.$$

We first create a benchmark for comparison by estimating the response through a Monte-Carlo simulation. Since the system is assumed to be initially at equilibrium for the unperturbed system (5.2), we initialize an ensemble of 100000 members by simply taking the final ensemble from the long simulation in Section 5.2.1. We first use a stochastic Monte-Carlo simulation by replacing $(\gamma/\epsilon)y_2$ in (5.4) with the stochastic term σdW_t with $\sigma^2 = 0.113$ as suggest in [20]. This yields a reduced model as suggested in [20],

$$(5.5) \quad dx = (x - x^3 - 20(x - 0.5) \exp(-100(x - 0.5)^2)) dt + \sigma dW_t$$

We apply the Euler-Maruyama scheme with integration time step $\delta t = 0.002$ to evolve this ensemble using (5.5). At any fixed time t we can then use the centered moments of the ensemble to produce a Monte-Carlo estimate of the response. We also tested the reduced model using our estimate of D to define $\sigma^2 = 2D = 0.107$ and the results were similar the those shown below.

We then apply the algorithm in Section 4 to estimate the response. This technique uses the same training data as Section 5.2.1 along with the perturbation δU given above to estimate the response without any parametric model. In the first 10 time units, density migrates quickly from the potential well centered at $x = 1$ into the new potential well centered at $x = 0.5$.

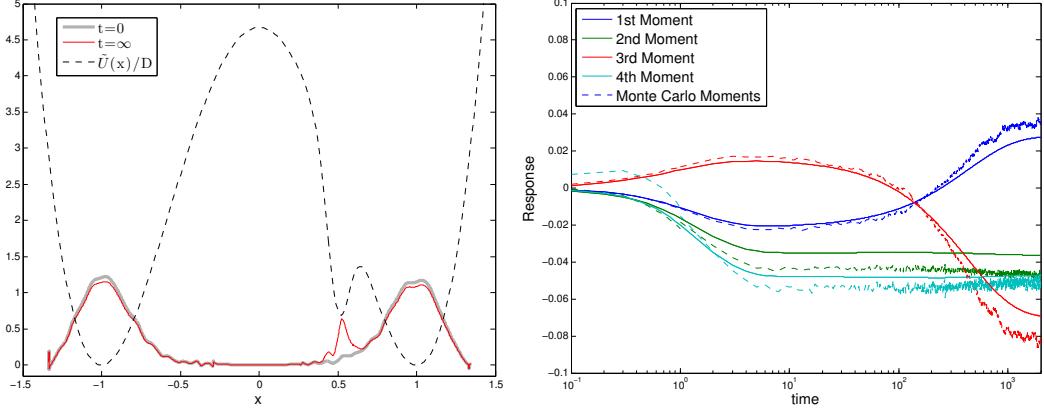


Figure 8. Left: Evolution of the density and the perturbed potential function. Right: Evolution of the moments compared to those of the stochastic Monte-Carlo simulation of the reduced model with $\sigma^2 = 0.113$.

Over the next 1000 time units, density migrates from the potential well centered at $x = -1$ into the wells at $x = 0.5$ and $x = 1$, which are now more stable due to the perturbation. In Figure 8 we summarize the evolution of the density and compare the response estimates of the nonparametric technique to the Monte-Carlo estimates of (5.5).

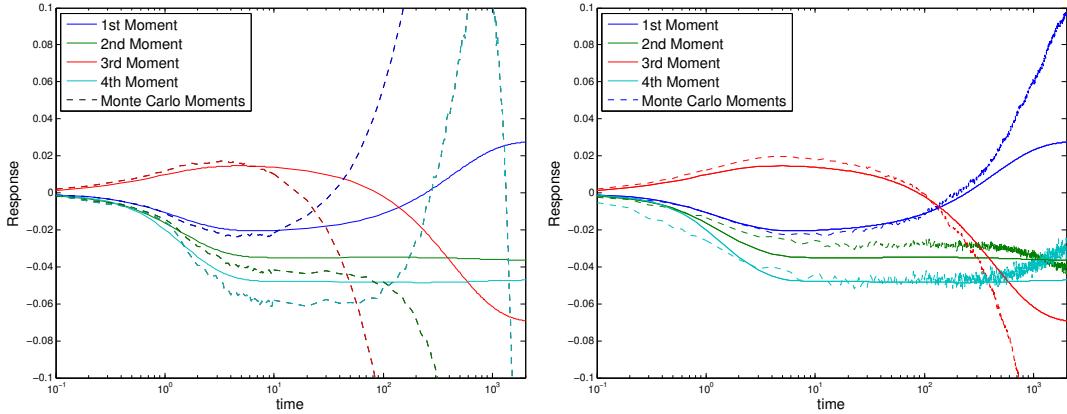


Figure 9. Left: Evolution of the moments compared to the full model with $\epsilon = \sqrt{0.1}$. Right: Evolution of the moments compared to the full model with $\epsilon = \sqrt{0.005}$. The evolution of the densities is shown in `fig9.mov` for $\epsilon = \sqrt{0.1}$.

Finally, we compare the non-parametric response estimate to a Monte-Carlo simulation of the full model (5.2) with perturbation (5.4). We initialize an ensemble of 100000 members as in the stochastic simulation, and then integrate the perturbed system using a fourth order Runge-Kutta scheme with $\delta t = 0.002$. As shown in Figure 9, for $\epsilon = \sqrt{0.1}$ this procedure only matches the evolution of the stochastically forced gradient flow for a short time. The evolution of the densities up to time $t = 10$ is shown for $\epsilon = \sqrt{0.1}$ in `fig9.mov`. By repeating the Monte-Carlo simulation with $\epsilon = \sqrt{0.005}$ and decreasing the integration time step to $\delta t = 0.0005$ we find better agreement for longer time, however the system still does not behave as a gradient

flow for very long times as shown in Figure 9.

6. Summary and Discussion. The techniques developed in this paper solve uncertainty quantification problems using only a training data set without any parametric model. The key tool is the diffusion maps algorithm developed in [6, 4], which uses the training data to build a non-parametric model for any stochastically forced gradient flow systems. The model is represented implicitly through a kernel matrix and which converges to the generator of the dynamical system in the limit of large data [4]. Using properties of gradient flow systems, we can use the eigenvectors of this kernel matrix to approximate the eigenfunctions of the corresponding Fokker-Planck operator. Subsequently, we solved three uncertainty quantification problems, forecasting, filtering, and response estimation, by projecting the associated PDE or SPDE onto this basis of eigenfunctions. In this basis the system becomes a linear ODE or SDE which we can then truncate onto a finite number of modes and solve analytically or approximate numerically.

A key advantage of this technique is that all the densities and eigenfunctions are represented by their values on the training data set, which alleviates the need for a grid. This is important because it is difficult to define a grid on an unknown manifold $\mathcal{M} \subset \mathbb{R}^n$. While our numerical examples focused on one-dimensional systems on $\mathcal{M} = \mathbb{R}$, the theory of [6, 4] allows the gradient flow system to evolve on any d -dimensional manifold of $\mathcal{M} \subset \mathbb{R}^n$. It is important to note that these techniques provably solve the full nonlinear uncertainty quantification problem, but only in the limit of infinite training data and an infinite number of eigenvectors. In practice, the amount of data and number of eigenvectors required will depend strongly on the intrinsic dimensionality of the underlying manifold. For high-dimensional systems the amount of data required will quickly make these techniques computationally infeasible even if such large data sets are available.

The techniques are also limited by the assumption of a gradient flow system with stochastic forcing which is isotropic and constant over the entire state space. Concurrent research in [5] suggests it may be possible to overcome these limitations using different kernels based on non-Euclidean norms which vary over the state space. Overcoming the restriction to gradient flow systems may also be possible using kernels which are not centered as shown in [5]. However, one limitation of the results of [5] is the need to estimate the drift and diffusion coefficients that are used to define the appropriate un-centered kernel. Finding provably convergent algorithms for estimating these coefficients can be difficult, particularly on non-compact domains as shown in [4]. Regardless of such developments, the fundamental limitation will likely remain the requirement of small intrinsic dimension of the manifold.

REFERENCES

- [1] N.U. AHMED AND S.M. RADAIDEH, *A powerful numerical technique solving zakai equation for nonlinear filtering*, Dynamics and Control, 7 (1997), pp. 293–308.
- [2] A. BAIN AND D. CRISAN, *Fundamentals of Stochastic Filtering*, Springer, New York, 2009.
- [3] M. BELKIN AND P. NIYOGI, *Laplacian eigenmaps for dimensionality reduction and data representation*, Neural Computation, 15 (2003), pp. 1373–1396.
- [4] T. BERRY AND J. HARLIM, *Variable bandwidth diffusion kernels*, Submitted to the Journal of Applied and Computational Harmonic Analysis, (2014). <http://arxiv.org/abs/1406.5064>.

- [5] T. BERRY AND T. SAUER, *Local kernels and geometric structure of data*, Submitted to the Journal of Applied and Computational Harmonic Analysis, (2014). <http://arxiv.org/abs/1407.1426>.
- [6] R. COIFMAN AND S. LAFON, *Diffusion maps*, Applied and Computational Harmonic Analysis, 21 (2006), pp. 5–30.
- [7] D. CROMMELIN AND E. VANDEN-EIJNDEN, *Reconstruction of diffusions using spectral data from time-series*, Communications in Mathematical Sciences, 4 (2006), pp. 651–668.
- [8] R. FREY, T. SCHMIDT, AND L. XU, *On galerkin approximations for the zakai equation with diffusive and point process observations*, SIAM Journal on Numerical Analysis, 51 (2013), pp. 2036–2062.
- [9] DI. GIANNAKIS AND A.J. MAJDA, *Nonlinear laplacian spectral analysis for time series with intermittency and low-frequency variability*, Proceedings of the National Academy of Sciences, 109 (2012), pp. 2222–2227.
- [10] D. GIVON, R. KUPFERMAN, AND A. STUART, *Extracting macroscopic dynamics: model problems and algorithms*, Nonlinearity, 17 (2004), p. R55.
- [11] M. HEIN AND J.-Y. AUDIBERT, *Intrinsic dimensionality estimation of submanifolds in R^d* , in Proceedings of the 22nd International Conference on Machine Learning, ICML '05, New York, NY, USA, 2005, ACM, pp. 289–296.
- [12] I.T. JOLLIFFE, *Principal component analysis*, Springer series in statistics, Springer-Verlang, 1986.
- [13] R.E. KALMAN AND R. BUCY, *New results in linear filtering and prediction theory*, Trans. AMSE J. Basic Eng., 83D (1961), pp. 95–108.
- [14] J.N. KUTZ, *Data-Driven Modeling & Scientific Computation: Methods for Complex Systems & Big Data*, OUP Oxford, 2013.
- [15] A LITTLE, Y.-M. JUNG, AND M. MAGGIONI, *Multiscale estimation of intrinsic dimensionality of data sets*, in Manifold learning and its applications : papers from the AAAI Fall Symposium, AAAI, 2009, pp. 26–33.
- [16] O.P.L. MAÎTRE AND O.M. KNIO, *Spectral Methods for Uncertainty Quantification: With Applications to Computational Fluid Dynamics*, Scientific Computation, Springer, 2010.
- [17] A.J. MAJDA, R.V. ABRAMOV, AND M.J. GROTE, *Information theory and stochastics for multiscale nonlinear systems*, CRM Monograph Series v.25, American Mathematical Society, Providence, Rhode Island, USA, 2005.
- [18] A.J. MAJDA, B. GERSHGORIN, AND Y. YUAN, *Low frequency response and fluctuation-dissipation theorems: Theory and practice*, J. Atmos. Sci., 67 (2010), pp. 1181–1201.
- [19] A.J. MAJDA AND J. HARLIM, *Filtering Complex Turbulent Systems*, Cambridge University Press, UK, 2012.
- [20] L. MITCHELL AND G.A. GOTZWALD, *Data assimilation in slow–fast systems using homogenized climate models*, Journal of the Atmospheric Sciences, 69 (2012), pp. 1359–1377.
- [21] H.N. NAJM, *Uncertainty quantification and polynomial chaos techniques in computational fluid dynamics*, Annual Review of Fluid Mechanics, 41 (2009), pp. 35–52.
- [22] E. PARZEN, *On estimation of a probability density function and mode*, The Annals of Mathematical Statistics, 33 (1962), pp. 1065–1076.
- [23] M. ROSENBLATT, *Remarks on some nonparametric estimates of a density function*, The Annals of Mathematical Statistics, 27 (1956), pp. 832–837.
- [24] S.R. SAIN AND D.W. SCOTT, *On locally adaptive density estimation*, Journal of the American Statistical Association, 91 (1996), pp. 1525–1534.
- [25] T.P. SAPSIS AND P.F.J. LERMUSIAUX, *Dynamically orthogonal field equations for continuous stochastic dynamical systems*, Physica D: Nonlinear Phenomena, 238 (2009), pp. 2347 – 2360.
- [26] T.P. SAPSIS AND A.J. MAJDA, *Statistically accurate low-order models for uncertainty quantification in turbulent dynamical systems*, Proceedings of the National Academy of Sciences, 110 (2013), pp. 13705–13710.
- [27] R.C. SMITH, *Uncertainty Quantification: Theory, Implementation, and Applications*, Computational Science and Engineering, SIAM, 2013.
- [28] D. G. TERRELL AND D. W. SCOTT, *Variable kernel density estimation*, Annals of Statistics, 20 (1992), pp. 1236–1265.

Appendix. Moments and Nonlinear Response of the Ornstein-Uhlenbeck Process.

Consider the following linear SDE,

$$(A.1) \quad dx = ((\alpha + a)x + b)dt + \sqrt{2D}dW_t,$$

with solutions given by Ornstein-Uhlenbeck processes centered at $\bar{x}_\infty \equiv -b/(\alpha + a)$. The Fokker-Plank equation for (A.1) is given by,

$$(A.2) \quad \frac{\partial}{\partial t} p^\delta = \mathcal{L}^* p^\delta = -\frac{\partial}{\partial x}(((\alpha + a)x + b)p^\delta) + D \frac{\partial^2}{\partial x^2} p^\delta,$$

and we can use this equation to find the evolution of the moments. First, the mean $\bar{x}(t) = \int x p^\delta(x, t) dt$ evolves according to,

$$\dot{\bar{x}} = \int x p_t^\delta dx = \int x \left(-\frac{\partial}{\partial x}(((\alpha + a)x + b)p^\delta) + D \frac{\partial^2}{\partial x^2} p^\delta \right) dx = (\alpha + a)\bar{x} + b.$$

Using the evolution of the mean, we can then find the evolution of the centered moments $M_n(t) = \int (x - \bar{x})^n p^\delta(x, t) dx$ using the same strategy,

$$\begin{aligned} \dot{M}_n &= \int (x - \bar{x})^n p_t^\delta - n\dot{\bar{x}}(x - \bar{x})^{n-1} p^\delta \\ &= \int n(x - \bar{x})^{n-1}((\alpha + a)x + b)p^\delta - n(x - \bar{x})^{n-1}((\alpha + a)\bar{x} + b)p^\delta + n(n-1)D(x - \bar{x})^{n-2}p^\delta \\ &= n(\alpha + a)M_n + n(n-1)DM_{n-2}. \end{aligned}$$

Noting that $M_0 = 1$ and $M_1 = 0$, we have the following solutions for the first four moments,

$$\begin{aligned} \bar{x}(t) &= (\bar{x}(0) - \bar{x}_\infty)e^{(\alpha+a)t} + \bar{x}_\infty \\ M_2(t) &= \left(M_2(0) + \frac{D}{\alpha + a} \right) e^{2(\alpha+a)t} - \frac{D}{\alpha + a} \\ M_3(t) &= M_3(0)e^{3(\alpha+a)t} \\ M_4(t) &= \left(M_4(0) + \frac{6D}{\alpha + a} \left(M_2(0) + \frac{D}{\alpha + a} \right) - \frac{3D^2}{(\alpha + a)^2} \right) e^{4(\alpha+a)t} \\ &\quad - \frac{6D}{\alpha + a} \left(M_2(0) + \frac{D}{\alpha + a} \right) e^{2(\alpha+a)t} + \frac{3D^2}{(\alpha + a)^2}. \end{aligned}$$

To find the analytic solutions for the response, $\delta\mathbb{E}[A(x)]$, we take the initial moments from the invariant distribution p_{eq} of the unperturbed system. Thus, $\bar{x}(0) = 0$, $M_2(0) = -D/\alpha$, $M_3(0) = 0$, and $M_4(0) = 3D^2/\alpha^2$ so plugging in these values and then subtracting them from the moments we have,

$$\begin{aligned} \delta\mathbb{E}[x] &= \bar{x}_\infty \left(1 - e^{(\alpha+a)t} \right) = \frac{-b}{\alpha + a} (1 - e^{(\alpha+a)t}) \\ \delta\mathbb{E}[(x - \bar{x})^2] &= \frac{aD}{\alpha(\alpha + a)} \left(1 - e^{2(\alpha+a)t} \right) \\ \delta\mathbb{E}[(x - \bar{x})^3] &= 0 \\ \delta\mathbb{E}[(x - \bar{x})^4] &= \frac{6aD^2}{\alpha(\alpha + a)^2} \left(e^{2(\alpha+a)t} - e^{4(\alpha+a)t} \right) + \left(\frac{3D^2}{\alpha^2} - \frac{3D^2}{(\alpha + a)^2} \right) \left(e^{4(\alpha+a)t} - 1 \right). \end{aligned}$$