IOP Publishing Inverse Problems

Inverse Problems **38** (2022) 035006 (31pp) https://doi.org/10.1088/1361-6420/ac3994

Graph-based prior and forward models for inverse problems on manifolds with boundaries

John Harlim[1], Shixiao W Jiang[2], Hwanwoo Kim[3,*] and Daniel Sanz-Alonso[3]

[1] Department of Mathematics, Department of Meteorology and Atmospheric Science, Institute for Computational and Data Sciences, The Pennsylvania State University, University Park, PA 16802, United States of America
[2] Institute of Mathematical Sciences, ShanghaiTech University, Shanghai 201210, People's Republic of China
[3] Department of Statistics, University of Chicago, Chicago, IL 60637, United States of America

E-mail: hwkim@uchicago.edu

Received 13 June 2021, revised 10 November 2021
Accepted for publication 15 November 2021
Published 31 January 2022

Abstract

This paper develops manifold learning techniques for the numerical solution of PDE-constrained Bayesian inverse problems on manifolds with boundaries. We introduce graphical Matérn-type Gaussian field priors that enable flexible modeling near the boundaries, representing boundary values by superposition of harmonic functions with appropriate Dirichlet boundary conditions. We also investigate the graph-based approximation of forward models from PDE parameters to observed quantities. In the construction of graph-based prior and forward models, we leverage the ghost point diffusion map algorithm to approximate second-order elliptic operators with classical boundary conditions. Numerical results validate our graph-based approach and demonstrate the need to design prior covariance models that account for boundary conditions.

Keywords: Bayesian inverse problem, manifold with boundary, ghost point diffusion map, graph-based prior

(Some figures may appear in colour only in the online journal)

*Author to whom any correspondence should be addressed.

1. Introduction

This paper develops manifold learning techniques to address two tasks in the numerical solution of PDE-constrained Bayesian inverse problems on manifolds with boundaries: (1) the design and approximation of Gaussian field priors for spatially-distributed PDE parameters; and (2) the approximation of forward maps from PDE parameters to PDE solutions. We introduce graph-based approximations of prior and forward models and numerically show their effectiveness in two test problems. The first one concerns the recovery of the diffusion coefficient of an elliptic PDE from pointwise noisy measurements of the PDE solution; the second one concerns the recovery of the initial condition of a heat equation from noisy measurements of the PDE solution at some positive time. Both of these problems have been widely used as test cases for Bayesian inversion on manifolds and Euclidean domains [6, 13, 21, 23, 25, 26, 44], but previous work has largely ignored the boundary effects that are the focus of this paper. The applied significance of elliptic and heat inverse problems is exemplified by [1, 52] and references therein. PDEs on manifolds arise in many applications, including granular flow [36], liquid crystals [47], biomembranes [20], computer graphics [3, 32], and brain imaging [34].

In the Bayesian approach to inverse problems [28, 44, 45], overviewed in section 2 below, inference on the PDE parameters is performed using a posterior distribution obtained by conditioning a user-chosen prior distribution to observed data. When the parameter of interest is a function, employing an adequate choice of prior is crucial: the prior determines the support of the posterior, and hence the space of parameters that can be recovered given sufficient data. In this work, we will focus on Matérn-type Gaussian field priors [33, 43] that have been widely adopted in inverse problems, statistics and machine learning due to their flexibility and computational efficiency, e.g. [7–10, 31, 37, 49]. We consider Matérn-type models on manifolds [31] and investigate their discretization using graphs, as well as their generalization to enhance flexibility near the boundaries. To our knowledge, all previous work on graph-based Matérn models, e.g. [4, 25, 39, 40], disregarded boundary effects, despite their importance in the discretization of Gaussian field priors in Euclidean settings [11, 17, 29]. Our numerical results will confirm that accounting for boundary conditions is important in the design of priors and in their graph-based approximation. While our emphasis is on PDE-constrained inverse problems, we expect that the boundary-aware graph-based Matérn priors that we introduce will also find applications in graph-based machine learning [40].

Approximation of the forward map requires solving PDEs on manifolds with a range of PDE parameters. There are numerous techniques to solve PDEs on manifolds, including the finite element method [19], the level set method [3], closest point representation [38], and the mesh-free radial basis function method [22]. In contrast to all of these methods, the kernel approach we will consider avoids the need to have some parameterization on the manifolds; we refer to [27] for a detailed discussion of the advantages and disadvantages of these PDE solvers. In this work, we introduce graph-based forward map approximations for elliptic and heat inversion on manifolds with boundaries. Our graph-based approach leverages the ghost point diffusion map (GPDM) algorithm [27] to approximate second-order elliptic operators with classical boundary conditions, which are used to define elliptic and heat forward models, and also the Matérn prior covariance. The main idea behind the GPDM algorithm is to extend the domain of the PDE through a set of artificially constructed *ghost points*. By extending the underlying manifold where the PDE was defined, one can treat the boundary of the original manifold as the interior of the extended manifold. The implication is that the kernel-based approximations that underpin the design of the GPDM method remain to be valid close to

the boundary. The GPDM algorithm has been generalized to solve time-dependent advection-diffusion equations on manifolds [50].

We close this introduction with an outline of the paper and a summary of our contributions.

- In section 2, we overview the Bayesian approach to inverse problems and provide a high-level summary of the proposed procedure. We also review the Bayesian approach in [26] for elliptic inversion on a closed (i.e. boundary-free and compact) manifold, and we lay out the forward map discretization for the heat inverse problem. We present elliptic and heat inverse problems in a parallel way, emphasizing that both can be treated in the same way under the proposed approach except for the necessary distinction in the discretization of the forward map.
- In section 3, we review the GPDM algorithm and introduce our boundary-aware Matérn-type priors and forward map approximations for elliptic and heat inversion.
- In section 4, we showcase implementations of the proposed methodology for elliptic and heat inversion on one and two-dimensional manifolds with boundaries. In addition, we demonstrate the enhanced flexibility of our proposed prior relative to existing graph-based approaches that ignore boundary effects.
- In section 5, we provide conclusions and some open directions for future work.

2. Background

In this section, we first give the necessary background on the Bayesian formulation of inverse problems [28, 42, 45] and the function space perspective [18, 44]. Next, we formulate the problem of interest, namely, PDE-constrained Bayesian inversion on manifolds. Lastly, we review the graph-based approach in [26] which is only applicable for closed manifolds.

2.1. Bayesian formulation of inverse problems

Suppose we have a parameter of interest $\theta \in \Theta$, observed data $y \in \mathbb{R}^M$ and a forward model $\mathcal{G}$, satisfying

$$y = \mathcal{G}(\theta) + \eta, \tag{2.1}$$

where $\eta \in \mathbb{R}^M$ is an observation noise. Our goal is to study the inverse problem of recovering θ from the observed data y. In the Bayesian framework, one endows probabilistic structure to any unknown quantity. Here this involves specifying a *prior* distribution μ for the parameter θ and a distribution ρ for the noise η. To facilitate our presentation, we assume throughout that $\eta \sim \rho = \mathcal{N}(0, \Gamma)$, where $\Gamma \in \mathbb{R}^{M\times M}$ is a given positive-definite covariance matrix. We also assume that θ and η are independent random variables. The solution of the inverse problem under the Bayesian framework is then the *posterior* distribution of θ given y, denoted μ^y. The posterior distribution allows to not only construct point estimates of the parameter of interest but also quantify the uncertainty in the parameter reconstruction. In practice, this often requires to resort to sampling techniques such as Markov chain Monte Carlo (MCMC).

If the parameter space is finite-dimensional, i.e. $\Theta \subseteq \mathbb{R}^N$, and the prior μ has Lebesgue density π, then the posterior μ^y has Lebesgue-density π^y given by

$$\pi^y(\theta) = \frac{1}{Z}\rho\left(y - \mathcal{G}(\theta)\right)\pi(\theta), \quad Z := \int \rho\left(y - \mathcal{G}(\theta)\right)\pi(\theta)\,\mathrm{d}\theta, \tag{2.2}$$

where $\rho\left(y - \mathcal{G}(u)\right)$ is the data *likelihood*. Under suitable assumptions [44], the characterization (2.2) of the posterior can be generalized to infinite dimensional parameter space Θ by writing

the posterior as a change of measure with respect to the prior:

$$\frac{\mathrm{d}\mu^y}{\mathrm{d}\mu}(\theta) = \frac{1}{Z}\,\rho\,(y - \mathcal{G}(\theta)) \propto \exp\left(-\Phi(\theta; y)\right), \tag{2.3}$$

where we have defined $\Phi(\theta; y) := \frac{1}{2}\|y - \mathcal{G}(\theta)\|_\Gamma^2$ and we set $\|\cdot\|_\Gamma := \|\Gamma^{-\frac{1}{2}}\cdot\|$.

2.2. General setting and approach

We are interested in the inverse problem of recovering a parameter function $\theta \in \Theta$ of a PDE defined on a manifold $\mathcal{M} \subset \mathbb{R}^D$ from noisy measurements of the PDE solution $u \in \mathcal{U}$ at given locations $\{\tilde{x}_m\}_{m=1}^M \subset \mathcal{M}$. Here both Θ and $\mathcal{U}$ are suitable function spaces. The data model (2.1) is therefore given by

$$y_m = u(\tilde{x}_m) + \eta_m, \quad m = 1, \ldots, M, \quad \eta = \{\eta_m\}_{m=1}^M \sim \mathcal{N}(0, \Gamma), \tag{2.4}$$

with the *forward model* $\mathcal{G} : \theta \mapsto (u(\tilde{x}_1), \ldots, u(\tilde{x}_M))$ defined as the composition of a *forward map* $\mathcal{F} : \theta \mapsto u$ from PDE input to PDE solution with an *observation map* $\mathcal{O} : u \mapsto (u(\tilde{x}_1), \ldots, u(\tilde{x}_M))$ from PDE solution to observed quantities. In this paper, we solely focus on pointwise observations and tacitly assume throughout that the solutions to the PDEs we consider can be evaluated pointwise. However, our methodology can deal with other types of observation map, e.g. defined by bounded linear functionals in L^2 [26]. We will discretize the forward map using a point cloud in $\mathcal{M}$, which is denoted by $\{x_n\}_{n=1}^N \supseteq \{\tilde{x}_m\}_{m=1}^M$, where $M \leqslant N$. Our approach for the computational implementation of the inverse problem is then summarized in the following four steps:

(a) **Prior specification**: specify a prior distribution for the infinite dimensional parameter $\theta \in \Theta$.

(b) **Prior and forward model** discretization:

1. Discretize using graph-based techniques the prior distribution μ to obtain μ_N, a prior distribution over $\theta_N := (\theta(x_1), \ldots, \theta(x_N)) \in \mathbb{R}^N$. Note that μ_N is a prior distribution over discrete functions θ_N defined on the point cloud $\{x_n\}_{n=1}^N$.
2. Discretize using graph-based techniques the given forward map $\mathcal{F} : \theta \to u$ to obtain $\mathcal{F}_N : \theta_N \mapsto u_N := \mathcal{F}_N(\theta_N) \in \mathbb{R}^N$, where u_N is an approximation to the solution of the PDE evaluated along the point cloud, i.e., $u_N \approx (u(x_1), \ldots, u(x_N))$. Furthermore, we define $\mathcal{G}_N(\theta_N) := (u_N(\tilde{x}_1), \ldots, u_N(\tilde{x}_M)) \in \mathbb{R}^M$ where $u_N(\tilde{x}_i)$ denotes the component of the vector $u_N \in \mathbb{R}^N$ that corresponds to $\tilde{x}_i \in \{x_n\}_{n=1}^N$.

(c) **Sampling**: use an MCMC algorithm to obtain samples from the posterior distribution over θ_N, given by

$$\frac{\mathrm{d}\mu_N^y}{\mathrm{d}\mu_N}(\theta_N) \propto \exp(-\Phi_N(\theta_N; y)), \quad \text{where } \Phi_N(\theta_N; y) := \frac{1}{2}\|y - \mathcal{G}_N(\theta_N)\|_\Gamma^2. \tag{2.5}$$

(d) **Interpolation**: if desired, extend the samples to functions on $\mathcal{M}$ with an interpolation map.

This four-step approach was introduced in [24, 25] and previous works on the discretization process include [26]. The primary focus of this paper is to contribute to the specification and discretization steps for PDE-constrained inverse problems on manifolds with boundaries.

Algorithm 1. Graph pCN.

Input: initial value $\theta_N^{(0)}$, number of samples J, parameter value $\zeta \in (0,1)$.
For $j = 1, \ldots, J$ **do**:

1. Propose $\tilde{\theta}_N^{(j)} = (1-\zeta^2)^{\frac{1}{2}}\theta_N^{(j)} + \zeta\,\xi_N^{(j)}$, where $\xi_N^{(j)} \sim \mathcal{N}(0, V_N)$.
2. Set

$$a(\theta_N^{(j)}, \tilde{\theta}_N^{(j)}) := \min\left\{1, \exp\left(\Phi_N(\theta_N^{(j)}; y) - \Phi_N(\tilde{\theta}_N^{(j)}; y)\right)\right\},$$

and let

$$\theta_N^{(j+1)} = \begin{cases} \tilde{\theta}_N^{(j)} & \text{with probability } a(\theta_N^{(j)}, \tilde{\theta}_N^{(j)}), \\ \theta_N^{(j)} & \text{with probability } 1 - a(\theta_N^{(j)}, \tilde{\theta}_N^{(j)}). \end{cases}$$

Output: samples of discrete functions $\{\theta_N^{(j)}\}_{j=1}^J$.

We propose (1) priors that accommodate for prescribed boundary conditions, (2) graph-based discretizations of such priors; and (3) graph-based discretizations of the forward PDEs supplemented with boundary conditions. These procedures will be introduced in section 3. In the remainder of this section, we address the sampling and interpolation steps, which are based on existing methodologies that are applicable in wide generality.

For the sampling step, we will use the graph pCN method [4, 25], summarized in algorithm 1. The covariance matrix V_N in the pCN proposal is tightly connected with the prior construction. In fact, this matrix will be precisely the covariance matrix of the discretized prior distribution. The parameter ζ controls the size of the proposed moves of the chain. For a large ζ value, we explore a wider region of the state space with higher number of rejections. The motivation for using the graph pCN method is that it shows robustness with respect to the level of discretization refinement determined by the value of N. We refer to [4, 5, 16, 25] for the theoretical and empirical justification of this robustness.

For the interpolation step, once we obtain finite-dimensional samples $\theta_N \in \mathbb{R}^N$ representing a parameter function evaluated on the point cloud $\{x_n\}_{n=1}^N$, we can extend the samples into functions on $\mathcal{M}$ using the K-NN interpolation map defined by

$$\theta(x) = \frac{1}{\mathsf{K}} \sum_{x_i \in N_{\mathsf{K}}(x)} \theta_N(x_i), \quad x \in \mathcal{M},\ x \notin \{x_n\}_{n=1}^N,$$

where $N_{\mathsf{K}}(x)$ is the set of K-nearest points in the point cloud $\{x_n\}_{n=1}^N$ to the point x. To find the K-nearest points, one can use the Euclidean distance in $\mathbb{R}^D$ or the geodesic distance on $\mathcal{M} \subset \mathbb{R}^D$, if available. We refer to [25] for more details and we note that other interpolation methods are possible.

2.3. Elliptic inverse problems and heat inversion on closed manifolds

In this section, we overview existing procedures for the prior specification and the discretization of prior and forward models on *closed* (i.e. compact and boundary-free) manifolds. We start by describing the two inverse problems used as running examples. Throughout this section, $\mathcal{M}$ will denote a d-dimensional smooth closed manifold isometrically embedded in $\mathbb{R}^D$.

Elliptic inverse problem. Consider the elliptic PDE

$$\mathcal{L}u := -\operatorname{div}(\kappa \nabla u) = f, \quad x \in \mathcal{M}, \tag{2.6}$$

where the divergence and gradient operators are defined with respect to the Riemannian metric inherited by $\mathcal{M}$ from $\mathbb{R}^D$. The goal of the elliptic inverse problem is to recover the diffusion coefficient κ given the right-hand side f (assumed to be smooth) and noisy pointwise observation of the solution u at M spatial locations $\{\tilde{x}_m\}_{m=1}^M \subset \mathcal{M}$. The data are therefore given by $\{y_m\}_{m=1}^M = \{u(\tilde{x}_m) + \eta_m\}_{m=1}^M$. We cast this problem into the general framework (2.1) defining the forward map $\mathcal{F} : \theta \mapsto u$, where $\theta := \log \kappa \in (-\infty, \infty)$. Previous kernel-based methods for elliptic inverse problems on manifolds [26] assumed $\mathcal{M}$ to be closed so that the elliptic PDE given in (2.6) is not supplemented with a boundary condition. Bayesian elliptic inverse problems on Euclidean domains are one of the standard model problems in uncertainty quantification [18, 23, 44]. We will refer to the operator $\mathcal{L}$ as the weighted Laplacian operator.

Heat inversion. Consider the heat equation

$$\begin{cases} u_t = -\Delta_{\mathcal{M}} u, & (x,t) \in \mathcal{M} \times [0,\infty), \\ u(x,0) = \theta(x), & x \in \mathcal{M}, \end{cases} \tag{2.7}$$

where $\Delta_{\mathcal{M}} := -\operatorname{div}(\nabla \cdot)$ is the Laplace–Beltrami operator on $\mathcal{M}$, which reduces to the negative of the conventional Laplace operator in Euclidean space. The goal of the inverse heat problem is to recover the initial heat θ defined on $\mathcal{M}$ from noisy pointwise observation of the heat at time $t^* > 0$ along M spatial locations $\{\tilde{x}_m\}_{m=1}^M \subset \mathcal{M}$. The data are therefore given by $\{y_m\}_{m=1}^M = \{u(\tilde{x}_m, t^*) + \eta_m\}_{m=1}^M$ where $u(\cdot, t^*)$ is the solution of the heat equation (2.7) at time t^*. The corresponding forward map is a heat equation solver, namely $\mathcal{F} : \theta \mapsto u$, where $u(x) := u(x, t^*)$ and θ is the initial condition. The function space formulation of Bayesian heat inversion was introduced in [21] on Euclidean domains, and graph-based formulations on closed manifolds were studied in [25]. Other than its natural physical interpretation, this inverse problem can also be viewed as a standard deconvolution problem arising in imaging applications.

2.3.1. Prior specification and discretization. Here we describe the specification of Matérn Gaussian field priors for the log-diffusion coefficient and the initial heat on closed manifolds. We will also overview their graph-based discretization. Recall that the Matérn model on a closed manifold $\mathcal{M}$ is defined [31] as the Gaussian measure

$$\mu = \mathcal{N}(0, V), \quad V = c(\tau I + \Delta_{\mathcal{M}})^{-s}, \tag{2.8}$$

where $\Delta_{\mathcal{M}} := -\operatorname{div}(\nabla \cdot)$ is the Laplace–Beltrami operator on $\mathcal{M}$ and $\tau > 0, s > \frac{d}{2}$ are two free parameters. The choice of normalizing constant

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

where $(\lambda_i)_{i=1}^{\infty}$ are the increasingly ordered eigenvalues of $\Delta_{\mathcal{M}}$, ensures that $\theta \sim \mu$ has unit marginal variance. Samples can be represented using the Karhunen–Loève (KL) expansion

$$\theta(x) = \sqrt{c} \sum_{i=1}^{\infty} (\tau + \lambda_i)^{-\frac{s}{2}} \zeta_i \phi_i(x), \quad x \in \mathcal{M}, \tag{2.9}$$

where $(\zeta_i)_{i=1}^{\infty}$ are i.i.d. standard normal random variables and $(\phi_i)_{i=1}^{\infty}$ are eigenfunctions of $\Delta_{\mathcal{M}}$ with corresponding eigenvalues $(\lambda_i)_{i=1}^{\infty}$. The parameter $\tau^{\frac{1}{2}}$ represents the inverse length-scale and allows to discern the significant terms in the KL expansion (2.9). The parameter s characterizes the almost-sure regularity of the samples. The requirement $s > \frac{d}{2}$ is motivated

by Weyl's law, see [12, theorem 72] and [15], which asserts that $\lambda_i \asymp i^{\frac{2}{d}}$, i.e., the asymptotic behavior of λ_i is equivalent to that of $i^{\frac{2}{d}}$. Thus, $s > \frac{d}{2}$ ensures that μ is a well-defined Gaussian measure in $L^2(\mathcal{M})$. Further increasing s allows to ensure higher-order sample path differentiability and Sobolev regularity [41, 44], and thereby the well-posedness of elliptic and heat inverse problems [26, 44]. The flexibility of the Matérn model can be enhanced by letting the inverse length-scale be a spatially varying function rather than a scalar parameter [31, 39].

To define a prior distribution over discrete functions defined along the point cloud $\{x_n\}_{n=1}^N \subset \mathcal{M}$, the paper [26] proposed to replace the Laplace–Beltrami operator $\Delta_{\mathcal{M}}$ in (2.8) by a graph Laplacian Δ_N constructed using the point cloud $\{x_n\}_{n=1}^N$. To be more specific, the discretized prior distribution is given by

$$\mu_N = \mathcal{N}(0, V_N), \quad V_N = c_N(\tau I + \Delta_N)^{-s},$$

where $\Delta_N \in \mathbb{R}^{N\times N}$ is a graph Laplacian constructed using $\{x_n\}_{n=1}^N$ and $\tau > 0$, $s > \frac{d}{2}$ are two free parameters. The role of the parameters τ and s is analogous to the infinite-dimensional case and these can be manually tuned, or learned from data using a hierarchical Bayesian approach [26, 39]. In practice, among different choices of graph Laplacian [48], the use of self-tuning graph Laplacian [51] was recommended in [26]. More specifically, in our numerical experiments we use a symmetric graph Laplacian given by

$$\Delta_N = I_N - A^{-\frac{1}{2}} S A^{-\frac{1}{2}}. \tag{2.10}$$

Here S is a similarity matrix and A is a diagonal matrix whose entries are respectively given by

$$S_{ij} = \exp\left(-\frac{|x_i - x_j|^2}{2d_i d_j}\right), \quad A_{ii} = \sum_{j=1}^N S_{ij},$$

where d_i is the distance between x_i and its kth closest neighbor. Same definition applies for d_j. Typically, the parameter k is tuned empirically. We refer to [48] for a review of graph Laplacians and to [39] for generalizations of this graph-based prior model and its connection to the Matérn family [43]. To obtain samples whose variance per node is one, the normalizing constant c_N can be set to be

$$c_N = \frac{N}{\sum_{n=1}^N (\tau + \lambda_n^{(N)})^{-s}},$$

where $(\lambda_n^{(N)})_{n=1}^N$ are the increasingly ordered eigenvalues of Δ_N. Samples θ_N from this prior can be represented via the KL expansion

$$\theta_N(x_i) = \sqrt{c_N} \sum_{n=1}^N (\tau + \lambda_n^{(N)})^{-\frac{s}{2}} \zeta_n \phi_n^{(N)}(x_i), \quad 1 \leqslant i \leqslant N, \tag{2.11}$$

where $(\zeta_n)_{n=1}^N$ are i.i.d. standard normal random variables and $(\phi_n^{(N)})_{n=1}^N$ are eigenvectors of Δ_N with corresponding eigenvalues $(\lambda_n^{(N)})_{n=1}^N$.

2.3.2. Forward map discretization. **Elliptic inverse problem.** For the discretization of the elliptic forward map, one can approximate the operator $\mathcal{L}$ in equation (2.6) by an integral operator [26], which can be subsequently approximated using a Monte Carlo sum. To be more specific, let

$$G_\epsilon(u(x)) := \epsilon^{-\frac{d}{2}} \int_{\mathcal{M}} h\left(\frac{|x-y|^2}{4\epsilon}\right) u(y)\, \mathrm{d}\mathcal{V}(y),$$

where $h(s) = 2^{-d}\pi^{-\frac{d}{2}}\exp(-s)$ and $\mathcal{V}$ denotes the volume form inherited by $\mathcal{M}$ from the ambient space $\mathbb{R}^D$. For a smooth u, it was shown in [14] that

$$G_\epsilon(u(x)) = u(x) + \epsilon(\omega u(x) - \Delta_{\mathcal{M}} u(x)) + \mathcal{O}(\epsilon^2), \quad x \in \mathcal{M}, \tag{2.12}$$

where ω is a function that depends only on the parametrization of the manifold $\mathcal{M}$.

Recall that $\Delta_{\mathcal{M}} := -\operatorname{div}(\nabla\cdot)$ and from (2.6),

$$\mathcal{L}u := -\operatorname{div}(\kappa\nabla u) = \kappa\Delta_{\mathcal{M}} u - \nabla u\cdot\nabla\kappa \tag{2.13}$$

$$= \sqrt{\kappa}\left(\Delta_{\mathcal{M}}(u\sqrt{\kappa}) - u\Delta_{\mathcal{M}}(\sqrt{\kappa})\right). \tag{2.14}$$

Use (2.12) on $\sqrt{\kappa}$ and $u\sqrt{\kappa}$ to obtain,

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

This yields

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

Motivated from this equation, we define

$$\mathcal{L}_\epsilon u(x) := \frac{\sqrt{\kappa(x)}}{\epsilon}\left(u(x)G_\epsilon\left(\sqrt{\kappa(x)}\right) - G_\epsilon(u(x)\sqrt{\kappa(x)})\right),$$

which can be rewritten as the following integral operator

$$\mathcal{L}_\epsilon u(x) := \frac{1}{\epsilon^{\frac{d}{2}+1}}\int_{\mathcal{M}} h\left(\frac{|x-y|^2}{4\epsilon}\right)\sqrt{\kappa(x)\kappa(y)}\,(u(x)-u(y))\,\mathrm{d}\mathcal{V}(y) \tag{2.15}$$

satisfying

$$\mathcal{L}_\epsilon u(x) = \mathcal{L}u(x) + \mathcal{O}(\epsilon), \quad x \in \mathcal{M}.$$

The kernel operator $\mathcal{L}_\epsilon$ can be approximated by Monte Carlo viewing the point cloud $\{x_n\}_{n=1}^N$ as manifold samples. Using importance sampling with approximate density q_ϵ, we have an approximation of $\mathcal{L}_\epsilon$ given by

$$\mathcal{L}_{\epsilon,N} u(x_i) := \frac{1}{\epsilon^{\frac{d}{2}+1}}\left(\frac{1}{N}\sum_{j=1}^N h\left(\frac{|x_i-x_j|^2}{4\epsilon}\right)\right.$$
$$\left.\times\sqrt{\kappa(x_i)\kappa(x_j)}\,q_\epsilon(x_j)^{-1}\left(u(x_i)-u(x_j)\right)\right), \quad i = 1,\ldots,N, \tag{2.16}$$

$$q_\epsilon(x_j) := \frac{1}{2^d\pi^{\frac{d}{2}}N\epsilon^{\frac{d}{2}}}\sum_{k=1}^N \exp\left(-\frac{|x_j-x_k|^2}{4\epsilon}\right)$$

is a kernel estimator of the density of the point cloud. One can write the discretized weighted Laplacian operator in (2.16) in a matrix form. Specifically, define a kernel matrix H with entries $H_{ij} := \exp\left(-\frac{|x_i - x_j|^2}{4\epsilon}\right)$ and define a vector Q with entries $Q_i = \sum_{j=1}^N H_{ij}$. Set the matrix W with entries $W_{ij} = \sqrt{\kappa(x_i)\kappa(x_j)}\, H_{ij} Q_j^{-1}$ and the diagonal matrix D with diagonal entries $D_{ii} = \sum_{j=1}^N W_{ij}$. Then the discretized weighted Laplacian $\mathcal{L}_{\epsilon,N}$ can be written as

$$\mathcal{L}_{\epsilon,N} = \frac{1}{\epsilon}(D - W). \tag{2.17}$$

For practical implementation, the bandwidth parameter ϵ can be empirically chosen so that it lies in the region where

$$\log(T(\epsilon)) = \log\left(\sum_{i,j=1}^{N,K} \exp\left(-\frac{|x_i - x_j(i)|^2}{4\epsilon}\right)\right) \tag{2.18}$$

is approximately linear [26]. Here $(x_j(i))_{j=1}^K$ are the K closest points to the point x_i. In practice, it was further observed that the maximum slope of $\log(T(\epsilon))$ often coincided with $\frac{d}{2}$, where d is the dimension of the underlying manifold $\mathcal{M}$ (see [2]).

Using the above discretization, one can obtain a discretized forward map, $\mathcal{F}_{\epsilon,N}$, which maps $\theta_N = (\log(\kappa(x_1)), \ldots, \log(\kappa(x_N)))$ to an N-dimensional vector u_N, which represents an approximate solution to (2.6) restricted to the point cloud. In other words, u_N is the minimal norm least-squares solution of

$$\mathcal{L}_{\epsilon,N} u_N = f_N,$$

where $f_N = (f(x_1), \ldots, f(x_N))$. Therefore, the discretized forward map is given by

$$\mathcal{F}_{\epsilon,N} : \theta_N \mapsto u_N = \mathcal{L}_{\epsilon,N}^{-1} f_N,$$

where $\mathcal{L}_{\epsilon,N}^{-1}$ denotes the pseudo-inverse. This allows us to write the discretized posterior distribution μ_N^y as a change of measure with respect to the discretized prior distribution μ_N in the following way:

$$\frac{\mathrm{d}\mu_N^y}{\mathrm{d}\mu_N}(\theta_N) \propto \exp\left(-\frac{1}{2}|y - \mathcal{G}_{\epsilon,N}(\theta_N)|_\Gamma^2\right),$$

where $\mathcal{G}_{\epsilon,N}(\theta_N) = (u_N(\tilde{x}_1), \ldots, u_N(\tilde{x}_M))$. The Lebesgue density of the posterior is given by $\pi_N^y(\theta_N) \propto \rho\left(y - \mathcal{G}_{\epsilon,N}(\theta_N)\right) \pi_N(\theta_N)$, where ρ and π_N are Gaussian densities $\mathcal{N}(0,\Gamma)$ and $\mathcal{N}(0, V_N)$ in $\mathbb{R}^M$ and $\mathbb{R}^N$, respectively. The graph pCN algorithm can then be used to obtain samples $\theta_N \sim \pi_N^y$ that can be extended into the underlying manifold using a K-NN map.

Heat inversion. To discretize the heat forward map, we replace the Laplace–Beltrami operator in (2.7) with the graph Laplacian Δ_N in (2.10) and solve

$$\begin{cases} \dfrac{\partial}{\partial t} u_N = -\Delta_N u_N, \quad t \in [0,\infty), \\ u_N(0) = \theta_N, \end{cases}$$

where $\theta_N \in \mathbb{R}^N$ represents the initial heat function restricted to the point cloud. Then the solution $u_N := u_N(t^*)$ at time t^* can be expressed as

$$u_N = \sum_{n=1}^{N} \langle \theta_N, \phi_n^{(N)} \rangle \mathrm{e}^{-\lambda_n^{(N)} t^*} \phi_n^{(N)},$$

where $(\lambda_n^{(N)}, \phi_n^{(N)})_{n=1}^N$ are the ordered eigenpairs of the graph Laplacian Δ_N. From this, we naturally obtain a discretization of the forward map given by

$$\mathcal{F}_N : \theta_N \mapsto u_N = \sum_{n=1}^{N} \langle \theta_N, \phi_n^{(N)} \rangle \mathrm{e}^{-\lambda_n^{(N)} t^*} \phi_n^{(N)}.$$

We can then write the discretized posterior distribution μ_N^y as a change of measure with respect to the prior distribution μ_N

$$\frac{\mathrm{d}\mu_N^y}{\mathrm{d}\mu_N}(\theta_N) \propto \exp\left(-\frac{1}{2}|y - \mathcal{G}_N(\theta_N)|_\Gamma^2\right),$$

where $\mathcal{G}_N(\theta_N) = (u_N(\tilde{x}_1), \ldots, u_N(\tilde{x}_M))$. The Lebesgue density of the posterior is given by $\pi_N^y(\theta_N) \propto \rho\left(y - \mathcal{G}_N(\theta_N)\right) \pi_N(\theta_N)$, and sampling and interpolation can be performed with the general methodology described previously.

3. Bayesian inverse problems on manifolds with boundaries

In this section, we propose a novel methodology for Bayesian inversion on manifolds with boundaries, addressing the design of priors that can reflect the given boundary conditions, the graph-based discretization of these priors, and the graph-based approximation of PDE-constrained forward maps supplemented with boundary conditions. To be concrete, we will focus on the following elliptic and heat inverse problems on one and two-dimensional manifolds with Dirichlet boundary conditions. Throughout this section and the rest of this paper, $\mathcal{M}$ will denote a smooth compact d-dimensional manifold isometrically embedded in $\mathbb{R}^D$ with boundary $\partial\mathcal{M}$. We denote $\mathcal{M}^o = \mathcal{M} \backslash \partial\mathcal{M}$.

Elliptic inverse problem. Consider the following elliptic PDE with Dirichlet boundary conditions,

$$\begin{cases} \mathcal{L}u := -\operatorname{div}(\kappa \nabla u) = f, & x \in \mathcal{M}^o, \\ u(x) = h(x), & x \in \partial\mathcal{M}, \end{cases} \tag{3.1}$$

where the divergence and gradient operators are defined with respect to the Riemannian metric inherited by the manifold $\mathcal{M}$ from $\mathbb{R}^D$. In our numerical examples, we will consider a one-dimensional semi-ellipse where $\partial\mathcal{M} = \{x_0, x_N\}$ and a two-dimensional semi-torus where $\partial\mathcal{M} = \mathcal{B}_1 \cup \mathcal{B}_2$ and $\mathcal{B}_i$, $i = 1, 2$ are circles. We will then denote by h_1 and h_2 the Dirichlet boundary conditions at $\mathcal{B}_1$ and $\mathcal{B}_2$. The goal of the elliptic inverse problem is to recover the diffusion coefficient κ given the right-hand side f and noisy pointwise observation of the solution u at M spatial locations $\{\tilde{x}_m\}_{m=1}^M \subset \mathcal{M}$. The data are given by $\{y_m\}_{m=1}^M = \{u(\tilde{x}_m) + \eta_m\}_{m=1}^M$ with the forward map $\mathcal{F} : \theta \mapsto u$, where $\theta := \log \kappa \in (-\infty, \infty)$. We refer to [27] for sufficient conditions on $\mathcal{M}$, κ, f and h to guarantee that the solution to (3.1) can be evaluated pointwise.

Heat inversion. Consider the following heat equation with Dirichlet boundary conditions,

$$\begin{cases} u_t(x,t) = -\Delta_{\mathcal{M}} u(x,t), & x \in \mathcal{M}^o, t > 0, \\ u(x,t) = h(x), & x \in \partial\mathcal{M},\ t \geqslant 0, \\ u(x,0) = \theta(x), & x \in \mathcal{M}. \end{cases} \tag{3.2}$$

Again the goal of the inverse heat problem is to recover the initial heat θ defined on $\mathcal{M}$ from noisy pointwise observation of the heat at time $t^* > 0$ along M spatial locations $\{\tilde{x}_m\}_{m=1}^M \subset \mathcal{M}$. The data are therefore given by $\{y_m\}_{m=1}^M = \{u(\tilde{x}_m, t^*) + \eta_m\}_{m=1}^M$. The corresponding forward map is a heat equation solver, namely $\mathcal{F} : \theta \mapsto u$, where $u(x) \equiv u(x, t^*)$ denotes the solution of (3.2) at time t^* and θ is the initial condition. For our numerical examples we will use a one-dimensional semi-ellipse and a two-dimensional semi-torus, and we will adopt the same notations as in the elliptic inverse problem.

3.1. Ghost point diffusion maps for Dirichlet boundary conditions

In this section, we give a short discussion on the construction of ghost points and the GPDM algorithm, which will be used in the discretizations of the prior and forward models. In this paper, we focus on a specific GPDM algorithm to approximate the weighted Laplacian operator, $\mathcal{L} = -\text{div}(\kappa\nabla\cdot)$ whose inputs are functions $u : \mathcal{M} \to \mathbb{R}$ that satisfy the Dirichlet boundary condition, $u(x) = h(x)$, for all $x \in \partial\mathcal{M}$. The discussion will focus on the algorithmic aspect that will be used in the forward map discretization. For other types of (possibly non-symmetric) second-order diffusion operators and boundary conditions, and the convergence analysis, we refer to [27, 50].

The key idea of the GPDM algorithm comes from the classical ghost point method [30] for solving PDEs with Neumann boundary condition using the finite-difference method. Particularly, the ghost points are constructed to improve the convergence rate in approximating the normal derivative at the boundary points. In the context of the GPDM algorithm, the ghost points are employed to overcome the biases induced by the graph Laplacian discretization near the boundary. Numerically, solving PDEs with the ghost point method requires the following two steps: (1) specification of ghost points; and (2) specification of function values at the ghost points. While these two steps are trivial when the geometry is Euclidean or known, they require nontrivial numerical algorithms and theoretical justification when the manifold is unknown in the sense that it can only be identified with finitely sampled point cloud data.

The GPDM algorithm addresses step (1) above by augmenting the sampled point cloud data on the manifold with a set of ghost points specified on the outer normal collar of the boundary. Theoretically, the GPDM algorithm extends the embedded manifold $\mathcal{M} \overset{\iota}{\hookrightarrow} \mathbb{R}^D$ with its collar neighbor $\Delta\mathcal{M}$ of a sufficiently large radius, such that the extended manifold $\mathcal{M} \cup \Delta\mathcal{M} \overset{\tilde{\iota}}{\hookrightarrow} \mathbb{R}^D$ is isometrically embedded and does not change the geometry of $\mathcal{M}$, i.e., $\tilde{\iota}|_{\mathcal{M}} = \iota$ (see lemma 3.5 in [27]). With this modification, the graph Laplacian construction in (2.16) is a consistent pointwise estimator of $\mathcal{L}$ for all points in $\mathcal{M}$, even for the points that are very close to the boundary $\partial\mathcal{M}$ since they are sufficiently far away from the boundary of the extended manifold, $\partial(\mathcal{M} \cup \Delta\mathcal{M})$, as illustrated in figure 1. Since our goal is to construct a forward map on the manifold $\mathcal{M}$, we need to specify the additional unknowns (the function values at these ghost points as we noted in the step (2)) by adding more equations. Specifically, we will impose a set of linear extrapolation equations, whose solution specifies the function values on the ghost points through function values on the point cloud.

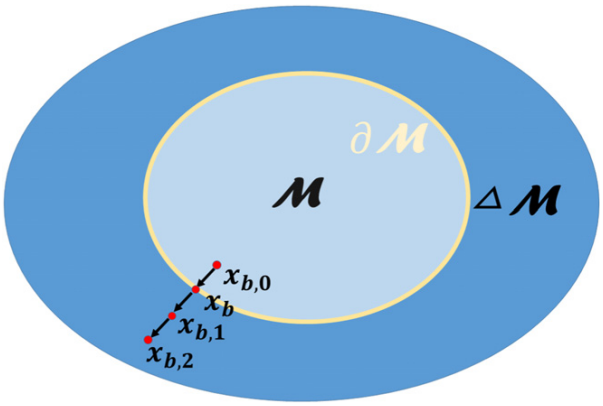

Figure 1. Construction of ghost points.

While there are various ways to realize the two steps above, in the following we will provide a simple numerical procedure for well-ordered data that is used in the numerical examples in this paper. For randomly sampled data or a higher convergence rate method, we refer interested readers to [27]. We should also point out that the presentation below uses much simpler notations compared to those in [27, 50] since we are only interested in the algorithmic aspect of the GPDM method. Given a point cloud $\{x_n\}_{n=1}^N \subset \mathcal{M}$, the GPDM estimator for $\mathcal{L}$ is constructed as follows:

(a) Specification of the ghost points. At each boundary point, $x_b \in \partial\mathcal{M}$:

1. Apply the secant line to approximate the normal vector at each boundary point. Following the notation in the illustration in figure 1, the unit normal vector is estimated via

$$v_b = \frac{x_b - x_{b,0}}{\|x_b - x_{b,0}\|}.$$

2. We specify K ghost points along the normal vector at each boundary point, $x_b \in \partial\mathcal{M} \cap \{x_n\}_{n=1}^N$,

$$x_{b,k} := x_b + \delta k v_b, \quad k = 1, \ldots, K,$$

where $\delta = \|x_b - x_{b,0}\|$ and $x_{b,0}$ is the closest point to x_b in $\mathcal{M}^o \cap \{x_n\}_{n=1}^N$.

(b) Specification of the function values at the ghost points. For each boundary point x_b, we impose the following extrapolation condition,

$$\begin{aligned} u(x_{b,1}) - 2u(x_b) + u(x_{b,0}) &= 0, \\ u(x_{b,2}) - 2u(x_{b,1}) + u(x_b) &= 0, \\ u(x_{b,k}) - 2u(x_{b,k-1}) + u(x_{b,k-2}) &= 0, \quad k = 3, \ldots, K. \end{aligned} \tag{3.3}$$

These algebraic equations are discrete analogs of matching the first-order derivatives along the estimated normal direction, v_b.

(c) Construction of the GPDM discrete estimator. Construct the graph Laplacian estimator as in (2.16) for the extended points, $\{x_n\}_{n=1}^N \cup \{x_{b,k}\}_{b,k=1}^{B,K}$. We point out that we also use a set

of algebraic equations identical to (3.3) to determine the function value of κ at the ghost points $\{x_{b,k}\}_{b,k=1}^{B,K}$. Let $\bar{N} = N + BK$ and obtain the corresponding weighted Laplacian matrix, $\mathcal{L}_{\epsilon,\bar{N}}$, whose construction is analogous to (2.17). The weighted Laplacian matrix can be rewritten as

$$\mathcal{L}_{\epsilon,\bar{N}} = \begin{pmatrix} \mathcal{L}^{(1)}_{\epsilon,\bar{N}} & \mathcal{L}^{(2)}_{\epsilon,\bar{N}} \\ \mathcal{L}^{(3)}_{\epsilon,\bar{N}} & \mathcal{L}^{(4)}_{\epsilon,\bar{N}} \end{pmatrix} \in \mathbb{R}^{\bar{N}\times\bar{N}}. \tag{3.4}$$

We will construct the GPDM matrix based on the sub-matrices $\mathcal{L}^{(1)}_{\epsilon,\bar{N}} \in \mathbb{R}^{N\times N}$, whose entries are constructed based on the affinity between pairs of the point cloud points in $\{x_n\}_{n=1}^N \subset \mathcal{M}$, and $\mathcal{L}^{(2)}_{\epsilon,\bar{N}} \in \mathbb{R}^{N\times BK}$, whose entries represent the affinity between an element of $\{x_n\}_{n=1}^N$ and a ghost point in $\{x_{b,k}\}_{b,k=1}^{B,K}$. Let $u_G = (u(x_{1,1}), \ldots, u(x_{B,K})) \in \mathbb{R}^{BK}$ and $u_N = (u(x_1), \ldots, u(x_{b,0}), \ldots, u(x_N)) \in \mathbb{R}^N$. Then we can write the solution of (3.3) in a compact form as $u_G = Gu_N$ such that $\mathcal{L}^{(1)}_{\epsilon,\bar{N}}u_N + \mathcal{L}^{(2)}_{\epsilon,\bar{N}}u_G = (\mathcal{L}^{(1)}_{\epsilon,\bar{N}} + \mathcal{L}^{(2)}_{\epsilon,\bar{N}}G)u_N$. Based on this observation, we define the GPDM estimator without boundary condition as

$$\tilde{\mathcal{L}}_{\epsilon,N} := \mathcal{L}^{(1)}_{\epsilon,\bar{N}} + \mathcal{L}^{(2)}_{\epsilon,\bar{N}}G \in \mathbb{R}^{N\times N}. \tag{3.5}$$

Denoting

$$\tilde{\mathcal{L}}_{\epsilon,N} = \begin{pmatrix} \tilde{\mathcal{L}}_{\epsilon,(N-B)\times(N-B)} & \tilde{\mathcal{L}}_{\epsilon,(N-B)\times B} \\ \tilde{\mathcal{L}}_{\epsilon,B\times(N-B)} & \tilde{\mathcal{L}}_{\epsilon,B\times B}, \end{pmatrix} \tag{3.6}$$

and splitting $u_N := (u_{N-B}, u_B)$ into function values at the interior and boundary points, respectively, the GPDM algorithm discretizes the Dirichlet problem in (3.1) as follows:

$$\begin{aligned} \tilde{\mathcal{L}}_{\epsilon,(N-B)\times(N-B)}u_{N-B} + \tilde{\mathcal{L}}_{\epsilon,(N-B)\times B}u_B &= f_{N-B}, \\ u_B &= h_B. \end{aligned} \tag{3.7}$$

Here, the components of the vector $f_{N-B} \in \mathbb{R}^{N-B}$ are the function values at the interior points, whereas the components of the vector $h_B \in \mathbb{R}^B$ are the function values at the boundary points.

3.2. Prior specification and discretization

In this section, we propose novel prior constructions to facilitate elliptic and heat Bayesian inversion on manifolds with boundaries. The priors we propose contain two terms. The first one is a Matérn-type Gaussian field with Dirichlet homogeneous boundary conditions, whose primary role is to capture the uncertainty on the parameter of interest in the interior of the manifold. The second term accounts for prior uncertainty along the boundary of the manifold, and can be omitted if the boundary values of the parameter are known.

For a one-dimensional manifold $\mathcal{M}$ with boundary $\partial\mathcal{M} = \{x_1, x_N\}$, i.e. a semi-ellipse in our numerical experiments, we propose using a prior defined as the law of

$$\theta = \underbrace{\frac{1}{\sqrt{\sum_{i=1}^{\infty}(\tau+\lambda_i)^{-s}}}\sum_{i=1}^{\infty}(\tau+\lambda_i)^{-\frac{s}{2}}\zeta_i\phi_i}_{\text{Interior term}} + \underbrace{\mu_1\psi_1 + \mu_2\psi_2}_{\text{Boundary term}}. \tag{3.8}$$

In the interior term, τ, s, and $(\zeta_i)_{i=1}^{\infty}$ play the same role as in (2.9), but now $(\lambda_i, \phi_i)_{i=1}^{\infty}$ are the ordered eigenpairs of the Laplace–Beltrami operator $\Delta_{\mathcal{M}}$ equipped with homogeneous Dirichlet boundary conditions, so that $\phi_i(x_1) = \phi_i(x_N) = 0$ for all i. Nonhomogeneous Dirichlet boundary conditions could also be considered. In the boundary term, $\mu_1, \mu_2 \sim \mathcal{N}(0,1)$ are independent of all other randomness and ψ_1, ψ_2 are harmonic functions defined by

$$\begin{cases} \Delta_{\mathcal{M}}\psi_1 = 0, \\ \psi_1(x_1) = 1, \\ \psi_1(x_N) = 0, \end{cases} \quad \text{and} \quad \begin{cases} \Delta_{\mathcal{M}}\psi_2 = 0, \\ \psi_2(x_1) = 0, \\ \psi_2(x_N) = 1. \end{cases} \tag{3.9}$$

Thus the random coefficients μ_1 and μ_2 in (3.8) respectively represent the uncertainty on the parameter value of interest at the two boundary points x_1 and x_N. Overall, the prior defined by (3.8) combines the flexibility of the Matérn model with homogeneous Dirichlet boundary conditions for interior reconstruction with additional flexibility in the boundary reconstruction. A simulation study that illustrates the increased flexibility afforded by the boundary term and our prior construction is given in section 4.

Following the same idea, for a two-dimensional manifold $\mathcal{M}$ with boundary $\partial\mathcal{M} = \mathcal{B}_1 \cup \mathcal{B}_2$, where $\mathcal{B}_1$ and $\mathcal{B}_2$ are disjoint regular closed curves, e.g. in our numerical examples $\mathcal{B}_1$ and $\mathcal{B}_2$ are boundary circles of a semi-torus, we define a prior as the law of

$$\theta = \underbrace{\frac{1}{\sqrt{\sum_{i=1}^{\infty}(\tau+\lambda_i)^{-s}}}\sum_{i=1}^{\infty}(\tau+\lambda_i)^{-\frac{s}{2}}\zeta_i\phi_i}_{\text{Interior term}} + \underbrace{\sum_{\ell=1}^{L}\mu_{1,\ell}\psi_{1,\ell} + \sum_{\ell=1}^{L}\mu_{2,\ell}\psi_{2,\ell}}_{\text{Boundary term}}. \tag{3.10}$$

The interior term is defined analogously as in the one-dimensional case, using the spectrum of the Laplace–Beltrami operator with homogeneous Dirichlet boundary conditions. For the boundary term, $\mu_{1,\ell}, \mu_{2,\ell} \sim \mathcal{N}(0,1)$ are independent of all other randomness, and $\{\psi_{1,\ell}\}_{\ell=1}^{L}$ and $\{\psi_{2,\ell}\}_{\ell=1}^{L}$ are harmonic functions on $\mathcal{M}$ defined by

$$\begin{cases} \Delta_{\mathcal{M}}\psi_{1,\ell} = 0, \\ \psi_{1,\ell}(x) = \tilde{\psi}_{1,\ell}(x), \quad x \in \mathcal{B}_1, \\ \psi_{1,\ell}(x) = 0, \quad x \in \mathcal{B}_2, \end{cases} \quad \text{and}$$

$$\begin{cases} \Delta_{\mathcal{M}}\psi_{1,\ell} = 0, \\ \psi_{2,\ell}(x) = 0, \quad x \in \mathcal{B}_1, \\ \psi_{2,\ell}(x) = \tilde{\psi}_{2,\ell}(x), \quad x \in \mathcal{B}_2, \end{cases} \quad \text{for } \ell = 1, \ldots, L, \tag{3.11}$$

where $\tilde{\psi}_{1,\ell}(x)$ and $\tilde{\psi}_{2,\ell}(x)$ are ordered eigenfunctions of the Laplace–Beltrami operator defined on the closed curves $\mathcal{B}_1$ and $\mathcal{B}_2$, respectively. The number L of basis-type functions for each boundary controls the flexibility of the prior along the boundary. Larger L allows to recover more frequencies of the parameter of interest along the boundary, but at the expense of introducing additional model parameters.

To discretize the prior defined in (3.8), we simply replace the role of the Laplace–Beltrami operator $\Delta_{\mathcal{M}}$ with a graph Laplacian as we did in subsection 2.3.1, but now taking care of

boundary conditions. Specifically, for one-dimensional inverse problems, our discretized prior distribution would have samples of the form

$$\theta_N \sim \frac{\sqrt{N}}{\sqrt{\sum_{n=1}^{N}(\tau+\lambda_n^{(N)})^{-s}}}\sum_{n=1}^{N}(\tau+\lambda_n^{(N)})^{-\frac{s}{2}}\zeta_n\phi_n^{(N)}+\mu_1\psi_1^{(N)}+\mu_2\psi_2^{(N)}, \quad (3.12)$$

where $(\lambda_n^{(N)},\phi_n^{(N)})$ are ordered eigenpairs of $\tilde{\Delta}_N$ defined as the $N\times N$ submatrix corresponding to the point cloud portion of the graph Laplacian (2.10) constructed using both the point cloud $\{x_n\}_{n=1}^N$ and ghost points $\{x_{b,k}\}_{b,k=1}^{B,K}$. From now on, we will refer to $\tilde{\Delta}_N$ as a *truncated graph Laplacian*. More specifically, given the point cloud and ghost points, we define $\tilde{\Delta}_N$ as an $N\times N$ submatrix of the $\bar{N}\times\bar{N}$ matrix $\Delta_{\bar{N}}$ with $\bar{N}:=N+BK$. Indices of the submatrix $\tilde{\Delta}_N$ correspond to that of the elements in the point cloud. We have observed such a construction provided an effective approximation to the Laplace–Beltrami operator on $\mathcal{M}$ with homogeneous Dirichlet boundary conditions in our numerical experiments. This construction avoids any potential idiosyncratic boundary behavior of eigenvectors when using a graph Laplacian constructed solely with point cloud data. Indeed, in our numerical experiments (see figure 6 and the associated discussion in section 4 below) we note that the eigenvectors of the self-tuned graph Laplacian (2.10) have spikes and oscillations near the boundary. Finally, $\psi_1^{(N)}$ and $\psi_2^{(N)}$ are the solutions of (3.9) evaluated along the point cloud. These solutions can be obtained by the GPDM algorithm as described in section 3.1.

An alternative attempt to discretize the homogeneous Dirichlet boundary condition was introduced in [46] where they employed truncation on the original point clouds without adding ghost points. In fact, the spectral convergence of the truncated graph Laplacian to the Dirichlet Laplacian on manifold with boundaries has recently been reported [35]. We should also point out that there were two discretizations of the Laplace–Beltrami operator introduced in this paper: (1) self-tuned normalized symmetric graph Laplacian given in (2.10) and (2) GPDM matrix given in (3.5). The eigenvectors of the former are orthogonal as the matrix is symmetric, whereas the eigensolutions of the latter are not the case since the GPDM matrix in (3.5) is not symmetric. With this fact, we clarify that we only use GPDM to solve linear problems (3.9) and do not use the eigenvectors of the corresponding GPDM matrix elsewhere.

For the two-dimensional graph-based prior discretization, we obtain a finite number of discretized functions for each boundary, denoted by $\{\psi_{1,\ell}^{(N)}\}_{\ell=1}^L$ and $\{\psi_{2,\ell}^{(N)}\}_{\ell=1}^L$, by solving (3.11) along the point cloud. One can again use the GPDM algorithm to obtain $\{\psi_{1,\ell}^{(N)}\}_{\ell=1}^L$ and $\{\psi_{2,\ell}^{(N)}\}_{\ell=1}^L$ where the boundary basis functions $\{\tilde{\psi}_{1,\ell}\}_{\ell=1}^L$ and $\{\tilde{\psi}_{2,\ell}\}_{\ell=1}^L$ are discretized using a self-tuned graph Laplacian constructed with all the point cloud elements that lie on the boundaries, which are two disjoint closed curves in our numerical examples. For instance, to obtain $\{\tilde{\psi}_{1,\ell}\}_{\ell=1}^L$ one can construct the self-tuned graph Laplacian solely using points in $\mathcal{B}_1\cap\{x_n\}_{n=1}^N$, where $\{x_n\}_{n=1}^N\subset\mathcal{M}$ is the point cloud of the manifold. Then $\{\tilde{\psi}_{1,\ell}\}_{\ell=1}^L$ can be chosen to be the first L eigenvectors of this self-tuned graph Laplacian. To summarize, in the two-dimensional case, samples from the proposed graph-based prior are defined by

$$\begin{aligned}\theta_N \sim{}& \frac{\sqrt{N}}{\sqrt{\sum_{n=1}^{N}(\tau+\lambda_n^{(N)})^{-s}}}\sum_{n=1}^{N}(\tau+\lambda_n^{(N)})^{-\frac{s}{2}}\zeta_n\phi_n^{(N)}\\ &+\sum_{\ell=1}^{L}\mu_{1,\ell}\psi_{1,\ell}^{(N)}+\sum_{\ell=1}^{L}\mu_{2,\ell}\psi_{2,\ell}^{(N)},\end{aligned} \quad (3.13)$$

where $(\lambda_n^{(N)},\phi_n^{(N)})_{n=1}^N$ are ordered eigenpairs of a truncated graph Laplacian.

3.3. Forward map discretization

Using the tools introduced in previous sections, here we propose graph-based discretizations of forward maps for elliptic and heat inverse problems on one and two-dimensional manifolds with boundary.

Elliptic inverse problem. For the inversion problem involving the elliptic PDE in (3.1), our goal is to learn the diffusion coefficient κ on $\mathcal{M}$. Correspondingly, analogous to the boundary-free setting, a forward map was given by $\mathcal{F}:\theta\mapsto u$, where $\theta=\log\kappa\in(-\infty,\infty)$ and u solves (3.1). In this case, we can use the GPDM algorithm introduced in section 3.1 to obtain the discretization of the forward map. Precisely, the discretized forward map is given by

$$\mathcal{F}_{\epsilon,N}:\theta_N\mapsto u_N=(u_{N-B},u_B)=\left(\tilde{\mathcal{L}}^{-1}_{\epsilon,(N-B)\times(N-B)}(f_{N-B}-\tilde{\mathcal{L}}_{\epsilon,(N-B)\times B}h_B),h_B\right),$$

where $\theta_N=(\log(\kappa(x_1)),\ldots,\log(\kappa(x_N)))$ and $\tilde{\mathcal{L}}^{-1}_{\epsilon,(N-B)\times(N-B)}$ denotes the pseudo-inverse. From this discretization process, we arrive at the relationship between the discretized posterior distribution μ^y_N and the discretized prior distribution μ_N, which is given by

$$\frac{\mathrm{d}\mu^y_N}{\mathrm{d}\mu_N}(\theta_N)\propto\exp\left(-\frac{1}{2}|y-\mathcal{G}_{\epsilon,N}(\theta_N)|^2_\Gamma\right),$$

where $\mathcal{G}_{\epsilon,N}(\theta_N)=(u_N(\tilde{x}_1),\ldots,u_N(\tilde{x}_M))$. For the sampling and interpolation steps, we follow the general methodology in section 2.2.

Heat inversion. Consider first the one-dimensional heat equation with Dirichlet boundary conditions given by (3.2) with $\partial\mathcal{M}=\{x_1,x_N\}$. Analogous to the prior construction, we introduce two time-independent functions ψ_1 and ψ_2 defined on $\mathcal{M}$ satisfying

$$\begin{cases}\Delta_{\mathcal{M}}\psi_1=0,\\ \psi_1(x_1)=1,\\ \psi_1(x_N)=0,\end{cases}\quad\text{and}\quad\begin{cases}\Delta_{\mathcal{M}}\psi_2=0,\\ \psi_2(x_1)=0,\\ \psi_2(x_N)=1.\end{cases}$$

Then we can obtain the solution of (3.2) by solving for w that satisfies,

$$\begin{cases}w_t(x,t)=-\Delta_{\mathcal{M}}w(x,t), & x\in\mathcal{M}^o,t>0,\\ w(x_1,t)=0,\ w(x_N,t)=0, & t\geqslant 0,\\ w(x,0)=\theta(x)-h(x_1)\psi_1(x)-h(x_N)\psi_2(x), & x\in\mathcal{M}.\end{cases}\tag{3.14}$$

To see this, notice that the function

$$u^*(x,t):=w(x,t)+h(x_1)\psi_1(x)+h(x_N)\psi_2(x)\tag{3.15}$$

is the solution of (3.2).

For the two-dimensional heat equation, let $\tilde{\psi}_{1,\ell}$ and $\tilde{\psi}_{2,\ell}$ be ordered eigenfunctions of the Laplace–Beltrami operator defined on $\mathcal{B}_1$ and $\mathcal{B}_2$, respectively. We can then write

$$h_1(x)=\sum_{\ell=1}^{\infty}a_\ell\tilde{\psi}_{1,\ell}(x),\quad x\in\mathcal{B}_1,\quad\text{and}\quad h_2(x)=\sum_{\ell=1}^{\infty}b_\ell\tilde{\psi}_{2,\ell}(x),\quad x\in\mathcal{B}_2,$$

for some real coefficients $\{a_\ell\}_{\ell=1}^\infty$ and $\{b_\ell\}_{\ell=1}^\infty$. The solution to (3.2) is then given by

$$u^*(x,t) := w(x,t) + \sum_{\ell=1}^{\infty} a_\ell \psi_{1,\ell}(x) + \sum_{\ell=1}^{\infty} b_\ell \psi_{2,\ell}(x), \tag{3.16}$$

where w is the solution of

$$\begin{cases} w_t(x,t) = -\Delta_{\mathcal{M}} w(x,t), & x \in \mathcal{M}^o, \quad t > 0, \\ w(x,t) = 0, & x \in \mathcal{B}_1, \quad t \geqslant 0, \\ w(x,t) = 0, & x \in \mathcal{B}_2, \quad t \geqslant 0, \\ w(x,0) = \theta(x) - \sum_{\ell=1}^{\infty} a_\ell \psi_{1,\ell}(x) - \sum_{\ell=1}^{\infty} b_\ell \psi_{2,\ell}(x), & x \in \mathcal{M}. \end{cases} \tag{3.17}$$

Here, $\psi_{1,\ell}$ is the harmonic function with boundary condition $\tilde{\psi}_{1,\ell}$ in $\mathcal{B}_1$ and 0 in $\mathcal{B}_2$, while $\psi_{2,\ell}$ is the harmonic function with boundary condition 0 in $\mathcal{B}_1$ and $\tilde{\psi}_{2,\ell}$ in $\mathcal{B}_2$. We have hence expressed the solution of the heat equation with non-homogeneous boundary conditions as a linear superposition of basis-like functions and the solution of homogeneous heat equation. Such a decomposition will play a key role in the discretization procedures.

Now suppose that in the one-dimensional case we are given an N-dimensional discrete representation θ_N of the initial heat distribution of the form

$$\theta_N = \psi_3^{(N)} + \mu_1 \psi_1^{(N)} + \mu_2 \psi_2^{(N)}, \quad \text{where}$$

$$\psi_3^{(N)} = \frac{\sqrt{N}}{\sqrt{\sum_{n=1}^N (\tau + \lambda_n^{(N)})^{-s}}} \sum_{n=1}^{N} (\tau + \lambda_n^{(N)})^{-\frac{s}{2}} \zeta_n \phi_n^{(N)},$$

which reflects our prior construction given in the previous section. To discretize the forward map of the one-dimensional heat equation, we replace the Laplace–Beltrami operator $\Delta_{\mathcal{M}}$ by $\tilde{\Delta}_N$ and solve for $w^{(N)} = (w_1, \ldots, w_N)$ satisfying

$$\begin{cases} \dfrac{\partial}{\partial t} w^{(N)} = -\tilde{\Delta}_N w^{(N)}, \\ w^{(N)}(0) = \psi_3^{(N)}. \end{cases} \tag{3.18}$$

This equation can be viewed as a discrete analog of (3.14) as the eigenvectors of $\tilde{\Delta}_N$ approximate those of the Laplace–Beltrami operator with homogeneous Dirichlet boundary condition. Given an initial condition $\psi_3^{(N)} = \theta_N - \mu_1 \psi_1^{(N)} - \mu_2 \psi_2^{(N)}$, the solution of the above initial value problem can be expressed as a linear combination of the eigenvectors of $\tilde{\Delta}_N$. In other words, the solution at time t is given by

$$w^{(N)}(t) = \sum_{n=1}^{N} \langle \phi_n^{(N)}, \psi_3^{(N)} \rangle \mathrm{e}^{-\lambda_n^{(N)} t} \phi_n^{(N)}.$$

From this, and the previous observations we made in (3.15), we naturally obtain the discrete approximation for the solution of (3.2) given by

$$u_N = w^{(N)} + \mu_1 \psi_1^{(N)} + \mu_2 \psi_2^{(N)},$$

which defines the discretized forward map $\mathcal{F}_N : \theta_N \mapsto u_N$.

Using the same argument, we can discretize the forward map for the two-dimensional heat equation. Suppose we are given an N-dimensional discrete representation of the initial heat function for two-dimensional heat inverse problem, denoted by θ_N, of the form

$$\theta_N = \psi_3^{(N)} + \sum_{\ell=1}^{L} \mu_{1,\ell}\psi_{1,\ell}^{(N)} + \sum_{\ell=1}^{L} \mu_{2,\ell}\psi_{2,\ell}^{(N)}, \quad \text{where}$$

$$\psi_3^{(N)} = \frac{\sqrt{N}}{\sqrt{\sum_{n=1}^{N}(\tau + \lambda_n^{(N)})^{-s}}} \sum_{n=1}^{N} (\tau + \lambda_n^{(N)})^{-\frac{s}{2}} \zeta_n \phi_n^{(N)}. \tag{3.19}$$

Similarly as in the one-dimensional case, with the observation (3.16), the discrete approximation for the solution of (3.2) in the two-dimensional case is given by

$$u_N = w^{(N)} + \sum_{\ell=1}^{L} \mu_{1,\ell}\psi_{1,\ell}^{(N)} + \sum_{\ell=1}^{L} \mu_{2,\ell}\psi_{2,\ell}^{(N)},$$

where $w^{(N)}$ denotes the solution of (3.18). Accordingly, the discretized forward map is given by $\mathcal{F}_N : \theta_N \mapsto u_N$. In both the one and two-dimensional cases the discretized posterior distribution μ_N^y and the discretized prior distribution μ_N satisfy the relationship

$$\frac{\mathrm{d}\mu_N^y}{\mathrm{d}\mu_N}(\theta_N) \propto \exp\left(-\frac{1}{2}|y - \mathcal{G}_N(\theta_N)|_\Gamma^2\right),$$

where $\mathcal{G}_N(\theta_N) = (u_N(\tilde{x}_1), \ldots, u_N(\tilde{x}_M))$. For the sampling and interpolation steps, we again follow the general methodology described in section 2.2.

4. Numerical results

In this section, we provide simulation results for the numerical solution of Bayesian inverse problems on manifolds with boundaries. We showcase our methodology for the elliptic inverse problem in section 4.1 and for the heat inverse problem in section 4.2. For each inverse problem, we consider one-dimensional and two-dimensional examples. In addition to validating our approach and providing implementation details, our numerical results will demonstrate the enhanced flexibility of our proposed priors near the boundary when compared with previous graph representations of Matérn priors. Specifically, we show the improved reconstruction achieved by our prior in a one-dimensional elliptic inverse problem, and we illustrate in a two-dimensional setting the emergence of artifacts near the boundary for the eigenfunctions of the graph Laplacian (2.10) used to define graph Matérn priors on closed manifolds in [26].

Our one-dimensional examples are set on a semi-ellipse and the two-dimensional examples are set on a semi-torus. For the semi-ellipse the embedding was given by

$$\iota(\alpha) = \begin{pmatrix} \cos\,\alpha \\ 3\,\sin\,\alpha \end{pmatrix}, \quad \alpha \in [0, \pi], \tag{4.1}$$

with Riemannian metric

$$g = \sin^2(\alpha) + 9\cos^2(\alpha). \tag{4.2}$$

The embedding for the semi-torus was given by

$$\iota(\alpha,\beta) = \begin{pmatrix} (2+\cos\ \alpha)\cos\ \beta \\ (2+\cos\ \alpha)\sin\ \beta \\ \sin\ \alpha \end{pmatrix}, \quad \alpha \in [0,2\pi],\ \beta \in [0,\pi], \tag{4.3}$$

where (α,β) are the intrinsic coordinates and the corresponding Riemannian metric is given by

$$g = \begin{pmatrix} 1 & 0 \\ 0 & (2+\cos\ \alpha)^2 \end{pmatrix}. \tag{4.4}$$

We use uniform grids to define point clouds over the given manifolds. For the one-dimensional semi-ellipse we used $N = 630$ points and for the two-dimensional semi-torus $N = N_1 \times N_2 = 1296$ points in a 36 by 36 grid. The boundary of the semi-ellipse corresponds to $\alpha = 0$ or $\alpha = \pi$, which necessitates to model boundary values of the parameter on $\alpha = 0, \pi$. For the semi-torus the boundary corresponds to $\beta = 0$ or $\beta = \pi$, which implies the need to model boundary values of the parameter along two boundary curves. In all of our numerical experiments we choose $N = M$, that is, we assume to have observations along the entire point cloud.

4.1. Elliptic inverse problem

For the one-dimensional elliptic inverse problem, we set the true PDE solution $u^\dagger$ to be

$$u^\dagger(\alpha) = \sin(\alpha),$$

and for the two-dimensional elliptic inverse problem, the true PDE solution was given by

$$u^\dagger(\alpha,\beta) = 10\ \sin(2\alpha)\cos(\beta).$$

While the one-dimensional solution satisfies homogeneous Dirichlet boundary condition at $\alpha = 0, \pi$, the two-dimensional solution has non-trivial Dirichlet boundary condition at $\beta = 0$ and π. According to the above true solution, the observations were given by

$$y_n = u^\dagger(\alpha_n) + \eta_n, \qquad \alpha_n = \iota^{-1}(x_n), \quad n = 1,\ldots,N,$$

where $\eta_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 0.01)$. We consider several choices of true input parameter κ, and for each choice we define the right-hand side of the PDE using the identity

$$f = -\text{div}(\kappa\nabla u) = -\frac{1}{\sqrt{\det g}}\partial_i\left(\kappa g^{ij}\partial_j u\sqrt{\det g}\right).$$

Note that here and henceforth we abuse notation by referring to $u^\dagger \circ \iota$ as $u^\dagger$.

4.1.1. One-dimensional manifold. For the one-dimensional elliptic inverse problem, we first augmented the given manifold point cloud data, adding 10 ghost points at each boundary point. To model the interior term from the proposed prior, we constructed a self-tuned graph Laplacian using both the manifold point cloud and ghost points. After obtaining the graph Laplacian, we truncated it to obtain a submatrix whose indices correspond to the manifold point cloud elements. Recall that this matrix was referred to as a truncated graph Laplacian in section 3.2. We used two nearest neighbors to construct the self-tuned graph Laplacian. For the two boundary terms, we used the GPDM algorithm to obtain two harmonic functions whose boundary

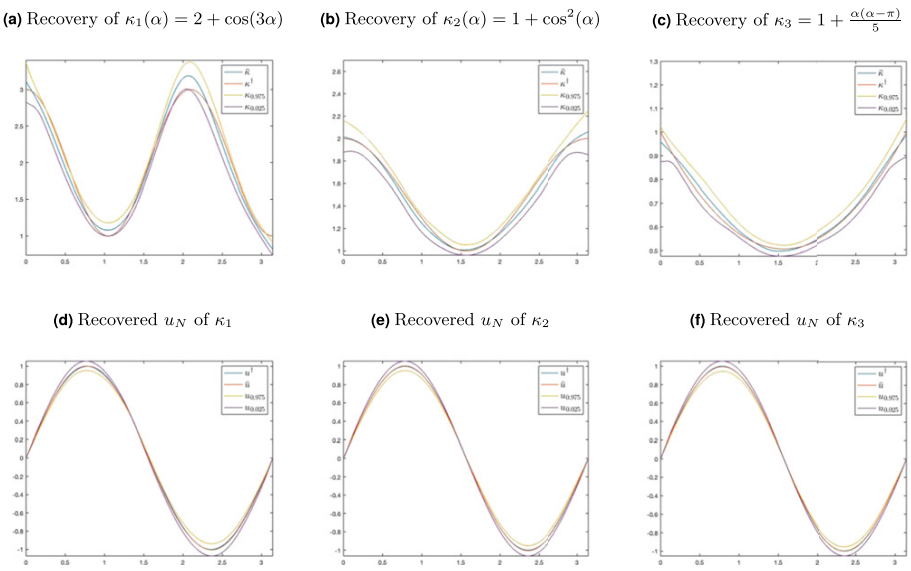

Figure 2. One-dimensional elliptic inverse problem on a semi-ellipse: top row: reconstruction of κ. Bottom row: solution of an elliptic PDE corresponding to the κ given right above.

values were either one or zero. These harmonic functions will allow us to model boundary values, i.e. values at $\alpha = 0, \pi$, of the diffusion coefficient. When constructing the weighted Laplacian matrix, we used 51 closest points in (2.18) and chose the value of ϵ which attained the maximum slope of $\log(T(\epsilon)) \approx \frac{1}{2}$.

We set the smoothness prior parameter to be $s = 4$, and the inverse length-scale parameter to be $\tau = 0.2$, where the choice of these values is based on empirical experiments. Hierarchical Bayesian formulations to learn these parameters could be considered [39]. In practice, instead of using all eigenvectors of the truncated graph Laplacian, one can use a sufficiently large subset of eigenvectors. In our simulation study, we used 20 eigenvectors of the truncated graph Laplacian to represent the interior term. Furthermore, to attain an acceptance rate between 40 and 60 percent, we used $\zeta = 0.01$ for the graph pCN algorithm. We computed a total of 10 000 MCMC iterations with an initial 5000 burn-in period. The results are shown in figure 2, where three different choices of diffusion coefficients are considered.

Observe from figure 2 that the three true diffusion coefficients considered lie, for the most part, inside of the 95 percent credible intervals. Moreover, the PDE solutions obtained using the recovered coefficients were all very close to the PDE solution with the true coefficients. To showcase the flexibility of our prior compared to the one proposed in [26], additional numerical experiments were conducted. While using the same forward map approximation given by the GPDM algorithm, we employed priors proposed in [26] where the graph Laplacian is constructed solely from the point cloud on the manifold. We first present the recovery results for $\kappa_1(\alpha) = 2 + \cos(3\alpha)$ and $\kappa_2(\alpha) = 1 + \cos^2(\alpha)$ with the same semi-ellipse manifold as in figure 2. All the parameter values for priors and pCN algorithms were set to be identical as before. The results are shown in figure 3.

(a) Recovery of $\kappa_1(\alpha) = 2 + \cos(3\alpha)$ **(b)** Recovery of $\kappa_2(\alpha) = 1 + \cos^2(\alpha)$

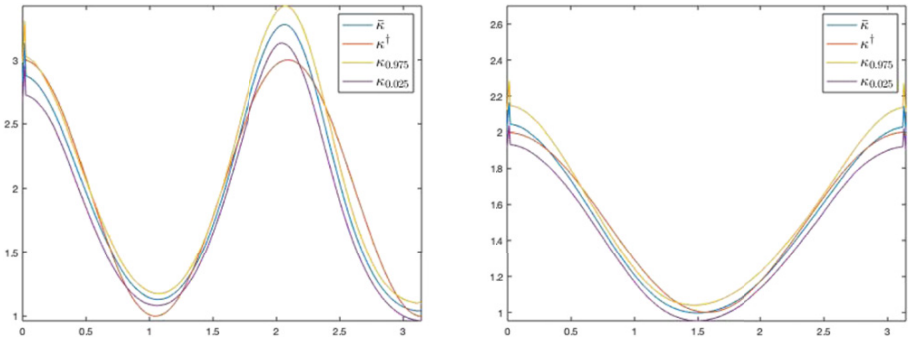

Figure 3. Elliptic inverse problem on a semi-ellipse with Matérn-type prior in [26].

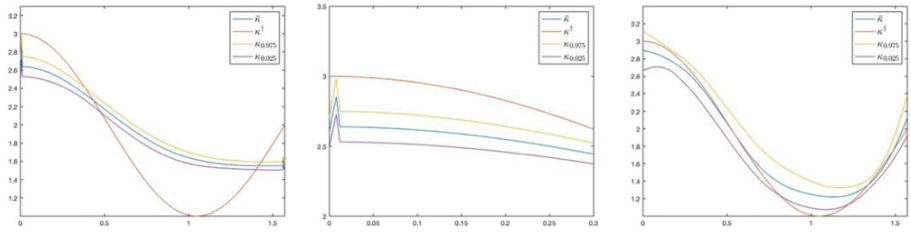

Figure 4. Comparison of the performance of two priors in the reconstruction of $\kappa_1(\alpha) = 2 + \cos(3\alpha)$ with $\alpha \in \left[0, \frac{\pi}{2}\right]$. Left: using the prior proposed in [26]. Middle: zoom-in of the left figure near the boundary. Right: using our proposed prior.

Figures 3(a) and (b) show that the recovered diffusion coefficients using the priors in [26] have artificial spikes near the boundary. The prior in [26] is only effective when the underlying manifold is closed, in which case the graph Laplacian used in [26] approximates the Laplace–Beltrami operator on the underlying manifold. However, when the underlying manifold has a boundary, the prior in [26] leads to the formation of spikes near the boundaries and it does not allow for flexible modeling of boundary conditions. To illustrate this point further, we performed the Bayesian inversion procedure for the one-dimensional elliptic equation on an ellipse restricted to the first quadrant (i.e. $\alpha \in \left[0, \frac{\pi}{2}\right]$) instead of the semi-ellipse. The true diffusion coefficient was set to be $\kappa = 2 + \cos(3\alpha)$. All the parameter values were again the same as before. Figure 4 demonstrates the flexibility of our proposed prior, while the reconstruction using the prior in [26] exhibits artifacts near the boundary and appears to incorrectly suggest a homogeneous Neumann boundary condition for κ. In the next subsection we provide further understanding of these phenomena by illustrating the different terms involved in the definition of our prior and the one in [26] in a two-dimensional setting.

4.1.2. Two-dimensional manifold. Similarly, as in the one-dimensional elliptic PDE, we augmented the given point cloud with 216 ghost points for each boundary. To define the boundary part of our prior, for each boundary (which is a circle in this example) we construct a self-tuned graph Laplacian using only the observations on the boundary with two-nearest neighbors. Then

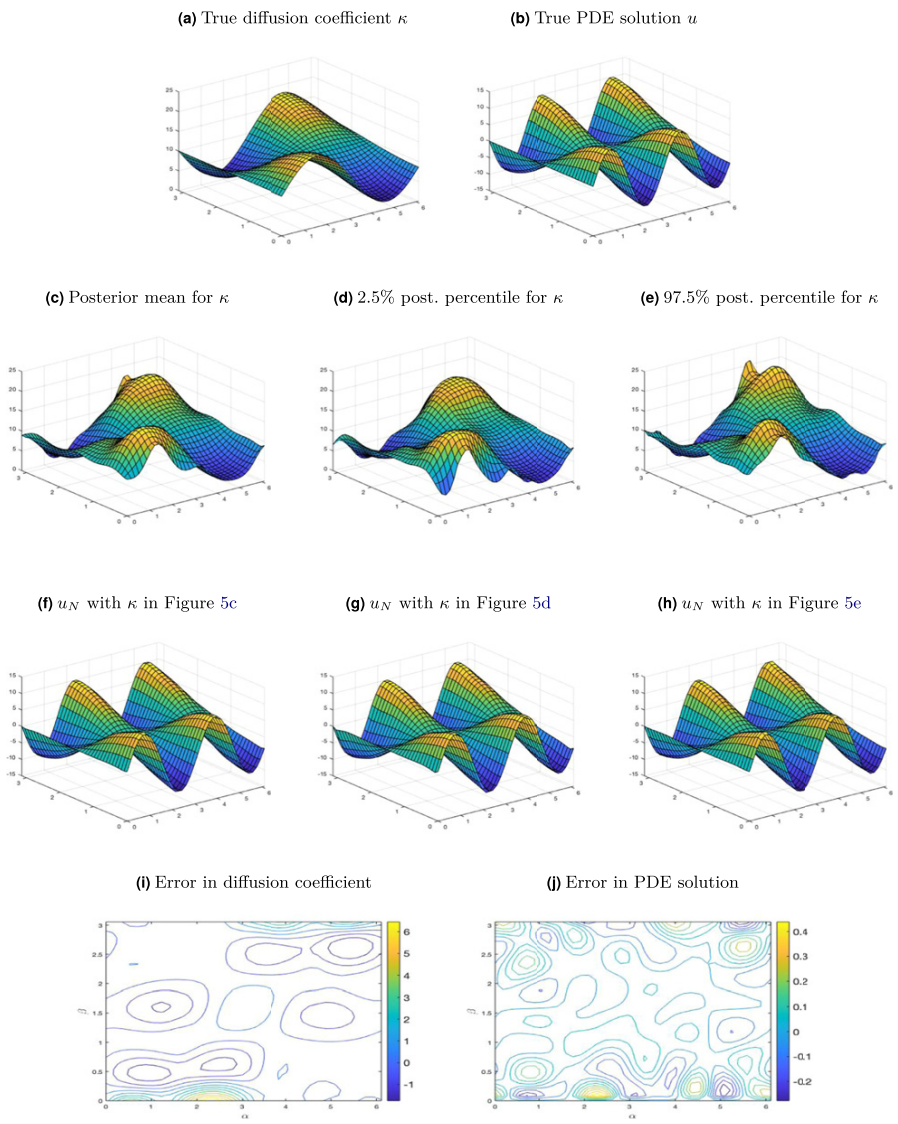

Figure 5. Elliptic inverse problem on a semi-torus for $\kappa(\alpha, \beta) = 10 + 8\sin(\alpha)\cos(\beta)$.

we obtain the 10 eigenvectors corresponding to the smallest 10 eigenvalues for each boundary. These eigenvectors will serve the role of discretized basis functions for each boundary. Solving 20 different Laplace-type equations in (3.11) using the GPDM algorithm, one can obtain 10 basis-like functions for each boundary, which would correspond to $\{\psi_{1,\ell}^{(N)}\}_{\ell=1}^{10}$ and $\{\psi_{2,\ell}^{(N)}\}_{\ell=1}^{10}$ in (3.13). These basis-like functions allow flexible model of functions in the vicinity of each boundary curve. In particular, in our example the true diffusion coefficient was set to $\kappa(\alpha, \beta) = 10 + 8\sin(\alpha)\cos(\beta)$. Therefore, the values we would like to capture along each boundary would be $\kappa(\alpha, 0) = 10 + 8\sin(\alpha)$ and $\kappa(\alpha, \pi) = 10 - 8\sin(\alpha)$. When constructing

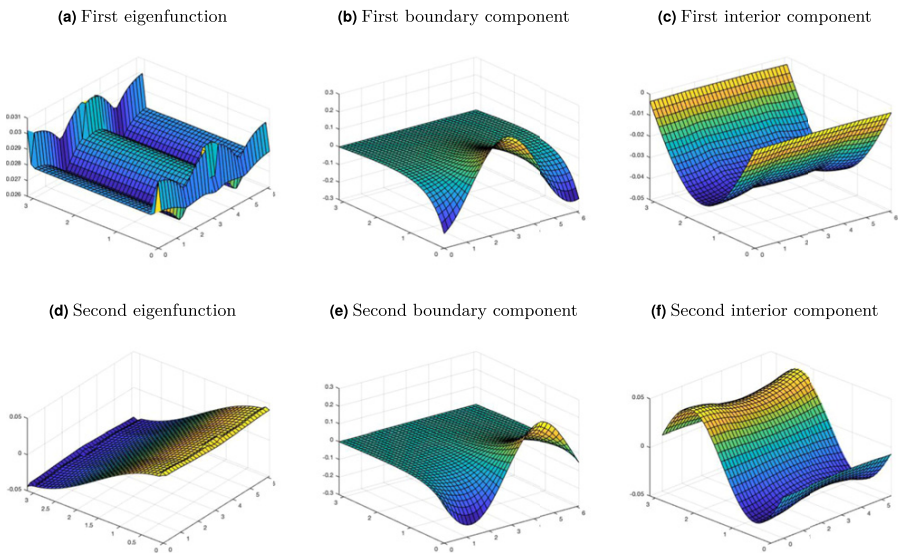

Figure 6. Representation of terms used to define prior draws on a semi-torus. Left column: first two eigenfunctions of the covariance matrix for the prior proposed by [26]. Middle column: first two (excluding the constant one) boundary components in our proposed prior. Right column: first two interior terms in our proposed prior.

the weighted Laplacian matrix, we used 128 closest points in (2.18) and chose the value of ϵ which attained the maximum slope of $\log(T(\epsilon)) \approx 1$.

For the interior part of the proposed prior, analogously to the one-dimensional case, we truncated a self-tuned graph Laplacian constructed using the augmented dataset and obtained a submatrix whose indices correspond to the elements in the point cloud. Taking eigenvalues and eigenvectors of this truncated graph Laplacian would give $(\lambda_n^{(N)}, \phi_n^{(N)})$ in (3.13). We used four-nearest neighbors to construct the self-tuned graph Laplacian. We set the smoothness prior parameter to be $s = 4$, and the inverse length-scale parameter to be $\tau = 0.24$, where the choice of these values is based on empirical experiments. To attain an acceptance rate between 40 and 60 percent, we used $\zeta = 0.001$ for the graph pCN algorithm. A total of 150 000 MCMC iterations with initial 75 000 burn-in iterations were run to obtain our results.

Starting from the top row of figure 5, the panels in the first row represent the true diffusion coefficient and the solution of the elliptic PDE based on the true diffusion coefficient. The second row represents, from left to right, the posterior mean of the MCMC samples, 2.5 percentile of MCMC samples and 97.5 percentile of MCMC samples. In other words, the second row portrays the credible interval for the true diffusion coefficient function. The third row corresponds to the solution of the elliptic PDE equation based on the diffusion coefficients given in the second row. The last two plots in the fourth row depict the difference between the true diffusion coefficient and the posterior mean of the MCMC samples, and the difference between the true solution with the one based on the posterior mean of the MCMC samples.

For the two-dimensional elliptic inverse problem, there is a larger error in the recovery of the true diffusion coefficient. This is unsurprising, since as one can see from the solutions corresponding to the MCMC samples of the diffusion coefficients, one can find several different diffusion coefficients that lead to approximate solutions that are close to the solution of the elliptic PDE, which is a manifestation of the ill-posedness of this elliptic inverse problem.

To illustrate the strength of the proposed prior relative to [26], figure 6 includes surface plots of (i) the first two eigenfunctions of graph Laplacian used to define the prior in [26]; (ii) the first two eigenfunctions, i.e., ϕ_1^N, ϕ_2^N of the truncated graph Laplacian used to define the interior term of our prior; and (iii) the boundary terms $\psi_{1,2}^N, \psi_{1,3}^N$ in (3.13). As one can see in the leftmost column in figure 6 the graph Laplacian used in the prior construction in [26] is not appropriate for manifolds with boundary. In particular, the spikes in the first eigenfunction can be explained by the use of a symmetric graph Laplacian and the fact that the degree of nodes close to the boundary is significantly different than the degree of nodes in the interior. However, the samples we propose do not possess such undesirable behavior as the boundary values are solely modeled by boundary components which are in the middle column of figure 6. The interior term is modeled by superposition of eigenfunctions of the homogeneous Dirichlet eigenvalue problem (rightmost column in figure 6) approximated using a truncated graph Laplacian.

4.2. Heat inverse problem

For the one-dimensional numerical simulations of heat inversion, given an initial heat function u_0, observations were obtained by

$$y_n = \tilde{u}(\alpha_n, t^*) + \eta_n, \qquad \alpha_n = \iota^{-1}(x_n), \quad n = 1, \ldots, N,$$

where $\eta_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 0.01)$ and $\tilde{u}$ represents an approximate solution of the heat equation given in (3.2). We assume that the boundary is given by $\{x_1, x_N\}$. For the one-dimensional semi-ellipse, we used the following explicit formula to compute the approximate solution over the point cloud at time $t^* > 0$, denoted by $\tilde{u}_N = (\tilde{u}(\alpha_1, t^*), \ldots, \tilde{u}(\alpha_N, t^*))$,

$$\tilde{u}_N := \sum_{n=1}^{N} \langle \phi_n^{(N)}, \psi_3^{(N)} \rangle \mathrm{e}^{-\lambda_n^{(N)} t^*} \phi_n^{(N)} + u_0(\alpha_1)\psi_1^{(N)} + u_0(\alpha_N)\psi_2^{(N)},$$

where $(\lambda_n^{(N)}, \phi_n^{(N)})_{n=1}^N$ are eigenpairs of the truncated graph Laplacian constructed using point cloud $\{x_n\}_{n=1}^N$ with two boundary points $\{x_1, x_N\}$. The definitions of $\psi_1^{(N)}$, $\psi_2^{(N)}$ and $\psi_3^{(N)}$ are given in section 3.2. Recall that $\psi_1^{(N)}$, $\psi_2^{(N)}$ are responsible for modeling values of the parameter at $\{x_1, x_N\}$ while $\psi_3^{(N)}$ determines the interior values of the parameter.

Similarly, for the two-dimensional semi-torus, given an initial heat function u_0, observations were obtained by

$$y_n = \tilde{u}((\alpha_{n_1}, \beta_{n_2}), t^*) + \eta_n, \quad (\alpha_{n_1}, \beta_{n_2}) =: \iota^{-1}(x_n), \ n = 1, \ldots, N,$$

where $\eta_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 0.01)$ and $\tilde{u}$ represents the approximate solution for (3.2). For the two-dimensional setting, the approximate solution

$$\tilde{u}_N = (\tilde{u}(\alpha_1, \beta_1, t^*), \ldots, \tilde{u}(\alpha_{N_1}, \beta_{N_2}, t^*))$$

over the point cloud at time $t^* > 0$, with $N = N_1 \times N_2$, is given by

$$\tilde{u}_N := \sum_{n=1}^{N} \langle \phi_n^{(N)}, \psi_3^{(N)} \rangle \mathrm{e}^{-\lambda_n^{(N)} t^*} \phi_n^{(N)} + \sum_{\ell=1}^{L} \tilde{\mu}_{1,\ell} \psi_{1,\ell}^{(N)} + \sum_{\ell=1}^{L} \tilde{\mu}_{2,\ell} \psi_{2,\ell}^{(N)},$$

where the leftmost term in the right-hand side is defined similarly as in the one-dimensional manifold setting. Likewise, the remaining finite summation terms involving $\{\phi_n^{(N)}\}_{n=1}^N$, $\{\psi_{1,l}^{(N)}\}_{l=1}^L$ and $\{\psi_{2,l}^{(N)}\}_{l=1}^L$ are responsible for modeling parameter values along each boundary curve. For the implementation, the true coefficients $\{\tilde{\mu}_{1,\ell}\}_{l=1}^L$ and $\{\tilde{\mu}_{1,\ell}\}_{l=1}^L$ were

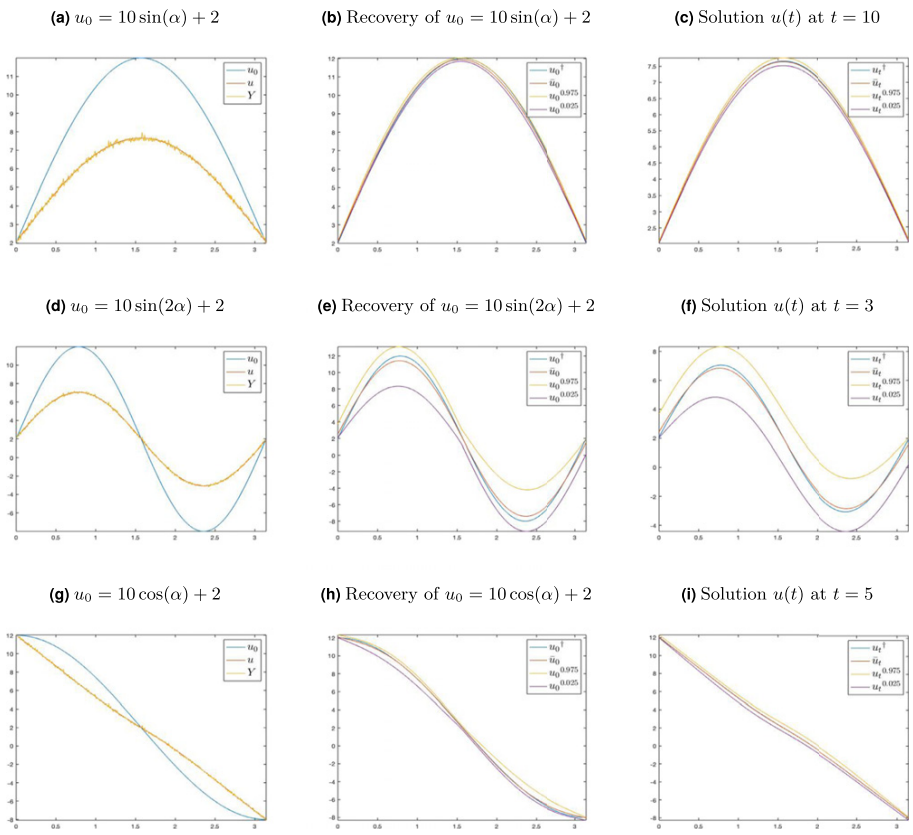

Figure 7. Heat inverse problem on a semi-ellipse.

obtained by regressing $u_0^{(N)} = (u_0(x_1), \ldots, u_0(x_N))$ on $\{\phi_n^{(N)}\}_{n=1}^N$, $\{\psi_{1,l}^{(N)}\}_{l=1}^L$ and $\{\psi_{2,l}^{(N)}\}_{l=1}^L$. In our numerical experiments, we constructed 10 basis-like functions for each boundary of the two-dimensional semi-torus, i.e., $L = 10$, and used 20 eigenvectors of the truncated graph Laplacian.

4.2.1. One-dimensional manifold. The prior construction was exactly the same as in the one-dimensional elliptic problem. And hence, the boundary values were only defined on two points i.e. values at $\alpha = 0, \pi$. We set the smoothness prior parameter to be $s = 6$, and the inverse length-scale parameter to be $\tau = 0.3$, where the choice of these values is based on empirical experiments. Furthermore, to attain an MCMC acceptance rate of roughly 50 percent, we used $\zeta = 0.005$ for the graph pCN algorithm. A total of 20 000 MCMC iterations with initial 10 000 burn-in iterations were run to obtain our results. For the heat equation, the degree of ill-posedness of the inverse problem is closely related to the time $t^* > 0$ at which we observe the data. The smaller the time $t^* > 0$ is, the easier the inversion.

We considered three different choices of initial heat functions. Starting from the left-most column of figure 7, plots in each column respectively represent initial heat function with observed data, true initial function with the posterior mean/2.5th and 97.5th percentile of MCMC samples, and the true solution with the solutions corresponding to the posterior mean,

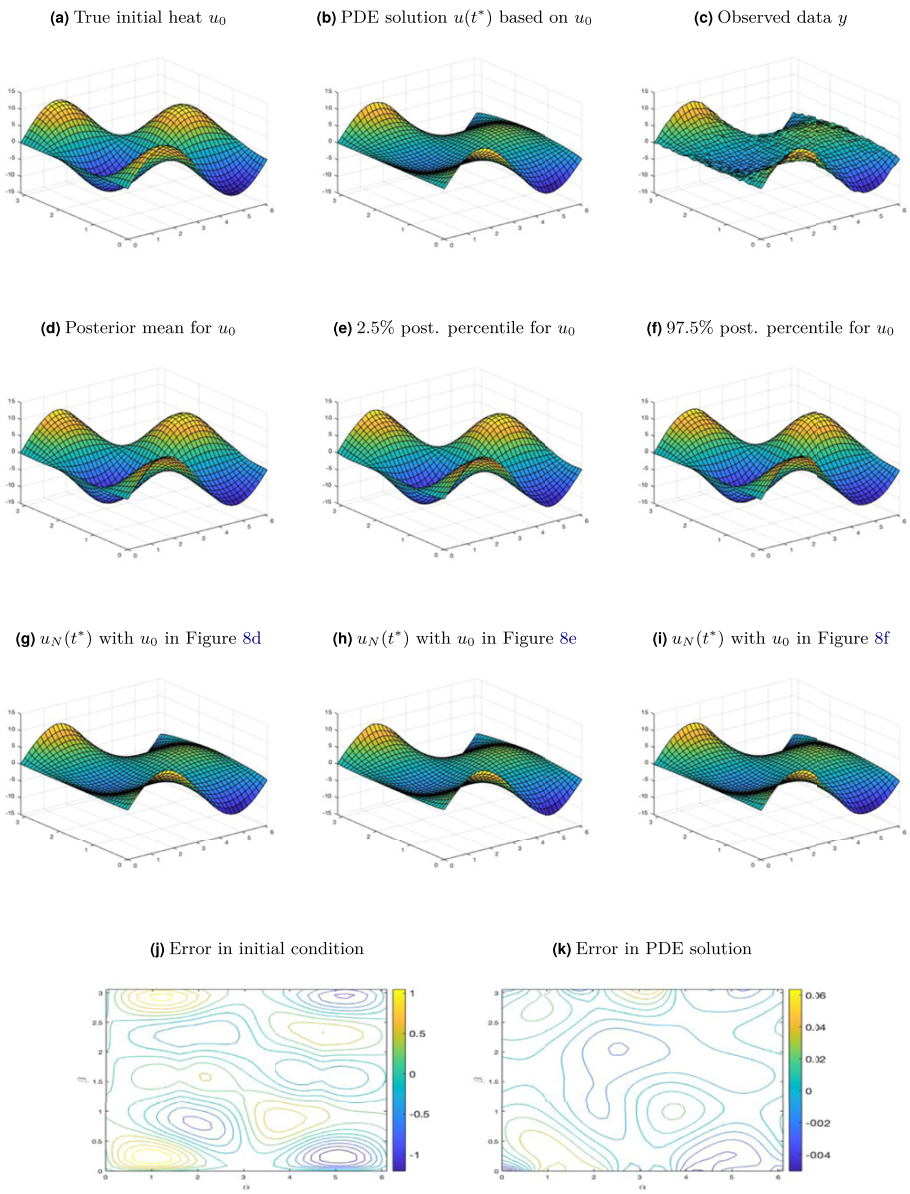

Figure 8. Heat inverse problem on a semi-torus: $u_0(\alpha, \beta) = 10\sin(\alpha)\cos(2\beta)$ with $\tau = 0.3$, $t^* = 5$.

2.5th and 97.5 percentile of MCMC samples. Specifically, the leftmost column in figure 7 represents the initial heat function u_0 with its corresponding noise-free observation u and noisy observation y. The middle column represents the true initial function $u_0^\dagger$ with the posterior mean estimate $\bar{u}_0$, 2.5th percentile $u_0^{0.025}$, and 97.5th percentile $u_0^{0.975}$ of MCMC samples. The rightmost column portrays the approximate solution based on the true initial heat function at

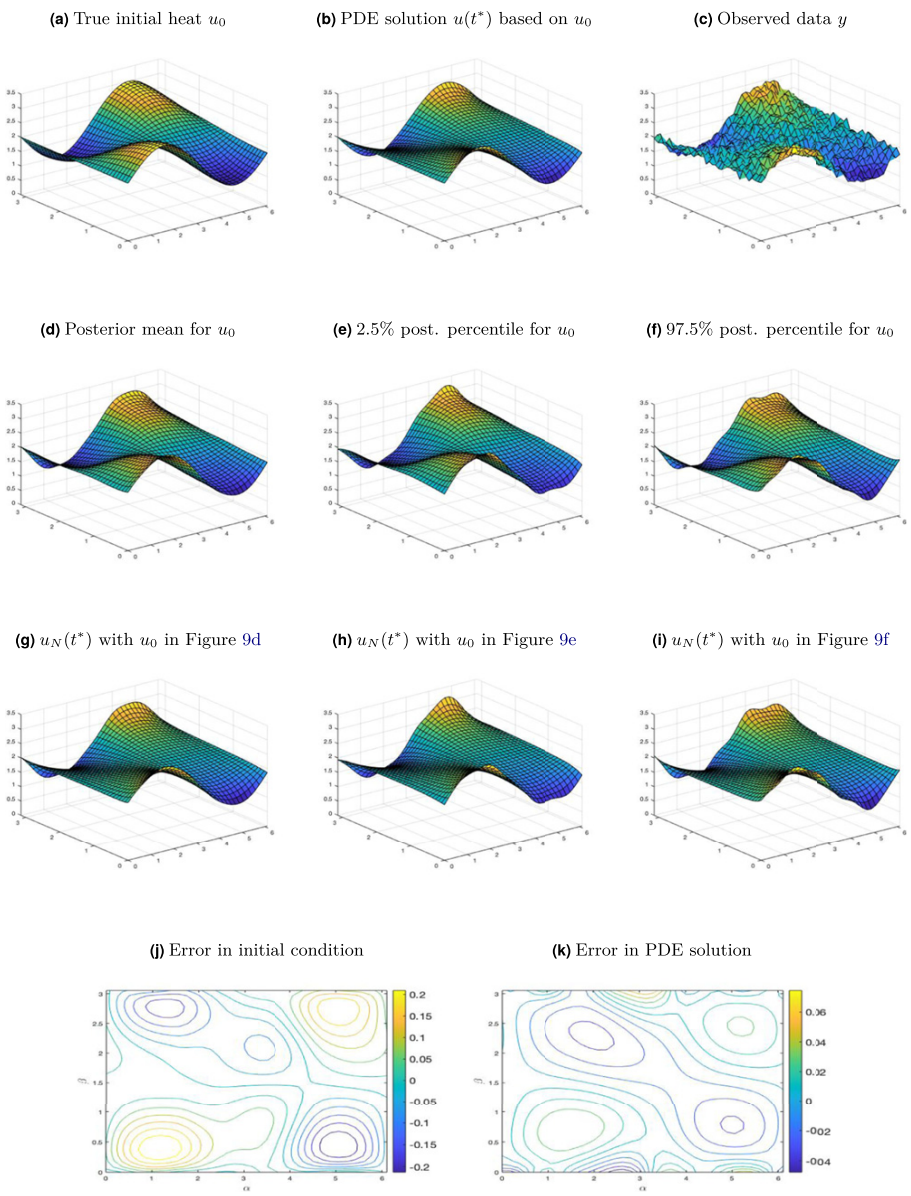

Figure 9. Heat inverse problem on a semi-torus: $u_0(\alpha, \beta) = 2 + \sin(\alpha)\cos(\beta)$ with $\tau = 0.012$, $t^* = 5$.

time t, denoted by $u_t^\dagger$ with approximate solutions based on the posterior mean estimate, 2.5th percentile and 97.5th percentile, respectively denoted by $\bar{u}_t$, $u_t^{0.025}$ and $u_t^{0.975}$.

Figure 7 shows that all of the true initial heat functions were captured inside of the 95 percent credible intervals for all three cases. Moreover, the corresponding solutions were very close to the true solution.

4.2.2. Two-dimensional manifold. The practical implementation of the two-dimensional prior was analogous to that of the two-dimensional elliptic inverse problem. We set the smoothness prior parameter to be $s = 4$, and the inverse length-scale parameter to be $\tau = 0.3$ or 0.012, where the choice of these values is based on empirical experiments. Furthermore, to attain an MCMC acceptance rate of roughly 50 percent, we used $\zeta = 0.006$ for the graph pCN algorithm. A total of 100 000 number of MCMC iterations with initial 50 000 burn-in iterations were run to obtain these results. We considered two initial heat functions: $u_0(\alpha, \beta) = 10\sin(\alpha)\cos(2\beta)$ in figure 8 and $u_0(\alpha, \beta) = 2 + \sin(\alpha)\cos(\beta)$ in figure 9. For $u_0(\alpha, \beta) = 10\sin(\alpha)\cos(2\beta)$, the initial heat function values along each boundary curve we would like to recover is $10\sin(\alpha)$. Similarly for $u_0(\alpha, \beta) = 2 + \sin(\alpha)\cos(\beta)$, our goal is to recover an initial heat function with values $2 + \sin(\alpha)$ or $2 - \sin(\alpha)$ corresponding to each boundary.

Starting from the top row of figures 8 and 9, the panels in the first row represent, from left to right, the true initial heat function, approximate solution of the heat equation based on the true heat initial function, noise-incorporated approximate solution, i.e. observation data. The second row represents from left to right, posterior mean of the MCMC samples, 2.5 percentile of MCMC samples, and 97.5 percentile of MCMC samples. In other words, the second row portrays the credible interval for the true initial heat function. The third row corresponds to the approximate solution of the heat equation based on the initial heat functions given in the second row. Finally, the last two plots in the fourth row depict the difference between the true initial function and the posterior mean of the MCMC samples, and the difference between the approximate solution based on the true initial function with the one based on the posterior mean of the MCMC samples. Figures 8 and 9 show that our proposed methodology led to reasonably successful recovery of the parameter of interest.

5. Conclusions and open directions

In this paper, we developed graph-based Matérn priors for solving Bayesian inverse problems on manifolds with boundaries. Our idea is to extend the Matérn priors introduced in [26], developed for elliptic PDEs on closed manifolds, by representing the boundary conditions via a set of functions obtained from solving Laplace equations on manifolds with appropriate Dirichlet boundary conditions. To solve PDEs on manifolds with boundaries, we employed the recently developed GPDM algorithm [27], which uses fictitious ghost points to remove the bias induced by integrating radial type kernels near the boundaries.

We validated this approach on two test problems. The first problem is an inversion of the diffusion coefficient of an elliptic PDE from the solution of the PDE corrupted by noise. The second problem is an inversion of the initial condition of a heat equation from noisy observation of the solution at a positive time. Based on our numerical simulations, we found positive results given the ill-posedness of the inverse problems we considered.

While the proposed method produces encouraging results, there are many open questions. First, we should point out that while the computational cost is independent of the ambient dimension, it scales exponentially as a function of intrinsic dimension (see [27] for detailed convergence rates for the forward maps). Since Bayesian inversion often requires to evaluate the forward map numerous times, it is of interest to improve on our graph-based approximations of the forward map by using computationally cheaper surrogate forward models and/or faster numerical solvers. Second, the method represents the hidden variables by a vector whose components are the function values of the variable of interest (e.g., diffusion coefficients) on the given point clouds. How to extend this to other points on the domain is of practical interest. Beyond these practical considerations, it is also of interest to understand the theoretical aspect

of such an approach, especially the effects of boundaries, extending the theoretical convergence result for closed manifolds in [26].

Acknowledgments

The research of JH was partially supported under the NSF Grant DMS-1854299. DSA is thankful for the support of NSF and NGA through the Grant DMS-2027056. The work of DSA was also partially supported by the NSF Grant DMS-1912818/1912802.

Data availability statement

The data that support the findings of this study are available upon reasonable request from the authors.

ORCID iDs

Hwanwoo Kim https://orcid.org/0000-0001-9875-6444
Daniel Sanz-Alonso https://orcid.org/0000-0002-5022-864X

References

[1] Beck J V, Blackwell B and Clair C R S Jr 1985 *Inverse Heat Conduction: Ill-Posed Problems* (Hoboken, NJ: Wiley)
[2] Berry T and Harlim J 2016 Variable bandwidth diffusion kernels *Appl. Comput. Harmon. Anal.* **40** 68–96
[3] Bertalmıo M, Cheng L-T, Osher S and Sapiro G 2001 Variational problems and partial differential equations on implicit surfaces *J. Comput. Phys.* **174** 759–80
[4] Bertozzi A L, Luo X, Stuart A M and Zygalakis K C 2018 Uncertainty quantification in graph-based classification of high dimensional data *SIAM/ASA J. Uncertain. Quantification* **6** 568–95
[5] Beskos A, Roberts G, Stuart A and Voss J 2008 MCMC methods for diffusion bridges *Stoch. Dyn.* **8** 319–50
[6] Bigoni D, Chen Y, Garcia Trillos N, Marzouk Y and Sanz-Alonso D 2020 Data-driven forward discretizations for Bayesian inversion *Inverse Problems* **36** 105008
[7] Bolin D 2014 Spatial Matérn fields driven by non-Gaussian noise *Scand. J. Stat.* **41** 557–79
[8] Bolin D and Kirchner K 2019 The rational SPDE approach for Gaussian random fields with general smoothness *J. Comput. Graph. Stat.* **29** 274–85
[9] Bolin D, Kirchner K and Kovács M 2018 Weak convergence of Galerkin approximations for fractional elliptic stochastic PDEs with spatial white noise *BIT Numer. Math.* **58** 881–906
[10] Bolin D, Kirchner K and Kovács M 2020 Numerical solution of fractional elliptic stochastic PDEs with spatial white noise *IMA J. Numer. Anal.* **40** 1051–73
[11] Calvetti D, Kaipio J P and Somersalo E 2006 Aristotelian prior boundary conditions *Int. J. Math. Comput. Sci.* **1** 63–81
[12] Canzani Y 2013 Analysis on manifolds via the Laplacian Lecture notes available at: http://math.harvard.edu/canzani/docs/Laplacian.pdf
[13] Chada N K, Chen Y and Sanz-Alonso D 2020 Iterative ensemble Kalman methods: a unified perspective with some new variants (arXiv:2010.13299)
[14] Coifman R R and Lafon S 2006 Diffusion maps *Appl. Comput. Harmon. Anal.* **21** 5–30
[15] Colbois B, El Soufi A and Savo A 2015 Eigenvalues of the Laplacian on a compact manifold with density *Commun. Anal. Geom.* **23** 639–70

[16] Cotter S, Dashti M and Stuart A M 2010 MCMC methods for functions: modifying old algorithms to make them faster *SIAM J. Numer. Anal.* **48** 322–45
[17] Daon Y and Stadler G 2018 Mitigating the influence of the boundary on PDE-based covariance operators *Inverse Problems Imaging* **12** 1083–102
[18] Dashti M and Stuart A M 2017 Bayesian approach to inverse problems *Handbook of Uncertainty Quantification* (Cham: Springer) pp 311–428
[19] Dziuk G and Elliott C M 2013 Finite element methods for surface PDEs *Acta Numer.* **22** 289–396
[20] Elliott C M and Stinner B 2010 Modeling and computation of two phase geometric biomembranes using surface finite elements *J. Comput. Phys.* **229** 6585–612
[21] Franklin J N 1970 Well-posed stochastic extensions of ill-posed linear problems *J. Math. Anal. Appl.* **31** 682–716
[22] Fuselier E J and Wright G B 2013 A high-order kernel method for diffusion and reaction-diffusion equations on surfaces *J. Sci. Comput.* **56** 535–65
[23] Garcia Trillos N and Sanz-Alonso D 2017 The Bayesian formulation and well-posedness of fractional elliptic inverse problems *Inverse Problems* **33** 065006
[24] García Trillos N and Sanz-Alonso D 2018 Continuum limits of posteriors in graph Bayesian inverse problems *SIAM J. Math. Anal.* **50** 4020–40
[25] Garcia Trillos N, Kaplan Z, Samakhoana T and Sanz-Alonso D 2020 On the consistency of graph-based Bayesian semi-supervised learning and the scalability of sampling algorithms *J. Mach. Learn. Res.* **21** 1–47
[26] Harlim J, Sanz-Alonso D and Yang R 2020 Kernel methods for Bayesian elliptic inverse problems on manifolds *SIAM/ASA J. Uncertain. Quantification* **8** 1414–45
[27] Jiang S W and Harlim J 2020 Ghost point diffusion maps for solving elliptic PDEs on manifolds with classical boundary conditions *Commun. Pure Appl. Math.* (accepted) https://doi.org/10.1002/cpa.22035
[28] Kaipio J and Somersalo E 2006 *Statistical and Computational Inverse Problems* vol 160 (New York: Springer)
[29] Khristenko U, Scarabosio L, Swierczynski P, Ullmann E and Wohlmuth B 2019 Analysis of boundary effects on PDE-based sampling of Whittle–Matérn random fields *SIAM/ASA J. Uncertain. Quantification* **7** 948–74
[30] LeVeque R J 2007 *Finite Difference Methods for Ordinary and Partial Differential Equations: Steady-State and Time-dependent Problems* vol 98 (Philadelphia, PA: SIAM)
[31] Lindgren F, Rue H and Lindström J 2011 An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach *J. R. Stat. Soc.* B **73** 423–98
[32] Macdonald C B and Ruuth S J 2010 The implicit closest point method for the numerical solution of partial differential equations on surfaces *SIAM J. Sci. Comput.* **31** 4330–50
[33] Matérn B 2013 *Spatial Variation* 2nd edn vol 36 (New York: Springer)
[34] Mémoli F, Sapiro G and Thompson P 2004 Implicit brain imaging *NeuroImage* **23** S179–88
[35] Peoples J W and Harlim J 2021 Spectral convergence of symmetrized graph Laplacian on manifolds with boundary (arXiv:2110.06988)
[36] Rauter M and Tuković Ž 2018 A finite area scheme for shallow granular flows on three-dimensional surfaces *Comput. Fluids* **166** 184–99
[37] Roininen L, Girolami M, Lasanen S and Markkanen M 2019 Hyperpriors for Matérn fields with applications in Bayesian inversion *Inverse Problems Imaging* **13** 1–29
[38] Ruuth S J and Merriman B 2008 A simple embedding method for solving partial differential equations on surfaces *J. Comput. Phys.* **227** 1943–61
[39] Sanz-Alonso D and Yang R 2020 The SPDE approach to Matérn fields: graph representations (arXiv:2004.08000)
[40] Sanz-Alonso D and Yang R 2020 Unlabeled data help in graph-based semi-supervised learning: a Bayesian nonparametrics perspective (arXiv:2008.11809)
[41] Sanz-Alonso D and Yang R 2021 Finite element representations of Gaussian processes: balancing numerical and statistical accuracy (arXiv:2109.02777)
[42] Sanz-Alonso D, Stuart A M and Taeb A 2018 Inverse problems and data assimilation (arXiv:1810.06191)
[43] Stein M L 2012 *Interpolation of Spatial Data: Some Theory for Kriging* (New York: Springer)
[44] Stuart A M 2010 Inverse problems: a Bayesian perspective *Acta Numer.* **19** 451–559

[45] Tarantola A 2015 *Inverse Problem Theory and Methods for Model Parameter Estimation* (Philadelphia, PA: SIAM)
[46] Thiede E H, Giannakis D, Dinner A R and Weare J 2019 Galerkin approximation of dynamical quantities using trajectory data *J. Chem. Phys.* **150** 244111
[47] Virga E G 2018 *Variational Theories for Liquid Crystals* (London: Chapman and Hall)
[48] Von Luxburg U 2007 A tutorial on spectral clustering *Stat. Comput.* **17** 395–416
[49] Wiens A, Nychka D and Kleiber W 2020 Modeling spatial data using local likelihood estimation and a Matérn to spatial autoregressive translation *Environmetrics* **31** e2652
[50] Yan Q, Jiang S W and Harlim J 2021 Kernel-based methods for solving time-dependent advection-diffusion equations on manifolds (arXiv:2105.13835)
[51] Zelnik-Manor L and Perona P 2005 Self-tuning spectral clustering *Advances in Neural Information Processing Systems* pp 1601–8
[52] Zimmerman D A *et al* 1998 A comparison of seven geostatistically based inverse approaches to estimate transmissivities for modeling advective transport by groundwater flow *Water Resour. Res.* **34** 1373–413