

Niu, M., Cheung, P., Lin, L., Dai, Z., Lawrence, N. and Dunson, D. (2019) Intrinsic Gaussian processes on complex constrained domains. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 81(3), pp. 603-627.

There may be differences between this version and the published version. You are advised to consult the publisher's version if you wish to cite from it.

This is the peer reviewed version of the following article Niu, M., Cheung, P., Lin, L., Dai, Z., Lawrence, N. and Dunson, D. (2019) Intrinsic Gaussian processes on complex constrained domains. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 81(3), pp. 603-627, which has been published in final form at <http://dx.doi.org/10.1111/rssb.12320>

This article may be used for non-commercial purposes in accordance with [Wiley Terms and Conditions for Self-Archiving](#).

<http://eprints.gla.ac.uk/202158/>

Deposited on: 07 November 2019

# Intrinsic Gaussian processes on complex constrained domains

Mu Niu

*Centre for Mathematical Sciences, School of Computing, Electronics and Mathematics,  
Plymouth University, Plymouth, UK.*

E-mail: mu.niu@plymouth.ac.uk

Pokman Cheung

*London, UK*

E-mail: pokman@alumni.stanford.edu

Lizhen Lin

*Department of Applied and Computational Mathematics and Statistics, The University of  
Notre Dame, USA.*

E-mail: lizhen.lin@nd.edu

Zhenwen Dai

Neil Lawrence

*The University of Sheffield and Amazon, UK.*

E-mail: zhenwend@amazon.com, lawrennd@amazon.co.uk

David Dunson

*Department of Statistical Science, Duke University*

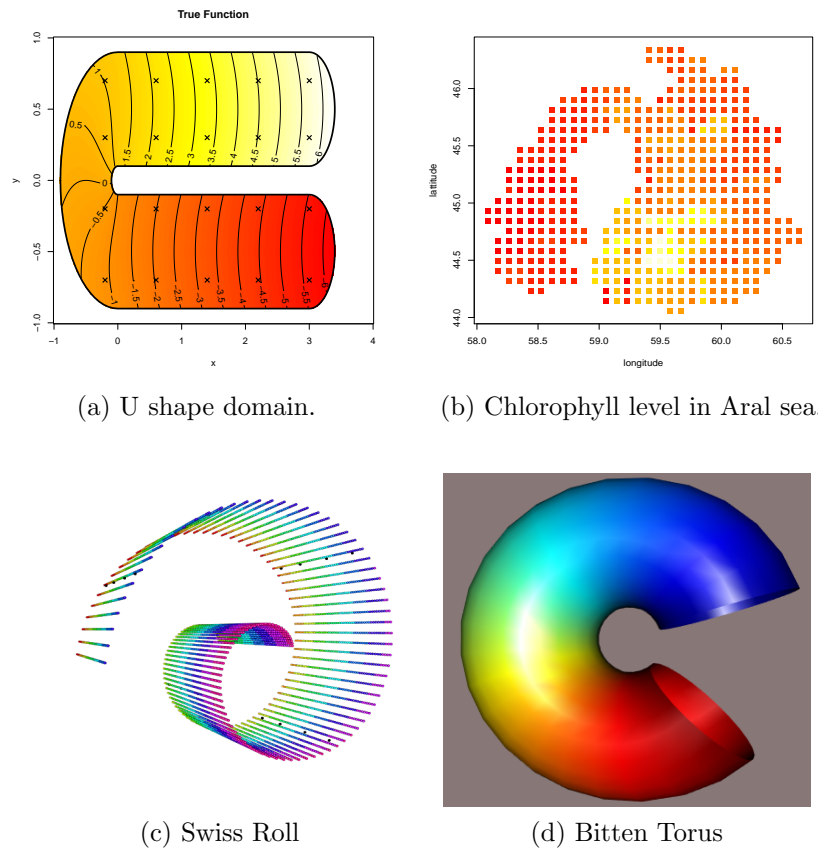
E-mail: dunson@duke.edu

**Summary.** We propose a class of intrinsic Gaussian processes (in-GPs) for interpolation, regression on manifolds with a primary focus on complex constrained domains or irregular-shaped spaces arising as subsets or submanifolds of  $\mathbb{R}$ ,  $\mathbb{R}^2$ ,  $\mathbb{R}^3$  and beyond. For example, in-GPs can accommodate spatial domains arising as complex subsets of Euclidean space. in-GPs respect the potentially complex boundary or interior conditions as well as the intrinsic geometry of the spaces. The key novelty of the proposed approach is to utilise the relationship between heat kernels and the transition density of Brownian motion on manifolds for constructing and approximating valid and computationally feasible covariance kernels. This enables in-GPs to be practically applied in great generality, while existing approaches for smoothing on constrained domains are limited to simple special cases. The broad utilities of the in-GP approach are illustrated through simulation studies and data examples.

*Keywords:* Brownian motion, Constrained domain, Gaussian process, Heat kernel, Intrinsic covariance kernel, Manifold

## 1. Introduction

In recent years it has become commonplace to collect data that are restricted to a complex constrained space. For example, data may be collected in a spatial domain but restricted to a complex or intricately structured region corresponding to a geographic feature, such as a lake. To illustrate, refer to Figure 1(b), which plots satellite measurements on chlorophyll levels in the Aral sea (Wood et al., 2008). In building a spatial map of chlorophyll levels in this sea, and in conducting corresponding inferences and prediction tasks, it is important to take into account the intrinsic geometry of the sea and its complex boundary. Traditional smoothing or modelling methods that do not respect the intrinsic geometry of the space, and in particular the boundary constraints, may produce poor results.



**Fig. 1.** Illustrative examples. In (a), a test function increases smoothly from the lower right to the upper right within the U-shaped boundary. Remote sensed chlorophyll data in the Aral sea from the SeaWifs satellite are shown in (b). The datasets for both (a) and (b) are from Wood et al. (2008). The Swiss Roll in (c) is a spiralling band in a three-dimensional Euclidean space. The Bitten Torus in (d) is constructed by removing the lower right part of a Torus. Synthetic datasets are considered on the surface of (c) and (d). Details for constructing (c) and (d) are available in the Supplementary Material.

For example, it is crucial to take into account the fact that pairs of locations having close Euclidean distance may be intrinsically far apart if separated by a land barrier. Refer in particular to the locations near longitude 58.5 and 59 in the southern region of the map in Figure 1(b). These locations have quite different chlorophyll levels due to the land barrier. However, usual smoothing or modelling approaches that do not account for the boundary would naturally provide close estimates of the chlorophyll level given their close spatial vicinity. The goal of this article is to provide a general methodology that can accommodate not just complex spatial subregions of  $\mathbb{R}^2$  (refer also to the U-shaped constraint in Figure 1(a)) but also complex subregions of higher-dimensional space ( $\mathbb{R}^3$  and beyond) and constraints, such as the Swiss roll in Figure 1(c) and the Bitten Torus in Figure 1(d).

To accommodate modelling on these broad and complex domains, we propose a novel class of *intrinsic* Gaussian processes (in-GPs). An intrinsic Gaussian process refers to a Gaussian process that employs the intrinsic Riemannian geometry of the manifold, including the boundary features and interior conditions. Note that this intrinsic notion of in-GPs is different from the *intrinsic random functions* defined in the seminal work of Matheron (1973), which refer to processes that have a more general form of stationarity than the usual second-order stationarity. in-GPs are designed to be useful in interpolation, regression on manifolds, with a particular emphasis on complex or difficult regions arising as submanifolds. A major challenge in constructing GPs on manifolds is choosing a valid covariance kernel - this is a non-trivial problem and most of the focus has been on developing covariance kernels specific to a particular manifold (e.g., Guinness and Fuentes (2016) consider low-dimensional spheres). Castillo et al. (2014) instead proposed to use randomly rescaled solutions of the heat equation to define a valid covariance kernel for reasonably broad classes of compact manifolds. They additionally provided lower and upper bounds on contraction rates of the resulting posterior measure. Unfortunately, they do not provide a methodology for implementing their approach in practice, and their proposed heat kernels are computationally intractable.

This article proposes a practical and general in-GP methodology, which uses heat kernels as covariance kernels. This is made possible by the major novel contribution of the paper, which is to utilise connections between heat kernels and transition densities of Brownian motion (BM) on manifolds to obtain algorithms for approximating covariance kernels. Specifically, the covariance kernels are approximated by first simulating a Brownian motion on the manifold or complex constrained space of interest, and then evaluating the transition density of the Brownian motion. The heat kernel generalizes the popular and well-studied squared exponential kernel to the manifold and arises from the Laplace operator, thus fully exploiting the intrinsic geometry of the space. We utilize a discretized version of Brownian motion on manifolds (without boundary) or reflective Brownian motion (RBM) for a Riemannian manifold with boundary. RBMs have been defined and thoroughly studied for Euclidean domains (Lions and Sznitman, 1984; Burdzy et al., 2004; Zhou et al., 2017). A  $C^2$  boundary guarantees the existence and uniqueness of a RBM (remark 3 of Zhou et al. (2017)). The transition density functions are the Neumann heat kernels of the domain (Xu, 1984).

Most current methods that can smooth noisy data over regions with a boundary can only be applied to spaces that are *subsets of*  $\mathbb{R}^2$ ; refer to Wood et al. (2008) and Ramsay

(2002). Sangalli et al. (2013) extended Ramsay (2002)’s smoothing spline method to model the brain surface arising as a subset of  $\mathbb{R}^3$  by first discretising the surface. The main idea in this literature is to develop smoothing splines that respect the boundary or interior constraints. Our in-GP approach is fundamentally different conceptually, while also having general applicability beyond two-dimensional examples. Although in-GPs have an increasing computational cost as the dimensionality of the space increases, due to the need to simulate Brownian motion, there is no discretisation of the space unlike methods proposed in Ramsay (2002) and Sangalli et al. (2013).

The related work includes Pelletier (2005) who extend kernel regression to a general Riemannian manifold. Bhattacharya and Dunson (2010) model a response and covariate on a manifold jointly using a Dirichlet process mixture model. The focus of our work on the other hand aims to generalise the powerful GP model to manifold-valued data. Although GPs have been extensively used in statistics and machine learning (see e.g., Rasmussen (2004)), these models can not be directly generalised to model data on manifolds, such as irregular shape spaces, due to the difficulty of constructing valid covariance kernels. Lin et al. (2018) propose *extrinsic covariance kernels* on general manifolds by first embedding the manifolds onto a higher-dimensional Euclidean space, and constructing a covariance kernel on the images after embedding. However, such embeddings are not always available or easy to obtain for complex spaces.

Aumentado-Armstrong and Siddiqi (2017) adopt a related idea of estimating the heat kernel for a sampled manifold (mesh or point cloud) from BM trajectories. No boundary condition is considered. Their approach can be summarised in three steps: 1) Construct a local surface for approximating the manifold using moving least squares. 2) Simulate BM trajectories specific to a local surface using SDEs with a local metric tensor. In order to move across different surfaces or charts, iteratively alternate between BM simulation and project the process onto a local surface. 3) Estimate the heat kernel from the BM trajectories using Kernel Density Estimation (KDE), expressed as a summation of Gaussian kernels. Each term is calculated based on the Euclidean distance between the BM sample paths and the target points. This is problematic when the Euclidean distance is small but the geodesic is big; for example, U shaped domains or regions of a manifold where the curvature is large. Ozakin and Gray (2009) show that the KDE estimator is poor and biased in this context.

In our approach, we estimate the heat kernel by simulating BM sample paths on manifolds with or without a boundary by using a global metric tensor. Our way of constructing the heat kernel estimator is completely different from that of Aumentado-Armstrong and Siddiqi (2017). Instead of using an approximation approach, such as relying on KDE, we develop a direct approach to estimate the heat kernel based on the definition of the BM transition probability on the manifold.

The paper is organised as follows. Section 2 introduces our construction of covariance kernels on manifolds and explores the connection between the heat kernel on a Riemannian manifold and the transition density of Brownian motion on the manifold. This connection is utilised in developing practical algorithms for approximating the heat kernel. Inference under in-GPs using the approximated heat kernel, including an extension to *sparse* in-GPs, is explained in Section 2.4. Properties of the heat kernel estimator are discussed in Section 2.5. Section 3 and 4 illustrate our in-GP methodology with various

simulation and data examples. Section 5 contains a discussion. Computational cost and algorithm complexity of the method are discussed in the Supplementary Material.

## 2. Intrinsic Gaussian process (in-GPs) on manifolds

### 2.1. in-GPs with heat kernel as the covariance kernel

We propose to construct intrinsic Gaussian processes (in-GPs) on manifolds and complex constrained spaces using the heat kernel as the covariance kernel. To be more specific, let  $M$  be a  $d$ -dimensional complete and orientable Riemannian manifold,  $\partial M$  its boundary that is continuous and  $C^1$  almost everywhere,  $\Delta_s$  the Laplacian-Beltrami operator on  $M$ , and  $\delta$  the Dirac delta function. A heat kernel of  $M$  is a smooth function  $K(x, y, t)$  on  $M \times M \times \mathbb{R}^+$  that satisfies the heat equation:

$$\frac{\partial}{\partial t} K_{heat}(s_0, s, t) = \frac{1}{2} \Delta_s K_{heat}(s_0, s, t), \quad \lim_{t \rightarrow 0} K_{heat}(s_0, s, t) = \delta(s_0, s), \quad s_0, s \in M,$$

where the initial condition holds in a distributional sense (Berline et al., 2003). If  $\partial M$  is empty,  $M$  admits a unique heat kernel. If  $\partial M$  is nonempty, multiple heat kernels exist, but the heat kernel becomes unique when we also impose a suitable condition along  $\partial M$ , such as the Neumann boundary condition:

$$\frac{\partial K}{\partial \mathbf{n}} = 0 \quad \text{along } \partial M, \quad (1)$$

where  $\mathbf{n}$  denotes a normal vector of  $\partial M$ .

Alternatively, a heat kernel can be viewed as an operator on  $L^2(M)$ :

$$f \mapsto \int_M K_{heat}(x, y, t) f(y) dy, \quad (2)$$

and as such is equivalent to  $e^{\frac{1}{2}t\Delta} f$ , with  $dy$  the infinitesimal Riemannian volume. The heat kernel is symmetric with  $K_{heat}(x, y, t) = K_{heat}(y, x, t)$ , and is a positive semi-definite kernel on  $M$  for any fixed  $t$ , and thus can serve as a valid covariance kernel for a Gaussian process on  $M$ . The *Neumann boundary condition* can be expressed as no heat transfer across the boundary  $\partial M$ .

If  $M$  is a Euclidean space  $\mathbb{R}^d$ , the heat kernel has a closed form corresponding to a time-varying Gaussian function:

$$K_{heat}(\mathbf{x}_0, \mathbf{x}, t) = \frac{1}{(2\pi t)^{d/2}} \exp \left\{ -\frac{\|\mathbf{x}_0 - \mathbf{x}\|^2}{2t} \right\}, \quad \mathbf{x} \in \mathbb{R}^d.$$

In addition, the heat kernel of  $\mathbb{R}^d$  can be seen as the scaled version of a radial basis function (RBF) kernel (or the popular squared exponential kernel) under different parametrisations:

$$K_{RBF}(\mathbf{x}_0, \mathbf{x}, l) = \sigma_r^2 \exp \left\{ -\frac{\|\mathbf{x}_0 - \mathbf{x}\|^2}{2l^2} \right\}, \quad \mathbf{x} \in \mathbb{R}^d.$$

Letting  $K_{heat}^t(x, y) = K_{heat}(x, y, t)$ , our in-GP uses  $K_{heat}^t(x, y)$  as the covariance kernel, where the time parameter  $t$  of  $K_{heat}$  has the same effect as that of the length-scale parameter  $l$  of  $K_{RBF}$ , controlling the rate of decay of the covariance. By varying the time parameter, one can vary the bumpiness of the realisations of the in-GP.

We use in-GPs to develop nonparametric regression and spatial process models on complex constrained domains  $M$ . Let  $\mathcal{D} = \{(s_i, y_i), i = 1, \dots, n\}$  be the data, with  $n$  the number of observations,  $s_i \in M$  the predictor or location value of observation  $i$  and  $y_i$  a corresponding response variable. We would like to do inferences on how the output  $y$  varies with the input  $s$ , including predicting  $y$  values at new locations  $s_*$  not represented in the training dataset. Assuming Gaussian noise and a simple measurement structure, we let

$$y_i = f(s_i) + \epsilon_i, \quad \epsilon_i \sim \mathcal{N}(0, \sigma_{noise}^2), \quad s_i \in M, \quad (3)$$

where  $\sigma_{noise}^2$  is the variance of the noise. This model can be easily modified to include parametric adjustment for covariates  $x_i$ , and to accommodate non-Gaussian measurements (e.g., having exponential family distributions). However, we focus on the simple Gaussian case without covariates for simplicity in exposition.

Under an in-GP prior for the unknown function  $f : M \rightarrow \mathfrak{R}$ , we have

$$p(\mathbf{f}|s_1, s_2, \dots, s_n) = \mathcal{N}(\mathbf{0}, \Sigma), \quad (4)$$

where  $\mathbf{f}$  is a vector containing the realisations of  $f(\cdot)$  at the sample points  $s_1, \dots, s_n$ ,  $f_i = f(s_i)$ , and  $\Sigma$  is the covariance matrix of these realisations induced by the in-GP covariance kernel. In particular, the entries of  $\Sigma$  are obtained by evaluating the covariance kernel at each pair of locations, that is,

$$\Sigma_{ij} = \sigma_h^2 K_{heat}^t(s_i, s_j). \quad (5)$$

Following standard practice for GPs, this prior distribution is updated with information in the response data to obtain a posterior distribution. Explicit expressions for the resulting predictive distribution are provided in Section 2.3.

**REMARK 2.1.** *We added an additional hyperparameter  $\sigma_h^2$  by rescaling the heat kernel for extra flexibility. The parameter  $\sigma_h^2$  plays a similar role as that of the magnitude parameter of an RBF kernel in the Euclidean space. As mentioned above, the parameter  $t$  replaces the length-scale parameter in an RBF or squared exponential kernel.*

The posterior distribution of  $f$  evaluated at locations  $\mathbf{S} = (s_1, \dots, s_n)$  has the following form:

$$\begin{aligned} f(s)|\mathcal{D} &\sim GP(m_{post}, \Sigma_{post}) \\ m_{post} &= \Sigma_{s, \mathbf{S}} (\Sigma_{\mathbf{S}, \mathbf{S}} + \sigma_{noise}^2 \mathbf{I})^{-1} \mathbf{y} \\ \Sigma_{post} &= \Sigma_{s, s} - \Sigma_{s, \mathbf{S}} (\Sigma_{\mathbf{S}, \mathbf{S}} + \sigma_{noise}^2 \mathbf{I})^{-1} \Sigma_{\mathbf{S}, s}, \end{aligned}$$

where  $\mathbf{y} = (y_1, \dots, y_n)$ .

One of the key challenges for inference using in-GPs with the construction in this section is that *closed form expressions for  $K_{heat}^t$  do not exist for general Riemannian*

*manifolds*. Explicit solutions are available only for very special manifolds, such as Euclidean spaces and spheres. Therefore, for most cases, one can not explicitly evaluate  $K_{heat}^t$  or the corresponding covariance matrices. To overcome this challenge and bypass the need to solve the heat equation directly, we utilise the fact that heat kernels can be interpreted as *transition densities of Brownian motion (BM) on  $M$* . Our recipe is to simulate Brownian motion on  $M$ , numerically evaluate the transition density of the Brownian motion, and then use the evaluation to approximate the kernel  $K_{heat}^t(s_i, s_j)$  for any pair  $(s_i, s_j)$ . The simulation of BM on Riemannian manifolds is discussed in Section 2.2. We also provide some background on Riemannian geometry and stochastic calculus on manifolds.

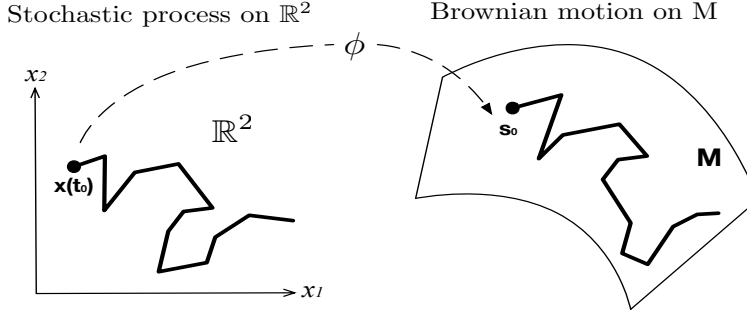
## 2.2. Simulating Brownian motion on manifolds

In order to estimate the transition density of Brownian Motion (BM) on  $M$ , we first need to simulate BM sample paths on  $M$ . Let  $\phi : \mathbb{R}^d \rightarrow M$  be a smooth local parameterisation of  $M$  around  $s_0 \in M$ . A demonstration of  $\phi$  is depicted in Figure 2. Let  $\mathbf{x}(t_0) \in \mathbb{R}^d$  be such that  $\phi(\mathbf{x}(t_0)) = s_0$ . In this paper, we assume the local parameterisation  $\phi$  is known. Examples of  $\phi$  are given in the Supplementary Material for the Swiss roll and Bitten torus. If  $\phi$  is unknown, Tosi et al. (2014) provide an approach to learn  $\phi$  by doing nonlinear dimension reduction using latent variable models.

The Riemannian manifold  $M$  is equipped with a metric tensor  $g$ . For example, if  $M$  is a submanifold of a Euclidean space, the induced metric tensor can be described in local coordinates as follows:

$$g_{ij}(\mathbf{x}) = \frac{\partial \phi}{\partial x_i}(\mathbf{x}) \cdot \frac{\partial \phi}{\partial x_j}(\mathbf{x}). \quad (6)$$

Based on its metric tensor, a Riemannian manifold has an associated Laplace Beltrami



**Fig. 2.** BM on  $M$  and its equivalent stochastic process in local coordinate system in  $\mathbb{R}^2$ .  $\phi : \mathbb{R}^2 \rightarrow M$  is a local parametrisation of  $M$ .

operator  $\Delta_s$ . In local coordinates,  $\Delta_s$  can be written as:

$$\Delta_s f = \frac{1}{\sqrt{G}} \frac{\partial}{\partial x^j} \left( \sqrt{G} g^{ij} \frac{\partial f}{\partial x^i} \right), \quad (7)$$



where  $G$  is the determinant of the matrix  $g$  and  $g^{ij}$  is the  $(i, j)$  element of its inverse. As we have seen in section 2.1, the heat kernel is a solution of the heat equation. The Laplace Beltrami operator is also the infinitesimal generator of Brownian Motion on the manifold. Let  $s_0$  be the starting point of the Brownian Motion  $S_t$  on manifold  $M$ , and introduce a function  $v : M \rightarrow \mathbb{R}$ . The expectation  $E_s(v(S_t)) = u(s_0, t)$  satisfies the heat equation,  $\partial u / \partial t = \frac{1}{2} \Delta_s u$ ,  $u(s_0, 0) = v(s_0)$ .

As in Figure 2, simulating a sample path of BM on  $M$  with starting point  $s_0$  is equivalent to simulating a stochastic process in  $\mathbb{R}^d$  with starting point  $\mathbf{x}(t_0)$ . The BM on a Riemannian manifold in a local coordinate system is given as a system of stochastic differential equations in the Ito form (Hsu, 1988, 2008):

$$dx_i(t) = \frac{1}{2} G^{-1/2} \sum_{j=1}^d \frac{\partial}{\partial x_j} \left( g^{ij} G^{1/2} \right) dt + \left( g^{-1/2} dB(t) \right)_i, \quad (8)$$

where  $g$  is the metric tensor of  $M$ ,  $G$  is the determinant of  $g$  and  $B(t)$  represents an independent BM in the Euclidean space. If  $M = \mathbb{R}^d$ ,  $g$  become an identity matrix and  $x_i(t)$  is the standard BM in  $\mathbb{R}^d$ . The first term of equation (8) is related to the local curvature of  $M$ . The second term relates to the position specific alignment of the BM by transforming the standard BM  $B(t)$  in  $\mathbb{R}^d$  based on the metric tensor  $g$ .

For simulating BM sample paths, the discrete form of equation (8) is first derived in equation (9). Specifically, the Euler-Maruyama method is used (Kloeden and Platen, 1992; Lamberton and Lapeyre, 2007) which yields:

$$\begin{aligned} x_i(t) &= x_i(t-1) + \frac{1}{2} \sum_{j=1}^d \left( -g^{-1} \frac{\partial g}{\partial x_j} g^{-1} \right)_{ij} \Delta t + \frac{1}{4} \sum_{j=1}^d (g^{-1})_{ij} \text{tr} \left( g^{-1} \frac{\partial g}{\partial x_j} \right) \Delta t + \left( g^{-1/2} dB(t) \right)_i \\ &= \mu(x_i(t-1), \Delta t)_i + \left( \sqrt{\Delta t} g^{-1/2} z^d \right)_i, \end{aligned} \quad (9)$$

where  $\Delta t$  is the diffusion time of each step of the BM simulation and  $z^d$  represents a  $d$ -dimensional standard normal random variable. The discrete form of the above stochastic differential equation defines the proposal mechanism of the BM with density

$$q(x(t)|x(t-1)) = \mathbb{N}(x(t)|\mu(x(t-1), \Delta t), \Delta t g^{-1}). \quad (10)$$

This proposal makes BM move according to the metric tensor. If the manifold  $M$  has boundary  $\partial M$ , we apply the Neumann boundary condition as in section 2.1 equation (1). It implies the simulated sample paths only exist within the boundary.

The simulation of BM from the discretisation of SDEs at points of singularity in the coordinate system (e.g. north pole of sphere) could be difficult. The drift term (the  $dt$  term in equation 8) may become too large for the simulated step to be a good approximation of the actual Brownian motion. A possible way to address this issue is to limit the size of drift in each simulation step by reducing the time step adaptively.

### 2.3. Numerical approximation of the heat kernel: exploiting connections with the transition density of Brownian motion

To explain explicitly the equivalence between the heat kernel and the transition density of the BM, let  $S(t)$  denote a BM on  $M$  started from  $s_0$  at time  $t = 0$ . The probability

of  $S(t) \in A \subset M$ , for any Borel set  $A$ , is given by

$$\mathbb{P}[S(t) \in A | S(0) = s_0] = \int_A K_{heat}^t(s_0, s) ds, \quad (11)$$

where the integral is defined with respect to the volume form of  $M$ . In this context, the Neumann boundary condition on the heat kernel corresponds to BM reflecting at the boundary. This can be approximated by pausing time and resampling the next step until it stays within the boundary. The difference between reflecting and resampling is small when the proposed BM step is not far from the boundary. Further discussions are provided in the Supplementary Material.

We approximate the heat kernel via approximating the integral in equation (11) by simulating BM sample paths and numerically evaluating the transition probability. Considering the BM  $\{S(t) : t > 0\}$  on  $M$  with the starting point  $S(0) = s_0$ , we simulate  $N$  sample paths. For any  $t > 0$  and  $s \in M$ , the probability of  $S(t)$  in a small neighbourhood  $A$  of  $s$  can be estimated by counting how many BM sample paths reach  $A$  at time  $t$ . Note that the BM diffusion time  $t$  works as the smoothing parameter. If  $t$  is large, the BM has higher probability to reach the neighbourhood of the target point and leads to higher covariance and vice versa. The *transition probability* is approximated as

$$\mathbb{P}[S(t) \in A | S(0) = s_0] \approx \frac{k}{N}, \quad (12)$$

where  $N$  is the number of simulated BM sample paths and  $k$  is the number of BM sample paths which reach  $A$  at time  $t$ . An illustrative diagram is shown in Figure 3. The *transition density* of  $S(t)$  at  $s$  is approximated as

$$K_{heat}^t(s_0, s) \approx \hat{K}^t = \mathbb{P}[S(t) \in A | S(0) = s_0] \approx \frac{1}{V(A)} \cdot \frac{k}{N}, \quad (13)$$

where  $V(A)$  is the Riemannian volume of  $A$ , which is parameterised with the radius of  $A$ , and  $\hat{K}^t$  is the estimated transition density. The error (numerical and Monte Carlo) of this estimator of the heat kernel is discussed in Section 2.5.

**REMARK 2.2.** *We are not aware of any rigorous definition of Reflective BM for a general Riemannian manifold with boundary. We conjecture that, given a suitable definition of reflective Brownian motion in a general Riemannian manifold with boundary (that generalizes the existing definition for a Euclidean domain), the RBM exists and is unique if the boundary is  $C^2$  (a.e.), and its transition density functions are the Neumann heat kernels.*

The in-GP can be constructed using the approximation in (13). The covariance matrix of the training data  $\Sigma_{\mathbf{ff}}$  can be explicitly obtained as follows: for the  $i$ th row of  $\Sigma_{\mathbf{ff}}$ ,  $N$  BM sample paths are simulated, with the starting point the  $i$ th data point indexed by the corresponding row. For each element of the  $i$ th row,  $\Sigma_{\mathbf{ff}}$ ,  $\hat{K}^t(s_i, s_j)$  is then estimated using (13). Algorithm 1 below provides details on how to generate  $\Sigma_{\mathbf{ff}}$ .

Optimisation of the kernel hyper parameters is discussed in section 2.6. Given in-GPs as the prior, one can then update with the likelihood to obtain the posterior distribution

**Algorithm 1** Simulating Brownian motion sample paths for estimating  $\Sigma$ 


---

```

1.1 Generate Brownian motion sample paths
for  $i = 1, \dots, N_d$  do                                ▷  $N_d$  is the size of data points
    for  $j = 1, \dots, N_{bm}$  do                            ▷  $N_{bm}$  is No. of sample paths
        for  $l = 1, \dots, T$  do                        ▷ T steps Brownian motion,  $T * \Delta t \rightarrow$  max diffusion time
            do                                          ▷ keep proposing x until value is within boundary
                 $q(x_{i,j}(l)|x_{i,j}(l-1)) \leftarrow \mathbb{N}(x_{i,j}(l)|\mu(x_{i,j}(l-1), \Delta t), \Delta t g^{-1})$     ▷ use eqn 10
            While  $x_{i,j}$  is located outside of M
    return  $\mathbf{x}$ 
1.2 Given a discrete choice of the diffusion time  $t \in \{\Delta t, 2 * \Delta t, \dots, T * \Delta t\}$ , the
covariance matrix  $\Sigma^t$  is estimated based on the BM simulation from 1.1.
for  $i = 1, \dots, N_d$  do
    for  $j = 1, \dots, N_d$  do
         $k = \text{which}(x(t) \in A_j)$                                 ▷ counting how many BM paths reach  $A_j$ 
         $K_{heat}^t(s_i, s_j) = \frac{k}{N_{BM} * V(A_j)}$                     ▷ use eqn 13
         $\Sigma_{ij}^t = \sigma_h^2 K_{heat}^t(s_i, s_j)$ 
    return  $\Sigma^t$ 

```

---

for inference. Let  $\mathbf{f}_*$  be a vector of values of  $f(\cdot)$  at some test points not represented in the training sample. The joint distribution of  $\mathbf{f}$  and  $\mathbf{f}_*$  is:

$$p(\mathbf{f}, \mathbf{f}_*) = \mathcal{N}\left(0, \begin{bmatrix} \Sigma_{\mathbf{f}\mathbf{f}} & \Sigma_{\mathbf{f}\mathbf{f}_*} \\ \Sigma_{\mathbf{f}_*\mathbf{f}} & \Sigma_{\mathbf{f}_*\mathbf{f}_*} \end{bmatrix}\right), \quad (14)$$

where  $\Sigma_{\mathbf{f}_*\mathbf{f}}$  is the covariance matrix for training and test data points. Each entry of the covariance matrix of the joint distribution can be calculated using equation (15):

$$\Sigma_{ij} = \sigma_h^2 \hat{K}^t(s_i, s_j). \quad (15)$$

For the same row of  $\Sigma_{\mathbf{f}\mathbf{f}}$  and  $\Sigma_{\mathbf{f}\mathbf{f}_*}$ , all elements can be estimated from the same patch of BM simulations which share the same starting points. No additional BM simulations are needed to estimate  $\Sigma_{\mathbf{f}_*\mathbf{f}_*}$ . The predictive distribution is derived by marginalising out  $\mathbf{f}$ :

$$p(\mathbf{f}_*|\mathbf{y}) = \int p(\mathbf{f}_*\mathbf{f}|\mathbf{y})d\mathbf{f} = \mathcal{N}\left(\Sigma_{\mathbf{f}_*\mathbf{f}}(\Sigma_{\mathbf{f}\mathbf{f}} + \sigma_{noise}^2 I)^{-1}\mathbf{y}, \Sigma_{\mathbf{f}_*\mathbf{f}_*} - (\Sigma_{\mathbf{f}\mathbf{f}} + \sigma_{noise}^2 I)^{-1}\Sigma_{\mathbf{f}\mathbf{f}_*}\right).$$

If one is only interested in the predictive mean, only  $\Sigma_{\mathbf{f}_*\mathbf{f}}$  and  $\Sigma_{\mathbf{f}\mathbf{f}}$  need to be estimated. The predictive variance of test points requires computing the covariance matrix  $\Sigma_{\mathbf{f}_*\mathbf{f}_*}^t$ . This requires extra BM simulations whose starting points are the test points. This could be computationally heavy if the number of test points is big. The sparse in-GP is introduced in the next section to handle this problem.

#### 2.4. Sparse in-GP on manifolds to reduce computation cost

The construction of in-GPs proposed in section 2.3 requires simulating BM sample paths at each data point. Although the BM simulations are embarrassingly parallelizable, the

computational cost can be high when the sample size is large. In addition, Gaussian processes face the well-known problem of high computational complexity  $O(n^3)$  due to the inversion of the covariance matrix. In this section, we propose to combine in-GPs with sparse Gaussian process approximations proposed by Quionero-Candela et al. (2007). We call the resulting construction *sparse in-GP*. By employing sparse in-GP, Brownian motion paths only need to be simulated starting at the induced points instead of every data point. The intuition behind this is that a lot of training data are located close together, implying that there may be a lot of redundant information. The inducing point approximation summarizes the training data into a small set of inducing points, so that inference could be done more efficiently.

The GP prior can be augmented with an additional set of  $m$  inducing points on  $M$  denoted as  $\mathbf{z} = [z_1, \dots, z_m]$ ,  $z_i \in M$  and we have  $m$  random variables  $\mathbf{u} = [f(z_1), \dots, f(z_m)]$ . The marginal prior distribution  $p(\mathbf{f}_*, \mathbf{f})$  remains unchanged after the model is rewritten in terms of the prior distribution  $p(\mathbf{u})$  and the conditional distribution  $p(\mathbf{f}_*, \mathbf{f}|\mathbf{u})$ :

$$p(\mathbf{f}_*, \mathbf{f}) = \int p(\mathbf{f}_*, \mathbf{f}, \mathbf{u}) d\mathbf{u} = \int p(\mathbf{f}_*, \mathbf{f}|\mathbf{u}) p(\mathbf{u}) d\mathbf{u}, \quad p(\mathbf{u}) = \mathcal{N}(0, \Sigma_{\mathbf{u}\mathbf{u}}), \quad (16)$$

where the distribution of  $\mathbf{u}$  is a multivariate Gaussian with mean zero and covariance matrix  $\Sigma_{\mathbf{u}\mathbf{u}}$ . The above augmentation does not reduce the computational complexity. For efficient inference, we adopt the Deterministic Inducing Conditional (DIC) approximation by Quionero-Candela et al. (2007), where  $\mathbf{f}_*$  and  $\mathbf{f}$  are assumed to be conditionally independent given  $\mathbf{u}$  and the relations between any  $\mathbf{f}$  and  $\mathbf{u}$  are deterministic:

$$p(\mathbf{f}_*, \mathbf{f}) \approx q(\mathbf{f}_*, \mathbf{f}) = \int q(\mathbf{f}_*|\mathbf{u}) q(\mathbf{f}|\mathbf{u}) p(\mathbf{u}) d\mathbf{u}, \quad (17)$$

$$q(\mathbf{f}|\mathbf{u}) = \mathbf{N}(\boldsymbol{\mu}_{\mathbf{f}}, 0), \quad \boldsymbol{\mu}_{\mathbf{f}} = \Sigma_{\mathbf{f}\mathbf{u}} \Sigma_{\mathbf{u}\mathbf{u}}^{-1} \mathbf{u}, \quad (18)$$

$$q(\mathbf{f}_*|\mathbf{u}) = \mathbf{N}(\boldsymbol{\mu}_{\mathbf{f}_*}, 0), \quad \boldsymbol{\mu}_{\mathbf{f}_*} = \Sigma_{\mathbf{f}_*\mathbf{u}} \Sigma_{\mathbf{u}\mathbf{u}}^{-1} \mathbf{u}. \quad (19)$$

The resulting sparse in-GP prior

$$q(\mathbf{f}, \mathbf{f}_*) = \mathbf{N} \left( 0, \begin{bmatrix} Q_{\mathbf{f}\mathbf{f}} & Q_{\mathbf{f}\mathbf{f}_*} \\ Q_{\mathbf{f}_*\mathbf{f}} & Q_{\mathbf{f}_*\mathbf{f}_*} \end{bmatrix} \right) = \mathbf{N} \left( 0, \begin{bmatrix} \Sigma_{\mathbf{f}\mathbf{u}} \Sigma_{\mathbf{u}\mathbf{u}}^{-1} \Sigma_{\mathbf{u}\mathbf{f}} & \Sigma_{\mathbf{f}\mathbf{u}} \Sigma_{\mathbf{u}\mathbf{u}}^{-1} \Sigma_{\mathbf{u}\mathbf{f}_*} \\ \Sigma_{\mathbf{f}_*\mathbf{u}} \Sigma_{\mathbf{u}\mathbf{u}}^{-1} \Sigma_{\mathbf{u}\mathbf{f}} & \Sigma_{\mathbf{f}_*\mathbf{u}} \Sigma_{\mathbf{u}\mathbf{u}}^{-1} \Sigma_{\mathbf{u}\mathbf{f}_*} \end{bmatrix} \right),$$

where  $Q$  is defined as  $Q_{\mathbf{a},\mathbf{b}} = \Sigma_{\mathbf{a},\mathbf{u}} \Sigma_{\mathbf{u},\mathbf{u}}^{-1} \Sigma_{\mathbf{u},\mathbf{b}}$ . Using algorithm 1,  $\Sigma_{\mathbf{u}\mathbf{u}}$ ,  $\Sigma_{\mathbf{u}\mathbf{f}}$  and  $\Sigma_{\mathbf{u}\mathbf{f}_*}$  are all obtained by estimating the transition density of BM simulation paths with inducing points as the starting points.

We then only need to simulate the BM sample paths starting from the inducing points. The total number of BM simulations is reduced from  $n \times N_{bm}$  to  $m \times N_{bm}$ , where  $m$  is the number of inducing points,  $n$  is the number of data points and  $N_{bm}$  is the number of Brownian motion sample paths given a single starting point. The complexity of inverting the covariance matrix is decreased from  $O(n^3)$  to  $O(n \times m^2)$ .

With the above approximation, the marginal distribution of the corresponding GP with a Gaussian likelihood is written as:

$$p(\mathbf{y}|\mathbf{f}) \approx q(\mathbf{y}|\mathbf{u}) = \prod_{i=1}^n \mathbf{N}(y_i | \Sigma_{f_i \mathbf{u}} \Sigma_{\mathbf{u}\mathbf{u}}^{-1} \mathbf{u}, \sigma_{noise}^2 \mathbf{I}). \quad (20)$$

The inducing points in the above marginal likelihood can be further marginalised out by substituting the definition of its prior distribution (16):

$$p(\mathbf{y}|\mathbf{s}_{induce}) = \int q(\mathbf{y}|\mathbf{u})p(\mathbf{u}|\mathbf{s}_{induce})d\mathbf{u} = \mathbf{N}(0, \Sigma_{\mathbf{f}\mathbf{u}}\Sigma_{\mathbf{u}\mathbf{u}}^{-1}\Sigma_{\mathbf{u}\mathbf{f}} + \sigma_{noise}^2\mathbf{I}). \quad (21)$$

With the above model, we can also obtain the predictive distribution as

$$q(\mathbf{f}_*|\mathbf{y}) = \mathbf{N}\left(Q_{\mathbf{f}_*\mathbf{f}}(Q_{\mathbf{f}\mathbf{f}} + \sigma^2\mathbf{I})^{-1}\mathbf{y}, Q_{\mathbf{f}_*\mathbf{f}_*} - Q_{\mathbf{f}_*\mathbf{f}}(Q_{\mathbf{f}\mathbf{f}} + \sigma^2\mathbf{I})^{-1}Q_{\mathbf{f}\mathbf{f}_*}\right). \quad (22)$$

There is a huge literature on reducing the matrix inversion bottleneck in GP computation (Schwaighofer and Tresp, 2002; Quiñonero-Candela and Rasmussen, 2005; Snelson and Ghahramani, 2006; Titsias, 2009). Recent approaches, such as Katzfuss and Guinness (2017), can achieve linear time computation complexity under certain conditions. However, such approaches require an analytical form of covariance kernel; to apply these methods, we would need to simulate BM paths at the training and prediction points. For this reason, we use DIC due to its avoidance of the need to estimate the diagonal elements of the covariance matrix.

### 2.5. Monte Carlo and numerical error for approximation of heat kernel

In this subsection, we discuss the error of our heat kernel estimator as defined in equation (13). We also consider numerical experiments in the special case of  $\mathbb{R}$  in which case the true heat kernel is known.

Consider a Brownian motion  $\{S(t) : t > 0\}$  on a Riemannian manifold  $M$  with  $S(0) = s_0$ . Fix some  $t > 0$  and  $s \in M$ . The probability density of  $S(t)$  at  $s$  is  $K_{heat}^t(s_0, s)$ . The true BM transition probability evaluated at a set  $A$  is given by  $p(A) = \mathbb{P}[S(t) \in A | S(0) = s_0] = \int_A K_{heat}^t(s_0, s)ds$ . The error of our estimator  $\hat{K}^t$  consists of two parts.

**Part I: Numerical error.** Choose local coordinates  $(r_1, \dots, r_d)$  near  $s$  with  $r_1(s) = \dots = r_d(s) = 0$  (for convenience of illustration) and a window size  $w$ . The heat kernel  $K_{heat}^t$  can then be approximated by

$$K^{t'} = \frac{1}{V(A)} \mathbb{P}[|r_i(S(t))| < w \text{ for } i = 1, \dots, d],$$

where  $V(A)$  denotes the volume of the region defined by  $\{|r_i| < w, i = 1, \dots, d\}$ . By Taylor expansion around  $s$ , we have

$$K^{t'} = K_{heat}^t + O(w^2). \quad (23)$$

Therefore, the approximation error increases (quadratically) with  $w$ , i.e., the order of magnitude of  $K^{t'} - K_{heat}^t$  is  $O(w^2)$ .

If  $M = \mathbb{R}^d$ , one can explicitly derive the error. Assuming the starting point of BM  $s_0$  is the origin for simplicity, the heat kernel  $K_{heat}^t$  on  $\mathbb{R}^d$  can be approximated as

$$\begin{aligned} K^{t'} &= \frac{1}{V(A)} \mathbb{P}[\|S(t) - s\| < w] = \frac{1}{V(A)} \int_A K_{heat}^t(s_0, s)ds, \\ &= \frac{1}{(2w)^d} \int_{s_1-w}^{s_1+w} \dots \int_{s_d-w}^{s_d+w} \exp\left(-\frac{\sum_{i=1}^d x_i^2}{2t}\right) dx_d \dots dx_1. \end{aligned} \quad (24)$$

Taylor expansion of equation (24) yields

$$K^{t'} - K_{heat}^t = \frac{\sum_{i=1}^d s_i^2 - d \cdot t}{6t} \cdot \frac{w^2}{t} + O\left(\frac{w^4}{t^2}\right). \quad (25)$$

Assuming  $w$  is small compared to  $\sqrt{t}$ , the order of magnitude of this error is  $O(w^2)$ .

REMARK 2.3. *For convenience in computing the integral in equation (24), a hyper-cube is used instead of the Euclidean ball. The order of magnitude of the error remains the same.*

**Part II: Monte Carlo error.** Given  $N_{BM}$  number of BM sample paths,  $K^{t'}$  is approximated by  $\hat{K}^t$ :

$$\hat{K}^t = \frac{1}{V(A)} \cdot \frac{k}{N_{BM}}, \quad \text{where } k \sim \text{Bin}(N_{BM}, V(A)K^{t'}). \quad (26)$$

Recall that  $k$  is the number of sample paths within  $\|S(t) - s\| < w$  and has binomial distribution with  $N_{BM}$  trials and probability of success  $V(A)K^{t'}$ . Here  $\frac{k}{N_{BM}}$  is the estimate of the transition probability of BM.

The expectation and the standard error of  $\hat{K}^t$  are:

$$E(\hat{K}^t(s_0, s)) = K^{t'} = K_{heat}^t(s_0, s) + O(w^2), \quad (27)$$

$$\begin{aligned} sd(\hat{K}^t) &= \frac{1}{N_{BM}V(A)} \cdot \sqrt{N_{BM} \cdot V(A)K^{t'}(1 - V(A)K^{t'})} \\ &\leq \sqrt{\frac{K^{t'}}{N_{BM}V(A)}} = O(N_{BM}^{-1/2}w^{-d/2}), \end{aligned} \quad (28)$$

The standard deviation decreases with  $w$  and  $N_{BM}$ . As  $w^2 \rightarrow 0$ ,  $E(\hat{K}^t(s_0, s)) = K_{heat}^t(s_0, s)$ , and also as  $\frac{w^{-d}}{N_{BM}} \rightarrow 0$ ,  $Var(\hat{K}^t(s_0, s)) = 0$ . The estimator  $\hat{K}^t(s_0, s)$  is asymptotically unbiased and consistent.

The optimal order of magnitude of  $w_{opt}$  can be calculated by minimising the sum of the numerical error and Monte Carlo error as described above. Specifically, for an arbitrary  $M$ , given a fixed number of BM simulations  $N_{BM}$ , one has

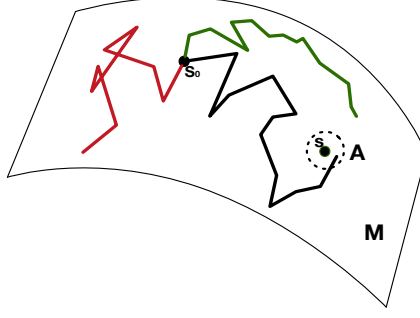
$$\mathcal{L}(w) = O(w^2) + O(w^{-d/2}). \quad (29)$$

In particular if  $M$  is  $R^d$ , an explicit expression of the error is available:

$$\mathcal{L}(w) = \sqrt{\frac{\hat{K}^t}{N \cdot (2w)^d}} + K_{heat}^t \frac{\sum_{i=1}^d s_i^2 - d \cdot t}{6t} \cdot \frac{w^2}{t}. \quad (30)$$

Given a pre-specified error level, the order of the minimum number of BM simulations  $N$  required can be derived. Refer to the Supplementary Material for the example of estimating the heat kernel in a one-dimensional Euclidean space.

Numerical accuracy of estimates for the special case of  $\mathbb{R}$  are shown in Table 1 and Figure 4. The true heat kernel  $K_{heat}^t(0, s)$  is calculated using equation 1 in the Supplementary Material at seventy equally spaced  $s \in (-9, 9)$ . The diffusion time is fixed as



**Fig. 3.** BM on a manifold  $M$ .  $s_0$  is the starting point of BM sample paths. The solid lines represent three independent BM sample paths from time 0 to  $t$ . The dashed circle represents a set  $A$ , which is a neighbourhood of a point  $s$  on  $M$ . In this example, only the black sample path reaches  $A$  at time  $t$  and the estimate of the transition probability  $p(S(t) \in A | S(0) = s_0)$  is  $\frac{1}{3}$ .

**Table 1.** Comparison of estimates of BM transition density and the heat kernel in  $\mathbb{R}$ . The table shows the median absolute error and median relative error between the true heat kernel  $K_{heat}^t$  and the numerical estimate of BM transition density. Values in brackets show the median absolute deviation.

No. Sample paths	$N_{BM}$	median absolute error	median relative error
3e+2		8.4e-3(8.9e-3)	24.6%(25.6e-1)
3e+3		2.8e-3 (2.9e-3)	6.4% (5.5e-2)
3e+4		7.2e-4 (6.8e-4)	1.6% (1.9e-2)
3e+5		4.7e-4 (3.8e-4)	1.3% (1.1e-2)

10. The transition probability of BM from the origin to the grid point  $s$  is estimated by counting how many BM paths reach the neighbourhood of  $s$  ( $[s - w, s + w]$ ) at time  $t$ . The transition density of BM at each grid point is then evaluated using equation (13). Using equation 7 in the Supplementary Material the order of magnitude of  $w_{opt}$  is derived as  $10^{-1}$ . We fix the radius  $w$  as 0.5 in equation (13).

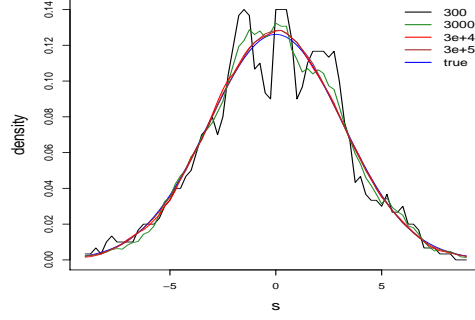
The number of BM simulation sample paths  $N_{BM}$  are selected from three hundred to three hundred thousand with increasing order of magnitude. The median of relative error decreases as  $N_{BM}$  increases and stabilises after thirty thousand. A similar pattern is observed for the median absolute error. Derivations for the transition density estimate of heat kernel in  $\mathbb{R}^2$  are shown in the Supplementary Material.

## 2.6. Optimising the kernel hyper parameters and comparison with an RBF kernel in $\mathbb{R}$

Given a diffusion time  $t$ , using algorithm 1 we can generate a covariance matrix  $\Sigma_{\mathbf{ff}}^t$  for the training data indexed by  $t$ . The log marginal likelihood function (over  $f$ ) is given by (Rasmussen, 2004):

$$p(\mathbf{y}|s) = \int p(\mathbf{y}|\mathbf{f})p(\mathbf{f}|s)d\mathbf{f} = -\frac{1}{2}\mathbf{y}^T(\Sigma_{\mathbf{ff}}^t + \sigma_{noise}^2I)^{-1}\mathbf{y} - \frac{1}{2}\log|\Sigma_{\mathbf{ff}}^t + \sigma_{noise}^2I| - \frac{N_d}{2}\log 2\pi. \quad (31)$$

The hyperparameters can be obtained by maximising the log of the marginal likelihood. The maximum of the BM diffusion time is set as  $T * \Delta t$ , where  $T$  is a positive integer,



**Fig. 4.** Comparison of estimates of BM transition density and the heat kernel in  $\mathbb{R}$ . The blue line represents the true heat kernel. Colored lines represent estimates of the BM transition density given different number of BM simulations ranging from 300 to  $3 \times 10^5$ .

**Table 2.** Comparison of estimates of kernel hyper parameters from Euclidean GP and in-GP in  $\mathbb{R}$ .

Case	Median estimates of $l$	Median estimates of $\sigma_r$
Truth	1	1
Euclidean GP	1.13(0.16)	0.94(0.36)
in-GP	1.15(0.2)	0.94(0.38)
p value	0.91	0.85

and  $\Delta t$  is the BM simulation time step as defined in (10).  $T$  covariance matrices  $\Sigma_{\mathbf{ff}}^{1..T}$  can be generated based on the BM simulations. Optimisation of diffusion time  $t$  can be done by selecting the corresponding  $\Sigma_{\mathbf{ff}}^t$  that maximises the log marginal likelihood. Estimation of  $\sigma_h$  given the smoothing parameter  $t$  follows using standard optimisation routines, such as quasi-Newton. For the sparse in-GP, the likelihood function is replaced by (21) and the hyper parameters can be obtained by similar procedures.

We compare the estimates of kernel hyperparameters from an Euclidean GP (the standard GP in the Euclidean space) and the in-GP in  $\mathbb{R}$  by applying both methods to ten sets of test data. Data sets are generated by sampling 20 data points from a multivariate normal distribution with mean zero and covariance  $\Sigma_{test}$ .  $\Sigma_{test}$  is produced by a standard RBF kernel with  $l = 1$  and  $\sigma_r = 1$ . In this case, the ground truth of the hyperparameters of the heat kernel is known.

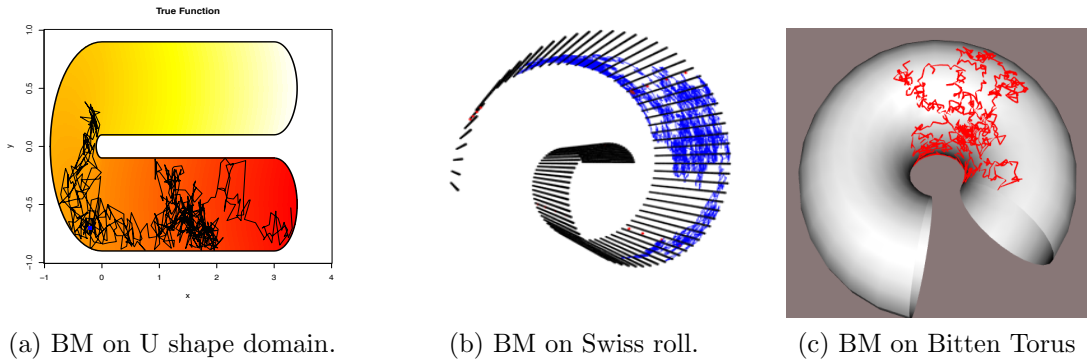
We simulate  $N_{bm} = 40,000$  BM sample paths for each testing data point. The estimates of hyper parameters  $t$  and  $\sigma_h$  are obtained by maximising (31). For the case of  $\mathbb{R}$ , the two methods should produce very similar results, since the heat kernel is equivalent to an RBF kernel in  $\mathbb{R}$ .

The result is shown in Table 2 which records the true value and the median estimates of kernel hyper parameter  $l$  and  $\sigma$ . Values in brackets show the median absolute deviation. The error bounds provide insights about the level of error that is introduced by random walk. The  $p$ -values of Wilcoxon tests indicate the difference in medians between the two methods are not significant.



### 3. Simulation studies

In this section, we carry out simulation studies for a regression model with true regression functions defined on a U shape domain, a 2-dimensional Swiss Roll embedded in  $\mathbb{R}^3$  and the Bitten Torus. The performance of in-GP is compared to that of an Euclidean GP (the standard GP as in Rasmussen (2004)) and the soap film smoother in Wood et al. (2008) for the U shape example. For the Swiss Roll and Bitten torus examples, the results from in-GP are compared with those from an Euclidean GP model. Examples of BM sample paths on the U shape domain, Swiss Roll and Bitten Torus are shown in Figure 5.

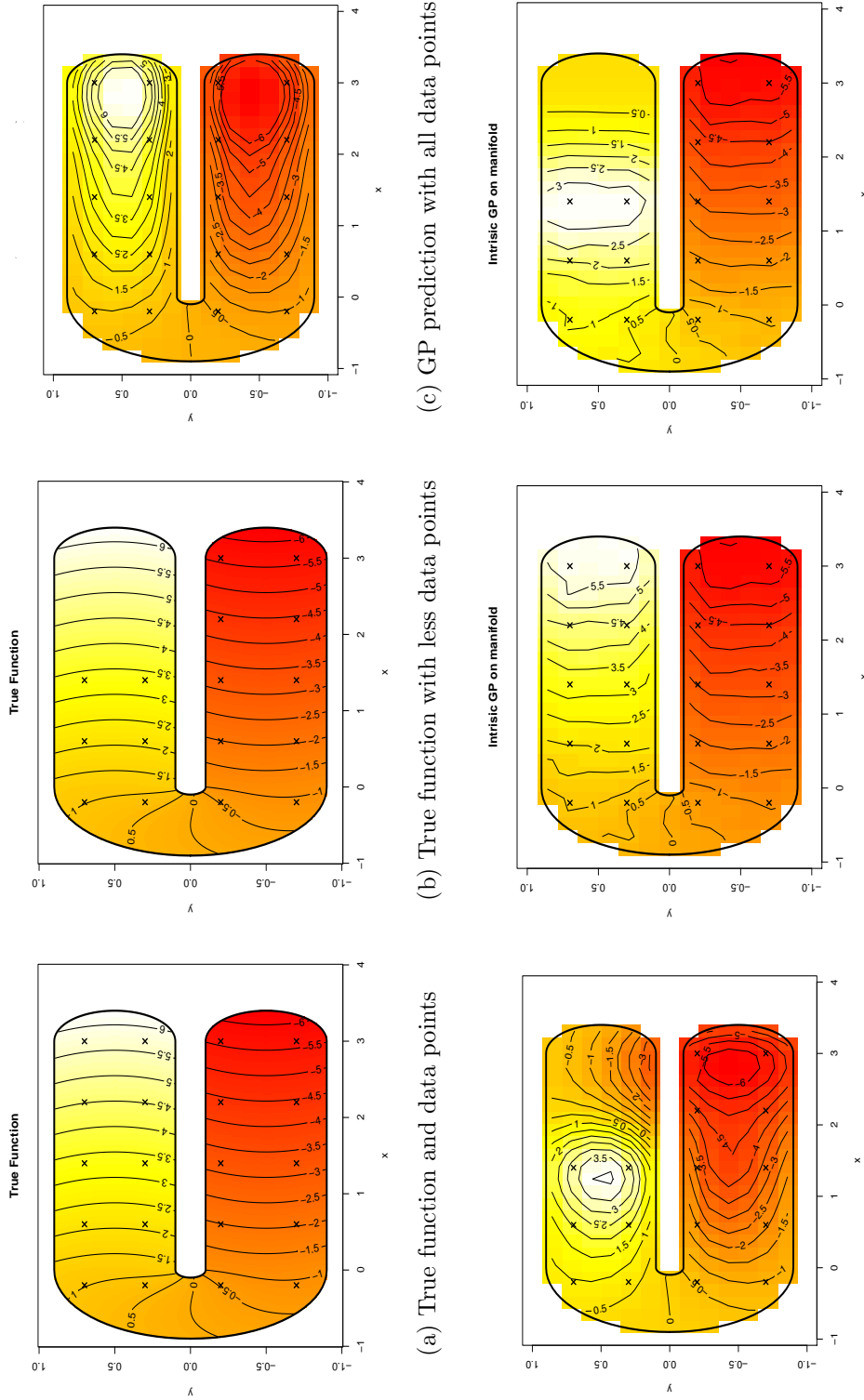


**Fig. 5.** Examples of BM sample paths are plotted as a black line on the U shape domain in (a), a blue line on the Swiss Roll in (b) and a red line on the Bitten Torus in (c).

#### 3.1. U shape example

A U shaped domain (see e.g., Wood (2001)), defined as a subset of  $\mathbb{R}^2$ , is plotted in Figure 6(a). The value of a test or regression function (i.e. the colour of the map) varies smoothly from the lower right corner to the upper right corner of the domain ranging from -6 to 6. The black crosses represent 20 observations which were equally spaced in both  $x$  and  $y$  directions within the domain of interest. The goal is to estimate the test function and make predictions at 450 equally spaced grid points within the domain.

Since the U shaped domain is defined as a subset of  $\mathbb{R}^2$ , the mapping function  $\phi$  in equation (2) is a constant. Therefore, BM reduces to the standard BM in the two-dimensional Euclidean space restricted within the boundary. When a proposed BM step hits the boundary, the proposed move is rejected. New proposal steps will be made until the proposed sample path locates within the boundary. The trajectory of a sample path (black line) of the BM is shown in Figure 5(a) with the blue dot serving as the starting location.



(c) GP prediction with all data points

(b) True function with less data points

(a) True function and data points

(f) in-GP prediction with less data points

(e) in-GP prediction with all data points

(d) GP prediction with less data points

**Fig. 6.** Comparison of in-GP and Euclidean GP in the U-shape example. The true function and data points are shown in (a). The Euclidean GP prediction and the in-GP prediction are plotted in (c) and (d). The contours in (c) are more squashed than (e). Four data points are removed from the upper right arm in (b). The Euclidean GP and in-GP prediction are shown in (d) and (f). The differences are exacerbated. It is clear that the Euclidean GP smooths across the boundary in (d).

**Table 3.** Comparison of the RMSE (root mean squared error) of predictive means for different methods on the U-shape domain. The table shows the mean of RMSE over 50 datasets. Values in brackets show the standard deviation.

<i>Case</i>	<i>Euclidean GP</i>	<i>in-GP</i>	<i>soap film smoother</i>
30db	1(0.01)	0.274(0.04)	0.271(0.22)
10db	1.36(0.17)	0.754(0.14)	0.747(0.37)

The heat map of the predictive mean of in-GP at the grid points is shown in Figure 6(e). The coloured contours of the prediction are similar to that of the true function in Figure 6(a). The contours of the Euclidean GP predictive mean in Figure 6(c) are more squashed, and the differences are exacerbated when certain observations are removed as in Figure 6(b). It is clear that the Euclidean GP smooths across the gap between the two arms of the domain (see Figure 6(d)). This is due to that fact that the upper arm and lower arm are close in Euclidean distance. In contrast, the in-GP, which takes into account the intrinsic geometry, does not smooth across the gap as seen in Figure 6(f). Given a fixed diffusion time, the transition probability of BM from points in the lower arm to points in the upper arm within the boundary is relatively small. This leads to lower covariance between these two regions and more accurate predictions.

The U shaped domain example has also been used for evaluating the performance of the soap film smoothers in Wood et al. (2008), in comparison with some other methods such as thin plate splines and the method of Ramsay (2002). Comparisons made in Wood et al. (2008) show that the soap film smoother outperforms the other methods. In our study, the in-GP, Euclidean GP, and soap film smoother are compared for different levels of signal-to-noise ratio. The values of the true function are perturbed by Gaussian noise with a standard deviation of 0.1 and 1 (signal-to-noise ratios are 30db and 10db, respectively) with 50 replicates for each noise level. For each of the replicates, different methods are applied to estimate the test function at the grid points. The mean and standard deviation of the mean-squared error (MSE) for these 50 replicates are reported in Table 3. The soap film smoother is constructed using 10 inner knots and 10 cubic splines. in-GP and soap film are both significantly better than Euclidean GP. There is no substantial difference in terms of the mean MSE between the two methods. However the standard deviation of the MSE for in-GP is substantially smaller for both noise levels. This indicates the prediction of in-GP is more robust.

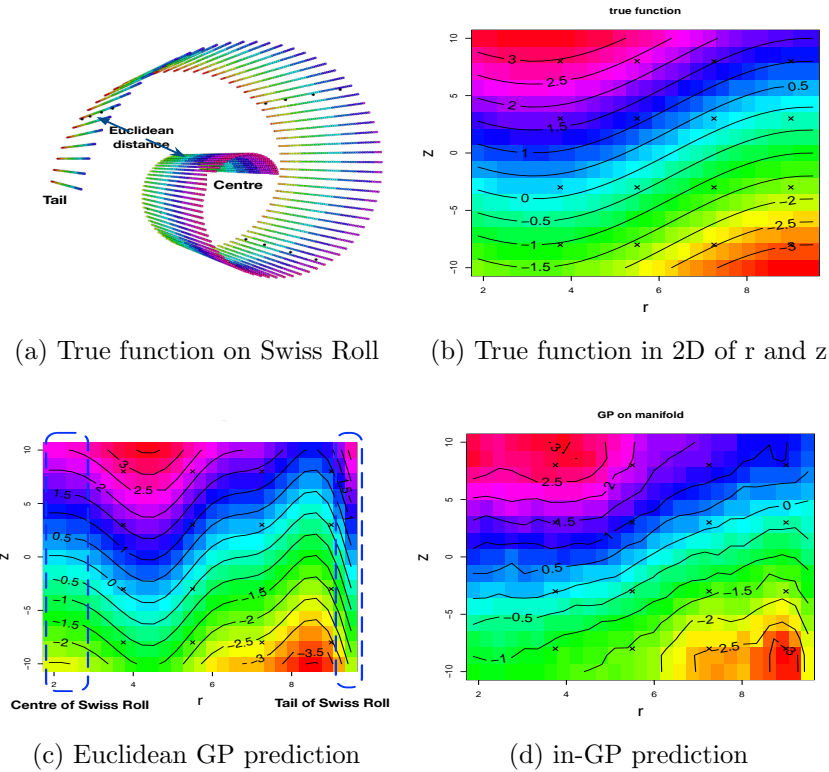
### 3.2. *Swiss Roll*

The in-GP model applies to general Riemannian manifolds and has much wider applicability to complex spaces beyond subsets of  $\mathbb{R}^2$ . Here we consider a synthetic dataset on a Swiss Roll which is two-dimensional manifold embedded in  $\mathbb{R}^3$ . The soap film method is only appropriate for smoothing over regions of  $\mathbb{R}^2$ , and hence cannot be applied here. A Swiss Roll is a spiralling band in a three-dimensional Euclidean space. A nonlinear function  $f$  is defined on the surface of Swiss Roll with

$$Y_i = f(x_i, y_i, z_i) + \epsilon_i,$$

where  $x_i, y_i, z_i$  are the coordinates of a point on the surface. The construction of the Swiss Roll and the derivation of the metric tensor are shown in the Supplementary Material.

The true function  $f$  is plotted in Figure 7(a). 20 equally spaced observations are marked with black crosses. For better visualisation, the true function is plotted in the unfolded Swiss Roll in the radius  $r$  and width  $z$  coordinates in Figure 7(b). The true function values are indicated by colour and with contours at the grid points.



**Fig. 7.** Comparison of the Euclidean GP and in-GP on Swiss Roll. The true function and data points (black dots) are plotted on the surface of Swiss Roll in (a). The true function is plotted in the two-dimensional unfolded Swiss Roll with coordinates  $r$  and  $z$  in (b). The GP predictions are plotted with contours in (c). The left end and right end of (c) marked by a blue dashed box are quite different from (b) in colour. The in-GP prediction is shown in (d). The prediction at the centre and the tail part of the Swiss Roll (right end and left end of (d)) has been improved.

We first applied an Euclidean GP to this example using an RBF kernel in  $\mathbb{R}^3$ . In order to visualise the differences between the prediction and the true function, the GP predictive mean is plotted in the unfolded Swiss Roll in Figure 7(c). The overall shape of contours is more wiggly comparing to the true function in Figure 7(b). In addition, the predictive mean is quite different from the truth in colour in certain regions. For example, the left end of Figure 7(c) marked by the blue dashed box corresponding to the centre of the Swiss Roll and the right end of Figure 7(c) corresponding to the tail

of the Swiss Roll. The prediction performance of the Euclidean GP in these regions is poor. This is because the Euclidean distance between the two regions is small as shown in Figure 7(a) while the geodesic distance between them (defined on the surface of Swiss Roll) is big.

In applying the in-GP to these data, the BM sample paths can be simulated using equation (9) using the metric tensor of the Swiss Roll. In particular, BM on the Swiss Roll can be modelled as the stochastic differential equation:

$$dr(t) = \frac{-2r}{(1+r^2)^2}dt + \frac{1}{2} \frac{2r}{(1+r^2)^2}dt + (1+r^2)^{-1/2}dB_r(t), \quad (32)$$

$$dz(t) = dB_z(t), \quad (33)$$

where  $B_r(t)$  and  $B_z(t)$  are two independent BMs in Euclidean space. A trace plot of a single BM sample path is shown in Figure 5(b). Following the procedure introduced in section 2.1, the predictive mean of in-GP is shown in Figure 7(d). The overall shape of the contour of the predictive mean is similar to that of the true function. The prediction at the centre and tail part of the Swiss Roll has been improved comparing to the results of the Euclidean GP. The root mean square error is calculated between the predictive mean and the true value at the grid points. It has been reduced from 0.53 (Euclidean GP) to 0.29 for the in-GP. The Neumann condition states that at any boundary point the heat kernel is stationary along the normal direction. This directly implies that the level curves of the in-GP prediction are orthogonal to the boundary. However, when the training data are far from the boundary, the level curve can be parallel to the boundary. For example, the right part of Figure 7 (d) corresponds to the tail part of the Swiss Roll. When  $r$  is big the distance on the surface of the Swiss Roll is bigger. The training data are far away from the boundary in the tail, so in-GP prediction tends to be close to the prior mean.

### 3.3. Bitten Torus

Here we consider another more substantial example: a Bitten Torus. The Torus is a two-dimensional manifold embedded in  $R^3$ . The three-dimensional coordinates can be parametrised by four variables:  $r$ —radius of tube,  $R$ —distance from centre of the tube to the centre of the torus, and  $(\theta, \phi)$  angles to parametrise the two full circles with  $\theta$  for angle of torus and  $\phi$  for angle of tube. In our case, we fix  $R$  and  $r$  and vary  $\theta$  and  $\phi$ . We removed the lower right part to construct the Bitten Torus. The Bitten Torus is not as ‘flat’ as the other examples considered above.

The value of the test function (i.e. the colour of the map, low value in dark blue and high value in dark red in Figure 8(a)) increases smoothly from 0.57 to 5.5 on the surface of the Bitten Torus. The true function and the noisy observations are plotted in Figure 8(a). 19 observations are marked with orange balls. 18 of the observations are evenly spaced and one additional observation is located near the centre of the Bitten Torus. Similar to the Swiss Roll example, the nonlinear function  $f$  is defined on the surface of the Bitten Torus with

$$Y_i = f(x_i, y_i, z_i) + \epsilon_i,$$

**Table 4.** Comparison of the RMSE of predictive means of two methods on the Bitten Torus. The table shows the mean of RMSE over 50 datasets. Values in bracket show the standard deviation.

Case	Euclidean GP	in-GP
30db	167.3(11.7)	25.8(11.6)
10db	229.2(93.4)	74.6(32.7)

where  $x_i, y_i, z_i$  are the coordinates of a point on the surface. More details on construction of the Torus and the derivation of the metric tensor are shown in the Supplementary Material. A demonstration of the BM on the Bitten Torus is shown in Figure 5(c).

We first applied an Euclidean GP to this example using an RBF kernel in  $\mathbb{R}^3$ . The GP predictive mean is plotted in Figure 8(b) with colour. Compared to the true function in Figure 8(a), the GP predictive mean is brighter (the colour is similar to yellow) in the centre area. The Euclidean distance between a data point and a grid point in the centre area is smaller than the geodesic distance on the Torus surface. The RBF kernel assigns bigger covariances between these points which makes the data point in the centre dominate the region.

By applying the in-GP to these data, the BM sample paths can be simulated using equation (9) with the metric tensor of the Bitten Torus. In particular, BM on the Bitten Torus can be modelled via the stochastic differential equations:

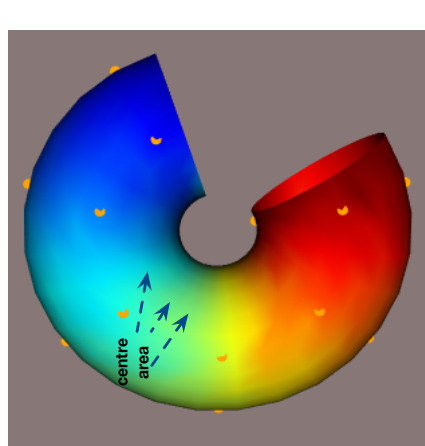
$$d\theta(t) = -\frac{1}{2}r^{-1} \sin \theta (R + r \cos \theta)^{-1} dt + r^{-1} dB_\theta(t) \quad (34)$$

$$d\phi(t) = |(R + r \cos \theta)^{-1}| dB_\phi(t). \quad (35)$$

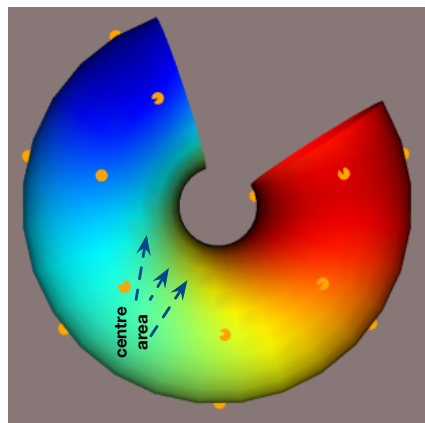
A trace plot of a single BM sample path is shown in Figure 5(c). Following the procedure in section 2.1, the predictive mean of in-GP is plotted in Figure 8(c). The in-GP prediction in the centre area looks more similar to the true function in Figure 8(a). Also the colour in the region near the lower bound is dark red, which is more similar to the true function comparing to the GP prediction in Figure 8(b).

For visualisation convenience, we have also plotted the function in two dimensions of  $\phi$  (angle of tube) and  $\theta$  (angle of torus) in Figure 8(d). The differences of the prediction in the centre area are clearer from the 2D contour plot of GP prediction in Figure 8(e) and in-GP prediction in Figure 8(f). In the 2D contour plots, the distance between  $\theta = 0$  and  $\theta = 2\pi$  is  $2\pi$  in  $R^2$ . However, given a fixed  $\phi$ ,  $\theta = 0$  and  $\theta = 2\pi$  represent the same point on the Torus and the distance between them is zero. Therefore, methods such as Sampson and Guttorp (1992) mapping the domain with a diffeomorphism to a regular region of  $R^k$  can lead to a big error for this case.

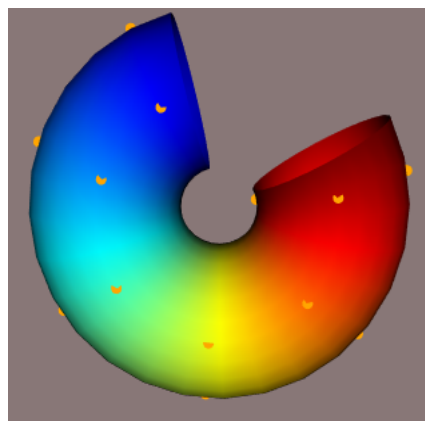
The performance of the Euclidean GP and in-GP are compared by varying the noise with different signal-to-noise ratios. The values of the true function are perturbed by Gaussian noise (30db and 10db) with 50 replicates for each noise level. For each of the replicates, different methods are applied to estimate the test function at some equally spaced grid points. The mean and standard deviation of the MSE for these 50 replicates are reported in Table 4. The prediction of in-GP is significantly better at all noise levels.



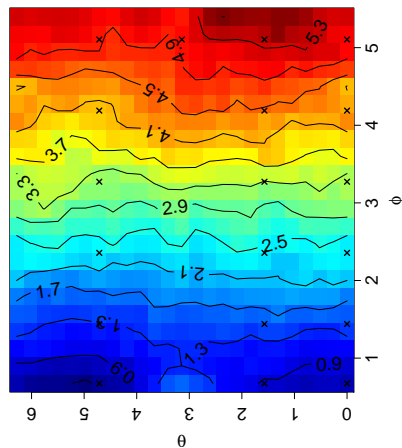
(c) In-GP prediction



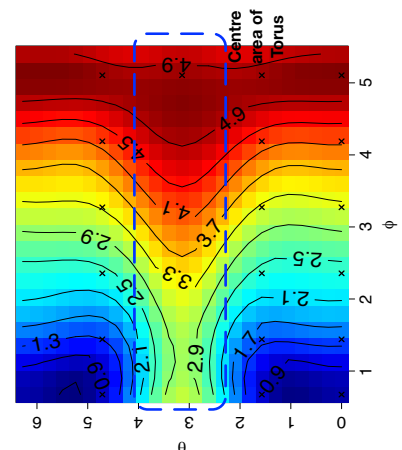
(b) GP prediction



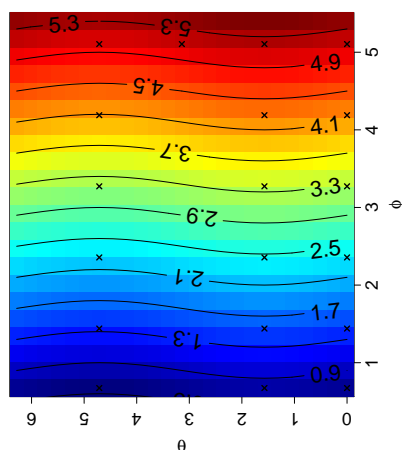
(a) True function with data points



(f) In-GP prediction in 2D of  $\phi$  and  $\theta$



(e) GP prediction in 2D



(d) True function in 2D of  $\phi$  and  $\theta$

**Fig. 8.** Comparison of the Euclidean GP and in-GP on Bitten Torus: The true function and data points (orange balls) are plotted on the surface of the Bitten Torus in (a). The function value increases smoothly from upper right to lower right of Torus. The GP and in-GP prediction are shown in (b) and (c). The in-GP prediction in the centre area looks similar to (a) comparing to the GP prediction in (b). The true function is plotted in the two-dimensional plane with coordinates  $\theta$  and  $\phi$ . The GP predictions are plotted with contours in (e). The middle part of (e) marked by the blue dashed box are quite different from (d). The in-GP prediction in shown in (f). The shape of the contour is similar to (d). The prediction at the middle part of (f), which corresponds to the centre area of (c), has been greatly improved.

We have also carried out experiments by removing the data point from the centre of the Torus. Details on more comparison results are also provided in the Supplementary Material.

#### 4. Application to chlorophyll data in Aral sea

In this section, we consider an analysis of remotely-sensed chlorophyll data at 485 locations in the Aral sea. The data are available from the `gamair` package (Wood, 2006) and are plotted in Figure 9(a). The level of chlorophyll concentration is represented by the intensity of the colour. The chlorophyll data from the satellite sensors are noisy and vary smoothly within the boundary but not across the gap corresponding to the isthmus of the peninsula. We applied different methods to estimate the spatial pattern of the chlorophyll density.

The  $\log$  of chlorophyll concentration is modelled as a function of the latitude and longitude coordinates of the measurement locations:

$$chl_i = f(lon_i, lat_i) + \epsilon_i,$$

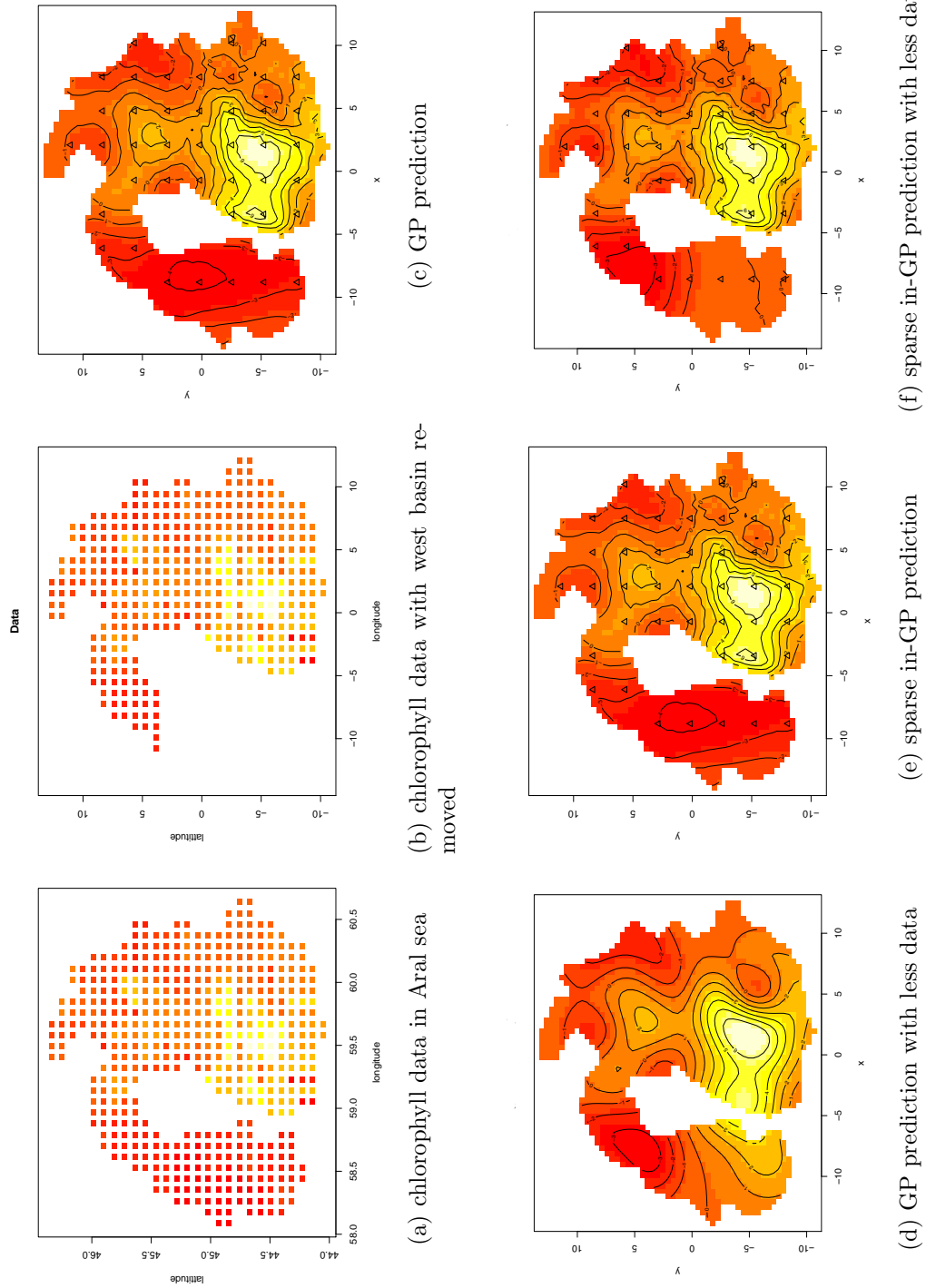
where  $lon_i$  and  $lat_i$  are standardised by subtracting the mean.

In order to reduce the computation cost, the sparse in-GP from section 2.4 is applied. 42 inducing points are introduced that are equally spaced within the boundary of the Aral sea (represented by small triangles in Figure 9(e)). The number of BM sample paths has been reduced from  $485 \times N_{BM}$  to  $42 \times N_{BM}$ , where  $N_{BM}$  is 20,000 in this example.

The predictive mean of the Euclidean GP is shown in Figure 9(c). As expected, the Euclidean GP smoothes across the isthmus of the central peninsula. Relatively high levels of chlorophyll concentration are estimated for the southern part of the eastern shore of the western basin of the sea, while all observations in this region have rather low concentrations. Similarly a decline in chlorophyll level towards the southern half of the western shore of the eastern basin is estimated, which is different from the pattern of the data in the region. On the other hand, the predictive mean using sparse in-GP does not produce these artefacts (see Figure 9(e)) and tracks the data pattern better. The values of the predictive variance are plotted as a heat map in Figure 1(a) in the Supplementary Material. The level curves are orthogonal to the boundary in the north and east part of Aral sea in Figure 9(e). The inducing points are more sparse in the west part of the Aral sea. Approximation errors, due to insufficiently dense inducing points and/or Monte Carlo errors in approximating reflective Brownian motion with resampling, can lead to non-orthogonality in some cases.

These artefacts become even more pronounced when the coverage of the data is uneven. In Figure 9(b) we removed most of the data points in the southern part of the western basin of the sea, and the same models are applied to this uneven dataset. Figure 9(d) shows the Euclidean GP extrapolation across the isthmus from the eastern basin of the sea. In contrast, the sparse in-GP estimates as plotted in Figure 9(f) do not seem to be affected by the data from the eastern side of the isthmus. The values of the predictive variance are plotted as a heat map in Figure 1(b) in the Supplementary Material.





**Fig. 9.** Comparison of in-GP and an Euclidean GP for chlorophyll data in Aral sea. The remotely sensed chlorophyll data at 485 locations in the Aral sea are plotted in (a). The datasets are available from the gamair package (Wood, 2006). The GP prediction and the sparse in-GP prediction are shown in (c) and (e). It is clear that the Euclidean GP smooths across the isthmus of the central peninsula in (c). The sparse in-GP tracks the data pattern better. If we removed the data points in the southern part of the western basin of the sea in (b), these artefacts are exacerbated for the Euclidean GP prediction in (d). The sparse in-GP prediction in (f) is not affected by the data from the eastern side of the isthmus.

Since most of the data points in the southern part of the western basin of the sea have been removed, the values of the variance estimates have increased in this region. In order to compare with the soap film approach, we divide the datasets into ten equal size batches randomly and iteratively pick one batch as the training data to make prediction for the other nine batches. The mean RMSE for in-GP is 17.9 with standard deviation 0.71. The mean RMSE for the soap film method is 17.2 with standard deviation 1.72. The statistical test shows no significant difference between these two methods. However, the standard deviation of RMSE for the in-GP is much smaller.

## 5. Discussion

Our work proposes a novel class of intrinsic Gaussian processes on manifolds and complex constrained domains employing the equivalence relationship between heat kernels and the transition density of Brownian motions on manifolds. One of the key features of the in-GP is to fully incorporate the intrinsic geometry of the spaces for inference while respecting the potentially complex boundary or interior constraints. To reduce the computational cost of simulating BM sample paths when the sample size is large, sparse in-GPs are developed leveraging ideas from the literature on fast computation in GPs in Euclidean spaces. The results in section 3 and 4 indicate that in-GP achieves significant improvement over usual GPs. Although we did not conduct a formal asymptotic study for in-GPs, with insights gained from the Euclidean GP with squared exponential kernel in the Euclidean space (see, e.g., van der Vaart and van Zanten (2009)), we expect in-GPs to yield posterior consistency with respect to appropriate neighbourhoods of the true regression function  $f_0$ . The focus of this article has been on developing in-GPs on manifolds with known metric tensors. There has been abundant interest in learning of unknown lower-dimensional manifold structure in high-dimensional data. in-GPs can be combined with these approaches for performing supervised learning on lower-dimensional latent manifolds.

## Codes and Supplementary Material

R code implementation of examples in Sections 3 and 4 are available on Github repository: <https://github.com/mu2013/Intrinsic-GP-on-complex-constrained-domain>. The Supplementary Material includes a description on the choice of the sample and window sizes for estimating the heat kernel in  $\mathbb{R}$  and  $\mathbb{R}^2$ . It also provides some details on the Brownian motions on the Swiss Roll and the Bitten Torus. A discussion on the differences between reflecting and resampling for the Brownian motion is also provided. The last section of the Supplementary Material is devoted to the discussion of the computational cost and algorithm complexity.

## Acknowledgement

Lizhen Lin acknowledges support for this article from NSF grants DMS CAREER 1654579 and IIS 1663870.

**References**

- Aumentado-Armstrong, T. and Siddiqi, K. (2017) Stochastic heat kernel estimation on sampled manifolds. *Computer Graphics Forum*, **36**, 131–138.
- Berline, N., Getzler, E. and Vergne, M. (2003) *Heat Kernels and Dirac Operators*. Grundlehren Text Editions. Springer Berlin Heidelberg.
- Bhattacharya, A. and Dunson, D. (2010) Nonparametric Bayes regression and classification through mixtures of product kernels. *Bayesian Analysis*, **9**, 145–164.
- Burdzy, K., Chen, Z.-Q. and Sylvester, J. (2004) The heat equation and reflected brownian motion in time-dependent domains. *Ann. Probab.*, **32**, 775–804.
- Castillo, I., Kerkycharian, G. and Picard, D. (2014) Thomas bayes walk on manifolds. *Probab. Theory Relat. Fields*, 665710.
- Guinness, J. and Fuentes, M. (2016) Isotropic covariance functions on spheres: Some properties and modeling considerations. *Journal of Multivariate Analysis*, **143**, 143–152.
- Hsu, E. P. (2008) A brief introduction to Brownian motion on a Riemannian manifold. *Lecture Notes*.
- Hsu, P. (1988) Brownian motion and Riemannian geometry. *Contemporary Mathematics*, **73**, 95–104.
- Katzfuss, M. and Guinness, J. (2017) A general framework for Vecchia approximations of Gaussian processes. *arXiv preprint arXiv:1708.06302*.
- Kloeden, P. E. and Platen, E. (1992) Higher-order implicit strong numerical schemes for stochastic differential equations. *Journal of Statistical Physics*, **66**, 283–314.
- Lamberton, D. and Lapeyre, B. (2007) *Introduction to Stochastic Calculus Applied to Finance*. CRC press.
- Lin, L., Mu, N., Chan, P. and Dunson, D. B. (2018) Extrinsic Gaussian processes for regression and classification on manifolds. *Bayesian Analysis*. In press.
- Lions, P. L. and Sznitman, A. S. (1984) Stochastic differential equations with reflecting boundary conditions. *Communications on Pure and Applied Mathematics*, **37**, 511–537.
- Matheron, G. (1973) The intrinsic random functions and their applications. *Advances in Applied Probability*, **5**, 439–468.
- Ozakin, A. and Gray, A. G. (2009) Submanifold density estimation. In *Advances in Neural Information Processing Systems 22* (eds. Y. Bengio, D. Schuurmans, J. D. Lafferty, C. K. I. Williams and A. Culotta), 1375–1382. Curran Associates, Inc.
- Pelletier, B. (2005) Kernel density estimation on Riemannian manifolds. *Statistics and Probability Letters*, **73**, 297–304.

- Quiñonero-Candela, J. and Rasmussen, C. E. (2005) A unifying view of sparse approximate Gaussian process regression. *Journal of Machine Learning Research*, **6**.
- Quiñonero-Candela, J., Rasmussen, C. and Williams, C. (2007) *Approximation Methods for Gaussian Process Regression*, 203–223. Neural Information Processing. MIT Press.
- Ramsay, T. (2002) Spline smoothing over difficult regions. *J. Royal Stat. Soc. Series B*, **64**, 307–319.
- Rasmussen, C. E. (2004) *Gaussian Processes in Machine Learning*, 63–71. Springer Berlin Heidelberg.
- Sampson, P. D. and Guttorp, P. (1992) Nonparametric estimation of nonstationary spatial covariance structure. *J. Amer. Statist. Assoc.*, **87**, 108–119.
- Sangalli, L. M., Ramsay, J. O. and Ramsay, T. O. (2013) Spatial spline regression models. *J. Royal Stat. Soc. Series B*, **75**, 681–703.
- Schwaighofer, A. and Tresp, V. (2002) Transductive and inductive methods for approximate gaussian process regression. In *Advances in Neural Information Processing Systems 15*, 953–960.
- Snelson, E. and Ghahramani, Z. (2006) Sparse gaussian processes using pseudo-inputs. In *Advances in Neural Information Processing Systems 18* (eds. Y. Weiss, B. Schölkopf and J. C. Platt), 1257–1264. MIT Press.
- Titsias, M. (2009) Variational learning of inducing variables in sparse gaussian processes. In *Proceedings of the Twelfth International Conference on Artificial Intelligence and Statistics* (eds. D. van Dyk and M. Welling), vol. 5 of *Proceedings of Machine Learning Research*, 567–574. PMLR.
- Tosi, A., Hauberg, S., Vellido, A. and Lawrence, N. D. (2014) Metrics for probabilistic geometries. In *Proceedings of the Thirtieth Conference on Uncertainty in Artificial Intelligence*, UAI'14, 800–808. AUAI Press.
- van der Vaart, A. W. and van Zanten, J. H. (2009) Adaptive bayesian estimation using a Gaussian random field with inverse gamma bandwidth. *Ann. Stat.*, **37**, 2655–2675.
- Wood, S. (2006) *Generalized Additive Models: an Introduction with R*. Boca Raton, Florida: Chapman and Hall/CRC, 1st edn.
- Wood, S. N. (2001) mgcv: Gams and generalized ridge regression for R. *R news*, **1**, 20–25.
- Wood, S. N., Bravington, M. V. and Hedley, S. L. (2008) Soap film smoothing. *J. Royal Stat. Soc. Series B*, **70**, 931–955.
- Xu, P. (1984) *Reflecting Brownian motion, boundary local time and the Neumann problem*. Stanford University.
- Zhou, Y., Cai, W. and Hsu, E. (2017) Computation of the local time of reflecting brownian motion and the probabilistic representation of the Neumann problem. *Communications in Mathematical Sciences*, **15**, 237–259.