Stationary Kernels and Gaussian Processes on Lie Groups and Homogeneous Spaces I: the compact case

Iskander Azangulov (University of Oxford), Andrei Smolensky (Neapolis University Pafos), Alexander Terenin (Cornell University), and Viacheslav Borovitskiy (University of Edinburgh)

Abstract

Gaussian processes are arguably the most important class of spatiotemporal models within machine learning. They encode prior information about the modeled function and can be used for exact or approximate Bayesian learning. In many applications, particularly in physical sciences and engineering, but also in areas such as geostatistics and neuroscience, invariance to symmetries is one of the most fundamental forms of prior information one can consider. The invariance of a Gaussian process' covariance to such symmetries gives rise to the most natural generalization of the concept of stationarity to such spaces. In this work, we develop constructive and practical techniques for building stationary Gaussian processes on a very large class of non-Euclidean spaces arising in the context of symmetries. Our techniques make it possible to (i) calculate covariance kernels and (ii) sample from prior and posterior Gaussian processes defined on such spaces, both in a practical manner. This work is split into two parts, each involving different technical considerations: part I studies compact spaces, while part II studies non-compact spaces possessing certain structure. Our contributions make the non-Euclidean Gaussian process models we study compatible with well-understood computational techniques available in standard Gaussian process software packages, thereby making them accessible to practitioners.

Stationary Kernels via Representation Theory

Theorem (Yaglom, 1961). *A Gaussian process $f \sim \mathrm{GP}(0,k)$ on a compact Lie group G is stationary with respect to left-and-right action of G on itself if and only if k is of form*

$$k(g_1, g_2) = \sum_{\lambda\in\Lambda} a^{(\lambda)} \operatorname{Re}\chi^{(\lambda)}(g_2^{-1}\bullet g_1)$$

where Λ is the set of all irreducible unitary representations, $a^{(\lambda)} \geq 0$ satisfy $\sum_{\lambda\in\Lambda} d_\lambda a^{(\lambda)} < \infty$. Moreover, for all λ, $\operatorname{Re}\chi^{(\lambda)}$ is a positive-definite function.

Theorem (Yaglom, 1961). *A Gaussian process $f \sim \mathrm{GP}(0,k)$ on a compact homogeneous space G/H is stationary with respect to the action of G if and only if k is of form*

$$k(g_1\bullet H, g_2\bullet H) = \sum_{\lambda\in\Lambda}\sum_{j,k=1}^{r_\lambda} a_{jk}^{(\lambda)} \operatorname{Re}\pi_{jk}^{(\lambda)}(g_2^{-1}\bullet g_1)$$

where the coefficients $a_{jk}^{(\lambda)} \in \mathbb{R}$ form symmetric positive semi-definite matrices $\mathbf{A}^{(\lambda)}$ of size $r_\lambda \times r_\lambda$ satisfying $\sum_{\lambda\in\Lambda} \operatorname{tr}\mathbf{A}^{(\lambda)} < \infty$, and $\pi_{jk}^{(\lambda)}$ are the zonal spherical functions. Moreover, for each individual $\lambda \in \Lambda$, the corresponding sum over j,k is positive semi-definite.

This work's goal: make calculating these expressions practical

Computational Techniques for General Kernels

Challenge 1: Λ is the set of all irreducible unitary representations of G, which is infinite. How should it be truncated?

- Idea: introduce an order using *signatures.*

Approach: the set of irreducible unitary representations of G is in explicit bijective correspondence with the set of *signatures* $\Lambda \subseteq \mathbb{Z}^m$.

Challenge 2: on a Lie group, how do we calculate the characters χ?

- Approach: apply the Weyl character formula.

Method: every character $\chi^{(\lambda)}$ with signature λ can be expressed as

$$\chi^{(\lambda)}(g) = \chi^{(\lambda)}(t) = \frac{P_1(t)}{P_2(t)} \qquad t = \tilde{g}^{-1}\bullet g\bullet\tilde{g}$$

where P_i are explicit polynomials whose order does not grow with λ.

Challenge 3: on G/H, what if $\pi_{jk}^{(\lambda)}$ is non-explicit?

- Idea: *generalized periodic summation.*

Theorem. *Let $f \sim \mathrm{GP}(0,k)$, where k is stationary and satisfies $\mathbf{A}^{(\lambda)} = a^{(\lambda)}\mathbf{I}$. Then*

$$k(g_1\bullet H, g_2\bullet H) = \int_H k_G(g_1\bullet h, g_2)\,\mathrm{d}\mu_H(h)$$

where μ_H is the Haar measure on H, and $f_G \sim \mathrm{GP}(0,k_G)$ with $k_G(g_1,g_2) = \sum_{\lambda\in\Lambda} a^{(\lambda)}\operatorname{Re}\chi^{(\lambda)}(g_2^{-1}\bullet g_1)$.

Challenge 4: can we efficiently sample the prior $f \sim \mathrm{GP}(0,k)$?

- Idea: *generalized random phase Fourier features.*

Lie group: for an explicit $K_k^{(\lambda)}$ we have

$$f(g) \approx \sum_{\lambda\in\tilde{\Lambda}} \frac{1}{\sqrt{S}} \sum_{s=1}^{S} w_s^{(\lambda)} K_k^{(\lambda)}(g, u_s) \quad u_s \sim \mu_G \quad w_s^{(\lambda)} \sim \mathrm{N}\left(0, \frac{a^{(\lambda)}}{d_\lambda}\right).$$

Homogeneous space: with the notation as above, we have

$$f(g\bullet H) \approx \sum_{\lambda\in\tilde{\Lambda}}\sum_{k=1}^{r_\lambda} \frac{1}{\sqrt{S}} \sum_{s=1}^{S} w_s^{(\lambda)} K_k^{(\lambda)}(g\bullet H, u_s) \qquad w_s^{(\lambda)} \sim \mathrm{N}\left(0, \frac{\alpha_k^{(\lambda)}}{d_\lambda}\right) \quad u_s \sim \mu_{G/H}.$$

Heat and Matérn Kernels

Define the *heat kernel* $k_{\infty,\kappa,\sigma^2}(x,x') = \frac{\sigma^2}{C_\kappa}\mathcal{P}(\kappa^2/2, x, x')$ as solution of

$$\frac{\partial\mathcal{P}}{\partial t}(t,\boldsymbol{x},\boldsymbol{x}') = \Delta_x \mathcal{P}(t,\boldsymbol{x},\boldsymbol{x}') \qquad \mathcal{P}(0,\boldsymbol{x},\boldsymbol{x}') = \delta(\boldsymbol{x}-\boldsymbol{x}')$$

where Δ is the Laplace–Beltrami operator. This generalizes the classical Euclidean squared exponential kernel.

- **Challenges**: how to compute this? What about Matérn?

Proposition. *The heat kernel on G is stationary, and is given by*

$$k_{\infty,\sqrt{2t},(4\pi t)^{-n/2}}(g_1,g_2) = \mathcal{P}(t,g_1,g_2) = \sum_{\lambda\in\Lambda} e^{-\alpha_\lambda t} d_\lambda \chi_\lambda(g_2^{-1}\bullet g_1)$$

where α_λ are eigenvalues of $-\Delta$, which are in explicit bijective correspondence with signatures $\lambda\in\Lambda$, and $d_\lambda \in \mathbb{N}$ are explicit constants.

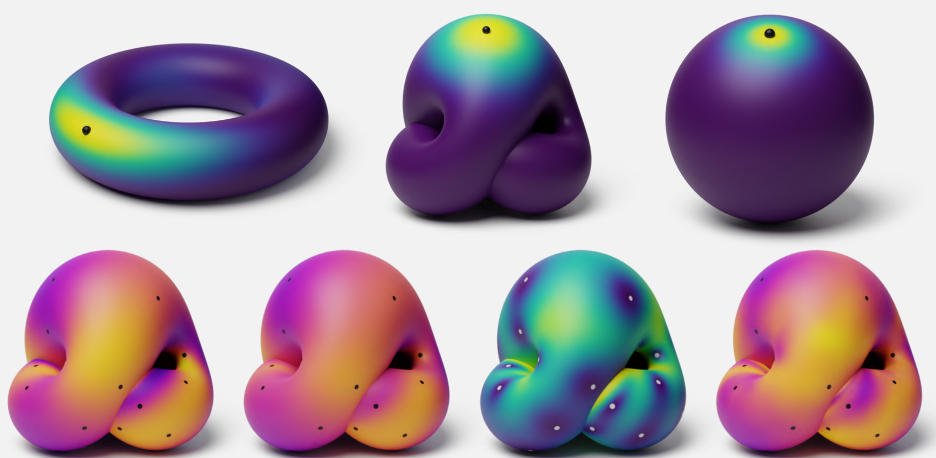
Proposition. *The heat kernel on G/H is stationary, and is given by*

$$k_{\infty,\sqrt{2t},(4\pi t)^{-n/2}}(g_1\bullet H, g_2\bullet H) = \sum_{\lambda\in\Lambda} e^{-\alpha_\lambda t} d_\lambda \sum_{j,k=1}^{r_\lambda} \pi_{jk}^{(\lambda)}(g_2^{-1}\bullet g_1).$$

Matérn kernels: we *define* them using the integral representation

$$k_{\nu,\kappa,\sigma^2}(x,x') = \frac{\sigma^2}{C_{\nu,\kappa}}\int_0^\infty u^{\nu-1+n/2} e^{-\frac{2\nu}{\kappa^2}u}\mathcal{P}(u,x,x')\,\mathrm{d}u.$$

Visuals: Kernel and Gaussian Process Regression



Check out the GeometricKernels Python package!