Stationary Kernels and Gaussian Processes on Lie Groups and Homogeneous Spaces II: non-compact symmetric spaces

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

Recall: on a *compact* homogeneous space, stationary kernels satisfy

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

where Λ is the set of all irreducible unitary representations, $a_{jk}^{(\lambda)} \in \mathbb{R}$ form 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.

Result. *Let G be a non-compact simple non-abelian Lie group. Then the only kernels on G stationary with respect to both left and right action of G on itself are identically constant.*

Key issue: viewing a Lie group G as a homogeneous space $G/\{e\}$, stationarity implies an infinite-dimensional analog of $\mathbf{A}^{(\lambda)} = a^{(\lambda)}\mathbf{I}$, but since we are in infinite dimension we cannot have $\sum_\lambda \operatorname{tr} \mathbf{A}^{(\lambda)} < \infty$.

- Approach: instead of G, study *symmetric spaces*, where $r_\lambda = 1$.

Theorem (Yaglom, 1961). *Let $X = G/H$ be a non-compact symmetric space, where G is a Lie group of type I. A Gaussian process $f \sim \mathrm{GP}(0, k)$ is stationary on X if and only if k is of form*

$$k(x, x') = k(g_1 \bullet H, g_2 \bullet H) = \Bbbk(g_2^{-1} \bullet g_1) = \int_\Lambda \pi^{(\lambda)}(g_2^{-1} \bullet g_1) \,\mathrm{d}\mu_k(\lambda)$$

for a function $\Bbbk : G \to \mathbb{R}$, where $g_1 \bullet H$ and $g_2 \bullet H$ are the cosets of x and x', Λ is the set of all—not necessarily finite dimensional—irreducible unitary representations of G with $r_\lambda = 1$, $\pi^{(\lambda)}$ are the zonal spherical functions, and μ_k is a nonnegative finite measure over Λ.

Computational Techniques for Symmetric Spaces

Challenge 1: Λ is the set of all irreducible unitary representations, which might be uncountable. How do we work with this numerically?

- Idea: leverage the *Iwasawa decomposition.*

Approach: there is an explicit surjective correspondence between a subset $\Lambda_A \subseteq \Lambda$ and the Euclidean space $\mathfrak{a}^*$, defined as the dual of the Lie algebra of the abelian part of the Iwasawa decomposition of G. Moreover,m for any stationary k, $\operatorname{supp} \mu_k \subseteq \Lambda_A$, and μ_k has a density.

Challenge 2: how do we integrate over μ_k and compute $\pi^{(\lambda)}$?

- Approach: work with Helgason's *spherical Fourier transform.*

Proposition. *Let $\Bbbk \in L^2(G)$. Then there is a non-negative $a^{(\lambda)} \in L^2(\mathfrak{a}^*, |c(\lambda)|^{-2} \,\mathrm{d}\lambda)$, where the L^2-weight function is defined in terms of Harish-Chandra's c-function, such that*

$$k(g_1 \bullet H, g_2 \bullet H) = \int_\Lambda \pi^{(\lambda)}(g_2^{-1} \bullet g_1) \,\mathrm{d}\mu_k(\lambda) = \int_{\mathfrak{a}^*} a^{(\lambda)} \pi^{(\lambda)}(g_2^{-1} \bullet g_1) |c(\lambda)|^{-2} \,\mathrm{d}\lambda$$

where $\pi^{(\lambda)}$ in the latter case are the zonal spherical functions obtained via the preceding bijective correspondence.

Result. *For $\lambda \in \mathfrak{a}^*$, and μ_H the normalized Haar measure, we have*

$$\pi^{(\lambda)}(g) = \int_H e^{(i\lambda+\rho)^\top a(h \bullet g)} \,\mathrm{d}\mu_H(h),$$

where $a(g) \in \mathfrak{a}$ satisfy $A(g) = \exp a(g)$, $\exp : \mathfrak{g} \to G$ is the exponential map, $\rho \in \mathfrak{a}^*$ is an explicit vector.

Overall approximation: symmetric space random Fourier features

$$\Bbbk(g) \approx \frac{\sigma^2}{L} \sum_{l=1}^{L} e^{(i\lambda_l+\rho)^\top a(h_l \bullet g)} \qquad \lambda_l \sim \frac{1}{\sigma^2}\mu_k \qquad h_l \sim \mu_H.$$

Challenge 3: can we efficiently sample the prior $f \sim \mathrm{GP}(0, k)$?

- Approach: random phase formula for zonal spherical functions.

Non-compact symmetric space: we have

$$f(x) \approx \frac{\sigma}{\sqrt{L}} \sum_{l=1}^{L} w_l e^{(i\lambda_l+\rho)^\top a(h_l \bullet x)} \quad \lambda_l \sim \frac{1}{\sigma^2}\mu_k \quad h_l \sim \mu_H \quad w_l \sim \mathrm{N}(0,1).$$

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

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

where Δ is the Laplace–Beltrami operator. This generalizes the classical Euclidean squared exponential kernel. Using this, we can obtain Matérn kernels as well through their integral representation.

- **Challenge**: how to compute this?

Proposition. *The heat and Matérn kernels on a non-compact symmetric space $X = G/H$ are stationary, and given by*

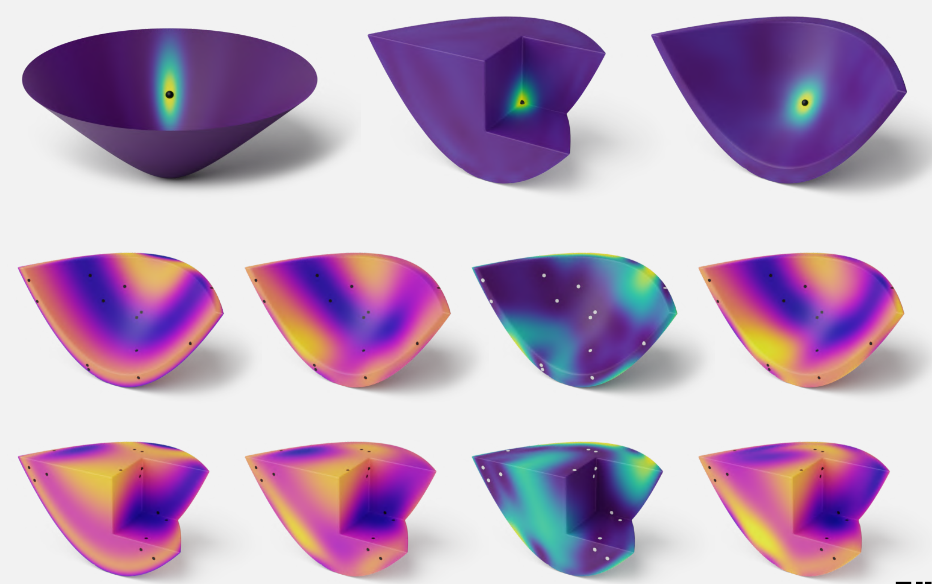
$$k_{\nu,\kappa,\sigma^2}(g_1 \bullet H, g_2 \bullet H) = \int_{\mathfrak{a}^*} \pi^{(\lambda)}(g_2^{-1} \bullet g_1) \,\mathrm{d}\mu_{\nu,\kappa,\sigma^2}(\lambda),$$

where the respective spectral measures $\mu_{\nu,\kappa,\sigma^2}$ have an explicit form.

Spectral measures μ_k for the heat kernel on specific spaces:

- **Hyperbolic space**: μ_k is a mixture of *scaled χ distributions.*
- **Space of PSD matrices**: μ_k is a very minor modification of the *Gaussian orthogonal ensemble* from random matrix theory!

Visuals: Kernel and Gaussian Process Regression



Check out the GeometricKernels Python package!