Gaussian Process on non-Euclidean Domain

Connections between Elliptical operators and RKHS

Yidan Xu

Department of Statistics
University of Washington

Table of contents

Motivations and Problem Setup

Why Gaussian Process (GP)?

GP has been used as a powerful tool in Bayesian non-parametric inference and serves as an alternative to NN for supervised and semi-supervised learning.

In the Bayesian optimisation framework, GP is often employed as a prior for a broad range of tasks in various fields, including:

- Spatial Statistics: Ecology, Climate science, Epidemiology;
- Dynamical System: Robotics, Reinforcement learning;
- Bayesian inverse problem: Medical imaging, Remote sensing, Ground prospecting;
- . . .

Stationary Matérn GP in Euclidean domain

Definition (Informal definition of GP)

If $f : \mathcal{X} \to \mathbb{R}$ such that $f \sim GP(m, k)$, with mean function m and Covariance function/kernel k, then for any subset $\boldsymbol{x} \in \mathcal{X}^n$, the random vector $\boldsymbol{f} = f(\boldsymbol{x})$ is multivariate Gaussian with mean $\boldsymbol{\mu} = m(\boldsymbol{x})$ and Gram matrix $K_{\boldsymbol{xx}} = k(\boldsymbol{x}, \boldsymbol{x})$.

The kernel k must be positive semi-definite, in the sense that for any $\boldsymbol{x} \in \mathcal{X}^n$, $K_{\boldsymbol{xx}}$ is positive semi-definite.

As a side note,

- We will assume a zero mean GP, with $m \equiv 0$.
- We will generally focus on stationary kernels, for which we can find a function $\ell : \mathcal{X} \to \mathbb{R}$ *s.t.* $k(x, x') = \ell(x - x')$.
- We will use GP and Gaussian Field (GF) interchangeably in the sequel (formal definition to come).

Example: Matérn family of kernels

For $x, x' \in \mathcal{X} = \mathbb{R}^d$, Matérn family of kernels admits the form

$$k_\nu(x, x') = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\sqrt{2\nu} \frac{\|x - x'\|}{\kappa} \right)^\nu K_\nu \left(\sqrt{2\nu} \frac{\|x - x'\|}{\kappa} \right) \tag{1}$$

with parameters ν, σ^2 and κ. K_ν is the modified Bessel function of the second kind.

A special example is squared exponential kernel (RBF kernel) given by $\nu \to \infty$,

$$k_\infty(x, x') = \sigma^2 \exp \left(-\frac{\|x - x'\|^2}{2\kappa^2} \right)$$

Extension to Manifold setting: A no-go attempt

On a m-dimensional compact Riemannian manifold $(\mathcal{M}, g)$ without boundary, one may wish to replace the Euclidean norm in the kernel with geodesic distance d_g *w.r.t.* g on $\mathcal{M}$.

This is generally not a well-defined kernel. For example,

Theorem (Feragen et al. Theorem 2)

For $(\mathcal{M}, g)$ defined above, if the geodesic squared exponential kernel below is positive semi-definite $\forall \kappa^2 > 0$, then $\mathcal{M}$ is isometric to an Euclidean space.

$$k(x, x') = \sigma^2 \exp\left(-\frac{d_g(x, x')^2}{2\kappa^2}\right)$$

Since Euclidean space is not compact, k is not well-defined.

SPDE approach to GP on Riemannian Manifold

An alternative would be finding the solution for some Elliptical SPDE on $(\mathcal{M}, g)$. This has been a relatively well-studied approach in Spatial statistics, with the following important characterisation of Matérn given by Whittle [7], [8].

SPDE approach: Matérn GP

For $\mathcal{X} = \mathbb{R}^d$, a GP $f(x)$ with covariance function (1) is the unique solution to the SPDE,

$$(\tau I - \Delta)^{\frac{s}{2}} f(x) = \mathcal{W}(x) \tag{2}$$

where $\tau = \frac{2\nu}{\kappa^2}$ and $s = \nu + \frac{d}{2}$. $\mathcal{W}$ denotes the spatial Gaussian white noise with unit variance. In addition, the marginal variance of $f(x)$ is

$$\sigma^2 = \frac{\Gamma(s - \frac{d}{2})}{(4\pi)^{\frac{d}{2}} \Gamma(s) \tau^{s - \frac{d}{2}}}$$

Some theoretical advantages:

- Easily extendable to non-stationary kernels: *employ general Elliptic operators* $\nabla \cdot (\gamma(x)\nabla)$ *in place of Laplace operator; and/or let* $\tau = \tau(x)$ *dependant on spatial variation.*

Some theoretical advantages:

- Easily extendable to non-stationary kernels:
  employ general Elliptic operators $\nabla \cdot (\gamma(x)\nabla)$ *in place of Laplace operator; and/or let* $\tau = \tau(x)$ *dependant on spatial variation.*
- Generalisation to compact Riemannian manifold is straight-forward:
  replace Δ *with Laplace-Beltrami operator* Δ_{LB} *w.r.t.* $(\mathcal{M}, g)$*; and define* $\mathcal{W}$ *w.r.t. Riemannian volume measure.*

Some theoretical advantages:

- Easily extendable to non-stationary kernels:
  employ general Elliptic operators $\nabla \cdot (\gamma(x)\nabla)$ *in place of Laplace operator; and/or let* $\tau = \tau(x)$ *dependant on spatial variation.*
- Generalisation to compact Riemannian manifold is straight-forward:
  replace Δ *with Laplace-Beltrami operator* Δ_{LB} *w.r.t.* $(\mathcal{M}, g)$*; and define* $\mathcal{W}$ *w.r.t. Riemannian volume measure.*
- It allows elegant extensions to Spatial-Temporal domain :
  consider Parabolic SPDE instead of Elliptical SPDE, for example stochastic heat equation.

Some theoretical advantages:

- Easily extendable to non-stationary kernels:
  employ general Elliptic operators $\nabla \cdot (\gamma(x)\nabla)$ *in place of Laplace operator; and/or let* $\tau = \tau(x)$ *dependant on spatial variation.*
- Generalisation to compact Riemannian manifold is straight-forward:
  replace Δ *with Laplace-Beltrami operator* Δ_{LB} *w.r.t.* $(\mathcal{M}, g)$*; and define* $\mathcal{W}$ *w.r.t. Riemannian volume measure.*
- It allows elegant extensions to Spatial-Temporal domain :
  consider Parabolic SPDE instead of Elliptical SPDE, for example stochastic heat equation.
- The solution is amenable to approximation with point-cloud data under manifold assumption [6]; and also compatible with sparse GP techniques for scalable training [2].

Gaussian Field and SPDE

Spectral property of Laplace-Beltrami Operator

Just as in Δ defined on bounded domain $G \subset \mathbb{R}^d$ with boundary conditions (usually Dirichlet), the Laplace-Beltrami operator defined on compact manifold also allows a spectral representation.

Theorem (Sturm-Liouville decomposition)

Let $(\mathcal{M}, g)$ be a compact Riemannian manifold without boundary, there exists an orthonormal basis $\{\phi_i\}_{i\in\mathbb{Z}^+}$ of $\mathcal{L}^2(\mathcal{M})$ such that $-\Delta_{LB}\phi_i = \lambda_i\phi_i$ with $0 = \lambda_0 < \lambda_1 \leq \cdots \leq \lambda_i \to 0$ as $i \to \infty$. Moreover, $-\Delta_{LB}$ admits the representation

$$-\Delta_{LB} f = \sum_{i=0}^{\infty} \lambda_i \langle f, \phi_i \rangle_{\mathcal{L}^2} \phi_i \tag{3}$$

which converges unconditionally in $\mathcal{L}^2(\mathcal{M})$.

Why this relates to RKHS

Recall Mercer's Theorem in the theory of RKHS,

Theorem (Mercer's Theorem, S.Saitoh 2016[5])

Assume $\mathcal{X}$ is a locally compact Hausdorff space equipped with a positive Borel measure μ. In addition, let $\mathcal{L}^2_\mu(\mathcal{X})$ be the separable Hilbert space of real square-integrable functions defined on $\mathcal{X}$. Consider the Integral operator for bilinear form $k \in \mathcal{L}^2_\mu(\mathcal{X} \times \mathcal{X})$,

$$(\mathcal{K}f)(y) = \int_{\mathcal{X}} f(x)k(x,y)\, d\mu(x), \quad \mathcal{K} : \mathcal{L}^2(\mathcal{X}) \to \mathcal{L}^2(\mathcal{X}), \tag{4}$$

and assume k satisfies the following assumptions:

1. $k(x,y) = k(y,x)\ \forall x, y \in supp(\mu)$;
2. $\int\int_{\mathcal{X}\times\mathcal{X}} k(x,y)f(x)f(y)\, d\mu(y) \geq 0,\ \forall f \in \mathcal{L}^2_\mu(\mathcal{X})$.

(Theorem continues...)

Then $\mathcal{K}$ admits a countable set of non-negative eigenvalues $\{\eta_i\}_i^\infty$ and corresponding orthonormal eigenfunctions $\{\phi_i\}_i^\infty$, where $\mathcal{K}f = \sum_i \eta_i \langle f, \phi_i \rangle_{\mathcal{L}^2} \phi_i$.

Convergence of the infinite sum is absolute and uniform on supp$(\mu \times \mu)$.

In fact, $\mathcal{K}$ gives an isomorphism from $\mathcal{L}^2(\mathcal{X})$ to a RKHS with reproducing kernel k, defined by

$$H_k = \left\{ f \in \mathcal{L}^2_\mu(\mathcal{X}) : f = \sum_i^\infty a_i \phi_i, \ \left(\frac{|a_i|^2}{\eta_i} \right) \in \ell^2 \right\},$$

with inner product of the form $\langle f, g \rangle_{H_k} = \sum_i^\infty \frac{a_i b_i}{\eta_i}$ where $f = \sum_i^\infty a_i \phi_i$, $g = \sum_i^\infty b_i \phi_i$, and $a_i, b_i = 0$ for $\lambda_i = 0$.

It's then straight-forward to see that

$$\mathcal{K}f = \sum_{i=0}^{\infty} \eta_i \langle f, \phi_i \rangle_{\mathcal{L}^2} \phi_i.$$

This has a similar form to (3) but not exactly.

Although $-\Delta_{LB}$ is positive and self-adjoint, it is however not bounded. Its spectral decomposition is in fact given by its resolvent $-(I + \Delta_{LB})^{-1}$, which is compact self-adjoint and positive definite[1] and therefore an Integral (Hilbert-Schmidt) operator in the form of (4).

The Borel function calculus subsequently gives the Sturm-Liouville decomposition of $-\Delta_{LB}$, where $\lambda_i = 1/\eta_i + 1$.

[1] *https://gauss.math.yale.edu/~mrm89/lecture5.pdf#page3*

Gaussian Measure

Consider a measurable space $(V, \mathcal{B}(V))$, where V is a linear topological space and $\mathcal{B}(V)$ the Borel σ-algebra of V. Then,

Definition (Gaussian Measure on Topological space, Def 3.2.4 [4])

A probability measure μ is called Gaussian if for every linear functional $\ell \in V'$ such that $\ell : V \to \mathbb{R}$, the probability measure $\ell^* \mu$ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ defined by

$$\ell^* \mu((a, b]) = \mu\{v \in V : \ell(v) \in (a, b]\}$$

is Gaussian. In addition, μ is called centered if $\ell^* \mu$ is centered for any $\ell \in V'$.

The Covariance function for centered Gaussian measure μ is

$$C_\mu(\ell, h) = \mathbb{E}_\mu[\ell h] = \int_V \ell(v) h(v) \, \mu(dv), \quad \ell, g \in V'.$$

In particular, for $V = \mathcal{H}$ a separable Hilbert space identified with its dual, i.e. $\mathcal{H} = \mathcal{H}'$, the following holds

Theorem (Thm 3.2.5 [4])

1. *If μ is a Gaussian measure on $\mathcal{H}$, then its covariance operator is nuclear.*
2. *Conversely, if $\mathcal{K}$ is a self-adjoint non-negative nuclear operator, then there exists a Gaussian measure μ on $\mathcal{H}$ such that $C_\mu = \langle \mathcal{K} f, g \rangle_{\mathcal{H}}$ for all $f, g \in \mathcal{H}$.*

Clearly, $\mathcal{K}$ is an Integral operator as in Mercer's Theorem, and is usually called the Covariance operator for μ.

In the case of 2., consider the set of eigenfunctions $\{\phi_k\}_{k\geq 1}$ of $\mathcal{K}$ with eigenvalues $\{\lambda_k\}_{k\geq 1}$, then define a Gaussian measure μ as the probability distribution of the random variable X, where

$$X = \sum_{k\geq 1} \sqrt{\lambda_k}\xi_k\phi_k.$$

and $\{\xi_k\}_k$ is a set of *i.i.d.* standard Gaussian r.v. This is called the Karhunen-Loève expansion of Gaussian random element X.

Generalised Gaussian Field

Definition (Zero-mean generalised Gaussian field, Def 3.2.10 [4])

A zero mean generalised Gaussian field $\mathfrak{X}$ over a Hilbert space $\mathcal{H}$ is a collection of Gaussian random variables $\{\mathfrak{X}(f) : f \in \mathcal{H}\}$ with

1. $\mathbb{E}\mathfrak{X}(f) = 0 \ \forall f \in \mathcal{H}$
2. There exists a bounded, linear, self-adjoint, non-negative $\mathcal{K}$ on $\mathcal{H}$ such that

$$C_\mu(\mathfrak{X}(f), \mathfrak{X}(g)) = \mathbb{E}[\mathfrak{X}(f)\mathfrak{X}(g)] = \langle \mathcal{K}f, g \rangle_{\mathcal{H}}$$

By the duality of $\mathcal{H}$, for any $\ell \in \mathcal{H}'$, $\exists! X \in \mathcal{H}$ s.t. $\ell(v) = \langle v, X \rangle_{\mathcal{H}} \forall v \in \mathcal{H}$[2]. Intuitively, $\mathfrak{X}$ is the functional ℓ, and X is the $\mathcal{H}$-valued random element with Gaussian measure μ. This is not generally true!

[2] This is due to the Riesz representation theorem.

Regularity of Generalised GF

Definition (Regularity, Def 3.2.13 [4])

A generalised field $\mathfrak{X}$ over Hilbert space $\mathcal{H}$ is regular if there exists an $\mathcal{H}$-valued random element $X \in \mathcal{L}^2$, i.e. $\mathbb{E}\|X\|_{\mathcal{H}}^2 < \infty$ such that

$$\mathfrak{X}(f) = \langle X, f \rangle_{\mathcal{H}} \quad \forall f \in \mathcal{H}.$$

In particular, we have the following important characterisation of generalised GF on seperable Hilbert space.

Theorem (Main characterisation, Thm 3.2.15 [4])

A generalised field $\mathfrak{X}$ over a separable Hilbert space $\mathcal{H}$ is regular iff $\mathcal{K}$ of $\mathfrak{X}$ is nuclear.

Upshot of the Main Characterisation

Fact

One can show for separable Hilbert space $\mathcal{H}$ with CONS $\{m_k\}_{k\geq 1}$, $\{\mathfrak{X}(m_k)\}_{k\geq 1}$ is a collection of *i.i.d.* standard Gaussian r.v.

From the Theorem, if $\mathcal{K}$ is nuclear with eigenfunction $\{\phi_k\}_k$ and eigenvalues $\{\lambda_k\}_k$, we may define *i.i.d.* standard Gaussian r.v. as

$$\xi_k = \frac{1}{\sqrt{\lambda_k}}\mathfrak{X}(\phi_k)$$

so the $\mathcal{H}$-valued Gaussian random element is defined by

$$X = \sum_k \langle X, \phi_k\rangle_{\mathcal{H}}\phi_k = \sum_k \mathfrak{X}(\phi_k)\phi_k = \sum_k \sqrt{\lambda_k}\xi_k\phi_k$$

Indeed, $\mathbb{E}\|X\|_{\mathcal{H}}^2 = \sum_k \lambda_k < \infty$ when $\mathcal{K}$ is nuclear.

In addition, for any $f, g \in \mathcal{L}^2(\mathcal{M}) = \mathcal{H}$,

$$\begin{aligned}
\mathbb{E}[\mathfrak{X}(f)\mathfrak{X}(g)] &= \mathbb{E}[\langle X, f\rangle\langle X, g\rangle] \\
&= \sum_k \sum_j f_k g_j \mathbb{E}[\langle X, \phi_k\rangle\langle X, \phi_j\rangle] \\
&= \sum_k f_k g_k \int\int \mathbb{E}[X(x)X(y)]\phi_k(x)\phi_k(y)\ dxdy \\
&= \sum_k f_k g_k \langle \mathcal{K}\phi_k, \phi_k\rangle \\
&= \sum_k \lambda_k f_k g_k = \langle \mathcal{K}f, g\rangle
\end{aligned}$$

where $k(x, y) := \mathbb{E}[X(x)X(y)]$ is the reproducing kernel, and $\mathcal{K}$ the Covariance operator for Gaussian measure μ (of X).

Then X is in fact a Gaussian random element in RKHS H_k associated to the kernel k.

Gaussian White Noise

When $\mathcal{K} = \mathrm{Id}$, $\mathfrak{X}$ is the Gaussian white noise on $\mathcal{H}$. In this case, the Gaussian field $\mathfrak{X}$ is not regular.

If $\mathcal{H} = \mathcal{L}^2(\mathcal{M})$, we denote such $\mathfrak{X}$ as $\mathcal{W}$. Clearly,

$$\mathbb{E}[\mathcal{W}(f)\mathcal{W}(g)] = \langle f, g \rangle, \quad \forall f, g \in \mathcal{L}^2(\mathcal{M})$$

Elliptical SPDE on Manifold

Definition (Solution to SPDE system, Def 4.2.1 [4])

Let $\mathcal{H}$ be a Hilbert space and $L : \mathcal{H} \to \mathcal{L}^2(\mathcal{M})$ be a bounded linear operator. Then the zero-mean generalised Gaussian random field $\mathfrak{X}$ over $\mathcal{H}$ is a solution to

$$L\mathfrak{X} = \mathcal{W} \tag{5}$$

if for every $g \in \mathcal{L}^2(\mathcal{M})$

$$\mathfrak{X}(L^* g) = \mathcal{W}(g).$$

We will generally assume (5) is an Elliptical SPDE, where L is an Elliptical operator (e.g. a function of Δ_{LB})

Theorem (Thm 4.2.2 [4])

If L is invertible, then a zero-mean generalised Gaussian field $\mathfrak{X}$ over $\mathcal{H}$ defined by

$$\mathfrak{X}(h) = \mathcal{W}((L^{-1})^* h)$$

is the unique solution to (5).

since we know the Integral operator for $\mathcal{W}$ is identity

$$\begin{aligned}
\mathbb{E}[\mathfrak{X}(f)\mathfrak{X}(g)] &= \mathbb{E}[\mathcal{W}((L_\kappa^1)^* f)\mathcal{W}((L_\kappa^{-1})^* g)] \\
&= \langle (L_\kappa^{-1})^* f, (L_\kappa^{-1})^* g \rangle_{\mathcal{L}^2(G)} \\
&= \langle f, L_\kappa^{-1}(L_\kappa^{-1})^* g \rangle_{\mathcal{L}^2(G)} \\
&= \langle f, \mathcal{K} g \rangle_{\mathcal{L}^2(G)}
\end{aligned}$$

Matérn Gaussian Field

Going back to the example for Matèrn GF, where

$$(\tau I - \Delta)^{\frac{s}{2}} f(x) = \mathcal{W}(x)$$

Then $\mathcal{K} = \tau^{s-\frac{m}{2}}(\tau I - \Delta)^{-s}$, the reproducing kernel admits the expansion (up to scaling of a constant)

$$k(x, y) = \mathbb{E}[X(x)X(y)] = \tau^{s-\frac{m}{2}} \sum_{k \geq 1} (\tau + \lambda_k)^{-\frac{s}{2}} \phi_k(x)\phi_k(y)$$

due to the Spectral property of Laplace operator and Borel functional calculus. The GF is of the form

$$f(x) = \tau^{\frac{s}{2}-\frac{m}{4}} \sum_{k} (\tau + \lambda_k)^{-\frac{s}{2}} \phi_k(x)\xi_k$$

and $f \sim GP(0, \tau^{s-\frac{m}{2}}(\tau I - \Delta)^{-s})$.

Extension to non-stationary Gaussian Field

Consider the Elliptical operator $\nabla \cdot (\gamma(x)\nabla)$, the SPDE

$$(\tau(x)I - \nabla \cdot (\gamma(x)\nabla))^{\frac{s}{2}} f(x) = \mathcal{W}(x)$$

gives a non-stationary second order structure for the Gaussian field. The Karhunen-Loève expansion for $f(x)$ is

$$f(x) = \tau(x)^{\frac{s}{2} - \frac{m}{4}} \gamma(x)^{\frac{m}{4}} \sum_k (\lambda_k)^{-\frac{s}{2}} \phi_k(x) \xi_k.$$

with marginal variance at each $x \in \mathcal{M}$ proportional to $\tau(x)^{\frac{s}{2} - \frac{m}{4}} \gamma(x)^{\frac{m}{4}}$. Then, $f \sim GP(0, [\tau I - \nabla \cdot (\gamma\nabla)]^{-s})$

In this case the eigenpairs $\{(\lambda_k, \phi_k)\}_k$ is from $\tau I - \nabla \cdot (\gamma\nabla)$.

Matérn GP on Riemannian Manifold (Borovitskiy et al. 2020)

Implementation details

- For d dimensional torus, there are closed form of the eigenfucntions and eigenvalues available, therefore the GF is known;
- For d dimensional Hyper-sphere, need to work with spherical harmonics directly. Low-rank approximation for the GF is required with truncated Karhunen-Loève expansion (HLE). [3]
- For general compact Riemannian manifold without boundary, numerical solver (FEM) is employed to find approximations to eigenpairs of Δ_{LB}, with truncated HLE.
  FEM is not generally applied to $d > 3$, due to the complication of mesh construction in high dimensions.

[3] According to [3], there's currently no stable spherical harmonic implementations available for high dimensional data.

Conclusion

[2] has not provided with a generally useful or applicable solution to high dimensional problems in the general compact Riemannian manifold setting.

What to look for next?

Graph Representation of Matérn GP (Sanz-Alonso & Yang 2020)

Contributions

- Introduced Gaussian Markov Random Field (GMRF) approximation of stationary and non-stationary GF, based on discrete SPDE (with graph Laplacian).
- Extension of GP on manifold to high-dimensional point-cloud data and graph-structured data with manifold assumption.
- Established rates of convergence for graph-based GF to the continuum counterpart in spectrum using tools from Optimal Transport.

Implementation details

- The graph Laplacian is constructed from point cloud data with KNN-graph or ϵ-neighborhood graph, and normalised by the probability density of the data estimated with KDE [4].
- The GMRF approximation comes from the sparse graph Laplacian in the SPDE, which gives an extension to the ICAR model from Besag[5] [1].

[4] There is, however, curse of dimensionality with KDE.

[5] ICAR is also a graph-based GF model $f \sim N(0, (D - W)^{-1})$, with unormalised graph Laplacian as its precision matrix.

Thanks for Listening!

Any Questions?

References i

J. Besag, J. York, and A. Mollié.
Bayesian image restoration, with two applications in spatial statistics.
Annals of the institute of statistical mathematics, 43(1):1–20, 1991.

V. Borovitskiy, A. Terenin, P. Mostowsky, and M. P. Deisenroth.
Matern gaussian processes on riemannian manifolds.
arXiv preprint arXiv:2006.10160, 2020.

V. Dutordoir, N. Durrande, and J. Hensman.
Sparse gaussian processes with spherical harmonic features.
arXiv preprint arXiv:2006.16649, 2020.

S. V. Lototsky and B. L. Rozovsky.
Stochastic partial differential equations.
Springer, 2017.

S. Saitoh and Y. Sawano.
Theory of reproducing kernels and applications.
Springer, 2016.

D. Sanz-Alonso and R. Yang.
The spde approach to matérn fields: Graph representations.
arXiv preprint arXiv:2004.08000, 2020.

P. Whittle.
On stationary processes in the plane.
Biometrika, pages 434–449, 1954.

P. Whittle.
Stochastic-processes in several dimensions.
Bulletin of the International Statistical Institute, 40(2):974–994, 1963.