SIAM/ASA J. UNCERTAINTY QUANTIFICATION
Vol. 8, No. 4, pp. 1414–1445

Kernel Methods for Bayesian Elliptic Inverse Problems on Manifolds*

John Harlim†, Daniel Sanz-Alonso‡, and Ruiyi Yang§

Abstract. This paper investigates the formulation and implementation of Bayesian inverse problems to learn input parameters of partial differential equations (PDEs) defined on manifolds. Specifically, we study the inverse problem of determining the diffusion coefficient of a second-order elliptic PDE on a closed manifold from noisy measurements of the solution. Inspired by manifold learning techniques, we approximate the elliptic differential operator with a kernel-based integral operator that can be discretized via Monte Carlo without reference to the Riemannian metric. The resulting computational method is mesh-free and easy to implement, and can be applied without full knowledge of the underlying manifold, provided that a point cloud of manifold samples is available. We adopt a Bayesian perspective to the inverse problem, and establish an upper bound on the total variation distance between the true posterior and an approximate posterior defined with the kernel forward map. Supporting numerical results show the effectiveness of the proposed methodology.

Key words. inverse problems, elliptic partial differential equations, manifold learning, graph Laplacian

AMS subject classifications. 65N21, 86A22, 58J05

DOI. 10.1137/19M1295222

1. Introduction.

Partial differential equations (PDEs) on manifolds are used to model a variety of physical and biological phenomena including pattern formation on biological surfaces, phase separation in biomembranes, tumor growth, and surfactants on fluid interfaces [26, 27, 15, 61]. In this paper we focus on the inversion of *elliptic* PDEs for two main reasons. First, elliptic PDEs are ubiquitous in applications and they are used, for instance, as simplified models for groundwater flow and oil reservoir simulation. The need to specify uncertain input parameters of these models leads naturally to the inverse problem of determining the permeability from the pressure under a Darcy model of flow in a porous medium [45, 43, 38, 48]. Second, elliptic models are widely used to test algorithms for forward propagation of uncertainty [29, 16, 2] and Bayesian inversion [56, 21, 32]. Despite the applied importance of elliptic inversion, the manifold setting that we consider is largely unexplored and may allow for more realistic modeling in applications. For example, the variables of interest in the groundwater flow problem may not be confined to a *flat* two-dimensional domain and knowledge of the underlying flow surface may be limited to a point cloud of landmark locations.

*Received by the editors October 24, 2019; accepted for publication (in revised form) May 11, 2020; published electronically November 24, 2020.
https://doi.org/10.1137/19M1295222
Funding: The work of the first author was partially supported by NSF Grants DMS-1619661 and DMS-1854299. The work of the second author was supported by the NSF Grants DMS-1912818 and DMS-1912802.

†Department of Mathematics, Department of Meteorology and Atmospheric Science, Institute for Computational and Data Sciences, The Pennsylvania State University, University Park, PA 16802-6400 USA (jharlim@psu.edu).
‡Department of Statistics, University of Chicago, Chicago, IL 60637 USA (sanzalonso@uchicago.edu.)
§Committee on Computational and Applied Mathematics, University of Chicago, Chicago, IL 60637 USA (yry@uchicago.edu).

The aim of this paper is to study the formulation and implementation of Bayesian inverse problems to learn input parameters of PDEs defined on manifolds. Specifically, we study the inverse problem of recovering the diffusion coefficient of a second-order, divergence-form elliptic equation, given noisy measurements of the solution. While our interest lies in solving the inverse problem, most of our efforts are devoted to studying the approximation of the *forward map* (the operator that takes the input parameter to the solution of the PDE). Several techniques to approximate the forward map have been proposed in the extensive literature on numerical methods for PDEs on manifolds. For example, finite element methods [25, 14, 11], level-set methods [7, 46], closest point methods [50], or mesh-free radial basis function methods [49]. The implementation detail of each of the existing methods is different, but a unifying theme is the need to have a representation of the manifold in order to approximate the differential operator. Unfortunately, these approaches are difficult to implement when one only has access to an unstructured point cloud of manifold samples and meshing is challenging, or when the dimension of the ambient space is large but the manifold dimension is moderate.

In this paper, we avoid the problems associated with the representation of the manifold by directly approximating the differential operator in the forward map with an appropriate kernel integral operator. With a consistent kernel approximation to the differential operator on the manifold, the numerical implementation can be performed naturally by discretizing the corresponding integral operator on a *point cloud* of manifold samples without further knowledge of the underlying manifold or its Riemannian metric. Building on this construction, we propose a fully discrete, mesh-free approach to the numerical solution of Bayesian inverse problems on point clouds. The idea of facilitating the discretization of PDEs on manifolds by an integral equation approximation can also be found in the recent papers [40, 41, 36], all of which build on manifold learning techniques and analyses. Our perspective in this paper and in [36, 6] is in contrast to the one taken in [3, 4, 18, 17, 5]. Rather than identifying the limiting continuum operator of different normalizations of graph Laplacians, our interest is to define a suitable kernel to approximate a given anisotropic diffusion operator on the underlying manifold.

We adopt a Bayesian approach to the inverse problem [39, 13, 56, 52], where we set a *prior* distribution on the unknown PDE input parameter, and condition on observed data to find its posterior distribution. The Bayesian approach is largely motivated by the following advantages. First, the posterior covariance and posterior confidence intervals may be used to quantify the uncertainty in the parameter reconstruction. Second, the Bayesian formulation leads to a well-posed inverse problem [44, 56, 52] by which a small perturbation on the data, the prior distribution, or the forward map leads to a small perturbation in the posterior solution. Our main theoretical result is an example of the well-posedness of the Bayesian formulation: we deduce a total variation error bound between the true posterior distribution and its kernel-based approximation from a new error bound between the forward map and its kernel approximation. The new forward map error bound, with a dependence on the diffusion coefficient, builds on existing results on the pointwise convergence kernel approximations to elliptic operators [17, 5].

The advantages of the Bayesian approach outlined above come with a cost: the need to specify a prior distribution on the unknown. The choice of prior is crucial as it determines the support of the posterior but, unfortunately, this choice is often only guided by ad hoc

and computational considerations. In this paper we consider a two-parameter family of log-Gaussian field priors on manifolds, defining the covariance through the Laplace–Beltrami operator on the manifold [42, 24]. The two prior parameters allow the specification of the smoothness and length-scale of prior (and hence posterior) draws, and the length-scale may be learned from data using a hierarchical approach as detailed in our numerical experiments. In addition to their flexibility, a further advantage of our choice of priors is that they allow the infusion of geometric information from the manifold on the reconstructed input by expressing it as a random combination of the first eigenfunctions of the Laplacian. Moreover, in the absence of a full representation of the manifold, these priors can be consistently discretized using a graph Laplacian [59, 33]. We refer to [8, 30, 33, 31] for recent applications and references on Gaussian processes on manifolds. From a computational viewpoint, the use of log-Gaussian priors and the existence of a continuum limit facilitate the design of Markov chain Monte Carlo (MCMC) algorithms that scale well with the size of the point cloud, as shown in a linear inverse problem in [31]. In this regard, a simple but powerful idea is to use a proposal kernel that satisfies detailed balance with respect to the prior [20].

Outline and main contributions. We close this introduction with an outline of the rest of the paper, summarizing the main contributions of each section.

- In section 2 we give a brief introduction to the Bayesian formulation of inverse problems, and formulate the problem on a manifold. The main novel contributions of this section are (i) to introduce a kernel-based approximation to the forward map and an associated approximation with the posterior in the continuum (subsection 2.2); and (ii) to employ the kernel approximations to formulate a Bayesian solution to the inverse problems on point clouds (subsection 2.3). These kernel and point cloud approximations are inspired by manifold learning and data analysis techniques.
- Section 3 contains the main theoretical contributions of this paper. Theorem 3.1 gives an error bound between the true and kernel-based forward maps, and Theorem 3.6 establishes a bound on the total variation distance between the posterior and its kernel-based approximation. The main novelty of Theorem 3.1 is to generalize the analysis in [17] to account for anisotropic diffusion and, more importantly, to explicitly track the dependence of the diffusion coefficient in the error bounds. Understanding this dependency is necessary in order to guarantee the closeness of the true and approximate posterior distributions shown in Theorem 3.6.
- In section 4 we discuss the practical implementation of the methods, provide guidelines for the choice of tuning parameters and for the posterior sampling, and conduct three numerical experiments of increasing difficulty to illustrate the applicability of our approach. We also consider in subsection 4.5 a hierarchical formulation of the inverse problem, where the prior length scale is learned from the data, when in the absence of such knowledge.
- We close in section 5 with conclusions and open directions for research that stem from our work.

Notation and setting. Throughout this paper $\mathcal{M}$ will denote an m-dimensional smooth Riemannian manifold embedded in $\mathbb{R}^d$. We will denote by $\mathcal{C}^k := \mathcal{C}^k(\mathcal{M})$ the space of k-times differentiable functions on $\mathcal{M}$ and by $\mathcal{C}^{k,\alpha} := \mathcal{C}^{k,\alpha}(\mathcal{M})$ the space of k-times differentiable functions whose kth partial derivatives are Hölder continuous with exponent α. We will

assume that the manifold $\mathcal{M}$ is compact and has no boundary, thus avoiding the technicalities necessary to deal with boundary conditions. The theoretical and computational investigation of the point cloud approximation to PDEs supplemented with boundary conditions is a topic of current research [40, 41, 36]. Due to the lack of boundary conditions, the elliptic problem that we consider is unique up to a constant. We will enforce uniqueness by working on the space $L^2_0 := L^2_0(\mathcal{M})$ of mean-zero square integrable functions on $\mathcal{M}$.

2. Bayesian inverse problems on manifolds and point clouds. We start in subsection 2.1 by recalling the Bayesian formulation of elliptic inverse problems in a given manifold. We then introduce in subsection 2.2 a kernel-based approximation to the forward map and a corresponding approximation to the posterior, both of which will be analyzed in section 3. Finally in subsection 2.3 we introduce a point cloud approximation of the kernel forward map leading to a formulation of the elliptic problem on point clouds without reference to the underlying manifold. We will investigate numerically the implementation of elliptic Bayesian inverse problems on point clouds in section 4.

2.1. Bayesian elliptic inverse problems on manifolds. We consider the elliptic equation

$$\mathcal{L}^\kappa u := -\mathrm{div}(\kappa \nabla u) = f, \qquad x \in \mathcal{M}, \tag{2.1}$$

where κ is a function on $\mathcal{M}$. Here and throughout, the differential operators are defined with respect to the Riemannian metric inherited by $\mathcal{M}$ from $\mathbb{R}^d$. We are interested in the inverse problem of determining the diffusion coefficient κ from noisy measurements of u of the form

$$y = \mathcal{D}(u) + \eta, \tag{2.2}$$

where the *observation map* $\mathcal{D} : L^2 \to \mathbb{R}^J$ will be assumed to be known. Examples of observation maps will be discussed in subsection 2.1.2. We adopt a Bayesian perspective to the inverse problem, described succinctly in what follows; we refer to [39, 56, 52] for a more detailed account. In short, the Bayesian formulation of inverse problems involves specifying a prior distribution π for the unknown PDE input κ and a distribution for the observation noise η. Once these distributions have been specified, the solution to the inverse problem is the posterior distribution of the variable κ conditioned on the observed data y. For simplicity of exposition, we will assume throughout that the observation noise is centered and Gaussian, $\eta \sim \mathcal{N}(0, \Gamma)$ for given positive definite $\Gamma \in \mathbb{R}^{J\times J}$.

Writing $\kappa = e^\theta$, we take a Gaussian prior for θ supported on a Banach space $\mathcal{B}$. A specific form of prior, widely used in applications, will be described in subsection 2.1.1. Our assumptions on κ and f in section 3 will guarantee the existence of a unique solution to (2.1) in the space L^2_0 of mean-zero square intergrable functions on $\mathcal{M}$. This, in turn, allows us to define a forward map $\mathcal{F} : \theta \in \mathcal{B} \mapsto u \in L^2_0$. Provided that the map $\mathcal{G} := \mathcal{D} \circ \mathcal{F} : \mathcal{B} \to \mathbb{R}^J$ is measurable and that the prior is supported on $\mathcal{B}$, the posterior π^y can be written as a change of measure with respect to the prior

$$\frac{d\pi^y}{d\pi}(\theta) \propto \exp\left(-\frac{1}{2}|y - \mathcal{G}(\theta)|^2_\Gamma\right), \tag{2.3}$$

with $|\cdot|^2_\Gamma := \langle \cdot, \Gamma^{-1} \cdot \rangle$. Equation (2.3) shows that the posterior distribution π^y is defined by reweighting the prior, favoring unknowns θ that produce a good match with the data

y through a likelihood function (the right-hand term), implied by (2.2) and the assumed Gaussian distribution of the noise η.

For our theoretical results in section 3 we will take $\mathcal{B} = \mathcal{C}^4$ and assume that $f \in \mathcal{C}^{3,\alpha}$ for $0 < \alpha < 1$. These assumptions guarantee [37] that almost surely with respect to the prior, the diffusion coefficient κ is uniformly elliptic, and the unique solution of (2.1) in L^2_0 lies in $\mathcal{C}^{5,\alpha}$, allowing us to establish a stability result for an approximation of the forward map. We believe, however, that these strong regularity conditions can be relaxed.

2.1.1. Matérn-type prior. Here we describe a choice of prior that is widely used in applications in the geophysical sciences and spatial statistics [55]; in subsection 2.3.1 we will introduce a point cloud approximation to this prior used in our numerical experiments in section 4. Since (2.3) implies that the support of the prior determines the support of the posterior, it should both capture the geometry of the manifold $\mathcal{M}$ and have enough expressivity to include a wide class of functions. This motivates us to choose the prior from a flexible two-parameter family of Gaussian measures on L^2. Precisely, we will consider priors of the form

$$\pi = \mathcal{N}(0, C_{\tau,s}), \quad C_{\tau,s} = c(\tau)(\tau I + \Delta_{\mathcal{M}})^{-s}, \tag{2.4}$$

where $\Delta_{\mathcal{M}} := -\mathrm{div}(\nabla\cdot)$ is the Laplace–Beltrami operator on $\mathcal{M}$, $\tau > 0, s > \frac{m}{2}$ are two free parameters, whose intuitive interpretation will be given below, and $c(\tau)$ is a normalizing constant. Let $\{(\lambda_i, \varphi_i)\}_{i=1}^{\infty}$ be eigenvalue-eigenvector pairs for $\Delta_{\mathcal{M}}$ with λ_i's increasing. Then by the Karhunen–Loéve expansion, random samples of π admit a series expansion

$$v = c(\tau)^{1/2} \sum_{i=1}^{\infty} (\tau + \lambda_i)^{-s/2} \xi_i \varphi_i, \tag{2.5}$$

where $\xi_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$ (i.i.d. is independent and identically distributed). The eigenfunctions of the Laplacian contain geometric information on the underlying manifold and, therefore, constitute a natural basis for functions on the manifold. By Weyl's law, $\lambda_i \asymp i^{2/m}$ and so the requirement $s > \frac{m}{2}$ is to ensure that samples from π belong to L^2 almost surely. Moreover, the parameter s controls the rate of decay of the coefficients and hence characterizes the regularity of the samples. The role of τ is more delicate. If we write the coefficients as $v_i := (\tau+\lambda_i)^{-s/2} = \tau^{-s/2}(1+\frac{\lambda_i}{\tau})^{-s/2}$, then we can see that the v_i's will be small for λ_i's that are much larger than τ. In particular, the only significant v_i's are those where the corresponding λ_i's are on the same order of τ and hence τ determines the significant basis functions in the expansion (2.5). Since the eigenfunctions $\{\varphi_i\}_{i=1}^{\infty}$ represent increasing frequencies, τ can be interpreted as a length-scale parameter. It can be seen from (2.5) that τ also affects the amplitude of the samples and this motivates us to choose the normalizing constant so that v has a fixed variance, which we set to be 1:

$$c(\tau) = \frac{1}{\sum_{i=1}^{\infty}(\tau + \lambda_i)^{-s}}. \tag{2.6}$$

Such priors are widely used when $\mathcal{M}$ is a domain in a Euclidean space and are related to the Whittle–Matérn distributions [24]. In [42] the authors also considered their extension

to manifolds. It can be shown that for s large enough, samples from π belong to $\mathcal{C}^k$ almost surely. And since the embedding of $\mathcal{C}^k$ into L^2 is continuous, the restriction of π to $\mathcal{C}^k$ is again a Gaussian measure [10]. Hence for our purpose, we choose s so that π is a Gaussian measure on $\mathcal{C}^4$. In particular we will need later in section 3 the result from Fernique's theorem [28] that there exists $\alpha > 0$ such that

$$\int_{\mathcal{B}} \exp\left(\alpha \|\theta\|_{\mathcal{C}^4}^2\right) d\pi(\theta) < \infty.$$

Remark 2.1. Choosing a prior with parameter τ that is far from the true length-scale of the unknown parameter would lead to poor Bayesian inversion. This can be problematic if such prior knowledge is not available, but may be at least partially alleviated by considering a hierarchical formulation specifying a joint prior on both τ and θ, so that the length-scale is learned from data simultaneously with the unknown θ; implementation details of the hierarchical formulation will be given in subsection 4.5.

2.1.2. Observation maps. Here we give two examples of observation maps that we shall consider. For theoretical considerations, we assume that the observation map is of the form $\mathcal{D}(u) = (\ell_1(u), \dots, \ell_J(u))^T$, where each ℓ_j is a bounded linear functional on L^2. A widely used example is the smoothed observation at a point x_j: $\ell_j(u) = \int K(x_j, x)u(x)dV(x)$, where K is a kernel such as the Gaussian kernel [9]; this type of observations arise in practice when the data are gathered from a collection of spatially distributed sensors located in the vicinity of landmark points x_j. Equally common is the pointwise evaluation [32, 23]: $\ell_j = u(x_j)$. Notice that pointwise evaluation is not a bounded linear functional on L^2 but can be approximated by smoothed observation arbitrarily well for continuous u's. We remark that the boundedness assumption of ℓ_j is only a technical one and for the numerical experiments in section 4 we will consider only pointwise evaluations.

2.2. Kernel approximation of the forward and inverse problem. In this subsection we introduce a kernel approximation $\mathcal{L}_\varepsilon^\kappa$ to the operator $\mathcal{L}^\kappa$. Instead of directly discretizing the differential operators on $\mathcal{M}$, the new kernel operator is defined by an integral that can be discretized by Monte Carlo integration as described in the next subsection. Our kernel approximation is inspired by the following construction and result found in [17].

Let

$$G_\varepsilon u(x) := \varepsilon^{-\frac{m}{2}} \int_{\mathcal{M}} h\left(\frac{|x-\tilde{x}|^2}{\varepsilon}\right) u(\tilde{x}) dV(\tilde{x}), \quad h(z) := \frac{1}{\sqrt{4\pi}} e^{-\frac{z}{4}},$$

where dV denotes the volume form inherited by $\mathcal{M}$ from the ambient space $\mathbb{R}^d$. Then Lemma 8 in [17] shows that, for u sufficiently smooth,

$$G_\varepsilon u(x) = u(x) + \varepsilon\big(\omega u(x) - \Delta_{\mathcal{M}} u(x)\big) + O(\varepsilon^2), \qquad x \in \mathcal{M}. \tag{2.7}$$

Here, $\Delta_{\mathcal{M}} := -\mathrm{div}(\nabla \cdot)$, and ω is a function that depends only on the embedding of $\mathcal{M}$. Now, note that by direct calculation

$$\mathcal{L}^\kappa u := -\mathrm{div}(\kappa \nabla u) = \sqrt{\kappa}\left[\Delta_{\mathcal{M}}(u\sqrt{\kappa}) - u\Delta_{\mathcal{M}}\sqrt{\kappa}\right], \tag{2.8}$$

and that the expansion (2.7) for $\sqrt{\kappa}$ and $u\sqrt{\kappa}$ yields

$$uG_\varepsilon\sqrt{\kappa} = u\sqrt{\kappa} + \varepsilon\big(\omega u\sqrt{\kappa} - u\Delta_{\mathcal{M}}\sqrt{\kappa}\big) + O(\varepsilon^2),$$
$$G_\varepsilon(u\sqrt{\kappa}) = u\sqrt{\kappa} + \varepsilon\big(\omega u\sqrt{\kappa} - \Delta_{\mathcal{M}}(u\sqrt{\kappa})\big) + O(\varepsilon^2).$$

Subtracting the second equation from the first equation above and using (2.8) gives that

$$uG_\varepsilon\sqrt{\kappa} - G_\varepsilon(u\sqrt{\kappa}) = \varepsilon\left[\Delta_{\mathcal{M}}(u\sqrt{\kappa}) - u\Delta_{\mathcal{M}}\sqrt{\kappa}\right] + O(\varepsilon^2) = \frac{\varepsilon}{\sqrt{\kappa}}\mathcal{L}^\kappa u + O(\varepsilon^2).$$

This equation motivates the following definition of the integral operator $\mathcal{L}^\kappa_\varepsilon$:

$$\begin{aligned}\mathcal{L}^\kappa_\varepsilon u(x) &:= \frac{\sqrt{\kappa(x)}}{\varepsilon}\left[u(x)G_\varepsilon\sqrt{\kappa}(x) - G_\varepsilon(u(x)\sqrt{\kappa}(x))\right] \\ &= \frac{1}{\sqrt{4\pi}\varepsilon^{\frac{m}{2}+1}}\int_{\mathcal{M}} \exp\left(-\frac{|x-\tilde{x}|^2}{4\varepsilon}\right)\sqrt{\kappa(x)\kappa(\tilde{x})}[u(x)-u(\tilde{x})]dV(\tilde{x}),\end{aligned}$$

which satisfies

$$\mathcal{L}^\kappa_\varepsilon u(x) = \mathcal{L}^\kappa u(x) + \mathcal{O}(\varepsilon), \qquad x \in \mathcal{M}.$$

We will make rigorous this formal derivation in section 3.

We next consider the following analogue to (2.1), defined by replacing the differential operator $\mathcal{L}^\kappa$ with the kernel approximation $\mathcal{L}^\kappa_\varepsilon$:

$$\mathcal{L}^\kappa_\epsilon u_\epsilon = f, \qquad x \in \mathcal{M}. \tag{2.9}$$

Lemma 3.2 below guarantees the existence of a unique weak solution $u_\varepsilon \in L^2_0$ to (2.9) provided that $f \in L^2$ and that the original PDE is uniformly elliptic. In other words, the solution u_ε satisfies

$$\int_{\mathcal{M}} \mathcal{L}^\kappa_\varepsilon u_\varepsilon v = \int_{\mathcal{M}} fv \quad \forall v \in L^2_0. \tag{2.10}$$

We define $\mathcal{F}_\epsilon$ as the map that associates $\theta = \log(\kappa)$ with the solution u_ε to (2.9). Denoting $\mathcal{G}_\epsilon = \mathcal{D} \circ \mathcal{F}_\epsilon$, the approximate posterior π^y_ϵ has the following form:

$$\frac{d\pi^y_\epsilon}{d\pi}(\theta) \propto \exp\left(-\frac{1}{2}|y - \mathcal{G}_\epsilon(\theta)|^2_\Gamma\right). \tag{2.11}$$

In section 3 we will establish a bound on the total variation distance between the posterior distribution π^y defined in (2.3) and its approximation π^y_ε. We note, however, that the sample-based discretization of the kernel operator $\mathcal{L}^\kappa_\varepsilon$—that we will introduce in the next subsection—will involve another layer of approximation not accounted for by the theory in section 3, but necessary in practice.

Remark 2.2. As will be seen in section 3, a weak solution to (2.9) is sufficient for all the results to hold. We remark that one can show, using a Fredholm alternative, the existence of a unique mean-zero strong solution with the additional condition that f has mean zero.

2.3. Kernel-based elliptic inverse problem on a point cloud. In this subsection we assume that we are given a point cloud $X = \{x_1, \dots, x_n\}$, sampled independently according to an unknown density q on $\mathcal{M}$, but that $\mathcal{M}$ is otherwise unknown. In applications, x_i may represent landmarks on the underlying manifold, that may correspond to sensor locations. We consider the inverse problem of determining the value of the unknown input parameter κ at the points $x_i \in \mathcal{M}$ given the observed data y. Again we will follow a Bayesian perspective, defining a suitable prior π_n over functions on the point cloud, as well as a sample-based approximation to the composition map $\mathcal{G}_\varepsilon = \mathcal{D} \circ \mathcal{F}_\varepsilon$. We discuss the priors in subsection 2.3.1 and the approximation to $\mathcal{G}_\varepsilon$ in subsection 2.3.2.

2.3.1. Prior on point cloud functions. We now present the choice of priors that we will use for our numerical experiments in section 4. These will be defined in analogy to (2.4), replacing $\Delta_{\mathcal{M}}$ by a graph Laplacian. More explicitly, given n points $x_1, \dots, x_n$, we set the prior to be

$$\pi_n = \mathcal{N}(0, C^n_{\tau,s}), \quad C^n_{\tau,s} = c_n(\tau)(\tau I + \Delta_n)^{-s}, \tag{2.12}$$

where $\Delta_n \in \mathbb{R}^{n\times n}$ is a graph Laplacian constructed with $x_1, \dots, x_n$ and $c_n(\tau)$ is a normalizing constant. We refer to [59] for a detailed account of graph Laplacians and to [53] for extensive theoretical and computational motivation for our choice of priors. Note that draws from π_n are functions defined intrinsically in the point cloud $\mathcal{M}_n$ rather than on the (unknown) manifold $\mathcal{M}$. The two paremeters τ and s play the same role as discussed above in (2.5). Again samples from π_n can be expressed by Karhunen–Loéve expansion,

$$v_n = c_n(\tau)^{1/2} \sum_{i=1}^{n} (\tau + \lambda_i^{(n)})^{-s} \xi_i \varphi_i^{(n)},$$

where $\{(\lambda_i^{(n)}, \varphi_i^{(n)})\}_{i=1}^n$ are the eigenvalue-eigenvector pairs for Δ_n and $\xi_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$. Similarly as in (2.6), we normalize the draws so that the variance per node is 1:

$$c_n(\tau) = \frac{n}{\sum_{i=1}^n (\tau + \lambda_i^{(n)})^{-s}}.$$

For practical considerations, we advocate to set Δ_n as the self-tuning graph Laplacian proposed in [62]. To illustrate the idea, let $X = \{x_1, \dots, x_n\}$ be the given point cloud. Then the symmetric graph Laplacian is constructed as the matrix

$$\Delta_n = I - A^{-1/2} S A^{-1/2}, \tag{2.13}$$

where $S \in \mathbb{R}^{n\times n}$ is a similarity matrix and A is a diagonal matrix with entries $A_{ii} = \sum_{j=1}^n S_{ij}$. We set the entries of the similarity matrix S to be

$$S_{ij} = \exp\left(-\frac{|x_i - x_j|^2}{2d(i)d(j)}\right),$$

where $d(i)$ is the distance from x_i to its kth nearest neighbor, and k is a tunable parameter. The idea is similar to the standard Gaussian similarities except that the local bandwidth parameter is allowed to change adaptively based on the density of the points x_i's. Moreover, the bandwidth parameter is specified through k, a positive integer, which can be easily tuned empirically.

2.3.2. Posterior on point cloud functions. In this subsection we discuss how to discretize the posterior by constructing a point cloud approximation of $\mathcal{G}_\varepsilon$. We first approximate $\mathcal{L}_\varepsilon^\kappa$ by discretizing the integral

$$\mathcal{I}u(x) := \int_{\mathcal{M}} \exp\left(-\frac{|x-\tilde{x}|^2}{4\epsilon}\right)\sqrt{\kappa(\tilde{x})}u(\tilde{x})dV(\tilde{x})$$

by a Monte Carlo sum with a reweighting which employs a kernel density estimation. Precisely, we have

$$\mathcal{I}u(x_i) \approx \frac{1}{n}\sum_{j=1}^{n} \exp\left(-\frac{|x_i-x_j|^2}{4\varepsilon}\right)\sqrt{\kappa(x_j)}u(x_j)q_\varepsilon(x_j)^{-1}, \tag{2.14}$$

where the approximate density, applying (2.7) that $G_\varepsilon q \approx q$ up to an error of order ε, is given by

$$q_\varepsilon(x_j) = \frac{1}{\sqrt{4\pi}n\varepsilon^{\frac{m}{2}}}\sum_{k=1}^{n} \exp\left(-\frac{|x_j-x_k|^2}{4\epsilon}\right).$$

The approximation in (2.14) can be interpreted as combining a kernel density estimation [60] with importance sampling [1, 51]. In section 4 we will use this observation, where the point clouds come from uniform grids. Then $\mathcal{L}_\varepsilon^\kappa u$ evaluated at the point cloud is approximated by

$$\mathcal{L}_\varepsilon^\kappa u(x_i) \approx \frac{1}{\sqrt{4\pi}n\varepsilon^{\frac{m}{2}+1}}\sum_{j=1}^{n} \exp\left(-\frac{|x_i-x_j|^2}{4\epsilon}\right)\sqrt{\kappa(x_i)\kappa(x_j)}q_\varepsilon(x_j)^{-1}[u(x_i)-u(x_j)] := L_{\varepsilon,n}^\kappa u(x_i). \tag{2.15}$$

More concisely, we can write $L_{\varepsilon,n}^\kappa$ in matrix form in a series of steps. Define P to be the kernel matrix with entries $P_{ij} = \exp\left(-|x_i-x_j|^2/4\varepsilon\right)$. Let Q be a vector with entries $Q_i = \sum_{j=1}^n P_{ij}$ and define W to be the matrix with entries $W_{ij} = P_{ij}\sqrt{\kappa(x_i)\kappa(x_j)}Q_j^{-1}$. Then we have

$$L_{\varepsilon,n}^\kappa = \frac{1}{\varepsilon}(D-W), \tag{2.16}$$

where D is a diagonal matrix with entry $D_{ii} = \sum_{j=1}^n W_{ij}$. Notice that the above construction resembles that of the unnormalized graph Laplacian. Indeed, if $\kappa \equiv 1$, then (2.16) is exactly the unnormalized graph Laplacian up to a factor of the density [17].

Given the above discretization, we consider the following analogue to (2.9), by replacing $\mathcal{L}_\varepsilon^\kappa$ with $L_{\varepsilon,n}^\kappa$ and restricting f to the point cloud:

$$L_{\varepsilon,n}^\kappa u_n = f_n, \tag{2.17}$$

where f_n is the n-dimensional vector with entries $f(x_i)$, or an approximation thereof when f is not smooth. One can see from (2.16) that $L_{\varepsilon,n}^\kappa$ is self-adjoint and positive semidefinite under the weighted inner product $\langle u, v\rangle_q := \frac{1}{n}\sum_{j=1}^n u(x_i)v(x_i)q_\varepsilon(x_i)^{-1}$. Hence $L_{\varepsilon,n}^\kappa$ admits a

nonnegative spectrum $\{\lambda_i\}_{i=1}^n$ with $\lambda_1 = 0$ and an orthonormal basis of eigenfunctions $\{v_i\}_{i=1}^n$ wih respect to $\langle \cdot, \cdot \rangle_q$, with $v_1 \equiv 1$. We then set the solution to be

$$u_n = \sum_{i=2}^{n} \frac{f_n^i}{\lambda_i} v_i, \tag{2.18}$$

where the $f_n = \sum_{i=1}^n f_n^i v_i$. Notice that the mean-zero condition of u translates into $\langle u, 1 \rangle_q = 0$, taking into account the density. By the orthogonality of the v_i's, we see that the solution u_n in (2.18) satisfies $\langle u_n, 1 \rangle_q = 0$ and moreover, $\{v_2, \dots, v_n\}$ forms a basis for $\ell_0^2 = \{v : \langle v, 1 \rangle_q = 0\}$, which is the discrete analogue of L_0^2. One can also check that u_n satisifies $\langle L_{\varepsilon,n}^{\kappa} u_n, v \rangle_q = \langle f, v \rangle_q$ for all $v \in \ell_0^2$, is consistent with (2.10). We remark that if in addition $\langle f, 1 \rangle_q = 0$, then u_n given by (2.18) is a strong solution of (2.17), in analogy to Remark 2.2.

Hence we can now define the discrete forward map $F_{\epsilon,n} : \mathbb{R}^n \mapsto \mathbb{R}^n$ as the map that associates $\theta_n = \log(\kappa_n) := \big(\log(\kappa(x_1)), \dots, \log(\kappa(x_n))\big)$ to the solution u_n. Approximating the pointwise observation map is straightforward. We may also approximate the smoothed observation map introduced in subsection 2.1.2 by Monte Carlo as follows:

$$\ell_j^{(n)}(u_n) = \frac{1}{n} \sum_{k=1}^{n} K(x_j, x_k) u_n(x_k) q_\varepsilon(x_k)^{-1}.$$

In either case, denoting $D_n(u_n) = \big(\ell_1^{(n)}(u_n), \dots, \ell_J^{(n)}(u_n)\big)^T$ and $G_{\varepsilon,n} = D_n \circ F_{\varepsilon,n}$, the graph posterior has the following form

$$\frac{d\pi_{\varepsilon,n}^y}{d\pi_n}(\theta_n) \propto \exp\left(-\frac{1}{2}|y - G_{\varepsilon,n}(\theta_n)|_\Gamma^2\right). \tag{2.19}$$

A full analysis of the convergence of the sample-based posteriors $\pi_{\varepsilon,n}^y$ to the ground-truth posterior π^y is beyond the scope of this paper. For a *linear* regression problem, the convergence of such graph-based posteriors has been established in [31] and [33] using spectral graph theory and variational techniques.

3. Analysis of kernel approximation to the forward and inverse problem.

In this section we study the error incurred by replacing the differential operator in the forward map by its kernel approximation, and the effect of such an error in the posterior solution to the Bayesian inverse problem. The approximation of the forward map is analyzed in subsection 3.1 and the approximation of the posterior in subsection 3.2.

3.1. Forward map approximation.

The main result of this subsection is the following theorem which bounds the difference between the solution to the PDE (2.1) and the solution to the kernel-based equation (2.9).

Theorem 3.1 (forward map approximation). *Suppose that $f \in \mathcal{C}^{3,\alpha}$ and $\kappa \in \mathcal{C}^4$, with κ bounded below by $\kappa_{\min} > 0$. Let u solve $\mathcal{L}^\kappa u = f$ and u_ε solve $\mathcal{L}_\varepsilon^\kappa u_\varepsilon = f$ weakly in L_0^2. Then for $\frac{1}{4} < \beta < \frac{1}{2}$ and ε small enough depending on β,*

$$\|u - u_\varepsilon\|_{L^2} \le C A(\kappa) \|f\|_{H^3} \varepsilon^{4\beta - 1},$$

where C is a constant depending only on $\mathcal{M}$ and

$$A(\kappa) = \sqrt{\kappa_{\min}^{-5} + \kappa_{\min}^{-6}\left(\|\kappa\|_{\mathcal{C}^3}^2 + \|\kappa\|_{\mathcal{C}^3}\right) + \kappa_{\min}^{-7}\left(\|\kappa\|_{\mathcal{C}^3}^2 + \|\kappa\|_{\mathcal{C}^3}\right)^2 + \kappa_{\min}^{-8}\left(\|\kappa\|_{\mathcal{C}^3}^2 + \|\kappa\|_{\mathcal{C}^3}\right)^3}\|\sqrt{\kappa}\|_{\mathcal{C}^4}^2.$$

The novelty is to generalize previous analysis [17, 36] to the case of anisotropic diffusions and, more importantly, to keep track of the dependence $A(\kappa)$ of the error bound on the diffusion coefficient κ. As we will show in subsection 3.2, understanding this dependence is a crucial ingredient in establishing an approximation result for the inverse problem.

The proof of Theorem 3.1 follows the classical numerical analysis argument of combining stability and consistency, coupled with an H^4 norm estimate for solutions to PDE (2.1). Lemma 3.2 below establishes the stability of solutions to the kernel-based equation (2.9), Lemma 3.3 shows consistency, and Lemma 3.5 shows an H^4 norm bound on solutions to (2.1). The proof of Theorem 3.1 will be given at the end of this subsection by combining these three lemmas. To streamline the presentation we postpone the proofs of the lemmas to an appendix.

Lemma 3.2 (stability). *The equation $\mathcal{L}_\varepsilon^\kappa u_\varepsilon = f$ with $f \in L^2$ and $\kappa \in L^2$ satisfying $\kappa(x) \geq \kappa_{\min}$ for a.e. $x \in \mathcal{M}$ has a unique weak solution $u_\varepsilon \in L_0^2$. Moreover, there is $C > 0$ independent of ε and κ such that*

$$\|u_\varepsilon\|_{L^2} \leq C\kappa_{\min}^{-1}\|f\|_{L^2}. \tag{3.1}$$

The next lemma makes rigorous the argument in subsection 2.2 and characterizes the error between $\mathcal{L}^\kappa$ and $\mathcal{L}_\varepsilon^\kappa$ by accounting for its dependence on κ.

Lemma 3.3 (consistency). *Let $u \in \mathcal{C}^4$ and $\kappa \in \mathcal{C}^4$. Then, for $\frac{1}{4} < \beta < \frac{1}{2}$ and ε sufficiently small depending on β, we have*

$$\|(\mathcal{L}_\varepsilon^\kappa - \mathcal{L}^\kappa)u\|_{L^2} \leq C\|u\|_{H^4}\|\sqrt{\kappa}\|_{\mathcal{C}^4}^2\varepsilon^{4\beta-1}.$$

Remark 3.4. In the proof of Lemma 3.3, found in the appendix, we cannot set $\beta = \frac{1}{2}$. However we can choose β arbitrarily close to $\frac{1}{2}$ so that the rate is essentially $O(\varepsilon)$. We remark that the proof of Lemma 3.3 suggests that the $\mathcal{C}^3$ assumption in [17] may not be sufficient.

The last lemma bounds the H^4 norm of the solution to (2.1) in terms of the diffusion coefficient κ.

Lemma 3.5 (H^4 norm bound). *Suppose that $\kappa \in C^4$ and $f \in C^{3,\alpha}$ with $0 < \alpha < 1$. Let $u \in \mathcal{C}^5$ be the zero-mean solution to the equation $\mathcal{L}^\kappa u = f$. Then*

$$\|u\|_{H^4}^2 \leq C\|f\|_{H^3}^2\left[\kappa_{\min}^{-5} + \kappa_{\min}^{-6}\left(\|\kappa\|_{\mathcal{C}^3}^2 + \|\kappa\|_{\mathcal{C}^3}\right) + \kappa_{\min}^{-7}\left(\|\kappa\|_{\mathcal{C}^3}^2 + \|\kappa\|_{\mathcal{C}^3}\right)^2 + \kappa_{\min}^{-8}\left(\|\kappa\|_{\mathcal{C}^3}^2 + \|\kappa\|_{\mathcal{C}^3}\right)^3\right],$$

where C is a constant that depends only on $\mathcal{M}$.

We are now ready to prove Theorem 3.1.

Proof of Theorem 3.1. Recall that we need to show that

$$\|u - u_\varepsilon\|_{L^2} \leq CA(\kappa)\|f\|_{H^3}\varepsilon^{4\beta-1},$$

where u solves $\mathcal{L}^\kappa u = f$ and u_ε solves $\mathcal{L}^\kappa u_\varepsilon = f$ weakly, and $A(\kappa)$ is defined in the statement of Theorem 3.1. Notice that in the weak sense

$$\mathcal{L}^\kappa_\varepsilon(u - u_\varepsilon) = \mathcal{L}^\kappa_\varepsilon u - f = \mathcal{L}^\kappa_\varepsilon u - \mathcal{L}^\kappa u.$$

Hence using Lemma 3.2 for the first inequality, and Lemma 3.3 for the second one noting that $f \in \mathcal{C}^{3,\alpha}$ implies that $u \in \mathcal{C}^5$ [37], we have

$$\|u - u_\varepsilon\|_{L^2} \le C\kappa_{\min}^{-1}\|(\mathcal{L}^\kappa_\varepsilon - \mathcal{L}^\kappa)u\|_{L^2} \le C\kappa_{\min}^{-1}\|u\|_{H^4}\|\sqrt{\kappa}\|^2_{\mathcal{C}^4}\varepsilon^{4\beta-1}.$$

The result follows by combining this inequality with the bound on $\|u\|_{H^4}$ derived in Lemma 3.5. ∎

3.2. Posterior approximation. In this subsection we characterize the total variation distance between the two posteriors:

$$\frac{d\pi^y}{d\pi}(\theta) = \frac{1}{Z}\exp\left(-\frac{1}{2}|y - \mathcal{G}(\theta)|^2_\Gamma\right), \qquad Z := \int \exp\left(-\frac{1}{2}|y - \mathcal{G}(\theta)|^2_\Gamma\right) d\pi(\theta),$$
$$\frac{d\pi^y_\varepsilon}{d\pi}(\theta) = \frac{1}{Z_\varepsilon}\exp\left(-\frac{1}{2}|y - \mathcal{G}_\varepsilon(\theta)|^2_\Gamma\right), \qquad Z_\varepsilon := \int \exp\left(-\frac{1}{2}|y - \mathcal{G}_\varepsilon(\theta)|^2_\Gamma\right) d\pi(\theta),$$

where Z and Z_ε are normalizing constants and recall that $\mathcal{G}(\theta) = (\ell_1(u), \dots, \ell_J(u))^T$ and $\mathcal{G}_\varepsilon(\theta) = (\ell_1(u_\varepsilon), \dots, \ell_J(u_\varepsilon))^T$, where ℓ_j's are bounded linear functionals on L^2.

The main result is Theorem 3.6 below. Its proof relies on Theorem 3.1 and a standard argument [56, 32, 52] for the analysis of approximations of Bayesian inverse problems. In particular, the proof makes use of the integrability of the function $A(\kappa)$ defined in Theorem 3.1 with respect to the prior π, guaranteed by Fernique's theorem [28].

Theorem 3.6 (posterior approximation). *Let π be a Gaussian measure on $\mathcal{C}^4$, and suppose that $f \in C^{3,\alpha}$ for $0 < \alpha < 1$. Then for any $\frac{1}{4} < \beta < \frac{1}{2}$ and ε sufficiently small depending on β,*

$$d_{\mathrm{TV}}(\pi^y, \pi^y_\varepsilon) \le C\varepsilon^{4\beta-1},$$

where C is constant depending only on $\mathcal{M}$.

The proof of Theorem 3.6 relies on the following lemma, whose proof can be found in the appendix by making use of the integrability of $A(\kappa)$ with respect to the prior.

Lemma 3.7. *For $\frac{1}{4} < \beta < \frac{1}{2}$ and ϵ small enough depending on β, we have*

$$\int \left|\exp\left(-\frac{1}{2}|y - \mathcal{G}_\varepsilon(\theta)|^2_\Gamma\right) - \exp\left(-\frac{1}{2}|y - \mathcal{G}(\theta)|^2_\Gamma\right)\right| d\pi(\theta) \le C\varepsilon^{4\beta-1},$$

where C is independent of ε.

Proof of Theorem 3.6. We have

$$\begin{aligned} d_{\mathrm{TV}}(\pi^y, \pi^y_\varepsilon) &= \int \left| \frac{1}{Z_\varepsilon} \exp\left(-\frac{1}{2}|y - \mathcal{G}_\varepsilon(\theta)|^2_\Gamma\right) - \frac{1}{Z} \exp\left(-\frac{1}{2}|y - \mathcal{G}(\theta)|^2_\Gamma\right) \right| d\pi(\theta) \\ &\leq \int \left| \frac{1}{Z_\varepsilon} - \frac{1}{Z} \right| \exp\left(-\frac{1}{2}|y - \mathcal{G}_\varepsilon(\theta)|^2_\Gamma\right) d\pi(\theta) \\ &\quad + \int \frac{1}{Z} \left| \exp\left(-\frac{1}{2}|y - \mathcal{G}_\varepsilon(\theta)|^2_\Gamma\right) - \exp\left(-\frac{1}{2}|y - \mathcal{G}(\theta)|^2_\Gamma\right) \right| d\pi(\theta) \\ &= \frac{|Z - Z_\varepsilon|}{Z} + \int \frac{1}{Z} \left| \exp\left(-\frac{1}{2}|y - \mathcal{G}_\varepsilon(\theta)|^2_\Gamma\right) - \exp\left(-\frac{1}{2}|y - \mathcal{G}(\theta)|^2_\Gamma\right) \right| d\pi(\theta) \\ &\leq \frac{2}{Z} \int \left| \exp\left(-\frac{1}{2}|y - \mathcal{G}_\varepsilon(\theta)|^2_\Gamma\right) - \exp\left(-\frac{1}{2}|y - \mathcal{G}(\theta)|^2_\Gamma\right) \right| d\pi(\theta). \end{aligned}$$

Then using Lemma 3.7 it follows that

$$d_{\mathrm{TV}}(\pi^y, \pi^y_\varepsilon) \leq C\varepsilon^{4\beta-1},$$

where C is independent of ε. ∎

Remark 3.8. Similarly as for Lemma 3.3 our proof fails when $\beta = \frac{1}{2}$. However one can choose β arbitrarily close to $\frac{1}{2}$ so that the rate in 3.6 is essentially $O(\varepsilon)$.

4. Numerical experiments. In this section we investigate numerically the point cloud formulation of the inverse problem introduced in subsection 2.3. We start in subsection 4.1 by considering some aspects of the implementation. Then in subsections 4.2, 4.3, and 4.4 we give three numerical examples, where the underlying manifold is chosen to be an ellipse, the torus, and a cow-shaped manifold. In subsection 4.5, we study a hierarchical approach where the prior length-scale parameter is learned from data.

4.1. Implementation. When it comes to practical applications, care must be taken when one chooses the parameters. Central in our kernel method is the parameter ε. While Theorem 3.1 characterizes the error in approximating $\mathcal{L}^\kappa$ with $\mathcal{L}^\kappa_\varepsilon$ and suggests the consistency of the estimator as ε goes to 0, in practice one cannot take ϵ too small as we explain now. One can indeed establish the consistency of $L^\kappa_{\varepsilon,n}$ with $\mathcal{L}^\kappa_\varepsilon$, using the same discrete estimation technique as in [5, 6, 36]. We should point out that while the resulting discrete error bound in [5, 6, 36] does not show any dependence on κ, which is needed for proving the convergence of the discrete posterior density estimate in (2.19) to (2.11), this result is sufficient for understanding the consistency of $L^\kappa_{\varepsilon,n}$ with $\mathcal{L}^\kappa_\varepsilon$. Specifically, for a point cloud with distribution characterized by density $q(x)$, defined with respect to the volume form inherited by $\mathcal{M} \subset \mathbb{R}^d$, the discrete estimate for fixed-bandwidth Gaussian kernel (e.g., see Corollary 1 of [5] with $\alpha = 1/2, \beta = 0$ in their setup) states that the sampling error for obtaining an order-ϵ^2 of the density q with q_ϵ is of order $\mathcal{O}(q(x_i)^{1/2} n^{-1/2} \epsilon^{-(2+m/4)})$ and the error between $L^\kappa_{\varepsilon,n}$ and $\mathcal{L}^\kappa_\varepsilon$ is of order $n^{-1/2}\epsilon^{-(1/2+m/4)}$. The fact that ϵ appears in the denominator of these estimates suggests that one cannot take ϵ too small in practice and it also implies that ϵ should be adequately scaled with the size of the data, n. Since a direct use of these estimates requires knowing the

constants that depend on the volume of $\mathcal{M}$ that are difficult to estimate, we instead adopt an automated empirical method for choosing ϵ. Precisely, we follow [19] and plot

$$T(\varepsilon) := \sum_{i,j} \exp\left(-\frac{|x_i - x_j|^2}{4\varepsilon}\right)$$

as a function of ε and choose ε to be in the region where $\log\big(T(\varepsilon)\big)$ is approximately linear.

In the following three subsections we demonstrate the local kernel method for solving inverse problems through three numerical examples. In the first two examples, the embeddings are known and we set the model analytically, i.e., we first choose the ground truth $\kappa^\dagger$ and $u^\dagger$, and then compute the corresponding f as

$$f = \operatorname{div}(e^\theta \nabla u) = \frac{1}{\sqrt{\det g}} \partial_i \left(e^\theta g^{ij} \sqrt{\det g} \partial_j u \right), \tag{4.1}$$

where g is the Riemannian metric on $\mathcal{M}$. The third example will be an artificial surface where the embedding is unknown. We will then generate the truth using our kernel method. We will use the preconditioned Crank–Nicolson (pCN) algorithm to sample from the posterior [21, 8, 31]. This is a Metropolis–Hastings algorithm with proposal mechanism to move from u_n to u_n^* given by

$$u_n^\star \sim \left(1 - c^2\right)^{1/2} u_n + c\xi^{(n)}, \quad \xi \sim \pi_n = \mathcal{N}(0, C_{\tau,s}^n), \tag{4.2}$$

where $c \in (0, 1)$ is a tuning parameter. Note that if $u_n \sim \pi_n$ then $u_n^* \sim \pi_n$ showing that the prior is invariant for this kernel. Moreover, it is not difficult to see that detailed balance holds, and as a consequence the Metropolis-Hastings accept/reject mechanism involves only evaluation of the likelihood function. The advantage of pCN in our setting over a standard random-walk or Langevin algorithm is that the rate of convergence of pCN does not deteriorate with n; this has been established rigorously for a linear inverse problem in [31].

Remark 4.1. At each iteration of the MCMC algorithm, the forward map involves an eigenvalue decomposition of a different matrix for different θ's as shown in subsection 2.3.2. Hence large n's are not favored for computational purposes and this can be an issue for high-dimensional $\mathcal{M}$'s where the number of points grow as n^m if one discretizes each dimension by n.

4.2. One-dimensional elliptic problem on an unknown ellipse. In this subsection we take $\mathcal{M}$ to be an ellipse with semimajor and semiminor axis of length $a = 3$ and 1, embedded through

$$\iota(\omega) = (\cos\omega, a\sin\omega)^T, \quad \omega \in [0, 2\pi], \tag{4.3}$$

and the Riemannian metric is

$$g_{11}(\omega) = \sin^2\omega + a^2\cos^2\omega.$$

The truth is set to be

$$\kappa^\dagger = 2 + \cos\omega, \quad u^\dagger = \cos\omega,$$

and right-hand side f in (2.1) is defined through (4.1). One can check that both $u^\dagger$ and f have mean zero, i.e., $\int_0^{2\pi} u^\dagger(\omega)\sqrt{g_{11}(\omega)}d\omega = \int_0^{2\pi} f(\omega)\sqrt{g_{11}(\omega)}d\omega = 0$. We generate the point cloud $\{x_1, \dots, x_n\}$ according to (4.3) from a uniform grid of ω of size $n = 400$. The observations are given as noisy pointwise evaluations at subsets of the point cloud:

$$\ell_j = u(x_j) + \eta_j, \qquad j = 1, \dots, J,$$

where $\eta_j \sim \mathcal{N}(0, \sigma^2)$ are assumed to be independent. We will take $J = 100$, 200, 400, respectively with noise size varying as $\sigma = 0.01, 005, 0.1$. As discussed in subsection 2.3.1, we construct the prior with a self-tuning graph Laplacian, using $k = 2$ neighbors. We empirically discover that such a choice of k gives the best spectral approximation towards the Laplace–Beltrami operator on the ellipse, which has spectrum $\{i^2\}_{i=0}^{\infty}$ with eigenfunctions $\{\sin(i\omega), \cos(i\omega)\}_{i=0}^{\infty}$. We also tune empirically the parameters in (2.12) as $\tau = 0.05$ and $s = 4$.

In Figure 1, we plot the posterior means as functions of $\omega \in [0, 2\pi]$ and the 95% credible intervals for different σ and J's. While the point cloud Bayesian solution is only defined at the discrete point cloud, to ease the visualization we represent the outcome as continuous functions defined on $\omega \in [0, 2\pi]$. We see that the truth is mostly captured in the Bayesian confidence intervals. To quantify the error of reconstruction, we compute the relative ℓ_2 distances between the posterior mean $\bar{\kappa}$ and the truth $\kappa^\dagger$. Moreover, we compute the solution $\bar{u}$ of (2.17) with $\bar{\kappa}$ and its relative ℓ_2 distance to the true solution $u^\dagger$. As shown in Table 1, the reconstruction error for $u^\dagger$ is much smaller than the relative noise level defined as $\sqrt{n}\sigma/\|u^\dagger\|_2$.

Remark 4.2. Since our prior is on $\theta = \log(\kappa)$, we are actually approximating $\log(2+\cos\omega)$ with trigonometric functions and hence the truth $\kappa^\dagger$ is not simply the combination of the first two eigenfunctions in the prior. In other words, although the truth $\kappa^\dagger$ is in the support of the prior, the fact that its coordinates in the eigenbasis do not decay like that in the expansion (2.5) makes it difficult to reconstruct.

Regarding Remark 4.2, we consider another prior with self-tuning graph constructed with $k = 0.2n$ points. This new graph Laplacian gives a worse spectral approximation to the Laplace–Beltrami operator in the underlying manifold, as its spectrum saturates instead of growing at the appropriate rate. In other words, the basis functions associated with the high frequencies will be given more weight in the expansion (2.5). This can be beneficial in practical applications since it effectively enlarges the support of the prior. Below in Figure 2, we solve the inverse problem using this new prior in the case $\sigma = 0.1$. The parameters are tuned empirically: $\tau = 0.75$, $s = 8$. It can be seen that although the reconstructions are rougher than those in Figure 1, they capture better the shape of $\kappa^\dagger$, with the help of the higher frequencies. Essentially, larger τ (corresponds to more nontrivial modes in the representation in (2.5)) gives less bias but larger variance, which is consistent with the theory of nonparameteric statistical estimation (e.g, section 1.7 of [58]).

Remark 4.3. We note that the inexact reconstruction is partially due to the ill-posedness of the elliptic inverse problem [47]. As can be seen in Table 1, the reconstruction error for $\bar{\kappa}$ is much larger than that for $\bar{u}$: a wide range of κ's around $\kappa^\dagger$ give solutions u which are "close enough" to $u^\dagger$ (within a range of order σ) that the algorithm cannot distinguish. When σ is

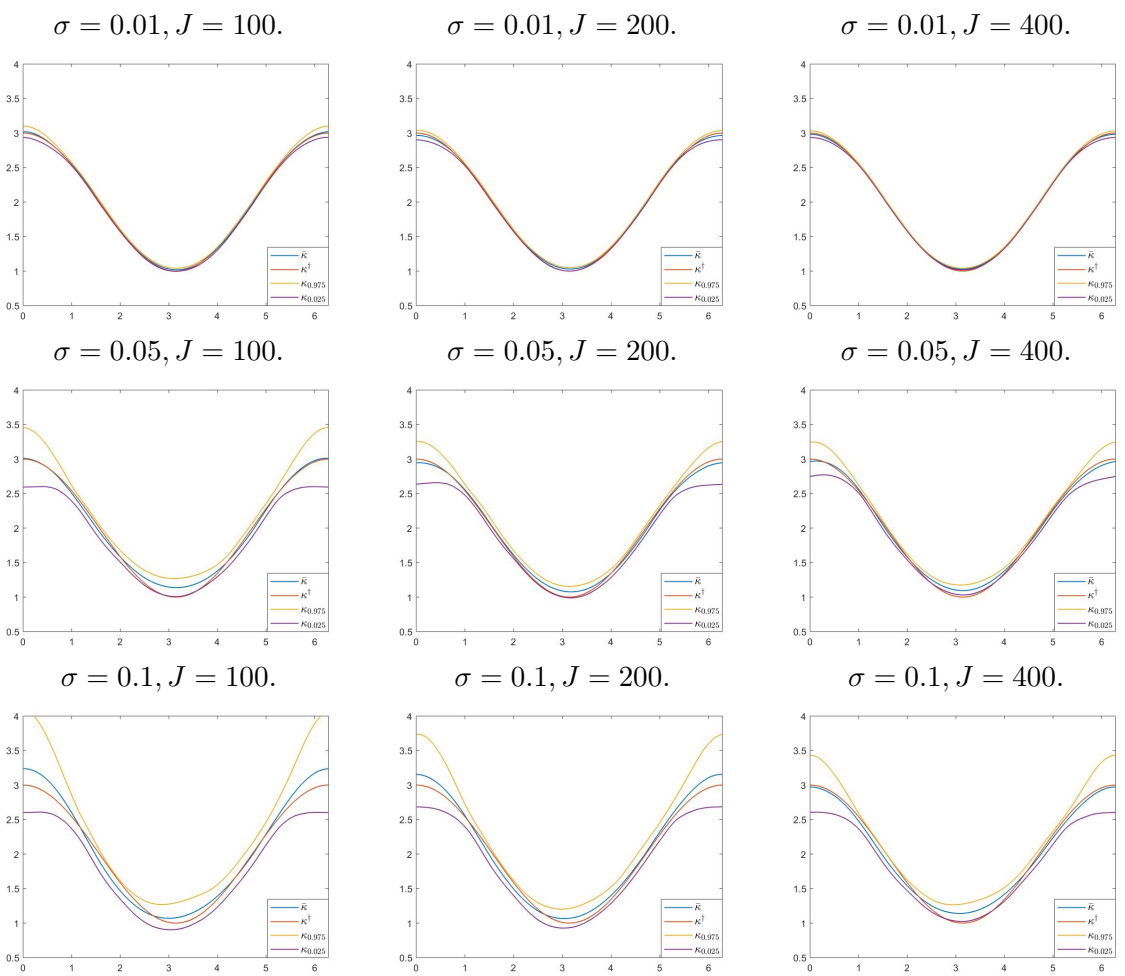

Figure 1. *Posterior means and 95% credible intervals for different σ's and J's. Here $\bar{\kappa}$, $\kappa_{0.025}$, and $\kappa_{0.975}$ represent the posterior mean, 2.5%, and 97.5% posterior quantiles, respectively.*

Table 1

Relative error of $\bar{\kappa}$ and $\bar{u}$ for different noise levels, σ's, and number of observations, J, where $\bar{\kappa}$ and $\bar{u}$ are the posteriors means for κ and u, respectively. In the last row, the relative noise level for each σ is reported for diagnostic purposes. Particularly, note that the reconstruction error for $u^\dagger$ is much smaller than the relative noise level.

σ	0.01			0.05			0.1		
J	100	200	400	100	200	400	100	200	400
$\frac{\|\bar{\kappa}-\kappa^\dagger\|_2}{\|\kappa^\dagger\|_2}$	0.60%	0.80%	0.62%	2.85%	1.96%	2.18%	5.46%	3.90%	3.45%
$\frac{\|\bar{u}-u^\dagger\|_2}{\|u^\dagger\|_2}$	0.26%	0.23%	0.23%	1.08%	0.83%	0.90%	1.70%	1.37%	1.70%
$\frac{\sqrt{n}\sigma}{\|u^\dagger\|_2}$	1.41%			7.07%			14.14%		

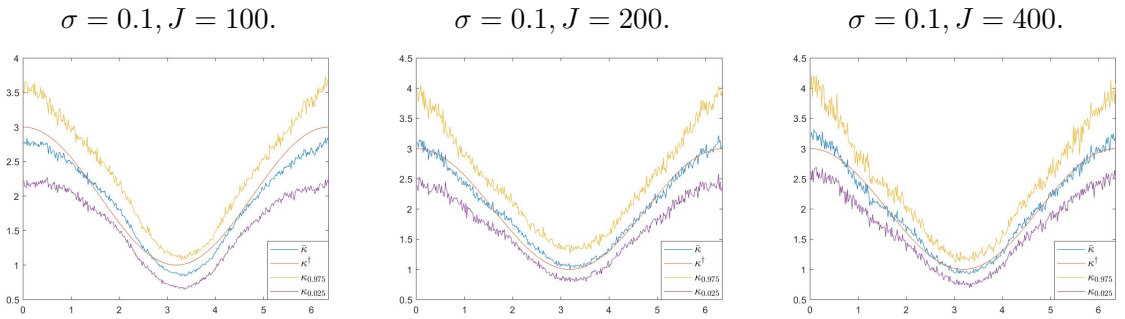

Figure 2. *Posterior means and* 95% *credible intervals for* $\sigma = 0.1$ *and different* J*'s. Here* $\kappa_{0.025}$ *and* $\kappa_{0.975}$ *represent the* 2.5% *and* 97.5% *posterior quantiles, respectively.*

large, such a tolerance is larger and the inverse problem becomes more difficult. This issue, together with Remark 4.2, explains why one cannot expect exact recovery of $\kappa^\dagger$ as seen in Figure 1.

4.3. Two-dimensional elliptic problem on an unknown torus. In this subsection we take $\mathcal{M}$ to be $\mathbb{T}^2$ embedded in $\mathbb{R}^3$ through

$$\iota(\omega_1, \omega_2) = \big((2 + \cos\omega_1)\cos\omega_2, (2 + \cos\omega_1)\sin\omega_2, \sin\omega_1\big)^T, \quad \omega_1, \omega_2 \in [0, 2\pi], \tag{4.4}$$

and the Riemannian metric is

$$g(\omega_1, \omega_2) = \begin{bmatrix} 1 & 0 \\ 0 & (2 + \cos\omega_1)^2 \end{bmatrix}.$$

The truth is set to be

$$\kappa^\dagger(\omega_1, \omega_2) = 2 + \sin\omega_1 \sin\omega_2, \quad u^\dagger = \sin\omega_1 \sin\omega_2,$$

and f is again specified through (4.1). One can check that both $u^\dagger$ and f have mean zero, i.e., $\int u^\dagger \sqrt{\det g} = \int f \sqrt{\det g} = 0$. For computational reasons as in Remark 4.1, we generate the point cloud according to (4.4) from a 20×20 uniform grid on $[0, 2\pi] \times [0, 2\pi]$ and the observations are given as noisy pointwise evaluations at all points. The graph Laplacian for the prior is constructed with $k = 16$ neighbors and again we empirically tune the parameters: $\tau = 0.08$ and $s = 6$. Unlike the ellipse case, we cannot get an almost exact spectral approximation given that we are only discretizing each dimension by 20 points. The eigenfunctions of the graph Laplacian are wiggly in this case, for which reason we need a large s to ensure sufficient decay of the spectrum to obstain relatively smooth reconstructions. In Figure 3, we plot the posterior means and the standard deviations as functions on $[0, 2\pi] \times [0, 2\pi]$; we note that the uncertainty is large when the function $\sin\omega_1 \sin\omega_2$ crosses 0. Table 2 quantifies the reconstruction error as usual and the reconstruction error for $u^\dagger$ is again much smaller than the noise levels, which are 2%, 10%, and 20%, respectively. However, the reconstruction error for $u^\dagger$ decreases with decreasing σ while the error for $\kappa^\dagger$ does the opposite, a manifestation of the ill-posedness.

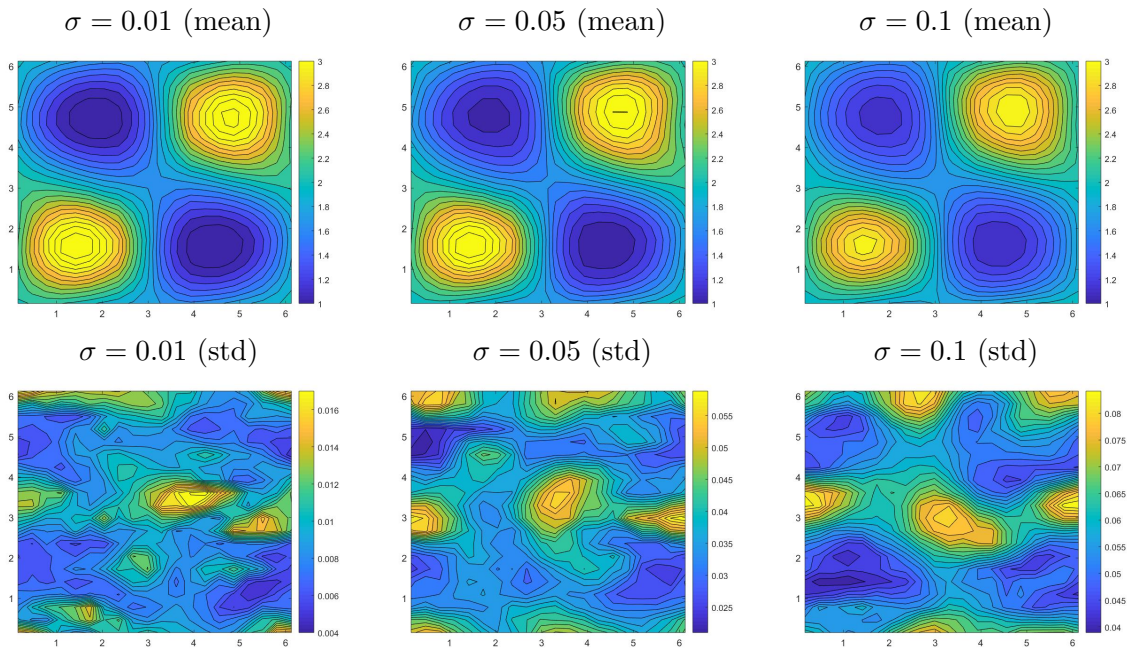

Figure 3. *Posterior means (first row) and standard deviations (second row) of posteriors for different σ's. horizontally. We have extended by interpolation the point cloud solution in order to ease visualization.*

Table 2

Relative error of $\kappa^\dagger$ and $\bar{u}$ for different noise level σ's. In the last row, the relative noise level is reported for diagnostic purpose. Particularly, note that the reconstruction error for $u^\dagger$ is much smaller than the relative noise level.

σ	0.01	0.05	0.1
$\frac{\|\bar{\kappa}-\kappa^\dagger\|_2}{\|\kappa^\dagger\|_2}$	8.56%	8.34%	6.94%
$\frac{\|\bar{u}-u^\dagger\|_2}{\|u^\dagger\|_2}$	1.8%	2.8%	4.8%
$\frac{\sqrt{n}\sigma}{\|u^\dagger\|_2}$	2%	10%	20%

Remark 4.4. For this example we solved (2.17) by taking the pseudoinverse $\hat{u}_n = (L^\kappa_{\varepsilon,n})^\dagger f_n$ instead, as this is numerically more stable than taking the eigenvalue decomposition of the asymmetric matrix $L^\kappa_{\varepsilon,n}$. Moreover, $\hat{u}_n$ is consistent with $u^\dagger$ for this specific problem as explained below. We have that $\hat{u}_n$ solves the following problem:

$$\hat{u}_n = \min\left\{\|u\|_2 : u \in \operatorname{argmin} \|L^\kappa_{\varepsilon,n} u - f_n\|_2\right\}. \tag{4.5}$$

The fact that f has zero mean implies $\langle f_n, 1\rangle_q = 0$ in the large n limit. Then (2.17) has a strong solution as mentioned in subsection 2.3.2 and so the characterization (4.5) implies that $\hat{u}_n$ is also a strong solution, with $\sum_{i=1}^n \hat{u}_n^i = 0$. Notice that the truth $u^\dagger$ also satisfies $\int u^\dagger = 0$ and hence makes $\hat{u}_n$ consistent.

4.4. Two-dimensional elliptic problem on an unknown artificial surface. In this subsection we consider an artificial dataset from Keenan Crane's 3D repository [22]. The dataset is

made of 2930 points sampled from a cow-shaped surface homeomorphic to the two-dimensional sphere. The purpose of this subsection is to demonstrate that our kernel method can be applied to more complicated manifolds. To avoid an inverse crime [39], we generate the truth using all 2930 points but solve the inverse problem on a subset of size $n = 1000$.

To be more precise, we generate $\kappa^\dagger$ from the Gaussian measure $\mathcal{N}(0, (\tau I + \Delta_{2930}))^{-s}$, where Δ_{2930} is the self-tuning graph Laplacian constructed with $k = 100$ neighbors and $\tau = 0.7$, $s = 6$. We then set $u^\dagger$ to be $10(\varphi_2 - c)$, where φ_2 is the second eigenvector of Δ_{2930} and c is a constant chosen below. The factor 10 in the definition of u is only to match the order of magnitude with $\kappa^\dagger$. We then set $f = L^{\kappa^\dagger}_{\varepsilon,2930} u^\dagger$. This would serve as our ground truth, and now we solve the inverse problem on a random subset X of $n = 1000$ points.

On the point cloud X, the truth becomes $\kappa^\dagger|_X$ and $u^\dagger|_X$. As mentioned above, the solution $u^\dagger|_X$ needs to have zero mean to be consistent with our theory. Since we do not have access to the Riemannian metric as in the previous two examples, we instead require $\langle u^\dagger|_X, 1\rangle_q = 0$, which gives the choice of c above. Again we consider noisy pointwise observations at all 1000 points. The inputs of the problem are now $f|_X$ and noisy $u^\dagger|_X$. The noise level is 10%, which gives $\sigma = 0.0186$. The prior that we use has the same parameter τ as used to obtain our synthetic truth, but is defined using a graph Laplacian on X. Namely, we consider

$$\pi_n = \mathcal{N}(0, (0.7I + \Delta_{1000})^{-6}),$$

where Δ_{1000} is constructed with $k = 80$ neighbors. Figure 4 shows the plots of the posterior mean, the truth, the error, and the standard deviation. The reconstruction errors are $\|\bar{\kappa} - \kappa^\dagger|_X\|_2/\|\kappa^\dagger|_X\|_2 = 9.73\%$ and $\|\bar{u} - u^\dagger|_X\|_2/\|u^\dagger|_X\|_2 = 16.24\%$. We remark that the large relative error for $u^\dagger$ is partly due to the fact that the point cloud of size 1000 does not approximate the original one well and in particular

$$L^{\kappa^\dagger|_X}_{\varepsilon,1000} u^\dagger|_X \neq f|_X. \tag{4.6}$$

When we solve for $\hat{u}$ in (4.6), i.e., solving

$$L^{\kappa^\dagger|_X}_{\varepsilon,1000} \hat{u} = f|_X,$$

we get $\|\hat{u} - u^\dagger|_X\|_2/\|u^\dagger|_X\|_2 = 17.01\%$ and this is the best one can hope for in terms of reconstructing $u^\dagger|_X$. Hence we see that the above relative error has already reached the limit of the method.

4.5. Hierarchical Bayesian formulation. As mentioned in subsection 2.1.1, the choice of τ is crucial and would require some prior knowledge of the length-scale of the function to be reconstructed. In this section, we demonstrate how one can learn the parameter τ together with κ through a hierarchical Bayesian approach proposed in [24]. We emphasize that the hierarchical approach may not be able to find the precise length-scale of the parameter to be reconstructed, and hence should only be applied when little prior knowledge on the length-scale is available. We will only focus on the point cloud inverse problem as in subsection 2.3.

We remark that our choice of priors in (2.4) and (2.12) differs from the ones used in [24] by the scaling constants. In the continuum space, the familiy of measures defined by (2.4) are

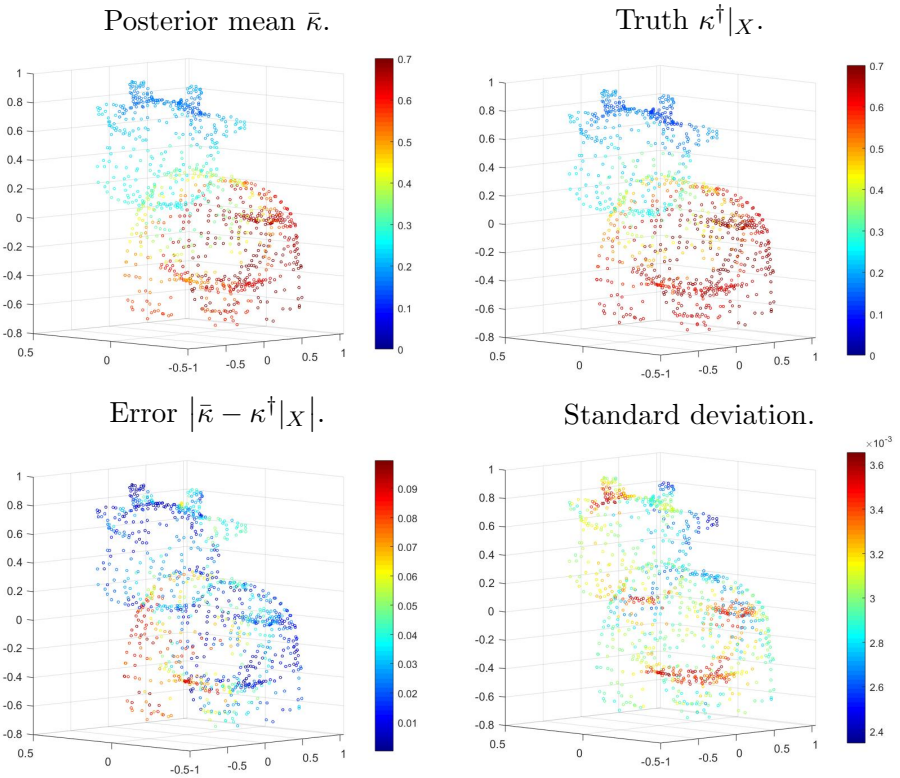

Figure 4. *Reconstruction for the cow-shaped manifold.*

mutually singular, which leads to technical difficulties when designing hierarchical methods. However, in the point cloud setting, the family of measures as in (2.12) are simply multivariate normal and are equivalent. The formulation in [24] then carries over.

The idea is to consider a joint prior on (θ_n, τ) that takes the form

$$\Pi(\theta_n, \tau) = \pi_0(\tau)\pi(\theta_n|\tau) := \pi_0(\tau)\pi_\tau(\theta_n),$$

where π_0 is a distribution on $\mathbb{R}^+$ and the conditional distribution π_τ is taken as in (2.12). Recall that π_τ has the form

$$\pi_\tau = \mathcal{N}(0, C^n_{\tau,s}), \quad C^n_{\tau,s} = c_n(\tau)(\tau I + \Delta_n)^{-s},$$

where $c_n(\tau)$ normalizes the draws so that $u \sim \pi_n$ satisfies $\mathbb{E}|u_n|^2 = n$. Now we can define the forward map $F_n : \mathbb{R}^n \times \mathbb{R}^+ \to \mathbb{R}^n$ that associates a pair (θ_n, τ) with the unique mean-zero solution u_n of (2.17). We notice that F_n is essentially the same as before except the additional τ that does not play a role in the definition. Denoting $G_n = D_n \circ F_n$, the joint posterior Π^y can be written as a change of measure with respect to the prior,

$$\frac{d\Pi^y}{d\Pi}(\theta_n, \tau) \propto \exp\left(-\frac{1}{2}|y - G_n(\theta_n, \tau)|^2_\Gamma\right). \tag{4.7}$$

4.5.1. Sampling. To sample from the joint posterior $\theta_n, \tau|y$, we will use a Metropolis within Gibbs sampling scheme by updating $\theta_n|\tau, y$ and $\tau|\theta_n, y$ alternately. Sampling of $\theta_n|\tau, y$ reduces to the nonhierarchical setting, where we use the pCN algorithm. Sampling of $\tau|\theta_n, y$ is more delicate since one first needs to make sense of this conditional distribution. Instead of making the presentation too involved, we will only present the algorithm and refer to [24] for more details. The idea is to use symmetric random-walk Metropolis–Hastings, with acceptance probability to accept the proposal τ, given the current chain value γ as

$$a(\tau, \gamma) = \exp\left(-\frac{1}{2}\left[H(\tau) - H(\gamma)\right]\right)\frac{\pi_0(\tau)}{\pi_0(\gamma)} \wedge 1, \tag{4.8}$$

where

$$H(\tau) = \sum_{i=1}^{n} \log \lambda_i(\tau) + \frac{\langle \theta, \varphi_i \rangle^2}{\lambda_i(\tau)}.$$

Here, $\{(\lambda_i(\tau), \varphi_i)\}_{i=1}^n$ are the eigenvalue-eigenfunction pairs of $C^n_{\tau,s}$.

From (4.8), we see that the algorithm favors length-scale τ's that give small values of $H(\tau)$. As τ approaches 0, the first eigenvalue of $(\tau I + \Delta_n)^{-s}$ goes to infinity and so the normalizing constant $c_n(\tau)$ approaches 0. However, the eigenvalues of $(\tau I + \Delta_n)^{-s}$, except the first one, do not change much since τ is now a much smaller quantity. Hence when multiplied by $c_n(\tau)$, they converge to 0. In other words $\lambda_i(\tau)$ approaches 0 as τ goes to 0 for $i \geq 2$, which then implies that $\sum_{i=1}^n \log \lambda_i(\tau) \to -\infty$. So the first term in H is minimized at $\tau = 0$, while the second term $\langle \theta, \varphi_i \rangle^2 / \lambda_i(\tau)$ favors large τ by a similar argument as above. Then the minimum of H is attained by balancing the two sums, using the coefficients $\langle \theta, \varphi_i \rangle$. Hence the algorithm will give a τ that is consistent with these coefficients, reflecting the length-scale of θ.

4.5.2. Numerical experiments. To demonstrate the hierarchical approach, we focus on the ellipse case with three truths of different length-scales: $\kappa^\dagger = e^{\cos(\omega)}$, $e^{\cos(5\omega)}$, $e^{\cos(8\omega)}$ with a fixed $f = \frac{1}{5}\sin\omega$. We fix f so that we are solving the same inverse problem with different underlying truth $\kappa^\dagger$'s. In this case we no longer have analytic solutions for u and we use the MATLAB PDE toolbox. The point cloud is generated as in subsection 4.2 with $n = 100$, with pointwise observations with noise level $\sigma = 0.01$ at all points. We take $\pi_0 = \mathcal{N}(2, 1)$ and $s = 4$. In the hierarchical setting, it takes longer for the chains to mix, where we run the chain for 3×10^7 iterations and use the last 5×10^6 samples for computations. We compare the hierarchical approach with the nonhierarchical one, where we use a same τ to reconstruct the three $\kappa^\dagger$'s. Figure 5 below shows the corresponding reconstructions.

The first row corresponds to the nonhierarchical approach where we fix $\tau = 2$, which is finely tuned to match the length-scale of $e^{\cos(5\omega)}$, for all three problems. The reconstruction is then acceptable for $e^{\cos(5\omega)}$ but is poorer for the other two. For $e^{\cos(\omega)}$, the reconstruction still fits the shape of the truth, but since the prior now has a length-scale much larger than that of $e^{\cos(\omega)}$, the reconstruction is oscillatory. On the other hand, the reconstruction of $e^{\cos(8\omega)}$ fails to capture the shape of the truth. This is because the prior now has a smaller length-scale than that of $e^{\cos(8\omega)}$, so that frequencies as high as $\cos(8\omega)$ barely belong to the prior. The second and third rows correspond to the reconstructions of the hierarchical approach and the corresponding sample paths for τ. We see that the credible intervals capture most

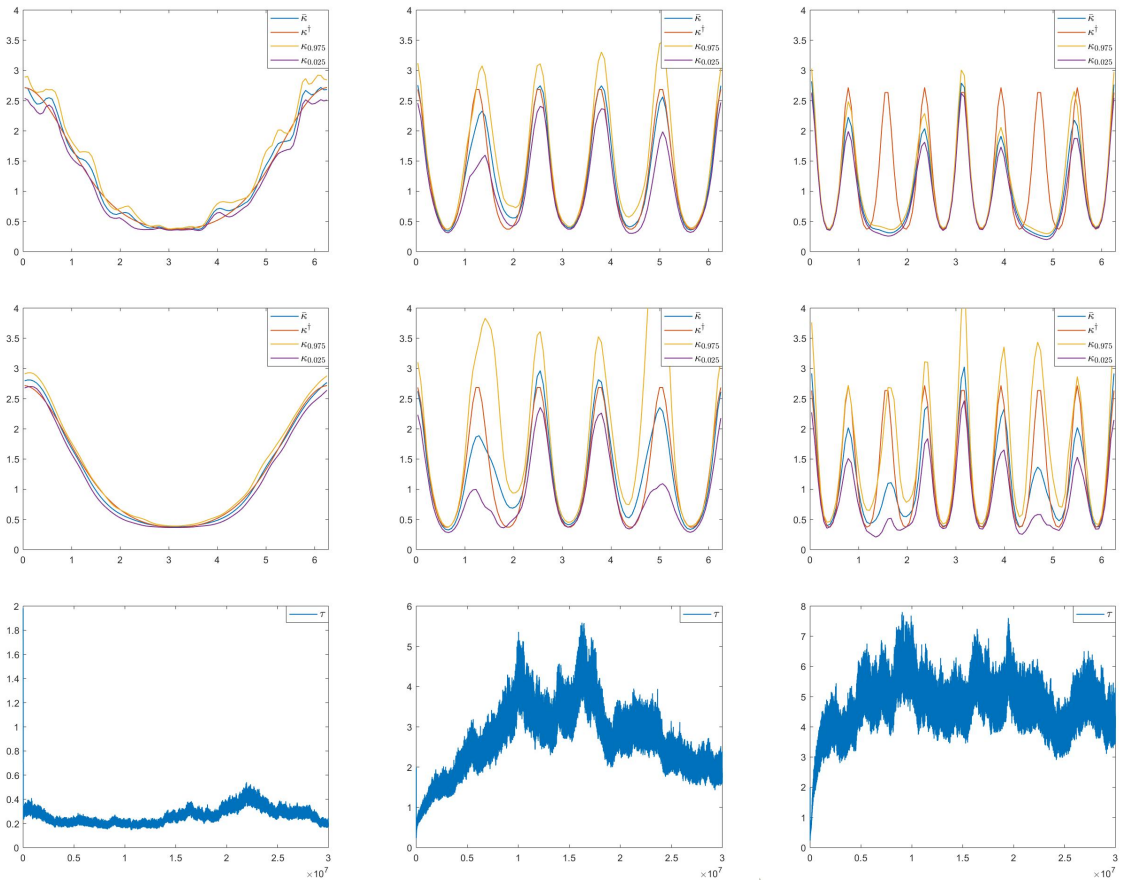

Figure 5. *Posterior means and* 95% *credible intervals for different truths. Figures are arranged so that the first two rows correspond to nonhierarchical and hierarchical, respectively, and the third row shows the sample paths for* τ*. The three columns represent the truths* $e^{\cos(\omega)}, e^{\cos(5\omega)}, e^{\cos(8\omega)}$*, respectively.*

of the truths and the reconstructions are much better for $e^{\cos(\omega)}$ and $e^{\cos(8\omega)}$ than with the nonhierarchical approach. Table 3 quantifies the reconstruction error. We notice that the hierarchical approach performs worse than the nonhierarchical one for $e^{\cos(5\theta)}$; this is because the chosen $\tau = 2$ agrees with the length-scale of the true diffusion coefficient. This fact suggests that the hierarchical approach only improves the performance when little prior knowledge of the length-scale is known. From the sample paths for τ's, we see that the chains have large variance and do not concentrate on a particular value. This is due to the ill-posedness of the inverse problem where κ's of different length-scales give equally good reconstructions of u and, hence, the algorithm cannot distinguish between them. So in general the algorithm may not give the precise length-scale but a possible range of it.

Remark 4.5. Notice that in the above the noise level has been set to be small. When the noise level σ is large, the performance of the hierarchical approach may be worse, as shown in Figure 6. The reason is that the algorithm sees only the noisy data, which is the truth $u^\dagger$ perturbed by noise. In other words, the length-scale of the data is corrupted by the noise,

Table 3

Relative error of $\bar{\kappa}$ and $\bar{u}$ for different truths. Here "NH" and "H" stand for nonhierarchical and hierarchical, respectively. In the last row, the relative noise level for each σ is reported for diagnostic purposes.

$\kappa^\dagger$	$e^{\cos(\omega)}$		$e^{\cos(5\omega)}$		$e^{\cos(8\omega)}$	
Method	NH	H	NH	H	NH	H
$\frac{\|\bar{\kappa}-\kappa^\dagger\|_2}{\|\kappa^\dagger\|_2}$	9.07%	3.69%	11.17%	17.41%	39.38%	28.82%
$\frac{\|\bar{u}-u^\dagger\|_2}{\|u^\dagger\|_2}$	0.84%	0.72%	0.73%	0.69%	0.87%	0.83%
$\frac{\sqrt{n}\sigma}{\|u^\dagger\|_2}$	1.87%		1.26%		1.27%	

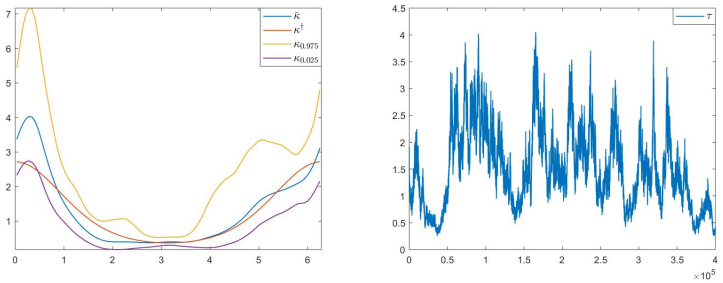

Figure 6. *Reconsturction of $e^{\cos(\omega)}$ and sample path for τ when $\sigma = 0.1$.*

which has length-scale converging to 0 ($\tau \to \infty$) in the large n and J limit if the noise is independent, i.e., $\Gamma = I$. As shown in Figure 6, the chain for τ oscillates in a wide range of values, suggesting that the data contain little information on this parameter.

5. Conclusions and future work.

- This paper introduced kernel-based methods for the solution of inverse problems on manifolds. We have shown through rigorous analysis that the forward map can be replaced by a kernel approximation while keeping small the total variation distance to the true posterior. Through numerical experiments we have shown that a point cloud discretization to the kernel approximation may allow one to implement the inverse problem on point clouds, without reference to the underlying manifold.
- An important question of theoretical interest when solving the inverse problem on the point cloud is how to choose optimally the kernel bandwith ε in terms of the number n of manifold samples. We conjecture that the convergence of the graph posteriors to the ground truth posterior, and guidelines on the choice of kernel bandwith, may be established by generalizing the spectral graph theory results in [12, 34] to anisotropic diffusions, and using the variational techniques introduced in [33, 31]. The analysis of these questions will be the subject of future work.
- We streamlined the presentation by working on a closed manifold with no boundary. We expect that the numerical and theoretical results may be extended to Neumann, Robin, and Dirichlet boundary conditions using the results and ideas in [36, 40, 54, 57].
- The practical success of the Bayesian approach is heavily dependent on the choice of prior. Here, we have used Matérn-type priors that are flexible models widely used in spatial statistics and the geophysical sciences [55]. While the hierarchical approach to

the inverse problem [24, 35] is effective for learning the prior length-scale from data in certain regimes as we have numerically shown, a more robust algorithm is needed and this merits an extensive further investigation.

- A topic of further research will be the extension of the kernel-based approximation to PDEs and inverse problems to other PDEs and ODEs beyond the elliptic model considered here.

Appendix A. Proofs of lemmas.

Proof of Lemma 3.2. The proof proceeds by a standard Lax–Milgram argument. Throughout, $C > 0$ denotes a constant independent of ε and κ that may change from line to line. Consider the bilinear and linear functionals

$$B : L_0^2 \times L_0^2 \to \mathbb{R}, \qquad (u, v) \mapsto \langle u, \mathcal{L}_\varepsilon^\kappa v\rangle, \qquad F : L_0^2 \to \mathbb{R}, \qquad v \mapsto \langle v, f\rangle.$$

Clearly, B and F are bounded. To show that B is coercive, note that by [54, Theorem 7.2] there exists $C > 0$ such that, for all $\varepsilon > 0$ $v \in L_0^2(\mathcal{M})$,

$$\langle v, \mathcal{L}_\varepsilon v\rangle \geq C\|v\|_{L^2}, \tag{A.1}$$

where

$$\mathcal{L}_\varepsilon v := \frac{1}{\sqrt{4\pi}\varepsilon^{\frac{m}{2}+1}} \int \exp\left(-\frac{|x-\tilde{x}|^2}{4\varepsilon}\right)[v(x) - v(\tilde{x})]dV(\tilde{x}).$$

It follows that, for $v \in L_0^2$,

$$\begin{aligned}
\langle v, \mathcal{L}_\varepsilon^\kappa v\rangle &= \frac{1}{\sqrt{4\pi}\varepsilon^{\frac{m}{2}+1}} \int\int \exp\left(-\frac{|x-\tilde{x}|^2}{4\varepsilon}\right)\sqrt{\kappa(x)\kappa(\tilde{x})}v(x)[v(x) - v(\tilde{x})]dV(\tilde{x})dV(x) \\
&= \frac{1}{\sqrt{4\pi}\varepsilon^{\frac{m}{2}+1}} \int\int \exp\left(-\frac{|x-\tilde{x}|^2}{4\varepsilon}\right)\sqrt{\kappa(x)\kappa(\tilde{x})}v(\tilde{x})[v(\tilde{x}) - v(x)]dV(\tilde{x})dV(x) \\
&= \frac{1}{2\sqrt{4\pi}\varepsilon^{\frac{m}{2}+1}} \int\int \exp\left(-\frac{|x-\tilde{x}|^2}{4\varepsilon}\right)\sqrt{\kappa(x)\kappa(\tilde{x})}|v(x) - v(\tilde{x})|^2dV(\tilde{x})dV(x) \\
&\geq \kappa_{\min}\frac{1}{2\sqrt{4\pi}\varepsilon^{\frac{m}{2}+1}} \int\int \exp\left(-\frac{|x-\tilde{x}|^2}{4\varepsilon}\right)|v(x) - v(\tilde{x})|^2dV(\tilde{x})dV(x) \\
&= \kappa_{\min}\langle v, \mathcal{L}_\varepsilon v\rangle \geq C\kappa_{\min}\|v\|_{L^2}^2,
\end{aligned}$$

establishing the coercivity of B. The existence and uniqueness of a weak solution, as well as the bound (3.1), follow from the Lax–Milgram theorem.

Proof of Lemma 3.3. Our proof follows the same argument as [17, Lemma 8] but keeps track of the coefficients of the higher order terms. Let

$$G_\varepsilon u(x) = \varepsilon^{-\frac{m}{2}} \int h\left(\frac{|x-\tilde{x}|^2}{\varepsilon}\right)u(\tilde{x})dV(\tilde{x}), \quad h(z) = \frac{1}{\sqrt{4\pi}}e^{-\frac{z}{4}}.$$

Let $0 < \beta < \frac{1}{2}$. We can localize the integration near x due to the exponential decay of e^{-x^2}:

$$\begin{aligned}
&\left|\varepsilon^{-\frac{m}{2}} \int_{\tilde{x}\in\mathcal{M}:|\tilde{x}-x|>\varepsilon^\beta} \exp\left(-\frac{|x-\tilde{x}|^2}{4\varepsilon}\right) u(\tilde{x}) dV(\tilde{x})\right| \\
\leq &\varepsilon^{-\frac{m}{4}} \|u\|_{L^2} \sqrt{\varepsilon^{-\frac{m}{2}} \int_{\tilde{x}\in\mathcal{M}:|\tilde{x}-x|>\varepsilon^\beta} \exp\left(-\frac{|x-\tilde{x}|^2}{2\varepsilon}\right) dV(\tilde{x})} \\
\leq &\varepsilon^{-\frac{m}{4}} \|u\|_{L^2} \sqrt{\int_{x+\sqrt{\varepsilon}\tilde{x}\in\mathcal{M}:|\tilde{x}|>\varepsilon^{\beta-1/2}} \exp\left(-\frac{1}{2}|\tilde{x}|^2\right) dV(\tilde{x})} \\
\leq &C\varepsilon^{-\frac{m}{4}} \|u\|_{L^2} \sqrt{\mathbb{P}\{\mathcal{N}(0,1) > \varepsilon^{\beta-1/2}\}} \\
\leq &C\varepsilon^{-\frac{m}{4}} \|u\|_{L^2} \exp\left(-c\varepsilon^{2\beta-1}\right) \leq C\|u\|_{L^2}\varepsilon^2,
\end{aligned}$$

where in the last inequality since $2\beta < 1$, $\exp(-c\varepsilon^{2\beta-1})$ decays faster than any polynomial in ε and in particular for ε small enough it decays faster than $\varepsilon^{2+\frac{m}{4}}$. Therefore,

$$G_\varepsilon u(x) = \varepsilon^{-\frac{m}{2}} \int_{\tilde{x}\in\mathcal{M}:|\tilde{x}-x|<\varepsilon^\beta} h\left(\frac{|x-\tilde{x}|^2}{\varepsilon}\right) u(\tilde{x}) dV(\tilde{x}) + O(\|u\|_{L^2}\varepsilon^2).$$

Now we Taylor expand u near x. Let $(s_1, \ldots, s_m)$ be the geodesic coordinates at x and $u(\tilde{x}) = u(\tilde{x}(s_1, \ldots, s_m)) = \tilde{u}(s_1, \ldots, s_m) = \tilde{u}(s)$. Then

$$\begin{aligned}
u(\tilde{x}) - u(x) = \tilde{u}(s) - \tilde{u}(0) = &\sum_{i=1}^m s_i \frac{\partial \tilde{u}}{\partial s_i}(0) + \frac{1}{2}\sum_{i=1}^m\sum_{j=1}^m s_i s_j \frac{\partial^2 \tilde{u}}{\partial s_i \partial s_j}(0) \\
&+ \frac{1}{6}\sum_{i=1}^m\sum_{j=1}^m\sum_{k=1}^m s_i s_j s_k \frac{\partial^3 \tilde{u}}{\partial s_i \partial s_j \partial s_k}(0) + \delta(s),
\end{aligned}$$

where

$$\delta(s) = \frac{1}{24}\int_0^1\int_0^1\int_0^1\int_0^1 \sum_{i=1}^m\sum_{j=1}^m\sum_{k=1}^m\sum_{\ell=1}^m s_i s_j s_k s_\ell \frac{\partial^4 \tilde{u}}{\partial s_i \partial s_j \partial s_k \partial s_\ell}(t_1 t_2 t_3 t_4 s) dt_1 dt_2 dt_3 dt_4.$$

Then expanding the fourth order term in $G_\varepsilon u$, we have

$$\begin{aligned}
|T_u(x)| &:= \left|\varepsilon^{-\frac{m}{2}} \int_{|s|<\varepsilon^\beta} h\left(\frac{|s|^2}{\varepsilon}\right)\delta(s) ds\right| \\
&\leq \frac{m^4\varepsilon^{4\beta}}{24}\int_0^1\int_0^1\int_0^1\int_0^1\int_{|s|<\varepsilon^\beta} \varepsilon^{-\frac{m}{2}} h\left(\frac{|s^2|}{\varepsilon}\right) \|\nabla^4 \tilde{u}(t_1 t_2 t_3 t_4 s)\| ds dt_1 dt_2 dt_3 dt_4 \\
&= \frac{m^4\varepsilon^{4\beta}}{24}\int_0^1\int_0^1\int_0^1\int_0^1\int_{|r|<t_1 t_2 t_3 \varepsilon^\beta} (t_1 t_2 t_3 t_4)^{-d}\varepsilon^{-\frac{d}{2}} h\left(\frac{|r|^2}{t_1^2 t_2^2 t_3^2 t_4^2 \varepsilon}\right) \|\nabla^4 \tilde{u}(r)\| dr dt_1 dt_2 dt_3 dt_4 \\
&:= \frac{m^4\varepsilon^{4\beta}}{24}\int_0^1\int_0^1\int_0^1\int_0^1\int_{|r|<t_1 t_2 t_3 t_4 \varepsilon^\beta} K(r,\tau)\|\nabla^4 \tilde{u}(r)\| dr d\tau,
\end{aligned}$$

where we have used the notation $\tau := (t_1, t_2, t_3, t_4)$ and $d\tau := dt_1 dt_2 dt_3 dt_4$. By interchanging the order of integration and noticing that $\nabla^4 \tilde{u}(r) = \nabla^4 u(x+\xi)$, where ξ's are directional vectors that are independent of x, we have

$$\begin{aligned}
\|T_u\|_{L^2}^2 &\le \left(\frac{m^4 \varepsilon^{4\beta}}{24}\right)^2 \int_{\mathcal{M}} \int_0^1 \int_0^1 \int_0^1 \int_0^1 \left[\int_{|r|<t_1t_2t_3t_4\varepsilon^\beta} K(r,\tau)\|\nabla^4 \tilde{u}(r)\| dr\right]^2 d\tau dV(x) \\
&\le \left(\frac{m^4 \varepsilon^{4\beta}}{24}\right)^2 \int_{\mathcal{M}} \int_0^1 \int_0^1 \int_0^1 \int_0^1 \left[\int_{|r|<t_1t_2t_3t_4\varepsilon^\beta} K(r,\tau) dr\right] \\
&\qquad \times \left[\int_{|r|<t_1t_2t_3t_4\varepsilon^\beta} K(r,\tau)\|\nabla^4 \tilde{u}(r)\|^2 dr\right] d\tau dV(x) \\
&= \left(\frac{m^4 \varepsilon^{4\beta}}{24}\right)^2 \int_0^1 \int_0^1 \int_0^1 \int_0^1 \left[\int_{|r|<t_1t_2t_3t_4\varepsilon^\beta} K(r,\tau) dr\right] \\
&\qquad \times \left[\int_{|r|<t_1t_2t_3t_4\varepsilon^\beta} K(r,\tau) \int_{\mathcal{M}} \|\nabla^4 u(x+\xi)\|^2 dV(x) dr\right] d\tau \\
&= \left(\frac{m^4 \varepsilon^{4\beta}}{24}\right)^2 \|\nabla^4 u\|_{L^2}^2 \int_0^1 \int_0^1 \int_0^1 \int_0^1 \left[\int_{|r|<t_1t_2t_3t_4\varepsilon^\beta} K(r,\tau) dr\right]^2 d\tau.
\end{aligned}$$

By a change of variable $r = t_1 t_2 t_3 t_4 \sqrt{\varepsilon} s$, we notice that

$$\int_{|r|<t_1t_2t_3t_4\varepsilon^\beta} K(r,\tau) dr = \int_{|r|<t_1t_2t_3t_4\varepsilon^\beta} (t_1t_2t_3t_4)^{-d} \varepsilon^{-\frac{d}{2}} h\left(\frac{|r|^2}{t_1^2 t_2^2 t_3^2 t_4^2 \varepsilon}\right) dr = \int_{|z|<\varepsilon^{\beta-1/2}} h(|z|^2),$$

which can be bounded by a constant independent of ε by the Gaussianity of $h(|z|^2)$. Hence

$$\|T_u\|_{L^2} \le C \|\nabla^4 u\|_{L^2} \varepsilon^{4\beta}, \tag{A.2}$$

where C is a constant that does not depend on u or ε. Now following the same argument as in [17] and keeping track of the derivatives of u, we have

$$G_\varepsilon u(x) = u(x) + \varepsilon\left[\omega(x)u(x) + \Delta u(x)\right] + R_u(x),$$

where

$$R_u(x) = T_u(x) + O\left(\left[\|u\|_{L^2} + \|\nabla u(x)\| + \|\nabla^2 u(x)\| + \|\nabla^3 u(x)\|\right]\varepsilon^2\right).$$

Applying G_ε to $u\sqrt{\kappa}$, we have

$$G_\varepsilon(u\sqrt{\kappa}) = u\sqrt{\kappa} + \varepsilon\left[\omega u\sqrt{\kappa} + \Delta(u\sqrt{\kappa})\right] + R_{u\sqrt{\kappa}}. \tag{A.3}$$

By expanding the derivatives of $u\sqrt{\kappa}$ and bounding them by the ∞-norms of κ and its derivatives, we have

$$R_{u\sqrt{\kappa}}(x) = T_{u\sqrt{\kappa}}(x) + O\left(\|\sqrt{\kappa}\|_{\mathcal{C}^4}\left[\|u\|_{L^2} + |u(x)| + \|\nabla u(x)\| + \|\nabla^2 u(x)\| + \|\nabla^3 u(x)\|\right]\varepsilon^2\right),$$

where

$$\|T_{u\sqrt{\kappa}}\|_{L^2} \leq C\|\sqrt{\kappa}\|_{\mathcal{C}^4}\|\nabla^4 u\|_{L^2}\varepsilon^{4\beta}.$$

By setting $u = 1$, we have

$$G_\varepsilon\sqrt{\kappa} = \sqrt{\kappa} + \varepsilon\left[\omega\sqrt{\kappa} + \Delta\sqrt{\kappa}\right] + O\left(\|\sqrt{\kappa}\|_{\mathcal{C}^4}\varepsilon^2\right). \tag{A.4}$$

By combining (A.3) and (A.4), we have

$$\mathcal{L}_\varepsilon^\kappa u = \frac{\sqrt{\kappa}}{\varepsilon}\left[uG_\varepsilon\sqrt{\kappa} - G_\varepsilon(u\sqrt{\kappa})\right] = \mathcal{L}^\kappa u + O\left(\sqrt{\kappa}u\|\sqrt{\kappa}\|_{\mathcal{C}^4}\varepsilon\right) + \frac{\sqrt{\kappa}R_{u\sqrt{\kappa}}}{\varepsilon}.$$

Hence it follows that

$$\|(\mathcal{L}_\varepsilon^\kappa - \mathcal{L}^\kappa)u\|_{L^2} \leq C\|u\|_{H^4}\|\sqrt{\kappa}\|_{\mathcal{C}^4}^2\varepsilon^{4\beta-1}.$$

Proof of Lemma 3.5. First we multiply the equation by u and integrate over $\mathcal{M}$. Integrating by parts, we get

$$\int fu = -\int \operatorname{div}(\kappa\nabla u)u = \int \kappa|\nabla u|^2 \geq \kappa_{\min}\|\nabla u\|_{L^2}^2.$$

By Hölder and Poincaré inequalities, there is a constant C that depends only on $\mathcal{M}$ so that

$$\|\nabla u\|_{L^2} \leq C\kappa_{\min}^{-1}\|f\|_{L^2} \leq C\kappa_{\min}^{-1}\|f\|_{H^3}, \tag{A.5}$$

$$\|u\|_{L^2} \leq C\kappa_{\min}^{-1}\|f\|_{L^2} \leq C\kappa_{\min}^{-1}\|f\|_{H^3}. \tag{A.6}$$

Now differentiating the equation with respect to x_k and testing against u_{x_k}, we get

$$\begin{aligned}\int f_{x_k}u_{x_k} = -\int \operatorname{div}\left(\frac{\partial}{\partial x_k}(\kappa\nabla u)\right)u_{x_k} &= \int \frac{\partial}{\partial x_k}(\kappa\nabla u)\cdot\nabla u_{x_k} \\ &= \int \kappa_{x_k}\nabla u\cdot\nabla u_{x_k} + \kappa\nabla u_{x_k}\cdot\nabla u_{x_k}.\end{aligned}$$

Using Young's inequality that $|ab| \leq \varepsilon a^2 + \frac{1}{4\varepsilon}b^2$ we get

$$\begin{aligned}\int f_{x_k}u_{x_k} &\geq -\|\kappa_{x_k}\|_\infty\left(\varepsilon\|\nabla u_{x_k}\|_{L^2}^2 + \frac{1}{4\varepsilon}\|\nabla u\|_{L^2}^2\right) + \kappa_{\min}\|\nabla u_{x_k}\|_{L^2}^2 \\ &= \left(\kappa_{\min} - \varepsilon\|\kappa_{x_k}\|_\infty\right)\|\nabla u_{x_k}\|_{L^2}^2 - \frac{\|\kappa_{x_k}\|_\infty}{4\varepsilon}\|\nabla u\|_{L^2}^2.\end{aligned}$$

Choosing $\varepsilon = \frac{\kappa_{\min}}{2(\|\kappa_{x_k}\|_\infty+1)}$ and rearranging terms, we get

$$\begin{aligned}\|\nabla u_{x_k}\|_{L^2}^2 &\leq \kappa_{\min}^{-2}\|\kappa_{x_k}\|_\infty\left(\|\kappa_{x_k}\|_\infty + 1\right)\|\nabla u\|_{L^2}^2 + 2\kappa_{\min}^{-1}\int f_{x_k}u_{x_k} \\ &\leq \kappa_{\min}^{-2}\|\kappa_{x_k}\|_\infty\left(\|\kappa_{x_k}\|_\infty + 1\right)\|\nabla u\|_{L^2}^2 + \kappa_{\min}^{-1}\|f_{x_k}\|_{L^2}^2 + \kappa_{\min}^{-1}\|u_{x_k}\|_{L^2}^2.\end{aligned}$$

Then we have

$$\|\nabla^2 u\|_{L^2}^2 = \sum_{k=1}^{m} \|\nabla u_{x_k}\|_{L^2}^2 \leq m\kappa_{\min}^{-2}\big(\|\nabla\kappa\|_\infty^2 + \|\nabla\kappa\|_\infty\big)\|\nabla u\|_{L^2}^2 + \kappa_{\min}^{-1}\|\nabla f\|_{L^2}^2 + \kappa_{\min}^{-1}\|\nabla u\|_{L^2}^2$$

$$\leq C\|f\|_{H^3}^2\left[\kappa_{\min}^{-3} + \kappa_{\min}^{-4}\big(\|\kappa\|_{\mathcal{C}^3}^2 + \|\kappa\|_{\mathcal{C}^3}\big)\right], \tag{A.7}$$

where we have used (A.5) and C only depends on $\mathcal{M}$. Moreover, we have used the fact that $\kappa_{\min} \geq e^{-\|u\|_\infty}$, which implies $\kappa_{\min}^{-n_1} \leq \kappa_{\min}^{-n_2}$ if $n_1 \leq n_2$. Now we bound the norm of the third derivatives by further differentiating the equation with respect to x_j and integrating against $u_{x_k x_j}$. We have again by Cauchy's inequality

$$\begin{aligned}
\int f_{x_k x_j} u_{x_k x_j} &= -\int \operatorname{div}\left(\frac{\partial^2}{\partial x_j \partial x_k}(\kappa\nabla u)\right) u_{x_k x_j} \\
&= \int \frac{\partial^2}{\partial x_j \partial x_k}(\kappa\nabla u)\cdot\nabla u_{x_k x_j} \\
&= \int \left[\kappa_{x_k x_j}\nabla u + \kappa_{x_k}\nabla u_{x_j} + \kappa_{x_j}\nabla u_{x_k} + \kappa\nabla u_{x_k x_j}\right]\cdot\nabla u_{x_k x_j} \\
&\geq \kappa_{\min}\|\nabla u_{x_k x_j}\|_{L^2}^2 - \|\kappa_{x_k x_j}\|_\infty\left(\varepsilon_1\|\nabla u_{x_k x_j}\|_{L^2}^2 + \frac{1}{4\varepsilon_1}\|\nabla u\|_{L^2}^2\right) \\
&\quad -\|\kappa_{x_k}\|_\infty\left(\varepsilon_2\|\nabla u_{x_k x_j}\|_{L^2}^2 + \frac{1}{4\varepsilon_2}\|\nabla u_{x_j}\|_{L^2}^2\right) - \|\kappa_{x_j}\|_\infty\left(\varepsilon_3\|\nabla u_{x_k x_j}\|_{L^2}^2 + \frac{1}{4\varepsilon_3}\|\nabla u_{x_k}\|_{L^2}^2\right) \\
&= \|\nabla u_{x_k x_j}\|_{L^2}^2\left(\kappa_{\min} - \varepsilon_1\|\kappa_{x_k x_j}\|_\infty - \varepsilon_2\|\kappa_{x_k}\|_\infty - \varepsilon_3\|\kappa_{x_j}\|_\infty\right) \\
&\quad - \frac{1}{4\varepsilon_1}\|\kappa_{x_k x_j}\|_\infty\|\nabla u\|_{L^2}^2 - \frac{1}{4\varepsilon_2}\|\kappa_{x_k}\|_\infty\|\nabla u_{x_j}\|_{L^2}^2 - \frac{1}{4\varepsilon_3}\|\kappa_{x_j}\|_\infty\|\nabla u_{x_k}\|_{L^2}^2.
\end{aligned}$$

Now choosing

$$\varepsilon_1 = \frac{\kappa_{\min}}{4(\|\kappa_{x_k x_j}\|_\infty + 1)}, \quad \varepsilon_2 = \frac{\kappa_{\min}}{4(\|\kappa_{x_k}\|_\infty + 1)}, \quad \varepsilon_3 = \frac{\kappa_{\min}}{4(\|\kappa_{x_j}\|_\infty + 1)}$$

and rearranging terms, we get

$$\begin{aligned}
\|\nabla u_{x_k x_j}\|_{L^2}^2 \leq{}& 2\kappa_{\min}^{-1}\left(\|f_{x_k x_j}\|_{L^2}^2 + \|u_{x_k x_j}\|_{L^2}^2\right) + 4\kappa_{\min}^{-2}\|\kappa_{x_k x_j}\|_\infty\big(\|\kappa_{x_k x_j}\|_\infty + 1\big)\|\nabla u\|_{L^2}^2 \\
&+ 4\kappa_{\min}^{-2}\|\kappa_{x_k}\|_\infty\big(\|\kappa_{x_k}\|_\infty + 1\big)\|\nabla u_{x_j}\|_{L^2}^2 + 4\kappa_{\min}^{-2}\|\kappa_{x_j}\|_\infty\big(\|\kappa_{x_j}\|_\infty + 1\big)\|\nabla u_{x_k}\|_{L^2}^2.
\end{aligned}$$

Then we have

$$\begin{aligned}
\|\nabla^3 u\|_{L^2}^2 &= \sum_{j=1}^{m}\sum_{k=1}^{m}\|\nabla u_{x_k x_j}\|_{L^2}^2 \\
&\leq 2\kappa_{\min}^{-1}\left(\|\nabla^2 f\|_{L^2}^2 + \|\nabla^2 u\|_{L^2}^2\right) + 4m^2\kappa_{\min}^{-2}\big(\|\nabla^2\kappa\|_\infty^2 + \|\nabla^2\kappa\|_\infty\big)\|\nabla u\|_{L^2}^2 \\
&\quad + 8m\kappa_{\min}^{-2}\big(\|\nabla\kappa\|_\infty^2 + \|\nabla\kappa\|_\infty\big)\|\nabla^2 u\|_{L^2}^2 \qquad \text{(A.8)} \\
&\leq C\|f\|_{H^3}^2\left[\kappa_{\min}^{-4} + \kappa_{\min}^{-5}\big(\|\kappa\|_{\mathcal{C}^3}^2 + \|\kappa\|_{\mathcal{C}^3}\big) + \kappa_{\min}^{-6}\big(\|\kappa\|_{\mathcal{C}^3}^2 + \|\kappa\|_{\mathcal{C}^3}\big)^2\right], \qquad \text{(A.9)}
\end{aligned}$$

where again we only keep the highest order in $\kappa_{\min}$. Differentiating further and applying a similar argument, gives that

$$\begin{aligned}\|\nabla^4 u\|_{L^2}^2 &\le 4\kappa_{\min}^{-1}\big(\|\nabla^3 f\|_{L^2}^2 + \|\nabla^3 u\|_{L^2}^2\big) + 16m^3\kappa_{\min}^{-2}\|\nabla u\|_{L^2}^2\big(\|\nabla^3\kappa\|_\infty^2 + \|\nabla^3\kappa\|_\infty\big)\\ &\quad + 48m^2\kappa_{\min}^{-2}\|\nabla^2 u\|_{L^2}^2\big(\|\nabla^2\kappa\|_\infty^2 + \|\nabla^2\kappa\|_\infty\big) + 48m\kappa_{\min}^{-2}\|\nabla^3 u\|_{L^2}^2\big(\|\nabla\kappa\|_\infty^2 + \|\nabla\kappa\|_\infty\big)\\ &\le C\|f\|_{H^3}^2\Big[\kappa_{\min}^{-5} + \kappa_{\min}^{-6}\big(\|\kappa\|_{\mathcal{C}^3}^2 + \|\kappa\|_{\mathcal{C}^3}\big) + \kappa_{\min}^{-7}\big(\|\kappa\|_{\mathcal{C}^3}^2 + \|\kappa\|_{\mathcal{C}^3}\big)^2 + \kappa_{\min}^{-8}\big(\|\kappa\|_{\mathcal{C}^3}^2 + \|\kappa\|_{\mathcal{C}^3}\big)^3\Big].\end{aligned} \tag{A.10}$$

The desired result follows by combining equations (A.6), (A.5), (A.7), (A.9), and (A.10).

Proof of Lemma 3.7. By Lipschitz continuity of e^{-x} when $x > 0$, we have

$$\begin{aligned}&\left|\exp\left(-\frac{1}{2}|y-\mathcal{G}_\varepsilon(\theta)|_\Gamma^2\right) - \exp\left(-\frac{1}{2}|y-\mathcal{G}(\theta)|_\Gamma^2\right)\right|\\ &\le \left|\frac{1}{2}|y-\mathcal{G}_\varepsilon(\theta)|_\Gamma^2 - \frac{1}{2}|y-\mathcal{G}(\theta)|_\Gamma^2\right|\\ &= \frac{1}{2}\left|(\mathcal{G}(\theta)^T + \mathcal{G}_\varepsilon(\theta)^T)\Gamma^{-1}(\mathcal{G}(\theta)-\mathcal{G}_\varepsilon(\theta)) + 2y^T\Gamma^{-1}(\mathcal{G}(\theta)-\mathcal{G}_\varepsilon(\theta))\right|\\ &\le \frac{1}{2}\|\Gamma^{-1}\|_2\Big(|\mathcal{G}(\theta)| + |\mathcal{G}_\varepsilon(\theta)|\Big)|\mathcal{G}(\theta)-\mathcal{G}_\varepsilon(\theta)| + \|\Gamma^{-1}\|_2|y||\mathcal{G}(\theta)-\mathcal{G}_\varepsilon(\theta)|,\end{aligned}$$

where $\|\Gamma^{-1}\|_2$ is the operator 2-norm of Γ^{-1}. By Theorem 3.1, Lemma 3.2, and (A.6),

$$\begin{aligned}|\mathcal{G}(\theta)-\mathcal{G}_\varepsilon(\theta)| &\le \sqrt{\sum\|\ell_j\|^2}\|u-u_\varepsilon\|_{L^2} \le C\sqrt{\sum\|\ell_j\|^2}A(\theta)\|f\|_{H^3}\varepsilon^{4\beta-1},\\ |\mathcal{G}(\theta)| + |\mathcal{G}_\varepsilon(\theta)| &\le \sqrt{\sum\|\ell_j\|^2}\big(\|u\|_{L^2} + \|u_\varepsilon\|_{L^2}\big) \le C\sqrt{\sum\|\ell_j\|^2}e^{\|\theta\|_\infty}\|f\|_{L^2},\end{aligned}$$

where u and u_ε are the zero-mean solutions associated with θ, and $\|\ell_j\|$ is the operator norm of ℓ_j. Here we have written A as a function of θ and use the fact that $\kappa_{\min} \ge e^{-\|\theta\|_\infty}$. So

$$\begin{aligned}&\int\left|\exp\left(-\frac{1}{2}|y-\mathcal{G}_\varepsilon(\theta)|_\Gamma^2\right) - \exp\left(-\frac{1}{2}|y-\mathcal{G}(\theta)|_\Gamma^2\right)\right|d\pi(\theta)\\ &\le C\|\Gamma^{-1}\|_2\varepsilon^{4\beta-1}\left[\sum\|\ell_j\|^2\|f\|_{H^3}^2\int e^{\|\theta\|_\infty}A(\theta)d\pi(\theta) + \sqrt{\sum\|\ell_j\|^2}\|y\|\|f\|_{H^3}\int A(\theta)d\pi(\theta)\right].\end{aligned}$$

It now suffices to show $\int\left(e^{\|\theta\|_\infty}\vee 1\right)A(\theta)d\pi(\theta) < \infty$. Since $\kappa = e^\theta$, we have

$$\|\kappa\|_{C^4} \le Ce^{\|\theta\|_\infty}\left(\|\theta\|_{\mathcal{C}^4} + \|\theta\|_{\mathcal{C}^4}^2 + \|\theta\|_{\mathcal{C}^4}^3 + \|\theta\|_{\mathcal{C}^4}^4\right),$$

where C is a constant depending on the dimension m and a similar relation is true for $\|\sqrt{\kappa}\|_{\mathcal{C}^4}$. Keeping only the highest order term in $e^{\|\theta\|_\infty}$, we have

$$\begin{aligned}A(\theta) &\le C\sqrt{P_1(\|\theta\|_{\mathcal{C}^4})e^{14\|\theta\|_\infty}}P_2(\|\theta\|_{\mathcal{C}^4})e^{\|\theta\|_\infty}\\ &\le C\sqrt{P_1(\|\theta\|_{\mathcal{C}^4})e^{14\|\theta\|_{\mathcal{C}^4}}}P_2(\|\theta\|_{\mathcal{C}^4})e^{\|\theta\|_{\mathcal{C}^4}},\end{aligned}$$

where P_1 and P_2 are polynomials. Since π is a Gaussian measure on $\mathcal{C}^4$, by Fernique's theorem [28],

$$\int \left(e^{\|\theta\|_\infty} \vee 1\right) A(\theta) d\pi(\theta) < \infty.$$

It follows that

$$\int \left| \exp\left(-\frac{1}{2}|y - \mathcal{G}_\varepsilon(\theta)|_\Gamma^2\right) - \exp\left(-\frac{1}{2}|y - \mathcal{G}(\theta)|_\Gamma^2\right) \right| d\pi(\theta) \leq C\varepsilon^{4\beta-1},$$

where C depends on Γ, y, f, and the ℓ_j's, but is independent of ε.

REFERENCES

[1] S. Agapiou, O. Papaspiliopoulos, D. Sanz-Alonso, and A. M. Stuart, *Importance sampling: Intrinsic dimension and computational cost*, Statist. Sci., 32 (2017), pp. 405–431.

[2] I. Babuska, R. Tempone, and G. E. Zouraris, *Galerkin finite element approximations of stochastic elliptic partial differential equations*, SIAM J. Numer. Anal., 42 (2004), pp. 800–825.

[3] M. Belkin and P. Niyogi, *Semi-supervised learning on Riemannian manifolds*, Mach. Learn., 56 (2004), pp. 209–239.

[4] M. Belkin and P. Niyogi, *Towards a theoretical foundation for Laplacian-based manifold methods*, in Learning Theory, Lecture Notes in Comput. Sci. 3559, Springer, Berlin, 2005, pp. 486–500.

[5] T. Berry and J. Harlim, *Variable bandwidth diffusion kernels*, Appl. Comput. Harmon. Anal., 40 (2016), pp. 68–96.

[6] T. Berry and T. Sauer, *Local kernels and the geometric structure of data*, Appl. Comput. Harmon. Anal., 40 (2016), pp. 439–469.

[7] M. Bertalmio, L.-T. Cheng, S. Osher, and G. Sapiro, *Variational problems and partial differential equations on implicit surfaces*, J. Comput. Phys., 174 (2001), pp. 759–780.

[8] A. L. Bertozzi, X. Luo, A. M. Stuart, and K. C. Zygalakis, *Uncertainty quantification in graph-based classification of high dimensional data*, SIAM/ASA J. Uncertain. Quantif., 6 (2018), pp. 568–595.

[9] D. Bigoni, O. Zahm, A. Spantini, and Y. Marzouk, *Greedy Inference with Layers of Lazy Maps*, preprint, https://arxiv.org/abs/1906.00031 (2019).

[10] V. I. Bogachev, *Gaussian Measures*, Math. Surveys Monogr. 62, American Mathematical Society, Providence, RI, 1998.

[11] A. Bonito, J. M. Cascón, K. Mekchay, P. Morin, and R. H. Nochetto, *High-order AFEM for the Laplace–Beltrami operator: Convergence rates*, Found. Comput. Math., 16 (2016), pp. 1473–1539.

[12] D. Burago, S. Ivanov, and Y. Kurylev, *A graph discretization of the Laplace-Beltrami operator*, J. Spectr. Theory, 4 (2014), pp. 675–714.

[13] D. Calvetti and E. Somersalo, *An Introduction to Bayesian Scientific Computing: Ten Lectures on Subjective Computing*, Surv. Tutor. Appl. Math. Sci. 2, Springer, Berlin, 2007.

[14] F. Camacho and A. Demlow, *L2 and pointwise a posteriori error estimates for FEM for elliptic PDEs on surfaces*, IMA J. Numer. Anal., 35 (2015), pp. 1199–1227.

[15] M. A. Chaplain, M. Ganesh, I. G. Graham, and G. Lolas, *Mathematical modelling of solid tumour growth: Applications of pre-pattern formation*, in Morphogenesis and Pattern Formation in Biological Systems, Springer, Tokyo, 2003, pp. 283–293.

[16] A. Cohen, R DeVore, and C. Schwab, *Analytic regularity and polynomial approximation of parametric and stochastic elliptic PDEs*, Anal. Appl., 9 (2011), pp. 11–47.

[17] R. R. Coifman and S. Lafon, *Diffusion maps*, Appl. Comput. Harmon. Anal., 21 (2006), pp. 5–30.

[18] R. R. Coifman, S. Lafon, A. B. Lee, M. Maggioni, B. Nadler, F. Warner, and S. W. Zucker, *Geometric diffusions as a tool for harmonic analysis and structure definition of data: Diffusion maps*, Proc. Natl. Acad. Sci. USA, 102 (2005), pp. 7426–7431.

[19] R. R. Coifman, Y. Shkolnisky, F. J. Sigworth, and A. Singer, *Graph Laplacian tomography from unknown random projections*, IEEE Trans. Image Process., 17 (2008), pp. 1891–1899.
[20] S. Cotter, M. Dashti, and A. M. Stuart, *Approximation of Bayesian inverse problems for PDEs*, SIAM J. Numer. Anal., 48 (2010), pp. 322–345.
[21] S. Cotter, G. O. Roberts, A. M. Stuart, and D. White, *MCMC methods for functions: modifying old algorithms to make them faster*, Statist. Sci., 48 (2013), pp. 424–446.
[22] K. Crane, *Keenans 3D Model Repository*, http://www.cs.cmu.edu/~kmcrane/Projects/ModelRepository
[23] M. Dashti and A. M. Stuart, *Bayesian approach to inverse problems*, in Handbook of Uncertainty Quantification, Springer, Cham, Switzerland, 2017, pp. 311–428.
[24] M. M. Dunlop, M. A. Iglesias, and A. M. Stuart, *Hierarchical Bayesian level set inversion*, Statist. Comput., 27 (2017), pp. 1555–1584.
[25] G. Dziuk and C. M. Elliott, *Finite element methods for surface PDEs*, Acta Numer., 22 (2013), pp. 289–396.
[26] C. Eilks and C. M. Elliott, *Numerical simulation of dealloying by surface dissolution via the evolving surface finite element method*, J. Comput. Phys., 227 (2008), pp. 9727–9741.
[27] C. M. Elliott and B. Stinner, *A surface phase field model for two-phase biological membranes*, SIAM J. Appl. Math., 70 (2010), pp. 2904–2928.
[28] X. Fernique, *Intégrabilité des vecteurs Gaussiens*, C. R. Acad. Sci. Paris Ser. A, 270 (1970), pp. 1698–1699.
[29] P. Frauenfelder, C. Schwab, and R. A. Todor, *Finite elements for elliptic problems with stochastic coefficients*, Comput. Methods Appl. Mech. Engrg., 194 (2005), pp. 205–228.
[30] T. Gao, S. Z. Kovalsky, and I. Daubechies, *Gaussian process landmarking on manifolds*, SIAM J. Math. Data Sci., 1 (2019), pp. 208–236.
[31] N. Garcia Trillos, Z. Kaplan, T. Samakhoana, and D. Sanz-Alonso, *On the Consistency of Graph-Based Bayesian Learning and the Scalability of Sampling Algorithms*, preprint, https://arxiv.org/abs/1710.07702 (2017).
[32] N. Garcia Trillos and D. Sanz-Alonso, *The Bayesian formulation and well-posedness of fractional elliptic inverse problems*, Inverse Problems, 33 (2017), 065006.
[33] N. Garcia Trillos and D. Sanz-Alonso, *Continuum limits of posteriors in graph Bayesian inverse problems*, SIAM J. Math. Anal., 50 (2018), pp. 4020–4040.
[34] N. Garcia Trillos, D. Sanz-Alonso, and R. Yang, *Local regularization of noisy point clouds: Improved global geometric estimates and data analysis*, J. Mach. Learn. Res., 20 (2019), pp. 1–37.
[35] C. J. Geoga, M. Anitescu, and M. L. Stein, *Scalable Gaussian process computations using hierarchical matrices*, J. Comput. Graph. Statist., 29 (2020), pp. 227–237.
[36] F. Gilani and J. Harlim, *Approximating solutions of linear elliptic PDE's on a smooth manifold using local kernel*, J. Comput. Phys., 395 (2019), pp. 563–582.
[37] D. Gilbarg and N. S. Trudinger, *Elliptic Partial Differential Equations of Second Order*, Springer, Berlin, 2015.
[38] M. A. Iglesias and C. Dawson, *The representer method for state and parameter estimation in single-phase Darcy flow*, Comput. Methods Appl. Mech. Engrg., 196 (2007), pp. 4577–4596.
[39] J. Kaipio and E. Somersalo, *Statistical and Computational Inverse Problems*, Appl. Math.. Sci. 160, Springer, New York, 2006.
[40] Z. Li and Z. Shi, *A convergent point integral method for isotropic elliptic equations on a point cloud*, Multiscale Model. Simul., 14 (2016), pp. 874–905.
[41] Z. Li, Z. Shi, and J. Sun, *Point integral method for solving Poisson-type equations on manifolds from point clouds with convergence guarantees*, Commun. Comput. Phys., 22 (2017), pp. 228–258.
[42] F. Lindgren, H. Rue, and J. Lindström, *An explicit link between Gaussian fields and Gaussian Markov random fields: The stochastic partial differential equation approach*, J. R. Stat. Soc. Ser. B Methodol., 73 (2011), pp. 423–498.
[43] R. J. Lorentzen, K. M. Flornes, and G. Nævdal, *History matching channelized reservoirs using the ensemble Kalman filter*, SPE Journal, 17 (2012), pp. 137–151.
[44] Y. Marzouk and D. Xiu, *A stochastic collocation approach to Bayesian inference in inverse problems*, Commun. Comput. Phys., 6 (2009), pp. 826–847.

[45] D. McLaughlin and L. R. Townley, *A reassessment of the groundwater inverse problem*, Water Res. Res., 32 (1996), pp. 1131–1161.
[46] F. Mémoli, G. Sapiro, and P. Thompson, *Implicit brain imaging*, NeuroImage, 23 (2004), pp. S179–S188.
[47] J. L. Mueller and S. Siltanen, *Linear and Nonlinear Inverse Problems with Practical Applications*, Comput. Sci. Eng. 10, SIAM, Philadelphia, 2012.
[48] J. Ping and D. Zhang, *History matching of channelized reservoirs with vector-based level-set parameterization*, Spe Journal, 19 (2014), pp. 514–529.
[49] C. Piret, *The orthogonal gradients method: A radial basis functions method for solving partial differential equations on arbitrary surfaces*, J. Comput. Phys., 231 (2012), pp. 4662–4675.
[50] S. J. Ruuth and B. Merriman, *A simple embedding method for solving partial differential equations on surfaces*, J. Comput. Phys., 227 (2008), pp. 1943–1961.
[51] D. Sanz-Alonso, *Importance sampling and necessary sample size: An information theory approach*, SIAM/ASA J. Uncertain. Quantif., 6 (2018), pp. 867–879.
[52] D. Sanz-Alonso, A. M. Stuart, and A. Taeb, *Inverse Problems and Data Assimilation*, preprint, https://arxiv.org/abs/1810.06191 (2018).
[53] D. Sanz-Alonso and R. Yang, *The SPDE Approach to Matérn Fields: Graph Representations*, preprint, https://arxiv.org/abs/2004.08000 (2020).
[54] Z. Shi and J. Sun, *Convergence of the Point Integral Method for Poisson Equation on Point Cloud*, preprint, https://arxiv.org/abs/1403.2141 (2014).
[55] M. L. Stein, *Interpolation of Spatial Data: Some Theory for Kriging*, Springer, New York, 1999.
[56] A. M. Stuart, *Inverse problems: A Bayesian perspective*, Acta Numer., 19 (2010), pp. 451–559.
[57] E. H. Thiede, D. Giannakis, A. R. Dinner, and J. Weare, *Galerkin approximation of dynamical quantities using trajectory data*, J. Chem. Phys., 150 (2019), 244111.
[58] A. Tsybakov, *Introduction to Nonparametric Estimation*, Springer Ser. Statist., Springer, New York, 2008.
[59] U. Von Luxburg, *A tutorial on spectral clustering*, Statist. Comput., 17 (2007), pp. 395–416.
[60] L. Wasserman, *All of Nonparametric Statistics*, Springer, New York, 2006.
[61] J.-J. Xu, Z. Li, J. Lowengrub, and H. Zhao, *A level-set method for interfacial flows with surfactant*, J. Comput. Phys., 212 (2006), pp. 590–616.
[62] L. Zelnik-Manor and P. Perona, *Self-tuning spectral clustering*, in Advances in Neural Information Processing Systems, MIT Press, Cambridge, MA, 2005, pp. 1601–1608.